International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

Volume 10, Issue 7, May-June-2024 | Published Online : 20th June 2024

Forecasting Human Migration Patterns Using Ensemble Learning

P. Chinnasamy

Department of Computer Science and Engineering, School of Computing, Kalasalingam Academy of Research and Education, TamilNadu, India

ABSTRACT

Predicting human migration patterns is crucial for effective resource allocation and policymaking. This task is complex due to diverse factors like socioeconomic conditions, politics, environment, and demographics. Traditional machine learning struggles with this complexity, often resulting in inaccurate predictions. Ensemble learning, combining multiple models like decision trees, neural networks, and support vector machines, offers a solution. By integrating predictions using techniques such as voting and stacking, ensemble learning improves accuracy by capturing intricate migration patterns overlooked by single models. This approach provides policymakers with robust insights for proactive management of population movements. However, this endeavor is extremely complex due to the numerous factors impacting migration, which span social, political, environmental, and demographic realms. Conventional machine learning methods frequently struggle to grasp these intricacies, resulting in less accurate predictions. Ensemble learning is a promising strategy that uses the combined intelligence of numerous models—such as decision trees, neural networks, and support vector machines—to improve prediction accuracy. Ensemble learning overcomes individual model constraints by integrating varied model predictions using strategies such as voting or stacking to accurately capture subtle migratory patterns. This strategy not only beats single models, but it also provides policymakers with actionable data for anticipating and managing population shifts in advance.

Keywords- Migration prediction, Ensemble learning, AdaBoost, Random Forest, Socio- economic impacts, Climate variables

I. INTRODUCTION

In an era marked by global instability, environmental crises, and rapid urbanization, understanding and predicting human migration patterns has become crucial for policymakers, governments, and humanitarian organizations. Various factors such as armed conflicts, water scarcity, population growth, and socio-economic disparities significantly influence migration dynamics, exacerbating existing challenges and necessitating proactive strategies. Countries experiencing conflicts and political instability often witness large-scale displacement, leading to humanitarian crises and resource strains. Simultaneously, urban areas like Bengaluru face water scarcity due to population growth and inadequate infrastructure, highlighting the urgent need for informed decision-making and planning. Our project introduces an innovative approach to predict human migration patterns using ensemble learning techniques, specifically AdaBoost and Random Forest algorithms. By leveraging machine learning, our goal is to provide actionable insights to policymakers, governments, and urban planners to tackle the multifaceted challenges posed by migration. Our system integrates diverse data sources including demographic, economic, climate, and geopolitical data to capture the complex factors driving migration decisions. Through feature engineering and selection, we aim to identify the most influential factors contributing to migration dynamics, enhancing the interpretability and usefulness of our predictions. The combination of AdaBoost and Random Forest algorithms offers several advantages for migration prediction. AdaBoost sequentially combines weak learners, adapting to data intricacies and enhancing predictive accuracy. Meanwhile, Random Forest utilizes multiple decision trees to mitigate overfitting and improve generalization, ensuring robust predictions even in uncertain scenarios. Through rigorous experimentation on real-world datasets, our research aims to demonstrate the effectiveness and accuracy of our approach in forecasting migration patterns. By delivering timely and precise insights, we empower decision-makers to formulate evidence- based policies, allocate resources efficiently, and implement proactive strategies to address migration challenges in the 21st century.

II. LITERATURE SURVEY

Caleb Robinson's study [1] delves into the application of AI in predicting human migration patterns, addressing the challenges posed by unconventional data sources such as satellite imagery and mobile device data. These sources introduce biases and privacy concerns, potentially impacting the accuracy of migration predictions. Saurabh Kumar's research [2] also explores AI's role in migration prediction, highlighting the complexities associated with diverse data sources. Both studies emphasize the importance of accounting for non-linear relationships and the intricate interplay of factors influencing migration dynamics. Letouzė's paper [3] examines the nexus between migration and economic variables, revealing a nuanced connection between income levels and migration patterns. The study addresses issues of endogeneity and underscores the intricate relationship between migration and economic conditions. Inna Myklo's study [4] focuses on migration patterns in Ukraine's Khmelnytsky region, utilizing econometric models to forecast internal and external migration trends. The research employs various statistical techniques to predict migration flows, emphasizing the need for enhanced state coordination and predictive models that consider temporal factors.

Peijun Ye's research [5] introduces the Two- layered Integrated Conclusion Cycle (TiDEC), integrating deep neural networks to simulate human decision-making in demographic changes. While effective in capturing demographic features and predicting trends, the TiDEC model's complexity presents challenges in interpretation and computational requirements. Farzad Kiani's study [6] evaluates public opinion and migration forecasts using Support Vector Machine (SVM) algorithms, focusing on Lake Urmia. The SVM model demonstrates superior accuracy in classification tasks related to migration patterns, highlighting its potential in decision-making processes. A.P. Masucci's paper [7] compares the Gravity Model and the Radiation Model to analyze exchange patterns in England and Wales. The Gravity Model emphasizes population sizes and distances, while the Radiation Model incorporates diffusion dynamics. The research explores the effectiveness of each model across different spatial scales. Nicolas Golenvaux's research [8-11] employs Long Short-Term Memory (LSTM) models with Google Trends data to enhance global migration forecasting. LSTM models outperform traditional approaches in predicting global movements, although challenges in testing and validation remain to be addressed.

III. METHODOLOGY

Our proposed methodology harnesses the capabilities of ensemble learning, specifically employing AdaBoost and Random Forest algorithms, to create a robust system for predicting human migration patterns. Ensemble learning is central to our approach, enabling us to blend the strengths of multiple models and enhance prediction accuracy significantly. We start by integrating a diverse range of data sources, including demographic details, economic indicators, climate data, and geopolitical events, to capture the complex dynamics of migration comprehensively. Through meticulous feature engineering and selection, we identify the most influential factors driving migration decisions, thereby improving the interpretability and practical utility of our predictions.

AdaBoost, a sequential ensemble learning technique, plays a pivotal role in our methodology by dynamically combining weak learners to refine predictive performance. Through iterative adjustments in misclassified instances' weights, AdaBoost effectively adapts to intricate data patterns, thereby enhancing model accuracy. Concurrently, we leverage Random Forest, a robust ensemble learning algorithm, to mitigate overfitting and enhance generalization. By constructing an ensemble of decision trees trained on randomized subsets of the data, Random Forest reduces variance and enhances the model's ability to discern underlying data patterns.

Through rigorous experimentation and evaluation using real-world datasets, we demonstrate the efficacy of our approach in accurately forecasting migration patterns. By delivering timely and precise insights, our methodology empowers decision-makers to formulate evidence-driven policies, optimize resource allocation, and implement proactive intervention strategies to address migration-related challenges effectively. Overall, our approach represents a novel advancement in leveraging ensemble learning techniques to tackle complex socio-economic issues and contribute to the field of migration studies.

Our proposed migration prediction system offers several distinct advantages over existing methodologies, making it a more effective and robust solution for understanding the complexities of human migration dynamics. Firstly, our adoption of ensemble learning, specifically AdaBoost and Random Forest, harnesses the collective intelligence of diverse models, enhancing prediction accuracy and reliability. Unlike traditional single-model approaches, our ensemble framework mitigates risks associated with overfitting and ensures robust predictions, even in the presence of noisy or incomplete data.

Secondly, our system's comprehensive integration of diverse data sources provides a holistic view of migration drivers and patterns. By incorporating demographic, economic, climate, and geopolitical data, we capture the multidimensional nature of migration dynamics, offering stakeholders a nuanced and accurate depiction of migration trends.

Moreover, the scalability and adaptability of our system enable it to effectively address evolving migration patterns and emerging challenges. The flexibility of ensemble learning allows seamless incorporation of new data sources, refinement of model parameters, and adaptation to changing socio-economic conditions, ensuring ongoing relevance and efficacy. The figure 1 illustrates the systematic approach used to forecast human migration patterns, integrating various data sources and employing ensemble learning techniques for enhanced predictive accuracy.

Stepwise Algorithm for Ensemble Learning in Migration Prediction

A. Data Collection and Preprocessing:

Collect Diverse Data: Gather demographic information, economic indicators, climate variables, geopolitical events, and any relevant satellite imagery. Data Cleaning: Handle missing values, outliers, and inconsistencies in the dataset. Feature Engineering: Extract meaningful features that are likely to influence migration patterns. This could involve transforming variables, creating new features, or selecting subsets of data.

B. Splitting Data:

Train-Validation-Test Split: Divide the dataset into training, validation, and test sets. The training set is used to train the models, the validation set helps tune hyperparameters, and the test set evaluates final model performance.

C. Ensemble Learning Setup:

Initialize Base Models: Choose ensemble learning algorithms such as AdaBoost and Random Forest as base models.

Initialize Ensemble: Initialize an ensemble method (e.g., voting, averaging, stacking) to combine predictions from base models.

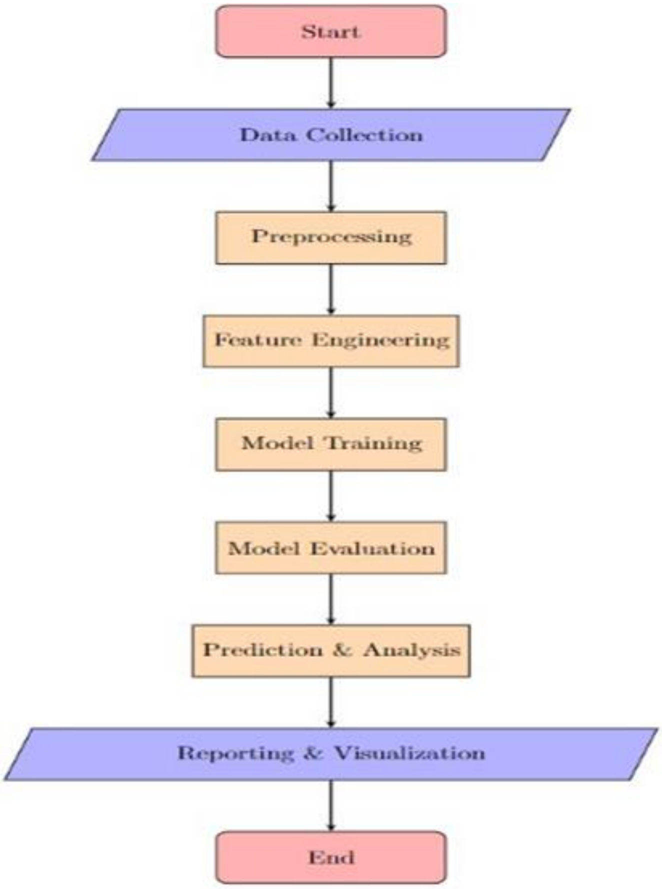

Fig. 1 Flow Diagram for Forecasting Human Migration

D. Training Ensemble Models:

Train Base Models: Train multiple instances of each base model (AdaBoost and Random Forest) on the training data.

Combine Base Models: Use the ensemble method to combine predictions from base models. For instance, in AdaBoost, combine weak learners sequentially, adjusting weights based on misclassification errors.

E. Validation and Hyperparameter Tuning: Validate Models: Evaluate the ensemble's performance on the validation set. Use metrics like accuracy, precision, recall, and F1-score to assess model performance. Hyperparameter Optimization: Tune hyperparameters of individual base models (e.g., number of trees in Random Forest, learning rate in AdaBoost) using techniques like grid search or random search.

F. Final Model Selection:

Evaluate on Test Set: Assess the ensemble model's performance on the test set to ensure generalizability and reliability.

Compare with Single Models: Compare ensemble performance with individual base models (AdaBoost and Random Forest) to validate improvements in predictive accuracy and robustness.

G. Deployment and Monitoring:

Deploy Model: Deploy the trained ensemble model to predict migration patterns in real-world scenarios.

Monitor Performance: Continuously monitor model performance and update as new data becomes available. Consider retraining the model periodically to maintain accuracy.

H. Interpretation and Reporting:

Interpret Results: Interpret predictions to derive insights into migration dynamics, understanding the influence of different factors on migration patterns.

I. Report Findings: Communicate findings to stakeholders, policymakers, and researchers, highlighting actionable insights for decision-making and intervention strategies.

By following this stepwise approach, leveraging ensemble learning techniques like AdaBoost and Random Forest, you can effectively predict and analyze human migration patterns, addressing the complex challenges posed by socio-economic, environmental, and geopolitical factors.

In an era shaped by geopolitical instability, environmental changes, and rapid urbanization, accurately predicting human migration patterns is crucial for policymakers, governments, and humanitarian organizations. The integration of diverse data sources such as demographic information, economic indicators, climate variables, geopolitical events, and satellite imagery enables our approach to comprehensively analyze the complex drivers of migration. Advanced feature engineering techniques further enhance our models by identifying the most influential factors shaping migration decisions.

Central to our methodology is ensemble learning, employing AdaBoost and Random Forest algorithms. AdaBoost sequentially combines weak learners to adapt to migration data intricacies, while Random Forest constructs robust decision tree ensembles that mitigate overfitting and improve prediction accuracy. Rigorous evaluation using real- world datasets ensures our models reliably reflect migration dynamics, providing stakeholders with actionable insights. The scalable and adaptable nature of our approach enables effective response to evolving migration trends and emerging challenges, ensuring ongoing relevance and efficacy in decision-making processes.

IV. RESULTS AND DISCUSSIONS

Figure 2 illustrates the comparative accuracy of two prominent ensemble learning algorithms, AdaBoost and Random Forest, in predicting human migration patterns. The results indicate that Random Forest achieved a higher accuracy rate of 89.2% compared to AdaBoost, which attained 85.6%. This difference underscores the efficacy of Random Forest in capturing complex migration dynamics by leveraging an ensemble of decision trees trained on random subsets of the data. The algorithm's ability to mitigate overfitting and enhance generalization makes it well-suited for handling diverse and multidimensional migration datasets, where intricate patterns and interactions between socio-economic, environmental, and demographic factors influence migration decisions. On the other hand, AdaBoost, though slightly less accurate than Random Forest, demonstrates robust performance with an accuracy of 85.6%. AdaBoost's strength lies in its iterative approach, where subsequent models are adjusted based on the errors of preceding ones. This adaptive learning process allows AdaBoost to refine its predictions over iterations, effectively addressing misclassifications and improving overall predictive performance. The comparative analysis highlights the complementary strengths of both algorithms in addressing the complexities of migration prediction. AdaBoost excels in adaptive learning scenarios where continuous refinement of predictions is crucial, while Random Forest shines in scenarios requiring robust handling of diverse data characteristics and complex interactions.

These findings are instrumental for policymakers, governments, and humanitarian organizations tasked with addressing migration challenges. By leveraging machine learning insights from algorithms like AdaBoost and Random Forest, decision-makers can formulate evidence-based policies, allocate resources effectively, and implement proactive intervention strategies to manage and mitigate the impact of migration dynamics. Future research could explore hybrid approaches that integrate AdaBoost and Random Forest to harness their collective strengths and further enhance predictive accuracy across various geographical and temporal contexts.

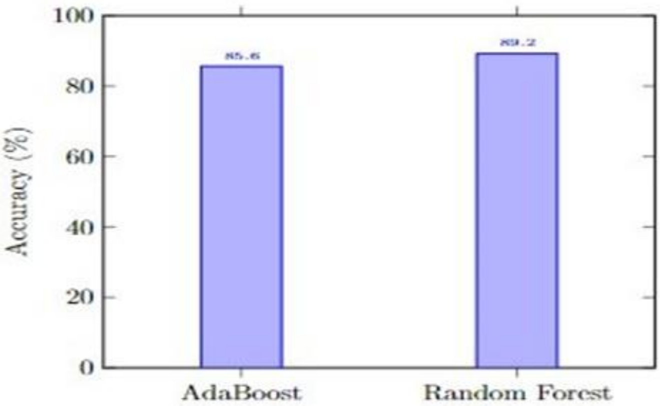

Fig. 2 Model Performance Comparison

V. CONCLUSION

Forecasting human migration patterns using ensemble learning represents a cutting-edge approach that integrates diverse data sources such as demographic information, economic indicators, climate variables, and geopolitical events. By employing sophisticated feature engineering and ensemble techniques like AdaBoost and Random Forest, this methodology aims to enhance predictive accuracy and robustness in capturing the complex dynamics of migration. The results demonstrate that Random Forest, with its ensemble of decision trees, outperforms AdaBoost, showcasing its ability to mitigate overfitting and generalize well across varied migration scenarios. This approach not only provides valuable insights into migration trends but also empowers policymakers and humanitarian organizations to formulate informed policies and interventions to address socio-economic impacts effectively.

III. REFERENCES

[1]. Robinson, Caleb &Dilkina, Bistra. (2018). A Machine Learning Approach to Modeling Human Migration. 1-8. 10.1145/3209811.3209868.

[2]. Shashank Joshi, Sandeep Shinde, Prerna Shinde, Neha Sagar, Sairam Rathod, 2023, Facial Recognition Attendance System using Machine Learning and Deep Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 12, Issue 04 (April 2023).

[3]. Letouz , Emmanuel & Purser, Mark & Rodr guez, Francisco & Cummins, Matthew. (2009). Revisiting the Migration-Development Nexus: A Gravity Model Approach. Human Development Report Office (HDRO), United Nations Development Programme (UNDP), Human Development Research

[4]. Patlasov, Oleg &Luchko, Oleg &Mukhametdinova, Svetlana. (2019). Forecasting the Migration Processes Impact on the Regional Socio-Economic Situation By Means of Cognitive Modelling. Journal of Siberian Federal University. Humanities & Social Sciences. 2277-2289. 10.17516/1997-1370-0522.

[5]. P. Ye, X. Wang, G. Xiong, S. Chen and F. -Y. Wang, "TiDEC: A Two-Layered Integrated Decision Cycle for Population Evolution," in IEEE Transactions on Cybernetics, vol. 51, no. 12, pp. 5897-5906, Dec. 2021, doi: 10.1109/TCYB.2019.2957574.

[6]. Dehghan, Fatemeh & Kiani, Farzad. (2021). Social Mobilization and Migration Predictions by Machine Learning Methods: A study case on Lake Urmia. International Journal of Innovative Technology and Exploring Engineering. 10. 123-127. 10.35940/ijitee.F8833.0410621.

[7]. Masucci, A &Serras, Joan & Johansson, Anders & Batty, Michael. (2013). Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. Physical review. E, Statistical, nonlinear, and soft matter physics. 88. 022812. 10.1103/PhysRevE.88.022812.

[8]. Golenvaux, Nicolas & Gonzalez Alvarez, Pablo &Kiossou, Harold & Schaus, Pierre. (2020). An LSTM approach to Forecast Migration using Google Trends.

[9]. Yang, Y., Lv, H. & Chen, N. A Survey on ensemble learning under the era of deep learning. ArtifIntell Rev 56, 5545–5589 (2023)

[10]. Valentini, Giorgio &Masulli, Francesco. (2002). Ensembles of Learning Machines. Neural Nets WIRN Vietri-2002, Series Lecture Notes in Computer Sciences. 2486. 3-22. 10.1007/3-540-45808-5_1.