ABESEC Ghaziabad
Department of Computer Science & Engineering

SYNOPSIS REPORT
(Session 2022-23)

Project Title: HR helping Hand: The Employee attrition and performance analysis.				
Project Type (application, product, research, review etc.)				
	Name	**Roll Number**	**Section**	**Signature**
Group member (1)	**Nishant Kumar**	**2100320100112**	**B**	
Group member (2)	**Mohd. Shahjahan**	**2100320100104**	**B**	
Group member (3)	**Rohit Sharma**	**2100320100133**	**B**	
Project Guide	**Dr. Meeta Chaudhary**			
Signature				
Date of submission				

1.1. Problem Introduction

1.1.1. Motivation

HR estimates the requirement of human resources in each part of organization and plan to recruit talented people, placing each employee's progression, employee's retention, attrition, salaries and other welfare benefits. Employees are the treasured property of any organization. However, if you quit your job unexpectedly, the company will cost a lot of money. Not only new employees are wasting money and time, but new employees are also spending time making profits for their companies.

1.1.2 Project Objective

This aims to identify the critical elements to contribute to employee attrition. Business heavily depend on employees so we here to predict or to find the probability of the employees attrition. We will use several algorithms such as logistic regression, decision tree, Gaussian Naive Bayes classifier using tenfold cross validation etc.

1.1.3 Scope of the Project

This project will help in HR related fields for analysis and decision making related to employee retention by analyzing performance of individual employee using various machine learning algorithms. It reduces paper work and save time for further recruiting process.

1.2 Related Previous Work

One of the branches of analytics is HR analytics, which is developing the system HR units in organizations function, principal to sophisticated proficiency, and improved outcomes overall. The usage of analytics by human resources for many years. Though the assortment, processing, and data analysis have been generally manual and specified the nature of HR dynamics, the approach has been constraining HR. The prospect to effort predictive analytics in categorizing the employees furthermost likely to grow promoted. Here we apply machine learning techniques to analyze the employee information for improving his/her position in the organization. Compensation and job performance information from revenue rates and personnel characteristics to payroll and service history, never before have HR executives had such liberated right to use to individual details. In this work, we are applying random forest classification, which facilitates employee classification based on their monthly income and informal way to execute analytics on data. Further, we use clustering techniques based on the performance metrics similarity to analyze employee performance.

In this paper [1] employee attrition is the loss of employees in a company caused by several factors, namely employees resigning, retiring, or other factors. This research aims to help the human resources department in the company to find out what factors influence the occurrence of employee attrition, by developing a new method of predicting attrition that this study aims to detect employee attrition in a company by implementing the Random Forest classification modeling. From the results of the tests that

have been carried out, Information Gain produces the highest accuracy value of 89.2%, while Select K Best produces an accuracy value of 87.8% and Recursive Feature Elimination produces an accuracy value of 88.8%

In paper [2] to find the probability of new employee attrition, various classification algorithms such as decision trees (DT) classifier, logistic regression (LR), random forests (RF), and K-means clustering are used. The accuracy provided by the decision tree is 97% the accuracy provided by random forest algorithm is 98% the accuracy provided by the logistic regression is 78%. This research aims to identify the most critical elements that contribute to employee attrition. Used ML algorithms for binomial classification problem are: decision trees, logistic regression, and random forest

The aim of paper [3] is to improving performance of algorithms. The usage of NB algorithm, KNN algorithm and SVM algorithm, prediction mannequin has been in contrast on the experimental groundwork and supply the end result of which algorithm is performing better. Naive Bayesian algorithms, K-nearest neighbors, and support vector machine data science techniques were applied to the predictions. Naive Bayes Algorithm predicted the developer turnover with the accuracy of 76%, K-Nearest Neighbor with the accuracy of 94% and Support Vector Machine with the accuracy of 96%

In this paper [4] there is using of various algorithms which are Gaussian Naïve Bayes Classifier, Decision Tree, Logistic Regression and Model tested on IBM analysis dataset best recall rate (0.54).It helps in HR activities optimizes and reduce critical issues by analyzing of data. The Techniques used for better error estimation are Holdout, Cross validation Scaling is used to avoid that one features is dominating

In paper [5] we study the concept of employee retention on key variables. Training, types of training and duration and check the effect of training or retention of employee satisfaction leads to employee dissatisfaction leads to high turnover. The objective is to identify the effect of training, its types and duration on retention. Main reason for employee retention are Training time and its duration, Support Interest among employee, workplace environment, Reward.

In paper [6] we study that most of the organization or companies have a formal performance evaluation system in which employee job performance is graded on a regular basis, usually once or twice a year. An Employee can improve their performance by way of monitoring the progression of their performance by way of Machine learning algorithms i.e. clustering algorithm and decision tree of data mining techniques can be used to find out the key characteristics of future prediction of an organization.

In paper [7] we study that Machine learning (ML) is scientific study of algorithms and statistical models that computer system use to perform a specific task without being explicitly programmed. Learning algorithms in many application that’s we make use of daily. There are various type of Machine learning techniques that are supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, multi-task learning, ensemble learning, neural network, instance based learning.

1.3 Software and Hardware requirements

Software used:

1) **Python**

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured, object-oriented and functional programming

2) **HTML**

HTML is a type of markup language. It encapsulates, or “marks up” data within HTML tags, which define the data and describe its purpose on the webpage.

3) **CSS**

Cascading Style Sheets is a style sheet language used for describing the presentation of a document written in a markup language such as HTML or XML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.

4) **JAVA Script**

JavaScript, often abbreviated to JS, is a programming language that is one of the core technologies of the World Wide Web, alongside HTML and CSS

Hardware used:

1) **Memory**

Computer memory is the storage space in the computer, where data is to be processed and instructions required for processing are stored. The memory is divided into large number of small parts called cells. Each location or cell has a unique address, which varies from zero to memory size minus one.

2) **RAM**

RAM is a temporary memory bank where your computer stores data it needs to retrieve quickly. RAM keeps data easily accessible so your processor can quickly find it without having to go into long-term storage to complete immediate processing tasks.

3) **CPU**

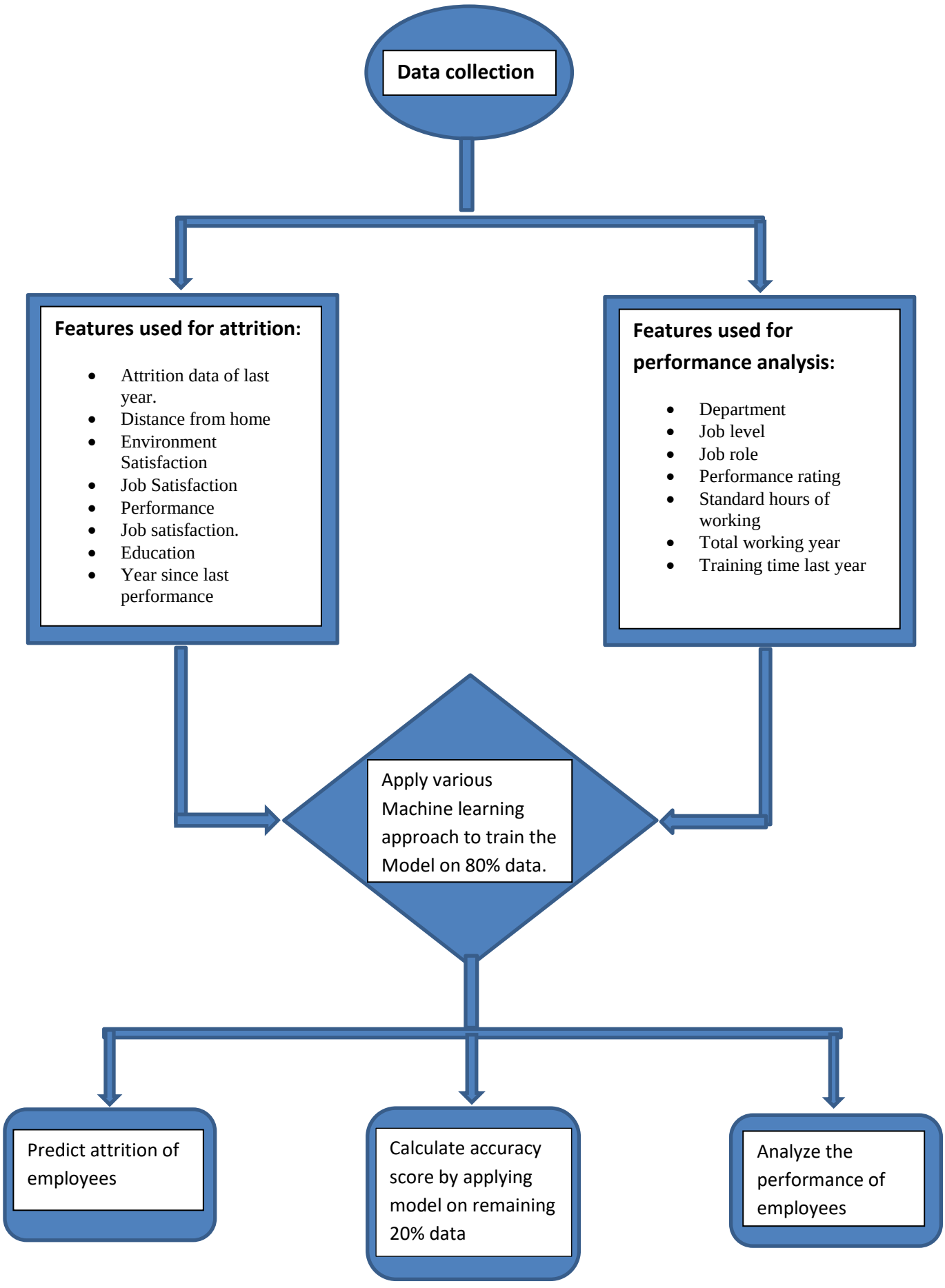
A central processing unit (CPU), also called a central processor, main processor or just processor, is the electronic circuitry that executes instructions comprising a computer program.

1.4 Proposed Method

We found that in previous model the work is dependent upon the accuracy on present time employee attrition and we are extending it by working on future aspect of company growth as well.
Then we collect the related information and data from internet. We have 30 features that are Age of the employee, Business travel, department, Distance from home, Education, Education Field, Employee Count ,Employee Number, Environment, Satisfaction gender, Job involvement, job level, Job role, Job satisfaction, Marital status, Monthly income, Number companies worked, Over18 , 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 since last promotion, years with current manager.
Then we clean the data and remove all wrong entries. Then we divide the data into two parts in 80:20 and use 80% for training and 20% data for testing
Use the following algorithms as follow:

- Random Forest classification
- Gaussian Naïve Bayes Classifier
- Decision Tree
- Logistic Regression
- K-means clustering
- Naive Bayesian algorithms

Use 20% testing data with the algorithms shown above after training with 80% data. Choose the best algorithm with highest accuracy score


Data collection
Features used for attrition:
Attrition data of last year.
Distance from home
Environment Satisfaction
Job Satisfaction
Performance
Job satisfaction.
Education
Year since last performance
Features used for performance analysis:
Department
Job level
Job role
Performance rating
Standard hours of working
Total working year
Training time last year
Apply various Machine learning approach to train the Model on 80% data.
Predict attrition of employees
Calculate accuracy score by applying model on remaining 20% data
Analyze the performance of employees

1.5 Deliverables

This project will be an advancement in the field of HR for the finding the probability of employee leaving. Although much work is done on machine learning methods, still there exist many challenges. The present methods do not achieve maximum results in attrition.

1.6 Stakeholders

Our project would be very helpful in the HR field in analysis and decision making in the company. It will help HR to initiate recruitment process for employee retention process and it can also help in company growth.

1.7 References

[1].Sindi Fatika Sari, Kemas Muslim Lhaksmana. Employee Attrition Predition Using Feature Selection with information Gain and Random Forest Classification.

[2].Elham Mohammed Thabit A. Alsaadi; Sameerah Faris Khlebus, Ashwak Alabaichi. Idenification of human resourse analytics using machine learning algorithms.

[3].Vinston Raja R; Srinath Doss; Ashok Kumar K. Analytics Approach of predicting Employee Attrition using data science techniques.

[4].Evanthia Faliagka, Kostas Ramantas, Athanasios Tsakalidis Application of Machine Learning Algorithms to an online Recruitment System.

[5].Ali A. Mahmoud, Tahini AL Shawabkeh, Walid A. Salameh, Ibrahim Al Amro Performance predicting in hiring process and performance appraisals using machine learning.

[6].Ananya Sarker, S.M. Shamim, Dr. Md. Shahiduz Zama & Md. Mustafizur Rahman Employee's Performance Analysis and prediction using K-means Clustering & Decision Tree Algorithm.

[7].Francesca Fallucchi, Marco Coladangelo, Romeo Giuliano and Ernesto William De Luca. Predicting Attrition Using Machine Learning Techniques.

[8].Sarah S. Alduayj, Kashif Rajpoot Predicting Employee Attrition Using Machine Learning.

[9].Jozsef Pap, Csaba Mako, Miklos Illessy, Zef Dedaj, Sina Ardadili, Bernat Torok and Amir Mosavi. Correlation Analysis of Factors Affecting Firm Performance and Employees Wellbeing: Application of Advanced Machine Learning Anaysis.

[10]. Batta Mahesh. Machine Learning Algorithms – A Review.

[11].Alina Kochling, Marius Claus Wehner. Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development.

[12].Ali Halawi, Nada Haydar. Effects of Training on Employees Performance: A case study of Bonjus and Khatib And Alami Companies.

[13].Osisanwo F.Y, Akinsola J.E.T, Awodele O, Hinmikaiye J.O, Olakanmi O. Akinjobi J.
Supervised Machine Learning Algoriths: Classificaton and Comparison.

[14].Thyago P. Carvalho, Fabrizzio A, A. M. N. Soares, Roberto Vista, Roberto da P. Francisco, Joao P. Basto, Symone G. S. Alcala. A systematic literature review of machine learning methods applied to predictive maintenance.

[15].V.kakulapati, Kalluri Krishna Chaitanya, Kolli Vamsi Guru Chaitanya & Ponugoti Akshay.
Predictive analysis of HR – A machine learning approach.