ABSTRACT

Work environment has always been a pivotal element in the personal growth of the employee. Employees improve their skill and knowledge in the workplace. Good working environment can bring more positive changes in the employees. Employee satisfaction analysis is a method to understand the employee's satisfaction levels in their carrier. The opinions given by the employees regarding the various factors such as satisfaction of salary, working hours, leave salary and some other factors related to workplace such as cleanliness, facilities etc. are used in this project. This project aims to understand the employees and analyse their job satisfaction levels using machine learning algorithms such as AdaBoostClassifier, XGBClassifier, SGDClassifier, CalibratedClassifierCV and OneVsRestClassifier.

INDEX

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

Employee satisfaction is a major issue in every companies. Because it impact on the growth of the company as well as the countries economy . As the working environment conditions affects an employees sentiments and work life, it determines the level of employee satisfactions. The employee satisfaction analysis is done using the reviews given by the employees regarding their workplace. This analysis will also helps in understanding the employees compatibility levels in a company and improve on it**.**

1.1 Aim and Objective

This project aims to analyzing the employees satisfaction based on the policies and facilities provided by the organization.

The objective of the project is to know regarding the employees sentiment towards the company. To know about the compatibility of employees in the workplace .To analyze the impact of working environment conditions on employees satisfaction levels. To analyze how the working environment conditions can be improved.

1.2 Problem Statement

The project is focused on providing a solution for understanding the employees sentiment towards the organization. As workplace significantly impact on the employees sentiments, it is required to analyze the reviews given by the employees in order to know regarding their satisfaction level.

CHAPTER 2

LITERATURE REVIEW

2.1 Work engagement and employee satisfaction in the practice of sustainable human resource management– based on the study of Polish employees [1]

Barbara Sypnienska et al. [1] described the importance of practising Sustainable human resource management(SHRM) in the workplace. They used a systematic literature review methodology. Considering the results of systematic literature review, the research model was built to determine the relationships between some factors such as employee workplace well-being, employee development, employee retention, and employee engagement. And they used regression model to understand these relations in a sustainable work environment. And they found that the employee engagement mediates relationships between employee workplace well-being, employee development and employee retention. And they concluded that, employee engagement in the workplace affects workplace well-being and employee satisfaction positively.

2.2 Work Environment and Job Satisfaction among Employees[2]

Anshika et al. [2] concluded that the working environment has a positive effect on employee's job satisfaction. And they said that giving importance to working environment will improve the company as well as its employees. Which will improve the economy of nation. Also, they stated that "having a comfortable work-life equilibrium act as a motivating factor towards work"

In this research paper author analyzed various other research papers and discussed regarding the impacts of work environment on employee's productivity, performance and mental well-being while completing their assigned task. And they mentioned that work environment is the place where employees with different abilities and motives work together towards achieving an identical goal. So. the workplace should be motivating for them in order to achieve anything better.

2.3 THE IMPACT OF WORKING ENVIRONMENT ON EMPLOYEES JOB SATISFACTION: A CASE STUDY OF PRIVATE SCHOOLS IN PESHAWAR CITY [3]

Lal Muhammad et al. [3] used regression model and mentioned that there is a direct positive association between work environment and job satisfaction. And they suggested that the supervisors must try to improve the level of communication between the coworkers. And the workload should be equally distributed among the them.

The study was conducted on the teachers of the private schools in the district of Peshawar. They made a survey to collect the information. They classified their samples based on the gender, age group, attitudes and opinions. They studied the relationship between the job satisfaction and some working environment related factors such as working hours, job safety and security, relation between the coworkers, workload and salary. And they concluded that in order to improve the employee's performance, an organization must focus on improving the working environment conditions.

2.4 EFFECT OF WORK ENVIRONMENT AND JOB SATISFACTION ON EMPLOYEE PERFORMANCE IN PT. NESINAK INDUSTRIES

G. Sailatha et al. [4] said that as poor working organizational cultures and rising occupational imbalances reduces the employee's satisfaction; it lowers the work motivation and productivity. They conducted a study on miliennial (a person who born between 1989-90) employees in Vijayawada city of Andhra Pradesh. A questionnaire was specially designed for the purpose of collecting the data. Data was also collected from the published sources like internet, research articles, books, journals, articles, magazines and other thesis. The sample size includes around 125 employees in the educational sector. The study was done to find the impact of the job satisfaction on work-life balance.

Using the regression model, they found that there is a linear relationship between employee retention, job satisfaction with work-life balance. And they concluded that the employees must try to improve on their work-life balance in order to improve on their job satisfaction levels.

2.5 Impact Of Working Environment On Job Satisfaction

Ovais Vohra et al. [5] conducted a study about the impact of work environment on job satisfaction. As they observed the employees from the different sectors such as educational institutions, communications etc. they noticed that the workplace atmosphere is crucial to job happiness. They used regression model to analyse and found a positive correlation between work environment and job satisfaction. And they concluded that in current world the employees are becoming more competent. As the increasing competition makes the work environment more challenging, it effects the employees. They also mentioned that employees are becoming more aware about their rights.

2.6 The Impact of Job Stress on Deviant Workplace Behavior: The Mediating Role of Job Satisfaction

Marwa Anis et al. [6] conducted a study on the impact of job stress on job satisfaction. Using the regression model, they observed a negative correlation between job stress and job satisfaction. Similarly, they found a negative correlation between job satisfaction and deviant workplace behaviour. According to their study there is an indirect positive relationship between job stress and deviant workplace behaviour. Also, they studied regarding the impact of job stress factors such as workload, role conflict, role ambiguity and work-to-family conflict. Here they concluded that these factors effects the productivity of employees in the workplace negatively.

2.7 The Role of Work Motivation and Work Environment in Improving Job Satisfaction

Muhammad Syafii A. Basalamah et al. [7] conducted a study relating the workplace environment, reward and recognition, teamwork, training and development with the job satisfaction. They made a reliability analysis and found that the contribution of workplace environment, reward and recognition, teamwork, training and development towards the job satisfaction are 23%, 12%, 45%, 21% respectively. Here result shows the positive relation of each factor with job satisfaction.

In this paper author concluded that the work motivation and work environment together have a positive effect on job satisfaction. So adequate facilities and support are necessary.

Also, the motivation from the both financial perspective and workplace are needed to improve the job satisfaction levels.

2.8 Impact of Working Environment on Job Satisfaction [8]

Rafia Hassan Taherdi et al. [8] used regression model to understand the working environment as a label of the physical social, secured and financial working environment. The study was conducted to find the relationship between working environment and job satisfaction. The work environment related factors such as working hours, job safety & security, relationship with co-workers, esteem needs and top management are taken for study. The data is gathered randomly from the employees of banks, telecommunication sector and universities in the city of Quetta Pakistan, through survey questionnaire. From each sector, 70 respondents were chosen.

And they found that the good working environment mostly depends on friendship and recognition. Also, the better working environment will lead to increased job satisfaction. They concluded that, as job satisfaction depends on the working environment, it is possible to improve the level of job satisfaction with a good relationship among the coworkers.

2.9 Work environment and job satisfaction among nurses in government tertiary hospitals in Nigeria [9]

Olusegun Emmannuel Avkinnale et al. [9] had conducted a study on Nurses in government hospitals of Nigeria using the regression model. Here they observed the relationship between the employee satisfaction and some factors such as Socio-political climate, administrative and managerial support, autonomy and responsibility, salary, supervision and working condition, recognition and achievement, advancement and promotion. And they concluded that the health care institutions should provide a mechanism that will make the health care workers satisfied at all the times; since they are vulnerable to certain diseases. Also, there should not be a socio-political barrier in the workplace. And the government and management of tertiary hospitals should ensure that organizational support is provided continuously to health-care workers in Nigerian hospitals. Management of hospitals need to also address the incidence of supervision of subordinates. And they suggested that there should not be partial supervision because its absence will create an atmosphere that will foster efficient socio-camaraderie at work.

2.10 Effect of work environment and job satisfaction on employee performance in pt. Nesinak industries[10]

Yuan Badrianto et al. [10] conducted a study on the employees of PT. Nesinak Industries in Bekasi of Indonesia. Using the multiple linear regression analysis, author studied the effect of work environment and job satisfaction on employee performance. The author observed that there is a positively relation between the workplace related factors and job satisfaction factors with employee performance. He noticed that workplace affects the employee performance by 58.3%; whereas job satisfaction affects by 42.3%. So, the work environment conditions have an indirect positive association with job satisfaction. Here they concluded that, both job satisfaction and good work environment conditions together shows the great impact on employee's performance.

CHAPTER 3

DATA DESCRIPTION

3.1 Codebook

An overview of all the columns in the original dataset is given below:

Sl. No.	Variable Name	Description	Data Type
1.	Name	Name of employee	Categorical
2.	Age	Age of employee	Numeric
3.	Gender	Gender of employee	Categorical
4.	Education Qualification	Qualification of employee	Categorical
5.	Department	Department name.	Categorical
6.	Job role	Employee job role	Categorical
7.	Salary category	Monthly income of employee	Numeric
8.	How long have you been working here?	Work experience of employee	Categorical
9.	Do you think company give concern for people or production ?	Company concern about people or production	Categorical
10.	Do you face any stress in your organization	Employees stress level	Categorical
11.	How will you rate the drinking water facility in your organisation ?	Satisfaction in drinking water	Categorical
12.	How will you rate the transportation facility in your organisation?	Satisfaction of transportation	Categorical
13.	How much are you satisfied with your salary?	Satisfaction of salary	Categorical
14.	Does your job match with your level of education attainment?	Employee job matching with education	Categorical
15.	How much are you satisfied with promotion policy adopted in the organization?	Satisfaction of promotion policy	Numeric

16.	How much are you satisfied with the working hours of your job?	Satisfaction pf Working hours	Categorical
17.	How much are you satisfied with your organization providing leave salary?	Satisfaction of leave salary	Numeric
18.	How much are you satisfied with training and other educational programs conducted by the organization?	Satisfaction of training program	Categorical
19.	How will you rate the cleanliness in your organization?	Satisfaction of cleanliness	Categorical

Table No. 1 Variable Description Table for Dataset before Preprocessing

An overview of all the columns in the modified dataset obtained after performing data preprocessing is given below

Sl. No.	**Variable Name**	**Description**	**Data Type**
1.	Age	Age of employee	Numeric
2.	Gender	Gender of employee	Numeric
3.	Education qualification	Qualification of employee	Categorical
4.	Department	Name of Department	Categorical
5.	Job Role	Title of the book.	Categorical
6.	Salary category	Department name.	Categorical
7.	Experience	Work experience of employee	Categorical
8.	Company_concern	Company concern about people or production	Numeric
9.	Face_stress	Employees stress level	Numeric
10.	Rate_drinking_water_facility	Satisfaction in drinking water	Numeric
11.	Rate_transportation_facility	Satisfaction of transportation	Numeric

12.	Satisfacation_of_salary	Satisfaction of salary	Numeric
13.	Job_match_with_level_of_education attainment	Employee job matching with education	Numeric
14.	Satisfacation_ promotion_policy	Satisfaction of promotion policy	Numeric
15.	Satisfacation_of_working_hours	Satisfaction pf Working hours	Numeric
16.	Satisfacation_of_leave_salary	Satisfaction of leave salary	Numeric
17.	Satisfacation_of_training_and_educational_programs	Satisfaction of training program	Numeric
18.	Rate_cleanliness	Satisfaction of cleanliness	Numeric

Table No. 2 Variable Description Table for Modified Dataset after Preprocessing

3.2 Structure of the Data

The initial step of the project is to load the dataset. The dataset used here is an excel file and it is loaded using pandas library.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from plotly.subplots import make_subplots
```

```
df=pd.read_excel('/content/employee.xlsx')
```

```
data=df.drop(['Name'], axis=1)
data
```

	Age	Gender	Education Qualification	Department	Job Role	Salary category	How long have you been working here?	Do you think company give concern for people or production ?	Do you face any stress in your organization	How will you rate the drinking water facility in your organisation ?	How will you rate the transportation facility in your organisation?	How much are you satisfied with your salary?	Does your job match with your level of education attainment?	How much are you satisfied with promotion policy adopted in the organization?	How much are you satisfied with the working hours of your job?	How much are you satisfied with your organization providing leave salary?	How much are you satisfied with training and other educational programs conducted by the organization?	How will you rate the cleanliness in your organization?
0	46	Male	Degree	Production	Chemist	15000-25000	6 Years	Production	No	Good	Satisfied	Netural	Yes	2	Satisfied	2	Satisfied	Medium
1	39	Male	Degree	Production	Clerk	25000-35000	8 years	Production	No	Average	Satisfied	Netural	No	3	Satisfied	3	Satisfied	Medium
2	55	Male	other	Marketing	Senior Executive manger	35000-45000	12 years	Production	Yes	Good	Highly Satisfied	Satisfied	Yes	4	Neutral	3	Satisfied	High
3	28	Male	Pg	Marketing	Junior Executive manger	25000-35000	2 Years	Production	No	Average	Satisfied	Netural	Yes	3	Neutral	3	Satisfied	Medium
4	31	Male	SSLC	Production	Operator	15000-25000	6 years	Production	No	Average	Satisfied	Netural	Yes	2	Satisfied	2	Netural	Medium
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
116	42	Female	other	Production	General worker	25000-35000	1 Years	People	No	Average	Satisfied	Netural	Yes	2	Neutral	4	Satisfied	Medium
117	38	Female	SSLC	Production	Worker	25000-35000	3 Years	Production	No	Good	Satisfied	Highly Satisfied	Yes	2	Satisfied	3	Satisfied	Medium
118	46	Male	SSLC	Production	General worker	25000-35000	2 Years	People	No	Good	Satisfied	Satisfied	No	3	Neutral	3	Satisfied	Medium
119	36	Female	Degree	Production	General Worker	35000-45000	6 yeras	Production	No	Average	Satisfied	Netural	No	1	Neutral	2	Satisfied	Medium
120	29	Female	Degree	Human Resource	Receptionist	15000-25000	2 years	People	No	Good	Satisfied	Netural	No	2	Satisfied	2	Satisfied	High

121 rows × 18 columns

The info() function is used to print information about the DataFrame. It contains details such as column name, count of non-null values and data types of all the columns in the DataFrame.

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 121 entries, 0 to 120
Data columns (total 18 columns):
 #   Column                                                                                Non-Null Count  Dtype
---  ------                                                                                --------------  -----
 0   Age                                                                                   121 non-null    int64
 1   Gender                                                                                121 non-null    object
 2   Education Qualification                                                               121 non-null    object
 3   Department                                                                            121 non-null    object
 4   Job Role                                                                              121 non-null    object
 5   Salary category                                                                       121 non-null    object
 6    How long have you been working here?                                                 121 non-null    object
 7   Do you think company give concern for people or production ?                          121 non-null    object
 8    Do you face any stress in your organization                                          121 non-null    object
 9    How will you rate the drinking water facility in your organisation ?                 121 non-null    object
 10  How will you rate the transportation facility in your organisation?                   121 non-null    object
 11   How much are you satisfied with your salary?                                         121 non-null    object
 12   Does your job match with your level of education attainment?                         121 non-null    object
 13   How much are you satisfied with promotion policy adopted in the organization?       121 non-null    int64
 14   How much are you satisfied with the working hours of your job?                       121 non-null    object
 15   How much are you satisfied with your organization providing leave salary?            121 non-null    int64
 16   How much are you satisfied with training and other educational programs conducted by the organization?  121 non-null    object
 17   How will you rate the cleanliness in your organization?                              121 non-null    object
dtypes: int64(3), object(15)
memory usage: 17.1+ KB
```

The function data.axes[] returns the number of values on a given axis. The following code is used to find the total number of rows in the data.

```
#rows
len(data.axes[0])
```

```
121
```

The following code is used to find the total number of columns in the data.

```
#columns
len(data.axes[1])
```

```
18
```

3.3 Missing Values and Treatment

The isnull().any() function is used to find the null values in the dataset. It returns a DataFrame with True for missing values and False for non null values.

```
data.isnull().any()
```

```
Age                                                                                    False
Gender                                                                                 False
Education Qualification                                                                False
Department                                                                             False
Job Role                                                                               False
Salary category                                                                        False
 How long have you been working here?                                                  False
Do you think company give concern for people or production ?                           False
 Do you face any stress in your organization                                           False
 How will you rate the drinking water facility in your organisation ?                  False
How will you rate the transportation facility in your organisation?                    False
 How much are you satisfied with your salary?                                          False
 Does your job match with your level of education attainment?                          False
 How much are you satisfied with promotion policy adopted in the organization?        False
 How much are you satisfied with the working hours of your job?                        False
 How much are you satisfied with your organization providing leave salary?             False
 How much are you satisfied with training and other educational programs conducted by the organization?    False
 How will you rate the cleanliness in your organization?                               False
dtype: bool
```

The rename() function is used to rename specified columns of the DataFrame.

```
data = data.rename(columns = {" How long have you been working here?": "Experience",
                              "Do you think company give concern for people or production ?":"Company_concern",
                              " Do you face any stress in your organization":"Face_stress",
                              " How will you rate the drinking water facility in your organisation ?":"Rate_drinking_water_facility",
                              "How will you rate the transportation facility in your organisation?":"Rate_transportation_facility",
                              " How much are you satisfied with your salary?":"Satisfacation_of_salary",
                              " Does your job match with your level of education attainment?":"Job_match_with_level_of_education attainment",
                              " How much are you satisfied with promotion policy adopted in the organization?":"Satisfacation_ promotion_policy",
                              " How much are you satisfied with the working hours of your job?":"Satisfacation_of_working_hours",
                              " How much are you satisfied with your organization providing leave salary?":"Satisfacation_of_leave_salary",
                              " How much are you satisfied with training and other educational programs conducted by the organization?":"Satisfacation_of_training_and_educational_programs",
                              " How will you rate the cleanliness in your organization?":"Rate_cleanliness "})
```

The drop() function is used to remove columns or rows from the DataFrame.

```
data=df.drop(['Name'], axis=1)
data
```

The replace() function is used to replace a particular value with another one.

```
data=data.replace("general worker","General Worker")
data=data.replace("worker","General Worker")
data=data.replace("Worker","General Worker")
data=data.replace("clerk","Clerk")
data=data.replace("chemist","Chemist")
data=data.replace("2 years","2 Years")
data=data.replace("6 years","6 Years")
data=data.replace("3 yr","3 Years")
data=data.replace("6 yeras","6 Years")
data=data.replace("12 years","12 Years")
data=data.replace("8 years","8 Years")
data=data.replace("1","1 Years")
data=data.replace("2 ","2 Years")
```

Label encoding is used to convert categorical variable into numerical variable form.

```
from sklearn.preprocessing import LabelEncoder
le= LabelEncoder()
new_data['Face_stress']=le.fit_transform(new_data['Face_stress'])
new_data['Gender ']=le.fit_transform(new_data['Gender '])
new_data['Company_concern']=le.fit_transform(new_data['Company_concern'])
new_data['Rate_drinking_water_facility']=le.fit_transform(new_data['Rate_drinking_water_facility'])
new_data['Rate_transportation_facility']=le.fit_transform(new_data['Rate_transportation_facility'])
new_data['Satisfacation_of_salary']=le.fit_transform(new_data['Satisfacation_of_salary'])
new_data['Job_match_with_level_of_education attainment']=le.fit_transform(new_data['Job_match_with_level_of_education attainment'])
new_data['Satisfacation_of_working_hours']=le.fit_transform(new_data['Satisfacation_of_working_hours'])
new_data['Satisfacation_of_training_and_educational_programs']=le.fit_transform(new_data['Satisfacation_of_training_and_educational_programs'])
new_data['Rate_cleanliness ']=le.fit_transform(new_data['Rate_cleanliness '])
new_data['Department']=le.fit_transform(new_data['Department'])
```

The modified dataset after performing all the preprocessing methods is printed below.

```
new_data = data
```

	Age	Gender	Education Qualification	Department	Job Role	Salary category	Experience	Company_concern	Face_stress	Rate_drinking_water_facility	Rate_transportation_facility	Satisfacation_of_salary	Job_match_with_level_of_education attainment
0	46	1	Degree	3	Chemist	15000-25000	6 Years	1	0	2	3	2	1
1	39	1	Degree	3	Clerk	25000-35000	8 years	1	0	0	3	2	0
2	55	1	other	2	Senior Executive manger	35000-45000	12 years	1	1	2	1	3	1
3	28	1	Pg	2	Junior Executive manger	25000-35000	2 Years	1	0	0	3	2	1
4	31	1	SSLC	3	Operator	15000-25000	6 years	1	0	0	3	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
116	42	0	other	3	General worker	25000-35000	1 Years	0	0	0	3	2	1
117	38	0	SSLC	3	Worker	25000-35000	3 Years	1	0	2	3	1	1
118	46	1	SSLC	3	General worker	25000-35000	2 Years	0	0	2	3	3	0
119	36	0	Degree	3	General Worker	35000-45000	6 yeras	1	0	0	3	2	0
120	29	0	Degree	1	Receptionist	15000-25000	2 years	0	0	2	3	2	0

121 rows × 18 columns

The duplicated() function is used to analyse duplicated values.

```
new_data.duplicated()
```

```
0      False
1      False
2      False
3      False
4      False
       ...
116    False
117    False
118    False
119    False
120    False
Length: 121, dtype: bool
```

The following is the information of the dataframe obtained after preprocessing.

```
new_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 121 entries, 0 to 120
Data columns (total 18 columns):
 #   Column                                          Non-Null Count  Dtype
---  ------                                          --------------  -----
 0   Age                                             121 non-null    int64
 1   Gender                                          121 non-null    int64
 2   Education Qualification                         121 non-null    object
 3   Department                                      121 non-null    int64
 4   Job Role                                        121 non-null    object
 5   Salary category                                 121 non-null    object
 6   Experience                                      121 non-null    object
 7   Company_concern                                 121 non-null    int64
 8   Face_stress                                     121 non-null    int64
 9   Rate_drinking_water_facility                    121 non-null    int64
 10  Rate_transportation_facility                    121 non-null    int64
 11  Satisfacation_of_salary                         121 non-null    int64
 12  Job_match_with_level_of_education attainment    121 non-null    int64
 13  Satisfacation_ promotion_policy                 121 non-null    int64
 14  Satisfacation_of_working_hours                  121 non-null    int64
 15  Satisfacation_of_leave_salary                   121 non-null    int64
 16  Satisfacation_of_training_and_educational_programs  121 non-null  int64
 17  Rate_cleanliness                                121 non-null    int64
dtypes: int64(14), object(4)
memory usage: 17.1+ KB
```

The description of each column in the data after preprocessing.

```
new_data.describe()
```

	Age	Gender	Department	Company_concern	Face_stress	Rate_drinking_water_facility	Rate_transportation_facility	Satisfacation_of_salary	Job_match_with_level_of_education attainment	Satisfacation_ promotion_policy	Satisfacation_of_working_hours
count	121.000000	121.000000	121.000000	121.000000	121.000000	121.000000	121.000000	121.000000	121.000000	121.000000	121.000000
mean	38.719008	0.661157	2.760331	0.727273	0.206612	1.107438	2.380165	2.214876	0.685950	2.694215	0.280992
std	9.201651	0.475284	1.016733	0.447214	0.406558	1.031194	0.787995	0.887004	0.466066	0.825055	0.451352
min	22.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
25%	31.000000	0.000000	2.000000	0.000000	0.000000	0.000000	2.000000	2.000000	0.000000	2.000000	0.000000
50%	38.000000	1.000000	3.000000	1.000000	0.000000	1.000000	3.000000	2.000000	1.000000	3.000000	0.000000
75%	46.000000	1.000000	3.000000	1.000000	0.000000	2.000000	3.000000	3.000000	1.000000	3.000000	1.000000
max	58.000000	1.000000	4.000000	1.000000	1.000000	3.000000	3.000000	3.000000	1.000000	5.000000	1.000000

CHAPTER 4

EXPLORATORY DATA ANALYSIS

4.1 Need and Importance

Exploratory data analysis (EDA) is used to analyse the dataset and summarize the characteristics ,using data visualization methods. Visuals in the form of graphs, images, diagrams or animations. It helps us to have a clear understanding of the dataset and helps to discover the relationships between them. There are three types of Exploratory Data Analysis Univariate, Bivariate and Multivariate Analysis. The human mind can comprehend the information more easily if it is presented in the form of visuals. Visualization make learning content more interesting and understanding. visual techniques to represent the Statistical data in diagrammatic and graphical presentations for the understanding of common men.

Performing exploratory data analysis prior to implementing algorithm helps us to have an understanding on the dataset. This help programmers to select suitable machine learning model for the dataset, resulting in more accuracy. Additionally, EDA aids in summarizing the data and presenting information in more appealing manner.

4.2 Univariate Analysis

Univariate analysis is performed on a single variable with the aim of understanding the features of that variable. This is one of the simplest analysis techniques.

4.2.1 EDA with categorical data

```
import seaborn as sns
sns.countplot(x='Satisfacation_of_salary',data=data).set(title="No.of employees satisfied with their salary")
```

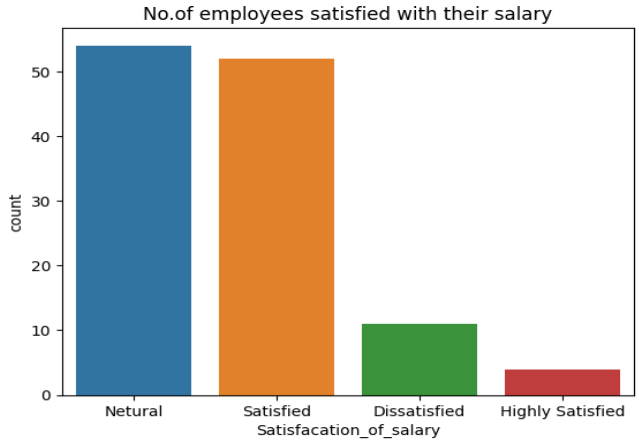

Fig. No. 1 Number of employees satisfied with their salary by using Count Plot

```
round(data['Department'].value_counts()/ data.shape[0]*100,2).plot.pie(autopct = '%1.1f%%',figsize =(7, 10)).set_title("Number of Employess in each Department")
```

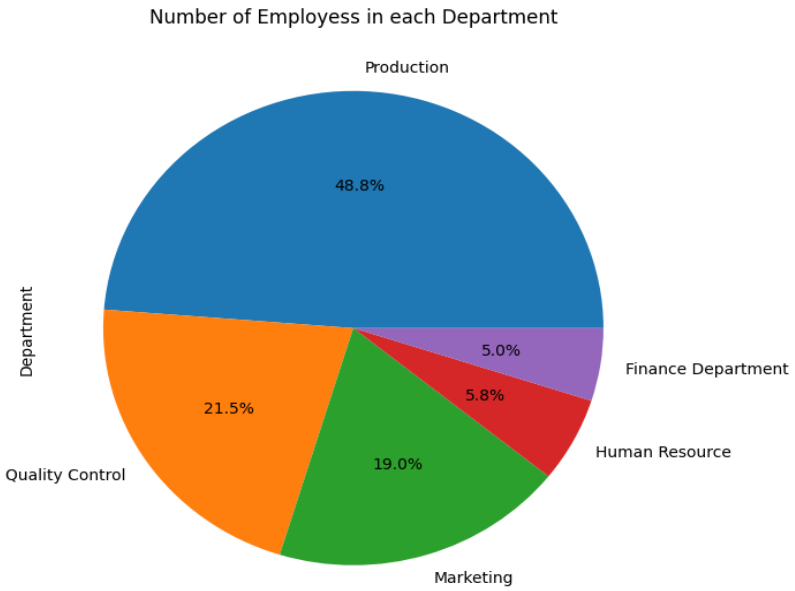

Fig. No. 2 Number of employees in each Department by using pie chart

4.3 Bivariate Analysis

Bivariate analysis is used to find the relation between two variables. The bivariate plots help to find the discrepancies between two variables as well.

4.3.1 EDA with categorical data

```
import plotly.express as px
px.histogram(data,x='Satisfacation_of_salary',color='Department',barmode='group',color_discrete_sequence= px.colors.qualitative.D3,title="Count of salary satisfaction level of employee in each Department")
```

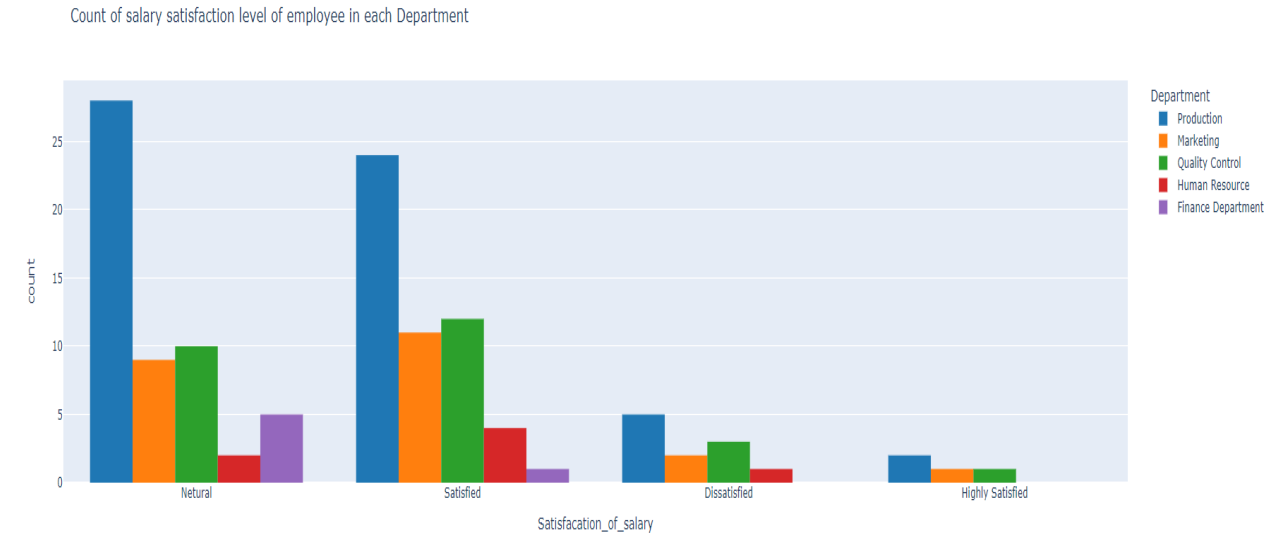

Fig. No. 3 Count the salary satisfaction of employee in each department by using histogram

```
#bivariate analysis
sns.displot(data=data, x="Age",hue="Department", kind="kde",height=7)
```

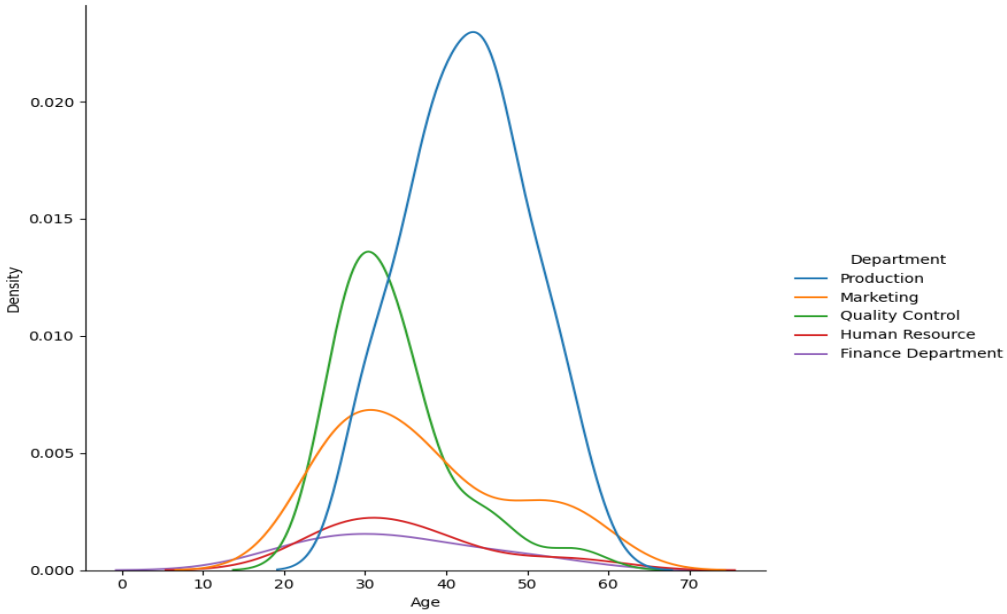

Fig. No. 4 Distribution of age group according to department

4.4 Multivariate Analysis

Multivariate analysis is used for analysing data having two or more variables. This is an extension of univariate and bivariate analysis.

4.4.1 EDA with Numeric data

```
#Multivariate Analysis
plt.figure(figsize=(10,5))
sns.heatmap(new_data.iloc[:,:-1].corr(), annot=True)
```

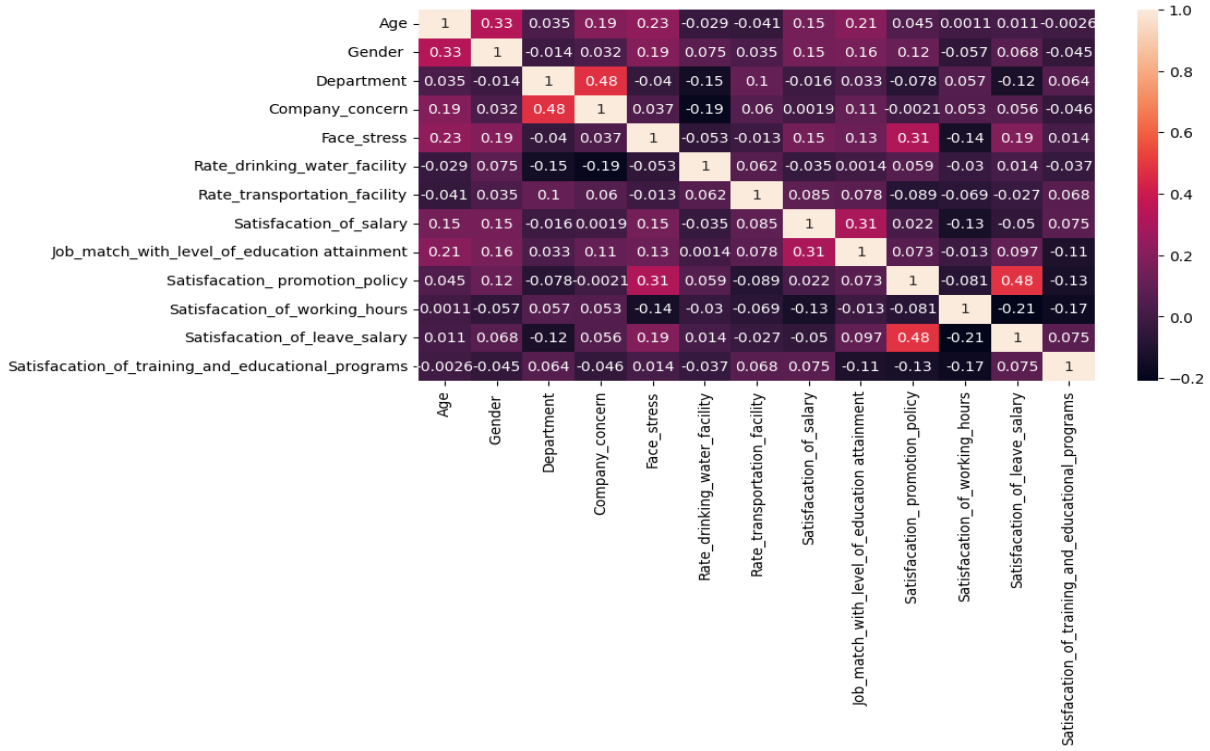

Fig. No. 5 Heatmap of Correlation

CHAPTER 5

METHODOLOGY

Workflow Diagram for Prediction

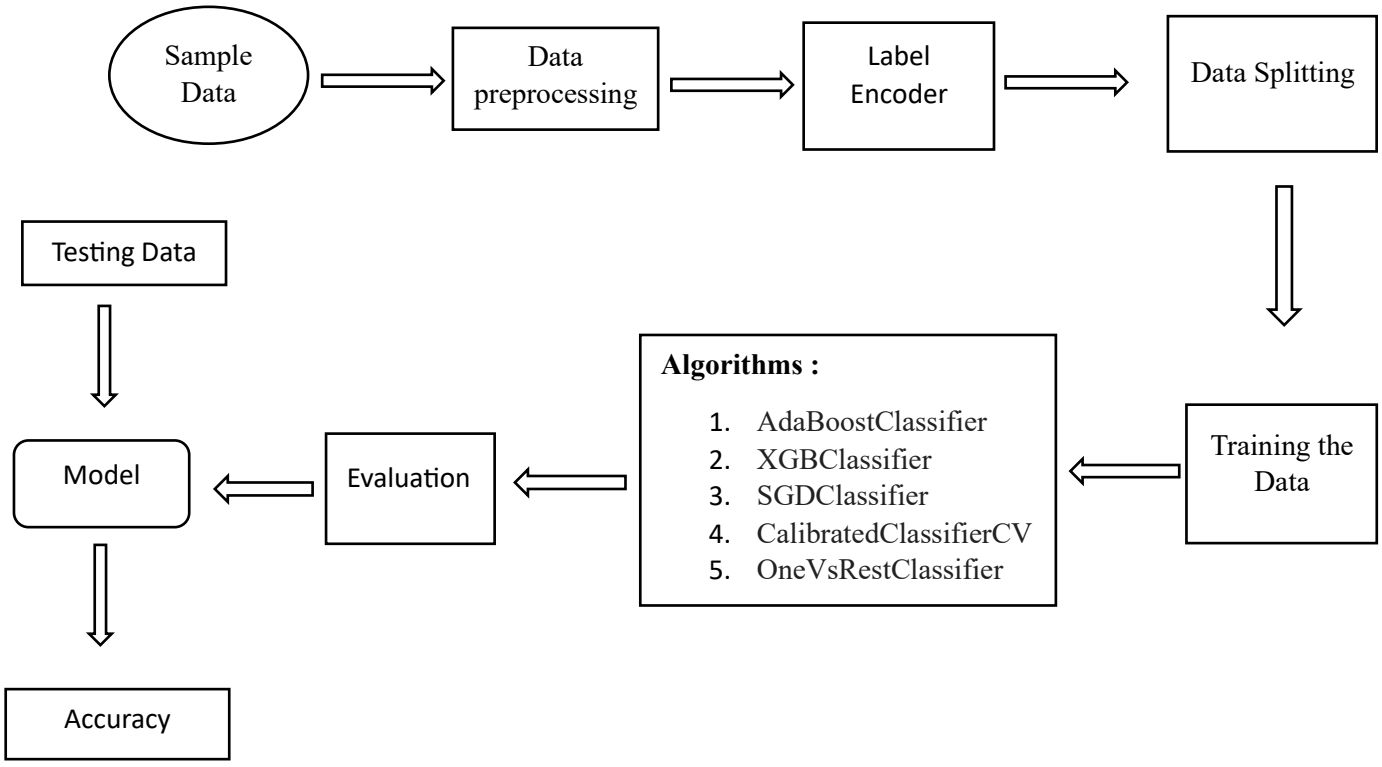

Fig 6: The workflow Diagram for prediction

The dataset is preprocessed using the various data preprocessing techniques available. Then the dataset is fitted with Label Encoder to change the categorical values into numerical values. The dataset is then split into training set and testing set. The training set is used to train the algorithms and the corresponding models are obtained. The model is then used for predicting the values of test dataset. The accuracy for each of the algorithms are measured.

5.1 Model I – AdaBoostClassifier

AdaBoost Algorithm is a supervised learning algorithm .It can be used for both classification and regression problems. It is a Boosting technique used as an Ensemble Method in Machine Learning. AdaBoost in machine learning is one of these predictive modelling techniques. It is an effective method for combining multiple weak or base learners into a single strong learner, and has been shown to have good generalization performance.

5.2 Model II - XGBClassifier

XGBoost is a supervised learning algorithm. It is used for both classification and regression problem. It attempts to accurately predict a target variable by combining an ensemble of estimates from a set of simpler and weaker models. XGBoost is applied for structured and tabular data. Itis an implementation of gradient boosted decision trees designed for speed and performance. XGBoost is an extreme gradient boost algorithm.

5.3 Model III - SGDClassifier

Stochastic Gradient Descent (SGD) is a variant of the gradient Descent algorithm that is used for optimizing machine learning models. It is a supervised learning algorithm. Stochastic Gradient Descent (SGD) is a machine learning algorithm that fits linear classifiers under convex loss functions such as SVM and Logistic Regression. It uses SGD as an optimization algorithm that updates the parameters of the function for each training instance. It is suitable for large-scale and sparse datasets and can be implemented using the SGDClassifier class in scikit-learn API .

SGDClassifier is a linear classifier optimized by the SGD. It supports multi-class classification by combining multiple binary classifiers in a “one versus all” (OVA) scheme. For each of the K classes, a binary classifier is learned that discriminates between that and all other K − 1 classes .

The cost function of Logistic Regression cannot be calculated directly, so we try to minimize it via Stochastic Gradient Descent, also known as Online Gradient Descent 1. In SGD, instead of using the entire dataset for each iteration, only a single random training example (or a small batch) is selected to calculate the gradient and update the model parameters. This random selection introduces randomness into the optimization process, hence the term “stochastic” in stochastic Gradient Descent

5.4 Model IV- Calibrated classifiercv

CalibratedClassifierCV is a class in the scikit-learn library that calibrates the output probabilities of a classifier. It does this by training a new classifier on the output probabilities of the original classifier. The new classifier is then used to predict the output probabilities of the original classifier. CalibratedClassifierCV uses a cross-validation approach to ensure unbiased data is always used to fit the calibrator. The calibration function maps the output probability scores to a more calibrated version that better reflects the true probabilities.

In other words, CalibratedClassifierCV is used to improve the accuracy of predicted probabilities by calibrating them to better reflect their true values. This is particularly useful when working with classifiers that produce poorly calibrated probability estimates, such as support vector machines (SVMs) and random forests.

CalibratedClassifierCV supports two types of probability calibration: parametric 'sigmoid' method (Platt's method) and nonparametric 'isotonic' method. The calibration function maps the output probability scores to a more calibrated version that better reflects the true probabilities.

5.5 Model V -One vs Rest

One-vs-Rest (OvR) is a heuristic method for using binary classification algorithms for multi-class classification. It involves splitting the multi-class dataset into multiple binary classification problems. For each of the K classes, a binary classifier is learned that discriminates between that and all other $K - 1$ classes.

The OvR strategy is also known as one-vs-all. It consists of fitting one classifier per class. For each classifier, the class is fitted against all the other classes. In addition to its computational efficiency (only n_classes classifiers are needed), one advantage of this approach is its interpretability.

Scikit-learn provides an implementation of the OvR strategy through the OneVsRestClassifier class. This class can be used to train a multi-class classifier on a dataset using any binary classifier as the base estimator.

CHAPTER 6

IMPLEMENTATION OF THE ALGORITHM

6.1 Implementation and Results

1. AdaBoostClassifier

The first algorithm used in this project is AdaBoostClassifier. 57.14% is the accuracy obtained using this algorithm**.**

```
from sklearn.ensemble import AdaBoostClassifier
adb = AdaBoostClassifier()
X= np.array( new_data['Face_stress'])
y= np.array( new_data['Satisfacation_of_salary'] )
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.05, random_state=0)
adb_model = adb.fit(X_train.reshape(-1, 1),y_train)
print("The accuracy of the model on validation set is", adb_model.score(X_test.reshape(-1, 1),y_test))
```

```
The accuracy of the model on validation set is 0.5714285714285714
```

2. XGBClassifier

The second algorithm used is XGBClassifier. 45.90% is the accuracy obtained using this algorithm.

```
from xgboost.sklearn import XGBClassifier
xgb = XGBClassifier()
X= np.array( new_data['Department'])
y= np.array( new_data['Satisfacation_of_salary'] )
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5, random_state=0)
y_train = le.fit_transform(y_train)
xgb.fit(X_train.reshape(-1,1),y_train)
y_Pre = xgb.predict(X_test.reshape(-1,1))
print('Accuracy', accuracy_score(y_test.reshape(-1,1), y_Pre))
```

```
Accuracy 0.45901639344262296
```

3. SGDClassifier

The third algorithm used is SGDClassifier. 71.42% is the accuracy obtained using this algorithm.

```
] from sklearn.linear_model import SGDClassifier
  sgd = SGDClassifier()
  X= np.array( new_data['Department'])
  y= np.array( new_data['Satisfacation_of_salary'] )
  from sklearn.model_selection import train_test_split
  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.05, random_state=0, shuffle=None)
  sgd.fit(X_train.reshape(-1,1), y_train)
  y_Pred = sgd.predict(X_test.reshape(-1,1))
  print('Accuracy1', sgd.score(y_test.reshape(-1,1), y_Pred))
```

```
Accuracy1 0.7142857142857143
```

4. CalibratedClassifierCV

The fourth algorithm used is CalibratedClassifierCV. 57.14% is the accuracy obtained using this algorithm.

```
from sklearn.calibration import CalibratedClassifierCV
cv = CalibratedClassifierCV()
X= np.array( new_data['Department'])
y= np.array( new_data['Satisfacation_of_salary'] )
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.05, random_state=100)
cd = cv.fit(X_train.reshape(-1, 1),y_train)
print("The accuracy of the model on validation set is", cd.score(X_test.reshape(-1, 1),y_test))
```

```
The accuracy of the model on validation set is 0.5714285714285714
```

5. OneVsRestClassifier

The fifth algorithm used for prediction is OneVsRestClassifier. 57.14% is the accuracy obtained using this algorithm.

```
from sklearn.multiclass import OneVsRestClassifier
from sklearn.linear_model import SGDClassifier
f= OneVsRestClassifier(SGDClassifier())
X= np.array( new_data['Department'])
y= np.array( new_data['Satisfacation_of_salary'] )
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.05, random_state=0)
clf = f.fit(X_train.reshape(-1, 1),y_train)
y_Pred = f.predict(X_test.reshape(-1,1))
print('Accuracy', accuracy_score(y_test, y_Pred))
```

```
Accuracy 0.5714285714285714
```

Model Accuracy Table

Algorithms	**Accuracy**
SGDClassifier	71.42 %
AdaBoostClassifier	57.14 %
CalibratedClassifierCV	57.14 %
OneVsRestClassifier	57.14 %
XGBClassifier	45.9 %

Table No. 3 Accuracy Table for the Algorithm

From the above table, it can be inferred that SGDClassifer algorithm returns the best accuracy value i.e., 71.42%.

Here it can be concluded that SGDClassifier algorithm is the best algorithm to understand regarding the employees satisfaction levels based on satisfaction levels based on satisfaction of salary, leave salary, working hours etc.

CHAPTER 7:

FUTURE SCOPE AND CONCLUSION

Employee satisfaction analysis is a research area that has a huge scope for development. Every organization including government and private sectors must focus on improving their employee's satisfaction levels. This project considers the reviews given by the employees about 'satisfaction of salary' and various other factors such as stress, working hours, leave salary etc. to analyse their sentiments towards the company. So, the project is supposed to understand the employees satisfaction levels along with their attitudes regarding the working environment. The implementation of the project on real time dataset obtained from "NICE Chemicals Pvt.Ltd , Kochi, kerala" Helped to understand the employees.

If an employee is satisfied always with his carrier-life then they will be a grace for there organization. More satisfied employees are more productive than the others. Only productive employees are beneficial to the company. Similarly, only productive companies are beneficial to the economy of a nation. Improving the quality of working environment also improves the employee's satisfaction levels. Along with this a company should have a system which supports the employees to balance their personal life and carrier life. When employee satisfaction analysis is done with analysing their sentiments towards the company, it is possible to get more accurate solutions. And it is possible to keep employees satisfied almost all the times.

REFERENCES

[1] Barbara Sypniewska, Małgorzata Baran, Monika Kłos," Work engagement and employee satisfaction in the practice of sustainable human resource management– based on the study of Polish employees",March 2023,p. 1069–1100 ,DOI: 10.1007/s11365-023-00834-9.

[2] Anshika Aggarwal, Devanshi Sharma, Pranika Vohra, Sneha Sharma, Vadini Sharma," Work Environment and Job Satisfaction among Employees", Volume 11, p. 2349-3429, January- March 2023, DOI: 10.25215/1101.126.

[3] Lal Muhammad, Jawad Rahim Afridi and Dr Rabia Ishrat "THE IMPACT OF WORKING ENVIRONMENT ON EMPLOYEES JOB SATISFACTION: A CASE STUDY OF PRIVATE SCHOOLS IN PESHAWAR CITY" 59(2): 168-183 March 2022, ISSN: 1553-6939

[4] G. Sailatha and Dr. R. S. V. Rama Swathi "EFFECT OF WORK ENVIRONMENT AND JOB SATISFACTION ON EMPLOYEE PERFORMANCE IN PT. NESINAK INDUSTRIES" 2022, Vol. 6, No. 6, 10542 – 10550

[5] Ovais Vohra and Mustafa Ozyesil "Impact Of Working Environment On Job Satisfaction" July 2022 Vol. 6, No. 6, 3496-3505

[6] Marwa Anis, Dina Emil "The Impact of Job Stress on Deviant Workplace Behavior: The Mediating Role of Job Satisfaction" 2022, 12, 123-134

[7] Muhammad Syafii A. Basalamah 1, Ajmal As'ad," The Role of Work Motivation and Work Environment in Improving Job Satisfaction", volume 1,p.2776-6365, July 2021 , DOI:10.52970/grhrm.v1i2.54.

[8] Rafia Hasan Taheri, Md. Shipon Miah and Md. Kamaruzzaman," Impact of Working Environment on Job Satisfaction ",volume 5, December 2020, DOI: 10.24018/ejbmr.2020.5.6.643

[9]Olusegun Emmanuel Akinwale and Olusoji James George, “Work environment and job satisfaction among nurses in government tertiary hospitals in Nigeria” ,volume 4, march 2020 , p. 71-92 , DOI 10.1108/RAMJ-01-2020-0002

[10] Yuan Badrianto and Muhamad Ekhsan “EFFECT OF WORK ENVIRONMENT AND JOB SATISFACTION ON EMPLOYEE PERFORMANCE IN PT. NESINAK INDUSTRIES” Volume 2 March 2020

[11] A. C. Müller and S. Guido, Introduction to machine learning with Python : a guide for data scientists. Beijing: O’reilly, 2017.

[12] geeksforgeeks “Data preprocessing” - https://www.geeksforgeeks.org/data-preprocessing-machine-learning-python

[13] geeksforgeeks “Label encoding” - https://www.geeksforgeeks.org/ml-label-encoding-of-datasets-in-python/

[14]Analytics vidhya “Adaboost” - https://www.analyticsvidhya.com/blog/2021/09/adaboost-algorithm-a-complete-guide-for-beginners/

[15]machine learning mastery “One vs rest” - https://machinelearningmastery.com/one-vs-rest-and-one-vs-one-for-multi-class-classification

[16] netlifyapp “SGD” https://michael-fuchs-python.netlify.app/2019/11/11/introduction-to-sgd-classifier/

[17]tutorialspoint “Stochastic gradient descent”.

https://www.tutorialspoint.com/scikit_learn/scikit_learn_stochastic_gradient_descent.htm

[18] Analytics vidhya “xgboost” - https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/