Final Project Report

1. Introduction

Cryptocurrency markets are known for their high volatility and unpredictable behavior. Rapid price fluctuations make it difficult for traders and investors to assess risk accurately. The objective of this project is to analyze historical cryptocurrency market data and develop a machine learning model capable of predicting price volatility. Accurate volatility prediction can help in risk management, portfolio planning, and informed decision-making.

---

2. Problem Statement

The cryptocurrency market experiences frequent and sharp price movements. Understanding and predicting volatility is essential for identifying high-risk periods. This project aims to build a machine learning-based system that predicts cryptocurrency volatility using historical price, volume, and market capitalization data.

---

3. Dataset Description

The dataset consists of historical daily cryptocurrency market data stored in a CSV file. Each record represents one trading day.

Dataset Attributes: - Timestamp - Date - Cryptocurrency Name - Open Price - High Price - Low Price - Close Price - Trading Volume - Market Capitalization

The dataset is time-series in nature and is used to extract patterns related to price movements and volatility.

---

4. Methodology

4.1 Data Preprocessing

- Removed unnecessary unnamed index column
- Checked and confirmed absence of missing values
- Converted date and timestamp columns to datetime format
- Sorted data based on date
- Applied feature scaling using StandardScaler on numerical columns

---

4.2 Feature Engineering

To enhance model performance, several new features were created: - Daily volatility using price range - Rolling volatility using 7-day and 14-day windows - Moving averages (7-day and 14-day) - Liquidity ratio (volume / market capitalization)

These features help capture market trends, price fluctuations, and trading activity.

4.3 Exploratory Data Analysis (EDA)

EDA was performed to understand data distribution and relationships: - Analyzed statistical summaries of numerical features - Observed high variation in price and volume - Performed correlation analysis using heatmaps - Identified strong correlation among price-based features

EDA insights guided feature selection and model design.

4.4 Model Building

The Random Forest Regressor algorithm was selected due to its robustness and ability to handle non-linear relationships. The dataset was split into training and testing sets using a 70:30 ratio.

4.5 Model Optimization

Hyperparameter tuning was performed using GridSearchCV to improve prediction accuracy. Different combinations of model parameters were tested, and the best-performing model was selected for final evaluation.

---

5. Model Evaluation

The model performance was evaluated using regression metrics: - R^2 Score - Mean Absolute Error (MAE) - Mean Squared Error (MSE)

The optimized model demonstrated improved accuracy compared to the base model, indicating effective learning of volatility patterns.

---

6. Results and Discussion

The trained model successfully predicted cryptocurrency volatility based on engineered features. Rolling volatility and liquidity-related features contributed significantly to model performance. The results confirm that historical price behavior and trading activity play a major role in volatility prediction.

7. Limitations

- The model is trained on historical data and may not capture sudden market shocks
- External factors such as news and regulatory events are not included
- Prediction accuracy may vary during extreme market conditions

---

8. Future Scope

- Incorporating sentiment analysis from news and social media
- Using deep learning models such as LSTM for time-series forecasting
- Extending the system to support multiple cryptocurrencies
- Deploying the model using a web-based interface

---

9. Conclusion

This project demonstrates an effective approach to cryptocurrency volatility prediction using machine learning. By combining data preprocessing, feature engineering, EDA, and model optimization, the system provides meaningful insights into market behavior. The proposed model can assist traders and analysts in understanding market risk and making informed decisions.

---

10. References

- Scikit-learn Documentation
- Pandas and NumPy Documentation
- Cryptocurrency Market Data Sources