Employee Attrition Prediction in an Organization Using Machine Learning Algorithms

Ajay Yadav[1], Prof, Sushma Ghose[2]

[1]*Student, Siddhant College of Engineering, Sudumbare*

[2]*Professor, Siddhant College of Engineering, Sudumbare*

Abstract: *The process of encouraging employees to stay with the organisation for as long as possible in order to get the most benefits to the business is known as employee retention. Employers have a responsibility to on-board their top talent. The majority of employees believe their value exceeds what they are paid. obviously there is a difference in perceptions of fair pay across individuals and organisations. Turnover may occur when there is too much of a difference and another opportunity presents itself. Thus, for businesses to succeed in the business world, it is vital to recruit and retain staff.*
This paper sheds light on system, that makes use of machine learning models such KNN, SVM, Regression and random forest to predict attrition that can help companies to retain the employees. Different employee-related features like Satisfaction Level, projects, Work-Life Balance, Workload, Work Environment and Compensation and Benefits are used that can be helpful to retain the employees. Furthermore, a comparison is done amongst machine learning algorithms, to discover the most accurate strategy for dealing with staff turnover.

Keywords: Employee Attrition Prediction, Machine learning Models, SVM, KNN, Random Forest, employee-related features, Employee Retention.

I.   Introduction

In this fiercely competitive corporate world, finding and retaining top talent is a challenge faced by organisations of all kinds. Employees are referred as an organization's "lifeline," and their contributions are essential to its expansion [1]. For organisations, predicting employee attrition is crucial since successful achievement of organisational goals depends on individuals who take their work seriously and make valuable contributions.

Employee attrition is a common phenomenon in the workforce and can have various implications for organizations. Employees leave the organization often due to reasons such as pursuing other career opportunities, dissatisfaction with the current job, personal reasons, or retirement or due to factors beyond their control, such as layoffs, restructuring, or termination for performance or disciplinary reasons [4,5,9].

Employee retention is the process of encouraging current employees to remain with the company; yet, in today's competitive environment, it can be challenging for organisations to manage and keep these personnel. Every company has to prioritise finding and keeping skilled workers since their knowledge and abilities are essential to succeed in the market. Moreover, when a company loses its best workers, it has a detrimental impact on both the profitability of the business and customer satisfaction. Therefore, it's important to identify the main issue causing individuals to abandon their jobs [5].

The purpose of this study is to predict those factors that bother employees to quit their jobs in advance, so that initiatives might be implemented to keep the employee. Employee reward programmes, flexible working hours,

timely promotions, career development programmes, and performance-based bonuses, compensation, wellness benefit, superior and subordinate relationships are the factors that play important role to attract and keep good employees.

Thus, the purpose of this study is to find out the factors responsible for an employee turnover and retention and then rank those factors in order of their importance and to predict employee turnover using historical data. Moreover, employee attraction and retention is done in an organisation to keep employees happy and motivated.

Thus employee attrition prediction is important affairs for:

- **Cost Saving [4]:** Identifying probable turnover enables organisations to take preventative measures to keep important staff. Recruiting and training new personnel can be expensive, thus avoiding excessive turnover can result in significant cost savings [5].
- **Enhance Work Productivity [5]:** Keeping attrition to a minimum help to ensure project and operational stability. Unexpected withdrawals can interrupt workflow, reduce production, and result in information losses. Predicting attrition enables organisations to mitigate these risks.
- **Workforce management:** By identifying employees who are more likely to leave, a company may concentrate on keeping and developing its most valuable personnel. This can involve specialised training, chances for professional advancement, and tailored retention tactics.
- **Improve the organization's image:** High personnel turnover can be detrimental to an organization's reputation. A corporation with a high attrition rate may be less likely to attract top talent. Predicting and resolving attrition effectively helps improve the organization's image as an employer of choice.

A. Research Contribution

It has been observed that the proper retention strategy must guarantee that employees' objectives are met in order to reduce employee turnover as a strategic problem. This research provided light on the various factors that contribute to employee attrition. Thus, investigating or identifying these factors supports in the development of a retention plan and contributes to the company's profitability. Thus the main contribution of the proposed paper is:

- Identifying significant factors contributing to employee turnover.
- To use machine learning models to predict attrition that can help companies to retain the employees.
- To compare the different ML models to find the most accurate technique for addressing employee turnover using historical data.

II. Literature Survey

In order to find the optimal model and avoid attrition, Mansor et al.[2] compared the performance of three machine learning techniques: Decision Tree (DT) classifier, Support Vector Machines (SVM) classifier, and Artificial Neural Networks (ANN) classifier. The IBM Human Resource Analytic Employee Attrition and Performance dataset was utilized for this purpose. Through the comparative analysis of the three classifiers, it was determined that the optimized SVM model demonstrated the highest accuracy rate of 88.87% in predicting employee attrition, outperforming the other models. Therefore, the primary focus of this research is on evaluating machine learning models for predicting employee turnover based on attrition data.

Using the IBM Watson dataset, Nandal et al.[3] offer an extensive strategy to estimate employee attrition utilising machine learning, ensemble approaches, and deep learning. Various techniques, including Logistic Regression Classifier, K-Nearest Neighbour (KNN), Decision Tree, Naïve Bayes, Gradient Boost, AdaBoost, Random Forest,

Stacking, XG Boost, "FNN (Feedforward Neural Network)", and "CNN (Convolutional Neural Network)" are used to predict employee attrition.

A methodology for forecasting employee attrition is put out by Chung et al.[4] so that talent management strategies can be implemented. Based on 30 characteristics that influence employee attrition, a prediction model was built using 1,470 records from the "IBM HR Analytics Employee Attrition & Performance data." In order to achieve this, eight prediction models including ensemble model, Random Forest, SVM, XGBoost, Logistic Regression, and Artificial Neural Network model were constructed and their performance was assessed.

Using an open-source (IBM HR) data set, Bada et al.[5] evaluated supervised machine learning techniques such logistic regression, decision trees, random forest classifiers, and XGBoost. Predicting whether the employees will leave their existing employers within the given time frame is the primary goal. The problem is designed to be a binary classification, wherein employees are categorised as either "will leave" (turnover) or "will not leave."
In order to forecast employee attrition in a company, Suvoj & Lydia [6] conducted an investigation focusing on the application of supervised machine learning algorithms such as Random Forest Classifier, Logistic Regression, Naive Bayes Classifier, AND XGBoost. They made use of a Kaggle data set that included details about a company's personnel. To identify the most effective model, the produced models were assessed using several evaluation metrics on the test set. The outcome demonstrated that, in comparison to the other two algorithms, the Random Forest classifier had an accuracy of 83%, while the XGBoost classifier had a better mean accuracy of 85.3%.

A survey on machine learning models for employee attrition prediction was conducted by Chourey et al. [7]. The writers argue that attrition is a serious issue in all organisations. An argument is presented for algorithms that use data mining techniques to forecast staff attrition. Furthermore, Mariam Al Akasheh [8], the author, resents a thorough and organised analysis of the machine learning (ML) methods employed over the previous ten years to forecast employee attrition. 52 pertinent, peer-reviewed studies that were released between 2012 and April 2023 in total were chosen. The findings show that more than 20 machine learning methods have been applied in different institutions to forecast staff churn.

III. Proposed System

Employee attrition refers to the process of employees leaving an organization, whether voluntarily (resignation) or involuntarily (termination). It is a common phenomenon in the workforce and can have various implications for organizations. The proposed system helps in predicting employee attrition, providing valuable insights that can guide decision-making and facilitate the creation of focused strategies to enhance employee retention and satisfaction. By analyzing historical data and identifying patterns, the system can take proactive measures to retain valuable talent. Different machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest and Support Vector Machines (SVM) are used to predict the employee attrition rate. Figure 1 shows the architecture of the proposed system. The working of the system is as follows:

The data was gathered from the Kaggle platform and pertains to various aspects of employees, such as demographics, job-related details, performance indicators, satisfaction levels, compensation and benefits, work environment, work-life balance, company culture and values, as well as other factors that could impact attrition.
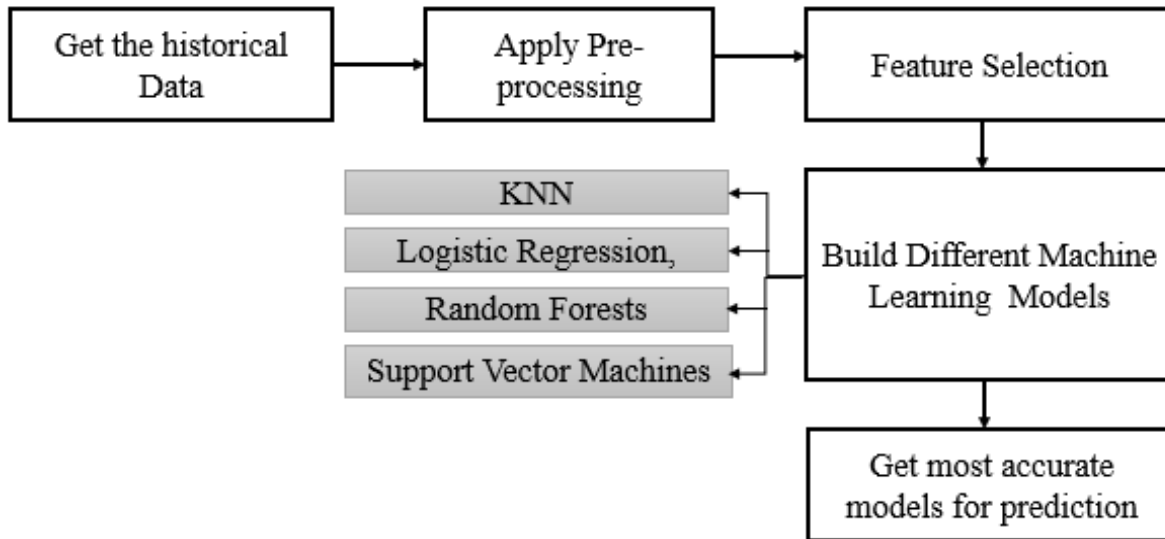
Figure 1 : System Architecture

In the realm of data analysis, data pre-processing and exploration play a crucial role. The data initially gathered may contain noise, necessitating pre-processing to refine the data. This includes addressing missing values and outliers. Additionally, exploring the data aids in comprehending the relationship between different features and the target variable, such as attrition. Feature engineering is subsequently employed to craft new features that may offer a more accurate representation of the underlying patterns.

The features such as demographics, job-related information, performance metrics, satisfaction, compensation and benefits, work environment, work-life balance, company culture and values are extracted from the data.

The employee attrition rate is predicted using many different kinds of machine learning algorithms, including logistic regression, random forests, support vector machines, and KNN. In order to evaluate the model's performance and choose the most accurate model for prediction, performance metrics are used. It has been shown that, among the all the ML model Random forest is outperforming all the other techniques in prediction of employee attrition rate.

A. Workflow of the proposed system

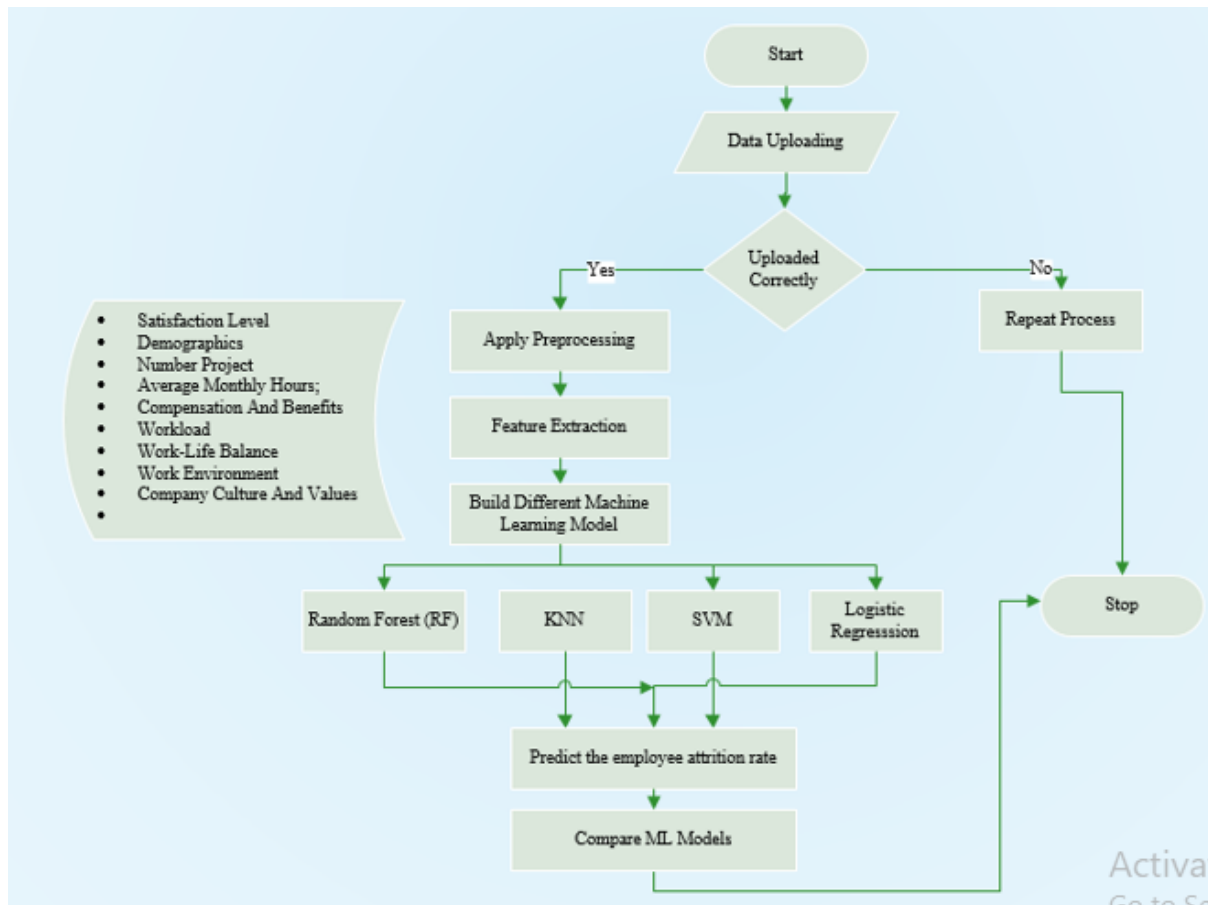Figure 2 describes the Workflow of the proposed system.

Figure 2: System Workflow

IV.   *Result Evaluation*

We have used WEKA (Waikato Environment for Knowledge Analysis) tool for evaluating the system performance. WEKA is a software library that contains machine learning algorithms, in order to solve practical data mining problems.

Random Forest:

First, we upload the dataset i.e. employee dataset from Kaggle to the WEKA and then run the Random forest classification algorithm. After providing the dataset as an input to the WEKA tool it gives the classification report that classifies the data in two classes as given in table I.

Class 0 – Employee will leave the organisation

Class 1 - Employee will not leave the organisation

Table I: Classification Report

Class	0	1
0	365	6
1	45	25

It helps to derive the values like true positive (TP), True Negative (TN), False positive (FP) and false Negative.
By using above values, the confusion matrix is calculated as illustrated in table.

Table II: TP, TN, FP and FN

	TP	TN	FP	FN
Class 0	365	25	45	6
Class 1	25	365	6	45

The values of TP, FP, and FN i.e truly positive, false positive and false negative can be used to calculate the
accuracy, precision, sensitivity and specificity using formulas mentioned below:

Precision=TP / (TP + FP)

Sensitivity = TP / (TP + FN)

Specificity = TN / (FP + TN)

Accuracy = TP+ TN/ TP+ TN+ FP+ FN

Table III: Confusion Matrix

Class	Accuracy	Precision	Sensitivity	Specificity
Class 0	88.44%	89.02%	98.38%	35.71%
Class 1	88.44%	80.65%	35.71%	98.38%

The graphical representation of the above table is represented in following graph. Thus the average c\accuracy of
the random forest for prediction of employee attrition rate is 88.44%.
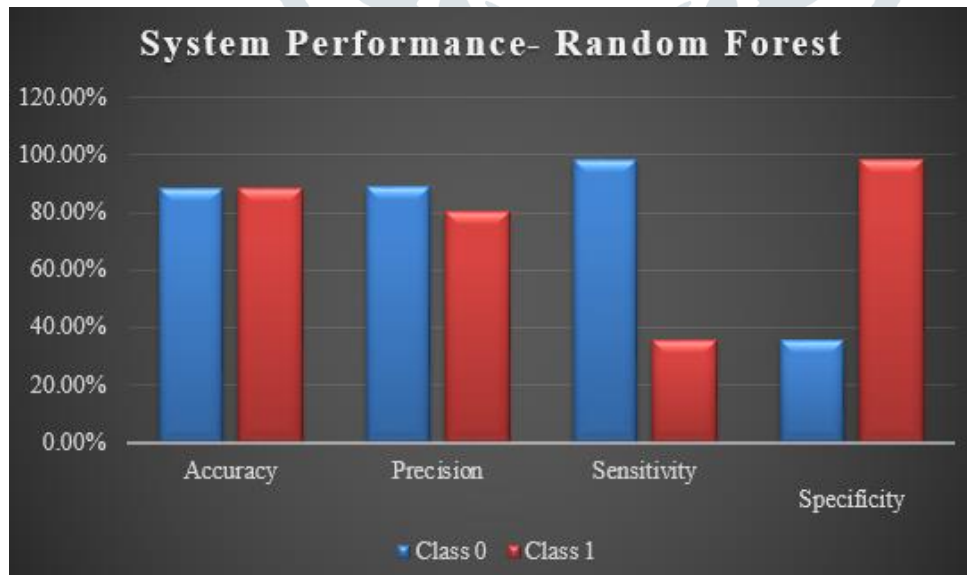

Figure 3 :System Performance

V.   *Conclusion*

Employee attrition prediction is not only about preventing turnover but also about fostering a positive work environment, retaining top talent, and strategically managing the organization's human capital for long-term success. The proposed system is beneficial for the company in order to predict the employee that likely willing to leave the organisation. Thus identifying the factors that bothering the employees help to make preventive strategies so that the turnover rate could be managed. The system shows that, demographics, job-related information, performance metrics, satisfaction, compensation and benefits, work environment, work-life balance, company culture and values plays important role in employee attrition and retention. The random forest proven to the best for predication the attrition rate with an accuracy of 88.44% as compared other mentioned techniques.

VI.   *References*

[1] Human Capital Management - Meaning and Important Concepts, available @ https://managementstudyguide.com/human-capital-management.htm

[2] Manju Nandal,Veena Grover,Divya Sahu,Mahima Dogra"Employee Attrition: Analysis of Data Driven Models",January 2024EAI Endorsed
    Transactions on Internet of Things 10

[3] Doohee Chung , Jinseop Yun b, Jeha Lee c, Yeram Jeon ,"Predictive model of employee attrition based on stacking ensemble learning",Expert Systems with Applications Volume 215, 1 April 2023, 119364.

[4] Olufunke Bada,Akinode John Lekan "EMPLOYEE ATTRITION PREDICTION USING MACHINE LEARNING ALGORITHMS",Conference: Proceedings of the 3rd International Conference, The Federal Polytechnic, Ilaro, 16th & 17th Aug, 2022.

[5] Suvoj Reddy Kovvuri& Lydia Sri Dommeti,"Employee Turnover Prediction A Comparative Study of Supervised Machine Learning Models" ,B.sc thesis is submitted to the Faculty of Engineering at Blekinge Institute of Technology
    in partial fulfilment of the requirements for the degree of Bachelor of Scienc in Computer Scienc.

[6] Chourey A, Phulre S, and Mishra S, "A survey paper on employee attrition prediction using machine learning techniques",2019.

[7] Mariam Al Akasheh a b, Esraa Faisal Malik c, Omar Hujran a b 1, Nazar Zaki,"A decade of research on machine learning techniques for predicting employee turnover: A systematic literature review",Expert Systems with Applications
    Volume 238, Part E, 15 March 2024, 121794.

[8] Vinston Raja,Srinath Doss ,Ashok Kumar k ,"Analytic Approach of Predicting Employee Attrition Using Data Science Techniques",DOI:10.21203/rs.3.rs-1833481/v1

[9] Andry Alamsyah; Nisrina Salma, "A Comparative Study of Employee Churn Prediction Model", IEEE International Conference on   Science and Technology, 2018