INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Comprehensive Study on Employee Promotion Using Classification Techniques

Deepali S. Akolkar*, Shubham S. Kand

Department of Statistics, Dr. D. Y. Patil Arts, Commerce & Science College, Pimpri, Pune 411018, Maharashtra, India

DOI: https://doi.org/10.51583/IJLTEMAS.2025.1413SP039

Received: 26 June 2025; Accepted: 30 June 2025; Published: 25 October 2025

Abstract: Promoting employees is an essential procedure in organizational frameworks that directly affects motivation, productivity, and retention of the workforce. This research investigates data-centric approaches for forecasting employee advancements through classification methods. A collection of 1,000 employee records was examined, featuring variables with 12 like education, age, training scores, last year's rating, and department. Following preprocessing to manage absent values and encode categorical variables, models such as Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Gaussian Naive Bayes (GNB) were created and assessed. Performance metrics including accuracy, precision, recall, specificity, and F1 score were utilized to evaluate model results. GNB proved to be the best model, achieving an accuracy of 83% on the test data, demonstrating resilience despite class imbalance. The study finds that statistical learning methods can greatly assist human resource departments in making informed, fair, and efficient decisions regarding promotions.

Keywords: Employee Promotion, Classification, Gaussian Naive Bayes, Human Resource Analytics, Data Mining, KNN, Logistic regression, Graphical visualization

I. Introduction

Employee promotion refers to the advancement in position with increased responsibilities, benefits, and authority. A well-structured and transparent promotion policy boosts employee morale, productivity, and organizational loyalty. This paper aims to harness classification-based machine learning techniques to identify patterns and predictors of promotion within a company, facilitating objective decision-making. The information used by secondary data.

Objectives

To Identify factors influencing promotions.

To Compare classification algorithms for promotion prediction.

To develop interpretable and accurate models for promotion forecasting.

II. Methodology:

Data Description:

The dataset used for the analysis is publicly available and was accessed via Google Drive. It consists of 1,000 rows and 12 columns (department, region, education, gender, No. of trainings, age, Previous year rating, length of service, awards won, average training score and is promoted) with each row representing an individual record (likely an employee or candidate). The primary focus of the dataset is the target variable is promoted, which is a binary classification label:

0: The individual was not promoted. 1: The individual was promoted.

Data Preprocessing:

Before applying machine learning models, several preprocessing steps were carried out to ensure data quality and consistency, Handling Missing Values, Missing data was imputed using appropriate statistical methods like Mean for numerical attributes, Mode for categorical variables, Median for skewed numerical attributes. All categorical features were converted into numeric format using Label Encoding, allowing the models to interpret these variables effectively. Any duplicate records present in the dataset were identified and removed to avoid biased or redundant training.

Modeling Techniques:

A variety of machine learning classification algorithms were applied to predict whether an individual would be promoted:

- Logistic Regression: A linear model suitable for binary classification, often used as a baseline model.
- Decision Tree: A non-linear model that splits data based on feature values, offering interpretability and flexibility.
- K-Nearest Neighbors (KNN): A distance-based algorithm that classifies based on the majority class among the k-nearest neighbors.
- Support Vector Machine (SVM): A powerful classifier that finds the optimal hyperplane to separate classes, especially

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

effective in high-dimensional spaces.

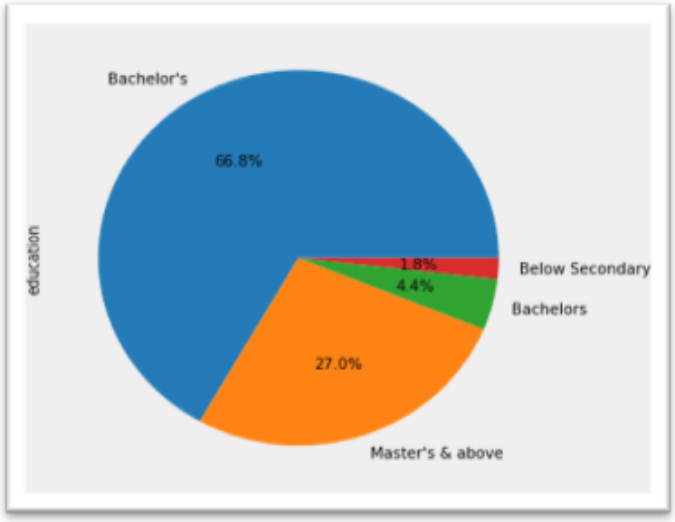
- Gaussian Naïve Bayes: A probabilistic classifier based on Bayes' theorem, assuming feature independence and normal distribution of features.

Evaluation Metrics

To assess the performance of each model, multiple evaluation metrics were used:

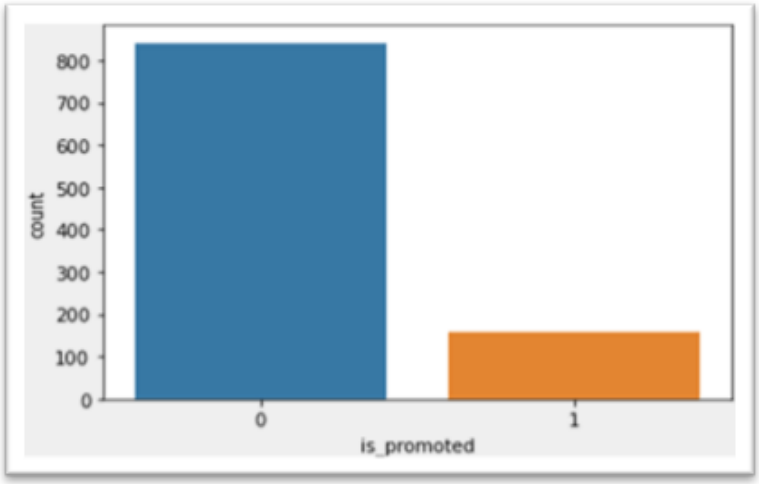
- Accuracy: The proportion of total correct predictions (both promoted and not promoted).
- Precision: The proportion of correctly predicted promotions out of all predicted promotions (focuses on false positives).
- Recall (Sensitivity): The proportion of actual promotions that were correctly predicted (focuses on false negatives).
- Specificity: The proportion of actual non-promoted individuals that were correctly identified.
- F1 Score: The harmonic mean of precision and recall, useful for balancing both metrics, especially when dealing with class imbalance.

Data Visualization:

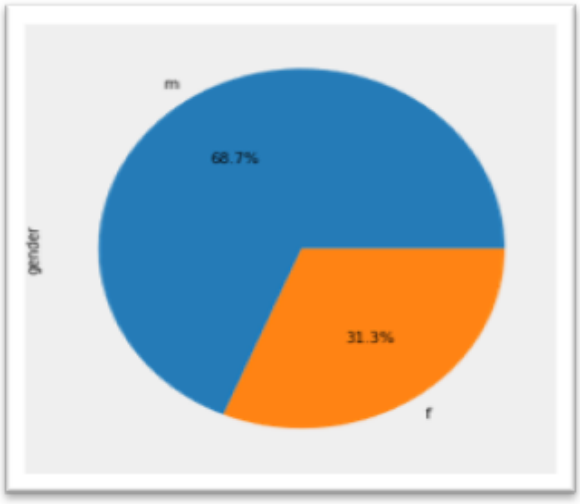


Education wise employees' chart

Department wise employees' chart

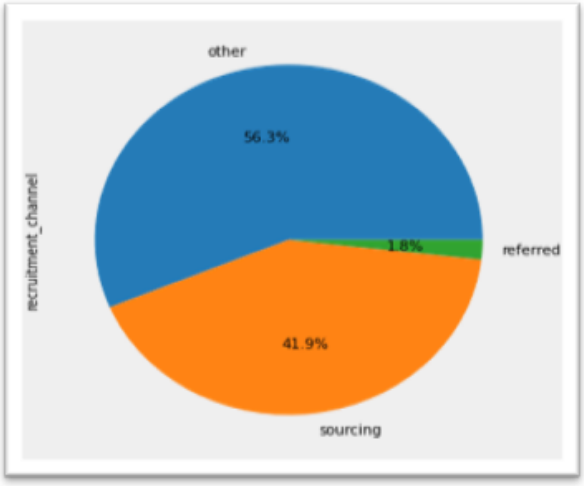


Count plot for is Promoted

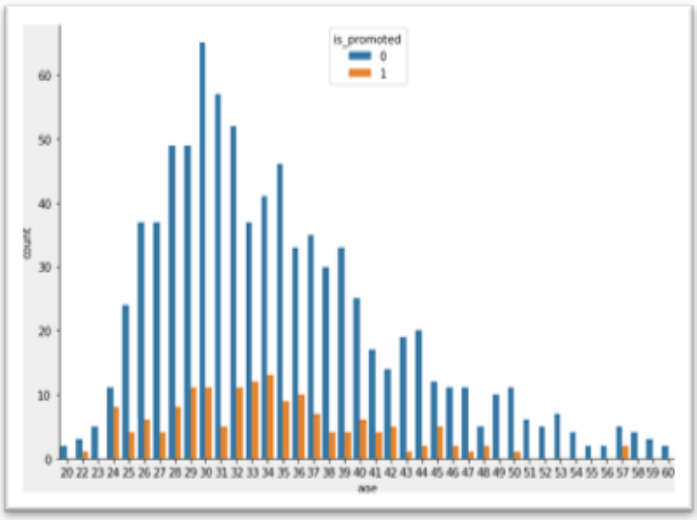


Gender wise Employees chart

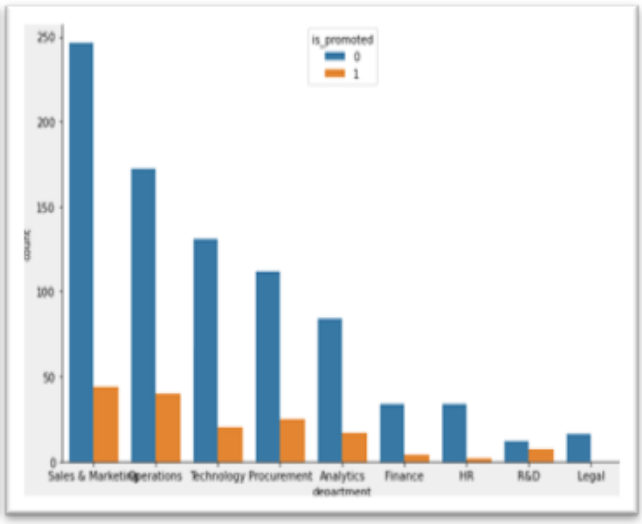
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025



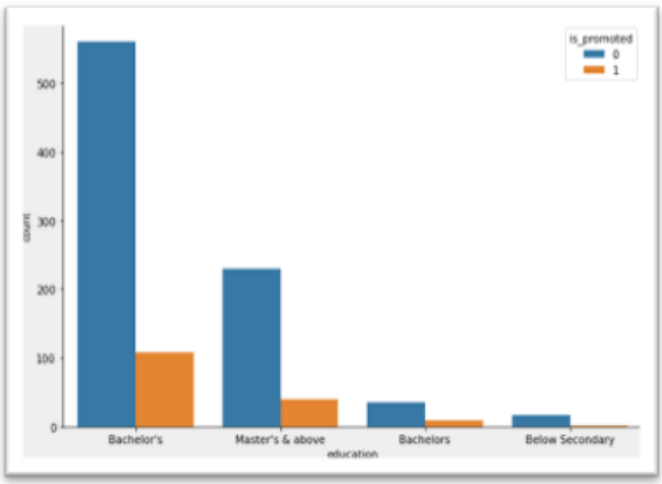
Recruitment Channel wise chart



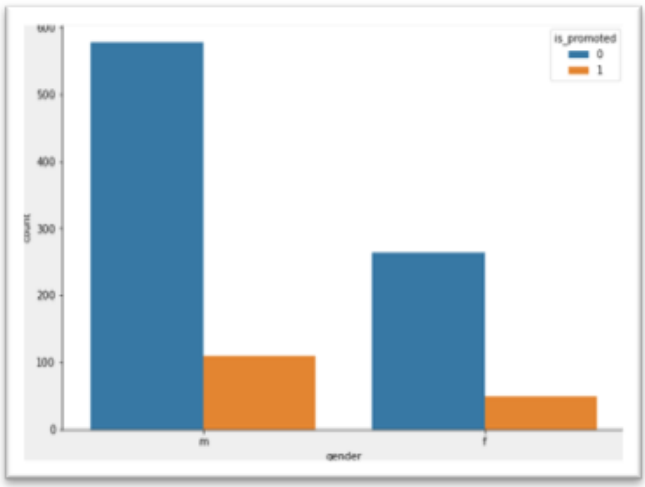
Age wise promotion chart



Department wise promotion chart

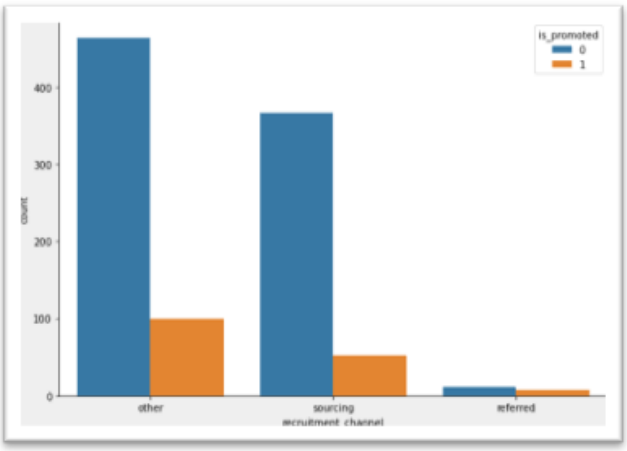


Education wise promotion of employee chart

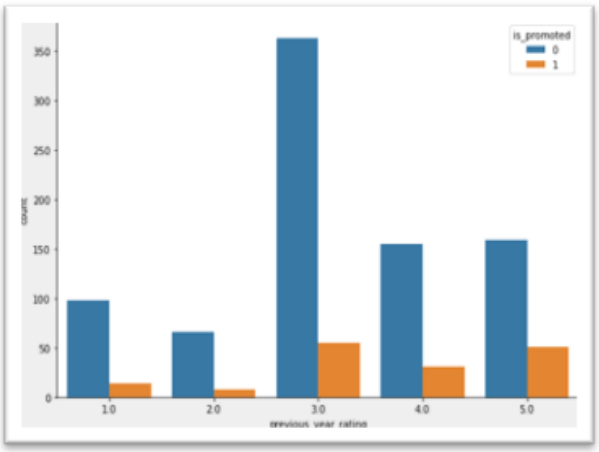


Gender wise promotion of employee chart

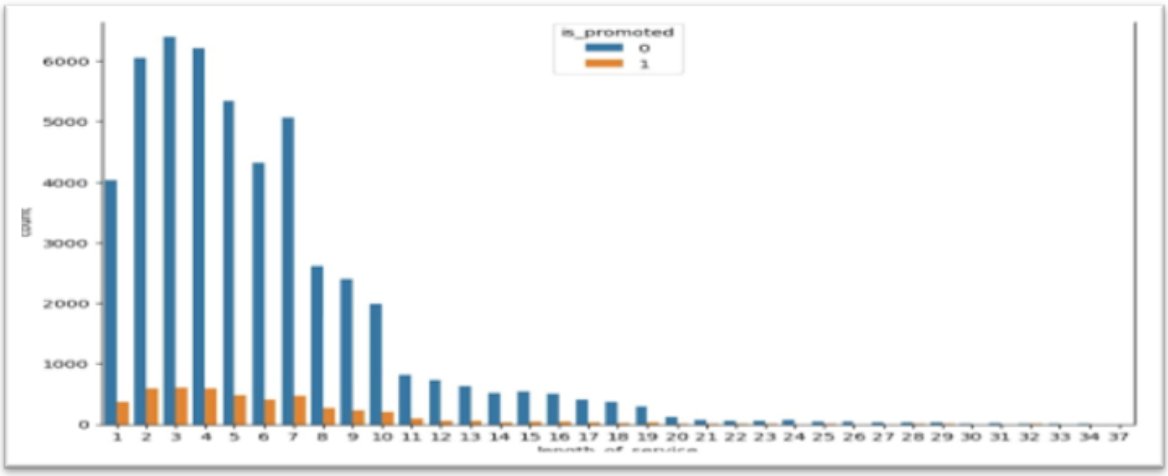
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025



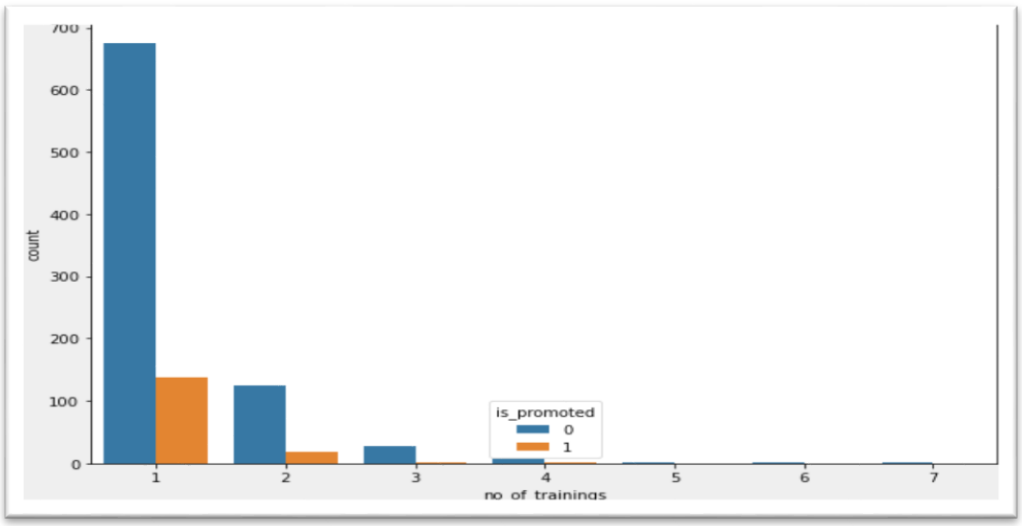
Recruitment channel wise promotion of employee chart



previous year rating wise promotion of employee chart



length of service wise promotion of employee's chart



Number of training wise promotion of employee chart

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Data Analysis:

Logistic Regression:

Classification Report:

Classificaton Report			precision	recall
0	0.85	1.00	0.92	682
1	0.00	0.00	0.00	118
accuracy			0.85	800
macro avg	0.43	0.50	0.46	800
weighted avg	0.73	0.85	0.78	800

Here, we can see that accuracy of logistic regression is 85%. It means 85% predicted values are corrected.

Decision tree:

The decision tree is an effective supervised machine learning method used for both regression and classification tasks. It is a graphical visual assistance for making predictions or decisions that represents a group of laws. Decision trees come in handy when it comes to staff advancement. due to their ability to identify complex relationships and patterns in the data and provide clear and easy to understand regulations. The decision tree algorithm splits the data recursively according to the values of the input features. Starting with a root node that represents the entire dataset, the dataset is subdivided at each internal node based on unique feature values.

III. Results:

The results we got after using the decision tree are as follows:

=100

Specificity=TN/(TN+FP)

=[118/(118+0)]=100

We got accuracy of the model using decision tree as 100%

Here, we can see that accuracy of decision tree is 100%. It means 100% predicted values are correct.

Classificaton Report			precision	recall
0	1.00	1.00	1.00	682
1	1.00	1.00	1.00	118
accuracy			1.00	800
macro avg	1.00	1.00	1.00	800
weighted avg	1.00	1.00	1.00	800

KNN:

Classificaton Report			precision	recall
0	0.79	0.96	0.87	159
1	0.12	0.02	0.04	41
accuracy			0.77	200
macro avg	0.46	0.49	0.45	200
weighted avg	0.66	0.77	0.70	200

Here, we can see that accuracy of KNN is 77%. It means 77% predicted values are correct.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Gaussian Naïve Bayes

Gaussian Naive Bayes (GNB) is a probabilistic classifier based on Bayes' Theorem with the assumption of normal (Gaussian) distribution for continuous features. It's called "naive" because it assumes independence between every pair of features.

Confusion matrix:

[678 4]

[112 6]

Class 0 (majority class): 678 correctly classified (True Negatives), 4 misclassified as class 1 (False Positives)

Class 1 (minority class): 6 correctly classified (True Positives), 112 misclassified as class 0 (False Negatives)

This shows a **severe class imbalance problem** and **poor performance on class 1 Class Precision Recall F1-Score Support**

0	0.86	0.99	0.92	682
1	0.60	0.05	0.09	118

Class 0 is predicted very well (high precision and recall).

Class1 has **low recall(0.05)** →the model is **missing most positive cases**.

Precision for class 1 (0.60) is relatively okay, meaning when it predicts class 1, it's correct 60% of the time

Macro avg F1-score = 0.51 indicates poor average performance across both classes, treating them equally.

Weighted avg F1-score=0.80 is better because it's dominated by class 0(majority class).

Accuracy=0.85

Model accuracy on training and testing:

Here, we can see that GAUSSIAN BAYES CLASSIFIER is the best model for this data because, it has 83% accuracy on testing data. So if we have the data of employees then we can predict promotion of employees using GAUSSIAN BAYES CLASSIFIER

	Model	Training	Testing
0	RANDOM FOREST	75.250	82.0
1	LOGISTIC REGRESSION	84.500	82.5
2	GAUSSIAN NAIVE BAYES CLASSIFIER	83.500	83.0
3	SVC	84.500	82.5
4	XGB	75.250	79.0
5	DECISION TREE	75.250	75.5
6	KNN	81.500	80.5
7	GradientBoostingClassifier	90.125	81.5

Result

The above analysis results that after getting values of training & testing by Random Forest, logistic regression, GNB, SVC XGB, Decision tree, KNN, and Gradient boosting classifier, the Gaussian Naïve Bayes classifier got highest accuracy in testing with 83%.

IV. Conclusion

Overall, we conclude that by **Gaussian Naïve Bayes classifier** the influencing factors observed as Awards won, Number of previous years KPI, Length of Service and education level. This study demonstrates the applicability of classification techniques in predicting employee promotions. GNB proved most effective due to its balance of simplicity and performance. HR departments can integrate such models for data-driven, equitable promotion decisions. The model calculates the likelihood of an employee being promoted based on individual feature contribution, assuming feature independence.

Declaration

There are no ethical issues.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

References

1. Ansari, N., &Vora, N. (2024). “Employee Promotion Evaluation and Prediction Using Machine Learning”. Journal of Information Technology and Digital World, 6(4), 317–332.
2. Asim, Y., Raza, B., Malik, A. K., Rathore, S., & Bilal, A. (2018). “Improving the Performance of Professional Blogger's Classification”. International Conference on Engineering Technologies, Mathematics, and Computing (iCOMET), Sukkur.
3. Huang, Y., Shum, M., Wu, X., & Xiao, J. Z. (2019), Discovery of Bias and Strategic Behavior in Crowdsourced Performance Assessment. arXiv.
4. Han, J., Kamber, M., & Pei, J. (2011),"Data Mining: Concepts and Techniques."
5. Ilwani, M., Nassreddine, G., & Younis, J. (2023). “Machine Learning Application on Employee Promotion”. Mesopotamian Journal of Computer Science.
6. Kelleher, J. D., Mac Carthy, M., & Tierney, B. (2015)."Fundamentals of Machine Learning for Predictive Data Analytics."
7. Nosratabadi, S., Zahed, R. K., Ponkratov, V. V., & Kostyrin, E. V. (2022). “Artificial Intelligence Models and Employee Lifecycle Management: A Systematic Literature Review”. arXiv.
8. Purwandari, B., Ruldeviyani, Y., &Ramdhani, T. W. (2016). “The Use of Data Mining Classification Technique to Fill in Structural Positions in Bogor Local Government”. International Conference on Advanced Computer Science and Information Systems (ICACSIS), Malang.
9. Shafie, S. S., Ooi, S. P., & Khaw, K. W. (2023). “Prediction of Employee Promotion Using Hybrid Sampling Method with Machine Learning Architecture”. Malaysian Journal of Computing.
10. Subramanian, P., & Suresh, R. (2018)."Employee promotion prediction using machine learning techniques. “Proceedings of the International Conference on Computer Communication and Informatics (ICCCI).
11. Parmar, D., et al. (2020)."Predictive Analytics in Human Resource Management using Machine
12. Learning Techniques."International Journal of Computer Applications.
13. **Link:** https://drive.google.com/file/d/1DjsdcaptBZ52BXM_NoxuHA2ciMGrgmVe/view?usp=drive_link For secondary data.