Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/

(Research Article)

Machine learning model for treasury bill yields prediction in Kenya

Kennedy Mung'are Njeri [1, *], Emma Anyika [2] and Kennedy Hadullo [2]

[1] *Department of Mathematical Sciences, The Cooperative University of Kenya, Karen, Nairobi, Kenya.*
[2] *Department of Computer Science and Information Technology, The Cooperative University of Kenya, Karen, Nairobi, Kenya.*

Global Journal of Engineering and Technology Advances, 2025, 24(03), 012-020

Publication history: Received on 25 July 2025; revised on 29 August; accepted on 01 September 2025

Article DOI: https://doi.org/10.30574/gjeta.2025.24.3.0259

Abstract

In this paper, we investigated the issue of forecasting the yields of treasury bills in the Kenyan financial market which is both volatile and complicated, and which traditional models of forecasting may fail because of non-linear behavior. We developed, trained and tested a hybrid machine learning model to improve the predictive power and stability of the model by integrating ARIMA to analyze linear trends, Support Vector Machines (SVM) to capture non-linear interdependencies, and Facebook Prophet (FB Prophet) to capture seasonality and handle missing data. The methodology consisted of obtaining information at the Central Bank of Kenya (CBK) of Treasury bill yields between July 2022 and June 2024. Models were trained and tested on performance measures, namely Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Scaled Error (MASE) and cross-validation was used to increase reliability. The findings indicated that Gaussian Copula ensemble model is a more effective model in predicting 364-day Kenyan Treasury bills yields. The hybrid model generated the least Mean Absolute Error (MAE) of 0.1187 compared to best-performing individual model, SVM which had an MAE of 0.1806. The paper concludes that this combination of linear, non-linear, and seasonal-trend models using the specific advantages of each model can offer more reliable and robust forecasts as compared to traditional ones. The model can assist in making intelligent decisions and risk management as well as formulating effective economic policies.

Keywords: Treasury bills; Ensemble model; Machine learning; Copula; Financial forecasting in Kenya

1. Introduction

Treasury bills (T-bills) hold an important place in Kenya's financial landscape as a safe, short-term investment tool through which an investor can access predictable returns. Issued by the government to raise funds and manage public expenditure, T-bills form a central point in the country's monetary policy, with their rates influencing broader economic dynamics like loan and savings rates (1,2). The problem that this study addresses is the inherent difficulty in predicting T-bill yields with accuracy, because of the volatile and complex nature of financial markets. The reason for this is that traditional forecasting models fall short in capturing the non-linearities that characterize these markets, resulting in poor investment choices and difficulties in formulating appropriate monetary policy (3).

The motivation for this study follows from the substantial need for better forecasting tools. For investors, accurate predictions are the basis for strategic decisions; for policymakers at the Central Bank of Kenya, they serve as crucial inputs for the management of national debt(4). Traditional models often struggle with the non-linear and volatile nature of financial data. This research addresses this gap by proposing a hybrid machine learning model that integrates the Autoregressive Integrated Moving Average (ARIMA) model, Support Vector Machines (SVM), and Facebook's Prophet (FB Prophet), leveraging the strengths of each to create a more robust predictive tool.(5)

* Corresponding author: Kennedy Mung'are Njeri

Copyright © 2025 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0.

The theoretical base of the study lies in the Efficient Market Hypothesis (EMH), which states that the price of an asset already indicates all the information available and, as a result, it is not possible to continually earn an excess return (6). While the EMH offers a general idea, it has been empirically demonstrated that market anomalies and non-linearity's are present in the markets and cannot be represented using the traditional linear models, demanding a more sophisticated approach (7).

Machine learning has emerged as a powerful tool in financial modeling due to its ability to handle non-linear data. ARIMA is a well-established statistical method for analyzing time-series data that assumes linear relationships (3,8). SVM, based on statistical learning theory, excels at capturing complex, non-linear patterns by transforming data into higher dimensions (9,10). FB Prophet is a modern forecasting tool designed to handle seasonality and trend changes effectively (11,12).

The idea of ensemble forecasts to enhance forecasting accuracy is well-known (13). Hybrid models that combine ARIMA and machine learning methods such as SVM have been proposed in the literature to be suitable for a broad range of financial forecasting purposes where the advantages of ARIMA in handling the linear aspects of financial time series can be combined with the ability of SVM to capture the non-linearities (14,15). The combination of such heterogeneous models is usually more powerful than any other individual model, as illustrated in several studies that combine ARIMA with LSTM and GARCH (5,16,17).

Copulas provide an elegant way of constructing these ensembles. A copula is a function that maps univariate marginal distributions into their full multivariate distribution which enables the modeling of complex dependence structure between different variables, in this case forecast errors of the individual models(18–20). By using a copula, the correlated errors between constituent models in an ensemble can be accounted for, resulting in a more robust final prediction(21).

This paper extends this work by proposing a hybrid forecasting system combining ARIMA, SVM, and FB Prophet with a Gaussian copula. The goal is to construct a model that combines the individual strengths of each method to give more accurate and robust predictions for Kenyan T-bill yields over the 364-day period than any one of the individual models would have been able to provide. The study period is about the 364-day T-bill yield from July 2022 to June 2024.

2. Materials and Methods

This study applies a quantitative time-series analysis approach to develop and evaluate a hybrid machine learning model for forecasting Kenyan Treasury bill yields of 364-days.

2.1. Data Sourcing and Preprocessing

This dataset was made by combining three different time series from Central Bank of Kenya (CBK), that is Treasury bill average rates, monthly 12-month inflation rates and the Central Bank Rate (CBR). The data covers the period of July 2022 to June 2024.

The preprocessing pipeline consisted of the following steps:

- Merging: The weekly T-bill data and the monthly inflation and CBR data at a less frequent frequency were aligned using a merge_asof function.
- Imputation: Forward-filling (ffill) was used to disseminate the last known values of inflation and CBR to subsequent dates in order to achieve a complete dataset.
- Feature Engineering: Lagged features were generated to give the models context on historical features. This comprised 12 lags of the target variable (TBill_364), and one lag (shifted by 4 weeks) of the inflation and CBR rates.
- Data Splitting: The final dataset was split chronologically into a training set (the first 80% of the data) and an evaluation set (the final 20%) to simulate a real-world forecasting scenario.

2.2. Individual Model Development

Three different time-series models were developed and tuned:

- ARIMA: The best orders (p, d, q) for the ARIMA model were automatically estimated by using the auto_arima function from the pmdarima library with the lowest Akaike Information Criterion (AIC). The model was estimated on the dependent variable with exogenous regressors (inflation, CBR and their lags).
- Support Vector Machine (SVM): A Support Vector Regressor (SVR) was implemented. Features were scaled using a RobustScaler to handle outliers effectively. An exhaustive hyperparameter search was conducted using three advanced techniques: GridSearchCV, RandomizedSearchCV, and Bayesian Optimization (BayesSearchCV). The best-performing parameters, as identified by Bayesian Optimization, were a linear kernel with C=18.9, epsilon=0.001, and gamma=100.0.
- FB Prophet: Tested out a grid of parameters for tuning the Prophet model: changepoint_prior_scale, seasonality_prior_scale, seasonality_mode ('additive' vs. 'multiplicative'). Exogenous variables such as lagged features were used as regressors to enhance the predictive ability of the model.

2.3. Copula-Based Ensemble Model

The forecasts from the three individual models were integrated into a final ensemble prediction using a Gaussian Multivariate Copula. The process was as follows:

- The individual models were used to generate predictions on the training data. The residuals (actual training values minus predicted training values) were calculated for each model.
- The Gaussian copula was fitted on a DataFrame containing the actual training values and the corresponding predictions from ARIMA, SVM, and Prophet. This allowed the copula to learn the joint probability distribution and the dependence structure of the model errors.
- For the evaluation set, the copula generated the final ensemble forecast by sampling from the learned conditional distribution, given the new predictions from the individual models.

2.4. Evaluation Metrics

All the models were evaluated rigorously using a set of performance metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Scaled Error (MASE). These measures were used as final model comparison measures, with our main focus on MAE.

3. Results and Discussion

This section details the empirical findings from the evaluation of the individual and ensemble forecasting models. The primary goal was to assess whether the proposed hybrid copula model could outperform its constituent parts in predicting 364-day Kenyan Treasury bill yields.

3.1. Individual Model Performance

The performance of the ARIMA, SVM, and FB Prophet models was first evaluated individually on the held-out test set. The results, summarized in the model evaluation table below in Table 2, reveal a competitive landscape.

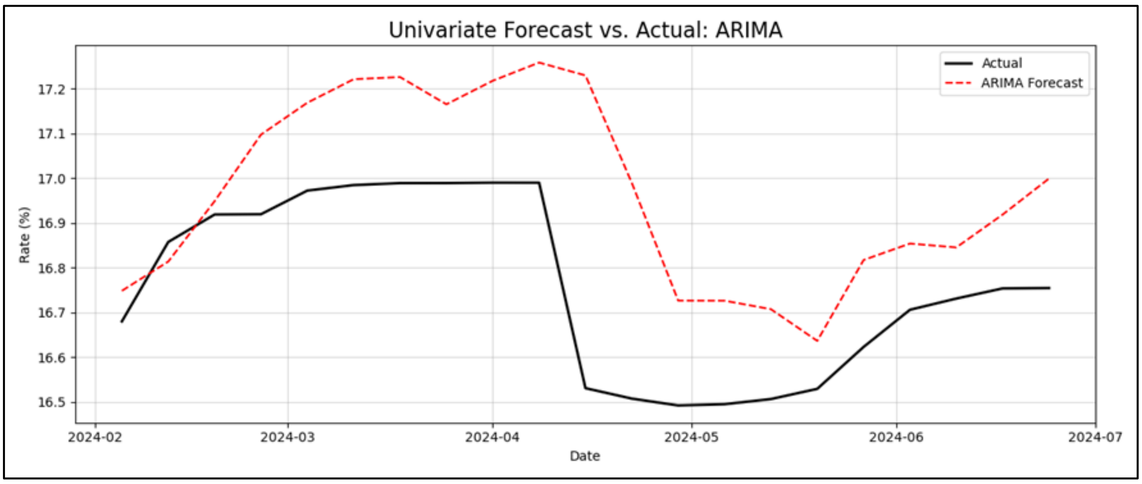

Figure 1 Univariate Forecast vs. Actual: ARIMA

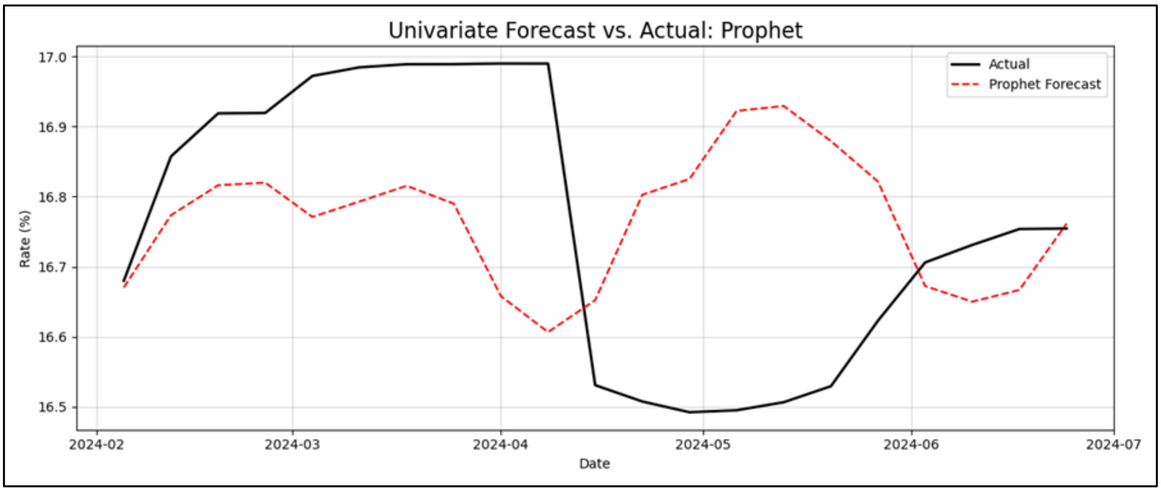

Figure 2 Univariate Forecast vs. Actual: Prophet

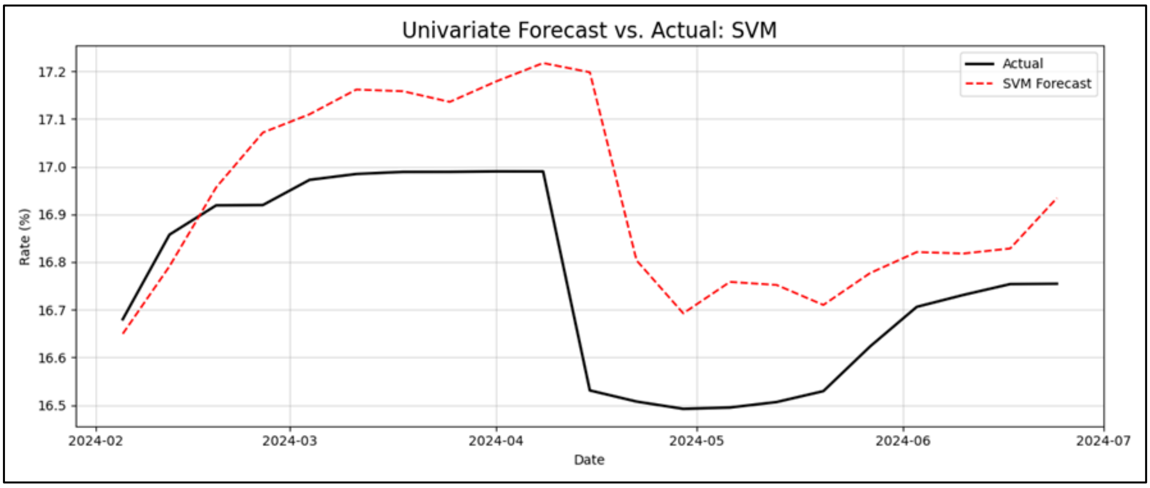

Figure 3 Univariate Forecast vs. Actual: SVM

The SVM model turned out to be the highest performing single model with a Mean Absolute Error (MAE) of 0.1806. It was able to outperform other models by effectively capturing non-linear relationships in the financial data and also by using rigorous hyperparameter tuning through Bayesian Optimization. The ARIMA model was also very good with MAE of 0.1982 in forecasting, which actually captures the linear autoregressive part of the time series. The FB Prophet model, which does a fantastic job of accounting for seasonality, had the largest MAE, which would indicate that the patterns in this dataset were less seasonal and more complex (with many dependencies and trends that the other models could better capture).

3.2. Residual Analysis and Motivation for Ensemble

A crucial step in justifying the ensemble approach was the analysis of the forecast errors (residuals) from the individual models. The correlation matrix of these residuals provided a compelling rationale for using a copula-based method.

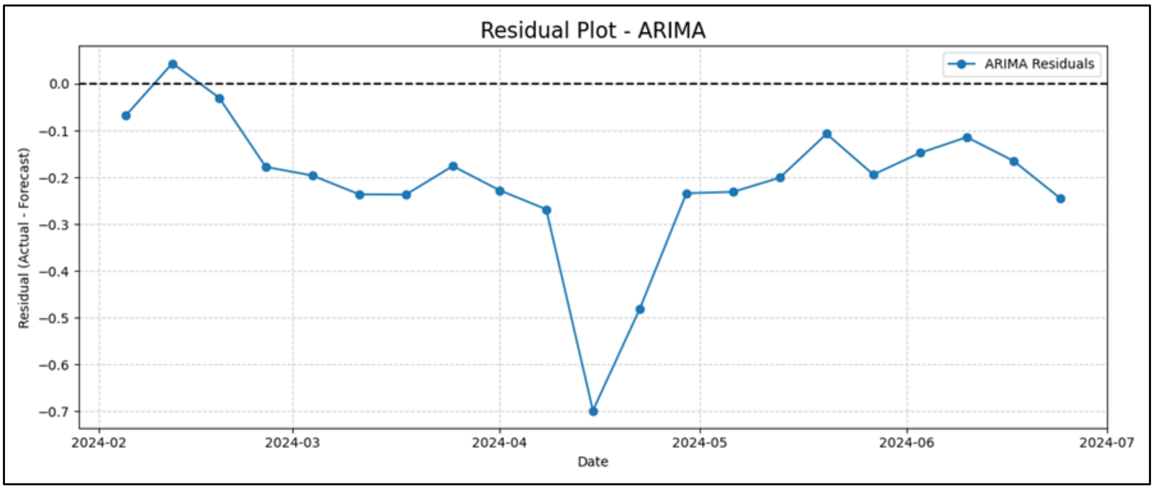

Figure 4 Residual Plot - ARIMA

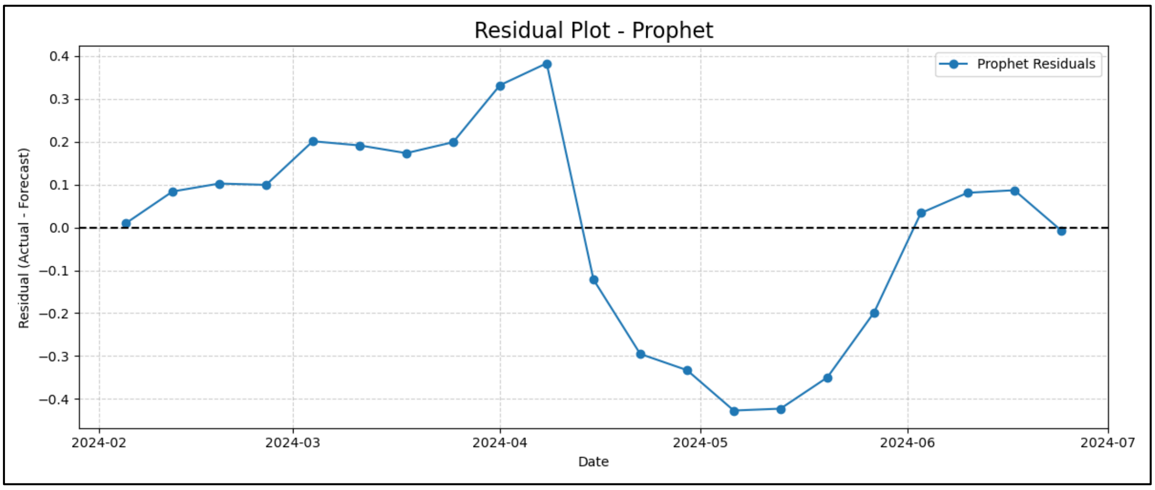

Figure 5 Residual Plot - Prophet

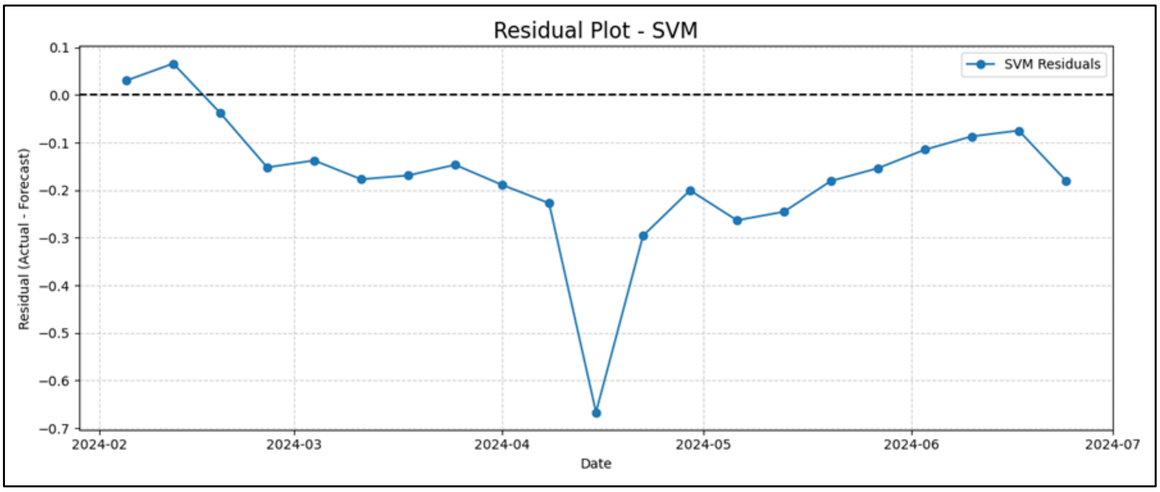

Figure 6 Residual Plot – SVM

Table 1 Residuals Correlation Matrix

	ARIMA_ residuals	Prophet_ residuals	SVM_ residuals
ARIMA_ residuals	1	0.1937	0.9359
Prophet_ residuals	0.1937	1	0.3072
SVM_ residuals	0.9359	0.3072	1

The Table 1 matrix revealed a very high positive correlation of 0.9359 between the residuals of the ARIMA and SVM models. This strong correlation indicates that these two models tended to make similar errors at the same time; when one over-predicted, the other was highly likely to do the same. A simple averaging ensemble would fail to correct for this shared bias. This is precisely the scenario where a copula model excels. By explicitly modeling this dependence structure, the copula can learn to adjust the combined forecast in a more intelligent way, correcting for the systematic, correlated errors of the individual models.

3.3. Ensemble Model Performance

The Gaussian Copula ensemble model was evaluated against the individual models, and the results confirmed its superiority. The ensemble achieved an MAE of 0.1187, representing a 34% improvement over the best individual model (SVM).

Table 2 Final Model Evaluation Metrics

Model	MAE	MSE	RMSE	MAPE	MASE
Gaussian Ensemble	0.1187	0.0284	0.1686	0.7116	1.1488
SVM	0.1806	0.0492	0.2218	1.0815	1.7477
Prophet	0.1967	0.0563	0.2372	1.1765	1.9037
ARIMA	0.2133	0.0659	0.2567	1.276	2.0639

As shown in Table 2, the evaluation metrics, the copula ensemble consistently outperformed all other models across every key metric, including MSE, RMSE, MAPE, and MASE. This demonstrates a significant enhancement in both accuracy and reliability. A direct comparison of the forecasted values against the actual Treasury bill rates for the evaluation period provides a clear view of model performance. Table 3 lists the predicted values from each model alongside the actual yield for each date in the evaluation set. Figure 7 graphically illustrates the data in Table 3, showing that the Gaussian Ensemble forecast tracks the actual yield movements more closely than any other model, particularly during the volatile period around April 2024

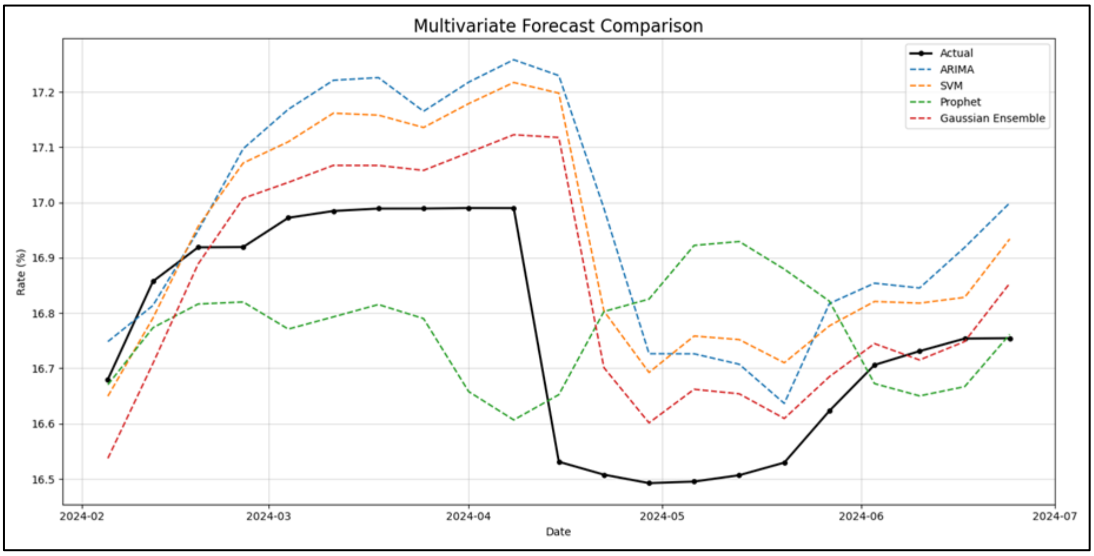

Figure 7 Multivariate Forecast Comparison

Table 3 Comparison of Actual vs. Predicted Yields (%)

ds	Actual	ARIMA	SVM	Prophet	Gaussian Ensemble
2024-02-05	16.6801	16.7484	16.6496	16.6704	16.5625
2024-02-12	16.8574	16.8137	16.7911	16.7735	16.7035
2024-02-19	16.9188	16.9485	16.9564	16.8163	16.8706
2024-02-26	16.9194	17.0972	17.0713	16.82	16.9984
2024-03-04	16.9722	17.1684	17.1096	16.771	17.0238
2024-03-11	16.9845	17.2209	17.1614	16.7931	17.0696
2024-03-18	16.9889	17.2257	17.1578	16.8154	17.078
2024-03-25	16.989	17.1649	17.1355	16.79	17.0364
2024-04-01	16.9899	17.2175	17.1786	16.658	17.1128
2024-04-08	16.9898	17.2581	17.2169	16.6067	17.1453
2024-04-15	16.531	17.2294	17.1976	16.6523	17.131
2024-04-22	16.5077	16.9892	16.8035	16.8027	16.7179
2024-04-29	16.4924	16.7265	16.6925	16.8252	16.6053
2024-05-06	16.4952	16.7263	16.7584	16.9223	16.6705
2024-05-13	16.5067	16.7074	16.752	16.9293	16.6487
2024-05-20	16.5295	16.6366	16.7101	16.8793	16.6159
2024-05-27	16.6231	16.8172	16.7769	16.8214	16.6818
2024-06-03	16.7061	16.8539	16.8209	16.6724	16.7381
2024-06-10	16.7311	16.8453	16.8178	16.6501	16.7303
2024-06-17	16.7538	16.9187	16.8284	16.6669	16.7364
2024-06-24	16.7545	16.9992	16.934	16.7615	16.8617

The discussion of these results highlights two key findings. First, no single model is universally optimal. While SVM was the strongest individual performer, its errors were highly correlated with ARIMA's. Second, and most importantly, explicitly modeling the dependence structure of these errors through a copula leads to a substantial improvement in forecasting accuracy. The success of the Gaussian Copula ensemble validates the central hypothesis of this research: that a hybrid model designed to leverage the strengths of diverse models while correcting for their shared weaknesses can provide a more powerful and robust tool for complex financial forecasting.

4. Conclusion

The volatile financial markets of developing countries like Kenya are a world far from what conventional forecasting models are able to represent. This study was able to successfully create a hybrid machine learning model, which offers a stronger solution to predict Treasury bill yields. By combining the features of ARIMA, SVM and FB Prophet in a copula-based framework, the model captures linear trend, non-linear dependence and seasonality in an effective way. The improvement in results with the hybrid model shown by a considerably smaller Mean Absolute Error, confirms the merit of such a combination approach. The results indicate that such a hybrid model can become an effective tool for investors, policymakers, and financial institutions to make better decisions and manage risk.

Compliance with ethical standards

Acknowledgments

I would like to thank all those people who in one way or another have helped to the successful completion of this work. First and foremost, I would like to appreciate Almighty God for the strength, endurance and wisdom in all my years of learning. I would like warmly and truly thank my supervisors of The Cooperative University of Kenya for their support, encouragement and ideas that assisted me in this job by offering valuable comments and encouragement before, during, and after this study. In conclusion, the support of my family and friends has been immobilizing; all your encouragement has made me to go on. This belief they have put in me has been my driving force in all the years of my academic pursuit.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Qing W. Treasury Bills. In: Dictionary of Contemporary Chinese Economics [Internet]. Springer, Singapore; 2025 [cited 2025 Aug 30]. p. 1811–3. Available from: https://link.springer.com/rwe/10.1007/978-981-97-4036-9_917

[2] Treasury Bills | CBK [Internet]. [cited 2025 Feb 19]. Available from: https://www.centralbank.go.ke/bills-bonds/treasury-bills/

[3] Sunki A, SatyaKumar C, Surya Narayana G, Koppera V, Hakeem M. Time series forecasting of stock market using ARIMA, LSTM and FB prophet. Satyanarayana K, Bobba PB, Perveen A, Debnath S, editors. MATEC Web Conf. 2024;392:01163.

[4] Mogaka FO. Effect Of Treasury Bill Rates On Capital Market Development: A Case Of Nairobi Securities Exchange [Internet] [Thesis]. University of Nairobi; 2019 [cited 2025 Aug 30]. Available from: http://erepository.uonbi.ac.ke/handle/11295/108707

[5] Jin YC, Cao Q, Wang KN, Zhou Y, Cao YP, Wang XY. Prediction of COVID-19 Data Using Improved ARIMA-LSTM Hybrid Forecast Models. IEEE Access. 2023;11:67956–67.

[6] Fama EF. Efficient Capital Markets: A Review of Theory and Empirical Work. J Finance. 1970 May;25(2):383.

[7] Malkiel BG. The Efficient Market Hypothesis and Its Critics. J Econ Perspect. 2003 Feb 1;17(1):59–82.

[8] Ariyo AA, Adewumi AO, Ayo CK. Stock Price Prediction Using the ARIMA Model. In: 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation [Internet]. 2014 [cited 2025 Aug 31]. p. 106–12. Available from: https://ieeexplore.ieee.org/document/7046047

[9] Vapnik VN. The Nature of Statistical Learning Theory [Internet]. New York, NY: Springer New York; 2000 [cited 2025 Feb 19]. Available from: http://link.springer.com/10.1007/978-1-4757-3264-1

[10] Van Gestel T, Suykens JAK, Baestaens DE, Lambrechts A, Lanckriet G, Vandaele B, et al. Financial time series prediction using least squares support vector machines within the evidence framework. IEEE Trans Neural Netw. 2001 July;12(4):809–21.

[11] Taylor SJ, Letham B. Forecasting at Scale. Am Stat. 2018 Jan 2;72(1):37–45.

[12] Sharma K, Bhalla R, Ganesan G. Time Series Forecasting Using FB-Prophet. 2022;

[13] Timmermann A. Chapter 4 Forecast Combinations. In: Elliott G, Granger CWJ, Timmermann A, editors. Handbook of Economic Forecasting [Internet]. Elsevier; 2006 [cited 2025 Aug 30]. p. 135–96. Available from: https://www.sciencedirect.com/science/article/pii/S1574070605010049

[14] Borrero JD, Mariscal J. Elevating Univariate Time Series Forecasting: Innovative SVR-Empowered Nonlinear Autoregressive Neural Networks. Algorithms. 2023 Sept 2;16(9):423.

[15] Luo Z, Jia X, Bao J, Song Z, Zhu H, Liu M, et al. A Combined Model of SARIMA and Prophet Models in Forecasting AIDS Incidence in Henan Province, China. Int J Environ Res Public Health. 2022 May 12;19(10):5910.

[16] Menculini L, Marini A, Proietti M, Garinei A, Bozza A, Moretti C, et al. Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices. Forecasting. 2021 Sept 15;3(3):644–62.

[17] Kemboi N, Yego NK, Kasozi J, Nkurunziza J. A Comparative Analysis of Machine Learning Models for the Prediction of Insurance Uptake in Kenya [Internet]. Preprints; 2023 [cited 2025 Feb 19]. Available from: https://www.authorea.com/users/604783/articles/634511-a-comparative-analysis-of-machine-learning-models-for-the-prediction-of-insurance-uptake-in-kenya?commit=ca181a5610d4400fb159b2579aad5b0522367774

[18] Van Vliet B. Abe Sklar's [Internet]. Rochester, NY: Social Science Research Network; 2023 [cited 2025 Feb 19]. Available from: https://papers.ssrn.com/abstract=4198458

[19] Dewick PR, Liu S. Copula Modelling to Analyse Financial Data. J Risk Financ Manag. 2022 Mar;15(3):104.

[20] Petropoulos F, Apiletti D, Assimakopoulos V, Babai MZ, Barrow DK, Ben Taieb S, et al. Forecasting: theory and practice. Int J Forecast. 2022 July;38(3):705–871.

[21] Kibet JJ. Modeling tail risks and systemic risks using copulas [Internet] [Thesis]. School of Mathematics, University of Nairobi; 2018 [cited 2025 Aug 30]. Available from: http://erepository.uonbi.ac.ke/handle/11295/104405