

Crypto Price Movement Classification

Virendra Beniwal

July 2025

1 Introduction

The goal of this project is to develop a machine learning pipeline that classifies the short-term price movement of a cryptocurrency (e.g., BTC/USD) based on recent market behavior. Although cryptocurrency markets are inherently noisy and difficult to predict, this project demonstrates the process of building a predictive model from scratch using historical data.

2 Problem Statement

We aim to classify the future 5-minute price return into one of six discrete bins. Each class corresponds to a return range:

- **Class 0:** Large drop (< -0.001)
- **Class 1:** Medium drop (-0.001 to -0.0005)
- **Class 2:** Small drop (-0.0005 to 0)
- **Class 3:** Small gain (0 to 0.0005)
- **Class 4:** Medium gain (0.0005 to 0.001)
- **Class 5:** Large gain (> 0.001)

These thresholds were chosen arbitrarily for demonstration purposes and can be adjusted to reflect different trading strategies.

3 Data Preprocessing and Feature Engineering

- Raw data consisted of OHLCV (Open, High, Low, Close, Volume) values per minute.
- Future return over the next 5 minutes was calculated and used to assign class labels.
- Features included price slopes over rolling windows of 5, 10, and 30 minutes.
- All features were scaled using `StandardScaler`.
- Non-numeric columns and missing/infinite values were dropped.

4 Modeling

A `RandomForestClassifier` from `scikit-learn` was used with the following hyperparameters:

- `n_estimators=100`
- `max_depth=8`
- `min_samples_leaf=10`
- `class_weight='balanced'`
- `random_state=42`

The model was trained on 80% of the data and evaluated on the remaining 20%.

5 Results

Accuracy and Log Loss

- **Accuracy:** 28.81
- **Log Loss:** 1.6662

Classification Report

	precision	recall	f1-score	support
0	0.27	0.58	0.37	16087
1	0.14	0.15	0.14	5929
2	0.37	0.02	0.04	8947
3	0.33	0.02	0.04	8917
4	0.19	0.09	0.12	6912
5	0.37	0.42	0.39	17998
accuracy			0.29	64790
macro avg	0.28	0.21	0.18	64790
weighted avg	0.30	0.29	0.24	64790

6 Visualization

The following plot shows the predicted class probabilities for each time step in the test set.

7 Conclusion

The model's accuracy and f1-scores are low, as expected given the volatility and randomness of cryptocurrency markets at minute-level granularity. However, the pipeline successfully demonstrates:

- Feature extraction from time-series financial data
- Binning continuous returns for classification
- Training and evaluating a multi-class probabilistic model

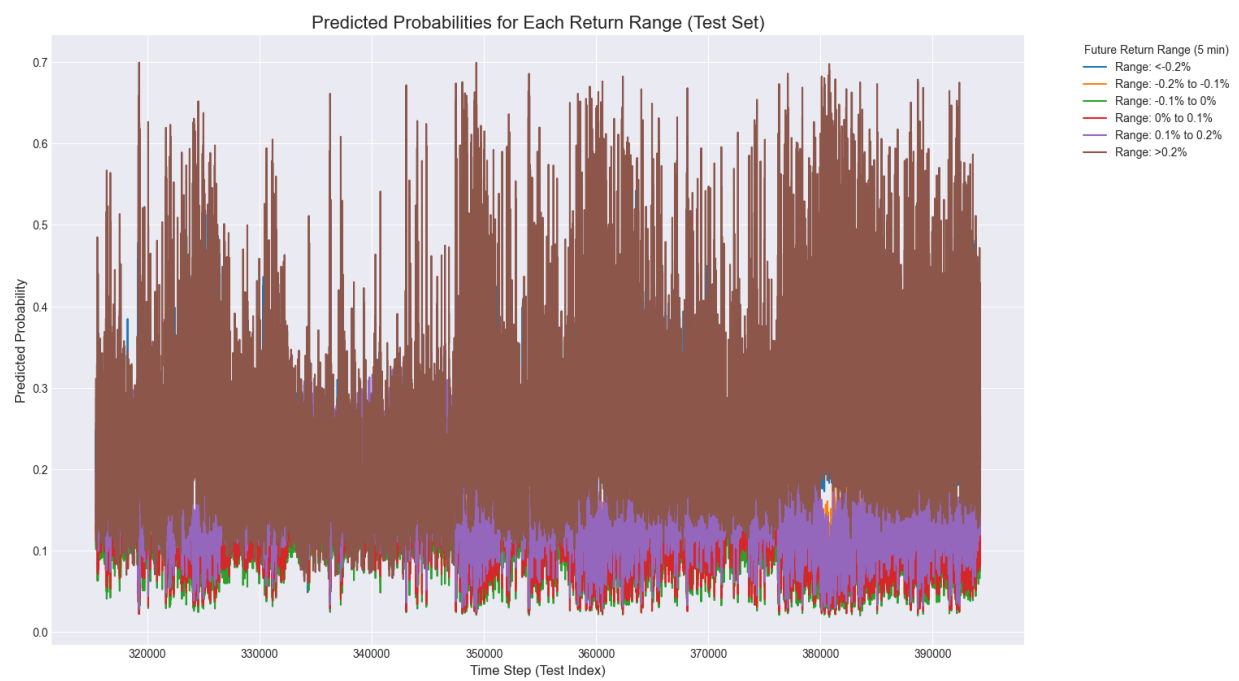


Figure 1: Predicted Class Probabilities Over Time (Test Set)