Data-Driven Approach to Forecasting Inflation in Bangladesh

Abstract—**The economy of Bangladesh has been on the rise over the past decade. One of the biggest factors contributing to the country's economic condition is the inflation rate. The increase in prices over a certain period is referred to as inflation. The project aimed to present a comprehensive analysis of forecasting the inflation rates in Bangladesh using Machine learning algorithms. The inflation rates are predicted using data over the last ten years. The data has been obtained from the Bangladesh Bureau of Statistics. This study predicts inflation rates based on the following factors: CPI Value, Currency against USD, Money Supply, Interest Rates, GDP (in BDT), and GDP Per Capita and compares the performance of numerous models, finding that XGBoost is the most efficient algorithm, with Mean Absolute Error (MAE) of 0.193 and Mean Squared Error (MSE) of 0.062. Similarly, the Random forest model also performed better with low evaluation metrics. The findings indicate the exciting prospects of machine learning approaches for economic forecasting, particularly in emerging nations like Bangladesh.**

Index Terms—**Money Inflation, Data Analysis, Machine Learning**

I. Introduction

The economic conditions and financial stability of a country are impacted by many factors. Bangladesh is a country that is steadily on the rise in terms of economy. But, one major factor affecting this growth is the yearly inflation rate. From the data obtained from the Bangladesh Bureau of Statistics, it is clear that the annual inflation rate is increasing rapidly. In 2024, the inflation rate reached over nine per cent. Therefore, monitoring and predicting future inflation rates is necessary. We use machine learning models to get the most accurate inflation forecasts. Machine learning can outperform regular econometric techniques [1]. We aim to improve on the existing works and produce even more accuracy. The project aims:

I. To find the future inflation rates using past data.

II. To find the best technique for inflation forecasting.

III. To help take precautions for future rise in prices.

Consumer price index (CPI): The consumer price index is the indicator of the average change in monthly prices paid by Bangladeshis. CPI, as further discussed in the paper is one of the major indicators of a country's inflation.

Currency against USD: The value of exchange rates has a huge influence on the inflation rates in a country. The rate of the US Dollar against the Bangladeshi Taka is steadily growing and has taken a big toll on the inflation rates in Bangladesh.

Money Supply (M1 and M2): The money supply is a big indicator of inflation. As the money supply grows faster compared to the economic growth, there is always a scope of increased inflation in the country.

Interbank rate: Interbank rate is another big factor influencing the inflation rates in Bangladesh. A low interbank rate can cause more demand. This, in turn will make people spend more and the inflation rate increases.

Gross Domestic Product: The project used both GDP in BDT and the GDP per capita. These factors show the change in the rate of inflation in the economy as a whole and also the change per capita.

Government spending: Government spending is directly proportional to the inflation rates. As the inflation rate rises, government spending also rises and vice versa.

The research used the above economic factors from the last ten years as the dataset for the project. Machine Learning models like, Random Forest Regressor, KNN, Support Vector Regressor (SVR), XGBoost, Decision Tree and Stacked ensemble model were used to forecast the inflation rate.

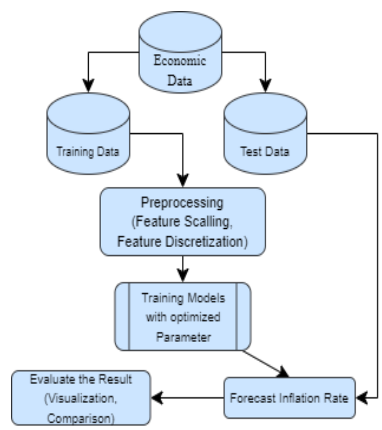

Fig. 1. Workflow Diagram

The research will compare the efficiency and accuracy of different models to provide the best possible prediction of future inflation rates. The following research paper will help to plan out future spending, living costs and many other factors to contribute positively to the economy of Bangladesh.

II. Literature Review

The research mainly considers papers that have done vital work in inflation forecasting using machine learning and deep learning. Most of the existing papers on inflation prediction have used time series models like ARIMA, SARIMA or AR to predict future inflation rates. With the advent of current trends in Machine Learning, researchers have started to use these methods to forecast inflation rates more accurately. In

2017, machine learning models like Lasso, adaptive Lasso, random forest and other models were used to predict inflation in Brazil [2]. Das et al. (2024) [3] found that Random forest and ANN models are more effective than time series or other linear models. They showed the need to use non-linear machine learning models to handle the unpredictability of feature variables. Araujo et al. (2023) [4] conducted the research with a large dataset, and the result showed that ML methods performed better in terms of mean-squared error than traditional econometric models. In Bangladesh, very few studies have been done on predicting inflation using ML models. The existing works have used very few models to predict inflation rates. Ismail et al. (2023) [5] found the AutoML model to be the best model for forecasting inflation rates in Bangladesh. Momo et al. (2021) [6] used SVR, RFR, decision tree, AdaBoosting, gradient boosting, and XGBoost for forecasting inflation, with AdaBoosting giving the best result. However, none of these studies use economic factors other than CPI (Consumer Price Index) for inflation forecasting. Acknowledging these limitations, our study aims to forecast the monthly inflation rates while considering other economic factors of Bangladesh.

III. Methodology

A. Dataset

The dataset obtained for this project was from the record contained in the Bangladesh Bureau of Statistics. The research used the following data to predict inflation: Consumer Price Index Value, Currency against USD, Money Supply M1, Money Supply M2, Interest Rates, Gross Domestic Product (in BDT), Interbank Rate, Government Spending and Gross Domestic Product Per Capita. The dataset used spans ten years. A total of 115 values have been used for each feature mentioned above.

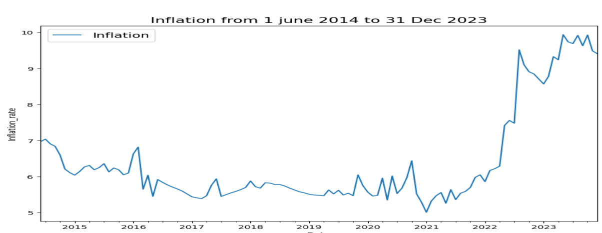

Fig. 2. Bangladesh Inflation Rate (6/1/2014-12/1/2023)

B. Empirical Models

1) Random Forest Regressor: A collection of different decision trees is called a Random Forest. This model is used with supervised machine learning where the target variable is labeled. Random forests also solve the problem of overfitting in decision trees by using the ensemble method. Random Forest was chosen because it gives good non-linearity, is lightweight computationally and prevents overfitting [7]. Economic data shows non-linear complex relationships between each other. These relationships cannot be captured by a simple non-linear model. Therefore, the Random forest model was chosen. *2) K-Nearest Neighbor Algorithm*: The K-Nearest Neighbor Algorithm is a model used to solve regression and classification problems. The K-Nearest Neighbor Algorithm finds the 'K' Nearest data points based on a given query point. Then, the model makes predictions based on the neighbors found. The Algorithm uses three different distance matrices to find the nearest points. Minkowski Distance is the one we used. The K-Nearest Neighbor Algorithm allows for the preservation of simplicity in the data. Therefore, this model has been used for predicting inflation [8].

3) Support Vector Regression (SVR): The Support Vector Regression Model is used for analyzing regression. SVR, along with proper tuning and optimization, is very useful for predicting inflation [9]. The SVR model finds an approximation between input and target variables that are continuous. The model does this while minimizing the prediction error.

4) XGBoost Algorithm: XGBoost or Extreme Gradient Boosting, is an algorithm which uses the principle of bagging. In this process, multiple decision trees are trained and the results are combined to find the optimum output. XGBoost is a great tool to predict inflation as it can get good results with non-linear data and is very robust [10].

5) Decision Tree: As the name suggests, a decision tree is a graphical representation of solutions. The branches of this tree are used to show the direction based on a decision and the nodes are used to show a decision. In this kind of project such as predicting inflation, a decision tree has the advantage of interpreting data intrinsically [11].

6) Stacking: Stacking is a process in which every model is used together weight-wise, which in turn, produces a new model. This process, therefore, produces better accuracy. For this project, we used the ensembled stack of three machine learning algorithms- Support Vector Regression (SVR), Random Forest Regressor and the K-Nearest Neighbor Algorithm. A stacking approach can help overcome the variance-bias problem while using machine learning algorithms separately to predict inflation [12].

IV. Experimental Analysis:

A. Dataset Preparation:

In this research, we predicted the Inflation rate based on different economic features. We obtained some data with yearly values (GDP, GDP per capita and Government Spending) and converted them into monthly values by dividing them by 12 for 12 months. Initially, we drew a correlation map by setting the threshold at 0.5 to drop features less important to our target variable (the inflation rate). In the beginning, we applied feature engineering techniques. Using the StandardScaler method ensured that each feature had a mean of zero and a standard deviation of 1. This is important for the proper functioning of an ML algorithm. Then another preprocessing technique, KBinsDiscretizer was used to convert continuous features into discrete bins. The use of categorical features in machine learning algorithms like Decision Tree and

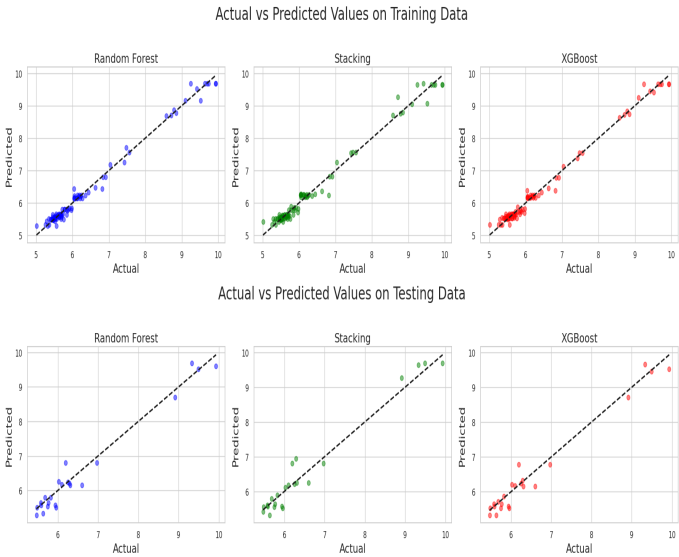

Fig. 3. Scatter plot displaying the performance of RFR, Stacked model, and XG-Boost. The diagonal line stands for the ideal prediction. All models performed very similarly.

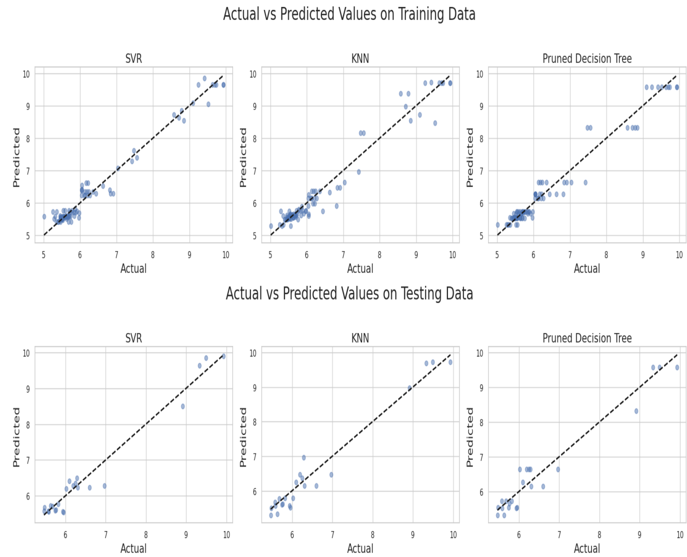

Fig. 4. Scatter plot displaying the performance of SVR, KNN, and Pruned-DT models. The wider scatter of points for the Pruned Decision Tree indicates higher prediction errors compared to SVR and KNN.

KNN significantly enhanced their performance by efficiently capturing patterns among data.

B. Models Implementation:

We applied the SVR model using The GridSearchCV, a hyperparameter optimization technique in ML. This was used to find the combination of parameters such as C, epsilon and kernel types (linear, rbf, polynomial and sigmoid), and optimize through a 5-fold cross-validation set using R^2 metric. We also applied the Random forest Regressor, Decision Tree (DT), KNN, Stacked ensemble model and XGBoost algorithm using standardized features obtained from the StandardScaler method, producing efficient performance without any hyper-parameter modification. For the KNN algorithm, we used the default parameter and achieved an effective solution. We also experimented with the decision tree pruning technique, which simplifies the decision tree by eliminating less informative parts, improving performance on unknown data and reducing overfitting. We pruned the DT with a maximum depth of 5, minimum samples per split of 10 and 5 samples per leaf to eliminate unwanted nodes. The Random forest regressor was applied with the default parameters. This model improved the prediction accuracy compared to a single decision tree, as it combines the predictions of multiple trees. We also applied the stacked ensemble model which combines SVR, Random Forest and KNN as base models. Gradient Boosting was used as the meta-learner. The stacked model improved prediction accuracy by utilizing the strengths of the base models. Finally, we implemented the XGBoost model with 100 trees and a 0.1 learning rate.

C. Result Analysis:

The research implemented six machine learning models to predict Bangladesh's inflation rate. To evaluate the models' performances, four important evaluation metrics were considered: R^2, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE). The seven economic factors used in this project have not been used together in existing works. This project contains numerically varying data for each factor; hence, there is a non-linearity or imbalance. Linear Machine Learning models could not comprehend the non-linearity of the dataset. Tables [I] and [II] demonstrate the models' overall performance in terms of training and testing data. The comparative analysis reveals that all the models were similarly effective. In training data, random forest regressor (RFR) performed best among these models in terms of R^2, MAE and MAPE. However, XGBoost models showed superior performance in test data and outperformed all other models in terms of the evaluation metrics used. XGBoost also achieved lower MSE for training data with almost the same performance as the random forest algorithm. Graphs were also plotted to visualize and compare the model performance on training and test data.

From Fig 3, it can be concluded that the Stacked model, XG-Boost and RFR models predicted the target most accurately. This is because their predicted values are aligned very close to the diagonal. In terms of generalization, XGBoost performed better. From Fig 4, we can see that SVR performed well for training data. However, the aforementioned models are more reliable, as SVR's plot shows more deviation of the predicted values from the diagonal. KNN, especially for larger values, also does not perform any better. The model with the lowest accuracy according to the plots is the Pruned Decision Tree, with the most noticeable deviations.

The deviation displays a higher percentage of error than SVR and KNN. The evaluation metrics also indicate the same. We also visualized model performances based on the residual

TABLE I
EVALUATION METRICS ANALYSIS ON TRAINING SET

Models	R^2	MAE	MAPE	MSE
RFR	**0.989**	**0.096**	**0.015**	0.018
SVR	0.970	0.160	0.025	0.049
KNN	0.953	0.188	0.027	0.079
Pruned-DT	0.962	0.182	0.027	0.064
Stacked EM	0.984	0.114	0.017	0.026
XGBoost	0.989	0.098	0.015	**0.017**

TABLE II
EVALUATION METRICS ANALYSIS ON TEST SET

Models	R^2	MAE	MAPE	MSE
RFR	0.964	0.200	0.030	0.065
SVR	0.962	0.206	0.030	0.070
KNN	0.956	0.228	0.035	0.080
Pruned-DT	0.945	0.268	0.041	0.100
Stacked EM	0.957	0.214	0.032	0.078
XGBoost	**0.966**	**0.193**	**0.029**	**0.062**

histogram. It's an effective tool for evaluating the performance of machine-learning models by showing skewness and bias. The normal distribution is an ideal distribution, which indicates that the model is capturing the fundamental trend in the data. Fig 5 revealed that the SVR model indicates a slightly right-skewed pattern, which means that it predicts the target values lower than the actual values. KNN also has a moderate right-skewed distribution, similar to SVR. This shape suggests that KNN may not operate consistently well in all conditions. The residual analysis suggests that Random Forest, Stacked model and XGBoost performed best among other models. They showed a very symmetric distribution and their distribution peak is centered near zero, indicating minimal bias and more accurate predictions. The Pruned-DT model might not be the best choice for this inflation problem as it tends to underestimate the target variable more than the other models.

V. CONCLUSION

This project analyzes the effectiveness of the various machine learning models in forecasting the inflation rate in Bangladesh. As the inflation rates of Bangladesh over the last 10 years do not show any seasonality or patterns, non-linear Machine Learning models were used to forecast inflation. After a comprehensive evaluation of six different models, XGBoost emerged as the most accurate predictor of the testing data, which indicates its ability to handle complicated economic patterns. It outperformed the other ML models for R^2 scores, MAE, MSE and MAPE in test data. However, the Random Forest model performed equally well with the training data, proving its endurance in learning from the patterns and securing the place of the best model in terms of all evaluation metrics. The findings of the research will help policymakers in making decisions on future spending, economic crises and other monetary.

In future works, authors can make use of even larger datasets by using data over a longer period and can implement hybrid models to analyze the patterns of changing inflation rates.

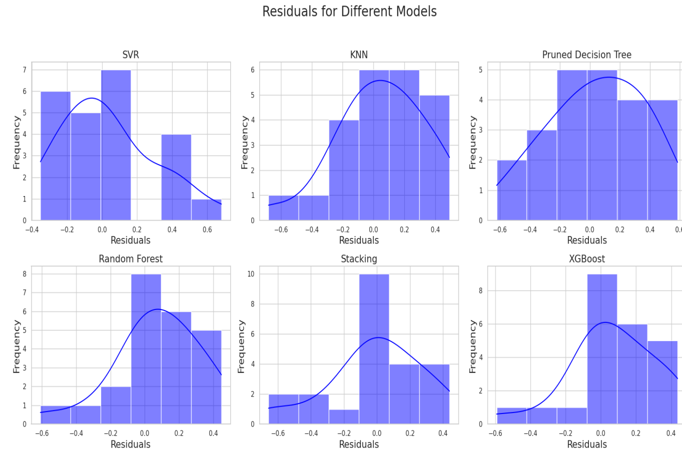

Fig. 5. Residual distributions for RFR, XGBoost and Stacked models demonstrate symmetric distributions, indicating improved predictive accuracy compared to other models.

REFERENCES

[1] Sendhil Mullainathan and Jann Spiess. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106, 2017.

[2] Márcio GP Garcia, Marcelo C Medeiros, and Gabriel FR Vasconcelos. Real-time inflation forecasting with high-dimensional models: The case of brazil. *International Journal of Forecasting*, 33(3):679–693, 2017.

[3] Pijush Kanti Das and Prabir Kumar Das. Forecasting and analyzing predictors of inflation rate: Using machine learning approach. *Journal of Quantitative Economics*, pages 1–25, 2024.

[4] Gustavo Silva Araujo and Wagner Piazza Gaglianone. Machine learning methods for inflation forecasting in brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, 4(2):100087, 2023.

[5] Labiba Binte Ismail, Mahira Ibnath Joytu, Tasnim Islam Plabon, and Muhammad Shafayat Oshman. Evaluation of machine learning models to forecast inflation: Bangladesh as a case study. In *2023 International Symposium on Networks, Computers and Communications (ISNCC)*, pages 1–6. IEEE, 2023.

[6] Shampa Islam Momo, Md Riajuliislam, and Rubaiya Hafiz. Forecasting of inflation rate contingent on consumer price index: Machine learning approach. In *Intelligent Computing and Innovation on Data Science: Proceedings of ICTIDS 2021*, pages 137–144. Springer, 2021.

[7] Todd E Clark, Seton Leonard, Massimiliano Marcellino, and Philipp Wegmüller. Weekly nowcasting us inflation with enhanced random forests. Technical report, Mimeo, 2022.

[8] Naima Khashimova and Manzura Buranova. Comparative analysis of machine learning algorithms for inflation rate classification and economic trend forecasting. In *Proceedings of the 7th International Conference on Future Networks and Distributed Systems*, pages 274–282, 2023.

[9] Erlin Mega Priliani, Anggyi Trisnawan Putra, and Much Aziz Muslim. Forecasting inflation rate using support vector regression (svr) based weight attribute particle swarm optimization (wapso). *Scientific Journal of Informatics*, 5(2):118–127, 2018.

[10] Yu-Shan Li, Ping-Feng Pai, and Ying-Lei Lin. Forecasting inflation rates be extreme gradient boosting with the genetic algorithm. *Journal of Ambient Intelligence and Humanized Computing*, 14(3):2211–2220, 2023.

[11] Codruț Ivașcu. Can machine learning models predict inflation? In *Proceedings of the International Conference on Business Excellence*, volume 17, pages 1748–1756, 2023.

[12] Emmanuel O Akande, Elijah O Akanni, Oyedamola F Taiwo, Jeremiah D Joshua, and Abel Anthony. Predicting inflation component drivers in nigeria: a stacked ensemble approach. *SN Business & Economics*, 3(1):9, 2022.