Machine Learning using Python

Project Report–[Hybrid ML Models for Stock and Bitcoin Prize Prediction]

College Name : Government College of Engineering -Bodinayakanur

Team Name : Jupyter Geniuses

Team Members:

1. 21CS103 - Almatheen Divan P
2. 21CS130 - Nitheesh Kumar K
3. 21CSTR301 - Mariyamagizhan B
4. 22CSL06 - Sriram S

1. Abstract :

The financial markets have witnessed a surge in the popularity of cryptocurrencies alongside traditional stock trading. This study explores the development of hybrid machine learning models to predict stock and Bitcoin prices, leveraging the strengths of various algorithms to enhance predictive accuracy. Utilizing Python as the primary programming language, we integrate multiple techniques, including regression analysis, decision trees, and neural networks, to create a robust framework. The dataset encompasses historical price data, trading volumes, and relevant economic indicators. By employing feature engineering and optimization techniques, we improve model performance and reduce over fitting. The results demonstrate that hybrid models outperform individual algorithms in terms of predictive accuracy and reliability. This research contributes to the growing field of financial forecasting by providing insights into effective methodologies for predicting market trends, ultimately aiding investors and traders in making informed decisions.

2. Model

Description: This section involves defining a set of machine learning models for classification and forecasting tasks. The models used are:

- **Logistic Regression:** A linear model for binary classification.
- **Support Vector Classification (SVC):** A model that finds the hyperplane maximizing the margin between classes. Uses a polynomial kernel in this case.
- **XGBoost Classifier:** An optimized gradient boosting algorithm.
- **Decision Tree Classifier:** A model that splits data based on feature values to make predictions.
- **Random Forest Classifier:** An ensemble of decision trees that improves accuracy by averaging predictions.
- **Gradient Boosting Classifier:** An ensemble technique that builds models sequentially to improve prediction accuracy**.**
- **AdaBoost Classifier:** An ensemble method that combines weak classifiers to form a strong classifier.
- **K-Nearest Neighbors (KNN):** A model that classifies data based on the majority class among its k-nearest neighbors.
- **Gaussian Naive Bayes:** A probabilistic classifier based on Bayes' theorem with Gaussian assumptions.
- **Quadratic Discriminant Analysis (QDA):** A classification technique based on quadratic decision boundaries.
- **Linear Discriminant Analysis (LDA):** A linear classification method that finds the linear combination of features that best separates classes**.**
- **Extra Trees Classifier:** An ensemble method that uses randomized decision trees to improve accuracy.

Purpose: The models listed will be trained and evaluated on Tesla and Bitcoin datasets to determine which model performs best for predicting future stock prices or market behavior.

3. Algorithm Implementation

Description: This section involves preparing data, training the models, and evaluating their performance.

Data Preparation:

1. **Tesla Data:**
    - The dataset is loaded, and missing values are checked.

- Date columns are converted to datetime objects.
- New features are engineered to capture additional insights.
- Features are standardized, and data is split into training and validation sets.

2. **Bitcoin Data:**
    - Similar preprocessing steps as Tesla data: loading, handling missing values, feature engineering, and standardization.
3. **Model Training:**
    - Each model is trained using the training set for both Tesla and Bitcoin data.
    - Model performance is evaluated using ROC AUC scores for both training and validation sets.

Forecasting with Prophet:

1. **Data Preparation for Prophet:**
    - Historical data is prepared by renaming columns and converting dates.
    - Prophet is used to forecast future values for Tesla, Bitcoin, and Microsoft stock prices.
2. **Forecasting:**
    - Prophet models are fit on historical data, and forecasts for the next 10 years are generated.

Purpose: The goal is to assess the performance of various machine learning models in predicting stock price movements and to provide forecasts for future values using Prophet.

4. Predication Comparison Report

Description: This section involves comparing the predictions of different models based on their performance metrics.

Evaluation Metrics:

- **ROC AUC Score:** Used to evaluate the models' ability to distinguish between classes. Higher scores indicate better model performance.

Confusion Matrices:

- Confusion matrices for each model are plotted to visualize the true positives, true negatives, false positives, and false negatives for both Tesla and Bitcoin datasets.

Forecast Visualization:

- Forecasts from Prophet are plotted alongside historical data to visualize predicted trends and uncertainties for Tesla, Bitcoin, and Microsoft stock prices.

Purpose: To provide a clear comparison of model performance and the quality of forecasts. This helps in selecting the best model and understanding the reliability of predictions.

5. Final Prediction

Description: This section summarizes the final predictions made by the models and the forecast results.

Model Performance:

- **Best Models:** Based on ROC AUC scores and confusion matrices, the models with the highest performance are identified.
- **Forecast Results:** Prophet's forecasts provide predictions of stock prices for the next 10 years for Tesla, Bitcoin, and Microsoft.

Purpose: To highlight the most effective models and the accuracy of long-term forecasts. This helps in making informed decisions based on predicted trends.

6. Conclusion

Description: This section provides a summary of the overall findings and insights derived from the analysis.

Findings:

1. **Model Performance:**
    - Certain models (e.g., Random Forest, XGBoost) might outperform others based on ROC AUC scores and confusion matrices.
    - Performance varies between Tesla and Bitcoin datasets, reflecting differences in data characteristics.
2. **Forecasting Results:**
    - Prophet provides a useful forecast for long-term trends, showing projected stock prices for Tesla, Bitcoin, and Microsoft.
    - The forecast includes uncertainty intervals, indicating the range within which future prices are expected to fall**.**

Insights:

- **Model Selection:** Based on performance metrics, some models are better suited for predicting stock price movements. Model selection should be guided by these metrics.
- **Forecasting Utility:** The forecasts from Prophet can guide long-term investment decisions by providing insights into future price trends.

Purpose: To summarize the results and offer actionable insights based on model performance and forecasts. This helps stakeholders in making informed decisions and understanding the limitations of the analysis.

SUMMARY:

The code performs a comprehensive analysis of Tesla, Bitcoin, and Microsoft stock data using various machine learning models and forecasting techniques. It begins by loading and preparing the data, including handling missing values, feature engineering, and standardizing features. The code then trains and evaluates multiple classification models, such as Logistic Regression and Random Forest, to predict stock price movements, using ROC AUC scores and confusion matrices for performance assessment. Additionally, it employs the Prophet library to forecast stock prices for the next decade, visualizing both historical data and future predictions with uncertainty intervals. This approach integrates predictive modeling and long-term forecasting to analyze and project stock trends effectively.