Akan Ndem

DS 4002

May 9th, 2025

Using Historical Stock Price with Linear Regression to Predict Future Performance

People say money makes the world go around and that could not be truer regarding the stock market. Many mathematicians, statisticians, software developers, etc. have realized a simple truth about the stock market. This truth is that people can squeeze value out of the stock market using advanced statistical analysis rather than typical investment techniques. Welcome to the world of quantitative finance!

Imagine, you have just started an internship at a quant firm (a financial firm that prioritizes quantitative analysis to get excess portfolio returns from the stock market), Your manager tasks you with building a quick, interpretable model to forecast future prices of a handful of stocks using historical data. You may be wondering what your goal is and that is to identify which companies are most predictable using simple statistical models.

Many trading strategies rely on predictive indicators from historical price data. Although machine learning dominates headlines, understanding and testing simpler models like linear regression remains essential. Linear regression and logistic regression rule the quantitative finance world as they are good at assessing stability, resistant to overfitting, and the signal quality is solid for financial data.

In this case study, you will explore stock data pulled via API, clean and normalize it, and use linear regression to predict prices across various companies. You will evaluate prediction performance using R^2 and compare predictability across different sectors, e.g. tech vs healthcare.

Your deliverable will include a set of prediction graphs, evaluation metrics, and a short reflection analyzing which stocks the model performed best on and why. For the explanation, try to use basic investment analysis or sentiment to understand other factors affecting stock price.