



Unidad Didáctica: Ejercicio de feedback 1

Matemáticas y Estadística para la IA

Para aplicar los conocimientos adquiridos durante la asignatura realizaremos un ejercicio de feedback, con la resolución de diferentes supuestos prácticos.

1. Objetivo

La retroalimentación consiste en una respuesta verbal, escrita o gestual por parte del docente con relación al desempeño o comportamiento de un estudiante. Su objetivo es potenciar al estudiante para que obtenga buenas calificaciones confiando en sus capacidades para resolver un problema

2. Enunciado

Eres un analista de datos contratado para explorar y extraer insights del dataset *"International Football Results from 1872 to 2024"*, disponible en Kaggle <https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017/data>

.Este dataset contiene información sobre más de 47,000 partidos internacionales de fútbol masculino, desde el primer partido oficial en 1872 hasta la actualidad, incluyendo datos de torneos históricos como la Copa Mundial de la FIFA y partidos amistosos.

Tu tarea es analizar este conjunto de datos utilizando las técnicas y herramientas aprendidas a lo largo del curso: estadística descriptiva, análisis EDA, modelos de clasificación y decisión, estimación bayesiana, depuración e imputación de datos, y redes neuronales avanzadas. Además, deberás responder preguntas clave sobre el impacto de jugar en casa, tendencias históricas, y los factores que influyen en los resultados.

Este análisis se realizará de manera individual y se entregará como un informe final.

Contexto del Dataset

El archivo principal, results.csv, incluye las siguientes columnas:

- **date**: Fecha del partido.
- **home_team** y **away_team**: Equipos locales y visitantes, respectivamente.
- **home_score** y **away_score**: Resultados finales, incluyendo tiempo extra (sin penales).
- **tournament**: Nombre del torneo.
- **city** y **country**: Ubicación del partido.
- **neutral**: Indica si el partido se jugó en un campo neutral (TRUE/FALSE).

Este dataset está diseñado para responder preguntas como:

- ¿Qué equipos dominan históricamente en casa y en torneos específicos?
- ¿Cómo han evolucionado las tendencias de goles anotados a lo largo de la historia?
- ¿Existe una ventaja clara al jugar como local?
- ¿Qué insights pueden extraerse mediante el uso de grafos o redes neuronales?

Tu análisis debe abordar estas preguntas e incluir visualizaciones y modelos predictivos basados en los datos.

Actividades y Retos

Exploración Inicial y Análisis Descriptivo

1. Exploración Inicial, EDA y Limpieza de Datos (1 puntos)

- Carga y explora el archivo results.csv:
 - Identifica las variables disponibles, tipos de datos y posibles errores (valores nulos, nombres duplicados).
 - Realiza una limpieza básica de los datos, incluyendo corrección de inconsistencias.
 - Implementa métodos de imputación para manejar datos faltantes.
 - Variables categóricas: imputación por moda.
 - Variables numéricas: media, mediana o interpolación.
- Exploración inicial EDA:
 - Identifica distribuciones principales (histogramas, boxplots, etc).
 - Analiza correlaciones iniciales entre variables cuantitativas (matriz de correlación).
 - Si fuese necesario, crea variables derivadas, como, por ejemplo: diferencia de goles, década del partido, etc.
 - Analiza relaciones categóricas, como frecuencia de victorias locales o torneos más populares, utilizando gráficos de barras y/o mapas de calor.
- Entrega un informe en .pdf (máximo 2 páginas) con:
 - Tablas descriptivas que incluyan estadísticas básicas (media, medianrango, etc.).
 - Ejemplos concretos de limpieza antes y después.
 - Visualizaciones iniciales para respaldar el EDA.
- 2. **Análisis Descriptivo Avanzado y EDA (1 punto)**
 - Responde preguntas clave utilizando análisis estadístico descriptivo:
 - ¿Cuál es la media de goles anotados por partido?
 - ¿Cuántos partidos se jugaron por década y en qué torneos?
 - ¿Qué equipos tienen el mayor rendimiento histórico jugando como local?
 - Presenta gráficos relevantes (barras, histogramas, boxplot, líneas de tiempo).
 - Entrega un informe con análisis descriptivo avanzado, preguntas clave respondidas y graficos relevantes.
- 3. **Reducción de Dimensionalidad (1 punto)**
 - Identificación de necesidades:
 - Determina si las variables derivadas son redundantes o altamente correlacionadas.
 - Si es necesario, usa el Análisis de Componentes Principales (PCA) para reducir dimensiones mientras mantienes información relevante.
 - Pasos de PCA:
 - Selecciona variables numéricas derivadas, como goles, diferencias de goles y métricas adicionales.
 - Calcula la varianza explicada por cada componente principal.
 - Visualiza los dos primeros componentes para identificar patrones históricos o grupales.
 - Aplicación en modelado:

- Utiliza las dimensiones reducidas como insumos para modelos predictivos.
- Entrega:
 - Visualizaciones del PCA y justificación del uso de dimensiones seleccionadas.

Análisis Avanzado y Modelos Predictivos

4. Ventaja de Localía y Probabilidades Condicionales (1 puntos)

- Usa el Teorema de Bayes para calcular:
 - La probabilidad de que un equipo gane si juega como local.
 - La probabilidad de empate en un campo neutral.
- Representa tus hallazgos en tablas y gráficos explicativos.

5. Modelos de Clasificación (2 puntos)

- Diseña un modelo para predecir el resultado de un partido (home_win, draw, away_win):
 - Selecciona variables relevantes como el torneo, localización, y goles anteriores.
 - Usa un modelo simple como árboles de decisión, regresión logística o clasificación bayesiana.
 - Evalúa el desempeño del modelo con métricas como precisión, F1-score y matriz de confusión.
- Entrega un informe en .pdf (máximo 3 páginas) con los pasos, resultados y gráficos de evaluación.

Redes y Técnicas Avanzadas

5. Análisis de Grafos (2 puntos)

- Crea un grafo donde:
 - Los nodos representen equipos.
 - Las aristas correspondan a los partidos jugados entre ellos.
- Calcula métricas como centralidad, grado y componentes conectados para identificar equipos clave.
- Visualiza el grafo y discute tus hallazgos en máximo 2 páginas.

6. Informe Final (2 puntos)

- Integra todos tus análisis en un informe final que contenga:
 - Exploración inicial y limpieza de datos.
 - Análisis descriptivo.
 - Resultados de análisis bayesiano.
 - Modelos predictivos.
 - Resultados avanzados de grafos.
- Máximo: 13 páginas en .pdf.

3. Instrucciones de entrega

- **Extensión:** Cumple con las indicaciones de la actividad.
- **Formato:**
 - Informe en PDF.
 - Código en un notebook de Jupyter o Google Colab.
 - Visualizaciones en PNG o JPG.
- **Nombre del fichero:** Feedback1_individual.pdf
- **Plataforma de entrega:** Aula virtual, sección Feedback. La fecha máxima de entrega será el jueves 06 de febrero de 2025.
- **Recursos Adicionales**
 - Kaggle - <https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017/data> Documentación de NLTK

WELCOME
TO
UAX

UAX

Universidad
Alfonso X el Sabio

GRACIAS

UAX.COM