



**UNIVERSIDAD  
DE ANTIOQUIA**

**COMPETENCIA M5 FORECASTING -  
ACCURACY**

Autor

Diego Armando Agudelo Ospina

Universidad de Antioquia

Facultad de Ingeniería, Departamento de Sistemas

Medellín, Colombia

2021



Competencia M5 Forecasting - Accuracy

**Diego Armando Agudelo Ospina**

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

**Especialista en Analítica y Ciencia de Datos**

Asesores (a):

**John Freddy Duitama Muñoz**

Universidad de Antioquia

Facultad de Ingeniería

Medellín, Colombia

2021.

## CONTENIDO

1. RESUMEN EJECUTIVO .....	3
----------------------------	---

## 1. RESUMEN EJECUTIVO

En este trabajo se encuentra una aproximación simple de 2 técnicas diferentes a LightBGM para la predicción múltiples series de tiempo, una de ellas tiene 2 aproximaciones, siendo una univariada y otra multivariada. El Mayor foco se dio a el análisis de la data y su comportamiento general y específico.

Esta competencia se basa en la predicción de periodos de ventas de 28 días en base a datos históricos de ventas, eventos especiales, bonos SNAP (Supplemental Nutrition Assistance Plan) y precios de venta por producto en cada tienda. La intención de la competencia es buscar la forma donde se obtenga la mayor precisión posible en la predicción de las ventas. El método de medición a utilizar seria WRMSSE (Weighted Root Mean Scale Squared Error), esto fue decisión de los organizadores, para medirla se requiere o crear una clase personal o subir los resultados obtenidos al calificador automático proveído.

Se encontrará que hay 3 estados (los cuales son California, Texas y Wisconsin), dentro de cada estado hay múltiples tiendas (entre 2 a 3 tiendas). Para cada una de estas tiendas tendremos las categorías de Household, Food y Hobbies. Después de este nivel vienen los departamentos por categoría, tendremos entre 1 a 3 departamentos y finalmente la cantidad de productos al final de este árbol. Tiene 4 niveles jerárquicos por encima de cada serie de tiempo y puede permitir agrupaciones para una mejor lectura de datos o incluso modelos a más alto nivel de ser requerido.

Para más información favor referirse a los documentos provistos por los organizadores para entender a profundidad el problema a solucionar y la meta deseada.

Las principales librerías de Python usadas para este proyecto son TensorFlow, FB Prophet, Matplotlib, Downcast, Sklearn, Seaborn, Pandas y Tqdm.

**Nota:** Tener en cuenta que el algoritmo sencillo de FB Prophet requiere un archivo Python separado para correr y hacer uso múltiples núcleos. El segundo algoritmo de FB Prophet tiene una duración superior a 24 horas de corrida.

Se recomiendan explorar la posibilidad de usar DeepAR y XGBoost para encontrar otros resultados por arboles jerárquicos.

En el siguiente repositorio se encuentran los notebooks y archivos Python que fueron usados para la ejecución del proyecto: <https://github.com/Daguos/M5---Forecasting>