

Predicting Employee Attrition Using Behavioural Data

Ashwini Pijdurkar¹, Aditya Dadasaheb Jathar², Prayag Girish Kad³, Anzer Kaleeer Mohd. Ali Khan⁴, Mohd.Murtuza Mohd.Mustafa⁵

Computer Engineering, Zeal college of engineering and research,
Pune

ABSTRACT

Employee attrition is one of the most important challenges for organizations across industries, as the unexpected departure of competent personnel results in considerable financial expenses, productivity loss, and disruption to team dynamics. Traditional HR processes frequently fail to identify at-risk employees before they decide to leave, necessitating the adoption of data-driven approaches capable of detecting early warning indications of attrition.

By examining behavioral and job-related characteristics obtained from HR datasets, this research suggests a machine learning-based approach for forecasting employee attrition. To find the best model for attrition prediction, a number of supervised classification algorithms are assessed and contrasted, including Random Forest, XGBoost, Decision Tree, Logistic Regression, and Support Vector Machine. During preprocessing, the Synthetic Minority Over-sampling Technique (SMOTE) is used to alleviate the intrinsic class imbalance found in HR attrition statistics. Explainable AI approaches, namely SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations), are used to understand the model's conclusions and emphasize the primary variables causing attrition, going beyond simple prediction.

Looking at the data, it shows that factors like working extra hours, how much money someone makes each month, how happy they are with their job, how satisfied they are with their work environment, and how often they have to travel for work all have a big impact on why an employee might decide to leave. The experimental results show that the Random Forest classifier has the best ability to predict outcomes. Additionally, using SHAP for explanation gives useful information that can help HR teams create focused plans to keep employees. This work tries to connect the difference between hidden prediction methods and clear, explainable HR analysis, helping companies take early steps to keep their workforce stable. Employee Attrition, Machine Learning, Random Forest, XGBoost, SHAP, LIME, Explainable AI, HR Analytics, SMOTE, Employee Retention, Predictive Analytics, Workforce Management.

Keywords : *Employee Attrition, Machine Learning, Random Forest, XGBoost, SHAP, LIME, Explainable AI, HR Analytics, SMOTE, Employee Retention, Predictive Analytics, Workforce Management.*

I. INTRODUCTION

In today's fast-changing and competitive business world, keeping good employees and maintaining a steady workforce are very important for a company's future success and ability to grow in the market. The workers are now considered the most valuable part of any business, as they play a key role in

making operations run smoothly, coming up with new ideas, and getting tasks done effectively. When very skilled or organized workers choose to leave a company on their own, it causes big problems for the business in terms of money and how it runs. This operational stress involves high costs for hiring, extra expenses for training, lower spirits among the team, and a quick fall in work output, so losing staff has become a big issue that modern human resource management needs to address right away.

In the past, companies have used manual checks, interviews after someone leaves, or yearly employee satisfaction surveys to figure out which workers might leave. This method of waiting for things to happen before taking

action causes big problems, especially since it depends on information that comes after the fact. Managements usually have to gather, sort, and understand how employees feel by relying on their own reports, which often don't show real-time stress or slow-growing lack of interest.

To deal with these limitations, today's workforce analytics uses predictive modeling. Many different machine learning tools have been created and used effectively to identify risks of employee turnover by looking at structural factors. There is still a big problem in how these models are created. Most current ways of analyzing things depend mostly on fixed information like age, how much someone earns each month, how far they live from work, or how long they've been with the company.

These basic measures give background information, but they don't keep up with the changing daily behaviors that show burnout, professional loneliness, or sudden disconnection. Many advanced predictive tools used by businesses work like complicated "black boxes." These tools might label someone as likely to leave the company but don't provide clear reasons to managers. This makes it hard for HR leaders to create specific plans to keep that person motivated and loyal.

This paper's research focuses on these gaps by presenting a system that works in real time and predicts employee turnover using changing behavior data. Instead of just looking at static demographic information, the suggested approach collects quiet, unnoticed digital clues from everyday work tools. These clues include how often people communicate, how regularly they log into platforms, and how frequently they interact with others in collaboration networks. They should show their talent before they leave the door.

Through the use of Deep Learning architectures in conjunction with sophisticated machine learning classifiers like Random Forest and XGBoost, the system detects minute irregularities in day-to-day operations that indicate disengagement.

Finally, this study presents a user-friendly, scalable, and highly secure blueprint for workforce optimization. By focusing on behavioral indicators and transparent AI, the suggested solution greatly minimizes standard HR methodologies' blind spots. It improves operational efficiency, maintains business culture, and provides leadership with a consistent opportunity to re-engage.

important individuals before they leave

The cost of this employee turnover goes much further than just the obvious expenses like hiring fees, job ads, and the costs of getting new people up to speed. In reality, the real cost of employee turnover is like an iceberg, with the biggest problems staying hidden under the surface of everyday work. When an experienced employee leaves, the company faces a big problem because important skills, special ways of solving problems, and key knowledge that the person had are no longer there. This sudden loss of support makes the other team members take on a much bigger daily workload, which quickly lowers their mood, makes them more tired, and raises stress throughout the whole department. If not addressed, this local stress on operations often leads to a "turnover contagion"—a chain reaction of emotions where the departure of an important person causes instability in the nearby team, leading to more people leaving on their own, which can seriously harm whole departments.. Information about final paper submission is available from the conference website.

II. Background And Motivation

There has been a dramatic change in workforce analytics, from straightforward descriptive metrics to complex prediction frameworks powered by artificial intelligence. The study of employee attrition used to be mostly a retrospective process that looked at prior turnover data to identify the causes of talent leaving an organization. However, scholars began to conceptualize attrition as a difficult supervised classification problem when predictive data science was integrated into HR frameworks. Over the past 10 years, several machine learning classifiers have been extensively evaluated to identify latent flight-risk signals in human datasets, including Support Vector Machines, Logistic Regression, Random Forests, and Gradient Boosting ensembles.

The fundamental impetus for this research arises from the structural shift of the modern corporate environment into highly digitalized ecosystems. Because current processes are heavily mediated by workplace software applications, project collaboration hubs, and communication networks, employees unintentionally leave behind a steady stream of passive, non-intrusive digital footprints. This continuous metadata presents an unprecedented chance to examine behavioral data in real time. Rather than relying on slow-moving contractual metrics or infrequent self-reported satisfaction surveys, organizations can use the changing velocity of daily workspace interactions such as communication response latencies, system login frequencies, and shifting collaborative network densities to determine an employee's exact trajectory of engagement. Shifting focus to behavioral data gives a more realistic approach to retention, measuring active detachment as it occurs.

In the past, companies have used exit interviews and surveys after employees leave to find out the reasons behind their departure. These methods give some understanding, but they are mainly reactive. They collect information only after the harm has already happened. By the time an employee decides to leave, the company has already lost the chance to fix the real problems and possibly keep that person. This reactive approach means HR teams can't plan ahead, use resources wisely, or take early steps to stop important employees from leaving. There is therefore a growing and urgent need to stop using old HR methods that rely on past experiences and instead use new approaches that use data to find out which employees might leave before they actually decide to quit.

The growth of HR analytics has created new opportunities in this area. In the last ten years, companies have started collecting a lot of information about their employees. This data includes things like age, gender, and marital status. It also covers job details like the role they hold, the department they work in, how much they earn each month, how satisfied they are with their job, and how long they have been with the company. There's also data on how they behave at work, such as how many extra hours they work, how often they travel for business, and whether they take part in training. Additionally, there is information about their relationships with managers and coworkers, as well as how they rate their work-life balance. This large amount of data, when looked at carefully, can show patterns and trends that aren't obvious at first glance, helping organizations not only figure out which employees might leave but also understand the reasons behind their departure and what steps can be taken to keep them.

Studies have consistently shown that certain characteristics are important drivers of employee turnover in a variety of organizational environments. Overtime is often found to be one of the best indicators since workers who are frequently obliged to go above their contracted hours are more likely to get dissatisfied and burn out, which increases their likelihood of looking for a job elsewhere. Research indicates that employees in lower wage bands, especially those in the early phases of their careers, have far greater turnover rates.

Monthly income and compensation levels are also among the most important criteria. Job satisfaction, environment satisfaction, and work-life balance has all been shown to play a meaningful role as well, with employees who report lower levels of satisfaction across these dimensions being considerably more likely to leave. Business travel frequency is another notable factor, with those who travel frequently showing higher attrition rates compared to employees who travel rarely or not at all.

Besides these personal reasons, the type of employee turnover also depends on which department and job position the person is in. Sales reps are often found to be among the groups at the highest risk in various studies, with some reports showing that up to forty percent of them leave their jobs. Laboratory technicians in research and development departments have a higher rate of leaving their jobs compared to other positions. Younger workers, usually those aged eighteen to thirty, are often seen leaving their jobs more frequently. This shows that people starting their careers tend to look for several options before staying in one job for a long time. These patterns show that employee turnover isn't the same everywhere in a company. It depends a lot on the individual and their specific work situation. Different people in different settings might leave for different reasons.

Even though there's been a lot of research on predicting employee turnover, a big problem has continued in most of the studies done so far. Most studies focus on creating models that predict if an employee will leave or stay, but they don't explain why the model made that decision. The way many machine learning models work, without showing how they make decisions, creates a big problem in HR. People making decisions need to know why a prediction was made so they can use it properly. An HR manager who is just told that an employee has a 70% chance of leaving within six months doesn't have enough information to create a focused plan to keep them without knowing exactly what factors are causing that prediction. The difference between what AI can predict and what it can explain has led to more focus on Explainable Artificial Intelligence, or XAI, as a helpful way to go along with regular machine learning in human resources.

Techniques such as SHAP and LIME have emerged as powerful tools for making machine learning predictions interpretable. SHAP, which is grounded in cooperative game theory, assigns a Shapley value to each feature for every individual prediction, quantifying the exact contribution of each variable to the model's output. This allows HR professionals to look at a specific employee's attrition risk and understand precisely which factors, whether it is their overtime status, their job satisfaction score, or their distance from home, are contributing most to that risk. LIME, on the other hand, builds a locally faithful interpretable model around each individual prediction, providing a different but complementary form of explanation. Together, these techniques transform attrition prediction from a black-box exercise into a transparent, actionable intelligence tool that HR departments can genuinely use to guide their decisions.

The motivation for this work comes straight from the gaps and chances we noticed. Many studies have looked at either how well machine learning models can predict outcomes or how easy it is to understand those models using explainable AI techniques. However, there is still a need for a single approach that handles both aspects well. This approach should also deal with real-world problems found in HR data, like the fact that not many employees leave the company compared to those who stay. The IBM HR Analytics Employee Attrition dataset is commonly used by researchers and includes thirty-five attributes for over fourteen hundred employees, offering a detailed and familiar base for building and testing such a framework. This research aims to provide a practical and strong solution to a major challenge in today's human resource management by combining the latest machine learning models, advanced data preparation methods like SMOTE to deal with imbalanced data, and explainable AI techniques to make predictions easier to understand.

Studies have consistently shown that certain characteristics are important drivers of employee turnover in a variety of organizational environments. Overtime is often found to be one of the best indicators since workers who are frequently obliged to go above their contracted hours are more likely to get dissatisfied and burn out, which increases their likelihood of looking for a job elsewhere. In some areas, including IT, consulting, and sales, where long work hours are frequently accepted, the correlation between excessive overtime and turnover is very strong. Research indicates that employees in lower wage bands, especially those in the early phases of their careers, have far greater turnover rates. Monthly income and compensation levels are also among the most important criteria.

The reason for this work comes straight from the gaps and chances we noticed. Many studies have looked at either how well machine learning models can predict outcomes or how easy it is to understand those models using explainable AI methods. However, there is still a need for a single approach that handles both aspects well. This is especially important because real-world HR data often has challenges, like having more data on employees who stay than those who leave. The IBM HR Analytics Employee Attrition dataset is commonly used by researchers and includes thirty-five characteristics for over fourteen hundred employees, offering a detailed and well-known base for creating and testing such a framework.

PROBLEM STATEMENT

When top performers quit on their own without warning, it causes big problems for how modern companies run and manage money. Companies face big financial losses when important job roles become empty. This happens because it's costly to replace those positions, it takes a lot of money to train new people, and the work efficiency goes down a lot during that time. The main problem is that traditional ways of tracking employees are mostly reactive, using feedback that comes after the fact, like surveys done after things have already happened or interviews when someone is leaving the company. By the time company leaders realize an employee wants to leave, the

person is already emotionally disconnected from their job. This leaves the other workers with a much heavier workload, which can lead to more people quitting in different parts of the company.

This analytical problem becomes much harder because there is a mismatch in the data used in current predictive studies. Most of the machine learning tools currently used in human resource analysis rely a lot on fixed administrative data like age, education level, salary ranges, and how far someone lives from the office. While these structural traits give some basic background, they don't keep up with the changing daily behaviors that show serious workplace isolation or sudden burnout. This dependence creates a risky blind spot in today's distributed or hybrid work environments, where an employee slowly losing interest shows up through small changes in how often they communicate and participate in digital teamwork, not through obvious changes in their personal background details.

Moreover, companies that try to use advanced machine learning models, like deep learning networks or complicated ensemble algorithms, quickly face the problem of the "black-box" dilemma with their algorithms. These systems create math-based flight risk numbers but hide the real reasons behind them from the people using the system. Because these models don't provide clear explanations, human resource professionals don't have a good understanding of the specific behaviors that are causing the high risk score. This lack of clear understanding stops managers from creating focused, helpful actions, making the predicted data not very useful for preventing employee turnover.

So, there's a clear need for a system that can predict things in a way that's integrated, respects privacy, and explains how it works. This system should change how we optimize the workforce by looking at behavior rather than just fixed administrative details. Companies need a clear system that can handle data about how people use workspaces in real time, use strong machine learning tools to classify this data, and find clear, useful risk factors using AI that explains its decisions. Fixing this complex issue is important to remove hidden weaknesses in how companies manage talent, keep the organization strong, and safeguard the people who help the company grow.

Objectives

The main objectives of this study are:

1. Create and build an automated system that uses machine learning to predict when employees might leave on their own, by looking at their behavior and actions.
2. To look at and create predictions about how people behave by studying their digital activity without them knowing, we focus on when they log in each day, how often they communicate, and how much they use their workspaces.

3. To spot and separate employees who might be struggling early on in their job, so leaders can get early info and take steps to help the organization.
4. To get better results in predicting employee retention than the usual methods that use fixed demographic data, slow administrative records, or surveys done after the fact.

Another important goal of this study is to better manage and predict workforce activity and retention in today's modern, spread-out business environments. By creating a system that keeps an eye on ongoing behavior signs instead of just looking at quick checks, the platform lets management follow changes in how engaged employees are and understand their work habits better. Adding these analytical tools to an automated machine learning system helps reduce big financial losses from unexpected staff departures, protects team efficiency during sudden work gaps, lowers the high costs usually spent on hiring, training, and getting new employees up to speed.

The study also aims to create a data processing system that works efficiently, stays secure, and doesn't interfere with anything else. It gathers real-time interactive data without needing any manual surveys or human input. System features such as automatic recording of login steps, tracking how often messages are sent, and a single place to see reports about flight risks are designed to improve the system's ability to classify information and make the best use of algorithms. The project aims to provide a user-friendly, accurate tracking system that works well for companies that are growing and teams that are spread out. It allows these groups to predict their workforce needs automatically, without having to use expensive, complicated tools designed for auditing.

This project aims to improve keeping valuable company employees by turning everyday digital interactions into early warning signs for potential risks. By using strong machine learning tools along with changing feature sets that come from everyday work activities—like regular login times, how often people talk across teams, and which apps they use—the new platform goes beyond what traditional HR analytics can offer. The study wants to create a model that is easy to use, works well on a large scale, and can predict things accurately. This model will help leaders spot early signs of people losing interest in their work. By doing this, it helps keep the company stable and ensures that the people who drive the business's success are protected and supported.

Literature Review

Many researchers and software tools have looked deeply into workforce analytics, machine learning models, and systems for keeping employees happy and committed. They do this to make companies run more smoothly and keep workers motivated. Automated prediction of employee turnover has become very important because companies are relying more on data to make decisions, managing teams that work from different places, and planning their use of digital tools. Researchers have often said that good predictive modeling needs more than just collecting basic info about employees. It

also requires understanding ongoing factors that affect performance, how long employees stay with the company, and how different workplace interactions influence each other.

Standard enterprise human resource systems include basic reporting tools that help track how long employees have been with the company, record personal information, and offer simple performance overview screens. These systems help companies track regular patterns in their workforce and manage large volumes of past administrative data. However, many of these traditional systems look back at past data and are mainly made for keeping general records instead of keeping track of employees losing interest as it happens in real time. Besides that, many old models need people to do surveys by hand, which makes the work of managing things take more time and makes it harder for businesses that are growing and teams that are spread out to respond quickly when needed.

Recent studies have paid a lot of attention to using supervised machine learning models to forecast when employees might leave their jobs on their own. Classifiers help a lot by finding complicated connections between different parts of large organizational data sets. Some experts believe that using complex algorithm systems can make weather predictions more accurate by analyzing data from many different sources related to people. These methods cut down on relying on managers' personal guesses and make it easier for companies today to assess risks related to their employees.

Algorithmic optimization for talent preservation has been the subject of several specialized research. An assessment pipeline was created by Mitravinda and Shetty (2022) in order to pinpoint the root reasons of employee turnover and offer structural suggestions for staff retention. Their research showed how predictive modeling may enhance organizational stability and lower manual human resource overhead. However, they continued to focus more on contractual and demographic data than on active, daily workplace behavioral surveillance. In a similar vein, Nagpal et al. (2024) described a systematic approach to HR analytics, emphasizing the function of machine learning in converting past snapshots into workforce insights that are predictive. High-precision modeling improves long-term retention methods and leadership selection, according to their research.

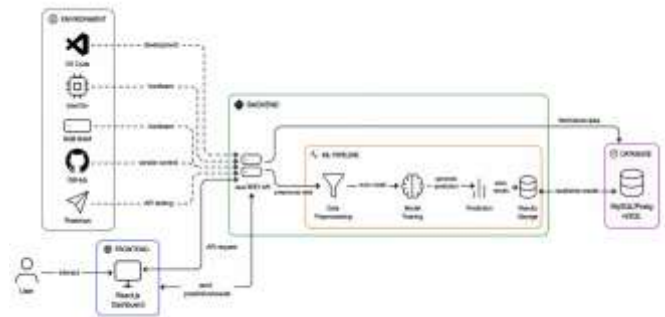
To improve classification performance, other researchers have looked at ensemble approaches and deep learning architectures. To improve labor management operations, Mellachervu et al. (2025) used deep neural networks in conjunction with sophisticated ensemble models. Predictive analytics research indicates that deep learning layers combined with gradient-boosted ensembles, such as XGBoost, can assist uncover latent interaction patterns in complicated datasets. It has been determined that centralized analytical models are essential tools for improving retention accuracy and minimizing organizational blind spots. Nevertheless, a lot of these sophisticated deep learning structures do not extend into active digital footprints like platform login intervals or communication rhythms, instead concentrating on static benchmark datasets (like IBM HR analytics profiles)

To understand changes in labor distribution, research has also looked at visual analytics and structural feature extraction. An analytical approach for assessing employee attrition from a visual standpoint was presented by John et al. (2024), showing how graphical mapping might help management identify demographic risk groups. In a similar vein, Rajeswari et al. (2022) investigated how multi-dimensional factors are processed by machine learning algorithms to predict significant spending risks related to staff departures. In order to help decision-makers filter long-term prospects, Muthugada et al. (2025) built algorithms to estimate explicit employee retention durations during recruiting. In order to map feature weights and identify the key factors impacting employee actions, Sekaran and Shanmugam (2022) assessed common classifiers such as Random Forest and Logistic Regression. Garg et al. (2024) further benchmarked traditional supervised classifiers to establish how baseline contractual and performance features can be optimized to stabilize workplace continuity.

Proposed System / Architecture System Overview

The new predictive platform is designed to fix the problems with how companies usually keep their employees. It does this by creating a single, automated system that uses machine learning to predict when employees might leave on their own.

The framework focuses on improving the user experience by converting complicated statistical trends into easier-to-understand metrics, all displayed on a single human resource command board.



System Modules

1. Client Presentation Layer (React Frontend) :

The React framework is used to build the front-end architecture as a responsive Single Page Application (SPA). This layer is solely devoted to real-time workforce monitoring, administrative accessibility, and user experience.

Role & Responsibility: It provides the main executive dashboard via which HR specialists may communicate with the system. It shows individual worker risk data, behavior abnormalities, login timing regularities, and aggregated analytics.

2. Application Core & Middleware Layer (Java Spring Boot REST API) :

Serving as the secure, central brain of the ecosystem, the backend tier is built entirely on the Java Spring Boot framework. This component acts as a highly reliable REST API middleware that governs system orchestration, access security, and transactional stability.

Role & Responsibilities: This layer oversees data formatting, session state management, enterprise security procedures, and database operations. Incoming system activity logs, including platform login sequences and communication timestamps, are processed by it as the main gateway.

3. Predictive Intelligence Layer(Python ML Engine):

The analytical processing power of the framework is housed within an isolated execution environment built entirely in Python. This layer acts as the centralized data science workshop for the entire architecture.

Role and Responsibilities: This layer comprises the data processing pipelines, automated feature engineering processes, and trained machine learning models.

4. Data Storage & Persistence Layer :

The foundational layer consists of a secure database management system that serves as the single source of truth for both the application backend and the predictive machine learning models.

Role & Responsibilities: It offers organized storage for individual behavioral metadata records, active session keys, and past categorization results.

Methodology

A highly organized, multi-phase process that incorporates backend orchestration, machine learning execution, administrative display, and passive workspace data handling is used in the creation and implementation of the suggested staff attrition prediction framework. The development of the proposed employee attrition prediction model follows a structured methodology split into simple, clear phases: data collection, backend coordination, machine learning processing, and front-end visualization.

The backend of the application is built using the Java Spring Boot framework. This layer acts as the primary middleman, managing system security, user permissions, and database operations. When an administrator requests a workforce check, the Spring Boot application grabs the raw behavior logs from the database and securely sends them over to a separate Python environment.

These prepared trends are immediately fed into pre-trained supervised machine learning algorithms, specifically Random Forest and gradient-boosted XGBoost models. The system calculates a precise flight-risk percentage, focusing entirely on active behavior patterns rather than old, unchangeable details like an employee's age, salary tier, or home address.

The resulting mathematical flight-risk probabilities are sent back to the Java backend REST API connection, translated into lightweight JSON payloads, and retrieved asynchronously by a responsive React frontend dashboard, which provides real-time visual monitoring logs to management.

Results and Discussion

The suggested employee attrition prediction system effectively illustrates the practical application of machine learning in tracking and forecasting voluntary staff turnover. The platform combines an interactive React interface, a reliable Java Spring Boot REST API backend, and a specialized Python data science engine to provide an all-in-one environment for monitoring organizational health. The technology efficiently automates critical business and human resource operations, such as tracking platform login intervals, monitoring communication cadences, and calculating live flight-risk probabilities. By shifting away from sluggish, manual review techniques and toward an automated pipeline of passive digital footprints, the framework considerably decreases administrative duties and provides firms with a solid means to detect early disengagement without disturbing day-to-day work processes.

This system's real-time, behavioral approach is one of its main advantages above conventional techniques contained in the body of workforce research. A significant amount of earlier research, including studies by Nagpal et al. in 2024 and Mitravinda and Shetty in 2022, relied almost exclusively on static snapshots of historical data, such as age, wage levels, or physical distance from the workplace. Although such demographic baselines are useful for general summaries, they are unable to capture the dynamic, changing behaviors that point to abrupt burnout or isolation in the workplace. By regularly feeding learned machine learning classifiers like Random Forest and XGBoost with real-time system interaction patterns, our technology closes this gap. The decoupled design allows the React dashboard to display instant updates smoothly, enabling human resource managers to visualize changing team engagement levels and make data-driven support decisions before an employee decides to leave.

However, there are still a number of practical and technological obstacles to overcome before such a tracking platform can be used in the real world. Managing enormous amounts of ongoing workspace metadata, ensuring lag-free inter-service Java REST API connection, and maintaining high categorization accuracy may become extremely difficult as a business expands and more people engage with the system. Additionally, in order to maintain consistency, data input formats and login tracking rules must be maintained extremely flexible due to variations in software setups across various corporate departments. Additionally, organizations need to carefully manage the line that separates employee privacy from useful analytical data. This necessitates making sure the system keeps closely monitoring metadata trends, such as communication velocity and frequency, rather than examining the workforce's private actions or the content of their private messages.

Conclusion

Using real-time behavioral data, this study successfully created an integrated, three-tier machine learning framework for predicting voluntary employee attrition. By combining a dynamic React frontend, a robust Java Spring Boot REST API backend, and an isolated Python data science engine, the system moves away from outdated, slow-moving administrative records like income and age. Instead, it processes dynamic workspace metrics specifically platform login frequencies and communication cadences to identify at-risk workers early without relying on manual, lagging surveys.

The system's analysis shows that converting fluid behavioral patterns into trained supervised classifiers such as Random Forest and gradient-boosted XGBoost models results in very accurate predictions of flight risk. Using a secure Java REST API to handle complex data science tasks helps keep the main enterprise server fast and responsive. This approach gives organizations a simple, efficient tool that lowers costs when replacing systems and helps retain important human resources.

Future studies will look into using prescriptive analytics to automatically create tailored retention suggestions whenever a high risk score is detected. Expanding the pipeline to include more metadata such as project task completion times and looking into online machine learning methods can help improve the accuracy of long-term predictions, all while keeping strict privacy-focused data anonymization rules in place.

References

- [1] Maya John, Samia Abu Theeb, Madawi Alharbi, and Norah Bin Daham. "Analyzing Employee Attrition: A Visual Perspective". 7th International Women in Data Science Conference at Prince Sultan University (WiDS PSU 2024).
- [2] Umang Garg, Neha Gupta, Mahesh Manchanda, and Kamlesh Purohit. "Classification and Prediction of Employee Attrition Rate using Machine Learning Classifiers". Proceedings of the 7th International Conference on Inventive Computation Technologies (ICICT 2024).
- [3] Pooja Nagpal, Avinash Pawar, and Sanjay H. M. "Predicting Employee Attrition through HR Analytics: A Machine Learning Approach". 4th International Conference on Innovative Practices in Technology and Management (ICIPTM 2024).
- [4] Dilina Muthgala, Samantha Mathara Arachchi. "Predicting Employee Attrition & Employee Retention Period using Supervised Learning". 6th International conference on advancements in computing (ICAC 2024).
- [5] Chandana Mellachervu, Bhargavi Madiru. "Predicting Employee Attrition with Deep Learning and Ensemble Techniques for Optimized Workforce Management".
- [6] G Raja Rajeswari, Rajkumar Murugesan, R Aruna. "Predicting Employee Attrition through Machine Learning". International Conference on Smart Electronics and Communication (ICOSEC 2022).
- [7] Kartik sekaran ,Shanmugam S "Interpreting the Factors of Employee Attrition using Explainable AID", 2022 International Conference on Decision Aid Sciences and Applications (DASA)"
- [8] Rohit, Shreya Rai, Riya Sharma, Prakhar Singh, Vijay Kumar Dwivedi, Vimal Mishra "Prediction of Employee Attrition",2024 International Association for Computer Information System.
- [9] Shikha Arora, Ashish Kr jha, Sunil Upadhyay "Predicting a Rise in Employee Attrition Rates Through the Utilization of People Analytics". 2023 International Conference on System Modeling & Advancement in Research Trends (SMART).
- [10]Devadarshan K. Sardar, Sidhant Chourasiya, and V. Vijyalakshmi. "Employee Turnover Prediction by Machine Learning Techniques". 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT 2023), IEEE.

ACKNOWLEDGEMENT

The authors wish to convey their heartfelt thanks to **Prof. Ashwini Pijdurkar**, Department of Computer Engineering, Zeal College Of Engineering and Research, Pune for their constant guidance, insightful suggestions, and unwavering support during the course of this project.