

Employee attrition Prediction based on Gradient Boosting Approach

ABSTRACT

In today's organizational landscape, predicting employee attrition has emerged as a critical concern. The departure of trained, technical, and pivotal staff members poses significant challenges, including financial setbacks incurred in their replacement. To address this, organizations harness current and historical employee data to discern prevalent attrition triggers. Employing established classification methodologies such as Decision Tree, Logistic Regression, Random Forest, Support vector machine, and Gradient boosting Algorithms are constructed using human resource data. Leveraging feature selection techniques, these models facilitate proactive measures to mitigate attrition risks. By accurately forecasting attrition, companies not only fortify their workforce stability but also enhance economic resilience through diminished human resource expenditures. This proactive approach not only aids in retaining valuable talent but also fosters sustainable growth by fostering an environment conducive to employee retention and organizational stability.

Keywords: Employing attrition, Machine learning, Gradient Boosting, Prediction, Feature selection

1. INTRODUCTION

Employee retention is a multifaceted issue influenced by various factors such as workplace environment, gender and pay equity and personal circumstances like family relocation or health concerns [1]. Attrition poses significant challenges for organizations, especially when skilled employees depart for better opportunities. This leads to financial losses and disrupts workflow. Analysis of present and historical employee data helps identify patterns and underlying causes of attrition, enabling proactive measures for retention. Understanding the complexities surrounding employee departures is crucial for implementing effective strategies to foster a supportive work environment and mitigate turnover, ultimately ensuring organizational stability and success [2].

Identifying employee attrition aids in predicting and addressing retention issues. Utilizing this data allows organizations to mitigate attrition rates and implement strategies for employee retention.

In our analysis, we employ various data classification methodologies to effectively address employee attrition. Decision Tree, characterized by its tree structure comprising branches, root nodes, and leaf nodes, facilitates attribute testing to determine class labels. Naive Bayes, leveraging the Bayes Theorem, assumes independence among features, allowing for classification based on individual feature probabilities [3]. For instance, in classifying fruits, features like color, shape, and size contribute independently to the probability of a fruit being identified as an apple. Logistic Regression, a statistical technique, evaluates datasets with independent variables to predict outcomes. By utilizing these methodologies, we can accurately identify patterns and factors contributing to employee attrition, enabling organizations to implement targeted retention strategies [4]. This comprehensive approach empowers decision-makers to proactively mitigate attrition rates and foster a supportive work environment conducive to employee retention and organizational success.

System architecture, the backbone of any software solution, undergoes meticulous design phases crucial for translating requirements into tangible systems. The design phase initiates the transformation from the problem domain to the solution domain, ensuring alignment with system requisites [5]. Its significance lies in its profound impact on software quality, influencing subsequent stages like testing and maintenance. Central to the design phase is the creation of a blueprint document outlining the system's structure, akin to an architectural plan guiding implementation, testing, and maintenance endeavors. This phase typically bifurcates into both detailed and conceptual system design [6].

Detailed Design, is known as high-level design, delineates the system's modules, their specifications, and their interrelations to achieve desired outcomes. Conceptual System Design delves deeper into the internal logic of each module outlined in System Design. While System Design focuses on module identification, Detailed Design delves into logic planning for each module [7].

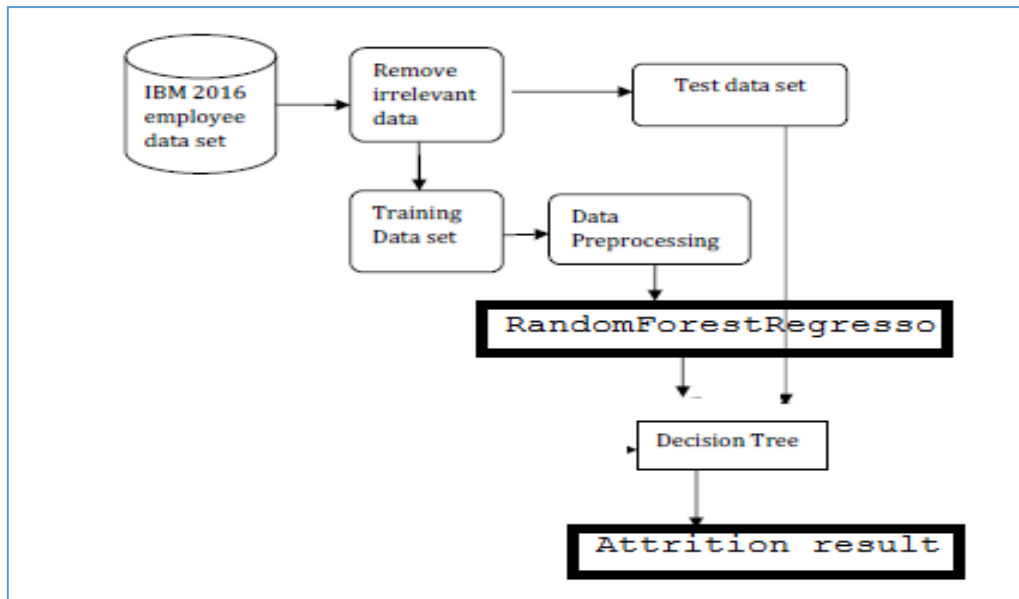


Figure 1. System Architecture

The following paper provides the literature review in section 2. Section 3 presents the methodology. Section 4 states the results and analysis. The final section concludes the paper.

2. LITERATURE REVIEW

Various software development tactics focus on the entire life cycle of systems development (the development life cycle), which includes the full process of creating or redesigning systems [8]. It involves systematic stages from initiation to maintenance, guiding engineers through planning, design, implementation, testing, and deployment. SDLC ensures structured, efficient development and optimization of systems [9].

Predictive analysis in current mechanisms often employs dictated data mining techniques. Knowing that it is prohibitively costly and challenging to hire back qualified workers, employee attrition exerts a significant financial, operational, and resource burden on firms, underscoring the necessity for innovative retention strategies [10].

Standard attrition identification advances by employing unstructured text data, which boosts scientific performance. using determining the presence of potential customers with greater dropout risk, this study aids those with marketing authority make

intelligent decisions that will increase the sustainability of client tactics and profitability [7].

Attrition among employees can be a major problem for businesses, especially when highly trained, technical, and critical employees leave for a better opportunity elsewhere. The increased interest in machine learning among company leaders and call centers necessitates that researchers investigate its application within businesses [11]. One of the most serious issues confronting business owners is the loss of talented employees. In a lot of research papers, there are machine learning models developed using various algorithms like Support Vector Machine, Decision trees, XGBoost, K Nearest Neighbors, ANN, Random Forest, etc. Machine learning has been used to predict employee behavior in several types of research. To predict employee performance, the authors employed decision trees (ID3 C4.5) and the Naive Bayes classifier in their research. They discovered that task title was the most important factor, while age had no discernible effect. The authors used a dataset of 1585 records and 27 characteristics to test manifold data mining techniques for predicting staff churn. They employed Bayes algorithms, SVM, logistic regression, decision trees, and random forests as machine learning methods. For all of these findings, a support vector machine with an accuracy of 85.12 percent should be considered.

3. METHODOLOGY

3.1 Gradient Boosting Algorithm

Gradient Boosting is a formidable boosting handle that turns consecutive weak learners into strong learners. Each new model is trained using gradient descent to minimize the loss function, such as the average squared error or entropy cross of the prior model. According to the current ensemble's predictions, the algorithm calculates the gradient of the loss function in every loop. It then trains a new weak model to minimize this gradient. After adding the new model's predictions to the ensemble, the procedure goes through until a stopping specification is established [10].

More research in the field of attrition may be found. Because the strategy to forecast employee attrition is quite similar to erosion, it enables us to predict alternative ways” [11] In [13], combining various training previous observations per employee from Training Data improves the predicted performance of retention models compared to

using simply the most relevant data. Another issue is that instead of obtaining several samples from the whole term of the individuals, they limit it to a small piece of data, implying that many jobs are once again eliminated.

4. RESULTS AND ANALYSIS

Employing established classification methodologies such as the Gradient boosting Algorithm is constructed using human resource data. Leveraging feature selection techniques, these models facilitate proactive measures to mitigate attrition risks. By accurately forecasting attrition, companies not only fortify their workforce stability but also enhance economic resilience through diminished human resource expenditures. This proactive approach not only aids in retaining valuable talent but also fosters sustainable growth by fostering an environment conducive to employee retention and organizational stability.

For implementation analysis, the data set is gathered from the Kaggle database, an open-access repository. Then trained data set Machine Learning models using the k-fold validation methodology, using 75 percent 25 percent dataset splits. XGBoost is the machine learning algorithm employed in research to pick the most accurate model out of all of them and compare their accuracies”.

Initially, the null values are checked in the dataset with particular functions. Then the outcome is depicted in the shape of a warmth map that showcases the null values in the dataset pictorially.

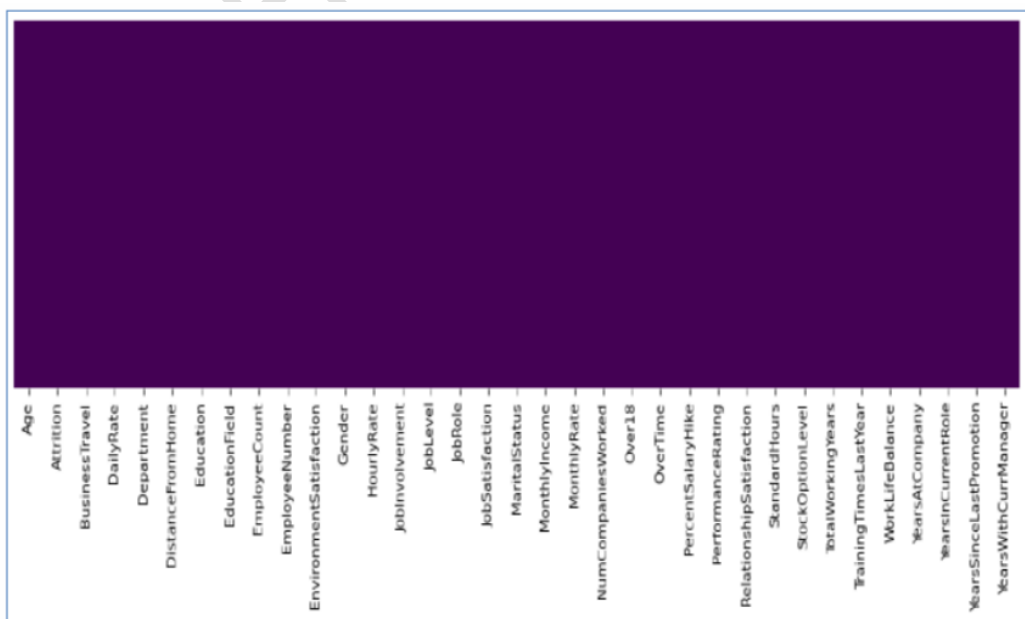


Figure 2: Heat Map

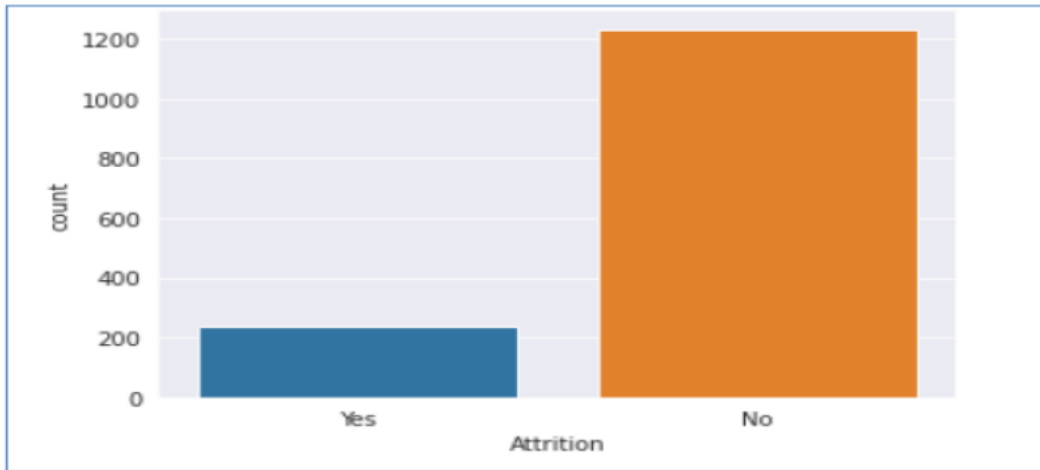


Figure 3: count of attritions

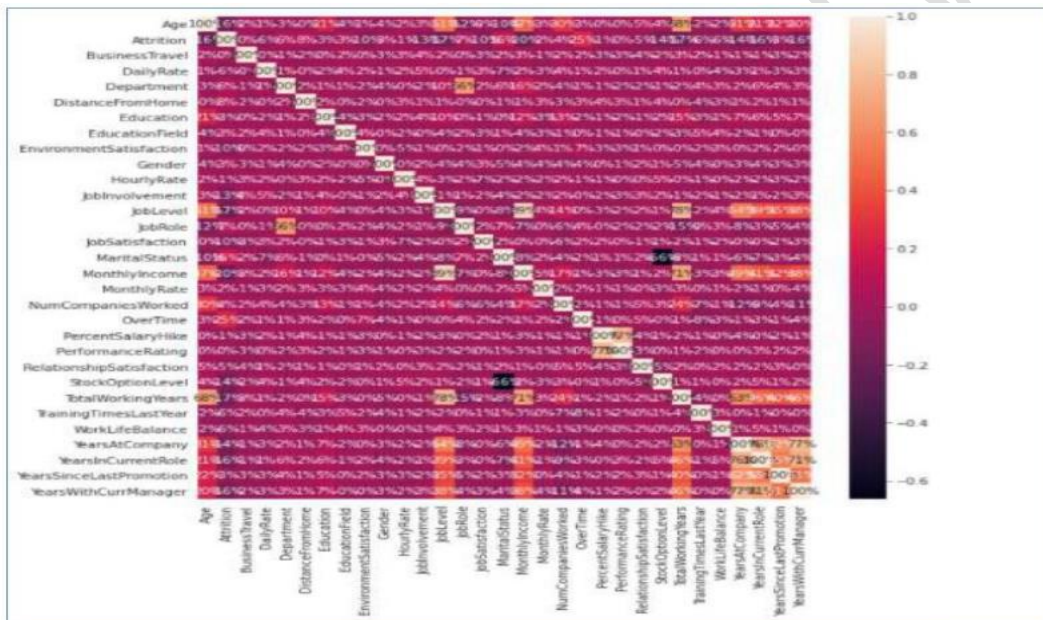


Figure 4: Correlation Matrix

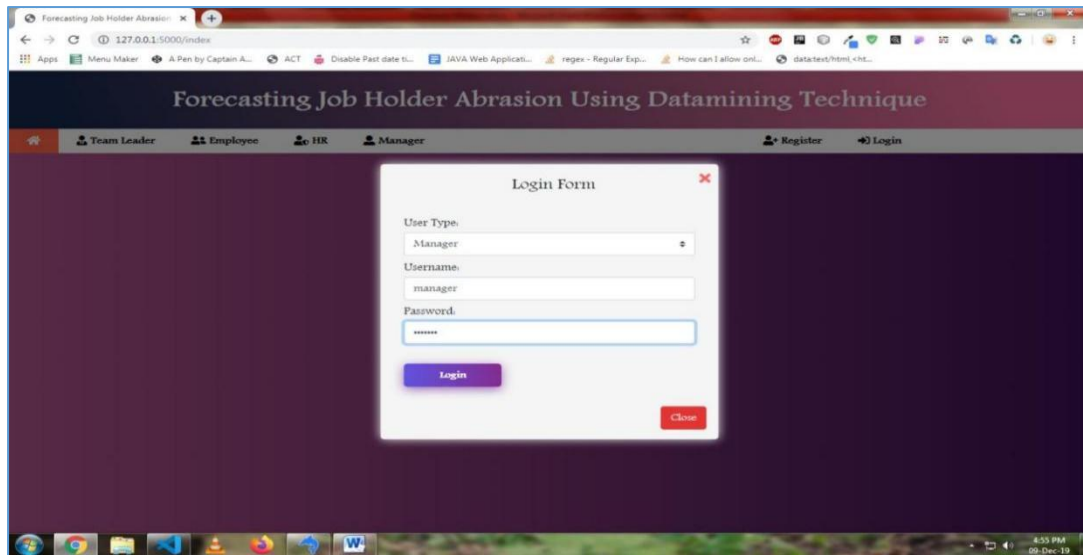


Figure 5. Login page

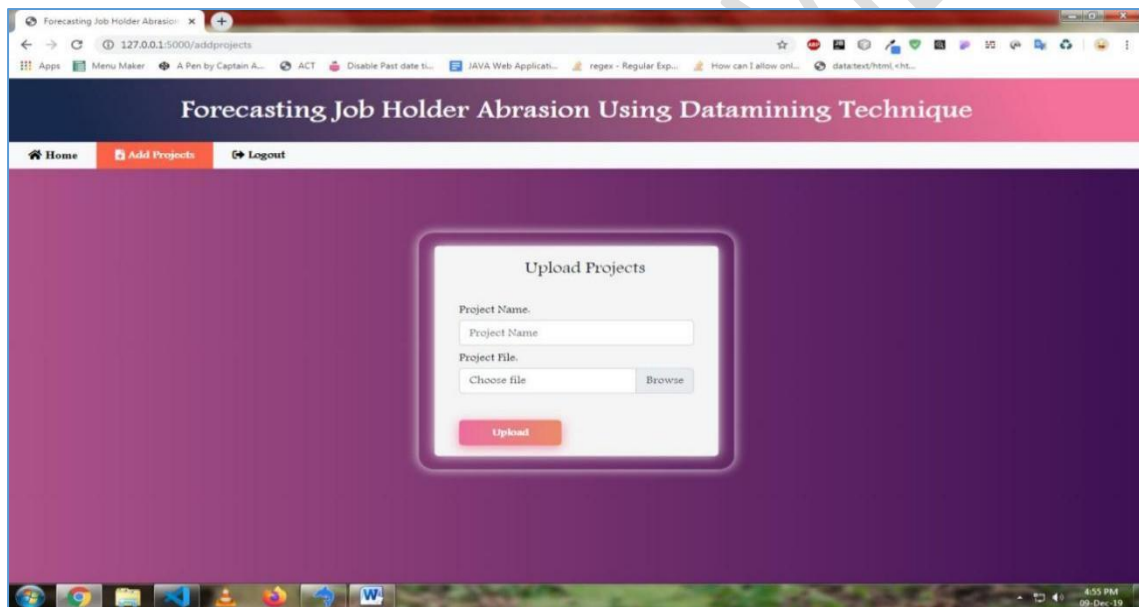


Figure 6. Given credentials

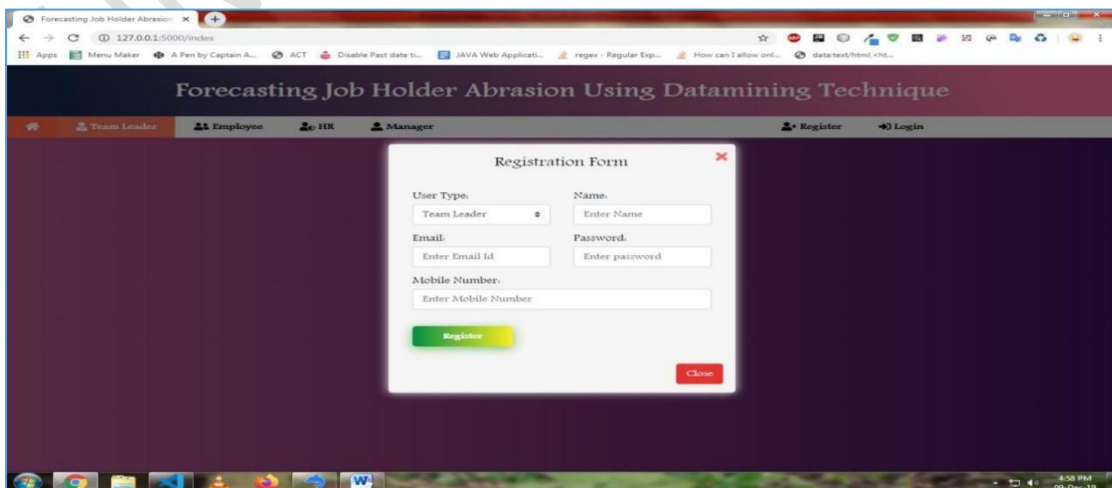
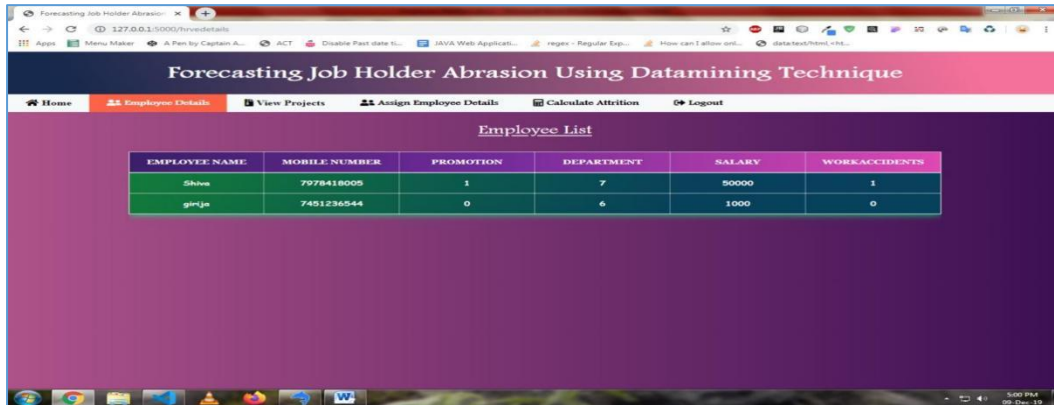


Figure 7. Fill Registration form



EMPLOYEE NAME	MOBILE NUMBER	PROMOTION	DEPARTMENT	SALARY	WORKACCIDENTS
Shiva	7978418005	1	7	50000	1
gmljn	7451236544	0	6	1000	0

Figure 8. Employee list

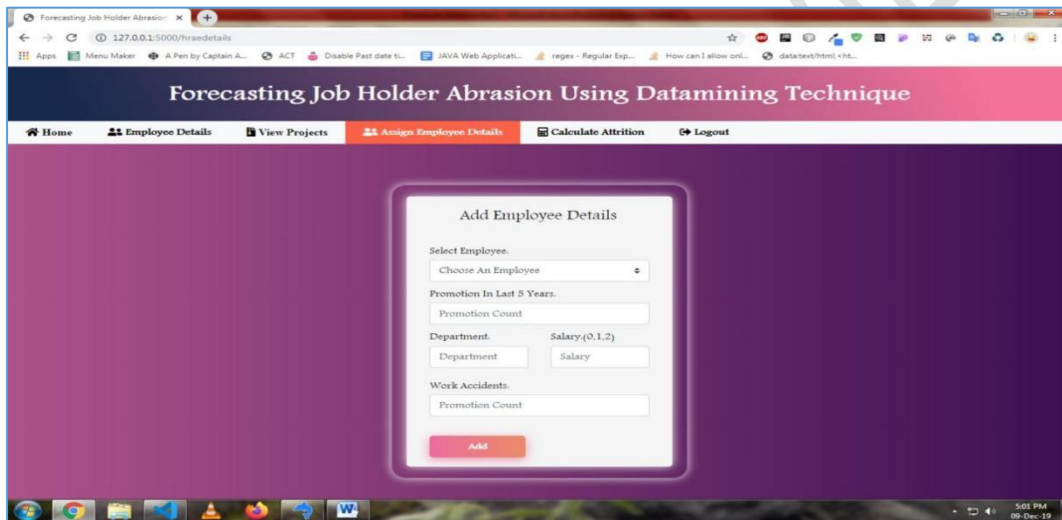


Figure 9. Adding Employee details to the form

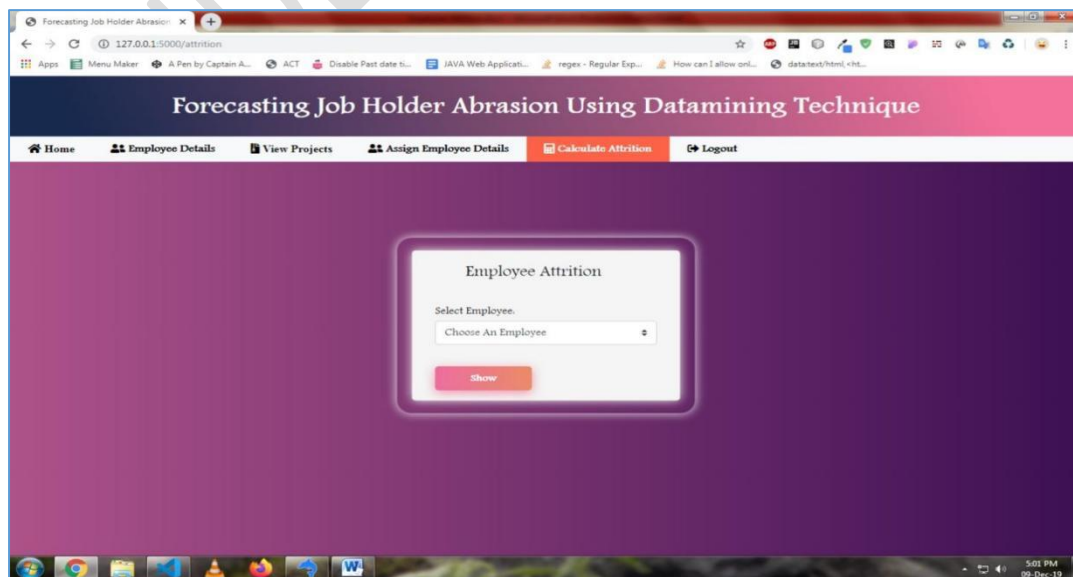


Figure 10. Employee attrition

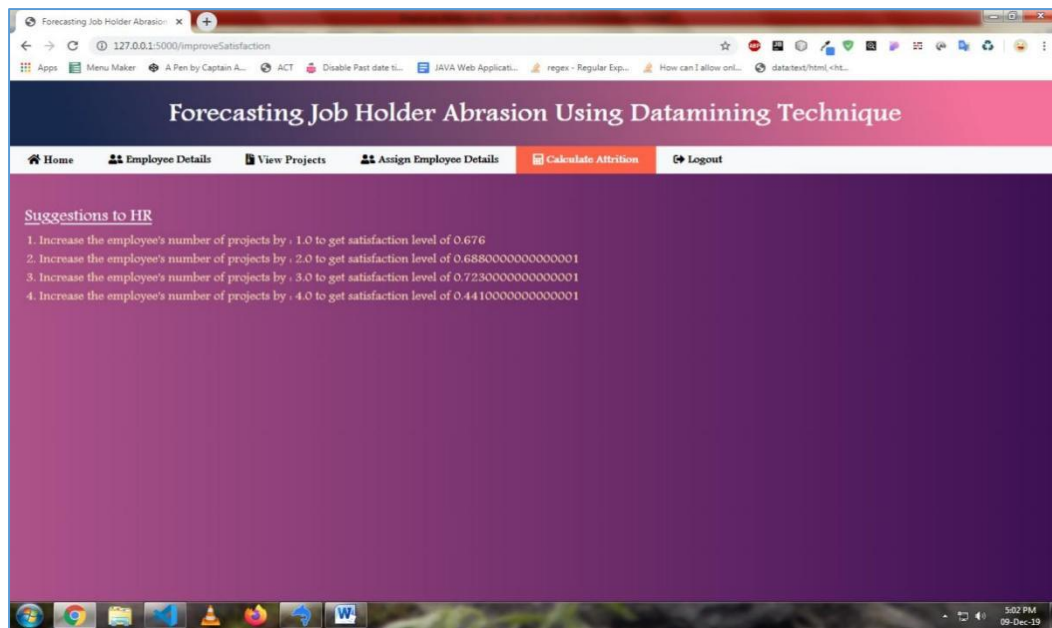


Figure 11. Suggestion to HR

5. CONCLUSION

Employee attrition poses significant challenges for organizations, leading to financial, temporal, and effort-related losses. The departure of trained and experienced personnel creates gaps that are hard to fill swiftly and economically. Recognizing its impact, organizations increasingly rely on analyzing past and current employee data to predict future attrition and understand turnover drivers. Through data extraction algorithms, accurate predictive models for attrition emerge, aiding in proactive retention strategies. This process extends beyond merely distinguishing potential leavers from stayers. Utilizing advanced data analysis techniques enables organizations to assign attrition probabilities to individual employees, facilitating targeted retention efforts and fostering a more stable workforce.

Future enhancements for the system entail continual adaptation to evolving user needs and technological advancements. Upgrades will ensure adaptability to changing environments and emerging technologies. Enhanced security measures, such as implementing single sign-on solutions, will address future security concerns effectively. Regular updates and advancements will maintain the system's relevance and functionality in an ever-changing landscape, ensuring optimal performance and user satisfaction.

REFERENCES

- [1] S. Kaur and R. Vijay, "Job Satisfaction – A Major Factor Behind Attrition or Retention in Retail Industry", *Imperial Journal of Interdisciplinary Research*, vol. 2, no. 8, 2016.
- [2] Subhani Shaik, "Analysis of Photo Plethysmography Signals with Artificial Neural Networks Using Curvelet Transform", *Journal of Adv Research in Dynamical & Control Systems*, Vol. 11, No. 1, 2019.
- [3] Shiva Keertan J and Subhani Shaik, "Machine Learning Algorithms for Oil Price Prediction", *International Journal of Innovative Technology and Exploring Engineering*, Volume-8 Issue-8, 2019.
- [4] KP. Surya Teja, Vigneswar Reddy, and Subhani Shaik, "Flight Delay Prediction Using Machine Learning Algorithm XGBoost", *Jour of Adv Research in Dynamical & Control Systems*, Vol. 11, No. 5, 2019.
- [5] Dr. Sunil Butada and Subhani Shaik, "IPL Match Prediction using Machine Learning", *IJAST*, Vol.29, Issue 5, April-2020.
- [6] Dr. R. Vijaya Kumar Reddy, Dr. Shaik Subhani, Dr. G. Rajesh Chandra, Dr. B. Srinivasa Rao, "Breast Cancer Prediction using Classification Techniques", *International Journal of Emerging Trends in Engineering Research*, Vol. 8, No.9,2020.
- [7] Mr. Sujan Reddy, Ms. Renu Sri and Subhani Shaik, "Sentimental Analysis using Logistic Regression", *International Journal of Engineering Research and Applications (IJERA)*, Vol.11, Series-2, July-2021.
- [8] Ms. Mamatha, Srinivasa Datta and Subhani Shaik, "Fake Profile Identification using Machine Learning Algorithms", *International Journal of Engineering Research and Applications (IJERA)*, Vol.11, Series-2, July-2021.
- [9] R. Vijaya Kumar Reddy, Subhani Shaik, B. Srinivasa Rao, "Machine learning based outlier detection for medical data" *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 24, No. 1, October 2021.
- [10] Dr. Vijayalakshmi K and Dr. Subhani Shaik, "Predicting employee attrition methodologies of K-fold technique" *I. J. Mathematical Sciences and Computing*, March 2023, 1, 23-36.
- [11] Subhani Shaik, P. Santhosh Kumar S. Vikram Reddy K. Sai Srinivas Reddy, and Sunil Bhutada, "Machine Learning based Employee Attrition Predicting", *Asian Journal of Research in Computer Science*, Volume 15, Issue 3, Page 34-39, March 2023.
- [12] Zhang, Y., & Li, J. (2023). "Utilizing Gradient Boosting for Employee Attrition Prediction: An Empirical Study." *Data Science and Management Review*, 5(3), 200-210.
- [13] Fernandes, S., & Pinto, A. (2023). "Predicting Employee Turnover with Advanced Machine Learning: A Gradient Boosting Perspective." *Journal of Organizational Behavior*, 44(5), 789-805.