



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 5, May 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

An Exploratory Study of Organizational Practices on Employee Attrition Rate in the field of IT

Mustaquim Areeb, Malik Sami, Muskan Jain, Vibhute Neil, S. P. Bholane(guide)

Dept. of Computer Engineering, Sinhgad College of Engineering, Pune, India

ABSTRACT: Every organisation has qualities such as productivity and strength that stand on the shoulders of its personnel. In today's competitive environment, keeping regular employees is a major difficulty for all organisations. Employee attrition is one of the most serious business issues in HR analytics. Companies invest much in staff training with the goal of maximising the benefits to the company in the future. When an employee departs the firm, the company suffers a loss of opportunity. These studies evaluate the employee attrition rate by associated factors such as Job Role, Overtime, and Job Level.

KEYWORDS: Classifier, Preprocessing, Employment Features, Attrition Rate, HR.

I. INTRODUCTION

Employees are the most precious asset and crucial resource in organisations, according to ch. The attrition rate is now determined by growing competitiveness and enhanced requirements in employees' competency. Employee attrition is seen as a severe concern for businesses. The expense of recruiting and training new personnel is prohibitively expensive. The company must look for, hire, and train new staff. Locking out experienced employees, particularly strong achievers, is difficult to manage and has a detrimental impact on an organization's success and performance. The study focuses on the aspects that may contribute to employee attrition control.

Employees are a company's most precious asset. However, if they left positions unexpectedly, it might cost any organisation a lot of money. Employee attrition is a reduction in personnel in any organisation caused by employees who willingly leave or retire. Employee turnover is the number of current employees who are re-placed by new employees during a given time period. It results in significant human resource spending by contributing to new employee recruiting, training, and development, as well as performance management. Once again, voluntary attrition is inescapable. As a result, we can undoubtedly lessen this problem greatly by improving staff morale and providing a pleasant working atmosphere.

The company's recruiting and termination criteria are used to calculate the rate of attrition. There are several reasons why employees may leave their jobs. The business terminology 'Turnover' and 'Attrition' are always at odds here. There are numerous types of 'turnover' in an organisation. 'Attrition' refers to the process of reducing the number of employees. These terminologies can be used interchangeably to analyse personnel statistics and other metrics required for manpower planning.

Employee turnover has become a major concern in organisations due to its negative effects on workplace self-esteem and efficiency. Organisations address this issue by anticipating the likelihood of staff turnover using machine learning techniques, allowing organisations to adopt proactive retention measures.

II. RELATED WORK

The updated methods utilising different data mining techniques are gathered to analyse the employee attrition rate at various stages. The following study on data mining for extracting the employee attrition rate employed in various models and a thorough assessment of various researchers' efforts are mentioned;

Data mining techniques were used by Qasem A, A. Radaideh, and EmanANagi et al. to create a classification model to forecast employee performance [5]. In their work, they used the CRISP-DM data mining methodology [6]. The primary



data mining tool utilised to construct the classification model, where a number of classification rules were constructed, was the decision tree. They validated the generated model; several experiments were conducted using real data collected from several companies. The model is used for predicting new employees' performance.

Ali EsmiaeeliSikaroudi, RouzbehGhousi, and Amir Mohammad EsmiaeeliSikaroudi, among others, used knowledge discovery techniques to actual manufacturing plant data. They chew over many characteristics of employees such as age, technical ability and job experience. They used to find out importance of data features is measured by Pearson Chi-Square test.

In order to help human resource professionals refocus on human capability criteria and improve the performance appraisal process of its human capital, John M. Kirimi, Christopher Moturi, et al. [8] created a prediction model for employee performance forecasting.

Extreme Gradient Boosting (XGBoost) approach, which is more reliable due to its regular-ization formulation, was investigated [9] by RohitPunnoose, Pankaj Ajit et al. [10] XGBoost is compared to six previously employed supervised classifiers using data from the HRIS of a major retailer to show that it is significantly more accurate in predicting employee turnover.Support Vector Machine (SVM)

An SVM is a supervised learning algorithm that implements the principles of sta- tistical learning theory [1] and can solve linear as well as nonlinear binary classification problems. [2] To achieve class separation, a support vector machine creates a hyper-plane or collection of hyper-planes in higher dimensional space. The assumption behind this is that the hyper-plane with the greatest separation from the nearest training data points of any class achieves a good separation; the higher the margin, the smaller the classifier's generalisation error.

A Discriminant analysis involves creating one or more discriminant functions so as to maximize the variance between the categories relative to the variance with the cate- gories [14]. Linear Discriminant Analysis is explained as deriving a variant or z-score, which is a linear combination of two or more independent variables that will discrimi- nate best between two (or more) different categories or groups. The z-scores calculated using the discriminant functions is then used to estimate the probabilities that a par- ticular member or observation belongs to a class. An important point to note with LDA is that the features used should be continuous or metric in nature.

III. METHODOLOGY

The implementation of the employee attrition rate prediction model involved three main components: data analysis and visualization, model building and prediction, and performance testing. For data analysis and visualization, the IBM Dataset obtained from Kaggle served as the dataset, and libraries such as Numpy, Pandas, Matplotlib, and Scikit-learn were utilized in the Python environment. Data preprocessing steps, including feature reduction, cleaning, binning, and balancing, were performed to prepare the data for analysis. Visualizations were created to gain insights into the data, identify trends, outliers, and understand the relationships among variables.

The model building and prediction phase began with splitting the dataset into a training set (75% of the data) and a test set (25% of the data) to train and evaluate the model. Feature selection techniques were applied to identify the most relevant attributes and reduce redundancy. The Random Forest algorithm, a powerful ensemble learning technique, was then implemented for both classification and regression tasks. Multiple decision trees were created, and their predictions were combined to obtain a more accurate and stable prediction.

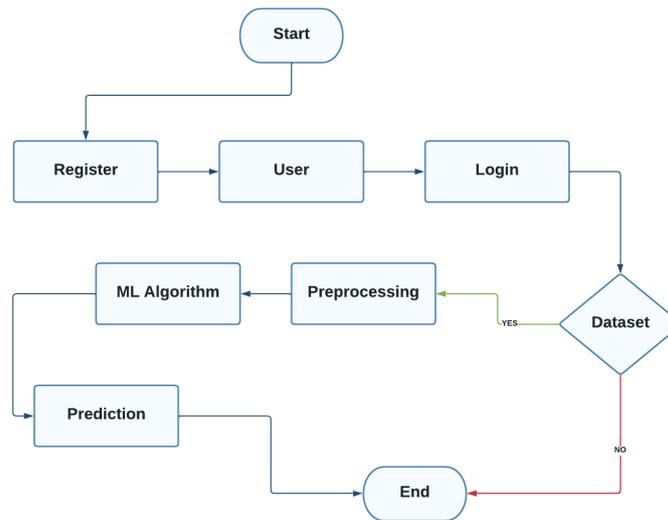


Fig-1: Architecture

To assess the performance of the model, various measures were considered. The accuracy of the model was evaluated using the 'accuracy_score' function, which compares the predicted attrition values with the actual attrition values. Additionally, a confusion matrix was generated to visualize the model's performance, providing insights into the number of correctly classified and misclassified records. The Receiver Operating Characteristic (ROC) curve was plotted to evaluate the model's performance, with the area under the curve (AUC) indicating the classifier's accuracy. The implementation of the employee attrition rate prediction model using the Random Forest algorithm showcased the practical application of machine learning techniques in addressing real-world challenges in human resources management. By leveraging data analysis, visualization, and advanced prediction models, organizations can gain valuable insights into attrition risks and make informed decisions regarding employee retention strategies. The implemented model provides a robust and accurate tool for predicting attrition and can be further enhanced and refined to meet specific organizational needs.

IV. APPLIED TOOLS AND TECHNOLOGIES

The Java Development Kit, or JDK 1.8 The Java Development Kit (JDK) is a software development environment for creating Java applications and applets. The Java Run-time Environment (JRE), an interpreter/loader (java), a compiler (javac), an archiver (jar), and a documentation generator (javadoc) are just a few of the Java development tools that are included.

A Java virtual machine (JVM) is a computer abstraction that allows a computer to run a Java programme. Specification, implementation, and instance are the three concepts that make up the JVM. A formal description of what a JVM implementation must do is provided in the specification. All implementations are guaranteed to be compatible by using a single standard.

An application that complies with the JVM standard is referred to as a JVM implementation. A process known as a JVM instance runs programmes that have been compiled into Java bytecode.

Databases

- The database is mostly used to store user information such as usernames and passwords. MYSQL GUI Browser was used to perform database functions.

Python

- Python is a server-side programming language used to construct web applications. Python's syntax enables programmers to construct programmes with fewer lines than other programming languages.



Random Forest Algorithm

A supervised learning technique called Random Forest is employed for both classification and regression.. Steps involved in RFA are as follows

A. Forest Formation at Random:

1. Choose "K" features at random from a total of "m" features, where $k < m$
2. Using the optimal split point, compute the node "d" among the "K" features.
3. Use the optimum split to divide the node into daughter nodes.
4. Repeat steps a-c until the "l" number of nodes is attained.
5. Build a forest by repeating steps a to d "n" times to build "n" trees.

B. Prediction Making Use of the Random Forest Classifier:

1. Takes the test characteristics and predicts the outcome using the rules of each randomly generated decision tree, then saves the projected outcome (target).
2. Determine the number of votes for each forecasted goal.
3. Take the most popularly anticipated target as the random forest algorithm's final prediction.

V. RESULTS

The project successfully delivers a comprehensive interface with interactive features, including a dashboard displaying key employee information, a user-friendly registration and login system, a contact form for user communication, and an efficient attrition predictor, enabling effective workforce management and decision-making.

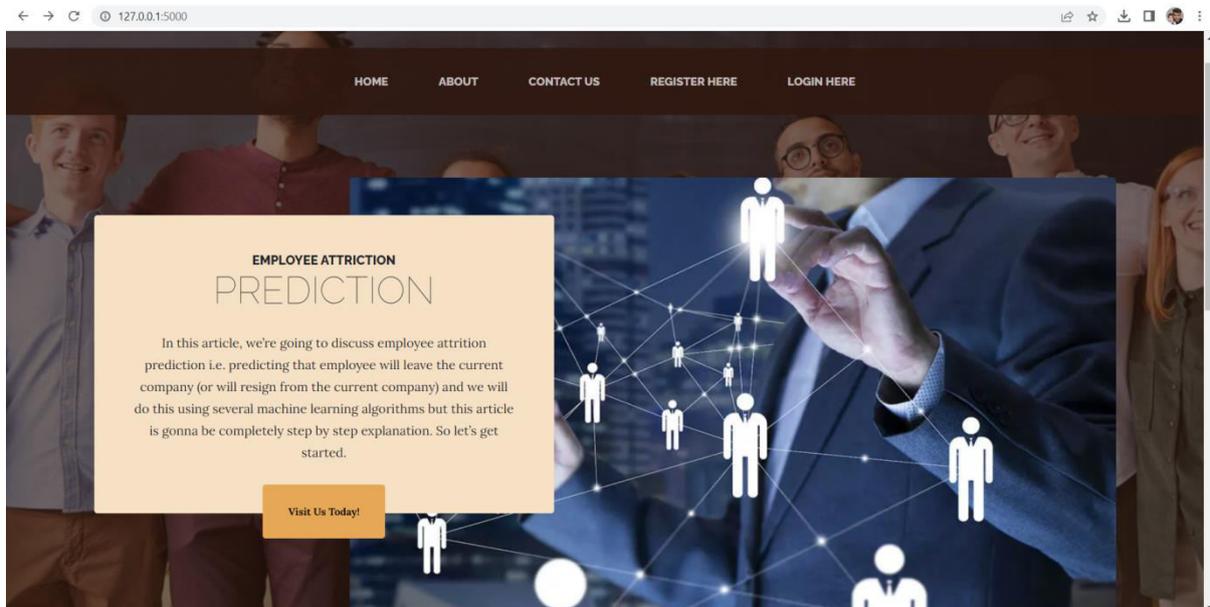


Fig-2: Home Page

The Home page encompasses various sections, including the home page, contact form, registration page, and dashboard, providing an interactive platform for users to navigate, communicate, and access different features.

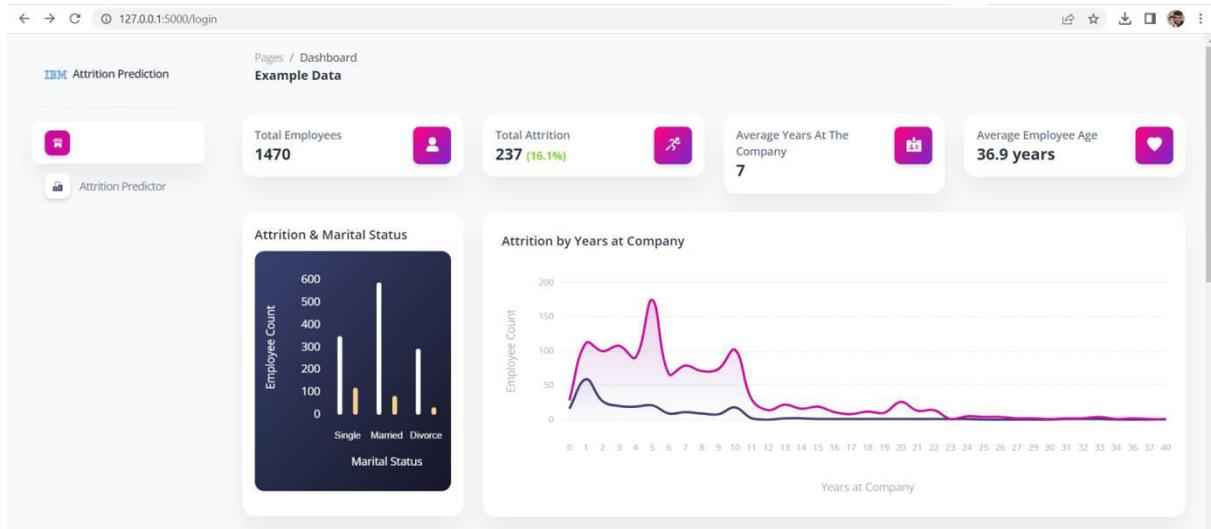


Fig-3: User Interface

This interface provides a dashboard displaying key information such as the number of employees, total attrition, average years at the company, and average employee age. It also provides visual representation of attrition with respect to attributes.

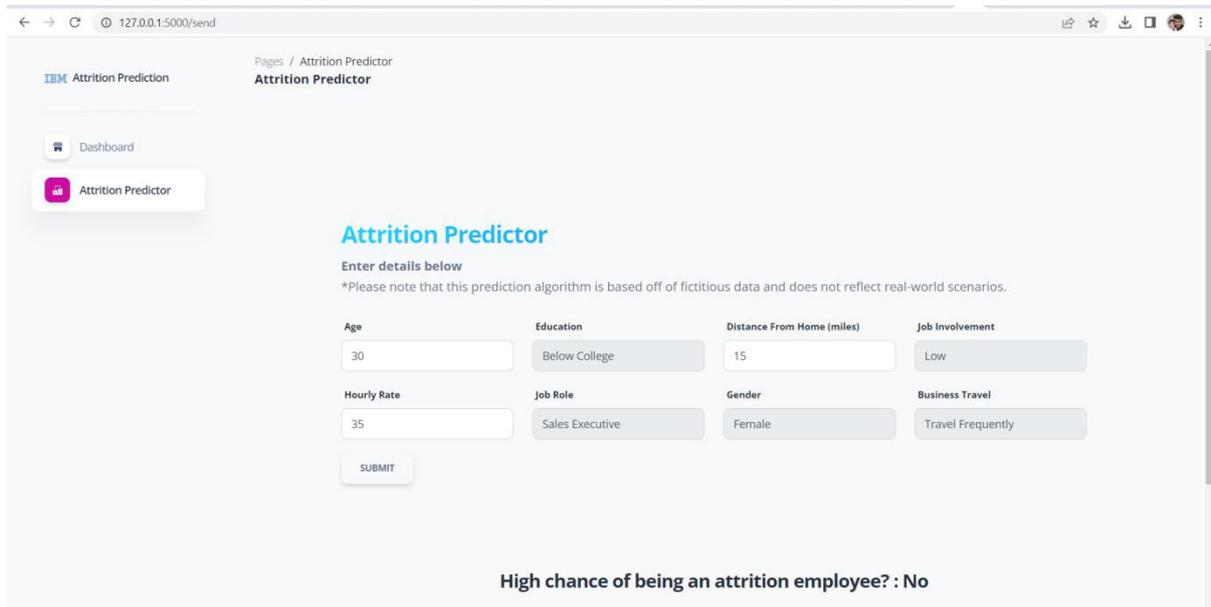


Fig-4: Attrition Predictor

The attrition predictor is a vital functionality within the project, where users can input employee details such as age, education, job involvement, and gender to predict the likelihood of attrition, providing valuable insights for workforce management and decision-making.

VI. CONCLUSION

For the majority of firms in the modern world, employee churn has been one of their top concerns. The organisation would have committed time and money, two of its most significant resources, in its employees, hence this issue has been impeding the firm's growth.



In order to forecast future attrition and identify the major variables influencing it, we examine employee data. The results show that a range of machine learning techniques can be used to establish and develop reliable models for comparing attrition rates. Finding employee attrition is only one part of our objective; we also want to identify the factors that influence it. The primary properties of the employee data can be examined using EDA (Exploratory Data Analysis) through visual representations like plots and graphs. Additionally, the choice of variables is crucial when developing a model because redundant and unnecessary variables can cause overfitting and sluggish computation. As a result, the Random Forest model uses linear Discriminant Analysis, an all-relevant variable selection technique, to pick variables, which improves model development. We can anticipate which employees are most likely to leave the company using a variety of machine learning techniques.

A comprehensive and universal programme that organisations can implement is a possible future direction. To reduce attrition rates, which will aid in both the growth of the company and the building of a strong bond between the workers and the company. A comprehensive and universal programme that organisations can implement is a possible future direction. To reduce attrition rates, which will aid in both the growth of the company and the building of a strong bond between the workers and the company.

REFERENCES

- [1] S. Jahan, "Human Resources Information System (HRIS): A Theoretical Perspective", Journal of Human Resource and Sustainability Studies, Vol.2 No.2, Article ID:46129, 2014.
- [2] C. Cortes and V. Vapnik, Support-vector networks. Machine learning, 20(3), 273-297, 1995
- [3] Rohit Punnoose, Pankaj Ajit, "Prediction of Employee Turnover in Organizations using Machine Learning Algorithms", (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 5, No.9, 2016.
- [4] "Guest Editorial: Special Issue on Early Prediction and Supporting of Learning Performance" IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 12, NO. 2, APRIL-JUNE 2019.
- [5] Renuka Agrawal, Jyoti Singh, and Zadgoankar .S, "Formative Assessment For Performance Evaluation Of Faculty Using Data Mining", International Journal Of Advances In Electronics And Computer Science, ISSN: 2393-2835.
- [6] Hossein Alizadeh, Buinzahra Branch and Islamic, 2016, "Introducing A Hybrid Data Mining Model To Evaluate Customer Loyalty", Engineering, Technology & Applied Science Research Volume. 6, No. 6, 1235- 1240.
- [7] Amir Mohammad Esmaieeli Sikaroudi, Rouzbehghousi and Ali Esmaieelisikaroudi, 2015 "A Data Mining Approach To Employee Turnover Prediction" (Case Study: Arak Automotive Parts Manufacturing), Journal Of Industrial And Systems Engineering Volume. 8, No. 4.



Impact Factor: 8.423



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details