Remittance Patterns and Economic Development

Uche, Miracle Anih, Nathan Chizoba Christian, Increase

Abstract
Remittances have grown in importance as a source of foreign capital for developing nations, but their impact on economic growth and poverty is still being debated. This research investigated the relationship between remittance patterns and economic development across a wide range of countries using a machine learning approach. Using six datasets obtained from World Bank's website, the study conducted an Exploratory Data Analysis (EDA), followed by data preprocessing and then modeling using various machine learning algorithms to predict GDP in relation varying remittance patterns between 1970 and 2022. The study identified that remittance received was proportionally related to GDP, while there was no clear linear relationship between remittance and unemployment rate or population size. Migration patterns indicated that countries with larger populations experienced lower migration. Increased migration correlated with higher remittance received, and there was a positive relationship between net migration and GDP. Additionally, correlations between various features were examined, revealing moderate positive relationships between population, GDP, remittance volatility, and remittance received, while remittance paid exhibited strong correlation with GDP and weak correlation with net migration. Following the model development and evaluation, XGBoost demonstrated strong performance across metrics. SVR and SGD were competitive but slightly less accurate, while Random Forest lagged in prediction accuracy compared to the other models. These findings will help policymakers make better decisions and promote economic development in remittance-dependent areas.

1 Introduction

Remittances have grown in importance as a source of foreign capital for developing nations, but their impact on economic growth and poverty is still being debated (Ekanayake & Moslares, 2020). According to the World Bank, remittances were $583 billion in 2017 and $624.5 billion in 2018, making them the most important external financial source for many impoverished countries (Olayungbo & Quadri, 2019). Workers' remittances are now the second most important private external funding source for developing nations after foreign direct investment (Goschin, 2014). Remittance rise is related to increased immigration and technical developments (Meyer & Shera, 2017). Scholars acknowledge remittances as key sources of savings and money, enabling investments in health, education, and entrepreneurship and, as a result, fueling economic growth (Anton, 2010; Woodruff, 2007; Woodruff & Zenteno, 2007; Yang, 2008). Furthermore, remittances contribute to the growth of the banking system, private sector lending, and poverty alleviation (Aggarwal et al., 2011; Misati & Nyamongo, 2011; Adams & Page, 2005; Gupta, Pattillo, & Wagh, 2009).

Remittances' impact on poor-country economic growth is unknown, with contradictory findings in the literature (Kumar et al., 2018; Meyer and Shera, 2017; Nyamongo et al., 2012; Feeny, Iamsiraroj, and McGillivray, 2014; Lim and Simmons, 2015). Some studies demonstrate a favourable association between remittances and economic growth, while others show a negative or no relationship. This research tackled the methodological shortcomings and investigated new views on the influence of remittances on economic development, building on the work of Team Bayes (Eke et al., 2023). This research investigated the relationship between remittance patterns and economic development across a wide range of countries.

2 Literature Review

Previous research has explored the effect of remittance patterns on economic development. The study by Feeny, Iamsiraroj and McGillivray (2014) evaluated the relationship between remittance inflows and per capita income growth. The study focused on Small Island Developing States (SIDS) using econometric analysis of data for 136 developing countries, including 25 SIDS, for the period 1971 to 2010. The study found a positive, statistically significant association between per capita income growth in and remittances in the region. This association was such that the impact of a 10% increase in remittances to these countries yielded as much as two additional percentage points in economic growth. Additionally, the study found that Growth in Pacific SIDS was found to be substantially lower in the absence remittance inflows. These countries grew on average by 0.94 percent from 1971 to 2010. As revealed in the study, lower per capita income growth in SIDS did was not evident when there were remittance inflows. The study, however, was unable to show exactly how remittances inflow have driven higher growth in SIDS, or the channels or mechanisms through which this outcome has arisen. The lack of predictive model for predicting the future effect of remittance on per capita income is a notable research gap.

Additionally, Meyer and Shera (2017) used annual panel data from 1999–2013 to explore the impact of worker remittances on economic growth of Albania and five regional countries. Multiple Linear Regression was used to make a predictive analysis on the effect of workers' remittances on economic growth. Different diagnostic tests were applied to confirm the major assumption of multiple regression analysis like multicollinearity, heteroskedasticity and autocorrelation. Findings showed that worker remittances positively and significantly contribute to the economic growth of six countries. Also, the study highlighted that workers' remittances positively affected economic growth using multiple linear regression analysis. However, the research was limited to employing only multiple linear regression without considerations for other machine learning algorithms to ensure reliability of their findings, hence choosing the best model for prediction. Furthermore, this study failed to create a production platform for general use.

Furthermore, the study at Hamoye AI Lab by a group of machine learning experts known as Team Bayes (HDSC WINTER, 2023) was carried out to evaluate the effect of remittance patterns on economic development (Team Bayes, 2022). They analyzed the relationship between remittance patterns and economic development leveraging key economic indicators such as GDP. They further performed model evaluation on their dataset using popular machine learning algorithms. Their results showed that the Random Forest model performed better than every other algorithm used in minimizing loss. However, team Bayes' research had significant limitations stemming from insufficient data cleaning techniques, resulting in data loss due to the removal of null values. Inadequate data wrangling and transformation hindered insights from exploratory data analysis. The study introduced bias by focusing solely on remittance values related to 2022 GDP, ignoring temporal trends from 1990 to 2022, and excluding key indicators like unemployment rate and poverty. The use of suboptimal visualizations, lack of relationship evaluation, and absence of feature engineering and interactive tools further constrained the study's impact and suggested avenues for improvement.

This research aimed to build on the identified research gaps to deepen investigation into the relationship between remittance patterns and economic development across a wide range of countries.

3 Methodology

The study conducted an Exploratory Data Analysis (EDA), followed by data preprocessing and then modeling using various machine learning algorithms to predict GDP in relation varying remittance patterns.

Data Description: Six datasets were downloaded from the World Bank website for this study. This include Remittance Received (World Bank Data, 2023c), Remittances Paid (World Bank Data, 2023d), GDP (% of remittances received) (World Bank Data, 2023b), Unemployment rate (World Bank Data, 2023f), Net Migration (World Bank Data, 2023a) and Population (World Bank Data, 2023e) datasets. A common column across each of the dataset was the country name, while each of the datasets contained columns of their respective observations between 1960 to 2022. Each dataset had a complementary metadata of unique countries, their regions and income groups.

3.1 Data Cleaning

This had to do with wrangling, cleaning, removal of duplicate values and removal of outliers from the data. Firstly, in all datasets, the years between 1960 to 1969 contain null values for all the datasets except for population and net-migration dataset. As a result, they were completely dropped for the year range across all the datasets to ensure homogeneity in the datasets and reduce bias in our model performance. Hence, the study proceeded with the rest of the datasets containing the year duration between 1970 to 2022, which is a 52-year period. Furthermore, random null values within the remaining years were filled with zeroes. Subsequently, comparisons were made between the country columns in each dataset with their accompanied Metadata to remove elements that were not country names. Following the process, only countries that were sovereign nations were retained in the dataset. Additionally, the column names were duly corrected to have representative meanings.

3.2 Data Transformation

Given that the six datasets had a common country name column, a comprehensive dataset was created by merging all their columns. Firstly, using Python's melt function, the columns of each data was collapsed into four columns, including year, country name, country code, and the unique column for each data, which was either remittance paid, remittance received, unemployment rate, net migration, GDP, and population. Following that, all the datasets were completely merged into each other based on the common country name column using the merge function.

3.3 Feature Engineering

New attributes were generated during this process to gain a full understanding of the connection between remittance patterns and economic development, as well as to enhance the predictive power of our models. Novel features such as remittance growth rate, remittance per capita, remittance volatility, and remittance-to-GDP ratio. The remittance growth rate was calculated by computing the percentage change in inward remittances over time. The remittance per capita was calculated by dividing the total amount of remittances received in each country by their total population size. The remittance volatility was calculated by taking the variance or standard deviation of remittance amounts to quantify the volatility or fluctuations in remittance flows. The remittance-to-GDP ratio was calculated by dividing the total remittance amounts received by a country (inward remittances) and by the country's GDP.

3.4 Exploratory Data Analysis (EDA)

Most countries in the dataset fell under the category of high-income countries, whereas the low-income countries were the least in count, as shown in figure 1. There has been a gradual and continuous increase in remittance flow (in and out) between 1970 and 2022 across the countries, as shown in figures 2 – 5. Countries in Europe and Central Asia had a peak cumulative remittance inflow of 210.7691 billion in 2022 while countries in North America received least remittance 8.36078 billion in 2022, as shown in figures 2 and 4. On the other hand, Countries in Europe and Central Asia paid the highest remittance in 2021 in the figure of USD 195.7677 billion. They received USD 203.776 billion the same year. While countries in Sub-Sahara Africa paid least in the figure of USD 3.093646 billion in 2022 and received USD 51.78041 billion in the same year, as shown in figures 3 and 5. North America received the least remittance, which was 7.938835B, in contrast to what they paid in 2022, which was 87.4648B, as shown in figures 2 – 5. This implies that countries in North America have sent about a 1000% increment of what they received. By countries, the USA (North America) paid the highest remittance (figure 5) while India received the highest remittance (figure 6). Increase in remittance received is directly proportional with GDP, as indicated in figure 8. There was no clear linear relationship between the remittance received by the countries and their respective unemployment rate, as indicated in figure 9. Also, there was no clear pattern between the population sizes of the countries and their remittance flows, whether inflow or outflow, as shown in figures 10 and 11. Figure 12 showed the varying patterns between population sizes and migration patterns. Countries with higher population sizes tend to experience less migration compared to the countries with lesser population sizes. Figure 13 presents that increased migration leads to increased remittance received by the countries, whereas figure 14 shows a positive relationship between net migration and GDP. Figure 15 shows the correlation between the features. There is moderately positive relationship between population, GDP, remittance volatility, and remittance received by the country. Remittance paid demonstrated a strong relationship with GDP, and weak correlation with net migration. There was a weak but positive relationship between remittance per capita and remittance-to-GDP ratio. There was a weak but positive association between net migration and GDP and remittance paid.

3.5 Data Visualization

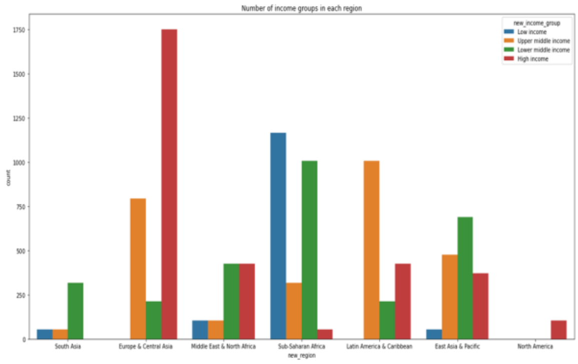

Figure 1. Number of income groups by region

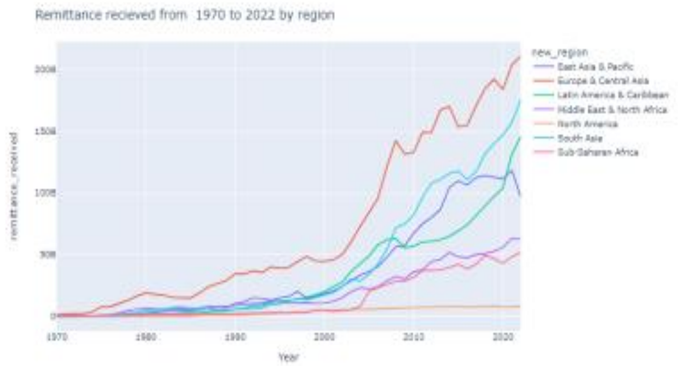

Figure 2. Remittances Received by different regions from 1970 to 2022

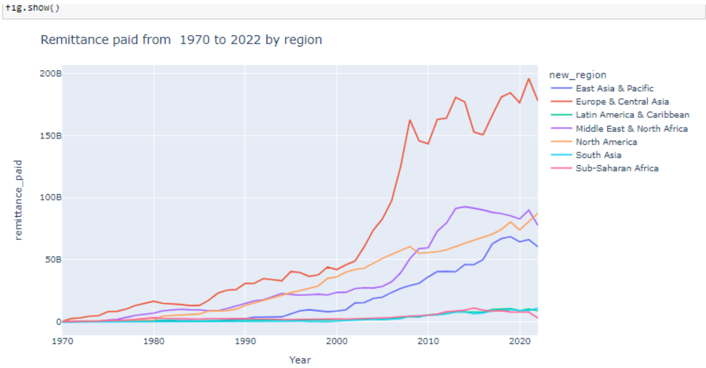

Figure 3. Remittances paid by different regions from 1970 to 2022

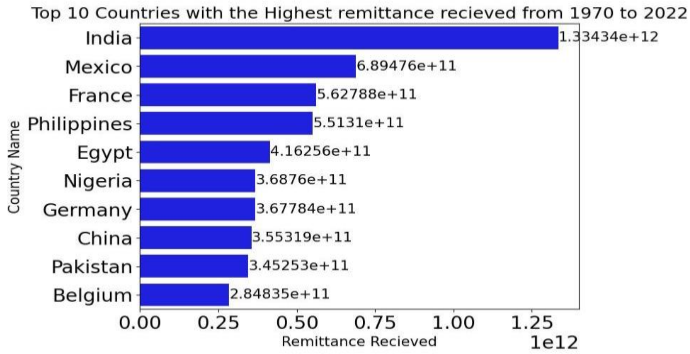

Figure 6. Top 10 Countries that Received Remittances from 1970 to 2022

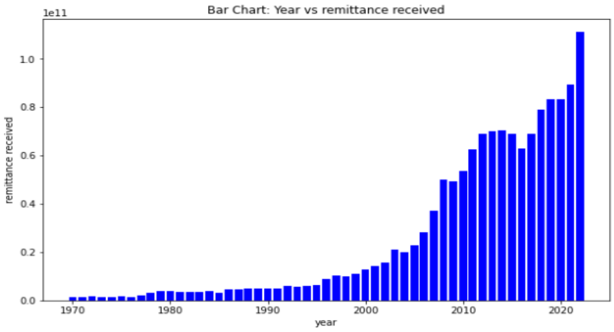

Figure 4. Incremental pattern of remittance received over time

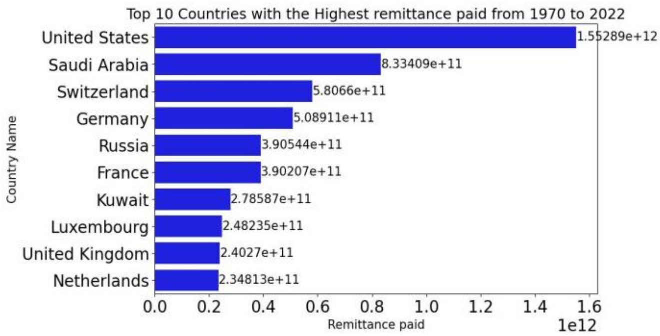

Figure 7. Top 10 Countries that Paid Remittances from 1970 to 2022

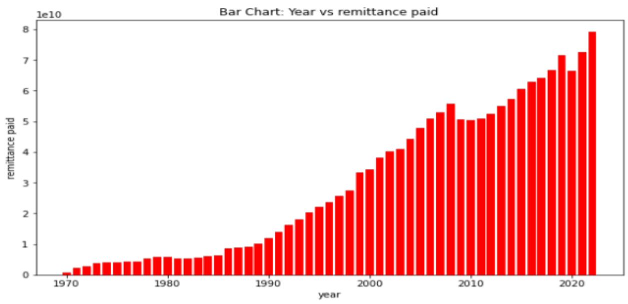

Figure 5. Incremental pattern of remittance paid over time

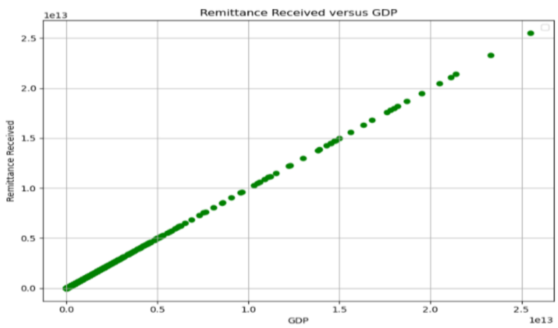

Figure 8. Positive relationship between remittance inflow and GDP

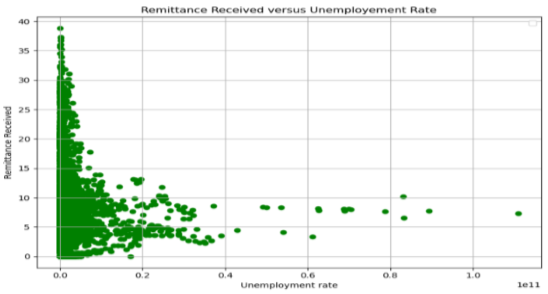

Figure 9. Relationship between remittance received and unemployment rate

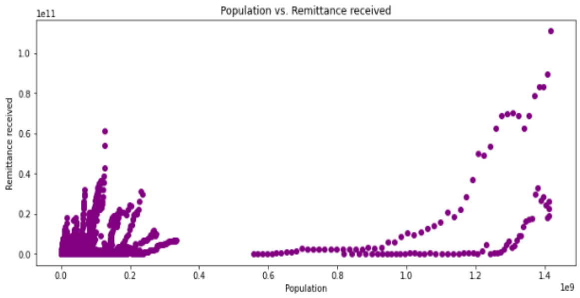

Figure 10. Relationship between remittances received and population size

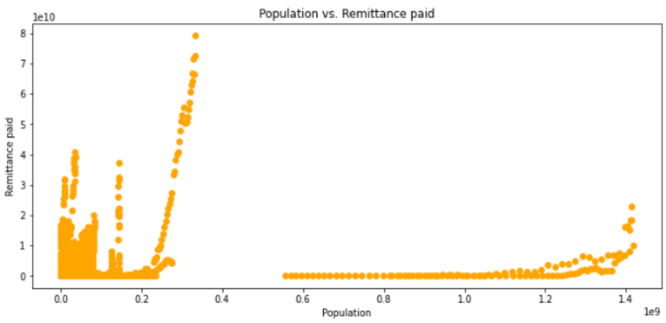

Figure 11. Relationship between remittance paid and population size

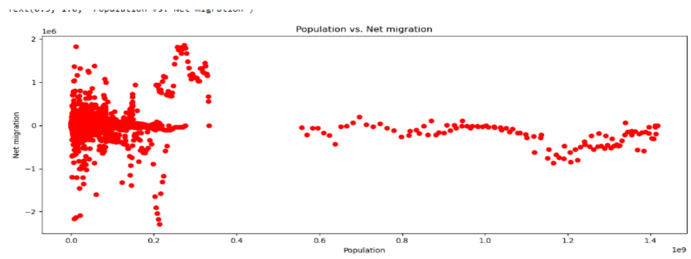

Figure 12. Relationship between net migration and population size

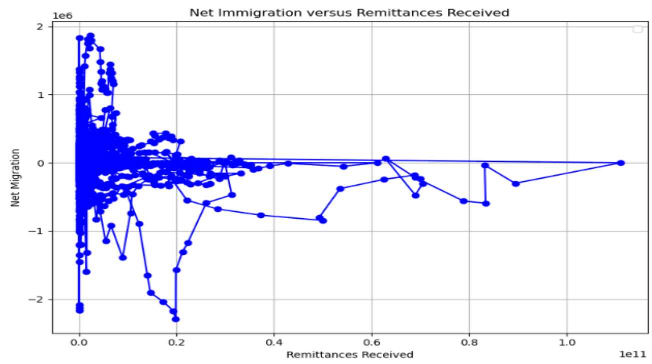

Figure 13. Relationship between remittances inflow and net migration

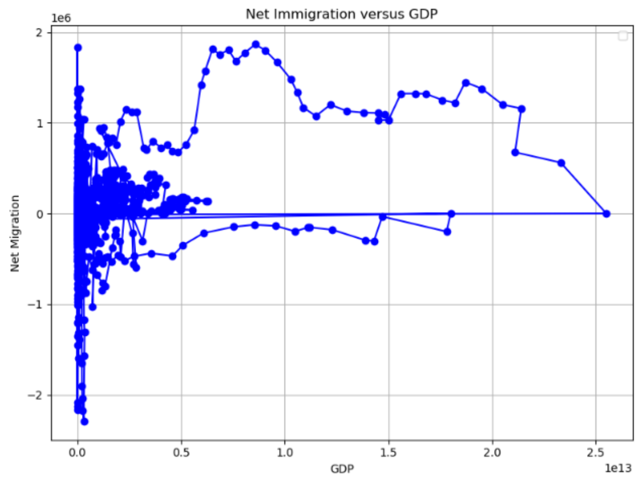

Figure 14. Relationship between GDP and net migration

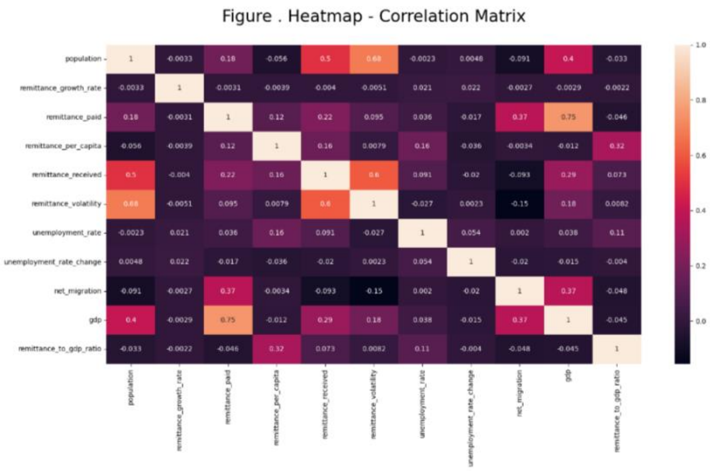

Figure 15. Correlation among the features in the dataset.

3.6 Modelling and Evaluation

Based on the results of the data analysis, a predictive model was developed using machine learning. This model was trained on historical data for the variables, and the accuracy was evaluated using various performance metrics.

3.6.1 Feature selection and segmentation

The process began by streamlining the dataset to remove redundancy. The 'country code' and 'year' columns were removed, which had no contribution to the data modelling. As a result, based on the EDA, features with higher correlation were selected for the modelling. Furthermore, the dataset was divided into two distinct sets of features, including numerical and categorical. This segmentation facilitated the preprocessing steps to each type of feature.

3.6.2 Standard Scaling of Numerical Features

Numerical features were standardized using standard scaling. This process ensured that all numerical features had a mean of zero and a standard deviation of one. Scaling was crucial to facilitate models like Support Vector Regression (SVR) and Stochastic Gradient Descent (SGD) that rely on distance metrics.

3.6.3 Binary Encoding of Categorical Features

Categorical columns were binary encoded to convert them into a numerical format suitable for machine learning models. This encoding method is particularly useful when dealing with high-cardinality categorical variables. Additionally, given that the country name column had 190 unique values, this was the best way of performing encoding on the data.

3.6.4 Data Splitting

Following the preprocessing phase, the data was partitioned into training and testing sets using the common split ratio where 70% was used for training and 30% used for testing. This separation ensured that our models learned from one portion of the data and was evaluated on novel data, gauging its real-world performance.

3.6.5 Model Selection

Given the inherent characteristics of our dataset, a regression model was built to evaluate the relationship between the variables. Four regression models were developing using the following algorithms, viz., random forest, a powerful ensemble learning method based on decision trees; XGBoost, an optimized gradient boosting algorithm known for its high performance; support vector regression (SVR), a regression technique that uses support vector machines to model relationships; and stochastic gradient descent (SGD), a simple and efficient optimization algorithm often used in linear regression. After the model training with the train data, their performance was evaluated the performance of various regression algorithms to determine the best fit for our prediction task.

3.6.6 Hyperparameter Tuning

Each regression model underwent a rigorous process of hyperparameter tuning to optimize their performance. A grid search cross-validation method was utilized to explore various hyperparameter combinations and select the best configuration.

3.6.7 Evaluation Metrics

To assess the predictive accuracy of the models, a range of evaluation metrics was utilized, including Root Mean Squared Error (RMSE), which measured the average deviation of predicted values from actual values, providing insight into the magnitude of errors; Mean Squared Error (MSE), which quantifies the average squared differences between predictions and actual values, emphasizing larger errors; and the Mean Absolute Percentage Error (MAPE), which calculates the average percentage difference between predicted and actual values and is useful for understanding the relative accuracy of predictions.

3.6.8 Model Deployment

After a comprehensive evaluation and the selection of the best model, the trained machine learning model was deployed using the Streamlit platform, a Python framework for generating interactive web apps. Because of the scalability and accessibility of this deployment, end users can produce real-time forecasts with the help of additional datasets. The model was serialised with Pickle and stored in ".pkl" format for easy loading and use in the web application. Users can enter their data, configure settings, and receive model predictions through the interactive Streamlit internet application with an easy-to-use interface. The application will be hosted on the GitHub platform to enable scalability, dependability, and worldwide access, allowing users from many locations to interact with the deployed model.

4 Results

Findings revealed that the XGBoost model performs very well across all metrics for both GDP and the unemployment rate predictions. The low values of MSE, MAE, RMSE, and the high R2 indicate accurate predictions and a good fit to the data. Concerning SVR and SGD, these models also show competitive performance with relatively low MSE, MAE, and RMSE values. The R2 values suggest that they explain a good portion of the variance, but they might not be as accurate as XGBoost. The Random Forest seems to perform less well than the other models. The higher MSE, MAE, and RMSE values indicate that its predictions are further from the actual values. However, the relatively high R2 for GDP prediction suggests that it captures a significant amount of variance in that context.

	Model			
Metrics	**XGBoost**	**SVR**	**SGD**	**Random Forest**
MSE GDP	3.9286723726077723e+21	0.1326770730871292	0.7590683979772747	1.7617888028636+22
MSE unemployment rate	35.68875996694728	0.3265895076598474	0.9192782196263962	6.85102210699681
MAE GDP	20935284409.64337	0.15540868432007296	0.3463495180672368	28709291610.964703
MAE unemployment rate	4.505499288708499	0.3155722250837432	0.7236693634433662	1.7257271686651291
RSME GDP	62679122302.46825	0.3642498772643983	0.8712453144650333	132732392537.14972
RSME unemployment	5.97400707027694835	0.5714621166499008	09587899768074322	2.6174457218816993
RSME unemployment rate	5.974007027694835	0.5714621166499008	0.9587899768074322	2.6174457218816993
R^2 GDP	0.9963363488194955	0.8601037266144557	0.1996347422942556	0.9835705830997394
R^2 unemployment rate	0.1335722989981264	0.6914211015116555	0.1313630406342149	0.1313630406342149
MAPE GDP	3.044560696873523e+24	0.805053256716889	3.47977176233185	1.7874883827059026e+24
MAPE unemployment rate	5957190470102564.0	2.5099678395662255	3.6646209975269	1428276726886811.2

5 Conclusion

This research comprehensively explored the interplay between remittance patterns and economic development across diverse countries. Through meticulous data analysis, including preprocessing and feature engineering, various machine learning models were employed to predict economic indicators like GDP and unemployment rate. The findings showcased the XGBoost model's outstanding accuracy, followed by competitive performances from SVR and SGD models. Although the Random Forest model exhibited weaker results, it still captured notable GDP variance. The study's outcomes offer valuable insights for informed decision-making in economic policy and development strategies, despite recognizing potential limitations. This research contributes to a nuanced understanding of the relationship between remittances and economic growth, paving the way for further refinements and advancements in predictive modelling and economic development strategies.

References

Aggarwal, R., Demirgüç-Kunt, A. & Peria, M. S. M. (2011). Do remittances promote financial development? *Journal of Development Economics*, 96(2): 255-264.

Anton, K. I. (2010). The impact of nutritional status of children in Ecuador. *International Migration Review*, 44, 269-299.

Ekanayake, E. M., & Moslares, C. (2020). Do remittances promote economic growth and reduce poverty? evidence from Latin American countries. *Economies*, 8(2), 1–26. https://doi.org/10.3390/ECONOMIES8020035

Eke, E. M., Akerekan, O., Buoro, O., Nsengiyumva, V., Jeffrey, O. & Uthman, A. (2023). Remittance patterns and economic development. *HDSC Winter '23 Capstone Project*, 1-10.

Feeny, S., Iamsiraroj, S. & McGillivray, M. (2014). Remittances and Economic Growth: Larger Impacts in Smaller Countries? *The Journal of Development Studies*, 50(8): 1055-1066.
Feeny, S., Iamsiraroj, S. & McGillivray, M. (2014). Remittances and Economic Growth: Larger Impacts in Smaller Countries? *The Journal of Development Studies*, 50(8), 1055-1066.

Goschin, Z. (2014). Remittances as an Economic Development Factor. Empirical Evidence from the CEE Countries. *Procedia Economics and Finance*, 10, 54–60. https://doi.org/10.1016/s2212-5671(14)00277-9

Gupta, S., Pattillo C. A., & Wagh, S. (2009). Effect of remittances on poverty and financial development in sub-Saharan Africa. *World development*, 37(1), 104-115.

Kumar, R. R., Stauvermann, P. J., Patel, A. & Prasad, S. (2018). The effect of remittances on economic growth in Kyrgyzstan and Macedonia: accounting for financial development. *International Migration*, 56(1), 95-126

Lim, S. & Simmons, W. O. (2015). Do remittances promote economic growth in the Caribbean community and common market? *Journal of Economics and Business*, 77, 42-59.

Meyer, D. & Shera, A. (2017). The impact of remittances on economic growth: An econometric model. *Economia*, 18(2), 147- 155.
Meyer, D. & Shera, A. (2017). The impact of remittances on economic growth: An econometric model. *Economia*, 18(2), 147- 155.

Misati, R. N. & Nyamongo, E. M. (2011). Financial development and private investment in sub-Saharan Africa. *Journal of Economics and Business*, 63(2): 139-151.

Nyamongo, E. M., Misati, R. N., Kipyegon, L. & Ndirangu, L. (2012). Remittances, financial development and economic growth in Africa. *Journal of Economics and Business*, 64(3), 240-260.

Olayungbo, D. O., & Quadri, A. (2019). Remittances, financial development and economic growth in sub Saharan African countries: evidence from a PMG-ARDL approach. *Financial Innovation*, 5(1), 1–25. https://doi.org/10.1186/s40854-019-0122-8

Team Bayes (2022). HDSC Winter Capstone Project Draft Article: Remittance Patterns and Economic Development. 1-10.

Woodruff, C. & Zenteno R. (2007). Migration networks and microenterprises in Mexico. *Journal of Development Economics*, 82(2): 509-528.

Woodruff, C. (2007). Mexican microenterprise investment and employment: The role of remittances. *Integration and Trade*, 11(27), 185-209.

World Bank Data (2023a). Net Migration https://data.worldbank.org/indicator/SM.POP.NETM

World Bank Data (2023b). Personal Remittances, Received (% of GDP) https://data.worldbank.org/indicator/BX.TRF.PWKR.DT.GD.ZS

World Bank Data (2023c). Personal Remittances, Received (Current US$) https://data.worldbank.org/indicator/BX.TRF.PWKR.CD.DT

World Bank Data (2023d). Personal, Paid (Current US$) https://data.worldbank.org/indicator/BM.TRF.PWKR.CD.DT

World Bank Data (2023e). Population, Total https://data.worldbank.org/indicator/SP.POP.TOTL

World Bank Data (2023f). Unemployment, total (% of total labor force) (modeled ILO estimate") https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS

World Bank (2019). Migration and Development Brief 31. Washington: The World Bank.

Yang, D. (2008). International migration, remittances and household investment: Evidence from Philippine migrants' exchange rate shocks. *The Economic Journal*, 118(528), 591-