Impact of Crude Oil on Indian Economy

MSc Research Project
Data Analytics

Ronit Kumar Pareta
Student ID: x21223645

School of Computing
National College of Ireland

Supervisor: Dr Ahmed Makki

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Ronit Kumar Pareta		
Student ID:	X21223645		
Programme:	Data Analytics	**Year:**	2022-2023
Module:	MSc Research Project		
Supervisor:	Dr Ahmed Makki		
Submission Due Date:	14-Aug-23		
Project Title:	Impact of Crude Oil Price on Indian Economy		
Word Count:	**21 Page 7171**		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Ronit Kumar Pareta

Date: 18-Aug-23

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Impact of Crude Oil Price on Indian Economy

Ronit Kumar Pareta
X21223645

Abstract

Crude oil price fluctuations have far-reaching effects on emerging economies like India which imports over 80% of its oil requirements. Understanding this impact is vital for macroeconomic stability. Despite extensive global literature, India-specific quantitative evidence on crude price-economy linkages remains limited. This research addresses this gap focusing on major Indian macroeconomic indicators. This thesis investigates the impact of international crude prices on key Indian macroeconomic indicators using linear regression, random forest, gradient boosting, and neural network models. The study collects extensive time series data and applies a model comparison approach, leveraging the strengths of different modelling techniques. Analysis of various important analyses provides new insights into transmission channels. Neural network model outperforms the other models with least value of mean squared error, mean absolute error, and root mean square error value. This thesis contributes significant India-focused empirical evidence, helping contextualize the macroeconomic effects of oil uncertainty. The insights can inform policy aimed at navigating oil price volatility and ensuring economic stability. The analysis focuses on macroeconomic indicators. Sectoral and external factors are excluded. Regular model updating is required to track evolving dynamics. In conclusion, this thesis enhances scholarly understanding of oil-economy linkages for India through rigorous quantitative analysis. The evidence holds relevance for policymakers and provides reference for future research.

Keywords: Indian GDP, Stock market, unemployment, crude oil volatility, machine learning models.

1 Introduction

In the landscape of global economics, the connection between fluctuation in crude oil price and key macroeconomic indicators stands as a pivotal topic of investigation. This thesis aims to delve into the intricate dynamics between fluctuations in crude oil prices and the consequential impact on various critical macroeconomic indicators within the Indian context.

When undertaking academic research, it becomes evident very quickly that the subject at hand has considerable significance because it could affect economic trajectories and form the criteria for political decisions. Crude oil has fundamentally altered how the world uses energy since the middle of the 1950s, contributing significantly to the industrialization and economic growth of numerous nations. The pertinence of oil goes beyond its merely being a source of energy and extends to include strategic importance inside geopolitical terrains. Furthermore, its influence extends to both developed and developing economies, having a significant and long-lasting impact on their complex dynamics.

The important work of emphasizes(Ahmad *et al.*, 2022) the crucial function that crude oil has taken on as a major energy source, effectively serving as the lifeblood and catalytic engine powering the industrialized world towards unparalleled growth. This is because crude oil has become a vital energy source, a role that is highlighted in the work. Parallel to this, the ground-breaking hypotheses put forth by (Shao and Hua, 2022) highlight the importance of oil as a strategic cornerstone in the field of international politics, resonating with its varied consequences across geopolitical domains. This body of information is especially pertinent for understanding how deeply oil's byproducts are woven into the daily fabric of Western culture. Oil plays a crucial part in addressing a variety of human requirements, from industry to everyday domestic tasks, as seen by the wide range of products it is used to create, from basic household necessities to essential industrial necessities.

The history of crude oil's impact on economies dates back several decades, witnessing fluctuations that have ranged from periods of abundance to scarcity. Throughout this history, the association between crude oil price volatility and its impact on economic indicators has been a matter of interest for economists and policymakers alike. The intricate connection between oil prices and economic performance has become increasingly evident, warranting detailed investigation to inform robust policy measures.

In the Indian context, the impact of crude oil price volatility on macroeconomic indicators has been evident in varying degrees over time. Empirical analyses have revealed that oil price fluctuations can lead to changes in output growth, inflation, and monetary policy responses (Masood and Syed Mohd Shahzeb, 2023). The intricacies of this impact demand a nuanced understanding to facilitate effective policy formulation that mitigates potential negative consequences and capitalizes on positive opportunities.

India appears as one of the most prominent consumers and procurers of crude oil. According to figures issued by the Ministry of Petroleum and Natural Gas[1], India's consumption of petroleum products maintained at a substantial 213.2 million tons, with imports reaching a record 809.2 mega tonnes of oil equivalent in the year 2018. Such pronounced reliance on crude oil underscores India's significance in the global energy landscape. Projections for India's energy demand, spanning from 2017 to 2040, anticipate a Compound Annual Growth Rate (CAGR) of 4.2%, affirming the nation's escalating energy requisites.

Addressing the impact of crude oil price volatility on Indian macroeconomic indicators carries multifaceted benefits. A comprehensive understanding of this relationship can aid policymakers in designing adaptive economic strategies that minimize the adverse effects of oil price shocks. Furthermore, insights derived from this research can foster economic stability, sustainable growth, and improved living standards by enabling effective risk management and resource allocation.

The complexities of this relationship lie in the intricate network of factors that mediate the impact of oil price fluctuations. Furthermore, the evolving global energy landscape and its potential transition to renewable and green energy sources introduce further dimensions to this intricate problem.

The research subject at the center of this study originates from the requirement to grasp the complex connection between crude oil price volatility and its impact on crucial macroeconomic indicators inside the Indian context.

1.1 Research questions

1) What is the relation between fluctuations in crude oil prices and the Indian Bombay Stock Exchange (BSE)?
2) What is the impact of crude oil price on Indian unemployment rate?
3) What is the impact of crude oil price on Indian inflation rate?
4) How does the speculation-driven price movement of crude oil influence India's Gross Domestic Product (GDP)?

[1] https://mopng.gov.in/en/documents/annual-reports

This study aims to contribute to educated policy choices that direct India's economy toward resilience, stability, and sustainable growth by analyzing the complex nature of this relationship. As the global energy landscape continues to evolve, a comprehensive understanding of these dynamics becomes imperative for shaping a prosperous economic future. However, some limitations exist the study's temporal scope is relatively short, possibly constraining generalizability. Furthermore, external factors like pandemics and government policies have not been fully considered.

1.2 Project Outline

This thesis is structured into seven core sections. The introduction provides background and motivations driving the research. Section 2 reviews relevant scholarly works linking crude oil prices and economic performance. The research design methodologies are elaborated in Section 3. Section 4 presents the model specifications and architectures. This is followed by Section 5 detailing the implementation of predictive analytics techniques. Experimental results and evaluations are covered in Section 6. Finally, Section 7 summarizes conclusions and proposes avenues for future work. The logical flow moves from framing the research context and questions to methodology, model development, implementation, results analysis, and synthesis. This organized structure facilitates clear communication of the study's purpose, approach, findings, and implications.

2 Related Work

The important connection between crude oil prices and a nation's economy has gained significant attention within the academic and policy-making arenas. The relationship between these two dynamic factors has been a subject of substantial examination, with scholars diving into its numerous dimensions and consequences in the context of the Indian economy, the impact of crude oil prices assumes essential relevance due to the nation's significant energy requirements and its status as a major oil importer. This section of the research study dives into the existing body of knowledge on the influence of crude oil prices on the Indian economy, presenting major findings and insights from earlier research projects.

2.1 Crude Oil and Economic Activities

(Periwal, 2023) undertakes an in-depth examination into the relationship between crude oil price changes and significant Indian macroeconomic indices like GDP, Index of Industrial Production (IIP), and Wholesale Price Index (WPI). Applying statistical and analytical approaches, the study shows the interplay between these variables and their impact on India's economic trajectory. A key contribution is the examination of how unpredictability in oil prices cascades through consumer behaviour and industrial strategies, uncovering real-world implications using concepts from behavioural economics. Methodologically, time series data analysis and regression techniques help quantify the influence of oil price volatility on macroeconomic factors. The study concludes by emphasizing crude oil's significant impact on key indicators like GDP, IIP, and WPI. However, limitations exist regarding temporal scope, global outlook, and sectoral analyses.

Shifting focus to Thailand, (Rafiq, Salim and Bloch, 2009) delve into oil price volatility's multifaceted relationship with unemployment, investment, and other critical macroeconomic dimensions between 1993-2006. The realized volatility method and vector autoregression models underpin the empirical analysis, complemented by structural break tests to account for the Asian Financial Crisis. Key contributions include rigorously quantifying volatility and utilizing sophisticated econometric techniques to model complex interrelationships. The study reveals oil price volatility's considerable influence on unemployment and investment in Thailand. Additionally, it underscores volatility's lasting impact on post-crisis budget deficits facilitated by the new exchange rate regime. However, the localized geographical context limits generalizability.

Widening the lens to 1991-2006, (Narayan and Narayan, 2007) comprehensively investigate crude oil price volatility dynamics using daily data and sub-sample analysis. The disaggregated analysis unearths asymmetry and shock persistence as core findings, challenging assumptions of homogeneous oil price behaviour. The study contributes empirically substantiated insights into volatility's nuanced subtleties, like negative shocks disproportionately reducing volatility. The authors thoughtfully interpret the evidence to highlight the need for incorporating regime shifts when modelling oil prices given their economic growth implications. However, the research is restricted to volatility metrics without examining fundamental supply-demand drivers.

Shifting focus to the 2000-2020 timeframe, (Xiuzhen, Zheng and Umair, 2022) shed light on oil price volatility's impact on global economic recovery, investment, and crises episodes like COVID-19. The vector autoregression framework enables a nuanced analysis of interrelationships. Key contributions include highlighting volatility transmission channels between oil and the economy and underscoring COVID-19's enduring influence. Findings suggest targeted policy interventions can mitigate excessive oil price declines. However, the absence of pre-2000 data limits historical context, while supply-demand factors remain unexplored.

Employing an innovative methodology, (Dong et al., 2019) use wavelet analysis to discern connections between oil prices and global economic activity. By circumventing assumptions about model specifications, the study offers an unbiased perspective. The technique reveals correlations between oil and economic activity over short-term horizons throughout the sample period, but less co-movement at lower frequencies. The lead-lag relationships uncovered enhance theoretical understanding of oil-economy interlinkages. However, the failure to account for other variables like supply-demand factors and the limited sample period are notable limitations.

Shifting focus to economic policy uncertainty, (Lyu et al., 2021) assesses the impact of global EPU shocks on Brent and WTI crude oil price volatility. The application of time-varying SVAR models provides insight into evolving dynamics, a key theoretical contribution. Findings reveal that EPU shocks can substantially increase oil price volatility, especially during crises. However, the exclusive focus on EPU overlooks other potential drivers, while methodological limitations persist.

In synthesis, these studies collectively enrich our understanding of how oil price unpredictability transmits across macroeconomic and financial systems. Sophisticated econometric approaches help quantify complex interrelationships and causal mechanisms.

However, geographical, temporal, and methodological constraints necessitate caution in interpreting and generalizing findings. As oil market turbulence persists, these works highlight the imperative of evidence-based policymaking to navigate uncertainty. Their empirical insights offer guideposts for economic stability and growth amidst an intricate global energy landscape.

2.2 Crude Oil Impact on Emerging and Developed Countries

Focusing on South Korea from 1988-2005, (Masih et al., 2011) make vital contributions by utilizing sophisticated time series techniques to uncover the evolving impact of oil price volatility on stock returns. Their methodological rigor in applying vector error correction models to capture complex interrelationships enables a nuanced understanding of this linkage over time. Key findings reveal the growing influence of unpredictability, necessitating adaptive risk management strategies. However, geographical confinement limits generalizability.

Shifting focus to China between 1999-2008, (Qianqian, 2011) provides unique insights by employing advanced error correction models to identify oil prices' detrimental effects on economic growth, exports and money supply. Leveraging cointegration analysis to quantify intricate equilibrium relationships represents a pivotal contribution. While illuminating oil's role in hindering stability, the absence of policy insights and narrow scope constrain applicability.

Offering an unprecedented century-scale analysis of major OECD economies, (van Eyden et al., 2019) uncover strong empirical evidence of oil volatility's adverse impact on economic growth. Methodologically, their application of sophisticated realized volatility measures and advanced panel estimators represents a significant value-add. The granular quantification of country-specific effects enhances theoretical understanding. However, the limited focus on direct growth linkages remains a constraint.

Spanning 1970-2014, (Diaz et al., 2016) conduct an intricate examination of oil volatility transmission to G7 stock markets using rigorous VAR models. Handling structural shifts and contrasting global versus national price effects represents vital contributions. The identification of pronounced linkages during high volatility periods holds significance. However, the focus on G7 economies limits wider generalizability.

Employing advanced causality testing, (Su et al., 2021) provides unique insights into asymmetric relationships between oil shocks and economic uncertainty in BRICS countries. The alignment with theoretical underpinnings enhances contextual relevance. Findings revealing country-specific nuances represent key empirical contributions. However, geographical confinement raises generalizability concerns.

In conclusion, these studies collectively advance multidisciplinary knowledge on complex oil-economy interlinkages by empirically highlighting volatility spill-over channels. Their sophisticated time series analyses to quantify intricate relationships and causal mechanisms represent major technical contributions. However, scope limitations underscore balanced interpretation. As oil uncertainty persists, these works highlight data-driven policy imperatives to maintain stability. Their insights enrich scholarly understanding, guiding resilience against multifaceted volatility risks.

2.3 Crude Oil and Stock Market

Exploring the US from 1973-2015, (Rahman, 2022) provides novel insights into asymmetric oil-stock relationships using sophisticated nonlinear structural VAR models. Capturing volatility intricacies within a bivariate framework represents a key technical contribution. Findings reveal oil's pivotal role in driving stock market asymmetry, enhancing theoretical understanding. However, the isolated focus on oil overlooks other influencing factors.

Shifting focus to India, (Ghosh and Kanjilal, 2016) significantly advance emerging market knowledge by applying threshold cointegration to uncover structural breaks in oil-stock linkages. The methodical incorporation of exchange rates in a multivariate setting represents a major value-add. Findings suggest the oil-stock relationship holds only in specific phases amidst structural changes. However, model assumptions and variable scope constraints persist.

Offering contemporary empirical evidence, (Jain and Biswal, 2016) employ advanced DCC-GARCH frameworks to elucidate intricate connections between gold, oil, exchange rates, and Indian stocks. The sophisticated causality testing provides unique insights into lead-lag relationships. Findings underscore policy imperatives to mitigate exchange rate and stock market fluctuations using oil and gold levers. However, an isolated focus on selected variables overlooks other potential drivers.

In summary, these studies enhance scholarly understanding of complex oil-economy interlinkages by pinpointing asymmetric and regime-dependent relationships. Their application of sophisticated econometric and time series methodologies to capture nuances represents a pivotal technical contribution. However, constraints regarding model assumptions and variable scope highlight the need for balanced interpretation. As oil-economy dynamics evolve, these works provide data-driven perspectives to guide policy aimed at navigating uncertainty and promoting stability. Their empirical insights enrich multidisciplinary knowledge, furthering evidence-based strategies.

2.4 Crude Oil Impact on GDP and Inflation

Focusing on Qatar, (Charfeddine and Barkat, 2020) provide novel insights into oil-economy dynamics by quantifying asymmetric price/revenue effects on macroeconomic performance and diversification. Their methodical incorporation of external factors within a structural VAR framework represents a vital technical contribution. Findings reveal the nuanced short-run impacts of positive oil shocks alongside the economy's long-term resilience, underscoring oil's dual role. However, the narrow geographical focus limits wider generalization.

Shifting focus to India, (Mohanty and John, 2015) significantly advance dynamic modeling by applying time-varying SVAR techniques to capture evolving inflation drivers. Permitting flexible parameter adjustments over time is a pivotal methodological contribution. Findings reveal frequently changing determinants, although monetary and fiscal policies consistently shape outcomes. However, the model excludes external factors overlooking recent developments.

Offering empirical evidence from 1993–2006, (Mandal, Bhattacharyya and Bhoi, 2012) quantify the increased oil price pass-through to Indian inflation and output, especially post-2002. Simulating complete deregulation scenarios represents a key analytical contribution. Findings confirm the disproportionate passthrough compared to global levels alongside macroeconomic implications. However, the simplicity of the VAR approach overlooks multifaceted relationships.

In summary, these studies enhance scholarly understanding of intricate oil-economy interlinkages by underscoring asymmetric and time-varying dynamics using sophisticated econometric approaches. Their novel applications of structural VAR and stochastic volatility techniques to capture nuances represent pivotal technical contributions. However, constraints regarding geographical focus, variable scope, and model assumptions highlight balanced interpretation needs. As the convoluted oil-economy relationship evolves, these works offer data-driven guideposts for policymaking aimed at stability amidst uncertainty. Their empirical insights advance knowledge, directing evidence-based strategies.

2.5 Conclusion

The studied study offers useful insights on oil price swings' diverse economic impacts globally. The studies apply rigorous analytical methodologies to demonstrate oil's complicated linkages with macroeconomic trends, financial markets, inflation, and fiscal parameters across countries. While providing in-depth perspectives on big economies, the literature finds information gaps regarding oil uncertainty consequences on emerging countries like India. This suggests areas for future research to deepen understanding of how oil price fluctuations may affect India's GDP, inflation, and fiscal situations. Overall, the studied research increases technical and contextual understanding of crude oil's economic repercussions, informing evidence-based decisions. They underline the need for sophisticated awareness of oil-economy links to ensure stability despite global energy landscape challenges.

3 Research Methodology

3.1 Methodology

This chapter discusses the techniques adopted to uncover the connections between crude oil price fluctuations and key economic factors in India. Statistical and econometric methodologies like regression and machine learning models will be leveraged to analyse the empirical correlations between variables based on the research goals and assumptions. The overall strategy aims to provide data-driven insights into how oil price volatility affects the Indian economy.

In this study, the link between crude oil prices and macroeconomic factors in India is examined using KDD. The KDD process is iterative and may involve refining data selection, cleaning, mining algorithms, and pattern evaluation to achieve accurate and relevant knowledge extraction. The process aims to ensure that the extracted knowledge is valid, novel, useful, understandable, and interesting.

Linear regression, random forest, gradient boosting, and neural network models will be employed for most relationships, while Gaussian Process Regression will be used for the correlation between crude oil prices and unemployment due to limited data availability. The approach comprises data collection, variables selection, econometric methodologies, and analytical frameworks used for this study. I am applying the above discussed models and comparing the results, the process of model comparison and evaluation to determine the most suitable fit for a specific problem involves a systematic assessment of various models based on their inherent strengths, weaknesses, and compatibility with the given problem's characteristics. The multifaceted nature of data, problem type, and additional factors underscores the importance of comprehensive model evaluation during predictive model construction.

Comparing the performance of different models provides important insights into their suitability for a particular dataset and problem. Quantitative metrics like Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) enable numerical assessment of model accuracy. Additionally, exploring diverse model types can uncover subtle relationships and patterns within the data that would otherwise remain hidden. Systematic comparison elucidates the nuanced strengths of different techniques. It highlights opportunities for

ensemble modelling by combining complementary approaches. Overall, model evaluation and contrast aids judicious selection and optimization to extract maximum value from data.
A deep understanding of the diverse errors generated by different models offers an additional layer of information to enhance overall results. For instance, if a model consistently overpredicts outcomes, it becomes imperative to consider adjustments to the model's parameters or the collection of new data to rectify this bias. This process of scrutinizing error types fosters the enhancement of predictive accuracy.
Lastly, the act of comparing models extends its utility to contingencies wherein the primary model might falter or exhibit subpar performance when faced with unseen data. Should such a scenario arise, the existence of alternative models that have undergone evaluation provides the flexibility to switch to a different model, ensuring robust performance on new and unexplored data.

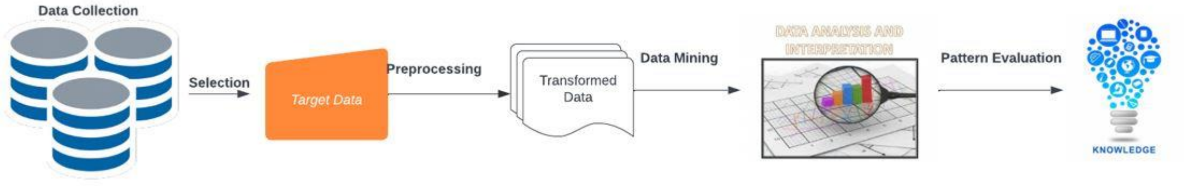

Figure 1 KDD Process

3.2 Data

For the research, I need the data for the crude oil price and economic indicators such as inflation, GDP (Gross Domestic Product), stock market index (BSE), and unemployment. I have collected the economic indicator and crude oil price data from open-source website Federal Reserve Economic Data (FRED). It is an online database consisting of millions of economic data time series from scores of national, international, public, and private sources. It is created and maintained by the Research Department at the Federal Reserve Bank of St. Louis. I have taken the 21 years historical monthly data for my analysis. And for unemployment I have taken 21 years historical annual data. I have collected the data from yahoo finance API. The Yahoo Finance API is a collection of libraries and methods designed to provide access to historical and real-time financial data across various markets and product. Please check the below table for more details.

Dataset	**Frequency**	**Format**	**Unit**
Global price of Dubai Crude[2]	Monthly	CSV	U.S. Dollars per Barrel
Inflation[3]	Monthly	CSV	Index
GDP[4]	Monthly	CSV	Index
BSE	Daily	Core Data frame	Index
Unemployment[5]	Yearly	CSV	Rate percentage
Global price of Dubai Crude	Yearly	CSV	U.S. Dollars per Barrel

[2] https://fred.stlouisfed.org/series/POILDUBUSDM
[3] https://fred.stlouisfed.org/series/INDCPIALLMINMEI
[4] https://fred.stlouisfed.org/series/INDLORSGPNOSTSAM
[5] https://fred.stlouisfed.org/tags/series?t=india%3Bunemployment

3.3 Data Preprocessing

The information is displayed in CSV (Comma-Separated Values) format. For several analytical contexts, such as the relationship between crude oil and the gross domestic product (GDP), inflation, the stock market, and unemployment, separate Jupyter Notebooks have been constructed for each dataset.

This separation aims to enhance comprehension and analysis. After data import, the pandas library is utilized to create data frames. The initial five records of each dataset are inspected. Subsequently, the column labelled 'POILDUBUSDM' is renamed as 'Crude_oil_price' within the crude oil data frame for improved interpretability.

Furthermore, the date column is standardized to an appropriate date format using the pandas `to_datetime` function. Statistical insights are derived from the data through the utilization of the `describe` function, offering measures such as mean, median, minimum, maximum, and quantiles. This comprehension of data facilitates the identification of potential outliers, a vital consideration as outliers can compromise model accuracy. A process is undertaken to identify and potentially remove these outliers. The values that deviate significantly from the mean and might be considered unusual or potentially erroneous. The threshold for identifying outliers is set at mean plus three times the standard deviation, which is a common approach in statistical analysis. We always don't remove the outliers if it is justifiable. Like in case of stock market index there is only one outlier, so it will not affect much, and we will not remove it.

For the Stock Market's BSE index, which exhibits daily frequency, the data is aggregated to a monthly interval by calculating the mean. Since the date column includes both date and time components, it is refined to only retain the date information.

3.4 EDA

For initial analysis, I plotted the data of stock market and crude oil to check the correlation. I have selected BSE index close column on y- axis and crude oil price on x- axis.

The scatter plot illustrates a positive correlation between BSE index and crude oil prices, implying the stock index tends to increase with rising oil prices and decrease when oil prices decline. However, the relationship is imperfect as evident by outliers deviating from the trend. While the plot indicates association between the two variables, the presence of deviations suggests factors beyond just crude prices also drive BSE fluctuations.

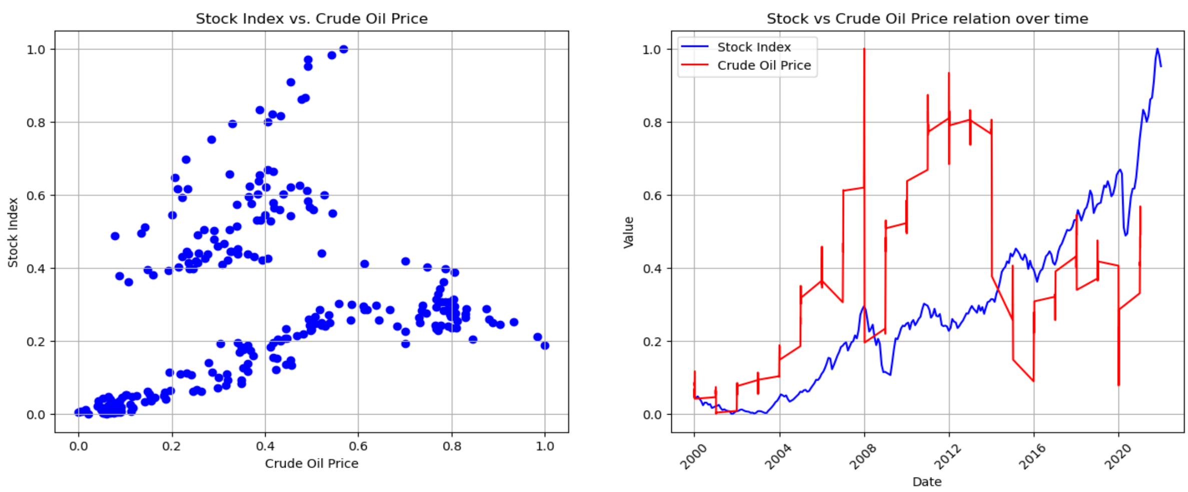

Figure 2 Correlation of stock market and crude oil prices

The line graph illustrates periods where the correlation between crude oil prices and BSE index weakened, notably during the 2008-2009 global financial crisis. The decoupling

suggests factors like the financial turmoil outweighed crude oil's influence on the index. While oil prices and Indian equities broadly move in tandem, world events can transiently disrupt this relationship. The plot highlights both proportionality and deviations in oil-stock correlation, underscoring contextual complexities. Though crude prices significantly sway BSE, equities comprise a nuanced system shaped by multifaceted global and domestic forces. The visualization provides perspective on real-world constraints governing the linkage between crude markets and Indian finance.
I have made a merged data frame which includes GDP rate from df_gdp data frame and crude oil price from crude oil data frame(df_crude). Then I checked for negative values. For initial analysis I plotted the scatter plot of GDP index versus crude oil price to see the correlation.

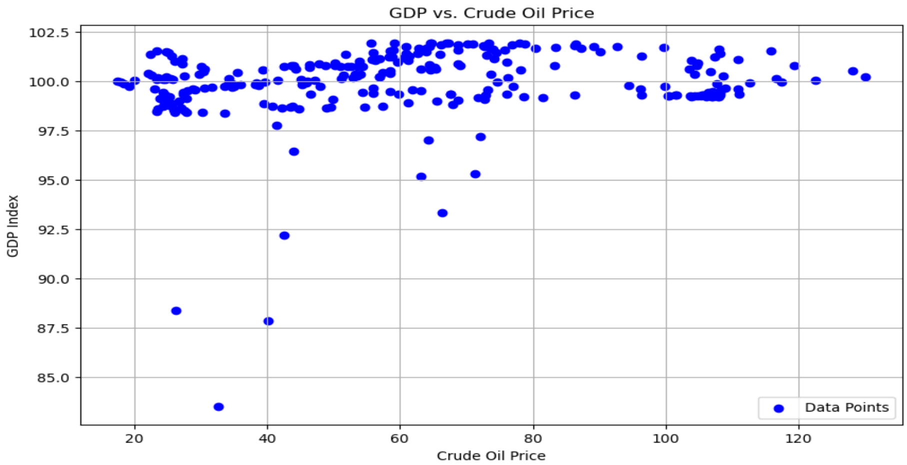

Figure 3 GDP and Crude Oil Price Scatter Plot

The visualization depicts the relationship between GDP index, representing national economic output, and crude oil prices. A proportional correlation is evident, implying gross domestic product tends to rise with increasing oil prices, and vice versa. However, deviations from perfect positive correlation indicate the presence of additional factors influencing macroeconomic cycles. While the two indices broadly align directionally, nuances exist in their correlation. The plot provides perspective into the connections between crude oil markets and aggregate demand-supply conditions but also underscores exercising prudence with causal extrapolations.

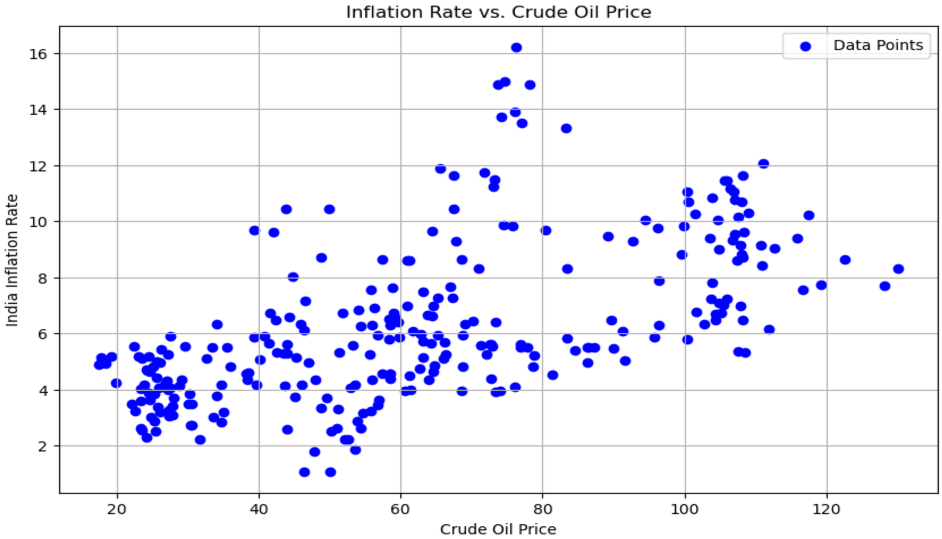

Figure 4 Correlation of inflation and crude oil price

The above scatter plot shows a positive relationship between inflation rates and crude oil prices in India. As oil prices increase, inflation also tends to go up. This makes sense because crude oil is used in many areas of the economy. When oil becomes more expensive, the costs of transportation, manufacturing, and other business activities also rise. This feeds into higher overall inflation. The plot shows correlation, but oil prices alone do not decide inflation rates. There are lots of other complicated reasons that make inflation go up and down. So in simple terms, the graph shows oil prices and inflation are related because oil affects production costs.

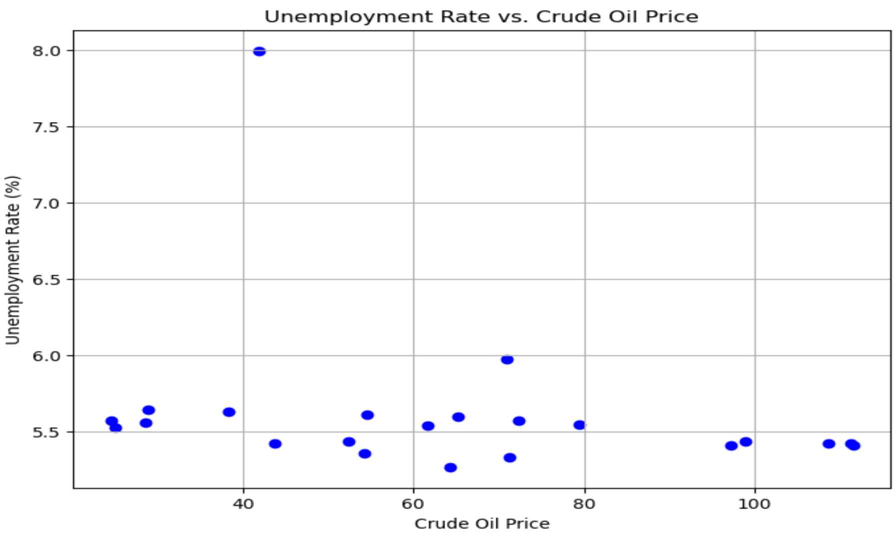

Figure 5 Relation between Crude oil price and Unemployment

The graph shows the relationship between the unemployment rate in India and the crude oil price. The data points show that there is a negative correlation between the two variables, meaning that when the unemployment rate increases, the crude oil price tends to decrease. However, the correlation is not perfect, and there are some data points that do not follow the trend. There are a few possible explanations for the negative correlation between the unemployment rate and the crude oil price. One possibility is that as the unemployment rate increases, there is less demand for oil, which drives down the price of crude oil. Another possibility is that as the unemployment rate increases, the government may take measures to stimulate the economy, such as increasing government spending or lowering interest rates, which can also lead to a decrease in the price of crude oil.

4 Design Specification

The research attempt is divided into four separate periods. In the initial stage, data is extracted from diverse sources, and further data preprocessing is conducted via Python programming aided by the pandas module. The preparation phase comprises analysis for discrepancies in data, identification of null values, and identifying of probable outliers. The processed data is then grouped into pandas data frames. The following important phase comprises the vital pursuit of Exploratory Data Analysis (EDA), wherein an attempt is made to identify underlying patterns natural to the dataset by applying basic graphical displays.
Transitioning onto the third step, the dataset is partitioned into training and test subsets, thus clearing the route for model building. The key point revolves around training machine learning models applying the training data subset. This training process is controlled within the machine learning framework, integrating models with the ability to recognize patterns and infer relationships from the data. The final phase consummates the project through extensive examination of the model's prediction efficiency. This evaluation covers diverse statistical tests, so developing a full knowledge of the model's performance and its agreement with the underlying data distribution.

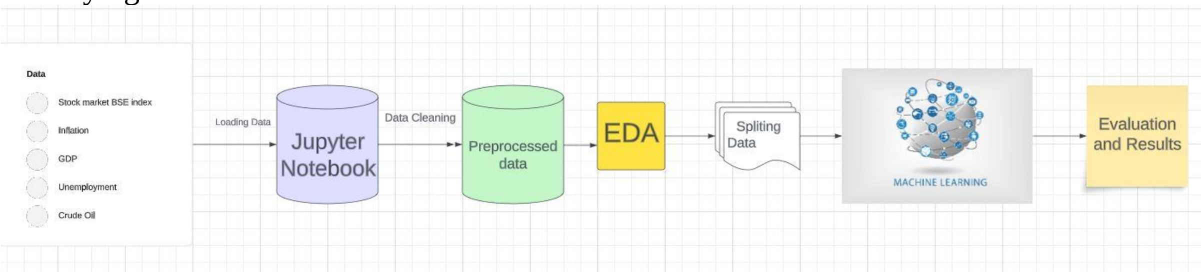

Figure 6 Design Specification

4.1 Algorithms

Linear regression stands as a foundational statistical method employed to model the linear interrelation between a feature variable and one or multiple target variables. This technique calculates coefficients to minimize the disparity between anticipated and observed values. When building numerous decision trees during training, the random forest ensemble supervised learning approach produces the class or mean prediction of the individual trees. In comparison to a single decision tree, it eliminates overfitting and increases accuracy.
Gradient boosting produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a staged fashion, iteratively improving the model by minimizing the loss function.
Gaussian process regression is a non-parametric Bayesian approach useful in modelling nonlinear relationships. It defines a probability distribution over possible functions fitting the data to make predictions along with uncertainty estimates. It works well with less dataset as well.
Neural networks are computational systems drawing inspiration from biological neural networks, proficient in recognizing intricate nonlinear associations between input and output data. They are structured in layers which transform input data through weighted connections to produce the output.
In summary, these techniques encompass both statistical and machine learning algorithms that can uncover and model different types of relationships between variables. Their nuanced

strengths can be leveraged based on the problem domain, data characteristics and desired interpretability.

5 Implementation

The effects of crude oil prices on macroeconomic variables in India are examined in this study using an ensemble modelling methodology. While more sophisticated methods like random forest, gradient boosting, and neural networks provide deeper examination of feature correlations, linear regression offers a baseline analysis. The variety of models tries to increase forecasting precision and clarify the nuanced relationships between crude oil volatility and important economic indicators. This approach pursues strong, data-driven insights into the mechanisms tying oil market swings with the real economy by utilizing the advantages of both statistical and machine learning methodologies. The ensemble approach extracts complex economic insights while improving model performance.

5.1 Software and Tools

To develop the outputs and the outcomes, the following prerequisites for the software and libraries have been used:
• Programming Language - Python
• IDE – Jupyter Notebook
• Python Libraries/Modules:
Pre-processing- Numpy, Pandas
Feature Engineering - matplotlib, Plotly
Modelling and Evaluation - Sklearn, tensorflow, GPY

5.2 Hardware

Processor: Intel(R) Core (TM) i5-8265U CPU @ 1.60GHz 1.80 GHz
RAM: 8.00 GB
Operating System: 64-bit operating system, x64-based processor
Device: Lenovo IdeaPad S145

5.3 Model Training for Crude Oil, BSE Indian Stock Index, Inflation, and GDP

After getting the cleaned data, we convert the data into the range of 0 to 1 using MinMaxScaler class from sklearn.preprocessing module This specialized scaler serves the purpose of transforming data to conform to a specific range, often spanning from 0 to 1.
The fit_transform method associated with the instance is employed. This function normalizes the column's values by reshaping them into a two-dimensional array using values.reshape(-1, 1), and then executing the scaling transformation. The resultant normalized values are subsequently reintegrated into the respective column of the Data Frame. Similarly, the fit_transform method linked to the respective instance is utilized to normalize the column within the Data Frame.

Normalization is valuable for bringing disparate features or columns to a consistent scale. This proves particularly advantageous for machine learning algorithms reliant on distance calculations or gradient descent optimization, as it enhances algorithm convergence and performance. The process of scaling values to a uniform range mitigates the potential dominance of a single feature due to its larger scale, contributing to more balanced and effective analyses.

The implementation begins by identifying the target variable (stock price) and input features (crude oil price). The data is split into training (80%) and test (20%) subsets using train_test_split to ensure rigorous evaluation on new data. The input features are reshaped into two-dimensional arrays for compatibility with machine learning models.

Hyperparameter optimization is performed using randomized grid search and cross-validation. This tunes model configurations like number of trees and tree depth for Random Forest, learning rate and loss function for Gradient Boosting. Grid search systematically evaluates combinations to determine ideal hyperparameters for maximum predictive accuracy.

Linear Regression provides a baseline model. A pipeline enables staged data processing and model building. Cross-validation identifies the optimal intercept parameter for best performance. This hyperparameter tuning refines the Linear Regression model.

For Neural Networks, a sequential model with fully connected layers is constructed. The input layer size matches the feature dimensionality. Hidden layers with ReLU activation generate interpretable representations. The output layer predicts the stock price. Model compilation and training on mean squared error loss tunes the network.

The optimized models are evaluated on the test set. Predicting unseen data and comparing it with actual values assesses generalizability and accuracy. This rigorous validation ensures models uncover meaningful patterns without overfitting and can effectively predict future data.

In summary, systematic hyperparameter optimization, structured validation, and controlled testing enable thorough and unbiased evaluation of model performance. This implementation provides a framework for data-driven analysis of crude oil's predictive relationship with Indian equities.

5.4 Model Training for Crude Oil vs Unemployment

I have selected target variable Unemployment rate as X and feature variable crude oil price as Y. The code first creates a radial basis function (RBF) kernel to model the relationships between data points in the Gaussian Process model. It then initializes a GPRegression instance using the GPy library, passing the input Crude Oil Price data, target Unemployment Rate data, and the RBF kernel. This sets up the model architecture. Finally, the optimize () method fine-tunes the kernel hyperparameters to minimize the prediction error and improve model fitting. In essence, the code leverages the RBF kernel flexibility, the GPRegression model, and hyperparameter optimization to train a Gaussian Process model that captures the intricacies between Crude Oil Prices as input and Unemployment Rates as target, to make accurate predictions.

Then the code generates predictions for the Unemployment Rate based on Crude Oil Prices using the trained Gaussian Process Regression model. It creates a testing set by sampling 20% of the data, selects the corresponding target values, and then predicts the mean Unemployment Rate and its associated variance for the testing set using the model.

6 Evaluation and Results

Following model training and evaluation, each prediction model's effectiveness is thoroughly examined. Researchers commonly use evaluation metrics like Mean Squared Error, Root Mean Squared Error, and Mean Absolute Error to assess model accuracy in regression and predictive modelling. MSE quantifies total error magnitude, penalizing larger errors due to squaring. RMSE expresses average error in original data units, enhancing interpretability. MAE calculates average absolute error, equally weighting all discrepancies. These metrics provide complementary insights into predictive performance - MSE emphasizes minimizing outliers, RMSE conveys errors in real-world terms, and MAE treats all mistakes equally (Plevris et al., 2022). Their extensive analysis in literature and ease of communication to both technical and non-technical audiences underpin their popularity. In summary, the well-rounded assessment, interpretability, and extensively documented characteristics make MSE, RMSE and MAE preferred choices for evaluating predictive models.

.

6.1 Experiment 1

I am applying all the four models on crude oil and stock market index data. With the lowest Mean Absolute Error (0.155), Root Mean Squared Error (0.192), and Mean Squared Error (0.036), the Neural Network model demonstrates superior accuracy over other techniques in predicting stock markets based on crude oil prices. The minimal errors across all metrics unanimously indicate the Neural Network's effectiveness in modelling this relationship.

The Random Forest and Gradient Boosting models also showed good performance with slightly higher errors compared to Neural Network.

The Linear Regression model had the worst performance with the highest errors - MAE of 0.208, RMSE of 0.237 and MSE of 0.056.

This indicates that the Non-linear models like Neural Networks, Random Forest and Gradient Boosting were better able to capture the relationship between crude oil prices and stock market compared to the linear Regression model.

Among the non-linear models, the Neural Network model had an edge over Random Forest and Gradient Boosting in terms of minimizing the magnitude of errors in predicting the stock market based on crude price data.

Model	**MAE**	**RMSE**	**MSE**
Linear Regression	0.208	0.237	0.056
Random Forest Regression	0.151	0.204	0.041
Gradient Boosting Regression	0.160	0.216	0.046
Neural Networks	0.155	0.192	0.036

Overall, the below graph shows that the neural network model is the most accurate model for predicting the stock index value. The random forest model is also a good option, followed by the gradient boosting model and the linear regression model.

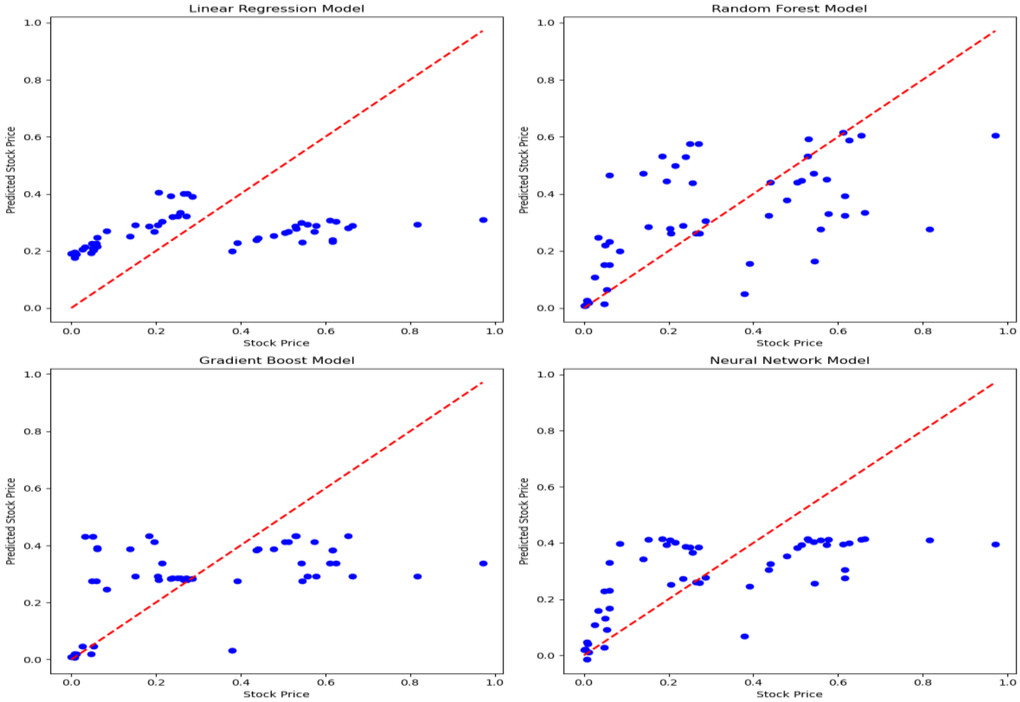

Figure 7 BSE Index prediction result

6.2 Experiment 2

In the experiment 2, I am evaluating the models which got trained on crude oil and GDP data for the prediction of GDP. The Gradient Boosting model, with the lowest Mean Absolute Error (0.061), Root Mean Squared Error (0.118) and Mean Squared Error (0.014), demonstrates the highest predictive accuracy for GDP based on crude oil prices. The minimized errors across all metrics indicate Gradient Boosting's effectiveness over Linear Regression, Neural Networks and Random Forest models in capturing the intricate oil price-GDP relationship.

Model	**MAE**	**RMSE**	**MSE**
Linear Regression	0.070	0.119	0.014
Random Forest Regression	0.076	0.126	0.016
Gradient Boosting Regression	0.061	0.118	0.014
Neural Networks	0.068	0.119	0.014

From the below graph also, we can see that the Gradient Boosting model is the most accurate.

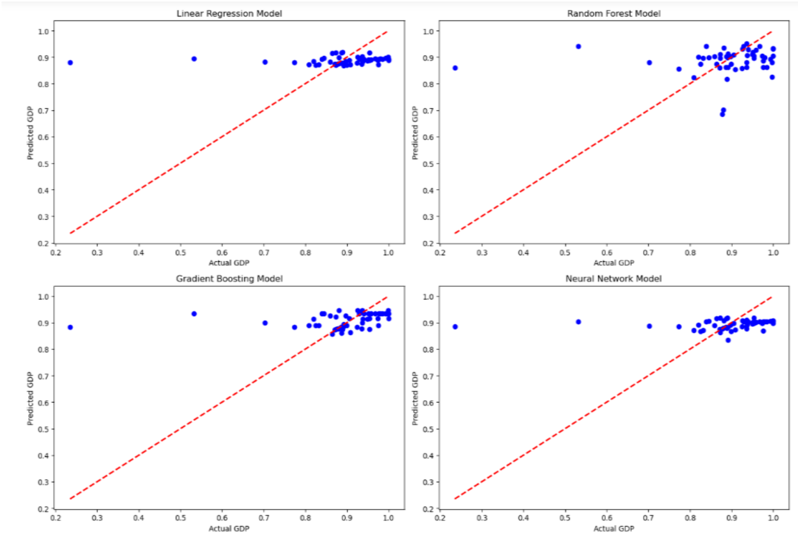

Figure 8 Prediction of GDP by various models

6.3 Experiment 3

I am evaluating models which got trained for prediction of inflation index based on crude oil price. The Linear Regression (MAE 0.107) and Neural Network (RMSE 0.135, MSE 0.018) models demonstrate the highest accuracy in predicting inflation based on crude oil prices, with minimal errors across key metrics. In comparison, Random Forest (MAE 0.110, RMSE 0.144, MSE 0.020) performs moderately while Gradient Boosting (MAE 0.122, RMSE 0.157, MSE 0.024) exhibits the largest deviations. The low error values confirm Linear Regression and Neural Networks' effectiveness in modelling the linkage between crude prices and inflation. From the below graph also, we can see neural model predictions is most accurate.

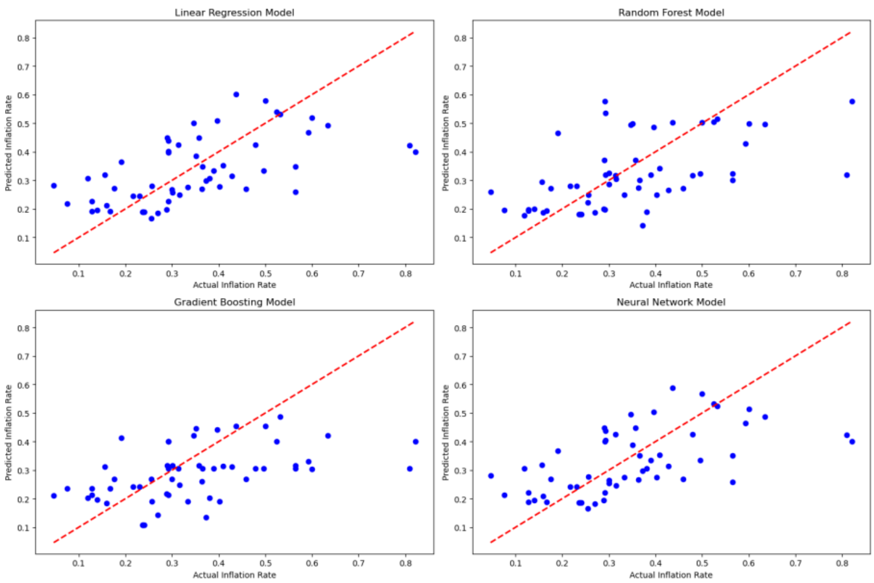

Figure 9 Inflation prediction result

Model	**MAE**	**RMSE**	**MSE**
Linear Regression	0.107	0.135	0.018
Random Forest Regression	0.111	0.144	0.020

Gradient Boosting Regression	0.122	0.157	0.024
Neural Networks	0.107	0.135	0.018

6.4 Experiment 4

In this experiment, I am evaluating the Gaussian Process Regression which got trained on unemployment and crude oil data with yearly frequency.
With low Mean Absolute Error (0.103), Root Mean Squared Error (0.130) and Mean Squared Error (0.017), the Gaussian Process Regression model demonstrates high accuracy in predicting unemployment from crude oil prices. The minimal errors indicate the model effectively captures the relationship, providing reliable unemployment rate forecasts based on crude price movements.
The below graph depicts a negative correlation between crude oil prices and unemployment rate, with higher oil prices associated with lower unemployment. This inverse relationship stems from oil price hikes reducing economic activity and jobs. However, uncertainty exists in the linkage as shown by the shaded range around the trendline, representing variance in possible unemployment rates for a given oil price. Overall, while increased oil costs tend to diminish employment, multifaceted real-world factors introduce volatility in the oil-unemployment relationship.

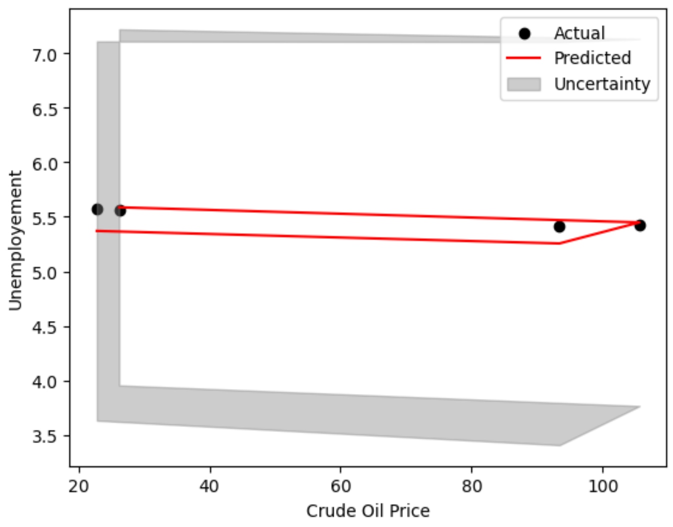

Figure 10 Unemployment Prediction

6.5 Discussion

In this study, the intricate relationship between crude oil prices and diverse Indian economic indicators is scrutinized using various machine learning models, including Linear Regression, Random Forest Regression, Gradient Boosting Regression, and Neural Networks. The outcomes offer valuable insights into this intricate interplay.

A recent investigation by (Sen, Dutta Choudhury, and Kumar Datta, 2023) employs deep learning to analyse a decade's worth of crude oil price data (2011-2020). Their approach delves into the impact of crude oil inventories and introduces financial instruments to comprehend their link with price fluctuations. Long Short-Term Memory (LSTM) techniques, highlighting deep learning's potential, are employed.

Similarly, (Daneshvar et al., 2022) leverage deep neural network architectures, particularly LSTM and Bi-LSTM models, to forecast Brent crude oil prices. Their findings align with the LSTM approach in the present study, reinforcing the potential of LSTM-based models for price prediction.
However, limitations exist. The study's temporal scope is relatively short, possibly constraining generalizability. Furthermore, external factors like pandemics and government policies have not been fully considered. Including variables such as exchange rates, interest rates, and geopolitical dynamics, and adopting advanced models like transformers, could enhance the study's efficiency.

7 Conclusion and Future Work

In this study, the intricate relationship between crude oil prices and the Indian economy has been meticulously examined through the application of various predictive models. The aim was to comprehend the impact of crude oil price fluctuations on key economic indicators, namely the Stock Market BSE Index, GDP Index, Inflation, and Unemployment. The results obtained from the different models have provided valuable insights into the dynamics of this complex interaction. The analysis of the crude oil and Stock Market BSE Index revealed that all models, including Linear Regression, Random Forest Regression, Gradient Boosting Regression, and Neural Networks, exhibited varying degrees of predictive accuracy. Among these, the Neural Networks demonstrated the lowest Mean Squared Error (MSE) of 0.036, Mean Absolute Error (MAE) of 0.151, indicating its superior predictive capability compared to the other models. This suggests that the stock market is significantly influenced by changes in crude oil prices, and the Neural Networks model stands out as a robust tool for such prediction tasks.
For the relationship between crude oil prices and the GDP Index, the Gradient Boosting Regression model emerged as the most effective, boasting a remarkably low MAE of 0.061. This finding suggests that the GDP of the Indian economy is notably affected by fluctuations in crude oil prices, and the Gradient Boosting model showcased its proficiency in capturing the intricate patterns within this relationship.
Furthermore, the study explored the correlation between crude oil prices and inflation. Among the models, Neural Networks and Linear Regression displayed consistent performance, both yielding a MAE of 0.107. These results underscore the significant influence of crude oil prices on inflation rates, and the stability of the Neural Networks and Linear Regression models in predicting this relationship.
The analysis of crude oil and unemployment employed the Gaussian Process Regression model, which returned a Mean Absolute Error of 0.103. The study reveals a reciprocal relationship between crude oil price fluctuations and Indian unemployment rates. It suggests that an increase in the unemployment rate corresponds with a decrease in crude oil prices. This alignment underscores the vulnerability of labor markets to external energy price shocks. This study's substantial progress in comprehending the impact of crude oil price fluctuations on the Indian economy is accompanied by avenues for future research. These encompass the exploration of microeconomic implications, analysis of external factors such as geopolitical events and technological advancements, utilization of advanced machine learning models, and the translation of insights into actionable policy recommendations. While this study enriches understanding, further research can refine predictions and provide comprehensive guidance for policymakers, businesses, and investors navigating the intricate Indian economic landscape amid volatile crude oil markets.

References

Ahmad, I. *et al.* (2022) ‘Macroeconomic effects of crude oil shocks: Evidence from South Asian countries’, *Frontiers in Psychology*, 13. Available at: https://doi.org/10.3389/fpsyg.2022.967643.

Charfeddine, L. and Barkat, K. (2020) ‘Short- and long-run asymmetric effect of oil prices and oil and gas revenues on the real GDP and economic diversification in oil-dependent economy’, *Energy Economics*, 86. Available at: https://doi.org/10.1016/j.eneco.2020.104680.

Daneshvar, A. *et al.* (2022) ‘Brent Crude Oil Price Forecast Utilizing Deep Neural Networchak Architectures’, *Computational Intelligence and Neuroscience*, pp. 1–13. Available at: https://doi.org/10.1155/2022/6140796.

Diaz, E.M., Molero, J.C. and Perez de Gracia, F. (2016) ‘Oil price volatility and stock returns in the G7 economies’, *Energy Economics*, 54, pp. 417–430. Available at: https://doi.org/10.1016/j.eneco.2016.01.002.

Dong, M. *et al.* (2019) ‘Revisiting global economic activity and crude oil prices: A wavelet analysis’, *Economic Modelling*, 78, pp. 134–149. Available at: https://doi.org/10.1016/j.econmod.2018.08.012.

van Eyden, R. *et al.* (2019) ‘Oil price volatility and economic growth: Evidence from advanced economies using more than a century’s data’, *Applied Energy*, 233–234, pp. 612–621. Available at: https://doi.org/10.1016/j.apenergy.2018.10.049.

Ghosh, S. and Kanjilal, K. (2016) ‘Co-movement of international crude oil price and Indian stock market: Evidences from nonlinear cointegration tests’, *Energy Economics*, 53, pp. 111–117. Available at: https://doi.org/10.1016/j.eneco.2014.11.002.

Jain, A. and Biswal, P.C. (2016) ‘Dynamic linkages among oil price, gold price, exchange rate, and stock market in India’, *Resources Policy*, 49, pp. 179–185. Available at: https://doi.org/10.1016/j.resourpol.2016.06.001.

Lyu, Y. *et al.* (2021) ‘Time-varying effects of global economic policy uncertainty shocks on crude oil price volatility: New evidence’, *Resources Policy*, 70. Available at: https://doi.org/10.1016/j.resourpol.2020.101943.

Mandal, K., Bhattacharyya, I. and Bhoi, B.B. (2012) ‘Is the oil price pass-through in India any different?’, *Journal of Policy Modeling*, 34(6), pp. 832–848. Available at: https://doi.org/10.1016/j.jpolmod.2012.06.001.

Masih, R., Peters, S. and De Mello, L. (2011) ‘Oil price volatility and stock price fluctuations in an emerging market: Evidence from South Korea’, *Energy Economics*, 33(5), pp. 975–986. Available at: https://doi.org/10.1016/j.eneco.2011.03.015.

Masood, T. and Shahzeb, S.M. (2023) ‘Exploring Macroeconomic Implications of Oil Price Fluctuations in India’, *The Indian Economic Journal*, 71(4), pp. 689–709. Available at: https://doi.org/10.1177/00194662221139353.

Mohanty, D. and John, J. (2015) ‘Determinants of inflation in India’, *Journal of Asian Economics*, 36, pp. 86–96. Available at: https://doi.org/10.1016/j.asieco.2014.08.002.

Narayan, P.K. and Narayan, S. (2007) ‘Modelling oil price volatility’, *Energy Policy*, 35(12), pp. 6549–6553. Available at: https://doi.org/10.1016/j.enpol.2007.07.020.

Periwal, A. (2023) ‘The Impact of Crude Oil Price Fluctuations on Indian Economy’, *International Journal for Research in Applied Science and Engineering Technology*, 11(4), pp. 3173–3202. Available at: https://doi.org/10.22214/ijraset.2023.50851.

Qianqian, Z. (2011) ‘The Impact of International Oil Price Fluctuation on China’s Economy’, *Energy Procedia*, 5, pp. 1360–1364. Available at: https://doi.org/10.1016/j.egypro.2011.03.235.

Rafiq, S., Salim, R. and Bloch, H. (2009) 'Impact of crude oil price volatility on economic activities: An empirical investigation in the Thai economy', *Resources Policy*, 34(3), pp. 121–132. Available at: https://doi.org/10.1016/j.resourpol.2008.09.001.
Rahman, S. (2022) 'The asymmetric effects of oil price shocks on the U.S. stock market', *Energy Economics*, 105. Available at: https://doi.org/10.1016/j.eneco.2021.105694.
Sen, A., Dutta Choudhury, K. and Kumar Datta, T. (2023) 'An analysis of crude oil prices in the last decade (2011-2020): With deep learning approach', *PLOS ONE*, 18(3), p. e0268996. Available at: https://doi.org/10.1371/journal.pone.0268996.
Shao, M. and Hua, Y. (2022) 'Price discovery efficiency of China's crude oil futures: Evidence from the Shanghai crude oil futures market', *Energy Economics*, 112, p. 106172. Available at: https://doi.org/10.1016/j.eneco.2022.106172.
Su, C.-W. *et al.* (2021) 'Does crude oil price stimulate economic policy uncertainty in BRICS?', *Pacific-Basin Finance Journal*, 66. Available at: https://doi.org/10.1016/j.pacfin.2021.101519.
Xiuzhen, X., Zheng, W. and Umair, M. (2022) 'Testing the fluctuations of oil resource price volatility: A hurdle for economic recovery', *Resources Policy* [Preprint]. Available at: https://doi.org/10.1016/j.resourpol.2022.102982.