Predicting stock price movement with machine learning. Project 4.

WELCOME

Aadhitya, Daryl, Gayatri and Jasmine

The Problem

Can we use past stock data, along with macroeconomic indicators to predict if stock price will increase or decrease?

This project aims to predict if ‘tomorrow’s’ price will increase or decrease based on data from the short window prior.

Table of Contents.

01

Why did we choose this topic?

TOPIC

Did you know?

Financial markets are dynamic and influenced by diverse factors, making traditional analysis methods less effective. Machine learning offers the ability to detect subtle patterns and trends in vast datasets.

Embracing machine learning aligns with industry trends, allowing for a competitive edge through the development of predictive models that adapt and learn from historical market behavior.

Disclaimer:
While machine learning models offer valuable insights, they are viewed as complementary tools rather than infallible predictors, recognizing the inherent uncertainties and risks associated with predicting stock prices.

02

How did we gather our data?

Stock prices are influenced by numerous factors from price movements, trader sentiment, company performance, macroeconomic factors etc. We needed to gather this data fast and for free.

A combination of APIs, python libraries and web scraping!

Stock Price Information

We used a library called yahoo_fin that scrapes Yahoo Finance

Company Financials

We used chromedriver to scrape the website stockanalysis.com

Stock Screening

We used NASDAQ stock screener to screen Large Cap stocks in North America

Economic Indicators

We used an API called Alphavantage that plugs into the F.R.E.D API

Technical Indicators

We used an API called finnhub.io. They also have a python library.

03

Let’s explore our data.

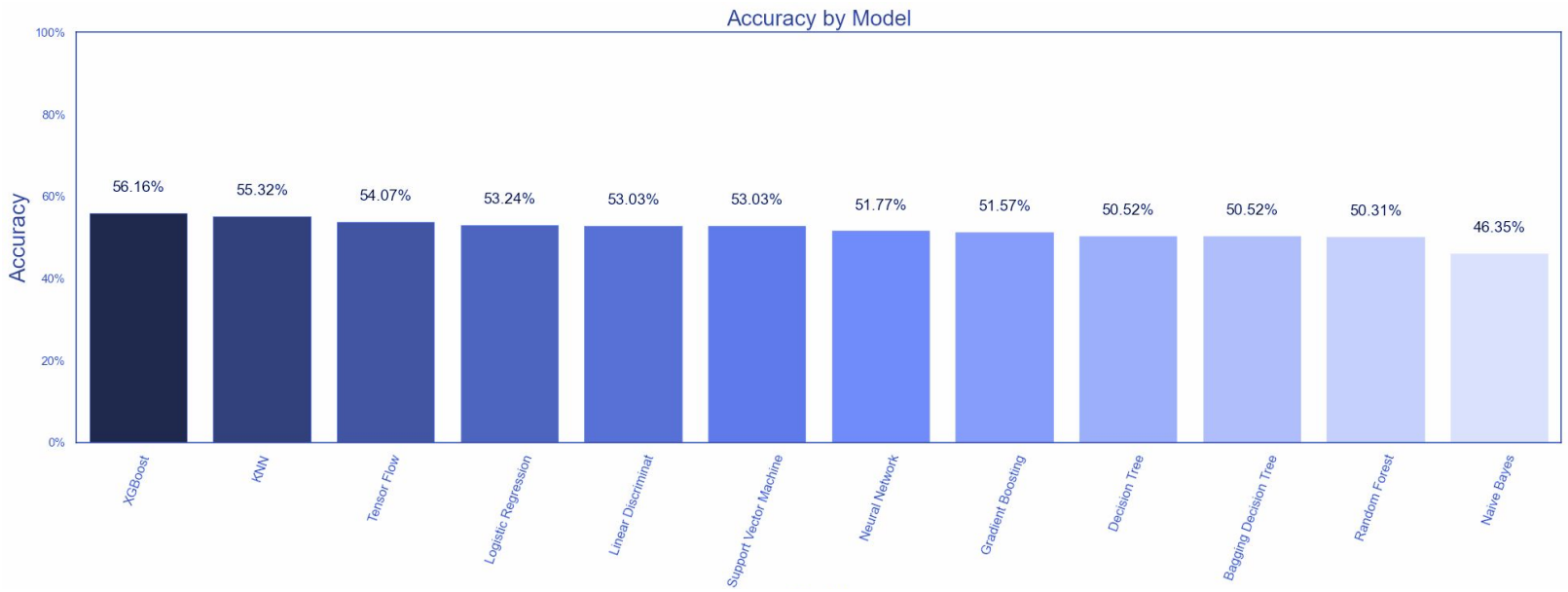
We built a feature set with 23 features and one binary label, to be predicted. Our label value is 1 if ‘tomorrow’s’ price increased (or stayed the same) and 0 if it decreased.

04

Let’s review our models.

We tested 12 different machine learning models, in order to figure out which would yield the best accuracy.

Plotting accuracy for each model



XGBoost Classification Report: Testing Data

	precision	recall	f1-score	support
0	0.46	0.41	0.43	220
1	0.54	0.58	0.56	259
accuracy			0.50	479
macro avg	0.50	0.50	0.50	479
weighted avg	0.50	0.50	0.50	479

XGBoost yielded the best accuracy.

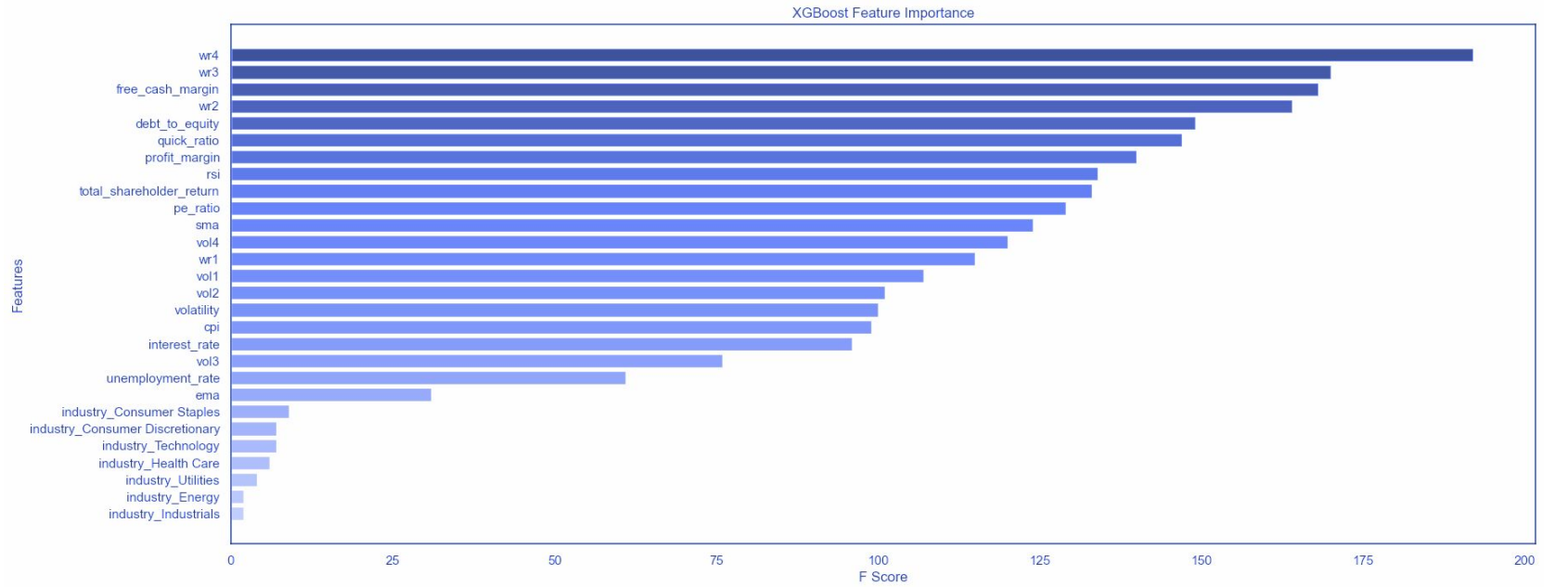
We can see that the model has a more difficult time with correctly recognizing when stock prices will decrease. Can we improve upon this with hyperparameter tuning?

05

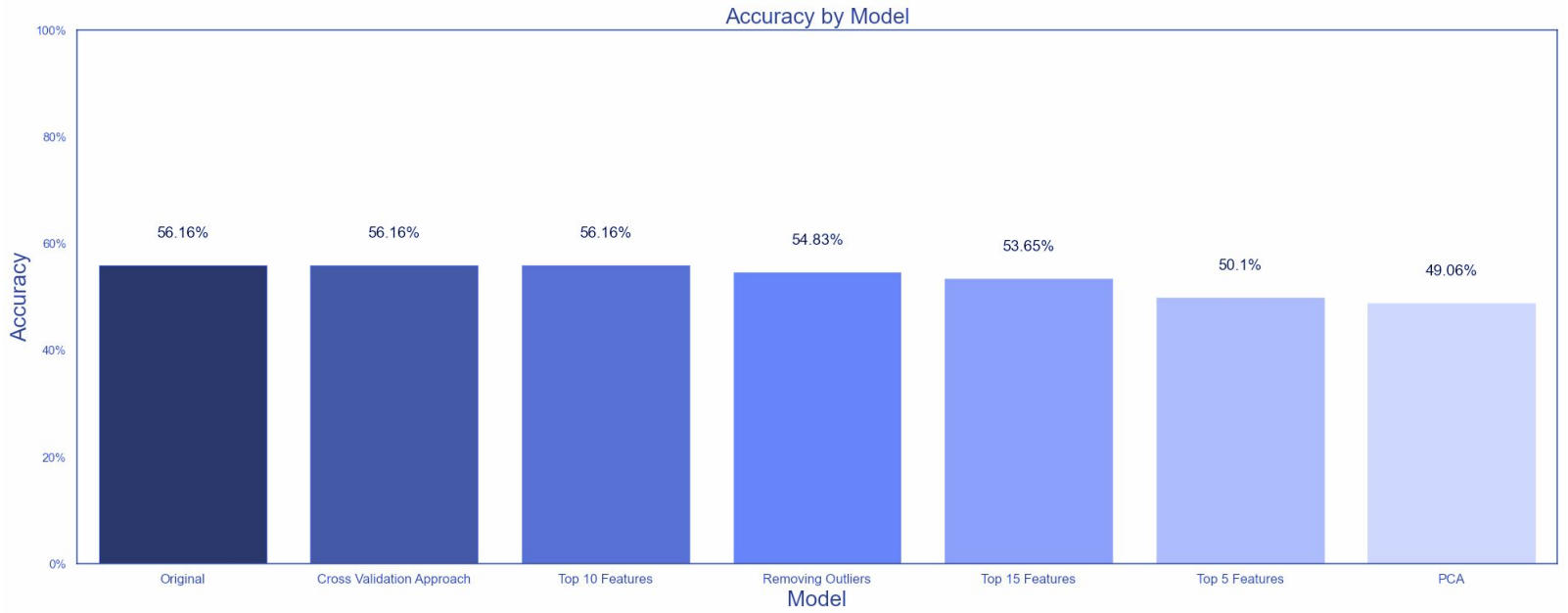
Hyperparameter tuning XGBoost.

We made several attempts to improve upon the original model’s accuracy score, from cross-validation, to reducing dimensionality to removing outliers

Observing feature importance



Accuracy after tuning models



06

Evaluating model performance.

Let's review the results.

Accuracy

The best accuracy we could achieve was 56.15%, even after tuning.

Model

After training 12 models, XGBoost had highest accuracy.

Features

The most important features in our data turned out to be price returns and financials

Tuning

Original model, hypertuned model and model with top 10 features yielded same results.

Precision/Recall

Precision rate of 46% for decrease, but 54% for increase or the same. Recall rate of 41% for decrease and 58% for increase or stay the same.

Overfitting

When reviewing many of our ML models, we observed that our training data had much higher accuracy than testing data

07

What were the challenges & limitations?

Most notably time constraints and resource limitations since we only used free data.

Major challenges

Stock market is very difficult to predict. Industry average is about 60%

We were limited to use free data. This meant we had a small dataset, with limits on data quality and features

We had a serious over-fitting issue

08

What can we do to improve our results?

How can we overcome the challenges we faced, and further improve upon our models?

Areas of improvement

More lines of data, higher quality of data and more features. Noting this is not free

Trying to use NLP models to capture market sentiment, including news and social media

Trying to find patterns in a niche may yield better results. Like this study.

09

What are the applications for real life?

Real World Applications

High Frequency Trading

Machine Learning can help in identifying patterns, and providing predictive insights to inform trading decision, especially for more subtle changes.

Portfolio Optimization

Machine learning can aid in portfolio optimization by analyzing diverse financial data, identifying patterns, and recommending optimal asset allocations to maximize returns while managing risk.

Risk Management

Machine learning can support risk management in trading by analyzing market data, predicting potential risks, and providing real-time insights to minimize financial losses.

Thank you!

Do you have any questions?