

Prediction of Employee attrition using Machine Learning Algorithm

Padmaja P¹, Dhanalakshmi R V², B.Sireesha³, Anand Patil⁴, K. Balanaga Gurunathan⁵

¹ Associate Professor, School of management science, Nalla Narasimha Reddy Education Society Group of institutions, Email: padmajaharinath1406@gmail.com

² Associate Professor, Department of Management Studies, New Horizon College of Engineering, Bengaluru, Email: dhanalakshmirv19@gmail.com

³ Assistant Professor, Department of Management Studies, Cambridge Institute of Technology, Bengaluru, Email: sireeshamba26@gmail.com

⁴ Associate Professor, School of Business and Management, Christ University, Bangalore, India, Email: anandp7@gmail.com

⁵ Professor in Finance and Research Head, School of Commerce and Management, Jain University, Bengaluru, Email: balanagagurunathan@yahoo.com

Received: 18.04.2024

Revised : 21.05.2024

Accepted: 27.05.2024

Abstract

Making decisions can play a big part in administration and can even point to the crucial element in the planning process. In order to retain talented workers, management must effectively handle the well-known issue of employee attrition. It's exciting to know that ML algorithms can accurately forecast staff attrition. The aim of this research was to identify the key variables affecting employee turnover and to forecast employee attrition with a high gradation of accuracy using machine and deep learning models. The Kaggle Depository provided the dataset that was used in this investigation. The dataset, which includes 24 attributes from 1,470 employees, was produced by IBM Analytics. In order to optimize the prediction accuracy of employee attrition, the dataset was preprocessed, balanced, and divided into three distinct sets: the train, valid, and test datasets. A number of tests were conducted to demonstrate the applicability of this research. This research used JASP software to analyze the data. The best machine learning model archived f1-scores of 82.6%, precision 81.8%, recall 84.7% and accuracy 84.6% for the prediction of employee attrition.

Keywords: Employee Attrition, K-Nearest Neighbor, Machine learning, Prediction

1. INTRODUCTION

One important aspect of how organizations contend with each other is staff productivity. Establishing and preserving a good atmosphere is the primary component in the contributions of dependable and cooperative personnel. The human resources department needs to investigate the backgrounds of its personnel in order to establish a welcoming atmosphere. Management is able to make better decisions and stop employee churn by examining these datasets. When workers leave a company for any reason, it's referred to as employee attrition. Attrition is a major issue for all of the largest multinational corporations. Attrition went from 10% to 20% between 2020 and 2021, according to research firm Gartner. Employee turnover has a range of effects on businesses. HR managers might greatly help in resolving this situation and preventing employee attrition by using machine learning classification models to assess whether an employee is likely to depart the organization. To ensure that prediction whether an employee would stay with the company or leave, we developed a system in this project that makes use of several machine learning categorization models.

2. Previous Studies

Researchers focused at employee churn from several perspectives. A few researchers looked at how employees behaved to find out what aspects of their decision affected them to stay or leave the organization. Researchers looked at the employee attrition problem from a variety of angles. G. Gabrani, A. Kwatra. A few academics examined worker behavior to identify the factors that influence employees' decisions to quit or remain with the company.

K-Nearest Neighbor

The supervised learning approach is the foundation of one of the simplest machine learning algorithms, k-Nearest Neighbor. The KNN method functions under the premise that the new data and the examples that already exist are comparable. The new case is categorized into the existing ones. All the data is stored via the KNN algorithm. A fresh data point is classified according to the similarity. This means that the KNN algorithm can quickly and simply classify newly discovered data into an appropriate category. The KNN technique is applicable to both classification and regression. It is primarily applied to classification issues. The algorithm KNN is non-parametric. It doesn't assume anything about the underlying data. The algorithm is referred to as a lazy learner because, instead of learning immediately from the training set, it stores the information and uses it for classification. The KNN method only stores the dataset and groups freshly acquired data into a subset that closely resembles the innovative data during the training phase.

Research Authors	Problem Studied	Techniques Studied	Recommended
Hong, Wei & SChen	Voluntary employee Turnover predictions	Logistic & probability regression model.	Logistic regression model.
G.Renuka	Employee Attrition	SVM, KNN, Decision tree	KNN & Light GBM
Marjorie Laura Kane sellers	Impacting variables on voluntary turnover of employees.	Binomial & logistic Regression	Binomial & logistic Regression

Rahul Y	Employee Attrition Prediction	KNN, Naïve Bayes	KNN
Pratik T	Customer Behaviour Analysis based on their purchasing	KNN	KNN
Saradhi and Palshikar	Data mining techniques in churning the employees	Naïve Bayes, SVM, Decision Trees	SVM

3. Problem Statement

Employee retention is a foremost factor in organizational development. Retaining current staff is less costly than hiring new ones, and firing personnel typically lowers the company's self-esteem. To forecast staff attrition, we are putting forth deep and machine learning methodologies.

4. Objectives

The objectives are:

- To increase the accuracy of Employee attrition using KNN
- To analyze the data set to find out the variables (features)
- To estimate the suggested models to determine the most effective method of employee attrition.

5. METHODOLOGY

Data set Collection

A prediction system's first step is to collect data and divide it into training and testing datasets. 20% of the dataset was used for testing, while 80% was used for training in this project. Employee datasets contain all the necessary attributes of employees to predict employee attrition.

Variable Selection

Variable Name	Data Type	Variable Name	Data Type
Age	Numeric	Distance from home	Numeric
Monthly Income	Numeric	Overtime	Catagorical
No.of Companies Worked	Numeric	Education	Categorical
Department	Categorical	PercentSalaryhike	Numeric
Gender	Categorical	PerformanceRating	Numeric
TotalWorkinghours	Numeric	Relationship satisfaction	Categorical
Joblevel	Categorical	EnvironmentSatisfaction	Categorical
Yearsat Company	Numeric	Stockoption level	Categorical
JobRole	Categorical	Trainingtimelastyear	Numeric
JobInvolvement	Categorical	Worklife balance	Categorical
Yearsat company	Numeric	Yearssincelast promotion	Numeric
Standardhours	Numeric	Job Satisfaction	Categorical

Data Preprocessing

Data pre-processing plays a major part in machine learning algorithms since it prepares the data in the necessary format. This keeps things from producing deceptive results. Missing values, skewed data, redundant data, and category values are all addressed via data pre-processing. Using label encoding helps manage dummy variables.

6. FINDINGS & DISCUSSION

Table 1.KNN Classification

K-Nearest Neighbors Classification

Nearest neighbors	Weights	Distance	n(Train)	n(Test)	Test Accuracy
3	rectangular	Euclidean	1176	294	0.847

According to the above table, 294 interpretations from the test set's indicator variables were utilized to calculate the prediction error, whereas 1176 data (or 80% of the 1470 total) are trained our k=3 closest neighbor model. The bar showed in Fig.1 further visualizes the data split information.

Data Split



Figure 1. Data split

Our k=3 closest neighbor model obtained 0.847 classification accuracy of a test set, as indicated by the output of the following figure. This means that we can accurately predict 84.7% of our holdout test data.

Table 2. Confusion Matrix

		Predicted	
		No	Yes
Observed	No	238	12
	Yes	33	11

The confusion matrix on the test set, which illustrates how the included observations were anticipated using the model, offers more details about the prediction accuracy. Table 2 showed that 238 "NO" and 11 "YES" were properly predicted by the algorithm, yielding a test accuracy of $238 + 12 / 294 = 0.8469$.

We are able to make firm predictions for a new case in the form of "YES, employee is going to churn" or "NO, employee is not going to churn" thanks to the training set. We are able to ration the degree of uncertainty in the prediction thanks to the test set.

Using the confusion matrix we are able to anticipate more precisely of the firm "YES, employee is going to churn with 0.47% chance" and "No, employee is not going to churn with 0.87% chance". Based on the result it is observed that, we have relative less employee attrition in our test compared to training set.

Evaluation Metrics

	Precision	Recall	F1 Score	Support	AUC
0	0.878	0.952	0.914	250	0.617
1	0.478	0.250	0.328	44	0.617
Average / Total	0.818	0.847	0.826	294	0.617

Note. Area Under Curve (AUC) is calculated for every class against all other classes.

Evaluation procedures that are used to rate a machine learning model's efficacy and performance. Based on the evaluation metrics results the KNN Classifier: F1-score (82.6%) which indicates the suggested model is good, Precision (81.8%), Recall (84.7%).

7. Future Research

In the future, the findings might be compared to the basic KNN model and the method could be modified to weight neighbors differently, giving closer neighbors a higher weight contribution to the fit instead of giving all neighbors the same weight.

8. CONCLUSION

This study presented the effect of Employee attrition on organizations, and why predicting it is crucial. This study further continued and outlined the KNN algorithm to solve the prediction problem. The results of the present study showed the superiority of the KNN classifier accuracy i.e., 84% and predictive effectiveness. It is a reliable technique that, when employed in its ideal configuration, produces correct results despite the noise in the dataset, which poses a important problem for machine learning algorithms. Therefore, accurately forecast employee attrition in an organization and allow HR to take the required steps to retain employees who are anticipated to be at danger of leaving, the authors advise using the KNN classifier.

REFERENCES

- [1] G. Gabrani, A. Kwatra. Machine Learning Based Predictive Model for Risk Assessment of Employee Attrition. In International Conference on Computational Science and Its Applications; Springer: Melbourne, Australia, pp. 189–201, 2018.
- [2] W. C. Hong, S. Y. Wei, and Y. F. Chen, "A comparative test of two employee turnover prediction models", International Journal of Management, 24(4), 808, 2007.
- [3] L. K. Marjorie, "Predictive Models of Employee Voluntary Turnover in a North American Professional Sales Force using Data-Mining Analysis", Texas, A&M University College of Education, 2007.
- [4] V. V. Saradhi and G. K. Palshikar, "Employee churn prediction", Expert Systems with Applications, 38(3), 1999-2006, 2011.
- [5] G. Renuka, "Employee Attrition Prediction Using Machine Learning", International Research Journal of Modernization in Engineering Technology and Science, Vol 4, Issue 5, May 2022.
- [6] Yedida, Rahul & Reddy, Rahul & Vahi, Rakshit & Jana, Rahul & Gv, Abhilash & Kulkarni, Deepti. (2018). Employee Attrition Prediction.
- [7] T.Pratik, "Customer Behavior Analysis: Identifying risky customers based on their purchased product on e-commerce", International Research Journal of Engineering & Technology, Vol.7, Issue 9, p- ISSN: 2395-0072, 2020.
- [8] M. Sameer, "Prediction Employee Attrition and Performance Using Deep Learning", Journal of Theoretical and Applied Information Technology, Vol. 100, No.21, ISSN: 1992-8645, 2022.

- [9] D. S. Rupa Chatterjee Das. Conceptualizing the Importance of HRA in Attrition Reduction. Int. 2 Res.J. Adv. Sci. Hub, vol. 2, pp.40–48, 2020.
- [10] Parv N Gandhi, “Employee Attrition Prediction using Machine Learning”, Journal of Emerging Technologies & Innovative Research, Vol. 7, Issue 9, ISSN-2349-5162, 2020.
- [11] Rohit P& Pankaj Ajit, “Prediction Of Employee Turnover In Organizations Using Machine Learning Algorithm”, International Journal Of Advanced Research In Artificial Intelligence (IJARAI) Vol. 5, Issue. 9, 2016.
- [12] Nagadevara.V, Srinivasan.V, and Valk.R, “Establishing a link between employee turnover and withdrawal behaviors: Application of data mining techniques”, Research and Practice in Human Resource Management, Vol.16, pp. 81-97, 2008.