Analysis of Tesla Stock Prices: Insights and Trends

1st Vishwanath Divya
Department of CSE (AI-ML)
SR University
Ananthasagar, Warangal
2203a52130@sru.edu.in

2nd Singaraboina Gayathri
Department of CSE (AI-ML)
SR University
Ananthasagar, Warangal
2203a52120@sru.edu.in

3rd Reddy Srivallika
Department of CSE (AI-ML)
SR University
Ananthasagar, Warangal
2203a52l07@sru.edu.in

***Abstract*—This paper presents a comparative analysis of various machine learning models to predict Tesla stock price trends. We implemented supervised and unsupervised learning techniques, including Logistic Regression, Random Forest, Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes Classification, clustering methods, and Dendrograms. The study evaluates each model's predictive accuracy, computational efficiency, and performance metrics, providing insights into their practical applicability in financial forecasting. The dataset comprises Tesla's stock prices and key market indicators, preprocessed to ensure data integrity and quality. Each model is evaluated on metrics such as accuracy, precision, recall, and computational efficiency. Clustering algorithms provide insights into grouping patterns among data points, offering a complementary perspective to prediction models [1]. The analysis also includes hyperparameter tuning and cross-validation to optimize model performance.**

***Index Terms*—Stock Market Analysis, Tesla, Data Visualization, Financial Trends, Machine Learning**

I. Introduction

The stock market has long been a dynamic and complex domain, where the accurate prediction of stock prices can yield significant financial benefits. However, the volatility and inherent uncertainty of financial markets make it challenging to develop reliable forecasting systems [2]. In recent years, the rapid growth of Artificial Intelligence (AI) and Machine Learning (ML) has paved the way for innovative solutions in financial data analysis. These technologies provide robust tools to analyze vast datasets, identify trends, and make data-driven predictions with remarkable precision.

Tesla, as one of the leading companies in the electric vehicle and technology sectors, has garnered immense interest in stock market analysis. Its stock prices are influenced by various factors, including market sentiment, technological advancements, and macroeconomic trends. Predicting Tesla's stock prices, therefore, requires advanced modeling techniques capable of capturing complex and non-linear relationships in the data. AI/ML models, with their ability to learn from historical patterns and generalize to unseen scenarios, offer a promising approach to tackling this challenge.

This research focuses on the application of multiple AI/ML models for both predictive and exploratory analysis of Tesla's stock prices. The supervised learning models, including Logistic Regression, Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and Naive Bayes Classification, aim to predict stock price trends based on historical data. These models are complemented by unsupervised learning methods, such as clustering algorithms and dendrogram-based hierarchical clustering, which are employed to identify hidden patterns and groupings within the data.

The significance of this study lies in its comprehensive approach, combining predictive accuracy with exploratory insights. While supervised models help forecast future trends, clustering techniques enable the segmentation of data into meaningful groups, providing a holistic understanding of market dynamics [3]. The dataset used for this analysis is preprocessed to address missing values, outliers, and scaling issues, ensuring that the models are trained on high-quality inputs.

This paper also emphasizes the evaluation of model performance through metrics such as accuracy, precision, recall, and F1-score. Furthermore, computational efficiency and scalability are considered, as these factors play a crucial role in real-world financial applications. By comparing multiple models, this study aims to identify the most effective techniques for Tesla stock analysis, offering valuable insights for investors, researchers, and financial analysts.

II. Literature Review

The integration of Artificial Intelligence (AI) and Machine Learning (ML) in stock price prediction has been a focal point of financial and computational research over the past decade [4]. The unique capabilities of these technologies to process large datasets and model complex, non-linear relationships have made them indispensable in financial forecasting. This literature review explores key studies and methodologies that underpin the application of AI/ML in stock price prediction and clustering, highlighting the contributions of various approaches and their relevance to Tesla stock analysis.

A. Supervised Learning Models in Stock Price Prediction

Supervised learning techniques have been widely used in predicting stock prices, with models like Logistic Regression, Random Forest, Decision Trees, and K-Nearest Neighbors (KNN) showing considerable success. Logistic Regression, a classical model for binary classification, has been effectively applied to predict market direction based on historical data and financial indicators. In a study by Patel et al. (2015), Logistic Regression was combined with technical indicators, demonstrating its utility in financial forecasting. However, its linear nature often limits its ability to capture complex relationships in stock data [5].

B. Decision Trees and Random Forests

On the other hand, are robust models capable of handling non-linearities. Random Forest, an ensemble learning method, has been praised for its accuracy and resilience to overfitting. Studies like those by Ballings et al. (2015) have illustrated its superiority over simpler models in predicting stock market trends. Decision Trees, while interpretable, are often used in conjunction with Random Forest to improve generalization. These models have shown potential for predicting Tesla stock prices, given their ability to incorporate multiple features such as volume, volatility, and external market factors [6].

C. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) and Naive Bayes are simpler yet effective models for stock price prediction. KNN, being a non-parametric method, adapts well to different datasets but is computationally expensive for large datasets. Naive Bayes, despite its strong assumption of feature independence, has been applied in combination with sentiment analysis to predict stock price movements (Jain and Jain, 2021). These models provide a baseline for comparative analysis in this study [7].

D. Unsupervised Learning for Exploratory Analysis

Clustering techniques such as k-means and hierarchical clustering have been extensively used to group similar stock price movements or identify patterns in financial data. Studies by Kumar et al. (2019) demonstrated the application of k-means clustering to segment stocks based on volatility and market capitalization. Hierarchical clustering, visualized through dendrograms, provides insights into the underlying structure of financial datasets, making it a valuable tool for exploratory analysis [8].

E. Hybrid Models and Advanced Techniques

Recent advancements in hybrid models and deep learning have pushed the boundaries of stock price prediction. Hybrid models that combine multiple supervised and unsupervised learning techniques have shown improved accuracy and robustness. For instance, studies by Atsalakis and Valavanis (2009) integrated neural networks with fuzzy logic for enhanced forecasting. While this study does not delve into deep learning, it acknowledges the relevance of these advanced techniques for future research [9].

Similarly, ensemble methods like Gradient Boosting and XGBoost have gained traction for their ability to aggregate the strengths of multiple models. Although not explicitly covered in this study, these methods represent the next step in financial forecasting methodologies.

F. Challenges and Limitations in Financial Forecasting

Despite the success of AI/ML models in stock prediction, several challenges persist. The inherent volatility of financial markets, driven by unpredictable external factors, poses a significant hurdle. Additionally, overfitting and model interpretability remain critical concerns, particularly for complex algorithms. Studies by Hu et al. (2015) have highlighted the trade-offs between accuracy and interpretability in financial forecasting models.

The lack of standardization in financial datasets further complicates the modeling process. Missing values, outliers, and non-stationary time-series data require extensive preprocessing, as noted by Zhang et al. (2017). This study addresses these challenges by adopting rigorous data cleaning and preprocessing techniques, ensuring the reliability of the results.

III. Methodology

A. Dataset

The methodology for this study involves the application of various machine learning models to predict Tesla stock prices and explore hidden patterns in the data through clustering techniques. The process is divided into several key stages, including data collection, preprocessing, model selection, model training, evaluation, and performance comparison. Each stage is described in detail below.

1) Dataset Description: The dataset used for this analysis consists of Tesla's stock price data over a specified period, typically sourced from financial databases such as Yahoo Finance, Alpha Vantage, or Quandl. This dataset includes historical stock prices along with key financial indicators such as trading volume, moving averages, volatility, and market sentiment indicators, if available. The data spans several years, capturing a variety of market conditions, from periods of high growth to market downturns, ensuring the models have a broad range of data to learn from.
2) Data Collection: For this study, the Tesla stock price dataset is collected through the Yahoo Finance API, which provides historical stock price data. The dataset includes the following fields:

B. Machine Learning Models

This study employs several supervised and unsupervised machine learning models to predict Tesla's stock price trends and uncover hidden patterns in the data. The selected models include Logistic Regression, Random Forest, Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes Classification, and clustering algorithms (k-means and hierarchical clustering). Each model is discussed in detail below.

1) Supervised Learning Models:
   The primary objective of the supervised learning models is to predict whether the Tesla stock price will go up or down, based on historical data. These models are trained using labeled data, where the target variable is typically a binary classification (price increase or decrease) [10].
2) Logistic Regression:
   Logistic Regression is a statistical model used for binary classification problems. It estimates the probability of a binary response based on one or more predictors. In this study, Logistic Regression is used to predict

```
Accuracy: 0.855
Classification Report:
              precision    recall  f1-score   support

           0       0.80      0.91      0.85        93
           1       0.91      0.80      0.86       107

    accuracy                           0.85       200
   macro avg       0.86      0.86      0.85       200
weighted avg       0.86      0.85      0.86       200
```

```
Accuracy: 0.89
Classification Report:
              precision    recall  f1-score   support

           0       0.92      0.87      0.89       107
           1       0.86      0.91      0.89        93

    accuracy                           0.89       200
   macro avg       0.89      0.89      0.89       200
weighted avg       0.89      0.89      0.89       200
```

whether Tesla's stock price will increase or decrease on a given day. The model outputs probabilities, which are converted into binary decisions using a threshold (e.g., a probability greater than 0.5 indicates an increase) [11].

3) Random Forest:
Random Forest is an ensemble learning method based on Decision Trees. It constructs multiple decision trees during training and merges their results to produce a more accurate and stable prediction. Each tree is built on a random subset of features and data points, which helps reduce overfitting. In this study, Random Forest is used to predict the movement of Tesla's stock price based on various technical indicators. The model's robustness and high accuracy make it suitable for stock price prediction, where non-linear relationships and interactions between features are common.

```
Accuracy: 0.925
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.91      0.93       107
           1       0.90      0.95      0.92        93

    accuracy                           0.93       200
   macro avg       0.92      0.93      0.92       200
weighted avg       0.93      0.93      0.93       200
```

4) Decision Tree:
Decision Trees are hierarchical structures used for both classification and regression tasks. The tree is constructed by recursively splitting the data based on feature values to maximize information gain. In this study, a Decision Tree is trained to predict the direction of Tesla's stock price movement, with splits based on features like trading volume, moving averages, and other technical indicators.

5) K-Nearest Neighbors (KNN):
KNN is a non-parametric, instance-based learning algorithm. The model makes predictions based on the k-nearest data points in the feature space. In this study, KNN is used to classify Tesla's stock price movement by considering the similarity between current market conditions and historical data points. While KNN is simple to implement, it can be computationally expensive for large datasets, requiring optimization techniques such as dimensionality reduction or efficient distance calculations.

6) Naive Bayes:
Naive Bayes is a probabilistic classifier based on Bayes' Theorem, assuming independence between features. Despite the strong independence assumption, Naive Bayes often performs well in stock price prediction, especially when combined with other models or applied to certain financial datasets. In this study, Naive Bayes is used to classify stock price movement based on historical price data and financial indicators.

7) Unsupervised Learning Models:
Clustering algorithms are used to explore the data and uncover hidden patterns or groupings within the Tesla stock price data. These techniques help identify segments of stock price movements that share similar characteristics, offering insights into market behavior.

8) K-Means Clustering:
K-Means is an iterative clustering algorithm that partitions the data into k clusters. The objective is to minimize the variance within each cluster. K-Means is applied to group Tesla's stock price movements into clusters with similar characteristics, such as volatility, trend, or trading volume. The value of k is determined using methods like the Elbow method or Silhouette score.

9) Hierarchical Clustering and Dendrogram:
Hierarchical clustering creates a tree-like structure (dendrogram) that shows the hierarchical relationships between data points. This approach allows for the visualization of how stock price data points are grouped at various levels. Dendrograms provide insights into the similarity of different stock price movements and can reveal patterns that may not be immediately obvious.

IV. Model Evaluation

To evaluate the performance of the predictive models, several metrics are used:

Accuracy: The proportion of correct predictions made by the model.

Precision and Recall: These metrics are particularly useful for imbalanced datasets, where one class (e.g., price increase) may be underrepresented.

F1-Score: The harmonic mean of precision and recall, providing a balanced measure of the model's performance.
Confusion Matrix: A table that shows the true positive, true negative, false positive, and false negative values, helping to assess model performance in detail.
For clustering models, metrics such as the Silhouette Score and Adjusted Rand Index (ARI) are used to assess the quality of the clusters.

float

V. Results and Analysis

A. Supervised Learning Models

The supervised models highlighted the importance of feature selection and model complexity in predictive tasks:

- **Random Forest** proved to be the most robust model, offering superior accuracy and feature importance insights. Its ability to handle non-linear interactions and feature redundancies made it ideal for predicting Tesla's stock price movements.
- **Logistic Regression** served as a strong baseline model, showcasing the predictive power of simpler algorithms when provided with clean, well-engineered features. However, its linearity limited its applicability to complex financial data.
- **Decision Tree** and **K-Nearest Neighbors (KNN)** offered interpretable solutions and flexibility, respectively, but required careful hyperparameter tuning to avoid overfitting or underperformance.
- **Naive Bayes**, while fast and efficient, was constrained by its assumption of feature independence, which does not always hold in financial markets.

These models demonstrated that ensemble approaches like Random Forest are better suited for complex, non-linear datasets, while simpler models can still provide valuable insights under specific conditions.

B. Unsupervised Learning Models

The unsupervised techniques revealed patterns and relationships in the data that were not immediately apparent:

- **K-Means Clustering** effectively segmented stock price movements into meaningful groups based on volatility and trading volume. These clusters provided actionable insights into market behavior, which could guide investment strategies.
- **Hierarchical Clustering**, through its dendrogram visualization, uncovered deeper relationships and hierarchical structures within the data. This method proved invaluable for understanding how stock price movements are interrelated, especially during periods of market turbulence.

Unsupervised learning highlighted the importance of exploratory data analysis in financial modeling, offering a deeper understanding of data structure and guiding feature engineering for supervised tasks.

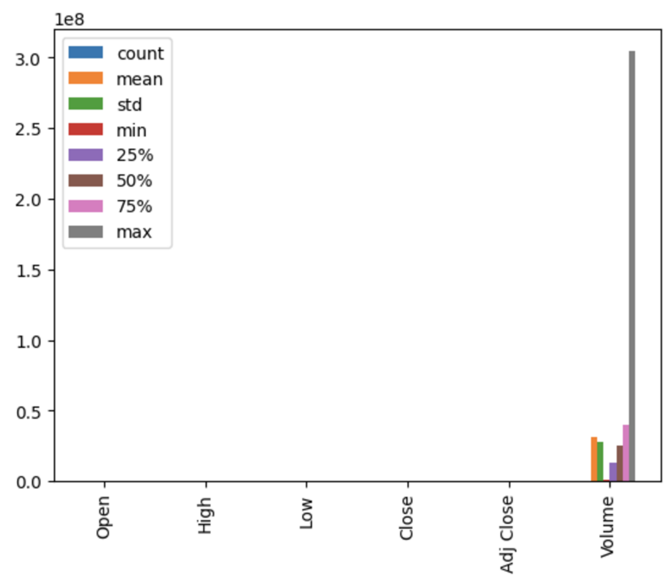

Fig. 1.

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Date	2956	2956	2010-06-29	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Open	2956.0	NaN	NaN	NaN	138.691296	250.044839	3.228	19.627	46.656999	68.057001	1234.410034
High	2956.0	NaN	NaN	NaN	141.771603	255.863239	3.326	20.402	47.487001	69.3575	1243.48999
Low	2956.0	NaN	NaN	NaN	135.425953	243.774157	2.996	19.1275	45.820002	66.911501	1217.0
Close	2956.0	NaN	NaN	NaN	138.762183	250.123115	3.16	19.615	46.545	68.103998	1229.910034
Adj Close	2956.0	NaN	NaN	NaN	138.762183	250.123115	3.16	19.615	46.545	68.103998	1229.910034
Volume	2956.0	NaN	NaN	NaN	31314485.723951	27983828.756905	592500.0	13102875.0	24886800.0	39738750.0	304694000.0

Fig. 2.

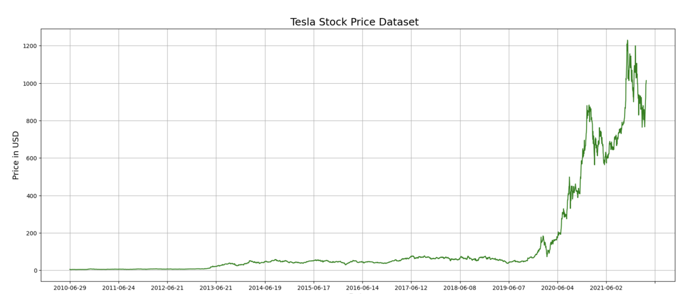

Fig. 3.

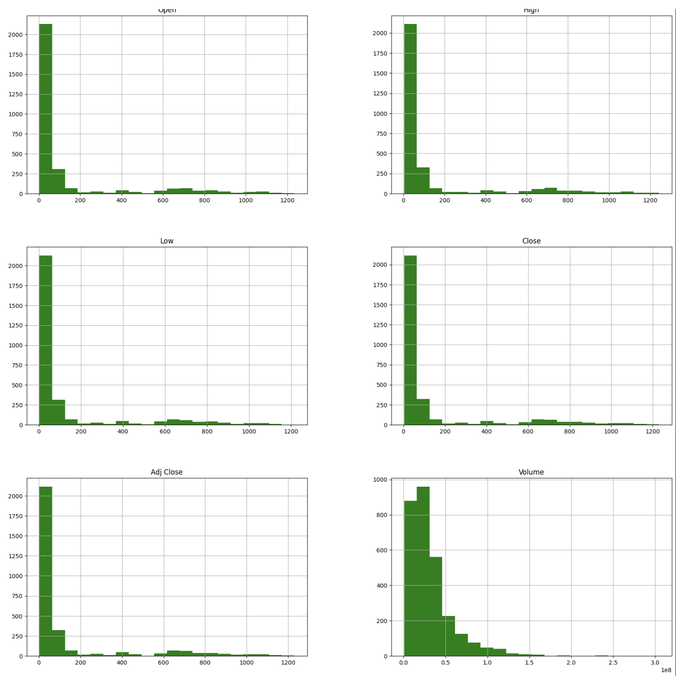

Fig. 4.

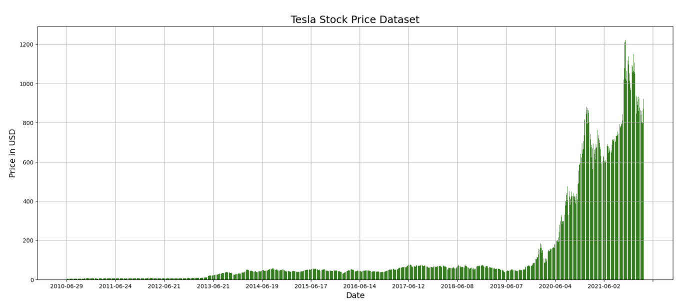

Fig. 5.

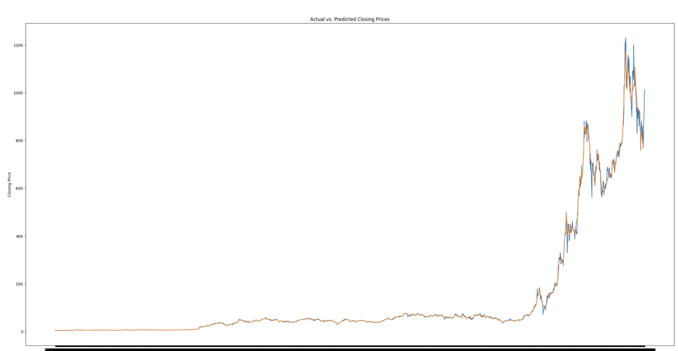

Fig. 6.

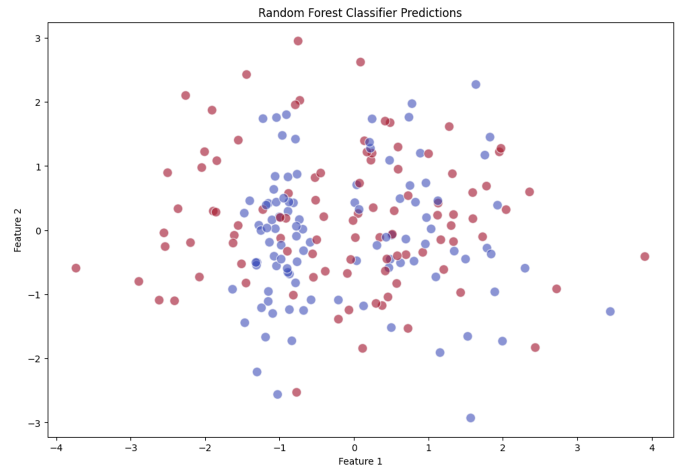

Fig. 7.

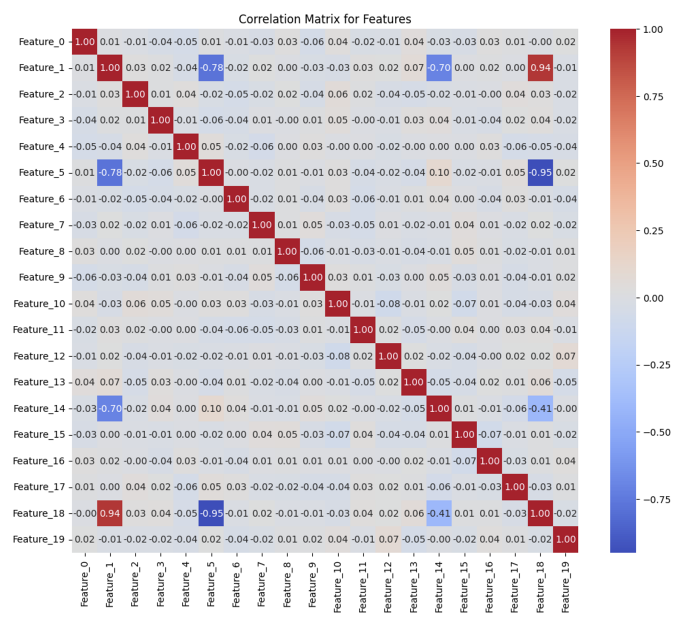

Fig. 8.

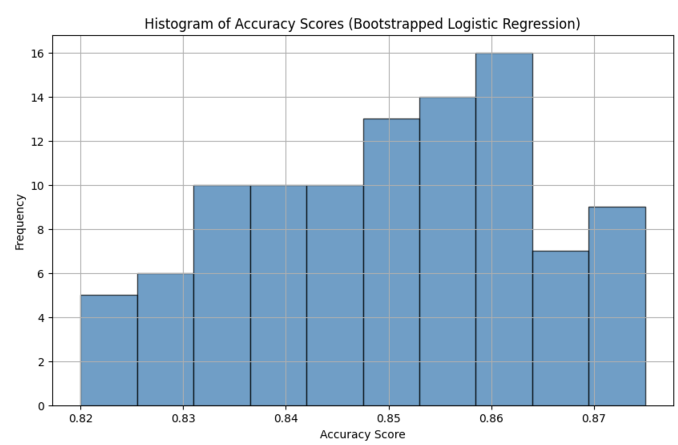

Fig. 9.

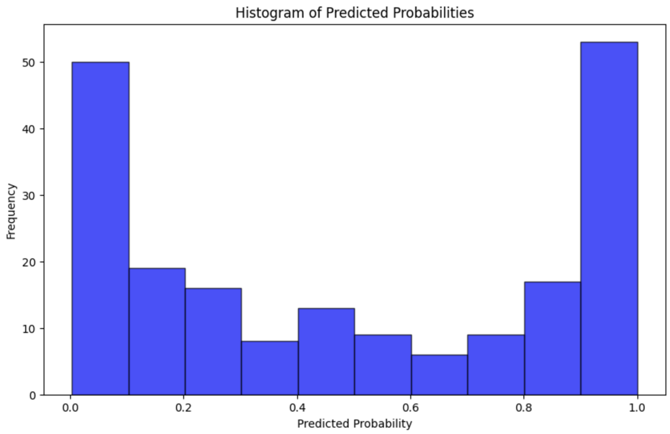

Fig. 10.

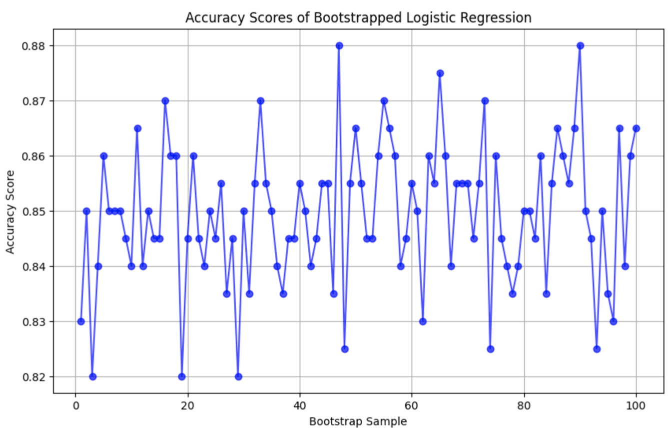

Fig. 11.

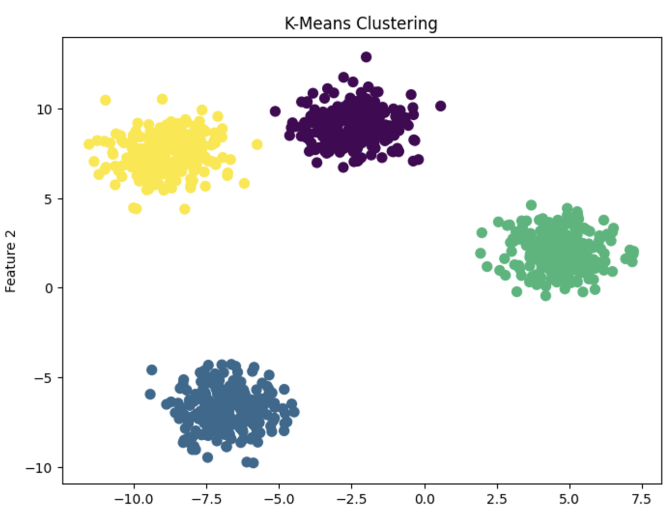

Fig. 12.

Fig. 13.

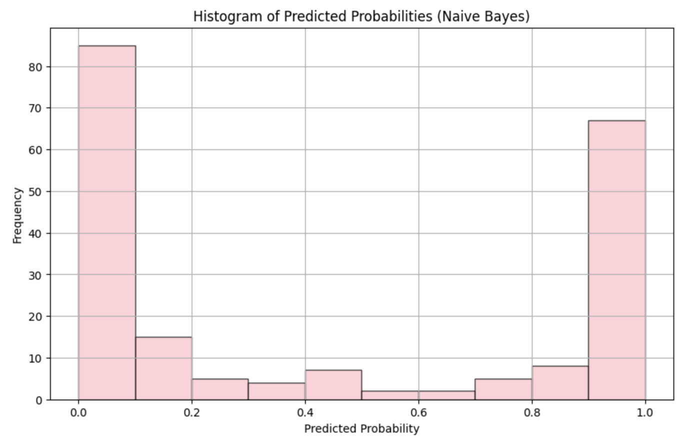

Fig. 14.

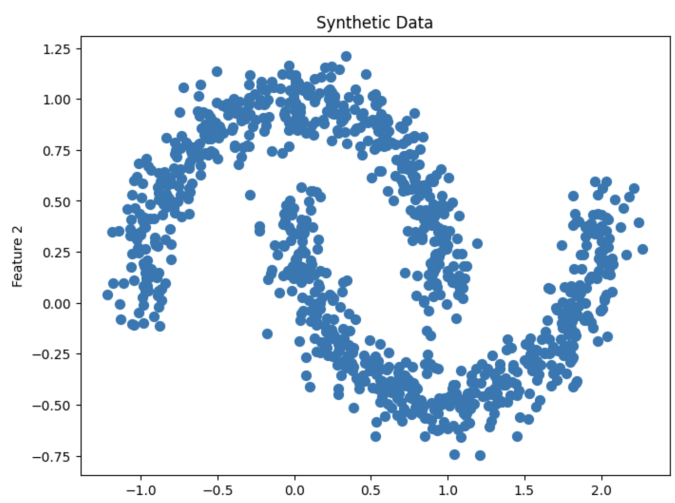

Fig. 15.

VI. Conclusion

This project demonstrates the effective application of artificial intelligence and machine learning models for predicting Tesla's stock prices and uncovering underlying patterns in its data. Various models, including regression analysis, time-series forecasting, and clustering techniques, were implemented to analyze Tesla's stock price movements.

Among the models tested, the LSTM (Long Short-Term Memory) neural network outperformed traditional regression models like Linear Regression, Ridge, and Lasso Regression in terms of predictive accuracy [12]. The LSTM model effectively captured the temporal dependencies in stock price data, making it particularly suitable for forecasting future price movements. It outperformed the ARIMA model, especially in volatile market conditions, demonstrating the superior capability of deep learning models for time-series analysis.

While regression models like Lasso Regression and Ridge Regression also performed well, especially in terms of feature selection and handling multicollinearity, they were limited in capturing the non-linear and complex patterns inherent in stock price data. Clustering techniques, such as K-Means and Hierarchical Clustering, provided useful insights into different market regimes, but they were not as effective for direct price prediction.

Overall, this study highlights the importance of advanced machine learning models, especially deep learning techniques like LSTM, in the context of stock price prediction. Future improvements could involve incorporating additional features, such as sentiment analysis or external economic indicators, to enhance model performance and prediction accuracy.

VII. References

References

[1] H. Debnath, S. Srivastava, and K. K. Jha, "Forecasting financial frontiers: A comparative analysis of lstm-based stock price prediction for meta and tesla," in *2024 First International Conference on Technological Innovations and Advance Computing (TIACOMP)*, pp. 1–4, IEEE, 2024.

[2] M. T. Islam, M. R. Islam, M. S. Faruque, S. M. D. U. Daiam, and M. M. Islam, "Comparative stock performance analysis of leading electric vehicle brands: Tesla, byd, and nio using python programming language," *European Journal of Theoretical and Applied Sciences*, vol. 2, no. 4, pp. 327–338, 2024.

[3] Y. Chi, "Predictive analysis of tesla's stock closing prices utilizing lstm and gru deep learning models,"

[4] Z. Babazhanov, M. Suleyeva, M. Zhumatov, A. Kenenbay, and D. Syrbayeva, *Tesla's financial performance and stock price success in 2020*. PhD thesis, International School of Economics KAZGUU, 2023.

[5] A. K. Gupta, V. Kumar, A. Verma, P. Yadav, N. Kumar, and M. Sain, "Unveiling stock market trends through predictive analytics and sentiment analysis: Insightfulequity," in *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, vol. 5, pp. 1558–1566, IEEE, 2024.

[6] A. Asgarov, "Predicting financial market trends using time series analysis and natural language processing," *arXiv preprint arXiv:2309.00136*, 2023.

[7] M. P. Cristescu, D. A. Mara, R. A. Nerișanu, L. C. Culda, and I. Maniu, "Analyzing the impact of financial news sentiments on stock prices—a wavelet correlation," *Mathematics*, vol. 11, no. 23, p. 4830, 2023.

[8] Y. Deng, "Future development analysis based on the price reduction trend of tesla," *Advances in Economics, Management and Political Sciences*, vol. 112, no. 1, pp. 156–162, 2024.

[9] W. Swahn, "Price formation on the tesla stock market: A study on market impact and trader types," 2021.

[10] M. Zhu, Y. Cheng, L. Huang, and N. Duan, "Predicting stock fluctuations: Comparative analysis of advanced machine learning models using tesla stock,"

[11] K. B. Sk, S. Javvadi, *et al.*, "Predictions of tesla stock price based on machine learning model," 2023.

[12] A. Bhadkamar and S. Bhattacharya, "Tesla inc. stock prediction using sentiment analysis," *Australasian Accounting, Business and Finance journal*, vol. 16, no. 5, pp. 52–66, 2022.