**Project Report**

**On**

**HR Helping Hand: The Employee Attrition and performance analysis**

**Submitted as partial fulfilment for the award of**

**BACHELOR OF TECHNOLOGY**

**DEGREE**

**Session 2022-23**

**In**

**Name of discipline**

**By**

**Nishant Kumar 2100320100112**

**Rohit Sharma 2100320100133**

**Mohd. Shahjahan 2100320100104**

**Under the guidance of**

**Dr. Meeta Chaudhry**

**ABES ENGINEERING COLLEGE, GHAZIABAD**

|  |  |  |
| --- | --- | --- |
| **D:\User Data\Desktop\nishant\Mini Project\presentation\abes logo.jpg** | APJAK logo |  |

**AFFILIATED TO**

**DR. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, U.P., LUCKNOW**

**(Formerly UPTU)**

**Project Report**

**On**

**HR Helping Hand: The Employee attrition and performance analysis**

**Submitted as partial fulfilment for the award of**

**BACHELOR OF TECHNOLOGY**

**DEGREE**

**Session 2022-23**

**In**

**Name of discipline**

**By**

**Nishant Kumar 2100320100112**

**Rohit Sharma 2100320100133**

**Mohd. Shahjahan 2100320100104**

**Under the guidance of**

**Dr. Meeta Chaudhry**

**ABES ENGINEERING COLLEGE, GHAZIABAD**

|  |  |  |
| --- | --- | --- |
| **D:\User Data\Desktop\nishant\Mini Project\presentation\abes logo.jpg** | APJAK logo |  |

**AFFILIATED TO**

**DR. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, U.P., LUCKNOW**

**(Formerly UPTU)**

**STUDENT’S DECLARATION**

We hereby declare that the work being presented in this report entitled

“HR Helping Hand: The employee attrition and performance analysis” is an authentic record of our own work carried out under the supervision of Dr. Meeta Chaudhry.

The matter embodied in this report has not been submitted by us for the award of any other degree.

**Dated:** **Signature of students**

Nishant Kumar

Rohit Sharma

Mohd. Shahjahan

**Department: CSE**

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

|  |  |
| --- | --- |
|  | Signature of Supervisor  (Dr. Meeta Chaudhry)  Associate Professor  (Computer Science & Engineering Department) |

CERTIFICATE

This is to certify that Project Report entitled “HR Helping Hand: The Employee Attrition and Performance Analysis” which is submitted by Nishant Kumar, Mohd. Shahjahan, Rohit Sharma in partial fulfilment of the requirement for the award of degree B. Tech. in Department of Computer Science & Engineering of Dr. A.P.J. Abdul Kalam Technical University, formerly Uttar Pradesh Technical University is a record of the candidate own work carried out by them under my supervision. The matter embodied in this thesis is original and has not been submitted for the award of any other degree.

**Supervisor** **Project Coordinator**  **HOD’s Signature**

**Date:**

**ACKNOWLEDGEMENT**

*It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B. Tech. Second Year. We owe special debt of gratitude to Professor Dr. Meeta Chaudhry Department of Computer Science & Engineering, ABESEC Ghaziabad for his constant support and guidance throughout the course of our work. Her sincerity, thoroughness, and perseverance have been a constant source of inspiration for us. It is only through her efforts that our endeavours have seen light of the day.*

*We also take the opportunity to acknowledge the contribution of Professor Dr. Divya Mishra, Head, Department of Computer Science & Engineering, ABESEC Ghaziabad for her full support and assistance during the development of the project.*

*We also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.*

*Signature:*

*Name*: Nishant Kumar

*Roll No.:* 2100320100112

*Date*:

*Signature:*

*Name*: Rohit Sharma

*Roll No.:* 2100320100133

*Date*:

*Signature:*

*Name*: Mohd. Shahjahan

*Roll No.:* 2100320100104

*Date*:

**ABSTRACT**

*At present, competitions between the organizations are at high. Every organization wants to the ahead of their competitors but to achieve that position they find difficulties, which is the loss of talented employees. As the employee is the main functioning body of an organization. The rate of growth of an organization or success and failure is relying on the employees and their performances. When a good/devoted employee left the organization without providing any previous information, it affects the growth of the organization, leading to increased costs, decreased productivity, and reduced competitiveness. This study is related to the finding prediction of employee attrition and their performance using machine-learning models.*

*The purpose of the model is to analyze the dataset of employees of an organization which we obtained from various employee surveys within organization and managerial reviews of employees and then calculate the probability of employee, whether the employee is leaving/resigning from the job or continue working with the same organization. Using various Machine-learning approaches, we improve the data analysis of human resource team to manage or build an excellent quality of workspace in an organization and remain in competition.*

**TABLE OF CONTENTS**

 Page

DECLARATION..................................................................................................... I

CERTIFICATE….....................................................................................................II

ACKNOWLEDGEMENTS...................................................................................... III

ABSTRACT.............................................................................................................IV

LIST OF TABLES....................................................................................................V

LIST OF FIGURES..................................................................................................VI

CHAPTER 1   INTRODUCTION

1.1 Problem Introduction

1.1.1 Motivation

1.1.2 Project Objective

1.1.3 Scope of the Project

1.2 Related Previous Work

1.3 Organization of the Report

CHAPTER 2   LITERATURE SURVEY

CHAPTER 3  SYSTEM DESIGN AND METHODOLOGY

3.1 System Design

3.1.1 System Architecture /Diagrammatical View

3.1.2 Flow Chart

3.2 Algorithm(s)

3.2.1 Random Forest Classification

3.2.2 Decision Tree

3.2.3 Logistic Regression

3.2.4 K-means clustering

3.2.5 Support Vector Machine

3.2.6 K-Nearest Neighbors

CHAPTER 4   Implementation and Results

4.1 Software and Hardware requirements

  4.2 Snapshots

CHAPTER 5 CONCLUSIONS

5.1 Performance Evaluation

5.2 Future Direction

REFERENCES... ........................................................................................  34

 APPENDIX ....................................................................................................

APPENDIX ........................................................................................................

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
|  | Name | Page No |
| Table 1 | Employee Survey Data. | 22 |
| Table 2 | Manager Survey Data | 23 |
| Table 3 | General Data | 23 |
| Table 4 | Test and Train score of different models | 29 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
|  | Name | Page No. |
| Fig. 1 | GUI (Graphical User Interface) | 15 |
| Fig. 2 | Main Window of HR Helping Hand | 16 |
| Fig. 3 | Employee Survey Form Window | 17 |
| Fig. 4 | Employee General Data Form Window | 17 |
| Fig. 5 | Manager Survey Form Window | 18 |
| Fig. 6 | Data Flow Chart | 19 |
| Fig. 7 | Python Libraries | 21 |
| Fig. 8 | Employee survey data distribution. | 22 |
| Fig. 9 | Manager Survey data distribution. | 23 |
| Fig. 10 | Distribution of employee attrition | 24 |
| Fig. 11 | Business Travel VS Attrition | 24 |
| Fig. 12 | Attrition VS Working years of employee | 25 |
| Fig. 13 | Attrition VS Training Time last year | 25 |
| Fig. 14 | Attrition VS Year at company | 26 |
| Fig. 15 | Attrition VS Year since last promotion | 26 |
| Fig. 16 | Attrition VS Year with current manager | 27 |
| Fig. 17 | Distribution of Attrition in department | 27 |
| Fig. 18 | Distribution of Attrition in different age | 28 |
| Fig. 19 | Distribution of Attrition for employee distance from home | 29 |

**LIST OF SYMBOLS**

[x] Integer value of x.

≠ Not Equal

 Belongs to

€ Euro- A Currency

\_ Optical distance

\_o Optical thickness or optical half thickness

**LIST OF ABBREVIATIONS**

AAM Active Appearance Model

ICA Independent Component Analysis

ISC Increment Sign Correlation

PCA Principal Component Analysis

ROC Receiver Operating Characteristics

**CHAPTER 1**

**INTRODUCTION**

Employee attrition refers voluntary departure of employees from organization. Attrition has a crucial impact on organizations, as it increases costs, reduces productivity, and decreases the competitiveness of organizations. Attrition also leads to a decline in morale among the remaining employees and affect the organization's ability to attract new talent.

When a well performance employee leaves the organization without any initial informed to the manager of that organization-faced difficulties. If the manager of organization has any previous information about the employee attrition, then he/she start hiring new employee process and focused on what are the parameters, which influenced the employee’s mentality and performance to resign from job.

To avoid the situation of facing rapid recruitment process, we prepare a model which helping the manager of organization to knowing what and who will be leaving organization so the manager prepares for the situation and take decision according to the stage. In this model, we obtain dataset of an organization which includes parameters like  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. After getting data we execute various algorithms like Decision tree, Logistic regression, Naïve Bayesian algorithm, Random Forest classification, Gaussian Naïve Bayes classier, k-means clustering to analyze the dataset and predicate the probability of employee’s attrition rate by taking the most accuracy score giving by algorithm.

* 1. **Problem Introduction**

When an employee leave who is performing well in the organization will leads to decline in company growth, effects the current projects, sets a negative impact on other employees and it take more time and wastage of money in training of new recruits and affect the image of organization. It is very difficult for HR management team to tackle this situation and to avoid the situation of facing rapid recruitment process.

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 depends on employees so we here to predict or to find the probability of the employee's 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. 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 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.

**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.

1.3 **Organization of the Report**

Now we are moving to chapter 2. In chapter 2, we read some immediate review on the research paper of related work. In chapter 3, we describe system design and the methodology and how the interface for user is looking and algorithms used in model. In chapter 4, we specify software and hardware requirement, dataset and the results. In the last chapter 5, the conclusion and the future description of project work is discussed.

**CHAPTER 2**

**LITERATURE SURVEY**

Studying various related literature and research papers, we find that different methods are used to predicate the attrition rate for various machine learning models are used provided with their accuracy rate to improve or compare the model for obtaining better results and in some papers, different techniques like data mining are also used for getting the best result of prediction. An interval-based survey is also taking place within the organization or feedback by employees, a managerial review is conducted about the effect of training and workspace and work environment in the organization, and to find the factors, which are responsible for employee attrition to overcome the attrition rate.

In 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 an accurate model for 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, which obtains 98%, the accuracy provided by the logistic regression is 78%.

In paper [3] 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 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 paper[4] There is using of various algorithms that 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

We study in paper[5] 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 is Training time and its duration, Support Interest among employee, workplace environment, Reward.

We study in paper[6] 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 uses to perform a specific task without being explicitly programmed. Learning algorithms in many applications that’s we make use of daily. There is 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.

**CHAPTER 3**

**SYSTEM DESIGN AND METHODOLOGY**

* 1. **System Design:**
     1. **System Architecture /Diagrammatical View**

Therefore, we design an interactive GUI (Graphical User Interface) is an interface through which a user interacts with electronic devices for both ends. At human resources team, they have an access of fully analyzed data of each every employee in form tables, charts, graphs and results of predicted attrition of company, track the performance of each employee and report the growth of organization or company with high accuracy, and improves decision-making process. On the other hand, at employee end GUI act as an interval based surveys forms, which collects data, then this collected data and historical data of employees is use to train our models which is build using scikit-learn which is a powerful machine learning module written in python. By using several inbuilt model evaluation methods. We select most accurate model, which is act as backend in our GUI.

USER

Graphical interface

Window Manager

Display Server

Kernel

Hardware

  Fig.1 GUI (Graphical User Interface)

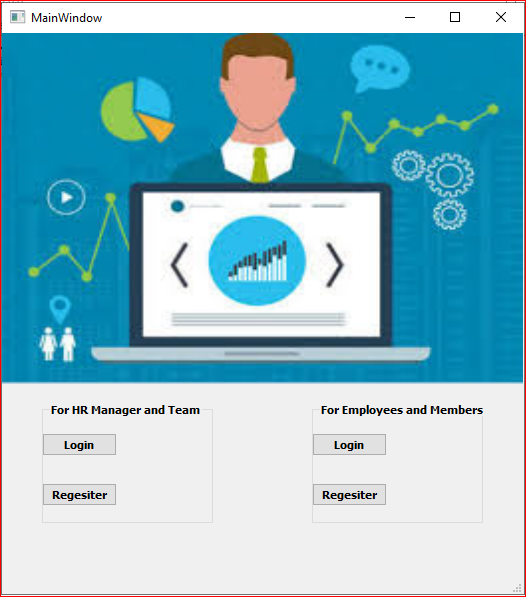


  Fig.2 Main Window of HR Helping Hand

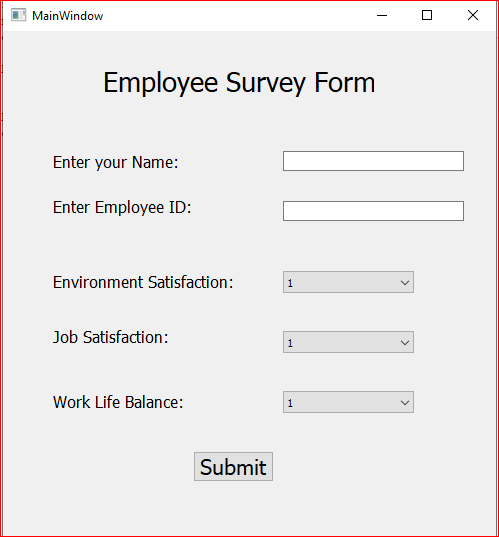


Fig.3 Employee Survey Form Window

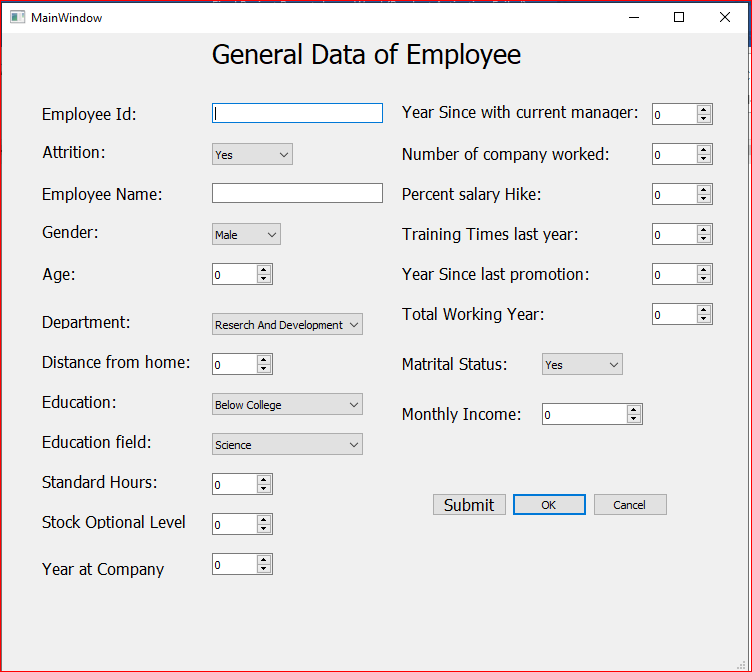


Fig.4 Employee General Data Form Window

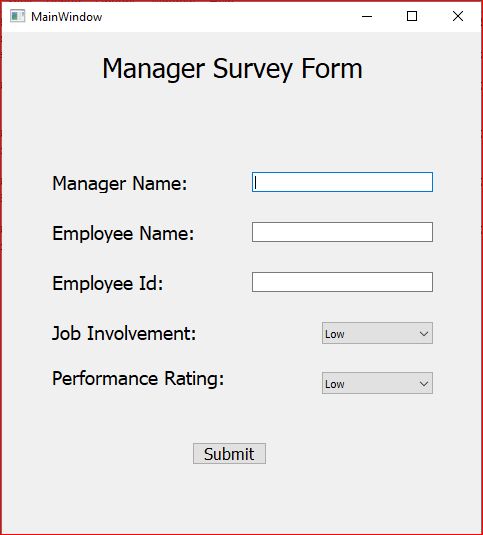


Fig.5 Manager Survey Form Window

**3.1.2 Flow Chart**

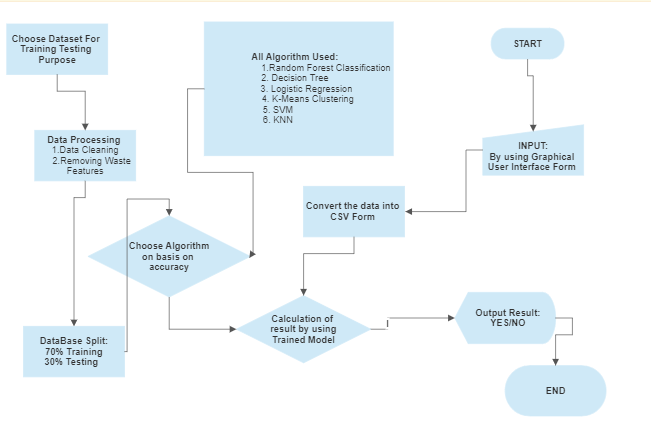


Fig.6 Data Flow Chart

* 1. **Algorithms**

Different models for prediction and classification using following machine learning algorithms.

* Decision Tree
* Random Forest Classification
* Logistic Regression
* K-means clustering
* SVM
* KNN

* + 1. **Decision Tree**

The decision tree is a popular machine-learning algorithm that is used for both classification and regression tasks. It is a model that predicts a target variable based on a series of decisions and their outcomes. The tree structure consists of decision nodes, leaf nodes, and branches.

**3.2.2 Random Forest Classification**

Random Forest is a supervised machine-learning algorithm that is used for classification, regression, and other tasks. It is an ensemble learning method that combines multiple decision trees and produces a final output by aggregating the outputs of individual trees.

**3.2.3 Logistic Regression**

Logistic Regression is a statistical method used for modeling the relationship between a binary dependent variable and one or more independent variables. It is commonly used for classification problems where the goal is to predict the probability of an event occurring.

**3.2.4 K-Mean Clustering**

K-means clustering is a popular unsupervised machine learning algorithm that is used to group or cluster data points into K clusters based on their similarity. The algorithm is based on the distance between data points and the centroid of the cluster to which the data point belongs.

**3.2.5 SVM**

Support Vector Machine (SVM) is a popular supervised machine-learning algorithm used for classification and regression tasks. The goal of SVM is to find the best hyperplane in a high-dimensional feature space that separates the data points into different classes.

**3.2.6 KNN**

K-Nearest Neighbors (KNN) is a popular machine-learning algorithm used for both classification and regression tasks. It is a non-parametric algorithm, which means it does not make any assumptions about the underlying data distribution.

**CHAPTER 4**

IMPLEMENTATION AND RESULTS

1. **Software and Hardware Requirements**

1.1 **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.  Major libraries used for project implementation:

1. **Pandas** is a python module used for reading data and data manipulation.
2. **Numpy** is a powerful python package used for computations and manipulating lists and tables of numerical data.
3. **Matplotlib**, **Seaborn**: is graphing tools used for data visualization.
4. **Scikit-Learn** is used for machine learning models.

5  **PyQt5** is used for building an interactive GUI.



Fig.7 Python Libraries

1.2 **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 considerable number of small parts called cells. Each location or cell has a unique address, which varies from zero to memory size minus one.

1.3 **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.

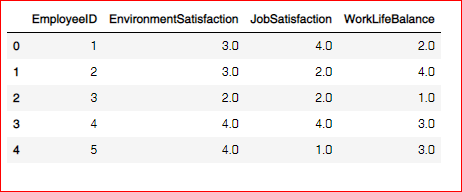
1.4 **CPU:**

 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.

**2. Snapshots**

2.1 Employee and Manager Surveys Data Analysis Inferences

Table-1 Employee Survey Data.



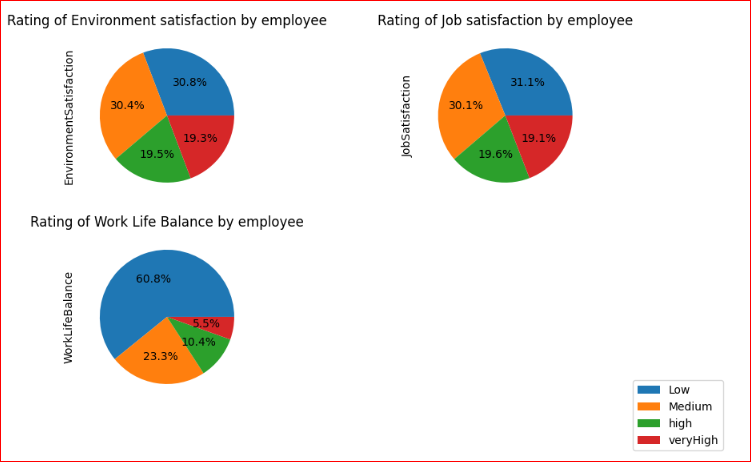
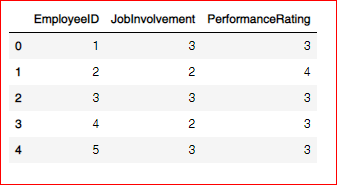


Fig.8 Employee survey data distribution.

Table-2 Manager Survey Data



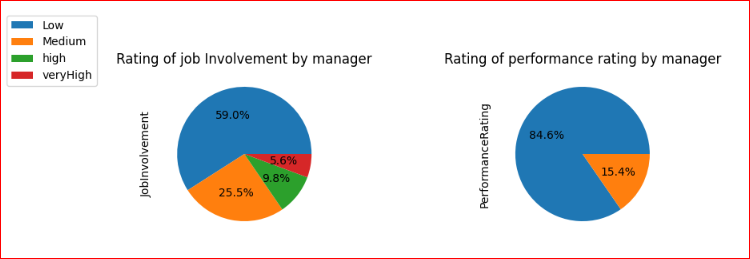
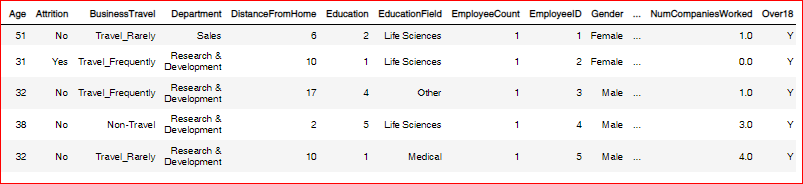


Fig.9 Manager Survey data distribution.

2.2 Employee general data analysis

Table-3 General Data



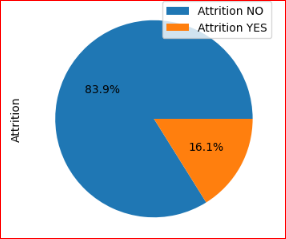


Fig.10 Distribution of employee attrition

## From the Pie Chart, we can infer that out of 4410 employees, 16.1% of the employees left their job due to some reasons whereas other 83.9% of the employees preferred to continue their job at the company.

## 

## Fig.11 Business Travel VS Attrition

## From the above data it is clear that Employees who travel rarely have more attrition rate followed by Employees who travel frequently

## 

## Fig.12 Attrition VS Working years of employee

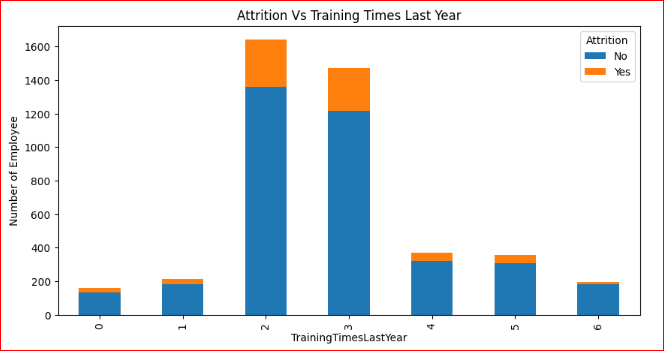


Fig.13 Attrition VS Training Time last year

#### It is observed that the employees who have two trainings last year quit their jobs most, followed by three trainings last year

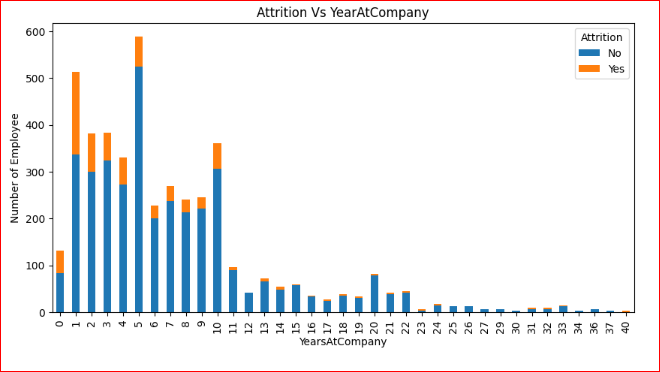


Fig.14 Attrition VS Year at company

It is observed that the newly arriving employees quit their jobs most, so more concern should be given to the freshers and their cause of leaving the company should be figured out.

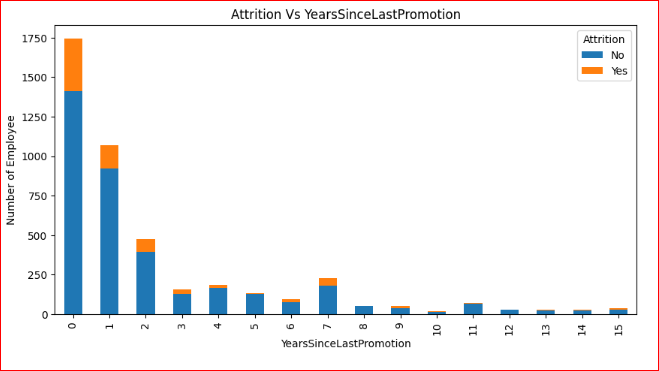


Fig.15 Attrition VS Year since last promotion

#### It is very clear that Employees who are in same post or not getting promoted tend to leave the company most

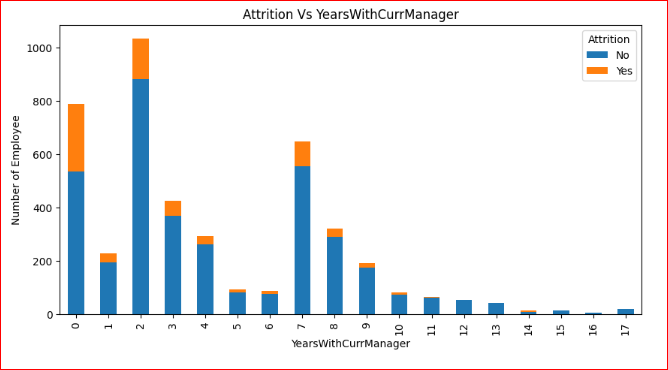


Fig.16 Attrition VS Year with current manager

#### From the above plot it is observed that attrition rate is higher in initial years

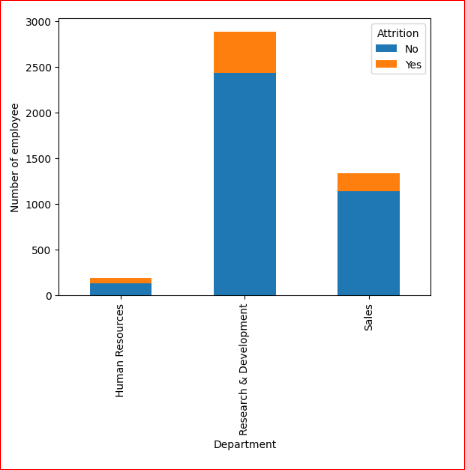


Fig.17 Distribution of Attrition in department

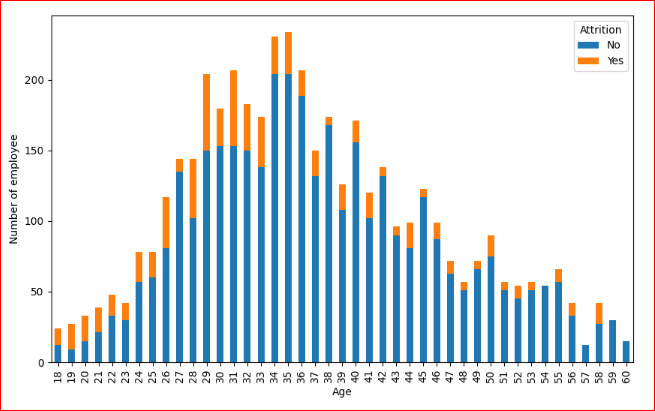


Fig.18 Distribution of Attrition in different age

#### In the above bar graph we observe that employee of age between 34-36 has higher attrition rate.

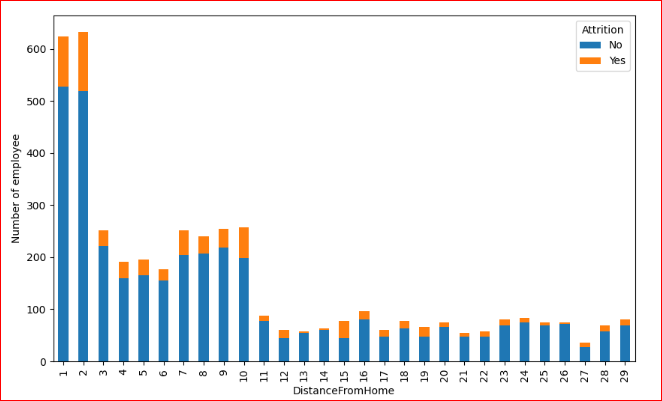
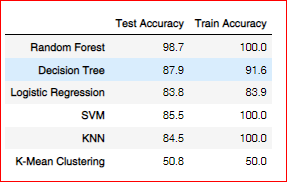


Fig.18 Distribution of Attrition for employee distance from home

It is observed that employee who lives closer to the office are more likely to have attrition in yes.

2.3 Model Evaluation Results

Table-4 Test and Train score of different models



Logistic Regression has the highest test score. Logistic Regression is used

**CHAPTER 5**

**CONCLUSION**

In conclusion, employee attrition can have a significant impact on an organization's success and long-term viability. The aim of this study was to identify factors that influence turnover and develop a predictive model to determine which employees are most likely to leave. The analysis of employee characteristics revealed that age, job level, education, monthly income, distance from home, percent salary increase, total working hours, training time in the previous year, years at the company, years since last promotion, and years with current manager are significant predictors of employee attrition. The developed prediction model accurately identifies employees who are at risk of quitting, providing employers with valuable information for retention.

**5.1 Performance Evaluation**

We applied six models to our dataset, including random forest classification, decision trees, logistic regression, SVM, KNN, and K-means clustering. Among these models, random forest classification had the highest test accuracy (98.715%), and we integrated it into our GUI.These predictive models can be integrated into existing HR procedures to proactively address retention challenges and enhance corporate success with further refinement and implementation.

**5.2 Future Direction**

* + In future, we includes more features to our GUI to make it more user friendly.
  + It will also predict the performance of employee with their historical data.
  + It will give suggestion to reduce employee attrition in organization.

**References**

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

 [2].Elham Mohammed Thabit A. Alsaadi; Sameerah Faris Khlebus, Ashwak Alabaichi. Identification of human resource 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.