International Journal of Enhanced Research in Management & Computer Applications
ISSN: 2319-7471, Vol. 13, Issue 4, April-2024, Impact Factor: 8.285
Presented at "ICRETETM-2024", Organized by GSMCOE, Pune, on 22nd - 23rd April 2024

Employee Turnover Prediction

Prof.Neelam Jadhav[1], Bhirud Gunjan[2], Barhe Harshal[3], Dendage Kishor[4], Shelar Pramod[5], Barole Chaitanya[6]

Assistant Professor, Dept. of Computer Engg, GS Moze College ofEngineering, Balewadi Pune, India.

ABSTRACT

Turnover intention is an employee's reported willingness to leave her organization within a given period oftime and is often used for studying actual employee turnover. Since employee turnover can have a detrimental impact on business and the labour market at large, it is important to understand the determinants of such a choice. We describeand analyse a unique European wide survey on employee turnover intention. A company's most valuable resource is its workforce, which includes each worker. Because of the crucial role that employees play in the success of an organization, measuring employee turnover rate has become one of the most important metrics that businesses are concentrating on in the modern era.

Keywords: Machine learning, Artificial intelligence, Data mining, Data analytics, Data visualization, Feature selection,Model stability Employee turnover.

INTRODUCTION

Any business can benefit from having more workers. When an individual begins working for a company, it is inevitable that they will quit at some point in the future for a variety of reasons. Attrition can be defined as the departure of any employee owing to events that may or may not be within their control, such as retirement, death, transfer, or the pursuit of better opportunities. When an employee is being hired, the hiring company invests a significant amount of time and a significant number of resources in the process [1]. When an employee's departure begins to have a negative impact on a company, it becomes a source of concern for everyone in the company, but particularly for its human resources department. Such a business not only suffers the loss of its competent experts because of the departure of qualified employees but must also rehire and educate the individual who replaces them [2]. This results in a decrease in the company's staff and has a negative impact on the company. There has been a tremendous uptick in opportunities across the board as a direct result of growing globalization, particularly in the period after the epidemic.

An employee makes the decision to leave one company and join another to pursue new opportunities and advance their professional development [3]. The loss of employees due to attrition has a negative impact on a company's operations for a limited amount of time. Incorporating artificial intelligence into the process of predicting attrition is one way tokeep workforce size stable while also cutting expenditures. This article presents a discussion of the numerous approaches that may be taken to forecast employee turnover, and it also includes an analysis of the most effective solution that was conducted by comparing different models.

TABLE AND DIAGRAM

satisfaction	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	l eft	promotion_last_5years	sales	salary
0.38	0.53	2	157	3	0	1	0	s ales	low
0.8	0.86	5	262	6	0	1	0	Sales	medium

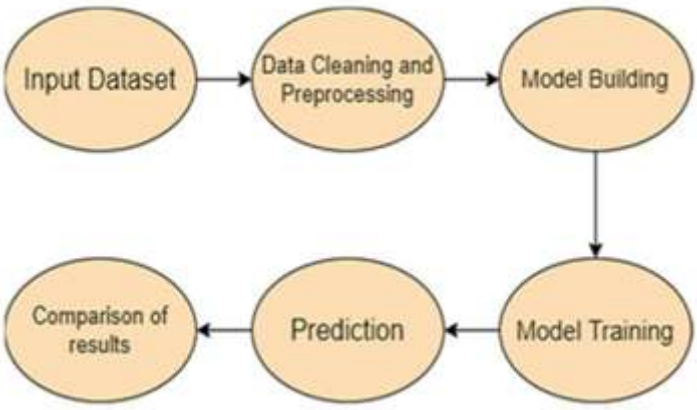
0.11	0.88	7	272	4	0	1	0	Sales	medium
0.72	0.87	5	223	5	0	1	0	Sales	Low
0.37	0.52	2	159	3	0	1	0	Sales	Low
0.41	0.5	2	153	3	0	1	0	Sales	Low
0.1	0.77	6	247	4	0	1	0	Sales	Low
0.92	0.85	5	259	5	0	1	0	sales	Low
0.89	1	5	224	5	0	1	0	Sales	Low
0.42	0.53	2	142	3	0	1	0	Sales	Low
0.45	0.54	2	135	3	0	1	0	Sales	Low
0.11	0.81	6	305	4	0	1	0	Sales	Low
0.84	0.92	4	234	5	0	1	0	Sales	Low
0.41	0.55	2	148	3	0	1	0	Sales	Low
0.36	0.56	2	137	3	0	1	0	sales	low

LITERATURE REVIEW

Several scholars have investigated the factors that lead to employee turnover as well as its consequences. According to one of the corresponding papers, ref [4], the upkeep of skilled and deserving employees is a crucial factor that HR departments need to pay attention to in order to be successful. The cited study identifiedthe most pertinent metrics that could be of assistance in the endeavour of forecasting employee turnover. It was pointed out that an employee's level of education and experience are directly proportionate to the numberof work opportunities that are available to them. It was also claimed that some of the most pleasant things that help in sustaining a workforce include having a decent work–life balance, having healthy connections with co-workers, having better policies, and so on. According to the findings of another study of this kind [5,6], for a business to realize the greatest possible profit, it must treat its workforce with the utmost respect and show a great deal of concern.

Sri Ranjitha Ponnuru and colleagues predicted staff turnover using Machine Learning algorithms. IBM HR research informed the forecast. They made predictions using Logistic Regression and obtained 85% accuracy[11]

METHODOLOGY



Machine Learning Algorithms

A. Logistic Regression: It is generally agreed that Logistic Regression is one of the most useful statistical models available. In addition, it is a well-known method of data mining that scientists and other researchers employ for the examination of both proportional and binary types of datasets. One of the things that sets logistic regression apart fromother types of regression is the fact that it may be used to analyse data from more than one class [8]. It is one of the most frequently used classification algorithms across the globe.

B. Decision Tree: The tree structure is visualized whenever the term "tree" is used in the vernacular of computer science. Root, branches, and leaves make up the components of a decision tree. It is common practice to refer to the root node as the parent node. Nodes are used to represent each characteristic, while branches are used to indicate the connections that are made between the nodes.

C. K-Nearest Neighbor(KNN): KNN is an algorithm for supervised machine learning that can be applied to issues involving classification as well as regression. The KNN algorithm makes predictions about the output by making use of knowledge about the input [16]. The input is then divided into the appropriate categories. The algorithm tends to look for the position that will provide the best results for a new datapoint to fit in. Following an analysis of the data points that were provided as input, a decision is made regarding the location of the new point. The algorithm that was employed in this study is as follows:

1. Step 1: The first step is to choose the number of K, i.e., the neighbourhood;
2. Step 2: The second step is to compute the distance (Euclidean);
3. Step 3: Using, positions of data points, locate the K individuals that are geographically closest to a given position;
4. Step 4: The fourth step is to tally the number of points earned in each category;
5. Step 5: The fifth step involves assigning the newly acquired points to a category in which the surrounding points are greater in number;
6. Step 6: Finish

D. Support Vector Machines (SVMs): Another popular variety of supervised machine learning models is called a support vector machine (SVM). Its primary application is in addressing classification difficulties, although it can also be utilized to address regression settings. The fundamental concept behind the method is to draw a line or establish a border that delineates n distinct groups or classifications inside the overall space [18]. As a new data point is added to this space, it can quickly locate its proper location within the categories that have been formed. A hyperplane is another name for the line that cuts across the middle of these classes. A classification problem is said to have a linear algorithm when it can be solved by drawing a single straight line. It is referred to as a non-linear support vector machine (SVM) when a straight line is insufficient, in which case a curved line is obtained instead [15].

E. Random Forest: A machine learning technique, Random Forest is utilized for solving problems involving regression and classification. It is something that has been passed down from the idea of ensemble learning. It is comparable to the use of decision trees. The dataset is split up into numerous subsets so that this technique can consider the many different types of trees. Because of this procedure, several outcomes are combined into a single result, which is then calculated as an average of all the component results. The accuracy of the algorithm improves in proportion to the number of trees and sub datasets that are taken into consideration. As a result of this behaviour, it can manage an extremely large number of datasets [10]. The steps of the algorithm are shown below:

1. Step 1: The first step is to pick K datapoints at random from the training set;
2. Step 2: Constructing decision trees for each subset is the second step;
3. Step 3: Choose the number N that will represent the number of decision trees;
4. Step 4: Execute S1 and S2 once again;
5. Step 5: According to the predictions made by each tree, assign each new datapoint to the appropriate category

CONCLUSION

MODEL	ACCURACY
Logistic Regression	87.71%
KNN Classifier	59.22%
Support Vector Machines	86.59%
Naive Bayes	83.24%
Decision Trees	80.45%
Random Forest	83.24%

.

REFERENCES

[1]. S. Jahan, "Human Resources Information System (HRIS): A Theoretical Perspective", Journal of Human Resourceand Sustainability Studies, Vol.2 No.2, Article ID:46129, 2014
[2]. Alcover, C.M.; Guglielmi, D.; Depolo, M.; Mazzetti, G. Aging-and-Tech Job Vulnerability: A proposed framework on the dual impact of aging and AI, robotics, and automation among older workers. Organ. Psychol. Rev.2021, 11, 175–201. [CrossRef]
[3]. Moldoveanu, M.; Narayandas, D. The future of leadership development. Harv. Bus. Rev. 2019, 97, 40–48.
[4]. Fallucchi, F.; Coladangelo, M.; Giuliano, R.; William De Luca, E. Predicting employee attrition using machinelearning techniques. Computers 2020, 9, 86. [CrossRef]
[5]. Saxena, U.R.; Sharma, P.; Gupta, G. Comprehensive Study of Machine Learning Algorithms for Stock MarketPrediction During COVID-19. J. Comput. Mech. Manag. 2023, 2, 1–7. [CrossRef]
[6]. Manikandan, R.; Maurya, R.K.; Rasheed, T.; Bose, S.C.; Arias-Gonzáles, J.L.; Mamodiya, U.; Tiwari, A. Adaptive cloud orchestration resource selection using rough set theory. J. Interdiscip. Math. 2023, 26, 311–320.[CrossRef]
[7]. Srivastava, P.K.; Kumar, S.; Tiwari, A.; Goyal, D.; Mamodiya, U. Internet of thing uses in materialistic ameliorate farming through AI. In Proceedings of the AIP Conference Proceedings, Jaipur, India, 6–7 May 2022;Volume 2782.
[8]. Ravula, A.K.; Ahmad, S.S.; Singh, A.K.; Sweeti, S.; Kaur, A.; Kumar, S. Multi-level collaborative framework decryption-based computing systems. In Proceedings of the AIP Conference Proceedings, Jaipur, India, 6–7 May2022; Volume 2782.
[9]. Ozdemir, F.; Coskun, M.; Gezer, C.; Gungor, V.C. Assessing Employee Attrition Using Classifications Algorithms. In Proceedings of the 2020 the 4th International Conference on Information System and Data Mining,Hawaii, HI, USA, 15–17 May 2020; pp. 118–122.
[10]. Shipe, M.E.; Deppen, S.A.; Farjah, F.; Grogan, E.L. Developing prediction models for clinical use using logisticregression: An overview. J. Thorac. Dis. 2019, 11, S574. [CrossRef] [PubMed]
[11]. Ravula, A.K.; Ahmad, S.S.; Singh, A.K.; Sweeti, S.; Kaur, A.; Kumar, S. Multi-level collaborative framework decryption-based computing systems. In Proceedings of the AIP Conference Proceedings, Jaipur, India, 6–7 May 2022; Volume 2782.
[12]. Ozdemir, F.; Coskun, M.; Gezer, C.; Gungor, V.C. Assessing Employee Attrition Using Classifications Algorithms. In Proceedings of the 2020 the 4th International Conference on Information System and Data Mining,Hawaii, HI, USA, 15–17 May 2020; pp. 118–122.
[13]. Shipe, M.E.; Deppen, S.A.; Farjah, F.; Grogan, E.L. Developing prediction models for clinical use using logisticregression: An overview. J. Thorac. Dis. 2019, 11, S574. [CrossRef] [PubMed]
[14]. Jijo, B.T.; Abdulazeez, A.M. Classification based on decision tree algorithm for machine learning. Evaluation 2021, 6, 7.
[15]. Reddy, E.M.K.; Gurrala, A.; Hasitha, V.B.; Kumar, K.V.R. Introduction to Naive Bayes and a Review on Its Subtypes with Applications. In Bayesian Reasoning and Gaussian Processes for Machine Learning Applications;Chapman and Hall/CRC: New York, NY, USA, 2022; pp. 1–14.
[16]. Ponnuru, S.; Merugumala, G.; Padigala, S.; Vanga, R.; Kantapalli, B. Employee attrition prediction using logisticregression. Int. J. Res. Appl. Sci. Eng. Technol. 2020, 8, 2871–2875. [CrossRef]
[17]. Mishra, S.; Mishra, D. Review of literature on factors influencing attrition and retention. Int. J. Organ. Behav.Manag. Perspect. 2013, 2, 435–444.
[18]. Allan, G. Qualitative research. In Handbook for Research Students in the Social Sciences; Routledge, TaylorFrancis: London, UK, 2020; pp. 177–189