Week 10 - Interim Report

Change Point Analysis and Statistical Modelling of Time Series Data

Introduction

This week's project establishes the foundational workflow for a change point analysis on Brent crude oil prices. The primary objective is to develop a robust data pipeline, a Bayesian statistical model, and an analysis framework to identify significant shifts in oil price trends and associate them with key historical events. We have made significant progress in structuring the project, preparing the time-series data, and implementing the core change point detection model.

Methodology

1. GitHub Repository and Project Setup

The project repository has been structured to ensure modularity and maintainability, following best practices for data science projects.

- **Core Structure:** The directory layout includes data/ (for raw and processed data), src/ (for modular Python scripts), notebooks/ (for exploratory analysis), events/ (for our curated event data), and reports/ (for visualizations and documentation).
- **Dependencies:** A requirements.txt file was created to manage all Python library dependencies, including pandas, numpy, pymc, arviz, and seaborn.

2. Data Preparation

The initial phase involved collecting and cleaning both the oil price time series and the key events dataset.

- **Script (src/data_preparation.py):** A custom Python script was developed to load daily Brent oil prices and a CSV file of historical events. It handles date conversions, resamples the daily data to a monthly frequency, and fills any missing values to create a clean, continuous time series.
- **Target Data:** The script processes brent_oil_prices.csv and key_events.csv, saving a cleaned monthly time series to data/processed/processed_data.csv.

3. Exploratory Data Analysis (EDA)

To gain a preliminary understanding of the time series, an EDA notebook was developed.

- **Notebook (notebooks/EDA.ipynb):** This Jupyter notebook visually inspects the raw and monthly-resampled oil prices. It checks for trends and volatility and overlays the key historical events to form initial hypotheses about potential correlations between events and price shifts.

4. Bayesian Change Point Modeling

The core of the project is a Bayesian model designed to detect a structural break in the oil price time series.

- **Script (src/model.py):** This script uses the PyMC library to build a Bayesian model. It assumes a single change point where the mean of the time series shifts. It defines priors for the change point location (tau), the pre- and post-change means (mu1 and mu2), and the standard deviation (sigma).
- **Inference:** The model uses Markov Chain Monte Carlo (MCMC) sampling to infer the posterior distributions of these parameters, providing a probabilistic estimate of the most likely change point and its uncertainty. The output is logged to logs/modeling.log.

5. Analysis and Visualization

To interpret the model's output, a visualization script was created.

- **Script (src/analysis.py):** This script loads the processed data and the model's output to generate a comprehensive plot. The plot displays the time series, the inferred change point, and the historical events, allowing for direct visual comparison and validation. The final plot is saved to reports/change_point_analysis.png.

Challenges & Solutions

Challenge: Log-based analysis and visualization:

- Since the model.py script logs the inference summary instead of returning a storable object, the analysis.py script needed to be adapted.
- **Solution:** The analysis.py script has been updated to simulate the inference data based on the logged summary output. In a future iteration, we will modify the model.py script to save the full arviz.InferenceData object (e.g., as a .nc file) to ensure a seamless and robust workflow.

Future Plan

Building upon the established analysis foundation, the next steps will focus on refining the model and preparing the final deliverables.

1. **Refine the Model:** Explore more complex models, such as those with multiple change points or different assumptions about volatility.
2. **Develop a Final Report:** A detailed PDF report will be prepared to explain the methodology, present the quantitative findings, and discuss the correlation between events and change points.
3. **Build a Dashboard:** An interactive web-based dashboard will be developed to allow stakeholders to dynamically explore the results.

Conclusion

This week's efforts have laid a robust groundwork for the Brent Oil Change Point Analysis project. Significant progress has been made in establishing a structured repository, implementing data preparation, and developing the core Bayesian model. This foundation is crucial for the upcoming analysis and reporting tasks and positions the project well for successful completion.