Mining Model for Employees Performance Prediction

Amira Abd El Hamed Kandil [1],   Sara M. Mosaad [2],

Mohamed Haggag[3]

Abstract

Managers and decision-makers in various industries now recognize the critical importance of Human Resource Management (HRM) in identifying highly qualified employees. This study explores the application of data mining techniques in predicting employee performance, leveraging HRM practices to effectively manage talent through comprehensive employee datasets and advanced algorithms. The primary goal is to develop a classification model using data mining techniques to provide managers and HR professionals with a data-driven tool for enhancing talent management and optimizing employee placement.

Previous studies often relied on intuition or anecdotal evidence rather than rigorous data mining techniques. This research addresses this gap by implementing a data-driven approach, achieving higher accuracy rates in predicting employee performance. The research question is tackled by constructing a classification model utilizing Decision Tree

[1] - Master's researcher Business Information Systems Department, Faculty of Commerce and Business Administration, Helwan University, Cairo, Egypt

[2] - Lecturer Faculty of Commerce & Business Administration, Department of Business Information Systems, Helwan University, Cairo, Egypt

[3] - Professor of Computer Science Department Faculty of Computers & Artificial Intelligence, Helwan University, Cairo, Egypt.

(DT), Naive Bayes, and Support Vector Machine (SVM), with the implementation carried out using Rapid Miner. The model incorporates demographic and work history factors to accurately predict employee performance.

The key impact of this research is the high accuracy rates of the classification model, ranging from 85.7% to 100%, depending on the algorithm used. By identifying critical factors influencing employee performance, the model enables managers to make informed, data-driven decisions. This leads to optimized employee placement, improved identification of high-potential individuals, and overall enhanced organizational effectiveness. Additionally, the model's predictive capabilities can reduce hiring risks and improve workforce productivity and engagement.

Keywords: Data Mining Techniques, Employee Performance, Predicting Model

Amira Kandil , Sara M. Mosaad , Mohamed Haggag

تقنيات استخراج البيانات للتنبؤ بأداء الموظفين

الملخص

يدرك المديرون وصناع القرار في مختلف الصناعات الآن الأهمية الحاسمة لإدارة الموارد البشرية (HRM) في تحديد الموظفين المؤهلين تأهيلا عاليا. تستكشف هذه الدراسة تطبيق تقنيات استخراج البيانات في التنبؤ بأداء الموظفين، والاستفادة من ممارسات إدارة الموارد البشرية لإدارة المواهب بشكل فعال من خلال مجموعات بيانات الموظفين الشاملة والخوارزميات المتقدمة. الهدف الأساسي هو تطوير نموذج تصنيف باستخدام تقنيات استخراج البيانات لتزويد المديرين ومحترفي الموارد البشرية بأداة تعتمد على البيانات لتعزيز إدارة المواهب وتحسين توظيف الموظفين.

اعتمدت الدراسات السابقة في كثير من الأحيان على الحدس أو الأدلة القصصية بدلاً من تقنيات استخراج البيانات الصارمة. يعالج هذا البحث هذه الفجوة من خلال تطبيق نهج قائم على البيانات، وتحقيق معدلات دقة أعلى في التنبؤ بأداء الموظفين. تمت معالجة سؤال البحث من خلال بناء نموذج تصنيف باستخدام Decision Tree (DT)، وNative Bayes، و Support Vector Machine (SVM)، مع التنفيذ باستخدام Rapid Miner. يتضمن النموذج العوامل الديموغرافية وعوامل تاريخ العمل للتنبؤ بدقة بأداء الموظف.

التأثير الرئيسي لهذا البحث هو معدلات الدقة العالية لنموذج التصنيف، والتي تتراوح من 85.7% إلى 100%، اعتمادًا على الخوارزمية المستخدمة. ومن خلال

تحديد العوامل الحاسمة التي تؤثر على أداء الموظفين، يمكّن النموذج المديرين من اتخاذ قرارات مستنيرة تعتمد على البيانات. ويؤدي هذا إلى تحسين توظيف الموظفين، وتحسين تحديد الأفراد ذوي الإمكانات العالية، وتعزيز الفعالية التنظيمية بشكل عام. بالإضافة إلى ذلك، يمكن للقدرات التنبؤية للنموذج أن تقلل من مخاطر التوظيف وتحسن إنتاجية القوى العاملة ومشاركتها.

الكلمات المفتاحية :تقنيات التنقيب عن البيانات، أداء الموظفين ، نموذج التنبؤ.

1. Introduction& LITERATURE REVIEW

Organizations demonstrate a keen interest in specific outcomes to assess their strengths and weaknesses within the current competitive environment. Strahmeier and Piazza (2015) [1] identify employee performance, turnover, and selection as crucial results for evaluation. Employee performance stands out as a pivotal element in various HRM processes and a key focus for organizations, as emphasized by Campbell and Wiernik (2015). The diverse applications of performance assessments underscore corporations' keen attention to their employees' performance appraisals. Campbell and Wiernik further highlight that research, legal compliance, advancement, compensation, high-stakes evaluations, feedback, and enhancement represent the primary objectives of performance appraisals. For example, employee performance evaluations play a vital role in guiding performance reviews and protect the legality of hiring decisions (Wiernik, B. M., & Campbell, J. P. (2015)) [2]. Organizations are very interested in more useful methods of assessing employee performance because of the previously mentioned reasons. Decision-makers in any organization should identify the important variables that could influence worker performance in a way that promotes the sustainability of the company. Managers employed a variety of approaches and procedures to identify these factors. Questionnaires are used by certain supervisors and others to assess employee performance.

Managers explore plausible factors that could lead proficient employees to resign from their positions through the utilization of statistical or mathematical frameworks. Furthermore, decision-makers utilize sophisticated machine learning and categorization algorithms. Certain decision-makers perceive the latter two approaches as superior due to their impartiality towards human involvement. The following study will introduce a classification model aimed at aiding decision-makers in identifying the fundamental factors that impact employee performance (SN Mishra, DR Lama, 2016) [3].

Data mining is considered a promising approach for discovering information and knowledge, especially due to the multifaceted nature of issues related to employee performance (Qasim et al., 2012) [4]. Finely formulated

Mining methodologies can facilitate the retrieval and extraction of recently acquired knowledge from comprehensive datasets (Jayadi, Pugh. (2018))[5]. A range of techniques can be used in data mining to extract relevant and interesting knowledge from data, which may provide solutions to various business dilemmas.

Large information blocks are scrutinized and assessed utilizing data mining methodologies to recognize pivotal patterns and trends. Various tasks benefit from this approach, such as filtering spam emails, detecting fraudulent activities, conducting database marketing, managing credit

risks, and analyzing user sentiments or opinions. Data mining encompasses techniques like classification, forecasting, prediction, and clustering, as delineated in studies by John M. Kirimi and Christopher A. Moturi (2016) [6], as well as Merry Grace T Li, Macrina Lazo, Ariel Kelly Balan, and Joel de Goma (2021).Classification is a supervised learning technique to find hidden patterns using the learning by-example approach [7].

Clustering is the process of categorizing data through unsupervised learning techniques such as k-means and the nearest neighbor algorithm. Conversely, classification involves supervised learning approaches that utilize various algorithms like decision trees to assign data points to predefined categories. The main objective is to identify the most compelling pattern that addresses the issue, irrespective of the specific mining methodology employed (S. E. Viswapriya, 2019)[8].

Data mining techniques to predict employee performance has become an essential area of research. By exploring large-scale data from various domains, organizations aim to improve their decision-making processes and enhance employee outcomes. Let's delve into some relevant studies and models:

Mohammad NourAljarrah(2021).The impact of Human Resources-linked Enterprise Resource Planning (ERP) systems on the evaluation of employee performance was thoroughly examined in a research conducted in Jordan.

The investigation focused on the Jordanian National Electric Power Company (NEPCO) as a specific case. Data was collected from 98 managers and department heads at NEPCO. Key areas of interest in this study included system quality, information quality, service quality, and user satisfaction. Findings indicated that user satisfaction, information quality, and system quality significantly influenced performance evaluations. Conversely, service quality did not display any noticeable effect on the outcomes of performance appraisals. The research recommends a strategic emphasis on enhancing system quality, information quality, and user satisfaction by providing improved support and training to enhance the effectiveness of performance assessments [9].

Qasem A. Al-Radaideh, Eman Al Nagi (2012).The study focus on human capital, which is very important to business management. Hiring highly qualified individuals who are expected to perform exceedingly well is the goal of organizations. Data mining has gained popularity recently as a means of finding useful information for users. One hundred thirty workers completed the questionnaire. The study builds a classification model for forecasting employee performance using data mining approaches. The model is developed using the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, with decision trees serving as the main tool. There are produced several classification rules. To predict the performance of new candidates, the model is validated using real data from many

firms. Performance prediction is greatly impacted by factors including degree, grade, job position, and type of university attended. Age doesn't seem to have much of an impact, but factors like marital status, gender, pay, years of experimenting, prior work experience, and job happiness do. This methodology can help businesses anticipate how recently hired staff will perform and reduce the risk of recruiting underperforming workers. Upcoming projects entail gathering more thorough data from different businesses, including databases for both present and previous workers [10].

In the 2014 investigation conducted by H. Jantan, Norazmah Mat Y., and Mohamed Rozuan N., the Support Vector Machine (SVM) methodology was applied to the classification of employee performance.

The primary aim of this research was to assess the capability of the SVM approach in recognizing the Essential data patterns required for the classification of employee achievement. Despite the satisfactory accuracy of the model as determined by the SVM technique, further enhancements are deemed necessary [11].

Utilizing decision-making approaches and an extensive personnel database, Lipsa Sadath (2013) conducted an investigation into the potential of automating and enhancing intelligent decision-making processes. The study concluded that the C4.5 approach exhibited superior accuracy levels. The primary objective of this research was to predict

employee performance through the application of optimal Knowledge Management (KM) strategies, leading to the establishment of a resilient HR infrastructure and efficient business processes [12].

Ahmed Samir, Mona Nasr, and Essam Shaaban (2019) conducted a study where they utilized data mining techniques to develop a classification model for predicting employee performance. The data used in the study was gathered from the Ministry of Egyptian Civil Aviation (MOCA) through a survey distributed to 145 employees. The researchers employed three primary data mining strategies, including the creation of a classification model and the identification of key factors that influence performance. Support Vector Machines, Naïve Bayes, and Decision Trees (DT) were among the techniques utilized in the study.. Experiments were conducted using WEKA methodologies to create a highly accurate model for decision-makers and HR professionals to predict and improve employee performance. The author's research revealed that Support Vector Machines (SVM) outperformed other data mining techniques in predicting employee performance. Across three experiments, SVM consistently achieved an impressive accuracy rate of 86.90%. These experiments were conducted using the WEKA toolkit, a reliable platform for data analysis[13].

Amarendra Deo, Yara Mahfouz (2023). Discussedthaty many companies struggle with evaluating employee performance effectively, particularly in large organizations.

This study addresses this challenge by proposing a system specifically designed for sales companies, utilizing Bayesian classification. Their system incorporates various employee data points like attendance, pay, revenue, and even gender for performance prediction. The initial model, trained on a small dataset of 30 employees and tested on a larger dataset of 400, demonstrated an accuracy of 65%. However, the authors acknowledge limitations and suggest improvements. They point out that incorporating more data points and potentially switching to algorithms like decision trees could significantly enhance accuracy, reaching up to 95% with sufficient data. Despite limitations, this study offers a promising framework for sales companies seeking a data-driven approach to evaluate and predict employee performance. Further development and testing are recommended to refine the system for real-world implementation. By condensing the information and focusing on key aspects like the problem, solution, results, limitations, and conclusion, this paraphrase creates a clearer and more concise paragraph [14].

ZarminaJaffara, Dr. WaheedNoorb ,ZartashKanwalc (2019) propose a data mining framework to predict significant factors related to employee turnover. They consider historical behavior (such as years at the company, overtime, and performance rating), demographics (including age, monthly income, and distance from home), and attitudinal characteristics (like environment satisfaction and job satisfaction). The study evaluates various algorithms,

achieving an average accuracy rate of 88%. Notably, the most influential factors causing turnover are the monthly rate, overtime, and employee age. Overall, this research provides valuable insights for understanding and addressing employee turnover using data mining techniques [15].

The study article is composed of multiple sections. The subsequent section offers a synopsis of various previous research works utilizing data mining techniques to predict factors influencing employee performance. The following section delves into the recommended model, data pre-processing, conducted experiments, and the results derived. The concluding section encapsulates the research summary alongside the findings.

2. PROPOSED METHODOLOGY

The suggested model can be partitioned into three distinct phases: input, processing, and output, in that order. The delineation of the model is outlined below:
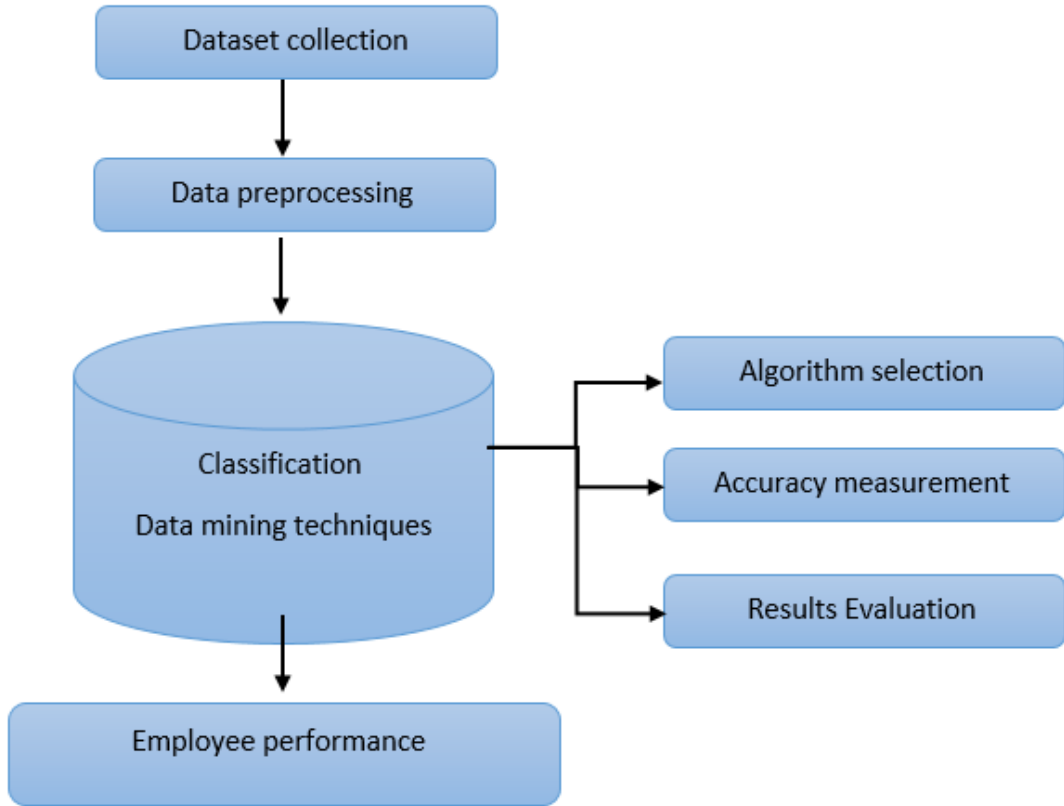
Fig. 1. Classification Model.

• **Phase1: Input**

Initially, the Information Technology department acquires the dataset containing behavioral, demographic, and attitudinal information to facilitate the prediction of Employee performance in alignment with the guidelines established by the Human Resources department. The concern of the organization lies in the attrition of employees to competitors. Subsequently, the data undergoes preprocessing, a stage in which it is manipulated and refined to ensure its usability.

• **Phase2: Processing**

To select the optimal algorithm for predictive purposes, an assessment of data mining algorithms will be conducted during this stage .Through the application of these methodologies, the prediction model uncovers latent attributes within the collected datasets. Finally, the suggested model compares the identified features to ascertain which one influences the process of employee performance prediction.

• **Phase3: Output**

The framework's output can be understood as a means of visualizing the outcome for the individuals responsible who operate as decision-makers, in the form of reports, enabling them to make the right decision at the right time.

3. CASE STUDY

The goal of this project is to develop a model for forecasting employee performance using data mining techniques to provide decision-makers in businesses with a solid answer that will enable them to make the appropriate decisions when needed. This model allows for predicting the appropriate job for the employee in addition to analyzing the performance evaluation criteria and selecting Prioritizing employee retention through appropriate measures and actions can significantly boost the company's profitability.

• Data selection and pre-processing

The initial stage in this framework involves the curation and refinement of data. Within this particular stage, we employ the HR Employee dataset from IBM to forecast employee productivity. A comprehensive overview of the dataset characteristics is presented in Table 2. Included in the dataset are details regarding the employees, as well as the variables that influence employee performance.

Behavioral such as:
- Years at the Company
- Over Time
- Performance Rating
- Years since Last Promotion
- Total Working Years

Demographics such as:

- Age
- Monthly Income
- Distance from Home
- Marital Status
- Education
- Gender

Attitudinal such as:

- Environment Satisfaction
- Job Satisfaction
- Work-Life Balance
- Number of Companies Work

Table 1. The dataset features a description

Attribute	Values
Age	18 to 60
Daily Rate	102 to 1499
Distance from Home	1 to 29
Hourly Rate	30 to 100
Monthly Income	1009 to 19973
Monthly Rate	2044 to 26999
Number of Companies Worked	0 to 9
Percent Salary Hike	11 to 25
Total Working Years	0 to 40
Training Times Last Year	0 to 6
Years at Company	0 to 36
Years in Current Role	0 to 17
Years since Last Promotion	0 to 15
Years with Current Manager	0 to 17
Department	H.R, Research &Development, Sales, IT
Education	1 to 5
Gender	Male, Female
Environment Satisfaction	1 to 4
Job Involvement	1 to 4
Job Level	1 to 5
Job Role	Manager, Sales Executive, Laboratory Technician, Research, Director, Human Recourses, Healthcare Representative, Manufacturing Director.
Job Satisfaction	1 to 4
Marital Status	Single, Married, Divorced
Over Time	Yes, No
Performance Rating	3,4
Work-Life Balance	1 to 4
Relationship Satisfaction	1 to 4

-
-
-

- **Classification**

The task of data analysis involves the process of identifying the model that represents and characterizes the layers and concepts within the data.

Classification, on the other hand, pertains to the determination of class categories based on a training dataset comprising notes that ascertain class membership.

Classification is reliant on machine learning, where entities are classified into predefined categories or groups within a dataset. The system is designed to group data elements based on classification. For instance, classification can be applied to worker records distinguishing between those who have left the company and those who have not. Various algorithms, such as the Generalized Linear Model, Deep Learning, Logistic Regression, Logistic Slope, and Support Vector Machine, were utilized to analyze the data. The study conducted tests employing classification algorithms like logistic regression, random forest, Fast Large Margin, and gradient-boosted trees using the data mining software, Rapid Miner.

- **Prediction**

Data mining techniques entail predicting or anticipating the relationship between independent and non-independent variables. Specifically, when utilizing sales forecasting methods to project future earnings, if revenue is identified as an independent variable, it may consequently become a dependent variable. Prediction, akin to classification, involves estimating the value of a numerical variable (e.g., purchase

amount) rather than a class (e.g., Purchaser or non-purchaser). While classification focuses on forecasting a Class, the concept of prediction pertains to predicting the value of a continuous variable [7].

4. RESULTS

In this part, we have used classification algorithms to know the most suitable algorithms in the process of predicting employee performance, and the most influential algorithms have been monitored that have achieved the highest percentage of accuracy such as:

Gradient Boosted Trees, Generalized Linear Model, and Decision Tree Achieved accuracy by (100 %), Naive Bayes achieved an accuracy by (99 %), Random Forest achieved an accuracy by (98.8 %), Logistic Regression Achieved accuracy by (97.9%), deep learning Achieved accuracy by (95.9 %), Support Vector Machine and Fast Large Margin Achieved accuracy by (85.7%). The results have been summarized in figure 2 for all the algorithms used and from the application and the results are summarized. The most appropriate classification algorithms have been known in the prediction process.
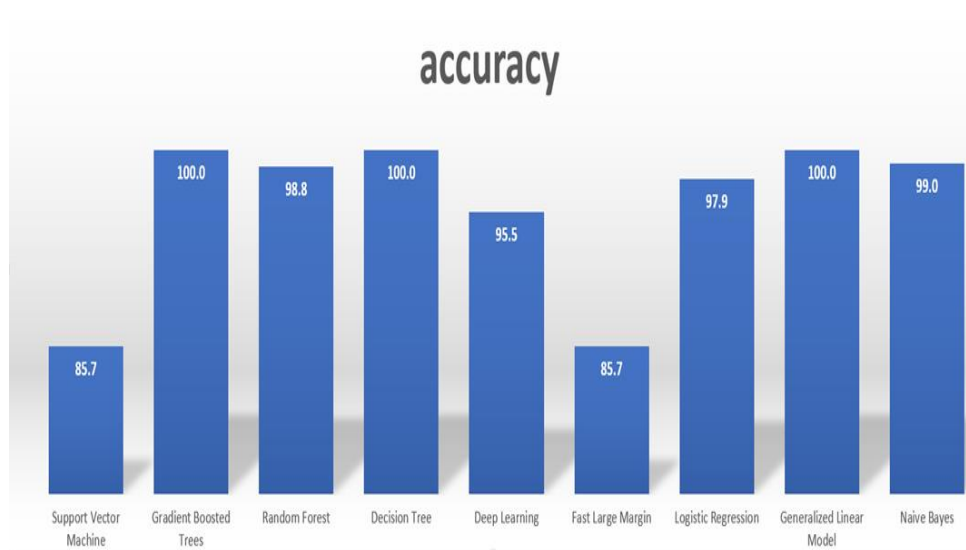
Fig. 2. The classification algorithms accuracy summary

• **Detailed Accuracy**

The accuracy of the details for each algorithm includes Recall and precision and F-measure.

• **Important factors by weights**

The paramount factors affecting the performance of employees in each algorithm are presented in this section, arranged by their relative importance. The subsequent variables are delineated as follows: Age, Monthly Income, Years at a Company, Total Working Years, and Training Time Previous Year, Department, Job Role, Education, Gender, and Years in Current

Monthly Rate. The percentage salary hike is displayed in Table 3.

Algorithm	accuracy	precision	recall	f-measure
Naive Bayes	99%	99.5%	99.4%	99.4%
Generalized Linear Model	100%	100%	100%	100%
Logistic Regression	97.9%	98.1%	99.4%	98.8%
Fast Large Margin	85.7%	85.7%	100%	92.3%
Deep Learning	95.5%	95.3%	99.7%	97.4%
Decision Tree	100%	100%	100%	100%
Random Forest	98.8%	100%	98.6%	99.3%
Gradient Boosted Trees	100%	100%	100%	100%
Support Vector Machine	85.7%	85.7%	100%	92.3%

Table 2. The Detailed algorithms Classification accuracy, F-Measure, recall, and precision

• **Age**

Table 4 demonstrates the substantial impact of age on employee performance. The mean age of the workforce is 37 years, as depicted in the diagram. The relevance of employee age to performance is depicted in (Fig. 3).

Minimum   (18)

Maximum   (60)

Average   (36.924)

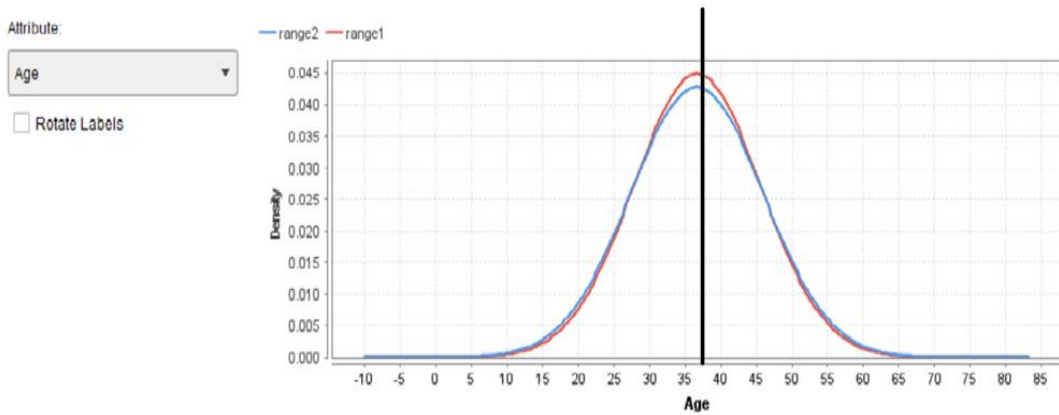## Fig. 3. The significant age for employee performance

Table 3. Significant features with weights

Attribute Algorithm	Weights				
	Age	Monthly Income	Total Working years	Years At a company	Percent Salary Hike
Naive Bayes	0.022	0.010	0.008	0.011	0.088
Generalized Linear Model	0.021	0.007	0.009	0.012	0.093
Logistic Regression	0.032	0.016	0.022	0.118	0.055
fast Large Margin	0.057	0.129	0.007	0.005	0.024
Deep Learning	0.088	0.137	0.015	0.075	0.102
Decision Tree	0.021	0.007	0.009	0.012	0.087
Random Forest	0.027	0.071	0.036	0.023	0.126
Gradient Bested trees	0.021	0.007	0.009	0.012	0.087
Support vector machine	0.00	0.001	0.001	0.00	0.00

• Monthly Income

One of the key determinants impacting employee productivity is the monthly salary received by the employee. An increase in the employee's monthly income corresponds to an enhancement in their work performance. The portrayal of the monthly income is illustrated in (Fig.4).
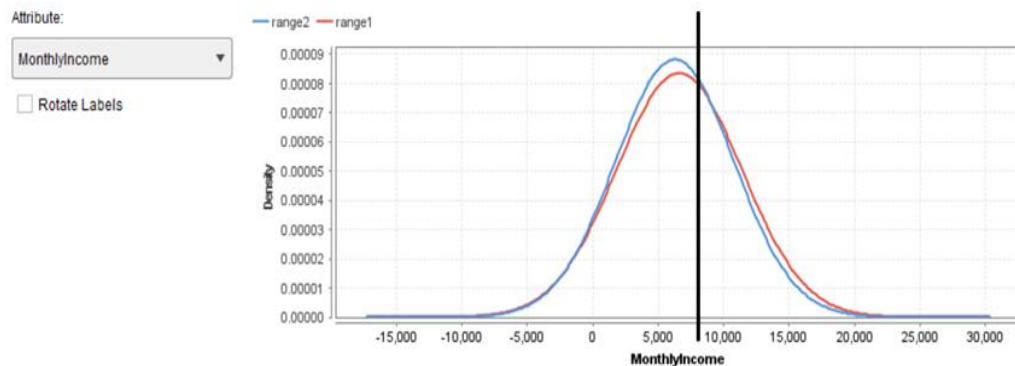
Minimum (1009)

Maximum (19999)

**Average** (6502.931)

Fig. 4. Monthly Income description.

• Total year of work

The duration of employment is a critical element influencing an individual's effectiveness in the workplace. As an employee accumulates more years within an organization, their level of professional expertise also increases. This relationship between years of work and performance is illustrated in Figure 5.

Minimum   (0)

Maximum   (40)

Average   (11.280)

Fig. 5. Working years' significant value in employee performance over
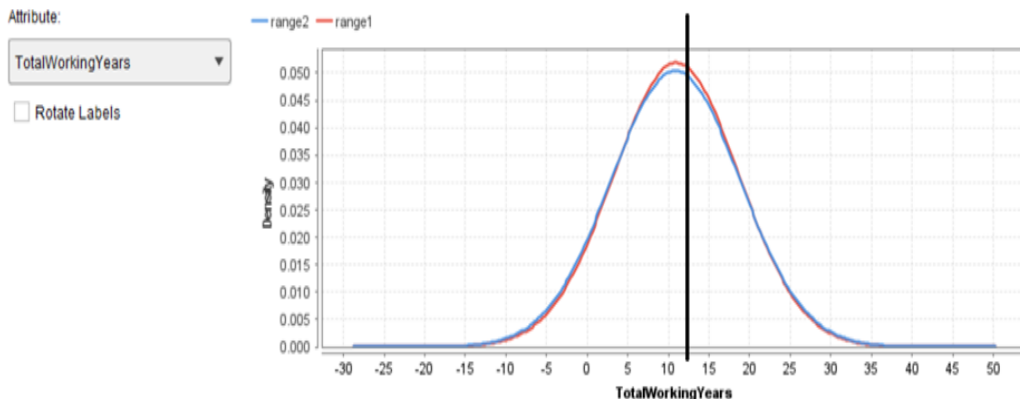
• Years at the company

The length of time spent in a job plays a crucial role in determining the efficiency of an individual in a professional setting. With each passing year within a company, an employee's competency level tends to enhance as well.

The correlation between tenure and job performance is visually represented in Figure 6.

Minimum   (0)

Maximum   (40)

Average   (7.008)

Fig.6. the significant working years at the same company
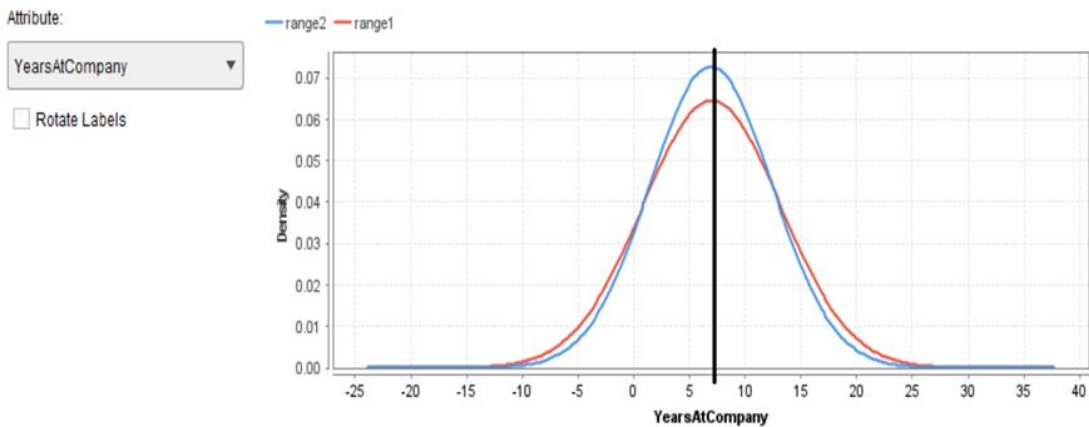
value

• **Percent salary Hike**

The Percent salary Hike is one of the most important factors that affect employee performance. The average salary increase percentage is about 18%, and the higher the percentage, the more productive the employee is. The salary percentage is described as shown in (Figure 7).

Minimum   (11)
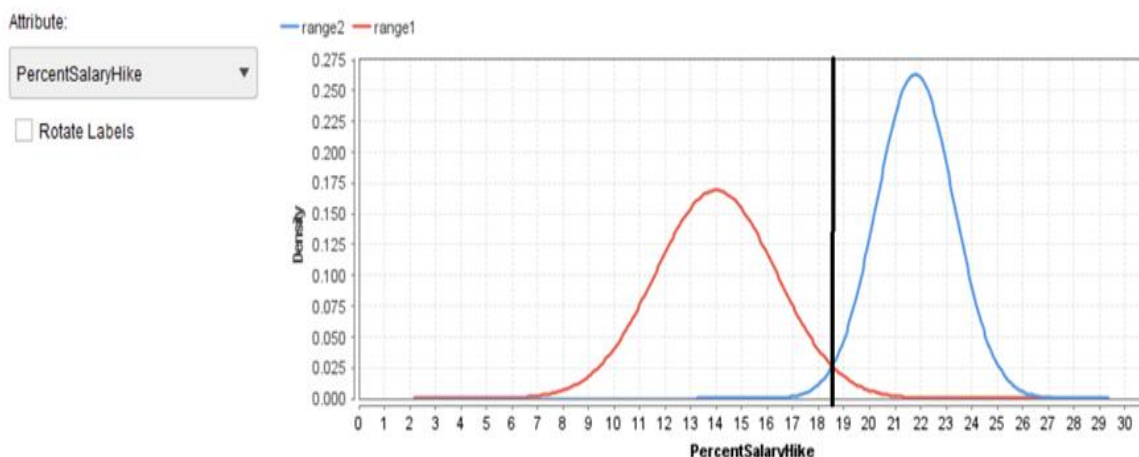
Maximum   (25)

Average   (18.210)

Fig.7. the significant percent salary Hike

Table 5. Significant features with weights

Attribute Algorithm	Weights				
	Training time last year	Departments	Job Satisfaction	Job involvement	Environment Satisfaction
Naive Bayes	0.008	0.018	0.010	0.019	0.004
Generalized Linear Model	0.008	0.015	0.007	0.020	0.003
Logistic Regression	0.026	0.016	0.019	0.042	0.001
fast Large Margin	0.092	0.004	0.007	0.009	0.098
Deep Learning	0.004	0.133	0.017	0.005	0.010
Decision Tree	0.008	0.014	0.006	0.021	0.003
Random Forest	0.028	0.024	0.012	0.074	0.018
Gradient Bested trees	0.008	0.014	0.006	0.021	0.003
Support vector machine	0.000	0.000	0.000	0.001	0.001

• **Training time last year**

Continuous training of employees leads to increased skills and efficiency, and therefore, the more training the employee has, the more competent he is in the job. The average number of Training times 3 years, Describe the training times as shown in (Figure 8).
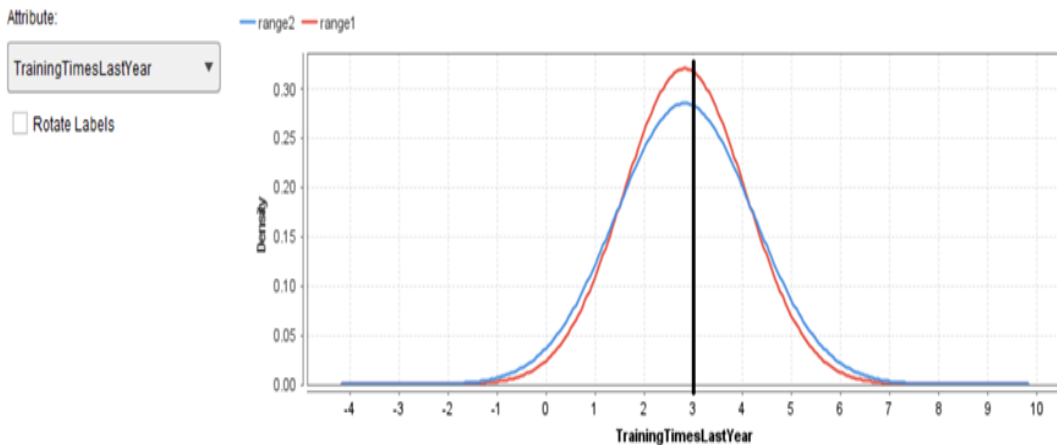
Minimum   (0)

Maximum   (6)

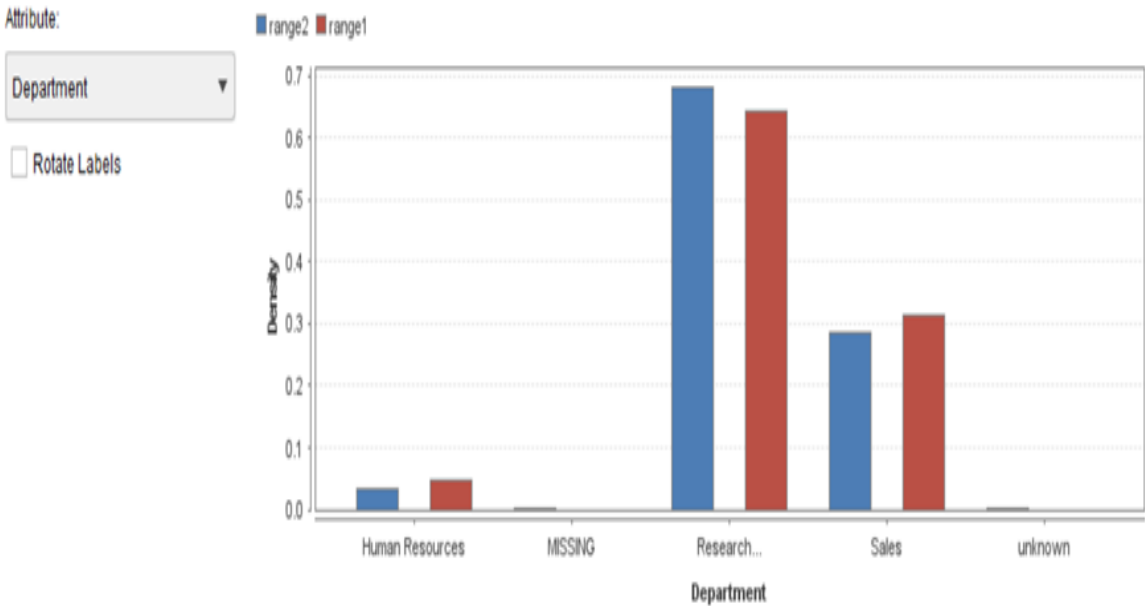Fig 8. The significant training time last year's value

- **Department**

As can be seen in (Figure 9), the department with the most   workers is the research and development department.

Research & Development   (65.37%)

Sales   (30.34%)

Human Resources   (4.29%)

Fig.9. Departments Description

• Job Satisfaction

Without question, one of the most crucial elements affecting workers' performance on the job is their level of job satisfaction. An employee is more connected to the organization and is less likely to just go to work if their job satisfaction score is greater. The description of job satisfaction is displayed in (Figure 10).

Minimum (1)

Maximum (4)

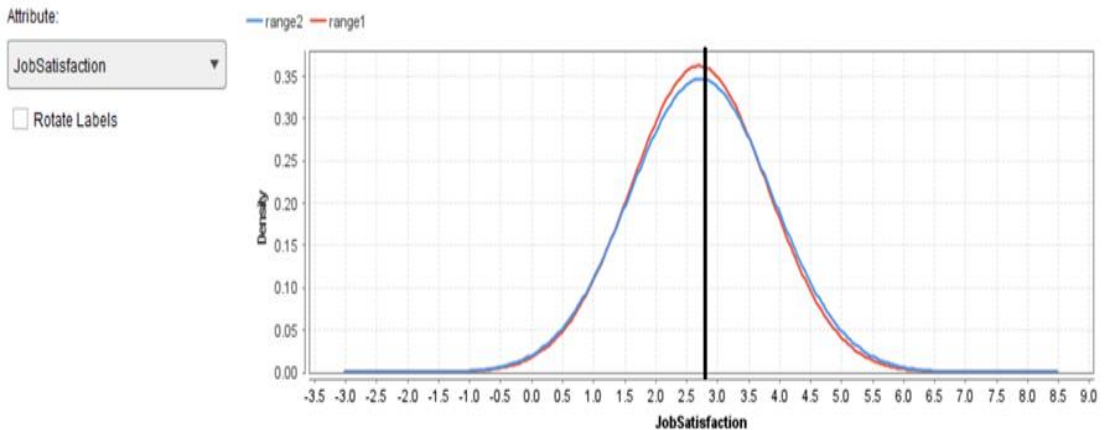**Average** (2.729)

Fig.10. Job Satisfaction significant value

• **Job involvement**

Job involvement is considered a critical factor influencing employee performance, with a direct correlation between the level of job engagement and the likelihood of employee performance. A detailed depiction of job involvement can be observed in Figure 11.

Minimum   (1)

Maximum   (4)

Average   (2.730)

Fig.11. Job involvement significant employee performance
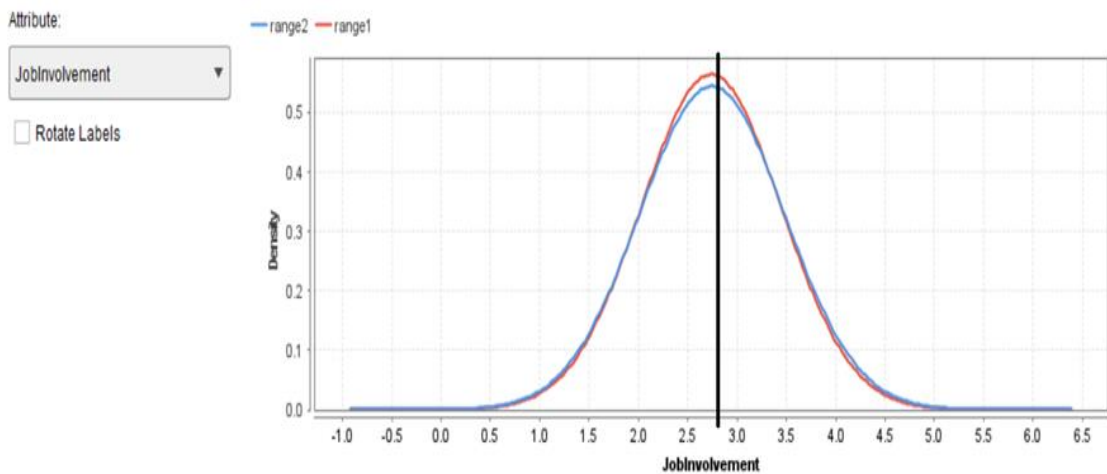
value

Environment Satisfaction

The positive correlation between satisfaction levels concerning capabilities and workplace conditions in the work environment and employee attachment to the organization, as well as the enduring impact of work, becomes more pronounced as satisfaction levels increase, as illustrated in Figure 12.

Minimum   **(1)**
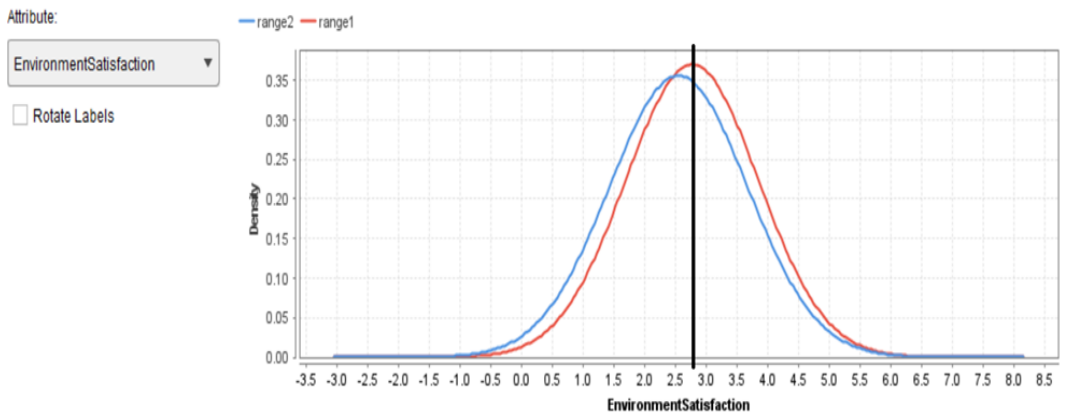
Maximum   **(4)**

Average   **(2.722)**

Fig.12. Environmental Satisfaction histogram

| Attribute / Algorithm	Weights				
	Education	Job Role	Years in current role	Monthly Rate	Gender
Naive Bayes	0.013	0.012	0.015	0.015	0.008
Generalized Linear Model	0.017	0.012	0.014	0.010	0.011
Logistic Regression	0.029	0.068	0.044	0.032	0.031
fast Large Margin	0.023	0.064	0.029	0.450	0.004
Deep Learning	0.012	0.042	0.051	0.045	0.025
Decision Tree	0.016	0.012	0.012	0.009	0.010
Random Forest	0.046	0.031	0.045	0.059	0.038
Gradient Bested trees	0.016	0.012	0.012	0.009	0.010
Support vector machine	0.001	0.00	0.00	0.00	0.001

Table 6. Significant features with weights

Education

The higher the employee's academic degree, the greater the employee's academic experience and education shown in (Figure 13).

Minimum   (1)

Maximum   (5)

Average   (2.913)

Fig.13. Education Description
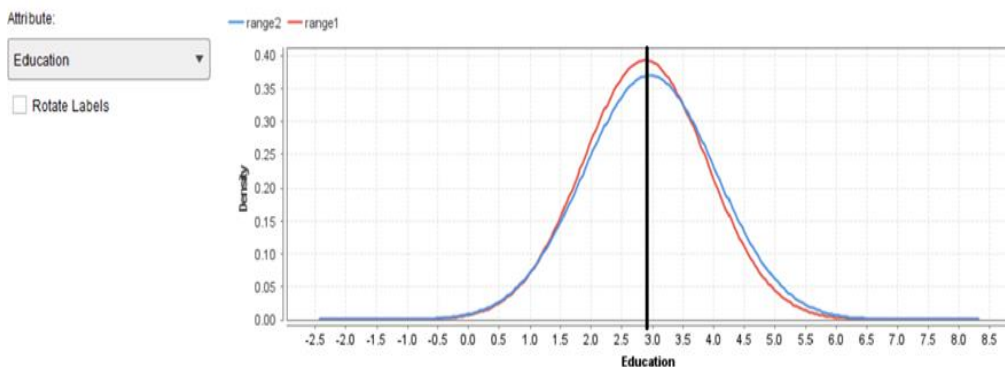
• Job Role

Job roles contain a range of roles such as Sales Director, and Sales Representative, Healthcare Representative, Human Resources (Fig.14) illustrating a description of the job roles.

Sales Executive   (22.18%)

Research Scientist   (19.86%)

Laboratory Technician   (17.62%)

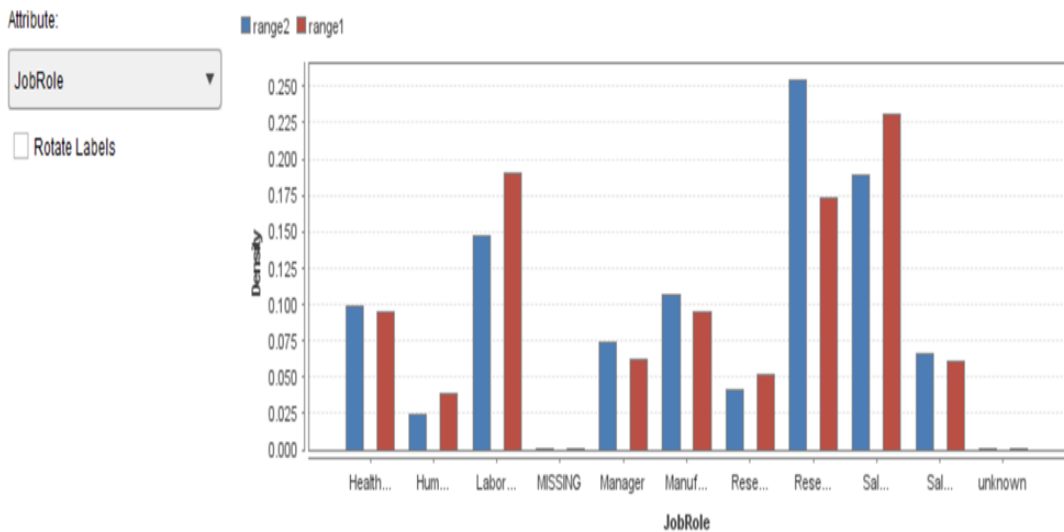|	
Manufacturing Director	**(9.86%)**
Healthcare Representative	**(8.91%)**
Manager	**(6.94%)**
Sales Representative	**(5.65%)**
Research Director	**(5.44%)**
Human Resources	**(3.54%)**

Fig.14. Education Description

• Years in current role

Years of current work are among the most important factors that affect an employee's performance. The more years an employee works in the field, the greater his competence and practical experience. Describe the years in the current role as shown in (Figure 15).

Minimum **(0)**

Maximum **(18)**

Average **(4.229)**

Fig.15. Years in current role
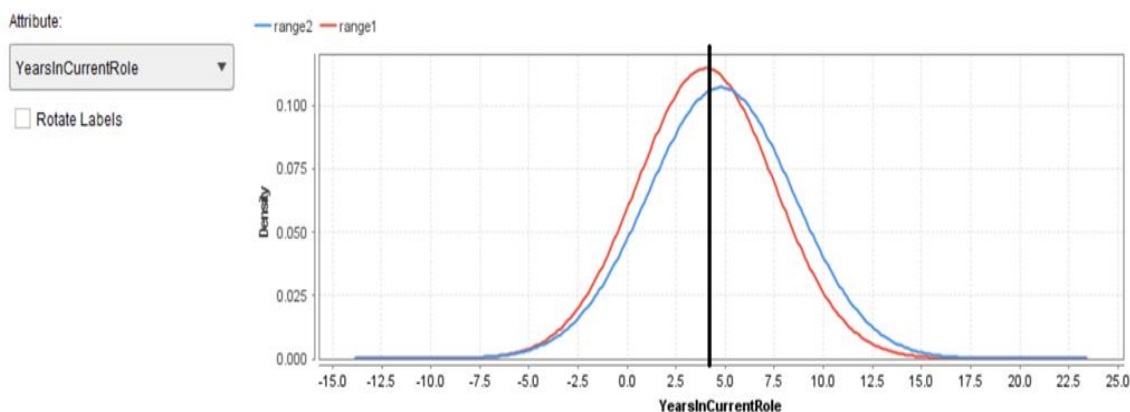

Description

• Monthly Rate

One of the crucial determinants influencing employee performance pertains to the average monthly income of the employee. A positive correlation is observed between the employee's monthly income and their performance level. The depiction of the monthly rate is elaborated in (Figure 16).

Minimum (2094)

Maximum (26999)

Average (14313.103)

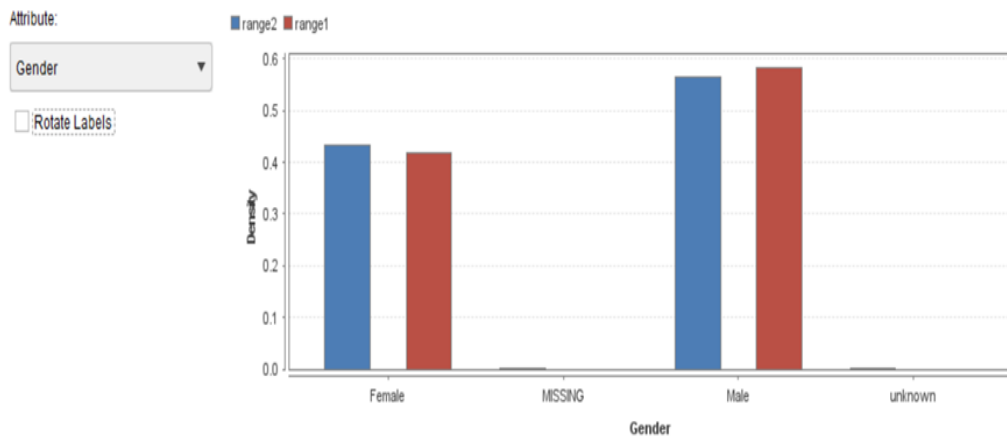**Fig.16. Monthly Rate Description**
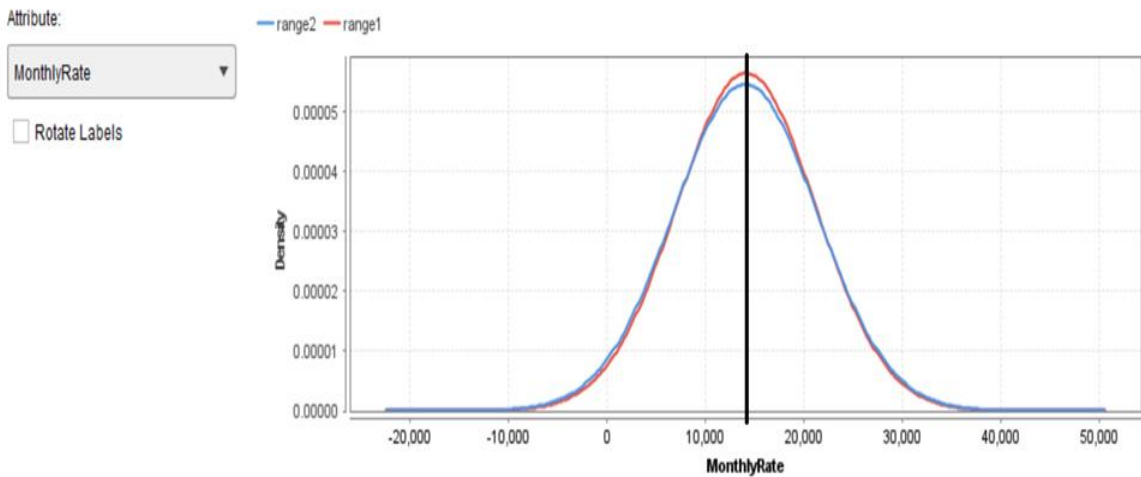
• **Gender**

Male   (882)   60%

Female   (588)   40%

One of the key elements influencing work performance is the employee's gender. It was found that the percentage of males was 60%. The percentage of females is 40%. Describe the employee's gender as shown in (Figure 17).

Fig.17. Monthly Rate Description

5. CONCLUSION

This study highlights the benefits of integrating data mining techniques into human resource management (HRM) practices to accurately identify and nurture skilled employees. Traditional methods of performance evaluation often rely on subjective intuition, which can be inconsistent. By employing data mining techniques and using a comprehensive dataset from IBM, a robust classification model that significantly enhances the accuracy of predicting employee performance is developed.

The model, which uses algorithms such as Decision Tree (DT), Naive Bayes, and Support Vector Machine (SVM), achieved high accuracy rates between 85.7% and 100%. It identifies critical factors like monthly income, salary adjustments, and tenure that influence employee performance. This data-driven approach enables HR professionals to make informed, fair evaluations and optimize employee placement. It also helps identify high-potential individuals for promotions and targeted development programs, reducing hiring risks and improving workforce productivity and engagement.

Incorporating data mining into HRM practices can enhance organizational effectiveness and competitiveness. Future research should expand the dataset and explore additional classification techniques to further improve the model's accuracy and applicability.

References

[1]　　　Strohmeier, S., & Piazza, F.. **Artificial Intelligence Techniques in Human Resource Management - A Conceptual Exploration. Intelligent Techniques in Engineering Management.** (2015).

[2]　Campbell, J. P., & Wiernik, B. M. **The modeling and assessment of work performance. Annual Review of Organizational Psychology and Organizational Behavior**, 2, 47–74. http://doi.org/10.1146/annurev-orgpsych-032414-111427. (2015).

[3]　SN Mishra, DR Lama. **A decision-making model for human resource management in organizations using data mining and predictive analytics.** International Journal of Computer Science and Information Security (IJCSIS) (2016).

[4]　Qasem et al., **"Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance",** in (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 2, (2012).

[5]　Jayadi, Puguh. **Use Of Inter-Employees Performance Evaluation Application on Employee Performance Improvement** (2018).

[6]　John M. Kirimi, Christopher A. Moturi**. Application of Data Mining Classification in Employee Performance Prediction.** International Journal of Computer Applications. 146, 7 (Jul 2016), 28-35. DOI=10.5120/ijca2016910883 (2016).

[7]　Merry Grace T Li, Macrina Lazo, Ariel Kelly Balan, and Joel de Goma. **"Employee Performance Prediction using Different Supervised Classifiers"** ProcIntConfIndEngOperManageIEOM Society, : pp. 6870–6876 (2021).

[8]　S. E. Viswapriya, (2019), **Survey on Predicting Performance of An Employee using Data Mining** Techniques, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 08, Issue 10 (October 2019).

[9]   MOHAMMADNOUR ALJARRAH, **"The Impact of Enterprise Resource Planning System of Human Resources on the Employees' Performance Appraisal in Jordan"**. WSEAS TRANSACTIONS on ENVIRONMENT and DEVELOPMENT DOI: 10.37394/232015.2021.17.35 (2021).

[10]   Qasem A. Al-Radaideh ,Eman Al Nagi**, "Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance** (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 2, 2012.

[11]   HamidahJantan, Norazmah Mat Yusoff and MohamadRozuan Noh , **"Towards Applying Support Vector Machine Algorithm in Employee Achievement Classification**", Proceedings of the International Conference on Data Mining, Internet Computing, and Big Data, Kuala Lumpur, Malaysia, 2014 ISBN: 978-1-941968-02-4 ©2014 SDIWC (2014).

[12]   LipsaSadath (2013), **"Data Mining: A Tool for Knowledge Management in Human Resource",** International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol-2, April 2013.

[13]   Mona Nasr, Essam Shaaban, and Ahmed Samir (2019**). A proposed Model for Predicting Employees' Performance Using Data Mining Techniques: Egyptian Case Study**. International Journal of Computer Science and Information Security (IJCSIS).

[14]   Amarendra Deo, Yara Mahfouz. (2023). **IMPLEMENTATION OF MACHINE LEARNING ALGORITHM FOR THE PREDICTION OF IT INDUSTRY EMPLOYEES PERFORMANCE.** International Research Journal of Modernization in Engineering Technology and Science. Volume:05/Issue:02/February-2023

[15]   ZarminaJaffara, WaheedNoorb ,and Zartash Kanwa. **Predictive Human Resource Analytics Using Data Mining Classification Techniques International** Journal of Computer (IJC) Volume 32, No 1, pp 9-20 (2019).