A
Project Report
On
"Employee Retention System"

Prepared by

Amit Chandravadiya (19DCS015)
Nisarg Chaniyara (19DCS016)
Anil Chavda (19DCS018)

Under the guidance of

Prof. Jesal Desai

A Report Submitted to

Charotar University of Science and Technology

for Partial Fulfillment of the Requirements for the

6th Semester Software Group Project-IV (CS357)

Submitted at

CSE

DEPSTAR

At: Changa, Dist: Anand – 388421

April 2022

CERTIFICATE

This is to certify that the report entitled "**Employee Retention System**" is a bonafied work carried out by **Mr. Amit Chandravadiya (19DCS015), Mr. Nisarg Chaniyara (19DCS016), Mr. Anil Chavda (19DCS018)** under the guidance and supervision of **Prof. Jesal Desai** for the subject CS357-**Software Group Project-IV** (CSE) of 6th Semester of Bachelor of Technology in **DEPSTAR** at Faculty of Technology & Engineering – CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate himself, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred to the examiner.

Prof. Jesal Desai
Assistant Professor – Computer
Science & Engineering, DEPSTAR
CHARUSAT, Changa, Gujarat.

Prof. Parth Goel
Head of Department – Computer
Science & Engineering, DEPSTAR
CHARUSAT, Changa, Gujarat.

Dr. Amit Ganatra
Principal, DEPSTAR
Dean, FTE
CHARUSAT, Changa, Gujarat.

ABSTRACT

Where there is any organization, there are Employees. One cannot imagine an organization without Employees. In today's competitive world, Employee leaves their organization for various reasons like they get greater opportunities in other organization. It would be very challenging for the organization if it fails to retain its employees. Basically, Employee Retention means to make the current Employees Sustainable by retaining them. Employee retention is not all about forcing Employees to stay by binding them with agreements or contracts but it is to figure out the factors that make them "Voluntarily stay". This project will show that how the employee retention will contribute towards employee performance and that will directly or indirectly improve organizations' profit. Also, this project will help to get knowledge about the role of employee retention in any organization. The main purpose of this project is to analyze and identify the major concerning factors affecting employee retention.

ACKNOWLEDGEMENT

We, the developer of a Machine Learning Project "**Employee Retention System**", with immense pleasure and commitment would like to present the project assignment. The development of this project has given us wide opportunity to think, implement and interact with various aspects of Machine Learning Skills as well as the new emerging technologies.

Every work that one completes successfully stands on the constant encouragement, good will and support of the people around. We hereby avail this opportunity to express our gratitude to number of people who extended their valuable time, full support and cooperation in developing the project.

We express deep sense of gratitude towards our Head of the CSE Department, Prof. Parth Goel and project guides Prof. Jesal Desai for the support during the whole session of study and development. It is because of them, that we were prompted to do hard work, adopting new technologies.

I am sincerely thankful to all the people at DEPSTAR who helped me complete the project in one way or the other.

They altogether provided me favourable environment, and without them it would not have been possible to achieve my goal.

Thanks,

Amit Chandravadiya (19DCS015)
Nisarg Chaniyara (19DCS016)
Anil Chavda (19DCS018)

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

1.1 PROJECT OVERVIEW

Success of any organization depends on its Employees; it can be said that Employees play a vital role in developing a long-term image of any organization. So, main task of any organization is to retain their employees by encouraging them. Employee Retention is nothing but Encouraging employees to stay in the organization for a longer period of time. This does not mean that forcefully binding them by contracts or agreements but by providing them the satisfaction and encouragement they need to make them stay voluntarily. Evolution/Advancements have led most organizations to become technology-driven but automation too requires Employee. In Any kind of automation or any technology-driven organization, Human resource is required.

Due to more and more competition in today's era, Skilled Employees are finding more and more opportunities. So, mostly that kind of Employee tends to leave their jobs. And here comes the terminology "Attrition Rate".

Attrition Rate: The rate at which employees leave an organization from a specific group over a particular period of time.

The main goal of our Project is to analyze and identify the major concerning factors affecting employee retention.

1.2 OBJECTIVES

The main objective of the project is to predict on what factors Employees tends to leave their organisation. Our Project will helps organisations to figure out how to retain their employees by dealing with the problem which is causing Employees to leave the organisation.

We want to

- Apply Machine Learning algorithms on this real-world problem like Employee Retention.
- Check the results which Machine Learning Algorithms provide by applying it in real-world scenarios.

1.3 SCOPE

The proposed system of the project will figure out the major concerning factors which are resulting Employees to leave their organisation.

The project will predict that the Employee will stay or leave the organisation on the basis of some major factors like;

a. salary
b. promotions in last 5 years (if any)
c. number of projects done
d. satisfaction level
e. work accidents (if any)
f. evaluation report

1.4 TOOLS & TECHNOLOGIES USED

Python	• Jupyter Notebook
Python Libraries	• NumPy • Pandas • Matplotlib • Seaborn • SkLearn
Front End	• HTML • CSS • BootStrap
Back End	• SQLite • Django Framework • Heroku • Github

Table 1. Tools and Technologies Used

CHAPTER 2: PROJECT MANAGEMENT

2.1 PROJECT PLANNING

2.1.1 Project Development Approach and Justification

The spiral model is a **risk-driven process model**. This SDLC model helps the group to adopt elements of one or more process models like a waterfall, incremental, waterfall, etc. The spiral technique is a combination of rapid prototyping and concurrency in design and development activities.

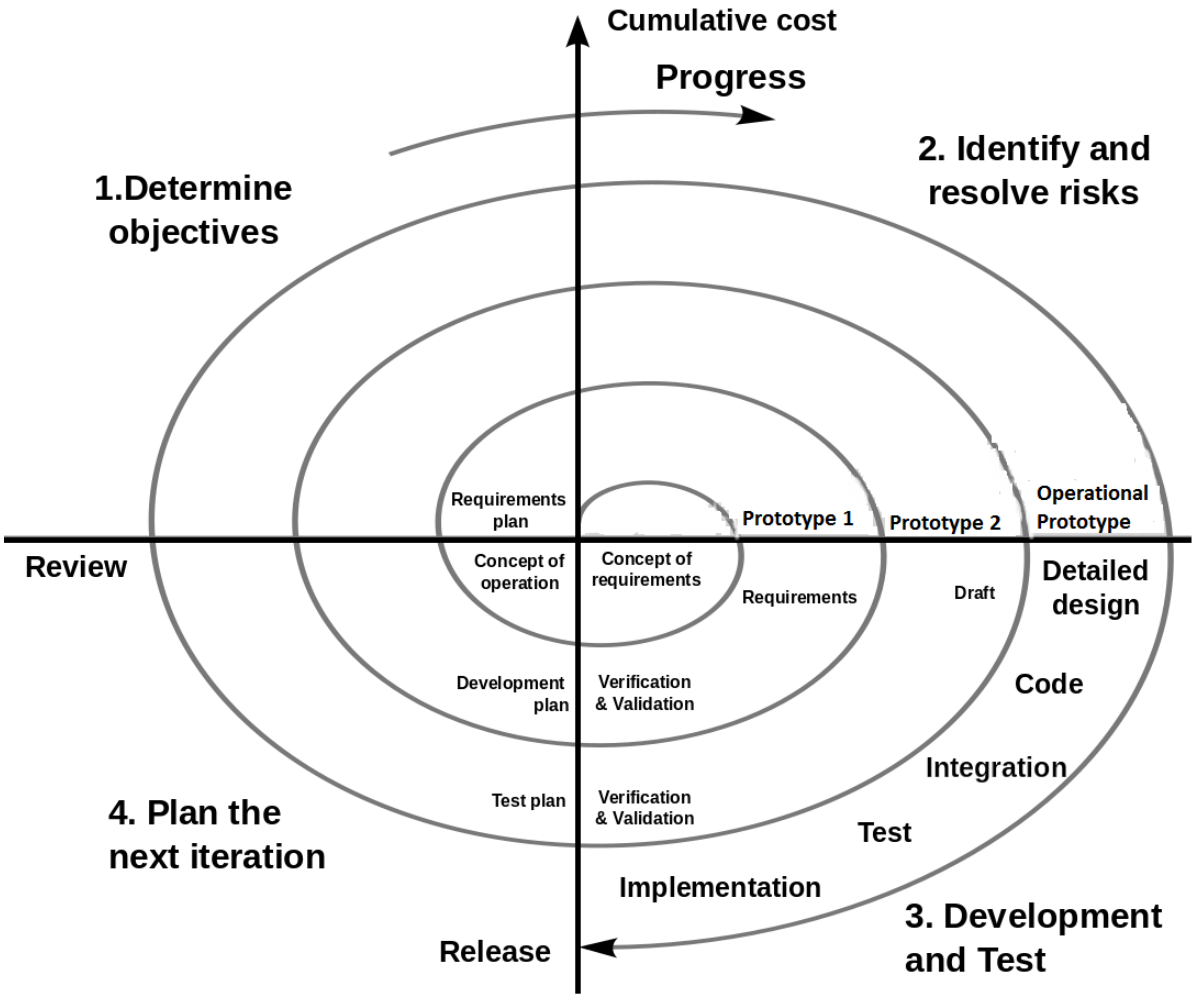

Fig 1. Spiral model

Each cycle in the spiral begins with the identification of objectives for that cycle, the different alternatives that are possible for achieving the goals, and the constraints that exist. This is the first quadrant of the cycle (upper-left quadrant).

The next step in the cycle is to evaluate these different alternatives based on the objectives and constraints. The focus of evaluation in this step is based on the risk perception for the project.

The next step is to develop strategies that solve uncertainties and risks. This step may involve activities such as benchmarking, simulation, and prototyping.

Advantages:

- Changing requirements can be accommodated.
- Allows extensive use of prototypes.
- Requirements can be captured more accurately.

Disadvantages:

- Management is more complex.
- End of project may not be known early.
- Not suitable for small or low risk projects and could be expensive for small projects.

2.1.2 Project Effort and Time, Cost Estimation

We have used basic COCOMO estimation technique for the estimation of our project. As our project team is combination of experienced and inexperienced members in developing such kind of applications. We concluded to include the project under semidetached category.

CHAPTER 3: SYSTEM REQUIREMENTS STUDY

3.1 USER CHARACTERISTICS

- **End Users**

  This system is mainly developed for HR of an organisation. So, it will be easier for them to predict the sustainability of the Employees, whether they will leave or not. Employee Retention Prediction System mainly focuses on the various reasons which leads Employees to leave the organisation.

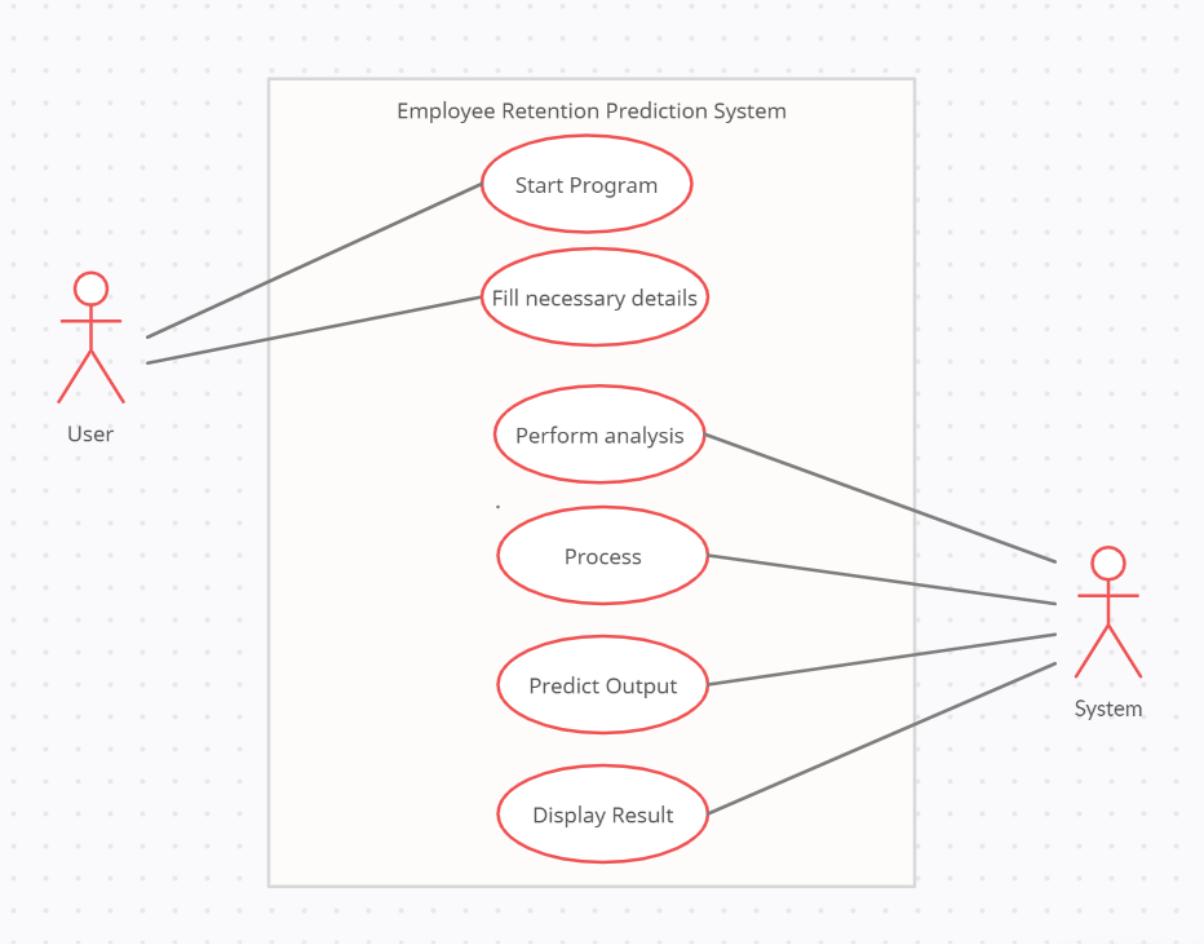

Fig 2. Use Case Diagram

3.2 HARDWARE AND SOFTWARE REQUIREMENTS

3.2.1 Hardware specification

- **For Developer**

CPU	Intel i5 10th gen
GPU	Nvidia GeForce GTX 1650
RAM	8 GB
ROM	128 GB SSD

Table 2. Hardware Specifications (Developer)

- **For User**

CPU	Intel core 2 duo
GPU	Intel UHD graphics
RAM	2 GB
ROM	20 GB

Table 3. Hardware Specifications (User)

3.2.2 Software specification

- **For Developer**

Jupyter Notebook Google Collab	Machine Learning Platform Run models in cloud Deploy models
Google Chrome	To run web-based system
Microsoft PowerPoint	Presentation of the Project
Microsoft Word	Creating Report of the Project
Kaggle	Get Dataset
Snipping Tool	For capturing screenshots
PyCharm	Web model
Visual Studio	Implementation and Deployment

Table 4. Software Specifications (Developer)

- **For User**

Operating System	Windows 7 and later
Web Browsers	Google Chrome Microsoft Bing Mozilla Firefox Opera

Table 5. Software Specifications (User)

3.3 DATA FLOW DIAGRAM

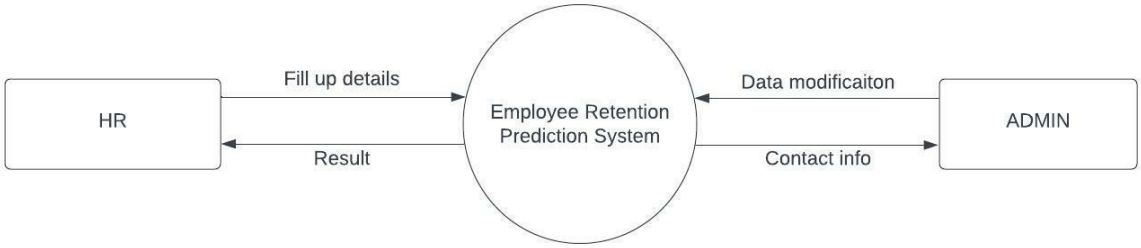

Fig 3. DFD (Level 0)

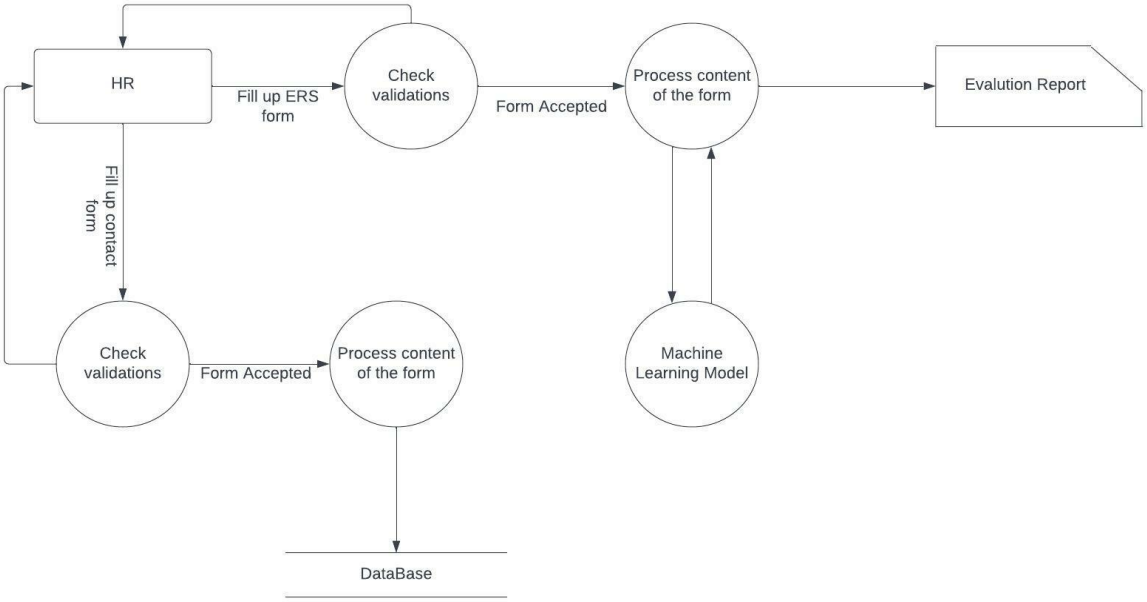

Fig 4. DFD (Level 1)

CHAPTER 4: PROPOSED SYSTEM

4.1 STUDY OF PROPOSED SYSTEM

For our Classification based project we will use Kaggle dataset, and there are total of 5 steps for Building our Model.

1. Data Cleaning

   Data cleaning refers to identifying and correcting errors in the dataset that may negatively impact a predictive model. Data cleaning is used to refer to all kinds of tasks and activities to detect and repair errors in the data.

2. Exploratory Data Analysis

   Exploratory Data Analysis (EDA) is an approach to analyse the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.

3. Data Pre-possessing

   Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data.

4. Model Building

   A machine learning model is built by learning and generalizing from training data, then applying that acquired knowledge to new data it has never seen before to make predictions and fulfil its purpose. Lack of data will prevent you from building the model, and access to data isn't enough.

5. Evaluation

   Model evaluation is the process of using different evaluation metrics to understand a machine learning model's performance, as well as its strengths and weaknesses.

6. Improvements which were needed

7. Deployment

   Deployment is the method by which you integrate a machine learning model into an existing production environment to make practical business decisions based on data.

4.2 FRAMEWORKS, TECHNIQUES & LIBRARIES USED

For Python	• Jupyter Notebook • Google Colab
Libraries	• Pandas • NumPy • Matplotlib • Seaborn • SKLearn
Front End	• HTML • CSS • BootStrap
Back End	• SQLite • Django Framework
Deployment	• Heroku • GitHub

Table 6. Frameworks, Techniques & Libraries Used

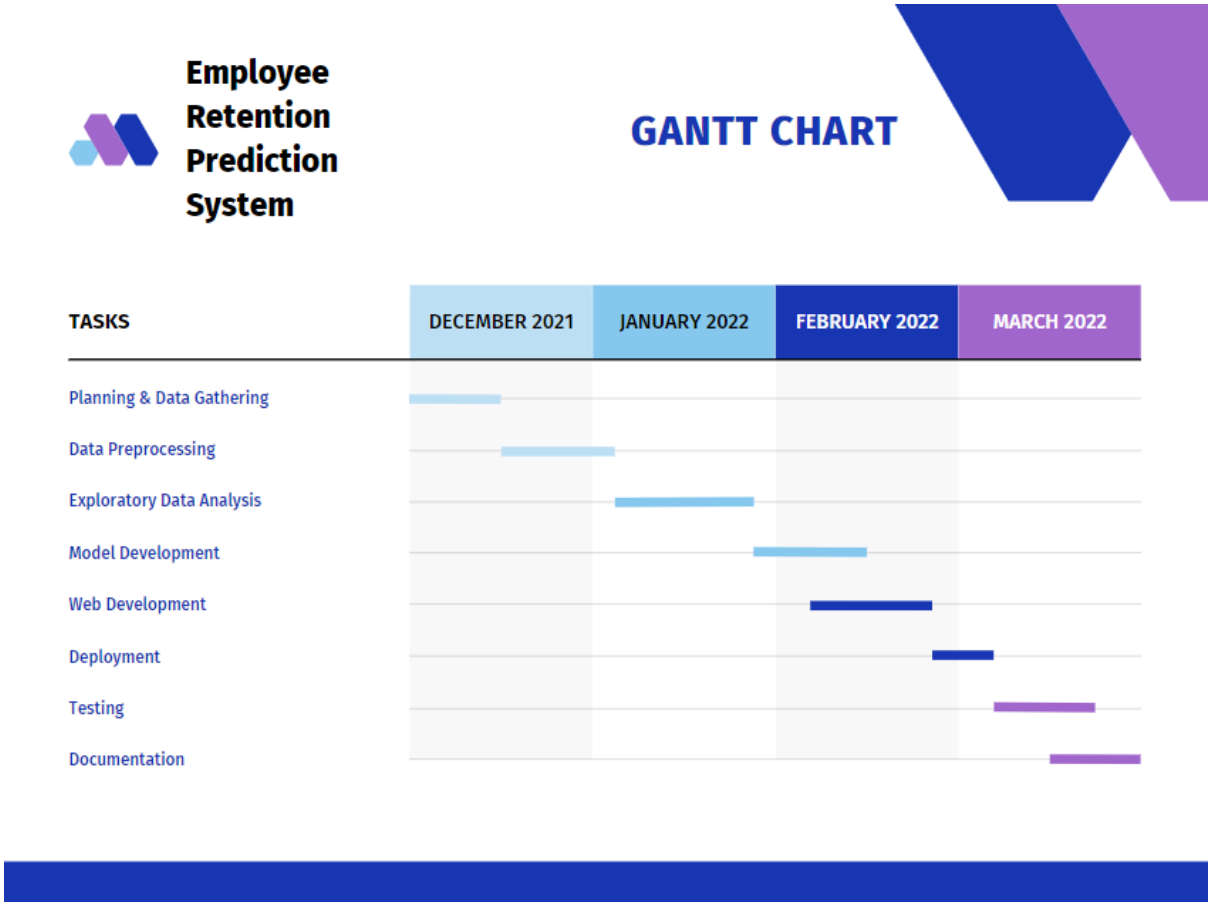

Fig 5. Gantt Chart

CHAPTER 5: SYSTEM ARCHITECTURE & DESIGN

5.1 EMPLOYEE RETENTION SYSTEM TECHNIQUE

Classification Algorithms which we have tested to find out which one is suitable for our Dataset:

1. Naïve Bayes Classifier
2. Logistic Regression
3. Support Vector Machine (SVM)
4. Decision Tree
5. Random Forest

Following are the screenshots which shows:

1. Data Preprocessing
2. Exploratory Data analysis (EDA)
3. Feature Engineering

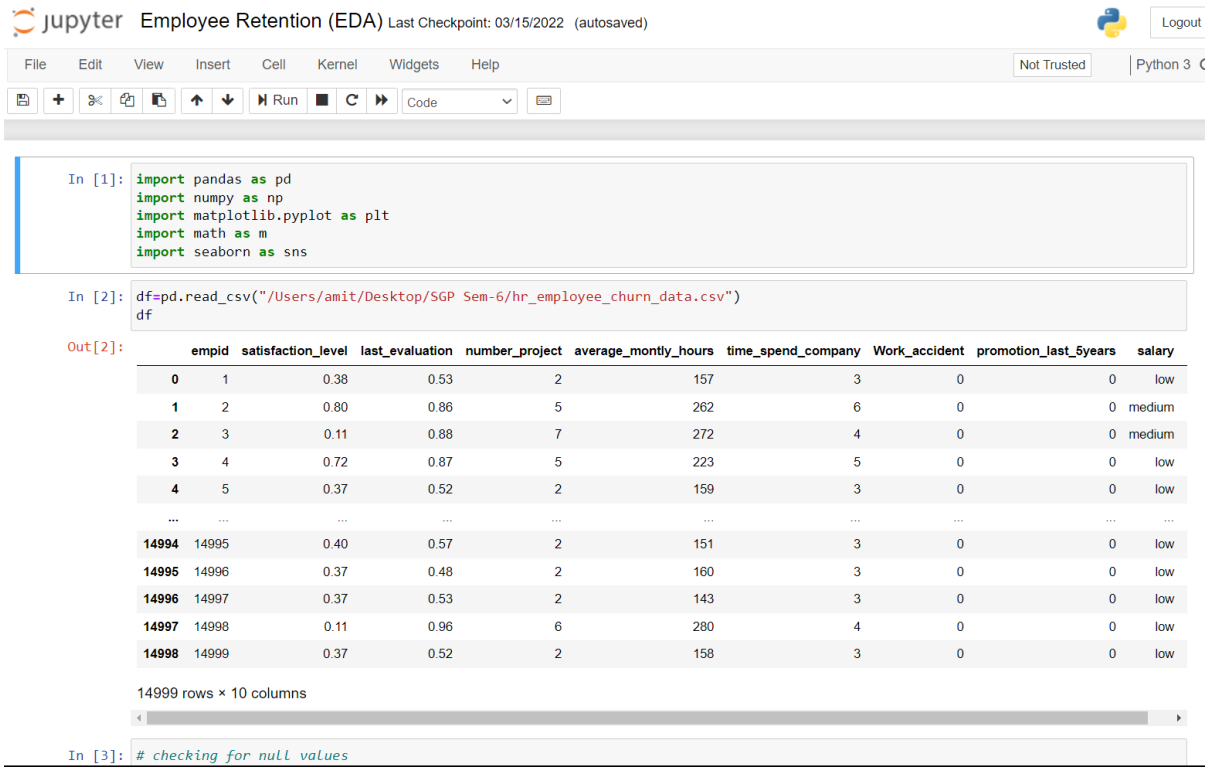

Fig 6. EDA

jupyter Employee Retention (EDA) Last Checkpoint: 03/15/2022 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

Run Code

In [3]:
```
# checking for null values
df.isnull().any()
```

Out[3]:
```
empid                    False
satisfaction_level        True
last_evaluation          False
number_project           False
average_montly_hours     False
time_spend_company       False
Work_accident            False
promotion_last_5years    False
salary                   False
left                     False
dtype: bool
```

In [4]:
```
# finding total null values
df.isnull().sum()
```

Out[4]:
```
empid                    0
satisfaction_level       2
last_evaluation          0
number_project           0
average_montly_hours     0
time_spend_company       0
Work_accident            0
promotion_last_5years    0
salary                   0
left                     0
dtype: int64
```

In [5]:
```
df2=df.dropna()
```

In [6]:
```
df2
```

Fig 7. EDA 1

jupyter Employee Retention (EDA) Last Checkpoint: 03/15/2022 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

Run Code

In [7]:
```
# checking for datatypes of column.
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 14997 entries, 0 to 14998
Data columns (total 10 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   empid                  14997 non-null  int64
 1   satisfaction_level     14997 non-null  float64
 2   last_evaluation        14997 non-null  float64
 3   number_project         14997 non-null  int64
 4   average_montly_hours   14997 non-null  int64
 5   time_spend_company     14997 non-null  int64
 6   Work_accident          14997 non-null  int64
 7   promotion_last_5years  14997 non-null  int64
 8   salary                 14997 non-null  object
 9   left                   14997 non-null  int64
dtypes: float64(2), int64(7), object(1)
memory usage: 1.3+ MB
```

In [8]:
```
# checking dimension of dataframe
df2.shape
```

Out[8]:
```
(14997, 10)
```

In [9]:
```
# checking statistics of the datasets
df2.describe()
```

Out[9]:

	empid	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	promotion_last_5years	14
count	14997.000000	14997.000000	14997.000000	14997.000000	14997.000000	14997.000000	14997.000000	14997.000000	
mean	7500.998733	0.612863	0.716128	3.803294	201.057411	3.498300	0.144629	0.021271	
std	4329.407522	0.248634	0.171165	1.232499	49.942638	1.460222	0.351738	0.144291	
min	1.000000	0.090000	0.360000	2.000000	96.000000	2.000000	0.000000	0.000000	
25%	3752.000000	0.440000	0.560000	3.000000	156.000000	3.000000	0.000000	0.000000	
50%	7501.000000	0.640000	0.720000	4.000000	200.000000	3.000000	0.000000	0.000000	

Fig 8. EDA 2

jupyter Employee Retention (EDA) Last Checkpoint: 03/15/2022 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

Run Code

```
In [10]: df2['Work_accident'].unique()
Out[10]: array([0, 1])

In [11]: df2['promotion_last_5years'].unique()
Out[11]: array([0, 1])

In [12]: df2['left'].unique()
Out[12]: array([1, 0])

In [13]: df2['salary'].unique()
Out[13]: array(['low', 'medium', 'high'], dtype=object)

In [14]: # checking total employee range based on the target column 'left'
         ax1=sns.countplot(x='left',data=df2)
         for p in ax1.patches:
             ax1.annotate(f'{p.get_height()}',(p.get_x()+0.32, p.get_height()+0.01))
         plt.show()
         # CONCLUSION: Number of Employee who are still working with the company are less as compared to Employees who left the company.
```

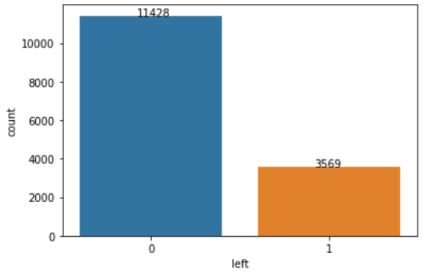

Fig 9. EDA 3

```
In [15]: # checking for salary feature.
         ax2=sns.countplot(x='salary',data=df2)
         for p in ax2.patches:
             ax2.annotate(f'{p.get_height()}',(p.get_x()+0.32, p.get_height()+0.01))
         plt.show()
         # CONCLUSION: Number of Employee with low salary are high as compared to other Employees with medium and high salary.
```

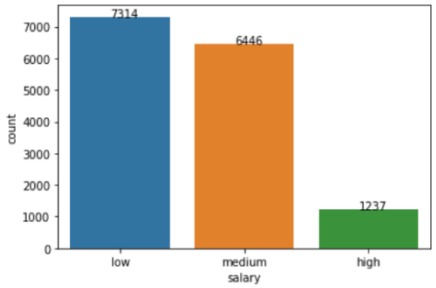

Fig 10. EDA 4

```
In [16]: # checking left and stayed on the basis of the salarya column.
g=sns.catplot(x='left',col='salary',kind='count',data=df2)
ax3= g.axes.ravel()

# l=[]
for i in ax3:
    for c in i.containers:
        labels = [f'{(v.get_height())}' for v in c]
        i.bar_label(c, labels=labels, label_type='edge')
plt.show()
# CONCLUSION: Employees with low salary have high churning rate.
```

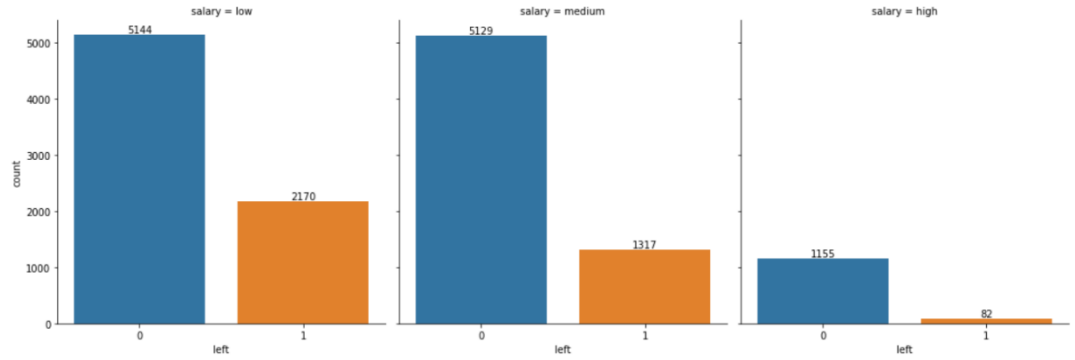

Fig 11. EDA 5

```
In [17]: # Checking count of the employes who got promoted in last years.
ax=sns.countplot(x='promotion_last_5years',data=df2)
for p in ax.patches:
    ax.annotate(f'{p.get_height()}',(p.get_x()+0.32, p.get_height()+0.01))
plt.show()
# CONCLUSION: Most of the employees who were not promoted they tends to leave the job.
```

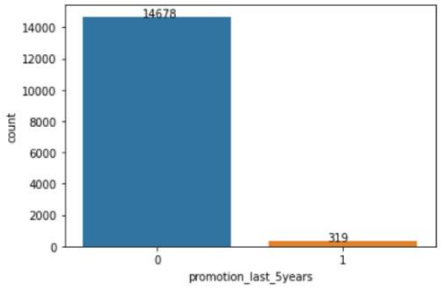

Fig 12. EDA 6

```
In [18]: # feature 'left' split over the feature 'promotion_last_5years'
g2=sns.catplot(x='left',col='promotion_last_5years',kind='count',data=df2)
ax4= g2.axes.ravel()

# This for loop is uesd to annotate the plot
for i in ax4:
    for c in i.containers:
        labels = [f'{(v.get_height())}' for v in c]
        i.bar_label(c, labels=labels, label_type='edge')
plt.show()
# CONCLUSION: Employee who are not promoted have high churning rate as compared to employees who are promoted.
```

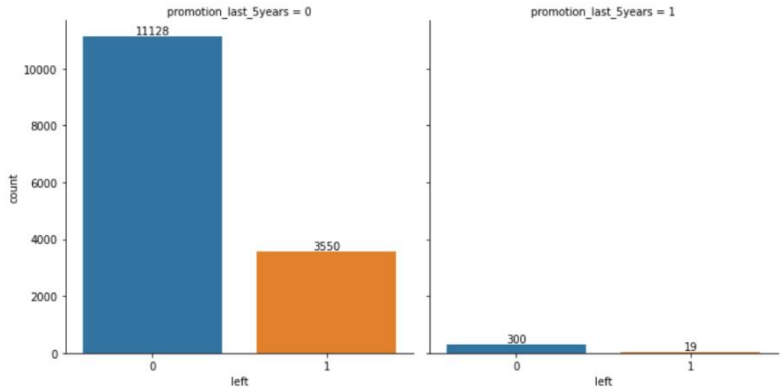

Fig 13. EDA 7

```
In [19]: # feature 'Work_accident'
ax5=sns.countplot(x='Work_accident',data=df2)
for p in ax5.patches:
    ax5.annotate(f'{p.get_height()}',(p.get_x()+0.32, p.get_height()+0.01))
plt.show()

# CONCLUSION: The employees who have high number of work accident tends to leave the company.
```

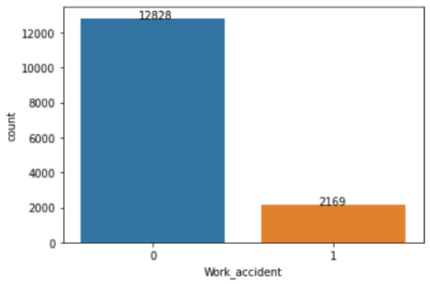

Fig 14. EDA 8

In [20]:
```
# feature 'left' split over the feature 'Work_accident'
g3=sns.catplot(x='left',col='Work_accident',kind='count',data=df2)
ax6= g3.axes.ravel()

# This for loop is uesd to annotate the plot
for i in ax6:
    for c in i.containers:
        labels = [f'{(v.get_height())}' for v in c]
        i.bar_label(c, labels=labels, label_type='edge')
plt.show()
# CONCLUSION: The employee who was not involved in any work accident tends to leave the Company more.

### NOTE : ###
# This feature is not that much important because logically the employee who was involved in the work accident will
# have higher probability of leaving the company. So, according to the conclusion of the Below graph it shows opp.
# result. So, due to this reason it makes this feature less important for our model.
```

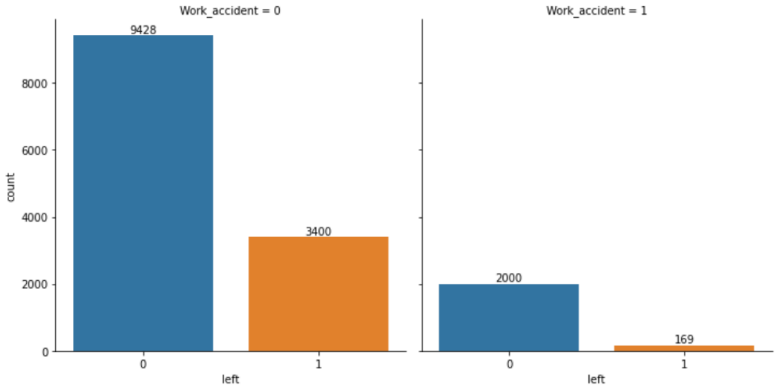

Fig 15. EDA 9

ANALYSIS OF THE NUMERICAL FEATURES

In [24]:
```
# Analysis of satisfaction_level feature
sns.distplot(df2['satisfaction_level'])
plt.show()
# CONCLUSION: Satisfaction level ranges between 0.1 to 1 and satisfaction level is highest at 0.1
```

```
/Users/amit/opt/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:2551: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

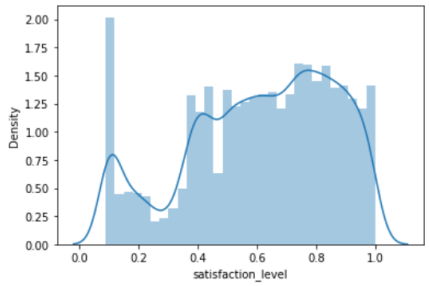

Fig 16. EDA 10

```
In [26]: # Boxplot to show distribution of satisfaction level over the Category left
sns.boxplot(x='left',y='satisfaction_level',data=df2)
plt.show()
# CONCLUSION: Employee left more who has less satisfaction level range from 0.1 - 0.5
```

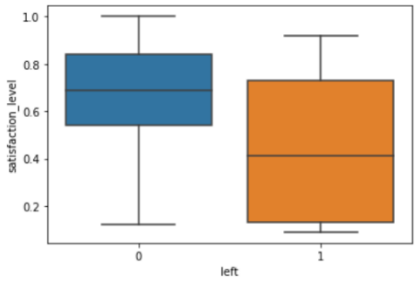

Fig 17. EDA 11

```
In [29]: # Analysis of number_project feature
sns.distplot(df2['number_project'])
plt.show()
# CONCLUSION: Project is distributed from 2-7 where it has high count at 4.
```

```
/Users/amit/opt/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:2551: FutureWarning: `distplot` is a deprecated
function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

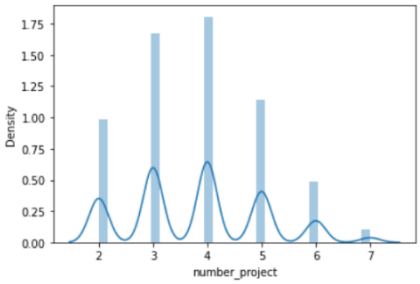

Fig 18. EDA 12

```
In [30]: # Boxplot to show distribution of number of projects over the Category left
sns.boxplot(x='left',y='number_project',data=df2)
plt.show()
# CONCLUSION: Employee left when project count is high and employee stays and work when project count is between 3-4
```

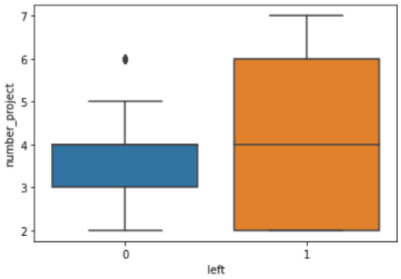

Fig 19. EDA 13

```
In [33]: # Analysis of average_monthly_hours feature
sns.distplot(df2['average_montly_hours'])
plt.show()
# CONCLUSION: count of working hour 150 is high in the range of 90-320 hrs.
```

```
/Users/amit/opt/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:2551: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

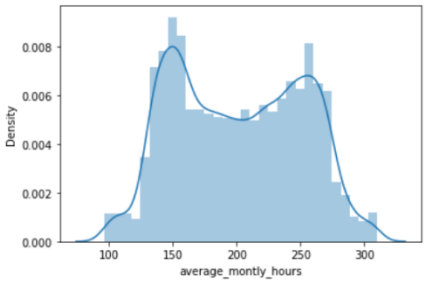

Fig 20. EDA 14

```
In [36]: # Boxplot to show distribution of average working hours  over the Category left
sns.boxplot(x='left',y='average_montly_hours',data=df2)
plt.show()
# CONCLUSION: Employee left more who has high average_monthly_hours value
```

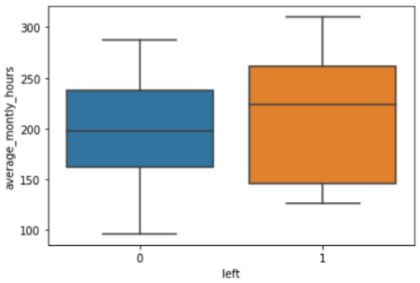

Fig 21. EDA 15

```
In [43]: # Correlation matrix
cor_matrix=df2.corr()
fig=plt.figure(figsize=(15,7))
sns.heatmap(cor_matrix,annot=True)
plt.show()
```

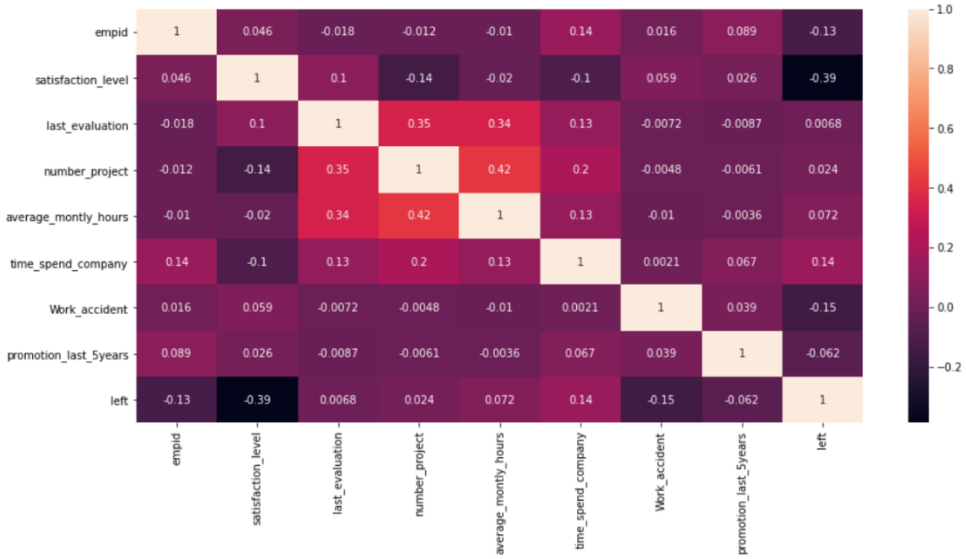

Fig 22. EDA 16

```
In [46]: # Pair plot for given dataframe
         sns.pairplot(df2)
         plt.show()
```

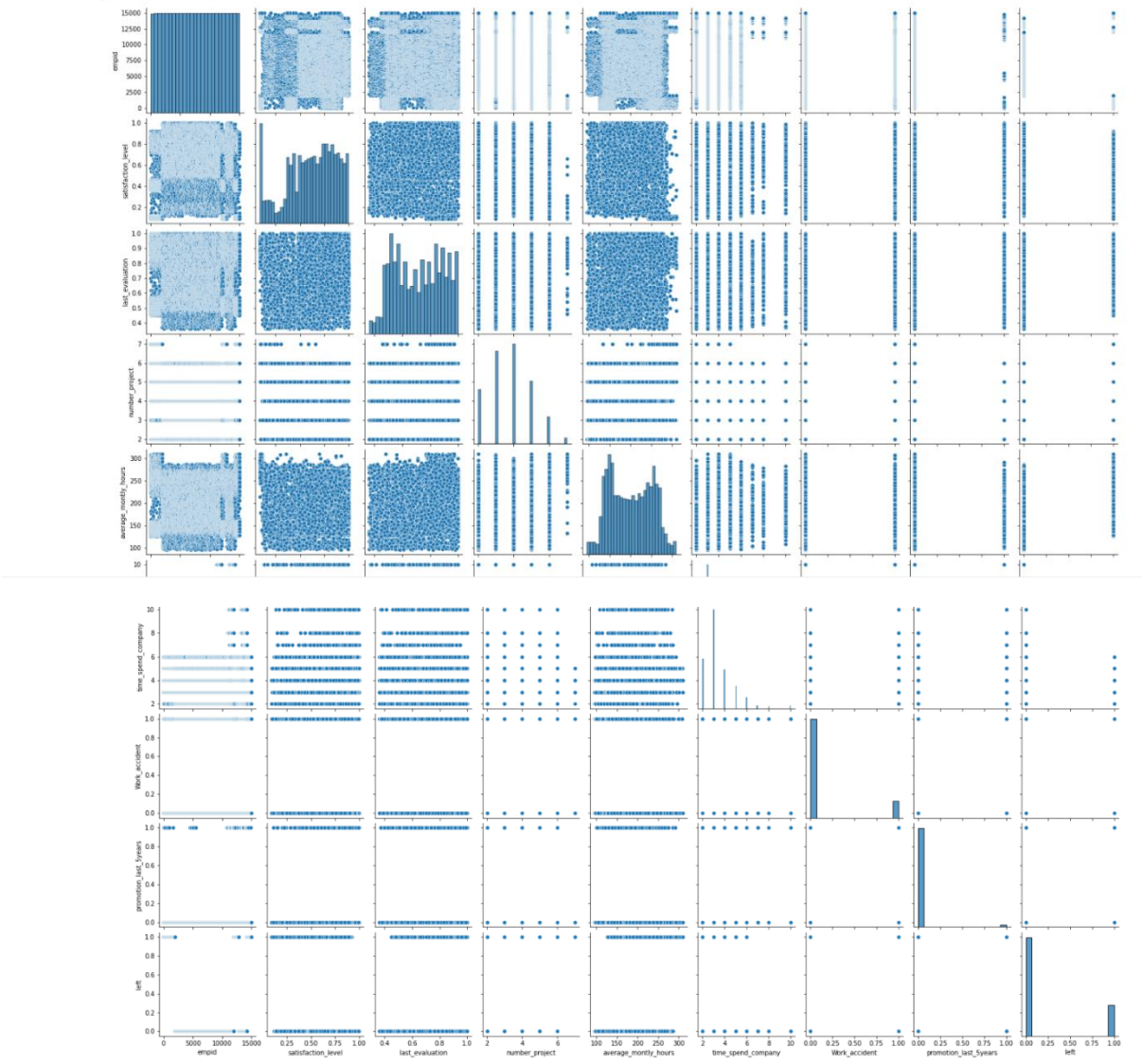

Fig 23. EDA 17

In [97]:
```
# Confusion matrix
cnf_mtx=confusion_matrix(y_test,pred1)
print(cnf_mtx)

# Classification report
clf_report1=classification_report(y_test,pred1)
print(clf_report1)
```

```
[[2120  162]
 [ 452  266]]
              precision    recall  f1-score   support

           0       0.82      0.93      0.87      2282
           1       0.62      0.37      0.46       718

    accuracy                           0.80      3000
   macro avg       0.72      0.65      0.67      3000
weighted avg       0.78      0.80      0.78      3000
```

Fig 24. Confusion Matrix - Logistic Regression

In [102…
```
# Confusion matrix
cnf_mtx3=confusion_matrix(y_test,pred3)
print(cnf_mtx3)

# Classification report
clf_report3=classification_report(y_test,pred3)
print(clf_report3)
```

```
[[1840  442]
 [ 153  565]]
              precision    recall  f1-score   support

           0       0.92      0.81      0.86      2282
           1       0.56      0.79      0.66       718

    accuracy                           0.80      3000
   macro avg       0.74      0.80      0.76      3000
weighted avg       0.84      0.80      0.81      3000
```

Fig 25. Confusion Matrix - Naïve Bayes Classifier

In [32]:
```
# Confusion matrix
cnf_mtx4=confusion_matrix(y_test,pred4)
print(cnf_mtx4)

# Classification report
clf_report4=classification_report(y_test,pred4)
print(clf_report4)
```

```
[[2280    2]
 [ 649   69]]
              precision    recall  f1-score   support

           0       0.78      1.00      0.88      2282
           1       0.97      0.10      0.17       718

    accuracy                           0.78      3000
   macro avg       0.88      0.55      0.52      3000
weighted avg       0.82      0.78      0.71      3000
```

Fig 26. Confusion Matrix - Support Vector Machine (SVM)

In [31]:
```
# confusion matrix
cnf_mtx5=confusion_matrix(y_test,pred5)
print(cnf_mtx5)

# classification matrix
clf_report5=classification_report(y_test,pred5)
print(clf_report5)
```

```
[[2245   37]
 [  18  700]]
              precision    recall  f1-score   support

           0       0.99      0.98      0.99      2282
           1       0.95      0.97      0.96       718

    accuracy                           0.98      3000
   macro avg       0.97      0.98      0.98      3000
weighted avg       0.98      0.98      0.98      3000
```

Fig 27. Confusion Matrix - Decision Tree

```
In [63]: # confusion matrix
         clf_mtx7=confusion_matrix(y_test,pred7)
         print(clf_mtx7)

         # classification report
         clf_report7=classification_report(y_test,pred7)
         print(clf_report7)
```

```
[[2281    1]
 [  19  699]]
              precision    recall  f1-score   support

           0       0.99      1.00      1.00      2282
           1       1.00      0.97      0.99       718

    accuracy                           0.99      3000
   macro avg       1.00      0.99      0.99      3000
weighted avg       0.99      0.99      0.99      3000
```

Fig 28. Confusion Matrix - Random Forest

Now, if we compare the accuracy of all five classification algorithms from confusion matrix and classification report, Random Forest gives best accuracy among all.

Hence, we have used Random Forest for our Final Model Development.

CHAPTER 6: SYSTEM DEPLOYMENT

6.1 DATA SOURCE

Dataset which we have used is collected from Kaggle.com and the link for the same is as follows;

HR Employee Retention | Kaggle

6.2 DEPLOYMENT ON HEROKU

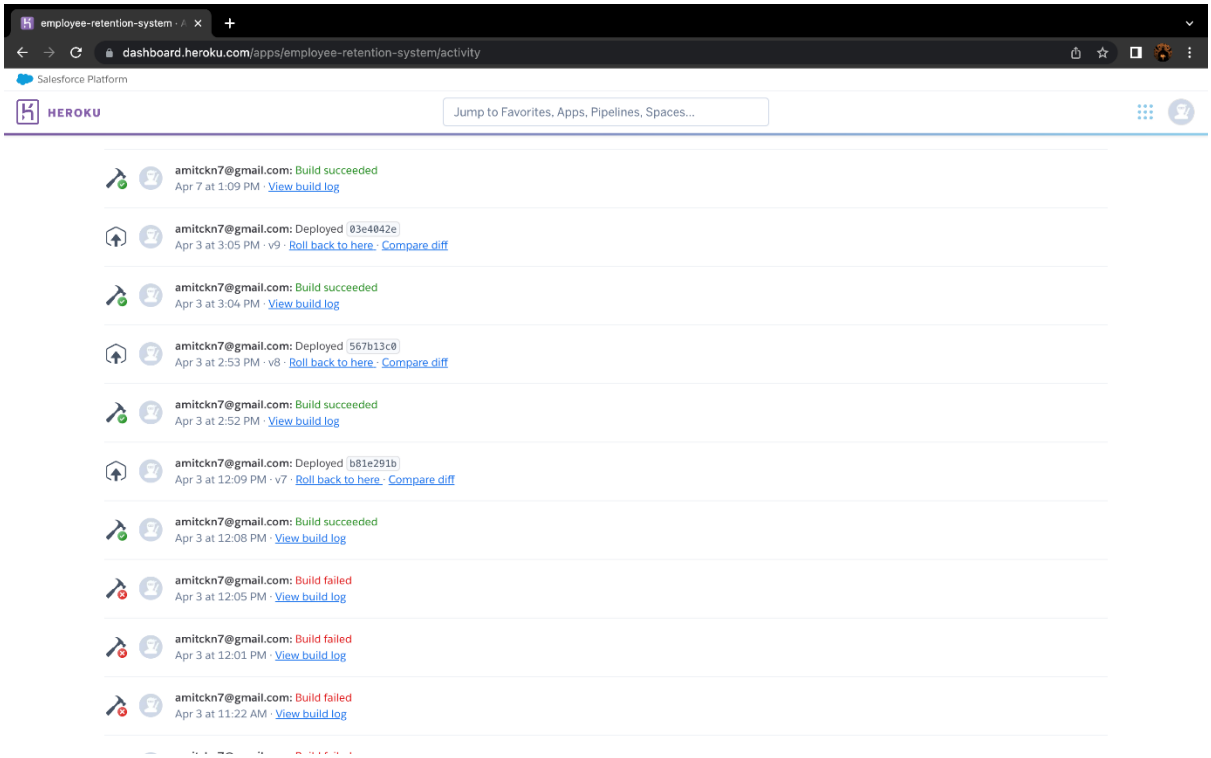

Fig 29. Deployment on Heroku

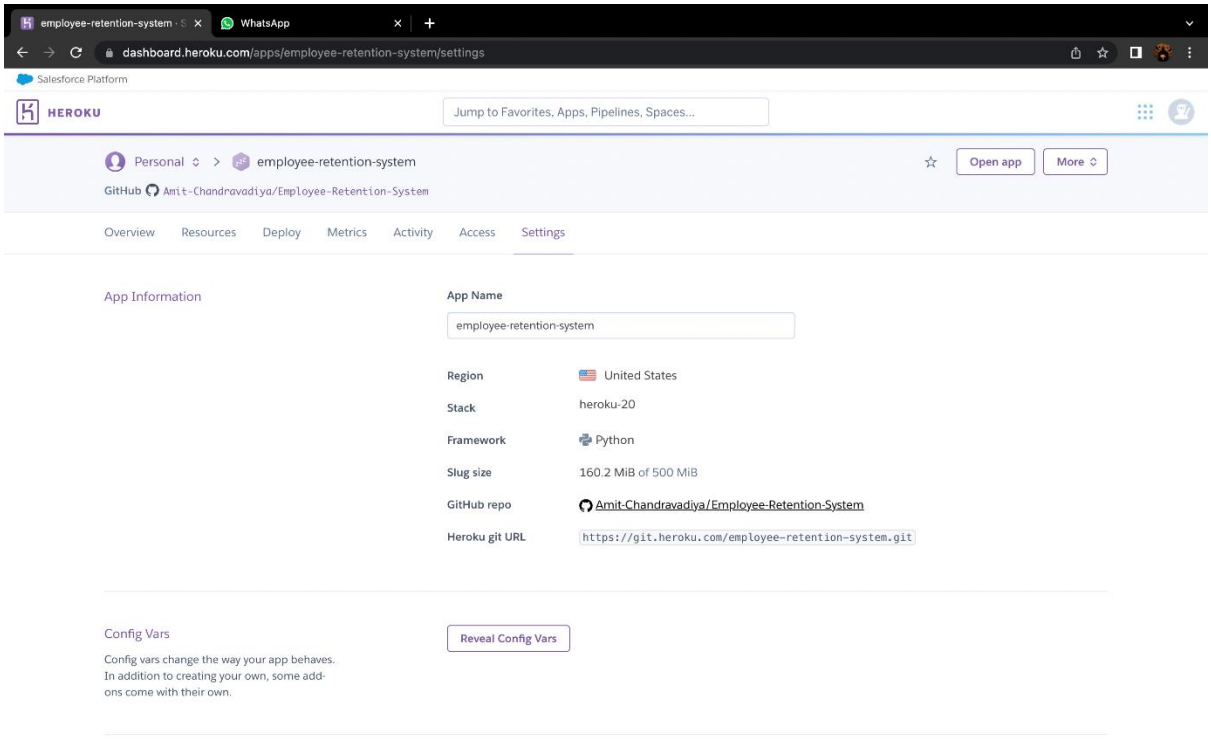

Fig 30. Final Deployment on Heroku

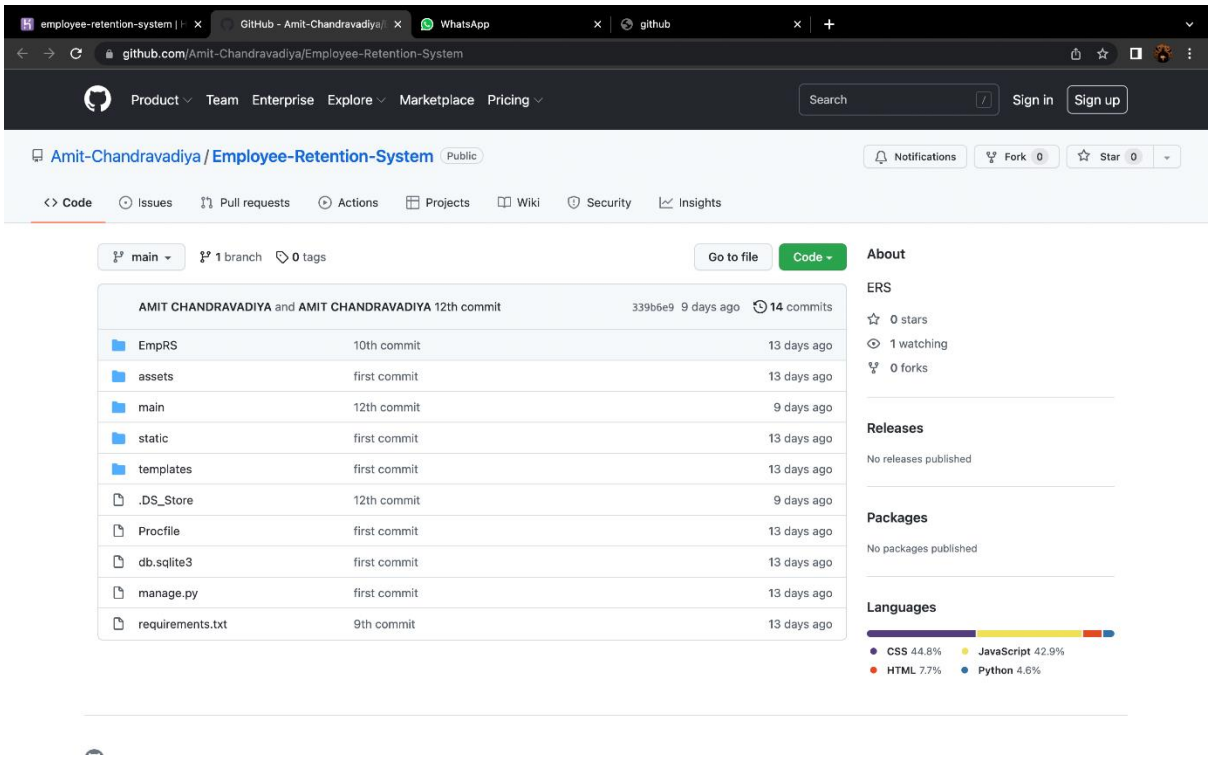

Fig 31. Project on GitHub

Currently, our project is live and the link for the same is as follows;

[Home Page (employee-retention-system.herokuapp.com)]

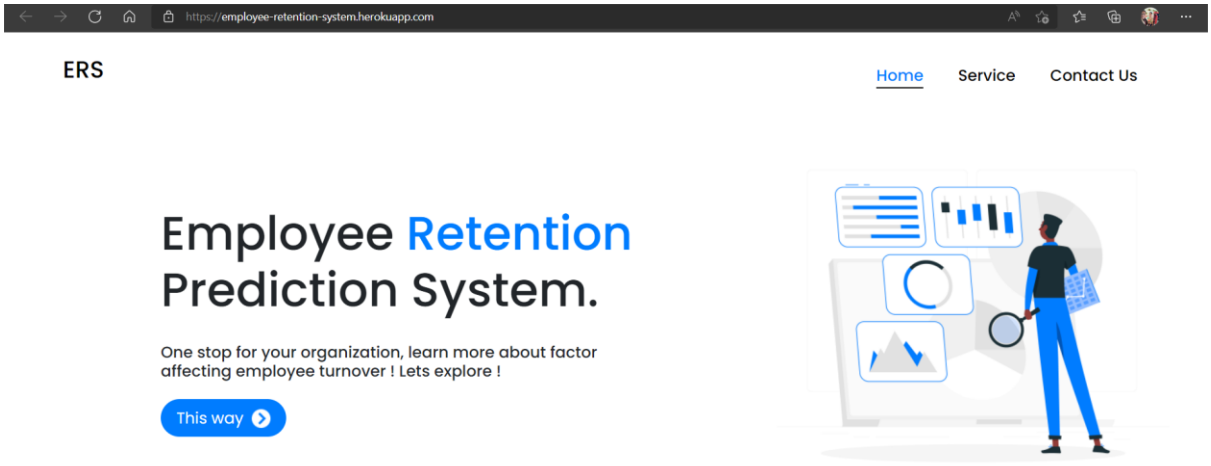

Fig 32. Home Page

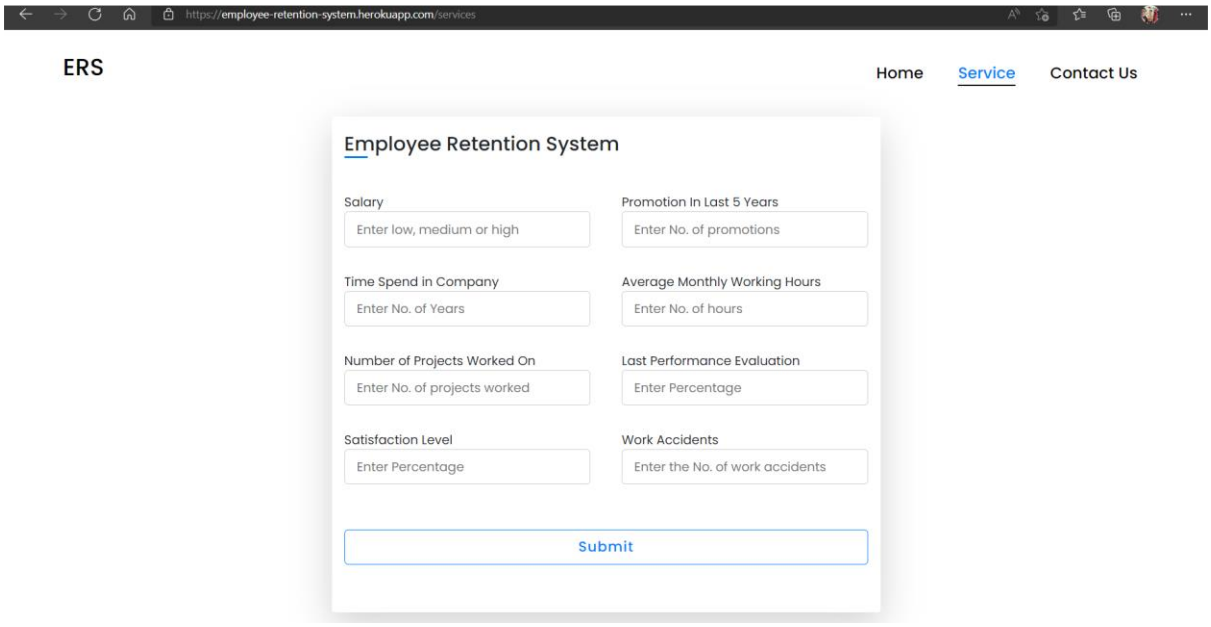

Fig 33. Service Page

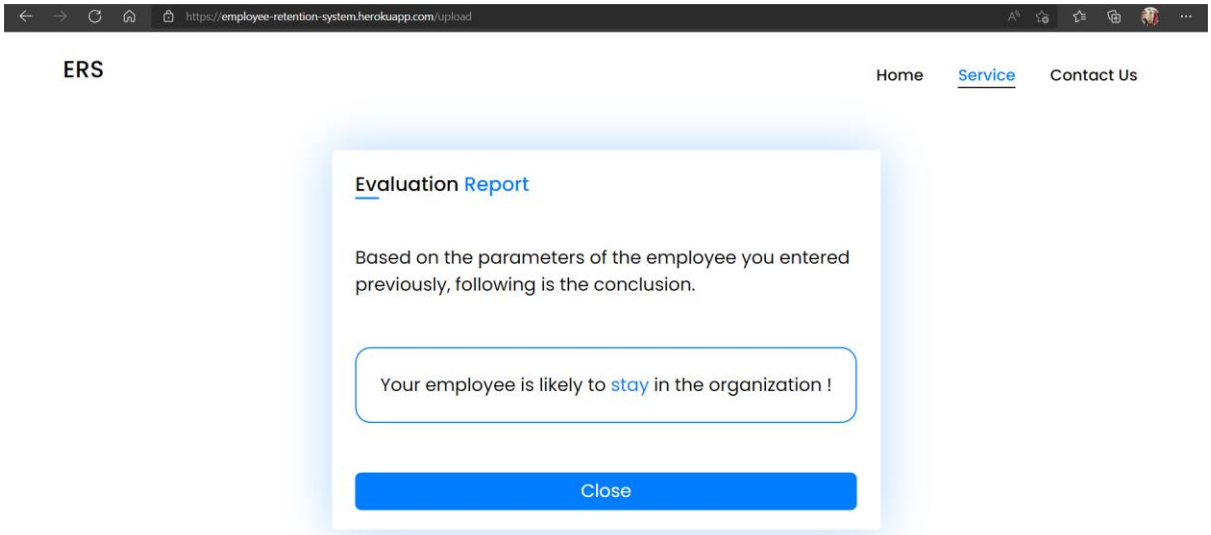

Fig 34. Output (Test case 1)

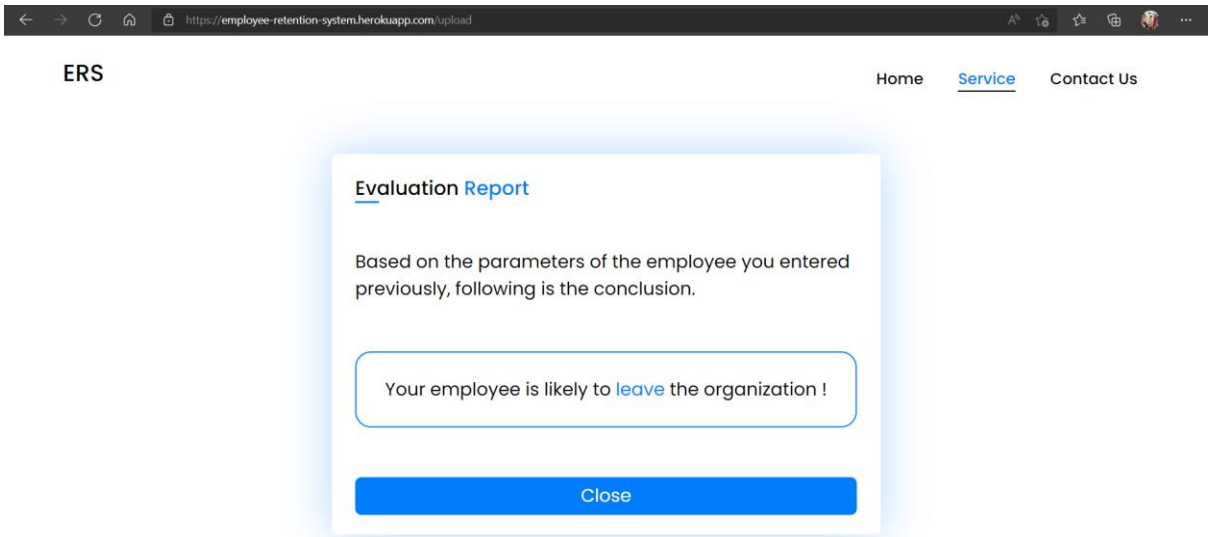

Fig 35. Output (Test case 2)

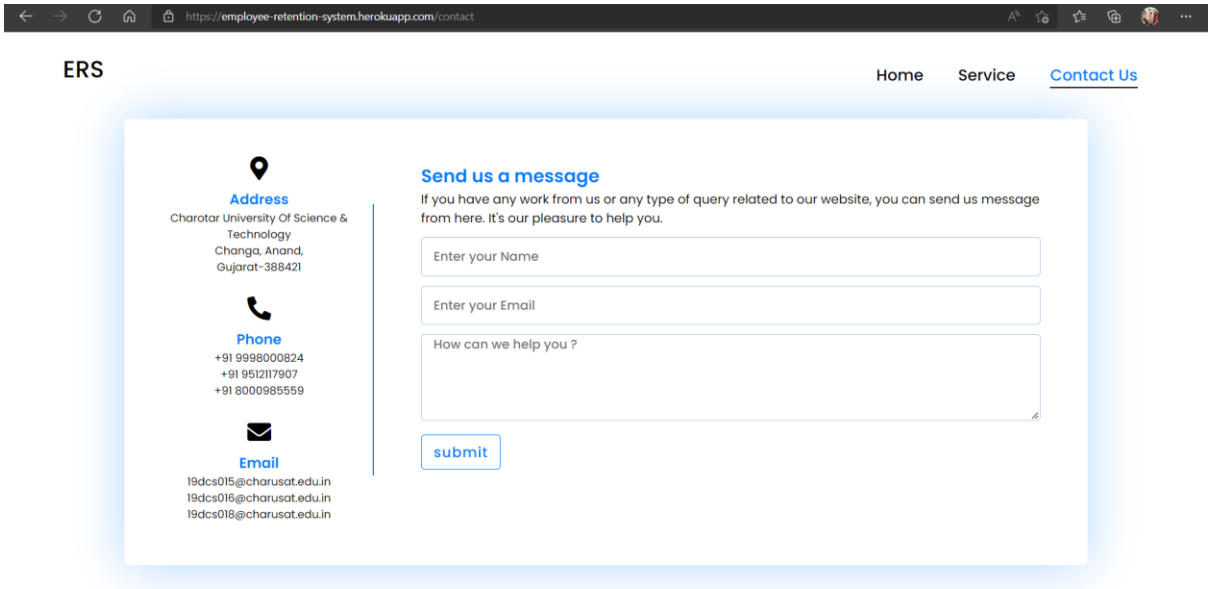

Fig 36. Contact Us page

CHAPTER 7: FUTURE ENHANCEMENT

7.1 FUTURE SCOPE

- As our project is just for some specific organisation, we are going to add some general and day-to-day parameters on which predictions can be made So that anyone from any organisation can use our project.
- Currently, we are working on the Research Work of our Project which will be completed in near Future.
- We are Planning to Publish our Research Work.
- We'll even optimize our Web-application to handle as many requests as possible and make it more responsive so it can be accessed through any Device.
- We are also Planning to make a Mobile Application of our Project.

7.2 LIMITATIONS AND CONSTRAINTS

- Currently, our project is just limited for some specific organisation, for that organisation there will be some company-specific parameters which some other organisation does not have. So, the other organisation cannot use our project.
- Our website is not much responsive as by minimising it to some extent, everything will be distorted.
- Also, it cannot handle multiple requests at a time.

CHAPTER 8: CONCLUSION

8.1 CONCLUSION

- The performance of a classification technique is affected by the quality of data source. Irrelevant and redundant features of data not only increase the elapse time, but also may reduce the accuracy of output. Each algorithm has its own advantages and disadvantages.
- There can be some parameters which are not providing significant contribution in classification, which can cause false predictions and can affect overall classification model to a great extent. So, we need to remove that parameter which is causing the same for better accuracy and correct results.
- Sometimes, by tuning hyperparameters in order to obtain greater accuracy can lead to decrease in accuracy. So, its better to use default algorithm without hyperparameter tuning.
- By doing this project we learnt the basics of Machine Learning which includes
    a. cleaning of data such as removing null values and useless columns
    b. extracting valuable insights from any dataset
    c. Comparing different algorithm which are to be used on the basis of confusion matrix and classification report
- We learnt the basics of web development like for front end (HTML, CSS, BootStrap) and for Backend (Django framework, SQLite)
- We also leant how to deploy any project on Heroku server

BIBLIOGRAPHY

1) https://www.kaggle.com/datasets/gummulasrikanth/hr-employee-retention
2) https://www.python.org/doc/
3) https://pandas.pydata.org/pandas-docs/stable/
4) https://numpy.org/doc
5) https://matplotlib.org/stable/index.html
6) https://scikit-learn.org/stable/index.html
7) https://seaborn.pydata.org
8) https://devdocs.io/html
9) https://devdocs.io/css
10) https://getbootstrap.com/docs/5.0
11) https://docs.djangoproject.com
12) https://docs.github.com/en
13) https://devcenter.heroku.com/categories/reference
14) https://sqlite.org/docs.html
15) Akuoko, O. K. (2012). "Employee Retention Strategies and Workers‟ Performance: General Views of Employees in Ashanti Region of Ghana",International Journal of Business and Management Tomorrow, Vol. 2 No. 8 ,1-19.
16) Beck, S. (2001). „Why Associates Leave and Strategies To Keep Them.‟ In American Lawyer Media L.P., v5, i2, pp. 23-27.
17) Daniel Eseme Gberevbie, (2010), "Organizational Retention Strategies and Employee Performance of Zenith Bank in Nigeria", African Journal of Economic and Management Studies, Vol. 1 Iss: 1.
18) Julia Christensen Hughes, Evelina rog, (2008),"Challenge of Employee Retention", International Journal of Contemporary Hospitality Management, Vol. 20, Issue 7.