



INFOTEC CENTRO DE INVESTIGACIÓN E  
INNOVACIÓN EN TECNOLOGÍAS DE LA  
INFORMACIÓN Y COMUNICACIÓN

DIRECCIÓN ADJUNTA DE INNOVACIÓN Y  
CONOCIMIENTO  
GERENCIA DE CAPITAL HUMANO  
POSGRADOS

## EXTRACCIÓN DE TÓPICOS EN COMENTARIOS DE

Solución Estratégica

Que para obtener el grado de MAESTRO EN  
CIENCIA DE DATOS E INFORMACIÓN

Presenta:

**Dulce Maria Reyes Lucas**

Asesor:

**(Nombre y Apellidos)**

Ciudad de México, junio, 2023.

## **Autorización de impresión**

## Agradecimientos

Agradezco a Dios por guiarme y darme fortaleza en cada paso que doy.

A mis padres por ser las personas que me han apoyado incondicionalmente y dado todo lo necesario, por haberme guiado y hacer una persona exitosa y de bien.

A todas las personas que de una y otra forma me apoyaron para poder lograr mis objetivos.

A todos los profesores que me han guiado con su conocimiento y apoyado, a esta institución que ha confiado en mí y me otorgo un lugar en esta reconocida maestría.

## Tabla de contenido

Tabla de contenido.....	5
Introducción.....	1
Capítulo 1. Objetivos y contenido.....	2
1.1 Tipos de aprendizaje.....	3
1.1.1 Aprendizaje supervisado.....	3
1.1.1.1 Árboles de decisión.....	3
1.1.1.2 Clasificación Naïve Bayes.....	4
1.1.1.3 Regresión por mínimos cuadrados.....	4
1.1.2 Aprendizaje no supervisado.....	5
1.1.2.1 Algoritmos de clustering.....	6
1.1.2.2 Componentes principales.....	6
1.1.2.3 Singular Value Decomposition (SVD).....	7
Capítulo 2. Fuente de datos y preprocesamiento.....	6
2.1 Nulla interdum ex erat.....	8
2.1.1 Mauris et nisl tempus, fringilla.....	9
Capítulo 3. Pellentesque mi mauris, tempor vel risus nec, cursus viverra mi. Aenean scelerisque at orci et llaamcorper.....	9
3.1 Cras velit purus, commodo condimentum finibus lobortis, interdum at lacus ..	9
3.1.1 Integer sed ipsum nisl. Sed sed blandit orci.....	10
Conclusiones.....	12
Bibliografía.....	13
ANEXO 1.....	15

Índice de términos.....	16
-------------------------	----

## Índice de figuras

## Índice de gráficos

## Índice de cuadros



## Siglas y abreviaturas

<b>LDA</b>	Latent Dirichlet Allocation
<b>SVD</b>	Singular Value Decomposition
<b>NPS</b>	Net Promoter Score

## Glosario

### “A”

**Aenean:** Sollicitudin sem lorem, nec tristique lacus auctor in. Etiam luctus purus non dui fringilla tempor. Suspendisse euismod urna id nisl cursus, nec tincidunt lacus sagittis. Sed mollis sem mattis ligula rutrum scelerisque. Ut mattis condimentum blandit. Curabitur ipsum mauris, ullamcorper non accumsan id, eleifend id velit. Nunc at semper massa, sit amet pellentesque metus. Pellentesque pulvinar eget mauris sit amet dapibus. Integer vel lorem ut sem pretium semper vel at est. Aenean vitae varius libero. Sed accumsan nisl eu nulla consectetur fringill.

### “B”

**Blandit:** Aenean laoreet ligula massa, ut varius lorem convallis ut. Integer at libero neque. Donec vestibulum neque in efficitur iaculis. Donec eros eros, porta suscipit auctor et, rutrum ut tortor.

**“C”**

**Congue:** Quam nibh convallis mauris, et tincidunt massa dolor maximus mauris. Fusce pretium lectus vitae aliquet aliquet. Fusce elit ligula, cursus eu velit eu, maximus tempus lorem.

## Introducción

En los últimos años se ha presenciado el increíble crecimiento y uso de las redes sociales, blogs y demás medios que guardan principalmente textos, estos datos no estructurados se han convertido en el mayor interés de empresas principalmente, debido a que los datos estructurados no son capaces de mostrar tal juicio o sentimiento como se describe en los comentarios de las personas que expresan su sentir respecto a una infinidad de temas.

La pandemia obligó a todos los sectores a evolucionar y ofrecer mejores servicios y a la distancia de un clic, poder realizar pagos, recargar saldo, consultar los movimientos de su cuenta, obtener un estado de cuenta, hacer una cita, renovar un servicio, pero principalmente a los bancos, de poder ofrecer la mayoría de los servicios sin necesidad de acudir a una sucursal, por ello, se requiere de una aplicación móvil cada vez más robusta, intuitiva, que cubra estas necesidades y que deje satisfecho al usuario.

Derivado de esta necesidad se presenta un proyecto que analizará los comentarios para identificar las debilidades de la aplicación móvil de Banco Santander.

# Capítulo 1

## Objetivos y contenido



## Capítulo 1. Objetivos y contenido

El objetivo principal del proyecto es implementar un análisis de sentimiento y modelos de texto de las reseñas en Google Play y Apple Store orientado a la aplicación móvil de Santander para detectar las debilidades y fortalezas de las funcionalidades de dicha aplicación.

Para lograr el objetivo principal se plantean una serie de objetivos específicos:

- 1) Realización de web scrapping hacia las tiendas de aplicaciones con Python y posteriormente el procesamiento para obtener un conjunto de documentos listos para poder analizarlos de manera correcta.
- 2) Aplicación del algoritmo de la distancia de Levenshtein para corregir las palabras mal escritas por el usuario.
- 3) Creación de modelos de texto para detectar los puntos de dolor en las funcionalidades de las aplicaciones móviles bancarias.
- 4) Generación de un modelo de análisis de sentimiento para detectar la polaridad de los comentarios de las tiendas de aplicaciones dado que la calificación por estrellas no aporta mucho sentido ni significado.
- 5) Categorización de los tópicos generados por funcionalidad.
- 6) Presentar los resultados y conclusiones a través de un dashboard en power bi.

Con el proyecto propuesto se pretende obtener los tópicos que describan las funcionalidades de la aplicación móvil de Santander, sus fortalezas y debilidades para potenciales mejoras, además, describir la polaridad de estos tópicos de acuerdo con sentimiento expresado en los comentarios para poder enfocar los esfuerzos en aquellos tópicos que tienen una polaridad mayormente negativa. Los resultados de esta investigación ayudarán a enfocar las estrategias de mejora de la aplicación del sector bancario pues actualmente me encuentro laborando en esta empresa y pretendo aplicar el análisis en mi actual posición.

El desarrollo de este proyecto será mediante la metodología KDD(Knowledge Data Discovery), en mi experiencia estoy familiarizada con la metodología SEMMA, sin embargo, me gustaría utilizar esta metodología.

Esta metodología considera los siguientes aspectos:

**Selección de los datos:** Los datos que usaré en este proyecto se encuentran en las tiendas de aplicaciones de Google y Apple Store, la forma de extracción será a través de una técnica llamada Web Scrapping y podremos extraer datos históricos. El proceso ya lo tengo listo.

**Preprocesamiento de los datos:** En esta parte de la metodología se realizará la limpieza de los datos, en la que se realizarán tareas como conversión a minúsculas, eliminación de acentos y caracteres especiales, stemming, stopwords, aplicación del algoritmo de Levenstein para la corrección de las malas escrituras, dejar palabras con una longitud mínima de 5 caracteres e identificar entidades. Generar diccionarios propios de stopwords y si es posible de sinónimos.

**Transformación de los datos:** Generar y entrenar modelos de extracción de tópicos y modelos de clasificación de sentimiento para alcanzar los objetivos.

**Interpretación y evaluación:** Para este paso se compararán los resultados obtenidos en los modelos de sentimiento con matriz de confusión o alguna otra métrica de desempeño. Mostrar los tópicos finales que tengan una interpretación más clara. Los resultados serán mostrados en un dashboard en la herramienta de Power BI.

## 1.1 Tipos de aprendizaje

Dentro de la ciencia de datos encontramos dos tipos de aprendizaje que son los que se aplican en los diferentes análisis en la generación de modelos e implementaciones.

### 1.1.1 Aprendizaje supervisado

El aprendizaje supervisado proviene de los ejemplos etiquetados en un conjunto de datos de entrenamiento en donde el analista provee al algoritmo un dataset conocido con entradas y salidas deseadas. El caso más simple es un dataset de clases positivas y negativas, después vienen las etiquetas multiclase, por ejemplo, positivo, negativo y neutro; y al final regresiones que vienen con entradas de datos continuos.

Por ejemplo, si queremos aprender la clase C de una familia de coches, nosotros tenemos un conjunto de autos ejemplo entonces encuestamos a un conjunto de personas, estas personas etiquetaran las imágenes, los carros que ellos creen son de la misma familia serán los casos positivos y el resto los negativos. Dadas estas clasificaciones se puede hacer una predicción, dado un ejemplo de carro que no ha sido visto antes por el modelo y basado en lo aprendido podemos identificar la clase a la que pertenece. Para que aprenda el modelo se puede añadir diferentes variables (atributos) para la correcta identificación del modelo.

#### 1.1.1.1 Árboles de decisión

Un árbol de decisión es una estructura jerárquica que implementa la estrategia de divide y conquista; se trata de un método no paramétrico. En este método la tarea es analizar los datos y clasificarlos, donde el modelo o clasificador es construido para predecir etiquetas de clase (categóricas), tanto como cuan seguro o riesgoso es un caso de fraudes en préstamos bancarios. Estas categorías pueden ser representadas a través de valores discretos donde el orden entre los valores no tiene significado alguno; por ejemplo, los valores 1, 2 y 3 podrían representar tratamiento A, tratamiento B y tratamiento C donde no hay un orden implicado entre un grupo de tratamientos.

La clasificación de datos es un proceso de dos pasos, consistiendo en un primer paso de "aprendizaje" donde el modelo de clasificación es construido y un segundo paso de "clasificación"

donde el modelo es usado para predecir nuevas etiquetas de clase. En el paso de aprendizaje o fase de entrenamiento donde un algoritmo de clasificación construye el clasificador a través de analizar o aprender del conjunto de entrenamiento hecho de un conjunto de tuplas y su asociada etiqueta de clase; una tupla  $X$  es representada por un vector de atributos  $n$ -dimensional  $X=(x_1, x_2, x_3, \dots, x_n)$ .

La estructura de un árbol de decisión es similar a un diagrama de flujo en donde cada nodo interno denota una prueba en un atributo, cada rama representa una salida de la prueba y cada nodo hoja mantiene una clase, el nodo más alto es la raíz del árbol.

El algoritmo comienza con una partición de datos de entrenamiento, una lista de atributos y un método de selección de atributos que representa un procedimiento heurístico para la selección de atributos que mejor discrimine las tuplas dadas acorde a la clase, este método puede ser Information Gain o Gini Index; criterios como Gini Index orilla el árbol resultante a ser binario, otros, como Information Gain permiten dos o más ramas crecientes desde un nodo; al aplicar un método de particionamiento este nos dará el mejor atributo para continuar las divisiones. Cuando el árbol de decisión es construido muchas de las ramas reflejarán anomalías en los datos de entrenamiento por ello se debe realizar la poda.

### 1.1.1.2 Clasificación Naïve Bayes

Los clasificadores Bayesianos con clasificadores estadísticos, estos pueden predecir probabilidades, como la probabilidad de que una tupla pertenezca a una clase específica, se asume que el efecto del valor de un atributo en una clase dada es independiente de los valores de otros atributos, este supuesto es llamado independencia condicional de clase.

A continuación, se listan los pasos que hay que realizar para poder utilizar el algoritmo Naive Bayes en problemas de clasificación como el mostrado en el apartado anterior.

- Convertir el conjunto de datos en una tabla de frecuencias.
- Crear una tabla de probabilidad calculando las correspondientes a que ocurran los diversos eventos.
- La ecuación Naive Bayes se usa para calcular la probabilidad posterior de cada clase.
- La clase con la probabilidad posterior más alta es el resultado de la predicción.

Aunque son unos clasificadores bastante buenos, los algoritmos Naive Bayes son conocidos por ser pobres estimadores, la presunción de independencia Naive muy probablemente no reflejará cómo son los datos en el mundo real. Cuando el conjunto de datos de prueba tiene una característica que no ha sido observada en el conjunto de entrenamiento, el modelo le asignará una probabilidad de cero y será inútil realizar predicciones.

### 1.1.1.3 Regresión por mínimos cuadrados

La regresión es una herramienta de modelación común para modelar la relación entre algunas variables explicativas y algunas reales o la variable dependiente; el método de los mínimos cuadrados se utiliza para calcular la recta de regresión lineal que minimiza los residuos, esto es, las diferencias entre los valores reales y los estimados por la recta, cuando hay varias variables



independientes nos encontramos ante un modelo de regresión lineal múltiple, mientras que cuando hay solo una hablaremos de la regresión lineal simple.

La regresión lineal requiere que la relación entre las variables sea lineal y puede representarse mediante la ecuación de la recta  $Y = \beta_0 + \beta_1 X$

son los coeficientes del modelo de regresión.  $\beta_0$  representa la constante del modelo (también llamada intercepto) y es el punto donde la recta corta el eje de ordenadas (el de las Y, para entendernos bien). Representaría el valor teórico de la variable Y cuando la variable X vale cero.

Por su parte,  $\beta_1$  representa la pendiente (inclinación) de la recta de regresión. Este coeficiente nos dice el incremento de unidades de la variable Y que se produce por cada incremento de una unidad de la variable X.

El problema es que la distribución de valores no se va a ajustar nunca de manera perfecta a ninguna recta así que, cuando vayamos a calcular un valor de Y determinado ( $y_i$ ) a partir de un valor de X ( $x_i$ ) habrá una diferencia entre el valor real de  $y_i$  y el que obtengamos con la fórmula de la recta. Ya nos hemos vuelto a encontrar con el azar, nuestro compañero inseparable, así que no tendremos más remedio que incluirlo en la ecuación:  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

El componente determinista lo marcan los dos primeros elementos de la ecuación, mientras que el estocástico lo marca el error en la estimación. Los dos componentes se caracterizan por su variable aleatoria,  $y_i$  y  $\varepsilon_i$ , respectivamente, mientras que  $x_i$  sería un valor determinado y conocido de la variable X.  $\varepsilon_i$  representa la diferencia entre el valor real de  $y_i$  en nuestra nube de puntos y el que nos proporcionaría la ecuación de la recta (el valor estimado, representado como  $\hat{y}_i$ ). Podemos representarlo matemáticamente de la siguiente forma y se conoce como residuo:  $e_i = y_i - \hat{y}_i$



### 1.1.2 Aprendizaje no supervisado

Aprendizaje no supervisado es sinónimo de clúster, el proceso de aprendizaje es no supervisado desde que los ejemplos de entrada no están etiquetados, típicamente, nosotros podríamos usar clustering para descubrir clases o grupos en los datos, patrones, por ejemplo, un modelo podría tomar como entrada un conjunto de imágenes de dígitos escritos, suponemos que encuentra 10 clúster tendríamos que describir cada grupo.

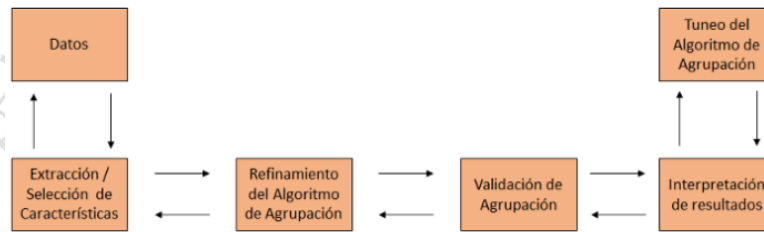
Las principales aplicaciones de aprendizaje no supervisado son:

- Segmentación de conjuntos de datos por atributos compartidos.
- Detección de anomalías que no encajan en ningún grupo.
- Simplificación de datasets agregando variables con atributos similares.

Estas técnicas se pueden condensar en dos tipos principales de problemas que el aprendizaje no supervisado trata de resolver. Estos son los problemas:

- Agrupación
- Reducción de la dimensionalidad

El proceso general que seguiremos al desarrollar un modelo de aprendizaje no supervisado se puede resumir en el siguiente cuadro:



### 1.1.2.1 Algoritmos de clustering

En términos básicos, el objetivo del clustering es encontrar diferentes grupos dentro de los elementos de los datos. Para ello, los algoritmos de agrupamiento encuentran la estructura en los datos de manera que los elementos del mismo clúster (o grupo) sean más similares entre sí que con los de clústeres diferentes.

Es el proceso de particionar el conjunto de datos en subconjuntos, cada subconjunto es un cluster, tal que los objetos dentro de un cluster son similares, comparten características y son disimilares entre clusters. Hay varios algoritmos de clustering, por ejemplo:

**Métodos de particionamiento:** Dado un conjunto de  $n$  objetos, un método de particionamiento construye  $K$  particiones de los datos, donde cada partición representa un cluster, los métodos de particionamiento adoptan la separación exclusiva, es decir, cada objeto solo puede pertenecer a un cluster, muchos de estos métodos están basados en la distancia.

**Métodos jerárquicos:** Estos crean una descomposición jerárquica de los datos y puede ser clasificado como aglomerativo o divisivo. El acercamiento aglomerativo comienza con cada objeto formando un grupo separado, este sucesivamente une los objetos o grupos cercanos el uno del otro hasta que todos los grupos son unidos en uno solo o hasta que se cumplan una determinada condición. El acercamiento divisivo comienza con todos los objetos en el mismo cluster y en cada iteración sucesiva un cluster es dividido en clusters más pequeños hasta que eventualmente cada objeto este en un cluster o que las condiciones de terminación se cumplan.

Un algoritmo muy conocido y común es K means.

### 1.1.2.2 Componentes principales

Análisis de componentes principales o PCA por sus siglas en inglés es un método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez que conserva su información. Supóngase que existe una muestra con  $n$  individuos cada uno con  $p$  variables ( $X_1, X_2, \dots, X_p$ ), es decir, el espacio muestral tiene  $p$  dimensiones. PCA permite encontrar un número de factores subyacentes ( $z < p$ ) que explican aproximadamente lo mismo que las  $p$  variables originales. Donde antes se necesitaban  $p$  valores para caracterizar a cada individuo, ahora bastan  $z$  valores. Cada una de estas  $z$  nuevas variables recibe el nombre de componente principal.

Cada componente principal se obtiene de una combinación lineal de las variables reales, el primer componente tendrá la mayor varianza, el proceso es, primero se resta a cada valor la media de la variable a la que pertenece para conseguir que tengan una media de cero después se resuelve el problema de optimización con el que se maximiza la varianza, una forma de hacerlo es con el cálculo de los eigenvalores y eigenvectores de la matriz de covarianzas.

Tanto la proporción de varianza explicada como la proporción de varianza explicada acumulada son dos valores de gran utilidad a la hora de decidir el número de componentes principales a utilizar en los análisis posteriores, la suma de la proporción de varianza explicada por los componentes es 1.

Al término del cálculo de los componentes se preguntará cuántos son los ideales y que me reducen la dimensionalidad de los datos, una forma consiste en evaluar la proporción de varianza explicada acumulada y seleccionar el número de componentes mínimo a partir del cual el incremento deja de ser significativo, en general tendremos  $n-1$  o  $p$  componentes dada una matriz de dimensiones  $n \times p$ .

### 1.1.2.3 Singular Value Decomposition (SVD)

Se trata de una factorización de esa matriz (original) en tres matrices  $A=UWV^t$ . Primero se calculan los valores singulares, haciendo  $AA^t$ , después se encuentran los vectores singulares. SVD se puede considerar como un método de proyección en el que los datos con  $m$  columnas (características) se proyectan en un subespacio con  $m$  o menos columnas, conservando la esencia de los datos originales. El SVD se usa ampliamente tanto en el cálculo de otras operaciones matriciales, como la matriz inversa, como también como un método de reducción de datos en el aprendizaje automático. En relación con el modelado de texto apoya con los problemas de alta dimensionalidad y dispersión; se define  $A$  como la matriz de término documento con  $m$  documentos y  $n$  términos, típicamente habrá más términos que documentos en el corpus, los valores singulares nos dan una medida de la importancia usada para decidir cuantas dimensiones mantener en la matriz.

The background features a complex, abstract design. On the left, there are several concentric, semi-circular gear-like structures with various internal patterns, including dashed lines and small rectangular notches. To the right of these, there are several horizontal and vertical lines, some with small circles or dots at their ends. A series of small, right-pointing triangles are arranged in a horizontal row. Another set of small triangles is arranged in a vertical column. There are also some dashed hexagonal shapes and a series of small circles at the bottom right. The overall style is technical and geometric, using a light gray color for the background elements.

## **Capítulo 2**

# **Preprocesamiento de los datos**

## Capítulo 2. Fuente de datos y preprocesamiento

La fuente de datos para este proyecto se encuentran en las tiendas de aplicaciones de Google y Apple Store, estos son las reviews (comentarios) de la aplicación Móvil de Santander México “SuperMóvil”, la forma de extracción será mediante la técnica de web scraping, en Python se apunta a la tienda de Google y de Apple Store a los ids mx.bancosantander.supermovil y id498944221 respectivamente que pertenecen a la aplicación SuperMóvil.

### **\*\*WEB SCRAPPING : GOOGLE PLAY STORE\*\***

```
#import play_scraper
import pandas as pd

import json

from tqdm import tqdm

import seaborn as sns
import matplotlib.pyplot as plt

from pygments import highlight
from pygments.lexers import JsonLexer
from pygments.formatters import TerminalFormatter

from google_play_scraper import Sort, reviews, app, reviews_all

import datetime
```

```
app_santander="mx.bancosantander.supermovil"
```

```
result, continuation_token= reviews(
    app_santander,
    lang='es',
    country='mx',
    sort=Sort.NEWEST,
    count=30000,
    filter_score_with=None
)
```

```
len(result)
data=pd.DataFrame(result)
data=data[data['content'].notna()]
```

```
data.shape
```

```
(30000, 10)
```

```
data["fecha"] = data["at"].dt.strftime("%Y-%m-%d")
data["hora"] = data["at"].dt.strftime("%H:%M:%S")
data["aniomes"] = data["at"].dt.strftime("%Y-%m")
```

```
data.aniomes.value_counts()
```

```
2021-12    8050
2022-01    7251
2022-02    5912
2022-03    5346
2021-11    3441
Name: aniomes, dtype: int64
```

## RESULTADOS

	reviewId	userName	userImage	content	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	fecha	hora	aniomes
2	gp:AOqpT	Antonio Martínez	<a href="https://play">https://play</a>	Muy buena app	0	5.62.3	2022-03-27 14:23:23			2022-03-27 14:23:23	2022-03	
3	gp:AOqpT	David Orozco	<a href="https://play">https://play</a>	Lenta y con muchos errores	0	5.62.3	2022-03-27 14:14:56			2022-03-27 14:14:56	2022-03	
4	gp:AOqpT	Diana Iizbeth Mejia Hernández	<a href="https://play">https://play</a>	Muy buena	0	5.62.3	2022-03-27 14:11:42			2022-03-27 14:11:42	2022-03	
5	gp:AOqpT	Hector Hernandez Cisneros	<a href="https://play">https://play</a>	Excelente	0	5.62.3	2022-03-27 13:48:36			2022-03-27 13:48:36	2022-03	
6	gp:AOqpT	Ana Mora	<a href="https://play">https://play</a>	Pésima y no me gusta que la app me rastree	0	5.62.3	2022-03-27 13:47:11			2022-03-27 13:47:11	2022-03	
7	gp:AOqpT	Jose María Ponce Becerril	<a href="https://play">https://play</a>	Buena	0	5.62.3	2022-03-27 13:44:32			2022-03-27 13:44:32	2022-03	
8	gp:AOqpT	ISRAEL ZAMORANO	<a href="https://play">https://play</a>	Pesi servicio en sucursal la app esta peor todo los días	0	5.62.3	2022-03-27 13:41:15			2022-03-27 13:41:15	2022-03	
9	gp:AOqpT	Cesar Octavio Loza Saucedo	<a href="https://play">https://play</a>	muy mala, pesima	0	5.62.3	2022-03-27 13:38:55			2022-03-27 13:38:55	2022-03	
10	gp:AOqpT	Daniel Ramirez Gama	<a href="https://play">https://play</a>	Excelente App	0	5.62.3	2022-03-27 13:35:01			2022-03-27 13:35:01	2022-03	
11	gp:AOqpT	Brandon Cruz	<a href="https://play">https://play</a>	Muy buena y rapida	0	5.62.3	2022-03-27 13:31:57			2022-03-27 13:31:57	2022-03	
12	gp:AOqpT	Hec Urizar	<a href="https://play">https://play</a>	Todo bien	0	5.62.3	2022-03-27 13:30:44			2022-03-27 13:30:44	2022-03	
13	gp:AOqpT	Naytali Torreblanca Hernández	<a href="https://play">https://play</a>	Es una app muy buena y muy confiable	0	5.62.3	2022-03-27 13:29:11			2022-03-27 13:29:11	2022-03	
14	gp:AOqpT	Eduardo Lira	<a href="https://play">https://play</a>	Una magnifica herramienta	0	5.62.3	2022-03-27 13:22:39			2022-03-27 13:22:39	2022-03	
15	gp:AOqpT	Iveth García	<a href="https://play">https://play</a>	Buena	0	5.62.3	2022-03-27 13:22:35			2022-03-27 13:22:35	2022-03	
16	gp:AOqpT	Nancy Martinez melquiades	<a href="https://play">https://play</a>	Excelenteapp	0	5.62.3	2022-03-27 13:15:52			2022-03-27 13:15:52	2022-03	
17	gp:AOqpT	Un usuario de Google	<a href="https://play">https://play</a>	Muy buena aplicación y de fácil uso	0	5.62.3	2022-03-27 13:14:11			2022-03-27 13:14:11	2022-03	
18	gp:AOqpT	ESTO PASA EN MÉXICO	<a href="https://play">https://play</a>	Me parece mal que si quiero hacer una transferencia	0	5.62.3	2022-03-27 13:08:26			2022-03-27 13:08:26	2022-03	
19	gp:AOqpT	Rubén dario Bojorquez patron	<a href="https://play">https://play</a>	Feliz con mi app Santander, una de las mejores	0	5.62.3	2022-03-27 12:52:32			2022-03-27 12:52:32	2022-03	
20	gp:AOqpT	AA 2	<a href="https://play">https://play</a>	Iban tan bien, por qué le tienen que andar moviendo	0	5.62.3	2022-03-27 12:50:49			2022-03-27 12:50:49	2022-03	
21	gp:AOqpT	Takero bailongo Chávez	<a href="https://play">https://play</a>	Buena a secas , tarde en arreglar mi problema , un poco	0	5.62.3	2022-03-27 12:50:32			2022-03-27 12:50:32	2022-03	

**\*\*WEB SCRAPING : APPLE STORE\*\* -- RESULTADOS**

```
import json
from app_store_scraper import AppStore
import numpy as np
```

```
sant=AppStore(country='mx', app_name='supermovil-santander-mexico',app_id='498944221')
sant.review(how_many=30000)
sant.reviews
```

```
df=pd.DataFrame(np.array(sant.reviews),columns=['review'])
df2=df.join(pd.DataFrame(df.pop('review').tolist()))
```

```
df2.head(3)
```

	title	date	rating	isEdited	review	userName	developerResponse
0	Notificaciones	2019-01-26 15:12:14	4	False	Porfavor agregen la opción de recibir notifica...	hdhsyehheje	NaN
1	Buena	2017-10-06 15:52:11	4	False	Es buena la aplicación pero falta información ...	Fabby Montes	NaN
2	Solución iPhone 11	2020-01-01 19:59:23	1	False	Para los que tienen el caso de que la App se c...	Arturo26390	NaN

-----FIN AVANCE

## 2.1 Nulla interdum ex erat

In pharetra nec orci in tincidunt. Nulla id libero ac sapien iaculis ultrices et vel odio. Aliquam pharetra leo tellus, vitae finibus mi ultricies vitae. Nulla interdum ex erat, non placerat ante iaculis non. Donec condimentum vestibulum neque, at viverra eros maximus porta. Nam eu interdum eros, ut hendrerit elit. Nullam purus justo, feugiat a sapien non, tempus varius eros.

Sed convallis scelerisque consectetur. Sed quis viverra risus. Vestibulum ligula tortor, cursus quis ipsum condimentum, elementum feugiat nulla. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris lacus elit, malesuada sit amet accumsan eu, varius at leo. Etiam consectetur a urna quis consectetur. In suscipit pretium scelerisque. Proin risus dui, luctus vitae lectus vitae, tempor vehicula mauris.

Pellentesque auctor quam in pulvinar laoreet. Aliquam rhoncus tempor augue, eget maximus quam tincidunt ut. Cras convallis congue tempor. Integer felis enim, imperdiet at facilisis ac, sodales a mi. Fusce nec vehicula velit, ac porttitor justo. Vivamus sagittis, ipsum id congue porta, sapien tortor pharetra libero, vel feugiat ante diam a sem. In mollis posuere pulvinar. Phasellus accumsan semper finibus.

Fusce vel massa elit. Vivamus ultrices lectus ac porttitor mattis. Curabitur tristique odio accumsan, porta mi nec, fermentum enim. Curabitur efficitur, elit a congue lacinia, mauris enim malesuada diam, non fringilla dui nulla sed lectus. Cras sit amet fringilla turpis. Curabitur tempus laoreet lectus, sed efficitur ante. Vestibulum auctor eros a purus vestibulum laoreet. Interdum et malesuada fames ac ante ipsum primis in faucibus. Vestibulum a nunc sit amet libero rutrum fringilla eu a lacus. Vestibulum quis erat turpis. Donec a lacus eget augue ultricies bibendum eu et ipsum.

### **2.1.1 Mauris et nisl tempus, fringilla**

Sed pharetra ante interdum ac. Ut metus felis, condimentum id vehicula sit amet, bibendum vel lacus. Vestibulum pharetra ac nulla in cursus. Integer bibendum, lorem ullamcorper ultricies ultrices, neque dui scelerisque nisi, at rhoncus diam arcu at ligula. Pellentesque laoreet iaculis lorem in cursus. Aliquam venenatis mollis lacus, aliquam aliquam ante. Integer consequat massa magna, et dapibus nisl lobortis quis. Mauris feugiat diam id tristique tristique.

Quisque a felis nisi. Quisque fermentum, nibh vel viverra aliquet, est nulla porta felis, ut ullamcorper dolor sapien quis nibh. Mauris et nisl tempus, fringilla nunc quis, mattis dolor.



## Capítulo 3

**Pellentesque mi mauris, tempor vel  
risus nec, cursus viverra mi.  
Aenean scelerisque at orci et  
ullamcorper**



## Capítulo 3. Pellentesque mi mauris, tempor vel risus nec, cursus viverra mi. Aenean scelerisque at orci et ullamcorper

Nulla egestas sem et massa finibus, sit amet ullamcorper tellus consequat. Nullam rutrum interdum scelerisque. Cras convallis rutrum justo at tempus. Sed lobortis rhoncus ligula at iaculis. Curabitur diam ante, luctus non condimentum nec, auctor quis nibh. Donec id ipsum dolor. Suspendisse potenti.

Fusce eleifend volutpat turpis nec scelerisque. Nunc lacinia porttitor pulvinar. Duis vel turpis elit. Maecenas bibendum nunc et molestie vulputate. Etiam auctor lorem tincidunt, egestas ipsum congue, tristique ante. Quisque ut ex at mi vestibulum molestie. Morbi finibus vehicula ante, hendrerit porta metus commodo in.

Nulla nisl eros, aliquam et tincidunt ut, feugiat in nibh. Nulla ultrices purus sit amet tellus sodales, in elementum lectus semper. Pellentesque sit amet egestas orci. Maecenas at euismod dui, id vestibulum massa. Proin iaculis semper arcu ut efficitur. Quisque non varius justo. Duis ullamcorper arcu at volutpat mattis. Vivamus porta porta urna sit amet pellentesque. Phasellus ac odio fermentum, semper erat et, eleifend nisl. Nunc non orci hendrerit, convallis nisl id, viverra felis. Duis pulvinar sem lorem, eu faucibus mauris blandit eget. Morbi fermentum mi in sodales luctus. Maecenas pretium mollis lorem vel lobortis. Maecenas quis erat erat. Proin in congue ligula, non tempus massa.

Sed eleifend nunc et augue pulvinar, dignissim scelerisque nisl interdum. Mauris a sapien non velit varius lacinia. Quisque et erat interdum, maximus ex ac, dictum magna. Etiam viverra vitae lacus non finibus. Sed lacinia nisi et odio blandit, ut convallis mi finibus.

### 3.1 Cras velit purus, commodo condimentum finibus lobortis, interdum at lacus

In pharetra nec orci in tincidunt. Nulla id libero ac sapien iaculis ultrices et vel odio. Aliquam pharetra leo tellus, vitae finibus mi ultricies vitae. Nulla interdum ex erat, non placerat ante iaculis non. Donec condimentum vestibulum neque, at viverra eros maximus porta. Nam eu interdum eros, ut hendrerit elit. Nullam purus justo, feugiat a sapien non, tempus varius eros.

Sed convallis scelerisque consectetur. Sed quis viverra risus. Vestibulum ligula tortor, cursus quis ipsum condimentum, elementum feugiat nulla. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris lacus elit, malesuada sit amet accumsan eu, varius at leo. Etiam consectetur a urna quis consectetur. In suscipit pretium scelerisque. Proin risus dui, luctus vitae lectus vitae, tempor vehicula mauris.

Pellentesque auctor quam in pulvinar laoreet. Aliquam rhoncus tempor augue, eget maximus quam tincidunt ut. Cras convallis congue tempor. Integer felis enim, imperdiet at facilisis ac, sodales a mi. Fusce nec vehicula velit, ac porttitor justo. Vivamus sagittis, ipsum id congue porta, sapien tortor pharetra libero, vel feugiat ante diam a sem. In mollis posuere pulvinar. Phasellus accumsan semper finibus.

Fusce vel massa elit. Vivamus ultrices lectus ac porttitor mattis. Curabitur tristique odio accumsan, porta mi nec, fermentum enim. Curabitur efficitur, elit a congue lacinia, mauris enim malesuada diam, non fringilla dui nulla sed lectus. Cras sit amet fringilla turpis. Curabitur tempus laoreet lectus, sed efficitur ante. Vestibulum auctor eros a purus vestibulum laoreet. Interdum et malesuada fames ac ante ipsum primis in faucibus. Vestibulum a nunc sit amet libero rutrum fringilla eu a lacus. Vestibulum quis erat turpis. Donec a lacus eget augue ultricies bibendum eu et ipsum.

### **3.1.1 Integer sed ipsum nisl. Sed sed blandit orci**

Sed pharetra ante interdum ac. Ut metus felis, condimentum id vehicula sit amet, bibendum vel lacus. Vestibulum pharetra ac nulla in cursus. Integer bibendum, lorem ullamcorper ultricies ultrices, neque dui scelerisque nisi, at rhoncus diam arcu at ligula. Pellentesque laoreet iaculis lorem in cursus. Aliquam venenatis mollis lacus, aliquam aliquam ante. Integer consequat massa magna, et dapibus nisl lobortis quis. Mauris feugiat diam id tristique tristique.

# Conclusiones



## Conclusiones

Maecenas commodo lacus et imperdiet vulputate. In non dictum justo, et accumsan ante. Ut felis mauris, aliquet ac rutrum nec, scelerisque laoreet felis. Proin vitae dui metus. Nam ac cursus purus, id molestie nisl. Sed vel ante eget odio pharetra faucibus. Integer non malesuada ex. Ut non dapibus purus. Cras varius turpis sed lectus facilisis, quis luctus ligula scelerisque. Etiam fermentum odio a sem ornare pretium. In at pharetra sem. In hac habitasse platea dictumst. Interdum et malesuada fames ac ante ipsum primis in faucibus.

Sed vestibulum orci non felis commodo, non vulputate velit suscipit. Nam viverra nulla vitae ipsum tincidunt, et elementum quam consectetur. Sed tristique suscipit lectus non sodales. Aenean ac aliquam lorem. Aenean eleifend nec ex id fringilla. Pellentesque aliquet semper arcu nec molestie. Phasellus scelerisque finibus nibh, ac luctus enim cursus sed. Integer leo tellus, eleifend et pharetra at, hendrerit sed metus. Sed eu eros vitae massa vehicula pellentesque. Pellentesque non fringilla eros, sed ultrices risus. Maecenas facilisis quis ante et suscipit.

Nullam vel elementum nisi, ut venenatis odio. Quisque lacinia enim odio, nec lobortis dolor viverra eu. Sed eu tempus ligula. Sed commodo est vel ligula pharetra, in maximus felis tempus. Nulla commodo eleifend ex, ut imperdiet felis blandit eget. Nulla egestas, dolor eget elementum hendrerit, nibh erat pretium augue, ac egestas justo ipsum ut eros. Cras dignissim, est sed rhoncus dapibus, nunc risus tempus leo, eget hendrerit lectus est vitae magna. Praesent cursus, ex sit amet tristique venenatis, nisi justo suscipit dolor, vitae consequat odio libero a tellus. Sed quis ullamcorper urna, ac aliquam nunc. Nunc ac ornare nunc, non sagittis leo.

Donec vel mauris convallis, accumsan mi sed, consectetur lacus. Suspendisse interdum consequat elit ut luctus. Donec volutpat ut diam et vehicula. Donec hendrerit leo vel metus sollicitudin finibus. Aenean nunc ipsum, scelerisque in vehicula vitae, commodo id nisl.

## Bibliografía

- ❖ SuperMóvil  
<https://play.google.com/store/apps/details?id=mx.bancosantander.supermovil&hl=es>
- ❖ SuperMóvil <https://apps.apple.com/mx/app/santander-super%C3%B3vil/id498944221>
- ❖ Chong. Sentiment Analysis and Topic Extraction of the Twitter Network of #Prayforparis
- ❖ Yeun Kim. Modeling Topic Extraction-based Sentiment Analysis Based on User Reviews
- ❖ Gabriele Pergola\_, Lin Gui, Yulan He. A Topic-Dependent Attention Model for Sentiment Analysis
- ❖ I.V. (2019, 6 noviembre). Tipos de gráficos y diagramas para la visualización de datos. ingeniovirtual.com. Recuperado 18 de marzo de 2022, de <https://www.ingeniovirtual.com/tipos-de-graficos-y-diagramas-para-la-visualizacion-de-datos/>
- ❖ Tableau Software. Maila Hardin, Daniel Hom, Ross Perez y Lori Williams. ¿Qué tabla o gráfico es el adecuado para usted? Recuperado 18 de marzo de 2022.
- ❖ IBM Analytics. Metodología Fundamental para la Ciencia de Datos.  
<https://www.ibm.com/downloads/cas/6RZMKDN8>



**ANEXOS**



## ANEXO 1

## Índice de términos

### “A”

Aliquam.....12

### “B”

Blandit.....3

### “C”

Consectetur.....7

### “D”

Donec.....12