

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

FACULTAD DE INGENIERÍA

ESCUELA DE INGENIERÍA INFORMÁTICA

**CLASIFICACIÓN AUTOMÁTICA DE LA INTENCIÓN
DEL USUARIO EN MENSAJES DE TWITTER**

MAURICIO ALEJANDRO MARTIS CÁCERES

TESIS DE GRADO

MAGÍSTER EN INGENIERÍA INFORMÁTICA

ENERO, 2012

Pontificia Universidad Católica de Valparaíso

Facultad de Ingeniería

Escuela de Ingeniería Informática

**CLASIFICACIÓN AUTOMÁTICA DE LA INTENCIÓN
DEL USUARIO EN MENSAJES DE TWITTER**

MAURICIO ALEJANDRO MARTIS CÁCERES

Profesor Guía: **Rodrigo Alfaro Arancibia**

Programa: **Magíster en Ingeniería Informática**

Enero, 2012

Dedicado a mis padres, por el amor
y apoyo incondicional.

Agradecimientos

A mis padres y hermanos, por el amor y confianza que me han entregado. Por estar preocupados, en todo momento, de que pudiese lograr cada uno de mis objetivos académicos.

A Sandra, mi amor, por su apoyo incondicional en los momentos que más lo necesité. Y por entregarme la mayor felicidad del mundo, nuestro hijo Nicolás.

A mi familia, amigos y compañeros por estar preocupados de que lograra mis objetivos y ofrecerme su ayuda cuando realmente los necesitaba.

Al profesor Rodrigo Alfaro, por orientarme y guiarme durante todo el proceso que me han permitido cumplir este objetivo.

A la Universidad, y principalmente a la Escuela de Ingeniería Informática, por la formación profesional entregada.

Índice

Resumen	iii
Abstract.....	iii
Lista de Figuras	iv
Lista de Tablas.....	v
1 Introducción.....	1
1.1 Descripción del Problema.....	2
1.2 Definición de Objetivos.....	3
1.2.1 Objetivo General	3
1.2.2 Objetivos Específicos	3
1.3 Estructura del Contenido	4
2 Marco Teórico	5
2.1 Clasificación Automática de Textos.....	5
2.1.1 Introducción.....	5
2.1.2 Definición de Clasificación de Textos	5
2.1.3 Representación de Documentos y Pre-procesamiento	6
2.1.4 Máquinas de Aprendizaje	9
2.2 Twitter	16
2.2.1 Introducción.....	16
2.2.2 Características de Twitter	19
2.3 Trabajos Relacionados.....	20
2.3.1 Resumen de Estudios Relacionados	22
3 Modelo Propuesto.....	23
3.1 Definición de la Taxonomía	23
3.1.1 Descripción de las Categorías	24
3.2 Clasificación Jerárquica.....	26
3.2.1 Jerarquía I	27
3.2.2 Jerarquía II.....	27
3.2.3 Pre-procesamiento	28
3.2.4 Representación	30
3.3 Modelo de Clasificación.....	30
4 Diseño de Experimentos.....	31
4.1 Dataset	31

4.1.1	Clasificación Manual de las Instancias.....	32
4.1.2	Análisis Descriptivo de los Datos.....	32
4.2	Herramientas utilizadas	35
4.3	Clasificadores	35
5	Desarrollo de Experimentos y Resultados.....	36
5.1	Resultados con BOW	37
5.2	Resultados con BOW Removiendo Stopwords.....	38
5.3	Resultados con BOW Aplicando Stemming	39
5.4	Evaluación de Resultados Utilizando Jerarquías.....	40
5.4.1	Resultados Sin Jerarquía.....	40
5.4.2	Resultados Jerarquía I.....	41
5.4.3	Resultados Jerarquía II	42
5.4.4	Resumen de Resultados Jerárquicos.....	43
6	Conclusiones.....	44
	Referencias	47
	Anexos	
A	Resultados Detallados.....	
A.1	Bag of Words.....	
A.2	Bag of Words Removiendo Stopwords	
A.3	Bag of Words Aplicando Stemming.....	
B	Extracto de Tweets por Categoría.....	
B.1	Reporte de Noticia	
B.2	Publicidad	
B.3	Opinión de Noticia.....	
B.4	Opinión General.....	
B.5	Mensaje Personal	
B.6	Compartir Ubicación / Evento	
B.7	Chat	
B.8	Pregunta	

Resumen

Las redes de micro-blogging son un servicio de comunicación entre usuarios que permite escribir mensajes cortos y transmitirlos a través de la red. Twitter es la red social de micro-blogging más popular en la actualidad, y carece de herramientas que permitan filtrar o clasificar los mensajes, por lo que el usuario recibe un flujo de mensajes con mucho ruido en el contenido. Este trabajo aborda la definición de una taxonomía que facilita la clasificación automática de mensajes en Twitter utilizando un enfoque de maquinas de aprendizaje.

Palabras-clave: Taxonomía, intención, Twitter, clasificación de textos cortos.

Abstract

Micro-blogging networks are a communication service between users that allows them to write short messages and broadcast them throw the network. Twitter is the most popular micro-blogging social network today, and it lacks of tools that allows to users to filter or classify their messages, so the user gets a message stream with a lot of noise into the content. This thesis propose the definition of a taxonomy that ease the automatic classification of messages in Twitter, using a machine learning approach.

Keywords: Taxonomy, intent, Twitter, short text classification.

Lista de Figuras

Figura 2.1 Visión general de los procesos en la clasificación de textos [6]	6
Figura 2.2 Hiperplano de margen máximo.....	11
Figura 2.3 Definición de hiperplano SVM	12
Figura 2.4 Hiperplano con margen blando	13
Figura 2.5 Caso linealmente no separable	14
Figura 2.6 Promedio de edad usuarios de Twitter [10]	16
Figura 2.7 Principales países que usan Twitter [10].....	17
Figura 2.8 Tipo de mensajes en Twitter [11]	18
Figura 2.9 Idiomas de los mensajes en porcentaje [11].....	18
Figura 3.1 Diagrama Jerarquía I	27
Figura 3.2 Diagrama Jerarquía II.....	28
Figura 5.1 Gráfico de exactitud (<i>accuracy</i>) con bag of words.	37
Figura 5.2 Gráfico de precisión por categorías.	37
Figura 5.3 Gráfico de exactitud con BOW removiendo stopwords.	38
Figura 5.4 Gráfico de precisión por categorías con BOW removiendo stopwords.....	38
Figura 5.5 Gráfico de exactitud con BOW aplicando stemming.....	39
Figura 5.6 Gráfico de precisión por categorías con BOW aplicando stemming.	39
Figura 5.7 Comparación de exactitud utilizando jerarquías	40
Figura 5.8 Gráfico comparativo SVM y Naive Bayes (Jerarquía I).....	41
Figura 5.9 Gráfico comparativo SVM y Naive Bayes (Jerarquía II)	42
Figura 5.10 Gráfico comparativo utilizando jerarquías y sin utilizar jerarquía (SVM) ..	43

Lista de Tablas

Tabla 2.1 Matriz de confusión.....	15
Tabla 3.1 Tabla comparativa entre categorías.....	24
Tabla 3.2 Ejemplo de tweets en la categoría Reporte de Noticia.....	24
Tabla 3.3 Ejemplo de tweets en la categoría Opinión de Noticia.....	24
Tabla 3.4 Ejemplo de tweets en la categoría Publicidad.....	25
Tabla 3.5 Ejemplo de tweets en la categoría Opinión General.....	25
Tabla 3.6 Ejemplo de tweets en la categoría Compartir Ubicación / Evento.....	25
Tabla 3.7 Ejemplo de tweets en la categoría Chat.....	25
Tabla 3.8 Ejemplo de tweets en la categoría Pregunta.....	26
Tabla 3.9 Ejemplo de tweets en la categoría Mensaje Personal.....	26
Tabla 4.1 Cantidad de instancias por categorías.....	31
Tabla 4.2 Cantidad de características por categoría.....	32
Tabla 4.3 Porcentaje de apariciones respecto del total por categoría.....	33
Tabla 5.1 Métricas de clasificadores sin usar jerarquía.....	40
Tabla 5.2 Métricas de clasificadores con Jerarquía I.....	41
Tabla 5.3 Métricas de clasificadores con Jerarquía II.....	42

1 Introducción

Debido a la gran cantidad de documentos en formato digital, y en una búsqueda continua para poder organizarlos de manera rápida y flexible, se ha generado un área de investigación en torno a la clasificación automática de los textos. Una fuente de información que requiere de herramientas que permitan filtrar y organizar su contenido de manera automática, son las redes de micro-blogging.

Las redes de micro-blogging son un sistema de comunicación muy popular en los últimos años. Este tipo de servicios se caracteriza principalmente porque permite a los usuarios publicar mensajes de texto cortos. Estos mensajes son enviados a los seguidores en tiempo real.

Este trabajo se enfoca en Twitter, uno de los servicios de micro-blogging más popular en la actualidad, con más de 500 millones de usuarios registrados [1] y más de 200 millones de usuarios activos [2]. Una de sus principales características es que limita el tamaño de un mensaje a 140 caracteres máximo.

En Twitter, los usuarios pueden suscribirse para recibir las actualizaciones de estado de otros usuarios. Estos mensajes son enviados en tiempo real y se concentran en un único flujo de información que el usuario recibe. Esto provoca una sobrecarga de información, dificultando el descubrimiento de información relevante. Un usuario puede estar interesado en seguir a otro porque este escribe de algún tópico de interés común, sin embargo, este recibirá todos los mensajes que el otro usuario escriba, a pesar que de estos algunos no estén relacionados al tópico de interés, y además, no existe una manera de poder filtrarlos automáticamente.

Twitter ha sido utilizado como fuente de información en distintos trabajos de investigación. En [3] se utilizó Twitter con el objetivo de monitorear la opinión de la gente mientras se transmitía en vivo el debate presidencial de Estados Unidos. Las estadísticas informaban que el sentimiento de los mensajes favorecían a Obama sobre McCain, quién resultó salir electo presidente. En [4] se investigó la interacción en tiempo real entre un evento objetivo, como un terremoto, en Twitter y se propuso un algoritmo para monitorear los *tweets* y detectar eventos. Además se propuso un modelo probabilístico espaciotemporal para detectar el centro y la trayectoria del evento.

Con el crecimiento continuo de Twitter como red social de micro-blogging, queda claro que es necesario dar solución al problema de la sobrecarga de información, es por eso que en esta tesis se propone definir una taxonomía que contenga las principales intenciones de los mensajes emitidos por los usuarios en Twitter. Además, utilizando un enfoque de máquinas de aprendizaje, se clasificará de manera automática los mensajes de Twitter en las categorías definidas.

1.1 Descripción del Problema

El presente proyecto surge del interés de estudiar el área relacionada con el procesamiento del lenguaje natural y la clasificación de los mensajes emitidos por los usuarios de las redes sociales, específicamente a través de la red de micro-blogging Twitter, que permitirá poder clasificar un mensaje de manera automática.

Uno de los primeros requisitos que surgen para clasificar la intención de los usuarios en los mensajes de Twitter, es la definición de una taxonomía que describa las categorías de las principales intenciones de los usuarios. Esta definición permitirá poder filtrar los mensajes de una manera rápida, fácil y automática para mejorar el filtrado de información y apoyar la toma decisiones.

El uso de máquinas de aprendizaje es fundamental para que la clasificación de los mensajes se realice de manera automática. Es por eso que se debe realizar una comparación de rendimiento de diferentes clasificadores (Máquinas de Soporte Vectorial, Naive Bayes) que son utilizados actualmente en los problemas de clasificación de textos, y así poder definir cuál de estos presenta mejores resultados y se desenvuelve mejor en un problema de clasificación de mensajes cortos, como es el caso de los mensajes que se emiten en Twitter.

Un ejemplo práctico de uso sería que una empresa pueda filtrar todos los mensajes de Twitter en donde se mencione su marca, durante un periodo de tiempo definido, y analizar sólo los mensajes que contengan opinión, para así poder tomar medidas que estime convenientes. Además, podría resolver las inquietudes de los usuarios identificando las preguntas y poder darles una respuesta con mayor rapidez.

La elección de Twitter como fuente de información para este trabajo, se debe a sus características únicas, entre ellas: el enorme volumen generado de información, ya que los usuarios escriben alrededor de 175 millones de mensaje al día. Por otra parte, el mensaje puede contener a lo más 140 caracteres, lo que supone un gran desafío, ya que la intención del usuario se encuentra concentrada en pocas palabras, lo que en conjunto al uso de un lenguaje más informal, resulta en un gran desafío.

1.2 Definición de Objetivos

1.2.1 Objetivo General

Proponer una taxonomía que defina las principales intenciones de los mensajes de Twitter y evaluar el desempeño de los clasificadores automáticos usando un enfoque de máquinas de aprendizaje.

1.2.2 Objetivos Específicos

Investigar y comprender el estado del arte de las técnicas utilizadas para la clasificación automática de textos.

Investigar estudios relacionados a la definición de taxonomías de la intención del usuario en Twitter.

Proponer un modelo utilizando un enfoque de máquinas de aprendizaje para la clasificación automática de los tweets.

Implementar y evaluar el rendimiento de los clasificadores utilizando un conjunto de datos reales.

1.3 Estructura del Contenido

El documento se organiza de la siguiente manera. El Capítulo 2 presenta el marco teórico, que introduce los conceptos utilizados en este trabajo, enfocado principalmente en la clasificación automática de textos, máquinas de aprendizaje y la red social de micro-blogging Twitter, además de describir los trabajos relacionados. El Capítulo 3 se expone la propuesta, se define la taxonomía, y un grupo de dos jerarquías que agrupan las categorías definidas, también se define la etapa de pre-procesamiento y la representación utilizada para los mensajes. Luego, en el Capítulo 4, se diseñan los experimentos que serán desarrollados y se expondrán los resultados en el Capítulo 5. Finalmente se presentan las conclusiones y el trabajo futuro que podría ser realizado.

2 Marco Teórico

En este capítulo se presentan los conceptos utilizados durante este trabajo. Se introduce a la clasificación automática de textos y además se presenta una visión general de la red social micro-blogging Twitter. Además se describen los estudios relacionados que conforman la base para la definición de la taxonomía de este trabajo.

2.1 Clasificación Automática de Textos

2.1.1 Introducción

La clasificación de textos, también conocida como categorización de textos, corresponde a la aplicación de etiquetas a textos escritos en lenguaje natural, a partir de un conjunto predefinido de categorías [5]. Dentro de algunas de las aplicaciones que abarcan la clasificación de textos se encuentran: indexación automática de documentos, filtrado de documentos, desambiguación de palabras y en general todo lo relacionado con la organización de documentos.

En sus comienzos, la clasificación automática de textos, era abarcada utilizando un enfoque basado en el conocimiento de expertos llamado ingeniería del conocimiento (Knowledge Engineering, KE), donde estos manualmente definían reglas para clasificar los documentos en una u otra categoría. Con el paso del tiempo este enfoque fue reemplazado por uno basado en las máquinas de aprendizaje (Machine Learning, ML). Las principales ventajas que presentaba este enfoque en comparación con el KE era que presentaba mejor precisión en comparación al conocimiento de los expertos y además prescindía de estos para construir los clasificadores.

2.1.2 Definición de Clasificación de Textos

Sebastiani [5] define la clasificación de textos como la tarea de asignar un valor booleano a cada par $\langle d_j, c_i \rangle \in D \times C$, donde D es un dominio de documentos y $C = \{c_1, \dots, c_{|C|}\}$ corresponde a un conjunto predefinido de categorías. Si a $\langle d_j, c_i \rangle$ se le asigna un valor verdadero T, indica la decisión de clasificar el documento d_j bajo la categoría c_i , por otra parte asignar un valor falso F indica la decisión de no clasificar d_j bajo c_i .

Definir la clasificación de textos de una manera más formal, corresponde a la tarea de aproximar la función objetivo desconocida $\check{\phi} : D \times C \rightarrow \{T, F\}$, que describe como un documento debe ser clasificado, por medio de una función $\phi : D \times C \rightarrow \{T, F\}$ llamado el clasificador, tal que $\check{\phi}$ y ϕ “coincidan lo más posible”.

Las categorías corresponden solo a etiquetas simbólicas, no existe conocimiento adicional de sus significados para ayudar a construir el clasificador. Además sólo se utiliza el texto del documento para poder clasificar estos en las categorías, no haciendo uso de algún tipo de conocimiento exógeno, como por ejemplo metadatos con información como la fecha de publicación, tipo de documento, fuente de publicación, entre otros.

En la Figura 2.1 se puede observar una visión general de los elementos y procesos necesarios para la clasificación automática de textos.

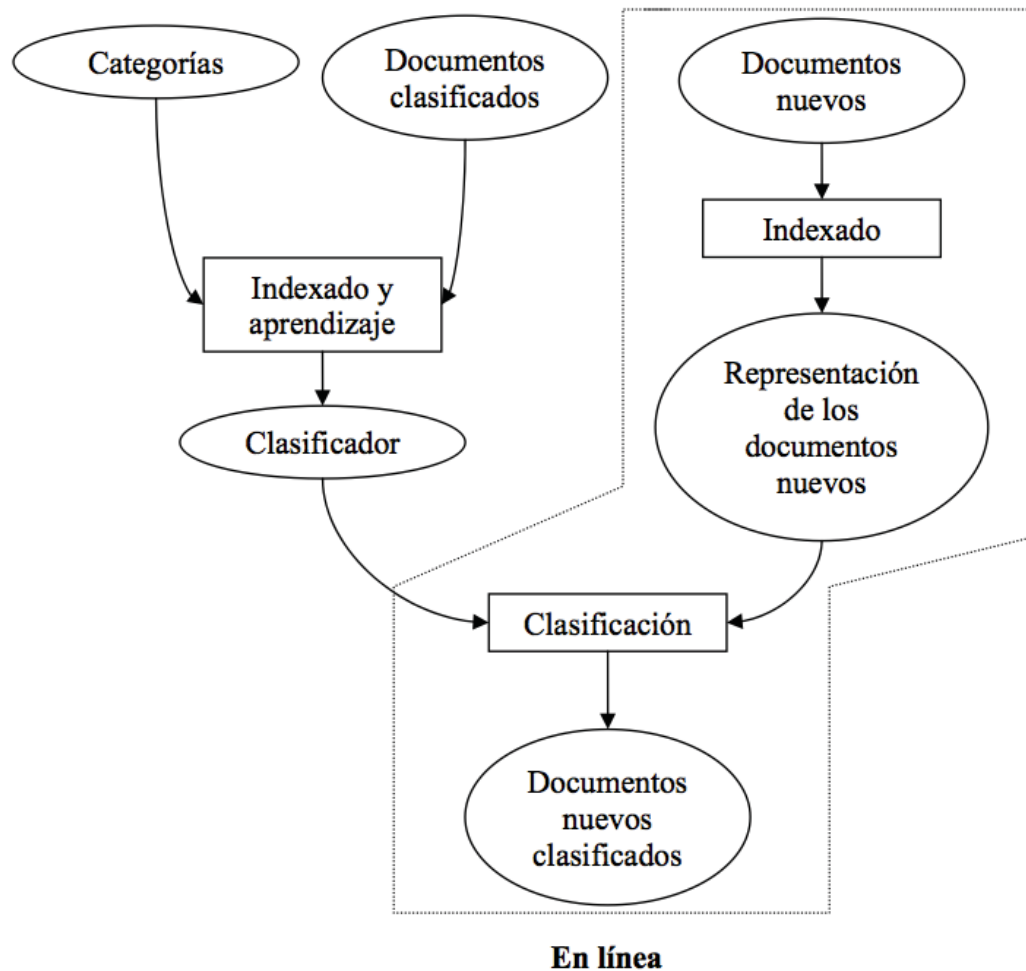


Figura 2.1 Visión general de los procesos en la clasificación de textos [6]

2.1.3 Representación de Documentos y Pre-procesamiento

2.1.3.1 Modelo de Espacio Vectorial

El modelo de espacio vectorial (Vector Space Model, VSM) o también conocido como bolsa de palabras (Bag of Words, BOW) representa cada documento como un vector. Cada vector representa un documento, donde cada componente de este corresponde a un término en el documento. Normalmente se ignora el orden de estas en el texto, por lo tanto, se pierde información gramatical.

Se puede considerar a D_i como el conjunto de documentos que contiene i documentos:
 $D_i = (d_{i1}, d_{i2}, \dots, d_{in})$

Cada dimensión en este vector corresponde a un índice del término d_i . El valor de cada componente dependerá del método utilizado para representar el peso de cada término. A continuación se describen cuatro formas de asignar pesos a las componentes.

Representación Binaria

En la representación binaria, también conocida como Presencia de Términos (Term Presence, TP), dado un conjunto D de documentos y un conjunto T de términos, el peso wd_{ij} del término t_i en el documento d_j corresponde a:

$$wd_{ij} = \begin{cases} 1, & \text{si } t_i \text{ aparece en } d_j \\ 0, & \text{en otro caso} \end{cases} \quad (2.1)$$

Por ejemplo, si se tiene los documentos:

- Documento A: “Hola Mundo”
- Documento B: “Chao Mundo”
- Documento C: “Hola a Todos”

Incluyendo en el vector las palabras *hola*, *mundo*, *chao*, *todos*, en el orden especificado, la representación binaria de estos documentos sería:

- $D_1 = (1,1,0,0)$
- $D_2 = (0,1,1,0)$
- $D_3 = (1,0,0,1)$

Term Frequency

La frecuencia (tf_{ij}) de un término (Term-Frequency, TF) i en un documento j se define como:

$$tf_{ij} = \frac{f_{ij}}{\max_i (f_{ij})} \quad (2.2)$$

Donde f_{ij} es el número de veces que el término i aparece en el documento j . La frecuencia es normalizada por la frecuencia del término más común en el documento.

Inverse Document Frequency

La frecuencia inversa del documento (Inverse Document Frequency, IDF) se usa para indicar el poder discriminativo de un término i . En general, términos que aparecen en muchos documentos distintos son menos indicadores de una categoría específica. Se define la frecuencia inversa del documento (idf_i) del término i como:

$$idf_i = \log_2\left(\frac{n}{df_i}\right) \quad (2.3)$$

Donde df_i es la frecuencia de documentos del término i lo que corresponde al número de documentos en el que aparece el término i . Se utiliza \log_2 para amortiguar el efecto relativo a tf_{ij} .

Term Frequency – Inverse Document Frequency

En Term Frequency – Inverse Document Frequency (TF-IDF) el proceso para calcular el peso w_{ij} requiere de la frecuencia del término i en el documento j , y la frecuencia inversa en el documento del término i .

El peso w_{ij} se calcula utilizando la medida TF-IDF que se define como:

$$w_{ij} = tf_{ij} \times idf_i \quad (2.4)$$

Si el término ocurre con mayor frecuencia en el documento j , pero raramente en el resto de los documentos, entonces se le asigna un valor alto.

2.1.3.2 Stop Words

Stop words se define como términos que se consideran irrelevantes respecto al tema principal del documento, además pueden aparecer frecuentemente en los documentos. Estas palabras incluyen determinantes, conjunciones, preposiciones, y su eliminación constituye uno de los pasos llevados a cabo comúnmente en la etapa de pre-procesamiento.

2.1.3.3 Stemming

Las palabras que aparecen en los documentos a menudo tienen muchas variantes morfológicas. Por lo tanto, se le realiza un proceso de normalización lingüística, en el cual las diferentes formas que puede adoptar una palabra son reducidas a una única forma común llamada *stem* o lema, el cual no necesita tener un significado. Ejemplo, se obtiene el lema “perr” para las palabras perro, perrera, perrito, perros.

2.1.4 Máquinas de Aprendizaje

En el enfoque de Máquinas de Aprendizaje (Machine Learning, ML), tal como se describe en [5], un proceso general inductivo (que se conoce como el *learner*) automáticamente construye un clasificador para una categoría c_i observando las características del conjunto de documentos que han sido previamente clasificados de manera manual bajo c_i o \bar{c}_i por un experto; a partir de estas características, el proceso inductivo extrae las características que un nuevo documento debe tener para ser clasificado bajo c_i .

Según este enfoque, el problema de clasificación es una actividad de aprendizaje supervisado, debido a que el proceso de aprendizaje es dirigido, o “supervisado”, por el conocimiento de las categorías y de las instancias de entrenamiento que pertenecen a ella.

Los clasificadores construidos en la actualidad utilizando técnicas ML, alcanzan buenos niveles de efectividad, haciendo la tarea de clasificación automática una alternativa a la clasificación manual.

2.1.4.1 Conjunto de Entrenamiento, Prueba y Validación.

El enfoque ML se basa en la existencia de un conjunto inicial de documentos previamente clasificados $\Omega = \{d_1, \dots, d_{|\Omega|}\}$ bajo el mismo conjunto de categorías $C = \{c_1, \dots, c_{|C|}\}$. Esto significa que los valores de la función $\check{\phi} : D \times C \rightarrow \{T, F\}$ se conocen para cada par $\langle d_j, c_i \rangle \in \Omega \times C$, donde $\Omega \subset D$. Un documento d_j se llama un *ejemplar positivo* de c_i si $\check{\phi}(d_j, c_i) = T$, y *ejemplar negativo* de c_i si $\check{\phi}(d_j, c_i) = F$.

Una vez que el clasificador se construye, es necesario evaluar su efectividad. En este caso, antes del proceso de clasificación, el conjunto inicial de documentos se divide en dos conjuntos, no necesariamente del mismo tamaño:

- a) Un conjunto de entrenamiento (y validación) $T_v = \{d_1, \dots, d_{|T_v|}\}$. Este corresponde al conjunto de documentos del que se observan las características con la que se construyen los clasificadores para las distintas categorías.
- b) Un conjunto de prueba $T_e = \{d_{|T_v|+1}, \dots, d_{|\Omega|}\}$. Este corresponde al conjunto de documentos que se utilizará para probar la efectividad del clasificador. Cada documento en T_e alimentará al clasificador, y la decisión del clasificador $\phi(d_j, c_i)$ se comparará con la decisión del experto $\check{\phi}(d_j, c_i)$; una medida de efectividad del clasificador estará basada en cuanto se igualan los valores $\phi(d_j, c_i)$ y $\check{\phi}(d_j, c_i)$.

Los documentos en T_e no pueden participar del proceso inductivo de construcción de los clasificadores; si esta condición no se cumpliera, los resultados obtenidos probablemente serían demasiado buenos, y la evaluación no tendría un carácter científico.

2.1.4.2 Construcción Inductiva de Clasificadores de Textos

La construcción inductiva de un clasificador de ranking para la categoría $c_i \in C$ usualmente consiste en la definición de una función $CSV_i: D \rightarrow [0,1]$ que, dado un documento d_j , retorna un estado de categorización para este, o sea un número entre 0 y 1 que representa la probabilidad para el hecho que d_j se debe clasificar bajo c_i .

La construcción de un clasificador “duro” se realiza mediante dos alternativas. El primero consiste en la definición de una función $CSV_i: D \rightarrow \{T, F\}$. El segundo consiste en definir una función $CSV_i: D \rightarrow [0,1]$, análogo a la usada para la clasificación por ranking, además de definir un umbral τ_i tal que $CSV_i(d_j) \geq \tau_i$ se interpreta como verdadero T , y cuando $CSV_i < \tau_i$ se interpreta como falso F .

2.1.4.3 Naive Bayes

El clasificador Naive Bayes es un clasificador probabilístico que se basa en aplicar el teorema de Bayes, que asume la independencia de las variables, es por eso que se le llama clasificador bayesiano ingenuo (*naive*).

Como se explica en [7], desde el punto de vista de la clasificación de textos se puede decir que se asume la independencia de las palabras, es decir, la probabilidad condicional de una palabra dada una clase se asume que es independiente de la probabilidad condicional de otras palabras dada esa clase.

Definiendo más formalmente, sea $\{1 \dots K\}$ el conjunto de clases posibles y $\{x_{i,1}, \dots, x_{i,m}\}$ el conjunto de valores de las características del ejemplo x_i el algoritmo Naive Bayes selecciona la clase que maximiza $P(k|x_{i,1}, \dots, x_{i,m})$:

$$\arg \max_x P(k|x_{i,1}, \dots, x_{i,m}) \approx \arg \max_x P(k) \prod_j P(x_{i,j}|k) \quad (2.5)$$

En donde $P(k)$ y $P(x_{i,1}, K)$ son probabilidades que se estiman a partir del corpus de aprendizaje mediante las frecuencias relativas.

Una de las debilidades conocidas de Naive Bayes es el problema para trabajar en espacios de gran dimensionalidad, es decir, con un gran número de características de aprendizaje [7].

2.1.4.4 Máquinas de Soporte Vectorial

Las máquinas de soporte vectorial (Support Vector Machines, SVM) son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik a mediados de los años 90, utilizados para clasificación y regresión. Es una nueva generación de algoritmos de aprendizaje basados en los avances en la teoría de aprendizaje estadístico, y aplicado a un gran número de aplicaciones del mundo real, como por ejemplo, clasificación de textos, reconocimiento de patrones en imágenes, entre otros [8].

En el caso linealmente separable, la idea clave de una SVM se puede explicar de la siguiente manera. Dado un conjunto de entrenamiento S que contiene puntos de dos clases, una SVM separa las clases a través de un hiperplano determinado por ciertos puntos de S , denominados vectores de soporte. En los casos separables, este hiperplano maximiza el margen, o el doble de la distancia mínima de ambas clases al hiperplano, y todos los vectores de soporte están a la misma distancia mínima desde el hiperplano (y son denominados vectores de margen). Ver Figura 2.2. En casos reales, las dos clases pueden no ser separables y ambos, el hiperplano y los vectores de soporte son obtenidos de la solución de un problema de optimización con restricciones. La solución es una compensación entre el margen mayor y el menor número de errores, compensación controlada por un parámetro de regularización [9].

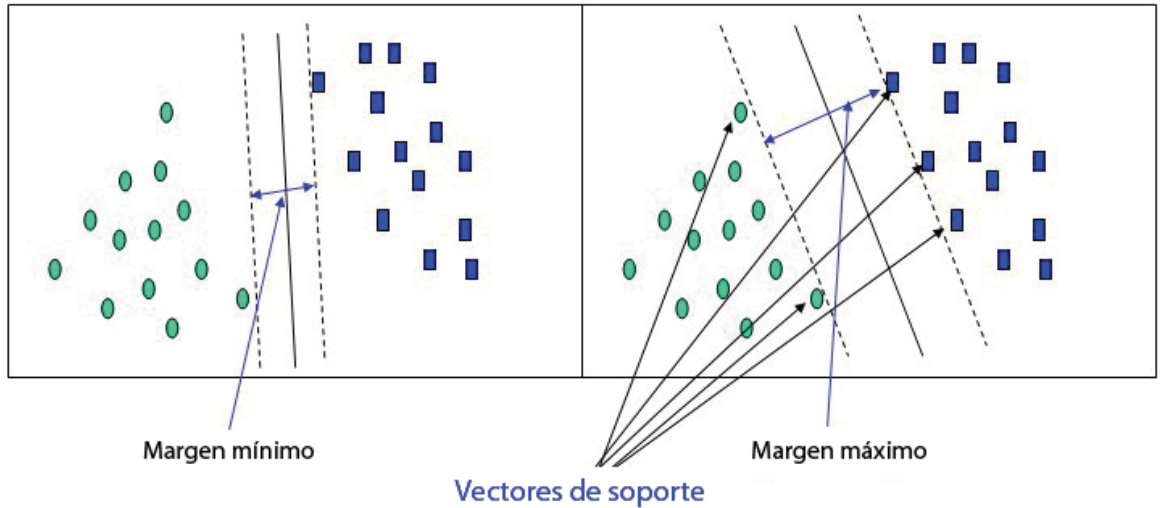


Figura 2.2 Hiperplano de margen máximo.

Dado un conjunto de entrenamiento P de la forma:

$$P = \{(x_i, c_i) \mid x_i \in \{-1, +1\}\}_{i=1}^{|D|} \quad (2.6)$$

Donde x_i corresponde a los vectores de los documentos de entrenamiento d_i y los valores que puede tomar c_i son $+1$ indicando pertenencia en los casos positivos y -1 indicando pertenencia en los casos negativos.

El hiperplano se define como un vector de dimensionalidad $|T| - 1$, donde $|T|$ representan el espacio de dimensión, equivalente al número de términos o palabras existentes. La manera mas simple de separar es a través de una línea recta que cumpla:

$$w \cdot x + b = 0 \quad (2.7)$$

En la Ecuación 2.7 el vector w es un vector normal perpendicular al hiperplano.

Debe satisfacer:

$$wx_i + b \geq 0 \text{ para los casos positivos } c_i = +1 \quad (2.8)$$

$$wx_i + b < 0 \text{ para los casos negativos } c_i = -1 \quad (2.9)$$

Estos pueden ser combinados en un conjunto de desigualdades:

$$y_i(\langle w \cdot x \rangle + b) \geq 1, \forall i \quad (2.10)$$

Al clasificar un nuevo documento, este es evaluado en la ecuación 2.7 y el valor resulta ser ≥ 0 , el elemento es clasificado en la clase positiva, en caso contrario no es clasificado.

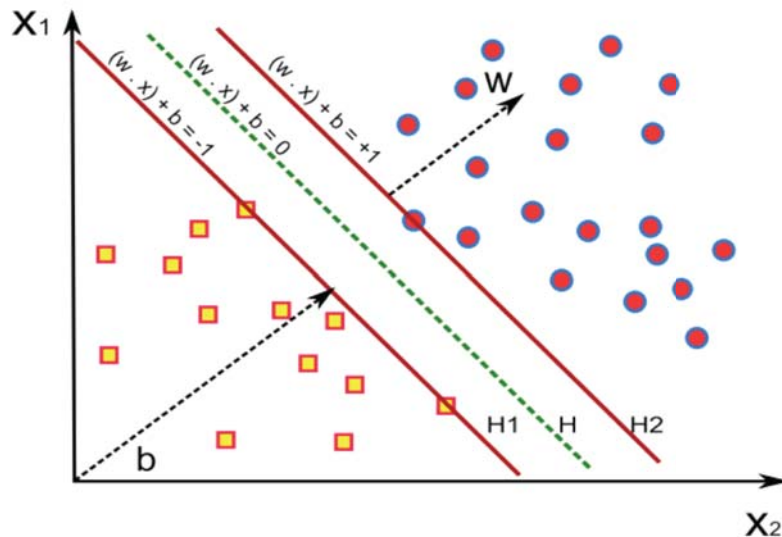


Figura 2.3 Definición de hiperplano SVM

No siempre es posible producir un hiperplano que separe completamente las dos categorías. Por lo tanto, las SVM manejan un parámetro C que controla la compensación entre los errores de entrenamiento y los márgenes rígidos, creando un margen blando (*soft margin*) que permita y a la vez penalice los errores en la clasificación. Ver Figura 2.4.

Es por eso que es necesario introducir las variables flojas no negativas $\xi_i (\geq 0)$ en la ecuación 2.10.

$$y_i(\langle w^T \cdot x \rangle + b) \geq 1 - \xi_i, \forall i \quad (2.11)$$

Mediante las variables ξ_i , la solución factible siempre existe. Para los datos de entrenamiento x_i , si $0 < \xi_i < 1$, los datos no poseen el margen máximo, pero pueden ser correctamente clasificados. Por otro lado, el ancho de este margen blando puede ser controlado por el parámetro de penalización C .

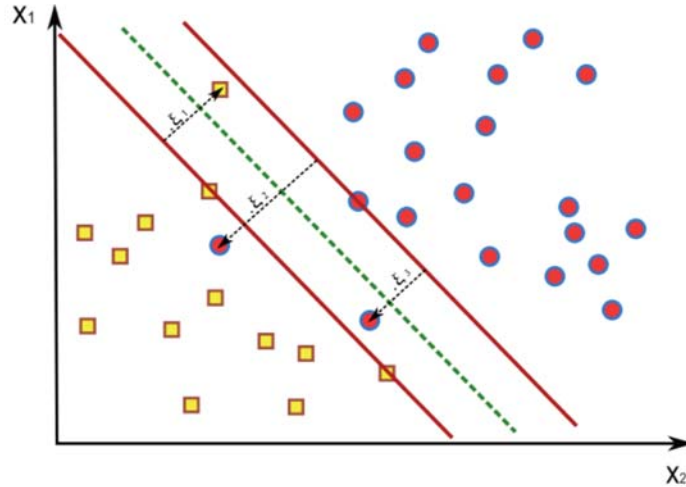


Figura 2.4 Hiperplano con margen blando

Hay casos en los que SVM no puede realizar una separación lineal en un conjunto de elementos, para solucionar esto SVM aplica las denominadas funciones Kernel que permiten trasladar el conjunto de elementos a otro espacio de mayor dimensionalidad llamado “espacio de características”.

Entre las funciones *Kernel* disponibles se tienen:

Kernel polinomial:

$$K(x_i, x_j) = (x_i \cdot x_j)^n \quad (2.12)$$

Kernel gaussiano, radial:

$$K(x_i, x_j) = \exp\left(\frac{-(x_i - x_j)^2}{2\sigma^2}\right) \quad (2.13)$$

Kernel perceptrón:

$$K(x_i, x_j) = \|x_i - x_j\| \quad (2.14)$$

Kernel sigmoidal:

$$K(x_i, x_j) = \tanh(x_i \cdot x_j - \theta) \quad (2.15)$$

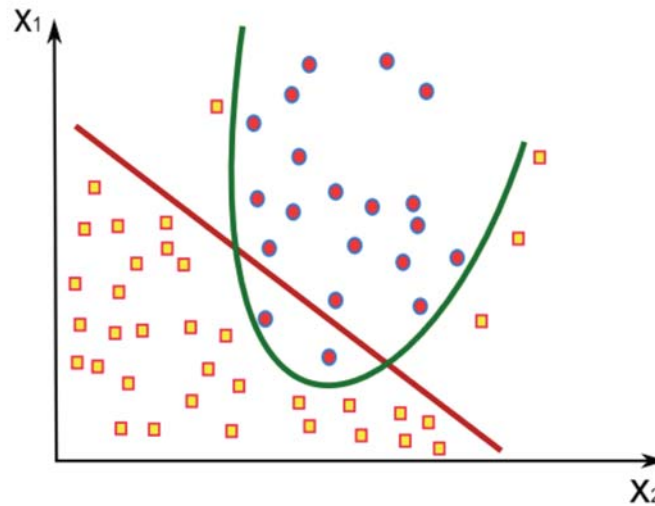


Figura 2.5 Caso linealmente no separable

SVM es un clasificador para dos clases, en caso de realizar clasificación multi-clase se pueden generar K clasificadores, ya sean 1-vs-1 o 1-vs-todos.

2.1.4.5 Medidas de Evaluación

Una vez que el clasificador realiza el proceso de clasificación, es necesario evaluar si esta se realizó correctamente. Para llevar a cabo esta tarea existen distintas medidas entre las cuales se encuentran la exactitud, la precisión, la cobertura y F-measure.

La exactitud (*accuracy*) es la métrica más intuitiva para medir el rendimiento y corresponde a la proporción de clasificaciones correctas frente al total de clasificaciones realizadas.

$$exactitud = \frac{VP + VN}{VP + FP + VN + FN} \quad (2.16)$$

La precisión (π) se entiende como el número de documentos correctamente clasificados entre todos los predichos pertenecientes a la clase:

$$\pi = \frac{VP}{VP + FP} \quad (2.17)$$

La cobertura (*recall*) (ρ) se entiende como el número de documentos que fueron clasificados correctamente del total de los documentos de la clase:

$$\rho = \frac{VP}{VP + FN} \quad (2.18)$$

F-measure (F_β) engloba a la precisión y la cobertura en una sola medida, y se describe como:

$$F_\beta = \frac{(1 + \beta^2)\pi \rho}{\beta^2\pi + \rho} \quad (2.19)$$

Dónde β controla la importancia relativa entre las dos medidas. Usualmente se utiliza $\beta = 1$, indicando igual importancia para la precisión y cobertura.

Los valores expresados en las fórmulas se obtienen de la matriz de confusión como se muestra a continuación:

Tabla 2.1 Matriz de confusión.

Categoría c_i		Real	
		Positivo	Negativo
Predicción	Positivo	VP_i	FP_i
	Negativo	FN_i	VN_i

2.2 Twitter

2.2.1 Introducción

Las redes de Microblogging son un sistema de comunicación muy popular en los últimos años. Este tipo de servicios permiten a la gente publicar mensajes de texto cortos y compartirlos con un conjunto de usuarios de su misma red. Con una limitación de 140 caracteres para compartir un mensaje, Twitter es en la actualidad la red social de microblogging más popular en Internet.

Según la información recolectada por website-monitoring [10] desde múltiples fuentes en noviembre del año 2012, Twitter cuenta con más de 500 millones de usuarios registrados, de los cuales sólo 140 millones son usuarios activos. Cada día se registran en promedio un millón de nuevos usuarios, lo que equivale a 11 cuentas por segundos aproximadamente.

Con respecto a las cuentas de Twitter, un 25 por ciento no tienen seguidores, y el 40 por ciento de las cuentas no han enviado nunca un mensaje. Por otro lado, el usuario promedio de Twitter tiene al menos 27 seguidores. En cuanto al género, el 45 por ciento de los usuarios de Twitter corresponden a hombres, siendo superado por las mujeres en un 10 por ciento.

En la Figura 2.6 se puede observar la edad promedio de los usuarios, en donde se aprecia que la mayoría de los usuarios corresponde a gente joven.

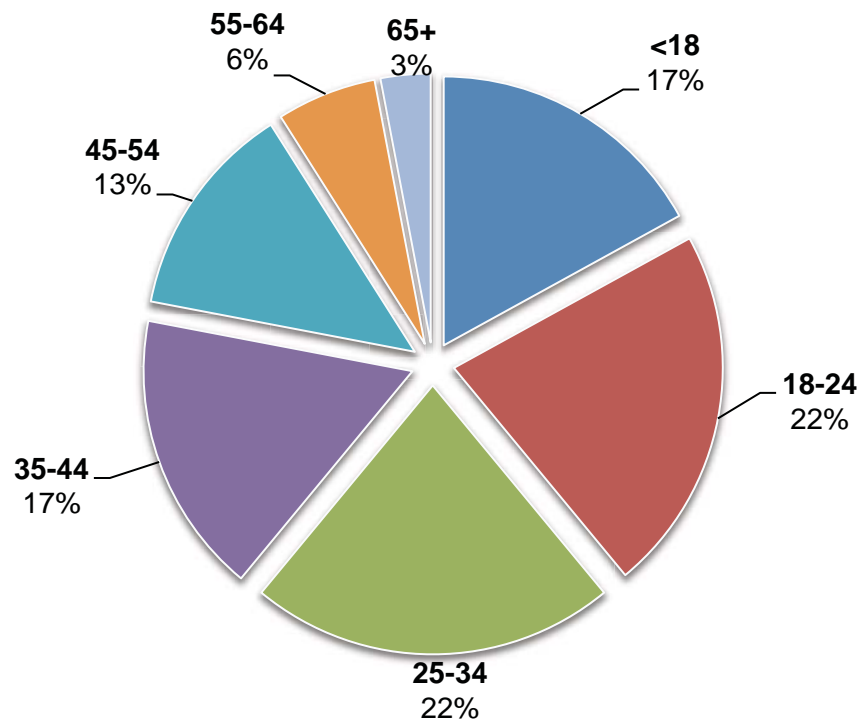


Figura 2.6 Promedio de edad usuarios de Twitter [10]

En la Figura 2.7 se observan los primeros 10 países en el ranking de uso de Twitter. Se puede apreciar que Estados Unidos lidera el uso de Twitter con casi un 51%, seguido por el Reino Unido (17,09%) y por Australia (4,09%).

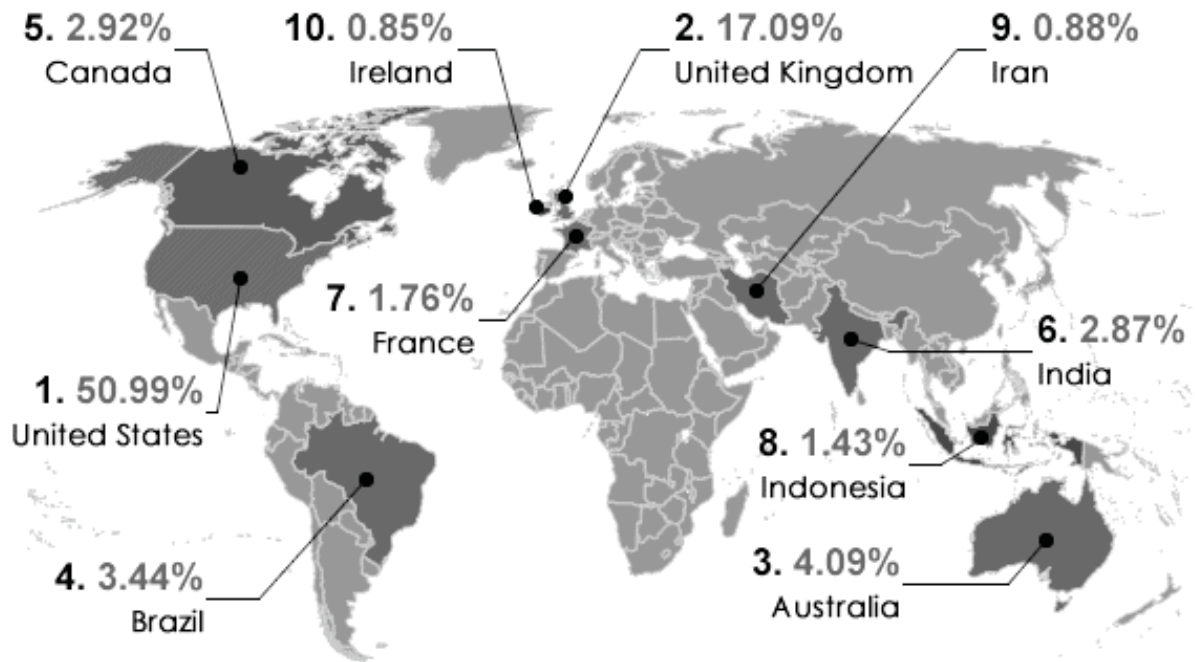


Figura 2.7 Principales países que usan Twitter [10]

Con respecto a los tweets, en un día de alta actividad, Twitter procesa cerca de 175 millones de tweets. En promedio se comparten alrededor de 750 tweets por segundo. El buscador de Twitter procesa mensualmente más de 24 billones de consulta, lo que es más que las búsquedas realizadas en Bing (4,1 billones) y Yahoo! (9,4 billones) juntas.

Según las estadísticas recabadas por [11] en el año 2010, el 31% de los mensajes que se publican en Twitter corresponde al estado actual de los usuarios, seguido por un 27% de conversaciones privadas. Un 10% de los mensajes enlaza a sitios de noticias o artículos de blogs, un 3% enlaza a imágenes y videos y otro 3% enlaza a contenido web de otro tipo. En la Figura 2.8 se muestra con más detalle los tipos de mensajes que se realizan en Twitter.

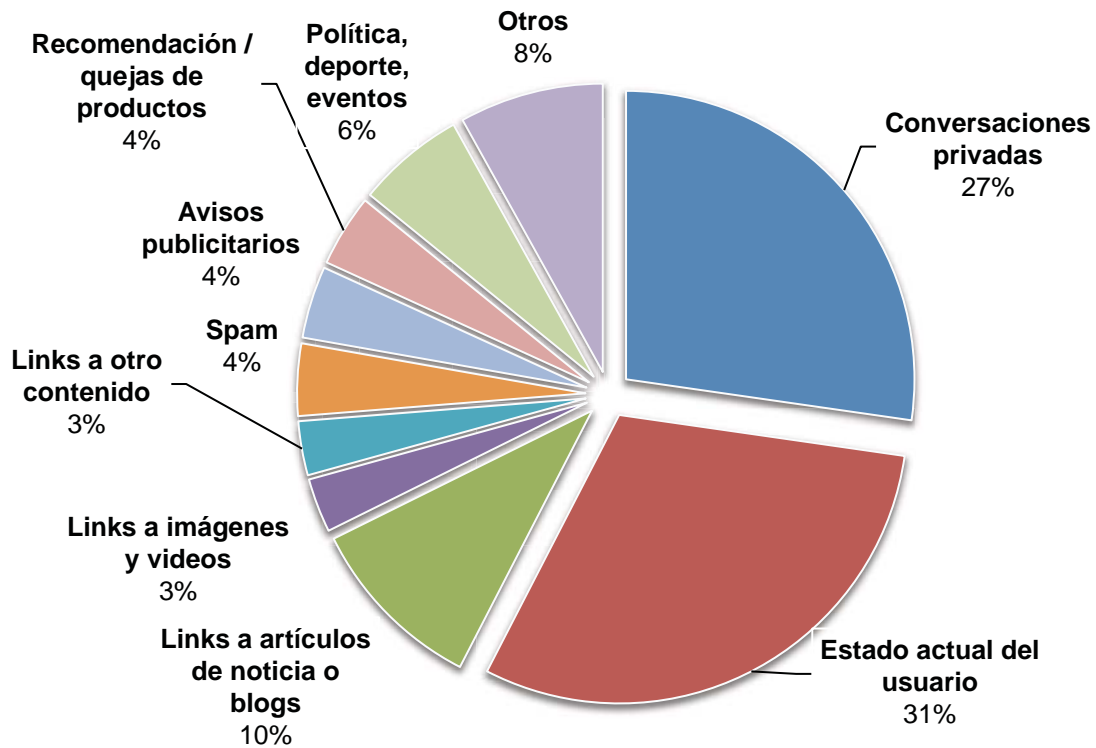


Figura 2.8 Tipo de mensajes en Twitter [11]

Los mensajes que se generan en Twitter son en su mayoría en inglés, seguidos por el portugués y español. En la Figura 2.9 se muestra un gráfico que muestra con mayor detalle esta información.

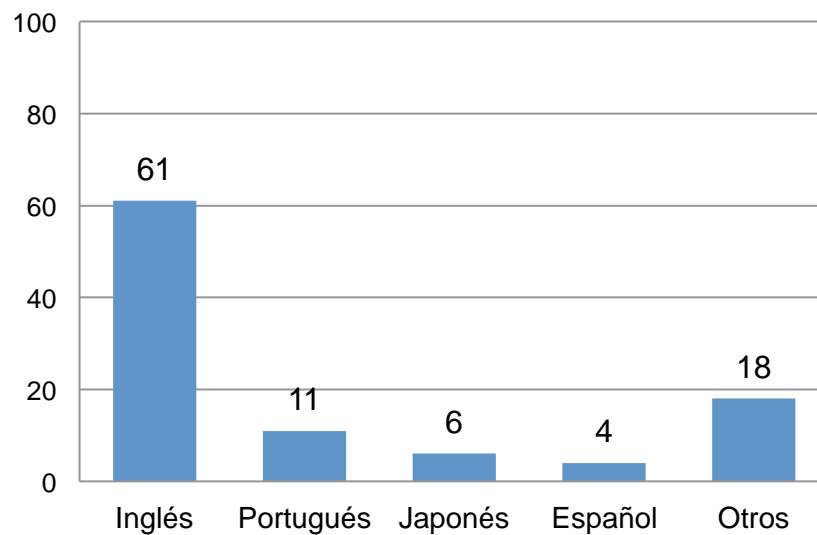


Figura 2.9 Idiomas de los mensajes en porcentaje [11]

2.2.2 Características de Twitter

Twitter presenta una serie de características que la hacen única. A continuación se describen sus características más importantes.

- **Seguidores**

Los seguidores (*followers* en inglés), corresponden a un conjunto de usuarios que deciden formar parte de la red de un usuario, es decir, cada vez que un usuario escriba un mensaje, todos sus seguidores serán informados de esta actualización de estado. Cabe señalar que no es necesario que ambos usuarios sean seguidores mutuamente, por lo tanto existe una relación asimétrica de amistad.

- **Tweets**

El término *tweet* se refiere a un mensaje publicado en Twitter. Durante esta tesis se usará indistintamente el término mensaje y tweet, ambos representarán un mensaje publicado en Twitter por un usuario.

- **Mencionar Usuarios**

En Twitter se puede mencionar a un usuario en un mensaje anteponiendo el símbolo “@” al nombre de este. Cuando se nombra a un usuario al comienzo del mensaje es común asumir que el mensaje va dirigido a él, por lo tanto existe una diferencia al mencionar a un usuario dentro de un mensaje y realizarlo al comienzo del este.

- **Retweet**

Realizar *retweet* corresponde a la acción de re-publicar un mensaje escrito por otro usuario. Además de utilizar la opción que provee Twitter para realizar retweet, es común que la gente anteponga RT a un mensaje escrito por otra persona, pudiendo además añadir algún comentario respecto de dicho mensaje, ya sea al comienzo o al final del mensaje.

- **Hashtag**

Debido a que no existe una manera en Twitter para categorizar un mensaje en base a etiquetas, los usuarios crearon una solución *ad hoc*. Cuando se desea clasificar mensajes relacionados, se utiliza un término preferiblemente corto anteponiéndole el signo “#”. Ejemplos de Hashtags: #SVM, #fb, #futbol, #política, etc

2.3 Trabajos Relacionados

En esta sección se describen algunos trabajos relacionados con esta investigación. Estos estudios abarcan la definición de taxonomías que definen de la intención de los mensajes en Twitter, y resultados obtenidos en la clasificación de mensajes de Twitter para mejorar el filtrado de información.

Why We Twitter: Understanding Microblogging Usage and Communities. [12]

Java *et al.* presentaron observaciones del uso de Twitter a partir del estudio de las propiedades topológicas y geográficas de Twitter. En cuanto a la distribución geográfica, la mayoría de los usuarios de Twitter se encuentran en América del Norte, seguidos por Europa, Asia, Oceanía y América del Sur.

Por otra parte, el estudio propuso un marco de trabajo para detectar la intención de los usuarios. Utilizando el algoritmo HITS, que fue desarrollado inicialmente para realizar el ranking de las páginas webs, se localizan los hubs y las autoridades en la red. A partir de esto se identifican tres categorías de intención de los usuarios: compartir información, buscar información y relación con amigos.

Finalmente, se propuso una breve taxonomía que identifica la intención de los usuarios en Twitter, estas categorías corresponden a: a) Charla Diaria: la mayoría de los usuarios en Twitter hablan de la rutina diaria o lo que la gente se encuentra haciendo actualmente; b) Conversación: corresponde a nombrar a usuarios al inicio de un mensaje, haciendo alusión a que el mensaje va dirigido a un usuario en particular, tornando este mensaje en una conversación; c) Compartir Información (enlaces): cerca del 13% de los mensajes analizados contenían enlaces en ellos, debido a la restricción de tamaño en los mensajes es común usar acortadores de enlaces en los mensajes para compartirlos; y por último, d) Reportar Noticias: muchos usuarios reportan noticias recientes o comentan acerca de estos eventos en Twitter, existen incluso agentes automatizados que publican noticias en Twitter.

Is it Really About Me?: Message Content in Social Awareness Streams. [13]

Naaman *et al.* examinaron las características de las actividades sociales y patrones de comunicación en Twitter. Usando mensajes de un total de 350 usuarios de Twitter y aplicando un análisis cuantitativo de los datos pretendían comprender la actividad de los individuos en esta red.

Los autores definieron el tipo de comunicación que ocurre en Twitter como flujos de conciencia o conocimiento social (Social Awareness Stream, SAS). Estos flujos tienen como característica 3 factores que los distinguen de otros tipos de comunicaciones: a) la naturaleza pública de la comunicación o conversación; b) la brevedad del contenido del mensaje; y, c) un espacio social altamente conectado.

En la investigación se definieron nueve categorías que se utilizaron para clasificar los últimos 10 mensajes de 350 usuarios, estas son: compartir información (*Information Sharing, IS*), auto promoción (*Self Promotion, SP*), Opinión/Quejas (*Opinion/Complaints, OC*), estado y pensamientos (*Statement and Random Thoughts, RT*), yo ahora (*Me Now, ME*), preguntas a

los seguidores (*Question to Followers, QF*), mantener presencia (*Presence Maintenance, PM*), anécdota personal (*Anecdote <me>, AM*) y anécdota de terceros (*Anecdote <others>, AO*).

Los resultados mostraban que el 40% de los mensajes son clasificados como ME (ej. “estoy viendo televisión”) seguido por un 20% en cada una de las categorías IS, OC y RT (cada mensaje podía ser clasificado en más de una categoría).

Este estudio deja en claro que es importante entender estos nuevos sistemas de comunicación, y que a la fecha, no se ha desarrollado un conocimiento profundo en torno a estos, pero si es de gran importancia en el ámbito de la investigación.

What do People Ask Their Social Networks, and Why?: A Survey Study of Status Message Q&A Behavior. [14]

Morris *et al.* exploraron el fenómeno de usar los mensajes en las redes sociales para realizar preguntas. Para esto realizaron una encuesta a 624 personas, pidiéndoles que compartieran las preguntas que han realizado y respondido en las redes sociales.

Más del 50% de los encuestados respondió que habían utilizado las redes sociales para obtener información realizando preguntas. Los encuestados además dieron ejemplos de preguntas que habían realizado o respondido en las redes sociales como Twitter, que debido a tener un límite en el largo de los mensajes, corresponden a preguntas concisas. El 80% de las preguntas fueron realizadas explícitamente y contenían el signo de interrogación. Las preguntas que no forman parte de ese porcentaje incluían frases como “me pregunto si ...”, “necesito ...” incluidas en por ejemplo “necesito una recomendación para comprarme un celular”.

La mayoría de las preguntas se dirigen implícitamente a la red completa de seguidores del usuario que la realiza (ej. ¿Debería reemplazar mi BlackBerry por un iPhone?). Incluso el 20% de las preguntas incluía la palabra “Alguien”, para indicar explícitamente que necesita una respuesta de cualquiera en su red de amigos.

Además en el estudio se identificó el tipo de preguntas que se realizan en las redes sociales. Las preguntas mas populares buscan una recomendación (29%) u opinión (22%), o sea, la mayoría busca respuestas subjetivas. Por otra parte un 17% busca respuestas objetivas. También con un 14% las preguntas retóricas buscan abrir un tema de conversación más que encontrar una respuesta como las otras categorías. Finalmente con menores porcentajes se encuentran las categorías invitaciones, favores, conexiones sociales y ofertas.

Finalmente, con relación al tópico de las preguntas, la gente está interesada en preguntar con respecto a tecnología (29%), entretenimiento (17%), hogar y familia (12%), entre otras categorías.

Short Text Classification in Twitter to Improve Information Filtering. [15]

Sriram propuso un conjunto de ocho características para clasificar mensajes de Twitter debido a que el método tradicional de bag-of-words (BOW) presenta limitaciones, ya que los textos cortos no proveen suficientes ocurrencias de palabras. Además definió un conjunto de siete categorías de intención de los mensajes de Twitter, las que incluían: a) noticias neutras,

b) noticias personales, c) opinión de noticias, d) opiniones, e) ofertas, f) eventos y g) mensajes privados, debido a la similitud de las categorías noticias neutras y noticias personales, y entre opiniones y opinión de noticias, estas fueron agrupadas en noticias y opiniones respectivamente.

Las ocho características (8F) que se definen son: información del autor, acortadores de palabras, información de tiempo y evento, opiniones, énfasis en las palabras, información estadística, información de dinero, nombrar a un usuario en el mensaje.

Durante la etapa de selección de tweets se eliminaron los mensajes que no estaban escritos en inglés, que contenían pocas palabras, y además, se eliminaron los mensajes que estaban compuestos sólo de un enlace (sin texto adicional).

Los algoritmos de clasificación utilizados fueron Naive Bayes, C4.5 y SMO (Sequential Minimal Optimization). Se comparó el rendimiento de los clasificadores utilizando como representación bag-of-words, bag-of-words en conjunto con las ocho características y sólo utilizando las características. Los mejores resultados se obtuvieron utilizando sólo 8F. En general Naive Bayes y SMO obtuvieron un rendimiento muy similar, pero SMO presentó mejores resultados cuando se utilizó 8F, por otra parte, la mejora de 8F en comparación con el uso de bag-of-words es mayor utilizando Naive Bayes.

2.3.1 Resumen de Estudios Relacionados

En la sección 2.3 se presentaron cuatro estudios relacionados con el tema principal de esta tesis, que es la definición de una taxonomía que permita establecer las categorías principales de intención de los usuarios en los mensajes de Twitter. A partir de las categorías expuestas en esa sección se puede definir una taxonomía que cubra las áreas relevantes para esta tesis.

El que se esté investigando entorno a este tema, sugiere que es bastante importante realizar más aportes relacionados con la clasificación de mensajes cortos, especialmente en redes sociales como Twitter.

Además de la definición de la taxonomía, es importante utilizar algoritmos de máquinas de aprendizaje que ayuden a la clasificación automática de los mensajes en las categorías definidas, con el objetivo de automatizar el proceso de clasificación y en un futuro realizar una clasificación en tiempo real de mensajes en las redes sociales.

A partir de estos estudios relacionados se desprende la propuesta de taxonomía que se presenta a continuación.

3 Modelo Propuesto

A continuación se presenta la propuesta que abarca cada una de las etapas para enfrentar el problema de clasificación de los mensajes de Twitter de acuerdo a la intención del usuario.

Primero se define la taxonomía que describe las categorías a las cuales puede pertenecer un mensaje dependiendo de la intención del usuario. Esta definición se desprende y es complementada a partir de las taxonomías definidas en estudios relacionados. Además se definen dos jerarquías que agrupan las categorías definidas y también se describe la etapa de pre-procesamiento y representación de los datos.

3.1 Definición de la Taxonomía

Uno de los primeros pasos que se debe realizar en esta tesis es definir una taxonomía que permita clasificar los mensajes de Twitter según la intención del usuario. Para cumplir con esto es necesario definir las categorías o clases que engloben las posibles intenciones de los mensajes. En la sección 2.3 se presentaron estudios relacionados que identificaron categorías de mensajes en Twitter e intencionalidad de los usuarios, que en conjunto forman el punto de partida para la definición de las categorías presentadas en esta sección.

A continuación se presentan las categorías identificadas y su relación con los estudios previos. Se identifican ocho categorías que incluyen las intenciones más comunes de los usuarios en los mensajes de Twitter y que son relevantes para este estudio.

- Reporte de Noticia
- Opinión de Noticia
- Publicidad
- Opinión General
- Compartir Ubicación / Evento
- Chat
- Pregunta
- Mensaje Personal

En la Tabla 3.1 se presenta una comparación entre las categorías propuestas en esta tesis con las identificadas en los estudios presentados en la sección 2.3.

Esta definición busca ser lo más general posible, para así poder ser utilizada en la más amplia variedad de problemas. A diferencia de los estudios previos que proponen categorías muy específicas, en esta propuesta se agrupan las categorías similares como se puede observar en Tabla 3.1. Como ejemplo, en el caso de la categoría Mensaje Personal, que incluye las actividades que una persona realiza a diario (Ejemplo: anécdotas personales, anécdotas de otros, mantener presencia), en la propuesta de Naaman *et al.* son consideradas categorías diferentes, en esta definición se consideran como una única categoría.

Por otra parte, también se ha buscado tener un mayor detalle en el ámbito de la información de noticias, diferenciando cuando una noticia es objetiva, cuando se emiten opiniones sobre estas y además en los casos que estas se consideran promociones.

Tabla 3.1 Tabla comparativa entre categorías

Categorías Propuestas	Categorías de Naaman [13]	Categorías de Java [12]	Categorías de Sriram [15]
Reporte de Noticia	Information Sharing	Reportar Noticia	Noticias
Opinión de Noticia	Information Sharing Opinions/Complaints		Opinión
Publicidad	Self Promotion	Compartir Información (Enlaces)	Ofertas
Opinión General	Opinion/Complaints	-	Opinión
Pregunta	Question to Followers	-	-
Mensaje Personal	Presence Maintenance Me Now Anecdote (me) Anecdote (Others)	Charla Diaria	Noticias
Compartir Ubicación / Evento	-	-	Eventos
Chat	-	Conversación	Mensaje privado

Finalmente se puede observar que esta definición enriquece y unifica las taxonomías definidas en los estudios anteriores, tomando en consideración aspectos que son de uso cotidiano en la red, como por ejemplo compartir un evento o ubicación geográfica con los seguidores, publicidad o promociones y comentar u opinar sobre noticias emitidas por otros usuarios.

3.1.1 Descripción de las Categorías

A continuación se presenta una descripción detallada de las categorías que se han identificado de la intención de los usuarios en los mensajes de Twitter, el objetivo es definir los tweets que serán considerados en cada una de las categorías.

- **Reporte de Noticia (RN):** Un reporte de noticia corresponde a una noticia emitida de manera objetiva. Por lo general emitida por una cuenta corporativa (CNN, Cooperativa, Terra, etc.) y acompañada de un hipervínculo a la noticia completa en el sitio web del informante. En este caso el tweet se encuentra escrito en lenguaje formal.

Tabla 3.2 Ejemplo de tweets en la categoría Reporte de Noticia.

Emisor	Mensaje
Cooperativa	Terremoto en Turquía alcanzó los 6 grados Richter http://bit.ly/IUOUUnL
Cooperativa	Peñarol acabó con el sueño de U. Católica en la Copa Libertadores http://bit.ly/mv1ix9

- **Opinión de Noticia (ON):** A diferencia del Reporte de noticia que es un mensaje objetivo, en la Opinión de Noticia se emite comentario, ya sea positivo o negativo, relativo al tema de la noticia. Esta no debería provenir de una cuenta corporativa ya que las noticias emitidas por estos son realizadas de manera objetiva.

Tabla 3.3 Ejemplo de tweets en la categoría Opinión de Noticia.

Emisor	Mensaje
EdxBigBro	RT @Cooperativa: RN llamó al Gobierno a “aplicar mano dura” a quienes causen destrozos el 21 de mayo http://bit.ly/kJIQ0K // como siempre!

- **Publicidad (PU):** Al igual que los Reportes de Noticias estos provienen de una cuenta corporativa. Usualmente podrían ser considerados spam, debido a que incluyen un hipervínculo al sitio de la oferta y palabras que hacen referencia a esta (ej. Descuento, dcto., gratis, etc.).

Tabla 3.4 Ejemplo de tweets en la categoría Publicidad.

Emisor	Mensaje
Santanderchile	Obtén 10% dcto. en la tasa al contratar tu Crédito de Consumo en Santander.cl. http://bit.ly/jiLIsT

- **Opinión General (OG):** Una opinión general representa el pensamiento del autor del tweet sobre algún tema en particular. A diferencia de la opinión de noticia, este usuario es quien comienza a opinar de un tema y no opinar sobre una noticia emitida a través de un tweet corporativo. Este mensaje suele estar escrito de manera más informal, acortando palabras, destacando sentimientos utilizando mayúsculas, haciendo uso de emoticones y otras técnicas que permitan demostrar opinión.

Tabla 3.5 Ejemplo de tweets en la categoría Opinión General.

Emisor	Mensaje
SoledadOnetto	Los del +56 etc me parece lo más latero. Esta noche no dormiré.
Wefesx	Jajajaja la película del 13 fue muy mala!! Lo único bueno fue ver a liv Tyler jajaja

- **Compartir Ubicación / Evento (CU):** Esta categoría hace referencia a compartir un mensaje que incluya información geográfica de dónde se encuentra el autor. Esto puede ser mediante el uso de servicios que permitan compartir la ubicación del usuario (ej. Foursquare) o informando de donde se encuentra sólo utilizando lenguaje natural (ej. “Me encuentro en...”, etc).

Tabla 3.6 Ejemplo de tweets en la categoría Compartir Ubicación / Evento.

Emisor	Mensaje
Jorge_Galvez	En la conferencia de James Adams (@ Pontificia Universidad Católica de Valparaíso) http://4sq.com/IMROlw
manuguisone	En la conferencia de Enrique Schewach... superó mis expectativas

- **Chat (CH):** Un mensaje perteneciente a la categoría de chat representa la conversación entre uno o más usuarios de Twitter. Esta conversación se manifiesta a través del uso de “@” que permite nombrar a un usuario dentro de un mensaje (ej. @usuario). Por lo general en la conversación se nombra al o los usuarios al comienzo del mensaje.

Tabla 3.7 Ejemplo de tweets en la categoría Chat.

Emisor	Mensaje
santanderchile	@giosalinas Debes solicitarlo en sucursal hasta las 14:00
mauriciomartis	@alberto_holts esa es la idea. No está nada de mal el nombre. Evaluaré la posibilidad.

- **Pregunta (PR):** Dentro de esta categoría pueden aparecer dos tipos de preguntas generalmente realizadas. Pregunta directa, que se realiza a uno o más usuarios específicos haciendo uso de la posibilidad de nombrar a un usuario dentro del mensaje. Pregunta a los seguidores, corresponde a una pregunta abierta a todos los seguidores, esta no posee ningún destinatario especificado como en el caso de la pregunta directa.

Tabla 3.8 Ejemplo de tweets en la categoría Pregunta.

Emisor	Mensaje
santanderchile	@nh_cuevas podemos ayudarte en algo?
SoledadOnetto	Bajando whatsapp. Experiencias que deseen compartir?

- **Mensaje Personal (MP):** Un mensaje personal corresponde a cualquier mensaje que sea compartido por un usuario de Twitter en el cual desee informar de asuntos personales y que no correspondan a ninguna de las otras categorías expuestas en esta taxonomía. Dentro de esta categoría se encuentran mensajes relacionados a la situación sentimental de la persona, información de lo que se encuentra haciendo, compartir algún tipo de anécdota, mantener presencia en la red, entre otros.

Tabla 3.9 Ejemplo de tweets en la categoría Mensaje Personal.

Emisor	Mensaje
SoledadOnetto	Parece el Valle de la Luna pero es Farellones antes del invierno blanco. Que paz! yfrog.com/h4uj6rzj
SoledadOnetto	Cerrada la oficina virtual. Abro mañana a las 8 am. Hora de ver “Hermanos” con Natalie Portman.
mauriciomartis	Viendo 40 y tantos... luego a dormir.

3.2 Clasificación Jerárquica

En esta sección se definen dos jerarquías que agrupan las ocho categorías definidas en la taxonomía. El método para la definición de estas dos jerarquías se explica a continuación.

La primera jerarquía agrupa las categorías según un criterio de similitud de las definiciones de las categorías, o sea, los mensajes que contienen Opinión General y las Opinión de Noticias se agrupan en una la categoría superior Opinión.

La segunda jerarquía agrupa las categorías de manera similar a la primera jerarquía, diferenciándose en que además se utiliza los resultados de la clasificación en base a las ocho categorías, en donde, a partir de la información que otorga la matriz de confusión se agrupan las categorías que parecen ser más similares según las instancias pertenecientes a una clase y que son mal clasificadas en otra clase. Como ejemplo se tiene que los mensajes de las categorías Mensajes Personal, Opinión General y Chat son agrupados en una categoría superior ya que es en donde los clasificadores tienen mayor problema al momento de clasificar. Es esperable que

3.2.1 Jerarquía I

En la Figura 3.1 se presenta la jerarquía definida a partir de la similitud en la definición de las categorías. Se agrupa los Reporte de Noticia y la Publicidad ya que estos son mensajes que se emiten principalmente por cuentas corporativas. Además, se agrupan Opinión de Noticia y Opinión General ya que son mensajes que están enfocados en la opinión. Por otro lado, los Mensaje Personal y Compartir Ubicación son considerados mensajes que el usuario escribe a diario, y finalmente Chat y Pregunta se consideran mensajes sociales.

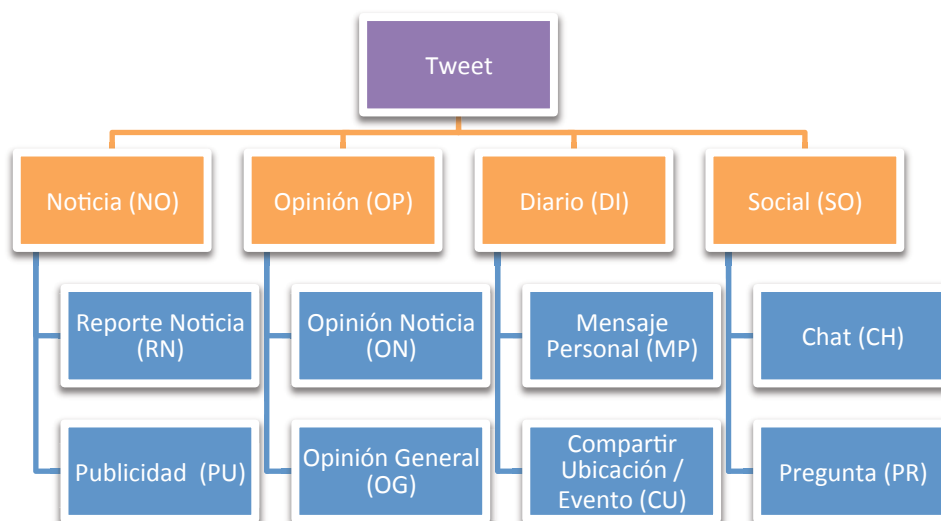


Figura 3.1 Diagrama Jerarquía I

3.2.2 Jerarquía II

En Figura 3.2 se muestra la jerarquía definida tomando en consideración los resultados obtenidos mediante la clasificación de las ocho categorías de la taxonomía, en donde las categorías se agrupan según la similitud que entrega la matriz de confusión.

La agrupación de dos o más categorías está dada por el siguiente criterio: en caso que el clasificador indique que las instancias de la clase B son clasificadas bajo la clase A, entonces estas se agruparán en una categoría superior. La excepción en esta jerarquía está dado por las categorías Compartir Ubicación y Pregunta que son agrupadas por el criterio definido para la jerarquía I.

Es esperable que los resultados obtenidos mediante esta definición sean superiores a los que entregarán los clasificadores utilizando la Jerarquía I. Esto se debe principalmente, a que el criterio de definición utiliza la similitud entre las categorías que entregan los clasificadores, sin embargo, se decide realizar esta definición para evaluar si la clasificación a un nivel más bajo del árbol jerárquico presenta también buenos resultados.

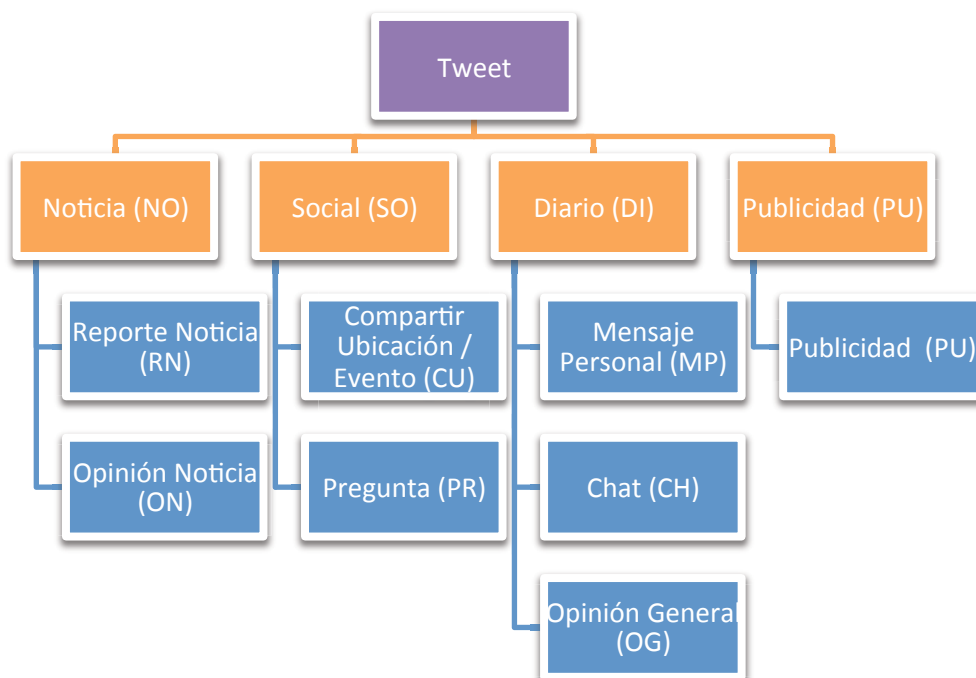


Figura 3.2 Diagrama Jerarquía II

3.2.3 Pre-procesamiento

Una vez que se han recolectado los datos (tweets), es necesario realizar el pre-procesamiento, con el objetivo de poder transformar el texto del mensaje en un vector de características.

Esta fase extrae la parte importante de los mensajes de Twitter, pero es necesario complementar las características comunes de los textos considerando las características específicas del dominio [16], por lo que a continuación se definen cuales son las características de los mensajes de Twitter que se recabarán en la etapa de pre-procesamiento.

Tomando como ejemplo el mensaje “@usuariou ¿Cómo estás? Te he estado llamando como 20 veces y no contestas! :(. Por favor visita este #sitio <http://bit.ly/x35dwd3>.”, a continuación se muestran las etapas de pre-procesamiento:

- **Identificación de Menciones:** En los mensajes es común hacer mención de otros usuarios con el fin de dirigirles un mensaje (chat), opinar sobre ellos o simplemente hacer alusión a ellos.
Ejemplo: “<MENCION> ¿Cómo estás? Te he estado llamando como 20 veces y no contestas! :(. Por favor visita este #sitio <http://bit.ly/x35dwd3>.”
- **Identificación de Emoción:** El uso de emoticonos es una forma de poder buscar polaridad en un mensaje y distinguir con mayor facilidad un mensaje positivo de

uno negativo. Todo emoticono se reemplaza por una palabra clave que no aparece en ningún mensaje de texto.

Ejemplo: “<MENCION> ¿Cómo estás? Te he estado llamando como 20 veces y no contestas! <EMOTICON>. Por favor visita este #sitio <http://bit.ly/x35dwd3>.”

- **Identificación de URL:** Debido a que las URL en los mensajes no entregan mucha información esta se reemplaza por la palabra clave que haga alusión a la URL.

Ejemplo: “<MENCION> ¿Cómo estás? Te he estado llamando como 20 veces y no contestas! <EMOTICON>. Por favor visita este #sitio <URL>.”

- **Identificación de puntuaciones:** Se identifican signos exclamación e interrogación.

Ejemplo: “<MENCION> <PR> Cómo estás <PR> Te he estado llamando como 20 veces y no contestas <EX> <EMOTICON>. Por favor visita este #sitio <URL>.”

- **Identificación de Hashtag:** Se identifican los Hashtag y se reemplazan por una palabra clave.

Ejemplo: “<MENCION> <PR> Cómo estás <PR> Te he estado llamando como 20 veces y no contestas <EX> <EMOTICON>. Por favor visita este <HASHTAG> <URL>.”

- **Eliminación de tildes, conversión a minúsculas y eliminación de signos de puntuación:** Se eliminan los tildes de las palabras y además se convierte el tweet a minúscula y finalmente se eliminan los signos de puntuación.

Ejemplo: “<MENCION> <PR> como estas <PR> te he estado llamando como 20 veces y no contestas <EX> <EMOTICON> por favor visita este <HASHTAG> <URL>.”

3.2.4 Representación

Para la representación de los mensajes se utilizó Bolsa de Palabras (Bag of Words, BOW). El vector de palabras generado contiene además características adicionales, relacionadas con las características específicas del dominio identificadas en la etapa de pre-procesamiento.

Las características adicionales que se utilizan están relacionado con los siguiente:

- a) **Mención de usuarios**
- b) **Uso de emoticonos**
- c) **Uso de URLs**
- d) **Uso de Hashtags**
- e) **Uso de signos de exclamación e interrogación**

La inclusión de estas características adicionales tiene el objetivo de tomar en consideración características específicas del dominio, para que aporten en la clasificación de los mensajes.

3.3 Modelo de Clasificación

A continuación se describe las fases que constituyen el modelo de clasificación:

1. Primero se obtienen los mensajes (tweets) desde Twitter. La construcción de este conjunto de datos se describe en la sección 4.1.
2. Luego se realiza una etapa de pre-procesamiento en donde el conjunto de datos se transforma a un modelo de espacio vectorial (bolsa de palabras) y se realizan procesamiento especiales, como por ejemplo remover stopwords y aplicar stemming. Este proceso se explicó en la sección 3.2.3.
3. Una vez concluida la fase de pre-procesamiento, al vector de palabras generado se le agregan características adicionales, con el objetivo de complementar la bolsa de palabras generada y mejorar así los resultados de los clasificadores. Las características adicionales se describieron en la sección 3.2.4.
4. Finalmente se realiza el proceso de clasificación en donde se comprueba el rendimiento de los clasificadores a través de distintas métricas. Los resultados obtenidos se presentan en la sección 5.

4 Diseño de Experimentos

4.1 Dataset

El conjunto de datos a emplear en esta investigación se creó a partir de 5222 mensajes de Twitter extraídos en distintos periodos de tiempo. Los mensajes obtenidos se encuentran escritos en idioma español y se ha elegido un subconjunto aleatorio para cubrir las categorías de manera homogénea.

En la Tabla 4.1 se puede observar la cantidad de instancias totales por categoría y las cantidades utilizadas.

Tabla 4.1 Cantidad de instancias por categorías

Categoría	Instancias Totales	
	Totales	Utilizadas
RN	2427	300
PU	417	300
ON	250	250
OG	316	300
MP	375	300
CU	659	300
CH	597	300
PR	334	300
Total	5222	2350

Los mensajes recabados para conformar el dataset se obtuvieron de la siguiente manera:

Banco Santander: mensajes emitidos por *@santanderchile* y mensajes que mencionan a *@santanderchile, santander*.

Pontificia Universidad Católica de Valparaíso: mensajes emitidos por *@comunidadpucv* y mensajes que mencionan *PUCV* y *@comunidadpucv*.

El Mercurio, La Tercera, Publimetro: mensajes emitidos por *@emol, @latercera, @laterceracom* y *@publimetrochile*. Mensajes que mencionan *@emol, el mercurio, emol, @latercera, @laterceracom, la tercera, publimetro* y *@publimetrochile*.

Movistar: mensajes emitidos por *@movistarchile* y mensajes que mencionan *@movistarchile* y *movistar*.

Mensajes al azar: un conjunto de mensajes al azar donde se mencionen palabras como: *#recomendar, chile, evento, Quilpué, Quillota, Rancagua, Santiago, Valparaíso, @fernandopaulsen, @biobio, #tecnologia, película, deporte, concierto, ipad, iPhone*.

Mensajes de usuarios al azar: además se extrajo un conjunto de mensajes de usuarios al azar.

4.1.1 Clasificación Manual de las Instancias

Los mensajes del dataset fueron clasificados manualmente bajo las categorías expuestas en la taxonomía descrita en el capítulo 3. Debido a que los mensajes deben pertenecer a una sola categoría, se definió un orden específico para la clasificación de los mensajes. El orden definido es el siguiente: 1) RN, 2) ON, 3) PU, 4) CU, 5) OG, 6) PR, 7) CH Y 8) MP.

La evaluación de los mensajes se hace en base a la definición descrita en el capítulo 3.1.1, o sea, para que el mensaje sea clasificado como RN, debe ser un mensaje de noticia, escrito en lenguaje formal e incluir un hipervínculo. Si el mensaje cumple estas características entonces es clasificado por el experto como un mensaje de Reporte de Noticia, en caso contrario se evalúa si pertenece a la siguiente categoría en la lista de prioridad y así sucesivamente.

Por consecuente, siguiendo esta definición, un mensaje que corresponda a la categoría Chat y que contenga opinión es considerado un mensaje de Opinión General (OG) en lugar de un mensaje de Chat (CH).

4.1.2 Análisis Descriptivo de los Datos.

En esta sección se presenta un breve análisis descriptivo de los 2350 registros que se utilizan en esta investigación.

Número de Características por Categoría

En la Tabla 4.2 se muestra la cantidad total de mensajes que poseen algunas de las características.

Las características que se consideraron corresponden a uso de hipervínculos (URLs), aparición de menciones (Mención), uso de emoticonos (Emoticonos), uso de Hashtags (Hashtag), uso de signos de pregunta (Pregunta) y de signos de exclamación (Exclamación).

Tabla 4.2 Cantidad de características por categoría

	Nº Reg.	URLs	Mención	Emoticonos	Hashtag	Pregunta	Exclamación
RN	300	300	14	0	15	1	0
PU	300	258	40	0	50	49	137
ON	250	80	199	20	30	36	75
OG	300	23	193	29	55	27	98
MP	300	20	54	31	59	12	54
CU	300	297	35	33	16	12	93
CH	300	28	298	33	8	9	110
PR	300	7	228	6	41	294	26
Total:		1013	1061	152	274	440	593

Porcentaje de Características por Categoría

En la Tabla 4.3 se aprecia el porcentaje de mensajes de una categoría que poseen una característica dada.

Tabla 4.3 Porcentaje de apariciones respecto del total por categoría

	URLs	Mención	Emoticonos	Hashtag	Pregunta	Exclamación
RN	100,0%	4,7%	0,0%	5,0%	0,3%	0,0%
PU	84,7%	13,3%	0,0%	16,7%	16,3%	45,7%
ON	86,0%	79,6%	8,0%	12,0%	14,4%	30,0%
OG	7,7%	64,3%	9,7%	18,3%	9,0%	32,7%
MP	6,3%	18,0%	10,3%	19,7%	4,0%	18,0%
CU	99,0%	11,7%	11,0%	5,3%	4,0%	31,0%
CH	9,3%	99,3%	11,0%	2,7%	3,0%	36,7%
PR	2,3%	76,0%	2,0%	13,7%	98,0%	8,7%

URLs

Como se observa en la Tabla 4.3, las categorías Reporte de Noticias (RN), Publicidad (PU), Opinión de noticias (ON) y Compartir Ubicación (CU) son las que más presentan hipervínculos en los mensajes.

En el caso de RN el 100% de los mensajes incluyen enlaces, esto confirma que los mensajes de esta categoría incluyen un enlace que complementa el tweet con un hipervínculo al sitio oficial para obtener mayor información noticiosa.

Los mensajes de la categoría CU incluyen hipervínculos en un 99% de los casos, esto comprueba que los usuarios utilizan servicios como Foursquare (check-in) para compartir su ubicación en un momento dado.

En el caso de ON se puede observar que el 86% de los mensajes incluyen URLs, esto se debe a que el usuario al comentar una noticia incluye en la mayor parte del mensaje la noticia de origen, la cual posee un enlace.

En cuanto a la categoría Publicidad, el 84,7% de los mensajes incluyen un hipervínculo, lo que significa que aparte de ofrecer una publicidad a los seguidores, incluye un enlace con información más detallada de la oferta. Es por eso que los mensajes de esta categoría podrían incluso considerarse Spam.

Finalmente se puede observar que el resto de las categorías casi no se hace uso de enlaces.

Menciones

Con respecto a la utilización de menciones, mensajes de Chat (CH) lidera el ranking con un 99,3%, seguido por ON con un 79,6%, Pregunta (PR) con un 76% y por Opinión General (OG) con un 64,3%.

El porcentaje de la categoría CH era esperable debido a que para emitir mensajes de chat los usuarios deben mencionar a los usuarios dentro de un mensaje.

Por otra parte, la categoría PR con un 76% indica que los usuarios además de realizar preguntas indirectas, en su mayoría realizan preguntas a sus seguidores.

Emoticonos

Los emoticonos son principalmente utilizados en las categorías Mensaje Personal (MP), Compartir Ubicación (CU) y Chat (CH). En estos mensajes los usuarios además de utilizar un lenguaje más informal para compartir sus mensajes, hacen uso de los emoticonos para expresar opinión o sentimientos.

También se puede observar que en las categorías Reporte de Noticias (RN) y Publicidad (PU) no se hace uso de emoticonos, esto se debe a que estos mensajes son emitidos principalmente por cuentas corporativas.

Hashtag

El uso de Hashtag ocurre principalmente en las categorías Publicidad (PU), Opinión General (OG), Mensaje Personal (MP) y Pregunta (PR). Esto brinda a los usuarios la posibilidad de categorizar los mensajes, para que puedan ser más fácilmente localizados por la comunidad de usuarios de Twitter, lo que aumenta la audiencia y el alcance de su emisión.

Signos de Exclamación e Interrogación

La posibilidad de utilizar signos de exclamación y signos de interrogación se puede ver utilizada principalmente en las categorías Preguntas (PR) y Publicidad (PU).

En donde el 45,7% de los mensajes de publicidad incluyen signos de exclamación para hacer énfasis en las ofertas.

Por otra parte en la categoría PR se puede notar que con un 98% se hace uso del signo de interrogación para realizar preguntas de manera explícita a los seguidores, que conjunto con la posibilidad de mención se realizan pregunta de manera explícita y directa a los usuarios.

4.2 Herramientas utilizadas

Para la implementación de la etapa de pre-procesamiento se utilizó Python, un lenguaje de programación multiplataforma y código abierto. Además se utilizó NLTK (<http://www.nltk.org>) un conjunto de librerías para el análisis de textos y procesamiento del lenguaje natural.

Para la clasificación se utilizó los clasificadores ofrecidos por la herramienta Weka (<http://www.cs.waikato.ac.nz/ml/weka/>), específicamente Máquinas de Soporte Vectorial y Naive Bayes.

Entre las diferentes funcionalidades que posee Weka, se utilizó un filtro llamado StringToWordVector, el cual convierte atributos de texto en un conjunto de atributos representando la información de aparición de las palabras (dependiendo del tokenizador) del texto contenido en la cadena de texto.

4.3 Clasificadores

Los clasificadores usados en esta investigación son Naive Bayes y Máquinas de Soporte Vectorial, ambos incluidos en Weka. Con respecto a SVM, se utilizó el algoritmo SMO para el entrenamiento del clasificador.

La implementación de SMO, reemplaza globalmente todos los valores perdidos y transforma atributos nominales en atributos binarios. Además normaliza todos los atributos por defecto. Los problemas multi-clase son resueltos usando clasificación de pares (1 vs 1) [17].

Para la división de los datos en conjunto de entrenamiento y prueba se *utiliza k-fold cross-validation*, dividiendo el conjunto de datos en 10 particiones en cada ejecución.

5 Desarrollo de Experimentos y Resultados

En esta sección se evaluará y comparará el rendimiento de los clasificadores Naive Bayes y Máquinas de Soporte Vectorial tomando como entrada el conjunto de tweets descritos anteriormente y clasificados manualmente en la taxonomía de 8 categorías definida en el capítulo 3. Adicionalmente, se identificará la incidencia del tipo de peso de términos utilizado en la representación bolsa de palabras.

Cada instancia contiene la categoría a la que pertenece, la bolsa de palabras, incluyendo las características adicionales, donde se asignan los valores de los pesos de cada palabra dependiendo si se utiliza la presencia de los términos (TP), frecuencia de los términos (TF) o la frecuencia del término inversa en el documento (TF-IDF).

Los resultados de los clasificadores SVM y Naive Bayes se agrupan y se realizan comparaciones de la siguiente manera:

- a) Resultados utilizando Bag of Words
- b) Resultados utilizando Bag of Words removiendo Stop Words
- c) Resultados utilizando Bag of Words removiendo Stop Words y aplicando Stemming

En cada muestra de resultados se presentan dos gráficos. Un gráfico muestra la exactitud de los clasificadores SVM y Naive Bayes según el tipo de peso que se aplica a los términos de la bolsa de palabras, estos pesos son TP, TF y TF-IDF. Además se presenta un gráfico de la precisión de los clasificadores según las categorías definidas.

Los resultados presentados en cuanto a exactitud (*accuracy*) representan el promedio de 10 ejecuciones de cada clasificador. Con respecto a los resultados y gráficos que muestran la precisión y cobertura de los clasificadores, se eligió una de las ejecuciones que presentara la exactitud más cercana al promedio.

5.1 Resultados con BOW

En todos los casos evaluados SVM tiene un rendimiento superior en comparación a Naive Bayes, alcanzando valores cercanos al 6% de mejora en la exactitud (*accuracy*) utilizando TF y TF-IDF. Como se puede observar en los gráficos de la Figura 5.1 es utilizando presencia de términos (TP) cuando Naive Bayes entrega mejor resultados (81,4%) en comparación con TF (79,4%) y TF-IDF (79,6%). Por otro lado, SVM presenta mejor resultados utilizando TF-IDF (86,26%), en comparación a TP (86,17%) y TF (86,18%).

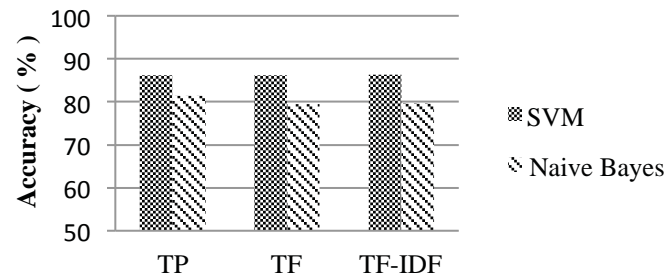


Figura 5.1 Gráfico de exactitud (*accuracy*) con bag of words.

Lo primero que se puede observar de los resultados obtenidos, es que la exactitud de los clasificadores, al utilizar distintos pesos de términos, permanece casi constante. Es por eso, que a continuación sólo se muestran los resultados utilizando Presencia de Términos (TP).

En el gráfico de la Figura 5.2 se compara la precisión de Naive Bayes y SVM utilizando Presencia de Términos (TP). Se puede observar que sólo la categoría pregunta (PR) obtiene una mayor precisión utilizando Naive Bayes en comparación con SVM.

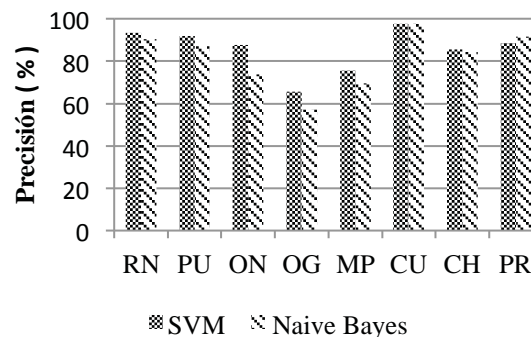


Figura 5.2 Gráfico de precisión por categorías.

Con respecto a la matrices de confusión adjuntas en el Anexo A.1 se puede resumir que los mensajes pertenecientes a la categoría Mensaje Personal cuando son mal clasificados, mayoritariamente son clasificados como Opinión General, sucediendo lo mismo en el caso

contrario. Esto se debe principalmente a la similitud de las palabras utilizadas en estas categorías, además de una utilización de un lenguaje más informal.

5.2 Resultados con BOW Removiendo Stopwords

Al igual que el caso anterior utilizando sólo BOW, al remover stopwords se puede observar en la Figura 5.3 que SVM posee un rendimiento mejor frente a Naive Bayes con una diferencia de entre el 5% y el 7%. Naive Bayes nuevamente presenta mejores resultados al utilizar TP (81,31%) como peso para los términos. Es por esto que a pesar que SVM presenta mejor resultados utilizando TF-IDF (86,3%), se realiza una comparación de la precisión por categorías utilizando TP (donde SVM obtiene un 86,04%).

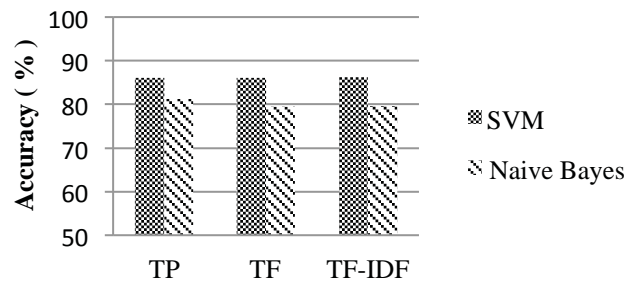


Figura 5.3 Gráfico de exactitud con BOW removiendo stopwords.

En el gráfico de la Figura 5.4 se compara la precisión de Naive Bayes y SVM utilizando Presencia de Términos (TP). Nuevamente se puede observar que sólo la categoría pregunta (PR) obtiene una mayor precisión utilizando Naive Bayes en comparación con SVM y que además Compartir Ubicación (CU) posee una precisión similar utilizando ambos clasificadores.

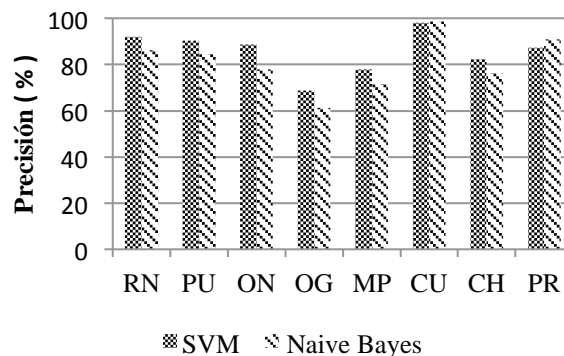


Figura 5.4 Gráfico de precisión por categorías con BOW removiendo stopwords.

Al comparar los resultados utilizando BOW y BOW removiendo stopwords los resultados permanecen en su mayoría constantes lo que significa que no es resulta de utilidad realizar esta fase de procesamiento. Al remover stopwords en un mensaje de Twitter que está compuesto de a lo más 140 caracteres, aproximadamente 30 palabras, se podría estar eliminando información relevante.

5.3 Resultados con BOW Aplicando Stemming

Los resultados obtenidos al aplicar stemming adicionalmente de remover stopwords son muy similares a los resultados anteriormente obtenidos entre SVM y Naive Bayes. SVM se posiciona por sobre Naive Bayes con un margen de 5%.

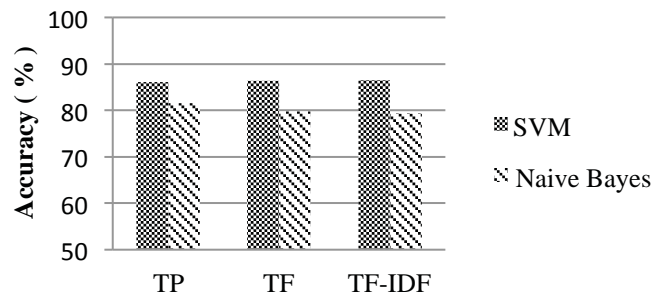


Figura 5.5 Gráfico de exactitud con BOW aplicando stemming.

Al comparar la precisión de Naive Bayes y SVM utilizando Presencia de Términos (TP) en el gráfico de la Figura 5.6, se observa que los resultados siguen la misma tendencia que los obtenidos anteriormente.

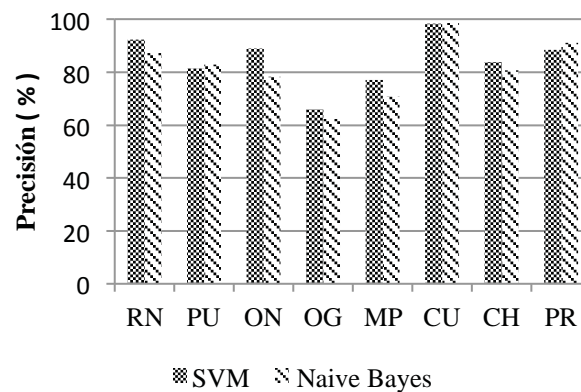


Figura 5.6 Gráfico de precisión por categorías con BOW aplicando stemming.

Al comparar los resultados con los obtenidos previamente en la sección 5.1 y la sección 5.2, se puede apreciar que la eliminación de stopwords y consecutiva aplicación de stemming no resulta en una mejoría de los resultados.

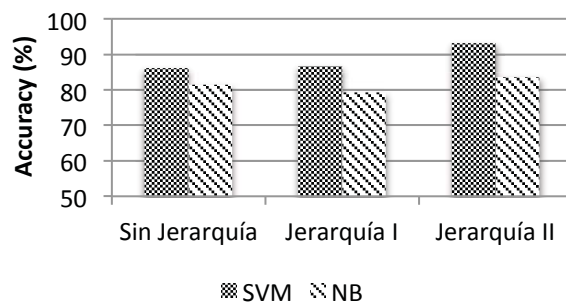
5.4 Evaluación de Resultados Utilizando Jerarquías

La representación utilizada en esta sección corresponde a BOW sin stopwords ni aplicar stemming y como peso de términos se utiliza Presencia de Términos (TP).

En la Figura 5.7 se aprecia la comparación de la exactitud de los clasificadores SVM y Naive Bayes obtenida utilizando las jerarquías definidas.

En los resultados se puede observar claramente que SVM entrega mejores resultados en comparación a Naive Bayes. Además, se puede apreciar que el agrupar las categorías de manera jerárquica entrega mejores resultados utilizando SVM, obteniendo mejoras de hasta un 7% en la exactitud de la clasificación agrupando en la Jerarquía II.

Los resultados obtenidos eran esperables, y se comprueba que la jerarquía II, al estar definida a partir la similitud de las categorías de acuerdo a los resultados obtenidos en la sección 5.1, presenta mejores resultados que la Jerarquía I.



	SVM	NB
Sin Jerarquía	0.861	0.814
Jerarquía I	0.866	0.792
Jerarquía II	0.931	0.836

a) Gráfico comparativo de exactitud de los clasificadores

b) Tabla de exactitud de los clasificadores

Figura 5.7 Comparación de exactitud utilizando jerarquías

5.4.1 Resultados Sin Jerarquía

Los resultados obtenidos muestran que la exactitud de SVM es 86,1% y de Naive Bayes 81,4%. Estos conforman la base para comparar con la clasificación jerárquica.

Tabla 5.1 Métricas de clasificadores sin usar jerarquía

a) SVM

	Precision	Recall	F-Measure
RN	0.933	0.97	0.951
PU	0.92	0.88	0.899
ON	0.878	0.952	0.914
OG	0.655	0.55	0.598
MP	0.755	0.74	0.747
CU	0.977	0.97	0.973
CH	0.855	0.887	0.871
PR	0.886	0.96	0.922

b) Naive Bayes

	Precision	Recall	F-Measure
RN	0.903	0.93	0.916
PU	0.87	0.867	0.868
ON	0.739	0.828	0.781
OG	0.571	0.547	0.559
MP	0.694	0.727	0.71
CU	0.976	0.957	0.966
CH	0.842	0.817	0.829
PR	0.917	0.843	0.878

5.4.2 Resultados Jerarquía I

Tal como se puede apreciar en la Figura 5.7, la exactitud de SVM es de 86,6% y de Naive Bayes es 79,2%. Este resultado corresponde a la ejecución en el primer nivel del árbol jerárquico, considerando las categorías NO, OP, DI y SO definidas en la sección 3.2.1.

En la Tabla 5.2 se aprecian las distintas métricas obtenidas por categorías para la jerarquía I. Para obtener estos valores se calculó a través del producto entre la precisión, cobertura o F-Measure del primer nivel del árbol y el obtenido en los niveles siguientes. Como ejemplo se tiene que la precisión de Noticia (NO) en el primer nivel del árbol es de 91,8% y la precisión de Reporte de Noticia (RN) en el segundo nivel es de 95,5% lo que entrega 88% que se muestra en la tabla. Este cálculo es análogo para los demás valores.

Además, siempre se utiliza el mismo clasificador en los niveles del árbol, o sea, cuando se utiliza SVM en el nivel superior, también se utiliza en el nivel inferior.

Tabla 5.2 Métricas de clasificadores con Jerarquía I

a) SVM				b) Naive Bayes			
	Precision	Recall	F-Measure		Precision	Recall	F-Measure
RN	0,88	0,94	0,91	RN	0,83	0,83	0,83
PU	0,92	0,89	0,90	PU	0,86	0,81	0,83
ON	0,74	0,74	0,74	ON	0,60	0,62	0,61
OG	0,78	0,71	0,73	OG	0,62	0,63	0,62
MP	0,85	0,86	0,85	MP	0,79	0,78	0,79
CU	0,86	0,85	0,85	CU	0,80	0,77	0,79
CH	0,86	0,88	0,87	CH	0,79	0,81	0,80
PR	0,85	0,88	0,87	PR	0,78	0,82	0,80

En la Figura 5.8 se presenta un gráfico comparativo entre el rendimiento de SVM y Naive Bayes para la jerarquía I. Este muestra la precisión entre las 4 categorías superiores definidas.

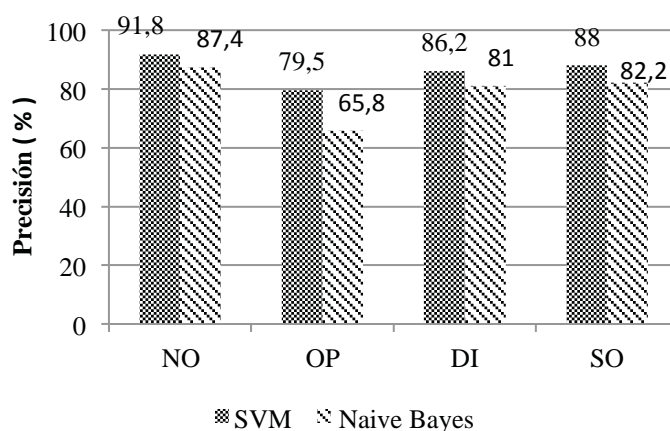


Figura 5.8 Gráfico comparativo SVM y Naive Bayes (Jerarquía I)

Principalmente se puede observar el bajo rendimiento al clasificar las opiniones (79,5% SVM y 65,8% Naive Bayes) comparado con la precisión de las Noticias (91,8% SVM y 87,4% Naive Bayes). Estos resultados eran esperable ya que en la categoría opinión se agruparon dos categorías que presentaban una baja precisión en los resultados previos obtenidos.

5.4.3 Resultados Jerarquía II

Tal como se puede apreciar en la Figura 5.7, la exactitud de SVM es de 93,1% y de Naive Bayes es 83,6%. Este resultado corresponde a la ejecución en el primer nivel del árbol jerárquico, considerando las categorías NO, SO, DI y PU definidas en la sección 3.2.2.

Tabla 5.3 Métricas de clasificadores con Jerarquía II

a) SVM				b) Naive Bayes			
	Precision	Recall	F-Measure		Precision	Recall	F-Measure
RN	0,90	0,97	0,93	RN	0,82	0,83	0,82
PU	0,94	0,86	0,89	PU	0,88	0,85	0,83
ON	0,88	0,98	0,93	ON	0,83	0,79	0,81
OG	0,67	0,60	0,63	OG	0,53	0,58	0,56
MP	0,74	0,73	0,73	MP	0,62	0,71	0,65
CU	0,92	0,95	0,94	CU	0,85	0,73	0,78
CH	0,83	0,82	0,82	CH	0,71	0,73	0,72
PR	0,93	0,95	0,94	PR	0,84	0,74	0,78

En la Figura 5.9 se presenta un gráfico comparativo entre el rendimiento de SVM y Naive Bayes para la jerarquía II. Este muestra la precisión entre las 4 categorías superiores definidas.

En este caso se puede observar que al agrupar las categorías según su similitud en base a los resultados obtenidos previamente, se mejora bastante el porcentaje de precisión. Además se observa SVM (93,1%) sigue siendo superior que Naive Bayes (83,6%) en cuanto a exactitud.

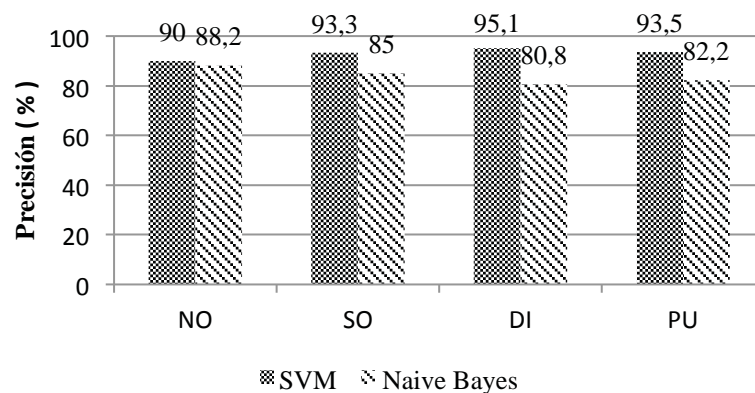


Figura 5.9 Gráfico comparativo SVM y Naive Bayes (Jerarquía II)

5.4.4 Resumen de Resultados Jerárquicos

En la Figura 5.10 se muestra el F-Measure por categoría, medida que toma en consideración la precisión y la cobertura, comparando así el rendimiento de SVM agrupando las categorías en base a la Jerarquía I, Jerarquía II y sin utilizar jerarquía.

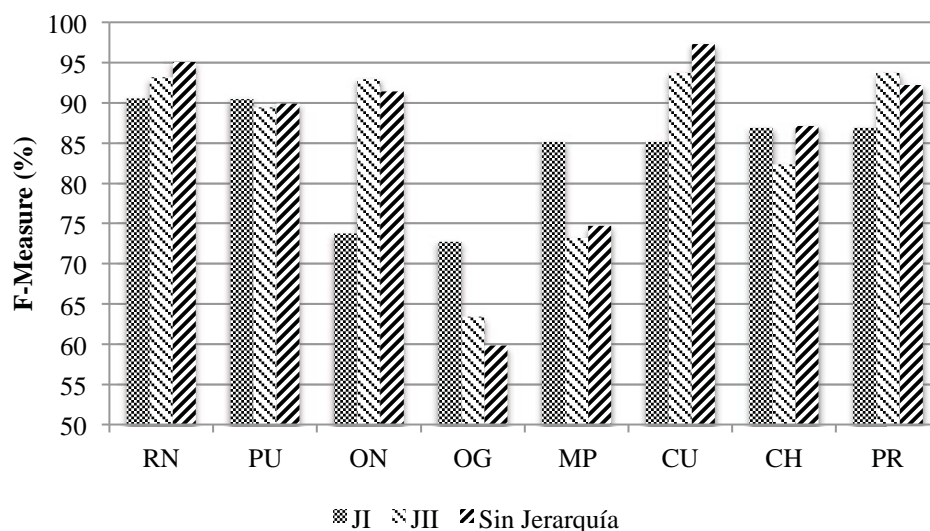


Figura 5.10 Gráfico comparativo utilizando jerarquías y sin utilizar jerarquía (SVM)

En el gráfico se puede observar que las categorías Reporte de Noticia y Compartir Ubicación / Evento son mejor clasificadas cuando no se agrupan con otras categorías mediante alguna de las jerarquías definidas. Por el contrario, se puede observar que las categorías Opinión General y Mensaje Personal son mejor clasificadas cuando se agrupan mediante la Jerarquía I, no así con la categoría Opinión de Noticias, que disminuye bastante el porcentaje correcto de clasificación utilizando la Jerarquía I.

Realizando una comparación a partir de las métricas precisión, cobertura y F-Measure se puede decir las jerarquías I y II tienen un comportamiento aceptable comparado con los resultados obtenidos sin jerarquía, lo que indica que ambas jerarquías entregan buenos resultados y son completamente válidas para poder agrupar los mensajes en categorías superiores.

6 Conclusiones

El trabajo presentado en esta tesis describe la problemática de la clasificación de mensajes de Twitter de acuerdo a la intención del usuario. Se ha descrito el problema, y se ha definido una propuesta que permite clasificar automáticamente los mensajes de Twitter. Esta propuesta incluyó la definición de una taxonomía en la que se definen las categorías que abarcan los mensajes de Twitter de acuerdo a su intención y un modelo de clasificación, incluyendo la etapa de pre-procesamiento, representación y clasificación, obteniendo resultados con distintos clasificadores.

Con respecto a la taxonomía de este trabajo, se propone un conjunto de categorías que agrupan las distintas propuestas evaluadas en los trabajos relacionados. La taxonomía abarca la mayoría de las categorías propuestas, reutilizando y complementando algunas definiciones.

La taxonomía, respecto a los resultados obtenidos, muestra una gran similitud entre las categorías reporte de noticias y opinión de noticias, esto se debe a que una opinión de noticia contiene en gran parte del mensaje el reporte de noticia del cual se está emitiendo juicio. Por otra parte, queda en evidencia que las categorías Compartir Ubicación / Evento y Reporte de Noticias son las que obtienen mejor precisión en comparación con el resto.

Adicionalmente, se definió dos jerarquías que agrupan las 8 categorías definidas en la taxonomía. Una de estas fue definida a partir de los resultados obtenidos anteriormente y la otra jerarquía reúne las clases a partir de la definición de estas categorías. Los resultados obtenidos con ambas jerarquías son considerados buenos y permiten poder clasificar los mensajes a distinto nivel de detalle, permitiendo distinguir los mensajes emitidos con la intención de informar sobre noticias de los mensajes emitidos con la intención de sociabilizar, y si se requiere, poder detallar a que sub categoría pertenecen.

La etapa de pre-procesamiento, como en cualquier proceso de minería de datos, es una de las más importante para la obtención de buenos resultados. Transformar los mensajes, realizar limpieza a los datos, y convertirlos para que el clasificador sea capaz de “entender” el texto es un proceso fundamental.

Se ha definido el uso de máquinas de aprendizaje para la clasificación automática de los mensajes, también siguiendo la línea de [15] se utilizarán características adicionales complementando Bag of Words, obteniendo mejoras de hasta un 6% en la exactitud de los clasificadores.

Respecto a los resultados, las Máquinas de Soporte Vectorial presentan mejores resultados que Naive Bayes para el tipo de clasificación expuesto en este trabajo, alcanzando hasta un 7% de diferencia en la exactitud del clasificador dependiendo del tipo de peso utilizado para los términos. También se puede concluir que el tipo de peso de términos resulta poco relevante para la clasificación de mensajes de Twitter. A pesar que TF-IDF pueda presentar una pequeña mejora con respecto a TP utilizando SVM, se prefiere el uso de Presencia de Términos debido a que el costo es mucho menor que el cálculo que implica obtener la Frecuencia de los Términos y más aun TF-IDF. La pequeña diferencia entre los resultados obtenidos utilizando TP, TF y TF-IDF se debe a que los mensajes de Twitter son

particularmente cortos, por lo que los tres pesos de términos se comportan prácticamente de la misma manera.

En relación a la categoría que mayor problema presenta en la clasificación, se puede concluir que tanto en Naive Bayes como en SVM esta categoría corresponde a Opinión General, alcanzando un 60% de precisión aproximadamente en ambos clasificadores. Por otra parte las categorías que presentan mejores resultados, tanto en Naive Bayes como en SVM corresponden a Compartir Ubicación/Evento, Reporte de Noticias y Publicidad.

A partir de las matrices de confusión expuestas en el Anexo A, se puede observar que un porcentaje importantes de mensajes que debiesen ser clasificados como Mensaje Personal son clasificados como Opinión General. Así mismo los mensajes que debiesen ser clasificados como Opinión General son clasificados como Mensaje Personal o Chat.

El problema en la clasificación en las categorías Mensaje Personal, Opinión General y Chat se debe a que el lenguaje utilizado en estas tres categorías es muy similar. Otro motivo que afecta la clasificación de estas categorías es el lenguaje informal utilizado, las faltas de ortografía, el uso de jergas y la abreviación de palabras, que causa que palabras que debiesen ser consideradas como la misma, se consideren como palabras diferentes.

Desde el punto de vista jerárquico, se puede concluir que una vez definida las dos jerarquías y comparar el rendimiento de los clasificadores Máquinas de Soporte Vectorial y Naive Bayes, la jerarquía II presenta un mejor desempeño en comparación a la jerarquía I.

El resultado obtenido era esperable, debido a que en la jerarquía II se agrupó las categorías mediante los resultados obtenidos en la sección 5.4. Realizando una comparación a partir de las métricas precisión, cobertura y F-Measure se puede decir que las jerarquías I y II tienen un comportamiento aceptable comparado con los resultados obtenidos sin jerarquía, lo que indica que ambas jerarquías entregan buenos resultados y son completamente válidas para poder agrupar los mensajes en categorías superiores.

Poder enfrentar el problema de la clasificación de la intención de los usuarios en los mensajes de Twitter desde un punto de vista jerárquico, permite clasificar los mensajes en distintos niveles de detalle. Por lo tanto, si se desea poder separar los mensajes sociales, noticiosos de los mensajes diarios y de opinión, se puede utilizar las 4 categorías definidas en las jerarquías I o II, lo que permite a su vez tener una granularidad más fina al poder bajar al siguiente nivel y obtener la clasificación de estos mensajes en una categoría mas específica.

Como conclusión general, los mejores resultados se obtienen utilizando Máquinas de Soporte Vectorial para la clasificación automática de la intención del usuario en los mensajes de Twitter. Utilizando Bag of Words para la representación de los mensajes, y debido a que no existe mayor diferencia en los resultados obtenidos con los distintos pesos de términos, se propone utilizar presencia de términos debido a que es el que presenta un menor costo computacional en comparación a frecuencia de términos.

Como trabajo futuro se propone realizar mejoras en el proceso de pre-procesamiento, enfocado principalmente en la corrección de palabras con falta de ortografía, detección y corrección de palabras acortadas e identificación del uso de jergas.

Además, se propone realizar un análisis descriptivo de los datos utilizando una representación en base a n-gramas, identificar cuales son los bigramas y trigramas que se repiten con mayor frecuencia, y además evaluar el rendimiento de los clasificadores utilizando dicha representación.

Finalmente, se propone complementar las características adicionales que se proponen en este estudio, enfocado principalmente en mejorar la precisión en las categorías con menor rendimiento.

Referencias

- [1] Ingrid Lunden. (2012, Julio) TechCrunch. [Online]. <http://techcrunch.com/2012/07/31/twitter-may-have-500m-users-but-only-170m-are-active-75-on-twitters-own-clients/>
- [2] Seth Fiegerman. (2012, Diciembre) Mashable. [Online]. <http://mashable.com/2012/12/18/twitter-200-million-active-users/>
- [3] Nicholas A. Diakopoulos and David A. Shamma, "Characterizing debate performance via aggregated twitter sentiment", in *ACM*, Atlanta Georgia, 2010.
- [4] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo, "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors", in *WWW2010*, North Carolina, 2010.
- [5] Fabrizio Sebastiani, "Machine Learning in Automated Text Categorization", in *ACM Computing Surveys* Volume 32 Issue 1, Marzo 2002.
- [6] Alberto Téllez Valero, "Extracción de Información con Algoritmos de Clasificación", INAOE, Tesis de Maestría 2005.
- [7] David Tomás Díaz, "Sistemas de Clasificación de Preguntas Basados en Corpus para la Búsqueda de Respuestas", in *Tesis Doctoral, Universidad de Alicante*, 2009.
- [8] Henok Girma, "A Tutorial on Support Vector Machine", in *Center of Experimental Mechanics, University of Ljubljana*, 2009.
- [9] Alessandro Verri Massimiliano Pontil, "Properties of Support Vector Machines", 1997.
- [10] Website-Monitoring. (2012) Twitter 2012 - Facts and Figures. [Online]. <http://www.website-monitoring.com/blog/2012/11/07/twitter-2012-facts-and-figures-infographic/>
- [11] Website-Monitoring. (2010) Twitter Facts and Figures History Statistics. [Online]. <http://www.website-monitoring.com/blog/2010/05/04/twitter-facts-and-figures-history-statistics/>
- [12] Akshay Java, Tim Finin, Xiaodan Song, and Belle Tseng, "Why We Twitter: Understanding Microblogging Usage and Communities", in *WebKDD/SNA-KDD '07 Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 2007.
- [13] Mor Naaman, Jeffrey Boase, and Chih-Hui Lai, "Is it Really About Me?: Message Content in Social Awareness Streams", in *CSCW '10 Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 2010.
- [14] Meredith Ringel Morris, Jaime Teevan, and Katrina Panovich, "What do People Ask Their Social Networks, and Why?: A Survey Study of Status Message Q&A Behavior", in *CHI '10 Proceedings of the 28th international conference on Human factors in computing systems*, 2010.
- [15] Bharath Sriram, "Short Text Classification in Twitter to Improve Information Filtering", in *Master of Science Thesis, The Ohio State University*, 2010.
- [16] Vipul Pandey and C.V.Krishnakumar Iyer, "Sentiment Analysis of Microblogs", SOFSEM

2012: Theory and Practice of Computers Science, 38th Conference on Current Trends in Theory and Practice of Computer Science, Enero 2012.

- [17] John Platt, "Fast Training of Support Vector Machines using Sequential Minimal Optimization", in *Advances in Kernel Methods - Support Vector Learning*, Advances in kernel methods pp. 185-209, Mit Press Cambridge, 1999.
- [18] Fermín L. Cruz, José A. Troyano, Fernando Enríquez, and Javier Ortega, "Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español", 2008, *Procesamiento del lenguaje Natural*, nº 41 (2008), pp. 73-80, Sociedad Española para el procesamiento del Lenguaje Natural.

Anexos

A Resultados Detallados

A.1 Bag of Words

Presencia de Términos

1) SVM

a) Matriz de Confusión

		Predicción							
		RN	PU	ON	OG	MP	CU	CH	PR
Real	RN	291	1	5	0	1	1	0	1
	PU	9	264	6	5	9	1	3	3
	ON	0	1	238	5	0	1	4	1
	OG	4	6	13	165	56	1	35	20
	MP	5	9	3	52	222	1	2	6
	CU	3	3	0	0	3	291	0	0
	CH	0	1	3	23	0	1	266	6
	PR	0	2	3	2	3	1	1	288

b) Métricas por Categorías

	Precision	Recall	F-Measure
RN	0.933	0.97	0.951
PU	0.92	0.88	0.899
ON	0.878	0.952	0.914
OG	0.655	0.55	0.598
MP	0.755	0.74	0.747
CU	0.977	0.97	0.973
CH	0.855	0.887	0.871
PR	0.886	0.96	0.922

2) Naive Bayes

a) Matriz de Confusión

		Predicción							
		RN	PU	ON	OG	MP	CU	CH	PR
Real	RN	279	3	17	0	1	0	0	0
	PU	9	260	14	3	10	0	3	1
	ON	15	4	207	13	6	3	1	1
	OG	1	12	19	164	60	0	30	14
	MP	3	14	5	51	218	4	2	3
	CU	2	1	5	1	4	287	0	0
	CH	0	3	9	36	3	0	245	4
	PR	0	2	4	19	12	0	10	253

b) Métricas por Categorías

	Precision	Recall	F-Measure
RN	0.903	0.93	0.916
PU	0.87	0.867	0.868
ON	0.739	0.828	0.781
OG	0.571	0.547	0.559
MP	0.694	0.727	0.71
CU	0.976	0.957	0.966
CH	0.842	0.817	0.829
PR	0.917	0.843	0.878

A.2 Bag of Words Removiendo Stopwords

Presencia de Términos

1) SVM

a) Matriz de Confusión

		Predicción							
		RN	PU	ON	OG	MP	CU	CH	PR
Real	RN	290	3	6	0	0	1	0	0
	PU	15	255	6	6	9	1	4	4
	ON	0	2	235	5	0	1	6	1
	OG	4	8	11	160	52	1	42	22
	MP	3	5	2	45	231	1	5	8
	CU	3	4	1	0	3	289	0	0
	CH	0	2	2	14	1	1	273	7
	PR	0	3	2	3	1	0	2	289

b) Métricas por Categorías

	Precision	Recall	F-Measure
RN	0.921	0.967	0.943
PU	0.904	0.85	0.876
ON	0.887	0.94	0.913
OG	0.687	0.533	0.6
MP	0.778	0.77	0.774
CU	0.98	0.963	0.971
CH	0.822	0.91	0.864
PR	0.873	0.963	0.916

2) Naive Bayes

a) Matriz de Confusión

		Predicción							
		RN	PU	ON	OG	MP	CU	CH	PR
Real	RN	278	6	14	0	1	1	0	0
	PU	14	253	13	5	8	2	4	1
	ON	22	2	203	9	5	0	8	1
	OG	3	15	15	134	64	0	54	15
	MP	3	13	5	35	232	1	6	5
	CU	3	3	3	0	2	289	0	0
	CH	0	3	6	21	1	0	265	4
	PR	0	4	2	15	11	0	11	257

b) Métricas por Categorías

	Precision	Recall	F-Measure
RN	0.861	0.927	0.892
PU	0.846	0.843	0.845
ON	0.778	0.812	0.795
OG	0.612	0.447	0.516
MP	0.716	0.773	0.744
CU	0.986	0.963	0.975
CH	0.761	0.883	0.818
PR	0.908	0.857	0.882

A.3 Bag of Words Aplicando Stemming

Presencia de Términos

1) SVM

a) Matriz de Confusión

		Predicción							
		RN	PU	ON	OG	MP	CU	CH	PR
Real	RN	287	4	6	0	1	2	0	0
	PU	12	260	4	7	11	0	3	3
	ON	0	0	238	7	0	1	3	1
	OG	5	6	12	167	50	1	39	20
	MP	3	7	3	49	226	0	4	8
	CU	3	3	0	2	1	291	0	0
	CH	0	2	3	20	2	1	266	6
	PR	1	2	2	1	2	0	2	290

b) Métricas por Categorías

	Precision	Recall	F-Measure
RN	0.923	0.957	0.939
PU	0.915	0.867	0.89
ON	0.888	0.952	0.919
OG	0.66	0.557	0.604
MP	0.771	0.753	0.762
CU	0.983	0.97	0.977
CH	0.839	0.887	0.862
PR	0.884	0.967	0.924

2) Naive Bayes

a) Matriz de Confusión

		Predicción							
		RN	PU	ON	OG	MP	CU	CH	PR
Real	RN	269	8	15	0	5	3	0	0
	PU	13	255	9	6	11	0	4	2
	ON	21	1	209	10	5	0	3	1
	OG	3	12	17	161	62	0	32	13
	MP	2	20	4	33	230	1	6	4
	CU	1	4	3	1	2	289	0	0
	CH	0	3	8	30	2	0	252	5
	PR	0	5	2	18	8	0	15	252

b) Métricas por Categorías

	Precision	Recall	F-Measure
RN	0.871	0.897	0.883
PU	0.828	0.85	0.839
ON	0.783	0.836	0.809
OG	0.622	0.537	0.576
MP	0.708	0.767	0.736
CU	0.986	0.963	0.975
CH	0.808	0.84	0.824
PR	0.91	0.84	0.873

B Extracto de Tweets por Categoría

B.1 Reporte de Noticia

Emisor	Mensaje
Comunidadpucv	DEMRE - Noticias: Plazo extendido para inscribirse en PSU hasta el 29 de julio: http://t.co/oZCFoP5 via @AddThis
Comunidadpucv	RT @ucvradio: Rector de la PUCV Claudio Elórtegui analizó panorama educativo y requerimientos que deben enfrentar las universidades http://t.co/ffa6YEF via @comunidadpucv
Comunidadpucv	Congreso Nacional: Rector destaca aporte de universidades públicas al desarrollo del país http://t.co/ffa6YEF via @comunidadpucv
Comunidadpucv	Estudiantes organizan feria laboral para personas discapacitadas y empresarios http://t.co/opPWVBm via @comunidadpucv
Comunidadpucv	EXPLORA CONICYT: concurso científico premia a escolares con expedición a la Antártica http://t.co/imqVExm via @comunidadpucv
Comunidadpucv	Investigación PUCV: textos de Lingüística persuaden más al lector que los de medicina y economía. http://t.co/qnUsckK via @comunidadpucv
Comunidadpucv	Las claves en el éxito de Alemania. Diario estrategia http://t.co/mhR7bxa
Comunidadpucv	Crean ingeniosa chaqueta porta-bebés para emergencias: http://t.co/3PMr59G via @AddThis
Comunidadpucv	Nueva plataforma virtual: Rectoría llama a defender carácter público de la PUCV http://t.co/JyS8zty via @comunidadpucv
Comunidadpucv	RT @pucvperiodismo: Congreso Latinoamericano de Estudiantes de Comunicación Social. La Paz, Bolivia. 18 al 21 de octubre de 2011 http://t.co/ffa6YEF via @comunidadpucv
Comunidadpucv	Niños aprenden con juegos a pensar como científicos: http://t.co/bDhRkqK via @AddThis
Comunidadpucv	PUCV logró el pentacampeonato en el balonmano de los Juegos Universitario-Navales #handball http://t.co/ust23V1 via @soyvalparaiso
Comunidadpucv	Ingeniería Comercial: estudiantes organizan muestra sobre tecnologías para la educación http://t.co/Wq9afKY via @comunidadpucv
Comunidadpucv	Columna de Pedro Santander, docente PUCV/Noticiarios de televisión: la misma cantinela todas las noches http://t.co/e8vtj1m via @elmostrador
Comunidadpucv	El próximo 9 de julio culmina plazo para inscribirse en beca que permite rendir gratis la PSU http://t.co/TRK54XO
Comunidadpucv	Una Margot Loyola de 92 años completa sus históricas investigaciones con el libro "La cueca" http://t.co/dNUUrtP via @emol
Emol	Presidente Piñera convoca a un Gran Acuerdo Nacional por la Educación http://j.mp/oIjjhh
Emol	Matías Fernández será sometido a exámenes para determinar la gravedad de su lesión. http://bit.ly/iKSRST
Emol	Matthei asegura ante ministro que no sabe quiénes bombardearon La Moneda el '73. http://bit.ly/mcASpK
Emol	CERC: Aprobación a Piñera cae a 35% y desaprobación sube a 53%. http://bit.ly/iMJ58X
Emol	Diego Rubio y su partida al Sporting de Lisboa: "Esto es un sueño". http://bit.ly/mHZRJp
Emol	La economía chilena crece más de lo esperado: Imacec llega a 7,3%. http://bit.ly/jPqYg0
Emol	Claudio Palma y Canal 13 ganan el primer duelo televisivo de #LaRoja en la Copa América http://j.mp/jPH22m #ca2011
Emol	Claudio Borghi: "Cuando nos salgan a atacar vamos a tener más opciones de hacer daño" http://j.mp/mNlpXe #LaRoja #ca2011
Emol	Van Lamoén clasifica al cuadro eliminatorio del mundial preolímpico de tiro al arco http://bit.ly/jQJMt
Emol	Chile y Japón firman acuerdo para proyecto de cooperación en prevención de tsunamis http://bit.ly/mxj3UD
Emol	Presidente del Palermo asegura que Pastore puede ir al Málaga de Pellegrini http://bit.ly/lcPvV
Emol	Programación de la 1 fecha del Grupo C de la Copa América http://bit.ly/mwAj05

Emol Popular modelo argentina lanza chocante campaña contra el aborto <http://bit.ly/kXIR92>

Emol Ex Kudai Pablo Holman participará en serie de MTV <http://bit.ly/lr4Ax4>

Emol Larroulet y crisis en Educación: "No estamos dispuestos a derogar el fin de lucro". <http://bit.ly/jBhgbm>

Emol Revista Time hace duro análisis del sistema educacional chileno. <http://bit.ly/ieQrfE>

Emol Internacional de Judo en Miami: Chile obtiene tres medallas de bronce en el US Open <http://bit.ly/kGRnWw>

Emol Ola de frío en Uruguay deja cinco indigentes muertos <http://bit.ly/kOYlpt>

Emol Dior presenta su primera colección sin John Galliano <http://bit.ly/jHVFyC>

Emol Kate Moss es avergonzada en su matrimonio por amiga de su esposo <http://bit.ly/jbqPHz>

Emol Fracasan las pruebas con los nuevos misiles supersónicos en Taiwán <http://bit.ly/mcGy4o>

Emol Descubiertos en Ucrania los segundos homo sapiens más antiguos de Europa <http://bit.ly/mmJblX>

Emol Más de diez millones de personas son afectadas por la sequía en el Cuerno del África <http://bit.ly/jVutcT>

Emol En 40 mil dólares subastan el legendario bolso de Margaret Thatcher <http://bit.ly/jdsWFO>

Emol Obama se reúne con líderes partidistas para analizar déficit <http://bit.ly/mIVXWL>

Emol En completa normalidad regresaron a la capital 180 mil vehículos <http://bit.ly/js7NLP>

Emol Denuncian violento desalojo en Universidad Alberto Hurtado <http://bit.ly/kyMsVb>

Emol Detienen a presunto jefe de las FARC en Ecuador <http://bit.ly/lloie5>

Emol Parejas gay podrán casarse en Nueva York a partir del 25 julio <http://bit.ly/j1IcBM>

Emol Claudio Bravo pide tranquilidad: "No hemos logrado nada aún a nivel de selección" <http://bit.ly/iyOmbX>

Emol Alicia Keys participará en producción de obra en Broadway <http://bit.ly/m0wsBD>

Emol PDI arresta a abuelo y primo de niña de 8 años que denunció ultrajes <http://bit.ly/jiE50Y>

Emol Carabineros despliega operativo en sector oriente de la capital tras atraco en Vitacura <http://bit.ly/kILNRs>

Laterceracom "El juicio en mi contra fue un show mediático y pasé por un verdadero circo romano", dijo Pilar Pérez a La Tercera <http://ow.ly/5vaEt>

Laterceracom Ultima Hora: Pilar Pérez habla desde la cárcel <http://ow.ly/5vaBJ>

Laterceracom Gaddafi amenaza con atacar Europa si la OTAN no detiene sus incursiones en Libia <http://ow.ly/5v9Kp>

Laterceracom Chile va camino al aeropuerto para viajar a Mendoza e iniciar aventura en Copa América. La transmisión de @LaTerceraTV en www.latercera.com.

Laterceracom Aquí puedes ver el frente a frente entre Nadal y Djokovic, los finalistas de #Wimbledon <http://ow.ly/5v6ox>

Laterceracom Camarera que acusó a ex jefe del FMI admitió haber mentado a la justicia <http://ow.ly/5v62D>

Laterceracom Por qué el Barça quiere a Sánchez: "Es capaz de engrasar la máquina perfecta de Guardiola" <http://ow.ly/5v60j>

Laterceracom ¿Ya viste el especial de Copa América?. Revisa acá todas las novedades del certamen que se inicia esta noche <http://ow.ly/5v5Tu>

Laterceracom Encuentran cuerpo de pescador que naufragó en Hualpén <http://ow.ly/5v5uY>

Laterceracom La Polar iniciará devolución en efectivo el 11 de julio a clientes que pagaron deuda repactada unilateralmente <http://ow.ly/5uXy7>

laterceracom Los vuelos de Lan a Buenos Aires y Montevideo, además de los que se dirigen a Neuquén y Bariloche están cancelados <http://ow.ly/5uWkt>

Laterceracom LAN vuelve a cancelar vuelos desde y hacia Argentina y Uruguay por cenizas volcánicas <http://ow.ly/5uWjy>

Laterceracom Acá detalles de la sentencia que deja en libertad condicional y con arraigo a ex jefe del FMI Dominique Strauss-Kahn <http://ow.ly/5uTCj>

B.2 Publicidad

Emisor	Mensaje
santanderchile	¡Sube tu video celebrando un gol de Chile junto a tu padre o hijo, y gana! participa en http://on.fb.me/AsadoSantander
santanderchile	Gana un asado para 12 personas, y vive la Copa América Argentina 2011 junto a @SantanderChile ¡PARTICIPA! http://on.fb.me/m9RuVj
santanderchile	Ya están abiertas las inscripciones para las Becas de Movilidad Libre Elección. Infórmate en Santander.cl sección Joven
santanderchile	GROUPON SANTANDER Paga \$5.000 y consume \$10.000 a la carta + 2 pisco sour Restaurant Embassy de PUERTO MONTT http://bit.ly/kpHqMM
santanderchile	Falta muy poco para que comience la #CopaAmérica, y para calentar motores tenemos un entretenido concurso ¡PARTICIPA! http://on.fb.me/INgYQl
santanderchile	Se acerca la #COPAMAERICA, y nosotros tenemos un entretenido #CONCURSO. Participa por una pelota autografiada x Pelé: http://on.fb.me/INgYQl
santanderchile	Ya están abiertas las inscripciones para las Becas de Movilidad Libre Elección. Infórmate en Santander.cl sección Joven http://bit.ly/mL2Xcd
santanderchile	Falta muy poco para que comience la #COPAMAERICA, y para calentar motores tenemos un entretenido #CONCURSO Participa: http://on.fb.me/INgYQl
santanderchile	Ya tenemos al ganador de nuestro concurso de la final de la Copa Santander Libertadores... ¿Serás tú? Averígualo aquí http://t.co/9YcuSY7
santanderchile	Ya están abiertas las inscripciones para las Becas de Movilidad Libre Elección. Más info #BecasSantanderUniversidades
santanderchile	¿Cuál será el resultado del partido entre Peñarol y Santos? Adivínalo, y participa por 1polera+ 1jockey http://bit.ly/SantanderLibertadores
santanderchile	Usa tu #TarjetaTUI, aprovecha los descuentos que tenemos para ti. Visita www.santander.cl sección Santander Joven .
santanderchile	La Copa Santander Libertadores llega a su fin, y nosotros te regalamos una camiseta y un jockey. ¡PARTICIPA! http://on.fb.me/mwjJtr
santanderchile	#CONCURSA ¿Quién será el Equipo campeón de la Copa Santander Libertadores, Santos o Peñarol? ¡Participa! Link http://on.fb.me/mwjJtr
santanderchile	Tu #TarjetaTUI acumula Puntos Club Movistar por tus compras. Visita www.santander.cl sección Santander Joven para descuentos y promociones
santanderchile	¿Te quedaste sin minutos en tu celular?, recarga con tu #TarjetaTUI en cajeros y a través de www.santander.cl
santanderchile	Usa tu #TarjetaTUI y aprovecha los descuentos que tenemos para ti. Visita www.santander.cl sección Santander Joven.
santanderchile	Recarga tu prepago @entel, @movistarchile o @clarochile_cl y el de quien quieras. Más información en http://bit.ly/kNBDXu
santanderchile	Ahora la TUI acumula Puntos Club Movistar por tus compras. Visita www.santander.cl sección Santander Joven para descuentos y promociones.
santanderchile	Recarga tu prepago @entel, @movistarchile o @clarochile_cl y el de quien quieras. Más información en http://bit.ly/kNBDXu
santanderchile	Paga \$54.600 en vez de \$140.000 x taller teórico y práctico de fotografía digital básica+ workshop Raúl Ernesto Encina http://bit.ly/IRjnzG
santanderchile	Revisa nuestro descuento Groupon de hoy exclusivo para clientes Santander. http://bit.ly/IRjnzG
santanderchile	#SuperVentaHipotecaria Propiedades a porcos únicos exclusivos para ti. Conócelas AQUÍ http://bit.ly/mCh8jJ
santanderchile	Paga \$16.000 en vez de \$32.800 x parrillada para 2+2 pisco sour o vaina + ensalada y más...En restaurant La Pradera http://bit.ly/IRjnzG
santanderchile	¿Sabes algo acerca de nuestras Becas de estudio Internacional?, infórmate en http://bit.ly/lqUItz
santanderchile	Recuerda cambiar la clave de tu #TarjetaTUI en cajeros Santander durante Mayo y participa en espectaculares sorteos. http://bit.ly/mvSw52
santanderchile	Aún puedes acumular opciones y ser uno de los 100 padres y 100 hijos que viajarán a la Copa América http://bit.ly/fMV3Z

santanderchile Contrata tu Crédito de Consumo y acumula 20 opciones para participar de los Sorteos de la Copa América <http://bit.ly/jiLlST>

santanderchile Compra todo lo que quieras de 6 a 12 cuotas sin interés con tus tarjetas de Crédito Santander. <http://bit.ly/mL3mvO>

santanderchile Si estás en el último año de Universidad esto es para ti <http://bit.ly/kdDpst>

santanderchile Compra todo lo que quieras de 6 a 12 cuotas sin interés con tus tarjetas de Crédito Santander. <http://bit.ly/mL3mvO>

santanderchile Sabemos que la decisión de comprar tu casa propia no es fácil. Cotiza tu hipotecario con nosotros <http://bit.ly/lfH5Skc>

laterceracom Buenos días! Acá puedes ver el diario La Tercera de hoy <http://ow.ly/5dtUj>

laterceracom PARTICIPA: ¿Quién ganará el título del Torneo de Apertura entre U. de Chile y U. Católica? <http://ow.ly/5droP>

laterceracom Buenos días! Acá puedes ver el diario La Tercera de hoy <http://ow.ly/5cxcE>

laterceracom Buenos días! Acá puedes ver el diario La Tercera de hoy <http://ow.ly/5bucE>

movistarchile Participa por un iPad 2 contándonos por qué eres un Nativo Digital <http://mvst.cl/nativos>

movistarchile ¡Aprovecha! Recarga tu celular en Gasco desde \$3500 y podrás obtener hasta 150 minutos gratis <http://mvst.cl/promore>

movistarchile ¡Buenos días! Tenemos más de 100 beneficios con el Club Movistar, ¿los conoces todos? <http://mvst.cl/dctoel>

movistarchile ¿Sabías que también puedes hacer rendir el saldo de tu prepago? Marca *303# y compra tu bolsa de SMS, MMS y minutos <http://mvst.cl/bols>

movistarchile ¿Vas a viajar? Puedes conocer los cambios de monedas marcando 0911 desde tu celular. Infórmate aquí <http://mvst.cl/0911>

movistarchile ¿Sabías que no sólo el Iphone puede utilizar Tethering? Entérate de más aquí: <http://mvst.cl/teth>

movistarchile Conoce el proyecto IPv6 que Telefónica tiene para Chile #worldipv6day <http://ipv6.movistar.cl>

movistarchile ¿Crees ser un Nativo Digital? Cuéntanos por qué y gana un iPad2 <http://mvst.cl/nativos>

movistarchile Tethering no tiene costo adicional a tu plan de datos. Si aún no lo tienes, aquí nuestras opciones para ti <http://mvst.cl/pmult>

movistarchile ¡Canjea Internet seguro por 30 días con 5.000 puntos Club Movistar! <http://mvst.cl/canjptos>

movistarchile ¿Sabías que hoy celebramos el IPv6 World Day? Entérate qué es y más en <http://mvst.cl/ipv6#worldipv6day>

movistarchile Aprende más sobre tethering en @movistarchile <http://mvst.cl/teth>

movistarchile ¡No te pierdas la final del fútbol chileno! Obtén TV Digital Prepago desde \$790 con CDF Premium ¡Sin costo por 1 mes! <http://mvst.cl/tvdprep>

movistarchile Si realizas una recarga sobre 5 mil pesos en tiendas Falabella, ¡Obtienes 30 minutos de regalo! ¿Qué esperas? <http://mvst.cl/promore>

movistarchile Hasta el 30/06 puedes disfrutar sin costo, de todo el deporte olímpico por las pantallas del CDO (494). Exclusivo clientes TV Digital

movistarchile @VitoAballay Puedes hacer Tethering con LG Optimus One. Es sin costo si tienes plan de datos, si no, el valor es \$350 la sesión de 1 hora

movistarchile ¡Buenos días! ¿Sabías que con el Plan Top de TV Digital puedes acumular 6.000 km LANPASS al año? Conócelo aquí <http://mvst.cl/plantop>

movistarchile Si te encanta la música no pierdas la oportunidad de canjear entradas dobles al @movistararena por sólo 5.000 puntos Club ¡Buenas noches!

movistarchile ¿Estás perdido? Marca 0911 desde tu celular y obtén la ubicación de calles en tu ciudad <http://mvst.cl/a0911>

movistarchile ¿Ya fuiste por tu Kit Prepago TV Digital? En la R.M: La Polar, Ripley, Paris, Tottus y Corona <http://mvst.cl/tvdprep>

movistarchile Recarga en OK Market desde \$3.500 y obtén tu premio marcando *777# ¡Aprovecha! <http://mvst.cl/promore>

movistarchile ¿Tienes todos los productos Movistar? Junta tus puntos y canjéalos si el titular de la cuenta es el mismo <http://mvst.cl/reqcanj>

B.3 Opinión de Noticia

fernandopaulsen	RT @VillalobosJara Chile, Uruguay y los negros. Me cagué de la risa. Vía E. Galeano: http://twitpic.com/5n7nt0u //Muy bueno.
fernandopaulsen	RT @carolinagutierr Mall de Valdivia prohíbe los besos. http://yfrog.com/h7jx9nxj //Debería haber una Marcha de los Besos en el mall.
fernandopaulsen	RT @sanfeliu Comercial de Wikileaks que se palanquea a Mastercard, no tiene precio http://t.co/LKCStnF //Brillante.
fernandopaulsen	RT @EduardoBerlin @mwaissbluth ¿Viste el monitoreo español? http://lockerz.com/s/112830335 //jajaja gran monitoreo
ginniasa	- Presidente de comisión por La Polar pide ampliar indagación a todo el retail financiero http://t.co/aJwsbNo vía @biobio / #eso, a todos!!
raul_canete	“@biobio: Presidente de comisión por La Polar pide ampliar indagación a todo el retail financiero http://t.co/CKPIK3J ” no es mala idea
carla_alvarez	biobiochile - Decretan alerta amarilla por sistema frontal en la Provincia de Iquique http://t.co/OMPjyst vía @biobio / HEAVY!!!
RNE_791	Hinchas penquistas siguieron partido de “La Roja” desde kiosko del paseo peatonal http://t.co/MAEsHSe vía @biobio // que buena!!
johnnyfritz	@biobio: Renca: Un fallecido y 2 heridos deja accidente en autopista Costanera Norte http://t.co/MrCUNa2 // una bicoca
Notechdex	RT @biobio: Autoridades de Antofagasta recomiendan a afectados por precipitaciones trasladarse a albergues http://t.co/Ie1oqpU // callate :v
juanbilliard	Polémica por decisión de Meteorología de Chile de sacar de la red página “tiempodelsur.cl” http://t.co/KW1nK8M vía @biobio Es nuestra!!!!
g3raldin3	@biobio PDI recupera notebook robado a hijo de diputada Mónica Zalaquett http://t.co/8XPfo6z / porque ES HIJO d DIPUTADO.!!!
danialeguzman	q raro RT: @biobio: PDI recupera notebook robado a hijo de diputada Mónica Zalaquett http://t.co/y07iYKr
pitufo_acido	RT @biobio: Senador Chadwik (UDI) asegura que no es momento para un cambio de gabinete http://t.co/WP1b1IF //q sae ese gil ohh
NegritaDYchile	OMG RT @biobio: Suben a 71 los fallecidos por cólera en República Dominicana http://rbb.cl/zmb
robertobaeza	Estudiantes de la UACH decidirán el martes si continua con la toma de campus en Valdivia http://t.co/3CrNEFh (vía @biobio) // fuerza cabros
Danodinos	“@Bomberos132: INTERNACIONAL. RT @biobio: Una ola de frío en Sudamérica deja al menos 22 muertos http://t.co/XJzwFQD ” //Por eso hay q ayudar
equisp	“@biobio: Condenan a 7 años de cárcel a sacerdote colombiano por pedofilia http://t.co/AgEums0 ” // y al otro lo mandan a rezar. ..#Karadima
pazitapaz	RT @biobio: Una ola de frío en Sudamérica deja al menos 22 muertos http://rbb.cl/zm5 // brrrr brrr
diegowillbefree	“@biobio: Yumbel organiza III Fiesta del Chanco en Misque http://t.co/E2XWs9s ” // De allá somos.
mauricio_leiva	Recién llegando a ver el partido. En en camino lo venía escuchando gracias a la gran Radio @biobio #CopaAmérica #CA2011 // Vamos Chile!
n_condorcet	RT @biobio: Vecinos y locatarios de Alameda piden a Intendente no autorizar marchas en centro de Santiago http://rbb.cl/zkq / y en vitacura?
mdsaldana	Crean parodia delo que le sucede a Félix Fumastre http://t.co/ZsuQydZ vía @PublimetroChile ... no llamen
Christianx3	Niños que nacen actualmente podrían vivir hasta 130 años http://t.co/AmjfOSj vía @PublimetroChile // OMG o.ó
estebaaaaam	Cumbre "Kitsch" juntará a Delfín, Wendy Sulca y la Tigresa del Oriente http://t.co/6D1e6r6 vía @PublimetroChile Empezaré a juntar plata :D
valecanetevidal	Me acabo d enterar q l alimentacion influye @PublimetroChile: [MUJERES] 6 infalibles para cuidar el pelo en invierno http://t.co/5IQdM8u
catinter	@PublimetroChile: Soledad Onetto se candidatea para ser parte de Tolerancia Cero"

http://bit.ly/noYinQ" // ☺ EXCELENTE!!! @SoledadOnetto

crixmonster Justin Bieber es segundo en ranking de "Mujeres más populares" <http://t.co/mYMn33Y> vía @PublimetroChile Jaj me mate de la risa! Gaga, unica

ramsefchile RT @PublimetroChile @ElGraficoChile: "La gente se merece la copa" <http://t.co/W4gl5LB> // demostramos estar para + cosas...#VIVACHILE

javcollao RT @PublimetroChile Cumbre "Kitsch" juntará a Delfín, Wendy Sulca y la Tigresa del Oriente <http://bit.ly/p17YHT> // junten agua!!!

paulinasalman @PublimetroChile: Soledad Onetto se candidatea para ser parte de Tolerancia Cero" <http://bit.ly/noYinQ> //nop! No es el perfil

Sheiscalled_Kri Sumastre se defiende: "La verdad es más turbia de lo que creen" <http://t.co/epw5aCp> vía @PublimetroChile // Cuaaaaaaticoo

Ercos [VIDEOS] Hinchas chilenos se "aprovechan" de las fans hot de Perú en Copa América <http://t.co/LRmLIQ2> vía @PublimetroChile// Picarones!! xD

romi_fdez @PublimetroChile: Regalarán \$10 milen bencina a conductores sobrios en Bellavista <http://bit.ly/olcYh0/ElColmo>! Presos a los curaos mejor

nata_ortiz RT: @PublimetroChile Regalarán \$10 mil en bencina a conductores sobrios en Bellavista <http://t.co/b9VhGPs> / que interesante!!

gabofigueroaa RT @PublimetroChile: Regalarán \$10 mil en bencina a conductores sobrios en Bellavista <http://bit.ly/olcYh0/wena>!

MauroRAndaur @PublimetroChile Regalarán \$10 mil en bencina a conductores sobrios en Bellavista <http://t.co/yjBphbT> / premio por actuar como corresponde

janaravena Columna de Juan Manual Astorga: "Menos diagnósticos, más soluciones" <http://t.co/h0HYYdC> vía @PublimetroChile // muy bueno

ChrisSepulveda1 Así era la "plantación" indoor que tenían los ex Yingo <http://t.co/ImOvY5s> vía @PublimetroChile pura tecnologia!!!

venom_oroichi RT @PublimetroChile Las 10 lesiones más comunes que provoca el sexo <http://t.co/UTZ8aVA> // faltan las mordeduras y arañazos =)

danayosoriom grande Yingo...semillero de grandes personas... <http://t.co/ETpXvpp> vía @PublimetroChile

freaknesss Giles qls xD RT: @PublimetroChile: Tres ex "Yingo" son detenidos por microtráfico y cultivo de marihuana <http://bit.ly/nSh6CO>

PalomaPopCorn "@PublimetroChile: Iggy Pop... a los 65 <http://t.co/QiOhp2W>"// wakala !! Esta pa la expo bodies

Nossferatux RT @PublimetroChile: Arzobispo Ezzati: "El Papa está preocupado por Chile" por caso Karadima <http://bit.ly/rlnmo> que la justucia actue!!

AscKoT Hinchas chilenos se "aprovechan" de las fans hot de Perú en Copa América <http://t.co/oDHkm8O> vía @PublimetroChile / Típico del Chileno xD

negramarsh Las 10 lesiones más comunes que provoca el sexo <http://t.co/5ABZsts> (vía @PublimetroChile) / No pude evitar reir!!...

MilyFlowers_ esta es la mano pa ganar plata aquí ... Las fiestas de divorcio son furor en Japón <http://t.co/Q4ShOIS> vía @PublimetroChile

La_Milly @PublimetroChile Mañalich afirma que paros estudiantiles ayudaron a frenar virus sincial <http://t.co/n7htTXY> // Pdte veale el lado bueno

Burich11 "@PublimetroChile: [VIDEO] Los Jaivas vuelven a cantar la poesía de Neruda en Machu Picchu <http://t.co/v01Xfi0>" la raaaaaja! Lo mejor!

marcelo_cruz_al Grandes Los Jaivas RT @PublimetroChile: [VIDEO] Los Jaivas vuelven a cantar la poesía de Neruda en Machu Picchu <http://bit.ly/qs6OA>

gleisita "@PublimetroChile: Revista: Justin Bieber es segundo en ranking de "Mujeres más populares" / Es mujer??yo juraba q era hombre :O

rodrigoinverso Ministro Lavín llama a los estudiantes a que lo "dejen trabajar tranquilo" <http://t.co/ZDjHGJu> vía @emol // Pobreshito

zauriel RT @Emol: Ministro Lavín llama a los estudiantes a que lo "dejen trabajar tranquilo" <http://j.mp/kEPy3g> // así con el dialogo

leopoldoquezada RT @MAURMASS: Rara cosa la denuncia contra Martinez por el @mostrador y la defensa por @emol #elmundoalreves

B.4 Opinión General

Emisor	Mensaje
Emol	RT @rodvithousand: falta mas precisión en la entrega #LaRoja #ca2011
Diewo_Z	Mas lenta que cajera del @santanderchile
Hudytan	Cada dia peor @santanderchile , voy a cerrar las tarjetas y ejecutivo ni siquiera pregunta razón, cero interés en retener a clientes...
melanie_goffard	super eficiente el banco @santanderchile , me mandaron las tarjetas nuevas a la casita :)
noesporAzar	El seguro por clonación de tarjeta o como quiera llamarle, es como el pico en @santanderchile. Me clavarón con 3 palos. Respuesta: relájase
alamo_Ka	Pucha que eficiente es mi banco. @santanderchile
PalomaArmijoB	@santanderchile... Uf! Pésimo servicio en sucursales y también a nivel de ejecutivos. Para variar... La verdad no me extraña...
juanpablo	todo es un problema con @santanderchile, siempre. qué desagradable.
xtephan	Señores Banco Santander: Arreglen su cagá de página! Pónganse Serios! No es un Blog! @santanderchile
Movlievka	@santanderchile Finalmente pude hacer la transacción, pero 4 horas despues de que necesitaba hacerla.- Y nadie respode, me parece pesimo!!!
MauricioFunk	hace días que el portal de @santanderchile anda medio reguleque !
Fdoaranc	Me encanta la forma de movilización que ha adoptado Arquitectura y Diseño de la PUCV: creativa, sensata e inteligente.
CuraCatri	Pucv es universidad particular y pública. Y hoy necesita dialogo. Y ser responsables los líderes sociales que son Ejemplo de los estudiantes
verors	No puedo creer lo mal que trabajan las secretarias de @pucv.....toda una semana perdida y los certificados PAGADOS también.
jcklenner	Así no más con Argentina en la Copa, esta desilusionando más que nuestro Monona SADP en el Torneo de Historia PUCV!
situsinversus94	PUCV y sus fichas de historia asquerosas
Cora__D	No debería sorprenderme la marcha de respaldo a Hugo Aguilar, pero es que en serio ya estoy mareada con el himno de Santander
envaladez	Ahora resulta que ya no tienen efectivo en la sucursal de #Santander en Torre de Mexicana. Seguro en España no hacen esto a sus clientes!
envaladez	Que pésimo servicio de #santander, es de los bancos que solo tienen 2 cajas en operación. Ojo @CondusefMX
AdrianVentosa	Increible las fiestas de #sanfermin en #tetuán #santander teneis que pasáros que hasta el día 10 esta todo operativo
mely_cantu	Como siempre, pésimo servicio #santander
Rockmunista_	Vuelven a repetir el reportaje de Obsolescencia programada en la 2. Simplemente magnífico. #recomendar
fernandopaulsen	@__Carolita_ Deliciosos, realmmmmmmmente deliciosos.
Rodribe	Gracias a @biobio por su impecable transmisión del partido. En mi viaje pude disfrutar cada momento junto a sus relatos :)
JuanBeto73	@biobio mas asqueroso qe su espionaje telefónico? Lo dudo
ruarui	que emoción!! fui retuiteada por @biobio
carla_valenx	@muozvaldes @biobio recién salí en auto y me devolví no eran mas de 5 o 7 metros de visibilidad...no se ve nada!!! horrible!!
AdelSanjim	@PDI_CHILE @biobio revise su face book y esta muy bueno.los felicito...sigan adelante
pancholarbarca	@biobio que pasa en costanera? Taco impresionante!!
Stefy_PatitOp	@biobio @KarenTVN @Tvn_Online @tvn_prog Hace 8 Horas En La Posta 3 HSJD Esperando Atencion Para Un Atropellado, QUE VERGÜENZA !
Zurda_K	@biobio pobrecitos :(

Kyba000	FUERA SUAZO!!!! por FAVOR @BIOBIO
josefa78	@biobio puffffffff de que sirve
danitamertino	@biobio Que saquen al arbitro! Es pésimo!
adela36	@biobio q pedazo de pelotudo este arbitro....
PabloMedinaBast	@biobio eso está bueno
jociito	@biobio se viene con furia el 14 !
tanyaureta	@PublimetroChile Chupete juega bien afuera, en Chile se descoloca, no brilla
jcandiat	@PublimetroChile ME GUSTA!!!
krusoe	@PublimetroChile es muy minita (aparte es rica) ...
aslyta	este ha sido el mejor partido de la copa america !!! #chile-Uruguay #vamoschile @PublimetroChile y me encanta borghi, tanto como alexis xD
JosChami	Nunca me aburriré de ver spiderman, amo esa película horriblemente
Lizyii	Es buenisima RT @Joalfred No me canso de ver la pelicula El Ilusionista.
nescavieres	Necesito recargar mi cel y la maldita pagina de @santanderchile me dice que no es posible !!! Rayos
Camivilches	Adorable, @santanderchile me viene a dejar la nueva tarjeta a la casa <3
acortesh_CIMAR	Increible lo lentas que son las cajas de @santanderchile
carlosdvt	@santanderchile sigue cobrandome comisiones por tarjetas sin uso. Que mal banco, mas aun, pesima atencion en sucursal ciudad empresarial
joaquinpineda	Para variar la Pág de @santanderchile está caída, menos mal que nos dicen que ellos se preocupan
paulinasalman	@santanderchile !A los q contraten seguro les dan kms Lan y a los q llevamos pagando años: nada! ni me pagan el fraude en mi cta corriente!
rtellojo	Si llamas al vox de @santanderchile y te dicen que te van a llamar en 20 min... no les crea :/
Myanezb	“@santanderchile: @MarioSchilling Si necesitas algo cuéntanos!”// es chiste?
cosito_panda	pq @santanderchile hace tan rapidas gestiones ante @MarioSchilling (sin ofender) si cuando 1 como comun mortal va al bco se demoran (cont)
gianninadodero	@santanderchile en su comercial de crédito de consumo fomenta el SINDROME DE ALIENACION PARENTAL
gianninadodero	“@santanderchile: @MarioSchilling Si necesitas algo cuéntanos!”/jajaja el poder de las redes sociales... @santanderchile asuman son pésimos!
luis_riquelme	“@alinunez72: “@santanderchile: @MarioSchilling Si necesitas algo cuéntanos!” /// notable este tweet lo q hace una red social !” es conocido
alinunez72	“@santanderchile: @MarioSchilling Si necesitas algo cuéntanos!” /// notable este tweet lo q hace una red social !!
SuarezParrao	RT @MarioSchilling: “@santanderchile: @MarioSchilling Si necesitas algo cuéntanos!” / sí, necesito que me entreguen mi vale vista. Perdón por no ser cliente.
unlocko	@santanderchile mi 2do intento en depositar un fondo mutuo! Si no pasa nada me voy de este banco!! Que x lo visto no necesita mi \$\$\$ #arica
ludwitcha	Muy choriado con @santanderchile #valdiviacl . 4 cajeros para "clientes" , 1 cajero pa'l perraje. 20 minutos pa' que me atiendan. #cortita
viruzck	Es impresentable, que desde ayer en @santanderchile NO esté disponible el pago en linea de tarjeta de crédito internacional
cat_osvaldo	Obtener la "superclave" en @santanderchile es como ir al metro y encontrarlo vacío -> el día del níspero te sucederá.
changachango	@santanderchile es el PEOR banco EVER!
Fefita_Camilla	@santanderchile pésima la atención en la sucursal de providencia con gel del canto llevo 1 hora y media esperando!!
panchoabasolo	@santanderchile ojala se pongan las pilas las ejecutivas de cuenta
andreawahr	En resumen, no estoy conforme con el servicio del banco @santanderchile

B.5 Mensaje Personal

Emisor	Mensaje
santanderchile	Estamos trabajando para solucionar algunos inconvenientes con santander.cl. Esperamos estar operando normalmente a la brevedad.
comunidadpucv	Hoy y mañana en Sausalito estarán entregando TNE que faltan de aquellos estudiantes que recibieron correo DPD. 10 a 13 y 14 a 16.30 hrs.
comunidadpucv	La TNE de todas las carreras las están entregando en campus Sausalito sala B 10
comunidadpucv	RT @eadpucv: Hoy 20:30 h. Exposición dibujos arquitecto Alex Moreno y lectura 3 poemas. En Sala de Música Ciudad Abierta @amereidas
comunidadpucv	RT @ucvradio: Escuela de Ingeniería Mecánica de la PUCV funciona normalmente luego de movilización estudiantil durante esta madrugada
comunidadpucv	Continúan suspendidas las actividades en la PUCV
comunidadpucv	La sesión de España Insólita de hoy lunes 6 de junio se suspende. Reprogramaremos nuestra cartelera y se lo informaremos @cinetecapucv
comunidadpucv	Hola a todos les recordamos que la cuenta @rector_pucv también es falsa. Rectoría no se comunica a través de este medio.
comunidadpucv	El perfil @RectorPUCV es falso. Rectoría no se comunica a través de este medio. #PUCV
comunidadpucv	Las actividades se encuentran suspendidas hasta el lunes en la PUCV. Estaremos entregando más información.
comunidadpucv	Se posterga el estreno del Documental del Mes en @cinetecapucv . Avisaremos la nueva fecha
comunidadpucv	Hoy a las 11.30 Feria Saludable en Curauma, además se contará con atención de salud gratuita. Organiza Vive Salud
comunidadpucv	Llegaron TNE de estudiantes de 1er año y cursos superiores que se matricularon en periodo oficial. Pueden retirar pases en sus UU.AA.
comunidadpucv	RT @ucvradio: Exclusivo: Subsecretario de educación confirma en vivo que TNE será extendida en su vigencia hasta el 15 de julio
comunidadpucv	Hoy Ciclo de Cine en Campus Curauma a las 14 hrs. "Inception" organiza DAE
comunidadpucv	No olviden que el próximo domingo 29 es el Día del Patrimonio Cultural! Aprovechen de visitar lugares emblemáticos de la Región d Valparaíso
comunidadpucv	Hoy, 11hrs Escuela de Periodismo inaugura año académico con clase magistral dictada por director de "El Mostrador", Mirko Macari Squella
comunidadpucv	RT @pucvperiodismo: En pocas horas inauguraremos nuestro año académico con la clase magistral de @mirkomacari, director de El @mostrador
comunidadpucv	Recuerden que a las 12:00 horas el poeta Elicura Chihuailaf estará en Sala Obra Gruesa #PUCV
comunidadpucv	RT @Estudiario: Pronto lanzaremos el sitio de VIVE SALUD donde podrás encontrar consejos y derribarás mitos al respecto.
comunidadpucv	Hoy a las 12hrs el destacado poeta nacional Elicura Chihuailaf, en la Sala Obra Gruesa de la PUCV Av. Brasil 2830, primer piso, Valparaíso
comunidadpucv	Buen día!! Que tengan una excelente semana!!
comunidadpucv	Última película de esta semana "El fotógrafo", con ella finaliza ciclo dedicado a Valparaíso en el cine.19 hrs. Gratis en @cinetecapucv
comunidadpucv	Si quieres conocer nuevas ofertas de empleo sigue a @exalumnosucv
comunidadpucv	Hoy a las 19.00 horas no se pierda lo mejor de CINE CHILENO. "Valparaíso mi amor" en @cinetecapucv . Entrada liberada.
comunidadpucv	RT @convivirmejor: Mañana lanzamos Web de Campaña de Convivencia Escolar en Valparaíso. :)
comunidadpucv	RT @TheMackayschool: Primer concierto de la extensión cultural de la PUCV
comunidadpucv	3 de mayo a las 19:00 horas en auditorio del colegio
comunidadpucv	Hoy además tenemos fútbol en Reñaca. A las 12.30 la PUCV se enfrentará a la UV. Todos invitados a alentar a nuestro equipo.

comunidadpucv Hoy a las 20:00 horas se enfrenta la PUCV v/s la Escuela Naval en voleibol varones en Casa Central. Vengan a apoyar a nuestra selección!!

comunidadpucv Y continúa el deporte... a las 14:00 horas futbol damas, Selección de la #PUCV v/s UMAR en MAGGI

comunidadpucv Hoy nuevamente hay que ir a apoyar a nuestros seleccionados. 14:00 horas tenis damas en Sausalito, PUCV v/s UPLA

comunidadpucv Hoy está toda la Comunidad #PUCV invitada al gimnasio de Casa Central a apoyar al equipo de voleibol damas que se enfrenta a la UPLA. 18.30

comunidadpucv Hoy a las 15.40 se realizará una nueva charla sobre alimentación saludable. "Almuerzo a la Carrera", sala G 1-1 del Edificio Gimpert

comunidadpucv Hoy continúa el ciclo España Insólita con la película Contactos de Paulino Viota Y mañana comienza ciclo de Peter Greenaway @cinetecapucv

comunidadpucv Que tengan una excelente semana Comunidad PUCV!!

comunidadpucv RT @cinetecapucv: Hoy: Estreno en Chile de El Documental del Mes "Su nombre es Sabine". En Valparaíso, 19:00 hrs. Entrada liberada.

comunidadpucv Hoy 20:00 hrs se enfrentan las selecciones de voleibol de la PUCV v/s UNAB por Juegos Dep. Universitarios Navales en Gimnasio Casa Central

comunidadpucv A minutos de comenzar el Foro sobre Educación en la #PUCV...Los esperamos

comunidadpucv Hoy a las 11:45 "Foro sobre Educación" en Salón de Honor de Casa Central. Los esperamos. #PUCV

comunidadpucv RT @Estudiario: Se encuentra disponible dinero de la Beca de Mantención correspondiente a abril

comunidadpucv Hoy la #PUCV saluda anticipadamente a sus funcionarios con actividades desde las 16:00 horas por el Día Internacional del Trabajo.

comunidadpucv Casi listos para comenzar el fin de semana...que tenga un buen día toda la @comunidadpucv . Saludos

comunidadpucv Si estás en Casa Central participa en la charla informativa que el ISEP y nuestro Programa de Intercambios ofrecerán. No faltes!

Emol A partir de las 20:55 estaremos transmitiendo en vivo el anuncio del Presidente Sebastián Piñera sobre sus nuevas políticas educativas

Emol Hoy a las 13:30 horas en Emol, el despacho en vivo de nuestros enviados especiales a Copa América con la previa de Chile-México.

Emol En pocos minutos estaremos conversando con Camila Vallejo sobre marcha de hoy y los alcances del movimiento estudiantil #marchaestudiantil

laterceracom A esta hora, Joaquín Lavín se reúne con Consejo de Rectores para analizar detalle de las propuestas entregadas anoche por Sebastián Piñera

laterceracom Estudiantes del Liceo de Aplicación se toman la sede de Renovación Nacional

laterceracom Ahora habla Sebastián Piñera en La Moneda y entrega más detalles sobre anuncios. "La educación es la madre de todas las batallas", dijo.

laterceracom Lavín: "Jóvenes del 40% de familias con menores recursos que entran a educación superior pasarán de 100 mil a 240 mil en este gobierno"

laterceracom Lavín: "Hay 110 mil jóvenes que están en Dicom porque no han podido pagar su crédito universitario"

laterceracom Lavín entrega más detalles a esta hora de los anuncios: "Queremos que el sistema de financiamiento sea el gran legado del Presidente Piñera"

laterceracom Roberto Cereceda será presentado hoy como refuerzo de Universidad Católica

laterceracom Cifras económicas: Dólar (cierre de ayer) \$464,7 | UF: \$21.904,46 | UTM (julio): \$38.441,00

laterceracom ¿El tiempo para hoy en Santiago? Nubosidad parcial variando a despejado, con 3° de mínima y 14° de máxima, según Dirección Meteorológica

laterceracom Desaprobación de Piñera aumenta seis puntos y llega a 53% en encuesta Cerc

laterceracom Actividad económica registró crecimiento de 7,3% en mayo y superó las expectativas

laterceracom Cifras económicas: Dólar (cierre de ayer) \$465,7 | UF: \$21.901,55 | UTM (junio): \$38.441,00

B.6 Compartir Ubicación / Evento

Emisor	Mensaje
PabloNachoLopez	Ahora el día esta lindo, así que mi animo también (@ Instituto De Biología PUCV) http://4sq.com/r2UWJe
constructorazul	En asamblea Ingeniería en Construcción PUCV.
PabloNachoLopez	Día gris (@ PUCV Campus Curauma) http://4sq.com/o6hcv3
Sikocio	continuando en la asamblea!!! (@ Facultad de Ingeniería PUCV) http://4sq.com/q087ew
Sikocio	viendo futurama esperando la asamblea!! (@ Facultad de Ingeniería PUCV) http://4sq.com/rpTfkp
ale_andres_pm	Trabajando... (@ Campus Sausalito PUCV) http://4sq.com/noFt0z
keeperjn	Ahora protesta de estudiantes en Plaza Anibal Pinto, Valpo #PUCV #USM #UPLA #UV
AnaMier95	En Santander
curropar	800 kilómetros, y aún me quedaban 130. En serio, usad carburantes de Shell o BP, la autonomía se dispara #recomendar http://4sq.com/oqgm1F
jpozuelos	A comer! Entrecot 8,40e #recomendar en Olot (@ Restaurant Can Guix) [pic]: http://4sq.com/jnV7GR
antoniolana	Comiendo con mi dreina #recomendar (@ Balbona) http://4sq.com/qYaa7t
Maitiita	=) (@ Tribal DDB Chile) http://4sq.com/qp6LkF
valehaddad	07/07 lindo día! (@ Ogilvy Chile) http://4sq.com/pOjp67
david_arostica	1 hora esperándote! (@ Biblioteca Nacional de Chile) http://4sq.com/pClp9b
El_Bachi	2 hrs y media y pa' la casita (@ NYK Sudamérica (Chile) Ltda.) http://4sq.com/qtOikx
debsmunoz	2 levantadas mas y es finde :) (@ everis Chile, of. Agustinas) http://4sq.com/nkvxJv
xtephan	2do día Seminario: "La Nueva Institucionalidad Tributaria en Chile: Desafíos y Esperanzas" http://4sq.com/pR3mww
Tocordo	9na despedida de @aycatalina con @delrevez viendo Uruguay-Chile (@ Antares) http://4sq.com/raUOng
jeannesita	Al fin!!.. Vamos Chile!.. (@ Ruby Tuesday) http://4sq.com/neo3CE
cangurosorte	al final no tuve clases... (@ Facultad de Artes Universidad de Chile) http://4sq.com/o4CdeJ
jsaintemarie	Alguna shit pa comer antes del cine (@ Unimarc Chile-España) [pic]: http://4sq.com/r2Vxt6
krlitos83	Alitas pa cenar (@ Chile's Buffalo Wings) http://4sq.com/ogOaVP
lejo_cangrejo	Almorzando algo civilizado (@ Centro Comercial Avenida Chile) http://4sq.com/oAXUYD
fdo_garcia	Almorzando en la previa de Chile-Uruguay (@ Facundo Restaurant) http://4sq.com/pHxCZ9
pabloenjoy	Almorzando en la previa del partido Chile-Uruguay (@ Istanbul w/ @jmvera @rqilodran) http://4sq.com/qAADVm
nicolasopazo	Almorzando más solo que Lavin... (@ Televisión Nacional de Chile) http://4sq.com/pWusHA
Elcapel	Almorzando su terrible de apañadora y yiiica Maruchana! (@ Axis Chile) [pic]: http://4sq.com/n219IQ
gmaluenda	almuerzo y seguir corriendo (@ SUBWAY® Metro Universidad de Chile) http://4sq.com/m8Ff0m
bikecl	Camino a Chile :) (@ Tocumen International Airport (PTY)) http://4sq.com/rr3B22
juanpablo1	Capacitación... :) (@ Samsung Electronics Chile Ltda.) http://4sq.com/pTgMrf
ProfeNavarrete	En casa con mi novia, mi hermano y su novia!!! Vamos Chile #VamosChileCTM (@ Edificio Nataniel Urbano) http://4sq.com/pxVURO
Cristian_NietoE	En Chile :) (@ Aeropuerto de Santiago Arturo Merino Benítez (SCL) w/ 22 others) http://4sq.com/qmmWj2

ovaldene	En clases :((@ DCC, U de Chile) http://4sq.com/pdlR0A
CaritoCat	En clases... (@ Facultad de Economía y Negocios Universidad de Chile w/ 2 others) http://4sq.com/nCpeim
eglezd	En comisión mixta sobre reajuste del salario mínimo. (@ Congreso Nacional de Chile) http://4sq.com/qyJwW
opalmab	En curso de Comunicación Efectiva y Relaciones Humanas (@ Dale Carnegie Chile) http://4sq.com/qpHeU3
Puchilla	En el living de mi casa esperando a la roja!! Vamos Chile mierda!! (@ California Cantina w/ 16 others) [pic]: http://4sq.com/pivg3A
paupolina	en el switch 2 con @natty y @carlosgrage viendo el partido (@ CNN Chile) http://4sq.com/ro0OYN
Daniela_Paz	En extensión horaria, cuantos pacientes llegaran entre tanta lluvia y partido de Chile??? (@ Cesfam Rendic) http://4sq.com/okgpp9
mchaparro	En focos group de puentes UC (@ Campus San Joaquín Pontificia Universidad Católica de Chile) http://4sq.com/nWToy7
ericbaez	en la jornada del Lab de tecnologías móviles del CEC LTE 4G ZTE (@ CEC - FCFM - Universidad de Chile) http://4sq.com/nXuKaD
Nednibz	En la pega (@ Jazzplat Chile) http://4sq.com/ndJWsl
sreyesc	En la previa del partido!! Vamos chile ctm!!! (@ Cassis) http://4sq.com/ntr5NW
JC_CIFUENTES	En las Oficinas de (@ ETC Engineer Chile Ltda.) http://4sq.com/p4tZsg
JC_CIFUENTES	En mi Oficina preparando reunión ;) (@ ETC Engineer Chile Ltda.) http://4sq.com/qX2hyB
vivirocco	En nuevas oficinas :) (@ Ybrant Digital Chile) http://4sq.com/lv7fHI
Jorge_Galvez	En presentación de BI, con mayor rating que el partido de Chile (@ Hotel Marina del Rey) http://4sq.com/nLl6E0
Matucoromina	En Reu (@ Escuela de Servicio Walmart Chile) http://4sq.com/n8SKAf
miguelarias	En reunión con @aprendefilmando (@ SerDigital Iberoamérica - Chile) http://4sq.com/nIWZvk
jorgejorquera	en reunion de estrategia digital (@ Haier Chile) http://4sq.com/oJpXqS
Janohead	En reunión del directorio IAB Chile. (@ TVN w/ 2 others) http://4sq.com/otTEWe
Tavosanchezac	En secretaría (@ Facultad de Derecho, Universidad Central de Chile w/ 2 others) http://4sq.com/nUZ1gQ
xtephan	En Seminario: "La nueva institucionalidad Tributaria en Chile: Desafíos y Esperanzas" http://4sq.com/qBGXHP
caroseg	En toma! (@ Casa Central Universidad de Chile) http://4sq.com/nLFCb
matymad	En una charla de emprendimiento global (@ Colegio de Ingenieros de Chile A.G.) http://4sq.com/pzDZyR
javier_gm	Resultados de los exámenes de sangre.. uuuh (@ Hospital Clínico Universidad de Chile) http://4sq.com/otx8As
Kiquete	Retirando la entrada para Chile Uruguay (@ Club Antonio Tomba) http://4sq.com/pO4uhQ
ptvalenz	Retirando pruebas de Short Story (@ Campus San Joaquín Pontificia Universidad Católica de Chile) http://4sq.com/ph2Pho
aventurachile	Reu con balmaceda (@ Nokia Chile) http://4sq.com/qMO8hs
javcollao	Reunión (@ PUMA Chile S.A.) http://4sq.com/pXaFTo
rolivaresb	Reunión de integración (@ Escuela de Servicio Walmart Chile w/ 2 others) http://4sq.com/ofMkQw
aeyzaguirre	Reunión mode on! (@ Asociación de Guías y Scout de Chile) [pic]: http://4sq.com/nmpQ5D
wkolic	Reunion! (@ Sanofi Chile) http://4sq.com/nl83kn
ritalin	Reuniones de descuento. Se nos viene el findeeeee (@ Laboratorio Chile) http://4sq.com/oRJAq4

B.7 Chat

Emisor

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

santanderchile

Mensaje

@xstephan De internet nos informaron que ahora esta funcionando correctamente.

@DonJuanra Muchas Gracias, nos contactaremos con la sucursal para enviarles tu Reclamo y evitar que vuelva suceder.

@PipeUrrea No, pero si necesitas conectarte al banco desde tu telefono puedes hacerlo en santandermovil.cl. Saludos!

@Palma_fco Nos indican que en este momento el sitio esta funcionando Ok, te agradecemos que entres de nuevo. Saludos!

@pgrez9 En este momento el sitio esta operando correctamente, te agradecemos que vuelvas a ingresar y nos cuentes si tienes algún problema

@tgarnham Hola te confirmamos que el sitio está operando correcatmente en este momento, te agradecemos que ingreses de nuevo y nos cuentes

@miguelburgos Nuestro sitio en estos momentos esta funcionando OK, te agradecemos que ingreses de nuevo, si tienes algún problema cuéntanos

@sebagodoy desde el extranjero llama al (56 2) 320 3000. Si necesitas mas informacion de emergencias bancarias revisa <http://bit.ly/qM92Bi>

@RommyRivas La información que necesitas ya está publicada en nuestro sitio web, puedes revisarla en <http://bit.ly/jkZgeO>

@alvaroneira Que bueno, cualquier cosa que necesites nos cuentas. Saludos y que tengas muy buen día!

@prod_chileno El PC de tu amigo debe estar afectado por un troyano que te redirecciono a un sitio de fraude. Lo que hiciste fue lo correcto.

@CArOlA_KaTiTa Te comunicaste con VOX para saber a qué corresponde el cobro? cuéntanos, para ver en que podemos ayudarte.

@jonyram Para susucibir tu cartola por email debes solicitarlo en cualquiera de nuestras sucursales. Saludos!!

@karenruthh Revisaremos lo que nos indicas. y te avisaremos para que puedas terminar tu inscripción.

@jonyram Para actualizar todos tus datos puedes haerlo en cualquiera de nuestras sucursales.

@Pepe_Schott Muchas Gracias, nos contactaremos con la sucursal para ver si sucede algo en particular.

@diegocajas Si, es solo hasta el 3 de julio!!

@nachobombastic Hola, debes contactarte con VOX al 600 320 3000 para bloquear tu tarjeta y solicitar una nueva. Saludos!

@rodrigo_bustos Hola, revisa el portal de Santander Inversiones,sale todo lo relacionado sobre FFMM, ingresa a <http://t.co/SNt5g6W> Saludos,

@ConyMujicaG hola, el máximo de giro con tu tarjeta de débito en cualquier cajero es de \$200.000 diarios. saludos,

@Pancha74 gracias

@Twin_Birds @santamarii hola pueden cuotizar sus compras Inter. en <http://t.co/xCq42p9> o llamndo a VOX antes del Venc.Estado d Cuenta TC.sl2

@Pancha74 hola,por favor envíalo a redessociales@santander.cl Recuerda,nunca en nuestros mails te pediremos claves ni tendra n links.Saludos

@DonPorotors Hola, qué raro. prueba borrando los temporales, no deberías tener problemas para ingresar a santander.cl desde tu PC. saludos,

@ZettaOne hola, nuestras tarjetas de débito son mastercard, la puedes ocupar en cualquier comercio de transbank. saludos,

@sebarod si puedes llamar al 02 3203000 saludos,

Hola @ZettaOne Te recomomendamos que te comuniques con VOX al 600 320 3000 para que te indiquen como proceder. Saludos.

Hola @anudmand Por ahora puedes conectarte a santandermovil.cl.

santanderchile	Hola @Reinosiano Debes comunicarte con VOX al 600 320 3000 para que te indiquen como proceder. Saludos!!
santanderchile	@SamEstbn Para cambiar tus tarjetas debes solicitarlo en VOX al 600 320 3000.
santanderchile	@diesmori la tarjeta TUI es equivalente a una cuenta Vista.
santanderchile	@loresalvom Te recomendamos que te contactes con VOX para que puedan revisar en línea de que se trata el tema. Saludos y que tengas buen día
santanderchile	@_javicardenas Si necesitas informacion de nuestra Tarjeta Santander Movistar puedes revisarla en http://bit.ly/eElcAy
santanderchile	@nicolicia Estás ingresando a un sitio de fraude, te agradecemos que te contactes con VOX al 600 320 3000 para que te indiquen procedimiento
Emol	@lesliona Muchas gracias. Saludos!! cc @felipegalvez
Emol	@huyoa Ese es El Mercurio, no @emol
Emol	@huyoa La intención es que entre todos conversemos sobre lo que está pasando. Saludos!
Emol	@nicolasantonioo Todo el plantel va a la banca, pero lo más probable es que no juegue. Saludos!
Emol	@nicolasantonioo Estamos confirmándolo
Emol	@natalabbe Por supuesto que nos importa, esto se trata de conversar sobre lo que pasa. Saludos!
Emol	@sebasecas Es la estimación de Carabineros. Saludos!
Emol	@UNPATAN Nuestra información es la que entregó Carabineros. Saludos!
Emol	@jsepulvedax Lo de 50 mil personas es una información de Carabineros. Luego hay que esperar balance final. Saludos! #marchaestudiantil
Emol	@chileeducacion @fraani_ Saludos!
Emol	@RodrigoSegura Son estimaciones entregadas por Carabineros. Saludos!
Emol	@FERNANDASOZA No, sólo intentamos generar una discusión en torno a éste y otros temas contingentes. Saludos!
Emol	@Marcelailapaz Saludos!
Emol	@felipegalvez No no, a ti
Emol	@ericaschilling Qué bien, para eso trabajamos. Saludos!!
Emol	@dan1221ko Aquí está http://j.mp/ioefNb Saludos!
Emol	@ivancaro921 Saludos!!
Emol	@LilySolis Jajaja, qué bueno que estemos conversando ya. Saludos!
Emol	@Manolo_Munoz_G Estilo olla?
juanitoal15	@santanderchile gracias a ustedes :)
geger	@santanderchile gracias por sus ayuda!
gerson_camus	Gracias!!! @santanderchile :)
PalomaArmijoB	@santanderchile no gracias, ya estoy haciendo los trámites para retirarme.
_Natilla	@santanderchile preferiría la transferencia electrónica, no puedo hacer llamadas desde el trabajo. Gracias por la respuesta, de todos modos.
geger	@santanderchile no, para pagar con redcompra!!
tgarnham	@santanderchile ya pude entrar. Gracias
pgrez9	@santanderchile gracias, al parecer el problema fue momentáneo..esta ok.
DaniiDR	@santanderchile Siganme para enviar un mensaje directo :) tengo unas consultas
alvaroneira	@santanderchile Solucionado. Mi ejecutiva me solucionó el problema al instante.
sylviaspikin	@biobio Las culpas son personales.
palobos	@biobio eso se viene.
DiegoFallau	@biobio con toda la razón del mundo!

B.8 Pregunta

Emisor	Mensaje
santanderchile	@Bridevaux hola gustavo, cuéntanos en qué podemos ayudarte?. saludos,
santanderchile	@Rodrigoy punto estimado Rodrigo, cuéntanos en qué podemos ayudarte. Tienes algún problema?. Saludos.
santanderchile	Hola @LuchoValenzuela, cuéntanos, ¿pudiste acceder? no tenemos otros reportes al respecto.
santanderchile	@germansims no puedes acceder a nuestro sitio?
santanderchile	@dpozol no tenemos otro reporte al respecto. Puedes intentarlo de nuevo?
santanderchile	@CatBulsara de qué sucursal es?
santanderchile	@CatBulsara Hola, Gracias por tu comentario!! podrias indicarnos el nombre de tu ejecutivo?
santanderchile	@EdoNamoncura llamaste al 600 20 10000?
santanderchile	@Fefita_Camilla Puedes indicarnos mas detalles?
santanderchile	@Twin_Birds Te refieres a la sucursal en Vitacura 4325?
santanderchile	@CabraLes a Te dieron alguna razón para negar la tarjeta?
santanderchile	@rosarittop No puedes comunicarte con VOX??
santanderchile	@xica_freak si perdiste la TUI puedes solicitar una nueva. Tu Universidad tiene sucursal Santander?
santanderchile	@fransilvag tu Universidad tiene sucursal Santander?
comunidadpucv	@dany serrano Por el momento todas las actividades. Estás interesada en saber de alguna actividad o sede en especial?
comunidadpucv	@Crispitoo Hola! en cual sede te encuentras??
comunidadpucv	@javiernarea Hola! En qué oficina te encuentras?
comunidadpucv	@Triniromerof Hola! acá puedes ver la malla http://bit.ly/IXgmNR qué estudias actualmente y dónde?
Emol	¿Qué les parecieron los anuncios del Presidente?
Emol	@Fariasariel @fraani_ Hola, podríamos publicar la foto? Saludos!
Emol	@nogometa10 en q sentido lo dices? Aalgun ej concreto?
Emol	@cristianfh @nachomunoz Qué necesitan bajar?
Emol	@sduran Todo bien etonces?
movistarchile	@stark José Ignacio, ¿te podemos ofrecer ayuda por este medio? ¡Saludos!
PublmetroChile	¿Se acuerdan del Minidisc de Sony? Bueno, ya no lo fabricarán más http://bit.ly/nyZw6V
joel_____	@santanderchile esta abajo.... o soy solo yo????
n_corsario	es solo a mi o la pagina de @santanderchile esta mala, despues de ingresar las credenciales??
FMaquieira	¿Está con problemas la transferencia electrónica del @santanderchile?
carmen_gloria	@santanderchile necesito bloquear la tarjeta como lo hago?
Hudytan	Existe banco que tenga peor atención de ejecutivos que @santanderchile ?
psepulve	@santanderchile como participo??
FMaquieira	@santanderchile estimados, hasta qué hora puedo hacer un depósito en sucursales? saludos.
geger	@santanderchile cual es el monto máx. ahora para la RedCompra con Uds???
tgarnham	@santanderchile la página no funciona, saben de algún problema?
pgrez9	@santanderchile tengo problemas al acceder a mi cuenta en linea, hay algún problema con la página ???
RommyRivas	@santanderchile gracias, sabes si mandarán un mail a sus clientes?
RommyRivas	@santanderchile tienen info sobre la carrera santander un techo para chile?

elrodrigoc	@javieraisidora alguien puede ayudar a esta cliente ? @santanderchile
goengocal	@santanderchile su página no funciona en Android 2.1 ??? No hay caso que pueda ingresar a través de mi móvil. Saben algo de eso ???
prod_chileno	@santanderchile no lo hice pq eso nunca me habia sucedido... es normal?
gabcarril	son de la pucv??
SebaChandia	@Pabloidc Buenas don Pablo! Acá bien. Disfrutando cada segundo de las vacaciones forzadas. No sabe si hay algo nuevo bajo el sol de la PUCV?
FranDanae	@lia_francisk como está la comunacha más motivá de la PUCV :P jajaja seguís metida en la toma? donde queman las papas? Jijijijji
deletrealopo	@nga2043 licenciado de la PUCV?
SusanCamposQ	@Paulo_CL Cecilia Montt, profesora de la PUCV, explicó que la alianza entre ambas instituciones se suscribió con el objetivo jaja será ella?
SebaFoka	RT @felipesalazar: Ayer pasé por la toma de la PUCV, y el cartel mas estúpido, era uno que se leía así "Mar para Bolivia" Qué tiene que ver eso?
SuiGenerius_20	@_FALOON_ faloon mi U , la PUCV en toma ya un mes...POR UNA EDUCACION DE CALIDAD.. NOS APOYAS?
Super_Lativa	:O cerrarán el semestre si se cumplen 2 meses en toma? ·#PUCV
franperalesf	@wsmithg y qué pasa con las privadas tradicionales con rol público como la PUCV? se debe permitir el lucro en ellas?
Konny_C	Elegí la uai por sobre la uls o la pucv. Despues elegí la unab por sobre la uv o la utalca. Tan mal estoy?
diegojota	@arammys_ Deslegitimizar al CAA EII en el CP frente a toda la PUCV pero pedir disculpar en el edificio?
Enrique_Cadiz	En realidad hay protesta AHORA en Plaza Anibal Pinto #Valapraiso #Valpo #PUCV ???????
MiLy_Vargas	@_c0te_ es en serio??? Cm es eso del impuesto a las privadas??? Y a las privadas cn fin publico cm la #PUCV???
mel2kill	@nediaz y privadas como la puc? la pucv? la utsfm?
acaso_importa	@miguelfigo amigo y como fue la experiencia en Santander? como lo trato el regional!
mrsalvaje	Alguien tiene referencias sobre unicaja? El mec obliga casi a pedir préstamo al Santander, pero se puede tb con aquella
CarraRuiz	@jjperajo mañana tengo boda en santander, hasta cuando estás? El finde que viene si que estoy a tope dando la cara jejeje
alexqardenaz	alguien tiene una tarjeta de credito #santander me urge!!
cherrylips9o	@Luuiii9o ya, ya lo vi. En Santander donde esta Sephora?
vanguardiacom	¿Creen ustedes que el Santander Fashion Week influye en el desarrollo de la región?.Su opinión es #Vanguardia
grumpyromero	@vivoencancun alguien sabe a que hora cierran el banco santander ????
JManuelParada	RT @Pamfie: @JManuelParada que haces Conde de santander??
vanesa_sina	Alguna mcflyer de santander?
ISEPnews	@coras_ca a vosotros! ¿Conocéis blogs de formación en psicología para recomendarnos? #ayuda #recomendar
aramdaucik	@salsadechiles alguna recomendación para cenar en una terraza (atico) en madrid capital? #recomendar
cinefilo	¿ #recomendar MyISAM o InnoDB ? Me interesa MyISAM por velocidad SELECT e InnoDB por integridad referencial, ¿ sugerencias ?
cmorer	¿Alguien sabe dónde comprar pizarras (de las negras) a medida en Madrid? @recomendar #recomendar :))
finareta	¿alguien sabe quién va ganando, @iescolar o @cccarnicero? ¿dónde lo puedo ver en streaming? #recomendar
Stulka	¿Algún buen documental de sectas destructivas? #recomendar