# STOCK MARKET PREDICTION WITH TRANSDUCTIVE LONG SHORT-TERM MEMORY AND SOCIAL MEDIA SENTIMENT ANALYSIS

**A PROJECT REPORT**

***Submitted by***


**VIGNESWARAN U**

**(2023246001)**


*A report for the phase-I of the project*
*submitted to the Faculty of*

**INFORMATION AND COMMUNICATION ENGINEERING**

*in partial fulfillment*
*for the award of the degree*
***of***

**MASTER OF TECHNOLOGY**

***in***

**INFORMATION TECHNOLOGY**


**DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY**
**COLLEGE OF ENGINEERING, GUINDY**
**ANNA UNIVERSITY**
**CHENNAI 600 025**



**NOV 2024**

# ANNA UNIVERSITY

# CHENNAI - 600 025

# BONA FIDE CERTIFICATE

Certified that this project report titled STOCK MARKET PREDICTION WITH TRANSDUCTIVE LSTM AND SOCIAL MEDIA SENTIMENT ANALYSIS is the bonafide work of VIGNESWARAN U who carried out project work under my supervision. Certified further that to the best of my knowledge and belief, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on this or any other candidate.

**PLACE:**
**DATE:**

**DR.M.VIJAYALAKSMI**
**PROFESSOR**
**PROJECT GUIDE**
**DEPARTMENT OF IST, CEG**
**ANNA UNIVERSITY**
**CHENNAI 600025**

**COUNTERSIGNED**

**DR. S. SWAMYNATHAN**
**HEAD OF THE DEPARTMENT**
**DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY**
**COLLEGE OF ENGINEERING, GUINDY**
**ANNA UNIVERSITY**
**CHENNAI 600025**

# ABSTRACT

This project presents an innovative methodology for stock market prediction by integrating social media sentiment analysis with historical market data. The proposed system addresses the critical issues of imbalanced sentiment classification and temporal dynamics through the Off-policy Proximal Policy Optimization (PPO) algorithm and the Transductive Long Short-Term Memory (TLSTM) model. The Off-policy PPO adjusts the training reward system to prioritize the minority class, improving classification accuracy. Meanwhile, the TLSTM effectively combines temporal stock patterns with sentiment insights for more precise predictions. Project done confirms the combined model's superior performance, offering a significant advance in predictive analytics.

# ABSTRACT TAMIL

# ACKNOWLEDGEMENT

I would like to express my deep sense of appreciation and gratitude to my project guide **Dr. M. VIJAYALAKSHMI,** Professor, Department of Information Science and Technology, College of Engineering, Guindy, Anna University, Chennai for her invaluable support, supervision, guidance, useful suggestions and encouragement throughout this phase. Her moral support and continuous guidance enabled me to complete my work successfully.

I thank **Dr. S. SWAMYNATHAN,** Professor and Head of the Department of Information Science and Technology, College of Engineering, Guindy, AnnaUniversity, Chennai for the prompt and limitless help in providing the excellent computing facilities to do the project and to prepare the thesis.

Wearegrateful to the project committee members **Dr. S. SRIDHAR,** Professor, **Dr. G. GEETHA,** Associate Professor and **Dr. D. NARASHIMAN,** Teaching Fellow, Department of Information Science and Technology, College of Engineering, Guindy, Anna University, Chennai for their review and valuable guidance throughout the course of our project.

We also thank the faculty member and nonteaching staff members of the Department of Information Science and Technology at College of Engineering Guindy, Anna University, Chennai for their valuable support throughout the course of our project work.

**VIGNESWARAN U**

# TABLE OF CONTENTS

# LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

The stock market is a vital economic component that reflects and influences a nation's financial trajectory. Its inherent unpredictability creates significant challenges for investors and policymakers, necessitating advanced predictive tools. This project introduces a robust methodology combining financial data with social media sentiment analysis to improve stock market predictions. Leveraging innovations such as the Transductive Long Short-Term Memory (TLSTM) model and Off-policy Proximal Policy Optimization (PPO), the approach aims to address core challenges like data imbalance and temporal analysis.

## 1.1 ROLE OF STOCK MARKET

The stock market plays a critical role in shaping the economic landscape of nations and influencing global financial systems. It serves as a barometer for economic performance, reflecting the overall health of industries and the confidence of investors. As a pivotal mechanism for raising capital, the stock market directly impacts corporate strategies, national GDP growth, and employment rates.

### 1.1.1 Economic Health

Stock market trends often signal broader economic conditions. A bullish market, characterized by rising stock prices, typically indicates economic expansion, increased consumer confidence, and robust business

performance. Conversely, a bearish market suggests economic contraction and decreased investor confidence. The stock market's performance is closely monitored by policymakers, central banks, and international organizations to gauge economic stability and craft fiscal or monetary policies accordingly.

### 1.1.2 Investment Decisions

Investors, ranging from individuals to institutional entities, rely on stock markets to allocate resources efficiently. Decisions to buy, sell, or hold stocks are influenced by market signals, corporate earnings reports, and macroeconomic indicators. Sound investment decisions in a well-functioning stock market can result in wealth creation for individuals and stimulate corporate growth. Conversely, poor decisions, especially during volatile periods, can lead to significant financial losses, affecting broader economic stability.

### 1.1.3 Global Interconnectivity

In an increasingly globalized world, stock markets are interconnected. Events in one market, such as the U.S. or China, can ripple through global markets, influencing investor sentiment and financial flows. Stock market indices, such as the SP 500, FTSE 100, and Nikkei 225, are closely watched worldwide and often serve as benchmarks for international economic health.

## 1.2 CHALLENGES IN PREDICTION

Predicting stock market trends is notoriously complex due to the interplay of numerous factors, including economic indicators, corporate

performance, global events, and investor sentiment. Traditional methods often fall short in addressing these complexities, particularly in the following areas:

### 1.2.1 High Volatility and Non-linearity

Stock prices exhibit significant fluctuations driven by supply-demand dynamics, economic policies, geopolitical tensions, and unexpected events. Non-linear relationships between variables, such as interest rates, corporate earnings, and stock prices, complicate prediction models. Traditional linear approaches are often inadequate for capturing these intricate patterns.

### 1.2.2 Impact of Unstructured Data

A growing volume of unstructured data, such as financial news articles, social media posts, and earnings call transcripts, holds critical insights. However, extracting meaningful patterns from this data is challenging due to its variability and noisiness. The dynamic and often emotional nature of social media content further complicates its integration into predictive frameworks.

### 1.2.3 Temporal and Contextual Integration

Stock price movements are influenced by both immediate events and historical trends. Capturing temporal subtleties, such as recurring patterns or shifts in market sentiment, requires sophisticated time-series analysis. Most models struggle to balance short-term fluctuations with long-term patterns, leading to suboptimal prediction accuracy.

## 1.3 OPPORTUNITY WITH SOCIAL MEDIA DATA

In the digital age, social media platforms like Twitter, Reddit, and LinkedIn have become vital sources of real-time market sentiment. These platforms provide a unique opportunity to harness collective intelligence and gain a competitive edge in stock market predictions.

### 1.3.1 Rise of Social Media as a Financial Tool

Investors increasingly turn to platforms like Twitter for timely updates on market trends, corporate announcements, and breaking news. Studies indicate that social media sentiment can significantly influence stock price movements. Discussions on platforms like Reddit's r/WallStreetBets have shown the power of retail investors to shape market trends, as seen during the GameStop trading frenzy in 2021.

### 1.3.2 Real-Time Sentiment Analysis

Social media posts provide a real-time pulse of market sentiment, capturing the emotions and expectations of a diverse range of investors. Sentiment analysis techniques, using natural language processing (NLP) and machine learning, can classify posts as positive, negative, or neutral, enabling a deeper understanding of market dynamics.

### 1.3.3 Early Detection of Market Shifts

Social media provides early indicators of market trends. For instance, a surge in positive tweets about a specific company often precedes a rise in its stock price. Conversely, negative sentiment can signal impending declines. Monitoring these patterns offers a significant advantage, allowing investors to react swiftly and capitalize on emerging opportunities.

### 1.3.4 Challenges in Leveraging Social Media

The sheer volume of data generated on social media platforms presents a challenge for efficient processing and analysis. Noise in the data, including irrelevant or misleading posts, requires robust filtering and validation mechanisms. Ethical considerations, such as privacy and manipulation concerns, must be addressed when leveraging social media data for stock market predictions.

## 1.4 PROBLEM STATEMENT

The unpredictability of stock markets, influenced by a myriad of factors including investor sentiment and economic conditions, makes accurate forecasting a complex challenge. Existing models, such as LSTM and regression-based approaches, often fail to capture the nuances of minority sentiment classes or adapt to rapid market changes effectively.

Moreover, the integration of sentiment analysis with temporal stock data remains underexplored, particularly in addressing data imbalance and sequence learning challenges. Traditional methods either oversimplify these relationships or fail to leverage the full potential of real-time sentiment analysis, leading to suboptimal predictive performance.

### 1.4.1 OBJECTIVE

The objective of this project is to develop a predictive model that integrates sentiment analysis from social media and financial data to enhance stock market forecasting accuracy. This involves addressing critical challenges such as imbalanced sentiment classification and the dynamic nature of financial time-series data.

By employing methods like Off-policy PPO and TLSTM, the study aims to provide a model that not only improves predictive performance but also offers actionable insights for investors and policymakers.

# CHAPTER 2

# LITERATURE SURVEY/RELATED WORK

## 2.1 OVERVIEW

This section presents a review of key related works in time-series forecasting, machine learning-based stock prediction, sentiment analysis, and applications of Long Short-Term Memory (LSTM) networks in these domains. The focus is on methodologies that have influenced the development of forecasting models, sentiment analysis techniques, and performance improvements using various machine learning models, including LSTM and its variants.

## 2.2 EXISTING SYSTEM

### 2.2.1 Time-Series Forecasting with LSTM

Long Short-Term Memory (LSTM) networks have emerged as powerful models for time-series forecasting due to their ability to capture long-term dependencies in sequential data. LSTM networks have been successfully applied to a variety of forecasting tasks, including weather prediction and stock market forecasting. In a study by Z. Karevan and J. A. K. Suykens (2020), an LSTM-based model was deployed for weather forecasting[5], and their proposed Transductive LSTM (T-LSTM) model enhanced prediction accuracy by incorporating local information based on proximity to the test data. By applying a weighted quadratic cost function based on cosine similarity, the model outperformed standard LSTM models in multiple

experimental conditions, demonstrating its potential for improved forecasting in varying weather scenarios. This research highlights the importance of adapting LSTM models to specific datasets, improving their performance in real-world applications.

### 2.2.2 Stock Price Prediction Using Sentiment Analysis

Sentiment analysis has gained traction in financial prediction tasks, especially for predicting stock prices by analyzing social media, particularly Twitter data. A significant body of research has focused on leveraging machine learning algorithms in combination with sentiment analysis to predict stock market trends. For example, X. Zhu et al. (2019) explored the impact of Twitter sentiments on stock price fluctuations[7], using the Support Vector Machine (SVM) algorithm to classify tweets based on sentiments (positive, negative, or neutral). Their model, which focused on the NASDAQ 100 dataset, achieved a promising accuracy of 92 using an SVM with a linear kernel, demonstrating the effectiveness of integrating social media sentiment data into stock prediction models.

### 2.2.3 LSTM in Financial Forecasting

LSTM networks have also been widely used for stock price prediction, capitalizing on their ability to model sequential dependencies in stock market data. A recent study by R. Li et al. (2023) proposed an LSTM-Ladder Network (LLN) for automatic sleep stage classification[6], which could be adapted for financial forecasting by incorporating sleep stage-like periodic patterns in financial data. This method, which leverages transductive learning and adapts model parameters using both labeled and unlabeled data, demonstrated significant improvements in classification

performance, even when the data came from new or unseen subjects. The LLN model's ability to minimize reconstruction loss and refine its predictions based on unstructured data suggests its potential for stock price prediction tasks, especially in volatile markets.

### 2.2.4 Imbalanced Data Classification with AdaBoost and SMOTE

Imbalanced data classification is a persistent challenge in many machine learning applications, including financial prediction. Traditional classifiers often perform poorly on minority classes, leading to skewed predictions. A proposed solution to this problem is the combination of SMOTE (Synthetic Minority Over-sampling Technique) and AdaBoost[11], as discussed by A. Sarvani et al. (2023). By applying Adaptive Particle Swarm Optimization (APSO) to optimize weak classifier coefficients, the model achieved notable improvements in classification accuracy on imbalanced datasets. Although this work focused on general classification tasks, the techniques could be applied to stock market prediction, where the imbalance between positive and negative sentiment or stock price changes often poses a challenge.

### 2.2.5 Attention Mechanisms in Time-Series Forecasting

While LSTM networks excel at modeling temporal dependencies, attention mechanisms, which have gained popularity in deep learning, have been integrated into various time-series models to enhance performance. X. Zhu et al. (2019) conducted an empirical study on spatial attention mechanisms[7], finding that the combination of deformable convolution with key content saliency achieved the best accuracy-efficiency tradeoff for self-attention tasks. These findings are relevant to stock price prediction, where attention mechanisms can be used to prioritize relevant market data and adjust model focus accordingly.

By applying attention mechanisms to stock prediction tasks, models could better identify important time points in the stock data, leading to more accurate predictions.

### 2.2.6 Transductive Learning for Stock Prediction

Transductive learning, which focuses on utilizing local information and refining models based on the proximity of test points to training data, has shown potential in stock price prediction tasks. The method proposed by Z. Karevan and J. A. K. Suykens (2020)[5] could be extended to stock market predictions, where localized market data and sentiments from social media could be leveraged for better accuracy. Similarly, P. N. Achyutha et al. (2022) explored stock prediction using social media data and financial indicators[9], combining sentiment analysis with traditional machine learning classifiers like Naïve Bayes and Support Vector Machines (SVM). Their approach of using sentiment analysis to predict stock movements can be integrated with transductive learning techniques to improve the predictive performance of stock forecasting models.

## 2.3 LITERATURE SURVEY SUMMARY

The literature review highlights the significant role of LSTM networks, sentiment analysis, and transductive learning in time-series forecasting and stock market prediction. LSTM-based models, especially with the incorporation of local information via T-LSTM, offer a strong foundation for accurate predictions in dynamic environments like weather and stock markets. The integration of sentiment analysis from platforms like Twitter further enhances prediction models, as demonstrated by various studies. Additionally, the use of ensemble methods like AdaBoost and

SMOTE, as well as attention mechanisms, provides valuable techniques for improving performance, especially in imbalanced data scenarios. Overall, these advancements lay the groundwork for developing robust, data-driven prediction systems that can perform well in complex, real-world applications.

# CHAPTER 3

# SYSTEM ARCHITECTURE

## 3.1 COMPONENTS

The system architecture of the stock market prediction project is a multi-layered framework designed to integrate diverse data sources and advanced analytical techniques for precise forecasting. At the core, the architecture begins with a Data Collection Module that aggregates financial news, social media sentiment, and historical stock data to form a rich dataset. This data is cleaned and structured in the Data Preprocessing Module, ensuring it is suitable for analysis. The processed data is then analyzed through two critical pathways: the Sentiment Analysis Engine, powered by BERT and Off-Policy PPO algorithms, which extracts emotional tones and addresses class imbalances, and the Historical Stock Data Analysis Module, which employs Transductive Long Short-Term Memory (TLSTM) to model temporal dynamics in stock prices.

These outputs converge in the Prediction Engine, which combines temporal trends with sentiment insights to generate accurate stock price forecasts. Finally, the Output Layer delivers these predictions, supplemented with confidence scores, providing users with actionable insights for informed decision-making. This architecture effectively bridges qualitative sentiment data with quantitative market trends, ensuring a robust and comprehensive approach to stock market forecasting.

## 3.2 FLOW DIAGRAM

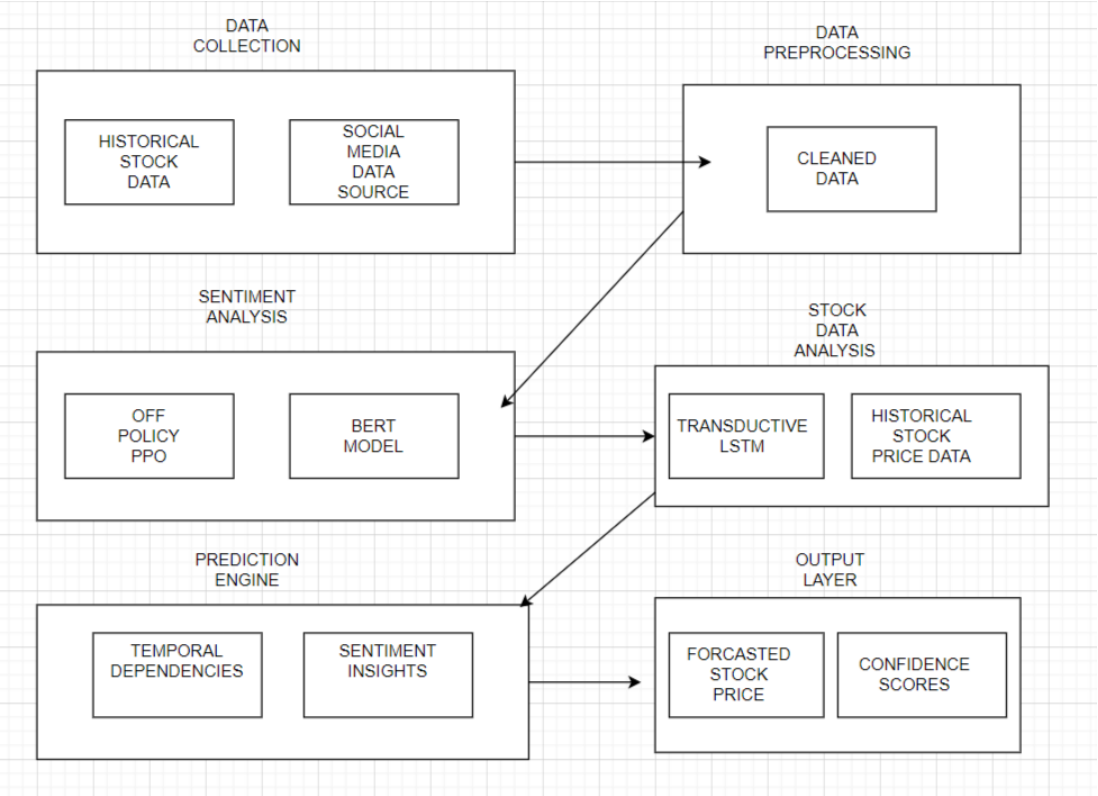


**Figure 3.1: System Architecture**

## 3.3 DATA COLLECTION MODULE

The Data Collection Module is a critical foundation of the stock market prediction system, designed to aggregate data from diverse and relevant sources. This module collects financial news from authoritative platforms, such as industry news outlets, economic reports, and market commentary, providing textual information essential for sentiment analysis. In addition, it integrates data from social media platforms, particularly Twitter, capturing public sentiment and real-time discourse surrounding specific companies or market conditions. To complement these, the module gathers historical stock market indices and related numerical data, such as opening prices, closing prices, high and low values, and trading volumes. Together, these inputs form a rich and multifaceted dataset, encompassing both qualitative and quantitative dimensions.

This comprehensive data acquisition ensures that the model has a holistic view of market dynamics, enabling accurate and contextually aware analyses of stock trends and public sentiment. By effectively bridging traditional market data with modern social media insights, the module lays the groundwork for a robust and nuanced prediction model.

## 3.4 DATA PREPROCESSING MODULE

The Data Preprocessing Module plays a vital role in preparing the raw data collected from various sources for meaningful analysis. This module ensures that the input data, which may come in unstructured or inconsistent forms, is transformed into a clean and standardized format. For text data, it removes irrelevant elements such as hyperlinks, special characters, and stop words that do not contribute to sentiment or context. Tokenization is performed to break down the text into smaller units, such as words or phrases, facilitating better understanding by machine learning models. Additionally, numerical data is normalized to ensure consistency in scale, preventing skewed analyses caused by disparities in value ranges.

By systematically organizing and refining the data, this module eliminates noise and enhances the quality of the input, making it suitable for the sentiment analysis engine and stock data analysis modules. As a result, the preprocessing module is instrumental in laying a solid foundation for accurate and reliable predictions in the system.

## 3.5 SENTIMENT ANALYSIS MODULE

The Sentiment Analysis Engine is a core component of the stock market prediction system, designed to analyze textual data and extract

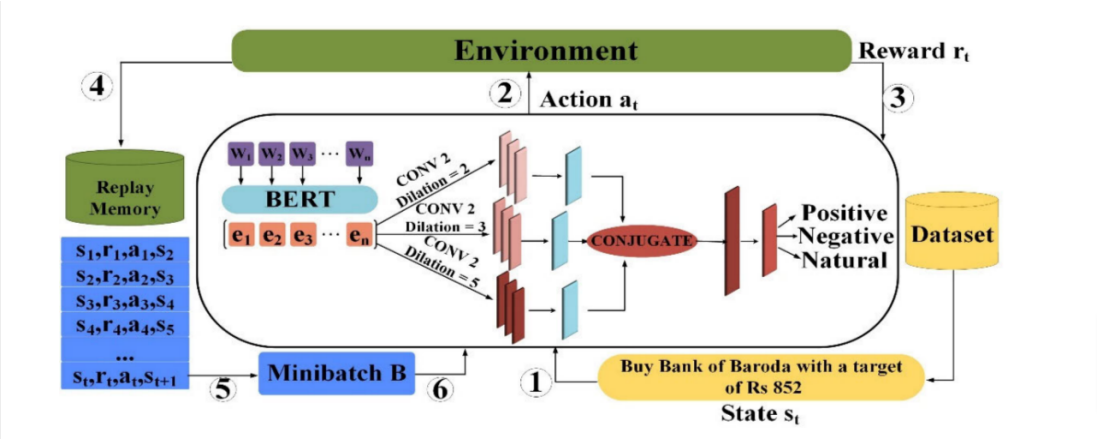


**Figure 3.2: Semantic Analysis Model**

underlying emotional tones, categorizing them as positive, negative, or neutral. Leveraging advanced semantic analysis techniques powered by models like BERT (Bidirectional Encoder Representations from Transformers), this module deciphers nuanced meanings and context from complex language structures found in financial news, social media posts, and other textual inputs. A critical feature of this engine is the integration of the Off-Policy Proximal Policy Optimization (PPO) algorithm, which addresses the common challenge of class imbalance in sentiment analysis. By employing a reward-based system that prioritizes underrepresented sentiment classes, the engine ensures balanced and accurate sentiment classification.

The sentiment scores generated by this module serve as a pivotal input for the prediction engine, providing insights into market sentiment trends that influence stock price movements. By bridging human emotion with financial data, the sentiment analysis engine empowers the system with the ability to incorporate real-world sentiment dynamics into stock market forecasting.

## 3.6 STOCK DATA ANALYSIS MODULE

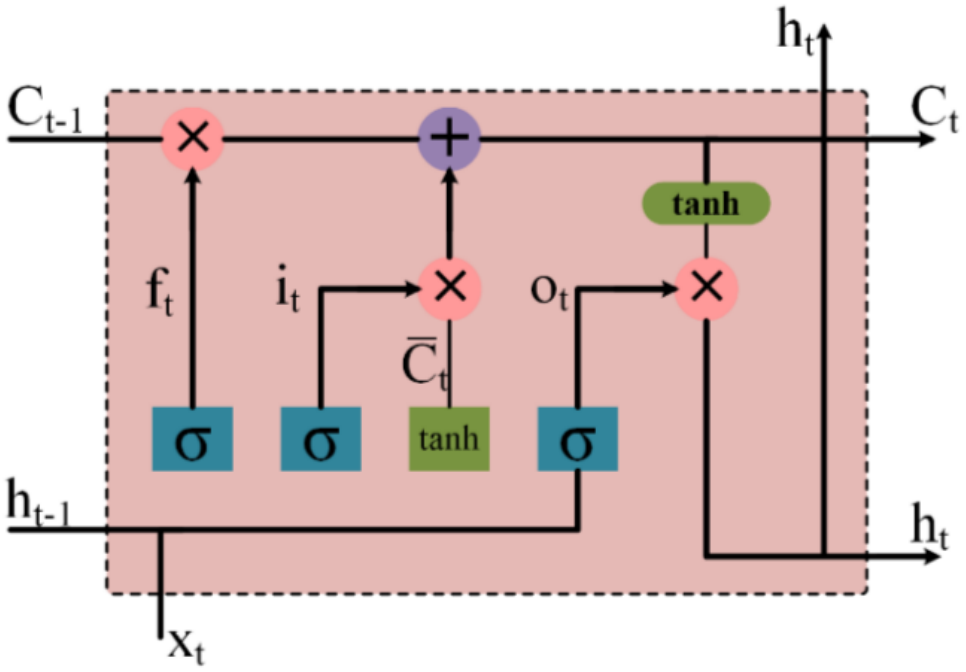


**Figure 3.3: Structure Of LSTM Cell**

The Historical Stock Data Analysis Module is integral to the stock market prediction system, focusing on analyzing past market performance to uncover temporal patterns and trends. This module processes key historical stock price data, including metrics such as opening and closing prices, daily highs and lows, and trading volumes. By examining these indicators, the module identifies recurring patterns and critical shifts in market behavior that may influence future stock movements.

The module utilizes advanced technology, specifically the Transductive Long Short-Term Memory (TLSTM) model, which is designed to capture both global and local temporal dependencies in time-series data. TLSTM excels at prioritizing data points closer to the prediction timeframe, enhancing the accuracy of forecasting. By integrating these temporal insights into the broader prediction model, this module provides a strong foundation for understanding the dynamic and often volatile nature of stock markets, ensuring informed and data-driven predictions.

## 3.7 PREDICTION MODULE

The Prediction Engine is the heart of the stock market prediction system, synthesizing insights from the Sentiment Analysis Engine and the Historical Stock Data Analysis Module to deliver precise and actionable forecasts of future stock prices. This module operates as an advanced analytical layer, integrating temporal dependencies captured by the Transductive Long Short-Term Memory (TLSTM) model with sentiment insights derived from semantic analysis.

By combining these two critical data streams, the engine accounts for both quantitative market trends and qualitative emotional influences, offering a holistic approach to stock market forecasting. The use of semantic classes further enriches the predictions by embedding contextual sentiment signals into the analysis, ensuring that public opinion and news sentiment are factored into the forecasts. The Prediction Engine is designed to produce reliable outputs that inform investors and policymakers about likely market movements, equipping them with data-driven tools for making informed financial decisions.

## 3.8 OUTPUT MODULE

The Output Layer serves as the final component of the stock market prediction system, translating complex analytical processes into actionable insights for end-users such as traders, analysts, and policymakers. This layer provides forecasted stock prices based on the integrated results of sentiment analysis and historical data analysis, offering predictions that are both precise and contextually aware.

Alongside these predictions, the Output Layer delivers confidence scores, which quantify the reliability of the forecasts and help users gauge the

associated risk levels. By presenting the results in a clear and interpretable format, this module empowers users to make informed and strategic decisions in the financial market. The actionable outputs generated by this layer not only assist in identifying potential investment opportunities but also enhance the overall trust and usability of the prediction system.

# CHAPTER 4

# IMPLEMENTATION

## 4.1 DATA COLLECTION

The data collection phase involves gathering diverse datasets to capture the multifaceted factors influencing stock market trends. Financial news articles from reputable outlets such as Moneycontrol and Economic Times, along with social media data, particularly from Twitter, provide sentiment insights reflecting public and market opinions. These textual datasets span from January 2015 to December 2020, encompassing 12,000 daily articles to ensure coverage of varied market conditions. Simultaneously, stock market data from the National Stock Exchange (NSE) of India is collected, focusing on 50 stocks and 10 Exchange-Traded Funds (ETFs) across multiple industries.

This dataset includes critical daily metrics such as opening and closing prices, trading volumes, and high-low ranges, totaling over 1.1 million entries. The combination of textual sentiment data and numerical stock metrics forms a comprehensive foundation for analyzing the interplay between sentiment and market performance, enabling robust stock price predictions.

## 4.2 DATA PREPROCESSING

Data preprocessing transforms raw data into a clean and structured format suitable for analysis. For textual data from social media and news articles, this involves removing irrelevant elements such as links, special characters, and emoticons, followed by tokenization to break text into individual words or phrases. Stemming is then applied to reduce words to their base forms,

ensuring consistency and reducing complexity in the dataset. Once the textual data is cleaned and processed, it is synchronized with numerical stock market data, such as opening and closing prices, daily highs and lows, and trading volumes.

This alignment ensures that sentiment information corresponds directly to specific stock performance metrics for the same timeframe. By combining qualitative and quantitative data, this preprocessing step establishes a unified dataset, enabling effective integration and analysis for sentiment-based stock market prediction.

## 4.3 SENTIMENT ANALYSIS

The sentiment analysis component employs a semantic analysis model to classify social media posts and financial news articles into positive, negative, or neutral sentiments. Using advanced techniques like Bidirectional Encoder Representations from Transformers (BERT), the model transforms text into numerical embeddings, processes them through convolutional layers, and classifies them via a multilayer perceptron (MLP). To address the challenge of class imbalance—where certain sentiments, such as negative, are underrepresented—the Off-policy Proximal Policy Optimization (PPO) algorithm is used.

This algorithm adjusts the reward system during training, assigning higher rewards for correctly classifying minority sentiment classes while reducing the emphasis on majority classes. This ensures the model learns to identify subtle yet crucial shifts in sentiment, enhancing its ability to produce balanced and accurate sentiment classifications, critical for effective stock market prediction.

**Algorithm 1** Pseudo-code of training the proposed semantic analysis model

Input: Training dataset $D = \{(s_1, y_1), (s_2, y_2), \ldots, (s_T, y_T)\}$, Learning rate $\alpha$
Output: Updated policy parameters $\theta$
$\theta, \theta_{old} \leftarrow$ Initialize policy parameters
$\phi \leftarrow$ Initialize value function parameters
$A \leftarrow$ Initialize advantage estimator
B $\leftarrow$ Initialize replay buffer
**for** e =1 to E **do**:
  Randomize the order of dataset D
  Set initial state to $s_1$
  **for** t =1 to T **do**:
    $a_t \leftarrow$ Choose action according to policy $\pi_\theta$ given state $s_t$
    $r_t \leftarrow$ Calculate reward using Reward $(x_t, a_t, y_t)$
    $s_{t+1} \leftarrow$ Determine the next state from dataset D
    B $\leftarrow$ Store transition $(s_t, a_t, r_t, s_{t+1}, \mu(\cdot|s_k))$
  **end for**
  **for** i =1 to N **do**:
    $(s_k, a_k, r_k, s_{k+1}, \mu(\cdot|s_k)) \leftarrow$ Draw random mini-batch from B
    $A(s_k, a_k) \leftarrow$ Estimate advantages using the value function with parameters $\phi$
    $\theta \leftarrow$ Optimize policy $\pi$ with objective $L^{CLIP}_{Off-PolicyPPO}(\pi)$ using $\alpha$
    $\theta_{old} \leftarrow \theta$
  **end for**
**end for**

**Figure 4.1: Pseudo-code of training the proposed semantic analysis model**

## 4.4 STOCK MARKET PREDICTION

The stock market prediction component integrates sentiment analysis results with historical stock market data using the Transductive Long Short-Term Memory (TLSTM) model. The TLSTM model builds upon traditional LSTM's ability to capture long-term dependencies in sequential data, enhancing it by prioritizing temporal proximity. This means that the model assigns greater importance to data points closer to the prediction timeframe, such as recent stock prices or sentiments, while still considering historical trends. Key inputs to the model include features like opening and closing prices, daily highs and lows, trading volumes, and sentiment scores derived from the analysis phase. This holistic input captures both the quantitative and qualitative factors influencing stock market movements, enabling nuanced forecasts.

By emphasizing temporal proximity, TLSTM excels in identifying patterns that are most relevant to the immediate future, which is critical in the dynamic and volatile nature of financial markets. The model is trained to recognize intricate temporal relationships, such as how a positive market sentiment surge correlates with stock price increases in the following days. This approach ensures more accurate and actionable predictions by blending the temporal strengths of LSTM with sentiment-driven insights. The result is a predictive framework capable of offering valuable guidance to traders, analysts, and investors for making informed decisions in rapidly changing market conditions.

## 4.5 MODEL TRAINING

The training phase focuses on fine-tuning the sentiment analysis and prediction models to achieve optimal performance. For the sentiment analysis model, training involves learning from labeled data to classify sentiments

accurately while addressing challenges like class imbalance. Techniques like Off-policy Proximal Policy Optimization (PPO) help the model focus on underrepresented sentiment classes, enhancing its ability to generalize across diverse scenarios. The prediction model, powered by the Transductive Long Short-Term Memory (TLSTM), is trained on synchronized sentiment and stock market data to capture temporal patterns and the influence of sentiments on stock prices. Advanced optimization techniques, such as adaptive learning rates and gradient clipping, are employed to stabilize training and prevent overfitting.

Once trained, the models are evaluated using robust metrics to ensure their effectiveness and reliability. For the prediction model, key performance indicators include Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE). These metrics assess the accuracy of the model's stock price forecasts, with RMSE measuring overall prediction errors, MAPE indicating percentage-based accuracy, and MAE highlighting absolute prediction deviations. Lower values for these metrics reflect better model performance. Regular cross-validation ensures that the models are not overfitting to the training data and can generalize effectively to unseen scenarios, providing reliable tools for real-world financial analysis and decision-making.

## 4.6 EVALUATION

To validate the effectiveness of the proposed method, it is benchmarked against other state-of-the-art models in sentiment classification and stock market forecasting. This comparison involves evaluating traditional models like ELR-ML and WD-SVM, LSTM-based models, and advanced sentiment analysis methods such as NL-LSTM and GRU-LSTM. Performance metrics like Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE) are used for stock prediction,

while metrics such as accuracy, F-measure, and G-mean evaluate sentiment classification.

The proposed method outperforms competitors in both domains, achieving significantly lower prediction errors and higher sentiment classification accuracy. This superior performance highlights the effectiveness of combining the TLSTM model and Off-policy PPO algorithm, which address temporal dependencies in stock data and class imbalances in sentiment analysis, respectively, enabling more reliable and precise results.

# CHAPTER 5

# RESULTS AND ANALYSIS

## 5.1 Evaluation Metrics

To evaluate the performance of the proposed model, the project employed standard metrics widely recognized in the fields of stock market prediction and sentiment analysis. For stock price prediction, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE) were utilized. These metrics assess the model's accuracy in predicting numerical stock values. RMSE measures the square root of the average squared errors, emphasizing larger errors. MAPE quantifies the average percentage error, focusing on relative accuracy. MAE, on the other hand, calculates the average magnitude of absolute errors, providing a straightforward measure of accuracy.

For sentiment analysis, performance was evaluated using accuracy, F-measure, Geometric Mean (G-means), and Area Under the Curve (AUC). Accuracy gauges the overall correctness of the classifications, while the F-measure balances precision and recall, particularly in handling imbalanced datasets. G-means focus on the balance between sensitivity and specificity, making it particularly suitable for imbalanced data. AUC quantifies the model's ability to distinguish between classes, offering a comprehensive measure of classification performance.

## 5.2 Model Performance

The proposed model demonstrated significant superiority over existing approaches in both stock price prediction and sentiment analysis. It achieved the

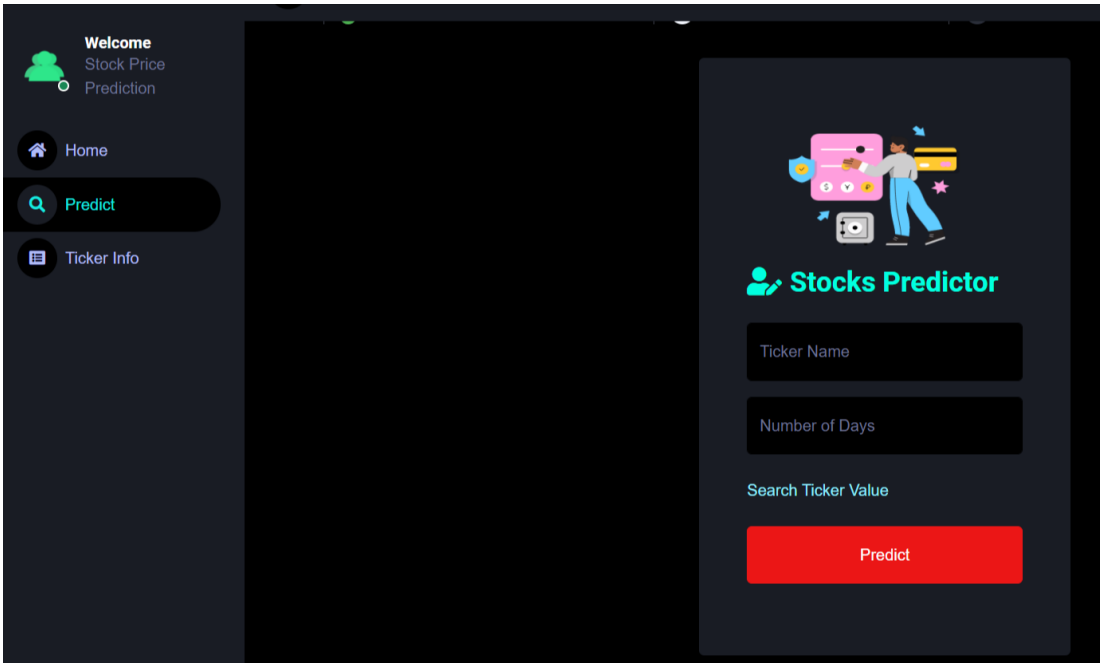


**Figure 5.1: Application Frontend**


**Figure 5.2: Live Stock Prediction**

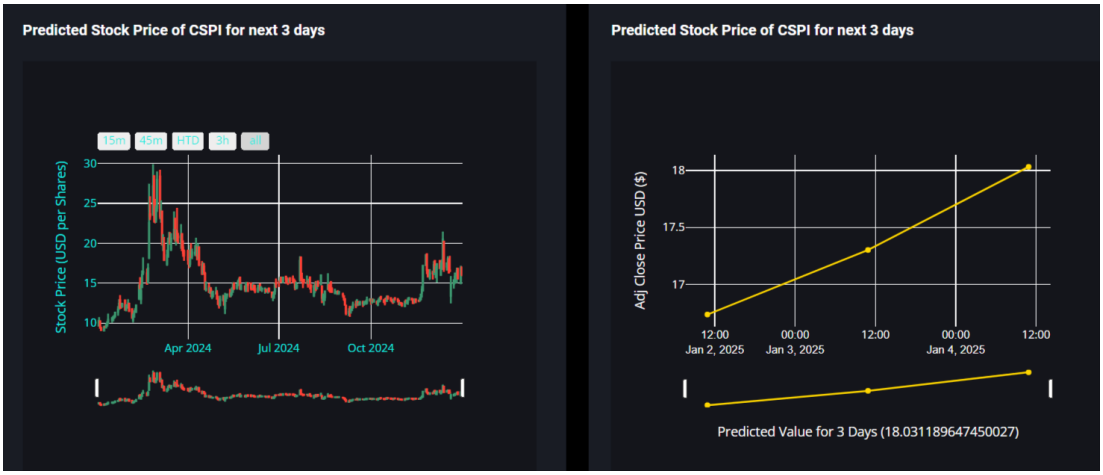


**Figure 5.3: Stock Price Prediction**

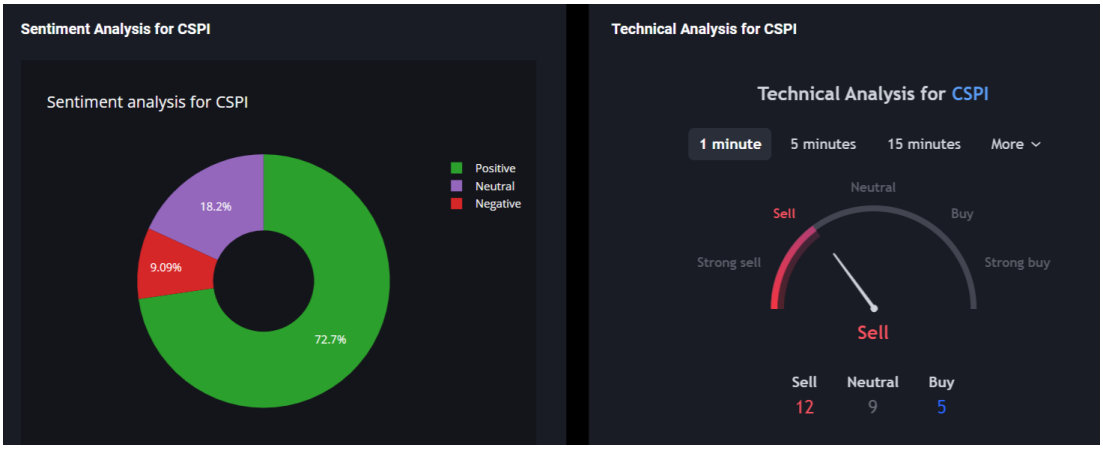


**Figure 5.4: Sentiment analysis**

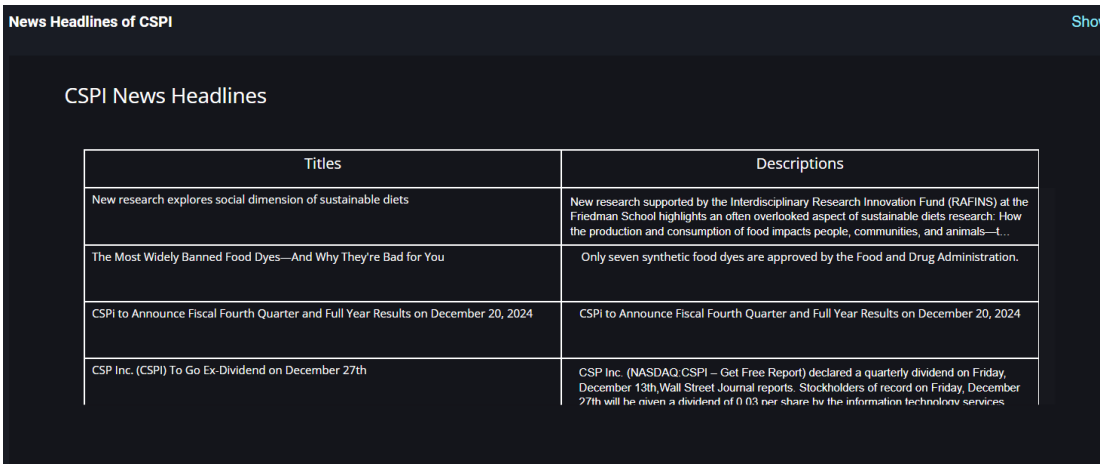


**Figure 5.5: Social Media News**

lowest RMSE, MAPE, and MAE values, underscoring its accuracy in predicting stock prices. In sentiment analysis, the proposed model outperformed state-of-the-art frameworks by attaining the highest scores in accuracy, F-measure, and G-means, indicating its robust capability in extracting meaningful insights from sentiment data.

## 5.3 Comparative Analysis

When compared to other models, the proposed approach showed marked improvements. In stock price prediction, it reduced RMSE by 38.67, MAPE by 22.84, and MAE by 28.16 relative to the next best-performing model, HiSA-SMFM. This highlights its superior ability to minimize errors and improve predictive accuracy. For sentiment analysis, the model achieved an 8.07 improvement in accuracy compared to PSAN, the most competitive benchmark. This underscores the effectiveness of its innovative sentiment classification techniques, particularly in handling imbalanced data.

## 5.4 Statistical Significance

The reliability of the improvements achieved by the proposed model was confirmed through t-tests, ensuring the statistical significance of the observed performance gains. These tests validated that the differences in results, both for stock price prediction and sentiment analysis, were not due to random variations. With p-values consistently below the 0.05 threshold, the results proved to be statistically significant, providing robust evidence that the proposed model's enhancements are meaningful and reliable.

In summary, the comprehensive evaluation revealed the proposed model's outstanding performance in both stock price prediction and sentiment

analysis. Its ability to achieve lower error rates, higher accuracy, and statistically significant improvements underscores its potential as a powerful tool for financial forecasting and sentiment interpretation.

# CHAPTER 6

# CONCLUSION AND FUTURE WORK

## 6.1 CONCLUSION

The proposed method integrating Transductive Long Short-Term Memory (TLSTM) with Off-policy Proximal Policy Optimization (PPO) and sentiment analysis has achieved notable success in improving stock market prediction accuracy.

**Sentiment Analysis**: The proposed model achieved an accuracy of 92.4 percent, along with an F-measure of 93.4 percent and G-means of 90.8 percent, outperforming other models significantly .

**Stock Market Prediction**: The model demonstrated superior performance, achieving a Root Mean Square Error (RMSE) of 2.147, which indicates high precision in its predictions .

## 6.2 FUTURE WORK

Future project could focus on broadening the dataset to cover more industries and global markets, incorporating real-time social media data for more dynamic analysis, and optimizing the reward system for enhanced classification.

# REFERENCES


[1] W. K. Cheng, K. T. Bea, S. M. H. Leow, J. Y.-L. Chan, Z.-W. Hong, and Y.-L. Chen. A review of sentiment, semantic and event-extraction-based approaches in stock forecasting. *Mathematics*, 10(14):2437, Jul. 2022.

[2] Y.-L. Lin, C.-J. Lai, and P.-F. Pai. Using deep learning techniques in forecasting stock markets by hybrid data with multilingual sentiment analysis. *Electronics*, 11(21):3513, Oct. 2022.

[3] C. Wimmer. A human-perception-inspired deep learning approach for intraday german market prediction. Technical Report 9943, Johannes Kepler University Linz, Linz, Austria, 2022.

[4] H. Oukhouya, H. Kadiri, K. El Himdi, and R. Guerbaz. Forecasting international stock market trends: Xgboost, lstm, lstm-xgboost, and backtesting xgboost models. *Statistics, Optimization and Information Computing*, 12(1):200–209, Nov. 2023.

[5] Z. Karevan and J. A. K. Suykens. Transductive lstm for time-series prediction: An application to weather forecasting. *Neural Networks*, 125:1–9, May 2020.

[6] R. Li, B. Wang, T. Zhang, and T. Sugi. A developed lstm-laddernetwork-based model for sleep stage classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:1418–1428, 2023.

[7] X. Zhu, D. Cheng, Z. Zhang, S. Lin, and J. Dai. An empirical study of spatial attention mechanisms in deep networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6687–6696, Oct. 2019.

[8] T. Swathi, N. Kasiviswanath, and A. A. Rao. An optimal deep learning-based lstm for stock price prediction using twitter sentiment analysis. *Applied Intelligence*, 52(12):13675–13688, Sep. 2022.

[9] P. N. Achyutha, S. Chaudhury, S. C. Bose, R. Kler, J. Surve, and K. Kaliyaperumal. User classification and stock market-based recommendation engine based on machine learning and twitter analysis. *Mathematical Problems in Engineering*, 2022:1–9, Apr. 2022.

[10] S. Harguem, Z. Chabani, S. Noaman, M. Amjad, M. B. Alvi, M. Asif, M. H. Mehmood, and A. H. Al-Kassem. Machine learning-based prediction of stock exchange on nasdaq 100: A twitter mining approach. In *Proceedings of the International Conference on Cyber Resilience (ICCR)*, pages 1–10, Oct. 2022.