A SURVEY ON EMPLOYEE TURNOVER PREDICTION USING MACHINE LEARNING

Kanchan Pradhan*1, Sumeet Suryawanshi*2, Abhishek Sawalkar*3, Samir Pathan*4

*1Professor, Computer Engg., JSPM's Bhivarabai Sawant Institute Of Technology and Research, Wagholi, Pune, Maharashtra, India.

*2,3,4Student, Computer Engg., JSPM's Bhivarabai Sawant Institute Of Technology and Research, Wagholi, Pune, Maharashtra, India.

ABSTRACT

Employee turnover is the measure of total number of employees who leave the company over a certain time frame. Along with fast development of economic and industrial growth, employee turnover phenomenon has become popular gradually in recent years. Employee turnover has been identified as a key issue for organizations because of its adverse impact on work place productivity and long-term growth strategies. It is considered a serious challenge for organizations and companies. Organizations need to strategize to reduce the turnover goals of the workers to have a competitive advantage over other organizations. Furthermore, Organizations need to grasp the major factors of employee turnover, and then take relevant measures to deal with this problem. This paper studies the machine learning algorithms that are used to predict employee turnover. Some of the popular algorithms used for this prediction are Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Logistic Regression. Organizations can use this predictive analysis to measure the number of employees that needs to be hired in place of the ones that are leaving.

Keywords: Employee Turnover, Machine Learning, Classification, Support Vector Machine, K-Nearest Neighbors, Decision Tree, Logistic Regression.

I.   INTRODUCTION

Employee churn became one of the biggest issues, particularly the churn of key talents or employees with the crucial roles. Organizations are impacted by turnover in at least 3 ways.

1. Firstly, Replacement hires costs paid by the company, expense of training, headhunter price for finding new talent that will replace the earlier would demand heavy investment.

2. Moreover, Performance of the organization decrease due to the lead time for hiring and training the replacement may not be able to fully realize potential or adapt to culture last but not least.

3. Further, Losing between 6 to 12 months competitiveness of the company, particularly with regard to the topic. Expertise in a certain field, which is difficult to locate on the job market, affects an organization's capacity and investor's confidence.

One of the approaches, to tackle this is by conducting an exit interview with a resigned employee and using a quantitative and qualitative survey form, traditional technique of finding factors that cause voluntary turnover. However, Employee do not always participate in off boarding processes, may not be truly forthcoming in the HR exit interview, and by the time the exit interview comes around it's too late to address the issues which caused the employee to leave in the first place. If the workforce is large enough, it is helpful to be able to predict which employees are at risk of leaving at any given time, how long they are expected to stay, and get a hint of which interventions may have a chance of reducing turnover.

II.   EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis allows us to describe the data into understandable form. A dataset from Kaggle is used to explain the term in lucid form.
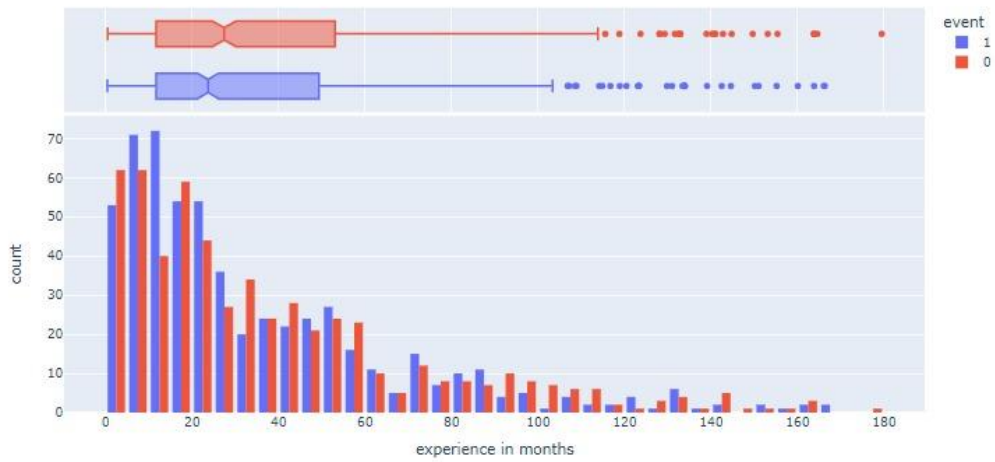
Above graph represents the relation between experience of the employee in months and the number of employees.
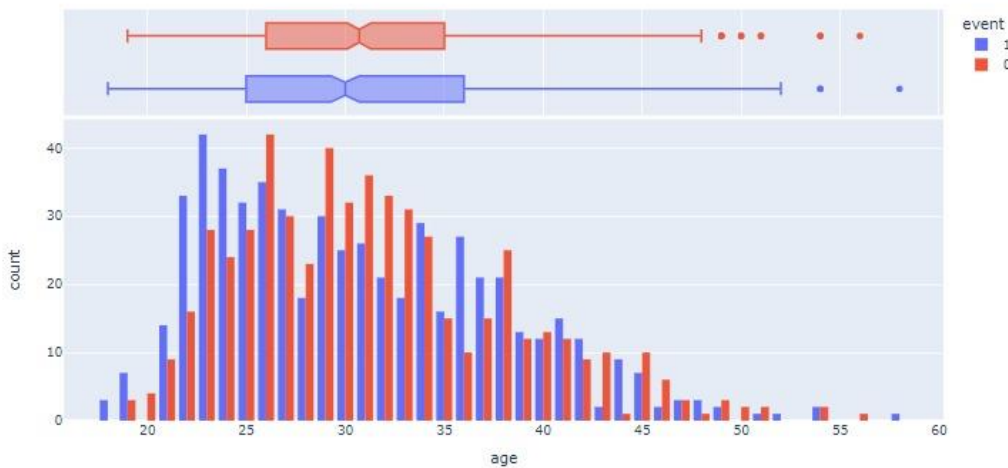
Similarly, this graph represents the relation between age of the employee and number of employees.

III.   ALGORITHMS ANALYSIS

Support Vector Machine:

Support Vector Machine, popularly abbreviated as SVM, is one of the most popular Supervised Learning Algorithms. It is used for Classification as well as Regression problems. This algorithm creates decision boundary that can segregate n-dimensional space into classes. This decision boundary is called as hyperplane which is created by support vectors.

SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable.
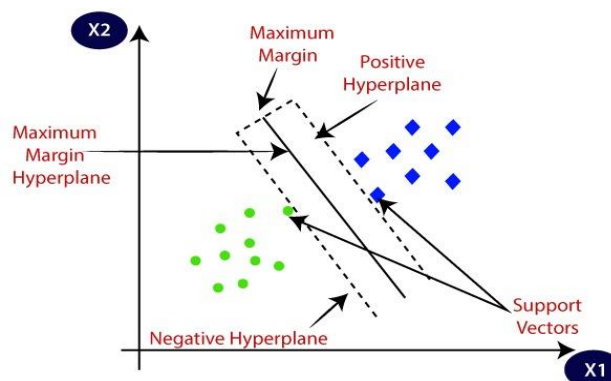
K-Nearest Neighbors:

K-Nearest Neighbors algorithm is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

Using 'feature similarity' to predict the values of new datapoints.
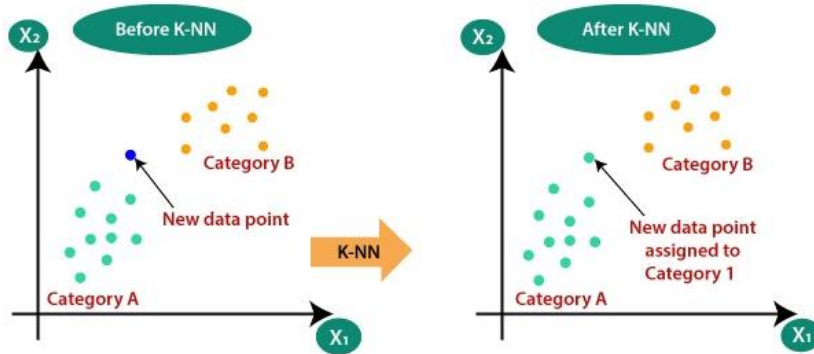

Decision Tree:

Decision Tree is a Supervised learning Technique that can be used for both Classification and Regression problems. As the name suggests, Decision Tree is a structure where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

All the decisions are performed on the basis of features of the given dataset. It is graphical representation that gets all the possible solutions to a problem/decision based on the given conditions.
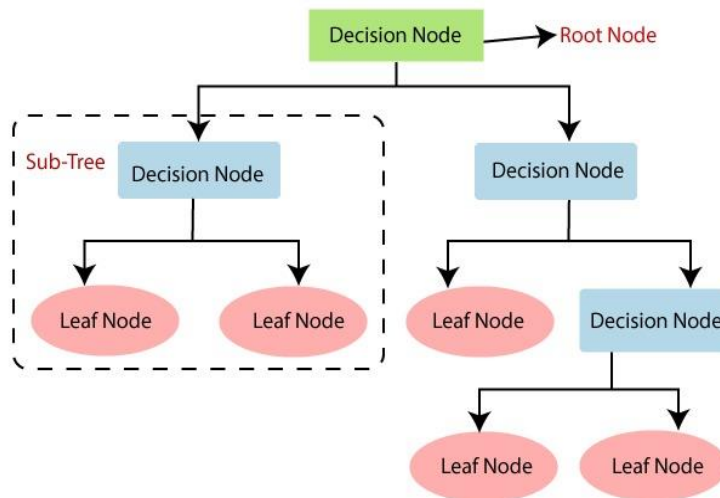
Logistic Regression:

Logistic Regression, as the name suggests, categorizes data based on True or False values. It also falls under classification problems category in Supervised Learning. The output is probabilistic values which lie between 0 and 1. The mapping of the predicted values to probabilities is done by a mathematical function called sigmoid function.
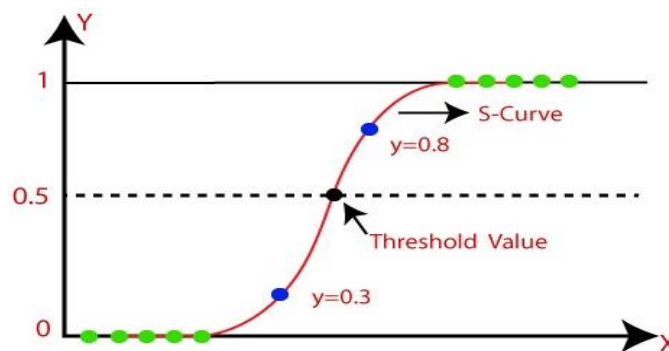
IV.   CONCLUSION

Employee turnover has been an issue for long. Regular development of the workforce will ensure that your employees are capable of adapting and meeting the demands of their jobs every day. If workforce of the organization, gets a sudden drop of employees it would surely come as a shock in most cases. This will result in losing the trust of the client, financial loses to recover from the loss of workforce and will eventually make organizations fearful to take on projects that depend highly on employees. However, with the early prediction of the turnover would give organizations enough time to react and prevent great losses.

V.   FUTURE WORK

Future work could include increasing size of dataset for more accurate performance. By training on such huge data, we could compare the accuracy, recall score, error rate, f1 score and conclude the best of the implemented algorithms. Using this accurate analysis, organizations will benefit greatly and would have sufficient time to implement retention strategies. This would result in more robust structure of the company.

VI.   REFERENCES

[1]   2021 13th International Conference on Knowledge and Smart Technology (KST) | 978-1-7281-7602-4/21/$31.00 ©2021 IEEE.

[2]   Qianwen Zhu, D Jiaxing Shang, Xinjun Cai, Linli Jiang" COXRF: Employee Turnover Prediction Based on Survival Analysis", 2019 IEEE Smart world, Ubiquitous Intelligence & Computing.

[3]   S K Monisaa Tharani[1], S N Vivek Raj[2]" Predicting Employee Turnover Intention in It&Ites Industry Using Machine Learning algorithms" Proceedings of The Fourth International Conference On I-SMAC IEEE Xplore.

[4]   Dishala Sandamini Rodrigo, Gayashini Shyanka Ratnayake" Employee Turnover Prediction System: With Special Reference to Apparel Industry in Sri Lanka", 2021 6th International Conference for Convergence in Technology (I2ct) Pune, India. Apr 02-04, 2021.

[5]   Xinjun Cai1,2, Jiaxing Shang1,2, Feiyi Liu1,2, Ziwei Jin1,2, Baohua Qiang3,4, Wu Xie5, And Liang Zhao6" Dbge: Employee Turnover Prediction Based on Dynamic Bipartite Graph Embedding" 10.1109/Access.2020.2965544, IEEE Access.

[6]   Dilip Singh Sisodia, Somdutta Vishwakarma, Abinash Pujahari" Evaluation of Machine Learning Models for Employee Churn Prediction" 978-1-5386-4031-9/17/$31.00 ©2017 IEEE.

[7]   Samarth Bhargava and Eli Yablonovitch" Lowering Hamr Near-Field Transducer Temperature Via Inverse Electromagnetic Design" IEEE Transactions on Magnetics, Vol. 51, No. 4, April 2015.

[8]   Heng Zhang1, Lexi Xu1, Xinzhou Cheng1, Kun Chao1, Xueqing Zhao2" Analysis and Prediction Employee Turnover Characteristics of Based on Machine Learning" 978-1-5386-8458-0/18/$31.00 ©2018 IEEE.

[9]   Apurva Mhatre1, Avantika Mahalingam2, Mahadevan Narayanan3, Akash Nair4, Suyash Jaju5 "Predicting Employee Attrition Along with Identifying High Risk Employees Using Big Data and Machine Learning" 2020 2nd International Conference on Advances in Computing.