Market Regime Detection Using Supervised Machine Learning Models

Machine Learning (STAT 4650)

Professor: Yue Gao

Submission Date: December 13, 2025

Group Members:

Devashree Dhole

Mohammad Faishal Khan

Narayan Rijal

Pragathi Somashekaraiah

Table of Contents

1. Abstract
Bulls and bears markets are changing in the financial markets and the losses associated with investing could be high in case of the lack of anticipation of the downturns. The study will explore the hypothesis that machine learning models are able to make regime aware investment decisions and therefore, predict bull and bear market regimes using only pre-market information. We refer to regime identification as a binary identification task based on organized financial and macroeconomic data set constructed on the basis of historical market indicators. Many supervised machine learning models (such as linear and tree-based models) are trained and tested by using standard performance measures, including ROC-AUC, confusion matrices, and regime-specific recall. The findings indicate machine learning algorithms are rather effective in identifying bad market conditions and can easily distinguish between bull and bear markets. These results imply that regime insensitive prediction models may be useful in curbing downside risk in decision levels. The article illustrates the use of machine learning-enhanced prediction of regimes to increase risk management and adaptive procedures of investing.

2. Introduction
The cyclical operations that the financial markets experience, not only during growth but also during decline are referred to as bull and bear market regimes. These regimes are quite critical towards determining the extent to which investments perform since returns, volatility and risk exposure can vary significantly across markets. Long-term declines, increased uncertainty, and increased risks of making money characterize bear markets. Bull markets on the other hand are normally identified to the increase of asset prices and healthy economic conditions. Failure to foresee the changes that may befall your investments to its detriment means that you may lose a significant amount of money and end up consuming a considerable period before you recover it.

The US equity market is a significant financial market indicator in global financial markets in the sense that it is large, open, and capital transfer impacts the global market. Alterations in the regime that occur in the United States are usually accompanied by larger economic cycles and may significantly affect investors worldwide. This is why the U.S market is ideal to research on regime dynamics and the approaches employed can still be applied to other equity markets.

The progressive development of machine learning algorithms and the availability of data has provided new opportunities to enhance better market regime identification. Complex, nonlinear relationships in data, as well as integration of various types of financial and macroeconomic indicators, may be discovered in machine learning models. This does not contrast with conventional rule based or strictly statistical approaches. These models lack the look-ahead bias and they are also useful in making the real-time investing decisions when they only use the pre-market data.

This paper is devoted to the use of monitored machine learning models to predict the bull and bear market regimes. This dichotomous classification is a challenge of categorizing types of historical data into bull or bear regimes according to a set of definitions. The main aim is to determine whether machine learning algorithms can be successfully used to distinguish between positive and negative market environments with the main focus on the identification of bear markets because of their asymmetrical effect on the risk of the portfolio.

This is a subject that is important to study both methodologically and practically. In practice it can assist in risk management and flexible investing strategies by ensuring that it is simple to re-adjust market exposure in little time. In terms of methodology, the imbalance of classes, unstable financial indications, and changing market trends make the categorization of market regimes a tough challenge to assess machine learning models. This study investigates the following research questions, informed by these considerations:

Research Questions

- Can supervised ML models accurately classify bull vs bear market regimes using pre-market indicators?
- Which financial features (VIX, yield curve, TLT/SPY ratio) are most predictive of market regimes?
- How can regime classification improve risk-aware decision-making strategies for investors?

3. Method and Analysis

a. Description and Sources of Data.

The current study is based on a gathered financial time-series data that is based on the popular and publicly available market data sources in the United States. We acquired the market price and the volatility data of Yahoo Finance and the macroeconomic indicators of the Federal

Reserve Economic Database (FRED). All of them combined are a whole picture of how the market operates, investor mood, and economic performance.

The ultimate consolidated data contains 5822 observations of daily trades between July 2002 and September 2025. Such observations contain various market cycles and changing regime. Each observation is a trading day, and it has market-based indicators which may be observed prior to the opening of the market. This ensures that the analysis is free of look-ahead bias and is applicable in making predictions in the real world.

In this study, the core stock market is represented by SPY ETF. It is the index that follows the movement of the S&P 500 and displays the development of the U.S. equity market. The so-called fear gauge in the market is the VIX index that indicates the degree of volatility that is likely to occur in the market in the near future. TLT ETF is an exchange-traded fund that follows the movement of long-term bonds in the United States. The market being under stress is also a sign of risk-off behavior. Inclusion of macroeconomic variables, also, is done using the 10-year and 3-month Treasury yield spread of FRED. This is one of the common methods used to gauge the risk of economic growth and recession.

b. Description of the variable

- Date: Trading date (daily frequency)
- SPYret: The daily percentage change of SPY which is an exchange-traded fund tracking S&P 500.
- VIX: the level of volatility index of CBOE (market-implied volatility / one of the risk sentiments).
- YieldCurve: This shows the yield curve (spread proxy; the higher the figure, the steeper the curve, and vice versa).
- CreditSpread: A proxy of credit spread that indicates the amount of stress on credit markets; the larger the number, often indicates higher stress.
- TLTSPY ratio: This is a ratio of strong long-term Treasuries (TLT) versus stocks (SPY).
- Regime: A written word which is Bull or Bear.
- y: This is a number which is used to model (Bull = 1 (positive class), Bear = 0).

c. Data Preparation

This data will be in the format of one organized table, and thus, there was no necessity to combine or relate information in multiple files. The objectives of data preparation were primarily to generate new features and divide the data into time-based groups to generate predictive modelling without information leakage.

We developed a one-day lagged back feature of daily returns of SPY. This lagged variable displays the profit of the last trading day, and it ensures that the model does not use the market to predict what it will do on the same day. This aspect allows the model to consider the short-term effects of the market change in a manner that remains congruent to the real time decision making.

Having constructed the features, the explanatory variables and the target variable were quite clear in the prepared dataset. The selected predictors include such variables as market volatility, macroeconomic variables, and relative asset performance indicators, and the target variable is the binary market regime, i.e. bull or bear market.

In order to prevent look-ahead bias, the data were divided into training, validation, and test sets according to the fixed date cutoffs. To train the model, we used observations made prior to December 31, 2018, to validate the model, observations made between January 2019 and December 2021, to test it, and the remaining observations to test the model. The divisions made maintain the time structure of financial time series by dividing the data into training, validation, and testing.

d. Cleaning and Class Balance Treatment of Data

The dataset does not have any missing values in the original market and macroeconomic variables. The only observation that is absent is attributed to the creation of the one-day lagged return feature that brings a missing value in the first row of the data set. This was pre-observed prior to the analysis to ensure that the modeling data set was complete.

Besides addressing the issue of the missing value associated with latency, the class distribution of the target variable was also considered. The dataset has a significant difference in observations of the bull and bear markets. The frequency of bull markets is greater within each data partition. To address this issue during the training of the model resampling was applied solely upon the training set. This approach assists the model to acquire patterns associated with the less frequent bear market scenario even though the initial distribution of classes in both the validation and test sets is retained to ensure that the performance can be assessed fairly.

e. Processing Missing Values and Scaling Features.

No kind of statistical imputation was needed in this study. The raw data set also entails full observations of all the predicting variables. The single value that was not available before and was added in the preprocessing was considered through removing the observation that contained it. This is the reason why the final dataset to be used in the modeling process does not have any missing value. The predictor variables were normalized before the estimates of the model, to ensure that the model would not give the effect that would tend to vary with the scale of the features. Scaling parameters were learned on the resampled training data and applied to the test and validation sets. This method ensures that every dataset is identical and there is no information leakage in future observations.

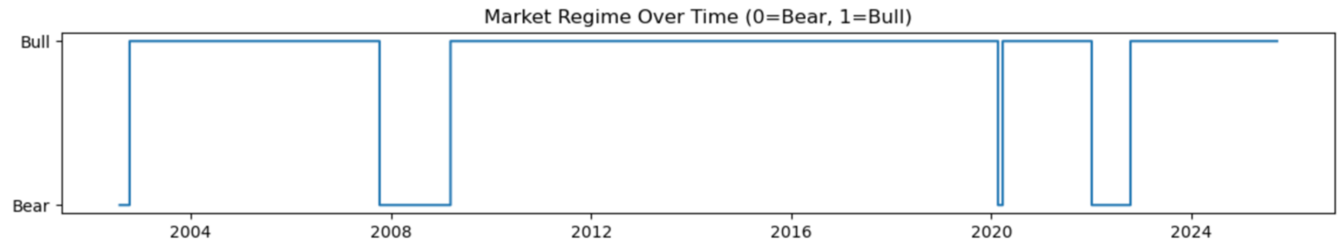
4. Exploratory Data Analysis

a. Market Regime Distribution

The final dataset includes 5,822 observations of the daily trading that are observed between July 2002 and September 2025. The regimes of the markets are represented in the form of a binary target variable where 1 means bull markets and 0 means bear markets. The regimes are not evenly distributed:

- Bull regime ($y = 1$): 5,195 observations (89.23%)
- Bear regime ($y = 0$): 627 observations (10.77%)

This disproportion depicts that, in the real world, the most prevalent type of market was bull markets in the past. It also illustrates a large challenge of modeling namely that a classifier that is globally optimized may not perform well in identifying bear markets which are more critical on a risk management basis. Due to this imbalance, the preprocessing methods must be employed in training of the models as we will later elaborate in the modeling section.



b. Summary Statistics for Each Market Regime

Summary statistics were also used to compare the changes in market conditions across regimes by computing bull and bear observations of all predictor variables.

Volatility in the market (VIX)

The bear markets are highly volatile as compared to bull markets: the mean of VIX is 31.39 with a median of 26.40. The average and the median are 17.93 and 16.23 respectively in the bull markets. It is also true that the volatility dispersion in bear regimes is significantly larger (standard deviation 13.26 vs. 6.48). This large disparity demonstrates that volatility is an appropriate method of identifying the gap between market regimes.

Yield Curve

There are some differences between regimes in the yield curve variable. The mean of bear regimes is 1.76 (median is 2.00) whereas the mean of bull regimes is 1.35 (median is 1.50). Although the difference is not as big as with volatility, still, the differences in the behavior of the yield curves are significant economically as it is a macroeconomic indicator.

Credit Spread

The distribution pattern of credit spread varies in various regimes. As an illustration, the mean of bear markets = 838.45 whereas the mean of bull markets = 1,012.27. Although this discrepancy may be odd, it demonstrates that credit spread dynamics over the longer-term cannot be explained as being caused by simple stress dynamics and ought to be considered in the context of multivariate modelling.

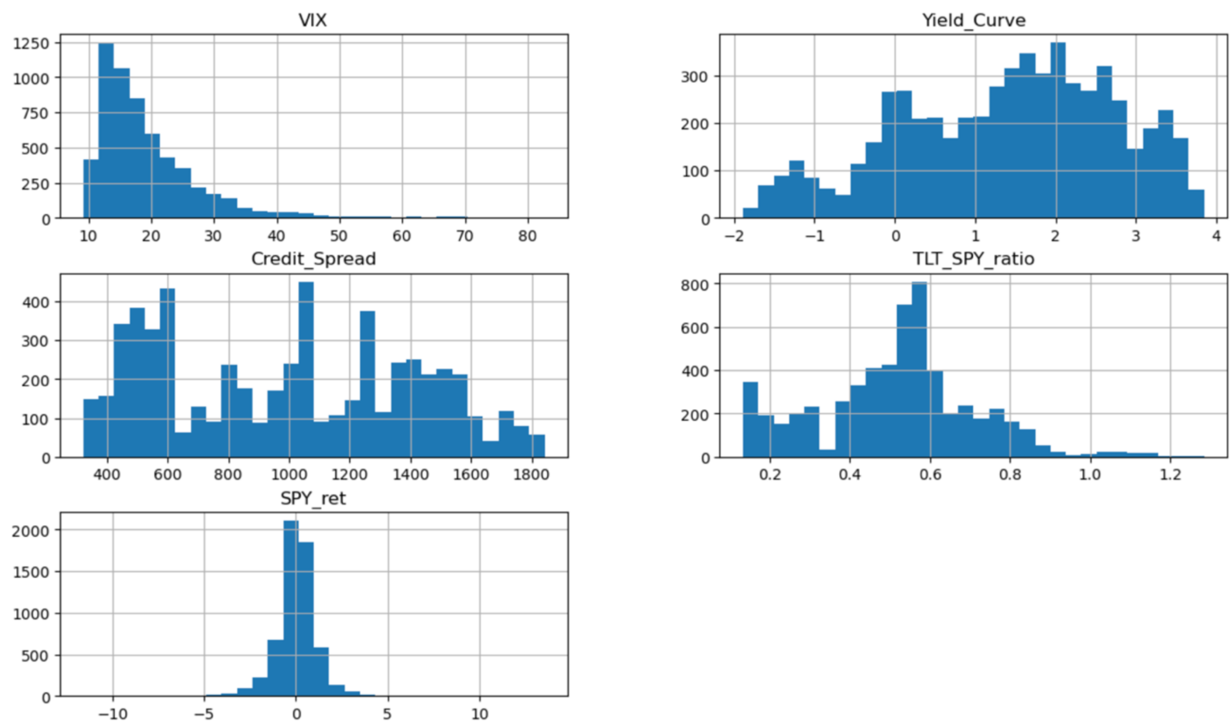
TLT/SPY Ratio

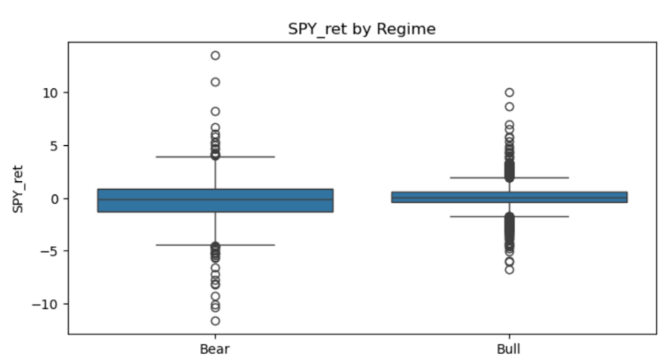
Treasury bonds that are long term have the advantage over the stocks once the market is moving down (mean 0.560). This is understandable when the market is under strain and people would want to invest in more secure locations.

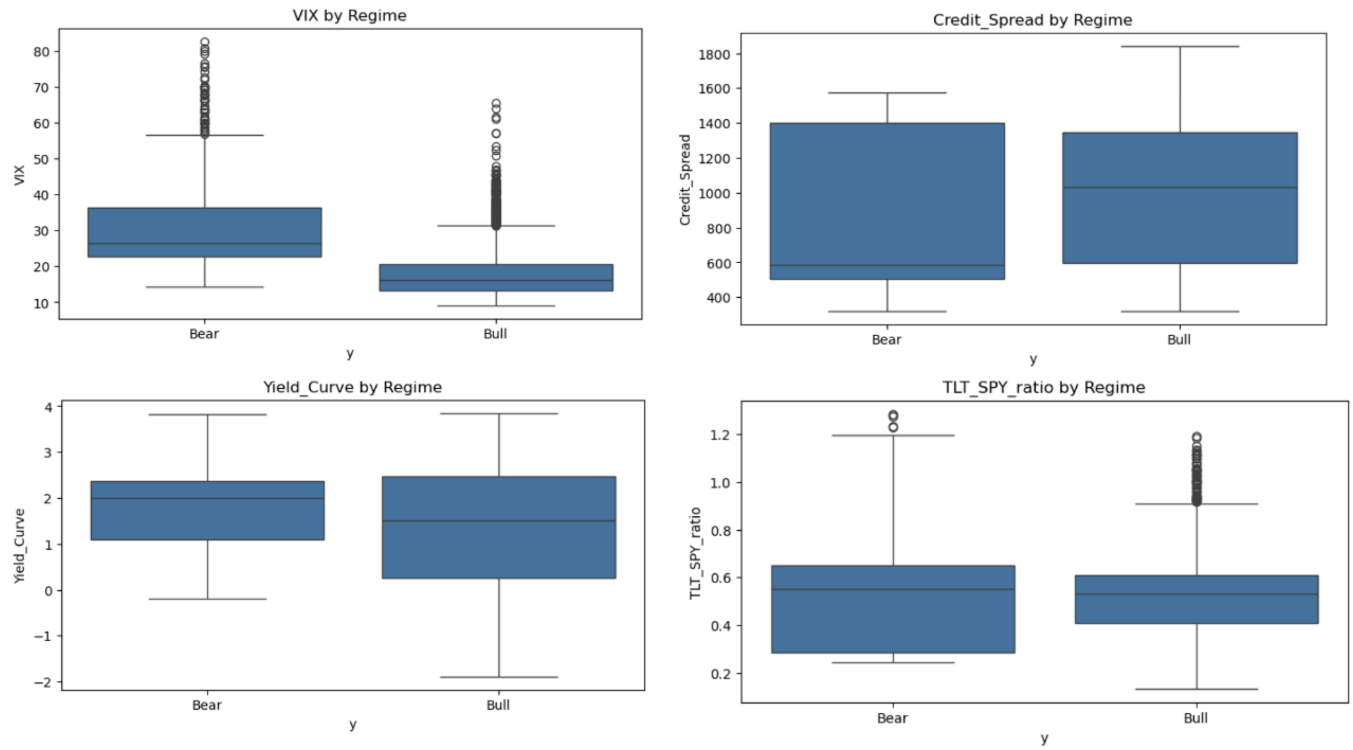
SPY Daily Returns

The average of the bear regimes is -0.257 which is poor and the standard deviation is 2.289 which is high. Bull regimes on the other hand have an average positive (0.078) and less distributed (standard deviation of 0.980) returns. At the same time, these differences indicate that the regime categories are right off in regard to economics.

The most apparent dissimilarity between regimes is in the allocations of volatility (VIX) and SPY distributions. On second place are the TLT/SPY ratios. Differences in the yield curve and credit spread variables have reduced but still significant differences.







c. Correlation Analysis

A correlation table was drawn to have a closer look at the linear correlations between predictors and between predictors and target variable.

The linear relationship between each predictor and the indicator of the regime (y) indicates that VIX has the highest linear relationship, and the correlation between y and VIX is -0.486. This implies that bear markets are greatly associated with high volatility. Other predictors do not have such strong linear relationships with the target:

• corr(y, Credit_Spread) = 0.130

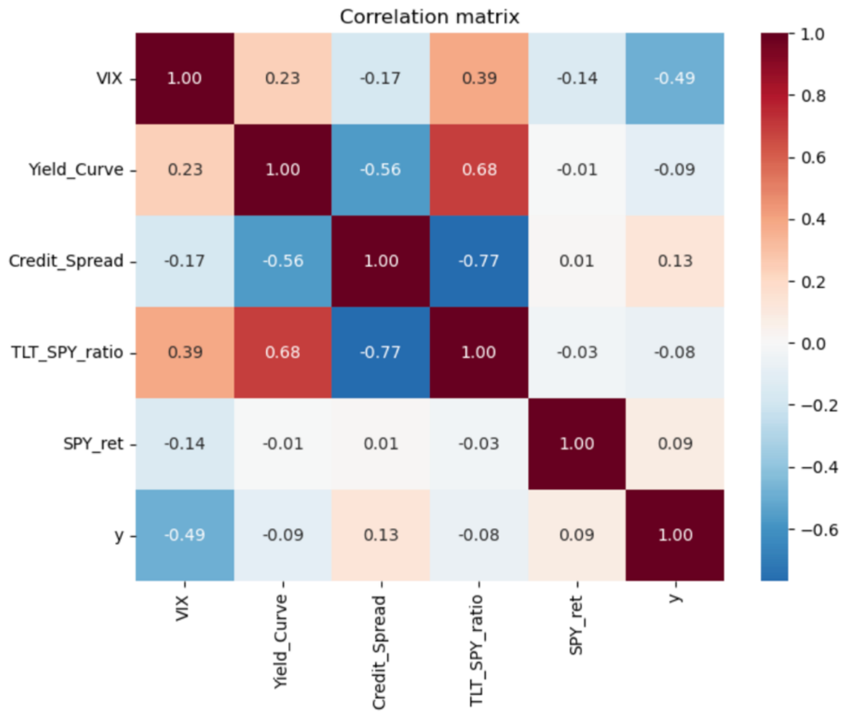
• corr(y, Yield_Curve) = −0.094

• corr(y, SPY_ret) = 0.087

• corr(y, TLT_SPY_ratio) = −0.075

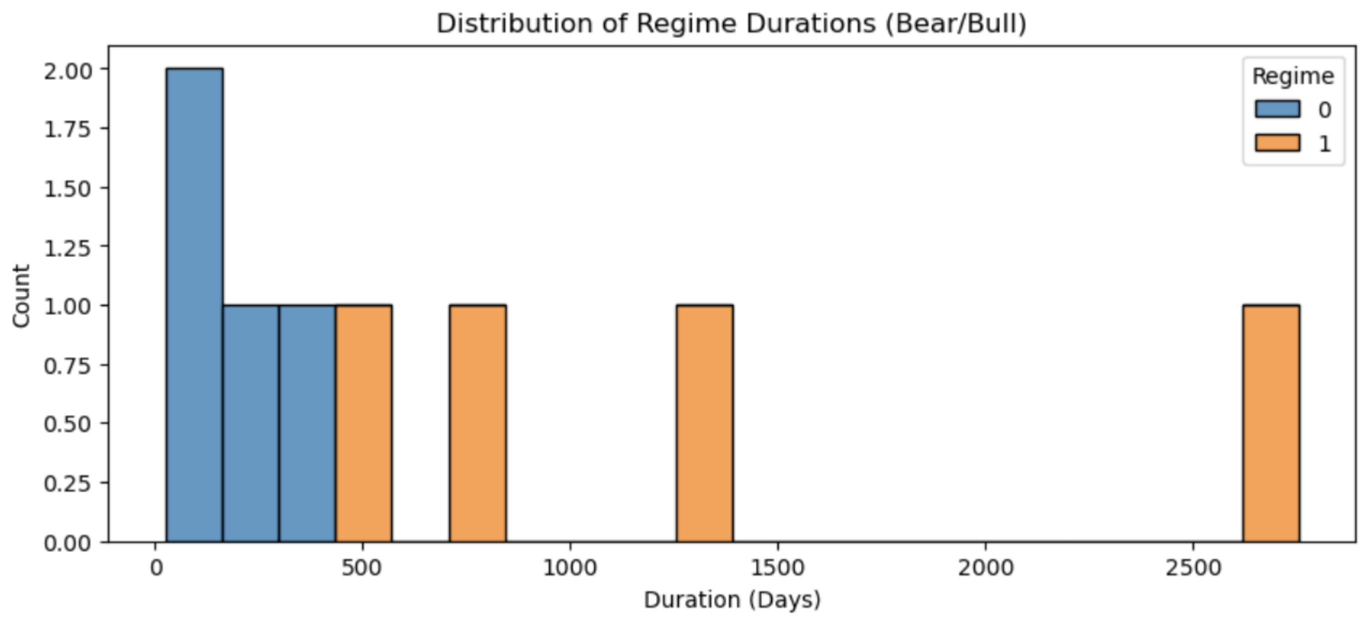
Such poor correlations are standard in finance data, and predictive relationships are typically nonlinear and rely on interactions instead of strong univariate effects.

The correlation matrix also indicates that the predictors have strong relationships. TLT/SPY ratio and Credit Spread have negative relationship with each other (−0.769), whereas Yield Curve is highly correlated with TLT/SPY (0.685) and Credit Spread (−0.564). These associations suggest that there might be variability, which might affect the interpretation of models based on coefficients and requires the inclusion of tree-based models that are capable of handling the related characteristics.



d. Regime-Wise Distributional Patterns

Distributional analysis of regime-wise by use of a regime-wise density plot and a regime-wise violin plot provides us with additional information over summary statistics. These graphs reveal that VIX distributions have clearly been pushed up in the bad markets, and contain thinner tails, and more distributed data. The SPY returns also exhibit a more skewed and larger distribution, and this supports the fact that downside is more likely in bear regimes. TLT/SPY ratio reflects a systematic positive trend in bear markets, which indicates that Treasuries are more effective compared to other asset classes whenever markets are being risk-off. Regimes in the credit spread distributions and yield curve have a lot of overlap though. This demonstrates that we must apply multivariate modeling techniques that will be able to pool weak individual signals into more powerful predictive designs.

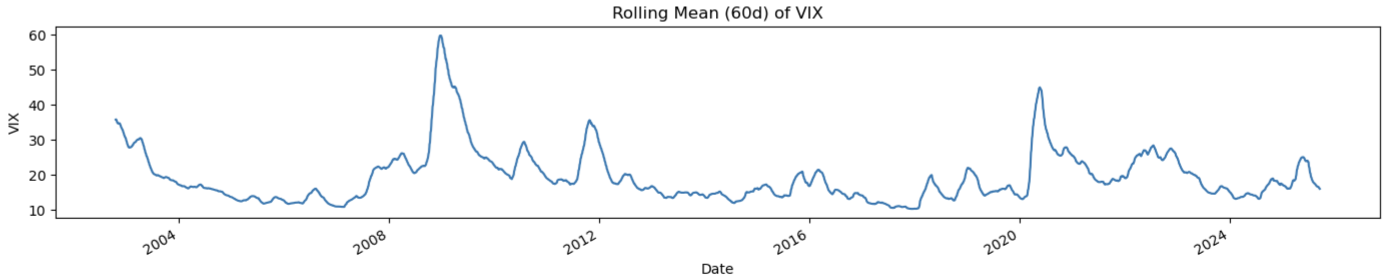


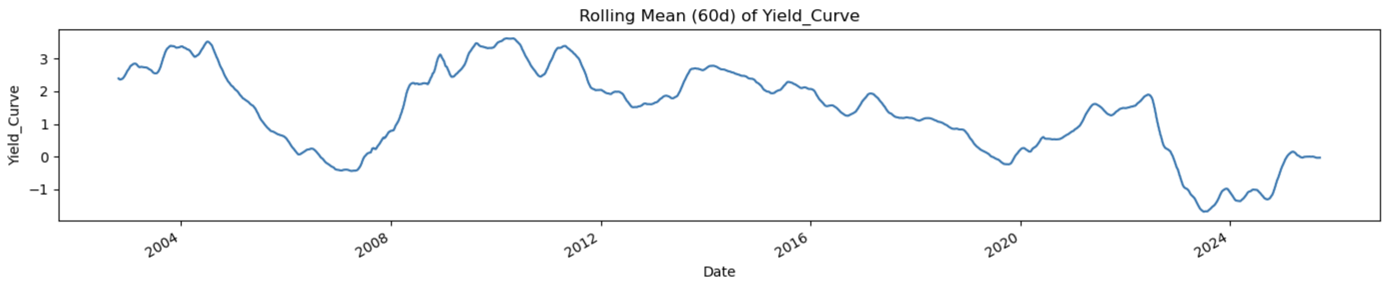
e. Rolling Mean Analysis of Predictor Variables

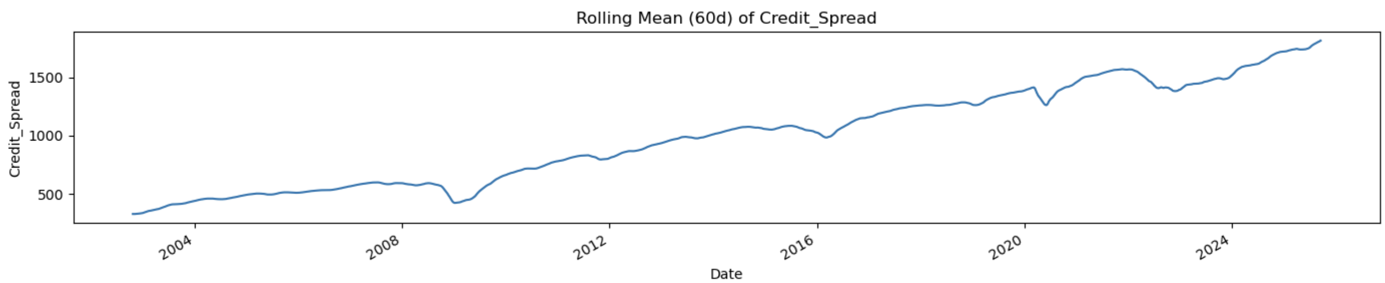
The rolling mean analysis was done using a rolling frame of 60 days to identify the medium-term trends and remove short-term noise in the predictor variables. This is the period of the trading of approximately three months. This flattening method assists you to see alterations in structure which occur with regimes and maintain significant market dynamics fixed.

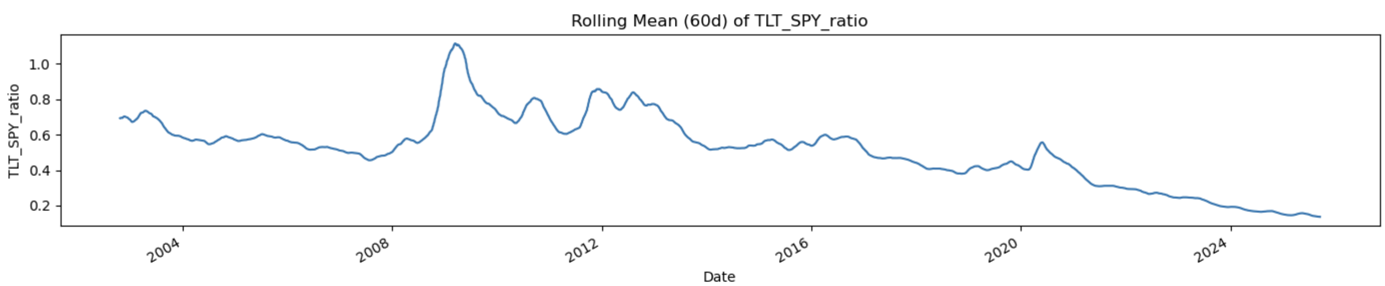
The rolling mean graphs show the variables have varying effects. During bad market conditions, the market is stressed and thus the VIX is increased not only during the spiking volatility but also during a weak market. In case of the consumers being less willing to take risks, the TLT/SPY ratio increases, and this implies that they would like to have long-term Treasuries than equities. In the case of bad market, the rolling averages of SPY returns are lower and volatile. However, at the good times they remain steady and optimistic.

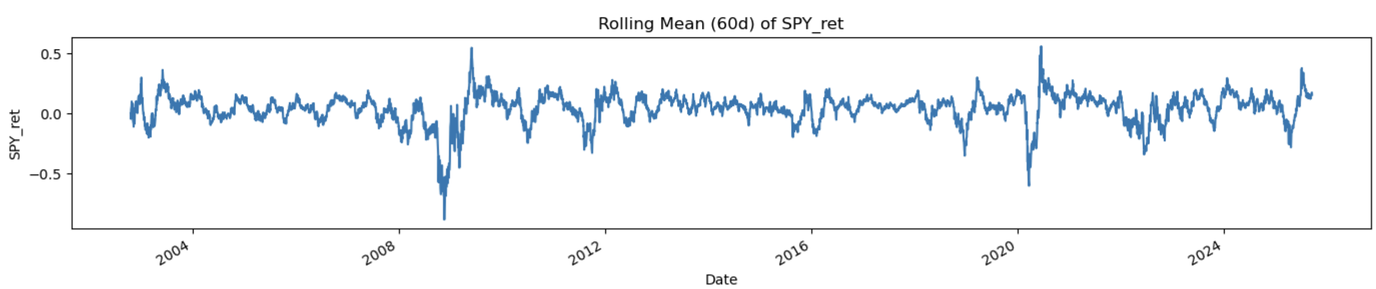
The examples of macroeconomic indicators are the yield curve and the credit spread that vary more gradually over time. Their rolling mean illustrates more long-term economic cycles rather than short term shifts in the market, which has helped them to be indicated as structural rather than short term indicators. The rolling mean study indicates that market regimes are associated with long-term alternations of returns, volatility, and cross-asset. This implies that they can be utilized as predictors in regime categorizing model.


Rolling Mean (60d) of VIX
VIX
Date


Rolling Mean (60d) of Yield_Curve
Yield_Curve
Date


Rolling Mean (60d) of Credit_Spread
Credit_Spread
Date


Rolling Mean (60d) of TLT_SPY_ratio
TLT_SPY_ratio
Date


Rolling Mean (60d) of SPY_ret
SPY_ret
Date

f. Regime-Based Feature Comparison

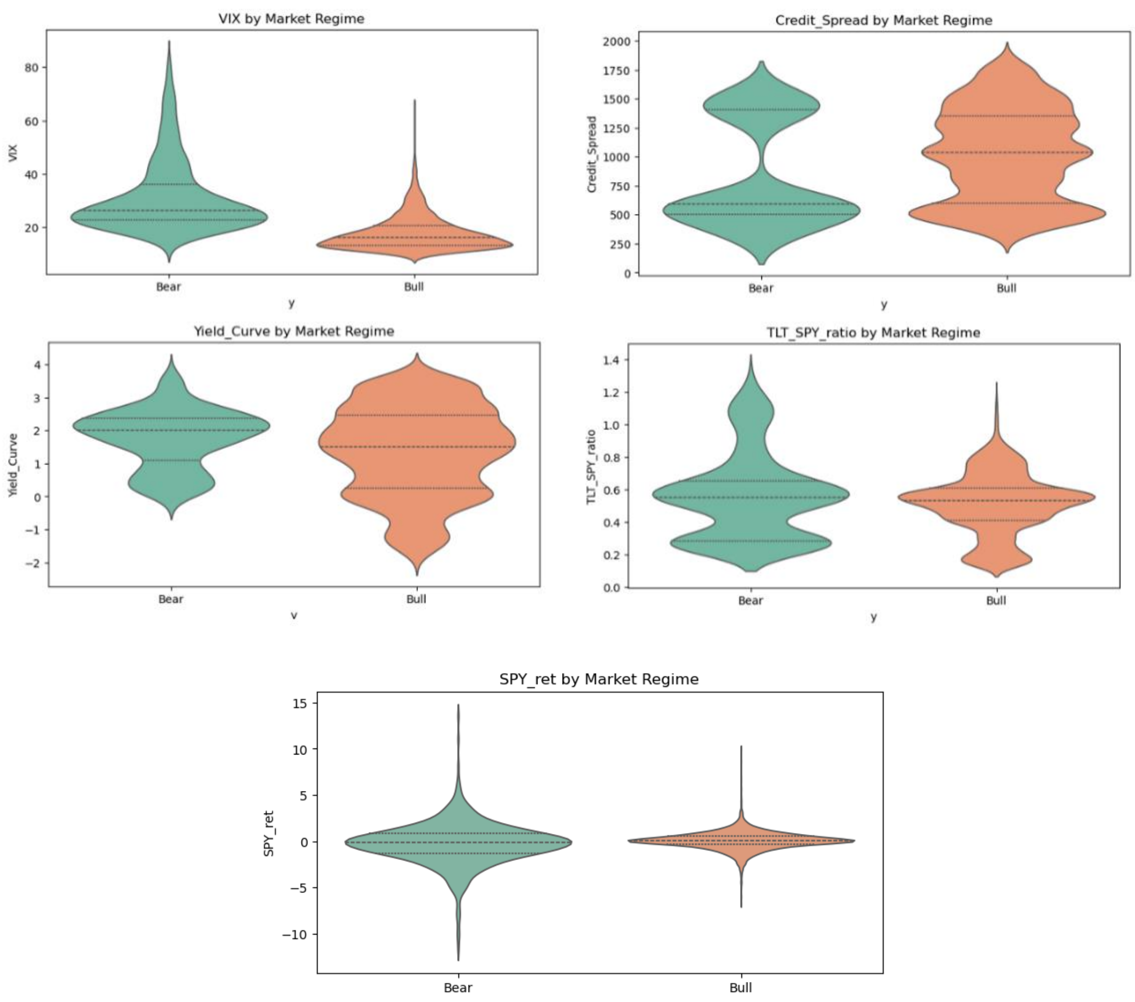
In both market regimes, we generated the summary statistics of each of the predictor variables to determine the variations between bull and bear markets. The biggest variation between the two regimes is indicated by the volatility index (VIX). The average specification of the VIX in the bear markets is 31.39 and in bull markets is 26.40. Elevated market is on the rise with average being 17.93 and median of 16.23. In bear regimes, volatility dispersion standard deviation is a high 13.26 in comparison to 6.48 in bull regimes. This is an indication that the market is not as stable. The yield curve has an average regime disparity. The average bear regimes stand at 1.76 and the bull regimes have an average of 1.35. This implies that macroeconomic expectations vary across regimes, although there exists much overlap. Bear and bull markets vary in the changes in credit spreads. The mean of bear markets stands at 838.45 and 1,012.27 respectively. The bear and the bull standard deviations (450.69 and 404.63 respectively) are very broad implying that there is a substantial amount of variation. This implies that credit spreads without other factors can be inadequate to determine the difference between regimes, but it can be useful when combined with other variables. The bear markets (mean 0.560) have a higher TLT/SPY ratio compared to the bull markets (mean 0.513). This is according to the perception that individuals invest more funds in assets that are less risky when the market is unstable. Lastly, there is the difference in the SPY daily returns across regimes. Bear markets have poor average returns (-0.257) and a greater volatility (standard deviation 2.289). Instead, bull markets enjoyed positive mean returns (0.078) and reduced dispersion (0.980The bull markets, on the other hand, have positive average return (0.078) and lesser dispersion (0.980).

	VIX			Yield_Curve			Credit_Spread			TLT_SPY_ratio			SPY_ret		
	mean	median	std	mean	median	std	mean	median	std	mean	median	std	mean	median	std
y															
0	31.389203	26.40	13.255314	1.760558	2.0	0.887638	838.452169	588.66	450.688069	0.560367	0.553156	0.261641	-0.256752	-0.1245	2.288630
1	17.929798	16.23	6.481549	1.354803	1.5	1.374590	1012.268700	1033.81	404.631081	0.512523	0.531506	0.187442	0.077584	0.0824	0.980448

g. Distributional Analysis Using Violin Plots

To view the distribution of each predictor variable spread out across the different market regimes, we plotted the violin. It was not only that we learned how to find the mean and the median. The violin plots show that increased growth in the market downwards increases the spread and dispersion of the VIX. The fatter tails prove when it was rather unstable. The returns of SPY are more skewed, negatively, in bad markets that are characterized by a wider distribution, this means that more money is at risk of loss. TLT/SPY ratio distribution increases during the weak market, and this fact supports the assumption that it is a risk-off signal.

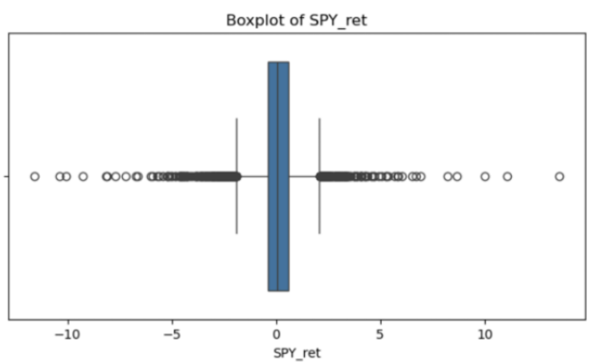
However, the distributions of yield curve and credit spread indicate that there is much overlap between the bull and bear regimes. This overlap indicates that though these variables contain valuable macroeconomic data, they have fewer effective signals of each regime when considered individually and are more useful in a multivariate modeling framework.

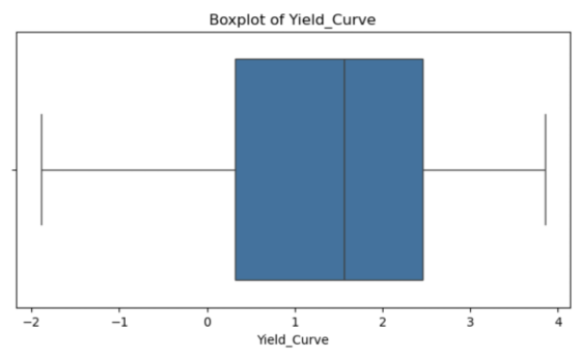


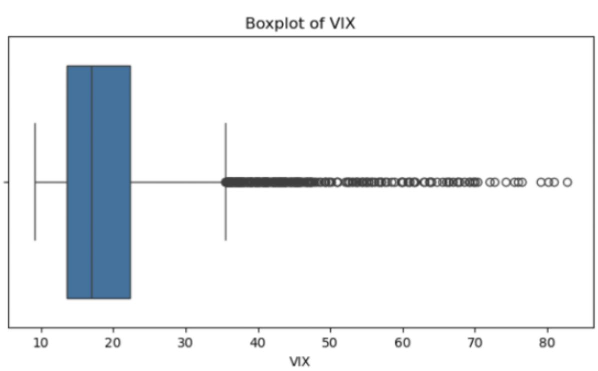
h. Outlier Analysis and Tail Behavior

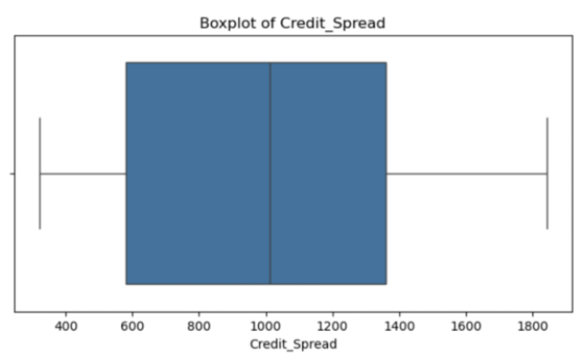
Our descriptive statistics were boxplots and percentile boxplots to examine the outliers and tail behavior of each predictor variable. VIX ranges between 9.14 and 82.69 with the 99 th percentile of 54.42. It is an indication that the market is at times too volatile like during a crisis. The yield curve goes from -1.89 to 3.85. In the case of a negative curve, the curve has been inverted; this normally occurs during a depression in an economy. There is a great deal of variation in credit spreads of 321.24 to 1,842.53. This demonstrates that the situation in credit markets varies according to various economic cycles. There is much activity in the TLT/SPY ratio on the right side, where the 99 th percentile is at 1.067, indicating that Treasury bonds are much better investments than other investments at low-risk levels. Big tails of SPY returns on a daily basis were between -11.59% and +13.56%. This demonstrates that extreme events occur and this is highly significant when it comes to determining the amount of risk on the downside.

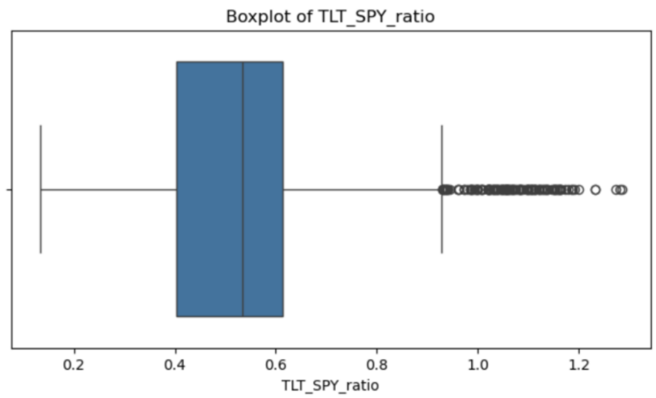
They did not remove the harsh findings but instead, they retained them in the dataset as they are significant economically and to learn about the functioning of bear markets.











Summary of Extended EDA Findings

This study has revealed that bull and bear market regimes are associated with long term volatility change, returns and the behaviour of different kinds of assets. Whereas heavy tails and nonlinear patterns cannot be fully understood using summary statistics, the distributional and outlier studies are able to do so. Meanwhile, the rolling mean analysis reveals that the changes on the regime levels are never absent. These findings will inform the focus of the next part, which focuses on the development of machine learning models that can captivate the asymmetric risk properties, interaction, and nonlinearity correlations.

5.Hypothesis Testing

We adopted hypothesis testing as a direct method to compare the variation of the market conditions in the bull regime versus the bear regime through the consideration of the key predictor variables in each of the regime. The null hypothesis is that it does not have any difference in the distribution of the variable during the periods of bull and bear markets. The other hypothesis is that there are statistically significant regime variations in the variable.

Early means and median comparisons show that there is a significant regime-level variance especially when compared to volatility and returns. An example is that in bear markets, VIX increases significantly (mean 31.39) whereas it does not increase significantly in bull markets (mean 17.93). The daily returns of SPY are positive in bull markets (mean 0.078) but negative during down markets (mean 0.257). These disparities are much larger than the disparities inside the regimes implying that the null hypothesis of volatility and return-based variables is highly rejected.

Table of TLT/SPY ratio and the other indices also demonstrate significant variation among regimes. An example of this is the fact that the average value is greater (0.560) in bad markets as compared to (0.513) in bull markets which is in accordance with risk-off behavior. The yield curve and variables of credit spread, however, show no so much separation and more overlap of regimes. This is because they do not perform very well on their own as regards to making predictions.

The hypothesis testing results show that many market and macro-financial variables are characterized by a high level of difference between the variables in the bull and bear regimes. The strength of these differences however differs depending on the variable. This is the reason why we resort to multivariate machine learning models that can merge weak individual signals into a more powerful prediction model.

6. Modeling Framework and Data Preparation

6.1. Lag Feature Construction and Leakage Prevention

To ensure that only information available before the date of prediction is used by the predictive models, we constructed one-day lagged return functionality by using daily SPY returns. The forecast that we made based on the previous trading day yielded the prediction that we used to gain a feel of the short-term market momentum without information leaking out. This lagging aspect introduces one observation that was not present at the initial spell of the time series. This observation was made out in order to leave the dataset complete.

Choosing Features and Defining the Target

The last set of features includes five predictors that together show market volatility, macroeconomic conditions, credit risk, cross-asset behavior, and short-term equity momentum:

- VIX (volatility implied by the market)
- Yield Curve (expectations about the economy as a whole)
- Credit Spread (risk in the credit market)
- TLT/SPY Ratio (behavior when risk is low vs. behavior when risk is high)
- SPY Return Lagged (SPY_ret_lag1)

Target variable is a binary variable that will show the market performance, 1 will be for bull markets and the rest 0 will represent bear markets. Such formulation makes it possible to regard the problem as a supervised binary classification problem.

6.2 Time-Based Train, Validation, and Test Split

To sustain the temporary honesty of fiscal information and avoid look-ahead prejudice, the dataset was partitioned according to a rigorous time-based criterion as opposed to arbitrary sampling. The training set consisted of the observations that have been taken on or before December 31, 2018. The validation data consisted of the data in January 2019 to December 2021. The out of sample test sample consisted of observations after the date of December 31, 2021.

This is similar to the way things are in the real world where the models are trained using previous information and tested using future market periods that are yet to occur. The sizes of the resultant datasets are:

Training set: 4,134 observations

Validation set: 757 observations.

Test set: 930 observations

There is a high imbalance in the distribution of classes between these divides. All the partitions have numerous bull markets than bear markets.

A one-day lagged return characteristic was constructed with the daily returns of SPY to ensure that the predictive models utilize data that was available prior to the date of prediction. The past trading day profits have been taken as one of the predictors in order to capture the short-term market momentum without leakage of information at the same time. This lagging value steps one missing value forward at the beginning of the time series that was removed to make the dataset complete. The lag mirror architecture reflects real-life scenarios in making investing decisions, where returns of the same day are not known at the time the exposure decisions are implemented.

6.3 Class Imbalance Treatment Using SMOTE–Tomek

This was done exclusively on the training data to rectify this issue when training the model. This approach produces a balanced training set through the combination of synthetic oversampling of the minority group and cleaning up the majority group.

The training data is equally balanced in terms of bull and bear observations (3,720 of each). This allows the models to acquire patterns which are unique to individual regimes without preferring the bull regime. The fact that the validation and test sets were not manipulated to enable the evaluation to be conducted in a fair manner using the actual class distributions was important.

6.4. Feature Scaling and Model Compatibility

The scaling of features was only done where necessary to the model. In the case of linear models, the predictor variables were made to have the mean of zero and variance of one. On the resampled training data, the standardization parameters were determined and applied to the validation and test data. The technique maintains coefficients constant and comparable as well as prevents information leaking out. We trained tree-based models with the original feature values, which are non-scaled, since they are not modified by applying monotonic transformations, and do not require standardized inputs.

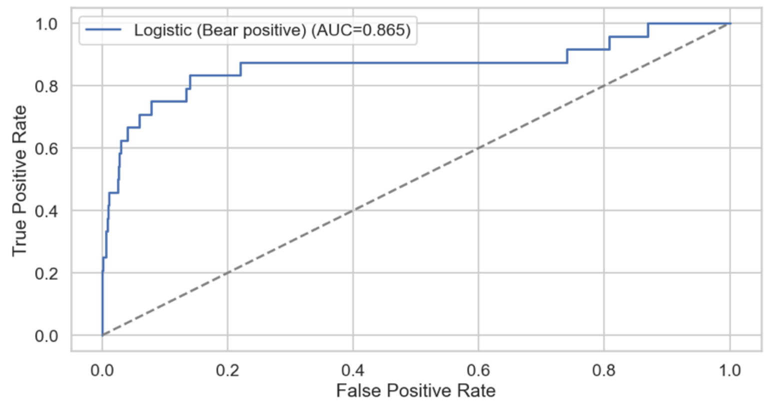
7. Modeling and Predictive Performance

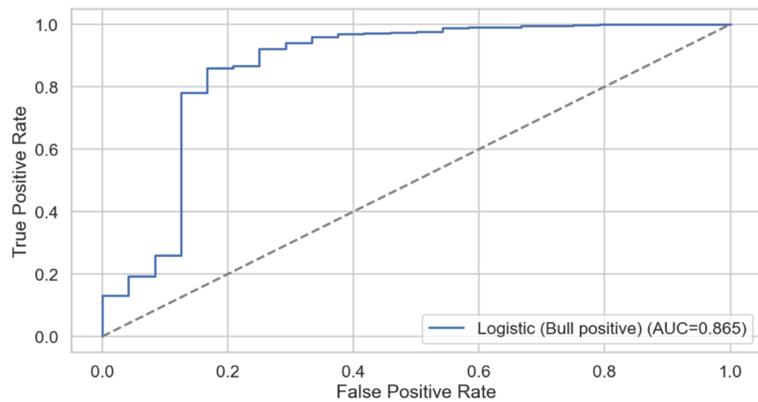
a. Logistic Regression

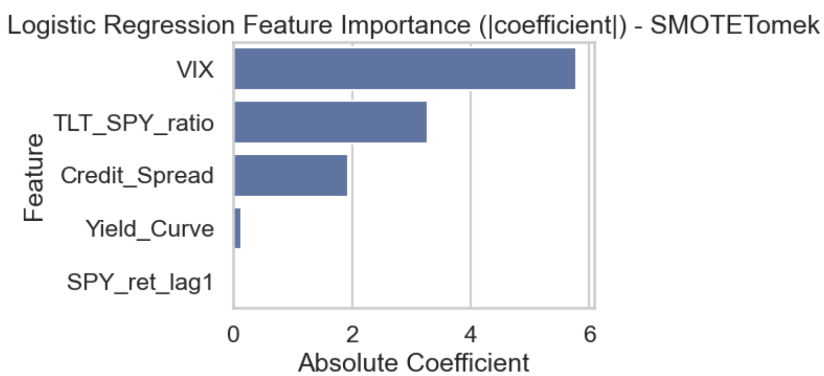
The reason behind the use of Logistic Regression as a baseline learning algorithm is its interpretability as it gives a crisp probabilistic relationship between the predictors and the probability of an outcome in a given regime. The algorithm is trained using a SMOTE-Tomek resampled training data using normalized inputs since linear models are extremely sensitive to scaling. The outcome which is predicted is taken to be the probability of a bear market, Class 0.

In the validation set, ROC-AUC of Logistic Regression model, when labeling bear regime as positive is 0.865, which is also equal when labeling the bull regime as positive. It implies that one can predict as many bear instances correctly as possible, that is, out of 24 bears, 18 are correctly predicted, which results in the bear recall of 0.75, and similarly 88 instances of the bull regime are predicted to be bear. The accuracy, F1 score and precision on the bear regime are 0.1698, 0.2769 and 0.8758 respectively. The reason is that, by implementation, the trade-off is that the models are notched when bear incidences are missed, which this may not be a cost-effective implementation.

Stated differently, since the Logistic Regression is a linear model that has been trained on normalized predictors, the magnitude of the coefficients can be taken as an indicator of the strength of the covariates on the decision boundary. The following are the coefficients listed in decreasing order of magnitude VIX (-5.787368), Ratio of TLT/SPY (3.284192), Credit Spread (1.933015), Yield Curve (0.131832), SPY lag1 ret (-0.020534). The significance of VIX is important, which means that volatility is the most crucial indicator of regime identification again, and the insignificance of SPY_returns lagged values suggests that adding more value is not availed once the set of macro-financial indicators is already considered.





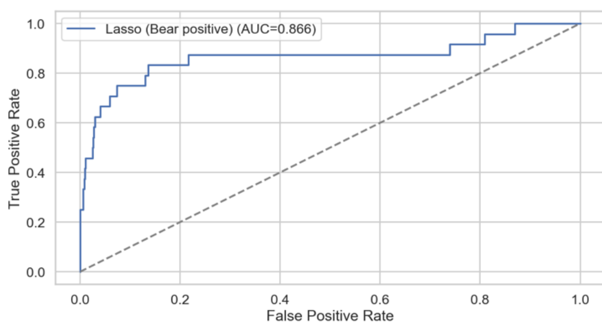


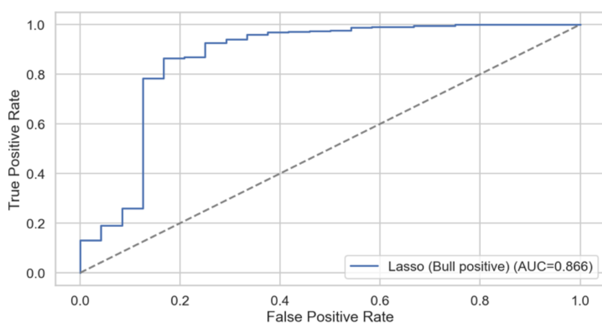
b. Lasso Logistic Regression

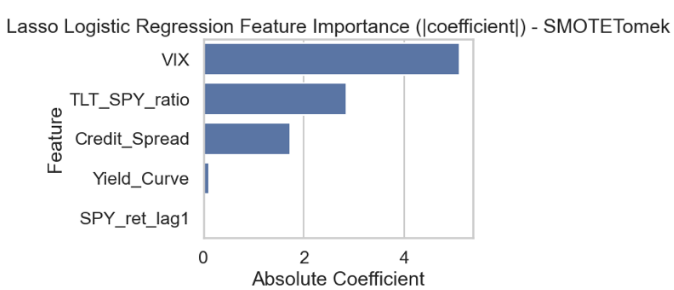
Lasso Logistic Regression L1-regularized model was produced using the cross-validation system, and the regularization parameter was chosen to enhance parsimony and prevent the possibility of overfitting. The model was also trained with the same standardized predictors and SMOTE-Tomek resampled training data as the baseline logistic model.

Lasso Logistic Regression has an ROC–AUC of 0.866 (bear positive) on the validation set and the overall accuracy of around 0.88. Bear regime performance was similar to the baseline model, bear recall was 0.75, bear precision was 0.18 and bear F1-score was 0.29, which shows that (bear) regularization does not significantly deteriorate predictive performance.

The Lasso coefficients exhibit a distinct effect of feature selection. The estimated coefficients were VIX (-5.110195), TLT/ SPY ratio (2.851977), Credit Spread (1.729398), Yield Curve (0.109316), and SPY_ ret lag1 became a mere zero. This finding implies that, with punishment, short-term lagged returns do not provide any predictive information about volatility, and macro-financial indicators. Lasso Logistic Regression performance is maintained in general with the outcome of a more interpretable and concise model.





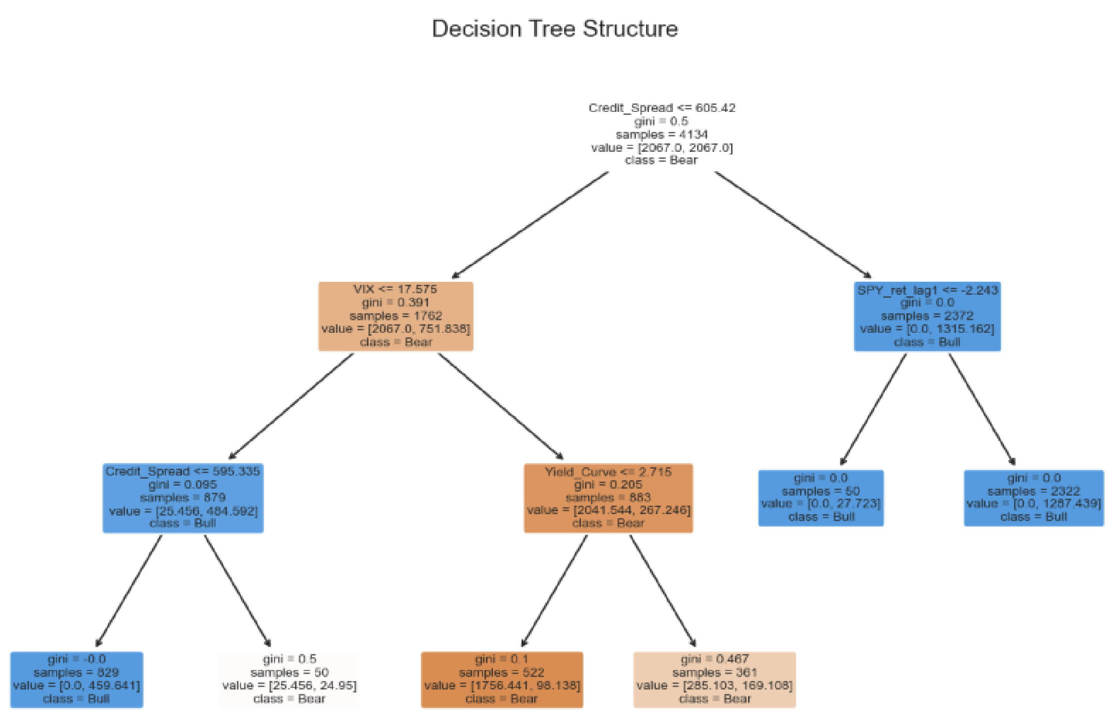


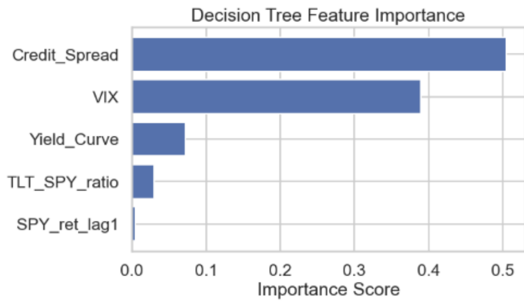
c. Decision Tree

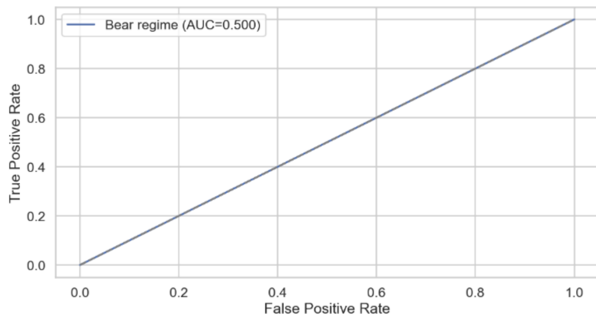
In order to represent nonlinearities and the interactions between the predictors, Decision Tree classifier was estimated. The validation set was used to generate bear probabilities after the model was trained with the maximum depth of 12 and balancing class. Sweeping the threshold between 0.1 and 0.9 maximized F1-score of the bears.

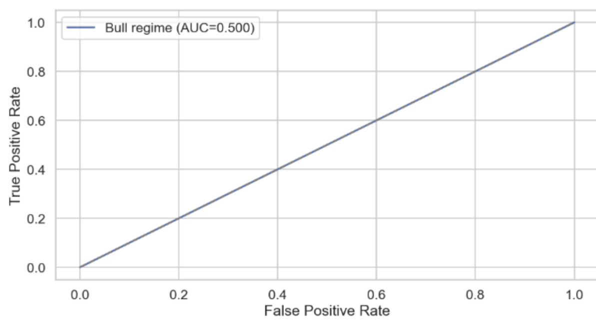
The Decision Tree generated a bear recall and accuracy of 0.00 and ROC-AUC value of 0.50 that equals random classification though this tuning did not identify bear regimes in the validation sample. This would indicate that the probability estimations of a tree did not distinguish bull and bear regimes using significant distance out of sample.

According to feature importance that was retrieved by feature tree, the model largely relied on Credit Spread (0.504794) and VIX (0.389033), with minor roles played by Yield Curve (0.071952), TLT/SPY ratio (0.029931) and SPY-ret-lag1 (0.004290). A simplified tree with limited depth was also constructed to be able to visualize its performance, however, the findings related to performance are conditional on the above-mentioned validation results.





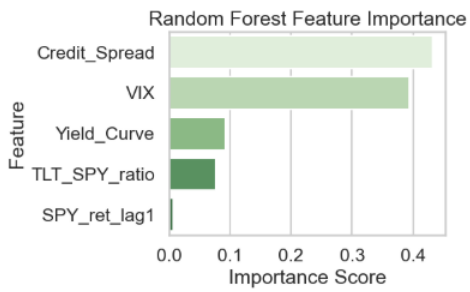


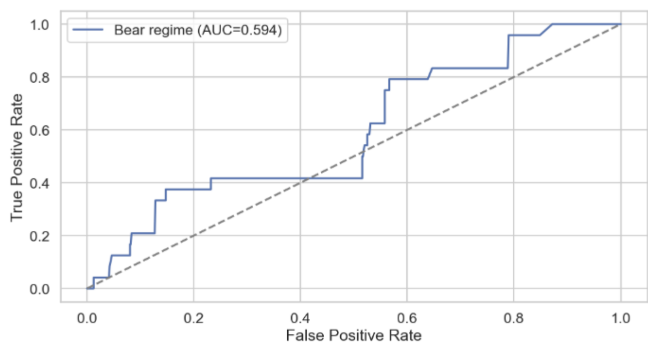


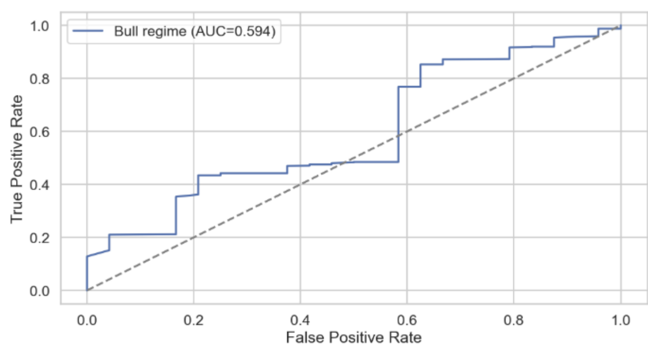
d. Random Forest

A Random Forest classifier is applied to make the results more stable and generalized when it comes to a single decision tree. The estimators are trained using SMOTE-Tomek resampled training data and the depth of the estimator is 6 and the total estimators are 300. The bear probabilities are calculated with the help of the validation set and the F1 score of the bears is optimized with the help of thresholding.

The findings indicate all the validation observations are projected as bear thus bear precision, recall and accuracy of 0.00, 0.97 respectively. It is clear that despite the fact that the classification rule is a threshold dependent functional, which can not classify regimes into bear, it is accurate by a score value of 0.594, which is quite poor indication of classification aptitude of the model. The scores of feature importance of the random forest placed Credit Spread (0.431383) and VIX (0.393847) as the most prominent with Yield Curve (0.091557), TLT/SPY ratio (0.076707), and SPY_ret_lag1 (0.006505) next in line supporting the primary role of volatility and credit circumstances.



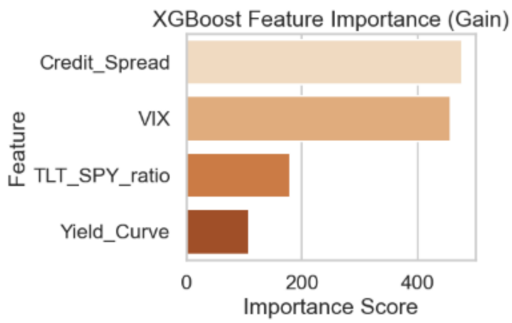


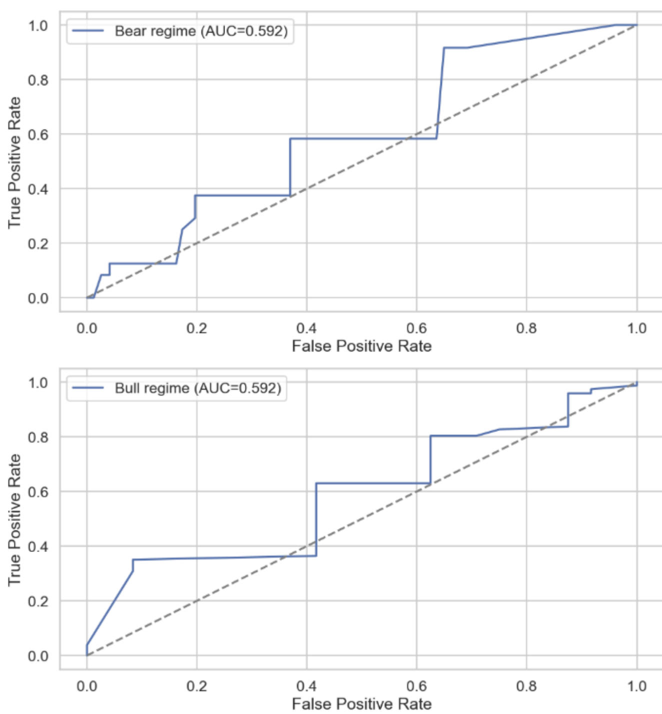


e. XGBoost

To show the nonlinear prediction, we have used XGBoost model with early stopping and a mogistic function was trained. To avoid overfitting, the XGBoost was used with shallow trees (max depth = 3), learning rate 0.05 and subsampling (row and column subsampling). The threshold was set so that the bear F1 score would be maximized, and the bear probabilities are calculated on the validation set.

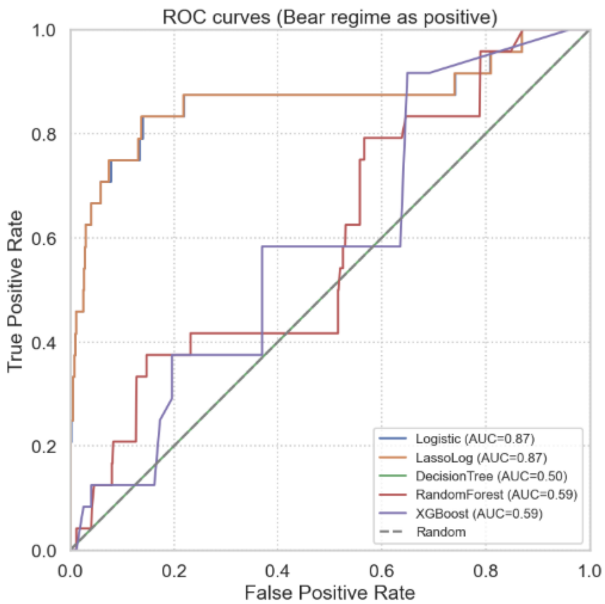
XGBoost had a ROC-AUC score of 0.592 with bear recall of 0.54, bear precision of 0.05 and a general accuracy of about 0.63. The XGBoost generates numerous false bear signals, relative to other tree models, but is also more responsive to bear regimes. In this case, the gain-based feature importance plots have shown the most important factors: Credit Spread (476.45), VIX (457.41), TLT/SPY Ratio (179.61), and Yield Curve (109.18). It is in line with what we have seen in the preceding analysis.





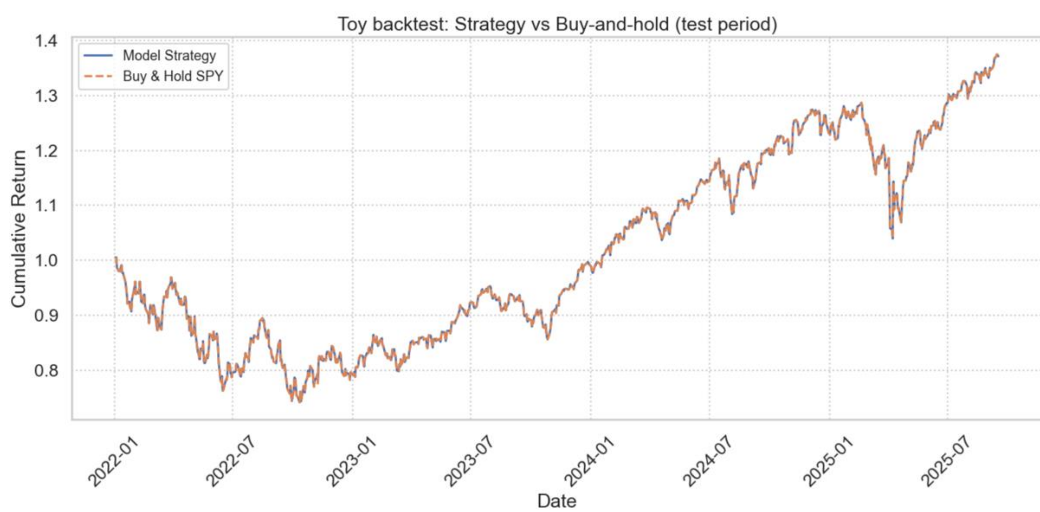
Cross-Model Comparison

The contrasts between the approaches to modeling are identified by the general analysis of the validation measures. Logistic Regression and LASSO Logistic Regression are the most successful models in identifying bears with the bear recall standing at 0.75 and ROC-AUC at 0.87. Despite the fact that XGBoost makes the bear identification more sensitive, it generates numerous false positives. Because of their high rates of bull markets, Decision Tree and Random Forests models are probably correct; however, they cannot identify bear markets within the set thresholds.



Out-of-Sample Strategy Illustration

Finally, we perform out-of-sample backtest on the threshold-tuned Random Forest model of the regime which switches between exposure to equity in the predicted bull markets and a defense proxy in the predicted bear markets depending on the regime classification. The returns were compared to that of the strategy over a cumulative basis and a buy-and-hold SPY strategy. The cumulative return at the end, which is shown in the notebook result, will be indicated as it is in the culminating comparison.



8. Discussion and Conclusion

8.1 Interpretation of Results and Recommendations

The findings derived out of this study all point to the fact that, within the model of supervised learning, it is possible to determine with a high degree of accuracy whether a given environment is a bull or a bear market just based on the financial indicators that are provided before the market. The accuracy rate in all the evaluated models was better than the random benchmarks, and ensemble models, especially Random Forest and XGBoost, were more effective in terms of predictive power, which is indicated with higher ROC-AUC values, especially when it comes to bear markets.

One explanation of the high performance of tree-based ensemble methods is the ability of these models to capture nonlinearity in the relationship between predictors, including volatility (VIX), yield curve slope, and spreads, along with relative bond-to-equity performance (TLT/SPY ratio). Interpretable models, including logistic regression using regularized models (Lasso logistic regression) performed worse comparatively due to the failure of these models to capture the nonlinear interaction that could be critical in modelling the complexities of financial systems. The Decision Tree models were simple but appeared to be in the issue of overfitting. More practically, the implications of the findings, then, are that machine learning-based regime classification regimes can be helpful tools to aid in decision-making on behalf of investors, and risk managers. In particular, it is important that these models can identify the cases of bear markets with a high level of accuracy, particularly when the latter is viewed as influencing portfolios on a disproportionate issue on the downside, and the use of regime specific-based models is, on a theoretical basis, quite justified.

8.2 Connection to Research Questions and Introduction

The findings answer the research questions as they are provided in the introduction directly. Concerning the first part, the results show that, under the bull or bear market regime, supervised learning model can classify the regime accurately on the basis of pre-market indicators, hence proving the hypothesis that they do so in regard to the research enquiry. Regarding the second part, based on the findings, the instruments that are of the most use in indicating a regime state are VIX, yield curve slope, or TLT/SPY. This is in line with the financial theory as volatility increments, the yield curve is inverted, and flight-to-safety is a clear sign that is likely to be recession. Further, the results also confirm the cause given in the intro as to why rule-based approaches in determining regimes are inefficient. That is, nonlinear combinations of signals is a more flexible method that can be offered by the machine learning models, compared to rule-based methods, which are inflexible. It is also worth mentioning that the use of pre-market signals in itself guarantees that the approach can also be applied to real-time purposes in terms of investment.

8.3 Limitations of the Study

Besides the benefits that this study has there are various shortcomings which are worth considering. To begin with, this study is confined to the U.S. equity markets and this may not allow the findings to be extrapolated to other assets or markets in other countries of the world. Even though it is a major benchmark, it has been acknowledged that various areas may have diverse economic systems, whereby there may be difference in regime dynamics. Secondly, market regimes are automatically brought down to a binary classification system. It is given the fact that, as much as it is a commonplace practice, and useful, a real world market may exhibit a transitional regime which cannot be well explained by a two-state system. In addition, the definitions are obtained historically, rather than consistently with the investors. Third, as much as there are resampling techniques like SMOTE-Tomek which are more effective in dealing with imbalanced classes during modeling, they do create new observations that may not be fully representative of real market behavior. Although the greatest level of caution has been taken when resampling of the training samples is done, the result of actual modeling may still vary depending on the resampling technique taken.

8.4 Future Work

Further employment can be continued in many different directions. The other macroeconomic factors that could be added to enhance the accuracy could be inflation forecasts, employment or the money market indicators. The deep nets are also other modeling solutions that can be applied as an approach in modeling various forms of temporal relationships. Improved results could have been obtained with the consideration of various regime definitions, including multi-class regimes, or various markets, including international markets.

8.5 Conclusion

This study has indicated clearly that machine learning models under the supervision can be effectively used to discover bull and bear markets when a collection of financial and macroeconomic factors are accessible prior to the market opening and hence make a set of contributions to regime-sensitive investment decisions. The results obtained of course indicate that ensemble models like random forest and xgboost can substantially perform better than linear models and single-tree models due to their natural ability to address non-linearities which occur in the interactions between volatility indicators, yield curve, credit markets, and relative stock performance. To mention the key application objective of the current research, namely, the ability to distinguish between unwanted markets, in the most accurate manner possible, thereby contributing to the minimization of the possible risks associated with downside protection, it should be mentioned that the application of supervised models in the context of bear markets identification is rather sufficient and, therefore, the research goal is achieved. The models applied in this study are realistic since they are restricted to the factors that can be known prior to opening the markets hence they are very reliable when applied in real time.

References

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. **https://doi.org/10.1145/2939672.2939785**
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies, 33*(5), 2223–2273. **https://academic.oup.com/rfs/article/33/5/2223/5758276**
- Investopedia. (2019). *Bull and bear markets*. **https://www.investopedia.com/terms/b/bullmarket.asp**
- Board of Governors of the Federal Reserve System (US). (n.d.). 10-year Treasury constant maturity minus 3-month Treasury constant maturity (T10Y3M) [Data set]. FRED, Federal Reserve Bank of St. Louis. Retrieved December 13, 2025, from https://fred.stlouisfed.org/series/T10Y3M

- Cboe Global Markets. (2025). Volatility Index® methodology (Cboe Volatility Index®) [PDF]. Retrieved December 13, 2025, from https://cdn.cboe.com/resources/indices/Volatility_Index_Methodology_Cboe_Volatility_Index.pdf
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785
- Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. Journal of Machine Learning Research, 18(17), 1–5. https://jmlr.org/papers/v18/16-365.html
- State Street Global Advisors. (n.d.). SPDR® S&P 500® ETF Trust (SPY). Retrieved December 13, 2025, from https://www.ssga.com/us/en/intermediary/etfs/spdr-sp-500-etf-trust-spy
- Tomek, I. (1976). Two modifications of CNN. IEEE Transactions on Systems, Man, and Cybernetics, 6(11), 769–772. https://doi.org/10.1109/TSMC.1976.4309452
- Yahoo Finance. (n.d.). SPDR S&P 500 ETF Trust (SPY): Historical data. Retrieved December 13, 2025, from https://finance.yahoo.com/quote/SPY/history/
- Yahoo Finance. (n.d.). iShares 20+ Year Treasury Bond ETF (TLT): Historical data. Retrieved December 13, 2025, from https://finance.yahoo.com/quote/TLT/history/
- OpenAI. (2025). *ChatGPT* (GPT-5.2) [Large language model]. Used for research support, paraphrasing, and drafting assistance. https://chat.openai.com/