Evaluation of Machine Learning Models in Classifying Women's Labor Force Participation in West Java

Indra Rivaldi Siregar[1], Windy Ayu Pratiwi[2], Adhiyatma Nugraha[3], Bagus Sartono[4], Aulia Rizki Firdawanti[5]
[1,2,3,4,5] Study Program of Statistics and Data Science, IPB University, Bogor Indonesia
E-mail: [1]siregarindra@apps.ipb.ac.id, [2]Windyayu@apps.ipb.ac.id, [3]1204adhiyatmanugraha@apps.ipb.ac.id, [4]bagusco@apps.ipb.ac.id, [5]auliafirda@apps.ipb.ac.id

Abstract

This study compares four classification models—Logistic Regression, Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (AdaBoost)—to predict women's labor force participation in West Java, using a dataset of 62 features. After feature selection, the dataset was reduced to 31 features, followed by modeling with the top 10 most important features from each model. Model performance, evaluated using Balanced Accuracy, F1-Score, and Cohen's Kappa, showed similar results, with RF and XGBoost slightly outperforming the others. However, the differences were not significant, indicating comparable predictive ability across models. The top 10 features from each model were averaged, and the five most influential features were selected. Key factors influencing women's employment status include household responsibilities, age, education, district minimum wage, and the age of the youngest child. The analysis found that 79.6% of unemployed women manage household duties, while employed women are less involved (18.9%). Age was significant, with employed women mostly in the 35-55 age range, correlating with older children and greater workforce participation. Additionally, employed women are more likely to come from regions with lower minimum wages, suggesting that economic necessity drives their labor market participation.

Keywords: female labor force, machine learning, classification, West Java

1. INTRODUCTION

The workforce is a primary resource that plays a crucial role in the development of a nation. In Indonesia, the labor force faces challenges due to the large supply of workers, attributed to the country's significant population, with 11 cities having populations exceeding one million [1]. Data published by Statistics Indonesia (BPS) in 2023 revealed that the population distribution by age and gender shows that the male and female populations are nearly equal across all age groups [2]. This indicates that half of Indonesia's population comprises women, highlighting the equally significant role and contribution of women in driving economic development. However, women's participation in the labor force still faces disparities. According to BPS, the female labor force participation rate (LFPR) in 2023 was recorded at only 54.52%, compared to 84.26% for men [3].

In West Java, women's labor force participation is a critical aspect to address. As one of the most populous provinces in Indonesia, data from the 2023 National Socioeconomic Survey (Susenas) highlights variations in female labor force participation in the region. The Female Labor Force Participation Rate (TPAK) in West Java reflects various factors influencing women's decisions to participate in the labor market, including education levels, domestic responsibilities, access to decent employment opportunities, and socio-cultural conditions. Women often encounter challenges in balancing their dual roles as homemakers and professionals [4]. This suggests that women's participation in the workforce is influenced not only by educational and economic factors but also by the complexity of their dual roles. Support from families, flexible work environments, and policies that promote work-life balance are crucial to assist women in overcoming these challenges.

Classification can be conducted using various machine learning algorithms. This study focuses on classification analysis using machine learning methods such as Random Forest, Logistic Regression, Adaptive Boosting (AdaBoost), and Extreme Gradient Boosting (XGBoost). Several related studies include research by [5] on diabetes classification using KNN, Naive Bayes, and logistic regression methods and [6] on traffic accident prediction in Surakarta using multinomial logistic regression. Other studies include the analysis of Gradient Boosting, XGBoost, and CatBoost for mobile phone classification by [7], the application of Haar-Like features and AdaBoost for coffee plant pest classification by [8], and the classification of toddler nutritional status using the Random Forest algorithm by [9].

This study presents a more comprehensive approach compared to prior related research. In addition to evaluating the performance of various classification methods, it aims to identify the key variables that significantly influence women’s labor force participation in West Java. As a result, this research contributes not only to the advancement of classification methodologies but also provides a deeper understanding of the factors that shape women’s participation in the labor market.

Thus, this study aims to analyze women's labor force participation in West Java Province from the 2023 Susenas survey using machine learning-based classification methods. This approach is expected to provide new insights into the factors influencing women's workforce participation and the characteristics of women who are actively employed. By understanding the patterns of women's labor force participation, the results of this research are expected to serve as a reference for the government and policymakers in formulating more inclusive and effective programs and policies to enhance women's workforce participation. It is hoped that this research will contribute not only theoretically to the advancement of knowledge but also practically in promoting sustainable and inclusive economic development in Indonesia.

2. RESEARCH METHOD

2.1 Data

This study utilizes the 2023 SUSENAS dataset from West Java, with employment status (employed or not employed) as the response variable. The data is filtered to include only females aged 15 years and older.

2.2 Features and Response

Feature selection was based on several references relevant to the topic, such as the studies by [10] and [11], which utilized features like fertility decline, education level, and provincial minimum wage. These factors are predicted to be the primary drivers of female labor force participation in both developed and developing countries.

Additionally, studies by [12] and [13] employed features such as internet access and societal norms in their research on women’s labor force participation. Further, research by [14] and [15] incorporated features like household income, psychological factors, mental health, access to technology, and economic conditions in their analyses. Based on these sources and careful consideration of variables with a strong association with female labor force participation, this study uses a total of 42 features and one response variable. The selected features are detailed below.

1. **Individual Information (14):** sequence number in household, relationship to household head, marital status, age, age at first marriage, ever pregnant, age at first pregnancy, household management, worry about food insecurity, calories per capita, protein per capita, fat per capita, carbohydrate per capita, income per capita.
2. **Education/Literacy Skills (5):** ability to read and write Latin script, ability to read and write Arabic script, currently attending school, public/private school, highest educational attainment.
3. **Technology Skills (9):** mobile phone user, computer user, internet user, internet access at own home, internet access outside of home, internet access in public places, mobile internet use, internet use for social media, ever learned ICT skills.
4. **Residential Location (1):** kota/desa (urban/rural).

5. **Household Conditions (6):** housing ownership status, main source of lighting, type of household electricity, primary source of household income, floor area, number of household members.
6. **Other Assets (4):** owns at least 10 grams of gold, owns a motorcycle, owns a car, owns land.
7. **Social Assistance Recipients (2):** recipient of village BLT assistance, recipient of regular local government assistance.

2.3 Feature Engineering

Feature Engineering is the process of designing and transforming features used to create a data representation that is appropriate for learning models [16].

1. As an example, to obtain the early marriage feature, based on Article 7 paragraph (1) of Law No. 16/2019, the minimum age for marriage is 19 years old. Therefore, for the feature of age at first marriage, individuals aged 0 to 19 years are categorized as "yes" and others as "no."
2. The features of total number of children, age of the youngest child, average age of children, and having toddlers are derived from the household status feature using filters for children ('biological/adopted children').
3. The feature of the percentage of household members with higher education is obtained as follows:

$$\frac{The\ number\ of\ household\ members\ with\ the\ highest\ education\ level}{Total\ number\ of\ household\ members} x100 \quad (1)$$

With the number of household members holding the highest level of education, which includes diploma, bachelor's, master's, and professional degrees. With the total number of features and responses, after incorporating Feature Engineering, amounting to 62 variables, these variables will undergo data exploration to identify the most significant and optimal variables for the analysis. The process will involve examining the relationships between the variables, testing their relevance, and selecting those that contribute most effectively to the model's predictive power. Below is the complete list of features used in this study:

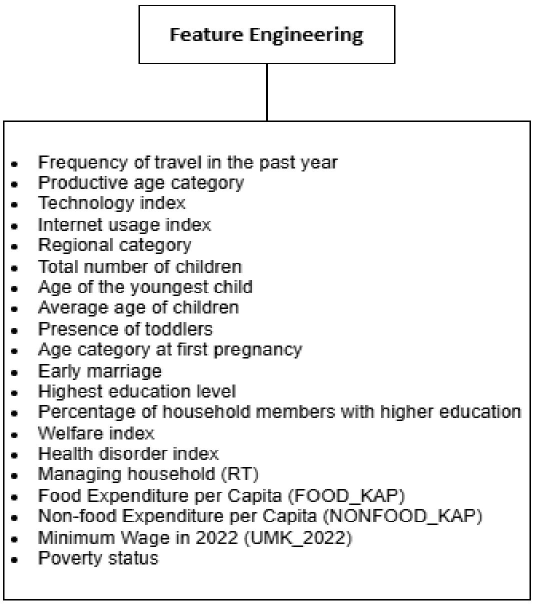


Figure 1. Feature Engineering

2.4 Features Selection

In the feature selection process, various checks will be performed to determine which features should be discarded and which are sufficiently potent for this study. In this study, feature selection will be conducted using several methods, such as boxplot analysis and Variance Threshold for numerical features, while for categorical features, the Chi-square test will be applied.

2.5 Preprocessing

The preprocessing process includes determining feature encoding, data splitting into training and testing sets, and standardization.

1. Feature Encoding

   In data mining, feature encoding is a crucial step to transform raw data into a more structured format, employing various methods to facilitate analysis. In this study, feature encoding is implemented using two techniques: Label Encoding and One-Hot Encoding. For the 'highest education level' feature, the categories are coded as follows: For the 'highest education level' feature, the categories are coded as follows: 'No Schooling' = 0, 'Elementary School' = 1, 'Secondary School' = 2, 'High School' = 3, 'Diploma' = 4, 'Bachelor’s Degree' = 5, and 'Postgraduate and Professional Degree' = 6. On the other hand, One-Hot Encoding represents each category as a binary vector, where only one bit is active to indicate a specific category, thus avoiding issues related to ordinality. Below is an illustration of One-Hot Encoding [17].

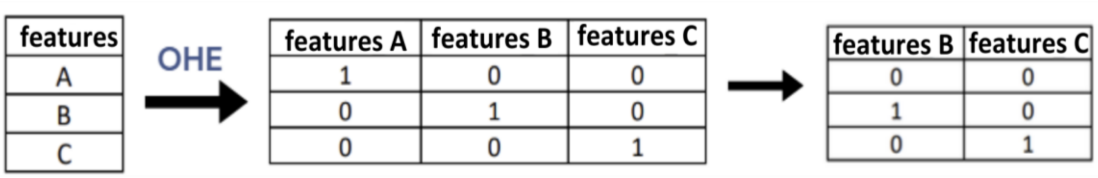


Figure 2. Illustration of one-hot encoding

2. Splitting Data

   The data is divided into two parts: training data and testing data. The primary goal of this division is to assess the model's performance when confronted with previously unseen data. The data splitting process can involve different proportions, depending on the analytical objectives. In this study, the data is split with an 80:20 ratio, where 80% is used for training and 20% is used for testing.

3. Standardization

   After performing the data split, the next step will be to standardize the data. Standardization of a dataset refers to the process of transforming the scale of data values such that the mean is 0 and the standard deviation is 1. This procedure ensures that each feature contributes equally to the analysis, thereby mitigating the effects of scale disparities among the variables [18].

2.6 Model Evaluation

This study utilizes four different machine learning model algorithms, namely Extreme Gradient Boosting (XGBoost), Random Forest, Adaptive Boosting (AdaBoost), and Logistic Regression.

1. Logistic Regression

   Logistic regression is a regression method used to model predictions and analyze the relationship between independent variables. This study uses a response variable representing employment status (working: yes/no) with the subject filter being women aged 15 years and above. This model is highly suitable for situations where the predicted outcome has only two possibilities (binary) and provides an estimation of the probability of a woman being employed based on specified characteristics or features [19]. The general equation for logistic regression is as follows:

$$P(Y=1)=\Pi=\left(\frac{e^{\beta_0+\beta_1 x_1}}{1+e^{\beta_0+\beta_1 x_1}}\right) \tag{2}$$

$$logit\ (P) = \beta_0 + \beta_1 x_1 \tag{3}$$

2. Random Forest

Random Forest (RF) is an ensemble learning method that combines multiple decision trees, each constructed from a bootstrapped subset of the training data. The trees are developed using a recursive splitting process, starting from the root node, and the splitting continues until predefined stopping criteria are reached. The predictive power of RF lies in aggregating the results of multiple weak learners (decision trees). Its performance is particularly strong when the trees in the forest have low correlation with one another. Further explanations and discussions on RF can be found in [20], [21]. Below is an illustration of Random Forest.

3. Extreme Gradient Boosting (XGBoost)

XGBoost, which stands for eXtreme Gradient Boosting, has become a leading ensemble learning algorithm that enhances traditional gradient boosting with various innovative features. Developed by Tianqi Chen, XGBoost introduces improvements in regularization, parallel computing, and tree optimization techniques, resulting in better performance and scalability compared to conventional gradient boosting methods. Overall, the literature demonstrates XGBoost's effectiveness as a versatile and powerful algorithm for classification tasks across different domains. Its ability to leverage advanced optimization techniques and handle complex data structures makes it the preferred choice for researchers and practitioners seeking high-performance solutions. In the context of this research, XGBoost is applied to classify the characteristics of the women labor force, aiming to understand the factors that determine whether a woman is employed or not [7].

4. Adaptive Boosting (AdaBoost)

Adaptive Boosting (AdaBoost) is one of the variants in the boosting algorithm family. AdaBoost is an ensemble learning method commonly used in boosting algorithms that combines the predictions from multiple models to improve overall performance. This approach is based on boosting techniques, where weak models (weak learners), typically simple decision trees, are combined to create a strong model (strong learner)[8].

2.6 Future Importance

Based on the objective of this study, which aims to understand the characteristics of working and non-working women, a Feature Importance analysis will be conducted. This analysis aims to identify which features have the greatest influence on the model's predictions, as well as which features can be ignored, thereby resulting in a more efficient and interpretable model.
Here is the research flow presented in a flowchart as shown below:

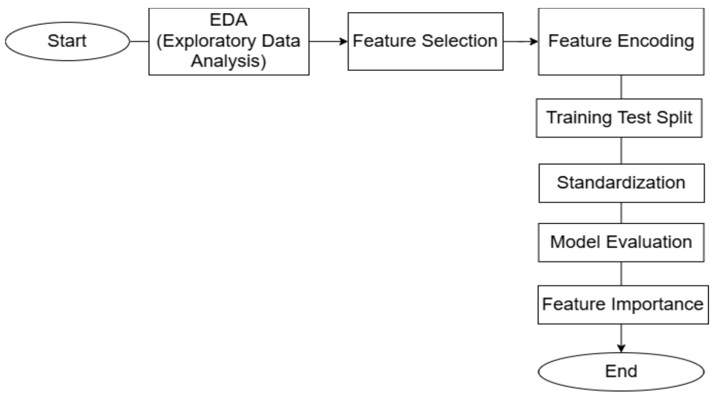


Figure 3. Workflow for evaluating machine learning models in classifying women's labor force participation in West Java

3. RESULTS AND DISCUSSION

3.1 Exploratory Data Analysis

In the data processing workflow, the initial step involved checking for the presence of missing values. Based on the analysis, some features in the dataset contained missing values with varying percentages. Features with missing values exceeding 5%, such as *public/private school*, were removed from the dataset as they were deemed to potentially affect the quality of the analysis. The "household electricity type" feature, with 4.5% missing values (1,430 entries), was imputed by filling missing values with "0 Watt," indicating non-PLN electricity sources without meters. This imputation aimed to reduce potential biases in the analysis resulting from missing data. Additionally, duplicate data checks were performed as part of the data-cleaning process. The results indicated that there were no duplicate entries in the dataset, eliminating the need for further steps to handle duplication. The absence of duplicate data ensures the integrity of the dataset to support further analysis.

An analysis was conducted to examine the distribution of worker percentages in West Java by gender for individuals aged 15 years and older. The results show that the percentage of working males reached 77.95%, significantly higher than females at 57.70%. Conversely, the percentage of non-working individuals was higher among females (42.30%) compared to males (22.05%). This significant disparity indicates a gender gap in workforce participation.

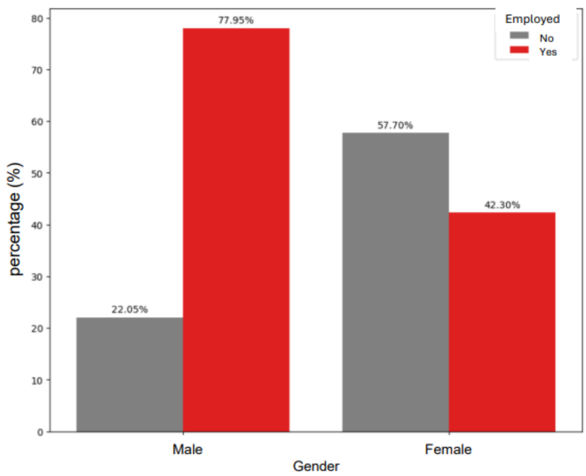


Figure 4. worker percentages in West Java

As part of the analysis, redundancy between variables was examined using a heatmap that visualizes the correlation among features in the dataset. We categorized the absolute Pearson correlation values into labels: L (Low) for values below 0.5, M (Medium) for values between 0.5 and 0.75, and H (High) for values above 0.75. The heatmap reveals that some variables exhibit high correlations, such as between 'food expenditure per capita' and 'non-food expenditure per capita', indicating a strong linear relationship between these variables. Such redundancy can cause issues in data dimension for modeling, so mitigation steps such as removing one of the variables. We also did the same analysis to drop other variables.

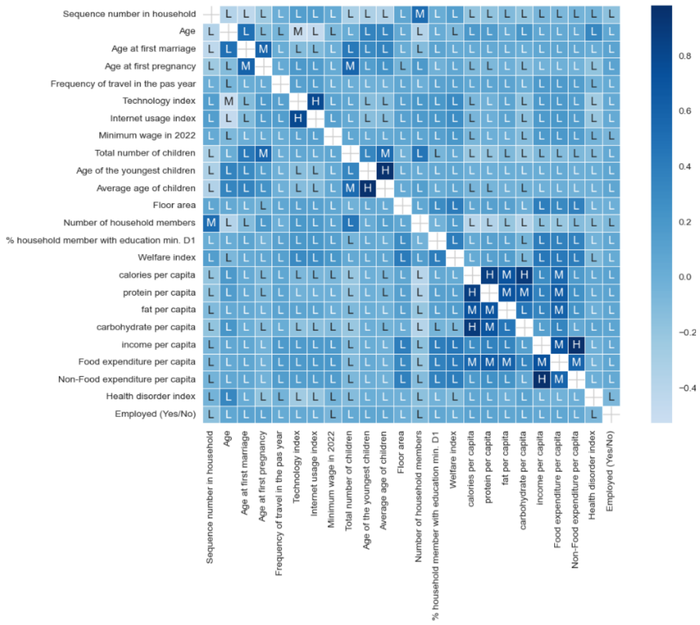


Figure 5. Heatmap

3.2 Feature Selection

a) Numerical

A boxplot was used to examine the differences in the behavior of each continuous predictor based on the female employment status response (employed/not employed). For example, the predictor "sequence number in household" shows a clear difference in median values between employed and non-employed women, indicating its potential as a feature for modeling. In contrast, predictors that do not show a median difference, such as "total number of children," were discarded and excluded from the modeling process. The same principle was applied to all individual continuous predictors.

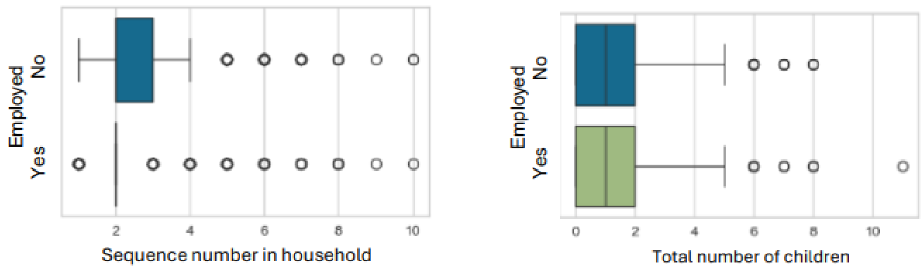


Figure 6. Feature selection using boxplot

After the data was normalized (range of 0-1) to ensure that differences in scale between variables did not affect the analysis, features with low variance were identified and removed using the Variance Threshold method. In this analysis, features with a variance less than 0.01 were deemed to contribute insignificantly to the data's variability and were dropped from the dataset. This step aims to enhance the model's efficiency in processing data.

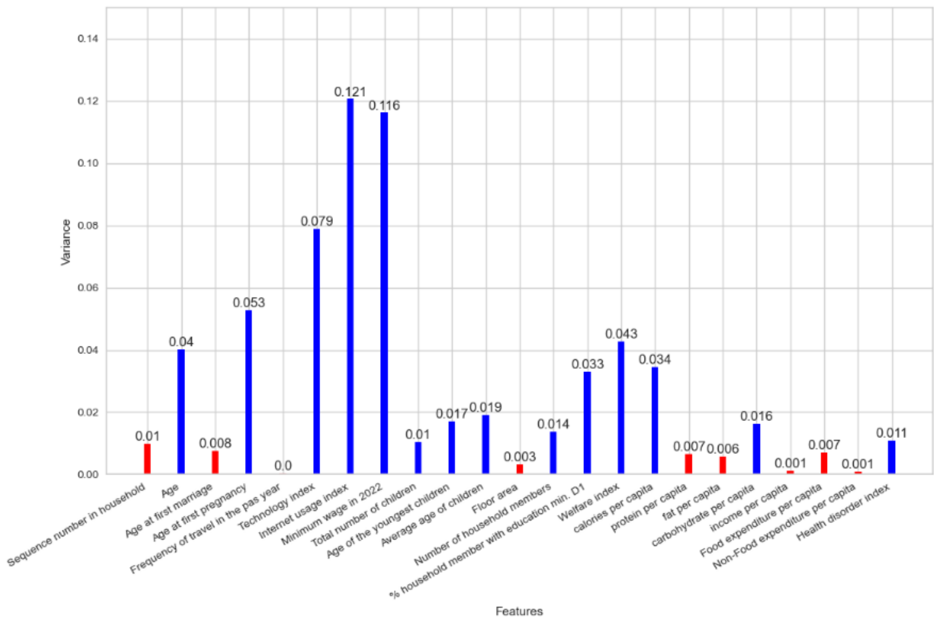


Figure 7. Variance threshold

b) Categorical

At this stage, categorical features were analyzed using the Chi-square test, which evaluates the relationship between independent variables (categorical features) and the target variable (classification target). During the Chi-square test, the proportion of each category within the variable was analyzed to determine whether there were significant differences in the target distribution. Features with significantly different distributions are considered relevant, while features with similar distributions are likely to be removed.

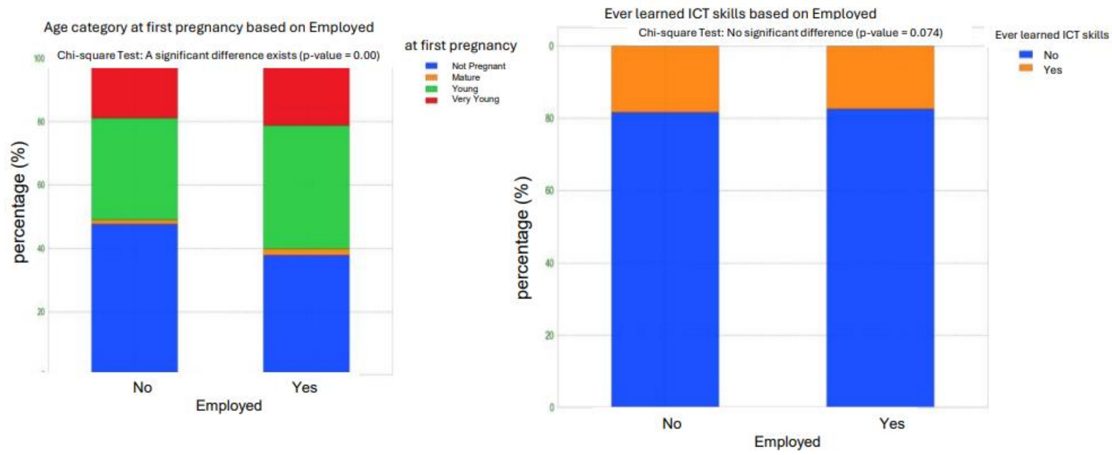


Figure 8. Feature Selection using Chi-square test

The figure illustrates the results of the Chi-square test, categorizing features into two main groups Different Proportions (Potential Features), Features that have a significant relationship with the target based on the Chi-square statistical test results and Similar Proportions (Features Likely to Be Removed), Features whose distribution concerning the target is considered insignificant and can be excluded from the model. After feature selection based on exploratory data analysis (EDA), including methods such as Variance Threshold, and Chi-square tests, a total of 31 features were selected as significant for further analysis. These features consist of a combination of numerical and categorical attributes, including:

Table 1. After feature selection

No	Feature	Type	No	Feature	Type
1	sequence number in household	numeric	17	productive age category	category
2	relationship to household head	category	18	minimum wage in 2022	numeric
3	marital status	category	19	regional category	category
4	age	numeric	20	age of the youngest child	numeric
5	ability to read and write Latin script	category	21	age category at first pregnancy	category
6	currently attending school	category	22	early marriage	category
7	highest educational attainment	category	23	housing ownership status	category
8	mobile phone user	category	24	type of household electricity	category
9	computer user	category	25	owns at least 10 grams of gold	category
10	internet access outside of home	category	26	owns a car	category
11	internet access in public places	category	27	owns land	category
12	mobile internet use	category	28	primary source of household income	category
13	internet use for social media	category	29	recipient of village BLT assistance	category
14	ever pregnant	category	30	number of household members	numeric
15	age at first pregnancy	numeric	31	poverty status	category
16	household management	category			

This feature selection process ensures that the features used in the analytical model are not only statistically relevant but also meaningful in the domain context, thus supporting better interpretation of the research results. The total data used in this process includes 31731 individual women.

3.3 Modeling

This study investigates the classification of women's employment status (employed vs. unemployed) using four classification models: Logistic Regression (Logistic), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (AdaBoost). The evaluation is conducted under two scenarios: utilizing all 31 selected features and a subset of the 10 most important features identified through feature importance rankings from each model.

Hyperparameter tuning is not performed for Logistic Regression, as it is not a machine-learning algorithm that requires hyperparameter optimization. Instead, it serves as a baseline model. For RF, XGBoost, and Adaboost, hyperparameter tuning is conducted using the k-fold cross-validation technique with k=5 on the training data. This ensures a robust and unbiased evaluation while keeping the testing data untouched during training and validation. The k-fold cross-validation technique divides the training dataset into equal-sized K subsets (folds). In each iteration, one-fold is the validation set, while the remaining k−1 folds form the training set. This process is repeated k times, with the model being validated on each fold once. The average performance across all folds provides a reliable estimate of the model's effectiveness.

Table 2. Model Evaluation

Metrics	31 Features				the 10 most important features			
	Logistic	RF	XGBoost	Adaboost	Logistic	RF	XGBoost	Adaboost
Bal Accuracy (Train)	0.86	0.89	0.87	0.87	0.86	0.87	0.87	0.86
F1-score (Train)	0.84	0.87	0.85	0.84	0.84	0.85	0.85	0.84
Kappa (Train)	0.74	0.79	0.76	0.74	0.73	0.76	0.75	0.74
Bal Accuracy (Test)	0.86	0.87	0.87	0.87	0.86	0.87	0.87	0.87
F1-score (Test)	0.84	0.85	0.85	0.85	0.84	0.85	0.85	0.85
Kappa (Model)	0.74	0.76	0.76	0.75	0.74	0.76	0.76	0.75

Balanced Accuracy evaluates the mean of sensitivity (true positive rate) and specificity (true negative rate). The findings indicate that all models exhibit consistent performance across both the training and test datasets, with Balanced Accuracy scores ranging from 0.86 to 0.89. Random Forest (RF) and XGBoost marginally outperform Logistic Regression and Adaboost in both scenarios (31 features and top 10 features), although the differences are not statistically significant. The F1-Score highlights the trade-off between precision and recall, making it an appropriate measure for binary classification tasks. All models achieve F1-Scores between 0.84 and 0.87 across training and testing datasets. RF and XGBoost achieve the highest scores, while Logistic and Adaboost perform only marginally lower, demonstrating their effectiveness. Cohen's Kappa measures the agreement between predicted and actual classifications, accounting for chance agreement. RF and XGBoost result in the highest Kappa scores (0.76 on testing) in the scenario of using the top ten features, followed closely by Adaboost. Logistic Regression demonstrates slightly lower Kappa values (0.74 on testing), though the differences are not substantial.

The reduction to ten features does not significantly impact the models' performance. Balanced Accuracy, F1-Score, and Kappa remain nearly identical between the two feature sets. This finding indicates that the selected ten features are sufficient to capture the underlying patterns in the data, allowing for a more computationally efficient and interpretable model. The results reveal no significant differences in performance among the four models. All models demonstrate comparable capabilities, with RF and XGBoost achieving slightly higher metrics. However, the performance differences are negligible, suggesting that Logistic Regression and Adaboost are also suitable for this task.

3.4 Modeling Feature Importance and Interpretation

Although RF and XGBoost achieve slightly higher scores mathematically but are not significant, we assume that all four models—RF, XGBoost, AdaBoost, and Logistic Regression—perform comparably overall. Initially, we extracted the top 10 most important features from each model and then normalized (scale 0 - 1) to ensure comparability across models. A feature that does not appear in other models will be assigned to be zero value. The merge of feature importance is presented in a single table, as shown in **Table 3**. We calculate the average importance of each feature across all models and rank the features based on their average scores for further analysis. This approach ensures a balanced and unbiased framework for integrating insights from multiple models while leveraging their complementary strengths.

Table 3. Merged Top 10 Feature Importance of All Models

Features	RF	XGBoost	Ada	Log
number of household members	0.02	0	0	0
primary source of household income_*investment*	0	0	0	0.1
primary source of household income_*remittance*	0.02	0.01	0.03	0.07
primary source of household income_*pension*	0	0	0	0.05
minimum wage in 2022	0.03	0	0.27	0
currently attending school	0.06	0.26	0.06	0.13
highest educational attainment	0.03	0	0.04	0
regional category	0.02	0	0.02	0
productive age category	0	0	0	0.06
household management	0.68	0.45	0.06	0.2
ever pregnant	0	0.13	0	0
relationship to household head_*child*	0	0.02	0	0
relationship to household head_*head*	0	0.01	0.03	0
relationship to household head_*parents/ parents in law*	0	0.04	0	0.11
relationship to household head_*housekeeper/driver*	0	0	0	0.1
marital status _*marriage*	0	0.03	0	0.05
age	0.1	0.03	0.3	0.13

age at first pregnancy	0.02	0.01	0.06	0
age of the youngest child	0.03	0	0.13	0

Referring to **Figure 9**, the five most significant features influencing the classification of women's employment status are evident. The highest-ranking feature is whether a woman is primarily responsible for managing household affairs, highlighting the dominant role of domestic responsibilities. The second most important feature is age, which likely reflects the varying labor force participation rates across different age groups. The third feature pertains to the woman's educational background, categorized as whether she has never attended school, is currently enrolled, or has completed her education, emphasizing the critical role of education in employment opportunities. The fourth feature is the district or city minimum wage, indicating the economic context and its influence on employment decisions. Lastly, the age of the youngest child emerges as a key factor, suggesting the impact of childcare responsibilities on women's labor force participation.

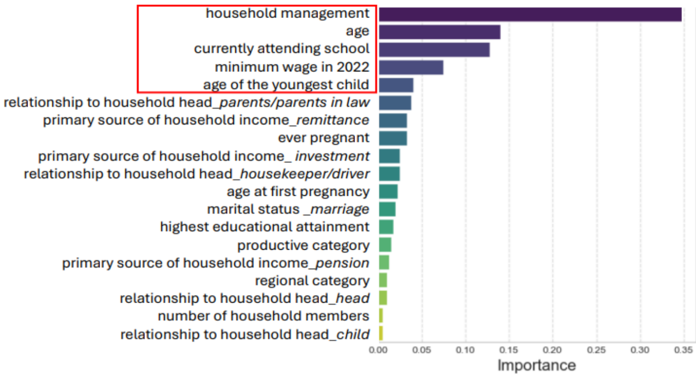
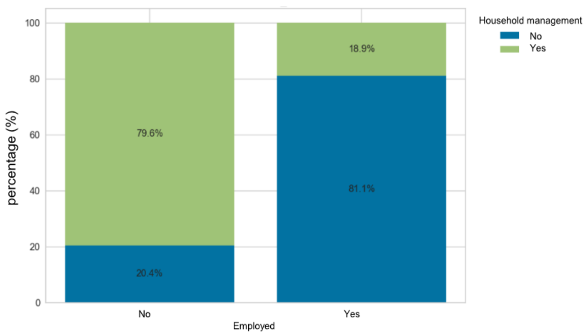


Figure 9. Average score of merged Top 10 feature importance of All Models

Table 4. Chi-Square Test (Homogeneity of proportions)

Feature	Chi-square Test
Household management	statistics: 11483.5 p-value: 0.00[a]
Age Group	statistics: 1350.6 p-value: 0.00[a]
Currently attending school	statistics: 1353.3 p-value: 0.00[a]
Regency minimum wage	statistics: 113.0 p-value: 0.00[a]



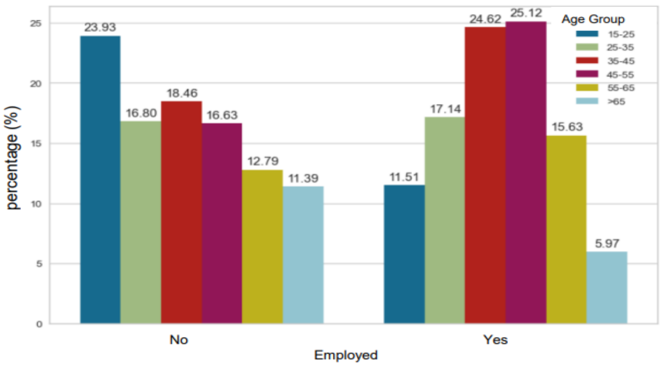


Figure 10: (a) Proportion of household managemen; **(b)** Proportion of age group

The analysis reveals that women who are not employed tend to take on household responsibilities, with 79.6% of them managing domestic affairs. In contrast, among employed women, the proportion of those managing household responsibilities compared to those who do not is 18.9% to 91.1%, respectively. A chi-square test is performed to statistically confirm this difference in proportions. The results indicate that the observed differences are significant at the 1% significance level, confirming the strong association between employment status and household responsibilities among women. This outcome is in line with [22], which asserts that household management/duties-such as washing clothes, cooking, cleaning, and caring for family members-are almost completely delegated to women.

We categorize age into several groups, as illustrated in **Figure 10(b)**. The data show that, among employed women, 24.62% are in the 35-45 age range, and 25.12% are in the 45-55 age range. This suggests that nearly half of employed women fall within the 35-55 years. These results align with the work by [15], which discovered that the peak of women's labor force participation in Indonesia occurs at around 45 years of age, followed by a gradual decline. In contrast, 23.93% of unemployed women, or approximately one-quarter, are in the 15-25 age group. A chi-square test at the 1% significance level further confirms a significant difference in the proportions between the two groups.

Table 5. Mean Age of Youngest Child based on Women's Age Group

Women's Age Group (year)	Mean Age of Youngest Child (year)
15-25	0.2
25-35	3
35-45	7.1
45-55	11.7
55-65	10.8
>65	6.8

We examine the relationship between the age of women and the age of their youngest child by calculating the mean age of the youngest child within each age category of women, as shown in the table above. Among employed women, nearly 50% fall within the 35-55 age range, which correlates with an average age of their youngest child between 7.1 and 11.7 years. It indicates that women in this age group are less likely to have infants or very young children, which could enable them to participate more actively in the workforce, as they no longer face the intensive caregiving duties associated with raising young children. On the other hand, unemployed women, predominantly aged 15-25, tend to have very young children, with an average age of 0.2 years. It aligns with the tendency of these women to focus on child-rearing rather than engaging in employment, as they have infants who require more attention and care. These findings are further corroborated by the study conducted by [15], which compared the decline in female labor force participation between Indonesia and the Netherlands. The study found that in Indonesia, the drop in participation during the childbearing years (post-childbirth) is notably more pronounced than in the Netherlands.

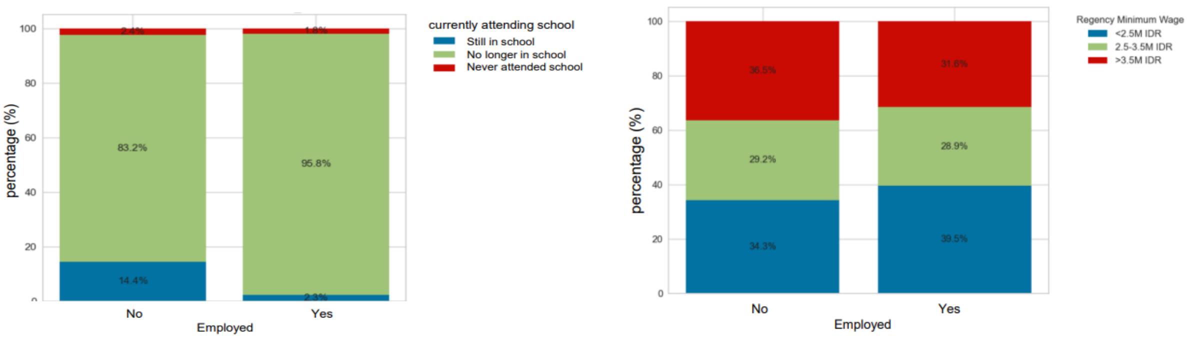


Figure 11: (a) Proportion od currently attending school; (**b)** Proportion of minimum wage

For women who are employed, approximately 95.8% have completed their education or are no longer attending school. In contrast, around 14% of women who are not employed are still currently enrolled in school. This result is in line with the study by [11], which found a positive relationship between the duration of schooling and the participation rate of women in the labor force in Indonesia.

Furthermore, a dominant portion of unemployed women come from regions with a minimum wage above 3.5 million IDR, accounting for 36.5%. In contrast, employed women are predominantly found in areas with a minimum wage below 2.5 million IDR, with a percentage of 39.5%. This finding contrasts with the study by [11], which demonstrates a positive relationship, where higher minimum wages are associated with an increase in the women's labor force participation rate in Indonesia. We suggest that these women are likely working to meet pressing economic needs. With lower wages, they may not generate sufficient income from other sources, making employment a critical necessity. These findings emphasize the connection between employment status and regional minimum wage levels, highlighting an important factor for labor market analysis. This relationship could be a key indicator for understanding the economic drivers of female workforce participation in different regions.

4. CONCLUSION

This study compares four classification models—Logistic Regression, Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (AdaBoost)—for predicting women's employment status. The models are evaluated using two scenarios: all 31 features and a subset of the top 10 most important features. Hyperparameter tuning was conducted for RF, XGBoost, and AdaBoost using k-fold cross-validation, while Logistic Regression served as a baseline model. The performance metrics, including Balanced Accuracy, F1-Score, and Cohen's Kappa, show that all models perform similarly, with RF and XGBoost slightly outperforming the others. However, the differences in performance are not substantial, leading us to conclude that, overall, the models are comparable in their predictive ability.

Analysis reveals that key factors influencing women's employment status include household responsibilities, age, education, district minimum wage, and the age of the youngest child. The results show that a significant portion of unemployed women (79.6%) manage domestic affairs, whereas employed women tend to have a lower involvement in household management (18.9%). Age also plays a crucial role, with employed women predominantly in the 35-55 age range, correlating with the age of their youngest child, suggesting that older women are less likely to have young children and, therefore, more able to participate in the workforce. Additionally, the study highlights that employed women tend to come from regions with lower minimum wages, indicating that economic necessity may drive their participation in the labor market.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Department of Statistics for their support and the Badan Pusat Statistik (BPS) for providing the data essential for this study.

REFERENCES

[1] World Population Review, "Indonesia Population 2024," World Population Review.

[2] Badan Pusat Statistik, "Jumlah Penduduk Menurut Kelompok Umur dan Jenis Kelamin, 2023," Badan Pusat Statistik.

[3] Badan Pusat Statistik, "Tingkat Partisipasi Angkatan Kerja Menurut Jenis Kelamin, 2021-2023," Jakarta, Nov. 2023.

[4] S. M. Berliana, D. Lukmi, A. Purbasari, S. Tinggi, and I. Statistik, "FAKTOR-FAKTOR YANG MEMENGARUHI JAM KERJA TENAGA KERJA WANITA BERSTATUS KAWIN DALAM SEMINGGU DI INDONESIA (ANALISIS DATA SAKERNAS 2014)," *Jurnal Ilmiah WIDYA*, vol. 3, 2016.
[5] D. Nasien *et al.*, "Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, Dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes," *JEKIN (Jurnal Teknik Informatika)*, vol. 4, no. 1, pp. 10–17, 2024.
[6] B. L. Fauzan, T. Agustin, and A. M. H. Mahmudah, "Prediksi Klasifikasi Kecelakaan Lalu Lintas di Kota Surakarta dengan Menggunakan Metode Regresi Logistik Multinomial," *Sustainable Civil Building Management and Engineering Journal*, vol. 1, no. 4, p. 9, Aug. 2024, doi: 10.47134/scbmej.v1i4.3159.
[7] A. F. L. Ptr, M. M. Siregar, and I. Daniel, "Analysis of Gradient Boosting, XGBoost, and CatBoost on Mobile Phone Classification," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 6, no. 2, pp. 661–670, Apr. 2024, doi: 10.47709/cnahpc.v6i2.3790.
[8] Supiyandi, M. Siddik Hasibuan, and H. Harahap, "PENERAPAN METODE HAAR-LIKE FEATURE DAN ALGORITMA ADABOOST DALAM PENENTUAN KLASIFIKASI HAMA TANAMAN KOPI," 2024. [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR
[9] P. Handayani, A. Charis Fauzan, and Harliana, "Komputer Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random Forest," *KLIK: Kajian Ilmiah Informatika*, vol. 4, no. 6, pp. 3064–3072, Jun. 2024, doi: 10.30865/klik.v4i6.1909.
[10] N. Nazah, J. Duasa, and M. I. Arifin, "Fertility and Female Labor Force Participation in Asian Countries; Panel ARDL Approach," *Jurnal Ekonomi & Studi Pembangunan*, vol. 22, no. 2, pp. 272–288, Oct. 2021, doi: 10.18196/jesp.v22i2.11142.
[11] Z. T. Rahayu, S. N. Sarfiah, and L. R. Indrawati, "Determinants of the Women's Labor Force Participation Rate in Indonesia."
[12] R. Heath *et al.*, "Female Labour Force Participation," *VoxDevLit*, vol. 11, no. 1, pp. 1–43, Feb. 2024.
[13] A. Rimkute and L. Sugiharti, "The Link Between Occupations, Labor Force Participation of Married Women, and Household Technology in Indonesia," *Journal of Population and Social Studies*, vol. 31, pp. 20–37, 2023, doi: 10.25133/JPSSv312023.002.
[14] D. Halim, M. B. O'sullivan, and A. Sahay, "OVERVIEW WORLD BANK GROUP GENDER THEMATIC POLICY NOTES SERIES: EVIDENCE AND PRACTICE NOTE INCREASING FEMALE LABOR FORCE PARTICIPATION."
[15] L. Cameron, D. Contreras Suárez, and W. Rowell, "Working Paper Series Female Labour Force Participation in Indonesia: Why Has It Stalled," 2018. [Online]. Available: http://www.melbourneinstitute.com
[16] C. Herdian, A. Kamila, and I. G. Agung Musa Budidarma, "Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label Encoding dan One-Hot Encoding Pada Metode Linear Regresi," *Technologia : Jurnal Ilmiah*, vol. 15, no. 1, p. 93, Jan. 2024, doi: 10.31602/tji.v15i1.13457.
[17] K. Aditya Ananta Wisnu Wardana and A. A. Mizwar Rahim, "Analisis Perbandingan Algoritma XGBoost Dan Algoritma Random Forest Untuk Klasifikasi Data Kesehatan Mental," *Jurnal Ilmu Komputer dan Pendidikan*, vol. 2, no. 5, pp. 808–818, Aug. 2024, Accessed: Dec. 03, 2024. [Online]. Available: https://journal.mediapublikasi.id/index.php/logic
[18] M. S. Gal and D. L. Rubinfeld, "Data standardization," *New York University Law Review*, vol. 94, no. 4, pp. 737–770, Oct. 2019, doi: 10.2139/ssrn.3326377.
[19] F. Faisal, I. R. Sari, I. Saraswati, and J. Heikal, "Analyzing the Influence of Product Characteristics on Customer Repurchase Intentions Using Binary Logistic Regression Method," *AKADEMIK: Jurnal Mahasiswa Humanis*, vol. 4, no. 3, pp. 1182–1190, 2024.

[20] H. Zhang and B. Singer, *Recursive Partitioning and Applications*, 2nd ed. New Yor: Springer Science & Business Media, 2010. [Online]. Available: www.springer.com/series/692
[21] L. Breiman, “Random Forests,” 2001.
[22] V. Y. D. Utari, “Unpaid Care Work in Indonesia: Why Should We Care?,” Jakarta, Oct. 2017. Accessed: Dec. 14, 2024. [Online]. Available:https://www.smeru.or.id/sites/default/files/events/24102017_fkp_ucw.pdf.