

ITAM - Métodos Estadísticos para C.Pol y R.I.

Proyecto Otoño 2022

Objetivo: Realizar regresiones lineales basadas en el conjunto de datos *House Prices*, que contiene datos que nos permiten estimar el precio de una casa basado en sus características.

Descripción: Por cada integrante del proyecto se deberá entregar un modelo lineal que busque explicar el precio de las casas. Equipos de mínimo 2 integrantes y máximo 4.

Entregables:

- **(60pts) Documento docx o pdf:**

- (20pts) Entre 2 y 3 hojas de introducción donde se realice un Exploratory Data Analysis (EDA) que nos permita identificar variables importantes.

- (30pts) Entre 1 y 2 hojas explicativas por modelo. Debe explicar todos los componentes vistos de Regresión Lineal Múltiple: $\hat{\beta}_i$, IC de β_i , Pruebas de hipótesis de β_i , Matriz de Varianzas-Covarianzas, IC para predicción media, para predicción a futuro, $\hat{\sigma}^2$, etc. Es importante que las explicaciones estén en contexto del problema.

- (10pts) 1 hoja de conclusión donde se comparen los modelos presentados y se concluya cual es el mejor modelo. Incluir su usuario dentro de plataforma Kaggle

El reporte escrito no debe contener código, solamente los resultados (graficas, tablas comparativas, etc)

- **(20 pts) Documento Rmd y PDF:** Se debe incluir el Notebook (Rmd) con sus secciones que correspondan al reporte escrito

- **(20 pts) Puntuación en plataforma Kaggle**

- **Opcionales:**

- 5 pts al mejor nombre de equipo

- 10 pts a la mejor puntuación de Kaggle

- 5 pts a la segunda mejor puntuación de Kaggle

Nota: Todo el proyecto debe ser congruente en todo su contenido; por ejemplo, los resultados del modelo de regresión deben estar de acorde a lo concluido del EDA.

Instrucciones:

1. Acceda al proyecto en la plataforma de Kaggle y registre a su equipo e integrantes.
2. Descargue los datos, son 4 archivos:
 - train.csv: estos datos contienen todas las columnas, incluyendo la columna a predecir
 - test.csv: estos datos contienen todas las columnas, excepto la columna a predecir. De estos datos se realizará la estimación del precio y se subirá a la plataforma
 - data.description.txt: descripción de las columnas
 - sample_submission.csv: contiene el formato en que deben subir los datos (solo se suben esas dos columnas)
3. Desarrolle los modelos en R y elija el mejor
4. Realice submit de las predicciones en el formato adecuado