Prediction of Nifty Sector Indices Using Macroeconomic Variables

Anay Malviya

May 25, 2024

Abstract

This report aims to predict the performance of Nifty sector indices using global and local macroeconomic variables. Data from the past 10 years was gathered and analyzed. Various models, including linear regression, Random Forest, ARIMA, SARIMA, and LSTM, were employed to predict which sector indices would outperform the Nifty 50. The results and efficacy of these models are discussed.

1 Introduction

The objective of this project is to analyze and predict the performance of various Nifty sector indices using historical macroeconomic data. Understanding the relationship between macroeconomic indicators and stock market performance can help investors make informed decisions. This study initially intended to focus on the last 10 years of data for indices like Nifty Auto, Nifty Bank, and others. However, due to time constraints and unsatisfactory results from the Nifty 50 analysis, the scope was limited to the Nifty 50 index.

2 Data Collection

2.1 Historical Nifty Sector Indices Data

Data for Nifty sector indices such as Nifty Auto, Nifty Bank, etc., was downloaded from `yfinance` for the period from 2011 to 2021.

2.2 Macroeconomic Variables

Macroeconomic data was collected from various sources including:

- World Bank
- Indian Fama French Momentum
- Macrotrends

The variables collected included GDP, GNI per capita, GNP, growth rate, manufacturing percentage of GDP, and annual inflation rate. The data was interpolated linearly to obtain monthly values.

3 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted, and the macroeconomic variables were interpolated linearly to obtain monthly data from annual data. Correlation analysis was then performed to identify significant predictors of the Nifty sector indices.

4 Approach

4.1 Current Approach

The approach taken in this study involved several steps:

- **Initial Focus on Nifty 50**: The initial focus was on predicting the Nifty 50 index performance using various macroeconomic variables as a foundational model.
- **Feature Engineering**: The plan was to perform feature engineering on this model to identify relevant features that could be used to predict the performance of other sector indices.
- **Baseline Model Development**: By developing a reliable model for the Nifty 50, the intention was to use it as a baseline to explore which sectors would outperform the Nifty 50 under different macroeconomic conditions.

4.2 Alternate Approach

In hindsight, a more effective approach might have been:

- **Integrated Modeling**: To model the sectors and the Nifty 50 together within the same framework.
- **Combined Analysis**: By incorporating sector indices and the Nifty 50 within the same model, it would be possible to observe the macroeconomic conditions under which specific sectors outperform the Nifty 50.
- **Comprehensive Understanding**: This approach could provide a more comprehensive understanding of the interactions between different sectors and the broader market index.

5 Modeling

Several models were used to predict the sector indices:

5.1 Linear Regression

A linear regression model was initially used and, as expected, it showed very limited predictive power.

5.2 Random Forest Regressor

A Random Forest Regressor was also employed, expecting it to capture non-linear relationships. However, the performance was not satisfactory. Random Forest was used to evaluate the contribution of each feature.

5.3 ARIMA

ARIMA models were used to incorporate time-series characteristics of the data. We searched for the best parameters and obtained the best ARIMA parameters: (0, 3, 4) with AIC: 1506.19. The best model had a Mean Squared Error (MSE) of 4633637.21 and an R-squared score of 0.378. Adding lagged Nifty 50 values as inputs, the best parameters obtained were (5, 4, 1). This model gave a Mean Absolute Error (MAE) of 1105.431, a Root Mean Squared Error (RMSE) of 1254.56, and R-squared score of 0.7748.

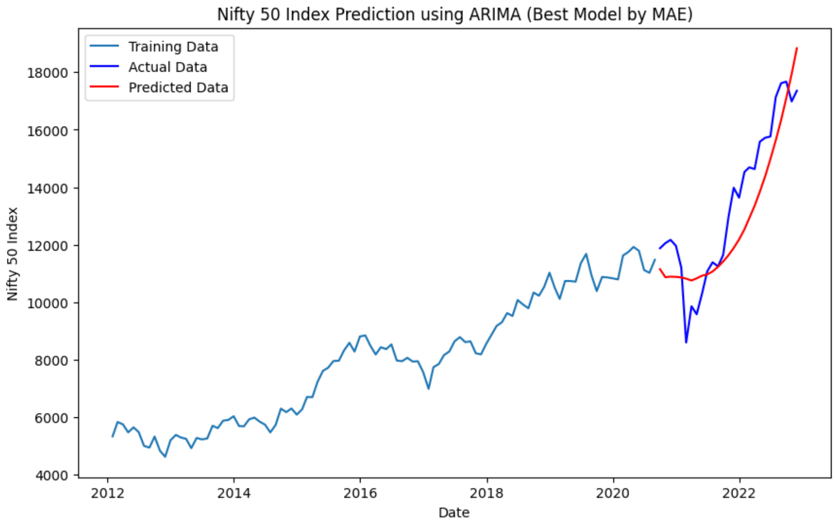

Figure 1: ARIMA Model Performance

5.4 LSTM

LSTM networks were applied to leverage their ability to capture long-term dependencies in the data. Using a single-lagged Nifty value, the training RMSE was 559.33 with an R-squared score of 0.929. However, the test RMSE was 2024.06 with an R-squared score of 0.491. Increasing the units led to overfitting, and adding more lagged values or dropping noisy features did not improve results.

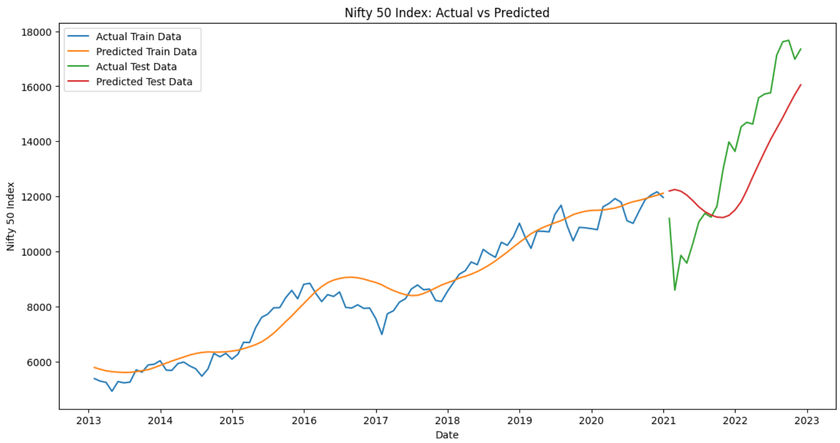

Figure 2: LSTM Model Performance

5.5 SARIMA

We also employed SARIMA models, searching for the best order and seasonal order. The best model had a Mean Absolute Error (MAE) of 1166.82 and a Root Mean Squared Error (RMSE) of 1556.80.

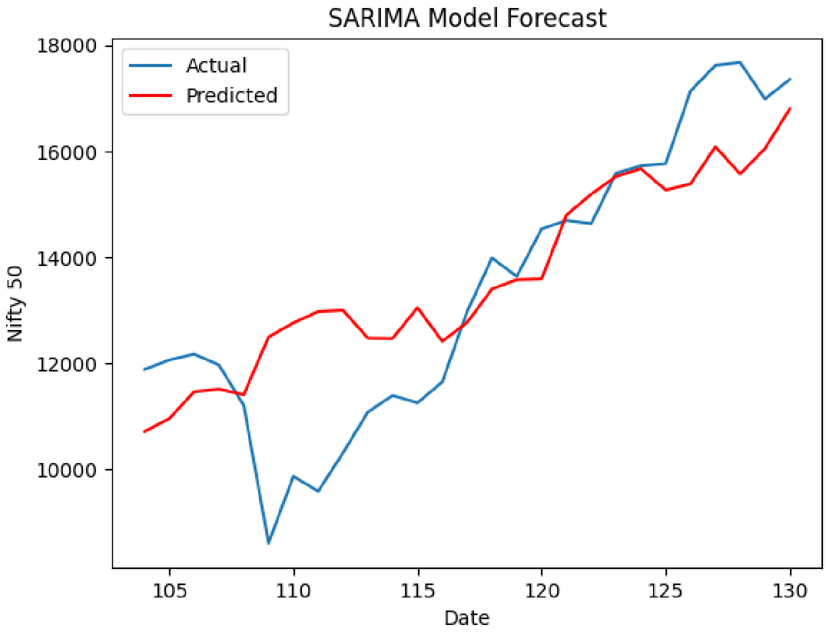

Figure 3: SARIMA Model Performance

6 Results

The models' performances were evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.

Model	MAE	RMSE	R-squared
ARIMA	1105.431	1254.56	0.7748
LSTM	...	2152.58	0.491
SARIMA	1166.82	1556.80	0.891

Table 1: Model Performance Metrics

Despite various modeling attempts, the predictions were not significantly better than baseline models. Incorporating lagged Nifty 50 values as inputs did not improve performance substantially. The unsatisfactory results can be attributed to several factors:

- **Complexity of Stock Market**: The stock market is influenced by numerous factors, including investor sentiment, political events, and sudden economic changes, which are difficult to capture with macroeconomic variables alone.
- **Data Quality and Granularity**: The interpolated monthly data from annual data may not accurately reflect the true monthly economic conditions, leading to high inaccuracies.

7 Conclusion

The study attempted to predict the performance of Nifty sector indices using macroeconomic variables and various modeling techniques. While the models provided some insights, none showed strong predictive capabilities. Future work could explore more sophisticated feature engineering, alternative modeling approaches, or a deeper dive into macroeconomic indicators' time lags.

References

[1] World Bank, `https://data.worldbank.org/`

[2] Indian Fama French Momentum, https://faculty.iima.ac.in/iffm/Indian-Fama-French-Momentum/

[3] Macrotrends, https://www.macrotrends.net/global-metrics/countries/IND/india/gdp-gross-domestic-product