**Machine Learning based Employee Attrition Prediction and Layoff Prediction System**

**IEEE BASE PAPER TITLE:**

**Analyzing Employee Retention Factors using Machine Learning**

**IEEE BASE PAPER ABSTRACT:**

This research proposal aims to employ machine learning techniques to analyze employee retention factors in Software Companies, recognizing its crucial role in organizational success and the potential costs of high turnover rates. Through Watson Analytics' advanced analytics capabilities, the study seeks to identify key factors contributing to employee attrition and retention, culminating in the development of a predictive model using machine learning algorithms. The outcomes are expected to include actionable recommendations to improve retention strategies, insights into the relative importance of different factors, and the creation of a data-driven, proactive talent management approach for Software Companies. By empowering organizations to retain top talent and fostering a positive work environment, this research envisions a transformative impact on long-term success and performance.

**OUR PROPOSED ABSTRACT:**

Employee attrition refers to the gradual reduction of a company's workforce due to employees leaving voluntarily, such as through resignation or retirement, without immediate replacement. It can impact organizational knowledge, morale, and productivity if not properly managed. Layoffs, on the other hand, are involuntary separations initiated by the employer, often due to economic downturns, restructuring, or cost-cutting measures. Both attrition and layoffs pose significant challenges to human resource management, necessitating proactive strategies to predict and mitigate their effects.

In this project we propose a new unique concept of integrating both Employee attrition and Employee layoff. The project titled "Machine Learning-based Employee Attrition Prediction and Layoff Prediction System" aims to leverage advanced machine learning techniques to accurately predict employee attrition and potential layoffs within organizations. Developed using Python for the backend and incorporating HTML, CSS, and JavaScript for the frontend, this system utilizes the Flask web framework to ensure seamless integration and deployment. The primary goal of this project is to provide organizations with valuable insights to preemptively address employee turnover and layoffs, thereby enhancing workforce management and strategic planning.

For predicting employee attrition, we employed two distinct machine learning models. The first model, the Random Forest Classifier, achieved a perfect training accuracy of 100% and an impressive testing accuracy of 98%. The second model, the Bagging Classifier, attained a training accuracy of 99% and a testing accuracy of 95%. The dataset used for this task consists of 1,470 records, encompassing a diverse range of 35 features. These features enable the models to capture various aspects influencing employee attrition, providing a comprehensive prediction framework.

In addition to attrition prediction, the system also focuses on predicting employee layoffs using two different machine learning models. The Gradient Boosting Regressor model achieved a Training Set Mean Absolute Error (MAE) of 0.2197 and a Testing Set MAE of 1.5444. Similarly, the Random Forest Regressor model achieved a Training Set MAE of 0.5434 and a Testing Set MAE of 1.4992. The dataset utilized for layoff prediction comprises 3,612 records with 9 features. These features provide a robust foundation for understanding and predicting layoff trends across different organizational contexts.

Overall, the "Machine Learning-based Employee Attrition Prediction and Layoff Prediction System" serves as a powerful tool for organizations, enabling them to anticipate and mitigate potential workforce challenges. By integrating accurate predictive models and comprehensive datasets, this system aids in strategic decision-making, ultimately fostering better management of human resources and contributing to organizational stability and growth.

**EXISTING SYSTEM:**

* In the existing system, the emphasis was on developing a reliable model for predicting employee attrition using machine learning techniques. The primary algorithm utilized for this purpose was the Support Vector Machine (SVM), a robust method known for its effectiveness in classification problems. The SVM model was trained and tested on a dataset containing various features related to employee demographics, job roles, and workplace satisfaction.
* In the existing system, during the development and evaluation phases, the SVM model was trained to recognize subtle indicators of potential employee departures. However, the model achieved an accuracy of 47%, reflecting the inherent challenges and complexities associated with predicting human behavior in a work environment. Despite the modest accuracy, this initial system laid the groundwork for understanding the critical factors contributing to employee attrition.
* The existing system's focus on employee attrition prediction demonstrated the potential of machine learning in human resource management. By identifying key predictors of employee turnover, the system aimed to assist organizations in implementing proactive measures to retain valuable talent. The insights gained from this phase provided valuable lessons and highlighted the need for more advanced models and a broader scope in predictive capabilities, paving the way for the current project's comprehensive approach to both employee attrition and layoff prediction.

**DISADVANTAGES OF EXISTING SYSTEM:**

* While the existing system marked a significant step towards understanding employee attrition through machine learning, it had several notable disadvantages that limited its effectiveness and scope.
* Firstly, the reliance on a single algorithm, the Support Vector Machine (SVM), constrained the system's predictive accuracy and robustness. Although SVM is a powerful classification tool, its performance in this application was suboptimal, achieving an accuracy of only 47%. This relatively low accuracy indicates that the model struggled to capture the complex and multifaceted nature of employee attrition, potentially due to the algorithm's sensitivity to the choice of kernel and parameters, as well as its limitations in handling noisy or overlapping data.
* Secondly, the existing system's narrow focus on employee attrition without addressing other critical aspects of workforce dynamics, such as layoffs, restricted its utility for comprehensive human resource management. By concentrating solely on predicting voluntary employee departures, the system overlooked the broader context of involuntary separations like layoffs, which are equally important for organizational planning and strategy.
* Additionally, the system's predictive capabilities were limited by the dataset's inherent characteristics. While the dataset included a wide range of features, the SVM model may not have fully leveraged these attributes due to its linear nature (unless a kernel trick was used), potentially missing out on complex, non-linear relationships within the data. Moreover, the model's performance could have been impacted by imbalanced data, where the number of employees who left the organization was significantly different from those who stayed, leading to biased predictions.
* Lastly, the existing system lacked the flexibility and adaptability needed to cope with changing organizational dynamics and workforce trends. As businesses evolve, the factors influencing employee attrition can shift, requiring models that can quickly adapt to new patterns and data. The static nature of the SVM-based system made it less responsive to such changes, limiting its long-term applicability and effectiveness.
* Overall, while the existing system provided a valuable starting point, its limitations in accuracy, scope, flexibility, and adaptability underscored the need for a more advanced and comprehensive approach, as seen in the current project.

**PROPOSED SYSTEM:**

* The proposed system builds upon the foundations of the existing system by significantly enhancing the scope and capabilities of predictive analytics within the realm of human resource management. This system integrates two critical predictive tasks: employee attrition prediction and employee layoff prediction, providing a comprehensive tool for organizations to manage and anticipate workforce changes.
* For employee attrition prediction, the proposed system employs two advanced machine learning models: the Random Forest Classifier and the Bagging Classifier. These models are designed to capture complex patterns and interactions within the data more effectively than the previously used Support Vector Machine. The Random Forest Classifier achieved a training accuracy of 100% and a testing accuracy of 98%, while the Bagging Classifier reached a training accuracy of 99% and a testing accuracy of 95%. The dataset used for this task consists of 1,470 records and includes a wide range of features such as Age, Business Travel, Daily Rate, Department, Distance From Home, Education, Education Field, Employee Count, Employee Number, Environment Satisfaction, Gender, Hourly Rate, Job Involvement, Job Level, Job Role, Job Satisfaction, Marital Status, Monthly Income, Monthly Rate, Number Companies Worked, Over18, Over Time, Percent Salary Hike, Performance Rating, Relationship Satisfaction, Standard Hours, Stock Option Level, Total Working Years, Training Times Last Year, Work Life Balance, Years At Company, Years In Current Role, Years Since Last Promotion, and Years With Current Manager.
* In addition to attrition prediction, the proposed system incorporates a robust framework for predicting employee layoffs, employing two machine learning models: the Gradient Boosting Regressor and the Random Forest Regressor. The Gradient Boosting Regressor achieved a Training Set Mean Absolute Error (MAE) of 0.2197 and a Testing Set MAE of 1.5444. The Random Forest Regressor, on the other hand, achieved a Training Set MAE of 0.5434 and a Testing Set MAE of 1.4992. The dataset for layoff prediction comprises 3,612 records and includes features such as company, location, industry, total laid off, percentage laid off, date, stage, country, and funds raised.
* Developed using Python for the backend, with HTML, CSS, and JavaScript for the frontend, and utilizing the Flask web framework, the proposed system ensures seamless integration and deployment. By leveraging advanced machine learning models and comprehensive datasets, the system aims to provide a detailed and accurate analysis of workforce dynamics, enabling organizations to make informed decisions and strategic plans.

**ADVANTAGES OF PROPOSED SYSTEM:**

* The proposed system offers several significant advantages that enhance its effectiveness and utility in predicting employee attrition and layoffs, making it a powerful tool for human resource management.
* One of the primary advantages is the use of advanced machine learning models, which significantly improve predictive accuracy and robustness. For employee attrition prediction, the Random Forest Classifier and Bagging Classifier models provide higher accuracy rates compared to the previously used Support Vector Machine. The Random Forest Classifier achieved a training accuracy of 100% and a testing accuracy of 98%, while the Bagging Classifier achieved a training accuracy of 99% and a testing accuracy of 95%. These models can capture complex patterns and interactions within the data, leading to more reliable predictions.
* Additionally, the inclusion of employee layoff prediction in the proposed system broadens its scope, making it a more comprehensive tool for managing workforce dynamics. By utilizing models such as the Gradient Boosting Regressor and the Random Forest Regressor, the system can accurately forecast potential layoffs. The Gradient Boosting Regressor achieved a Training Set Mean Absolute Error (MAE) of 0.2197 and a Testing Set MAE of 1.5444, while the Random Forest Regressor achieved a Training Set MAE of 0.5434 and a Testing Set MAE of 1.4992. This dual focus on attrition and layoffs allows organizations to address both voluntary and involuntary employee separations effectively.
* The proposed system's integration of comprehensive datasets for both attrition and layoff prediction is another notable advantage. Furthermore, the system's development using Python for the backend, along with HTML, CSS, and JavaScript for the frontend, and the Flask web framework ensures seamless integration and user-friendly deployment. This technological stack provides a flexible and scalable platform that can be easily adapted to different organizational needs and integrated with existing HR systems.
* Overall, the proposed system's advanced predictive models, comprehensive scope, rich datasets, and robust technological infrastructure offer significant advantages over the existing system, providing organizations with a powerful tool to anticipate and manage workforce changes effectively.

**SYSTEM ARCHITECTURE:**

Random Forest Classifier.

Predicted Result: Employee Attrition Prediction

Input Employee Details Dataset

Performance Analysis and Graph

Bagging Classifier.

Gradient Boosting Regressor

Predicted Result: Number of Layoff Prediction

Input Company Details Dataset

Performance Analysis and Graph

Random Forest Regressor

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium i3 Processor.
* Hard Disk : 500 GB.
* Monitor : 15’’ LED.
* Input Devices : Keyboard, Mouse.
* Ram : 8 GB.

**SOFTWARE REQUIREMENTS:**

* Operating System : Windows 10 / 11.
* Coding Language : Python 3.12.0.
* Web Framework : Flask.
* Frontend : HTML, CSS, JavaScript.

**REFERENCE:**

Moshiur Rahman; Md Rashedul Islam; Partho Bala; Abdus Sattar, “Analyzing Employee Retention Factors using Machine Learning”, IEEE CONFERENCE, 2024. DOI: 10.1109/ICAECT60202.2024.10469009