



FACULTAD DE INGENIERÍA
INGENIERÍA CIVIL INFORMÁTICA

**Análisis de sentimientos en redes sociales orientado a la
percepción de la calidad de servicios de internet, redes
móviles, tv cable y electricidad.**

Proyecto de Título para optar al grado de Ingeniero Civil en Informática.

Autor:

Tomás Ignacio Garcés Chaparro

Profesor guía: Carlos Gómez-Pantoja, PhD.

Santiago de Chile

Mayo, 2019

Agradecimientos

A mi madre, Nancy Chaparro Espinoza, y a mi padre, Leonel Garcés Bórquez, por el apoyo, el amor, la confianza y las enseñanzas que me han entregado siempre. Su dedicación y esfuerzo ha sido y siempre será fuente de inspiración y motivo de admiración.

Al equipo de Scikit Learn por ofrecer una herramienta de uso libre donde el desempeño, documentación y facilidad de uso son claves.

Índice general

CAPÍTULO 1	10
INTRODUCCIÓN	10
MOTIVACIÓN	10
CONTRIBUCIÓN DE LA TESIS	12
ESTRUCTURA DE LA TESIS	13
CAPÍTULO 2	15
DESCRIPCIÓN DEL PROBLEMA	15
2.1 CONTEXTUALIZACIÓN	15
2.1.1 <i>Diagrama de Ishikawa</i>	16
2.2 ENUNCIADO DEL PROBLEMA	16
2.3 OBJETIVOS E HIPÓTESIS	17
2.3.1 <i>Objetivo general</i>	17
2.3.2 <i>Objetivos específicos</i>	17
<i>i. Tabla de métricas</i>	18
	3

<i>ii. Hipótesis</i>	18
2.4 METODOLOGÍA DE TRABAJO	19
2.4.1 El método científico	19
2.4.2 El método del estudio de caso	20
2.5 Herramientas y entorno de trabajo	21
2.6 ALCANCES	21
CAPÍTULO 3	22
MARCO TEÓRICO	22
3.1 TWITTER	22
3.2 WEB SCRAPING	23
3.3 BIG DATA	24
3.4 DATA MINING	25
3.5 INTELIGENCIA ARTIFICIAL	25
3.6 MACHINE LEARNING	27
3.6.1 Supervised Learning	27
3.6.2 Unsupervised Learning	29
3.6.3 Reinforcement Learning	30
3.7 Modelos Clasificadores	30

3.7.1 Naive Bayes	30
3.7.1.1 Multinomial Naive Bayes:	31
3.7.2 Decision Trees	32
3.7.3 SVM	33
3.7.4 Logistic Regression	33
3.8 Métricas de desempeño para clasificadores	35
3.8.1 Accuracy	35
3.8.2 Matthews Correlation Coefficient	35
3.8.3 ROC AUC SCORE	36
3.8.4 Precision	37
3.8.5 Recall	37
3.8.6 F1-Score	37
CAPÍTULO 4	38
ESTADO DEL ARTE	38
4.1 PERCEPCIÓN DE CALIDAD DE SERVICIOS POR LOS CONSUMIDORES	38
4.1.1 <i>Supermarket self-checkout service quality, customer satisfaction, and loyalty: Empirical evidence from an emerging market</i>	38
4.1.2 <i>How to Improve Perceived Service Quality by Increasing Customer Participation</i>	42

4.2 ANÁLISIS DE SENTIMIENTO U OPINIÓN EN RRSS.	45
4.2.1 Tesis: Herramientas de análisis de opinión en redes sociales virtuales.	45
4.2.2 Techniques and applications for sentiment analysis	46
4.2.3 New avenues in opinion mining and sentiment analysis.	54
4.2.4 State of the art for sentiment analysis.	62
4.3 LO QUE OFRECE EL MERCADO	73
4.3.1 ProCalidad	74
4.3.2 Servqual	74
4.3.4 Brand24	76
4.3.5 Hootsuite	77
4.3.6 Natural Language de Cloud	77
4.3.7 IBM Cloud – Natural Language Understanding	79
4.3.8 Amazon Comprehend	80
4.4 COMPARACIÓN Y ANÁLISIS	81
CAPÍTULO 5	82
Descripción de la solución	82
5.1 Descarga de Twitts	83
5.2 Entrenamiento y Testing	85

5.3 Análisis de sentimientos	87
5.4 Muestreo de resultados	88
Capítulo 6	89
6. Pruebas y resultados de la solución	89
6.1 Pruebas y Resultados para Claro Chile	89
6.1.1 Etiquetado para Claro Chile	89
6.1.2 Entrenamiento y evaluación de modelos predictivos para Claro Chile	92
6.1.2.1 MultinomialNB	94
6.1.2.2 DecisionTreeClassifier	95
6.1.2.3 SVM	96
6.1.2.4 LogisticRegression	97
6.1.2.5 Análisis de modelos	98
6.1.2.6 Comparación de modelos	101
6.1.3 Modelos predictivos en data no etiquetada para Claro Chile	102
6.1.3.1 LogisticRegression	103
6.1.3.2 SVM	105
6.1.3.3 Comparación LogisticRegression vs SVM	106
6.1.4 Análisis de Sentimientos	108

6.1.4.1 Septiembre-2018	109
6.1.4.2 Octubre-2018	116
6.1.4.3 Noviembre-2018	120
6.1.4.4 Diciembre-2018	124
6.2 Pruebas y Resultados para Enel Chile	128
6.2.1 Etiquetado para Enel Chile	128
6.2.2 Entrenamiento y evaluación de modelos predictivos para Enel Chile	132
6.2.2.1 MultinomialNB	133
6.2.2.2 DecisionTreeClassifier	134
6.2.2.3 SVM	135
6.2.2.4 LogisticRegression	136
6.2.2.5 Análisis de modelos	137
6.2.2.6 Comparación de modelos	141
6.2.3 Modelo predictivo LogisticRegression en data no etiquetada para Enel Chile	142
6.2.4 Análisis de Sentimientos	146
6.2.4.1 Septiembre-2018	147
6.2.4.2 Octubre-2018	151
6.2.4.3 Noviembre-2018	155

6.2.4.4 Diciembre-2018	160
Capítulo 7	164
7. Conclusiones	164
7.1 Trabajos Futuros	168
7.1.1 Machine Learning	168
7.1.2 Análisis de Sentimientos	169
7.1.3 Limpieza, normalización y tratamiento de texto en lenguaje natural	171
Bibliografía	174
Anexos	178
Anexo 1: Planificación utilizada en el trabajo de tesis.	178
Anexo 2: Experimentación de aprendizaje no-supervisado (descartado)	178
Clustering	178
K-means	180

Capítulo 1

1. Introducción

1.1. Motivación

En la última década, y a medida que pasa el tiempo, cada vez más el Internet es considerado una parte integral de nuestra vida diaria, entregando acceso a información, comunicación global y fuentes de entretenimiento.

Desde su creación (internet), los desarrolladores están en una búsqueda constante de explotar al máximo las cualidades de este medio de comunicación, implementando softwares que entreguen a los usuarios una conectividad de mejor calidad (estable, manejable y rápida).

Con esta misión en mente, desde finales de los 90', se han visto diversas plataformas que buscan conectar personas. En sus inicios, aplicaciones como AOL Instant Messenger (1997) daban el primer paso, ofreciendo un chat en tiempo real entre contactos. Más adelante, llegaría MySpace (2003), plataforma online que no sólo permitía conectarte con tus contactos, sino que permitía al usuario crear su perfil y generar contenido.

El año 2004 se lanza Facebook, concebida originalmente como una plataforma para conectar estudiantes universitarios, hoy en día con cerca de 2.167 millones de usuarios activos al mes (la red social con mayor cantidad de usuarios activos al mes actualmente).

Dos años después (2006) se inaugura Twitter. Una red social de microblogging, cuya misión principal, a diferencia de Facebook, es conectar ideas, opiniones y temas en lugar de conectar personas.

Y así, a medida que pasa el tiempo, han ido saliendo al mercado distintas RRSS (redes sociales), como Whatsapp (1.300 millones de usuarios activos al mes) e Instagram (más de 800 millones de usuarios activos al mes), todas con muchos parecidos, pero también diferencias.

El distinto enfoque de cada RRSS, y por consecuencia, lo diferente de los perfiles de los usuarios que son parte de una red social u otra, es el motivo de mayor peso al elegir la red social en la que se trabajaría para esta tesis.

Dado que lo que se busca en este trabajo de tesis es automatizar el análisis de sentimiento en base a una opinión en redes sociales, se escogió Twitter como plataforma de la que se extraerá la data. Esto dado las facilidades que ofrece para el acceso y almacenamiento de la información disponible.

Una vez definida la red social en la que se trabajará, solo queda preguntar: ¿Es valiosa la información disponible?, ¿Se puede acceder a esa información?, ¿Es posible automatizar la extracción y análisis de miles y miles de publicaciones?, ¿Es posible, mediante a lo mencionado anteriormente, orientar el desarrollo de una empresa a un enfoque centrado en el grado de satisfacción del cliente?, la respuesta a todas esas preguntas es sí, y se pretende demostrar con este trabajo de tesis.

Mediante el uso de técnicas como *Web Scraping* para la búsqueda y descarga de información, desarrollo de códigos para el tratamiento del texto, y

clasificadores y algoritmos de IA (inteligencia artificial) para el análisis de sentimientos y automatización en base a aprendizaje supervisado, para finalmente representar mediante el uso de gráficos los resultados obtenidos.

Para el análisis de sentimiento, es indispensable que se adapte el algoritmo al contexto en el que se esté trabajando. Esto, dado que, al momento de limpiar el texto, dependerá del contexto las palabras que se eliminen por considerar que no aportan información o sentimiento a la oración.

Dado el enfoque de la tesis, considerando la importancia de tomar en cuenta la opinión de las personas en nuevas vías como Twitter, y cómo servicios como internet han pasado a ser indispensables, es que se ha decidido analizar la opinión sobre la calidad del servicio que entregan distintas empresas del rubro internet, cable, redes móviles y electricidad. Para esto, se ha verificado manualmente la cantidad de publicaciones que se realizan sobre estos temas en Twitter, y se comprueba que tiene una actividad constante, por lo que se decide extraer la información de mensual.

A continuación, en el presente informe se presentará el completo desarrollo de la idea de tesis, explicando metodologías, aspectos técnicos, teorías y resultados.

1.2. Contribución de la tesis

La contribución del proyecto de título es la creación de un software que permita recopilar información pública de Twitter, la procese y entregue de manera

resumida los resultados del análisis de sentimiento realizado. Específicamente las contribuciones de este trabajo son:

- Implementación de un Bot que extrae “*Tuits*” asociados a una búsqueda específica, utilizando *Web Scraping*.
- Creación de un software que permite la limpieza, estandarización, y tratamiento del texto en idioma español.
- Desarrollo de software de IA (inteligencia artificial) que realiza el trabajo de aprendizaje para la detección de sentimientos. Dicho software desarrollado de manera escalable que permita extrapolar su aplicabilidad a distintas temáticas (política, TV, juegos, etc....) y con un mayor trabajo, a distintos idiomas.

1.3. Estructura de la tesis

El capítulo 2 contiene la descripción del problema a tratar, además de un análisis e información correspondiente al contexto a tratar. Junto con esto, se definen los objetivos del proyecto, además de su alcance e hipótesis.

En el capítulo 3 se presentan los términos y conceptos de carácter más técnicos que han sido necesarios para realizar y comprender el trabajo de tesis.

El capítulo 4 consiste en describir y analizar empresas, herramientas, estudios o trabajos que abarquen la misma temática con propósitos similares. Esto con el fin de comparar lo que puede ser encontrado hoy en día con la solución propuesta. Además, de nutrir tanto al estudiante en el desarrollo del mismo

como al lector con el propósito de estar mejor preparados para la comprensión del proyecto de tesis. Cabe señalar que dada la metodología ocupada, toda la documentación utilizada y citada en este trabajo han de asumirse como “la verdad”.

En el capítulo 5 se describe de manera detallada la solución y todas sus partes, junto con la explicación del análisis que justifica las decisiones tomadas que culminan en el flujo de trabajo construido.

El capítulo 6 consiste en la exposición y análisis de los resultados obtenidos al haber aplicado lo expuesto en el capítulo 5. Se ha procurado exponer los resultados con su correspondiente análisis y explicación de modo tal que permita la total comprensión por parte del lector tanto en los ¿Por qué? como en los ¿Para qué? que puedan surgir sobre lo desarrollado.

En el capítulo 7 se presentan conclusiones que den fin al trabajo realizado y al informe del mismo. Se expone un breve resumen de los resultados obtenidos y se comparan con los objetivos iniciales de la tesis. Se analiza el trabajo realizado en cuanto a desafíos y problemas enfrentados, en base a los cuales se proponen posibles trabajos futuros a realizar, a través de los cuales se busca mejorar los resultados obtenidos o dar mayor robustez a la solución.

En la bibliografía se exponen los documentos científicos, trabajos investigativos, tesis y otro documentos utilizados durante el desarrollo del trabajo de tesis.

Finalmente se adjuntan a modo de anexos la planificación de tareas versus el tiempo (carta gantt) y el tiempo real utilizado, además de una alternativa de

solución en base a aprendizaje no supervisado que fue descartado por no aportar a alcanzar los objetivos propuestos.

Capítulo 2

2. Descripción del problema

2.1 Contextualización

Twitter es una red social que cuenta con más de 320 millones de usuarios activos al mes y se caracteriza por concentrar opiniones de la gente (al estilo microblogging) respecto a distintos tópicos.

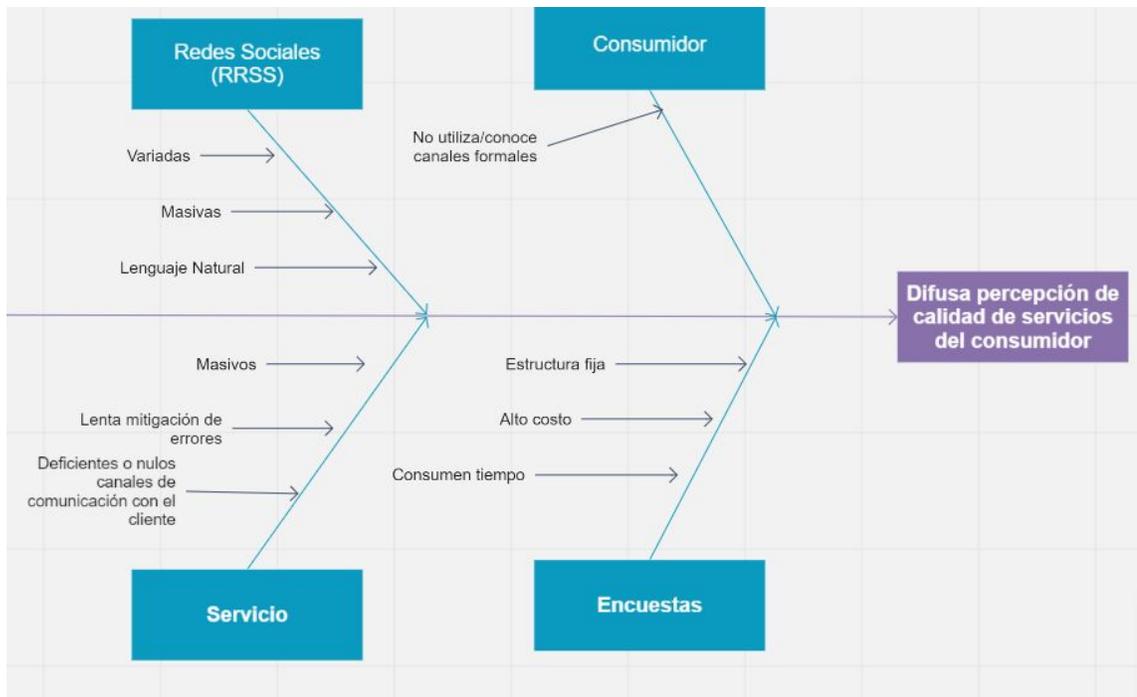
Los usuarios pueden publicar sobre lo que deseen en el momento que lo deseen, sin mayores restricciones más que la cantidad de caracteres del post (280 caracteres), llegando así de manera rápida y fácil a sus *followers* (*seguidores*).

Pero el alcance de lo que se publica en Twitter no termina con los seguidores. Esta red social se caracteriza por su contenido público, ya sea porque un seguidor “*retuiteó*” un post o porque dicho post cuenta con un *Hashtag*, el alcance puede ser de millones de usuarios.

El proyecto de título tiene como finalidad proponer e implementar un sistema automatizado de descarga de “*Tuits*”, limpieza de texto, análisis y detección de

sentimientos basado en un aprendizaje previo. Además, se mostrarán con gráficos y de manera resumida los resultados generales del análisis.

2.1.1 Diagrama de Ishikawa



*Figura 1: Diagrama de Ishikawa. Fuente: Elaboración propia.

2.2 Enunciado del problema

Internet es considerado hoy en día indispensable para que muchas tareas y procesos puedan ser llevados a cabo. Por esto, es un bien cada vez más accesible y se espera que en un futuro no muy lejano esté al alcance de todos y todas. Así, el uso de internet se ha expandido, y si en un principio se utilizó

únicamente con fines académicos, comerciales o militares, hoy en día es muy utilizado para conectar personas, entretener y compartir ideas.

En redes sociales como Twitter, se ven miles y miles de opiniones, pero solo pocas logran destacar y ser vistas y/o tomadas en cuenta, especialmente si no eres una persona popularmente conocida. A pesar de aquello, Twitter es una plataforma muy popular al momento de realizar críticas, quejas y denuncias de manera no oficial.

En base a lo anterior, y con la intención de incentivar el desarrollo de empresas del rubro de TV, internet y electricidad hacia un enfoque que priorice la satisfacción del cliente con el producto o servicio entregado, es que se pretende desarrollar un software que permita recopilar, analizar y resumir la opinión de los usuarios sobre las empresas proveedoras.

2.3 Objetivos e hipótesis

2.3.1 Objetivo general

Incentivar un enfoque empresarial orientado a la satisfacción del cliente a partir de una herramienta que realice un análisis automatizado de las opiniones manifestadas en redes sociales.

2.3.2 Objetivos específicos

- 1) Determinar un conjunto de herramientas para análisis de sentimiento para lenguaje natural en español.
- 2) Identificar los principales factores de reclamo en redes sociales.

- 3) Establecer un conjunto de indicadores para representar datos agrupados de RRSS

i. Tabla de métricas

Objetivos Específicos	Métrica/Unidad	VAM	CEM
OE1	Número de Herramientas	Desconocido	N>=3
OE2	Frecuencia de palabras	Desconocido	F>= 10%
OE3	Número de Indicadores	Desconocido	N>=3

**Tabla 1: Métricas de objetivos específicos. Fuente: Elaboración propia.*

ii. Hipótesis

Las vías formales por las cuales un consumidor debe manifestar su descontento o insatisfacción con el servicio por el que paga resultan ser largos trámites burocráticos que consumen tiempo y parecen no ofrecer potencial beneficio al consumidor.

Las redes sociales como Twitter son una potencial plataforma de agrupación de opiniones de los consumidores sobre los servicios que consumen y muestran la reputación de las empresas según la opinión de miles de clientes.

A través del desarrollo de la herramienta de análisis de sentimiento de Twitter, las empresas podrán conocer las opiniones de sus clientes de manera rápida, automática e inteligente, permitiendo mejorar la calidad del servicio entregado.

2.4 Metodología de trabajo

La metodología escogida para este trabajo es una mezcla de dos metodologías: el método científico y el método de estudio de caso. Si bien el método científico definirá principalmente el cómo llevar a cabo la investigación, es decir, los pasos a seguir; El método de estudio de caso se acerca más a los objetivos de esta investigación, por lo que se tomaron ciertos elementos de ella.

A continuación se abordan ambas metodologías, explicando sus componentes y funcionamiento:

2.4.1 El método científico

El método científico, así como todas las metodologías de investigación, ofrece un camino o guía de cómo alcanzar el objetivo deseado de manera objetiva y sistemática. En este caso particular, se compone de 6 pasos o etapas y debe cumplir con los siguientes requisitos:

- Empirismo: El estudio debe ser real y objetivo.
- Repetibilidad: El estudio debe ser confirmado al ser repetido.
- Aceptabilidad: El investigador presupone la aceptación de lo publicado anteriormente.
- Publicidad: Los descubrimientos deben darse a conocer.

Para cumplir con aquello se definen los siguientes pasos o etapas:

- Definición y planteamiento del problema: Pregunta para la cual no se encuentra respuesta. Es necesario que sea resoluble y debe ser formulado en términos adecuados.
- Formulación de la hipótesis: Es la “verdad provisional” o cómo se explica el problema a la luz de lo que se sabe. Las hipótesis se pueden formular como objetivos o resultados que se quieren conseguir.
- Recogida y análisis de datos: Comprobación empírica tras recogida de datos.

- Confrontación de los datos con la hipótesis: El investigador debe realizar un análisis comparativo entre lo que esperaba obtener y lo obtenido.
- Conclusiones y generalización de los resultados: Si los datos avalan la hipótesis será confirmada. En caso contrario se concluirá que en las circunstancias contempladas la hipótesis no ha sido confirmada y/o se devolverá a la segunda etapa proponiendo una nueva y coherente solución al problema.
- Nuevas predicciones: Esta etapa es añadida por algunos autores y hace referencia a nuevos problemas que surgirían de los resultados obtenidos.

2.4.2 El método del estudio de caso

El método de estudio de caso es una herramienta valiosa de investigación, y su mayor fortaleza radica en que a través del mismo se mide y registra la conducta de las personas involucradas en el fenómeno estudiado, mientras que los métodos cuantitativos sólo se centran en información verbal obtenida a través de encuestas por cuestionarios (Yin, 1989).

Además, en el método de estudio de caso los datos pueden ser obtenidos desde una variedad de fuentes, tanto cualitativas como cuantitativas; esto es, documentos, registros de archivos, entrevistas directas, observación directa, observación de los participantes e instalaciones u objetos físicos (Chetty, 1996).

Por otra parte, Yin (1994, citado en Chetty (1996)) argumenta que el método de estudio de caso ha sido una forma esencial de investigación en las ciencias sociales y en la dirección de empresas, así como en las áreas de educación, políticas de la juventud y desarrollo de la niñez, estudios de familias, negocios internacionales, desarrollo tecnológico e investigaciones sobre problemas sociales.

Para la aplicación de esta metodología a este trabajo se definió:

- Conducta de personas a ser estudiada: La manifestación del pensar acerca de la percepción de la calidad de servicios de empresas proveedoras de ISP, TV y electricidad en redes sociales.

- Fenómeno estudiado: Las empresas estudiadas ofrecen servicios pagados que prometen: disponibilidad, calidad (performance) y honestidad, frente a lo que los/las clientes/as manifiestan incumplimiento de servicios, cobros indebidos, servicios deficientes y hostigamiento comercial de parte de la empresa.
- Fuente de datos: Plataforma de redes sociales, específicamente Twitter.

2.5 Herramientas y entorno de trabajo

Las herramientas utilizadas para el desarrollo del proyecto de título son:

- Sistema Operativo Ubuntu 16.04.
- Sistema Operativo Windows 10.
- Lenguaje de programación Python 3.0.
- Sublime Text.
- Microsoft Office 365 (Word, Excel & PowerPoint).

- Google Drive.

2.6 Alcances

El desarrollo del proyecto de tesis se limitará a lo siguiente:

- El software de extracción de información será desarrollado utilizando *web scraping* en Python 3 y funcionará en cualquier sistema Linux con las bibliotecas señaladas correctamente instaladas.
- El software de extracción de información funcionará sólo en la página de Twitter y con una búsqueda particular.

- El software de análisis de sentimientos será desarrollado en Python 3 y funcionará sólo con comentarios en español, en el contexto de esta tesis (servicios) y en cualquier sistema Linux con las bibliotecas señaladas correctamente instaladas.
- El software de análisis de sentimientos debe ser desarrollado de manera tal que sea escalable y, con cierto trabajo, aplicable a otros contextos y/o idiomas.
- Los resultados serán desplegados en forma de gráficos, tablas u otros elementos de soporte visual.

Capítulo 3

3. Marco Teórico

En este capítulo se abordarán los principales temas de carácter teórico que fueron necesarios para desarrollar (y, por ende, entender) este trabajo de tesis.

3.1 Twitter

Es una red social de microblogging fundada el año 2006 que hoy en día cuenta con más de 320 millones de usuarios activos al mes, lo que la transforma en una de las redes sociales más populares de la historia. Como toda red social es hoy en día una popular forma de conectar y comunicar personas a través de una plataforma que permite compartir contenido en formato de texto e imágenes. A diferencia de otras redes sociales como Facebook, el uso principal de Twitter va más allá de conectar personas. Twitter es considerada una red social de debate, puesto que busca que los usuarios compartan sus ideas,

agrupándolas a través de palabras claves (o *hashtags*), permitiendo así generar espacios de discusión virtuales.

Twitter es quizás la red social más utilizada al minuto de realizar trabajos investigativos, especialmente de data masiva. Esto debido a que no sólo es una red popular, tiene una restricción de Twitts de 240 caracteres y el acceso a su información es completamente público, por lo que acceder y descargar la información que el sitio aloja resulta sencillo.

Diversos estudios se han realizado en base a la data que Twitter ofrece, desde el análisis sobre los usuarios y sus comportamientos, hasta predicciones de comportamiento, ubicación geográfica del “*tuit*” e incluso predicciones políticas.

3.2 Web Scraping

Es una técnica utilizada para extraer data de un sitio web de manera automatizada. Es utilizada muy frecuentemente en contextos de *big data* y *data mining*. Debido a su utilidad, existen códigos y bibliotecas de acceso libre que cuentan con todas las funciones necesarias para acceder a los sitios web, extraer la información e incluso visualizar la interacción del “bot” con el sitio web. Con el material disponible, es posible simular el comportamiento humano con acciones como scroll, click, rellenar campos, completar formularios, etc.

Actualmente, dichas bibliotecas y códigos se encuentran disponibles para lenguajes de programación como Python y Java, y están implementados a alto nivel, por lo que resulta fácil de trabajar para cualquier programador medianamente experimentado.

3.3 Big Data

Es el término utilizado para referirse a una cantidad de datos voluminosa y compleja que los softwares de procesamiento de datos tradicionales no pueden tratarlos. El tamaño de la data con la que se va a trabajar conlleva grandes desafíos, que son: capturar la data, almacenarla y analizarla.

Hoy en día es un área de investigación sin límites y en constante ascenso. El año 2014 Google procesaba más de 24 petabytes de data al día y el mundo enviaba 400 millones de tuits al día. Toda esa información requiere ser procesada por herramientas de análisis y algoritmos complejos para entregar información con valor y significado.

Data de semejante magnitud ya no es sorpresa para ninguna empresa, ya que son muchos los campos en los que se puede trabajar con *Big Data*, tales como: bancos, agricultura, química, *data mining* (que será abarcado en el punto 3.4), *cloudcomputing* (Computación en la nube), finanzas, marketing, transporte, salud, etcétera.

La examinación y procesamiento de vastos sets de datos es denominada “Analítica de *Big Data*” (o *Big Data Analytics*). La importancia de este rubro es que, los resultados de este análisis pueden guiar a un marketing más efectivo, nuevas oportunidades de ingresos, mejor servicio al cliente, mejoras en la eficiencia operacional y ventaja competitiva sobre empresas o instituciones de la competencia. Razones que impulsan a las empresas a enfocarse e invertir más en la analítica de *Big Data* para su negocio.

3.4 Data Mining

Es el término utilizado para referirse a un campo interdisciplinario, que consiste en utilizar distintos métodos de distintas áreas de investigación (principalmente *Machine Learning* (abarcado en 3.6) y estadísticas) para extraer conocimiento o información valiosa de data del mundo real.

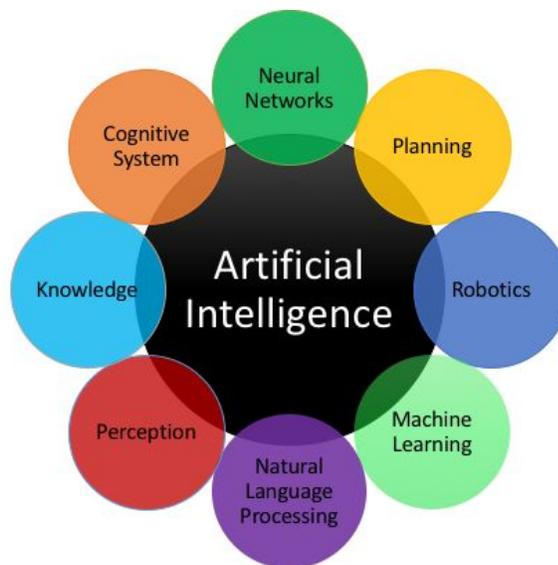
Data mining se refiere al paso clave de un proceso más amplio llamado “descubrimiento de conocimiento en bases de datos” (*Knowledge Discovery in Databases*). *Data Mining* es el proceso en el que efectivamente se extrae conocimiento de los datos, mientras que el descubrimiento de conocimiento incluye varios pasos de pre-procesamiento (preparación de la data) y post-procesamiento (refinamiento del conocimiento). El objetivo de los métodos de preparación de la data es transformar la data para facilitar la aplicación de uno o muchos algoritmos de *data mining*. Por el otro lado, los métodos de refinamiento de conocimiento buscan validar y refinar el conocimiento descubierto.

3.5 Inteligencia Artificial

Durante años se ha hablado sobre la Inteligencia Artificial (IA). En un comienzo abarcada por la ciencia ficción, en películas y comics, hoy en día una realidad que parece no tener límites y que genera debate en el campo científico.

Es el término utilizado para referirse a un área de investigación y desarrollo ligada a las ciencias de la computación y la robótica, que comprende toda inteligencia demostrada por las máquinas. Así como la inteligencia natural la

poseemos humanos y animales, es posible su simulación a través de códigos, permitiendo a máquinas emular las facultades únicas de razonamiento con las que cuenta el ser humano (dicha máquina es llamada agente inteligente). Por lo tanto, el principal objetivo de este campo es darle a la máquina características de conocimiento, razonamiento, resolución de problemas, percepción, aprendizaje, planificar, habilidad de manipular y mover objetos, etcétera.



**Figura 2: Artificial Intelligence, Fuente:*

<https://www.besttechguru.com/future-technology-explained-what-is-artificial-intelligence/>

Otras áreas de investigación ligadas con la AI son big data, data mining, machine learning, knowledge engineering y robótica.

La comprensión y avance de la inteligencia artificial afecta a todos y todas. Hoy en día es difícil pensar en contextos en los que no esté presente. Ya sea en

supermercados, transporte, salud, videojuegos, educación, redes sociales, bancos e incluso instituciones militares. Su aplicabilidad es multidisciplinaria y su uso es prácticamente ilimitado, y sus beneficios son bastante significativos, ya que ayuda a optimizar tiempo y dinero además de guiar y orientar el desarrollo de ciertos procesos en el área industrial y empresarial.

3.6 Machine Learning

Es una rama de la inteligencia artificial, producto de las ciencias de la computación, que pretende estudiar el desarrollo de máquinas capaces de aprender. Lo que se busca, es estudiar el reconocimiento de patrones y razonamiento probabilístico para que así la máquina pueda resolver problemas de un contexto específico a pesar de que varíe el input (en su contenido más no en su estructura). El *machine learning* en general se entiende como el estudio del proceso en el que se colecta la data y se desarrolla una serie de algoritmos que permite a la máquina aprender a resolver los problemas a los que se enfrenta.

Para lograrlo, existen 3 tipos de *machine learning*, que serán descritos a continuación.

3.6.1 Supervised Learning

El aprendizaje supervisado consiste en enseñarle a la máquina a partir de data clasificada. La clasificación de la data consiste en tener un set de datos en lenguaje natural y clasificarlo en alguna clase. Comúnmente estas clases son: positivo, negativo y neutral. Este trabajo se realiza de manera manual y

minuciosa ya que dependiendo de qué tan bien estén clasificados los datos y la cantidad de datos clasificados que se posean será la calidad del aprendizaje de la máquina y por lo tanto el desempeño posterior. Es importante que los datos a clasificar pertenezcan al contexto real en el que se desempeñará la máquina o sean lo más parecido a la realidad posible. Una gran cantidad de datos muy variada puede nutrir de mejor manera los algoritmos estadísticos que generarán el modelo predictivo a testear.

Posteriormente, se realiza lo necesario para la etapa de entrenamiento. Una vez que se tiene la data clasificada, se divide el total para ser utilizado en dos etapas diferentes del aprendizaje: el entrenamiento y la prueba. Para fraccionar los datos, una diversidad de estudios y docentes recomiendan distribuir 80% para entrenar y 20% para probar.

Se denomina supervisado ya que, el modelo predictivo debe ser validado con el set de prueba, que está clasificado al igual que el set de entrenamiento. Para generar el modelo predictivo, existen algoritmos de conocimiento público, que utilizan distintas heurísticas, metodologías y criterios para “aprender” del set de pruebas, estos son denominados modelos de clasificación y algunos de los más conocidos son: Clasificador Naive Bayes, Máquina de soporte vectorial (SVM por sus siglas en inglés) y regresión logística.

La implementación de estos clasificadores requiere en su output (salida), ciertos indicadores que permiten establecer qué tan bueno fue el desempeño del clasificador.

Algunos indicadores que pueden ser analizados a simple vista son la precisión (cuántos fueron clasificados correctamente) y el recall (de los clasificados positivos, cuántos eran realmente positivos).

3.6.2 Unsupervised Learning

En español conocido como aprendizaje no supervisado, es un tipo de *machine learning* que requiere una entrada (datos que se ingresan al modelo de clasificación) menos trabajada en comparación al aprendizaje supervisado. Esto ya que en el aprendizaje no supervisado la data de entrada fue tratada pero no está clasificada. Para aprender de datos no clasificados, la máquina debe procesar la información e ir clasificando o agrupando en base a las similitudes y diferencias de la información. Un modelo de clasificación muy conocido en aprendizaje no supervisado es el Clustering y se explicará a modo de ejemplo.

El clustering consiste en agrupar la data disponible en base a sus diferencias y similitudes. Cada grupo es denominado cluster. Por ejemplo, si se desea identificar si un comentario en redes sociales es o no político y se cuenta con datos clasificados previamente, se deberá utilizar aprendizaje no supervisado para entrenar a la máquina. La máquina analiza toda la data y la representa en un plano cartesiano virtual n-dimensional. Se calcula la distancia que separa a todos los puntos entre sí y son agrupados. Luego, con los datos agrupados, solo queda asignar la clase correspondiente a cada grupo.

3.6.3 Reinforcement Learning

Conocido como aprendizaje por refuerzo y es un tipo de *machine learning* más complejo de entender e implementar que los otros dos. En este tipo de aprendizaje, la máquina interactúa directamente con el ambiente o contexto real en el que se trabajará. La máquina realiza distintas acciones y analiza el resultado de su interacción. Si esta es positiva, recibe una recompensa (reward +1) y si es negativa recibe un castigo (reward -1).

Un ejemplo es el caso de enseñarle a una máquina a jugar Super Mario Bros. La máquina recibe un primer estado, que consiste en el frame inicial del juego. La máquina realiza alguna de las acciones disponibles para con el contexto y recibe el siguiente frame del juego. Si el personaje sigue vivo, recibe un reward +1. Si el personaje muere, recibe un reward -1. Este proceso entrega como salida una secuencia de estados, acciones y rewards.

3.7 Modelos Clasificadores

3.7.1 Naive Bayes

Uno de los modelos clasificadores utilizados para este trabajo es el Multinomial Naive bayes, pero para definirlo, es necesario primero explicar qué es Naive bayes.

En *machine learning*, los clasificadores Naive Bayes (o bayes ingenuo) son una familia de clasificadores probabilísticos basados en aplicar el teorema de bayes con fuertes suposiciones de independencia entre sus *features*. Ha sido intensamente estudiado desde los 1950s y se mantiene como un popular

método base para la categorización de texto, especialmente si es combinado con métodos más avanzados como SVM.

El teorema de bayes, en teoría probabilística y estadísticas, describe la probabilidad de un evento, basado en conocimiento previo de condiciones que pueden estar relacionadas con dicho evento (H. Zhang, 2004. The optimality of Naive Bayes).

Matemáticamente, el teorema de bayes se define de la siguiente manera:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

**Figura 3: Fórmula matemática del teorema de bayes. Fuente: Elaboración Propia.*

Dicha fórmula se lee de la siguiente manera: “La probabilidad de que ocurra A dado B es igual a la probabilidad de que ocurra B dado A, por la probabilidad de que ocurra A, dividido en la probabilidad de que ocurra B”.

3.7.1.1 Multinomial Naive Bayes:

Es una versión más de la familia de clasificadores bayes, aplicada a la distribución multinomial. Una distribución multinomial hace referencia a que si, por ejemplo, en una distribución binomial se refiere a experiencias con resultados dicotómicos (claro u oscuro, vivo o muerto, positivo o negativo, etc...), en una distribución multinomial se habla de experiencias con r posibles resultados, cada uno con una probabilidad asignada (H. Zhang, 2004. The optimality of Naive Bayes).

El modelo de clasificación MultinomialNB implementa el algoritmo de naive bayes para data distribuida de forma multinomial y es una de las dos clásicas variaciones de naive bayes para la clasificación de texto. La distribución es parametrizada por vectores $\theta_y = (\theta_{y1}, \dots, \theta_{yn})$ para cada clase y , donde n es el número de *features* y θ_{yi} es la probabilidad $P(x_i|y)$ de que el *feature* i aparezca en un caso perteneciente a la clase y .

El parámetro θ_y es estimado a partir de una versión relajada de la máxima verosimilitud.

3.7.2 Decision Trees

Los Decision Trees, o árboles de decisión en español, es un método de aprendizaje supervisado no paramétrico utilizado para clasificación y regresión. El objetivo es crear un modelo que predice el valor de una variable objetivo por medio de aprendizaje de reglas de decisión simples inferidas de la data de las *features* o variables.

Algunas ventajas de los árboles de decisión son:

- Simples de entender e interpretar. Los árboles pueden ser visualizados.
- Requieren poca data para su preparación. Otras técnicas usualmente requieren normalización de la data, variables “tontas” deben ser creadas y valores en blanco deben ser removidos. Note que este módulo no permite valores faltantes.
- Capaz de manejar tanto data numérica como categórica. Otras técnicas usualmente son especializada en analizar datasets que cuentan con sólo un tipo de variables.
- Capaz de manejar problemas con múltiples salidas.

Entre las desventajas de los árboles de decisión destacan:

- Crean árboles imparciales si una de las clases domina. Por lo tanto, se recomienda balancear los datasets previo a ingresarlos al modelo.
- Los árboles de decisión pueden ser inestables debido a que pequeñas variaciones en la data pueden resultar en la generación de un árbol completamente diferente. Este problema es mitigado utilizando los árboles de decisión dentro de un conjunto.
- Situaciones de aprendizaje de árboles de decisión pueden crear árboles sobre-complejizados que no generalizan bien la data. Esto se llama *overfitting*. Mecanismos como podado, declarar el número mínimo de muestras requeridos al nodo hoja o declarar la profundidad máxima de un árbol son necesarios para evitar este problema.

El clasificador de árboles de decisión, así como con otros clasificadores, en la biblioteca Scikit Learn es capaz de realizar clasificación multi-clase en el dataset.

La implementación de este clasificador toma como entrada dos arreglos: un arreglo “X” que almacena las instancias de entrenamiento, y un arreglo “Y” de valores enteros que almacena las etiquetas de clase de las instancias de

entrenamiento (Pedregosa et al. 2011. “Scikit Learn: Machine Learning in Python”).

3.7.3 SVM

Las máquinas de soporte vectorial (SVM por sus siglas en inglés) son un conjunto de métodos de aprendizaje supervisado utilizados para clasificación, regresión y detección de anomalías o valores atípicos.

Matemáticamente hablando, una máquina de soporte vectorial construye un hiperplano o un set de hiperplanos en un espacio dimensional alto o infinito. Intuitivamente, una buena separación es lograda por el hiperplano que tiene la mayor distancia al punto más cercano de la data de entrenamiento de cualquier clase (dicho plano también es llamado margen funcional), en general mientras mayor sea el margen menor es el error de generalización del clasificador.

SVM suele ser utilizado en casos de clases binarias o dicotómicas, pero permite el uso en escenarios multiclase. Para la clasificación multiclase, si n_class es el número de clases, entonces $n_class * (n_class - 1) / 2$ clasificadores son construidos y cada uno entrena con data de dos clases (Pedregosa et al. 2011. “Scikit Learn: Machine Learning in Python”).

Un ejemplo de la nomenclatura que se puede utilizar en svm es la siguiente:

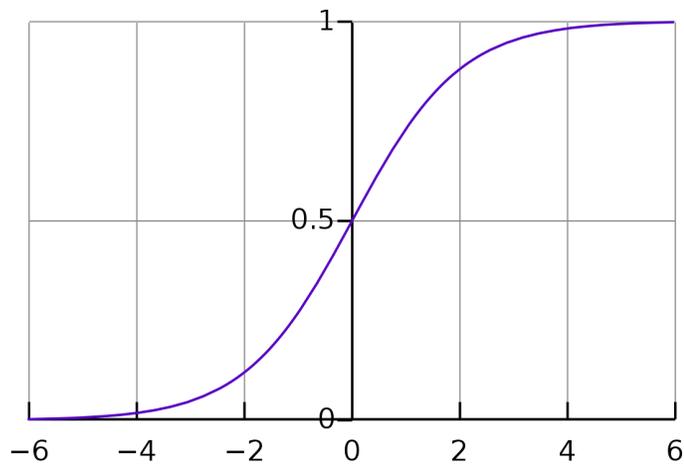
label feature_1:value_1 feature_2:value_2 feature_n:value_n.

Donde *label* es la etiqueta o clase correspondiente a la instancia, *features* son las componentes y *value* el valor de las componentes. En la clasificación de texto, *feature* representa cada palabra presente en el texto y *value* su frecuencia.

3.7.4 Logistic Regression

La regresión logística, a pesar de su nombre, es un modelo de clasificación lineal en lugar de una regresión como tal. Es también conocida en la literatura como *logit regression*, *maximum-entropy classification (Max Ent)*, o *log-linear classifier*. En este modelo, las probabilidades que describen la posible salida de una instancia particular son modelados utilizando una función logística (Pedregosa et al. 2011. “Scikit Learn: Machine Learning in Python”).

Una función logística es una curva sigmoid común (forma de S):



*Figura 4: curva de sigmoid. Fuente: Documentación Scikit.

, cuya ecuación es:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

*Figura 5: Fórmula matemática de Logistic Regression. Fuente: Documentación Scikit.

Donde:

- e = Número de Euler,
- X_0 = El x -valor del punto medio de sigmoid,
- L = El valor máximo de la curva,
- k = El índice logístico de crecimiento o lo abrupto de la curva.

En estos modelos, el valor objetivo (o de etiqueta) se espera que sea una combinación lineal de las variables de entrada. En notación matemática, si \hat{y} es el valor inferido:

$$\hat{y}(w, x) = w_0 + w_1 x_1 + \dots + w_p x_p$$

*Figura 6: Fórmula del valor inferido para Logistic Regression. Fuente: Documentación Scikit.

Aquí, $w_0, w_1, w_2, \dots, w_p$ son los coeficientes de regresión del modelo y son calculados por el estimador de máxima verosimilitud y x_1, x_2, \dots, x_p son los *features* de variables independientes.

La implementación de este modelo en Scikit Learn permite instancias binarias, una vs el resto, o regresión logística multinomial con regularización opcional L2 o L1.

3.8 Métricas de desempeño para clasificadores

3.8.1 Accuracy

Es el cómputo de predicciones correctas de un clasificador. Puede ser expresado como fracción o como un valor entero. En su representación fraccional, si se obtiene el valor 1,0 quiere decir que el desempeño fue perfecto y todas las predicciones fueron correctas. De lo contrario, si se obtiene el valor 0,0 es que todas las predicciones fueron incorrectas (Pedregosa et al. 2011. “Scikit Learn: Machine Learning in Python”).

Si \hat{y}_i es el valor predicho de la muestra i , e y_i es el valor real correspondiente, luego la fracción de predicciones correctas sobre n_{samples} como cantidad de muestras es:

$$\text{accuracy}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1(\hat{y}_i = y_i)$$

**Figura 7: Fórmula matemática del Accuracy. Fuente: Documentación Scikit.*

3.8.2 Matthews Correlation Coefficient

Es utilizado en machine learning para medir la calidad de la clasificación binaria (2 clases). Toma en cuenta los verdaderos y falsos positivos y negativos y es generalmente considerado como una medida balanceada la cual puede incluso ser utilizada cuando las clases son de tamaños muy diferentes. El MCC es un valor de coeficiente de correlación que fluctúa entre los valores +1 y -1. El valor +1 indica una predicción perfecta, el valor 0 representa una predicción aleatoria y el valor -1 representa una predicción inversa (Pedregosa et al. 2011. “Scikit Learn: Machine Learning in Python”).

Para el caso binario, el MCC se define como:

$$MCC = \frac{tp \times tn - fp \times fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}}$$

**Figura 8: Fórmula matemática de Matthews Correlation Coefficient. Fuente: Documentación Scikit.*

Donde:

- tp = verdaderos positivos,
- tn = verdaderos negativos,
- fp = falsos positivos,
- fn = falsos negativos.

Para el caso multiclase, donde el número de clases posibles es mayor a 2, existe una generalización del MCC que se define en términos de la matriz de confusión C para K clases. Para simplificar la definición considere las siguientes variables intermedias:

- $t_k = \sum_i^K C_{ik}$ el número real de ocurrencia de la clase k,
- $p_k = \sum_i^K C_{ki}$ el número de veces que la clase k fue predicha,
- $c = \sum_k^K C_{kk}$ el número de muestras correctamente predichas,
- $s = \sum_i^K \sum_j^K C_{ij}$ el número total de muestras.

Luego, el MCC para casos multiclase se define como:

$$MCC = \frac{c \times s - \sum_k^K p_k \times t_k}{\sqrt{(s^2 - \sum_k^K p_k^2) \times (s^2 - \sum_k^K t_k^2)}}$$

**Figura 9: Fórmula matemática MCC para casos multiclase. Fuente: Documentación Scikit.*

3.8.3 ROC AUC SCORE

El ROCK AUC SCORE hace referencia sobre el valor del área bajo la curva ROC.

La curva ROC (receiver operating characteristic) es una muestra gráfica que ilustra el desempeño de un clasificador binario. Es creado a partir de la gráfica de la fracción de TPR (true positive rate) v/s la fracción de FPR (false positive rate), donde TPR son los verdaderos positivos por sobre todos los positivos, y FPR son los verdaderos negativos por sobre todos los negativos (Pedregosa et al. 2011. "Scikit Learn: Machine Learning in Python").

La función del AUC ROC es resumir la curva ROC en un único valor. Este valor fluctúa entre 0 y +1. Donde +1 indica una clasificación perfecta, 0.5 indica una clasificación aleatoria y 0 indica una clasificación inversa.

El uso más común del ROC AUC es para la comparación de modelos.

3.8.4 Precision

Es una métrica de desempeño que fluctúa entre los valores +1 y 0, donde +1 es el mejor y 0 el peor. Se define como la habilidad del clasificador de no etiquetar como positivo una instancia que es negativa (Pedregosa et al. 2011. "Scikit Learn: Machine Learning in Python"), y se calcula de la siguiente manera:

$$\text{Precision} = TP / (TP + FP)$$

3.8.5 Recall

Es una métrica de desempeño que fluctúa entre los valores +1 y 0, donde +1 es el mejor y 0 el peor. Se define como la habilidad del clasificador de encontrar todas las instancias positivas (Pedregosa et al. 2011. "Scikit Learn: Machine Learning in Python"), y se calcula de la siguiente manera:

$$\text{Recall} = TP / (TP + FN)$$

3.8.6 F1-Score

Es otra medición del accuracy de una prueba. Se emplea en la determinación de un valor único ponderado de la precision y recall (Pedregosa et al. 2011. "Scikit Learn: Machine Learning in Python").

Su valor se considera una media armónica que combina los valores de la precision y recall, de tal forma que:

$$F_1 = \left(\frac{\text{recall}^{-1} + \text{precision}^{-1}}{2} \right)^{-1}$$

*Figura 10: Fórmula matemática de F1-Score. Fuente: Documentación Scikit.

Capítulo 4

4. Estado del Arte

4.1 Percepción de calidad de servicios por los consumidores

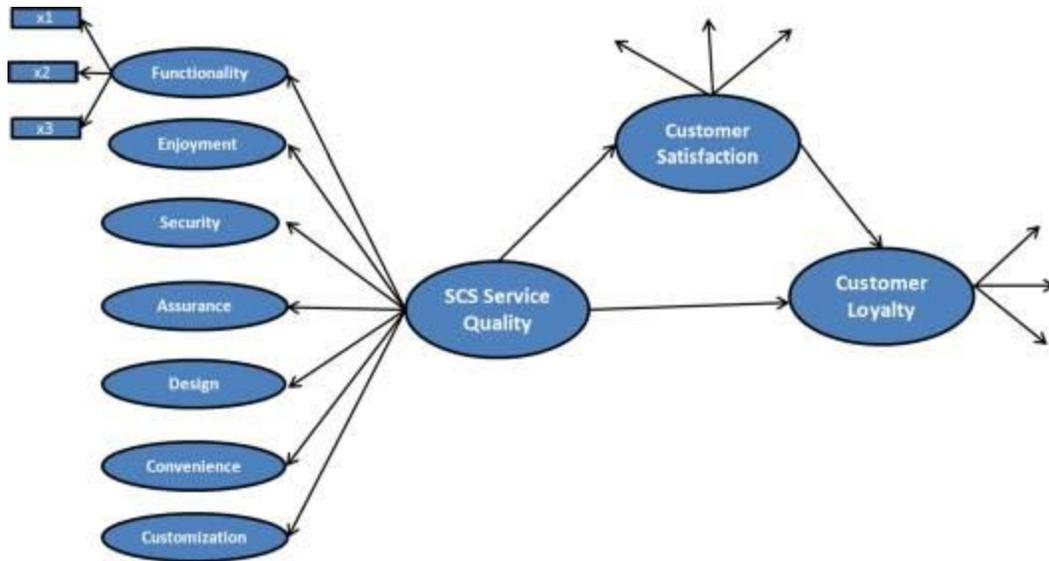
4.1.1 Supermarket self-checkout service quality, customer satisfaction, and loyalty: Empirical evidence from an emerging market

Este paper analiza y compara 4 herramientas de análisis de la calidad de servicio aplicado a un contexto de tecnologías de auto-servicio en tiendas y supermercados, más específicamente el auto-cobro al pasar por caja.

De acuerdo con estudios, si en 1999 en Estados Unidos sólo el 6% de los supermercados ofrecían sistemas self-checkout, la cifra aumenta a 35% en 2003 y llega a casi 95% en 2007. Un reporte del 2013 de IHL (empresa que entrega servicio de negocio inteligente personalizado), muestra que aproximadamente un 15-40% del valor de las transacciones diarias y el 12-30% del valor diario del dólar de supermercados son manejados por servicios check-outs.

A pesar de que la motivación de las tiendas es reducir costos, eficiencia, flexibilidad, productividad y mejora en las funciones de la empresa, es imperativo examinar la experiencia de compra del cliente, para mejorar el servicio entregado, lograr satisfacción del cliente y ganar con esto su lealtad.

Koten y Armstrong (1996) definen la satisfacción del cliente como “el nivel de cómo se siente una persona comparando el desempeño percibido del producto o servicio versus las expectativas propias del consumidor”. Bitner y Zeithaml (2003) declararon que la satisfacción del cliente es la evaluación de un producto o servicio en términos de si cumple sus necesidades y expectativas o no.



*Figura 11: Modelo conceptual de la investigación. Fuente: Demirci, F. Kara, A. (2014).

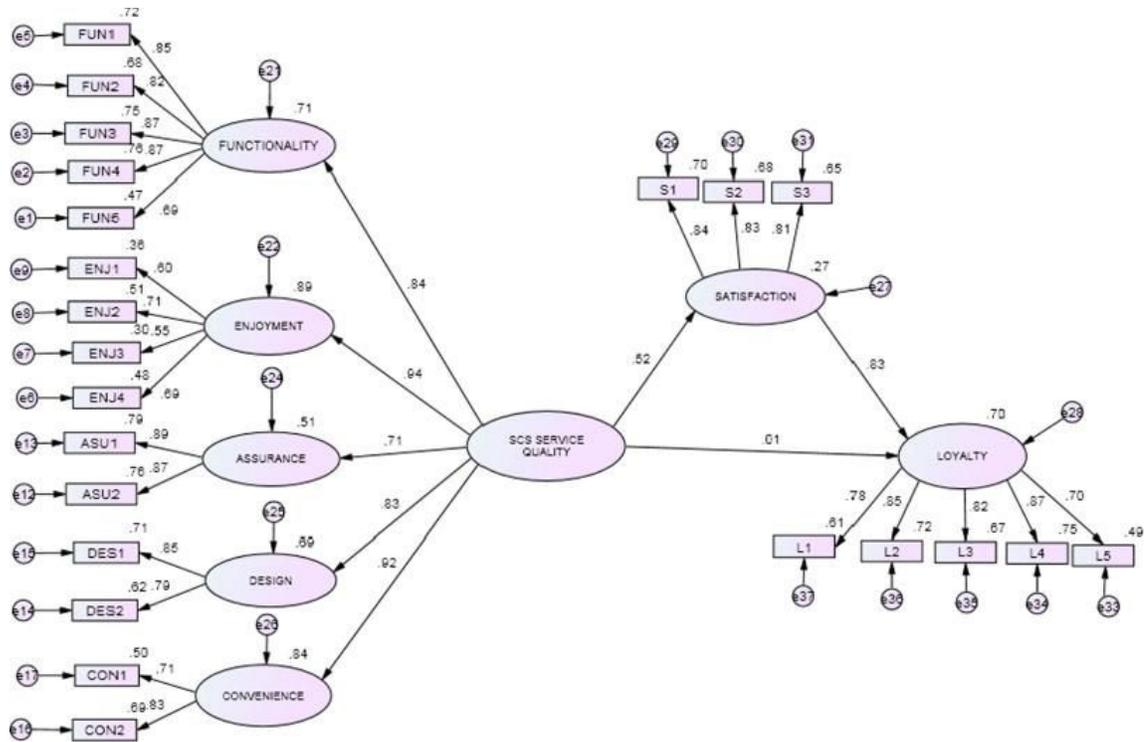
SERVQUAL y SERVPERF son escalas originalmente diseñadas para posicionar la interacción cliente-empleado, pero no tecnologías de auto-servicio. Por esto, investigadores argumentan que estas no puedan representar las diferentes facetas de la calidad de servicio de sistemas self-checkout.

Durante la última década, se han realizado distintos intentos para desarrollar una escala de medición de calidad para el contexto específico de servicios basados en tecnología. Por ejemplo, Zeithaml et al. (2002), produjo 11 dimensiones para evaluar la entrega de servicio de calidad de servicios

electrónicos. Estas dimensiones incluyen: acceso, facilidad de navegación, eficiencia, flexibilidad, confiabilidad, personalización, seguridad, responsividad, garantía/seguridad, estética de sitio, y conocimiento de precios. Así también, Barnes y Vidgen (2001) han extendido la escala SERVQUAL y establecieron un índice WebQual incluyendo las siguientes 7 dimensiones: confiabilidad, competencia, responsividad, acceso, credibilidad, comunicación y comprensión del individuo.

Para medir la satisfacción del cliente hemos adoptado las escalas de “Bloemer and Ruyter (1998)”, las cuales miden la satisfacción usando escalas del tipo “Likert” (*Overall, I am very satisfied with the company; This store confirms my expectations; I'm really satisfied with the service quality of this store; I'm really satisfied with the service quality of self-checkout in this store*).

Finalmente, la lealtad del cliente fue medida usando la escala adoptada por “Cronin et al (2000)”, la cual mide la lealtad usando cinco escalas del tipo “Likert” (*The probability that I will shop at this store again is; The likelihood that I would recommend this store to a friend is; If I had to do it again, I would still shop at this store; I say positive things about this supermarket to other people; I consider this supermarket as my primary supermarket*).



*Figura 12: Modelo estructural testeado. Fuente: Demirci, F. Kara, A. (2014).

Limitaciones y futuras investigaciones:

Cabe destacar que la data utilizada para el estudio fue recolectada durante el proceso de compra (shopping) y que la cadena de supermercados utilizada fue una de las con mayor cuota del mercado en Turquía.

A pesar de que hemos utilizado un instrumento de medición válido de SST (tecnologías de auto-servicio), SSTQUAL, algunas dimensiones de la escala original debieron ser podadas debido a su aplicabilidad en el contexto de self-checkout en los supermercados. Por esto, este estudio no entrega confirmación suficiente de la estabilidad de SSTQUAL. Recomendamos el

desarrollo de instrumentos de medición más específicos para los self-checkout systems utilizados en supermercados.

Además, si bien nos enfocamos en la calidad de servicio centrados en las cajas de auto-servicio, cabe señalar que detrás de toda experiencia de un servicio, hay muchos factores y/o actores que podrían no ser reconocidos o nombrados por el cliente, pero que tienen gran impacto en la calidad del servicio entregado.

4.1.2 How to Improve Perceived Service Quality by Increasing Customer Participation

Un marco conceptual es desarrollado para examinar el impacto de la participación del cliente en la entrega del servicio en la percepción de la calidad del servicio entregado. Se propone que, para servicios específicos, cómo aquellos que resultan familiares a los clientes o requieren bajo grado de experticia, una mayor participación por parte del cliente generalmente llevará a una mejor percepción de la calidad del servicio.

Tanto los efectos directos como indirectos de la participación del cliente son examinados. Un modelo casual es propuesto para abarcar estas relaciones y sirve como fundamentación para futuras investigaciones empíricas. La importancia de la calidad de servicio para el cliente y por tanto para los proveedores de servicios es incuestionable.

Los clientes están en una constante búsqueda, y últimamente han estado exigiendo mayor calidad de servicio (Rabin 1983; Berry, Zeithalm, and Parasuraman 1985; Sherden 1988).

Para los proveedores de servicios la calidad de servicio está directamente relacionada con la imagen, ventas y ganancias (Berry, Bennet, and Brown 1989; Buzzel and Gale 1987; Lewis and Booms 1983; Rothman 1983; Rudie & Wansley 1985).

Dado el avance económico de Estados Unidos, no sorprende que el "Marketing Science Institute" les ha dado una muy alta prioridad a estudios relacionados con la "percepción y evaluación de la calidad de servicio" (Marketing Science Institute 1986).

Las investigaciones sobre la calidad de servicio se han centrado en determinantes que se originan en el proveedor mientras que las variables relacionadas con el cliente no reciben mucha atención. Aun así, dado la interacción natural de los servicios, los clientes pueden afectar en la calidad del servicio entregado.

Ha habido investigaciones y discusiones que relacionan la participación del cliente con el tiempo de espera y el tiempo de espera con la calidad del servicio (Maister 1985; Langeard et al. 1981; Bateson 1985b) pero ningún marco conceptual ha sido sugerido para relacionar las 3 variables.

Calidad de servicio percibido

Bonner y Nelson (1985) señalan que la percepción de la calidad de servicio del cliente suele diferir de la evaluación del productor. Jacoby y Olson (1985) apoyan esta diferencia introduciendo dos conceptos separados de calidad:

Calidad real u objetiva y la calidad percibida o subjetiva.

Ellos proponen que la calidad percibida está más relacionada a la experiencia de venta, por lo que requiere mayor atención. Zeithalm (1988) define la calidad percibida como el juicio del cliente sobre la excelencia general o superioridad de un producto (o servicio).

El modelo desarrollado en este paper está basado en la calidad percibida en lugar de la “real”.

La calidad de servicio percibida puede ser operacionalizada usando una escala de múltiples atributos para medir la percepción del cliente de varios atributos de la calidad de servicio (cf. Gronroos 1983; Klaus 1985; Parasuraman, Zeithalm, and Berry 1985).

Lovelock y Young (1979) creen que aumentar la participación del cliente no solo aumenta la productividad, sino que además resulta en un mejor servicio para el cliente. A modo de ejemplo, señalan que los buffet de auto-servicio no solo ahorran costos de HH, sino que además les da la oportunidad de elegir “Qué quieren, la cantidad que quieren y sin retraso”.

En base a esto se propone:

- 1) Para servicios familiares para el cliente o de baja experticia, un incremento en la participación del cliente dará como resultado un incremento en la calidad de servicio percibida.
- 2) Un aumento en el tiempo de espera real causará un aumento en la percepción del tiempo de espera.

- 3) Un decremento en el tiempo de espera percibido resultará en un aumento en la percepción de la calidad de servicio por parte de los clientes.

Un famoso comercial de Federal Express en los principios de 1980 señaló que “esperar es frustrante, desmoralizador, agonizante, agravante, molesto, demandante de tiempo e increíblemente costoso” (Fortune 1980; Maister 1985). Dicho mensaje aplica a todos los servicios de cualquier ámbito, y cuanto antes las empresas sean conscientes del valor de aquello, podrán evidenciar el impacto positivo que un buen servicio puede tener tanto para la empresa como para los/las clientes/as.

4.2 Análisis de sentimiento u opinión en RRSS.

4.2.1 Tesis: Herramientas de análisis de opinión en redes sociales virtuales.

El año 2014, Iván Pliouchtchai, publica su memoria de tesis para optar al título de ingeniero civil en computación en la universidad de Chile. Este trabajo de tesis tiene como objetivo predecir con el mayor grado de certeza posible los resultados de las elecciones presidenciales para el periodo 2014-2018 en base a los comentarios publicados en Twitter.

Para el análisis de sentimiento se estudiaron 2 técnicas: Método Estadístico y Método Ontológico. El método estadístico requiere de datos clasificados para entrenar a la máquina. Mientras que, con el método ontológico, se busca obtener el sentimiento de un texto dividiéndolo en oraciones y en ellas se identifican nombres, verbos y adjetivos para identificar si la oración pertenece o

no a una determinada temática. Luego, se analizan las palabras claves que permiten determinar el sentimiento asociado al sujeto de la oración.

Las principales características de este trabajo son:

1. Utiliza biblioteca *freeling* para tratar el texto libre.
2. Analiza expresiones con ubicación geográfica del tweet.
3. Se analiza el funcionamiento de Twitter (uso) en cuanto a la hora del día.
4. Sistema orientado al análisis de menciones de los candidatos a la presidencia de Chile para el periodo 2014-2018 cuyas elecciones se realizaron en noviembre-diciembre del año 2013.
5. Enfoque Ontológico.

En cuanto a los resultados de los experimentos realizados, se realizó un análisis de los Twitts 7 días antes de la segunda vuelta. Dicho análisis arrojó una popularidad del 61% para Michelle Bachelet (quien obtuvo 62% en las elecciones) y un 39% para Evelyn Matthei (quien obtuvo un 38% en las elecciones).

4.2.2 Techniques and applications for sentiment analysis

El análisis de sentimiento (o minería de opinión) es definido como la tarea de encontrar las opiniones de autores sobre entidades específicas. El proceso de toma de decisiones de las personas es afectado por las opiniones formadas por líderes pensadores y personas ordinarias.

Cuando una persona quiere comprar un producto en línea, él o ella típicamente empezará buscando críticas y opiniones de otras personas escritas en las

varias ofertas disponibles en la web. El análisis de sentimientos es una de las áreas de investigación más activas en las ciencias de la computación.

Es común clasificar las oraciones en dos clases principales de acuerdo a la subjetividad: oraciones objetivas que contienen información de hechos y oraciones subjetivas que contienen opiniones, creencias, y puntos de vista sobre temas o entidades específicos. Aquí, se centrará el análisis en las oraciones subjetivas.

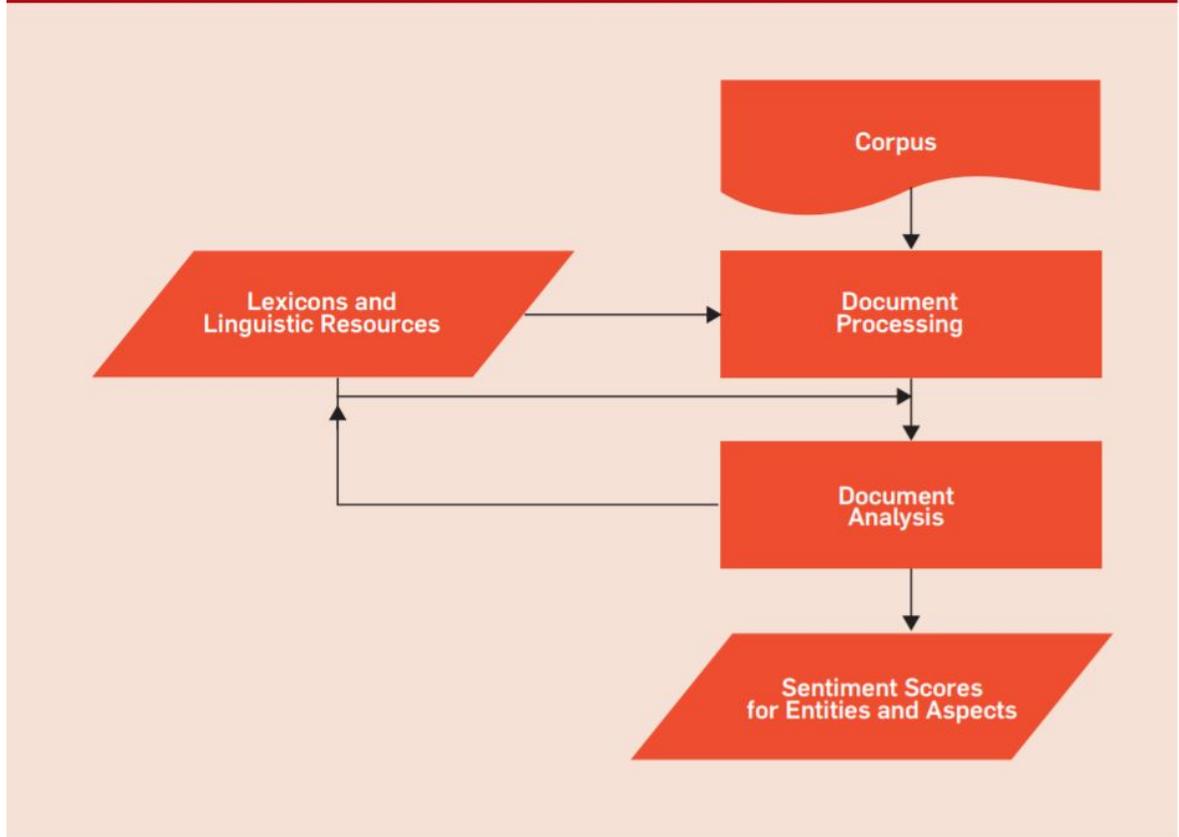
En este documento, se presentan los principales problemas relacionados al análisis de sentimiento y algunas de las técnicas utilizadas para resolverlos, luego se analizan algunas de las principales áreas de aplicación en las que se utiliza el análisis de sentimientos hoy en día.

Este análisis, se concentra en cinco problemas específicos dentro del campo de análisis de sentimiento:

- Análisis de sentimiento a nivel de documento;
- Análisis de sentimiento a nivel de oración;
- Análisis de sentimiento basado en aspecto;
- Análisis de sentimiento comparativo; y,
- Adquisición de lexicón de sentimiento.

La siguiente imagen, muestra una arquitectura general de un sistema genérico de análisis de sentimiento:

Figure 1. Architecture of a generic sentiment analysis system.



**Figura 13: Arquitectura de un sistema genérico de análisis de sentimiento. Fuente: Fieldman, R. (2013).*

Análisis a nivel de documento:

Es la forma más simple de análisis de sentimiento y se asume que el documento contiene una opinión sobre un objeto principal por el autor del documento.

Existen dos acercamientos principales al análisis de sentimiento en documentos: aprendizaje supervisado y aprendizaje no-supervisado. El acercamiento supervisado, asume que existe un set finito de clases en los

cuales el documento debe ser clasificado y hay data de entrenamiento disponible para cada clase. Por otro lado, el aprendizaje no-supervisado a nivel de documento, se basa en determinar la orientación semántica (SO por sus siglas en inglés) de frases específicas dentro del documento. Si la orientación semántica promedio de estas es por sobre un límite predefinido el documento es clasificado como positivo y del caso contrario como negativo.

Asumimos que conocemos la identidad de la entidad a discutir en la oración.

Antes de analizar la polaridad de la oración debemos determinar si las oraciones son objetivas o subjetivas. Sólo las oraciones subjetivas serán analizadas. La mayoría de los métodos utilizan acercamientos supervisados para clasificar las oraciones en las dos clases (positiva y negativa).

Un acercamiento único basado en los “cortes mínimos” fue propuesto por “Pang y Lee” (2004). La premisa principal de su acercamiento es que las oraciones vecinas deberían tener la misma clasificación de subjetividad

Recientes investigaciones muestran que es recomendable tratar distintos tipos de oraciones con distintas estrategias. Las oraciones que requieren una estrategia única incluyen las oraciones condicionales, oraciones de pregunta y oraciones sarcásticas. El sarcasmo es extremadamente difícil de detectar y se presenta mayormente en contextos políticos.

Análisis de sentimiento basado en aspecto

Como sea, en muchos casos la gente habla de entidades que tienen muchos aspectos (atributos) y ellos tienen diferentes opiniones acerca de esos

aspectos. Esto suele ocurrir en reseñas de productos o foros de discusiones dedicados a categorías específicas de productos (como autos, cámaras, smartphones, e incluso drogas farmacéuticas).

El análisis de sentimiento basado en aspecto es el problema de investigación que se centra en el reconocimiento de todas las expresiones de sentimiento dentro de un documento entregado y los aspectos a los que se refieran.

El acercamiento clásico, que es utilizado por muchas compañías comerciales, para la identificación de todos los aspectos de un corpus de las reseñas de productos es extraer todas las “frases sustantivo” y luego dejar todas las “frases sustantivo” cuya frecuencia sea mayor a la de un indicador determinado experimentalmente.

Un acercamiento consiste en reducir el ruido encontrado en las “frases sustantivo”. La idea principal es medir para cada candidato FS el PMI (Pointwise Mutual Information, una medición de la asociación utilizada en teoría de la información y estadísticas) con frases que están altamente relacionadas con el producto de la categoría (como teléfonos, impresoras o cámaras). Sólo esas FS que tienen un PMI por sobre el indicador aprendido son retenidas.

Por ejemplo, para la categoría de impresoras sería “la impresora viene con” o “la impresora tiene”.

Otro acercamiento para la identificación del aspecto es utilizar un separador de frases por dependencia que utiliza expresiones de sentimiento conocidas para encontrar aspectos adicionales (incluso los no frecuentes).

Sólo se ha discutido la identificación de aspectos explícitos, es decir, aspectos que son mencionados explícitamente en la oración. Como sea, hay muchos aspectos que no son mencionados explícitamente y pueden ser inferidos de expresiones de sentimiento que los mencionan implícitamente.

Estos son llamados aspectos implícitos y ejemplos de ellos son el peso, que puede ser inferido del fragmento “este teléfono es muy pesado”, o el tamaño que puede ser inferido del fragmento “la cámara es algo compacta”.

Una forma de extraer aquellos aspectos implícitos es propuesta por Z. Hai et al. (2011) donde una regla de asociación de coocurrencia de dos fases es utilizada para relacionar los aspectos implícitos (expresiones de sentimiento) con aspectos explícitos.

Con estos dos sets se puede utilizar un algoritmo simple que indica la polaridad de cada expresión de sentimiento basado en un léxico de sentimiento, cambiadores de sentimientos (como las palabras de negación), y un manejo especial de conjunciones adversativas como el “pero”.

La polaridad final de cada aspecto está determinada por un promedio con carga de las polaridades de todas las expresiones de sentimiento inversamente cargada por la distancia entre el aspecto y la expresión de sentimiento.

Análisis de sentimientos comparativo

En muchos casos, los usuarios no entregan una opinión directa sobre un producto, pero en su lugar entrega una opinión de comparación, como los siguientes casos de Edmonds.com:

“La 300 C Touring se ve mucho mejor que la Magnum”, “Yo conducí el Honda Civic, no se maneja mejor que el TSX, ni siquiera cerca”. El objetivo del sistema de análisis de sentimientos en este caso es identificar las oraciones que tienen opiniones comparativas, y de extraer las entidades preferidas en cada opinión.

Uno de los papers pioneros en el análisis de sentimiento comparativo es el de Jindal y Liu (2006). Este paper encontró que utilizando un número relativamente pequeño podemos cubrir el 98% de todas las opiniones comparativas.

Estas palabras son:

- Adjetivos adverbios comparativos como: “more”, “less”, y palabras terminadas en “er” como “lighter”.
- Adjetivos y adverbios superlativos como “most”, “least”, y palabras terminadas en “est” como “finest”.
- Frases adicionales como “favor”, “exceed”, “outperform”, “prefer”, “than”, “superior”, “inferior”, “number one”, “against”.

Dado que estas palabras llevan a un alto recall, pero una baja “precisión”, un clasificador naive bayes fue utilizado para filtrar las oraciones que no contienen opinión comparativa.

El clasificador usó patrones secuenciales como características (features). Los patrones secuenciales fueron descubiertos por el “*class sequential rule*” (CSR) algoritmo de minado. Un algoritmo simple para identificar las entidades

preferidas basado en el tipo de comparación usada y la presencia de negación está descrito en Ding et al. (2009).

Adquisición del léxico de sentimientos

Como vimos en la discusión previa, el léxico de sentimiento es el recurso más crucial para la mayoría de los algoritmos de análisis de sentimiento. Aquí hago una breve mención de unos pocos acercamientos para la adquisición del léxico.

Existen 3 opciones para adquirir el léxico:

- Acercamientos manuales, en los que la gente programa el lexico a mano
- Acercamiento basado en diccionarios, en el cual un set de palabras “semilla” es expandido usando recursos como WordNet.
- Y el acercamiento basado en corpus, en el cual un set de palabras “semilla” es expandido utilizando un gran “corpus” de documentos de un mismo dominio.

Conclusión:

Este artículo revisa los principales problemas en el campo de análisis de sentimiento y comenta varios algoritmos que buscan resolver dichos problemas. Muchos de los sistemas de análisis de sentimiento utilizan técnicas simplistas para así evitar desafíos abiertos y por lo tanto su desempeño deja mucho que desear. Entregar soluciones satisfactorias a estos problemas va a hacer del área del análisis de sentimiento algo mucho más amplio.

4.2.3 New avenues in opinion mining and sentiment analysis.

Las opiniones de otros pueden ser cruciales al momento de tomar una decisión o elegir entre múltiples opciones. Cuando esas elecciones involucran recursos de valor (como invertir tiempo y dinero para comprar productos o servicios) la gente suele acudir a las experiencias pasadas de sus pares.

Hasta hace poco, las principales fuentes de información eran los amigos y revistas o páginas web especializadas. Ahora, la 'web social' entrega nuevas herramientas para crear y compartir ideas eficientemente con todos los que están conectados a la World Wide Web.

Minar opiniones y sentimientos en lenguaje natural (NL) es desafiante, ya que requiere un profundo entendimiento de lo implícito y explícito, regular e irregular, y las reglas sintácticas y semánticas del lenguaje. Los investigadores de análisis de sentimientos tienen dificultades con los problemas no resueltos de NLP (natural language processing): resolución de correferencia, manejo de negaciones, resolución de anáforas, reconocimiento de entidades nombradas, y desambiguación del sentido de las palabras.

Además, existen varias herramientas para ayudar a las compañías a extraer y analizar la información de tendencias de gran escala sobre productos; entre estas herramientas se incluye SenticNet (www.sentic.net) , Factiva (www.dowjones.com/factiva) y Converseon (www.converseon.com).

Estos métodos, se basan principalmente en partes de texto en las cuales las personas expresan estados emocionales de manera explícita y por lo tanto esas herramientas no pueden capturar el sentimiento u opinión implícita del crítico.

Common Sentiment Analysis Task

La tarea básica de minería de opinión es la clasificación de polaridad. Ésta ocurre cuando un pedazo de texto expresando una opinión sobre un tema particular es clasificado con uno de dos sentimientos opuestos (ej: like y dislike).

Si el texto no contiene opiniones fuertes o cubre más de un tema o problemática, surgen nuevos desafíos, como la detección de subjetividad e identificación del objetivo de la oración.

Actualmente, la minería de opinión y el análisis de sentimiento se basan en la extracción de vectores para representar las características más salientes e importantes del texto. Podemos usar este vector para clasificar las características más relevantes. Dos características comúnmente utilizadas son 'frecuencia de término' y 'presencia'.

La presencia es un vector de características de valor binario en el cual las entradas indican solo si el termino ocurre (valor 1) o no (valor 0).

Es posible agregar otras características 'término-basadas' al vector de características. 'Posición' hace referencia a como la posición de un *token* en una unidad de texto puede afectar el sentimiento del texto. Luego, podemos considerar la presencia de *n-grams* (típicamente *bigrams* y *trigrams*) como

características útiles. Algunos métodos también se basan en la distancia entre los términos.

Algunos investigadores han desarrollado otras técnicas de mapeo de texto que asignan etiquetas a categorías predefinidas o números reales representando el grado de polaridad. Estos acercamientos están estrictamente limitados por el dominio y tema. Además, la mayoría de las investigaciones de análisis de sentimiento se concentra en texto escrito en inglés y, consecuentemente, la mayoría de las fuentes desarrolladas (como *sentiment lexicon* y *corpora*) están en inglés. Aplicar esta investigación a otros idiomas es un problema de adaptación de dominio.

From heuristics to Discourse Structure

En algunos acercamientos de aprendizaje no supervisado, el léxico de sentimiento o *sentiment lexicon* es generado y luego utilizado para detectar el grado de positividad y subjetividad.

El léxico de sentimientos identifica la polaridad o subjetividad preferente de un término o frase. Los primeros trabajos trabajaban en su mayor parte con heurísticas lingüísticas.

Luego, trabajos posteriores se concentraron en propagar el valor de las palabras semilla (*seed art*, de los cuales la polaridad es conocida) hacia términos que co-ocurren con ellas en otras relaciones definidas en WordNet (una gran base de datos de léxico en inglés).

From Coarse- to Fine-Grained Analysis

Al principio, Bo Pang y sus colegas clasificaban documentos enteros por una polaridad general positiva o negativa. Estos documentos fueron mayoritariamente supervisados, manualmente etiquetados, como críticas de películas o productos indicando explícitamente una opinión general positiva o negativa. Las opiniones no ocurren únicamente a nivel de documento ni están limitados a un único objetivo o valencia. Un documento puede contener opiniones positivas y negativas sobre uno o más tópicos. Por esto, trabajos posteriores adoptaron un análisis de opinión a nivel de segmento que utiliza técnicas basadas en grafos para distinguir secciones sentimentales de las no sentimentales.

Incluso acercamientos a nivel de oración fallan al momento de descubrir sentimientos acerca de una entidad o sus aspectos. Para corregir eso, otros investigadores adoptaron un acercamiento a nivel de aspecto, en la cual una opinión consiste en objetivos y los sentimientos relacionados a ellos.

From Keywords to Concepts

Podemos agrupar los acercamientos existentes en cuatro grupos: *keyword spotting*, *lexical affinity*, *statistical methods*, y *concept-based techniques*.

Keyword Spotting

A pesar de ser el acercamiento más *ingenuo*, la accesibilidad y economía de la identificación de palabras clave lo hace popular. Este método clasifica el texto

en la categoría afectada basado en la presencia de palabras no-ambiguas como *happy, sad, afraid and bored* (feliz, triste, asustado/a y aburrido/a).

La identificación de palabras claves es débil en dos áreas: no puede reconocer con certeza palabras de afecto negadas y se centra en características superficiales.

Lexical Affinity

Este acercamiento no solo detecta las palabras de afecto obvias, este además asigna a palabras arbitrarias una probable “afinidad” a emociones particulares. Por ejemplo, la afinidad de léxico podría asignarle a la palabra ‘accident’ un 75% de probabilidad de afectar negativamente al objetivo. A pesar de que comúnmente realiza una excepcional identificación de palabras clave, hay dos problemas principales con este acercamiento.

Primero, las oraciones negadas (I avoided an accident) engañan a la afinidad lexical, porque operan sólo a nivel de palabra. Segundo, las probabilidades de la afinidad lexical suelen estar orientada a un texto de un género en particular, dictado por la fuente lingüista de la *corpora*. Esto hace difícil el desarrollo de un modelo independiente del dominio reutilizable.

Statistical Methods

Este acercamiento, que incluye la inferencia bayesiana y las máquinas de soporte vectorial es popular en la clasificación de texto sentimentado. Los

investigadores utilizan métodos estadísticos en proyectos tales como el clasificador de películas de Pang y muchos otros.

Alimentando un algoritmo de *machine-learning* con un gran cuerpo de entrenamiento de texto afectivo, el sistema podría no solo aprender del valor de afectividad de las palabras claves (como en el acercamiento de identificación de palabras claves), sino que además tomar en cuenta la valencia de otras palabras claves arbitrarias (similar al *lexical affinity*), puntuación, y frecuencias de palabras co-ocurrentes.

Generalmente, los métodos estadísticos son semánticamente débiles, lo que significa que individualmente—con la excepción de palabras de afecto obvias—elementos de léxico o coocurrencia de los modelos estadísticos tienen poco valor predictivo. Como resultado, los clasificadores estadísticos de texto sólo funcionan bien cuando reciben texto de entrada suficientemente vasto.

Así que, mientras estos métodos pueden ser capaces de clasificar en cuanto a afecto el texto de un usuario en el nivel de página o párrafo, no suelen funcionar bien en unidades más pequeñas de texto como las oraciones o cláusulas.

Concept-based Approaches

Estos métodos utilizan ontología Web o redes semánticas para alcanzar el análisis semántico del texto. Esto ayuda al sistema a agarrar la información afectiva y conceptual asociada con las opiniones en lenguaje natural. Confiando en bases de conocimiento semántico vastas, tales acercamientos se alejan de utilizar contadores de palabras clave y co-ocurrencia de palabras, y en lugar de

eso se basa en las características y significados implícitos asociados a los conceptos del lenguaje natural.

Superior a las técnicas puramente sintácticas, los acercamientos basados en conceptos pueden detectar sentimientos sutilmente expresados. Acercamientos basados en conceptos dependen fuertemente en la profundidad y anchura de la base de conocimientos que utiliza. Sin una fuente comprensiva que engloba el conocimiento humano, un sistema de minería de opinión tendrá dificultad al momento de captar la semántica del texto en lenguaje natural. Además, lo típico de las bases de conocimiento—eso es, el hecho de que contienen solo información típica asociada a los conceptos—limita su capacidad de manejar matices semánticos.

Su representación arreglada/plana, finalmente, sitúa barreras en inferencias de características afectivas y semánticas asociadas con los conceptos.

Conclusiones:

Gradualmente, la investigación del análisis de sentimientos se está distinguiendo como un campo separado, encontrándose entre el NLP (procesamiento del lenguaje natural) y NLU (entendimiento del lenguaje natural). Se proyecta la investigación del análisis de sentimiento moviéndose hacia un análisis de texto en lenguaje natural basado en contenido, concepto y contexto, acompañado de técnicas de análisis eficientes en cuanto a tiempo y aplicables a grandes cantidades de data de fuentes sociales (E. Cambria et al., 2013).

En lugar de reunir opiniones aisladas sobre un tema u objeto, los usuarios generalmente prefieren comparar *features* específicas de diferentes productos (por ejemplo, el Iphone 5 versus el Galaxy S3). Para poder realizar estas comparaciones, los investigadores deben construir bases comprensivas de conocimiento general para detectar *features* y bases de sentido común para detectar polaridad. Dichas bases de sentido común serán clave para deconstruir apropiadamente el texto en lenguaje natural en sentimientos—por ejemplo, evaluando el concepto “habitación pequeña” como negativo para una crítica de hotel y “cola pequeña” como positivo para una oficina de correos, o “ve a leer el libro” como positivo para una crítica de libro, pero negativo para una crítica de película.

La web ha cambiado de “sólo leer” a “leer y escribir”. Esta evolución ha creado usuarios entusiastas que interactúan y comparten a través de redes sociales, comunidades online, blogs, wikis, y otros medios colaborativos. A pesar del significativo progreso, la minería de opinión y el análisis de sentimientos aún están buscando su propia voz como un campo interdisciplinario. Ingenieros y científicos del área de la computación utilizan técnicas de *machine learning* para la clasificación automática de sentimientos de video, voz, texto y fisiología.

La minería de opinión y el análisis de sentimiento están intrínsecamente ligados a las ciencias afectivas que buscan entender las emociones humanas. Un conocimiento más completo debe ser combinado con métodos de razonamiento que estén más profundamente inspirados en el pensamiento humano y la psicología. Esto llevará a un mejor entendimiento de las opiniones en lenguaje natural y a una más eficiente disminución de la separación entre la información

multimodal (no estructurada) y la información procesable por máquina (estructurada).

Combinar teorías científicas de emociones con los prácticos objetivos ingenieriles de analizar sentimientos en textos en lenguaje natural llevará a acercamientos más bio-inspirados para el diseño de sistemas inteligentes de minería de opinión capaces de manejar conocimiento semántico, realizar analogías, aprender nuevo conocimiento afectivo, y detectar, percibir y “sentir” emociones.

4.2.4 State of the art for sentiment analysis.

Hay un tremendo interés en el análisis de sentimientos en distintas disciplinas, tales como Comercio (Jansen et al., 2009), Salud (Chew and Eysenbach, 2010), y manejo de desastres (Verma et al, 2011; Mandel et al., 2012).

El texto describe cómo se creó el estado del arte de dos clasificadores SVM, uno para detectar el sentimiento de mensajes como Tweets y SMS (tarea a nivel de mensaje) y otra para detectar el sentimiento de un término dentro de un mensaje (tarea a nivel de término).

Dichos clasificadores, creados en el contexto de una competencia internacional de ejercicios de evaluación semántica (SemEval-2013) (Wilson et al., 2013).

Algunas de las condiciones establecidas para la competencia son las siguientes: el sentimiento puede ser una de tres posibilidades: positivo, negativo

o neutral. Los organizadores crearon y compartieron Tweets con etiqueta de sentimiento para entrenamiento, desarrollo y testeo.

La postulación del paper se mantuvo primera en ambas tareas, obteniendo un macro-averaged F-Score de 69.02 en la tarea a nivel de mensaje y 88.93 en la tarea a nivel de término. Los organizadores entregaron además un segundo set de testeo, compuesto de mensajes de SMS (Short Message Service). No se hizo entrega de mensajes SMS para entrenamiento.

Lexicons o Sentiment Lexicons son listas de palabras con asociaciones a sentimientos positivos y negativos. Es importante tener claro estos conceptos ya que serán mencionados reiteradamente en el desarrollo del documento.

Los léxicons manualmente creados que se utilizaron incluyen el NRC Emotion Lexicon (Mohammad and Turney, 2010; Mohammad and Yang, 2011) cerca de 14.000 palabras, el MPQA Lexicon (Wilson et al., 2005) cerca de 8.000 palabras, y el Bing Liu Lexicon (Hu y Liu, 2004) cerca de 6.800 palabras.

Nuevo automáticamente generado lexicon de sentimiento específicamente para Tweets.

NRC hashtag Sentiment Lexicon

Algunas palabras en los tweets están especialmente marcadas con un *hashtag* (o etiqueta) (#) para indicar el tema o sentimiento. Mohammad (2012) mostró que las palabras de emociones '*hashtageadas*' tales como goce, tristeza, enojo y sorpresa son buenos indicadores de que el Tweet como un todo, incluso sin la palabra de emoción '*hashtageada*', está expresando el mismo sentimiento o

emoción. Adaptaron esa idea para crear un gran corpus de Tweets positivos y negativos.

Consultaron la API de Twitter cada cuatro horas de Abril a Diciembre de 2012 en búsqueda de Tweets que tengan ya sea un hashtag de una palabra positiva o negativa.

Una colección de 78 palabras semilla relacionadas muy de cerca a *positivo* o *negativo* tales como #good, #excellent, #bad y #terrible se utilizaron (32 positivos y 36 negativos)- Estos términos fueron escogidos de entradas para positivo y negativo en el *Roget's Thesaurus* (un diccionario de sinónimos).

Un set de 775.000 Tweets fue utilizado para generar un vasto lexicón de asociación palabra-sentimiento. Un Tweet fue considerado positivo si este tenía una de las 32 palabras semilla positivas hashtageadas, y negativa si este contenía una de las 36 palabras semilla negativas hashtageadas.

El *association score* para un término w fue calculado de estos Tweets pseudo-etiquetados como se muestra a continuación:

$$score(w) = PMI(w, positive) - PMI(w, negative)$$

*Figura 14: Fórmula del Score. Fuente: Saif M. Mohammad, Kiritchenko, S., Zhu, X. (2013).

Donde PMI significa *pointwise mutual information* (punto de información mutua). Un puntaje positivo indica asociación con sentimientos positivos, mientras que un puntaje negativo indica una asociación con un sentimiento negativo. La magnitud es un indicativo del grado de la asociación. El léxico final al que nos

referiremos como NRC *Hashtag Sentiment Lexicon* tiene entradas para 54.129 *unigrams* y 316.531 *bigrams*. Las entradas fueron además generadas para pares de *unigram-unigram*, *unigram-bigram*, y *bigram-bigram* que fueran no necesariamente contiguos en el corpus del Tweet. Se eliminaron pares con ciertas puntuaciones, símbolos '@' y algunas *function-words* fueron removidas. El Lexicón tiene entradas para 308.808 pares no-contiguos.

Tarea: Detectando el sentimiento de un mensaje de manera automática

Clasificador y features

Entrenaron una máquina de soporte vectorial (SVM) (fan et al., 2008) con la data de entrenamiento entregada. SVM es un algoritmo del estado del arte demostrado ser efectivo en las tareas de clasificación de texto y robusto en grandes espacios de *features*. El kernel lineal y el valor para el parámetro $C=0.005$ fueron escogidos vía cross-validation de la data de entrenamiento.

Tokenizaron y etiquetaron parte-del-discurso de los Tweets con la herramienta Carnegie Mellon University (CMU) (Gimpel et al., 2011). Cada Tweet fue representado como un vector de *feature* creado del siguiente grupo de *features*:

- *Ngrams* de palabras: presencia o ausencia de secuencias contiguas de 1, 2, 3 y 4 tokens; *ngrams* no contiguos (*ngrams* con un token reemplazado por *);
- *ngrams* de carácter: presencia o ausencia de secuencias contiguas de 3, 4 y 5 caracteres;
- *All-caps*: el número de palabras con todos los caracteres en mayúscula;
- POS (*arto f speech*): el número de ocurrencias de la etiqueta de cada parte-del-discurso;
- Hashtags: número de hashtags;

- Lexicons: los siguientes sets de *features* fueron generados para cada uno de los 3 *Sentiment Lexicon* manualmente construidos (NRC emotion lexicon , MPQA, Bing liu lexicon). Sets de *features* separados fueron producidos para *unigrams* , *bigrams*, y pares no contiguos. Las *Lexicon Features* fueron creadas para todos los tokens en el Tweet, para cada etiqueta de parte-del-discurso, para hashtags, y para los tokens *all-caps*. Para cada token w y emoción o polaridad p , utilizamos el puntaje de sentimiento/emoción $\text{score}(w,p)$ para determinar:
 - Conteo total de tokens en el Tweet con $\text{score}(w,p) > 0$;
 - Puntaje total = $\sum_{w \in \text{Tweet}} \text{score}(w,p)$;
 - El puntaje máximo = $\max_{w \in \text{Tweet}} \text{score}(w,p)$;
 - El puntaje del último token en el Tweet con $\text{score}(w,p) > 0$;
- Puntuación:
 - El número de secuencias contiguas de signos de exclamación, signos de pregunta, y ambos de exclamación y pregunta;
 - Si es que el último token contiene un símbolo de exclamación o pregunta;
- Emoticons: La polaridad de un emoticon fue determinada con una expresión regular adoptada del Script de tokenización de Christopher Potts.
 - Presencia o ausencia de emoticones positivos o negativos en cualquier posición del Tweet;
 - Si es que el último token es un emoticón positivo o negativo;
- Palabras elongadas: número de palabras con un carácter repetido más de 2 veces, por ejemplo “sooo”.
- Clusters: la herramienta de *pos-tagging* de CMU entrega los clusters de token producidos con el algoritmo de clustering ‘*Brown*’ en 56 millones de Tweets en inglés. Estos 1.000 clusters sirven como una representación alternativa del contenido de un Tweet, reduciendo lo esparcido del espacio de tokens;
 - La presencia o ausencia de tokens de cada uno de los 1000 clusters.
- Negación: el número de contextos negados. Siguiendo (Pang et al., 2002), definimos un contexto negado como un segmento de un Tweet que comienza con una palabra de negación (ejemplo: no, shouldn’t) y termina con uno de los simbolos de puntuación: “,” , “.,” , “:,” , “;,” , “!,” , “?” . Un contexto negado afecta el *ngram* y las *features* del Lexicon: agregamos

el sufijo “_NEG” a cada palabra que sigue a la palabra de negación (“perfect” se vuelve “perfect_NEG”). El sufijo “_NEG” es también agregado a *features* de polaridad y emociones (“POLARITY_positive” se vuelve “POLARITY_positive_NEG”). La lista de palabras de negación fue adoptada del tutorial de sentimientos de Christopher Potts.

EXPERIMENTOS

Entrenaron el clasificador SVM en el set de 9.912 Tweets anotados (8.258 en el set de entrenamiento y 1.654 en el set de desarrollo). Aplicaron el modelo al set de testeo de 3.813 Tweets sin mirar. El mismo modelo fue aplicado sin cambios al otro set de prueba de 2.094 mensajes de SMS. El puntaje de resultado final utilizado por los organizadores fue el *macro-averaged* F-Score de las clases positivas y negativas. Los resultados obtenidos por su sistema en el set de entrenamiento (cross-validation de 10 partes), el set de desarrollo (cuando entrenado en el set de entrenamiento) y set de testeo (cuando es entrenado con ambos set el de entrenamiento y el de desarrollo) son mostrados en la tabla 2.

Table 2: Message-level Task: The macro-averaged F-scores on different datasets.

	Classifier	Tweets	SMS
Training set:	Majority	26.94	-
	SVM-all	67.20	-
Development set:	Majority	26.85	-
	SVM-all	68.72	-
Test set:	Majority	29.19	19.03
	SVM-unigrams	39.61	39.29
	SVM-all	69.02	68.46

Table 3: Message-level Task: The macro-averaged F-scores obtained on the test sets with one of the feature groups removed. The number in the brackets is the difference with the *all features* score. The biggest drops are shown in bold.

Experiment	Tweets	SMS
all features	69.02	68.46
all - lexicons	60.42 (-8.60)	59.73 (-8.73)
all - manual lex.	67.45 (-1.57)	65.64 (-2.82)
all - auto. lex.	63.78 (-5.24)	67.12 (-1.34)
all - Senti140 lex.	65.25 (-3.77)	67.33 (-1.13)
all - Hashtag lex.	65.22 (-3.80)	70.28 (1.82)
all - ngrams	61.77 (-7.25)	67.27 (-1.19)
all - word ngrams	64.64 (-4.38)	66.56 (-1.9)
all - char. ngrams	67.10 (-1.92)	68.94 (0.48)
all - negation	67.20 (-1.82)	66.22 (-2.24)
all - POS	68.38 (-0.64)	67.07 (-1.39)
all - clusters	69.01 (-0.01)	68.10 (-0.36)
all - encodings (elongated, emoticons, punctuations, all-caps, hashtags)	69.16 (0.14)	68.28 (-0.18)

*Tabla 2 y 3: Resultados experimentos. Fuente: Saif M. Mohammad, Kiritchenko, S., Zhu, X. (2013).

La tabla 3 muestra los resultados de los experimentos de extracción en los que se repite el mismo proceso de clasificación, pero se remueve un grupo *feature* a la vez. Las *features* más influyentes para ambas bases de datos resultaron ser las *features* de léxico de sentimiento, las cuales entregaron ganancias de más del 8.5%. Es interesante notar que los Tweets se beneficiaron en mayor parte del lexicón de sentimiento automático (lexicón de NRC hashtag y de Sentiment140), mientras que el set de SMS se benefició más del lexicón manual (MPQA, NRC emotion y el de Bing Liu). Entre los lexicones automáticos, tanto el de Hashtag Sentiment como el de Sentiment140 contribuyeron aproximadamente lo mismo en la mejora de desempeño en el set de Tweets. Como era de esperar, el impacto de los ngrams en los dataset de SMS fue menos importante dado que el modelo clasificador fue entrenado con

Tweets solamente. Es importante señalar que prestar atención a las negaciones mejoró el desempeño en ambos datasets.

Tarea: Detectando el sentimiento de un término en un mensaje.

El objetivo de esta tarea es detectar si un término (palabra o frase) dentro de un mensaje transmite un sentimiento positivo, negativo o neutral. Cabe destacar que el mismo término puede expresar distintos sentimientos en contextos diferentes.

Clasificador y features

Se entrenó el SVM utilizando los paquetes de LibSVM (Chang and Lin, 2011) y kernel lineal. En cross-validation de 10 partes sobre el set de entrenamiento, el kernel lineal tuvo un desempeño mejor que el de otros kernels implementados en LibSVM así como el clasificador de entropía máxima. El modelo aprovecha una variedad de *features*, que serán descritos a continuación:

- Ngrams de palabras:
 - Presencia o ausencia de unigrams, bigrams, y el string de la palabra completa del término objetivo;
 - Unigrams y bigrams de inicio y término;
- Ngrams de caracteres: presencia o ausencia de dos y tres prefijos y sufijos de todas las palabras en un término objetivo (el término objetivo puede ser una secuencia multi-palabra);
- Palabras alargadas: presencia o ausencia de palabras alargadas (por ejemplo: “Soooo”);
- Emoticones: el número y categorías de los emoticones que el término contiene;
- Puntuación: presencia o ausencia de secuencias de puntuación tales como “?!” y “!!!”;
- Mayúsculas:

- Ya sea que todas las palabras en el objetivo empiezan con una letra en mayúscula seguida de una en minúscula;
- Ya sea que las palabras del objetivo están todas en mayúscula (potencial nombre de una entidad);
- Palabras stop (stopwords): Si es que un término se compone sólo de stop-words. Si es así los *features* separados indican si es que hay 1, 2, 3 o más stop-words;
- Largo:
 - Largo del término objetivo (número de palabras);
 - El largo promedio de palabras (número de caracteres) de un término;
 - Un *feature* binario indicando si es que un término contiene palabras largas;
- Negación: Similar a los descritos en la tarea a nivel de mensaje. Cuando una palabra de negación sea encontrada inmediatamente antes del objetivo o dentro del objetivo, las polaridades de todos los tokens después de la negación fueron cambiados;
- Posición: Ya sea un término esté al inicio, final o alguna otra posición.
- Lexicons de sentimiento: utilizamos lexicones automáticamente creados (NRC Hashtag, Sentiment140) así como lexicones creados manualmente (NRC Emotion ,MPQA, Bing Liu).
 - Conteo total de tokens en el término objetivo con puntaje de sentimiento más grande que 0;
 - La suma de puntajes de sentimiento para todos los tokens en el objetivo;
 - El puntaje de sentimiento máximo;
 - El puntaje de sentimiento distinto de cero del último token en el objetivo;
- Separación de término: cuando un término contiene un hashtag hecho de múltiples palabras (por ejemplo #biggestdaythisyear), se divide el hashtag en sus palabras componentes;
- Otros:
 - Si un término contiene un usuario de Twitter;
 - Si un término contiene una URL.

Las *features* de más arriba fueron extraídas del término objetivo así como también del resto del mensaje (el contexto). Para unigrams y bigrams, se utilizaron 4 palabras en cada lado del objetivo como contexto.

EXPERIMENTOS

Se entrenó el clasificador SVM sobre los 8.891 términos anotados en tweets (7.756 en el set de entrenamiento y 1.135 términos en el set de desarrollo). Se aplicó el modelo a los 4.435 términos en el set de tweets de prueba (test). El mismo modelo fue aplicado sin cambios al otro set de prueba de 2.334 términos en mensajes SMS sin revisar. El puntaje de línea de fondo (base) utilizado por los organizadores de tareas fue el macro-promediado F-Score de las clases positivas y negativas.

Los resultados en los sets de entrenamiento (cross-validation de 10 partes), y los sets de prueba (entrenados en los sets combinados de tweets de entrenamiento y desarrollo) son mostrados en la tabla 4. La tabla muestra además resultados de línea base obtenidos por un clasificador mayoritario que siempre predice la clase más frecuente como salida, y un resultado adicional de línea base obtenidos utilizando un SVM y *features* unigram por si solas. Su postulación obtuvo un macro-aproximado F-Score de 88.93 en el set de tweets y recibió el primer lugar entre 29 postulaciones de 23 equipos participantes. Incluso sin adaptación específica a data de SMS, su postulación de SMS obtuvo un segundo lugar con un F-score de 88.00. Cabe destacar que el desempeño es significativamente mejor en las tareas a nivel de término versus las tareas a nivel de mensaje. Esto se debe en mayor parte debido a las *features* de ngrams. Se analizó la data etiquetada proporcionada para determinar por qué los ngrams tienen tanta relevancia en esta tarea. Se encontró que el número de tokens de prueba ya vistos dentro de los objetivos de la data de entrenamiento fue del 85.1%. Además, la proporción promedio de las instancias

pertencientes a la polaridad más dominante de un término objetivo versus el número total de instancias de ese término objetivo fue de 0.808.

Table 4: Term-level Task: The macro-averaged F-scores on the datasets. The official scores of our submission are shown in bold. SVM-all* shows results after a bug fix.

	Classifier	Tweets	SMS
Training set:	Majority	38.38	-
	SVM-all	86.80	-
Development set:	Majority	36.34	-
	SVM-all	86.49	-
Test set:	Majority	38.13	32.11
	SVM-unigrams	80.28	78.71
	official SVM-all	88.93	88.00
	SVM-all*	89.10	88.34

Table 5: Term-level Task: The F-scores obtained on the test sets with one of the feature groups removed. The number in brackets is the difference with the *all features* score. The biggest drops are shown in bold.

Experiment	Tweets	SMS
all features	89.10	88.34
all - ngrams	83.86 (-5.24)	80.49 (-7.85)
all - word ngrams	88.38 (-0.72)	87.37 (-0.97)
all - char. ngrams	89.01 (-0.09)	87.31 (-1.03)
all - lexicons	85.15 (-3.95)	83.70 (-4.64)
all - manual lex.	87.69 (-1.41)	86.84 (-1.5)
all - auto lex.	88.24 (-0.86)	86.65 (-1.69)
all - negation	88.38 (-0.72)	86.77 (-1.57)
all - stopwords	89.17 (0.07)	88.30 (-0.04)
all - encodings (elongated words, emoticons, punctns., uppercase)	89.16 (0.06)	88.39 (0.05)
all - target	72.97 (-16.13)	68.96 (-19.38)
all - context	85.02 (-4.08)	85.93 (-2.41)

*Tabla 4 y 5: Resultados experimentos. Fuente: Saif M. Mohammad, Kiritchenko, S., Zhu, X. (2013).

La tabla 5 muestra los F-scores de la extracción de *features*. Note que las *features* de ngrams fueron las más útiles. El lexicón de sentimiento es el siguiente grupo más útil—quitarlos llevaría a una caída del F-score de 3.95 puntos para el set de Tweets y 4.64 para el set de SMS. Modelar las negaciones mejora el F-score en 0.72 puntos en el set de Tweets y 1.57 puntos en el set de SMS.

Las últimas dos filas de la Tabla 5 muestra los resultados obtenidos cuando las *features* son extraídas sólo del objetivo (y no su contexto) y de cuando son

extraídas sólo de su contexto (y no del objetivo mismo). Resulta interesante notar que a pesar de que el contexto puede afectar la polaridad del objetivo, utilizar las *features* del objetivo por sí sólo resulta substancialmente más útil que utilizar las *features* de contexto por sí solas. Sin embargo, agregar *features* de contexto mejora el F-score en 2-4 puntos.

Conclusiones:

Se crearon dos clasificadores SVM del “estado del arte”, uno para detectar el sentimiento de mensajes y uno para detectar el sentimiento de un término dentro de un mensaje. Su postulación en la data de Tweet se mantuvo primero en ambas subtareas de la competición “Detecting Sentiment in Twitter” de SemEval-2013. Se implementaron una variedad de *features* basados en la forma superficial y categorías lexicales. Las *features* de lexicón de sentimiento (tantos las generadas manualmente como automáticamente) en conjunto con las *features* de ngrams (ambos ngrams de palabras y de carácter) llevaron a la mayor ganancia en desempeño.

4.3 Lo que ofrece el mercado

Aquí se mencionan servicios disponibles en el mercado y se describen brevemente. Cabe destacar que, en los servicios relacionados al análisis de sentimiento, no es posible analizar en un aspecto técnico el funcionamiento de los sistemas dado que en esto reside su propuesta de valor.

4.3.1 ProCalidad

ProCalidad es una organización sin fines de lucro, fundada el año 2001. Es una iniciativa conjunta con la Universidad Adolfo Ibáñez y Praxis Customer Engagement. Esta iniciativa tiene como objetivo el fortalecimiento en las empresas chilenas de un decisivo enfoque hacia el cliente, que aumente gradualmente su nivel de competitividad.

¿Qué hace ProCalidad?

1. Conocen y hacen pública la opinión de los chilenos en relación con el servicio que reciben por lo que pagan.
2. Crean un movimiento de los ejecutivos hacia un enfoque de excelencia en la gestión de clientes.
3. Fomentan la competitividad de los diferentes sectores económicos del país.

Variables Medidas:

- Satisfacción general.
- Satisfacción dado precio.
- Recomendación.
- Recompra.
- Existencia de problemas.

4.3.2 Servqual

El Modelo SERVQUAL de Calidad de Servicio fue elaborado por Zeithaml, Parasuraman y Berry cuyo propósito es mejorar la calidad de servicio ofrecida por una organización. Utiliza un cuestionario tipo que evalúa la calidad de un servicio a lo largo de cinco dimensiones:

Criterios utilizados:

- 1) Elementos tangibles: Apariencia de las instalaciones equipos, empleados y materiales de comunicación.
- 2) Fiabilidad: Habilidad de prestar el servicio prometido de forma fiable y cuidadosa tal como se ha prometido con error cero.
- 3) Capacidad de respuesta o velocidad de respuesta: Disposición y voluntad para ayudar a los usuarios y proporcionar un servicio rápido.
- 4) Seguridad o garantía: Conocimiento y atención mostrados por los empleados y sus habilidades para inspirar credibilidad y confianza.
- 5) Empatía: Atención individualizada que ofrecen las empresas a sus clientes.

El cuestionario Servqual se resume en 3 secciones:

En la primera, se interroga al cliente sobre expectativas como tal. Lo que éste opina sobre lo que un servicio determinado debe ser. Esto se hace mediante 22 declaraciones. Sobre ellas, el usuario debe situar, en una escala de 1 a 7, el grado de expectativa para cada una de dichas declaraciones.

En la segunda, se recoge la percepción del cliente respecto al servicio que presta la empresa. Es decir, hasta qué punto considera que la empresa posee las características descritas en cada declaración.

Finalmente, otra sección, situada entre las dos anteriores, cuantifica la evaluación de los clientes respecto a la importancia relativa de los cinco criterios, lo que permitirá ponderar las puntuaciones obtenidas.

¿Qué se espera obtener a través de Servqual?

- Una calificación de la calidad del servicio: Índice de Calidad del Servicio (ISC).
- Lo que desean los consumidores
- Lo que encuentran los consumidores
- Las lagunas de insatisfacción
- Ordena los defectos de calidad

4.3.4 Brand24

Nació a partir de un startup y ofrece un servicio de monitoreo inteligente en internet para marcas, productos o negocios de todos los tamaños. Hoy en día más de 100.000 compañías lo utilizan para monitorear lo que denominan como “huella o rastro digital” (*digital footprint*) a nivel mundial y ha sido premiado en categorías como: mejor startup (premios Aulery, 2012), mejor servicio (“relations with costumers” Ekomersy, 2014), mejor app web (“the next web startup awards”, 2013) y mejor aplicación móvil (“best mobile app for companies” mobile trends, 2014).

Entre sus principales características destacan:

- Análisis de sentimiento
- Índice de autoridad para cada participante de la discusión
- Notificaciones dentro de la app y vía email de cambios significativos en cantidad y contenido de los comentarios de una discusión
- Un sistema organizado de *streaming* de las menciones de tu compañía para una interacción más rápida
- Gráficos de volumen de discusiones

4.3.5 Hootsuite

Es una herramienta de gestión de redes sociales, cuyo fin principal es potenciar al máximo la efectividad de las distintas plataformas, además de facilitar la programación y análisis de estas. Sus clientes son pequeñas, medianas y grandes empresas o personas naturales que quieran lograr un mayor alcance de público. Además, tienen una sección de recursos donde ofrecen material de estudio gratuito con cientos de archivos y guías para ayudar a tener éxito en redes sociales. Y para quienes buscan profesionalizarse en redes sociales, ofrecen cursos (gratuitos y pagados) y certificaciones de distintos niveles.

En resumen, Hootsuite ofrece:

- 1) Programación: Mantén activa tu presencia en redes sociales 24/7 al programar de manera automática cientos de publicaciones de redes sociales en todas tus cuentas.
- 2) Administración de contenido: Mantén un mensaje consistente en tus perfiles sociales. Publica tu contenido desde la nube (iCloud). Administra tu contenido con las etiquetas, el motor de búsqueda y el uso de estadísticas dentro de la plataforma.
- 3) Análisis social: Evalúa el impacto de tus campañas en redes sociales con informes personalizados.

4.3.6 Natural Language de Cloud

Es un potente producto de análisis de texto del área de inteligencia artificial y *machine learning* de Google. Consiste en una API REST que descubre la

estructura y el significado del texto a través de modelos de aprendizaje automático ya preparados y de modelos personalizados que se crean fácilmente con AutoML Natural Language.

Entre las características principales de este producto destacan:

- Extracción y análisis de información valiosa de clientes, permitiendo conocer información útil sobre la acogida de cierto producto o la experiencia de los usuarios mediante la detección de entidades y análisis de opiniones mantenidas con los clientes por correo electrónico, chat o redes sociales.
- Asistencia multimedia en varios idiomas, combinando esta herramienta con la API *Google Cloud Speech*, es posible extraer información valiosa de conversaciones de audio o combinando con la API *Vision* para entender documentos escaneados. Para extraer y comprender las opiniones escritas en distintos idiomas, es posible traducir primero utilizando la API *Cloud Translation*.
- Clasificación de contenido y gráfico de relaciones, a través del uso de indicadores del sistema de análisis sintáctico para crear gráficos de relaciones de las entidades que se extraigan a partir de noticias o artículos de Wikipedia.
- Modelos de aprendizaje profundo de Google, utilizando en cada búsqueda la tecnología de aprendizaje automático de los módulos de “Búsqueda” y “Asistente de Google”. La “Búsqueda” utiliza esta tecnología para encontrar respuestas concretas a las preguntas de los usuarios, mientras que el “Asistente” la usa para entender el lenguaje natural.

4.3.7 IBM Cloud – Natural Language Understanding

Esta API de IBM permite a los desarrolladores analizar características semánticas de un texto de entrada, incluyendo categorías, conceptos, emociones, entidades, palabras clave (*keywords*), metadata, relaciones, roles semánticos y sentimientos. Su modo de uso consiste en enviar peticiones a la API con un texto, HTML, o una URL pública, y especificar una o más de las características mencionadas a analizar.

Si bien el idioma inglés es el único que soporta todas las características, hay una lista de idiomas con soporte limitados tales como: árabe, chino, francés, español, alemán, entre otros. Hay 6 principales aristas en las que Watson puede mejorar la manera en la que se desempeña una organización:

- Búsqueda inteligente sistemática y automatizada, que acelera el proceso de búsqueda de información para dar uso a ese tiempo en actuar y tomar medidas.
- Enriquecimiento de las interacciones reduciendo tiempos de respuesta, aumentando el número de transacciones, y hacer de cada interacción algo significativo y productivo utilizando Autodesk.
- Anticipar y adelantarse a interrupciones utilizando IA para monitorear constantemente la condición de los sistemas de los que depende su negocio para asegurarse de que los problemas no interrumpen su trabajo.
- Recomendar con confianza utilizando “data inteligente” para construir una relación de confianza con el cliente y su negocio.

- Experticia en escala, combinando la experticia de cada empleado con el último aprendizaje de la industria, “cada uno de ustedes sabrá tanto como todos ustedes”.
- Detectar responsabilidades y migrar riesgos entrenando sistemas para entender y mantenerse al día con regulaciones en constante cambio y obligaciones de privacidad.

4.3.8 Amazon Comprehend

Es un servicio de procesamiento de lenguaje natural (NLP) que usa el aprendizaje automático para encontrar información y relaciones en textos. No se requiere experiencia en aprendizaje automático.

Amazon Comprehend utiliza el aprendizaje automático para ayudarle a descubrir la información y las relaciones en sus datos no estructurados. El servicio identifica el idioma del texto; extrae frases, nombres de lugares, personas, marcas o eventos clave; comprende el grado de positividad o negatividad del texto; analiza el texto mediante tokenización y categorías gramaticales; y organiza automáticamente una colección de archivos de texto por tema. También puede utilizar las funciones de AutoML en Amazon Comprehend para crear un conjunto personalizado de entidades o modelos de clasificación de texto que se adapten de forma exclusiva a las necesidades de su organización.

Amazon Comprehend es completamente administrado, por lo que no hay servidores que aprovisionar ni modelos de aprendizaje automático que construir, formar o implementar. Solo pagará por lo que utilice: no se requieren pagos mínimos ni compromisos iniciales.

Sus principales funciones son:

- Extracción de frases claves.
- Análisis de opiniones.
- Análisis sintáctico.
- Reconocimiento de entidad.
- Detección de idioma.

- Clasificación personalizada.
- Compatibilidad con varios idiomas (inglés, francés, alemán, italiano, portugués y español).

4.4 Comparación y Análisis

	Interacción Presencial	Interacción Remota	Interacción Forzada	Interacción Natural
Servqual	X		X	
Procalidad	X		X	
Tesis Iván Pliouchtchai		X		X
Brand24		X		X
Hootsuite		X		X
NN		X		X

*Tabla 6: Comparación con el Estado del Arte. Fuente: Elaboración propia.

	Discriminación Etaria	Discriminación Geográfica	Reglas de Expresión	Soporte de idioma Español	Ofrece un servicio a las empresas
Servqual			X	X	X
Procalidad	X	X	X	X	X
Tesis Iván Pliouchtchai				X	
Brand24					X
Hootsuite					X
NN				X	X

*Tabla 7: Comparación con el Estado del Arte (parte 2). Fuente: Elaboración propia.

Principales diferencias:

En comparación a los métodos de análisis de calidad revisados en 4.1 este trabajo tiene un alcance menos profundo dado que involucra NLP. Por esto, el análisis de calidad que puede realizar este trabajo se limita a identificar los comentarios negativos y sus más recurrentes causas.

En cuanto a lo abarcado en 4.2 hay que destacar que para este trabajo se ocuparon modelos estadísticos y/o herramientas de clasificación ya desarrolladas mientras que los trabajos descritos creaban sus propios modelos en base al análisis de la gramática y sintaxis del lenguaje. Cabe destacar que el acercamiento de los trabajos en 4.2 tienen mayores probabilidades de llegar a mejores resultados, pero a su vez requieren un tiempo de trabajo significativamente mayor.

Finalmente, en comparación a lo que ofrece el mercado, según lo revisado en 4.3, el trabajo realizado es para un nicho muy específico. Las herramientas del mercado trabajan con múltiples empresas y han tenido que robustecer su servicio agregando más funcionalidades que sólo el análisis de sentimientos que ofrece este trabajo, tales como asistencia multimedia, búsqueda inteligente, monitoreo fuera de Twitter, gestión de redes sociales, etc.

Capítulo 5

5. Descripción de la solución

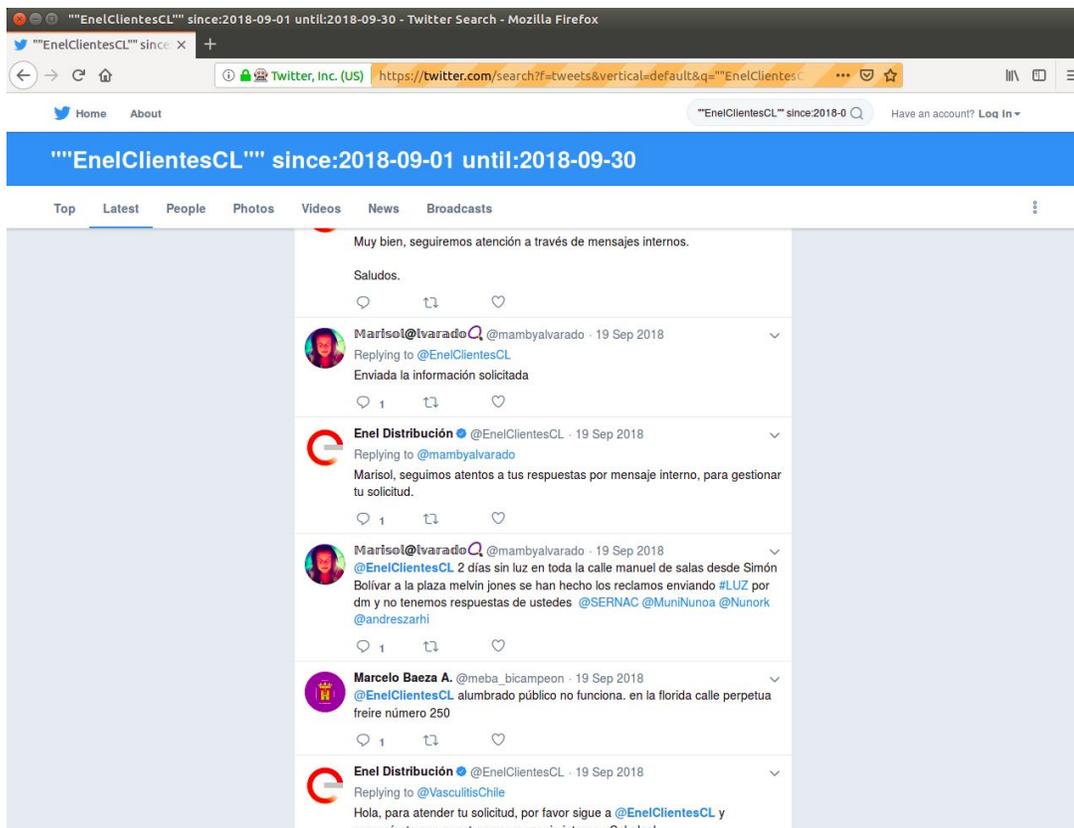
La solución propuesta consiste en desarrollar un sistema de análisis de sentimientos para la percepción de calidad de servicios de internet y electricidad según los/as clientes/as en Twitter.

Como se ha visto en el estado del arte, este tipo de problemas se puede abarcar de muchas formas diferentes, ya sea por el alcance, tiempo disponible u objetivos a cumplir. Para el desarrollo del proyecto, se establecieron 3 principales etapas que serán descritas de manera detallada a continuación.

NOTA: Cabe destacar que todo lo desarrollado para este proyecto en cuanto a programación se realizó en sistema operativo Ubuntu 16.04.

5.1 Descarga de Twitts

Para descargar los Twitts, se utilizó la herramienta Web Scrapping. Se desarrolló el bot que realiza la descarga en Python 3, utilizando bibliotecas Geckodriver, phantomjs y BeautifulSoup. Esto bajo la guía del libro “Web Scraping with python^[18]”.



**Figura 18: Ejemplo de descarga de Twitts de septiembre para Enel Chile. Vista del navegador siendo controlado remotamente por el código 'twitter.py', Fuente: Elaboración Propia.*

El bot se diseñó de la siguiente manera:

- 1) Ingresar a la URL especificada (búsqueda incluida en la URL).
- 2) Realizar ingreso a una cuenta de Twitter.
- 3) Realizar tantos Scrolls como hayan sido definidos.
- 4) Descarga de Twitts cargados con los siguientes atributos: nombre de usuario, alias, fecha de publicación, idioma, likes, compartidos y mensaje (contenido del Twitt).

Los Twitts descargados son sometidos a funciones de limpieza de texto requerida para el uso de la información, los que luego son almacenados en un archivo Json y además en una base de datos MySql y archivos formato “txt” para cada búsqueda. Cabe señalar que las funciones de limpieza de texto pueden ser modificadas en cada iteración en base a un nuevo “ruido” encontrado en el formato del texto, también conocido como *Stopwords*.

5.2 Entrenamiento y Testing

Una vez lista la base de datos se procede con el entrenamiento. Se ha elegido el mecanismo que utiliza la máquina de soporte vectorial (SVM). Esto consiste en expresar la data de la que se quiere aprender a una notación vectorial.

```
1 | 1:1 2:1 3:1 4:1 5:1 6:1 7:1 8:1 9:1 10:1 11:1 12:1 13:1 14:1
2 2 1:1 10:1 14:1 15:1 16:1 17:1 18:2 19:1 20:1 21:1 22:1 23:2 24:1 25:1 26:1 27:1 28:1
3 2 10:1 14:1 21:1 23:1 29:1 30:1 31:1 32:1 33:1 34:1 35:1 36:1 37:1
```

**Figura 19: Ejemplo de notación SVM para Mallet. Elaboración Propia*

La notación vectorial funciona de la siguiente manera: El primer número indica el número de línea del documento y cada línea representa un Twitt. El segundo, indica la clase o etiqueta asignada a cada Twitt (recordar que esto es aprendizaje supervisado). Luego viene la notación vectorial. Esta notación funciona de la manera *feature:value*, donde el *feature* corresponde a una ID única asignada a cada palabra y el *value* indica la frecuencia con la que esa palabra se encuentra en el Twitt.

Para realizar el entrenamiento se escogió la herramienta Scikit learn. Scikit learn es una biblioteca para machine learning en Python. Es una herramienta

muy útil que incluye varios algoritmos de clasificación, regresión y análisis de grupos entre los cuales están las máquinas de soporte vectorial (SVM), *Random Forest*, *Gradient boosting*, *K-means*, entre otros. Scikit learn es reconocida por ser una biblioteca simple y eficiente para el análisis y la minería de datos, *open-source*, comercialmente usable, aplicable a varios contextos y con incorporación de bibliotecas *NumPy*, *Scipy* y *matplotlib*. Grandes e importantes empresas utilizan Scikit learn, como por ejemplo: **Spotify** para la recomendación de música, **Evernote** para todo el proceso de *machine learning*, **Booking.com** para la recomendación de hoteles y destinos a sus usuarios además de la detección de reservaciones fraudulentas, **Change.org** utiliza el clasificador *RandomForest* para dirigir destinos de emails que llegan a millones de usuarios en el mundo cada semana, **OKCupid** lo utiliza en sus sistemas para evaluar y mejorar su sistema de matchmaking (emparejamiento), entre otros.

Para el entrenamiento de la máquina se definieron las siguientes etapas:

- 1) Etiquetado: Como ya se definió en el marco teórico, para realizar aprendizaje supervisado es necesario que la data de la que se quiere aprender esté etiquetada. Para esto, se diseñó y desarrolló un programa en Python que abre y muestra cada Twitt descargado y solicita al usuario que ingrese la etiqueta asociada. Las opciones son representadas con números, donde el número "1" representa que es un Twitt negativo, y el número "0" que no es negativo (positivo o neutral) y, para el caso de Enel Chile, "2" para la clase informativa. Esta es la parte más lenta y tediosa del aprendizaje supervisado, ya que no sólo hay que revisar una gran cantidad de datos (la mayor que sea posible), sino que hay que

asegurarse de etiquetar correctamente, ya que esto tiene un gran impacto en el posterior desempeño del clasificador.

- 2) Entrenamiento: Primero fue necesario crear la bolsa de palabras. Para esto, la biblioteca *Scikit learn* ofrece herramientas de fácil integración y uso. Para la bolsa de palabras se utilizó *Counter()*, herramienta que trabaja revisando las palabras ingresadas a un vector, y realiza un conteo de frecuencias y le asigna una ID a cada palabra. Luego, utilizando vectores nuevamente, se almacena cada pregunta con su correspondiente etiqueta. Finalmente, *Scikit learn* recorre los vectores manteniendo las características de cada Twitt en base a la posición que toman en dicho vector. Esto se entrega a funciones de “*transform*” que permiten pasar los datos al formato de entrada del algoritmo. Finalmente, se almacenan en variables que son ingresadas a cada modelo clasificador.
- 3) Testing: Para el testing, se entregan los datos de prueba, que no fueron entregados al modelo clasificador en la etapa de entrenamiento, al modelo y comienza la clasificación. El algoritmo compara las etiquetas inferidas con las etiquetadas por el usuario.
- 4) Despliegue de resultados: El programa utilizado para el entrenamiento, entrega como salida tablas y métricas con información de desempeño y un archivo que almacena el modelo clasificador entrenado (para posterior uso en data no clasificada).

5.3 Análisis de sentimientos

El análisis de sentimiento se realizó por cada Twitt, tanto en los set de entrenamiento como los de prueba. Se implementó el análisis de sentimientos en tres niveles: Unigrams, Bigrams y Trigrams en estado puro en base a su

frecuencia, es decir, para cada set de datos, se extrajo de manera automática los n-grams ya mencionados que más se repetían como método de detección de la/s principal/es motivación/es de los Twitts.

5.4 Muestreo de resultados

Para el muestreo de resultados del sistema de *machine learning* desarrollado se manejaron los siguientes parámetros y/o métricas:

- Etiquetado: Cantidad de Twitts negativos y no negativos etiquetados por el usuario.
- Entrenamiento y Testing: Matriz de confusión, precision, recall, accuracy, F1-score, Matthews correlation coefficient y ROC AUC SCORE (área bajo la curva ROC).
- Prueba de modelos entrenados en nueva data: Cantidad de Twitts clasificados como negativos y no negativos **por el modelo**.

Para el muestreo de resultados del sistema de análisis de sentimientos desarrollado se consideraron las siguientes métricas:

- Unigrams: Despliegue de las palabras con mayor frecuencia en los Twitts, descartando las Stopwords.
- Bigrams: Despliegue de los pares de palabras consecutivas con mayor frecuencia en los Twitts.
- Trigrams: Despliegue de los tríos de palabras consecutivas con mayor frecuencia en los Twitts.

El muestreo de resultados del trabajo de análisis de sentimientos se realizó a través de gráficos excel automáticamente generados. Esto se realizó utilizando la biblioteca *'xlswriter'* que permite crear archivos excel utilizando muchas funciones como tablas, gráficos, fórmulas, etcétera.

Capítulo 6

6. Pruebas y resultados de la solución

6.1 Pruebas y Resultados para Claro Chile

6.1.1 Etiquetado para Claro Chile

Para el etiquetado, se desarrolló un programa en el lenguaje Python para hacer de esta una tarea más fácil y rápida. El algoritmo recorre los Twitts previamente señalados uno por uno, lo despliega en pantalla y consulta al usuario por la etiqueta que corresponde, ofreciendo 2 opciones: Negativo y No Negativo. Se consideró la idea de utilizar 3 etiquetas (Positivo, negativo, neutral), pero considerando la data que hay que etiquetar y el tiempo que eso conlleva, se prefirió dejar como 2 etiquetas para que la data sea más representativa.

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/WebScraping/Etiquetado.py
Linea: 1270 comentario: clarochile_cl sernac subtel_chile estimados esto es normal todo el dia lo mismo nadie a
tiende ni contesta solo unos dm que envian sin ningun contenido senores clarochile_cl que nefastos y ratas que son
mirenpictwittercomcowxlhum9x
Ingrese (1) si es NEGATIVO, (0) de lo contrario
```

**Figura 20: Etiquetado.py, Twitt real. Fuente: Elaboración Propia*

Para el etiquetado, en el caso de la empresa Claro Chile, se descargaron los Twitts de Octubre y Noviembre de 2018, de los cuales se descartaron todos los Twitts provenientes de las diversas cuentas de la misma empresa ya que lo que se busca es obtener la opinión de los usuarios. La data descargada de octubre corresponde a 2.356 Twitts y la de noviembre corresponde a 2.533 Twitts, dando un total de 4.889 de los cuales 2.601 fueron etiquetados y 2.288 (correspondiente al 47%) fueron descartados por ser off-topic para la finalidad de esta investigación, entre las que destacaron: conversaciones entre usuarios, ofertas de otras compañías, campañas publicitarias, promoción de la Teletón, conversaciones de programas de TV, entre otros.

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/WebScraping/Etiquetado
Linea: 1221 comentario: en el comercial de clarochile_cl donde estan las cabras futbolistas en el momento que se van a la cancha se nota clarito que a una no le llevaron zapatos de futbol me dio lata por ella es la primera que se ve y se le notan sus calcetas blancas con rojo claro
Ingrese (1) si es NEGATIVO, (0) de lo contrario

```

**Figura 21: Ejemplo de Twitt real descartado por off-topic. Fuente: Elaboración Propia.*

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/WebScraping/Etiquetado
Linea: 1602 comentario: clarochile_cl nuevamente muy molesta con los servicios de claro ya que no tengo solucion con mis mensajes de textos ayer twittie aqui mismo y me borraron la publicacion por un cambio de plan que no fue con mi consentimiento solo queda ir al sernac ya que tampoco contesta el call
Ingrese (1) si es NEGATIVO, (0) de lo contrario

```

**Figura 22: Ejemplo de Twitt real etiquetado como negativo. Fuente: Elaboración Propia.*

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/WebScraping/Etiquetado
Linea: 1241 comentario: clarochile_cl buenos dias necesito saber si esta disponible iphone xs max en su cursal a metros de down town
Ingrese (1) si es NEGATIVO, (0) de lo contrario

```

**Figura 23: Ejemplo de Twitt real etiquetado como no negativo. Fuente: Elaboración Propia.*

De los 2.601 Twitts etiquetados, 1.776 (68.3%) fueron etiquetados como negativos y 825 (31.7%) como no negativos.

6.1.2 Entrenamiento y evaluación de modelos predictivos para Claro Chile

Para el proceso de *machine learning*, como ya fue señalado, se basó en la biblioteca de python Scikit learn. A grandes rasgos para el entrenamiento y testing sólo se requirieron 3 bibliotecas:

- **OS:** Biblioteca de python que permite interactuar con componentes del sistema operativo. Fue utilizada para poder recorrer carpetas y acceder a los archivos que contienen los Twitts previamente descargados.
- **Collections:** Biblioteca de python que ofrece contenedores de datos alternativos de alto rendimiento. Fue utilizada para poder crear la bolsa de palabras.
- **Sklearn:** Biblioteca de python creada para desarrollar sistemas de *machine learning* con muchas funcionalidades y modelos listos para su uso. Fue utilizada para importar modelos predictivos, entrenarlos, testarlos, obtener las métricas de rendimiento de los mismos y almacenar los modelos entrenados.

Cada iteración entrega las métricas señaladas de desempeño, y se seleccionó el modelo de cada clasificador con mejor desempeño para la clasificación de data no etiquetada.

Cabe destacar que para cada clasificador, se realizaron 7 iteraciones sobre la misma data. Esto dado que al momento de entrenar, se utilizó un 20% de la

data para testing que fue seleccionado de manera aleatoria en cada iteración, por lo que un mismo modelo sobre una misma data podría tener distinto desempeño dependiendo de qué datos se dejaban para entrenar y cuáles para validar.

A considerar:

- Para simplificar notaciones y aplicación de métricas, así como la comprensión de los resultados, se expresó como “positiva” la clase “no negativa”.
- Para cada modelo se realizaron 7 iteraciones, en todas ellas se utilizaron los parámetros por defecto. Las figuras adjuntas para cada modelo representan el peor y mejor desempeño obtenido respectivamente. Para tomar esta decisión se basó en las métricas Accuracy, MCC y ROC AUC SCORE principalmente, y en todas ellas se refleja el mejor desempeño en la segunda figura mostrada.
- Para escoger el mejor modelo clasificador, se utilizó como punto de comparación la iteración de mejor desempeño de cada modelo. Esto quiere decir, que al momento de comparar, se escogió el mejor desempeño obtenido tanto para SVM como para DecisionTreeClassifier, LogisticRegression y MultinomialNB.
- Es posible identificar a simple vista que para todos los casos, el desempeño sobre la clase negativa fue mucho mejor que la “positiva”. Los posibles motivos de esto se analizarán más adelante.

```

86 #features_train , features test, etiqueta train, etiqueta test
87 x_train, x_test, y_train, y_test = tts(features, labels, test_size=0.2)

```

*Figura 24: Scikit.py, línea 87 en la que se define el tamaño del set de prueba al momento de entrenar.

Fuente: Elaboración Propia.

6.1.2.1 MultinomialNB

```

1 -----
2 Iteracion: 1
3 Matriz de Confusion:
4 [[ 73  96]
5  [ 16 336]]
6 Accuracy: 0.7850287907869482
7 Matthews Correlation Coefficient: 0.48075652794826257
8 Classification Report:
9           precision    recall  f1-score   support
10
11 Clase Positiva      0.82      0.43      0.57      169
12 Clase Negativa      0.78      0.95      0.86      352
13
14   micro avg          0.79      0.79      0.79      521
15   macro avg          0.80      0.69      0.71      521
16   weighted avg       0.79      0.79      0.76      521
17
18 ROC AUC SCORE:
19 0.693249058633674
20 -----

```

*Figura 25: Resultado desempeño clasificador MultinomialNB sobre data clasificada, iteración 1. Fuente:

Elaboración Propia.

```

115 -----
116 Iteracion: 7
117 Matriz de Confusion:
118 [[ 87  70]
119  [ 10 354]]
120 Accuracy: 0.8464491362763915
121 Matthews Correlation Coefficient: 0.6208237523242103
122 Classification Report:
123           precision    recall  f1-score   support
124
125 Clase Positiva      0.90      0.55      0.69      157
126 Clase Negativa      0.83      0.97      0.90      364
127
128   micro avg          0.85      0.85      0.85      521
129   macro avg          0.87      0.76      0.79      521
130   weighted avg       0.85      0.85      0.83      521
131
132 ROC AUC SCORE:
133 0.7633337999580039

```

*Figura 26: Resultado desempeño clasificador MultinomialNB sobre data clasificada, iteración 7. Fuente:

Elaboración Propia.

6.1.2.2 DecisionTreeClassifier

```
20 -----
21 Iteracion: 2
22 Matriz de Confusion:
23 [[111  57]
24 [ 67 286]]
25 Accuracy: 0.761996161228407
26 Matthews Correlation Coefficient: 0.4641145997121518
27 Classification Report:
28           precision    recall  f1-score   support
29
30 Clase Positiva      0.62      0.66      0.64       168
31 Clase Negativa      0.83      0.81      0.82       353
32
33   micro avg         0.76      0.76      0.76       521
34   macro avg         0.73      0.74      0.73       521
35   weighted avg      0.77      0.76      0.76       521
36
37 ROC AUC SCORE:
38 0.7354562929987858|
39 -----
```

*Figura 27: Resultado desempeño clasificador DecisionTreeClassifier sobre data clasificada, iteración 2.

Fuente: Elaboración Propia.

```
39 -----
40 Iteracion: 3
41 Matriz de Confusion:
42 [[125  43]
43 [ 41 312]]
44 Accuracy: 0.8387715930902111
45 Matthews Correlation Coefficient: 0.6298896476009562
46 Classification Report:
47           precision    recall  f1-score   support
48
49 Clase Positiva      0.75      0.74      0.75       168
50 Clase Negativa      0.88      0.88      0.88       353
51
52   micro avg         0.84      0.84      0.84       521
53   macro avg         0.82      0.81      0.81       521
54   weighted avg      0.84      0.84      0.84       521
55
56 ROC AUC SCORE:
57 0.8139501551328747
58 -----
```

*Figura 28: Resultado desempeño clasificador DecisionTreeClassifier sobre data clasificada, iteración 3.

Fuente: Elaboración Propia.

6.1.2.3 SVM

```
58 -----
59 Iteracion: 4
60 Matriz de Confusion:
61 [[123 35]
62 [ 52 311]]
63 Accuracy: 0.8330134357005758
64 Matthews Correlation Coefficient: 0.6182384393078535
65 Classification Report:
66           precision    recall  f1-score   support
67
68 Clase Positiva      0.70      0.78      0.74      158
69 Clase Negativa      0.90      0.86      0.88      363
70
71   micro avg         0.83      0.83      0.83      521
72   macro avg         0.80      0.82      0.81      521
73   weighted avg      0.84      0.83      0.84      521
74
75 ROC AUC SCORE:
76 0.8176151619764969
77 -----
```

**Figura 29: Resultado desempeño clasificador SVM sobre data clasificada, iteración 4. Fuente: Elaboración Propia.*

```
1 -----
2 Iteracion: 1
3 Matriz de Confusion:
4 [[140 26]
5 [ 38 317]]
6 Accuracy: 0.8771593090211133
7 Matthews Correlation Coefficient: 0.7234097971587818
8 Classification Report:
9           precision    recall  f1-score   support
10
11 Clase Positiva      0.79      0.84      0.81      166
12 Clase Negativa      0.92      0.89      0.91      355
13
14   micro avg         0.88      0.88      0.88      521
15   macro avg         0.86      0.87      0.86      521
16   weighted avg      0.88      0.88      0.88      521
17
18 ROC AUC SCORE:
19 0.8681656202273885|
20 -----
```

**Figura 30: Resultado desempeño clasificador SVM sobre data clasificada, iteración 1. Fuente: Elaboración Propia.*

6.1.2.4 LogisticRegression

```
58 -----
59 Iteracion: 4
60 Matriz de Confusion:
61 [[121  43]
62 [ 34 323]]
63 Accuracy: 0.8522072936660269
64 Matthews Correlation Coefficient: 0.6527816105417902
65 Classification Report:
66           precision    recall  f1-score   support
67
68 Clase Positiva      0.78      0.74      0.76       164
69 Clase Negativa      0.88      0.90      0.89       357
70
71   micro avg         0.85      0.85      0.85       521
72   macro avg         0.83      0.82      0.83       521
73   weighted avg      0.85      0.85      0.85       521
74
75 ROC AUC SCORE:
76 0.8212833914053427
77 -----
```

*Figura 31: Resultado desempeño clasificador LogisticRegression sobre data clasificada, iteración 4.

Fuente: Elaboración Propia.

```
39 -----
40 Iteracion: 3
41 Matriz de Confusion:
42 [[129  28]
43 [ 34 330]]
44 Accuracy: 0.8809980806142035
45 Matthews Correlation Coefficient: 0.7206848164006068
46 Classification Report:
47           precision    recall  f1-score   support
48
49 Clase Positiva      0.79      0.82      0.81       157
50 Clase Negativa      0.92      0.91      0.91       364
51
52   micro avg         0.88      0.88      0.88       521
53   macro avg         0.86      0.86      0.86       521
54   weighted avg      0.88      0.88      0.88       521
55
56 ROC AUC SCORE:
57 0.8641247287744103
58 -----
```

*Figura 32: Resultado desempeño clasificador LogisticRegression sobre data clasificada, iteración 3.

Fuente: Elaboración Propia.

6.1.2.5 Análisis de modelos

- MultinomialNB:

Para la clase negativa, de los 364 ejemplares del set de validación, 10 (2.74%) fueron clasificados de manera errónea. Para esta clase, el *precision* de 0.83 indica que tiene muy pocas probabilidades de etiquetar como negativo una muestra positiva, mientras que el *recall* de 0.97 indica que hay casi nulas probabilidades de etiquetar como positiva una instancia negativa.

En cuanto a la clase positiva, de las 157 muestras realmente positivas, 70 (44.58%) fueron clasificadas erróneamente como negativas. En base al *precision* y el *recall* obtenidos para la clase positiva, se concluye que el clasificador tiene pocas probabilidades de etiquetar como positiva una muestra negativa, pero al mismo tiempo pocas probabilidades de identificar todas las instancias positivas como tal, es decir, que podría etiquetar como negativa una muestra positiva.

- DecisionTreeClassifier:

Respecto a la clase negativa, de los 353 ejemplares realmente negativos, 41 (11.61%) fueron clasificados de manera errónea. El *precision* con un valor de 0.88 es una buena señal, considerando que está más cerca del 1.00 que del 0.5, lo que indica que es poco probable que identifique como positiva una muestra negativa. Un caso muy similar se presenta para el *recall* que obtuvo el mismo valor de 0.88, lo que indica que hay las mismas pocas probabilidades de

identificar una muestra negativa como positiva como de identificar una positiva como negativa.

Según lo obtenido para la clase positiva, de las 168 muestras, 43 (25.59%) fueron clasificadas de manera errónea. Como era de esperar, según lo analizado para la clase negativa, este modelo también entregó valores muy similares de *recall* y *precision* para la clase positiva, tomando los valores 0.74 y 0.75 respectivamente, por lo que las conclusiones entre las relaciones de ambas clases se mantiene según lo expresado en el análisis sobre la clase negativa de este modelo.

- SVM:

En cuanto al desempeño sobre la clase negativa, SVM etiquetó erróneamente 38 (11.98%) muestras del total de 317 ejemplares realmente negativos. Con un *precision* de 0.92 y un *recall* de 0.89 este modelo indica que, para la clase negativa, tiene mayor certeza que incertidumbre al momento de etiquetar.

Por otro lado, para la clase positiva, de las 166 muestras realmente positivas, 26 (15.66%) fueron etiquetadas de manera errónea. Si bien el modelo tuvo un peor desempeño para la clase positiva que la negativa, es posible apreciar que hay una correlación entre los desempeños de las distintas clases, ya que, al ser clases dicotómicas, tiene sentido que las fortalezas y falencias para identificar una clase se vean reflejadas en las fortalezas y falencias de su clase opuesta. Aplicado a este caso específico (modelo y clase), se verifica que si en la clase negativa el *precision* > *recall*, en su clase opuesta ocurre lo contrario (*recall* > *precision*).

- LogisticRegression:

Del total de 364 ejemplares de la clase negativa, 34 (9.34%) fueron clasificados de manera errónea por LogisticRegression. Para esta clase, el modelo obtuvo los valores de 0.92 en *precision* y 0.91 en *recall*. Estos son muy buenos resultados, muy cercanos al 1.00 (desempeño perfecto), lo que indica que para la clase logró encontrar la mayor parte de los ejemplares y que, lo etiquetado es de muy buena calidad (nivel de certeza al etiquetar).

En cuanto a su desempeño para la clase positiva, del total de 157 ejemplares, 28 (17.83%) fueron etiquetados incorrectamente. Esto permite llegar a la aseveración de que la clase positiva obtuvo peor desempeño que la negativa para todos los modelos utilizados, para lo cual se postulan 2 posibles causas: la primera, que es la clase con menor representación en la data de entrenamiento, por lo que el modelo tiene menos data de la que aprender; la segunda, que al ser dos clases representadas en una (positivo y neutral), puede haber dificultado la tarea de asociar las *features* a la clase, debido a que podía darse la relación de más *features* en menos muestras. Además, considerar que algunas *features* podían aplicar tanto para la clase positiva como negativa, lo que suele generar dificultades en el campo del NLP (procesamiento de lenguaje natural).

6.1.2.6 Comparación de modelos

	LogisticRegression (3)	SVM (1)	DecisionTreeClassifier (3)	MultinomialNB (7)
Accuracy	0,880	0,877	0,838	0,846
MCC	0,72	0,723	0,629	0,62
ROC AUC SCORE	0,864	0,868	0,813	0,763

**Figura 33: Tabla comparativa de desempeño de modelos. Fuente: Elaboración Propia.*

Como se puede apreciar en la tabla, hay 2 modelos clasificadores que se descartaron de manera inmediata. DecisionTreeClassifier y MultinomialNB, ambos en su iteración con mejores resultados (3 y 7 respectivamente), entregaron los peores resultados en cuanto a métricas de desempeño. Esto resultó ser una sorpresa para el caso de MultinomialNB, dado que por lo investigado en la literatura y trabajos científicos, es un modelo muy utilizado en la predicción sobre texto y ha demostrado dar buenos resultados cuando es utilizado con SVM (Jin Huanh et al. (2003); Dr. S. Vajayarani, Mr. S. Dhayanand (2015); Selina S. Y. Ng et al. (2014)). Una posible explicación de estos resultados puede ser lo desbalanceado de las clases, donde se observa una alta predominancia de la clase negativa por sobre la positiva. Para esto la literatura (ej: Eibe Frank, Remcko R. Bouckaert, 2006) ha estudiado medidas como la normalización del conteo de palabras para clases de distinto tamaño en la parte de preprocesamiento, lo que ha dado considerables mejoras de desempeño del clasificador. Además, se propone realizar tantas pruebas como sea posible modificando los parámetros del modelo hasta encontrar una iteración con mejor desempeño, lo que fue descartado por limitaciones de tiempo. Por otro lado, la literatura también hace referencia a asunciones erróneas de parte del clasificador por sobre los features (Karl-Michael Schneider, 2005) lo que hace sentido para el contexto ya en el proceso de

etiquetado se comprobó que muchas palabras se repetían o aplicaban para ambas clases, lo que habría confundido al modelo clasificador.

Por otro lado, la elección del mejor modelo para este contexto ha sido más complejo de lo esperado. Como muestra la tabla, LogisticRegression fue el modelo con mayor *Accuracy* de todos, lo que indica que es el que tuvo una mayor cantidad de predicciones correctas. En cuanto a SVM, tuvo el mayor *MCC* y *ROC AUC SCORE* dentro de los modelos comparados, lo que tiene suma importancia dado que ambas métricas son muy utilizadas al momento de comparar modelos por ser balanceadas y más significativas que el *accuracy*, dado que también consideran la data incorrectamente clasificada.

En una vista rápida sería lógico determinar que SVM es definitivamente el modelo con mejor desempeño dentro de los estudiados, pero, si se observan las cifras más de cerca, se evidencia que lo que separa a SVM y LogisticRegression no es mayor a 0,004 en todas las métricas señaladas. Dado que la brecha es increíblemente estrecha y que la data sobre la que serán utilizados estos modelos no supera los 10.000 Tweets (en total), se ha considerado que la brecha es despreciable y, por lo tanto, se utilizarán ambos modelos para clasificar la data no etiquetada de los meses posteriores (Diciembre y Enero).

6.1.3 Modelos predictivos en data no etiquetada para Claro Chile

Para probar los modelos previamente entrenados, se desarrolló otro código Python que, utilizando la misma biblioteca Scikit learn, importa los modelos e inicia el proceso de etiquetado automático, según lo aprendido, sobre la data no clasificada. El modelo entrega el número de Twitts clasificados como Negativos

y el número de Twitts clasificados como No Negativos además de la clasificación en detalle mostrando cada Twitt y la etiqueta asignada. Esto se realiza con la data descargada por mes, por lo que es posible señalar en porcentajes la reputación otorgada por los usuarios en dicho mes.

Como ya fue mencionado, para obtener las predicciones se utilizó LogisticRegression y SVM dado que fueron los modelos con mejor desempeño, el cual era muy sobresaliente por sobre MultinomialNB y DecisionTreeClassifier y a la vez despreciable la superioridad de uno sobre el otro.

6.1.3.1 LogisticRegression

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo
-----
nuevamente estoh5sin el servicio de llamadas por clarochile claro
Negativo!
-----
ayudamovistarcl hasta cuando xux espero q arreglen internet llevo 2 meses esperan
do ofrecen cosas y despues nada nos vamos vtrchile clarochile_cl
Negativo!
-----
clarochile_cl miclaro_cl se puede sacar una segunda linea movil que sea cuenta e
xacta
Positivo/Neutro
-----
clarochile_cl buenas tardes necesito dar de baja el servicio de internet que ten
ia contratado en su cmpnia pero ya llevamos 4 nros de atencion y no lo hemos podi
do lograr porque se corta el llamado
Negativo!
-----
es una paja pero es peor cuando te persiguen por deudas que no son tuyas
Negativo!
-----
clarochile_cl otra vez con problemas de servicio claronofunciona cauquenes
Negativo!
Total Positivos: 1075 Total Negativos: 1745
tomas@tomas-OMEN-by-HP-Laptop:~/Escritorio/Universidad/Tesis/Codigo Twitter/Codig
o Descarga KC (este)/WebScraping/Etiquetado$
```

*Figura 34: Prueba de modelo LogisticRegression (3) entrenado con la data etiquetada de Octubre-2018 y Noviembre-2018, evaluando la data de Septiembre-2018. Fuente: Elaboración Propia.

Como se muestra en la figura, de un total de 2.820 Twitts, el modelo que en su entrenamiento obtuvo un 88% de *accuracy*, etiquetó 1.745 como Negativos y 1.075 como No Negativos. Esto quiere decir que en septiembre, aproximadamente el 61.87% de los Twitts publicados fueron identificados como negativos y un 38.13% como positivos/neutrales .

```

tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo D
-----
clarochile_cl estimados urgente necesito orden en la cuenta no pude pagar uno de m
is dos lineas y estoy sin servicio en ambas tienen un gran desorden

Negativo!
-----
subsecretaria pamela gidi pangidi buenas noches a traves de que medio podriamos ha
cerle llegar los antecedentes de este caso para asi acelerar el proceso de investi
gacion y facilitarle el trabajo al vecino nos parece muy serio y grave lo ocurrido
estaremos a la espera

Positivo/Neutro
-----
plop

Positivo/Neutro
-----
bienvenido al dick en el ojo

Positivo/Neutro
-----
pisapapeles_net huawei p20 lite con la ultima actualizacion c25 permite las llama
as sobre 4g lte en clarochile_cl huaweimobileclpictwittercom5hgqrjojcm

Positivo/Neutro
-----
listo

Positivo/Neutro
-----
compra mejor una tarjeta 4g

Positivo/Neutro
-----
hola claroteayuda clarochile_cl estoy en brasil y necesito contratar roaming no
lo hice en chile como lo hago es un poco urgente gracias

Positivo/Neutro
-----
clarochile_cl asi es supera la ficcuon pagamos un servicio y este no funciona plop

Negativo!
Total Positivos: 568 Total Negativos: 733
tomas@tomas-OMEN-by-HP-Laptop:~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo
Descarga KC (este)/WebScraping/Etiquetados

```

**Figura 35: Prueba de modelo LogisticRegression (3) entrenado con la data etiquetada de Octubre-2018 y Noviembre-2018, evaluando la data de Diciembre-2018. Fuente: Elaboración Propia.*

En cuanto a la data de Diciembre, de un total de 1.301 Twitts, 733 (56.34%) fueron clasificados como negativos y 568 (43.66%) como positivos/neutrales.

6.1.3.2 SVM

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo
-----
expreso mi molestia con clarochile_cl no puedo portar un numero prepago claro a ot
ra compania porque aparece que no soy el titular del numero he ido mas de una vez
a las oficinas de claro y aun no entregan una solucion pesimo servicio a los clien
tes sernac subtcl_chile

Negativo!
-----
gracias llamo ahora adios movistar

Positivo/Neutro
-----
hola franjd22 si necesita ayuda comercial puede llamar al 6008009000 y desde celul
ares al 2 23100200 un asesor respondera a sus preguntas muchas gracias

Positivo/Neutro
-----
nuevamente estoh5sin el servicio de llamadas por clarochile claro

Negativo!
-----
ayudamovistarcl hasta cuando xux espero q arreglen internet llevo 2 meses esperand
o ofrecen cosas y despues nada nos vamos vtrchile clarochile_cl

Negativo!
-----
clarochile_cl miclaro_cl se puede sacar una segunda linea movil que sea cuenta ex
acta

Positivo/Neutro
-----
clarochile_cl buenas tardes necesito dar de baja el servicio de internet que tenti
a contratado en su cmpnia pero ya llevamos 4 nros de atencion y no lo hemos podido
lograr porque se corta el llamado

Negativo!
-----
es una paja pero es peor cuando te persiguen por deudas que no son tuyas

Negativo!
-----
clarochile_cl otra vez con problemas de servicio claronofunciona cauquenes

Negativo!
Total Positivos: 1103 Total Negativos: 1717
tomas@tomas-OMEN-by-HP-Laptop:~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo
Descarga KC (este)/WebScraping/Etiquetado$
```

**Figura 36: Prueba de modelo SVM (1) entrenado con la data etiquetada de Octubre-2018 y Noviembre-2018, evaluando la data de Septiembre-2018. Fuente: Elaboración Propia.*

Como se muestra en la figura, de un total de 2.820 Twitts, el modelo que en su entrenamiento obtuvo un 87.7% de *accuracy*, etiquetó 1.717 como Negativos y 1.103 como No Negativos. Esto quiere decir que en septiembre, aproximadamente el 60.88% de los Twitts publicados fueron identificados como negativos y un 39.12% como positivos/neutrales.

```

tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo
-----
clarochile_cl estimado urgente necesito orden en la cuenta no pude pagar uno de m
is dos lineas y estoy sin servicio en ambas tienen un gran desorden
Negativo!
-----
subsecretaria pamele gidi pangidi buenas noches a través de que medio podríamos ha
cerle llegar los antecedentes de este caso para así acelerar el proceso de investi
gación y facilitarle el trabajo al vecino nos parece muy serio y grave lo ocurrido
estaremos a la espera
Positivo/Neutro
-----
plop
Positivo/Neutro
-----
bienvenido al dick en el ojo
Positivo/Neutro
-----
pisapapeles_net huawei p20 lite con la última actualización c25 permite las llama
das sobre 4g lte en clarochile_cl huaweimobileclpictwittercom5hgqrjocm
Positivo/Neutro
-----
listo
Positivo/Neutro
-----
compra mejor una tarjeta 4g
Positivo/Neutro
-----
hola claroteayuda clarochile_cl estoy en brasil y necesito contratar roaming no
lo hice en chile como lo hago es un poco urgente gracias
Positivo/Neutro
-----
clarochile_cl así es supera la ficción pagamos un servicio y este no funciona plop
Negativo!
Total Positivos: 574 Total Negativos: 727
tomas@tomas-OMEN-by-HP-Laptop:~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo
Descarga KC (este)/WebScraping/Etiquetado$

```

**Figura 37: Prueba de modelo SVM (1) entrenado con la data etiquetada de Octubre-2018 y Noviembre-2018, evaluando la data de Diciembre-2018. Fuente: Elaboración Propia.*

En cuanto a la data de Diciembre, de un total de 1.301 Twitts, 727 (55.88%) fueron clasificados como negativos y 574 (44.12%) como positivos/neutrales.

6.1.3.3 Comparación LogisticRegression vs SVM

septiembre-2018	LogisticRegression (3)	SVM (1)	Diferencia	Promedio
Negativos	1745	1717	28	1731
Positivos/Neutrales	1075	1103	-28	1089

**Tabla 6: Tabla comparativa LogisticRegression vs SVM sobre data correspondiente a septiembre-2018. Fuente: Elaboración Propia.*

Como era de esperar, al igual que sobre la data de *testing*, ambos modelos presentaron resultados muy similares sobre la data de septiembre. Su diferencia fue menor al 1%.

diciembre-2018	LogisticRegression (3)	SVM (1)	Diferencia	Promedio
Negativos	733	727	6	730
Positivos/Neutrales	568	574	-6	571

**Tabla 7: Tabla comparativa LogisticRegression vs SVM sobre data correspondiente a diciembre-2018.*

Fuente: Elaboración Propia.

Al igual que sobre la data de septiembre, la diferencia fue despreciable. Sólo 6 Twitts fueron etiquetados de manera diferente entre ambos modelos, lo que representa un 0.46% de los 1.301 Twitts.

Tanto para la data de septiembre como para la de diciembre, la diferencia se produjo en que SVM etiquetó como negativa data que LogisticRegression etiquetó como positiva/neutral.

Un posible motivo de esto puede ser que para el entrenamiento, SVM tuvo mayor soporte en la clase positiva/neutral que LogisticRegression, lo que llevaría a pensar que esa diferencia se produjo por el hecho de que SVM tenga una (levemente) mayor incertidumbre que LogisticRegression al momento de etiquetar algo como negativo, por lo que resulta lógico pensar que esas pocas etiquetas que pasaron de positivo/neutral a negativo entre un modelo y otro corresponden a un etiquetado erróneo por parte del modelo SVM, a pesar de haber tenido un (levemente) mejor desempeño general sobre la data de prueba.

A raíz de las pruebas realizadas sobre la data de septiembre-2018 y diciembre-2018, se puede agregar que:

- La cantidad de Twitts **negativos** emitidos por los usuarios mes a mes para Claro Chile fluctúa entre el 55.88% y el 61.87%.
- Del 100% de Twitts mostrados en la búsqueda “ClaroChile”, en promedio un 37.9% corresponde a cuentas de Claro Chile, entre los cuales destacan respuestas a dudas e inquietudes, además de quejas de usuarios y publicidad de promociones de productos y/o servicios.

6.1.4 Análisis de Sentimientos

Se realizaron pruebas de análisis de sentimientos para toda la data descargada, es decir, los Tweets utilizados para entrenamiento y los utilizados para prueba. A continuación, se muestra el resultado de cada prueba.

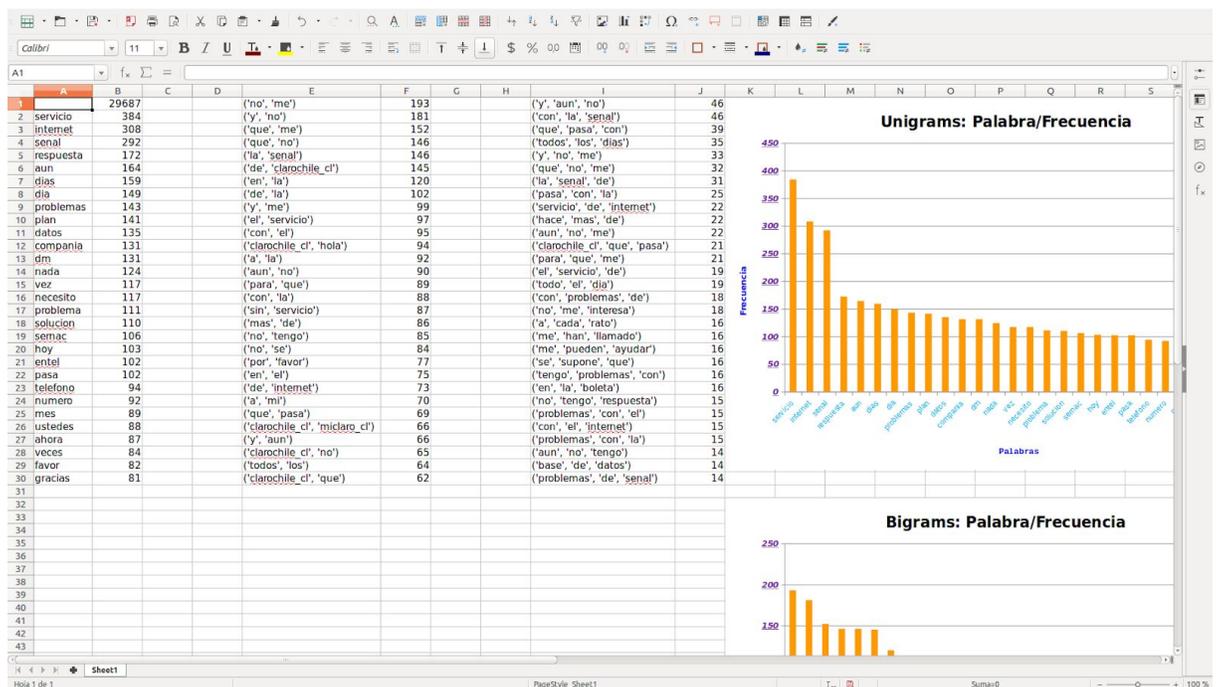
Es importante señalar que el código de análisis de sentimiento es el que más tiempo de ejecución toma en el flujo de todo el proyecto de investigación (excluyendo la descarga a nivel mensual), esto ya que como se explicó anteriormente, para obtener los uni-grams, se pasan las palabras por un filtro de stopwords (palabras sin valor), por lo que en la práctica, el algoritmo realiza (para el caso de septiembre) la comparación de 55.480 palabras (provenientes de los Tweets) con las 224 definidas como stopwords, para así eliminar las coincidentes. Para este caso particular, el algoritmo tomó 453.72 segundos, lo que equivale a 7.56 minutos.

Notar además que el análisis realizado es progresivo, por lo que se intentará evitar la redundancia de análisis mes a mes a menos de ser estrictamente

necesario ya que se espera identificar tantos factores que motivan los Tweets como sea posible.

A continuación se muestran los resultados generados de manera automática en gráficos excel:

6.1.4.1 Septiembre-2018



*Figura 38: Vista general de archivo excel generado por el código `Sentiment_Analysis.py` para la data de septiembre-2018. Fuente: Elaboración Propia.

Como se aprecia en la figura, la salida del código es un archivo excel que contiene una selección de los “X” n-grams con mayor frecuencia, con su

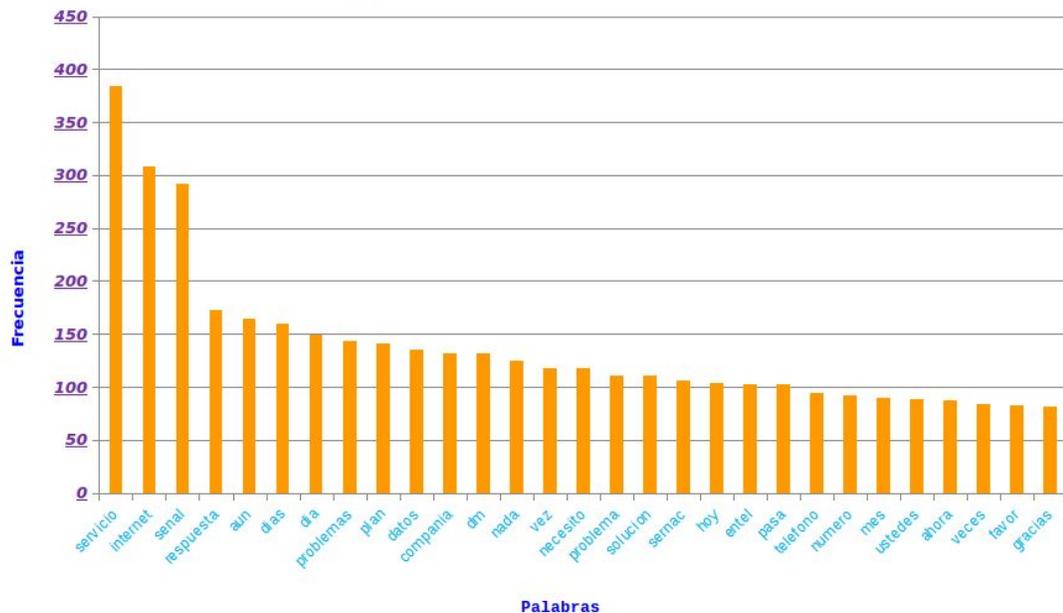
respectivo valor y gráfico representativo. Para el ejemplo de la figura, se escogieron los 30 uni, bi y tri-grams de mayor frecuencia.

NOTA: Por motivos de espacio, el gráfico de tri-grams quedó fuera de la imagen, pero se encuentra debajo del gráfico de bi-grams.

	29687		('no', 'me')	193		('y', 'aun', 'no')	46
servicio	384		('y', 'no')	181		('con', 'la', 'senal')	46
internet	308		('que', 'me')	152		('que', 'pasa', 'con')	39
senal	292		('que', 'no')	146		('todos', 'los', 'dias')	35
respuesta	172		('la', 'senal')	146		('y', 'no', 'me')	33
aun	164		('de', 'clarochile cl')	145		('que', 'no', 'me')	32
dias	159		('en', 'la')	120		('la', 'senal', 'de')	31
dia	149		('de', 'la')	102		('pasa', 'con', 'la')	25
problemas	143		('y', 'me')	99		('servicio', 'de', 'internet')	22
plan	141		('el', 'servicio')	97		('hace', 'mas', 'de')	22
datos	135		('con', 'el')	95		('aun', 'no', 'me')	22
compania	131		('clarochile cl', 'hola')	94		('clarochile cl', 'que', 'pasa')	21
dm	131		('a', 'la')	92		('para', 'que', 'me')	21
nada	124		('aun', 'no')	90		('el', 'servicio', 'de')	19
vez	117		('para', 'que')	89		('todo', 'el', 'dia')	19
necesito	117		('con', 'la')	88		('con', 'problemas', 'de')	18
problema	111		('sin', 'servicio')	87		('no', 'me', 'interesa')	18
solucion	110		('mas', 'de')	86		('a', 'cada', 'rato')	16
semac	106		('no', 'tengo')	85		('me', 'han', 'llamado')	16
hoy	103		('no', 'se')	84		('me', 'pueden', 'ayudar')	16
entel	102		('por', 'favor')	77		('se', 'supone', 'que')	16
pasa	102		('en', 'el')	75		('tengo', 'problemas', 'con')	16
telefono	94		('de', 'internet')	73		('en', 'la', 'boletera')	16
numero	92		('a', 'mi')	70		('no', 'tengo', 'respuesta')	15
mes	89		('que', 'pasa')	69		('problemas', 'con', 'el')	15
ustedes	88		('clarochile cl', 'mclaro cl')	66		('con', 'el', 'internet')	15
ahora	87		('y', 'aun')	66		('problemas', 'con', 'la')	15
veces	84		('clarochile cl', 'no')	65		('aun', 'no', 'tengo')	14
favor	82		('todos', 'los')	64		('base', 'de', 'datos')	14
gracias	81		('clarochile cl', 'que')	62		('problemas', 'de', 'senal')	14

*Figura 39: Vista tablas n-grams de archivo excel generado por el código *Sentiment_Analysis.py* para la data de septiembre-2018. Fuente: *Elaboración Propia*.

Unigrams: Palabra/Frecuencia



*Figura 40: Vista gráfico uni-grams de archivo excel generado por el código `Sentiment_Analysis.py` para la data de septiembre-2018. Fuente: Elaboración Propia.

Los uni-grams aplicados a un contexto pueden entregar mucha información. A primera vista en el gráfico se pueden reconocer inmediatamente 4 aspectos muy importantes a analizar y que, coincidentemente, son los con mayor frecuencia y por ende, el principal motivo de los Tweets: servicio, respuesta, internet y señal.

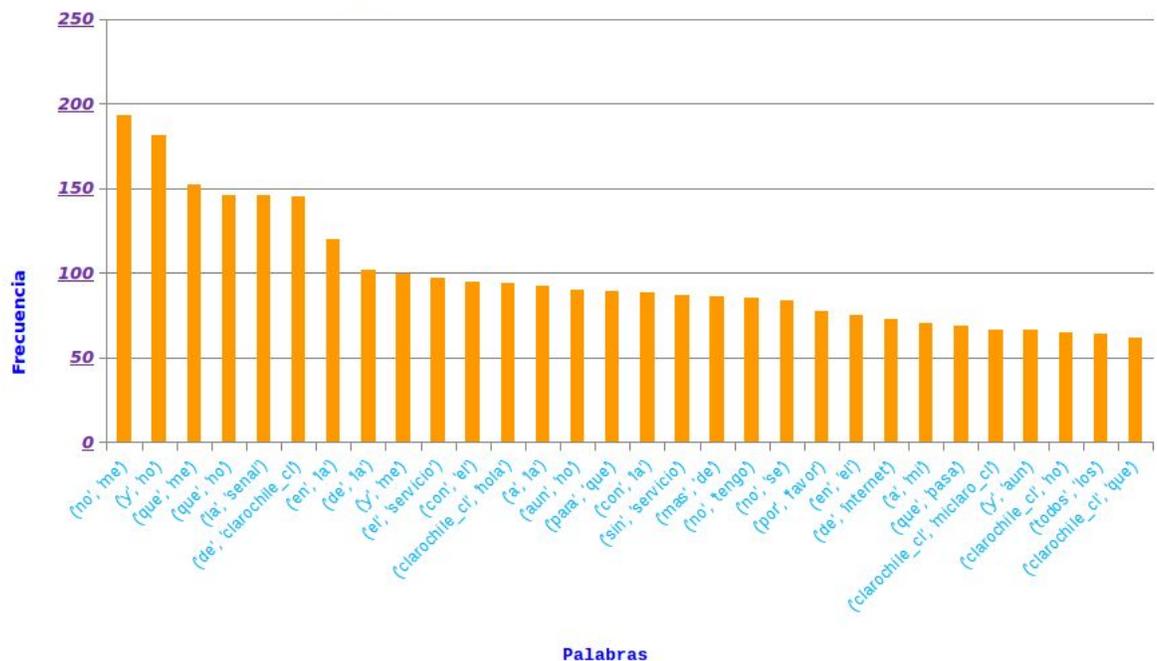
Además, palabras como “sernac” (servicio nacional del consumidor) y “problema” son un indicador directo de una denuncia por parte del consumidor o cliente.

A raíz de la revelación estadística en base a la polaridad de los Tweets obtenida en la etapa de prueba de modelos, se sabe que más de la mitad de los

comentarios son negativos. Teniendo esto en cuenta, se le puede dar valor a palabras que de por sí podrían no entregar valor, como por ejemplo las palabras relacionadas con el tiempo: día, días, mes, o palabras como: aún, vez, veces.

Para quien realiza el etiquetado, es bastante fácil suponer los contextos en los que se presentan dichas palabras: “aún no tengo respuesta”, “llevo todo el día sin internet”, “días con este problema”, “he llamado X veces”, etcétera. Pero aún así, los uni-grams son poco concluyentes, por lo que se avanza a la siguiente etapa: los bi-grams.

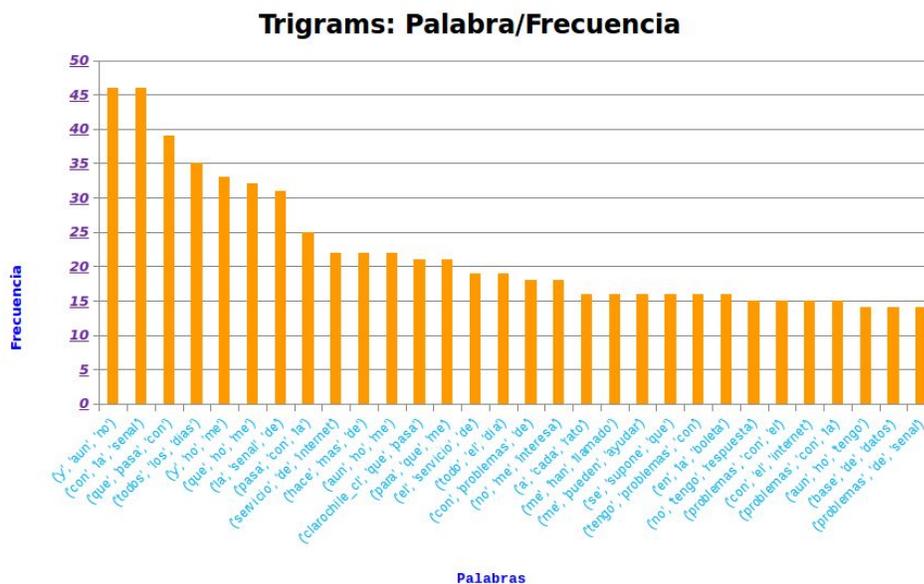
Bigrams: Palabra/Frecuencia



*Figura 41: Vista gráfico bi-grams de archivo excel generado por el código Sentiment_Analysis.py para la data de septiembre-2018. Fuente: Elaboración Propia.

Retomando palabras encontradas en la selección de uni-grams, resulta interesante destacar: “la señal”, “el servicio”, “aún no”, “no tengo”, “y aún”, “qué pasa”, “todos los”. Aquí resulta un poco más claro el mensaje que la gente busca entregar y su polaridad. Por ejemplo, todos los “aún” indican que hay una acción o condición primaria, que produce que el usuario espere que ocurra o se cumpla algo que sigue sin ocurrir.

Bi-grams un poco más claros son “sin servicio” y “qué pasa”, si se conecta con otros bi-grams de igual o similar frecuencia, es posible asumir que se presentaron en los mismos Tweets, como por ejemplo: “sin servicio” y “de internet”; “aún no” y “no tengo”; “no tengo” y “el servicio”. Y si se unen más, se pueden formar frases como: “qué pasa”, “con la”, “la señal”.



*Figura 42: Vista gráfico tri-grams de archivo excel generado por el código `Sentiment_Analysis.py` para la data de septiembre-2018. Fuente: Elaboración Propia.

Los tri-grams son sin duda los n-grams que mayor información entregan sin ser necesario un análisis, por lo que, si los analiza alguien inmerso en el contexto, es capaz de sacar información relevante y concluyente de las opiniones sobre la empresa.

Ahora además, es posible unir el análisis realizado a los uni y bi-grams con lo que entregan los tri-grams. El tri-gram con mayor frecuencia es “y aún no” y se pudo apreciar en los n-grams anteriores que las palabras “aún” y “respuesta” eran de las con mayor frecuencia. Aquí se hace evidente que los usuarios pasan tiempo esperando que haya alguna respuesta o solución por parte de Claro Chile para algún “problema” que están presenciando.

Los dos siguientes tri-grams con mayor frecuencia son “qué pasa con” y “con la señal”. Para este punto del análisis no hay duda que el mayor motivo de los Tweets son los problemas con la señal. Continuando en orden de frecuencia aparece “todos los días” lo que indica que el motivo del Tweet, sea cual sea, es algo que el usuario vive día a día, lo que puede resultar molesto, frustrante y generar una muy mala imagen de la compañía para el cliente, lo que no sólo podría motivar su cambio de compañía, sino que reduce considerablemente las posibilidades de que recomiende la compañía a un colega, amigo/a o cercano/a.

Nuevamente, resulta interesante analizar n-grams que tienen la misma frecuencia o una frecuencia similar, ya que resulta válido asumir que se presentaron en los mismos Tweets. Considerando aquello, tri-grams con frecuencias muy similares son: “el servicio de”, “todo el día”, “con problemas de”. De a poco se van corroborando asunciones, como que los usuarios están “con problemas de” durante “todo el día”. Y, ya que vimos que la señal es el

principal motivo de los Tweets, se podría asumir que la siguiente frase representa a una gran cantidad de usuarios: “ ‘el servicio de’ internet ha estado ‘con problemas de’ señal ‘todo el día’ “. Siguiendo con el mismo tipo de análisis, los tri-grams “me han llamado” y “a cada rato” tienen exactamente la misma frecuencia. Luego de haber etiquetado la data, teniendo conocimiento del contexto estudiado, una queja muy frecuente entre los usuarios son los constantes llamados de call center, ya sea para ofrecer promociones o para el cobro de servicios. De hecho, con una frecuencia levemente mayor, aparece “no me interesa” y “base de datos”, ambas frases presentes al momento de etiquetar en Tweets que solicitan que cesen los continuos llamados que no hacen más que hostigar, por lo que piden ser borrados de la base de datos por falta de interés en lo que ofrecen.

Un poco más abajo aparece “en la boleta” lo que, estando familiarizado con el contexto, se debe a que muchos Tweets de la data etiquetada (que no incluía el mes de septiembre) hacían referencia a cobros indebidos en la boleta, lo que genera desconfianza y sensación de fraude para el cliente.

Finalmente, los últimos tri-grams desplegados son bastante claros en lo que buscan expresar: “no tengo respuesta”, “problemas con el”, “con el internet”, “problemas con la” y “problemas de señal”.

Finalmente, para el mes de septiembre, se definen como principales problemas:

- Los tiempos de espera de los clientes y la capacidad de respuesta de Claro Chile.
- La calidad de la señal.
- La calidad del internet.

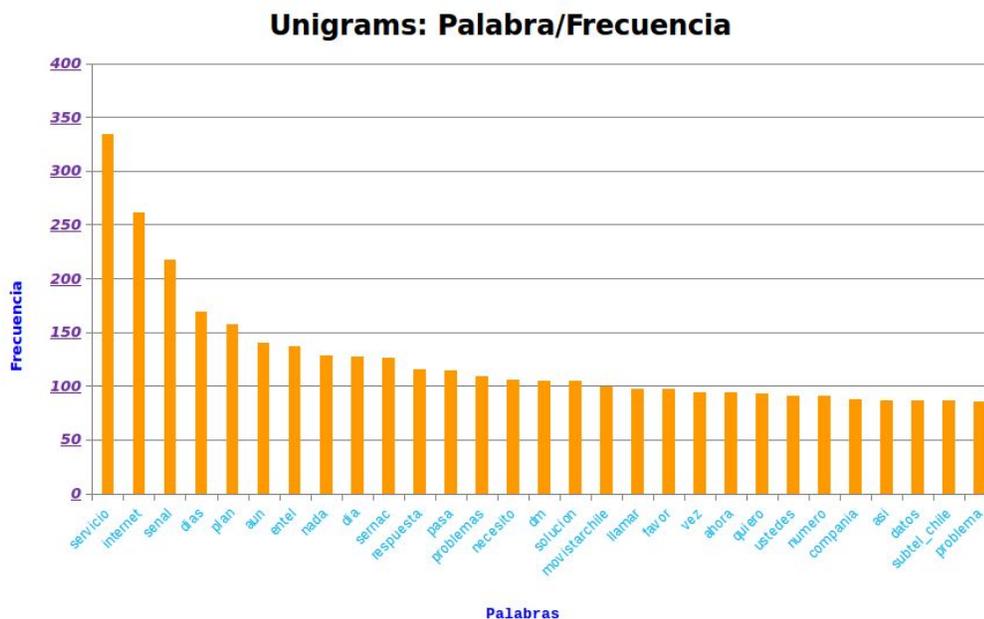
- El acoso realizado de manera telefónica por el call center.
- Los cobros indebidos en las boletas.

6.1.4.2 Octubre-2018

De aquí en adelante se omitirá la vista general del documento excel, priorizando los listados de n-grams y sus respectivos gráficos.

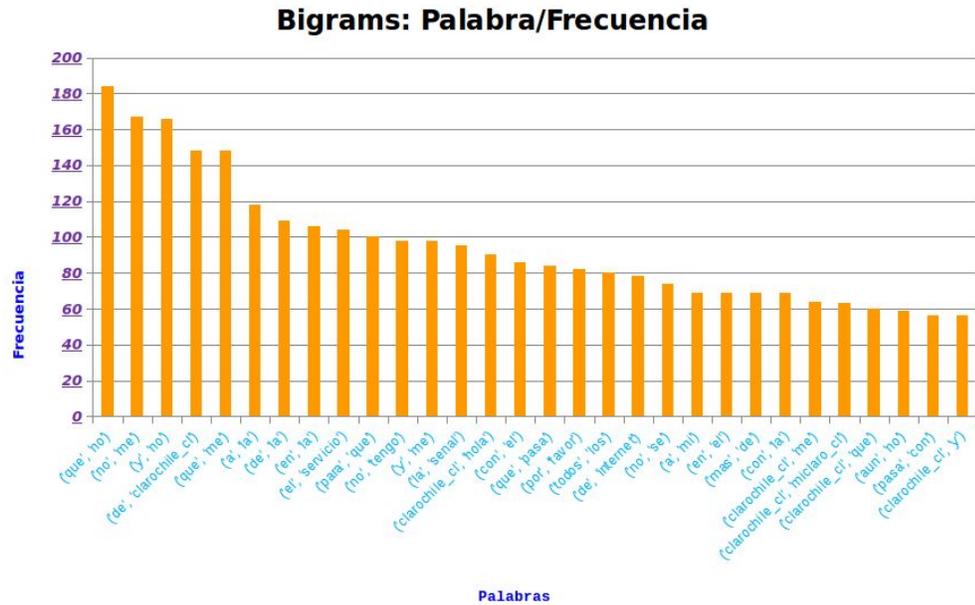
	28281		('que', 'no')	184		('que', 'pasa', 'con')	51
servicio	334		('no', 'me')	167		('todos', 'los', 'dias')	39
internet	261		('y', 'no')	166		('con', 'la', 'senal')	37
senal	217		('de', 'clarochile cl')	148		('la', 'senal', 'de')	33
dias	169		('que', 'me')	148		('que', 'no', 'me')	28
plan	157		('a', 'la')	118		('y', 'no', 'me')	26
aun	140		('de', 'la')	109		('clarochile cl', 'que', 'pasa')	26
entel	137		('en', 'la')	106		('y', 'aun', 'no')	25
nada	128		('el', 'servicio')	104		('todo', 'el', 'dia')	24
dia	127		('para', 'que')	100		('pasa', 'con', 'la')	23
semac	126		('no', 'tengo')	98		('que', 'no', 'tengo')	22
respuesta	115		('y', 'me')	98		('con', 'el', 'internet')	21
pasa	114		('la', 'senal')	95		('no', 'me', 'interesa')	20
problemas	109		('clarochile cl', 'hola')	90		('no', 'tengo', 'senal')	20
necesito	106		('con', 'el')	86		('con', 'el', 'servicio')	19
dm	105		('que', 'pasa')	84		('el', 'servicio', 'de')	19
solucion	105		('por', 'favor')	82		('para', 'que', 'me')	19
movistarchile	99		('todos', 'los')	80		('velocidad', 'de', 'internet')	19
llamar	97		('de', 'internet')	78		('servicio', 'de', 'internet')	18
favor	97		('no', 'se')	74		('que', 'dejen', 'de')	18
vez	94		('a', 'mi')	69		('no', 'me', 'llamen')	18
ahora	94		('en', 'el')	69		('dar', 'de', 'baja')	18
quiero	93		('mas', 'de')	69		('a', 'mi', 'me')	17
ustedes	91		('con', 'la')	69		('me', 'llamen', 'mas')	17
numero	91		('clarochile cl', 'me')	64		('dejen', 'de', 'llamarme')	17
compania	87		('clarochile cl', 'miclaro cl')	63		('hace', 'mas', 'de')	16
asi	86		('clarochile cl', 'que')	60		('a', 'la', 'sucursal')	16
datos	86		('aun', 'no')	59		('dejen', 'de', 'llamar')	15
subtel_chile	86		('pasa', 'con')	56		('no', 'tengo', 'internet')	14
problema	85		('clarochile cl', 'y')	56		('aun', 'no', 'tengo')	14

*Figura 43: Vista tablas n-grams de archivo excel generado por el código Sentiment_Analysis.py para la data de octubre-2018. Fuente: Elaboración Propia.



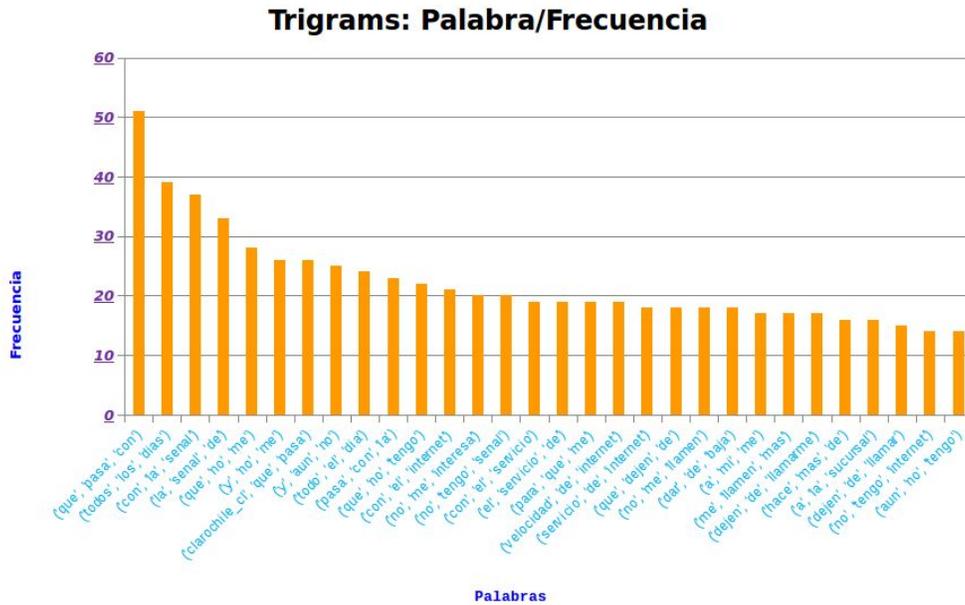
**Figura 44: Vista gráfico uni-grams de archivo excel generado por el código Sentiment_Analysis.py para la data de octubre-2018. Fuente: Elaboración Propia.*

Nuevamente “servicio”, “internet” y “señal” son la principal motivación de los Tweets. Además, siguen destacando los uni-grams ya analizados “aún”, “días” y “día”. Aparece un nuevo concepto en comparación a la data de septiembre que es “subtel_chile” (subsecretaría de telecomunicaciones de Chile). Al igual que “sernac”, es una manera directa de manifestar una queja o reclamo ya que ambas son instituciones reguladoras que deben velar por el correcto cumplimiento en los servicios que ofrecen estas empresas.



**Figura 45: Vista gráfico bi-grams de archivo excel generado por el código Sentiment_Analysis.py para la data de octubre-2018. Fuente: Elaboración Propia.*

En el caso de los bi-grams de octubre, luego de lo analizado para septiembre, podemos destacar: “aún no”, “qué pasa”, “la señal”, “el servicio”, “todos los” y “con la”. Ahora que conocemos mejor el escenario que se está investigando, si unimos algunos de estos bi-grams y completamos frases acordes al contexto podría quedar de la siguiente manera: “qué pasa” “con la” “la señal” y “el servicio” “todos los” días reclamo y “aún no” tengo respuesta.



**Figura 46: Vista gráfico tri-grams de archivo excel generado por el código Sentiment_Analysis.py para la data de octubre-2018. Fuente: Elaboración Propia.*

A continuación se mostrarán algunos de los tri-grams de mayor frecuencia ordenados de forma tal que el mensaje sea aún más claro:

“Qué pasa con” “la señal de” “todos los días” “que no me” “que no tengo” “no tengo señal”. “No me interesa” “no me llamen” “me llamen mas” “dejen de llamarme”. “Aún no tengo” “no tengo internet”.

En base a lo analizado, las principales motivaciones manifestadas durante octubre-2018 son:

- Problemas con el internet.
- Problemas con la señal.
- Problemas e inconvenientes frecuentes (todos los días).

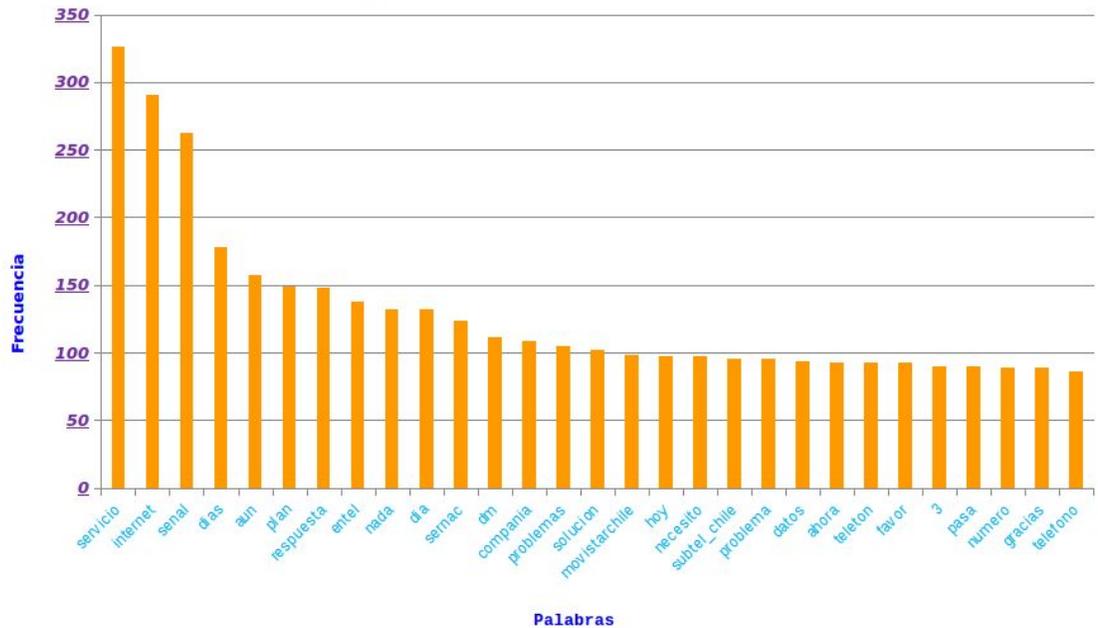
- Reiterado acoso telefónico por parte del Call Center.
- Deficiente capacidad de respuesta de Claro Chile.
- Reclamos directos a instituciones reguladoras (sernac y subtel_chile).

6.1.4.3 Noviembre-2018

	29718		('de', 'clarochile_cl')	160		('la', 'senal', 'de')	44
servicio	326		('en', 'la')	144		('teleton', 'y', 'clarochile_cl')	38
internet	290		('que', 'no')	138		('que', 'pasa', 'con')	38
senal	262		('no', 'me')	137		('todos', 'los', 'dias')	37
dias	178		('y', 'no')	136		('con', 'la', 'senal')	33
aun	157		('que', 'me')	129		('y', 'aun', 'no')	27
plan	149		('de', 'la')	129		('que', 'no', 'me')	24
respuesta	148		('la', 'senal')	129		('el', 'servicio', 'de')	24
entel	137		('a', 'la')	112		('todo', 'el', 'dia')	24
nada	132		('en', 'el')	99		('de', 'teleton', 'y')	23
dia	132		('para', 'que')	98		('a', 'traves', 'de')	23
sernac	123		('el', 'servicio')	94		('servicio', 'de', 'internet')	22
dm	111		('con', 'la')	88		('pasa', 'con', 'la')	22
compania	108		('por', 'favor')	82		('clarochile_cl', 'que', 'pasa')	19
problemas	105		('clarochile_cl', 'hola')	80		('y', 'no', 'me')	19
solucion	102		('y', 'me')	79		('senal', 'de', 'internet')	18
movistarchile	98		('no', 'se')	78		('senal', 'de', 'clarochile_cl')	17
hoy	97		('mas', 'de')	76		('hace', 'mas', 'de')	16
necesito	97		('de', 'internet')	75		('aun', 'no', 'me')	16
subtel_chile	95		('con', 'el')	74		('de', 'forma', 'remota')	14
problema	95		('clarochile_cl', 'que')	71		('en', 'la', 'innovaton')	13
datos	93		('clarochile_cl', 'miclaro_cl')	71		('de', 'clarochile_cl', 'y')	13
ahora	92		('miclaro_cl', 'clarochile_cl')	70		('ya', 'les', 'envie')	13
teleton	92		('servicio', 'de')	70		('les', 'envie', 'un')	13
favor	92		('todos', 'los')	67		('rehabilitacion', 'de', 'foma')	13
3	90		('senal', 'de')	64		('tengo', 'problemas', 'con')	13
pasa	90		('a', 'mi')	63		('desarrollar', 'plataformas', 'que')	12
numero	89		('no', 'tengo')	61		('por', 'teleton', 'y')	12
gracias	89		('que', 'pasa')	60		('con', 'problemas', 'de')	12
telefono	86		('lo', 'mismo')	60		('me', 'dicen', 'que')	12

*Figura 47: Vista tablas n-grams de archivo excel generado por el código Sentiment_Analysis.py para la data de noviembre-2018. Fuente: Elaboración Propia.

Unigrams: Palabra/Frecuencia

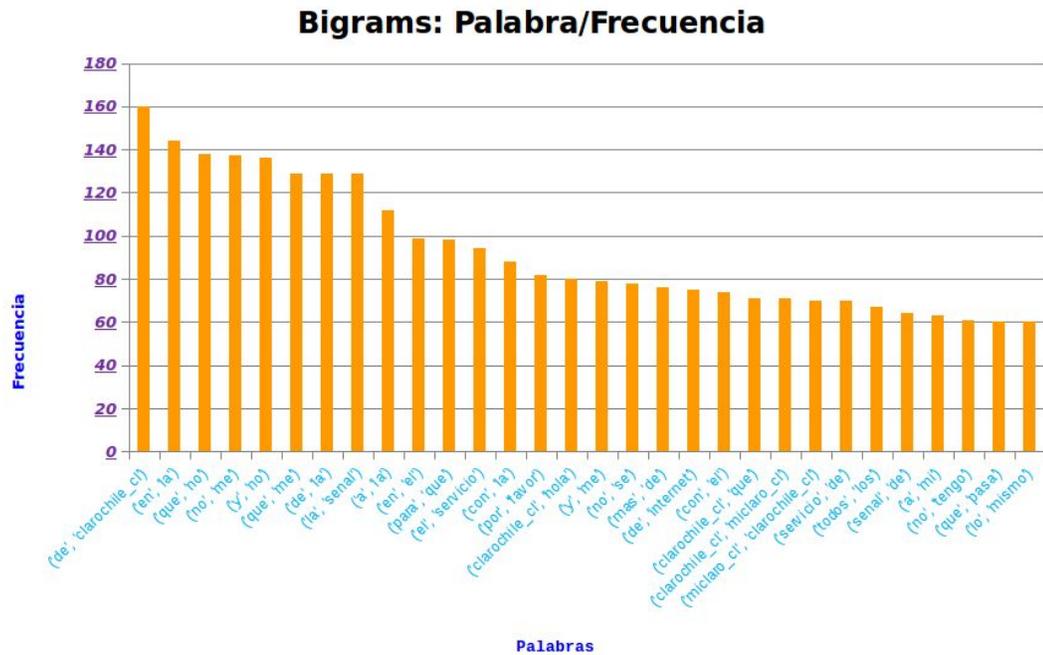


*Figura 48: Vista gráfico uni-grams de archivo excel generado por el código *Sentiment_Analysis.py* para la data de noviembre-2018. Fuente: Elaboración Propia.

Noviembre es el tercer mes de análisis de esta investigación y al parecer los problemas de los clientes son constantes. Por tercer mes consecutivo, “servicio”, “señal” e “internet” son los uni-grams de mayor frecuencia. Así mismo, “sernac” y “subtel_chile” se mantienen presentes en la mayoría de los Tweets.

Es importante destacar que al momento de etiquetar la data de noviembre, la cantidad de Tweets off-topic aumentó considerablemente, en gran parte, por la promoción a la Teletón (fundación de rehabilitación de niños/as y jóvenes con

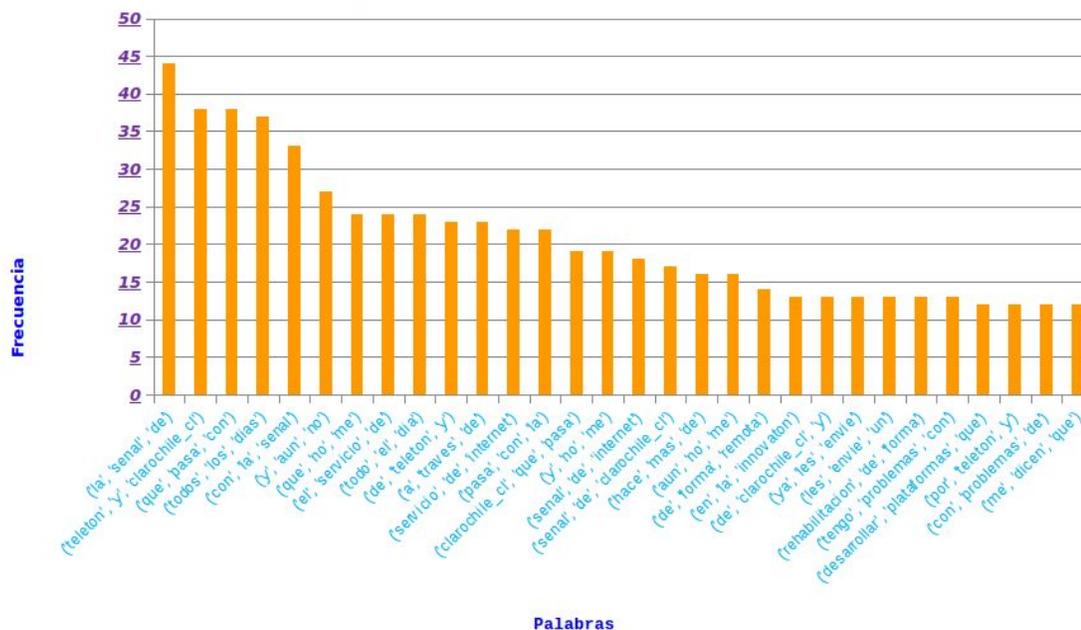
discapacidad) que toma lugar el 30 de noviembre. Esto es evidenciado al encontrar la palabra “teleton” dentro de los uni-grams de mayor frecuencia.



*Figura 49: Vista gráfico bi-grams de archivo excel generado por el código `Sentiment_Analysis.py` para la data de noviembre-2018. Fuente: Elaboración Propia.

Para los bi-grams de este mes, resulta crucial estar inmerso en el contexto, ya que así se pueden unir en base a lo visto en el análisis de los meses anteriores o durante el proceso de etiquetado de la data. Por ejemplo: “qué pasa” “con la” “la señal” “de internet” “no tengo” “todos los” “lo mismo”.

Trigrams: Palabra/Frecuencia



*Figura 50: Vista gráfico tri-grams de archivo excel generado por el código `Sentiment_Analysis.py` para la data de noviembre-2018. Fuente: Elaboración Propia.

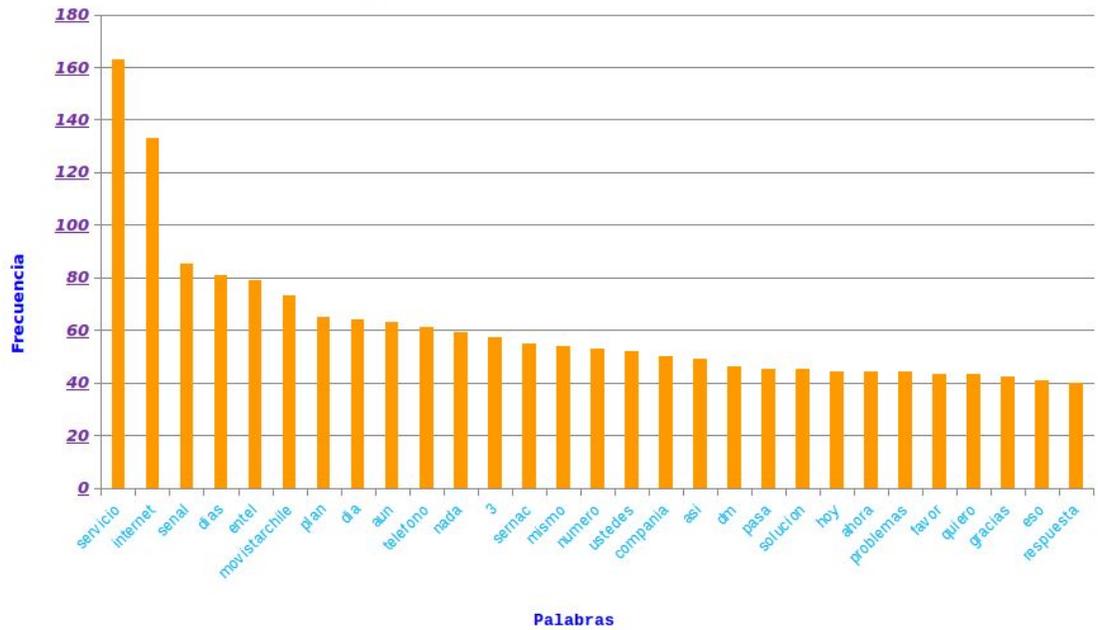
Si bien los tri-grams entregan información de por sí, por el hecho de que con 3 palabras es posible formar una frase acotada, al unir (incluso de manera intuitiva) algunos de los tri-grams de mayor frecuencia es posible obtener un mensaje aún más claro, como se muestra a continuación: “clarochile_cl que pasa” “que pasa con” “la señal de” “el servicio de” “todo el dia” “con problemas de” “ya les envie” “les envie un” “hace más de” “y aún no” “aún no me”.

6.1.4.4 Diciembre-2018

	13692	('que', 'no')	78	('que', 'pasa', 'con')	19
servicio	163	('y', 'no')	74	('el', 'servicio', 'de')	15
internet	133	('de', 'clarochile cl')	71	('clarochile cl', 'que', 'pasa')	15
senal	85	('que', 'me')	71	('y', 'aun', 'no')	14
dias	81	('no', 'me')	62	('servicio', 'de', 'internet')	13
entel	79	('en', 'el')	61	('jueves', '3', 'preventa')	12
movistarchile	73	('el', 'servicio')	55	('que', 'no', 'me')	12
plan	65	('a', 'la')	50	('no', 'me', 'interes')	12
dia	64	('servicio', 'de')	48	('clarochile cl', 'y', 'viernes')	11
aun	63	('de', 'la')	44	('para', 'que', 'me')	11
telefono	61	('en', 'la')	44	('y', 'viernes', '4')	11
nada	59	('para', 'que')	43	('con', 'el', 'internet')	11
3	57	('con', 'el')	40	('3', 'preventa', 'clarochile cl')	11
semac	55	('no', 'se')	38	('preventa', 'clarochile cl', 'y')	11
mismo	54	('clarochile cl', 'que')	38	('me', 'dicen', 'que')	11
numero	53	('y', 'me')	37	('el', 'jueves', '3')	11
ustedes	52	('clarochile cl', 'hola')	37	('todos', 'los', 'dias')	11
compania	50	('todos', 'los')	36	('desde', 'el', 'jueves')	11
asi	49	('la', 'senal')	36	('con', 'la', 'senal')	10
dm	46	('lo', 'mismo')	35	('la', 'senal', 'de')	10
pasa	45	('por', 'favor')	34	('los', 'de', 'clarochile cl')	9
solucion	45	('clarochile cl', 'miclaro cl')	34	('de', 'baja', 'el')	9
hoy	44	('clarochile cl', 'me')	32	('tengo', 'problemas', 'con')	9
ahora	44	('mas', 'de')	32	('todos', 'los', 'meses')	9
problemas	44	('que', 'se')	31	('puedes', 'adquirir', 'en')	9
favor	43	('el', 'internet')	30	('adquirir', 'en', 'puntoticket')	9
quiero	43	('que', 'pasa')	30	('les', 'envie', 'dm')	8
gracias	42	('no', 'tengo')	30	('y', 'no', 'me')	8
eso	41	('lo', 'que')	28	('dar', 'de', 'baja')	8
respuesta	40	('call', 'center')	28	('problemas', 'con', 'el')	8

*Figura 51: Vista tablas n-grams de archivo excel generado por el código `Sentiment_Analysis.py` para la data de diciembre-2018. Fuente: Elaboración Propia.

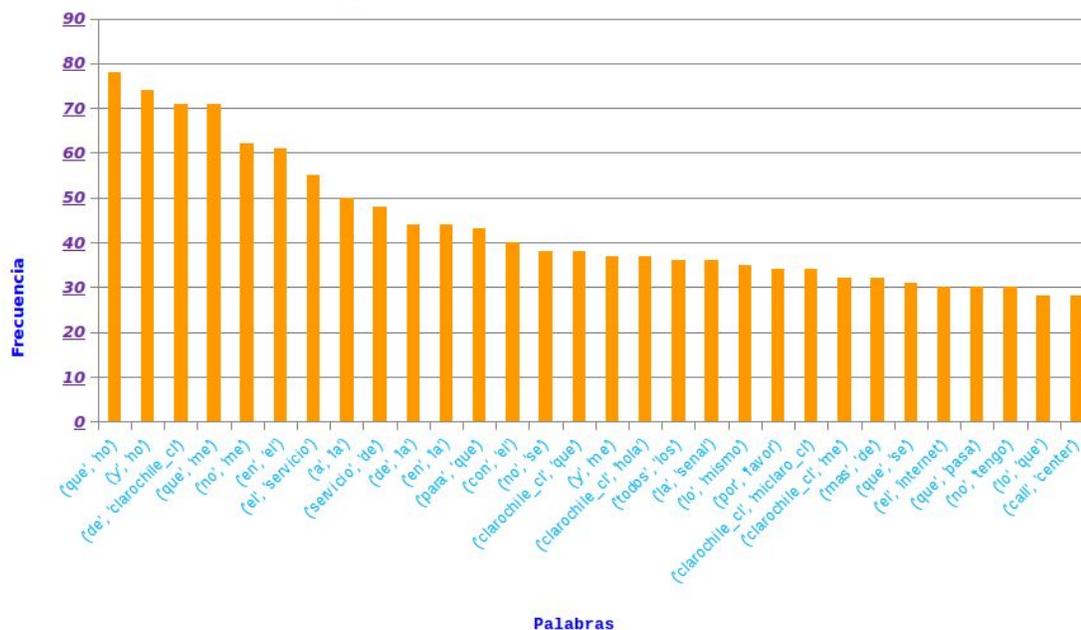
Unigrams: Palabra/Frecuencia



**Figura 52: Vista gráfico uni-grams de archivo excel generado por el código `Sentiment_Analysis.py` para la data de diciembre-2018. Fuente: Elaboración Propia.*

En el cuarto y último mes de estudio para Claro Chile, se comprueba que “senal”, “internet” y “servicio” son una constante al momento de escribir Tweets a Claro. Desaparece “subtel_chile”, pero se mantiene “sernac”, lo que consta la seriedad con la que los clientes publican sus reclamos.

Bigrams: Palabra/Frecuencia



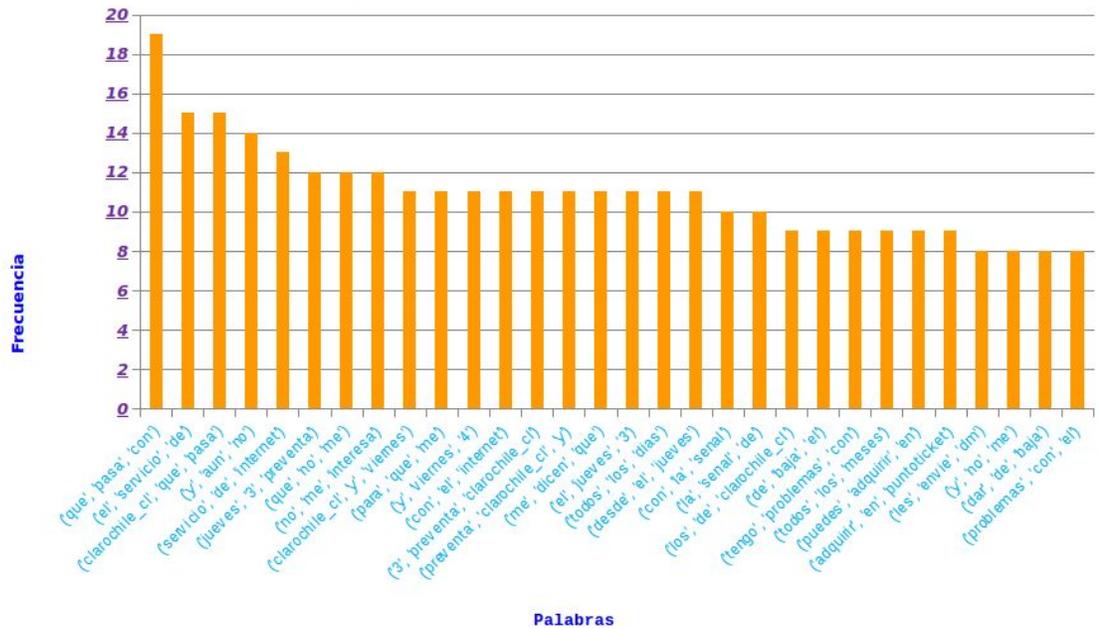
*Figura 53: Vista gráfico bi-grams de archivo excel generado por el código `Sentiment_Analysis.py` para la data de diciembre-2018. Fuente: Elaboración Propia.

Lo primero que resulta importante resaltar de los bi-grams encontrados este mes es la mención explícita de “call center”. Se puede deducir de partida que este mes el acoso y hostigamiento telefónico hacia los clientes no cesó.

Luego, si ordenamos los bi-grams de la forma que lo hicimos los meses pasados, se puede ver que se repite la misma situación:

“Qué pasa” “con el” “no tengo” “el internet” “la señal”.

Trigrams: Palabra/Frecuencia



*Figura 54: Vista gráfico tri-grams de archivo excel generado por el código `Sentiment_Analysis.py` para la data de diciembre-2018. Fuente: Elaboración Propia.

Finalmente, uniendo algunos de los tri-grams de diciembre obtenemos:

“Qué pasa con” “el servicio de” “servicio de internet” “tengo problemas con” “problemas con el” “todos los días” “todos los meses”. “Dar de baja” “de baja el” “servicio de internet”.

Finalmente, los principales motivos de los Tweets de diciembre para Claro Chile son:

- Problemas con los servicios.
- Problemas con el internet.

- Deficiente capacidad de respuesta por parte de Claro Chile.
- Solicitar dar de baja servicios.
- Acoso telefónico por parte del call center.
- Denuncias al sernac.

6.2 Pruebas y Resultados para Enel Chile

6.2.1 Etiquetado para Enel Chile

Para el etiquetado, se desarrolló un programa en el lenguaje Python para hacer de esta una tarea más fácil y rápida. El algoritmo recorre los Twitts previamente señalados uno por uno, lo despliega en pantalla y consulta al usuario por la etiqueta que corresponde, ofreciendo 3 opciones: Negativo, Positivo/Neutro e Informativo, donde:

- Negativo: Twitts que explícitamente hacen una crítica o denuncia a Enel y la calidad de servicio percibida.
- Positivo/Neutro: Twitts que expresan comentarios positivos o neutrales hacia Enel. Como por ejemplo “muchas gracias por la rápida respuesta” o “ya envié el dm”. Se consideraron como neutrales también los Twitts que informan la reposición del suministro, esto ya que se puso énfasis en separar los Twitts que informan corte de suministro de los que informan su reposición.
- Informativo: Twitts que solicitan o entregan información sobre cortes de luz y/o suministro o situaciones relacionadas a los postes de tendido

eléctrico, árboles, medidores, etcétera. En resumen, lo relacionado a una alteración de la continuidad del servicio.

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/WebSc
Línea: 1295 comentario: enelclientescl que se han imaginado cortan la luz y luego la dan e inmedi
atamente la cortan nuevamente ustedes van a correr con los gastos de los artefactos que se puedan que
mar pesima su gestion
Ingrese (1) si es NEGATIVO, (0) si es POSITIVO/NEUTRO, (2) si es INFORMATIVO
█
```

**Figura 55: Ejemplo de Twitt real etiquetado como Negativo. Fuente: Elaboración Propia.*

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/WebSc
Línea: 1310 comentario: valle grande lampa sin luz nuevamente enelclientescl me impresiona como
puede llegar a ser tan habitual su mal servicio comlnhsqrnfd
Ingrese (1) si es NEGATIVO, (0) si es POSITIVO/NEUTRO, (2) si es INFORMATIVO
█
```

**Figura 56: Ejemplo de Twitt real etiquetado como Negativo. Fuente: Elaboración Propia.*

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/WebSc
Línea: 1626 comentario: estimados ya se soluciono el problema gracias por la pronta respuesta era un
corte interno
Ingrese (1) si es NEGATIVO, (0) de lo contrario
█
```

**Figura 57: Ejemplo de Twitt real etiquetado como Positivo/Neutral. Fuente: Elaboración Propia.*

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/WebSc
Línea: 1258 comentario: enelclientescl estimados estamos sin luz en colina sector san jose las enc
inas al puente
Ingrese (1) si es NEGATIVO, (0) si es POSITIVO/NEUTRO, (2) si es INFORMATIVO
█
```

**Figura 58: Ejemplo de Twitt real etiquetado como Informativo. Fuente: Elaboración Propia.*

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/WebSc
Línea: 1313 comentario: enelclientescl estimados hay un corte de luz en quillicura en sector santa l
uisa con las torres alguna informacion gracias alertaquillicura muniquillicura
Ingrese (1) si es NEGATIVO, (0) si es POSITIVO/NEUTRO, (2) si es INFORMATIVO
█
```

**Figura 59: Ejemplo de Twitt real etiquetado como Informativo. Fuente: Elaboración Propia.*

Cabe señalar que en un principio se comenzó a etiquetar la data con las mismas 2 etiquetas utilizadas sobre la data de Claro Chile, pero luego de haber etiquetado alrededor de 1.000 Twitts, se concluyó que para este contexto 2 etiquetas no alcanzaban a representar bien la polaridad o categoría de los Twitts, lo que generaría confusión para el proceso de aprendizaje y llevaría a conclusiones erróneas en los resultados. Esto se debe a que muchos Twitts (según lo etiquetado en primera instancia, la mayoría) estaban siendo etiquetados como negativos cuando sólo informaban sobre cortes en distintos sectores y, luego de un análisis, se consideró necesario separar esos Twitts de los que realmente hacen alusión a la calidad del servicio entregado. Esto ya que es importante señalar que por el tipo de servicio que entrega Enel Chile, su servicio puede verse afectado por factores externos y de los cuales no tienen control (como desastres naturales o intervenciones de terceros), por lo que se debe poner más atención en cómo enfrentan estas situaciones y cómo percibe la gente su capacidad de reacción y resolución de problemas, además de la atención al cliente.

Por otro lado, descargando la data y excluyendo los Twitts provenientes de la misma empresa, para el mes de noviembre se descargaron 7.295 Twitts, cantidad mucho mayor a la obtenida para Claro Chile en el mismo mes. Por esto, se consideró que de contar con el tiempo, se tendría la cantidad de data suficiente para que resulte representativa para cada clase entrenando sólo con la data de dicho mes (o parte de ella).

De los 7.295 Twitts, 3.711 (50.88%) fueron descartados por venir de parte de cuentas de la misma empresa Enel Chile. Luego, de los 1.843 revisados para la labor de etiquetado, 1.210 (65.65%) fueron efectivamente etiquetados y considerados para la labor de entrenamiento. Dentro de los 633 (34.35%)

omitidos, se encuentran Twitts off-topic como conversaciones entre usuarios y Twitts iguales o de mucha similaridad. La omisión de dichos Twitts es debido que a veces, al ocurrir un corte o problemas de suministros, mucha gente del mismo sector escribía para comentar el mismo problema y con prácticamente las mismas palabras. Esto provocaría overfitting y podría asociar la clase del Twitt a la localidad o sector del incidente en lugar de lo realmente acontecido.

De los 1.210 Twitts etiquetados, 162 (13.38%) se etiquetaron positivos (o neutrales), 178 (14.71%) negativos y 869 (71.91%) informativos.

6.2.2 Entrenamiento y evaluación de modelos predictivos para Enel Chile

Respecto al código y desarrollo del proceso de aprendizaje supervisado para Enel Chile, se utilizaron los mismos procedimientos, bibliotecas y modelos clasificadores. Fuera de eso, vale la pena destacar:

- Para simplificar notaciones y aplicación de métricas, así como la comprensión de los resultados, se expresó como “positiva” la clase que representa Twitts positivos y neutrales, “negativa” los Twitts que hacen alusión directa a una crítica a la empresa e “informativo” los Twitts que solicitan o entregan información sobre hechos relacionados a alteraciones del servicio.
- Para cada modelo se realizaron 7 iteraciones, en todas ellas se utilizaron los parámetros por defecto. Las figuras adjuntas para cada modelo representan el peor y mejor desempeño obtenido respectivamente. Para tomar esta decisión se basó en las métricas Accuracy y MCC principalmente, dando mayor importancia al MCC por ser una métrica

balanceada. Es importante destacar que para este caso no se utilizó la métrica ROC AUC SCORE, esto debido a que es una métrica exclusiva de clasificación binaria y en este caso se utilizaron 3 clases.

- Para escoger el mejor modelo clasificador, se utilizó como punto de comparación la iteración de mejor desempeño de cada modelo. Esto quiere decir, que al momento de comparar, se escogió el mejor desempeño obtenido tanto para SVM como para DecisionTreeClassifier, LogisticRegression y MultinomialNB.
- Es posible identificar a simple vista que para todos los casos, el desempeño sobre la clase informativa fue mucho mejor que la positiva y negativa. Los posibles motivos de esto se analizarán más adelante.

6.2.2.1 MultinomialNB

```

115 -----
116 Iteracion: 7
117 Matriz de Confusion:
118 [[ 13  2 17]
119 [  1 23 20]
120 [  1  3 162]]
121 Accuracy: 0.8181818181818182
122 Matthews Correlation Coefficient: 0.5868866290539688
123 Classification Report:
124                precision    recall  f1-score   support
125
126  Clase Positiva         0.87      0.41      0.55         32
127  Clase Negativa         0.82      0.52      0.64         44
128  Clase Informativa      0.81      0.98      0.89        166
129
130      micro avg           0.82      0.82      0.82        242
131      macro avg           0.83      0.63      0.69        242
132      weighted avg        0.82      0.82      0.80        242

```

**Figura 60: Resultado desempeño clasificador MultinomialNB sobre data clasificada, iteración 7. Fuente: Elaboración Propia.*

```

1 -----
2 Iteracion: 1
3 Matriz de Confusion:
4 [[ 20  2 17]
5 [  1 21 11]
6 [  1  4 165]]
7 Accuracy: 0.8512396694214877
8 Matthews Correlation Coefficient: 0.6557787073740503
9 Classification Report:
10                precision    recall  f1-score   support
11
12  Clase Positiva      0.91      0.51      0.66        39
13  Clase Negativa      0.78      0.64      0.70        33
14  Clase Informativa   0.85      0.97      0.91       170
15
16      micro avg        0.85      0.85      0.85       242
17      macro avg        0.85      0.71      0.75       242
18      weighted avg     0.85      0.85      0.84       242
19
20 -----

```

**Figura 61: Resultado desempeño clasificador MultinomialNB sobre data clasificada, iteración 1. Fuente: Elaboración Propia.*

6.2.2.2 DecisionTreeClassifier

```

115 -----
116 Iteracion: 7
117 Matriz de Confusion:
118 [[ 21  0  6]
119 [  5  7 16]
120 [  8 15 164]]
121 Accuracy: 0.7933884297520661
122 Matthews Correlation Coefficient: 0.45702119761376725
123 Classification Report:
124                precision    recall  f1-score   support
125
126  Clase Positiva      0.62      0.78      0.69        27
127  Clase Negativa      0.32      0.25      0.28        28
128  Clase Informativa   0.88      0.88      0.88       187
129
130      micro avg        0.79      0.79      0.79       242
131      macro avg        0.61      0.63      0.62       242
132      weighted avg     0.79      0.79      0.79       242

```

**Figura 62: Resultado desempeño clasificador DecisionTreeClassifier sobre data clasificada, iteración 7. Fuente: Elaboración Propia.*

```

20 -----
21 Iteracion: 2
22 Matriz de Confusion:
23 [[ 32  0  5]
24 [  4 10 19]
25 [  4 15 153]]
26 Accuracy: 0.8057851239669421
27 Matthews Correlation Coefficient: 0.5607086897535872
28 Classification Report:
29           precision    recall  f1-score   support
30
31  Clase Positiva         0.80     0.86     0.83         37
32  Clase Negativa         0.40     0.30     0.34         33
33  Clase Informativa       0.86     0.89     0.88        172
34
35      micro avg           0.81     0.81     0.81        242
36      macro avg           0.69     0.69     0.68        242
37      weighted avg        0.79     0.81     0.80        242
38
39 -----

```

**Figura 63: Resultado desempeño clasificador DecisionTreeClassifier sobre data clasificada, iteración 2.*

Fuente: Elaboración Propia.

6.2.2.3 SVM

```

20 -----
21 Iteracion: 2
22 Matriz de Confusion:
23 [[ 20  3  7]
24 [  3 14 16]
25 [  7  2 170]]
26 Accuracy: 0.8429752066115702
27 Matthews Correlation Coefficient: 0.5993368606492797
28 Classification Report:
29           precision    recall  f1-score   support
30
31  Clase Positiva         0.67     0.67     0.67         30
32  Clase Negativa         0.74     0.42     0.54         33
33  Clase Informativa       0.88     0.95     0.91        179
34
35      micro avg           0.84     0.84     0.84        242
36      macro avg           0.76     0.68     0.71        242
37      weighted avg        0.83     0.84     0.83        242
38
39 -----

```

**Figura 64: Resultado desempeño clasificador SVM sobre data clasificada, iteración 2. Fuente:*

Elaboración Propia.

```

115 -----
116 Iteracion: 7
117 Matriz de Confusion:
118 [[ 29  1  2]
119 [  3 17 13]
120 [  2  7 168]]
121 Accuracy: 0.8842975206611571
122 Matthews Correlation Coefficient: 0.7227559183449921
123 Classification Report:
124                precision    recall  f1-score   support
125
126  Clase Positiva      0.85      0.91      0.88        32
127  Clase Negativa      0.68      0.52      0.59        33
128  Clase Informativa  0.92      0.95      0.93       177
129
130      micro avg      0.88      0.88      0.88       242
131      macro avg      0.82      0.79      0.80       242
132      weighted avg   0.88      0.88      0.88       242

```

**Figura 65: Resultado desempeño clasificador SVM sobre data clasificada, iteración 7. Fuente: Elaboración Propia.*

6.2.2.4 LogisticRegression

```

39 -----
40 Iteracion: 3
41 Matriz de Confusion:
42 [[ 20  3  6]
43 [  3 11 16]
44 [  4  5 174]]
45 Accuracy: 0.8471074380165289
46 Matthews Correlation Coefficient: 0.587493403841766
47 Classification Report:
48                precision    recall  f1-score   support
49
50  Clase Positiva      0.74      0.69      0.71        29
51  Clase Negativa      0.58      0.37      0.45        30
52  Clase Informativa  0.89      0.95      0.92       183
53
54      micro avg      0.85      0.85      0.85       242
55      macro avg      0.74      0.67      0.69       242
56      weighted avg   0.83      0.85      0.84       242
57
58 -----

```

**Figura 66: Resultado desempeño clasificador LogisticRegression sobre data clasificada, iteración 3. Fuente: Elaboración Propia.*

```

1 -----
2 Iteracion: 1
3 Matriz de Confusion:
4 [[ 29  0  7]
5 [  2 21 20]
6 [  0  0 163]]
7 Accuracy: 0.8801652892561983
8 Matthews Correlation Coefficient: 0.7528094372479645
9 Classification Report:
10                precision    recall  f1-score   support
11
12  Clase Positiva      0.94      0.81      0.87         36
13  Clase Negativa      1.00      0.49      0.66         43
14  Clase Informativa   0.86      1.00      0.92        163
15
16      micro avg       0.88      0.88      0.88        242
17      macro avg       0.93      0.76      0.82        242
18      weighted avg    0.89      0.88      0.87        242
19
20 -----

```

**Figura 67: Resultado desempeño clasificador LogisticRegression sobre data clasificada, iteración 1.*

Fuente: Elaboración Propia.

6.2.2.5 Análisis de modelos

- MultinomialNB:

Este modelo presenta su mejor desempeño sobre la clase informativa. De los 170 Twitts de dicha clase, etiquetó 165 de manera correcta y sólo 5 (2.94%) de manera incorrecta. De esos 5, 4 errores fueron ocasionados por confundir un Twitt que es informativo con uno negativo. Esto puede deberse a que algunos de los Twitts previamente etiquetados como negativos, partían informando el problema (como un Twitt informativo), y al final realizaban la crítica a la compañía. Por otro lado, el buen desempeño sobre la clase puede atribuirse a que era la clase mejor representada en los Twitts de entrenamiento (con mayor soporte), por lo que tenía más información para aprender a identificar la clase.

En cuanto al desempeño sobre la clase negativa, de los 33 ejemplares, 12 (36.36%) fueron etiquetados erróneamente, siendo 11 de esos erróneamente clasificados confundidos con los Twitts informativos. Nuevamente, se asume que la confusión entre ambas clases se produce porque la crítica suele ir acompañada de información asociada a un problema con el suministro, lo que confunde al clasificador.

Para la clase positiva, de los 39 ejemplares, 19 (48.71%) fueron etiquetados de manera errónea, siendo la clase con peor desempeño de este modelo. Sólo 2 ejemplares fueron confundidos con la clase negativa, un bajo valor que era de esperarse dado que al ser clases “opuestas” se espera que el contenido de los Twitts sea en su mayoría muy disímil. Los otros 17 ejemplares fueron confundidos con la clase informativa. Esto puede deberse a dos factores: el primero, que algunos Twitts que informaban problemas comentaban que la luz ya había vuelto pero que se volvía a ir, y los Twitts positivos también informaban que ya había regresado el suministro junto con un mensaje de agradecimiento; el segundo, la cantidad de ejemplares para cada una de las clases, dado que los ejemplares de la clase positiva son muy pocos en comparación a los de la clase informativa, lo que habría llevado al clasificador a preferir, ante la duda, la etiqueta informativa por sobre la positiva.

- DecisionTreeClassifier:

Para la clase informativa, de los 172 ejemplares, 19 (11.04%) fueron etiquetados de manera errónea, 15 de estos etiquetados como negativos. Los posibles motivos de confusión entre estas clases fueron discutidos en el modelo anterior.

En cuanto al desempeño sobre la clase negativa, de los 33 ejemplares, 23 (69.69%) fueron etiquetados de manera errónea, siendo la clase con el peor desempeño de este modelo. Nuevamente, la mayor confusión se produjo con la clase informativa.

Finalmente, en la clase positiva, de los 37 ejemplares, sólo 5 (13.51%) fueron etiquetados de manera errónea, todos ellos a causa de confusión con la clase informativa, lo que habla muy bien de cómo el modelo diferenció la clase positiva de la negativa.

- SVM:

Para la clase informativa, de los 177 ejemplares, 9 (5.08%) fueron etiquetados de manera errónea, siendo la mayor confusión producida con la clase negativa. De los 33 ejemplares de la clase negativa, 16 (48.48%) fueron etiquetados de manera errónea, de los cuales 13 se confundieron con la clase informativa. Esta es la clase con peor desempeño de este modelo, mostrando que el modelo tiene poca capacidad de identificar un Twitt negativo como tal, prefiriendo etiquetarlo como la clase mayoritaria (informativa).

Finalmente para la clase positiva, de los 32 ejemplares realmente positivos, 3 (9.37%) fueron etiquetados de manera errónea, de los cuales 2 fueron confundidos con la clase informativa.

- LogisticRegression:

Para la clase informativa, de los 163 ejemplares realmente informativos, el 100% fue etiquetado de manera correcta, lo que no sólo la hace la clase de

mejor desempeño del modelo, sino que es el modelo con mejor desempeño sobre la clase de todos los modelos analizados. Esto quiere decir que este modelo fue capaz de identificar un Twitt informativo siempre que se enfrentó a uno, lo que no significa que no pueda identificar un Twitt no informativo como tal (lo que se ve reflejado en la *precision* de la clase).

En el caso de la clase negativa, de las 43 muestras, 22 (51.16%) fueron clasificadas de manera errónea, de los cuales 20 fueron confundidos con la clase informativa y sólo 2 con la positiva (o neutral). Esto lo hace el tercer clasificador de cuatro clasificadores en los que la clase negativa es la con peor desempeño. Este mal desempeño sobre la clase que se repite a lo largo de 3 clasificadores puede deberse principalmente a dos motivos: primero, que los ejemplares representantes de la clase eran muy pocos, por lo que los modelos no tenían suficiente información para aprender de manera correcta y; segundo, que, como ya se mencionó, la diferencia entre un Twitt etiquetado como negativo y uno etiquetado como informativo es de muy pocas palabras, dado que en ambos se puede entregar información y lo único que los diferencia son unas pocas palabras hacia la empresa como por ejemplo: “pésimo servicio”, “muy mala atención al cliente”, “sinverguenzas”, etcétera.

Respecto a la clase positiva, de los 36 ejemplares realmente positivos, 7 (19.44%) fueron etiquetados erróneamente, todos ellos confundidos con la clase informativa.

6.2.2.6 Comparación de modelos

	LogisticRegression (1)	SVM (7)	DecisionTreeClassifier (2)	MultinomialNB (1)
Accuracy	0,880	0,884	0,805	0,851
MCC	0,752	0,722	0,56	0,655

*Tabla 8: Tabla comparativa de desempeño de modelos. Fuente: Elaboración Propia.

Como muestra la tabla, DecisionTreeClassifier obtuvo los peores resultados tanto para el Accuracy como para el MCC, por lo que fue descartado inmediatamente. Con esto, DecisionTreeClassifier se confirma como el peor clasificador para el contexto de este trabajo de investigación para ambas empresas. Esto se adjudica a una importante desventaja que caracteriza a este clasificador y que ya fue mencionado en el marco teórico que es que se caracterizan en presentar un desempeño deficiente en casos donde las clases no están balanceadas. En el caso de Enel Chile, al momento de etiquetar, la clase informativa representó casi un 72% del total etiquetado, versus las otras dos clases (positiva y negativa) que no alcanzaban el 15% cada una.

Nuevamente resulta una sorpresa que MultinomialNB no haya sobresalido en ninguna de las métricas comparadas ni en su mejor iteración. Los posibles motivos de esto fueron analizados en la sección de Claro Chile.

Al igual que para el caso de Claro Chile, SVM y LogisticRegression fueron los modelos con mejor desempeño. Si bien SVM obtuvo el mayor Accuracy de todos los modelos, se descartó la posibilidad de ser utilizado sobre la data no etiquetada por dos principales motivos:

- Si bien obtuvo el mayor Accuracy, en esta métrica lo que separa a SVM de LogisticRegression es de sólo 0,004.
- Como se mencionó anteriormente, se privilegió la métrica MCC por sobre Accuracy y cualquier otra ya que esta es una métrica balanceada. Mientras Accuracy mide las predicciones correctas sobre el total, MCC incluye en su cálculo las predicciones erróneas, por lo que, un mayor MCC indica también que las predicciones incorrectas fueron menores.

6.2.3 Modelo predictivo LogisticRegression en data no etiquetada para Enel Chile

A continuación se muestra el resultado de la clasificación del modelo LogisticRegression (1) entrenado con parte de la data de noviembre probado sobre la data no etiquetada (septiembre, octubre y diciembre).

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitler/Codigo Descarga KC (este)/WebSc
-----
gracias x la respuesta
Positivo/Neutro
-----
lo sepero la baja de voltaje no se puede reportar por mensaje interno
Informativo
-----
enelclientescl tengo problemas de voltajevitacura vespuciosolo prenden algunas ampolletas en forma ten
uedesde las 0630 am
Informativo
-----
enelclientescl los valores de tension de la red estan muy altos en las condes
Informativo
-----
enelclientescl el transformador que esta afuera de mi casa ha sonado todo el dia
Informativo
-----
nada en mi bandeja
Informativo
-----
enelclientescl alguna interrupcion de servicios en la comuna de las condes
Informativo
-----
que dice enelclientescl al respecto
Informativo
Total Positivos: 793 Total Negativos: 483 Total Informativos: 2899
tomas@tomas-OMEN-by-HP-Laptop:~/Escritorio/Universidad/Tesis/Codigo Twitler/Codigo Descarga KC (este)/
WebScraping/Etiquetado$
```

**Figura 68: Prueba de modelo LogisticRegression (1) entrenado con la data etiquetada de Noviembre-2018, evaluando la data de Septiembre-2018. Fuente: Elaboración Propia.*

Como se aprecia en la figura, de los 4.175 Twitts descargados de septiembre, 2.899 (69.43%) fueron clasificados como informativos, 483 (11.56%) como negativos y 793 (19.01%) como positivos/neutrales.

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/WebSc
eneelclientescl estimados no hay luz en el pasaje san patricio quien es encargado de ello

Informativo
-----
alguna respuesta a mi inquietud de la baja de voltaje

Informativo
-----
ok los sigo

Positivo/Neutro
-----
saludos

Positivo/Neutro
-----
eneelclientescl buenas noches en pastora rosa astete 600 quilicura hay casas que tienen luz y otras que
no ademas de tener bajas de voltaje que es lo que pasa

Informativo
-----
eneelclientescl sin luz en los juncos con los copihues la comuna de lo prado

Informativo
-----
eneelclientescl restimados yo si me encuentro con suministro electrico lo que me preocupa es la carta q
ue me llego que nnoo han tenido acceso a mi propiedad abrazo

Informativo
-----
eneelclientescl villa los maitenes en quilicura con poco voltaje saben que esta pasando

Informativo
Total Positivos: 570 Total Negativos: 569 Total Informativos: 2209
tomas@tomas-OMEN-by-HP-Laptop:~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/
WebScraping/Etiquetado$
```

**Figura 69: Prueba de modelo LogisticRegression (1) entrenado con la data etiquetada de Noviembre-2018, evaluando la data de Octubre-2018. Fuente: Elaboración Propia.*

Para el caso de octubre, del total de 3.348 Twitts descargados, 2.209 (65.97%) fueron clasificados como informativos, 569 (16.99%) como negativos y 570 (17.04%) como positivos/neutrales.

```
tomas@tomas-OMEN-by-HP-Laptop: ~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/WebSc
respon den siempre la misma estupidez los problemas de enel son frecuentes espero que le llegue una bue
na multa enelclientescl susanajimenezs sec_cl sebastianpinera lavinjoaquin

Negativo!
-----
acabo de ver su tuits alcalde ya lo comente y le dije gracias por dar mas info que enelclientescl

Positivo/Neutro
-----
sin luz en alonso de camargo y sebastian elcano nuevamente en menos de un mes otro corte

Informativo
-----
gracias por la info alcalde ni enelclientescl fue capaz de informar que habia pasado

Positivo/Neutro
-----
ok

Positivo/Neutro
-----
esta en mis tuits no hay hora de reposicion todavia

Informativo
-----
ya les envie la info que me solicitaron enelclientescl enel_chile lavinjoaquin ya vamos 1 hra con laluz
cortada sin info sobre que paso ni cuando se solucionara

Negativo!
-----
mis abuelos dicen que se cayo un arbol sobre los cables de luz y sobre un autoque estan tratando de sa
car a las personas del auto es asi

Informativo
-----
a que hora estiman la reposicion del servicio

Informativo
-----
ahora si

Informativo
-----
cuando devuelta la luz en sector colonel cano todoloquequeriaparamidomingo

Informativo
-----
gracias por responder

Positivo/Neutro
-----
por favor a que hora vuelve seb elcano con colon

Informativo
Total Positivos: 1058 Total Negativos: 829 Total Informativos: 4853
tomas@tomas-OMEN-by-HP-Laptop:~/Escritorio/Universidad/Tesis/Codigo Twitter/Codigo Descarga KC (este)/
WebScraping/Etiquetado$
```

**Figura 70: Prueba de modelo LogisticRegression (1) entrenado con la data etiquetada de Noviembre-2018, evaluando la data de Diciembre-2018. Fuente: Elaboración Propia.*

Finalmente, para la data de diciembre, de los 6.740 Twitts descargados, 4.853 (72%) fueron clasificados como informativos, 829 (12.29%) como negativos y 1.058 (15.71%) como positivos/neutrales.

A modo de concluir de manera general según lo etiquetado por el modelo LogisticRegression (1) que obtuvo 0.880 Accuracy y 0.752 MCC en su fase de entrenamiento/validación:

- Octubre fue el mes con mayor porcentaje de Twitts negativos.
- En todos los meses, el porcentaje de Twitts positivos/neutrales era mayor que el de los Twitts negativos.
- La clase informativa fue predominante en todos los meses estudiados, su soporte fluctuó entre el 69.43% y el 72% del total de Twitts de cada mes.

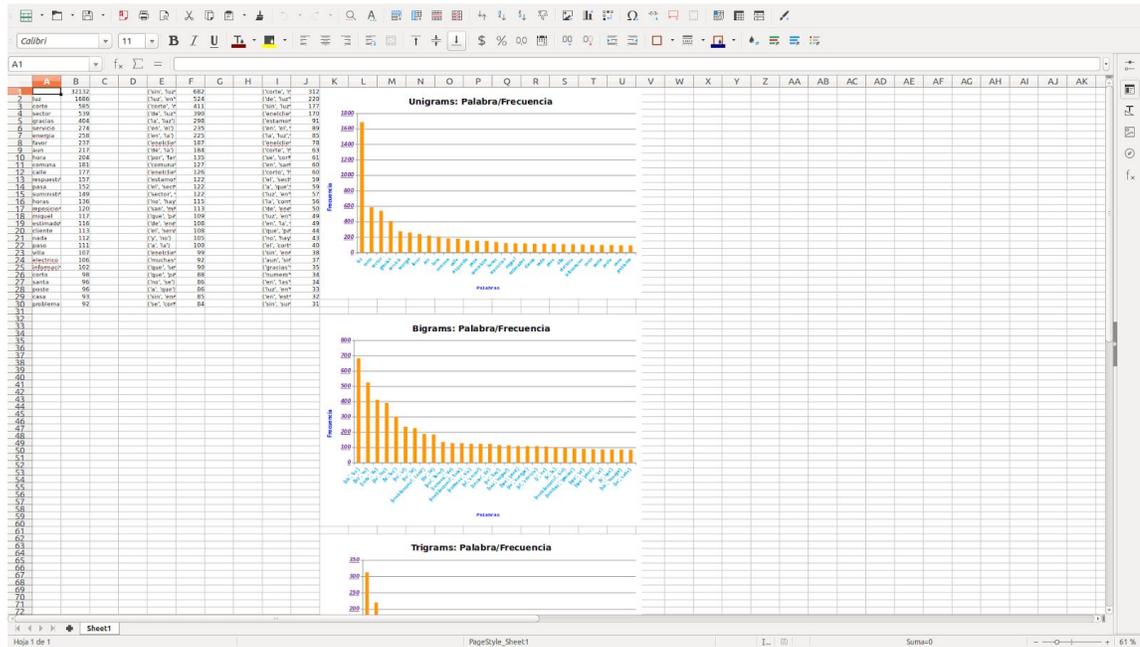
6.2.4 Análisis de Sentimientos

En el análisis de sentimientos para Enel Chile se utilizó el mismo procedimiento que para Claro Chile. La data analizada comprende lo publicado entre septiembre y diciembre del 2018, esto es, la data utilizada para entrenamiento y la data utilizada para probar el modelo clasificador.

Notar además que el análisis realizado es progresivo, por lo que se intentará evitar la redundancia de análisis mes a mes a menos de ser estrictamente necesario ya que se espera identificar tantos factores que motivan los Tweets como sea posible.

A continuación se muestran los resultados generados de manera automática en gráficos excel:

6.2.4.1 Septiembre-2018

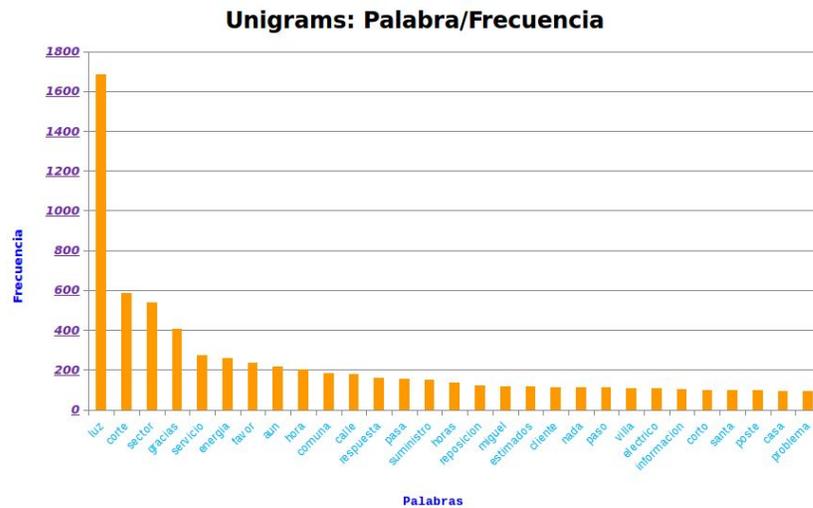


*Figura 71: Vista general de archivo excel generado por el código `Sentiment_Analysis_enel.py` para la data de septiembre-2018. Fuente: Elaboración Propia.

Como se aprecia en la figura, la salida del código es un archivo excel que contiene una selección de los “X” n-grams con mayor frecuencia, con su respectivo valor y gráfico representativo. Para el ejemplo de la figura, se escogieron los 30 uni, bi y tri-grams de mayor frecuencia. El análisis de todos los n-grams es enfocado en los comentarios negativos dado que se busca encontrar las principales falencias de la empresa.

	32132	('sin', 'luz')	682	('corte', 'de', 'luz')	312
luz	1686	('luz', 'en')	524	('de', 'luz', 'en')	220
corte	585	('corte', 'de')	411	('sin', 'luz', 'en')	177
sector	539	('de', 'luz')	390	('enelclientescl', 'corte', 'de')	170
gracias	404	('la', 'luz')	298	('estamos', 'sin', 'luz')	91
servicio	274	('en', 'el')	235	('en', 'el', 'sector')	89
energía	258	('en', 'la')	225	('la', 'luz', 'en')	85
favor	237	('enelclientescl', 'corte')	187	('enelclientescl', 'sin', 'luz')	78
aun	217	('de', 'la')	184	('corte', 'de', 'energía')	63
hora	204	('por', 'favor')	135	('se', 'corto', 'la')	61
comuna	181	('comuna', 'de')	127	('en', 'san', 'miguel')	60
calle	177	('enelclientescl', 'hola')	126	('corto', 'la', 'luz')	60
respuesta	157	('estamos', 'sin')	122	('el', 'sector', 'de')	59
pasa	152	('el', 'sector')	122	('a', 'que', 'hora')	59
suministro	149	('sector', 'de')	122	('luz', 'en', 'la')	57
horas	136	('no', 'hay')	115	('la', 'comuna', 'de')	56
reposicion	120	('san', 'miguel')	113	('de', 'energía', 'en')	50
miguel	117	('que', 'pasa')	109	('luz', 'en', 'el')	49
estimados	116	('de', 'energía')	108	('en', 'la', 'comuna')	49
cliente	113	('el', 'servicio')	108	('que', 'pasa', 'con')	44
nada	112	('y', 'no')	105	('no', 'hay', 'luz')	43
paso	111	('a', 'la')	100	('el', 'corte', 'de')	40
villa	107	('enelclientescl', 'sin')	99	('sin', 'energía', 'eléctrica')	38
eléctrico	106	('muchas', 'gracias')	92	('aun', 'sin', 'luz')	37
informacion	102	('que', 'se')	90	('gracias', 'por', 'la')	35
corto	98	('que', 'paso')	88	('numero', 'de', 'cliente')	34
santa	96	('no', 'se')	86	('en', 'las', 'condes')	34
poste	96	('a', 'que')	86	('luz', 'en', 'san')	33
casa	93	('sin', 'energía')	85	('en', 'estacion', 'central')	32
problema	92	('se', 'corto')	84	('sin', 'suministro', 'eléctrico')	31

*Figura 72: Vista tablas n-grams de archivo excel generado por el código `Sentiment_Analysis_enel.py` para la data de septiembre-2018. Fuente: Elaboración Propia.

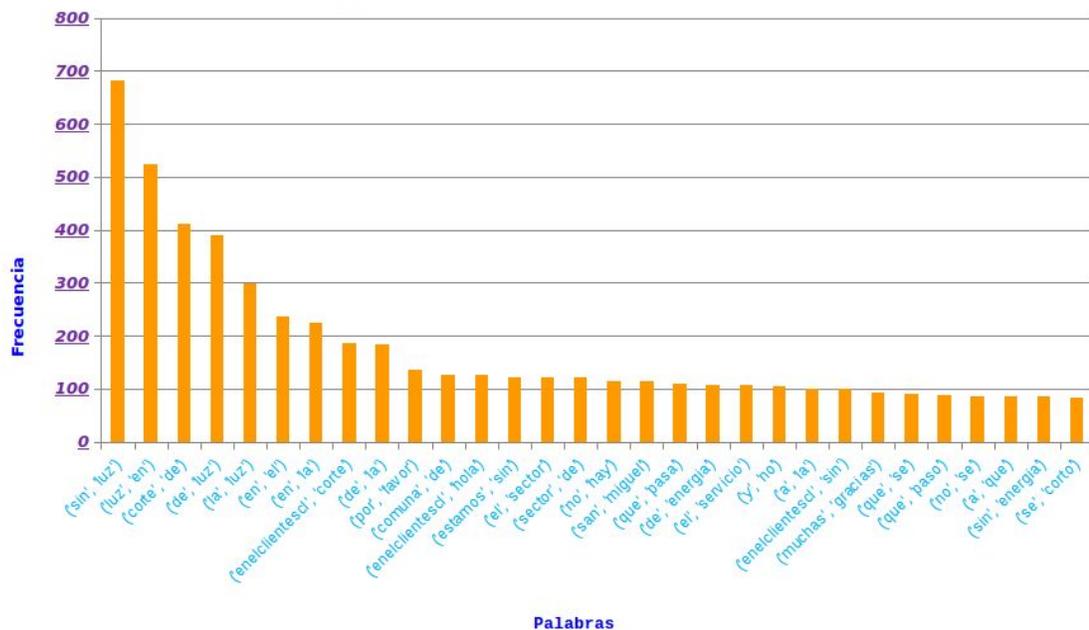


*Figura 73: Vista gráfico uni-grams de archivo excel generado por el código `Sentiment_Analysis_enel.py` para la data de septiembre-2018. Fuente: Elaboración Propia.

Si se analizan los uni-grams en busca de comentarios negativos o críticas a la empresa a analizar, es importante poner énfasis en los uni-grams “aún”,

“horas”, “respuesta”, “nada”, “cuánto” y “reposición”. En base a esto, se pueden sacar conclusiones preliminares sobre los principales motivos de los Twitts negativos. Las referencias al tiempo como “aún” y “horas” pueden entenderse como una espera de los clientes, algún problema que lleva horas y aún no tiene solución o respuesta. Por otro lado “cuanto” y “reposición” pueden entenderse como una falta de información de parte de la empresa a los clientes en cuanto al motivo de la interrupción del suministro y cuánto tiempo tomará su reposición.

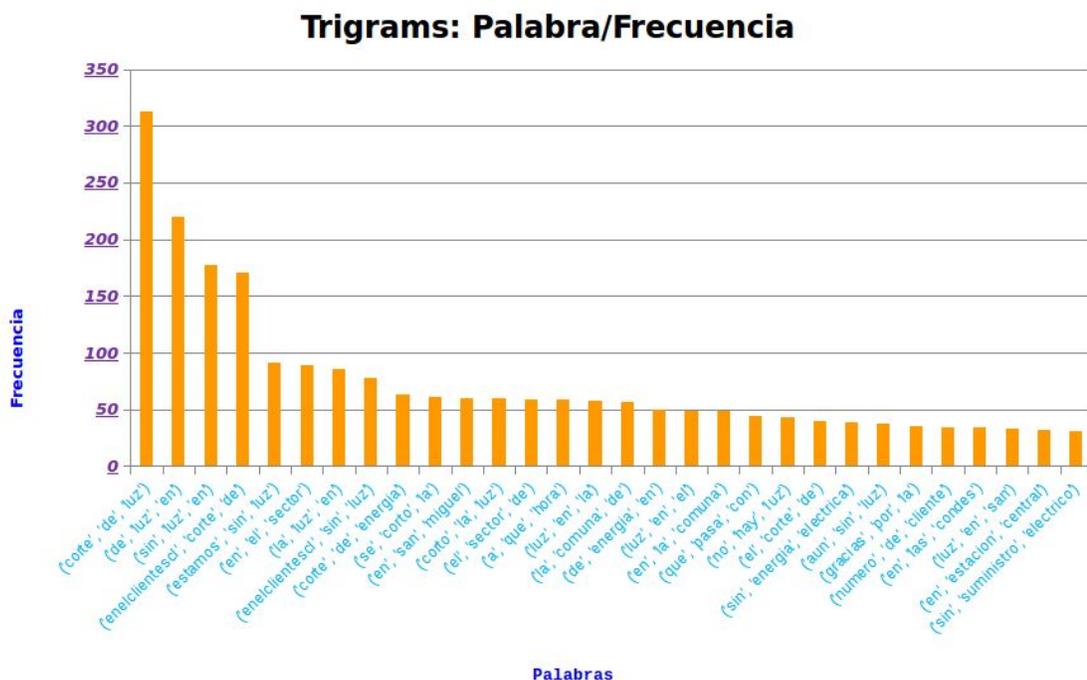
Bigrams: Palabra/Frecuencia



*Figura 74: Vista gráfico bi-grams de archivo excel generado por el código `Sentiment_Analysis_enel.py` para la data de septiembre-2018. Fuente: Elaboración Propia.

Para el caso de los bi-grams, se hace más evidente la predominancia de la clase informativa por sobre el resto. Más allá de realizar comentarios que puedan apuntar a algo negativo sobre la empresa, los bi-grams de mayor

frecuencia informan “sin, luz” , “corte, de” , “de, luz”, “en, el”, “en, la”, “comuna, de”, “sector, de”. Concentrándose en los comentarios posiblemente negativos, “que, pasa”, “que, pasó” y “a, qué”. Dichos bi-grams podrían representar como motivación de los Twitts sea, nuevamente, la falta de información sobre lo ocurrido, qué pasa con la luz, que pasó con el corte de suministro, a qué se debe el corte de luz o a qué hora se repondrá el servicio.



**Figura 75: Vista gráfico tri-grams de archivo excel generado por el código Sentiment_Analysis_enel.py para la data de septiembre-2018. Fuente: Elaboración Propia.*

Como ya se mencionó anteriormente, los tri-grams son los n-grams de esta investigación que mayor información entregan. Nuevamente, se confirma la clase informativa como la clase mayoritaria, siendo los tri-grams con mayor frecuencia los que informan sobre cortes de luz, corte de energía, estar sin

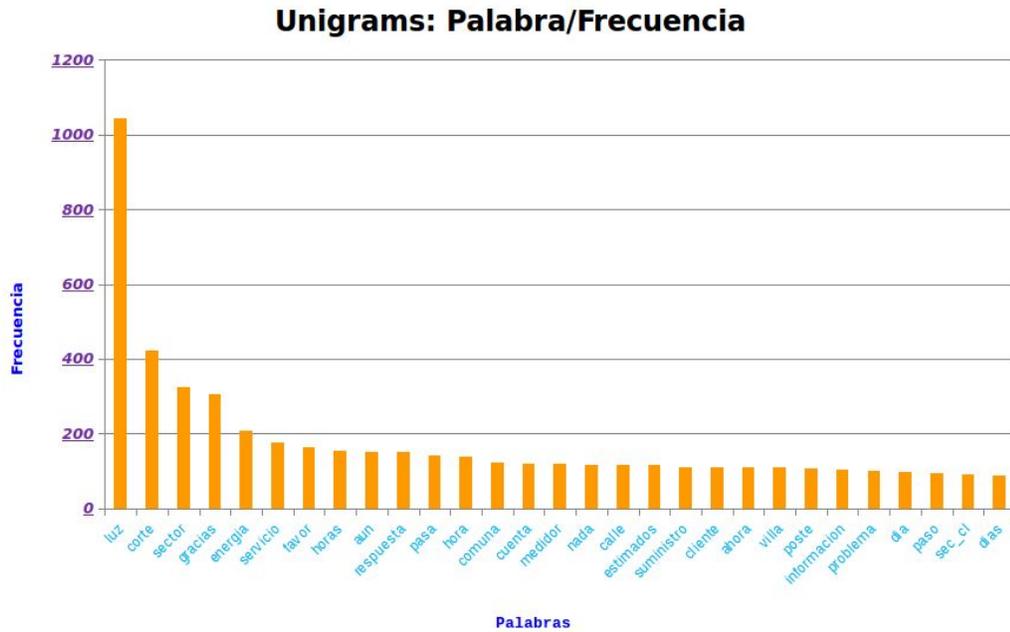
electricidad, sin energía, etcétera. En cuanto a los Twitts que pueden hacer tener connotación negativa, nuevamente aparecen los tri-grams como “a qué hora”, “qué pasa con” y “aún sin luz”. De estos tri-grams es posible señalar como principal motivo de los Twitts (negativos) en septiembre:

- Falta de información de parte de Enel Chile frente a cortes del suministro.
- Desconocimiento de los clientes sobre a qué se deben los cortes de suministro.
- Desconocimiento de los clientes sobre los tiempos estimados de reposición.
- Tiempos de reposición que toman más de lo esperado por los clientes.

6.2.4.2 Octubre-2018

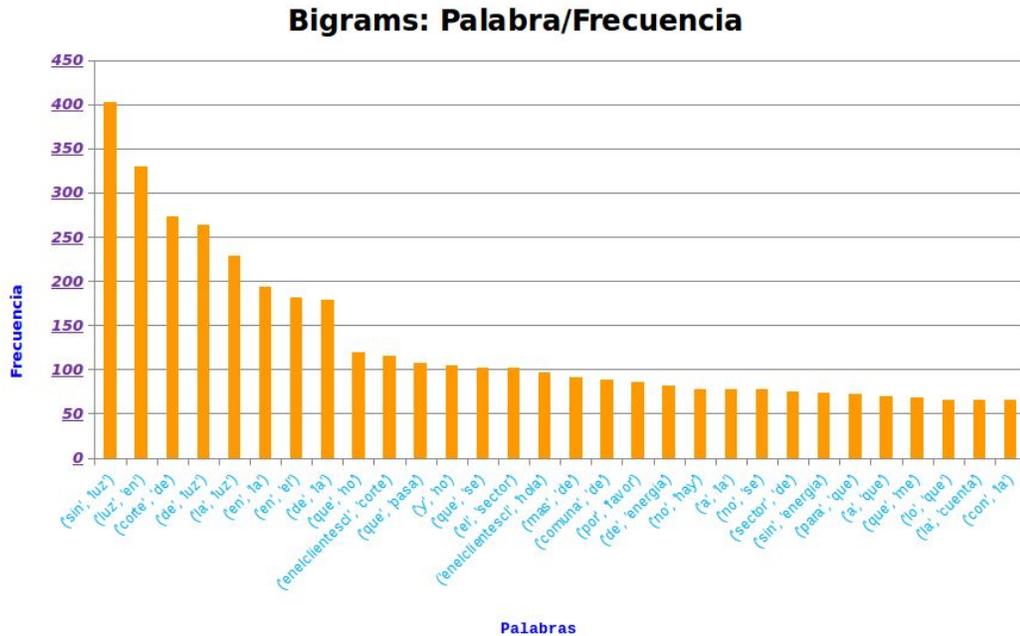
	30392	('sin', 'luz')	402	('corte', 'de', 'luz')	198
luz	1044	('luz', 'en')	330	('de', 'luz', 'en')	134
corte	421	('corte', 'de')	273	('sin', 'luz', 'en')	114
sector	325	('de', 'luz')	263	('enelclientescl', 'corte', 'de')	93
gracias	304	('la', 'luz')	229	('en', 'el', 'sector')	69
energía	207	('en', 'la')	193	('la', 'luz', 'en')	59
servicio	177	('en', 'el')	181	('corte', 'de', 'energía')	54
favor	162	('de', 'la')	178	('enelclientescl', 'sin', 'luz')	45
horas	154	('que', 'no')	119	('el', 'sector', 'de')	43
aun	149	('enelclientescl', 'corte')	116	('estamos', 'sin', 'luz')	42
respuesta	149	('que', 'pasa')	107	('horas', 'sin', 'luz')	38
pasa	141	('y', 'no')	105	('enelclientescl', 'que', 'pasa')	38
hora	139	('que', 'se')	102	('que', 'pasa', 'con')	37
comuna	123	('el', 'sector')	102	('a', 'que', 'hora')	37
cuenta	119	('enelclientescl', 'hola')	97	('la', 'comuna', 'de')	36
medidor	118	('mas', 'de')	91	('en', 'la', 'comuna')	36
nada	117	('comuna', 'de')	88	('luz', 'en', 'la')	34
calle	117	('por', 'favor')	86	('sin', 'luz', 'y')	33
estimados	114	('de', 'energía')	82	('luz', 'en', 'el')	32
suministro	110	('no', 'hay')	78	('sin', 'energía', 'eléctrica')	32
cliente	109	('a', 'la')	78	('se', 'corto', 'la')	31
ahora	108	('no', 'se')	78	('corto', 'la', 'luz')	30
villa	108	('sector', 'de')	75	('de', 'energía', 'en')	29
poste	107	('sin', 'energía')	74	('el', 'corte', 'de')	27
información	102	('para', 'que')	72	('numero', 'de', 'cliente')	25
problema	101	('a', 'que')	69	('un', 'corte', 'de')	25
día	98	('que', 'me')	68	('no', 'hay', 'luz')	25
paso	92	('lo', 'que')	65	('de', 'la', 'pirámide')	24
sec_cl	90	('la', 'cuenta')	65	('bosques', 'de', 'la')	23
días	87	('con', 'la')	65	('gracias', 'por', 'la')	23

*Figura 76: Vista tablas n-grams de archivo excel generado por el código `Sentiment_Analysis_enel.py` para la data de octubre-2018. Fuente: Elaboración Propia.



**Figura 77: Vista gráfico uni-grams de archivo excel generado por el código `Sentiment_Analysis_enel.py` para la data de octubre-2018. Fuente: Elaboración Propia.*

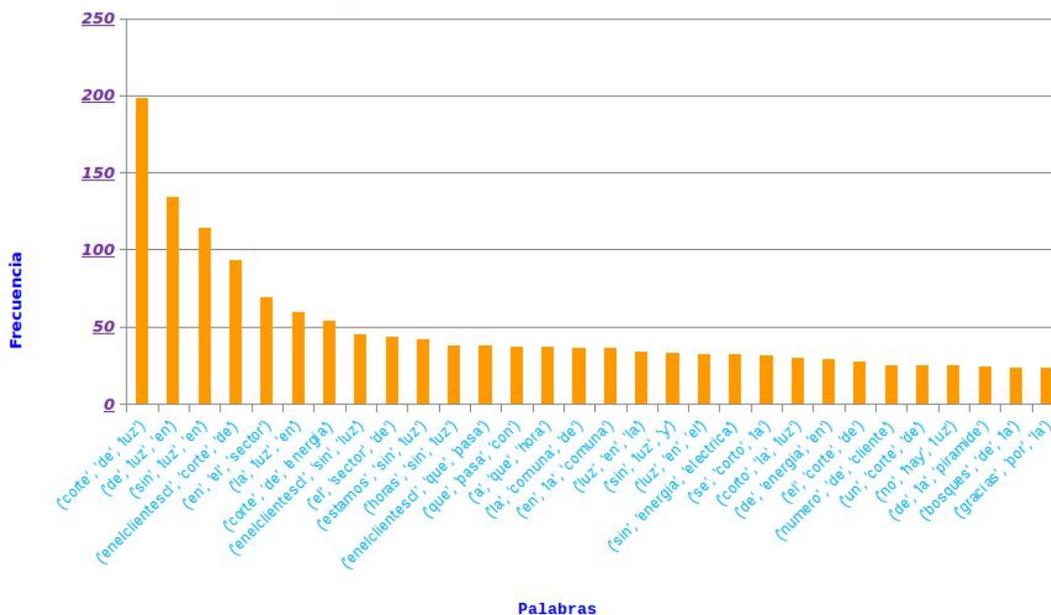
Para los uni-grams del mes de octubre, se presentan nuevamente “aún”, “poste” y “respuesta”. Aparecen nuevos uni-grams en los que es importante poner atención como “medidor”, “problema” y “medidores”. En base a los nuevos uni-grams es posible señalar como principal motivo fallas con los medidores y una alta presencia de problemas.



**Figura 78: Vista gráfico bi-grams de archivo excel generado por el código `Sentiment_Analysis_enel.py` para la data de octubre-2018. Fuente: Elaboración Propia.*

Para este mes algunos de los bi-grams que valen la pena destacar son “que pasa”, “la cuenta” y “a que”. Dos de estos bi-grams, bajo el contexto en el que se están estudiando, evidencian la incertidumbre que sienten los clientes, como por ejemplo, “qué pasa con el suministro / qué pasa con la luz” o “a qué se debe el corte / a qué hora vuelve la luz”. El otro bi-gram no se presentó en el mes de septiembre y tiene suma importancia con este y cualquier servicio: “la cuenta”. De manera preliminar, y con el conocimiento del contexto adquirido en la etapa de etiquetado, es posible pensar que los Twitts que hablan sobre la cuenta hagan alusión a cobros indebidos, lo que es un gran problema ya que genera desconfianza en los clientes.

Trigrams: Palabra/Frecuencia



*Figura 79: Vista gráfico tri-grams de archivo excel generado por el código `Sentiment_Analysis_enel.py` para la data de octubre-2018. Fuente: Elaboración Propia.

En cuanto a los tri-grams, nuevamente los con mayor frecuencia son los informativos, lo que valida la predominancia de la clase. En cuanto a las posibles motivaciones de los Twitts negativos, “horas sin luz” es de los tri-grams con mayor frecuencia. Resulta válido relacionar el tri-gram al contexto negativo ya que en ningún caso es bueno pasar horas sin luz o sin un suministro o servicio por el que se está pagando. Nuevamente, aparecen los “qué pasa” y en este caso aparecen en dos tri-grams: “enelclientescl qué pasa” y “qué pasa con”, lo que hace referencia al desconocimiento por parte de los clientes de lo que ocurre con el suministro y los posibles motivos de la alteración del mismo.

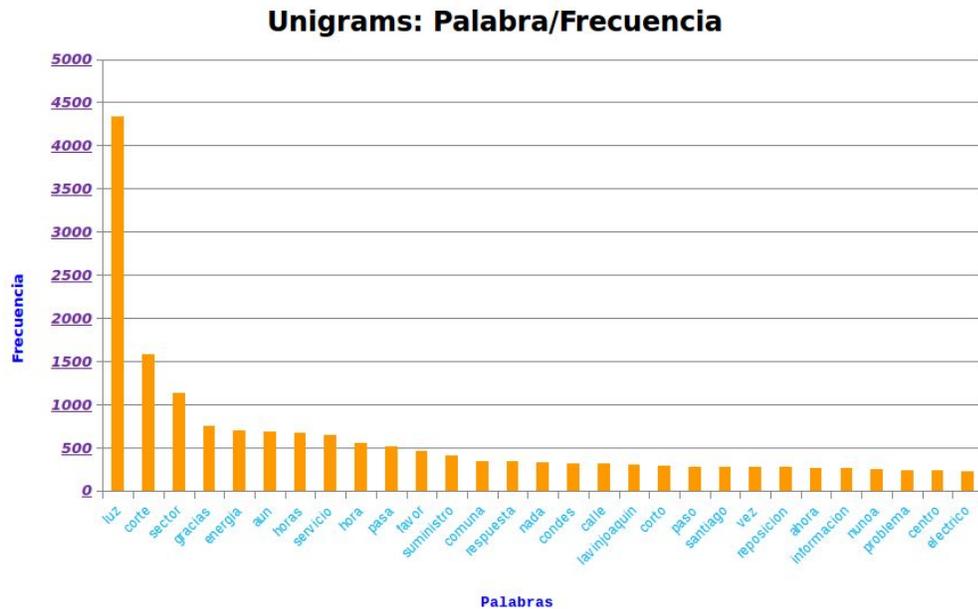
Finalmente, las principales motivaciones para el mes de octubre son:

- Desconocimiento por parte de los clientes sobre las alteraciones al servicio.
- Falta de información de parte de la empresa frente a incidentes que afecten a la continuidad del servicio.
- Problemas con las cuentas, posiblemente cobros indebidos.
- Altas menciones a los medidores y postes, ambos artefactos de los que debe hacerse cargo Enel Chile.

6.2.4.3 Noviembre-2018

	81818	('sin', 'luz')	1899	('corte', 'de', 'luz')	797
luz	4331	('luz', 'en')	1305	('de', 'luz', 'en')	522
corte	1578	('corte', 'de')	1068	('sin', 'luz', 'en')	465
sector	1130	('de', 'luz')	958	('enelclientesci', 'corte', 'de')	341
gracias	748	('la', 'luz')	864	('la', 'luz', 'en')	219
energía	694	('en', 'la')	538	('corte', 'de', 'energía')	190
aun	680	('en', 'el')	461	('estamos', 'sin', 'luz')	175
horas	675	('de', 'la')	406	('enelclientesci', 'sin', 'luz')	171
servicio	642	('que', 'pasa')	404	('se', 'corto', 'la')	167
hora	553	('enelclientesci', 'corte')	380	('corto', 'la', 'luz')	164
pasa	514	('las', 'condes')	318	('en', 'el', 'sector')	164
favor	458	('no', 'hay')	292	('aun', 'sin', 'luz')	152
suministro	404	('de', 'enelclientesci')	274	('horas', 'sin', 'luz')	151
comuna	347	('el', 'corte')	265	('a', 'que', 'hora')	149
respuesta	337	('mas', 'de')	265	('que', 'pasa', 'con')	148
nada	327	('de', 'energía')	264	('en', 'las', 'condes')	144
condes	321	('enelclientesci', 'que')	258	('enelclientesci', 'que', 'pasa')	133
calle	316	('que', 'se')	256	('el', 'corte', 'de')	133
lavinoaguin	297	('el', 'sector')	256	('luz', 'en', 'la')	128
corto	282	('por', 'favor')	253	('el', 'sector', 'de')	117
paso	280	('el', 'servicio')	250	('no', 'hay', 'luz')	115
santiago	280	('que', 'no')	245	('corta', 'la', 'luz')	109
vez	280	('estamos', 'sin')	239	('de', 'energía', 'en')	108
reposicion	273	('y', 'no')	238	('se', 'corta', 'la')	108
ahora	264	('sin', 'energía')	238	('la', 'comuna', 'de')	107
informacion	262	('comuna', 'de')	237	('sin', 'energía', 'eléctrica')	101
nunca	250	('horas', 'sin')	235	('en', 'la', 'comuna')	96
problema	234	('se', 'corto')	235	('seguimos', 'sin', 'luz')	93
centro	230	('que', 'hora')	234	('sin', 'luz', 'y')	90
eléctrico	228	('a', 'la')	228	('luz', 'en', 'sector')	87

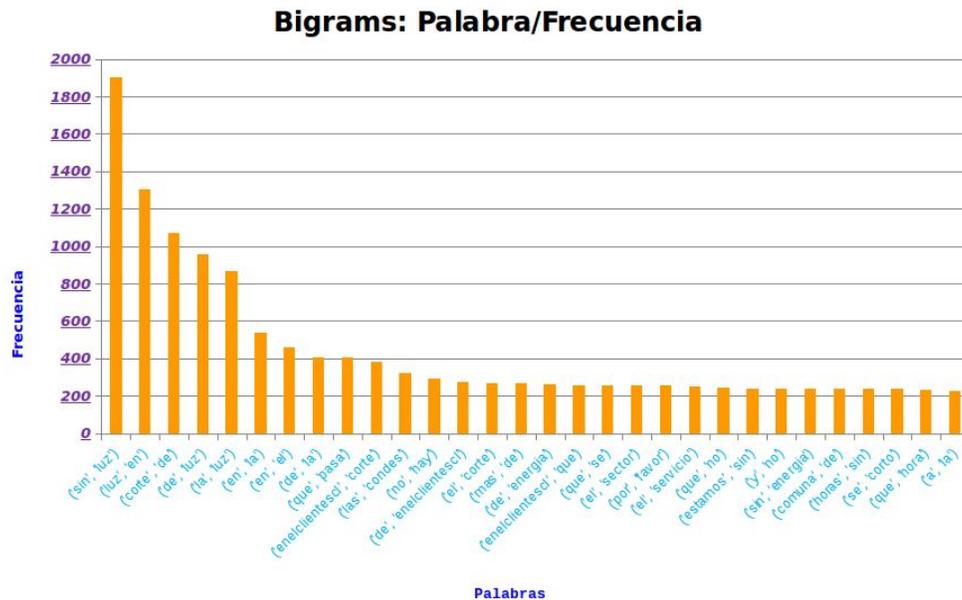
*Figura 80: Vista tablas n-grams de archivo excel generado por el código `Sentiment_Analysis_enel.py` para la data de noviembre-2018. Fuente: Elaboración Propia.



**Figura 81: Vista gráfico uni-grams de archivo excel generado por el código Sentiment_Analysis_enel.py para la data de noviembre-2018. Fuente: Elaboración Propia.*

Los uni-grams de noviembre, analizados en búsqueda de motivaciones negativas y críticas o quejas contra la empresa, pueden entregar información importante sobre lo que habría originado que noviembre sea el mes con mayor cantidad de Twitts de los meses estudiados. Se identifican como relevantes los siguientes uni-grams: “aún”, “horas”, “pasa”, “respuesta”, “nada”, “vez”, “reposición”, “información”, “problema”. Con conocimiento del contexto es posible realizar asunciones como: los uni-grams como “respuesta”, “pasa”, “reposición”, “información”, “problema” son palabras utilizadas en Twitts de contexto se incertidumbre y duda, anteriormente clasificados como informativos; los uni-grams como “aún”, “horas”, “nada”, “vez” son palabras utilizadas en Twitts que hablan del tiempo que le toma a la empresa solucionar los incidentes, de lo reiterado que es para el cliente la situación manifestada; y de

la poca o nula manifestación de Enel ya sea en redes sociales o en el lugar de los hechos.

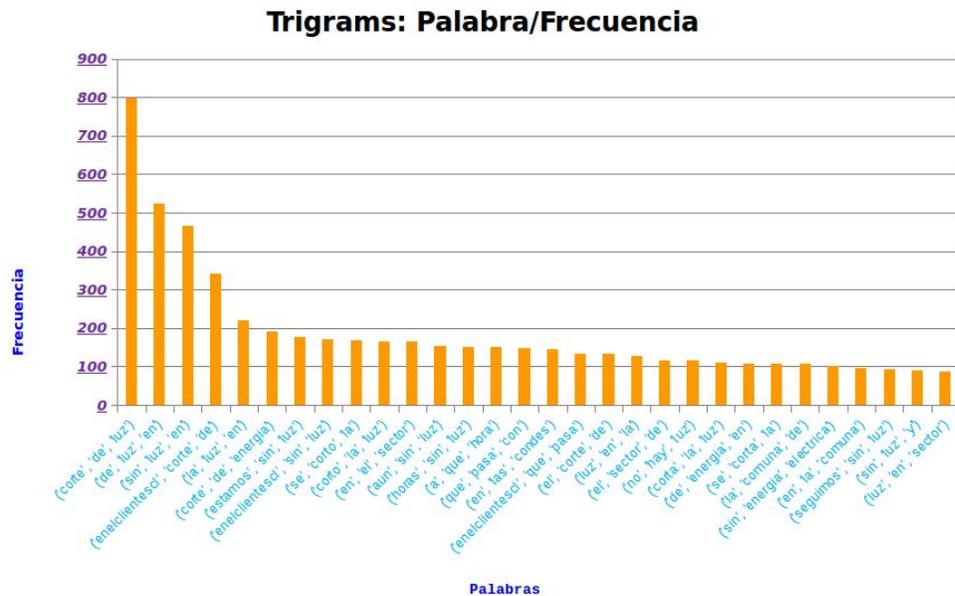


**Figura 82: Vista gráfico bi-grams de archivo excel generado por el código Sentiment_Analysis_enel.py para la data de noviembre-2018. Fuente: Elaboración Propia.*

Para los bi-grams, el “qué, pasa” nuevamente tiene un gran valor de frecuencia, lo que expresa la incertidumbre de los clientes frente a interrupciones del suministro energético. Resulta interesante recalcar la alta presencia del “más, de”, dado a que puede ser una referencia de tiempo que, según lo visto en la etapa de etiquetado, puede aplicar principalmente a una situación: referencia a cuánto tiempo ha pasado desde que se cortó el suministro o cuánto tiempo han pasado sin respuesta desde que hicieron su reclamo, pero siempre un comentario con respecto al tiempo de índole negativo. Esta idea se sustenta con otro bi-gram de gran frecuencia que es el “horas, sin”, que si se utiliza el análisis de que n-grams de similar frecuencia tienen mayores probabilidades de

provenir de los mismos Twitts, es posible formar “más de” X (número cualquiera) “horas sin” “sin luz” (el bi-gram de mayor presencia).

Por otro lado, es importante cuando en los distintos n-grams analizados, se observan referencias a lugares geográficos, más específicamente comunas, ya que una alta presencia de esta puede indicar dos cosas: primero, que el corte de suministro en ese sector afectó a mucha gente; segundo, que el corte de suministro se mantuvo por un largo periodo de tiempo, lo que generaba insistencia de los afectados a modo de presionar a la empresa y acelerar la entrega de una solución tanto como sea posible.



**Figura 83: Vista gráfico bi-grams de archivo excel generado por el código Sentiment_Analysis_enel.py para la data de noviembre-2018. Fuente: Elaboración Propia.*

Los tri-grams de noviembre, señalan nuevamente sensaciones de incertidumbre por parte de los clientes en frases como “enelclientescl qué pasa”, “qué pasa

con”, “a qué hora”. Además, se ven tri-grams que hacen referencia a la duración de los problemas o cortes de suministro como “horas sin luz”, “aún sin luz” o “seguimos sin luz”. Finalmente, vuelve a destacar la mención a “en las condes” como uno de los tri-grams de mayor frecuencia.

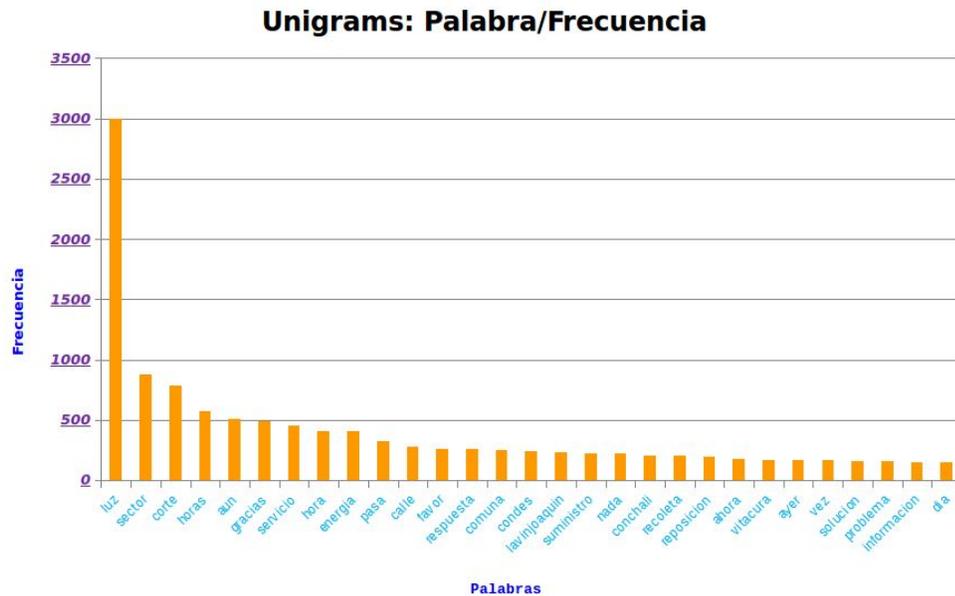
Con esto, es posible concluir señalando como principales motivaciones de noviembre lo siguiente:

- Cortes de energía o suministro prolongados.
- Sector de la comuna de Las Condes como principal víctima de problemas o cortes de suministro.
- Desconocimiento por parte de los clientes sobre los motivos de los cortes de energía.
- Desconocimiento por parte de los clientes sobre el tiempo que tardará la reposición del servicio.

6.2.4.4 Diciembre-2018

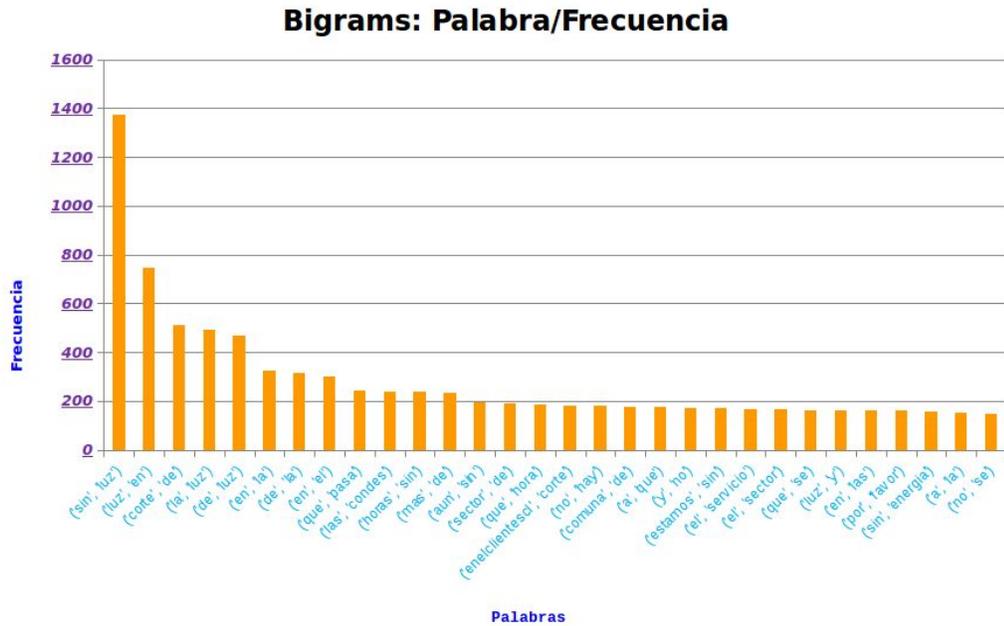
	53101		('sin', 'luz')	1371		('corte', 'de', 'luz')	375
luz	2996		('luz', 'en')	746		('sin', 'luz', 'en')	310
sector	874		('corte', 'de')	510		('de', 'luz', 'en')	242
corte	784		('la', 'luz')	490		('horas', 'sin', 'luz')	181
horas	569		('de', 'luz')	467		('enelclientescl', 'corte', 'de')	157
aun	510		('en', 'la')	326		('estamos', 'sin', 'luz')	137
gracias	484		('de', 'la')	315		('aun', 'sin', 'luz')	135
servicio	452		('en', 'el')	302		('a', 'que', 'hora')	126
hora	408		('que', 'pasa')	242		('en', 'el', 'sector')	122
energia	402		('las', 'condes')	238		('la', 'luz', 'en')	122
pasa	319		('horas', 'sin')	238		('sin', 'luz', 'y')	102
calle	273		('mas', 'de')	234		('en', 'las', 'condes')	97
favor	261		('aun', 'sin')	195		('enelclientescl', 'sin', 'luz')	97
respuesta	254		('sector', 'de')	188		('que', 'pasa', 'con')	87
comuna	250		('que', 'hora')	187		('corte', 'de', 'energia')	86
condes	240		('enelclientescl', 'corte')	181		('el', 'sector', 'de')	81
lavinjoaquin	229		('no', 'hay')	179		('no', 'hay', 'luz')	81
suministro	224		('comuna', 'de')	177		('seguimos', 'sin', 'luz')	79
nada	223		('a', 'que')	177		('corto', 'la', 'luz')	74
conchali	202		('y', 'no')	173		('la', 'comuna', 'de')	72
recoleta	200		('estamos', 'sin')	170		('se', 'corto', 'la')	71
reposicion	192		('el', 'servicio')	165		('luz', 'en', 'la')	68
ahora	175		('el', 'sector')	164		('sin', 'luz', 'desde')	64
vitacura	165		('que', 'se')	163		('el', 'corte', 'de')	64
ayer	164		('luz', 'y')	163		('luz', 'en', 'sector')	59
vez	162		('en', 'las')	160		('de', 'energia', 'en')	57
solucion	157		('por', 'favor')	159		('luz', 'en', 'el')	55
problema	152		('sin', 'energia')	157		('en', 'la', 'comuna')	54
informacion	149		('a', 'la')	149		('enelclientescl', 'que', 'pasa')	52
dia	148		('no', 'se')	147		('hasta', 'que', 'hora')	52

*Figura 84: Vista tablas n-grams de archivo excel generado por el código `Sentiment_Analysis_enel.py` para la data de diciembre-2018. Fuente: Elaboración Propia.



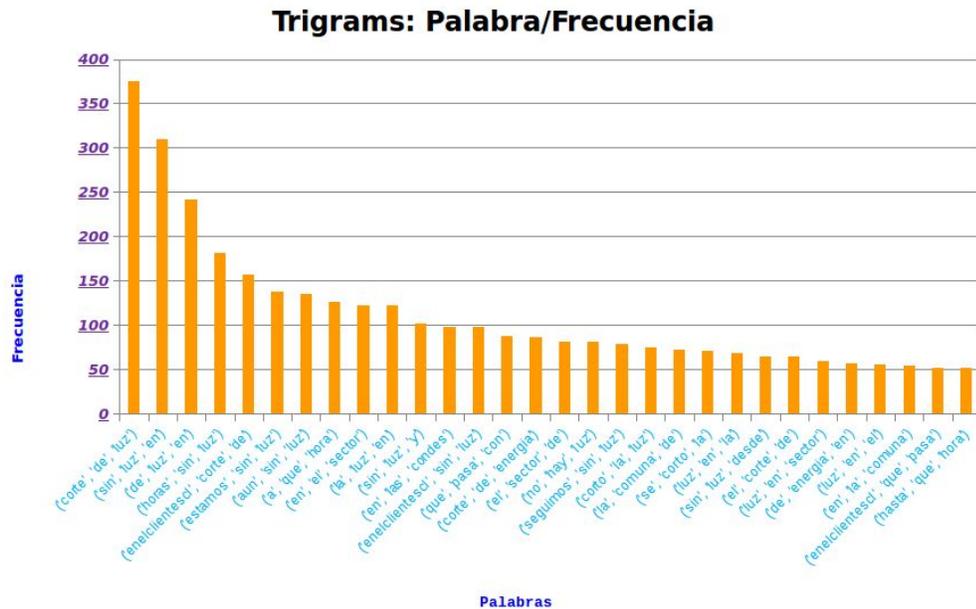
**Figura 85: Vista gráfico uni-grams de archivo excel generado por el código Sentiment_Analysis_enel.py para la data de diciembre-2018. Fuente: Elaboración Propia.*

Diciembre es el último mes estudiado, y en lo que a uni-grams respecta, nuevamente destacan: “horas”, “aún”, “información”, “reposición”, “problema” y “respuesta”, que ya fueron analizados en los meses pasados. Para este caso resulta importante analizar los uni-grams que hacen referencia a los sectores afectados como “condes”, “lavinjoaquin” (alcalde de las condes), “vitacura”, “conchalí” y “recoleta”. Como se señaló anteriormente, las referencias a lugares afectados como comunas, especialmente con tanta frecuencia como lo es “las condes” y su respectivo alcalde “joaquin lavín”, son un indicativo de que cualquiera sea el problema presentado, se ha mantenido por más del tiempo esperado o que los clientes no tienen información ni respuestas sobre el problema presentado.



**Figura 86: Vista gráfico bi-grams de archivo excel generado por el código Sentiment_Analysis_enel.py para la data de diciembre-2018. Fuente: Elaboración Propia.*

Siguiendo el análisis realizado en los uni-grams, en los bi-grams se presenta nuevamente con alta frecuencia la referencia a “las condes”. Esto es un indicio de que para el mes de diciembre, en dicho sector se presentaron muchos problemas o un problema de muy larga duración y en el que los clientes no tenían información al respecto. En esta línea, siguiendo la estrategia de asumir que los n-grams de similar frecuencia pueden pertenecer al mismo Twitt, se genera lo siguiente: “qué pasa” en “las condes” “horas sin” “sin luz”.



**Figura 87: Vista gráfico tri-grams de archivo excel generado por el código Sentiment_Analysis_enel.py para la data de diciembre-2018. Fuente: Elaboración Propia.*

Finalmente, los tri-grams sirven para validar o invalidar las asunciones realizadas en los uni y bi-grams. En este caso, los tri-grams validan 3 asunciones: primero, “a qué hora”, “hasta que hora”, “qué pasa con” son tri-grams que indican desconocimiento de parte de los clientes sobre qué ocurre y en cuanto tiempo se solucionará; segundo, “en las condes” está en el puesto 12 de los tri-grams con mayor frecuencia, lo que valida las asunciones realizadas para esa zona en los n-grams anteriores; tercero, “seguimos sin luz” “sin luz desde” hacen referencia a cortes de energía prolongados.

Con esto, las principales motivaciones de los Twitts de diciembre para Enel Chile son:

- Desconocimiento por parte de los clientes de qué ocasiona los cortes de suministro.
- Desconocimiento por parte de los clientes de cuánto tardará la reposición del suministro.
- Varios cortes de luz o cortes muy prolongados en la comuna de Las Condes.
- Cortes de luz se prolongan por más del tiempo esperado por los clientes.

Capítulo 7

7. Conclusiones

Este trabajo de investigación se enfocó en el estudio de dos principales temáticas: el análisis de sentimientos para lenguaje natural y la medición de la calidad de servicios. El resultado de este trabajo incluye un estudio basado en la literatura y papers científicos donde se investigaron distintas escalas de medición para la calidad de servicios y cómo esto impacta a la empresa (proveedora), cómo la mala calidad de un servicio puede generar sentimientos negativos en los clientes como frustración y enojo, lo que disminuye las probabilidades de que el cliente recomiende la empresa a un conocido o que vuelva a elegir la empresa para obtener el mismo servicio. O por otro lado, como una buena percepción de la calidad de servicio de una empresa puede producir que el cliente recomiende dicha empresa a sus pares y aumente la probabilidad de volver a acudir a la empresa en ocasiones futuras.

En cuanto al análisis de sentimientos, se investigaron distintos acercamientos y enfoques sobre cómo ponerlo en práctica, los desafíos que presenta, posibles soluciones y casos de experimentos con sus respectivos resultados.

Con lo aprendido, se trabajó en la construcción de un flujo de trabajo que consiste en una serie de softwares desarrollados en el lenguaje de programación Python que automatizan la realización de todo el proceso (por partes) incluyendo la captura de información (Web Scraping), la limpieza y tratamiento del texto (normalización), el entrenamiento de la máquina (machine learning) y finalmente el análisis de sentimientos a nivel de n-grams, todo esto con grandes sets de datos (Big Data).

En cuanto a la descarga de Twitts, para Claro Chile el mes de diciembre tuvo la menor cantidad de Twitts mientras que el mes de septiembre tuvo la mayor. Para Enel Chile, noviembre fue el mes con mayor cantidad de Twitts y octubre el con menor.

Para el proceso de entrenamiento, para el caso de ambas empresas (Claro y Enel), **LogisticRegression** fue el clasificador con mejor desempeño y en ambos casos SVM fue el segundo mejor clasificador, con resultados muy cercanos a los de LogisticRegression.

Luego, probando los modelos de LogisticRegression entrenados para cada empresa en data no clasificada, los resultados arrojaron que los comentarios negativos para Claro Chile fluctuaban entre el 56.34% y el 61.87%. En cambio, para Enel Chile, los comentarios negativos fluctuaban entre el 11.56% y el 16.99%.

Es importante señalar que las diferencias de porcentaje de Twitts negativos entre ambas empresas están justificadas debido a condiciones intrínsecas de cada servicio. En el caso de Claro Chile, la responsabilidad tanto de las fallas del servicio como de su reposición va por parte de la empresa. Esto se fundamenta con la alta presencia de comentarios señalando problemas de señal móvil o en su señal de internet (móvil y hogar), en ambos casos es deber de Claro Chile solucionarlo y cumplir con la entrega de un servicio 100% funcional en todas sus versiones para todos sus clientes. En el caso de Enel Chile, como ya se mencionó en la etapa de resultados, sus servicios pueden verse afectados por intervenciones de terceros como un choque automovilístico con un poste de electricidad, o desastres naturales como caída de árboles sobre cables de tendido eléctrico o fuertes lluvias. Adicional a esto, para Claro Chile se utilizaron sólo 2 clases, mientras que para Enel Chile se utilizaron 3.

Luego de identificar la polaridad de los Twitts, se identificaron las principales motivaciones de carácter negativo al momento de escribir los Twitts para cada mes. A continuación se presentan los motivos principales a nivel general para cada empresa.

Claro Chile:

- Señal nula o de mala calidad.
- Servicio de internet nulo o de mala calidad (móvil y hogar).
- Los tiempos de espera de los clientes y la deficiente capacidad de respuesta de Claro Chile.
- Alta frecuencia de problemas (“todos los días”).
- Reiterado acoso de manera telefónica por parte del callcenter.
- Cobros indebidos en las boletas.

- Solicitar dar de baja los servicios.
- Reclamos directos a través de menciones a instituciones reguladoras (sernac y subtel_chile).

Enel Chile:

- Falta de información de parte de Enel Chile frente a cortes del suministro.
- Desconocimiento de los clientes sobre el motivo que origina los cortes de suministro.
- Desconocimiento de los clientes sobre los tiempos estimados de reposición.
- Tiempos de reposición que toman más de lo esperado por los clientes.
- Problemas con las cuentas, posiblemente cobros indebidos.
- Cortes de energía o suministro prolongados.
- Gran cantidad de menciones a comunas específicas (problemas focalizados).

Estas motivaciones fueron extraídas de más de 10.000 Twitts para Claro Chile y más de 20.000 Twitts para Enel Chile. Debido a la limpieza de stopwords que se realiza, la ejecución de los códigos de análisis de sentimientos (n-grams) tomó hasta 40 minutos para el caso de noviembre de Enel. Esto quiere decir que el código de n-grams permitió tener información semi-concluyente respecto a las motivaciones de los Twitts en 40 minutos, una fracción mínima de lo que significa revisar uno a uno los más de 10.000 Twitts del mes de noviembre.

Este trabajo, a pesar de haberse desarrollado a un nivel relativamente básico, logró demostrar la importancia del machine learning en el big data, evidenciar su impacto en términos de beneficio versus tiempo, dar muestra fundamentada

de el beneficio empresarial que un flujo de trabajo como este puede generar, exponer y analizar los principales desafíos del trabajo de texto en lenguaje natural y cómo atacarlos, potenciar Twitter y las redes sociales en general como medios válidos de manifestación para el público y canales válidos para ser un medio intercomunicador entre la empresa y el cliente permitiendo abstraer lo que el cliente opina y a través de esto conocer las principales falencias de la empresa, entregando mayor nivel competitivo.

7.1 Trabajos Futuros

Durante el desarrollo de este trabajo se presentaron problemas y dificultades de las cuales muchas, por motivos de tiempo y/o capacidad de procesamiento, fueron dejados fuera (en cuanto a implementación) pero se realizó la labor investigativa y analítica de cómo enfrentarlos y darles posible solución.

Dentro del análisis de los resultados obtenidos, y comparándolo con los resultados esperados, resulta propicio separar esta sección en tres partes principales donde es posible realizar mejoras que tendrían un gran impacto en la calidad de los resultados obtenidos.

7.1.1 Machine Learning

Como se mencionó anteriormente, para la etapa de entrenamiento de la máquina, en todos los modelos se utilizaron los parámetros por defecto (*default*). Lo que ocasionó que en cada iteración los resultados fueran diferentes fue la data con la que eran entrenados, específicamente el proceso en el que se divide la data clasificada en el grupo utilizado para entrenar y el grupo utilizado para validar. Se tomó esta decisión en base a que tanto en los estudios y

papers revisados como en la teoría del Machine Learning se da énfasis al gran impacto que tiene en los resultados posteriores lo bien o mal que represente la data de entrenamiento a la data (no-clasificada) a la que se enfrentará el modelo posteriormente, además de que con la herramienta Scikit era un parámetro fácil de modificar.

Dicho esto, surge una mejora evidente, y es que al encontrar el set de entrenamiento con mejores resultados dentro de N iteraciones, proceder a realizar más iteraciones con ese mismo set de entrenamiento pero variando los parámetros de cada modelo. Esta tarea requiere un gran tiempo para realizar pruebas, dado que los distintos modelos ofrecen muchos parámetros a modificar, entre los cuales destacan: parámetros de penalización, parámetros que cambian las heurísticas a utilizar, parámetros de balanceo de clases, parámetros de variaciones a la función de decisión, parámetros de variación de semillas, parámetros de elección de algoritmos resolutivos (que varían según el tamaño del dataset), parámetros de paralización de procesamiento para los distintos núcleos del procesador, etcétera.

7.1.2 Análisis de Sentimientos

Para este trabajo, se decidió realizar el análisis de sentimientos a nivel de datasets (mensuales) en lugar de realizarlo a nivel de Twitt, ya que eso de una u otra forma lo realizaba el modelo clasificador en el proceso de prueba. El proceso realizado entrega como resultado un archivo excel que entrega los n-grams de mayor frecuencia encontrados en los Twitts de cada mes, lo que es sumamente útil ya que reduce la tarea de leer miles de Twitts a sólo leer tablas con, en este caso, los 30 n-grams de mayor frecuencia. De todos modos, los resultados siguen requiriendo de una persona que los interprete y de

preferencia dicha persona tiene que tener conocimientos del contexto al que se está enfrentando.

Para solucionar esto, se propone un diccionario de polaridad de palabras ajustado al contexto. Esto se refiere a la creación de un documento de texto que tenga un listado de palabras (tan grande como sea posible) y dicha palabra tenga asignada una polaridad ya sea en formato de texto (positivo, negativo, neutral) o con un número que represente dicha polaridad. El fin de esto es que, luego de encontrados los n-grams de mayor frecuencia, se comparen con este diccionario y resalte de manera automática los que tengan mayor importancia y entreguen mayor información que en este caso serían los negativos. Nuevamente, esta mejora requiere tiempo de implementación y de cómputo bastante elevados, ya que comprende la comparación de la cantidad “X” de n-grams de mayor frecuencia con un diccionario de palabras tan grande como sea posible. Esto propone nuevos desafíos en cuanto a la implementación como lo es la optimización en cuanto a heurísticas y estructuras de datos para la comparación, así como la paralización de procesamiento (hilos) para un ahorro de tiempo de cómputo.

Otros desafíos que se dejaron fuera y que tendrían un gran impacto en el análisis de sentimientos son:

- La consideración de negaciones.
- Distinción de palabras en el idioma español por género (abarcado en el siguiente punto).
- Consideración de n-grams de contexto (diferenciar “servicio la raja” a “servicio como la raja”).
- Detección de sarcasmo (muy difícil de detectar).

- Unificar las distintas conjugaciones de palabras (abarcado en el siguiente punto).

7.1.3 Limpieza, normalización y tratamiento de texto en lenguaje natural

Finalmente, pero no menos importante, de hecho, posiblemente la etapa más importante de todo el proceso realizado para este trabajo, están los desafíos y dificultades que presenta trabajar con texto en lenguaje natural.

Extraer data de Twitter o de cualquier red social con la finalidad de realizar un análisis a nivel de Big Data trae consigo enfrentarse a un contexto de libre expresión no formal, donde pueden pasarse por alto reglas ortográficas y de sintaxis ya sea de manera voluntaria o involuntaria. A pesar de las herramientas que entregan los navegadores o los sistemas operativos para la corrección de errores ortográficos, la cantidad de mensajes mal escritos es muy grande y se evidencio en la etapa de etiquetado de este trabajo.

Todos estos errores generan ruido al momento de tratar la información, dentro del principal ruido encontrado destacan:

- Error ortográfico.
- Repetición voluntaria de caracteres (Ej: “me tienen chatoooo”, “nooooo”).
- Error de tipeo que resulta en la concatenación de palabras o caracteres de palabras distintas.
- Concatenación de palabras a modo de hashtag. (Ej: “mal servicio” escrito como #malservicio).

La propuesta de solución al problema de repetición de caracteres es utilizar una biblioteca de Python llamada *itertools*. Esta biblioteca incluye una función llamada *groupby* que permite remover los caracteres repetidos de un string (cadena de caracteres) y mantener el orden original de los mismos a través de la agrupación de los caracteres no repetidos de manera adyacente.

En cuanto a los errores ortográficos, se propone una solución que involucra un leuario en español y el uso de la distancia Levenshtein. El leuario de una lengua contiene todas las palabras (lemas) que forman dicha lengua. El leuario español oficial lo provee la RAE (real academia española) pero lamentablemente no es algo de acceso abierto y público. Por suerte, Ismael Olea se dió el trabajo y subió a la plataforma GitHub un leuario en español compuesto de 87.899 palabras.

La distancia Levenshtein se define como “el valor de distancia que describe el número mínimo de inserciones, sustituciones o supresiones requeridos para transformar un string (fuente) a otro string (objetivo). La idea es calcular la distancia Levenshtein entre las palabras de un Twitt y el leuario en español, y reemplazar la palabra del Twitt por la del leuario a la que más se acerque. La implementación de esta solución debe ser de manera inteligente, ya que puede darse el caso de que haya más de una palabra objetivo a la misma distancia de la palabra de origen, además de comparar de manera previa las palabras de los Twitts con las del leuario para evitar la futura comparación con palabras que ya pertenezcan a él, asegurándose de comparar sólo las palabras que no pertenecen a la lengua española. Pero las dificultades de implementación de esta solución no terminan ahí, ya que del modo que sea implementada, incurrirá en una gran cantidad de comparaciones. A modo de ejemplo, los Twitts descargados sobre Enel Chile para el mes de diciembre contienen en total

104.359 palabras, que en teoría deberían compararse cada una con las 87.900 palabras del lecionario español.

Además, se estudió un método denominado Stemmer. Su nombre proviene de la palabra *stem* (tallo o tronco en inglés) y hace referencia a las raíces u origen de las palabras. Su funcionamiento se describe como el remover los afijos (prefijos, sufijos, infijos) morfológicos de una palabra, dejando sólo la palabra raíz. Para el caso del lenguaje de programación Python, la biblioteca *nltk* tiene una función llamada *SnowballStemmer* que tiene implementada la función con soporte para los idiomas: danés, holandés, finlandés, francés, alemán, húngaro, italiano, noruego, portugués, romano, ruso, inglés, español y sueco. Cabe señalar que la implementación de esta función debe ser adaptada para funcionar de manera correcta en este contexto, ya que muchas veces puede retornar como palabra raíz palabras que no pertenezcan al lecionario español como tal, por ejemplo, con la palabra “encontramos” retorna “encontr”. Una buena implementación de esta solución podría ayudar a disminuir los *features* que en español separan las palabras por género, donde “aburrido” y “aburrida” son palabras distintas que expresan exactamente lo mismo.

Bibliografía

[1] Axel Bruns & Stefan Stieglitz (2013) Towards more systematic *Twitter* analysis: metrics for tweeting activities, *International Journal of Social Research Methodology*. Recuperado de <https://www.tandfonline.com/doi/abs/10.1080/13645579.2012.756095> .

[2] Mejía, JC. (2018). *Estadísticas de redes sociales 2018: Usuarios de Facebook, Twitter, Instagram, Youtube, LinkedIn, Whatsapp y otros*. Recuperado de https://www.juancmejia.com/marketing-digital/estadisticas-de-redes-sociales-usuarios-de-facebook-instagram-linkedin-twitter-whatsapp-y-otros-infografia/#5_Usuarios_activos_de_Twitter .

[3] Anderson, E. L., Steen, E., & Stavropoulos, V. (2017). Internet use and Problematic Internet Use: A systematic review of longitudinal research trends in adolescence and emergent adulthood. *International Journal of Adolescence and Youth*, 22(4), 430-454. Recuperado de <https://www.tandfonline.com/doi/abs/10.1080/02673843.2016.1227716> .

[4] Culotta, A., Ravi, N. K., & Cutler, J. (2015). Predicting the demographics of Twitter users from website traffic data. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence, AAAI 2015 and the 27th Innovative Applications of Artificial Intelligence Conference, IAAI 2015* (Vol. 1, pp. 72-78). AI Access Foundation. Recuperado de <https://www.scholars.northwestern.edu/en/publications/predicting-the-demographics-of-twitter-users-from-website-traffic> .

- [5] Russel, S. Dewey, D. Tegmark, M. (2015). *Research Priorities for Robust and Beneficial Artificial Intelligence*. Recuperado de <https://www.aaai.org/ojs/index.php/aimagazine/article/view/2577/2521> .
- [6] O'Leary, D. (2013). *Artificial Intelligence and Big Data*. Recuperado de <https://ieeexplore.ieee.org/abstract/document/6547979/> .
- [7] Robert, C. (2014). *Machine Learning, a Probabilistic Perspective*, CHANCE, DOI: 10.1080/09332480.2014.914768.
- [8] Bottou, L. (2013). *From machine learning to machine reasoning. An essay*. DOI: DOI 10.1007/s10994-013-5335-x.
- [9] Scharkow, M. (2011). *Thematic content analysis using supervised machine learning: An empirical evaluation using German online news*. DOI: 10.1007/s11135-011-9545-7.
- [10] Habernal, I. Ptacek, T. Steinberger, J. (2013). *Sentiment Analysis in Czech Social Media Using Supervised Machine Learning*. Recuperado de <http://www.aclweb.org/anthology/W13-1609> .
- [11] Wuest, T., Irgens, C. & Thoben, KD. J Intell Manuf (2014). *An approach to monitoring quality in manufacturing using supervised machine learning on product state data*. Recuperado de <https://doi.org/10.1007/s10845-013-0761-y> .

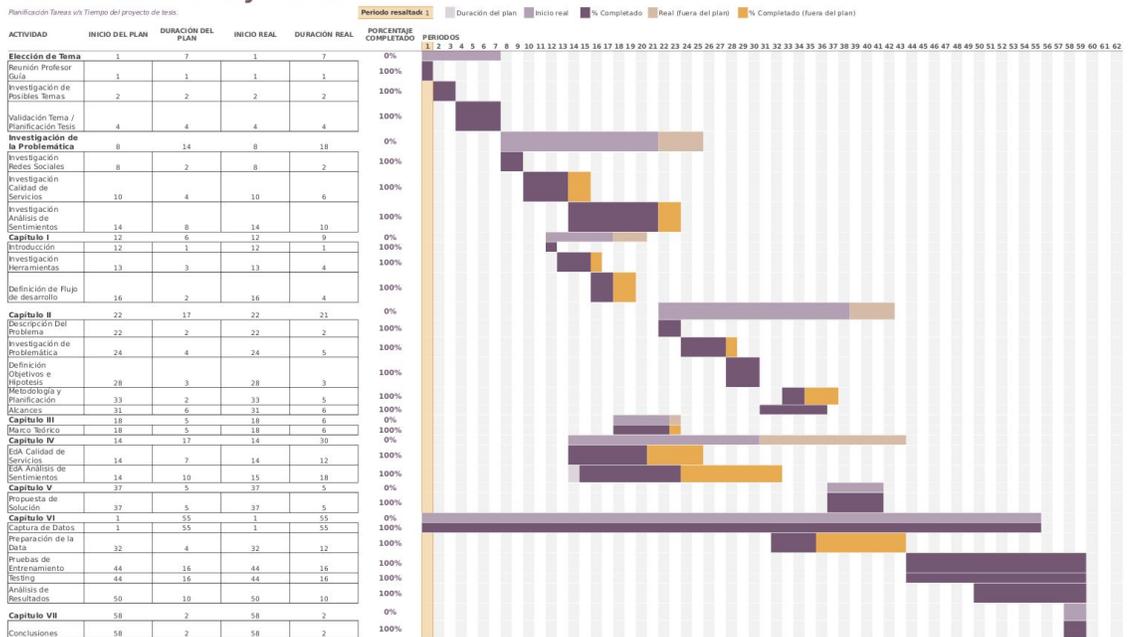
- [12] Srivastava, N. (2015). *Unsupervised Learning of Video Representations using LSTMs*. University of Toronto, Canada. Recuperado de <http://proceedings.mlr.press/v37/srivastava15.pdf> .
- [13] Demirci, F., Kara, A. (2014). *Supermarket self-checkout service quality, customer satisfaction, and loyalty: Empirical evidence from an emerging market*. Recuperado de <https://www.sciencedirect.com/science/article/pii/S0969698913000829>.
- [14] Dabholkar, P. A. (2014). *How to Improve Perceived service quality by Increasing Customer Participation*. *Developments in Marketing Science: Proceedings of the academy of marketing science*. Doi: 10.1007/978-3-319-13254-9_97.
- [15] Feldman, R. (2013). *Techniques and applications for sentiment analysis*. *Communications of the ACM*. Doi: 10.1145/2436256.2436274.
- [16] Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). *New avenues in Opinion Mining And sentiment analysis*. *IEEE Intelligent Systems*. Doi: 10.1109/MIS.2013.30/.
- [17] Saif M. Mohammad, Kiritchenko, S., Zhu, X. (2013). *NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets*. National research council Canada. Recuperado de <https://arxiv.org/abs/1308.6242>.
- [18] Mitchell, R. (2018). *Web Scraping with Python: Collecting More Data from the Modern Web*. " O'Reilly Media, Inc."

[19] Pedregosa *et al.* (2011). “*Scikit-learn: Machine learning in Python*”, *Journal of Machine Learning Research* 12, pp. 2825-2830.

Anexos

Anexo 1: Planificación utilizada en el trabajo de tesis.

Carta Gantt - Proyecto de Tesis



*Figura 88: Diagrama de carta gantt de la planificación del proyecto de tesis. Fuente: Elaboración Propia.

Anexo 2: Experimentación de aprendizaje no-supervisado (descartado)

Clustering

Se conoce como la tarea de dividir la población o datos representados en un plano n-dimensional correspondiente a los n-features que representen la data. Se hace de manera tal que los datos que quedan en los mismos grupos sean

más similares a los otros puntos pertenecientes a ese grupo que a los de otros grupos. En pocas palabras, el objetivo es segregarse grupos con similares características y asignarlos a clusters.

Existen muchos métodos para lograr el objetivo de agrupar la data, cada método sigue distintas reglas para definir la similaridad entre los datos. En base a su funcionamiento, estos algoritmos pueden separarse en 4 tipos:

- Modelos de conectividad: como su nombre sugiere, estos modelos se basan en la noción de que los datos más cercanos entre sí en el espacio de datos muestran una mayor similitud entre ellos que con la data más alejada. Estos modelos tienen dos mecanismos de funcionamiento. En el primero, se comienza clasificando toda la data en clusters distintos (uno por cada dato) y luego segregarlos a medida que la distancia disminuye. En el segundo, todos los datos son clasificados en un único cluster y luego particionados a medida que la distancia aumenta. Además, la función de distancia es subjetiva. Estos modelos son fáciles de interpretar pero carecen de escalabilidad para manejar grandes sets de datos. Ejemplos de estos modelos son los algoritmos de clustering jerárquico (hierarchical algorithm) y sus variantes.
- Modelos de centroides: estos son algoritmos de clustering iterativos en los que la noción de similaridad entre los datos se obtiene a partir de la cercanía de los datos con el centro del cluster. El algoritmo de clustering “K-means” es un algoritmo popular que pertenece a esta categoría. En estos modelos, el número de clusters requeridos para el final de la ejecución del algoritmo debe ser declarado de antemano, por lo que resulta importante tener conocimiento previo de los sets de datos. Estos modelos corren de manera iterativa para encontrar el óptimo local.
- Modelos de distribución: estos modelos de clustering están basados en la noción de qué tan probable es que los datos en el cluster pertenezcan a la misma distribución (por ejemplo: normal, gaussiana). Estos modelos suelen caer en el *overfitting*. Un ejemplo popular de estos modelos es el algoritmo “expectation-maximization” el cual utiliza distribución normal multivariable.
- Modelos de densidad: estos modelos se basan en la detección de áreas con la mayor concentración o densidad de puntos y dónde están separados por áreas vacías o puntos escasos. Aíslan varias regiones de

densidad diferente y asignan los datos dentro de esas regiones al mismo cluster. Ejemplos populares de modelos de densidad son DBSCAN y OPTICS.

K-means

Como ya fue mencionado, este algoritmo pertenece a los modelos de centroides. Este algoritmo realiza clusters de la data intentando separar las muestras en n grupos de la misma diferencia, minimizando el criterio conocido como *Inertia* que consiste en el cálculo de la distancia de cada punto al centroide del cluster más cercano. Este algoritmo requiere que el número de clusters sea especificado de antemano. Escala muy bien ante un gran número de muestras y ha sido utilizado en un gran número de áreas de aplicación en distintos campos.

El funcionamiento de este algoritmo puede explicarse en los siguientes 5 pasos:

1. Especificar el número K de clusters deseados.
2. Inicializar K centroides en el espacio (pueden ser puntos aleatorios).
3. Por cada muestra, ver qué centroide es el más cercano y asignarlo.
4. Por cada centroide, mover el centroide al centro de los puntos asignados.
5. Repetir los últimos 2 pasos hasta que la asignación de centroides no sufra cambios.

El algoritmo K-means divide un set de N muestras X en K clusters disjuntos C , cada uno descrito por la media μ_j de las muestras en el cluster. Las medias son comúnmente llamadas “centroides”; notar que no son, en general, puntos de X , aunque viven en el mismo espacio. El algoritmo K-means busca escoger los centroides que minimizan la inercia:

$$\sum_{i=0}^n \min_{\mu_j \in C} (\|x_i - \mu_j\|^2)$$

*Figura 89: Fórmula matemática de minimización de ‘inercia’. Fuente: Documentación Scikit.

Este algoritmo fue probado y descartado por los deficientes resultados entregados. Se realizaron 20 iteraciones del algoritmo y en ninguna fue capaz de separar las clases de la manera que se necesita para la investigación. Por lo mismo, se optó por utilizar aprendizaje supervisado etiquetando la data de

manera manual. Es un trabajo muy lento y tedioso pero permite obtener resultados ajustados al contexto y por ende poder sacar posteriores conclusiones en base a lo que se busca.

De todas formas, es importante señalar que los parámetros utilizados fueron:

- `n_clusters = 3`
- `init = 'k-means++'`
- `max_iter = 300`
- `n_init = 3`
- Otros parámetros = default.

Donde:

- **n_clusters** es el número de clusters que se desea tener, es decir, la cantidad de grupos o clases en las que se espera sea separada la data.
- **init** define el método de inicialización de centroides de clusters. Scikit tiene definido `k-means++` como un método inteligente de acelerar la convergencia. De no definir un valor, `k-means++` es el valor considerado por defecto.
- **max_iter** es el número máximo de iteraciones para el algoritmo K-means para cada ejecución del algoritmo.
- **n_init** es el número de veces que el algoritmo K-means correrá con diferentes semillas de centroides. El resultado final será la mejor salida de las `n_init` ejecuciones consecutivas en términos de inercia.