



FACULTAD DE INGENIERÍA Y CIENCIAS APLICADAS

DESARROLLO DE UNA SOLUCIÓN PARA LA VISUALIZACIÓN Y
ANÁLISIS DE SENTIMIENTO EN BASE A LOS MENSAJES QUE
REALIZAN LOS USUARIOS A UNA EMPRESA MEDIANTE UNA RED
SOCIAL.

AUTOR

Daniel Jose Calatrava Hurtado

AÑO

2019



FACULTAD DE INGENIERÍA Y CIENCIAS APLICADAS

DESARROLLO DE UNA SOLUCIÓN PARA LA VISUALIZACIÓN Y ANÁLISIS
DE SENTIMIENTO EN BASE A LOS MENSAJES QUE REALIZAN LOS
USUARIOS A UNA EMPRESA MEDIANTE UNA RED SOCIAL.

Trabajo de Titulación presentado en conformidad con los requisitos
establecidos para optar por el título de Ingeniero en Sistemas de Computación
e Informática

Profesor Guía

Ph.D. Mario Salvador González Rodríguez

Autor

Daniel Jose Calatrava Hurtado

Año

2019

DECLARACIÓN DEL PROFESOR GUÍA

Declaro haber dirigido el trabajo, Desarrollo de una solución para la visualización y análisis de sentimiento en base a los mensajes que realizan los usuarios a una empresa mediante una red social, a través de reuniones periódicas con el estudiante Daniel Jose Calatrava Hurtado, en el semestre 201910, orientando sus conocimientos y competencias para un eficiente desarrollo del tema escogido y dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación.

Mario Salvador González Rodríguez

Doctor en Informática

Pasaporte: 0958376345

DECLARACIÓN DEL PROFESOR CORRECTOR

Declaro haber revisado este trabajo, Desarrollo de una solución para la visualización y análisis de sentimiento en base a los mensajes que realizan los usuarios a una empresa mediante una red social, del estudiante Daniel Jose Calatrava Hurtado, en el semestre 201910, dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación.

Bernarda Cecibel Sandoval Romo

Master en Ciencias de la Computación

CI: 1709974453

DECLARACIÓN DE AUTORÍA DEL ESTUDIANTE

Declaro que este trabajo es original, de mi autoría, que se han citado las fuentes correspondientes y que en su ejecución se respetaron las disposiciones legales que protegen los derechos de autor vigentes.

Daniel Jose Calatrava Hurtado

Pasaporte: 066812348

AGRADECIMIENTOS

A mi familia y a Dios, por haberme dado la oportunidad de formarme en esta prestigiosa universidad y haber sido mi apoyo durante todo este tiempo. A mi tutor Mario González Ph.D y a mi correctora Bernarda Sandoval MsC, por haberme guiado, no solo en la elaboración de este trabajo de titulación, sino a lo largo de mi carrera universitaria y haberme brindado el apoyo para formarme profesionalmente.

DEDICATORIA

A mis padres Ramon e Ingrid quienes con su amor, paciencia y esfuerzo me han permitido llegar a cumplir hoy esta meta, A mi hermano Brian por su apoyo incondicional durante todo este proceso, A toda mi familia porque con sus oraciones, consejos y palabras de aliento hicieron de mí una mejor persona y de una u otra forma me acompañan en todas mis metas.

RESUMEN

El presente trabajo de titulación busca dar solución a los problemas que tienen las distintas empresas, en su relación con los clientes, entendiendo que, existe un gran porcentaje de ellas que están presentes en Twitter, así como usuarios que utilizan esta aplicación para conocer a las empresas y distintas marcas en el mercado. Aquellas empresas que no utilizan estos medios no establecen una relación adecuada con sus clientes o no conocen exactamente el nivel de satisfacción por parte de los usuarios, lo que puede ocasionar pérdidas monetarias y de reputación para estas marcas. Por lo que particularmente se busca analizar el comportamiento de los usuarios en los ámbitos de un producto, evento, empresa u organización, este sistema se desarrolla a partir de la unión de ciertas herramientas de Microsoft, que finalmente permitirá generar reportes para analizar los mensajes enviados por los usuarios en la red social Twitter y donde se crearán distintas medidas o métricas que facilitarán al cliente el entendimiento de los usuarios. Para el desarrollo de este trabajo de titulación, se escogieron las metodologías Scrum y CRISP-DM, debido a que son metodologías bastante adaptables para la corrección y direccionamiento de la solución y a su vez con la metodología CRISP-DM se puede desarrollar eficientemente un proyecto de minería de datos. Con el desarrollo de este proyecto de titulación se espera que las empresas puedan tener un enfoque más adecuado del comportamiento de los usuarios hacia productos, eventos, personas y la misma empresa, a través de los mensajes que estos envían mediante la red social Twitter.

ABSTRACT

This degree work seeks to solve the problems that different companies have, in their relationship with customers, understanding that there is a large percentage of them that are present on Twitter, as well as users who use this application to know the companies and different brands in the market. Those companies that do not use these means do not establish an adequate relationship with their customers or do not know exactly the level of satisfaction on the part of the users, which can cause monetary and reputation losses for these brands. For what is particularly sought to analyze the behavior of users in the fields of a product, event, company or organization, this system is developed from the union of certain Microsoft tools, which will finally allow generating reports to analyze the messages sent by the users in the social network Twitter and where different measures or metrics will be created that will facilitate the client's understanding of the users. For the development of this titling work, the Scrum and CRISP-DM methodologies were chosen, because they are quite adaptable methodologies for correcting and directing the solution and, in turn, with the CRISP-DM methodology, a project can be efficiently developed. data mining. With the development of this titling project it is expected that companies can have a more adequate approach to the behavior of users towards products, events, people and the same company, through the messages they send through the social network Twitter.

ÍNDICE

1. Capítulo I: Introducción.....	1
1.1 Antecedentes.....	1
1.2 Objetivo General.....	5
1.3 Objetivos específicos.....	5
1.4 Alcance	6
1.5 Justificación	8
2. Capítulo II: Marco Teórico.....	9
2.1 Introducción	9
2.2 Metodologías	9
2.2.2 Metodología Scrum.....	9
2.2.2 Metodología CRISP-DM	11
2.3 Historia de la Empresa	22
2.4 Análisis de Sentimiento	23
2.5 Programas y Herramientas	27
2.5.1 Power BI Desktop, Web y Mobile	27
2.5.2 Microsoft Flow.....	29
2.5.3 Microsoft Azure.....	32
2.5.4 Microsoft SQL Server	33
2.5.5 Microsoft Servicios Cognitivos.....	34
2.6 Conclusión del Capítulo.	35
3. Capítulo III: Desarrollo de Solución	35
3.1 Introducción.	35
3.2 Criterio de obtención de datos.	36

3.3	Metodología para análisis de sentimientos.....	37
3.4	Análisis de Estructura Tecnológica	38
3.5	Conclusión del Capítulo.	39
4.	Capitulo IV: Implementación	40
4.1	Introducción	40
4.2	Diagrama de Solución.	40
4.3	Configuración y Creación de Servicios en Azure.....	42
4.4	Configuración y Creación de flujos de trabajo en Microsoft Flow.....	49
4.5	Conclusión del Capítulo.	58
5.	Capítulo V: Resultados.	59
5.1	Introducción	59
5.2	Análisis de Sentimientos de Organización o Empresa	59
5.3	Análisis de Sentimientos para un Producto	65
5.4	Análisis de Sentimientos para Evento.....	70
5.5	Conclusión del Capítulo.	74
6.	CONCLUSIONES Y RECOMENDACIONES.	75
6.1	Conclusiones	75
6.2	Recomendaciones.....	76
	REFERENCIAS.....	77
	ANEXOS	83

1. Capítulo I: Introducción

1.1 Antecedentes

Con la cultura de blogs y redes sociales, la minería de opinión y el análisis de sentimientos se convirtieron en un campo de interés para muchas investigaciones. Se presentó una descripción muy amplia del trabajo existente en (Pang y Lee, 2008). En su encuesta, los autores describen las técnicas existentes y enfoques para una recuperación de información orientada a la opinión. Sin embargo, no muchas investigaciones en minería de opinión consideraron blogs e incluso mucho menos abordaron el microblogging.

En (Yang et al., 2007), los autores usan blogs web para construir un corpus para el análisis de sentimientos y el uso de emociones. Los iconos asignados a las publicaciones del blog como indicadores del estado de ánimo de los usuarios.

Los autores aplicaron SVM (Support Vector Machine) y CRF (Conditional Random Fields) para clasificar los sentimientos a nivel de la oración y luego investigaron varias estrategias para determinar el sentimiento general del documento. Como resultado, la estrategia ganadora se define considerando el sentimiento de la última oración como el sentimiento a nivel de documento.

J. Read en (Read, 2005) usó emoticonos como “:-)” y “:-(”, Para formar un conjunto de entrenamiento para la clasificación de sentimientos. Para este propósito, el autor recopiló textos que contenían emoticonos de los grupos de noticias de Usenet. El conjunto de datos fue dividido en "positivo" (textos con emoticonos felices) y "negativo" (Textos con emoticonos tristes o enojados).

Clasificadores limitados por la vulnerabilidad: SVM y Naive Bayes, pudieron obtener hasta un 70% de precisión en el conjunto de prueba. En (Go et al., 2009), los autores utilizaron Twitter para recopilar datos de entrenamiento y luego realizar una búsqueda de sentimientos. El enfoque es similar a (Read, 2005). Los autores construyen cuerpos mediante el uso de emoticonos para obtener muestras "positivas" y "negativas", y luego usar varios clasificadores. El mejor

resultado fue obtenido por el clasificador Naive Bayes con una medida de información mutua para la selección de características. Los autores pudieron obtener hasta un 81% de precisión en sus equipos de prueba. Sin embargo, el método mostró un mal desempeño con tres clases (“negativa”, “positiva” y “neutral”).

El análisis del sentimiento es un área en crecimiento del procesamiento lenguaje natural con investigación que va desde la clasificación a nivel de documento (Pang and Lee 2008) hasta el aprendizaje de la polaridad de palabras y frases (por ejemplo, (Esuli et al., 2006)). Dadas las limitaciones de caracteres en los tweets, la clasificación del sentimiento de los mensajes de Twitter es muy similar al análisis de sentimientos a nivel de oraciones (por ejemplo, (Yu y Hatzivassiloglou, 2003; Kim y Hovy, 2004)); Sin embargo, el lenguaje informal y especializado utilizado en tweets, así como la naturaleza misma del dominio de microblogging hace que el análisis del sentimiento de Twitter sea una tarea muy diferente. Características comunes al microblogging. (por ejemplo, los emoticones) también son comunes, pero ha habido poca investigación sobre la utilidad de los recursos de sentimientos existentes desarrollados en datos que no son de microblogging. Los investigadores también han comenzado a investigar varias formas de recoger automáticamente los datos de entrenamiento y adaptarlos a sus necesidades. Varios investigadores confían en los emoticones para definir sus datos de entrenamiento (Pak y Paroubek 2010) y explotar los sitios existentes de Twitter para recoger estos datos. (Davidov, Tsur y Rappoport 2010) también utilizan hashtags para crear datos de entrenamiento, pero limitan sus experimentos a la clasificación de sentimiento / no-sentimiento, en lugar de Clasificación de polaridad de 3 vías.

Inteligencia de negocios según (Costa y Souza, 2012) se compone de dos pasos. Primero, extraer, transformar y cargar datos desde fuentes no estructuradas (por ejemplo, sitios web de redes sociales, correos electrónicos) o estructurados (por ejemplo, ERP (enterprise resource planning), CRM (Customer relationship management)), lo que da como resultado un depósito de datos que incluye un repositorio de datos integrado, orientado a temas, variante de tiempo y no -volátil.

Las empresas deben tener una comprensión completa de las opiniones y necesidades de sus clientes sobre los productos o servicios que ofrecen, pero enfrentan el desafío de tratar con textos no estructurados desde la fuente de las opiniones y necesidades de los clientes. Los sentimientos de los productos y servicios de los consumidores como lo indica (Aalderks y Falls, 2012) ahora no solo son una fuente de comentarios y referencias de los clientes, sino una fuente de servicios al cliente, inteligencia comercial y gestión de la reputación de la marca del producto. Según lo indicado por (Grimes, 2008), algunos de los desafíos y necesidades que hacen que las organizaciones quieran responder a problemas fundamentales en la voz del cliente son:

- ¿Están los clientes satisfechos con los servicios, productos y soporte?
- Qué piensan los clientes de los productos y servicios que ofrecen los competidores
- Qué influye en el mercado y cómo se propagan las opiniones
- Qué les gusta a los clientes
- Qué problemas tienen los clientes
- Qué características adicionales le gustaría tener al cliente y cuáles están dispuestos a pagar.

Estos desafíos incluyen el manejo del ruido y la vinculación con datos estructurados (Venkata et al. 2009). La naturaleza ad-hoc y la documentación concisa de las técnicas adoptadas por los investigadores, que ha demostrado ser limitada en el sentido de que se adaptan laboriosamente a las complejidades de los datos subyacentes y los casos de uso previstos (Alberto et al. 2012).

La inteligencia empresarial descrita por (Venkata et al. 2009) implica el uso de tecnologías y metodologías para la recopilación, integración y análisis de información relevante en una empresa con el fin de tomar mejores decisiones en esa empresa. Los criterios para una verdadera inteligencia empresarial según lo identificado por Bogza (2008) son:

- Amplitud: funciones y tecnologías de acuerdo con una organización,

- Profundidad: Difundir en todos los niveles funcionales y organizacionales,
- Integridad: Plataforma de extremo a extremo,
- Analítica avanzada: Entregar información predictiva no solo retrospectiva,
- Calidad de los datos: garantizar que se utilizan los datos correctos.

Los estudios de la industria indicados por (Chen y Storey, 2012) muestran que la inteligencia de negocios se está convirtiendo cada vez más en un aspecto importante tanto en las comunidades empresariales como en las áreas académicas en las últimas décadas. IBM Tech Trends Reports (2011) identificó la analítica de negocios como una de las cuatro principales tendencias tecnológicas de más de 4,000 profesionales de la tecnología de la información de 93 países en la década de 2010. Las empresas, especialmente aquellas que ofrecen servicios y productos para consumidores, pueden querer saber las opiniones y sentimientos de sus consumidores y opiniones sobre sus productos y servicios para que puedan tomar decisiones estratégicas con el fin de atraer a más clientes, mantener clientes existentes o ventaja competitiva sobre los rivales. La información utilizada para analizar los sentimientos y las opiniones de los clientes se obtiene principalmente de fuentes como los clientes o usuarios que revisan los servicios y productos en Amazon (Aalderks and Falls, 2012). La inteligencia de negocios como un subtipo de análisis de sentimientos ha evolucionado con el advenimiento de las tecnologías en la web, desde la web 1.0, 2.0 hasta la 3.0 (Chen y Storey, 2012).

En base a los estudios ya realizados por otros autores e investigadores, se quiere realizar un sistema que permita el análisis de sentimientos de los usuarios, que interactúan mediante el envío de mensajes en una red social con el fin de generar reportes que permitan a las organizaciones conocer el comportamiento y lo que se está pensando sobre temas específicos, esto con el fin de analizar el nivel de satisfacción de los usuarios, ya sea de un producto de la empresa o de un tema sobre el que se esté trabajando, todo esto aplicado a la inteligencia de negocios que permitirá realizar el respectivo ETL (Extract, Transform and Load) y posteriormente el uso de herramientas de análisis de la información, que

respectivamente facilitarán a las empresas el uso adecuado de estos datos para mejorar los niveles de satisfacción de los usuarios.

De esta manera surge la idea de generar un sistema que se adapte a estas especificaciones para la red social Twitter, la cual tiene un gran uso por parte de las personas para generar calificaciones y envió de mensajes constantes, asimismo, con ayuda de la empresa patrocinadora de este trabajo, la cual posee conocimiento amplio y un desarrollo en el área de la inteligencia artificial, inteligencia de negocios y minería de datos, este sistema permitirá dar paso a una iniciativa mayor para las demás redes sociales en la implementación de un sistema de análisis de sentimientos más elaborado y con métricas o indicadores que permitan a la empresa, cliente o usuario, generar un análisis más adecuado con la información recolectada.

Actualmente existen herramientas con una cantidad amplia de funcionalidades, capaces de conectarse a una gran variedad de orígenes de datos, algunas de estas herramientas son: Tableau, Qlik, Power BI (Business Intelligence), entre otras herramientas. Cada una de estas tiene distintas ventajas y será necesario implementarlas según el giro de negocio. Asimismo, herramientas como Microsoft Flow, Microsoft Azure, Microsoft Bot Framework, Twitter Developer, son igualmente algunos medios que serán utilizados para crear el sistema.

1.2 Objetivo General

Desarrollar una solución para la visualización y análisis de sentimientos que los usuarios realizan hacia una entidad en base a los mensajes que envían mediante la red social Twitter en tiempo real.

1.3 Objetivos específicos

- Analizar la estructura tecnológica a implementar para conectarse con Twitter y extraer los datos necesarios para el sistema.
- Diseñar la estructura tecnológica para conectarse con Twitter y extraer los datos necesarios para los reportes.

- Diseñar la arquitectura de solución para la visualización y análisis de sentimiento con las distintas herramientas de Microsoft, que permitirán crear una inteligencia de negocios, a partir de los mensajes enviados por los usuarios en la red social Twitter.
- Implementar la arquitectura de solución para la visualización y análisis de sentimiento.
- Sincronizar las herramientas para que se pueda acceder a los datos capturados en tiempo real.
- Crear atributos derivados que permitan analizar la información importante dentro de los reportes.
- Analizar que componentes serán utilizados en cada reporte, para estructurar la información acorde a los requerimientos de la empresa.
- Seleccionar los atributos que compondrán cada componente dentro del tablero.
- Filtrar la información mostrada en los componentes del tablero para generar las métricas necesarias de la empresa.
- Construir los reportes con los componentes seleccionados, para verificar que se muestren correctamente.
- Sincronizar la información capturada en intervalos de 30 minutos, para que se vea reflejada en los distintos componentes del reporte.
- Facilitar a la empresa los resultados adquiridos, que les permitirán tomar las acciones necesarias para mejorar los estados de satisfacción de los usuarios.

1.4 Alcance

El desarrollo de este sistema permitirá a las empresas tener un conocimiento más efectivo de los niveles de satisfacción que tienen los usuarios en algún tema en específico y con esto desarrollar estrategias más eficientes, que les permitan aumentar positivamente estos niveles. De igual manera con el uso de la inteligencia de negocios se podrá generar un proceso de ETL, para luego con herramientas de BI generar los distintos tipos de tableros que permitan a la empresa estudiar el comportamiento de los usuarios por:

- Un Tweet o mensaje enviado a la red social.
- Porcentaje de aceptación de la empresa.
- Porcentaje de Tweets generados diariamente.
- Frecuencia en fechas de mensajes enviados
- Cantidad de usuarios que generaron mensajes por fecha

Cada indicador y reporte que se cree será en base a las necesidades de la empresa y lo más importante es que se realizaran en tiempo real, realizando actualizaciones dentro del sistema a intervalos de 30 minutos, sin embargo, estos intervalos se pueden modificar a necesidad de la empresa.

La elaboración de este sistema que utiliza inteligencia de negocios se realizará mediante el uso de varias herramientas de Microsoft como lo son Power BI, Microsoft Flow, Azure, Servicios Cognitivos, SQL Server, Twitter Developer, cada una de estas, aportará lo necesario para crear un sistema de calidad, que pueda ser implementado a largo plazo, en las distintas redes sociales existentes en el mercado.

Es importante resaltar, que este proyecto no tiene como alcance el ser implementado en las distintas redes sociales más conocidas en el mercado. Este sistema será implementado para la red social Twitter. Otro punto importante, es que este sistema no garantiza que los usuarios eleven sus niveles de satisfacción con la empresa, ya que se busca analizar el comportamiento de los usuarios por sus mensajes enviados y a partir de ello crear indicadores que faciliten a la empresa estudiar al usuario y crear metodologías que permitan aumentar esa satisfacción.

Se mostrarán las ventajas y desventajas de las herramientas, se expondrán reportes de los beneficios que tendrán las empresas al momento de utilizar el sistema, de manera que en un futuro se pueda implementar para comercializar el sistema a gran escala, a su vez los indicadores de ser necesario serán calculados de manera directa en las herramientas de BI, utilizando el lenguaje DAX (Data Analysis Expressions).

1.5 Justificación

El presente trabajo busca dar solución a los problemas que tienen las distintas empresas, en su relación con los clientes, entendiendo que, existe un gran porcentaje de ellas que están presentes en Twitter, así como usuarios que utilizan esta aplicación para conocer a las empresas y distintas marcas en el mercado. Aquellas empresas que no utilizan estos medios no establecen una relación adecuada con sus clientes o no conocen exactamente el nivel de satisfacción por parte de los usuarios, lo que puede ocasionar pérdidas monetarias y de reputación para estas marcas. Por ende, se busca solucionar estos problemas mediante la implementación de un sistema que analice estos comportamientos, para permitir a la empresa, crear métodos que aumenten los niveles de satisfacción.

Actualmente en Ecuador las empresas buscan superar a sus competencias de una manera eficaz, enfocándose en una adaptación constante de las estrategias más efectivas utilizadas en el mercado con el fin de atraer la mayor cantidad de clientes, por ello muchas empresas con el uso de redes sociales logran ampliar ese rango poblacional, sin embargo obvian uno de los aspectos más importantes en estos medios que es la comunicación constante y relacional con el usuario, lo que ocasiona un desinterés por las personas y un descenso de la reputación de la empresa, ya que muchos usuarios se rigen por estos indicadores en las redes sociales y sus calificaciones con comentarios positivos.

Mediante este proyecto se pretende brindar apoyo a las empresas que tienen un uso de la red social Twitter para que tengan un conocimiento de lo que sienten los usuarios sobre la empresa o algún tema trabajado, de igual manera esto permitirá reducir tiempos y dinero en la implementación de encuestas, reportes, para saber estos niveles de satisfacción y aceptación, los cuales analizan un rango poblacional considerablemente pequeño, mientras que con el sistema a desarrollar se estarían analizando grandes cantidades de datos sobre los usuarios y en muy corto tiempo.

Al final de este proyecto se pretende mostrar el desarrollo de una solución para analizar mediante inteligencia artificial e inteligencia de negocios, los niveles de satisfacción y aceptación de los usuarios a las empresas sobre temas específicos o productos de la empresa. Además, se pretende introducir a las empresas que no tengan conocimientos de estas herramientas, lo cual les permitirá facilitar su trabajo y desenvolvimiento con los usuarios.

2. Capítulo II: Marco Teórico

2.1 Introducción

En el presente capítulo, se realizará una explicación a detalle de cada una de las metodologías y herramientas a utilizar para implementar el proyecto. Se describirán las metodologías de desarrollo Scrum y minería de datos CRISPDM (Cross-Industry Standard Process for Data Mining), las cuales serán utilizadas para el desarrollo de este trabajo. De igual manera, se describirá el funcionamiento de las herramientas a utilizar y como se relacionan con el aplicativo final.

2.2 Metodologías

2.2.2 Metodología Scrum

Scrum es una metodología o marco de administración de proyectos ágil que se utiliza principalmente para proyectos de desarrollo de software con el objetivo de ofrecer nuevas capacidades de software cada 2 a 4 semanas. Es uno de los enfoques que influyeron en el Manifiesto Agile, que articula un conjunto de valores y principios para guiar las decisiones sobre cómo desarrollar software de mayor calidad más rápido (Collabnet, 2018). Con Scrum las buenas prácticas para trabajar en conjunto son necesarias, con el fin de mejorar los distintos resultados (Pérez, 2015).

Scrum cuenta con sprints, que tienen como objetivo obtener un entregable funcional que se presentará al cliente y este lo validará dando retroalimentación al equipo para el siguiente Sprint (Schwaber, 2016).

En Scrum existen roles como: El Equipo Scrum, el cual está conformado por

- Dueño de Producto (Product Owner).
- Equipo de Desarrollo (Development Team).
- Scrum Máster.

El Equipo Scrum busca tener creatividad, flexibilidad y aumentar la productividad del trabajo.

El Product Owner es el actor clave del proyecto, generalmente un cliente interno o externo, o un portavoz del cliente. Solo hay un Product Owner que transmite la misión y visión general del producto que el equipo está creando (Schwaber, 2016).

Por su parte, el equipo de desarrollo es un grupo multifuncional, autoorganizado, armado con todas las habilidades para proporcionar los entregables al completar cada sprint. Scrum amplía la definición del término "desarrollador" más allá de los programadores para incluir a cualquiera que participe en la creación del incremento entregado (Schwaber, 2016).

El Scrum Máster es el líder de servicio del Product Owner, el equipo de desarrollo y la organización. Sin autoridad jerárquica sobre el equipo, sino más bien un facilitador que garantiza al equipo se adhiera a la teoría, las prácticas y las reglas de Scrum. Además, es el encargado de ser el nexo entre el Equipo Scrum y las personas externas al equipo y lograr una comunicación eficaz entre todas las partes interesadas (Collabnet, 2018).

Las características más importantes de Scrum son la disminución del costo de implementación de un proyecto, cuando este se ha incrementado demasiado, la reducción de tiempos de entrega, solución de problemas complejos, facilidad de

adaptación del equipo de trabajo al estilo que plantea, creación de un ambiente laboral colaborativo, motivación del equipo de trabajo (Queralt, 2014).

2.2.2 Metodología CRISP-DM

CRISP-DM ayuda a organizar rutas de proyectos, resultados y anotaciones, de acuerdo con las fases de un proyecto normal de minería de datos. Puede realizar informes en cualquier momento durante el proyecto en función de las notas de rutas y fases de CRISP-DM (IBM, 2012).

Los objetivos de esta metodología son:

- Aplicar y comprender de mejor manera la Minería de Datos y sus resultados basándose en un proceso jerárquico.
- Dar cumplimiento de objetivos desde el punto de vista empresarial dando preferencia a la comprensión del negocio.
- Desarrollar proyectos de minería de datos mediante un proceso estandarizado.
- Minimizar los costos que implica un proyecto de minería de datos en las empresas

El modelo de CRISP-DM es flexible y se puede personalizar fácilmente. CRISP-DM permite crear un modelo global de minería de datos que se adapte a necesidades concretas de las organizaciones (IBM, 2012).

La metodología CRISP-DM, está dividida en 4 niveles de abstracción organizados de forma jerárquica, en tareas que van desde el nivel más general, hasta los casos más específicos y organiza el desarrollo de un proyecto de minería de datos, en una serie de seis fases, como se muestra en la Figura 1:

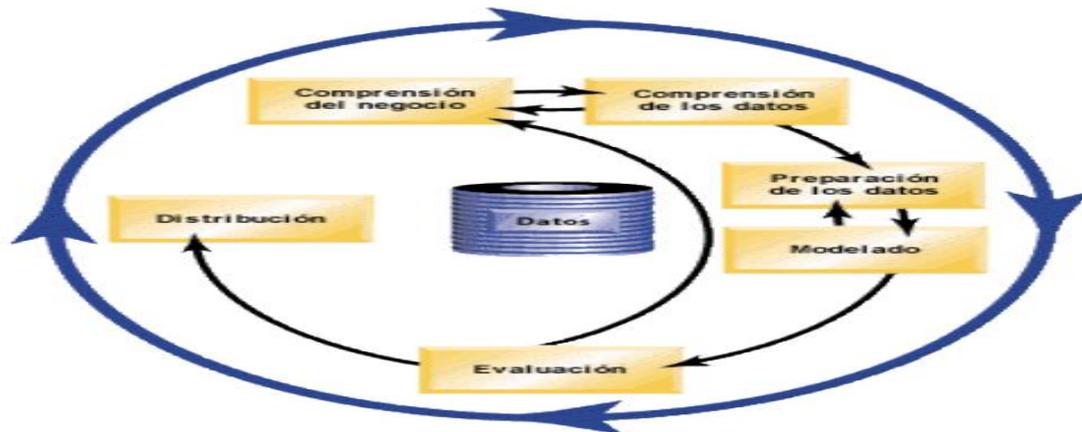


Figura 1. Ciclo de vida CRISP-DM.

Tomado de: (IBM, 2012).

La sucesión de fases no es necesariamente rígida. Cada fase es estructurada en varias tareas generales de segundo nivel y las tareas generales se proyectan a tareas específicas, donde finalmente se describen las acciones que deben ser desarrolladas para situaciones específicas (IBM, 2012).

Etapa 1: Comprensión del negocio.

La primera etapa del proceso CRISP-DM es comprender lo que desea lograr desde una perspectiva empresarial. Su organización puede tener objetivos y limitaciones en competencia que deben estar adecuadamente equilibrados. El objetivo de esta etapa del proceso es descubrir factores importantes que podrían influir en el resultado del proyecto. Ignorar este paso puede significar que se pone mucho esfuerzo en producir las respuestas correctas a las preguntas incorrectas (IBM, 2012).

Lo que se desea tener como resultado para este tipo de proyecto es:

- **Establecer objetivos:** esto significa describir el objetivo principal desde una perspectiva empresarial.
- **Producir un plan de proyecto:** se describirá el plan para lograr la extracción de datos y los objetivos comerciales. El plan debe especificar los pasos a realizar durante el resto del proyecto, incluida la selección inicial de herramientas y técnicas.

- **Criterios de éxito empresarial:** se presentarán los criterios a utilizar para determinar si el proyecto ha tenido éxito desde el punto de vista empresarial. Idealmente, estos deberían ser específicos y mensurables.

En la etapa 1 se debe tomar en cuenta ciertos aspectos que permitirán comprender el negocio de una manera eficaz:

Evaluar la situación actual.

Esto implica una búsqueda de datos más detallada sobre todos los recursos, restricciones, suposiciones y otros factores que se debe tener en cuenta al determinar el objetivo de análisis de datos y plan de proyecto.

- 1. Inventario de recursos:** enumerar los recursos disponibles para el proyecto, que incluyen:
 - a. Personal (expertos en negocios, expertos en datos, soporte técnico, expertos en minería de datos).
 - b. Datos (extractos fijos, acceso a datos en vivo, almacenados u operativos).
 - c. Recursos informáticos (plataformas hardware).
 - d. Software (herramientas de minería de datos, otro software relevante).
- 2. Requisitos, supuestos y restricciones:** enumerar todos los requisitos del proyecto, incluido el calendario de finalización, la comprensibilidad y la calidad de los resultados requeridos, y cualquier inquietud sobre la seguridad de los datos, así como cualquier problema legal. Hay que asegurar que se tiene permiso para utilizar los datos. Hacer una lista de las suposiciones hechas por el proyecto. Estas pueden ser suposiciones sobre los datos que pueden verificarse durante la extracción de datos, pero también pueden incluir suposiciones no verificables sobre el negocio relacionado con el proyecto. Listar las restricciones en el proyecto. Estas pueden ser restricciones en la disponibilidad de recursos, pero también pueden incluir limitaciones tecnológicas como el tamaño del conjunto de datos que es práctico utilizar para modelar.

- 3. Riesgos y contingencias:** enumere los riesgos o eventos que podrían retrasar el proyecto o hacer que falle. Enumere los planes de contingencia correspondientes.

Determinar los objetivos de minería de datos.

Un objetivo de negocio establece objetivos en la terminología de negocios. Un objetivo de minería de datos establece los objetivos del proyecto en términos técnicos.

- 1. Criterios de éxito empresarial :** describir los resultados previstos del proyecto que permiten el logro de los objetivos comerciales.
- 2. Criterios de éxito de la minería de datos :** definir los criterios para un resultado exitoso del proyecto en términos técnicos. Al igual que con los criterios de éxito empresarial , puede ser necesario describirlos en términos subjetivos, en cuyo caso se debe identificar a la persona o personas que realizan el juicio subjetivo.

Producir el plan de proyecto.

Describir el plan previsto para alcanzar los objetivos de minería de datos y, por lo tanto, lograr los objetivos comerciales. El plan debe especificar los pasos a realizar durante el resto del proyecto, incluida la selección inicial de herramientas y técnicas.

- 1. Plan de proyecto-** Enumerar las etapas que se ejecutarán en el proyecto, junto con su duración, recursos requeridos, entradas, salidas y dependencias. Cuando sea posible, haga explícitas las iteraciones a gran escala en el proceso de minería de datos. Como parte del plan del proyecto, también es importante analizar las dependencias entre el calendario y los riesgos. Marcar los resultados de estos análisis explícitamente en el plan del proyecto, idealmente con acciones y recomendaciones si los riesgos se manifiestan. Decidir en este punto qué estrategia de evaluación se utilizará en la fase de evaluación. El plan de proyecto será un documento dinámico. Al final de cada fase, se debe

revisar el progreso y los logros y actualizar el plan del proyecto en consecuencia. Los puntos de revisión específicos para estas actualizaciones deben ser parte del plan del proyecto.

2. **Evaluación inicial de herramientas y técnicas** : al final de la primera fase, debe realizar una evaluación inicial de herramientas y técnicas. Es importante evaluar las herramientas y técnicas al inicio del proceso, ya que la selección puede influir en todo el proyecto.

Etapa 2: Comprensión de los datos.

La segunda etapa del proceso CRISP-DM requiere que se adquieran los datos enumerados en los recursos del proyecto. Esta colección inicial incluye la carga de datos, si esto es necesario para la comprensión de los datos. Si se adquiere múltiples fuentes de datos, se debe considerar cómo y cuándo se integrarán (IBM, 2012).

- **Informe inicial de recopilación de datos** : enumerar las fuentes de datos adquiridas junto con sus ubicaciones, los métodos utilizados para adquirirlos y cualquier problema encontrado. Registrar los problemas que se hayan encontrado y cualquier resolución conseguida. Esto ayudará tanto a la replicación futura de este proyecto como a la ejecución de proyectos futuros similares.

Describir datos.

Examinar las propiedades de los datos adquiridos e informar sobre los resultados obtenidos.

- **Informe de descripción de datos** : describir los datos que se han adquirido, incluido su formato, su cantidad, las identidades de los campos y cualquier otra característica que se haya descubierto. Evaluar si los datos adquiridos satisfacen los requisitos.

Explorar datos.

Durante esta etapa, se debe abordar las preguntas de minería de datos utilizando técnicas de consulta, visualización de datos e informes. Estos pueden incluir:

- Distribución de atributos clave.
- Relaciones entre pares o pequeños números de atributos.
- Resultados de agregaciones simples.
- Propiedades de subpoblaciones significativas.
- Análisis estadísticos simples.

Estos análisis pueden abordar directamente los objetivos de minería de datos. También pueden contribuir o refinar la descripción de los datos y los informes de calidad, y alimentar la transformación y otros pasos de preparación de datos necesarios para un análisis más detallado.

- **Informe de exploración de datos** : describir los resultados de la exploración de datos, incluidos los primeros hallazgos o hipótesis iniciales y su impacto en el resto del proyecto. Si corresponde, se puede incluir gráficos y diagramas, para indicar las características de los datos que sugieren un examen más detallado de los subconjuntos de datos.

Verificar la calidad de los datos.

Examinar la calidad de los datos, abordando preguntas tales como:

- ¿Los datos están completos?
- ¿Es correcto o contiene errores y, si hay errores, qué tan comunes son?
- ¿Hay valores faltantes en los datos? Si es así, ¿cómo se representan, ¿dónde se producen y qué tan comunes son?

Informe de calidad de datos.

Listar los resultados de verificación para la calidad de los datos. Si existen problemas de calidad, sugerir posibles soluciones. Las soluciones a los

problemas de calidad de los datos generalmente dependen en gran medida tanto de los datos como del conocimiento empresarial.

Etapa 3: Preparación de los datos.

Esta es la etapa del proyecto en la que se decide los datos que se van a utilizar para el análisis. Los criterios que se pueden usar para tomar esta decisión incluyen la relevancia de los datos para la extracción de datos, la calidad de los datos y también las limitaciones técnicas, como los límites en el volumen de datos o los tipos de datos (IBM, 2012).

Limpieza de los datos.

Este paso implica mejorar la calidad de los datos al nivel requerido por las técnicas de análisis. Esto puede implicar la selección de subconjuntos limpios de los datos, la inserción de valores predeterminados adecuados o técnicas más ambiciosas, como la estimación de datos faltantes mediante el modelado.

- **Informe de limpieza de datos** : describir qué decisiones y acciones se tomarán para abordar los problemas de calidad de los datos. Considerar cualquier transformación de los datos realizados para fines de limpieza y su posible impacto en los resultados del análisis.

Construir los datos requeridos.

Este paso incluye operaciones constructivas de preparación de datos, como la producción de atributos derivados o registros nuevos completos, valores transformados para atributos existentes.

- **Atributos derivados** : estos son atributos nuevos que se construyen a partir de uno o más atributos existentes en el mismo registro, por ejemplo, puede usar las variables de longitud y ancho para calcular una nueva variable de área.
- **Registros generados** : aquí se describe la creación de registros completamente nuevos. Por ejemplo, es posible que necesite crear

registros para los clientes que no hicieron compras durante el año pasado. No había ninguna razón para tener tales registros en los datos sin procesar, pero para fines de modelado podría tener sentido representar explícitamente el hecho de que los clientes particulares no realizaron compras.

Integrar datos.

Estos son métodos mediante los cuales la información se combina desde múltiples bases de datos, tablas o registros para crear nuevos registros o valores.

- **Datos combinados** : fusionar tablas se refiere a unir dos o más tablas que tienen información diferente sobre los mismos objetos.
- **Agregaciones** : las agregaciones se refieren a las operaciones en las que se calculan nuevos valores al resumir la información de múltiples registros y / o tablas.

Etapas 4: Modelado.

En esta etapa seleccionar la técnica de modelado real que se utilizará. Aunque es posible que se seleccione una herramienta durante la fase de comprensión del negocio, en esta etapa se debe seleccionar la técnica de modelado específica. Si se aplican múltiples técnicas, realice esta tarea por separado para cada técnica (IBM, 2012).

- **Técnica de modelado** : documentar la técnica de modelado real que se utilizará.
- **Supuestos de modelado** : muchas técnicas de modelado hacen supuestos específicos sobre los datos. Registre los supuestos realizados.

Generar diseño de prueba.

Antes de crear un modelo, se debe generar un procedimiento o mecanismo para probar la calidad y la validez del modelo. Por ejemplo, en tareas de minería de

datos supervisadas, como la clasificación, es común usar tasas de error como medidas de calidad para los modelos de minería de datos. Por lo tanto, normalmente separar el conjunto de datos en conjuntos de pruebas y entrenamientos, construir el modelo en el conjunto de entrenamiento y estimar su calidad en el conjunto de pruebas por separado (IBM, 2012).

- **Diseño de prueba** : describir el plan previsto para la capacitación, prueba y evaluación de los modelos. Un componente principal del plan es determinar cómo dividir el conjunto de datos disponible en los conjuntos de datos de capacitación, prueba y validación.

Construir modelo.

Ejecutar la herramienta de modelado en el conjunto de datos preparado para crear uno o más modelos.

- **Configuración de parámetros** : Enumerar los parámetros y sus valores elegidos, junto con la justificación de la elección de la configuración de parámetros.
- **Modelos** : estos son los modelos reales producidos por la herramienta de modelado, no un informe sobre los modelos.
- **Descripciones de modelos** : describir los modelos resultantes, informar sobre la interpretación de los modelos y documentar cualquier dificultad encontrada con sus significados respectivos.

Modelo de evaluación.

En este paso, se debe clasificar los modelos y evaluarlos de acuerdo con los criterios de evaluación. Los objetivos de negocio y los criterios de éxito empresarial. En la mayoría de los proyectos de minería de datos, una sola técnica se aplica más de una vez y los resultados de la minería de datos se generan con varias técnicas diferentes (IBM, 2012).

- **Evaluación del modelo** : resumir los resultados de esta tarea, enumerar las cualidades de los modelos generados y clasificar la calidad en relación con los demás modelos.
- **Configuraciones de parámetros revisadas** : de acuerdo con la evaluación del modelo, revisar las configuraciones de parámetros y ajustarlas para la próxima ejecución de modelado. Enumerar la construcción y evaluación de modelos hasta que se verifique que se ha encontrado los mejores modelos. Documentar todas las revisiones y evaluaciones.

Etapa 5: Evaluación.

Durante este paso, se evaluará el grado de cumplimiento del modelo con los objetivos del negocio y se tratará de determinar si existe algún motivo comercial por el que este modelo sea deficiente. La fase de evaluación también implica evaluar cualquier otro resultado de minería de datos que se haya generado. Los resultados de la extracción de datos incluyen modelos que están necesariamente relacionados con los objetivos comerciales originales y todos los demás hallazgos que no están necesariamente relacionados con los objetivos originales del negocio, pero también pueden revelar desafíos adicionales, información o sugerencias para futuras direcciones (IBM, 2012).

- **Evaluación de los resultados de la extracción de datos** : resumir los resultados de la evaluación en términos de criterios de éxito empresarial, incluida una declaración final sobre si el proyecto ya cumple con los objetivos comerciales iniciales.
- **Modelos aprobados** : evaluar los modelos con respecto a los criterios de éxito empresarial, los modelos generados que cumplen con los criterios seleccionados se convierten en los modelos aprobados.

Proceso de revisión.

En este punto, los modelos resultantes parecen ser satisfactorios y permiten satisfacer las necesidades del negocio. Ahora es apropiado que se realice una

revisión más exhaustiva del trabajo de extracción de datos para determinar si hay algún factor o tarea importante que de alguna manera se haya pasado por alto (IBM, 2012).

- **Revisión del proceso** : resumir la revisión del proceso y resaltar las actividades que se han perdido y las que deberían repetirse.

Determinar los siguientes pasos.

Dependiendo de los resultados de la evaluación y de la revisión del proceso, ahora se decide cómo proceder. ¿Finaliza este proyecto y pasa a la implementación, inicia nuevas iteraciones o configura nuevos proyectos de minería de datos? También se debe hacer un balance de los recursos y presupuesto restantes, ya que esto puede influir en las decisiones (IBM, 2012).

- **Lista de acciones posibles** : enumerar las posibles acciones futuras, junto con las razones a favor y en contra de cada opción.
- **Decisión** : describir la decisión sobre cómo proceder, junto con la justificación.

Etapas 6: Implementación.

En la etapa de implementación, se tomará los resultados de la evaluación y se determinará una estrategia para la implementación. Si se ha identificado un procedimiento general para crear los modelos relevantes, este procedimiento se documenta en esta fase para su posterior despliegue. Tiene sentido considerar las formas y los medios de implementación durante la fase de comprensión del negocio, ya que la implementación es absolutamente crucial para el éxito del proyecto. Aquí es donde el análisis predictivo realmente ayuda a mejorar el aspecto operativo de su negocio (IBM, 2012).

- **Plan de implementación** : resumir la estrategia de implementación, incluidos los pasos necesarios y cómo realizarlos.

Plan de seguimiento y mantenimiento.

La supervisión y el mantenimiento son cuestiones importantes si el resultado de la extracción de datos se convierte en parte de la actividad diaria. La preparación cuidadosa de una estrategia de mantenimiento ayuda a evitar períodos innecesariamente largos de uso incorrecto de los resultados de la extracción de datos. Para monitorear el despliegue de los resultados de la extracción de datos, el proyecto necesita un plan detallado del proceso de monitoreo. Este plan tiene en cuenta el tipo específico de despliegue (IBM, 2012).

- **Plan de monitoreo y mantenimiento** : resuma la estrategia de monitoreo y mantenimiento, incluidos los pasos necesarios y cómo realizarlos.

Producir informe final.

Al final del proyecto, escribir un informe final. Dependiendo del plan de implementación, este informe puede ser solo un resumen del proyecto y sus experiencias (si aún no se han documentado como una actividad en curso) o puede ser una presentación final y completa de los resultados de la extracción de datos (IBM, 2012).

- **Informe final** : este es el informe final escrito del compromiso de minería de datos. Incluye todos los entregables previos, resumiendo y organizando los resultados.
- **Presentación final** : una reunión al final del proyecto en la que se presentan los resultados al cliente.

2.3 Historia de la Empresa

Grupo Business IT, es un grupo empresarial innovador, especialista en servicios informáticos de alta tecnología, con presencia en varios países de Latinoamérica (Grupo Business IT,2016)

MISIÓN

Ser el grupo empresarial más innovador de Latinoamérica en Tecnologías de Información, entregamos lo mejor de nosotros y así construimos una empresa sustentable donde nos apasiona trabajar (Grupo Business IT,2016)

VISIÓN

Ser reconocidos en el año 2019 como uno de los principales Grupos Tecnológicos de AMÉRICA LATINA, contando para ello con las empresas # 1 en los diferentes segmentos que atendemos (Grupo Business IT,2016)

Cuentan con un equipo de más de 100 profesionales de primer nivel para la ejecución de sus proyectos. Apoyan a sus clientes a cumplir sus metas empresariales a través de soluciones tecnológicas “Green IT” enmarcadas en la mejora continua de nuestro servicio. Green IT les permite a las organizaciones reducir sus costos operacionales, incrementar la eficiencia y mejorar la productividad a través de prácticas de IT sostenibles y amigables con el ambiente. Sus oficinas están ubicadas en Ecuador (Quito, Guayaquil y Loja) y Panamá (Grupo Business IT,2016).

2.4 Análisis de Sentimiento

El AS (Análisis de Sentimiento) o OM (Minería de Opiniones), es el estudio computacional de las opiniones, actitudes y emociones de las personas hacia una entidad. La entidad puede representar individuos, eventos o temas. Las dos expresiones AS o OM son intercambiables, expresan un significado mutuo. Sin embargo, algunos investigadores opinan que OM y AS tienen nociones ligeramente diferentes (Mikalai, 2012). Minería de Opiniones extrae y analiza la opinión de las personas acerca de una entidad, mientras que el análisis de sentimientos identifica el sentimiento expresado en un texto y luego lo analiza. Por lo tanto, el objetivo de AS es encontrar opiniones, identificar los sentimientos que expresan y luego clasificar su polaridad como se muestra en figura 2 (Themis Palpanas, 2012).

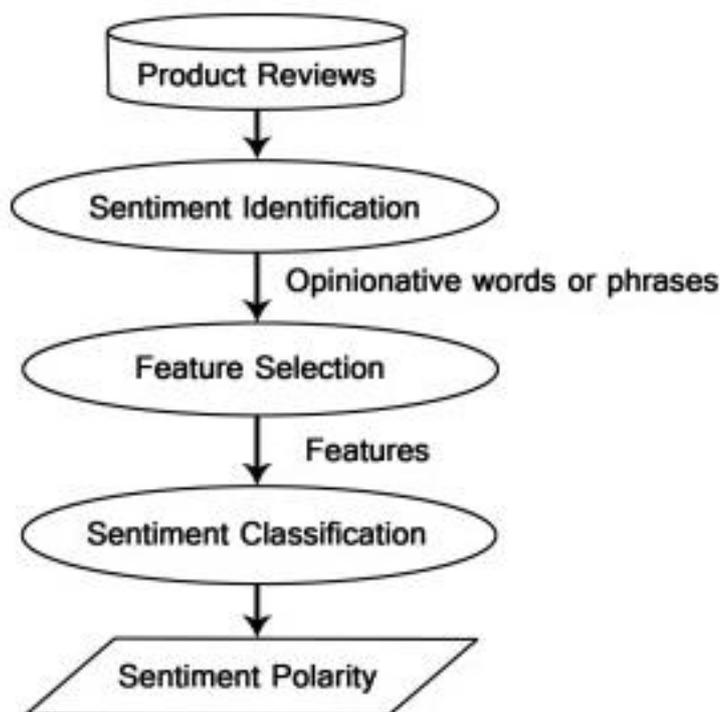


Figura 2. Proceso de análisis de sentimientos en revisiones de productos.

Tomado de: (Themis Palpanas, 2012).

El análisis de sentimientos puede considerarse un proceso de clasificación como se muestra en la figura 2. Hay tres niveles de clasificación principales en AS: nivel de documento, nivel de oración y nivel de aspecto. A nivel de documento AS tiene como objetivo clasificar un documento de opinión que exprese un sentimiento positivo o negativo. Considera todo el documento como una unidad de información básica. A nivel de oración AS apunta a clasificar el sentimiento expresado en cada oración. El primer paso es identificar si la oración es subjetiva u objetiva. Si la oración es subjetiva, AS determinará si la oración expresa opiniones positivas o negativas (Mikalai, 2012). Wilson et al. (2005) han señalado que las expresiones de sentimiento no son necesariamente de naturaleza subjetiva. Sin embargo, no hay una diferencia fundamental entre las clasificaciones a nivel de documento y de oración porque las oraciones son solo documentos breves (Liu, 2012).

La clasificación de texto en el nivel del documento o en el nivel de la oración no proporciona las opiniones necesarias sobre los detalles necesarios sobre todos

los aspectos de la entidad que se necesitan en muchas aplicaciones. Para obtener estos detalles; necesitamos ir al nivel de aspecto. El nivel de aspecto en AS tiene como objetivo clasificar el sentimiento con respecto a los aspectos específicos de las entidades. El primer paso es identificar las entidades y sus aspectos. Los poseedores de opinión pueden dar diferentes opiniones para diferentes aspectos de la misma entidad como esta frase "La calidad de voz de este teléfono no es buena, pero la duración de la batería es larga". Esta frase aborda los dos primeros tipos de AS.

Los conjuntos de datos utilizados en AS son un tema importante en este campo. Las principales fuentes de datos son las revisiones de productos. Estas revisiones son importantes para los propietarios de negocios, ya que pueden tomar decisiones comerciales de acuerdo con los resultados del análisis de las opiniones de los usuarios sobre sus productos (Yu et al., 2013). Las fuentes de reseñas son principalmente sitios de revisión y análisis de productos. La AS no solo se aplica en las revisiones de productos, sino que también se puede aplicar en mercados de valores (Hagenau et al., 2013), artículos de noticias, o debates políticos (Tao et al., 2012). En los debates políticos, por ejemplo, se podría averiguar las opiniones de la gente sobre ciertos candidatos electorales o partidos políticos. Los resultados de las elecciones también pueden predecirse a partir de opiniones y sentimientos políticos. Los sitios de redes sociales y los sitios de microblogging se consideran una muy buena fuente de información porque las personas comparten y discuten sus opiniones sobre un determinado tema libremente (Maks, 2012). También se utilizan como fuentes de datos en el proceso de AS.

La analítica de texto utiliza un algoritmo de clasificación de aprendizaje automático para generar un puntaje de sentimiento entre 0 y 1. Los puntajes más cercanos a 1 indican un sentimiento positivo, mientras que los puntajes cercanos a 0 indican un sentimiento negativo. El modelo está prediseñado con un extenso cuerpo de texto con asociaciones de sentimientos. Actualmente, no es posible proporcionar sus propios datos de entrenamiento. El modelo utiliza una combinación de técnicas durante el análisis del texto, incluido el procesamiento

del texto, el análisis de la parte del habla, la colocación de palabras y las asociaciones de palabras (Steen, 2018).

El análisis de sentimientos se realiza en todo el documento, en lugar de extraer el sentimiento para una entidad particular en el texto. En la práctica, existe una tendencia a mejorar la precisión de la puntuación cuando los documentos contienen una o dos oraciones en lugar de un gran bloque de texto (Steen, 2018).

Durante una fase de evaluación de la objetividad, el modelo determina si un documento en su conjunto es objetivo o contiene sentimiento. Un documento que se requiere evaluar y no avanza a la fase de detección de sentimiento, resulta en una puntuación de 0.50, sin procesamiento adicional. Para los documentos que continúan en proceso, la siguiente fase genera una puntuación por encima o por debajo de .50, según el grado de sentimiento detectado en el documento (Steen, 2018).

El análisis de sentimientos produce un resultado de mayor calidad cuando le das trozos de texto más pequeños para trabajar. Esto es opuesto a la extracción de frases clave, que funciona mejor en bloques de texto más grandes. Para obtener los mejores resultados de ambas operaciones, considere la posibilidad de reestructurar las entradas en consecuencia (Steen, 2018). No obstante, como menciona Heidi (Steen, 2018) *“el tamaño del documento debe ser inferior a 5.000 caracteres por documento, para permitir un análisis correcto y con un resultado por encima o por debajo de los 0.50”*.

Otro punto importante, es comprender como se realiza el análisis de sentimiento en el texto y para ello es importante construir características significativas para el clasificador de texto, y la lista de características incluye varias investigaciones de vanguardia:

- **N-grams:** denota todas las apariciones de n palabras consecutivas en el texto de entrada los elementos pueden ser fonemas, sílabas, letras, palabras o pares de bases según la aplicación. El valor preciso de n puede variar según los escenarios, pero

es común seleccionar $n = 2$ o $n = 3$. Con $n = 2$, para el texto "el zorro marrón rápido", se generarían los siguientes n-gramas: ["el rápido", "marrón rápido", "zorro marrón"] (Parimi, 2015).

- **El POS (etiquetado de parte voz):** es el proceso de asignar una parte de voz a cada palabra en el texto de entrada. También se calculan características basadas en la presencia de emoticonos, puntuación y mayúsculas (superior o inferior) (Parimi, 2015).
- **Las incrustaciones de palabras:** son un desarrollo reciente en el procesamiento del lenguaje natural, donde las palabras o frases que son sintácticamente similares se asignan juntas, por ejemplo. el término gato se mapearía más cerca del término perro, que al término coche, ya que tanto los perros como los gatos son animales. Las redes neuronales son una opción popular para construir este tipo de mapeo. Para el análisis de sentimientos, empleamos redes neuronales que también codifican la información de sentimientos asociada. Las capas de la red neuronal se utilizan como características para el clasificador (Parimi, 2015).

2.5 Programas y Herramientas

2.5.1 Power BI Desktop, Web y Mobile

Inteligencia de negocios o BI, por sus siglas en inglés, según el Data Warehouse Institute, lo define como la combinación de tecnología, herramientas y procesos que permite transformar los datos almacenados en información, esta información en conocimiento y este conocimiento dirigido a un plan o una estrategia comercial. La inteligencia de negocios debe ser parte de la estrategia empresarial, esta le permite optimizar la utilización de recursos, monitorear el cumplimiento de los objetivos de la empresa y la capacidad de tomar buenas decisiones para así obtener mejores resultados (Oracle, 2015).

Es importante visualizar de alguna forma que comprende una arquitectura de inteligencia de negocios. La figura 3 representa esta arquitectura. Si se analiza este diagrama de izquierda a derecha. Los primeros dibujos representan las

distintas fuentes de datos (Cubos esbase, bases de datos Oracle, Sql Server, mainframe, archivos planos, archivos xml, hojas de Excel, entre otros) que pueden utilizarse para extraer los datos de múltiples fuentes simultáneamente. El segundo dibujo representa el proceso de ETL. Este proceso donde se definen las fuentes heterogéneas y los campos que se van a utilizar, si necesitan algún tipo de modificación y/o transformación y donde se quiere ubicar estos datos, este proceso se le conoce como “mapping”. El tercer dibujo representa el repositorio de datos. En este repositorio se encuentran los datos transformados representados visualmente en modelos multidimensionales, dimensiones y tablas de datos. Existe un proceso entre el repositorio de datos y la interfaz de acceso al usuario, este es el motor de BI que me permite habilitar componentes, administrar consultas, monitorea procesos, cálculos, métricas. La interfaz de acceso a usuarios permite interactuar con los datos, representar de forma gráfica aquellos resultados de las consultas y los indicadores de gestión que fueron contruidos (Oracle, 2015).

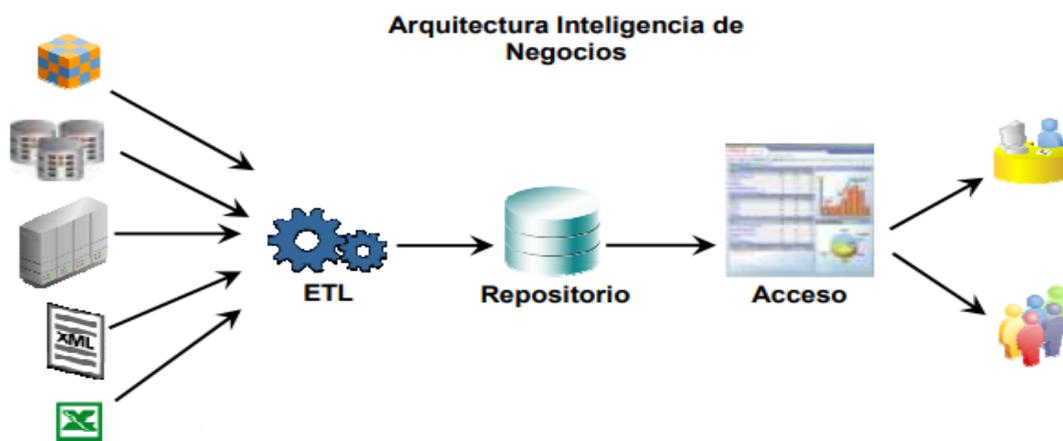


Figura 3. Arquitectura Inteligencia de Negocios.

Tomado de: (Oracle, 2015).

Esta herramienta es capaz de centralizar, agilizar y simplificar tareas que de otro modo serían interminables. Los procesos de diseño y creación de repositorios e informes de inteligencia empresarial requieren de herramientas de inteligencia de negocios para hacer que este trabajo realmente sea rentable. Es muy sencillo crear modelos de datos eficaces, junto con increíbles informes gracias a Power

BI Desktop. Su funcionamiento y capacidades son sumamente sencillos de adquirir (Makesoft, 2017).

Power BI podrá integrarse con las soluciones de Microsoft que ya se utilizan como Office 365, Sharepoint, Excel y Teams. De igual manera podrá conectarse con cientos de orígenes de datos para ver su información en paneles e informes en tiempo real, después podrá compartir su información con toda la organización para promover acciones inteligentes (PowerBI, 2018).

Otro aspecto importante de Power BI es que podrá acceder desde cualquier dispositivo con la aplicación nativa Power BI Mobile para anotar los informes fácilmente y mantener al equipo de trabajo informado en todo momento (PowerBI, 2018).

Para los analistas de los datos, Power BI permitirá ir del conocimiento a la acción rápidamente. La conexión a cientos de orígenes, preparación sencilla de los datos y creación de informes se realizará en cuestión de minutos (PowerBI, 2018).

Para profesionales de TI, la administración simplificada, cumplimiento y protección de los datos mientras se da acceso a los conocimientos que la gente necesita (Microsoft, 2018).

Para desarrolladores, las aplicaciones cobran vida a través de los datos. Presentación de informes atractivos, de alta fidelidad, en cualquier dispositivo, gracias a la fácil incrustación de objetos visuales de datos interactivos (Microsoft, 2018).

2.5.2 Microsoft Flow

Un flujo de trabajo representa tanto el modelado como la automatización de un proceso de negocio (Cotos, 2005). La Workflow Management Coalition (WFMC, 2004), define flujo de trabajo como *“la automatización de un proceso de negocio, en su totalidad o en parte, durante el cual documentos, información o tareas*

pasan de un participante a otro para actuar, de acuerdo a un conjunto de reglas de procesamiento”.

La automatización del proceso de negocio tiene dos partes: una parte estática y una parte dinámica. La parte estática viene representada por el modelo del flujo de trabajo obtenido a partir del proceso de negocio. La parte dinámica viene representada por las instancias del flujo de trabajo (Taboada, 2005).

Un diagrama de flujo de trabajo, como se muestra en la figura 4, brinda un panorama gráfico del proceso de negocio. Por medio de símbolos y formas estandarizados, el flujo de trabajo muestra de forma detallada cómo se completa un trabajo de principio a fin. También muestra quién es responsable del trabajo, en qué punto del proceso. Diseñar un flujo de trabajo implica llevar a cabo primero un análisis exhaustivo de dicho flujo, lo cual puede exponer debilidades potenciales. Un análisis del flujo de trabajo puede ayudarte a definir, estandarizar e identificar áreas críticas del proceso (Lucidchart, 2016).

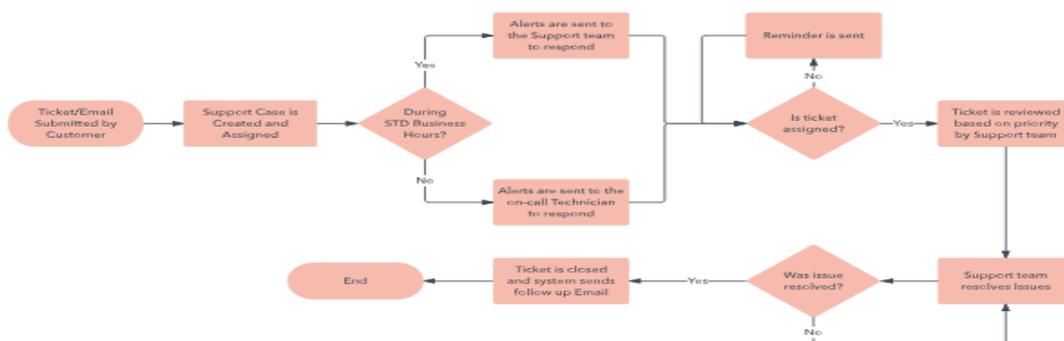


Figura 4. Diagrama de flujo de trabajo.

Tomado de: (Lucidchart, 2016).

Microsoft Flow es una herramienta que permite crear flujos de trabajo automatizados entre sus aplicaciones y servicios favoritos para obtener notificaciones, sincronizar archivos, recopilar datos y mucho más (Microsoft, 2017).

Microsoft Flow, como se muestra en la figura 5, se encuadra dentro de las aplicaciones PowerUser junto con PowerApps y PowerBI y las tres tienen en común que permiten crear distintos tipos de aplicaciones de forma rápida, sin necesidad de escribir código y son de fácil distribución (Hernández, 2016).

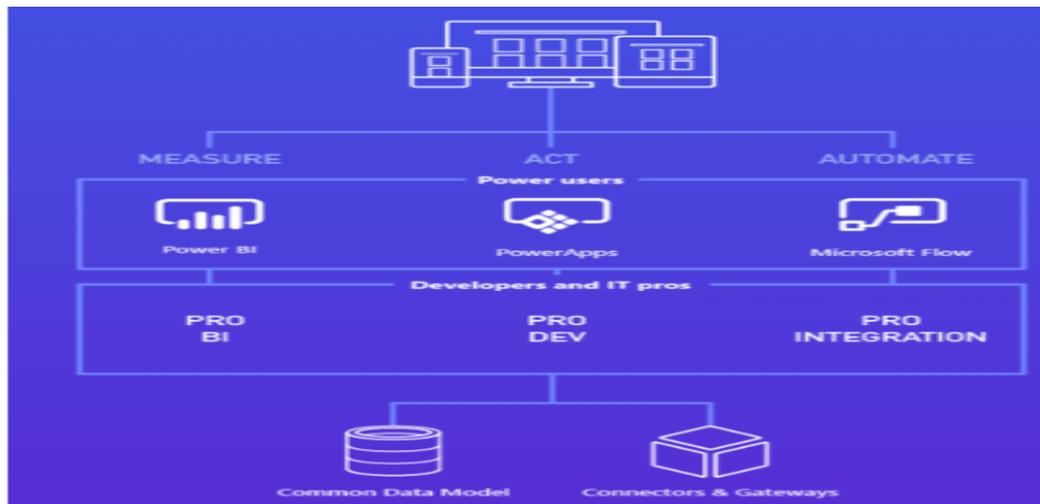


Figura 5. Arquitectura Microsoft Flow.

Tomado de:(Hernández, 2016).

Como sucede en PowerApps y PowerBI, Microsoft Flow no se queda atrás en cuanto a conectividad y permite conectar los flujos de trabajo casi con cualquier servicio o aplicación que podamos encontrar en el mercado. Desde redes sociales como Yammer o Twitter, con servidores de correo como Outlook o contra sistemas CRM como Dynamics 365 o Salesforce (Hernández, 2016).

2.5.3 Microsoft Azure

Para entender esta herramienta primero es importante conocer ciertos conceptos que permitirán comprenderla de manera eficaz. En esta tecnología se quiere explicar estos tres conceptos y cómo se pueden aplicar en la organización. Esta herramienta posee servicios apalancados en la nube y en las velocidades de internet, estos conceptos pueden visualizarse en la figura 6.

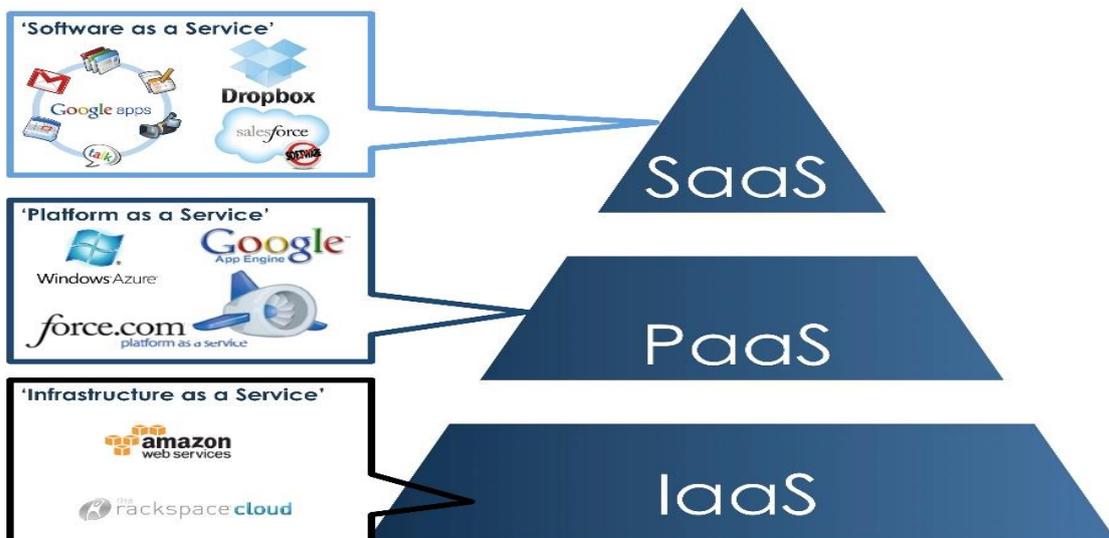


Figura 6. Servicios en la nube.

Tomado de:(Barabas, 2015).

Microsoft Azure, es un conjunto en constante expansión de servicios en la nube, para ayudar a la organización a satisfacer las necesidades comerciales. Le otorga la libertad de crear, administrar e implementar aplicaciones en una red mundial enorme con herramientas y marcos favoritos por los usuarios (Azure, 2018).

Azure, ofrece servicios que pueden ayudar a implementar y administrar las soluciones existentes, así como ofrecer soluciones personalizadas o listas. Además, puede encontrar una cartera de servicios experimentados (Azure, 2018).

Azure, es compatible con tecnologías de código abierto, por lo que puede usar las herramientas y tecnologías que prefiera. Ejecutar casi cualquier aplicación,

usando los orígenes de datos, los sistemas operativos y en los dispositivos compatibles (Azure, 2018).

Azure, es la única nube híbrida coherente, tiene más regiones que cualquier proveedor de nube, ofrece una productividad de desarrollo sin comparación y otorga una cobertura conforme a las normativas globales, incluido el cumplimiento de requisitos del RGPD (Reglamento general de protección de datos) (Azure, 2018).

2.5.4 Microsoft SQL Server

Un sistema de gestión de bases de datos permite a una persona organizar, almacenar y recuperar datos de una computadora. Es una forma de comunicarse con la "memoria almacenada" de una computadora (Foote, 2017).

Una base de datos, como una recopilación de información, se puede organizar para que un Sistema de administración de bases de datos pueda acceder y extraer información específica (Foote, 2017).

Los sistemas RDBMS (Sistema de gestión de bases de datos relacionales), eran una forma eficiente de almacenar y procesar datos estructurados. Luego, las velocidades de procesamiento se hicieron más rápidas y los datos "desestructurados" (arte, fotografías, música, etc.) se convirtieron en un lugar mucho más común. Los datos no estructurados, no son relacionales ni tienen esquemas, y los sistemas de gestión de bases de datos relacionales simplemente no fueron diseñados para manejar este tipo de datos (Foote, 2017).

Actualmente, las bases de datos están empezando a tomar una lógica aún más compleja. Otra característica que se está expandiendo, es el concepto de separación de ubicación del concepto abstracto de la base de datos en sí. Esta característica, permite que una base de datos resida en más de una ubicación y se consulte como una unidad continua. Dichas instancias se denominan bases de datos distribuidas o federadas (Berg, 2013).

SQL Server es un RDBMS de Microsoft que está diseñado para el entorno empresarial. SQL Server se ejecuta en T-SQL (Transact -SQL), un conjunto de extensiones de programación de Sybase y Microsoft que añaden varias características a SQL estándar, incluyendo control de transacciones, excepción y manejo de errores, procesamiento fila, así como variables declaradas (Rouse, 2015).

2.5.5 Microsoft Servicios Cognitivos.

Los servicios cognitivos permiten a las empresas que carecen del personal de investigación, la infraestructura y el presupuesto necesarios para la IA (Inteligencia Artificial), seleccionar herramientas de IA para sus necesidades (Diezma, 2017).

Los servicios cognitivos, pueden mejorar las habilidades de estos agentes para ayudar con lo siguiente:

- Escala para apoyar el crecimiento.
- Hacer recomendaciones inteligentes.
- Realizar traducción de idiomas.
- Utilice la visión artificial para reconocer a los usuarios a partir de imágenes y contenido moderado.

La incorporación de la inteligencia artificial en sitios web y aplicaciones es una tendencia poderosa y se está volviendo cada vez más prominente en estos días. Un ejemplo sería los chatbots que interactúan con los visitantes de las páginas web. Estos son agentes de software automatizados que pueden hablar en el lenguaje natural del usuario cuando él quiere preguntar algo, aprender sobre ventas y soporte. (Cabot, 2018).

Con los Servicios cognitivos de Microsoft, se pueden crear algoritmos potentes con líneas de código mínimas. Existen diferentes API (application programming interface) que ayudan a procesar imágenes a través de la visión por computadora, el habla, la extracción de conocimientos, la búsqueda web y el procesamiento de lenguaje natural. Estos brindan acceso a tecnologías

altamente actualizadas para incorporar tareas basadas en IA en plataformas, aplicaciones y servicios. (Noon, 2018).

Estos servicios cognitivos son API súper inteligentes que permitirían lanzar aplicaciones inteligentes a clientes potenciales. A través de esto, puede crear sistemas que puedan ver, escuchar, hablar y comprender a las personas en su propio lenguaje natural y utilizar el mismo método de comunicación para relacionarse con ellos. (Cabot, 2018).

2.6 Conclusión del Capítulo.

En conclusión, al utilizar las metodologías Scrum y CRISP-DM, se puede tener un proceso de desarrollo rápido, organizado y productivo. Además, la obtención de los datos se realizará de la manera correcta y eficaz. Es importante resaltar que las herramientas y servicios utilizados pertenecientes a Microsoft, son bastantes sólidas, lo que permite tener una alta disponibilidad y escalabilidad para atender las peticiones de los usuarios. Adicionalmente, es necesario mencionar que, a pesar de las limitaciones por licenciamiento de cada una de las herramientas, esto no impide el desarrollo de este proyecto, ya que cada una de estas licencias están siendo proporcionadas por la empresa patrocinadora.

3. Capítulo III: Desarrollo de Solución

3.1 Introducción.

En el presente capítulo, se detallará las metodologías implementadas para la obtención de los datos y el análisis de sentimientos, de igual manera se explicarán los pasos utilizados de la metodología CRISP-DM, que los utilizados en este proyecto son: La comprensión del negocio y de los datos, la preparación de los datos y su distribución, para conseguir un desarrollo adecuado del proyecto. Con la aplicación de esta metodología, se busca reconocer los distintos tipos de sentimientos encontrados en mensajes de texto enviados por los usuarios en la red social Twitter y los criterios que los clientes y/o empresas puedan necesitar para evaluar cada uno de los resultados.

3.2 Criterio de obtención de datos.

La obtención de datos es un paso muy importante en la minería de datos, ya que a partir de la información recolectada se generarán los análisis respectivos que se necesitan para evaluar el negocio. Es por lo que según IBM SPSS Modeler (2012), *“Se debe dedicar tiempo a explorar las expectativas de su organización con respecto a la minería de datos. Intente implicar a la mayor cantidad de personas que sea posible en estas discusiones y documente los resultados. Conozca las razones comerciales para que sus esfuerzos en minería de datos aseguren que todos los usuarios están de acuerdo antes de asignar recursos.”*

Para el presente trabajo, la información, requerimientos, necesidades y expectativas de la organización fueron proporcionadas por la empresa Grupo Business IT, la cual requería un sistema capaz de analizar el comportamiento de los usuarios ante sus productos y servicios ofrecidos a cada uno de sus clientes. Es por lo que mediante la red social Twitter se implementó un aplicativo capaz de capturar los tweets generados por un usuario a partir de una etiqueta, mención, nombre de usuario o palabra específica.

A partir de esta primera versión, surgió la necesidad de evaluar el comportamiento de los usuarios en otras áreas de negocio, donde la información podía ser recolectada masivamente por ser desencadenantes de viralizaciones. Los reportes que se analizarán vienen dados por tres ámbitos importantes, los cuales son:

- Un producto.
- Un evento importante a nivel mundial.
- Una organización o empresa.

Luego de analizar qué datos eran importantes recolectar, para generar las evaluaciones necesarias para la empresa, desde la red social Twitter se pudo obtener datos como: Texto del Tweet, Recuento de Favoritos, Recuento de Retweets, Recuento de Seguidores, Ubicación, Puntuación de Sentimiento, entre otros. Información que permite ser analizada en las herramientas de

inteligencia de negocios, para crear reportes que la empresa pueda utilizar para comprender al usuario.

3.3 Metodología para análisis de sentimientos.

La analítica de texto utiliza un algoritmo de clasificación de aprendizaje automático para generar un puntaje de sentimiento entre 0 y 1. Sin embargo, para este trabajo se crearon nuevas métricas para expandir el rango de sentimientos que pueden tener los usuarios, esto motivado a que en investigaciones y proyectos similares, se utilizaron 3 clasificadores comunes: negativo, positivo y neutral, para analizar que mensajes estaban en los distintos clasificadores, de igual manera se crearon rangos definidos para agruparlos según el puntaje de 0 a 1, estos rangos son:

- Valores comprendidos entre $0 > 0.4$, Negativo.
- Valores comprendidos entre $0.4 \geq 0.7$, Neutral.
- Valores comprendidos entre $0.7 > 1$, Positivo.

El análisis de sentimientos se realiza en todo el documento, en lugar de extraer el sentimiento para una entidad particular en el texto. En la práctica, para este trabajo se buscó seguir las recomendaciones que indican la teoría expuesta en capítulos anteriores donde existe una tendencia a mejorar la precisión de la puntuación cuando los documentos contienen una o dos oraciones en lugar de un gran bloque de texto.

Para el presente trabajo, muchos de los tweets que se analicen tendrán esta característica donde su puntaje sea de 0.50, ya que algunos mensajes vienen con información no estructurada, como lo son imágenes, enlaces a páginas webs. Sin embargo, aquellos mensajes que no pertenezcan a este grupo, se generará una puntuación por encima o por debajo de 0.50, según el grado de sentimiento detectado por el servicio cognitivo en el texto. No obstante, se busca realizar un filtrado de estos tweets con puntajes de 0.50, de manera que se pueda obtener los mensajes que permitan generar los reportes necesarios para las empresas.

El análisis de sentimientos produce un resultado de mayor calidad cuando le das trozos de texto más pequeños para trabajar. Es por lo que se utiliza la red social Twitter, debido a que esta limita los caracteres de sus mensajes a 240 por tweet, facilitando el análisis para el servicio cognitivo.

3.4 Análisis de Estructura Tecnológica

Para este trabajo se utilizó el Azure ML(Machine Learning) Text Analytics basado en los estudios y comparaciones realizadas por distintas fuentes, donde se denota la calidad de análisis de sentimientos del texto con respecto a otras herramientas. Esta herramienta de Microsoft tiene buen desempeño en los conjuntos de datos analizados, por lo que se quiere comparar con ofertas externas. Con la evaluación del desempeño frente a dos servicios externos: el motor de análisis de sentimientos NLP (Neuro-Linguistic Programming) de Stanford y una herramienta comercial popular. (Parimi, 2015). Aquí se muestran los puntos de referencia comparativos:

- En los conjuntos de datos que comprenden tweets, Azure ML Text Analytics fue un 10-20% mejor en la identificación de tweets con sentimientos positivos frente a los negativos. Se utilizaron datos de tweets de Sentiment140 y CrowdScale. Aquí está la comparación de los tres sistemas en el área bajo la curva ROC (Receiver operating characteristic):

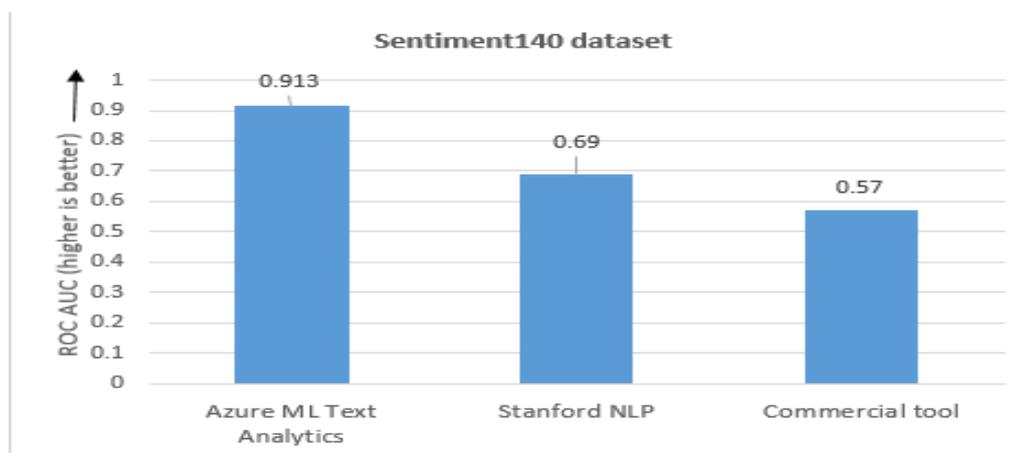


Figura 7. Comparación Sentimen140 dataset.

Tomado de: (Parimi, 2015).

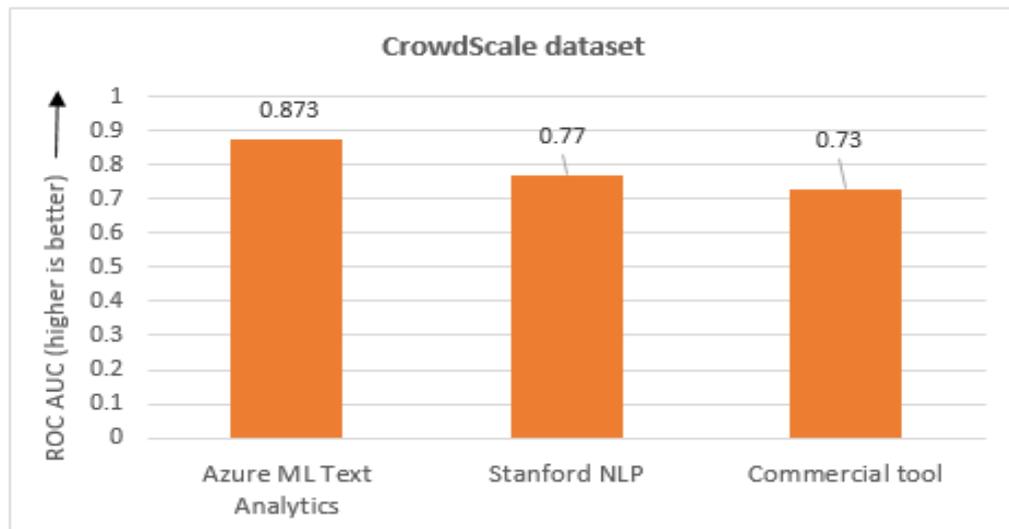


Figura 8. Comparación CrowdScale dataset.

Tomado de: (Parimi, 2015).

Como se puede observar en las Figura 7 y Figura 8, Azure Machine Learning Text Analytics supera a otras ofertas en formas de texto cortas y largas para la tarea de análisis de sentimientos. Este servicio también puede extraer frases clave, que denotan los principales puntos de conversación en el texto.

3.5 Conclusión del Capítulo.

En conclusión, la obtención de los datos fue uno de los factores más importantes para este trabajo, ya que, sin esta información, los siguientes pasos no se pueden realizar, limitándose la demostración y el análisis de sentimientos que se desea graficar en los distintos reportes. Adicionalmente, este análisis se construyó en base a lo investigado y según los parámetros más utilizados por los distintos investigadores, sin embargo, observando lo amplio y preciso de estos datos se elaboraron otras métricas, que serán detalladas en los capítulos de implementación y resultados, para evaluar el comportamiento de los usuarios ante los ámbitos descritos previamente.

4. Capítulo IV: Implementación

4.1 Introducción

A continuación, se detallará el diagrama de solución, configuración y desarrollo del aplicativo para cada una de las herramientas utilizadas. Además, se describirán los resultados obtenidos y como comprender cada uno de los reportes generados en Power BI.

4.2 Diagrama de Solución.

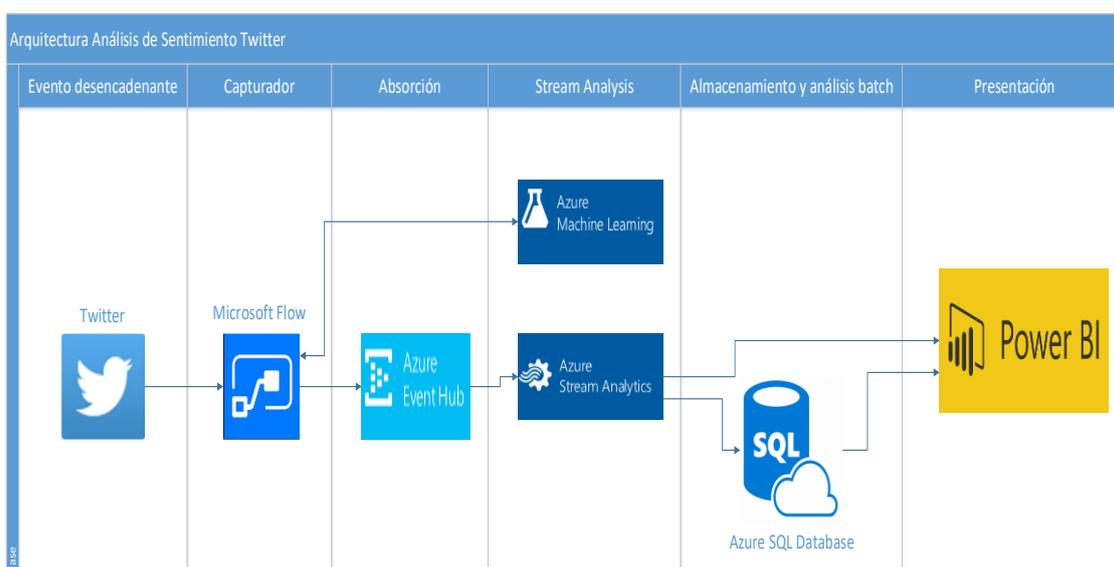


Figura 9. Arquitectura para el Análisis de sentimientos en Twitter.

La estructura del proyecto, como se muestra en la figura 9, consiste en el desarrollo y la implementación de un flujo de trabajo en Microsoft Flow, que permite la captura de los tweets generados por un usuario, según sea el parámetro de captura: Etiqueta, Nombre de Usuario, Palabras Específicas. Para el presente trabajo se implementaron 3 flujos de trabajo que permiten la captura de los tres ámbitos que se desean analizar: Productos, Eventos y Organizaciones/Empresas. No obstante, este trabajo tenía las intenciones de ser implementado para la empresa Grupo Business IT, que proveyó las licencias y herramientas necesarios para la elaboración del proyecto, sin embargo, al realizar las capturas de los datos se notó que la información no sería suficiente para analizarla, por ende, se reemplazó el estudio de la empresa por los casos

descritos anteriormente, los cuales son desencadenantes de información que permitirán un análisis correcto por la herramienta de inteligencia de negocios

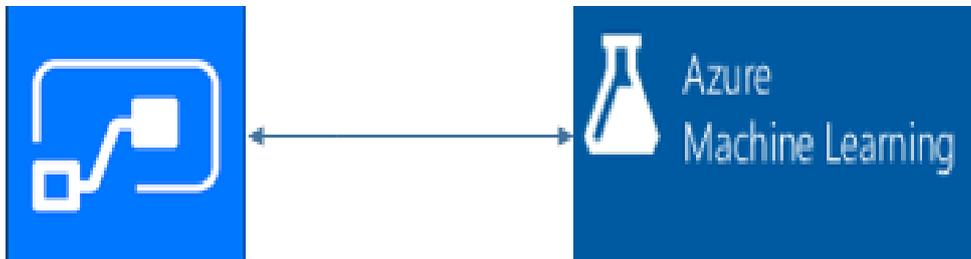


Figura 10. Conexión de MS Flow y Azure Machine Learning.

Como se muestra en la Figura 9, desde el Twitter se genera un tweet que es capturado por MS Flow. El MS Flow se conecta al servicio cognitivo de Azure para el análisis de texto y se obtiene el puntaje luego de analizar el tweet generado por el usuario, luego se envía la información nuevamente al MS Flow como se muestra en la Figura 10.



Figura 11. Conexión de MS Flow con Azure Event Hub y Azure Stream Analytics.

Como se muestra en la figura 11, el aplicativo captura la información en tiempo real y de manera masiva, ya que en Twitter se generan millones de mensajes diariamente, por ello es necesario que los eventos generados sean ingeridos lo más rápido posible y en el menor tiempo, con el uso del Event Hub de Azure es posible este requerimiento, ya que como indica Microsoft (Azure,2018) “*Event Hubs es un servicio de ingesta de datos en tiempo real totalmente administrado que es sencillo, confiable y escalable. Hacer streaming de millones de eventos por segundo desde cualquier origen para crear canalizaciones de datos dinámicos y responder a los desafíos empresariales de inmediato*”. Y de esta

manera, pasar esta información por el siguiente servicio de Azure Stream Analytics, el cual permite un procesamiento de datos para un sinfín de eventos y que proporciona el análisis en tiempo real.



Figura 12. Conexión de Azure Stream Analytics con SQL Database y Power BI.

A continuación, se debe guardar esta información en una base de datos que será consumida por el Power BI para realizar los reportes necesarios por la empresa para analizar el comportamiento de los usuarios. Para ello, como se ilustra en la Figura 12, la información que se capturó analizó y estructuró, es enviada a una base de datos en la nube para proporcionar disponibilidad en todo momento. Por último, esta información es consumida por el Power BI y estructurada en los distintos índices que permitirán analizar a detalle cada uno de los datos obtenidos previamente, estos índices son construidos a partir de las columnas estructuradas en la base de datos y las creadas en la herramienta por lenguaje DAX para encontrar información importante que se puede mostrar más organizada en cada uno de los reportes.

4.3 Configuración y Creación de Servicios en Azure.

Uno de los pasos más importantes para el desarrollo de este proyecto, es la configuración de cada una de las herramientas utilizadas, por lo que lo primordial es conocer cómo realizarlo y entender a detalle cada uno de los programas.

Primero, se necesita tener una cuenta personal Outlook o Hotmail, con la que se podrá registrar en los distintos ambientes. Teniendo una cuenta personal de correo electrónico, se procede a configurar la primera herramienta Azure.

Para ello debe dirigirse al portal de Azure y registrarse con el correo creado.(Figura 13)

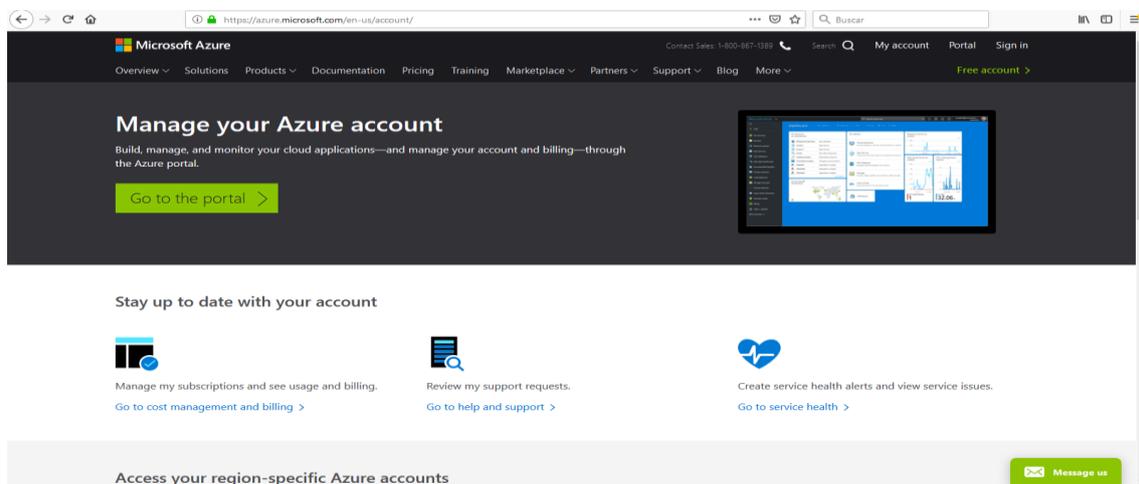


Figura 13. Portal de Azure

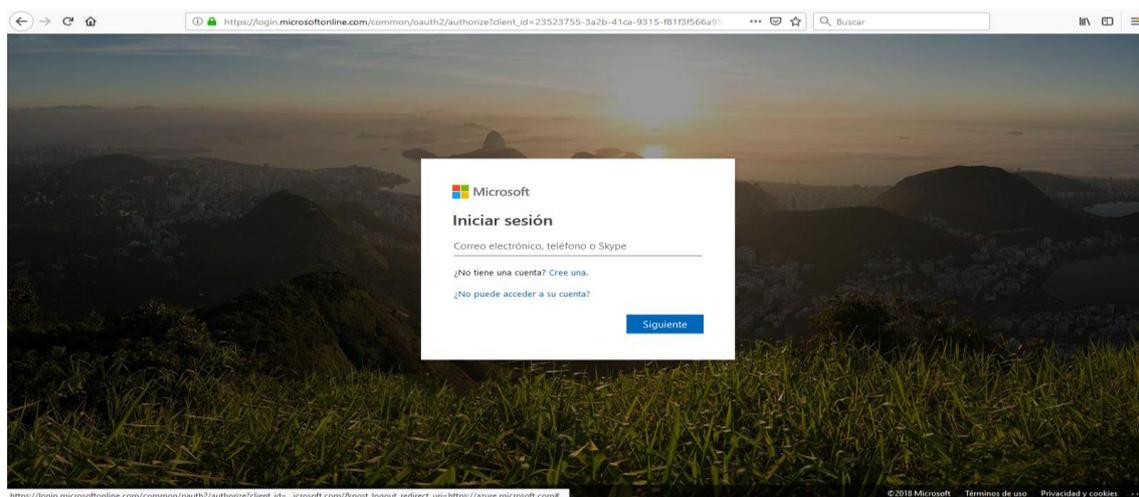


Figura 14. Portal de Inicio de Sesión Office 365.

Como se muestra en la figura 14, con el fin de unificar todas las herramientas y simplificar el tema de los correos electrónicos, Microsoft a través de Office 365 permitió a sus usuarios tener todas las herramientas asociadas a una misma dirección de correo electrónico, facilitando el tema de inicio de sesión y para mantener la información en todo lugar con cualquier dispositivo conectado a Office 365.

Luego de iniciar sesión, se redirecciona al portal de Azure donde por defecto posee una licencia gratuita o de prueba y para poder consumir ciertos servicios necesarios en este proyecto, es importante que agregue crédito a su cuenta de Azure por algún medio de pago u obteniendo una suscripción **Visual Studio Enterprise – MPN (Microsoft Partner Network)**.

La ventaja de una suscripción **Visual Studio Enterprise – MPN**, es que adicionalmente de poseer un crédito mensual por 150\$, tendrá otros beneficios para las distintas herramientas de Microsoft, como se muestra en la figura 15.

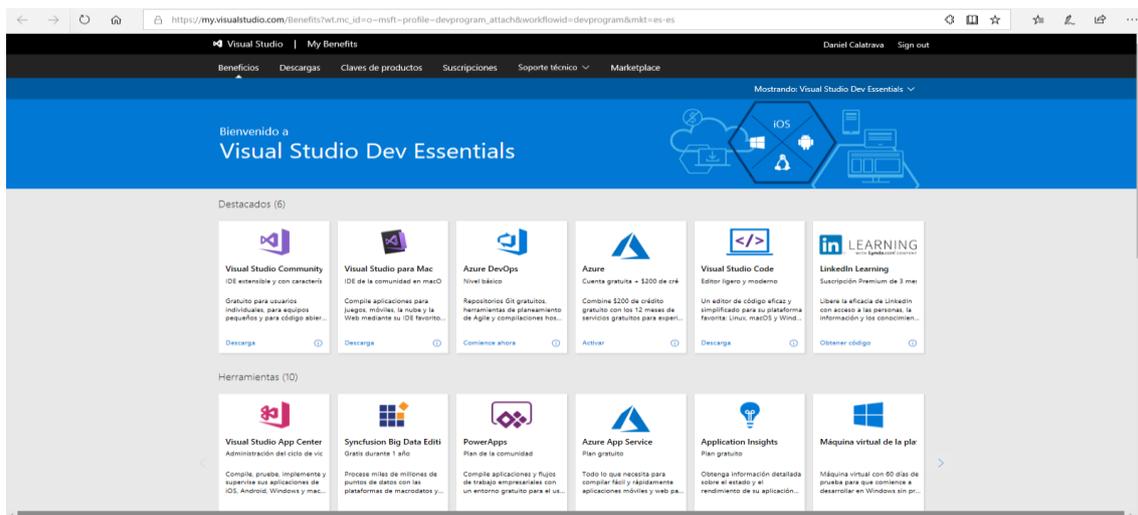


Figura 15. Visual Studio Dev Essentials

En el portal de Azure se necesita crear dos servicios que son muy importantes para el desarrollo del proyecto, el Servicio Cognitivo de Análisis de Texto y el servicio de base de datos SQL, como se muestra en la figura 16.

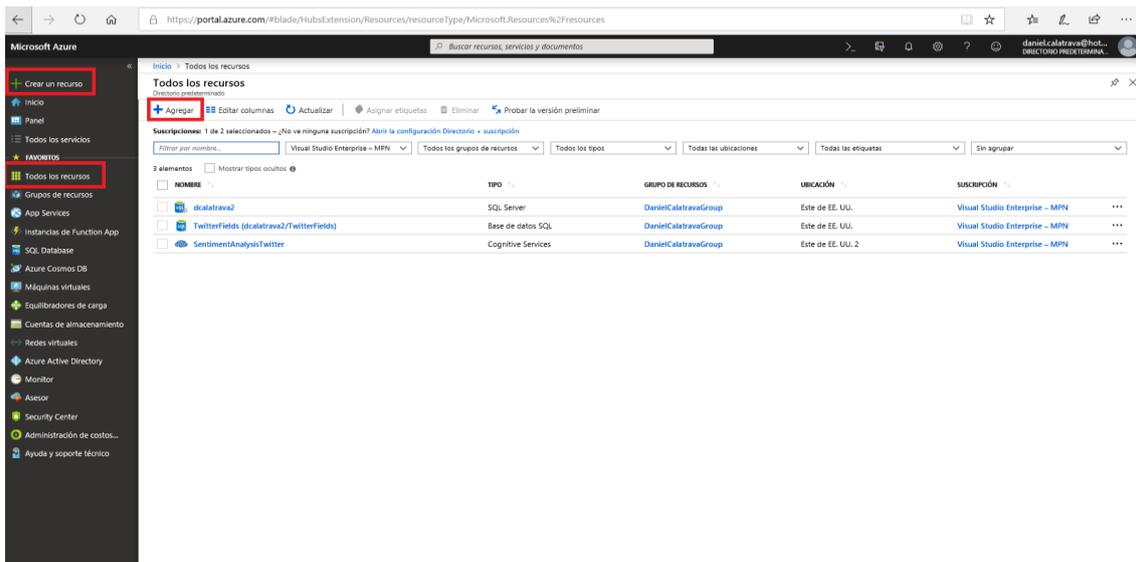


Figura 16. Portal de Azure – Recursos y Servicios.

En esta página, como se muestra en la figura 17, se presentan varias maneras de crear estos recursos, uno de ellos es dentro del menú lateral derecho la opción (+ Crear Recurso), otra es seleccionar la opción “Todos los Recursos” y dar clic en Agregar. Lo que abrirá una ventana para ubicar el servicio requerido dentro de Azure.

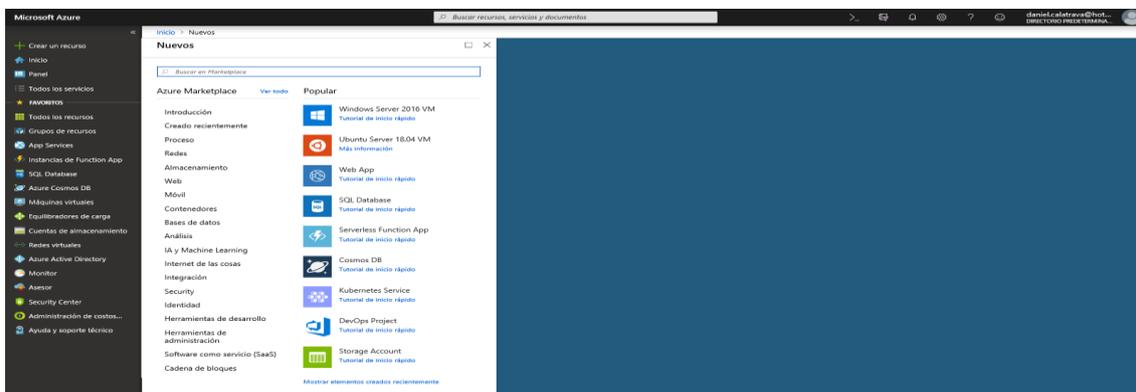


Figura 17. Portal de Azure – Creación de recursos o servicios.

Se procede a crear los servicios mencionados anteriormente y como se muestra en la figura 18.

- Text Analytics del Cognitive Services
- SQL Database

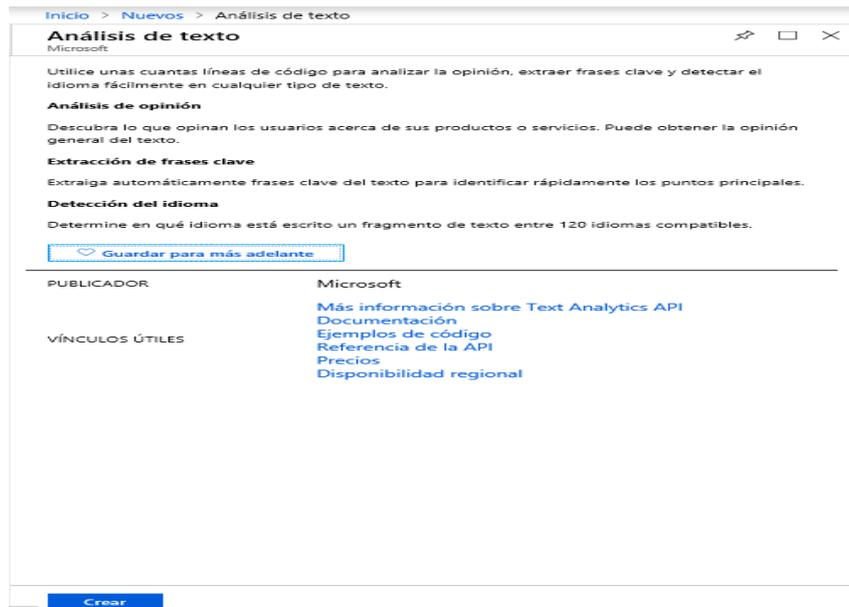


Figura 18. Servicio de Análisis de texto Azure.

Al momento de crear el servicio, le solicitará ciertos datos necesarios para su configuración, entre ellos:

- **Nombre:** ingresar un nombre que sea referencial.
- **Suscripción:** colocar el tipo de suscripción que usted posee, si no tiene deberá añadir un método de pago para que se le adjudique, si se ingresa una tarjeta de crédito, la suscripción que se creara es "Pay as you Go".
- **Ubicación:** se recomienda colocar Este U Oeste de EE. UU., ya que son los servidores más económicos y más estables para Latinoamérica.
- **Plan de tarifa:** se recomienda verificar la calculadora de Azure, la cual según el servicio le indicará, el precio a consumir y las transacciones diarias, mensuales, que se podrán consultar. Por defecto para este proyecto se utilizó el plan F0, un plan gratuito que maneja 5000 llamadas mensuales y es un servicio por un mes.
- **Grupo de Recursos:** este es un grupo donde se añaden los servicios que se utilizan en el proyecto y con este medio se puede organizar en caso de consumir otros servicios en Azure que no pertenezcan al proyecto, si no posee se recomienda crear uno nuevo donde puede organizar los servicios.

Posteriormente, se redirige a una ventana mostrada en la figura 19, donde podrá observar todas las métricas necesarias del servicio y donde se obtendrá las contraseñas y la URL necesaria para una configuración dentro del Microsoft Flow.

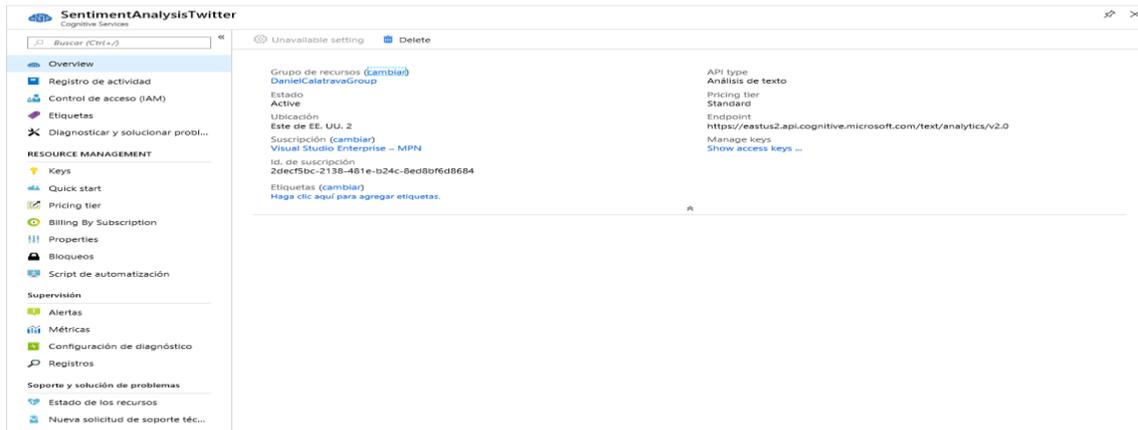


Figura 19. Servicio de Análisis de Texto – Panel de configuración.

Como se muestra en la figura 20, para el servicio de base de datos SQL, se debe seguir los mismos pasos anteriores y tener en cuenta que el plan de tarifa debe adecuarse a las características necesarias por el cliente o empresa, ya que, a mayor capacidad, rendimiento y velocidad, mayor es el costo.

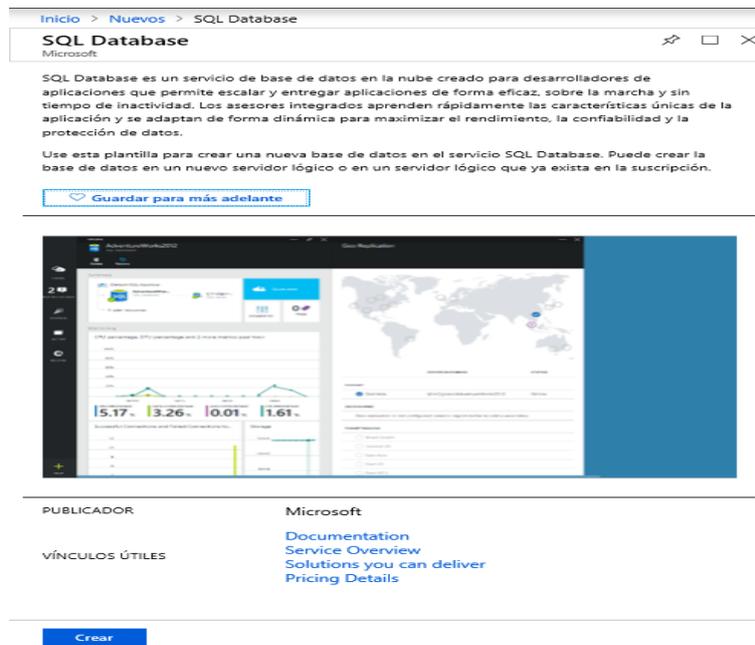


Figura 20. Servicio de Base de Datos SQL.

A continuación, como se muestra en la figura 21, se verifica que la conexión a la base de datos se pueda realizar desde el programa de SQL Server 2017 o versiones anteriores, para validar que la configuración del servicio en Azure fue correcta. Para ello colocar el nombre del servidor, el usuario debe ser el creado durante la configuración del servicio, se selecciona el tipo de autenticación “**SQL Server Authentication**”.

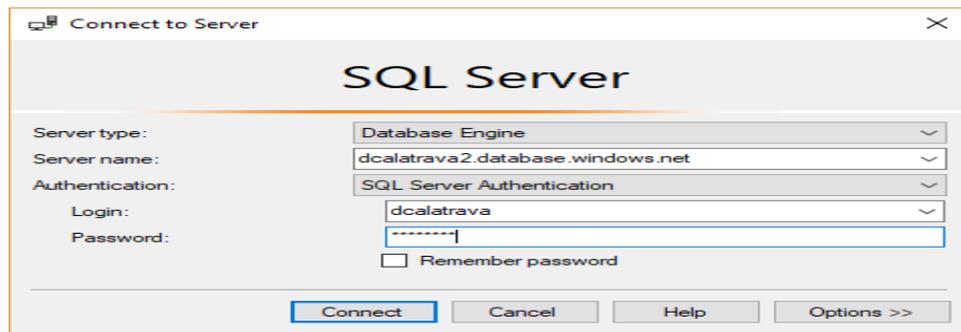


Figura 21. Conexión BD SQL en SQL Management Studio.

Al verificar que la conexión es exitosa, se puede validar que la base de datos que se creó aparezca y como se muestra en la figura 22, se crean las tablas que se necesitaran en el proyecto para cada uno de los ámbitos a evaluar o analizar. Para este proyecto se crearon 3 tablas:

- Evento Boca Vs River
- Organización o Empresa – Eléctrica de Quito
- Producto – Coca Cola

Cada una de las columnas que se necesitan en las tablas se añadirán en otro apartado, motivado a que se debe obtener las columnas del servicio Twitter API para seleccionar las necesarias en cada tabla.

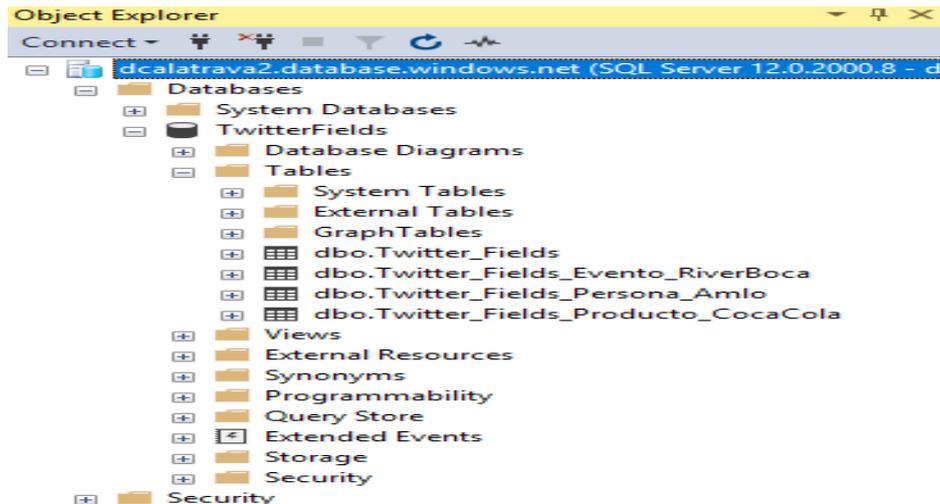


Figura 22. Base de Datos con tablas.

4.4 Configuración y Creación de flujos de trabajo en Microsoft Flow.

El Microsoft Flow, es la herramienta que va a permitir capturar los mensajes de Twitter mediante una etiqueta, nombre de usuario o frase/palabras, de igual manera a través de una acción en concreta se podrá consumir el servicio, previamente configurado para el análisis de texto en Azure, por otra parte, esta herramienta de Microsoft permite visualizar las columnas que se obtienen del Twitter API, facilitando la lectura de los datos y con lo que se podrá construir cada tabla dentro del SQL Server.

Prosiguiendo con lo explicado, se debe crear un nuevo flujo y para ello, luego de iniciar sesión, es necesario dirigirse a “**Mis Flujos**” y se redirecciona a una página con los flujos que se han implementado, luego seleccionar la opción “**Nueva**” y “**Crear desde cero**”, como se muestra en la figura 23.

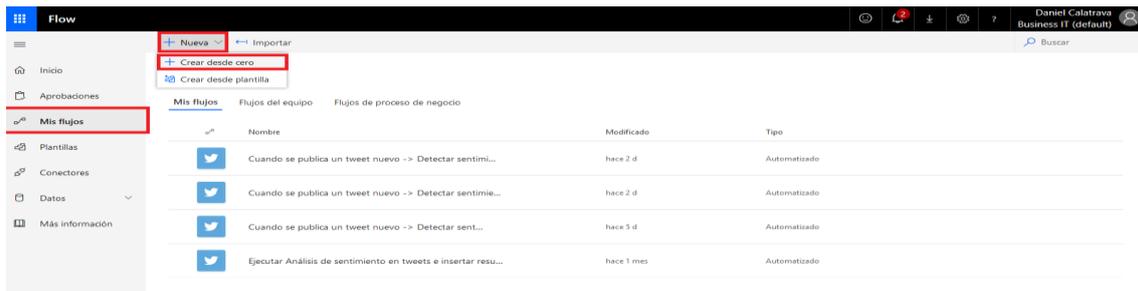


Figura 23. Microsoft Flow, Cartera de Flujos.

Para el presente trabajo se crearon 3 flujos para cada ámbito a evaluar mencionados anteriormente, cada uno de ellos con las configuraciones necesarias y los datos más importantes a recolectar, que permitirán la creación de reportes más precisos. Los parámetros a utilizar para la captura de los datos no fueron obtenidos por miembros de las empresas o eventos a evaluar, sino de un análisis realizado por el desarrollador del trabajo, para evidenciar los beneficios de las herramientas y probar su correcto funcionamiento. Esto con el fin de presentar los distintos reportes a la empresa patrocinadora y que analice la posibilidad de financiar el proyecto para su implementación y oferta en las distintas empresas del país

Lo primero a realizar es buscar el desencadenador del flujo, para este proyecto es la creación de un nuevo tweet, posterior se debe especificar en base a que termino de búsqueda se desea capturar el tweet. como se muestra en las figuras 24 y 25.



Figura 24. Microsoft Flow – Acción de desencadenamiento de flujo.

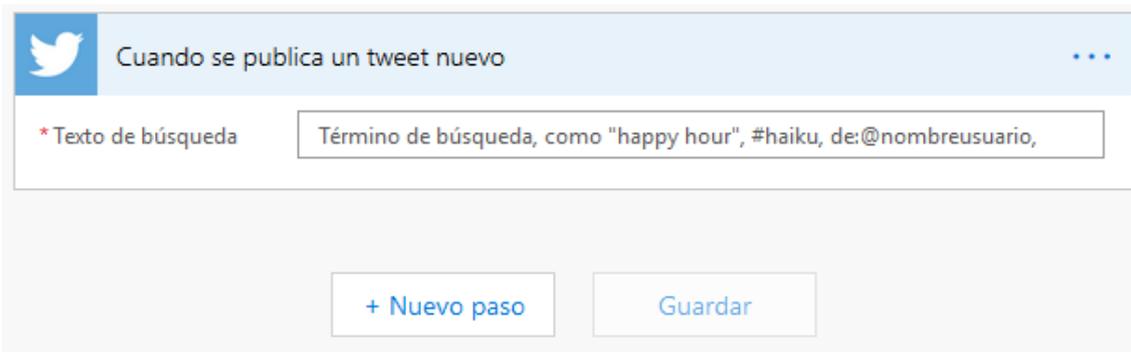


Figura 25. Microsoft Flow – Configuración de acción.

Para el proyecto se realizó a partir de etiquetas o hashtags, lo importante es ubicar cuales son los más comunes o más utilizados por los usuarios, de igual manera se puede configurar para frases o nombres de usuarios, con el fin de capturar la mayor información posible. Para uno de los flujos creados en este trabajo, se puede apreciar en la Figura 26, se utilizaron una variedad de etiquetas y el usuario respectivo. Las configuraciones se pueden realizar con los distintos condicionales OR y AND.

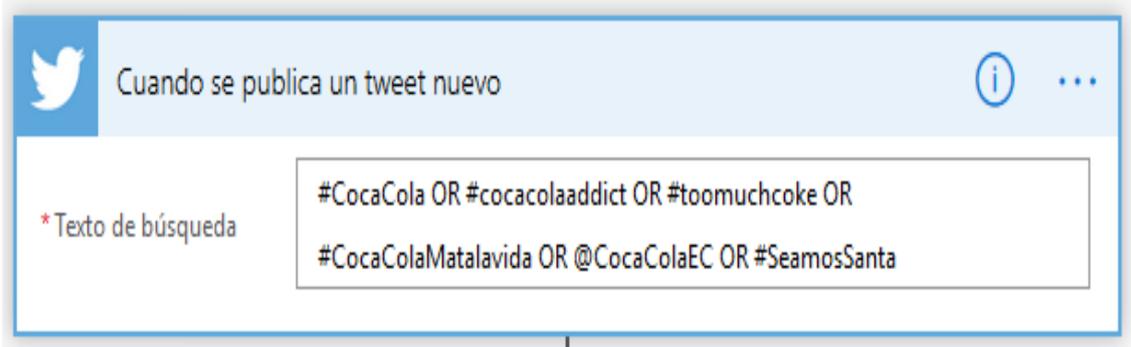


Figura 26. Microsoft Flow – Configuración de acción en producto Coca-Cola.

Continuando con el flujo, se debe crear la acción o nuevo paso para consumir el servicio cognitivo de análisis de texto que se creó en Azure y para ello se va a necesitar ciertos datos que se ubican dentro del panel de configuración del respectivo servicio y que se puede consultar desde el portal de Azure, como se muestra en la figura 27 y 28.

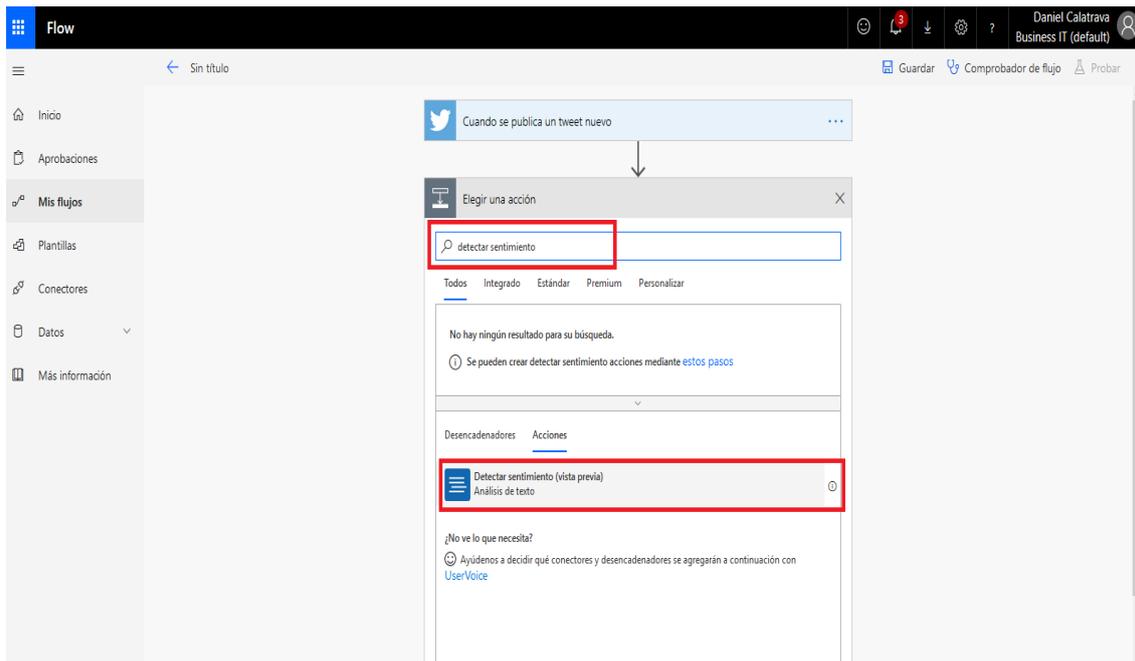


Figura 27. Microsoft Flow – Acción detectar sentimiento.

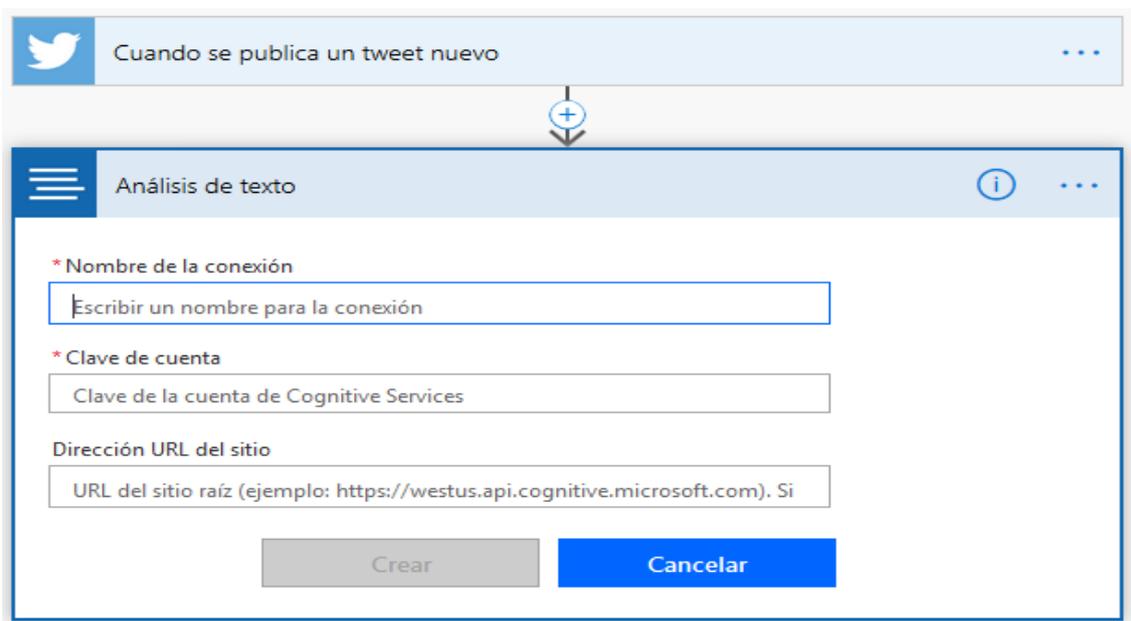


Figura 28. Microsoft Flow – Configuración de acción detectar sentimiento.

Dentro del portal de Azure, dentro del servicio de análisis de texto, se ubica la opción "Overview", en esta opción se podrá visualizar la "Dirección URL del sitio" que se consume para este servicio, de igual manera se puede acceder a las

claves de cuenta y el nombre de usuario en la opción “Manage keys” como se muestra en la figura 29.

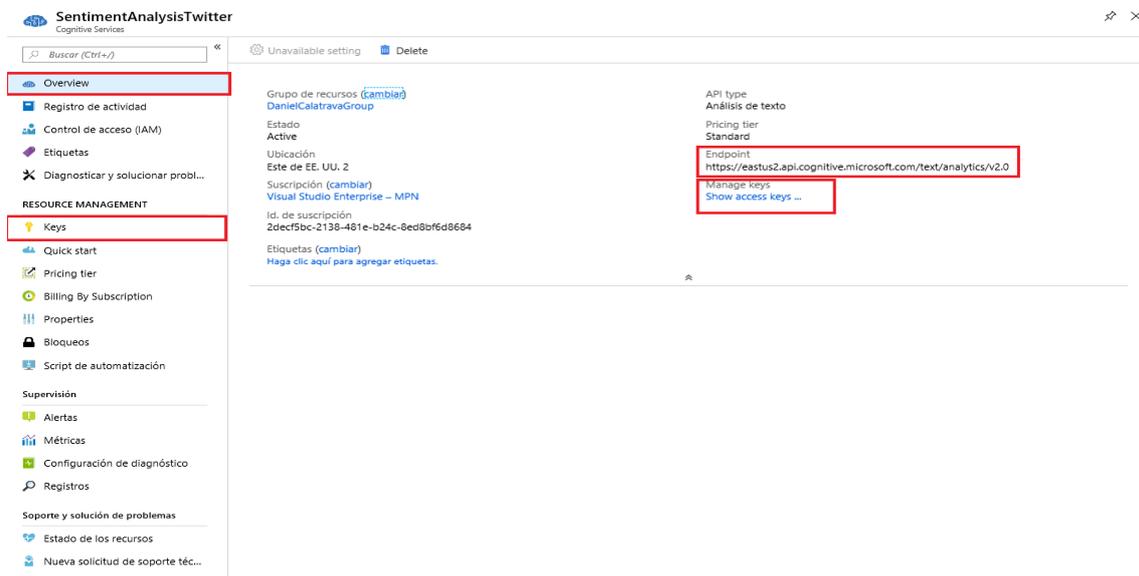


Figura 29. Microsoft Azure – Panel de Servicio Análisis de Texto.

Luego de configurar correctamente la conexión al servicio, como se muestra en la figura 30, es importante seleccionar el texto que se desea analizar dentro del detector de sentimiento. Para ello, se debe dar clic en “Texto”, donde se despliega un recuadro con el contenido dinámico que se puede acceder desde el Twitter API, seleccionar entre las columnas listadas el “Texto del Tweet”, .

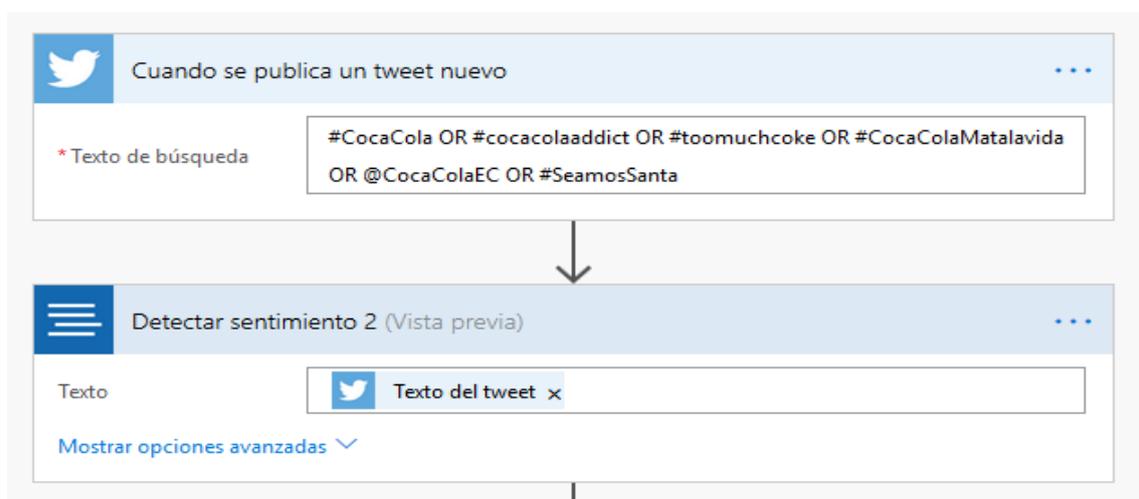


Figura 30. Microsoft Flow – Detectar Sentimiento para Texto del Tweet.

Finalmente, se debe crear una acción para insertar en la base de datos la información necesaria para cada ámbito a evaluar, para ello previamente se debe crear las columnas en las tablas a partir del contenido dinámico obtenido del Twitter API por el Microsoft Flow, desde el SQL Server crear las columnas necesarias y seleccionar en esta nueva acción la columna respectiva a insertar, como se muestra en la Figura 31 y 32.

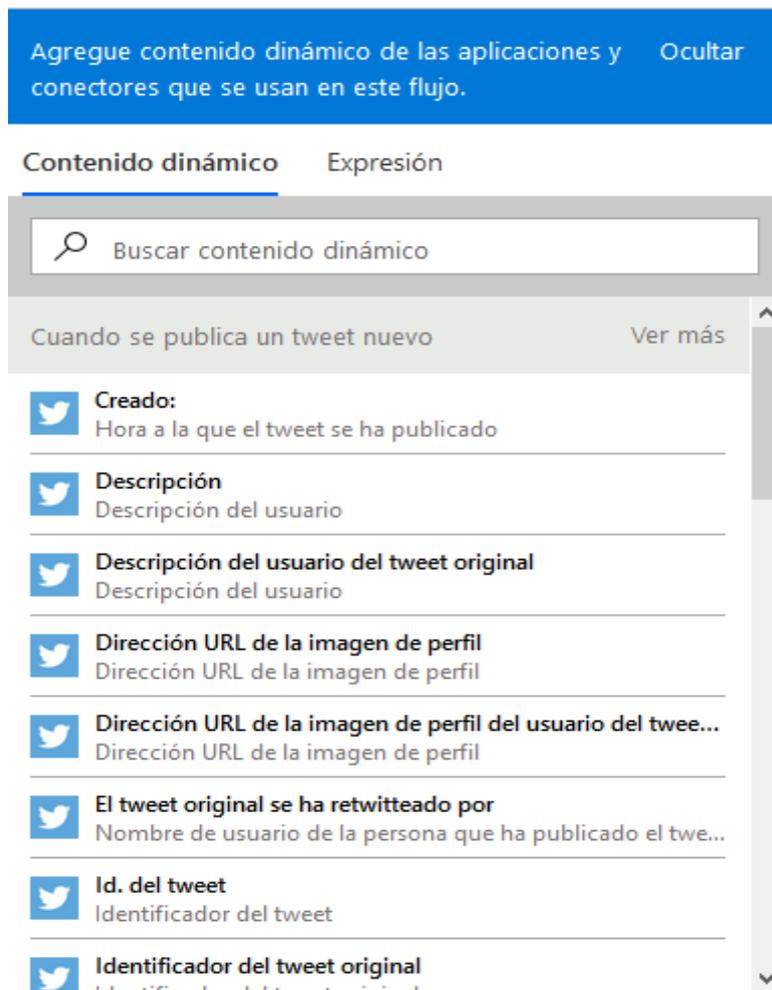


Figura 31. Microsoft Flow – Contenido Dinámico Twitter API.

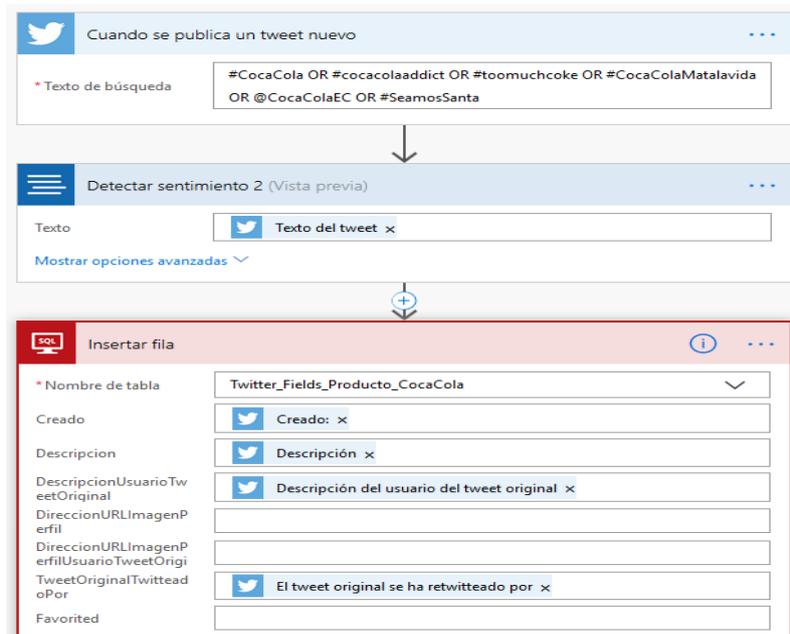


Figura 32. Microsoft Flow – Acción de Insertar fila en SQL Server DB.

Estos pasos son importantes para conseguir capturar la información de Twitter y analizar los mensajes generados por los usuarios a empresas, productos, eventos y personas. Luego de verificar que las acciones del flujo no tengan ningún error, se procede a guardar y a verificar que este habilitado para su desencadenamiento. Para probar, se debe ir a la cuenta de Twitter y generar un nuevo tweet con el termino de búsqueda configurado o esperar la generación de un tweet por un usuario de la red social, si todo está correcto, dentro del historial de flujo se observará la ejecución del flujo de trabajo, como se muestra en la Figura 33, y dentro de la base de datos se insertará la información.



Figura 33. Microsoft Flow – Panel de Flujo de trabajo.

Para concluir las configuraciones del aplicativo, desde la herramienta Power BI Desktop, como se muestra en la figura 34, se accede a la información capturada en la base de datos, para ello se debe ingresar al aplicativo e iniciar sesión con la misma cuenta donde se tenga la licencia correspondiente que permitirá generar sin inconvenientes los distintos reportes necesarios para analizar la información.

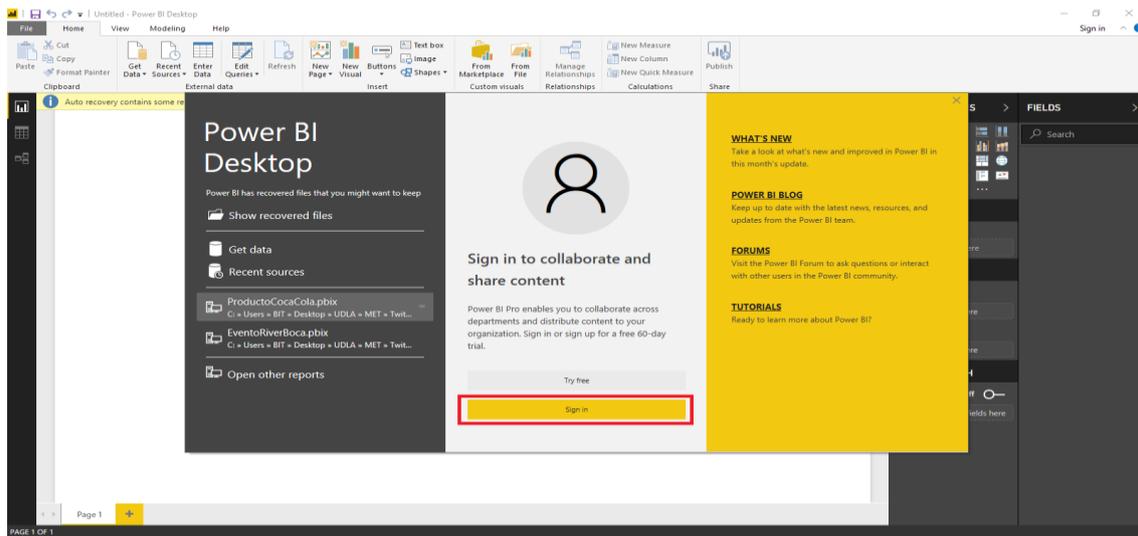


Figura 34. Power BI – Panel de Inicio de Sesión.

Posteriormente, se debe obtener los datos desde la base de datos, seleccionar la opción “Get Data” u “Obtener Información”, desde el menú superior. Como se muestra en la figura 35, esta herramienta permite obtener datos de múltiples fuentes, lo que no limita el funcionamiento ni el análisis de información y se ajusta a las necesidades del desarrollador o analista. Para este trabajo se obtienen los datos del SQL Server, en este punto se debe colocar todos los datos correspondientes para conectarse con el servidor de base de datos y a la tabla adecuada.

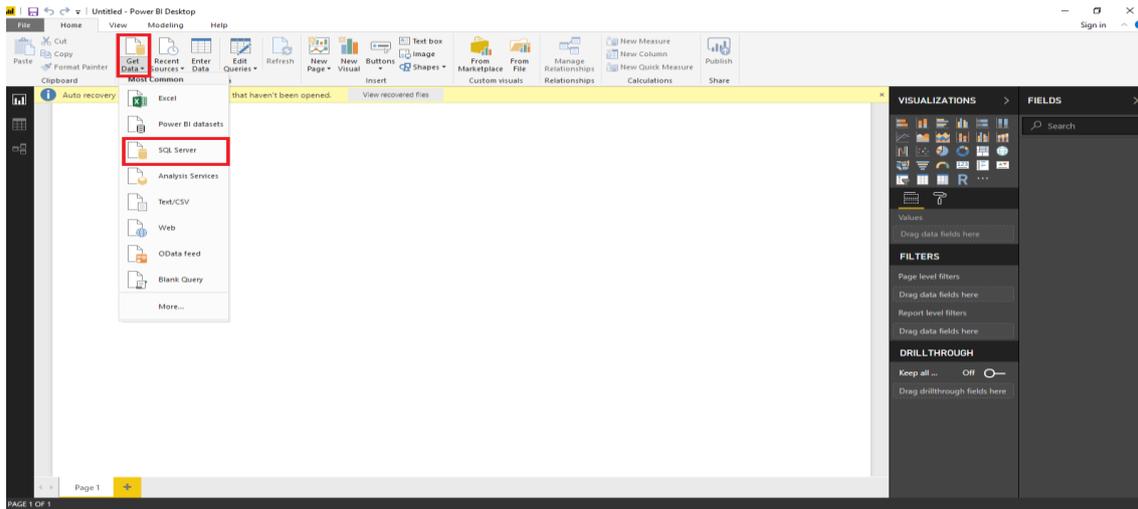


Figura 35. Power BI – Panel de Inicio/Obtener datos SQL Server.

Desde el menú lateral izquierdo se puede observar los datos de la base de datos y como está estructurada, por otro lado, desde el mismo menú donde se obtuvo la información se encuentra una opción que permite generar nuevas medidas, columnas condicionadas, entre otros, que facilitaran el análisis y visualización de los datos para el usuario final, como se muestra en la figura 36. Adicional, en esta nueva ventana se pueden observar los formatos de las columnas y cambiar según sea la necesidad, muchas son las opciones posibles con las que se puede trabajar para mejorar el trabajo a implementar.

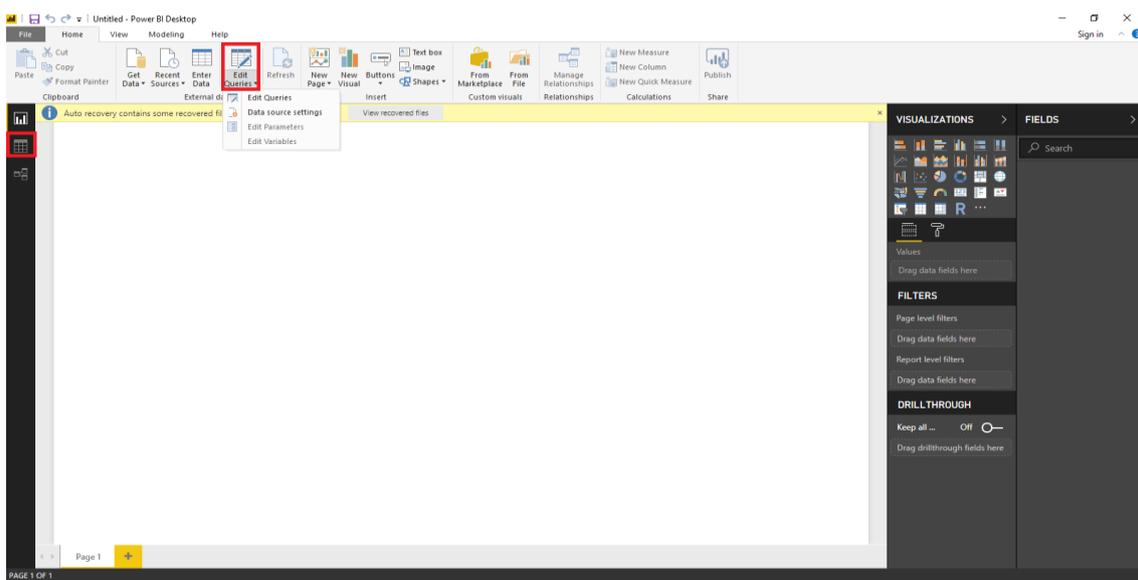


Figura 36. Power BI – Panel de Inicio/Tabla de datos/Edición de Tabla.

Por último, se debe ir creando los distintos reportes o visualizadores que graficarán la información lo más ordenada posible y según los parámetros que se configuren, adicional dentro de estos reportes se puede ir jugando con las distintas columnas hasta conseguir el dato deseado, es muy importante que la lógica tenga sentido ya que de no ser así la información no se mostrara correctamente. Como se detalla en la Figura 37, desde el menú lateral derecho se encuentran tanto las columnas a agregar por reporte, como una caja con las distintas herramientas a utilizar, si no se consigue alguna deseada desde el mercado de aplicaciones se puede importar y conseguir el deseado, de igual manera, se observa los filtros y opciones disponibles para configurar los reportes a gusto del usuario final.

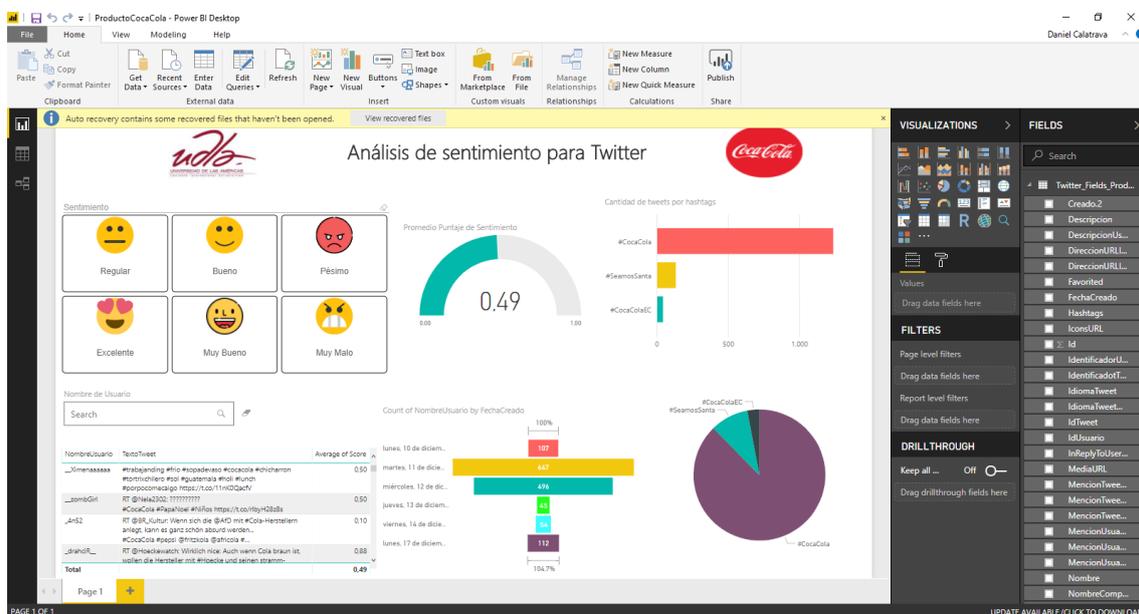


Figura 37. Power BI – Panel de reporte.

4.5 Conclusión del Capítulo.

En conclusión, la configuración de cada una de las herramientas utilizadas en este trabajo es muy importante para el desarrollo de este aplicativo, ya que, cada una de ellas cumple una función muy específica para generar el análisis de sentimientos, luego por inteligencia de negocios es estructurada con el fin de comprender los datos de una manera fácil y eficaz.

5. Capítulo V: Resultados.

5.1 Introducción

A continuación, se mostrarán los distintos resultados obtenidos en cada ámbito analizado, donde se realizará un desglose de cada herramienta usada y como se puede comprender sin dificultad. De igual manera, las herramientas utilizadas fueron escogidas por el desarrollador, a partir de la experiencia previa en trabajos de inteligencia de negocios y lo que permitió tomar la decisión de los componentes mas adecuados a ser añadidos en cada uno de los reportes. Esto con el fin de demostrar la potencialidad de la herramienta y con el fin de presentar los distintos usos de cada componente.

5.2 Análisis de Sentimientos de Organización o Empresa

Para el siguiente trabajo, se realizó un análisis de sentimientos, para la empresa Eléctrica de Quito, con el fin de comprender el comportamiento de los usuarios hacia esta empresa, a través de las distintas métricas y reportes utilizados en la herramienta de Power BI. Es importante resaltar que la información es capturada en tiempo real y se irá actualizando el tablero por intervalos definidos de 30 minutos.

Como se explicó en capítulos anteriores luego de realizar las configuraciones para el aplicativo se configuraron los siguientes parámetros para la captura de mensajes en Twitter para la empresa Eléctrica de Quito:

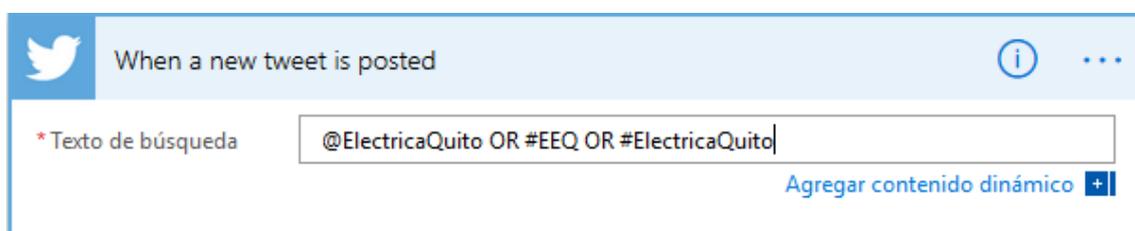


Figura 38. Microsoft Flow – Parámetros de captura de mensajes para Empresa EEQ.

En la Figura 38, se puede observar que los parámetros de búsquedas son concretos para lograr conseguir la información más acertada para esa empresa. Para ello con dirigirse a la base de datos y verificar que la información se esté guardando correctamente en base a estos parámetros configurados.

The screenshot shows a SQL Server query window with the following SQL code:

```

1 /***** Script for SelectTopNRows command from SSMS *****/
2 SELECT TOP (1000) [Id]
3     , [Creado]
4     , [Descripcion]
5     , [TextoTweet]
6     , [DescripcionUsuarioTweetOriginal]
7     , [DireccionURLImagenPerfil]
8     , [DireccionURLImagenPerfilUsuarioTweetOriginal]
9     , [TweetOriginalTwitteadoPor]
10    , [Favorited]
11    , [IdUsuario]
12    , [IdTweet]
13    , [IdentificadorUsuarioTweetOriginal]
14    , [IdentificadorTweetOriginal]
15    , [IdiomaTweet]
16    , [IdiomaTweetOriginal]
17    , [InReplyToUserId]
18    , [MediaURL]

```

The results window displays the following data:

Id	Creado	Descripcion	Texto Tweet	DescripcionUsua	
1	10787	2018-10-17T19:29:26.000Z	Operador Maquinaria Pesada	RT @ElectricaQuito: En el sector de las Antenas del Pichi...	Siempre junto a t
2	11024	2018-10-18T13:20:24.000Z	Ing. en Marketing NACIONAL es la Gloria del Fútbol!!!	RT @ElectricaQuito: La Empresa @ElectricaQuito cuenta...	Siempre junto a t
3	11242	2018-10-19T13:26:11.000Z	God's daughter Chubby soul. Veggie-Nutritionist-Human Be...	RT @LaChichoY: Querido @MauricioRodasEC @Municip...	mejor vivir sin mi
4	11378	2018-10-19T22:10:31.000Z	Cuenta Oficial de Twitter del Ministerio de Energía y Recurs...	RT @ElectricaQuito: Toma precauciones durante la temp...	Siempre junto a t
5	11637	2018-10-23T02:16:38.000Z	Cuenta oficial de la Empresa Pública Flota Petrolera Ecuato...	RT @ElectricaQuito: La @ElectricaQuito instala en el cabl...	Siempre junto a t
6	11809	2018-10-25T02:45:52.000Z	Odio la injusticia y la mentira y odio sentirme tan impotente	RT @IvanGualpa: @Lenin @ElectricaQuito Y el atracado...	
7	12029	2018-10-26T18:38:30.000Z	TRABAJAMOS PARA ILUMINAR TU VIDA Cuenta Oficial ...	RT @ElectricaQuito: Desde el año 2014, la @ElectricaQu...	Siempre junto a t
8	12228	2018-10-27T22:45:38.000Z	Siempre junto a ti	@flakita_ldu Buenas tardes, la EEQ le saluda. Debe ingre...	NULL
9	12505	2018-10-30T16:53:57.000Z	Siempre junto a ti	RT @CELEC_EP: @CELEC_EP implementa ingeniería e...	Somos la Corpor
10	12764	2018-11-02T01:05:19.000Z	Siempre junto a ti	RT @CELEC_EP: #Datosdegeneración 11 de noviembre...	Somos la Corpor
11	15204	2018-11-18T12:32:18.000Z	Siempre junto a ti	@Gringamiller Un gusto atenderle. Saludos.	NULL
12	18197	2018-12-02T14:28:25.000Z	Siempre junto a ti	@ElectricaQuito buenos días por favor en pusuqui no ten...	NULL
13	19778	2018-12-22T15:58:24.000Z	Nerd reprimido. Emprendedor, desde 1 de febrero del 2017!!!	@ElectricaQuito estamos desde las 8 sin luz en Tanda, qu...	NULL
14	12996	2018-11-04T16:10:23.000Z	Somos la Corporación Eléctrica del Ecuador. Tenemos 14 u...	RT @ElectricaQuito: Instalamos 186 luminarias en el parq...	Siempre junto a t
15	13242	2018-11-06T15:27:06.000Z	Siempre junto a ti	@PamAguasY Estimada, reportaremos su solicitud al área ...	NULL
16	13576	2018-11-08T23:19:56.000Z	Vivo en un país maravilloso, lleno de paraísos y parajes úni...	@ElectricaQuito @Lenin @Arconel_ec El edificio de junto...	NULL

Figura 39. SQL Server – Registros insertados en tabla de base de datos de Empresa EEQ.

Luego de verificar la información en la base de datos, como se muestra en la figura 39, se construyeron los distintos reportes que se detallan a continuación:

- **Chiclet Slicer:** con este visualizador se puede categorizar los puntajes definidos por el servicio cognitivo y desde el Power BI construir distintas imágenes a partir de estos valores, adicional se pueden organizar columnas que tengan información corta y definida como los idiomas.

Para este tablero de la empresa EEQ (Eléctrica de Quito), las categorías por imagen, como se muestra en la figura 40, fueron definidas por los siguientes parámetros:

- Valores comprendidos entre $0 > 0.4$, Negativo.
- Valores comprendidos entre $0.4 \geq 0.7$, Neutral.
- Valores comprendidos entre $0.7 > 1$, Positivo.

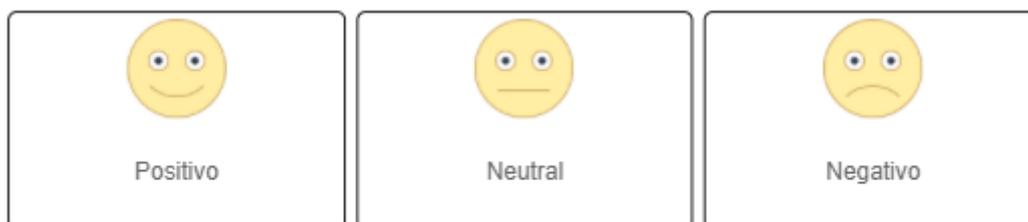


Figura 40. Power BI – Visualizador 1 para tablero de EEQ.

Otra manera de utilizar este visualizador para filtrar por el idioma del texto, como se muestra en la figura 41, para ello se debe tener una columna con el idioma del texto, lo que permitirá automáticamente agregar en este reporte, es importante resaltar que se puede filtrar los idiomas que se desean mostrar ya que es posible que entre la información recolectada existan tweets de otros países y con idiomas distintos al español e inglés.



Figura 41. Power BI – Visualizador 2 para tablero de EEQ.

- **Gráfico de columnas agrupadas:** con este visualizador se puede evaluar distintas métricas dentro de la información recolectada, como se muestra en la figura 42 y 43, sin embargo, para este trabajo se utilizó con el fin de definir la cantidad de tweets Positivos, Negativos y Neutrales dentro de la base de datos y para definir cuantos tweets eran generados mensualmente.

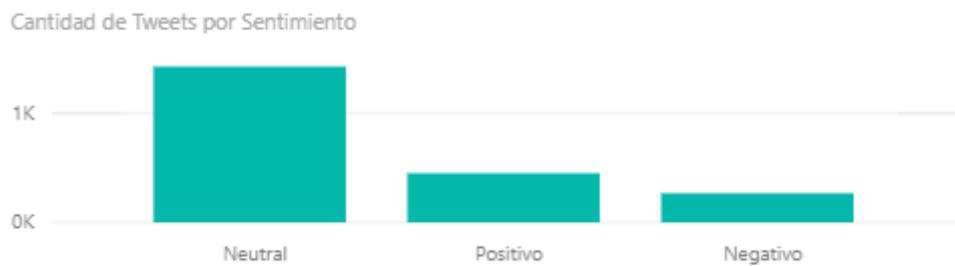


Figura 42. Power BI – Visualizador 3 para tablero de EEQ.

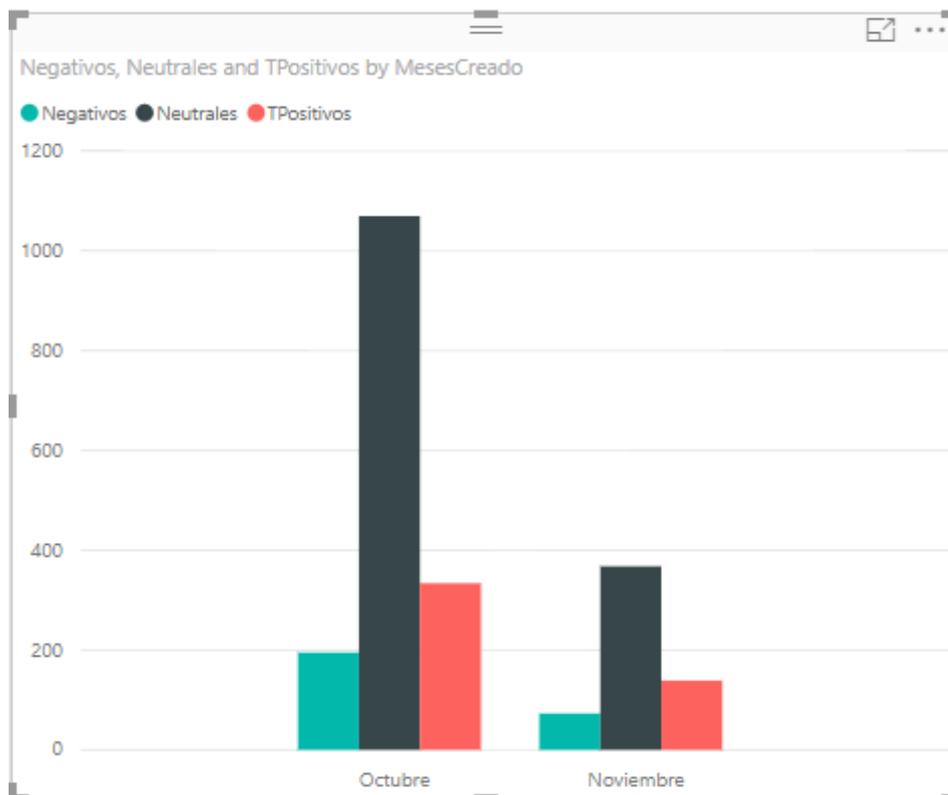
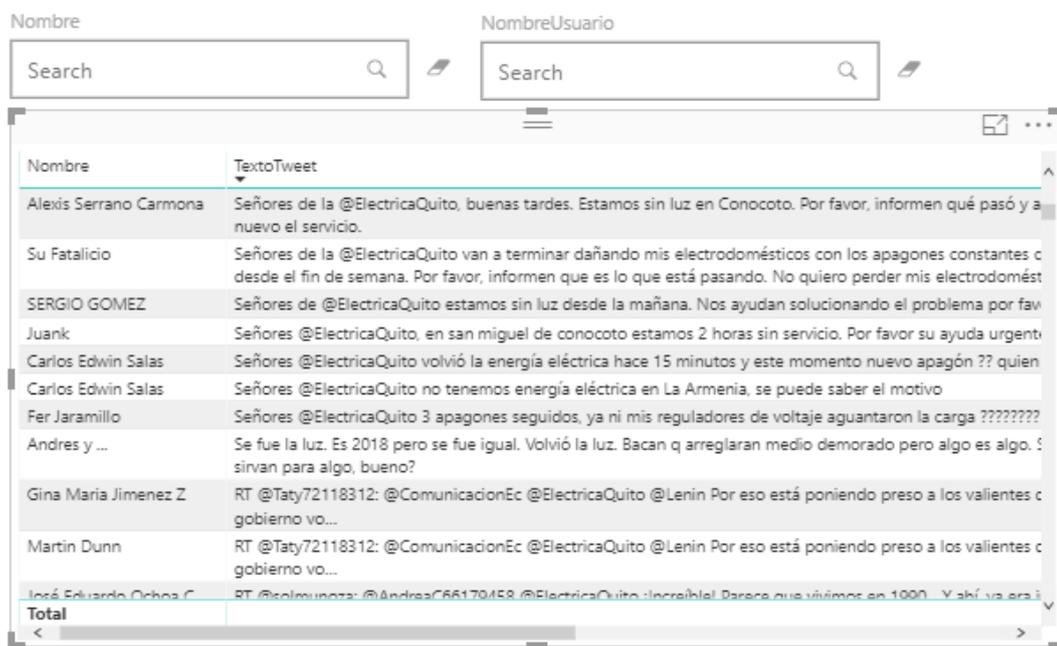


Figura 43. Power BI – Visualizador 4 para tablero de EEQ.

- **Tabla y Filtro de texto:** con estos visualizadores se listan las distintas columnas que serán importantes para el usuario al momento de buscar un tweet en específico y para ello se tiene el visualizador de filtro de texto, el cual a partir de una columna que se encuentre en la tabla realice un filtro de la información. Para este tablero, como se muestra en la figura 44, se utilizaron con el fin de observar los puntajes por tweet, los textos de los tweets, los usuarios que realizaron el tweet, la ubicación de donde provino el tweet.



Nombre	TextoTweet
Alexis Serrano Carmona	Señores de la @ElectricaQuito, buenas tardes. Estamos sin luz en Conocoto. Por favor, informen qué pasó y a nuevo el servicio.
Su Fatalicio	Señores de la @ElectricaQuito van a terminar dañando mis electrodomésticos con los apagones constantes desde el fin de semana. Por favor, informen que es lo que está pasando. No quiero perder mis electrodomésticos.
SERGIO GOMEZ	Señores de @ElectricaQuito estamos sin luz desde la mañana. Nos ayudan solucionando el problema por favor.
Juank	Señores @ElectricaQuito, en san miguel de conocoto estamos 2 horas sin servicio. Por favor su ayuda urgente.
Carlos Edwin Salas	Señores @ElectricaQuito volvió la energía eléctrica hace 15 minutos y este momento nuevo apagón ?? quien es el responsable?
Carlos Edwin Salas	Señores @ElectricaQuito no tenemos energía eléctrica en La Armenia, se puede saber el motivo?
Fer Jaramillo	Señores @ElectricaQuito 3 apagones seguidos, ya ni mis reguladores de voltaje aguantaron la carga ????????
Andres y ...	Se fue la luz. Es 2018 pero se fue igual. Volvió la luz. Bacan q arreglaran medio demorado pero algo es algo. ¿sirvan para algo, bueno?
Gina Maria Jimenez Z	RT @Taty72118312: @ComunicacionEc @ElectricaQuito @Lenin Por eso está poniendo preso a los valientes con el gobierno voy...
Martin Dunn	RT @Taty72118312: @ComunicacionEc @ElectricaQuito @Lenin Por eso está poniendo preso a los valientes con el gobierno voy...
José Eduardo Ordoñez C	RT @bolmuna: @AndresC66170458 @ElectricaQuito ¡Increíble! Parece que vivimos en 1900. Y ahí va otra vez la luz.
Total	

Figura 44. Power BI – Visualizador 5 para tablero de EEQ.

- **Mapa:** con este visualizador se puede saber con exactitud la ubicación del usuario que generó un tweet, sin embargo, para este tablero lo que se puede observar es una posible ubicación dentro del país, ya que desde el Twitter API no se puede obtener la ubicación exacta del usuario que generó el tweet a menos que tenga habilitada esa opción, como se muestra en la figura 45.



Figura 45. Power BI – Visualizador 6 para tablero de EEQ.

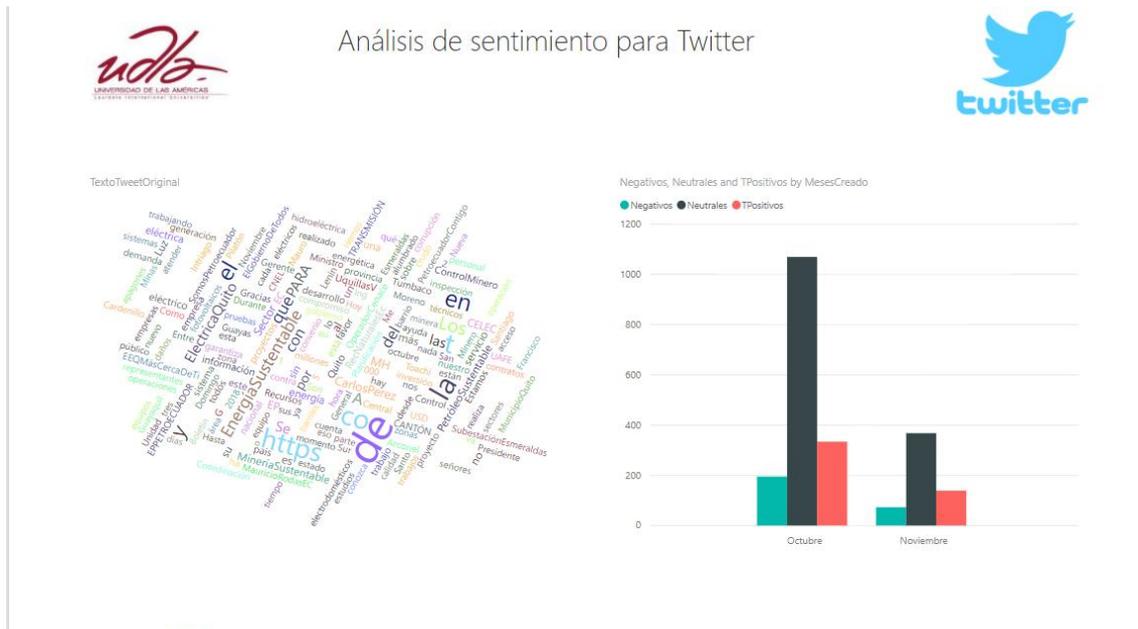


Figura 48. Power BI – Tablero 2 para empresa EEQ.

5.3 Análisis de Sentimientos para un Producto

Para el siguiente trabajo se realizó un análisis de sentimientos para un producto conocido mundialmente como Coca-Cola, esto con el fin de comprender el comportamiento de los usuarios hacia este producto. Es importante resaltar que la información es capturada en tiempo real y se irá actualizando el tablero por intervalos definidos de 30 minutos.

Como se explicó previamente luego de realizar las configuraciones para el aplicativo se configuraron los siguientes parámetros para la captura de mensajes en Twitter para el producto Coca-Cola:

Cuando se publica un tweet nuevo

* Texto de búsqueda

#CocaCola OR #cococolaaddict OR #toomuchcoke OR
#CocaColaMatalavida OR @CocaColaEC OR #SeamosSanta

Figura 49. Microsoft Flow – Parámetros de captura de mensajes para Coca-Cola.

En la Figura 49, se puede observar que se agregaron parámetros definidos por la cantidad de tweets generados por los usuarios para esas etiquetas y que previamente se analizó en Twitter cuales de estos eran los más viables para conseguir la información necesaria a evaluar.

Luego de verificar la información en la base de datos, se construyó los distintos reportes que se detallan a continuación:

- **Chiclet Slicer:** con este visualizador se categorizo el puntaje obtenido por el tweet con una parametrización distinta a la utilizada en el tablero de la empresa Eléctrica de Quito. Para este tablero Coca-Cola, como se muestra en la figura 50, las categorías por imagen fueron definidas por los siguientes parámetros:

- Valores comprendidos entre 0 a 0.2, Pésimo.
- Valores comprendidos entre 0.2 a 0.5, Muy Malo.
- Valores comprendidos entre 0.5 a 0.6, Regular.
- Valores comprendidos entre 0.6 a 0.8, Bueno.
- Valores comprendidos entre 0.8 a 0.9, Muy Bueno.
- Valores comprendidos entre 0.9 a 1, Excelente.

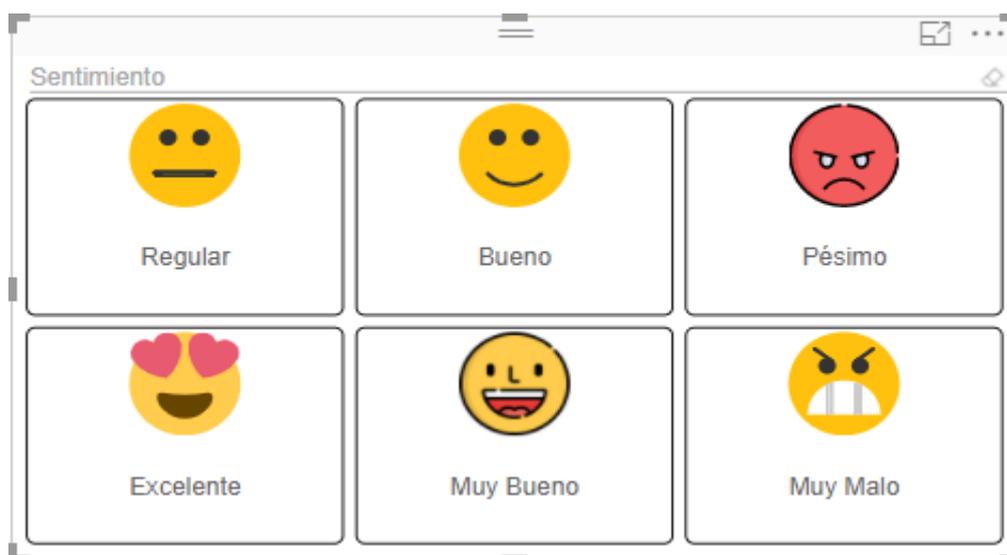


Figura 50. Power BI – Visualizador 1 para tablero de Coca-Cola.

Esto permite verificar que se puede adaptar el formato de evaluación de los puntajes generados por el servicio cognitivo a conveniencia, lo que facilita ciertos análisis para el usuario.

- **Gráfico de columnas agrupadas:** con este visualizador se puede evaluar distintas métricas dentro de la información recolectada, sin embargo, para este tablero se utilizó con el fin de analizar la cantidad de tweets que se generaron por Hashtags o Etiquetas. como se muestra en la figura 51.

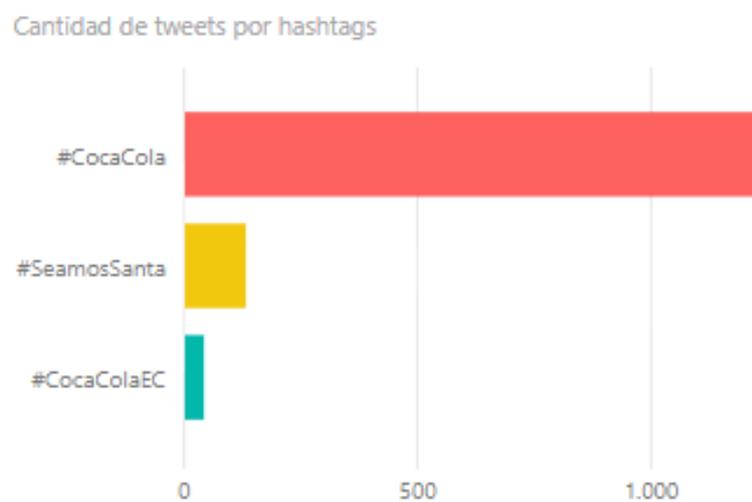


Figura 51. Power BI – Visualizador 2 para tablero de Coca-Cola.

- **Tabla y Filtro de texto:** con estos visualizadores se listan las distintas columnas que serán importantes para el usuario al momento de buscar un tweet en específico y para ello se tiene el visualizador de filtro de texto, el cual a partir de una columna que se encuentre en la tabla realice un filtro de la información, como se muestra en la figura 52.

NombreUsuario	TextoTweet	Average of Score
_4n52	RT @BR_Kultur: Wenn sich die @AfD mit #Cola-Herstellern anlegt, kann es ganz schön absurd werden... #CocaCola #pepsi @fritzkola @africola #...	0,10
drahciR	RT @Hoeckewatch: Wirklich nice: Auch wenn Cola braun ist, wollen die Hersteller mit #Hoecke und seinen stramm-braunen Kamerrraden der #AfD...	0,88
erincal	Compré flores para mi mamá ?? #SeamosSanta @thompkimberly https://t.co/Fjlo4KCNHi	0,98
_gregorilg	RT @BR_Kultur: Wenn sich die @AfD mit #Cola-Herstellern anlegt, kann es ganz schön absurd werden...	0,10
Total		0,47

Figura 52. Power BI – Visualizador 3 para tablero de Coca-Cola.

- **Embudo o Funnel:** con este visualizador se puede analizar los tweets o la cantidad de usuarios que generaron un tweet en un determinado día, esto permite analizar el comportamiento de los usuarios en la semana, como se muestra en la figura 53.

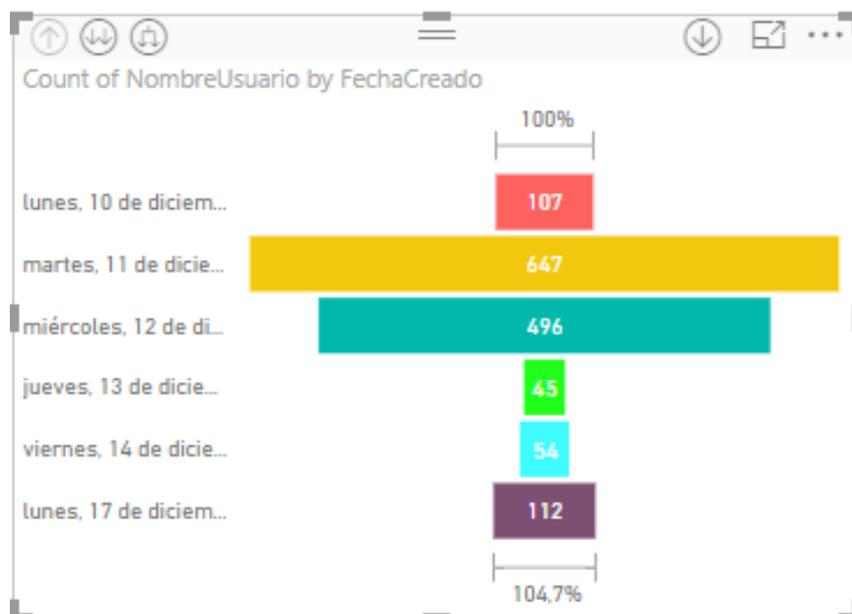


Figura 53. Power BI – Visualizador 4 para tablero de Coca-Cola.

- **Indicador o Gauge:** a partir de este visualizador se busca obtener el promedio de sentimientos que se han generado con los tweets hasta el momento y con ello saber el nivel de satisfacción de los distintos usuarios hacia el producto, como se muestra en la figura 54.

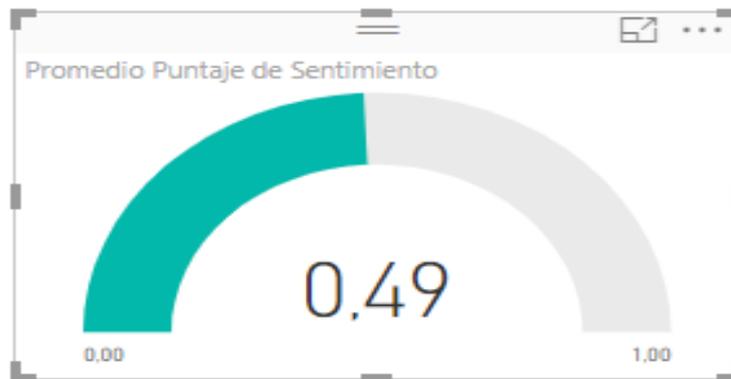


Figura 54. Power BI – Visualizador 5 para tablero de Coca-Cola.

- **Gráfico circular:** al igual que el gráfico en barras, con este visualizador se puede observar la cantidad de hashtags o etiquetas generadas en los distintos tweets. Esto ayuda a comprender que se puede construir un mismo gráfico a través de otros visualizadores y no limitándose a los por defecto del Power BI, como se muestra en la figura 55.

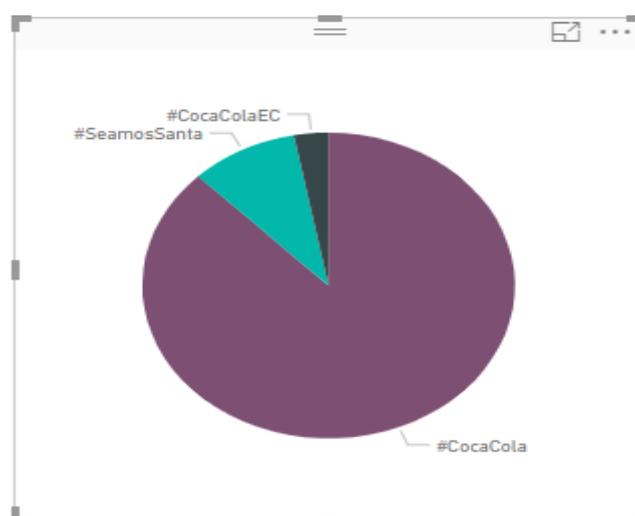


Figura 55. Power BI – Visualizador 6 para tablero de Coca-Cola.

Finalmente, luego de construir cada uno de los reportes para el producto seleccionado, se obtiene un tablero como se muestra en la Figura 56.

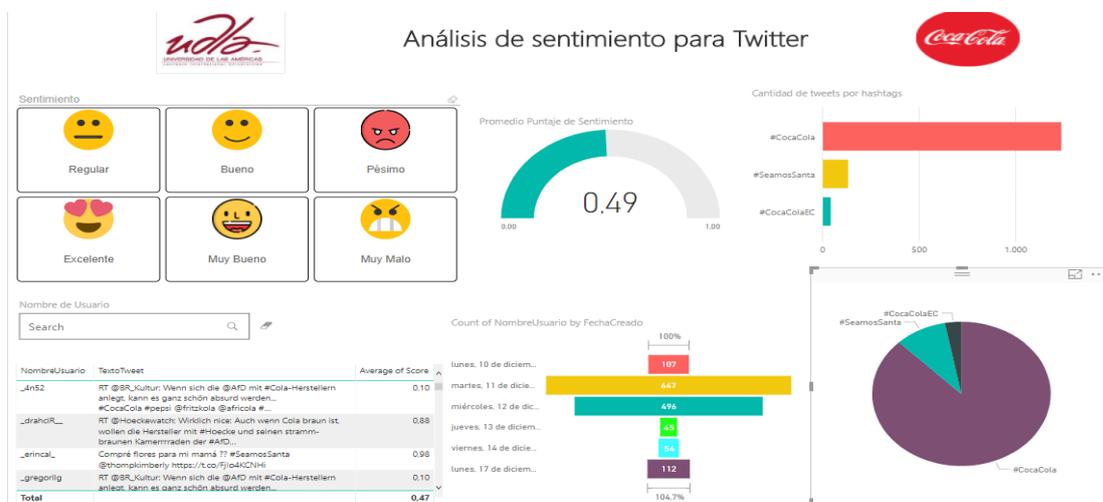


Figura 56. Power BI – Tablero para producto Coca-Cola.

5.4 Análisis de Sentimientos para Evento

Para el siguiente trabajo se realizó un análisis de sentimientos para un evento futbolístico entre dos grandes rivales que se disputaban la copa libertadores River Plate Vs Boca Juniors, esto con el fin de comprender el comportamiento de los usuarios hacia cualquiera de los dos equipos. Es importante resaltar que la información es capturada en tiempo real y se irá actualizando el tablero por intervalos definidos de 30 minutos.

Como se explicó previamente luego de realizar las configuraciones para el aplicativo, se configuraron los siguientes parámetros para la captura de mensajes en Twitter para el evento de River Plate Vs Boca Juniors:

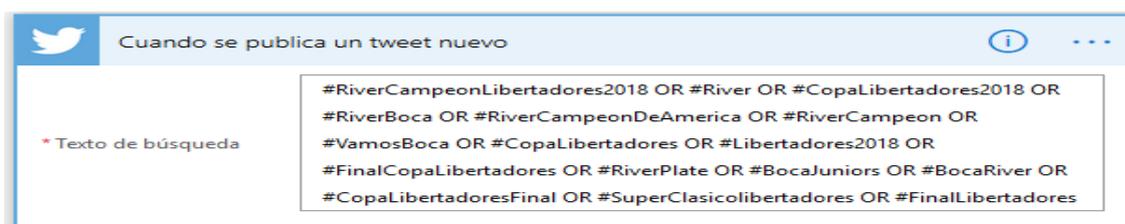


Figura 57. Microsoft Flow – Parámetros de captura de mensajes para RiverVsBoca.

En la Figura 57, se puede observar que se agregaron parámetros definidos por la cantidad de tweets generados por los usuarios para esas etiquetas y que previamente se analizó en Twitter cuales de estos eran los más viables para conseguir la información necesaria a evaluar previo al partido que se disputaría en una fecha pactada.

Luego de verificar la información en la base de datos, se construyen los distintos reportes que se detallan a continuación:

- **Chiclet Slicer:** con este visualizador se puede categorizar los puntajes definidos por el servicio cognitivo y desde el Power BI construir distintas imágenes a partir de estos valores, como se muestra en la figura 58.

Para este tablero del evento entre River Vs Boca, las categorías por imagen fueron definidas por los siguientes parámetros:

- Valores comprendidos entre 0 a 0.4, Negativo.
- Valores comprendidos entre 0.4 a 0.7, Neutral.
- Valores comprendidos entre 0.7 a 1, Positivo.



Figura 58. Power BI – Visualizador 1 para tablero de RiverVsBoca.

Esto permite verificar que se puede adaptar el formato de evaluación de los puntajes generados por el servicio cognitivo a conveniencia, lo que facilita ciertos análisis para el usuario.

- **Gráfico de columnas agrupadas:** con este visualizador se puede evaluar distintas métricas dentro de la información recolectada, sin embargo, para este tablero se utilizó con el fin de analizar la cantidad de tweets que se

generaron por Hashtags o Etiquetas. Además, se desea contabilizar los tweets positivos y negativos generados por los distintos usuarios en Twitter para el evento, como se muestra en la figura 59, 60 y 61.

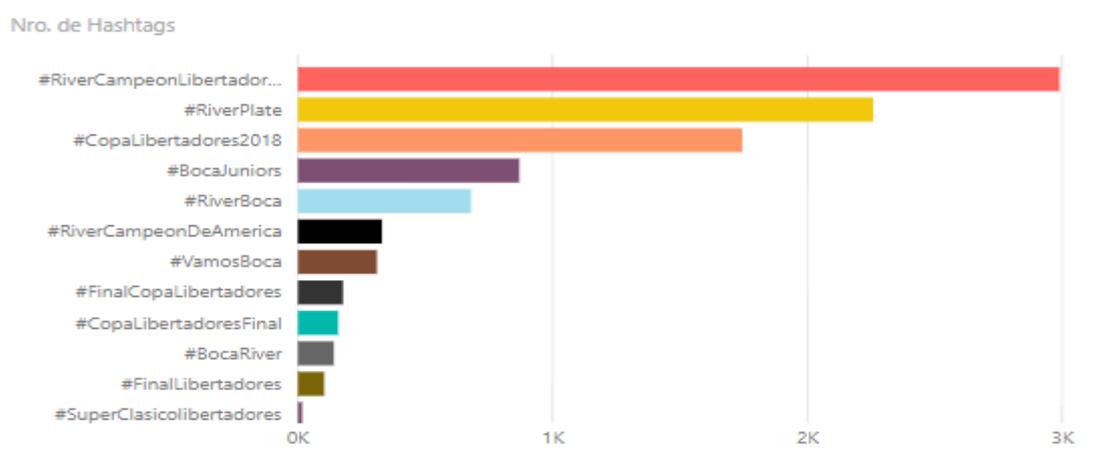


Figura 59. Power BI – Visualizador 2 para tablero de RiverVsBoca.



Figura 60. Power BI – Visualizador 3 para tablero de RiverVsBoca.

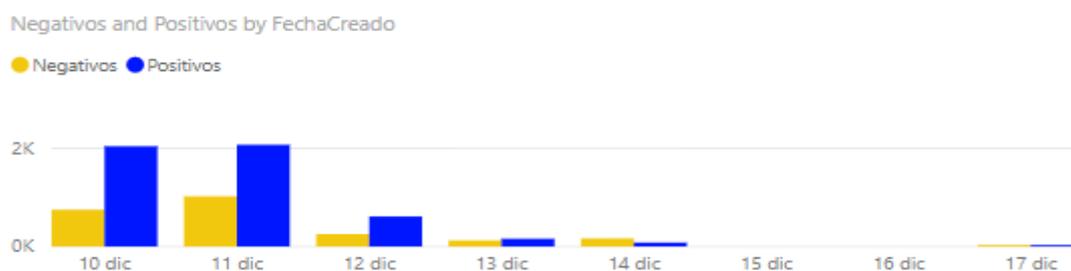


Figura 61. Power BI – Visualizador 4 para tablero de RiverVsBoca.

En la Figura 61, se puede observar que el grafico detalla los tweets positivos y negativos generados por fecha, adicional se observa como poco a poco fue

disminuyendo el índice de tweets esto motivado a que el evento empieza a quedar en el pasado.

- **Tabla:** con estos visualizadores se listan las distintas columnas que serán importantes para el usuario al momento de buscar un tweet en específico, como se muestra en la figura 62.

NombreUsuario	TextoTweet	Average of Score
petritamallorca	#River ?????? https://t.co/svE1SHsdO2	1,00
Loretulkan	#RiverCampeon #CopaLibertadores2018 ?????? https://t.co/Y3IB2zOwf2	1,00
arrua_soledad	#rivercampeonlibertadores ?????????????? #River #juntossomosmasgrandes #Gracias ?? #goodvibes ??? ? https://t.co/XzyhqLvlhu	1,00
jorariass	#RiverCampeonLibertadores2018 ????	1,00
solcigrandis	??? ?s?? s????s? ? ??? ?s?? ?s???? ??? s?????? ????? #rivercampeónlibertadores2018 https://t.co/DUIXfpehshz	1,00
FutbolPermitido	????#RiverCampeonLibertadores2018 ??? https://t.co/kVYnDF3IYR	1,00
FlorBaezCrespin	?????? #RiverCampeonLibertadores2018 https://t.co/jV81T6N65	1,00
Florr_ponce	?????? #RiverCampeonLibertadores2018 https://t.co/nv22i1oCtI	1,00
Total		0,63

Figura 62. Power BI – Visualizador 5 para tablero de RiverVsBoca.

- **Indicador o Gauge:** a partir de este visualizador la cantidad de tweets generados tanto para River Plate, como para Boca Juniors y a partir de ello se observa que existe un índice bastante alto de tweets generados para River, esto dado por ser el campeón de la copa libertadores en 2018, como se muestra en la figura 63.

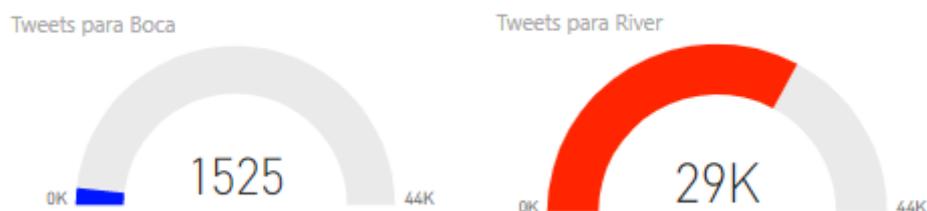


Figura 63. Power BI – Visualizador 6 para tablero de RiverVsBoca.

Finalmente, luego de construir cada uno de los reportes para el evento seleccionado, se obtiene un tablero como se muestra en la Figura. 64.

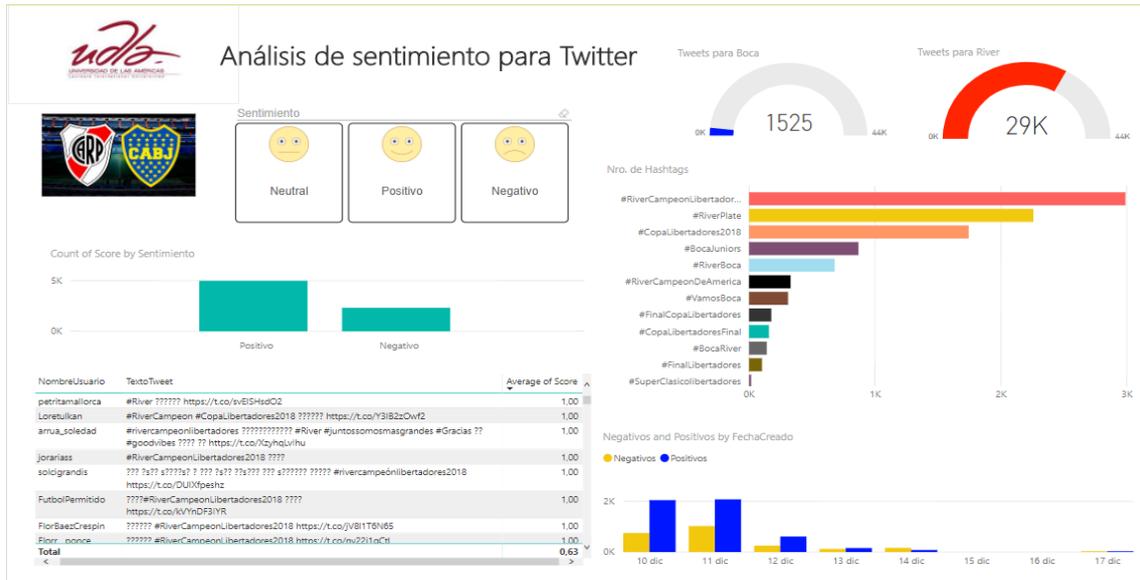


Figura 64. Power BI – Tablero para el evento de RiverVsBoca.

5.5 Conclusión del Capítulo.

En conclusión, los resultados obtenidos indican que la información recolectada permite al cliente comprender si la empresa, producto o evento, está en capacidad de satisfacer a los usuarios de la red social Twitter. Además, a través de los distintos visualizadores que ofrece Power BI se pueden adaptar los reportes a la necesidad del usuario o cliente, lo importante es definir los requerimientos y la información que se desea analizar. Estos visualizadores implementados en cada reporte fueron escogidos por el desarrollador y no por personas pertenecientes a las empresas, ya que se buscó demostrar el potencial de la aplicación de inteligencia de negocios para analizar los datos.

6. CONCLUSIONES Y RECOMENDACIONES.

6.1 Conclusiones

Se realizó el desarrollo de un aplicativo para la inteligencia de negocios y análisis de sentimientos con las herramientas de Microsoft, donde se concluyó que posee herramientas poderosas para la implementación del proyecto, sin embargo, se tienen muchas restricciones que pueden limitar su desarrollo, entre ellas están el tema de las licencias, para este proyecto se contó con el apoyo de una empresa auspiciante que proveyó cada una de estas licencias.

Se realizó el desarrollo del reporte para la empresa patrocinadora, pero los datos no eran suficientes para generar un análisis adecuado de la información capturada, por ende, se decidió expandir la implementación a los tres ámbitos descritos anteriormente, los cuales tenían la capacidad de masificar o viralizar los datos que se capturaron para este proyecto.

Se desarrollaron distintos tableros con el fin de analizar distintos ámbitos de trabajo, que permitían adaptar la información a cada uno de los reportes. Después de comprender cada uno de los visualizadores y como se pueden utilizar con la información recolectada se concluyó que es importante tener requerimientos definidos por un cliente, con el fin de estructurar el tablero acorde a las necesidades del cliente.

Se amplió el conocimiento sobre la herramienta de Power BI y su valor para analizar los datos obtenidos en los distintos ámbitos, ya que permite visualizar, editar y filtrar la información de una manera bastante amigable e intuitiva, lo que facilitó la implementación de cada tablero desarrollado.

Se logró conocer las limitaciones de cada herramienta, esto facilitó el desarrollo del aplicativo, sin embargo, algunas herramientas como Azure, limitan el funcionamiento al bloquear o deshabilitar los servicios utilizados. Adicionalmente si no se tiene buen conocimiento de Azure y sus servicios, el consumo monetario puede ser muy elevado, sobrepasándose los límites estipulados, deshabilitando la licencia o suscripción adquirida. Por ende, se debe investigar a fondo cual servicio se debe adquirir y con el plan que se ajuste a las necesidades.

6.2 Recomendaciones

Se recomienda fortalecer la comunicación con los clientes, con el fin de comprender sus necesidades en relación a sus productos o a su empresa, ya que esto ayuda a definir los requerimientos correctos, esto permitirá que el reporte se adapte a las necesidades del usuario y pueda obtener los beneficios esperados con este aplicativo de análisis de sentimientos.

Es muy importante verificar el presupuesto que posee el cliente para obtener este aplicativo, ya que a partir de eso se puede adaptar las distintas licencias y poder desarrollar sin inconvenientes. Sin embargo, se debe explicar los limitantes de cada licenciamiento y el alcance que tendrá el aplicativo de análisis de sentimientos.

Se recomienda investigar cada una de las herramientas hasta comprenderlas en su gran parte, es muy importante ya que de no ser el caso se pueden producir muchos errores y problemas que afecten la sincronización entre las distintas herramientas.

Los resultados que se presentan en los distintos reportes sobre el análisis de sentimiento para un producto, evento y empresa, aumentan el conocimiento necesario para el desarrollo de inteligencia de negocios en otros campos y adicional proporcionan información valiosa que permite a las empresas generar estrategias para mejorar su relación con los clientes.

REFERENCIAS.

- Aalderks, D., Falls, S. (2012). *Sentiment Analysis, Analyzing the Polarity of Communication*. Recuperado el 6 de noviembre de 2018 de <https://web.njit.edu/~da225/media/Assignment%207&8%20Sentiment%20Analysis.pdf>.
- Alberto, B., Nicolas, B., and Latifa, G., (2012). *Workshop on Mining Unstructured Data (MUD) "mining unstructured data is like fishing in muddy waters"!* 2012 19th Working Conference on Reverse Engineering, IEEE Computer Society Conference Publishing Services (CPS). Recuperado el 6 de noviembre de 2018 de <http://www.computer.org/cps>.
- Azure. (2018). Que es Azure. Recuperado el 10 de noviembre de 2018 de <https://azure.microsoft.com/es-es/overview/what-is-azure/>
- Barabas, J. (2015). Una Perspectiva de IBM: IaaS vs PaaS vs SaaS. Recuperado el 12 de noviembre de 2018 de <https://www.ibm.com/cloud-computing/mx-es/learn-more/iaas-paas-saas/>
- Berg, K., Seymour, T., Goel, R. (2013). *History of Databases*. Recuperado el 7 de noviembre de 2018 de https://www.researchgate.net/publication/298332910_History_Of_Databases
- Bogza, R.M., Dorin, Z., (2008). "Business intelligence as a competitive differentiator", *AQTR, 2008, International Conference on Automation, Quality and Testing, Robotics, International Conference on Automation, Quality and Testing, Robotics 2008*, pp. 146-151, doi:10.1109/AQTR.2008.

- Cabot, S. (2018). Por qué debería comenzar a incorporar los servicios cognitivos de Azure en sus aplicaciones. Recuperado el 8 de noviembre de 2018 de <https://hackernoon.com/why-you-should-start-incorporating-azure-cognitive-services-in-your-apps-4b3786a5a58e>
- Collabnet. (2018). Que es metodología scrum. Recuperado el 6 de noviembre de 2018 de <https://resources.collab.net/agile-101/what-is-scrum>
- Costa, P., Souza, F. (2012). *Towards Integrating Online Social Networks and Business Intelligence. International Conferences Web Based Communities and Social Media (IADIS WBCSM 2012)*. Recuperado el 10 de diciembre de 2018 de <http://homepages.dcc.ufmg.br/~fabricio/download/iadis-wbc2012.pdf>.
- Chen, H., Storey, V. (2012). *Business intelligence and analytics: from big data to big Impact*. MIS Quarterly, 36(4), 1–24.
- Davidov, D., Tsur, O., Rappoport, A. (2010). *Enhanced sentiment learning using twitter hashtags and smileys*. Recuperado el 12 de diciembre de 2018 de <http://www.aclweb.org/anthology/C10-2028>.
- Diezma, P. (2017). ¿Qué nos ofrece la inteligencia artificial y los servicios cognitivos? Recuperado el 15 de noviembre de 2018 de <http://www.zerintia.com/blog/que-nos-ofrece-la-inteligencia-artificial-y-los-servicios-cognitivos/>
- Esuli, A., Sebastiani, F. (2006). SentiWordNet: *A publicly available lexical resource for opinion mining*. Recuperado el 13 de diciembre de 2018 de <http://nmis.isti.cnr.it/sebastiani/Publications/LREC06.pdf>
- Foote, K. (2017). *A Brief history of Database Management*. Recuperado el 17 de noviembre de 2018 de <http://www.dataversity.net/brief-history-database-management/>
- Francia, J. (2017). ¿Qué es Scrum? Recuperado el 21 de noviembre de 2018 de <https://www.scrum.org/resources/blog/que-es-scrum>

- Go, A., Huang, L., Bhayani, R. (2009). *Twitter sentiment analysis. Final Projects from CS224N for Spring 2008/2009 at The Stanford Natural Language Processing Group*. Recuperado el 22 de noviembre de 2018 de <http://hughchristensen.co.uk/papers/socialNetworking/Twitter%20Sentiment%20Analysis.pdf>
- Grimes, B. (2008). *Voice of the Customer. Text Analytics for the Responsive Enterprise. Business Intelligence Network. Powell Media and Alta Plana Corporation*. Recuperado el 22 de noviembre de 2018 de <http://altaplana.com/BIN-VOCTextAnalyticsReport.pdf>.
- Hagenau, M., Liebmann, M., Neumann, D. (2013). *Automated news reading: stock price prediction based on financial news using context-capturing features*. *Decis Supp Syst*, 5(3), pp. 145-156.
- Hernández, S. (2016). *Microsoft Flow: creando flujos de trabajo en la nube*. Recuperado el 8 de noviembre de 2018 de <https://blogs.encamina.com/por-una-nube-sostenible/microsoft-flow-creando-flujos-trabajo-la-nube/>
- IBM Corporation. (2012). *Manual CRISP-DM de IBM SPSS Modeler*. Recuperado el 29 de octubre de 2018 de: <ftp://ftp.software.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- Kim, S.-M., Hovy, E. (2004). *Determining the sentiment of opinions*. Recuperado el 23 de noviembre de 2018 de <http://www.aclweb.org/anthology/C04-1200>
- Liang-Chih Yu, Jheng-Long Wu, Pei-Chann Chang, Hsuan-Shou Chu. (2013). *Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news*. *Knowl-Based Syst*, 41 (2), pp. 89-97.

- Liu, B. (2012). *Sentiment analysis and opinion mining*. Recuperado el 21 de noviembre de 2018 de <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>
- Lucidchart. (2016). Diagrama de Flujo de Trabajo. Recuperado el 8 de noviembre de 2018 de <https://www.lucidchart.com/pages/es/que-es-un-diagrama-de-flujo-de-trabajo>
- Makesoft Technologies. (2017). Como funciona Power BI Desktop. Recuperado el 27 de noviembre de 2018 de <https://www.makesoft.es/es/como-funciona-power-bi-desktop/>
- Maks, I., Vossen, P. (2012). *A lexicon model for deep sentiment analysis and opinion mining applications*. *Decis Support Syst*, 53 (12), pp. 680-688.
- Mikalai, T., Palpanas, T. (2012). *Survey on mining subjective data on the web* *Data Min Knowl. Discov*, 24 (5), pp. 478-514.
- Oracle Corporation. (2015). Inteligencia de Negocios. Recuperado el 9 de noviembre de 2018 de https://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/317529_esa.pdf
- Pang, B., and Lee, L. (2008). *Opinion mining and sentiment analysis*. *Foundations and Trends in Information Retrieval* 2(1-2):1–135.
- Parimi, N. (2015). Introducción a la analítica de texto en Azure ML. Recuperado el 13 de noviembre de 2018 de <https://blogs.technet.microsoft.com/machinelearning/2015/04/08/introducing-text-analytics-in-the-azure-ml-marketplace/>

- Pak, A., Paroubek, P. (2010). *Twitter as a corpus for sentiment analysis and opinion mining*. Recuperado el 12 de noviembre de 2018 de https://www.researchgate.net/publication/220746311_Twitter_as_a_Corpus_for_Sentiment_Analysis_and_Opinion_Mining
- Pérez, L. (2015). ¿Cómo Ayuda la Metodología Scrum a la Gestión de Proyectos de TI?. Recuperado el 12 de noviembre de 2018 de <https://searchdatacenter.techtarget.com/es/cronica/Como-ayuda-lametodologia-Scrum-a-la-gestion-de-proyectos-de-TI>
- Power BI. (2017). Documentación de Power BI. Recuperado el 15 de diciembre de 2018 de <https://docs.microsoft.com/es-es/powerbi/>
- Queralt, E. (2014). ¿Qué es el Método SCRUM y que Beneficios y Ventajas va a Aportar a Mi Negocio? Recuperado el 27 de noviembre de 2018 de <https://www.openiebs.com/blog/que-es-scrum-beneficios-ventajas-paranegocio-agile-scrum/>
- Read, J. (2005). *Using emoticons to reduce dependency in machine learning techniques for sentiment classification*. Recuperado el 09 de diciembre de 2018 de <https://dl.acm.org/citation.cfm?id=1628969>
- Rouse, M. (2015). SQL Server. Recuperado el 04 de noviembre de 2018 de <https://searchdatacenter.techtarget.com/es/definicion/SQL-Server>
- Schwaber, K., Sutherland, J. (2016). La Guía de Scrum. *En La Guía Definitiva de Scrum*. (p.5-9).
- Steen, H. (2018). Como detectar sentimiento en la analítica de texto. Recuperado el 01 de diciembre de 2018 de <https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/how-tos/text-analytics-how-to-sentiment-analysis>

- Taboada, J., Cotos, J. (2005). *Sistemas de Información Medioambiental*. Recuperado el 19 de diciembre de 2018 de <https://books.google.com.ec/books?id=FEBhY2xmmT8C&pg=PA176&dq=que+es+un+flujos+de+trabajo&hl=es&sa=X&ved=0ahUKEwivxtOP5YTfAhXw01kKHArwBqkQ6AEITjAH#v=onepage&q=que%20es%20un%20flujos%20de%20trabajo&f=false>
- Venkata, L., Tanveer, A., Shajith, I., Shantanu, G., Mukesh, K. (2009). *Business Intelligence from Voice of Customer*. Recuperado el 16 de noviembre de 2018 de <https://ieeexplore.ieee.org/abstract/document/4812540>
- Wilson T, Wiebe J, Hoffman P. (2005). *Recognizing contextual polarity in phrase-level sentiment analysis*. Recuperado el 15 de noviembre de 2018 de <https://people.cs.pitt.edu/~wiebe/pubs/papers/emnlp05polarity.pdf>
- Xu T., Qinke P., Yinzhao C. (2012). *Identifying the semantic orientation of terms using S-HAL for sentiment analysis*. *Syst*, 35 (2), pp. 279-289
- Yang, C., Hsin-Yih Lin, K., Hsin-His, C. (2007). *Emotion classification using web blog corpora*. In *WI '07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 275–278, Washington, DC, USA. IEEE Computer Society.
- Yu, H., Hatzivassiloglou, V. (2003). *Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences*. Recuperado el 23 de noviembre de 2018 de https://www.researchgate.net/publication/215470804_Towards_answering_opinion_questions_Separating_facts_from_opinions_and_identifying_the_polarity_of_opinion_sentences.

ANEXOS

Índice de Abreviaturas

API:	Application Programming Interface
AS:	Análisis de Sentimiento
BI:	Business Intelligence
CRF:	Conditional Random Fields
CRISPDM:	Cross-Industry Standard Process for Data Mining
CRM:	Customer Relationship Management
DAX:	Data Analysis Expressions
EEQ:	Eléctrica de Quito
ETL:	Extract, Transform and Load
ERP:	Enterprise Resource Planning
IA:	Inteligencia Artificial
IBM:	International Business Machines
ML:	Machine Learning
MPN:	Microsoft Partner Network
MS:	Microsoft
NLP:	Neuro-Linguistic Programming
OM:	Minería de Opiniones
POS:	Etiquetado de Parte Voz
RDBMS:	Sistema de Gestión de Bases de Datos Relacionales
RGPD:	Reglamento General de Protección de Datos
ROC:	Receiver Operating Characteristic
SVM:	Support Vector Machine
T-SQL:	Transact -SQL

