

MCD407
Ingeniería de datos
Proyecto de fin de curso

5 de diciembre de 2024
M.Sc. Leydi Cruz

Relación Clima y Humedad del Suelo

Se cuenta con 2 datasets, uno de ellos contiene datos climáticos recopilados durante las últimas dos gestiones y el otro dataset contiene además de los valores de temperatura y humedad ambiente valores de humedad del suelo.

El dataset que contiene valores de humedad de suelo contiene información desde brote de cultivos de maíz hasta su etapa de llenado de grano. A partir de la información contenida en ambos datasets se buscar hallar qué variables climáticas podrían estar relacionadas con la humedad del suelo, y el impacto de ciertas variables climáticas sobre el cultivo. Para ello realizar:

1. (a) Un análisis exploratorio en ambos dataset, luego verificar ¿existe posibilidad alinear los períodos temporales en ambos?. **(30 Puntos)**
- (b) Limpiar los dos dataset (detectar y corregir errores en ambos datasets ej. valores faltantes, duplicados, etc.) para garantizar que los datos ruidosos en ambos datasets han sido eliminados por completo. Finalmente alinear los periodos temporales en ambos. **(40 Puntos)**
- (c) Evaluar correlaciones entre las variables climáticas y la humedad del suelo. **(15 Puntos)**
- (d) Probar modelos lineales simples para predecir la humedad del suelo a partir de variables climáticas. **(10 Puntos)**
- (e) Sabiendo que las condiciones ideales de crecimiento para el maíz son entre temperaturas mayores a 10 °C y menores a 35 °C, crees que la temperatura ambiente fue favorable en la campaña agrícola en cuestión?. **(10 Puntos)**

- (f) Sabiendo que las condiciones ideales de crecimiento para el maíz durante todo el ciclo agrícola son de 400mm de precipitación pluvial, crees que lluvia caída en fue favorable en la campaña agrícola en cuestión?. (10 Puntos)

Rúbrica de Evaluación para el Proyecto de Fin de Curso

En el desarrollo del proyecto se espera que se cubran los siguientes puntos:

Criterio	Descripción
Definición del Problema	Claridad en la formulación del problema, objetivos bien definidos y descripción adecuada de los datos proporcionados.
Preprocesamiento	Limpieza de los datos: manejo de valores faltantes, detección/corrección de outliers y validación del dataset.
Metodología	Aplicación adecuada de técnicas analíticas y uso de herramientas relevantes (NumPy, Pandas, etc.). Estructura clara en el flujo de trabajo.
Análisis Estadístico	Cálculo y presentación de estadísticas descriptivas, correlaciones y otros análisis relevantes al problema planteado.
Visualización	Uso de gráficos claros y efectivos (Matplotlib, Seaborn, etc.) para respaldar el análisis y comunicar hallazgos.
Conclusiones e Insights	Resumen de los hallazgos, interpretación de resultados y conexión con el problema original. Incluye respuesta clara a las preguntas planteadas.
Presentación	Claridad y organización del reporte o presentación final. Código bien comentado y estructurado.