

# UNIVERSIDAD NACIONAL AGRARIA LA MOLINA

ESCUELA DE POSGRADO

MAESTRÍA EN ESTADÍSTICA APLICADA



PROYECTO DE TESIS :

## PERCEPCIÓN DEL CONSUMIDOR LIMEÑO SOBRE LAS TIENDAS POR DEPARTAMENTOS EN LA RED SOCIAL TWITTER USANDO ANÁLISIS DE SENTIMIENTOS

Presentado por  
*Jaime Gómez Marín*

Docente  
Ph.D. Augusto BERNUY ALVA

La Molina, 2019

# Índice general

<b>1. Introducción</b>	<b>2</b>
<b>2. Justificación de la investigación</b>	<b>5</b>
<b>3. Objetivos de la investigación</b>	<b>6</b>
3.1. Objetivos Generales . . . . .	6
3.2. Objetivos Especificos . . . . .	6
<b>4. Formulación de hipótesis</b>	<b>7</b>
<b>5. Marco Teórico</b>	<b>8</b>
5.1. Twitter . . . . .	8
5.2. Data Mining . . . . .	8
5.3. Text Mining . . . . .	9
5.4. Análisis de Sentimiento . . . . .	10
5.4.1. Técnicas de Análisis de Sentimiento . . . . .	10
<b>6. Metodología</b>	<b>11</b>
<b>7. Cronograma</b>	<b>12</b>
<b>8. Presupuesto</b>	<b>13</b>
<b>9. Colaboradores</b>	<b>14</b>

# Capítulo 1

## Introducción

En los últimos 10 años, la revolución que ha causado el empleo de teléfonos inteligentes en la población ha permitido masificar el uso del internet y podríamos decir sin temor a equivocarnos, la democratización de su uso. La consecuencia es que las personas están más comunicadas y/o conectadas en tiempo real a través de comunidades en internet, las cuales son más conocidas con el nombre de “redes sociales”, el caso más emblemático de su empleo en una revolución política sucedió entre los años 2011 y 2012, con el movimiento denominado “Primavera Árabe” (1), donde las redes sociales tuvieron una vital importancia para la organización de este movimiento, siendo las redes sociales más empleadas Facebook y Twitter.

Twitter como se mencionó anteriormente es una red social donde un usuario, que previamente se ha registrado, puede enviar mensaje de texto sobre un tema que él considere importante o puede interactuar con otros mensajes dando a conocer su punto de vista; el texto enviado es conocido como tweets y tiene un máximo de 280 caracteres. En la página de Twitter, la empresa se define a sí misma con las siguientes palabras “lo que está pasando en el mundo y los temas sobre los que está hablando la gente.” (2). Un componente opcional en los tweets es que incluye su geolocalización, lo cual es una pieza importante en el análisis del comportamiento porque nos permite determinar la opinión de las personas por zonas geográficas.

La situación en el Perú no es ajena, de acuerdo al estudio de mercado realizado por la consultora CPI (Compañía Peruana de Estudios de Mercados y Opinión Pública S.A.C.) sobre el uso de redes sociales en Lima de agosto del 2018, se evidencia el uso de Facebook como la red social mas popular con el 72.7%, seguido de WhatsApp al 68.5%, Instragram al 25.0% y como



Figura 1.1: Redes Sociales mas usada por los limeños - Agosto 2018 - Fuente : CPI

cuarto puesto a Twitter con el 12.7 %.

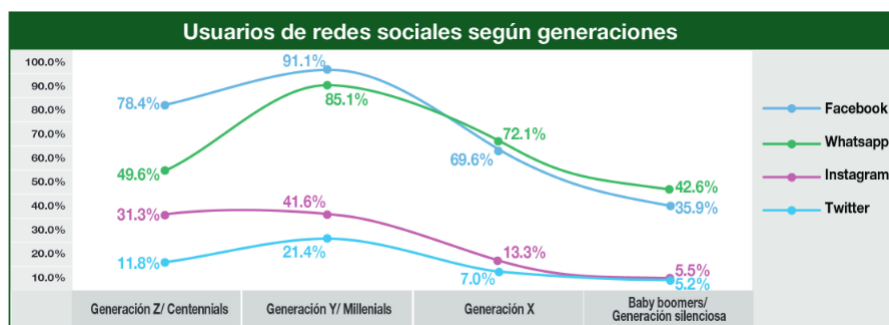


Figura 1.2: Uso de redes sociales según generación - Agosto 2018 - Fuente : CPI

## Capítulo 2

### Justificación de la investigación

Actualmente no existen estudios sobre las preferencias de los limeños sobre las tiendas por departamentos empleando técnica de Análisis de Sentimientos aplicados a la red social Twitter, esto se ha verificado al realizar una búsqueda en el banco de datos de trabajos de investigación del SUNEDU.

El estudio permitirá entender las opiniones que tienen los consumidores sobre los centros comerciales agrupadas por su localización geográfica dentro de Lima, para tal fin se ha considerado las 4 divisiones que tiene Lima: Lima Sur, Lima Norte, Lima Este y Lima Centro.

El estudio tiene como finalidad crear una aplicación informática que pueda ser usado por las empresas para desarrollar estrategias de mercado orientado a conocer a sus potenciales consumidores.

## Capítulo 3

# Objetivos de la investigación

### 3.1. Objetivos Generales

Clasificar las opiniones de los consumidores aplicando análisis de sentimiento sobre los centros comerciales a través de los tweets generados por los usuarios.

### 3.2. Objetivos Especificos

- Uso de herramientas de Machine Learning para el Análisis de Sentimiento en la red Social Twitter.
- Recolección y tratamiento de los datos de la red social Twitter.

## Capítulo 4

### Formulación de hipótesis

La hipótesis consiste en que es posible aplicar el análisis de sentimiento en la red social Twitter sobre las opiniones de la población limeña de los más importantes centros comerciales de la Ciudad para poder ser usado en estrategias de mercado.



# Capítulo 5

## Marco Teórico

Se definen las tecnologías y conceptos teóricos que se utilizarán en el desarrollo del presente trabajo.

### 5.1. Twitter

Es una plataforma de servicio de microblogging que permite el envío de mensajes de texto en un tamaño máximo de 280 caracteres denominados tweets, los usuarios pueden subscribirse a los tweets de otros usuarios y hacerles seguimientos, esta acción es conocida como seguidores o followers. Por defecto los mensajes son públicos y pueden contener la zona geográfica de la persona que ha enviado el tweet.

### 5.2. Data Mining

El Data Mining es un campo de la estadística y las ciencias de la computación relacionado con el proceso de descubrir patrones en grandes volúmenes de conjuntos de datos. Utiliza métodos de inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos

La forma para aplicar Data Mining se basa en las siguientes etapas:

- Comprensión del Negocio.
- Comprensión de los Datos.

- Preparación de los datos.
- Modelado
- Evaluación
- Desarrollo

### 5.3. Text Mining

Es el análisis de información no estructurada, la cual se puede encontrar en redes sociales, para tal fin se emplea técnicas de lingüística, modelamientos estadísticos y técnicas de aprendizaje para descubrir conocimientos que no existen explícitamente en ningún texto de la colección, pero que surgen al relacionar el contenido de muchos de ellos.

Se suelen aplicar a encuestas de opinión, encuestas de satisfacción, libros de reclamación, etc.

La forma para aplicar Text Mining se basa en las siguientes etapas:

- Preparar texto para el análisis
- Extraer conceptos
- Aplicar el análisis de enlace de texto
- Construir categorías
- Desplegar modelos predictivos

Los beneficios del uso del text mining es que nos permite identificar hechos o datos puntuales a partir del texto de los documentos, agrupándolo en clustering e igualmente determinar el tema o temas tratados en los documentos mediante la categorización automática de los textos y crear redes de conceptos.

- El text mining se puede aplicar en:
- Resumen automático de textos

- Detección de fraudes
- Tendencial electorales
- Análisis de Sentimiento
- Clasificación de textos.

## 5.4. Análisis de Sentimiento

Es el uso del procesamiento del lenguaje natural, análisis de texto y lingüística computacional para identificar y extraer información subjetiva de los recursos. Desde el punto de vista de la minería de textos, el análisis de sentimientos es una tarea de clasificación masiva de documentos de manera automática, en función de la connotación positiva o negativa del lenguaje ocupado en el documento. Es importante mencionar que estos tratamientos generalmente "se basan en relaciones estadísticas y de asociación, no en análisis lingüístico".

### 5.4.1. Técnicas de Análisis de Sentimiento

Según Medhat, las principales técnicas de análisis de sentimiento se dividen en dos grandes grupos: las que se basan en aprendizaje automático (machine learning approach) y las que se basan en diccionarios (lexicon-based approach).

La siguiente figura muestra las 2 técnicas:

Una vez mostrado los métodos de análisis, en la presente investigación se usará la estrategia de los algoritmos de aprendizaje automático que tiene relación con el campo de la informática y más en concreto, de la inteligencia artificial.

El procesamiento de datos es a través del aprendizaje para tal fin se extrae patrones de comportamiento a partir de las entradas recibidas, y en base a dicha información aprendida o asimilada, realice la evaluación de nuevas entradas. Los algoritmos internos que constituyen la base de este aprendizaje tienen un fuerte componente estadístico y algebraico, con la consiguiente capacidad de cálculo.

# Capítulo 6

## Metodología

# Capítulo 7

## Cronograma

# Capítulo 8

## Presupuesto

# Capítulo 9

## Colaboradores