# Introduction:

Idea of regime change is natural and intuitive. Regimes identified by econometric methods often correspond to different periods in regulation, policy and other secular changes. In other words, different regimes will correspond to periods of high and low volatility, and bull and bear market periods.

Regime switching models will capture stylized behavior of many financial series including fat tails, persistently occurring periods of turbulence followed by periods of low, skewness and time-varying correlations. Through trend filtering algorithm, we will identify recession periods and periods of stock market crashes.

# My project objectives:

This project presents Logistic Regression, Decision Tree Classifier, Random Forest Classifier and XGBoost Classifier algorithms built on Historical values of the Asset classes to predict Regimes which acts as Target or Dependent variable with respect to time and attempts to interpret the Bullish and Bearish periods in stock markets.

# Target Audience:

My proposed solution will be useful for retail investor who is a non-professional investor and who lack the education and will lose their money for most of the times.

# Hypothesis, Key questions investigated:

1) How am I able to predict Bullish regimes (represented by1) and Bearish regimes (represented by 0) with regards to asset classes returns?

2) Am I able to take better decisions by understanding issue as in 1.what is my accuracy of predictions?

# Data Sources, Modeling Techniques, EDA used:

The data is obtained from Federal Reserve Economic Data. FRED is an online database consisting of hundreds of thousands of economic data time series from

Scores of national, international, public, and private sources.

# What has been done Earlier and what would

# be my Benchmarks

The key determinant of whether the market is bull or bear is not just the market's knee-jerk reaction to a particular event, but how it's performing over the long term. Small movements only represent a short-term trend or a market correction. Whether or not there is going to be a bull market or a bear market can only be determined over a longer time period.

However, not all long movements in the market can be characterized as bull or bear. Sometimes a market may go through a period of stagnation as it tries to find direction. In this case, a series of upward and downward movements would actually cancel-out gains and losses resulting in a flat market trend.

In a bull market, there is strong demand and weak supply for securities. In other words, many investors wish to buy securities but few are willing to sell them. As a result, share prices will rise as investors compete to obtain available equity. In a bear market, the opposite is true i.e. more people are looking to sell than buy. The demand is significantly lower than supply and, as a result, Assets classes’ prices drop.

It is also a known fact that investor psychology and sentiment affect whether the market will rise or fall.

Because the businesses whose Asset classes are invested on the exchanges are participants in the greater economy, it is already known that the stock market and the economy are strongly linked.

Based on these known factors, Market analysts come up with their own fundamental Analysis with some key Market ratios and Technical indicators. This has been already done earlier which roughly give an estimate of the Bullish and Bearish Regimes.

I have put forth my Key questions investigated as above. I would like to know what are my accuracies in able to predict Bullish and Bearish regimes by using the Asset class returns as Independent variables and Regimes period(Bullish as 0,Bearish as 1)as Target variable.

My major milestone targets would be trying simpler and easy to understand solutions which can get most effortlessly applied by any non-professional investor and which still helps them to optimize their portfolios.

# Major Action Items Implemented:

Assumptions of Time series Model are that Data should be stationary. Non stationary data may show the trend. But we should convert non stationary data to stationary data to do better forecasting. We will convert the features into stationary form by applying the necessary transformations namely Differencing and Log transformation. Then we will Add 1, 3, 6,9,12 months lags of the features.

Then I have employed different Modeling algorithms on the transformed data. I am using Regime as Target or Dependent Variable to predict bullish or bearish Regimes.

FRED is an online database which provided me the actual returns data for the Asset classes. It also provided Regime\_index1.The index is a benchmark market index that represents the weighted average of all the Asset classes’ returns. Based on Bullish or Bearish trends, Regime\_index1 marks the Bullish Regime as 0 and Bearish period as 1.Similarly Regime\_index2, Regime\_index3, Regime\_index4, Regime\_index5 is obtained.

I have used Logistic Regression modeling several times by tuning hyper parameters differently by varying hyper parameter penalty as None, l1 and l2 and also modifying Inverse regularization parameter C differently as 0.0001, 0.01, 0.1, 1, 10 and 100.For the Logistic Regression Modeling I have also defined the hyper parameters solver as 'saga'and max\_iter as 100.

I have also used Decision Tree Classifier Modeling several times by tuning its hyper parameters differently such as max\_depth as 3, 5, 8, 10,splitter as best and random,min\_samples\_split as 2, 3 and 5.

I have then also used Random Forest Classifier Modeling several times by tuning its hyper parameters differently such as random\_state as 42, max\_depth as 3, 5, 8 and 10,n\_estimators as 100,200 and 400.

I have then also used XGBoost Classifier Modeling several times by tuning its hyper parameters differently such as booster as gbtree,max\_depth as 3, 5, 8 and 10,n\_estimators as 100,200 and 400,random\_state as 42,objective as binary: logistic.

Then I determine the best solution possible from each of the six different types of Modeling Algorithms namely Logistic Regression modeling with penalty as Null, Logistic Regression modeling with penalty as L1, Logistic Regression modeling with penalty as L2, Decision Tree Classifier Modeling, Random Forest Classifier Modeling and XGBoost Classifier Modeling.

I have used 3 different evaluation Metrics namely classification accuracy (ACC), quadratic probability score (QPS) ,Matthew's Correlation Coefficient (MCC) and Area under the ROC curve to determine the best Prediction Models.

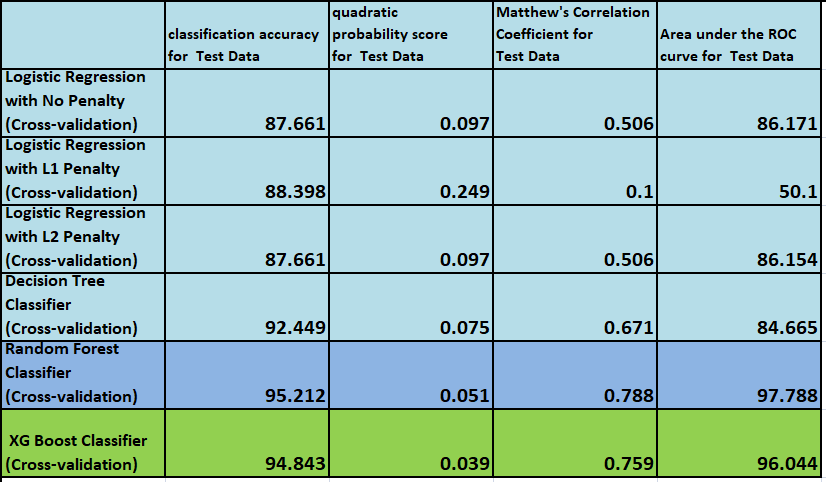
Finally I am able to determine the Modeling Algorithms giving the most accurate predictions about Target variable namely Regime to predict a period as Bullish (0) or Bearish i.e. Recession period (1).

Regime Prediction Based on Regime\_index1:

I have used Asset classes as Independent Variables.

I have used Regime\_index1 as Target Variable to predict a period as Bullish (0) or Bearish i.e. Recession period (1).based on the returns of all asset classes as Independent variables.

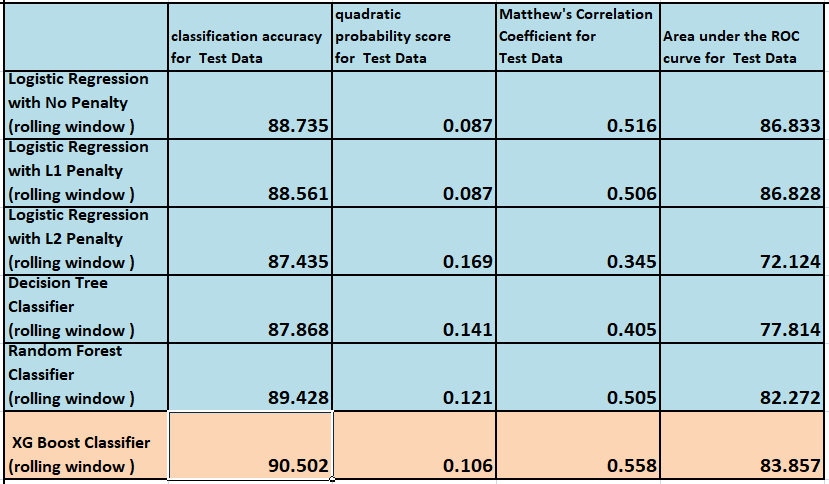
My Leader Board with Cross-Validation gives me the following results:



Leader Board-comparison of Metrics for Predictions on Regime\_index1 (Bullish or Bearish) against Asset classes by different prediction Models using Cross-Validation Technique

Clearly it can be observed that Accuracy for Test data is highest in case of Random Forest Classifier using Cross-Validation Technique. Quadratic probability score is also comparatively lesser. Matthew’s Correlation coefficient for test data is also significantly higher than other Machine Learning Models whereas Area under the ROC curve for the test data is the Highest. So we can use Random Forest Classifier Model to predict Regime and Recession Zones against the Asset Classes used as Independent variables.XG Boost classifier is coming as the next closer Model to Random Forest Classifier in terms of Metrics Comparisons.

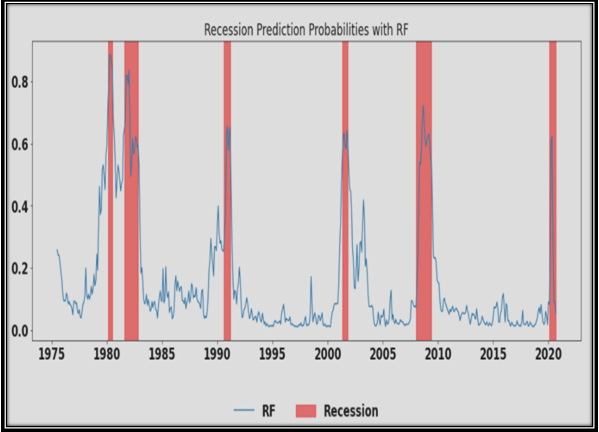
My Leader Board with Rolling-Window gives me the following results:



Leader Board-comparison of Metrics for Predictions on Regime\_index1 (Bullish or Bearish) against Asset classes by different prediction Models using Rolling-Window Technique

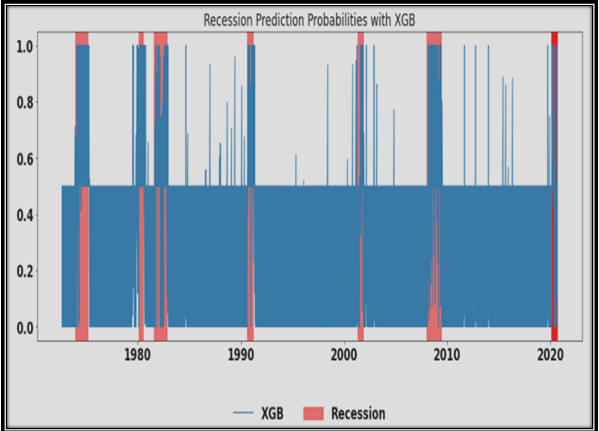
Clearly it can be observed that Accuracy for Test data is highest in case of XG Boost Classifier using Rolling Window Technique. Quadratic probability score is also comparatively lesser. Matthew’s Correlation coefficient for test data is also significantly higher than other Machine Learning Models whereas Area under the ROC curve for the test data is the Highest. So we can use XG Boost Classifier using Rolling Window Technique to predict Regime and Recession Zones against the Asset Classes used as Independent variables.

Based on Random Forest Classifier using Cross-Validation Technique, following is the predictions for Regime\_index1 (Bullish and Bearish Regimes and Recession Zones) against the Asset Classes used as Independent variables.



The Insight drawn would be that since the Bearish Regimes prediction probability crosses 0.5 during periods in 1980s, short period in and around 1982-1983, around 1990, around 2001, around 2010 and around 2020.So investor need to not only avoid making investments during these periods but should also consider the options of withdrawing riskier investment made during Bullish regimes to evade Monetary losses.

Based on XG Boost Classifier using Rolling Window Technique, following is the predictions for Regime\_index1 (Bullish and Bearish Regimes and Recession Zones) against the Asset Classes used as Independent variables.



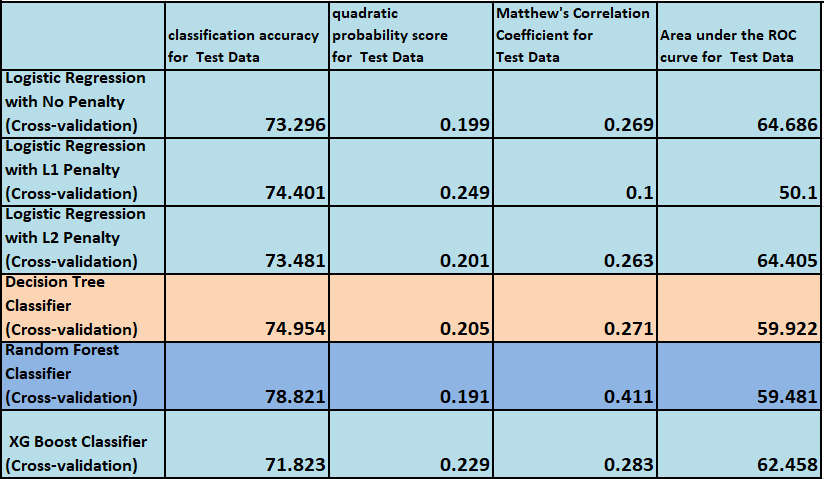
The Insight drawn would be that since the Bearish Regimes prediction probability crosses 0.5 during periods in 1980s, short period in and around 1982-1983, around 1990, around 2001, around 2010 and around 2020.So investor need to not only avoid making investments during these periods but should also consider the options of withdrawing riskier investment made during Bullish regimes to evade Monetary losses.

Regime Prediction Based on Regime\_index2:

I have used Asset classes as Independent Variables.

I have used Regime\_index2 as Target Variable to predict a period as Bullish (0) or Bearish i.e. Recession period (1).based on the returns of all asset classes as Independent variables.

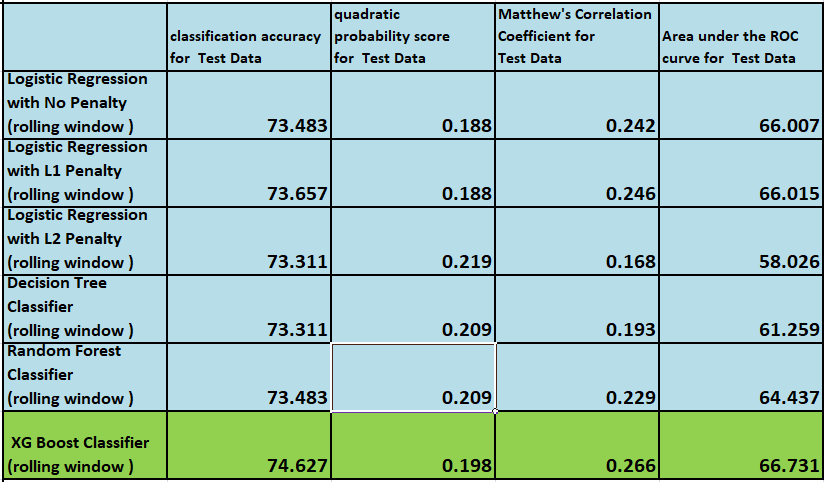
My Leader Board with Cross-Validation gives me the following results:



Leader Board-comparison of Metrics for Predictions on Regime\_index2 (Bullish or Bearish) against Asset classes by different prediction Models using Cross-Validation Technique

Clearly it can be observed that Accuracy for Test data is highest in case of Random Forest Classifier using Cross-Validation Technique. Quadratic probability score is the least. Matthew’s Correlation coefficient for test data is the highest compared to other Machine Learning Models whereas Area under the ROC curve for the test data is fairly acceptable. So we can use Random Forest Classifier Model to predict Regime and Recession Zones against the Asset Classes used as Independent variables.XG Boost classifier is coming as the next closer Model to Random Forest Classifier in terms of Metrics Comparisons.

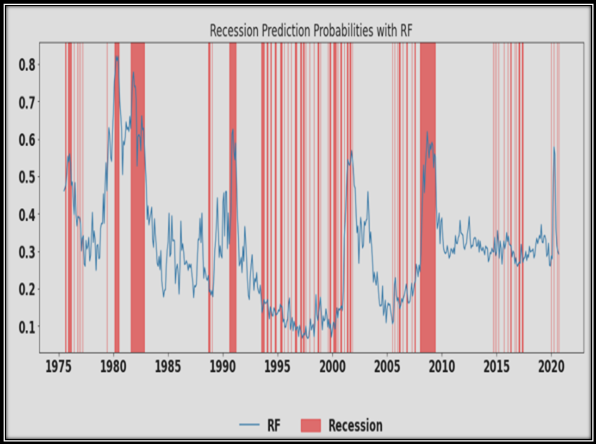
My Leader Board with Rolling-Window gives me the following results:



Leader Board-comparison of Metrics for Predictions on Regime\_index2 (Bullish or Bearish) against Asset classes by different prediction Models using Rolling-Window Technique

Clearly it can be observed that Accuracy for Test data is highest in case of XG Boost Classifier using Rolling Window Technique. Quadratic probability score is also comparatively lesser. Matthew’s Correlation coefficient for test data is the highest compared to other Machine Learning Models whereas Area under the ROC curve for the test data is the Highest. So we can use XG Boost Classifier using Rolling Window Technique to predict Regime and Recession Zones against the Asset Classes used as Independent variables.

Based on Random Forest Classifier using Cross-Validation Technique, following is the predictions for Regime\_index2 (Bullish and Bearish Regimes and Recession Zones) against the Asset Classes used as Independent variables.

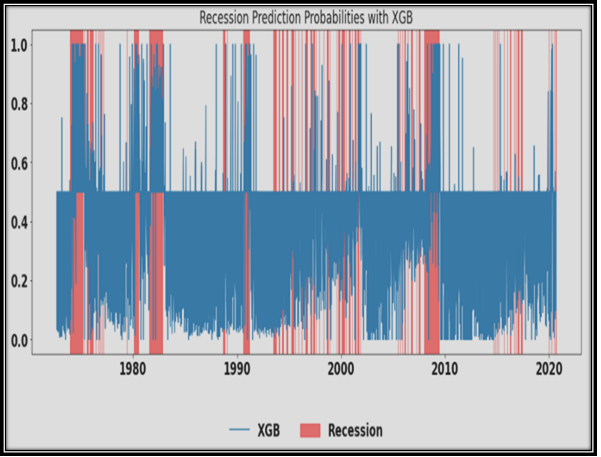


The Insight drawn would be that since the Bearish Regimes prediction probability crosses 0.5 during periods in 1975,in 1980s, short period in and around 1982-1983, lean recession wave around 1988,periods around 1990, lean recession waves between 1993-2001 and 2006-2009, periods around 2010, lean recession waves between 2015-2017 and around 2020.

So investor need to not only avoid making investments during thicker Recession periods but also consider the options of withdrawing riskier investment made during Bullish regimes to evade monetary losses.

However, investor may continue to Hold older investments during lean recession wave and resist the urge to start any newer investments.

Based on XG Boost Classifier using Rolling Window Technique, following is the predictions for Regime\_index2 (Bullish and Bearish Regimes and Recession Zones) against the Asset Classes used as Independent variables.



The Insight drawn would be that since the Bearish Regimes prediction probability crosses 0.5 during periods in 1975,in 1980s, short period in and around 1982-1983, lean recession wave around 1988,periods around 1990, lean recession waves between 1993-2001 and 2006-2009, periods around 2010, lean recession waves between 2015-2017 and around 2020.

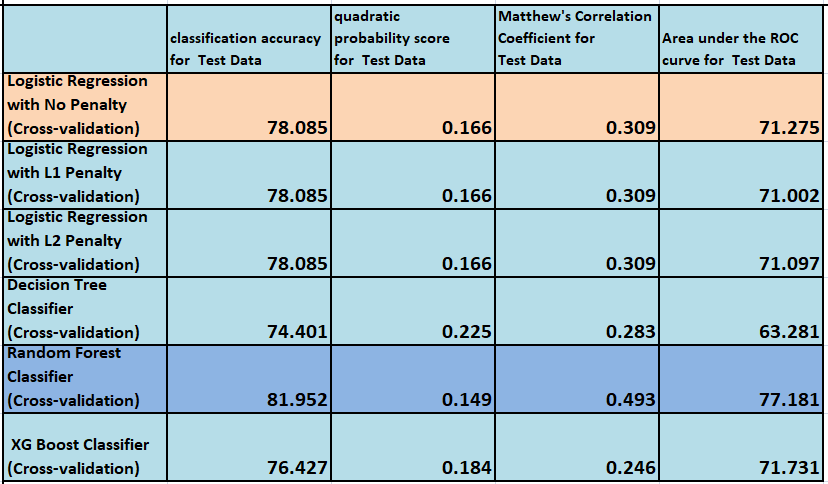
So investor need to not only avoid making investments during thicker Recession periods but also consider the options of withdrawing riskier investment made during Bullish regimes to evade monetary losses. However, investor may continue to Hold older investments during lean recession wave and resist the urge to start any newer investments.

Regime Prediction Based on Regime\_index3:

I have used Asset classes as Independent Variables.

I have used Regime\_index3 as Target Variable to predict a period as Bullish (0) or Bearish i.e. Recession period (1).based on the returns of all asset classes as Independent variables.

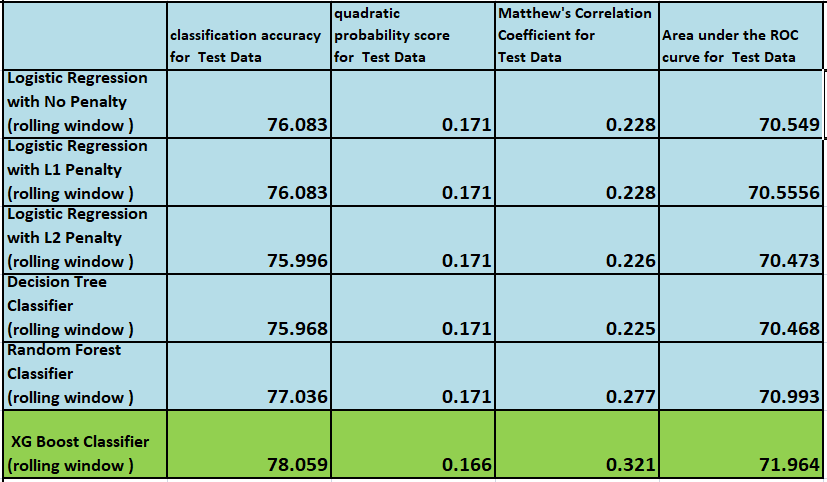
My Leader Board with Cross-Validation gives me the following results:



Leader Board-comparison of Metrics for Predictions on Regime\_index3 (Bullish or Bearish) against Asset classes by different prediction Models using Cross-Validation Technique

Clearly it can be observed that Accuracy for Test data is highest in case of Random Forest Classifier using Cross-Validation Technique. Quadratic probability score is the least. Matthew’s Correlation coefficient for test data is the highest compared to other Machine Learning Models whereas Area under the ROC curve for the test data is the highest. So we can use Random Forest Classifier Model to predict Regime and Recession Zones against the Asset Classes used as Independent variables. Logistic Regression classifier with no penalty as hyper parameter is coming as the next closer Model to Random Forest Classifier in terms of Metrics Comparisons.

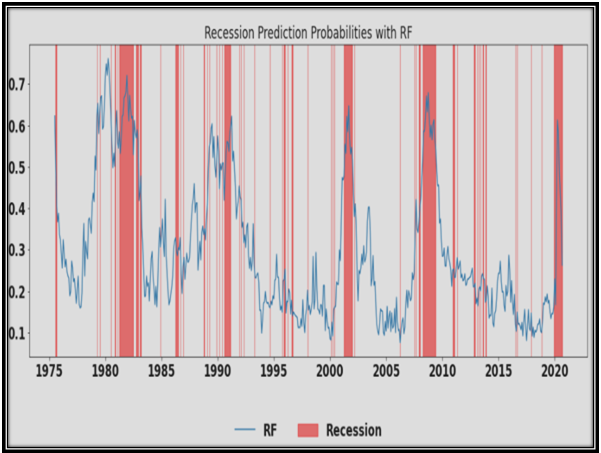
My Leader Board with Rolling-Window gives me the following results:



Leader Board-comparison of Metrics for Predictions on Regime\_index3 (Bullish or Bearish) against Asset classes by different prediction Models using Rolling-Window Technique

Clearly it can be observed that Accuracy for Test data is highest in case of XG Boost Classifier using Rolling Window Technique. Quadratic probability score is also the least. Matthew’s Correlation coefficient for test data is the highest compared to other Machine Learning Models whereas Area under the ROC curve for the test data is the Highest. So we can use XG Boost Classifier using Rolling Window Technique to predict Regime and Recession Zones against the Asset Classes used as Independent variables.

Based on Random Forest Classifier using Cross-Validation Technique, following is the predictions for Regime\_index3 (Bullish and Bearish Regimes and Recession Zones) against the Asset Classes used as Independent variables.

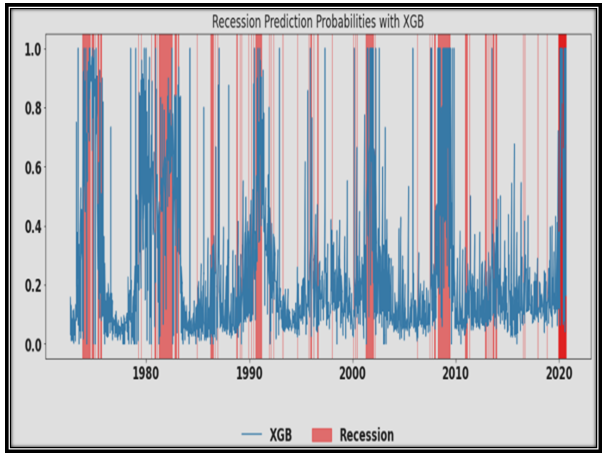


The Insight drawn would be that since the Bearish Regimes prediction probability crosses 0.5 during periods in 1975, lean recession waves in 1979,1981,short period in and around 1982-1983, periods around 1985,lean recession waves between 1988-1989,periods around 1990,lean recession waves in 1991, 1992 , 1993, 1995, 1996,1999,periods around 2001 and 2010, lean recession waves in 2012, 2014, 2015, 2017,2018,2019 and periods around 2020.

So investor need to not only avoid making investments during thicker Recession periods but also consider the options of withdrawing riskier investment made during Bullish regimes to evade monetary losses.

However, investor may continue to Hold older investments during lean recession wave and resist the urge to start any newer investments.

Based on XG Boost Classifier using Rolling Window Technique, following is the predictions for Regime\_index3 (Bullish and Bearish Regimes and Recession Zones) against the Asset Classes used as Independent variables.



The Insight drawn would be that since the Bearish Regimes prediction probability crosses 0.5 during periods in 1975,in 1980s, short period in and around 1982-1983, lean recession wave around 1988,periods around 1990, lean recession waves between 1993-2001 and 2006-2009, periods around 2010, lean recession waves between 2015-2017 and around 2020.

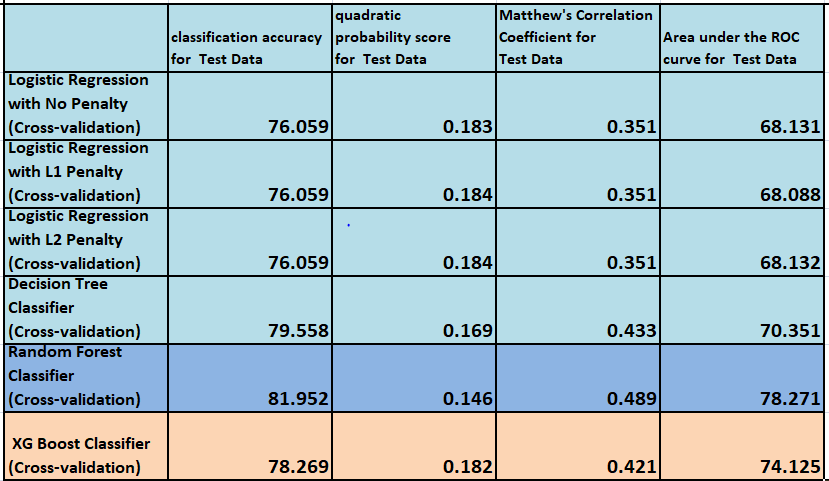
So investor need to not only avoid making investments during thicker Recession periods but also consider the options of withdrawing riskier investment made during Bullish regimes to evade monetary losses. However, investor may continue to Hold older investments during lean recession wave and resist the urge to start any newer investments.

Regime Prediction Based on Regime\_index4:

I have used Asset classes as Independent Variables.

I have used Regime\_index4 as Target Variable to predict a period as Bullish (0) or Bearish i.e. Recession period (1).based on the returns of all asset classes as Independent variables.

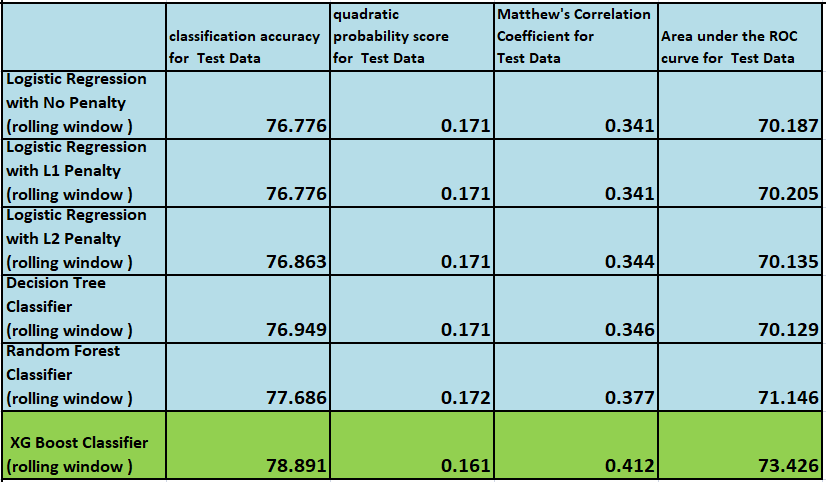
My Leader Board with Cross-Validation gives me the following results:



Leader Board-comparison of Metrics for Predictions on Regime\_index4 (Bullish or Bearish) against Asset classes by different prediction Models using Cross-Validation Technique

Clearly it can be observed that Accuracy for Test data is highest in case of Random Forest Classifier using Cross-Validation Technique. Quadratic probability score is the least. Matthew’s Correlation coefficient for test data is the highest compared to other Machine Learning Models whereas Area under the ROC curve for the test data is the highest. So we can use Random Forest Classifier Model to predict Regime and Recession Zones against the Asset Classes used as Independent variables. XG Boost classifier using Cross-Validation Technique is coming as the next closer Model to Random Forest Classifier in terms of Metrics Comparisons.

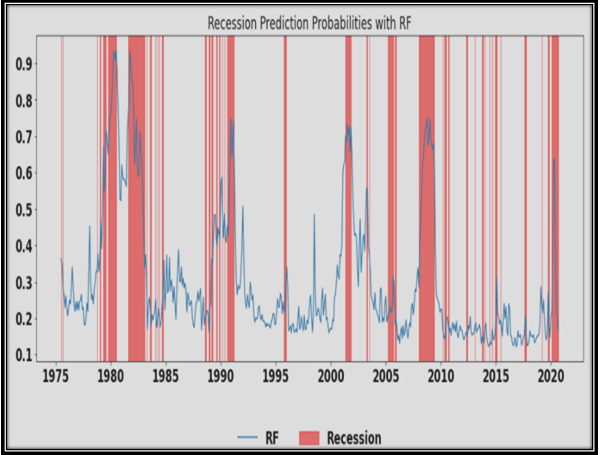
My Leader Board with Rolling-Window gives me the following results:



Leader Board-comparison of Metrics for Predictions on Regime\_index4 (Bullish or Bearish) against Asset classes by different prediction Models using Rolling-Window Technique

Clearly it can be observed that Accuracy for Test data is highest in case of XG Boost Classifier using Rolling Window Technique. Quadratic probability score is also the least. Matthew’s Correlation coefficient for test data is the highest compared to other Machine Learning Models whereas Area under the ROC curve for the test data is the Highest. So we can use XG Boost Classifier using Rolling Window Technique to predict Regime and Recession Zones against the Asset Classes used as Independent variables.

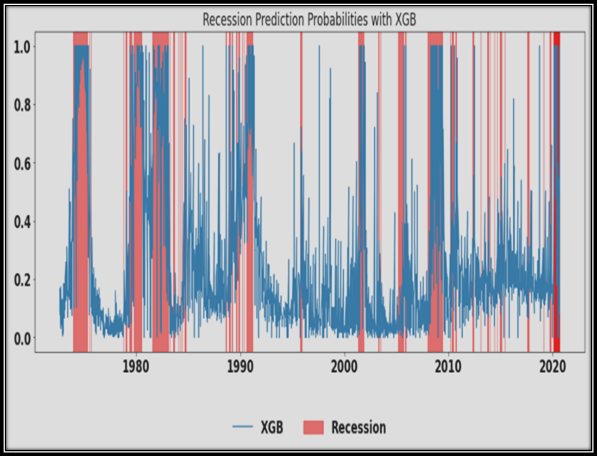
Based on Random Forest Classifier using Cross-Validation Technique, following is the predictions for Regime\_index4 (Bullish and Bearish Regimes and Recession Zones) against the Asset Classes used as Independent variables.



The Insight drawn would be that since the Bearish Regimes prediction probability crosses 0.5 during periods in 1980s, short period in and around 1982-1983, lean recession wave in 1984,1985,1989,periods around 1990,1995,2000,lean recession waves in 2002, periods around 2005,2010, lean recession waves in 2011,2013,2014,2015,2017 and around 2020.

So investor need to not only avoid making investments during thicker Recession periods but also consider the options of withdrawing riskier investment made during Bullish regimes to evade monetary losses. However, investor may continue to Hold older investments during lean recession wave and resist the urge to start any newer investments.

Based on XG Boost Classifier using Rolling Window Technique, following is the predictions for Regime\_index4 (Bullish and Bearish Regimes and Recession Zones) against the Asset Classes used as Independent variables.



The Insight drawn would be that since the Bearish Regimes prediction probability crosses 0.5 during periods in 1975, 1980,1982,2000,2010 and around 2020. Lean recession waves could also be observed during certain periods in the period ranging from 1975-2020.

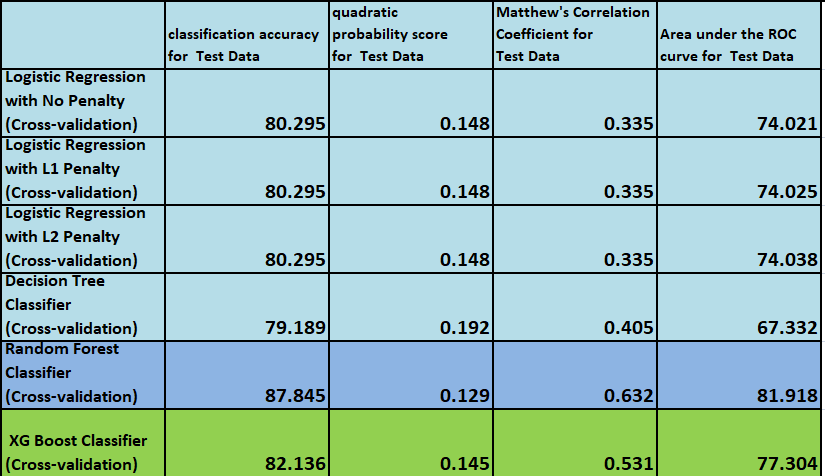
So investor need to not only avoid making investments during thicker Recession periods but also consider the options of withdrawing riskier investment made during Bullish regimes to evade monetary losses. However, investor may continue to Hold older investments during lean recession wave and resist the urge to start any newer investments.

Regime Prediction Based on Regime\_index5:

I have used Asset classes as Independent Variables.

I have used Regime\_index5 as Target Variable to predict a period as Bullish (0) or Bearish i.e. Recession period (1).based on the returns of all asset classes as Independent variables.

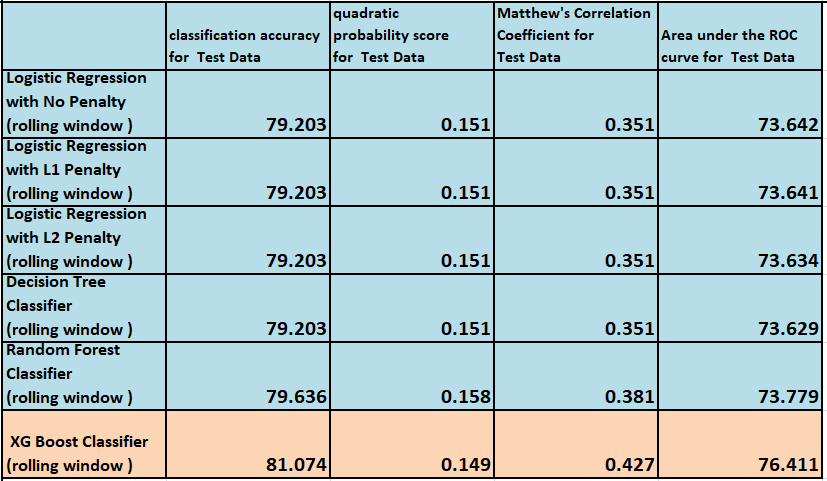
My Leader Board with Cross-Validation gives me the following results:



Leader Board-comparison of Metrics for Predictions on Regime\_index5 (Bullish or Bearish) against Asset classes by different prediction Models using Cross-Validation Technique

Clearly it can be observed that Accuracy for Test data is highest in case of Random Forest Classifier using Cross-Validation Technique. Quadratic probability score is the least. Matthew’s Correlation coefficient for test data is the highest compared to other Machine Learning Models whereas Area under the ROC curve for the test data is the highest. So we can use Random Forest Classifier Model to predict Regime and Recession Zones against the Asset Classes used as Independent variables. XG Boost classifier using Cross-Validation Technique is coming as the next closer Model to Random Forest Classifier in terms of Metrics Comparisons.

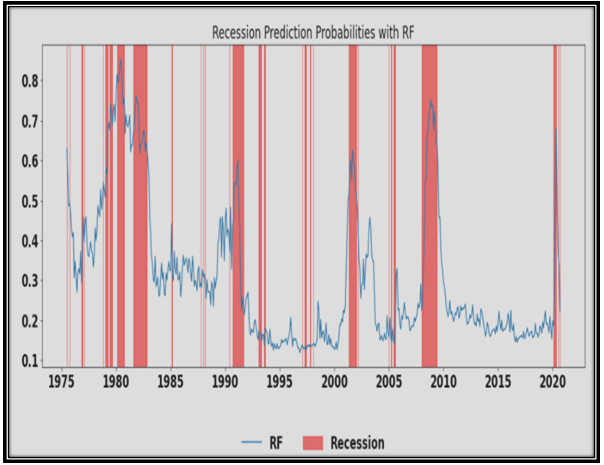
My Leader Board with Rolling-Window gives me the following results:



Leader Board-comparison of Metrics for Predictions on Regime\_index5 (Bullish or Bearish) against Asset classes by different prediction Models using Rolling-Window Technique

Clearly it can be observed that Accuracy for Test data is highest in case of XG Boost Classifier using Rolling Window Technique. Quadratic probability score is also the least. Matthew’s Correlation coefficient for test data is the highest compared to other Machine Learning Models whereas Area under the ROC curve for the test data is the Highest. So we can use XG Boost Classifier using Rolling Window Technique to predict Regime and Recession Zones against the Asset Classes used as Independent variables.

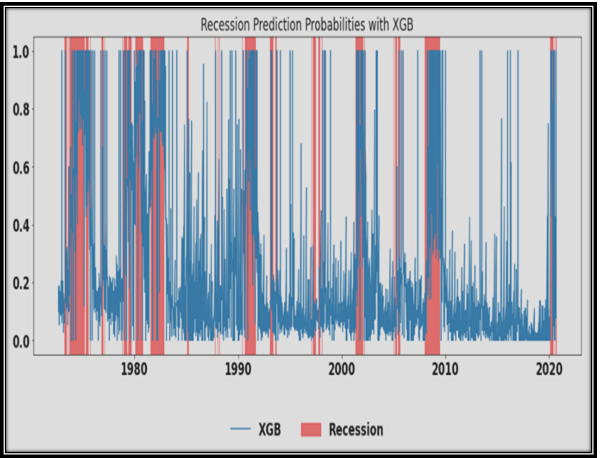
Based on Random Forest Classifier using Cross-Validation Technique, following is the predictions for Regime\_index5 (Bullish and Bearish Regimes and Recession Zones) against the Asset Classes used as Independent variables.



The Insight drawn would be that since the Bearish Regimes prediction probability crosses 0.5 during periods in 1980s, short period in and around 1982-1983, lean recession wave in 1985, 1988,periods around 1990, lean recession waves in 1992,1997, periods around 2000,2010, lean recession waves in 2005 and around 2020.

So investor need to not only avoid making investments during thicker Recession periods but also consider the options of withdrawing riskier investment made during Bullish regimes to evade monetary losses. However, investor may continue to Hold older investments during lean recession wave and resist the urge to start any newer investments.

Based on XG Boost Classifier using Rolling Window Technique, following is the predictions for Regime\_index5 (Bullish and Bearish Regimes and Recession Zones) against the Asset Classes used as Independent variables.



The Insight drawn would be that since the Bearish Regimes prediction probability crosses 0.5 during periods in 1975, 1980,1982,1990,2000 and around 2010. Lean recession waves could also be observed during certain periods in the period ranging from 1975-2020.

So investor need to not only avoid making investments during thicker Recession periods but also consider the options of withdrawing riskier investment made during Bullish regimes to evade monetary losses. However, investor may continue to Hold older investments during lean recession wave and resist the urge to start any newer investments.