Analyzing Trends Between Crude Oil Prices and Airline Performance Data

Aditya Kalyan Jayanti (aj8582@cs.rit.edu)
Advisor: Dr. Carol Romanowski (cjrcms@rit.edu)

Computer Science

Abstract

- Airline industry is a competitive market with a small profit margin. In the past year, twenty-four airlines worldwide have ceased operations.
- **Higher oil prices**, **frequent cancellations**, **delays** have an adverse impact on airline profitability.
- The goal is to analyze market trends and develop an early warning **prototype** system for airlines.
- By analyzing trend of crude oil prices and airline performance using forecasting algorithms and natural language processing.

Introduction

- Fuel constitutes 20-30% of an airline operating costs.
- $1 increase in crude oil result in 1% increase in fuel costs for an airline, reducing profits.
- Flight delays cost the airline industry $22 billion each year, reduced the US GDP by $4 billion in 2007.
- Forecasting trends of crude oil and flight delays provide insights to sustain business through extreme volatility.

Data Collection & Preparation

- **Web scraping news articles** related to “crude” oil news. Source: “oil news articles” section on oilprice.com.
- **Crude oil price data:** open, high, low, and volume. Source: FRED Economic Data, WTI crude oil, Oklahoma.
- **Airline performance data:** 5M rows, 43 columns Source: Bureau of Transportation Statistics (BTS).

Historical Trend of Crude Oil Prices

- Observed Duration: 2012 – 2016
- Reason: 53.97% percentage change during this period.
- Fig.1 represents a graph on analyzing historical crude oil price trend; Followed by extracting news articles.

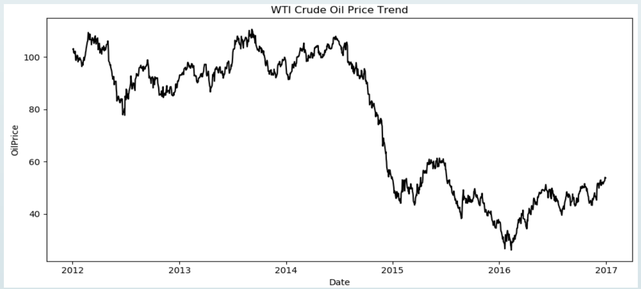

Fig 1. Historical Trend of crude oil prices.

Methodology

- Analyzed crude oil price trend using sentiment of unstructured data obtained through web scraping.
- Set of pretrained models, works well with social media data, reduces complexity to **linear time**.
- **Data cleaning process:** eliminating URLs, invalid string data, ordering news articles by timestamp.
- Computed compounded sentiment value using **VADER**, grouped by mean over observed duration.
- Fig.2 shows trend of news sentiment, downward trend closely related to plunge observed in Fig. 1.
  - Positive Sentiment: Rise in crude oil prices.
  - Negative Sentiment: Fall in crude oil prices.
- Drop in fuel prices, boosted airlines earnings and increased profits.

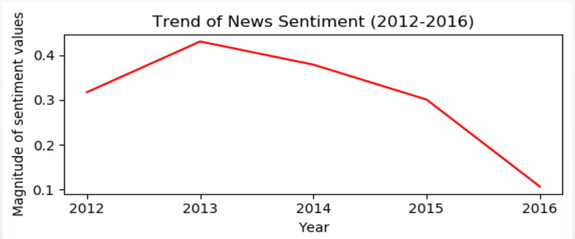

Fig 2. Mean of sentiment values over news articles.

- **Facebook Prophet API**: Seasonal forecasting on time series data, works well for non-linear trends.
- **MAE** on average 2.62 for first 41 days of forecast.
- Fig. 3 consists of original data (black), oil forecast model (blue), change points (vertical dashed red), confidence interval (light blue region), trend with seasonality removed (solid red).

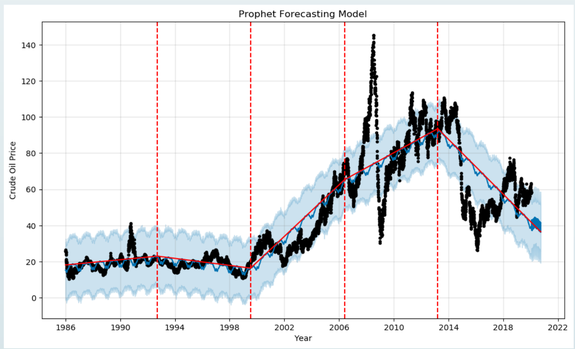

Fig 3. Forecasting using Prophet.

- **Forecasting** assists with data-driven decisions, including whether to **hedge fuel** prices or not.
- High inefficiency in the transport sector, increases cost of associated businesses.
- **Aircraft delays** are classified as follows: 61% (air carrier, late arrival), 32% (National Airspace System), 6% (weather conditions), 1% (security).

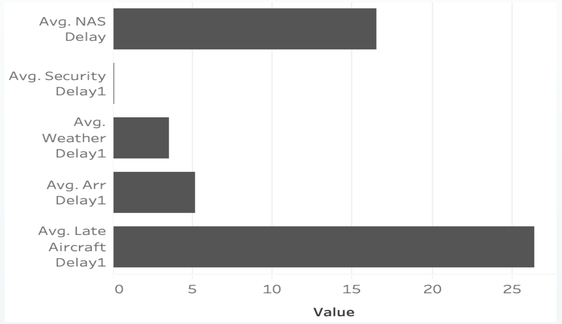

Fig 4. Analyzing the cause of aircraft delays.

- Delays are consistently attributed to late arrival due to carrier and the aviation infrastructure.
- By 2025, US airline traffic increases by 30%.
- Without upgrades to NAS, delays are expected to increase. Reducing profits substantially.
- Fig. 5 shows average delays by month in 2019.

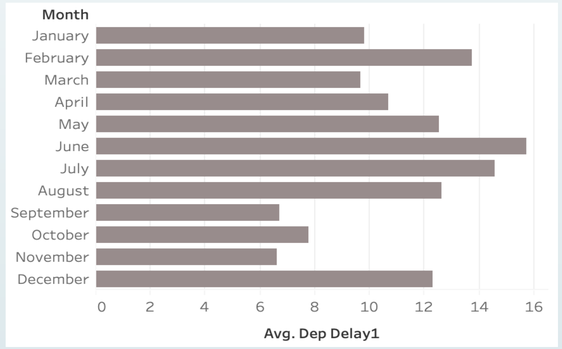

Fig 5. Average delays by month.

- Most delays occur during the summer months.
- Crude oil prices are highest during these months.
- Reduced airline profits during peak travel season
- Airlines require 75% occupancy to break even.

Results

- **Predictive analysis** retrieved from this project can contribute in the form of a prototype, that helps to identify operational variables that impact **profitability** of airlines.
- Achieved a **Mean Absolute Error** (MAE) of 2.599.
- Fig.6 describes the performance of forecasting as a function of time. Forecast error increases with time.

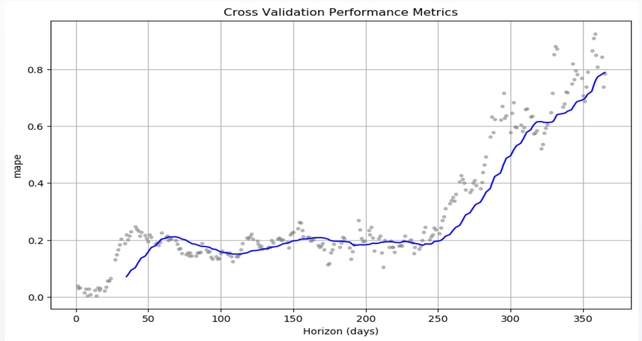

Fig 6. Performance of forecasting using MAPE

Limitations

- This model has a relatively high error rate for forecasts more than 50 days, uses news article from one website.
- High error rate for forecasts during unusual events.
- The possible solution is to extract features and analyze sentiment of news articles using a CNN classifier.

Future Work

- Develop a data pipeline to periodically scrape updated data (oil prices, airline performance, news articles).
- Potentially explore the possibility of chunking data from a large file for multiprocessing.
- Feature grouping based on Latent Dirichlet Allocation (LDA) for distinguishing effects from various news topics.

References

[1] X. Li, W. Shang, and S. Wang. Text-based crude oil price forecasting: A deep learning approach. International Journal of Forecasting, 35(4):1548–1560, 2019.
[2] M. Oussalah and A. Zaidi. Forecasting weekly crude oil using twitter sentiment of u.s. foreign policy and oil companies data. In 2018 IEEE International Conference on Information Reuse and Integration (IRI).