Employee Attrition Prediction: A Data-Driven HR Decision Support System

A. Samriddhi Kathane[1], B. Samruddhi Tonghale[2], C. Shreya Bommarapu[3], D. Titiksha Tijare[4], E. Palak Bhimke[5], F. Prof. Rahul Suryawanshi[6]
[1,2,3,4,5]*Student, G H Raisoni College of Engineering and Management, Nagpur*
[6]*Assistant Professor, G H Raisoni College of Engineering and Management, Nagpur*

***Abstract*— Employee Attrition has become a major challenge for organizations these days, because it causes the loss of trained and experienced employees that leads to increased recruitment costs, reduced productivity and an overall decline in the team performance. Understanding different factors that contribute to this attrition factor and an early prediction to this can help the organizations in retention of the employees. The study examines colorful factors similar as job satisfaction, payment, workload, performance conditions, work-life balance and career growth openings to understand their influence on waste. Different types of Machine learning models such as Logistic Regression, Decision Trees, Random Forest and XGBoost are trained and compared to find the best prediction model. The results show that ensemble models like Random Forest perform particularly well, offering strong and harmonious predictions. Along with prediction, the design also highlights the most important factors contributing to waste, giving HR brigades precious perceptivity to improve hand engagement and retention. Overall, this exploration demonstrates how data- driven approaches can support smarter HR opinion and helps associations maintain a stable and motivated pool.**

Keywords: Employee Attrition Prediction, Machine Learning, XGBoost, Explainable Artificial Intelligence (XAI), SHAP Values, Workforce Overview, Employee Retention, Decision Tree, Random Forest, Support Vector Machine(SVM), HR Predictive Analytics

I. INTRODUCTION

Employee attrition is one of the major issues faced by organizations today. The key to long-term success for the companies is that the companies should get experienced employees, and retain their skills. If an employee leaves an organisation it causes a lot of losses not only financially but the team's morale, workflow, and overall the performance will be slowed. It is important to understand why an employee leaves and how attrition can be predicted, to help an HR.

HR teams traditionally were using surveys, personal judgements, and observation to evaluate the employees satisfaction and predict their turnover. This method often falls short, cannot cover all the hidden patterns, and subtle indicators of a reason that an employee would leave.

Now organisations have digital data and machine learning algorithms to improve their surface trends which were not that easy to find. This helps an HR person to find out which employee is likely to leave an organisation and help create new opportunities for predicting attrition more accurately.

This project has various factors on which we can predict the attrition of an employee such as: job satisfaction, salary level, workload, timings, marital status, age, education, work-life balance, gender, environment of the office, etc. By considering the above mentioned attributes we can find the best methodology for the prediction of the attrition using several machine learning algorithms and compare the results. This project not only considers the attributes but also focuses on understanding different features to make the model more helpful to the HRs. The objective of the research is to make the result practical. The study shows how the model explains the decisions that are taken by the model and which factors are responsible for this. Then an insight is generated which helps to improve the satisfaction of the employees, and strengthens the strategies, and improves the workplace.

This project not only reduces the employees turnover but also motivates the workforce, and makes it more stable. Also this machine learning feature helps HR to figure out things properly and in a more practical way.

II. METHODOLOGY

The steps we applied for building the Employee Attrition Prediction system follow as - data clean up, building models, testing them, then putting them live. This whole process mixes smart algorithms with tools that show how decisions are made, so results stay reliable, easy to grasp and ready to use.

A. Dataset Collection

The data for this project comes from workers files that hold personal details, job info, performance ratings, also behaviour notes. Main factors are age, sex, position at work, pay level, time spent with company,how well they mix life and duties, happiness with their role, commute length, along with past promotions. All these factors included together help to predict who might leave the company.

B. Data Preprocessing

Data cleaning was done first to boost accuracy before putting the information into the model. After that, missing bits were filled using estimates. Then, values were adjusted in a common range for the calculation. After that, the categories were turned into numerical data so that the computers could process them. At last, the dataset is split into chunks for testing and learning.
Filling Gaps : The empty or broken entries from the data are fixed or removed out to make the data normalized.
Turning categories into numbers : Different methods like one-hot encoder or label coding were used to change textual data like job type or marriage status into numerical data. Instead of words, we used numerical data for effective working of the system. For each category a matching number was assigned based on its group. Because of this, details like work position or relationship data became usable.
Rescaling number data : We adjusted the values using standard methods so that they would fit into a common range, which helped us to keep things consistent across features.
Feature Reduction : Unwanted traits are removed, so that things run faster and work better.
Dealing with uneven classes : SMOTE is used to make extra samples for rare cases, so that stayers and leavers end up equally. Instead of just adding data, it created realistic examples by blending nearby points. This ensured that the model learned well without bias.
The data was divided - 80% for training and 20% for testing.

C. Model Development

For choosing the best model for the project, we tried to check different models to see their performance. We tried the following types of models:

- Logistic Regression
- Random Forest
- Gradient Boosting
- Decision Tree
- Support Vector Machine(SVM)
- K-Nearest Neighbors(KNN)
- XGBoost(Extreme Gradient Boosting)

Models were trained on cleaned-up data, then adjusted according to ideal settings. After observing each model, we found out that XGBoost worked better than others. We then chose it because it not only deals well with organized information but also maps complex patterns while avoiding noise.

D. Prediction and Validation

Once our models were trained, we checked every model with the remaining testing data. Following are the factors we determined in the testing of the model:

- Accuracy
- Precision
- Recall
- F1-Score

The factors are used to predict how well a model can guess which employee is likely to quit the company. Since XGBoost gave the best prediction of all the models, we included it in the final prediction of our project.

E. Ways to make AI easier to understand

To make things easier, we used several different ways to explain AI, such as showing how decisions are being made or explaining each step of the model instead of just listing them.
SHAP(Shapley Additive Explanations) : Provide a clear detail about the importance of each feature for each case, one by one.

LIME : Helps generate localized explanations for specific predictions. Also provides insights of how each part affects the choices and where the output comes from.
SHAPASH : It gives clear visuals to HR, which they can actually make use of. These tools help the HR to view why the employee may quit, so that they can rely on these forecasts before taking any action.

F.Working and Setting up the system
The final version of the working model is deployed as a web app that was built with Django, it helps the HR to use the model without any problem. The model includes factors like :
Uploading employee CSV datasets
Automatic Prediction creation
Probability of leaving of individual employee
Visual dashboards using Chart.js and AG-Grid
PDF report generation
Maintaining the records of employees safe in a MySQL setup
The overall model used for the attrition prediction ensures that the tools work well, stay easy to use and at the same time integrates with the human resource environment.

III. RESULT AND DISCUSSION

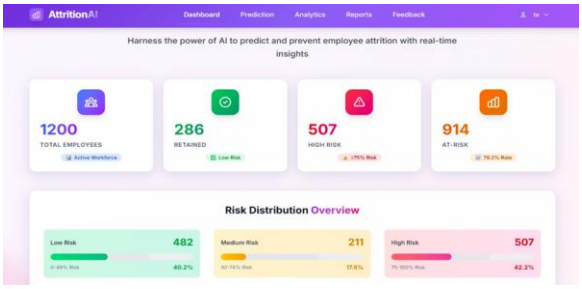

Fig. 1 Analytics Dashboard Page

This is the main dashboard page which is visible after the user login. This gives an overview of the organization's employee status.

Fig. 2 Workforce Overview of Risk, Attrition and Satisfaction level of employees

Showcase the risk distribution, attrition by gender and their satisfaction levels. Everything is presented in a different color so that it becomes easy to understand.

Fig. 3 Hiring Analysis and Comparison of Hiring and Analysis.

Here the red and green color bar graph gives a clear representation of how many people were hired and how many of them were turnover.

This project Employee Attrition Prediction model gives positive results and helps the organisation to manage the turnover. It provides an attrition risk score and the attribute on which the result has been generated. This helps HR to identify which employee is at a high risk and why.
By having clear understanding and analysis of the data by examining the age, experience levels, job roles etc helps the organisations to improve the work conditions.
The web page has a dashboard which makes it easy to understand the results in a more simple way. Managers can quickly get an overview of the risk levels of the employees through simple charts and graphs, which helps the HR manager to make better further decisions.
At last the results show a predictive analysis on which the HR team can work and improve their policies and reduce the resignations, and prevent the retention, and help long-term organisational growth.

IV. CONCLUSION

This project demonstrates the machine learning techniques in predicting the attrition based on the historical data. After comparing with a lot of methods

like the Logistic Regression, K Nearest Neighbour (kNN), XGBoost, Support vector machine (SVM), and Random Forest, XGBoost was giving a higher accuracy score than others so we chose it.

This model is able to capture complex patterns by using key factors like salary, job satisfaction, age, gender, etc as these were the more significant topics and on that basis the employees who are at high risk were identified. The project was having an frontend as Django, Chart.js, AG-Grid and MySQL for database and machine learning model as backend.

The future scope for this project can be working on the accuracy as it was not 100% accurate, working with real-time data can make the system work more efficiently.

This is the robust, scalable solution for the attrition of the employee. This fills the gaps between the traditional methods and modern approach and helps to reduce the turnover and increase the retention. This model also increases the satisfaction level of the employees.

REFERENCES

[1] Mohiuddin, Karishma, Mirza Ariful Alam, Mirza Mohtashim Alam, Pascal Welke, Michael Martin, Jens Lehmann, and Sahar Vahdati. "Retention is all you need." In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 4752-4758. 2023.

[2] Ma, Xiaoye, Weiheng Liu, Changyi Zhao, and Liliya R. Tukhvatulina. "Can Large Language Model Predict Employee Atrition?." In *Proceeding of the 2024 5th International Conference on Computer Science and Management Technology*, pp. 1164-1172. 2024.

[3] Iparraguirre-Villanueva, Orlando, Luis Chauca-Huete, Rosas Prieto-Chavez, and Cleoge Paulino-Moreno. "Employee attrition prediction using machine learning models." In *Proceedings of the 22nd LACCEI International Multi-Conference for Engineering, Education and Technology (LACCEI2024): "Sustainable Engineering for a Diverse, Equitable, and Inclusive Future at the Service of Education, Research, and Industry for a Societ, Latin American and Caribbean Consortium of Engineering Institutions*. 2024.

[4] Alshiddy, Muneera Saad, and Bader Nasser Aljaber. "Employee attrition prediction using nested ensemble learning techniques." *International Journal of Advanced Computer Science and Applications* 14, no. 7 (2023).

[5] Guerranti, Filippo, and Giovanna Maria Dimitri. "A comparison of machine learning approaches for predicting employee attrition." *Applied Sciences* 13, no. 1 (2022): 267.

[6] Fallucchi, Francesca, Marco Coladangelo, Romeo Giuliano, and Ernesto William De Luca. "Predicting employee attrition using machine learning techniques." *Computers* 9, no. 4 (2020): 86.

[7] Gazi, Md Sumon, Md Nasiruddin, Shuvo Dutta, Rajesh Sikder, Chowdhury Badrul Huda, and Md Zahidul Islam. "Employee attrition prediction in the usa: A machine learning approach for hr analytics and talent retention strategies." *Journal of Business and Management Studies* 6, no. 3 (2024): 47-59.

[8] Jain, Nishant, Abhinav Tomar, and Prasanta K. Jana. "A novel scheme for employee churn problem using multi-attribute decision making approach and machine learning." *Journal of Intelligent Information Systems* 56, no. 2 (2021): 279-302.

[9] Konar, Karabi, Saptarshi Das, Samiran Das, and Sanjay Misra. "Employee Attrition Prediction Using Bayesian Optimized Stacked Ensemble Learning and Explainable AI." *SN Computer Science* 6, no. 6 (2025): 672.

[10] Yin, Ziqing, Baojun Hu, and Shuhan Chen. "Predicting employee turnover in the financial company: A comparative study of catboost and xgboost models." (2024).

[11] Djarum, Danny Hartanto, Zainal Ahmad, and Jie Zhang. "Reduced Bayesian Optimized Stacked Regressor (RBOSR): A highly efficient stacked approach for improved air pollution prediction." *Applied Soft Computing* 144 (2023): 110466.

[12] Zhang, Yiting, Ziling Cai, and Hongyang Feib. "Using Machine Learning: Based on the Psychological." In *Proceedings of the 2024 2nd International Conference on Digital Economy and Management Science (CDEMS 2024)*, vol. 292, p. 341. Springer Nature, 2024.

[13] Ma, Xiaoye; Liu, Weiheng; Zhao, Changyi; Tukhvatulina, Liliya R. "*Can Large Language Model Predict Employee Attrition?"* Published in: *Proceedings of the 5th International*

Conference on Computer Science and Management Technology 2024

[14] Ahmed, Esmael, and Mohammed Omer. "Predicting Employee Attrition Using Artificial Neural Networks: A Comparative Study of Machine Learning Models and Imbalanced Data Handling Techniques." *Available at SSRN 5105905*.

[15] Atique, Md Monir Ahammod Bin, Md Nesarul Hoque, and Md Jamal Uddin. "Employee attrition analysis using CatBoost." In *International Conference on Machine Intelligence and Emerging Technologies*, pp. 644-658. Cham: Springer Nature Switzerland, 2022.