IDEATION FOR CRUDE OIL PRICE PREDICTION

Crude oil prices typically fluctuate based on seasonal demand and supply. Most recently, the COVID-19 pandemic caused crude price changes through a drop in demand. While economic recovery is underway, oil prices continue to be affected by global uncertainties.Future oil prices will depend greatly on innovations in energy, transportation, and other industries as societies work to become less fossil fuel dependent.Crude oil prices typically fluctuate based on seasonal demand and supply. Most recently, the COVID-19 pandemic caused crude price changes through a drop in demand. While economic recovery is underway, oil prices continue to be affected by global uncertainties.Oil prices are affected by several factors that include everything from weather to economic and political instabilities.

Reasons for Today's Volatile Oil Prices

- Oil prices used to have a predictable seasonal swing. They spiked in the spring as oil traders anticipated high demand for summer vacation driving. Once demand peaked, prices dropped in the fall and winter.
- Global supply and prices are also affected greatly by geopolitical conflict and civil unrest.
- Oil prices are more volatile today due to many factors, but five are the most influential.

The Russian Invasion Of Ukraine

Russia is the third-largest producer of liquid fuels and petroleum, so when the country invaded Ukraine in late February 2022, it had immediate impact on Brent crude oil futures prices.10 As the conflict continued, the prices of crude oil settled in out on an upward trajectory, reaching nearly $130/b in early March, and staying well above $100/b into April.7

The choice of the four media platforms included in the analysis was founded on their size – in terms of number of Users/media sources and produced content (the bigger the data source, the more the information that can be analysed) –,By their reputation, and by the possibility to get free access to the data. Each specific choice is better supported in the Following.Twitter is a worldwide popular platform, which offers a social networking and microblogging service. It enables its Users to update their status in tweets, to follow people they are interested in, and to communicate with them directly. Twitter users include, but are not limited to, commodity traders, politicians, companies, activists, and major news Outlets, as well as casual users. Twitter describes itself as "a real-time information network that connects you to the Latest information about what you find interesting". It is a social awareness system, which delivers a fragmented mix of Information, enlightenment, entertainment, and engagement from a range of sources . In 2012, Twitter had more Than 100 million registered users posting more than 340 million tweets per day1, updated to be 310 million monthly Active users on March 2016. Thus, Twitter popularity has drawn more and more researchers' attention from different.Disciplines, to understand its usage and community structure, influence of users and information propagation, and its Prediction power and potential application to other areas]. Twitter data have been collected using the social network And semantic analysis software Condor, developed by Galaxy Advisors3. It enables the visualisation, measurement, and Analysis of the communication structure in social networks, as well as the analysis of the use of language over time.We collected data from Twitter by fetching all the tweets that contained the search term "crude oil price". Collected Tweets were then filtered, excluding all non-relating posts –

for instance, those referring to cooking or lubricating oils. Analysing the tweets, four Twitter variables have been determined, i.e. Number of tweets per day, Sentiment, Complexity and Emotionality.In addition to Twitter, Wikipedia, the well-known online encyclopaedia, has been included in this research. Currently, Wikipedia is the largest knowledge repository on the web. Wikipedia is available in dozens of languages, and Its English version is the largest of all with more than 400 million words in over one million articles [62]. It is also Densely structured: its articles have in total hundreds of millions of links. These connections link the topics being Discussed, and provide an environment which fosters serendipitous gathering of information [63]. This huge amount of Information and links provides a real opportunity to help unfold the world history and explore upcoming events. Different from Twitter, the quality of information in Wikipedia is controlled and consequently higher. Within the Wikipedia research community, findings are constantly published and verified, and the reputation of an author grows With the reputation of his contributions [64]. The distribution of users' reputation in Wikipedia shows that saboteurs and Inexpert users are quite a minority compared to high reputation users [65].

FOr Wikipedia, the daily views count of the pages relating to Crude Oil Price, such as "Benchmark (crude oil)", "World oil market chronology", "2000s energy crisis", and others, were collected. For the final study, two pages have Been considered – "Price of oil" and "OPEC" – as their traffic statistics were the most significant in terms of correlation With the dependent variable.Google is the first search engine in the world, which makes it one of the most reliable resources for investigating web Search queries. Since 2004, Google has been providing three data sources that can be useful for social science: Google Trends, Google Correlate, and Google Consumer Surveys [66]. Google Trends is commonly used in "now-casting" or in The prediction of the present, the very near future, and the very near past [67].In this study, the search queries were set to be the same as the page titles of Wikipedia, mainly for two reasons: (1) to Show how different public audiences can produce different predictive activity using the same keywords on two different Platforms; (2) to show the response time difference between the two platforms. An analogous procedure to that of Wikipedia has been applied to gather Google Trends data. This data is available directly on the Google website, which Allows filtering for location, time range, and keywords. The location has been set to "worldwide", whilst the queried Keywords have been "Price of oil" and "OPEC". Other possible keyword combinations have been tested, obtaining less Significant results.The Global Data on Events, Location and Tone (GDELT) is the final platform considered. The GDELT Project is an Open source repository of news articles, which is continuously updated and made available to researchers through an Application program interface. Initially, the GDELT Project was a coded dataset of 200 million geo-located events; now, It has been updated to be 400 million events spanning over more than 12,900 days4. The dataset includes more than 300 Different types of events: therefore it reveals all that has happened in place and time, since 1979. Kwak and An [68] Referred to it as a tale of the world. The GDELT Project is described as an initiative to construct a catalogue of human Societal-scale behaviour and beliefs across all countries of the world. It connects every person, organization, location, Count, theme, news source, and event across the planet into a single massive network. The database relies on tens of Thousands of broadcasts, print and online news sources from every corner of the globe [69]. Including GDELT in the Analysis is important to control for world events related to the WTI Crude Oil Price and to have a proxy for media Activities on newspapers.GDELT data was obtained by crawling the Global Knowledge Graph (GKG) dataset available on the project website. The daily number of newspaper articles covering "WTI Crude Oil Price" search criterion and the count of all Organizations names or advisory councils mentioned all over the world during the period of the study have been Extracted.