

Employee Attrition Rate and analysis

1. INTRODUCTION

Project Overview

The aim of the project is to analyze and predict employee attrition within a company using machine learning techniques. The goal is to identify factors that contribute to employee turnover and develop predictive models to forecast attrition

Purpose

The purpose of analyzing employee attrition rate using machine learning techniques is to gain insights into the factors that contribute to employee turnover within an organization. By leveraging machine learning algorithms, businesses can identify patterns and relationships in large datasets to predict and understand why employees are leaving. This analysis can help organizations take proactive measures to retain valuable employees, improve employee satisfaction, and reduce turnover rates.

2. LITERATURE SURVEY

Existing problem

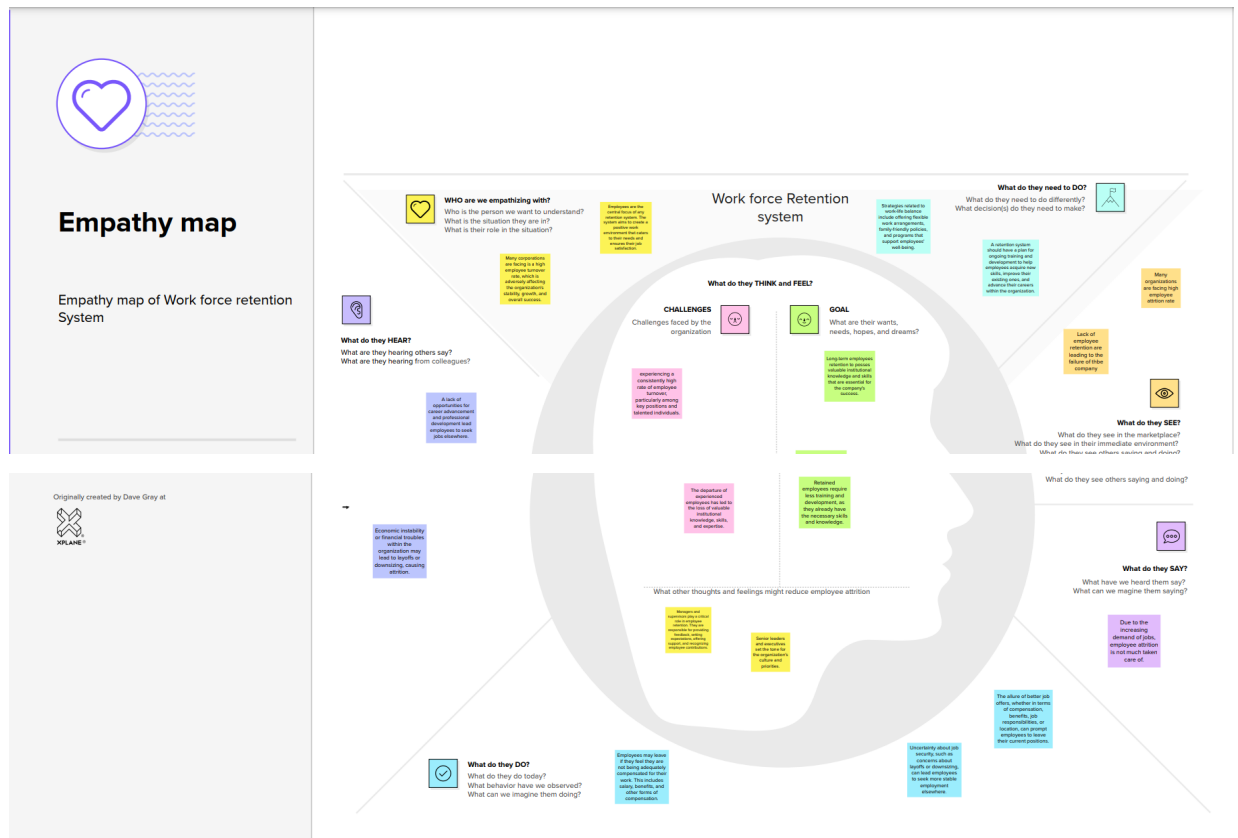
The existing problem of employee attrition rate refers to the challenge that organizations face in retaining their employees. High employee attrition rates can have negative consequences for businesses, including increased recruitment and training costs, reduced productivity, and a loss of institutional knowledge. Understanding the reasons behind employee attrition is crucial for organizations to implement effective retention strategies

Problem Statement Definition

Many well-established companies operating in a competitive industry, have been facing persistent challenges related to employee turnover. High employee attrition rates have resulted in increased recruitment and training costs, disruption of workflow, and a negative impact on overall productivity

3. IDEATION & PROPOSED SOLUTION

Empathy Map Canvas

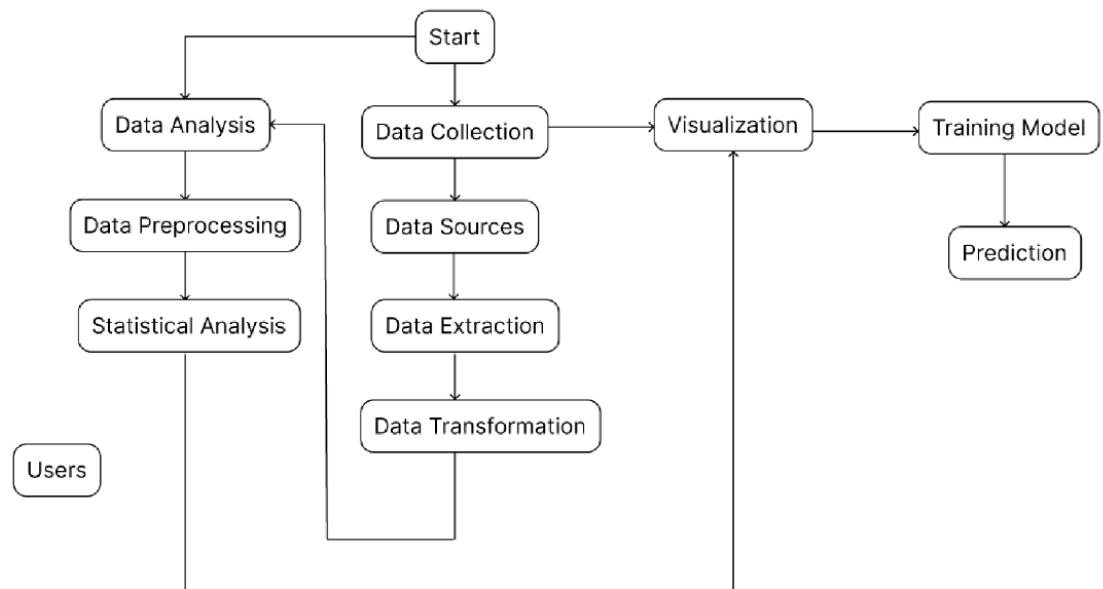


Ideation & Brainstorming

A machine learning model using predictive analytics will be developed analyzing previous data of employees attrition and job satisfaction. Based on the feedback and data analysis, develop a set of targeted retention strategies and initiatives.

4. PROJECT DESIGN

Data Flow Diagrams



User Stories

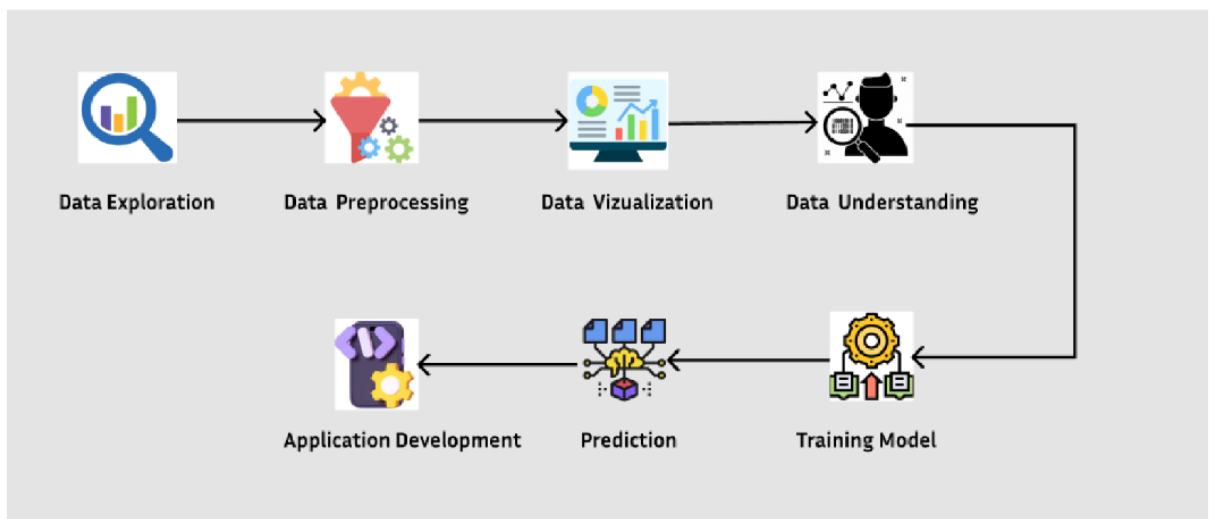
User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard Upon successful registration, a confirmation email is sent	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email I can click on the confirmation link in the email to verify my account.	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail	I can register and access the dashboard using Gmail Login.	Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password	I can access my account / dashboard after successful login.	High	Sprint-1
	Dashboard	USN-6	As a user (HR), I can view the employee status and well being.	- I can see statistics on employee status. - I can filter the displayed data. - Data is up-to-date and accurate.	High	Sprint-2
		USN-7	As a user (HR), I can the employee status between different domains.	- I can select domains for comparison. - The system provides clear and meaningful comparisons.	Medium	Sprint-2

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Web user)	Registration	USN-8	As a web user, I can register for the application by providing my email, password, and confirming the password.	I can access my account/dashboard upon successful registration.	High	Sprint-1
	Login	USN-9	As a web user, I can log into the application using my email and password.	After successful login, I can access my account/dashboard.	High	Sprint-1
	Dashboard	USN-10	As a web user, I can access a user-friendly dashboard that displays employee attrition statistics.	I can filter, customize, and visualize data effectively on the dashboard. The displayed data is up-to-date and accurate.	High	Sprint-2
Customer Care Executive	Access to User Data	USN-11	As a customer care executive, I can access user data for the purpose of assisting users.	Access is limited to user-related data and is in compliance with data protection regulations.	High	Sprint-2
	User Support	USN-12	As a customer care executive, I can provide support to users, answer queries, and resolve issues via a support interface.	Users receive timely and helpful assistance.	High	Sprint-2
Administrator	User Management	USN-13	As an administrator, I can manage user accounts, including adding, modifying, and deactivating accounts.	User data is kept up-to-date and secure.	High	Sprint-2
	Data Management	USN-14	As an administrator, I can manage data sources, ensuring data accuracy and integrity.	Data sources are regularly updated and reliable.	High	Sprint-2
	Security	USN-15	As an administrator, I can implement and manage security measures to protect user data and the system.	Data is secure and complies with relevant regulations.	High	Sprint-3

Solution Architecture

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.
- Define features, development phases, and solution requirements.
- Provide specifications according to which the solution is defined, managed, and delivered.



5. PROJECT PLANNING & SCHEDULING

Sprint Planning & Estimation

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Use the below template to create product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	2
Sprint-1	Registration	USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	2
Sprint-2	Registration	USN-3	As a user, I can register for the application through Gmail	2	Medium	2
Sprint-1	Login	USN-4	As a user, I can log into the application by entering email and password	1	High	2
Sprint-2	Dash board	USN-5	As a user, I can view the progress and status using the dashboard interface	2	High	2
Sprint-2	Dashboard	USN-6	As a user I can track the analysis of the ongoing research	2	high	2
Sprint-3	Result	USN-7	As a user , I can finally view my analysis report and also get the solution of the employee attrition rate with accuracy percentage.	1	High	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	20	6 Days	28 Oct 2023	29 Oct 2023	20	29 Oct 2023
Sprint-2	20	6 Days	31 Oct 2023	05 Nov 2023	20	06 NOV 2023
Sprint-3	20	6 Days	07 Nov 2023	11 Nov 2023	20	11 NOV 2023

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

#	Column	Non-Null Count	Dtype
0	Age	1470 non-null	int64
1	Attrition	1470 non-null	object
2	DailyRate	1470 non-null	int64
3	Department	1470 non-null	object
4	DistanceFromHome	1470 non-null	int64
5	Education	1470 non-null	int64
6	EducationField	1470 non-null	object
7	EnvironmentSatisfaction	1470 non-null	int64
8	Gender	1470 non-null	object
9	HourlyRate	1470 non-null	int64
10	JobInvolvement	1470 non-null	int64
11	JobLevel	1470 non-null	int64
12	JobRole	1470 non-null	object
13	JobSatisfaction	1470 non-null	int64
14	MaritalStatus	1470 non-null	object
15	MonthlyIncome	1470 non-null	int64
16	MonthlyRate	1470 non-null	int64
17	NumCompaniesWorked	1470 non-null	int64
18	OverTime	1470 non-null	object
19	PercentSalaryHike	1470 non-null	int64
20	PerformanceRating	1470 non-null	int64
21	RelationshipSatisfaction	1470 non-null	int64
22	StockOptionLevel	1470 non-null	int64
23	TotalWorkingYears	1470 non-null	int64
24	TrainingTimesLastYear	1470 non-null	int64
25	WorkLifeBalance	1470 non-null	int64
26	YearsAtCompany	1470 non-null	int64
27	YearsInCurrentRole	1470 non-null	int64
28	YearsSinceLastPromotion	1470 non-null	int64
29	YearsWithCurrManager	1470 non-null	int64

```
In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
In [17]: df=pd.read_csv("C:/Users/abine/Downloads/Employee-Attrition.csv")
```

```
In [66]: from sklearn.preprocessing import LabelEncoder
```

```
In [67]: le=LabelEncoder()
```

```
In [68]: columns_to_encode=['Attrition','Department','EducationField','Gender','JobRole','MaritalStatus','OverTime']
```

```
In [69]: for columns in columns_to_encode:
df[columns] = le.fit_transform(df[columns])
```

```
In [39]: df.head()
```

```
Out[39]:
```

	Age	Attrition	DailyRate	Department	DistanceFromHome	Education	EducationField	EnvironmentSatisfaction	Gender	HourlyRate	...	PerformanceRating
0	41	1	1102	2	1	2	1	2	0	94	...	
1	49	0	279	1	8	1	1	3	1	61	...	
2	37	1	1373	1	2	2	4	4	1	92	...	

3	33	0	1392	1	3	4	1	4	0	56	...
4	27	0	591	1	2	1	3	1	1	40	...

5 rows × 30 columns

```
In [40]: df.corr()['Attrition']
```

```
Out[40]: Age                -0.159205
Attrition                 1.000000
DailyRate                -0.056652
Department                0.063991
DistanceFromHome         0.077924
Education                -0.031373
EducationField            0.026846
EnvironmentSatisfaction  -0.103369
Gender                   0.029453
HourlyRate               -0.006846
JobInvolvement           -0.130016
JobLevel                 -0.169105
```

```
JobRole                 0.067151
JobSatisfaction          -0.103481
MaritalStatus            0.162070
MonthlyIncome            -0.159840
MonthlyRate              0.015170
NumCompaniesWorked       0.043494
OverTime                 0.246118
```

```
In [41]: from sklearn.preprocessing import MinMaxScaler
```

```
In [42]: scaler=MinMaxScaler()
```

```
In [43]: from sklearn.linear_model import LogisticRegression
```

```
In [44]: from sklearn.model_selection import train_test_split
```

```
In [45]: x=df.drop(columns=['Attrition'],axis=1)
```

```
In [46]: y=df['Attrition']
```

```
In [70]: X_train,X_test,Y_train,Y_test=train_test_split(x,y,test_size=0.3,random_state=101)
```

```
In [71]: X_train=scaler.fit_transform(X_train)
```

```
In [72]: X_test=scaler.transform(X_test)
```

```
In [73]: lr=LogisticRegression()
```

```
In [74]: lr.fit(X_train,Y_train)
```

```
Out[74]: LogisticRegression
LogisticRegression()
```

```
In [75]: prediction=lr.predict(X_test)
```

```
In [76]: from sklearn.metrics import confusion_matrix,classification_report
```

```
In [77]: print(confusion_matrix(Y_test,prediction))
```

```
[[361  10]
 [ 51  19]]
```

```
In [78]: print(classification_report(Y_test,prediction))
```

	precision	recall	f1-score	support
0	0.88	0.97	0.92	371
1	0.66	0.27	0.38	70
accuracy			0.86	441
macro avg	0.77	0.62	0.65	441
weighted avg	0.84	0.86	0.84	441

```
In [79]: from sklearn.metrics import accuracy_score
```

```
In [63]: accuracy_score(Y_test,prediction)
```

```
Out[80]: 0.8616780045351474
```

```
In [64]: import pickle
```

```
# Make sure the 'lr' object is defined
file_path = 'Employee Attrition.pkl'
with open(file_path, 'wb') as file:
    pickle.dump(lr, file)
```

7. PERFORMANCE TESTING

Performace Metrics

```
In [79]: >> from sklearn.metrics import accuracy_score
```

```
In [80]: >> accuracy_score(Y_test,prediction)
```

```
Out[80]: 0.8616780045351474
```

```
In [64]: >> import pickle

# Make sure the 'lr' object is defined

file_path = 'Employee Attrition.pkl'

with open(file_path, 'wb') as file:
    pickle.dump(lr, file)
```

8. RESULTS

Output Screenshots

9. ADVANTAGES & DISADVANTAGES

Advantages of using machine learning approaches to address the problem of employee attrition:

1. Data-driven insights: Machine learning algorithms can analyze vast amounts of employee data, including demographics, performance metrics, and job satisfaction surveys. This can provide valuable insights into the factors that contribute to attrition and help organizations make informed decisions.
2. Prediction and proactive measures: Machine learning models can be trained to predict the likelihood of employee attrition based on various parameters. This allows organizations to identify at-risk employees and take proactive measures to retain them, such as offering mentoring programs, career development opportunities, or personalized retention strategies.
3. Cost-effective: By identifying the factors that contribute to attrition, machine learning can help organizations allocate resources more effectively. This can result in cost savings by reducing turnover and minimizing recruitment and training expenses.

Disadvantages of using machine learning approaches to address the problem of employee attrition:

1. Data quality and availability: Machine learning models heavily rely on high-quality and relevant data. If the data used for training the models is incomplete, inaccurate, or biased, it can lead to unreliable predictions and ineffective strategies.

2. Ethical considerations: Machine learning algorithms are only as good as the data they are trained on. If the models are trained on biased data, they can perpetuate existing biases and discrimination, leading to unfair decisions or interventions in the workplace.

3. Human interpretation and intervention: While machine learning algorithms can provide insights and predictions, human input is still crucial for interpretation and decision-making. Organizations must ensure that human judgment and expertise are integrated into the process to avoid overly relying on automated systems and potentially missing important contextual factors.

10. CONCLUSION

Overall, machine learning offers tremendous potential in addressing the problem of employee attrition, but it should be used as a tool to augment human decision-making rather than replacing it entirely.

11. FUTURE SCOPE

Machine learning algorithms can be used to build predictive models that analyze various employee attributes, such as job satisfaction, performance metrics, and engagement levels, to forecast the likelihood of attrition. These models can help organizations identify high-risk employees or departments and take proactive measures to retain them.

12. APPENDIX

Source Code

GitHub & Project Demo Link

