

INFORME MISION 2 - CASO 2

Informe de Análisis y Predicción de Precipitaciones Mensuales

CURSO INTELIGENCIA ARTIFICIAL – EXPLORADOR

VIRTI01-1-Inteligencia artificial Explorador

Básico - 2024 – 3 – L1- G4

PRESENTADO POR GRUPO 8

Oscar David Guerra Hernández

David Tapias Valencia

Diana Moya

Juan Fernando Diez Zapata

Juan Pablo Dueñas M

Sandra Milena Gomez

Walkira Heredia Abril

DOCENTE:

CARLOS PINEDA

NOVIEMBRE 2024

Informe de Análisis y Predicción de Precipitaciones

Introducción

El análisis de datos históricos de precipitaciones es esencial para mejorar la planificación en sectores sensibles al clima, como la agricultura y el manejo de recursos hídricos. Este informe detalla el proceso de construcción y evaluación de un modelo predictivo que estima la precipitación mensual en diciembre utilizando datos de los 11 meses anteriores. Se aplicaron técnicas de machine learning, como Regresión Lineal y Random Forest Regressor, para obtener predicciones precisas.

El estudio abarca la exploración de datos, limpieza, preparación de características, entrenamiento y ajuste del modelo, y finalmente, la evaluación y predicción con aplicaciones prácticas claras.

Objetivo

Predecir las precipitaciones mensuales de diciembre basándose en datos históricos de los 11 meses anteriores, agrupados por cuencas. Este modelo tiene aplicaciones prácticas en la planificación agrícola, la gestión del agua y el monitoreo climático.

Metodología

1. Exploración y limpieza de los datos

- Carga del dataset: Se utilizó el archivo *Precipitaciondt.csv*. Se ajustaron codificaciones para garantizar su correcta lectura.
- Imputación de valores nulos:
 - Los valores numéricos, como en la columna JUNIO, se imputaron con la media.
 - Los valores categóricos, como CUENCA, se imputaron con el modo (valor más frecuente).
- Codificación: Se transformaron las columnas categóricas (CUENCA, ESTACIÓN, MUNICIPIO) en variables dummy mediante *One-Hot Encoding*.

2. Agrupación y preparación de datos

- Se calcularon los promedios de precipitaciones mensuales por cuenca.
- Las características para el modelo incluyeron las precipitaciones de ENERO a NOVIEMBRE (X), mientras que DICIEMBRE fue definido como el objetivo (Y).

3. Entrenamiento del modelo de predicción

- **División del conjunto de datos:**
 - 80% para entrenamiento y 20% para prueba.
- **Modelos utilizados:**
 - Modelo base: Regresión Lineal.
 - Modelo optimizado: Random Forest Regressor, ajustado mediante *GridSearchCV*.

4. Ajuste de hiperparámetros

Los mejores parámetros obtenidos fueron:

- `n_estimators`: 100 árboles.
- `max_depth`: Sin límite (árboles completos).
- `min_samples_split`: 2 muestras.
- `min_samples_leaf`: 1 muestra.

5. Evaluación del modelo

- Métricas utilizadas:
 - RMSE (Raíz del Error Cuadrático Medio): 7.69.
 - R^2 : 0.77, indicando que el modelo explica el 77% de la variación en los datos.
- Validación cruzada: Se utilizó para evitar sobreajuste y garantizar la generalización del modelo.
- Visualización: Se generaron gráficos comparativos entre predicciones y valores reales.

6. Predicción de diciembre

- El modelo predijo una precipitación de 85.16 mm para diciembre en una cuenca específica.

7. Guardado del modelo

- Se almacenó el modelo entrenado como *modelo_precipitaciones.pkl* para su uso futuro.

Resultados y Análisis

1. Exploración de datos

- Las estadísticas descriptivas iniciales mostraron variabilidad entre cuencas y meses.
- La limpieza y codificación eliminaron inconsistencias en los datos categóricos.

2. Evaluación de resultados

- Gráfico de predicciones vs. valores reales: Demostró una correlación adecuada, con pequeñas desviaciones.
- Rendimiento del modelo: El Random Forest superó significativamente al modelo base en precisión, con un RMSE bajo y un R^2 alto.

Limitaciones

- Dependencia de datos históricos: Las predicciones no consideran factores climáticos externos no incluidos en los datos, como eventos extremos.
- Agrupación por cuencas: Podría generar pérdida de información específica de estaciones o municipios.

Propuestas de mejora

1. Incluir variables climáticas adicionales, como temperatura o humedad.
2. Probar otros modelos avanzados, como Gradient Boosting o Redes Neuronales.
3. Ampliar la base de datos con registros más recientes y detallados.
4. Incorporar un análisis de sensibilidad para entender el impacto de cada mes en las predicciones.

Conclusiones

- El modelo Random Forest, optimizado con *GridSearchCV*, demostró ser adecuado para predecir precipitaciones mensuales con un error razonable (RMSE = 7.69).
- Este análisis es útil para aplicaciones en la agricultura y la gestión hídrica, ayudando a mitigar riesgos asociados al clima.
- El modelo entrenado fue guardado para facilitar su reutilización en futuros análisis.