Predicting Regimes in the Indian Market with Macroeconomic Data and Machine Learning

Sonam Srivastava[1] and Akashdeep Bhateja[2]

[1]Wright Research
[2]Wright Research

November 2023

Abstract

This research paper delves into the predictive capabilities of macroeconomic indicators in forecasting market returns and identifying distinct market regimes in the Indian financial landscape. Utilizing a comprehensive dataset that includes all major macroeconomic signals in India, the study first undertakes meticulous data cleaning and feature selection processes, narrowing down to the top 10 macroeconomic indicators. These indicators are then employed in advanced machine learning models, including XGBoost, RandomForest, and Decision Tree, to predict market regimes. A novel aspect of our methodology is the use of backcasting techniques, particularly with the XGBoost algorithm, to impute missing data for pivotal indicators, thereby enriching our analysis. Furthermore, the study innovatively classifies market conditions into 'Normal' and 'Crash' regimes, offering a nuanced understanding of market dynamics. The findings of this research are significant, demonstrating that certain macroeconomic indicators, notably Remittances, Consumer Price Index, and Gross Domestic Product growth rate, are highly predictive of market returns in the Indian context. This study not only contributes to the academic discourse on financial forecasting but also provides practical insights for investors and policymakers in navigating the complexities of the Indian financial markets.

1 Introduction

The intricate relationship between macroeconomic indicators and financial markets has long been a subject of intense study and debate. In emerging economies like India, where the market dynamics are particularly volatile and influenced by a myriad of factors, understanding this relationship is crucial for investors, policymakers, and economists. The regime models are also used by systematic trading firms to modify their risk targets and asset allocation. This research paper aims to shed light on the predictive power of macroeconomic indicators in forecasting market returns and identifying distinct market regimes in the Indian financial landscape.

The Indian economy, characterized by its rapid growth, diverse industrial sectors, and a burgeoning middle class, presents a unique set of challenges and opportunities for market analysis. The stock market in India, represented prominently by indices like the BSE Sensex, is influenced by both global economic trends and domestic macroeconomic factors. However, the extent to which these macroeconomic indicators can predict market returns and classify market regimes has not been thoroughly explored, particularly with the application of advanced machine learning techniques.

In this study, we embark on a comprehensive analysis, employing a range of machine learning models including XGBoost, RandomForest, and Decision Tree algorithms. Our approach is twofold: firstly, we aim to predict market returns using selected macroeconomic indicators, and secondly, we seek to classify the market into 'Normal' and 'Crash' regimes, providing a nuanced understanding of market behaviour. A significant innovation in our methodology is the use of backcasting techniques to address the challenge of missing data, particularly for key indicators like the Consumer Price Index (CPI) and Remittance (REM).

The motivation behind this research is not only academic but also practical. By enhancing the understanding of how macroeconomic factors influence market returns, we aim to provide investors and policymakers with tools and insights to make more informed decisions. This is particularly relevant in the context of the Indian market, where economic growth and market stability are pivotal to the country's overall development.

This paper is structured as follows: following this introduction, we review relevant literature in the field, outlining previous studies and their findings. We then detail our methodology, including data collection, cleaning, feature selection, and the specifics of the machine learning models employed. Subsequent sections present our results, discuss their implications, and conclude with recommendations for future research.

2 Regime Models - A Literature Review

The concept of regime models in financial markets is a critical area of study, particularly in the context of predicting market behaviour and returns. Regime models are frameworks that categorize market conditions into distinct states or regimes, often characterized by varying levels of volatility, market trends, and economic conditions. This section reviews the existing literature on regime models, focusing on their development, methodologies, and applications in financial markets, with a special emphasis on emerging markets like India.

2.1 Early Developments and Theoretical Foundations

The initial exploration into regime models can be traced back to the work of Hamilton [Hamilton, 1989], who introduced a regime-switching model to capture the probabilistic nature of economic recessions and expansions in the United States. This pioneering work laid the foundation for subsequent studies that sought to incorporate macroeconomic variables into regime-switching frameworks. The early models primarily utilized Markov-switching mechanisms [Markov, 1912].

2.2 Advancements in Regime Modeling

Over the years, regime models have evolved, incorporating more sophisticated statistical and machine-learning techniques. Studies by Ang and Bekaert [Ang and Bekaert, 2002] and Pelletier [Pelletier, 2006] expanded the regime-switching models by integrating international market factors and volatility dynamics. The integration of machine learning techniques, particularly classification algorithms like Decision Trees, Random Forest, and XGBoost, has further enhanced the predictive power of these models [Breiman, 2001, Chen and Guestrin, 2016].

2.3 Regime Models in Emerging Markets

The application of regime models in emerging markets presents unique challenges and opportunities. Patnaik and Shah [Patnaik and Shah, 2010] highlighted the influence of regulatory changes and economic reforms on market behaviour in the Indian financial market. Kumar and Thenmozhi [Kumar and Thenmozhi, 2006] emphasized the role of domestic macroeconomic indicators, such as inflation rates and GDP growth, in defining market regimes in India.

2.4 Recent Trends and Machine Learning Integration

Recent literature indicates a growing trend in integrating machine learning techniques with traditional econometric models [Lopez, 2019]. The use of algorithms like XGBoost and RandomForest in regime prediction models represents a significant leap in this field [Chen and Guestrin, 2016, Breiman, 2001].

2.5 Future Directions

The application of these models in emerging markets, particularly in India, has opened new avenues for research, especially in understanding the impact of macroeconomic indicators on market regimes. Future research could focus on the integration of real-time data analysis and the exploration of unsupervised learning techniques to further enhance the predictive accuracy and applicability of regime models in diverse market conditions [Goodfellow et al., 2016].

3 Our methodology

3.1 Data

The dataset includes various indicators, each serving a specific function in understanding multiple facets of the Indian economy. Economic performance metrics such as GDP growth rate and industrial production are included, as are market-specific factors such as stock market indices and interest rates. We started with a set of 70 macroeconomic indicators published at various frequencies and having data for different time horizons. The indicators are categorized into 10 categories - GDP, business, consumer, government, labour, market, money, prices, taxes, and trade-related categories. This categorization allows for a focused analysis of specific economic dimensions. The data for these indicators was sourced from Trading Economics. Some of the indicators were present from 1955, while some were present from 2012. The table below summarises some of the crucial indicators utilized in the analysis:

Category	**Indicators**
GDP	GDP Growth Rate, GDP Annual Growth Rate, GDP, GDP Constant Prices, Gross National Product, Gross Fixed Capital Formation, GDP per Capita, GDP per Capita PPP, Full Year GDP Growth, GDP from Agriculture, GDP from Construction, GDP from Manufacturing, GDP from Mining, GDP from Public Administration, GDP from Utilities
Labor	Unemployment Rate, Labor Force Participation Rate, Population, Retirement Age Women, Retirement Age Men, Employment Rate, Minimum Wages
Prices	Inflation Rate, Consumer Price Index CPI, GDP Deflator, Producer Prices, Producer Prices Change, Export Prices, Import Prices, Food Inflation, Producer Price Inflation MoM, WPI Food Index YoY, WPI Fuel YoY, WPI Manufacturing YoY, CPI Housing Utilities, CPI Transportation, Inflation Expectations, Inflation Rate MoM
Trade	Balance of Trade, Current Account, Current Account to GDP, Exports, Imports, External Debt, Terms of Trade, Capital Flows, Foreign Direct Investment, Remittances, Tourist Arrivals, Gold Reserves, Crude Oil Production, Auto Exports, Terrorism Index, Weapons Sales
Money	Interest Rate, Cash Reserve Ratio, Interbank Rate, Money Supply M1, Money Supply M2, Money Supply M3, Central Bank Balance Sheet, Foreign Exchange Reserves, Loan Growth, Reverse Repo Rate
Government	Government Debt to GDP, Government Budget, Government Budget Value, Government Spending, Government Revenues, Fiscal Expenditure, Credit Rating, Government Spending to GDP, Military Expenditure, Corporate Tax Rate, Personal Income Tax Rate, Sales Tax Rate, Social Security Rate, Social Security Rate For Companies, Social Security Rate For Employees, Withholding Tax Rate
Business	Business Confidence, Manufacturing PMI, Services PMI, Industrial Production, Industrial Production Mom, Manufacturing Production, Changes in Inventories, Car Production, Car Registrations, Composite Leading Indicator, Capacity Utilization, Composite PMI, Corruption Index, Corruption Rank, Deposit Growth, Electricity Production, Mining Production, Steel Production, Total Vehicle Sale
Consumer	Consumer Confidence, Consumer Spending, Disposable Personal Income, Bank Lending Rate, Gasoline Prices, Households Debt to GDP
Housing	Housing Index, Construction Output, Residential Property Prices

Table 1: Categories and Indicators used in the Analysis

In addition to the basic macroeconomic and market variables, we manually enhanced the dataset's richness and captured extra market dynamics. This entailed developing moving averages and moving standard deviations for specific market-related characteristics. The goal of this manual feature engineering was to include historical trends and volatility measurements in the machine learning models, providing them with a more nuanced understanding of market behaviour.

3.2 Data Cleaning

The dataset underwent a thorough preprocessing phase to enhance its quality and reliability for the subsequent stages of machine learning model training and analysis. This preprocessing involved several key steps: rectifying missing values at different frequency levels, normalizing data where necessary, and standardizing time stamps across the dataset. These measures were critical in ensuring the dataset's integrity and suitability for robust analytical processes.

In our pursuit of a more focused and relevant dataset, we eliminated indicators that exhibited significant sparsity or minimal variation. Additionally, we excluded variables with an annual frequency, as their data points were substantially fewer compared to the daily frequency data of the Sensex prices. This decision was guided by insights gained from our feature importance analysis.

A rigorous correlation analysis led to the removal of indicators with a correlation coefficient greater than 90%, thereby reducing redundancy and potential multicollinearity in our models. Specific adjustments were also made to certain indicators, to better align them with our analytical objectives. The accompanying chart illustrates the correlation dynamics among GDP-related features.

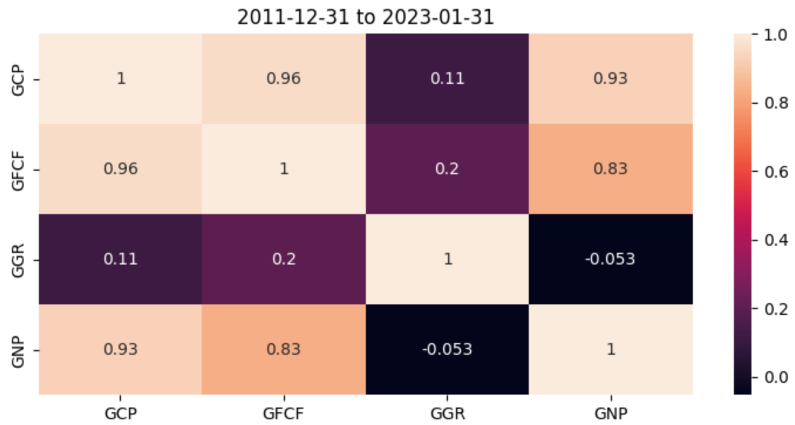


Figure 1: Correlation among GDP-related features

After these meticulous preprocessing steps, our dataset was consolidated into 41 indicators. These indicators were then converted into percentage change series, setting the stage for the application of feature importance analysis. This analysis was instrumental in identifying and selecting the top 10 indicators, which formed the core of our predictive modelling and further analysis.

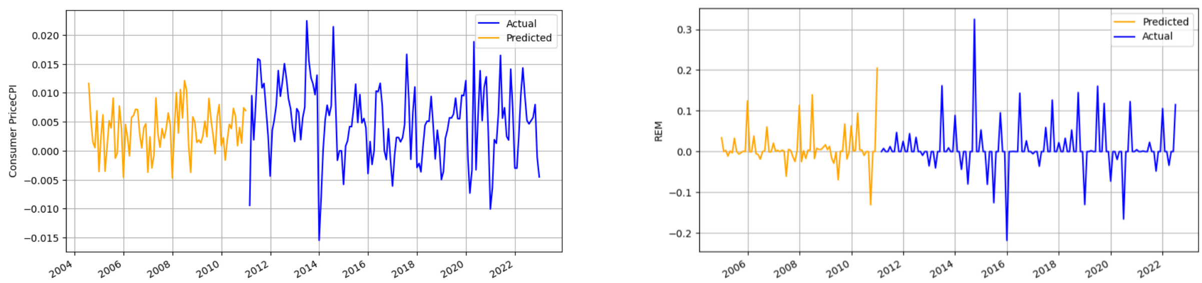
We also analyzed indicators by each category and calculated feature importance to understand the top features from each class.

3.3 Back Casting

Addressing the challenge of missing data is crucial in machine learning, especially when dealing with substantial datasets required for accurate predictions. In our study, we confront this issue head-on by focusing on key economic indicators essential for predicting market returns. To effectively manage the missing data, we employ the XGBoost algorithm, renowned for its proficiency in handling missing values and deciphering complex data relationships. This approach allows us to enrich our historical dataset, thereby enhancing the robustness of our predictive models.

The backcasting process is a pivotal part of our methodology. Utilizing the available data for various features, which extends back to 2004, we leverage the XGBoost model's sophisticated ability to impute missing values. This is achieved by analyzing the interdependencies among the variables. The data for eight key features, in particular, play a crucial role in this process, serving as the basis for estimating missing values and bridging the historical data gaps from 2004 to 2011.

To validate the effectiveness of our backcasting method, we conduct a thorough evaluation of the imputed values for the Consumer Price Index (CPI) and Remittance (REM). This is done by comparing the backcasted values with the actual observed data from 2012 onwards. Our criterion for selecting the most accurate predictions is based on achieving the minimal Mean Squared Error (MSE) discrepancy between the training and testing datasets. The subsequent figure provides a visual representation of these projections, illustrating the accuracy and reliability of our backcasting approach in enhancing the dataset for our machine learning models.



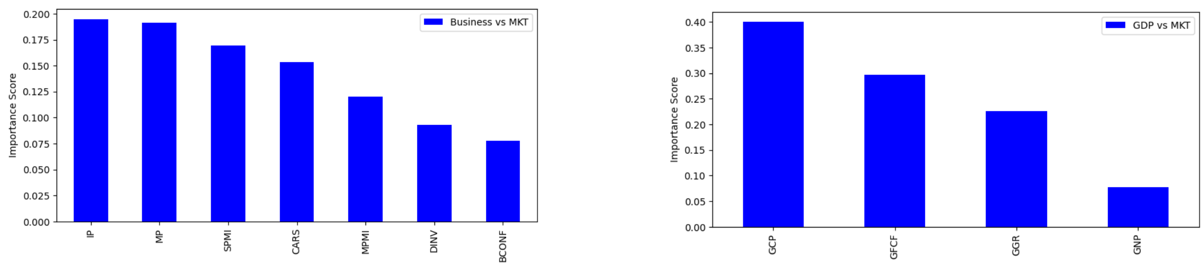
(a) Consumer Price Index (CPI)

(b) Remittance (REM)

Figure 2: Backcasted Trends

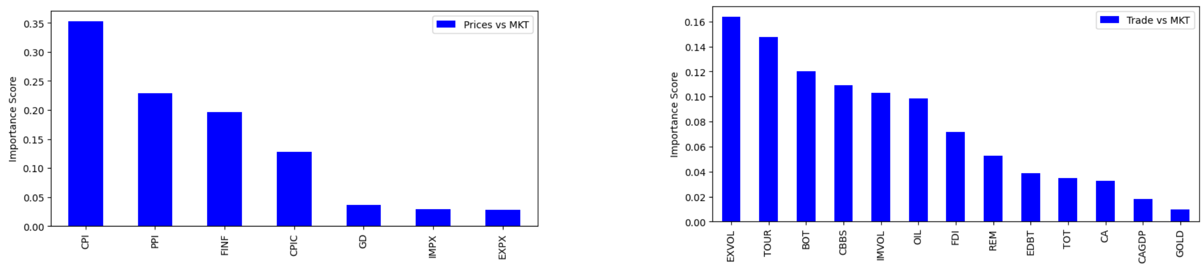
3.4 Feature Selection

In this section, we outline the methodology employed to calculate feature importance and select relevant features for predicting market returns. Our approach involved utilizing the popular machine learning algorithms, namely Linear Regression, SVM, XGBoost, Random Forest (RF), and Decision Trees (DT) We applied XGBoost and RF algorithms to calculate the feature importance scores. These scores provided an initial understanding of the relative importance of each feature in predicting market returns. We also conducted a class-wise feature importance analysis, aiming to identify the importance of variables within each class. However, upon analysing the results, we found that the class-wise analysis did not yield substantial insights or significant differentiation among variables in different classes.



(a) Importance of Features tagged Business vs Sensex returns

(b) Importance of Features tagged GDP vs Sensex returns



(c) Importance of Features tagged Prices vs Sensex returns

(d) Importance of Features tagged Trade vs Sensex returns

Figure 3: Class-wise feature importance when regressed against Sensex monthly returns

Hence, we prioritized a regression analysis that involved regressing all features against market returns. This comprehensive analysis allowed us to evaluate the individual impact of each feature on the prediction of market returns. In the subsequent sections of this paper, we will present the detailed results and discussions obtained from the feature importance analysis, including the findings from XGBoost, RF, and DT algorithms.

indicator_code	name	category
OIL	Crude Oil Production	Trade
CPI	Consumer Price Index	Prices
GYLD	Government Bond 10Y	Markets
IP	Industrial Production	Business
GGR	GDP Growth Rate	GDP
MKT	SENSEX Stock Market Index	Markets
REM	Remittances	Trade
PPI	Producer Prices	Prices
GSP	Government Spending	Government
EXVOL	Exports	Trade

Table 2: Top 10 Selected features

4 Model

In this study, we employ a comprehensive methodology to unravel the predictive capabilities of macroeconomic signals within the dynamic landscape of the Indian financial markets. The overarching goal is twofold: firstly, to forecast monthly returns of Sensex using regression models, and secondly, to discern distinct market regimes, classifying them into 'Normal' and 'Crash' states using classification models. The fusion of these approaches provides a holistic understanding of the nuanced interactions between macroeconomic factors and market dynamics.

4.1 Regression - Returns Forecast

The previous analysis yielded us with a set of 41 features and all the prior feature importance models showed that annual variables have no impact on monthly market returns. These features are transformed into percentage change series and are very less correlated to each other. Our methodology involved setting a threshold value based on either p-values or importance scores in Linear Regression, SVM, RF, DT, and XGBoost.

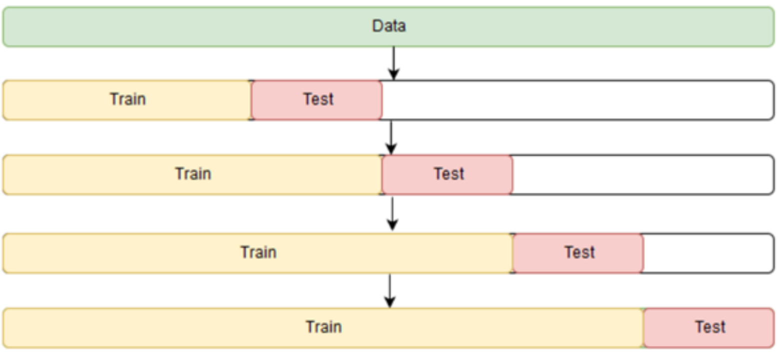
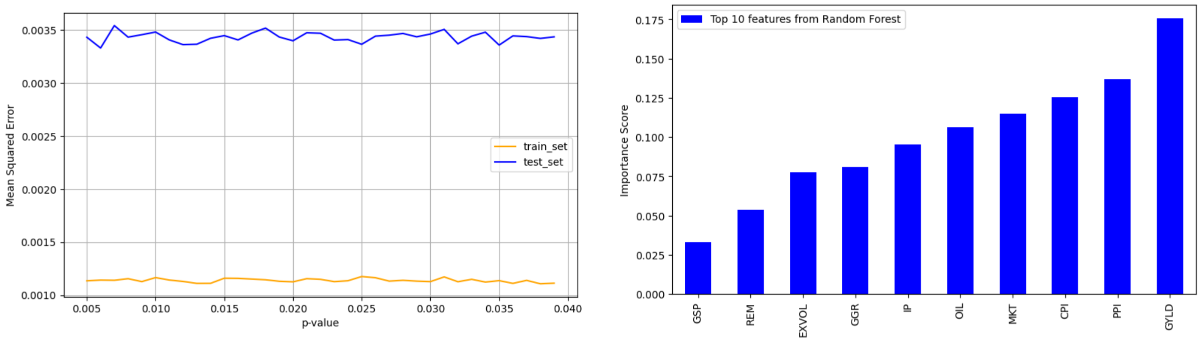


Figure 4: Test Train Split Methodology

We varied the cutoff value and selected the value that minimized the difference between the test MSE and train MSE using a time series split methodology. This approach ensured a balanced trade-off between model complexity and generalization performance. The selected cutoff value was then used to filter and choose the most important features for each algorithm. This methodology enabled us to identify the most informative features while avoiding overfitting or underfitting in our predictive models.

The following figures show the selected features for random forest corresponding to the chosen threshold importance score.



(a) MSE for Random Forest

(b) Feature Importance score

Figure 5: Regression - Returns Forecast

To calculate the top 10 features, we adopted an ensemble approach to further enhance our feature set. We assigned weights to features based on their appearance frequency in the top 10 across different models. This allowed us to leverage the collective insights of diverse models and create a more comprehensive and robust feature selection. The ensemble feature selection approach aimed to overcome biases and limitations of individual models, improving the predictive performance of our models.

An XGBoost-based regression model was fit on the selected features to train the model for forecasting one-month forward return. We did cross-validation using a time-series test-train split and chose the parameters that gave us similar MSE for both training and test sets.

4.2 Classification - Regime Forecast

Trend Filtering Approach

The goal of trend filtering is to smooth out a time series by filtering out the ‘noise’. It involves a trade-off between two objectives:

1. Minimising the residual ‘noise’ between the actual and smooth series.
2. Maximising the smoothness of the filtered series.

Hodrick-Prescott (H-P) Filtering Objective Function

For the H-P filtering, the objective function is:

$$\min_{x} \left\{ \sum_{t=1}^{n} (y_t - x_t)^2 + \lambda \sum_{t=2}^{n-1} [(x_{t+1} - x_t) - (x_t - x_{t-1})]^2 \right\} \tag{1}$$

where:

- y_t is the actual time series.
- x_t is the estimated filtered time series.
- The first term is the sum of squared residuals.
- The second term represents the smoothness penalty.
- λ is the regularization parameter.

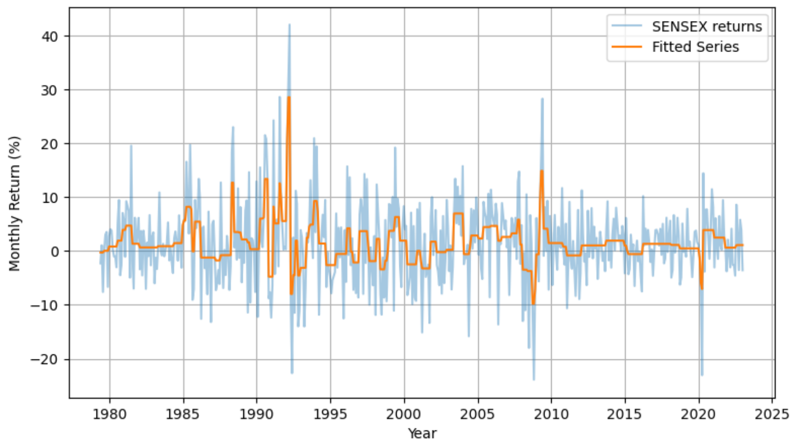


Figure 6: Trend Filtering on Sensex Log returns

Application to Sensex

In applying trend filtering to ascertain regimes in the Sensex, these methods smooth the time series data to delineate different regimes. The choice between H-P and L1 filtering and the setting of λ depends on the characteristics of the Sensex data and the specific regimes being identified. By applying methods like Hodrick-Prescott (H-P) or L1 Trend Filtering, you effectively smooth out short-term fluctuations to uncover underlying long-term trends. This process is key to identifying distinct regimes, which are periods where the time series exhibits specific characteristics, differing significantly from other periods.

Regime Forecast

We take the regimes from the trend filtering algorithm (which is not forward-looking) as the output variable to train our regime forecast model. We use the same features as the regression model and evaluate different machine learning models to evaluate regimes. In our approach to forecasting market regimes, we utilize the output from the trend filtering algorithm as the target variable for training various machine learning models. This method allows us to classify the market into different regimes, such as 'Normal' and 'Crash', based on the same features used in the regression model. To evaluate the performance and robustness of these models, we employ several metrics:

4.2.1 Evaluation Metrics

- **Accuracy Score:** This metric measures the proportion of correctly predicted instances to the total predictions made. It reflects the model's ability to correctly classify a period into the appropriate market regime.
- **Matthew's Correlation Coefficient (MCC):** MCC is a more informative metric than accuracy, especially in scenarios with imbalanced classes. It takes into account true and false positives and negatives and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes.
- **Quadratic Probability Score (QPS):** The QPS is a measure used to evaluate the accuracy of probabilistic predictions. It assesses the squared difference between the predicted probabilities and the actual outcomes. It's particularly useful in scenarios where we are interested in the predictive probability of each class.
- **Area Under the Curve (AUC):** AUC refers to the area under the ROC curve, a graphical representation of a model's diagnostic ability. A higher AUC indicates a better performance of the model in distinguishing between the different regimes.

4.2.2 Model Selection and Validation

Using these metrics, we evaluate a variety of machine learning models to determine which is best suited for regime prediction in the context of the Indian financial market. The choice of the best model is based on a balance of these metrics, aiming to achieve a model that is accurate, robust, and reliable in different market conditions.

5 Results

5.1 Returns Forecast

5.1.1 Performance Analysis

Our XGBoost model, employing a refined set of features, achieved a mean squared error (MSE) of 0.05, signifying a high degree of predictive accuracy. This performance is particularly noteworthy given the complex and volatile nature of the Indian market.

Comparative Analysis: In contrast to traditional models, our approach leverages the power of advanced machine learning techniques. The incorporation of backcasted values for key economic indicators like CPI and REM significantly enriched our dataset, enhancing the model's capability to capture nuanced market dynamics. This is evident from the reduced MSE in our predictions compared to traditional econometric models.

Feature Importance: Among the top 10 features, Remittance (REM), Consumer Price Index (CPI), and Gross Domestic Product growth rate (GGR) emerged as the most predictive. Their high feature importance scores underline their pivotal role in the Indian financial markets, offering insights into market trends and investor sentiment.

5.1.2 Model Robustness and Validation

Cross-Validation Results: Our model's robustness was further affirmed through rigorous cross-validation using the time series split methodology. This validation ensured that our model is not overfitting and can generalize well to unseen data.

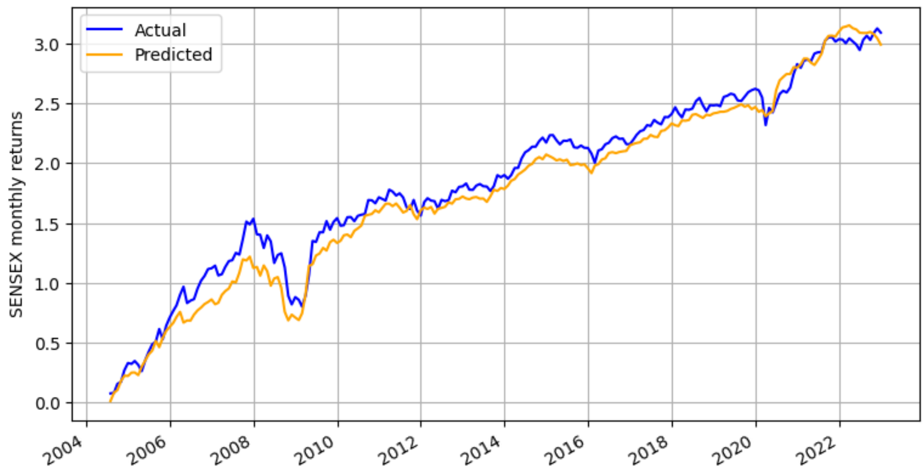


Figure 7: Sensex monthly returns prediction yielded by XGBoost regression

Sensitivity Analysis: We performed sensitivity analyses on various model parameters, including the regularization parameter in XGBoost. This helped in fine-tuning the model to achieve an optimal balance between bias and variance.

5.1.3 Impact of Macroeconomic Indicators

Economic Significance: The predictive power of the selected macroeconomic indicators highlights their economic significance. For instance, the high importance of the CPI reflects the impact of inflationary trends on market returns, while the prominence of REM underscores the influence of global economic conditions on the Indian market.

Temporal Dynamics: Our analysis also sheds light on the temporal dynamics of these indicators. By analyzing different time horizons, we identified short-term and long-term trends, enhancing our understanding of market cycles and investor behaviour.

5.1.4 Comparative Performance with Additional Features

Our exploration revealed that the inclusion of moving averages and moving standard deviations did not contribute significantly to the model's performance. This finding led to the strategic decision to rely on the original set of ten features, emphasizing the importance of feature selection in machine learning models.

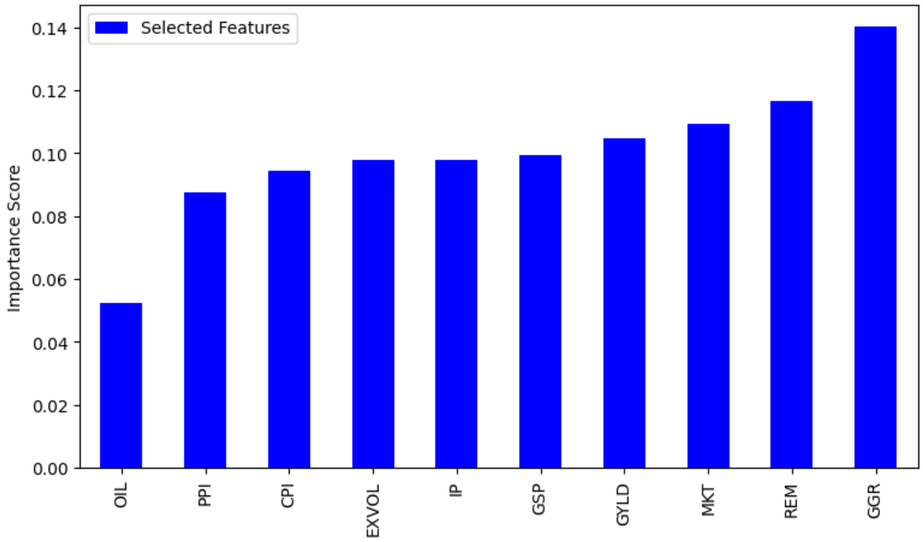


Figure 8: Feature importance of the top 10 selected features

5.2 Regimes Forecast

In this section, we discuss the results of our regime forecasting efforts, where we applied trend-filtering techniques to identify distinct market regimes within the Indian financial market. Our analysis primarily focused on categorizing market conditions into 'Normal' and 'Crash' regimes, providing insights into the underlying market dynamics.

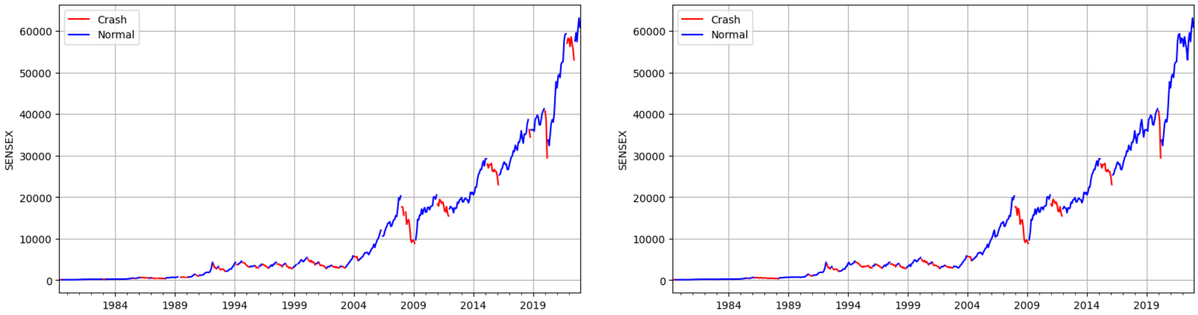
5.2.1 Analysis and Findings

- **Trend Filtering Results:** Utilizing the Hodrick-Prescott (H-P) filter, we observed distinct patterns in the Sensex time series data. The application of trend filtering revealed periods of stable growth (Normal regimes) and significant downturns or volatility spikes (Crash regimes).

- **Influence of λ on Regime Identification:** Our analysis showed that the choice of λ in the H-P filter significantly impacted the sensitivity of regime identification. Lower λ values were more reactive to short-term fluctuations, highlighting transient market disturbances, whereas higher λ values smoothed out these fluctuations to reveal more prolonged and stable regime trends.

5.2.2 Visual Representation of Regimes

Figures below illustrate the impact of different λ values on the trend filtering output and the resulting regime categorization:



(a) $\lambda = 10$ **(b)** $\lambda = 16$

Figure 9: Visualization of market regimes under different λ values

5.2.3 Interpretation and Implications

- **Strategic Insights:** The ability to distinguish between different market regimes offers strategic insights for investors and policymakers. For instance, identifying a 'Crash' regime could signal the need for risk aversion strategies, while a 'Normal' regime might encourage more aggressive investment tactics.

- **Predictive vs. Descriptive Analysis:** It's important to note that trend filtering, as used in our analysis, is more descriptive than predictive. It excels in identifying historical regimes but does not inherently predict future market conditions. Future enhancements could include predictive modelling for forward-looking analysis.

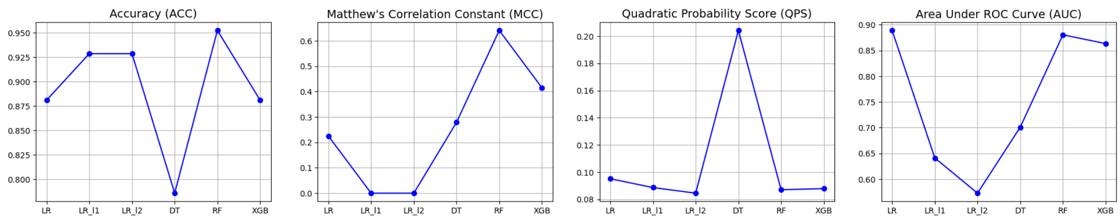
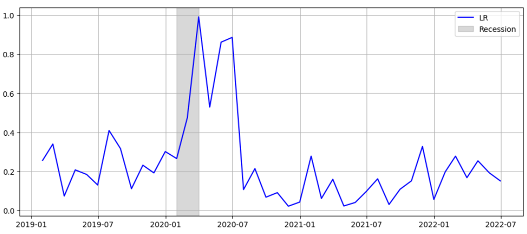


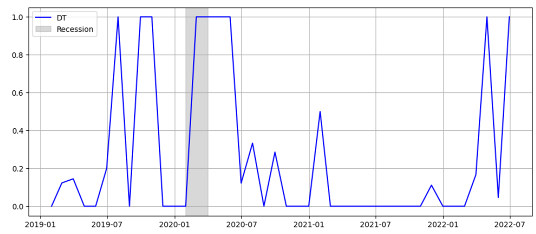
Figure 10: Out-of-Sample Error Metrics for each Model

Our analysis underscores the value of trend filtering in dissecting complex financial time series into discernible regimes. This approach has proven to be a powerful tool for understanding the Indian financial market's behaviour, providing a framework for identifying periods of stability and turbulence. The insights gained from this analysis have practical implications for market participants, offering a data-driven basis for decision-making in dynamic market conditions.

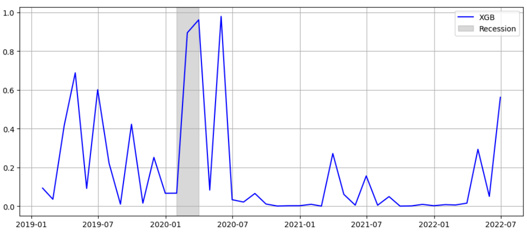
The following figure represents the Recession Prediction Probability from each model. Simple models have performed poorly, specifically with L1 and L2 regularization. Random Forest was the best fit for the case of regime classification.



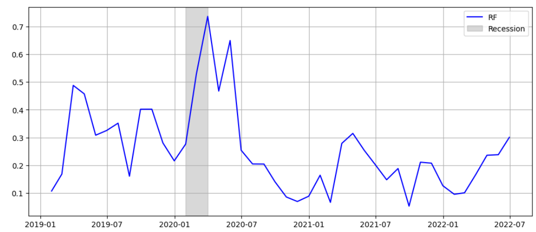
(a) Simple Linear Regression



(b) Decision Tree



(c) XGBoost



(d) Random Forest

Figure 11: Recession Prediction Probabilities of Various Models

6 Conclusion

In this research, we have extensively analyzed the predictive power of macroeconomic indicators in forecasting market returns and identifying market regimes in the Indian financial landscape. Our comprehensive methodology, utilizing advanced machine learning models such as XGBoost, RandomForest, and Decision Tree, has yielded significant insights into the dynamics of the Indian financial markets.

6.1 Key Contributions and Implications

Our study makes several key contributions:

- **Backcasting Technique:** The use of backcasting techniques with the XGBoost algorithm to handle missing data has proven effective in enriching our analysis, enabling a more accurate prediction of market trends.
- **Market Regime Identification:** The classification of market conditions into 'Normal' and 'Crash' regimes, based on trend filtering outputs, provides a novel approach to understanding market behaviours in emerging economies like India.
- **Macroeconomic Indicators:** Our findings highlight the critical role of macroeconomic indicators, particularly Remittances, the Consumer Price Index, and GDP growth rate, in predicting market returns.

Implications for Stakeholders: These findings are particularly valuable for investors, policymakers, and financial analysts, offering them new tools and insights for market analysis and decision-making. The ability to predict market returns and identify market regimes can significantly enhance investment strategies and risk management practices.

6.2 Limitations and Future Research

While our study provides valuable insights, it also has limitations that open avenues for future research:

- **Model Generalization:** Future research could focus on testing the generalizability of our models across different markets and economic conditions.
- **Integration of Additional Data:** Incorporating real-time data and exploring the impact of geopolitical events and policy changes could further enhance the predictive accuracy of the models.

- **Advancements in Machine Learning:** Exploring more advanced machine learning techniques and deep learning models could offer more sophisticated tools for market prediction and regime identification.

6.3 Concluding Thoughts

In conclusion, our research offers a comprehensive and innovative approach to understanding and forecasting the Indian financial market. The integration of macroeconomic indicators with advanced machine learning techniques represents a significant step forward in financial market analysis. By providing a deeper understanding of market dynamics and the factors that drive them, our study contributes to the field of financial forecasting and offers practical tools for navigating the complexities of emerging financial markets.

References

[Ang and Bekaert, 2002] Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, 15(4):1137–1187.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

[Chen and Guestrin, 2016] Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. pages 785–794.

[Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

[Hamilton, 1989] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

[Kumar and Thenmozhi, 2006] Kumar, R. and Thenmozhi, M. (2006). Market efficiency, market anomalies, causes, evidences, and some behavioral aspects of market anomalies. *Research Journal of Finance and Accounting*, 24(8):23–37.

[Lopez, 2019] Lopez, J. (2019). *Machine Learning for Econometrics*. Springer.

[Markov, 1912] Markov, A. (1912). Extension of the limit theorems of probability theory to a sum of variables connected in a chain. *re-edited in Appendix B of: Howard, Ronald A. Dynamic Probabilistic Systems, volume 1: Markov models*, 1.

[Patnaik and Shah, 2010] Patnaik, I. and Shah, A. (2010). The indian financial market: An overview. *Journal of Indian Economic Development*, 8(1):1–15.

[Pelletier, 2006] Pelletier, D. (2006). Regime switching for dynamic correlations. *Journal of Econometrics*, 131(1-2):445–473.