

# Universidad ORT Uruguay

## Facultad de Ingeniería

### Análisis de datos masivos en redes sociales

Entregado como requisito para la obtención del título de  
Licenciatura en Sistemas

Mario Camerota - 117716

Elisa Rodríguez - 151148

Dani Vianna - 152326

Tutor: Helena Garbarino

2015

## Declaración de autoría

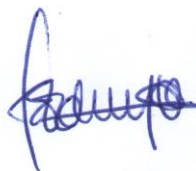
Nosotros, Mario Camerota, Elisa Rodríguez y Dani Vianna, declaramos que el trabajo que se presenta en esta obra es de nuestra propia mano. Podemos asegurar que:

- ✚ La obra fue producida en su totalidad mientras realizábamos el Proyecto Final de Carrera;
- ✚ Cuando hemos consultado el trabajo publicado por otros, lo hemos atribuido con claridad;
- ✚ Cuando hemos citado obras de otros, hemos indicado las fuentes. Con excepción de estas citas, la obra es enteramente nuestra;
- ✚ En la obra, hemos acusado recibo de las ayudas recibidas;
- ✚ Cuando la obra se basa en trabajo realizado conjuntamente con otros, hemos explicado claramente qué fue contribuido por otros, y qué fue contribuido por nosotros;
- ✚ Ninguna parte de este trabajo ha sido publicada previamente a su entrega, excepto donde se han realizado las aclaraciones correspondientes.



Mario Camerota

10/09/2015



Elisa Rodríguez

10/09/2015



Dani Vianna

10/09/2015

## Agradecimientos

Queremos expresar nuestro más sincero agradecimiento a las personas que nos acompañaron a lo largo de nuestro proyecto.

A los integrantes de la Universidad ORT Uruguay, en especial a nuestra tutora Helena Garbarino quien nos guió durante las diferentes etapas del proyecto; a Rosana Fernández del CIE, quien nos orientó en los comienzos para justificar la viabilidad de este viaje y a Pedro Salvetto por el apoyo brindado desde el comienzo del proyecto.

A Sergio Garrido, experto de dominio, por dedicar invaluable horas de su tiempo para darnos su *feedback* y aportar grandes ideas.

A nuestras familias y parejas, quienes quedaron relegadas durante la realización del proyecto y muchas veces teniendo que cambiar sus rutinas para poder darnos espacio para reunirnos y dedicarnos tiempo oficiando de revisores externos de nuestro proyecto.

A todos ellos, muchas gracias.

## Abstract

Hoy día, junto con la explosión de medios sociales tales como blogs, foros y redes sociales, se observa una tendencia creciente y masiva de usuarios de Internet que comparten libremente sus opiniones, ideas, actividades, así como también sus experiencias personales y estados emocionales. Esta información es cada vez más valiosa para las empresas ya que, a través de ésta, se puede obtener una visión sobre el sentimiento que despierta una marca, un producto lanzado al mercado, su reputación, entre otros.

El objetivo de este proyecto surge de la necesidad de la empresa ColectaTe de consolidarse en el mercado; adquiriendo herramientas que asistan a la toma de decisiones en base al análisis masivo de datos capturados en redes sociales y que permitan generar ventajas competitivas, así como captar y fidelizar clientes mediante tecnologías móviles.

El proyecto cuenta con una instancia de investigación de las tecnologías involucradas, que permitió un mejor entendimiento de los conceptos involucrados en *Big Data* y aplicaciones móviles. En base a ello, se desarrollaron una serie de procesos que permiten extraer, analizar, almacenar y visualizar grandes volúmenes de datos, presentando los resultados al usuario en un formato que permita la detección de tendencias y temas que requieran atención inmediata, de forma rápida y sencilla.

En conclusión, se generó una solución que permite obtener información actual y en tiempo real relevada de las redes sociales que es de interés para la empresa, y que le permite detectar tendencias de mercado para cumplir con los objetivos planteados.

## Palabras Clave

A continuación se listan las palabras claves más importantes del sistema, para que el documento pueda ser indexado por los buscadores.

Análisis de Datos Masivos, Análisis de Datos en Redes Sociales, Análisis de Sentimiento, Big Data, Social Media Analysis, Sentiment Analysis

# Índice

1.	Introducción .....	18
1.1.	Descripción del proyecto .....	18
1.1.1.	Propósito .....	18
1.1.2.	Objetivos del proyecto .....	18
1.1.3.	Alcance del proyecto .....	19
1.1.4.	Descripción del equipo .....	20
1.1.5.	Características tecnológicas .....	20
1.1.6.	Experiencia del equipo con la tecnología.....	21
2.	Análisis estratégico de la empresa y su entorno.....	22
2.1.	Análisis del entorno .....	22
2.1.1.	Sector industrial .....	22
2.1.2.	Análisis de las 5 fuerzas de Porter.....	22
2.1.3.	Análisis PESTEL .....	23
2.1.4.	Análisis FODA .....	25
2.2.	Análisis del modelo de negocio .....	26
3.	Especificación de requerimientos .....	27
3.1.	Estrategia de relevamiento .....	27
3.2.	Proceso de ingeniería de requerimientos .....	27
3.3.	Descripción general de la solución .....	27
3.4.	Alcance de la solución .....	28
3.5.	Actores.....	29
3.6.	Requerimientos funcionales.....	30
3.7.	Requerimientos no funcionales.....	31
4.	Diseño Arquitectónico.....	33
4.1.	Descripción general .....	33
4.2.	Atributos de calidad.....	33
4.3.	Tecnologías y herramientas.....	35
4.4.	Requerimientos mínimos de hardware .....	36
4.5.	Vistas de arquitectura.....	36
4.6.	Vista de base de datos.....	39
4.7.	Vistas de procesos .....	42
4.7.1.	Extracción .....	42
4.7.2.	Análisis.....	45
4.7.3.	Reportes .....	50

4.8.	Diseño aplicación móvil .....	55
4.9.	Justificación de la arquitectura y tecnologías .....	56
4.9.1.	Plataforma de Big Data .....	56
4.9.2.	Plataforma de almacenamiento.....	56
4.9.3.	Decisiones de diseño .....	57
5.	Solución Funcional .....	60
5.1.	Introducción.....	60
5.2.	Ejecución de Procesos .....	60
5.3.	Agendado de Procesos .....	61
5.4.	Clasificación de Sentimiento .....	62
5.5.	Análisis Sentimiento .....	63
5.6.	Oportunidades de Mejora .....	63
5.7.	Análisis de Tendencias.....	65
5.8.	Análisis Actualidad.....	66
5.9.	Análisis Sentimiento en Tiempo Real .....	67
5.10.	Aplicación móvil .....	69
6.	Gestión del Proyecto .....	70
6.1.	Metodología .....	70
6.1.1.	Gestión del proyecto .....	70
6.1.2.	Desarrollo de software.....	71
6.2.	Gestión de esfuerzo.....	72
6.2.1.	Estimación .....	72
6.2.2.	Seguimiento esfuerzo.....	72
6.3.	Gestión del cronograma .....	74
6.4.	Gestión de alcance.....	76
6.4.1.	Product Backlog.....	76
6.4.2.	Seguimiento alcance .....	77
6.5.	Gestión de riesgos .....	78
6.5.1.	Identificación de riesgos.....	78
6.5.2.	Análisis cualitativo de riesgos .....	80
6.5.3.	Disparadores, mitigación y contingencia .....	83
6.6.	Gestión de calidad .....	85
6.6.1.	Objetivos de calidad .....	86
6.6.2.	Plan de calidad .....	87
6.6.3.	Tareas y responsabilidades .....	88

6.6.4.	Estándares y guías .....	89
6.6.5.	Métricas de proceso y producto .....	90
6.6.6.	Revisión y plan de auditoría .....	91
6.6.7.	Evaluación y testeo .....	92
6.6.8.	Calidad de datos .....	93
6.6.9.	Resolución de incidentes y acciones correctivas .....	94
6.7.	Gestión de la comunicación .....	94
6.8.	Gestión de configuración.....	95
6.8.1.	Elementos de SCM .....	95
6.8.2.	Herramientas.....	95
6.8.3.	Control de versiones .....	96
6.8.4.	Control de cambios .....	96
6.9.	Gestión de adquisiciones.....	96
6.9.1.	Hardware.....	96
6.9.2.	Software .....	96
7.	Conclusiones, lecciones aprendidas y proyección a futuro .....	98
7.1.	Conclusiones.....	98
7.2.	Lecciones aprendidas .....	98
7.3.	Oportunidades de mejora y líneas de trabajo futuras .....	99
8.	Referencias Bibliográficas .....	101
9.	Anexos.....	105
9.1.	Anexo 1 - Descripción Canvas.....	105
9.2.	Anexo 2 - Investigación de las herramientas.....	110
9.2.1.	Plataformas de Big Data .....	110
9.2.2.	Plataformas de almacenamiento .....	120
9.2.3.	Herramientas de visualización .....	120
9.2.4.	Herramientas de Empaquetado Móvil.....	122
9.3.	Anexo 3 - Cómo KNIME cumple con el proceso de Big Data?.....	124
9.4.	Anexo 4 - Análisis de sentimiento (Sentiment Analysis) .....	126
9.5.	Anexo 5 - Latent Dirichlet Allocation (LDA).....	128
9.6.	Anexo 6 - Solicitud de permisos Facebook.....	129
9.7.	Anexo 7 - Plan de gestión de riesgos.....	132
9.8.	Anexo 8 - Técnica de estimación .....	134
9.9.	Anexo 9 - Product Backlog.....	136
9.10.	Anexo 10 - Estimación Product Backlog .....	148

9.11.	Anexo 11 - Estimación tareas Product Backlog .....	150
9.12.	Anexo 12 - Sprint Review .....	156
9.13.	Anexo 13 - Retrospectiva .....	158
9.14.	Anexo 14 - Estándares de desarrollo .....	162
9.15.	Anexo 15 - Revisiones y mediciones .....	164
9.16.	Anexo 16 - Resultados - Plan de testing .....	169
9.16.1.	Resultados - Pruebas funcionales.....	169
9.16.2.	Resultados - Pruebas de rendimiento .....	176
9.16.3.	Resultados - Pruebas de integridad de datos.....	177
9.17.	Anexo 17 - Encuesta de Satisfacción.....	179
9.18.	Anexo 18 - Descripción de metanodos KNIME .....	181
9.19.	Anexo 19 - Detección de dispositivos .....	202
9.20.	Anexo 20 - Nomenclatura de flujos en Knime .....	203

## Índice de tablas

Tabla 1.1	Integrantes del equipo	20
Tabla 1.2	Experiencia del equipo en las tecnologías	21
Tabla 2.1	Análisis FODA	25
Tabla 3.1	Requisitos de tiempo para relevamiento	27
Tabla 3.2	Alcance de la solución	28
Tabla 3.3	Exclusiones de alcance de la solución	29
Tabla 3.4	Análisis de interesados	29
Tabla 3.5	Requerimientos funcionales big data	30
Tabla 3.6	Requerimientos funcionales aplicación móvil	30
Tabla 3.7	Requerimientos no funcionales big data	31
Tabla 3.8	Requerimientos no funcionales aplicación móvil	32
Tabla 4.1	Atributos de calidad Big Data	34
Tabla 4.2	Atributos de calidad Aplicación Móvil	35
Tabla 4.3	Tecnologías y herramientas	35
Tabla 4.4	Requerimientos de hardware	36
Tabla 4.5	Catálogo de elementos de base de datos	41
Tabla 4.6	Decisión guardar datos	57
Tabla 4.7	Decisión separación procesos	58
Tabla 4.8	Decisión estrategia aplicación móvil	59
Tabla 6.1	Desviación de esfuerzo	74
Tabla 6.2	Cronograma de proyecto	76
Tabla 6.3	Alcance por sprint	76
Tabla 6.4	Prioridad user stories	77
Tabla 6.5	Seguimiento por sprint	77
Tabla 6.6	Referencia Burndown chart	78
Tabla 6.7	Análisis resultados por sprint	78
Tabla 6.8	Riesgos identificados	79
Tabla 6.9	Magnitud de riesgos	81
Tabla 6.10	Estrategías genéricas de riesgos	82
Tabla 6.11	Plan de acción de riesgos	85
Tabla 6.12	Objetivos de calidad proceso y producto	86
Tabla 6.13	Objetivos de calidad big data y aplicación móvil	86
Tabla 6.14	Plan de calidad	88
Tabla 6.15	Tareas y responsabilidades por rol	88
Tabla 6.16	Responsables por artefactos del proyecto	89
Tabla 6.17	Métricas de proceso	90
Tabla 6.18	Métricas de producto	91
Tabla 6.19	Métricas atributos de calidad	91
Tabla 6.20	Plan de revisiones	91
Tabla 6.21	Plan de pruebas de funcionalidad	92
Tabla 6.22	Plan pruebas de rendimiento	92
Tabla 6.23	Plan pruebas de integridad de datos	93
Tabla 6.24	Plan de calidad de datos	94
Tabla 6.25	Plan de comunicación	95
Tabla 7.1	Lecciones aprendidas	99

Tabla 9.1 Comparación plataformas de almacenamiento .....	120
Tabla 9.2 Escala de impacto riesgos.....	132
Tabla 9.3 Escala de probabilidad de ocurrencia riesgos .....	133
Tabla 9.4 Escala de ocurrencia en el tiempo riesgos .....	133
Tabla 9.5 Matriz de estrategias genéricas riesgos .....	133
Tabla 9.6 Estimación product backlog .....	149
Tabla 9.7 Estimación sprint 1 .....	151
Tabla 9.8 Estimación sprint 2 .....	153
Tabla 9.9 Estimación sprint 3 .....	154
Tabla 9.10 Estimación sprint 4 .....	155
Tabla 9.11 Retrospectiva sprint 1.....	158
Tabla 9.12 Retrospectiva sprint 2.....	159
Tabla 9.13 Retrospectiva sprint 3.....	160
Tabla 9.14 Retrospectiva sprint 4.....	161
Tabla 9.15 Resultado revisión sprint 1 .....	164
Tabla 9.16 Resultado medición sprint 1 .....	165
Tabla 9.17 Resultado revisión sprint 2 .....	165
Tabla 9.18 Resultado medición sprint 2.....	166
Tabla 9.19 Resultado revisión sprint 3 .....	166
Tabla 9.20 Resultado medición sprint 3.....	167
Tabla 9.21 Resultado revisión sprint 4 .....	167
Tabla 9.22 Resultado medición sprint 4.....	168
Tabla 9.23 Casos de prueba análisis sentimiento .....	169
Tabla 9.24 Datos y pruebas análisis sentimiento.....	170
Tabla 9.25 Resultado análisis sentimiento .....	170
Tabla 9.26 Casos de prueba oportunidades de mejora (tópicos) .....	171
Tabla 9.27 Datos y pruebas oportunidades de mejora (tópicos).....	171
Tabla 9.28 Tópicos esperados prueba oportunidades de mejora (tópicos) corrida 1 .....	172
Tabla 9.29 Resultados prueba oportunidades de mejora (tópicos) corrida 1 .....	172
Tabla 9.30 Tópicos esperados prueba oportunidades de mejora (tópicos) corrida 2 .....	172
Tabla 9.31 Resultados prueba oportunidades de mejora (tópicos) corrida 2 .....	172
Tabla 9.32 Tópicos esperados prueba oportunidades de mejora (tópicos) corrida 3 .....	173
Tabla 9.33 Resultados prueba oportunidades de mejora (tópicos) corrida 3 .....	173
Tabla 9.34 Datos y pruebas oportunidades de mejora (tag cloud) .....	174
Tabla 9.35 Resultados prueba oportunidades de mejora (tag cloud) .....	174
Tabla 9.36 Casos de prueba análisis actualidad .....	174
Tabla 9.37 Datos y prueba análisis actualidad .....	175
Tabla 9.38 Resultados prueba análisis actualidad .....	175
Tabla 9.39 Datos y pruebas filtrado por ubicación .....	175
Tabla 9.40 Resultados prueba filtrado por ubicación .....	176
Tabla 9.41 Resultados prueba pertinencia de post extraídos.....	176
Tabla 9.42 Casos de prueba rendimiento .....	177
Tabla 9.43 Resultado prueba rendimiento .....	177
Tabla 9.44 Casos de prueba integridad de datos .....	178
Tabla 9.45 Resultados prueba integridad de datos .....	178
Tabla 9.46 Resultados encuesta de satisfacción .....	180

## Índice de Imágenes

Imagen 2.1 Modelo Canvas .....	26
Imagen 2.2 Modelo Canvas .....	26
Imagen 4.1 Vista de módulos procesos Big Data .....	37
Imagen 4.2 Vista ejecución procesos Big Data.....	38
Imagen 4.3 Vista asignación procesos Big Data .....	39
Imagen 4.4 Modelo Entidad Relación .....	40
Imagen 4.5 Flujo Extracción Twitter.....	42
Imagen 4.6 Flujo Extracción Twitter ColectaTe.....	43
Imagen 4.7 Flujo Extracción Twitter Stream .....	44
Imagen 4.8 Flujo Sentiment Analysis .....	45
Imagen 4.9 Flujo Sentiment Analysis Streaming.....	47
Imagen 4.10 Flujo Análisis Actualidad.....	48
Imagen 4.11 Flujo Análisis de Tendencias.....	49
Imagen 4.12 Flujo Análisis Oportunidades de Mejora .....	50
Imagen 4.13 Flujo reporte Análisis de Sentimiento .....	50
Imagen 4.14 Flujo reporte Análisis Actualidad .....	51
Imagen 4.15 Flujo reporte Análisis de Tendencias .....	52
Imagen 4.16 Flujo reporte Análisis de Oportunidades de mejora .....	53
Imagen 4.17 Dataset reporte Análisis de Sentimiento en Tiempo Real .....	54
Imagen 4.18 Arquitectura aplicación móvil .....	55
Imagen 5.1 Login Knime .....	60
Imagen 5.2 Menú procesos servidor Knime.....	61
Imagen 5.3 Iniciar ejecución proceso en Knime .....	61
Imagen 5.4 Menú agendar procesos Knime .....	61
Imagen 5.5 Panel agendar procesos Knime .....	62
Imagen 5.6 Reporte análisis sentimiento gráfica torta.....	63
Imagen 5.7 Reporte análisis sentimiento gráfica barras.....	63
Imagen 5.8 Reporte oportunidades de mejora Tag Cloud.....	64
Imagen 5.9 Reporte oportunidades de mejora gráfica barras.....	64
Imagen 5.10 Reporte oportunidades de mejora tabla .....	65
Imagen 5.11 Reporte análisis de tendencias tag cloud.....	65
Imagen 5.12 Reporte análisis de tendencias gráfica barras .....	66
Imagen 5.13 Reporte análisis de tendencias tabla .....	66
Imagen 5.14 Reporte análisis actualidad gráfica barras .....	67
Imagen 5.15 Reporte análisis actualidad tabla .....	67
Imagen 5.16 Reporte análisis sentimiento en tiempo real - selección reporte.....	68
Imagen 5.17 Reporte análisis sentimiento en tiempo real - gráfica barras y torta .....	68
Imagen 5.18 Reporte análisis sentimiento en tiempo real - datos adicionales.....	69
Imagen 5.19 Reporte análisis sentimiento en tiempo real - tweets.....	69
Imagen 5.20 Menú de usuario ColectaTe .....	69
Imagen 5.21 Pantalla ingreso a ColectaTe .....	69
Imagen 5.22 ColectaTe instalado en un dispositivo móvil.....	69
Imagen 6.1 Ciclo Scrum [16].....	70
Imagen 6.2 Ciclo Big Data [17] .....	71
Imagen 6.3 Ciclo desarrollo ágil [43] .....	71

Imagen 6.4 Gráfica variación esfuerzo por tarea .....	72
Imagen 6.5 Gráfica % esfuerzo por tarea .....	73
Imagen 6.6 Gráfica distribución horas por tarea .....	73
Imagen 6.7 Gráfica desviación sprint backlog.....	74
Imagen 6.8 Cronograma de proyecto .....	75
Imagen 6.9 Burndown Chart .....	77
Imagen 6.10 Gráfica de riesgos .....	82
Imagen 9.1 Sandbox HDP .....	110
Imagen 9.2 HDP cliente web .....	110
Imagen 9.3 Archivos generados por Flume conteniendo datos extraídos de Twitter.....	112
Imagen 9.4 Datos extraídos.....	113
Imagen 9.5 Informatica 1 .....	114
Imagen 9.6 Informatica 2 .....	114
Imagen 9.7 Informatica 3 .....	114
Imagen 9.8 Informatica 4 .....	114
Imagen 9.9 Informatica 5 .....	115
Imagen 9.10 Informatica 6 .....	115
Imagen 9.11 Informatica 7 .....	115
Imagen 9.12 Informatica 8 .....	115
Imagen 9.13 Informatica 9 .....	116
Imagen 9.14 Talend 1 .....	116
Imagen 9.15 Talend 2 .....	117
Imagen 9.16 Talend 3 .....	117
Imagen 9.17 Talend 4 .....	117
Imagen 9.18 Talend 5 .....	118
Imagen 9.19 KNIME Workflow para extracción de datos de Twitter .....	118
Imagen 9.20 Configuración del componente "Twitter API Connector" .....	119
Imagen 9.21 Configuración del componente "Twitter Search" .....	119
Imagen 9.22 Resultado de la extracción .....	119
Imagen 9.23 Conectores KNIME Big Data .....	124
Imagen 9.24 Usuarios Twitter .....	126
Imagen 9.25 Solicitud de permisos Facebook 1 .....	129
Imagen 9.26 Solicitud de permisos Facebook 2 .....	129
Imagen 9.27 Rechazo de permisos Facebook .....	130
Imagen 9.28 Mail IDATHA .....	131
Imagen 9.29 T-Shirt size [44].....	134
Imagen 9.30 Planning Poker [45] .....	134
Imagen 9.31 Template reporte .....	163
Imagen 9.32 Resultados pruebas filtrado por ubicación .....	176
Imagen 9.33 Metanodo Cant Extraídos .....	181
Imagen 9.34 Metanodo Cant Valorados .....	182
Imagen 9.35 Metanodo Conexión MySQL .....	183
Imagen 9.36 Metanodo Extracción Twitter y filtrado.....	183
Imagen 9.37 Metanodo Fecha Ejecución Proceso .....	184
Imagen 9.38 Metanodo Filtrar Hashtags .....	185
Imagen 9.39 Metanodo Filtrar Productos y Servicios .....	186
Imagen 9.40 Metanodo Filtrar Tweets.....	186

Imagen 9.41 Metanodo Generar String Consulta .....	188
Imagen 9.42 Metanodo Generar Tabla Análisis a Guardar.....	188
Imagen 9.43 Metanodo Generar TagCloud.....	189
Imagen 9.44 Metanodo Generar Tópicos .....	190
Imagen 9.45 Metanodo Guardar análisis en BD .....	191
Imagen 9.46 Metanodo Guardar análisis negativo en BD .....	191
Imagen 9.47 Metanodo Guardar análisis positivo en BD .....	192
Imagen 9.48 Metanodo Guardar Cabezal en BD.....	193
Imagen 9.49 Metanodo Guardar Fecha Ultimo Tweet .....	193
Imagen 9.50 Metanodo Guardar Sentimiento en Tweet.....	194
Imagen 9.51 Metanodo Guardar tópicos negativo en BD .....	194
Imagen 9.52 Metanodo Guardar tópicos positivo en BD .....	195
Imagen 9.53 Metanodo Leer Credenciales Twitter.....	196
Imagen 9.54 Metanodo Leer Productos y Servicios.....	196
Imagen 9.55 Metanodo Obtener Cabezal .....	197
Imagen 9.56 Metanodo Obtener Fecha Ultimo Tweet.....	197
Imagen 9.57 Metanodo Obtener Tweets de BD .....	198
Imagen 9.58 Metanodo Ocurrencia de palabras .....	199
Imagen 9.59 Metanodo Preprocesamiento datos .....	199
Imagen 9.60 Metanodo Valoración Sentimiento.....	200

## Glosario

Término	Definición
Apache	Licencia de <i>software</i> libre creada por la <i>Apache Software Foundation</i> (ASF), que requiere la conservación del aviso de copyright y el disclaimer, permitiendo el uso y distribución del código fuente para <i>software</i> libre y <i>software</i> propietario.
API	API (del inglés <i>Application Programming Interface</i> - Interfaz de Programación de Aplicaciones) es un conjunto de especificaciones de comunicación entre componentes <i>software</i> . Se trata del conjunto de llamadas al sistema que ofrecen acceso a los servicios del sistema desde los procesos y representa un método para conseguir abstracción en la programación, generalmente (aunque no necesariamente) entre los niveles o capas inferiores y los superiores del <i>software</i> .
APK	Un archivo con extensión <i>.apk</i> ( <i>Application Package File</i> ) es un paquete para el sistema operativo Android. Este formato es una variante del formato JAR de Java y se usa para distribuir e instalar componentes empaquetados para la plataforma Android para <i>smartphones</i> y <i>tablets</i> .
BD	Base de Datos - Una base de datos es un “almacén” que nos permite guardar grandes cantidades de información de forma organizada para que luego se pueda encontrar y utilizar fácilmente.
BI	<i>Business Intelligence</i> es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios.
<i>Big Data</i>	<i>Big Data</i> se define como el conjunto de herramientas informáticas destinadas a la manipulación, gestión y análisis de grandes volúmenes de datos de todo tipo los cuales no pueden ser gestionados por las herramientas informáticas tradicionales.
Corpus	Conjunto de datos, textos u otros materiales sobre determinada materia que pueden servir de base para una investigación o trabajo
CSS	Es un lenguaje utilizado en la presentación de documentos HTML. Un documento HTML es llamado coloquialmente “una página web”. Es decir que el lenguaje CSS sirve para organizar la presentación y aspecto de una página web.
CSV	Los ficheros CSV (del inglés <i>comma-separated values</i> ) son un tipo de documento sencillo para representar datos en forma de tabla, en las que las columnas se separan por comas (o punto y coma en donde la coma es el separador decimal) y las filas por saltos de línea.

DataSet	Conocido también como un conjunto de datos, es una colección de datos habitualmente tabulada. En general y en su versión más simple, un conjunto de datos corresponde a los contenidos de una única tabla de base de datos o una única matriz de datos estadística, donde cada columna de la tabla representa una variable en particular, y cada fila representa a un miembro determinado del conjunto de datos en cuestión.
ETL	Los procesos ETL son un término estándar que se utiliza para referirse al movimiento y transformación de datos. Se trata del proceso que permite mover datos desde múltiples fuentes, reformatearlos y cargarlos en otra base de datos con el objetivo de analizarlos.
FODA	El Análisis FODA es una metodología de estudio de la situación competitiva de una empresa o solución dentro de su mercado y de las características internas de la misma, a efectos de determinar sus Debilidades, Amenazas, Fortalezas y Oportunidades.
Frameworks	<i>Framework</i> , o “marco de trabajo” define, en términos generales, un conjunto estandarizado de conceptos, prácticas y criterios para enfocar un tipo de problemática particular que sirve como referencia, para enfrentar y resolver nuevos problemas de índole similar.
Hashtags	Una etiqueta o <i>hashtag</i> es una cadena de caracteres formada por una o varias palabras concatenadas y precedidas por una almohadilla o numeral (#).
HDP	<i>Hortonworks Data Platform</i> es una plataforma para manejo de datos empresariales que permite el almacenamiento, procesamiento y análisis de grandes volúmenes de datos en diferentes formatos y provenientes de distintas fuentes.
HTML	<i>Hypertext Markup Language</i> es el lenguaje que se emplea para el desarrollo de páginas de internet. Está compuesto por una serie de etiquetas que el navegador interpreta y da forma en la pantalla.
ISP	<i>Internet Service Provider</i> o proveedor de acceso a Internet, es una compañía que ofrece acceso a Internet. Normalmente, la conexión con el ISP tiene lugar a través de una conexión de acceso telefónico (línea telefónica) o una conexión de banda ancha.
LDA	<i>Latent Dirichlet Allocation</i> , es un modelo probabilístico generativo para las colecciones de datos discretos como corpus de textos.
Metanodo	Conjunto de nodos de Knime agrupados para facilitar la visualización de un proceso.
ODBC	<i>Open DataBase Connectivity</i> (ODBC) es un estándar de acceso a las bases de datos. El objetivo, es hacer posible el acceso a cualquier dato desde cualquier aplicación, sin importar qué sistema de gestión de bases de datos (DBMS) almacene los datos.

<i>Open Source</i>	Es el término inglés con el que se conoce al <i>software</i> desarrollado y cuyo código distribuido libremente.
PB	<i>Product Backlog</i>
PDF	Acrónimo inglés de <i>Portable Document Format</i> . Se trata de un formato de almacenamiento de documentos.
PESTEL	Es un instrumento de planificación estratégica para definir el contexto de una campaña, emprendimiento o proyecto. Analiza factores externos políticos, económicos, sociales, tecnológicos, ambientales y jurídicos que pueden influir.
PHP	PHP ( <i>Hypertext Preprocessor</i> ) es un lenguaje de código abierto muy popular especialmente adecuado para el desarrollo web y que puede ser incrustado en HTML.
REST	La Transferencia de Estado Representacional ( <i>Representational State Transfer</i> ) es un estilo de arquitectura software para sistemas hipermedia distribuidos como la <i>World Wide Web</i> .
SCM	La gestión de configuración de <i>software</i> (SCM) es una actividad de auto protección que se aplica a lo largo del proceso de la Ingeniería de <i>Software</i> y que tiene como objetivo identificar, organizar y controlar las modificaciones que sufre el <i>software</i> que construye un equipo.
SP	<i>Story Points</i>
SQA	SQA es un conjunto de actividades sistemáticas y planeadas para asegurar que los procesos y productos de <i>software</i> cumplen con los requerimientos, estándares y procedimientos.
<i>Stream</i>	Es la transmisión continua de datos que es consumida en paralelo mientras se descarga. La palabra <i>streaming</i> se refiere a una corriente continuada, que fluye sin interrupción.
<i>Tag Cloud</i>	Una nube de palabras o nube de etiquetas es una representación visual de las palabras que conforman un texto, en donde el tamaño es mayor para las palabras que aparecen con más frecuencia.
URL	Un localizador de recursos uniforme o URL - siglas en inglés <i>deuniform resource locator</i> .
US	<i>User Stories</i> .
XLS	Extensión de archivo por defecto del formato Microsoft Excel 2003 o versiones anteriores.
XML	Acrónimo en inglés de <i>eXtensible Markup Language</i> ("lenguaje de marcas extensible"), es un metalenguaje extensible de etiquetas desarrollado por el <i>World Wide Web Consortium</i> (W3C).

# 1. Introducción

## 1.1. Descripción del proyecto

### 1.1.1. Propósito

ColectaTe es una empresa creada en 2014 cuyo objetivo es brindar servicios para organizar y participar en colectas y/o colectivos a través de internet.

Actualmente cuentan con un sitio web, mediante el cual ofrecen la posibilidad de crear, organizar y gestionar colectas de forma práctica y segura, así como también realizar depósitos en dichas colectas de forma *online* (mediante la modalidad de pagos electrónicos). Permite también invitar a participar y compartir colectas a través de distintas redes sociales.

Si bien la empresa cuenta con una primer versión 100% funcional del *software* de gestión, necesita:

- ✚ Nuevas funcionalidades para una segunda versión del sitio
- ✚ Herramientas que asistan en la toma de decisiones a los directores de la empresa, respondiendo a las siguientes preguntas mediante el análisis de datos masivos capturados en las redes sociales [1]:
  - ✓ Qué productos/servicios son los más consumidos
  - ✓ Qué productos/servicios se pueden ofrecer conjuntamente como venta cruzada
  - ✓ Qué reputación tiene la empresa para con los clientes
  - ✓ De qué manera se puede mejorar el servicio
- ✚ Componente móvil que permita acceder a las funcionalidades necesarias a los clientes y participantes de las colectas.

### 1.1.2. Objetivos del proyecto

Por ser ColectaTe un emprendimiento reciente, requiere de herramientas que le faciliten consolidarse en el mercado mediante:

- ✚ Identificación de nichos de mercado sólidos
- ✚ Fidelización de clientes
- ✚ Diferenciarse de competidores y productos sustitutos

El proyecto se enfoca en construir herramientas que le permitan a la empresa tomar decisiones para alcanzar los objetivos arriba mencionados. Estas herramientas se van a enfocar en los siguientes aspectos:

#### Optimizar estrategia de contenidos

Determinar nuevas oportunidades de negocio captando y almacenando datos de las redes sociales: analizar de qué se habla y qué productos/servicios se consumen más seguido de manera de optimizar la venta cruzada, prediciendo cuales serían más apropiados para los clientes y ofreciendo una mejor combinación de los mismos.

Analizar los círculos de los clientes en redes sociales para poder sugerir realizar colectas en el sitio web de la empresa.

### Venta directa y atención al cliente

Establecer relaciones de confianza con el cliente analizando datos de conversaciones referentes a la empresa, de manera de tener un mejor manejo de las quejas/disconformidades relativas al servicio y poder así mejorarlo. Determinar causas de fracaso o problemas.

### Visualización de los datos

Ofrecer el despliegue de datos a partir de fuentes de información. Presentar los resultados al usuario en un formato que permita la detección de tendencias y temas que requieran atención inmediata, de forma rápida y sencilla. Se buscará resolver el reto de crear una componente visual para la ejecución estratégica.

### Aplicación móvil

Desarrollo de un componente móvil que permita acceder a las funcionalidades necesarias a los clientes y participantes de las colectas.

## **1.1.3. Alcance del proyecto**

En lo que refiere a *Big Data* [1] (procesamiento de grandes volúmenes de información con el fin de obtener un beneficio comercial):

- ✚ Investigación y selección de las distintas herramientas de procesamiento de grandes volúmenes de datos disponibles en el mercado
- ✚ Investigación de los *frameworks*, *APIs* y herramientas ofrecidas por las diferentes redes sociales para la extracción de sus datos
- ✚ Investigación y selección de diferentes alternativas para la persistencia de datos
- ✚ Diseño e implementación de los procesos de *Big Data* requeridos para alcanzar los objetivos planteados
- ✚ Ejecución de los procesos y generación de los resultados analizados a ser entregados a ColectaTe en formato de reporte. Dichos reportes se entregarán durante la ejecución del proyecto, culminando el compromiso junto a la fecha de finalización del mismo
- ✚ Generación de documentación funcional y técnica a ser entregada a ColectaTe (apartados de Especificación de Requerimientos y Diseño Arquitectónico del presente documento)

En cuanto a la aplicación móvil:

- ✚ Investigación de las diferentes alternativas para el desarrollo de la aplicación móvil, seleccionando una y justificándola en base a los requerimientos relevados
- ✚ Desarrollo y pruebas de la aplicación móvil a ser entregada

- ✚ Generación de documentación funcional y técnica a ser entregada a ColectaTe (apartados de Especificación de Requerimientos y Diseño Arquitectónico del presente documento)

#### 1.1.4. Descripción del equipo

El equipo de proyecto se forma por 3 estudiantes que se encuentran cursando el último año de la carrera Licenciatura en Sistemas de la Universidad ORT Uruguay.

Los integrantes del equipo son:

Nombre y Apellido	N° de Estudiante	Rol
Mario Camerota	117716	Ingeniero de Requerimientos, Arquitecto
Elisa Rodríguez	151148	Responsable de SQA
Dani Vianna	152326	Gerente de Proyecto, SCM

Tabla 1.1 Integrantes del equipo

Si bien se asignaron roles entre los integrantes del equipo, las tareas se realizaron mayoritariamente entre todos los integrantes y el resultado de cada una de ellas fue revisado y avalado por cada uno de los integrantes para su presentación.

La asignación de roles se realizó con el único objetivo de reducir la posibilidad de dejar tareas sin hacer por problemas de comunicación o entendimiento entre los integrantes del equipo.

Los roles de desarrollador y *tester* no se asignan a ningún integrante como tal, ya que las responsabilidades y tareas intrínsecas a los mismos fueron realizadas por todos los integrantes de igual forma.

#### 1.1.5. Características tecnológicas

El sistema web actual de ColectaTe se encuentra desarrollado en HTML5, php y Javascript. Se utiliza MySQL como motor de base de datos.

El hosting del sitio se encuentra en <http://www.hostinger.es/>, el cual provee tres ambientes separados, desarrollo, *testing* y producción, contando todos ellos con bases de datos y aplicaciones independientes.

El primer desafío consiste en investigar y definir un sistema *Big Data* [1]. Este proceso consta de varias etapas: captación de los datos, almacenamiento, análisis y visualización [2] [3].

Para ello se seleccionó la plataforma Knime, como sistema para diseñar y ejecutar los procesos de *Big Data* junto a MySQL como plataforma de persistencia de datos. En el Anexo 2 - Investigación de las herramientas se encuentra una descripción detallada sobre la investigación realizada de las herramientas de *Big Data* y bases de datos, mientras que en la sección 4.9 - Justificación de la arquitectura y tecnologías se encuentra detallado el motivo de la elección de cada una de ellas.

El segundo desafío consiste en generar una aplicación móvil para las diferentes plataformas móviles que le permita a ColectaTe obtener una ventaja competitiva, así como captar y retener clientes.

### 1.1.6. Experiencia del equipo con la tecnología

La experiencia de los integrantes del equipo a nivel tecnológico se conforma de la siguiente manera:

	<b>Mario Camerota</b>	<b>Elisa Rodríguez</b>	<b>Dani Vianna</b>
<i>Big Data</i>	Nulo	Nulo	Nulo
Base de datos	Medio	Medio	Medio
Desarrollo web	Medio	Bajo	Bajo
Desarrollo móvil	Nulo	Nulo	Nulo

Tabla 1.2 Experiencia del equipo en las tecnologías

## 2. Análisis estratégico de la empresa y su entorno

### 2.1. Análisis del entorno

El análisis del entorno se realizó mediante las 5 fuerzas de Porter [4] y el análisis PESTEL [5].

#### 2.1.1. Sector industrial

El sector industrial correspondiente es el sector servicios. A su vez, dentro de este sector, la empresa se enfoca en el sector financiero.

ColectaTe se encuentra en este sector ya que se relaciona de forma directa con bancos (Santander, BROU, BBVA, entre otros), pasarelas de pagos (MVDCobros), clientes y usuarios finales.

#### 2.1.2. Análisis de las 5 fuerzas de Porter

##### 2.1.2.1. Poder de Negociación de los Clientes

- ✚ Los “costos de cambio” son bajos, para el cliente (organizadores de colectas que tienen la necesidad de recaudar dinero) no tiene un costo significativo realizar diferentes colectas en diferentes proveedores de este tipo de servicio. Existe un costo mínimo desde el punto de vista de funcionalidades ofrecidas, ya que cada web es diferente y ofrece más o menos información (también puede realizarse de forma artesanal), pero concretamente el costo es bajo.
- ✚ Dentro de la “diferenciación de insumos”, la diferenciación del servicio ofrecido es importante y diferente entre los distintos proveedores del servicio.
- ✚ Un punto importante es que, a pesar de los puntos anteriores, el cliente no tiene muchas opciones para elegir ya que existen pocos proveedores; particularmente en Uruguay no hay en este momento.

Esto hace que, el “Poder de Negociación de los Clientes” sea **Bajo**.

##### 2.1.2.2. Amenaza de Nuevos Ingresos

- ✚ Las “barreras de entrada” son bajas. No se necesita de una gran inversión inicial. Por lo tanto, es bajo el costo de implementación.
- ✚ No es necesario poseer tecnologías inaccesibles (o complicadas de obtener), ya que se trata de un sitio web y una aplicación móvil las cuales son económicamente accesibles y técnicamente sencillas de desarrollar en base a la disponibilidad de la oferta actual.
- ✚ No existe en la actualidad una “Identidad en la Marca” reconocida en el mercado.

Esto hace que, la “Amenaza de Nuevos Ingresos sea **Alta**”.

### 2.1.2.3. Poder de Negociación de los Proveedores

- ✚ Según datos recabados y analizados en Uruguay, no existe una Empresa que brinde un servicio igual.
- ✚ En otros países de la región, como Argentina, Brasil, Chile, Puerto Rico y México hay muy pocas Empresas (en la actualidad) ofreciendo y brindando el servicio.
- ✚ Otro factor importante es la escasez de los mismos. Por ejemplo las “pasarelas de pago” son muy pocas y todas imponen condiciones difíciles de negociar.
- ✚ En la actualidad para los proveedores ColectaTe no es una empresa clave en cuanto a la rentabilidad obtenida por ellos.

Esto hace que, el “Poder de Negociación de los Proveedores sea **Alto**”.

### 2.1.2.4. Amenaza de Productos Sustitutos

- ✚ Un factor importante es el beneficio que ofrece el servicio. Se diferencia de forma específica sobre productos sustitutos existentes como por ejemplo “hacer un regalo”.
- ✚ La dificultad de obtener la lealtad de los usuarios y clientes, en definitiva lograr en una primera etapa fidelización, generar confianza y prestigio.

Esto hace que, la “Amenaza de Productos Sustitutos sea **Alta**”.

### 2.1.2.5. Intensidad de la Rivalidad

- ✚ En la actualidad, existe competencia:
  - ✓ Redes de cobranza (físicas)
  - ✓ Organizadores y recolectores artesanales
- ✚ Hay una alta amenaza de productos sustitutos y alta amenaza de nuevos ingresos.
- ✚ Existe una baja posibilidad de negociación con proveedores, y esto puede desmotivar el ingreso de competencia.
- ✚ Si en un mediano plazo ingresa más rivalidad, los márgenes de ganancia deberán disminuir.

Esto hace que, la “Intensidad de la Rivalidad en la actualidad sea **Media**”.

## 2.1.3. Análisis PESTEL

### 2.1.3.1. Político

En este punto, no se encontraron aspectos relevantes (estabilidad, política fiscal, reglamentos, políticas sociales). Nada de lo mencionado anteriormente afecta el caso de estudio.

### 2.1.3.2. Económico

En el aspecto económico, teniendo en cuenta el sector donde se desarrollará la investigación lo más relevante es que:

- ✚ La política monetaria se mantiene estable en Uruguay y en países de nuestro interés (en principio, países de América Latina) [6].
- ✚ La inflación entre 2013 y 2015 en Uruguay se encuentra entre un 8% y 9% [7] [8].
- ✚ La tasa de desempleo en América Latina ha bajado fuertemente. De todas maneras, en la actualidad (mayo de 2015) ha subido, pero no de forma significativa [9].

### 2.1.3.3. Social

Los países de América Latina tienen costumbres similares (ejemplo: Argentina, Chile y Brasil). A modo de ejemplo, si bien en base a averiguaciones se detectó que en Argentina no existe un lugar donde “crear” una colecta para un matrimonio, a diferencia de Uruguay donde existen las redes de cobranza; sí se recauda el dinero de otra forma.

### 2.1.3.4. Tecnológico

En el aspecto tecnológico, es importante destacar el crecimiento de la telefonía móvil y de *smartphones* (aplicaciones para iPhone de la empresa Apple) y Android (sistema operativo de Google) de manera muy fuerte.

### 2.1.3.5. Ecológico

En este punto, no han encontrado aspectos relevantes que interfieran en el estudio (leyes de protección, residuos, consumo de energía).

### 2.1.3.6. Legal

Desde el punto de vista Legal, no hay inconvenientes mayores en el sector que se está estudiando. La empresa cuenta con el aval del Banco Central del Uruguay especificando la “no regulación” por el tipo de negocio. Desde el punto de vista de la “Salud y seguridad” todos los involucrados cumplen los requisitos y requerimientos necesarios.

## 2.1.4. Análisis FODA

Análisis Interno	
Fortalezas	Debilidades
<ul style="list-style-type: none"> <li>+ Innovación (nueva forma de hacer colectas)</li> <li>+ Practicidad</li> <li>+ Comodidad</li> <li>+ Seguridad</li> <li>+ Ahorro de tiempo</li> <li>+ Aprovechar el uso de las tarjetas de débito gracias a la ley de inclusión financiera</li> <li>+ Colaboración “espontánea”</li> </ul>	<ul style="list-style-type: none"> <li>+ Nueva forma de hacer “Colectas”</li> <li>+ Inconvenientes para escalar (inversión)</li> <li>+ Difícil para que los usuarios “confíen”</li> <li>+ Desconocimiento de la marca</li> <li>+ Lograr prestigio demostrando diferencias</li> <li>+ Poder abarcar todo lo que hay que hacer (tecnológicamente, estratégicamente, documentación, entre otros) siendo muy pocas personas las involucradas</li> </ul>
Análisis Externo	
Oportunidades	Amenazas
<ul style="list-style-type: none"> <li>+ No existe un servicio igual en Uruguay</li> <li>+ Integración con empresas y servicios como WOOW y Cluval para sugerir que comprar luego de realizada la colecta, o para enviar a su sitio al beneficiario</li> <li>+ Integración con empresas como WOOW y Cluval para que desde su propio sitio, puedan ofrecer como medio de pago a “ColectaTe” y crear una colecta de forma automática para comprar lo seleccionado</li> <li>+ Integración con empresas como PencaDeportiva para ofrecer un servicio más integrado y “colaborativo” entre las empresas</li> <li>+ Ofrecer una nueva vía de colaboración a organizaciones benéficas que permita llegar a más personas</li> <li>+ Expandir y ampliar el segmento de mercado pudiendo salir con el servicio del país gracias a lo escalable que es</li> </ul>	<ul style="list-style-type: none"> <li>+ La simplicidad de la solución, dado que es fácilmente replicable</li> <li>+ El porcentaje de comisión</li> <li>+ El “cambio” cultural que implica realizar la transición de colectas artesanales a colectas electrónicas</li> <li>+ La posibilidad de que empresas ya existentes incluyan el servicio</li> </ul>

Tabla 2.1 Análisis FODA

## 2.2. Análisis del modelo de negocio

En la Imagen 2.1 se muestra el modelo de negocios de ColectaTe. Ver Anexo 1 - Descripción Canvas para una explicación detallada del modelo de negocios. Cada color representa una relación directa entre las diferentes secciones del Canvas, siendo el color verde común a todos.

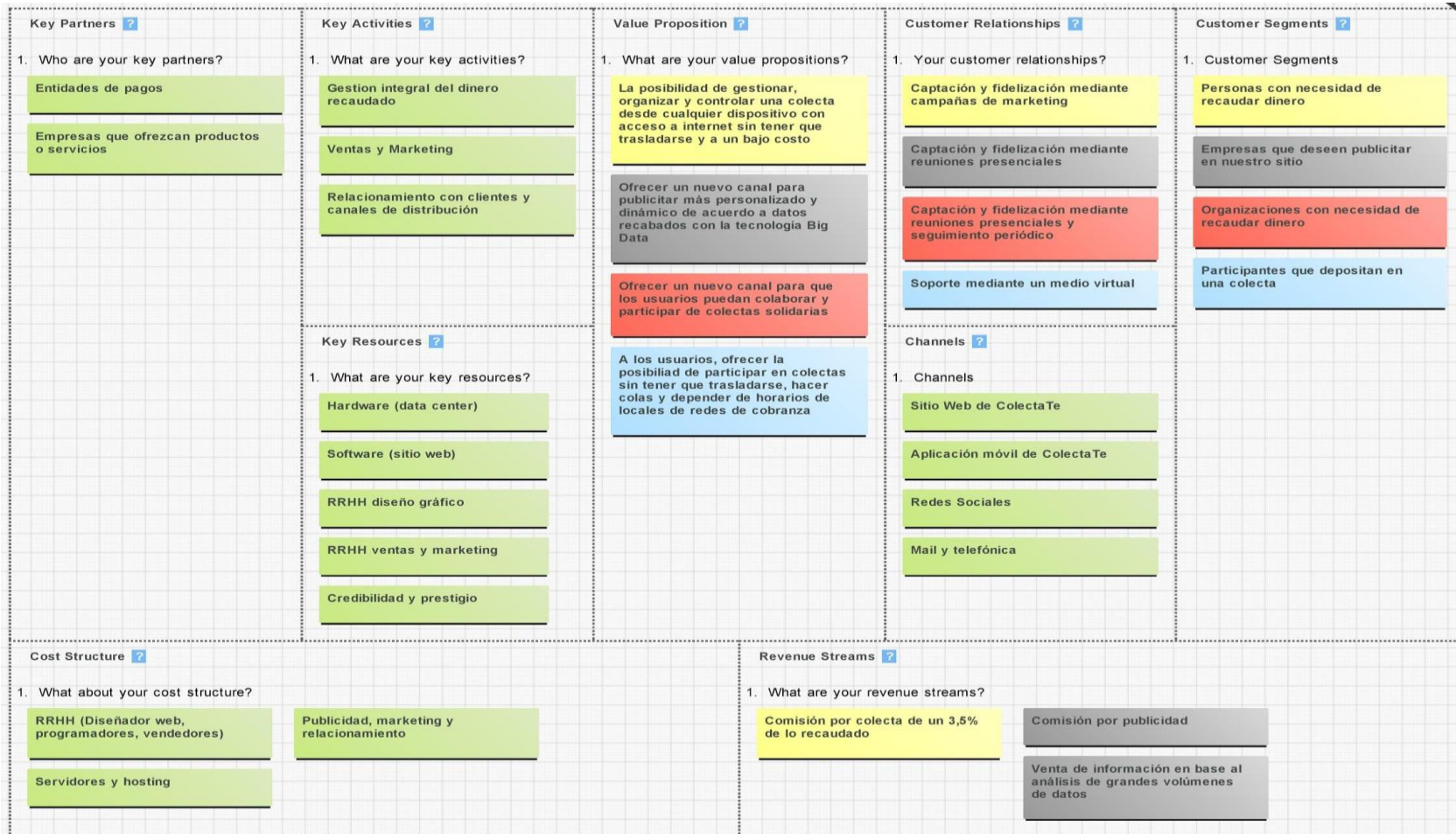


Imagen 2.1 Modelo Canvas

## 3. Especificación de requerimientos

### 3.1. Estrategia de relevamiento

En base a lo acordado, se coordinó una estrategia para poder realizar el relevamiento de forma eficaz y conveniente para ambas partes. En primera instancia, se agendaron entrevistas presenciales con el experto del negocio, las mismas se establecieron con un duración estimada que va entre los 30 minutos y 1 hora.

Fecha reunión	Tema	Duración
4/5/2015	Requerimientos de proceso de <i>Big Data</i>	60 minutos
8/5/2015	Requerimientos de proceso de <i>Big Data</i>	60 minutos
11/5/2015	Requerimientos de aplicación móvil	45 minutos
25/5/2015	Validación y cierre de alcance	30 minutos

Tabla 3.1 Requisitos de tiempo para relevamiento

Para la realización de las mismas, se cuenta con las preguntas a realizar en cada una de ellas, ya elaboradas intentando maximizar la calidad y cantidad de información obtenida para el correcto entendimiento y relevamiento de la información.

Se utilizó el correo electrónico como herramienta para la evacuación de dudas puntuales luego de realizadas las reuniones.

### 3.2. Proceso de ingeniería de requerimientos

El proceso de ingeniería de requerimientos se basó en la metodología ágil Scrum (ver 6.1.1 Gestión del proyecto) y consistió de varias etapas. La primera fue el relevamiento (explicado en el punto anterior). De acuerdo a los objetivos planteados, se acordaron una serie de entrevistas para aclarar y comprender el alcance real de lo solicitado. En paralelo, se documentaron los requerimientos y se comenzó a confeccionar el *Product Backlog* (PB). Es decir, se comenzó a identificar las *User Stories* (US), inicialmente en forma de *Epics* (las cuales se corresponden con los requerimientos funcionales) con sus condiciones de satisfacción, detallando más en profundidad las US a realizar en el *Sprint* próximo.

La etapa de validación de requerimientos tiene un alto grado de interacción con el experto del negocio. Con él, a través de reuniones virtuales y consultas vía mail, se realizó el armado y validación del PB. Para realizar un cierre formal del mismo, se agendó una reunión con el experto del negocio donde se revisó formalmente el PB y se acordó el mismo como alcance del proyecto.

### 3.3. Descripción general de la solución

La solución se compone de dos macro-entregables. Por un lado, la aplicación móvil y por otro lado, la implementación de los procesos de análisis de datos (*Big Data*).

La aplicación móvil facilitará a los usuarios de ColectaTe la gestión de las colectas y la participación en las mismas. Para ello, se requiere que la aplicación cuente con las mismas funcionales del sitio web actual de la empresa.

El sitio web ofrece opciones relacionadas a:

- ✚ la creación y gestión de colectas desde el punto de vista de los usuarios finales
- ✚ la realización de depósitos en colectas
- ✚ la gestión y administración desde el punto de vista de la empresa

Por otra parte, la implementación de los procesos de análisis de datos (*Big Data*) le permitirá a la empresa analizar el mercado e identificar tendencias a través de las cuales podrá realizar nuevos negocios, generar valor agregado para los usuarios ofreciendo productos y servicios de interés e identificar oportunidades de mejora para ofrecer un servicio de mayor calidad.

Para esto es necesario:

- ✚ Extraer el flujo de datos de las redes sociales
- ✚ Analizar los datos extraídos para:
  - ✓ detectar los productos y servicios más populares
  - ✓ entender la visión de los clientes con respecto a la empresa
- ✚ Presentar los datos de una forma clara y amigable facilitando la toma de decisiones

### 3.4. Alcance de la solución

Se encuentra dentro del alcance del proyecto:

Sistema	Alcance
Aplicación móvil	Creación y gestión de colectas desde el punto de vista de los usuarios finales
	Realización de depósitos en colectas
<i>Big Data</i>	Extraer el flujo de datos de las redes sociales (Facebook y Twitter)
	Análisis de los datos extraídos
	Presentación de los datos

Tabla 3.2 Alcance de la solución

Queda fuera del alcance del proyecto:

Sistema	Alcance
Aplicación móvil	Implementación de funcionalidad para la gestión y administración desde el punto de vista de la empresa

	Disponibilizar la aplicación en las tiendas de productos de las plataformas móviles
	Soporte post-entrega
<i>Big Data</i>	Conexión con otras redes sociales diferentes a Facebook y Twitter
	Instalación y configuración de servidores para la ejecución de los procesos en <i>Big Data</i>
	Licencias de plataforma de <i>Big Data</i> para la ejecución en producción de los procesos
	Desarrollo de aplicación(es) para la parametrización de los procesos de <i>Big Data</i>
	Soporte post-entrega

Tabla 3.3 Exclusiones de alcance de la solución

### 3.5. Actores

Se identificaron los actores detallados a continuación para los cuales se considera que pueden ser afectados tanto de forma positiva como negativa por las actividades realizadas en el proyecto.

Para cada interesado se identificó el grado de interés y poder en base a una escala de 3 niveles: alto, medio y bajo.

Interesado	Interés	Poder
Experto del negocio / ColectaTe	Alto	Alto
Equipo de proyecto	Alto	Alto
Tutor	Alto	Alto
Organizaciones creadoras de colectas solidarias	Medio	Bajo
Empresas con las que ColectaTe tiene convenio	Medio	Bajo
Usuarios creadores de colectas	Medio	Bajo
Usuarios depositantes de colectas	Bajo	Bajo
Medios de pago	Bajo	Bajo
Empresa proveedora de plataforma de <i>Big Data</i>	Bajo	Medio
Plataforma de distribución (Apple store, Google play)	Bajo	Medio

Tabla 3.4 Análisis de interesados

### 3.6. Requerimientos funcionales

La lista de requerimientos funcionales se divide en dos grupos, los relacionados con *Big Data* y los relacionados con la aplicación móvil. Ver Anexo 9 - Product Backlog para el detalle de los requerimientos funcionales.

#### 3.6.1.1. Big Data

ID	Requerimiento
RF-BD01	Proveer un mecanismo que permita analizar cuáles son los productos y servicios más populares en las redes sociales de los usuarios de Uruguay a partir de los productos ya ofrecidos por las empresas en el mercado
RF-BD02	Proveer un mecanismo que permita identificar las tendencias de productos y servicios que actualmente no son ofrecidos por ColectaTe y que podrían ser una nueva oportunidad de negocio
RF-BD03	Proveer un mecanismo que permita conocer la visión de los clientes respecto a la calidad del servicio ofrecido por ColectaTe en base al <i>sentiment analysis</i> (análisis de sentimiento) en las redes sociales
RF-BD04	Proveer un mecanismo que permita determinar las oportunidades de mejora en base al <i>sentiment analysis</i> (análisis de sentimiento) en las redes sociales

Tabla 3.5 Requerimientos funcionales big data

#### 3.6.1.2. Aplicación móvil

ID	Requerimiento
RF-AM01	<p>Se requiere una aplicación móvil que tenga las mismas funcionalidades que el sitio web <a href="http://www.colectate.com.uy">www.colectate.com.uy</a>:</p> <ul style="list-style-type: none"><li>✚ Registro y <i>login</i> de usuario</li><li>✚ Abrir una colecta</li><li>✚ Depositar en una colecta</li><li>✚ Invitar a participar</li><li>✚ Retirar fondos</li><li>✚ Compartir en redes sociales (Facebook y Twitter)</li><li>✚ Página principal con información sobre colectas activas</li><li>✚ Poder visualizar detalle de colectas creadas</li><li>✚ Poder visualizar detalle de depósitos realizados en una colecta</li><li>✚ Poder visualizar detalle de retiros realizados</li><li>✚ Gestionar perfil de usuario</li></ul>

Tabla 3.6 Requerimientos funcionales aplicación móvil

### 3.7. Requerimientos no funcionales

La lista de requerimientos no funcionales se divide en dos grupos, los relacionados con *Big Data* y los relacionados con la aplicación móvil.

#### 3.7.1.1. Big Data

ID	Categoría	Requerimiento
RNF-BD01	General	La solución debe cubrir todo el ciclo de vida de <i>Big Data</i> (ver sección 6.1.2 - Desarrollo de software para una descripción detallada del ciclo de vida), siendo valorada la utilización de una herramienta que cubra el ciclo completo
RNF-BD02	Portabilidad	La solución y herramientas deben ejecutarse y funcionar en entorno Windows 7 o superior de 32 y 64 bits
RNF-BD03	Eficiencia	La ejecución completa del proceso (contemplando la extracción de datos, análisis y generación de reportes) no debe superar las 2 horas sin tomar en cuenta las demoras o tiempos por intervención humana  Se toma como carga normal de extracción un máximo de 18 mil tweets y como carga normal del proceso de análisis un máximo de 126.000 tweets
RNF-BD04	Disponibilidad	El sistema debe estar disponible un 90% sobre el tiempo total definido como 24 horas por 7 días a la semana
RNF-BD05	Escalabilidad	El sistema debe ser capaz de soportar aumentos en el volumen de datos sin mayor impacto en la eficiencia. Se debe asegurar que es capaz de soportar el doble de carga de datos de lo normal sin errores críticos y sin superar el umbral de eficiencia definido en el RNF-BD03  Tomando como carga normal de datos los volúmenes definidos en el RNF-BD03, se toma como doble de carga de datos el valor 252.000 tweets
RNF-BD06	Usabilidad	La presentación de los datos debe ser de forma amigable y que facilite su interpretación para la toma de decisiones

Tabla 3.7 Requerimientos no funcionales big data

### 3.7.1.2. Aplicación móvil

ID	Tipo	Requerimiento
RNF-AM01	Portabilidad	La aplicación deberá correr en dispositivos móviles con sistema operativo Android 4.0, con pantalla de 4 pulgadas en adelante
RNF-AM02	Seguridad	La solución debe contar con la habilidad de resistir intentos de uso no autorizados mientras se continúa ofreciendo los servicios a los usuarios legítimos, permitiendo solo operar a usuarios registrados (el usuario tiene una cuenta creada en el sistema) y autenticados (el usuario ingresó su usuario y contraseña en el <i>login</i> y los datos fueron validados por el sistema)
RNF-AM03	Eficiencia	Toda acción que se realice en la aplicación móvil y que sea interna al sistema no debe superar los 3 segundos
RNF-AM04	Disponibilidad	La aplicación debe estar disponible y funcionando un 99,5% sobre el tiempo total definido como 24 horas por 7 días a la semana
RNF-AM05	Interoperabilidad	La aplicación debe ser capaz de intercambiar información con los medios de pagos correctamente y que las funcionalidades donde estos intervienen se ejecuten de forma satisfactoria desde el dispositivo móvil
RNF-AM06	Usabilidad	El sistema debe ofrecer facilidad de uso y aprendizaje, debe ser intuitivo y atractivo para el usuario, el acceso a las funcionalidades debe ser de fácil ubicación de componentes y contar con la ayuda correspondiente.

Tabla 3.8 Requerimientos no funcionales aplicación móvil

## 4. Diseño Arquitectónico

### 4.1. Descripción general

A nivel de diseño, la estrategia se divide en dos grupos:

- ✚ *Big Data*
- ✚ Aplicación móvil

Para el diseño arquitectónico de los procesos de *Big Data* se optó por un diseño en 3 capas (datos, lógica, presentación), en un entorno cliente-servidor.

La capa de datos, la lógica y parte de la capa de presentación (donde se generan los reportes) se ejecutan del lado del servidor. Mientras que la visualización de los reportes se ejecuta del lado del cliente, ya sea en un entorno web o de escritorio (visualización de los reportes en formato pdf a modo de ejemplo).

Ver apartado “Vistas de la arquitectura” de esta sección para más detalles y diagrama de arquitectura.

En cuanto a la aplicación móvil se optó por realizar el empaquetado del sitio web para el sistema operativo Android, lo que permite rápidamente contar con una versión funcional reutilizando el código existente del sitio web “*responsive*” con el cual ya contaba ColectaTe. Ver apartado “Diseño Aplicación móvil” de esta sección para más detalles y diagrama de arquitectura.

### 4.2. Atributos de calidad

Con el fin de que la solución satisfaga de manera exitosa los requerimientos funcionales y no funcionales establecidos, el diseño arquitectónico fue realizado conjuntamente con el análisis de los atributos de calidad involucrados [10].

El análisis realizado de cada uno de los atributos, sentó las bases para las posteriores decisiones de arquitectura.

A continuación se describen los principales atributos considerados en la solución tanto para los procesos de *Big Data* como para la aplicación móvil.

#### 4.2.1.1. Big Data

Atributo	Estrategia
Portabilidad	Se seleccionó la plataforma Knime, la cual se puede ejecutar tanto en entorno Windows XP (o superior) o Linux de 32 o 64 bits.  La plataforma StyleScope Agile se ejecuta en entorno Windows XP (o superior) de 32 o 64 bits.

Eficiencia	<p>El diseño modular de la solución permite ejecutar los diferentes componentes en diferentes servidores lógicos y físicos, lo cual permite aplicar estrategias de eficiencia según la capacidad de ColectaTe para adquirir <i>hardware</i> más potente.</p> <p>En adición a esto, Knime cuenta con opciones para procesar los datos directamente en memoria sin necesidad de bajarlos a disco, lo cual puede ser utilizado en caso de necesitar mejorar la performance al procesar grandes volúmenes de datos a la vez.</p> <p>En cuanto a los componentes desarrollados, los mismos fueron generados para la extracción y análisis en tiempo real, lo que implica que procesarán un tweet a la vez, no impactando significativamente en los valores generales de eficiencia.</p>
Disponibilidad	<p>El diseño modular de la solución permite ejecutar los diferentes componentes en diferentes servidores lógicos y físicos, lo cual permite aplicar estrategias de redundancia según la capacidad de ColectaTe para adquirir el <i>hardware</i> redundante.</p> <p>No se toman en cuanto las interrupciones de servicios del ISP.</p>
Escalabilidad	<p>El diseño modular de la solución permite ejecutar los diferentes componentes en diferentes servidores lógicos y físicos, lo cual permite escalar el <i>hardware</i> según la capacidad de ColectaTe para adquirirlo y así incrementar la capacidad de procesamiento.</p> <p>Por otro lado, en base a los requerimientos funcionales se establece la carga normal del sistema, y en base a esto se diseñan los procesos para ser capaces de procesar del doble de datos dentro del umbral de eficiencia.</p>
Usabilidad	<p>Se realizarán diferentes tipos de gráficos (barras, líneas, torta) entendiendo que la representación de los datos en forma gráfica ofrece mensajes más claros facilitando el entendimiento de las conclusiones para la toma de decisiones.</p> <p>Los gráficos tenderán a ser sencillos, con poco texto y específicos para un conjunto de datos con un único denominador común.</p> <p>Cada gráfico irá acompañado de una tabla que presenta los mismos datos del gráfico para despejar cualquier duda que surja de la lectura del mismo.</p>

Tabla 4.1 Atributos de calidad Big Data

#### 4.2.1.2. Aplicación Móvil

Atributo	Estrategia
Portabilidad	Mediante la realización del empaquetado del sitio web, se construye una aplicación ejecutable en entornos Android 2.3 o superior.
Seguridad	Mediante la realización del empaquetado del sitio web, se cubren los mismos requisitos de seguridad que los del propio sitio web.
Eficiencia	Mediante la realización del empaquetado del sitio web, se cubren los mismos requisitos de eficiencia que los del propio sitio web.
Disponibilidad	Mediante la realización del empaquetado del sitio web, se cubren los mismos requisitos de disponibilidad que los del propio sitio web.
Interoperabilidad	Mediante la realización del empaquetado del sitio web, se cubren los mismos requisitos de interoperabilidad que los del propio sitio web.
Usabilidad	Dado que la plataforma de ejecución es sustancialmente diferente desde el punto de vista del tamaño del <i>display</i> y mecanismo de interacción, se prevé la adaptación de los estilos (.css) del sitio web para reordenar, ocultar y cambiar el diseño visual de algunos componentes para mejorar la usabilidad para los dispositivos móviles.

Tabla 4.2 Atributos de calidad Aplicación Móvil

### 4.3. Tecnologías y herramientas

Se detallan a continuación las tecnologías y herramientas utilizadas a lo largo del proyecto y los anexos donde se encuentra la investigación realizada para cada una de ellas.

Tecnología	Nombre de la herramienta	Anexo
Plataforma de <i>Big Data</i>	KNIME	Anexo 3 - Cómo KNIME cumple con el proceso de <i>Big Data</i> ?
Motor de base de datos	MySQL	9.2.2 - Plataformas de almacenamiento
Herramienta de visualización en tiempo real	StyleScope Agile	9.2.3 - Herramientas de visualización
Empaquetado de aplicación móvil	Apache Cordova	9.2.4 - Herramientas de Empaquetado Móvil

Tabla 4.3 Tecnologías y herramientas

#### 4.4. Requerimientos mínimos de hardware

Los procesos se ejecutaron y testearon en una maquina con las características que se detallan en la tabla a continuación.

Característica	Especificación
Procesador	Quadcore 1.9 Ghz
Memoria	8 GB
Disco Duro	300 GB disponibles
Video	Cualquiera
Arquitectura	64 bit
Sistema Operativo	Windows 8.1

Tabla 4.4 Requerimientos de hardware

Para asegurar el cumplimiento de los requerimientos no funcionales de eficiencia y escalabilidad, se sugiere instalar el servidor en donde correrán los procesos en una máquina con similares características.

En adición, para cumplir con el requerimiento de eficiencia es necesario realizar la siguiente configuración sobre la instalación de Knime:

1. En el directorio de instalación de Knime, localizar el archivo Knime.ini
2. Dentro del archivo, la línea -Xmx512m cambiarla por -Xmx4g
3. Dentro del archivo, en la línea -XX:MaxPermSize el valor 256m cambiarlo por 3g

#### 4.5. Vistas de arquitectura

##### 4.5.1.1. Vista de Módulos

En la Imagen 4.1 se pueden ver los diferentes módulos o procesos (extracción, análisis, reportes) que componen la solución según la capa (datos, lógica, presentación) a la que pertenecen y cómo se comunican entre sí.

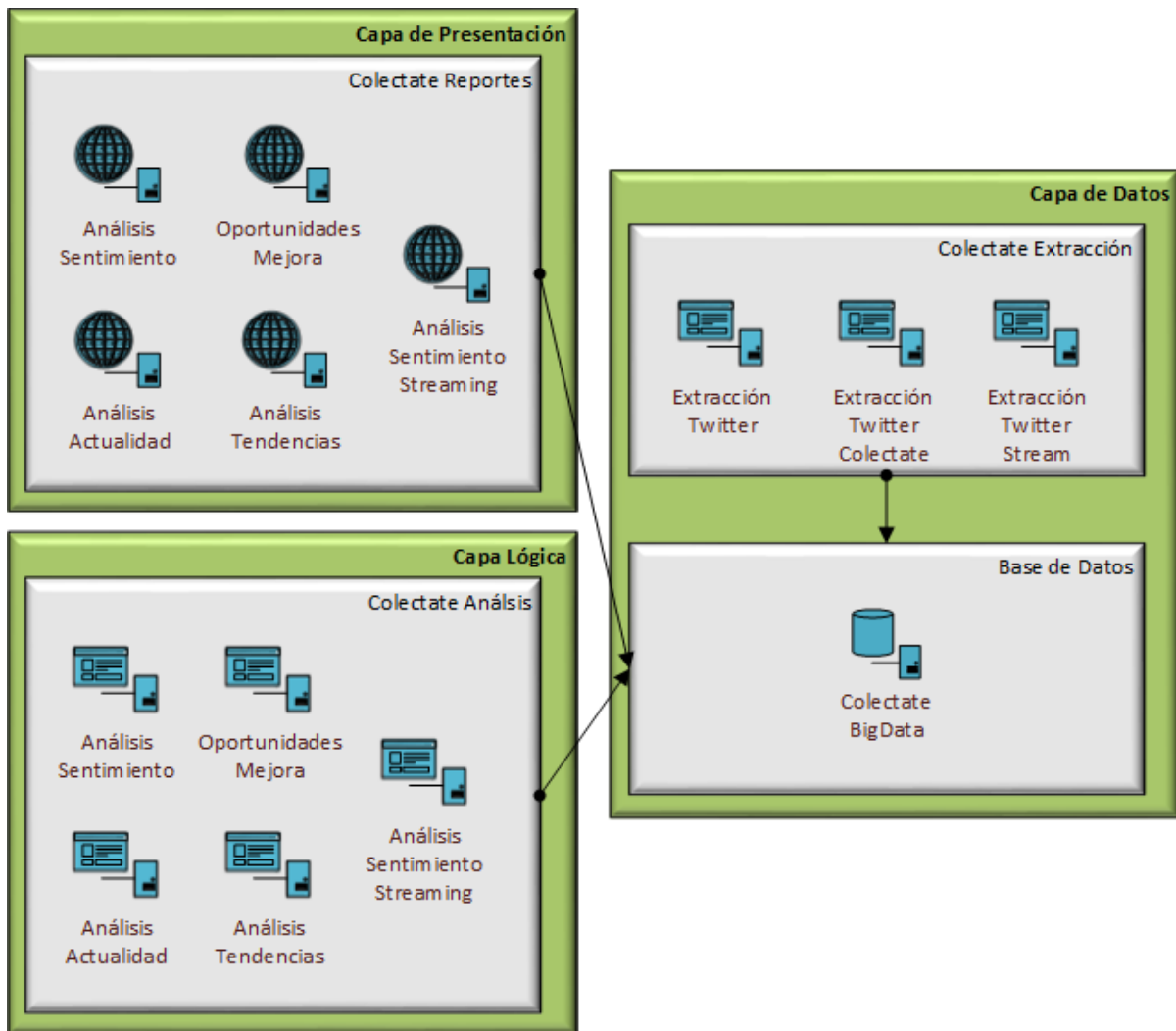


Imagen 4.1 Vista de módulos procesos Big Data

#### 4.5.1.2. Vista de Ejecución

En la Imagen 4.2 se pueden apreciar las diferentes plataformas en las que se ejecutan las diferentes capas de la solución.

- ✚ La capa de datos se encuentra conformada por la base de datos sobre MySQL y los procesos de extracción de tweets que almacenan los datos en dicha BD y que se ejecutan sobre Knime.
- ✚ La capa lógica contiene los proceso de análisis, los cuales se ejecutan sobre Knime
- ✚ La capa de presentación contiene los reportes, los cuales se generan en procesos ejecutados en Knime o mediante el *dashboard* de tiempo real el cual se ejecuta sobre StyleScope Agile de InetSoft.

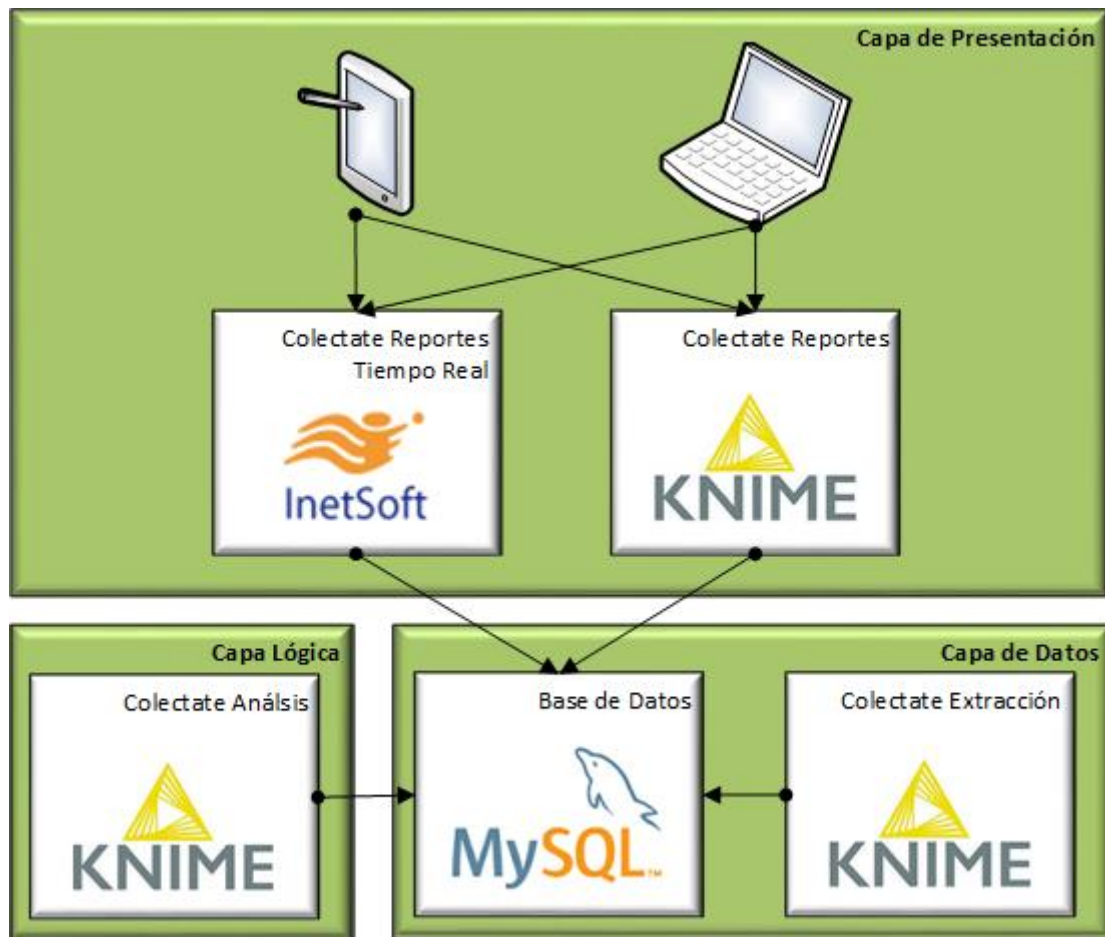


Imagen 4.2 Vista ejecución procesos Big Data

#### 4.5.1.3. Vista de Asignación

En la Imagen 4.3 se visualiza en qué servidores se deben desplegar cada uno de los módulos o procesos para su correcta ejecución.

Cabe destacar que si bien cada los diferentes módulos corren sobre diferentes plataformas o servidores lógicos, a nivel físico, pueden ejecutarse en el mismo servidor.

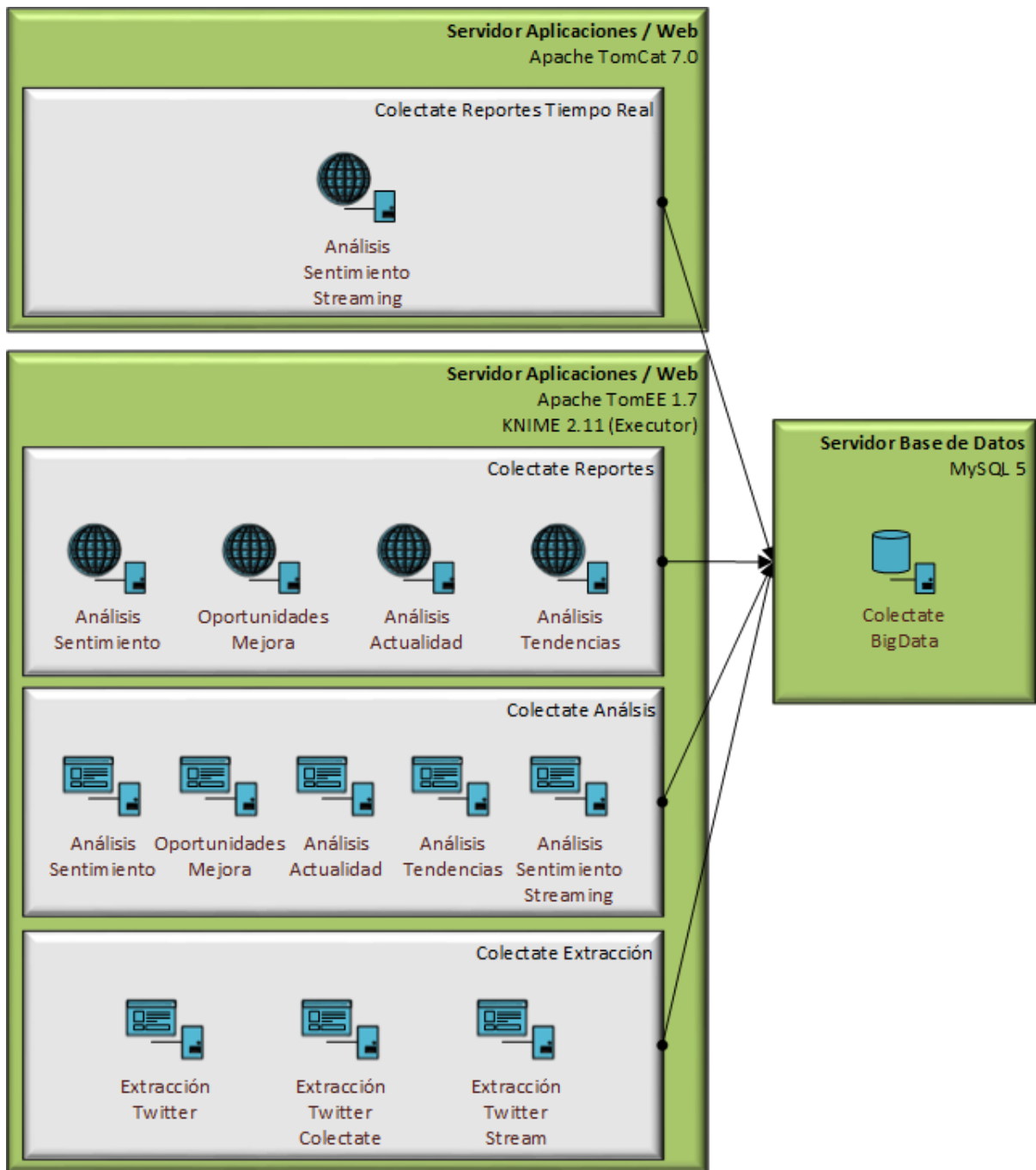


Imagen 4.3 Vista asignación procesos Big Data

## 4.6. Vista de base de datos

### 4.6.1.1. Modelo Entidad Relación

En la Imagen 4.4 se puede apreciar el diseño de la base de datos para almacenar la información relacionada con los procesos de *Big Data*.

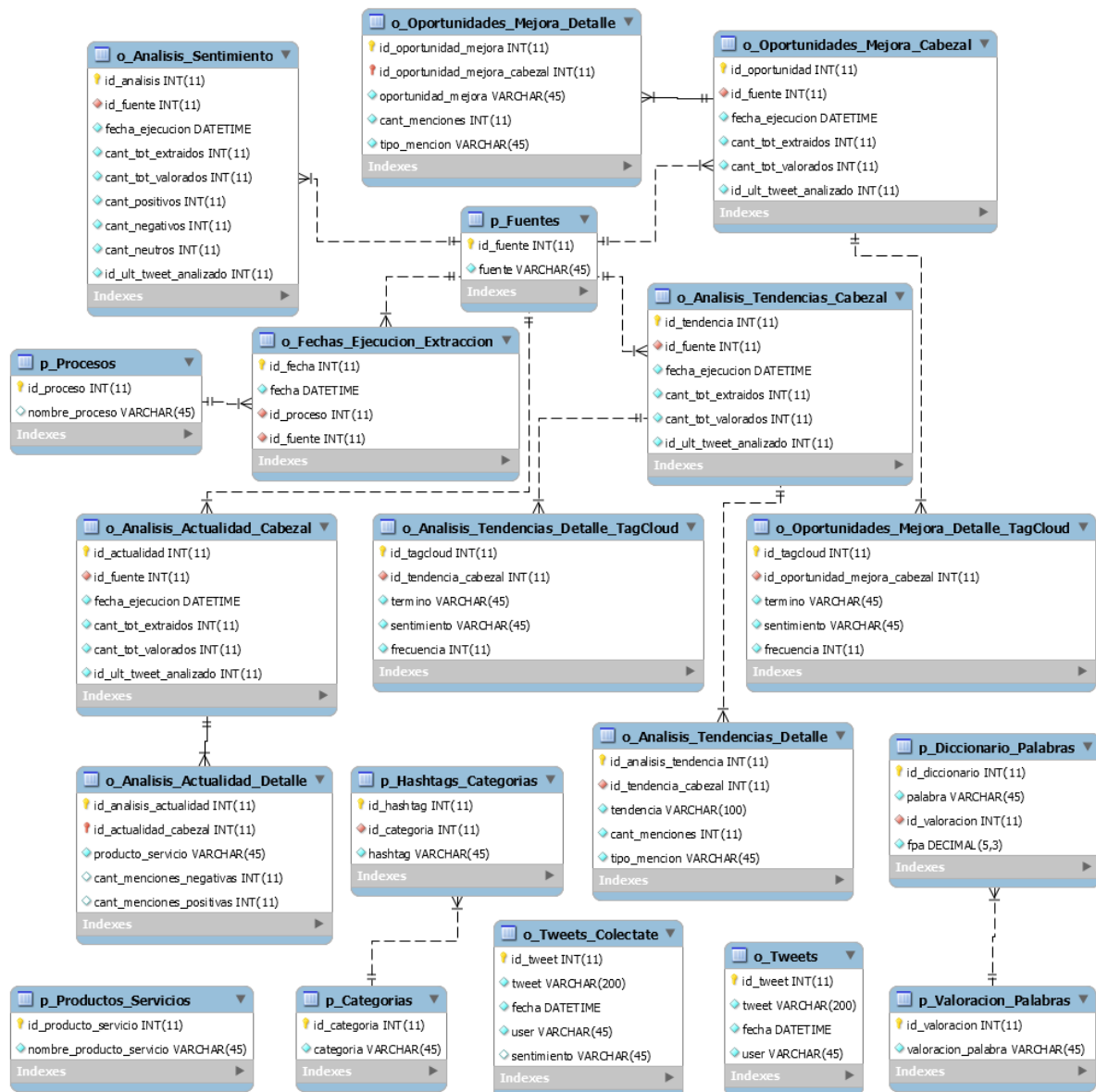


Imagen 4.4 Modelo Entidad Relación

#### 4.6.1.2. Catálogo de elementos

Elemento	Descripción
o_Analisis_Sentimiento	Contiene la información consolidada del análisis de sentimiento
o_Oportunidades_Mejora_Cabezal	Contiene la información consolidada del análisis de las oportunidades de mejora
o_Oportunidades_Mejora_Detalle	Contiene la información detallada por tópicos del análisis de las oportunidades de mejora

o_Oportunidades_Mejora_Detalle_TagCloud	Contiene la información detallada por palabra del análisis de las oportunidades de mejora
o_Analisis_Tendencias_Cabezal	Contiene la información consolidada del análisis de tendencias
o_Analisis_Tendencias_Detalle	Contiene la información detallada por tópico del análisis de tendencias
o_Analisis_Tendencias_Detalle_TagCloud	Contiene la información detallada por palabra del análisis de tendencias
o_Analisis_Actualidad_Cabezal	Contiene la información consolidada del análisis de actualidad
o_Analisis_Actualidad_Detalle	Contiene la información detallada por palabra del análisis de actualidad
o_Fechas_Ejecucion_Extraccion	Contiene las fechas históricas en que se ejecutaron los procesos de extracción de datos
o_Tweets	Almacena los tweets extraídos por hashtags
o_Tweets_Colectate	Almacena los tweets extraídos que hacen referencia a ColectaTe
p_Procesos	Tabla de parámetros donde se listan los procesos a nivel funcional
p_Fuentes	Tabla de parámetros donde se listan las posibles fuentes de extracción
p_Diccionario_Palabras	Tabla de parámetros que contiene las palabras claves utilizadas para clasificar el sentimiento en positivo o negativo
p_Valoracion_Palabras	Tabla de parámetros que contiene los tipos de sentimientos posibles de las palabras
p_Hashtags_Categorias	Tabla de parámetros que contiene los hashtags utilizados para la extracción de datos
p_Categorias	Tabla de parámetros que contiene las categorías en que se trata de agrupar las opiniones de los usuarios al interpretar su contexto
p_Productos_Servicios	Tabla de parámetros utilizada para contar la ocurrencia de palabras relacionadas con ciertos productos o servicios

Tabla 4.5 Catálogo de elementos de base de datos

## 4.7. Vistas de procesos

Se detalla y describen a continuación todos los flujos construidos para la realización del proyecto. Para información detallada sobre los metanodos contenidos en cada uno de los flujos, referirse al Anexo 18 - Descripción de metanodos KNIME. En el Anexo 20 - Nomenclatura de flujos en Knime se explica la semántica utilizada por los diferentes componentes que forman los flujos.

### 4.7.1. Extracción

#### 4.7.1.1. Extracción Twitter

El proceso “Extracción Twitter” (Imagen 4.5) es el encargado de realizar la extracción de tweets en función de una lista de *hashtags* (lista provista por ColectaTe sobre los temas de interés para la empresa). El filtrado por *hashtag* permite extraer tweets que tienen relación con el tema que se está queriendo analizar.

Dicha lista se levanta de la base de datos, y se genera una consulta que es luego enviada para la extracción en Twitter.

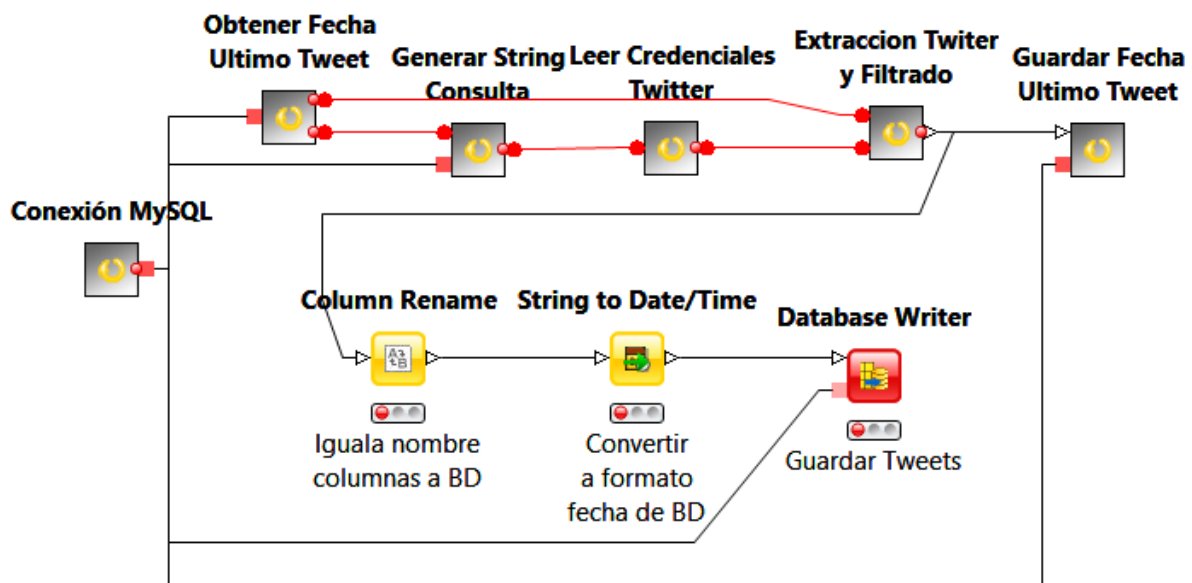


Imagen 4.5 Flujo Extracción Twitter

La ejecución se realiza de la siguiente forma:

1. Se realiza la conexión a la base MySQL a través de la lectura de un archivo “Propiedades.txt”, el cual contiene el nombre del servidor remoto, el nombre de la base de datos y las credenciales de acceso.
2. Se obtiene la fecha del último tweet extraído para evitar la extracción de tweets obtenidos en alguna ejecución anterior.
3. Se genera el *string* de la consulta, en función de la lista de *hashtags* obtenidos desde la base de datos, la fecha del último tweet extraído y el parámetro “lang:es” para obtener tweets únicamente en español

4. Se leen las credenciales de acceso a la aplicación Twitter desde el archivo “Credenciales Twitter Extracción Sin Filtrar.txt”
5. Se realiza la conexión a Twitter y se extraen los tweets en función de la consulta generada anteriormente. Estos datos extraídos son luego filtrados y pre-procesados.
6. Finalmente se almacenan los tweets extraídos y la fecha del último tweet.

La extracción de tweets se realiza a través de la REST API de Twitter. Esta API provee una lista de Tweets que coincidan con una consulta específica, pero posee ciertas restricciones y requerimientos los cuales están detalladas en el sitio de Twitter [11].

#### 4.7.1.2. Extracción Twitter ColectaTe

El proceso “Extracción Twitter ColectaTe” (Imagen 4.6) es el encargado de realizar la extracción de tweets que mencionan el nombre de la empresa.

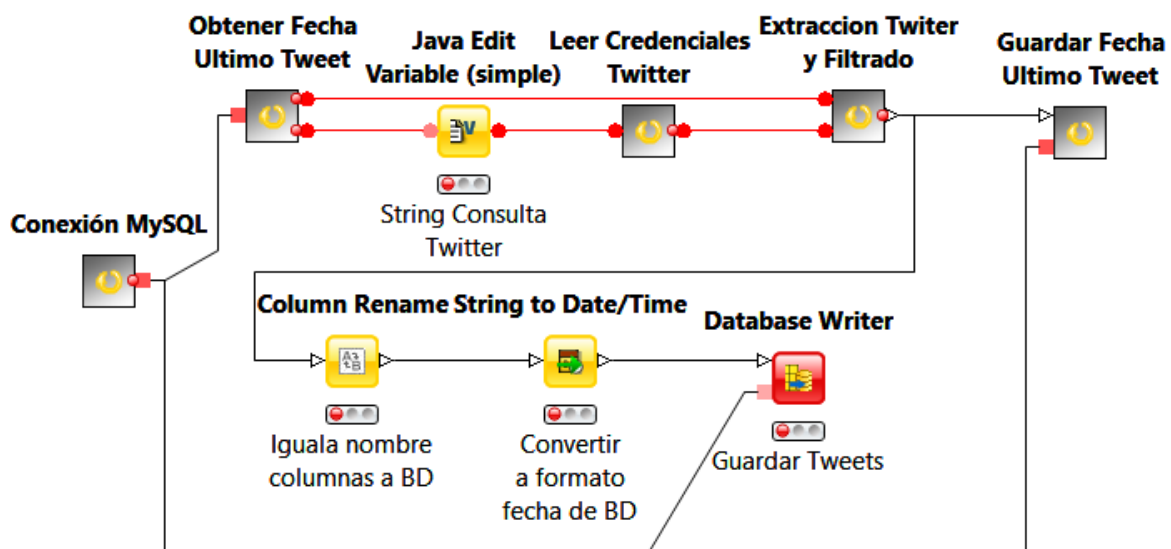


Imagen 4.6 Flujo Extracción Twitter ColectaTe

La ejecución se realiza de la misma manera que la anterior, con las siguientes salvedades:

1. Se leen las credenciales de acceso a la aplicación Twitter desde el archivo “Credenciales Twitter Extracción Filtrado.txt”
2. El *string* de la consulta solo incluye el nombre de la empresa, tomándose las siguientes posibilidades: “ColectaTe”, “Colecta te” o “Colecta-te” y el parámetro “lang:es” para obtener tweets únicamente en español

La extracción se realiza a través de la misma API que en el flujo anterior, por lo cual las restricciones también aplican en este flujo.

#### 4.7.1.3. Extracción Twitter Stream

El proceso de “Extracción Twitter *Stream*” (Imagen 4.7) se encarga de realizar la extracción en tiempo real de los tweets.

Es importante destacar, que Knime como herramienta, no se encuentra optimizada para realizar tareas donde un nodo se ejecuta de forma persistente, sino que funcione como un flujo de trabajo, donde un nodo es ejecutado luego que el anterior ha concluido su ejecución y generado la salida correspondiente.

Es por esta razón, que para la implementación del proceso de extracción en tiempo real, fue necesario desarrollar componentes a medida para Knime (nodos en lenguaje java) que permiten conectarse, escuchar los tweets y extraer los datos de los usuarios que los enviaron.

Los componentes desarrollados fueron los denominados:

- ✚ ColectaTe Conexión: establece conexión persistente con los servidores de Twitter.
- ✚ ColectaTe Streaming: escucha de forma continua la conexión establecida, a la espera de que Twitter envíe uno o más tweets donde se mencionen los términos de búsqueda.
- ✚ ColectaTe Usuario: extrae los datos de los usuarios que enviaron los tweets extraídos.

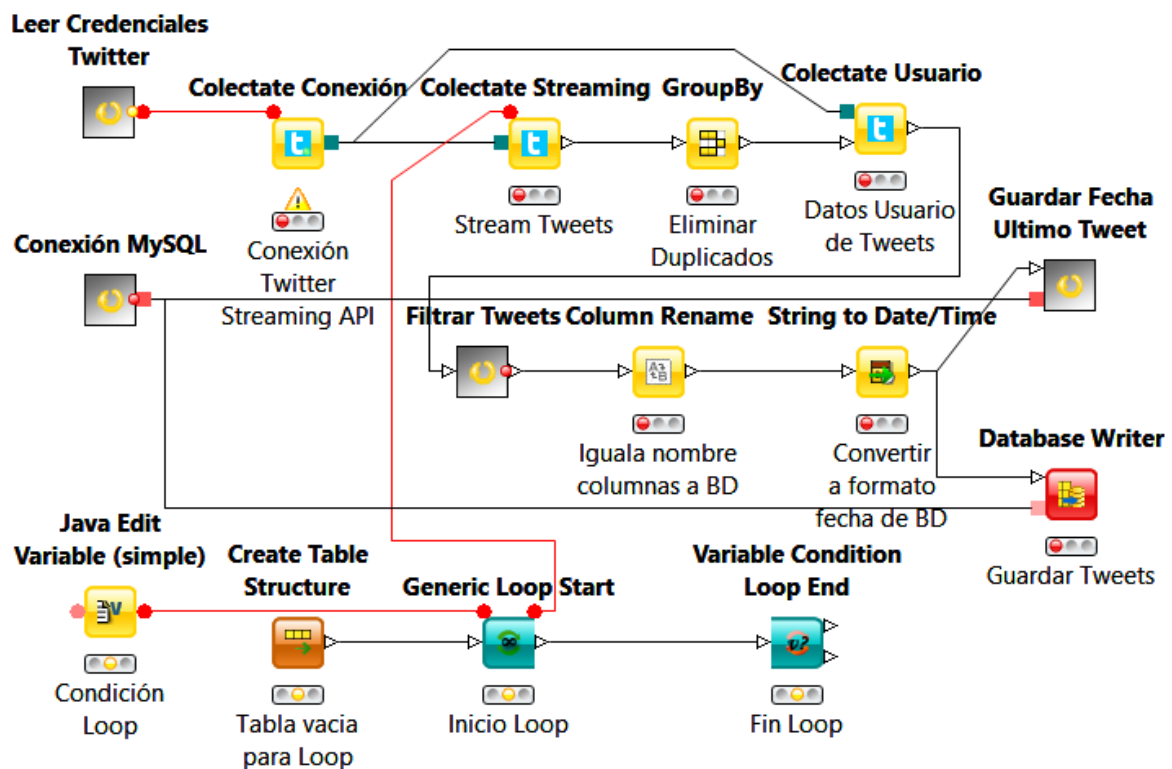


Imagen 4.7 Flujo Extracción Twitter Stream

El ciclo de ejecución sigue el siguiente orden:

1. Se establece conexión persistente con los servidores de Twitter
2. Se inicia bucle
  - a) Se procede a la escucha del canal
  - b) Twitter envía tweet
  - c) Se obtienen los datos de los usuarios

- d) Se aplica el filtro (ver Anexo 18 - Descripción de metanodos KNIME - Filtrar Tweets)
  - e) Se guarda la fecha del último tweet para futuras extracciones
  - f) Se guarda el tweet y los datos de usuario
3. Fin de la iteración actual y comienzo de una nueva iteración (vuelve al punto 2)

Es importante destacar que Twitter no asegura el envío del 100% de los tweets mediante la *Streaming API* [12]. Para alcanzar dicha efectividad se ofrece un servicio pago ofrecido mediante terceros llamado Firehose.

## 4.7.2. Análisis

### 4.7.2.1. Análisis de Sentimiento (*Sentiment Analysis*)

El proceso de análisis de sentimiento (Imagen 4.8) es el encargado de realizar el análisis de sentimiento de los últimos tweets extraídos, y guardar los resultados en la base.

En el Anexo 4 - Análisis de sentimiento (*Sentiment Analysis*) se encuentra una descripción de la investigación realizada sobre el análisis de sentimiento.

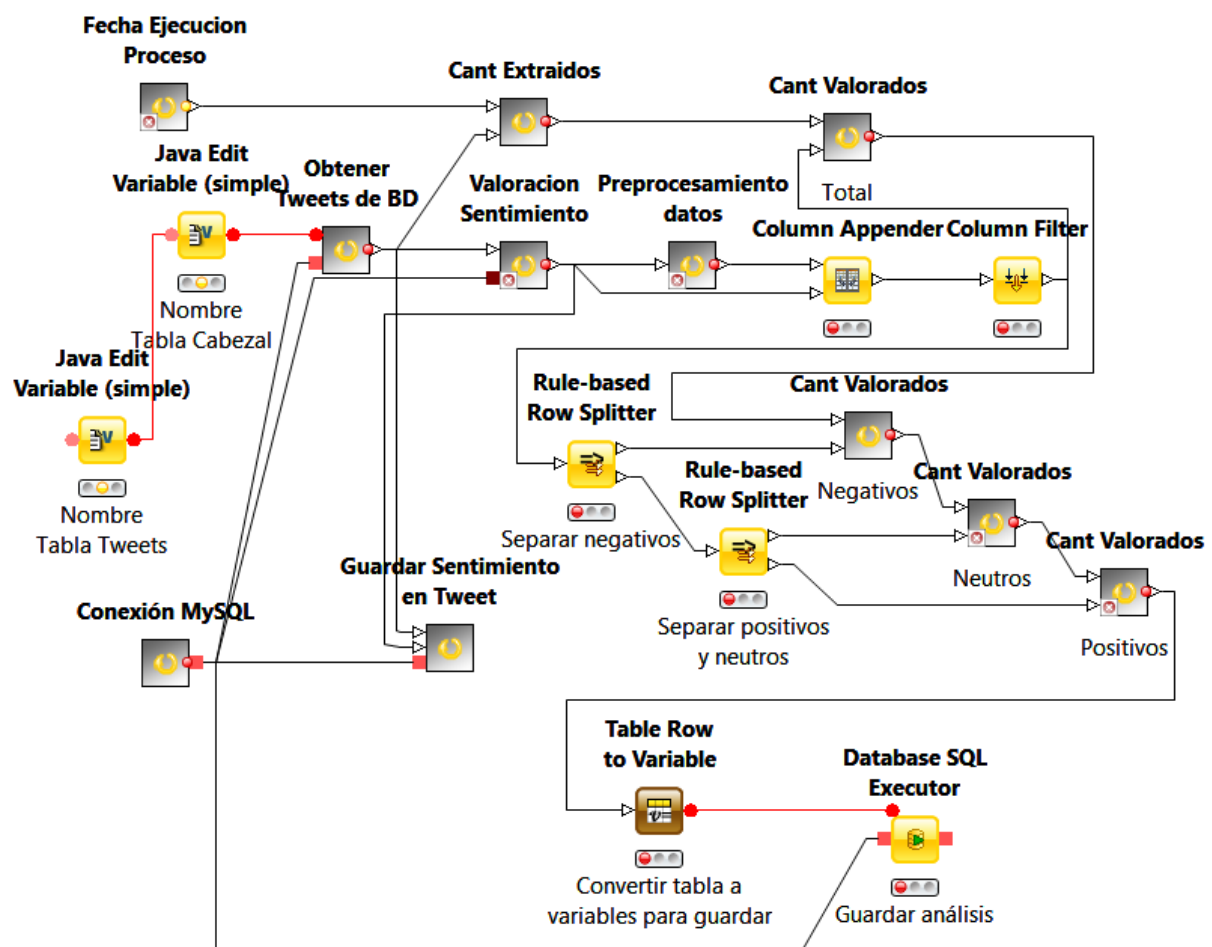


Imagen 4.8 Flujo Sentiment Analysis

La ejecución del flujo se realiza de la siguiente forma:

1. Se realiza la conexión a la base MySQL a través de la lectura de un archivo "Propiedades.txt", el cual contiene el nombre del servidor remoto, el nombre de la base de datos y las credenciales de acceso.
2. Se obtiene la fecha del último tweet analizado, y se recuperan todos los tweets almacenados que sean posteriores a dicha fecha.
3. Se procede a realizar la valoración del sentimiento (ver Anexo 18 - Descripción de metanodos KNIME - Valoración Sentimiento)
4. Luego, se realiza el pre procesamiento de datos (ver Anexo 18 - Descripción de metanodos KNIME - Preprocesamiento datos)
5. Se guarda el sentimiento obtenido por cada tweet en la base.
6. Se obtiene la cantidad de tweets positivos, negativos, neutros, extraídos y valorados y la fecha de ejecución del proceso
7. Se almacena en la base los datos anteriormente calculados.

#### 4.7.2.2. Análisis de Sentimiento *Streaming*

En proceso de análisis de sentimiento *streaming* (Imagen 4.9) pretende realizar el análisis de sentimiento de los tweets extraídos a través del proceso de extracción "Extracción Twitter *Stream*", y guardar los resultados en la base.

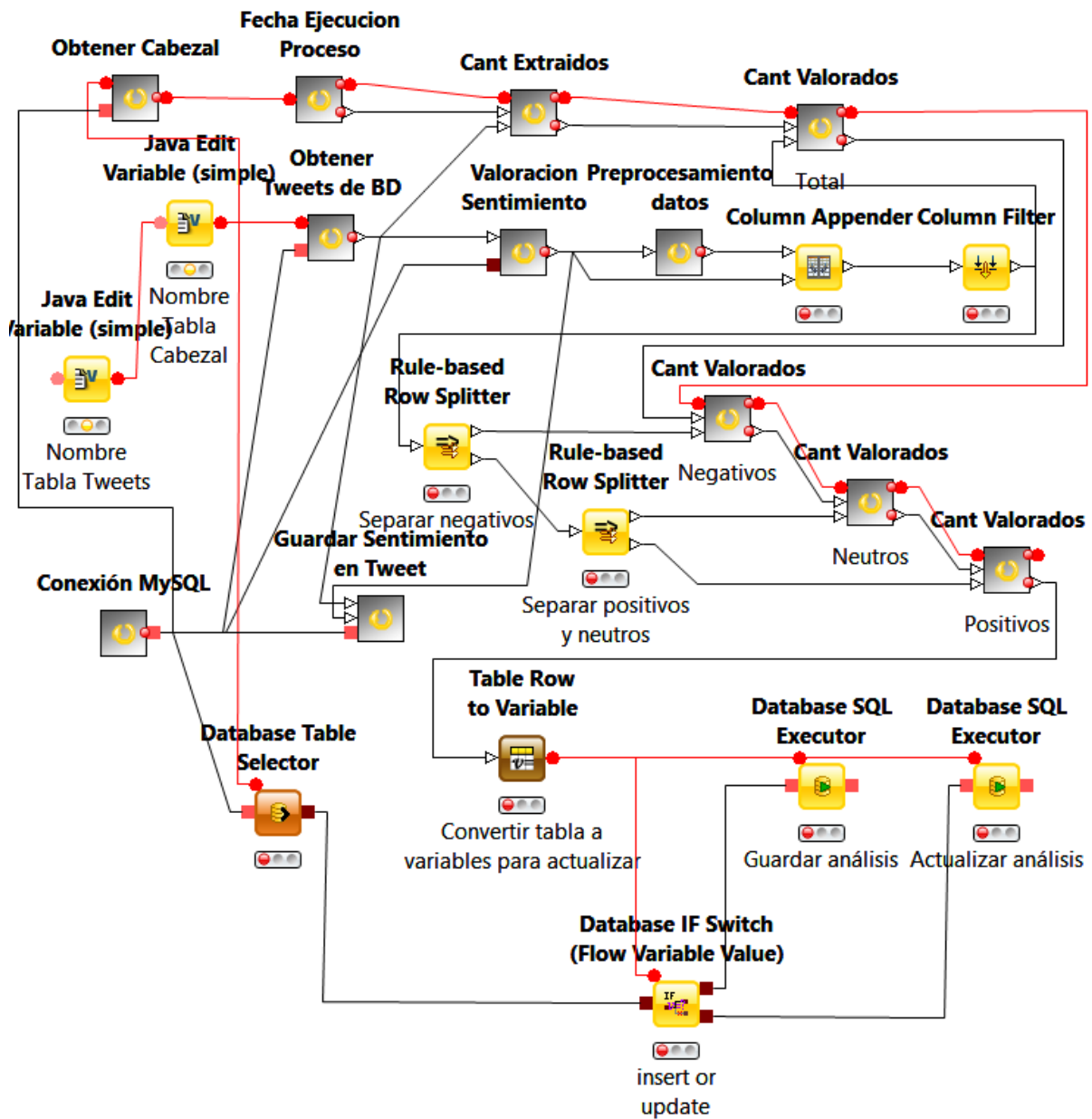


Imagen 4.9 Flujo Sentiment Analysis Streaming

La ejecución del flujo se realiza de la misma forma que el anterior, con la siguiente diferencia: el guardado del análisis en la base se realiza a través de un “IF Switch”, el cual determina que si el último tweet analizado tiene una fecha que corresponde a la misma semana de ejecutado el análisis, se realizará un *update* de los datos. En caso que la semana sea anterior, se realizará el *insert* de una nueva tupla.

#### 4.7.2.3. Análisis Actualidad

El proceso análisis actualidad (Imagen 4.10) es el encargado de analizar los tweets para determinar la popularidad de ciertos productos y servicios, de interés para ColectaTe.

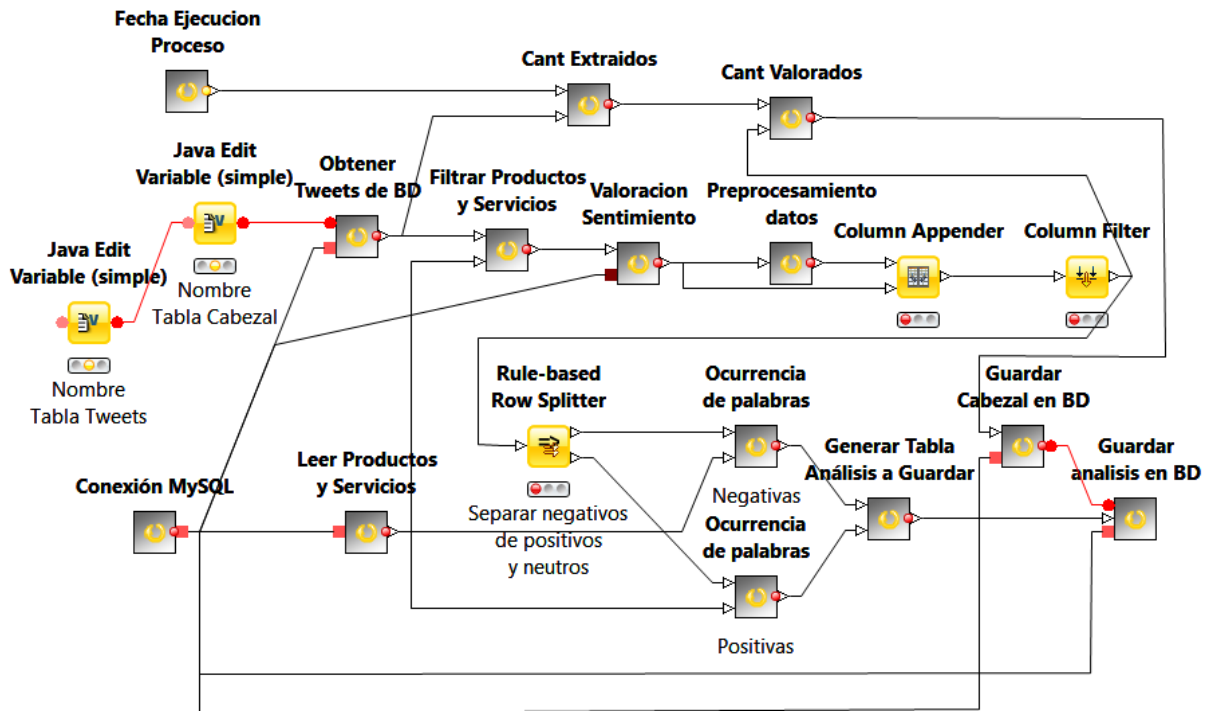


Imagen 4.10 Flujo Análisis Actualidad

La ejecución del flujo se realiza de la siguiente forma:

1. Se realiza la conexión a la base MySQL a través de la lectura de un archivo "Propiedades.txt", el cual contiene el nombre del servidor remoto, el nombre de la base de datos y las credenciales de acceso.
2. Se obtiene de la base la lista de productos / servicios de interés para la empresa.
3. Se obtiene la fecha del último tweet analizado, y se recuperan todos los tweets almacenados que sean posteriores a dicha fecha.
4. Para cada tweet, se identifica si el mismo contiene o no alguno de los productos o servicios de interés para ColectaTe. Los que no mencionan ninguno de los productos / servicios, no serán tenidos en cuenta para el análisis.
5. Se realiza un análisis de sentimiento de cada tweet, de manera de poder identificar en el análisis cuáles productos / servicios son mencionados de forma positiva o negativa.
6. Se realiza el pre procesamiento de datos (ver Anexo 18 - Descripción de metanodos KNIME)
7. Se realiza el conteo de ocurrencias de cada producto / servicio (negativos y positivos) para determinar la popularidad
8. Se almacena en la base de datos el cabezal del análisis con los siguientes datos: fecha de ejecución del análisis, cantidad de tweets extraídos, cantidad de tweets valorados, identificador del último tweet analizado.
9. Se almacena en la base de datos el resultado del análisis de popularidad, conteniendo los productos / servicios mencionados y su respectiva cantidad de menciones, con una referencia al cabezal del análisis.

#### 4.7.2.4. Análisis de Tendencias

El proceso de análisis de tendencias (Imagen 4.11) es el encargado de detectar los temas que las personas están “hablando” en las redes sociales actualmente.

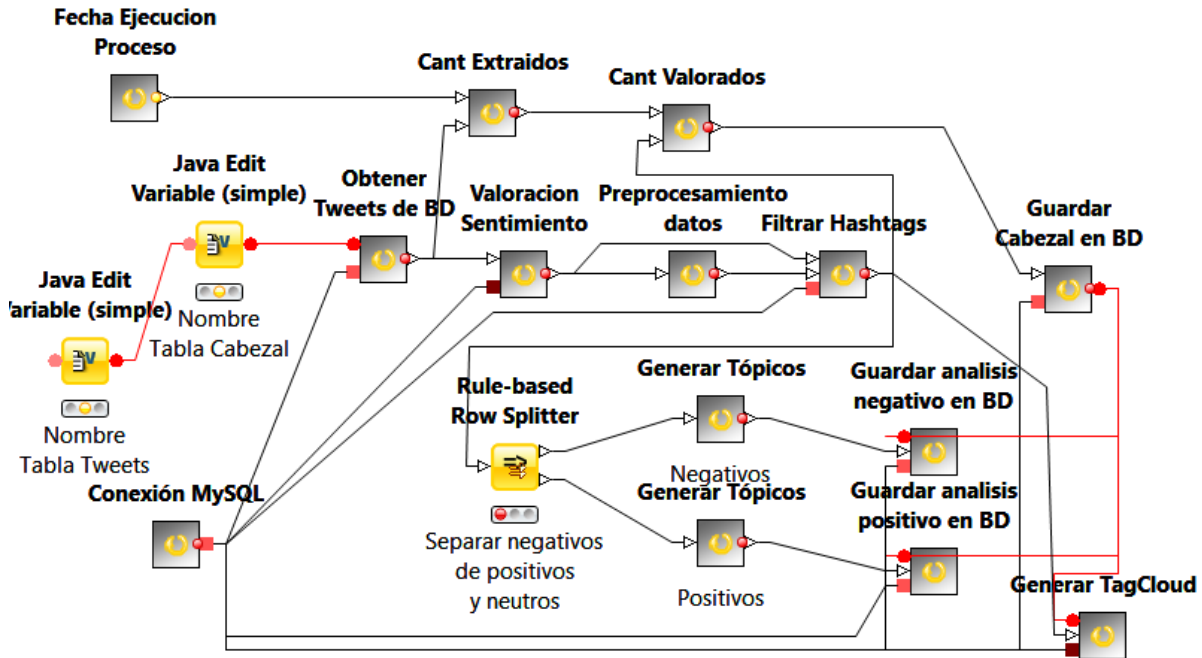


Imagen 4.11 Flujo Análisis de Tendencias

La ejecución del flujo se realiza de la siguiente forma:

1. Se realiza la conexión a la base MySQL a través de la lectura de un archivo “Propiedades.txt”, el cual contiene el nombre del servidor remoto, el nombre de la base de datos y las credenciales de acceso.
2. Se obtienen de la base de datos los tweets que aún no han sido analizados
3. Se realiza un análisis de sentimiento para diferenciar la tendencias positivas y negativas
4. Se realiza un filtrado por *hashtag* para que no sean incluidos en el análisis como tendencia
5. Se generan los tópicos: identificación de categorías tanto positivas como negativas (ver Anexo 18 - Descripción de metanodos KNIME - Generar Tópicos). Se contabilizan para determinar el nivel de popularidad.
6. Por otra parte, se calculan las frecuencias de las palabras mencionadas en los tweets extraídos y se almacenan en la base de datos.
7. Guardado del cabezal de análisis conteniendo la fecha de ejecución del análisis, la cantidad de tweets extraídos y la cantidad de tweet valorados
8. Guardado del análisis con los tópicos detectados y la cantidad de menciones de los mismos, separado por menciones negativas y menciones positivas

#### 4.7.2.5. Análisis Oportunidades de Mejora

El proceso análisis oportunidad de mejora (Imagen 4.12) pretende identificar las oportunidades de mejora del servicio brindado por la empresa.

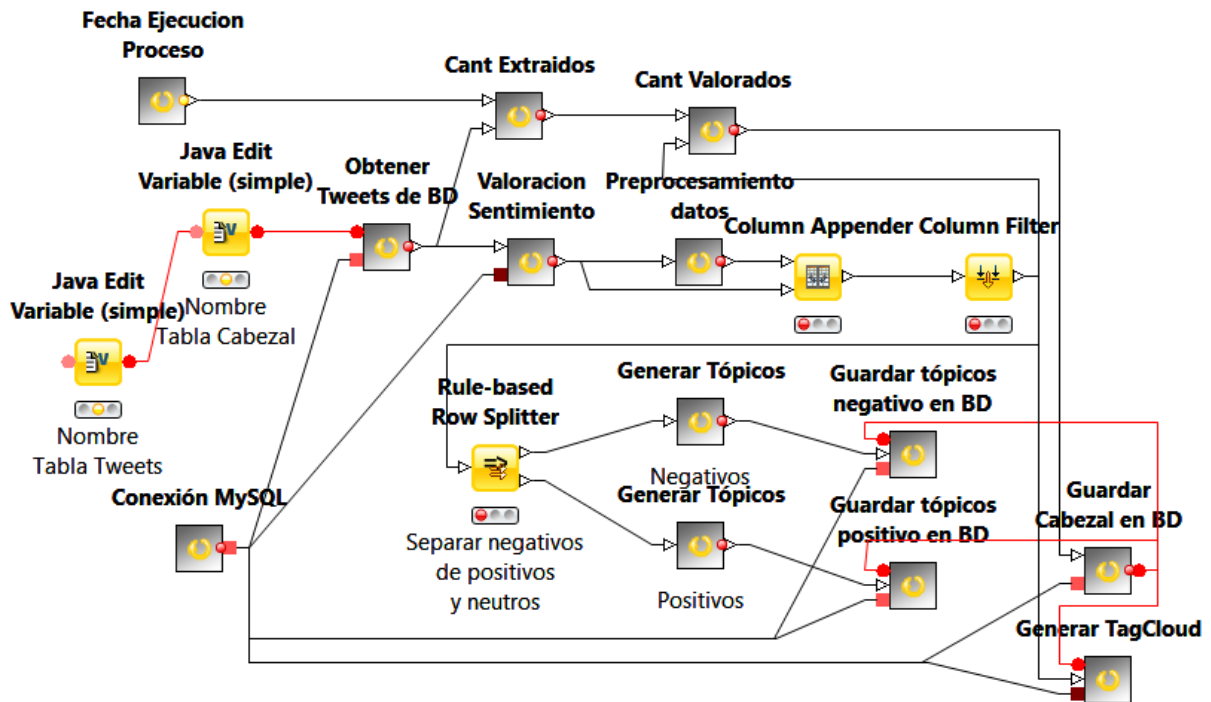


Imagen 4.12 Flujo Análisis Oportunidades de Mejora

La ejecución se realiza exactamente de la misma manera que la anterior, la diferencia siendo que los tweets a analizar son aquellos que mencionen a la empresa (Tweets extraídos a través del proceso “Extracción Twitter ColectaTe”)

### 4.7.3. Reportes

#### 4.7.3.1. Análisis de Sentimiento

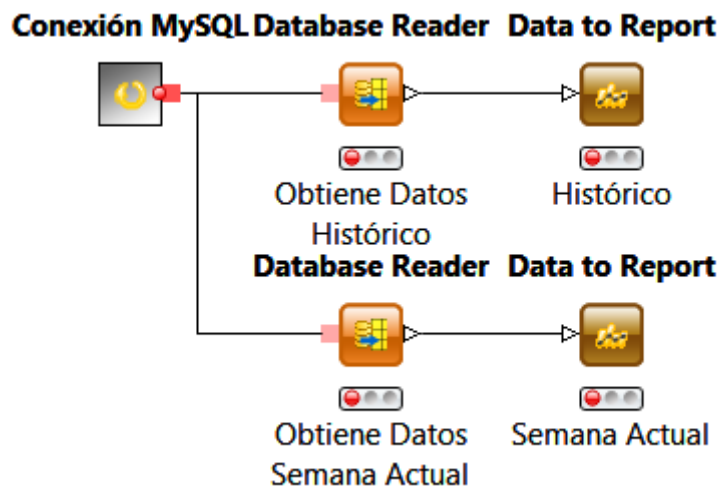


Imagen 4.13 Flujo reporte Análisis de Sentimiento

Para la generación del reporte (Imagen 4.13), se realiza la conexión a la base MySQL y se obtienen los resultados del análisis almacenado en la tabla *o\_Analisis\_Sentimiento* correspondientes a la semana de ejecución del último proceso, y los resultados correspondientes a las últimas 5 ejecuciones para conformar los resultados del histórico.

#### 4.7.3.2. Análisis Actualidad

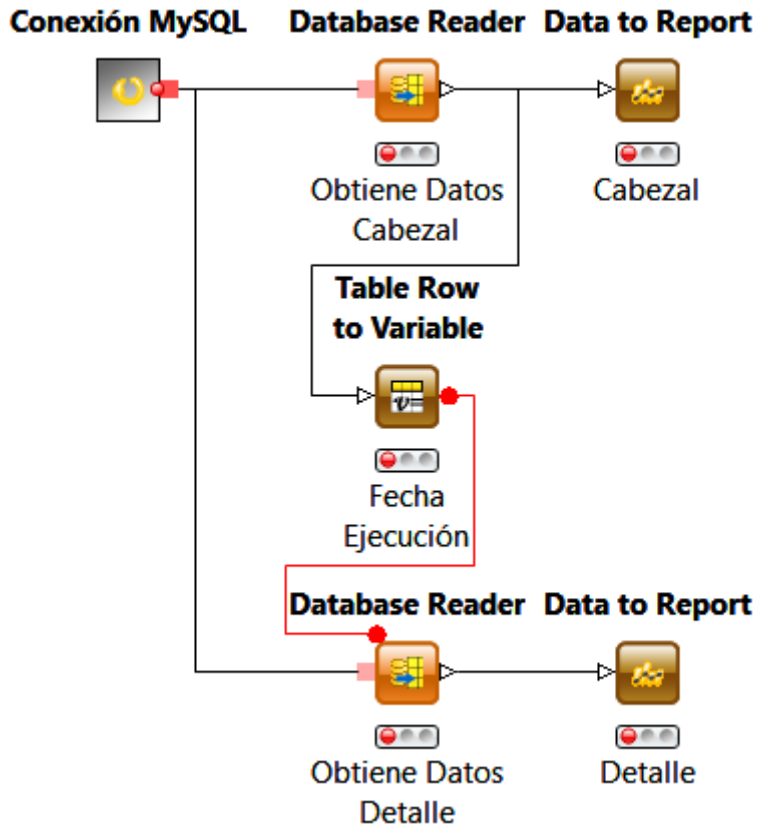


Imagen 4.14 Flujo reporte Análisis Actualidad

Para la generación del reporte (Imagen 4.14), se realiza la conexión a la base MySQL y se obtienen los resultados del análisis almacenado en las tablas *o\_Analisis\_Actualidad\_Cabecal* y *o\_Analisis\_Actualidad\_Detalle* correspondientes a la última ejecución del análisis.

#### 4.7.3.3. Análisis de Tendencias

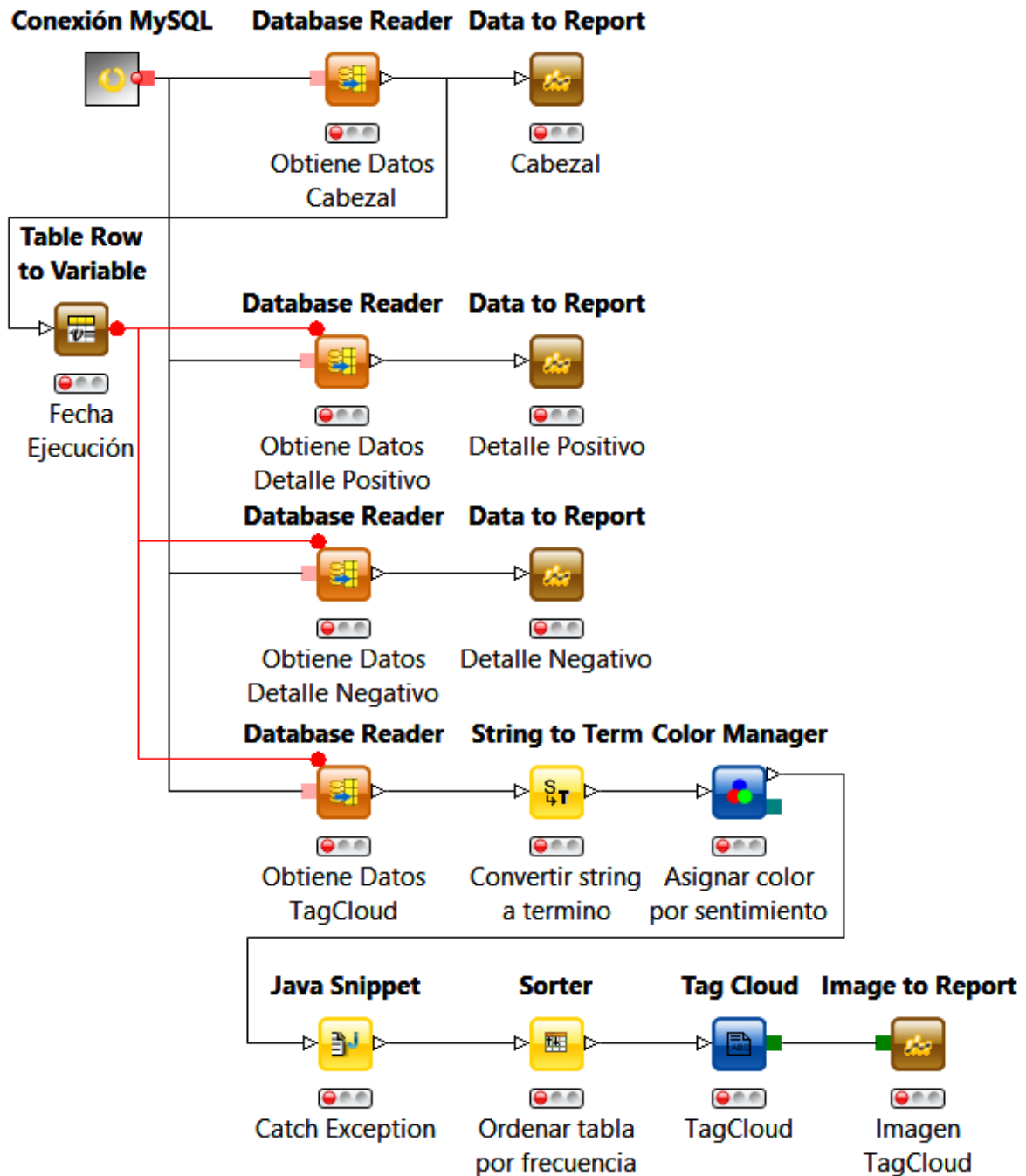


Imagen 4.15 Flujo reporte Análisis de Tendencias

Para la generación del reporte (Imagen 4.15 Imagen 4.15 Flujo reporte Análisis de Tendencias), se realiza la conexión a la base MySQL y se obtienen los resultados del análisis almacenado en las tablas *o\_Analisis\_Tendencias\_Cabezal* y *o\_Analisis\_Tendencias\_Detalle* (conteniendo las tendencias positivas y las tendencias negativas) correspondientes a la última ejecución del análisis.

Por otro lado, se obtienen las frecuencias almacenadas en la tabla *o\_Analisis\_Tendencias\_Detalle\_TagCloud*, asignándole un color según el sentimiento detectado para la generación del *Tag Cloud*.

#### 4.7.3.4. Análisis de Oportunidades de mejora

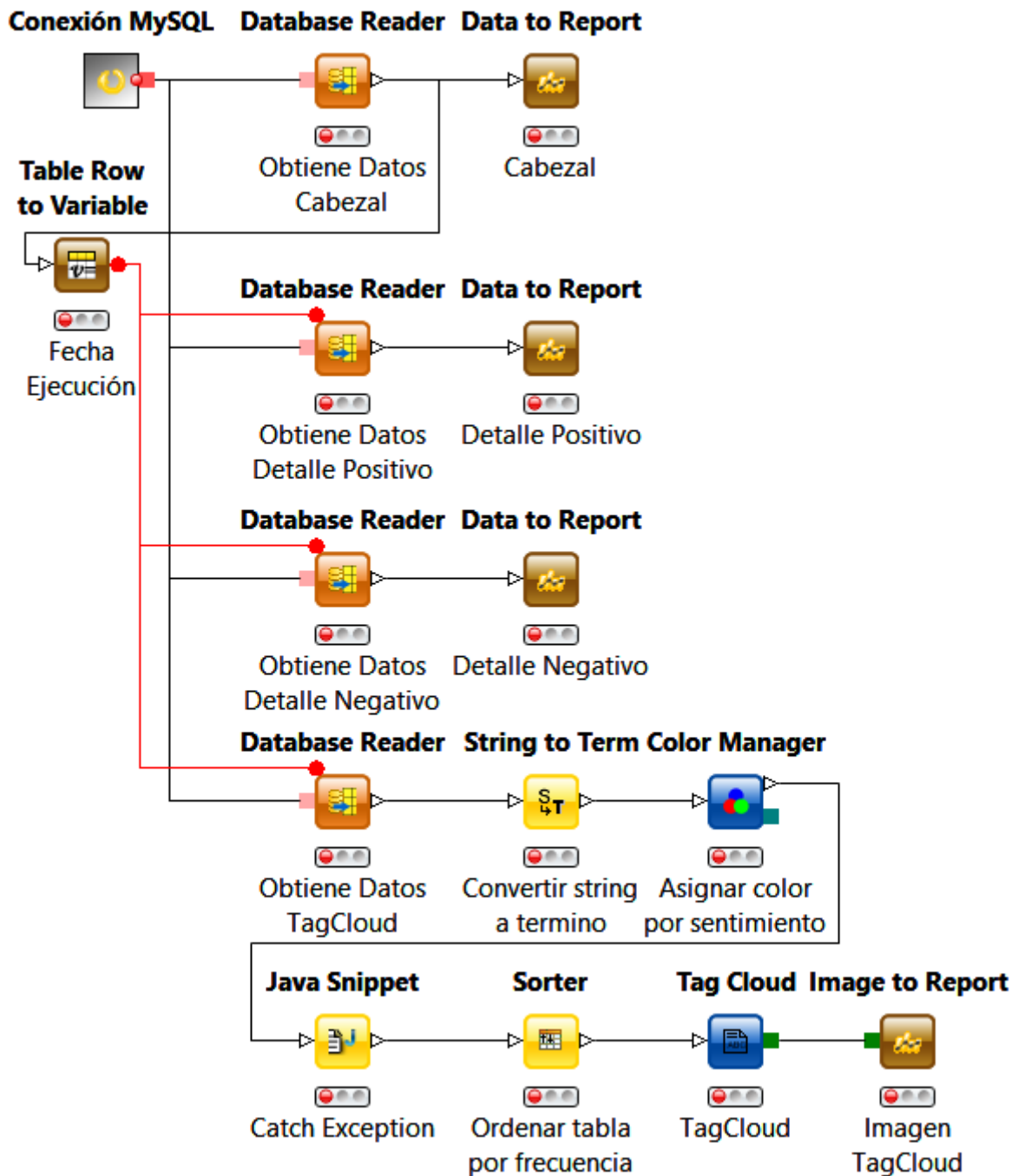


Imagen 4.16 Flujo reporte Análisis de Oportunidades de mejora

La generación de este reporte (Imagen 4.16) se realiza de la misma forma que el reporte anterior, la única diferencia siendo las tablas consultadas para obtener los resultados de los análisis. En este caso se consultan las tablas *o\_Oportunidades\_Mejora\_Cabecal*, *o\_Oportunidades\_Mejora\_Detalle* y *o\_Oportunidades\_Mejora\_Detalle\_TagCloud*.

#### 4.7.3.5. Análisis de Sentimiento en Tiempo Real

Para la generación del *dashboard* de análisis de sentimiento en tiempo real se generó un *dataset* en la herramienta StyleScope Agile, el cual consume los datos ya analizados y consolidados desde la base de datos.

El *dataset* cuenta con tres tablas internas, las cuales se ven en la Imagen 4.17.

o_Analisis_Sentimiento (Primary)					
Fech	cant_tot_extraidos	cant_tot_valorados	Positivo	Neutro	Negativo
2015-07-13	31	26	6	20	0
2015-07-20	25	25	10	15	0
2015-07-27	44	44	15	24	5
2015-08-03	26	26	11	12	3

o_Tweets_Colectate			
fec	Fecha_Analisis	user	tweet
2015-08-18 16:23:3	2015-08-18	XXXX	XXXXXXXX

Query1		
Fecha_Anali	Sentimiento	Menciones
2015-07-13 00:00:00	Positivo	6
2015-07-13 00:00:00	Neutro	20
2015-07-13 00:00:00	Negativo	0
2015-07-20 00:00:00	Positivo	10
2015-07-20 00:00:00	Neutro	15
2015-07-20 00:00:00	Negativo	0
2015-07-27 00:00:00	Positivo	15
2015-07-27 00:00:00	Neutro	24
2015-07-27 00:00:00	Negativo	5
2015-08-03 00:00:00	Positivo	11
2015-08-03 00:00:00	Neutro	12
2015-08-03 00:00:00	Negativo	3

Imagen 4.17 Dataset reporte Análisis de Sentimiento en Tiempo Real

La tabla **o\_Analisis\_Sentimiento** simplemente levanta los datos analizados y consolidados de la base de datos, sin realizar ningún tratamiento o modificación.

La tabla **Query1** realiza un pivot sobre la tabla **o\_Analisis\_Sentimiento** dejando fija la columna de la Fecha\_Analisis y generando una tupla por cada sentimiento (Positivo, Negativo, Neutro) y totalizando la cantidad de menciones para cada uno por fecha.

Esto fue necesario para poder desplegar la información en la gráfica de barras apiladas (ver sección Solución Funcional).

La tabla **o\_Tweets\_Colectate** se utiliza para desplegar la información de fecha, usuario y tweet según lo seleccionado en el *dashboard*, permitiéndole a ColectaTe tomar acciones en tiempo real.

## 4.8. Diseño aplicación móvil

El trabajo se realizó en dos partes. El primer objetivo fue detectar desde qué dispositivo se está consumiendo el sitio (*smartphone*, *tablet*, *pc*) y en base a este dato, la segunda parte se concentra en redirigir al consumidor al sitio que se desea mostrar. En este caso, no se redirecciona a otra página, sino que se muestran u ocultan atributos del *index.html* (página principal).

Primero, se descarga e instala un archivo (*Mobile-Detect.php*) en el servidor y se incluye en el "*index.html*". Esto permite detectar desde que dispositivo (*smartphone* o *tablet*) se está conectando el usuario.

Para la segunda parte, se incluyó código PHP en la página principal "*index.html*" para que un usuario pueda visualizar lo que corresponde según el dispositivo en el que se encuentra. Para información detallada sobre el código modificado ver Anexo 19 - Detección de dispositivos.

**Mobile-Detect.php** es una clase PHP utilizada para detectar dispositivos móviles. Utiliza la cadena de agente combinada con cabeceras HTTP específicas para detectar el entorno móvil.

En la Imagen 4.18 se puede apreciar el diseño de forma gráfica.

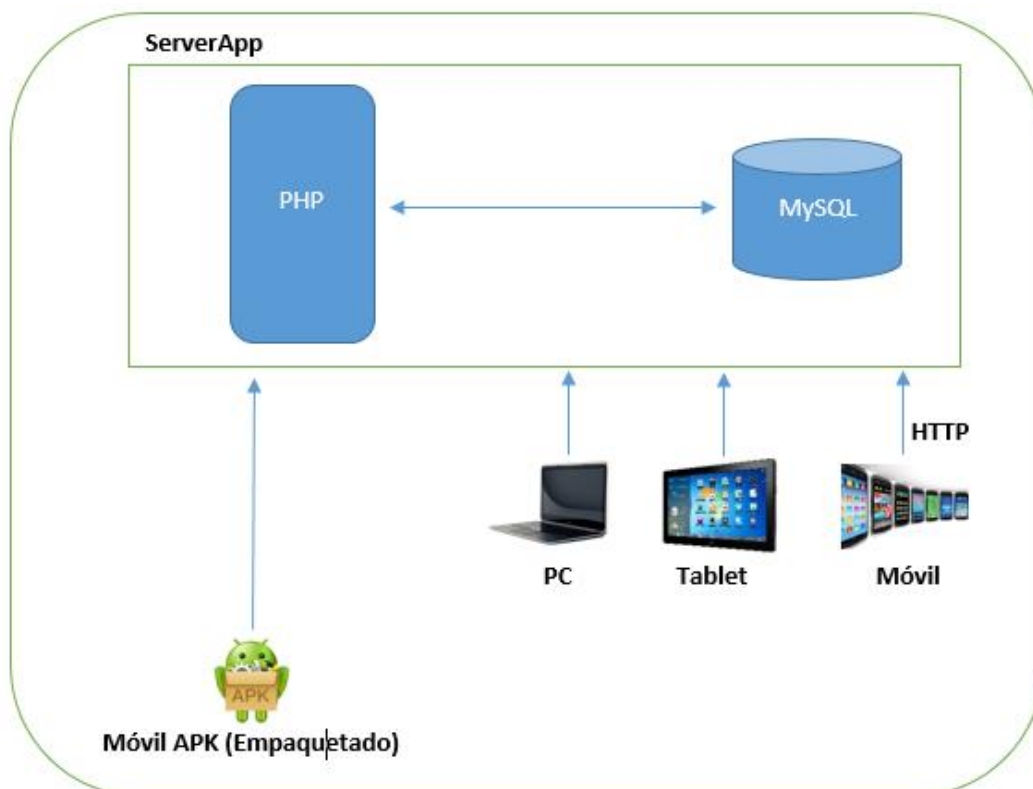


Imagen 4.18 Arquitectura aplicación móvil

## 4.9. Justificación de la arquitectura y tecnologías

### 4.9.1. Plataforma de Big Data

Para tomar una decisión sobre la plataforma a utilizar, se analizaron diferentes herramientas teniendo en cuenta todas las etapas del proceso de *Big Data*: extracción, almacenamiento, análisis y visualización.

Existe una gran variedad de plataformas, licenciadas y *open source*, con diferentes enfoques, como por ejemplo algunas con más énfasis en la extracción, otras en la visualización.

Según el análisis basado en los factores mencionados a continuación, **se seleccionó la plataforma KNIME** ya que con el alcance inicial definido es la herramienta que más se adecúa a nuestras necesidades:

- ✚ Ejecución de operaciones de ETL (extracción)
- ✚ Conectividad para almacenamiento, unificación con la plataforma seleccionada
- ✚ Funcionalidad para realizar procesamiento y análisis de datos
- ✚ Visualización según el objetivo planteado para usuario de negocio

Para esta selección, se priorizaron los puntos anteriormente nombrados más que el tipo de licenciamiento, asegurando la íntegra ejecución de todo el proceso *Big Data*.

Para ver la comparativa técnica de las plataformas de *Big Data* analizadas, ver 9.2.1 - Plataformas de Big Data.

### 4.9.2. Plataforma de almacenamiento

Desde el punto de vista del almacenamiento, es importante considerar el tipo de información que se necesita persistir y el momento (o paso) dentro del proceso de *Big Data* cuando se necesita hacerlo.

Existen tres grupos de información:

- ✚ Información paramétrica: este tipo de información es utilizada durante el proceso de *Big Data* para comparar, calcular, filtrar y ordenar los resultados intermedios de cada fase del proceso. El volumen de esta información es escaso y puede ser actualizado previo a la ejecución del proceso, durante la ejecución por el propio proceso o posterior al mismo.
- ✚ Información cruda o en proceso de transformación: este tipo de información refiere a la extraída de las fuentes de datos diversas y que para su procesamiento es requerido gran cantidad de espacio en disco y memoria, dado que se accede y modifica durante la ejecución del proceso, luego es descartada.
- ✚ Información procesada o analizada: es la información resultante de la ejecución finalizada del proceso de *Big Data*. La misma se almacena en los últimos pasos del proceso y tiene un carácter de información consolidada.

Para la información cruda, la plataforma de *Big Data* seleccionada (Knime) cuenta con mecanismos y herramientas propios de almacenamiento de esta información.

En cuanto a la información paramétrica y la información analizada, se optó por la **utilización del motor de bases de datos MySQL**, debido a los siguientes factores:

- ✚ Tipo de licenciamiento
- ✚ Límite de tamaño de base de datos y/o tablas
- ✚ Límite de utilización de memoria
- ✚ Límite de utilización de procesadores
- ✚ Unificación de tecnología con las ya utilizadas por ColectaTe

ColectaTe actualmente utiliza MySQL como motor de base de datos para sus sistemas.

Para ver la comparativa técnica de los motores de bases de datos analizados, ver 9.2.2 - Plataformas de almacenamiento.

### 4.9.3. Decisiones de diseño

Se presentan a continuación las decisiones de diseños más relevantes que se tomaron en cuenta para el desarrollo de la solución.

#### 4.9.3.1. Big Data

##### Decisión de guardar o no datos crudos en la extracción de datos

	Guardar	No Guardar
Necesidad de reprocesar los datos ya extraídos	+1	-1
Requerimiento de espacio de almacenamiento	-1	+1
<i>Overhead</i> e impacto en la performance del proceso	-1	+1
Flexibilidad a la hora de la implementación de proceso	+1	-1
Criticidad / necesidad de los datos crudos	+1	-1
<b>TOTAL</b>	<b>+1</b>	<b>-1</b>

Tabla 4.6 Decisión guardar datos

Conclusión: en base al análisis se decide guardar los datos.

##### Decisión de separación de procesos o unificación de los mismos

Realizar un único proceso que extraiga los datos y luego se ejecuten las fases de procesamiento y análisis en paralelo de cada requerimiento o realizar procesos por separado para cada requerimiento

	Separar	No Separar
Duplicación de datos extraídos	-1	+1
Mantenibilidad	+1	-1
Eficiencia extracción	-1	+1
Eficiencia análisis	+1	-1
Escalabilidad	+1	-1
<b>TOTAL</b>	<b>+1</b>	<b>-1</b>

Tabla 4.7 Decisión separación procesos

Conclusión: en base al análisis se decide separar los procesos.

#### 4.9.3.2. Aplicación móvil

Para el diseño de la aplicación móvil, se consultó con el experto Nicolás Fornaro. Durante la reunión se discutieron las diferentes alternativas de desarrollo, las cuales se redujeron a las 3 siguientes:

- ✚ Desarrollo nativo para cada plataforma
- ✚ Desarrollo web multiplataforma
- ✚ Aplicación web empaquetada para móvil

Desarrollo nativo para cada plataforma: implica desarrollar una aplicación para cada plataforma seleccionada (Android, iOS, Windows Phone) por separado. Esto requiere conocimientos técnicos específicos de los *frameworks* para desarrollo de cada plataforma lo que brinda la posibilidad de usar los recursos locales para mejorar la experiencia del usuario.

Desarrollo web multiplataforma: implica desarrollar una única aplicación web que se ejecuta en todas las plataformas seleccionadas a través de un *framework* de terceros; siendo éste el que se encarga de adaptar el *look and feel* en cada plataforma de ejecución.

Aplicación web empaquetada para móvil: implica generar una única aplicación que ejecuta un navegador apuntando hacia el sitio web que se desplegará en el mismo. El usuario no percibe que está sobre un navegador, dado que no se ven las barras y comandos que normalmente se visualizan en ellos.

Las 3 presentan sus ventajas y desventajas.

	Nativo	Web Multiplataforma	Empaquetado
Mejora la experiencia del usuario mediante utilización de las funcionalidades del teléfono como cámara, gps y sms.	+1	+1	-1
Conocimiento de las tecnologías	-1	-1	+1
Multiplataforma	-1	+1	-1
Visualización en pantalla ( <i>look and feel</i> )	+1	+1	-1
Tiempo de desarrollo y gestión	-1	-1	+1
Fecha límite del proyecto	-1	-1	+1
Prioridad en el negocio	-1	+1	-1
Prioridad en el proyecto	-1	-1	+1
Complejidad del desarrollo	-1	-1	+1
<b>TOTAL</b>	<b>-5</b>	<b>-1</b>	<b>+1</b>

Tabla 4.8 Decisión estrategia aplicación móvil

**Conclusión:** en base al análisis se decide seguir la estrategia de aplicación web empaquetada para móvil. En adición a esto, ColectaTe nos manifestó su interés por perseguir la estrategia de empaquetado dado que la aplicación móvil no es prioritaria en este momento y los costos de mantenimiento deben ser mínimos (basándose en que ya cuentan con el conocimiento necesario para mantener una aplicación empaquetada y no una aplicación nativa).

## 5. Solución Funcional

### 5.1. Introducción

Desde el punto de vista funcional, la solución generada se divide en dos partes: la aplicación móvil y los procesos Big Data. Estos últimos se dividen en 2 grupos:

- ✚ Procesos *batch*
- ✚ Procesos en tiempo real

Ambos grupos cuentan con componentes de extracción, análisis y reportes, según se describen en la sección de Diseño Arquitectónico del presente documento.

En el caso de los procesos *batch*, cada componente es ejecutado de forma independiente y no depende de otro, si bien, para que existan datos para reportar, se debe primero analizar, y para analizar de deben primero extraer.

Para los procesos en tiempo real, la independencia es la misma, salvo que para ver los resultados en tiempo real, los componentes deben estar corriendo en paralelo.

### 5.2. Ejecución de Procesos

Para el acceso de los usuarios se proporciona una interfaz web, la cual es accesible mediante el siguiente *link*, donde localhost se debe reemplazar por el nombre de red del servidor donde se encuentra instalado Knime:

<http://localhost:8080/com.knime.enterprise.server>

Luego de ingresar usuario y contraseña (Imagen 5.1)



Imagen 5.1 Login Knime

Se debe proceder a seleccionar un proceso que se desee ejecutar desde el menú lateral (Imagen 5.2)



Imagen 5.2 Menú procesos servidor Knime

Luego presionar el botón *Start* para ejecutar el proceso (Imagen 5.3)

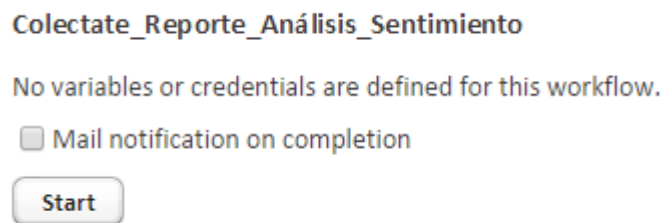


Imagen 5.3 Iniciar ejecución proceso en Knime

### 5.3. Agendado de Procesos

Los procesos de extracción y análisis deben ser agendados en el servidor para que corran de forma recurrente uno o más días en específico a una hora dada sin intervención manual.

El agendado se realiza desde la interfaz de escritorio del servidor, realizando clic derecho sobre el proceso que se desea agendar y eligiendo la opción *Execute* (Imagen 5.4)

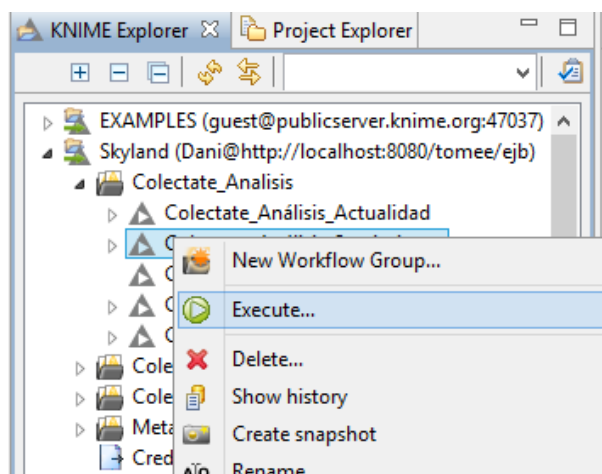


Imagen 5.4 Menú agendar procesos Knime

Por último se debe programar la ejecución (Imagen 5.5)

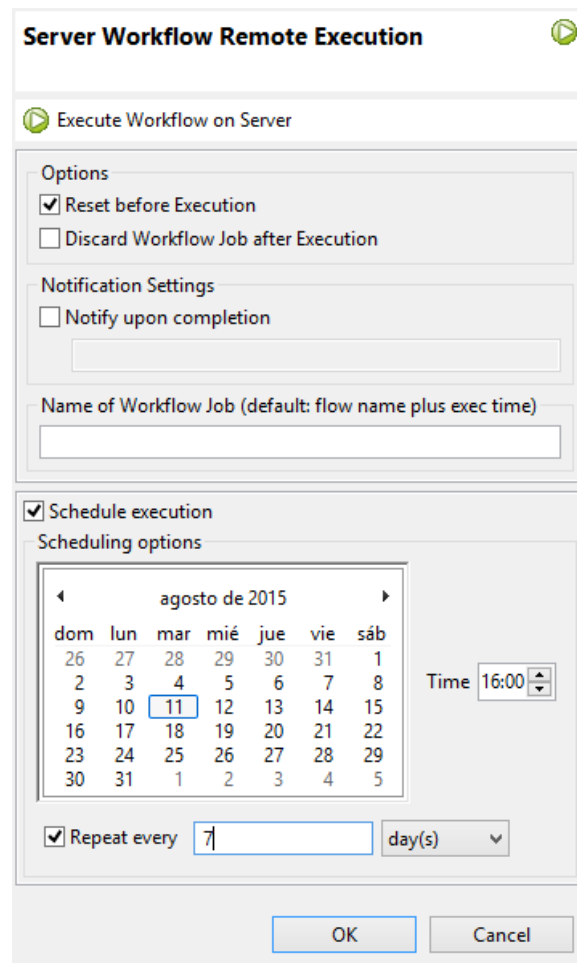


Imagen 5.5 Panel agendar procesos Knime

## 5.4. Clasificación de Sentimiento

Uno de los pasos de los módulos de análisis de los procesos es la clasificación del sentimiento. Esto se hace tanto para presentar los resultados finales de forma sencilla y agrupada según la opinión de los usuarios o para filtrar los resultados en los casos que solo interesa analizar resultados de una de las clasificaciones.

El detalle técnico del proceso de clasificación de sentimiento se puede encontrar en la sección de Diseño Arquitectónico del presente documento.

A nivel funcional, el sentimiento se clasifica y agrupa en 3 categorías:

- ✚ Positivo: representa las opiniones favorables o positivas de los usuarios
- ✚ Negativo: representa las opiniones desfavorables o negativas de los usuarios
- ✚ Neutro: representa las opiniones de los usuarios que no muestran inclinación a alguna de las anteriores

## 5.5. Análisis Sentimiento

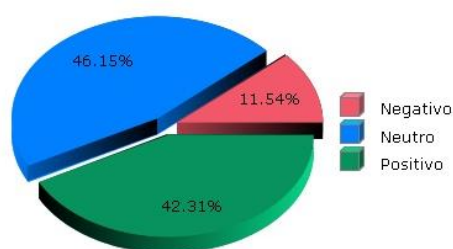
El reporte de análisis de sentimiento (*sentiment analysis*) permite identificar de forma sencilla la percepción general de los usuarios de ColectaTe según lo que comentan en las redes sociales.

El reporte se divide en dos áreas, en la primera (Imagen 5.6) se muestra el resumen consolidado del último período de análisis, agrupando los resultados según el sentimiento.



### Reporte de Analisis de Sentimiento

#### Resumen semanal



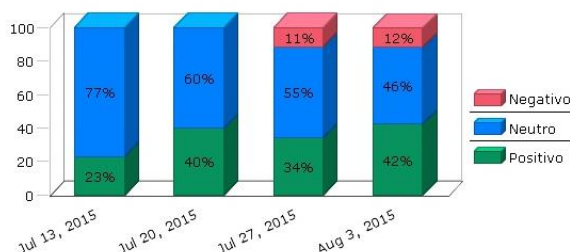
Fecha ejecución:	Aug 3, 2015
Total extraídos:	26
Total valorados:	26
Positivos:	11
Neutros:	12
Negativos:	3

Imagen 5.6 Reporte análisis sentimiento gráfica torta

En la gráfica de torta se puede apreciar el porcentaje de cada clasificación sobre el total de opiniones relevadas en el período.

En la tabla contigua se encuentran los datos generales, como la fecha del análisis, el total de opiniones extraídas, el total de valoradas y la cantidad de opiniones de cada clasificación.

#### Tendencia últimas 5 semanas



Fecha ejecución	Total extraídos	Total valorados	Positivos	Neutros	Negativos
Aug 3, 2015	26	26	11	12	3
Jul 27, 2015	44	44	15	24	5
Jul 20, 2015	25	25	10	15	0
Jul 13, 2015	31	26	6	20	0

Imagen 5.7 Reporte análisis sentimiento gráfica barras

En la segunda área del reporte (Imagen 5.7) se encuentra la tendencia del análisis de sentimiento. En la misma se pueden apreciar hasta un máximo de 5 análisis anteriores, ordenados de forma decreciente por fecha, comenzado desde el actual.

## 5.6. Oportunidades de Mejora

El reporte de oportunidades de mejora presenta la opinión de los usuarios sobre ColectaTe en Twitter, de forma que se pueda realizar una trazabilidad entre las mismas y las áreas o servicios ofrecidos por la empresa.

La información desplegada en el reporte corresponde únicamente al último período de análisis y se muestra en dos formatos.

El primer formato es en la sección *Tag Cloud* (Imagen 5.8), donde listan las palabras de cada una de las opiniones de los usuarios y se presentan en forma de nube, donde las más mencionadas (con mayor frecuencia de repetición entre las opiniones) se despliegan con mayor tamaño y hacia el centro de la nube.



## Reporte de Oportunidades de Mejora

### Resumen semanal - Tag Cloud

Colaboren Creamos Acabo  
 Encontré Gente hablaron web  
 Necesito invito recomendar forma visitar  
 lindo recomiendo colaborar práctico brinda sitio  
 servicio crear Colectate colectas probaron  
 lindos todos online colectivo colecta bueno  
 Participé más Caninos fácil participar podamos  
 conocer Utilicé nueva Colecta opción bien pagina  
 invitamos alguien nuevo animalitos nuestro

Fecha ejecución:	Aug 3, 2015
Total extraídos:	26
Total valorados:	26

Imagen 5.8 Reporte oportunidades de mejora Tag Cloud

A su vez, las palabras con significado positivo se muestran en color verde, las palabras negativas en color rojo, mientras que las neutras en color azul.

### Resumen semanal - Comentarios Positivos

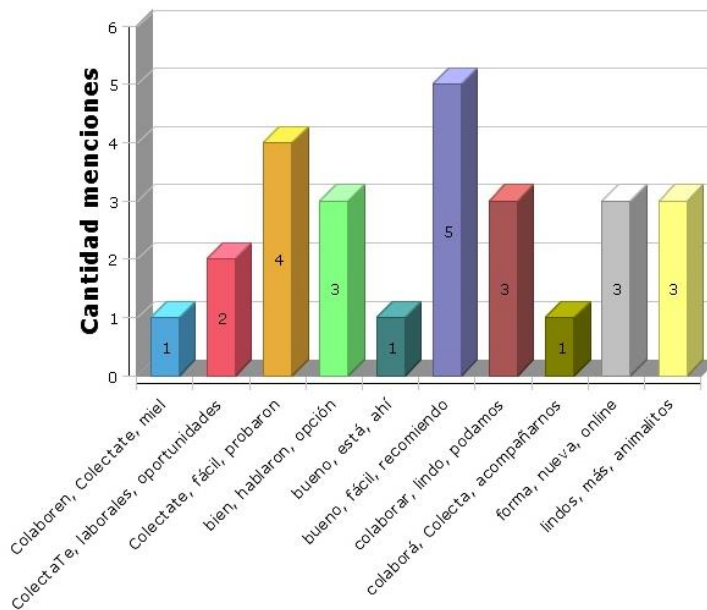


Imagen 5.9 Reporte oportunidades de mejora gráfica barras

En el segundo formato los resultados se muestran en formato de gráfica de barras (Imagen 5.9), donde cada barra representa un tópico (ver apartado de Diseño Arquitectónico para definición de tópico y forma de generación) y su cantidad de menciones.

Se muestra además una tabla con la misma información de la gráfica para facilitar su lectura (Imagen 5.10).

Con el uso de tópicos se pretende aportar un contexto a cada palabra y así facilitar la identificación de las oportunidades de mejora, es por esto que los resultados en este caso se dividen en dos secciones, los tópicos positivos y neutros por un lado, y los negativos por otro.

Oportunidad de Mejora	Cantidad de Menciones
bueno, fácil, recomiendo	5
Colectate, fácil, probaron	4
forma, nueva, online	3
bien, hablaron, opción	3
lindos, más, animalitos	3
colaborar, lindo, podamos	3
ColectaTe, laborales, oportunidades	2
bueno, está, ahí	1
colaborará, Colecta, acompañarnos	1
Colaboren, Colectate, miel	1

Imagen 5.10 Reporte oportunidades de mejora tabla

## 5.7. Análisis de Tendencias

El reporte de Análisis de Tendencias trata de identificar y detectar los temas de los que las personas están hablando en las redes sociales.

La información desplegada en el reporte corresponde únicamente al último período de análisis y se muestra en dos formatos.

El primer formato es en la sección *Tag Cloud* (Imagen 5.11), donde listan las palabras de cada una de las opiniones de los usuarios y se presentan en forma de nube, donde las más mencionadas (con mayor frecuencia de repetición entre las opiniones) se despliegan con mayor tamaño y hacia el centro de la nube.



Imagen 5.11 Reporte análisis de tendencias tag cloud

A su vez, las palabras con significado positivo se muestran en color verde, las palabras negativas en color rojo, mientras que las neutras en color azul.

En el segundo formato los resultados se muestran en formato de gráfica de barras (Imagen 5.12), donde cada barra representa un tópico (ver apartado de Diseño Arquitectónico para definición de tópico y forma de generación) y su cantidad de menciones.

### Resumen semanal - Tendencias positivas

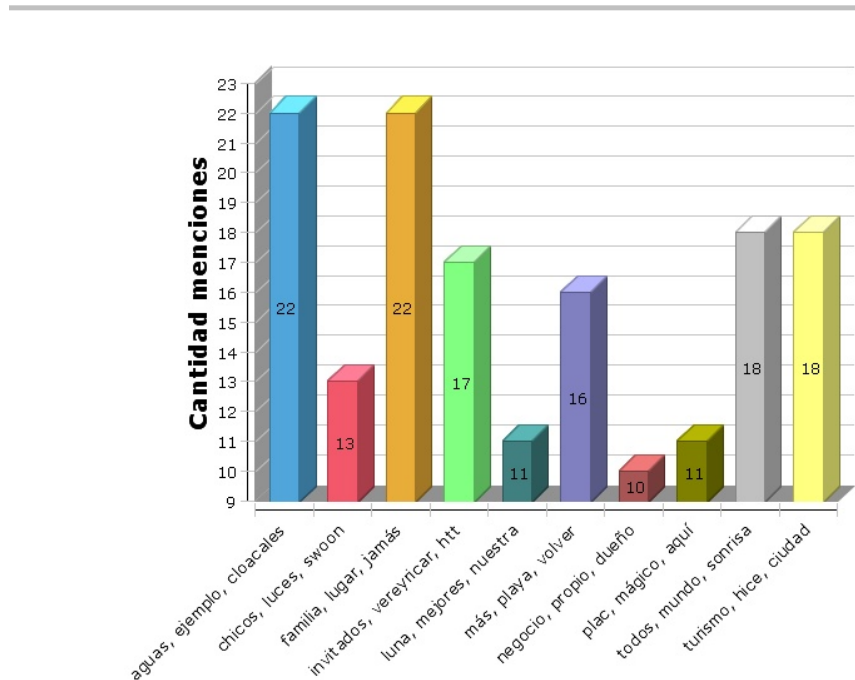


Imagen 5.12 Reporte análisis de tendencias gráfica barras

Se muestra además una tabla con la misma información de la gráfica para facilitar su lectura (Imagen 5.13).

Con el uso de tópicos se pretende aportar un contexto a cada palabra y así facilitar la identificación de las oportunidades de mejora, es por esto que los resultados en este caso se dividen en dos secciones, los tópicos positivos y neutros por un lado, y los negativos por otro.

Tendencia	Cantidad de Menciones
aguas, ejemplo, cloacales	22
familia, lugar, jamás	22
todos, mundo, sonrisa	18
turismo, hice, ciudad	18
invitados, vereyricar, htt	17
más, playa, volver	16
chicos, luces, swoon	13
plac, mágico, aquí	11
luna, mejores, nuestra	11
negocio, propio, dueño	10

Imagen 5.13 Reporte análisis de tendencias tabla

## 5.8. Análisis Actualidad

El reporte de Análisis Actualidad identifica la opinión de las personas en las redes sociales respecto a una serie de temas o palabras dadas, por ejemplo, ciertos productos o servicios, ciertas marcas, ciertos eventos, entre otros.

La información desplegada en el reporte corresponde únicamente al último período de análisis y se muestra en formato de gráfica de barras, separando las opiniones positivas de las negativas por color (Imagen 5.14).

**Resumen semanal**

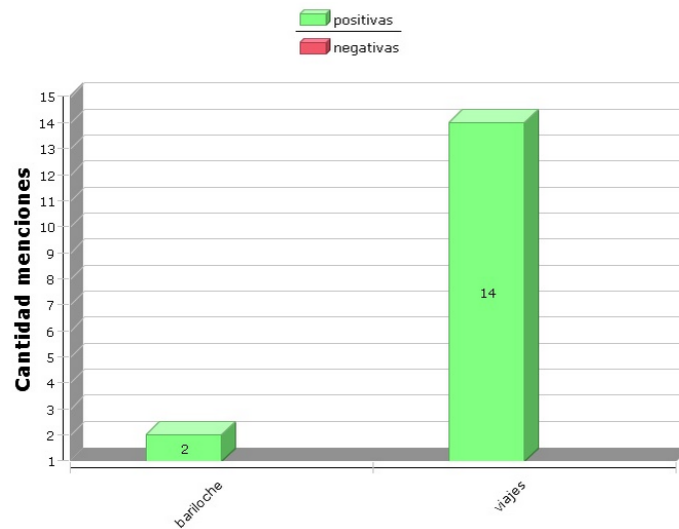


Imagen 5.14 Reporte análisis actualidad gráfica barras

Se muestra además una tabla con la misma información de la gráfica para facilitar su lectura (Imagen 5.15).

Fecha ejecución:	Aug 4, 2015
Total extraídos:	161
Total valorados:	31

Producto / Servicio	Cantidad Menciones Positivas	Cantidad Menciones Negativas
viajes	14	
bariloche	2	

Imagen 5.15 Reporte análisis actualidad tabla

### 5.9. Análisis Sentimiento en Tiempo Real

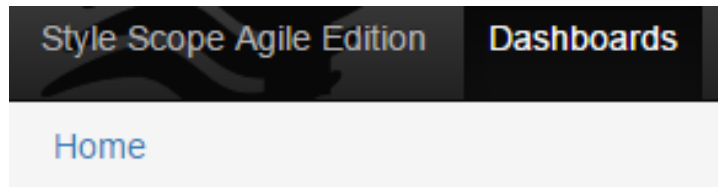
El reporte de análisis de sentimiento en tiempo real permite visualizar de forma sencilla y gráfica la opinión de los usuarios respecto a ColectaTe y de esta forma facilitar la toma de decisiones y realización de acciones en tiempo real.

Para acceder al *dashboard* se debe ingresar a la siguiente url, donde localhost es el nombre de red del servidor donde se ha instalado StyleScope Agile

http://localhost:35100/

Luego de logueado, se debe seleccionar el *dashboard* correspondiente (Imagen 5.16)

El *Dashboard* cuenta con 3 secciones principales: Tendencia de sentimiento, Resumen período actual (o seleccionado), Lista de opiniones tomadas en cuenta para la tendencia por período.



En la gráfica de barras apilada se muestra la tendencia del sentimiento respecto a ColectaTe según los análisis ejecutados de forma satisfactoria en el período delimitado por el selector de rango de fechas (Imagen 5.17)

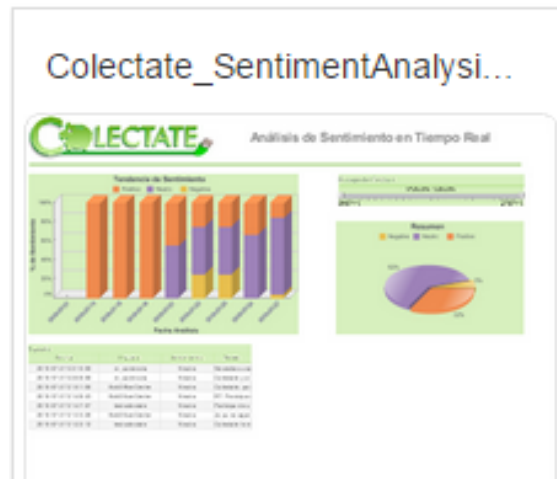


Imagen 5.16 Reporte análisis sentimiento en tiempo real - selección reporte

Al ser este un *dashboard* que muestra la información en tiempo real, a medida que nuevas opiniones son tomadas en cuenta, la última barra correspondiente a la fecha actual es actualizada con los nuevos porcentajes.

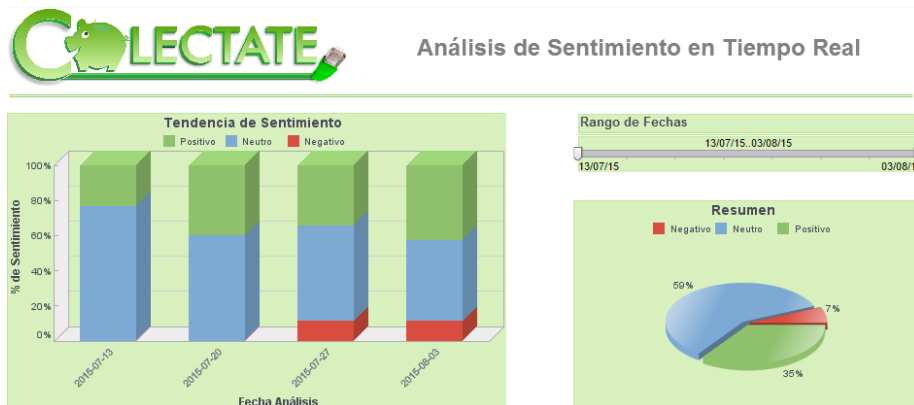


Imagen 5.17 Reporte análisis sentimiento en tiempo real - gráfica barras y torta

Esta actualización se realiza cada 5 segundos, pero a esto se le debe sumar el tiempo que demora la información en ser enviada por Twitter y ser procesada para calcular el sentimiento.

La gráfica de torta muestra un consolidado total del sentimiento según los datos seleccionados en la gráfica de tendencia. Si no se ha seleccionado ningún período en particular, se toma el total de los períodos por defecto.

Al mover el mouse por encima de las barras de la gráfica de tendencia, se puede observar una tabla emergente que muestra la información general tomada en cuenta para el análisis (Imagen 5.18).

Fecha	Total Extraídos	Total Valorados	Positivo	Neutro	Negativo
2015-07-13	31	26	6	20	0

Imagen 5.18 Reporte análisis sentimiento en tiempo real - datos adicionales

De la misma forma que las gráficas se actualizan, la tabla con las opiniones de los usuarios, la fecha de la misma, el sentimiento calculado y el usuario como tal, son mostrados y actualizados a medida que son analizados para permitir tomar acciones en tiempo real sobre las mismas (Imagen 5.19).

Tweets				
Fecha	Usuario	Sentimiento	Tweet	
2015-07-27 22:10:08	m_camerota	Neutro	Necesitas una opción online para crear una colecta? Colectate en :	
2015-07-27 22:09:23	m_camerota	Neutro	Colectate que buena opción para crear y participar de colectas online	
2015-07-27 22:08:58	m_camerota	Neutro	Colectate y crea tu primera colecta online	
2015-07-27 22:05:38	BckOfficeCentre	Neutro	Dónde está el experto de negocio de Colectate ?	
2015-07-27 21:51:06	BckOfficeCentre	Neutro	Colectate, participá y a daní le gustan las empanadas de la ribera SI SI SI dice que le encantan	
2015-07-27 21:48:43	BckOfficeCentre	Neutro	RT Participa de colectas 100% online ■ Colectate ahora..... :	
2015-07-27 21:47:57	testcolectate	Neutro	Participa de una colectate de Colecta-te y sentite mejor contigo mismo	

Imagen 5.19 Reporte análisis sentimiento en tiempo real - tweets

Es importante destacar, que de querer medir el impacto de una campaña dada sin tomar en cuenta el histórico, lo único que se debe hacer es reducir el filtro en el selector de fechas y dejar solamente la fecha actual seleccionada o en su defecto la fecha desde el comienzo de la campaña hasta la fecha actual.

## 5.10. Aplicación móvil

Para la aplicación móvil se realizó un empaquetado compuesto primeramente por las mismas funcionalidades ofrecidas actualmente en el sitio web de ColectaTe. Se realizaron algunas modificaciones en la página principal tomando en cuenta los atributos de calidad, particularmente el de usabilidad [13] [14], descritos en el apartado de diseño arquitectónico.

En la Imagen 5.22, se muestra la aplicación de ColectaTe instalada en un dispositivo móvil. En la Imagen 5.21, se muestra la pantalla de ingreso al sistema. Por último, en la Imagen 5.20 se muestra el menú de usuario una vez ingresado al sistema.

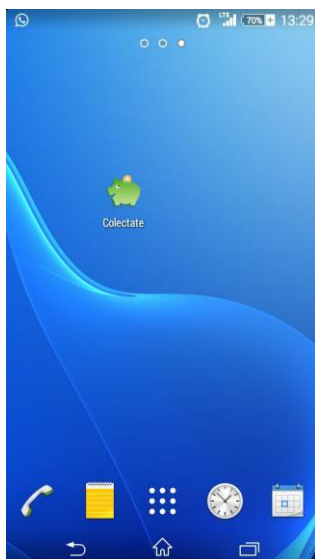


Imagen 5.22 ColectaTe instalado en un dispositivo móvil



Imagen 5.21 Pantalla ingreso a ColectaTe



Imagen 5.20 Menú de usuario ColectaTe

## 6. Gestión del Proyecto

### 6.1. Metodología

#### 6.1.1. Gestión del proyecto

Para llevar adelante la gestión de este proyecto se implementa el *framework* Scrum [15], en base en las siguientes características:

- ✚ No se cuenta con el conocimiento técnico en las diferentes tecnologías y plataformas necesarias para desarrollar los sistemas requeridos
- ✚ Se requiere constante interacción con el equipo de negocio de ColectaTe en formato de consultoría sobre el negocio para realizar la definición de los requerimientos

La Imagen 6.1 ejemplifica el ciclo Scrum utilizado:

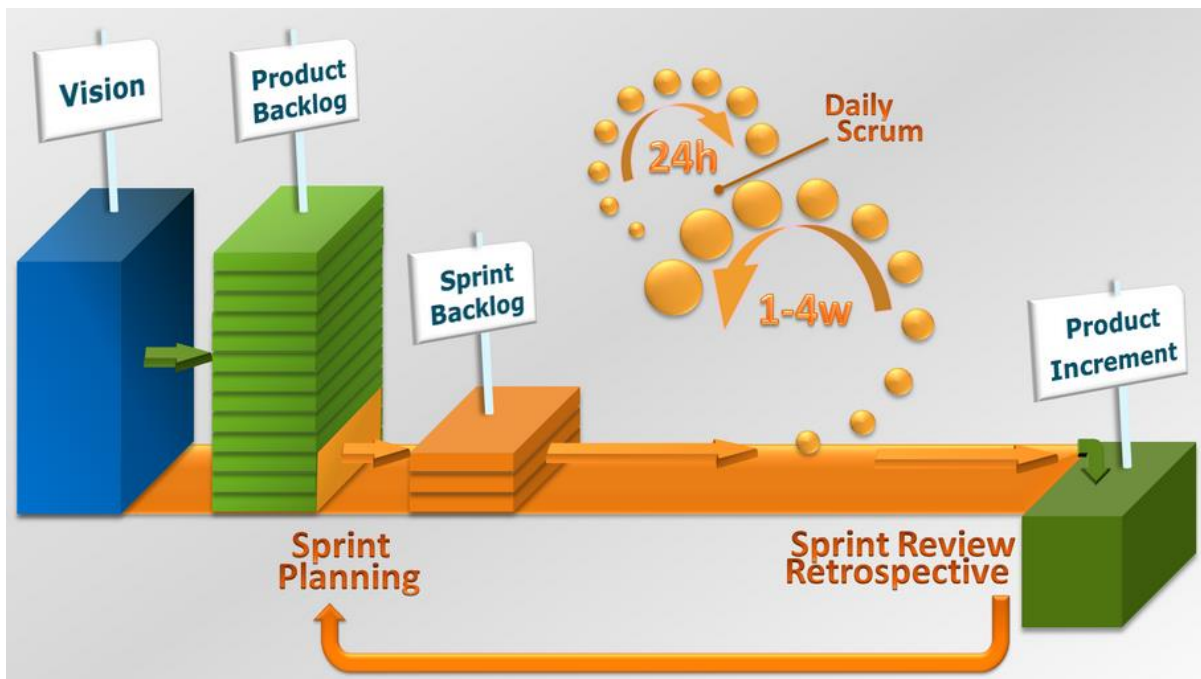


Imagen 6.1 Ciclo Scrum [16]

A partir de la visión del problema a resolver, se define el *Product Backlog*. Luego se define el *Sprint Backlog* previa estimación de los elementos del *Product Backlog*. La estimación se realiza utilizando la estrategia definida en este documento. Se comienza con el *Sprint* de 2 semanas, manteniendo reuniones periódicas, de sincronización entre los miembros del equipo. Al final del *Sprint*, se obtiene una versión del sistema, la cual es presentada en formato de Demo al equipo de negocio de ColectaTe para validación. Esta validación sirve como entrada para actualizar el *Product Backlog* para la siguiente iteración. Como último paso del ciclo, se realiza la retrospectiva, la cual sirve de entrada para redefinir la visión del problema si es necesario y comenzar un nuevo ciclo.

## 6.1.2. Desarrollo de software

Desde el punto de vista de *software*, el proyecto se divide en dos ramas:

- ✚ *Big Data*
- ✚ Aplicación móvil

La rama de lo relacionado a *Big Data* sigue una metodología de 4 fases [2] [3] cíclicas, que se ejemplifica en la Imagen 6.2.

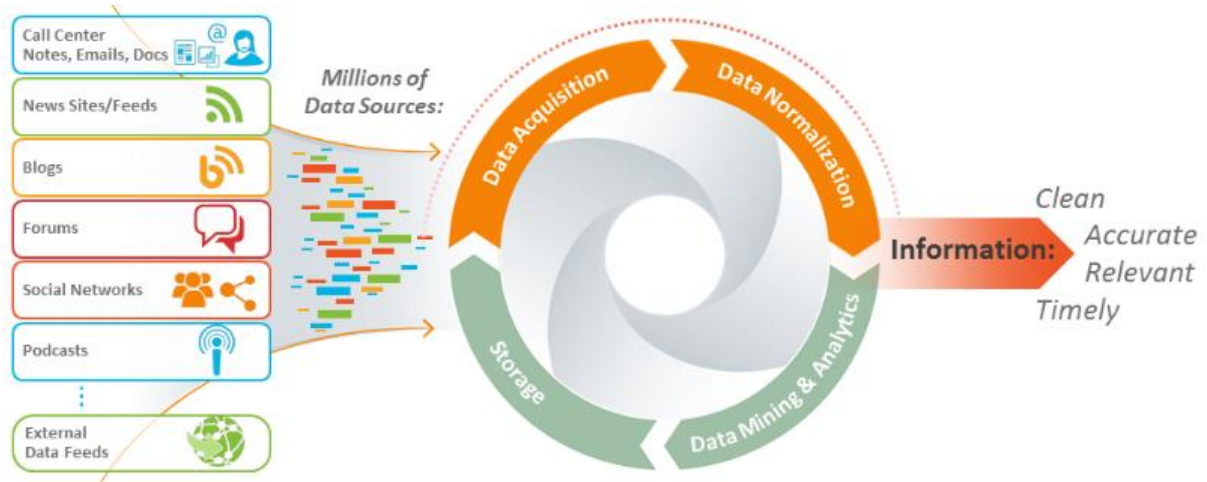


Imagen 6.2 Ciclo Big Data [17]

La fase inicial se corresponde con la identificación y obtención de la información de las diferentes y variadas fuentes de datos.

La segunda fase consiste en la normalización, consolidación y análisis de la información.

Luego de realizado el análisis se procede al almacenamiento de la misma en la tercer fase, repitiendo este ciclo hasta obtener los resultados esperados y así proceder a la cuarta fase, denominada visualización, y que consiste en distribuir a los usuarios el análisis de los datos de una forma que sea útil y sencilla para la toma de decisiones

Para la rama relacionada con el desarrollo de la aplicación móvil se utiliza el ciclo de vida del desarrollo ágil [18] que se ejemplifica en la Imagen 6.3.

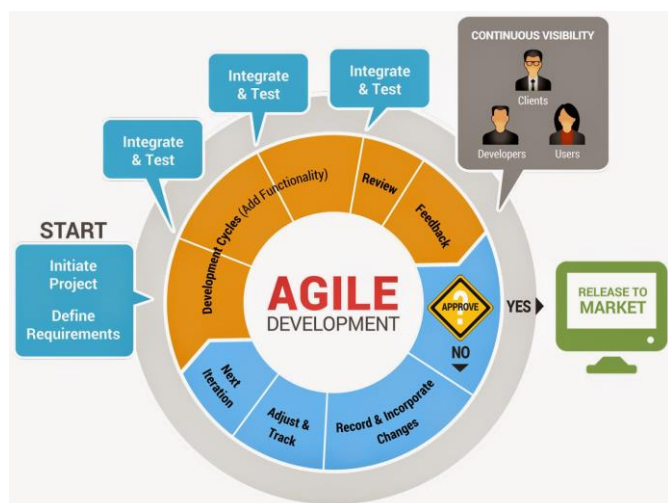


Imagen 6.3 Ciclo desarrollo ágil [43]

Desde el punto de vista del *software*, este ciclo de vida es un modelo iterativo incremental que permite agregar funcionalidades sobre la marcha del desarrollo. Esto requiere para su implementación exitosa, un diseño simple, lo que se logra entre otras cosas mediante actividades constantes de refactorización.

## 6.2. Gestión de esfuerzo

### 6.2.1. Estimación

Para la estimación de las *User Stories* se utilizaron tres técnicas encadenadas: (i) *Relative Mass Order*, (ii) *T-Shirt Size*, (iii) *Planning Poker*. Ver Anexo 8 - Técnica de estimación para más detalles sobre dicha metodología.

Ver Anexo 10 - Estimación *Product Backlog* para detalles de estimación de las *User Stories*.

### 6.2.2. Seguimiento esfuerzo

#### 6.2.2.1. Esfuerzo por tareas

A continuación se detalla el esfuerzo invertido en las diferentes tareas durante las etapas del proyecto.

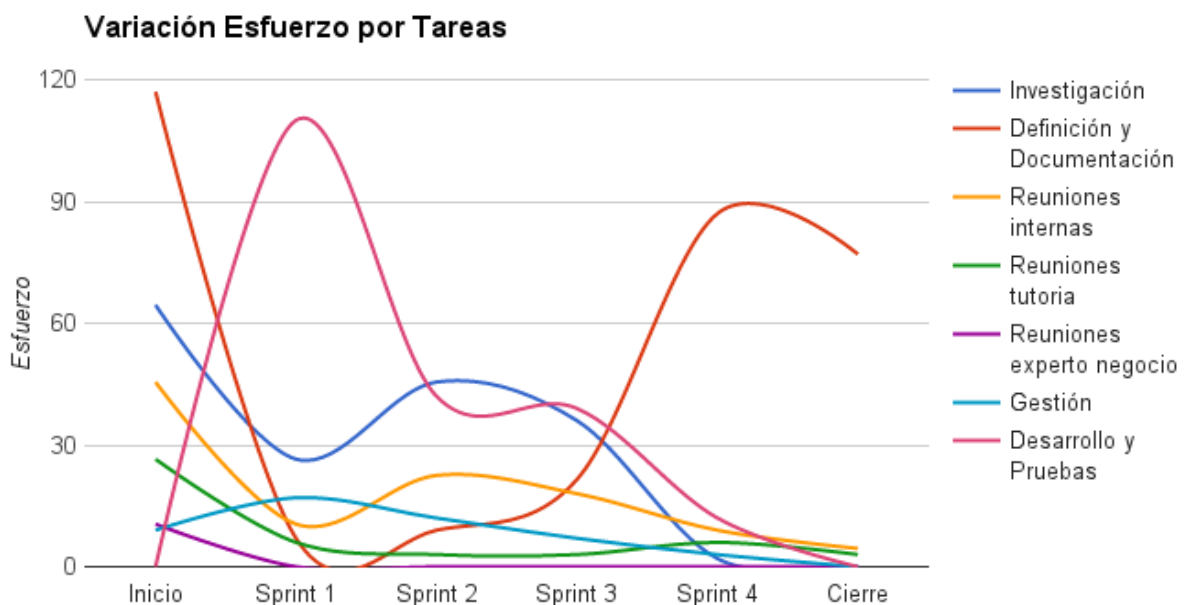
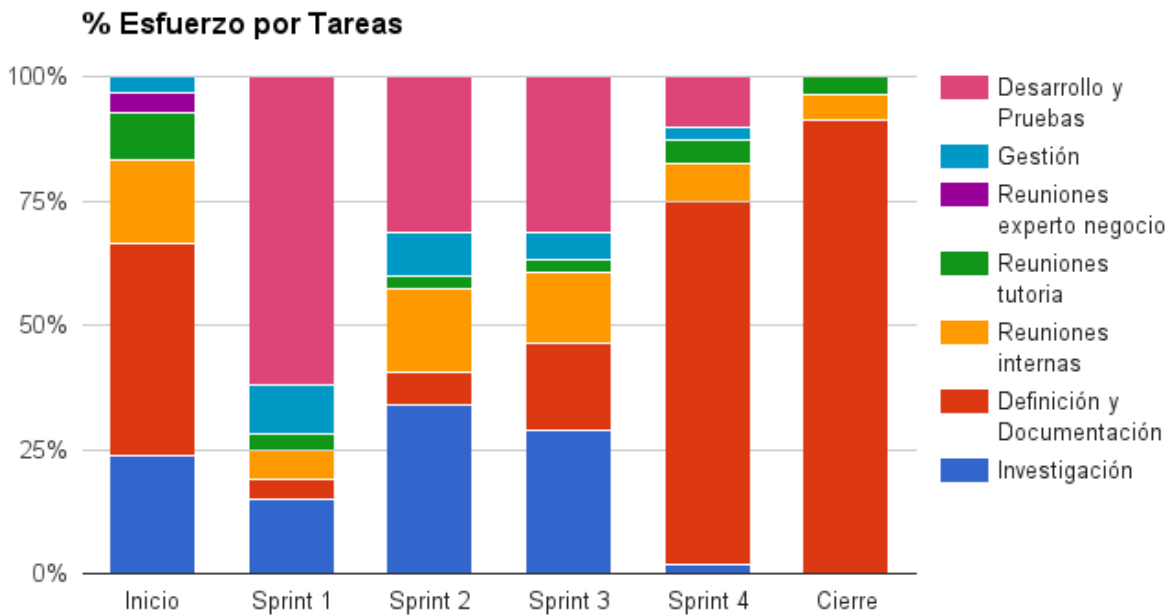


Imagen 6.4 Gráfica variación esfuerzo por tarea

De la Imagen 6.4 se hace notoria la reducción de horas invertidas por el equipo luego de finalizado el primer sprint. Eso se debió a que el equipo detectó que sería insostenible continuar invirtiendo esa misma cantidad de horas a lo largo de las restantes etapas del proyecto.

En la Imagen 6.5 se puede apreciar el esfuerzo invertido por tarea para cada fase en proporción al total de esfuerzo de cada fase.

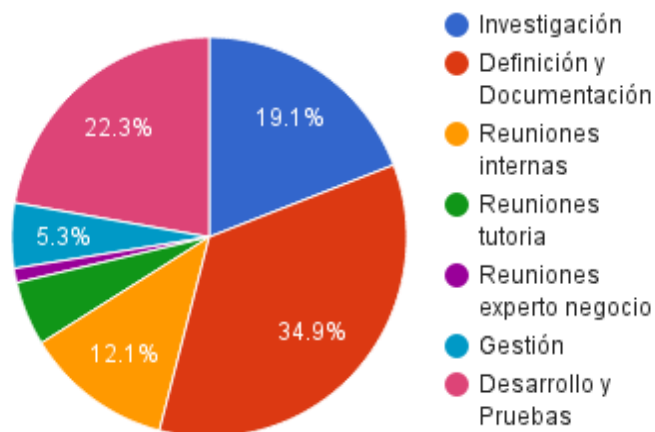


*Imagen 6.5 Gráfica % esfuerzo por tarea*

Los puntos importantes a destacar son:

- ✚ La cantidad de horas dedicadas a investigación se concentró sobre el inicio del proyecto y durante el segundo y tercer sprint, coincidiendo con la investigación de las herramientas y el cambio de alcance relacionado con el análisis en tiempo real.
- ✚ Durante el sprint 4, la mayor cantidad de horas se dedicó a documentación en lugar de tareas de desarrollo y pruebas como se podría esperar.

### Distribución Horas por Tarea



*Imagen 6.6 Gráfica distribución horas por tarea*

En la Imagen 6.6 se puede apreciar que a lo largo del proyecto, la mayor cantidad de horas fue invertida en definiciones y documentación, seguido por tareas de desarrollo y pruebas, quedando en tercer lugar tareas relacionadas a la investigación.

### 6.2.2.2. Esfuerzo por User Stories

A continuación se detalla la desviación entre las horas estimadas para las tareas de cada *Sprint* y las horas reales invertidas en dichas tareas. En la Imagen 6.7, se hace notoria la desviación del esfuerzo en el tercer sprint, lo cual se debió al foco del mismo en los requerimientos de análisis en tiempo real, lo cual fue necesario investigar, siendo dichas tareas de investigación, subestimadas, ya que en sprints previos no habían sido incluidas en la estimación.

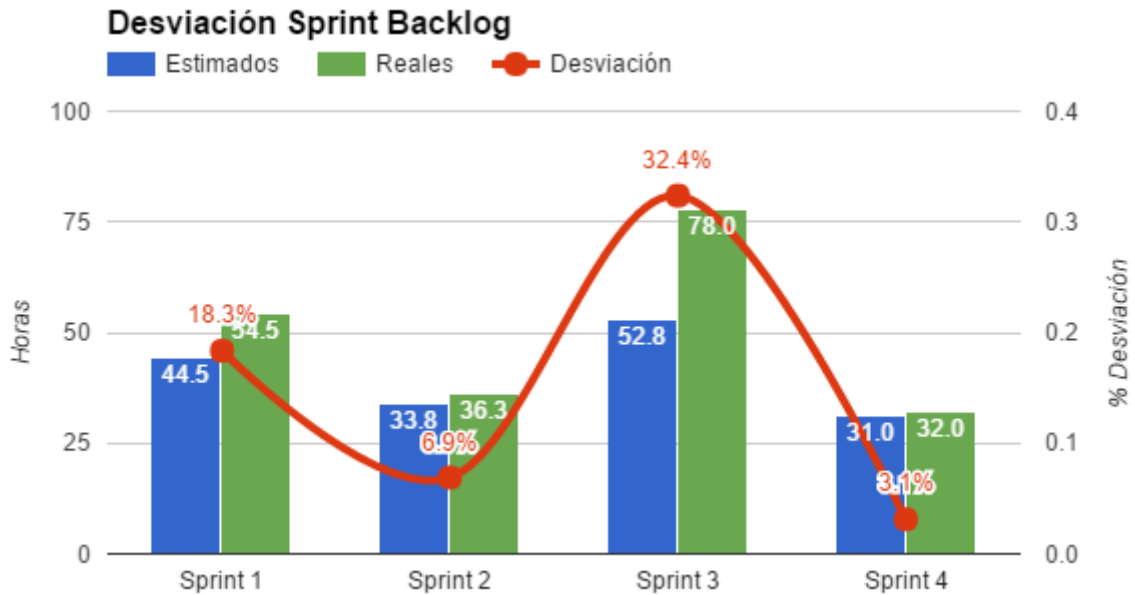


Imagen 6.7 Gráfica desviación sprint backlog

	Horas Estimadas	Horas Reales	Desviación
<b>Sprint 1</b>	44.5	54.5	18.3%
<b>Sprint 2</b>	33.8	36.3	6.9%
<b>Sprint 3</b>	52.8	78.0	32.4%
<b>Sprint 4</b>	31.0	32.0	3.1%
<b>Total</b>	<b>162.0</b>	<b>200.8</b>	<b>15.2%</b>

Tabla 6.1 Desviación de esfuerzo

Ver Anexo 11 - Estimación tareas *Product Backlog* para detalle de las tareas por *User Story*.

### 6.3. Gestión del cronograma

A continuación se detalla la planificación del calendario para el proyecto. Se debe tomar en cuenta que el calendario es fijo en su comienzo y fin, no pudiendo variar estas fechas pero si la planificación del período comprendido entre ambas.

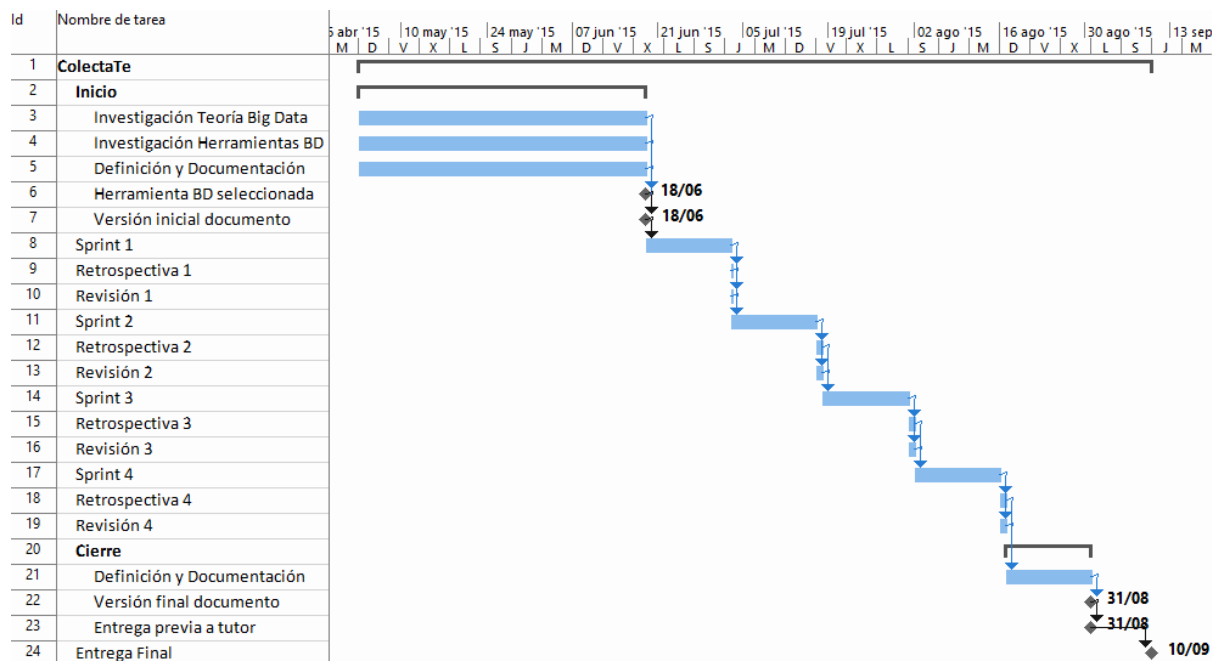


Imagen 6.8 Cronograma de proyecto

Nombre de tarea	Duración	Comienzo	Fin
ColectaTe	130,6 días	sáb 02/05/15	jue 10/09/15
Inicio	47 días	sáb 02/05/15	jue 18/06/15
Investigación Teoría <i>Big Data</i>	47 días	sáb 02/05/15	jue 18/06/15
Investigación Herramientas BD	47 días	sáb 02/05/15	jue 18/06/15
Definición y Documentación	47 días	sáb 02/05/15	jue 18/06/15
Herramienta BD seleccionada	0 días	jue 18/06/15	jue 18/06/15
Versión inicial documento	0 días	jue 18/06/15	jue 18/06/15
<i>Sprint 1</i>	14 días	jue 18/06/15	jue 02/07/15
Retrospectiva 1	0,5 días	jue 02/07/15	jue 02/07/15
Revisión 1	0,5 días	jue 02/07/15	jue 02/07/15
<i>Sprint 2</i>	14 días	vie 03/07/15	vie 17/07/15
Retrospectiva 2	0,5 días	vie 17/07/15	vie 17/07/15
Revisión 2	0,5 días	vie 17/07/15	vie 17/07/15
<i>Sprint 3</i>	14 días	sáb 18/07/15	sáb 01/08/15
Retrospectiva 3	0,5 días	sáb 01/08/15	sáb 01/08/15

Revisión 3	0,5 días	sáb 01/08/15	sáb 01/08/15
<i>Sprint 4</i>	14 días	dom 02/08/15	dom 16/08/15
Retrospectiva 4	0,5 días	dom 16/08/15	dom 16/08/15
Revisión 4	0,5 días	dom 16/08/15	dom 16/08/15
Cierre	14 días	lun 17/08/15	lun 31/08/15
Definición y Documentación	14 días	lun 17/08/15	lun 31/08/15
Versión final documento	0 días	lun 31/08/15	lun 31/08/15
Entrega previa a tutor	0 días	lun 31/08/15	lun 31/08/15
Entrega Final	0 días	jue 10/09/15	jue 10/09/15

Tabla 6.2 Cronograma de proyecto

En los Anexo 12 - *Sprint Review* y Anexo 13 - Retrospectiva para detalles de las ceremonias.

## 6.4. Gestión de alcance

### 6.4.1. Product Backlog

#### 6.4.1.1. User Stories por Sprint

<i>Sprint</i>	Cantidad	<i>User Stories</i>
<i>Sprint 1</i>	6	US-BD11, US-BD14, US-BD15, US-BD16, US-BD17, US-BD18
<i>Sprint 2</i>	7	US-BD01, US-BD03, US-BD06, US-BD07, US-BD08, US-BD09, US-BD10
<i>Sprint 3</i>	3	US-BD19, US-BD20, US-BD21
<i>Sprint 4</i>	16	US-AM01, US-AM02, US-AM03, US-AM04, US-AM05, US-AM06, US-AM07, US-AM08, US-AM09, US-AM10, US-AM11, US-AM12, US-AM13, US-AM14, US-AM15, US-BD22
Cancelado	5	US-BD02, US-BD04, US-BD05, US-BD12, US-BD13

Tabla 6.3 Alcance por sprint

Ver Anexo 9 - *Product Backlog* para listado detallado de las *User Stories*, sus criterios de validación y su prioridad.

#### 6.4.1.2. User Stories por prioridad

<i>Sprint</i>	Cantidad US	Cantidad SP
Alta	13	72

Media	3	38
Baja	16	19
Cancelado	3	0

Tabla 6.4 Prioridad user stories

Ver Anexo 9 - *Product Backlog* para listado detallado de las *User Stories*, sus criterios de validación y su prioridad.

### 6.4.2. Seguimiento alcance

En el siguiente gráfico se puede apreciar el progreso en el desarrollo de las *User Stories* (US) que conforman el *Product Backlog* (PB).

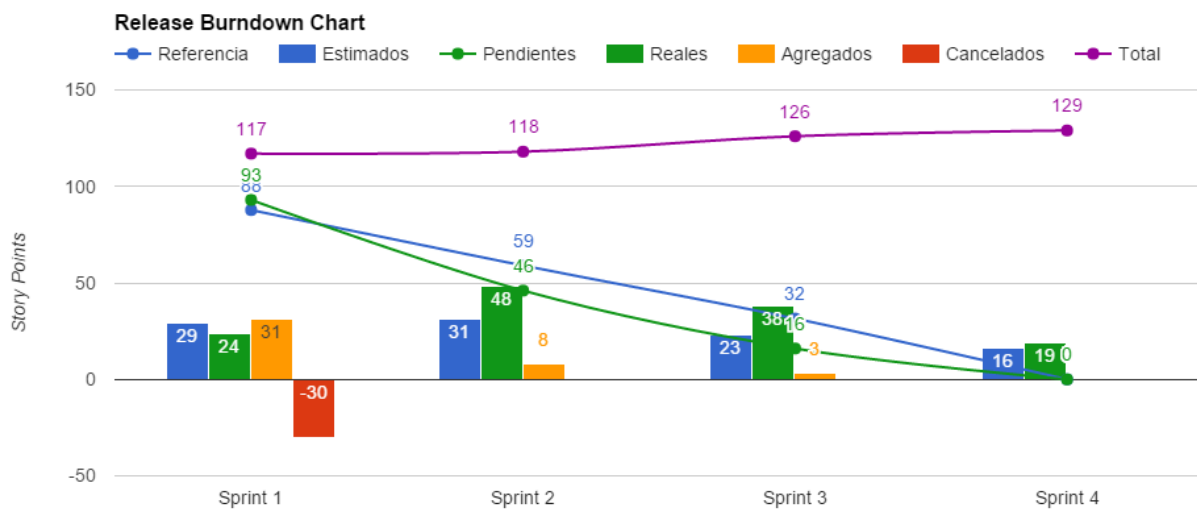


Imagen 6.9 Burndown Chart

	Referencia	Estimados	Pendientes	Reales	Agregados	Cancelados	Total
<b>Sprint 1</b>	88	29	93	24	31	-30	117
<b>Sprint 2</b>	59	31	46	48	8		118
<b>Sprint 3</b>	32	23	16	38	3		126
<b>Sprint 4</b>	0	16	0	19			129

Tabla 6.5 Seguimiento por sprint

Referencia	Comentario
—●— Referencia	SP estimados en el PB luego de concluido el <i>Sprint</i>
■ Estimados	SP estimados a desarrollarse en el <i>Sprint</i>
—●— Pendientes	SP restantes en el PB luego de concluido el <i>Sprint</i>





 Reales	SP reales desarrollados en el <i>Sprint</i>
 Agregados	SP agregados al finalizar el <i>Sprint</i>
 Cancelados	SP cancelados / eliminados al finalizar el <i>Sprint</i>
 Total	SP totales del PB en el <i>Sprint</i>

Tabla 6.6 Referencia Burndown chart

## Análisis









<b>Sprint</b>	<b>Comentario</b>
<i>Sprint 1</i>	<ul style="list-style-type: none"> <li> Se cancelaron 30 SP al finalizar el <i>Sprint</i> debido a la cancelación de las US relacionadas con Facebook</li> <li> Se agregaron 31 SP relacionadas con el nuevo requerimiento de análisis en tiempo real</li> <li> Con el fin de mantener la unicidad funcional de los procesos dentro del <i>Sprint</i> y desarrollar un ciclo completo (extracción, análisis y visualización), se desarrollaron menos SP de las mencionadas en la referencia.</li> </ul>
<i>Sprint 2</i>	<ul style="list-style-type: none"> <li> Se agregaron 8 SP luego de re-estimar el PB en base a la experiencia del primer <i>Sprint</i></li> <li> Con el fin de mantener la unicidad funcional de los procesos dentro del <i>Sprint</i> y desarrollar un ciclo completo (extracción, análisis y visualización), se desarrollaron más SP de las mencionadas en la referencia.</li> </ul>
<i>Sprint 3</i>	<ul style="list-style-type: none"> <li> Se agregaron 3 SP luego de re-estimar el PB en base a la experiencia del segundo <i>Sprint</i></li> <li> Con el fin de mantener la unicidad funcional de los procesos dentro del <i>Sprint</i> y desarrollar un ciclo completo (extracción, análisis y visualización), se desarrollaron más SP de las mencionadas en la referencia.</li> </ul>
<i>Sprint 4</i>	 Sin comentarios.

Tabla 6.7 Análisis resultados por sprint

## 6.5. Gestión de riesgos

### 6.5.1. Identificación de riesgos

<b>ID</b>	<b>Descripción Riesgo</b>
<b>RSK01</b>	No lograr cumplir con los acuerdos realizados de tiempo por retrasos en las validaciones de los requerimientos por parte del experto del negocio
<b>RSK02</b>	Requerimientos poco definidos o ambiguos debido a la falta de claridad en las respuestas recibidas durante la fase de relevamiento de requerimientos

<b>RSK03</b>	Cambios de alcance sobre la marcha del proyecto debido a que el tiempo real requerido para la completitud de los requerimientos no coincide con el tiempo definido para el proyecto
<b>RSK04</b>	No lograr cumplir con los acuerdos realizados de alcance y tiempo por falta de compromiso y apoyo del experto en el negocio
<b>RSK05</b>	No lograr cumplir con los acuerdos realizados de alcance, tiempo, calidad por falta de conocimiento de los integrantes de equipo en las tecnologías relacionadas
<b>RSK06</b>	Incumplimiento del alcance debido a la poca disponibilidad de horas del equipo
<b>RSK07</b>	Incremento en la cantidad de horas requeridas para la finalización del <i>Release</i> en comparación con las estimadas debido al re-trabajo por baja calidad del resultado
<b>RSK08</b>	No lograr cumplir con los acuerdos realizados de alcance, tiempo debido a cambios de alcance o cambio en prioridades
<b>RSK09</b>	No lograr cumplir con los acuerdos realizados de alcance, tiempo debido a estimaciones incorrectas
<b>RSK10</b>	No lograr cumplir con los acuerdos de calidad, alcance debido a la alta complejidad en la interacción del sistema con las redes sociales
<b>RSK11</b>	No lograr cumplir con los acuerdos de calidad, alcance debido a negación de permisos necesarios para la extracción de datos desde la redes sociales
<b>RSK12</b>	No lograr cumplir con los acuerdos realizados de alcance, tiempo debido a cambios sobre la marcha en la estrategia de desarrollo de la aplicación móvil (desarrollo vs empaquetado)
<b>RSK13</b>	No lograr cumplir con los acuerdos de calidad debido a alcanzar conclusiones obvias o esperadas en base al análisis resultante de los procesos
<b>RSK14</b>	No lograr cumplir con los acuerdos realizados de alcance, tiempo, calidad debido a complicaciones técnicas de la herramienta seleccionada para diseñar y ejecutar los procesos de <i>Big Data</i>
<b>RSK15</b>	No lograr cumplir con los acuerdos de calidad debido a alcanzar conclusiones erróneas en base a la mala calidad de los datos obtenidos y tratados de las redes sociales
<b>RSK16</b>	No lograr cumplir con los acuerdos realizados de alcance, calidad debido a la creación de leyes que restrinjan el uso de la información provista por los usuarios en las redes sociales
<b>RSK17</b>	No lograr cumplir con los acuerdos realizados de alcance, calidad debido a reducción de la disponibilidad de alguno de los miembros del equipo a causa de imprevistos como viajes laborales o enfermedades.

Tabla 6.8 Riesgos identificados

Para información detallada sobre el plan de riesgos ver Anexo 7 - Plan de gestión de riesgos.

## 6.5.2. Análisis cualitativo de riesgos

### 6.5.2.1. Magnitud

Inicio				Iteración 1				Iteración 2				Iteración 3				Iteración 4				
ID	I	PO	OT	Magnitud	I	PO	OT	Magnitud	I	PO	OT	Magnitud	I	PO	OT	Magnitud	I	PO	OT	Magnitud
01	3	0,2	0	0,6	3	0,2	0	0,6	2	0,2	0	0,4	2	0,2	0	0,4	0	0,0	0	0,0
02	3	0,2	1	0,6	3	0,2	1	0,6	2	0,2	1	0,4	2	0,2	1	0,4	0	0,0	0	0,0
03	4	0,2	1	0,8	4	0,2	1	0,8	3	0,2	1	0,6	3	0,2	1	0,6	0	0,0	0	0,0
04	4	0,2	0	0,8	3	0,2	0	0,6	3	0,2	0	0,6	3	0,2	0	0,6	0	0,0	0	0,0
05	5	0,6	1	3,0	5	0,6	1	3,0	5	0,6	1	3,0	3	0,4	1	1,2	3	0,2	1	0,6
06	4	0,0	1	0,0	4	0,0	1	0,0	3	0,0	1	0,0	3	0,0	1	0,0	0	0,0	0	0,0
07	2	0,4	1	0,8	2	0,8	1	1,6	2	0,8	1	1,6	2	0,4	1	0,8	2	0,2	1	0,4
08	4	0,2	1	0,8	4	0,6	1	2,4	4	0,6	1	2,4	3	0,2	1	0,6	0	0,0	0	0,0
09	4	0,8	1	3,2	4	0,6	1	2,4	3	0,4	1	1,2	2	0,4	1	0,8	0	0,0	0	0,0
10	4	0,4	0	1,6	4	0,4	0	1,6	4	0,4	0	1,6	0	0,0	0	0,0	0	0,0	0	0,0
11	5	0,6	0	3,0	5	1,0	0	5,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
12	3	0,2	1	0,6	3	0,2	1	0,6	3	0,2	1	0,6	3	0,2	1	0,6	0	0,0	0	0,0
13	2	0,4	2	0,8	2	0,4	2	0,8	2	0,6	2	1,2	3	0,6	2	1,8	3	0,4	2	1,2
14	4	0,2	1	0,8	4	0,6	1	2,4	4	0,6	1	2,4	0	0,0	1	0,0	0	0,0	0	0,0

<b>15</b>	2	0,4	2	0,8	2	0,4	2	0,8	2	0,2	2	0,4	2	0,2	2	0,4	2	0,4	2	0,8
<b>16</b>	3	0,2	2	0,6	3	0,2	2	0,6	3	0,2	2	0,6	3	0,2	2	0,6	0	0,0	0	0,0
<b>17</b>	3	0,2	0	0,6	3	0,2	0	0,6	3	0,2	2	0,6	3	0,2	2	0,6	0	0,0	0	0,0

Tabla 6.9 Magnitud de riesgos

### 6.5.2.2. Estrategia genérica

Inicio		Iteración 1		Iteración 2		Iteración 3		Iteración 4		
ID	Magnitud	Estrategia	Magnitud	Estrategia	Magnitud	Estrategia	Magnitud	Estrategia	Magnitud	Estrategia
<b>01</b>	0,6	Aceptar	0,6	Aceptar	0,4	Aceptar	0,4	Aceptar	0,0	Aceptar
<b>02</b>	0,6	Aceptar	0,6	Aceptar	0,4	Aceptar	0,4	Aceptar	0,0	Aceptar
<b>03</b>	0,8	Mitigar	0,8	Mitigar	0,6	Aceptar	0,6	Aceptar	0,0	Aceptar
<b>04</b>	0,8	Mitigar	0,6	Aceptar	0,6	Aceptar	0,6	Aceptar	0,0	Aceptar
<b>05</b>	3,0	Evitar	3,0	Evitar	3,0	Evitar	1,2	Mitigar	0,6	Aceptar
<b>06</b>	0,0	Aceptar	0,0	Aceptar	0,0	Aceptar	0,0	Aceptar	0,0	Aceptar
<b>07</b>	0,8	Aceptar	1,6	Mitigar	1,6	Mitigar	0,8	Aceptar	0,4	Aceptar
<b>08</b>	0,8	Mitigar	2,4	Evitar	2,4	Evitar	0,6	Aceptar	0,0	Aceptar
<b>09</b>	3,2	Evitar	2,4	Evitar	1,2	Mitigar	0,8	Aceptar	0,0	Aceptar
<b>10</b>	1,6	Mitigar	1,6	Mitigar	1,6	Mitigar	0,0	Aceptar	0,0	Aceptar
<b>11</b>	3,0	Evitar	5,0	Evitar	0,0	Aceptar	0,0	Aceptar	0,0	Aceptar

<b>12</b>	0,6	Aceptar	0,6	Aceptar	0,6	Aceptar	0,6	Aceptar	0,0	Aceptar
<b>13</b>	0,8	Aceptar	0,8	Aceptar	1,2	Aceptar	1,8	Mitigar	1,2	Mitigar
<b>14</b>	0,8	Mitigar	2,4	Evitar	2,4	Evitar	0,0	Aceptar	0,0	Aceptar
<b>15</b>	0,8	Aceptar	0,8	Aceptar	0,4	Aceptar	0,4	Aceptar	0,8	Aceptar
<b>16</b>	0,6	Aceptar	0,6	Aceptar	0,6	Aceptar	0,6	Aceptar	0,0	Aceptar
<b>17</b>	0,6	Aceptar	0,6	Aceptar	0,6	Aceptar	0,6	Aceptar	0,0	Aceptar

Tabla 6.10 Estrategias genéricas de riesgos

### 6.5.2.3. Gráfica de riesgos

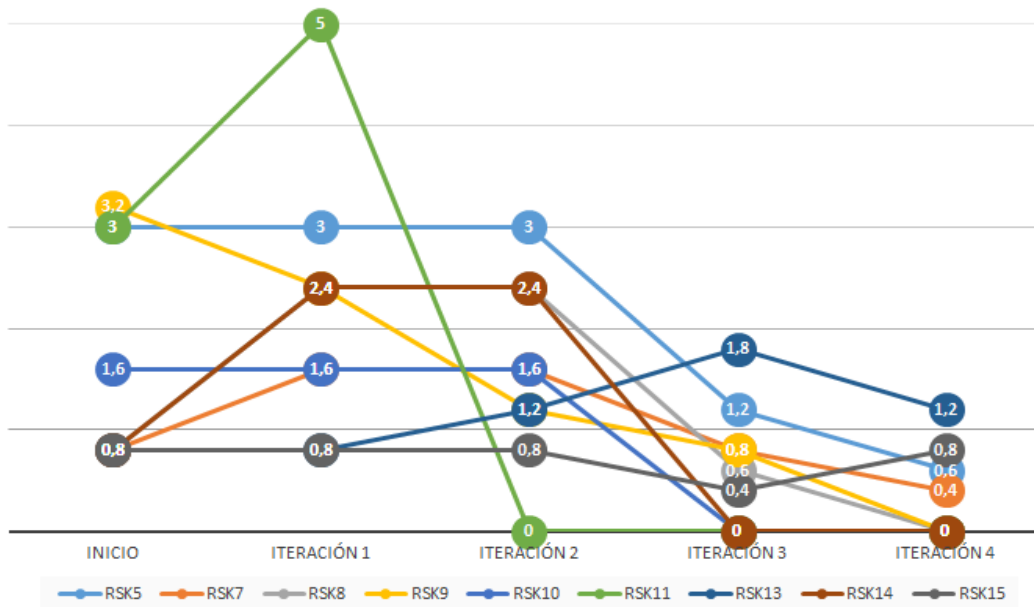


Imagen 6.10 Gráfica de riesgos

Para un detalle sobre la materialización del riesgo 11, ver Anexo 6 - Solicitud de permisos Facebook.

### 6.5.3. Disparadores, mitigación y contingencia

ID	Plan de acción
01	<u>Disparador</u> : Retraso de un 10% en las tareas de validación por parte del experto de negocio cuando pasan a formar parte del camino crítico del proyecto
02	<u>Disparador</u> : 50% de requerimientos en los que se haya tenido que realizar más de 3 interacciones con el experto de negocio para aclarar el requerimiento y lograr un entendimiento completo del mismo
03	<u>Disparador</u> : Desviación del +-20% en el avance del proyecto según el <i>Burndown Chart</i>
	<u>Mitigación</u> : Definir línea base del alcance y renegociar con el experto de negocio y el tutor al finalizar el primer <i>sprint</i>
	<u>Contingencia</u> : Acordar un nuevo alcance con el experto de negocio y tutor del proyecto
04	<u>Disparador</u> : 50% de las dudas no son respondidas en menos de 48 horas de enviadas o 2 reuniones de continuo sin la participación del experto de negocio
	<u>Mitigación</u> : Establecer y acordar el plan de comunicación con el experto de negocio y acordar los tiempos de respuesta previo al inicio de la fase de relevamiento, según se acordó en la carta de compromiso de la empresa
	<u>Contingencia</u> : Escalar el problema al tutor
05	<u>Disparador</u> : Desviación del +-20% en el avance del proyecto según el <i>Burndown Chart</i>
	<u>Mitigación</u> : Acelerar la curva de aprendizaje mediante la colaboración y capacitación conjunta con otros grupos de proyecto que se encuentran trabajando sobre temas similares. Definir un período de investigación al inicio del proyecto con el fin de comprender los desafíos, las posibles herramientas y generar un piloto que cubra los aspectos básicos para determinar la factibilidad técnica
	<u>Contingencia</u> : Invertir esfuerzo en investigación y capacitación por parte de los miembros del equipo. Evaluar un posible cambio de plataforma o utilización de herramientas complementarias que faciliten la tarea en cierto aspecto. Renegociar el alcance con el experto de negocio y tutor en base a las posibilidades técnicas del equipo
06	<u>Disparador</u> : Desviación del +20% del total de <i>Story Points</i> (en <i>User Stories</i> agregadas por terceros) al final de cada <i>Sprint</i> comparado con el total de <i>Story Points</i> al inicio del proyecto
	<u>Mitigación</u> : Acordar previamente con el experto de negocio y tutor que los cambios de alcance o prioridades serán incorporados en base a la capacidad de desarrollo del equipo ( <i>velocity</i> ) y el ajuste del <i>backlog</i> tomando en cuenta la duración fija del proyecto

	<u>Contingencia</u> : Redefinir el alcance en base a las prioridades del experto de negocio y tutor, tomando en cuenta que el proyecto tiene una duración fija y el equipo tiene una capacidad de desarrollo dada ( <i>velocity</i> )
07	<u>Disparador</u> : Desviación superior al 20% de esfuerzo en comparación al esfuerzo del <i>sprint</i> para la solución de errores de las <i>User Stories</i> entregadas en el mismo
	<u>Mitigación</u> : Verificar la calidad de la salida resultante de la ejecución de la tarea, una vez finalizada la misma
	<u>Contingencia</u> : Utilizar tiempo del colchón para retrabajo
08	<u>Disparador</u> : 40% de los requerimientos sufren cambios luego de comenzadas las iteraciones, o se agregan más de 20% de nuevos requerimientos en comparación al <i>sprint</i> luego de la <i>product review</i>
	<u>Mitigación</u> : Evaluar junto al experto de negocio / tutor si el cambio solicitado puede evitarse mediante la realización de un procedimiento diferente por parte del usuario
	<u>Contingencia</u> : Replanificar / renegociar el alcance del proyecto
09	<u>Disparador</u> : 20% de las tareas definidas para cada <i>sprint</i> no se logran realizar en el tiempo estimado (desviación de +- 10% por tarea)
	<u>Mitigación</u> : Estimar las tareas de cada <i>sprint</i> en base a la retroalimentación de <i>sprints</i> anteriores
	<u>Contingencia</u> : Utilizar tiempo del colchón estimado
10	<u>Disparador</u> : Si transcurrido el 50% del tiempo del primer <i>sprint</i> no se logró la interconexión y extracción de datos de las redes sociales
	<u>Mitigación</u> : Realizar pilotos antes del comienzo del primer <i>sprint</i> para identificar la factibilidad técnica de la interconexión con las redes sociales
	<u>Contingencia</u> : Evaluar un posible cambio de plataforma o utilización de herramientas complementarias que faciliten la tarea de interacción y extracción de datos desde las redes sociales
11	<u>Disparador</u> : Negación de permisos de forma reiterada (más de 4 solicitudes denegadas)
	<u>Mitigación</u> : Basar la solicitud de permisos en base a experiencias previas de solicitudes de terceros. Redactar las solicitudes de forma cuidadosa tratando de minimizar que sean rechazadas
	<u>Contingencia</u> : Evaluar la extracción de datos de redes sociales o fuentes alternativas
12	<u>Disparador</u> : Identificación de la inviabilidad técnica en la realización de la aplicación móvil mediante la estrategia de Empaquetado antes del comienzo del tercer <i>Sprint</i>

13	<u>Disparador</u> : <i>Feedback</i> negativo por parte del experto de negocio sobre las conclusiones obtenidas del análisis de los datos
	<u>Mitigación</u> : Realizar demostraciones de los resultados intermedios al experto de negocio para obtener <i>feedback</i> e implementar acciones en caso de obtener resultados obvios
	<u>Contingencia</u> : Explorar alternativas diferentes para extracción y filtrado de datos. Mejorar el algoritmo de análisis de datos. Generar un corpus específico para ColectaTe
14	<u>Disparador</u> : Si transcurrido el 50% del tiempo del primer <i>sprint</i> se identificaron restricciones o complicaciones técnicas
	<u>Mitigación</u> : Realizar pilotos antes del comienzo del primer <i>sprint</i> para identificar la factibilidad técnica de la realización de las funcionalidades críticas
	<u>Contingencia</u> : Evaluar un posible cambio de plataforma o utilización de herramientas complementarias que permitan realizar las tareas requeridas
15	<u>Disparador</u> : Desviación del 20% en alguna de las métricas identificadas en el plan de calidad respecto a Calidad de Datos
16	<u>Disparador</u> : Promulgación de una propuesta de ley o decreto relativo a la privacidad de los datos electrónicos
17	<u>Disparador</u> : Ausencia de uno o más miembros del equipo en más de 5 <i>daily meetings</i> de continuo

Tabla 6.11 Plan de acción de riesgos

## 6.6. Gestión de calidad

El propósito de esta sección es establecer y especificar las actividades y procedimientos que se llevarán a cabo para asegurar que la calidad de la solución a construir sea la esperada por el cliente. Se detallarán las normas y estándares a utilizar, las tareas para la revisión de los entregables y procedimientos para registrar los defectos encontrados para darle seguimiento hasta su resolución.

A su vez, todos los entregables que se realicen, deberán cumplir con los requerimientos relevados y establecidos. Se definirán métricas de validación de la calidad, que serán utilizadas durante la ejecución de las actividades de SQA, para asegurar el nivel de aceptación de los entregables.

Nota: Se considera retrabajo todo aquello en lo que sea requerido invertir esfuerzo para corregir o cambiar algo sobre lo que ya se había invertido esfuerzo previamente y que no se encuentra de acuerdo a lo estipulado, acordado y firmado como parte del proyecto inicialmente. Por otro lado, se considera un cambio de alcance, todo aquello en lo que es requerido invertir esfuerzo para realizar o cambiar algo que no se encuentra dentro de lo estipulado, acordado y firmado como parte del proyecto inicialmente.

### 6.6.1. Objetivos de calidad

Se describen a continuación los objetivos de calidad definidos desde el punto de vista del proceso y producto.

Calidad del proceso	Forma de evaluación	Objetivo
Requisitos completados por iteración	<i>Sprint backlog</i>	100%
Horas de retrabajo	<i>Sprint backlog</i> / Registro de horas	<20%
Desviación de avance	<i>Product backlog</i>	<5%
Desviación de esfuerzo	<i>Sprint backlog</i>	<5%
Calidad del producto	Forma de evaluación	Objetivo
Efectividad de las revisiones técnicas	Registro de revisiones	>98%
Efectividad de las revisiones del cliente	<i>Sprint reviews</i>	<1%
Satisfacción del cliente con respecto al producto en general	Encuesta de satisfacción	80% >= "Bueno"
Correctitud del análisis de sentimiento	<i>Test funcional</i>	>80%
Correctitud del análisis de oportunidades de mejora	<i>Test funcional</i>	>90%
Correctitud del análisis de actualidad	<i>Test funcional</i>	>99%
Calidad del filtrado por ubicación	<i>Test funcional</i>	>90%
Pertinencia de los <i>posts</i> extraídos con respecto al tema analizado	<i>Test funcional</i>	>70%

Tabla 6.12 Objetivos de calidad proceso y producto

Atributos de calidad - <i>Big Data</i>	Forma de evaluación	Objetivo
Eficiencia	<i>Test de rendimiento</i>	>99%
Escalabilidad	<i>Test de rendimiento</i>	+100% de carga
Usabilidad (determinada por el usuario)	Encuesta de satisfacción	>="Bueno"
Atributos de calidad - App móvil	Forma de evaluación	Objetivo
Usabilidad (determinada por el usuario)	Encuesta de satisfacción	>="Bueno"

Tabla 6.13 Objetivos de calidad big data y aplicación móvil

Para información detallada sobre los atributos de calidad, ver sección "Requerimientos no funcionales".

Para información sobre los resultados de la encuesta de satisfacción, ver Anexo 17 - Encuesta de Satisfacción.

### 6.6.2. Plan de calidad

Se elaboró un plan de calidad acorde a una metodología ágil que oficie como apoyo al equipo de proyecto en la definición de la forma de trabajo y de funcionamiento como equipo [19].

Práctica	Desarrollo	Objetivos/Fundamentos	Rol responsable
Planificación de iteraciones	Realizar la planificación de las iteraciones de 2 semanas de duración	Se busca definir los requisitos de mayor prioridad, definir el compromiso, la meta de la iteración y la planificación de tareas. Además, obtener <i>feedback</i> por parte del experto de negocio luego de finalizada la iteración para que sirva como input adicional a las siguientes iteraciones.	Equipo
Liberación frecuente de entregables	Generar un entregable al final de cada iteración	Se busca liberar pequeños entregables de manera frecuente de funcionalidades, de manera que el cliente pueda ver el avance real en el desarrollo del producto	Equipo
Diseño	Realizar el diseño de flujos lo más simple posible	Se busca que los flujos construidos sean de diseño simple para facilitar su comprensión	Equipo
Pruebas unitarias	Realizar pruebas a medida que se desarrolla la solución	Se busca que el <i>tester</i> / desarrollador realice pruebas lo más temprano posible en la iteración para disminuir el re-trabajo	Equipo
Refactorización	Modificación de los nodos del flujo en el transcurso del desarrollo de la solución	Se busca simplificar y mejorar los flujos una vez probados, alterando su estructura sin cambiar su comportamiento, pero mejorando la comprensión del mismo para facilitar su mantenimiento en el futuro	Equipo
Integración continua	A medida que se realizan los distintos entregables, integrarlos	Se busca que a medida se vayan liberando los entregables, se realicen pruebas de integración.	Equipo

14 horas semanales	Cantidad de horas que pueden trabajar los programadores	Se busca que al no trabajar horas extra, se mejore el ánimo del equipo para aumentar la productividad	Equipo
Comunicación con el experto del negocio	Comunicación a demanda con el experto	Se busca que el cliente participe en el proyecto de manera que esté siempre informado de los avances, y comunique tanto los cambios necesarios como las consultas que puedan surgir.	Experto del negocio
Estándares de codificación	Generar los flujos guiándose en estándares específicos definidos para la realización del proyecto	Se busca que todos los integrantes del equipo cumplan con ciertas reglas de codificación (nombre de nodos y generación de metanodos) para facilitar en la comprensión de los flujos.	Equipo

Tabla 6.14 Plan de calidad

### 6.6.3. Tareas y responsabilidades

Los roles responsables del aseguramiento de calidad son los siguientes:

Rol	Descripción
Gerente de Calidad	Responsable de las funciones de aseguramiento de calidad
Gerente de Proyecto	Debe asegurar la implementación de las actividades de calidad, aprobar el plan SQA y coordinar la resolución de problemas encontrados
Experto de negocio	Define las expectativas de calidad y determina si el producto es aceptado
Testers	Ejecutan las pruebas de <i>software</i>

Tabla 6.15 Tareas y responsabilidades por rol

La responsabilidad principal del gerente SQA para el aseguramiento de la calidad es chequear que los procedimientos definidos cumplan con los requisitos de calidad establecidos en el plan y que los estándares especificados están siendo correctamente respetados.

Las tareas de calidad a realizar comprenden:

- ✚ Realización del Plan de SQA
- ✚ Identificación de los requisitos de calidad
- ✚ Revisar actividades de verificación de ajuste al proceso
- ✚ Asegurar el apego a los estándares definidos

- ✚ Definición de métricas
- ✚ Recolección de mediciones
- ✚ Validación de las *User Stories* al comienzo de cada iteración
- ✚ Realización de revisiones
- ✚ Indicar acciones correctivas en caso de errores encontrados
- ✚ Análisis causal
- ✚ Prueba de *software*

Como parte de las actividades de calidad, se revisarán los productos detallados a continuación:

Producto	Responsable
<i>Product Backlog</i>	Gerente de requerimientos
<i>Sprint Backlog</i>	Gerente de requerimientos
Alcance del sistema	Gerente de Proyecto
Descripción de la arquitectura	Arquitecto / Desarrollador
Plan de proyecto	Gerente de Proyecto
Plan de riesgos	Gerente de Proyecto
Plan de estimaciones	Gerente de Proyecto

Tabla 6.16 Responsables por artefactos del proyecto

#### 6.6.4. Estándares y guías

Se referencian a continuación los marcos normativos que han sido utilizados durante el desarrollo del proyecto.

##### Estándares de calidad

- ✚ ISO/IEC 25010:2011 - Requisitos y evaluación de la calidad de sistemas y del *software* [20]
- ✚ IEEE 730:1998 - Norma para el Plan de Aseguramiento de la Calidad de *Software* [21]
- ✚ IEEE 1028:2008 - Norma para Revisiones de *Software* y Auditorías [22]

##### Estándares de desarrollo

En el Anexo 14 - Estándares de desarrollo se encuentran detallados los estándares utilizados durante el proyecto:

- ✚ Guía de utilización de componentes y armado de flujos en la herramienta KNIME.
- ✚ Nomenclatura para el modelo físico de la base de datos
- ✚ Modelo de formato estándar para la generación de reportes.

### 6.6.5. Métricas de proceso y producto

Se describen a continuación las métricas de proceso y producto a ser evaluadas y monitoreadas durante el transcurso del proyecto.

Métricas del proceso	Fórmula	Periodicidad
Requisitos completados por iteración	$\frac{US\ Completadas - US\ Estimadas}{US\ Estimadas\ Sprint}$	Sprint
Horas de retrabajo	$\frac{Horas\ de\ retrabajo}{Horas\ totales\ del\ Sprint}$	Sprint
Velocity	$\frac{\sum SP\ de\ US\ Completas}{Cantidad\ Sprints\ Completos}$	Sprint
Desviación esfuerzo	$\frac{Esfuerzo\ Actual - Esfuerzo\ Estimado}{Esfuerzo\ Estimado\ Sprint}$	Sprint
Desviación de avance	$\frac{SP\ Completados - SP\ Estimados}{SP\ Estimados\ Sprint}$	Sprint
Desviación de alcance por Sprint	$\frac{\sum SP\ de\ CRs\ Sprint}{\sum SP\ Estimadas\ Sprint}$	Sprint

Tabla 6.17 Métricas de proceso

Métricas del producto	Fórmula	Periodicidad
Efectividad de las revisiones técnicas	$\frac{Cantidad\ defectos\ revisiones}{Cantidad\ total\ defectos\ detectados}$	Sprint
Efectividad de las revisiones del cliente	$\frac{Cantidad\ defectos\ detectados\ cliente}{Cantidad\ total\ defectos\ detectados}$	Sprint
Satisfacción del cliente con respecto al producto	Resultado encuesta de Satisfacción	Release
Correctitud del análisis de sentimiento	$\frac{Cantidad\ casos\ de\ prueba\ OK}{Total\ casos\ de\ prueba}$	Release
Correctitud del análisis de oportunidades de mejora	$\frac{Cantidad\ casos\ de\ prueba\ OK}{Total\ casos\ de\ prueba}$	Release
Correctitud del análisis de actualidad	$\frac{Cantidad\ casos\ de\ prueba\ OK}{Total\ casos\ de\ prueba}$	Release
Calidad del filtrado por ubicación	$\frac{Cantidad\ casos\ de\ prueba\ OK}{Total\ casos\ de\ prueba}$	Release

Pertinencia de los <i>posts</i> extraídos con respecto al tema analizado	$\frac{\text{Cantidad casos de prueba OK}}{\text{Total casos de prueba}}$	Release
--	---	---------

Tabla 6.18 Métricas de producto

Métricas atributos de calidad - Big Data	Fórmula	Periodicidad
Eficiencia	Tiempo procesar [ (cantidad datos diario promedio) * 7 ] < 120 minutos	Release
Escalabilidad	Tiempo procesar [ (cantidad datos diario promedio) * 7 * 2 ] < 120 minutos	Release

Tabla 6.19 Métricas atributos de calidad

### 6.6.6. Revisión y plan de auditoría

En esta sección se detallan los tipos de revisiones que se llevarán a cabo durante el ciclo de vida del proyecto, sus objetivos y mecanismo de implementación.

Tipo de revisión	Objetivo	Mecanismo de implementación
Revisión técnica	Evaluar la conformidad del producto con las especificaciones	Se revisa al finalizar cada iteración si el trabajo planificado fue realizado y si los criterios de verificación se cumplen satisfactoriamente.
Revisión de la calidad de los productos	Detectar desviaciones en los objetivos de calidad definidos	Se revisa al finalizar cada iteración si los objetivos de calidad definidos se cumplen satisfactoriamente.
Inspección de ajuste al proceso	Inspeccionar si los productos obtenidos cumplen con los estándares definidos	Se inspeccionan los productos para verificar que cumplan con los estándares definidos. Se documentan las desviaciones encontradas y se comunica al equipo para que las mismas sean corregidas.

Tabla 6.20 Plan de revisiones

En el Anexo 15 - Revisiones y mediciones se encuentran documentadas las revisiones llevadas a cabo.

### 6.6.7. Evaluación y testeo

El objetivo de esta sección es describir los tipos de pruebas de *software* que se ejecutarán para asegurar que el sistema satisface las funcionalidades especificadas, y el cumplimiento con el grado de calidad requerido definido anteriormente. Para cada una de las pruebas se detallará el objetivo, la técnica de ejecución así como los criterios que determinan si la prueba fue exitosa o no.

#### 6.6.7.1. Pruebas de funcionalidad

Las pruebas de funcionalidad se realizarán a través de la herramienta en forma de pruebas de caja negra. Pretende verificar que los flujos se comportan como especificado.





<b>Objetivo</b>	Verificar el cumplimiento de los requerimientos funcionales del sistema, asegurar que el mismo acepta y devuelve los datos correctamente en función de las reglas del negocio establecidas. Determinar los porcentajes de exactitud de los análisis.
<b>Técnica</b>	<ul style="list-style-type: none"> <li> Ejecutar cada uno de los flujos creados para el análisis y tratamiento de datos, especialmente los que correspondan a los procesos de análisis más relevantes.</li> <li> Ejecutar los flujos de extracción y filtrado de ubicación</li> <li> Definir un conjunto de datos de prueba junto con el resultado esperado que permita comparar con el resultado obtenido</li> <li> Definir configuraciones de nodos alternativas para determinar cuál es la que se acerca al objetivo de correctitud</li> </ul>
<b>Criterios de éxito</b>	Todos los procesos relevantes fueron ejecutados y los resultados se encuentran dentro de los objetivos de calidad definidos.

Tabla 6.21 Plan de pruebas de funcionalidad

#### 6.6.7.2. Pruebas de rendimiento

Las pruebas de rendimiento se realizarán en un entorno que contenga los requerimientos de *hardware* y *software* previamente establecidos. Pretenden verificar que los requerimientos de eficiencia han sido alcanzados.




<b>Objetivo</b>	Verificar tiempos de ejecución de los flujos en función del volumen de datos y comprobar se encuentran dentro de los criterios fijados.
<b>Técnica</b>	<ul style="list-style-type: none"> <li> Estimar un promedio de datos extraídos en un caso de uso normal y extrapolar según los requerimientos de escalabilidad establecidos</li> <li> Ejecutar cada uno de los flujos</li> <li> Comparar los resultados con respecto al requerimiento de eficiencia establecido</li> </ul>
<b>Criterios de éxito</b>	Todos los procesos de análisis fueron ejecutados y los resultados se encuentran dentro de los objetivos de calidad definidos.

Tabla 6.22 Plan pruebas de rendimiento

### 6.6.7.3. Pruebas de integridad de datos

Las pruebas de integridad se realizarán de forma independiente a los flujos, es decir, se ejecutarán sólo los componentes específicos relacionados a la interacción con la base de datos.

<b>Objetivo</b>	Asegurar que el acceso a la base de datos desde la herramienta funciona correctamente y no corrompe datos.
<b>Técnica</b>	<ul style="list-style-type: none"> <li>✚ Ejecutar cada uno de los nodos que realice un <i>insert / update</i> de tabla</li> <li>✚ Inspeccionar la base para asegurar que los datos fueron almacenados correctamente</li> </ul>
<b>Criterios de éxito</b>	Todas las ejecuciones funcionan como esperado sin corrupción de datos

Tabla 6.23 Plan pruebas de integridad de datos

En el Anexo 16 - Resultados - Plan de *testing* se puede visualizar los casos de prueba ejecutados y los resultados obtenidos.

### 6.6.8. Calidad de datos

Considerando que los datos que son extraídos para la realización de los análisis son generados por usuarios de internet en forma de texto libre, se debe realizar un filtrado y depuración de los mismos para asegurar un correcto almacenamiento en la base de datos y a su vez, para obtener una mejor exactitud en los procesos de análisis, ya que de ellos dependerá la toma de decisiones de la empresa.

Las actividades de calidad de datos se realizarán en cada *Sprint* formando parte de las actividades definidas para la realización de las *User Stories*.

Se enfocarán en los siguientes aspectos:

Tipo de actividad	Objetivo
Filtrado de datos en base a la localización del usuario	Se requiere que el análisis sea realizado sólo para usuarios de Uruguay.
Filtrado de datos en base al lenguaje (español)	Se requiere que el análisis sea realizado sólo para <i>posts</i> en español.
Pertinencia de los <i>posts</i> extraídos con respecto al tema analizado	Asegurar que los comentarios de usuarios que son extraídos de la fuente y analizados, tienen relación con el tema que se pretende analizar.
Filtrado de palabras "ruido"	Filtrar palabras que no aporten valor para la realización de los análisis, a saber: palabras de menos de 3 letras, pronombres, conectores, conjunciones y adverbios.

Filtrado de puntuación	Filtrar caracteres de puntuación que no aporten valor a la realización del análisis: puntos, punto y coma, dos puntos, signos de exclamación y signos de interrogación.
Filtrado de caracteres especiales	Filtrar caracteres fuera del rango ASCII para evitar fallas en el almacenamiento de los datos en la base.
Filtrado de URLs	Filtrar información adicional que incluya el usuario en su <i>post</i> que no aporte valor a los análisis ( <i>links</i> a otras páginas).
Filtrado de usuarios	Filtrar menciones a otros usuarios dentro del <i>post</i> .
Filtrado de <i>hashtags</i>	Filtrado de <i>hashtags</i> utilizados para la extracción de datos de la fuente de manera de evitar que sean tenidos en cuenta en los análisis de tendencia.
Descarte de <i>post</i> duplicados	Al realizarse dos tipos de extracción, controlar que no se almacenen <i>post</i> duplicados que puedan interferir con los resultados de los análisis
Almacenado	Controlar que los datos extraídos son completamente almacenados en la base (no son truncados, cortados o con palabras faltantes)

Tabla 6.24 Plan de calidad de datos

### 6.6.9. Resolución de incidentes y acciones correctivas

El responsable SQA documentará los hallazgos durante las actividades de aseguramiento de la calidad del producto. Dicho reporte tiene como objetivo registrar las discrepancias encontradas para su correcto monitoreo, y le será comunicado al equipo de proyecto para posteriormente implementar las acciones correctivas más adecuadas.

## 6.7. Gestión de la comunicación

Se detalla a continuación el plan de comunicación definido para el proyecto.

Qué se comunica	Quién lo comunica	A quién se comunica	Cómo se comunica	Frecuencia de comunicación
Requerimientos, Evacuación de dudas sobre requerimientos	Equipo	Equipo	Reunión presencial, mail, videoconferencia	Según sea necesario
Dudas sobre requerimientos	Equipo	Referente de negocio, Tutor	Reunión presencial, mail	Según sea necesario



Estado del <i>Product Backlog</i> ( <i>Backlog refinement, estimation</i> )	Equipo	Equipo	Videoconferencia	Con cada ceremonia
Estado del <i>Sprint Backlog</i> ( <i>Sprint Planning</i> )	Equipo	Equipo	Videoconferencia	Con cada ceremonia
Estado del proyecto, avance, planes, bloqueos ( <i>Daily meetings</i> )	Equipo	Equipo	Videoconferencia	2 o 3 veces a la semana
Bloqueos	Equipo	Tutor	Reunión presencial, mail	Según sea necesario
Revisión del producto ( <i>Sprint Review</i> )	Equipo	Referente de negocio, Tutor	Reunión presencial	Con cada ceremonia
Evaluación del <i>Sprint</i> ( <i>Retrospective</i> )	Equipo	Equipo, Tutor	Videoconferencia	Con cada ceremonia

Tabla 6.25 Plan de comunicación

## 6.8. Gestión de configuración

### 6.8.1. Elementos de SCM

Se identifican dos elementos que requieren control de configuración:

-  Documentación
-  Componentes de *software*

En caso de la documentación, para cada uno de los elementos identificados se utilizará el mecanismo de control de versiones de la herramienta utilizada.

En caso de los componentes de *software*, la nomenclatura es la preestablecida por la herramienta a utilizar, donde de ser posible, se deberá seguir el formato XXX\_Vx.y, en dónde XXX es el nombre del elemento y Vx.y la versión.

El número identificador de versión estará compuesto de la siguiente forma: x.y, en donde “x” representa un cambio severo en la versión e “y” un cambio menor.

Se definirá una línea base de los elementos identificados por cada iteración.

### 6.8.2. Herramientas

El control de versionado de los documentos se realizará a través de la herramienta Gdrive, la cual permite visualizar de forma sencilla los cambios realizados a cada documento y recuperar versiones anteriores del mismo en base a la fecha de modificación.

Las herramientas utilizadas para el versionado y control de la configuración para los elementos de *software* varían según el tipo de elemento:

- ✚ Elementos de *Big Data*: la propia aplicación Knime contiene un mecanismo que permite generar *snapshots* de los procesos, siendo posible volver el proceso al estado de uno de estos *snapshots*.
- ✚ Elementos de aplicación móvil: se utilizará Gdrive dado que la tecnología utilizada para generar el apk del empaquetado no requiere desarrollar código, sino que empaqueta el sitio web que ya posee la empresa.

### 6.8.3. Control de versiones

En el caso de la documentación, las actividades del control de versiones se realizan de forma automática para cada cambio realizado sobre los documentos.

Para los componentes de *software*, antes de realizar cualquier modificación, se debe generar una versión que permita recuperar el estado inicial antes del cambio en caso de problemas.

### 6.8.4. Control de cambios

En caso que se proponga un cambio no menor sobre el *software* (que amerite un análisis de impacto), se deberá informar a todo el equipo y posteriormente evaluar las consecuencias de su implementación (impacto en el plan de proyecto, en el acuerdo con el cliente y en el plan de riesgos), y los elementos para llevarlo a cabo. Si el cambio es viable, se determinará en qué momento será realizado, y cuándo será liberado.

Todo cambio en la documentación deberá ser detallado en el documento (especificando fecha, autor y motivo del cambio), e informar al resto del equipo por correo electrónico. Se entiende como cambio toda modificación que se realice y que afecte las líneas base.

## 6.9. Gestión de adquisiciones

### 6.9.1. Hardware

Para la realización del proyecto no ha sido necesaria la adquisición de *hardware*. Se han utilizado las computadoras personales de los integrantes del equipo como plataformas de desarrollo y servidor de aplicaciones, así como los dispositivos móviles propios para realizar las pruebas de los desarrollos móviles.

### 6.9.2. Software

Para *Big Data* se ha optado por la utilización de la solución de *software* Knime ([www.knime.org](http://www.knime.org)), de la cual se han utilizado 3 componentes [23]:

- ✚ KNIME *Analytic Platform*: de uso gratuito
- ✚ KNIME *Server*: licenciado, para lo cual se ha gestionado con Knime una licencia con fines académicos hasta el 31/12/2015. La misma se instala en el servidor.

- ✚ KNIME *Big Data Extension*: licenciado, para lo cual se ha gestionado con Knime tres licencias con fines académicos hasta el 31/12/2015. Las mismas se instalan en las máquinas de los equipos de proyecto.

Para la visualización de la información en tiempo real se optó por la solución de InterSoft:

- ✚ StyleScope Agile: licenciado, se gestionaron 3 licencias con fines académicos hasta el 31/12/2015. Las mismas se instalan en las máquinas de los equipos de proyecto.

## 7. Conclusiones, lecciones aprendidas y proyección a futuro

### 7.1. Conclusiones

Los objetivos planteados al inicio del proyecto fueron cumplidos satisfactoriamente gracias al esfuerzo de cada uno de los integrantes del equipo y la buena disposición y colaboración de las personas involucradas de forma directa o indirecta.

Esto permitió entregar un prototipo funcional, que si bien aún se puede mejorar, aportará valor al negocio y le será de utilidad a ColectaTe para afianzarse como empresa.

En lo que respecta a la gestión, y considerando la importancia de la misma en nuestra carrera, la elección de la metodología fue adecuada para el proyecto y nos permitió poner en práctica los conocimientos sobre metodologías ágiles aprendidos durante la carrera y que nos serán útiles a nivel profesional en el futuro.

En cuanto a las tecnologías involucradas, en particular todo lo relacionado con *Big Data* y tomando en cuenta el desconocimiento de la misma debido a su reciente explosión a nivel de mercado, fue muy gratificante a lo largo del proyecto poder investigarla y llegar a resultados tangibles. Esto nos permitirá optar por nuevos caminos con gran potencial profesional a futuro.

Para concluir, podemos decir que la experiencia fue muy positiva y enriquecedora. El proyecto nos ha dejado un gran aprendizaje a nivel personal y profesional, donde reafirmamos que la buena comunicación, trabajo en equipo y relacionamiento entre los integrantes del equipo de proyecto fue un factor clave para el éxito del mismo e hizo su transcurso ameno.

### 7.2. Lecciones aprendidas

Durante el transcurso del proyecto nos enfrentamos a diversas situaciones que nos enriquecieron personal y profesionalmente. A continuación detallamos un conjunto de lecciones aprendidas que creemos nos van a ser útiles en futuros proyectos.

#### Planificación

- ✚ Cuando un proyecto se enfrenta a tecnologías nuevas es importante planificar correctamente el tiempo de investigación y capacitación. En nuestro proyecto nos enfrentamos a tecnologías totalmente nuevas para todos los miembros del equipo y debido al corto plazo del proyecto se trató de lograr un balance entre el alcance y el tiempo dedicado a investigación, el cual no fue suficiente para abarcar todos los aspectos de las tecnologías involucradas.
- ✚ Para mitigar la falta de tiempo para investigación de las tecnologías, recurrimos a expertos y referentes en las mismas que nos ahorraron tiempo y nos permitieron validar hipótesis para continuar con nuestro proyecto. Esto nos demostró que la visión de expertos, ya sea a nivel tecnológico como de negocio en etapas tempranas del proyecto permite generar una planificación más acertada.

✚ A su vez, debido al desconocimiento de la tecnología y al corto tiempo de investigación, se aprendió que a veces es mejor establecer objetivos menos ambiciosos pero alcanzables para luego redefinirlos o proponer nuevos; esto en contraposición a fijar los objetivos ideales desde el comienzo.

#### Gestión

- ✚ Respecto a las estimaciones se aprendió que es importante tratar de incluir las horas necesarias de investigación en la planificación de los *Sprint*, para lograr cumplir con el alcance de cada *Sprint* de forma correcta y sin necesidad de invertir esfuerzo extra.
- ✚ En cuanto a la gestión de riesgos se definieron riesgos que de pronto no eran tan relevantes para el proyecto y se invirtió esfuerzo en su gestión, cuando el mismo se podría haber dedicado a temas más relevantes; por ejemplo los riesgos RSK12 y RSK16 para mencionar algunos.
- ✚ Los cambios de alcance solicitados durante el transcurso del proyecto dejaron ver la falta de capacitación durante la carrera en ciertas aptitudes indispensables dentro de la gestión de proyectos como es la habilidad de Negociación; necesaria para ser capaz de redefinir y renegociar los cambios de alcance con el cliente y manejar las expectativas del mismo.

#### Comunicación

- ✚ Durante el inicio del proyecto la mayor cantidad de reuniones internas fue realizada de forma virtual, pasando a mayor cantidad de reuniones presenciales a medida que avanzó el proyecto. De esta forma se incrementó la sinergia entre los integrantes del equipo, generando un sentimiento de mayor productividad, mejor calidad del trabajo realizado y mejorando la relación personal de los miembros.
- ✚ El punto anterior nos dejó ver que la buena comunicación y relación entre los integrantes del equipo fue un factor indispensable para el éxito del proyecto.

#### Desarrollo

- ✚ En cierto punto del desarrollo fue necesario invertir horas en retrabajo para mejorar la mantenibilidad de los procesos de *Big Data*, esto en parte se debió al desconocimiento de las herramientas y plataformas utilizadas, pero por otra parte también debido al no seguimiento del plan de calidad en el paso a paso del desarrollo. Esto llevó a que los costos de los cambios para cumplir con la calidad fueran mayores de lo esperado; y dejó ver la importancia de ejecutar las tareas según la planificación desde todo punto de vista.

Tabla 7.1 Lecciones aprendidas

### 7.3. Oportunidades de mejora y líneas de trabajo futuras

En el caso que ColectaTe esté interesado en continuar con la utilización de la herramienta, se puede profundizar en las siguientes oportunidades de mejora:

- ✚ Generar un *front end* que permita, de forma sencilla, realizar la administración y configuración de los parámetros dinámicos y diccionarios de datos utilizados en los procesos de *Big Data*
- ✚ Incluir otros países y geografías dentro del análisis de *Big Data*

- + Incluir otras redes sociales dentro del análisis de *Big Data*
- + Mejorar el análisis de sentimiento / análisis a través de aprendizaje
- + Mejorar la eficiencia y mantenibilidad de los flujos
- + Realizar una parametrización dinámica en función de la cantidad de tweets extraídos y analizados, y la cantidad de tópicos que se quieren obtener de ellos
- + Mejorar calidad de datos y procesos de análisis
  - ✓ Palabras que son seguidas de puntuaciones sin espacio entre la palabra y la puntuación y donde la puntuación es repetida más de una vez.
  - ✓ Palabras negadas donde entre la negación y la palabra negada no hay otras palabras o existe más de una palabra entre ambas.
  - ✓ Cuando existen igual número de palabras positivas y negativas, donde el resultado debería ser neutro, se arroja un resultado aleatorio dependiendo del orden en que se detecta el sentimiento de las palabras.
  - ✓ Detección de ironía en el análisis de sentimiento
  - ✓ Identificación dinámica de sentimiento de una palabra en base al contexto
- + Reducción de costos de licenciamiento
- + App móvil: se detectó que se puede mejorar la mantenibilidad cambiando la estrategia de visualización en la página HTML principal.

## 8. Referencias Bibliográficas

- [1] A. De Mauro, M. Greco y M. Grimaldi, «What is Big Data? A Consensual Definition and a Review of Key Research Topics,» de *4th International Conference on Integrated Information*, 2014.
- [2] Big Data Value Association, «European Big Data Value Strategic Research & Innovation,» *Big Data Value Europe*, p. 7, Ene 2015.
- [3] Cognizant, «Big Data's Impact on the Data Supply Chain,» *Cognizant 20-20 Insights*, p. 5, May 2012.
- [4] M. Porter, *Competitive Strategy: Techniques for Analyzing Industries and Competitors*, New York: Free Press, 1980.
- [5] D. Martínez y A. Milla, *Análisis del entorno*, Madrid: Ediciones Díaz de Santos, 2012.
- [6] Fondo Monetario Internacional, «Perspectivas económicas – Las Américas, desafíos crecientes,» *Estudios económicos y financieros*, Abr 2014.
- [7] A. Ordoqui, «El top 10 de la inflación en América Latina: Venezuela y la Argentina, a la cabeza,» Infobae, 16 Ene 2014. [En línea]. Available: <http://www.infobae.com/2014/01/16/1537515-el-top-10-la-inflacion-america-latina-venezuela-y-la-argentina-la-cabeza>.
- [8] [www.datosmacro.com](http://www.datosmacro.com), «IPC de Uruguay,» Jul 2015. [En línea]. Available: <http://www.datosmacro.com/ipc-paises/uruguay>.
- [9] Presidencia – República Oriental del Uruguay, «Tasa de desempleo se situó en 6.1%,» 12 Dec 2014. [En línea]. Available: <http://www.presidencia.gub.uy/comunicacion/comunicacionnoticias/desempleo-diciembre-2013>.
- [10] L. Bass, P. Clements y R. Kazman, *Software Architecture in Practice*, Addison Wesley, 2003.
- [11] Twitter, Inc, «GET search / tweets,» [En línea]. Available: [dev.twitter.com/rest/reference/get/search/tweets](https://dev.twitter.com/rest/reference/get/search/tweets).
- [12] Twitter, Inc, «The Streaming APIs,» [En línea]. Available: <https://dev.twitter.com/streaming/overview>.

- [13] M. Mascheroni, C. Greiner, R. Petris, G. Dapozo y M. Estayno, «Calidad de Software e Ingeniería de Usabilidad,» de *XIV Workshop de Investigadores en Ciencias de la Computación*, 2012.
- [14] X. Ferré, «Principios Básicos de Usabilidad para Ingenieros Software,» Universidad Politécnica de Madrid, Madrid.
- [15] Scrum Alliance, «Learn About Scrum,» [En línea]. Available: <https://www.scrumalliance.org/why-scrum>.
- [16] Capgemini, «The Scrum Perspective to Agile,» May 2011. [En línea]. Available: <https://www.capgemini.com/blog/capping-it-off/2011/05/the-scrum-perspective-to-agile>.
- [17] J. Nolan, «The Promise of Big Data Still Requires a Human Touch,» Lionbridge, Sep 2013. [En línea]. Available: <http://blog.lionbridge.com/enterprise-crowdsourcing/2013/09/16/the-promise-of-big-data-still-requires-a-human-touch/>.
- [18] «Agile Methodology,» [En línea]. Available: <http://agilemethodology.org/>.
- [19] A. Alvarez, «Elaboración de planes de calidad en proyectos de software,» de *XI Congreso Argentino de Ciencias de la Computación*, Concordia, Entre Ríos, 2005.
- [20] ISO/IEC, *25010:2011 - Systems and software Quality Requirements and Evaluation (SQuaRE)*, ISO/IEC, 2011.
- [21] IEEE, *730-1998 - Standard for Software Quality Assurance Plans*, IEEE, 1998.
- [22] IEEE, *1028:2008 - IEEE Standard for Software Reviews and Audits*, IEEE, 2008.
- [23] KNIME, «KNIME Product Matrix,» [En línea]. Available: <https://www.knime.org/products/product-matrix>.
- [24] Microsoft, «Características compatibles con las ediciones de SQL Server 2014,» [En línea]. Available: <https://msdn.microsoft.com/es-es/library/cc645993.aspx>.
- [25] MySQL, «Windows Platform Limitations,» [En línea]. Available: <http://dev.mysql.com/doc/refman/5.7/en/limits-windows.html>.
- [26] Jaspersoft, «How to auto-refresh a html-Report,» [En línea]. Available: <http://community.jaspersoft.com/questions/850671/how-auto-refresh-html-report>.
- [27] KNIME, «Integrating Big Data is as Easy as 1, 2, 3 ...4,» [En línea]. Available: <https://www.knime.org/blog/integrating-big-data-is-as-easy-as-1-2-3-4>.

- [28] eMarketer, «Latin American Countries Among Fastest-Growing Twitter Markets Worldwide,» 14 Jun 2014. [En línea]. Available: <http://www.emarketer.com/Article/Latin-American-Countries-Among-Fastest-Growing-Twitter-Markets-Worldwide/1010891>.
- [29] N. Jindal y B. Liu, *Opinion Spam and Analysis*, Chicago: Department of Computer Science, University of Illinois, 2008.
- [30] E. Cambria, B. Schuller, Y. Xia y C. Havasi, «New Avenues in Opinion Mining and Sentiment Analysis,» *IEEE Intelligent Systems*, vol. 28, 2013.
- [31] B. Pang, L. Lee y S. Vaithyanathan, «Thumbs up? Sentiment Classification using Machine Learning Techniques,» *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 79-86, 2002.
- [32] B. Liu, Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing, N. Indurkha and F. J. Damerau, 2010.
- [33] B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.
- [34] D. Blei, A. Ng y M. Jordan, «Latent Dirichlet allocation,» *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003.
- [35] D. Hu, *Latent Dirichlet allocation for Text, Images and Music*, San Diego: Department of Computer Science - University of California.
- [36] D. Blei, *Introduction to probabilistic topic models*, Princeton University.
- [37] S. Grant y J. Cordy, «Estimating the optimal number of latent concepts in source code analysis,» de *International Working Conference on Source Code Analysis and Manipulation*. IEEE Computer Society, 2010.
- [38] B. D. Heinz y L. Dawn, *Understanding LDA for Software Engineering*, Maryland: Loyala University .
- [39] Facebook, «Permissions with Facebook Login,» [En línea]. Available: <https://developers.facebook.com/docs/facebook-login/permissions/v2.4>.
- [40] M. Porter, «Spanish stemming algorithm,» [En línea]. Available: <http://snowball.tartarus.org/algorithms/spanish/stemmer.html>.
- [41] I. Diaz, G. Sidorov y S. Suárez, «Creación y evaluación de un diccionario marcado con emociones y ponderado para el español,» *Onomázen: revista semestral de lingüística, filología y traducción*, vol. 29, pp. 31-46, Jun 2014.

- [42] RAE, «Diccionario prehispanico de dudas,» 2005. [En línea]. Available: <http://lema.rae.es/dpd/srv/search?id=vvMk91OVTD6CnaLO7C>.
- [43] Flockofbirds, «Summer Training - Agile Software Development - ,» [En línea]. Available: <http://www.flockofbirds.nl/summer-training/>.
- [44] R. Chandane, «Agile: What I learnt so far (slide 6),» Abr 2014. [En línea]. Available: <http://www.slideshare.net/rohan.chandane/agile-presentation-what-i-learnt-so-far>.
- [45] hingedthinker, «Agile moment: How to estimate story points using planning poker,» Oct 2014. [En línea]. Available: <http://blogagility.marshallguillory.us/2014/10/29/agile-moment-how-to-estimate-story-points-for-a-user-story-using-planning-poker/>.

## 9. Anexos

### 9.1. Anexo 1 - Descripción Canvas

#### 9.1.1.1. Situación Inicial

Analizando el mercado uruguayo de colectas y tomando en cuenta solo tres tipos de éstas (casamientos, cumpleaños de 15 y nacimientos) se llegó a una estimación de 30.000 colectas anuales. Asumiendo que por colecta participan en promedio 50 personas aportando u\$s20 cada una, se calcula que se recaudan aproximadamente u\$s30.000.000 anualmente.

Cabe mencionar que se dejó por fuera del análisis otros tipos de colectas como pencas, bautismos, viajes, despedidas, entre otros, lo cual aumenta significativamente el tamaño del mercado potencial estimado; así como si se considerara también mercados regionales (Argentina 120 mil y en Brasil 1 millón de casamientos anuales).

Actualmente, cuando una persona desea crear una colecta a través de una red de cobranza, debe trasladarse físicamente hasta el local, depender de los horarios y días de atención, realizar trámites y colas; tampoco tiene una forma fácil de invitar a participar o compartir la existencia de la colecta. A su vez, según la red de cobranza elegida para realizar una colecta, se paga entre un 2,8% y un 3% de lo recaudado.

Si por el contrario, se crea una colecta de forma “artesanal” (es decir, juntar el dinero llevando un control manual de la misma), aparecen nuevos problemas: la responsabilidad de juntar el dinero y guardarlo (en muchas ocasiones grandes montos recaudados generan inseguridad), perseguir a los participantes para que paguen, informar a todos cuánto se ha recaudado y luego posiblemente decidir qué hacer con ello.

Asimismo, para las personas que desean participar de una colecta, deben también hacerlo trasladándose a una red de cobranzas, y no se tiene una buena trazabilidad de su colaboración. En caso de participar de colectas artesanales, debe recordar realizar la colaboración que en su momento se comprometió ofrecer, tener la plata justa, y/o trasladarse para hacer llegar su participación.

Por otro lado, no se ofrecen posibilidades (en servicios existentes) de elegir qué hacer con lo recaudado: obtener el dinero, una tarjeta de regalo, un vale de compras, un cheque o tener la posibilidad de seleccionar un regalo de forma centralizada.

Tampoco se ofrece la posibilidad de compartir y difundir fácilmente a través de las diferentes redes sociales para que la información de una colecta llegue a lugares de forma masiva (por ejemplo para las colectas solidarias).

Hasta ahora, no existe un servicio online donde poder crear, organizar, gestionar, participar, invitar y/o compartir colectas de diversas categorías; donde cada persona, desde la comodidad que desee (sólo contando con acceso a internet) pueda realizar cualquiera de las tareas y actividades mencionadas anteriormente, mediante la utilización de medios de pagos electrónicos.

### 9.1.1.2. Segmentos del mercado

El modelo de negocios distingue entre cuatro grandes grupos de clientes para los cuales se genera valor:

- + Personas desde 15 años en adelante, con acceso a internet y con necesidad de recaudar dinero (por el motivo que sea) en cualquier lugar de Uruguay. Se tomó la referencia de los 15 años en adelante dado que hoy día las entidades bancarias dan la posibilidad de obtener una cuenta o tarjeta de débito.
- + Empresas en Uruguay que deseen publicitar en nuestro sitio opciones para la adquisición de productos o servicios con el dinero recaudado en las diferentes colectas.
- + Organizaciones de beneficencia en Uruguay con necesidad de recaudar dinero (por el motivo que sea) en cualquier lugar de Uruguay.
- + Participantes que depositan en una colecta en Uruguay, son ellos quienes depositan. Dentro de Uruguay, que tengan o no tarjetas de débito y/o acceso a internet, dado que se puede también depositar mediante las redes de cobranzas físicas.

### 9.1.1.3. Propuesta de valor

ColectaTe es una solución digital para simplificar y facilitar la gestión e impulsar la participación en colectas de particulares y organizaciones.

Lo que se ofrece es un sitio que desde el lugar que cada uno prefiera (*PC, notebook, tablet, smartphone*), sin tener que trasladarse, pueda realizar la gestión integral de una colecta de forma centralizada.

El mismo ofrece los siguientes beneficios:

#### Para particulares (organizadores y participantes)

- + La posibilidad de crear, organizar, gestionar, participar, invitar y compartir colectas de forma online
- + Se brinda la opción de elegir qué hacer con lo recaudado: obtener el dinero, una tarjeta de regalo, un vale de compras, o tener la posibilidad de seleccionar un regalo
- + Las transacciones se realizan de forma segura
- + Mediante el uso de medios electrónicos se reduce el riesgo de sufrir robos en la vía pública

#### Para organizaciones

- + Ofrecer un nuevo canal para que los usuarios puedan colaborar y participar de colectas solidarias, brindando funcionalidades que faciliten la organización y gestión de colectas de forma online

Adicionalmente, en base al análisis y explotación de información obtenida a través de procesamiento de grandes volúmenes de datos (*Big Data*), se pueden obtener los siguientes beneficios:

#### Para empresas

- + Contar con un nuevo canal de ventas
- + Mejorar la estrategia de contenidos de la empresa
- + Mejorar la venta cruzada detectando productos y servicios consumidos comúnmente mediante la evaluación de tendencias de diferentes mercados
- + Predecir los elementos más apropiados para los clientes que hacen uso de ColectaTe y ofrecer una mejor combinación de los mismos

#### Para ColectaTe

- + Adaptar el servicio a las necesidades específicas de los diferentes clientes
- + Afianzar la relación con los clientes mediante el análisis de las conversaciones referentes a la empresa para tener un mejor manejo de quejas y disconformidades

#### **9.1.1.4. Canales de distribución**

El canal principal desde donde se brinda el servicio es una plataforma web (sitio web ColectaTe), en donde se publican las colectas, donde se comunican novedades y se le brinda al usuario la información de contacto de la empresa.

El servicio también será brindado a través de una aplicación móvil con las funcionalidades básicas para la gestión de colectas y atención al cliente.

Se cuenta con Redes Sociales, Facebook, Twitter, LinkedIn y Google+ como medios para compartir, comunicar y promocionar las colectas creadas o información en general. También se pueden generar Tweets y publicarlos en la web de manera de tener a los clientes y usuarios siempre informados.

La comunicación en todo sentido con los clientes y usuarios es vía mail. O al menos esta es la vía más importante que la que se cuenta. Al registrarse, a cada persona se le solicita un teléfono para ante cualquier situación o inconveniente, poder contar con otra vía de comunicación.

En el caso de las empresas u organizaciones, las mismas serán atendidas por ejecutivos comerciales con el fin de facilitar la comunicación e interacción para cerrar acuerdos.

#### **9.1.1.5. Relaciones con los clientes**

Las principales acciones se basan en la captación y fidelización de clientes, que requieran recolectar dinero, mediante campañas de marketing uno a uno en medios virtuales.

En cuanto a las empresas, la captación y fidelización de las mismas mediante reuniones presenciales para ofrecer nuestros servicios y lograr que publiciten en el sitio de ColectaTe.

Referente a las organizaciones, también se les brindará una atención personalizada.

Relación con los usuarios que depositan en colectas, sólo mediante un medio virtual (mail o teléfono) para actividades de soporte ante problemas con los depósitos en colectas.

#### 9.1.1.6. Fuentes de ingreso

Para todas las colectas creadas se cobrará una comisión del 3,5% de lo recaudado.

Respecto a las empresas, se cobrará una comisión que puede variar dependiendo del tamaño de la publicidad y la cantidad de días publicada. Estas empresas serán identificadas en base al procesamiento de grandes volúmenes de datos, determinado tópicos de interés y cruzando la información con las empresas que proveen estos servicios o productos.

También derivado de este estudio, la idea es poder contar con información para estas empresas, por ejemplo: “Se está hablando mucho de electrodomésticos, ¿te interesa ofrecer tus heladeras para que se compren directo desde nuestro sitio?”. De la venta de esta información se pretende obtener ingresos.

#### 9.1.1.7. Recursos clave

Dado que el negocio se basa en un entorno web, es crítica la disponibilidad del sistema, por lo tanto es imprescindible contar con una muy buena infraestructura a nivel de *hardware (data center)*.

Por el mismo motivo, el sitio web en sí, su diseño y usabilidad es un aspecto clave para captar y fidelizar clientes. De esto se desprende la importancia del personal de diseño gráfico.

Dado el negocio y la importancia de conseguir empresas que deseen publicitar en el sitio de ColectaTe, como organizaciones que quieran utilizar ColectaTe como otra vía de recaudación, es un recurso crítico el equipo de personas de ventas y *marketing*.

A su vez, la credibilidad y prestigio de la empresa es importante y se construye mostrando la colaboración con socios y proveedores de renombre en el mercado, así como brindando un servicio de calidad.

#### 9.1.1.8. Actividades clave

Lista de actividades claves:

- ✚ Entrega de lo recaudado a personas y organizaciones debe ser tal cual lo estipulado: darle confianza y seguridad al cliente que la recaudación será entregada en tiempo y forma cuando éste lo solicite, de manera de afianzar las relaciones para con ellos.
- ✚ Promocionar el valor agregado del servicio que se ofrece en referencia a lo existente actualmente: resaltar los beneficios y mejoras que presenta la utilización de ColectaTe en función de los servicios ofrecidos por la competencia como forma de promoción.

- ✚ El relacionamiento con los clientes y los canales de distribución: hacer especial énfasis en la relación con el cliente, intentando encontrar oportunidades de mejora de la plataforma web, mostrar al cliente una imagen de confianza y seriedad.

#### 9.1.1.9. Socios clave

Son socios claves las diferentes entidades de pagos (BanRed, SistarBanc, MVDCobros), ya que a través de las mismas se realiza todo el proceso de pago desde el depósito hasta el retiro del recaudo. A su vez, colaborar con estos, ayuda a generar credibilidad en la empresa y el servicio que se brinda.

Otros socios claves son las empresas que ofrezcan productos o servicios que puedan ser una solución para cambiar dinero por regalos.

#### 9.1.1.10. Estructura de costos

El principal costo se encuentra relacionado con los recursos humanos:

- ✚ Se requiere de la contratación de un diseñador web para mantener y mejorar la usabilidad del sitio web
- ✚ Programadores y *testers*
- ✚ Vendedores y equipo de *marketing*

Además, costos de infraestructura (servidores y *hosting*).

Se evalúa además la posibilidad de invertir en publicidad, *marketing* y relacionamiento.

## 9.2. Anexo 2 - Investigación de las herramientas

### 9.2.1. Plataformas de Big Data

#### 9.2.1.1. Hortonworks

Hortonworks es una empresa estadounidense enfocada en el desarrollo y soporte de Hadoop.

Ofrece una solución *open source* basada en Hadoop (HDP - Hortonworks *Data Platform*) para manejo de datos empresariales que permite el almacenamiento, procesamiento y análisis de grandes volúmenes de datos en diferentes formatos y provenientes de distintas fuentes.

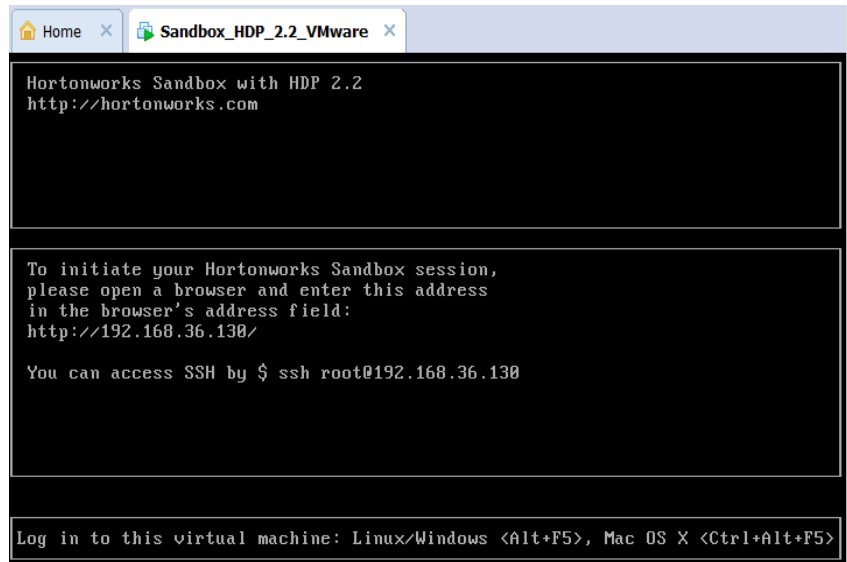


Imagen 9.1 Sandbox HDP

Se descargó desde la web oficial de Hortonworks la herramienta para utilizar desde un Sandbox, el cual se levantó en el entorno de virtualización VM Ware (Imagen 9.1).

Se accede al sistema desde un navegador web a través de la IP provista una vez que la máquina virtual esté levantada (Imagen 9.2).

Para la realización de la prueba, de debió instalar Flume, un servicio provisto por Apache el cual permite la extracción de datos del *streaming* de Twitter (por ejemplo) para ser almacenados y posteriormente, procesados.

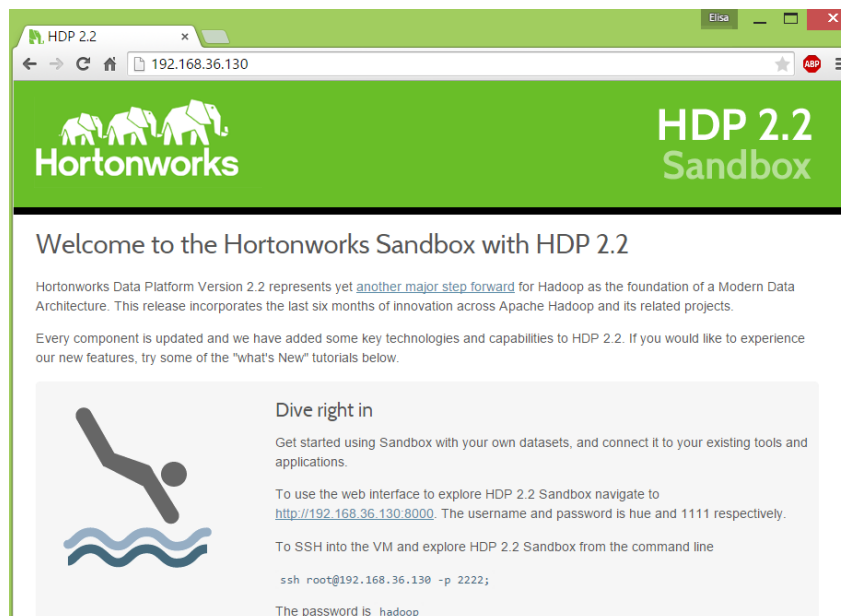


Imagen 9.2 HDP cliente web

La instalación se llevó a cabo a través de la ejecución del siguiente comando en la máquina virtual:

```
> yum install flume
```

Para la extracción de los datos, se debió descargar el archivo .jar “flume-sources-1.0-SNAPSHOT.jar” el cual contiene el código fuente necesario para conectarse a la API de Twitter, extraer los tweets e insertarlos en la base de datos de HDP.

El .jar debe ser copiado a la máquina virtual en el directorio /usr/lib/flume/lib

Por otro lado, se debe editar el archivo de configuración “flume.conf” localizado en el directorio “/etc/flume/conf” de la siguiente manera:

### **Definición de los componentes de Apache Flume para el agente “TwitterAgent”**

```
TwitterAgent.sources = Twitter
```

```
TwitterAgent.channels = MemChannel
```

```
TwitterAgent.sinks = HDFS
```

### **Configuración de la fuente del agente y claves de autenticación y acceso a la API de Twitter:**

```
TwitterAgent.sources.Twitter.type = com.cloudera.flume.source.TwitterSource
```

```
TwitterAgent.sources.Twitter.channels = MemChannel
```

```
TwitterAgent.sources.Twitter.consumerKey = BRJHZSKfD936JtJhckkULbP6c
```

```
TwitterAgent.sources.Twitter.consumerSecret =  
jxESJBjtCnQLWZlWV3XQFGbnOMeLXP9pg7KKDke5pdWv8mPf14
```

```
TwitterAgent.sources.Twitter.accessToken = 1920612620-  
fhudjWmkji4YCaBV5XBYhOVg11ni403R92hAXaf
```

```
TwitterAgent.sources.Twitter.accessTokenSecret =  
hWwhOYIASKJRi73jtcFVsh5u7hEFBywqRqPGhxfukjgix
```

### **Definición de las palabras clave a buscar dentro de los tweets:**

```
TwitterAgent.sources.Twitter.keywords = Europa
```

### **Definición del componente “Sinks” y carpeta de destino de los datos extraídos:**

```
TwitterAgent.sinks.HDFS.channel = MemChannel
```

```
TwitterAgent.sinks.HDFS.type = hdfs
```

```
TwitterAgent.sinks.HDFS.hdfs.path = hdfs://192.168.36.130:8020/usr/hdp/2.2.0.0-  
2041/flume/tweets/%Y/%m/%d/
```

```
TwitterAgent.sinks.HDFS.hdfs.fileType = DataStream
```

*TwitterAgent.sinks.HDFS.hdfs.writeFormat = Text*

*TwitterAgent.sinks.HDFS.hdfs.batchSize = 1000*

*TwitterAgent.sinks.HDFS.hdfs.rollSize = 0*

*TwitterAgent.sinks.HDFS.hdfs.rollCount = 10000*

### Definición del componente "Channel":

*TwitterAgent.channels.MemChannel.type = memory*

*TwitterAgent.channels.MemChannel.capacity = 10000*

*TwitterAgent.channels.MemChannel.transactionCapacity = 1000*

**La extracción de los datos se realizó ejecutando el siguiente comando:**

*flume-ng agent --conf-file /etc/flume/conf/flume.conf --name TwitterAgent*

Los datos extraídos se pueden visualizar desde el cliente web, en el File Browser (Imagen 9.3 e Imagen 9.4).

### Conclusiones:

Licencia: Gratuita / Open Source

Requerimientos:  
Entorno de virtualización  
VMWare / Hyper-V / VirtualBox. OS 64 bits, 4 GB RAM mínimo

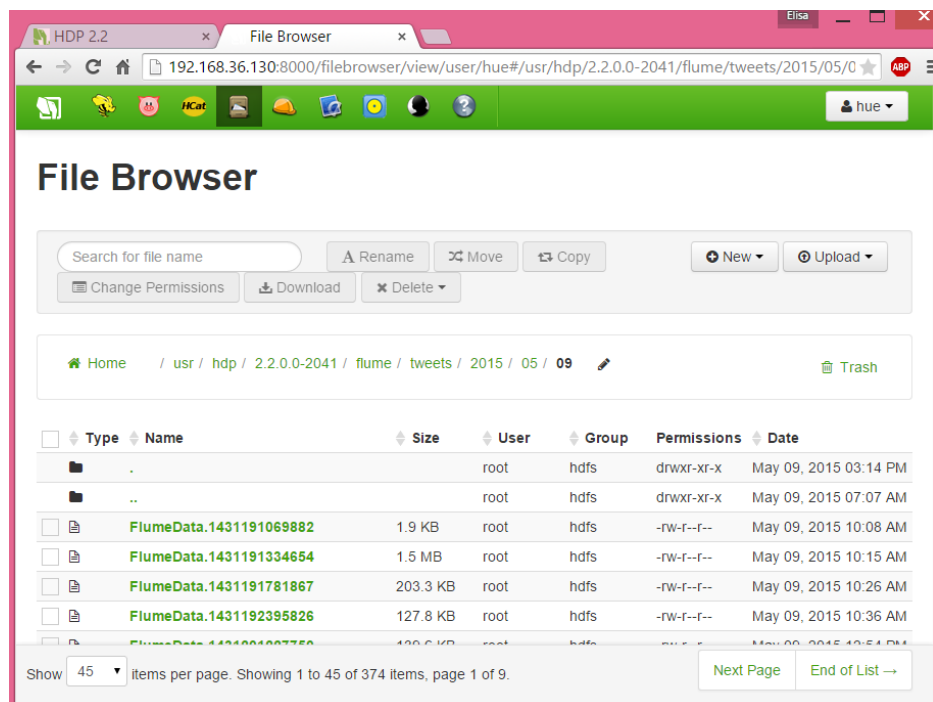


Imagen 9.3 Archivos generados por Flume conteniendo datos extraídos de Twitter

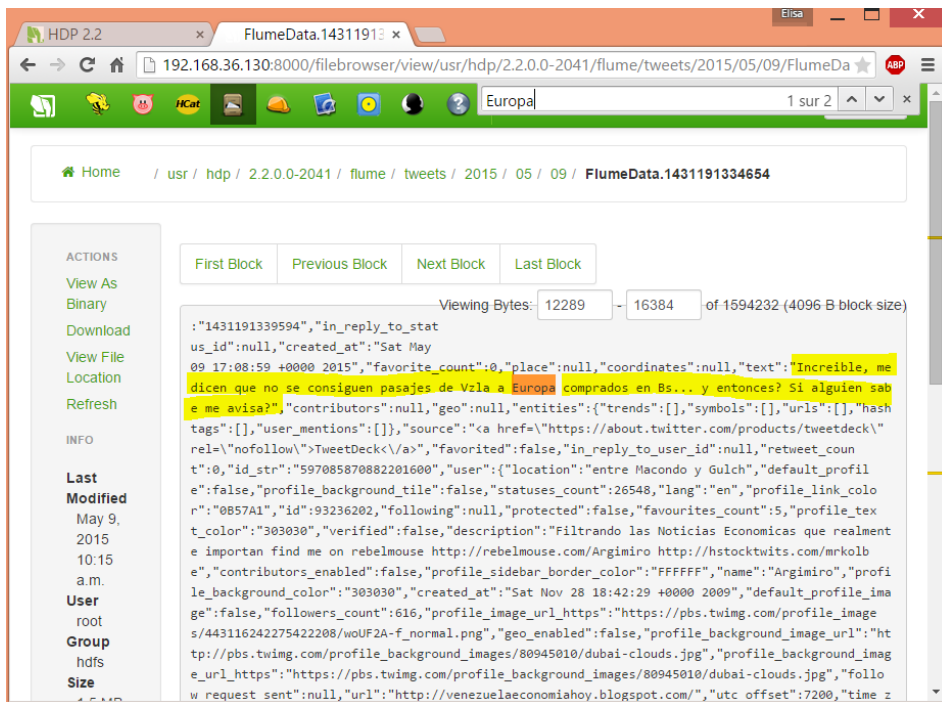


Imagen 9.4 Datos extraídos

Facilidad de uso: Baja. Se debe proceder de forma manual por línea de comando.

Configuración mediante archivos de configuración. No es intuitivo, requiere mucho tiempo en capacitación.

Creación de componentes propios: Requiere conocimiento avanzado en Apache y

programación Java.

Integración con herramientas de terceros: Se integra con varias bases de datos y diversas herramientas de BI a través de ODBC

Completitud del proceso *Big Data*: Todo menos visualización

### 9.2.1.2. Informatica

Informatica PowerCenter es una plataforma enfocada en la extracción, transformación y carga o ETL (*Extract, Transform, Load* en inglés).

La misma se encuentra desarrollada sobre J2EE y Hadoop, y su versión PowerCenter Express es gratuita, si bien cuenta con funcionalidades limitadas.

Link al sitio web:

<https://www.informatica.com/products/data-integration/powercenter/powercenter-express.html#fbid=BmvAJiosegX>

Para la conexión y extracción de datos desde redes sociales es necesario la instalación de un paquete adicional que integra los conectores necesarios en la aplicación. El paquete se denomina PowerCenter for Social Media y se puede descargar desde:

[https://community.informatica.com/solutions/informatica for social media](https://community.informatica.com/solutions/informatica%20for%20social%20media)

Previo a la prueba de extracción desde Twitter se creó una aplicación de Twitter desde el sitio <http://dev.twitter.com/apps> y se obtuvieron los datos de acceso necesarios (*Consumer Key*, *Consumer Secret*, *Access Token*, *Access Secret*).

Para la prueba se siguieron los siguientes pasos:

1. Se creó una conexión con Twitter (Imagen 9.5)

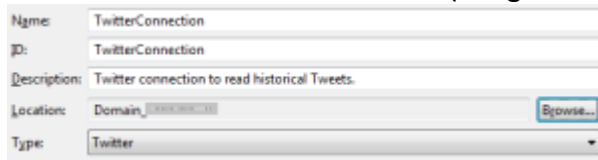


Imagen 9.5 Informatica 1

2. Se cargaron las credenciales de acceso (*Consumer Key*, *Consumer Secret*, *Access Token*, *Access Secret*) (Imagen 9.6)

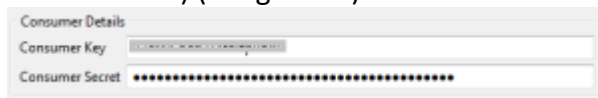


Imagen 9.6 Informatica 2

3. Se creó un objeto de datos de tipo Twitter (Imagen 9.7)

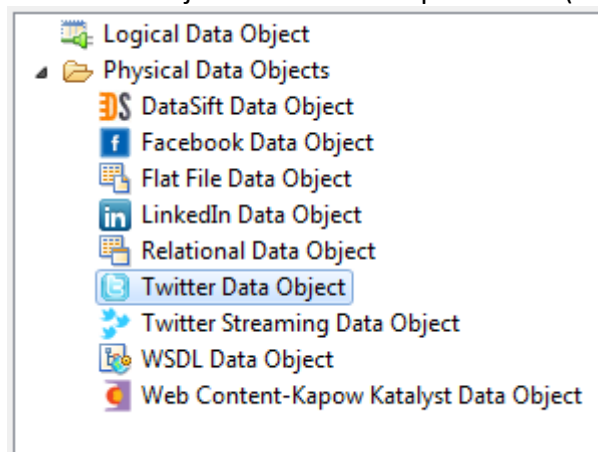


Imagen 9.7 Informatica 3

4. En el objeto se seleccionó la conexión creada en previamente (Imagen 9.8)

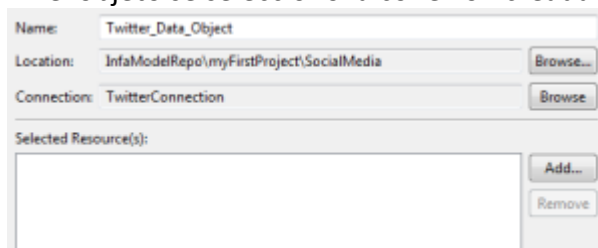


Imagen 9.8 Informatica 4

5. Luego se seleccionó el tipo de información a obtener desde Twitter, para este caso se van a obtener tweets, por lo tanto se necesita invocar el servicio "Search" (Imagen 9.9)

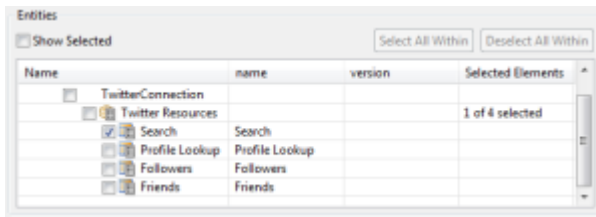


Imagen 9.9 Informatica 5

6. Se procede luego a crear una operación sobre el objeto de datos de Twitter creado previamente (Imagen 9.10)

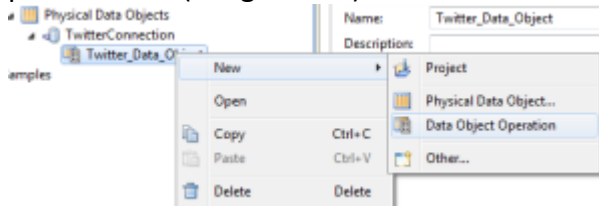


Imagen 9.10 Informatica 6

7. En la operación se selecciona una o más de las funciones habilitadas en el objeto, en este caso la función de "Search" (Imagen 9.11)

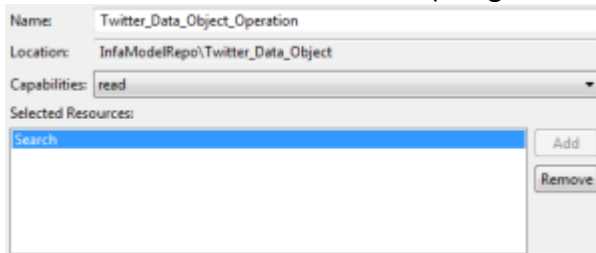


Imagen 9.11 Informatica 7

8. Por defecto, se mapean uno a uno los campos entre desde la API de Twitter y el objeto de salida que crear Informatica para recibir los resultados, eso puede ser editado manualmente (Imagen 9.12)

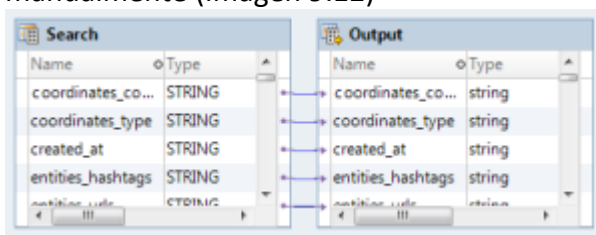


Imagen 9.12 Informatica 8

9. En la ventana de propiedades del componente se carga el parámetro *Query* con el texto a buscar
10. Se crea un objeto para guardar los datos en un archivo de texto plano
11. Se conecta el Objeto de Twitter con el Objeto para guardar los resultados
12. Se ejecuta el proceso de extracción y en la vista de visualización de datos se puede ver el resultado extraído (Imagen 9.13)

Object	Object	Source	Name	Imported	User	User
8	None	web	Peter Parker Resurrected in Scott's Amazing Spider-Man! SCOOOO... None	None	None	None
9	None	va href="http..."/> <td>Peter Parker Resurrected in Scott's Amazing Spider-Man! Manual II... None</td> <td>None</td> <td>None</td> <td>None</td>	Peter Parker Resurrected in Scott's Amazing Spider-Man! Manual II... None	None	None	None
9	None	va href="http..."/> <td>Peter Parker Resurrected in Scott's Amazing Spider-Man! Manual II... None</td> <td>None</td> <td>None</td> <td>None</td>	Peter Parker Resurrected in Scott's Amazing Spider-Man! Manual II... None	None	None	None
9	None	va href="http..."/> <td>Is the amazing Spider-Man 2 has the four villains wif None</td> <td>None</td> <td>None</td> <td>None</td>	Is the amazing Spider-Man 2 has the four villains wif None	None	None	None
20	None	va href="http..."/> <td>RT @CBR: BRUJING Peter Parker Resurrected in @DavidDini's 'Waa... None</td> <td>None</td> <td>None</td> <td>None</td>	RT @CBR: BRUJING Peter Parker Resurrected in @DavidDini's 'Waa... None	None	None	None
21	None	va href="http..."/> <td>On March and/or April 2014 The Amazing Spider-Man comic returns... None</td> <td>None</td> <td>None</td> <td>None</td>	On March and/or April 2014 The Amazing Spider-Man comic returns... None	None	None	None
22	None	va href="http..."/> <td>I'm watching The Amazing Spider-Man for the first time and I gotta... None</td> <td>None</td> <td>None</td> <td>None</td>	I'm watching The Amazing Spider-Man for the first time and I gotta... None	None	None	None
22	None	web	So Peter Parker, the real Peter Parker is coming back in April with 4... None	None	None	None

Imagen 9.13 Informatica 9

## Conclusiones

Luego de la prueba se identificó que Informatica PowetCenter Express es una herramienta excelente para ejecutar operaciones de ETL, pero no cuenta con funciones para realizar pre procesamiento de datos, no permite realizar análisis sobre los mismos ni tampoco provee de mecanismos de visualización avanzados para usuario de negocio.

### 9.2.1.3. Talend

Talend *Open Studio for Big Data* es una plataforma que permite simplificar las operaciones de ETL, integración y análisis de datos.

Se encuentra desarrollado sobre J2EE y Hadoop y es de uso gratuito. Las fuentes de los componentes y nodos son *open source* y es posible extender sus funcionalidades o desarrollar nuevas si se desea.

Sitio web: <https://www.talend.com>

Previo a la prueba para extraer información de Twitter, fue necesario descargar un componente gratuito desarrollado por un tercero para poder conectarse con los servicios de Twitter y extraer datos. Sitio de descarga:

<http://gabrielebaldassarre.com/talend/twitter-components-talend/>

Para la prueba se siguieron los siguientes pasos:

1. Se creó una conexión con Twitter (Imagen 9.14)

**New job**

Add a job in the repository

Name: ExtraccionTwitter

Purpose: Piloto Extraccion Twitter

Descripción: Piloto Extraccion Twitter - Colectate

Imagen 9.14 Talend 1

2. Se generó un proceso básico para conectarse a Twitter, extraer tweets y guardarlos en un .csv (Imagen 9.15)

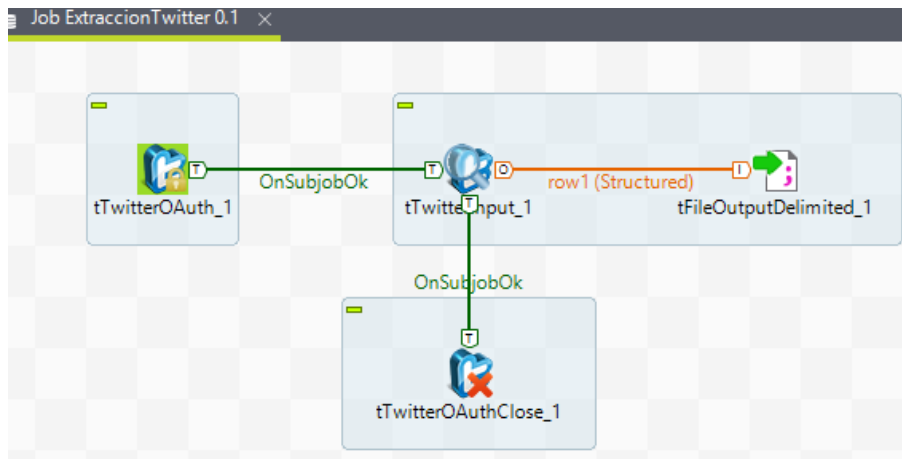


Imagen 9.15 Talend 2

3. La conexión a Twitter se configuró de la siguiente forma (Imagen 9.16)

Consumer Key "4VYq....."

Consumer Secret "8UK....."

Access Token "7C....."

Access Token Secret "....."

!!!CONNECTION\_TYPE.NAME!!! Twitter API \*

Imagen 9.16 Talend 3

4. El componente *Input* se configuró con los parámetros a obtener de Twitter y la palabra de búsqueda (Imagen 9.17)

Schema Built-In Edit schema Select a c

Column	Operation
tweetId	Status ID
tweetFecha	Creation date
TweetUsuario	Sender's Screen Name
TweetTexto	Text

Logical operator used to combine conditions And \*

Operator	Query
Including term(s)	"Colectate"

Imagen 9.17 Talend 4

5. Se agregaron además los componentes para cerrar la conexión con Twitter y bajar los resultados a un .csv (Imagen 9.18)

```

629793681512529920;07-08-2015;m_camerota;RT @colectateuy: Orga
629769311402491904;07-08-2015;Vanesnandez;RT @colectateuy: Orga
629752263171551232;07-08-2015;colectateuy;Organizando la fiest
629054800819286016;05-08-2015;colectateuy;En #colectata #colec
629052912417800192;05-08-2015;m_camerota;En #colectata #colect
628688632598130689;04-08-2015;lety_luna;Ingresá a http://t.co/
627937842380075008;02-08-2015;m_camerota;@gmalvasio Muchas gra

```

Imagen 9.18 Talend 5

## Conclusiones

Es una herramienta relativamente sencilla de utilizar, gratuita y que cuenta con facilidades para realizar extracción de datos desde Twitter ya que un tercero desarrolló y ofrece gratuitamente los nodos para esto. No cuenta con componentes desarrollados para extraer datos de otras redes sociales.

Cuenta con funciones que facilitar las operaciones en 3 de las 4 fases del ciclo *Big Data*: extracción, procesamiento y análisis. No cuenta con funciones de visualización.



### 9.2.1.4. KNIME

Knime (Konstanz Information Miner) es una plataforma *open source* para el análisis, reporte e integración de datos. Permite realizar minería de datos, análisis de tendencias y predicción de resultados.

Está basado en Eclipse y programado sobre J2EE.

Sitio web oficial de KNIME Analytics Platform: <https://www.knime.org/>

Para la prueba de extracción de datos de Twitter, se realizó el diagrama tal como se muestra en la Imagen 9.19, utilizando los siguientes componentes:

-  Twitter API Connector
-  Twitter Search

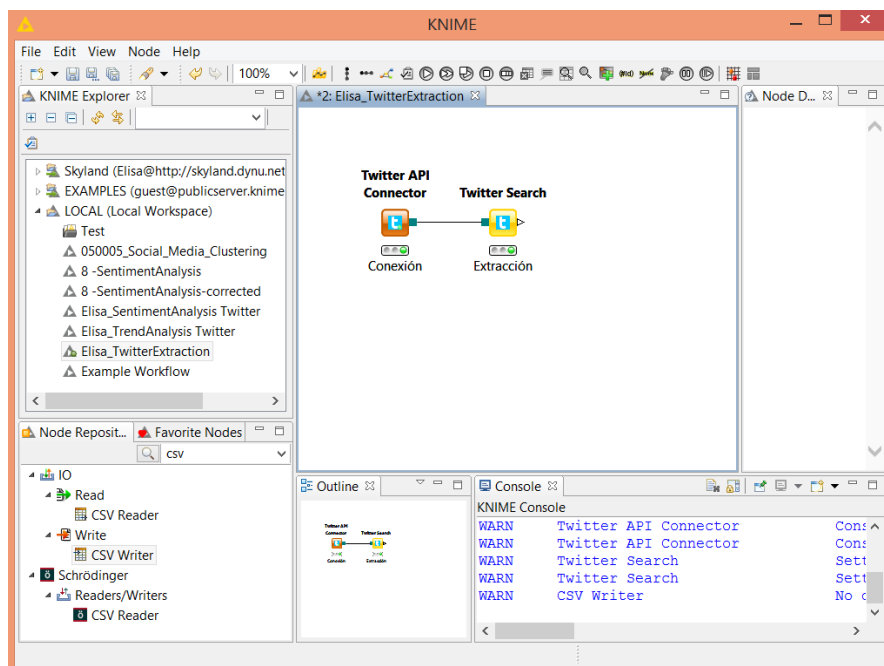


Imagen 9.19 KNIME Workflow para extracción de datos de Twitter

Los componentes fueron configurados de la siguiente forma:

En el nodo “Conexión”, se especificación los detalles para autenticación con la API de Twitter (Imagen 9.20).

En el nodo “Extracción”, se especificaron las palabras claves a ser buscadas y la cantidad máxima de tweets a ser extraídos (Imagen 9.21).

Para extraer los datos se ejecutaron todos los nodos, obteniéndose el resultado mostrado en la Imagen 9.22.

## Conclusiones

Licencia gratuita / *Open Source* con ciertos componentes pagos no indispensables como el Servidor y conectores extra para *Big Data*.

Es una herramienta muy fácil de usar que permite la visualización de los resultados en tiempo de desarrollo y sobre todo posee funcionalidades que abarcan las 4 principales fases del ciclo *Big Data*, desde la extracción hasta la visualización.

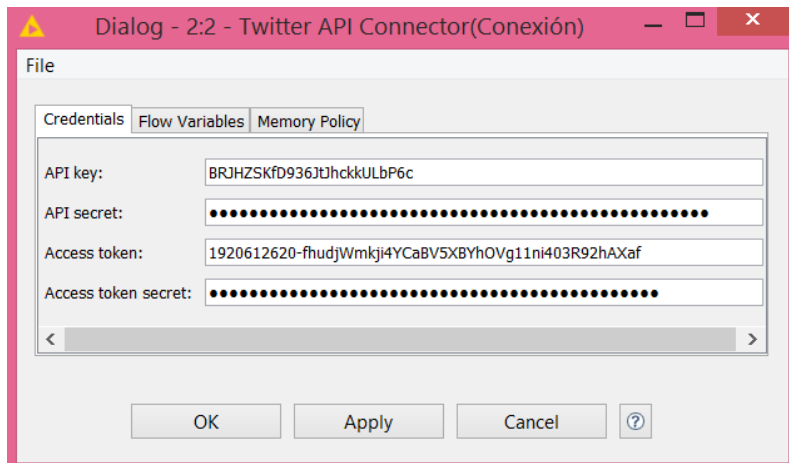


Imagen 9.20 Configuración del componente “Twitter API Connector”

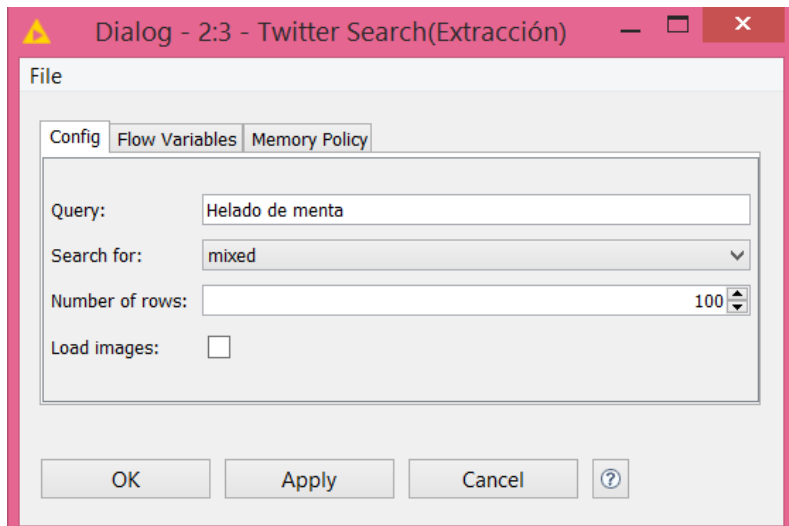


Imagen 9.21 Configuración del componente “Twitter Search”

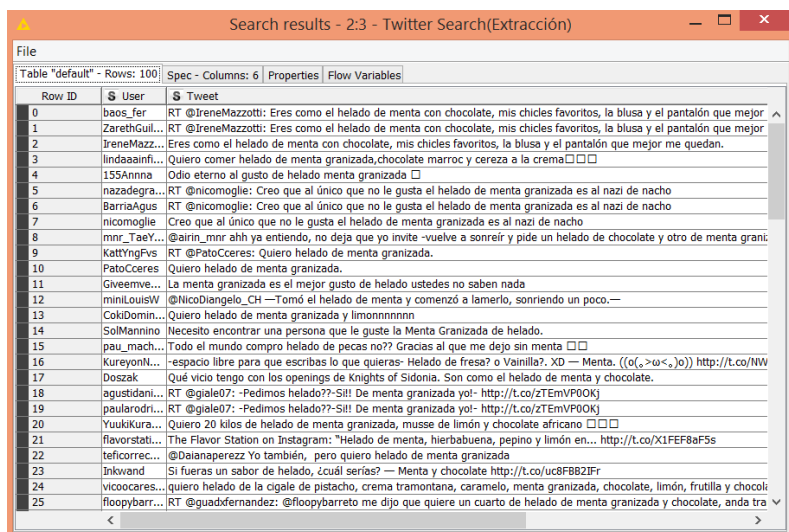


Imagen 9.22 Resultado de la extracción

## 9.2.2. Plataformas de almacenamiento

A continuación se encuentra una comparación de las características principales entre SQL Server Express [24] y MySQL [25] que impactan en la decisión de la elección del motor de bases de datos para el proceso de *Big Data* de este proyecto.

Característica	SQL Server Express	MySQL
<b>Versión</b>	2014	5.7
<b>Licencia</b>	Gratuita	Gratuita
<b>Límite de tamaño de base de datos y/o tablas</b>	10 GB	256 TB o límite de sistema de archivos
<b>Límite de utilización de memoria</b>	1 GB	Límite de sistema de 64 bits
<b>Límite de utilización de procesadores</b>	1 procesador físico con múltiples procesadores lógicos	Sin límites

Tabla 9.1 Comparación plataformas de almacenamiento

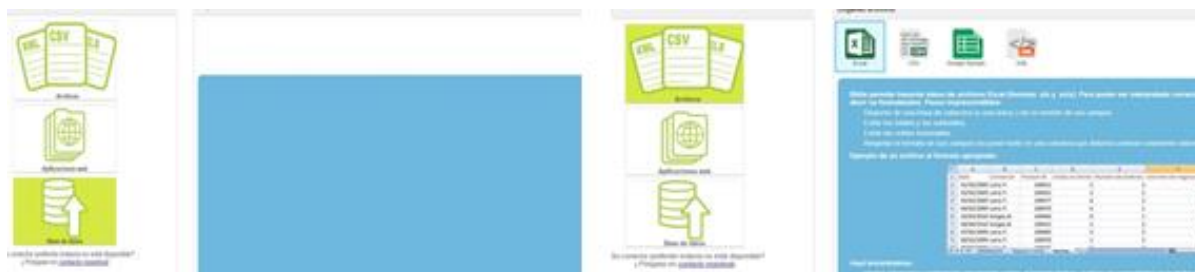
## 9.2.3. Herramientas de visualización

### 9.2.3.1. Bittle

Es una herramienta que permite crear cuadros de mando e informes en línea de forma rápida y sencilla para que pueda gestionar de forma eficaz la actividad empresarial de la organización.



Ya sea que la actividad esté en *Marketing*, Ventas, Recursos Humanos, Finanzas, Compras o de TI, Bittle le permitirá realizar un seguimiento de las métricas y no se requieren conocimientos informáticos técnicos...



En la versión gratuita no es posible realizar la conexión con una base de datos MySQL. En la versión empresarial, se encuentra poca variedad en el diseño y armado de los reportes.

### 9.2.3.2. ClicData

La versión gratuita, nuevamente no tiene habilitada la conexión a una base MySQL. De todas maneras, se realizó el registro y pruebas conectando a Google Drive para ver el funcionamiento.

Aquí se detectó que el *auto-refresh* no es parametrizable, sino que se realiza de forma automática cada hora.



### 9.2.3.3. Dash

Es cuadros de mando en tiempo real para su sitio web, su negocio y su vida. Su primer salpicadero es libre para siempre. Podrás añadir *widgets* que conectan a los servicios populares o fuentes de datos personalizados. A continuación, verás la luz del tablero con actualizaciones en tiempo real.



En esta herramienta no es posible la conexión con MySQL bajo ninguna de sus versiones (*free* o *paga*).

### 9.2.3.4. JasperSoft

Jaspersoft Studio es una herramienta de código abierto, es un diseñador



de informes gratuito basado en Eclipse para JasperReports y JasperReports Server. Se pueden crear diseños muy sofisticados que contienen gráficos, imágenes, sub-informes, tablas de contingencia y mucho más. Acceder a sus datos a través de JDBC, TableModels, JavaBeans, fuentes XML, Hibernate, CSV y personalizados. Luego, se pueden publicar los informes en PDF, RTF, XML, XLS, CSV, HTML, XHTML, texto, DOCX, u OpenOffice.

JasperReports Server es un servidor de informes independientes e integrables. Proporciona informes y análisis que se pueden incrustar en una aplicación web o móvil, así como operan como un centro de información central para la empresa mediante la entrega de información de misión. JasperReports Server está optimizado para compartir, proteger y gestionar de forma centralizada los informes de Jaspersoft y vistas analíticas.

En esta herramienta no se encontraron problemas al conectar con MySQL, se logró realizar todo en forma exitosa pero, a la hora de realizar los reportes de forma automática se detectó que no es posible [26].

### 9.2.3.5. InetSoft

InetSoft Style Scope™ es una aplicación fácil de usar, es un *software* de escritorio interactivo que incluye capacidades de informes en tiempo real. Es una edición de “Style Intelligence” que se centra en el seguimiento y la exploración de los datos de negocio mediante la combinación de una plataforma de inteligencia de datos con tecnología de visualización, que sirve tanto a las empresas y proveedores de soluciones.



En el núcleo de la plataforma de inteligencia de datos InetSoft Data Block™ es la tecnología de datos que permite el mapeo y crea bloques de rendimiento afinado y de datos de seguridad controlado que se pueden transformar y montar por los usuarios de negocio para las preguntas de negocio en tiempo real.

Los usuarios de negocios consiguen autoservicio personalizable, apuntando y haciendo clic intuitivamente tienen acceso visual a la información. Los usuarios avanzados y los científicos de datos tienen la capacidad de trabajar con todos los datos que necesitan sin tener que depender de TI.

## 9.2.4. Herramientas de Empaquetado Móvil

### 9.2.4.1. Intel XDK

Ofrece una solución de desarrollo para crear aplicaciones móviles multiplataforma utilizando HTML, PHP, CSS y JavaScript.



Esta ha sido la herramienta seleccionada para realizar el empaquetado. El mismo, se realiza una única vez, ya que luego, cada cambio realizado en el sitio web se ve reflejado al instante en la aplicación. Es posible crear nuevas versiones pero solo en caso de realizar cambios propios de la aplicación, o sea, utilizar un servicio del dispositivo, como por ejemplo que la aplicación se visualice sólo de forma horizontal.

### 9.2.4.2. Apache Cordova

Es un conjunto de APIs de dispositivos que permiten a un desarrollador de aplicaciones móviles poder acceder a la función del dispositivo nativo como la cámara o acelerómetro desde JavaScript. Combinado con un marco de interfaz de usuario, tales como jQuery Mobile o Dojo Mobile o Sencha Touch, esto permite que una aplicación de teléfono inteligente que se desarrolló con sólo HTML, CSS y JavaScript.



Realizada la investigación de la herramienta, la misma presenta complejidades técnicas que en esta primera etapa no serán necesarias abarcar ya que, como se menciona anteriormente, Apache Cordova permite acceder a funciones nativas del dispositivo no necesarias para nuestro objetivo.

#### 9.2.4.3. Android Creator

Es una herramienta que de forma gratuita permite crear una aplicación nativa para dispositivos móviles y *tablets* Android. Sólo es necesario ingresar ciertos datos requeridos en el asistente de la misma. Al realizar una investigación más detallada y pruebas de funcionamiento no se logró el funcionamiento necesario para cumplir con el objetivo, ya que al contener el sitio un fuerte componente de PHP el resultado no es el esperado y por lo tanto, no se cumplió el objetivo planteado.



### 9.3. Anexo 3 - Cómo KNIME cumple con el proceso de Big Data?

Cada vez más personas están descubriendo que la cantidad de datos en bruto disponibles recogidos por un sistema puede crecer exponencialmente rápido, alcanzando rápidamente un tamaño muy grande y un alto número de características, en ocasiones la clasificación de estos datos se conoce como "*Big Data*". En el caso de conjuntos de datos muy grandes, puede ser útil aprovechar las grandes prestaciones de la plataforma de datos, especialmente para ejecutar procedimientos ETL.

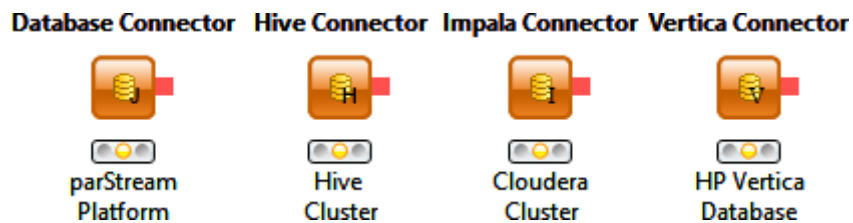


Imagen 9.23 Conectores KNIME Big Data

La conexión a una plataforma de datos grandes en general puede ser bastante compleja, puede requerir de piezas de código dedicado. Sin embargo, KNIME ofrece una serie de nodos conectores para la conexión a bases de datos en general y para las plataformas de grandes volúmenes de datos en particular (ver Imagen 9.23). Algunos nodos conectores han sido diseñados para plataformas específicas de datos grandes, difíciles de codificar y ocultar los más complejos detalles de configuración. Por tanto, estos nodos conectores dedicados proporcionan una ventana de configuración muy simple que sólo requiere los parámetros básicos de acceso, como las credenciales. Dichos nodos conectores están disponibles bajo KNIME Extensión *Big Data*, que se puede adquirir en la tienda KNIME.

Después que la conexión se ha establecido, una serie de nodos cooperadores SQL continúan para seleccionar la tabla original, realizan las consultas SQL requeridas, y recuperan los datos resultantes.

Hay muchas plataformas de *Big Data* disponibles para el analista de datos y, a menudo la elección de uno u otro sólo puede hacerse después de la prueba a fondo de la velocidad y las dificultades de implementación. El uso de nodos conectores KNIME hace dicha prueba fácil y rápida. Un flujo de trabajo puede contener un número de diferentes conectores para tantas plataformas de grandes datos diferentes.

En cada momento, sólo una de las conexiones establecidas se canaliza en la siguiente cascada de nodos cooperadores SQL. Por lo tanto, la ejecución de los nodos de ayuda de SQL en una plataforma de datos sólo requiere la conexión de un nuevo nodo conector a la cascada de nodos cooperadores SQL.

En resumen, la integración de una plataforma de *Big Data* en KNIME es muy sencillo [27]:

1. Arrastre y suelte el nodo conector adecuado en el flujo de trabajo para conectarse a la plataforma de *Big Data* de elección.
2. Configure el nodo conector con los parámetros necesarios para acceder a los datos sobre la plataforma de *Big Data*, es decir: credenciales, dirección URL del servidor y la configuración específica de la plataforma.

3. Definir la consulta SQL para realizar las operaciones de ETL con la ayuda de los nodos de manipulación de SQL. La manipulación de nodos SQL ayuda a construir la consulta SQL correcta.
4. Finalmente, la ejecución de un nodo de recuperación de datos (*Database Connection* - nodo *Table Reader*) permite la recuperación de datos de acuerdo con la consulta SQL previamente construido.

## 9.4. Anexo 4 - Análisis de sentimiento (Sentiment Analysis)

Hoy día, junto con la explosión de medios sociales tales como blogs, foros y redes sociales, se observa una tendencia creciente y masiva de usuarios de Internet que comparten libremente sus opiniones, ideas y actividades, así como también sus experiencias personales y estados emocionales.

Según una estimación del sitio eMarketer [28], 33.7 millones de usuarios de América Latina utilizan Twitter en el año 2015, proyectándose 43.3 millones de usuarios para el año 2018 (ver Imagen 9.24).

La información publicada diariamente por estos usuarios puede tener contenido valioso y relevante para una empresa [29], ya que, a través de ésta, se

puede obtener una visión sobre el sentimiento que despierta una marca, un producto lanzado al mercado, su reputación, entre otros, siendo de interés extraer dichos *posts* y analizarlos para poder tomar decisiones en base a los resultados (modificar su estrategia de marketing, corregir un problema con la experiencia del cliente, o tomar otras acciones). Es evidente además el efecto que las opiniones públicamente expresadas pueden llegar a tener sobre otros potenciales clientes: a la hora de comprar un producto o solicitar un servicio, las personas tienden a buscar en internet la opinión de otras que ya lo han comprado o consumido, y ésta puede afectar directamente (positiva o negativamente) en su intención de comprarlo o no, o peor aún, obtenerlo en la competencia.

La idea detrás del análisis de sentimiento (*Sentiment Analysis*) es, dado un post escrito por un usuario determinado, en una red social determinada y sobre un concepto o tema concreto determinado, detectar su polaridad [30]; es decir, si el mismo es de carácter positivo o negativo. La información textual se puede clasificar básicamente en 2 categorías: hechos y opiniones [31].

Dada la complejidad y las distintas connotaciones y significados que puede tomar el lenguaje escrito, el problema de analizar el hecho / opinión a través de un proceso computacional es particularmente compleja y toma distintos matices. En primer lugar, requiere una diferenciación de la polaridad de las distintas palabras del lenguaje (palabras explícitamente negativas y palabras explícitamente positivas así como sus diferentes grados); pero además, un entendimiento de lo explícito y lo implícito, expresiones irónicas o sarcásticas, palabras ambiguas, manejo de la negación, uso de lenguaje popular o de diferente dialecto, faltas de ortografía, inversión y repetición de letras, así como también opiniones regulares o comparativas, o incluso falsas dado el anonimato que ofrece Internet.

Algunos ejemplos:

<b>Twitter Users in Latin America, by Country, 2012-2018</b>							
<i>millions</i>							
	2012	2013	2014	2015	2016	2017	2018
Brazil	6.6	9.2	12.7	14.6	15.8	16.8	17.6
Mexico	3.6	5.6	7.7	9.5	10.6	11.6	12.6
Argentina	1.4	2.6	3.7	5.0	5.7	6.2	6.6
Other	2.7	3.3	3.9	4.6	5.2	5.8	6.5
<b>Latin America</b>	<b>14.3</b>	<b>20.7</b>	<b>28.0</b>	<b>33.7</b>	<b>37.4</b>	<b>40.4</b>	<b>43.3</b>

*Note: internet users who access their Twitter account via any device at least once per month; numbers may not add up to total due to rounding*  
*Source: eMarketer, May 2014*

172992 www.eMarketer.com

Imagen 9.24 Usuarios Twitter

- ✚ “No hay peor atención al cliente que la de @fibertel. Los odio” (Tweet de @lacopecope, 23/06/2015): opinión explícitamente negativa.
- ✚ “Microsoft no ofrecerá actualizaciones de Windows 10 a copias piratas” (Tweet de @DiarioContraste, 27/06/2015): Es negativo como reseña pero neutral como descripción del producto.
- ✚ “Gracias a los q pensáis q subir el aire acondicionado es mejorar la atención al cliente” (Tweet de @beautyblog\_es, 27/06/2015): Comentario irónico.
- ✚ “Increiiiiiiiiible el concierto de @MelendiOficial” (Tweet de @anajm\_11, 08/08/2015): opinión explícitamente positiva con letras repetidas.
- ✚ “Orrible es que te digan que solo hay pepsi y no hay Coca-cola” (Tweet de @SuperaTNoSoMel, 11/04/2015): opinión con faltas de ortografía, positiva para Coca-Cola pero negativa para Pepsi.

Se han realizado varios estudios respecto al análisis de sentimiento en texto escrito y a distintos niveles [32] [33]:

A nivel de documento, en la cual se clasifica si la polaridad global de un documento en donde se expresa una opinión sobre una entidad en particular es negativa o positiva (por ejemplo, dada la reseña de un producto en particular, determinar si el reseñante expresa una opinión en general positiva o no).

A nivel de oración, en la cual se determina si la polaridad de una oración es positiva o negativa.

A nivel de característica, en la cual se determina si la opinión sobre un determinado rasgo o aspecto del producto o servicio es positiva o negativa. Este nivel, de granularidad más fina que los anteriores, tiene como objetivo aspectos más precisos en el contexto de la opinión.

Ejemplo 1: “La calidad de la llamada en un iPhone es buena, pero la batería tiene una vida muy corta”.

En este ejemplo, se pueden diferenciar dos aspectos del producto: la calidad de la llamada (positiva) y la vida de la batería (negativa).

Ejemplo 2: “Si bien no tienen un muy buen servicio, me encanta comer en este restaurante”. En este ejemplo, el sentimiento y tono general del comentario es positivo, sin embargo se menciona el servicio como negativo, y el restaurante en sí, como positivo [33].

## 9.5. Anexo 5 - Latent Dirichlet Allocation (LDA)

LDA es un modelo probabilístico propuesto por Blei, Ng y Jordan en 2003, el cual determina categorías en base a la distribución de las palabras contenidas en un conjunto de documentos [34]. Ha sido originalmente utilizado en documentos de texto, pero se ha extendido su aplicación a varios dominios [35].

Se considera que un corpus de documentos, (representado como un conjunto de palabras), contiene una cantidad  $K$  de tópicos. El algoritmo determinará los tópicos contenidos en dichos documentos, cada uno de ellos representado por un conjunto de palabras, bajo una distribución de Dirichlet. Estos tópicos se determinan basándose en la premisa de que documentos que traten de un mismo tema, tienen una alta probabilidad que mencionan términos similares; por lo cual, se puede determinar la frecuencia en la cual estos términos co-ocurren entre los distintos documentos de un corpus [35].

Según Blei [36], existen 3 supuestos con respecto al algoritmo a tener en cuenta. Por un lado, el orden de las palabras dentro del documento no es relevante. Por otro lado, el orden de los documentos no es relevante. Finalmente, el número de tópicos se asume conocido y fijo. Para éste último supuesto existen estudios y técnicas para determinar el número óptimo de tópicos para mejorar los resultados del algoritmo [37], ya que un número muy bajo de tópicos puede llevar a una generalización y asignación de documentos que no estén relacionados entre sí, mientras que un número muy alto puede fallar en asignar documentos que sí estén relacionados.

Considerando que:

- ✚  $\theta_{td}$  es la probabilidad de que el documento  $d$  contenga el tópico  $t$
- ✚  $\phi_{tw}$  es la probabilidad de que la palabra  $w$  sea parte del tópico  $t$ ,

la probabilidad de detectar  $w$  en  $d$  es el peso promedio de las probabilidades de palabras por tópico, en donde los pesos son los tópicos probables por documento.

Dado que los pesos cambian por cada documento, la distribución varía de documento a documento, actualizándose los tópicos asignados. La actualización se basa en la prevalencia de las palabras en los tópicos elegidos y la prevalencia de los tópicos en los documentos [38].

## 9.6. Anexo 6 - Solicitud de permisos Facebook

Proceso y dificultad para extraer datos de Facebook:

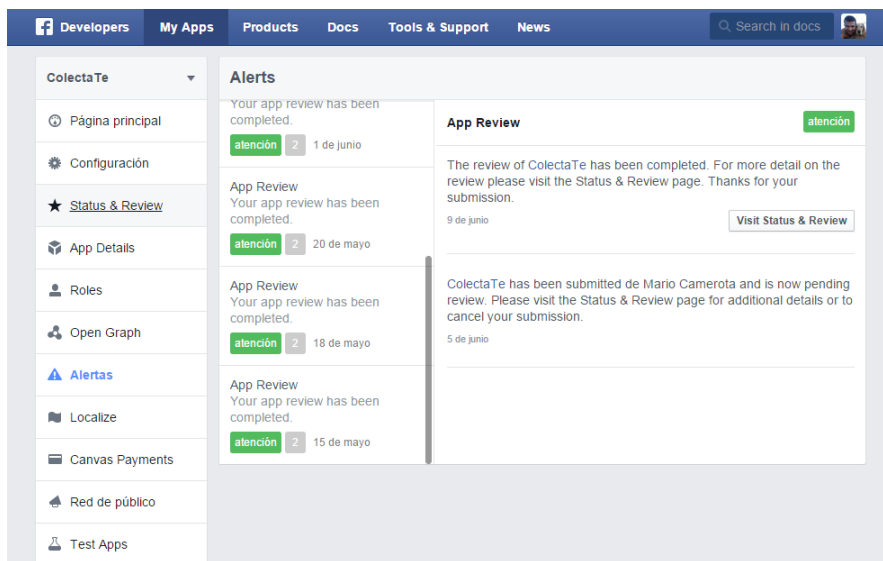


Imagen 9.25 Solicitud de permisos Facebook 1

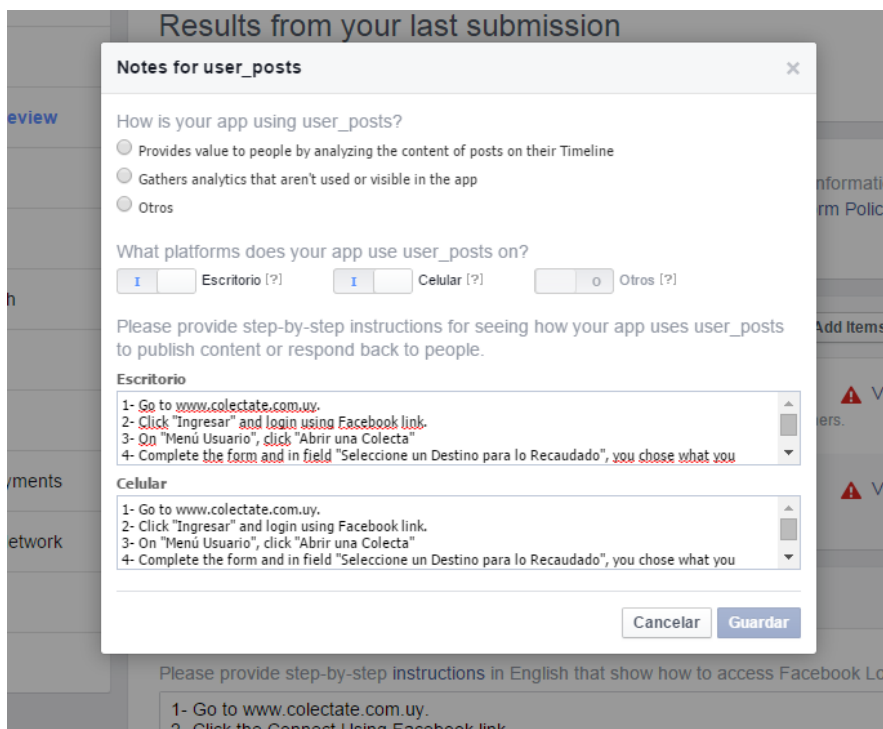


Imagen 9.26 Solicitud de permisos Facebook 2

Facebook otorga a cada aplicación creada (en nuestro caso ColectaTe) permisos aprobados por defecto [39]. Esto quiere decir que sin solicitarlos los administradores de la aplicación cuentan con estos datos de forma inmediata. Los mismos son:

**email:** Proporciona acceso a la dirección de correo electrónico principal de la persona.

**public\_profile:** Proporciona acceso a la información básica de una persona, incluyendo el nombre, el apellido, foto de perfil, el sexo y rango de edad.

**user\_friends:** Proporciona acceso a la lista de amigos que también utilizan su aplicación de una persona.

Por lo tanto, se comenzó con el proceso de solicitar a Facebook los permisos necesarios para poder realizar el análisis comprendido en el alcance de nuestro proyecto [39]. Para esto, es necesario que nos otorguen más permisos y Facebook solicita una explicación detallada por cada permiso solicitado.

Los permisos solicitados fueron:

**user\_posts:** Proporciona acceso a los puestos en la línea de tiempo de una persona, incluyendo sus propios mensajes y los mensajes de los demás.

**user\_status:** Proporciona acceso a los estados de una persona. Estos son los mensajes en Facebook que no incluyen enlaces, vídeos o fotos.

**read\_stream:** Proporciona acceso a leer los mensajes en una persona *News Feed*, o los mensajes en su perfil.

**user\_events:** Proporciona acceso de sólo lectura a los acontecimientos que una persona está organizando o ha organizado.

Los permisos fueron solicitados cinco veces (ver Imagen 9.25 e Imagen 9.26), cada una de ellas fue rechazada por Facebook con una respuesta muy general como se muestra en la Imagen 9.27.

Facebook para Uruguay, no tiene un contacto telefónico ni un lugar donde realizar reclamos

y/o consultas. Analizando e investigando por diferentes vías, se llegó a lo siguiente: hasta la GRAPH API 2.0, si se podía obtener los datos de los “posts” de las personas y de ésta manera realizar el análisis de sentimiento (*sentiment analysis*).

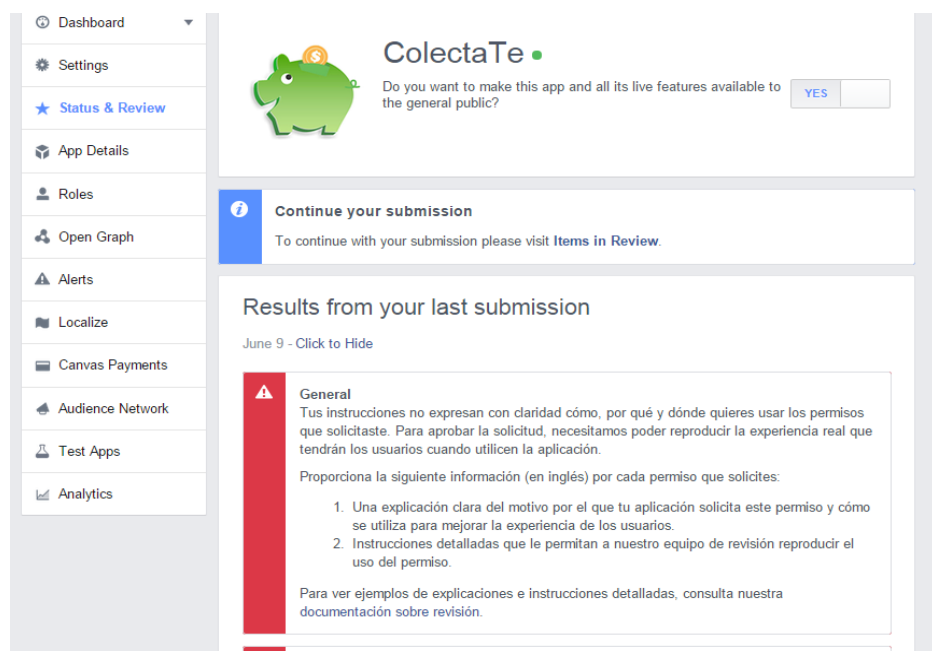


Imagen 9.27 Rechazo de permisos Facebook

A partir de la 2.3 (que es la actual y no se puede obtener una versión anterior) esto ya no es posible. Por lo tanto, aun obteniendo los permisos solicitados (algo que tampoco ha ocurrido) sólo será posible acceder a los *posts* de las personas que han dado “Me Gusta” en la aplicación, que son muy pocos en la actualidad.

A continuación se describe cómo se solicitaron los permisos (se deben solicitar en inglés):

1. *Go to [www.colectate.com.uy](http://www.colectate.com.uy).*
2. *Click "Ingresar" and login using Facebook link.*
3. *On "Menú Usuario", click "Abrir una Colecta"*
4. *Complete the form and in field "Seleccione un Destino para lo Recaudado", you chose what you would like to do with the money you raise. With this information we will be able to provide more options in this combobox for the user to choose based on general interest of our customers and prioritize them based on their interests.*

Se consultó a un experto del área, Mauricio Clausen, Co-Founder de IDATHA, empresa que se dedica a *Media Analytics, Business Intelligence y Data Science Consulting* y nos confirmó las dificultades para poder extraer datos desde Facebook según se ve en la Imagen 9.28.

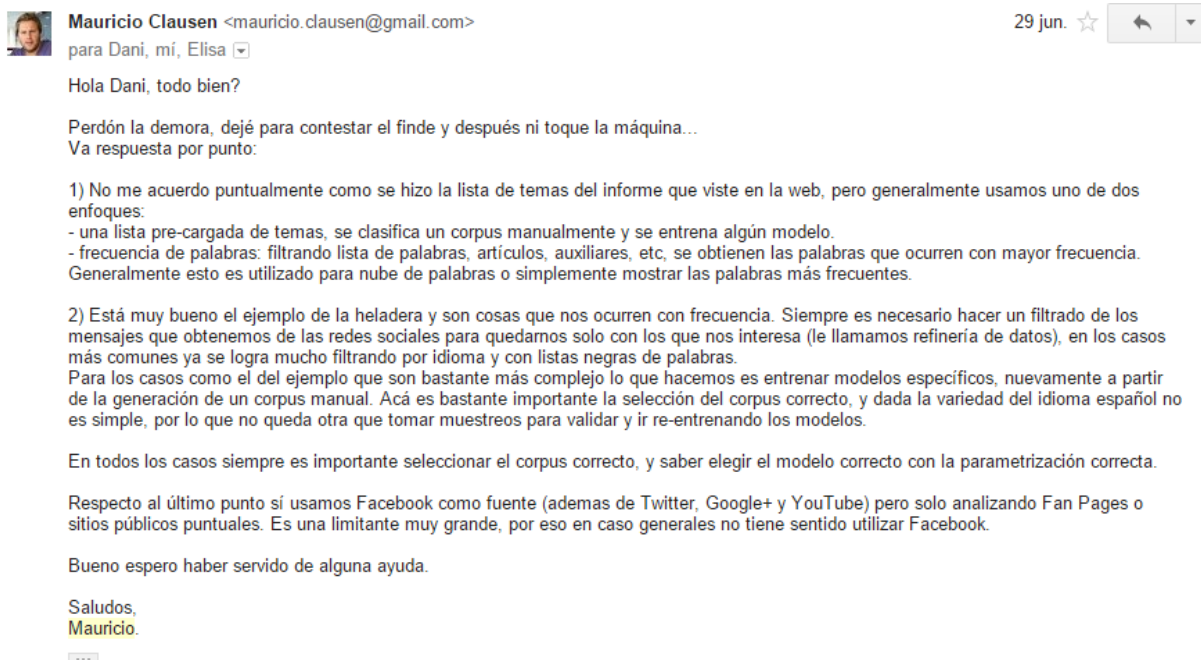


Imagen 9.28 Mail IDATHA

## 9.7. Anexo 7 - Plan de gestión de riesgos

### 9.7.1.1. Enfoque para la gestión de riesgos

El plan de gestión de riesgos describe el proceso para realizar gestión de los riesgos inherentes al proceso de desarrollo de *software* y a las actividades dentro del alcance del proyecto.

En la etapa de planificación los integrantes del equipo desarrollan el plan de gestión de riesgos, en conjunto con los interesados del proyecto. En estas reuniones se define la escala a utilizar de impacto, ocurrencia y ocurrencia en el tiempo, así como las estrategias genéricas.

En adición a esto, se definen las tareas relacionadas con la gestión de riesgos y se realiza un análisis primario de riesgos, identificando la magnitud de los riesgos, los disparadores y acciones de mitigación y contingencia, en base al análisis cualitativo.

Posterior a esto, en la etapa de ejecución, se realiza el seguimiento de los riesgos y se comunica su estado.

### 9.7.1.2. Escalas

#### Escala de impacto (I)

Valor	Referencia
0	Ninguno
1	Marginal
2	Poco importante
3	Importante (puede retrasar el proyecto)
4	Crítica (puede detener el proyecto)
5	Catastrófica (fracaso del proyecto)

Tabla 9.2 Escala de impacto riesgos

#### Escala de probabilidad de ocurrencia (PO)

Valor	Referencia
0,0	No probable
0,2	Poco probable
0,4	Probable
0,6	Muy probable
0,8	Altamente probable
1,0	Se convierte en problema

Tabla 9.3 Escala de probabilidad de ocurrencia riesgos

**Escala de ocurrencia en el tiempo (OT)**

Valor	Referencia
0	Corto Plazo
1	Mediano Plazo
2	Largo plazo

Tabla 9.4 Escala de ocurrencia en el tiempo riesgos

**9.7.1.3. Matriz de estrategias genéricas**

En base a la magnitud de cada riesgo se le asigna una estrategia genérica preestablecida en la siguiente tabla. Dependiendo del tipo de estratégica es el tipo de acción que se debe realizar.

Impacto	Probabilidad	Poco probable (0,2)	Probable (0,4)	Muy probable (0,6)	Altamente probable (0,8)	Problema (1,0)
Marginal (1)		Aceptar	Aceptar	Aceptar	Aceptar	Mitigar
Poco importante (2)		Aceptar	Aceptar	Aceptar	Mitigar	Mitigar
Importante (3)		Aceptar	Mitigar	Mitigar	Mitigar	Evitar
Crítica (4)		Mitigar	Mitigar	Evitar	Evitar	Evitar
Catastrófica (5)		Mitigar	Evitar	Evitar	Evitar	Evitar

Tabla 9.5 Matriz de estrategias genéricas riesgos

## 9.8. Anexo 8 - Técnica de estimación

Para realizar la estimación de los requerimientos se utilizan variantes de tres técnicas:

- ✚ *Planning poker*
- ✚ *T-Shirt size*
- ✚ *Relative mass evaluation*

Se realiza un taller de escritura, donde participan todos los miembros del equipo, aquí se acuerda cómo se escribirían las *User Stories* (US) con sus condiciones de satisfacción.

Las mismas son escritas en post-it y luego son ordenadas en orden de tamaño (*Relative mass valuation*) por los miembros del equipo. Se trata de comenzar por una pequeña y luego se van colocando las demás en base a la comparación relativa con las demás.

Una vez finalizado el punto anterior, se procede a utilizar la técnica *T-Shirt size*, para lo cual a cada US, el equipo le asigna una talla de remera según su complejidad. Se debe tomar en cuenta el orden definido en el paso anterior, si bien el mismo puede cambiar.



Imagen 9.29 T-Shirt size [44]

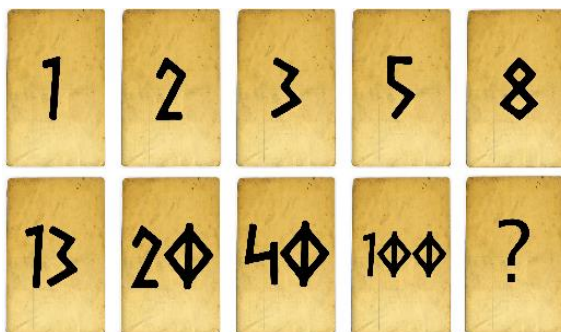


Imagen 9.30 Planning Poker [45]

Para finalizar se utiliza la técnica *Planning poker*, donde la estimación de cada US se realiza en *Story Points*, estableciendo el siguiente rango 1, 2, 3, 5, 8, 13, 20, 40, 100 de SP.

Dado que la técnica *Planning poker* se basa en la complejidad relativa entre US, se comienza estimando la más sencilla, a la que se le pone el menor valor, y luego se van tomando las restantes US y asignando un valor en base a qué

tanto más complejas son comparadas con las anteriores ya estimadas.

Este proceso se repite por cada US, donde cada participante en la estimación le asigna un valor y luego las discrepancias se justifican y se llega a un acuerdo de estimación común para cada US.

A la hora de realizar la estimación de un *Sprint*, se compromete la realización de ciertas US en base a la disponibilidad del equipo tomando en cuenta cuantos SP se pueden realizar en dicha disponibilidad.

Dado que en el primer *Sprint*, no se cuenta con estimaciones anteriores, no se puede realizar una estimación primaria comparando con *Sprints* previos, pero en los subsiguientes *Sprints* la estimación anterior es tomada en cuenta.

Para realizar la estimación de la confección de la documentación, se utiliza el juicio experto. Para realizar la misma se genera una instancia donde los integrantes del equipo se reúnen y cada uno propone un esfuerzo estimado para cada ítem.

En base a las diferencias, se procede a explicar las razones del esfuerzo sugerido por cada miembro y repetir los pasos hasta llegar a un acuerdo para cada ítem a estimar.

## 9.9. Anexo 9 - Product Backlog

ID	Prioridad	User Story
RF-BD01		Proveer un mecanismo que permita analizar cuáles son los productos y servicios más populares en las redes sociales de los usuarios de Uruguay a partir de los productos ya ofrecidos por las empresas en el mercado ( <i>Epic</i> )
RF-BD02		Proveer un mecanismo que permita identificar las tendencias de productos y servicios que actualmente no son ofrecidos por ColectaTe y que podrían ser una nueva oportunidad de negocio ( <i>Epic</i> )
US-BD01	Alta	Como dueño, quiero acceder al flujo de tweets que hacen referencia a una o más palabras claves de usuarios de Uruguay con el objetivo de analizar los mismos para descubrir nuevos productos y/o servicios y ofrecerlos en la lista de las acciones a realizar con el dinero recaudado en las colectas
		<ul style="list-style-type: none"> <li>- Verificar que los datos extraídos se componen de nuevos datos no extraídos anteriormente y datos ya extraídos en la ejecución previa del proceso</li> <li>- Verificar que luego de descartar los datos ya extraídos previamente, no existen datos con un valor menor al valor del atributo de control</li> <li>- Verificar que los datos extraídos sean provenientes de usuarios de Uruguay</li> <li>- Verificar que los datos extraídos hacen referencia a una o más palabras claves</li> </ul>
US-BD02	Cancelado	Como dueño, quiero acceder al flujo de <i>posts</i> de Facebook que hacen referencia a una o más palabras claves de usuarios de Uruguay con el objetivo de analizar los mismos para descubrir nuevos productos y/o servicios y ofrecerlos en la lista de las acciones a realizar con el dinero recaudado en las colectas
		<ul style="list-style-type: none"> <li>- Verificar que los datos extraídos se componen de nuevos datos no extraídos anteriormente y datos ya extraídos en la ejecución previa del proceso</li> <li>- Verificar que luego de descartar los datos ya extraídos previamente, no existen datos con un valor menor al valor del atributo de control</li> <li>- Verificar que los datos extraídos sean provenientes de usuarios de Uruguay</li> <li>- Verificar que los datos extraídos hacen referencia a una o más palabras claves</li> </ul>

US-BD03	Alta	Como dueño, quiero acceder al flujo de tweets de usuarios de Uruguay con el objetivo de analizar los mismos para descubrir nuevos productos y/o servicios y ofrecerlos en la lista de las acciones a realizar con el dinero recaudado en las colectas
		<ul style="list-style-type: none"> <li>- Verificar que los datos extraídos se componen de nuevos datos no extraídos anteriormente y datos ya extraídos en la ejecución previa del proceso</li> <li>- Verificar que luego de descartar los datos ya extraídos previamente, no existen datos con un valor menor al valor del atributo de control</li> <li>- Verificar que los datos extraídos sean provenientes de usuarios de Uruguay</li> </ul>
US-BD04	Cancelado	Como dueño, quiero acceder al flujo de <i>posts</i> de Facebook de usuarios de Uruguay con el objetivo de analizar los mismos para descubrir nuevos productos y/o servicios y ofrecerlos en la lista de las acciones a realizar con el dinero recaudado en las colectas
		<ul style="list-style-type: none"> <li>- Verificar que los datos extraídos se componen de nuevos datos no extraídos anteriormente y datos ya extraídos en la ejecución previa del proceso</li> <li>- Verificar que luego de descartar los datos ya extraídos previamente, no existen datos con un valor menor al valor del atributo de control</li> <li>- Verificar que los datos extraídos sean provenientes de usuarios de Uruguay</li> </ul>
US-BD05	Cancelado	Como dueño, quiero unificar los datos extraídos de ambas fuentes con el objetivo de facilitar el análisis y visualización de los mismos para descubrir nuevos productos y/o servicios y ofrecerlos en la lista de las acciones a realizar con el dinero recaudado en las colectas
		<ul style="list-style-type: none"> <li>- Verificar que la cantidad de tuplas unificadas de ambas fuentes sea igual a la suma de las tuplas extraídas de ambas fuentes por separado luego de realizar el descarte</li> </ul>
US-BD06	Alta	Como dueño, quiero analizar los datos procesados obteniendo una tabla con la lista completa de los mismos junto al grado de popularidad con el objetivo de facilitar su visualización para descubrir nuevos productos y/o servicios y ofrecerlos en la lista de las acciones a realizar con el dinero recaudado en las colectas
		<ul style="list-style-type: none"> <li>- Verificar que los datos analizados se basan en opiniones positivas, negativas y neutras</li> <li>- Verificar que la tabla obtenida contenga las siguientes columnas: identificador, cantidad de menciones</li> </ul>

		- Verificar que se haya aplicado correctamente la calidad de datos, habiéndose descartado palabras de menos de 3 caracteres, puntuación, caracteres especiales, <i>links</i> .
US-BD07	Alta	Como dueño, quiero analizar los datos procesados obteniendo una tabla con la lista de productos y servicios de interés de ColectaTe junto al grado de popularidad con el objetivo de facilitar su visualización para descubrir nuevos productos y/o servicios y ofrecerlos en la lista de las acciones a realizar con el dinero recaudado en las colectas
		- Verificar que los datos analizados se basan en opiniones positivas, negativas y neutras de los productos y/o servicios - Verificar que los datos analizados mencionan al menos un producto o servicio de la lista de interés dada por ColectaTe - Verificar que la tabla obtenida contenga las siguientes columnas: identificador de producto o servicio, cantidad de menciones - Verificar que se haya aplicado correctamente la calidad de datos, habiéndose descartado palabras de menos de 3 caracteres, puntuación, caracteres especiales, <i>links</i> .
US-BD08	Alta	Como dueño, quiero almacenar los datos analizados con el objetivo de consolidarlos semanalmente para descubrir nuevos productos y/o servicios y ofrecerlos en la lista de las acciones a realizar con el dinero recaudado en las colectas
		- Verificar que el valor del campo "Fecha de actualización" sea igual a la fecha de la ejecución del último proceso de almacenado
US-BD09	Alta	Como dueño, quiero poder visualizar de forma semanal un reporte que muestre un gráfico con el top 10 de los productos y servicios más populares entre los usuarios de Uruguay en las redes sociales; y que contenga además una tabla con la lista completa de los mismos junto al grado de popularidad con el objetivo de descubrir nuevos productos y/o servicios y ofrecerlos en la lista de las acciones a realizar con el dinero recaudado en las colectas
		- Verificar que el reporte contiene el gráfico con los productos y servicio más populares - Verificar que el reporte contenga una tabla con cuatro columnas: identificación del producto o servicio, cantidad de menciones positivas, cantidad de menciones negativas - Verificar que el reporte contenga una tabla que muestre la cantidad total de registros analizados

US-BD10	Alta	Como dueño, quiero poder visualizar de forma semanal un reporte que muestre un gráfico con el top 10 de los productos y servicios más populares de una lista dada entre los usuarios de Uruguay en las redes sociales; y que contenga además la tabla con la lista completa de los mismos junto al grado de popularidad con el objetivo de ofrecer los más populares en la lista de las acciones a realizar con el dinero recaudado en las colectas
		<ul style="list-style-type: none"> <li>- Verificar que el reporte contiene el gráfico con los productos y servicio más populares de la lista dada</li> <li>- Verificar que el reporte contenga una tabla con cuatro columnas: identificación del producto o servicio, cantidad de menciones positivas, cantidad de menciones negativas</li> <li>- Verificar que los productos y servicios mencionados en la tabla según la columna "identificación" se corresponden con los definidos en la lista dada</li> <li>- Verificar que el reporte contenga una tabla que muestre la cantidad total de registros analizados</li> </ul>
RF-BD03		Proveer un mecanismo que permita conocer la visión de los clientes respecto a la calidad del servicio ofrecido por ColectaTe en base al análisis de la información ( <i>Sentiment Analysis</i> ) en las redes sociales ( <i>Epic</i> )
RF-BD04		Proveer un mecanismo que permita determinar las oportunidades de mejora en base al análisis de la información ( <i>Sentiment Analysis</i> ) en las redes sociales ( <i>Epic</i> )
US-BD11	Alta	Como dueño, quiero acceder al flujo de tweets que hacen referencia a ColectaTe con el objetivo de analizar los mismos y entender la opinión de los usuarios respecto a la empresa e identificar oportunidades de mejora
		<ul style="list-style-type: none"> <li>- Verificar que los datos extraídos se componen de nuevos datos no extraídos anteriormente y datos ya extraídos en la ejecución previa del proceso</li> <li>- Verificar que luego de descartar los datos ya extraídos previamente, no existen datos con un valor menor al valor del atributo de control</li> <li>- Verificar que los datos extraídos hacen referencia a la palabra ColectaTe</li> </ul>
US-BD12	Cancelado	Como dueño, quiero acceder al flujo de <i>posts</i> de Facebook que hacen referencia a ColectaTe con el objetivo de analizar los mismos y entender la opinión de los usuarios respecto a la empresa e identificar oportunidades de mejora
		<ul style="list-style-type: none"> <li>- Verificar que los datos extraídos se componen de nuevos datos no extraídos anteriormente y datos ya extraídos en la ejecución previa del proceso</li> </ul>

		<ul style="list-style-type: none"> <li>- Verificar que luego de descartar los datos ya extraídos previamente, no existen datos con un valor menor al valor del atributo de control</li> <li>- Verificar que los datos extraídos hacen referencia a la palabra ColectaTe</li> </ul>
US-BD13	Cancelado	Como dueño, quiero unificar los datos extraídos de ambas fuentes con el objetivo de facilitar el análisis y visualización de los mismos para entender la opinión de los usuarios respecto a la empresa e identificar oportunidades de mejora
		<ul style="list-style-type: none"> <li>- Verificar que la cantidad de tuplas unificadas de ambas fuentes sea igual a la suma de las tuplas extraídas de ambas fuentes por separado luego de realizar el descarte</li> </ul>
US-BD14	Alta	Como dueño, quiero poder analizar los datos procesados obteniendo información sobre la percepción y opinión ( <i>Sentiment Analysis</i> ) de los usuarios con el objetivo de facilitar su visualización para entender la opinión de los usuarios respecto a la empresa e identificar oportunidades de mejora
		<ul style="list-style-type: none"> <li>- Verificar que la tabla obtenida contenga las siguientes columnas: cantidad de menciones, cantidad de menciones positivas, cantidad de menciones negativas</li> <li>- Verificar que se haya aplicado correctamente la calidad de datos, habiéndose descartado palabras de menos de 3 caracteres, puntuación, caracteres especiales, <i>links</i>.</li> </ul>
US-BD15	Alta	Como dueño, quiero analizar los datos procesados obteniendo una lista categorizada de oportunidades de mejora con el objetivo de facilitar su visualización para entender la opinión de los usuarios respecto a la empresa e identificar oportunidades de mejora
		<ul style="list-style-type: none"> <li>- Verificar que la lista obtenida contenga las siguientes columnas: identificador, cantidad de menciones, cantidad de menciones positivas, cantidad de menciones negativas</li> <li>- Verificar que se haya aplicado correctamente la calidad de datos, habiéndose descartado palabras de menos de 3 caracteres, puntuación, caracteres especiales, <i>links</i>.</li> </ul>
US-BD16	Alta	Como dueño, quiero almacenar los datos analizados con el objetivo de consolidarlos semanalmente para entender la opinión de los usuarios respecto a la empresa e identificar oportunidades de mejora
		<ul style="list-style-type: none"> <li>- Verificar que el valor del campo "Fecha de actualización" sea igual a la fecha de la ejecución del último proceso de almacenado</li> </ul>

US-BD17	Alta	Como dueño, quiero poder visualizar de forma semanal un reporte que muestre un gráfico con la información sobre la percepción y opinión ( <i>Sentiment Analysis</i> ) con el objetivo de entender la opinión de los usuarios respecto a la empresa
		<ul style="list-style-type: none"> <li>- Verificar que el reporte contiene el gráfico con el análisis de sentimiento</li> <li>- Verificar que el reporte contenga una tabla que muestre la cantidad total de registros analizados</li> </ul>
US-BD18	Alta	Como dueño, quiero poder visualizar de forma semanal un reporte que muestre un gráfico de las oportunidades de mejora de forma categorizada según la cantidad de menciones de cada categoría, con el objetivo de entender la opinión de los usuarios respecto a la empresa e identificar oportunidades de mejora
		<ul style="list-style-type: none"> <li>- Verificar que el reporte contiene el gráfico con el top 10 de las opiniones negativas más mencionadas</li> <li>- Verificar que el reporte contenga una tabla con dos columnas: opiniones negativas, cantidad de menciones</li> <li>- Verificar que el reporte contenga una tabla que muestre la cantidad total de registros analizados</li> </ul>
US-BD19	Media	Como dueño, quiero acceder al flujo de tweets en tiempo real que hacen referencia a ColectaTe con el objetivo de analizar los mismos y entender la opinión de los usuarios respecto a la empresa y tomar acciones inmediatas
		<ul style="list-style-type: none"> <li>- Verificar que los datos se extraen de Twitter en tiempo real</li> <li>- Verificar que los datos extraídos se componen de nuevos datos no extraídos anteriormente y datos ya extraídos en la ejecución previa del proceso</li> <li>- Verificar que luego de descartar los datos ya extraídos previamente, no existen datos con un valor menor al valor del atributo de control</li> <li>- Verificar que los datos extraídos hacen referencia a la palabra ColectaTe</li> </ul>
US-BD20	Media	Como dueño, quiero poder analizar los datos procesados en tiempo real obteniendo información sobre la percepción y opinión ( <i>Sentiment Analysis</i> ) de los usuarios con el objetivo de facilitar su visualización para entender la opinión de los usuarios respecto a la empresa y tomar acciones inmediatas
		<ul style="list-style-type: none"> <li>- Verificar que los datos se analizan en tiempo real luego de ser extraídos</li> <li>- Verificar que la tabla obtenida contenga las siguientes columnas: cantidad de tweets extraídos, cantidad de tweets valorados, cantidad de menciones positivas, cantidad de menciones negativas, cantidad de menciones neutras</li> </ul>

		<ul style="list-style-type: none"> <li>- Verificar que se haya aplicado correctamente la calidad de datos, habiéndose descartado palabras de menos de 3 caracteres, puntuación, caracteres especiales, <i>links</i>.</li> <li>- Verificar que los datos se guardan en tiempo real luego de ser analizados</li> <li>- Verificar que el valor del campo "Fecha de actualización" sea igual a la fecha de la ejecución del último proceso de almacenado</li> </ul>
US-BD21	Media	<p>Como dueño, quiero poder visualizar en tiempo real un gráfico con la información sobre la percepción y opinión (<i>Sentiment Analysis</i>), además de la información específica del tweet con el objetivo de entender la opinión de los usuarios respecto a la empresa y tomar acciones inmediatas</p>
		<ul style="list-style-type: none"> <li>- Verificar que los datos que ingresan se reflejan en el/los gráficos en tiempo real</li> <li>- Verificar que el reporte contiene el gráfico con el análisis de sentimiento</li> <li>- Verificar que el reporte contiene los datos específicos de los tweets</li> <li>- Verificar que el reporte contenga una tabla que muestre la cantidad total de registros analizados</li> </ul>
US-BD22	Baja	<p>Como dueño, quiero mejorar el resultado obtenido como parte de los procesos con el objetivo de facilitar su visualización para descubrir nuevos productos y/o servicios y ofrecerlos en la lista de las acciones a realizar con el dinero recaudado en las colectas</p>
		<ul style="list-style-type: none"> <li>- Verificar que no existan datos duplicados a la hora de analizar</li> <li>- Verificar se excluyen los hashtags de extracción a la hora de ejecutar el proceso</li> </ul>
RF-AM01		Se requiere una aplicación móvil que tenga las mismas funcionalidades que el sitio web <a href="http://www.colectate.com.uy">www.colectate.com.uy</a> ( <i>Epic</i> )
US-AM01	Baja	<p>Como dueño, quiero que la aplicación cuente con un mecanismo que permita que los usuarios se registren con el objetivo de poder identificarse y autenticarse en la aplicación</p>
		<ul style="list-style-type: none"> <li>- Verificar que al registrarse al sistema aparezca una pantalla donde se le solicite a la persona el nombre, apellido, mail, contraseña, teléfono, <i>checkbox</i> de aceptación de términos y condiciones y el botón de "registrarse". Además un botón de Facebook que permita el registro a través del mismo</li> <li>- Verificar que los campos no se encuentren vacíos</li> <li>- Verificar que el campo "mail" contenga una dirección con formato válido</li> <li>- Verificar que el campo "teléfono" sea únicamente numérico</li> <li>- Verificar que el campo "contraseña" tenga un mínimo de 6 caracteres alfanuméricos</li> </ul>

		<ul style="list-style-type: none"> <li>- Verificar que el <i>checkbox</i> de aceptación de términos sea seleccionado</li> <li>- Verificar que al presionar "registrarse" se muestra un mensaje de error cuando algunos de los campos tiene datos incorrectos</li> <li>- Verificar que al presionar "registrarse" si todos los datos se encuentran correctos se crea el usuario</li> <li>- Verificar que al presionar "registrarse" si el usuario ya existe muestre un mensaje de error y no se cree el usuario</li> </ul>
US-AM02	Baja	<p>Como dueño, quiero que los usuarios se deban identificar en la aplicación antes de comenzar a utilizarla con el objetivo de hacer uso de las funcionalidades de la aplicación</p>
		<ul style="list-style-type: none"> <li>- Verificar que al ingresar al sistema aparezca una pantalla donde se le solicite a la persona el mail, la contraseña y el botón "ingresar". Además un botón de Facebook que permita el logueo a través del mismo. También brindar un <i>link</i> para solicitar la contraseña en caso de olvido y un <i>link</i> a la funcionalidad de registrarse</li> <li>- Verificar que los campos no se encuentren vacíos</li> <li>- Verificar que el campo "mail" contenga una dirección con formato válido</li> <li>- Verificar que al presionar "ingresar" se muestra un mensaje de error cuando algunos de los campos tiene datos incorrectos</li> <li>- Verificar que al presionar "ingresar" si todos los datos se encuentran correctos se inicia sesión</li> <li>- Verificar que al presionar "ingresar" si el usuario no existe o la contraseña es errónea muestre un mensaje de error y no deje iniciar sesión</li> </ul>
US-AM03	Baja	<p>Como dueño, quiero que la aplicación cuente con una página principal con información general con el objetivo de facilitar la difusión de las colectas activas y aumentar la participación en las mismas</p>
		<ul style="list-style-type: none"> <li>- Verificar que la página principal cuente con el campo "buscar"</li> <li>- Verificar que la página principal cuente con una sección de "colectas solidarias" donde se pueda ver información de las colectas solidarias activas actualmente</li> <li>- Verificar que se visualice el menú de usuario</li> <li>- Verificar que se visualice el botón para crear colectas</li> </ul>
US-AM04	Baja	<p>Como dueño, quiero que la aplicación cuente con un menú con el objetivo de facilitar el acceso a los usuarios a las diferentes funciones de la aplicación</p>
		<ul style="list-style-type: none"> <li>- Verificar que existe un menú con las siguientes funcionalidades: abrir una colecta, depositar en colecta, retirar fondos, invitar, mis colectas, mis depósitos, mis retiros y mi perfil</li> </ul>

US-AM05	Baja	Como usuario, quiero poder cerrar sesión en la aplicación con el objetivo de asegurar que no se puedan realizar acciones a mi nombre en la aplicación
		<ul style="list-style-type: none"> <li>- Verificar que en el menú de usuario existe la opción de "cerrar sesión"</li> <li>- Verificar que al cerrar sesión se redirecciona al usuario a la página principal y que se visualizan en ella los botones "registrarse" e "ingresar"</li> </ul>
US-AM06	Baja	Como usuario, quiero poder abrir una colecta con el objetivo de que otros usuarios depositen dinero en ella
		<ul style="list-style-type: none"> <li>- Verificar que al acceder a la función de "abrir colecta" se visualizan los siguientes campos: número de colecta, fecha límite, motivo de la colecta, título de la colecta, descripción de la colecta, nombre y apellido del beneficiario, destino de lo recaudado</li> <li>- Verificar que si se elige destino de lo recaudado con valor "cheque" o "tarjeta regalo OCA" o "tarjeta regalo <i>gift card</i>", se visualiza: domicilio de entrega</li> <li>- Verificar que si se elige destino de lo recaudado con valor "depósito en cuenta corriente" o "depósito en caja de ahorro", se visualiza: seleccione bando, ingrese número de cuenta</li> <li>- Verificar que se visualicen las opciones de "mostrar a los participantes el monto recaudado" y "mostrar a los participantes el promedio recaudado"</li> <li>- Verificar que se visualice el botón "crear colecta"</li> <li>- Verificar que al presionar "crear colecta" se muestra un mensaje de error cuando algunos de los campos tiene datos incorrectos</li> <li>- Verificar que al presionar "crear colecta" si todos los datos se encuentran correctos se crea la colecta</li> <li>- Verificar que luego de creada la colecta se le despliega al usuario un mensaje de confirmación y se visualizan las opciones para compartir la colecta en redes sociales y correo electrónico</li> </ul>
US-AM07	Baja	Como usuario, quiero poder depositar en una colecta con el objetivo de colaborar a su causa
		<ul style="list-style-type: none"> <li>- Verificar que una vez seleccionada la colecta en la cual se desea participar se visualizan los detalles de la colecta y los siguientes campos para realizar el depósito: nombre del depositante, mensaje para el beneficiario, moneda, monto a depositar</li> <li>- Verificar que se visualizan las formas de pago, de las cuales se puede seleccionar una sola mediante los <i>radio buttons</i></li> <li>- Verificar que se visualiza el botón "participar"</li> <li>- Verificar que al presionar "participar" se muestra un mensaje de error cuando algunos de los campos tiene datos incorrectos</li> </ul>

		<ul style="list-style-type: none"> <li>- Verificar que al presionar "participar" si todos los datos se encuentran correctos se redirecciona a la url correspondiente según el medio de pago seleccionado</li> <li>- Verificar que luego de realizado el pago, se redirecciona al usuario a la página del menú de usuario</li> </ul>
US-AM08	Baja	<p>Como usuario, quiero poder invitar a participar a otras personas en una colecta con el objetivo de aumentar la cantidad de personas que colaboran con la causa</p>
		<ul style="list-style-type: none"> <li>- Verificar que una vez seleccionada la opción "invitar" del menú usuario, se visualizan los siguientes campos: colecta, mails de los participantes a invitar y mensaje</li> <li>- Verificar que se visualiza el botón "enviar invitación"</li> <li>- Verificar que al presionar el botón, se despliega un mensaje de confirmación</li> <li>- Verificar que al presionar aceptar, el usuario es redirigido al menú usuario</li> <li>- Verificar que el correo electrónico fue enviado correctamente</li> </ul>
US-AM09	Baja	<p>Como usuario, quiero poder retirar los fondos recaudados de las colectas que administro con el objetivo de hacer usufructo del dinero</p>
		<ul style="list-style-type: none"> <li>- Verificar que una vez seleccionada la opción "retirar fondos" del menú usuario, se visualiza una tabla mostrando las colectas activas del usuario y los siguientes campos: colecta e importe a retirar</li> <li>- Verificar que se visualiza el botón "retirar fondos"</li> <li>- Verificar que al ingresar un monto mayor al saldo de la cuenta, se despliega un mensaje de error</li> <li>- Verificar que al presionar el botón, se despliega un mensaje indicando que se realizará el pago del monto establecido</li> </ul>
US-AM10	Baja	<p>Como usuario, quiero poder visualizar los detalles de las colectas que administro con el objetivo de tener un panorama general del estado de las mismas</p>
		<ul style="list-style-type: none"> <li>- Verificar que al ingresar al menú "mis colectas" se visualiza una tabla que las colectas activas del usuario con la siguiente información: número de colectas, fecha de inicio y cierre, beneficiario, destino de lo recaudado, monto bruto, costo comisión, monto neto, saldo, <i>link</i> a ver depósitos, <i>link</i> a ver retiros</li> <li>- Verificar que se muestra una tabla de colectas finalizadas con la misma información de la tabla de colectas activas</li> </ul>
US-AM11	Baja	<p>Como usuario, quiero poder visualizar los detalles de los depósitos realizados en las colectas que administro con el objetivo de saber quiénes depositaron en las colectas</p>

		<p>- Verificar que al ingresar al <i>link</i> "ver depósitos" desde la pantalla de datos generales de la colecta se despliega una tabla con el desglose de los depósitos: número de participación, nombre del depositante, fecha, monto bruto, costo comisión, monto neto, moneda, medio de pago, estado</p> <p>- Verificar que al final de la tabla se muestran los totales de las columnas de monto bruto, costo comisión, monto neto</p> <p>- Verificar que se muestra una tabla con el desglose de los pagos pendientes, con los mismos campos que la tabla de los depósitos realizados</p>
US-AM12	Baja	<p>Como usuario, quiero poder visualizar los detalles de los depósitos que he realizado en las diferentes colectas con el objetivo de tener un histórico de mi participación en las colectas</p> <p>- Verificar que al ingresar al <i>link</i> "mis depósitos" se visualiza una tabla con los campos: número de colecta, beneficiario, fecha, monto bruto, estado</p>
US-AM13	Baja	<p>Como usuario, quiero poder visualizar los detalles de los retiros que he realizado en las colectas que administro con el objetivo de llevar un control financiero y controlar que el saldo sea correcto</p> <p>- Verificar que al ingresar al <i>link</i> "ver retiros" desde la pantalla de datos generales de la colecta se despliega una tabla con el desglose de los retiros: número colecta, beneficiario, fecha, monto retirado</p>
US-AM14	Baja	<p>Como usuario, quiero poder compartir en redes sociales diferentes colectas con el objetivo de aumentar la cantidad de personas que colaboran con la causa</p> <p>- Verificar que una vez seleccionada la opción "compartir" (de una de las redes sociales) en una colecta se abre una nueva pestaña con la página de la red social correspondiente, con el texto predefinido invitando a participar de la colecta y con el <i>link</i> de la misma.</p> <p>- Verificar que al presionar el botón correspondiente para compartir (según la red social), se comparte la colecta en la red social</p>
US-AM15	Baja	<p>Como usuario, quiero poder gestionar mi perfil de usuario con el objetivo de mantener mi información actualizada</p> <p>- Verificar que al acceder a la función de "mi perfil" se visualizan los siguientes campos: correo electrónico, nombre, apellido, contraseña (enmascarada), teléfono, número de documento, dirección, fecha de nacimiento, sexo, correo electrónico secundario y los botones "ir al menú usuario" y "actualizar"</p>

		<ul style="list-style-type: none"><li>- Verificar que el campo correo electrónico no se pueda modificar y los demás si sean editables</li><li>- Verificar que al presionar el botón "ir al menú usuario" se redirecciona al menú de usuario sin guardar cualquier cambio que se haya realizado</li><li>- Verificar que al presionar el botón "actualizar", se muestra un mensaje de error cuando algunos de los campos tiene datos incorrectos</li><li>- Verificar que al presionar el botón "actualizar, si todos los datos se encuentran correctos actualizan los datos del perfil del usuario, se muestra un mensaje de confirmación y se redirecciona al menú de usuario</li></ul>
--	--	--

## 9.10. Anexo 10 - Estimación Product Backlog

ID	Relative Mass Order	T-Shirt Size	SP Sprint 1	SP Sprint 2	SP Sprint 3	SP Sprint 4
US-BD01	2	1. XS	1	5	5	5
US-BD02	3	4. L	8	0	0	0
US-BD03	2	1. XS	1	1	1	1
US-BD04	3	4. L	8	0	0	0
US-BD05	2	2. S	3	0	0	0
US-BD06	5	6. XXL	20	20	20	20
US-BD07	5	5. XL	13	13	13	13
US-BD08	2	2. S	2	5	5	5
US-BD09	4	3. M	5	2	2	2
US-BD10	4	3. M	5	2	2	2
US-BD11	2	1. XS	1	1	1	1
US-BD12	3	4. L	8	0	0	0
US-BD13	2	2. S	3	0	0	0
US-BD14	5	3. M	3	3	3	3
US-BD15	5	4. L	8	8	8	8
US-BD16	2	2. S	2	2	2	2
US-BD17	4	3. M	5	5	5	5
US-BD18	4	3. M	5	5	5	5
US-BD19	5	5. XL	0	20	20	20
US-BD20	2	3. M	0	5	5	5
US-BD21	3	3. M	0	5	13	13
US-BD22	2	2. S	0	0	0	2
US-AM01	2	1. XS	1	1	1	1
US-AM02	2	1. XS	1	1	1	1
US-AM03	3	2. S	2	2	2	3
US-AM04	2	1. XS	1	1	1	1

US-AM05	2	1. XS	1	1	1	1
US-AM06	2	1. XS	1	1	1	1
US-AM07	2	1. XS	1	1	1	1
US-AM08	2	1. XS	1	1	1	1
US-AM09	2	1. XS	1	1	1	1
US-AM10	2	1. XS	1	1	1	1
US-AM11	2	1. XS	1	1	1	1
US-AM12	2	1. XS	1	1	1	1
US-AM13	2	1. XS	1	1	1	1
US-AM14	2	1. XS	1	1	1	1
US-AM15	2	1. XS	1	1	1	1

Tabla 9.6 Estimación product backlog

## 9.11. Anexo 11 - Estimación tareas Product Backlog

### 9.11.1.1. Sprint 1

ID	Tarea ID	Tarea	Estimación Tarea	Esfuerzo Real
US-BD11	1	Creación aplicación Twitter	0.5	0.25
	2	Obtención <i>token</i>	0.5	0.25
	3	Crear paquete de trabajo	0.5	0.25
	4	Conexión con Twitter	0.5	0.25
	5	Extracción de datos de Twitter	0.5	1
	6	DQ: Filtrado de datos en base a locación	1	2
	7	Guardar fecha tweet más nuevo	1	3
	8	Filtrar tweets ya procesados	1	3
	9	Verificar el cumplimiento con criterios de validación	1	0.5
US-BD14	1	DQ: Filtrar palabras inválidas	2	3
	2	Cargar diccionario de palabras	4	3
	3	Taggear palabras según diccionario	2	3
	4	Calcular valoración media del tweet	1	1
	5	Verificar el cumplimiento con criterios de validación	1	0.5
US-BD15	1	DQ: Filtrar palabras inválidas	1	2
	2	Identificar tópicos	4	8
	3	Contar cantidad menciones por tópico	1	2
	4	Obtener fecha de proceso	1	2
	5	Verificar el cumplimiento con criterios de validación	1	0.5
US-BD16	1	Consolidar y totalizar datos del proceso	2	3
	2	Guardar datos analizados	1	3
US-BD17	1	Agregar fuente datos semana actual	1	2
	2	Agregar gráfica semana actual	1	2

	3	Agregar tabla datos semana actual	1	1
	4	Agregar fuente datos histórico	2	1
	5	Agregar gráfica histórico	2	1
	6	Agregar tabla datos histórico	1	1
	7	Formatear reporte	1	2
	8	Verificar el cumplimiento con criterios de validación	1	0.5
US-BD18	1	Agregar fuente datos	2	0.5
	2	Agregar gráfica	1	0.5
	3	Agregar tabla datos	1	0.5
	4	Formatear reporte	2	0.5
	5	Verificar el cumplimiento con criterios de validación	1	0.5

Tabla 9.7 Estimación sprint 1

#### 9.11.1.2. Sprint 2

ID	Tarea ID	Tarea	Estimación Tarea	Esfuerzo Real
US-BD01	1	Crear paquete de trabajo	0.25	0.25
	2	Conexión con Twitter	0.25	0.25
	3	Cargar lista de productos y servicios	2	1.5
	4	Generar <i>string</i> consulta Twitter	2	4
	5	Extracción de datos de Twitter	0.25	0.25
	6	DQ: Filtrado de datos en base a locación	0.25	0.25
	7	Guardar fecha tweet más nuevo	0.25	0.25
	8	Filtrar tweets ya procesados	0.25	0.25
	9	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-BD03	1	Crear paquete de trabajo	0.25	0.25
	2	Conexión con Twitter	0.25	0.25
	3	Extracción de datos de Twitter	0.25	0.25

	4	DQ: Filtrado de datos en base a locación	0.25	0.25
	5	Guardar fecha tweet más nuevo	0.25	0.25
	6	Filtrar tweets ya procesados	0.25	0.25
	7	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-BD06	1	DQ: Filtrar palabras inválidas	0.5	0.25
	2	Filtrar tweets por categoría	4	3
	3	Generar tópicos	0.5	1
	4	Totalizar menciones por producto y servicios	2	1
	5	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-BD07	1	DQ: Filtrar palabras inválidas	0.25	0.25
	2	Filtrar tweets por categoría	4	3
	3	Taggear palabras según lista de productos y servicios	0.5	3
	4	Totalizar menciones por producto y servicios	2	3
	5	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-BD08	1	Consolidar y totalizar datos del proceso	3	1.5
	2	Guardar datos analizados	3	5
US-BD09	1	Agregar fuente datos	0.5	0.25
	2	Agregar gráfica	0.5	1.5
	3	Agregar tabla datos	0.5	0.25
	4	Formatear reporte	0.5	0.5
	5	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-BD10	1	Agregar fuente datos	0.5	0.25
	2	Agregar gráfica	0.5	0.25
	3	Agregar tabla datos	0.5	0.25
	4	Formatear reporte	0.5	0.5

	5	Verificar el cumplimiento con criterios de validación	0.5	0.5
--	---	---	-----	-----

Tabla 9.8 Estimación sprint 2

### 9.11.1.3. Sprint 3

ID	Tarea ID	Tarea	Estimación Tarea	Esfuerzo Real
US-BD19	1	Crear paquete de trabajo	0.25	0.25
	2	Desarrollar conexión con Twitter <i>Streaming</i> API	7	4
	3	Desarrollar componente recepción tweets	12	22
	4	Desarrollar componente consulta datos usuario	4	5
	5	Generar flujo de recepción tweets	1	3
	6	DQ: Filtrado de datos en base a locación	0.25	0.25
	7	Guardar tweets en la base de datos local	0.25	1
	8	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-BD20	1	Obtener tweets no procesados	1	2
	2	Valorar sentimiento de tweets no procesados	0.25	0.25
	3	DQ: Filtrar palabras inválidas	0.25	0.25
	4	Levantar estadísticas de sentimientos	1	0.5
	5	Recalcular sentimiento	4	5
	6	Guardar datos analizados	1	3
	7	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-BD21	1	Investigar herramienta de visualización	10	18
	2	Levantar datos de la BD de forma recurrente	2	3
	3	Generar gráfica de sentimiento	2	3
	4	Mecanismo de actualización recurrente de la gráfica	1	0

	5	Visualizar datos de los tweets y usuarios por sentimiento	4	6
	6	Verificar el cumplimiento con criterios de validación	0.5	0.5

Tabla 9.9 Estimación sprint 3

#### 9.11.1.4. Sprint 4

ID	Tarea ID	Tarea	Estimación Tarea	Esfuerzo Real
US-AM01	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM02	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM03	1	Generar empaquetado para Android	10	8
	2	Editar HTML página principal para detectar desde que sistema operativo se está ejecutando	2	3
	3	Editar .css para excluir o incluir video en página principal en base al sistema operativo	1	1
	4	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM04	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM05	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM06	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM07	1	Generar empaquetado para Android	0	0

	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM08	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM09	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM10	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM11	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM12	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM13	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM14	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-AM15	1	Generar empaquetado para Android	0	0
	2	Verificar el cumplimiento con criterios de validación	0.5	0.5
US-BD22	1	Descartar datos duplicados en el proceso de extracción	3	2
	2	Excluir los hashtags de extracción en el proceso de análisis de tendencias	1	2
	3	Verificar calidad de los procesos	6	8
	4	Verificar el cumplimiento con criterios de validación	0.5	0.5

Tabla 9.10 Estimación sprint 4

## 9.12. Anexo 12 - Sprint Review

### 9.12.1.1. Sprint 1

Resultado de la revisión de la iteración con el tutor del proyecto:

- + *Product Backlog*
  - ✓ Se comentó que el *product backlog* se encontraba correcto en líneas generales.
  - ✓ Se sugirió redactar la frase “análisis de sentimiento” de otra forma que clarifique el significado dado que en castellano no es un término utilizado comúnmente
- + Alcance
  - ✓ Se solicitó un cambio de alcance relacionado con la obtención, análisis y visualización de los datos en tiempo real
- + Gestión de Riesgos
  - ✓ Se sugirió agregar un riesgo relacionado con la falta de disponibilidad de los integrantes del equipo de proyecto

### 9.12.1.2. Sprint 2

Resultado de la revisión de la iteración con el tutor del proyecto:

- + Calidad
  - ✓ Se enfatizó en la importancia del manejo de calidad de datos lo cual permitirá obtener resultados más certeros en la ejecución de los análisis.
- + Gestión
  - ✓ Se estableció que una vez materializado un riesgo pasa a tener valor cero, dado que se tomó alguna medida para aceptarlo o contener el impacto.
- + Documentación
  - ✓ Se sugirió realizar una breve explicación, a modo de marco teórico, sobre los temas principales abordados en el proyecto, tales como: “*Sentiment Analysis*” y “*Big Data*”.
- + Producto
  - ✓ Se revisaron los reportes generados en base a los procesos desarrollados en el *sprint*; en especial el *tag cloud* incluido en los mismos, lo cual mejoró la usabilidad de los reportes.

### 9.12.1.3. Sprint 3

Resultado de la revisión de la iteración con el tutor del proyecto:

- + Producto
  - ✓ En el análisis en tiempo real sería recomendable poder visualizar o no el histórico. Esto permite enfocarse en el resultado de una campaña particular, sin que los datos históricos interfieran sobre los actuales.

- ✓ Se debería contemplar la opción de analizar otros mercados que influyeran en el mercado uruguayo.

#### + Calidad

- ✓ Se sugirió realizar alguna prueba estadística que permita validar los datos utilizados.

#### 9.12.1.4. Sprint 4

Resultado de la revisión de la iteración con el tutor del proyecto:

#### + Aplicación móvil

- ✓ Algunos iconos parecen fuera de lugar cuando se visualiza el sitio web desde un dispositivo móvil. De haber tiempo se podría mejorar un poco el diseño.
- ✓ Está bien que no se visualice el video de la pantalla inicial cuando se carga el sitio desde un dispositivo móvil.

#### + Gestión

- ✓ No parecería ser necesario generar una métrica de satisfacción de usuario dado que se basaría en un solo usuario.
- ✓ Si se sugiere generar una encuesta de pocas preguntas para entender la opinión del usuario.

## 9.13. Anexo 13 - Retrospectiva

### 9.13.1.1. Sprint 1

Resultado de la retrospectiva de la iteración entre los integrantes del equipo:

Puntos que salieron bien	
Criterio	Evaluación
Planificación	La estimación de tiempo para la creación del <i>product backlog</i> y los planes de gestión de riesgos, comunicaciones, y estrategia de estimaciones fue adecuada.
Comunicación	Se logró una muy buena dinámica de comunicación en el equipo, utilizando varias vías o herramientas para esto. Se generaron reuniones de estado, tanto virtuales como presenciales.
Desarrollo	Se cumplió con el alcance establecido para el primer <i>sprint</i> . La curva de aprendizaje de la herramienta utilizada estuvo dentro de los tiempos estimados para el mismo, lo que permitió avanzar con el desarrollo de forma eficaz según lo planificado.
Puntos a mejorar	
Criterio	Evaluación
Planificación	Se estimaron las horas de las tareas del <i>Sprint</i> sin contemplar la cantidad de personas que estarían trabajando en ellas, se tomó como que una sola persona estaría trabajando en ellas. Se detectó que algunas tareas de desarrollo se pueden dividir entre los miembros del equipo y no es necesario que los tres estén trabajando sobre lo mismo.
Comunicación	No hay comentarios.
Desarrollo	Si bien se obtuvo un aprendizaje aceptable de la herramienta, la profundización y entendimiento de sus componentes ayudaría a mejorar la productividad en la construcción de los procesos. Se puede mejorar la calidad de los entregables ahondando en la investigación de las buenas prácticas relacionadas con el análisis masivo de datos.

Tabla 9.11 Retrospectiva sprint 1

### 9.13.1.2. Sprint 2

Resultado de la retrospectiva de la iteración entre los integrantes del equipo:

Puntos que salieron bien	
Criterio	Evaluación
Planificación	Se redujo la desviación entre la estimación de las horas por tarea y las horas reales del <i>Sprint</i> . Se separaron tareas entre los integrantes de equipo para trabajar en paralelo, mayormente en temas de investigación.
Comunicación	Se incrementó la cantidad de reuniones presenciales lo que mejoró la relación en el grupo y la dinámica de trabajo. El resto se desarrolló de la misma forma al <i>sprint</i> anterior.
Desarrollo	Se cumplió con el alcance establecido para el segundo <i>sprint</i> . En base al conocimiento adquirido de la herramienta y la tecnología, se logró ser más eficiente en el diseño y construcción de los procesos definidos para el <i>sprint</i> , esto se refleja en que para el segundo <i>sprint</i> , se desarrollaron <i>User Stories</i> estimadas en más <i>Story Points</i> que para el primer <i>sprint</i> , pero la cantidad de horas requeridas para las mismas fue menor.
Puntos a mejorar	
Criterio	Evaluación
Planificación	Si bien se separaron tareas entre los miembros del equipo, la separación de las mismas no fue del todo balanceada.
Comunicación	No hay comentarios.
Desarrollo	Se puede refinar la estimación de las tareas del <i>sprint</i> tomando en cuenta a la hora de estimar, las horas de investigación requeridas previas a la realización de la tarea.

Tabla 9.12 Retrospectiva sprint 2

### 9.13.1.3. Sprint 3

Resultado de la retrospectiva de la iteración entre los integrantes del equipo:

Puntos que salieron bien	
Criterio	Evaluación
Planificación	Se tomaron acciones efectivas que permitieron reducir la magnitud de los riesgos considerablemente, llevando la mayoría de éstos a valores que permiten tomar una estrategia de aceptación de los mismos.
Comunicación	La comunicación interna al equipo se mantuvo al mismo nivel que en los <i>sprints</i> anteriores, siendo clara y tomando acciones en conjunto entre los integrantes del equipo.
Desarrollo	Se logró cumplir con el alcance del tercer <i>sprint</i> si bien los requerimientos incluidos en el mismo requirieron gran cantidad de horas de investigación y desarrollo de componentes para los que no fue posible basarnos en el aprendizaje adquirido en <i>sprints</i> anteriores.
Puntos a mejorar	
Criterio	Evaluación
Planificación	Se concluyó el <i>sprint</i> con la mayor variación de esfuerzo al momento. Esto se debió al foco del <i>sprint</i> en los requerimientos de análisis en tiempo real y a la inclusión de las tareas de investigación. Se detectó que las tareas de investigación fueron subestimadas.
Comunicación	No hay comentarios.
Desarrollo	Dado que se incluyó una nueva herramienta para realizar la visualización en tiempo real de la información, se considera que se puede mejorar la calidad y usabilidad del reporte en tiempo real generado si se profundiza en el aprendizaje y entendimiento de sus componentes.

Tabla 9.13 Retrospectiva sprint 3

#### 9.13.1.4. Sprint 4

Resultado de la retrospectiva de la iteración entre los integrantes del equipo:

Puntos que salieron bien	
Criterio	Evaluación
Planificación	En este último <i>Sprint</i> la cantidad de horas estimadas fue considerablemente menor, lo cual permitió cumplir con el alcance definido sin inconvenientes. La desviación entre la cantidad de horas estimadas y la cantidad de horas reales fue la menor en comparación con los otros <i>sprints</i> .
Comunicación	En base a los antecedentes positivos en cuanto a la comunicación, se continuó con la misma dinámica con la que se venía trabajando.
Desarrollo	Se logró cumplir con el alcance del cuarto <i>sprint</i> a pesar de la naturaleza de las actividades definidas, ya que no fue posible basarnos en el aprendizaje adquirido en <i>sprints</i> anteriores.
Puntos a mejorar	
Criterio	Evaluación
Planificación	Se notó la disparidad en la cantidad de horas invertidas en los diferentes <i>sprints</i> .
Comunicación	No hay comentarios.
Desarrollo	Se detectó que se puede mejorar la mantenibilidad cambiando la estrategia de visualización en la página HTML principal.

Tabla 9.14 Retrospectiva sprint 4

## 9.14. Anexo 14 - Estándares de desarrollo

### 9.14.1.1. Guía de utilización de componentes KNIME

Se detallan a continuación algunos lineamientos a seguir para la construcción de flujos en KNIME:

- ✚ Los nodos seleccionados para el flujo deben tener una breve descripción de no más de 3 o 4 palabras, que dé una idea general sobre la función que cumple dicho nodo.
- ✚ Los nodos que cumplan un propósito en común, deben ser agrupados utilizando la opción de generación de “metanodos” (*Collapse into Metanode*).
- ✚ Los “metanodos” deben ser realizados de manera tal que puedan ser reutilizados en otros nodos
- ✚ Los conectores deben seguir una secuencia lógica en la medida de lo posible, en concordancia con el proceso que se está construyendo.
- ✚ Las uniones de cada nodo no deben solaparse o solaparse lo menos posible con otros nodos, y deben permitir la lectura de las descripciones de cada uno de ellos.
- ✚ Se debe evitar, en la medida de lo posible, configurar nodos con valores “hardcodeados”.
  - ✓ Las conexiones a la base de datos se deben hacer desde un archivo de configuración.
  - ✓ Los datos de conexión a las fuentes de datos (credenciales Twitter) deben ser definidos en un archivo de configuración

### 9.14.1.2. Nomenclatura para el modelo físico de la base de datos

La nominación de las tablas se definió de la siguiente manera:

- ✚ Los nombres deben ser descriptivos y en plural.
- ✚ Si hay espacio entre palabras deberán ser separadas por un guión bajo “\_”
- ✚ Las tablas que contengan datos que no sean modificadas por los flujos, deben ser identificadas con el prefijo “p\_”: paramétricas.
- ✚ Las tablas que contengan datos que son modificados por los flujos, deben ser identificadas con el prefijo “o\_”: operacionales.

La nominación de los campos se definió de la siguiente manera:

- ✚ Los campos deben ser definidos en minúsculas
- ✚ Los nombres deben describir el dato que van a almacenar.
- ✚ Si se usan abreviaturas, deberán ser mnemotécnicas
- ✚ Las claves primarias o foráneas deberán identificarse con el sufijo “id\_”

### 9.14.1.3. Formato de reportes

Todos los reportes deben seguir el siguiente formato:

- ✚ Cabezal con logo de la empresa del lado izquierdo y nombre del reporte del lado derecho
- ✚ Las secciones del reporte deben estar diferenciadas por un separador, y un subtítulo de la sección
- ✚ Debe indicar la fecha de ejecución del reporte

Reporte ejemplo (Imagen 9.31):



**Título reporte**

---

**Título subsección**

---

**Gráfico**

*Imagen 9.31 Template reporte*

## 9.15. Anexo 15 - Revisiones y mediciones

### 9.15.1.1. Sprint 1

Tipo de revisión	Elementos revisados	Comentarios / Hallazgos
Revisión técnica	Extracción_Twitter_Col ectate	- Error al insertar los datos en la base cuando los tweets contienen caracteres fuera del rango ASCII
	Análisis_Sentimiento	- Problema con el manejo de la negación: se implementó mecanismo para tomar en cuenta al menos uno de los casos de negación - El diccionario de palabras no es adecuado. No contempla todas las palabras del idioma español.
	Reporte_Analisis_Sentimiento	- La fecha de ejecución del análisis incluye una letra "T" entre la fecha y la hora.
	Análisis_Oportunidades_Mejora	- La cantidad de tópicos detectados no es representativo de la realidad; la cantidad de tópicos configurado no es la adecuada para la detección de categorías en base a un volumen de datos dinámico.
	Reporte Oportunidades_Mejora	- A nivel visual, los tópicos no clarifican la información obtenida. Se analiza incluir algún elemento adicional que permita visualizar de mejor manera las oportunidades de mejora detectadas
Inspección de ajuste al proceso	Extracción_Twitter_Col ectate Análisis_Sentimiento Análisis_Oportunidades_Mejora Reporte_Analisis_Sentimiento Reporte Oportunidades_Mejora	- Falta nombrar algunos nodos - Falta agrupar metanodos en común

Tabla 9.15 Resultado revisión sprint 1

### Revisión de la calidad de los productos - resultados mediciones

Métricas del proceso	Resultado	Cumplimiento
Requisitos completados por iteración	100%	OK
Horas de retrabajo	13.7%	OK
Velocity	54.5	-

Desviación esfuerzo	18.3%	KO
Desviación de avance	-4.3%	OK
Desviación de alcance por <i>Sprint</i>	0.8%	-
<b>Métricas del producto</b>	<b>Resultado</b>	<b>Cumplimiento</b>
Efectividad de las revisiones técnicas	84%	KO
Efectividad de las revisiones del cliente	16%	KO

Tabla 9.16 Resultado medición sprint 1

### 9.15.1.2. Sprint 2

Tipo de revisión	Elementos revisados	Comentarios / Hallazgos
Revisión técnica	Extracción_Twitter	- La estrategia de extracción utilizada desaprovecha gran porcentaje de tweets extraídos
	Análisis_Actualidad	Sin comentarios
	Análisis_Tendencias	- La cantidad de tópicos detectados no es representativo de la realidad; la cantidad de tópicos configurado no es la adecuada para la detección de categorías en base a un volumen de datos dinámico.
	Reporte Análisis_Tendencias	Sin comentarios
	Reporte Análisis_Actualidad	Sin comentarios
Inspección de ajuste al proceso	Extracción_Twitter Análisis_Actualidad Análisis_Tendencias Reporte Análisis_Tendencias Reporte Análisis_Actualidad	- Falta nombrar algunos nodos

Tabla 9.17 Resultado revisión sprint 2

### Revisión de la calidad de los productos - resultados mediciones

Métricas del proceso	Resultado	Cumplimiento
Requisitos completados por iteración	100%	OK
Horas de retrabajo	11%	OK
<i>Velocity</i>	36	-

Desviación esfuerzo	6.9%	KO
Desviación de avance	54.8%	KO
Desviación de alcance por <i>Sprint</i>	6.8%	-
<b>Métricas del producto</b>	<b>Resultado</b>	<b>Cumplimiento</b>
Efectividad de las revisiones técnicas	100%	OK
Efectividad de las revisiones del cliente	0%	OK

Tabla 9.18 Resultado medición sprint 2

### 9.15.1.3. Sprint 3

Tipo de revisión	Elementos revisados	Comentarios / Hallazgos
Revisión técnica	Extraccion_Twitter_Stream	<ul style="list-style-type: none"> <li>- Solapamiento de los procesos de extracción <i>streaming</i> y <i>batch</i></li> <li>- Proceso de <i>streaming</i> no guarda la fecha de extracción recibida</li> </ul>
	Análisis_Sentimiento_Streaming	<p>Análisis incorrecto en los siguientes escenarios:</p> <ul style="list-style-type: none"> <li>- palabras seguidas de puntuaciones sin espacio entre la palabra y la puntuación y donde la puntuación es repetida más de una vez.</li> <li>- palabras negadas donde entre la negación y la palabra negada no hay otras palabras o existe más de una palabra entre ambas.</li> <li>- cuando existen igual número de palabras positivas y negativas, donde el resultado debería ser neutro, se arroja un resultado aleatorio dependiendo del orden en que se detecta el sentimiento de las palabras.</li> </ul>
	Reporte Análisis Sentimiento en tiempo real	- Problemas al visualizar algunos datos en el <i>dashboard</i> que nos hicieron realizar cambios en el proceso de KNIME
Inspección de ajuste al proceso	<ul style="list-style-type: none"> <li>- Extraccion_Twitter_Stream</li> <li>- Análisis_Sentimiento_Streaming</li> <li>- Reporte Análisis Sentimiento en tiempo real</li> </ul>	Sin comentarios

Tabla 9.19 Resultado revisión sprint 3

## Revisión de la calidad de los productos - resultados mediciones

Métricas del proceso	Resultado	Cumplimiento
Requisitos completados por iteración	100%	OK
Horas de retrabajo	15.3%	OK
<i>Velocity</i>	36,6	-
Desviación esfuerzo	32.4%	KO
Desviación de avance	65.2%	KO
Desviación de alcance por <i>Sprint</i>	2.3%	-
Métricas del producto	Resultado	Cumplimiento
Efectividad de las revisiones técnicas	100%	OK
Efectividad de las revisiones del cliente	0%	OK

Tabla 9.20 Resultado medición sprint 3

### 9.15.1.4. Sprint 4

Tipo de revisión	Elementos revisados	Comentarios / Hallazgos
Revisión técnica	App móvil	- Algunos íconos se muestran fuera de lugar
Inspección de ajuste al proceso	No aplica	

Tabla 9.21 Resultado revisión sprint 4

## Revisión de la calidad de los productos - resultados mediciones

Métricas del proceso	Resultado	Cumplimiento
Requisitos completados por iteración	100%	OK
Horas de retrabajo	12.5%	OK
<i>Velocity</i>	32.25%	-
Desviación esfuerzo	3.1%	OK
Desviación de avance	18.75%	KO
Desviación de alcance por <i>Sprint</i>	0%	-
Métricas del producto	Resultado	Cumplimiento
Efectividad de las revisiones técnicas	0%	KO
Efectividad de las revisiones del cliente	100%	KO

Tabla 9.22 Resultado medición sprint 4

## 9.16. Anexo 16 - Resultados - Plan de testing

### 9.16.1. Resultados - Pruebas funcionales

#### 9.16.1.1. Test - Análisis Sentimiento

##### Definición de pruebas

Caso	Sentimiento	Descripción	Cantidad
TC01	Positivo	Texto que utilice únicamente palabras con connotación positiva y ninguna negativa	2
TC02	Positivo	Texto que utilice únicamente una palabra con connotación negativa y dos positivas	2
TC03	Negativo	Texto que utilice únicamente palabras con connotación negativa y ninguna positiva	3
TC04	Negativo	Texto que utilice únicamente una palabra con connotación positiva y dos negativas	2
TC05	Negativo	Texto que utilice únicamente una palabra con connotación positiva pero negada con “no” previamente dándole significado negativo	2
TC06	Neutro	Texto que NO utilice ninguna palabra con connotación positiva y ninguna negativa	2
TC07	Neutro	Texto que utilice únicamente una palabra con connotación positiva y una negativa	2

Tabla 9.23 Casos de prueba análisis sentimiento

##### Datos y Pruebas

Caso	Texto	Sentimiento Esperado	Sentimiento Identificado
TC01	Colectate está re <b>bueno</b> , me encanta!!!	Positivo	Positivo
TC01	Colectate es fácil de usar, es <b>excelente</b> !	Positivo	Positivo
TC02	Colectate funciona <b>horrible</b> pero el diseño es <b>feliz</b> y <b>agradable</b>	Positivo	Positivo
TC02	Es <b>lindo</b> y <b>emotivo</b> que Colectate ayude a animales pero el servicio es <b>malo</b> ...	Positivo	Positivo
TC03	Colectate es <b>horrible</b> !	Negativo	Negativo
TC03	El servicio de Colectate es <b>malo</b> !!!	Negativo	Neutro

TC03	El servicio de Colectate es <b>malo</b> !!!	Negativo	Negativo
TC04	El sitio de Colectate me <b>encanta</b> , pero funciona <b>horrible</b> , anda <b>mal</b>	Negativo	Negativo
TC04	La <b>generosidad</b> de Colectate es inspiradora, pero el servicio al cliente es <b>malo</b> y me dejó <b>descontento</b>	Negativo	Negativo
TC05	Colectate no es divertido	Negativo	Negativo
TC05	El resultado de la colecta <b>no fue aceptable</b>	Negativo	Negativo
TC06	Colectate ofrece un servicio de colectas online	Neutro	Neutro
TC06	Colaborará con las colectas solidarias mediante Colectate!	Neutro	Neutro
TC07	Colectate funciona <b>horrible</b> pero el diseño es <b>agradable</b>	Neutro	Negativo
TC07	La <b>generosidad</b> de Colectate es inspiradora, pero el servicio al cliente es <b>malo</b>	Neutro	Positivo

Tabla 9.24 Datos y pruebas análisis sentimiento

## Resultado

Total Casos	Total OK	Fórmula	% Correctitud
TC = 15	TO = 12	$(TO / TC) * 100$	80%

Tabla 9.25 Resultado análisis sentimiento

Se detectó que el proceso no detecta correctamente el sentimiento en los siguientes casos:

- ✚ Palabras que son seguidas de puntuaciones sin espacio entre la palabra y la puntuación y donde la puntuación es repetida más de una vez.
- ✚ Palabras negadas donde entre la negación y la palabra negada no hay otras palabras o existe más de una palabra entre ambas.
- ✚ Cuando existen igual número de palabras positivas y negativas, donde el resultado debería ser neutro, se arroja un resultado aleatorio dependiendo del orden en que se detecta el sentimiento de las palabras.

### 9.16.1.2. Test - Oportunidades de mejora (Generador de tópicos)

#### Definición de pruebas

Caso	Tópico / Sentimiento	Descripción	Cantidad
TC01	"Servicio" / positivo	Texto que mencione "servicio" de forma positiva	4

TC02	“Web” / negativo	Texto que mencione “web” de forma negativa	4
TC03	“App” / positivo	Texto que mencione “app” de forma positiva	4

Tabla 9.26 Casos de prueba oportunidades de mejora (tópicos)

## Datos y Pruebas



Caso	Texto	Tópico esperado
TC01	El servicio es muy bueno, son muy atentos.	“Servicio” Positivo
TC01	El servicio es excelente!	
TC01	Colectate funciona bien, el servicio es adecuado.	
TC01	El servicio de Colectate es adorable. Tienen mucho afán.	
TC02	El sitio web es horroroso, no funciona correctamente.	“Web” Negativo
TC02	La web es terrible! El último cambio que hicieron es pésimo.	
TC02	Me causa rechazo la página web de la empresa.	
TC02	El diseño de la web es pobre. Da pena.	
TC03	La app es extasiante, un encanto.	“App” Positivo
TC03	La app es excelente!	
TC03	La app funciona bien.	
TC03	La app de Colectate es adorable. Tienen mucho afán.	

Tabla 9.27 Datos y pruebas oportunidades de mejora (tópicos)

## Resultados

### 1er corrida

Configuración del nodo

-  Nro tópicos: 10
-  Cantidad palabras por tópico: 3

Tópico esperado	Identificado (Si/No)
“Servicio” Positivo	No
“Web” Negativo	Sí
“App” Positivo	No

Tabla 9.28 Tópicos esperados prueba oportunidades de mejora (tópicos) corrida 1



Total Casos	Total OK	Fórmula	% Correctitud
TC = 3	TO = 1	$(TO / TC) * 100$	33%

Tabla 9.29 Resultados prueba oportunidades de mejora (tópicos) corrida 1

Comentarios: Los tópicos “Servicio” y “App” esperados fueron identificados como 1 sólo tópico.

## 2da corrida

Configuración del nodo

-  Nro tópicos: 2
-  Cantidad palabras por tópico: 3

Tópico esperado	Identificado (Si/No)
“Servicio” Positivo	Sí
“Web” Negativo	Sí
“App” Positivo	Sí

Tabla 9.30 Tópicos esperados prueba oportunidades de mejora (tópicos) corrida 2


Total Casos	Total OK	Fórmula	% Correctitud
TC = 3	TO = 3	$(TO / TC) * 100$	100%


Tabla 9.31 Resultados prueba oportunidades de mejora (tópicos) corrida 2

Comentarios: El tópico “Web” esperado fue identificado como 2 tópicos distintos.

## 3da corrida

Configuración del nodo

-  Nro tópicos: 3

 Cantidad palabras por tópico: 3

Tópico esperado	Identificado (Si/No)
“Servicio” Positivo	Sí
“Web” Positivo	Sí
“App” Positivo	Sí

Tabla 9.32 Tópicos esperados prueba oportunidades de mejora (tópicos) corrida 3

Total Casos	Total OK	Fórmula	% Correctitud
TC = 3	TO = 3	$(TO / TC) * 100$	100%

Tabla 9.33 Resultados prueba oportunidades de mejora (tópicos) corrida 3

## Conclusiones

El algoritmo requiere tener un conocimiento previo o una idea aproximada de cuántos tópicos se encuentran en la totalidad de los datos extraídos; esto determinará la configuración del nodo.

El número de tópicos 10 es el valor que se definió para el nodo. Si bien el porcentaje para este caso de prueba es bajo y no cumple con el objetivo de calidad definido, se tomará este valor estándar ya que se desconoce de antemano los tópicos a los que van a referir los tweets que se extraerán.

### 9.16.1.3. Test - Oportunidades de mejora (Tag Cloud)

#### Datos y Pruebas

Se toman en cuenta sólo las palabras que tienen más de 1 mención.

Palabras	Cantidad de menciones esperadas	Cantidad de menciones obtenidas
Web	4	4
Servicio	4	4
App	4	4
Funciona	3	3
Colectate	3	3
Bien	2	2

Afán	2	2
mucho	2	2
adorable	2	2
excelente	2	2

Tabla 9.34 Datos y pruebas oportunidades de mejora (tag cloud)

## Resultados

Total Casos	Total OK	Fórmula	% Correctitud
TC = 10	TO = 10	$(TO / TC) * 100$	100%

Tabla 9.35 Resultados prueba oportunidades de mejora (tag cloud)

### 9.16.1.4. Test - Análisis Actualidad

#### Definición de pruebas

Caso	Producto / Servicio	Descripción	Cantidad
TC01	Ferrugem	Texto que mencione "Ferrugem" de forma positiva	2
TC02	Woow	Texto que mencione "web" de forma negativa	2
TC03	Champagne	Texto que mencione "Champagne" de forma positiva	2
TC04	Lavarropas	Texto que mencione "Lavarropas" de forma positiva	2

Tabla 9.36 Casos de prueba análisis actualidad

#### Datos y Pruebas

Caso	Texto	Cantidad esperada	Cantidad obtenida
TC01	Hice una colecta y me fui a Ferrugem! Fue increíble!	2	2
TC01	Mi luna de miel en Ferrugem, qué emoción!		
TC02	Compré un voucher en Woow para un regalo, fue tedioso y atroz.	2	2
TC02	Horrible la atención en Woow. Un desperdicio de plata.		

TC03	Colaboren con mi colectivo en Champagne!		
TC03	A celebrar mi boda! (Colectivo en Champagne - <i>click</i> en el siguiente <i>link</i> )	2	2
TC04	Ayudame a comprar mi lavarropas a través de esta Colecta!		
TC04	Miren mi lavarropas nuevo! Estoy feliz. Gracias!	2	2

Tabla 9.37 Datos y prueba análisis actualidad

## Resultado

Total Casos	Total OK	Fórmula	% Correctitud
TC = 4	TO = 4	$(TO / TC) * 100$	100

Tabla 9.38 Resultados prueba análisis actualidad

### 9.16.1.5. Test - Filtrado por ubicación

Total Tweets extraídos	520
Tweets con lenguaje “es”	442
Restantes	78
Tweets con <i>Timezone</i> “Br”, “Bs. As.”	53
Restantes	25
Tweets con locación: Montevideo, Uruguay	19
Restantes	6
Otras locaciones: Yaracuy, Venezuela, Ecuador	6

Tabla 9.39 Datos y pruebas filtrado por ubicación

## Resultado

En la gráfica siguiente que se muestra en la Imagen 9.32 se pueden ver los resultados de las pruebas.

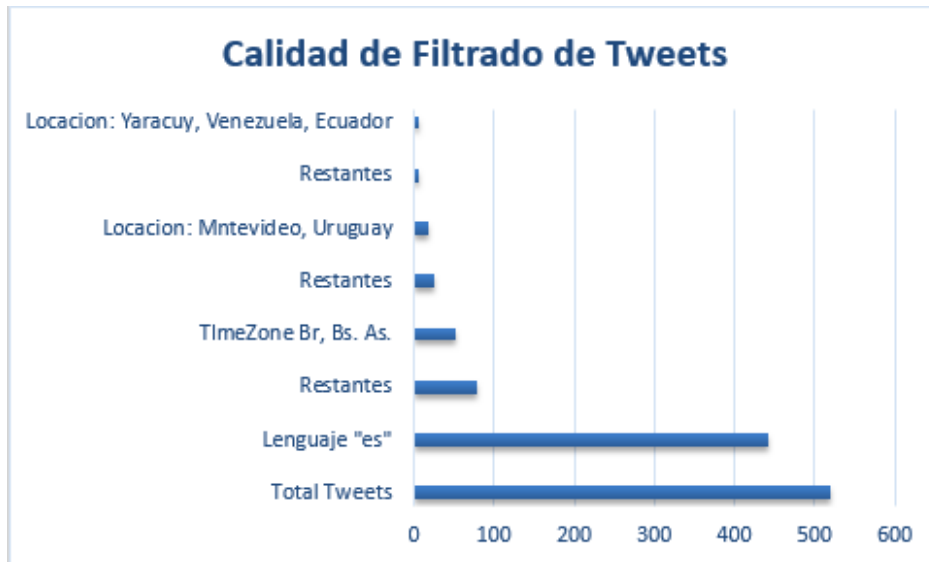


Imagen 9.32 Resultados pruebas filtrado por ubicación

Total Tweets	Total OK	Fórmula	% Correctitud
TT = 520	TO = 514	$(TT / TO) * 100$	99

Tabla 9.40 Resultados prueba filtrado por ubicación

#### 9.16.1.6. Test - Pertinencia de los post extraídos con respecto al tema analizado

Se analiza el contenido de los tweets extraídos uno a uno en base a los *hashtags* definidos y la interpretación de los responsables de realizar el test.

#### Resultado

Total Casos	Total OK	Fórmula	% Correctitud
TT = 520	TO = 226	$(TO / TC) * 100$	43

Tabla 9.41 Resultados prueba pertinencia de post extraídos

### 9.16.2. Resultados - Pruebas de rendimiento

#### Definición de pruebas

Proceso	Cantidad Datos	Tiempo
Análisis Sentimiento	252.000	120 minutos
Oportunidades Mejora	252.000	120 minutos
Analisis Actualidad	252.000	120 minutos
Analisis Tendencia	252.000	120 minutos

Tabla 9.42 Casos de prueba rendimiento

En base a extracciones de prueba de tweets con la configuración actual de filtrado por *hashtags*, se determinó que diariamente se extraen un promedio de 12 mil a 18 mil tweets.

Dada la sugerencia de agendar una o más extracciones diarias y agendar una vez a la semana los procesos de análisis, se define para las pruebas la siguiente cantidad de tweets en un caso normal:  $18.000 \times 7 = 126.000$  tweets

Dado que el requerimiento no funcional de escalabilidad establece que se debería ser capaz de procesar el doble de carga normal sin superar el umbral de tiempo máximo se definen las pruebas de la siguiente forma:  $126.000 \times 2 = 252.000$  tweets

La prueba se realizó en un entorno cuyas especificaciones coinciden con los requerimientos definidos en la sección “Requerimientos de *hardware*”.

### Resultado

Proceso	Cantidad Datos	Tiempo	Resultado
Análisis Sentimiento	270.000	52 minutos	Ok
Oportunidades Mejora	270.000	67 minutos	Ok
Analisis Actualidad	270.000	118 minutos	Ok
Analisis Tendencia	270.000	82 minutos	Ok

Tabla 9.43 Resultado prueba rendimiento

### 9.16.3. Resultados - Pruebas de integridad de datos

Proceso	Nombre nodo / metanodo	Test
Analisis Actualidad	Guardar Análisis en BD	OK
	Guardar cabezal en BD	OK
Analisis Tendencia	Guardar Análisis positivo en BD	OK
	Guardar Análisis negativo en BD	OK
	Guardar cabezal en BD	OK
	Guardar análisis <i>Tag cloud</i> en BD	OK
Análisis sentimiento Análisis Sentimiento <i>Streaming</i>	Guardar sentimiento en Tweet	OK
	Guardar análisis	OK

	Actualizar análisis	OK
Oportunidades Mejora	Guardar Tópicos positivos en BD	OK
	Guardar Tópicos negativos en BD	OK
	Guardar cabezal en BD	OK
Colectate Extracción Twitter	Guardar fecha último Tweet	OK
Colectate Extracción Twitter Colectate	Guardar fecha último Tweet	OK
Colectate Extracción Twitter <i>Streaming</i>	Guardar fecha último Tweet	OK

Tabla 9.44 Casos de prueba integridad de datos

Total Casos	Total OK	Fórmula	% Correctitud
TC = 15	TO = 15	$(TO / TC) * 100$	100%

Tabla 9.45 Resultados prueba integridad de datos

## 9.17. Anexo 17 - Encuesta de Satisfacción

1) ¿Las expectativas generadas al inicio del proyecto luego de planteados los objetivos del mismo, fueron cubiertas?

	Pobre		Regular		Bueno		Muy Bueno	X	Excelente
--	-------	--	---------	--	-------	--	-----------	---	-----------

Al inicio del proyecto se plantearon objetivos basados en optimizar la estrategia de contenidos, venta directa y atención al cliente, visualización de los datos y la aplicación móvil. Consideramos y confirmamos que el cumplimiento de los mismos hay sido exitoso. Contamos al finalizar el proyecto con reportes y visualizaciones que nos ayudan a obtener los beneficios propuestos al inicio del mismo. También la aplicación, que nos dará más oportunidades para la captación de clientes y un mejor servicio a los ya existentes.

2) ¿Cómo considera la calidad de los siguientes reportes entregados? Si es posible, realice un breve comentario.

### Los Reportes de Sentimiento:

	Pobre		Regular		Bueno	X	Muy Bueno		Excelente
--	-------	--	---------	--	-------	---	-----------	--	-----------

### Reporte de Oportunidades de Mejora:

	Pobre		Regular		Bueno	X	Muy Bueno		Excelente
--	-------	--	---------	--	-------	---	-----------	--	-----------

### Reporte de Tendencias:

	Pobre		Regular		Bueno	X	Muy Bueno		Excelente
--	-------	--	---------	--	-------	---	-----------	--	-----------

### Reporte de Análisis Productos y Servicios:

	Pobre		Regular		Bueno	X	Muy Bueno		Excelente
--	-------	--	---------	--	-------	---	-----------	--	-----------

**Comentario:** Los reportes obtenidos son de la calidad esperada. Los reportes mencionados recaban información trascendente para nosotros. Nos ofrecen datos valiosos como: conocer qué dice la gente de ColectaTe, que productos y servicios de nuestro interés se mencionan junto a su valoración, nos permiten una fácil visualización y por ende, nos facilitan la tarea de la toma de decisiones.

3) ¿Cómo considera la calidad de los datos en tiempo real entregados? Si es posible, realice un breve comentario.

#### Datos en tiempo Real

	Pobre		Regular		Bueno		Muy Bueno	X	Excelente
--	-------	--	---------	--	-------	--	-----------	---	-----------

**Comentario:** La vista obtenida en el análisis de sentimiento en tiempo real, es una excelente herramienta para poder realizar acciones de forma instantánea, al momento (o muy poco después) de que vemos la información. Si es positivo, entonces destacarlos, si es negativo, tratar de rápidamente revertir esa opinión. Podremos lograrlo, gracias a esta vista.

4) ¿Cómo considera la calidad de la aplicación móvil? Si es posible, realice un breve comentario.

#### Aplicación Móvil

	Pobre		Regular	X	Bueno		Muy Bueno		Excelente
--	-------	--	---------	---	-------	--	-----------	--	-----------

**Comentario:** La aplicación móvil nos abre oportunidades no alcanzadas hasta el momento. La realidad actual, muestra que las personas cada vez más consumen servicios desde sus Smartphone y dispositivos. Esto significa, que podremos entonces promocionar y difundir el uso de la misma logrando captar más clientes y usuarios en base a un uso muy fácil e intuitivo.

5) ¿Cree que las “Oportunidades de Mejora” planteadas son realmente de utilidad? ¿Le gustaría implementar alguna de ellas en el corto plazo? ¿Cuál?

	Pobre		Regular		Bueno		Muy Bueno	X	Excelente
--	-------	--	---------	--	-------	--	-----------	---	-----------

Las oportunidades de mejora planteadas por el equipo de proyecto, son de máximo interés por ColectaTe ya que nos permitirán mejorar la calidad de nuestro servicio y entender cada vez más a nuestros clientes y usuarios. Particularmente nos gustaría implementar en el corto plazo la “Mejora en el análisis de sentimiento / Análisis a través de aprendizaje”.

#### Resultados:

Preguntas	Total Pobre	Total Regular	Total Bueno	Total Muy Bueno	Total Excelente	Objetivo Calidad
Generales	-	-	-	-	2	OK
<i>Big Data</i>	-	-	-	4	1	OK
App móvil	-	-	1	-	-	OK

Tabla 9.46 Resultados encuesta de satisfacción

## 9.18. Anexo 18 - Descripción de metanodos KNIME

<b>Nombre del metanodo</b>	Cant Extraídos
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad, Colectate_Análisis_Sentimiento, Colectate_Análisis_Sentimiento_Streaming, Colectate_Análisis_Tendencias, Colectate_Oportunidades_Mejora

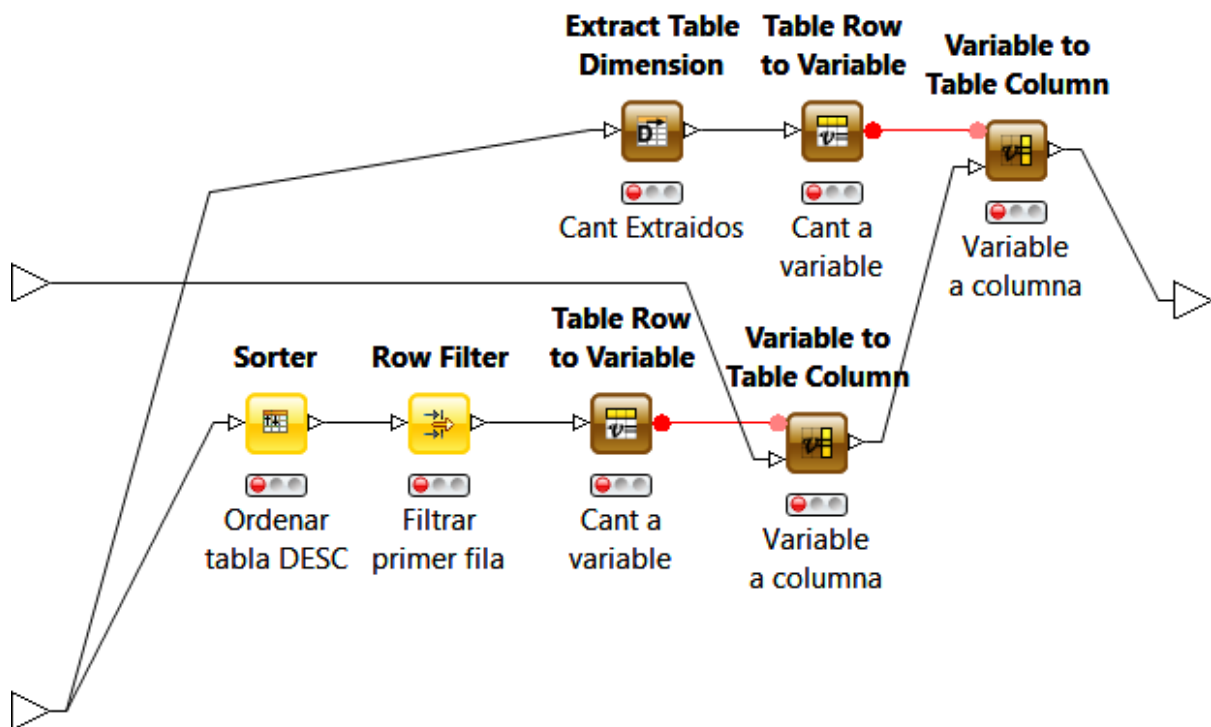


Imagen 9.33 Metanodo Cant Extraídos

En este metanodo, se reciben los tweets obtenidos desde la BD y la fecha de ejecución del proceso, con la finalidad de obtener la cantidad de tweets extraídos para su valoración.

<b>Nombre del metanodo</b>	Cant Valorados
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad, Colectate_Análisis_Sentimiento, Colectate_Análisis_Sentimiento_Streaming, Colectate_Análisis_Tendencias, Colectate_Oportunidades_Mejora

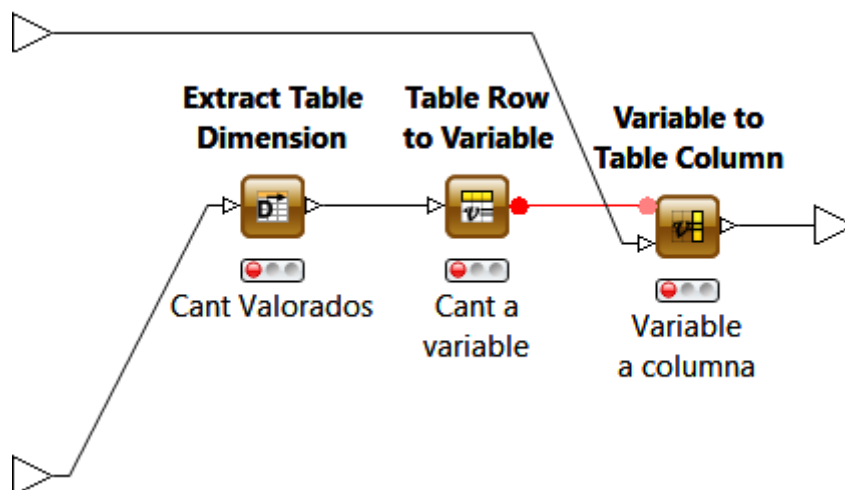


Imagen 9.34 Metanodo Cant Valorados

El metanodo llamado Cant Valorados, nos permite cuantificar los datos ya valorados y procesados. Es utilizado en distintos procesos y también con distintas finalidades. En ocasiones se utiliza para contar la cantidad de valorados en “total”, pero también es utilizado para contar la cantidad de “Valorados negativos”, “Valorados positivos” y “Valorados neutros”.

<b>Nombre del metanodo</b>	Conexión MySQL
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad, Colectate_Análisis_Sentimiento, Colectate_Análisis_Sentimiento_Streaming, Colectate_Análisis_Tendencias, Colectate_Oportunidades_Mejora, Colectate_Extracción_Twitter, Colectate_Extracción_Twitter_Colectate, Colectate_Extracción_Twitter_Stream, Colectate_Reporte_Análisis_Actualidad, Colectate_Reporte_Análisis_Sentimiento, Colectate_Reporte_Análisis_Tendencias, Colectate_Reporte_Oportunidades_Mejora

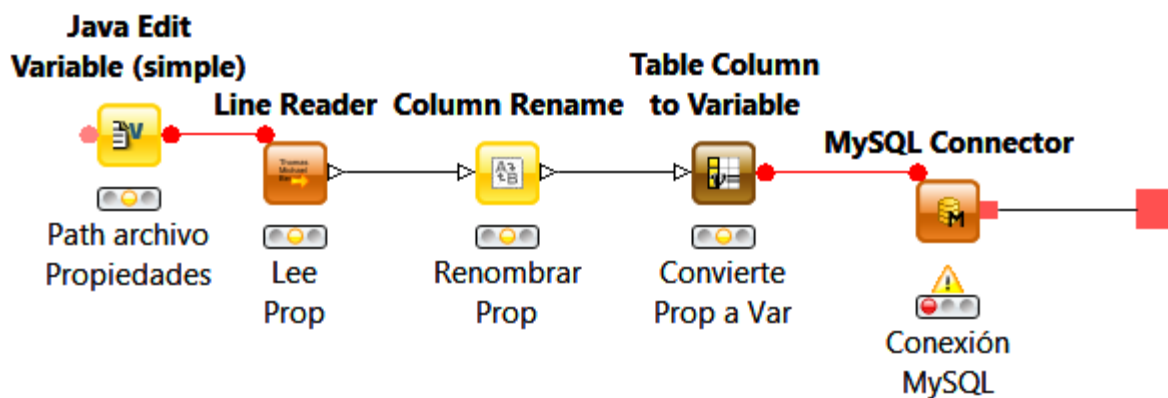


Imagen 9.35 Metanodo Conexión MySQL

En este metanodo se realiza la conexión con la base de datos. En nuestro caso, MySQL. La conexión se realiza utilizando un archivo “Propiedades” que contiene el nombre del servidor, el nombre de la base de datos, el usuario y la contraseña para poder realizar la conexión. De esta manera, es más eficiente la mantenibilidad ya que ante cualquier cambio se modifica dicho archivo y no es necesario modificar ninguno de los procesos en KNIME.

<b>Nombre del metanodo</b>	Extracción Twitter y filtrado
<b>Flujos en lo que se encuentra presente</b>	Colectate_Extracción_Twitter, Colectate_Extracción_Twitter_Colectate,

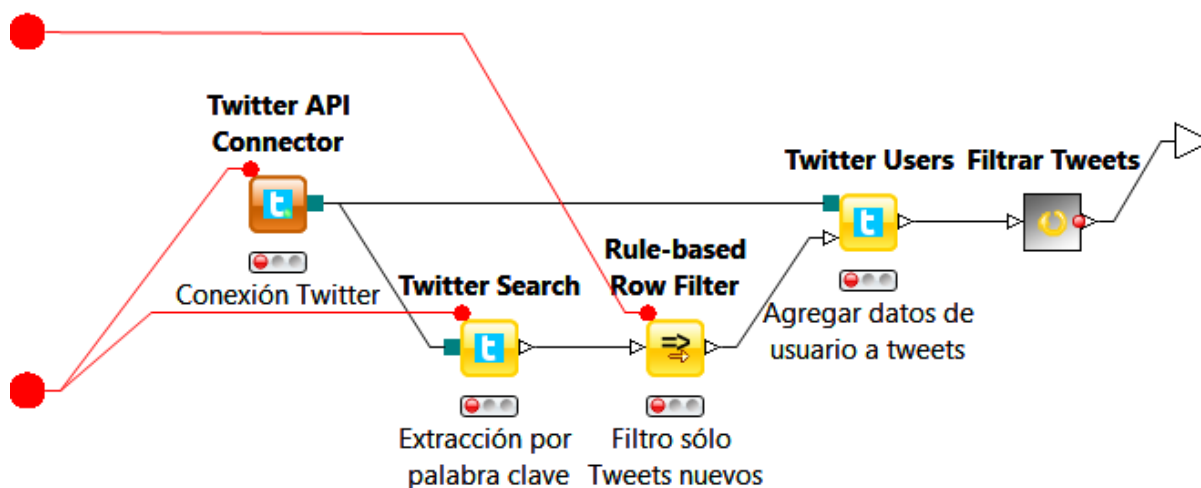


Imagen 9.36 Metanodo Extracción Twitter y filtrado

Este metanodo contiene los componentes necesarios para realizar la conexión y extracción de tweets a través de la REST API de Twitter. Esta API provee una lista de Tweets que coincidan con una consulta específica, y posee ciertas restricciones y requerimientos los cuales están detalladas en el sitio de Twitter [11].

<b>Nombre del metanodo</b>	Fecha Ejecución Proceso
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad, Colectate_Análisis_Sentimiento, Colectate_Análisis_Sentimiento_Streaming, Colectate_Análisis_Tendencias, Colectate_Oportunidades_Mejora,

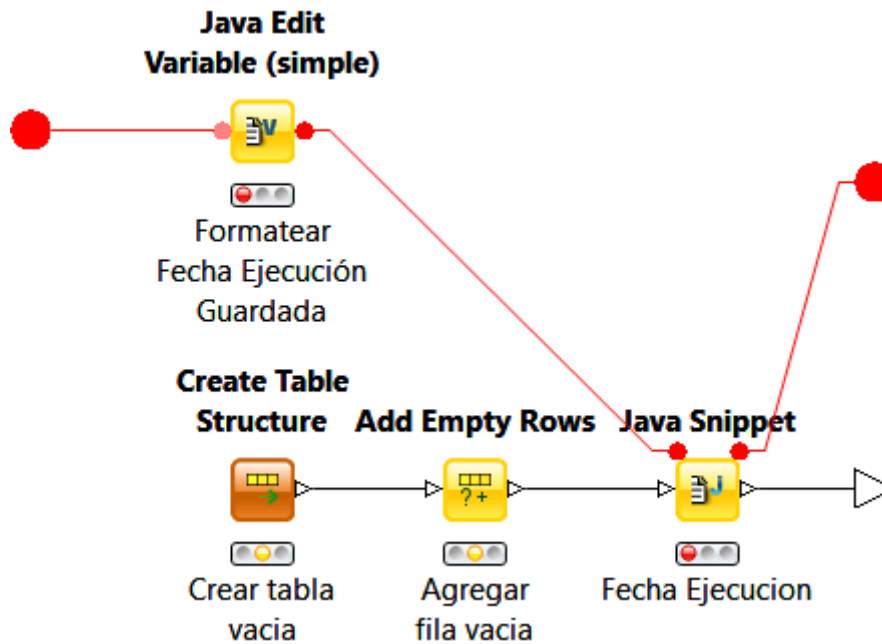


Imagen 9.37 Metanodo Fecha Ejecución Proceso

El objetivo del metanodo mostrado anteriormente es: recibiendo la fecha de ejecución del proceso decidir si corresponde realizar una actualización de los datos (si la fecha obtenida es igual a la ya guardada) y esto significa sumar los datos nuevos a los guardados. O registrar un nuevo ingreso en la base (si las fechas no son iguales) esto por otra parte, significa crear una nueva fila de datos en la tabla. Los datos a los que se hace referencia son: cantidad de extraídos, cantidad total de valorados, cantidad de valorados negativos, cantidad de valorados positivos y cantidad de valorados neutros.

<b>Nombre del metanodo</b>	Filtrar <i>Hashtags</i>
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Tendencias

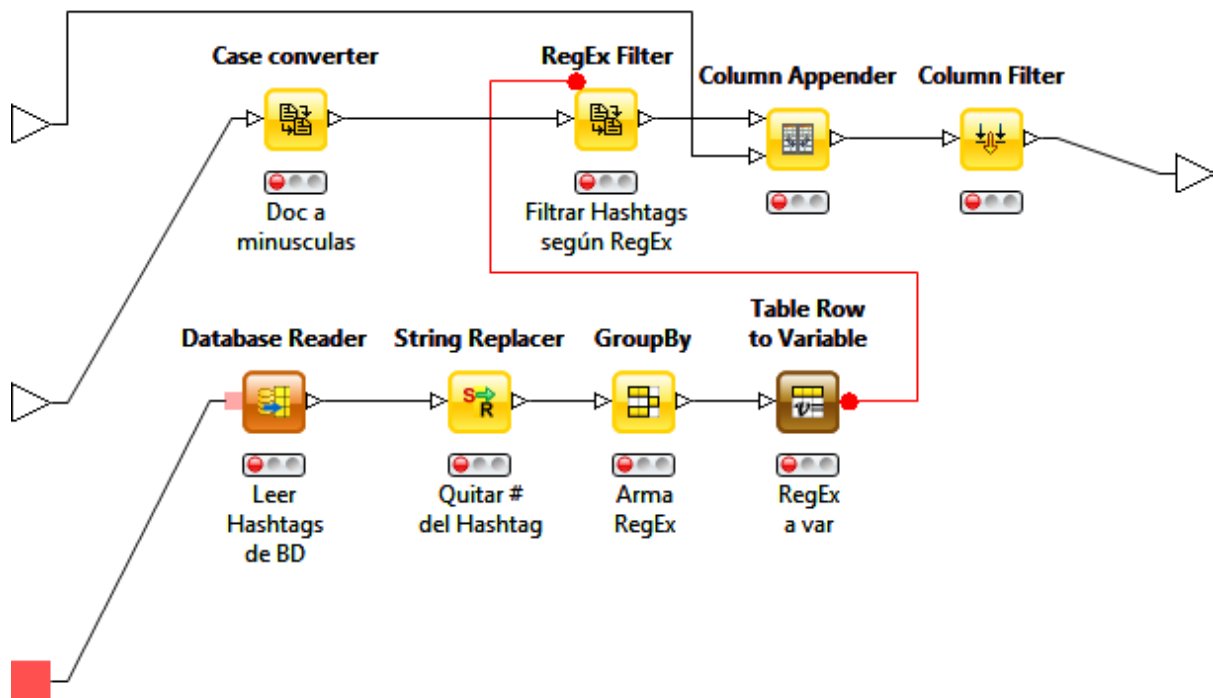


Imagen 9.38 Metanodo Filtrar Hashtags

En este metanodo, se realiza el filtrado completo del *hashtags*. Por un lado, se quita el símbolo #, pero también se quita del proceso, luego de procesar los datos y realizar la valoración de los mismos la palabra que compone el *hashtag*. Se cuenta con una lista dada de los mismos, y éstos son los que se quitan del análisis a realizar.

<b>Nombre del metanodo</b>	Filtrar Productos y Servicios
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad

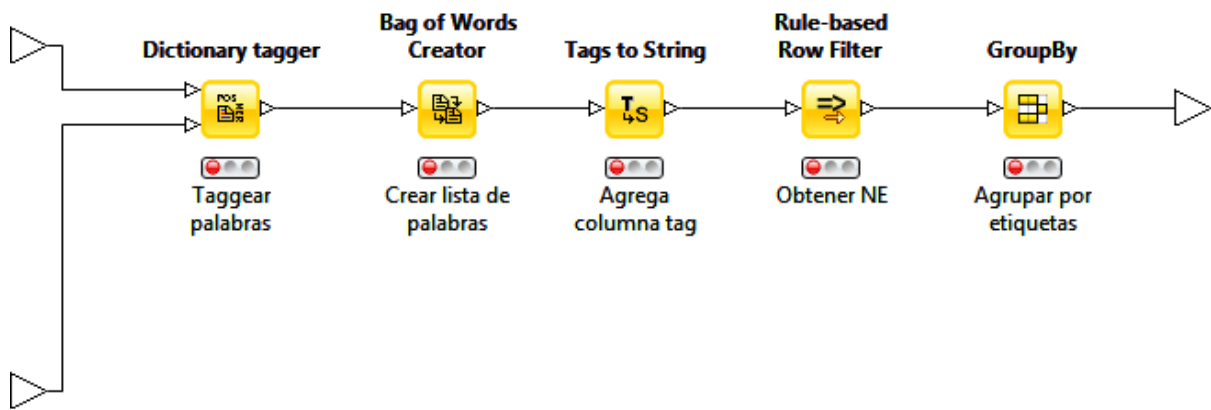


Imagen 9.39 Metanodo Filtrar Productos y Servicios

En éste metanodo se filtra una lista de productos y servicios. Luego de obtener la misma (dada por el experto de negocios) se realiza el filtrado de los tweets que contienen al menos un servicio o producto de los que contiene la lista. Luego, se sigue adelante en el proceso sólo con los datos que se quieren analizar.

<b>Nombre del metanodo</b>	Filtrar Tweets
<b>Flujos en lo que se encuentra presente</b>	Colectate_Extracción_Twitter_Stream Colectate_Extracción_Twitter

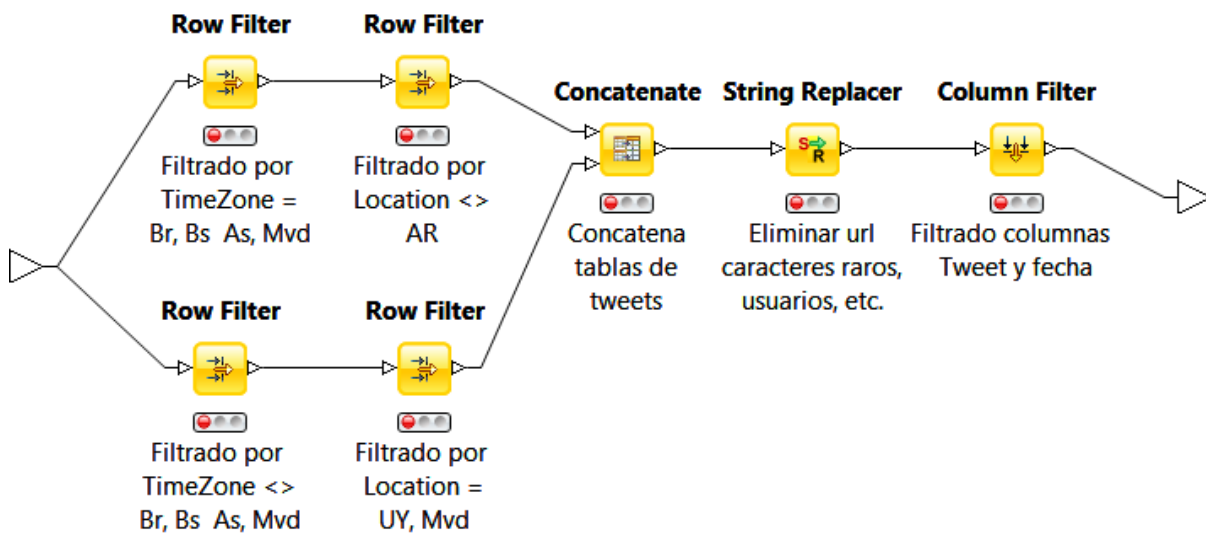


Imagen 9.40 Metanodo Filtrar Tweets

En este metanodo se realiza el filtrado tweets extraídos utilizando la siguiente estrategia:

- ✚ La lista de tweets se separa en dos listas disjuntas tomando como criterio de separación inicial, el *timezone* del usuario que generó el tweet. Dado que para Uruguay no existe un *timezone* específico entre las opciones de Twitter, se detectó que la mayoría de los usuarios tienen configurado por defecto los *timezone* de Buenos Aires o Brasilia.
- ✚ Si el usuario tiene alguno de estos *timezone*, se filtra luego por Locación, descartando los que tengan específicamente como locación Argentina.
- ✚ Si el usuario NO tiene alguno de estos *timezone*, y en la locación tiene referenciado algún valor diferente a Uruguay se filtra la tupla.
- ✚ Luego se unifican los conjuntos resultantes ya filtrados.

A nivel de negocio, cuando se desea modificar el ámbito geográfico tomado en cuenta para la extracción y análisis, se debe proceder a modificar la parametrización de este metanodo, incluyendo o excluyendo términos comúnmente utilizados para hacer referencia a dicha locación.

Para mejorar la calidad de los datos a ser almacenados en la base, se realizó un filtrado de caracteres especiales así como también de URLs puestas por el usuario en el tweet, ya que estos fallaban en la inserción en las tablas.

Para lograr este cometido, se utilizó la expresión regular “[^x00-xFF](@\S+)(http)(\S+)”

[^x00-xFF]: filtra cualquier carácter fuera del rango ASCII

(@\S+): filtra el símbolo “@” pegado con cualquier letra (usuarios)

(http): filtra la ocurrencia de las letras “http”

(\S+): filtra cualquier cadena de caracteres con formato “//url.com.uy”

<b>Nombre del metanodo</b>	Generar <i>String</i> Consulta
<b>Flujos en lo que se encuentra presente</b>	Colectate_Extracción_Twitter

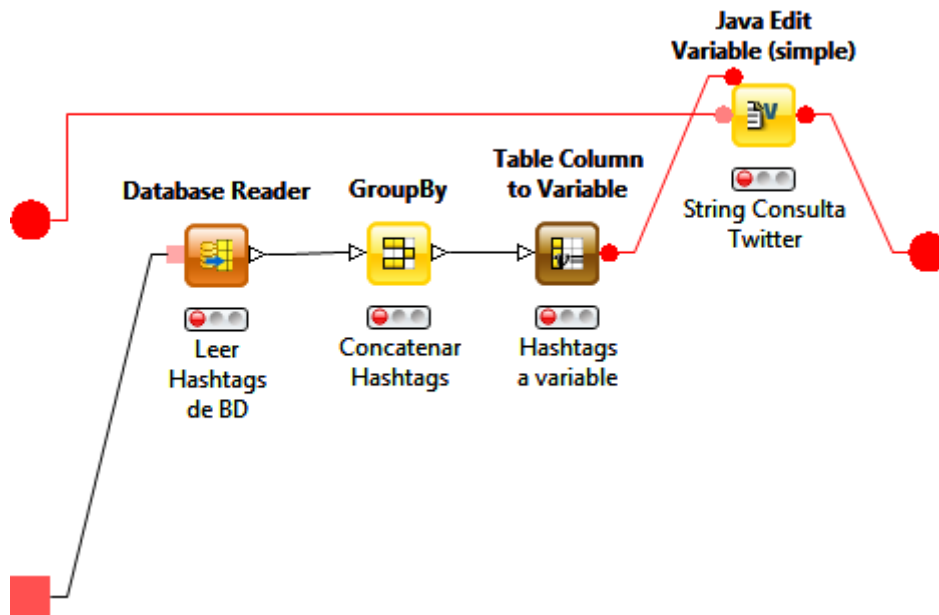


Imagen 9.41 Metanodo Generar String Consulta

Este metanodo cumple con el objetivo de leer la lista de *hashtags* obtenida desde la base de datos, agruparla y convertirla en un formato que luego se utilizará para extraer tweets del interés deseado.

<b>Nombre del metanodo</b>	Generar Tabla Análisis a Guardar
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad

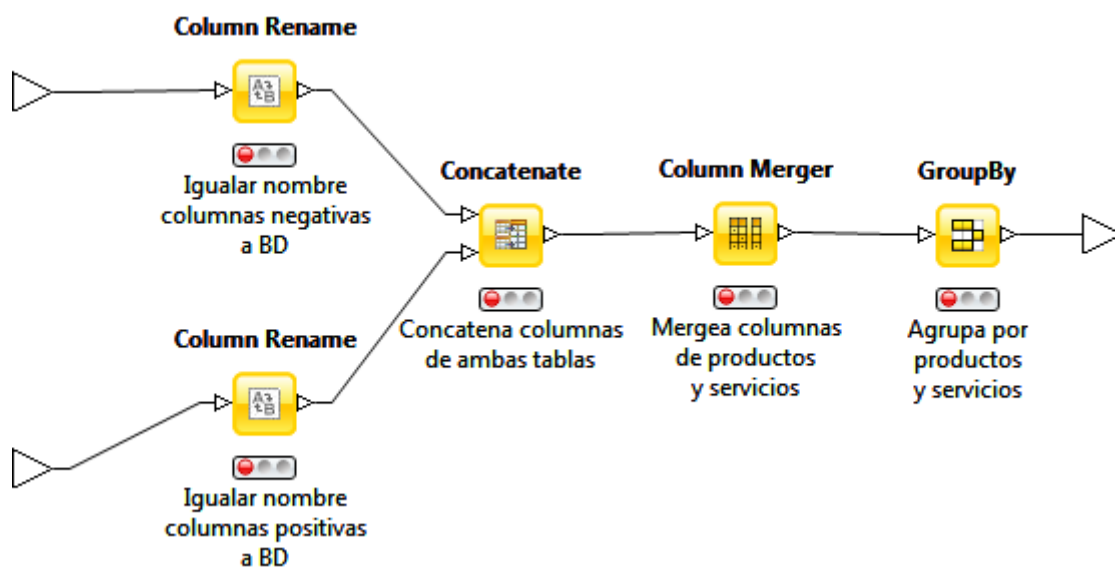


Imagen 9.42 Metanodo Generar Tabla Análisis a Guardar

Este metanodo, tiene la tarea de generar la tabla que luego se guardará en la base de datos. A él llegan dos tablas, luego de una conversión concatena las columnas de ambas tablas y por último agrupa según la lista dada de productos y servicios de manera de obtener para cada uno de ellos la cantidad de menciones.

<b>Nombre del metanodo</b>	Generar TagCloud
<b>Flujos en lo que se encuentra presente</b>	Colectate_Oportunidades_Mejora, Colectate_Análisis_Tendencias

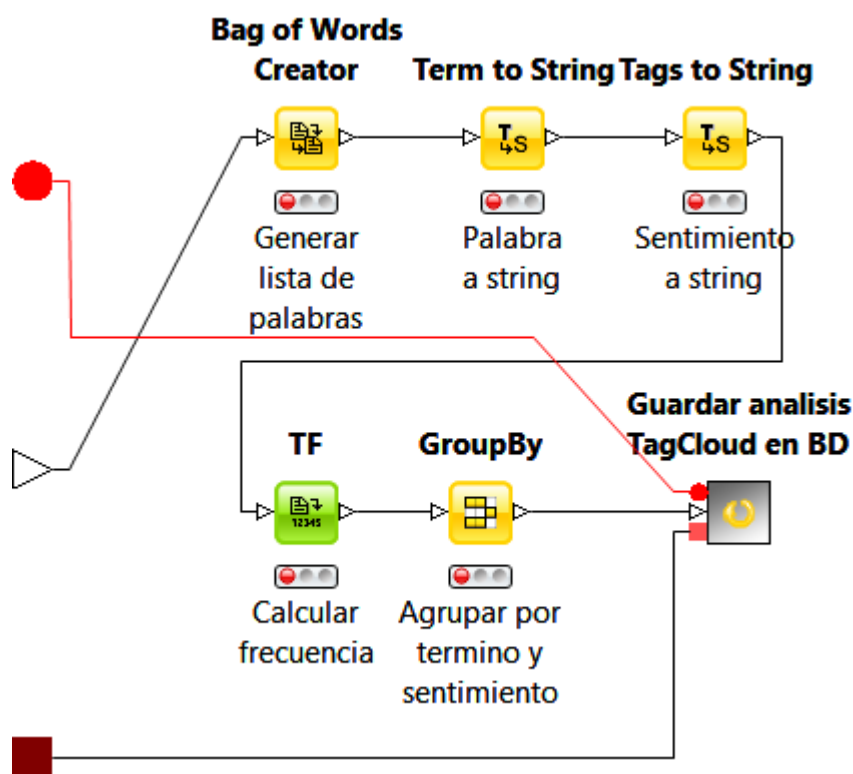


Imagen 9.43 Metanodo Generar TagCloud

En el metanodo llamado “Generar TagCloud”, se genera la lista de palabras con su respectivo sentimiento. Se calcula la frecuencia de los mismos para determinar gráficamente la importancia de cada palabra (a más frecuencia, más veces esa palabra fue encontrada y por lo tanto es más visible gráficamente en el TagCloud).

<b>Nombre del metanodo</b>	Generar Tópicos
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Tendencias, Colectate_Oportunidades_Mejora

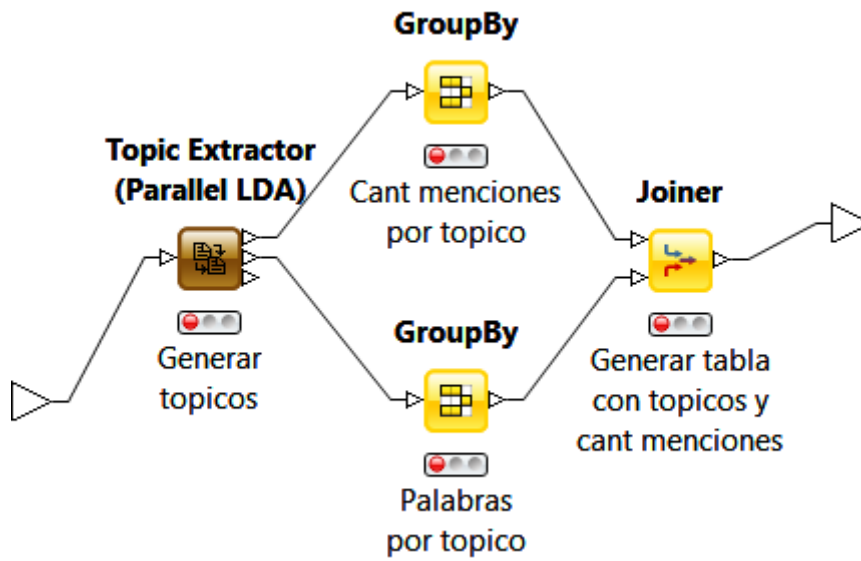


Imagen 9.44 Metanodo Generar Tópicos

Para la generación de tópicos (identificación de categorías), se utilizó un algoritmo probabilístico llamado “Parallel LDA (Latent Dirichlet Allocation)”. En el Anexo 5 - *Latent Dirichlet Allocation* (LDA) se encuentra una descripción de la investigación realizada sobre dicho algoritmo.

Para cada uno de los tópicos generados, se cuenta la cantidad veces que es mencionado, así como también la cantidad de ocurrencias de los términos contenidos en cada tópico.

<b>Nombre del metanodo</b>	Guardar análisis en BD
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad

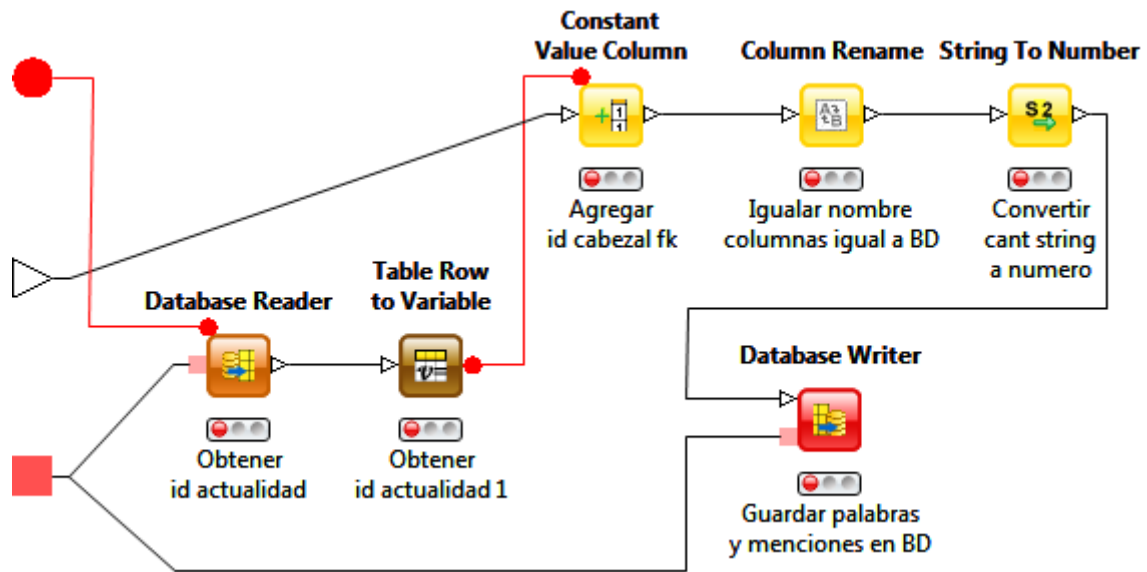


Imagen 9.45 Metanodo Guardar análisis en BD

Este metanodo es el encargado de guardar en la base de datos el análisis efectuado partiendo de una lista de servicios y productos de interés. Por lo tanto, aquí se persiste tanto ese dato (servicio o producto) como también la cantidad de menciones del mismo para luego visualizar los mismos en un reporte. Los datos se guardan en la tabla o\_Análisis\_Actualidad\_Detalle.

<b>Nombre del metanodo</b>	Guardar análisis negativo en BD
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Tendencias

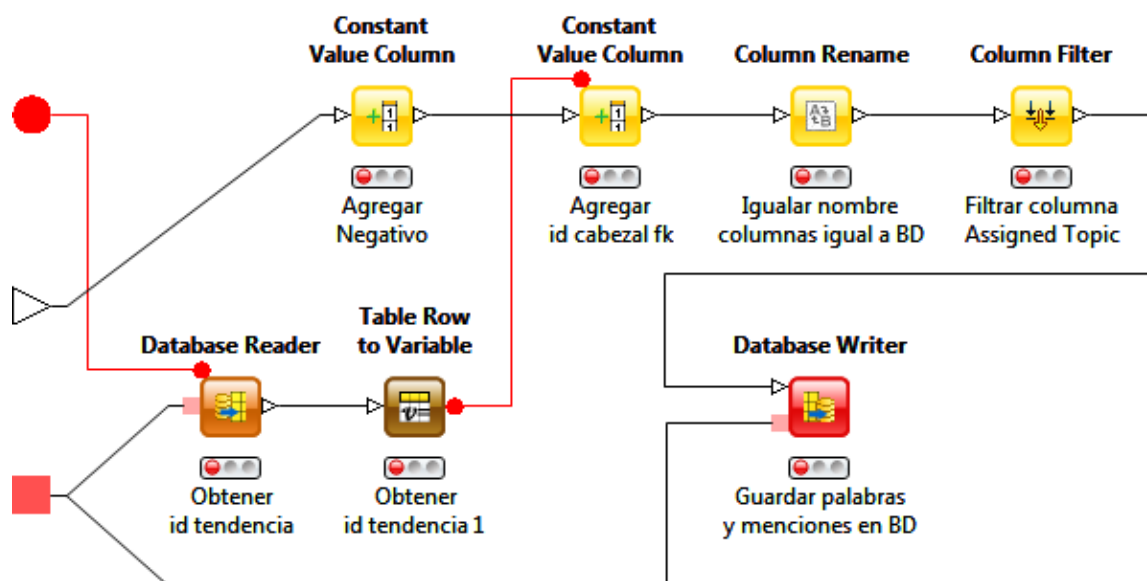


Imagen 9.46 Metanodo Guardar análisis negativo en BD

El metanodo “Guardar análisis negativo en BD”, guarda el análisis generado con la cantidad de tópicos creados y la cantidad de palabras agrupadas en cada uno de ellos. Como el título lo menciona, en este caso los tópicos generados agrupan palabras valoradas negativamente. Los datos se guardan en la tabla o\_Análisis\_Tendencias\_Detalle.

<b>Nombre del metanodo</b>	Guardar análisis positivo en BD
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Tendencias

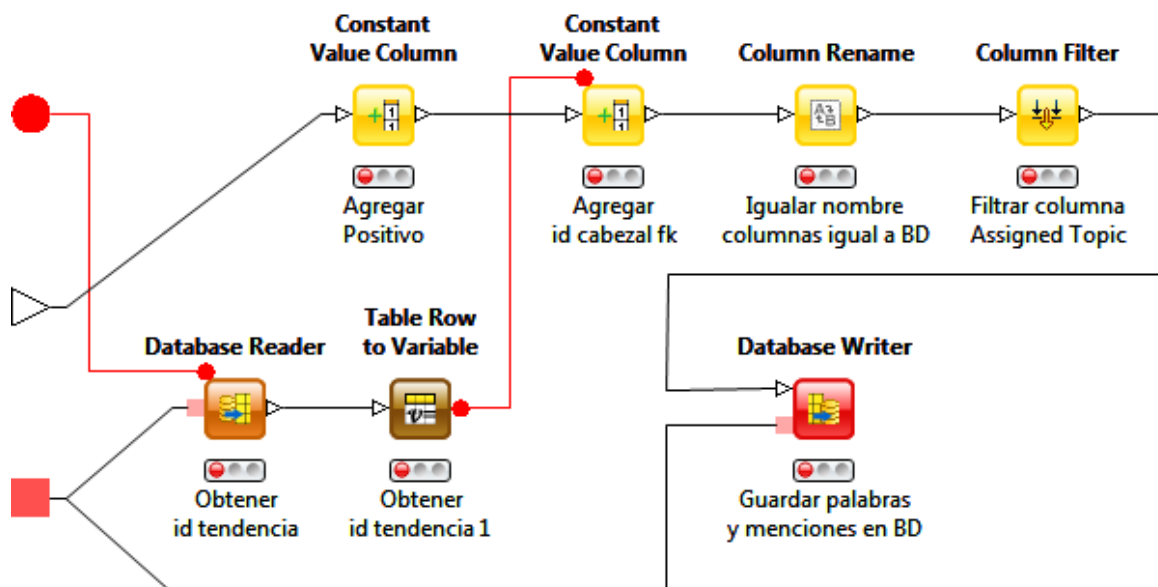


Imagen 9.47 Metanodo Guardar análisis positivo en BD

El metanodo “Guardar análisis positivo en BD”, guarda el análisis generado con la cantidad de tópicos creados y la cantidad de palabras agrupadas en cada uno de ellos. Como el título lo menciona, en este caso los tópicos generados agrupan palabras valoradas positivamente. Los datos se guardan en la tabla o\_Análisis\_Tendencias\_Detalle.

<b>Nombre del metanodo</b>	Guardar Cabecal en BD
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad, Colectate_Análisis_Tendencias, Colectate_Oportunidades_Mejora,

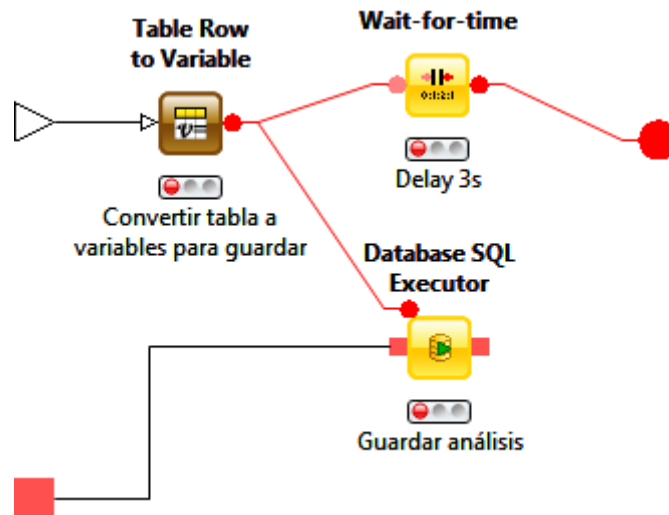


Imagen 9.48 Metanodo Guardar Cabezal en BD

El metanodo “Guardar Cabezal en BD” tiene por objetivo insertar en la base de datos el análisis realizado con los siguientes datos: fuente extraída, fecha de ejecución, cantidad total extraída, cantidad total valorados y la identificación del último tweet analizado.

<b>Nombre del metanodo</b>	Guardar Fecha Ultimo Tweet
<b>Flujos en lo que se encuentra presente</b>	Colectate_Extracción_Twitter, Colectate_Extracción_Twitter_Colectate, Colectate_Extracción_Twitter_Stream

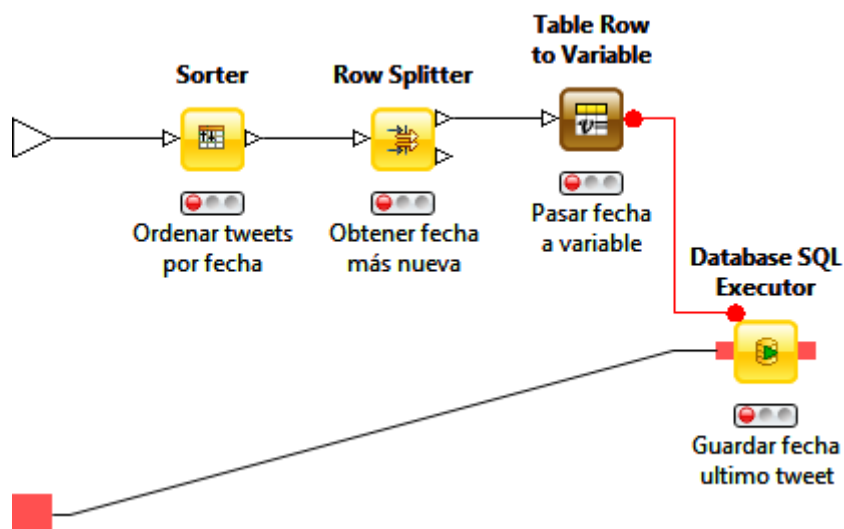


Imagen 9.49 Metanodo Guardar Fecha Ultimo Tweet

En este metanodo, el objetivo es conseguir la fecha del último tweet analizado (es decir la fecha más nueva) e insertarla en la base de datos en la tabla o\_Fecha\_Ejecucion\_Extraccion para luego poder utilizarla.

<b>Nombre del metanodo</b>	Guardar Sentimiento en Tweet
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Sentimiento, Colectate_Análisis_Sentimiento_Streaming

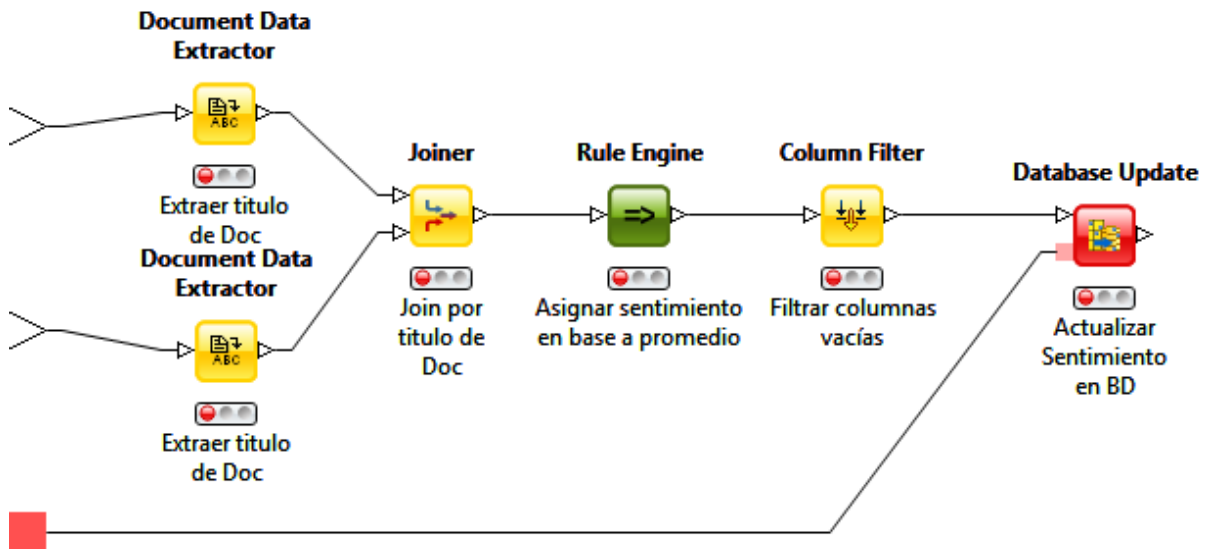


Imagen 9.50 Metanodo Guardar Sentimiento en Tweet

“Guardar Sentimiento en Tweet” es un metanodo creado con el objetivo de asignar un sentimiento al tweet y actualizar la tabla correspondiente en la base de datos guardando en la misma, el tweet junto a su sentimiento (neutro, negativo, positivo).

<b>Nombre del metanodo</b>	Guardar tópicos negativo en BD
<b>Flujos en lo que se encuentra presente</b>	Colectate_Oportunidades_Mejora

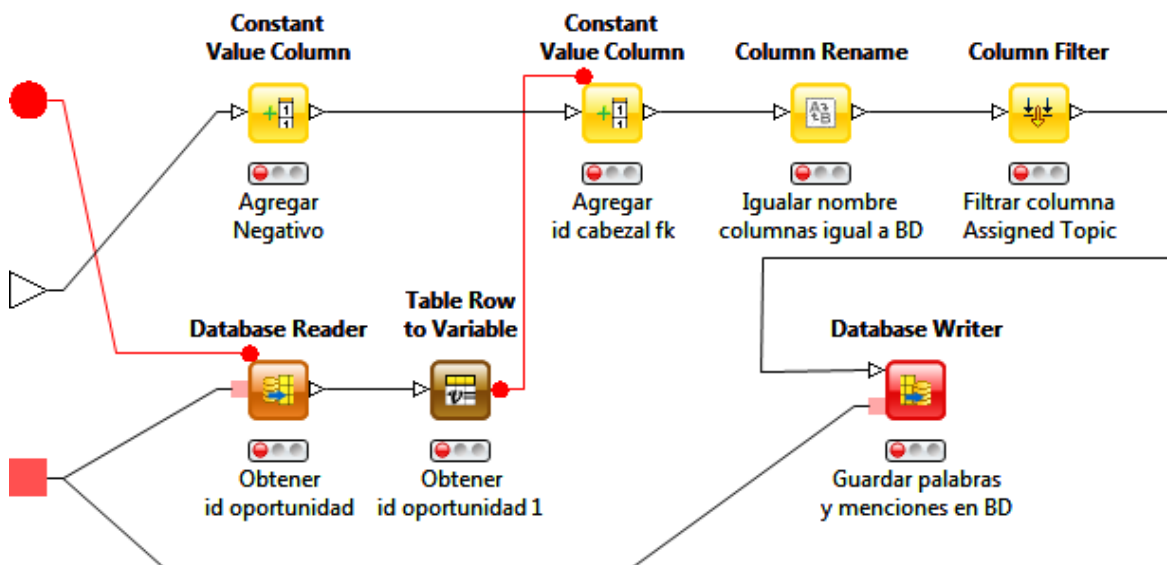


Imagen 9.51 Metanodo Guardar tópicos negativo en BD

El metanodo “Guardar tópicos negativo en BD”, guarda el análisis generado con la cantidad de tópicos creados y la cantidad de palabras agrupadas en cada uno de ellos. Como el título lo menciona, en este caso los tópicos generados agrupan palabras valoradas negativamente. Los datos se guardan en la tabla o\_Oportunidades\_Mejora\_Detalle.

<b>Nombre del metanodo</b>	Guardar tópicos positivo en BD
<b>Flujos en lo que se encuentra presente</b>	Colectate_Oportunidades_Mejora

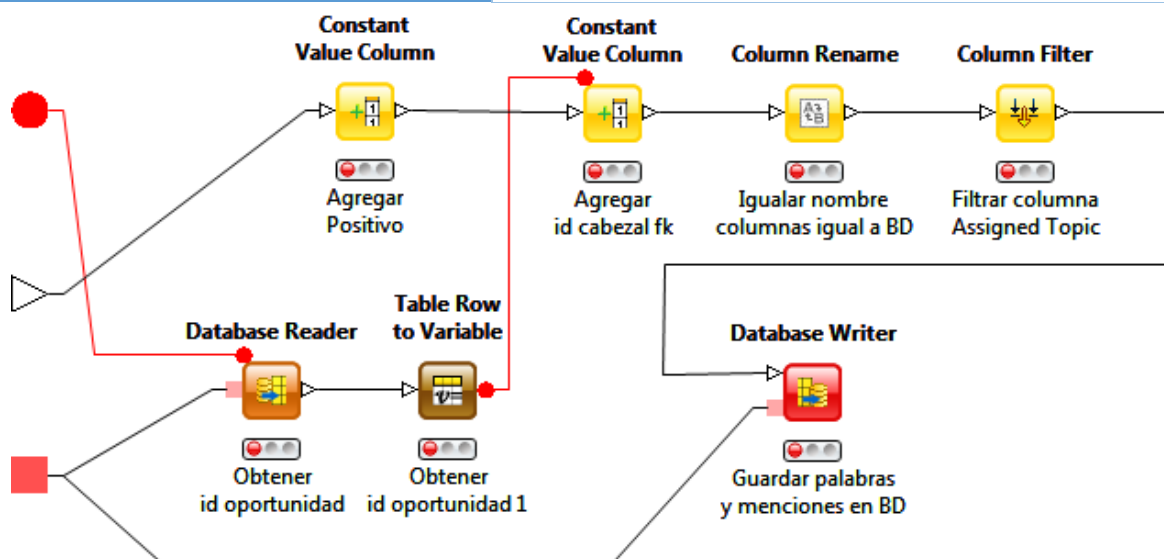


Imagen 9.52 Metanodo Guardar tópicos positivo en BD

El metanodo “Guardar tópicos positivo en BD”, guarda el análisis generado con la cantidad de tópicos creados y la cantidad de palabras agrupadas en cada uno de ellos. Como el título lo menciona, en este caso los tópicos generados agrupan palabras valoradas positivamente. Los datos se guardan en la tabla o\_Oportunidades\_Mejora\_Detalle.

<b>Nombre del metanodo</b>	Leer Credenciales Twitter
<b>Flujos en lo que se encuentra presente</b>	Colectate_Extracción_Twitter, Colectate_Extracción_Twitter_Colectate, Colectate_Extracción_Twitter_Stream

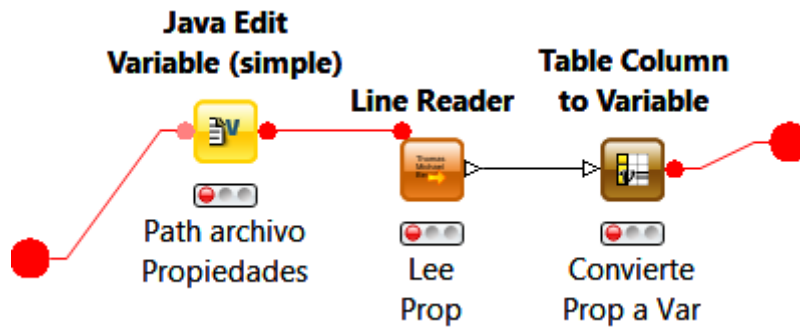


Imagen 9.53 Metanodo Leer Credenciales Twitter

Este metanodo se encarga de leer desde un archivo llamado "CredencialesTwitterExtraccionSinFiltrar" las credenciales para acceder a la aplicación para luego proceder a la extracción de los tweets.

<b>Nombre del metanodo</b>	Leer Productos y Servicios
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad

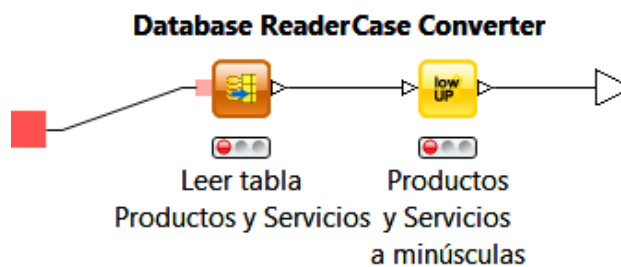


Imagen 9.54 Metanodo Leer Productos y Servicios

Leer la lista de Productos y Servicios es la tarea a realizar por este metanodo. Luego de realizar la conexión a la base de datos, éste selecciona la lista completa de productos y servicios de la tabla correspondiente, la misma es p\_Productos\_Servicios.

<b>Nombre del metanodo</b>	Obtener Cabezal
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Sentimiento_Streaming

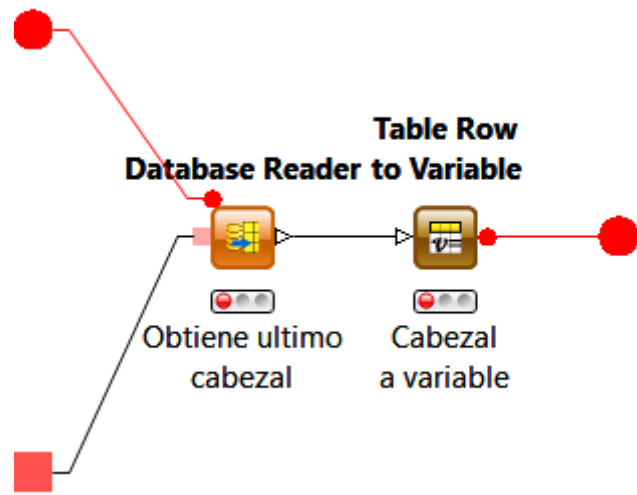


Imagen 9.55 Metanodo Obtener Cabezal

La finalidad del metanodo "Obtener Cabezal" consiste en seleccionar los datos correspondientes a: fecha de ejecución, cantidad total de extraídos, cantidad total de valorados, cantidad de positivos, negativos y neutros que tengan la máxima fecha de ejecución.

<b>Nombre del metanodo</b>	Obtener Fecha Ultimo Tweet
<b>Flujos en lo que se encuentra presente</b>	Colectate_Extracción_Twitter, Colectate_Extracción_Twitter_Colectate

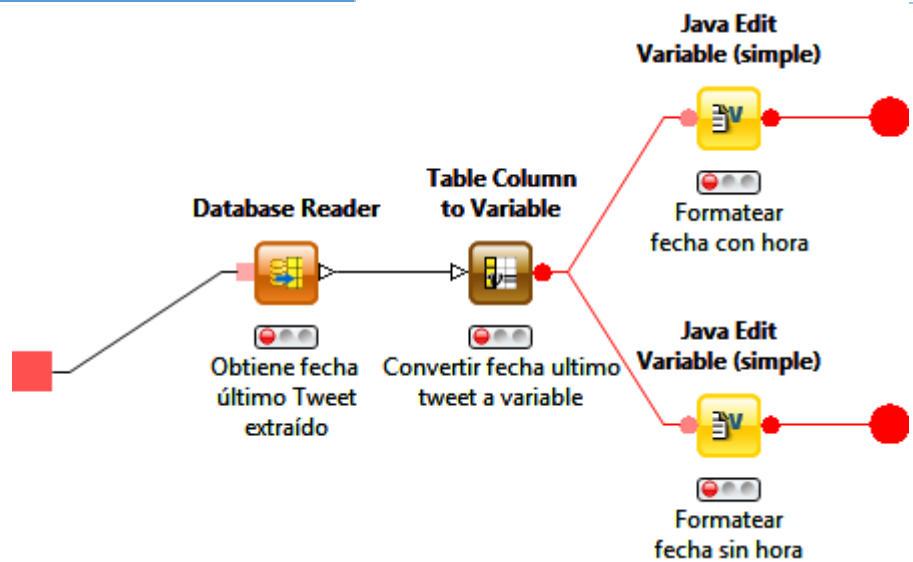


Imagen 9.56 Metanodo Obtener Fecha Ultimo Tweet

En el metanodo mostrado, la finalidad es obtener la fecha del último tweet extraído. Es por esto, que se selecciona de la base de datos llamada o\_Fechas\_Ejecucion\_Extracción la máxima fecha (o sea la más nueva) correspondiente al proceso de extracción.

<b>Nombre del metanodo</b>	Obtener Tweets de BD
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad, Colectate_Análisis_Sentimiento, Colectate_Análisis_Sentimiento_Streaming, Colectate_Análisis_Tendencias, Colectate_Oportunidades_Mejora

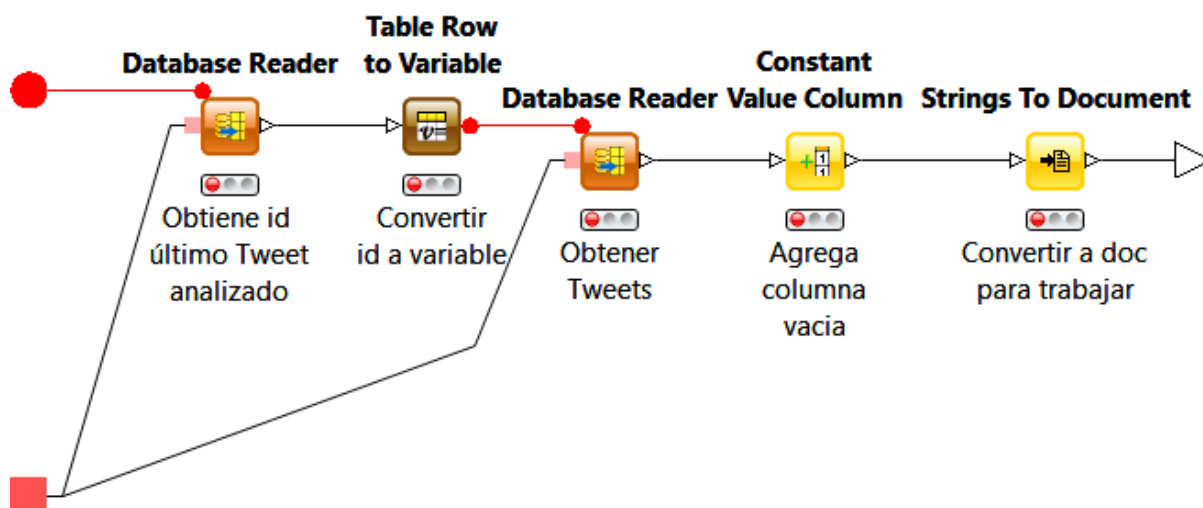


Imagen 9.57 Metanodo Obtener Tweets de BD

“Obtener Tweets de BD”, consiste en obtener todos los tweets que aún no se han analizado. Esto se consigue identificando el último tweet analizado (obteniendo su id), para luego solicitar a la base de datos que entregue todos los tweets que tienen fecha superior e éste para poder analizarlos.

<b>Nombre del metanodo</b>	Ocurrencia de palabras
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad

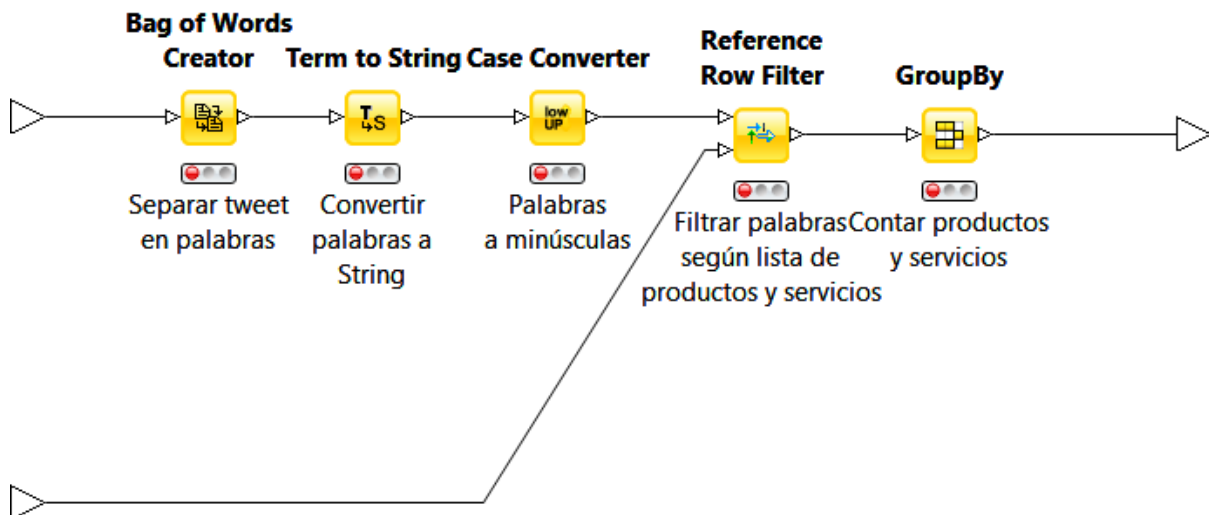


Imagen 9.58 Metanodo Ocurrencia de palabras

La “Ocurrencia de palabras”, se utiliza para detectar menciones positivas y negativas partiendo de una lista dada de productos y servicios de interés. La finalidad de este metanodo, es poder obtener la cantidad de ocurrencias de las mismas diferenciando si se mencionan positiva o negativamente.

<b>Nombre del metanodo</b>	Preprocesamiento datos
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad, Colectate_Análisis_Sentimiento, Colectate_Análisis_Sentimiento_Streaming, Colectate_Análisis_Tendencia, Colectate_Oportunidades_Mejora,

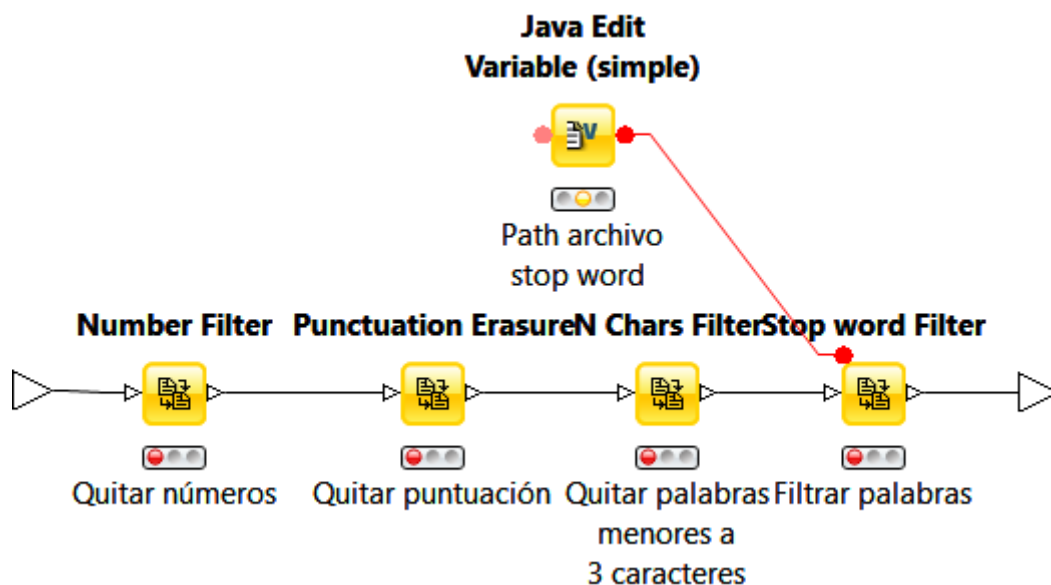


Imagen 9.59 Metanodo Preprocesamiento datos

Siendo el análisis de sentimiento el proceso base para los demás análisis, resulta de interés realizar un tratamiento de los datos extraídos para eliminar términos “ruido”. Este metanodo cumple este propósito al filtrar: números, puntuación, palabras menores a 3 caracteres y palabras contenidas en el archivo “*Stop words*” [40]. Este archivo contiene conectores, pronombres, conectores, adverbios, algunos verbos y sus conjugaciones, así como también la lista de palabras de 3 letras en español.

<b>Nombre del metanodo</b>	Valoración Sentimiento
<b>Flujos en lo que se encuentra presente</b>	Colectate_Análisis_Actualidad, Colectate_Análisis_Sentimiento, Colectate_Análisis_Sentimiento_Streaming, Colectate_Análisis_Tendencia, Colectate_Oportunidades_Mejora,

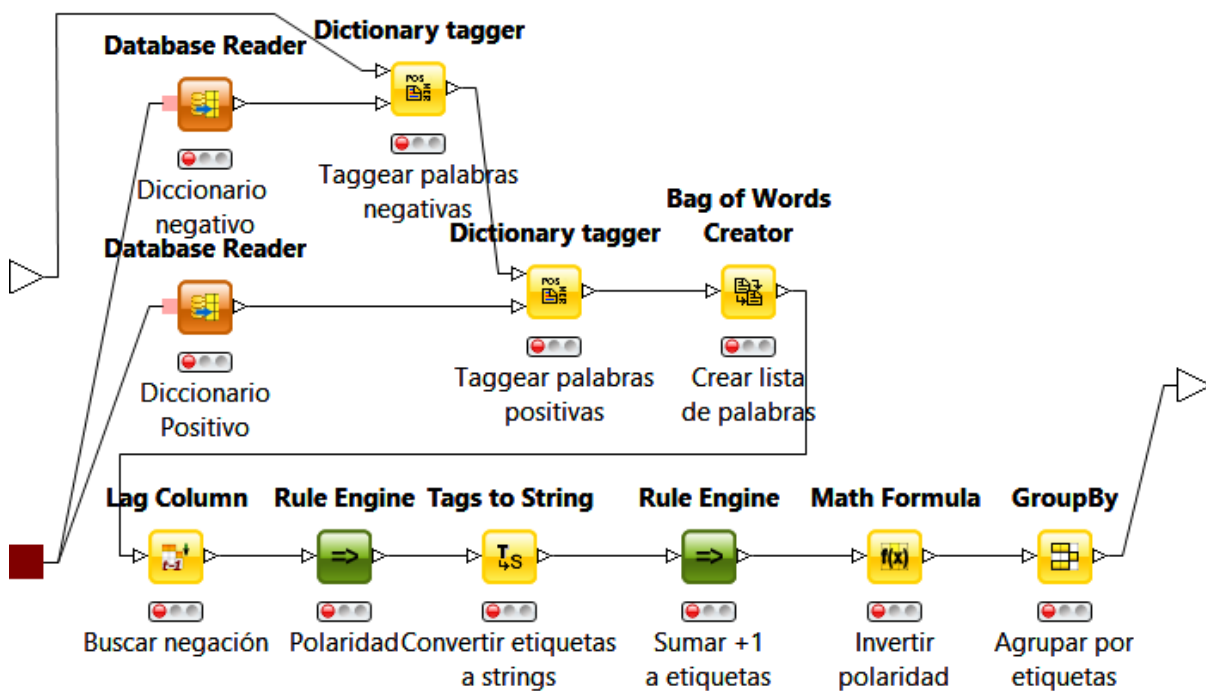


Imagen 9.60 Metanodo Valoración Sentimiento

Este metanodo está compuesto por: nodos relativos a la identificación de palabras con contenido emocional, nodos relativos al manejo de la negación y nodos relativos al cálculo del sentimiento general del tweet analizado.

Para el cálculo del sentimiento, se realizó una identificación de las palabras con contenido emocional presentes en cada tweet y se valoraron en base a un diccionario de palabras. Dicho diccionario en idioma español, fue confeccionado por el Laboratorio de Lenguaje Natural y Procesamiento de texto del Instituto Politécnico Nacional de México. En el mismo, se asigna un valor sentimental a palabras seleccionadas siguiendo una serie de criterios de valoración que clasifican cada palabra en una categoría dada [41].

En el caso de las palabras negadas, se hacía necesario identificar la ocurrencia del adverbio de negación “no”, el cual niega un adjetivo o verbo que, si aparecieran por sí solos, se contarían como palabras positivas causando que el cálculo final del sentimiento del tweet sea incorrecto.

Si bien existen otras formas de negación en el lenguaje español [42], el caso que se tomó en cuenta para el análisis es el siguiente: en caso de que una palabra “n” identificada por el diccionario esté precedida por un “no” en la posición n-2, la polaridad de dicha palabra se invertirá.

Ejemplo: *Este restaurante es caro, y **no** me gusta.*

En este caso, la palabra “gusta” fue identificada a través del diccionario como una palabra de carácter positivo, pero al ser precedida por el adverbio “no” en la posición n-2, la polaridad de esta palabra se invierte.

Finalmente, para el cálculo total del sentimiento, se le asigna a cada palabra identificada el valor 0, +1 o -1, y se realiza el cálculo de la media para cada tweet, obteniéndose un valor final que determina si el tweet es positivo, negativo o neutro.

Este método no es exacto ya que no se toman en cuenta los casos más complejos, sin embargo los test arrojaron como resultado un 80% de correctitud en el análisis.


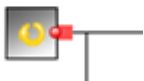








## 9.19. Anexo 19 - Detección de dispositivos

Este, es un resumen del código incluido en el índice del sitio para poder detectar desde donde se está consumiendo el mismo. En el siguiente ejemplo, se puede detectar si es un móvil o una *tablet*.

```
1 | <?php
2 |     include('Mobile_Detect.php'); // acuérdate de incluir la ruta correcta
3 |
4 |     $detect = new Mobile_Detect();
5 |
6 |     if ($detect->isMobile()) { //cualquier dispositivo móvil
7 |     }
8 |     if ($detect->isTablet()) { //cualquier dispositivo tipo tableta
9 |     }
```

## 9.20. Anexo 20 - Nomenclatura de flujos en Knime

A continuación se muestran las principales figuras utilizadas en el desarrollo de los flujos en Knime, explicando para cada una de ellas su interpretación.

Figura	Interpretación
 <p><b>Database Writer</b> Guardar Tweets</p>	<p>Nodo. En la parte superior en negrita se muestra el nombre del nodo en Knime según su función técnica. En la parte inferior se muestra el nombre funcional asignado al nodo.</p>
 <p><b>Conexión MySQL</b></p>	<p>Metanodo (agrupación de nodos).</p>
	<p>Sentido del flujo de datos entre nodos</p>
	<p>A la izquierda se encuentra el conector de entrada de datos a un nodo. A la derecha se encuentra el conector de salida de datos de un nodo.</p>
	<p>A la izquierda en color rosado se encuentra el conector de entrada de conexión con base de datos.</p>
	<p>A la izquierda se encuentra el conector de salida de datos de un nodo personalizado para ColectaTe. A la derecha se encuentra el conector de entrada de datos de un nodo personalizado para ColectaTe.</p>
	<p>A la izquierda se encuentra el conector de salida de variables de un nodo. A la derecha se encuentra el conector de entrada de variables de un nodo.</p>
	<p>Nodo pendiente de configuración.</p>
	<p>Nodo configurado y pendiente de ejecución.</p>
	<p>Nodo ejecutado exitosamente.</p>