Project Report

Stock Market Prediction Using Machine Learning

Project Report

What is the Stock Market?

A stock market is a public market where you can buy and sell shares for publicly listed companies. The stocks, also known as equities, represent ownership in the company. The stock exchange is the mediator that allows the buying and selling of shares.

Importance of Stock Market

- Stock markets help companies to raise capital.
- It helps generate personal wealth.
- Stock markets serve as an indicator of the state of the economy.
- It is a widely used source for people to invest money in companies with high growth potential.

1. Statement of Purpose

Stock market prediction is one of the most challenging tasks due to the unpredictability of financial markets. This project aims to analyze stock price movements using machine learning models to determine whether the closing price of Tesla stocks will rise or fall. By leveraging data preprocessing, feature engineering, and predictive modeling, this study contributes to understanding financial forecasting and improving investment strategies.

---

2. Project Questions and Answers

Q1: What is the objective of this project?

The objective is to develop a machine learning model that can predict stock price movement based on historical data.

Q2: Which models are used for prediction?

The models used are Logistic Regression, Support Vector Classifier (SVC), and XGBoost Classifier.

Q3: What features are used in the prediction?

Features include:

- open-close: Difference between opening and closing price.
- low-high: Difference between lowest and highest price.
- is_quarter_end: Indicator for the quarter-end period.

Q4: How is the model performance evaluated?

Model performance is assessed using:

- ROC AUC Score
- Confusion Matrix
- Training and Validation Accuracy

Q5: What are the limitations of this approach?

- Stock prices are influenced by external factors such as news, political events, and economic indicators, which are not included in this model.

---

3. Study Methodology

Step 1: Data Collection

- The dataset contains Tesla stock price data, including open, high, low, close, and volume values.

Step 2: Data Preprocessing

- Missing values are checked and handled.
- Redundant columns are removed.
- Feature engineering is applied to extract meaningful insights.

Step 3: Data Visualization

- Closing prices are plotted over time.
- Feature distributions and outliers are analyzed.
- Correlation heatmaps help identify relationships between features.

Step 4: Feature Engineering

- Extracts day, month, year from the date.
- Computes new features like open-close and low-high.
- Creates a target variable for classification (1: Price increase, 0: Price decrease).

Step 5: Model Training and Evaluation

- Data is split into training and validation sets.
- Models are trained and evaluated using ROC AUC scores and confusion matrices.

---

4. Libraries Used

- **NumPy**: For numerical computations.
- **Pandas**: For data manipulation and analysis.
- **Matplotlib & Seaborn**: For data visualization.
- **Scikit-learn**:
    - train_test_split – Splits data into training and testing sets.
    - StandardScaler – Standardizes the dataset.
    - LogisticRegression, SVC – Machine learning models.
    - metrics – Model evaluation metrics.
- **XGBoost**: An efficient gradient boosting algorithm.
- **Warnings Filter**: Ignores unnecessary warnings.

---

5. Data Exploration & Preprocessing

Loading and Understanding the Data

- The dataset is loaded using Pandas.
- df.head() is used to inspect the first few rows.
- df.shape provides the dataset's dimensions.
- df.describe() gives statistical insights.
- df.info() helps check for missing values.

Visualizing Stock Prices

- The Close price of Tesla stocks is plotted to observe trends.
- Checking for redundant columns like Adj Close, which is removed.
- df.isnull().sum() ensures no missing values.

Feature Distributions and Outliers

- **Histograms** show the distribution of features (Open, High, Low, Close, Volume).
- **Boxplots** identify outliers in stock price movements.

Feature Engineering

- Extracts day, month, and year from the Date column.
- Creates an is_quarter_end column to mark quarter-end periods.
- Aggregates data by year to observe trends.
- Creates new features:
    - open-close: Difference between opening and closing price.
    - low-high: Difference between lowest and highest price.
    - target: 1 if the next day's closing price is higher, otherwise 0.
- A **correlation heatmap** helps identify highly correlated features.

---

6. Model Training

Data Preparation

- Selects open-close, low-high, and is_quarter_end as features.
- Uses StandardScaler to normalize the data.
- Splits the data into training and validation sets.

Machine Learning Models Used

1. **Logistic Regression**: A simple linear model for classification.
2. **Support Vector Classifier (SVC)**: A polynomial kernel-based classifier.
3. **XGBoost Classifier**: A powerful ensemble learning model.

Model Training and Evaluation

- Each model is trained on the dataset.
- Performance is measured using **ROC AUC score**.
- A **confusion matrix** visualizes classification performance.

---

7. Insights from the Project

1. **Feature Engineering is Essential**
    - Simple transformations like open-close and low-high are valuable.
    - Quarter-end features help capture market trends.

2. **Preprocessing Matters**
    - Normalizing data improves model performance.
    - Splitting data ensures better generalization.
3. **Model Performance**
    - Logistic Regression and SVC had lower accuracy.
    - XGBoost performed better due to its ability to capture complex patterns.
4. **Stock Prices are Unpredictable**
    - More financial indicators might improve prediction accuracy.

---

8. Conclusion

This project demonstrates how machine learning can be applied to financial markets. While predictions show promising results, stock prices remain highly volatile. Future improvements could include adding more technical indicators, sentiment analysis, and using deep learning models.

---

9. Future Work

- Incorporate **more technical indicators** (e.g., Moving Averages, RSI, MACD).
- Use **LSTM (Long Short-Term Memory)** models for time-series prediction.
- Integrate **news sentiment analysis** to capture external factors.

Final Thoughts

- **Machine learning can predict stock trends**, but market volatility makes it challenging.
- **Feature selection, hyperparameter tuning, and deep learning** could further improve accuracy.