CRUDE OIL PRICE PREDICTION

Project Report

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

in

Electrical Engineering

by

ADARSH AGRAWAL (B200272EE)
RAJAN YADAV (B200297EE)

तमसो मा ज्योतिर्गमय

Department of Electrical Engineering

NATIONAL INSTITUTE OF TECHNOLOGY CALICUT

September 2023

CERTIFICATE

This is to certify that the report entitled "**CRUDE OIL PRICE PREDICTION**" is a bonafide record of the **Project** done by **ADARSH AGRAWAL** (*Roll No.*: **B200272EE**) and **RAJAN YADAV** (Roll No**.: B200297EE**) under my supervision, in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Electrical Engineering** from **National Institute of Technology Calicut**, and this work has not been submitted elsewhere for the award of a degree.

Dr. JEEVAMMA JACOB
(Guide)
Professor
Dept. of Electrical Engineering

Dr. PREETHA P.
HOD, Dept. of Electrical Engineering

Place: NIT Calicut
Date: 08-December-2023

ABSTRACT

This report tells about crude oil price prediction and its profound implications for global financial markets, energy economics, and policy formulation. It explores the application of deep learning techniques to improve the accuracy of crude oil price forecasts. Drawing from an extensive literature survey, the study traces the evolution of predictive models, emphasizing the transition from traditional econometric methods to data-driven deep learning approaches. It underscores the critical role of selecting and preprocessing high-quality datasets, encompassing historical price data, macroeconomic indicators, geopolitical events, and environmental factors. The report comprehensively explores various deep learning architectures, including statistical models, auto-regressive integrated moving average (ARIMA), long short-term memory (LSTM) networks, convolutional neural networks (CNNs), and hybrid models, along with innovative techniques like attention mechanisms and reinforcement learning. Crucially, the results demonstrate that deep learning consistently outperforms traditional methods, reducing mean absolute error (MAE) and mean squared error (MSE), highlighting its ability to capture intricate relationships between crude oil price movements and influencing factors. In conclusion, this report provides a comprehensive overview of the state-of-the-art in crude oil price prediction using deep learning, aiming to equip decision-makers with precise forecasts for informed actions in the dynamic energy sector and financial markets.

CONTENTS

LIST OF ABBREVIATIONS

ARIMA	Auto Regressive Integrated Moving Average
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
CNN	Convolutional Neural Network

CHAPTER 1

INTRODUCTION

Predicting crude oil prices is a complex and crucial task with far-reaching implications for economies and industries worldwide. Crude oil, often referred to as "black gold," is one of the most crucial and volatile commodities in the global economy. Its price fluctuations have far-reaching consequences, impacting industries, governments, and the daily lives of individuals. In an era defined by global energy demands, geopolitical tensions, and environmental concerns, the ability to forecast crude oil prices accurately holds immense value. Crude oil is not just a commodity; it's a cornerstone of modern civilization, driving transportation, energy generation, and various industrial processes. These price fluctuations are influenced by a multitude of factors, such as supply and demand dynamics, geopolitical events, weather conditions, and economic indicators. Predicting these price movements is like deciphering a puzzle with countless interlocking pieces, making it a prime candidate for a multimodal prediction system.

Traditionally, experts in finance and energy have relied on a combination of fundamental analysis, technical indicators, and market sentiment to forecast crude oil prices. However, with the advent of machine learning (ML) and deep learning (DL) techniques, there is an opportunity to leverage data-driven approaches to enhance the accuracy of these predictions.

The ARIMA model, an acronym for Auto Regressive Integrated Moving Average, is a traditional time-series analysis technique that has been widely used for decades to forecast financial and economic data, including crude oil prices. ARIMA combines autoregressive (AR) and moving average (MA) components with differencing to make time series data stationary. This approach helps capture underlying trends and seasonality in the crude oil price data, providing valuable insights for short to medium-term predictions.

Statistical data analysis plays a pivotal role in understanding the broader economic and geopolitical context surrounding crude oil prices. Factors like interest rates, inflation, GDP growth, and political stability can significantly impact oil markets. By incorporating these variables into predictive models, analysts can gain a comprehensive view of the forces at play and their potential effects on crude oil prices.

In recent years, the advent of deep learning has revolutionized the field of time-series forecasting. LSTM networks, a type of recurrent neural network (RNN), have emerged as a powerful tool for capturing long-term dependencies in sequential data. When applied to crude oil price prediction, LSTM models can uncover complex patterns and relationships that may not be evident through traditional statistical methods. Their ability to process vast amounts of historical price data makes them well-suited for both short-term and long-term forecasting.

On the other hand, CNNs, primarily known for their prowess in image recognition, have also found utility in the domain of crude oil price prediction. By transforming price data into image-like representations, CNNs can uncover spatial patterns and anomalies that may be crucial for predicting market movements. This novel approach showcases the versatility of deep learning techniques when applied to unconventional data types.

Combining these various models and techniques into a multimodal system provides a holistic approach to crude oil price prediction. It allows analysts to harness the strengths of each method, mitigating the limitations of individual models. The synergy of ARIMA, statistical data analysis, LSTM, and CNN in a cohesive system empowers analysts to make more informed decisions, whether in financial trading, energy planning, or policy formulation.

In conclusion, predicting crude oil prices is a complex endeavor that demands a multidisciplinary approach. Leveraging the power of ARIMA, statistical data analysis, LSTM, and CNN in a multimodal system represents a cutting-edge strategy to tackle this challenge. As the global energy landscape continues to evolve, the ability to forecast crude oil prices accurately becomes increasingly essential for businesses, governments, and individuals alike. The pursuit of more accurate predictions is not merely an academic exercise but a critical element in navigating the volatile and ever-changing world of energy markets.

CHAPTER 2
MOTIVATION

The motivation behind undertaking this project is driven by several key factors:

1. **Economic Impact:** Crude oil prices have a profound influence on the global economy. Sudden price fluctuations can disrupt financial markets, impact inflation rates, and have cascading effects on various industries. Accurate price predictions can assist businesses and governments in proactively addressing economic challenges.
2. **Risk Management:** Businesses operating in energy-dependent sectors, such as transportation, manufacturing, and agriculture, face significant risk exposure to crude oil price movements. Developing reliable prediction models can help these industries hedge against price volatility and optimize their operations.
3. **Investment Decisions:** Investors and financial institutions often include commodity i.e. like crude oil in their portfolios. Having access to accurate price forecasts can aid investors in making informed asset allocation and trading decisions.
4. **Energy Policies:** Governments formulate energy policies and taxation strategies based on crude oil price expectations. Reliable predictions can guide policymakers in creating stable and sustainable energy policies.
5. **Technological Advancements:** Recent advancements in machine learning and data analytics have opened up new possibilities for improving the accuracy of crude oil price predictions. Leveraging these technologies can result in more sophisticated models and better forecasting capabilities.
6. **Environmental Impact:** Sustainable energy transition and environmental concerns are increasingly affecting the energy sector. Predicting oil price trends can help in planning and implementing cleaner and more efficient energy solutions.
7. **Business Strategy:** Companies involved in crude oil exploration, production, and distribution requires precise market insights to make strategic decisions regarding investments, production levels, and pricing strategies.
8. **Research and Innovation:** The field of predictive analytics and machine learning is continuously evolving. Conducting research in crude oil price prediction contributes to the advancement of these methodologies and their application in real-world scenarios.

In light of these motivations, this project aims to harness the power of data-driven techniques to develop a sophisticated and accurate crude oil price prediction model. By doing so, it seeks to address the challenges and opportunities associated with the dynamic and globally significant crude oil market.

CHAPTER 3

LITERATURE SURVEY

Crude oil price prediction is a critical and challenging task with far-reaching implications for various sectors, including global financial markets, energy economics, and policy formulation. To gain insights into the evolving landscape of crude oil price forecasting, this literature survey delves into a wide range of research articles and academic papers. It seeks to provide a comprehensive overview of the methodologies, models, and techniques employed in the pursuit of more accurate predictions. This survey explores traditional econometric models, machine learning approaches, and deep learning methods, shedding light on their strengths, weaknesses, and applicability. Additionally, it examines the selection of relevant datasets, feature engineering strategies, and the emergence of innovative techniques in this field. Through this literature survey, we aim to gain a deeper understanding of the state-of-the-art in crude oil price prediction and its significance in guiding decision-making processes in an ever-dynamic energy and financial landscape.

3.1 The ARIMA (Auto Regressive Integrated Moving Average) model is a widely used time series forecasting method, and it has also found application in predicting crude oil prices. Here's an overview of how the ARIMA model works and its application in crude oil price prediction:

3.1.1 ARIMA Model Basics:

ARIMA is a combination of three components:

1. **Auto Regressive (AR) -** The model considers the relationship between the current data point and past data points in a time series.

2. **Integrated (I) -** This component accounts for differencing the time series data to make it stationary, i.e., removing trends and seasonality.

3. **Moving Average (MA) -** It involves modeling the relationship between the current data point and past prediction errors.

3.1.2 Application of ARIMA Model in Crude Oil Price Prediction:

1. **Data Preparation:** The first step in applying an ARIMA model to predict crude oil prices is to gather historical price data. This dataset should ideally be stationary, but if it's not, differencing can be applied to remove trends and seasonality.

2. **Model Identification:** This step involves determining the appropriate order of the ARIMA

model, typically denoted as (p, d, q), where:

- p (Auto Regressive Order): It represents the number of past data points to include in the model.

- d (Integrated Order): It indicates the number of differencing required to make the time series stationary.

- q (Moving Average Order): It specifies the number of past forecast errors to include in the model.

This order can be identified through data analysis, autocorrelation plots, and partial autocorrelation plots.

3. **Model Estimation:** Once the order (p, d, q) is determined, the ARIMA model is estimated using historical data. The model aims to capture the relationships between past observations and forecast future values.

4. **Model Evaluation:** The performance of the ARIMA model is assessed using various statistical measures, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Additionally, visual inspection of the model's fit to the data can provide insights into its accuracy.

5. **Forecasting**: With the ARIMA model in place, it can be used to make future predictions of crude oil prices. These predictions can assist decision-makers in financial markets, energy companies, and government agencies in making informed decisions and policy formulation.

6. **Model Refinement:** ARIMA models may require periodic updates and refinements as new data becomes available. This ensures that the model remains accurate in a dynamic crude oil market.

7. **Risk Assessment:** Crude oil price predictions using ARIMA models should be accompanied by risk assessments, taking into account external factors like geopolitical events, environmental factors, and economic indicators, which can significantly impact oil prices.

In conclusion, the ARIMA model is a valuable tool for predicting crude oil prices by capturing time series patterns and trends. However, it should be used in conjunction with other sources of information and risk assessments for more robust and accurate predictions in the volatile world of crude oil markets.

3.2 Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) that has proven effective in various time series forecasting tasks, including crude oil price prediction. Here's an overview of how LSTM models work and their application in predicting crude oil prices:

3.2.1 LSTM Model Basics:

LSTM is a deep learning architecture designed to capture and learn patterns and dependencies in sequential data, making it well-suited for time series forecasting. Key features of LSTM include its ability to remember and use past information over long sequences, as well as its resistance to the vanishing gradient problem, which can hinder training in traditional RNNs.

3.2.2 Application of LSTM Model in Crude Oil Price Prediction:

1**. Data Preparation:** To apply an LSTM model for crude oil price prediction, historical price data is collected and prepared as a time series dataset. It is crucial to format the data in a way that the model can learn and make predictions effectively.

2. **Sequence Length and Features:** Sequences of historical price data are created with a specified length (e.g., past 30 days of prices) as input features for the LSTM model. Other relevant features, such as macroeconomic indicators, geopolitical events, or environmental factors, can also be included to improve prediction accuracy.

3. **Normalization and Scaling:** Data preprocessing steps like normalization and scaling are often applied to ensure that all input features have the same scale and that the model can learn effectively.

4. **Model Architecture:** The LSTM model is designed with input layers, LSTM layers, and output layers. The LSTM layers are essential for capturing temporal dependencies in the data. The number of LSTM layers and the number of units in each layer can be adjusted based on the complexity of the data and the problem.

5. **Training:** The LSTM model is trained using historical data, where it learns the patterns and relationships between past price sequences and future prices. The model uses back propagation through time to adjust its weights during training.

6. **Validation and Testing:** The model's performance is evaluated on validation data and test data to assess its accuracy and generalization capabilities. Common evaluation metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

7. **Hyper parameter Tuning:** Fine-tuning the LSTM model may involve adjusting hyper parameters such as the learning rate, batch size, and the number of epochs to optimize its performance.

8. **Prediction:** Once the LSTM model is trained and validated, it can be used to make future predictions of crude oil prices. These predictions can assist in decision-making processes

related to trading, investment, or policy formulation.

9. **Monitoring and Updating:** LSTM models may require periodic updates to adapt to changing market conditions and incorporate new data. Continuous monitoring and retraining are essential for maintaining accuracy.

In conclusion, LSTM models offer a powerful approach to predicting crude oil prices by leveraging their ability to capture complex temporal patterns. When combined with relevant external factors and effective data preprocessing, LSTM models can provide valuable insights and forecasts for stakeholders in the energy sector and financial markets. However, they should be used in conjunction with other analytical tools and risk assessments to make informed decisions in the dynamic crude oil market.

Some related works are mentioned in the form of table no: 1

Work Accomplished:

In the course of this project, extensive groundwork has been laid to comprehensively understand the landscape of crude oil price prediction. We have diligently conducted a series of studies and analyses to build a strong foundation for our research. The key achievements and undertakings include:

1. Literature Review and Background Research: We commenced our project by conducting an exhaustive review of the existing literature on crude oil price prediction. This phase involved scrutinizing academic publications, industry reports, and historical data to gain insights into the challenges, methodologies, and trends in this field**.**

2. Statistical Models and Techniques: Classical statistical methods were examined, including regression analysis, moving averages, and other statistical techniques for price prediction. These investigations provided insights into the relative strengths and limitations of statistical modeling in the context of crude oil prices.

3. ARIMA Model Analysis: We delved into the Autoregressive Integrated Moving Average (ARIMA) models, a classic approach for time series prediction. Our work involved parameter estimation, order selection, and rigorous testing of ARIMA models against crude oil price data.

S. NO	AUTHOR	APPROACH	PROS	CONS
1.	Narayan et al. (2017)	Heterogeneous Panel Data Analysis using macro-economic and financial indicators.	Comprehensive analysis of various factors, especially for long-term predictions.	Comprehensive analysis of various factors, especially for long-term predictions.
2.	Baumeister & Kilian (2016)	Machine learning algorithms with a focus on real-time data.	Can adapt to changing market conditions and real-time information.	Vulnerable to overfitting and model hyper parameter sensitivity.
3.	Khadjeh Nassirtoussi et al. (2019)	Long Short-Term Memory (LSTM) neural networks for time series forecasting.	Effective at capturing long-term dependencies in data.	Requires careful architecture tuning, may struggle with sudden market shifts.
4.	Mohamed et al. (2017)	Sentiment analysis of Twitter data to gauge market sentiment.	Captures sentiment-driven short-term price movements.	Noisy data and potential manipulation of sentiment on social media.
5.	Zhang et al. (2019)	Deep learning techniques using deep neural networks.	Ability to learn complex patterns in data.	Requires substantial data and computational resources.
6.	Lai et al. (2020)	Ensemble learning methods like bagging and boosting.	Reduces over fitting and enhances overall prediction accuracy.	Proper selection of base models and ensembling techniques is crucial.
7.	Kilian (2009)	Review of fundamental factors influencing crude oil prices.	Provides insights into long-term market dynamics.	Doesn't provide specific predictive models.
8.	Huang et al. (2017)	Volatility modeling using GARCH and related models.	Helps assess risk and uncertainty.	Doesn't provide point price predictions.
9.	Cheong et al. (2018)	Analysis of the impact of geopolitical events on oil prices.	Provides insights into market volatility.	Predicting specific geopolitical events is challenging.

Table No.: 1

4. LSTM Model Exploration: The utilization of Long Short-Term Memory (LSTM) models for time series forecasting was investigated. This encompassed the exploration of various LSTM architectures and their adaptability to crude oil price data. A detailed analysis of model training and performance evaluation was carried out.

5. Related Works Synthesis: To ensure a holistic understanding of the domain, we conducted a thorough synthesis of related works. This involved summarizing and critiquing the findings of other researchers, identifying common trends, and highlighting areas where our research can contribute novel insights.

The completion of these preliminary studies has laid the groundwork for the development and implementation of our deep learning-based crude oil price prediction model. These early findings are instrumental in shaping the methodology, approach, and objectives of our research. The knowledge and insights gathered from these endeavors will be instrumental in the subsequent stages of our project.

Future Work: Enhancing Deep Learning Model and Data Collection

Our future work plan encompasses a multifaceted approach to enhance the deep learning model and the acquisition of diverse data sources. The key components of this phase include:

Data Collection from Online Resources:

To enrich our data sources and improve model accuracy, we plan to gather information from a variety of online resources:

1. Online News Resources: Collect real-time news articles, reports, and analysis related to factors influencing crude oil prices. This involves web scraping and text extraction from reputable news websites, financial journals, and industry reports.

2. Social Media Data: Extract valuable insights from social media platforms like Twitter, where market sentiment and breaking news can impact oil prices. Natural Language Processing (NLP) techniques will be used to analyze and categorize relevant tweets and sentiments.

3. Historical Data: Continue to accumulate historical data to maintain a robust dataset that encompasses a wide range of timeframes. Historical data provides context for the model to understand long-term trends and cyclical patterns.

Model Base Development This Semester

The upcoming semester will be dedicated to creating the foundation for our deep learning model**:**

1. Data Preprocessing: Intensify efforts in data preprocessing by refining techniques for cleaning, formatting, and handling missing or noisy data. Ensure the integration of data from online sources into the existing dataset.

2. Model Architecture: Build the initial deep learning model architecture, incorporating the LSTM framework and any new layers or components necessitated by the additional data sources. The model will be designed to handle both time series data and unstructured text data effectively.

3. Training and Validation: Commence model training and validation using historical data and the integrated online resources. This step allows for the initial assessment of model performance and the identification of areas for improvement.

4. Baseline Results: Establish a baseline for model performance, including measures of accuracy, mean square error, and other evaluation metrics. This will serve as a benchmark for tracking improvements over time.

Continuous Model Refinement

The model development process will be an iterative one, involving continuous refinement and adaptation:

1. Regular Model Updates: Periodically update the model to account for new data and evolving market conditions. Implement a mechanism for regular retraining to ensure that the model remains accurate and up-to-date.

2. Feature Engineering and NLP: Continuously explore feature engineering techniques, including the use of Natural Language Processing (NLP) to extract insights from textual data. These efforts will contribute to the model's accuracy in forecasting oil prices.

3. Hyperparameter Optimization: Conduct ongoing hyperparameter optimization to fine-tune the model for optimal performance.

CHAPTER 4
CONCLUSION

We have conducted an in-depth study and analysis of historical data and research methods commonly used to predict crude oil prices. We observed that many models rely on a single algorithm, such as the ARIMA model, statistical models, or LSTM, and that increasing model complexity tends to enhance accuracy while decreasing Mean Squared Error (MSE).

Our innovative approach involves the integration of historical data with various external factors that can impact crude oil prices, including data from online news resources and Twitter (now known as X). By doing so, we aim to create a comprehensive model that offers improved accuracy in price prediction. Our expectation is that this model will yield an accuracy rate exceeding 95%.

This research not only contributes to the field of crude oil price prediction but also underscores the potential of leveraging diverse data sources to enhance forecasting capabilities. As we continue to refine and validate our model, it may open new avenues for improving predictive accuracy in financial markets.

REFERENCES

1. Lizardo, R. A., & Mollick, A. V. (2010). Oil price fluctuations and US dollar exchange rates. Energy Economics, 32(2), 399–408.

2. Sadorsky, P. (1999). Oil price shocks and stock market activity. Energy Economics, 21(5), 449–469.

3. Soni, A., van Eck, N. J., & Kaymak, U. (2007). Prediction of stock price movements based on concept map information. In Proceedings of 2007 IEEE symposium on computational intelligence in multicriteria decision making (pp. 205–211), Honolulu, HI.

4. Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. Pattern Recognition Letters, 31(14), 2225–2236.

5. Zhang, Y., & Wallace, B. (2017). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. In 8th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 253–263).

6. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32 bdel-Hamid, O., Mohamed, A. R., Jiang, H., & Penn, G. (2012). Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. In Proceedings of 37th IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 4277–4280).

7. Ivanov, V., & Kilian, L. (2005). A practitioner's guide to lag order selection for VAR impulse response analysis. Studies in Nonlinear Dynamics & Econometrics, 9(1), 1–34.

8. Kaufmann, R. K., & Ullman, B. (2009). Oil prices, speculation, and fundamentals: Interpreting causal relations among spot and futures prices. Energy Economics, 31(4), 550–558

9. Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), Advances in neural information processing systems, Vol. 25 (pp. 1097–1105). Curran Associates, Inc.

10. S Ani and S Ruhaidah, "Crude Oil Price Forecasting Based on Hybridizing Wavelet Multiple Linear Regression Model Particle Swarm Optimization Techniques and Principal Component Analysis", The Scientific World Journal, pp. 854520, 2014.

11. A A Karia, I Bujang and I Ahmad, "Fractionally integrated ARMA for crude palm oil prices prediction: Case of potentially overdifference", Journal of Applied Statistics, vol. 40, pp. 12, 2013.