



Universidad Autónoma de Querétaro: Facultad de Informática

Disciplina: Ingeniería en Ciencia y Analítica de Datos

Código: CAD22

Profesor(a): Dr. Martín Muñoz Mandujano

Semestre: 5

Curso: Modelos de Pronostico.

Series de Tiempo

Lee atentamente las instrucciones:

- Esta práctica se realizará en parejas.
- Pueden emplear cualquier modelo que deseen.

1. Descripción del Proyecto Parcial: Pronóstico de Temperatura Diaria

El conjunto de datos **Daily Minimum Temperatures** contiene una serie de tiempo univariante con las **temperaturas mínimas diarias** registradas en la ciudad de Melbourne, Australia, a lo largo de varios años. En este proyecto el objetivo es **construir un modelo de pronóstico de series de tiempo** que permita predecir la temperatura mínima diaria futura a partir del historial disponible.

Este proyecto busca que los estudiantes apliquen los conocimientos adquiridos sobre análisis de datos, preprocesamiento de series de tiempo, construcción y evaluación de modelos de pronóstico, así como la comunicación de hallazgos mediante un reporte técnico.

1.1. Archivos Disponibles

Se proporciona únicamente el siguiente archivo:

- **1_Daily_minimum_temps.csv**: Archivo que contiene la serie histórica de temperaturas mínimas diarias.

Cada equipo deberá, a partir de este único archivo, realizar la división en conjuntos de entrenamiento y prueba directamente en su código (no se proporcionan archivos **train.csv** ni **test.csv** separados).

1.2. Descripción de las Columnas

El dataset incluye las siguientes variables:

- **Date**: Fecha de la observación diaria (formato de cadena de texto que deberá convertirse a tipo fecha/tiempo).
- **Temp**: Temperatura mínima diaria registrada en grados Celsius.

1.3. Indicaciones del Proyecto

En este proyecto parcial, los estudiantes deberán:

1. Exploración y análisis inicial de la serie de tiempo:

- Cargar y visualizar los datos contenidos en **1_Daily_minimum_temps.csv**.
- Convertir la columna **Date** a un tipo de dato de fecha y ordenar cronológicamente la serie.
- Identificar valores faltantes, rangos de fechas y estadísticas descriptivas básicas de la temperatura (mínimo, máximo, media, mediana, etc.).
- Realizar visualizaciones de la serie de tiempo (gráfica de línea) y, de ser posible, explorar patrones de tendencia, estacionalidad o ciclos.

2. Preparación y preprocesamiento de los datos:

- Manejar valores faltantes o atípicos en la serie (si los hubiera), justificando la estrategia utilizada.

- Dividir la serie en **conjunto de entrenamiento** y **conjunto de prueba** respetando el orden temporal (sin mezclar ni barajar observaciones).
- Aplicar técnicas de escalado o normalización de la variable **Temp** cuando el modelo lo requiera.
- Generar características relevantes para el modelo de pronóstico, por ejemplo:
 - Ventanas temporales (*lags*) de la temperatura de días anteriores.
 - Promedios móviles u otras transformaciones que consideren la dependencia temporal.
 - Variables derivadas de la fecha (mes, día del año, etc.), si se consideran útiles.

3. Construcción del modelo de pronóstico:

- Definir claramente el **horizonte de pronóstico** que se desea abordar (por ejemplo, predecir la temperatura de los próximos 7, 14 o 30 días), justificando dicha elección.
- Entrenar al menos un **modelo de pronóstico** basado en **Redes Neuronales** (por ejemplo, modelos tipo MLP con ventanas, LSTM, GRU, etc.); sin embargo, pueden comparar contra otros modelos clásicos de series de tiempo si lo consideran conveniente.
- Ajustar la arquitectura y los hiperparámetros del modelo (número de neuronas, capas, épocas de entrenamiento, tamaño de ventana, etc.) en función del rendimiento obtenido.

4. Evaluación del modelo:

- Evaluar el desempeño del modelo en el conjunto de prueba usando métricas adecuadas para pronóstico de series de tiempo, tales como:
 - *MAE* (Mean Absolute Error).
 - *RMSE* (Root Mean Squared Error).
 - *MAPE* (Mean Absolute Percentage Error), si tiene sentido en el contexto.
- Analizar errores típicos del modelo (por ejemplo, días en los que el pronóstico se desvía mucho de la realidad) y proponer posibles explicaciones.

5. Generación del pronóstico y visualización de resultados:

- Generar los pronósticos para el horizonte definido utilizando el conjunto de prueba.
- Presentar gráficas comparando **valores reales** vs. **valores pronosticados** en el conjunto de prueba.
- Incluir al menos una tabla o resumen numérico donde se muestren algunos de los últimos días de la serie con su valor real y el valor pronosticado.
- (Opcional) Extender el pronóstico algunos días más allá del final de la serie original y discutir la plausibilidad de los resultados.

6. Entrega final (reporte y código):

- **Reporte en PDF** donde se describan:
 - Introducción y objetivo del proyecto.
 - Descripción del dataset y de la serie de tiempo.
 - Metodología seguida (preprocesamiento, modelo elegido, configuración del experimento).
 - Resultados obtenidos (métricas, gráficas, tablas de pronóstico).
 - **Conclusiones** y posibles líneas de mejora del modelo.
- **Código fuente completo y documentado**, ya sea anexado al reporte (en apéndice) o compartido mediante un repositorio, de forma que se pueda reproducir el análisis y el pronóstico.