A STUDY ON PREDICTION OF EMPLOYEE ATTRITION USING MACHINE LEARNING AT PLR PROJECTS PVT. LTD.

Submitted

In partial fulfilment of the requirements for the award of the degree of

MASTER OF BUSINESS ADMINISTRATION
(BUSINESS DATA ANALYTICS)

IN

SCHOOL OF MANAGEMENT STUDIES

BY

KATARI PAVAN

[Regd.No:220A1E0403]

Under the supervision of
Dr.T.NARAYANA REDDY

M.A(Eco.),MBA, Ph.D

Professor & Head, JNTUA SMS

SCHOOL OF MANAGEMENT STUDIES

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY
ANANTAPUR

ANANTAPURAMU-515001

ANDHRA PRADESH

INDIA

2022-2024

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY

ANANTAPUR – 515002

SCHOOL OF MANAGEMENT STUDIES

BONAFIDE CERTIFICATE

This is to certify that the project report, entitled **"A STUDY ON PREDICTION OF EMPLOYEE ATTRITION USING MACHINE LEARNING AT PLR PROJECTS PVT.LTD."** Submitted to the **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, ANANTAPUR** in Partial fulfilment of the requirements for the award of the Degree of **MASTER OF BUSINESS ADMINISTRATION IN BUSINESS DATA ANALYTICS (MBA-BDA)** is a record of original research work done by **KATARI PAVAN** , during the period 18TH January 2024 To 5th may 2024 of his study in the **MASTER OF BUSINESS ADMINISTATION** at **SCHOOL OF MANAGEMENT STUDIES, JAWAHARLAL NEHRU TECHNOLOGICAL UNIIVERSITY ANANTAPUR,** under the supervision of **Dr . T. NARAYANA REDDY M.A(Eco.),MBA, Ph.D Professor and HOD**. and the report has not formed the basis for the award of any Degree / Diploma / Associate ship / Fellowship or other similar title to any candidate of any University.

Internal guide

Head

JNTUA SMS

Viva voce examination held on: _____

Internal Examine

External Examiner

Date:
Place: ANANTAPURAMU.

PLR
PROJECTS

PLR Projects Pvt. Ltd.

TO WHOM SO EVER IT MAY CONCERN

This is to certify that Mr. Katari Pavan Reg No.220A1E0403student of MBA(BDA)

studying at Jawaharlal Nehru Technological University AnanthapurAs aPart of MBA(BDA)programme he has successfully completed his Project work on A study on prediction of employee attrition at PLR PROJECTS PVT LTD.

Duration :45 Days

During the above period he has been conducted his maximum satisfaction to the

company authorities .

PLR Projects Pvt. Ltd
Hot Mix Plant
Kallur. Chittoor (Dt.)

An ISO 9001-2008 Certified Company

Corporate Office : Inwinex Towers, 3B, 3rd Floor, 130, Road No.2, Plot No.2, Banjara Hills, Hyderabad - 500 034. A.P. India. Tel : 040-23540246, 23540247 Fax : 23540249 E-mail: info@plrprojects.com, www.plrprojects.com
Regd. Office : 10-34, Maruthi Nagar, M.R. Palli, Tirupati, Chittoor - 517 002. Ph : (Off) 2240088 (Resi)2241148, Fax : 0877-2242875

A STUDY ON PREDICTION OF EMPLOYEE ATTRITION USING MEACHINE LEARNING AT PLR PROJECTS PVT.LTD

ORIGINALITY REPORT

22%	19%	7%	16%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

5%

★ Submitted to Southern New Hampshire University - Continuing Education

Student Paper

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

HEAD
School of Management Studies
JNTUA Anantapur
Ananthapuramu-515002 (A.P.)

DECLARATION

I hereby declare that the project report entitled **"A STUDY ON PREDICTION OF EMPLOYEE ATTRITION USING MACHINE LEARNING AT PLR PROJECTS PVT. LTD."** is record of independent research work and is carried out by me during the period of my study, submitted to **SCHOOL OF MANAGEMENT STUDIES**, **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, ANANTPUR,** under the guidance of **Dr . T. NARAYANA REDDY Professor and HOD** M.A(Eco.),MBA, Ph.D. School of Management Studies, **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY** for the award of the degree of **MASTER OF BUSINESS ADMINISTRATION (2022-2024)** is my original work, has not been submitted to any other university for the award of any degree.

Date : KATARI PAVAN

Place: ANANTAPURAMU 220A1E0403

ACKNOWLEDGEMENT

I express my sincere gratitude to **Dr. T. NARAYANA REDDY PhD, MA(Eco),** Head, School of Management Studies for his constant support and encouragement.

I feel it a great privileges to express my sincere gratitude to my project guide

Dr . T. NARAYANA REDDY Professor and HOD School of Management Studies for his valuable guidance in preparation of this project work.

I also thank our faculty members **Dr. B.C. Lakshmanna, Dr. P. Basaiah, Dr. P. Jaya rami Reddy, K. Srilatha, Dr. D. Varalakshmi, Dr.P.Viswanath, Dr. P. Sravan Kumar, Dr. D. Jakir Hussain, Dr. K. Nagaraju.** for their keen interest and constructive criticism, which played vital role in completing my project on time

I would like to express my heartfelt thanks to all the respondents who gave valuable information to me.

I also extend my thanks to all my family members and my friends who helped me in completing this project work successfully.

KATARI PAVAN

220A1E0403

ABSTRACT

Employee attrition is costly and disrupts workflows. This project explores using machine learning models to predict which employees are at risk of leaving. We'll gather HR data on factors like job satisfaction, performance appraisal, and grievance appraisal system . After cleaning and preparing the data, we'll train various machine learning models to identify patterns associated with employee departures. Evaluating these models will help us choose the most accurate one for predicting future attrition. By pinpointing high-risk employees, this project aims to develop predictive model to predictive the employee attrition and to develop targeted retention strategies, fostering a happier and more stable workforce.

TABLE OF CONTENTS

LIST OF GRAPHS

LIST OF TABLES

LIST OF FIGURES AND GRAPHS

CHAPTER-1

INTRODUCTION

1.1 INTRODUCTION:

HR analytics is a data-driven approach to managing people at work. HR analytics, also known as people analytics, workforce analytics, or talent analytics, it revolves around analyzing people problems using data to answer critical questions about your organization. This enables better and data-driven decision-making.

Employee attrition is defined as employees leaving their organizations for unpredictable or uncontrollable reasons. Many terms make up attrition, the most common being are termination, resignation, planned or voluntary retirement, structural changes, long-term illness, layoffs.

Types of Attrition :

A. Voluntary attrition : It refers to employees leaving a company willingly, often for personal reasons or better opportunities elsewhere. It impacts workforce stability, productivity, and morale. Understanding its causes, such as job dissatisfaction or career advancement, is crucial for organizations to implement effective retention strategies and maintain a skilled workforce.

B. Involuntary attrition : It involves employees leaving a company due to factors beyond their control, such as layoffs, restructuring, or termination. It can result from economic downturns, organizational changes, or performance issues. Managing involuntary attrition requires careful planning, communication, and support to minimize negative impacts on both employees and the organization.

C. Compulsory attrition : Compulsory attrition involves reducing workforce size through involuntary means like layoffs or terminations. It's often used for cost-cutting or restructuring but can impact morale and organizational culture negatively.

D. Natural attrition: Natural attrition refers to the gradual reduction in the size of a workforce over time through voluntary means such as resignations, retirements, or employees leaving for other reasons. Unlike compulsory attrition, which involves involuntary actions like layoffs, natural attrition occurs as a result of employees making individual decisions to leave the organization.

Calculation of employee attrition :

$$\frac{\text{(Number of employee departures)}}{\text{(Average number of employees)}} \times 100 = \text{Attrition rate (percentage)}$$

Neural Networks :

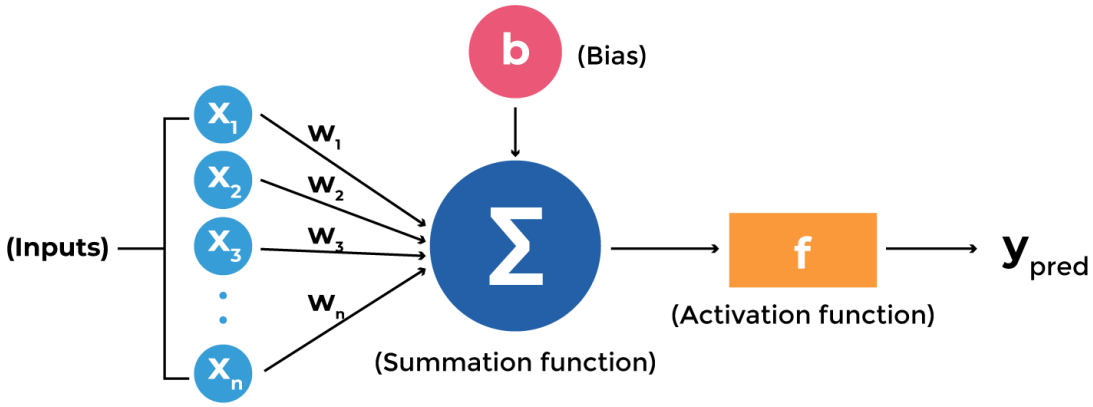


Fig. 1 Block Diagram of Neural Networks

Inputs : Inputs are the Values that to inserted to the neuron. Number of input neurons is depends on number of independent variables or Number of input variables.

Weights : Weights are the backbone of artificial neural networks, enabling them to learn from data and make predictions.

Bias : Bias is considered a systematic error that occurs in the machine learning model itself due to incorrect assumptions in the ML process. Technically, we can define bias as the error between average model prediction and the ground truth.

Activation Function : Activation function is the process of applying function to the neurons.

Summation Function : The summation function sums up all the inputs and adds bias to it.

Output : The output layer is the final layer in the neural network where desired predictions are obtained

Process of Neural Networks :

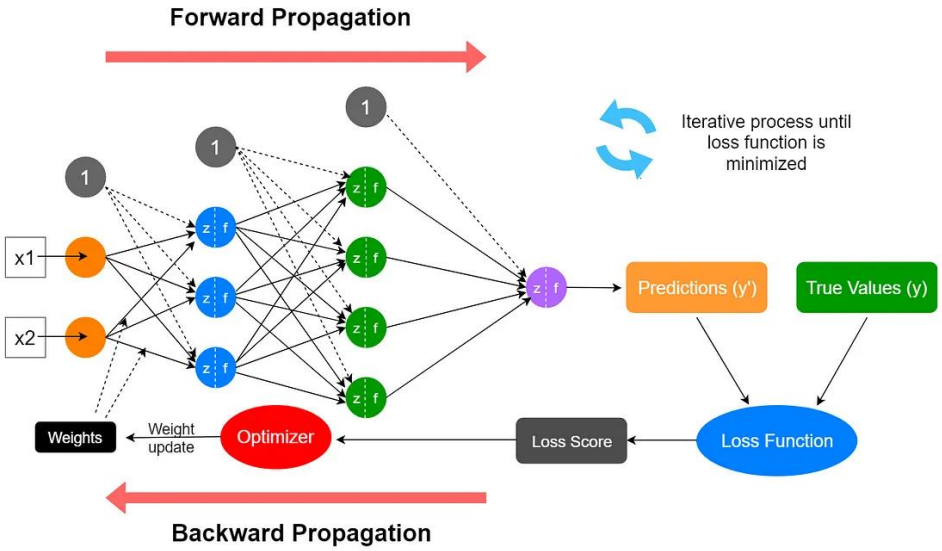


Fig. 2 Process of Neural Networks

Forward Propagation : Forward propagation is where input data is fed through a network, in a forward direction, to generate an output. The data is accepted by hidden layers and processed, as per the activation function, and moves to the successive layer.

Backward Propagation : Backpropagation is the process of reducing error's by adjusting or updating parameters is called as backward propagation.

Echoes : Echoes indicate that how many times the data should be processed.

Loss Function : The sum of error's from all the observation's is called loss function. It is also called as cost function or error function..

Optimizer : optimizer is a function or an algorithm that adjusts the attributes of the neural network, such as weights and learning rates. It helps in reducing the overall loss and improving accuracy.

Types of optimizer :

1) **Binary cross entropy :** Binary Cross Entropy is a loss function used in deep learning to measure the difference between predicted binary outcomes and actual binary labels. It quantifies the dissimilarity between probability distributions, aiding model training by penalizing inaccurate predictions. It's widely used in tasks like binary classification, where the goal is to categorize data into two classes.

2) **categorical cross entropy :** Categorical Cross Entropy is also known as Softmax Loss. It's a SoftMax activation plus a Cross-Entropy loss used for multiclass classification. Using this loss, we can train a Convolutional Neural Network to output a probability over the N classes for each image.

Types of Neural Networks :

1)Perceptron : Using of step function as activation function is called Perceptron. Perceptron is first artificial neuron.

2)Feed Forward Neural Networks(FNN) : Feed forward neural networks are type of neural networks in which nodes do not form loops/cycle. This type of neural network is also known as a multi-layer neural network as all information is only passed in forward direction only. During data flow, input layer receives data, which travel through hidden layers, and exit from output layers.

3)Convolution Neural Network (CNN) : Convolution Neural Network are useful for finding patterns in images to recognize objects, classes, and categories. They can also be quite effective for classifying audio, time-series, and signal data. Convolution Neural Network learns directly from data.

4)Recurrent Neural Networks (RNN) : Recurrent Neural Network is a type of Neural Network where the output from the previous step is fed as input to the current step.

5)Long Short Term Memory Networks (LSTM) : Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) capable of learning and remembering information over time. This makes it a powerful algorithm for tasks like natural language processing, speech identification, and time series forecasting.

Types of Activations Functions :

Activation Function	Formula	Range	Domain	Suitable Layer
Sigmoid	$f(x) = \frac{1}{1+e^{-x}}$	(0,1)	(-∞,∞)	Output layer for binary classification
Tanh	$f(x) = \tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$	(-1,1)	(-∞,∞)	Hidden layers in neural networks
ReLU	$f(x) = \max(0, x)$	[0, ∞)	(-∞,∞)	Hidden layers in deep neural networks
Leaky ReLU	$f(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha x & \text{if } x \leq 0 \end{cases}$ Where α is a small positive constant (typically, a small value like 0.01)	(-∞,∞)	(-∞,∞)	Hidden layers, where a small positive slope is desired
ELU	$\text{ELU}(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha(e^{x} - 1), & \text{otherwise} \end{cases}$	(- α,∞)	(-∞,∞)	Hidden layers with potential to mitigate dead neurons
PReLU	$\text{PReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha x, & \text{otherwise} \end{cases}$ In PReLU, unlike Leaky ReLU where the slope(α)is fixed, here , α is learned during training.	(- α,∞)	(-∞,∞)	Hidden layers with adaptability to negative slopes
Softmax	$\text{softmax}(x_i) = \frac{e^{x_i}}{\sum_j^n e^{x_j}}$	(0,1)	(-∞,∞)	Output layer for multi-class classification
SELU	$\text{SELU}(x) = \lambda \times \begin{cases} x, & \text{if } x > 0 \\ \alpha e^{x} - \alpha, & \text{otherwise} \end{cases}$ where α is the scale parameter and λ is the stability parameter	(λ·α,∞)	(−∞,∞)	Hidden layers in deep neural networks

Table-4.1 Types of Activation functions

Logistic regression :

Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analysing the relationship between one or more existing independent variables. Logistic regression is a statistical method used for binary classification tasks. It models the relationship between a set of independent variables and a binary outcome using the logistic function. Unlike linear regression, logistic regression predicts the probability that an instance belongs to a specific class (typically 0 or 1). The logistic function transforms the output of a linear combination of input features into a value between 0 and 1, representing the probability. Parameters are estimated through techniques like maximum likelihood estimation or gradient descent. Logistic regression is widely applied in various fields due to its simplicity, interpretability, and effectiveness in binary classification tasks.

Random Forest :

Random Forest is a powerful ensemble learning technique used for classification and regression tasks. It constructs multiple decision trees during training and combines their predictions to produce a final output. Each tree is trained on a random subset of the data and features, reducing overfitting and improving generalization. Random Forest utilizes the wisdom of crowds by aggregating predictions from diverse trees, yielding robust results. It's highly scalable, handles high-dimensional data well, and is less prone to overfitting compared to individual decision trees. Random Forest finds applications in various domains such as finance, healthcare, and marketing, owing to its versatility and performance.

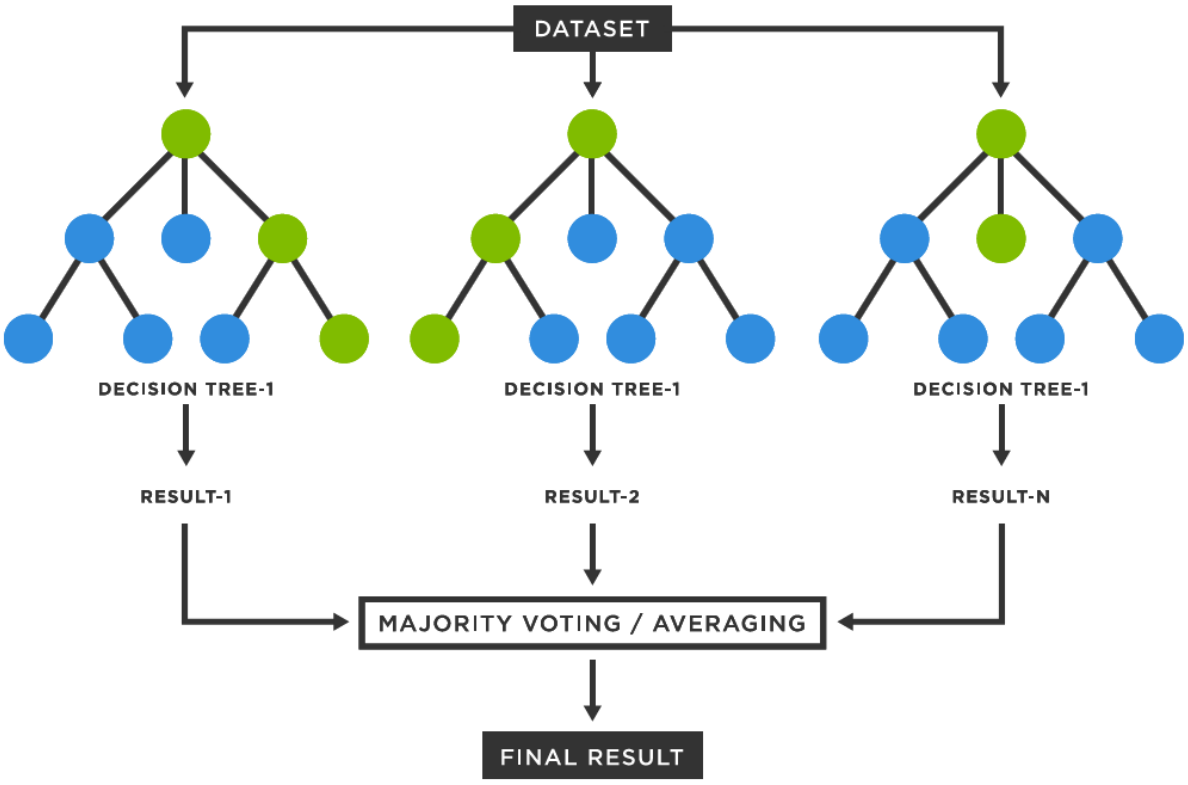


Fig.4 Random forest

1.2 REVIEW OF LITERATURE

A review of literature on employee attrition in HR analytics provides valuable insights into the factors influencing employee turnover, the methodologies employed for analysis, and the strategies adopted for retention. Numerous studies have examined the multidimensional nature of attrition, considering both individual and organizational factors.

Firstly, research consistently highlights the importance of demographic variables such as age, gender in predicting attrition. For example, younger employees and those with shorter tenures are often found to have higher turnover rates. Additionally, job-related factors such as job satisfaction, workload, and appraisal system are significantly impact attrition.

Methodologically, studies employ various statistical techniques to analyse attrition patterns. Machine learning algorithms such as logistic regression, random forests and deep learning model such as feed forward neural networks are commonly utilized for prediction and classification tasks. These methods enable researchers to identify key predictors of attrition and develop robust models for forecasting turnover.

Furthermore, literature emphasizes the role of HR analytics in proactively managing attrition. Data-driven approaches enable organizations to leverage internal and external data sources to identify at-risk employees and implement targeted retention interventions. For instance, sentiment analysis of employee feedback, social network analysis, and predictive modelling are employed to detect early warning signs of attrition and inform strategic HR decisions.

Moreover, research highlights the reasons that are making to employee attrition. By addressing the underlying causes of attrition, organizations can enhance employee engagement, satisfaction, and loyalty, thereby reducing turnover rates.

Overall, the literature on employee attrition in HR analytics underscores the significance of understanding the complex interplay of individual, job-related, and organizational factors driving turnover. By employing advanced analytical techniques and leveraging HR data, organizations can develop proactive retention strategies to mitigate attrition and foster a more engaged and productive workforce.

CHAPTER – 2

INDUSTRY PROFILE

AND

COMPANY PROFILE

2.1 INDUSTRY PROFILE :

The infrastructure industry encompasses a wide range of sectors crucial for societal functionality and economic development. It includes transportation (roads, bridges, airports, railways), energy (power plants, transmission lines), water and sanitation (dams, pipelines, sewage systems), and telecommunications (internet, mobile networks).

Infrastructure plays a foundational role in supporting commerce, facilitating trade, and improving living standards. In recent years, there has been a growing emphasis on sustainable infrastructure development, incorporating environmentally friendly practices and resilience against climate change impacts.

Governments worldwide are investing in infrastructure projects to stimulate economic growth, create jobs, and address aging infrastructure challenges. The India Infrastructure Sector Market size is estimated at USD 204.06 billion in 2024, and is expected to reach USD 322.27 billion by 2029, growing at a CAGR of 9.57% during the forecast period (2024-2029).

2.2 COMPANY PROFILE:

Name of the organization	PLR PROJECTS PVT.LTD
Founder	P.Padmavatiamma
Founded year	2006
Managing director	G.Swarnalatha
Turnover	350 crores
Employees	650
Location	Maruthi Nagar, M.R palli circle, Tirupati.

Source : Company website https://www.plrprojects.com

CHAPTER – 3

RESEARCH METHODOLOGY

3.1 NEED OF THE STUDY

The study on employee attrition aims to proactively identify factors influencing attrition, enabling organizations to implement targeted interventions for improved workforce management and long-term organizational success. The present study is taken for analysis of employee attrition using machine learning at PLR PROJECTS PVT.LTD.

3.2 SCOPE OF THE STUDY

The study covers to analyze employee attrition patterns, Enabling organization to optimize workforce management and foster a more engaged and stable workforce at PLR PROJECTS PVT.LTD.

3.3 OBJECTIVES OF THE STUDY

- To study various factors contributing to employee attrition at PLR PROJECTS PVT.LTD.
- To develop an predictive model to predict employee attrition at PLR PROJECTS PVT.LTD.
- To analyze employee attrition using machine learning at PLR PROJECTS PVT.LTD.

3.4 HYPOTHESIS

Null Hypothesis (H0) :

There is no significant difference between employee attrition actual values and predicted values.

Alternative Hypothesis (H1) :

There is a significant difference between employee attrition actual values and predicted values.

3.5 RESEARCH METHODOLOGY

- The study is based on both primary data and secondary data.
- The primary data is collected through structured questionnaire.
- The secondary data is collected from company annual reports, website.
- Website : https://www.plrprojects.com

Sampling technique	:	Random sampling
Sample size	:	549
Methods of data analysis	:	Multivariant analysis

3.6 TOOLS AND TECHNIQUES

Tool	:	Machine learning,
		Tableau.
Techniques	:	Logistic regression,
		Feedforward neural networks,
		Random forest,
		Tableau functions.

3.7 LIMITATIONS OF THE STUDY

- The developed models may not give accurate results all the time.
- The study is confined to employee attrition at PLR PROJECTS PVT. LTD.

CHAPTER – 4

DATA ANALYSIS

AND

INTERPRETATION

```
# importing Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from sklearn.metrics import accuracy_score, classification_report
```

```
[2] # uploading Data file
    Data = pd.read_csv('Employee Attrition  (Responses) - Form Responses 1-2.csv')
```

```
[3] Data.head(5)
```

	Timestamp	Employee ID	Age	Year of Joining	Gender	Education Qualification	Marital Status	Employee Department	Job satisfaction	Job involment	Satisfaction on work environment	Sa
0	NaN	PLR001	44	2018	Female	Under Graduation	Unmarried	Engineering	5	2	5	
1	NaN	PLR002	28	2023	Female	Under Graduation	Unmarried	Engineering	5	4	5	
2	NaN	PLR003	29	2018	Female	Under Graduation	Unmarried	Engineering	5	1	5	
3	NaN	PLR004	30	2022	Female	Under Graduation	Married	Engineering	5	5	5	
4	NaN	PLR005	41	2023	Female	Under Graduation	Married	Engineering	1	2	1	

```
Data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 549 entries, 0 to 548
Data columns (total 19 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Timestamp                       8 non-null      object
 1   Employee ID                     549 non-null    object
 2   Age                             549 non-null    int64
 3   Year of Joining                 549 non-null    int64
 4   Gender                          549 non-null    object
 5   Education Qualification         549 non-null    object
 6   Marital Status                  549 non-null    object
 7   Employee Department             549 non-null    object
 8   Job satisfaction                549 non-null    int64
 9   Job involment                   549 non-null    int64
 10  Satisfaction on work environment  549 non-null  int64
 11  Satisfaction of package         549 non-null    int64
 12  Relation with colleagues        549 non-null    int64
 13  Performance appraisal system    549 non-null    int64
 14  Grivenance appraisal system     549 non-null    int64
 15  Work life balance               549 non-null    int64
 16  Promotion                       549 non-null    object
 17  Attrition                       549 non-null    object
 18  Reason for attrition            122 non-null    object
dtypes: int64(10), object(9)
memory usage: 81.6+ KB
```

```
#Grouping of age
Data['Age_Group'] = pd.cut(Data['Age'], bins=[18, 25, 35, 45, 55, 65], labels=['21-26', '27-32', '33-38', '39-44', '44-50'])
Data.head()
```

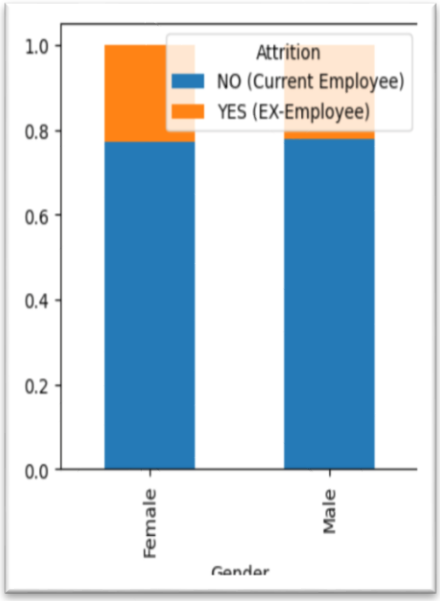
	Employee Department	Job satisfaction	Job involment	Satisfaction on work environment	Satisfaction of package	Relation with colleagues	Performance appraisal system	Grivenance appraisal system	Work life balance	Promotion	Attrition	Reason for attrition	Age_Group
	Engineering	5	2	5	4	5	4	5	1	NO	YES (EX-Employee)	Family Problem	33-38
	Engineering	5	4	5	4	5	4	5	1	NO	YES (EX-Employee)	Family Problem	27-32
	Engineering	5	1	5	4	5	4	5	2	NO	YES (EX-Employee)	Ineffective Grievance Appraisal system	27-32
	Engineering	5	5	5	5	5	5	1	2	NO	YES (EX-Employee)	Ineffective Grievance Appraisal system	27-32
	Engineering	1	2	1	4	1	4	1	4	NO	YES (EX-Employee)	Ineffective Grievance Appraisal system	33-38

Objective-1:

To study various factors contributing to employee attrition at PLR PROJECTS PVT.LTD.

```
#Bivariant Analysis Categorical Data variable Vs Target Variables
Gender = pd.crosstab(Data['Gender'],Data['Attrition'])
Gender.div(Gender.sum(1).astype(float),axis=0).plot(kind='bar',stacked=True,figsize=(4,4))

Employee_Department= pd.crosstab(Data['Employee Department'],Data['Attrition'])
Employee_Department.div(Employee_Department.sum(1).astype(float),axis=0).plot(kind='bar',stacked=True,figsize=(4,4))

Performance_appraisal_system= pd.crosstab(Data['Performance appraisal system'],Data['Attrition'])
Performance_appraisal_system.div(Performance_appraisal_system.sum(1).astype(float),axis=0).plot(kind='bar',stacked=True,figsize=(4,4))

Satisfaction_on_work_environment= pd.crosstab(Data['Satisfaction on work environment'],Data['Attrition'])
Satisfaction_on_work_environment.div(Satisfaction_on_work_environment.sum(1).astype(float),axis=0).plot(kind='bar',stacked=True,figsize=(4,4))

Grivenance_appraisal_system= pd.crosstab(Data['Grivenance appraisal system'],Data['Attrition'])
Grivenance_appraisal_system.div(Grivenance_appraisal_system.sum(1).astype(float),axis=0).plot(kind='bar',stacked=True,figsize=(4,4))
```

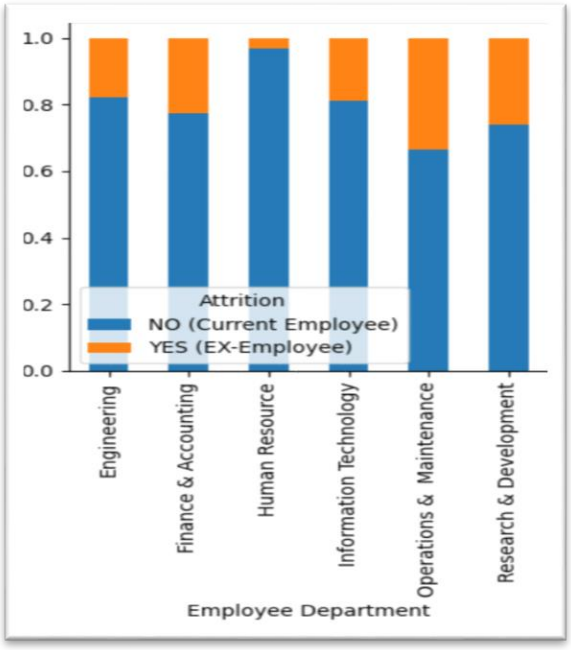
Output :



Graph **: 4.1** Gender vs Attrition

Interpretation :

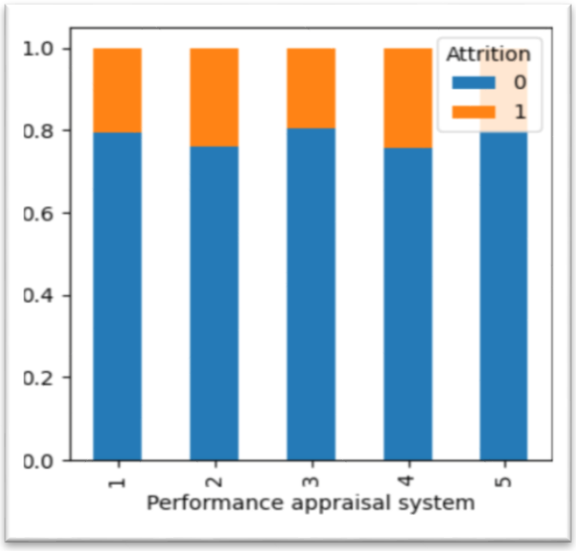
- From above graphs we can observe that attrition in male is high as compared to female.



Graph : 4.2 Employee department Vs Attrition

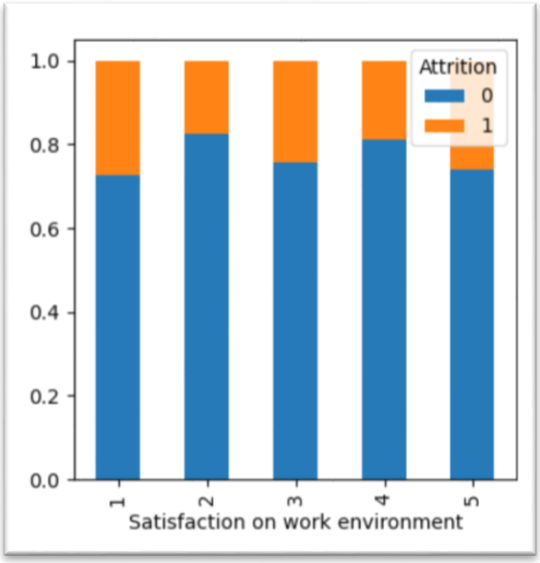
Interpretation :

- Operations and maintenance department as high attrition level.



Graph : 4.3 Performance appraisal system

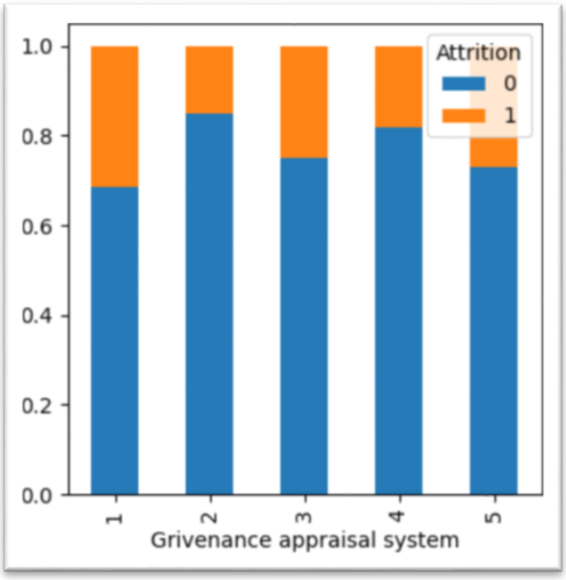
Interpretation : Most of the employees rated performance appraisal system with 4 and 5 rating.



Graph : 4.4 Work environment Vs Attrition

Interpretation :

Most of employees rated that satisfaction on work environment as satisfied.



Graph : 4.5 Grievance appraisal system Vs Attrition

Interpretation :

From grievance appraisal system graph shows that most of the employees are dissatisfied and very dissatisfied with grievance appraisal system.

Data pre-processing :

One hat encoding categorical data:

```
Data.head(5)
```

Gender	Education Qualification	Marital Status	Employee Department	Job satisfaction	Job involment	Satisfaction on work environment	Satisfaction of package	Relation with colleagues	Performance appraisal system	Grivenance appraisal system	Work life balance
0	1	0	1	5	2	5	4	5	4	5	1
0	1	0	1	5	4	5	4	5	4	5	1
0	1	0	1	5	1	5	4	5	4	5	2
0	1	1	1	5	5	5	5	5	5	1	2
0	1	1	1	1	2	1	4	1	4	1	4

Table 4.2 One hat encoding

Source : Machine learning

Exploratory data Analysis

```
Data.describe()
```

	Age	Year of Joining	Job satisfaction	Job involment	Satisfaction on work environment	Satisfaction of package	Relation with colleagues	Performance appraisal system	Grivenance appraisal system	Work life balance
count	549.000000	549.000000	549.000000	549.000000	549.000000	549.000000	549.000000	549.000000	549.000000	549.000000
mean	33.792350	2018.750455	3.754098	3.748634	3.739526	3.734062	3.774135	3.737705	3.737705	3.710383
std	7.822195	2.776601	1.233507	1.236099	1.224560	1.235997	1.226229	1.215944	1.206906	1.235833
min	21.000000	2014.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
25%	27.000000	2017.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000
50%	34.000000	2019.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000
75%	40.000000	2021.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000
max	47.000000	2023.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

Table 4.3 Exploratory data Analysis

Source : Machine learning

Objective -2 :

- **To develop the predictive model to predict employee attrition at PLR PROJECTS PVT.LTD.**

Model Building : Logistic regression

Model performance evaluation :

```
# Model performance evaluation
print("Accuracy Score:", accuracy_score(y_test, y_pred))
print("Classification Report:")
print(classification_report(y_test, y_pred))
```

```
Accuracy Score: 0.7727272727272727
Classification Report:
              precision    recall  f1-score   support

           0       0.77      1.00      0.87        85
           1       0.00      0.00      0.00        25

    accuracy                           0.77       110
   macro avg       0.39      0.50      0.44       110
weighted avg       0.60      0.77      0.67       110
```

Interpretation :

Logistic regression model is capturing the relationship between data variables with an accuracy of 77%.

Model Building : Feed Forward neural networks

```
[ ] X = Data.drop('Attrition', axis=1)
    y = Data['Attrition']

[ ] # Split the data into train and test sets
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

[ ] # Standardize the features
    scaler = StandardScaler()
    X_train_scaled = scaler.fit_transform(X_train)
    X_test_scaled = scaler.transform(X_test)

    # Define the neural network architecture
    model = Sequential([
        Dense(64, activation='relu', input_shape=(14,)),
        Dense(32, activation='relu'),
        Dense(1, activation='sigmoid')
    ])
```

```
[ ] # Compile the model
    model.compile(optimizer='adam',
                  loss='binary_crossentropy',
                  metrics=['accuracy'])

[ ] y_train = tf.keras.utils.to_categorical(y_train, num_classes=2)

[ ] print(model)

    <keras.src.engine.sequential.Sequential object at 0x7a139cde8730>

[ ] print(X_train_scaled.shape)
    print(y_train.shape)

    (439, 14)
    (439, 2)

[ ] y_pred = model.predict(X_test_scaled)
    y_pred = np.argmax(y_pred, axis=1)

    4/4 [==============================] - 0s 3ms/step
```

```
# Save the model
model.save('attrition_model.h5')

# Load the model
model = tf.keras.models.load_model('attrition_model.h5')

# Make predictions on new data
new_data = X_test_scaled[:5]
predictions = model.predict(new_data)

# Print the predictions
for i, prediction in enumerate(predictions):
  if prediction > 0.5:
    print(f"Employee {i+1}: Attrition Risk")
  else:
    print(f"Employee {i+1}: No Attrition Risk")

1/1 [==============================] - 0s 56ms/step
Employee 1: No Attrition Risk
Employee 2: No Attrition Risk
Employee 3: No Attrition Risk
Employee 4: Attrition Risk
Employee 5: No Attrition Risk
```

Model performance evaluation :

```
[ ] from sklearn.metrics import accuracy_score, classification_report

[ ] y_true = y_test.astype(int)
    y_pred = y_pred.astype(int)

    print("Accuracy Score:", accuracy_score(y_true, y_pred))
    print("Classification Report:")
    print(classification_report(y_true, y_pred))

    Accuracy Score: 0.7727272727272727
    Classification Report:
                  precision    recall  f1-score   support

               0       0.77      1.00      0.87        85
               1       0.00      0.00      0.00        25

        accuracy                           0.77       110
       macro avg       0.39      0.50      0.44       110
    weighted avg       0.60      0.77      0.67       110
```

Interpretation :

Using of sigmoid function at outer layer gives output in binary form. Feedforward neural network gives accuracy score of 0.77% indicating that it correctly predicts outcomes for 77% of the data it was tested on.

Model Building : Random forest

```
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(Data.drop('Attrition',axis=1), Data['Attrition'], test_size=0.25, random_state=0)

# Standardize the data
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
X_train = imputer.transform(X_train)
X_test = imputer.transform(X_test)

# Create a Random Forest classifier
classifier = RandomForestClassifier(n_estimators=100, criterion='entropy', random_state=0)

# Train the classifier
classifier.fit(X_train, y_train)

from sklearn.impute import SimpleImputer
imputer = SimpleImputer(missing_values=np.nan, strategy='mean')
imputer.fit(X_train)
# Make predictions on the test set
y_pred = classifier.predict(X_test)
```

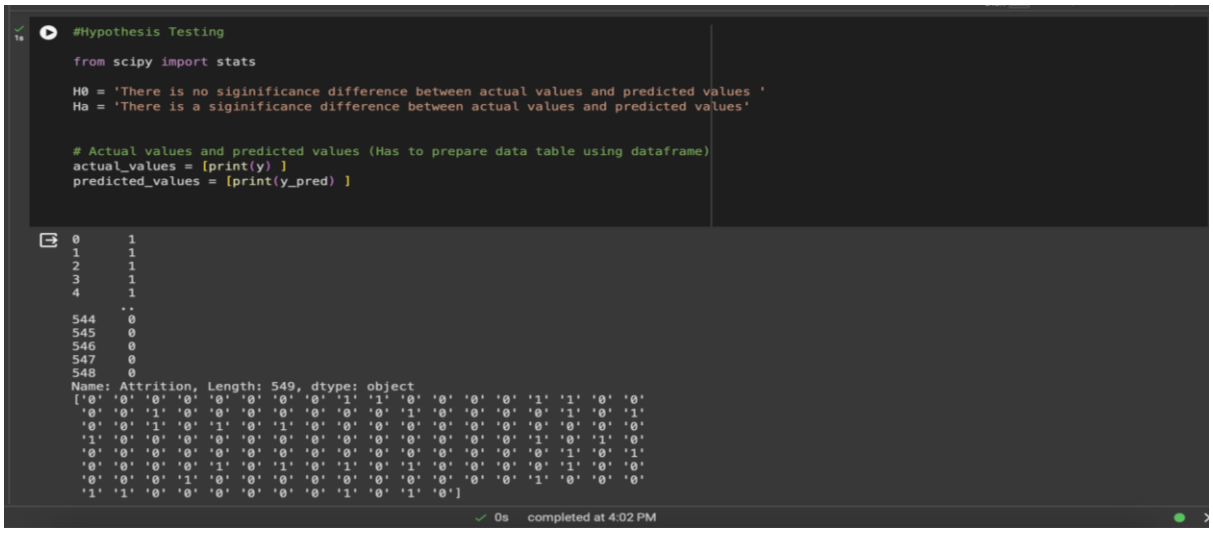
Model performance evaluation :

```
[46] classifier.fit(X_train, y_train)
     y_pred = classifier.predict(X_test)
     accuracy = accuracy_score(y_test, y_pred)
     print('Accuracy:', accuracy)
     print(classification_report(y_test, y_pred))

     Accuracy: 0.9782608695652174
                  precision    recall  f1-score   support

               0       0.97      1.00      0.99       108
               1       1.00      0.90      0.95        30

        accuracy                           0.98       138
       macro avg       0.99      0.95      0.97       138
    weighted avg       0.98      0.98      0.98       138
```

Interpretation :

Random forest model gives an accuracy of 97%. It means that the model is effectively capturing patterns and relationships in the data it was trained on.

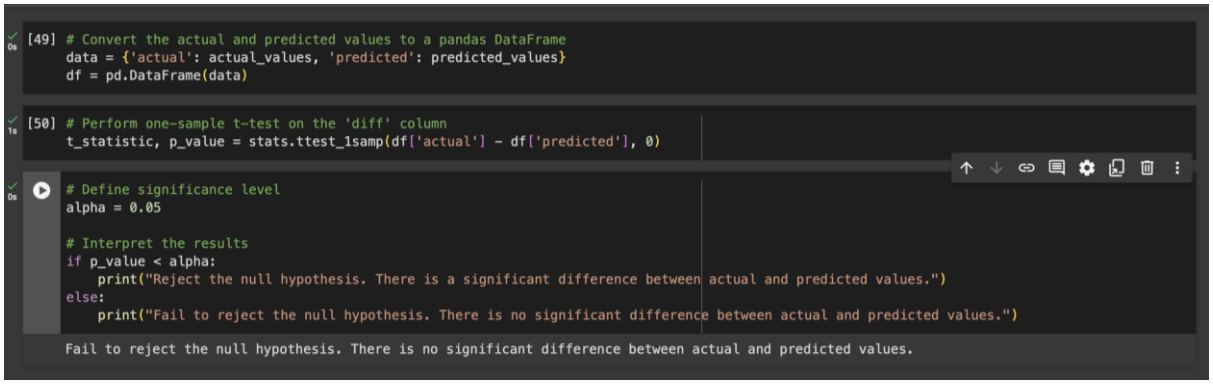
Hypothesis Testing :



```
#Hypothesis Testing

from scipy import stats

H0 = 'There is no siginificance difference between actual values and predicted values '
Ha = 'There is a siginificance difference between actual values and predicted values'


# Actual values and predicted values (Has to prepare data table using dataframe)
actual_values = [print(y) ]
predicted_values = [print(y_pred) ]
```

```
0      1
1      1
2      1
3      1
4      1
      ..
544    0
545    0
546    0
547    0
548    0
Name: Attrition, Length: 549, dtype: object
['0' '0' '0' '0' '0' '0' '0' '0' '1' '1' '0' '0' '0' '0' '1' '1' '0' '0'
 '0' '0' '1' '0' '0' '0' '0' '0' '0' '0' '1' '0' '0' '0' '0' '1' '0' '1'
 '0' '0' '1' '0' '1' '0' '1' '0' '0' '0' '0' '0' '0' '0' '0' '0' '0' '0'
 '1' '0' '0' '0' '0' '0' '0' '0' '0' '0' '0' '0' '0' '0' '1' '0' '1' '0'
 '0' '0' '0' '0' '0' '0' '0' '0' '0' '0' '0' '0' '0' '0' '1' '0' '1' '1'
 '0' '0' '0' '0' '1' '0' '1' '0' '1' '0' '1' '0' '0' '0' '0' '1' '0' '0'
 '0' '0' '0' '1' '0' '0' '0' '0' '0' '0' '0' '0' '0' '0' '1' '0' '0' '0'
 '1' '1' '0' '0' '0' '0' '0' '0' '1' '0' '1' '0']
```

0s completed at 4:02 PM



```
[49] # Convert the actual and predicted values to a pandas DataFrame
     data = {'actual': actual_values, 'predicted': predicted_values}
     df = pd.DataFrame(data)

[50] # Perform one-sample t-test on the 'diff' column
     t_statistic, p_value = stats.ttest_1samp(df['actual'] - df['predicted'], 0)

     # Define significance level
     alpha = 0.05

     # Interpret the results
     if p_value < alpha:
         print("Reject the null hypothesis. There is a significant difference between actual and predicted values.")
     else:
         print("Fail to reject the null hypothesis. There is no significant difference between actual and predicted values.")

     Fail to reject the null hypothesis. There is no significant difference between actual and predicted values.
```

Interpretation : Results of hypothesis testing stating that there is no significant difference between actual values and predicted values. It means we have to accept the null hypothesis.

Objective -3

- To analyze employee attrition at PLR PROJECTS PVT.LTD.

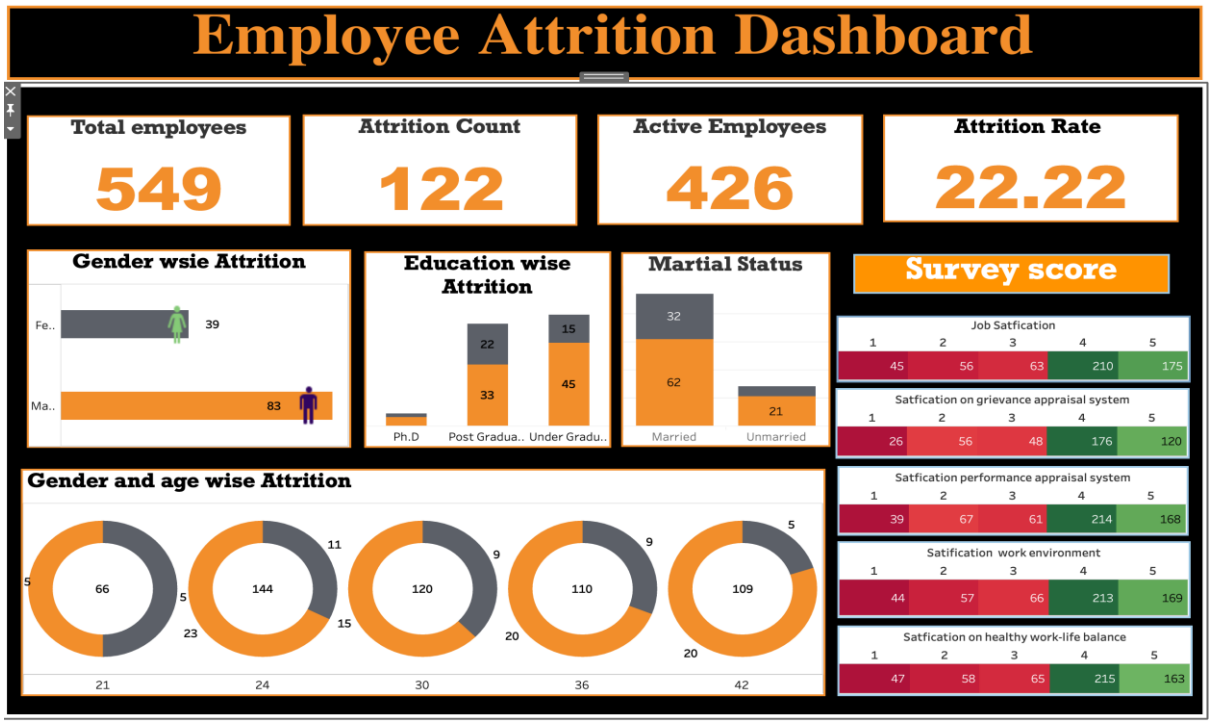


Fig.4.4 Employee Attrition Dashboard

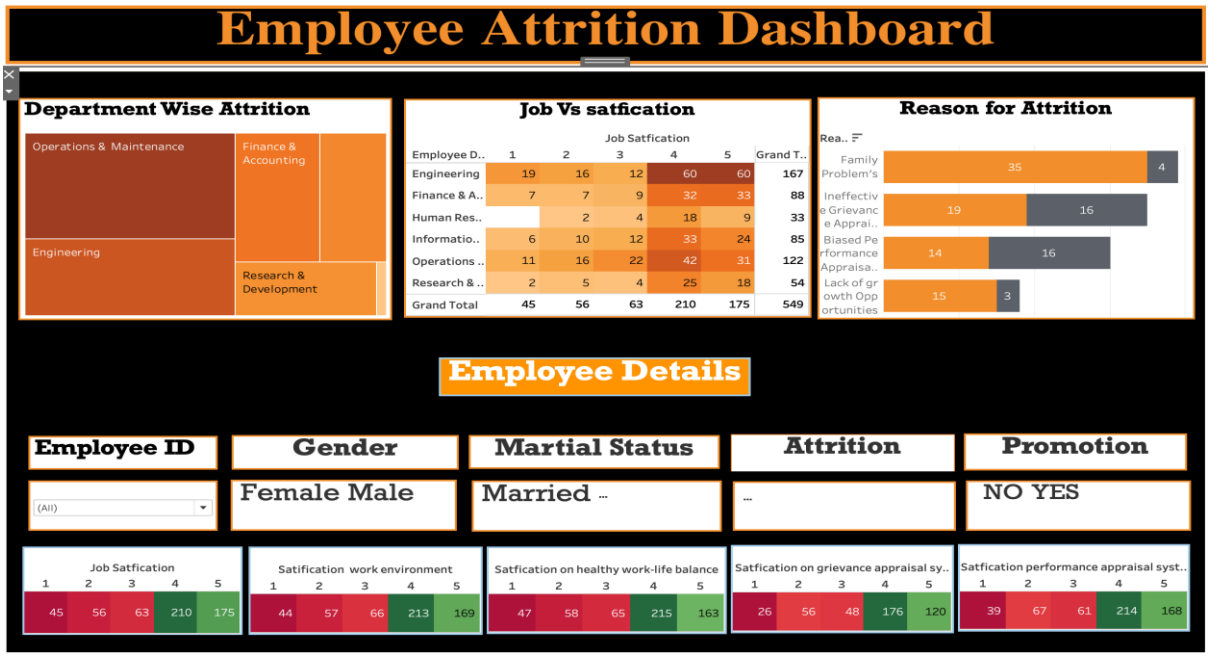


Fig.4.4.1 Employee Attrition Dashboard

Interpretation :

- Employee attrition count is 122 and employee attrition rate is 22.22.
- Operations & Maintenance has the highest number of employees leaving the company (41), followed by Engineering (30) and Finance & Accounting (20).
- The most common reason for employees leaving is family problem(39), followed by ineffective grievance appraisal system (35), and biased performance appraisal system (30).
- Employees in Engineering are the most satisfied with their jobs, followed by Information Technology (33) and Human Resources (33).

CHAPTER – 5

FINDINGS,

SUGGESTIONS &

CONCLUSION

5.1 FINDINGS OF THE STUDY :

- By Conducting Bi-variant analysis identified that job satisfaction, performance appraisal system, and work-life balance factors are strongly associated with employee attrition.
- Random forest model gives accuracy score of 0.97,which indicates that random forest is best model among remaining models Logistic regression 77% and Feedforward neural networks 77%.
- Developed machine learning models achieved accuracy of 97% in predicting employee attrition, indicating the effectiveness of the predictive model.
- Employee attrition count is 122 and attrition rate is 22.22 %.
- Reason behind the most of the employees leaving organization is due to family problems (39).
- Discovered that operations and maintenance departments have higher attrition rates due to factors like workload, ineffective grievance appraisal system (35).

5.2 SUGGESTIONS

- The project establishes a data-driven approach to employee retention. This allows company to move beyond intuition and focus on measurable factors that impact employee decisions.
- Most of employee attrition is happen due to ineffective grievance appraisal system and performance appraisal system company has to relook and solve this both issues.
- Company can introduce HR analytics for all people related activities.
- It suggested that to Conduct exit interviews with departing employees to understand their reasons for leaving and gather valuable feedback. Analyze this feedback to identify common themes and areas for improvement within the organization.
- Regularly assess employee satisfaction and engagement through surveys and feedback mechanisms. Use this data to identify areas for improvement and implement changes accordingly.

5.3 CONCLUSION

Our project on employee attrition at PLR projects Pvt. Ltd., has provided actionable insights for the organization, enabling proactive intervention strategies to mitigate attrition risks. Through comprehensive data analysis, we have identified several key predictors of employee attrition, including but not limited to job satisfaction, tenure, salary, and performance ratings. The project highlight factors influencing employee attrition. This knowledge allows companies to develop targeted retention strategies, fostering a more engaged and productive workforce. We have successfully developed a robust predictive model that utilizes machine learning algorithms to forecast employee attrition with a high degree of accuracy. This model serves as a valuable tool for identifying employees at risk of leaving the organization. By leveraging these insights, the company can implement targeted retention initiatives tailored to the needs of at-risk employees. By identifying employees at risk of leaving, companies can take proactive steps to retain them. This can save significant costs associated with recruitment, training, and lost productivity. The project establishes a data-driven approach to employee retention. This allows company to move beyond intuition and focus on measurable factors that impact employee decisions.

BIBLOGRAPHY

BIBLOGRAPHY

1) Kishori Singh, Reetu Singh. A study on employee attrition : Effects and causes. International journal of research in engineering, science and management Volume-2,Issue-8,August-2019.
2) Chandrakant Varma, Dr Chandrahauns R Chavan. A case of HR analytics – To understand effect on employee turnover. JETIR Volume 6, Issue 6, June 2019.

Annexure:

Employee ID ?

Year of joining ?

Gender of the employee ?

A) Male B) Female C) Prefer not to say

Education Qualification of the employee ?

A) Higher Education(+2) B) Under Graduation C) Post Graduation D) Ph.D.

Marital status of the employee ?

A) Married B) Unmarried C)Other's

Employee Department ?

A) Human Resource B)Engineering C)Research & Development
D)Finance & Accounting E) Information Technology F) Operations & Maintenance

How much you satisfied with your job ?

1 Strongly dissatisfied

2 Dissatisfied

3 Neutral

4 Satisfied

5 Strongly dissatisfied

How much you satisfied with your job involvement ?

1 Strongly dissatisfied

2 Dissatisfied

3 Neutral

4 Satisfied

5 Strongly dissatisfied

How much you satisfied with work environment ?

1 Strongly dissatisfied

2 Dissatisfied

3 Neutral

4 Satisfied

5 Strongly dissatisfied

How much you satisfied with your package?

1 Strongly dissatisfied

2 Dissatisfied

3 Neutral

4 Satisfied

5 Strongly dissatisfied

How much you satisfied with your colleagues ?

1 Strongly dissatisfied

2 Dissatisfied

3 Neutral

4 Satisfied

5 Strongly dissatisfied

How much you satisfied with performance appraisal system?

1 Strongly dissatisfied

2 Dissatisfied

3 Neutral

4 Satisfied

5 Strongly dissatisfied

How much you satisfied with grievance appraisal system?

1 Strongly dissatisfied

2 Dissatisfied

3 Neutral

4 Satisfied

5 Strongly dissatisfied

How much you satisfied with work life balance ?

1 Strongly dissatisfied

2 Dissatisfied

3 Neutral

4 Satisfied

5 Strongly dissatisfied

Are you promoted ?

Yes No

17)Is Attrition ?

A)Yes(EX-Employee) B)No (Current Employee)

18)Reason for attrition ?

A)Better Opportunity

B) Dissatisfied with current salary

C) Biased Performance Appraisal System

D) Biased Performance Appraisal System

E) Ineffective Grievance Appraisal system

F) Family Problems

G) Lack of growth Opportunities