

Predicción de Escasez de Agua en Presas Utilizando Modelos de Series Temporales y Técnicas de Clustering

Yossi Khebzou Dabbah A01028125
Jacques López Aldana A01746203
Darío López Cornú A01750876

1. Introducción

La escasez de agua es uno de los problemas más críticos y apremiantes a nivel mundial, y México se enfrenta a desafíos significativos en la gestión de sus recursos hídricos. Las presas juegan un papel vital al regular el flujo de agua para diversos usos, incluidos el consumo humano, la agricultura y la producción de energía hidroeléctrica. Sin embargo, factores como el cambio climático, la variabilidad en las precipitaciones y el aumento de la demanda están generando incertidumbre sobre la disponibilidad futura de agua. Es por esto que la necesidad de herramientas predictivas para planificar la gestión de las presas es cada vez más urgente.

El objetivo principal de este estudio es desarrollar un sistema predictivo basado en técnicas de series temporales que permita anticipar escenarios de escasez de agua en presas. Para ello, se emplean modelos estadísticos como ARIMA y Prophet, los cuales son ampliamente reconocidos en el campo de la predicción de series temporales. Estos modelos se utilizan para predecir los niveles de almacenamiento de agua a partir de datos históricos de precipitación, evaporación, y extracciones de agua. Además, se utiliza el algoritmo de K-Means para identificar grupos de presas que presentan comportamientos similares en cuanto a su respuesta a factores climáticos y de extracción, facilitando así la planificación estratégica en diferentes regiones.

Una de las innovaciones clave de este trabajo es la capacidad de proporcionar información predictiva no solo para la gestión de las presas existentes, sino también para futuros proyectos de construcción de presas. Los tomadores de decisiones, ingenieros y urbanistas que planeen la construcción de nuevas presas podrán hacerlo basándose en predicciones científicas rigurosas, que les permitirán evaluar el comportamiento esperado del almacenamiento de agua y las posibles fluctuaciones estacionales en las regiones seleccionadas. Este enfoque reduce la incertidumbre en la planificación de infraestructuras hidráulicas, proporcionando una base sólida para inversiones a largo plazo.

El análisis se centra en presas ubicadas en diferentes regiones de México y utiliza datos históricos desde el año 2000, lo que permite captar tanto las variaciones estacionales como las tendencias a largo plazo. Las herramientas utilizadas en este estudio permiten a las autoridades anticiparse a situaciones de escasez crítica y a gestionar de manera más eficiente los recursos disponibles. A través de una combinación de análisis predictivo y técnicas de agrupación, este proyecto ofrece un marco científico sólido para mejorar la toma de decisiones en la gestión y planificación de presas.

2. Trabajo Relacionado

El modelo ARIMA (Autoregressive Integrated Moving Average) es uno de los enfoques más tradicionales y ampliamente utilizados en la predicción de series temporales. Ha sido aplicado con éxito en la predicción de niveles de agua en embalses, donde los patrones históricos son claves para proyectar el comportamiento futuro. Estudios recientes, como los de Gupta y Kumar (2020), muestran cómo ARIMA puede ser combinado con modelos de aprendizaje automático para mejorar la precisión en la predicción de los flujos de entrada diarios de agua en embalses. Este enfoque ha sido utilizado con éxito en lugares como la presa Angat en Filipinas, donde los patrones estacionales juegan un papel crucial en la gestión del almacenamiento de agua (Liponhay, Legara, Monterola, 2022). Asimismo, Lee et al. (2020) han demostrado la efectividad de ARIMA en la predicción de los flujos de entrada mensuales en la presa del río Soyang en Corea del Sur, lo que subraya su utilidad en la planificación y gestión de recursos hídricos.

Prophet, desarrollado por Facebook, ha ganado popularidad en la predicción de fenómenos naturales con estacionalidad, como las precipitaciones y las extracciones de agua en sistemas hidráulicos. Hamzah et al. (2021) aplicaron Prophet para modelar series temporales en estudios hidrológicos, encontrando que su capacidad para tratar datos faltantes y modelar estacionalidades anuales lo convierte en una herramienta útil en la predicción de comportamientos a largo plazo en embalses y presas.

Por otro lado, el algoritmo de K-Means ha sido aplicado con éxito para agrupar presas y embalses en función de la similitud de sus series temporales. Rashid et al. (2018) utilizaron K-Means para identificar patrones en embalses que compartían comportamientos similares frente a factores externos como la precipitación y la evaporación, lo que permite a los gestores de recursos hídricos realizar una planificación más eficiente en situaciones de escasez de agua. Esta técnica de agrupamiento ofrece una ventaja significativa en la identificación de presas con respuestas similares a las condiciones climáticas, mejorando así la toma de decisiones y la asignación de recursos en momentos críticos.

3. Recopilación de Datos

Los datos utilizados en este estudio han sido obtenidos de la plataforma de datos abiertos del Gobierno de México. Estos datos provienen de diversas fuentes, incluidas las estaciones de monitoreo ubicadas en presas y embalses en todo el país. La recopilación incluye información clave relacionada con los niveles de almacenamiento de agua, precipitaciones, evaporación, y extracciones de agua, lo cual es fundamental para realizar predicciones sobre la escasez futura en las presas.

La CONAGUA proporciona datos históricos detallados que abarcan desde el año 2000 hasta el presente, lo que permite captar tanto las variaciones estacionales como las tendencias a largo plazo. Estos datos han sido recolectados y almacenados en formato CSV, facilitando su manipulación y análisis en este proyecto. Todos los conjuntos de datos utilizados están disponibles públicamente y se han documentado de manera exhaustiva. Los archivos se han descargado directamente desde la plataforma de datos abiertos de la CONAGUA y estarán accesibles en el repositorio de GitHub de este proyecto para cualquier usuario que desee replicar o mejorar los análisis realizados.

4. Estructura de los Datos

Cada archivo CSV contiene información de una categoría específica, como precipitación, evaporación, o extracciones. La estructura de los datos es uniforme y sigue un formato tabular, donde cada fila corresponde a un registro diario de una presa en particular. A continuación se presenta la descripción de los principales archivos utilizados:

- `almacenamiento.csv`: Contiene datos sobre los niveles de almacenamiento de agua de cada presa. Las columnas incluyen la fecha de medición y el volumen de agua almacenado en millones de metros cúbicos (mm^3).
- `precipitacion.csv`: Incluye los registros diarios de precipitación en milímetros (mm) para cada presa monitoreada.
- `evaporacion.csv`: Proporciona datos de evaporación diaria en milímetros (mm) para cada presa, lo que permite modelar la pérdida de agua por factores atmosféricos.
- `extracciones por obra de toma.csv`: Recoge los volúmenes de agua extraídos diariamente a través de obras de toma para diversos usos.
- `extracciones por vertedor.csv`: Incluye datos sobre las extracciones de agua realizadas a través de vertederos de las presas.

Cada uno de estos archivos contiene una columna que indica la fecha de medición (en formato YYYY-MM-DD) y las columnas correspondientes a los nombres

de las presas (identificadas por sus códigos). Los datos están organizados en formato de serie temporal, lo que facilita su integración en los modelos predictivos como ARIMA, Prophet, y técnicas de clustering como K-Means.

5. Disponibilidad de los Datos

Todos los datos empleados en este proyecto estarán disponibles en el siguiente repositorio de GitHub: <https://github.com/yossitec/Proyecto-Sustentabilidad>,

6. Metodología

El análisis realizado en este proyecto se centra en la predicción y agrupación de datos hidrológicos para diferentes presas utilizando series temporales. Para ello, se implementaron tres enfoques principales: ARIMA, Prophet y K-Means Clustering. A continuación, se describen los pasos específicos y el código desarrollado para llevar a cabo este análisis.

6.1. Preprocesamiento de Datos

Antes de realizar cualquier análisis predictivo o de agrupamiento, se llevó a cabo un preprocesamiento de los datos. Este paso fue crucial para limpiar y transformar los datos en un formato adecuado para los modelos.

El código utilizado para cargar y preprocesar los datos fue el siguiente:

```
def cargar_datos(ruta, columna_fecha='fecha'):
    df = pd.read_csv(ruta)
    df[columna_fecha] = pd.to_datetime(df[columna_fecha])
    df.set_index(columna_fecha, inplace=True)
```

Se emplearon archivos CSV provenientes del Gobierno de México, los cuales incluyen datos de almacenamiento, precipitaciones, evaporación y extracciones de agua. El preprocesamiento se encargó de gestionar valores faltantes, mediante la imputación de medias o valores más cercano, dependiendo de cuántos datos faltaran y asegurar que las series temporales estuvieran correctamente alineadas.

6.2. Predicción con ARIMA

Para la predicción de variables hidrológicas, como el nivel de almacenamiento o las extracciones de agua, se utilizó el modelo ARIMA. Este modelo fue seleccionado debido a su capacidad para capturar tendencias y patrones en series temporales. Se implementó de la siguiente manera:

```
def hacer_prediccion_arima(datos, columna_presa, pasos_prediccion=12):
    presa_data = datos[['fecha', columna_presa]].set_index('fecha').dropna()
    model = ARIMA(presa_data, order=(5,1,0))
```

```

model_fit = model.fit()
predicciones = model_fit.forecast(steps=pasos_prediccion)

```

En este código, los datos históricos de una presa específica se seleccionan y se ajustan a un modelo ARIMA. Posteriormente, se generan predicciones para los próximos 12 meses. Este enfoque es particularmente útil cuando se dispone de una serie temporal continua y robusta de datos históricos, como es el caso de las presas monitorizadas por el Gobierno de México.

6.3. Predicción con Prophet

El modelo Prophet se utilizó para predecir fenómenos con estacionalidad clara, como la precipitación y la evaporación. Prophet fue elegido por su capacidad para manejar datos con estacionalidad anual y por su flexibilidad ante datos faltantes o irregulares. La implementación del modelo Prophet fue la siguiente:

```

def aplicar_prophet_interactivo(dataset_name, presa_name, dias_prediccion=365, anio_inicio=2000):
    datos_filtrados_por_anio = datos[datos.index.year >= anio_inicio]
    df_prophet = datos_filtrados_por_anio.reset_index().rename(columns={'fecha': 'ds', presa_name: 'y'})
    modelo = Prophet()
    modelo.fit(df_prophet)
    futuro = modelo.make_future_dataframe(periods=dias_prediccion)
    predicciones = modelo.predict(futuro)
    fig = modelo.plot(predicciones)

```

Prophet fue utilizado para predecir fenómenos a largo plazo, como la evolución de las precipitaciones o las extracciones de agua. El modelo se ajusta a los datos históricos de una presa y permite realizar predicciones a un horizonte de hasta 365 días. Esto es particularmente útil para planificar la disponibilidad de recursos hídricos en el futuro.

6.4. Agrupación con K-Means Clustering

Además de las predicciones, se utilizó K-Means Clustering para agrupar presas con comportamientos similares en función de sus datos hidrológicos. Este enfoque permitió identificar patrones entre las presas, lo que facilita que la gente que quiera construir una presa, sepa a qué características se asimilaría la suya y distribución de recursos. La implementación fue la siguiente:

```

def aplicar_kmeans_interactivo(dataset_name, n_muestras=50, n_clusters=4, anio_inicio=2000):
    datos_filtrados_por_anio = datos[datos.index.year >= anio_inicio]
    muestra_presas = datos_filtrados_por_anio.sample(n=n_muestras, axis=1)
    scaler = StandardScaler()
    data_scaled = scaler.fit_transform(muestra_presas.values.T)
    kmeans = KMeans(n_clusters=n_clusters, random_state=42)
    labels = kmeans.fit_predict(data_scaled)

```

K-Means agrupa las presas en diferentes clusters en función de la similitud en sus series temporales. Esto es útil para identificar presas que responden de manera similar a factores como las precipitaciones o la evaporación, y para generar estrategias de gestión más eficientes.

6.5. Visualización Interactiva

Para facilitar la comprensión y el análisis de los datos, se implementaron visualizaciones interactivas utilizando ipywidgets. Estas visualizaciones permiten al usuario seleccionar el dataset, la presa y otros parámetros, y observar tanto los datos históricos como las predicciones generadas por los modelos. El siguiente código muestra cómo se habilitaron estas interacciones:

```
dataset_selector = widgets Dropdown(options=list(datasets.keys()), description='Dataset:')
presa_selector = widgets Dropdown(options=precipitacion.columns[1:], description='Presa:')
boton_prediccion = widgets.Button(description="Predecir")
display(dataset_selector, presa_selector, boton_prediccion)
```

Este enfoque interactivo permite al usuario explorar diferentes escenarios y visualizar los resultados de manera intuitiva, lo que facilita la toma de decisiones basada en datos.

7. Resultados

El programa desarrollado en este proyecto ha demostrado funcionar exitosamente en la predicción de fenómenos hidrológicos y en la agrupación de presas con características similares. Se utilizaron modelos predictivos como ARIMA y Prophet, así como el algoritmo de K-Means para el análisis de datos de series temporales, con resultados altamente satisfactorios. A continuación se resumen los principales resultados obtenidos:

7.1. Predicción de Escasez de Agua

El modelo ARIMA se aplicó a los datos históricos de almacenamiento, precipitación, evaporación, y extracciones de agua de varias presas mexicanas. Las predicciones generadas por este modelo han mostrado una alta precisión en el corto plazo, lo que permite anticipar posibles situaciones de escasez de agua. De manera similar, el modelo Prophet ha demostrado ser efectivo en la predicción a largo plazo, capturando tendencias estacionales y ajustándose a las fluctuaciones anuales de variables hidrológicas.

Las predicciones generadas permiten a los administradores de recursos hídricos prever escenarios de escasez crítica con antelación, y tomar decisiones informadas sobre la distribución y gestión de recursos. Además, este enfoque puede ser aplicado en la planificación de nuevas infraestructuras hidráulicas, como la construcción de presas, brindando a los ingenieros y urbanistas información clave sobre los posibles comportamientos futuros del almacenamiento de agua en una región específica.

7.2. Agrupación de Presas por Comportamientos Similares

La implementación del algoritmo K-Means ha permitido agrupar presas en función de sus patrones de comportamiento ante variables externas, como la precipitación y la evaporación. Esto es particularmente útil para optimizar la gestión de recursos a nivel regional, ya que las presas que comparten características similares pueden ser gestionadas de manera conjunta. Los resultados muestran que las presas con condiciones climáticas similares responden de manera análoga a los cambios en el entorno, lo que sugiere que los gestores hídricos pueden aplicar estrategias de manejo uniforme a los grupos de presas identificados.

7.3. Aplicaciones y Usos Potenciales

El programa desarrollado no solo se aplica exitosamente a la predicción de fenómenos hidrológicos en presas existentes, sino que también puede ser utilizado en la planificación de nuevas infraestructuras. Los ingenieros y urbanistas que planean construir presas o expandir las capacidades de almacenamiento de agua podrán beneficiarse de este análisis predictivo para anticipar el rendimiento de la infraestructura en escenarios futuros.

Además, el modelo es aplicable a diversas áreas dentro de la gestión del agua, incluyendo:

- Predicción de niveles de agua en embalses y presas, permitiendo una gestión más eficiente de los recursos hídricos.
- Identificación de presas con respuestas similares a fenómenos climáticos, optimizando la asignación de recursos en tiempos de escasez.
- Planificación a largo plazo de infraestructuras hidráulicas, considerando escenarios futuros de demanda de agua y disponibilidad de recursos.
- Análisis de la sostenibilidad de recursos hídricos, con la posibilidad de anticipar problemas de escasez y permitir la toma de decisiones proactiva.

7.4. Conclusión

El programa ha demostrado ser eficaz tanto para la predicción de escasez de agua como para la agrupación de presas en función de sus respuestas hidrológicas. Estos resultados proporcionan una herramienta valiosa para los gestores de recursos hídricos, que podrán optimizar la distribución del agua y anticipar problemas futuros. Asimismo, el modelo es una base sólida para futuros estudios y análisis sobre la sostenibilidad del agua en diferentes regiones de México.