ETHEREUM WEEKLY VOLATILITY PREDICTION USING MARKET, SENTIMENT, AND ON-CHAIN DATA

Yunhui Dong

INTRODUCTION

This project focuses on predicting Ethereum weekly volatility in the short term. By combining market data (ETH price from Yahoo Finance), investor sentiment (Fear & Greed Index), and on-chain usage metrics (Ethereum transaction fees) over the past 500 days, we develop models to predict whether Ethereum's price will increase in the following seven days.

Though visualization ans statistical analysis, the project aims to identify the most influential factors behind price movements, understand the relationships between sentiment, network activity, and market behavior, and evaluate how effectively machine learning models can capture these patterns.

DATA SOURCE

Date Range: most recent 500 days

| Dataset	Type	Description	Purpose
ETH price from Yahoo Finance	API Call	Daily Ethereum price data, including Close, High, Low, Open, Volume	Capture market behavior
Fear & Greed Index from CoinMarketCap	API Call	Daily fear & greed index, including Value(fear & greed index), Value_classification	Quantify market sentiment
Ethereum transaction fees from DefiLlama	API Call	Daily total transaction fees, including Value(transaction fees)	Represent on-chain activity and network usage

OVERVIEW

- **Ethereum Price**
ETH experienced large fluctuations over the 500-day period, dropping sharply in early 2025, rebounding strongly mid-year and entering another decline toward late 2025.

- **Fear & Greed Index**
Investor sentiment showed repeated cycles of fear and greed, with sharp dips often aligning with market pullbacks and steady recoveries during price rebounds.
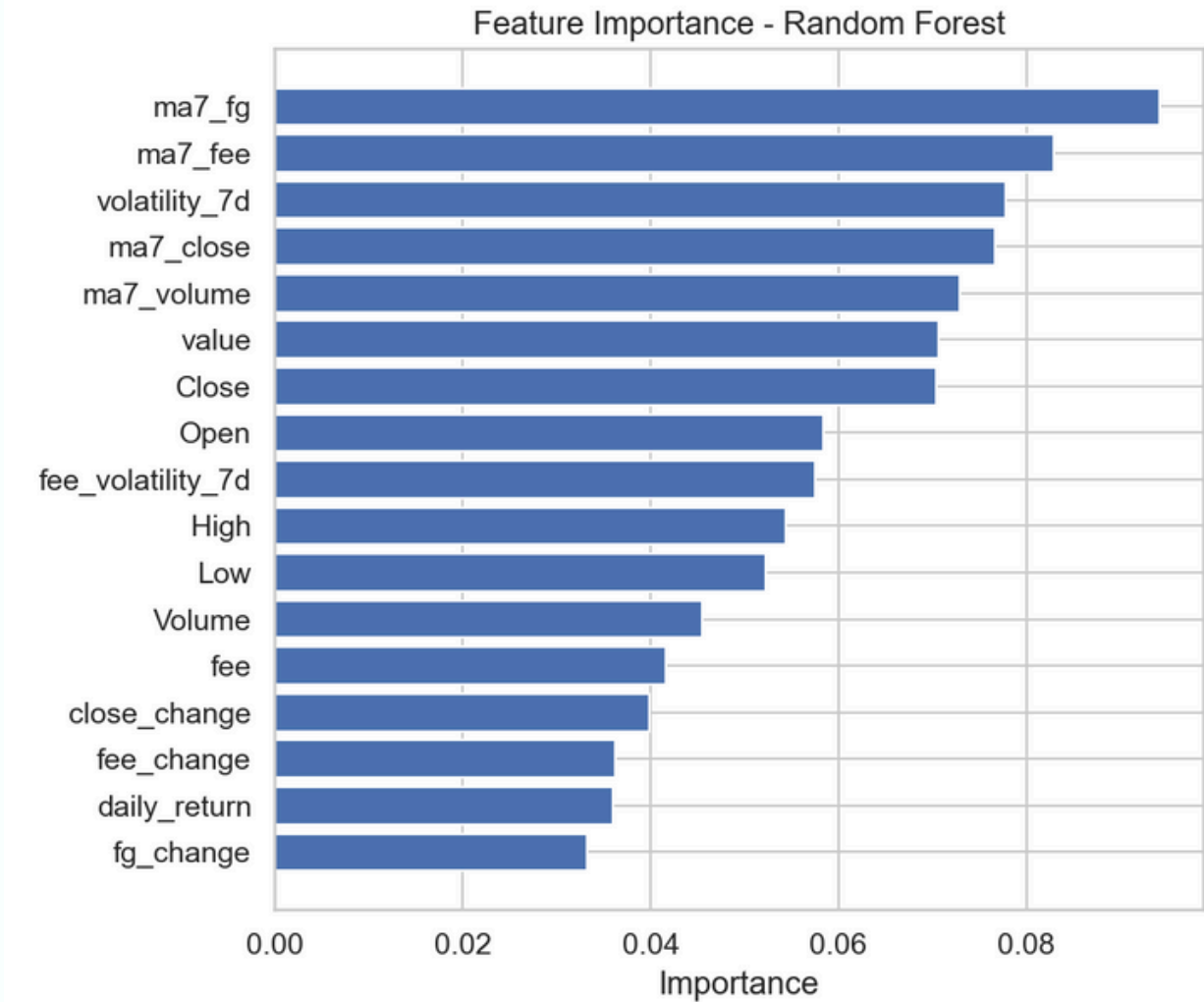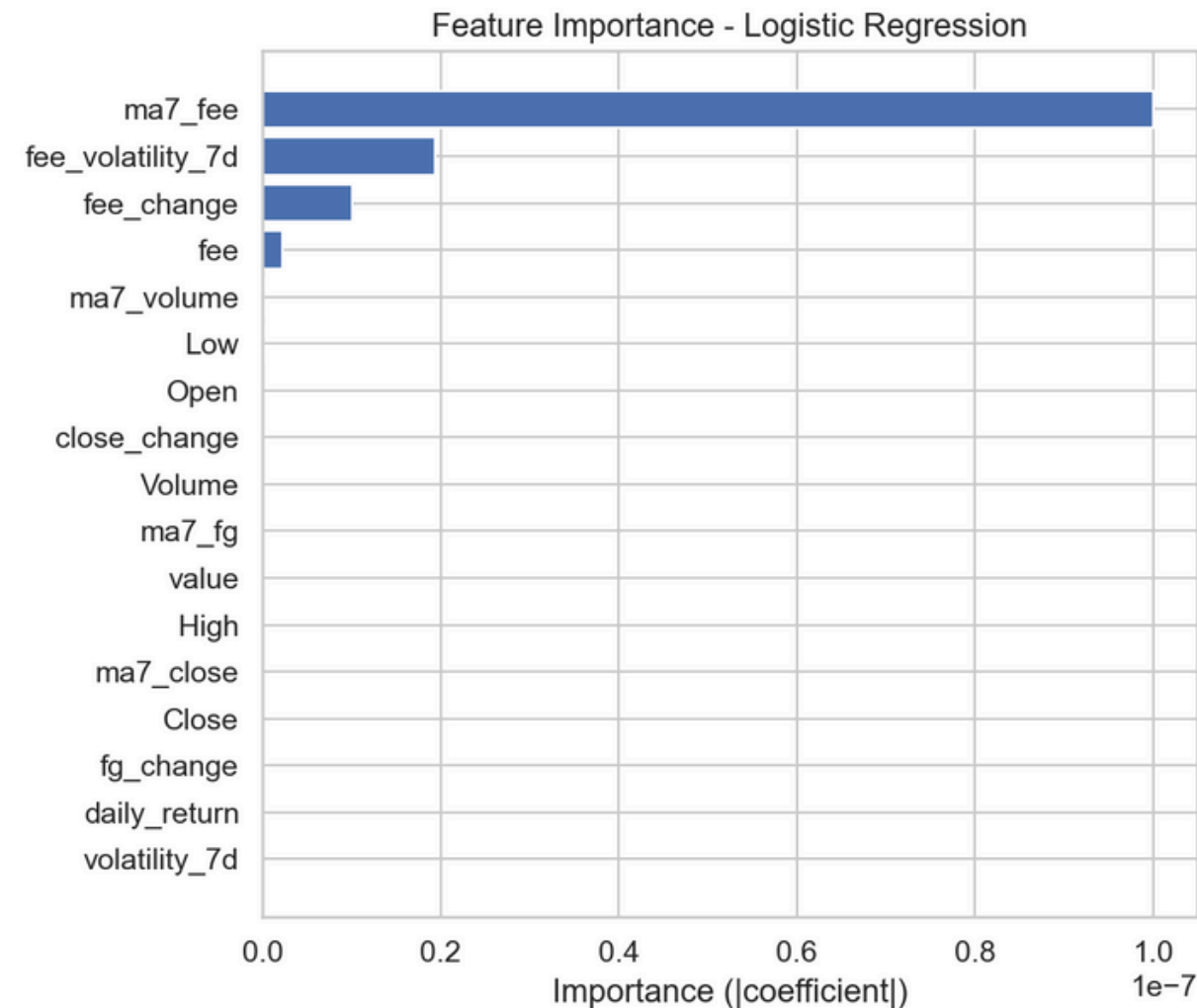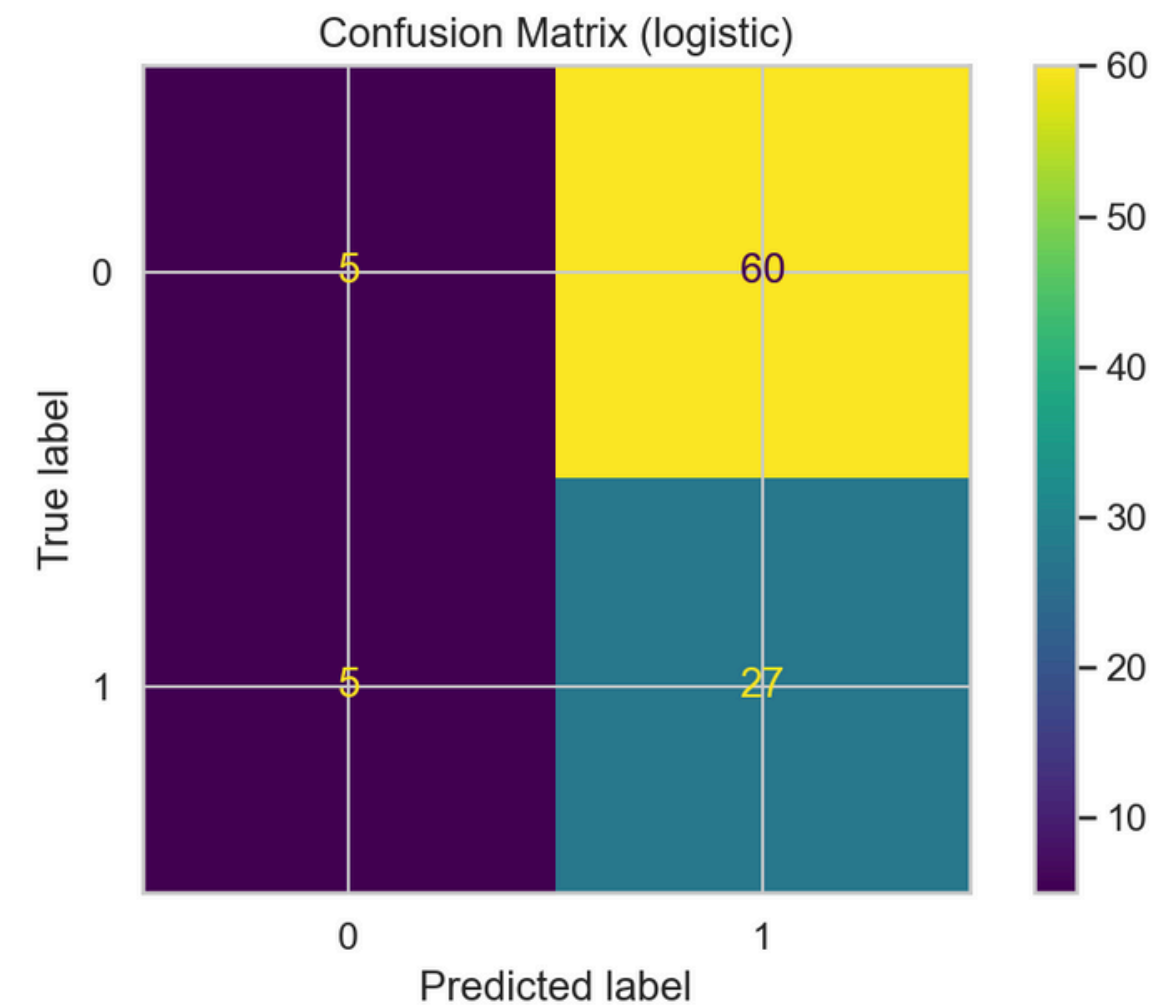
- **Transaction Fees**
Daily transaction fees fluctuated heavily indicating several periods of intense on-chain activity (often preceding short-term volatility in ETH price).

ETH Price Trend (Last 500 Days)

Fear & Greed Index Trend

Ethereum Fees Trend (On-chain Usage Proxy)
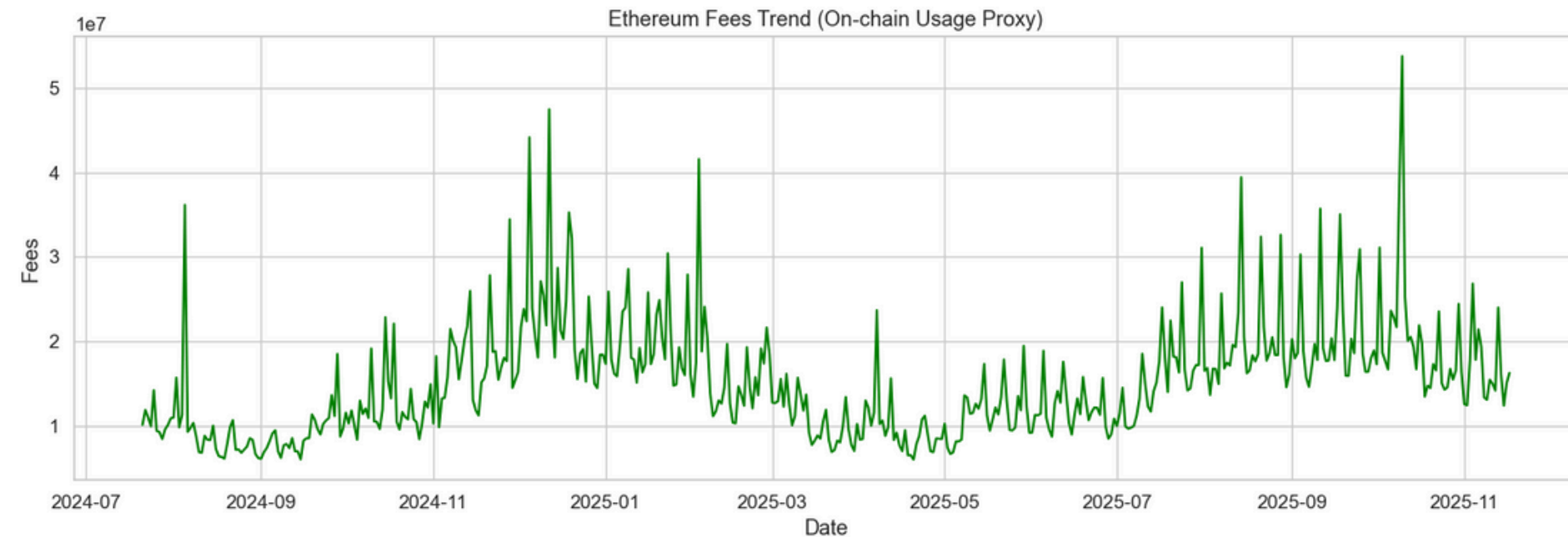
MODEL FEATURES & PREDICTIVE POWER
Which Factors Matter?

- On-chain and sentiment signals (MA7 fees, MA7 FG, volatility) rank among the strongest predictors.
- Random Forest captures non-linear relationships, while Logistic Regression reveals weak but interpretable linear effects.
- Confusion matrix shows the difficulty of predicting 'up' weeks.

# STATISTICAL STRUCTURE OF THE ETHEREUM MARKET
## Why Prediction Is Hard?
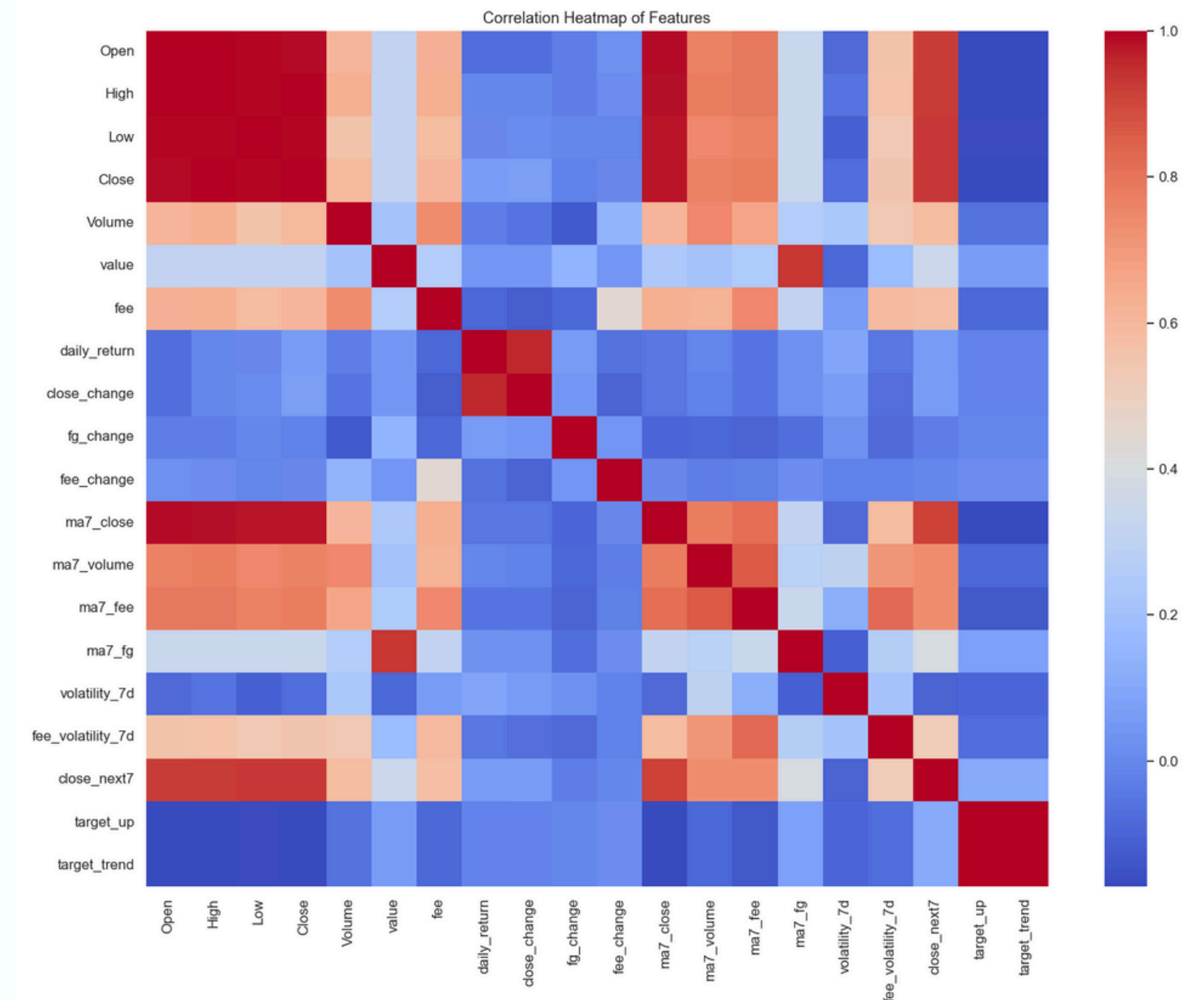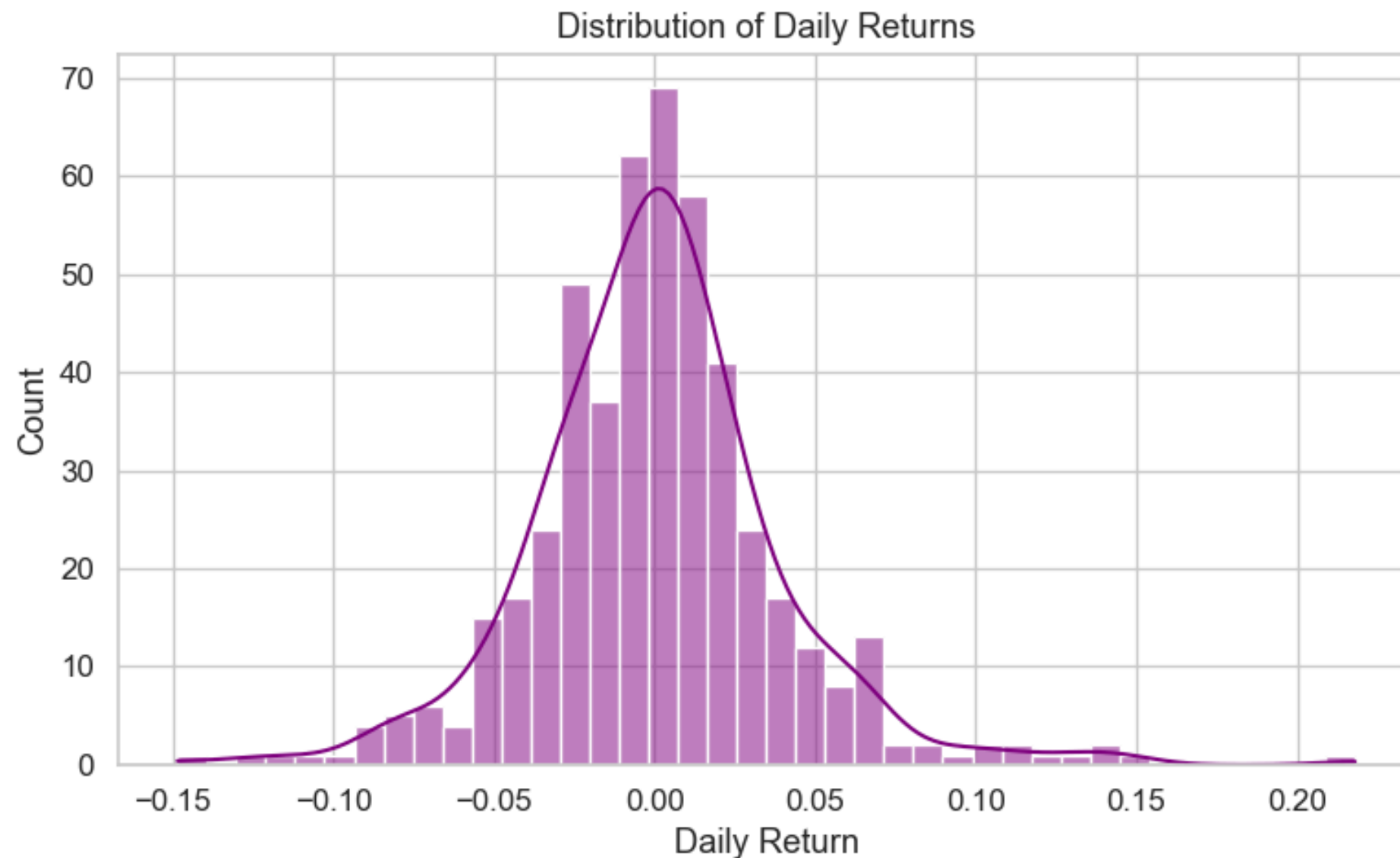
- Daily returns follow a heavy-tailed distribution, with many small moves and occasional jumps.
- Strong correlations exist among price-based features.
- On-chain and sentiment features have weaker correlations (complementary but noisy signals).

PREDICTION BEHAVIOR & MODEL LIMITATIONS
How the Model Behaves in Reality?

- Predicted probabilities cluster around 0.5–0.7, showing limited confidence due to noisy short-term signals.
- Timeline comparison reveals frequent mismatches, especially during trend reversals or high-volatility periods.
- Model tends to overpredict 'up' movements, reflecting crypto's upward bias but also creating false positives.

Predicted Probability Distribution (logistic)

True vs Predicted target_up (logistic)
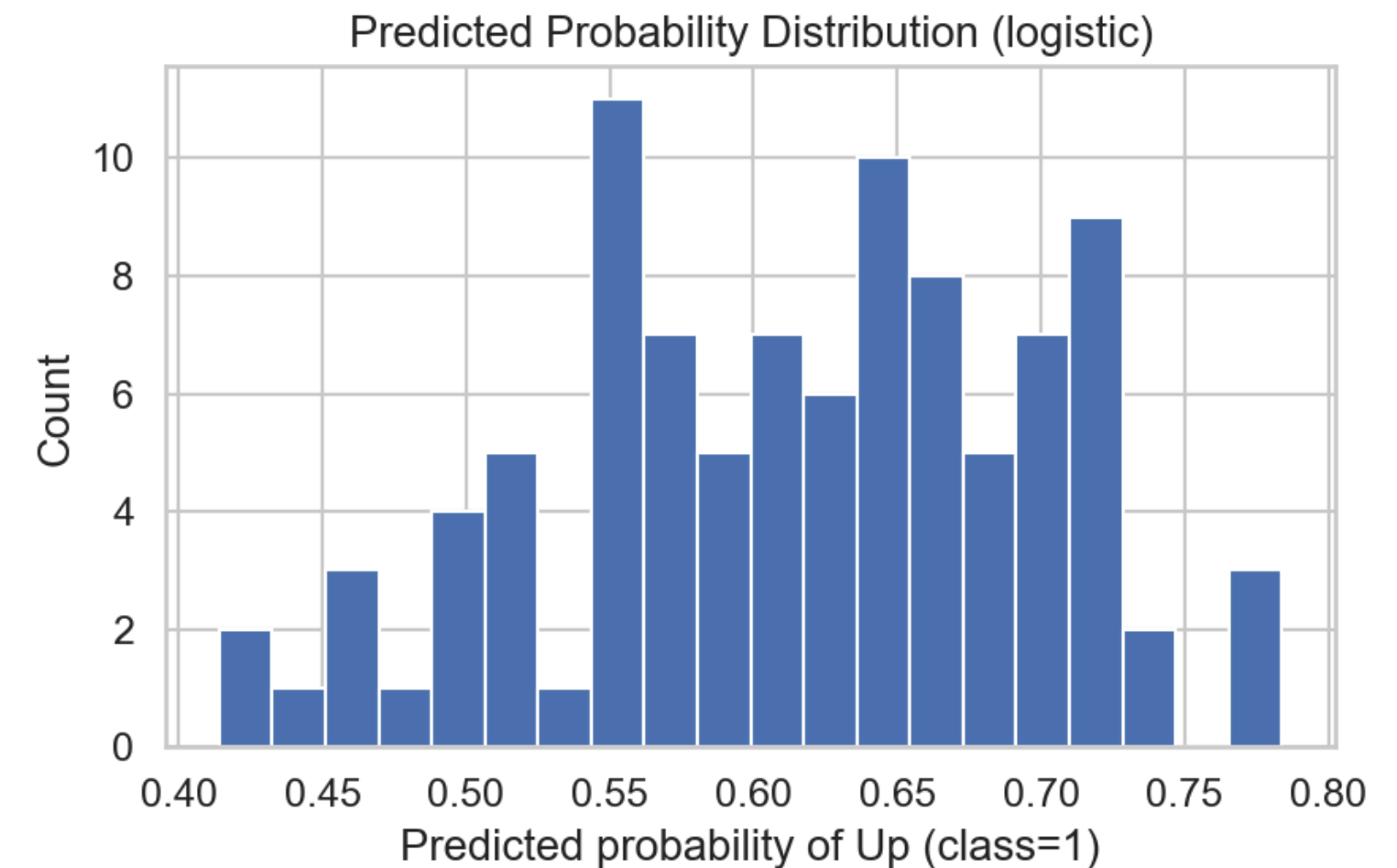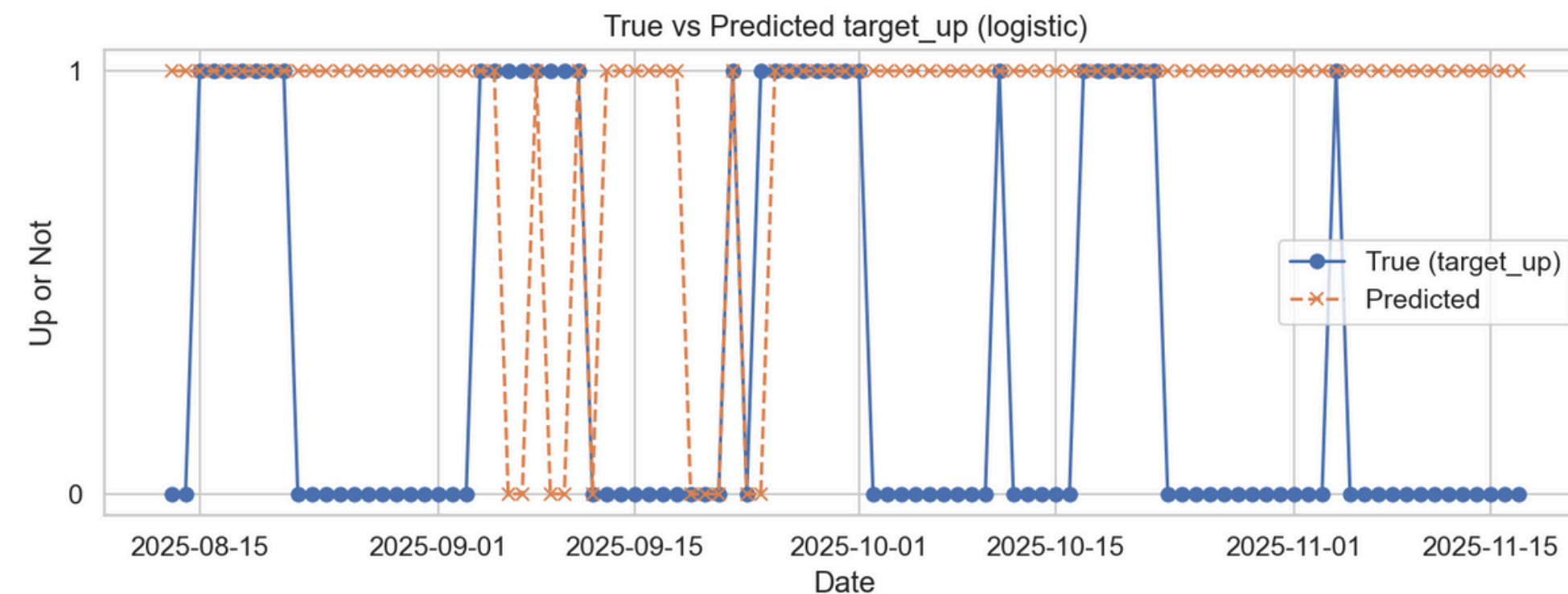
KEY CHALLENGES IN PREDICTING WEEKLY ETH MOVEMENTS

1. Market volatility is extremely high and patterns are unstable
   - Cryptocurrency prices are highly volatile and noisy, making weekly price movements extremely difficult to predict.
   - Short-term pumps/dumps break historical patterns and reduce model reliability.
2. Limited data for weekly prediction (only 485 samples)
   - Weekly-labeling compresses 500 daily rows into 485 usable rows.
   - This small dataset can lead to overfitting and prevent the model from learning deeper nonlinear relationships.
3. Difficult to detect a positive week due to class imbalance & weak signals
   - "Up" weeks occur less consistently, causing the model to default to predicting 0 (no increase).
   - Sentiment and on-chain metrics help, but signals are still too weak to achieve high recall.

THANK YOU