

CD3002C – Inteligencia Artificial con Impacto Empresarial

Febrero – Junio 2024

Evidencia 2 - Preparación y modelado de los datos

1. Genera un archivo R markdown que reporte tu proceso de limpieza de los datos. El archivo debe incluir lo siguiente:

- ❖ Descripción de la estructura general de los datos previo a la limpieza.
- ❖ ¿Qué variables tienen valores nulos y cuántos?
- ❖ Describir el manejo que se hizo de los valores nulos
 - Si se imputaron variables, especificar como se realizó la imputación y por qué.
 - Si se decidió eliminar alguna variable especificar por qué.
- ❖ Describir las diferentes etapas de limpieza, si se realizaron filtrados, eliminaciones, generación de nuevas variables etc. y cuantos registros se eliminan tras cada ronda de limpieza.
- ❖ Si se generaron nuevas variables mencionar cuáles fueron, qué cálculo se realizó para generarlas y por qué se tomó la decisión de crearlas.
- ❖ Especificar las reglas de negocio utilizadas en el proceso de limpieza.
- ❖ Describir la estructura final de los datos una vez finalizado el proceso de limpieza.

** Nota: Estos elementos no deben aparecer en este orden necesariamente, pero deben estar presentes en el reporte del proceso de limpieza.

2. Genera un archivo R markdown para reportar el proceso de diseño de tus modelos de inteligencia artificial. Se debe incluir lo siguiente:

- ❖ Al menos tres modelos de Inteligencia Artificial utilizando técnicas para datos estructurados
- ❖ Al menos tres modelos de Inteligencia Artificial utilizando técnicas para datos no estructurados
- ❖ Reportar la estructura de cada uno de los modelos generados, que variables se seleccionaron para cada modelo y por qué.
- ❖ Reportar las medidas de eficiencia de cada uno de los modelos.
- ❖ Interpretar los resultados de los modelos (cuando sea posible).
- ❖ Realizar una comparativa entre modelos y seleccionar el mejor. Establecer claramente cuál fue el criterio para seleccionar un modelo como el mejor.

❖ **SP1- Forecast Ventas**

Modelos Regresivos

1. Series de Tiempo
2. Regresión Múltiple
3. Support Vector Machine

❖ **SP2 – Bajas de Capital Humano**

Modelos de Clasificación

1. Modelo CART
2. Regresión Logística
3. Naive Bayes

❖ **SP3 – Datos No Estructurados**

Algoritmos NoSQL

❖ Conclusiones y Hallazgos

❖ Referencias

Fecha de Entrega: Semana 15-17
Formato de Entrega: Archivo R – Markdown (html / PDF)