**EMPLOYEE QUIT PREDICTION USING MACHINE LEARNING TECHNIQUES**

Sunil Kumar Swain1,Sk Nafish Alam2,, Manoj Kumar Behera3, DR.S. Chakravarty4

1, 2, 3,4 Dept. of CSE, Centurion University of Technology & Management, Odisha, 752050, India[1](mailto:1darshanray8555@gmail.com) [1sunilkumarswain645@gmail.com,2shaikhnafishalam786@gmail.com,3](mailto:1sunilkumarswain645@gmail.com,2shaikhnafishalam786@gmail.com,3)[manojbehera@cutm.ac.in](mailto:manojbehera@cutm.ac.in), 4[sujata.chakravarty@cutm.ac.in](mailto:sujata.chakravarty@cutm.ac.in)

**Abstract** - Employees are considered as backbone of an organization. Success or failure of the organization depends on the employees who work for an organization. The organizations have to face the problems when trained, skilled and experienced employees leave the organization for better opportunities. The study was mainly undertaken to identify the dissatisfaction factor of employees and for what reasons they prefer to change their jobs. Once the dissatisfaction factor/s of employees has/have been identified, the organizations can take actions accordingly and it may help them to reduce the attrition rate. We basically trying to build a system which will help the employee for attrition based on Employee dataset. We generated heatmap to show the relations between the attributes. For prediction purpose, we have used four different machine learning algorithms such as KNN (K-Nearest Neighbor), SVM (Support Vector Machine), Decision Tree, Light GBM This paper suggest reasons which optimize the employee attrition in any organization.

***Keyword:*** Organization, Attrition, Predict Employee Attrition, Employee Attrition dataset, Machine Learning Algorithm

1. **INTRODUCTION**

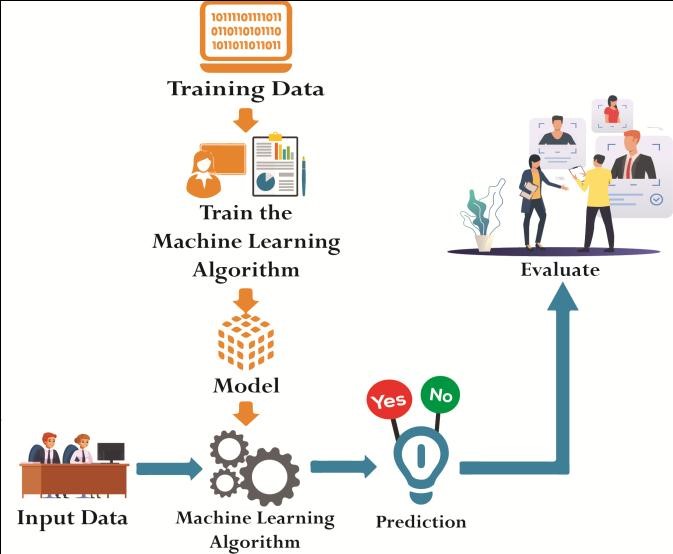
Employee quit is a reduction high in any organization where employees resign [1] . Employees plays a vital role in assets of any organization. It’s important to know whether the employees are dissatisfied or are there any other reasons for leaving the job. Those days are better for an opportunities, employees are eager and to jump from one organization to other. but if they leave the job it might handle huge losses in organization . New hiring will consume money and time, and also the freshly hired employees take time to make the respective organization profitable. Retention of skilled and hardworking employees is one of the most critical challenges faced by many organizations, we developed a system that uses employee data to analyze reasons for employee attrition. This application is applicable for employees who have completed their probation period. If the employee has recently joined the organization, then it is difficult to predict their dissatisfaction factors as they are not a confirmed employee before their allocated probation period. Hence, by improving employee satisfaction and providing a desirable working environment, we can certainly reduce this problem significantly [1].  The management has reduce an HR analytics firm to understand what factors they should focus on, in order to check weakening. After word , they want to know what changes they should make to their workplace, in order to get most of their employees to stay. Also, they want to know which of these variables is most important and needs to be addressed right away When an Employee leaves an organization, the reasons are determined by a variety of factors, some of the reason of leaving the organization could be better-paying job outside, a bad relationship with boss, pursuing higher studies, relocating due to family reasons, fired from organization, job Dissatisfaction, salary not as per expectation, poor relationship with team members, poor working environment, lack of opportunity for career development, overtime, workload etc. In order to tackle this issue.

This system is able to predict which employee may leave an organization with what reason, so that they can take several corrective actions in order to ensure that employees stay in the organization and can reduce the attrition. Some of the employee are strategies to control corrosion are motivating employees, expose employees to newer roles, taking constant feedback from employees, etc. We applied different machine learning algorithms such as SVM (Support Vector Machine), KNN (K-Nearest Neighbor), Decision Tree and LGBM classifier. Graphical representation is also provided for better understanding of insights.

* 1. **Literature Review**

Employee weakening is mostly the normal way of people out of an organization, due to career job change, relocation, illness and so on [2]. Employee weakening is the percentage of employees for what so reasons employee leaving the organization. Employees can leave the organization for many personal as well as professional reasons. So basically there are two types of turnover, one is voluntary turnover which is decided by the employee, and the other type of turnover is decided by the company and that is why it is called involuntary turnover [6]. Involuntary turnover generally happens when performance of the employee is not up to the expectations. Retention is also necessary for the growth and stability of an organization [6]. The high attrition rate causes when there are more employment opportunities in the market. Currently the employee weakening is one of the major issue faced by HR managers. There are so many working employees who are not satisfied due to one of the aspect which is not fulfilled by the organization which results in higher weakening rate.

1. **DESIGN AND ARCHITECTURE**



The proposed system consists of different machine learning algorithms. To build model, we take employee dataset which includes all past and present records of the employees, then we perform data preprocessing .We have divided dataset into two parts one is train data and second one is test data. Most of the data is used for training and smaller portion of data is used for testing. The aim of training is to make a prediction correctly as often as possible. The test data is used to see how well the machine can predict new answers and to validate machine learning model behavior.

Afterward using different machine learning algorithm we have build the model Furthermore, user can choose algorithm according to their choice and check the result.. After evaluating result the reason behind the weakening is also given by the system.

**3.TECHNOLOGIES USED IN THE PROPOSED SYSTEM**

**3.1 Machine Learning**

Machine Learning is vast important technology about data analysis for quality prediction and evaluation. There are no of types of algorithms in machine learning which are used to predict the relevant class of new or unseen data. In our system we used different machine learning algorithms to find out the reasons for employee weakening.

**3.1.1 K-Nearest Neighbors**

K-Nearest Neighbor is considered a lazy learning algorithm that classifies data sets based on their similarity with neighbors. It is one of the most fundamental and simple classification methods and one of the best choices for a classification study of the data [7]. The classification using KNN involve determining neighboring data points and then deciding the class based on the classes of the neighbors.

**3.1.2 Support Vector Machine**

Support Vector Machine is kind of classification technique. It is a model used for classification and regression problems. It can solve linear and non-linear problems. The idea of SVM is simple: The algorithm creates a line or a hyper plane which separates the data into classes [9]. When unknown data is given as input it predicts which class it belongs to. The margin between the hyper plane and the support vectors are as large as possible to reduce the error in classification.

**3.1.3 Decision Tree**

As the name implies all decision tree techniques recursively separate observations into branches to construct a tree for the purpose of improving the prediction accuracy. Decision tree is a conventional algorithm used for performing classifications based on the decisions made in one stage. This provides tree structured representation of the decision sets [10].

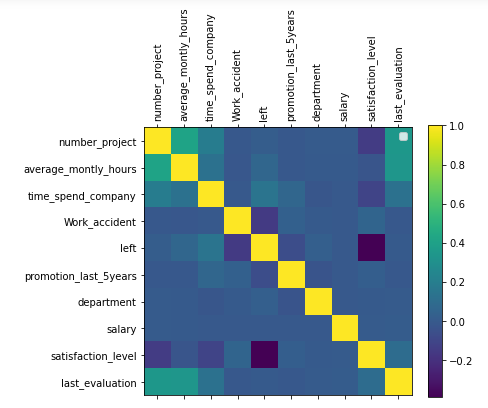
**3.1.4 Light GBM**

Light GBM is a speed, dispense, high speed gradient boosting structure based on decision tree algorithm, used for classify, classification and machine learning tasks. Better performance than any other boosting algorithm: It produces much more combination trees by admirer leaf wise split detain rather than a level-wise approach which is the main factor in realize higher accuracy.

**3.2 Dataset Analysis**

Data collection refers to the collection of relevant data from all available sources to perform analysis. The data used for this employee attrition analysis was obtained from Kaggle Website [11]. This data set contains 1499 records and 11attributes. The categorical values are converted to numeric values in order to make the classification algorithm more effectual.

**3.2.1 Heatmap**

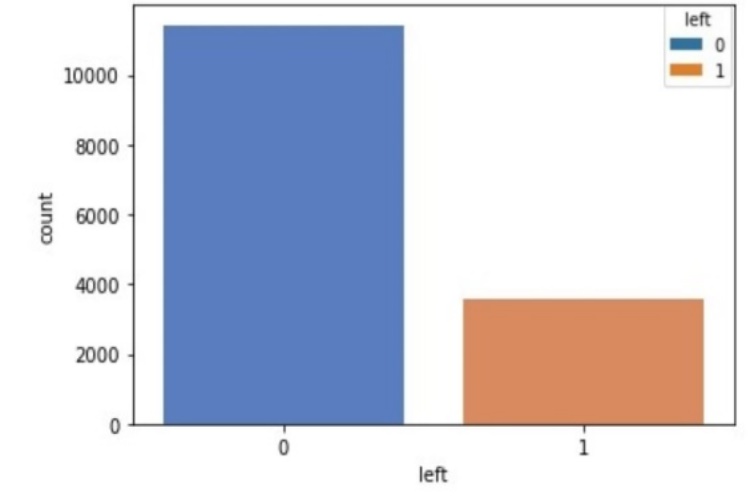
****

**Fig 3.1**

The aboveFig 3.1 represents the heatmap which helps to identify attributes with the strong or weak correlation between them.

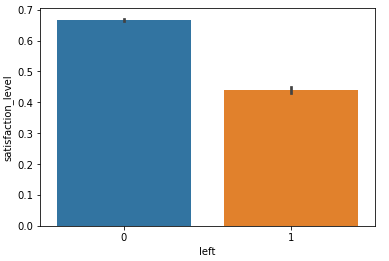
### Some graphs with explanation

Here are several graphs generated by the system with respect To attrition:

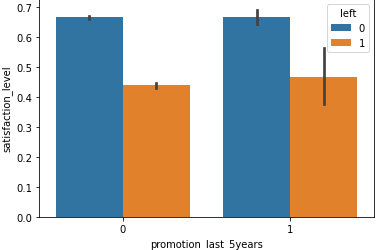


**Fig 3.2**

Above Fig 3.2. represents the bar graph of the employee data. there are total 14999 no of employees. 1 represent total no of employee left and 0 represent the total number of employee stay in the company.



**Fig 3.3**

Above Fig 3.3. represents the bar graph of the satisfaction level vs Left . satisfaction level is the literally main reason that the person left the company .hear in Sea born 0 and 1 with color full bar graph in last past five years whose satisfaction level is more than 60% they did not left the company but whose satisfaction level is below 60% they left the company. ****

**Fig 3.4**

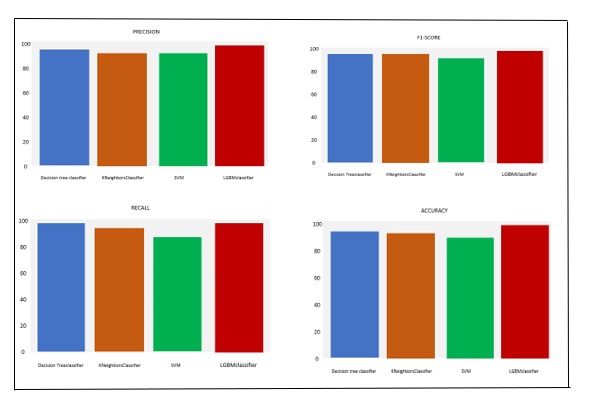
Above Fig 3.4 represents that some employees whose satisfaction level is less than 50% and they did not get promotion then they left the company and who did not got promoted but their satisfaction level was more than 65% they stay in the company in second table some employees who get promoted but their satisfaction level is less than 50% they left the company, satisfaction level is the main reason that the person left the company.

# 4.RESULT AND EVALUATION

In above dataset, there are various attributes like department, number of projects, work accident, satisfaction level, last evaluation etc. Based on these values, model which was build with the help of different machine learning algorithms which will predict whether employees will leave the organization or not. The predicted values are compared with test values to calculate the accuracy of the each algorithm. The table given below describes various factors, so we can easily conclude which algorithm is best for our model. From the table, we can infer that LGBM Classifier gives highest accuracy on the HR Employee Attrition dataset whereas SVM gives the lowest accuracy for the same dataset.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Attributes/Model | KNN | SVM | Decision Tree | LGBM Classifier |
| Accuracy | 98 | 96.53 | 98.06 | 99.13 |
| Precision | 0.94 | 0.94 | 0.95 | 0.98 |
| Recall | 0.97 | 0.91 | 0.96 | 0.97 |
| f1-score | 0.96 | 0.93 | 0.96 | 0.98 |

**Table: shows Results of Different Classifier**



**Fig 4.1- Performs comparison of Different plot**

**5.CONCLUSION**

This paper find out which machine learning algorithm is bring out well in predicting the employees who are likely to leave the respective organization. From the result, we can analysis that LGBM Classifier performs better than the other classifiers. We observed that, the cause of employee attrition is because of both external and internal factors. This study might help organization for knowing the factors of employee attrition and can take appropriate steps to minimize the weakening rate.

**REFERENCES**

[1] Dilip Singh Sisodia, Somdutta Vishwakarma, Abinash Pujahari, “Evaluation of machine learning models for employee churn prediction,” International Conference on Inventive Computing and Informatics (ICICI 2017).

[2] M.Sudheer Kumar, Obulesu Varikunta, K.Ramakrishna, “Employee Attrition and Retention Strategies in Manufacturing: An Empirical Study in Amara Raja Batteries Limited,”International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8, Issue-7, May, 2019, pp. 2962-2968.

[3] Mrs. Jaya Sharma, “Employee Attrition and Retention in a Cut-Throat Competitive Environment in India: A Holistic Approach,” Paripex - Indian Journal of Research (PIJR), Volume 4, Issue 2, Feb 2015.

[4] Dr. B. Latha Lavanya, “A Study on Employee Attrition: Inevitable yet Manageable,”International Journal of Business and Management Invention, Volume 6, Issue 9, September. 2017, pp. 38-50.

[5] Heng Zhang , Lexi Xu , Xinzhou Cheng , Kun Chao , Xueqing Zhao, “Analysis and Prediction of Employee Turnover Characteristics based on Machine Learning,” The 18th International Symposium on Communications and Information Technologies (ISCIT 2018).

[6] Diwakar Singh, “A Literature Review on Employee Retention with Focus on Recent Trends,” International Journal of Scientific Research in Science and Technology (IJSRST 2019), Volume 6, Issue 1, pp. 425-431. [7] L. E. Peterson (2009), “K-nearest neighbor,” Scholarpedia, vol. 4, no. 2, p. 1883. [Online]. Available: <http://www.scholarpedia.org/article/Knearest_neighbour>.

[8] Random Forest Algorithm. [Online]. Available: https://www.javatpoint.com/machine-learningrandom-forest-algorithm

[9] Support Vector Machines (SVM). [Online]. Available: <https://towardsdatascience.com/httpsmedium-com-pupalerushikesh-svm-f4b42800e989>

[10] S. Saranya, J. Sharmila Devi, “Predicting Employee Attrition Using Machine Learning Algorithms and Analyzing Reasons for Attrition,” International Journal of Advanced Engineering Research and Technology (IJAERT), Volume 6, Issue 9, September 2018, pp. 475-478. [11] Kaggle, “HR-Employee-Attrition.” [Online]. Available: https://www.kaggle.com/pavansubhasht/ibm-hranalytics-attrition-dataset