SSOE Crude Oil Forecasting

- Energy Staple of the World
- Contributes 1/3rd of the worlds energy consumption
- Important to businesses that heavily rely on fuel , such as airlines, plastic producers, agricultural etc.
- One of the most Actively traded commodities
- Price is extremely sensitive to the geopolitical events .

Problem Statement

- The aim is to predict the direction of change of BRENT values and predict if the market will be bullish or bearish.

Business Background

- SSOE is a leading bulk commodity trading solutions provider and logistics partner to clients around the world.
- They are mostly involved in physical trading of commodities
- They also have a derivatives trading desk which trades in Crude Oil and iron ore which started 3 years back.
- Their risk committee has a lot of restrictions so they are only allowed to make 1 trade per day (buy or sell)

Why Analytics ?

- SSOE trades only on their market understanding
- Not able to relate the magnitude of relation of the commodities with all the variables.
- Apart from the existing variables, we can also include more variables.
- We at GDNA can bring all the different variables and relate them to the crude oil prices.
- Also we can use techniques like NLP to gather the true market sentiment of the commodity.

Why NLP ?

- Certain Events that affect the price cannot be captured through numeric data.
- The cause and effect relationship between impactful news and market movements can be directly observed during the release of big news.eg. Even a single Tweet from Trump on oil can cause massive price variations.
- Easily scalable across different verticals.

- News information gives the true nature of the market of any commodity
- There are several ways to get the news information :
    - PDF
    - Tweets
    - News API
- Sentiment Analysis of the news can give us a sense of the market sentiment of the commodity
- Topic Modelling can help cluster similar news together

Collecting Twitter Data

- Python-twitter API helps us to scrape the tweets
- Tweets can be collected through a search keyword or from specific accounts
- With the actual text, we can also get information on number of likes and number of retweets
- With the help of the business, we were able to narrow down the search accounts to around 25-30 and specific keywords that give less noisy Tweets.
- For eg. If we just use the keyword ‘Crude Oil’ we were also getting tweets about oil advertisements and podcasts related to Crude Oil , which would add noise . By using proper combination of the keywords and accounts , noisy Tweets were avoided.
- Some of the important accounts are :
    - EIAgov
    - IEA
    - PlattsOil
    - WorldOIl
- An important step in news analytics is to clean the tweets of any html links, remove the stopwords.
- So we removed all the URL’s, usernames, hashtags and repeating characters.

Limitations of Python-Twitter API

- No access to Tweets older than 7 days i.e. No historical data without purchasing twitter premium developers account (expensive)
- A Maximum of 100 tweets can be scraped in a request

So to overcome these limitations, we began searching for other ways to get the Tweets.

Found a library **GetOldTweets3**, which has no restriction on historical Tweets, and also gives other metadata that the older API provided

Advantages of GetOldTweets3:

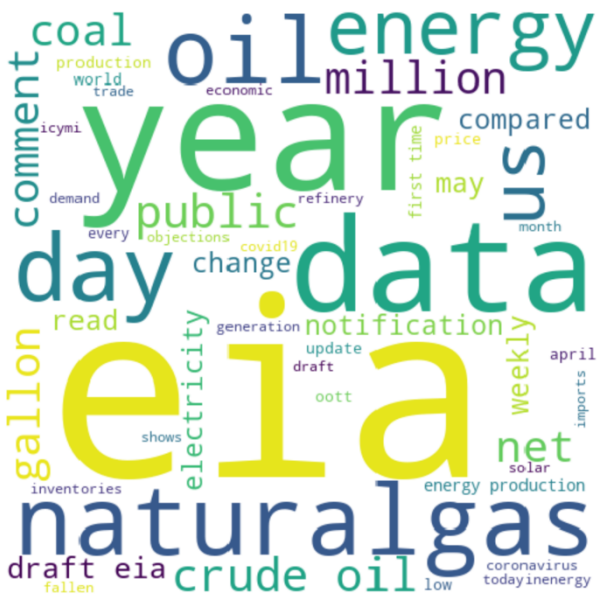
- Can get the Top Tweets automatically
- Can set the reference location to get tweets
- All possible tweets can be scraped
- Scrape through multiple users at a time

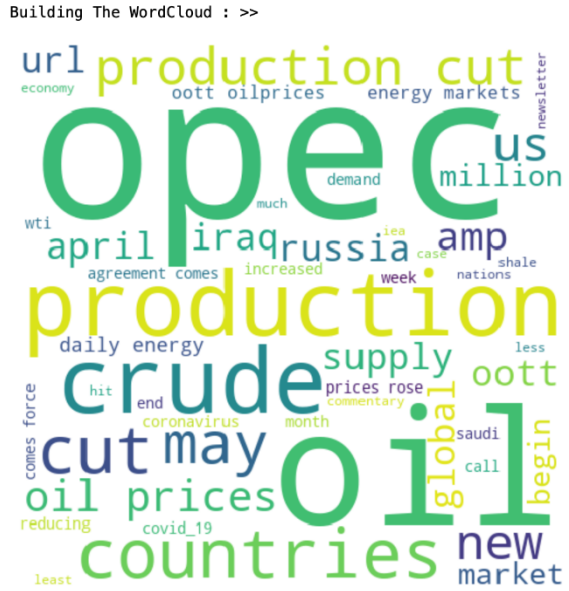
Sample Tweets

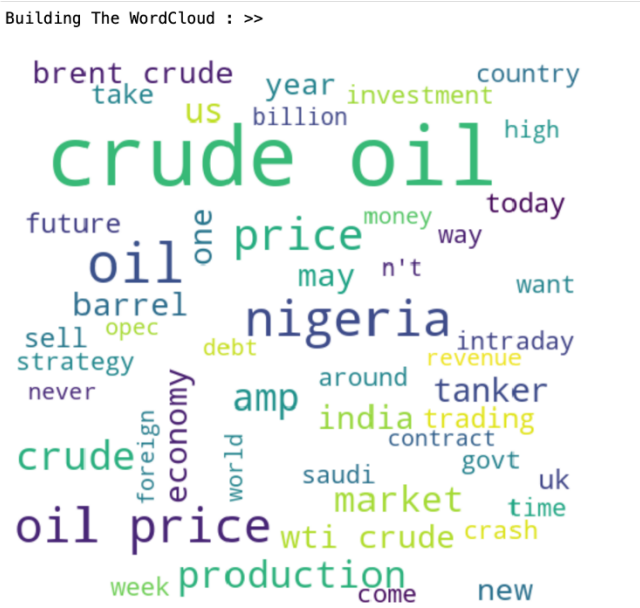
	Date	username	Tweet	Retweet-Count	Fav-Count
0	2017-01-29 23:54:06	HInnov8	Campaigners demand answers after leak closes part of North Sea oil well http://j.mp/2jHcMEB	0	0
1	2017-01-29 23:32:07	CanadaAc	Global oil demand will continue to grow, we must also replace about 3.5% of current oil production every year lost to depletion and decline.	2	5
2	2017-01-29 23:17:09	JaroGiesb	Balance of supply and demand according to the Opec report is 31.59mb/d #yeg #yyc #abpoli #cdnpoli #yxh #yql #ymm #opec #oil #Uspoli #imf	1	0
3	2017-01-29 23:02:35	LeinerKO	If we can reduce the oil drilled, that will force demand down, and price up. The environment will be better off. #shellno	0	0
4	2017-01-29 22:55:10	JaroGiesb	Projected demand for 1Q17 - 94.59 mb/d According to the Opec report. #yeg #yyc #abpoli #cdnpoli #yxh #yql #ymm #opec #oil #Uspoli #imf	1	0
5	2017-01-29 22:50:22	Wilderne	Join us in our fight & SPEAK UP to demand Congress stop trying to dismantle oil & gas reforms! https://wilderness.org/methane	30	45
6	2017-01-29 22:49:36	JaroGiesb	World Oil demand 4th-Q 95.3 mb/d acording to the Opec report. #yeg #yyc #abpoli #cdnpoli #yxh #yql #ymm #opec #oil #Uspoli #imf	1	1
7	2017-01-29 22:40:18	ADIAnalyt	Join ADI's upcoming events! Webinar on oil and gas outlook (https://goo.gl/h0OOkN) and forum on gas / NGL demand (https://goo.gl/jJfZac)	0	0
8	2017-01-29 22:17:07	JeromeOL	Actus Mer/Sea News: Campaigners demand answers after leak closes part of North Sea oil well - @guardianeco http://dlvr.it/NDtXSr	0	0
9	2017-01-29 22:03:07	IndustryW	BP sees a future of slowing oil demand growth, abundant supplies http://buff.ly/2k5jmn0	2	0
10	2017-01-29 22:00:11	JonDeeOz	BP calls peak coal - also sees renewables grow and cuts long-term oil and gas demand. http://www.theaustralian.com.au/business/mining-energy/bp-sees-renewables-grow-cuts-longterm-oil-gas-demand/news-story/111cbdb74c524b0ccc935f7e107d1620	0	1

Word Cloud

- Using the cleaned Tweets, we plotted some WordClouds which often provide a bigger picture of the market
- Its clear from them that most of the tweets that time were talking about the impact of Covid-19 on the oil market and the oil production cut in Nigeria







Sentiment Analysis

- Sentiment Analysis of the twitter Data can provide us a sense of the market direction.
- Using Vader Sentiment , we tried to analyse a sample of the Tweets.
- Although not trained on Crude Oil specific dictionary, Vader is Able to predict the emotion in the tweets upto a good accuracy

	Text	Sentiment-Score
0	U.S. oil prices have crashed below $0 a barrel. This is not a problem. We should not try to prop up oil prices. We should not bail out oil companies. We should accelerate our shift to solar, wind, and other non-fossil fuels	0.3089
1	We will never let the great U.S. Oil & Gas Industry down. I have instructed the Secretary of Energy and Secretary of the Treasury to formulate a plan which will make funds available so that these very important companies and jobs will be secured long into the future!	0.5706
2	Trump filled up the Strategic Petroleum Reserve a month ago with oil at $30 a barrel. \n\nRight now, it\u2019s going for -$10 a barrel. \n\nAnother brilliant move by our dealmaker-in-chief.	0.5859
3	FIRST IN @IndianExpress: Hammered by the global drop in oil demand due to #COVID19 shutdowns & an oil price war, energy major Schlumberger withdraws job offers to fresh IIT graduates https://t.co/sN9iTJYaxy	-0.4767
4	CME to allow listing of negative oil options https://t.co/JRxiiogeZI https://t.co/sriPsVTOW0	-0.4215
5	Here is what #R350 can do for you, people should stop complaining from a point of privileges	-0.1027
6	Oil tankers are gathered off California\u2019s coast with nowhere to go	0
7	RT @spectatorindex: JUST IN: Brent crude oil price falls to $16.5, the lowest in over twenty years.	-0.743
8	RT @HeavierThanMost: negative oil means the oil is turning back into dinosaurs	-0.5719
9	Never thought I'll see oil prices tanking like this. Unbelievable.\n\n#OilMarketCrash	0.3612
10	RT @CNBC: Saudi Arabia is the winner from oil's historic price plunge, analysts say https://t.co/catqu2Wmxw	0.5859
11	ACCC directs petrol stations to pass on record-low oil prices https://t.co/ojRmlXFI0a	0
12	RT @Pun_Starr: Rupee is at an all-time low.\nCrude oil price is at an all-time low.\nOnly communal hate & bigotry are at an all-time high.	-0.5719
13	In the US, the oil billionaires are overwhelmingly republican. If they go bankrupt, the US is likely to change its climate policy (although oil would be cheaper to burn).\n\nThus, let us hope for two new decades of low energy prices!	-0.3595
14	RT @iimcomic: USA has a lot of oil now.\n\nExplains why the US is on a mission to destroy itself.	-0.5423
15	Oil prices plummeted below zero for the first time ever on Monday, marking the industry's worst crisis in a generation. How did this happen and what comes next? https://t.co/GUM2JyB7fR (Video via @QuickTake) https://t.co/z2AAqn5YAm	-0.8481

Topic Modelling

- On further discussion with the business, we wanted to find out the Top Topics which were in the Tweets, which would later help us in manual labelling .
- Topic Modelling is a method of unsupervised classification of documents, similar to clustering on numerical data, which finds natural groups of items(topics)
- We can choose the number of topics and number of words in each topic to get better results
- The Topics again talk about the effects of the coronavirus lockdown on the demand of oil, supply cuts, tensions between OPEC and Russia.

Topics found via LDA:

Topic #0:
oil crude inventories rise output trade sources cuts venezuela nigeria

Topic #1:
oil crude demand russia production tensions prices palm industry june

Topic #2:
crude oil price prices 34 futures 35 update source wti

Topic #3:
oil crude prices market despite lower gold cheer aims struggle

Topic #4:
oil crude high new week price 10 demand buy wti

Topic #5:
oil crude price million india api build states storage store

Topic #6:
crude oil gold gas trading silver 00 commodities natural markets

Topic #7:
oil crude prices north long supply short cuts wti american

Topic #8:
crude oil gt prices demand day nigeria recovery coronavirus lockdowns

Topic #9:
crude oil saudi china arabia russia supplier prices igr april

Modelling (NLP)

- Initially we thought of making a classifier with Demand, Supply, Price Up or Down and Long-term/Short-term classes.
- But realised that it would require a lot of manual labelling of the Tweets with the help of business
- On exploring, we found a great research paper that used the change in price of a commodity with the news articles to predict the direction of each day
- So for the baseline model, we took the change in BRENT prices and categorized the labels into 1, -1 and 0.
- Extracted almost 20,000 tweets from the year 2018
- Achieved an accuracy of 60% for the model
- Next Step is to model for 1,00,000 lakh tweets which have already been scraped. With more amount of data, we expect a higher accuracy.
- We also look to provide the business with not just Up or Down price change, but also the magnitude of the change which helps get a better sense of the market

Quantitative Model

Data Preparation

- Variables consists of both daily and weekly data (interpolated to daily level)
- Taking lags : 2 lags for the weekly variables and 5 lags for daily variables
- Took daily and weekly change values to incorporate for the directional changes
- Including technical indicators like RSI, MACD, EMA, Bollinger Bands.

Modelling

- 1 day ahead forecast
- Training on 2013-2018, Tuning on Jan 2019, Testing on Feb, March, April
- We used XgBoost to train the model and Hyperopt for the Hyper-Parameter Tuning, initially with MAPE as the objective function and then incorporating direction also.
- In terms of metric, after multiple iterations and optimizations, we were getting an Average MAPE of 2-3% and directional accuracy of 55%. Considering the volatility of the series, the directional accuracy we are getting is good for baseline model
- Looking for a better prediction, we were also interested in the range of predictions and not just a single estimate, so we also tried approaches like Random Forest Quantile Regression.

Results Summary

Iteration description	% times correct direction					
1 XGBOOST with fundamnetals and change using MAPE as metric	46.0%					
2 XGBOOST with fundamnetals and change using MAPE and Direction as metric	54.0%	Jan	Feb	March	April	
		62%	60%	33%	61%	
3 XGBOOST with all variables and Technical Indicators using MAPE and Direction as n	55.3%	Jan	Feb	March	April	
		52%	60%	57%	52%	

Thank You