

PROYECTO DE CURSO – FASE 2

Objetivo

Construir algoritmos de clasificación para comparar el rendimiento de cada algoritmo así como evaluar distintos hiper-parámetros de cada modelo para buscar la mejor solución a un problema de clasificación.

Requisitos

1. Desarrollo de modelo de clasificación: deberá desarrollar y analizar los algoritmos de clasificación listados a continuación.
 - a. Naive Bayes
 - b. LDA
 - c. Regresión logística
 - d. SVM.
 - e. Árboles de decisión
 - f. Random forest
 - g. Análisis de discriminante lineal
 - h. Análisis de discriminante cuadrático
 - i. AdaBoost
 - j. Gradient Boosting
 - k. XGBoost
 - l. LGBM
2. Para cada algoritmo deberá entrenar y evaluar al menos cinco configuraciones de hiper-parámetros diferentes en los casos que aplique.
3. Listado de modelos finales: Una vez entrenados y evaluados sus modelos de clasificación, deberá utilizar un set de entrenamiento y testing para entrenar y realizar el registro del modelo y devolver una tabla ordenada considerando del mejor al peor modelo.

Entregables – Fase 2

- Link de github incluyendo:
 - Dataset seleccionado original.
 - Dataset resultante del proceso de ingeniería de características.
 - **Notebook #1:** con Análisis Exploratorio con tablas y gráficas de sus datos de la fase 1
 - **Notebook #2:** con el análisis, procedimiento y construcción del Pipeline de ingeniería de características de la fase 1.
 - **Notebook #3:** entrenamiento y evaluación de modelos de clasificación detallando los pasos seguidos y la configuración de hiper-parámetros.
 - **Archivo Excel con Listado de Modelos:** Archivo de excel con el listado de los modelos probados en el orden que usted considere al finalizar el ejercicio de train-test, deberá incluir una columna con el índice del modelo, el nombre del modelo, lista de hiper-parámetros, métrica resultante, tiempo de entrenamiento, fecha y hora de finalización del entrenamiento.
 - Readme del proyecto y la información incluida en el repositorio.