IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, May 2023

Analysis of Employee Attrition using Bug Tracking System

Devesh Garg[1], Ayushi Singh[2], Simran Chauhan[3], Navneet Kumar[4], Swapnil Tiwari[5]
Assistant Professor, Department of Computer Science and Engineering[1]
Students, Department of Computer Science and Engineering[2,3,4,5]
Raj Kumar Goel Institute of Technology, Ghaziabad, India

Abstract: *Employee turnover, or the rate at which employees leave the organization,[6] can have a significant impact on a company's productivity, culture, and profitability. Understanding the factors that contribute to attrition is crucial for organizations to develop effective employee retention and talent management strategies. Employee turnover analysis has traditionally been based on HR data such as surveys, exit interviews[1] and performance reviews. However, this study presents a new method that can provide better insights into employee engagement and job satisfaction by integrating data from an error. By combining this data with employee data and related HR metrics, relationships and patterns can be identified to reveal key factors contributing to the impact.[3] Additionally, error systems can provide insight into collaboration and communication within a group. Through this analysis, organizations can gain a deeper understanding of the factors that influence bullying and develop strategies to improve employee engagement, job satisfaction and performance.*

Keywords: Bug Tracking System

I. INTRODUCTION

Attrition, that is, the voluntary or involuntary departure of employees from an organization causes serious problems for companies in terms of productivity, loss of knowledge, and the cost of finding a job and acquiring new skills. Understanding the root causes of attrition is critical for organizations to develop effective strategies.[7] The routine process of analyzing customer churn relies on HR data such as surveys, exit interviews and performance reviews. However, there are opportunities to explore other data that can provide better insight into the phenomenon of attrition.

A potentially useful source of untapped information is the bug tracker, a software tool widely used in the software industry to manage and track problems or bugs. This information includes transmission of errors, resolution times, and collaboration patterns between partners.

Analyzing the frequency and severity of mistakes made by employees can provide insight into the complexity and satisfaction of their job. It unresolved issues, or recurring issues that are causing employee dissatisfaction and eventual injury. Additionally, coordination information from error systems can reveal team dynamics, communication patterns, and relationship issues that can affect competitive costs.

II METHODOLOGY

1) K-Nearest Neighbor Algorithm

The K-Nearest Neighbors (KNN) algorithm is a supervised algorithm for classification and regression operations. It is a simple and intuitive algorithm that makes predictions based on the similarity between input data and recorded data in the training process.

In the KNN algorithm, "K" represents the number of nearest neighbors to be considered when making predictions. Given the new input data, the algorithm identifies the K nearest neighbor from the training data based on distance measures such as Euclidean distance.[4] The predicted class or value for the new data is then determined by the majority vote (for classification) or average (for regression) of the text or neighboring K's results.

For the classification task, class labels are assigned to data points based on the K most common neighbor class. For the regression task, the output value is approximate towers from the mean or weighted average of the target difference between neighbors K.[5]

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 8, May 2023

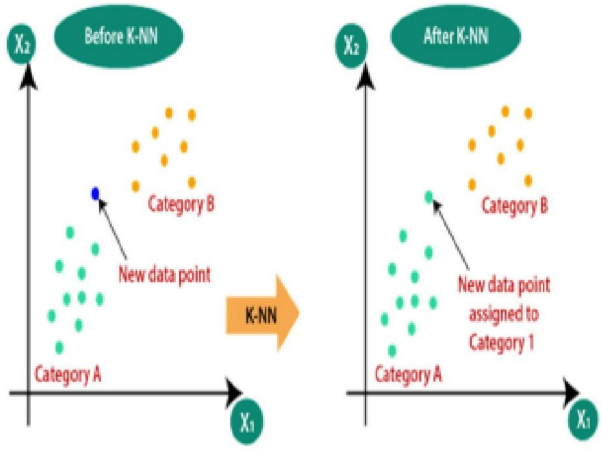

Figure 1 : KNN Algorithm [7]

2) Logistic Regression

Logistic regression is a tracking algorithm for binary classification where the purpose is to estimate the probability that an input belongs to one of two classes. Despite its name, logistic regression is a classification algorithm, not a regression algorithm.

The logistic regression algorithm uses a logistic function, also known as the sigmoid function, to model the relationship between input variables (features) and a binary target variable.[1] The sigmoid function maps each real value to a value between 0 and 1 that represents the probability that the input will fall into the positive class.

After training, the logistic regression model is represented by a set of learning coefficients (weights) and interaction time for each feature.

Given the new input data, calculate the weight of the features using the learning coefficients and use the logistic (sigmoid) function to estimate the probability of joining a good class. Generally, an input is classified into one of two classes using a probability of 0.5. Test the performance of a logistic regression model by testing its predictions on separate datasets.

Common criteria for binary classification include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic (ROC) curve.Experiment with different hyperparameters, such as increasing the power constant (for example, L1 or L2 constant) to improve the performance of your logistic regression model. [3]Regularization helps prevent overfitting and improves generalization.

Logistic regression provides interpretable results because the coefficients associated with each feature indicate the direction and strength of their effects on the predicted outcome.A positive coefficient indicates a positive correlation with class, and a negative coefficient indicates a negative correlation.

Logistic regression can be used to identify employees to predict whether an employee will leave or remain with an organization based on their characteristics or characteristics.

By training a logistic regression model on historical data on employees who have left or not left,[6] the model can learn how these characteristics affect those involved.

By using logistic regression to analyze employee turnover, organizations can understand the factors that cause turnover and take critical steps to improve employee retention. Interpretation of the model allows stakeholders to understand the root causes of abuse and make informed decisions based on predicting the consequences.

3) Decision Tree

The decision tree algorithm is a maintenance algorithm for two classification tasks. It creates a flowchart model called a decision tree, where each node represents a feature or behavior, each branch represents a decision based on that feature, and each leaf represents a list (in distribution) or predictive values (in regression).

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, May 2023

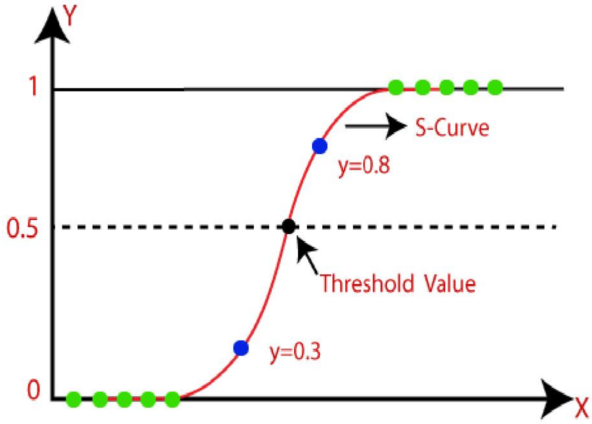

Figure 2: Logistic Regression Algorithm [7]

A decision tree algorithm iteratively classifies data based on selected features. At each step, it selects features and corresponding points in each classification that maximize class separation (in classification) or minimize impurity or variance (in regression).[2] The algorithm decides when to stop distributing more data, whether the maximum number of nodes in the tree has been reached, the minimum number of nodes in the leaf has been reached, or that further allocation does not improve the performance of trees. . model. The decision tree algorithm works on categorical features by creating branches for each value and chooses discrete points as the starting point for the numeric features.

The performance of the decision model is measured by appropriate criteria such as accuracy, precision, recall, F1 score, or mean squared error, depending on the task. To prevent spoilage, pruning techniques that simplify logging by removing redundant branches or combining similar leaves can be used. Editing methods such as limiting the maximum depth or increasing the penalty can also be used.

Decision trees can be used in employee analysis to understand what causes employees to leave the organization. By building decision-making models based on historical data, organizations can understand the key factors affecting employee turnover and make informed decisions to improve the management of these employees.

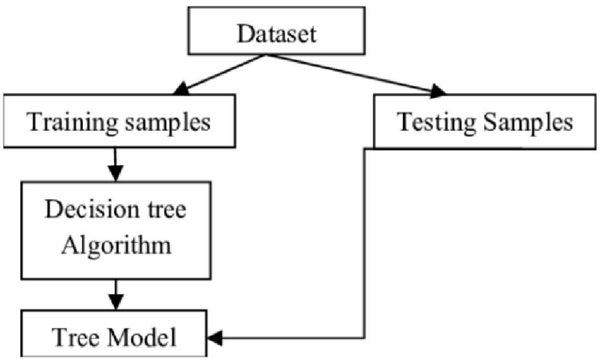

Figure 3: Decision Tree Algorithm [8]

III LITERATURE

[2] Wie-Chiang H., Ruey-Ming C. (2007), Legal workforce forecasting models are essential for early detection of unplanned departures and therefore give managers ample time to address management issues related to turnover. Logit and Probit models have been successfully used to solve inequality and regression problems. However, the possibility of using it for the prediction of voluntary change has not yet been explored. Therefore, a sample of voluntary turnover data from a central Taiwanese automobile company was used and the results show that the proposed model has a high predictive power.[7] Thus, these two models also offer alternative methods for estimating employee turnover in human resource management.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, May 2023

[3] V. Rama Devi, “Employee engagement is a two way street": The purpose of this article is to highlight the importance of building an insurance culture and to discuss the driving force behind employee engagement. Design/methodology/methodology This article presents the author's personal views based on research findings and effectiveness. Findings This article supports the view that talent acquisition and retention is the foundation of an organization's ability to be profitable, control costs, grow through acquisitions, innovate, create new products and services, and find new markets.

A company culture built on teamwork, quality work, employee care, growth opportunities, flexible working, and positive leadership and management support staff. Benefits This article argues that in today's competition, companies must recognize the importance of leaders in employee retention.Authenticity/Value This article shows why employee engagement depends on the two-way relationship between employer and employee.

[5] Jonathan P. Doh, Richard R. Smith, Stephen A.Stumpf, Walter G Tymon Jr.," Pride and professionals: The effectiveness of HR programs designed to prevent the loss of new professionals has rarely been rigorously analyzed for their success and cultural use. To answer these questions, we conducted a study to examine professionals' perceptions of the effectiveness of knowledge management in their organizations in India.

Design/Method/Methodology A sample of 9,301 (4,811 participants) from 28 companies in 32 workplaces across India were able to participate in this study. Of the respondents, 2,723 were new professionals. The actual change data of the participants were obtained one year after the first survey.The results showed that there is a positive relationship between performance management, professional development, leader support and leadership behaviors and organizational satisfaction.

IV PROPOSED WORK

As we demonstrate in this report, we can use data to explain situations using machine learning.

Gather employee information, including demographics, employment history, performance reviews, job satisfaction surveys, and other information that may indicate potential breach damage. This information may be obtained from human data, research, or other sources.

Clean up and pre-populate data by processing missing values, negatives, and inconsistencies. Perform appropriate data transformations, such as converting categorical variables to numerical representations. Perform data analysis to understand data and identify patterns or relationships between different components. Visualize data with graphs, histograms, and summary statistics to understand distributions and relationships in the data.

Identify the most important factors that will affect employee turnover. Use techniques such as correlation analysis, importance ranking, or domain recognition to select a set of features with predictive power. Some commonly used algorithms for this task include logistic regression, decision trees, KNN. Selection should be based on the nature of the data, the complexity of the problem, and the need for model interpretation or accuracy. Demonstrate the selected model of the training process and evaluate its performance in the benchmark using appropriate metrics such as accuracy, precision, recall, F1 score or area under the ROC curve (AUC-ROC). Optimize the model by tuning hyperparameters or using methods such as cross validation to deliver potential.

V RESULT DISCUSSION

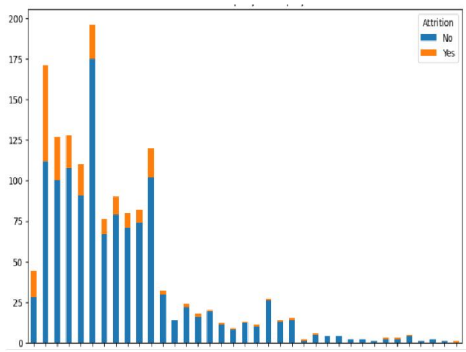

Figure 4: Employees working in organization

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, May 2023

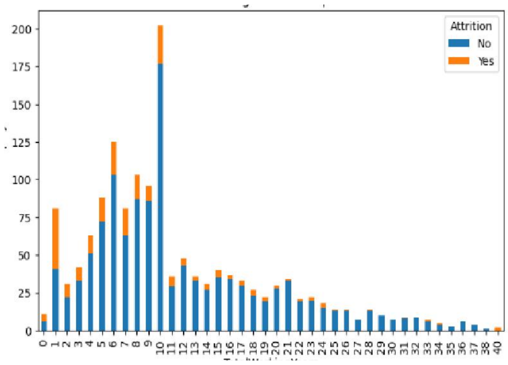

Figure 5: Employees leaving the organization

A. Experimental Setup

The system was executed on a Windows 10 (64-bit) PC with an Intel Core i5-6200U processor running at a clock speed of 2.30 GHz and 8GB of RAM.

Dataset : HR Employee Attrition from Kaggle.

B. Comparison Result

In this section the three machine learning algorithms are compared which are logistic regression, KNN algorithm and Decision tree. They are based on the test dataset of the employee and the train data of the employee working in the organization. The accuracy of these algorithm is compared and results are reported.

Test	Accuracy	Train Accuracy
Logistic Regression	87.228261	83.212341
KNN	85.597826	100.000000
Decision Tree	81.250000	100.000000

Table 1: Comparison of tet and train accuracy

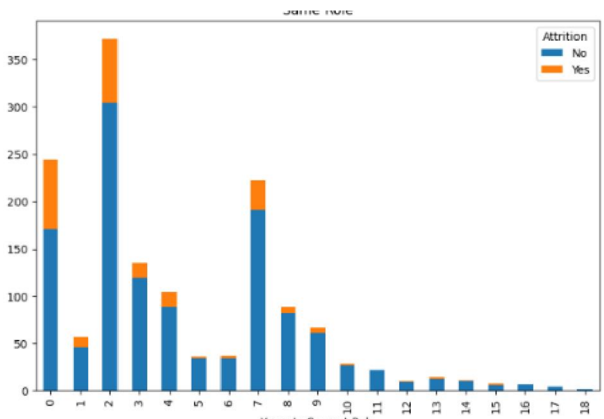

Figure 7 : Employees without promotion

DOI: 10.48175/IJARSCT-10298

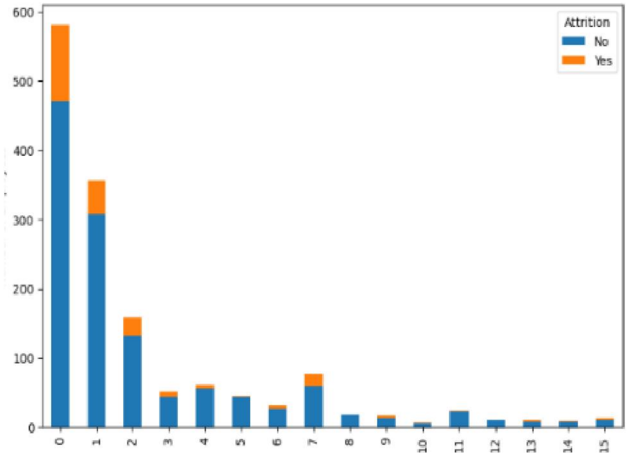

Figure 8 : Employees with promotion

VI CONCLUSION

The analysis of employee turnover plays an important role in understanding and addressing the factors that cause employees to leave an organization. Using a data-driven approach and machine learning, organizations can gain better insights into design trends and drivers, allowing them to take critical steps to improve employee retention and engagement. Through the analysis of employee turnover, organizations can identify important factors that affect earnings, such as job satisfaction, pay, life balance, career development, and leadership. By understanding these factors, organizations can implement strategies and interventions to address root causes of abuse.

Machine learning algorithms, such as logistic regression, decision trees, or random forests, can be employed to build predictive models for employee attrition. These models can effectively identify employees who are at a higher risk of attrition, allowing organizations to take preemptive actions to retain them. By utilizing such models, organizations can make data-driven decisions and allocate resources to retain valuable employees.

Furthermore, the insights gained from employee attrition analysis can inform the development of effective retention programs and policies. Organizations can tailor their initiatives based on the identified factors influencing attrition, creating personalized career development plans, improving work-life balance, enhancing compensation and benefits packages, and fostering a positive organizational culture.

REFERENCES

[1] Kotsiantis, S.B. (2007) Supervised Machine Learning: A Review of Classification Techniques. Informatica 31,249-268.

[2] Wie-Chiang H., Ruey-Ming C. (2007); A Comparative Test of Two Employee Turnover Prediction Models. International Journal of Management; June 2007; Vol.24 No.2; pp. 216 – 229.

[3] V. Rama Devi, “Employee engagement is a two way street": Human Resource management International digestVol. 17 NO. 2 2009, pp. 3-4

[4] Piyali Ghosh and Rachita Satyawadi," Who stays with you? Factors predicting employees’ intention to stay": International Journal of Organizational Analysis Vol. 21 No. 3, 2013 pp. 288-312.

[5] Jonathan P. Doh, Richard R. Smith, Stephen A.Stumpf, Walter G Tymon Jr.," Pride and professionals: retaining talent in emerging economies." Journal of business strategy, VOL. 32 NO. 5 2011, pp. 35-42.

[6] Kumar, N., Chauhan, R., & Singh, R. (2020). Employee attrition prediction using machine learning. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.

[7]https://www.javatpoint.com/logistic-regression-in-machine-learning.

[8]https://www.researchgate.net/figure/Workflow-of-Decision-Tree-New-instances-are-classified-from-the-root-node-and-sorted_fig1_315917143