Hydropower forecasting based on machine learning model

Prepared by

Yahya Ebrahim Elkholy

Supervised by

Dr. Mohammed Nasef

Menoufia University – Faculty of Science

Mathematics and Computer Science Department

(2024)

Acknowledgements

First and foremost, I thank God Almighty for enabling me to complete this work and for granting me success in this endeavor.

Secondly, I would like to express my deep gratitude to my family for their constant and devoted support, which has brought me to this stage.

Thirdly, I extend my sincere thanks and appreciation to my research supervisor, Dr. Mohamed Nasif, for his continuous and ongoing guidance, support, and valuable expertise throughout the duration of this project. His extensive knowledge and dedication in the fields of computer science and artificial intelligence have significantly influenced the development of this research and its final form. His insightful feedback and constructive criticism have greatly enhanced the quality of this graduation book.

I would also like to express our heartfelt appreciation to the Faculty of Science at Menoufia University for providing the necessary resources and facilities for conducting this research. The conducive academic environment and access to advanced technology have been crucial in carrying out our investigation into the school attendance system.

Thank you

–Yahya Elkholy–

Abstract

The world's climate has changed dramatically in recent years due to industrial development. Climate change significantly impacts hydropower plants, with one of the negative effects being the reduction in hydropower generation. This challenge is critical given the importance of hydropower in meeting increasing energy demands. The main purpose of this study is to predict future hydropower generation (2019-2024) for Norway. To achieve this goal, it is essential to study the water flow rates from previous years. After conducting thorough research to find a sufficient dataset, we obtained a comprehensive dataset that includes water flow data from 27 European countries from 1991 to 2019, with weekly readings. We aim to predict water flow and hydropower generation for the next five years using artificial intelligence (AI) and machine learning (ML) tools. Utilizing these technologies will enable us to provide well-informed recommendations to improve water resource management and optimize hydropower utilization.

Contents

Item ……………………………………………………page No.

list of Figures

Chapter 1

introduction

Chapter 1
Introduction

1.1 overview

Time-series Forecasting (TSF) occurs when you make scientific predictions based on historical time stamped data.

TSF Applications (Examples)

A retailer may be interested in predicting future sales at an SKU (stock keeping unit) level for planning and budgeting.

A small merchant may be interested in forecasting sales by store, so it can schedule the right resources (more people during busy periods and vice versa).

A software giant like Google may be interested in knowing the busiest hour of the day or busiest day of the week so that they can schedule server resources accordingly.

The health department may be interested in predicting the cumulative COVID vaccinations administered so that they can further predict when herd immunity is expected to kick in.

Predicting the TESLA stock price using time-series forecasting techniques.

TSF Approaches

- Classical / Statistical Models —Moving Averages, Exponential Smoothing, ARIMA, SARIMA, TBATS (Stats Models)
    - ARIMA (Auto Regressive Integrated Moving Average) Model
- Machine Learning —Linear Regression, XGBoost, Random Forest, or any ML model with reduction methods (Facebook Prophet)
- Deep Learning —RNN, LSTM (Neural Prophet) [1]

Hydro power is a renewable energy source that relies on the natural water cycle. It does not consume fossil fuels and does not produce harmful carbon emissions, making it a sustainable and environmentally friendly option. Hydro power significantly contributes to electricity generation in many countries, especially those with abundant water resources. It provides stable and reliable energy compared to other renewable sources like wind and solar power, which depend on weather conditions.

Hydro power plants can be used to regulate electrical loads, as they can store energy by pumping water to higher elevations and then use it to generate electricity when needed. This enhances the stability of power grids. Hydro power plants are often part of comprehensive water management systems that include flood control, irrigation water supply, and water quality improvement.

Predicting hydro power inflow enhances planning for water resource management and electricity generation, enabling better decision-making for dam and plant operations. Accurate predictions help avoid electricity generation shortages or surpluses, maintaining grid stability and reducing waste. They also assist in preparing for emergencies like floods or droughts, improve operational efficiency by optimizing maintenance and performance, aid in understanding climate change impacts on water resources, and boost economic returns by optimizing market operations. Overall, accurate predictions ensure the sustainable and efficient use of hydro power, balancing energy demand and natural resource preservation.

1.2 objective

In this project, we aim to forecast hydro power inflow for the next five years using the latest time series prediction techniques. We will employ two primary models [Facebook prophet – statistical models]:

Facebook Prophet:

Facebook Prophet is a state-of-the-art forecasting model known for its accuracy and robustness. Its performance metrics for our data are as follows:

Statistical Models:

We also employ traditional statistical models, which have been in use for nearly 50 years since their development in the 1970s. Although these models are well-established, they performed relatively poorly compared to the Prophet model. The performance metrics are as follows:

The project is organized as follows:

1.3 Data Analysis

We will begin with an in-depth analysis of the data to understand its characteristics and trends. This step is crucial for extracting meaningful insights and preparing the data for modeling.

1.4 Modeling

Next, we will use both the Facebook Prophet and traditional statistical models to make predictions. We will then compare their results to evaluate their performance.

This structured approach will provide a clear understanding of the data and the effectiveness of different forecasting methods. By comparing a modern, advanced model with a traditional one, we aim to highlight the advancements in time series forecasting and their practical implications for predicting hydro power inflow.

Chapter 2
Data analysis

Chapter 2

Data analysis

2.1 Dataset Information

I will conduct a detailed data analysis of water flow in 27 European countries to extract valuable information and insights for better understanding the data.

The dataset contains information on hydro power inflow for various European countries and includes the following columns:

year: The year

week: The week of the year

country_iso2: The country code using ISO 3166-1 alpha-2

inflow_mwh: Hydro power inflow in megawatt-hours (MWh)

2.2 Structure of the Jupyter Notebook:

1. Importing Libraries**:** Importing necessary libraries for data analysis and visualization [numpy – pandas – matplotlib – seaborn – plotly].
2. Importing Data**:** Loading the hydro power inflow data.
3. Data Wrangling: Cleaning and preparing the data for analysis.
4. Dealing with Date and Time**:**
   - Creating a new column [Date] from [year + week].
   - Creating a new column [Season] from [year + week].
5. Data Mapping**:**
   - Mapping data based on week.
   - Mapping data based on season.

   The interactive map: ../interactive_map.html

2.3 Exploratory Data Analysis (EDA):

1. Descriptive Statistics**:**
   - Overview of variables (mean, median, standard deviation, etc.).

2. Sorting Countries**:**
    - Sorting countries by inflow using group by.
3. Comparative Analysis**:**
    - Comparing minimum, maximum, and average values for:
        - Country
        - Year
        - Season
        - Week

Note: All values are calculated using the mean to avoid the influence of outliers.

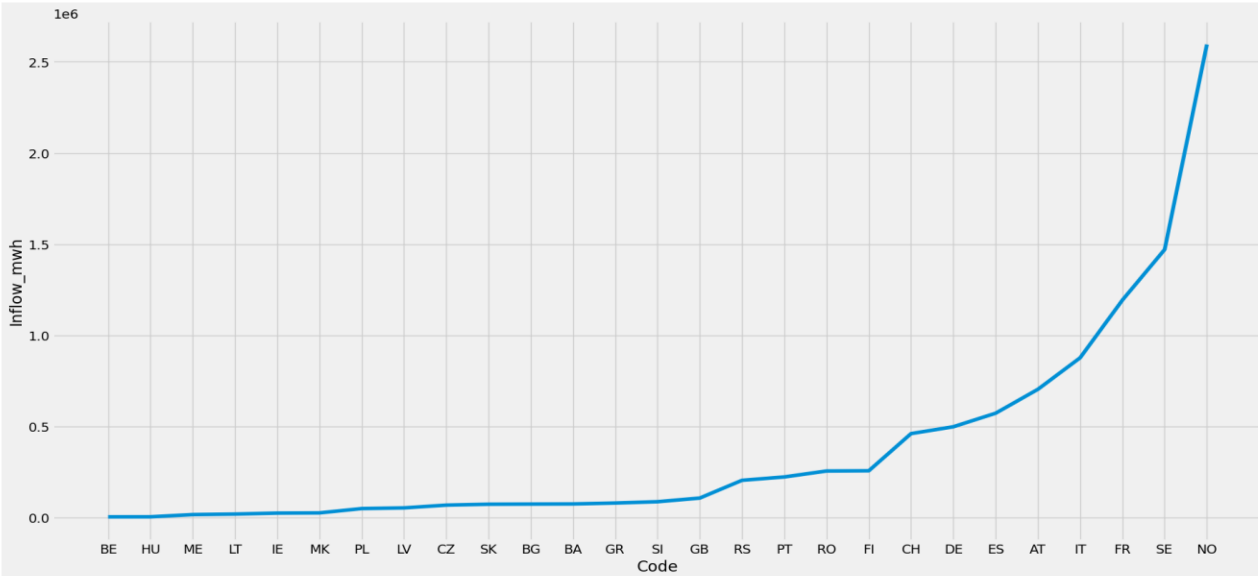


Figure 2.1: Sorting countries by inflow using group by hydro power inflow.

From Figure 2.1, it is evident that Norway excels in hydro power inflow due to its mountainous terrain, abundance of rivers and lakes, and heavy rainfall. Its significant investments in infrastructure and modern technology, along with supportive government policies, enhance its production capacity. The variation in hydro power inflow among European countries is attributed to differences in terrain, climate, investment in infrastructure, technological capabilities, and government policies and regulations supporting the use of water resources for energy generation.

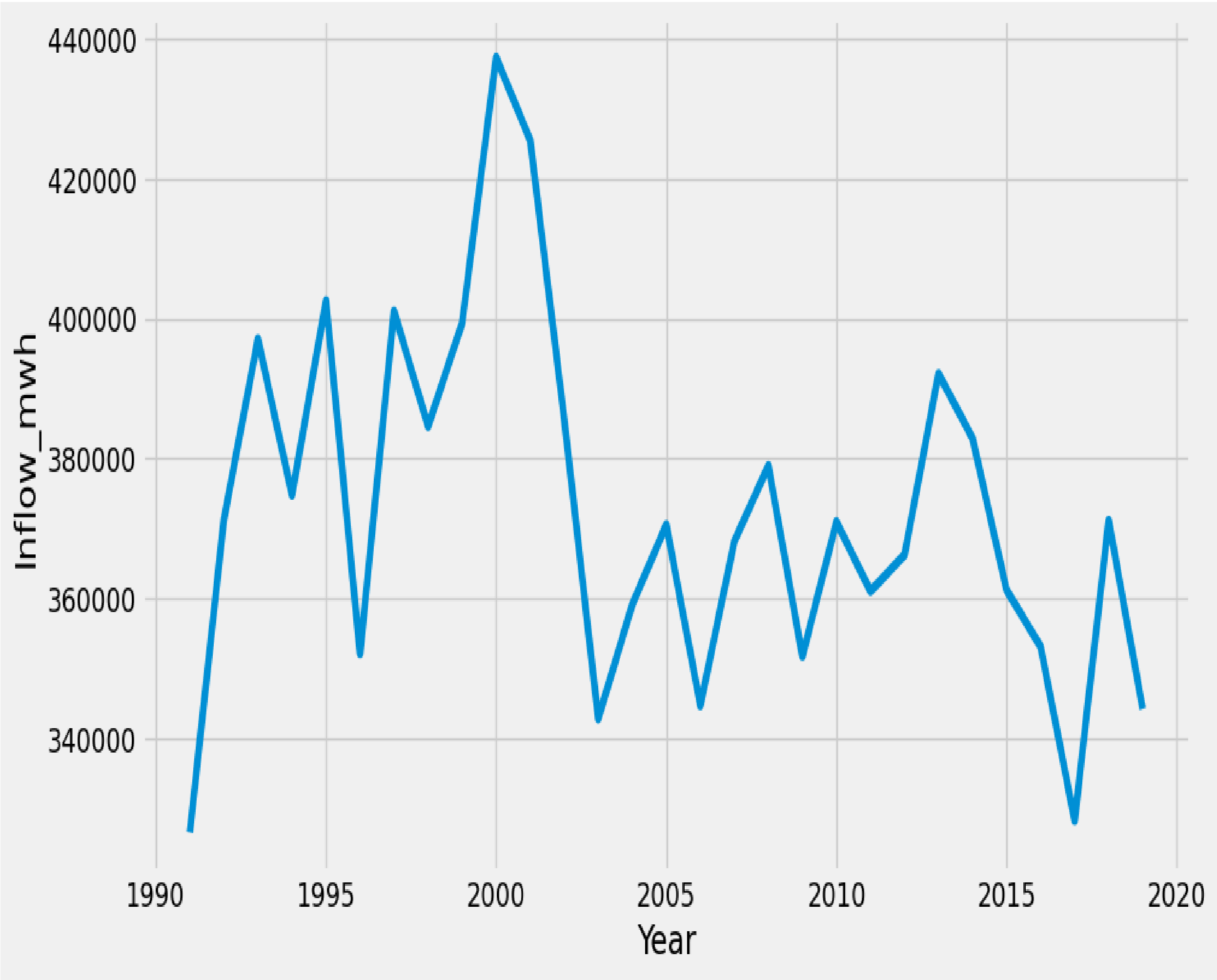


Figure 2.2: Sorting Inflow by Year

From Figure 2.2:

the maximum year of Inflow all over the world is 2000 with [437347.4450614518] mwh. the minimum year of Inflow all over the world is 1991 with [326526.79436490417] mwh.

the year containing the average value is 2010 with [370931.9589297571] mwh.

El Niño phenomenon:

El Niño is a climate phenomenon that regulates temperature in the Pacific region, affecting global economies. At times, it can lead to increased precipitation in certain regions, including Europe.[2]

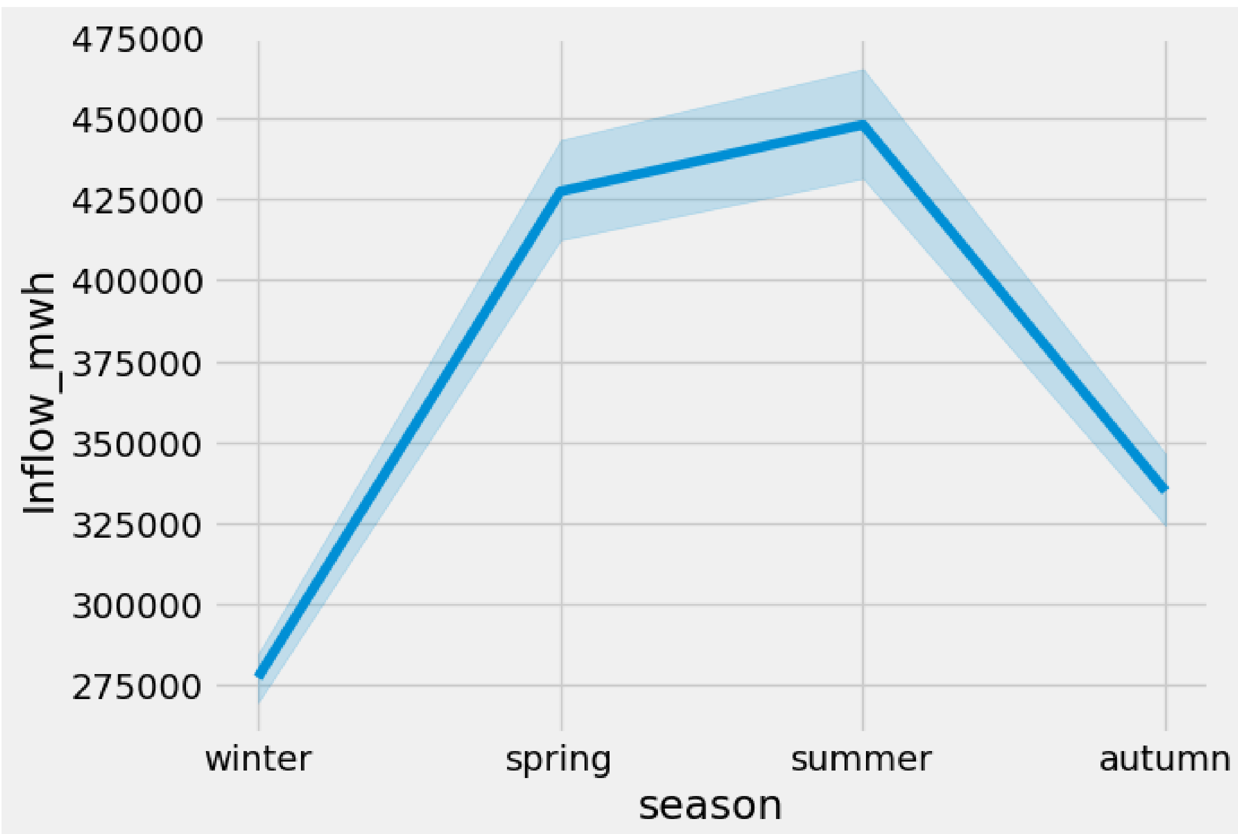


Figure 2.3: Sorting Inflow by Season

Based on Figure 2.3, the order of seasons in terms of water flow in Europe is as follows, from highest to lowest:

1. **Summer:** Highest average water flow. This is likely due to snowmelt in the mountains and increased rainfall in some areas.

2. **Spring:** Second highest average water flow. Snowmelt combined with warming temperatures contributes to increased water flow.
3. **Autumn:** Moderate water flow, less than summer and spring. This is a result of autumn rains that increase river and lake levels.
4. **Winter:** Lowest average water flow. Water is often frozen in many areas, reducing its flow.

Explanation of the results:

- **Summer:** Shows the highest average water flow, likely due to snow melting in the mountains and increased rainfall in some regions.
- **Spring:** Ranks second, with snowmelt and the onset of warmth contributing to increased water flow.
- **Autumn:** Shows moderate flow due to autumn rains that increase river and lake levels.
- **Winter:** Shows the lowest average water flow, as water is often frozen in many regions, reducing flow.

These results illustrate that water flow is significantly influenced by seasons and associated climatic variations. This should be considered when planning water resource management and hydroelectric power generation strategies.

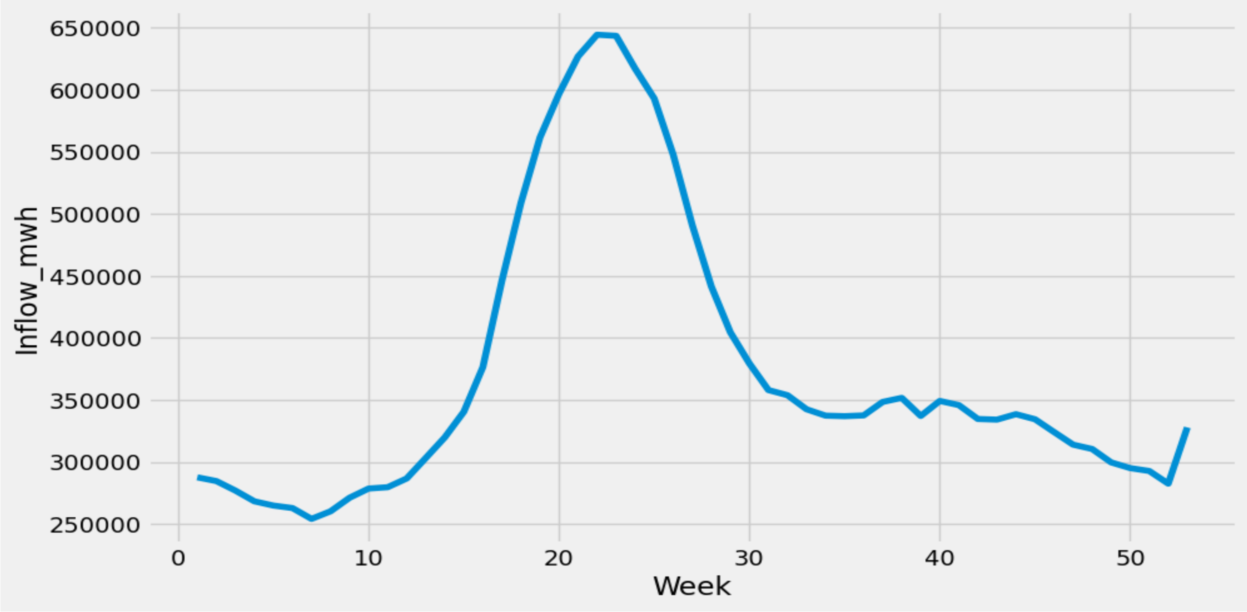


Figure 2.4: Sorting Inflow by Week

Based on Figure 2.4, The increase and decrease in water flow across seasons can be attributed to several factors that affect a specific region:

1. **Snowmelt:**
   - In spring (from week 11 to week 22), snow in the mountains begins to melt due to rising temperatures, leading to increased water flow.
   - Once most of the snow has melted, water flow begins to decrease.
2. **Rainfall:**
   - During autumn, there are often heavy rains that increase river and lake levels, contributing to higher water flow.
   - As the rainy season ends, water flow starts to decrease.
3. **Natural seasonal cycles:**
   - Nature follows natural seasonal cycles that affect water flow, such as snowmelt and annual rainfall.
4. **Climate variations:**
   - Climate changes can impact precipitation patterns and snowmelt, leading to fluctuations in water flow over the years.

Based on these factors, the increase and decrease in water flow across seasons can be a result of these natural phenomena. The specific weeks mentioned (from week 11 to week 22) encompass crucial periods related to snowmelt and rainfall, which lead to increased water flow, while other periods experience decreased flow due to other climatic factors.

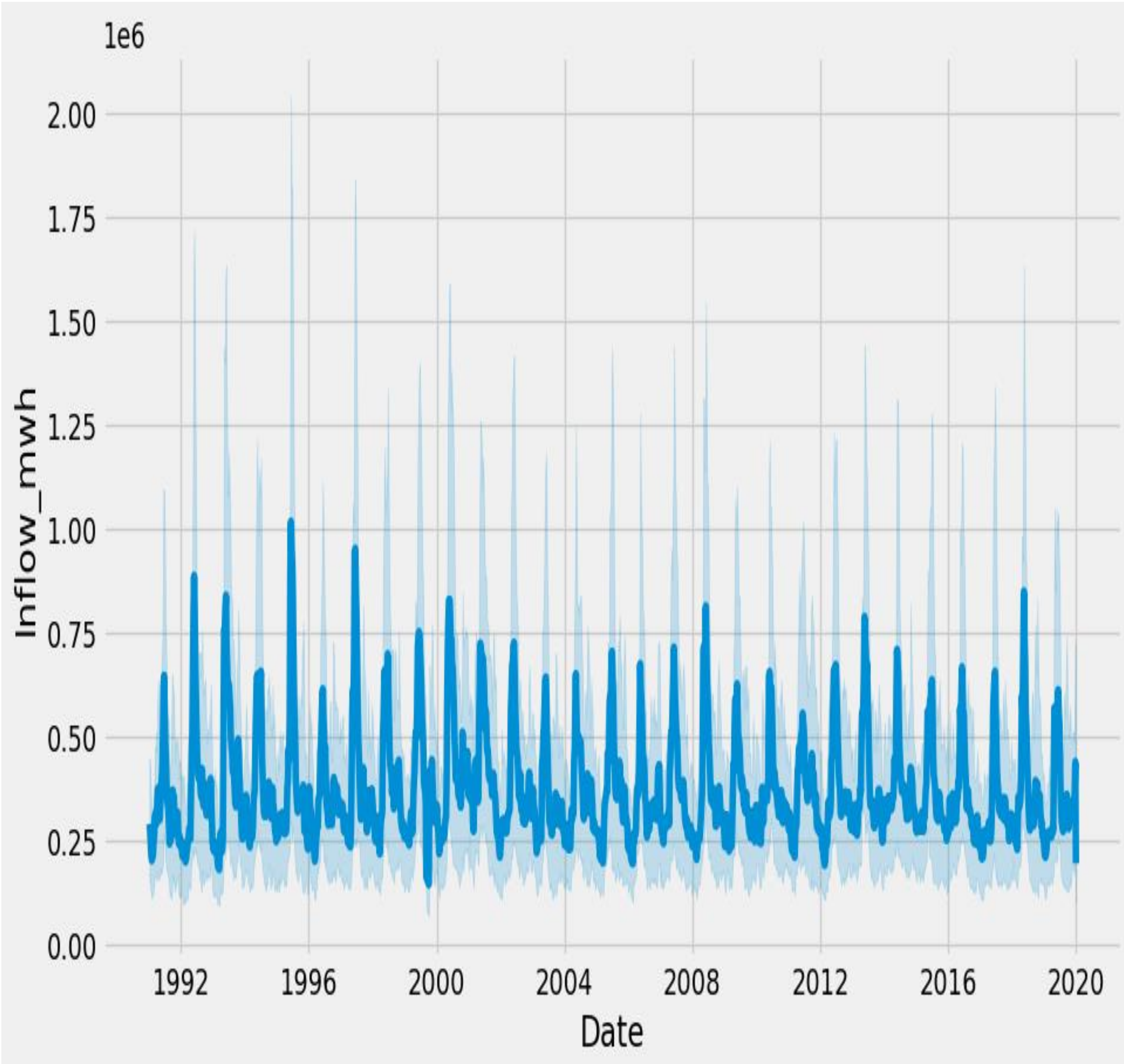


Figure 2.5: Line plot for European countries data

This figure shows the graphical representation of the dataset.

2.4 case study

We selected a portion of the data, specifically the data for Norway, because it is most suitable for the study due to its high values as previously mentioned. Therefore, we will apply data analysis and model building on it.

Explanation for Norway's Excellence in Hydropower Flow and the Differences Between Countries

1. **Norway as a Leader in Hydropower Flow:** Norway excels in hydropower flow for several reasons:

   Geographical and Climatic Nature:

   - **Mountains:** Norway has many high mountains that help in collecting rainwater and snow, which flow into rivers and lakes.
   - **Rivers and Lakes:** Norway contains numerous large rivers and lakes that contribute significantly to hydropower production.
   - **Heavy Rainfall:** Coastal areas in Norway receive substantial rainfall throughout the year, increasing water flow.

   Infrastructure and Technology:

   - **Investment in Hydropower:** Norway has heavily invested in building hydroelectric power plants and developing the technology used in this field.
   - **Operational Efficiency:** Norway uses modern, highly efficient techniques in the operation and maintenance of hydroelectric power plants.

   Government Policies:

   - **Government Support:** The Norwegian government significantly supports the hydropower sector through environmental and regulatory policies.
   - **Legislation:** Laws that encourage the use of renewable energy sources enhance the use of hydropower.

2. **Differences Between Countries in Hydropower Flow:** The disparity in hydropower flow among European countries is due to several factors:

   Terrain and Geography:

   - **Mountainous Countries:** Countries with mountains and rivers like Norway and Switzerland have substantial water resources, contributing to large amounts of hydropower production.
   - **Low-lying Countries:** Countries with flat, low-lying terrain, such as Belgium and the Netherlands, have less favorable water resources for hydroelectric power production.

   Climate:

   - **Rain and Snow:** Areas that receive significant amounts of rain and snow, such as Norway and Sweden, have higher water flow compared to regions with less rainfall.

   Infrastructure:

   - **Investment and Development:** Countries that have heavily invested in building and developing hydroelectric power plants, like Norway and France, have higher hydropower production capacity.
   - **Technological Capabilities:** Countries that use advanced technologies in operating power plants have higher efficiency in energy production.

   Policies and Regulations:

   - **Government Support:** Countries receiving strong government support for renewable energy achieve higher productivity.
   - **Legislation:** The presence of supportive legislation for using water resources in power generation enhances productivity.

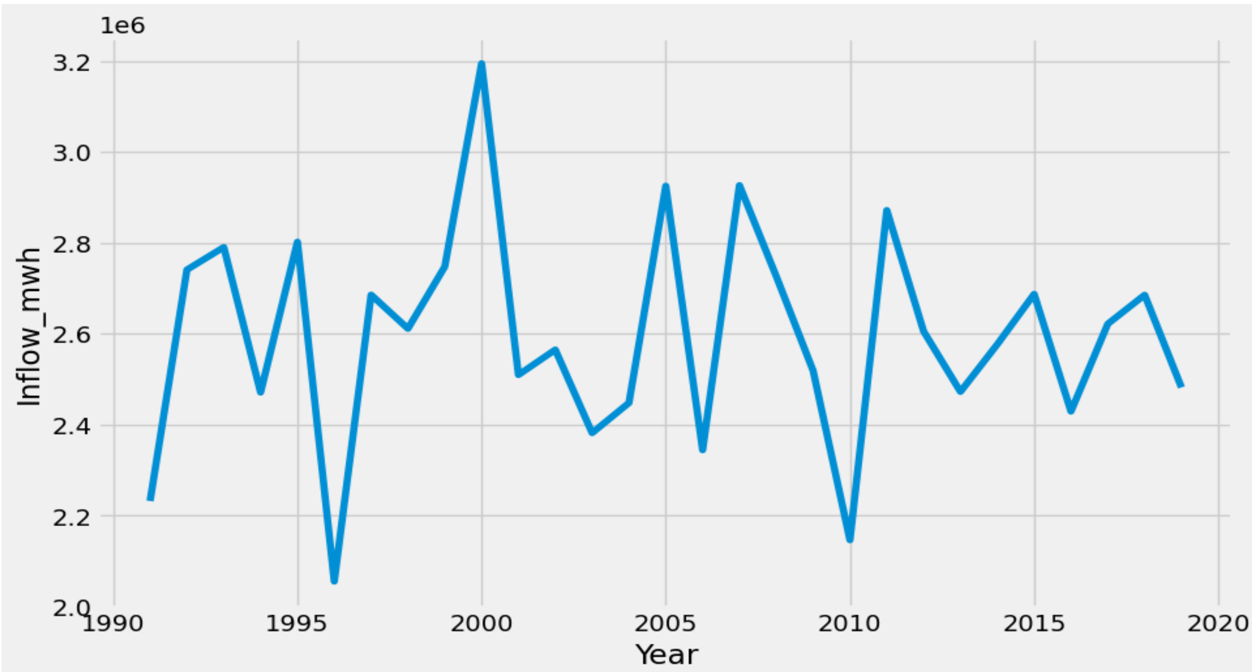


Figure 2.6: Sorting Inflow by Year of Norway

From Figure 2.6:

the maximum year of Inflow all over the world is 2000 with [3194881.3700283593] mwh.

the minimum year of Inflow all over the world is 1996 with [2055612.1639743575] mwh.

the year containing the average value is 2012 with [2605620.956505481] mwh.

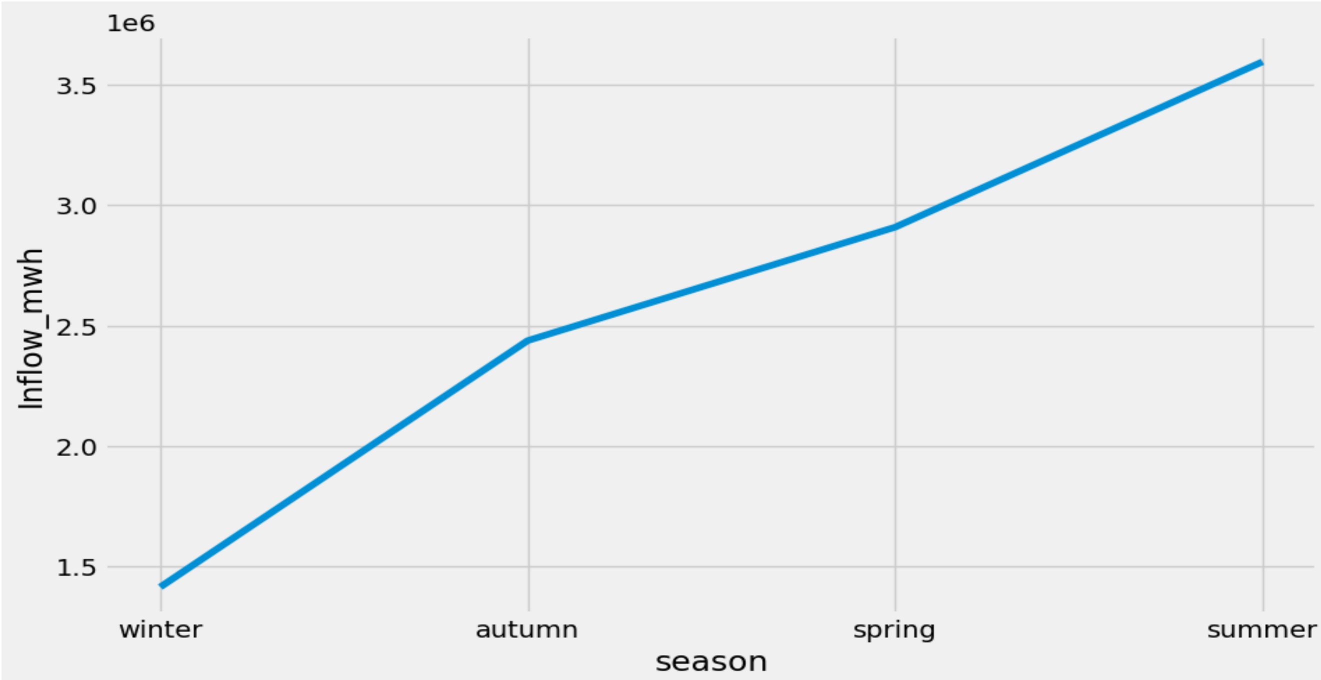


Figure 2.7: Sorting Inflow by Season of Norway

From Figure 2.7:

Summer: Highest average water flow.

Spring: Second highest average water flow.

Autumn: Moderate water flow.

Winter: Lowest average water flow.

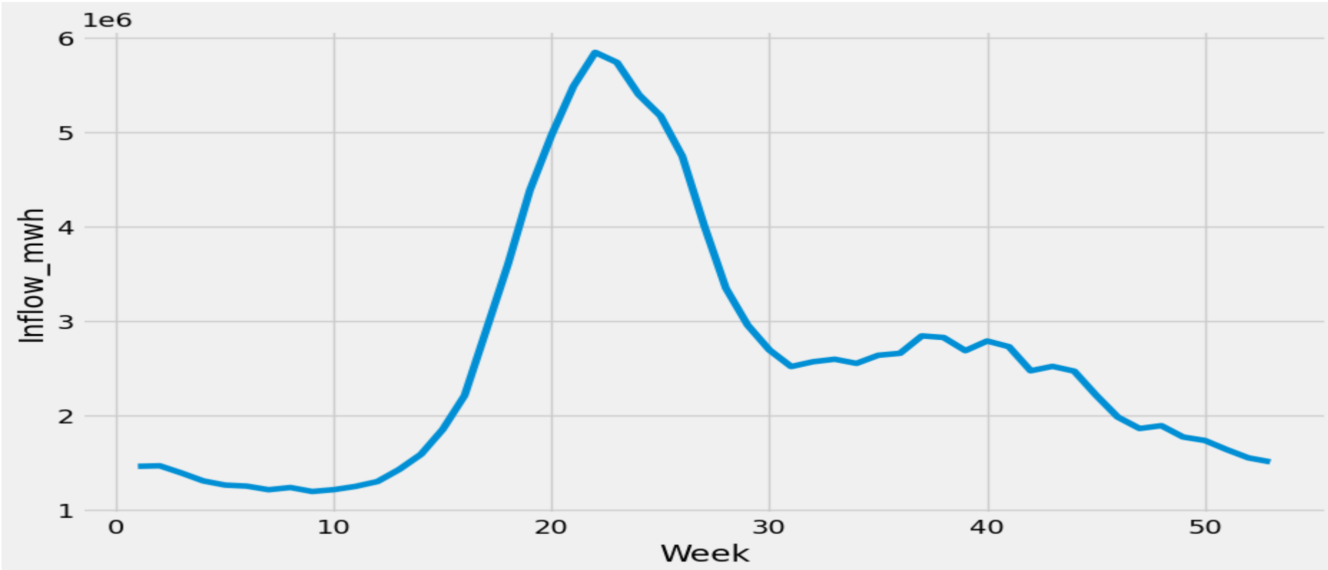


Figure 2.8: Sorting Inflow by Week of Norway

From Figure 2.8, the increase and decrease in water flow across seasons can be a result of these natural phenomena. The specific weeks mentioned (from week 11 to week 22) encompass crucial periods related to snowmelt and rainfall, which lead to increased water flow, while other periods experience decreased flow due to other climatic factors.

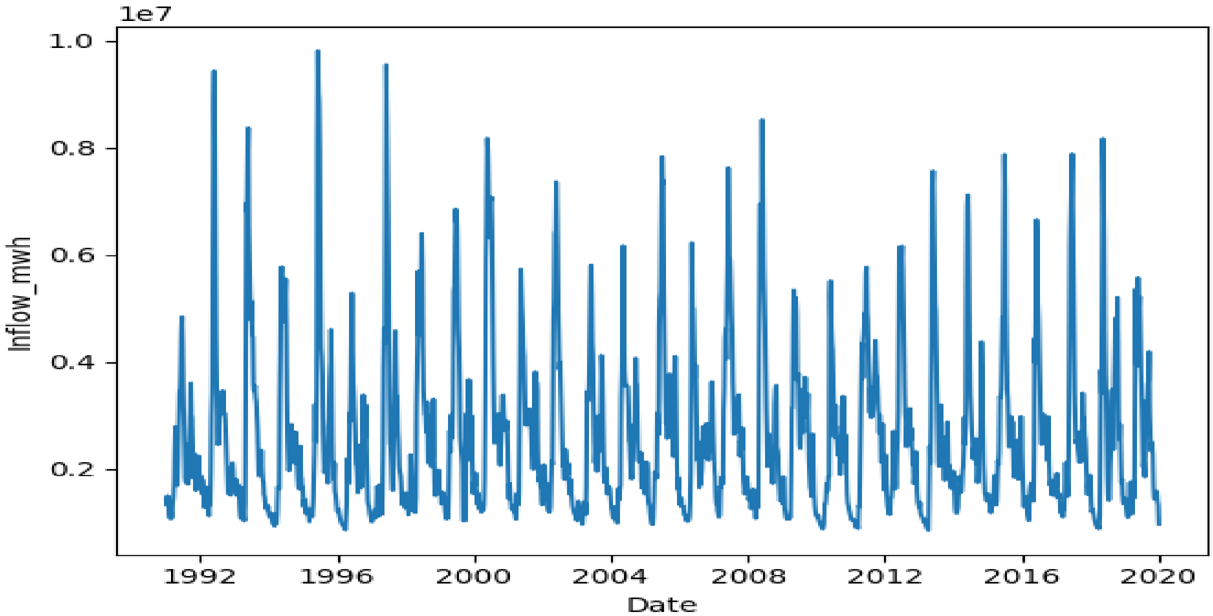


Figure 2.9: Line plot for Norway data

This figure shows the graphical representation of the dataset of Norway.

Chapter 3
Modeling

Chapter 3

Modeling

3.1 Time Series Forecasting

What is Time Series Forecasting**?** Time series forecasting is the process of using historical time-based data to predict future values. This type of forecasting relies on analyzing patterns and trends in the data over time to predict what will happen in the future. Time series are used in a wide range of fields such as economics, finance, weather, and sales.[3]

Time-series Forecasting Basics

•A time-series is a sequence where a metric is recorded over regular time intervals.

Hourly, daily, weekly, monthly, quarterly and annual.

•Why?

Time series analysis shows how data changes over time, and good forecasting can identify the direction in which the data is changing.

3.2 Advantages:

- **Accuracy of Prediction:** When time series data is well-available, predictions can be very accurate.
- **Automatic Detection of Seasonal Trends:** Time series models can detect seasonal patterns and trends over time.
- **Integration with Other Models:** They can be combined with other models to improve prediction accuracy.

3.3 Disadvantages:

- **Need for Extensive Historical Data:** Accurate time series forecasting requires large amounts of historical data.
- **Sensitivity to Sudden Changes:** Time series models can be ineffective in predicting sudden changes or unexpected events.
- **Mathematical Complexity:** Understanding and applying the mathematical and statistical models used can be complex.

3.4 Methods Used in Time Series Forecasting:

Traditional Statistical Models:

- **Simple Moving Average (SMA):** Uses the moving average of historical values to predict future values.
- **Autoregressive Integrated Moving Average (ARIMA):** A model used to analyze time series patterns and address trends and seasonality.

Linear Models:

- **Simple Linear Model:** Used to determine the linear relationship between time variables.
- **Multiple Linear Model:** Enhances prediction accuracy by using additional explanatory variables.

Facebook Prophet: An important and widely used tool for time series forecasting.

In this project we will apply two different models:

3.5 two different models

1- Autoregressive Integrated Moving Average (ARIMA)
2- Facebook Prophet.

1- Autoregressive Integrated Moving Average (ARIMA)[4]

Traditional Statistical Models in Time Series Forecasting:

1. Simple Moving Average (SMA):

What is the Simple Moving Average? The Simple Moving Average is a technique used to smooth a time series by calculating the average of data over a specific time period. The moving average for each time period is calculated by taking the simple average of the data values over that period.

How to Calculate SMA:

- Determine the number of time periods n to be used for calculating the moving average.
- For each point in the time series, calculate the simple average of the values over the previous n periods.

Uses:

- **Smoothing Data:** SMA helps reduce the impact of random fluctuations and noise in the data.
- **Forecasting:** SMA can be used to predict future values based on historical patterns.

Advantages:

- **Simplicity:** Easy to understand and implement.
- **Effective Smoothing:** Helps in visualizing the overall trends in the data.

Disadvantages:

- **Lagging:** The moving average may lag behind rapid changes in trends.

- **Sensitivity to Period Selection:** Results depend significantly on the chosen number of time periods.

2. Autoregressive Integrated Moving Average (ARIMA):

What is the ARIMA Model? ARIMA is a statistical model used to analyze and forecast time series. The model comprises three components: Autoregression (AR), Integration (I), and Moving Average (MA).

Explanation of ARIMA Components:

- **Autoregression (AR):** Represents the relationship between current values and previous values in the time series. It uses past values to predict current values.[5]

  AR(p) formula

  $$y t = \delta + \varphi 1 y t - 1 + \varphi 2 y t - 2 + \ldots + \varphi p y t - 1 + A t$$

  - y t - 1 , y t - 2…y t - p are the past series values (lags),
  - At is white noise (i.e. randomness),
  - where φ is the process mean Integration (I) [3].

- **Moving Average (MA):** Represents the relationship between current values and past errors. This component is used to smooth out the impact of random fluctuations.

ARIMA(p, d, q) Model:

- p: Number of time periods in the AR component.
- d: Number of differencing required to achieve stationarity.
- q: Number of time periods in the MA component.

How to Use the ARIMA Model:

1. **Examine the Data:** Check for stationarity and use differencing if the series is non-stationary.
2. **Determine Parameters:** Identify appropriate values for ppp, ddd, and qqq using tools like Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).
3. **Estimate the Model:** Use historical data to estimate the model coefficients.
4. **Forecast:** Use the model to predict future values.

Advantages:

- **Flexibility:** Can handle a variety of time series by adjusting parameters.
- **Accuracy:** Provides accurate forecasts when used correctly.

Disadvantages:

- **Complexity:** Can be difficult to understand and apply for beginners.
- **Need for Historical Data:** Requires substantial historical data to analyze patterns and make forecasts.

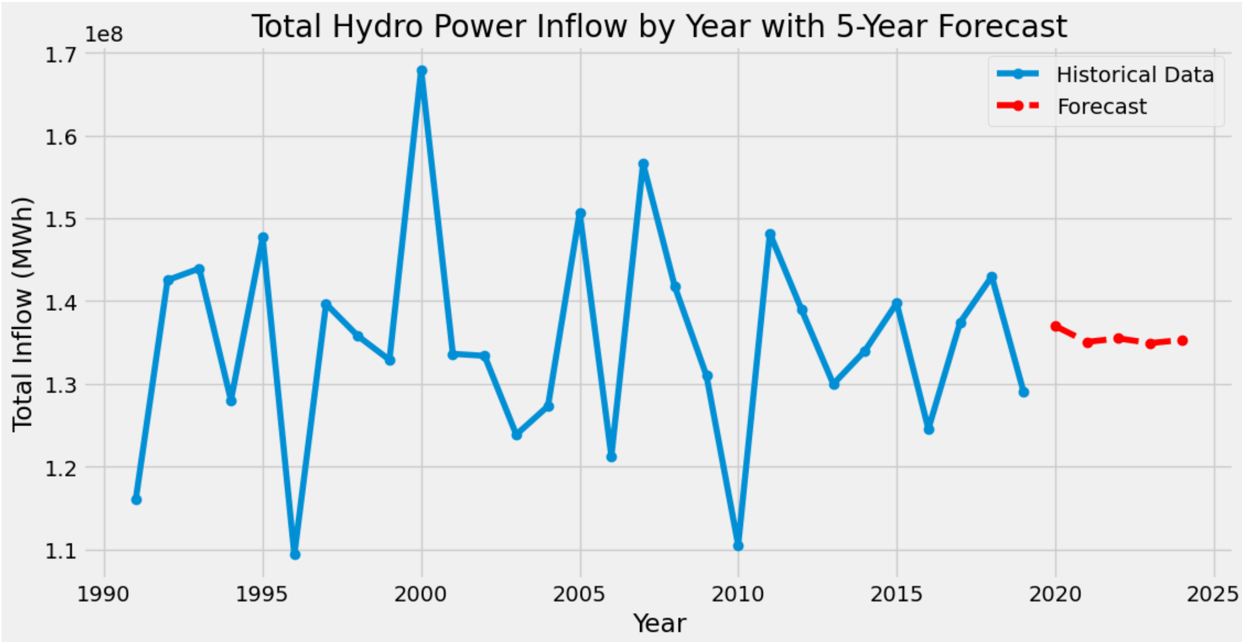


Figure 3.1: Line plot for forecasting data

In Figure 3.1, I plotted two lines: one for the historical data and the other for the forecast representing the predictions for the next five years. This allows me to assess the model's efficiency by comparing its performance to the original data, which showed weak predictive performance.

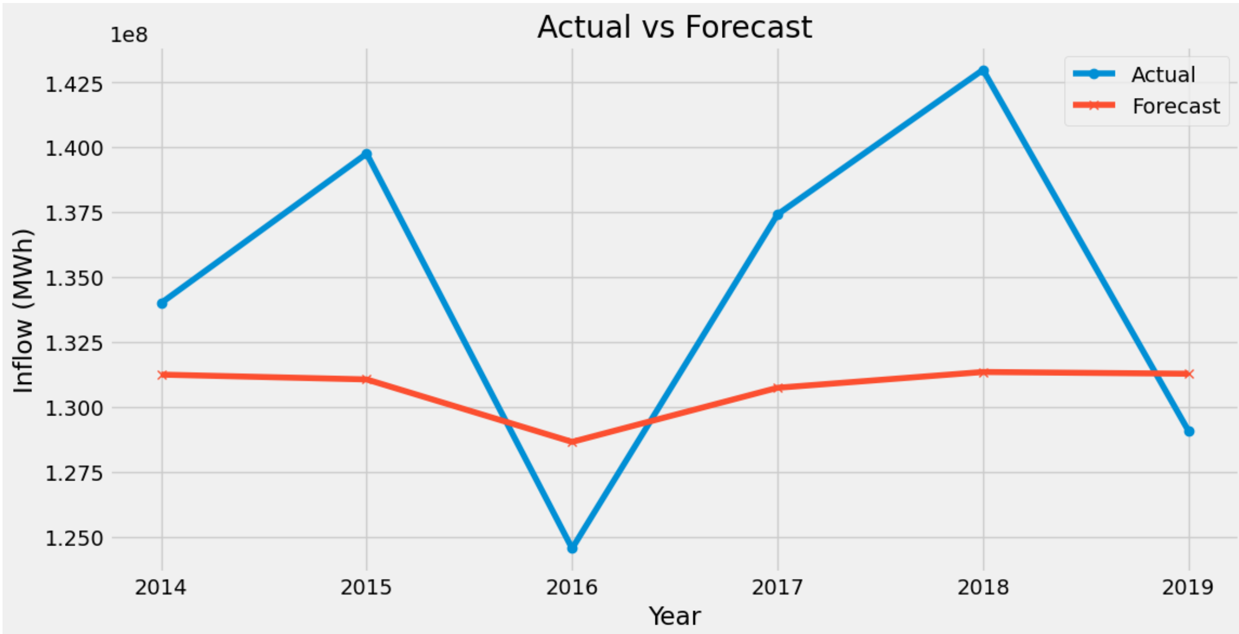


Figure 3.2: Actual data vs forecast data

Figure 3.4: Line plot for evaluating the model to determine its accuracy and measure its efficiency by dividing the data into three parts: train, validate, and test, with the validation process conducted on the last 5 years of actual data.

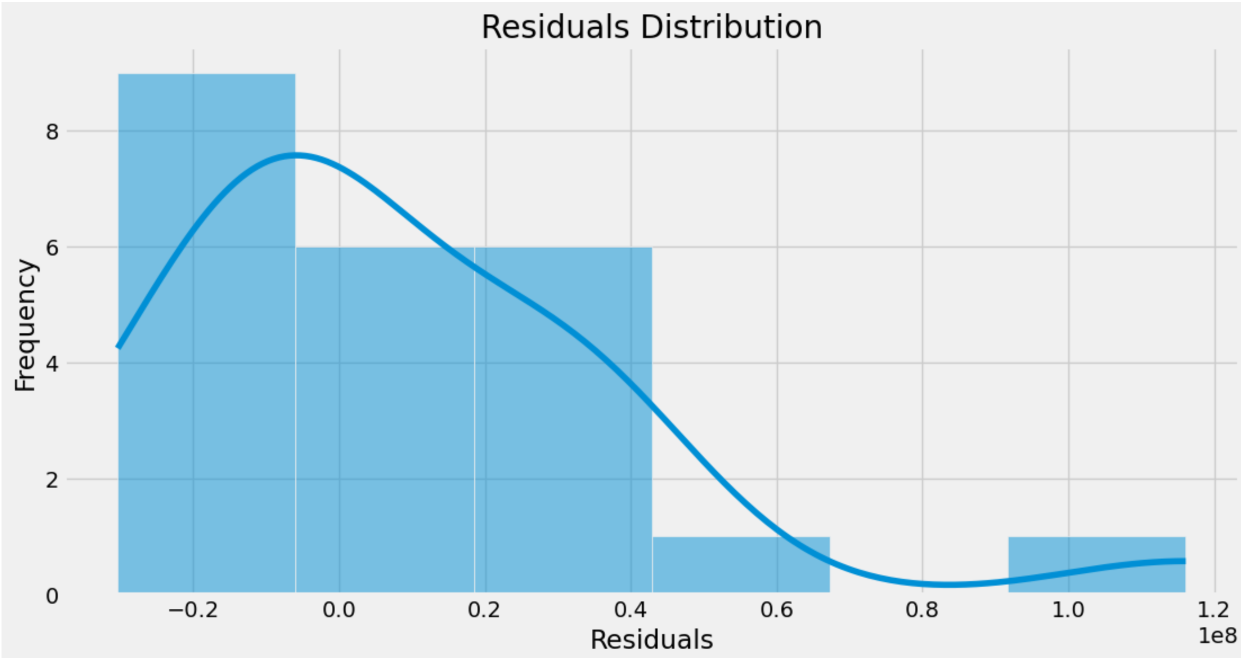


Figure 3.3: residual analysis

Residual analysis is the process of evaluating the differences between the actual values and the predicted values made by a statistical model or a machine learning model. Residuals are the differences between the actual values and the predicted values given by the model. The mathematical formula for residuals is:

Residual=Actual Value−Predicted Value

From Figure 3.3: A residual plot is a graph that displays residuals on the vertical axis (Y-axis) versus the predicted values or inputs on the horizontal axis (X-axis). There are two main types of residual plots:

1- Residuals vs. Fitted Values Plot

2- Histogram of Residuals

2- Facebook Prophet

What is Facebook Prophet**?** Facebook Prophet is an open-source tool for time series forecasting developed by Facebook's data team. It is designed to be easy to use and provide accurate forecasts while accounting for seasonal factors and holidays.[6]

Advantages:

- **Ease of use:** Prophet is designed to be easy to use, even for beginners in time series forecasting.
- **Flexibility:** It can handle data with multiple seasonal trends and account for holiday effects.
- **High accuracy:** It provides accurate forecasts by using an additive model that combines trends, seasonality, and holidays.
- **Handling missing data:** Prophet can handle missing data and irregular time intervals.
- **Interpretability:** The model provides clear interpretations of the different components of the forecast, such as seasonal trends and holiday effects.

Disadvantages:

- **Limited customization:** Despite its flexibility, advanced users might find limitations in customizing the model according to their specific needs.
- **Large data size:** It can be slow with very large datasets.

Relation to Machine Learning: Prophet is not a traditional machine learning model but a statistical model based on predictive analytics. However, it leverages some concepts from machine learning, such as model optimization and performance evaluation.

Methods Used in Facebook Prophet: Prophet uses an additive model consisting of three main components:

- **Trends:** Identifies the overall direction in the data over time.
- **Seasonality:** Detects recurring seasonal patterns, such as monthly or yearly trends.
- **Holidays:** Accounts for the effects of holidays or special events that may impact the data.

How to Use Facebook Prophet: Prophet can be easily used with programming languages like Python and R.

Facebook Prophet is a powerful and flexible tool for time series forecasting, characterized by its ease of use and high accuracy. Its design for interpretability and handling of missing data and irregular time intervals makes Prophet an excellent choice for a wide range of time series forecasting applications.

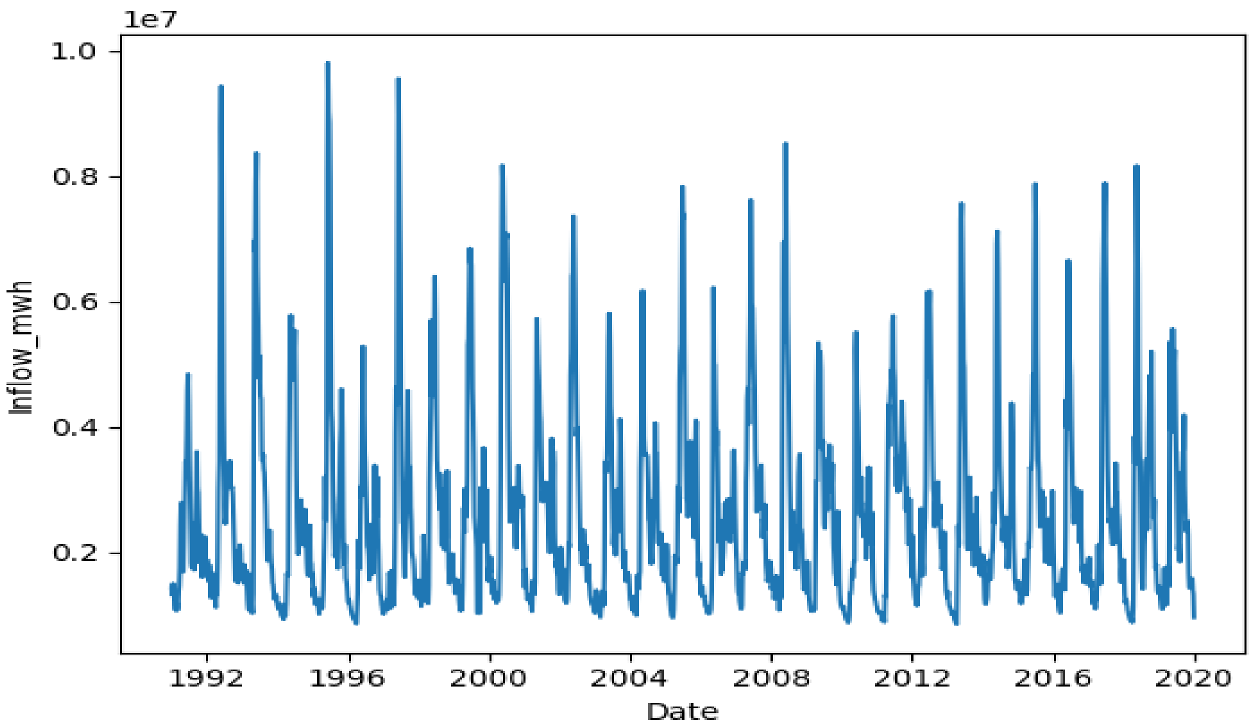


Figure 3.4: Actual data

From Figure 2.14: A plot showing the normal distribution of the data.

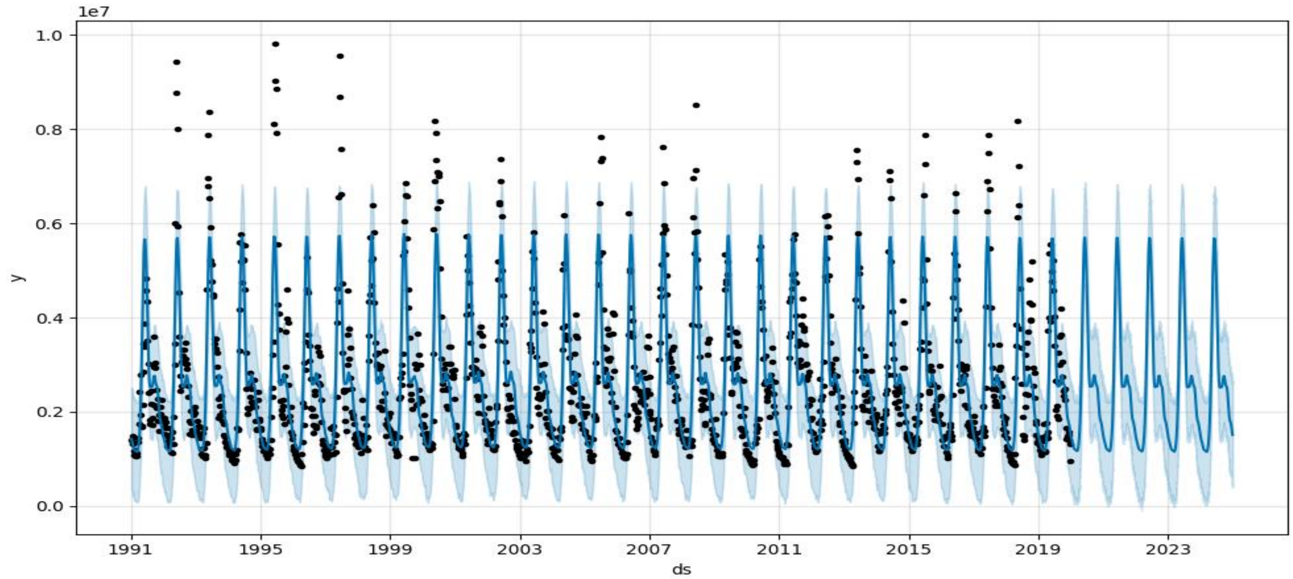


Figure 3.5: forecast data.

Figure 2.15 : A plot showing the actual data followed by the predicted values to facilitate visual comparison

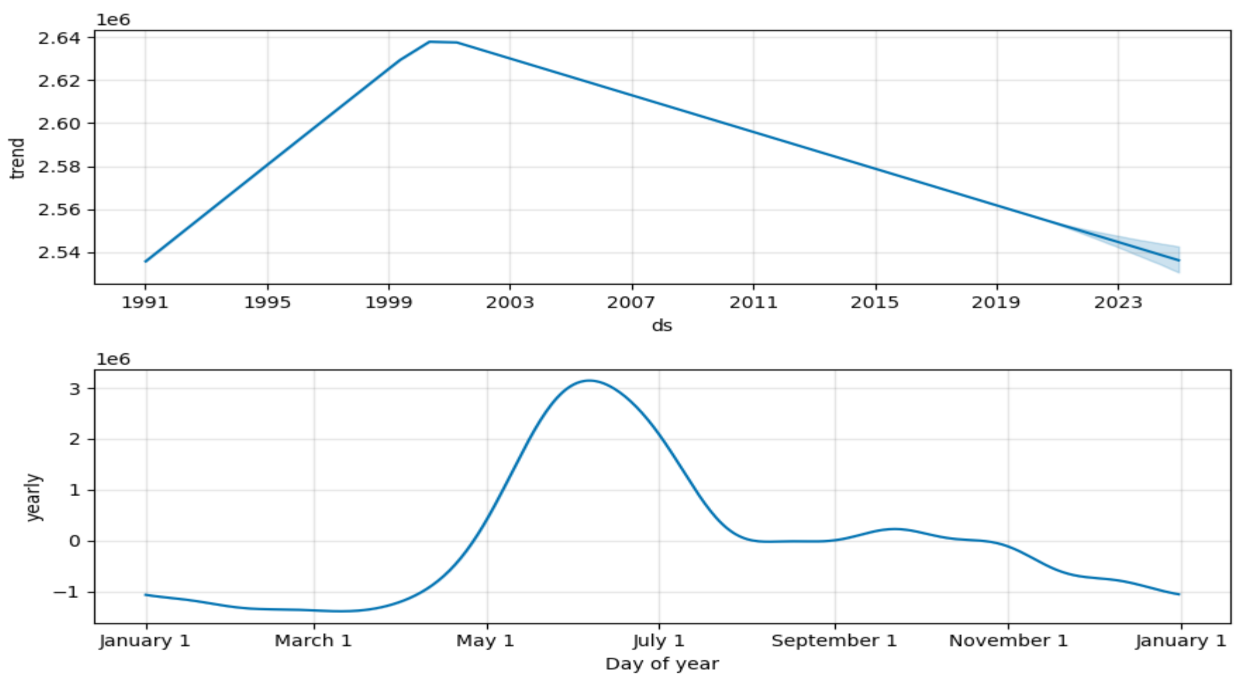


Figure 3.6: components of forecast

Figure 3.6: Forecast components Which indicate that the average is low, meaning that the amount of water flowing globally decreases each year compared to previous years, which is a dangerous indicator

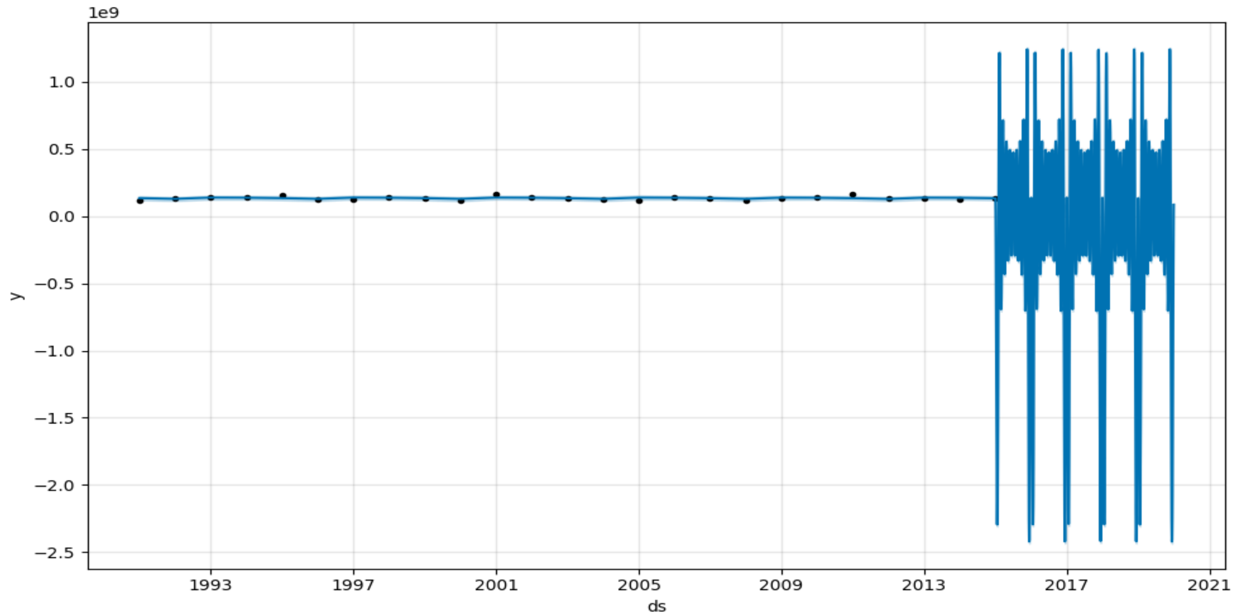


Figure 3.7: evaluation

Figure 3.7: Line plot for evaluating the model to determine its accuracy and measure its efficiency by dividing the data into three parts: train, validate, and test, with the validation process conducted on the last 5 years of actual data.

3.6 Comparisons result

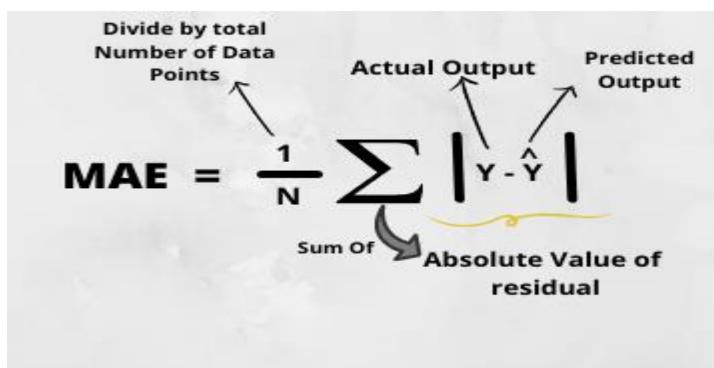
1- Comparison result prediction between ARIMA model and prophet model.

Model performance metric	***ARIMA***	***Prophet***
MAE	***6,013,296.22***	***2,572,479.40***
MSE	***47,502,549,861,569.48***	***14,722,721,937,004.97***
RMSE	***6,892,209.36***	***3,837,019.93***
MAPE	***4.38%***	***1.82%***

*3.7 metrics evaluation:***[7]**

1. MAE (Mean Absolute Error)

The Mean Absolute Error is a measure used to determine the average absolute differences between actual values and predicted values. MAE shows how close predictions are to the actual values without considering the direction of the errors.



2. MSE (Mean Squared Error)

The Mean Squared Error is a measure used to determine the average squared differences between actual values and predicted values. MSE gives greater weight to larger errors compared to MAE.

$$MSE = \frac{1}{n}\sum\underbrace{\left(y-\hat{y}\right)^2}_{\text{The square of the difference between actual and predicted}}$$

3. RMSE (Root Mean Squared Error)

The Root Mean Squared Error is a measure used to determine how close predictions are to the actual values, and it gives significant weight to larger errors but expresses the error in the same units as the variable being studied.

$$\mathbf{RMSE} = \sqrt{\mathbf{MSE}}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j-\hat{y}_j)}$$

4. MAPE (Mean Absolute Percentage Error)

The Mean Absolute Percentage Error is a measure used to determine the average absolute differences between actual values and predicted values as a percentage of the actual values. MAPE gives an idea of the magnitude of the error compared to the size of the data.

$$MAPE = \frac{\overset{\text{Multiplying by 100\% converts to percentage}}{\boxed{100\%}}}{n}\sum\left|\frac{\overbrace{y-\hat{y}}^{\text{The residual}}}{\underset{\text{Each residual is scaled against the actual value}}{\boxed{y}}}\right|$$

Source of data

https://data.jrc.ec.europa.eu/dataset/297b6de3-8c78-4da1-97d4-4c766dc4260d#dataaccess

References

1- Time Series Forecasting Methods

https://www.datacamp.com/tutorial/tutorial-time-series-forecasting

2- El Niño phenomenon:

https://www.noaa.gov/understanding-el-nino

3- time series

https://www.datacamp.com/tutorial/tutorial-time-series-forecasting

4- Statistical Models

https://www.datacamp.com/tutorial/tutorial-time-series-forecasting

5- Autoregressive Model:

https://www.statisticshowto.com/autoregressive-model/#:~:text=The%20AR%20%28p%29%20model%20is%20defined%20by%20the,%2B%20%CF%86%20p%20y%20t-1%20%2B%20A%20t

6- Facebook prophet:

https://facebook.github.io/prophet/

7- metrics evaluation:
https://www.analyticsvidhya.com/blog/2021/05/know-the-best-evaluation-metrics-for-your-regression-model/#h-mean-absolute-error-mae

الملخص

لقد تغير مناخ العالم بشكل كبير في السنوات الأخيرة بسبب التطور الصناعي. يؤثر تغير المناخ بشكل كبير على محطات الطاقة الكهرومائية، ومن الآثار السلبية لذلك هو تقليل توليد الطاقة الكهرومائية. هذا التحدي حاسم نظرًا لأهمية الطاقة الكهرومائية في تلبية الطلب المتزايد على الطاقة. الهدف الرئيسي من هذه الدراسة هو التنبؤ بتوليد الطاقة الكهرومائية في المستقبل (2019-2024) لدولة النرويج. لتحقيق هذا الهدف، من الضروري دراسة معدلات تدفق المياه من السنوات السابقة. لقد حصلنا على مجموعة بيانات شاملة تتضمن بيانات تدفق المياه من 27 دولة أوروبية من عام 1991 إلى عام 2019، مع قراءات أسبوعية. يهدف المشروع إلى التنبؤ بتدفق المياه وتوليد الطاقة الكهرومائية للسنوات الخمس التاليه باستخدام أدوات الذكاء الاصطناعي (AI) والتعلم الآلي .(ML) ستمكننا هذه التقنيات (تحليل السلاسل الزمنيه – التنبؤ بالسلاسل الزمنيه) من تقديم توصيات مدروسة لتحسين إدارة الموارد المائية وتحسين استخدام الطاقة الكهرومائية.