ISSN 2277-2685
IJESR/April-June. 2025/ Vol-15/Issue-2s/625-634
Sangannagare Chandana *et. al.*, / International Journal of Engineering & Science Research

Employee Attrition Prediction

[1]A Hima Bindu, [2]Sangannagare Chandana, [3]Gundla Harshitha, [4]Mittapelli Hasmitha
[1]Assistant professor, Department of CSE, Bhoj Reddy Engineering College for Women, India
[2,3,4]B.Tech Students, Department of CSE, Bhoj Reddy Engineering College for Women, India

ABSTRACT

The attrition of employees is the problem faced by many organisations, where valuable and experienced employees leave the organisation on a daily basis. Many businesses around the globe are looking to get rid of this serious issue. The main objective of this research work is to develop a model that can help to predict whether an employee will leave the company or not. The essential idea is to measure the effectiveness of employee appraisal and satisfaction rates within the company, which can help to reduce the attrition rate of employees.

Nowadays, Employee Attrition prediction has become a major problem in the organisations. Employee Attrition is a big issue for the organisations, especially when trained, technical and key employees leave for a better opportunity from the organisation. This results in financial loss to replace a trained employee. Therefore, we use the current and past employee data to analyse the common reasons for employee attrition. For the prevention of employee attrition, we applied well known classification methods, that is, Logistic Regression, SVM, KNN, Random Forest, XG boost methods on the human resource data. For this we implement a feature selection method on the data and analyse the results to prevent employee attrition. This is helpful to companies to predict employee attrition, and also helpful to their economic growth by reducing their human resource cost.

1. INTRODUCTION

Employee Attrition is a reduction in manpower in any organisation where employees resign Employees are the valuable assets of any organisation. It's necessary to know whether the employees are dissatisfied or are there any other reasons for leaving the respective job. These days for better opportunities, employees are eager to jump from one organisation to another. But if they leave their jobs unexpectedly, it may cause a huge loss for the organisation. New hiring will consume money and time, and also the freshly hired employees take time to make the respective organisation profitable. Retention of skilled and hardworking employees is one of the most critical challenges faced by many organisations.This system is able to predict which employee may leave an organisation with what reason, so that they can take several corrective actions in order to ensure that employees stay in the organisation and can reduce attrition. Some of the employee retention strategies to control attrition are motivating employees, exposing 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 Neighbour), Decision Tree and Random Forest. Graphical representation is also provided for better understanding of insights.Regularly monitoring employee satisfaction, offering professional growth opportunities, and ensuring competitive compensation packages can enhance employee loyalty. By predicting potential attrition early, companies can implement proactive retention measures, such as targeted training, mentoring programs, or changes in leadership

styles. Moreover, fostering a culture of open communication and work-life balance can significantly reduce the likelihood of employees seeking opportunities elsewhere.

2. REQUIREMENT ANALYSIS

Functional Requirements

Modules:

- User:

Inputs the Data : HR inputs employee data

Views Result: It predicts the result of employee attrition

Non-Functional Requirements

- **Performance**: The system's performance is measured by its response time and throughput, ensuring fast and efficient processing of user requests and predictions.
- **Reliability**: The system guarantees high availability, fault tolerance, and robust backup and recovery mechanisms to ensure continuous operation and minimize downtime.
- **Security**: The system incorporates data encryption, access control measures, and audit logging to protect sensitive information and monitor user activities for potential security breaches.
- **Usability**: The system is designed with an intuitive user interface, complemented by comprehensive documentation and training materials to enhance user experience and ease of use.
- **Maintainability**: The system is built with modularity, high code quality, and thorough testing to facilitate easier maintenance and quick updates or fixes.
- **Portability**: The system ensures platform independence, allowing it to run on different operating systems, and provides flexibility in deployment across diverse environments.

3-DESIGN

Architecture

An architecture description is a formal description and representation of a system organized in a way that supports reasoning about the structure of the system. Architecture is of two types. They are

(1) Software Architecture

(2) Technical Architecture

Software Architecture

Software architecture defines the high-level structure of a software system, outlining components, their interactions, and technology choices. It ensures the system is scalable, maintainable, and aligned with business needs.

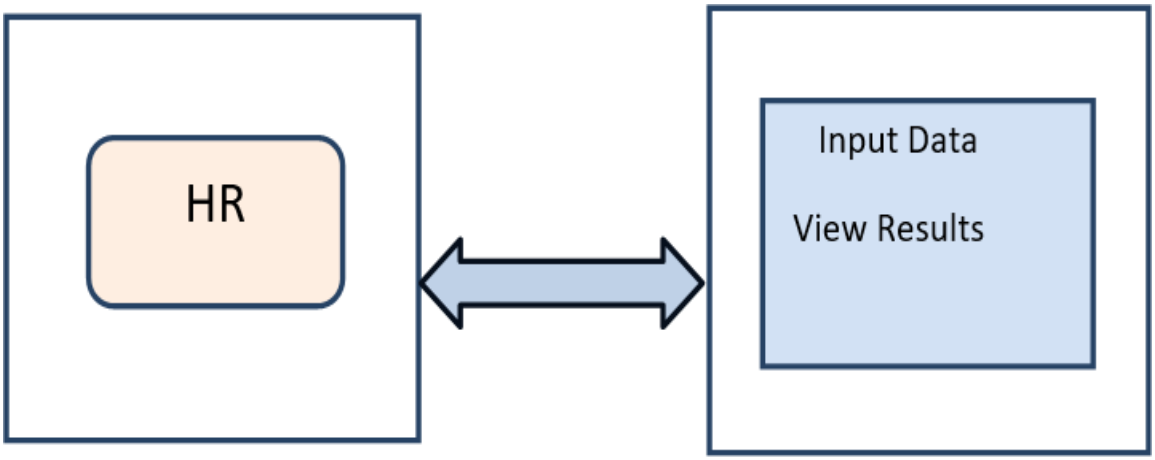


Fig 3.1 Software Architecture

Technical Architecture

It refers to the design and organization of the physical components of a system, including processors, memory, and input/output devices. It defines how these components interact to support system functionality and performance.

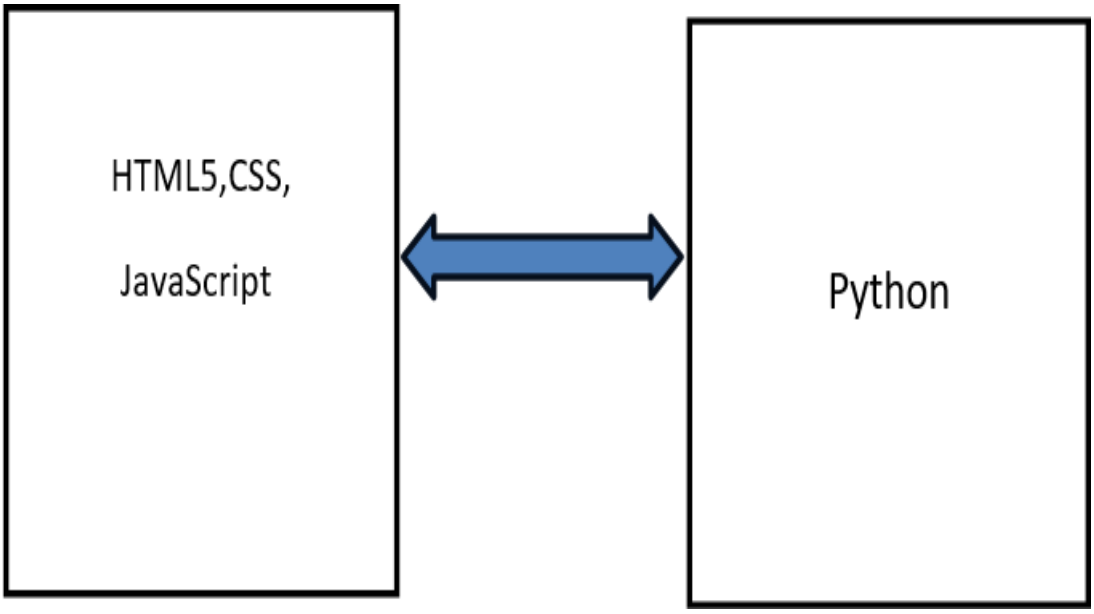


Fig 3.2 Technical Architecture

The proposed employee attrition prediction system begins by taking the employee dataset as input, which includes attributes such as employee salary, job role, performance metrics, and satisfaction levels. The data is pre-processed to handle missing values and outliers, ensuring consistency and quality. Feature selection techniques are applied to identify the most relevant attributes, followed by the application of ML algorithms to select the optimal representative attributes. These attributes are then used to train different algorithms, including Logistic Regression, Support Vector Machine (SVM), and Random Forest, etc. among others. The performance of these algorithms is evaluated using metrics such as accuracy, recall, precision, and F-measure.

Data Flow Diagram

A Data Flow Diagram (DFD) is a visual representation of how data flows through a system, highlighting processes, data storage, and external entities. It consists of four key components: processes, data stores, data flows, and external entities. The DFD shows how data is input, processed, stored, and output by the system. There are multiple levels of DFDs, with Level 0 providing a high-level view and subsequent levels offering more detailed breakdowns of processes. DFDs are useful for simplifying system understanding, facilitating communication among stakeholders, and ensuring clear documentation of system operations. They help both technical and non-technical users grasp the functionality of a system. The DFD is widely used in system analysis and design,

providing clarity in data management.By breaking down complex systems into hierarchical levels,

from high-level overviews to detailed analyses, DFDs enhance system understanding, streamline communication among stakeholders, and ensure accurate documentation for effective system design and analysis.

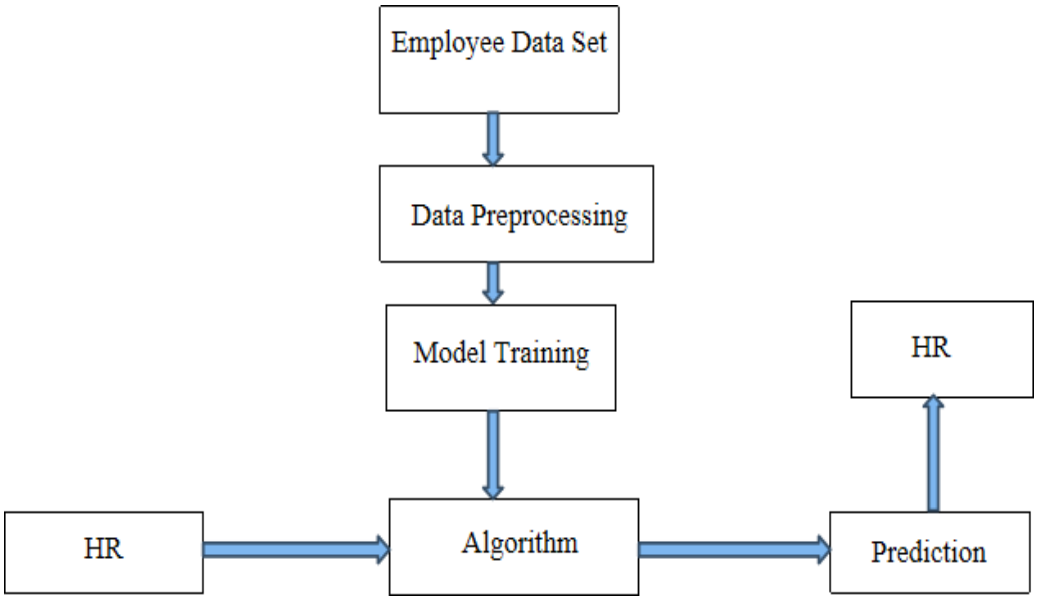


Fig 3.3 data flow diagram

4-IMPLEMENTATIION

Technologies

This system is developed using Python programming language.

Python

Python is one of the most popular programming languages today, known for its simplicity and extensive features.It was created by Guido van Rossum, and released in 1991. Its clean and straightforward syntax makes it beginner-friendly, while its powerful libraries and frameworks makes it perfect for developers. It can be used on a server to create web applications. It can be used to handle big data and perform complex mathematics.Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.Python can be treated in a procedural way, an object-oriented way or a functional way.Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose.Python is a versatile, high-level programming language that emphasizes simplicity and readability. It is widely used in various fields, including web development, data science, artificial intelligence, and machine learning, making it ideal for projects.

5-TESTING

Testing is a critical phase in the software development life cycle that ensures the system functions as intended and meets all specified requirements. For the breast cancer prediction project, testing was conducted to verify the

accuracy, reliability, and performance of the system, focusing on its ability to provide correct predictions and handle various scenarios effectively.

Testing Metrics

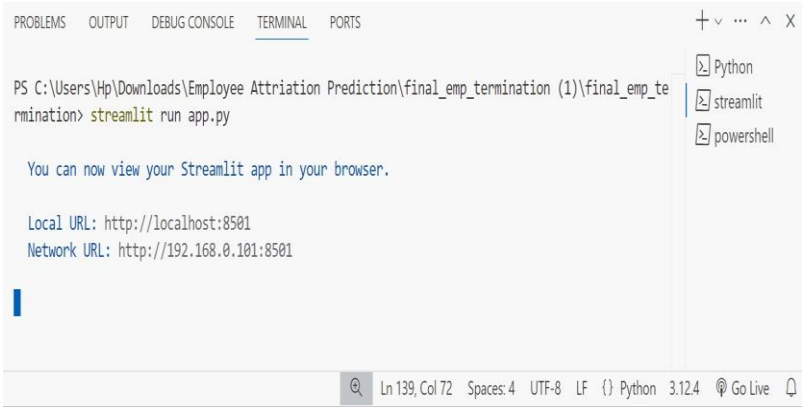
1. **Accuracy**: Proportion of correct predictions made by the system.
2. **Precision**: Ratio of correctly predicted positive observations to total predicted positives.
3. **Recall**: Measure of the model's ability to identify all actual positive cases.
4. **F1-Score**: Harmonic mean of precision and recall, used to balance both metrics.
5. **Execution Time**: Time taken to process input and return predictions

Test Cases

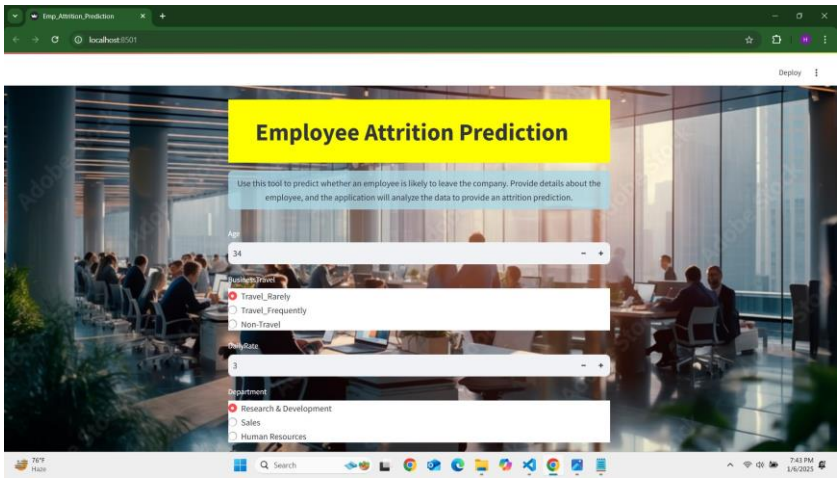
Tested	Test name	Inputs	Expected output	Actual Output	status
1	Load Dataset	Csv file	Read dataset	Load dataset	success
2	Split dataset	Train70% and test30%	Divide the training set and testing set	Split train and Test	success
3	Train Model with Feature Selection	Train dataset	8 best features selected	8 best features selected	success
4	Test Data with Attrition Input	Test column with high attrition values	Predicted as attrition	Predicted as attrition	success
5	Test Data with Non-Attrition Input	Test column with low attrition values	Predicted as no attrition	Predicted as no attrition	success
6	Test Data with Invalid Input	Test column with invalid datatype	Error: Invalid input type	Error: Invalid input type	success

Table 5.2 Test cases

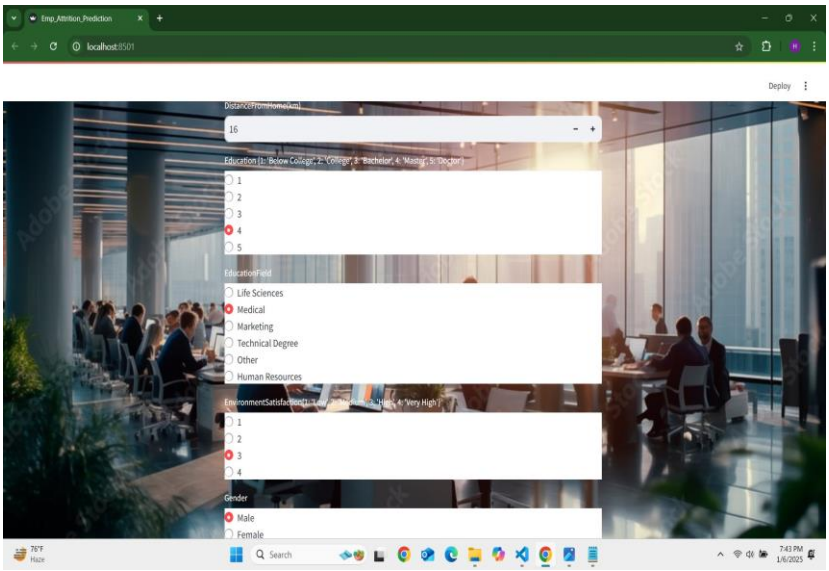
6-SCREENSHOTS



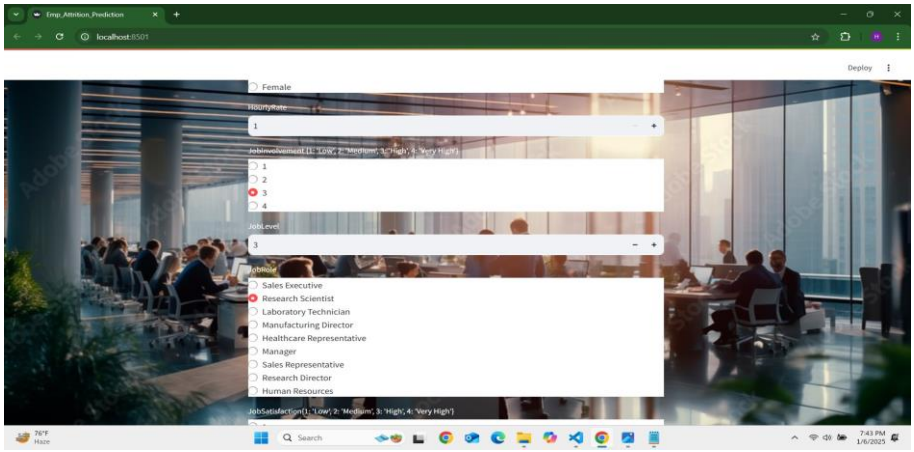
Screenshot 6.1 Link Generated for Webpage



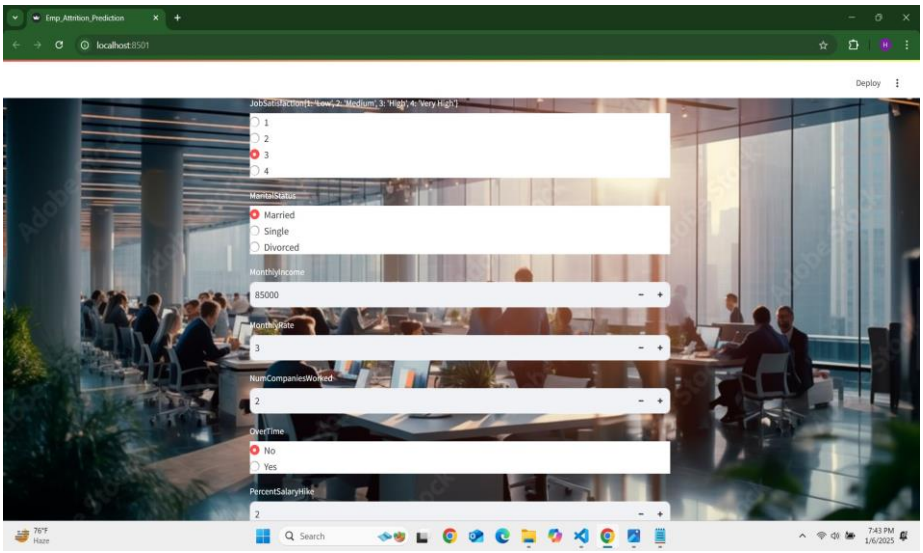
Screenshot 6.2 Image Showing values entered by user



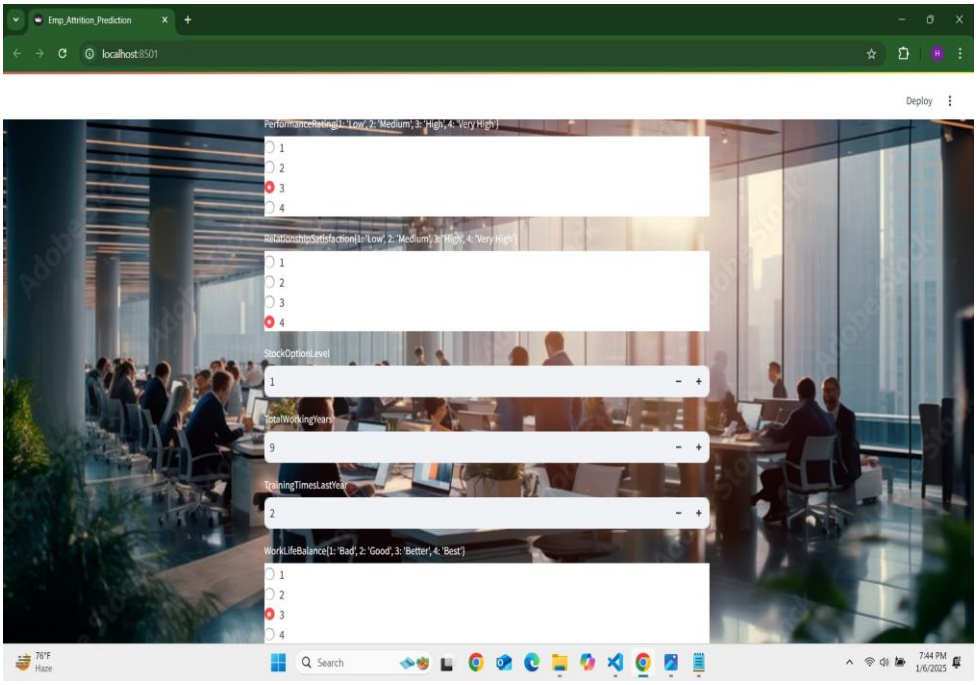
Screenshot 6.3 Image Showing another values entered by user



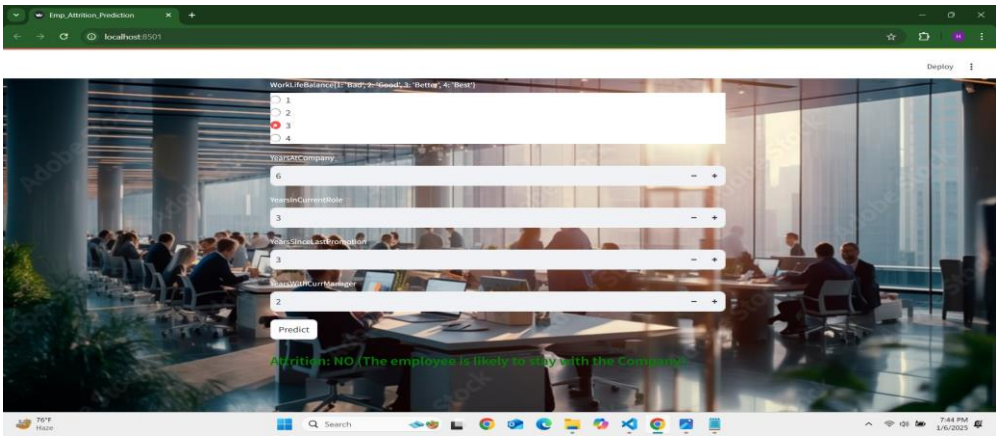
Screenshot 6.4 Image Showing another values entered by user



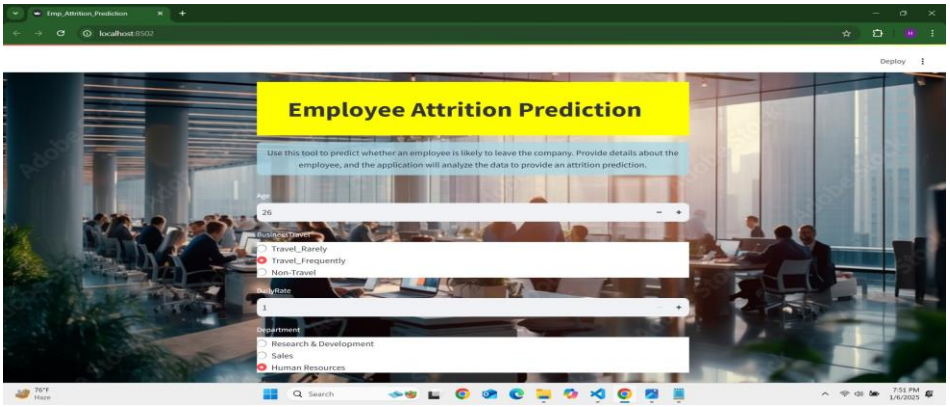
Screenshot 6.5 Image Showing another values entered by user



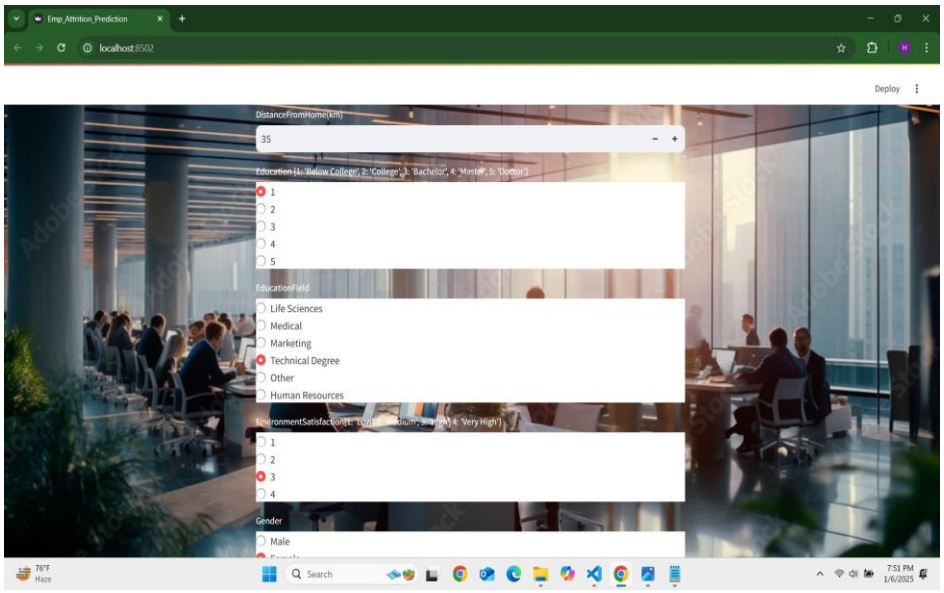
Screenshot 6.6 Image Showing another values entered by user



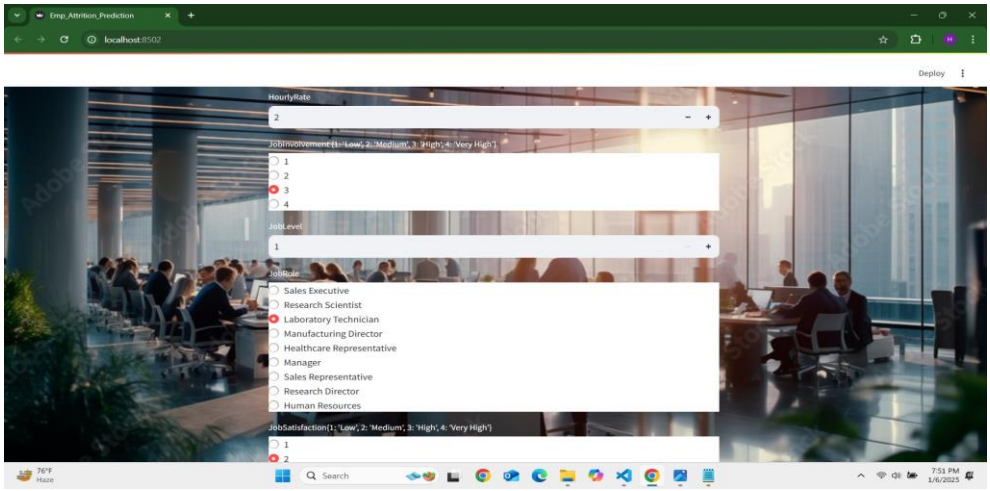
Screenshot 6.7 Image Showing Result1



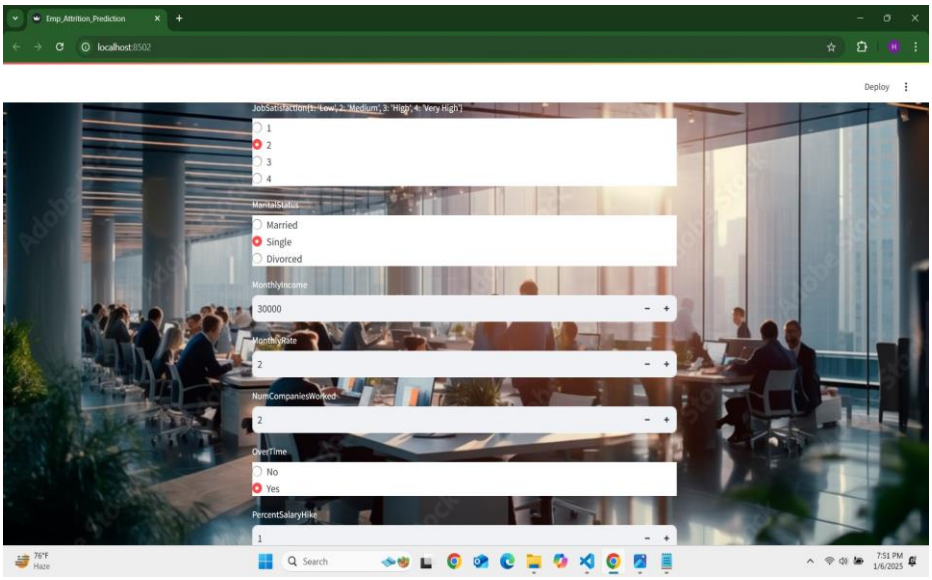
Screenshot 6.8 Image Showing values entered by user



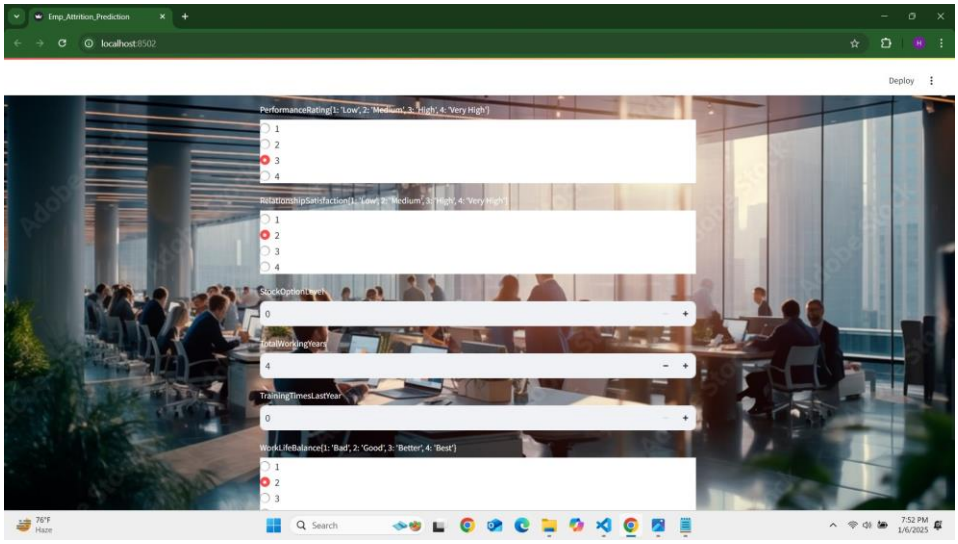
Screenshot 6.9 Image Showing another values entered by user



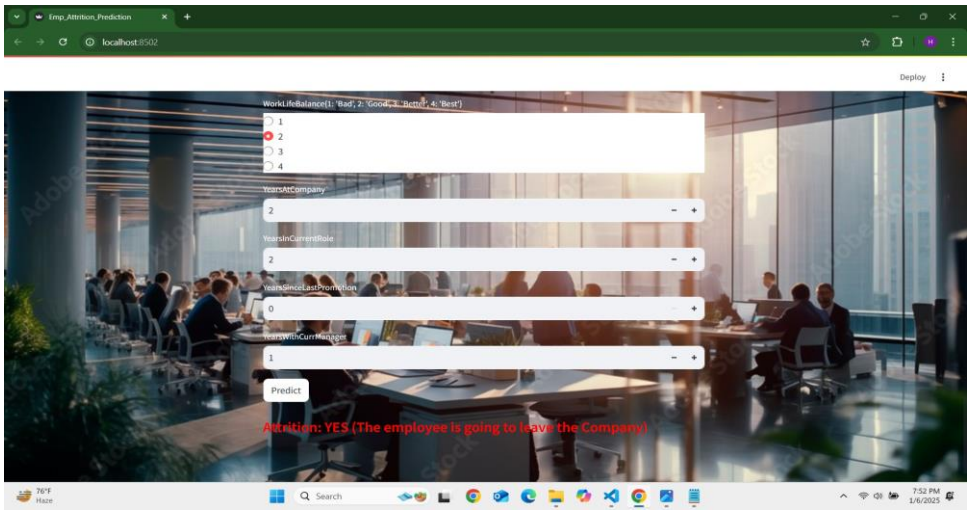
Screenshot 6.10 Image Showing another values entered by user



Screenshot 6.11 Image Showing another values entered by user



Screenshot 6.10 Image Showing another values entered by user



Screenshot 6.13 Image Showing result2

7-CONCLUSION AND FUTURE SCOPE

Conclusion

Employee attrition can affect an organisation in many ways like goodwill, revenues and cost in terms of both time and money. The predictive attrition model helps in not only taking preventive measures, but also making better hiring decisions. In this study implementation of various classification methods helps in predicting whether a particular employee might leave the organization in the near future by deriving trends in the employee's past data. It was intuited that salary or other financial aspects like promotions are not the sole reasons behind the attrition of employees. These models can help us in prioritising the features with higher impact in attrition of an employee and the possible reasons behind it so that HR can make appropriate decisions for the retention process. The main purpose of this research is to build reliable and accurate models which can optimise the hiring and retention cost of quality employees. This could be done by determining the attrition status of employees under consideration by using the appropriate data mining techniques.

Future Scope

The future of employee attrition prediction using machine learning lies in improving workforce planning, enhancing employee engagement, and enabling proactive retention strategies. It includes integration with advanced technologies like NLP, AI-powered decision-making, and predictive analytics. Real-time monitoring, ethical AI practices, and industry-specific customization will drive adoption.

8-REFERENCES

[1]D. S. Sisodia, S. Vishwakarma and A. Pujahari, "Evaluation of machine learning models for employee churn prediction," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 1016-1020, doi: 10.1109/ICICI.2017.8365293.

[2] S. Alduayj and K. Rajpoot, "Predicting Employee Attrition using Machine Learning." 2018 International Conference on Innovations in Information Technology (IIT), Al Ain, United Arab Emirates, 2018, pp. 93-98, doi: 10.1109/INNOVATIONS.2018.8605976.

[3]R. Khare, D. Kaloya, C. K. Choudhary, and G. Gupta, "Employee attrition risk assessment using logistic regression analysis,"

[4] M. L. Kane-Sellers, Predictive models of employee voluntary turnover in a North American professional sales force using data-mining analysis

[5]V. V. Saradhi and G. K. Palshikar, "Employee attrition prediction," Expert Systems with Applications, vol. 38, no. 3, pp. 1999-2006, 2011.

[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), Volume6, Issue 1, pp. 425-431