451 Feature Engineering: Programming Assignment 1

Prepared by Shruti Kalaskar

October 5, 2025

1. Problem Description

The objective of this project is to build a machine learning model that predicts the direction (up or down) of next-day returns for a financial asset using daily price and volume data. Inspired by a prior analysis on WTI oil futures, this implementation focuses on **Apple Inc. (AAPL)** stock and explores the predictive value of engineered features from historical prices.

Using a combination of feature engineering, subset selection based on the Akaike Information Criterion (AIC), and XGBoost classification, we aim to explore whether short-term price movements can be accurately forecasted.

2. Data Preparation and Feature Engineering

Data was retrieved using the yfinance Python package, covering AAPL stock from **January 1, 2010** to **October 5, 2025**. The raw data includes:

- Open
- High
- Low
- Close
- Volume

From this, we engineered **15 features** based on financial domain knowledge:

Category	Features
Lagged Close	CloseLag1, CloseLag2, CloseLag3
High–Low	HMLLag1, HMLLag2, HMLLag3
Open–Close	OMCLag1, OMCLag2, OMCLag3
Volume	VolumeLag1, VolumeLag2, VolumeLag3
Exponential MA	CloseEMA2, CloseEMA4, CloseEMA8

The **target variable** is a binary label:

- 1 if the next day's return is **positive**
- 0 otherwise

3. Feature Selection Using AIC

- We used an **exhaustive wrapper method** to evaluate all possible subsets of the 15 features based on AIC. The best single feature according to AIC was OMCLag2. However, recognizing the limitations of single-feature models, we constructed a **5-feature subset** using both AIC insights and financial intuition:
- ['OMCLag2', 'VolumeLag1', 'CloseLag2', 'HMLLag1', 'CloseEMA8']
- This set attempts to capture short-term return structure (lagged close), volatility (HML), momentum (EMA), and volume-based trading signals.

4. Model Training and Cross-Validation

We trained an **XGBoost classifier** using a 5-fold **time series cross-validation** strategy, which avoids lookahead bias. A randomized grid search was used to tune hyperparameters such as:

- max_depth
- learning_rate
- min_child_weight
- n_estimators
- subsample

Despite thorough tuning, the resulting model showed weak predictive power.

5. Model Evaluation

The final evaluation on the holdout fold produced an **AUC score of 0.512**, only marginally better than random guessing (0.50). Most predictions were concentrated in one class, limiting precision and recall for upward movements.

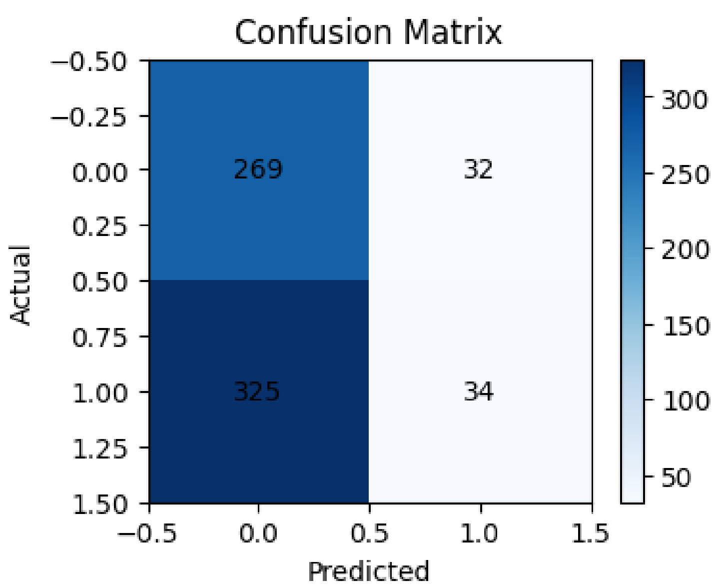

Figure: Confusion Matrix for Final Model

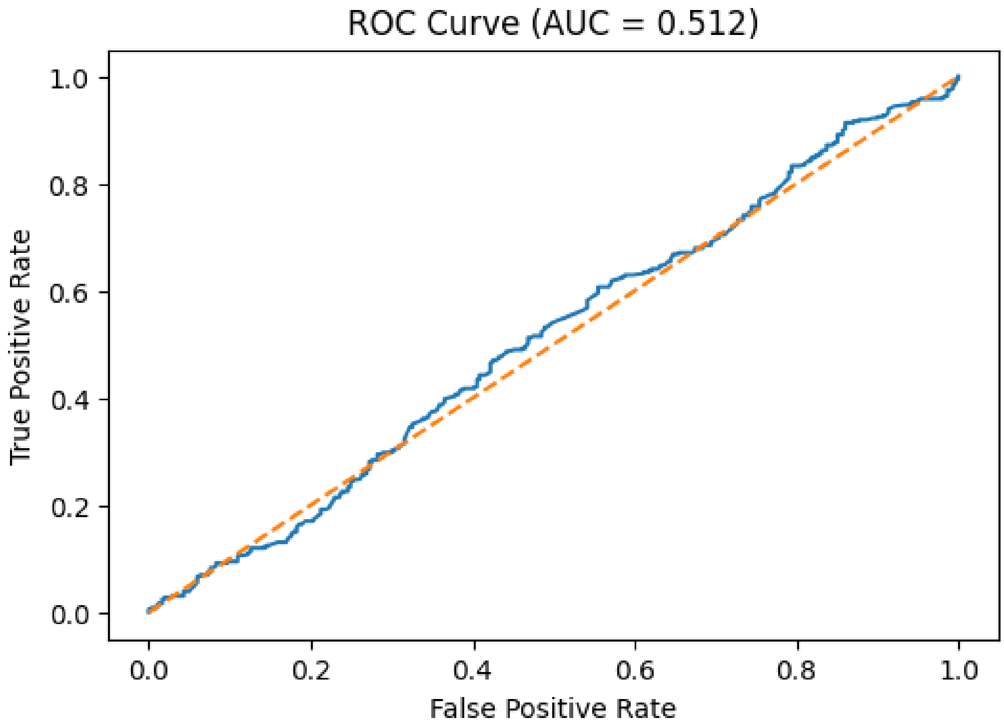

Figure : ROC Curve with AUC = 0.512

6. Conclusion

This implementation successfully reproduces a full modeling pipeline using time-series-aware machine learning. However, the poor model performance highlights several key challenges:

- AAPL is a highly liquid, efficiently priced asset, making short-term prediction difficult.
- Lag-based technical features may not offer enough signal to beat randomness.
- A better-performing model may require incorporating external data (news sentiment, macro indicators) or switching to assets with more volatility (e.g., TSLA, BTC-USD).