 **STOCK PREDICTION**

INTERNSHIP REPORT

***Submitted by***

### 

### **VIJAYALAKSHMI N AC20UCS146**

Under Supervision of

Mr. Ravidas Bhat S,

Titan Engineering and Automation ,Hosur

(Duration:10/05/2023 To 10/06/2023)

*A report submitted in partial fulfillment of the requirements for the award of degree of*

**BACHELOR OF ENGINEERING**

**in**

##### **COMPUTER SCIENCE AND ENGINEERING**

**ADHIYAMAAN COLLEGE OF ENGINEERING (Autonomous)**

**HOSUR – 635 130, TAMIL NADU**

**JUNE - 2023**

# ADHIYAMAAN COLLEGE OF ENGINEERING

**(Autonomous)**

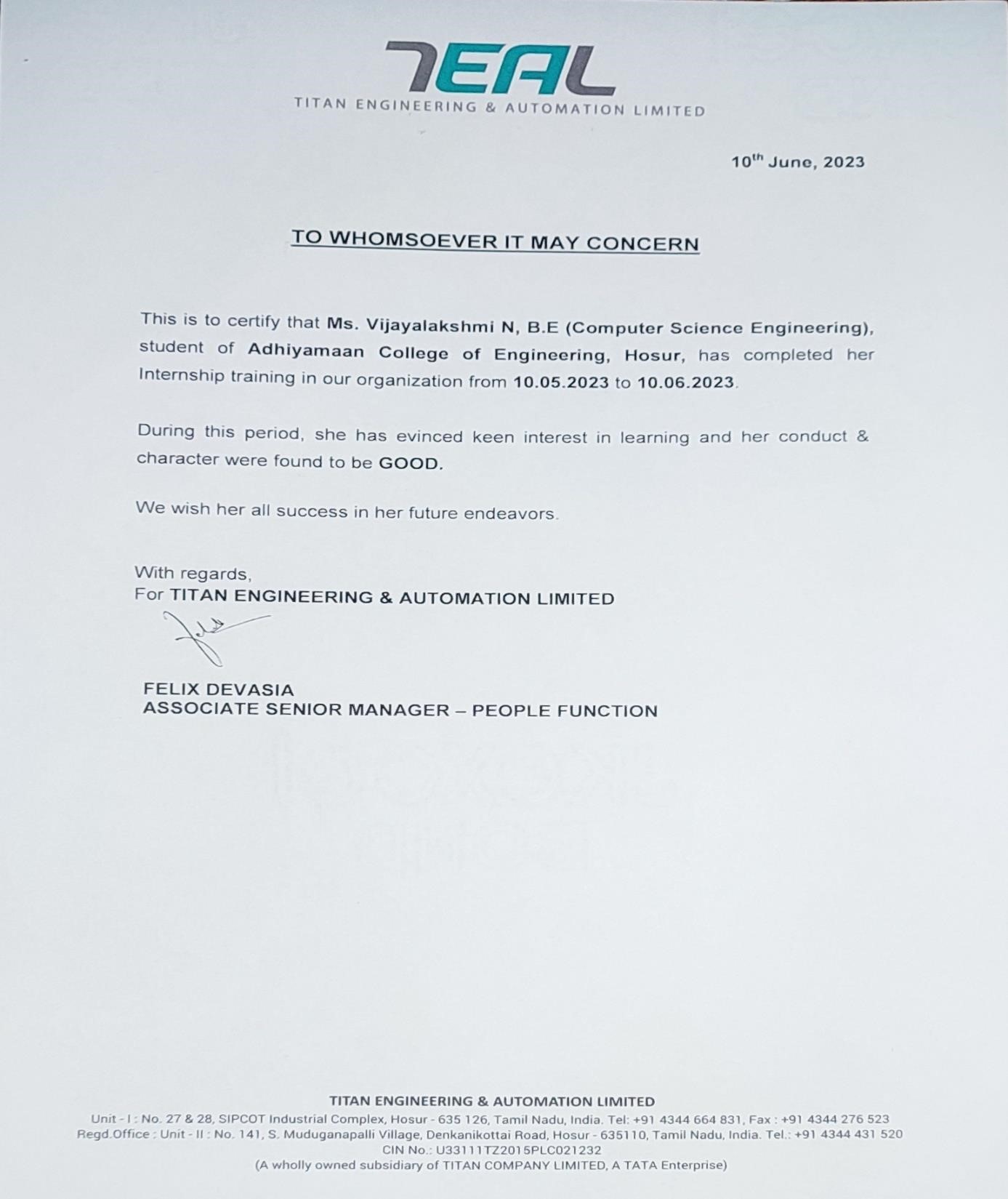
**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



This is to certify that the “**Internship report**” on “**STOCK PREDICTION USING MACHINE LEARNING”** submitted by **VIJAYALAKSHMI N (AC20UCS146)** during 2022-2023 academic year at **Titan Engineering and Automation, Hosur.**

Department Internship Coordinator Head of the Department

**CERTIFICATE FROM ORGANIZATION**



**ACKNOWLEDGEMENT**

We thank the almighty for granting me the wisdom, strength and grace to complete this project and for being with me in every step that we took in order to complete the project successfully.

With honor we place my sincere thanks and gratitude to **Dr. G. RANGANATH, M.E., Ph.D.,** Principal, Adhiyamaan College of Engineering (Autonomous), Hour, for this valuable opportunity given to us.

We are thankful to **Dr. G. FATHIMA, M.E., Ph.D.,** Head of the Department, Department of Computer Science and Engineering, Adhiyamaan College of Engineering, for her most valuable guidance, advice and encouragement in all stages of our project.

We would like to extend our gratitude to **Mr. R. VIKRAM, M.E.,** Assistant Professor, Department of Computer Science and Engineering, Adhiyamaan College of Engineering, for his continuous guidance and suggestions throughout the project.

Last but not the least, I would like to thank my family, friends and well-wishers who have assisted us directly or indirectly in the completion of my project.

**ABSTRACT**

Stock prediction plays a vital role in financial markets, aiding investors, traders, and financial institutions in making informed decisions. This abstract provides a concise summary of the methodologies, challenges, and trends in stock prediction. It discusses various approaches, including traditional statistical methods and modern machine learning techniques, highlighting their strengths and limitations in capturing the complex dynamics of stock markets. It begins by emphasizing the significance of stock prediction and its impact on investment strategies. It highlights the limitations of traditional methods that rely on historical price and volume data, such as moving averages and technical indicators. This sets the stage for the introduction of machine learning algorithms as a promising alternative for accurate and robust stock prediction. The abstract then explores the application of machine learning techniques in stock prediction. It discusses the use of supervised learning algorithms, such as artificial neural networks (ANNs), support vector machines (SVMs), and random forests, in modelling the relationships between historical stock data and relevant features, including financial indicators, news sentiment, and macroeconomic factors. It emphasizes the ability of machine learning models to capture non-linear patterns and exploit large volumes of data for improved forecasting accuracy. Furthermore, the abstract addresses the challenges associated with stock prediction, such as data preprocessing, feature selection, model overfitting, and evaluation metrics. It highlights the importance of robust feature engineering, careful model selection, and rigorous evaluation techniques to ensure reliable and unbiased predictions. It also emphasizes the need for domain expertise and continuous model refinement to adapt to changing market conditions. Lastly, the abstract discusses emerging trends and future directions in stock prediction. It highlights the potential of deep learning models, reinforcement learning, and ensemble techniques in enhancing prediction accuracy and adaptability. It also acknowledges the importance of incorporating interpretability and explain ability in machine learning models to gain trust and facilitate decision-making in financial markets. In conclusion, this abstract provides an overview of stock prediction methodologies, challenges, and trends. It underscores the advantages of machine learning techniques in capturing complex market dynamics and highlights the need for continuous research and development in this field. It serves as a valuable resource for researchers, practitioners, and investors interested in leveraging machine learning for stock market forecasting and decision-making.

**INDEX**

**S. No CONTENTS PAGE NO**

**1 INTRODUCTION 1**

1.1 Overview 1

1.2 About the Organization 2

1.3 Problem Statement 3

1.4 Objectives 4

**2 SYSTEM ANALYSIS 5**

2.1 Existing System 6

2.2 Proposed System 7

**3 METHODOLOGIES 8**

3.1 Architectural Diagram 9

3.2 Modules 10

**4 TECHNOLOGY USED 12**

**5 IMPLEMENTATION 14**

**6 RESULT 16**

**7 CONCLUSION 23**

**8 FUTURE WORK 24**

**1 INTRODUCTION**

Stock prediction is a challenging task that has captivated the interest of investors, financial analysts, and researchers for decades. The ability to accurately forecast stock prices is considered crucial in making informed investment decisions and maximizing returns. Traditional methods of stock analysis, such as fundamental analysis and technical analysis, have their limitations when it comes to predicting the complex dynamics of the stock market. However, with the advancements in machine learning and data analysis techniques, there has been a growing interest in using these approaches to enhance stock prediction accuracy.

* 1. **OVERVIEW**

The goal of stock prediction is to leverage historical market data, such as stock prices, trading volumes, financial indicators, news sentiment, and other relevant factors, to develop models that can generate accurate forecasts. These forecasts are valuable for investors as they can help identify potential investment opportunities, optimize portfolio allocation, and manage risks. This approach involves analysing the financial health and performance of a company to determine its intrinsic value and predict future stock prices. Fundamental factors such as earnings, revenue, growth prospects, industry trends, and management quality are considered in this analysis. This method focuses on studying historical price patterns, trading volumes, and market trends to forecast future price movements. Technical analysts use various tools and indicators, such as moving averages, support and resistance levels, chart patterns, and momentum indicators, to make predictions. With the advent of machine learning techniques, researchers and practitioners have been applying algorithms and models to predict stock prices. Machine learning models can learn from historical data and identify patterns and relationships that may not be evident through traditional analysis. Regression models, time series analysis, support vector machines, random forests, and deep learning architectures are commonly used in this context. This approach involves analysing news articles, social media sentiment, and other textual data to gauge market sentiment and its impact on stock prices.

* 1. **ABOUT THE ORGANIZATION**

**Titan Engineering & Automation Limited:**

Titan Engineering & Automation Limited is an engineering and automation company that specializes in providing innovative solutions for various industries. Here's an overview of Titan Engineering & Automation Limited

**Company Background:**

Titan Engineering & Automation Limited is a global company with headquarters in India. It was founded with the vision to revolutionize industries through advanced engineering and automation technologies. The company has established itself as a trusted partner for delivering cutting-edge solutions.

**Areas of Expertise:**

Titan Engineering & Automation Limited offers a wide range of services and solutions in the field of engineering and automation. Their expertise spans across multiple industries, including manufacturing, automotive, aerospace, robotics, oil and gas, and more.

**Product Development:**

The company excels in developing customized and innovative products to meet specific industry requirements. They have a dedicated team of engineers, designers, and technologists who work closely with clients to understand their needs and design solutions that address their challenges. From concept to final product, Titan Engineering & Automation Limited ensures high-quality and reliable deliverables.

**Automation Solutions:**

Titan Engineering & Automation Limited offers comprehensive automation solutions to streamline manufacturing processes and improve efficiency. They specialize in the design and implementation of robotic systems, industrial automation, control systems, and machine vision solutions. Their expertise in automation helps clients achieve cost savings, increase production output, and enhance overall operational effectiveness.

**Turnkey Projects:**

The company has extensive experience in handling turnkey projects, where they take full responsibility for the entire project lifecycle. They provide end-to-end solutions, including project planning, design, engineering, procurement, installation, commissioning, and ongoing support. This approach ensures seamless project execution and customer satisfaction.

**Research and Development:**

Titan Engineering & Automation Limited emphasizes continuous research and development to stay at the forefront of technological advancements. They invest in exploring new technologies, methodologies, and tools to enhance their offerings and provide innovative solutions to their clients. This commitment to R&D enables them to tackle complex challenges and deliver state-of-the-art solutions.

* 1. **PROBLEM STATEMENT**

The aim of this project is to develop a predictive model that accurately forecasts the future prices or trends of a given stock in the financial market. The stock market is characterized by high volatility, making it challenging to predict the future movements of stock prices. Therefore, the objective is to build a robust and accurate model that can assist investors and traders in making informed decisions. The goal of this project is to develop a robust and accurate stock prediction model that can provide reliable forecasts of stock prices or trends. The model should incorporate historical stock market data, financial indicators, news sentiment, and other relevant information to capture the underlying patterns and dynamics of the market. The prediction model should be evaluated based on its accuracy, interpretability, and ability to provide valuable insights for investment decision-making. The research aims to explore and develop machine learning algorithms, such as regression models, time series analysis, and deep learning architectures, that can handle non-linearities and capture the inherent complexities of stock price movements.

**1.4 OBJECTIVES**

**Develop accurate prediction models**:

The primary objective is to develop machine learning models that can accurately predict stock prices. This involves exploring various algorithms, such as regression models, time series analysis, and deep learning architectures, to capture the complex patterns and relationships in stock market data. The aim is to improve prediction accuracy compared to traditional methods and provide reliable forecasts for investment decision-making.

**Incorporate relevant features:**

Another objective is to identify and incorporate relevant features or input variables that can enhance the prediction performance. This includes selecting appropriate financial indicators, market data, news sentiment, and other factors that are known to impact stock prices. The goal is to improve the predictive power of the models by considering the most influential features.

**Handle non-linear relationships:**

Traditional linear models may not capture the non-linear relationships inherent in stock market data. Hence, the objective is to explore and develop machine learning algorithms that can effectively handle non-linear relationships and capture the complexities of stock price movements. Techniques such as deep learning, ensemble methods, and support vector machines can be employed to capture these non-linear patterns.

**Adapt to changing market conditions:**

The stock market is dynamic and influenced by various external factors, such as economic indicators, geopolitical events, and investor sentiment. An objective is to develop models that can adapt to changing market conditions and incorporate new information as it becomes available. This allows the models to adjust their predictions in real-time and provide up-to-date forecasts that align with the evolving market dynamics.

**Evaluate and refine models:**

Evaluation of the prediction models is crucial to assess their performance and identify areas for improvement. Objective evaluation metrics, such as accuracy, mean absolute error, and root mean square error, can be used to measure the performance of the models against actual stock prices.

**2 SYSTEM ANALYSIS**

System analysis for stock prediction involves using various techniques and methodologies to analyse and predict the future performance of stocks.

**Requirement Analysis:**

Identify the goals and objectives of the stock prediction system, such as accurate price forecasting, risk management, or portfolio optimization. Determine the specific requirements of the system, considering factors such as the desired prediction horizon, data sources.

**Data Analysis:**

Analyse the available historical stock market data, including price, volume, financial indicators, and news sentiment. Assess the quality and reliability of the data, identifying any missing values, outliers, or inconsistencies.

**System Design:**

The overall architecture of the stock prediction system, including the data flow, processing steps, and model integration. Select appropriate machine learning algorithms and models based on the analysis of data and requirements.

**Model Development:**

Develop and train machine learning models using the identified algorithms and techniques. Optimize the models by tuning hyperparameters and conducting cross-validation to improve their performance. Validate the models using appropriate evaluation metrics and statistical techniques to assess their accuracy and reliability.

**Integration and Deployment:**

Develop a user-friendly interface or application for users to interact with the system, providing inputs and receiving predictions. Deploy the system to the intended users or stakeholders, considering security, privacy, and regulatory compliance.

**2.1 EXISTING SYSTEM**

**Linear Regression:**

This is a simple and commonly used machine learning algorithm for stock prediction. It tries to establish a linear relationship between the input features (such as historical prices, trading volume, etc.) and the target variable (future stock price). However, linear regression may not capture complex patterns in stock market data.

**Long Short-Term Memory (LSTM) Neural Networks:**

LSTM is a type of recurrent neural network (RNN) that can capture temporal dependencies in sequential data, making it suitable for time series analysis like stock prediction. LSTM models can consider historical stock prices as a sequence and predict future prices based on the patterns learned from the data.

**Convolutional Neural Networks (CNN):**

While CNNs are mainly used for image recognition, they can also be applied to stock prediction tasks. CNNs can analyse stock market data represented as images (e.g., candlestick charts) and learn meaningful patterns from them.

**Reinforcement Learning:**

In RL, an agent learns to take actions in an environment to maximize a reward signal. The agent can learn to make trading decisions based on historical stock data and optimize its actions over time.

**DISADVANTAGES:**

* Existing stock prediction systems often employ complex algorithms and models that require significant computational resources.
* It may struggle to consistently achieve high prediction accuracy rates, especially in highly volatile or irrational market conditions.
* The performance of stock prediction systems heavily depends on the quality and relevance of the input data.
* Overfitting occurs when a model performs well on the training data but fails to generalize accurately to new, unseen data.

**2.2 PROPOSED SYSTEM**

**Data Collection:**

Gather historical stock data for the target company or companies of interest. This data can include daily or intraday stock prices, trading volume, financial ratios, news sentiment, and any other relevant features.

**Feature Selection:**

Select the most informative features that are likely to have an impact on stock prices. This can be done through statistical analysis, correlation analysis, or using machine learning techniques like feature importance estimation

**Model Training:**

Choose an appropriate machine learning algorithm based on the nature of the problem and the available data. As mentioned earlier, algorithms like linear regression, SVM, random forest, LSTM, or CNN can be considered. Split the data into training and testing sets and train the model on the training data

**ADVANTAGES**

* These systems help investors make informed decisions based on objective analysis rather than relying solely on intuition or speculation.
* Automated stock prediction systems can analyse vast amounts of data, process it quickly, and generate predictions in a timely manner.
* It can assist in risk management by providing insights into potential market fluctuations or downside risks.
* It can help optimize investment portfolios by suggesting asset allocation strategies, identifying diversification opportunities, or recommending buy/sell decisions based on predicted stock performance.

**3 METHODOLOGIES**

**Supervised Learning:**

Supervised learning algorithms are widely used for stock prediction. These algorithms learn from labelled historical data, where the input features (e.g., past stock prices, trading volumes, technical indicators) are used to predict the target variable (e.g., future stock prices, price movement direction

**Time Series Analysis:**

Stock prices exhibit temporal dependencies, making time series analysis techniques valuable for predicting future prices. Autoregressive Integrated Moving Average (ARIMA) models are frequently used for capturing trends, seasonality, and fluctuations in stock prices. Other time series models, such as exponential smoothing methods and state space models, can also be employed.

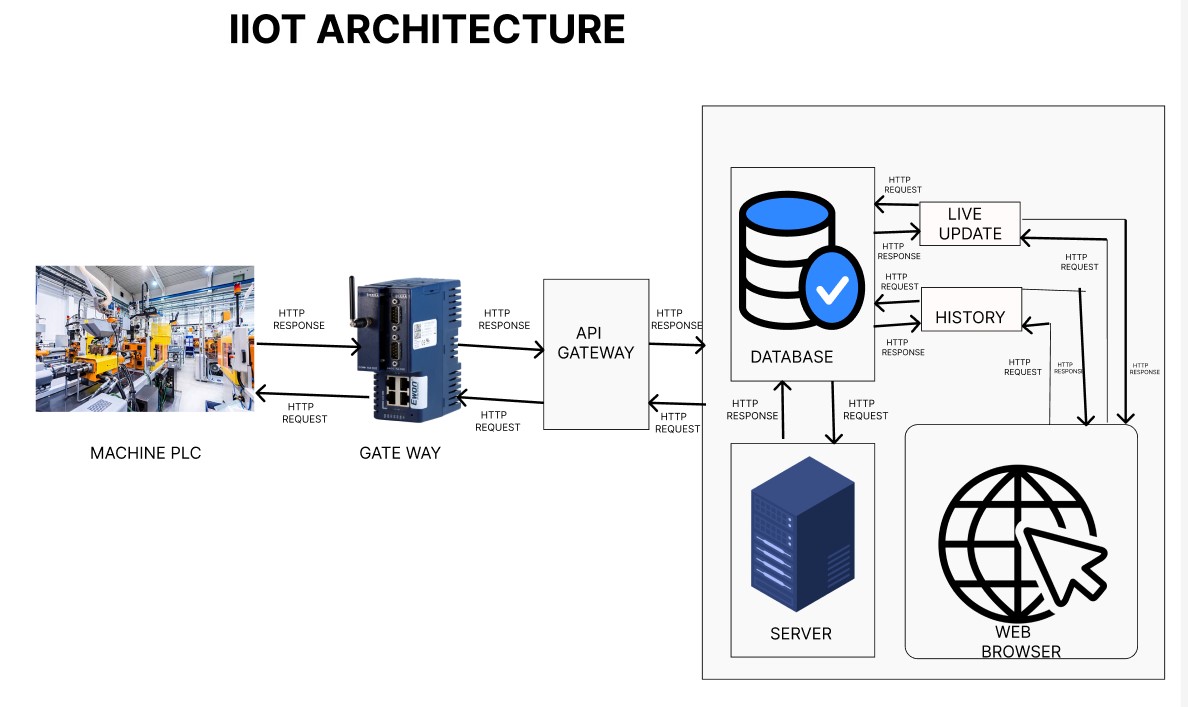
**Reinforcement Learning:**

Reinforcement learning (RL) can be used to develop stock trading agents that learn optimal trading strategies over time. RL agents interact with the stock market environment, take actions (e.g., buy, sell, hold), and receive rewards or penalties based on their performance. By optimizing their actions through trial and error, RL agents can learn strategies that aim to maximize long-term returns

**Sentiment Analysis:**

Stock prices can be influenced by news, social media sentiment, and other textual data. Sentiment analysis techniques can be used to analyse news articles, financial reports, social media posts, and other textual sources to extract sentiment scores or sentiment trends. These sentiment scores can then be incorporated as features in machine learning models to improve stock prediction. It's important to note that the choice of methodology depends on the specific requirements, available data, and the nature of the stock prediction problem. Experimentation and iterative refinement are often necessary to determine the most effective approach.

**3.1** **ARCHITECTURAL DIAGRAM**



**3.2 MODULES**

**There are six types of modules:**

* Data Collection Module
* Data pre-processing Module
* Feature Selection Module
* Model Training Module
* Model Evaluation Module
* Prediction Module

**Data Collection Module:**

Responsible for gathering historical and real-time stock market data from various sources, such as financial APIs, set data providers, or web scraping techniques. This module ensures a consistent and reliable data feed for the prediction system.

**Data pre-processing** :

Module: Performs data cleaning tasks to handle missing values, outliers, and inconsistencies in the collected data. Extracts relevant features from the raw data, such as technical indicators (e.g., moving averages, MACD.

**Feature Selection Module:**

Conducts an analysis of the available features and selects a subset of the most relevant ones for stock prediction. May employ techniques like correlation analysis, feature importance ranking, or domain expertise to determine the most impactful features.

**Model Training Module:**

Selects an appropriate machine learning model for stock prediction, such as linear regression, support vector machines (SVM), random forest, recurrent neural networks (RNN), or long short-term memory (LSTM) networks.

**Model Evaluation Module:**

Assesses the trained model's performance using appropriate evaluation metrics, such as accuracy, Compares the model's predictions against actual stock prices or market movements to measure its effectiveness.

**Prediction Module:**

Collects real-time or near-real-time stock market data for prediction. The choice of the module depends on your specific needs, the complexity of the models you want to build, and your familiarity with the library.

Stock prices exhibit time-dependent behaviour, making time series analysis an important module. Techniques such as autoregressive integrated moving average (ARIMA), exponential smoothing, or more advanced methods like long short-term memory (LSTM) networks can be used to model and forecast stock price movements based on historical patterns.

**4 TECHNOLOGY USED**

**Programming Languages:**

**Python:** Python is widely used for machine learning due to its extensive libraries and frameworks, such as NumPy, Pandas, scikit-learn, TensorFlow, and Keras. Python's versatility, extensive libraries, and active developer community make it a preferred language for stock prediction technology. It offers a wide range of tools and frameworks to handle different aspects of the prediction process, from data collection to model building and evaluation.

**Machine Learning Libraries and Frameworks:**

**scikit-learn:**

A comprehensive machine learning library in Python that provides various algorithms for regression, classification, and clustering tasks.

**TensorFlow:**

An open-source deep learning framework that allows building and training neural networks for stock prediction.

**Keras:**

A high-level neural network library that runs on top of TensorFlow, providing a user friendly interface for designing and training deep learning models.

**Data Processing and Analysis:**

**NumPy:**

A fundamental library for numerical computations in Python, used for handling arrays and mathematical operations on large datasets efficiently.

**Pandas:**

A versatile library for data manipulation and analysis, offering data structures like Data Frames that facilitate pre-processing and feature engineering.

**SQL:**

Structured Query Language is commonly used for managing and querying large datasets in relational databases, often employed to store and retrieve financial data.

**Data Visualization:**

**Matplotlib:**

A popular plotting library in Python for creating various types of charts, graphs

**Linear Regression:**

Used to establish a relationship between input variables and stock prices.

**Support Vector Machines (SVM):**

Classifies data points into different stock market trends.

**Natural Language Processing (NLP):**

NLP techniques are used to analyse textual data from news articles, social media posts, and financial reports. By extracting sentiment, identifying key events, and understanding market-related information, NLP helps in assessing the impact of news on stock prices.

**5 IMPLEMENTATION**

**Importing library files:**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sb

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler

from sklearn.linear\_model import LogisticRegression

from sklearn.svm import SVC

from xgboost import XGBClassifier

from sklearn import metrics

**Importing Dataset:**

df = pd.read\_csv('D:\ml\TSLA.csv')

df.head()

**Exploratory Data Analysis:**

plt.figure(figsize=(15,5))

plt.plot(df['Close'])

plt.title('Tesla Close price.', fontsize=15)

plt.ylabel('Price in dollars.')

plt.show()

**Feature Engineering:**

splitted = df['Date'].str.split('-', expand=True)

df['day'] = splitted[1].astype('int')

df['month'] = splitted[0].astype('int')

df['year'] = splitted[2].astype('int')

df.head()

**Data Splitting and Normalization:**

features = df[['open-close', 'low-high', 'is\_quarter\_end']]

target = df['target']

scaler = StandardScaler()

features = scaler.fit\_transform(features)

X\_train, X\_valid, Y\_train, Y\_valid = train\_test\_split(

features, target, test\_size=0.1, random\_state=2022)

print(X\_train.shape, X\_valid.shape)

**Model Development and Evaluation:**

models = [LogisticRegression(), SVC(

kernel='poly', probability=True), XGBClassifier()]

for i in range(3):

models[i].fit(X\_train, Y\_train)

print(f'{models[i]} : ')

print('Training Accuracy : ', metrics.roc\_auc\_score(

Y\_train, models[i].predict\_proba(X\_train)[:,1]))

print('Validation Accuracy : ', metrics.roc\_auc\_score(

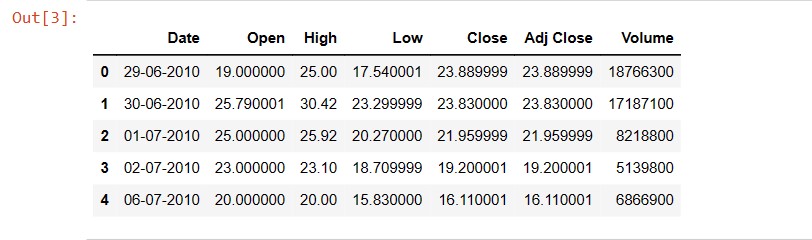
Y\_valid, models[i].predict\_proba(X\_valid)[:,1]))

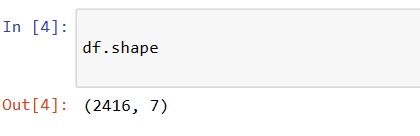
print()

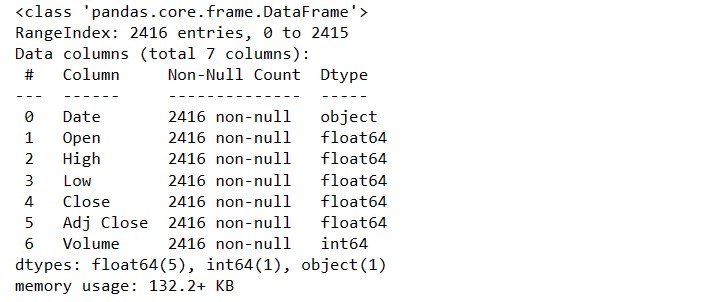
**6 RESULT**

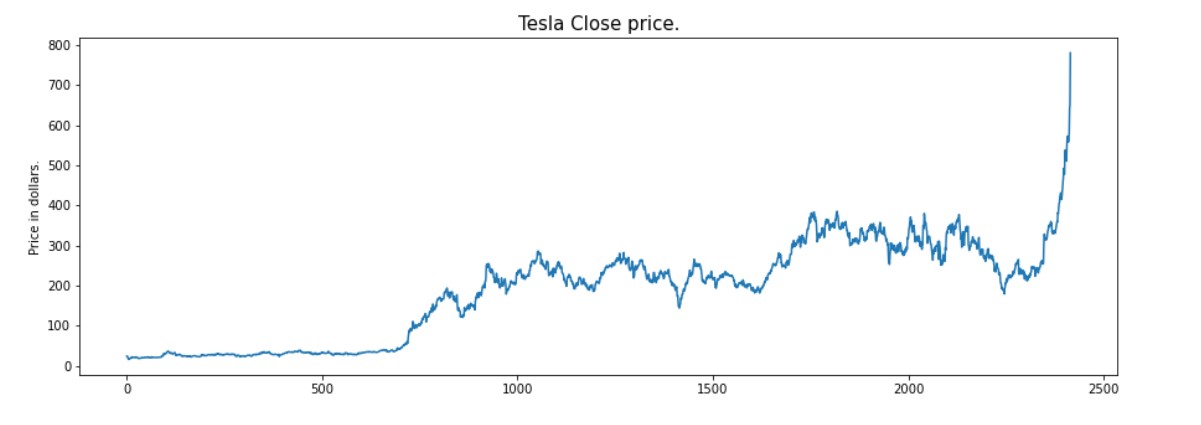
The result of stock prediction refers to the outcome or prediction generated by a stock prediction model or system. The result typically includes forecasts or estimates of future stock prices, market trends, or other relevant indicators. The accuracy and reliability of the result can vary depending on the quality of the prediction model, the data used, and the dynamic nature of the stock market.

It may provide predictions or estimates of future stock prices. These predictions can be in the form of specific price levels for a given time period or price ranges indicating the expected movement of stock prices. The model may generate predictions for individual stocks or broader market indices.





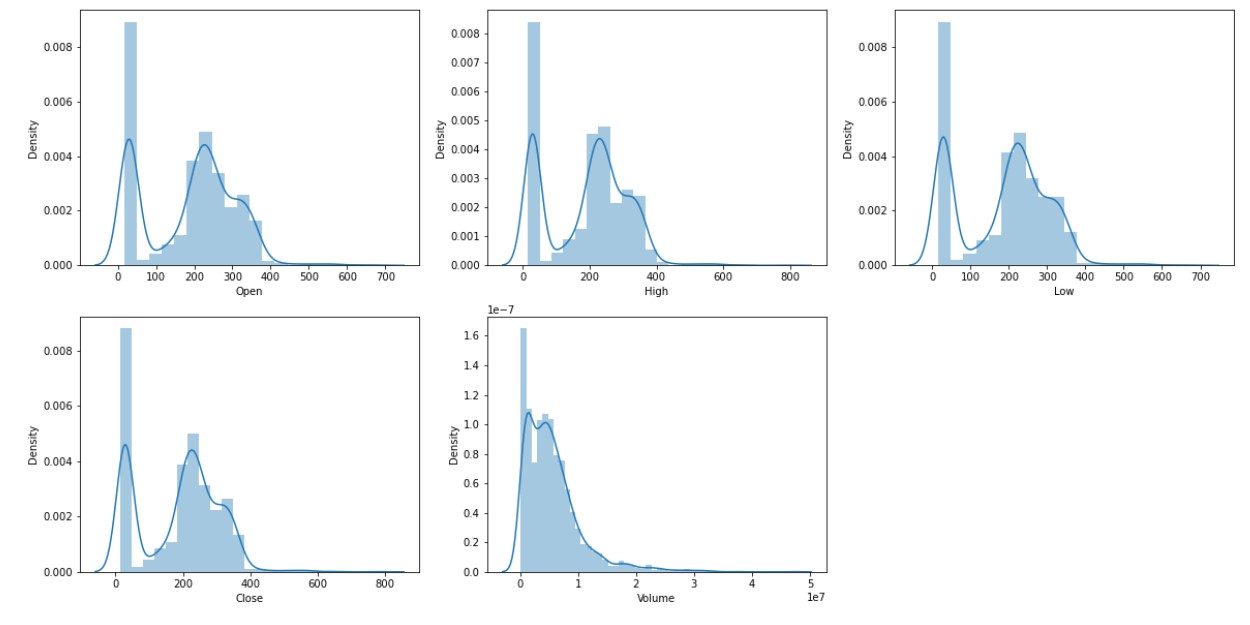


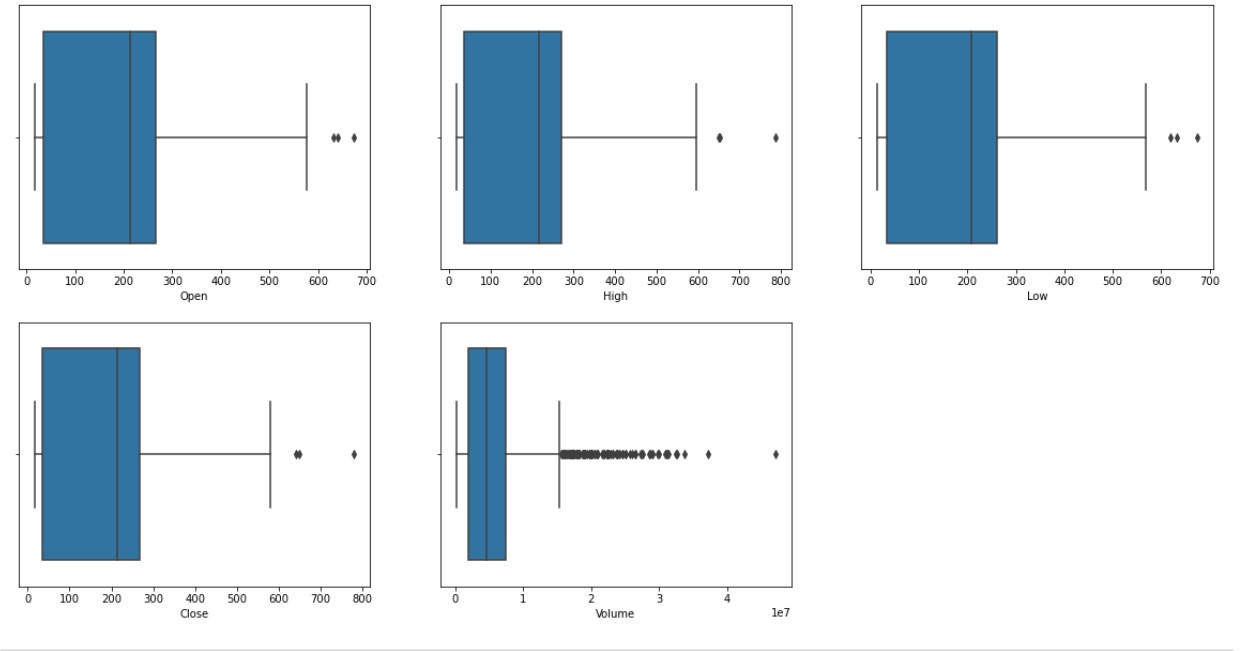




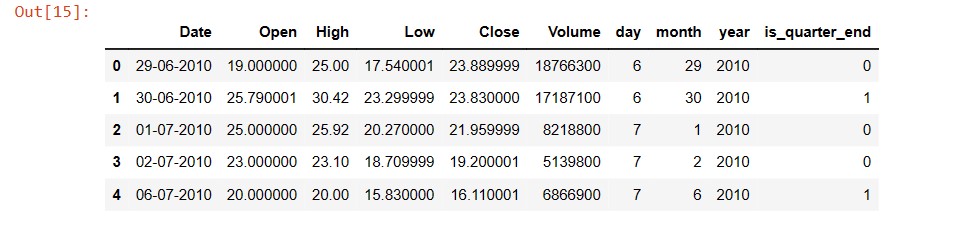


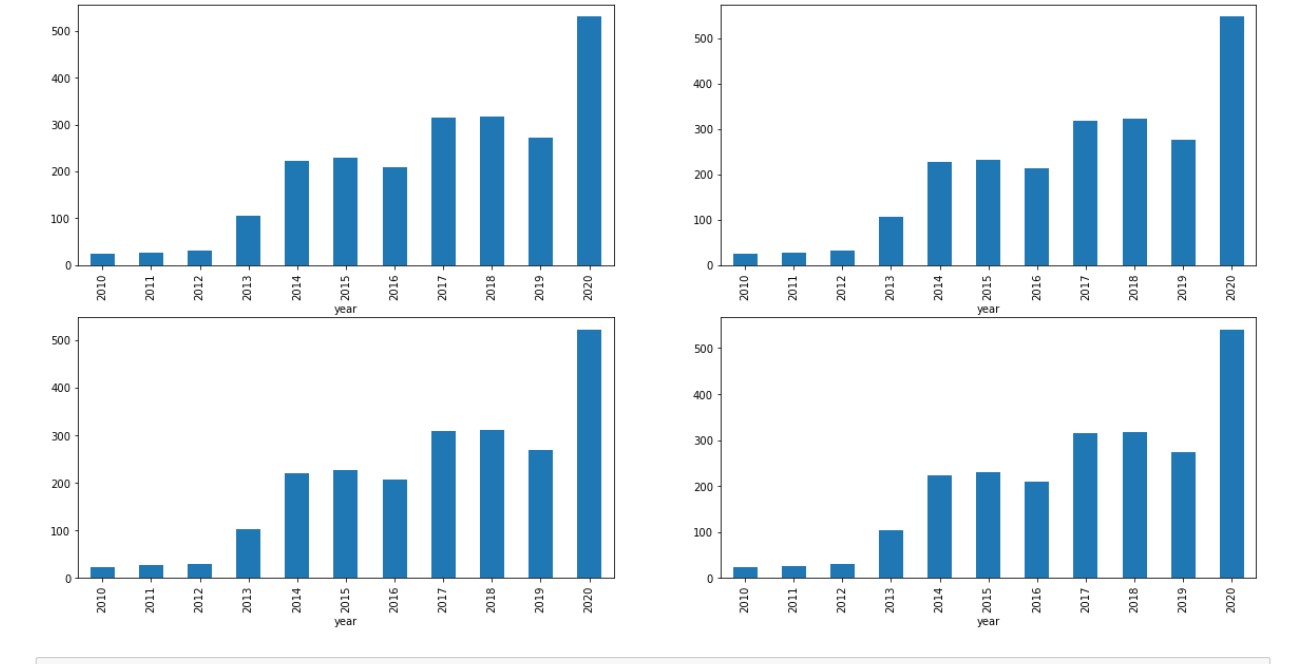


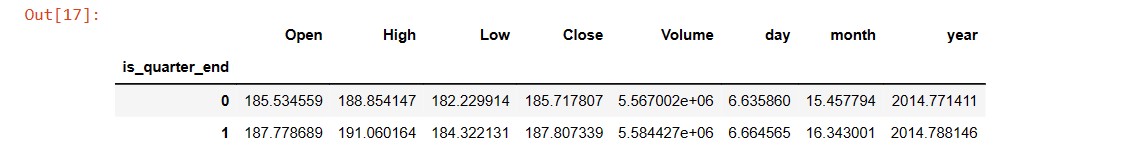


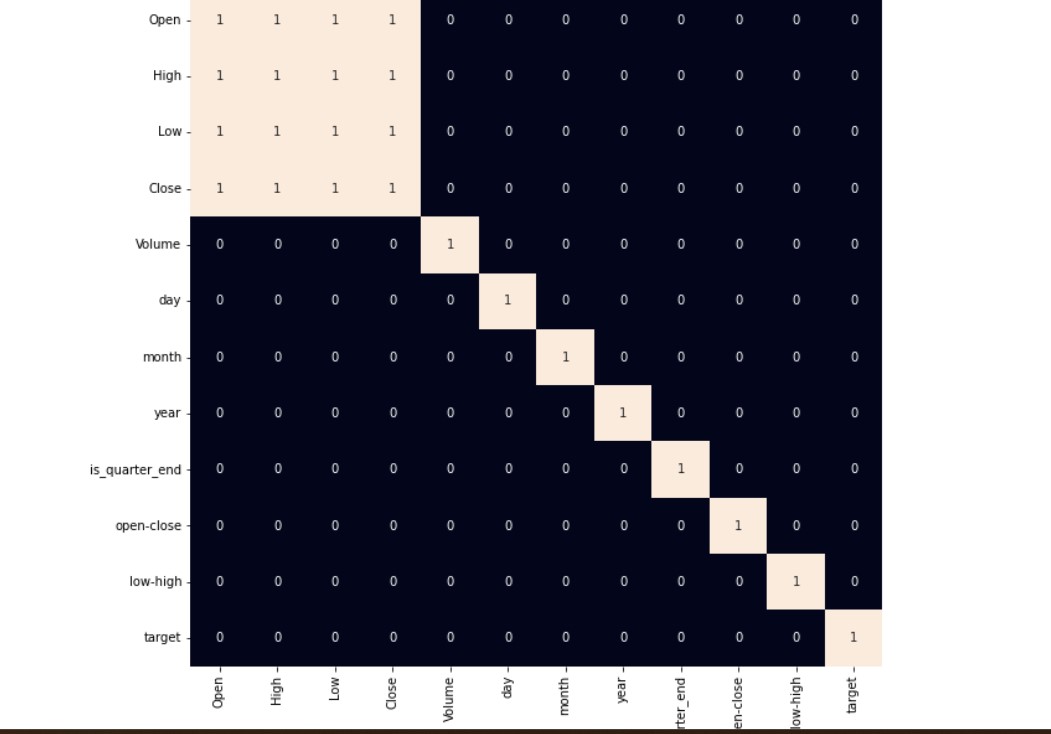
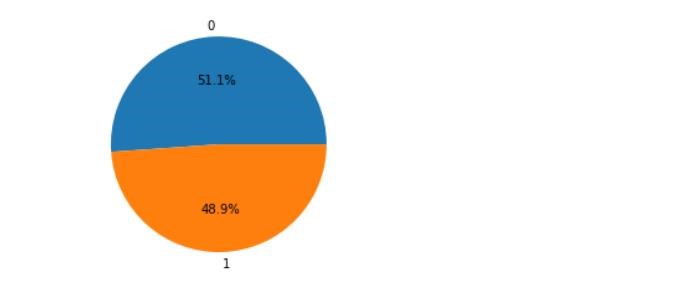


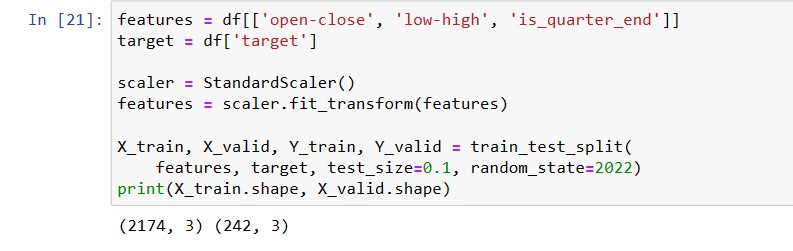


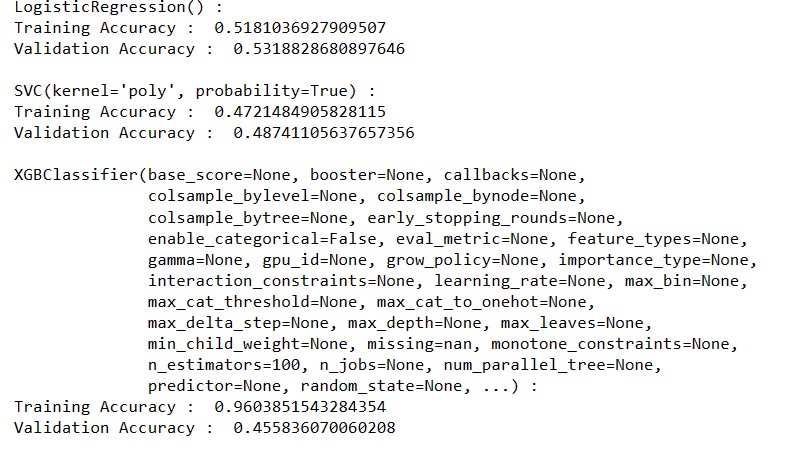


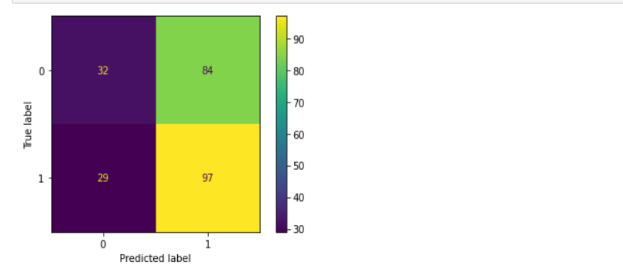












**7 CONCLUSION**

Stock prediction using machine learning is a challenging task and does not guarantee 100% accurate results. Machine learning models can provide valuable insights and potentially improve predictions compared to random guessing or simple heuristics. The performance of the model depends on factors such as data quality, choice of features, model architecture, and market conditions during the prediction period. Evaluation metrics like MSE, RMSE, or MAE are used to assess the accuracy of the predictions. Machine learning models should be used as one component of a comprehensive investment strategy, along with other methods like fundamental analysis and expert opinions. It's important to follow a systematic approach involving data collection, pre-processing, feature engineering, model selection, training, evaluation, deployment, and continuous monitoring. Stock prices are influenced by various factors beyond historical price data, such as economic indicators, news events, and investor sentiment. It's crucial to acknowledge the limitations of machine learning models and the inherent unpredictability of financial markets. Machine learning models should be used in conjunction with thorough research and consultation with financial professionals. Remember, stock prediction using machine learning is a tool that can assist in decision making, but it should not be the sole basis for investment decisions. We can observe that the accuracy achieved by the state-of-the-art ML model is no better than simply guessing with a probability of 50%. Possible reasons for this may be the lack of data or using a very simple model to perform such a complex task as Stock Market prediction.

**8 FUTURE WORK**

Future work involves exploring and incorporating alternative data sources beyond traditional financial and market data. This includes leveraging non-traditional data such as social media sentiment, news sentiment, satellite imagery, web scraping, or other unstructured data sources. Future work involves developing more sophisticated feature engineering techniques to extract relevant and meaningful features from raw data. This may involve applying advanced natural language processing (NLP) techniques for textual data, identifying informative technical indicators, or using dimensionality reduction methods to capture the most relevant information. Ensemble methods, which combine predictions from multiple models, have shown promise in stock prediction. Future work involves exploring novel ensemble techniques and model combination strategies to improve prediction accuracy and mitigate the risk of relying on a Single Mode Ensemble methods, in stock prediction. It aims to develop methods that provide insights into the factors influencing predictions and the rationale behind specific decisions. This can aid in building trust, understanding model limitations, and meeting regulatory requirements. Overall, future work in stock prediction combines advancements in data science, machine learning, deep learning, and financial domain expertise to develop more accurate, reliable, and interpretable models for predicting stock prices and supporting investment decision-making.