

Google analytics customer revenue prediction

**Gabriel Jaime Zapata**

**Valeria Quiroz Gomez**

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título de:

**Especialista en analitica y ciencia de datos**

Asesores (a):

Julian David Arias Londoño - Doctor en ingeniería

Línea de Investigación:

Nombrar la línea de investigación en la que se enmarca la tesis o trabajo de investigación

Universidad de Antioquia

Facultad de ingeniería, Departamento de ingeniería de sistemas.

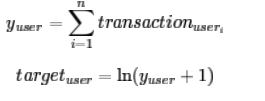
Medellín, Colombia

2021.

1. **Descripción del problema**

Se cree que un porcentaje de clientes pequeños son quienes generan la mayor cantidad de ingresos de las organizaciones, especialmente aquellas que se dedican al comercio. Por esto, las empresas se enfrentan al desafío de realizar campañas de marketing con correctas estrategias, enfocadas en clientes que de verdad generen ingresos y estén dispuestos a pagar el costo de los artículos ofrecidos.

Por lo anterior, el objetivo de este trabajo de investigación es realizar un análisis que permita predecir el registro natural de la suma de todas las transacciones por usuario que ingresa a la tienda de Google, expresado matemáticamente de la siguiente manera:



Para lograr este objetivo se analiza un conjunto de datos en un periodo de tiempo de XXXXXX de los clientes de Google Merchandise Store,Google Merchandise Store es un sitio web de comercio electrónico que vende productos de la marca Google, también conocido como GStore; con el fin de brindar un mejor uso a los presupuestos de marketing que permita el enfoque en sus campanas y posteriormente mejorar las ventas en la tienda. En este caso, el conjunto de datos está disponible públicamente en Kaggle.com. Además la medida de error utilizada es el error cuadrático medio (RMSE) y el error absoluto medio - MAE por sus siglas en inglés.

1. **Descripción del dataset**

El conjunto de datos que se analiza están dados por un conjunto de datos de entrenamiento y un conjunto de test. Cada fila del conjunto de datos es una visita a la tienda por parte de un usuario, es importante mencionar que un registro de visita no necesariamente implica una compra. La columna transactionRevenue contiene la información de ingresos de los usuarios que pretendemos predecir, cabe destacar que esta subcolumna existe solo para los datos de entrenamiento. En el conjunto de datos se tiene XXX variables, con XXX cantidad de filas. En las variables otorgadas en el dataset se considera información tal como: la fecha de visita a la tienda, un identificador único por usuario, el canal a través del cual el usuario ingresa a la tienda, el estado civil del usuario, especificaciones del dispositivo por el cual accede a la tienda, información geográfica del usuario, el total de visitas a la tienda, número de visitas del usuario a la tienda, en caso de ser la primera vez, automáticamente es 1, entre otras variables.

1. **Objetivos**

Predecir el registro natural de la suma de todas las transacciones por usuario en la tienda de compras online de Google (Gstore)

Construir modelos que permitan llevar a cabo la predicción de los ingresos por usuario.

1. **Marco teórico**

**Marco metodológico**

En Machine learning existen diversos algoritmos que permiten realizar predicciones sobre datos e información, para este caso se utilizan los modelos de GradientBoostingRegressor, RandomForestRegressor, LightGBM y redes neuronales. Los cuales tienen diferentes maneras de hacer regresiones y brindan resultados que suelen ser atractivos a la hora de resolver el problema planteado.

Al ser un ejercicio publicado en Kaggle.com hay varios competidores que lo han resuelto, especialmente en el lenguaje de programación R, además la competencia ya finalizó, por tanto se puede consultar los mejores resultados que se obtuvieron por parte de los concursantes en la página oficial del concurso.

1. **Resultados**

**6.1 capítulos**

1. **Conclusiones**
2. **Referencias bibliográficas**
3. **Agradecimientos**
4. **Anexos**

[**https://www.kaggle.com/shrmis/google-analytics-customer-revenue-prediction**](https://www.kaggle.com/shrmis/google-analytics-customer-revenue-prediction)

[**https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html**](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html)

[**https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html**](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html)

[**https://www.kaggle.com/c/ga-customer-revenue-prediction/rules**](https://www.kaggle.com/c/ga-customer-revenue-prediction/rules)

[**https://lightgbm.readthedocs.io/en/latest/**](https://lightgbm.readthedocs.io/en/latest/)

[**https://lightgbm.readthedocs.io/en/latest/Python-Intro.html**](https://lightgbm.readthedocs.io/en/latest/Python-Intro.html)