

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



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



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 be 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.	21
Table 2	Manager Survey Data	22
Table 3	General Data	22
Table 4	Test and Train score of different models	28

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

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 with various parameters .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.

1.1 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 furthestmost 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

HR analytics is a field that aims to enhance the performance of HR units in enterprises, leading to better overall results. Analytics has been a part of human resources for many years, but it was traditionally done manually, with limited impact. By utilizing automated selection, processing, and data analysis techniques, HR analytics has become more effective and dynamic, allowing for more advanced proficiency in HR 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 refers to the departure of employees from an organization due to various reasons, such as resignations, retirements, and other factors. This research aims to identify employee attrition within a company by utilizing a Random Forest classification model, which can help the human resources department determine the factors that contribute to employee turnover. The study found that Information Gain has the highest accuracy value of 89.2% among the tested methods, while Pick K Best and Recursive Feature Elimination have accuracy values of 87.8% and 88.8%, respectively.

In paper [2] Various classification methods were used to determine the probability of new employee attrition, including decision trees (DT) classifier, logistic regression (LR), random forests (RF), and K-means clustering. The results indicated that the random forest algorithm achieved the highest accuracy of 98%, followed by the decision tree with an accuracy of 97% and the logistic regression with an accuracy of 78%.

In paper [3] This research aims to identify the most significant factors that contribute to employee attrition. Various machine learning algorithms such as decision trees, logistic regression, and random forests are used to improve the performance of binomial classification problems. To compare their efficacy, the NB algorithm, KNN method, and SVM algorithm were tested experimentally, and the results were analyzed. Support vector machines, K-nearest neighbors, and naive Bayesian algorithms were used to predict developer turnover. The results show that the Naive Bayes Algorithm achieved an accuracy rate of 76%, while the K-Nearest Neighbor had an accuracy rate of 94%, and the Support Vector Machine had an accuracy rate of 96%.

In paper[4] Multiple algorithms, including Gaussian Naive Bayes Classifier with the highest recall rate, Decision Tree, Logistic Regression, and Model (0.54), are being utilized. Analyzing the data assists in optimizing HR activities and reducing critical issues. Two methods, namely Holdout and Cross-validation, are used to achieve more accurate error estimation. Scaling is used to prevent any one feature from dominating the analysis.

We study in paper[5] This study focuses on the concept of staff retention and aims to identify the crucial factors that influence it. Training, including the type, duration, and impact of training, is examined to determine its effect on employee satisfaction and its potential to reduce high turnover and employee discontent. The primary objective is to determine the type, amount, and impact of training on employee retention. Factors such as the duration of training, employee support, workplace culture, and rewards are considered essential in promoting employee retention.

We study in paper[6] It is common for organizations and businesses to have a formal system in place to evaluate employee performance, often conducted once or twice a year. Machine learning algorithms, such as clustering algorithms and decision trees, which are data mining techniques, can be used to identify important characteristics for future prediction of an organization. By tracking the progress of their performance, employees can improve their performance.

In paper [7] The scientific field of machine learning (ML) explores mathematical formulas and statistical models used by computers to perform tasks without explicit programming. Learning algorithms are incorporated into several everyday programs. Various machine learning approaches include instance-based learning, reinforcement learning, multi-task learning, ensemble learning, neural networks, and supervised learning.

CHAPTER 3

SYSTEM DESIGN AND METHODOLOGY

3.1 System Design:

3.1.1 System Architecture /Diagrammatical View

Therefore, we design an interactive GUI 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.

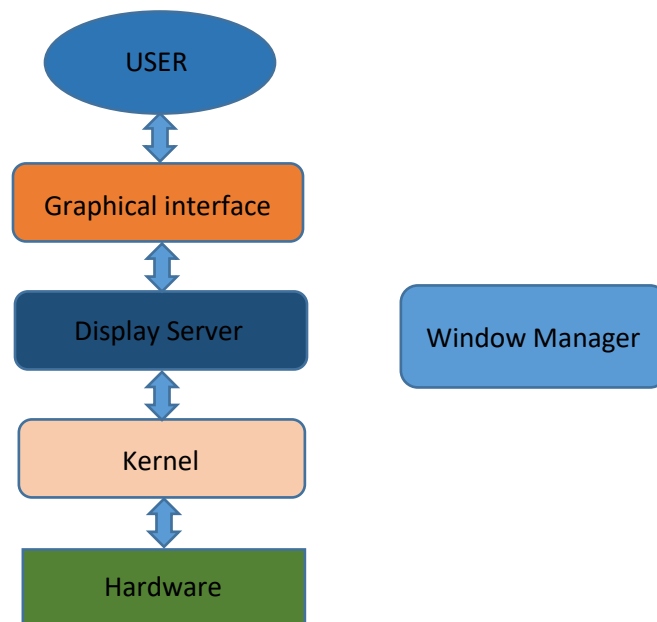


Fig.1 GUI (Graphical User Interface)

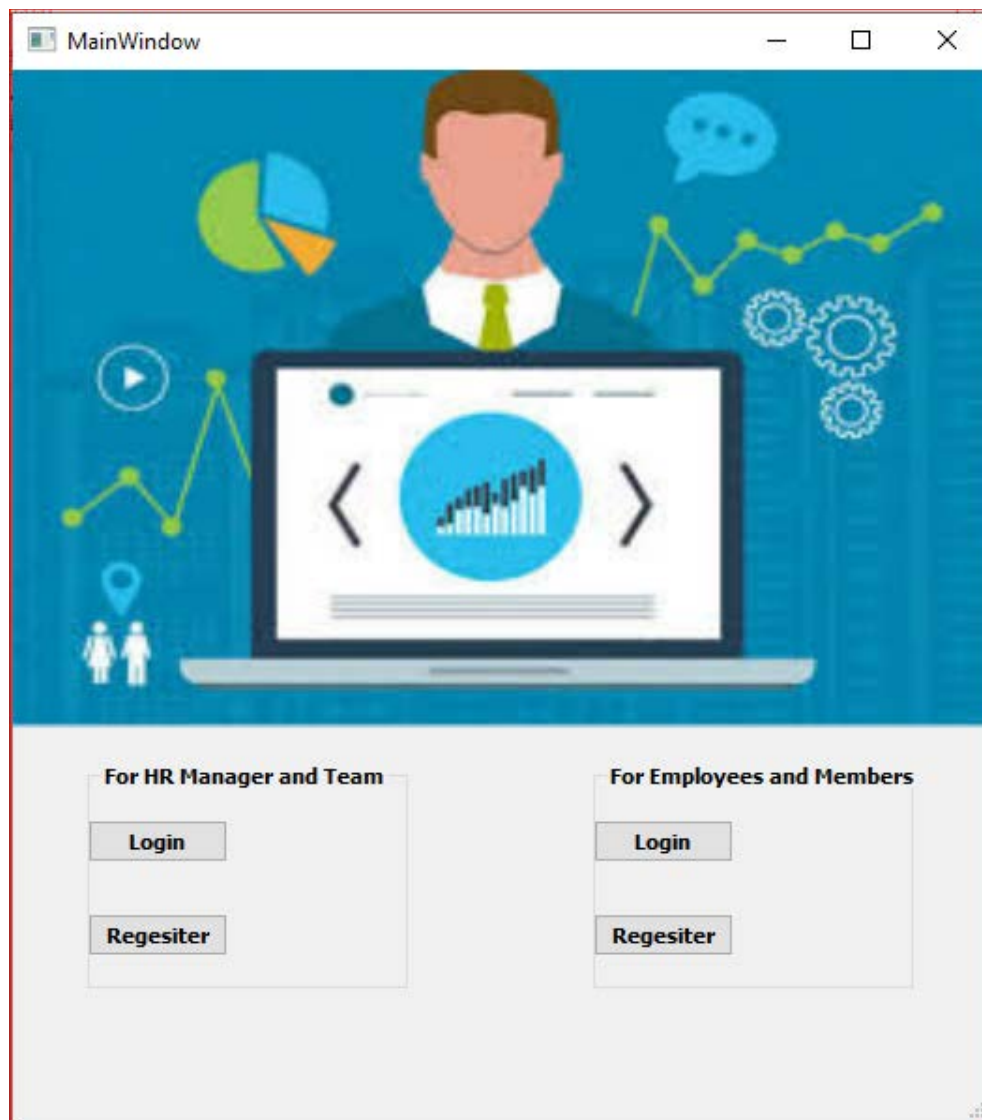
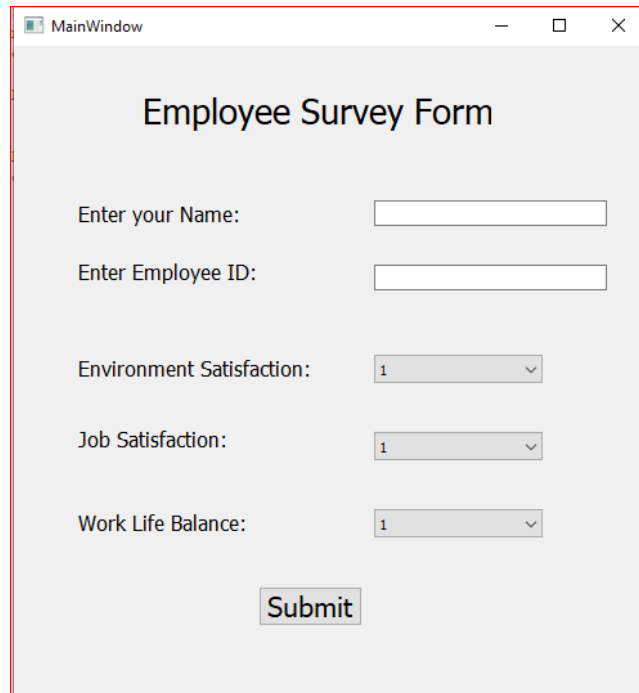
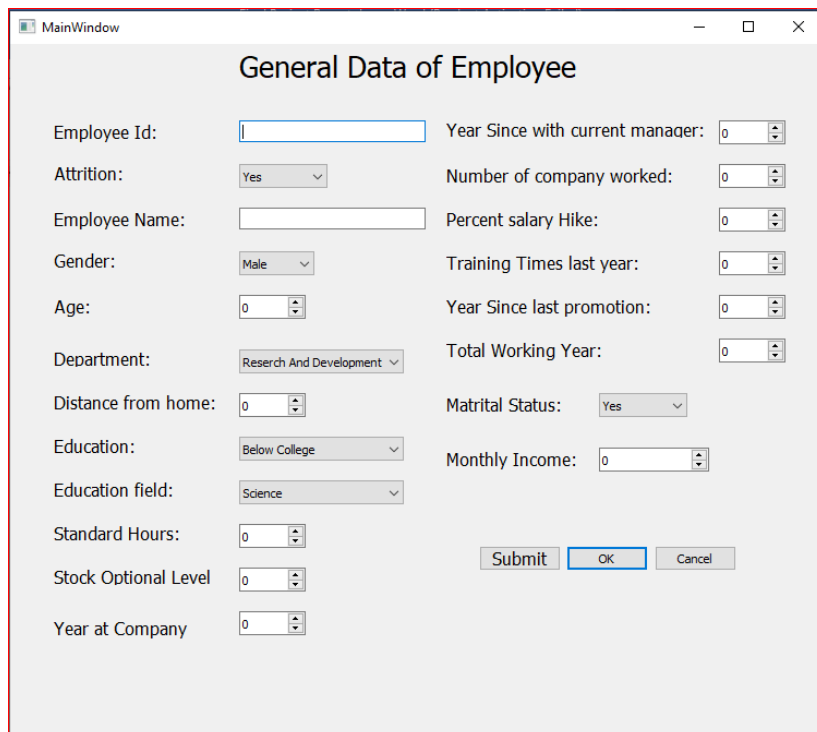


Fig.2 Main Window of HR Helping Hand



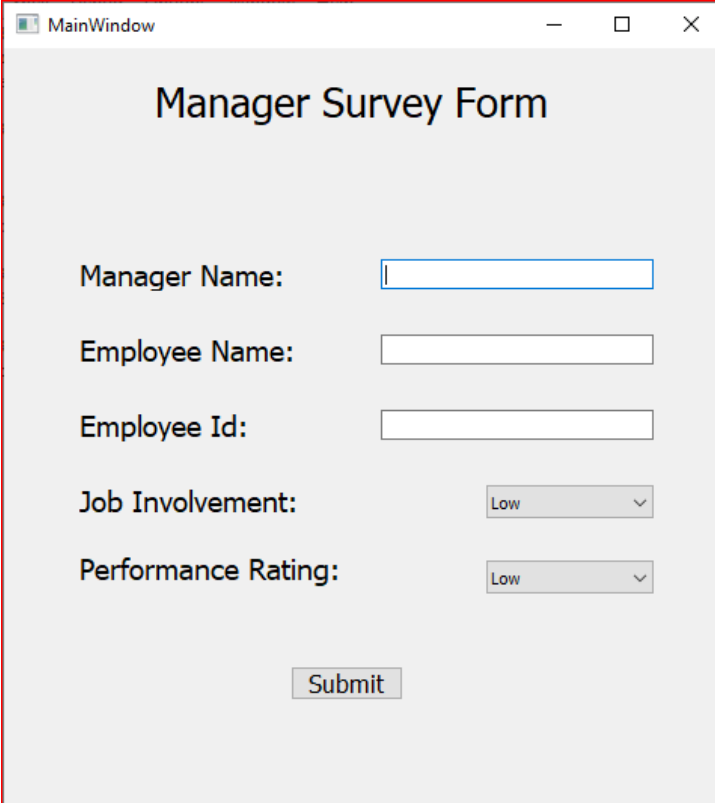
The image shows a software window titled "MainWindow" with a title bar containing standard minimize, maximize, and close buttons. The main content area has a light gray background and is titled "Employee Survey Form" in a large, bold, black font. Below the title, there are five input fields arranged vertically, each with a label to its left: "Enter your Name:" followed by a text box; "Enter Employee ID:" followed by a text box; "Environment Satisfaction:" followed by a dropdown menu showing "1"; "Job Satisfaction:" followed by a dropdown menu showing "1"; and "Work Life Balance:" followed by a dropdown menu showing "1". At the bottom center of the form is a rectangular button labeled "Submit".

Fig.3 Employee Survey Form Window



The image shows a software window titled "MainWindow" with a title bar containing standard minimize, maximize, and close buttons. The main content area has a light gray background and is titled "General Data of Employee" in a large, bold, black font. Below the title, there are two columns of input fields. The left column includes: "Employee Id:" with a text box; "Attrition:" with a dropdown menu showing "Yes"; "Employee Name:" with a text box; "Gender:" with a dropdown menu showing "Male"; "Age:" with a spinner box showing "0"; "Department:" with a dropdown menu showing "Research And Development"; "Distance from home:" with a spinner box showing "0"; "Education:" with a dropdown menu showing "Below College"; "Education field:" with a dropdown menu showing "Science"; "Standard Hours:" with a spinner box showing "0"; "Stock Optional Level" with a spinner box showing "0"; and "Year at Company" with a spinner box showing "0". The right column includes: "Year Since with current manager:" with a spinner box showing "0"; "Number of company worked:" with a spinner box showing "0"; "Percent salary Hike:" with a spinner box showing "0"; "Training Times last year:" with a spinner box showing "0"; "Year Since last promotion:" with a spinner box showing "0"; "Total Working Year:" with a spinner box showing "0"; "Matrital Status:" with a dropdown menu showing "Yes"; and "Monthly Income:" with a spinner box showing "0". At the bottom right of the form are three buttons: "Submit", "OK", and "Cancel".

Fig.4 Employee General Data Form Window



The image shows a software window titled "MainWindow" with a standard Windows-style title bar (minimize, maximize, close buttons). The window contains a "Manager Survey Form" with the following fields:

- Manager Name:** A text input field with a blue border.
- Employee Name:** A text input field with a black border.
- Employee Id:** A text input field with a black border.
- Job Involvement:** A dropdown menu currently showing "Low".
- Performance Rating:** A dropdown menu currently showing "Low".
- Submit:** A button located at the bottom center of the form.

Fig.5 Manager Survey Form Window

3.1.2 Flow Chart

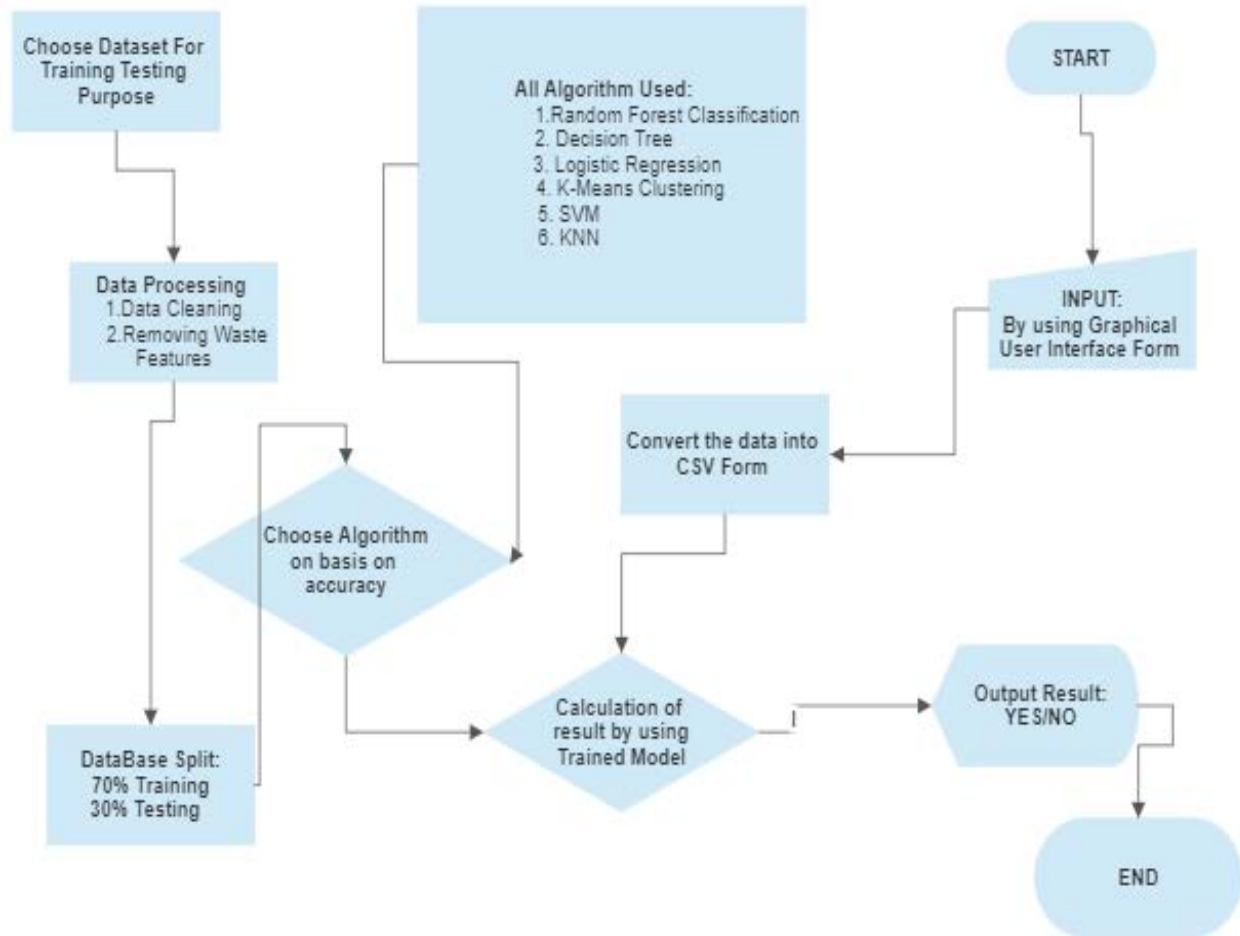


Fig.6 Data Flow Chart

3.2 Algorithms

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

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

3.2.1 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 cross platform open source coding language which contains powerful libraries and modules. 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

2. Snapshots

2.1 Employee and Manager Surveys Data Analysis Inferences

Table-1 Employee Survey Data.

EmployeeID	EnvironmentSatisfaction	JobSatisfaction	WorkLifeBalance
0	1	3.0	4.0
1	2	3.0	2.0
2	3	2.0	1.0
3	4	4.0	3.0
4	5	4.0	1.0

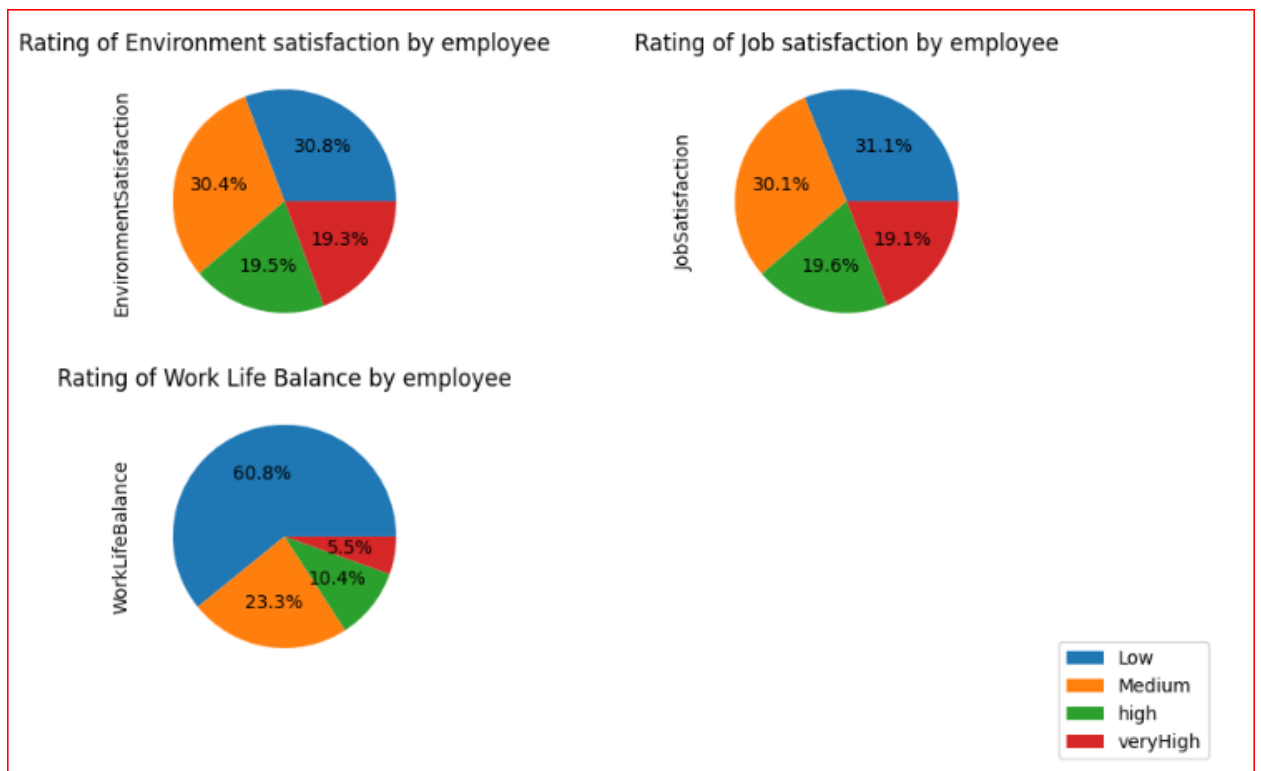


Fig.8 Employee survey data distribution.

Table-2 Manager Survey Data

EmployeeID	JobInvolvement	PerformanceRating
0	1	3
1	2	2
2	3	3
3	4	2
4	5	3

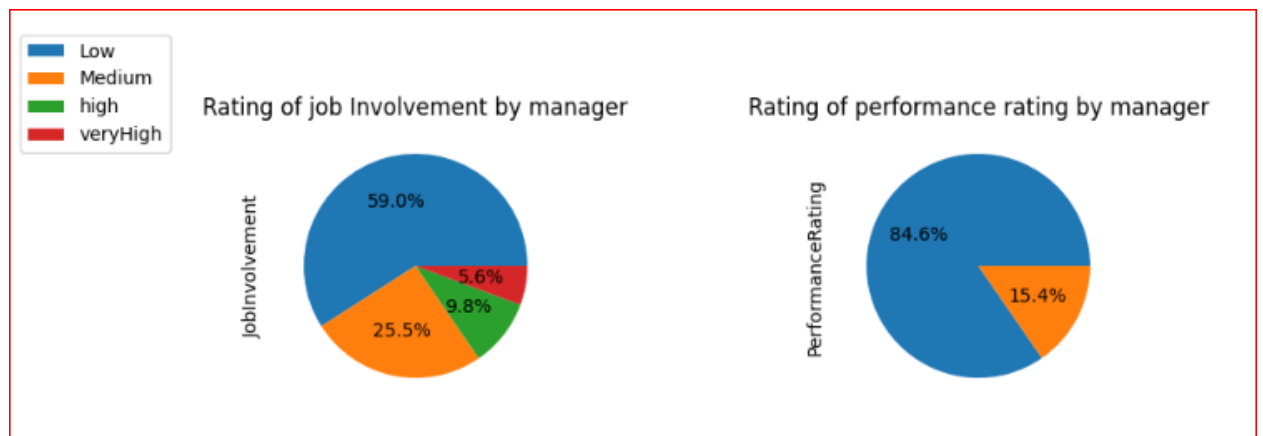


Fig.9 Manager Survey data distribution.

2.2 Employee general data analysis

Table-3 General Data

Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeID	Gender	...	NumCompaniesWorked	Over18
51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	1	Female	...	1.0	Y
31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	2	Female	...	0.0	Y
32	No	Travel_Frequently	Research & Development	17	4	Other	1	3	Male	...	1.0	Y
38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	4	Male	...	3.0	Y
32	No	Travel_Rarely	Research & Development	10	1	Medical	1	5	Male	...	4.0	Y

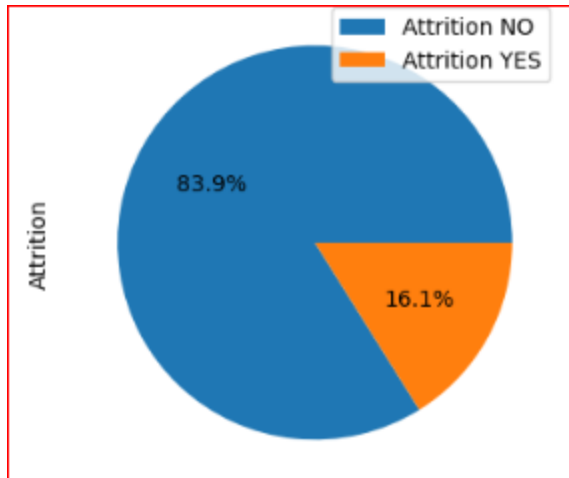


Fig.10 Distribution of employee attrition

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.

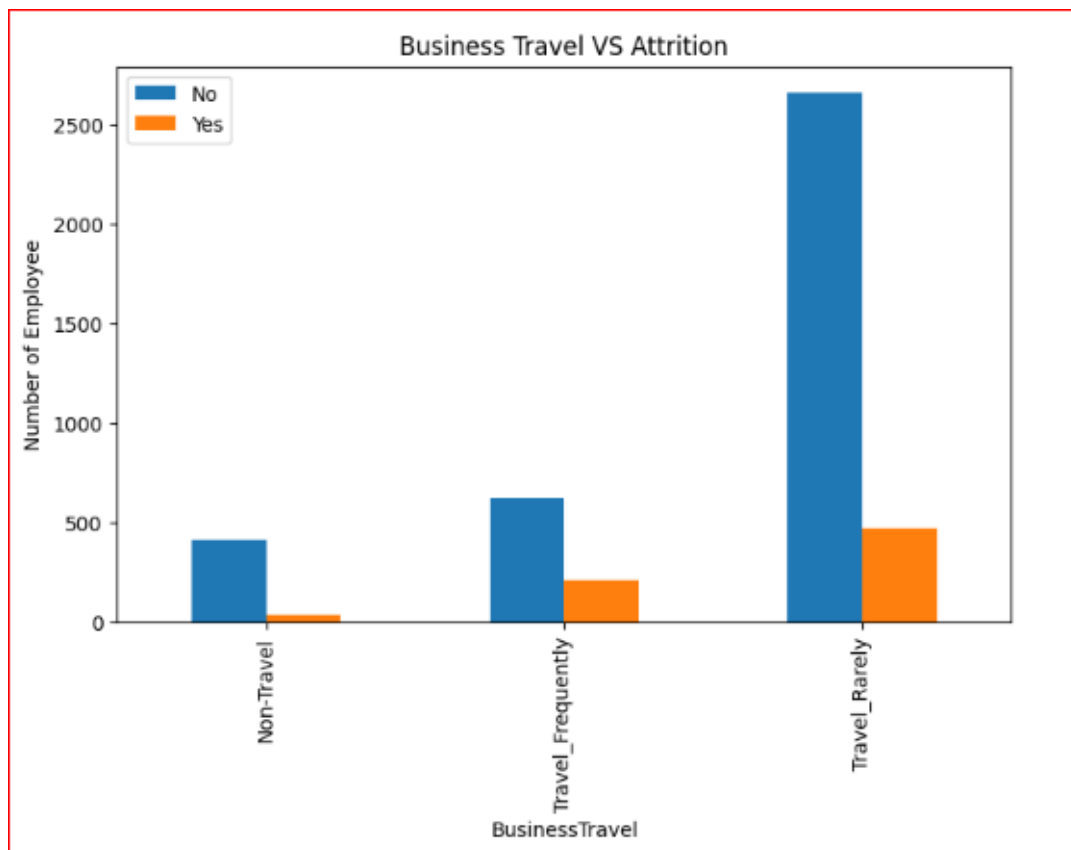


Fig.11 Business Travel VS Attrition

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

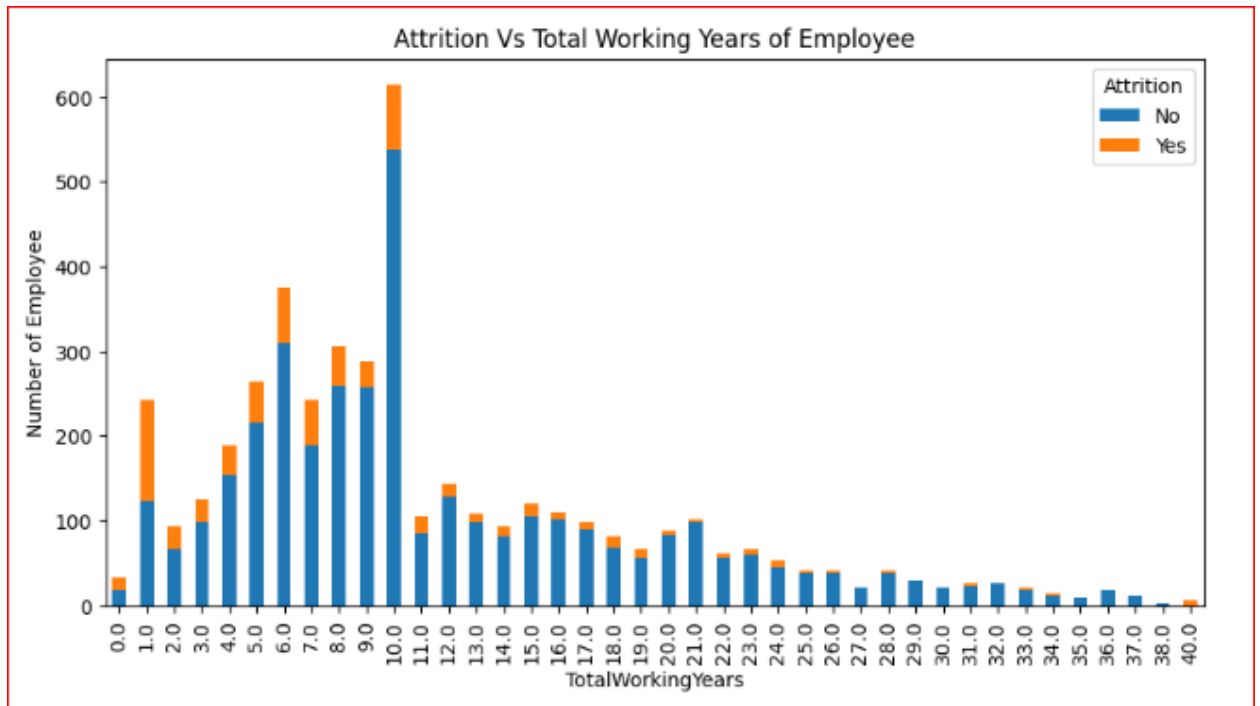


Fig.12 Attrition VS Working years of employee



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

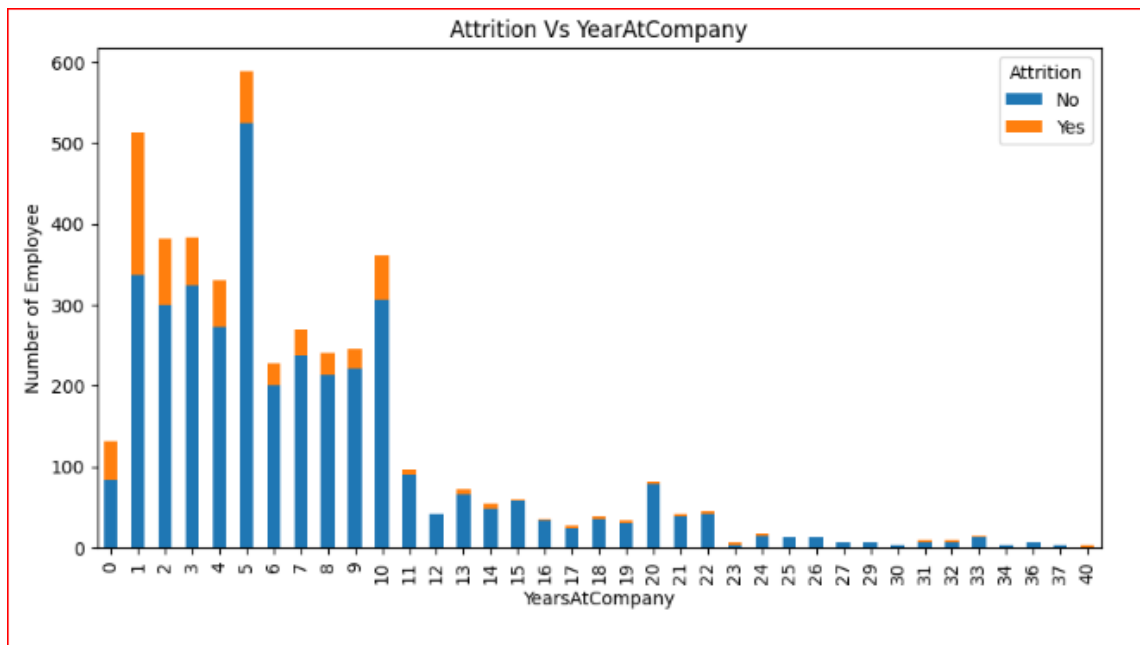


Fig.14 Attrition VS Year at company

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.

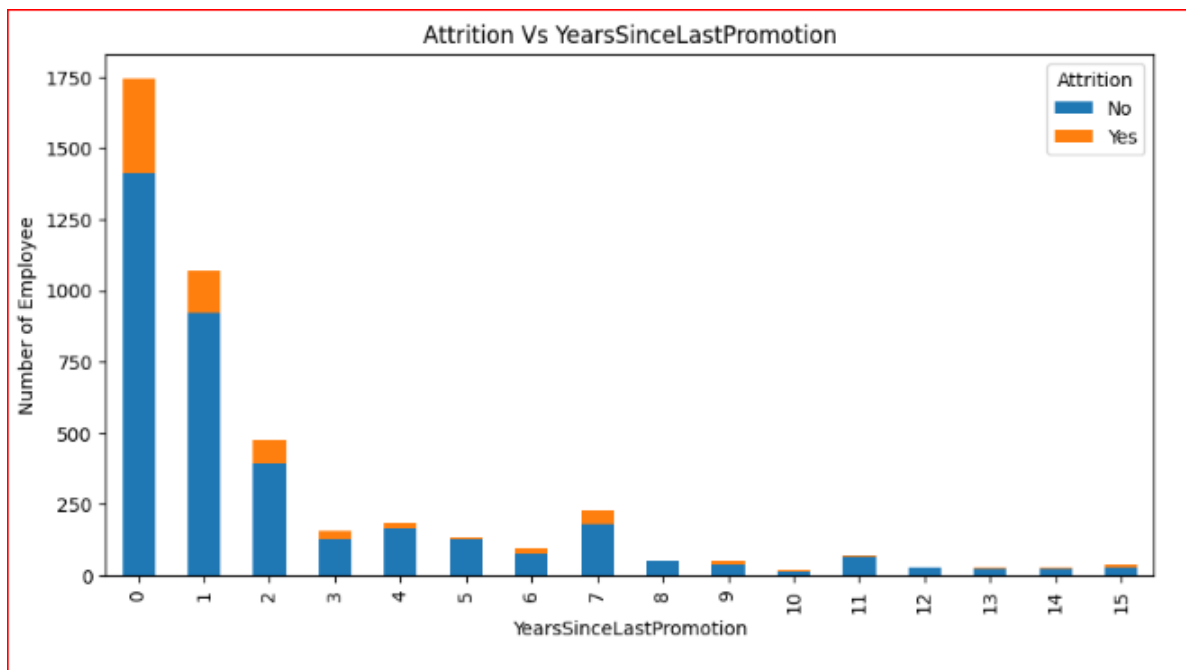


Fig.15 Attrition VS Year since last promotion

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

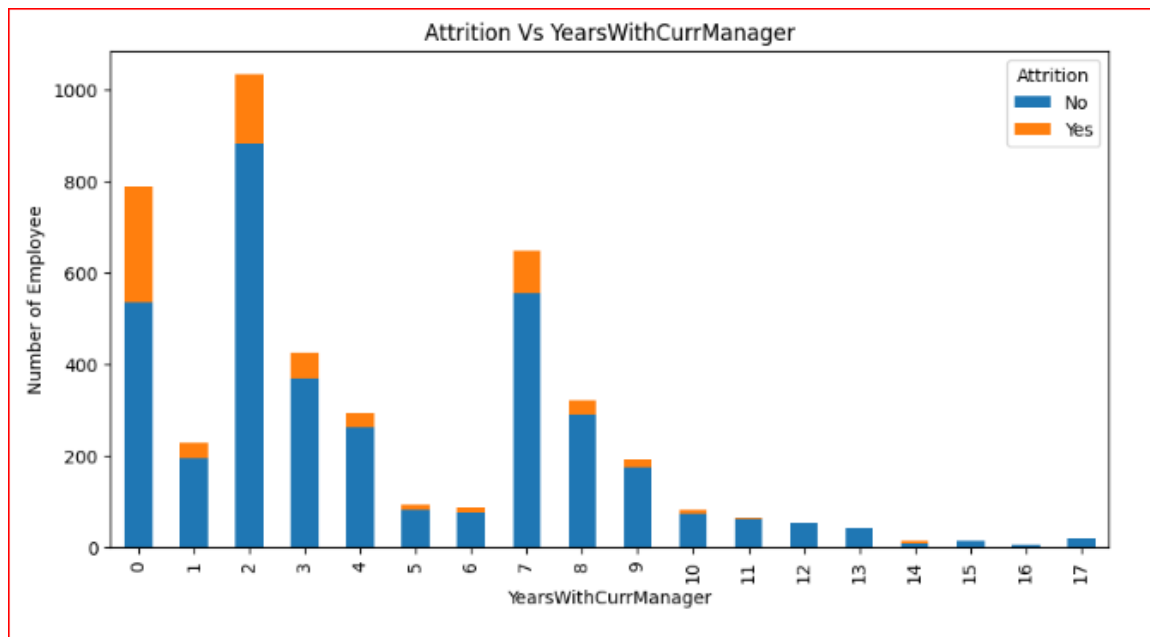


Fig.16 Attrition VS Year with current manager

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

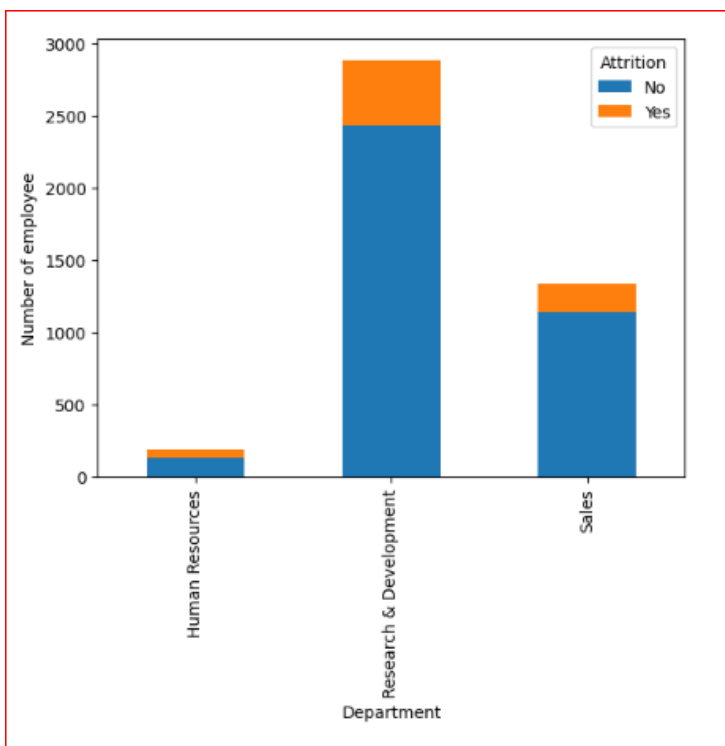


Fig.17 Distribution of Attrition in department

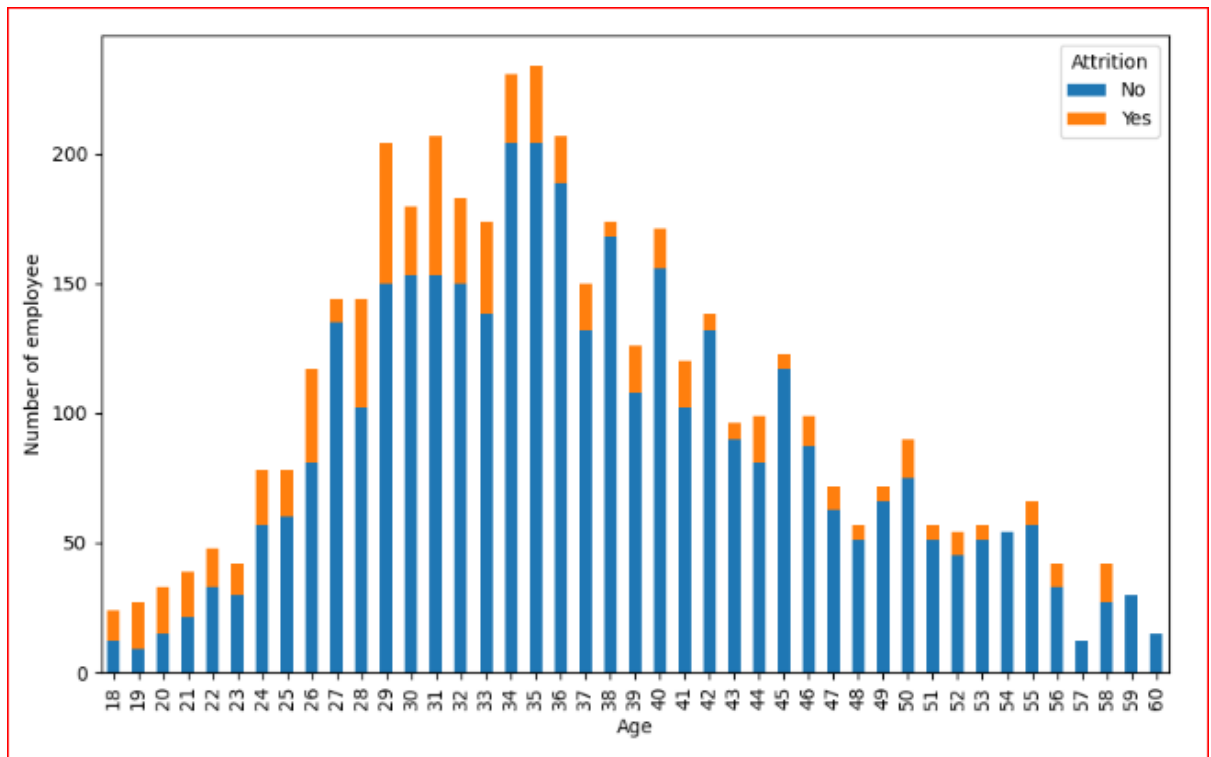


Fig.18 Distribution of Attrition in different age

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

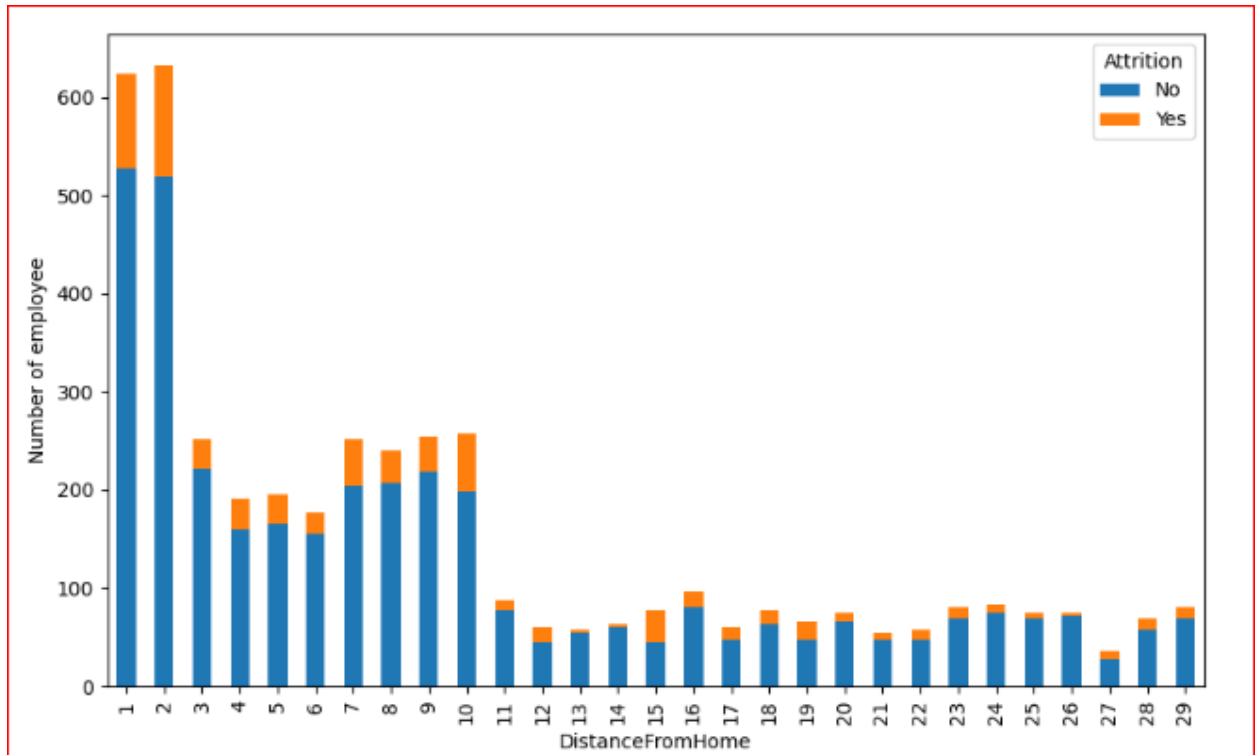


Fig.18 Distribution of Attrition for employee distance from home

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

	Test Accuracy	Train Accuracy
Random Forest	98.7	100.0
Decision Tree	87.9	91.6
Logistic Regression	83.8	83.9
SVM	85.5	100.0
KNN	84.5	100.0
K-Mean Clustering	50.8	50.0

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.