10 Academy: Artificial Intelligence Mastery

Week 10 Challenge Document

Date: 19 Feb - 25 Feb 2025

# Change point analysis and statistical modelling of time series data - detecting changes and associating causes on time series data

**Business objective**

The main goal of this analysis is to study how important events affect Brent oil prices. This will focus on finding out how changes in oil prices are linked to big events like political decisions, conflicts in oil-producing regions, global economic sanctions, and changes in Organization of the Petroleum Exporting Countries (OPEC) policies. The aim is to provide clear insights that can help investors, analysts, and policymakers understand and react to these price changes better.

## Situational Overview (Business Need)

You are a data scientist at **Birhan Energies,** a leading consultancy firm specialising in providing data-driven insights and strategic advice to stakeholders in the energy sector. With a mission to help clients navigate the complexities of the global energy market, Birhan Energies focuses on delivering actionable intelligence that supports decision-making processes for investors, policymakers, and energy companies.

You are tasked with analyzing how big political and economic events affect Brent oil prices. Understand how political decisions, conflicts in oil-producing areas, international sanctions, and OPEC policy changes affect the market.

The oil market is very unstable. This makes it hard for investors to make good decisions, manage risks, and maximize returns. Policymakers need detailed analysis to create strategies for economic stability and energy security. Energy companies need accurate price forecasts to plan operations, control costs, and secure supply chains.

As a data analyst at **Birhan Energies**, you are tasked with:

1. Finding key events that have significantly impacted Brent oil prices over the past decade.
2. Measuring how much these events affect price changes.
3. Providing clear, data-driven insights to guide investment strategies, policy development, and operational planning.

## **Data**

The [data](https://drive.google.com/file/d/1gK9oGSPnpWDFO_belCbbPw3M6BVueHoP/view?usp=drive_link) set contains historical Brent oil prices. It includes daily prices from May 20, 1987, to September 30, 2022.

**Data fields**

* **Date:** Represents the date of the recorded Brent oil price. Each entry is formatted as ‘day-month-year’ (e.g., 20-May-87). The dataset covers daily prices from May 20, 1987, to September 30, 2022.
* **Price**: This column represents the price of Brent oil on the corresponding date. The price is recorded in USD per barrel.

## 

## **Learning Outcomes:**

Skills:

* Statistical Modelling
* Using PyMC3 - a standard Bayesian modelling package in Python
* Running and modifying a moderately large Python package

Knowledge:

* Probability distributions and choosing the relevant one for a given task
* Bayesian inference
* Monte Carlo Markov Chain
* Model comparison
* Policy analysis

Communication:

* Reporting to government bodies

## **Team**

Tutors:

* Mahlet
* Rediet
* Kerod
* Elias
* Emitinan
* Rehmet

## **Key Dates**

* Discussion on the case - Wednesday 19 Feb 2025. Use #all-week10 to pre-ask questions.
* Interim Solution - 20:00 UTC on Friday 21 Feb 2025.
* Final Submission - 20:00 UTC on Tuesday 25 Feb 2025

# Instructions

## Objectives:

The global (business) objective is divided into the following sub-objectives

* Defining the data analysis workflow
* Understanding the model and data
* Extracting statistically valid insights in relation to the business objective

# Task 1

Task 1 focuses on defining the data analysis workflow and understanding the model and data. This involves planning the analysis steps needed to achieve the project’s objective and ensuring a clear understanding of the key concepts related to the project.

1. Defining the Data Analysis Workflow:
   1. Clearly outline the steps and processes involved in analyzing the Brent oil prices data.
   2. Ensure a comprehensive understanding of how the data is generated, sampled, and compiled.
   3. Understand the model inputs, parameters, and outputs.
   4. Identify and state any assumptions and limitations of the analysis.
   5. Determine the main media channels and formats for communicating results to stakeholders.
2. Understanding the Model and Data:
   1. Read the main references related to the project to grasp the key concepts and models being used.
   2. Familiarize yourself with models suitable for time series analysis, such as ARIMA or GARCH, and how they relate to the Brent oil prices data.
   3. Explain the purpose and application of these models in the context of analyzing price fluctuations.
   4. Identify the processes that generate the data and how these are modelled by the chosen time series models.
   5. Describe the expected outputs and limitations of the analysis, and how predictions are made.

## 

## Task 2

1. Adapt the Knowledge from Task 1.1 to Analyze Brent Oil Prices:
   1. Build on the foundational understanding of time series analysis developed in Task 1.1.
   2. Apply this knowledge to analyze the historical Brent oil prices data.
2. Utilize Additional Statistical and Econometric Models as Needed to Refine the Analysis:
   1. Explore advanced time series models such as VAR (Vector Autoregression) for multivariate time series analysis.
   2. Consider regime-switching models like Markov-Switching ARIMA to capture different market conditions.
   3. Implement machine learning models like LSTM (Long Short-Term Memory) networks for capturing complex patterns and dependencies in the data.
3. Explore Other Potential Factors Influencing Oil Prices:
   1. Economic Indicators:
      1. GDP (Gross Domestic Product): Analyze the correlation between GDP growth rates of major economies and oil prices.
      2. Inflation Rates: Examine how inflation in key economies impacts oil demand and prices.
      3. Unemployment Rates: Investigate the relationship between unemployment rates and oil consumption patterns.
      4. Exchange Rates: Assess the effect of currency fluctuations, especially the USD, on oil prices.
   2. Technological Changes:
      1. Advancements in Extraction Technologies: Study the impact of technologies like hydraulic fracturing (fracking) and deep-sea drilling on oil supply.
      2. Renewable Energy Developments: Analyze how growth in renewable energy sources affects oil demand and prices.
      3. Efficiency Improvements: Evaluate how improvements in fuel efficiency and alternative energy usage influence oil consumption.
   3. Political and Regulatory Factors:
      1. Environmental Regulations: Investigate how stricter environmental regulations and carbon pricing affect oil production and prices.
      2. Trade Policies: Study the impact of trade agreements, tariffs, and embargoes on oil markets.

Adapting the Model to a New Scenario:

1. Apply the Analysis Framework to Different Scenarios or Related Datasets:
   1. Extend the analysis to other commodities or related markets, such as natural gas or coal.
   2. Compare and contrast the factors influencing different energy markets.
2. Extend the Model to Incorporate New Variables or Data Sources:
   1. Integrate additional data sources such as economic reports, technological advancements, and regulatory changes.
   2. Include macroeconomic variables and indices to enhance the model’s predictive power.
3. Validate the Model’s Performance in Predicting Future Price Movements and Other Outcomes:
   1. Backtest the model using historical data to assess its accuracy and robustness.
   2. Conduct out-of-sample testing to validate the model’s predictive capabilities.
   3. Use cross-validation techniques to ensure the model’s generalizability.

Suggested Approach:

1. Data Collection:
   1. Gather comprehensive datasets on economic indicators, technological changes, and political factors.
   2. Utilise reliable sources such as the World Bank, IMF, IEA, and industry reports.
2. Data Preprocessing:
   1. Clean and preprocess the data to ensure consistency and accuracy.
   2. Handle missing values, outliers, and anomalies appropriately.
3. Exploratory Data Analysis (EDA):
   1. Perform EDA to identify patterns, trends, and relationships in the data.
   2. Use visualisations to gain insights into the interactions between different variables and oil prices.
4. Model Building:
   1. Develop multiple models to capture different aspects of the data.
   2. Use time series models, econometric models, and machine learning algorithms as appropriate.
5. Model Evaluation:
   1. Evaluate the models using performance metrics such as RMSE, MAE, and R-squared.
   2. Compare the models to identify the best-performing ones.
6. Insight Generation:
   1. Interpret the model outputs to generate actionable insights.
   2. Provide clear recommendations based on the analysis results.

### Task 3: Developing an Interactive Dashboard for Data Analysis Results

Build a Dashboard Application using Flask (backend) and React (frontend) to visualize the results of the analysis, helping stakeholders explore how various events affect Brent oil prices.

* Key Components of the Dashboard:
  + Backend (Flask)**:**
    - Develop APIs to serve data from the analysis results, making it accessible for the React frontend.
    - Handle requests for different datasets, model outputs, and performance metrics.
    - Integrate data sources for real-time updates (optional, if necessary).
  + Frontend (React)**:**
    - Create an intuitive and user-friendly interface to display the analysis results.
    - Design interactive visualizations to show how different events correlate with changes in oil prices.
    - Include features like filters, date ranges, and comparisons, allowing users to explore data around specific events or time periods.
    - Ensure responsiveness for various devices (desktop, tablet, mobile).
    - React tools for charts([Recharts](https://recharts.org/en-US/), [React Chart.js 2](https://www.npmjs.com/package/react-chartjs-2), [D3.js](https://d3js.org/))
* Key Features:
  + Present historical trends, forecasts, and correlations with events (e.g., political decisions, conflicts, economic sanctions).
  + Allow users to see how specific events influenced Brent oil prices, with features like "event highlight" to visualize spikes or drops in prices.
  + Enable users to filter data, select date ranges, and drill down into details for deeper insights.
  + Display key indicators like volatility, average price changes, and model accuracy metrics (e.g., RMSE, MAE).

# Tutorials Schedule

In the following, the colour **purple** indicates morning sessions, and **blue** indicates afternoon sessions.

## Wednesday: Data Analysis Techniques

* Introduction to the challenge (Mahlet)
* Data Science Workflow (Rediet)

## Thursday: Change Point Analysis

* Change Point Analysis (Kerod)

## Friday: Bayesian inference

* Bayesian inference and Monte Carlo Markov Chain (Elias)

## Monday: Working With react

* How to build react dashboards (Emitinan)

# 

# Interim Submission

* Interim report - Covering task - 1
* Link to your GitHub code.

## Feedback

You may not receive detailed comments on your interim submission but will receive a grade.

# Final Submission

* A blog post entry (which you can submit for example to Medium publishing) or a pdf report.
* Link to your Github code, and make sure to screenshots demonstrating anything else you have done.

## Feedback

You will receive comments/feedback in addition to a grade.

# 

# References

**Data Science Workflow**

1. <https://www.datascience-pm.com/data-science-workflow/>
2. <https://towardsdatascience.com/mastering-the-data-science-workflow-2a47d8b613c4>
3. <https://www.knowledgehut.com/blog/data-science/data-science-workflow>
4. <https://medium.com/codex/a-comprehensive-guide-to-master-the-data-science-workflow-739295117d67>

**Change Point Analysis**

1. <https://towardsdatascience.com/change-point-detection-a-bayesian-approach-8eb3cfca4a6e>
2. <https://eecs.wsu.edu/~cook/pubs/kais16.2.pdf>
3. <https://jagota-arun.medium.com/change-point-detection-in-time-series-bcf01409010e>
4. <https://www.iese.fraunhofer.de/blog/change-point-detection/>
5. [Change Point Detection in Time Series](https://www.youtube.com/watch?v=JrOnOcnkR-8)

**Bayesian Change point detection**

1. [Bayesian Changepoint Detection with PyMC3](https://cscherrer.github.io/post/bayesian-changepoint/)
2. [Model Comparison using PyMC3](https://pymc3.readthedocs.io/en/latest/notebooks/model_comparison.html)

**Bayesian inference and Monte Carlo Markov Chain**

1. [As a introduction to Bayesian statistics and the python package (PyMC3)](https://docs.pymc.io/notebooks/api_quickstart.html)
2. <https://warwick.ac.uk/fac/sci/statistics/staff/academic-research/steel/steel_homepage/bayesiantsrev.pdf>
3. <https://www.embecosm.com/2021/12/18/forget-arima-going-bayesian-with-time-series-analysis/>
4. <https://towardsdatascience.com/bayesian-structural-time-series-interruption-method-5018761db92b>
5. <https://machinelearningmastery.com/markov-chain-monte-carlo-for-probability/>
6. <https://towardsdatascience.com/monte-carlo-markov-chain-mcmc-explained-94e3a6c8de11>

**Alternative Data Source**

1. [Petroleum Trade: Imports From OPEC Countries](https://www.eia.gov/totalenergy/data/browser/index.php?tbl=T03.03C#/?f=M)
2. [Energy Consumption Transportation Sector](https://www.eia.gov/totalenergy/data/browser/index.php?tbl=T02.01B#/?f=M)

**React dashboard templates**

1. <https://github.com/flatlogic/react-dashboard>
2. <https://github.com/creativetimofficial/light-bootstrap-dashboard-react?tab=readme-ov-file>
3. <https://github.com/flatlogic/react-dashboard>
4. <https://github.com/marmelab/react-admin>
5. <https://github.com/coreui/coreui-free-react-admin-template>