

## Contents

<b>1</b>	<b>Procesamiento y análisis de datos</b>	<b>2</b>
1.1	Datos utilizados . . . . .	2
1.2	Sobre la empresa . . . . .	2
1.3	Lenguajes y librerías utilizados . . . . .	2
1.4	Repositorio de Github . . . . .	3
<b>2</b>	<b>Breve análisis del set de Datos</b>	<b>4</b>
<b>3</b>	<b>Featuring Engineering</b>	<b>4</b>
<b>4</b>	<b>Clasificadores</b>	<b>4</b>
<b>5</b>	<b>Tuning</b>	<b>4</b>
<b>6</b>	<b>Ensamble</b>	<b>4</b>
<b>7</b>	<b>Conclusiones generales</b>	<b>5</b>

# 1 Procesamiento y análisis de datos

En esta sección se introduce brevemente el producto a analizar y las herramientas que se utilizaron para realizar el análisis y la predicción requerida.

## 1.1 Datos utilizados

Se estudiaron datos provistos por la empresa Trocafone, analizando un conjunto de eventos de web analytics de usuarios que visitaron [www.trocafone.com](http://www.trocafone.com), su plataforma de e-commerce de Brasil.

## 1.2 Sobre la empresa

Trocafone es un side to side Marketplace para la compra y venta de dispositivos electrónicos que se encuentra actualmente operando en Brasil y Argentina.

La empresa realiza distintas actividades que van desde la implementación de plataformas de trade-in (conocidos en la Argentina como Plan Canje), logística directa y reversa, reparación y recertificación de dispositivos (refurbishing) y venta de productos recertificados por múltiples canales (e-commerce, marketplace y tiendas físicas).

Para conocer más de su modelo de negocio, pueden visitar el siguiente artículo:

<https://medium.com/trocafone/el-maravilloso-mundo-de-trocafone-5bdc5761856b>

## 1.3 Lenguajes y librerías utilizados

- Se utilizó como lenguaje de programación **Python3**.
- Para las visualizaciones, se utilizaron las librerías **Matplotlib** y **Seaborn**.
- Como editor se utilizó **Jupyter Lab** (o **Jupyter Notebook**)
- Para el manejo de DataFrames, se eligió **Pandas** como librería a utilizar.
- Se utilizaron algunas herramientas como `std` o `argsort` de la librería **Numpy**. `calendar`

- Se importaron diferentes métricas como "accuracy score", "f1 score", "precision score", "recall score", "roc auc score" de la librería **sklearn**.
- Se utilizaron diferentes búsquedas de hiperparametros, entre ellas "Grid-Search" y "RandomizedSearchCV" de **sklearn**.
- Para poder realizar cross-validation, se importó "StratifiedKFold" de **sklearn**.
- Se realizó Clustering mediante la funcion "KMeans" de la librería **sklearn**.
- Los siguientes clasificadores fueron importados: "xgboost", "lightgbm", "RandomForestClassifier", "CatBoostClassifier".
- Se utilizo como ensamble de clasificadores a "VotingClassifier".

## 1.4 Repositorio de Github

Para el trabajo en conjunto del equipo, se utilizo un repositorio en github, donde se encuentran todos los archivos necesarios del análisis y predicciones y este informe propiamente dicho.

Link: <https://github.com/emabrea/7506-DATOS-TP2.git>

## **2 Breve análisis del set de Datos**

### **3 Featuring Engineering**

## 4 Clasificadores

## 5 Tuning

## **6 Ensamble**



## 7 Conclusiones generales

Al finalizar el presente trabajo, nos propusimos destacar ciertos aspectos del mismo, de tal modo que permitan a la empresa tener una visión más clara de su progreso.

Los principales puntos a tratar son:

- Las ventas de celulares vienen aumentando con los meses, pero lentamente, aproximadamente 50 ventas más por mes.
- Las campañas y avisos publicitarios deben publicarse durante la semana, y no en el fin de semana. Además, dichas campañas deben mostrarse a partir de las 16:00, recordando que la mayor actividad ocurre a la noche.
- El día martes es un buen candidato para publicitar, tal vez porque los lunes son muy cargados de trabajo, y recién el martes uno tiene más libertad.
- Continuar con la campaña de subscripciones, luego de realizar una compra, pues en el último mes viene dando resultados.
- Si bien la marca más comprada es Samsung, deberían profundizar en la venta de iPhone, que son los más visitados, pues están de moda por ser considerados de alta gama. Esto no implica no seguir con los modelos de Samsung, simplemente abrir el panorama.
- Si bien los celulares de 16gb son los más vendidos, prestar atención con los de 32gb, que podrían ser los próximos protagonistas. A aquellos con 8gb, no brindar tanto soporte como antes.
- La condición “Bueno” sigue siendo la más visitada y comprada, tener esto en cuenta. Cada mes más usuarios compraron a partir de una campaña, pero como se vió en el análisis, la única con resultados significativos es la de Google. Por lo tanto, creemos que debería continuarse con esta modalidad, pero habría que evaluar la relación costo beneficio de las otras campañas publicitarias.
- Dar soporte especializado al navegador Google Chrome, tanto en su versión de escritorio como en su versión mobile, los otros no son tan significativos a nivel cantidad de usuarios.

- Por su parte, las tablets están en desuso, por lo tanto, enfocarse en los smartphones principalmente, que van a superar las computadoras de escritorio pronto, en términos de uso.
- Si bien la mayoría de los usuarios son de Brasil, se noto una creciente actividad en los Estados Unidos. Por lo tanto, una propuesta es traducir la página web al inglés, para promover la visita de personas de dicho país. También notamos un posible negocio futuro en Argentina, donde se debe hacer hincapié en la diferencia de precio con respecto a otros vendedores.
- Por último, se debe volver a hacer un análisis de datos en unos meses, y verificar el progreso de la empresa en base a las recomendaciones propuestas, para determinar el correcto camino.

## References

- [1] Trocafone website, [www.trocafone.com](http://www.trocafone.com).
- [2] NumPy - NumPy, <http://www.numpy.org/>.
- [3] Python Data Analysis Library, <https://pandas.pydata.org/>.
- [4] Matplotlib: Python plotting — Matplotlib 3.0.0 documentation, [matplotlib.org](https://matplotlib.org).
- [5] seaborn: statistical data visualization — seaborn 0.9.0 documentation, [seaborn.pydata.org](https://seaborn.pydata.org)
- [6] Folium information, <https://github.com/python-visualization/folium>
- [7] GeoPy's documentation, <https://geopy.readthedocs.io/en/stable/>