Employee Attrition Prediction Using Machine Learning Techniques

Sushma Mali
IINTM, Janakpuri, New Delhi, India
sushmamalikiitm@gmail.com

Anamika Rana
MSI, Janakpuri, New Delhi, India
anamica.rana@gmail.com
orcid.org/0000-0002-6201-7831

Madhu Chauhan
IINTM, Janakpuri, New Delhi, India.
myself_madhu26@yahoo.com

DOI: http://dx.doi.org/10.31642/JoKMC/2018/110211
Received Apr. 2, 2024. Accepted for publication May. 19, 2024

Abstract: *Effective employees are regarded as the most precious assets and the foundation of any business. Companies invest a lot of money in staff training programs because they believe they will pay off in the long run. Therefore, it is crucial to keep a long-term, promising staff; as has been seen throughout the years, this is one of HR's most challenging duties. This study aims to determine the key variables influencing employee attrition and develop a machine learning model to predict employee attrition based on the variables given. This will make it easier for management to spot workers who are likely to quit, allowing them to take preventative measures and make wise decisions about appraisal and recognition. The four most popular classification algorithms utilized in this study were KNN, Naive Bayes, Random Forest and Logistic Regression. It was shown that Naive Bayes classifier outperformed the others by 89% in terms of accuracy and produced more accurate predictions.*

Keywords: Naive Bayes, SVM, KNN, Logistic Regression and Machine Learning, Employee Attrition

I. INTRODUCTION

Employee attrition refers to the phenomenon of employees leaving an organization over a specific period. It's often measured as the percentage of employees who leave the company voluntarily or involuntarily. Attrition can have a significant impact on an organization, leading to increased costs associated with hiring and training new employees, a potential loss of institutional knowledge, and disruptions in workflow[1]. There are two main types of employee attrition[2],[3]:

A. Voluntary Attrition

Resignation: Employees leave the organization voluntarily for various reasons, including personal reasons, career advancement opportunities elsewhere, dissatisfaction with the work environment, or better job offers.

Retirement: Employees may leave due to retirement.

B. Involuntary Attrition

Termination: Employees are let go by the organization for reasons such as poor performance, violation of company policies, or restructuring.

Layoffs: Employees are laid off due to organizational changes, financial constraints, or a decrease in the demand for certain roles.

Reducing employee attrition is a common goal for organizations because high turnover rates can have negative consequences. Managing an organization's employee retention rate is currently regarded as one of the difficult and important responsibilities carried out by the HR departments. Due to its negative effects on an organization's success and reputation, it has grown to be a significant problem.

In cases of high retention rates, the following are the primary difficulties that businesses frequently encounter[4],[5]:

Replacement Costs: It can be costly to find, hire, and train new personnel. Frequent hiring due to high turnover rates raises the expenses of recruitment and onboarding.

Loss of Knowledge and Expertise: Experienced employees possess valuable knowledge and expertise. High turnover can result in a loss of institutional knowledge, making it challenging for the organization to maintain a consistent level of performance.

Impact on Team Dynamics: Frequent turnover can disrupt team dynamics. Constant changes in team composition may lead to decreased cohesion, communication challenges, and a negative impact on overall team performance.

Decreased Productivity: Constantly having to bring new employees up to speed can lead to decreased productivity. New hires often take time to reach the same level of efficiency as experienced employees.

Negative Impact on Morale: High turnover can negatively affect the morale of remaining employees. It may create an atmosphere of uncertainty and insecurity, leading to decreased job satisfaction.

Impact on Customer Relationships: In customer-facing roles, frequent turnover can negatively impact relationships with clients or customers. Consistent personnel changes may lead to a lack of continuity in service delivery.

Time and Resource Allocation: HR departments and management spend significant time and resources on recruitment and onboarding when turnover rates are high. This can divert attention from other strategic initiatives.

Difficulty in Succession Planning: High turnover rates can make succession planning challenging. Identifying and preparing internal candidates for key roles becomes difficult when employees are leaving frequently.

Impact on Organizational Culture: A high turnover rate can contribute to a negative organizational culture. It may lead to a perception that the company is not a stable or desirable place to work.

Competitive Disadvantage: Organizations with high turnover rates may struggle to attract top talent. A reputation for frequent turnover can make it difficult to compete for skilled professionals in the job market.

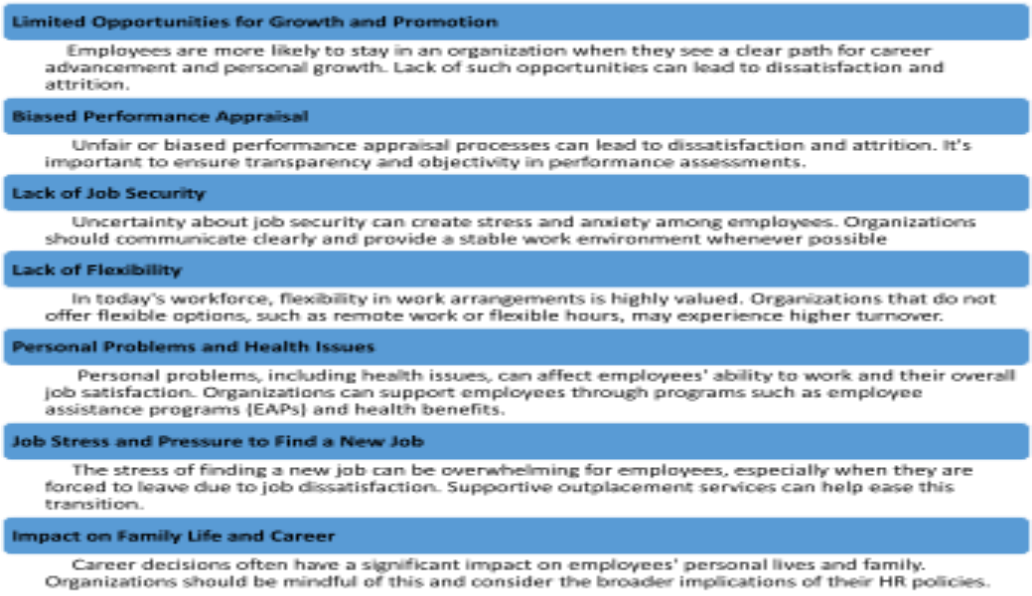


Fig 1: Numerous Factors Affecting the Organizations' *Lower Retention Rates [6],[7]*

Addressing these factors involves a combination of HR policies, organizational culture, and management practices. Strategies for improving retention rates may include:

- Offering clear career development opportunities.
- Conducting fair and transparent performance appraisals.
- Providing job security and stability.
- Implementing flexible work arrangements.
- Offering support programs for employees facing personal challenges.
- Establishing a welcoming and upbeat workplace.
- Encouragement of work-life harmony.
- Offering chances for growth and training.
- Promoting candid criticism and honest dialogue.

By addressing these factors and adopting retention-focused strategies, organizations can work towards creating a workplace where employees are

more likely to stay and thrive. The value of a business is maintained by increasing its staff and improving decision-making by HR departments. They therefore frequently rely on data-driven judgments that produce better HR analytics, which in turn represent the worth and development of the firm. Data analytics, especially through the use of machine learning, is a powerful tool for HR departments to enhance decision-making, improve employee retention, and contribute to the overall success of the organization. It's a transformative approach that aligns HR practices with the data-driven culture embraced by many successful businesses[8].

Machine learning is a technology that allows algorithms to accurately predict results by using previously learnt data as input. By using statistical techniques and algorithms, machine learning models can be developed to produce predictions or classifications and to uncover important insights in the field of data analytics. Ideally, these insights lead to actions that impact important growth metrics in applications and companies. Traditional models like logistic regression, which frequently yield inconsistent results, were frequently utilized as ML models in HR analytics. Due to the improved accuracy and effectiveness of machine learning (ML) to address issues, we use supervised machine learning models in this work to predict employee attrition[9].

The key contributions of this study include:

- An examination of several factors impacting employee attrition.
- Machine learning techniques used for categorization problems are used to forecast if an employee is likely to quit given the supplied factors.
- The models' accuracy is evaluated using accuracy estimation methods, and employers are advised to pick the best classifier.

II. Literature Review

There are numerous studies that concentrate on different factors that lead to high employee attrition rates. Using data analytics and machine learning models, other research attempted to predict employee attrition based on a variety of employee characteristics. This section summarizes a few of those studies to offer an overview of the various methods that researchers have employed and the results.

Table 1: Literature Review

Sr. No	Reference	Review
1	[10]	Building a classification model using decision trees is a common approach in data mining and machine learning. Decision trees are interpretable and can be easily converted into a set of rules, making them valuable for generating classification rules.
2	[11]	Using the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology and data mining techniques to predict and understand employee performance is a systematic and structured approach.
3	[12]	Using XGBoost, an extreme gradient boosting technique, for predicting employee turnover is a sound choice. XGBoost is known for its high performance, scalability, and regularization capabilities, making it particularly effective in scenarios where accuracy is crucial. The regularization formulation in XGBoost helps prevent overfitting and enhances model generalization.
4	[13]	Both tangible and intangible costs associated with workforce turnover. The attributes of workers such as job position, overtime, and work level, can indeed have a significant impact on attrition.
5	[14],[15]	Proposed work sounds promising and focused on providing valuable insights to the human resources department for better employee retention strategies. The factors most likely to affect an employee's choice are extra hours, job level, and monthly pay.
6	[16]	With the help of the IBM HR Employee Attrition & Performance dataset, the discrepancy in the obtained data was discovered. The correlation between the continuous variables in the model was displayed using the correlation plot and histogram visualization throughout the data

		exploration phase. The Synthetic Minority Oversampling Technique, or SMOTE, was employed to maintain equilibrium in the Attrition class.
7	[17]	The rationale of using data mining techniques, specifically Weka, to understand the factors influencing attrition of human resources is both practical and beneficial.
8	[18]	The recommendation of the KNN classifier for accurately predicting employee attrition, backed by evidence of its superiority in the study, has practical implications for HR decision-making. It's essential to ensure that the implementation aligns with the organization's needs and ethical standards.

III. METHODOLOGY

To create a machine learning (ML) model for a given problem statement, there are three main processes involved: data preprocessing, model training, and accuracy estimate. In this section, the approach utilized to complete this work and the data set used are briefly discussed. The approach taken for this paper is shown in Figure 2.

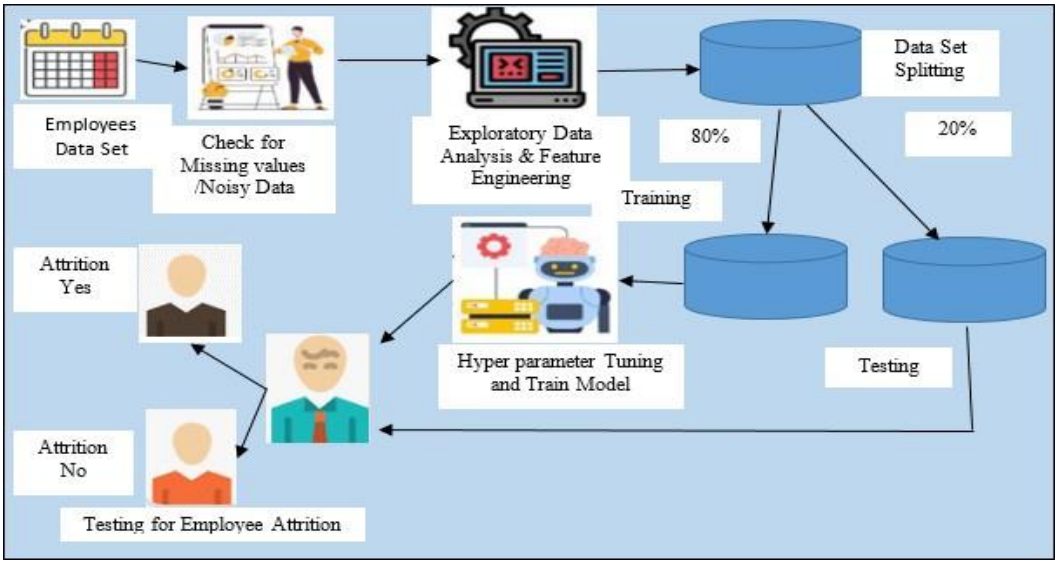


Fig 2: Our proposed research study's methodology analysis for predicting employee attrition

The data set taken from the Kaggle with 1470 employee records in the Employee Attrition dataset, each having 13 columns and one dependent variable. The true dependent variable (or target) for which models must be constructed is the final characteristic, "Attrition." String columns are also present in the dataset as shown in Table 2. We utilized an encoding technique to convert categorical properties into numerical ones because most machine learning algorithms can only create models from numerical data. In order to achieve this, we used label encoding rather than one-hot encoding because label encoding does not add or introduce additional columns to the dataset.

Table 2: Details of Dataset used

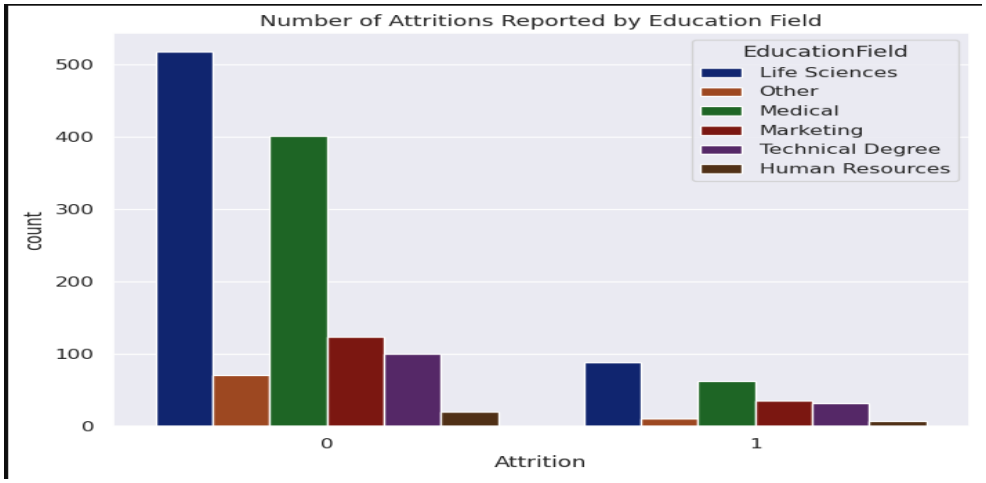
Feature	Data type	Range of values	Remarks
Age	Integer	18-60	Age of employee
Department	String		Department of work
Distance from home	Integer	1-30	In miles

Education	Integer	1-5	1-Below College; 2-College; 3-Bachelor; 4-Master; 5-Doctor
Education Field	String	Technology/Science/Medical /etc	
Environment	Integer	1- 4	1-Low; 2-Medium; 3-High; 4-Very High;
Job satisfaction	Integer	1- 4	1-Low; 2-Medium; 3-High; 4-Very High
Marital Status	String	Single/Married/Divorced	
Monthly Income	Integer	1000-20000	
No of companies worked	Integer	0-10	Number of companies worked prior to IBM
Work Life Balance	Integer	1- 4	1-Bad; 2-Good; 3-Better; 4-Best
Years At Company	Integer	0-40	Current years of service in IBM
Attrition	String	Yes / No	Target / Dependent Variable

Managing missing values constitutes a critical phase in the data preprocessing workflow, and furnishing explicit explanations regarding their treatment enhances the transparency and replicability of the analysis. The full dataset was then checked for any missing values and irregularities, and those were properly handled. Imputation method is used to replacing the missing values. This can be achieved through statistical method such as mode imputation.

The dataset was divided into 80:20, or 80% for training the model and the remaining 20% for testing, as part of the data preprocessing process. Before creating models, feature analysis was finalized to determine the relationships between the dataset's various attributes. In Figures 3, a few analysis results are displayed.

Python was selected as the programming language for implementing the employee attrition project due to its versatility, rich ecosystem of libraries for data analysis and machine learning, and ease of use. Python provides powerful tools such as pandas for data manipulation, scikit-learn for machine learning algorithms, and matplotlib/seaborn for data visualization, making it well-suited for the various tasks involved in analyzing and modeling employee attrition



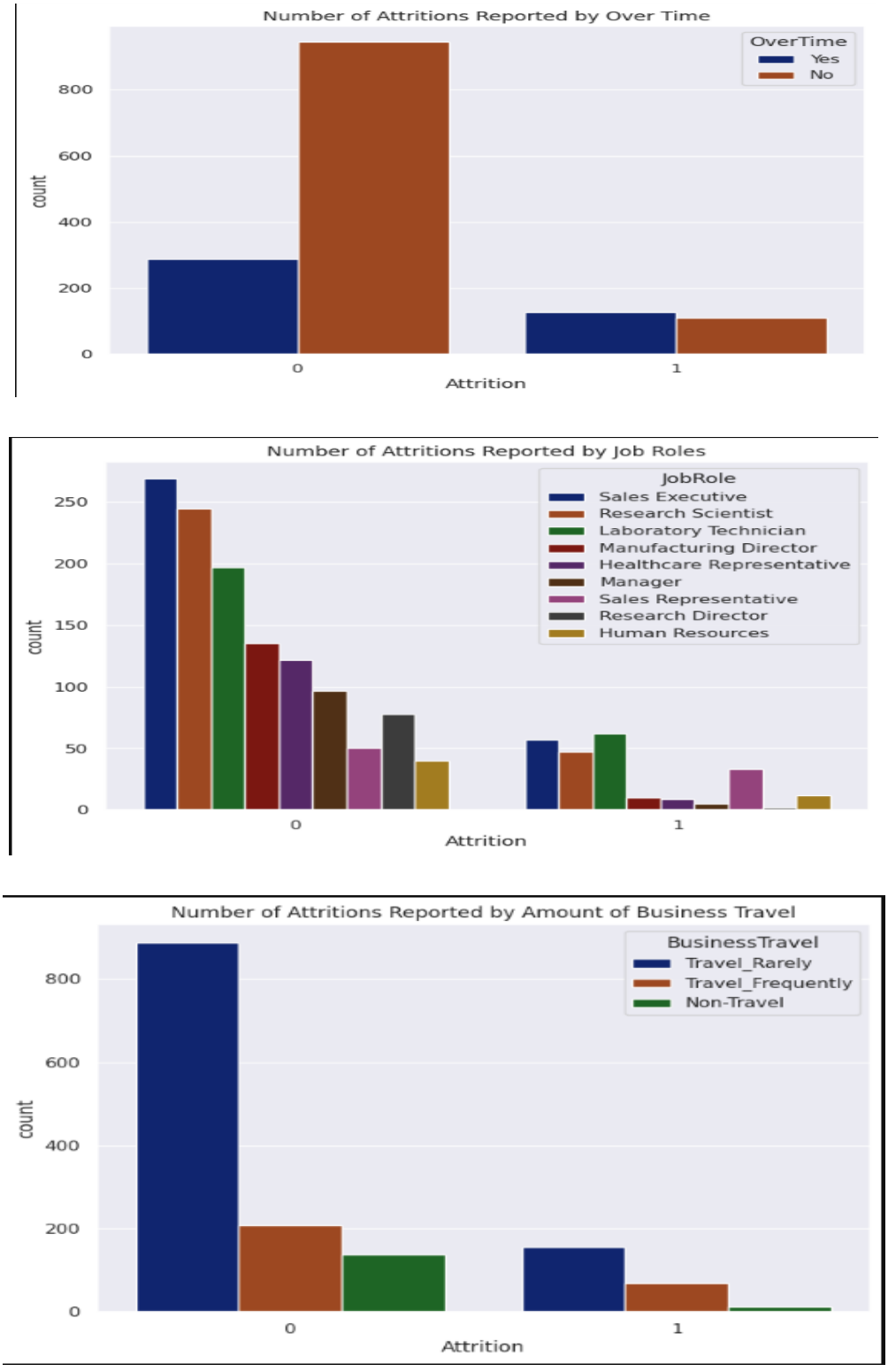


Fig 3: Number of Attritions reported by Education field, Over Time, Job Roles and Business Travel

IV. Machine Learning Approaches

K-Nearest Neighbor (KNN), Naive Bayes (NB), Random Forest (RF), and Linear Regression are four widely recognized supervised machine learning classifiers chosen for their efficacy in classification and regression tasks. KNN operates by identifying the 'k' closest data points to a given instance in the feature space and predicting its class based on the majority class among its neighbors. In binary classification problems like Employee Attrition, where predictions entail binary outcomes (e.g., whether an employee will leave or stay), KNN can excel. It thrives in scenarios where distinct clusters or groups exist in the data, as instances of one class often cluster together in the feature space.
Additionally, Naive Bayes is selected for its simplicity and effectiveness in handling high-dimensional data, making it suitable for binary classification tasks. Meanwhile, Random Forest is favored for its robustness to overfitting, capacity to handle high-dimensional data, and ability to capture complex relationships within the data. In contrast, other algorithms like Decision Trees, Support Vector Machines (SVM), and Neural Networks were not selected because Decision Trees are interpretable and can capture complex interactions in the data, they are prone to overfitting, especially with high-dimensional data, SVM perform well on high dimensional data and neural network require large amounts of data. The operation of these classifiers is briefly covered in this section.

A. Naïve Bayes: The Naïve Bayes algorithm, which relies on the Bayes theorem, is a supervised learning technique for classification problems. The Naive Bayes method is built on the concept of probability. This classifier bases its operation on the idea of probability and learns the probabilities associated with each data point (or object), together with the attributes and class to which it belongs. It is also referred to as a probabilistic classifier as a result. Out of all the potential outcomes, probability aids in the prediction of an event's occurrence.

The mathematical formulation of Bayes' theorem is given by the equation (1):

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} \quad (1)$$

Where P (B) is not equal to zero and A and B are events.

In essence, we are looking for the probability of event A given the truth of event B. Evidence also refers to Event B. P(A) is the prior probability of A (the probability of an event occurring before supporting evidence is observed). An attribute value of an unidentified instance (in this case, event B) serves as the proof. P (A|B) is the a posteriori probability of B, or the likelihood of an event following the discovery of evidence[19],[20].

Table 3: Comparative study of ML approaches

Sr. No.	Approach	Advantage	Disadvantage
1	Naive Bayes	● Simple and Easy to implement. ● Requires a small amount of training data to estimate the necessary parameters	● The Naive assumption may not hold in all cases. ● It doesn't perform well on highly correlated features.
2	KNN	● Simple to implement ● Robust to the noisy training data ● More effective if the training data is large	● Always needs to determine the value of K which may be complex sometime ● The computation cost is high because of calculating the distance between the data points for all the training samples.
3	Random Forest	● Capable of performing both classification and regression tasks. ● can handle large dataset with high dimensionality ● enhance the accuracy of the model and prevents the overfitting issue.	● Not suitable for a regression task.
4	Logistic Regression	● Easier to implement, interpret and very efficient to train.	● May lead to overfitting when no. of observation is less than the no. of

		• fast at classifying unknown records • Good accuracy for many datasets and it performs well when the data set is linearly separable.	features. • Not capable of solving non-linear data and in real-world linearly separable data is hard to find.

B. K-Nearest Neighbors (KNN): A straightforward yet effective supervised machine learning approach for classification and regression problems is K-Nearest Neighbors (KNN). Being a non-parametric and lazy learning method, it delays the generalization stage until it receives fresh data points to categorize and makes no assumptions about the underlying data distribution. It fits nicely with supervised learning, where a model is trained on a collection of labeled data with the expectation that it would correctly predict the label of test data. Pattern recognition, categorization issues (such spotting fraud or spam emails), and intrusion detection are where KNN is most commonly employed. KNN is non-parametric since it makes no underlying assumptions about the distribution of the data under investigation. It bases its operation on the closeness of the data points; hence, when a new data point needs to be grouped, it determines its class by calculating the distance to the closest data points, then adds the new data point to the class with the shortest distance as shown in Figure 4. The number of neighbors for which distance must be determined is indicated by the value of "k" [19],[20]. Steps followed by KNN as shown in Figure 5.

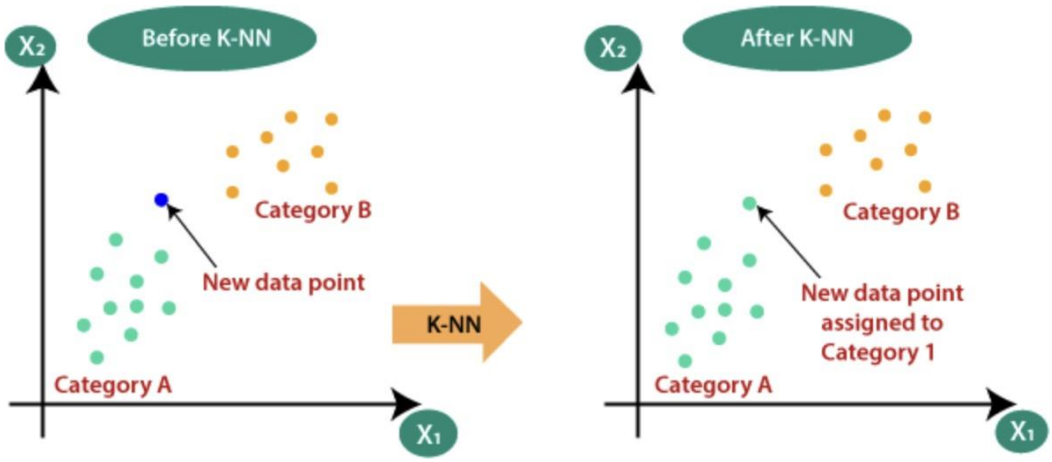


Fig 4: K-Nearest Neighbors (KNN) Classifier [22]

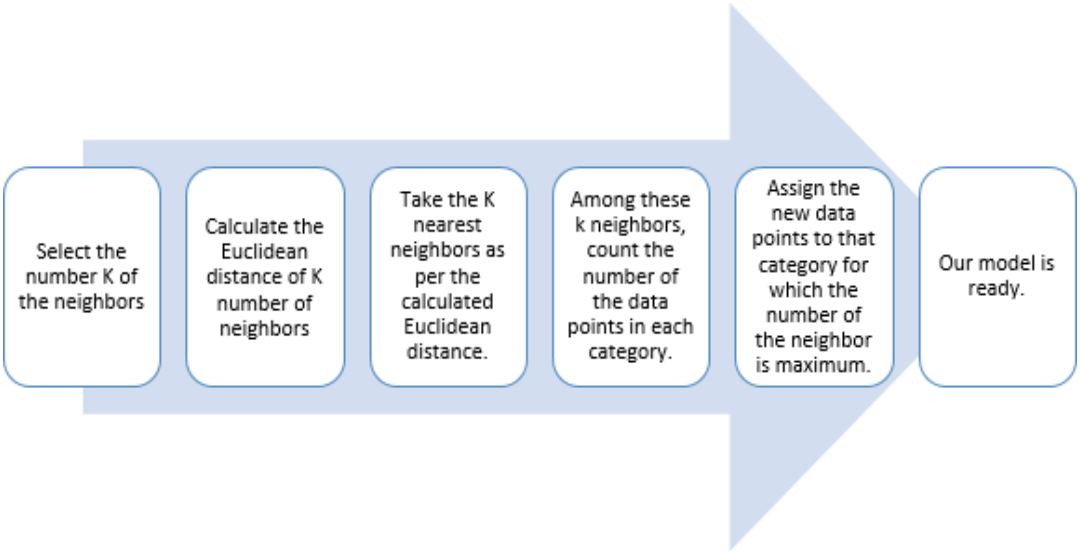


Fig 5: Steps of K-NN Classifier

C. Random Forest: Random Forest is a well-known machine learning component of the supervised learning approach. Figure 6 displays the working of Random Forest where regression and classification-based machine learning challenges can be solved with it. Its foundation is the idea of ensemble learning, a method for merging multiple classifiers to tackle difficult issues and enhance model performance. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." The random forest makes predictions based on the votes of the majority of projections rather than relying only on one decision tree. It does this by using forecasts from each decision tree. With more trees in the forest, overfitting and higher accuracy are avoided [21].

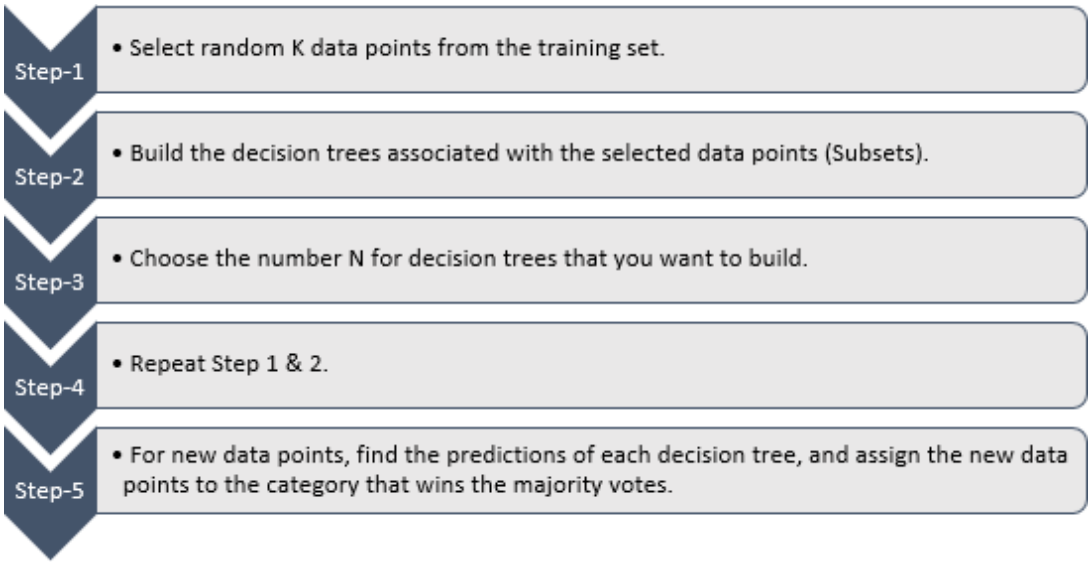


Fig 6: Working of Random Forest Algorithm

D. Logistic Regression (LR)

In binary classification situations, where the goal is to predict a binary outcome based on one or more independent variables (features), logistic regression (LR) is a statistical technique. To translate the linear combination of input features into a probability value between 0 and 1, the logistic function, also known as the sigmoid function, is employed in the logistic regression classification model. Equation (2) gives the following description of the logistic function:

$$Sigmoid\,(z) = \frac{1}{1+e^{-z}} \qquad (2)$$

Where z is the input features' linear combination with matching weights:

$$Z = w_0 + w_1x_1 + w_2x_2 + \ldots + w_mx_m \qquad (3)$$

In the above equation (3), w_0, w1, w2,…,w_m are the model's coefficients (weights), and $x_1,x_2,\ldots,x_m$ are the input features.

The logistic function maps the probability P(y=1) that the binary outcome is 1 (positive class) to the linear combination z. In cases when P(y=1) is greater than or equal to 0.5, the logistic regression model predicts the positive class (1); in other cases, it predicts the negative class (0).

Finding the best-fit correlations between the target variables (dependent variables) and predictor factors (independent variables) in a binary classification problem is the main objective of logistic regression. It makes an effort to determine the likelihood that a specific instance will belong to either the positive or negative class [23].

V. Results and Methodology

Once the models were trained on the training data set, they were assessed on a separate test bed that was not part of the training data set. In addition to other evaluation criteria including recall, precision, and f1-score, the models' accuracy was evaluated. The most favorable outcomes from the multiple-run trials were chosen for analysis and comparison.

Table 4: Classifier comparisons for predicting employee attrition

	KNN			**Naïve Bayes**			**Random Forest**			**Logistic Regression (LR)**		
	Attrition : No	**Attrition : Yes**	**Avg**	**Attrition : No**	**Attrition : Yes**	**Avg**	**Attrition : No**	**Attrition : Yes**	**Avg**	**Attrition : No**	**Attrition : Yes**	**Avg**
Precision	0.84	0.15	**0.49**	0.9	0.42	**0.66**	0.83	0.49	**0.66**	0.85	0.23	**0.54**
Recall	0.91	0.06	**0.48**	0.94	0.21	**0.57**	0.97	0.07	**0.52**	0.86	0.12	**0.49**
F1- Score	0.88	0.07	**0.47**	0.96	0.28	**0.62**	0.86	0.15	**0.505**	0.81	0.11	**0.46**

Table 4 displays the findings from studies and a comparison of metrics used for accuracy estimates. The Naive Bayes classifier generated the best results because it outperformed the other three classifiers in terms of precision and F1-score. Figure 5's accuracy results are displayed, with NB achieving 89% accuracy in prediction, Random Forest achieving 83% while KNN & Logistic Regression earned the lowest accuracy with 82% and 79% accurate predictions displayed in Figure 7.

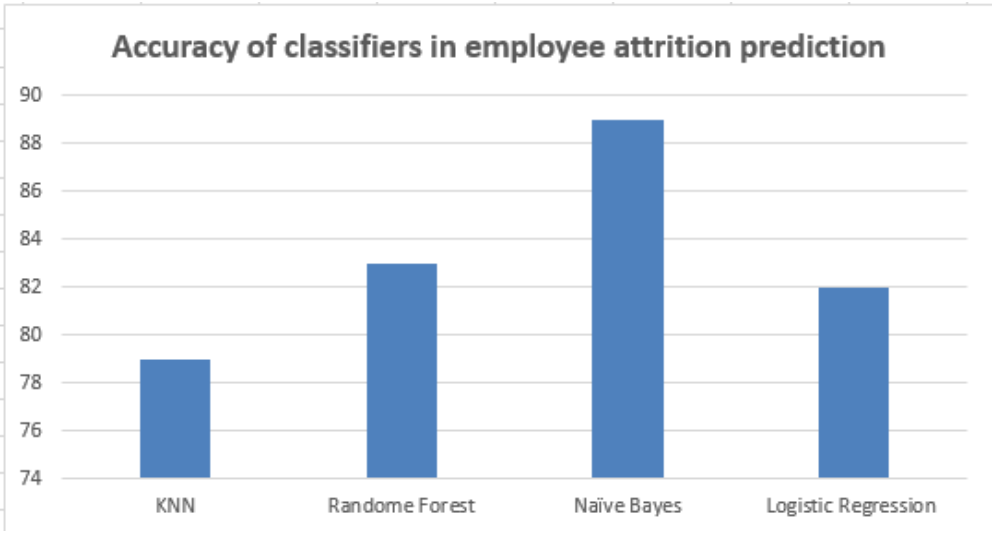


Fig 7: Accuracy Achieved by Different Classifiers

Observations

Based on the data analysis, we observed the following point in the dataset:

- Compared to medical, life sciences, and other sectors, personnel with technical degrees, human resources, and marketing backgrounds had a higher likelihood to leave their jobs.
- Compared to employees who don't work extra, 30% of overtime employees suffer attrition. Regular business travel is another factor in attrition.
- Sales representatives make up roughly 40% of attrition employees, followed by lab technicians and roles in human resources.
- Attrition is also significantly influenced by the level of stock options. For example, employees are less likely to attrite if the employer gives them more stock options.
- The data demonstrates that an employee is less likely to leave the organization the more invested they are in their work.

Implications

Based on our observations, we can draw the following implications for companies:

- **Tailoring Support Based on Job Roles:** Recognizing the diverse needs of employees with different backgrounds is a strategic approach. Technical, human resources, and marketing roles may require unique forms of support, training, and professional development.
- **Reducing Overtime for Job Satisfaction:** Limiting overtime can indeed contribute to increased job satisfaction by promoting a healthy work-life balance. This can lead to improved employee morale, productivity, and retention.
- **Incentives for Specific Positions:** Providing additional incentives, such as reduced work hours and flexibility, for sales representatives, laboratory technicians, and human resources positions can help address the specific challenges and demands associated with these roles. This may attract and retain talent in these critical areas.
- **Offering More Stock Options:** Introducing stock options is a common strategy to align employees' interests with the company's success. This can be a powerful motivator, especially in positions where employees' contributions significantly impact the company's performance.
- **Increasing Employee Involvement:** Fostering a sense of involvement and contribution can enhance job satisfaction and overall engagement. This can be achieved through clear communication, involving employees in decision-making processes, and recognizing their achievements.

Implementing these recommendations requires a thoughtful and well-planned approach. Here are some considerations:

- **Employee Surveys and Feedback:** Regularly gather feedback from employees to understand their needs and preferences. This can help tailor strategies to address specific concerns.
- **Flexibility Programs:** Develop flexible work arrangements that suit the nature of each job. This could include remote work options, flexible hours, or compressed workweeks.
- **Communication and Training:** Ensure clear communication about the benefits and incentives offered. Additionally, provide ongoing training and development opportunities to enhance employees' skills and job satisfaction.
- **Performance Recognition:** Implement effective performance recognition programs to acknowledge and reward employees for their contributions. This can include both monetary and non-monetary rewards.
- **Policy Development:** Develop and communicate policies that support the well-being of employees, such as guidelines on overtime, work hours, and stock option programs.
- **Monitoring and Adjusting:** Regularly assess the impact of implemented strategies and be willing to adjust based on feedback and changing circumstances.

By taking a comprehensive and employee-centric approach, companies can create a positive work environment that attracts, retains, and motivates talent across various job roles.

VI. Conclusion

For HR departments, keeping experienced people in a company has become a major challenge, and they are looking for better strategies to deal with this issue. Organizations can lose a lot of money through employee attrition since it costs a lot to replace their knowledge and productivity. We used a variety of machine learning classifiers in this study to forecast employee attrition based on predetermined parameters. Employers may then take preventative actions to ensure that this didn't happen by using this information to predict target employees who are likely to depart. We created models utilizing KNN, RF, NB and Logistic Regression classifiers and discovered that Naive Bayes performed the best in terms of accuracy, outperforming the others. On this dataset, the proposed model was essentially accurate and sufficiently met the desired outcome.

References

[1] S. S. Alduayj and K. Rajpoot, "Predicting employee attrition using machine learning," in *2018 international conference on innovations in information technology (iit)*, 2018, pp. *93–98*.

[2] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. William De Luca, "Predicting employee attrition using machine learning techniques," *Computers*, vol. 9, no. 4, p. 86, 2020.

[3] A. Raza, K. Munir, M. Almutairi, F. Younas, and M. M. S. Fareed, "Predicting employee attrition using machine learning approaches," *Appl. Sci.*, vol. 12, no. 13, p. 6424, 2022.

[4] P. Ajit, "Prediction of employee turnover in organizations using machine learning algorithms," *Algorithms*, vol. 4, no. 5, p. C5, 2016.

[5] M. Subhashini and R. Gopinath, "Employee attrition prediction in industry using machine learning techniques," *Int. J. Adv. Res. Eng. Technol.*, vol. 11, no. 12, pp. 3329–3341, 2020.

[6] A. Qutub, A. Al-Mehmadi, M. Al-Hssan, R. Aljohani, and H. S. Alghamdi, "Prediction of employee attrition using machine learning and ensemble methods," *Int. J. Mach. Learn. Comput*, vol. 11, no. 2, pp. 110–114, 2021.

[7] F. K. Alsheref, I. E. Fattoh, and W. M Ead, "Automated Prediction of Employee Attrition Using Ensemble Model Based on Machine Learning *Algorithms*," *Comput. Intell. Neurosci.*, vol. 2022, 2022.

[8] S. Najafi-Zangeneh, N. Shams-Gharneh, A. Arjomandi-Nezhad, and S. Hashemkhani Zolfani, "An improved machine learning-based employees attrition prediction framework with emphasis on feature selection," *Mathematics*, vol. 9, no. 11, p. 1226, 2021.

[9] P. K. Jain, M. Jain, and R. Pamula, "Explaining and predicting employees' attrition: a machine learning approach," *SN Appl. Sci.*, vol. 2, pp. 1–11, 2020.

[10] V. V. Saradhi and G. K. Palshikar, "Employee churn prediction," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 1999–2006, *2011*.

[11] C. Singh and A. Gopal, "Per for mance Analysis of Faculty using Data Mining Techniques," *Int. J. Comput. Sci. Appl. Issue*, pp. 140–144, *2010*.

[12] R. Punnoose and P. Ajit, "Prediction of Employee Turnover in Organizations using Machine Learning Algorithms," *Int. J. Adv. Res. Artif. Intell.*, vol. 5, no. 9, pp. 22–26, 2016, doi: 10.14569/*ijarai*.2016.050904.

[13] R. Chakraborty, K. Mridha, R. N. Shaw, and A. Ghosh, "Study and prediction analysis of the employee turnover using machine learning approaches," in *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), 2021*, pp. 1–6.

[14] S. Al-Darraji, D. G. Honi, F. Fallucchi, A. I. Abdulsada, R. Giuliano, and H. A. Abdulmalik, "Employee attrition prediction using deep neural networks," *Computers*, vol. 10, no. 11, pp. 1–11, 2021, doi: 10.3390/*computers10110141*.

[15] S. M. Arqawi *et al.*, "Predicting Employee Attrition and Performance Using Deep Learning," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 21, pp. 6526–6536, 2022.

[16] K. K. Mohbey, "Employee's attrition prediction using machine learning approaches," in *Machine Learning and Deep Learning in Real-Time Applications*, IGI Global, 2020, pp. 121–128.

[17] I. F. Alsuahim, F. A. Alotaibi, M. A. AlAsiri, M. S. Alkharji, and S. A. Alharthi, "Predicting employee attrition using machine learning," *Proc. Int. Conf. Ind. Eng. Oper. Manag.*, no. November, pp. 1007–1014, 2019.

[18] R. Yedida, R. Reddy, R. Vahi, R. Jana, A. GV, and D. Kulkarni, "Employee Attrition Prediction," 2018, [Online]. Available: http://arxiv.org/abs/1806.10480

[19] S. Aggarwal, M. Singh, S. Chauhan, M. Sharma, and D. Jain, "Employee attrition prediction using machine learning comparative study," in *Intelligent Manufacturing and Energy Sustainability: Proceedings of ICIMES 2021*, 2022, pp. 453–466.

[20] L. S. Ganthi, Y. Nallapaneni, D. Perumalsamy, and K. Mahalingam, "Employee Attrition Prediction Using Machine Learning Algorithms," in *Proceedings of International Conference on Data Science and Applications: ICDSA 2021, Volume 1*, 2022, pp. *577*–596.

[21] Rana, A., Malik, S., Chauhan, M., & Kumar, P. (2023, September). Heart Disease Predictive Analysis Using Machine Learning Approaches. In 2023 6th International Conference on Contemporary Computing and Informatics (*IC3I*) (Vol. 6, pp. 2361-2367). IEEE.

[22] Javatpoint, "K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint." https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning (accessed Oct. 05, 2023).

[23] Iyortsuun, N. K., Kim, S. H., Jhon, M., Yang, H. J., & Pant, S. (2023, January). A review of machine learning and deep learning approaches on mental health diagnosis. In Healthcare (Vol. 11, No. 3, p. 285). MDPI.