

Forecasting Weekly Crude Oil Using Twitter Sentiment of US Foreign Policy and Oil Companies Data

A. Zaidi

Natural Language Processing
Group, University of
Cambridge, Cambridge
CB3 0FD, UK
Ahmed.Zaidi@cl.cam.ac.uk

M. Oussalah

Centre for Ubiquitous Computing,
Faculty of Information technology,
University of Oulu,
Oulu, Finland.
Mourad.Oussalah@oulu.fi

Abstract - *The drop in crude oil price during late 2014 has had a significant impact on all nations. While some countries have reaped the benefits of low oil prices, others have suffered greatly. As a result, it is no surprise that many academics have attempted to develop reliable models to forecast crude oil price. In the age of information and social media, the role of Twitter and Facebook has become increasingly more relevant in understanding our environment. Many academics have exploited this wealth of data to extract features including sentiment and word frequency to build reliable forecasting models for financial instruments such as stocks. These methodologies, however, remain unexplored for the prediction of crude oil prices. The purpose of this investigation to develop a novel model that uses sentiment of United States foreign policy and oil companies' to forecast the direction of weekly WTI crude oil prices. The investigation is divided into three parts: 1) a methodology of collecting tweets relevant to US foreign policy and oil companies'; 2) a statistical analysis of the novel features using Granger Causality Test; 3) the development and evaluation of three machine learning classifiers including Naïve Bayes, ANNs, and SVM to predict the direction of weekly WTI crude oil. The findings of the statistical analysis showed strong correlation between the novel inputs and WTI crude oil price. The results of the statistical tests were then used in the development of the predictive model. SVM was found to provide best forecasting performance. Furthermore, using these novel features, the predictive accuracy exceeded that of existing models mentioned in literature.*

Keywords: Text mining, Sentiment analysis, Time series analysis, Prediction.

I INTRODUCTION

The drop in crude oil price during late 2014 has had a significant impact on all nations. While some countries have

reaped the benefits of low oil prices, others have suffered greatly. In light of the importance of crude oil to the global economy, it is not surprising that economists have spent a great deal of resources trying to predict its movements. Forecasting future prices will allow companies¹ mitigate their risks against price fluctuations.

Crude oil prices are impacted by a variety of external and internal factors. Some factors include production rates, foreign sanctions, economic growth and seasonal consumption. Traditionally, economists have focused on historical oil prices, supplementary commodity prices (e.g. coal, natural gas), inventory levels, and financial instruments prices (futures and options) as a method of predicting crude oil prices. Few academics have modelled the price of oil as a function of non-oil variables (e.g. US dollar exchange rate gap). However, there still remains a gap on the impact of the overall sentiment of foreign policy on oil price, specifically US foreign policy. For instance, the US-led sanctions imposed against Iran in mid-2012 reduced their oil production from 2.4 million barrels a day to 1.4 million barrels a day, which trivially impacted the overall oil price. Traditional approaches for oil price prediction can be categorized into **quantitative** and **qualitative** models [1-3]. Quantitative models, which include time-series, financial and structural models of historical data are more suited for short to medium term oil price forecasting. For instance, Pindyck [14] used an autoregressive model to forecast crude oil prices from 1887 to 1996, although with quite a poor accuracy level. Radchenko [15] built on Pindyck's model by adding an OPEC behavior, which slightly increases the

¹ Particularly Airline companies as oil prices represent are large aspect of their total cost

forecasting accuracy. Silva et al. [18] implemented a hidden Markov model (HMM) to forecast medium term crude oil price movements and achieved a mean forecasting accuracy of 57%. Chin et al. [3] examined energy futures prices to accurately forecast future spot prices. The findings suggested that future prices are unbiased predictors of spot prices and outperform time-series models. Qualitative models combine the inputs from quantitative models with additional inputs such as isolated events e.g. natural disasters, political factors (e.g., elections, revolutions) to boost the forecasting accuracy. In this respect, Wang et al. [21] used a novel approach called TEI@I to predict WTI (West Texas Intermediate) crude oil prices. The author investigated the effect of infrequent events extracted using web-text mining, on oil prices. The results indicated that the non-linear TEI@I produced superior results to the linear ARIMA model. Similarly, inventory clues extracted through text mining techniques have been reported by Ye et al. [23-24]. The use of machine learning like approach has been reported in [22, 25]. Ghaffari and Zare [6] used historical oil spot prices and Adaptive Network-based Fuzzy Inference System (ANFIS) to forecast the direction of the daily WTI crude oil prices from 5/1/2007 to 5/31/2007, achieving a 68% accuracy rate. Genetic algorithm based strategy has been advocated in [20]. Nevertheless, despite an increase in accuracy rate in recent achievements, both qualitative and quantitative methods suffer from the lack of long-term prediction capability as well as the inappropriate use of available events such political factors, legislation changes and natural disasters.

In the age of information and social media, websites such as Facebook and Twitter contain a plethora of knowledge that can be used to understand market behaviour and extract sentiment of individuals, companies and even countries. Studies have shown that Twitter sentiment of companies can be used as inputs for predicting future stock prices. For instance, Zhang et al. [26] used sentiment of tweets to identify the directional shift in Dow Jones, S&P 500, and NASDAQ. Gilbert and Karahalios [7] used emotions to predict the stock market movement. See also discussions in [4] on generic Twitter based sentiment analysis. However, such an approach, to the best of our knowledge has not been successfully applied to crude oil price prediction. Therefore, this study aims to fill the gap in this domain by applying social media forecasting methods with inputs from US foreign policy (USFP) and oil companies (OC)' sentiment to forecast the direction of crude oil prices. In other words, the inputs of our developed model (s) are based on previously

unexplored but theoretically sound factors that impact crude oil: political factors (US foreign policy), oil company behaviour, and OPEC behaviour². The investigation will use social media as a primary source of data for the extraction of these variables. As literature shows that sentiment has been used to successfully forecast the price of financial instruments, the investigation will use sentiment as the key indicator of US foreign policy and oil company behaviour. Especially, the study aims to: i) Develop and evaluate a novel methodology and tool to extract and store US foreign policy and oil companies' sentiment from Twitter into a relational database; ii) Conduct a statistical causation and correlation study between US foreign policy and oil companies' sentiment with weekly WTI crude oil prices; iii) Build and evaluate a predictive model to forecast weekly WTI crude oil prices using US foreign policy and oil companies' as an input; iv) Create a hedging strategy using the predictive model developed to mitigate risk for companies; v) Identify limitations of the investigation and opportunities for future works.

Section 2 of this paper details the methodology employed. Implementation issues are discussed in Section 3, while testing and exemplification are reported in Section 4.

II METHODOLOGY

As already pointed out in the introduction section, Twitter sentiment of **USFP** and **OC** is used as input to forecasting the directional shift of weekly WTI crude oil prices. Figure 1 summarizes the key milestone stages in order to achieve the forecasting task.

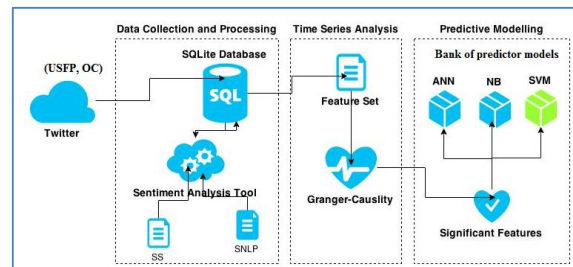


Fig.1. General scheme of the Oil forecasting task

A Data Collection

Three rationale assumptions were hypothesized for collecting data related to USFP and OC:

- Only tweets from credible or influential Twitter accounts are deemed as relevant information when forecasting crude

² The lack of OPEC behaviour consideration was identified as a detriment in ARIMA models by Radchenko [13]

oil. This claim suggests that tweets about USFP or OC from unreliable and non-influential users on Twitter have no significant impact on the WTI crude oil prices.

- USFP data on Twitter is represented by the tweets made by US foreign policy and strategy think tanks' Twitter accounts. This claim suggests that US foreign policy and strategy think tank tweets are a reliable source and representative of USFP data
- OC data on Twitter is represented by the tweets made by the largest oil companies' and associations' Twitter accounts, suggesting that such organizations are most reliable source and representative of OC data.

Therefore, for USEP data, a list of all the foreign policy and strategy think tanks based in the United States was acquired (EIA.gov). For each member of the list, a Twitter account was searched. In the case where there was no Twitter account or its account is inactive, the member was removed from the list. Otherwise, the username of the member was recorded. This process was repeated for the first 100 think tanks. The final list was comprised of 76 think tanks that were actively operating Twitter accounts. On the other hand, for OC dataset, the relevant oil companies and associations, a list of the 25 biggest oil companies by revenue was retrieved from Forbes. This, along with the rankings from Platts 250 was used to produce a list of 52 oil companies and associations that were operating active Twitter accounts. Additionally, well known energy administrations such as the United States Energy Information Administration (USEIA) were also included in the list. Besides, data was also collected for Justin Bieber's username that will be used as a control variable later on. Once the list of usernames was finalized, a SQLite database was setup to store the relevant tweets that are collected. Next, the Twitter Search API³ was employed to retrieve the Tweets from the identified entities. The overall process of data acquisition is shown in Figure 2.

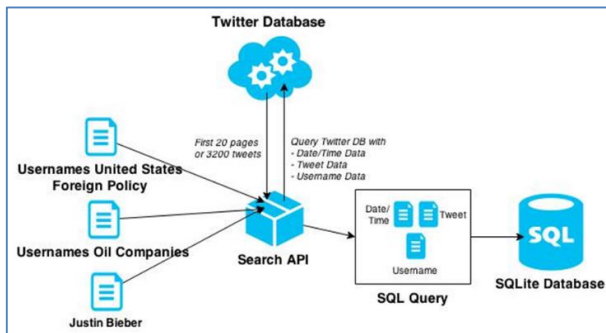


Fig. 2. Overview of Data collection task

Especially, a user_timeline approach was adopted to retrieve tweets where the first 20 pages or the 3,200 most recent tweets (whichever comes first) for each username are identified. After the tweets collection was completed, the data in the SQLite database was standardized for compatibility purposes. This includes formatting the dates, removing duplicate tweets, URL links, uncommon characters, among others.

B. Sentiment Analysis

The tweets contained in the SQLite database were analysed for sentiment strength. This process is completed using two tools: SentiStrength (lexicon-based) (SS)⁴ and Stanford NLP Sentiment (SNLPS) analyser⁵ (machine learning-based, see [19]). Both tools have publicly available Java libraries. Especially, in order to make the processes of sentiment analysis more efficient, a Java tool was created (See Figure 3 for the block diagram). The tool retrieves the tweets from the SQLite database and determines the sentiment strength of each tweet using both SNLPS and SS analysers. The output from both analysers is then incorporated into a SQL query. The SQL query executes and updates the database with the sentiment for each tweet. Strictly speaking, the SS sentiment outputs, for each phrase, are two integers (positive and negative strengths) ranging from 1 (no sentiment) to 5 (high sentiment). We therefore consider the difference between positive and negative sentiment as the overall score for the corresponding phrase. On the other hand, the NSLPS carries out extensive pre-processing tasks including lower-cased, stop-word removal, HTML tags, and non-English removal.

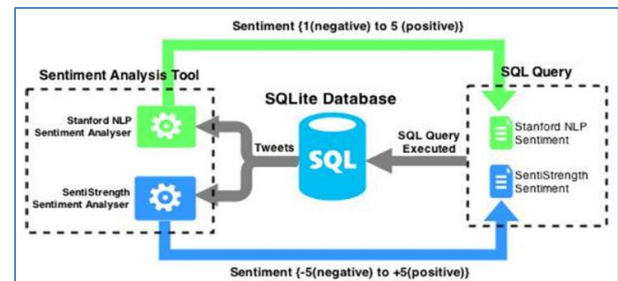


Fig. 3. Sentiment analysis block diagram description

Literature review suggested that OPEC behavior (e.g., [12]) is a key impacting factor of WTI crude oil price. One novel method of quantifying OPEC behaviour is by monitoring the frequency of references to OPEC members in USFP tweets. For the purpose of our study, features of interest include date, frequency of "oil" (including its various related terms), OPEC members, sentiment score, and full tweet message.

³ <https://dev.twitter.com/rest/public/search>

⁴ <http://senticstrength.wlv.ac.uk/>

⁵ <https://nlp.stanford.edu/sentiment/>

On the other hand, the WTI oil price were obtained from the United State Energy Information. The motivation for use of the frequency of references to “oil” is due to alternative studies where the references to a particular word can indicate future values of certain products or stocks (e.g., [4], [26]).

C Feature Selection

Since prior studies have shown that certain word frequencies correlate with the rise and fall of financial instrument prices, this justifies the decision to investigate the correlation between the frequency of “oil” and WTI crude oil price. Using these novel inputs, a comprehensive time-series analysis was conducted. More formally, the question of interest is to determine whether the sentiment of USFP and OC and frequency of references to “oil” and OPEC members in USFP tweets provide *statistically* significant information to forecast weekly WTI crude oil prices. By doing so, we can determine the optimal inputs that can be used to build the predictive model. The study uses weekly crude oil prices as opposed to daily as long term horizon forecast is more useful for regulators than short term. Given that the data set is limited to approximately 4 years, monthly prices would reduce our sample size significantly, thereby, potentially preventing us from obtaining accurate and reliable results. For this purpose, a statistical based approach is designed. Four main null hypotheses were tested.

- USFP sentiment is *not* statistically significant in forecasting weekly WTI crude oil prices
- OC sentiment is *not* statistically significant in forecasting weekly WTI crude oil prices
- Frequency of references to “oil” is *not* statistically useful in forecasting weekly WTI crude oil prices
- Frequency of references to OPEC members is *not* statistically useful in forecasting weekly WTI crude oil prices

Justin Bieber’s sentiment is *not* statically significant in forecasting weekly WTI crude oil prices. **Control Variable** – we know for certain the Bieber’s sentiment should not have any impact on WTI crude oil prices. Therefore the robustness of our methodology can be tested by using this variable.

In order to determine the significance of each feature, we used the Granger-Causality test [9]. The latter is a statistical hypothesis test that uses a bivariate linear regression method to model a stochastic process. The test uses two time-series, X and Y . Time series X is said to “Granger-cause” time series Y if time-series Y can be better predicted using historical values of both time series X and Y as opposed to the historical values of Y in isolation.

The Granger-Causality Test [9] can be defined by the following equation:

$$Y_t = \alpha_1 + \sum_{i=1}^m \beta_{t-i}^1 Y_{t-i} + \sum_{i=1}^n \beta_{t-i}^2 X_{t-i} + \varepsilon_t$$

$$X_t = \alpha_2 + \sum_{i=1}^n \beta_{t-i}^3 X_{t-i} + \sum_{i=1}^m \beta_{t-i}^4 Y_{t-i} + \gamma_t$$

The variable n and m represent the maximum number of lags to be considered in the model for inputs X and Y , respectively. The matrix β represents the coefficients of the model. ε_t and γ_t are two uncorrelated white noise series, representing the residual values of the time series (prediction error). An f-test is used on the statistical model to determine whether or not time-series X “Granger-causes” times series Y . If the **f-statistic** outputted from the test is greater than the **critical value** then the **null hypothesis** can be rejected. If the null hypothesis is rejected then the underlying feature is deemed statistically significant, and, thereby, will be incorporated into the predictive model. Table 1 lists the features considered in this test. The results of the Granger-Causality Test have been recorded in Table 2. We only present the results pertaining to some attributes because of page restriction. The graphical representations of each of some attributes mentioned in Table 1 can be seen in Figure 4. The table reports evaluation of *lag*, *f-statistic*, and *critical value*. The lag is the number of weeks that the feature takes to have an impact on the price of weekly WTI crude oil. The f-statistic is the output of the f-test. If the value of the f-statistic is larger than the critical value, then at the specified lag, the null hypothesis can be rejected⁶.

Table 1. Summary of features

Feature	Contents	Value size and type
usfp_senti	United States foreign policy sentiment (SentiStrength)	218x1 d
usfp_nlp	United States foreign policy sentiment (Stanford NLP)	218x1 d
usfp_oilfreq	frequency of the term “oil” in United States foreign policy tweets	218x1 d
usfp_opecfreq	frequency of references to OPEC members in United States foreign policy tweets	218x1 d
oil_senti	oil companies’ sentiment (SentiStrength)	218x1 d
oil_nlp	oil companies’ sentiment (Stanford NLP)	218x1 d
jb_senti	Justin Bieber sentiment (SentiStrength)	66x1 d
jb_nlp	Justin Bieber sentiment (Stanford NLP)	66x1 d
jb_date	weeks corresponding to the tweets available from Justin Bieber’s twitter account	66x1 cell
jb_oilprice	weekly WTI crude oil prices corresponding to the weeks from jb_date	66x1 d
oil_price	weekly WTI crude from 3/1/2015 to 2/3/2015	218x1 d
dates	start of week dates between 3 rd January 2011 and 2 nd March 2015	218x1 cell

The results of the Granger-Causality Test using USFP sentiment from the SS analyser in Table 2 showed that the f-statistic was greater than the critical value at lags of 1-10. Therefore, we can reject the null hypothesis about USFP

⁶ If the null hypothesis is rejected, then the feature is statically significant and can be used as an input in the predictive model.

sentiment as SS analyser does not provide statically significant information in forecasting weekly WTI crude oil price. The rejection of the null hypothesis indicates that USFP as per SS “Granger-causes” the future value of WTI crude oil and as a result would be an acceptable input for the forecasting model. Table 2 also shows that lag 6 and 7 have the largest f-statistic of 7.29 suggesting that the impact of USEP sentiment likely takes 6 to 7 weeks to work through the weekly price of WTI crude oil. Comparing the results of Table 2 with the graph in Figure 4 we can see a clear visual correlation with the weekly price of WTI crude oil and the USFP sentiment as per SS analyser. The weekly price of WTI crude oil and USFP sentiment remain relatively constant until around October 2014, when the price of crude oil begins to drop along with the sentiment. Towards the middle of February 2015, the sentiment of USFP begins to pick up and so does the price of WTI crude oil. This suggests a positive correlation between the two variables.

Table 2. Granger-Causality Test – statistical significance (f-statistic) at lags of 1 – 10 weeks between weekly WTI crude oil price and USFP as per SS analyser (alpha 0.05; bold = significant; *highest f-statistic)

Lag (n)	f-statistic	critical value
1	3.9817	3.8851
2	4.1303	3.8853
3	4.1303	3.8853
4	4.1303	3.8853
5	4.1303	3.8853
6	7.8631*	2.1422
7	7.8631*	2.1422
8	4.5240	3.8866
9	4.5240	3.8866
10	4.5240	3.8866

Repeating the above reasoning for other attributes suggested that each feature tested using Granger-Causality Test is statistically significant in forecasting WTI crude oil prices at lags of 6 and 7. Although, some features were statistically significant in other lag values as well, the lags of 6 and 7 consistently returned the largest f-statistic. A slight increase of the f-statistic at lag 7 indicates that the latter can be considered as the optimal number of lags between weekly WTI crude oil and all the features used in this investigation. In order to verify the robustness of this finding, we selected a variable that is for sure fully disconnected from weekly WTI crude oil prices, as a control variable.

The variable chosen for this test was Justin Bieber’s sentiment on Twitter. The f-statistic values outputted from the Granger-Causality Test performed on Bieber’s sentiment, determined by both SS and SNLP sentiment analysers, revealed no correlation between sentiment and

weekly WTI crude oil price. Therefore the null hypothesis cannot be rejected and thus supporting the credibility of the initial finding. The results of this investigation are highlighted in Table 3

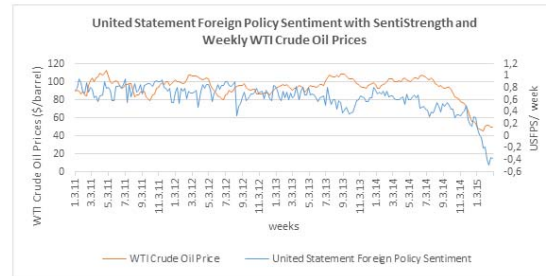


Figure 2. Relationship between weekly WTI crude oil price and United States foreign policy sentiment as per SS analyser.

Table 3. - Granger-Causality Test – statistical significance (f-statistic) at lags of 1 – 10 weeks between weekly WTI crude oil price and Justin Bieber’s sentiment as per SS analyser (alpha 0.05; bold = significant; *highest f-statistic) and SNLP.

Lag (n)	f-statistic (SS)	f-statistic (SNLP)	critical value (SS)	critical value (SNLP)
1	0.5451	0.8603	3.9934	3.9934
2	0.4743	1.4933	3.9959	3.9959
3	0.4743	1.4933	3.9959	3.9959
4	0.4743	1.4933	3.9959	3.9959
5	0.4743	1.4933	3.9959	3.9959
6	0.4743	1.4933	3.9959	3.9959
7	0.4743	1.4933	3.9959	3.9959
8	0.4743	1.4933	3.9959	3.9959
9	0.4743	1.4933	3.9959	3.9959
10	0.4743	1.4933	3.9959	3.9959

This analysis reveals the existence of an inverse correlation between the frequency of “oil” and OPEC references and WTI crude oil. It also shows that all the attributes tested in this investigation provide some statistically significance or “unique” information about the future values of WTI crude oil price at the lag of 7. As a result, all of the attributes are statistically significant inputs for the predictive model. It also shows that the overall time that a change in feature’s value takes to impact WTI crude oil price is 7 weeks. This is in line with the theoretical nature of demand and supply of oil. For example, a threat of a sanction may not immediately impact the supply of oil and thus the price of oil. A study conducted by GlobalPetrolPrices.com and Kimberley Amadeo⁷, a US economy expert, identified a 7 week lag between the change of WTI crude oil price and the price of gasoline.

⁷ <https://www.thebalance.com/how-are-oil-prices-determined-3305650>

D Predictive model

As a result of the previous analysis, all USFP and OC sentiment features are taken as inputs for our predictive model. Besides, based on the finding regarding the lag of 7, the predictive model will be built to forecast the price oil 7-weeks ahead. We hypothesize that the direction of weekly WTI crude oil prices *can* be *effectively* predicted using **machine learning techniques** and modelling weekly WTI crude oil price as a function of USFP and OC sentiment and frequency of “oil” and OPEC member references. In other words, the model attempts to solve a **binary classification problem** answering the question “What will the directional shift of the weekly WTI crude oil price in seven weeks? (Increase or a Decrease). For this purpose, the dataset is (automatically) labelled to account for the {increase, decrease} classes. Besides, the second order statistics of the attributes have been accounted for, see Table 4.

Table 4. Overview of features employed for forecasting system

Attribute	Description
oil_price	Weekly WTI Oil Price
usfp_senti	United State foreign policy sentiment as per SS
usfp_senti_var	Variance of United States foreign policy sentiment as per SS
usfp_senti_std	Standard deviation of United states foreign policy sentiment as per SS
usfp_nlp	United State foreign policy sentiment as per SNLP
usfp_nlp_var	Variance of United States foreign policy sentiment as per SNLP
usfp_nlp_std	Standard deviation of United states foreign policy sentiment as per SNLP
usfp_oil_freq	Frequency of "oil" in United States foreign policy tweets
usfp_opec_freq	Frequency of OPEC member references in United States foreign policy tweets
usfp_oil_var	Variance of frequency of "oil" in United States foreign policy tweets
usfp_opec_var	Variance of frequency of OPEC member references in United States foreign policy tweets
usfp_oil_std	Standard deviation of frequency of "oil" in United States foreign policy tweets
usfp_opec_std	Standard deviation of Frequency of OPEC member references in United States foreign policy tweets
oil_senti	Oil companies' sentiment as per SS
oil_nlp	Oil companies' sentiment as per SNLP
oil_senti_var	Variance of oil companies' sentiment as per SS
oil_senti_std	Standard deviation of oil companies' sentiment as per SS
oil_nlp_var	Variance of oil companies' sentiment as per SNLP
oil_nlp_std	Standard deviation of oil companies' sentiment as per SNLP
Classification {Increase, Decrease}	Class

Finally, given the availability of powerful machine learning toolkits allowing testing of several algorithms simultaneously, we employed Weka package [13] to assess the prediction capabilities of the already implemented classifiers. The initial result conducted on a randomly selected partition of the dataset with a 10-fold cross-validation method and default parameters of individual classifiers showed the emergence of three classifiers in terms of prediction accuracy. This consists of Support Vector Machine (SVM), Naïve Bayes and Multi-layer perception (ML) whose number of layers and neurons were initially tuned up [16]. Therefore, the overall architecture adopted for the prediction consists in combining the outcome of these three classifiers through a simple majority-voting mechanism in order to yield the final outcome of the prediction model [11].

III EXPERIMENT RESULTS

In order to quantifier the prediction results, standard precision and recall metric were used. These are defined below

$$Precision = \frac{true\ positive}{true\ positive + false\ positive}$$

$$Recall = \frac{true\ positive}{true\ positive + false\ negative}$$

After using a first bench of dataset (60% of total dataset) to train and optimize the model parameters, the classifiers were tested with the remaining bench of dataset. The output of the results are highlighted in Table 5 for the three classifiers.

Table 5 - Accuracy of classifier performance by class {CA = classification accuracy, ROC = ROC Area, 1 lag = 1 week, *best performing classifier, bold =lag with highest accuracy, Pre=Precision, Rec=Recall}

Classifiers	Lag 1				Lag 7			
	CA (%)	ROC	Pre	Rec	CA (%)	ROC	Prec	Rec
SVM*	55.71	0.52	0.56	0.56	74.29	0.74	0.75	0.74
Increase	50.00	0.52	0.52	0.50	83.80	0.74	0.72	0.78
Decrease	60.50	0.52	0.59	0.61	63.60	0.74	0.78	0.70
Naïve Bayes	52.86	0.54	0.55	0.53	67.14	0.69	0.78	0.97
Increase	68.80	0.54	0.49	0.69	40.50	0.69	0.94	0.41
Decrease	39.50	0.54	0.60	0.40	97.00	0.69	0.59	0.67
MLP	44.29	0.44	0.44	0.44	61.43	0.72	0.61	0.62
Increase	84.40	0.44	0.44	0.84	67.60	0.72	0.63	0.68
Decrease	10.50	0.44	0.44	0.11	54.50	0.72	0.60	0.55

The results in Table 5 showed that the lag 7 performs significantly better in all classifiers. This supports the results from the previous section that stressed that lag 7 is the amount of time that the features take to impact the WTI crude oil price. The results also revealed the SVM is the best classifier with accuracy of 74.29%. Naïve Bayes and MLP have achieved overall accuracies of 67.14% and 61.43%. These classification accuracies (CA) are based on an unseen data set with 70 instances (SET 2). SVM performed slightly better in classifying Increase instances than Decrease instances with 83.80% and 63.60% accuracy respectively. In contrast, Naïve Bayes observed a steep gap between its Increase CA and Decrease CA with values of 40.50% and 97.00% respectively. MLP model was better at classifying Increase instances.

The precision and recall of SVM model at lag 7 revealed that it strikes an adequate balance between the number of correct instances and the number of instances classified for both *Increase* and *Decrease* class (Increase {P = 0.72, R = 0.78}, Decrease {P = 0.78, R = 0.70}). In contrast, Naïve Bayes achieves precision of 0.94 for the *Increase* class but with only 0.41 recall, suggesting that it classifies a large number of *Decrease* instances as *Increase*.

In order to compare our results with some of the state of art oil price prediction models, Table 6 summarizes some of the key findings in this course.

Table 6. Comparison of state of art methods

Contributors	Approach	Freq	Period	Acc.
Morana [12]	Semi parametric approach	Daily	11/21/1998 to 1/21/1999	46.6 7%
Gori et al. [8]	Adaptive Neuro Fuzzy Inference System (ANFIS)	Monthly	2/1999 to 12/2003	45.7 6%
Fan et al. [5]	Genetic algorithm	Daily	6/27/2005 to 7/26/2005	54.5 4%
Ghaffari & Zare [6]	ANFIS	Daily	5/1/2007 to 5/31/2007	68.1 8%
Li et al. [10]	Least square support vector regression (LSSVR)	Monthly	1/2/2002 to 3/20/2009	52.5 2%
Chen [1]	Linear Regression	Monthly	1/1991 to 8/2012	65.0 0%
Shambora and Rossiter [17]	ANNs	Daily	1/1/1998 to 12/31/2003	53.1 0%
Our Model	Majority (SVM, NB, NN)	Weekly	1/3/2011 to 3/2/2015	76.2 9%*

From the results we can conclude that the novel inputs extracted from Twitter (USFP and OC sentiment, frequency of “oil” and OPEC references) are better predictors of future prices WTI crude oil prices than the previous inputs mentioned in literature. Furthermore, this study supports the

claim that non-linear models have superior forecasting abilities when it comes to crude oil predictions.

V CONCLUSION

This investigation allowed us to identify a gap in the existing methodologies of predicting crude oil. The literature showed that the current models relied on historical oil prices, financial instruments, oil variables (e.g. inventory levels), and irregular and infrequent events (i.e. elections, natural disasters) as inputs for their forecasting models. However, non-traditional methods but increasingly popular methods such as social media data analysis, word frequencies had not been explored as inputs to forecast crude oil, thus justifying further exploration.

Through the use of Twitter API, SQL, a methodology to extra and store US foreign policy and oil companies’ sentiment was achieved. The statistical study supported our selection of attributes and led to a novel discovery of the seven week lag. Using the encouraging results of the statistical study, the attributes with statistically significant correlation were used as inputs to the forecasting model. Building the model using three different types of classifiers (Naïve Bayes, SVM, and ANN), it was found that SVM returned the highest classification accuracy of 74%. The latter has been raised to 76% by using the majority voting strategy of the three classifiers. Through, the use of unexplored inputs, we were able to achieve and overall classification accuracy superior to that of existing models. Despite the superior performance, there are several **limitations** of our predictive model.

- The model relies on a limited number of Twitter accounts to retrieve the data. In the long term, these Twitter accounts may become more or less active, thus, potentially impacting the performance of the model.
- The model relies on sentiment, which is still only 80.7% accurate in the case of Stanford NLP and 64% in the case of SentiStrength. There is a large margin for error in sentiment which can ultimately impact or skew the accuracy of the model.
- The model uses a variety of different features to forecast the price of oil. However, are all of those features equally important? The relative contribution of each attribute to the overall classification accuracy can be explored as a next step for this investigation. Furthermore, can this similar methodology be applied to other grades of oil as well such as Brent or Dubai? Referring back to the original aim of the investigation on attempting to improve on the existing methods for forecasting the directional shift of crude oil price by introducing **Twitter** sentiment of **US foreign policy (USFP)** and **oil companies’ (OC)** as an input into the machine learning forecasting model for crude oil prices. Nonetheless, additional investigation still need to take place to test the robustness of the model.

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for their valuable suggestions because of which the technical quality of the work presented in this paper has improved. This work is partially supported by EU Marie Skłodowska-Curie grant No 645706 and EU grant 770469-Cutler.

REFERENCES

1. Chen, S.-S., 2014. Forecasting Crude Oil Price with Oil-Sensitive Stocks. *Western Economic Association International*, 52(2), pp. 830-844.
2. Chernenko, S., Schwarz, K. & Wright, J., 2004. The Information Content of Forward and Futures Prices: Market Expectations and the Price of Risk. *FRB International Finance Discussion Paper*.
3. Chin, M., Leblanch, M. & Coibion, O., 2005. The Predictive Content of Energy Futures: An Update on Petroleum, Natural Gas, Heating Oil and Gasoline. *HBER*.
4. Diakopoulos, N. & Shamma, D., 2010. Characterizing debate performance via aggregated Twitter sentiment. New York, ACM, pp. 1195-1198.
5. Fan, Y., Zhang, Y.-J., Tsai, H. & Wei, Y., 2008. Estimating Value at risk of crude oil price and its spillover effect using the GED-GARCH approach. *Energy Economics*, Volume 30, pp. 3156-3171.
6. Ghaffari, A. & Zare, S., 2009. A novel algorithm for Prediction of Crude Oil Price Variation Based on Soft Computing. *Energy Economics*, Volume 31, pp. 531-536.
7. Gilbert, E. & Karahalios, K. (2010). Widespread Worry and the Stock Market. 4th International AAAI Conference on Weblogs and Social Media (ICWSM), 2010
8. Gori, F., Ludovisi, D. & Cerritelli, P., 2007. Forecast of Oil Price and Consumption in the Short term Under Three Scenarios: Parabolic, Linear and Chaotic behavior. *Energy Economics*, Volume 32, pp. 1291-1296.
9. Granger C. J., 1969, Investigating Causal relationships by Econometrics Models and Cross Spectral Methods, *Econometrica*, Vol. 37, 1969, pp. 425-435.
10. Li X., Kaijian H, Kin K. L., and Yingchao Z., 2014, Forecasting Crude Oil Price with Multiscale Denoising Ensemble Model, *Mathematical Problems in Engineering*, Volume 2014
11. MacKay, D. J. C., 2003, *Information Theory, Inference and Learning Algorithms*, Cambridge Press
12. Morana, C., 2001. A Semiparametric Approach to Short-term Oil Price Forecasting. *Energy Economics*, Volume 23, pp. 325-338.
13. OpenCV, 2015. *Introduction to Support Vector Machines*. [Online] Available at: http://docs.opencv.org/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html [Accessed 1 April 2015].
14. Pindyck, R., 1999. The Long-run Evolution of Energy Prices. *The Energy Journal*, Volume 20, pp. 1-27.
15. Rachenko, S., 2005. The Long-run Forecasting of Energy Prices Using the Model of Shifting Trend. *University of North Carolina at Charlotte*.
16. Saedsayad.com, 2015. *Artificial Neural Network*. [Online] Available at: http://www.saedsayad.com/artificial_neural_network.htm [Accessed 1 April 2015].
17. Shambora, W. & Rossiter, R., 2007. Are There Exploitable Inefficiencies in the Futures Market for Oil?. *Energy Economics*, Volume 29, pp. 18-27.
18. Silva, E. G. S., Legey, L., L. F. & Silva, E. A. S., 2010. Forecasting oil price trends using wavelets and hidden Markov models. *Energy Economics*, Volume 32, pp. 1507-1519.
19. Socher, R., Perelygin, A., Wu, J. Y. & Jason Chuang, 2013. *Recursive Deep Models for Semantic Compositionality over a sentiment treebank*. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, p 1631–1642, Stroudsburg, PA, October. Association for Computational Linguistics.
20. Tehrani, R. & Khodayar, F., 2011. A hybrid optimized artificial intelligent model to forecast crude oil using genetic algorithm. *African Journal of Business Management*. Vol.5(34), pp. 13130-13135.
21. Wang, S., Yu, L. & Lai, K., 2004. A Novel Hybrid AI System Framework for Crude Oil Price Forecasting. *Lecture Notes in Computer Science*, Volume 3327, pp. 233-242.
22. Xie, W., Yu, L., Xu, S. & Wang, S., 2006. A new method for Crude Oil Price Forecasting Based on Support Vector Machines. *International Conference on Computational Science*, pp. 444-451.
23. Ye, M., J. Z. & J. S., n.d. Forecasting Short-run Crude Oil Price Using High and Low Inventory Variables. *Energy Policy*, Volume 34, pp. 2736-2743.
24. Ye, M., Zyren, J. & Sore, J., 2002. Forecasting Crude Oil Spot Price Using OECD Petroleum Inventory Levels. *International Advances in Economic Research*, Volume 8, pp. 324-334.
25. Yu, L., Wang, S. & Lai, K. K., 2008. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*.
26. Zhang, Xue, Fuehres, Hauke, and Gloor, Peter, 2010, Predicting stock market indicators through Twitter – “I hope it is not as bad as I fear”, In Collaborative Innovations Networks Conference, 1-8, Savannah, GA.