Integrated Dashboards with Predictive Analytics

Oil Arbitrage and Price Forecasting

Yiwen (Elaine) Xu, Chien An Liu, Christian L Rodriguez, Kevin Zhu, Stephen Kim, Son Tran

Introduction/Motivation

This project aims to address a significant gap in the oil and gas trading space by consolidating global supply chain data into a single, user-friendly application that can support oil arbitrage decisions. Oil arbitrage is currently conducted through three main strategies: supply and demand analysis to take advantage of price differentials, price signals to take advantage of market patterns, and market sentiment to anticipate price movement. Current oil arbitrage practices are limited by the industry's resistance to adopt data analytic tools despite an overwhelming amount of available data and a reliance on disparate data sources which require manual effort to view together.

To overcome these challenges, the dashboard created merged data from relevant sources and the analytics based on the BEAST model gave more insights and better forecasts. It offers traders a centralized platform with integrated analysis tools for better decision-making, which is currently missing in the public domain. It will also save time and improve operational efficiency for commodities traders. All above makes this project important to conduct.

Our Approach

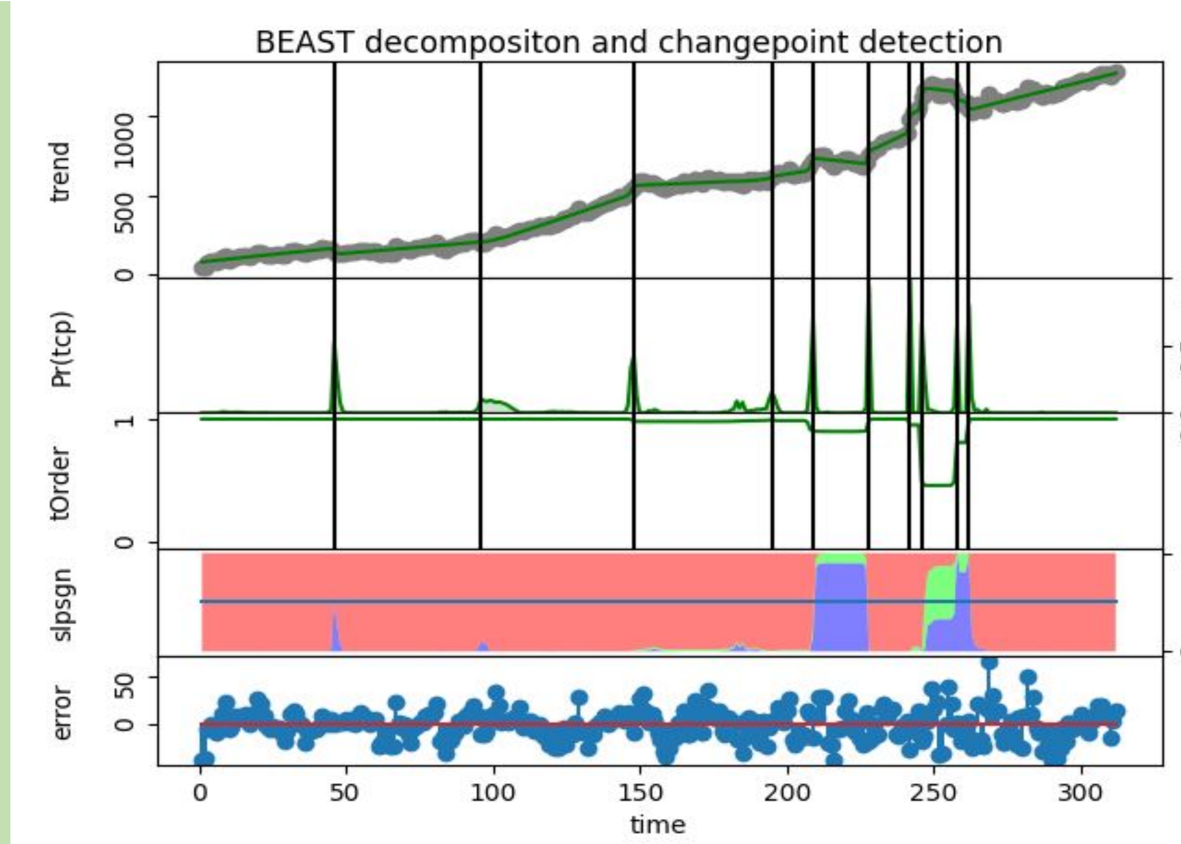
What are the team's approaches and how do they work?

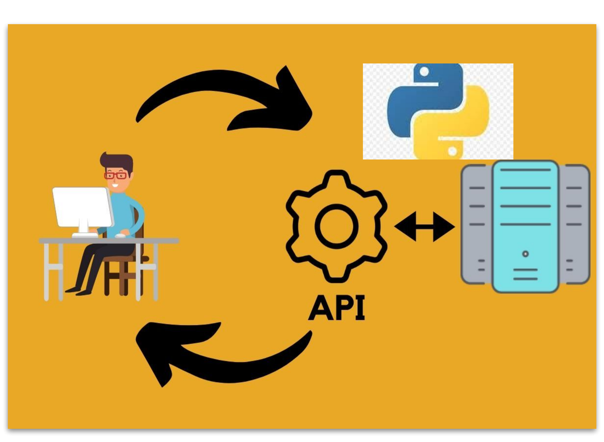
The team settled on a Bayesian change-point detection and time series decomposition (BEAST) model to detect trend, noise, seasonality, and critically identify stepwise changes in these features.

Complementing this, we utilize the ARIMA model, renowned for its excellence in short-term forecasting (Kim, 2018). For our forecasting efforts, we split the available data into a 70/30 ratio for training and testing. We explored two methodologies: a basic ARIMA model with predefined parameters (1, 1, 1) for autoregressors, differencing, and moving averages, and the auto-ARIMA model from the pmdarima library, which optimizes these parameters automatically.

Why do they solve the problem and what is new?

Our approach empowers users to tailor their analytical processes to align with their unique requirements, thanks to a modular architecture that supports the integration of custom models and visualization techniques. This flexibility encourages innovation as users experiment with new methodologies, while centralized data management ensures easy access to data from a single, unified source. Moreover, the user-friendly interface provided by Power BI allows individuals of all skill levels to effortlessly navigate and explore data. Our dashboard, featuring the BEAST and ARIMA models as a proof of concept, offers significant advantages over traditional industry standards. It not only provides customizable analytical capabilities but also ensures seamless data management, empowering users to meet their specific needs more effectively.





Data

How data was collected?

Data was collected through Energy Aspects API with the team developing a custom get_metadata method to efficiently extract thousands of data series in a structured and convenient way.

What are the data characteristics?

The dataset covers 93 countries across seven categories, including crude oil, oil products, NGLs, liquids, condensate, biofuels, and power. It spans from 01/01/2000 to the present, with data available in daily, weekly, or monthly frequencies. Featuring millions of structured time-series (temporal) points with actual, forecasted values or both, the dataset ensures integrity through rigorous processing by the team, replacing invalid points with NaN and filling gaps using forward and backward methods.

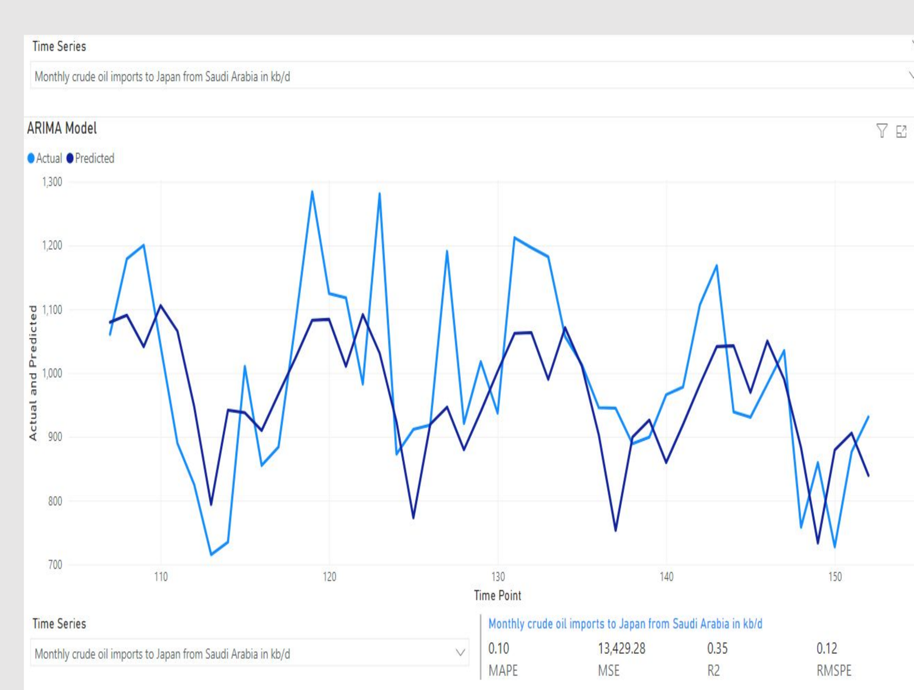
Experiments and Results

How was the project evaluated: The BEAST and ARIMA forecasts are evaluated using a range of measures of prediction accuracy including R-Squared, Mean Absolute Percentage Error (MAPE), Root Mean Squared Percentage Error (RMSPE) and Mean Squared Error (MSE), as well as, visual comparison of actual vs forecasted values through dashboards. The evaluation via accuracy measures is presented by histograms across product categories providing an overall assessment of the performance. Computational power requirements for different methods are also assessed.

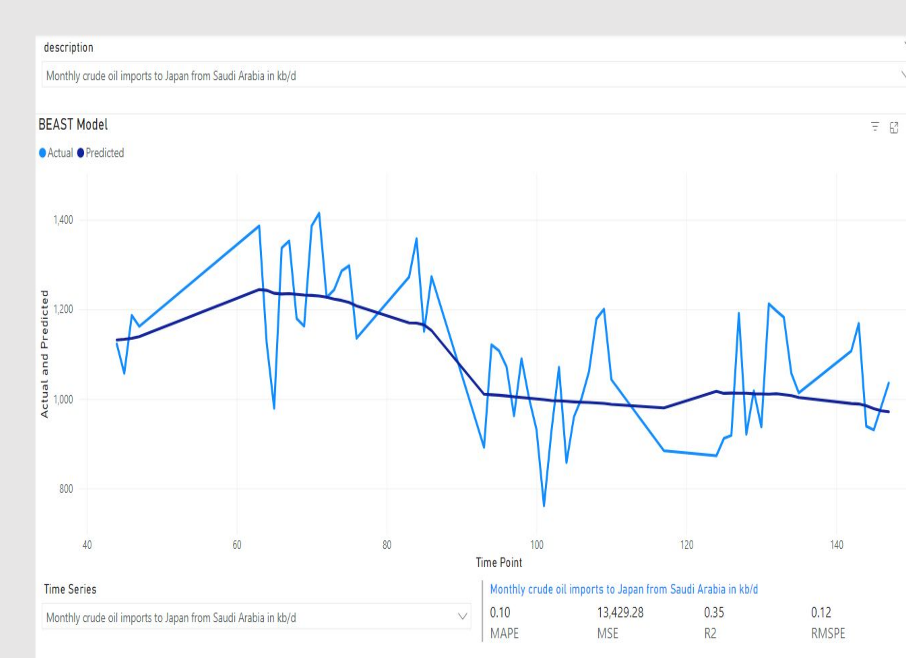
What are the results and how do the methods compared to other methods: BEAST consistently achieves excellent back-test accuracy with approx. 70% of crude oil data series, for instance, showing R-Squared above 0.6 and MAPE and RMSPE mostly between 0 and 0.2, indicating a very low forecast error. BEAST balances computational efficiency with strong prediction performance.

Auto - ARIMA performed with 70/30 data splits displays a less favourable results than expected compared to BEAST models with significantly lower R-Squared and MAPE and RMSPE mostly between 0 and 0.4, from the category of crude oil for example and even lower R-Squared for other categories. In the meantime, the required computation capacity is more demanding.

ARIMA on a rolling basis shows a strongly high accuracy with an error ratio of almost zero which indicates a highly overfitting model and is not a good comparable to BEAST considering BEAST forecasting mostly with data splits rather on a rolling basis. Also it requires a high computation capacity since it requires 100x more computational capacity than the data-splitting ARIMA.



ARIMA



BEAST