International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 15, No. 1, March 2026, pp. 189~197
ISSN: 2252-8776, DOI: 10.11591/ijict.v15i1.pp189-197

An integration clustering and multi-target classification approach to explore employability and career linearity

Nadzla Andrita Intan Ghayatrie, Devi Fitrianah
Department of Computer Science, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Feb 23, 2025
Revised Oct 17, 2025
Accepted Nov 5, 2025

Keywords:

Data augmentation
Gender gap
K-means least trimmed square
Multi-target random forest
Women in science, technology, engineering, and mathematics

ABSTRACT

This study analyzes job placement waiting times and job linearity among female science, technology, engineering, and mathematics (STEM) graduates using clustering and multi-target classification (MTC) models. The K-means least trimmed square (LTS) algorithm, known for its robustness against outliers, was employed for clustering. With k = 2 and a trimming percentage of 30%, the model achieved a silhouette score of 77%, resulting in two distinct clusters: ideal and non-ideal. To enhance the dataset for classification, synthetic data was generated using the adaptive synthetic (ADASYN)-gaussian method. Principal component analysis (PCA) was used for visualization purposes, along with overlapping histograms, to illustrate that the synthetic data distribution closely resembled the original. For classification, a random forest (RF) model was used to predict both jobs waiting time and job linearity. Hyperparameter tuning produced an optimal model with a classification accuracy of 92%. Cross-validation (CV) confirmed the model's robustness, with F1-micro and F1-macro scores of 94% and 93%, respectively. Results show that although women in STEM are underrepresented, 73% of the female alumni analyzed belonged to the short job waiting group. Furthermore, a strong negative correlation between GPA and job waiting time suggests that higher-GPA graduates tend to secure employment more quickly.

Corresponding Author:

Nadzla Andrita Intan Ghayatrie
Computer Science Department, BINUS Graduate Program, Master of Computer Science
Bina Nusantara University
11480, Jakarta, Indonesia
Email: nadzla.ghayatrie@binus.ac.id

1. INTRODUCTION

The underrepresentation of women in science, technology, engineering, and mathematics (STEM) remains a pressing concern in global gender disparity. According to [1], the labor force participation gap among individuals aged 15 and above reaches 27.1%, with women comprising 46.8% and men 73.9%. Women also hold only 26.2% of parliamentary seats compared to 73.8% for men, highlighting the urgent need to address the persistent global imbalance in workforce participation. In STEM fields specifically [2], female representation increased only slightly-from 26.1% in 2016 to 28.6% in 2024-while women's participation in non-STEM sectors has neared 50%. This contrast underscores the importance of further research into female STEM graduates' employment outcomes, particularly job waiting time and job linearity (the relevance of their work to their field of study).

Machine learning (ML) [3], offers powerful tools for clustering and classification, bolstered by data augmentation techniques to tackle class imbalance [4], [5]. As depicted in Figure 1, female alumni are significantly outnumbered by males. However, there is hope. To counter this imbalance, oversampling methods such as adaptive synthetic (ADASYN) [6] and SMOTE [7] are commonly employed. ADASYN generates synthetic data in sparse minority regions through linear interpolation, while its Gaussian-enhanced variant (ADASYN-Gaussian) introduces greater diversity and resistance to outliers. This study harnesses the potential of ADASYN-Gaussian to balance the minority (Female alumni) class, offering a promising path to address gender disparities in STEM.

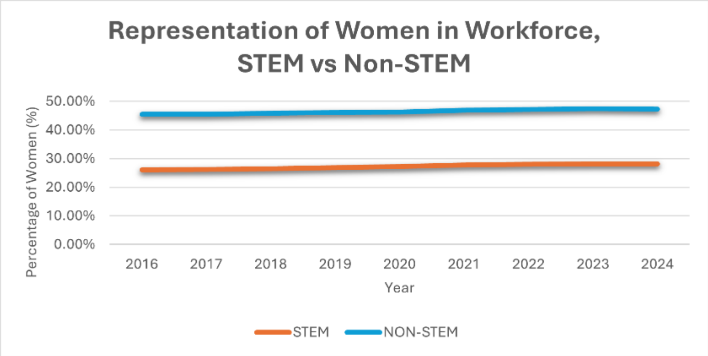

Figure 1. Comparison of women in STEM and non-STEM dields

Beyond clustering job waiting times, this research also examines job linearity, which refers to the degree of predictability in job waiting times, using both unsupervised and supervised approaches. Clustering is used to group job waiting durations, while classification predicts both jobs waiting category and job linearity [8]. Prior studies have explored similar themes. For instance, [9] applied fuzzy c-means (FCM) to cluster job waiting times into “fast,” “moderate,” and “slow” categories, followed by C4.5 decision tree classification, achieving only 86% accuracy without k-value optimization. Another study [10] employed fuzzy clustering to identify employment factors, emphasizing job stability (40.7%) and employment rate (34.4%) as key influences. A premilinary version of this research [11], proposed a combined clustering and multi-target classification (MTC) framework, achieving 77% accuracy and a silhouette score of 0.61. Building on that, the present research enhances performance through optimized model selection, using K-means LTS for its robustness to outliers-validated across ten benchmark datasets [12].

This study predicts two target variables simultaneously, requiring a MTC approach [13], [14]. In [15], Prior research supports random forest (RF) effectiveness in predicting graduate employability. Similarly, [16] used genetic algorithms (GAs) to predict career success (87.61% fit) and [17] applied Naïve Bayes with 90% accuracy on limited data. Comparative evaluations [18], have further shown RF superior performance, robustness to noise, and resistance to overfitting, making it a reliable choice for this study’s outlier-heavy dataset.

This paper presents a comprehensive integrated modeling framework that combines ADASYN-Gaussian augmentation, K-means LTS clustering, and multi-target random forest (MTRF) classification. Unlike earlier works that focused on single-target prediction or models lacking outlier resilience, this study jointly models job waiting time and job linearity in a unified pipeline. To our knowledge, no previous research has examined these dual employment outcomes for female STEM graduates using such an integrated approach.

2. RESEARCH METHOD

This study consists of three main stages: data preparation, K-means LTS clustering, and the MTRF classification process, as illustrated in Figure 2. In the data preparation phase, the pre-processing step ensures the dataset is free from missing values and duplicate records. Once the data is clean, data augmentation is conducted to address the significant imbalance between male and female alumni records. This study applies the ADASYN-gaussian technique to generate synthetic data for the minority class (female alumni). After augmentation, the data selection process is performed to isolate only female alumni records for further modeling.

Next, the K-means LTS clustering process is applied to group the job waiting times of the selected alumni. This involves hyperparameter tuning to determine the optimal number of clusters (K) and the best trimming percentage. The final stage is the classification process, where the grouped data is used as input for a MTRF model to predict both jobs waiting time and job linearity simultaneously.

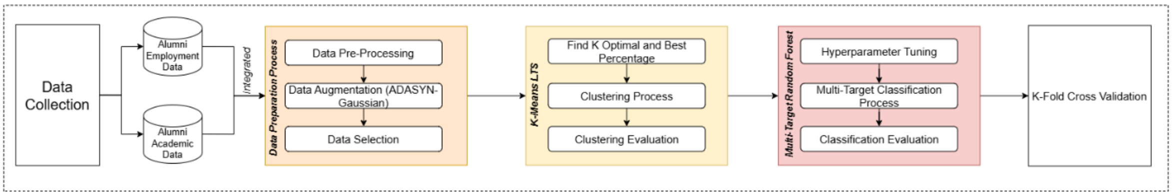

Figure 2. Research framework

2.1. Data collection

The data used in this study was sourced from several private universities in Indonesia. Two types of datasets were collected: alumni academic records and post-graduation employment data. The data collection process involved sending formal requests to each university, accompanied by a letter of agreement stating that the data would be used strictly for research purposes, would remain confidential, and would not reveal the identity of any participating institutions. Upon approval and the signing of a memorandum of understanding (MoU), the universities provided the requested datasets, which were then processed for research analysis. Based on the collected data, it was found that female alumni made up only 15% of the total records, confirming the class imbalance addressed through augmentation.

2.2. Data preparation process

2.2.1. Data pre-processing

Data pre-processing was conducted to ensure data quality and eliminate anomalies [19]. This stage involved several steps, including checking for null values, removing duplicate records, and identifying outliers, all of which are essential to avoid suboptimal modeling outcomes. Outlier detection was performed using the interquartile range (IQR) method and visualized through boxplots generated with the matplotlib library [20].

2.2.2. Data augmentation

ADASYN is a data augmentation technique that generates synthetic samples through linear interpolation between minority class instances and their nearest neighbors [6]. However, this linear approach may be inadequate when the minority class exhibits a complex data distribution [21], potentially leading to synthetic data that lacks diversity or fails to reflect the underlying distribution accurately. To address this limitation, ADASYN-Gaussian, a modified version of ADASYN, incorporates the Gaussian distribution to generate synthetic data. This study adopts the multivariate Gaussian distribution, suitable for two or more data dimensions, as defined in (1).

$$p(x;\ \mu,\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \quad (1)$$

The parameters k and β are critical in the augmentation technique. The parameter k determines how far and similar the generated synthetic samples are. Incorrect k values can lead to overfitting, as augmentation can be a regularizer controlling model complexity [22]. The parameter β adds variation through controlled noise while preserving the overall relevance of the data [23]. Selecting the appropriate β value ensures that the augmented data is relevant and does not deviate significantly from the original data. ADASYN-gaussian automatically tunes these parameters, making applying this technique and generating synthetic data straightforward.

2.2.3. Data selection

The data selection process was conducted to isolate and retain only the records of female alumni after the augmentation step. Given that female alumni constituted only 15% of the initial dataset, data augmentation was essential to address this significant imbalance prior to modeling. Following the augmentation process, which considered the original distribution of female alumni data, only the relevant female alumni records were filtered and selected for use in the subsequent modeling stages.

2.3. K-means LTS

K-means least trimmed square (LTS) [12] is a modification of the K-means algorithm that trims outliers after clusters are formed. Clustering is employed to group alumni based on similarities in their characteristic within the dataset. This approach helps uncover the natural structure of the data and provides initial insights into the number and nature of the resulting groups. These clusters serve as the basis for further analysis, which is then followed by classification stages. Clustering is performed prior to classification to better understand the distribution and structure of the data, especially when labels such as employment patterns or behavior are not explicity available. In this model, the optimal number of clusters (k) and the trimming proportion are determined through tuning functions provided within the algorithm's class.

Altough centroid initialization of K-means can influence the final clustering due to its susceptibility to local optima [24], this issue is mitigated in K-means LTS. In this method, clustering is performed using multiple random initializations, and the best result is selected based on the lowest total within-cluster variance. Furthermore, the trimming process removes a specified proportion of data points with the highest residuals after clustering, which helps reduce the influence of poorly initialized centroids and outliers. These two mechanisms (repeated initialization and trimming) contribute to a more robust and stable clustering outcome.

2.3.1. Find K optimal and best percentage

This process is a parameter tuning step required for the modified K-means model. Using the trimming concept as in [25], K-means LTS requires an optimal outlier trimming percentage. Tuning is performed by assessing the silhouette score, and the parameter combination that yields the best silhouette score is selected for modeling.

2.3.2. Clustering process

K-means LTS looks like robust trimmed K-means (RTKM) [25] utilizing the concept of trimming outliers based on LTS. In K-means LTS [12], the process involves sorting the clustered dataset by its distance to the centroid and trimming the farthest data points (outliers) based on a specific percentage. Data falling within the farthest n% from the centroid are trimmed. This algorithm manages complexity well, with main steps such as centroid initialization and cluster ID assignment $\left(O(n \times k)\right)$, calculating data distances to the centroid $\left(O(n \times k \times d)\right)$, sorting the distances for outlier trimming $\left(O(n * \log(n))\right)$, and reassigning clusters $\left(O(n \times k)\right)$. Given this complexity, K-means LTS is suitable for medium-sized datasets, balancing efficiency and computation.

2.3.3. Clustering evaluation

Clustering results are evaluated using the silhouette score, suitable for datasets lacking a training set for model evaluation [26]. The silhouette score measures the clustering quality by comparing an element's average distance to members of its cluster $a(x_i)$ and the nearest other cluster $b(x_i)$. The score ranges from -1 to 1, with high positive values indicating well-clustered elements, while negative values suggest elements might be incorrectly clustered [27]. A maximal silhouette score $(s(x_i) = 1)$ is achieved when elements are close to their cluster and far from others.

2.4. MTRF

2.4.1. Hyperparameter tuning

Hyperparameter tuning is crucial in ML modeling, especially for tree-based models with numerous parameters [28]. It refers to building an optimal model by configuring hyperparameters through a search strategy [29]. This study uses Grid search for hyperparameter tuning in the MTRF model. Grid search is a decision-theoretical approach involving exhaustive search over a set range of hyperparameter values [30].

Grid search uses cross-validation (CV) for each combination of parameters. The data is split into 10 parts, *'n_splits'* and for each part, one section is used as test data, while the rest is training data. After testing all parameter combinations, Grid search selects the best combination based on the evaluation score 'scoring'. This study uses F1-micro and F1-macro scores as the evaluation metrics for multi-target modeling [31]. This study uses four that are significant for model performance. *'n_estimators'* refers to the number of trees, aiming to boost model performance by reducing variance. *'max_depth'* controls the maximum depth of each decision tree. Deeper trees are more prone to overfitting [32], so this study limits the depth to 10, 20, and 30. *'min_sample_split'* sets the minimum number of samples required to split a node, and *'min_sample_leaf'* [33] controls the minimum number of samples required at each leaf node. The values [1], [2], [4] balance capturing granular patterns and preventing overfitting.

2.4.2. MTC process

The classification model is a crucial step in this study, where ML algorithms are applied to predict and categorize alum data based on their waiting time for a job and the linearity of their career. This study uses the MTRF since two target variables need to be classified. According to [34], MTC works as follows,

– The training dataset consists of triplets (x_i, t_j, y_{ij}), where $y_{ij} \in Y$ describes the relationship between instance x_i and target t_j.
– During training, n different instances and m different targets, where n and m are finite numbers. The scores y_{ij} from the training data are arranged in an $n \times m$ matrix, often incomplete due to missing values
– The score set Y is one-dimensional and consists of nominal, ordinal, or fundamental values.
– This process aims to predict scores for each instance-target pair (x, t) in the set $X \times T$.

2.4.3. Classification evaluation

Evaluating multi-target models cannot rely on standard classification metrics like accuracy, precision, recall, or F1-score. In MTC, these metrics must be slightly modified, resulting in micro averaging [35] and macro averaging [36]. As explained [11], micro-averaging sums the values in the confusion matrix to produce accuracy, precision, recall, and F1-score. Macro averaging takes the average of each value in the confusion matrix. Micro averaging is useful when overall performance is prioritized, while macro averaging is preferred when each target variable is equally important.

2.5. K-fold cross validation

CV is an important ML and statistical modeling method that selects models, assesses performance, and tunes hyperparameters [37]. This method involves splitting the data into training and testing sets, training the model on the training set, and evaluating its performance with the testing set. This study uses 10-fold CV, which means there are 10-folds with different training and testing sets in each fold. The sample index allows data index to be part of the train set in fold 1 and part of the test set in fold 2. In this process, the metrics used to assess the performance of each fold are F1-micro and F1-macro scores.

3. RESULTS AND DISCUSSION

3.1. Data augmentation

As mentioned in section 3, ADASYN-gaussian uses gaussian distribution to generate synthetic data. The parameters k and beta used in this stage are k = 4 and β = 0.5, with the most optimal F1-score being 0.8405. Unlike SMOTE [38], ADASYN-gaussian focuses on generating synthetic data based on its density. Therefore, the number of minority synthetic data is not equal to the majority data. In this study, balancing the data between males and females is not as important because only female data is used for modeling, as mentioned in section 3. Figure 3 shows the overlapping histogram visualization for the features *'gpa'* and *'waiting_time'*. In the visualization of the *'gpa'* feature, the synthetic data generated has a higher density near a GPA of 4.0. Both distributions have similar shapes, showing that the synthetic data follows the original distribution. For the *'waiting_time'* feature, most of the original data is concentrated around a waiting time of 0 – 5 months. Both distributions (original and synthetic data) taper off similarly as the waiting time increases beyond 20 months.

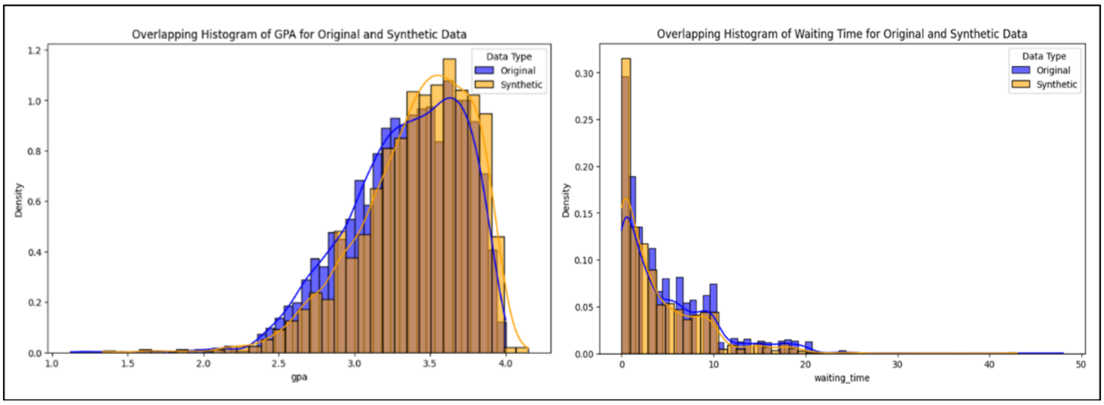

Figure 3. Overlapping histogram for 'gpa' feature (left) and 'waiting_time' feature (right)

3.2. Clustering process

One of the benefits of using K-means LTS in clustering modeling is that K-means LTS is a robust model against outliers. The presence of outliers in the data is substantial, especially in the *'GPA,'* *'length_study,'* and *'waiting_time'* features. Using k = 2 and a trimming percentage of 30%, as described in section 3, the number of outliers is significantly reduced. However, a considerable number of outliers remain in the *'length_study'* which is expected, as the average study duration for female alumni in this dataset is 4 years, a standard timeframe for undergraduate studies in Indonesian universities.

Figure 4 displays the visualization of the plot results for each formed cluster. In the 2D plot, clusters 0 and 1 are clearly well-separated. Similarly, the 3D plot shows two distinct clusters positioned closely but still separable. With $k = 2$ and a 30% trimming rate, the K-means LTS model achieves a silhouette score of 0.77, indicating strong clustering performance. Although further optimization is possible, this score demonstrates sufficient separation between the two clusters. The clusters formed can be interpreted as follows,

- Cluster 0 (544 members): a group of female alumni with long job waiting times, lower GPAs than Cluster 1 (average GPA 3.24), and a study duration slightly longer at an average of 4.13 years.
- Cluster 1 (1,483 members): a group of female alumni with short job waiting times, higher GPAs (average GPA 3.6), and a study duration of 4.11 years.

For comparison, applying standard K-means clustering with the same number of clusters $(k = 2)$ results in a lower silhouette score of 0.63, indicating less distinct separation. This comparison emphasizes the effectiveness of K-means LTS in handling outliers and improving clustering quality within this dataset.

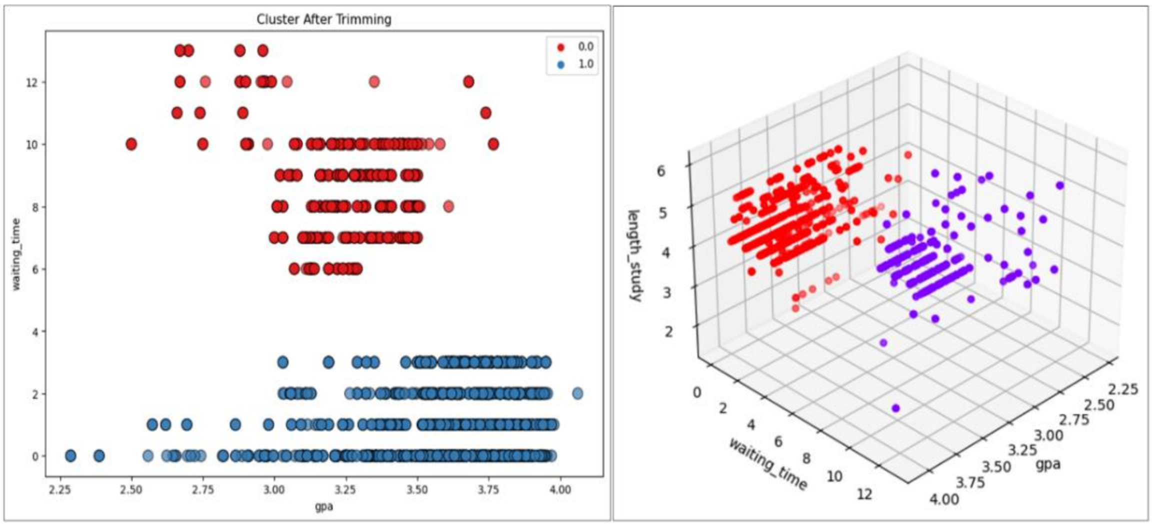

Figure 4. Visualization of the formed clusters plot, (left) plot with 2 dimensions, and (right) plot with 3 dimensions

3.3. MTC process

The best model architecture from the hyperparameter tuning performed *'max_depth'* = 20, *'n_estimators'* = 300, *'min_samples_leaf'* = 1, and *'min_sample_split'* = 2. From this architecture, the best score achieved is 92% or 0.9188. Table 1 shows the results of each metric used, and the micro average provides a higher result than the macro average.

Table 1. MTRF modelling results of hyperparameter tuning

Metrics	Micro average	Macro average
Precision	0.930612	0.912371
Recall	0.913828	0.894089
F1-score	0.922144	0.903203

The above table combines the results for both target variables, linearity and waiting time. From calculation of confusion matrix using the accuracy formula $((TP + TN)/(TP + F = +TN + FN)))$, the accuracy for the *'linearity'* target feature is 81% or 0.809, and for the *'waiting_time'* feature, it is 100% or 1.0.

3.4. 10-fold cross validation

From the previous hyperparameter tuning results, the 10-fold validation process used the *best_estimator_* and was saved as the best MTRF model for this study. The data is used for CV evaluation after augmentation and clustering processes. Hence, the required features are already available. The data will be split into 10-folds with equal proportions of train and test in each fold. Table 2 shows the evaluation results of F1-score s, both micro and macro, for each fold generated during the CV process. The best values were achieved in the seventh fold, with a micro score of 0.939 and a macro score of 0.929. The consistency of scores between folds is critical in assessing model stability [37]. The values generated in Table 2 show consistency, indicating that the MTRF model is stable and provides satisfying evaluation results.

Table 2. F1-micro and F1-macro results of each fold

n-fold	Micro average	Macro average
1-fold	0.91017964	0.8902439
2-fold	0.93279022	0.92105263
3-fold	0.91170431	0.8919598
4-fold	0.91836735	0.90654206
5-fold	0.91053678	0.88341969
6-fold	0.92337165	0.9
7-fold	0.93962264	0.92920354
8-fold	0.92116183	0.9
9-fold	0.91358025	0.890625
10-fold	0.93256262	0.91935484

4. CONCLUSION

This study investigated the modeling of clustering and MTC to analyze job waiting time and job linearity among female STEM alums in Indonesia. Using K-means LTS for clustering, we conducted hyperparameter tuning and found the optimal number of clusters k = 2 with a trimming percentage of 30%, yielding a silhouette score of 0.77. Based on these results, we identified two distinct alum groups: those with fast and long job waiting times. Notably, 73% of female alums belonged to the fast job waiting group. The resulting clusters were further used in a MTRF classification model to predict both job waiting time category and job linearity. The best performance was achieved using a tuned RF classifier with parameters max_depth = 20, n_estimators = 300, min_samples_leaf = 1, and min_samples_split = 2. Evaluation using 10-fold CV revealed strong and consistent results, with F1-micro reaching 94% and F1-macro reaching 93% in the best fold. The classification of job waiting time yielded exceptionally high performance, achieving 100% accuracy, precision, recall, and F1-score. In contrast, the job linearity classification achieved lower accuracy (81%) due to the complexity of that target and the limited features available.

Compared to previous works that applied fuzzy logic to classify job outcomes and that used GAs for optimization, our approach demonstrated higher interpretability and improved classification performance through a combination of outlier-robust clustering (K-means LTS) and multi-target modeling. Additionally, while many prior studies treated classification tasks independently, this work highlights the value of modeling multiple interrelated outcomes simultaneously using MTRF. This study, however, is not without limitations. First, the features used to predict job linearity may not fully capture the real-world factors influencing job relevance, such as job type, industry alignment, or personal career preferences. The dataset was also limited to female graduates from a small sample of private universities, which may affect the generalizability of the findings.

Future studies should consider expanding the feature set with richer academic and employment history data, including qualitative information from surveys or interviews. Additionally, applying alternative multi-target learners such as deep neural networks or ensemble methods may improve generalizability, especially for more complex targets like job linearity. In summary, this research demonstrates that combining clustering (K-means LTS) and MTRF modeling can effectively uncover patterns in graduate employment outcomes, particularly in job waiting time. Despite the underrepresentation of women in STEM, the data reveal that a significant portion of female alumni successfully enter the workforce quickly and show strong academic performance. These findings support the notion that female STEM graduates possess competitive capabilities and resilience in the job market.

ACKNOWLEDGMENTS

This research was conducted under the auspices of Bina Nusantara University. I sincerely thank the faculty, staff, and all contributors for their valuable support and assistance. This paper also utilized AI tools to assist with translation.

FUNDING INFORMATION

This research received no grant from any funding agency in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Nadzla Andrita Intan Ghayatrie	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Devi Fitrianah	✓	✓		✓		✓	✓	✓		✓		✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY

The data used in this study are not publicly available due to strict confidentiality agreements with multiple partner universities. Sharing of this data is restricted as per the signed agreements and institutional policies.

REFERENCES

[1] UNDP, "Gender inequality index," pp. 156–159, 2013, doi: 10.18356/c791776a-en.
[2] W. E. Forum, "Global gender gap report," *WEF*, 2023, [Online]. Available: https://www3.weforum.org/docs/WEF_GGGR_2023.pdf.
[3] P. P. Liang, A. Zadeh, and L. P. Morency, "Foundations and trends in multimodal machine learning: principles, challenges, and open questions," *ACM Computing Surveys*, vol. 56, no. 10, 2024, doi: 10.1145/3656580.
[4] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, 2022, doi: 10.1016/j.gltp.2022.04.020.
[5] C. Khosla and B. S. Saini, "Enhancing performance of deep learning models with different data augmentation techniques: a survey," *Proceedings of International Conference on Intelligent Engineering and Management, ICIEM 2020*, pp. 79–85, 2020, doi: 10.1109/ICIEM48762.2020.9160048.
[6] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: adaptive synthetic sampling approach for imbalanced learning," *Proceedings of the International Joint Conference on Neural Networks*, pp. 1322–1328, 2008, doi: 10.1109/IJCNN.2008.4633969.
[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.
[8] L. Rokach and O. Maimon, "Clustering methods," *Data Mining and Knowledge Discovery Handbook*, pp. 321–352, 2006, doi: 10.1007/0-387-25465-x_15.
[9] D. Fitrianah and Y. Harwani, "Alumni absorption prediction framework based on machine learning (in *Indonesian*)," *Engineering, MAthematics and Computer Science (EMACS) Journal*, vol. 2, no. 1, pp. 21–28, 2020, doi: 10.21512/emacsjournal.v2i1.6251.
[10] Y. Jing and J. Wang, "Influencing factors analysis of college employment based on fuzzy clustering," *Advances in Transdisciplinary Engineering*, vol. 47, pp. 1206–1213, 2024, doi: 10.3233/ATDE231305.
[11] N. A. I. Ghayatrie, W. P. Sari, D. Fitrianah, and E. Chew, "Analyzing STEM women's employability: a machine learning framework with clustering and multi-target classification," *ICoCSETI 2025 - International Conference on Computer Sciences, Engineering, and Technology Innovation, Proceeding*, pp. 750–755, 2025, doi: 10.1109/ICoCSETI63724.2025.11019987.
[12] T. Estella, N. Andrita Intan Ghayatrie, and A. Wibowo, "Outlier handling in clustering: a comparative experiment of K-means, robust trimmed K-means, and K-means least trimmed squared," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 1029–1039, 2024, doi: 10.12785/ijcds/160175.
[13] F. K. Nakano, K. Pliakos, and C. Vens, "Deep tree-ensembles for multi-output prediction," *Pattern Recognition*, vol. 121, 2021, doi: 10.1016/j.patcog.2021.108211.
[14] D. Xu, Y. Shi, I. W. Tsang, Y. S. Ong, C. Gong, and X. Shen, "Survey on multi-output learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 7, pp. 2409–2429, 2020, doi: 10.1109/TNNLS.2019.2945133.
[15] Y. Yusof, F. Baharom, N. Jamaludin, and S. Badroddin, "Binary-based classification models for Malaysian graduates employment analysis," pp. 501–513, 2025, doi: 10.1007/978-3-031-91485-0_39.
[16] R. Pico-Saltos, J. Garzás, A. Redchuk, P. Escandón-Panchana, and F. Morante-Carballo, "Role of alumni program in the prediction of career success in an ecuadorian public University," *Applied Sciences (Switzerland)*, vol. 12, no. 19, 2022, doi: 10.3390/app12199892.

[17] R. Amalia and A. Wibowo, "Prediction of the waiting time period for getting a job using the naive Bayes algorithm," *International Research Journal of Advanced Engineering and Science*, vol. 5, no. 2, pp. 225–229, 2020.
[18] M. Basgalupp, R. Cerri, L. Schietgat, I. Triguero, and C. Vens, "Beyond global and local multi-target learning," *Information Sciences*, vol. 579, pp. 508–524, 2021, doi: 10.1016/j.ins.2021.08.022.
[19] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, and G. Li, "Computational intelligence enabling the development of efficient clinical decision support systems: Case study of heart failure," *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications*, pp. 123–133, 2018, doi: 10.1016/B978-0-12-813314-9.00005-0.
[20] J. Hunter, D. Dale, E. Firing, and M. Droettboom, "matplotlib.pyplot.boxplot," *matplotlib*.
[21] X. Wang, J. Xu, T. Zeng, and L. Jing, "Local distribution-based adaptive minority oversampling for imbalanced data classification," *Neurocomputing*, vol. 422, pp. 200–213, 2021, doi: 10.1016/j.neucom.2020.05.030.
[22] T. Dao, A. Gu, A. J. Ratner, V. Smith, C. De Sa, and C. Ré, "A kernel theory of modern data augmentation," *36th International Conference on Machine Learning, ICML 2019*, vol. 2019-June, pp. 2755–2775, 2019.
[23] Z. Duan, C. Wang, and W. Zhong, "SSGCL: Simple social recommendation with graph contrastive learning," *Mathematics*, vol. 12, no. 7, 2024, doi: 10.3390/math12071107.
[24] A. A. Abdulnassar and L. R. Nair, "Performance analysis of Kmeans with modified initial centroid selection algorithms and developed Kmeans9+ model," *Measurement: Sensors*, vol. 25, 2023, doi: 10.1016/j.measen.2023.100666.
[25] O. Dorabiala, J. N. Kutz, and A. Y. Aravkin, "Robust trimmed K-means," *Pattern Recognition Letters*, vol. 161, pp. 9–16, 2022, doi: 10.1016/j.patrec.2022.07.007.
[26] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, 2021, doi: 10.3390/e23060759.
[27] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*, 2020, pp. 747–748, doi: 10.1109/DSAA49011.2020.00096.
[28] F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated machine learning*. Cham: Springer International Publishing, 2019.
[29] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020, doi: 10.1016/j.neucom.2020.07.061.
[30] F. Abbas *et al.*, "Optimizing machine learning algorithms for landslide susceptibility mapping along the karakoram highway, gilgit baltistan, pakistan: a comparative study of baseline, bayesian, and metaheuristic hyperparameter optimization techniques," *Sensors*, vol. 23, no. 15, 2023, doi: 10.3390/s23156843.
[31] S. Wu, Y. Chen, Z. Li, J. Li, F. Zhao, and X. Su, "Towards multi-label classification: next step of machine learning for microbiome research," *Computational and Structural Biotechnology Journal*, vol. 19, pp. 2742–2749, 2021, doi: 10.1016/j.csbj.2021.04.054.
[32] B. P. Koya, S. Aneja, R. Gupta, and C. Valeo, "Comparative analysis of different machine learning algorithms to predict mechanical properties of concrete," *Mechanics of Advanced Materials and Structures*, vol. 29, no. 25, pp. 4032–4043, 2022, doi: 10.1080/15376494.2021.1917021.
[33] K. Hoffman, J. Y. Sung, and A. Zazzera, "Multi-output random forest regression to emulate the earliest stages of planet formation," *2021 IEEE Systems and Information Engineering Design Symposium, SIEDS 2021*, 2021, doi: 10.1109/SIEDS52267.2021.9483749.
[34] W. Waegeman, K. Dembczyński, and E. Hüllermeier, "Multi-target prediction: a unifying view on problems and methods," *Data Mining and Knowledge Discovery*, vol. 33, no. 2, pp. 293–324, 2019, doi: 10.1007/s10618-018-0595-5.
[35] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: an overview," 2020, [Online]. Available: http://arxiv.org/abs/2008.05756.
[36] K. Takahashi, K. Yamamoto, A. Kuchiba, and T. Koyama, "Confidence interval for micro-averaged F 1 and macro-averaged F 1 scores," *Applied Intelligence*, vol. 52, no. 5, pp. 4961–4972, 2022, doi: 10.1007/s10489-021-02635-5.
[37] L. A. Yates, Z. Aandahl, S. A. Richards, and B. W. Brook, "Cross validation for model selection: a review with examples from ecology," *Ecological Monographs*, vol. 93, no. 1, 2023, doi: 10.1002/ecm.1557.
[38] J. Brandt and E. Lanzén, "A comparative review of SMOTE and ADASYN in imbalanced data classification," *2021*, p. 42, 2020, [Online]. Available: https://www.diva-portal.org/smash/record.jsf?pid=diva2:1519153.

BIOGRAPHIES OF AUTHORS

Nadzla Andrita Intan Ghayatrie received his first degree from Bina Nusantara University, Computer Science, Jakarta, in 2024. She has also master's degree from Bina Nusantara University, Jakarta, Indonesia, in 2025. Her research interests center around addressing gender gap issues through innovative applications of data science, ML, and AI. She can be contacted at email: nadzla.ghayatrie@binus.ac.id.

Devi Fitrianah is a lecturer and researcher at the Master of Computer Science Department, Bina Nusantara University. She earned her B.Sc. in Computer Science (2000) from Bina Nusantara University, and her M.IT (2008) and Ph.D. in Computer Science (2015) from Universitas Indonesia. In 2014, she joined the PRIPGIS Lab, Michigan State University, USA, through a sandwich program, and is currently a fellow researcher at Eureka Robotics Lab, Cardiff Metropolitan University, UK. Her research interests include data mining, ML, AI, and applied remote sensing. She can be contacted at email: devi.fitrianah@binus.ac.id.