

Predicción de Condiciones Climáticas en Puebla

Objetivo de Desarrollo Sostenible (ODS): Acción por el Clima (ODS 13)

Contexto:

El cambio climático es una de las mayores amenazas a la estabilidad global, afectando a todas las naciones y comunidades. El ODS 13, "Acción por el Clima", tiene como objetivo abordar este desafío urgente mediante la adopción de medidas para combatir el cambio climático y sus impactos. Este proyecto busca contribuir a este objetivo mediante la predicción precisa de condiciones climáticas en Puebla.

Dataset y Preprocesamiento

Recopilación de Datos:

Se recopilaron datos meteorológicos de Hermanos Serdán International Airport en Puebla. Para abordar las variaciones temporales, el conjunto de datos se dividió en cinco periodos de tiempo. Este enfoque permitió una comprensión más profunda de las tendencias climáticas a lo largo de décadas.

Tareas de Preprocesamiento:

Manejo de Datos Faltantes: Se eliminaron columnas con más del 70% de valores faltantes para garantizar la integridad de los datos.

Imputación de Valores Faltantes: Se aplicaron estrategias específicas (mediana para numéricos, moda para categóricos) para abordar los valores faltantes y mantener la coherencia en el conjunto de datos.

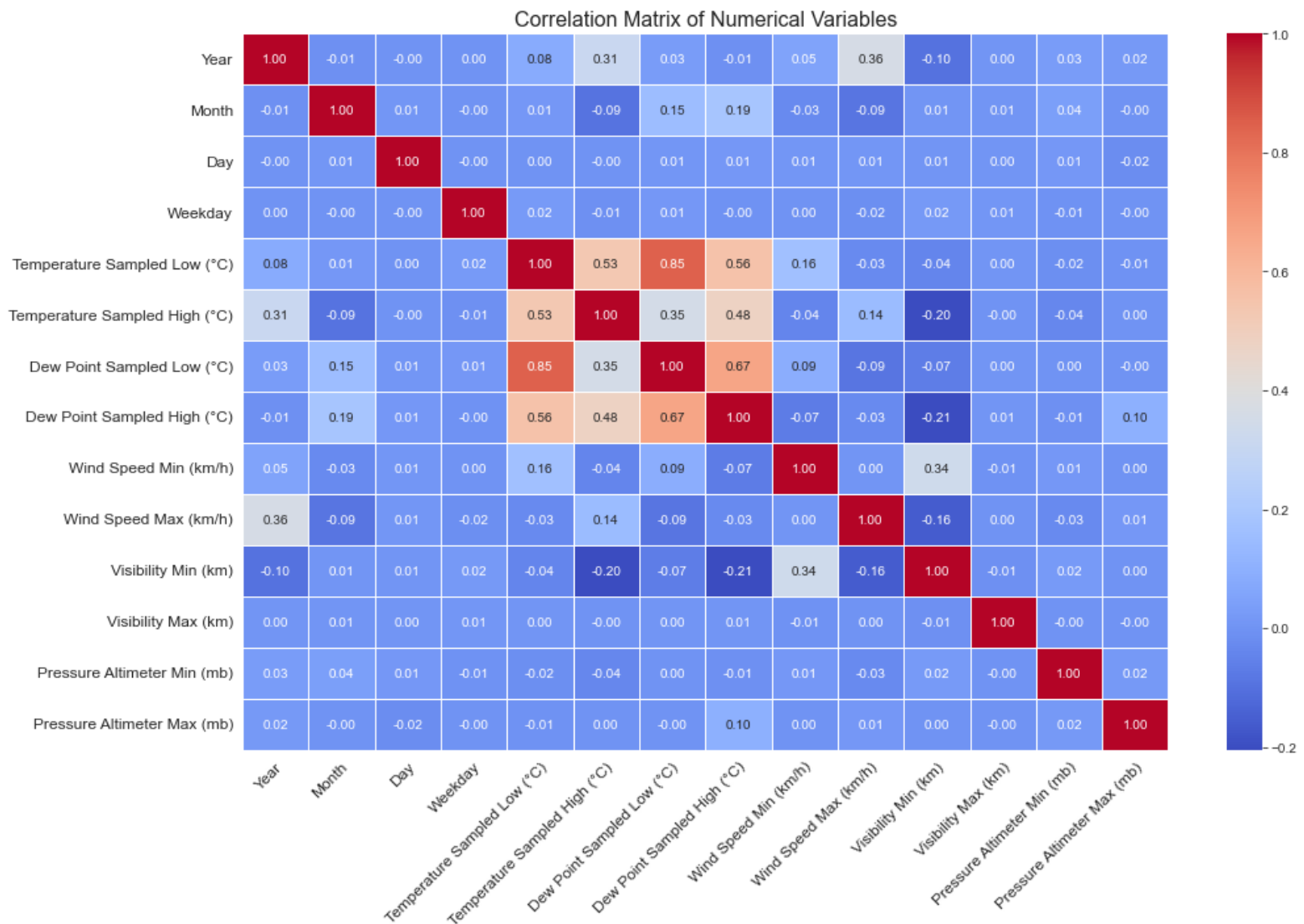
Características Temporales: Se extrajeron características temporales como año, mes, día y día de la semana para capturar posibles patrones estacionales.

Normalización de Unidades: Se ajustaron las unidades de ciertas variables para facilitar la interpretación y comparación.

Exploración y Visualización de Datos

Análisis Exploratorio:

Se exploraron las distribuciones de variables numéricas y se visualizó la correlación entre ellas. Estas exploraciones permitieron identificar posibles relaciones y patrones que podrían influir en las predicciones.



Resultados Clave:

Correlaciones Significativas: Se identificaron relaciones significativas entre ciertas variables, como la temperatura y la visibilidad.

Patrones Estacionales: La exploración reveló patrones estacionales que podrían ser cruciales para la precisión de las predicciones.

Modelos de Machine Learning

Linear Regression

El modelo de regresión lineal se entrenó y evaluó. Las métricas de desempeño en el conjunto de prueba sugieren un posible sobreajuste, ya que las predicciones en datos no vistos tienen un error mayor.

Métricas de Entrenamiento: MAE: 3.64, MSE: 2130.25, R²: 3.3%

Métricas de Prueba: MAE: 5.64, MSE: 84108.11, R²: 3.0%

Sobreajuste Potencial: Las métricas de prueba indican un posible sobreajuste, ya que el modelo tiene un rendimiento inferior en datos no vistos.

Áreas de Mejora: Se recomienda explorar variables adicionales y considerar modelos más complejos para mejorar la generalización.

Random Forest

Se implementó un modelo de Random Forest y se observó una mejora en el desempeño en comparación con la regresión lineal. Sin embargo, la diferencia entre las métricas de entrenamiento y prueba indica cierto sobreajuste.

Métricas de Entrenamiento: MAE: 1.043, MSE: 372.62, R^2 : 90.0%

Métricas de Prueba: MAE: 5.036, MSE: 85522.65, R^2 : 32.1%

Desempeño en Entrenamiento: El modelo se ajusta bien a los datos de entrenamiento, explicando el 90% de la variación en ciertas métricas.

Desafíos en Generalización: Aunque el desempeño es sólido en entrenamiento, las métricas de prueba sugieren desafíos en la generalización a nuevos datos.

Random Forest con Parámetros Óptimos

Se realizó una búsqueda de hiper parámetros para mejorar el modelo de Random Forest. Aunque hubo mejoras, las métricas aún indican posibles problemas de generalización.

Métricas de Entrenamiento: MAE: 2.27, MSE: 1691.55, R^2 : 51.7%

Métricas de Prueba: MAE: 4.801, MSE: 84159.96, R^2 : 33.5%

Mejoras Moderadas: Aunque hubo mejoras, las métricas indican que la generalización a datos no vistos sigue siendo un desafío.

Gradient Boosting

Se exploró el modelo Gradient Boosting, aunque las métricas sugieren cierta capacidad predictiva, aún hay desafíos en la generalización.

Métricas de Entrenamiento: MAE: 3.09, MSE: 1971.99, R^2 : 34.18%

Métricas de Prueba: MAE: 5.21, MSE: 84162.36, R^2 : 30.73%

Capacidad Moderada: El modelo muestra cierta capacidad predictiva, pero las métricas sugieren limitaciones en la generalización.

Análisis Integral:

- **Sobreajuste Observado:** La diferencia significativa entre métricas de entrenamiento y prueba indica la posibilidad de sobreajuste.
- **Revisión de Variables:** Se sugiere revisar las variables utilizadas para predecir la variable objetivo y explorar nuevas características.
- **Exploración Continua:** Modelos más avanzados y técnicas de ingeniería de características podrían mejorar la capacidad predictiva.

