



The Good, the Bad, and the Ugly

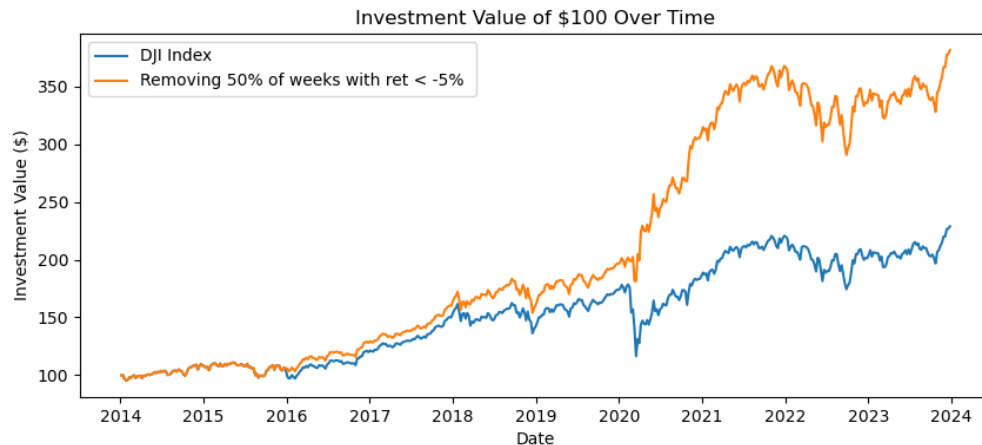
MFE 431 – FINAL PRESENTATION

Aditya Singh, Dhruv Shah, Samin Martinrazm, Vinamra Rai, Yasaswy Aluri

UCLA Anderson



Introduction



Key Objective: Identifying periods of high risk (Imminent Danger) and low risk in DJIA.

Idea: Utilize ML techniques to classify market conditions and predict extreme negative returns

Expected Outcome: Improved market timing and risk management through accurate forecasting of significant market downturns.

Intuition Behind Strategy

Decoding Market Regimes:

- **Challenge:** Traditional indicators like moving averages and volatility are unreliable alone.
- **Solution:** Combine multiple weak signals using ML techniques : Classifiers/Decision Trees.

Machine Learning Magic:

- **ML Predictor:** Transforms weak signals into a powerful predictor by capturing complex interactions.
- **Three Regimes:**
 - **Business as Usual (BAU):** Steady market.
 - **Goldilocks:** Smooth sailing with positive returns.
 - **Imminent Danger:** High risk of market drops.
- **Strategic Moves:**
 - **Avoiding Pitfalls:** Reduce exposure during Imminent Danger to dodge downturns.
 - **Seizing Opportunities:** Boost exposure in Goldilocks periods for maximum gains.

Impact: Data-Driven Decisions for smarter, dynamic portfolio management.

Data Sources

- **Bloomberg**

- Usage: The Raven sentiment index, RSI for 5-day and 14-day data

- **Chicago Board Options Exchange (CBOE)**

- Usage: Put – Call Ratio of S&P 500 over the past 5-days

- **Yahoo Finance**

- Usage: Historical Returns data starting 1st January, 2000. Train Dataset – 2000 to 2017

- **Federal Reserve (Fed)**

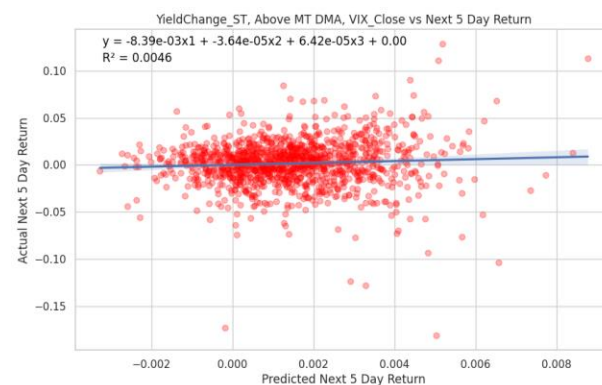
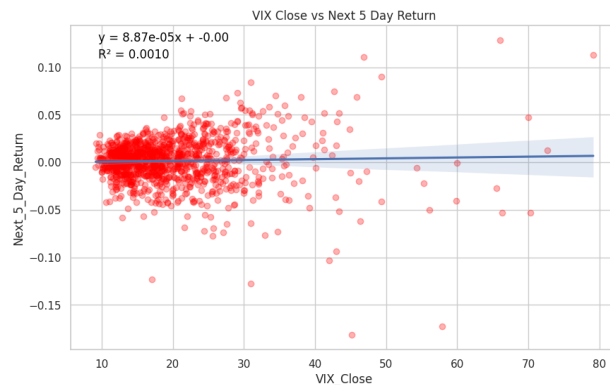
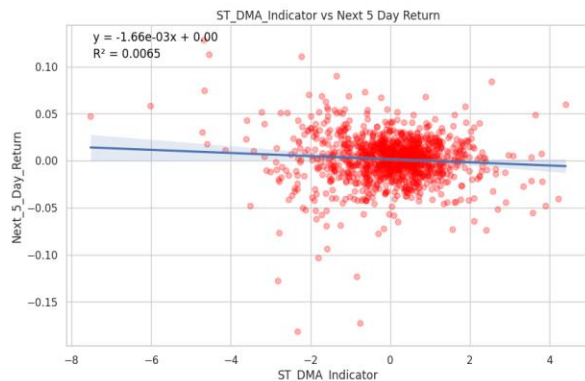
- Usage: News Sentiment Index created from various news sources using Federal Reserve NLP Model

Variables Description

Variable	Definition
RSI5	Relative Strength Index (5-day)
RSI14	Relative Strength Index (14-day)
MT Historical Return	DJI return, past 20-day
Return Difference	DJI return past five day - DJI return past 20 day
ST_DMA_Indicator	Percent difference of DJI index from its 5-day moving average
MT_DMA_Indicator	Percent difference of DJI index from its 20-day moving average
LT_DMA_Indicator	Percent difference of DJI index from its 50-day moving average
YieldChange_MT	Change in 10-year yield over past 20-days
YieldChange_ST	Change in 10-year yield over past 5-days
Varex_eig1	Percent of cross-sectional variance explained by first principal component of correlation matrix
Varex_eig_top5	Percent of cross-sectional variance explained by first five principal components of correlation matrix
Varex_5-1	Varex_eig_top5 minus Varex_eig1
Coeffs_mean	The average of the pairwise stock correlation
Coeffs_std	The standard deviation of the pairwise stock correlation
Above ST DMA	The percentage of stocks that are currently above their 10-day moving average
Above MT DMA	The percentage of stocks that are currently above their 20-day moving average
Raven Sentiment Index	Sentiment index for all news related to DJI & SP500 published by Ravenpack
Fed Daily News Sentiment	Sentiment index using 24 major news articles published by Federal Reserve
Put Call Ratio	Indicator to gauge investor sentiment measuring the volume of put options traded relative to call options
VIX	Measure of the stock market's expectation of volatility based on S&P 500 index options
DJI Volume	Total number of shares traded during a specific period (usually a day) for all the constituent stocks of the DJIA

Empirical Analysis : Exploratory Regressions

- **Individual Variable Limitations:** The scatter plots clearly demonstrate that individual variables are not reliable predictors of short-term index returns.
- **Combination of Variables:** Even the combination of three variables does not significantly enhance predictive accuracy.
- **Robust Prediction Model:** To achieve a robust prediction model for market downturns, we utilize an aggregation of 18 weakly correlated variables.
- **Comprehensive Approach:** This approach harnesses collective insights, offering a much stronger predictive capability for identifying potential market crashes.

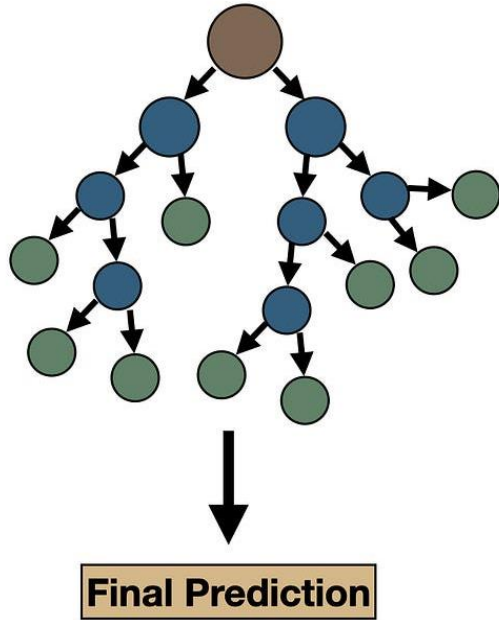


Empirical Analysis : Introducing ML Models

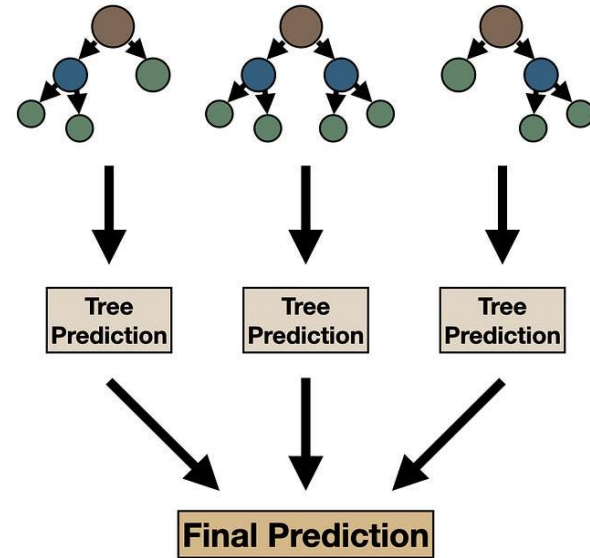
- **Individual Variable Limitations:** Individual variables may lack predictive power on their own.
- **Combined Variable Approach:** Using the variables collectively to train machine learning models.
- **ML Models Utilized:**
 - Decision Tree: Simple, interpretable models for classification and regression tasks
 - KNN Classifier: Non-parametric method used for classification and regression.
 - Gradient Boosting: Powerful ensemble technique that builds models sequentially to correct the predecessors' errors.
 - XG Boost: An optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable.
- **Enhanced Predictive Power:** This comprehensive approach significantly enhances the predictive capabilities of the models.

Comparison Between Models

Decision Tree

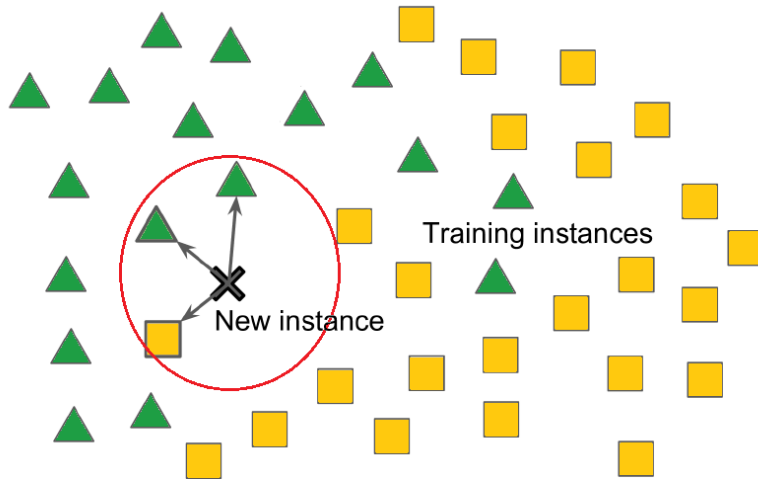


Random Forest

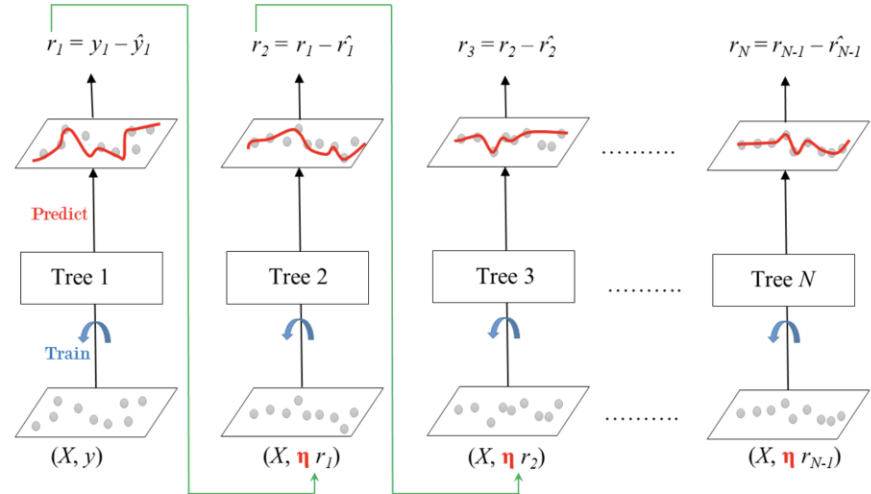


Comparison Between Models

KNN Classifier

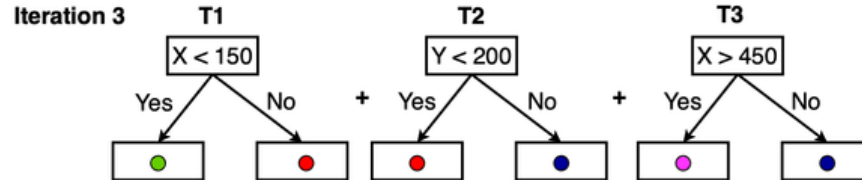
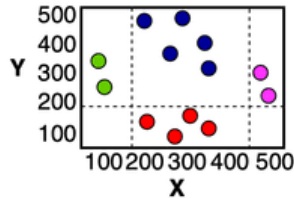
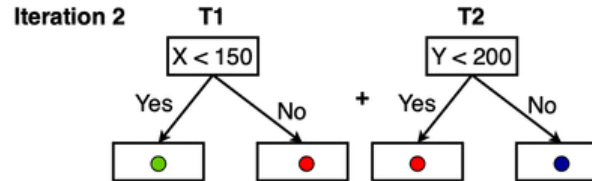
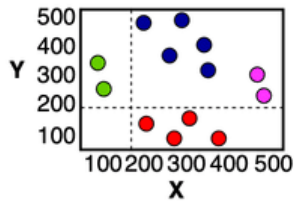
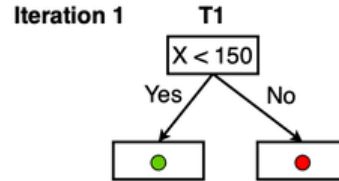
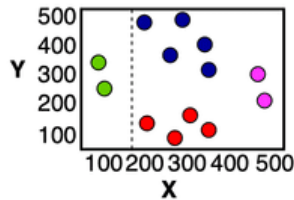


Gradient Boosting Classifier



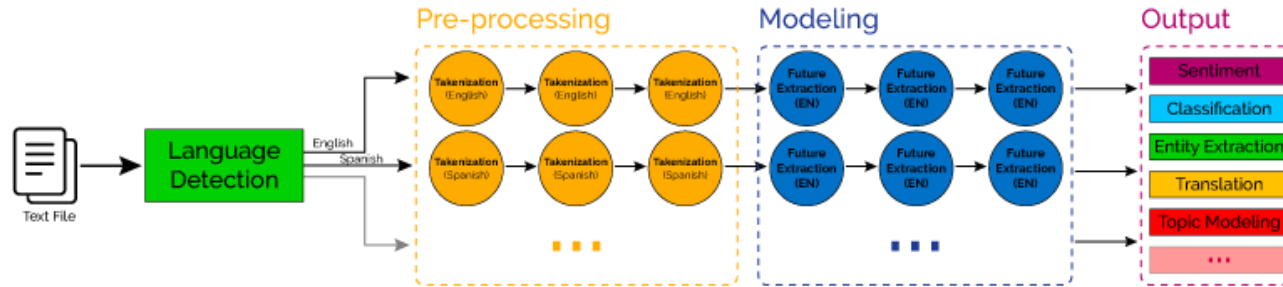
Comparison Between Models

XG Boost



Natural Language Processing

Classical NLP



- Sentiment Index Examples – Fed and Ravenpack

Investment Strategy

Step 1

Classify our predicted returns in of the three categories: Goldilocks, Business as usual and Imminent Danger.

Step 2

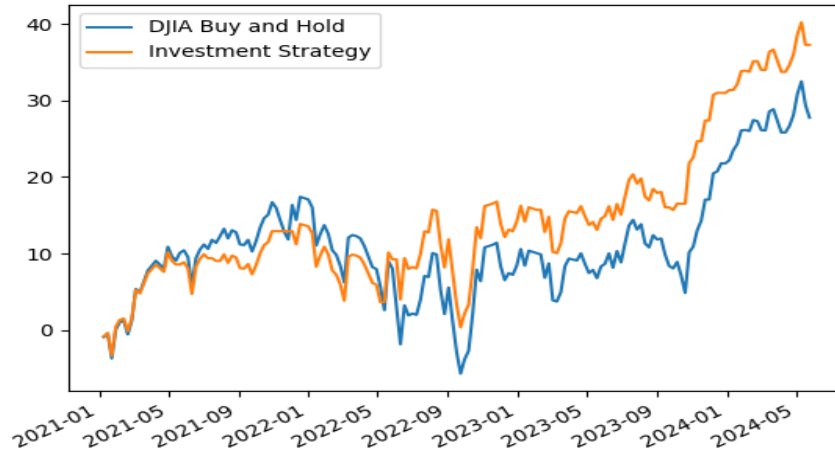
- Goldilocks Period: Hold 1 unit of the Dow Jones Industrial Average (DJI) when the model forecasts a Goldilocks period.
- Imminent Danger: Hold 0 units of DJI when the model forecasts Imminent Danger.
- Other Times: During other periods, maintain 0.9 units of DJI.

Step 3

We plot our cumulative returns predicted by the model vs the actual predicted returns along with the metrics of the particular Machine Learning model used.

Alternatively, this model can also be used to opportunistically hedge a long portfolio to reduce the equity risk. When the model forecasts Imminent Danger, the hedging ratio is the highest. On the other hand, when the model forecasts Goldilocks zone, the hedge ratio is the lowest.

Results: K-Nearest Neighbors



A	Goldilocks Period
B	Imminent Danger
C	Other times/Business as Usual

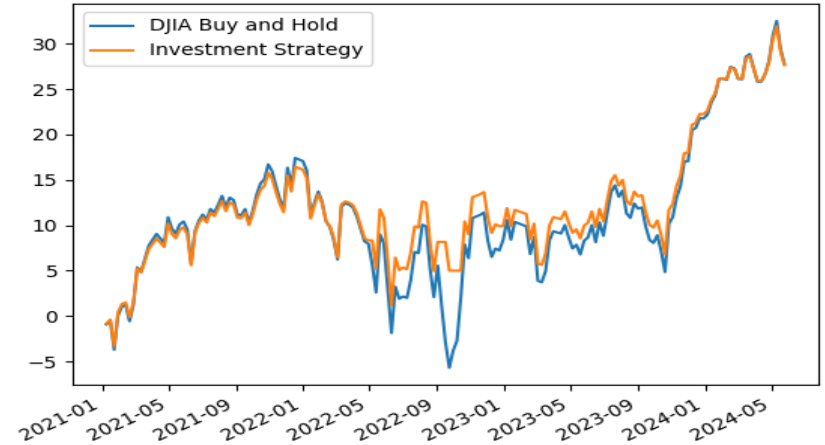
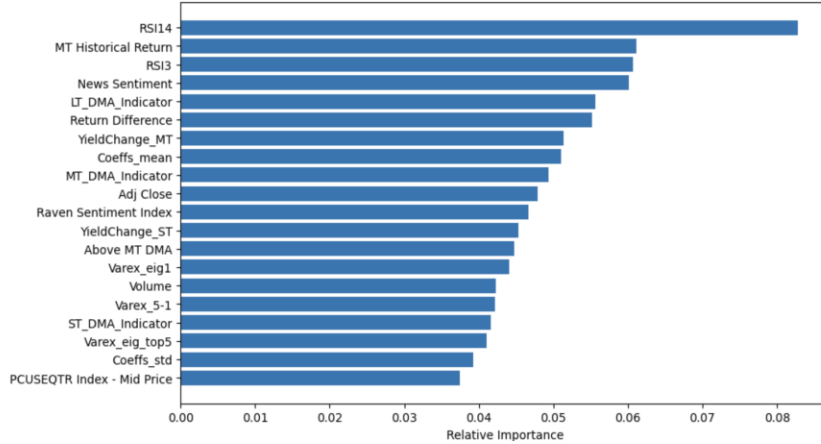
Accuracy = 60.23%				
	Precision	Recall	f1-Score	Support
A	0.23	0.26	0.24	23
B	0.09	0.12	0.10	17
C	0.77	0.73	0.75	131

Predicted \ Actual	A	B	C
A	6	1	16
B	3	2	12
C	17	19	95

Performance Stats	DJI Buy and Hold	Investment Strategy
Ann. Returns	8.52%	10.48%
Ann. Volatility	14.70%	13.02%
Sharpe Ratio	0.58	0.81

Results: Random Forest

Feature Importances for Random Forest

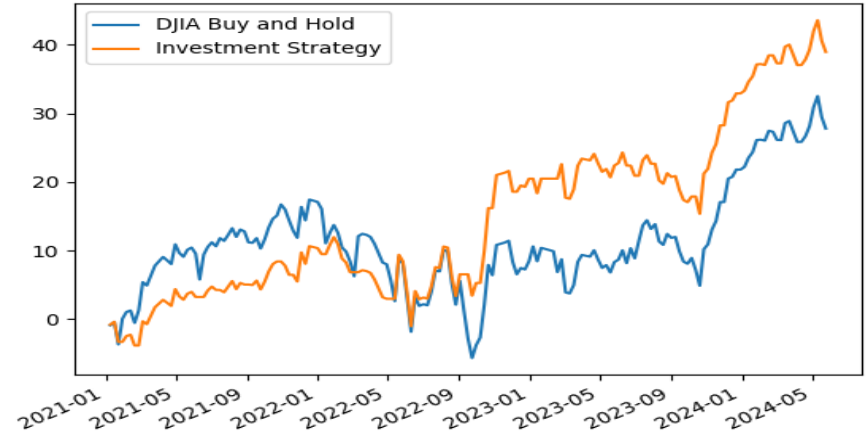
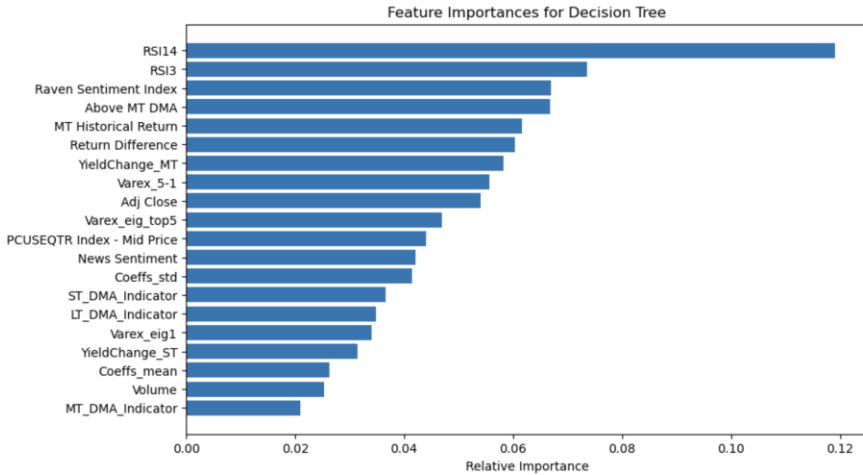


Accuracy = 71.93%				
	Precision	Recall	f1-Score	Support
A	0.19	0.17	0.18	23
B	0.40	0.12	0.18	17
C	0.81	0.89	0.85	131

Predicted	A	B	C
Actual	A	B	C
A	4	0	19
B	6	2	9
C	11	3	117

Performance Stats	DJI Buy and Hold	Investment Strategy
Ann. Returns	8.52%	8.27%
Ann. Volatility	14.70%	12.99%
Sharpe Ratio	0.58	0.64

Results: Decision Tree

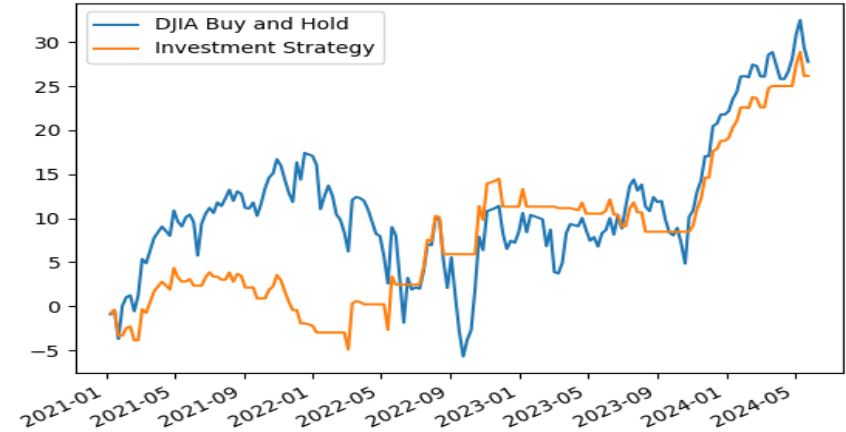
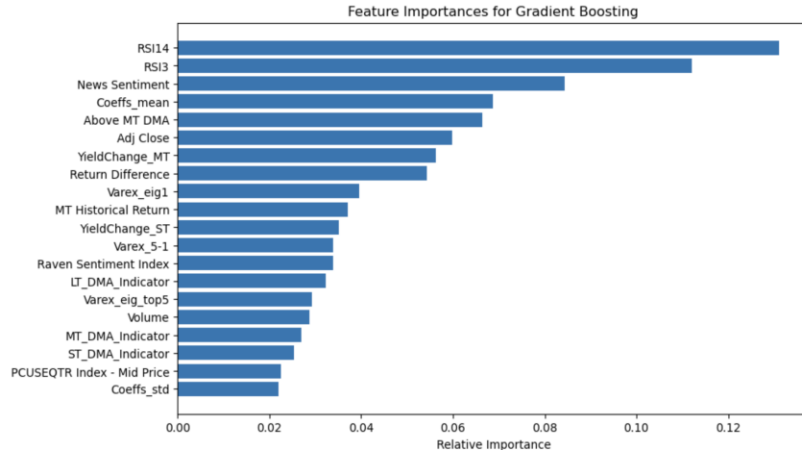


Accuracy = 59.06%				
	Precision	Recall	f1-Score	Support
A	0.26	0.30	0.28	23
B	0.14	0.29	0.19	17
C	0.82	0.68	0.74	131

Predicted	A	B	C
Actual	A	B	C
A	7	5	11
B	3	5	9
C	17	25	89

Performance Stats	DJI Buy and Hold	Investment Strategy
Ann. Returns	8.52%	9.92%
Ann. Volatility	14.70%	12.07%
Sharpe Ratio	0.58	0.82

Results: Gradient Boosting



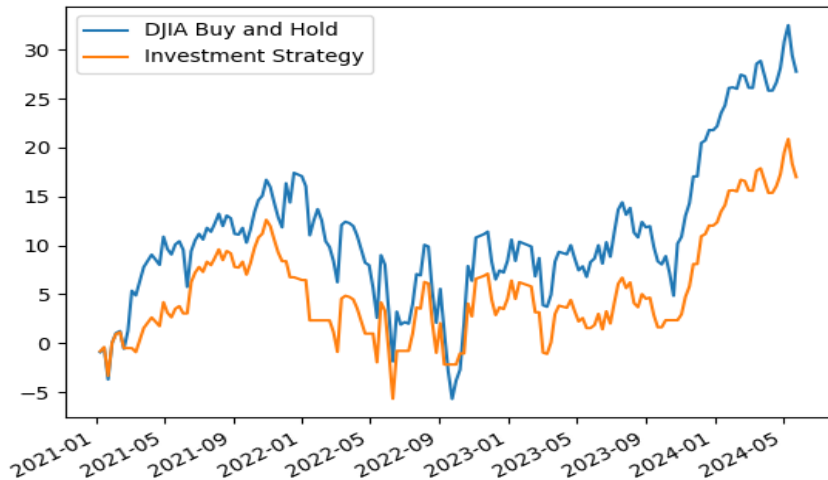
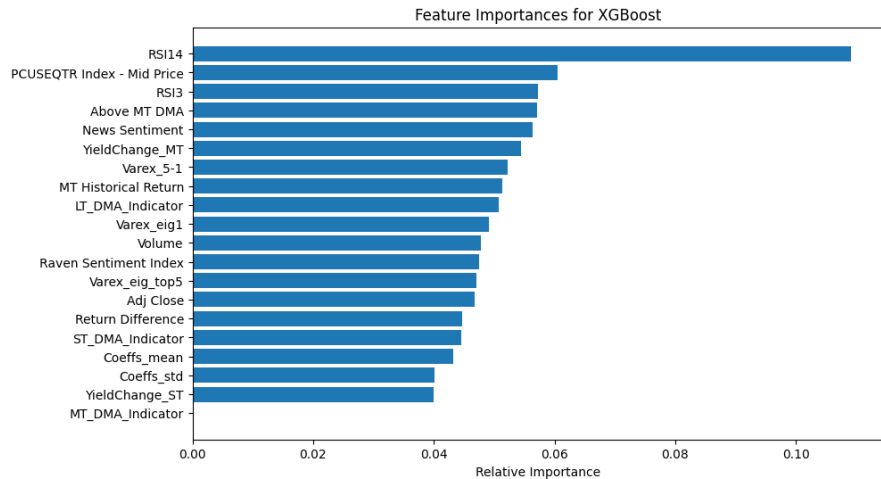
Accuracy = 53.22%

	Precision	Recall	f1-Score	Support
A	0.50	0.22	0.30	23
B	0.16	0.71	0.26	17
C	0.88	0.56	0.69	131

Predicted \ Actual	A	B	C
A	5	12	6
B	1	12	4
C	4	53	74

Performance Stats	DJI Buy and Hold	Investment Strategy
Ann. Returns	8.52%	8.16%
Ann. Volatility	14.70%	9.27%
Sharpe Ratio	0.58	0.88

Results: XGBoost



Accuracy = 66.67%				
	Precision	Recall	f1-Score	Support
A	0.23	0.13	0.17	23
B	0.12	0.18	0.14	17
C	0.82	0.82	0.82	131

Predicted	A	B	C
Actual	A	B	C
A	3	5	15
B	5	3	9
C	5	18	108

Performance Stats	DJI Buy and Hold	Investment Strategy
Ann. Returns	8.52%	5.49%
Ann. Volatility	14.70%	12.06%
Sharpe Ratio	0.58	0.46

Concerns with the Strategy

- **Prolonged Directional Movement in Market** : Utility of a regime prediction model is limited in a one-way upside market
- **Transaction Costs**: We haven't taken into account the slippage/transaction costs associated with hedging /closing positions
- **Structural Breaks in Market** : Model relies entirely on historical train data and can give wrong results when economic regime changes
- **External Factors** : Rare events are often not predictable by historical data-based models.
- **Data Dependency and Complexity** : Hard to explain models to general public. Garbage in garbage out could be an issue

Future Work

- **Additional Variables** : Adding more fundamental/macro variables can improve accuracy
- **Ensemble Learning Methods**: Investigate the use of ensemble learning techniques that combine multiple models to potentially improve prediction accuracy and reduce overfitting
- **Explore other segments / Geographies** : Look out for inefficient markets
 - Example from India: Drawdown Protection in Nifty 50 (Jan 2007 – Mar 2023)
 - Results using Microsoft's LightGBM model

		Actual		
		Goldilocks	BAU	Imminent Danger
Predicted	Goldilocks	40	24	0
	BAU	37	652	8
	Imminent Danger	1	10	27

- Source: Kotak Institutional Equities

