IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

ISSN
INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

INNO SPACE
SJIF Scientific Journal Impact Factor

doi
crossref

निस्केयर
NISCAIR

🌐 www.ijarety.in   ✉ editor.ijarety@gmail.com

Prediction of Employee Turnover with Machine Learning

Mr. Mavunuri Ajay Kumar[1], Mr. Pothuganti Abhinay[2], Ms. Potuganti Sandhya[3], Ms. Todeti Sindhu[4]

Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, India[1]

Student, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, India[2,3,4]

ABSTRACT: The determination of an employee to leave an organization is a significant factor impacting the performance of the organization, as it affects overall profitability. To gain a competitive advantage, impacting an employee's intent to leave, management can intervene with strategic policies and decisions to reduce the intent to leave and increase employee engagement. This project uses machine learning algorithms to predict an employee's intention to leave the organization in the near future and identifies the significant features impacting the employee's intention to leave. Data was collected from 416 employees working in IT and ITES companies using convenience sampling and structured questionnaires. Additionally, text mining was used to analyze open-ended questionnaire responses, mining frequently used words and employee sentiments. The study found that among the classification algorithms used for predicting employee turnover intention, XGBoost

I. INRODUCTION

Today turnover is one of the major problems faced by industry across the world. It is the most burning issue for the industry, and high turnover rates lead to many issues in the boundary of the organization like losing the talents and knowledge, cost related to training and administration, and recruitment. It is observed that many attributes lead to the turnover of an employee. Which includes working environment, job satisfaction, employer's behavior, job timing, and most important is salary or incentives. Also, the prediction model plays an essential role in finding the behavior of employees.

Timely delivery of any service or product is the primary goal of any organization in recent days due to high competition in industries. If a talented employee leaves unexpectedly, the company is not able to complete the task at defined times. It may become the reason for the loss of that company. Therefore, companies are interested in knowing the employee's attrition. They can make a proper substitute or arrangements earlier.

II. LITERATURE SURVEY

*Anjali Avinash Kalse(2024)*Employee turnover poses a significant challenge for organizations, impacting productivity, morale, and overall business performance. The application of machine learning (ML) techniques for predictive analysis of employee turnover and the development of effective retention strategies. Utilizing historical employee data, including demographic information, job roles, performance metrics, and engagement surveys, a predictive model is constructed using advanced ML algorithms. 1. Introduction Employee turnover, a pervasive issue in contemporary workplaces, presents multifaceted challenges for organizations. The repercussions extend beyond mere human resource fluctuations, as turnover entails substantial financial costs, loss of institutional knowledge, and a palpable dip in team morale. The intricate dynamics of employee departure necessitate innovative solutions, and one promising avenue is the integration of machine learning into workforce management. This advanced technology offers a data-driven approach to predict turnover patterns, enabling organizations to preemptively address issues and implement tailored retention strategies. By analyzing historical data, machine learning algorithms can discern patterns and correlations that might elude traditional methods. Factors such as job satisfaction, career development opportunities, and interpersonal relationships can be quantitatively assessed to generate predictive models. These models, in turn, empower organizations to identify at-risk employees and proactively implement measures to mitigate turnover risks.

*Al-Radaideh et. al(2022)*Human capital is of a high concern for companies' management where their most interest is in hiring the highly qualified personnel which are expected to perform highly as well. Recently, there has been a growing interest in the data mining area, where the objective is the discovery of knowledge that is correct and of high benefit for users. In this paper, data mining techniques were utilized to build a classification model to predict the performance of

employees. To build the classification model the CRISP-DM data mining methodology was adopted. Decision tree was the main data mining tool used to build the classification model, where several classification rules were generated. To validate the generated model, several experiments were conducted using real data collected from several companies. The model is intended to be used for predicting new applicants' performance.

Alao(2020) Employee turnover is a serious concern in knowledge-based organizations. When employees leave an organization, they carry with them invaluable tacit knowledge which is often the source of competitive advantage for the business. In order for an organization to continually have a higher competitive advantage over its competition, it should make it a duty to minimize employee attrition. This study identifies employee related attributes that contribute to the prediction of employees' turnover in organizations. Three hundred and nine (309) complete records of employees of one of the Higher Institutions in Nigeria who worked in and left the institution between 1978 and 2006 were used for the study. The demographic and job-related records of the employee were the main data which were used to classify the employee into some predefined turnover classes. Waikato Environment for Knowledge Analysis (WEKA) and See5 for Windows were used to generate decision tree models and rule-sets. The results of the decision tree model and rule-sets generated were then used for developing a predictive model that was used to predict new cases of employee attrition. A framework for a software tool that can implement the rules generated in this study was also proposed. Keywords: Employee Attrition, Decision Tree Analysis, Data Mining

Alduayj (2021) Data is the new oil. More and more data is being captured and stored across industries and this is changing society and, therefore, how businesses work. Traditionally, BI tried to give an answer to the general question: what has happened in my business? Today, companies are involved in a digital transformation that enables the next generation of BI: Advanced Analytics (AA). With the right technologies and a data science team, businesses are trying to give an answer to a new game changer question: what will happen in my business?

We are already listening how AA is helping to increase profits in many companies. However, some businesses are late in the adoption of AA, while others are trying to adopt AA but are just failing for various reasons. Clear Peaks is already helping many businesses to adopt AA, and in this blog article we will review, as an illustrative example, an AA use case involving Machine Learning (ML) techniques to help HR departments to retain talent.

Employee turnover refers to the percentage of workers who leave an organization and are replaced by new employees. A high rate of turnover in an organization leads to increased recruitment, hiring and training costs. Not only it is costly, but qualified and competent replacements are hard to find. In most industries, the top 20% of people produce about 50% of the output

III. METHODOLOGY

The study of this project is to develop a machine learning model to predict employee turnover, leveraging a combination of data preprocessing, feature engineering, and advanced machine learning techniques. This aims to improve prediction accuracy and provide actionable insights for human resources (HR) departments to proactively address employee retention issues. By leveraging advanced feature extraction, machine learning models, and real-time prediction through a Flask-based web application, this system will provide HR departments with valuable insights, enabling them to take proactive measures to retain employees and reduce turnover rates.

DISADVANTAGES OF EXISTING SYSTEM:
* Logistic regression fails to predict a continuous outcome.
* Logistic regression may not be accurate if the sample size is too small.
* Logistic regression has a major limitation in the assumption of linearity between the dependent variable and the independent variables.
* Less inclined to over-fitting but it can over fit in high dimensional datasets.

PROPOSED SYSTEM
From the comparison of the classification algorithms, the accuracy of the XG boost is 0.94 which shows that the model correctly predicts 94 percentage of the dataset. Precision value of XG boost algorithm is 92% which shows the percentage of correct positive predictions to the total predicted positives. Recall value of XG boost algorithm is 97% which shows the percentage of correct positive predictions to the total positives and f1 score is the harmonic mean of precision and recall shows the accuracy of predicting the model with low false positive and the false-negative rate

which is to be 0.94. The area under the curve for the XG boost algorithm is 0.97 which shows that 97% accurate of performance of the model that distinguishes between positive and negative classes. Thus XG Boost gives better results compared to other models.. The precision value is about 89% and recall value of about 86% for both the algorithms.

ADVANTAGES OF PROPOSED SYSTEM:

- XG Boost can achieve superior prediction accuracy compared to individual models.
- XG Boost is designed for efficiency, especially with large datasets.
- XG Boost techniques to prevent over fitting. Over fitting occurs when a model becomes too specific to the training data and performs poorly on unseen data.

SYSTEM ARCHITECTURE:
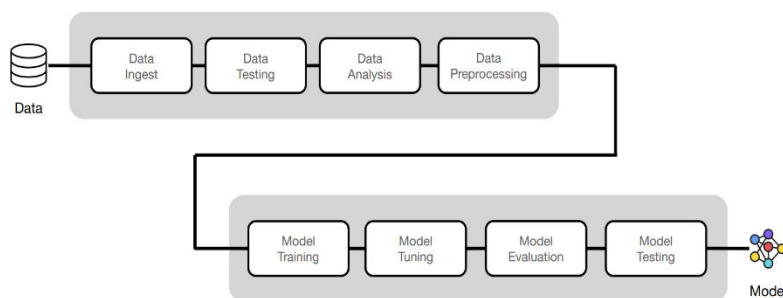
MODULES:

1. **Data set** : A data set is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question.
2. **Pre-Processing** : Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.
When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way.
3. **Splitting :** Data splitting is the act of partitioning available data into. two portions, usually for cross-validatory purposes. one portion of the data is used to develop a predictive model. and the other to evaluate the model's performance.
   - **Training Data:** Used for train the model or given as input to the to the learning model
   - **Testing Data**: Used for test the model or given as input to the model for prediction.
4. **Apply Algorithm :** In this we are using support vector machine algorithm to predict accuracy. It is a non-probabilistic supervised machine learning approaches used for classification and regression. It assigns a new data member to one of two possible classes. It defines a hyperplane that separates n-dimensional data into two classes
5. **Accuracy :**Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.
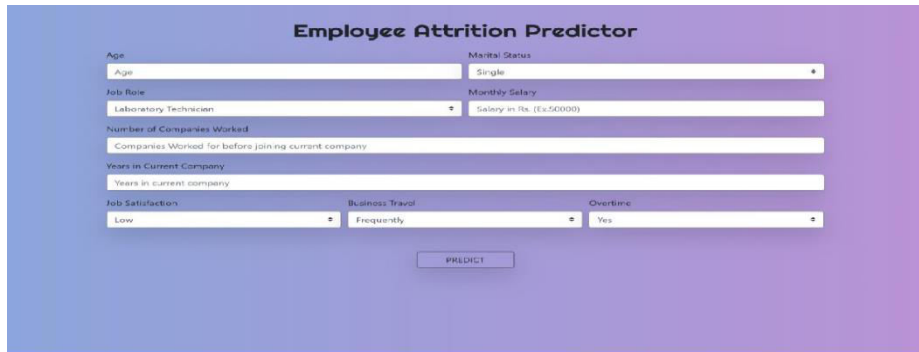
IV. IMPLEMENTATION

XG Boost Algorithm:
XG Boost is a powerful machine learning algorithm that falls under the category of ensemble methods, specifically gradient boosting. It's known for its exceptional performance in various tasks. XG Boost achieves its effectiveness by combining the strengths of multiple weaker models, typically decision trees. These individual trees are sequentially built, where each one focuses on correcting the errors of the previous ones. XG Boost is a versatile and powerful machine learning algorithm that has become a go-to choice for various prediction tasks. Its ability to handle large datasets, achieve high accuracy, and effectively address common challenges like missing values make it a valuable tool for many data science applications.
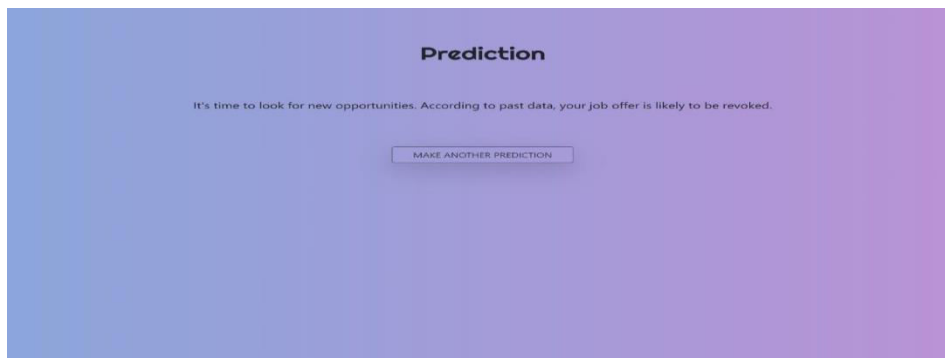
V. EXPERIMENTAL RESULTS

HOME PAGE:

EXPLANATION: The home page is the main landing page of a website, serving as the gateway for users to explore its content and features. It typically includes a welcoming message and a clean, visually appealing design that reflects the brand identity.

TEST RESULT PAGE:

EXPLANATION: This page indicates the prediction of status of an employee in the organization. Based on the featured data it gives whether the employee is in safe condition or not.

VI. CONCLUSION

Employee turnover has been identified as a significant problem for any organization. High performance and talented employee turnover are considered as a loss for that organization. Finding a substitute for that employee is a time-consuming task. In this work, the performance of various machine learning approaches is evaluated on the HR dataset. Here five different approaches are compared. Based on the accuracy measurement, logistic regression well performed for this dataset. It has higher precision, recall, and accuracy. The result of the turnover prediction will be helpful for an organization to reduce the turnover rate of their company.

VII. FUTURE ENHANCEMENT

In our future work, it has higher precision, recall, and accuracy. The result of the turnover prediction will be helpful for an organization to reduce the Turnover rate of their company. Enhancing the prediction of employee turnover in the future requires a combination of advanced data analytics, machine learning techniques, and integration of external and internal organizational factors.

Include external factors such as market trends, industry benchmarks, economic conditions, and competitive data for a comprehensive analysis. Gather qualitative data through surveys, performance reviews, and exit interviews to gain insights into employee sentiment. Implement systems to collect real-time employee engagement data through tools like

pulse surveys or productivity monitoring systems. Implement systems to collect real-time employee engagement data through tools like pulse surveys or productivity monitoring systems

REFERENCES

1. Ajit, P. (2016). Prediction of employee turnover in organizations using machine learning algorithms.Algorithms, 4(5), C5.
2. Al-Radaideh, Q. A., & Al Nagi, E. (2012). Using data mining techniques to build a classification model for predicting employees performance. International Journal of Advanced Computer Science and Applications, 3(2).
3. Alao, D., & Adeyemo, A. B. (2013). Analyzing employee turnover using decision tree algorithms. Computing, Information Systems, Development Informatics and Allied Research Journal, 4.
4. Alduayj, S. S., & Rajpoot, K. (2018). Predicting Employee Turnover using Machine Learning. 2018International Conference on Innovations in Information Technology (IIT), 93–98. 10.1109/INNOVATIONS.2018.8605976
5. Cortes, C., &Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. doi:10.1007/BF00994018.
6. Ho, T. K. (1995). Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition, 1, 278–282.
7. Jantan, H., Hamdan, A. R., & Othman, Z. A. (2010). Human talent prediction in HRM using C4. 5 classification algorithm. International Journal on Computer Science and Engineering, 2(8), 2526–2534.
8. Kaur, S., & Vijay, M. R. (2016). Job satisfaction-A major factor behind turnover of retention in retail McKinley Stacker, I. V. (2015). IBM was ton analytics. Sample data: HR employee turnover and performance [Data file]. McKinley Stacker.
9. Moncarz, E., Zhao, J., & Kay, C. (2009). An exploratory study of US lodging properties' organizational practices on employee turnover and retention. International Journal of Contemporary Hospitality Management, 21(4), 437-458.doi:10.1108/09596110910955695
10. Morgan, J. N., &Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. Journal of the American Statistical Association, 58(302), 415–434. doi:10.1080/01621459.1963.10500855
11. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., &... . (2011). Scikit learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825–2830.
12. Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Academic Press.
13. Rane, A., & Kumar, A. (2018, July). Sentiment Classification System of Twitter Data for {US} Airline Service Analysis. In 2018 {IEEE} 42nd Annual Computer Software and Applications Conference({COMPSAC}). IEEE. 10.1109/compsac.2018.00114
14. Raschka, S. (2015). Python machine learning. Packet Publishing Ltd.
15. Saradhi, V. V., &Palshikar, G. K. (2011). Employee churn prediction. Expert Systems with Applications,38(3), 1999–2006. doi:10.1016/j.eswa.2010.07.134
16. Sexton, R. S., McMurtrey, S., Michalopoulos, J. O., & Smith, A. M. (2005). Employee turnover: A neural network solution. Computers & Operations Research, 32(10), 2635– 2651. doi:10.1016/j.cor.2004.06.022
17. Sikaroudi, E., Mohammad, A., Ghousi, R., &Sikaroudi, A. (2015). A data mining approach to employee turnover prediction (case study: Arak automotive parts manufacturing). Journal of Industrial and Systems Engineering, 8(4), 106–121.
18. Tzeng, H.-M., Hsieh, J.-G., & Lin, Y.-L. (2004). Predicting Nurses' Intention to Quit with a Support
19. Vector Machine: A New Approach to Set up an Early Warning Mechanism in Human Resource Management. CIN: Computers, Informatics, Nursing, 22(4), 232–242. PMID:15494654
20. Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., & Zhu, X. (2018). Employee turnover prediction with machine learning: A reliable approach. Proceedings of SAI Intelligent Systems Conference, 737–758.

IJARETY

International Journal of Advanced Research in Education and Technology

🌐 www.ijarety.in   ✉ editor.ijarety@gmail.com