CD3002C Inteligencia Artificial con Impacto Empresarial
Módulo 3 – Modelos de IA para Datos Estructurados
Febrero – Junio 2024

Actividad 2 – Modelos de Clasificación

Instrucciones: A partir de la asignación durante la sesión de clase de las siguientes bases de datos **i)** bank_application_data o ii) bank_marketing_data, realizar las instrucciones 1 – 6. En el desarrollo del archivo de R-Markdown, por favor incluir *data storytelling* de los resultados del análisis exploratorio de los datos (EDA) así como la interpretación de los resultados estimados.

Lectura Sugeridas:

What is Classification in Machine Learning and Why Is It Important?
https://emeritus.org/blog/artificial-intelligence-and-machine-learning-classification-in-machine-learning/

5 Classification Algorithms for Machine Learning
https://builtin.com/data-science/supervised-machine-learning-classification

What is K-Means Algorithm and How it Works?
https://towardsmachinelearning.org/k-means/

1) Brevemente responder con tus propias palabras 2 de las siguientes 3 preguntas:

i) ¿Qué es Supervised Machine Learning y cuáles son algunas de sus aplicaciones en análisis de clasificación?

ii) ¿Cuáles son los principales algoritmos de Supervised Machine Learning - Classification? Brevemente describir con tus propias palarbas 4 – 6 de los principales algoritmos de Supervised Machine Learning - Classification.

iii) Respecto a la selección de los resultados de los modelos de clasificación ¿Qué es la matriz de confusión? ¿Qué es el estadístico Kappa? ¿Cuál es la relación entre AUC y ROC Curve?

2) Desarrollar Análisis Exploratorio de los Datos (EDA) que incluye los siguientes elementos:

a. Identificación de NA's
b. Reemplazo de NA's
c. Medidas descriptivas
b. Medidas de dispersión
c. Identificación de patrones y/o tendencias en los datos mediante el uso de gráficos incluyendo bar plots, line plots, pie plots, histogramas, matriz de correlación, box plot, scatter plot, qq-plot, etc Mostrar al menos 4 – 6 gráficos.
d. Posible transformación de variables de interés (e.g., log, estandarización, normalización)

3) A partir de los resultados de EDA describir la especificación del modelo de clasificación a estimar. Brevemente, describir cómo es el posible impacto y/o relación de cada una de las variables control de interés sobre la principal variable de estudio.

4) Estimación de cada uno de los siguientes modelos de Supervised Machine Learning - Clasificación:

a. Logistic Regression
b. Decision Trees
c. Support Vector Machine (SVM)
d. K – Means Clustering
e. KNN
f. Naïve Bayes
g. Random Forest

6) Evaluación y Selección de Modelo de Clasificación

a. Matriz de Confusión
b. Estadístico Kappa
c. AUC
d. ROC Curve

7) Desarrollar una breve descripción de los 6 – 10 principales hallazgos de:

a. EDA
b. Modelo de Clasificación seleccionado. Brevemente, interpretar los resultados estimados.

Fecha de Entrega: Martes 12 de Marzo 2024 a las 11:59 PM (Vía Canvas)
Formato de Entrega: R – Markdown (html o pdf)
Formato de Entrega: Individual | Incluir Nombre Completo al inicio del archivo