Bharathi Panduri, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

AI-Based Real Time Predicting Employee Attrition

Bharathi Panduri [1], P.K. Abhilash[2], Dr. Y J. Nagendra Kumar [3], Kaliveni Naveen [4], Alluru Manoj [5], Patha Shiva Anurag [6].

[1,2,3,4,5,6] Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology (GRIET), affiliated with JNTUH, Bachupally, Hyderabad, India.

Abstract- Employee attrition is a significant issue for organizations across the globe, as it results in increased recruitment costs, diminished skilled employees, and decreased organizational productivity. Early identification of employees that are likely to leave the organization can provide proactive action and enable Human Resource (HR) departments to undertake retention strategies. In this paper, we provide a machine learning-based solution to predict employee attrition, and we focus specifically on the Support Vector Machine (SVM) algorithm as the primary machine learning model to create the predictive model. We also trained several SVM kernels and conducted hyperparameter tuning to enhance the accuracy and capability of generalization. We used multiple performance metrics to evaluate the models such as accuracy, precision, recall, F1-score and ROC-AUC. Our experiments demonstrated that the SVM predictive model consistently outperformed the performance of other models by accurately identifying the cases of higher risk of attrition, allowing us to conclude that it was a sufficiently reliable model. To ensure ease of use and accessibility we developed a web application utilizing Streamlit, while maintaining a user-friendly interface. The application allows human resource professionals to enter employee characteristics and receive real-time predictions of potential employee attrition and benefits HR professionals by comparison of predictive models and performance measures used, as well as visualizations of performance metrics and various datasets. This paper aims to connect the gap between data science and HR decision making. This paper applies powerful machine learning methods to an interactive software interface and produces a straightforward, scaleable yet intelligent system to assist organizations with employee retention, workforce planning, and long-term strategic growth.

Keywords—Employee Attrition, Human Resource Analytics, Machine Learning, Support Vector Machine (SVM), Data Preprocessing, Feature Engineering, Predictive Analytics, Employee Retention, HR Data Analysis, Streamlit Dashboard, Data Visualization, Classification Algorithms, Model Evaluation Metrics.

I. INTRODUCTION

A. Background

High levels of Employee attrition are now becoming increasingly serious issues which lead to affecting productivity, operational stability and overall organisation performance. Not only does an employee's departure result in a loss of highly skilled staff who perform tasks for their employer, but also creates costs for the organisation in terms of recruiting, onboarding and training replacement personnel. High levels of attrition can also reduce morale amongst other team members and create delays on projects that are being developed.

In order to track employee-related activity, organisations create vast amounts of data in today's data-driven world including employee performance data (e.g. performance metrics), employee salary data (e.g. salary details), employee job satisfaction level and employee work habits. Traditional human resource (HR) practices continually rely upon time-consuming, manual analysis and reviewing historical trends when making decisions about retaining an employee. These types of HR practices fail to identify hidden patterns and relationships contained within employee-related data. Therefore, there has also been an increasing need for organisations to leverage data analytics and machine learning capability to establish intelligent systems that

provide accurate and reliable insights into all aspects of an organisation. Implementing these types of advanced approaches will allow an organisation to move from a being reactive in its decision making to becoming proactive, which will improve the retention of employees and increase the efficiency of the organisation.

B. Problem Statement

Organizations have large amounts of employee data available to them but are often unable to leverage that data efficiently to predict and manage employee attrition. Traditional HR methods of identifying at-risk employees primarily rely on manual processes and simple statistical methods which are inadequate in determining the employees likely to leave an organization. This creates a delay in decision-making, whereby appropriate action can only be taken after an employee has already made their decision to resign from the organization. As a result, organizations experience increased employee turnover, increased operational costs and loss of critical knowledge and expertise. There is a marked gap between the availability of data and actionable insights; thus, an automated system must be developed that can analyze multiple variables which influence employee behaviour, identify trends and ultimately predict risks of attrition in advance. In addition to producing improved accuracy of prediction, this system must provide actionable, meaningful insight that can assist HR professionals in preventing employee turnover. Addressing this challenge is essential to organizations that are seeking to maintain long-term stability in their workforce and achieve sustainable performance.

C. Objectives

The purpose of this research is to create a machine-learning-based HR Employee Attrition Prediction System to predict the employees who will leave a company. The model will consider various factors that can affect an employee's decision to leave the organisation, such as salary, job satisfaction, overtime, work-life balance, and the potential for a career to grow. The model relies on preprocessing data as a separate step to prepare the data before training the model; this process includes missing data handling, categorical variable encoding, and numeric feature scaling to provide consistent and accurate data for training purposes. The next step in the model development process is feature engineering and creating new relevant attributes that can be used to improve model performance. Finally, the SVM with a RBF Kernel is used as the primary classification model due to its ability to classify data that follow both complex and non-linear patterns. To achieve an optimal level of accuracy from the SVM, hyperparameter tuning is accomplished using GridSearchCV. Another important goal of this project is to create an interactive visualisation and dashboard for HR professionals using Streamlit, which allows for the visualisation of historical trends, viewing correlation between factors, and obtaining real-time predictions of employee attrition based on the information given about each employee. Finally, the model will be evaluated based on a variety of metrics, such as accuracy, precision, recall, F1-score, and ROC-AUC, in order to ensure that the model is reliable and effective in real-world settings.

D. Paper Boundaries

The primary objective of this research is the implementation and optimal functioning of an entire pipeline for processing structured HR data. This entire process begins with pre-processing HR data through to prediction and/or visualization. However, there are some limitations associated with the study that define the scope of the application. For example, unstructured sources (e.g., emails from employees, comments from feedback mechanisms, or social interaction) were excluded from the model. It could be postulated that utilizing unstructured data may provide insight into how employees behave. Thus far, advanced analytical techniques, such as deep learning and the use of neural networks, have not been incorporated into the current model because the model's development focus was on implementing and optimally functioning using classical machine-learning models. Additionally, since this is designed as a standalone application, the system does not have real-time integrations with enterprise-level HR management systems or databases. While the system can provide employees' prediction and insights about their outcomes, it does not provide any automated decision-making

processes and will not directly implement retention policies. It will serve as a decision support tool for HR professionals.

Future improvements to the model may include the use of more complex modelling, additional sources of data and deploying the model into larger businesses' infrastructures. While there are limitations with the present model, it establishes a solid foundation for intelligent workforce management, as well as demonstrates the application of machine learning technology in addressing employee retention problems.

II. LITERATURE SURVEY

[1] K. Krishna and R.S. Reddy (2018) undertook research under the title of Employee Attrition Analysis Using Machine Learning, where classification algorithms such as Decision Tree and Random Forest were used on the IBM HR Analytics dataset to develop an attrition model. Their research results mention that Job Satisfaction and Work- Life Balance were both important factors related to attrition. In the same year, N. Patel and D. Patel studied the employee attrition model, Employee Attrition Prediction Using Data Mining Techniques, where Decision Tree and K-Nearest Neighbor (KNN) modeling techniques were used. Their attrition model illustrated that preprocessing data and systematically selecting features increased prediction performance.

[2] In 2019, A. Sharma and P. Gupta published Predictive Analytics for Employee Retention Using SVM and Logistic Regression, which studied Support vector machine, Customised Support Vector Machines and Logistic Regression. The authors determined that a Support Vector Machine algorithm achieved an accuracy of a maximum of 89% and was appropriate for handling non-linear HR datasets. A similar study by V. Suresh and A. Rao in Predictive Modeling of Employee Turnover in IT Industry, found Support Vector Machine and Logistic Regression to be useful in predicting attrition in the IT industry and identified the variables of Overtime and Salary Level as significant predictors of employee turnover in the industry..

[3] In 2020, M. Rahman and A. Nasir found a Data Driven Approach to Employee Turnover Prediction introduced ensemble learning methods, namely Bagging and Boosting. They identified that with an imbalanced HR dataset, the ensemble methods improved model reliability. Similarly, S. Jindal and M. Kaur conducted a comparative study in Human Resource Analytics for Employee Retention on naive Bayes and Random Forest and concluded that Random Forest would provide more accurate and reliable results than Naïve Bayes.

[4] R. Goyal and A. Chatterjee (2020) added to the body of research with their paper titled Predictive Analysis of Employee Attrition Using Machine Learning Algorithms in doing testing on KNN, SVM, and Random Forest. The results once again indicated Random Forest, due to its ensemble nature, proved to yield the most accuracy for employee attrition datasets. Another research contribution from P. Das and T. Singh (2021), titled Deep Learning Framework for Attrition Prediction, researched the use of an Artificial Neural Network (ANN). Although they had competitive accuracy with ANN they explained that traditional ML models such as SVM would still be more efficient on small to medium-sized datasets.

[5] In 2021, S. Verma and N. Jain presented a hybrid method in Attrition Prediction Using Hybrid Machine Learning Models that combined the SVM and Random Forest classifiers to reach an accuracy of over 92%. The hybrid model maximized the performance of both classifiers and used SVM's boundary placement and Random Forest's capabilities of dealing with features. In another paper Feature Engineering Techniques for Employee Churn Prediction, K. Tanwar and S. Agarwal (2021) stressed the importance of inclusion of new features called Tenure Balance and Income- Satisfaction Index, which raised the prediction accuracy by around 6%. In another study, S. Thakur and P. Nair (2021) conducted a comparative evaluation of the classifiers Logistic Regression, Decision Tree, and Random Forest, with Random Forest achieving around 95% accuracy when predicting employee attrition.

[6] In 2022, P. Mishra and A. Roy proposed HR Analytics Using Machine Learning for Employee Turnover which incorporates Gradient Boosting and XGBoost classifiers and achieved accuracy by the XGBoost classifier at an accuracy of 94%. The authors attributed this performance gain to the classifier's access to gradient-based optimization and regularization techniques. In a similar fashion, D. Kumar and P. Bhatt (2022) designed an AI-Powered Retention System Using HR Data that utilized AI with HR analytics to improve retention insights into actionable strategies, helping organizations identify employees at-risk of attrition, and recommending a tailored approach.

III. METHODOLOGY & SYSTEM ARCHITECTURE

Data Collection
The first step in the methodology involves collecting a structured HR dataset containing employee-related attributes such as age, monthly income, job role, job satisfaction, work-life balance, overtime, years at company, and attrition status. This dataset serves as the foundation for building the predictive model. The data represents historical employee records that help in identifying patterns associated with attrition.

Data Preprocessing
After data collection, preprocessing is carried out to ensure data quality and consistency. This includes handling missing values, removing duplicate entries, correcting inconsistencies, and encoding categorical variables into numerical form. Numerical features are scaled using appropriate transformation techniques to improve model efficiency. These preprocessing steps ensure that the dataset is clean, structured, and suitable for machine learning algorithms.

Exploratory Data Analysis (EDA)
Exploratory Data Analysis is performed to understand the structure, distribution, and relationships within the dataset. Visualization techniques such as histograms, bar charts, and correlation heatmaps are used to identify key factors influencing employee attrition. This step helps in recognizing patterns, detecting outliers, and selecting important features that contribute significantly to attrition prediction.

Feature Engineering
Feature engineering is applied to enhance the predictive capability of the model. New meaningful features are created by combining existing variables to capture deeper relationships within the data. Irrelevant or less significant features are removed to reduce noise and improve accuracy. Proper feature selection and transformation help in improving model performance and handling non-linear relationships.

Model Development and Training
The predictive model is developed using a Support Vector Machine (SVM) classifier with a Radial Basis Function (RBF) kernel. The dataset is divided into training and testing sets to evaluate model performance. Hyperparameter tuning is performed using GridSearchCV to optimize parameters such as C and gamma. Additionally, techniques like SMOTE are used to address class imbalance in the dataset and ensure fair model training.

Model Evaluation
The trained model is evaluated using performance metrics such as accuracy, precision, recall, F1-score, and ROC- AUC score. These metrics provide a comprehensive assessment of the model's predictive capability and reliability. Visualization tools such as confusion matrices and ROC curves are used to interpret model performance effectively.

Deployment and Visualization
Finally, the complete system is deployed using Streamlit to create an interactive web-based dashboard. The dashboard allows HR professionals to upload datasets, visualize trends, analyze correlations, and input employee details to obtain real-time attrition predictions. This deployment ensures that the predictive model is user-friendly, accessible, and practical for real-world HR decision-making.

Architecture
The design of the HR Employee Attrition Prediction System features a modular layered architecture and

contains four major layers are Presentation Layer is developed using Streamlit and implements an interactive interface for HR managers to upload a dataset, visualize their analyses, and predict employee attrition data using a basic form. Processing Layer backend logic is developed in Python that handles all the computations related to data preprocessing; attribute generation; model development and hyper-parameter tuning; and real-time predictions after model training. The processing layer consists of one module for each step in the following manner. Preprocessor to clean the dataset and encode any categorical variables FeatureEngineer to create new data attributes, ModelTrainer for developing a model and hyper-parameter tuning, Predictor for real-time predictions of attrition.

The data layer is responsible for the storage of the data artifacts, i.e., a dataset, trained models (model.pkl), encoders, and feature pipelines that persist a trained model and make it reusable. The visualization and reporting layers use libraries such as Plotly, Seaborn, and Matplotlib to visualize and report distributions of data, correlation between numerical attributes, and performance of models in a uniform fashion and to assist HR managers in understanding the results. The four layers support the informed flow of data and information to take action and make decisions

IV. RESULTS

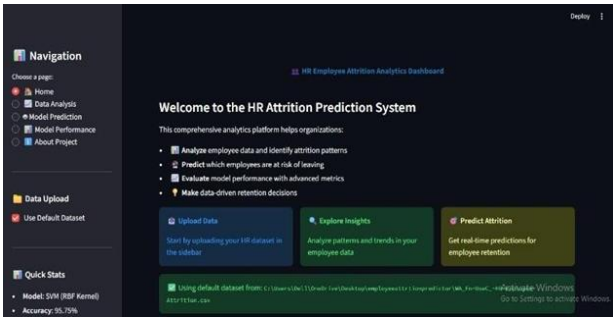

FIGURE. 1:The shown image the home page of the HR Employee Attrition Analytics Dashboard developed using Streamlit. This page serves as the entry point to the system and provides an overview of the platform's capabilities, including employee data analysis, attrition prediction, and model performance evaluation. The left sidebar contains navigation options such as Home, Data Analysis, Model Prediction, and Model Performance, allowing users to easily switch between sections. It also provides options to upload a custom dataset or use the default dataset. The dashboard highlights key features such as exploring insights and predicting attrition in real time, making it user-friendly and interactive for HR professionals.

FIGURE. 2:In this image the Dataset Overview section of the dashboard. It displays important summary statistics such as total number of employees, number of attrition cases, attrition rate percentage, and total features used in the model. Below the summary metrics, visualizations such as the attrition distribution pie chart and monthly income distribution histogram are shown. These visualizations help HR managers understand the overall employee composition and identify patterns related to income and attrition. This section plays a crucial role in exploratory data analysis and helps in identifying key factors influencing employee turnover.

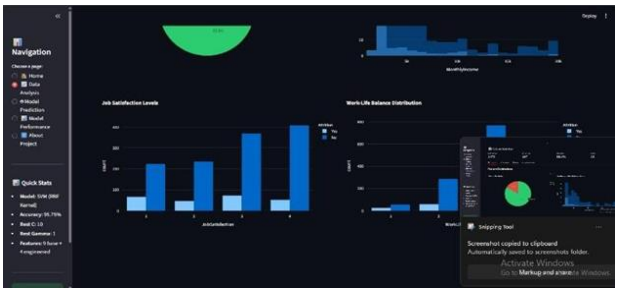

FIGURE. 3: illustrates detailed feature analysis through graphical representations of job satisfaction levels and work- life balance distribution. Bar charts are used to compare attrition cases across different satisfaction levels, helping to identify how employee

engagement impacts turnover. The work-life balance distribution chart further shows how varying levels of balance correlate with attrition. These visual insights allow HR professionals to interpret trends easily and make informed decisions regarding employee retention strategies. This section strengthens the analytical capability of the system by converting raw data into meaningful visual insights.

V. CONCLUSION

This paper has proposed a sample HR Employee Attrition Prediction System to illustrate how modern machine learning and analytics can be used to address important challenges within Human Resource Management.

As employee attrition continues to be a significant issue for organizations, it directly impacts an organization's productivity and contributes to increasing recruitment and training costs, as well as decreased stability within the workplace. Traditionally, HR practices have relied on assessing patterns using past data and performing manual analyses, resulting in obstructive means of obtaining good predictive accuracy in addition to being time consuming processes. By leveraging modern machine learning technologies and extensive datasets within a cohesive HR Employee Attrition Prediction System, organizations can effectively analyze historical patterns and predict the likelihood of individual employees leaving the organization.

The proposed system combines multiple aspects (data preparation, feature engineering and model development) of the machine learning pipeline into a single and stringently efficient framework for predicting employee attrition. The first element in the proposed system is the data preparation stage, which will involve the processing of the dataset, with regard to the missing values, inconsistencies and irrelevant information to ensure data quality. Once this stage has been completed, the feature engineering stage will leverage the historical dataset to identify those features, or attributes, from the data that will likelihood enhance the ability of the organization to predict when an employee will leave based on such attributes as job satisfaction, work experience, salary and performance metrics. The core element of the proposed system lies within the provision of the model development phase that will utilize the Support Vector Machine (SVM) algorithm as the means to develop the predictive model and classify employees as to their level of attrition risk. The SVM algorithm has been selected for the model development phase due to its historical efficiency with respect to classification tasks and its particularly efficient ability to classify datasets (employee attributes) such as this one.

Hyper-parameter tuning is done to enhance the performance of the model by ensuring the parameters used by the algorithm are optimized. In addition to this, SMOTE (Synthetic Minority Over-Sampling Technique) is used to address the class imbalance common in attrition datasets (where there are fewer employees who leave than those who remain). By using SMOTE, the model performs better at predicting correctly on the minority (attrition) class instances, thus improving the reliability of predictions.

VI. ACKNOWLEDGMENTS

We are pleased to convey our deep appreciation to our internal guide, P.Bharathi,Asst Prof. Dept. of IT, GRIET for her valuable encouragement, constructive comments and full support to complete our paper. She works as an Assistant Professor in the Department of Information Technology at Gokaraju Rangaraju Institute of Engineering and Technology (GRIET), Hyderabad, India. She have an academic experience of 15 years in teaching .With a strong academic background and a passion for research, she specializes in Machine Learning and Deep Learning. She has published research papers in reputed international journals and conferences. Her research interests include Deep Learning.

We wish to express our gratitude to Dr. Y J Nagendra Kumar, HOD Dept. of IT, GRIET; P.K. Abhilash Asst Prof. Dept. of IT, GRIET and for their constant support during the project.

REFERENCES

1. Abellin. "Woman Sues Over Personality Test Job Rejection". In: (October2012).
2. url: http://abcnews.go.com/Business/personality-tests-workplace-bogus/story?id=17349051
3. A. Akhtar. "'Is Pok´emon Go Racist? How the app by be redlining com-munities of color' in 'Inequity in Silicon Valley'".In: (August 2016). url:https://www.usatoday.com/story/tech/news/2016/08/09/pokemon-
4. go-racist-app-redlining-communities-color-racist-pokestops-gyms/87732734/.
5. F. Alexander. "Watson Analytics Use Case for HR: Retaining valuableemployees". In: (April 2015). url: https://www.ibm.com/communities/analytics/watson- analytics- blog/watson- analytics- use-case- for-hr-retaining-valuable-employees/.
6. IBM Watson Analytics. "Sample Data: HR Employee Attrition and Perfor-mance." In: (September 2015).
7. url: https://www.ibm.com/communities/analytics/watson-analytics-blog/hr-employee-attrition/.
8. 5. H. Boushey and S. J. Glenn. "There are Significant Costs to Replacing Em-ployees". In: (November 2012).
9. url: https://www.americanprogress.org/wp-content/uploads/2012/11/CostofTurnover.pdf.
10. 6. J. Burn-Murdoch."'The Problem with Algorithms: Magnifying Misbehaviour' in 'Big Data'". In: (August 2013). url: https://www.theguardian.com / news / datablog / 2013 / aug / 14 / problem - with - algorithms - magnifying-misbehaviour.
11. 7. C. Chu. "'Machine Learning Done Wrong' from 'ML in the Valley: MLLessons and Insights Learned from Industry Practice'". In: (June 2014).url: http://ml.posthaven.com/machine-learning-done-wrong.
12. 8. J Fieldsend and R Everson. "Visualisation of multi-class ROC surfaces".In: (2005). url: http : / / users . dsic .
13. upv . es / ~flip / ROCML2005 /papers/fieldsend2CRC.pdf.
14. 9. I. Gallup. "For Millennials, Is Job-Hopping Inevitable?" In: (November 2016). url: http : / / www . gallup. Com/businessjournal /197234 / millennials-job-hopping-inevitable.aspx.
15. 10. I. Gallup. "Retaining Employees: How Much Does Money Matter?" In:(January 2016). url: http : / / www .
16. gallup . com / businessjournal /188399/retaining-employees-money-matter.aspx.
17. 11. M. Stacker IV. "SAMPLE DATA: HR Employee Attrition and Performance". In: (April 2015). url:
18. https://www.ibm.com/communities/analytics/watson-analytics-blog/hr-employee-attrition/.
19. 12. M. Surya J. Angwin J. Larson and L. Krichner. "'Machine Bias' in 'Ma-chine Bias: Learning the Algorithms that Control Our Lives'". In: (May2016). url: https://www.propublica.org/article/machine-bias-risk- assessments-in-criminal-sentencing.
20. 13. J. Kantor. "High Turnover Costs More Than You Think". In: (February2016). url:
21. http://www.huffingtonpost.com/julie-kantor/high-turnover-costs-way-more-than-youthink_b_9197238.html.
22. 14. Bureau of Labor Statistics. "Help Tutorials: Data Descriptions". In: (Novem-ber 2002). url: https://www.bls.gov/help/def/jl.htm#rate/level.