Project Report Format

Team ID	**PNT2022TMID53427**
Project Name Project -	**Corporate Employee Attrition Analytics**
Team Leader	**Yogitha Vijayakumar**
Team Members	**Shiva Sankari C, Sneha Priyadharshini V, Soundharya G**

1. INTRODUCTION

1.1 Project Overview

An Data Analytical model is created for the organizations to understand the factors or causes of increasing employee attrition.This model will also provide the areas where the organization could improve for employee retention.

1.2 Purpose

The purpose of this project is to create a data analytical model which can predict the attrition rate for each employee and find out the factors which lead an employee to leave the organization.

It will help organizations to understand the reasons for attrition to decrease attrition or plan in advance the hiring of the new candidate.

It's important to decrease attrition rate because costs associated with losing valuable employees whom you'd like to retain can be staggering.

This scope of the project extends to companies of midsize to large size companies and to all industries.

2. LITERATURE SURVEY

2.1 Existing Problem

Employee attrition is a recurring problem for the HR department. Employee turnover has recently increased dramatically. It is critical for employers to understand whether their employees are unsatisfied or have other reasons for leaving. It is usually prudent to explore the main source of a condition before taking drastic action.
Employees nowadays are more ready than ever to jump from one business to another in searching for a better chance. Employee attrition has become a critical issue in the majority of enterprises.There is no accurate way to determine the root cause of the issue or to deal with it.

2.2 References

1. Fatma Ozdemir, Mustafa Coskun, Cengiz Gezer, V. Cagri Gungor, 2020:Assessing Employee Attrition Using Classifications Algorithms

2. Kishori Singh,Reetu Singh, 2019:A Study on Employee Attrition:Effects and Causes

3. Rama Krishna Garigipati (Koneru Lakshmaiah Education Foundation, India), Kasula Raghu (Mahatma Gandhi Institute of Technology, India) and K. Saikumar (Koneru Lakshmaiah Education Foundation, India),2022:Detection and Identification of Employee Attrition Using a Machine Learning Algorithm

4. Saswat Barpanda and Athira S,2022:Cause of Attrition in an Information Technology-Enabled Services Company: A Triangulation Approach

5. Sabha Yousuf Khan,2019:Study on the Most Determining Factor of Employee Attrition I.E. Age Factor

6. M.S.Kamalaveni, S.Ramesh, T.Vetrivel,2019: A review of literature on employee retention

2.3 Problem Statement Definition

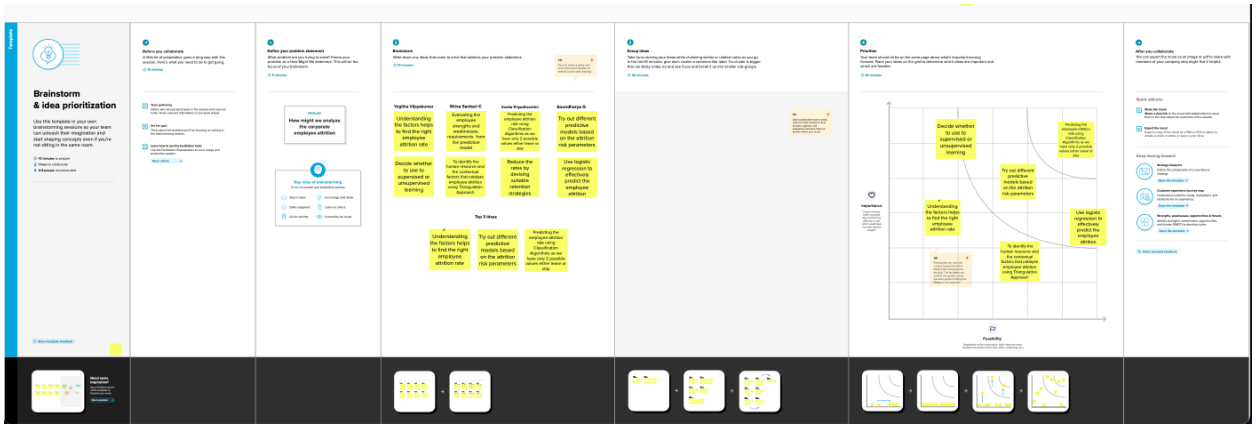
The key to success in any organization is attracting and retaining top talent. As an HR analyst one of the key tasks is to determine which factors keep employees at the company and which prompt others to leave. Given in the data is a set of data points on the employees who are either

currently working within the company or have resigned. The objective is to identify and improve these factors to prevent loss of good people.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

3.2 Ideation & Brainstorming

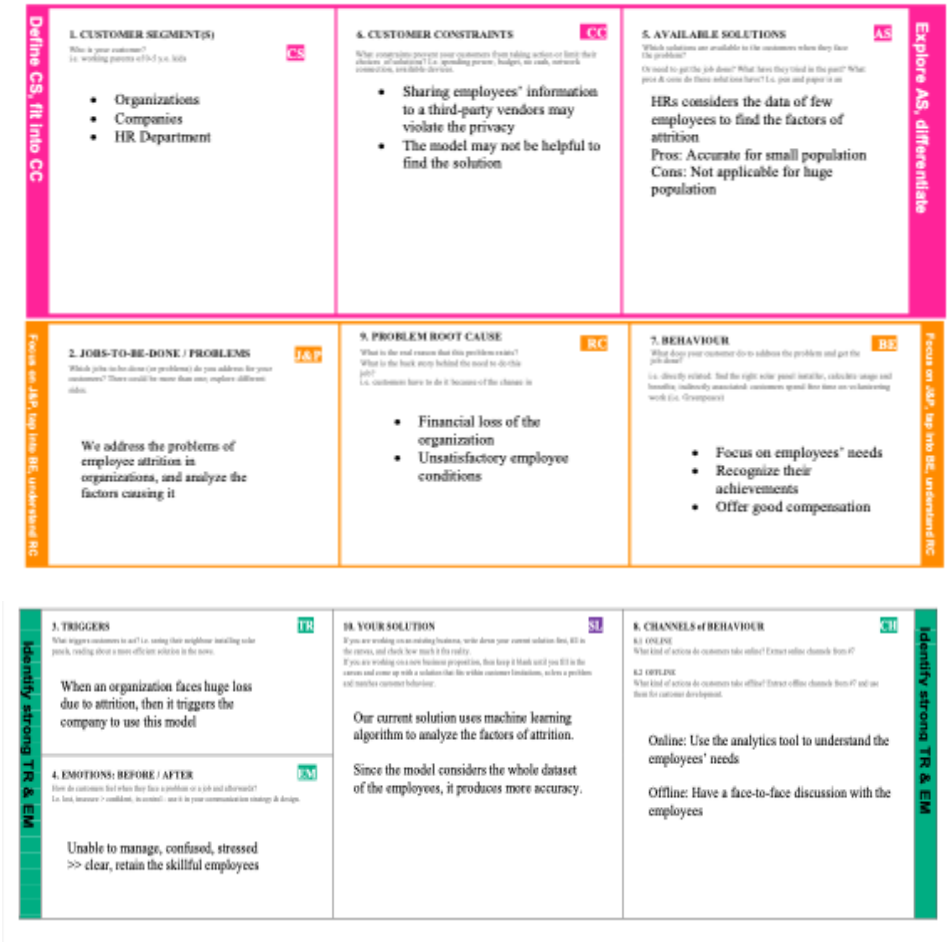


3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	For an organisation to be successful it is important to attract and retain top talents.In order to do that, an organisation must determine the factors and the cause of an employee to leave or stay.
2.	Idea / Solution description	Using algorithms to find the factors to analyze and understand the underlying pattern to improve on the factors leading to retention of employees.
3.	Novelty / Uniqueness	Right before the employee joins, the analytical system would use the factors like gender,age,work duration in previous companies and etc to categorize whether the employee would stay for long term or short term hence controlling the attrition.
4.	Social Impact / Customer Satisfaction	This analytical system would help the organisations to function steadily,maintain their reputation and avoid financial loss due to employee attrition.
5.	Business Model (Revenue Model)	Many organisations use 'standard' attrition rates such as 4% for employees under the age of 25 and above, 3% for those between the ages of 31 and 44, and 2% for those 45 and beyond.Attrition of employees leads to decreased productivity.
6.	Scalability of the Solution	This analytical system is only applicable to employees with experience, organizations like hospitals,IT,airline,etc and results may not be helpful for small scale start-ups because all they need is to get their work done.

3.4 Problem Solution Fit

Project Title: Corporate Employee Attrition Analytics | Project Design Phase-I - Solution Fit Template | Team ID: PNT2022TMID5342



4. REQUIREMENT ANALYSIS:

Requirement's analysis, also called requirements engineering, is the process of determining user expectations for a new or modified product. These features, called requirements, must be quantifiable, relevant and detailed. In software engineering, such requirements are often called functional specifications and are an important aspect of project management.

4.1 Functional requirements

The aim of our model is to study factors like salary, superior – subordinate relationship, growth opportunities, facilities, policies and procedures, recognition, appreciation, suggestions, co-workers by which it helps to know the Attrition level in the organizations and factors relating to retain them.The models helps to

- To know the satisfactory level of employees towards their job and working conditions.
- To identify the factors which make employees dissatisfied about company's policy and norms.
- To find the areas where companies are lagging behind.

Registration:

- As a HR, I can register for the application by entering my email, password, and confirming my password.
- As a HR, I will receive a confirmation email once I have registered for the application.
- As a HR, I can register for the application through LinkedIn.
- As a HR, I can register for the application through Gmail.
-

Login:

- The HR can log into the application by entering email & password.
- He/She can access the already account from this.

Dashboard:

- The HR can use the dashboard to upload data.
- He/She can update employee data.

Cleaning:

- The HR can clean the data uploaded.
- He/She can avoid the noisy data.

Processing:

The HR can process the input data using a suitable model.

Prediction:

HR can predict the result of employee attrition.

Visualization:

- The HR can use it to visualize the results and see the attrition rate.
- I can see the results in the form of graph,bar etc

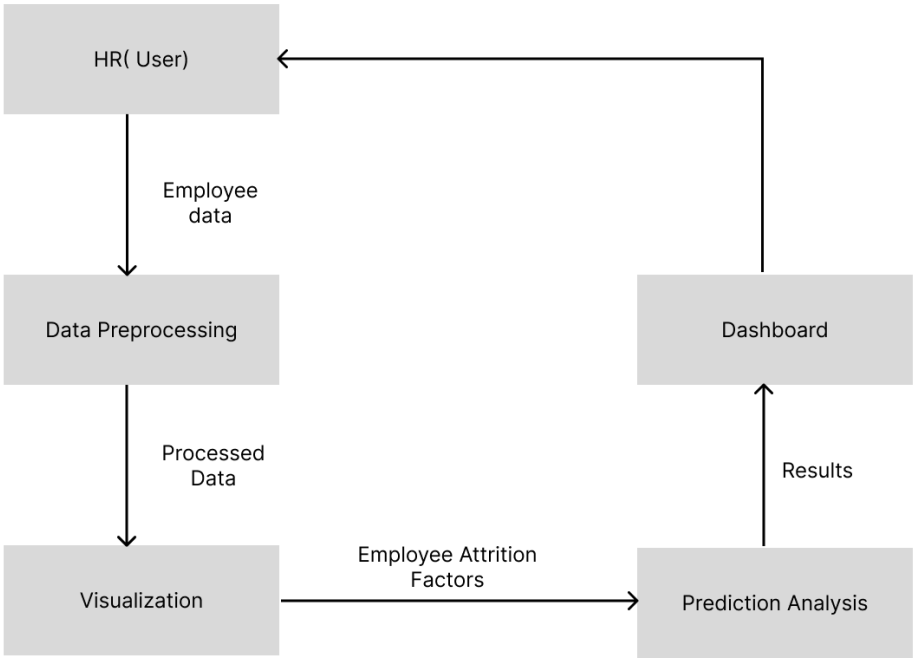
4.2 Non-Functional requirements:

Usability:
The proposed system should be easy for the user to operate, enter data, and interpret the output.
Compatibility:
The proposed system should be compatible with all web browsers.
Security:
There is a need for a proper and encrypted login authentication for head chef and admin as employee sensitive information as well as inventory should be protected from hacking.
Flexibility:
If need arises in the future, software can be modified to change the requirements.
Maintainability:
Software can be easily repaired if a fault occurs.
Portability: Software can be easily installed on devices and would run smoothly according to the requirement.

5. PROJECT DESIGN

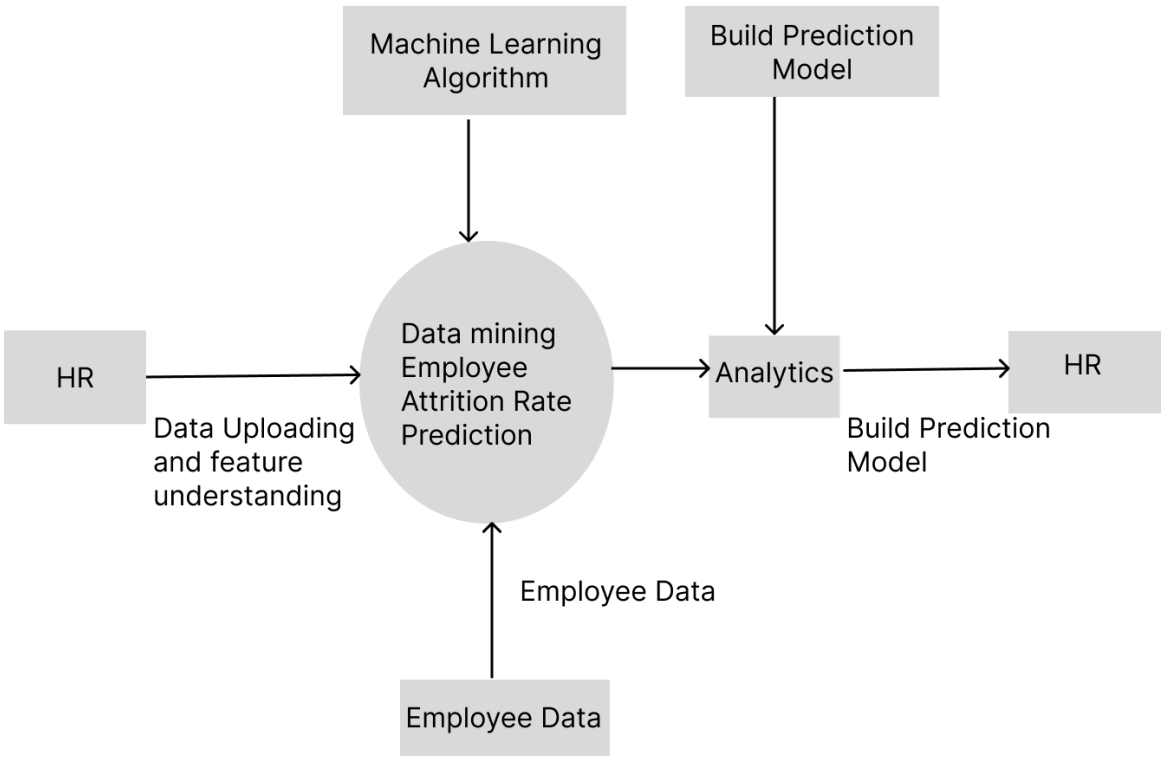
5.1 Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



5.2 Solution & Technical Architecture:

Solution architecture is the process of developing solutions based on predefined processes, guidelines and best practices with the objective that the developed solution fits within the enterprise architecture in terms of information architecture, system portfolios, integration requirements and many more.



5.3 User Stories:

User Type	Functional Requirements (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (HR)	Registration	USN-1	As a HR, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a HR, I will receive confirmation	I can receive confirmation email & click	High	Sprint-1

			email once I have registered for the application	to confirm		
		USN-3	As a HR, I can register for the application through LinkedIn	I can register & access the dashboard with LinkedIn	Low	Sprint-2
		USN-4	As a HR, I can register for the application through Gmail	I can register & access the dashboard with Gmail	Medium	Sprint-1
	Login	USN-5	As a HR, I can log into the application by entering email & password	I can access my account	High	Sprint-1
	Dashboard	USN-6	As a HR, I can use the dashboard to upload the data	I can update employee data	High	Sprint-1
	Cleaning	USN-7	As a HR, I can clean the data uploaded	I can avoid the noisy data	High	Sprint-2
	Processing	USN-8	As a HR, I can able to process the input data using a suitable model	I can process the input data	High	Sprint-2

	Predict	**USN-9**	**As a HR, I can able to predict the result of employee attrition**	**I can predict the employee attrition rate**	**High**	**Sprint-2**
	Visualize	**USN-10**	**As a HR, I can able to use to visualize the results and see the attrition rate**	**I can see the results in the form of a graph,bar and etc**	**Medium**	**Sprint - 3**

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation:

Sprint	**Functional Requirement (Epic)**	**Story Points**	**Priority**	**Team Members**
Sprint-1	Registration	2	High	Yogitha, Shiva Sankari,Soundharya, Sneha Priyadharshini
Sprint-1	Login	5	High	Yogitha,Soundharya
Sprint-2	Dashboard	10	High	Shiva Sankari, Sneha Priyadharshini
Sprint-3	Cleaning	5	High	Yogitha,Shiva Sankari

Sprint-3	Processing	5	High	Soundharya,Sneha Priyadharshini
Sprint-4	Predict	4	High	Shiva Sankari, Soundharya
Sprint- 4	Visualize	6	Medium	Sneha Priyadarshini, Yogitha

6.2 Sprint Delivery Schedule:

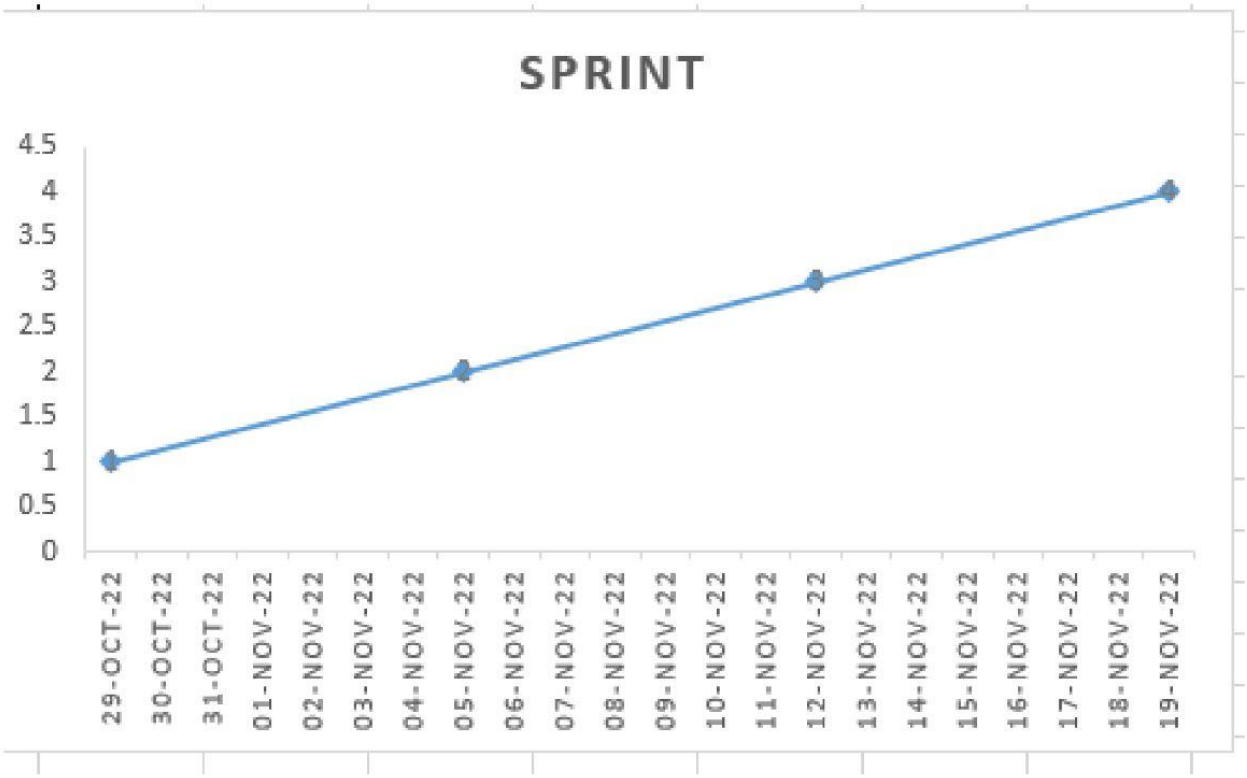
Project Tracker, Velocity & Burndown Chart: (4 Marks)

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	10	6 Days	24 Oct 2022	29 Oct 2022	10	29 Oct 2022
Sprint-2	10	6 Days	31 Oct 2022	05 Nov 2022	10	
Sprint-3	10	6 Days	07 Nov 2022	12 Nov 2022	10	
Sprint-4	10	6 Days	14 Nov 2022	19 Nov 2022	10	

Velocity:
Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

$$\text{AV} = \frac{Sprint\ duration}{velocity} = \frac{6}{10} = 0.6$$

6.3 Reports from JIRA



7. CODING & SOLUTIONING (Explain the features added in the project along with code)

7.1 Feature 1

Random Forest Classifier

Random Forest is a powerful algorithm in Machine Learning. It is based on the Ensemble Learning technique (bagging).

```
import pandas as pd
import numpy as np
from lightgbm import LGBMClassifier
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import make_column_transformer, make_column_selector
from sklearn.pipeline import make_pipeline
from sklearn.ensemble import RandomForestClassifier, VotingClassifier
import joblib


df = pd.read_csv('general_data.csv')

df.drop(['EmployeeID','EmployeeCount','Over18','StockOptionLevel'],axis=1,inplace=True)

df.drop_duplicates(inplace=True)

X = df.drop('Attrition', inplace=False, axis=1)
y = df.Attrition

ohe = OneHotEncoder(sparse=False, handle_unknown='ignore')

ct = make_column_transformer((ohe, make_column_selector(dtype_include='object')),
                             remainder='passthrough')


rnd = RandomForestClassifier(n_estimators=500,min_samples_split=80,min_samples_leaf=2,
                             max_features='log2',max_depth=8,random_state=42)
```

7.2 Feature 2

LGBM Classifier

Light GBM is a fast, distributed, high-performance gradient boosting framework that uses a tree-based learning algorithm. It also supports GPU learning and is thus widely used for data science application development.

Light GBM splits the tree leaf-wise with the best fit whereas other boosting algorithms split the tree depth-wise or level-wise rather than leaf-wise. In other words, Light GBM grows trees vertically while other algorithms grow trees horizontally.

```
import pandas as pd
import numpy as np
from lightgbm import LGBMClassifier
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import make_column_transformer, make_column_selector
from sklearn.pipeline import make_pipeline
from sklearn.ensemble import RandomForestClassifier, VotingClassifier
import joblib


df = pd.read_csv('general_data.csv')

df.drop(['EmployeeID','EmployeeCount','Over18','StockOptionLevel'],axis=1,inplace=True)

df.drop_duplicates(inplace=True)

X = df.drop('Attrition', inplace=False, axis=1)
y = df.Attrition

ohe = OneHotEncoder(sparse=False, handle_unknown='ignore')

ct = make_column_transformer((ohe, make_column_selector(dtype_include='object')),
                             remainder='passthrough')


rnd = RandomForestClassifier(n_estimators=500,min_samples_split=80,min_samples_leaf=2,
                             max_features='log2',max_depth=8,random_state=42)

lgbm = LGBMClassifier(random_state=42)

vc = VotingClassifier(estimators=[('rnd', rnd), ('lgbm', lgbm)],voting='soft', n_jobs=-1)

pipe = make_pipeline(ct,vc)

pipe.fit(X, y)

filename = 'model.pkl'
joblib.dump(pipe, filename)
```

8 TESTING

8.1 Test Cases

LoginPage_TC_001
LoginPage_TC_002
LoginPage_TC_003
LoginPage_TC_004
LoginPage_TC_005
Prediction
Make_Another_Prediction
Logout

Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Act
LoginPage_TC_OO1	Functional	Home Page	Verify user is able to see the	Registered User	1.Enter URL and click go	https://shopenzer.com/	Login/Signup popup should display	W
LoginPage_TC_OO2	UI	Home Page	Verify the UI elements in	Registered User	1.Enter URL and click go	https://shopenzer.com/	Application should show below UI	W
LoginPage_TC_OO3	Functional	Home page	Verify user is able to log into		1.Enter URL(https://shopenzer.com/)	Username:	User should navigate to user account	W
LoginPage_TC_OO4	Functional	Login page	Verify user is able to log into		1.Enter URL(https://shopenzer.com/)	Username: chalam@gmail	Application should show 'Incorrect	W
LoginPage_TC_OO4	Functional	Login page	Verify user is able to log into		1.Enter URL(https://shopenzer.com/)	Username:	Application should show 'Incorrect	W
LoginPage_TC_OO5	Functional	Login page	Verify user is able to log into		1.Enter URL(https://shopenzer.com/)	Username: chalam	Application should show 'Incorrect	W
Prediction	Functional	Home Page	Verify the user is able to easily access the UI to predict the attrition		Enter the required valuesClick Predict	Dataset provided by IBM	Predict the Attrition as Yes/No	
ake_Another_Predicti	Functional	Result Page	Verify the user is able to easily access the UI to predict the attrition af		Click Make Another PredictionGive the	Dataset provided by IBM	Predict the Attrition as Yes/No	
Logout	Functional	Logout Page	Verify the user is able to logout and redirected to login page again		Click Logout in result pageRedirected t	Not required	Logout from the Prediction Software	

8.2 User Acceptance Testing

2. 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	Subtotal
By Design	5	0	0	5
Duplicate	0	0	0	0
External	2	3	0	5
Fixed	7	3	0	10
Not Reproduced	0	0	0	0
Skipped	0	0	0	0
Won't Fix	0	0	0	0
Totals	14	6	0	20

3. 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	5	0	0	5
Client Application	5	0	0	5
Security	5	0	0	5
Outsource Shipping	5	0	0	5
Exception Reporting	5	0	0	5
Final Report Output	5	0	0	5
Version Control	2	0	0	2

9. RESULTS

9.1 Performance Metrics

Performance Metrics of the Model is evaluated using Confusion Matrix

```
    cm = confusion_matrix(y_test, pred)
    print("CONFUSION MATRIX\n",cm)
    print("\nCLASSIFICATION REPORT\n",classification_report(y_test,pred))
    print("\nACCURACY SCORE : ",accuracy_score(y_test, pred))
```

```
CONFUSION MATRIX
 [[239   7]
 [ 51   3]]

CLASSIFICATION REPORT
              precision    recall  f1-score   support

          No       0.82      0.97      0.89       246
         Yes       0.30      0.06      0.09        54

    accuracy                           0.81       300
   macro avg       0.56      0.51      0.49       300
weighted avg       0.73      0.81      0.75       300


ACCURACY SCORE :  0.8066666666666666


ACCURACY OF THE MODEL IS 80.67%
```

10. ADVANTAGES & DISADVANTAGES

10.1 ADVANTAGES

Random Forest is based on the bagging algorithm and uses Ensemble Learning technique. It creates as many trees on the subset of the data and combines the output of all the trees. In this way it reduces overfitting problems in decision trees and also reduces the variance and therefore improves the accuracy.

Non linear parameters don't affect the performance of a Random Forest unlike curve based algorithms. So, if there is high nonlinearity between the independent variables, Random Forest may outperform as compared to other curve based algorithms.

Light GBM uses a histogram-based algorithm i.e it buckets continuous feature values into discrete bins which fasten the training procedure.

10.2 DISADVANTAGES

Random Forest creates a lot of trees (unlike only one tree in case of decision tree) and combines their outputs. By default, it creates 100 trees in the Python sklearn library. To do so, this algorithm requires much more computational power and resources. On the other hand decision tree is simple and does not require so much computational resources

Light GBM splits the tree leaf-wise which can lead to overfitting as it produces many complex trees.

11. CONCLUSION

On the whole, this project was a useful experience. We have gained new knowledge and skills and achieved several of my learning goals. We got insight into professional practice. We learned the different facets of working . We experienced that self exploration, as in many organizations, is an important factor for the progress of projects. Related to our study we learned more about employee attrition rate prediction and the various approaches and algorithms to achieve the same. There is still a lot to discover and to improve. The methods used at the

moment are still not standardized and a consistent method is in development. Furthermore we have experienced that it is of importance of each strategy and how the other one is better than the current algorithm and in which application. We found that the internship is not one-sided, but it is a way of sharing knowledge, ideas and opinions and implementing the same to get results. The internship was also good to find out what our strengths and weaknesses are. This helped me to define what skills and knowledge. We believe that our time spent in learning and surfing regarding various algorithms and the mathematics behind was well worth it and contributed to finding an acceptable solution to build a model and predict the employee's attrition rate. Two main things that we've learned are the importance of time-management skills and self-motivation. At last this project has given us new insights and motivation to pursue a career in the machine learning domain.

12. FUTURE SCOPE

The goal with employee attrition and retention is to strike the right balance of holding on to top talent while accepting that some level of attrition is healthy; employee attrition analytics enables organizations to find that balance.In future this model could take feedbacks from employees and analyze them to understand what the employees needs to keep the healthy balance intact.New and different models can be used to analyze the data leading to a improved accuracy of the prediction.

13. APPENDIX

Source Code

Link to the source code: Source Code Link Github

GitHub & Project Demo Link

Github Repository Link: Github Link
Project Demo Link: Demo Video Link