Optimizing Random Forest Performance for Regional Gender Classification via Bayesian Optimization

Blerina Boçi[1]

Aurora Simoni[2]

1Department of Mathematics, Faculty of Information Technology
University "Aleksandër Moisiu", Durrës, Albania
2Department of Applied Mathematics, Faculty of Natural Science
University of Tirana, Albania

Received: 2 October 2025 / Accepted: 15 November 2025 / Published: 26 November 2025

Doi: 10.56345/ijrdv12n319

Abstract

This study investigates the application of Bayesian Optimization (BO) to enhance the performance of a random forest classifier (RFC) for gender-based regional classification in Albania. The model was applied to both the original demographic dataset and an extended simulated dataset that preserved the same structural relationships among demographic variables. The datasets included prefecture-level demographic indicators such as total population, male and female counts, year, and area. A unified machine learning pipeline was implemented, incorporating feature standardization, label encoding, and class imbalance correction through the SMOTENC technique embedded within cross-validation folds. Bayesian Optimization with BayesSearchCV was applied to automatically select the optimal hyperparameters of the random forest classifier, optimizing for the F1-macro score. The performance of the optimized RFC was compared with the default configuration using multiple evaluation metrics, including accuracy, balanced accuracy, F1-macro, precision, recall, and ROC-AUC, all supported by 95% bootstrap confidence intervals. The optimized model consistently outperformed the baseline on both datasets, achieving higher predictive accuracy and more balanced classification performance. Results from the original dataset demonstrated that the BO-tuned model provided significantly improved accuracy and sensitivity across classes, while the simulated dataset yielded even stronger generalization with improved stability and narrower confidence intervals. The comparison between real and simulated data confirmed that Bayesian Optimization effectively adapts model complexity to different data scales while preserving interpretability. These findings highlight the potential of BO as a reliable and automated hyperparameter tuning strategy for improving machine learning (ML) models in demographic and socio-economic studies, particularly when datasets are small, imbalanced, or limited in scope.

Keywords: *Gender dominance, class imbalance, classification, random forest classifier, Bayesian Optimization, SMOTENC*

1. Introduction

The population of a country is a key indicator of its social and economic development. Population dynamics are shaped by birth and death rates, migration patterns, and a complex interplay of demographic, economic, political, and cultural factors. Natural growth reflects the balance between births and deaths, while migration can substantially alter a country's demographic structure. Moreover, public policies, employment opportunities, living standards, and access to education play crucial roles in determining long-term population trends. This analysis is based on population change and gender distribution data provided by INSTAT, which offer valuable insights into Albania's demographic and socioeconomic

patterns.

In Albania, these factors have notably influenced population shifts since the early 1990s, a period marked by profound social and economic transformation (Lerch, 2016). Due to the patriarchal structure of Albanian society, emigration has predominantly involved men, while independent female migration has remained limited (Gëdeshi & King, 2018). This male-dominated outmigration has significantly affected the gender balance, with many women remaining in rural areas as men migrated abroad. Recent demographic analyses indicate that Albania is entering the final phase of its demographic transition, characterized by declining fertility, population aging, and rising divorce rates. As noted in the report, "Albania has entered the last stage of the demographic transition and/or the second demographic shift... fertility and marriage rates have declined, while divorces have increased" (Ármás, 2022).

Building upon these demographic patterns, this study applies a machine learning approach to classify regions according to gender dominance using structured demographic and geographic data. As a continuation of our previous work, which applied Bayesian Optimization to regression models (Boçi & Simoni, 2025), the present research extends its application to a classification problem, employing the same optimization strategy for gender dominance prediction.

The dataset includes variables such as prefecture, male and female population counts, year, and computed population density. The target variable is binary, indicating whether the male or female population is dominant in a given region. To address class imbalance, where one gender category occurs more frequently, we employ the synthetic minority oversampling technique for nominal and continuous features (SMOTENC), an extension of the SMOTE framework (Chawla et al., 2002). This method ensures that both classes are adequately represented during training by generating synthetic samples for the minority class while appropriately handling mixed categorical and continuous variables.

For the classification task, we utilize the random forest classifier, a robust ensemble learning method known for its accuracy and interpretability (Breiman, 2001). To further enhance performance and mitigate overfitting, Bayesian Optimization (BO) is incorporated for hyperparameter tuning, providing an efficient and probabilistic search for optimal model configurations (Snoek et al., 2012).

2. Related Work

Gender classification has long been a topic of significant research interest, owing to its relevance in enhancing face recognition systems (Moghaddam & Yang, 2002), and its broader applications in human–computer interaction, soft biometrics (Jain et al., 2004; Dantcheva et al., 2011), and beyond.

Some previous studies have focused on employing optimization techniques to fine-tune model hyperparameters and enhance performance, whereas others have primarily investigated the application of various machine learning classifiers for gender prediction tasks. In (Murtadha et al., 2017), the effectiveness of using grid search to optimize random forest parameters for gender classification tasks involving voice-based features was demonstrated, highlighting the critical role of hyperparameter tuning in improving model performance.

Previous research (Verma et al., 2019) has applied multiple machine-learning classifiers, including random forest (RF), support vector machine (SVM), and artificial neural networks (ANN), to predict student gender based on attitudes toward ICT and MT. The study employed techniques such as SMOTE and various cross-validation methods, with random forest achieving the highest accuracy.

Anime image classification is an emerging task in computer vision, particularly for gender recognition. One study (Mulyana & Pramansah, 2022) employed a random forest classifier in conjunction with Gray-Level Co-occurrence Matrix (GLCM) texture analysis to classify anime characters by gender. The model achieved high accuracy (95%), with the Homogeneity5 feature identified as the most relevant for improving performance. GridSearchCV was employed for hyperparameter tuning, improving the classifier's performance despite the inherent complexity and variability of anime image textures.

Another study (Thiojaya & Qomariyah, 2023) investigated gender classification based on human gait patterns using joint angles estimated through MediaPipe pose detection. The authors compared several machine learning algorithms, including support vector machines (SVM), Gaussian naive Bayes (GNB), random forest (RF), and XGBoost. Among these, the XGBoost model achieved the highest classification accuracy of 97%, demonstrating the effectiveness of gait-based analysis for identifying gender characteristics.

In (Ibarra-Vazquez et al., 2023), machine learning models were applied to forecast gender based on students' self-perceived competencies in open education, encompassing knowledge, skills, and values. The study employed cross-validation techniques and interpretable machine learning approaches to identify significant features, confirming that

student perceptions can effectively predict gender within educational contexts. Notably, the random forest model exhibited the highest accuracy, highlighting its reliability in capturing the relationship between open education competencies and gender. The use of interpretable models such as decision tree (DT), random forest (RF), convolutional neural network (CNN), and light gradient boosting machine (LGBM) facilitated the identification of key predictive features, providing valuable insights for educational strategies and enabling comparison between interpretable and non-interpretable approaches.

A recent study (Arif et al., 2024) applied the random forest algorithm for gender classification using ECG data from the ECG ID database, achieving the highest accuracy with filtered signals. The findings highlight the importance of noise reduction and hyperparameter tuning in improving classification performance. The study also provides valuable insights into the challenges and limitations of applying random forest to biometric data such as ECG for gender classification.

Unlike previous gender classification studies that primarily relied on biometric or image-based data such as handwriting, facial features, or gait patterns, the present research employs machine learning on regional demographic and geographic data. The dataset includes variables such as prefecture, year, male and female population counts, total population (MF), area, and population density. This approach introduces a novel perspective by leveraging structured demographic indicators rather than biometric information and integrating Bayesian Optimization to enhance both classification performance and model interpretability.

3. Methodology

This study is guided by two primary research objectives. The first objective is to evaluate the effectiveness of a random forest classifier (RFC) integrated within a preprocessing and resampling pipeline that includes feature standardization and the synthetic minority oversampling technique for nominal and continuous features (SMOTENC)**,** applied directly within cross-validation folds to prevent data leakage. The second objective is to assess how automatic hyperparameter tuning through Bayesian Optimization (BO), implemented via the BayesSearchCV framework, enhances model performance compared to the default random forest configuration.

Fig. 1 presents the schematic representation of the entire algorithmic workflow.

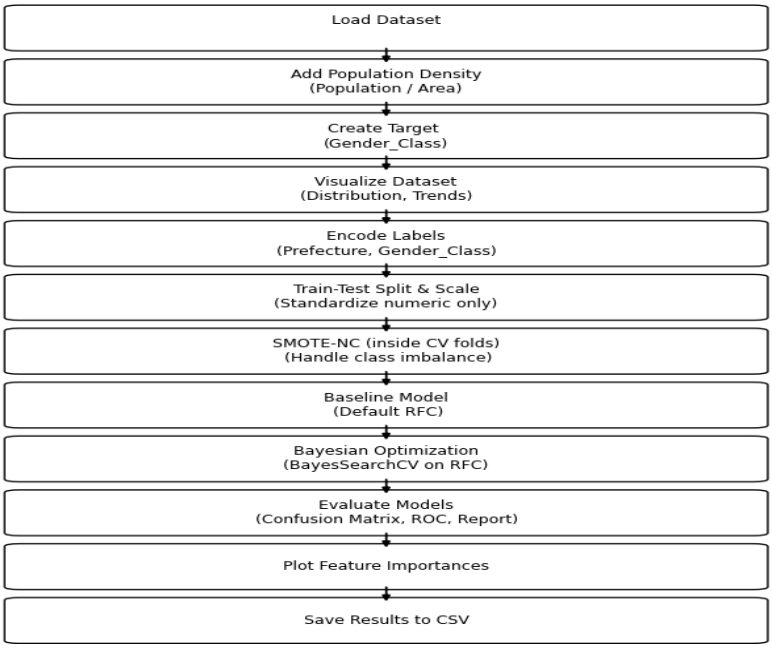


Figure 1. Schematic representation of the Random Forest–Bayesian Optimization–SMOTENC workflow

3.1 Data Description

A structured dataset containing demographic and geographic information for Albanian prefectures, spanning the years 2014 to 2025, is utilized in this research. The data, sourced from INSTAT Albania, includes annual records published on January 1st of each year, reflecting gender-based population distributions across regions. Annual estimation population is based on cohort component method. This consistent reporting framework allows for reliable year-over-year comparisons to analyse demographic trends and regional shifts.

The dataset includes the following variables:

- Prefecture: The name of the administrative region.
- Total population (MF): The combined number of males and females in the region. This was used to compute population density.
- Male population (M): The number of males in the region.
- Female population (F): The number of females in the region.
- Year: The calendar year of the record, ranging from 2014 to 2025.
- Area: The area of each region in square kilometers, used to calculate population density.

Table 1. Summary of demographic indicators by prefecture

Prefecture	MF	M	F	Year	Area (km^2)
Berat	145097	73502	71595	2014	1798
Dibër	138321	70919	67402	2014	2586
…	…	…	…	…	…
Tiranë	759981	367099	392882	2025	1652
Vlorë	143999	71827	72172	2025	2706

Overall, the dataset presented in Table 1 provides a regional-level overview of gender distribution across Albania, forming the foundation for training and evaluating the gender dominance classification model.

3.2 Data Visualization

Figure 2 illustrates population trends by gender across Albanian prefectures from 2014 to 2025, using data obtained from INSTAT.

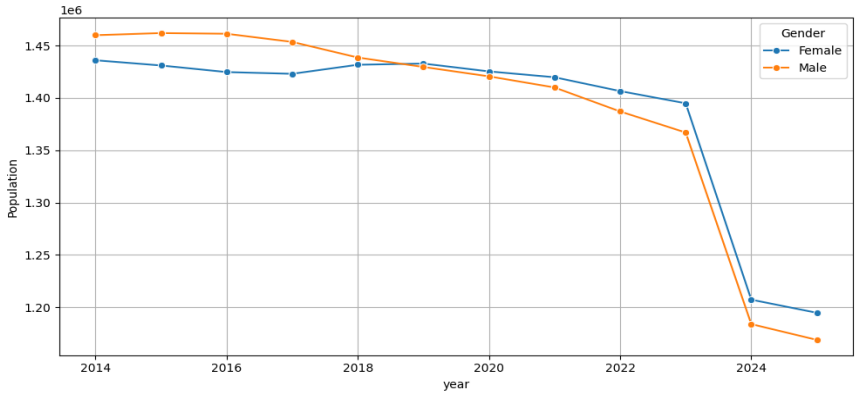


Figure 2. Population trends by gender over time in Albania (2014-2025)

Understanding regional gender distribution is crucial for informing public policy, guiding resource allocation, and supporting social planning. According to INSTAT, as of January 1st, 2025, Albania's population was recorded at

2,363,314, representing a 1.2% decrease compared to the previous year. This decline reflects a continuation of the overall downward trend observed throughout the study period, indicating ongoing demographic changes that may affect gender distribution across regions and require appropriate policy responses.

Figure 3 presents a comparative analysis of the male and female population across Albanian prefectures for the years 2014 and 2025. Each prefecture displays four bars, representing male and female populations in both years. Prefectures like Tirana continue to hold the highest population density, while regions such as Kukës and Gjirokastër show consistently lower numbers.

The dataset includes 86 male-dominant and 58 female-dominant instances, indicating a higher prevalence of male dominance within the sample, Figure 4. However, as the analysis covers data across several years, this observed dominance may not fully represent broader regional gender trends, where females sometimes outnumber males. This highlights the importance of considering temporal and local demographic variations when interpreting the results.

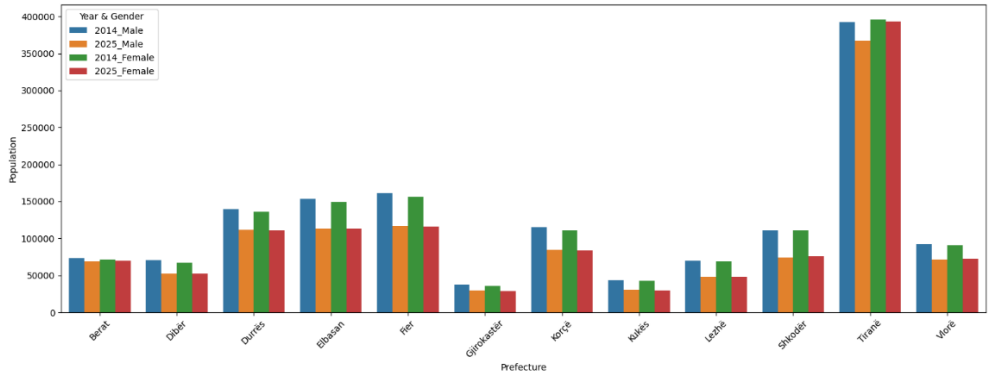


Figure 3. Comparison of male and female population by prefecture in 2014 and 2025.

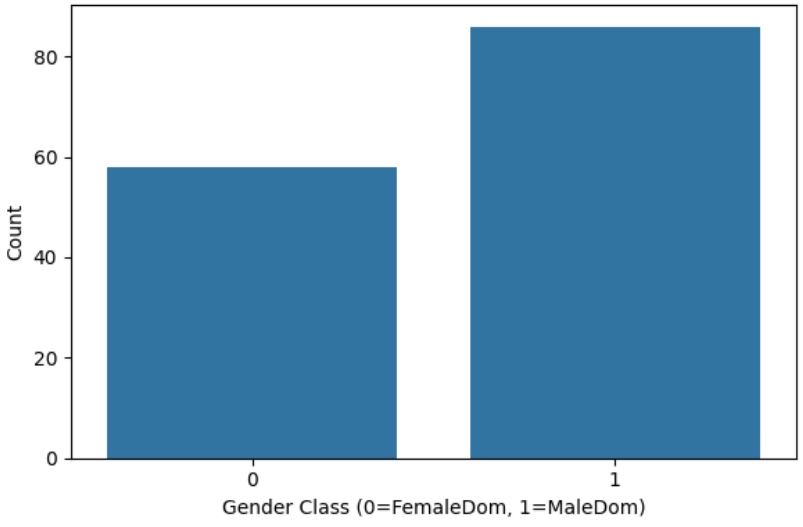


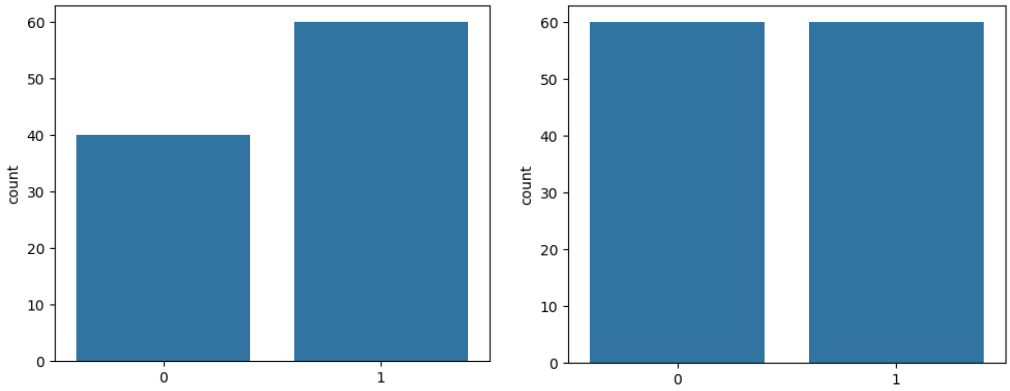
Figure 4. Gender class distribution in the entire dataset

3.3 Data Processing and Splitting

A derived feature, Population_Density, was computed to represent the spatial density of the population, which may reflect migration trends or urbanization patterns. The target variable (Gender_Class) was defined according to gender dominance: regions where the male population exceeded the female population were labeled male dominant (1), while the remaining regions were labeled female dominant (0). Numeric features (M, F, Year, and Population_Density) were standardized using StandardScaler, whereas the categorical feature (Prefecture) was label-encoded and retained within the ColumnTransformer. The target variable was also categorical and encoded separately using LabelEncoder. To prevent data leakage, the SMOTENC resampling method was integrated into the machine learning pipeline, ensuring that oversampling was applied only to the training data during cross-validation. The dataset was then divided into training (70%) and testing (30%) subsets using stratified sampling to maintain class proportions. This preprocessing and splitting strategy ensured standardized numeric features, preserved categorical information, and maintained a balanced class distribution for model training and evaluation. Given the limited size of the original dataset, a simulated dataset was also generated to assess the pipeline's robustness and reproducibility under controlled conditions.

3.4 Class Imbalance Handling

During the initial model tuning phase, the classifier exhibited suboptimal performance, largely due to the imbalance between gender classes in the dataset and the limited amount of available data. To address this issue, SMOTENC (Chawla et al., 2002; Iscra et al., 2024) was applied to achieve a more balanced class distribution before model training. Figure 5 presents the class distribution before and after applying the SMOTENC technique, illustrating how the imbalance between classes was addressed.



a) b)

Figure 5. Class distribution before (a), and after (b) applying SMOTENC

3.5 Simulated Dataset

The original prefecture-level demographic dataset was extended to cover the years 2000-2080 by extrapolating linear trends of total population (MF) and male proportion (M/MF) for each prefecture. Random noise proportional to the residual variance of the fitted models was added to preserve realistic fluctuations, while the area of each prefecture remained constant. The final simulated dataset contained 972 yearly records across 12 prefectures. Figure 6 shows the overall trend and distribution of the simulated dataset used in this study, illustrating how the generated data replicate the structural patterns of the original dataset.

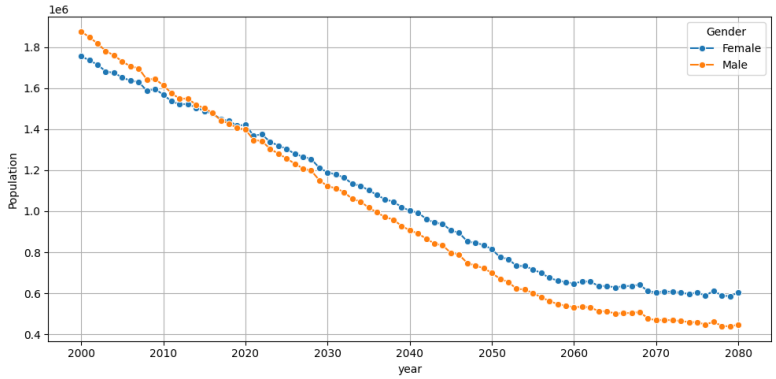


Figure 6. Simulated population trends by gender in Albania (2000–2080)

3.6 Random Forest Classifier

Random forest is an ensemble learning method that constructs multiple decision trees using resampled subsets of the training data. Each tree is trained on a different subset obtained through bootstrap sampling, and the final prediction is made by majority voting for classification tasks or averaging for regression tasks, thereby enhancing both accuracy and robustness (Breiman, 2001). It is a well-known algorithm that often delivers strong performance even with default settings (Probst et al., 2019).

3.7 Hyperparameter Optimization

During model training, Bayesian Optimization was implemented using the BayesSearchCV framework to automatically tune the key hyperparameters of the random forest classifier. The optimization process employed a 10-fold stratified cross-validation strategy to ensure balanced evaluation of both gender classes (Kohavi, 1995). The F1-macro score was used as the optimization objective, as it provides a balanced assessment of performance in imbalanced datasets.

Table 2. Default and optimized BO ranges for the Random Forest Classifier

Hyperparameters	Default RFC	Parameter space (BO – original dataset)	Parameter space (BO–simulated dataset)
n_estimators	100	(60, 200)	(100, 800)
max_depth	5	(3, 15)	(5 ,30)
min_samples_split	2	(2, 10)	(2, 50)
min_samples_leaf	1	(1, 10)	(1, 20)

Parameter optimization is an effective method to improve the accuracy of machine learning algorithms. Compared to grid search and random search (Jain & Gupta, 2022), Bayesian Optimization significantly reduces the number of evaluations required to find optimal solutions (Wu et al., 2019; Shahriari et al., 2016; Mesquita & Marques, 2024). Table 2 presents the default and optimized hyperparameter settings of the random forest classifier (RFC) obtained through Bayesian Optimization (BO) for both the original and simulated datasets.

For the simulated dataset, the parameter ranges were broadened to allow deeper tree structures and increased model flexibility, accommodating the larger sample size relative to the original dataset.

3.8 Evaluation Metrics

Model evaluation typically relies on the confusion matrix to assess performance on test data. The confusion matrix summarizes the prediction results with the following components:

- True Positives (TP): Correctly identified positive observations.
- True Negatives (TN): Correctly identified negative observations.
- False Positives (FP): Negative observations incorrectly classified as positive.
- False Negatives (FN): Positive observations incorrectly classified as negative.

To address class imbalance, standard accuracy can be an unreliable indicator of model performance, as it may overestimate performance on the majority class. Therefore, the random forest classifier was evaluated using multiple complementary metrics, including accuracy, balanced accuracy, precision (macro-average), recall (macro-average), F1 (macro-average), and the area under the receiver operating characteristic curve (AUC-ROC). Additionally, 95% bootstrap confidence intervals (CIs) were computed for each metric to assess the statistical robustness of the model's performance (Rink & Brannath, 2025).

The formulas for the evaluation metrics used are:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \quad (4)$$

Metrics like the F1 Score and AUC-ROC are preferred in such contexts because they provide a more balanced and informative evaluation for classification tasks involving class imbalance (Saito & Rehmsmeier, 2015).

4. Results and Discussion

4.1 Performance Comparison

The performance of the default random forest classifier (RFC) was compared with that of an optimized version trained using Bayesian Optimization (BO). In both configurations, SMOTENC was integrated within the pipeline to resample the minority class during each cross-validation fold, thereby addressing class imbalance and preventing data leakage. The baseline model employed the default hyperparameter settings, serving as a reference for assessing the improvements achieved through optimization. Conversely, the optimized model combined SMOTENC with Bayesian Optimization, implemented via BayesSearchCV, to automatically fine-tune the key hyperparameters. Both models were evaluated on the same original, imbalanced test set to ensure a fair and consistent comparison.

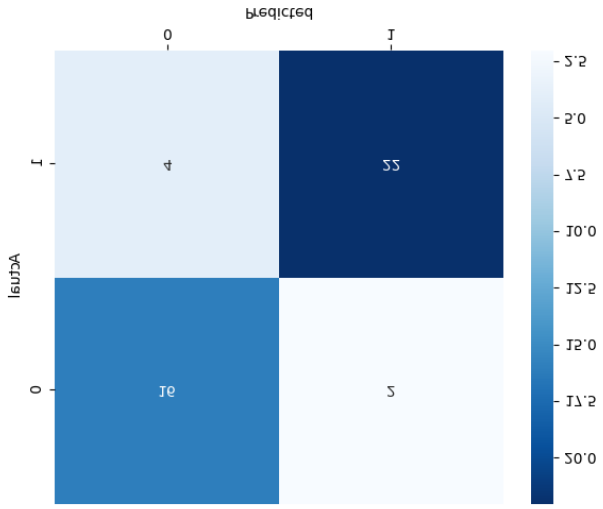


Figure 7. Confusion matrix of default model on test data

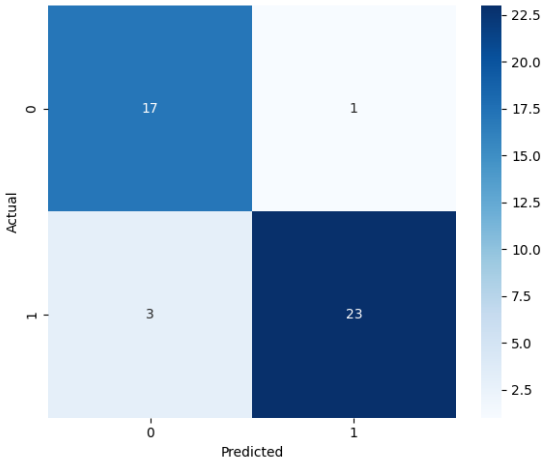


Figure 8. Confusion matrix of optimized model on test data

The confusion matrix for the default model reflects its baseline performance, Figure 7, while the matrix for the optimized model, tuned using Bayesian Optimization Figure 8, illustrates the improvements achieved through hyperparameter tuning.

The optimized model outperformed the default model across all key evaluation metrics, as shown in Table 3. The optimized model achieved an overall accuracy of 91%, representing an improvement from 86% obtained with the default configuration. Additionally, the error rate decreased from approximately 14% to 9%, demonstrating the effectiveness of hyperparameter tuning via BO in enhancing the model's predictive performance.

Table 3. Classification metrics comparison

Metric	Class	Default model	Optimized model
Precision	0	80%	85%
	1	92%	96%
Recall	0	89%	94%
	1	85%	88%
F1-Score	0	84%	89%
	1	88%	92%
Accuracy		86%	91%

Table 4. Bootstrapped test metrics

Metric	Mean Value	95% Confidence Interval
Accuracy	0.9091	(0.8182, 0.9773)
Balanced Accuracy	0.9145	(0.8250, 0.9821)
F1-macro	0.9074	(0.8120, 0.9772)
Precision-macro	0.9042	(0.8080, 0.9808)
Recall-macro	0.9145	(0.8250, 0.9821)
ROC-AUC (binary)	0.9679	(0.9150, 1.0000)

After comparing the classification performance of the default and optimized random forest models, additional validation was conducted using bootstrapped confidence intervals to quantify the uncertainty of each performance metric. Table 4 presents the mean values and corresponding 95% confidence intervals computed from 2000 bootstrap resamples of the test predictions. These results demonstrate the reliability and stability of the optimized model, with all intervals showing narrow bounds and consistently high scores across Accuracy, Balanced Accuracy, F1-macro, Precision-macro, Recall-macro, and ROC-AUC.

4.2 Additional Validation with Simulated Data

The same experimental code and pipeline were used for both datasets, ensuring consistency in preprocessing, resampling, model training, and evaluation. The only adjustment applied concerned the BO search space, which was broadened for the larger simulated dataset to allow higher model capacity while retaining the same optimization procedure and evaluation criteria.

The simulated dataset included 695 female-dominant and 277 male-dominant regional observations, maintaining a realistic class proportion consistent with demographic trends.

Table 5 presents the classification metrics comparison for both classes, highlighting the model's ability to accurately distinguish gender dominance across the simulated population.

Table 5. Classification metrics comparison on simulated data

Metric	Class	Default model	Optimized model
Precision	0	96%	99%
	1	89%	96%
Recall	0	96%	99%
	1	89%	96%
F1-Score	0	96%	99%
	1	89%	96%
Accuracy		94%	98%

The random forest model was optimized using the F1-macro score as the objective function, and the Gini criterion was maintained to ensure interpretability via feature importance analysis. As shown in Table 3. (original dataset) and Table 5. (simulated dataset), Bayesian Optimization substantially improved the performance of the random forest classifier compared with the default configuration.

When applied to the simulated dataset, the optimized model achieved high stability, reaching an overall accuracy of 97.9 %, with nearly perfect precision and recall across both gender classes. As shown in Table 6, the 95 % confidence intervals for all metrics were very narrow, indicating strong model consistency. These results confirm that the Bayesian-optimized Random Forest with SMOTENC generalizes reliably and maintains robust performance under larger, simulated data conditions.

Table 6. Bootstrapped test metrics on simulated data

Metric	Mean Value	95% Confidence Interval
Accuracy	0.9795	(0.9623, 0.9932)
Balanced Accuracy	0.9748	(0.9507, 0.9940)
F1-macro	0.9748	(0.9528, 0.9919)
Precision-macro	0.9748	(0.9508, 0.9948)
Recall-macro	0.9748	(0.9507, 0.9940)
ROC-AUC (binary)	0.9958	(0.9911, 0.9991)

4.3 Feature Importance

To interpret the optimized Random Forest classifier, Gini-based feature importance was computed to assess each predictor's contribution to reducing node impurity. As shown in Figure 9, Prefecture was the most influential feature, followed by Year and Population_Density, highlighting the significance of regional and temporal factors in explaining gender dominance patterns. Overall, the results confirm that the optimized model leverages both categorical and continuous features in a manner consistent with domain expectations.

Figure 10 illustrates the Gini-based feature importance for the optimized random forest trained on the simulated dataset. The results show that Year is the most influential predictor, followed by Population_Density and Prefecture, indicating that temporal and spatial factors strongly shape gender distribution patterns in the simulated data.

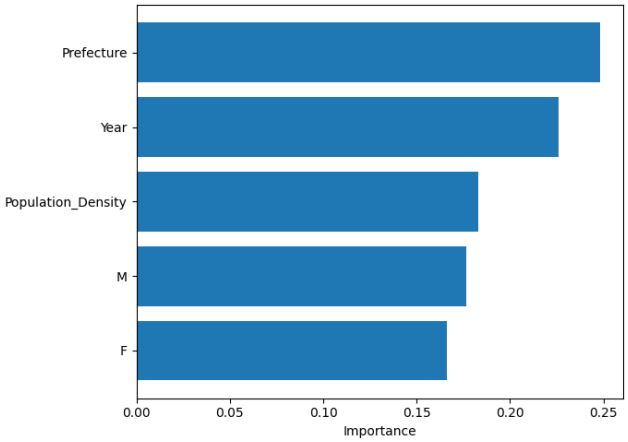


Figure 9. Feature importance (Gini) of the optimized Random Forest Classifier

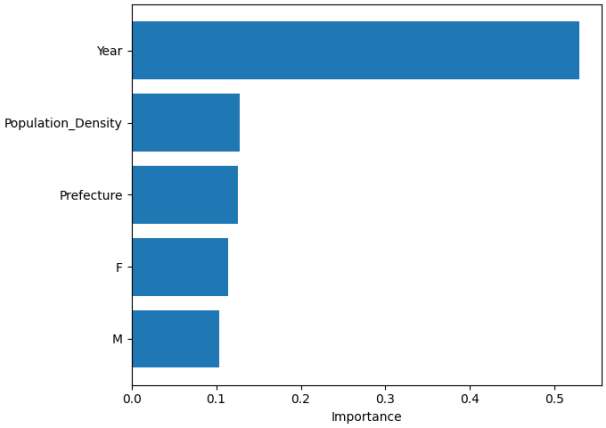


Figure 10. Feature importance (Gini) of the optimized Random Forest Classifier on simulated data

5. Conclusion

The proposed RF–BO–SMOTENC framework demonstrated strong and consistent performance in classifying regional gender dominance across Albania. Using the original dataset (144 entries), Bayesian Optimization significantly improved the random forest model's accuracy from 86% to 91%, reducing the error rate from 14% to 9% and yielding higher precision and F1-scores for both classes. When applied to the simulated dataset (972 entries), the same experimental pipeline achieved even higher stability, reaching an overall accuracy of 98% with narrow bootstrapped confidence intervals across all evaluation metrics.

These findings confirm that automatic hyperparameter tuning through Bayesian Optimization effectively enhances model generalization, while the integration of SMOTENC ensures balanced learning across gender categories. The consistent results between real and simulated data validate the robustness and scalability of the proposed framework for demographic classification tasks. These improvements support more accurate predictions of regional gender distribution, aiding policymakers in planning targeted interventions in healthcare, education, and resource allocation. Such applications promote balanced and equitable regional development.

6. Future Work

Future research should compare the proposed RF–BO–SMOTENC framework with alternative machine learning models such as support vector machine (SVM), XGBoost, or other ensemble and kernel-based classifiers to evaluate their relative performance in binary gender dominance classification. Such comparative analyses would help determine whether Bayesian Optimization consistently provides advantages in improving accuracy, interpretability, and generalization when applied to regional demographic data.

References

Arif N. H., Faisal M. R., Farmadi A., Nugrahadi D., Abadi F., Ahmad U. A. (2024). An Approach to ECG-based Gender Recognition Using Random Forest Algorithm. Journal of Electronics, Electromedical Engineering, and Medical Informatics, vol. 6, no. 2, pp. 107-115. https://doi.org/10.35882/jeeemi.v6i2.363

Ármás, J. (2022). An Emptying State? Demographic and International Migration Trends in Albania. FOREIGN POLICY REVIEW, *21*(1), 71-88. https://doi.org/10.47706/KKIFPR.2022.1.71-88

B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, N. De Freitas, "Taking the human out of the loop: A review of Bayes Search", Proceedings of the IEEE, Vol. 104, No. 1, 2016, pp. 148-175. doi: 10.1109/JPROC.2015.2494218

Boçi, B., & Simoni, A. (2025). Refining wage predictions with machine learning and Bayesian optimization. Mathematics and Statistics, 13(5), 279–296. https://doi.org/10.13189/ms.2025.130503

Breiman L. (2001). Random forests. Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. https://doi.org/10.1023/A:1010933404324

Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, vol. 16, no. 1, pp. 321. https://doi.org/10.1613/jair.953

Dantcheva A., Elia P., Ross A. (2011). What else does your biometric data reveal? A survey on soft biometrics. IEEE Transactions on Information Forensics and Security, vol. 11, no. 3, pp. 441–467. doi: 10.1109/TIFS.2015.2480381

Gëdeshi, I., & King, R. (2018). New trends in potential migration from Albania. Friedrich Ebert Stiftung.

Ibarra-Vazquez G., Ramírez-Montoya M. S., Buenestado-Fernández M. (2023). Forecasting Gender in Open Education Competencies: A Machine Learning Approach," in IEEE Transactions on Learning Technologies, vol. 17, pp. 1224-1235. doi: 10.1109/TLT.2023.3336541

Iscra, K., Biscontin, A., Miladinovic, A., Bonini, A., Furlanis, G., Prandin, G., ... & Ajčević, M. (2024). Optimizing machine learning models for classification of stroke patients with epileptiform EEG pattern: the impact of dataset balancing techniques. Procedia Computer Science, 246, 4600-4609. https://doi.org/10.1016/j.procs.2024.09.324

Jain A. K., Ross A., Prabhakar S. (2004). An introduction to biometric recognition. IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 4–20. doi: 10.1109/TCSVT.2003.818349

Jain S. K., Gupta A. K. (2022). Application of random forest regression with hyper-parameters tuning to estimate reference evapotranspiration. International Journal of Advanced Computer Science and Applications, vol. 13, no. 5, pp. 742–750. DOI:10.14569/IJACSA.2022.0130585

Kohavi R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In International Joint Conference on Artificial Intelligence, vol. 14, no. 2, pp. 1137-1145.

Lerch M. (2016). Internal and International Migration Across the Urban Hierarchy in Albania. Population research and policy review, 35(6), 851–876. https://doi.org/10.1007/s11113-016-94042

Mesquita F., Marques G. (2024). An explainable machine learning approach for automated medical decision support of heart disease. Data & Knowledge Engineering, vol. 153, pp. 102339. https://doi.org/10.1016/j.datak.2024.102339

Moghaddam B., Yang M. H. (2002). Learning gender with support faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 707–711. doi: 10.1109/34.1000244

Mulyana D.I., Pramansah V.V., (2022). Gender Classification for Anime Character Face Image Using Random Forest Classifier Method and GLCM Feature Extraction. JUITA: Jurnal Informatika, vol. 10, no. 2, pp.243-250. https://doi.org/10.30595/juita.v10i2.13833

Probst P., Boulesteix A. L., Bischl B. (2019). Tunability: Importance of hyperparameters of machine learning algorithms. Journal of Machine Learning Research, vol. 20, no. 53, pp. 1-32. https://doi.org/10.48550/arXiv.1802.09596

Ramadhan, M. M., Sitanggang, I. S., Nasution, F. R., & Ghifari, A. (2017). Parameter tuning in random forest based on grid search method for gender classification based on voice frequency. DEStech transactions on computer science and engineering, *10*(2017).

Rink, P., & Brannath, W. (2025). Post-selection confidence bounds for prediction performance. Machine Learning, *114*, 82. https://doi.org/10.1007/s10994-024-06632-w

Saito T., Rehmsmeier M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PloS one, vol.10, no. 3, pp. e0118432. https://doi.org/10.1371/journal.pone.0118432

Snoek J., Larochelle H., Adams R. P. (2012*).* Practical Bayesian optimization of machine learning Algorithms. In Advances in neural information processing systems, pp. 2951–2959.

Thiojaya, J., Qomariyah, N. N. (2023*).* Classification of human gender through machine learning analysis of gait. *In* Proceedings of the 2023 International Seminar on Intelligent Technology and Its Applications (ISITIA) (pp. 382–387). IEEE. https://doi.org/10.1109/ISITIA59021.2023.10221166

Verma C., Illés Z., Stoffová V. (2019). Gender prediction of Indian and Hungarian students towards ICT and mobile technology for the real-time. International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9S3, pp.1260-1264.

Wu J., Chen X. Y., Zhang H., Xiong L. D., Lei H., Deng S. H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. Journal of Electronic Science and Technology, vol. 17, no. 1, pp. 26-40. https://doi.org/10.11989/JEST.1674-862X.80904120