Commodities and Their Effects on the Technology Sector

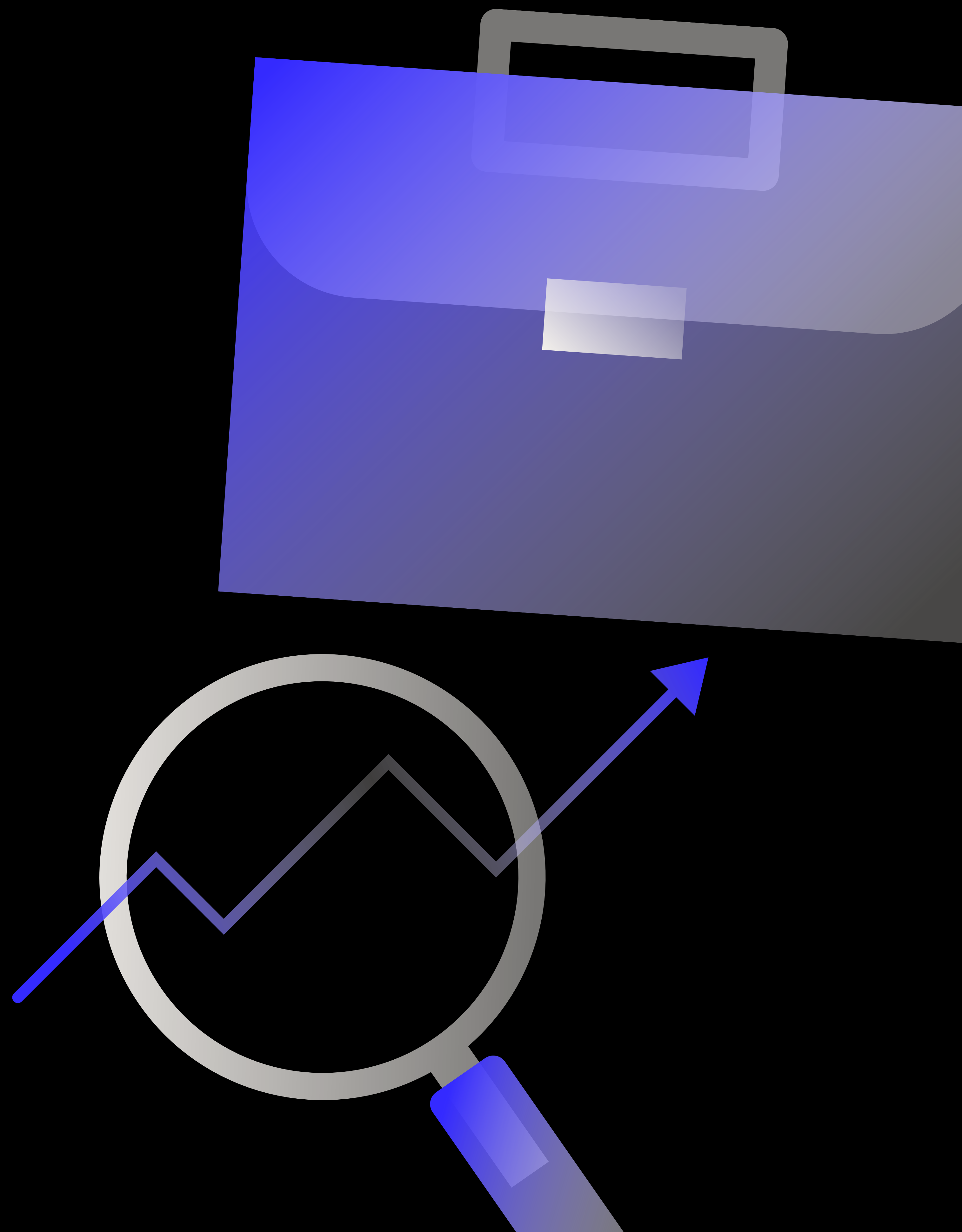
Darren Nguyen, Zayaan Atif, and Bryan Zhao

Introduction

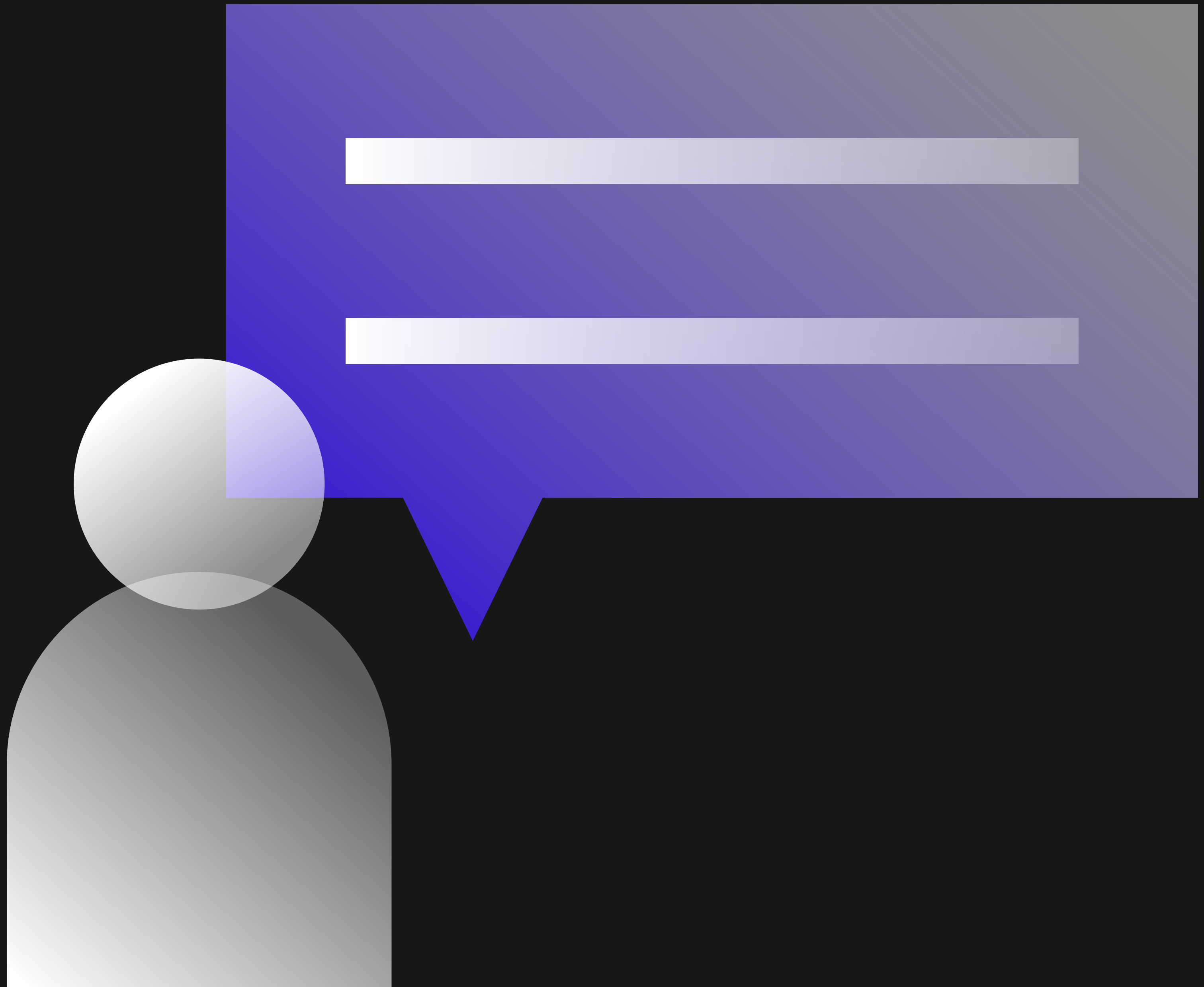
Tools for investment analysis are ever-changing. Investors are finding new ways to gather, analyze, and predict movements within the stock market. Within the discipline of quantitative finance, machine learning and neural networks have emerged as crucial methods for data analysis, attracting significantly more attention in recent years.

One overlooked factor when investing in stocks is commodity prices, some of which play an important role in the technology sector. For example, gold, silver, copper, and platinum are all utilized to manufacture technological infrastructure. As a result, a trading strategy focused on the technological sector could be developed based on these commodities.

The goal of this project was to predict movements in the technology sector using information about commodity prices.



Essential Questions



1. Which model is most effective in predicting an index’s movement based on commodity price changes?
2. Which commodities are most important to the technology sector?
3. What is the most important commodity to market sentiment, and how does this relate to our models’ findings?
4. Can we develop a trading strategy using our model?

Data Sources

yfinance

We pulled 10-year historical price data from the yfinance API across 9 different commodities and an information technology index fund.

We also pulled recent market sentiment from posts across the following subreddits: 'r/news', 'r/worldnews', 'r/breakingnews', 'r/globalnews', 'r/wallstreetbets', 'r/stockmarket', 'r/stocks', 'r/trading', 'r/daytrading', 'r/economics', and 'r/economy'.

Commodity	Name
GLD	SPDR Gold Shares (Spot Price)
SLV	iShares Silver Trust (Spot Price)
PPLT	abrdn Physical Platinum Shares (Spot Price)
CPER	United States Copper Index Fund (Futures)
USO	United States Oil Fund (Futures)
USNG	United States Natural Gas Fund (Futures)
WEAT	Teucrium Wheat Fund (Futures)
SOYB	Teucrium Soybean Fund (Futures)
CORN	Teucrium Corn Fund (Futures)

ETF	Name
VGT	Vanguard Information Technology Index Fund

Data Pre-Processing

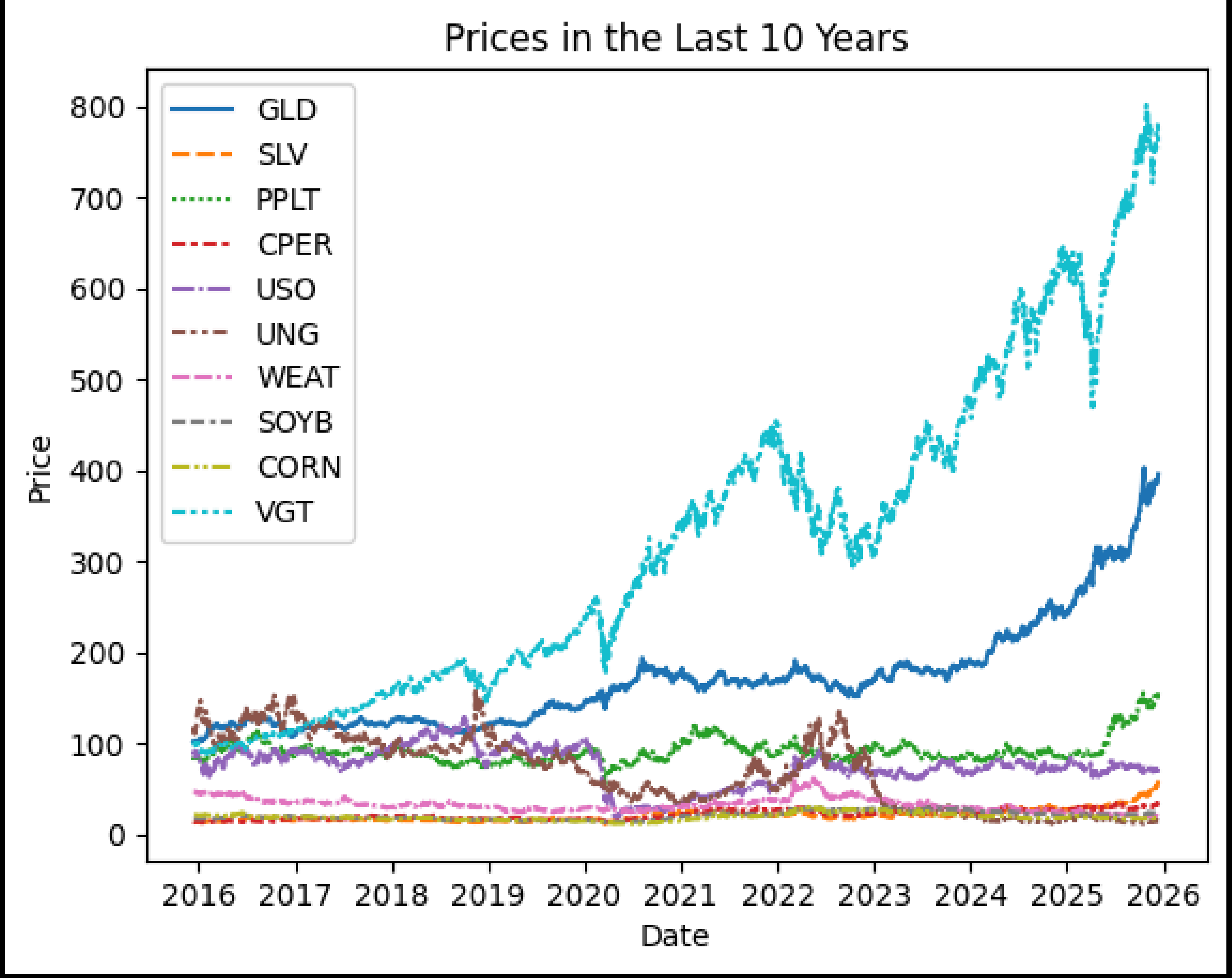
We plotted the price data over the last 10 years.

With the price data, we calculated the percent changes to reflect relative movement for each ETF.

We applied a Standard Scalar to utilize in some of our models.

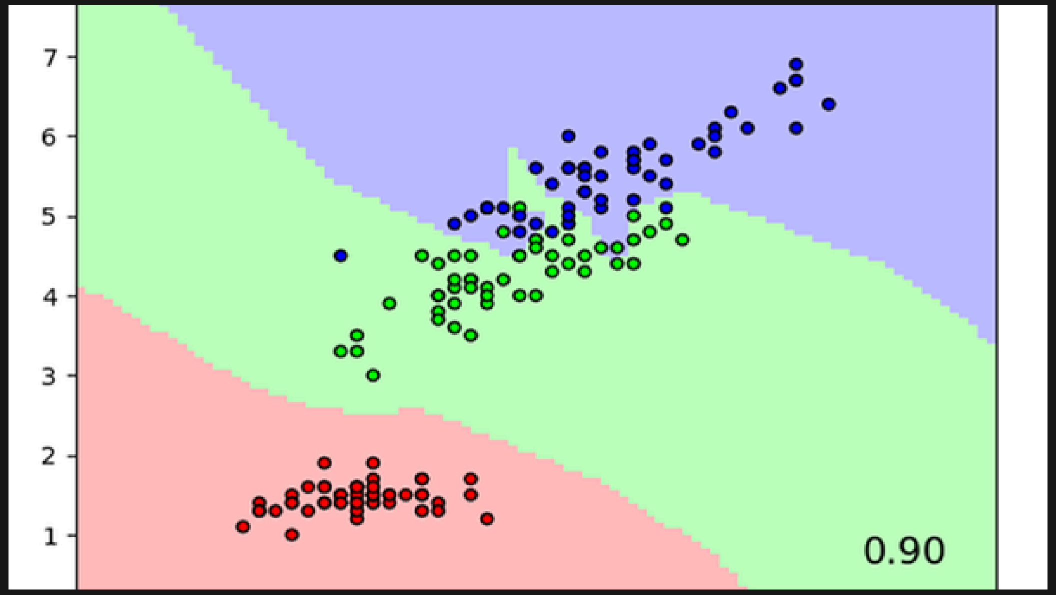
Then, because our goal was to predict VGT's direction of movement, we represented any positive percent change as a 1 and represented any negative (or zero) percent change as 0. This was our target column.

Finally, we applied a train, test, and split to the data to accurately develop our models.



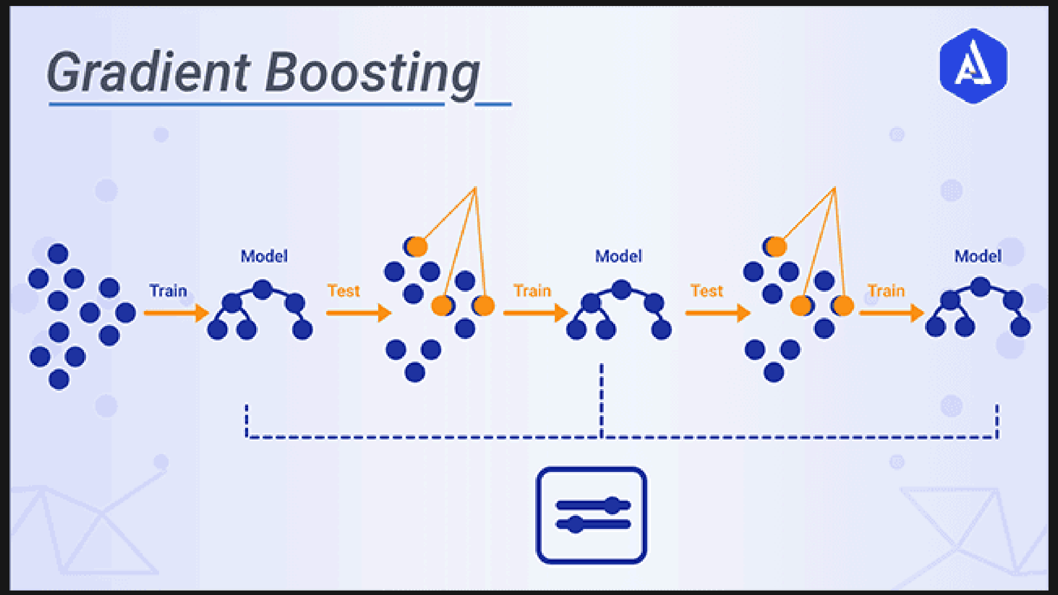
Supervised Learning Models

K-Nearest-Neighbors (KNN)



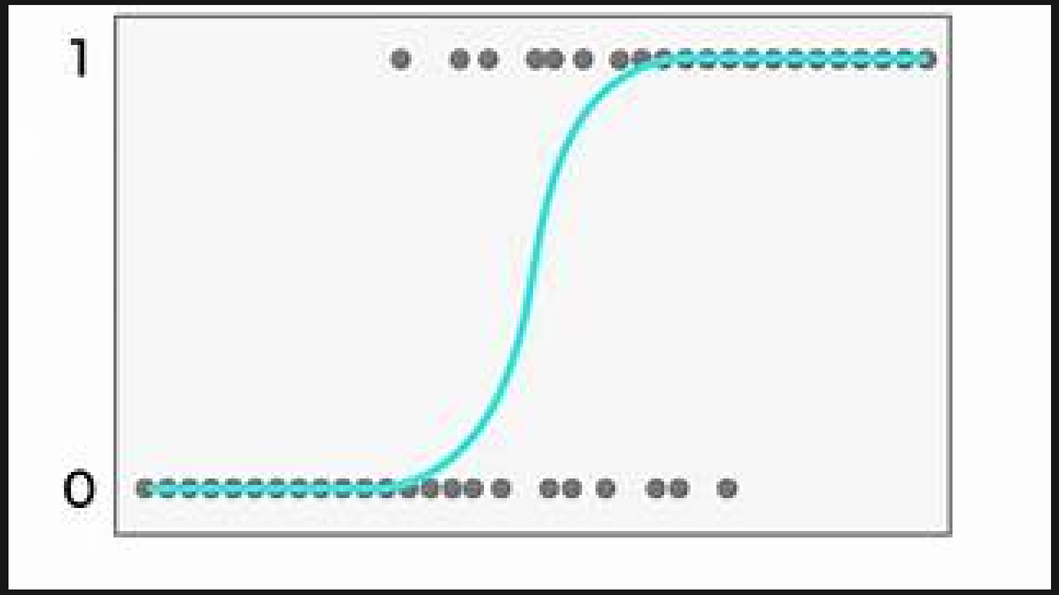
Looks at neigherest neighbors to classify target

Extreme Gradient Boosting Classifier (XGBC)



Uses many tree models to create one strong model

Logistic Regression (LR)



Using a sigmoid to calculate the probability of an outcome

Optimizing & Evaluating the Models

Evaluating

To evaluate our models, we used both a precision and an F-Beta score. This study only applies to buying VGT shares, so it would be bad to predict VGT increases when it actually goes down (incurring losses), but not as bad to predict VGT decreases when it actually goes up (missing gains). Because investors want to avoid losses, we chose precision as a scoring metric. However, if investors avoid all losses, they would never make any money. The F-Beta score provided a metric that mixed both precision and recall, balancing loss aversion and gain aversion. We used a beta of 0.5 to prioritize precision.

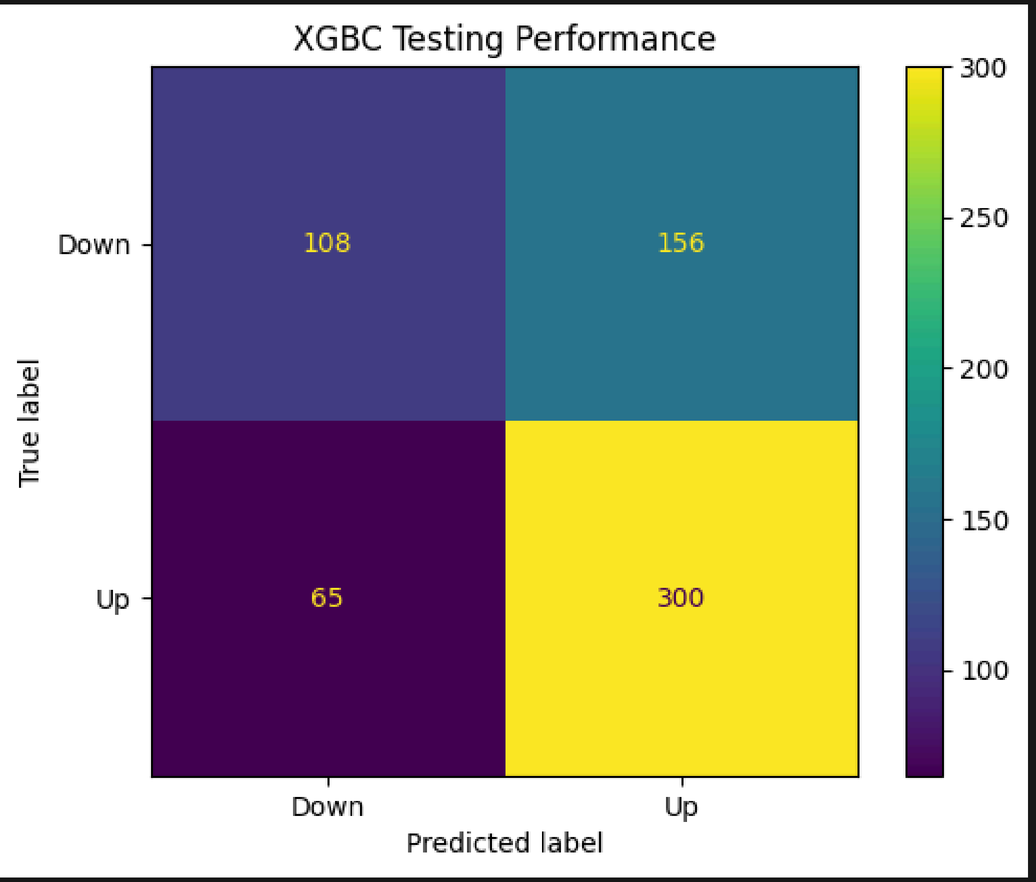
Optimizing

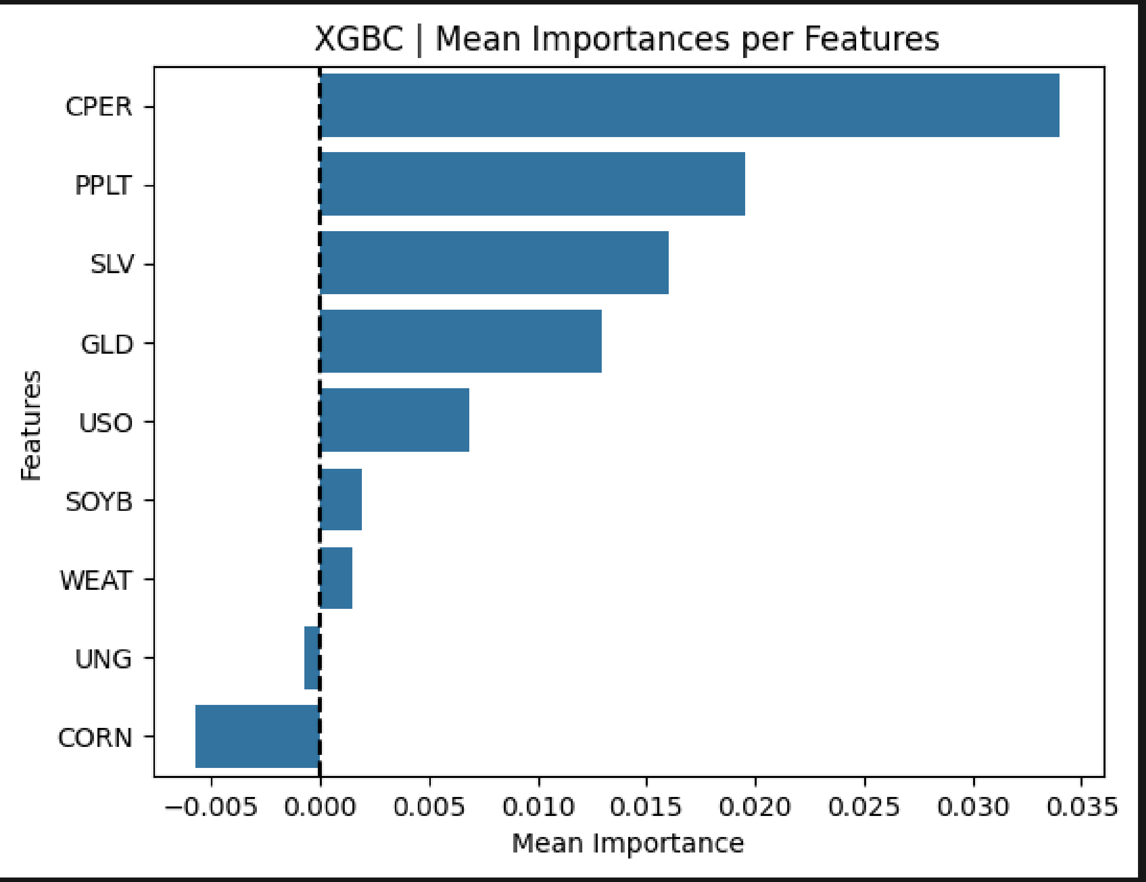
To determine the optimal parameters, we employed GridSearch on the nearest neighbors for KNN and max depth for XGB.

Determining Feature Importances

To compare commodities and determine which were useful features, we used permutation importance. This method shuffles feature values around and calculates the effects on model performance. The most important features are the ones that drop performance the most when randomized. However, we were only able to utilize this method on the supervised learning models, not the artificial neural network.

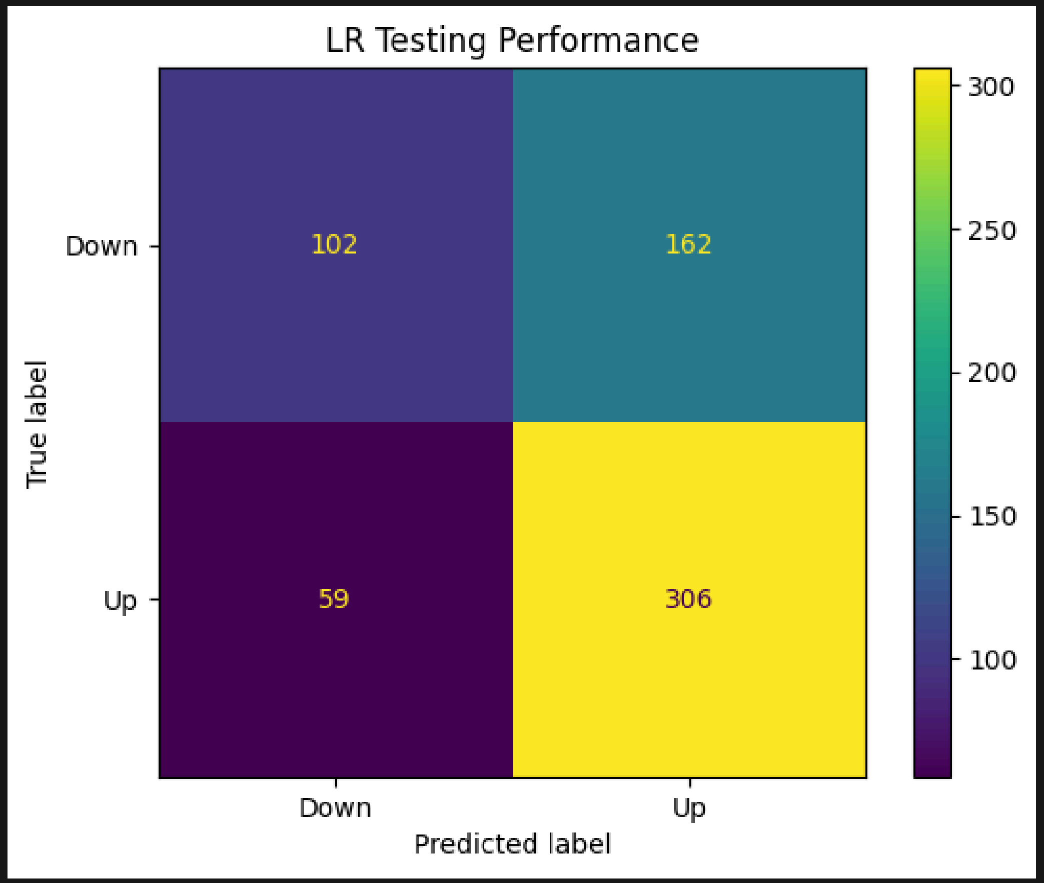
Extreme Gradient Boosted Classification

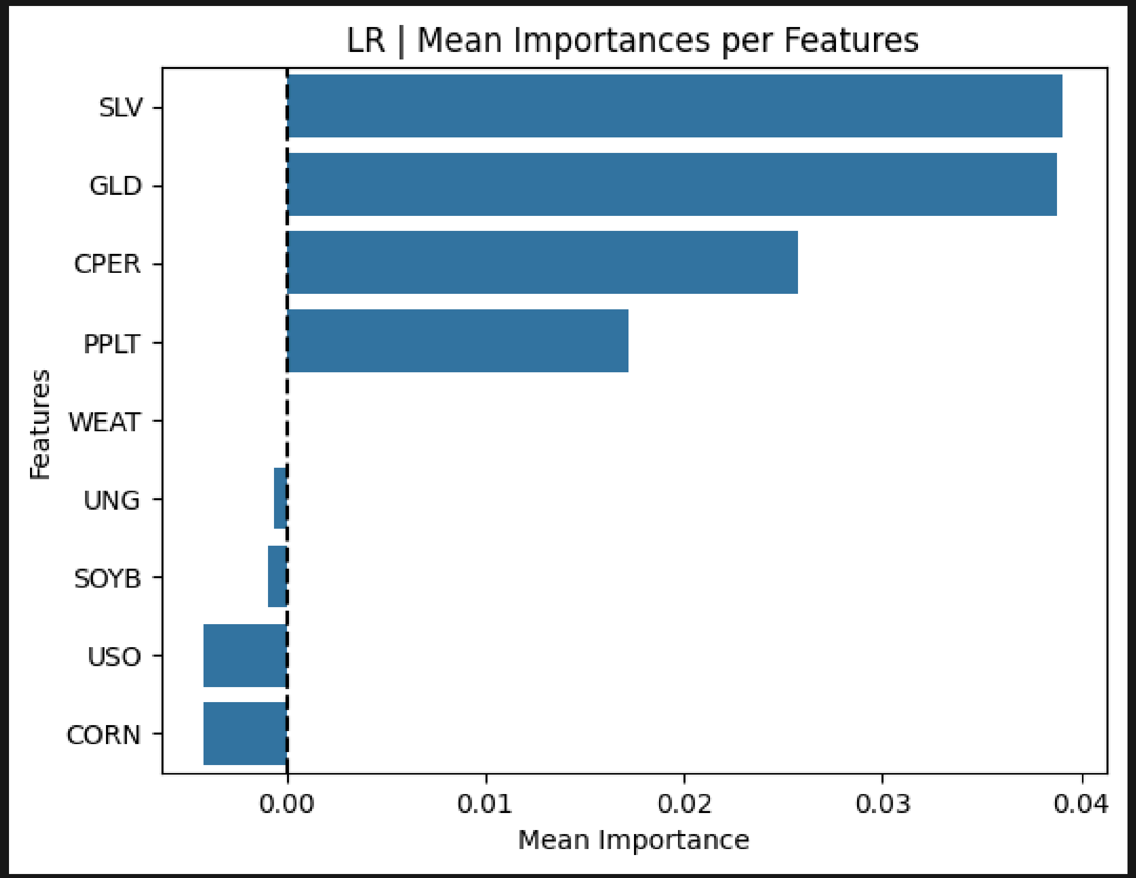




- Precision Score: 65.8%
- F-Beta Score: 68.5%
- High Mean Importances: Copper, Platinum, Silver, Gold
- Low Mean Importances: Corn, Natural Gas, Wheat, Soybean, Oil

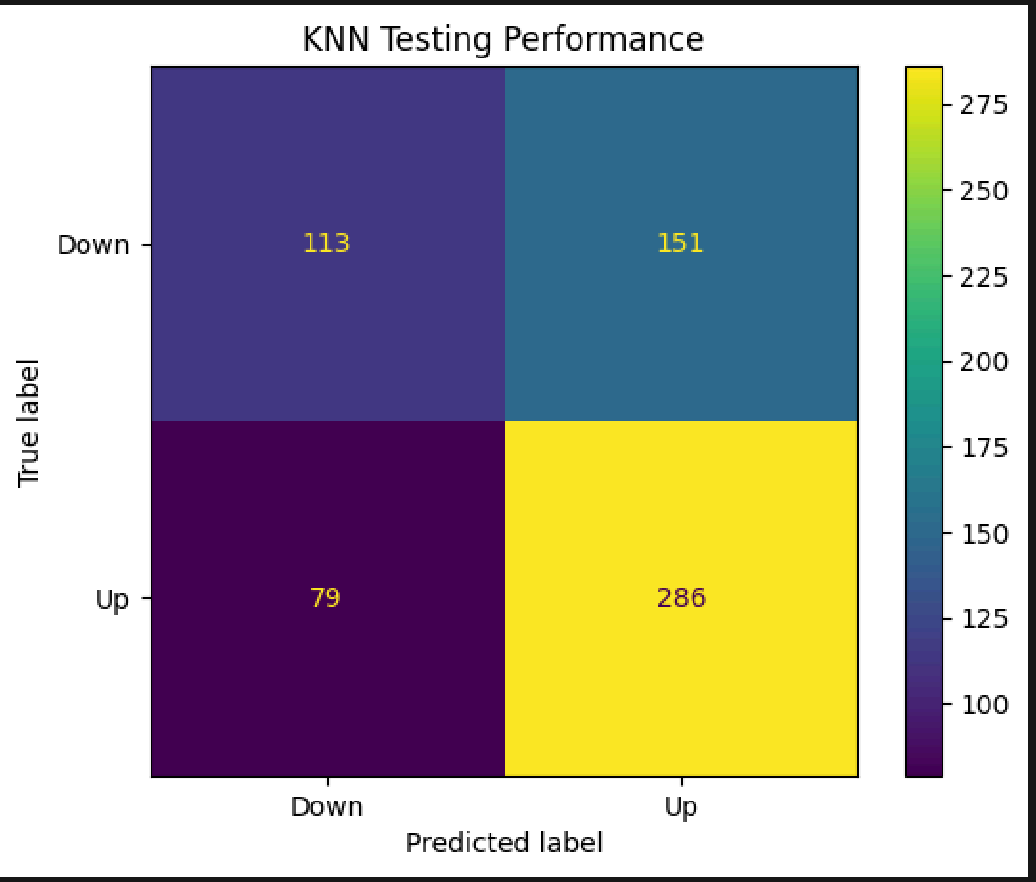
Logistic Regression

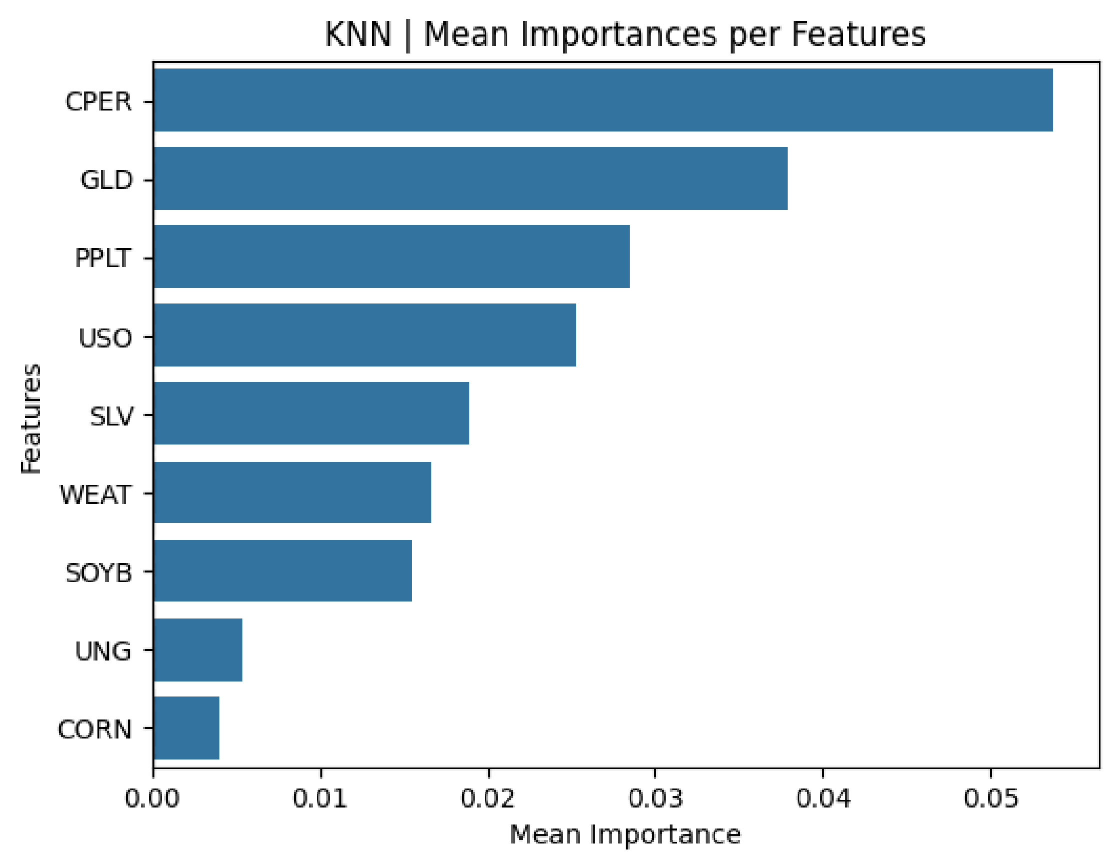




- Precision Score: 65.4%
- F-Beta Score: 68.4%
- High Mean Importances: Silver, Gold, Copper
- Low Mean Importances: Corn, Oil, Soybean, Natural Gas, Wheat

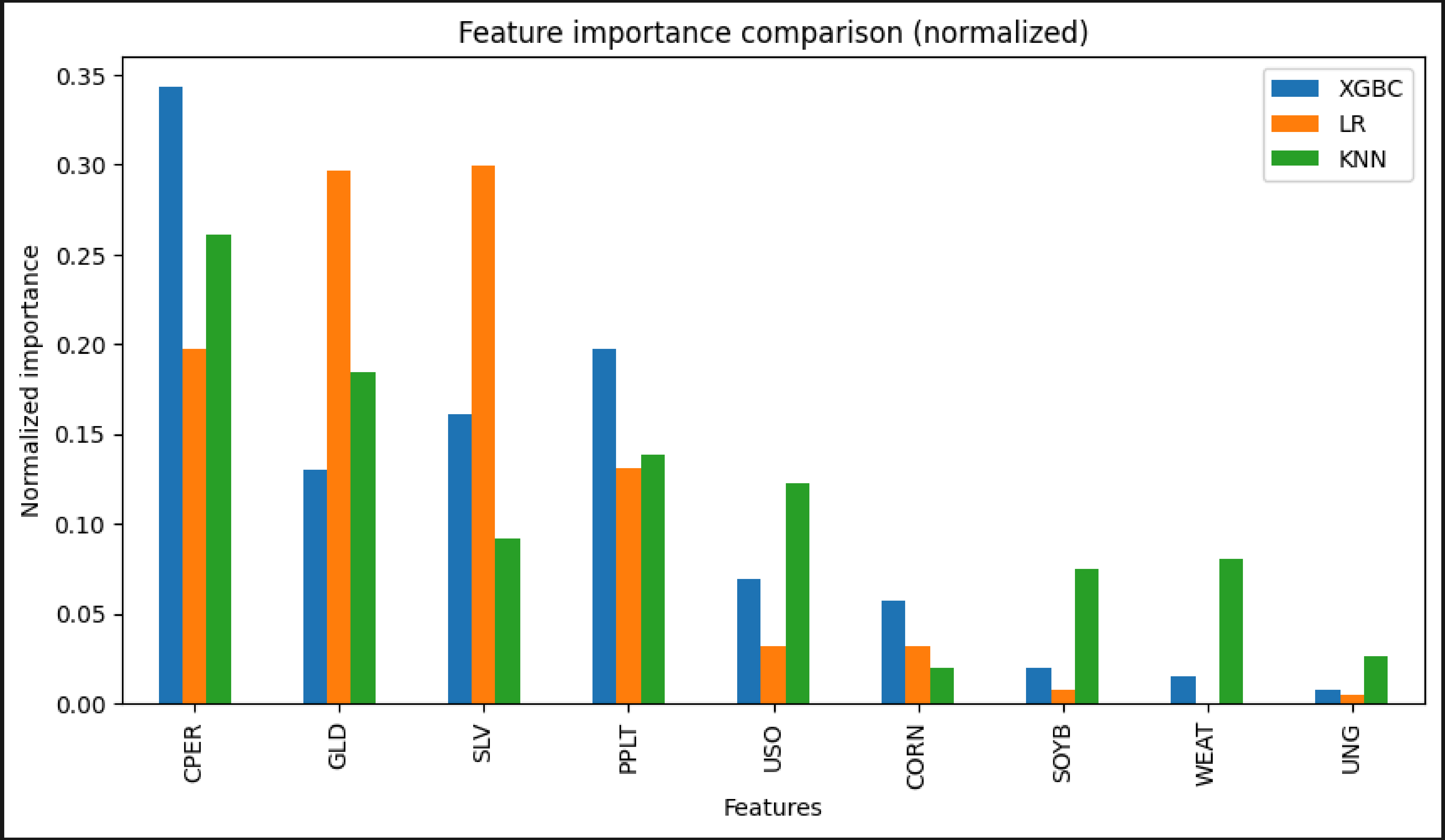
K-Nearest Neighbors Classification



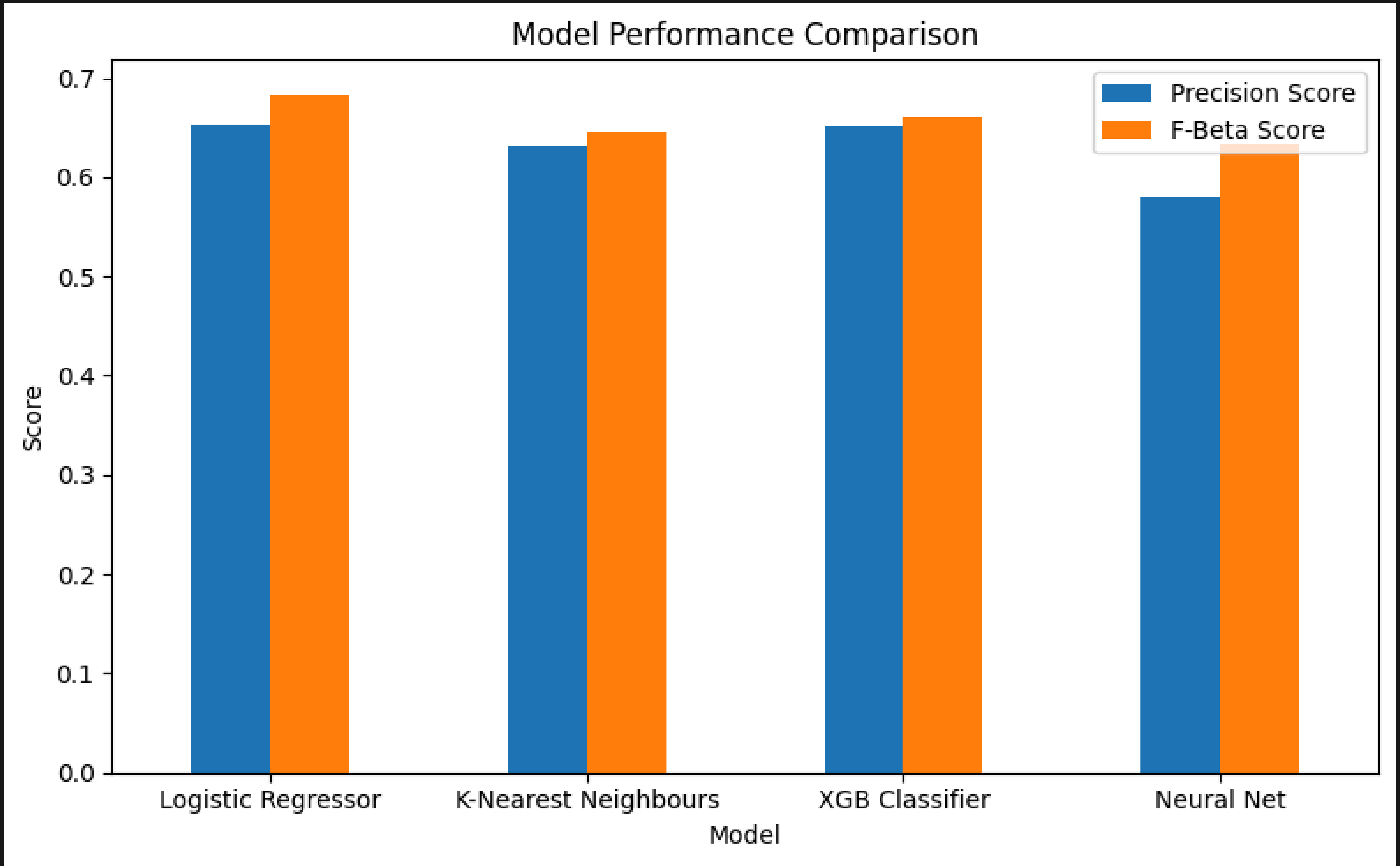


- Precision Score: 65.4%
- F-Beta Score: 67.7%
- High Mean Importances: Copper, Gold, Platinum, Oil
- Low Mean Importances: Corn, Natural Gas, Soybean, Wheat, Silver

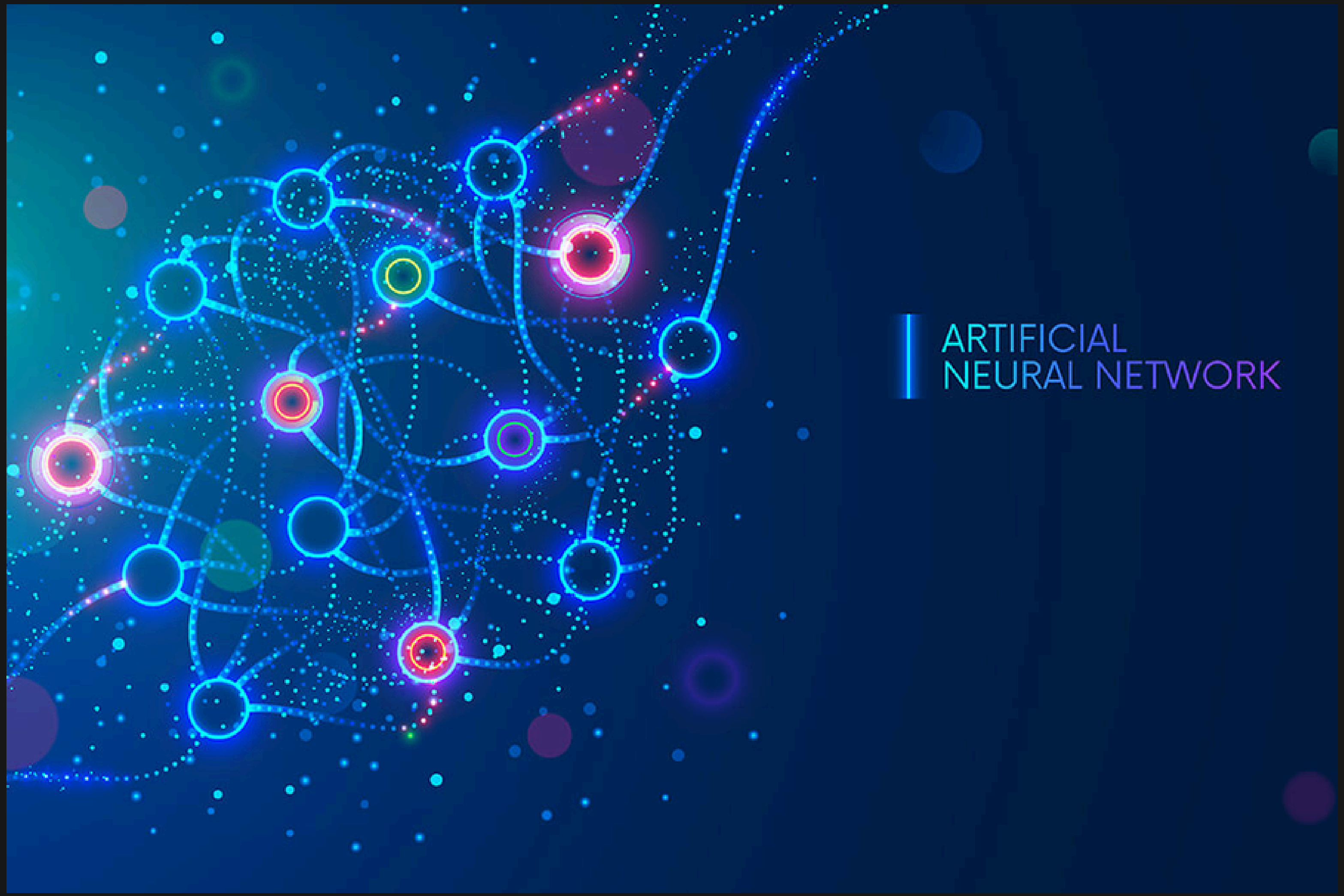
Comparing Feature Importances

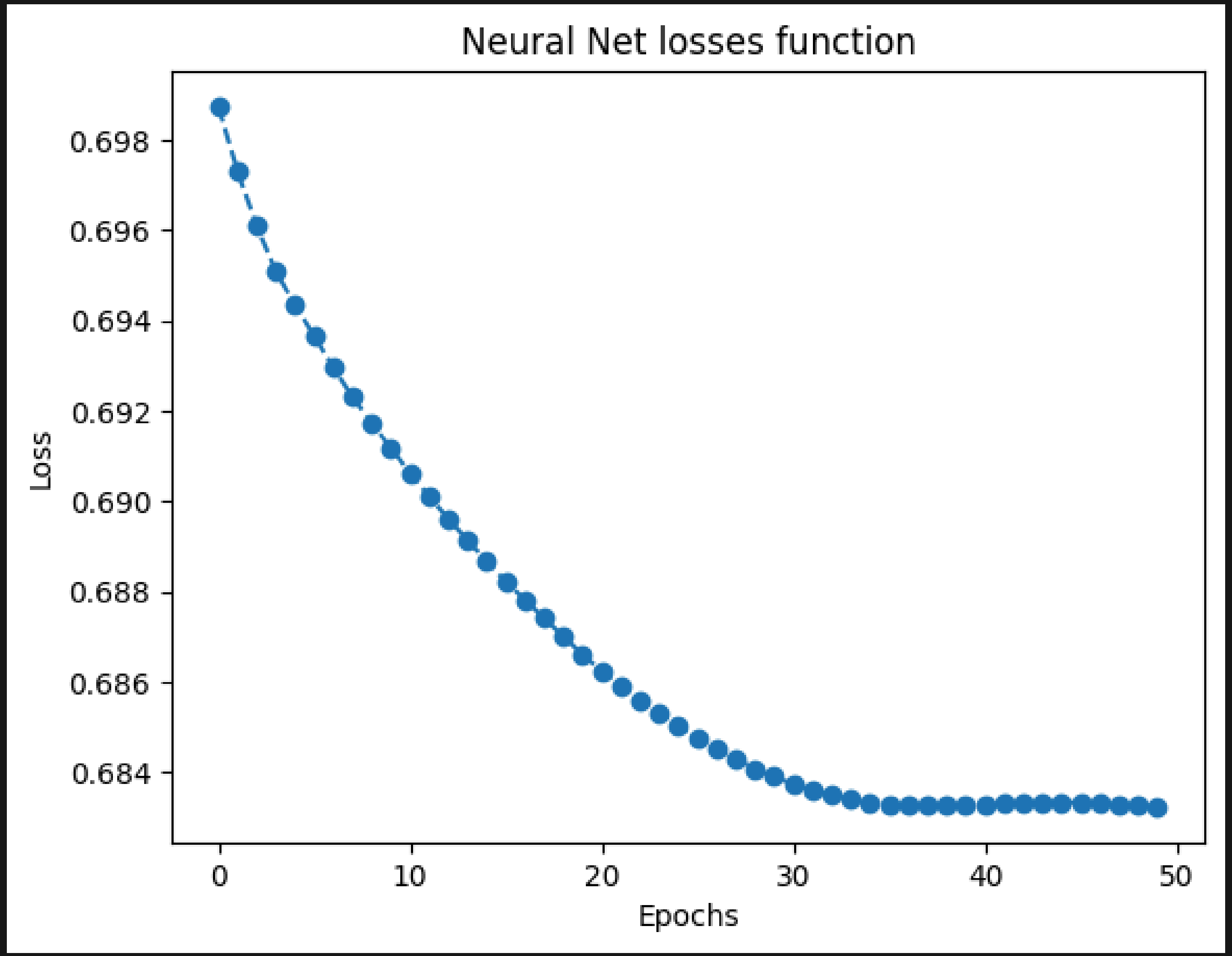


Comparing Peformance Scores



Artificial Neural Network



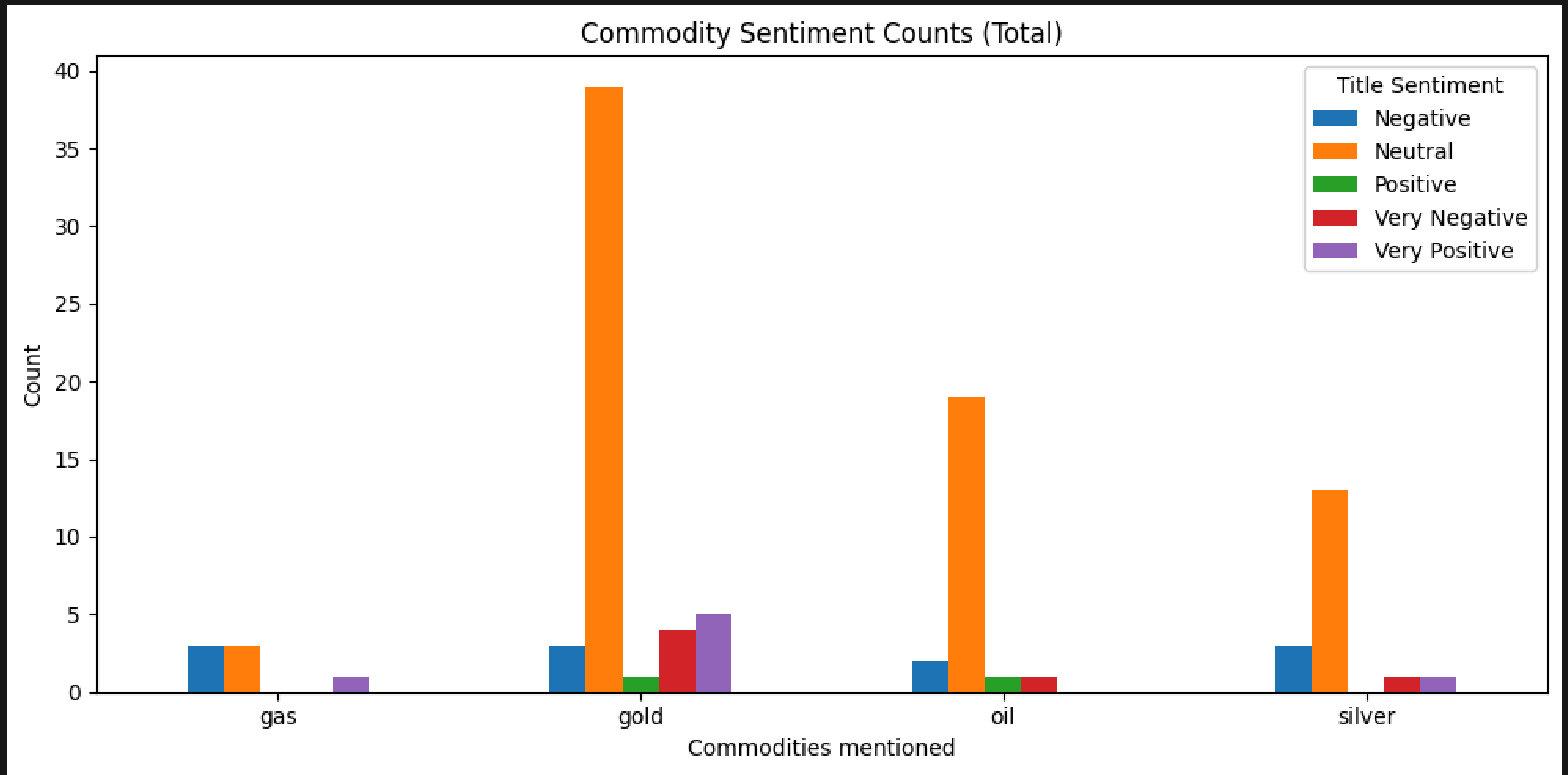


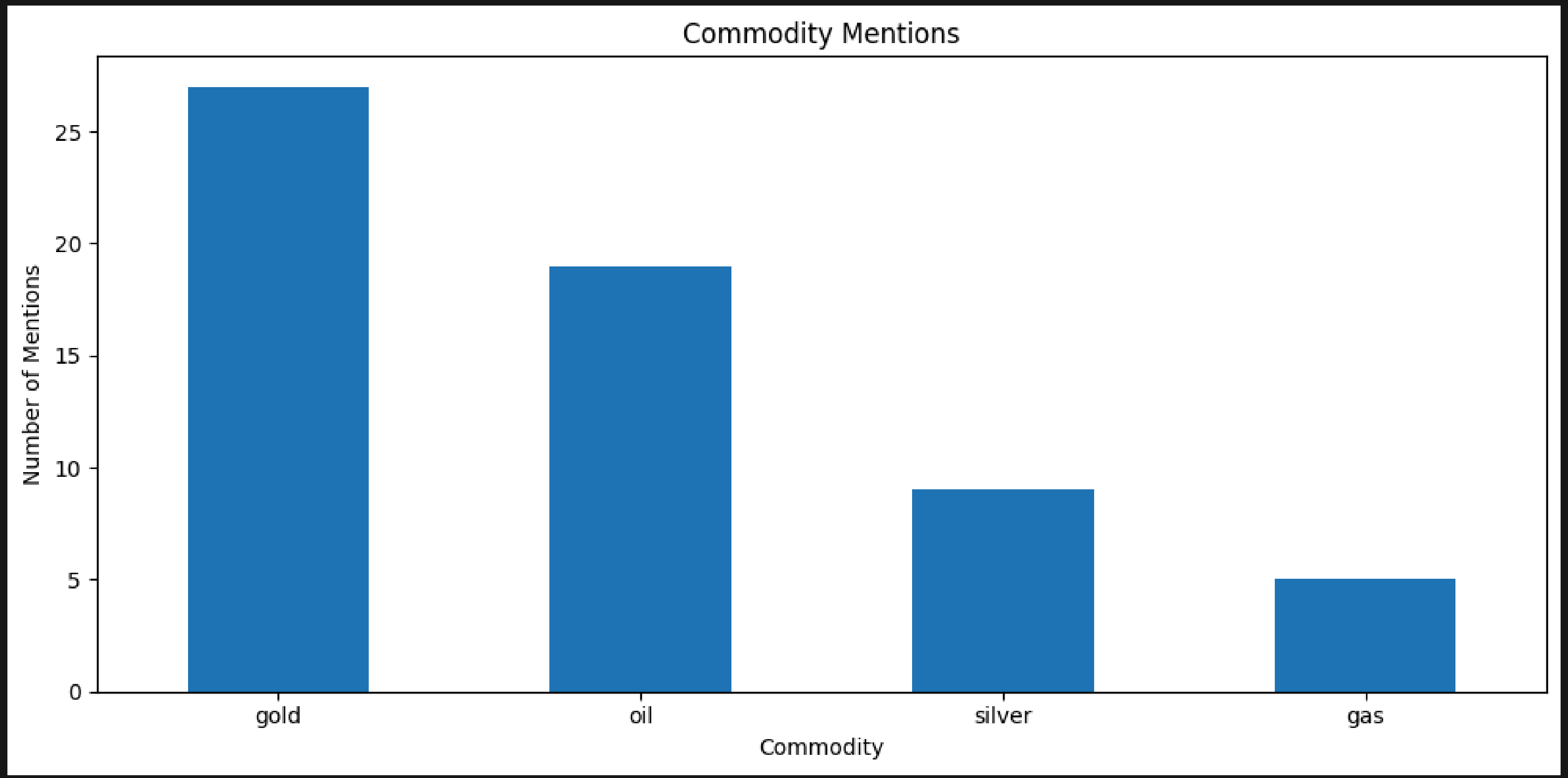
Uses layers of gradients to create different models and paramters to model the data

- Precision Score: 58.0%
- F-Beta Score: 63.3%

Sentiment Analysis

We used the “tabularisai/multilingual-sentiment-analysis” to analyze the sentiment of the title and body of each Reddit post. ranking the texts as either “Very Negative”, “Negative”, “Neutral”, “Positive”, or “Very Positive”.





Gold was the most mentioned commodity, followed by oil and silver. Surprisingly, there was a lot of discussion about gold and silver, but not as much about copper, which was the most important variable for most of our models. However, gold and silver may be highly mentioned due to idiomatic phrases or other uses, not just stock-related.

Answering Essential Questions

Which model is most effective in predicting an index's movement based on commodity price changes?
The best-performing model was the Extreme Gradient Boosted Classification. It had a precision score of 65.8% and an F-Beta score of 68.5%. However, all 3 models had scores that were vary similar, varying by only a few tenths of a percentage point. Meanwhile, the artificial neural network for classification performed the worst.

Which commodities are most important to the technology sector?
Across all models, precious metal commodities typically had the largest mean importance, with copper, silver, and gold being the most important. The least important features were the agricultural commodities as we expected. Interestingly, fuel commodities such as natural gas and oil did not seem very important, contrary to our initial beliefs.

What is the most important commodity to market sentiment, and how does this relate to our models' findings?
According to our webscraping and sentiment analyzer, gold, oil, silver, and gas were the most mentioned commodities. Interestingly, copper, the most important feature in our model, was not a commonly mentioned stock.

Can we develop a trading strategy using our models?
Although our models performed better than the baseline, about 65% precision compared to 57%, our models would not directly be used for a trading strategy. It is important to note that we used the closing prices for VGT for the same days as our commodity prices. This makes our models more focused on predicting the movement of the technology sector given the commodity price changes for that day, rather than directly predicting future VGT prices. However, the results showed important precious metals to track, something that can be further explored in a trading strategy.

Conclusion

Through optimizing and evaluating multiple models to predict the movement of the technology sector based on the movement of commodity prices, we found that the Extreme Gradient Boosting Classifier was the most effective model in predicting the movement of the technology sector, whilst minimizing type I errors. Additionally, in terms of feature importances, precious metal commodities were the most important, which aligned with the findings from our sentiment analysis models.

To further improve our model, we want to expand our data scope and sentiment analyzer. First, our data was very narrow in scope and lacked a prospective view on the stock. Additionally, our sentiment analyzer was not fine-tuned for our goal in analyzing stocks and commodities specifically, making our market research difficult and inaccurate in certain scenarios. In minimizing errors and selecting a more effective model, we believe that we can utilize these findings in developing a trading strategy.

** Note: Figures in the code may be different from the paper and presentation due to new market data. The data shown in the pictures of this paper and the presentation came from December 10th, 2025.