Date	19TH November
Team ID	PNT2022TMID53194
Project Name	Corporate Employee Attrition Analysis
Industry mentor	Shanawaz Anwar
Faculty mentor	DR.S.Mohanvalli

Mmebers:
R.DIVYA DARSHNI
GAYATHRI.R
LOKESHWARAN
MOHAMMED RIYAS
PREETHA.S

Project Report Format

1.INTRODUCTION

1.1 Project Overview

Employee attrition is referred as reduction in number of employees in an organization. For IT industry, employee attrition has become a known challenge since last 2 decades. Employees leave the organization for various reasons. A few reasons are, demand of high salary, change in technology or role, professional challenges etc.

High attrition leads to expense over multiple attributes and functions in the company. Recruitment, Training and Development costs increases overall cost on the employees. The core reason of this attrition could be mismatch in expectations of organization and expectations of employees from each other. This project is used to analyze the attrition reasons as well as understand the expectation of employees from the organization.

1.2 Purpose

Attrition analysis contributes to the details generated by HR managers on employees leaving the company. The metrics offer accuracy in terms of the reasons given by employees themselves. Apart from this, a wider avenue for change and dynamism also emerges from analysis of attrition.

It brings to fore the cause of employee disengagement.

2.LITERATURE SURVEY

2.1 Existing problem

A quiet significant amount of works related to the Attrition of Employees using Machine Learning algorithms have been made. An efficient attrition rate prediction has been made by using various algorithms some of them include Logistic Regression, Decison Tree, Random Forest Classifier etc. It can be seen in results that each algorithm has its strength to register the defined objectives.

The model incorporating PAM had the ability to calculate based on various attributes such as age, sex, marital status, education level, work experience, distance from hometown, etc. and generates various levels of risk of attrition.It didnt use any one algorithm but depending on the organizational contexts, different models have to be tried and evaluated before making the final selection. But the accuracy that was obtained in such PDM model was far more less than the new upcoming model .

2.2 References

- Hardik P. K. (2016) , “a study on employee attrition: with special reference to kerala it industry”. IMPACT: International Journal of Research in Business Management. 75-82
- Bodjrenou Kossivi, Ming Xu, Bomboma Kalgora (May 2016), Study on Determining Factors of Employee Retention. Open Journal of Social Sciences, Vol.4 No.5, May 30, 2016

- Brijesh Kishore Goswami, Sushmita Jha (April 2012), “Attrition Issues and Retention Challenges of Employees” , International Journal of Scientific & Engineering Research Volume 3, Issue 4, April-2012 1 ISSN 2229-5518
- Vivek Sinha, (March 10, 2011) - Attrition is Indian firms’ new worry – Vivek Sinha, Hindustan Times, (March 10, 2011) Lucknow Edition
- Sabitha Niketh (March 2008), Attrition: A Global Problem, HRM Review, March 2008 Issue, Pg. no. 64-67, ICFAI University Press, Hyderabad

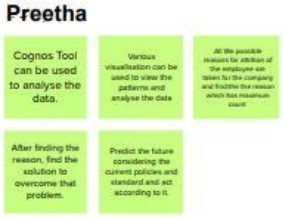
2.3. Problem Statement Definition

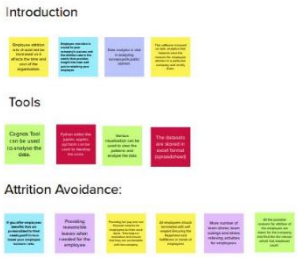
- Over the past two years, this type of analytic practice has become indispensable. Global labour markets have swung dramatically due to the COVID-19 pandemic.
- In addressing the ongoing challenges of the pandemic and the rise of remote work, employee attrition analytics will remain important to organizations seeking to retain top talent.
- Predictive analytics capability enables the design of an employee retention model to keep these valuable employees engaged and on board.
- Employee attrition analytics is specifically focused on identifying why employees voluntarily leave, what might have prevented them from leaving, and how we can use data to predict attrition risk.
- There are actually two types of attrition problems: too little and too much.
- The more talented the worker, the greater the consequences of attrition: Replacing an individual employee typically cost one or two the worker’s annual salary.
- Even if a good employee leaves as a result of “graduating” into a job with a client, if they become a great ambassador for the company, it can be a positive loss

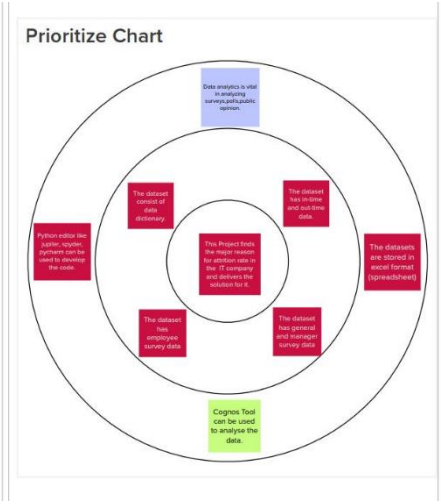
3.IDEATION AND PROPOSED SOLUTION

3.1.Empathy Map

3.2.Brainstorming and Ideation Process







3.3. Proposed solution

S.No.	Parameter	Description

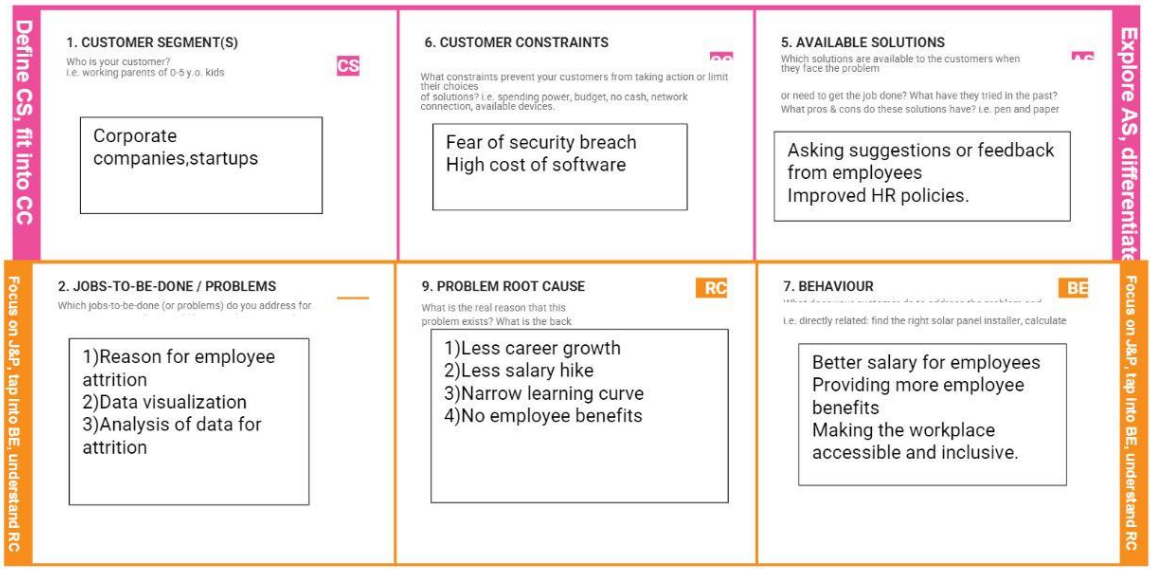
1.	Problem Statement (Problem to be solved)	Over the past two years, this type of analytic practice has become indispensable. Global labour markets have swung dramatically due to the COVID-19 pandemic. In addressing the ongoing challenges of the pandemic and the rise of remote work, employee attrition analytics will remain important to organizations seeking to retain top talent. Employee attrition analytics is specifically focused on identifying why employees voluntarily leave, what might have prevented them from leaving, and how we can use data to predict attrition risk.

2.	Idea / Solution description	Data analysis can be used to establish internal employee turnover benchmarks. Tracking these benchmarks over time can reveal how the employee experience is changing for better or worse, if the reasons employees are leaving have changed, or if the attrition pattern or time cycle is different. These benchmarks will illustrate whether the actions the organization is taking to reduce attrition are effective, alerting leaders and managers to make adjustments or take different targeted actions if needed. We can use K nearest algorithm to load, visualize, pre process the data .Using KNeighborsClassifier for finding the best number of neighbour with the help of misclassification error.
3.	Novelty / Uniqueness	We consider three types of attritions here and try to solve the problems of overcoming it. Voluntary- When an employee leaves the company for a better job opportunity or career growth or more pay, and leaves on his own. Involuntary- If an employee is terminated from a job due to some ethical issue or lack or performance. Sometimes, a degrowing business also forces employees to quit the job, which leads to a higher rate of people leaving. Retirement- Once an employee finishes his/her tenure at a company and retires. This is mostly a natural attrition that occurs and

		companies are prepared with succession planning.
4.	Social Impact / Customer Satisfaction	This helps the corporate in learning the reasons for attrition, understanding different types of attrition, trying to limit the attrition through various techniques.
5.	Business Model (Revenue Model)	This project would be a profitable one for the corporate as Attrition is something which every company faces especially in this post covid period
6.	Scalability of the Solution	Initially this model is focused on a small number of companies in the development phase. Once its successful ,the number of users increases so we can use cloud for higher storage of the large datasets of each company

3.4 Problem Solution fit

The Problem-Solution Fit simply means that we have found a problem with our customer and that the solution we have realized for it actually solves the customer's problem. It helps entrepreneurs, marketers and corporate innovators identify behavioural patterns and recognize what would work and why. The purpose is to solve complex problems in a way that fits
the state of your customers and succeed fasterand increase your solution adoptionbytapping int o existing mediumsand channels of behaviour



4. REQUIREMENT ANALYSIS

Functional requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Gmail
FR-2	User Authentication and Confirmation	Authenticate the user trying to login using database Confirmation via Email Confirmation via OTP
FR-3	Attrition analysis	Employee attrition analysis by biannual or quarterly performance appraisal, Identifying the team where the major resignation happens (three types of attrition-voluntary, involuntary, retirement)
FR-4	Employee management	Validating and managing the registered employee details.

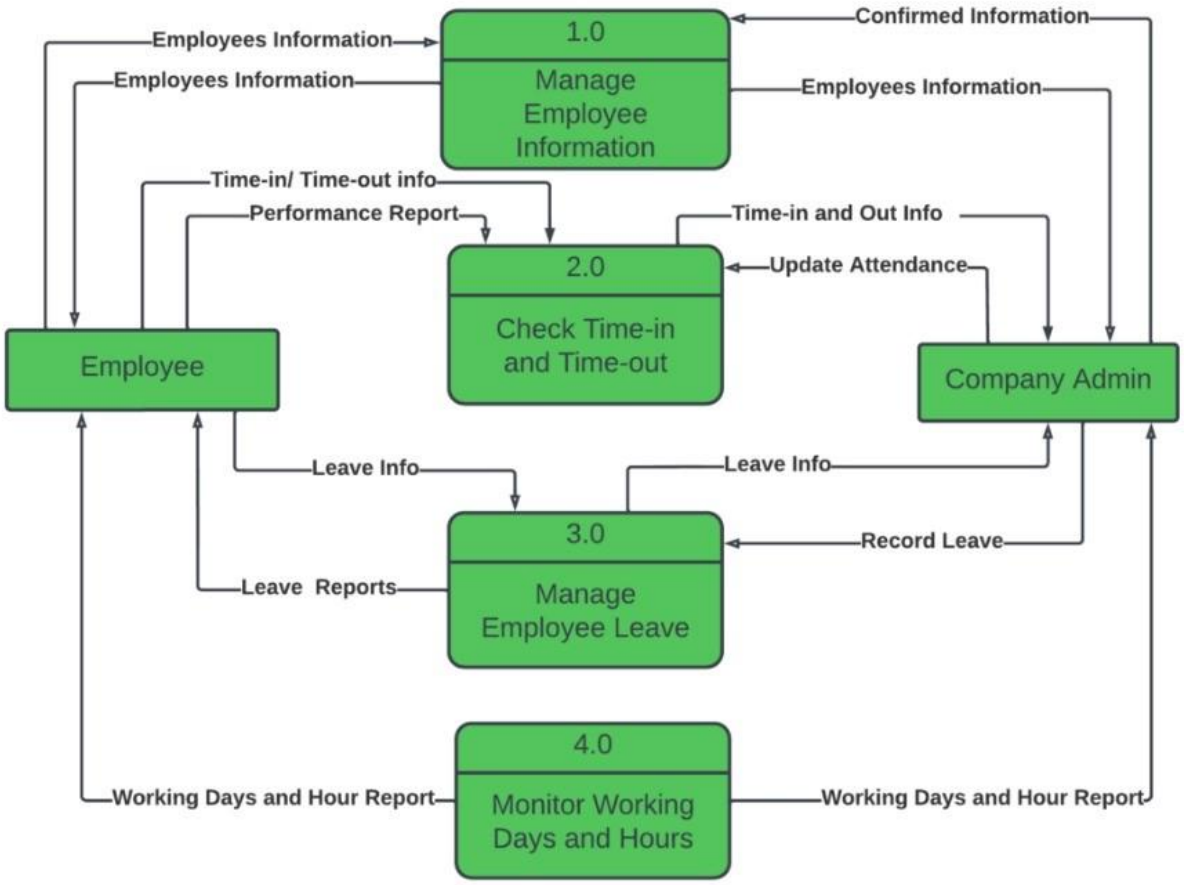
FR-5	Progress management	Maintaining the progress of each employee for Appraisals
FR-6	Link	It is used for predicting the likely attrition factors. We'll be taking the following route for this analysis - Getting our input and storing it, Select the necessary attributes forthe Prediction, Creating Dashboard, Report & Stories ,Predicting our results, Showcase the results with the help of dashboard, Report & Stories

Non-Functional requirements

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	This Data Visualization shall be easy to use for all users with minimal instructions. 100% of the languages on the graphical user interface (GUI) shall be intuitive and understandable by non-technical users.
NFR-2	**Security**	The user of the system should be provided the surety that their account details are secure.
NFR-3	**Reliability**	The Link shall be operable in all conditions. The system must be less prone to errors.
NFR-4	**Performance**	The performance of the system must assist the system's quality.
NFR-5	**Portability**	The link shall be portable to all operating platforms. Therefore, this link should not depend on the different operating systems.

5.PROJECT DESIGN

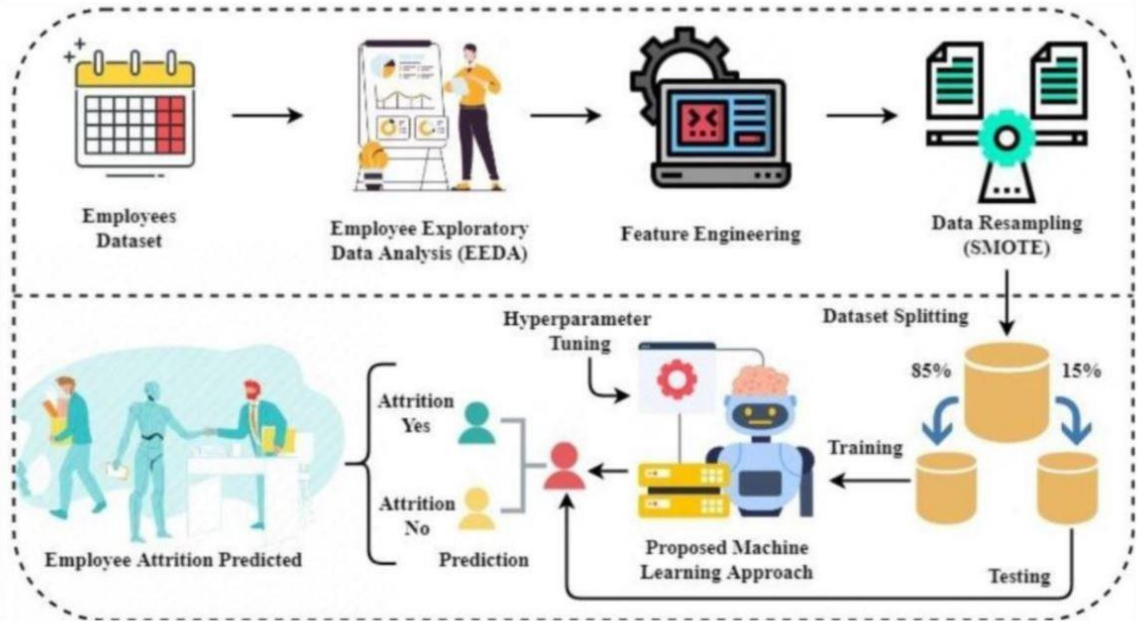
5.1 Data Flow Diagrams



DATA FLOW DIAGRAM LEVEL 1

5.2 Solution & Technical Architecture

Solution Architecture



Technical Architecture

S. No	Component	Description	Technology
1.	User Interface	How user interacts with application e.g.Web UI, MobileApp, Chatbot etc.	HTML, CSS, JavaScript / Angular Js /ReactJs etc.
2.	Database	Data Type,Configurations etc.	MySQL, NoSQL,etc.
3.	Cloud Database	Database Service on Cloud and Storing the datasets uploaded	IBM DB2, IBM Cloudant etc.
4.	Machine Learning Model	Purpose of MachineLearning Model	Training the model
5.	Infrastructure (Server / Cloud)	Application Deployment on LocalSystem / CloudLocal Server Configuration:	Local, Cloud Foundry, Kubernetes, etc.

5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Employees	Registration	USN-1	The employees can register to be a part of the Organisation by entering in form and getting authenticated via email.	I can access my account / dashboard	High	Sprint-1
		USN-2	As an employee, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As an employee, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-1
		USN-4	As an employee, I can register for the application through Gmail	I can get a verification link through email	Medium	Sprint-1
	Login	USN-5	As a employee, I can log into the application by entering email & password	I can enter the application	High	Sprint-2
	About	USN-6	I can click on the "About" to get the idea on employee attrition rate prediction using parameters like work environment, sentiments of employee, daily employee engagement and work progress.	I can get an idea about the project	Low	Sprint-2
	Launch	USN-7	As a HR, I can upload various analyzed parameters from the computer.	I can choose any employee ('s all parameters) from my device	High	Sprint-2
	Predict	USN-8	As a HR, I can perform prediction using predict button	I can view the employee's parameters on the dashboard along with the attrition rate.	High	Sprint-3
		USN-9	I can also upload csv format of employee retention parameters from cloud.	I can view the employee's parameters on the dashboard along with attrition rate	Medium	Sprint-3

6. PROJECT PLANNING & SCHEDULING

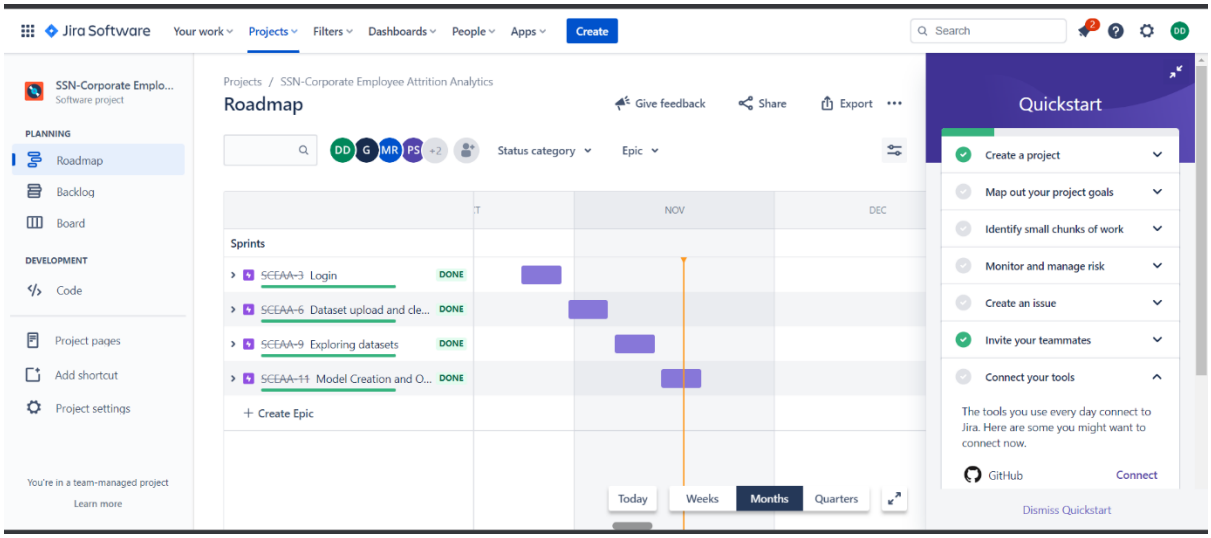
6.1 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Login	USN-1	As a user, I should be able to login in the application and the view the final output	5	High	Lokeshwaran
Sprint-2	Dataset upload and cleaning	USN-2	The analyst should be able to upload the dataset clean the dataset	2	Medium	Divya darshni
Sprint-3	Exploring dataset	USN-3	The analyst performs exploratory analysis on the data to analyze the important factors for attrition.	5	Medium	Preetha
		USN-4	The analyst presents the data using analytical tools like charts and graphs.	4	Medium	Mohammed Riyaz
Sprint-4	Model Creation and Output	USN-5	The analyst creates a model and use to predict the attrition rate and prediction is done through the website.	5	High	Gayathri

6.2 Sprint Delivery Schedule

`Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	5	6 Days	24 Oct 2022	29 Oct 2022	5	29 Oct 2022
Sprint-2	5	6 Days	31 Oct 2022	05 Nov 2022	5	05 Nov 2022
Sprint-3	5	6 Days	07 Nov 2022	12 Nov 2022	5	12 Nov 2022
Sprint-4	5	6 Days	14 Nov 2022	19 Nov 2022	5	19 Nov 2022

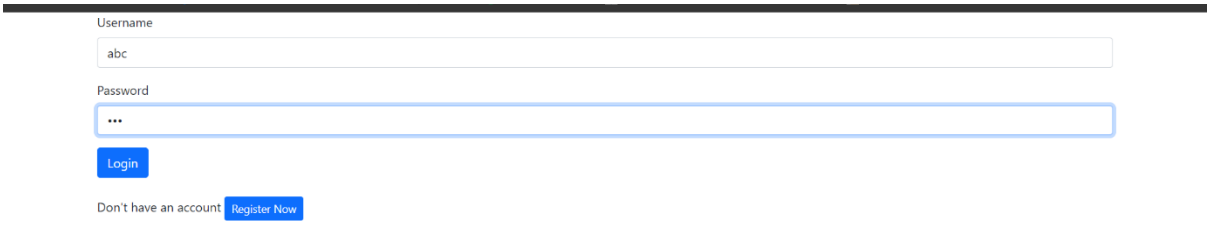
6.3 Reports from JIRA



7. CODING & SOLUTIONING

Feature 1 - Login

LOGIN



Feature 2 - Accuracy rate (Attriton)

DECISION TREE -

Accuracy :81%

```
[ ] from sklearn.model_selection import train_test_split
    X_train, X_test, y_train, y_test = train_test_split(X, y,test_size=0.25,random_state =1)

[ ] from sklearn.tree import DecisionTreeClassifier

    tree_classifier = DecisionTreeClassifier(max_depth = 10, random_state = 42)
    tree_classifier.fit(X_train,y_train)

    DecisionTreeClassifier(max_depth=10, random_state=42)

[ ] y_tree_pred = tree_classifier.predict(X_test)

[ ] from sklearn.metrics import confusion_matrix, accuracy_score, f1_score, precision_score, recall_score
    acc = accuracy_score(y_test, y_tree_pred)

[ ] acc

    0.8179347826086957
```

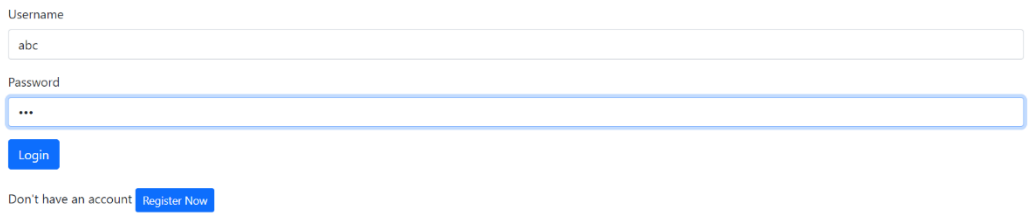
RANDOM FOREST -
Accuracy :83%

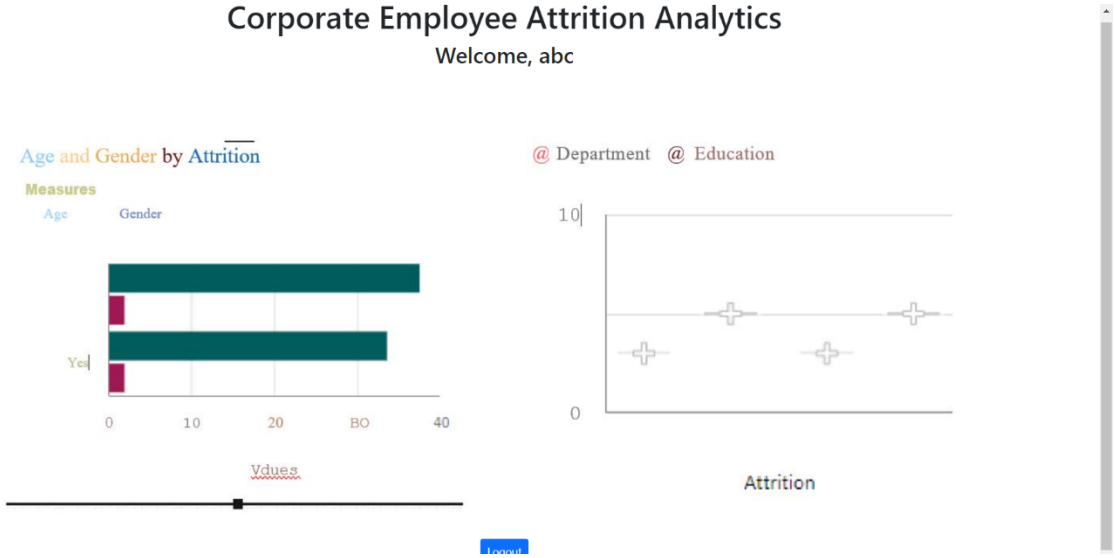
```
from sklearn.ensemble import RandomForestClassifier
forest = RandomForestClassifier(n_estimators = 100, criterion = 'entropy', random_state = 0)
forest.fit(X_train, Y_train)

RandomForestClassifier(criterion='entropy', random_state=0)

forest.score(X_train, Y_train)

0.8378684807256236
```

8. TESTING

8.1 Test Cases

Sucessfully Login will make the website go to next page-Home page





Model performance testing

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visualizations / Graphs - 10
2.	Data Responsiveness	Good
3.	Amount Data to Rendered (DB2 Metrics)	-
4.	Utilization of Data Filters	Yes for filtering out datasets with null value as they do not contribute to the prediction
5.	Effective User Story	No of Scene Added - 8
6.	Descriptive Reports	No of Visualizations / Graphs - 7

8.2 User Acceptance Testing

Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the project at the time of the release to User Acceptance Testing (UAT).

Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	4	2	3	20
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	24	14	13	26	77

Test Case Analysis

This report shows the number of test cases that have passed, failed, and untested

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	7	0	0	7
Client Application	51	0	0	51
Security	2	0	0	2
Outsource Shipping	3	0	0	3
Exception Reporting	9	0	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

9. RESULTS

9.1 Performance Metrics

- Hours worked: 50 hours
- Stick to Timelines: 100%
- Stay within budget: 100%

- Consistency of the product: 85%
- Efficiency of the product: 85%
- Quality of the product: 85%

10. ADVANTAGES & DISADVANTAGES

Advantages:

Identifying attrition can really help the company in identifying where they are going wrong and correcting it

This project has -

Smooth User Interface and Accuracy is achieved quickly

Disadvantages:

Random forest can be used for both classification and regression tasks,butit is no more suitable for Regression tasks.

This analysis is only based on the dataset or data provided,so it has to be perfectly correct.

11. CONCLUSION

Research findings suggest that attrition reasons in IT organizations primarily revolve around professional growth and challenges in the organization. Although economic factors happen to the most influential factor, professionals may settle for second best criteria of their preference that is career growth and supportive work policies in the organization.

On the other hand, candidates who aspire to have a better job than the one in hand are more interested in securing the next job. Young talent wants to work on latest technology and functional domain. IT professionals who are young career makers are less influenced by Brand name or geographical area. Most of the IT professionals look for challenging role and position in the organization. Candidates as well as senior professionals believe that challenging work motivate them to maintain the interest in the work life.

This overview of the project conveys the idea that numerous methods have been investigated for detecting the attrition rate. Big data,machine learning, and data mining can be used to great success to analyse the prediction model with the highest degree of accuracy. The primary goal of this project is calculate employee attrition in an orgamization .

12. FUTURE SCOPE

A future update shall comprise of section for upload datasets even which have null values and still get the attrition rate. The obtained output can be further processed and sent to smart devices to provide necessary information. Constant monitoring can provide necessary data to recommend to use it incase an emergency.Also it can developed as an app and hosted so that anywhere any organization can use it.

13. APPENDIX

Source Code:https://github.com/IBM-EPBL/IBM-Project-17040-1659627084/tree/main/Project%20Development%20Phase/Sprint%201/Login%20Source%20codes

GitHub &Project Demo Link : https://github.com/IBM-EPBL/IBM-Project-17040-1659627084