

Final ML Project Report

Project Title:

Cryptocurrency Trend Analysis Using Machine Learning

1. Objective

To analyze and predict trends in cryptocurrency prices using machine learning by combining datasets from different dates, preprocessing them, training predictive models, and deploying a basic user interface.

2. System Architecture (HLD Summary)

Key Modules:

1. Data Collection
2. Data Preprocessing
3. Exploratory Data Analysis (EDA)
4. Feature Engineering
5. Model Training & Evaluation
6. Prediction & Inference
7. GUI Deployment via Gradio

Technology Stack:

- Language: Python
- Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, gradio
- Platform: Jupyter Notebook
- Data Source: CoinGecko (CSV files for March 16 & 17, 2022)

3. Low-Level Design (LLD Summary)

3.1 Data Collection:

```
df_16 = pd.read_csv("coin_gecko_2022-03-16.csv")  
df_17 = pd.read_csv("coin_gecko_2022-03-17.csv")
```

```
df = pd.concat([df_16, df_17], ignore_index=True)
```

3.2 Data Preprocessing:

- Handle missing values (dropna or fillna)
- Remove duplicates
- Type conversion (e.g., strings to floats)
- Rename columns if needed

3.3 EDA:

- Visualize price changes
- Correlation heatmaps
- Rank distributions
- Identify top gainers and losers

3.4 Feature Engineering:

- price_change_percentage_24h
- market_cap_rank
- total_volume
- Date/time parsing if applicable

3.5 Model Training:

- Algorithms used: Linear Regression, Decision Trees, Random Forest
- Data split using train_test_split

3.6 Model Validation:

- Metrics used:
 - Mean Absolute Error (MAE)
 - Mean Squared Error (MSE)
 - Root Mean Squared Error (RMSE)
 - R^2 Score

3.7 Hyperparameter Tuning:

- Used Grid Search or Randomized Search
- Optimization via Cross-Validation (CV)

3.8 Model Testing & Validation:

- Validate performance on test data
- Ensures generalization capability

3.9 Deployment:

- Local UI using Gradio
- Accepts inputs and provides predictions in real-time

4. Pipeline Flow Summary

Data Collection



Data Preprocessing



Exploratory Data Analysis (EDA)



Feature Engineering



Model Training



Hyperparameter Tuning



Model Evaluation



Model Testing



Deployment (Gradio GUI)

5. Final Outcomes

- Cleaned Dataset: Merged and preprocessed
- Insights: Trend visualization and correlation patterns
- Model Output: Predictive results with performance metrics
- UI Deployment: Gradio interface for live testing

6. Assumptions & Constraints

- Data from CoinGecko is consistent in structure
- Two-day data limits long-term prediction accuracy
- Cryptocurrency data is inherently volatile