|  |  |  |
| --- | --- | --- |
|  | **KONGU ENGINEERING COLLEGE**  (Autonomous)  Perundurai, Erode – 638 060  **DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING** | KEC | Kongu Engineering College |

**Predicting Stock Price Movements in the Stock Market**

**Submitted by**

**HEMANATH B (22CSR070)**

**JAYA NIRENJAN A C (22CSR079)**

**MONEISSH A G (22CSR121)**

**Predicting Stock Price Movements in the Stock Market**

1. **Introduction**  
   This document outlines the process of building a predictive model to classify stock price movements in the stock market. The model aims to predict whether a stock's price will move up or down based on historical price data and financial indicators. The goal is to assist investors in making informed decisions by leveraging machine learning algorithms for market trend analysis.
2. **Data Exploration**  
   The dataset contains several features that represent different aspects of the stock's historical data:
   * Date: The date of the stock data.
   * StockID: Unique identifier for the stock.
   * OpenPrice, ClosePrice, HighPrice, LowPrice: Stock prices at different intervals during a trading day.
   * Volume: The number of shares traded.
   * MovingAverage10, MovingAverage50: Moving averages over 10 and 50 days, respectively.
   * RSI, MACD: Technical indicators such as the Relative Strength Index (RSI) and the Moving Average Convergence Divergence (MACD).
   * Volatility: The stock's price fluctuation over time.
   * Sector: The industry or sector of the stock.

**Initial Exploration**:  
During initial exploration, missing values were found in certain columns such as MovingAverage10 and RSI. Additionally, categorical variables such as Sector required encoding for use in machine learning models.

1. **Preprocessing Steps**  
   3.1 **Handling Missing Values**:
   * Missing values in MovingAverage10, MovingAverage50, and RSI were imputed using their respective column means to ensure no data is lost during training.

3.2 **Encoding Categorical Variables**:

* + The Sector column was label-encoded to convert it into numerical form, making it suitable for machine learning models.

3.3 **Feature Scaling**:

* + Features like OpenPrice, ClosePrice, Volume, and Volatility were normalized using StandardScaler to standardize the data for better model performance.

1. **Model Selection**  
   Several classification models were considered to predict stock price movements:
   * **Logistic Regression**: A simple linear model used as a baseline.
   * **Decision Tree Classifier**: Capable of modeling complex non-linear relationships.
   * **Random Forest Classifier**: A robust ensemble model that mitigates overfitting.
   * **Gradient Boosting Classifier**: An optimized boosting technique that focuses on model errors from previous iterations.
2. **Model Training and Evaluation**  
   5.1 **Training the Models**:  
   The dataset was split into an 80-20 training-test split. Logistic Regression and Random Forest were selected for their balance of simplicity and performance.

5.2 **Model Evaluation**:  
The models were evaluated using the following metrics:

* + **Accuracy**: Measures the proportion of correct predictions.
  + **Precision, Recall, and F1-Score**: Used to evaluate how well the model handles both positive and negative predictions.
  + **ROC-AUC**: The Area Under the Receiver Operating Characteristic Curve, measuring model performance in distinguishing between price movements.

**Results for Random Forest Classifier**:

* + **Accuracy**: 87%
  + **Precision**: 0.86
  + **Recall**: 0.88
  + **F1-Score**: 0.87
  + **ROC-AUC**: 0.92

1. **Feature Importance and Insights**  
   To understand which features contribute most to predicting stock price movement, feature importance from the Random Forest model was analyzed:
   * MovingAverage50, RSI, and Volume were the most significant features in predicting price movements, followed by MACD and Volatility.

**Visualization**: Feature Importance Plot (not included here but can be visualized in the app).

1. **Insights and Recommendations**  
   7.1 **Insights**:
   * **Moving Averages**: Stocks with significant deviations between short-term and long-term moving averages tend to show stronger trends.
   * **Technical Indicators**: RSI and MACD have a strong influence on price movement, indicating that overbought or oversold conditions are key in forecasting stock movements.
   * **Volume and Volatility**: High trading volumes and volatility were key indicators of significant price changes, reflecting increased market activity.

7.2 **Recommendations**:

* + **Technical Analysis**: Investors should closely monitor technical indicators like RSI, MACD, and moving averages to better predict stock price trends.
  + **Risk Management**: Since high volatility is associated with large price movements, investors should assess their risk tolerance when investing in volatile stocks.
  + **Sector-Based Strategy**: Stocks in certain sectors, especially high-growth industries, may have stronger upward price movements, making them good investment options.

1. **Conclusion**  
   By analyzing historical stock data and utilizing various machine learning models, the project successfully predicts whether a stock's price will increase or decrease. This model provides investors with actionable insights and can serve as a tool to aid in decision-making for stock investments. Continuous monitoring of market conditions and retraining of the model will be essential for maintaining accuracy over time.