ABSTRACT

Employee Attrition is one of the biggest business problems in HR Analytics. The study reveals to attrition of employees in the company. Through the study has been assessed that the employees are having safe and comfortable working environment in their company. Predicting whether a particular employee may leave or not will help the company to make preventive decisions. Companies invest a lot in the training of the employees keeping in mind the returns they would provide to the company in the future. If an employee leaves the company, it is the loss of opportunity cost to the company. The continued growth of the company depends upon in attrition their valuable employees who are the pillars of the organization. This project works on employee attrition prediction using machine learning by random forest algorithm. Employees are the most important part of an organization. The company can also go for introducing new incentives schemes, transport facility, accommodation facility and to increase the bonus amount which helps the organization to motivate their employees to work even more and this brings best result from the employees. Employee attrition is a major cost to an organization and predicting such attrition is the most important requirement of the Human Resources department in many organizations.

CHAPTER 1
INTRODUCTION

INTRODUCTION

1.1 Overview

Employee attrition is also known as Labour attrition. Attrition defined as employment loss such as sudden resignation, personal health , or other similar reasons. Losing an talented and well trained employee drastically effects the organization regarding in making an employee more skilful. The attrition rate tends to vary from skilled and unskilled labours .Whenever there is a hiring of new employee then at that period of time there is increase in cost of recruitment and training.It is the keen responsibility of the HR manager to hire a well natured , faithful, trained and workaholic employees are required to run a successful organization.he employee should have a best knowledge about his work which is assigned to him so as to providing preventive techniques that are required to decrease the attrition rate and update it to Manager ,upgrade the company abundantly.

1.2 Problem Definition

Employees plays the role of a huge asset in any company and they are those who work for the vision and mission of the same. So consistency of their presence and work is crucial for the company’s growth in the competitive world. Therefore it is important for an HR analyst to predict the factors of dissatification faced during the work so that employees wish to be retained.

CHAPTER-2

LITERATURE SURVEY

LITERATURE SURVEY

Corportate Employee Attrition Analysis

Shenghuan Yang-Jiangxi University of Finance and Economics, Md Tariqul Islam Syracuse University

This utilized Random Forest and K-means Clustering to select important features that had obvious impact on the employee attrition. Firstly, according to Random Forest results, monthly income, age, the number of companies worked are the main reasons why people choose to resign. Then older people, high job level, high job satisfaction, high monthly income, number of companies worked, these kinds of people are not likely to go based on the clustering result of K-means Clustering.

This study found that females' attrition was 0.659 times than that of males, married and divorced people were 0.427 and 0.304 times than people who were single, respectively. Besides, the attrition of people who travelled frequently was 2.4 times higher than that of people who rarely travelled. Finally, there are other interesting findings in our study in terms of number of companies worked, people who worked in 2-4 companies are less likely to leave, the female attrition rate is less than male after working for six companies, and people who earned Doctor's Degree are almost always having the lowest attrition rate.

Attrition Analysis in a Leading Sales Organisation in India
Mamta Mohapatra, International Management Institute, New Delhi, Nikita Lamba Genpact, New Delhi, India

The paper highlights the importance for sales organizations to realise their attrition rates and identify factors leading to it. The sales employees directly interact with the

customers on daily basis. For more satisfied customers, removing job dissatisfiers is considered essential. Thereafter, the results of the survey are discussed. The survey identifies some variables like politics, role ambiguity and supervisor related issues that play a major role in influencing the attrition rate in a sales company. The impact of these variables for employees in different locations (facing different problems), different tenure range and grades is discussed.

However, there are a number of limitations to this study. The impact of these variables across genders, educational background, experience range and performance levels can be measured. After an in-depth analysis, the results can be generalized in the context of sales industry in India. This data can be helpful for the organisations which are striving to identify the influencers in employee attrition amongst their sales people. It can help the companies design better retention strategies, thereby, reducing attrition costs.

Employee Attrition Prediction Using Machine Learning Algorithms
Lok Sundar Ganthi, Yaswanthi Nallapaneni, Deepalakshmi Perumalsamy & Krishnakumar Mahalingam Conference paper, 23rd November 2021.

This paper used machine learning approaches to forecast to prevent the aforementioned scenarios. With the aid of certain relevant data, workers who want to quit the organization can be exploited. Finding the characteristics that motivate workers to leave their job. Utilizing the categorization algorithms, namely Decision, to forecast employee attrition rate extreme gradient, tree, Random-forest, K-Nearest Neighbors, and neural networks. This paper tried boosting and Ada-Boosting. Additionally, this inference paper implemented regularisation for each algorithm to determine the appropriate criteria to forecast the attrition rate of
employees taking 35 feature HR-data set from the Kaggle website,
which includes 34 of them
This attrition feature, includes Yes/No answers in both independent and one

dependent feature. In this paper, and are going through different steps to finally obtain an accuracy of 88% with good precision and recall values.

Singh, Moninder, etc al. "An analytics approach for proactively combating voluntary attrition of employees." 2012 IEEE 12th International Conference on Data Mining Workshops. IEEE, 2012

This paper discusses a proactive approach to lowering employee attrition. This is particularly crucial for businesses with sizable service divisions because the unexpected departure of key members can result in significant losses in terms of lost productivity, missed deadlines, and hiring expenses for replacements. The proactive compensation increases given to atrisk employees is the main retention strategy examined in this paper. In order to make the best use of any limited funds that may be available for this purpose, the paper approach uses data mining like clustering to identify employees at risk of attrition and weighs the cost of attrition/replacement of an employee against the cost of retaining that employee. This allows the action to be targeted toward employees with the highest potential returns on investment. The retention action was carried out in two phases. The first phase involved around 7500 employees in all but one of the business areas considered. the second phase was carried out a couple of months later and involved roughly 12000 employees from the previously left out business area. The total net benefit estimated by the company's HR department is approximately 150% during the 2012 calendar year. This estimate is based on the assumption of a certain average ’success’ attrition rate amongst the targeted employees, based on a limited retention action that had been carried out in a prior year.

CHAPTER-3
SYSTEM ANALYSIS

SYSTEM ANALYSIS

3.1 Existing System

Employee attrition prediction by using various classification algorithms like logistic regression, LDA, SVM, KNN to predict the probability of attrition of any new employee. Deep learning technique along with some preprocessing steps to improve the prediction of employee attrition. Several factors lead to employee attrition. Such factors are analyzed to reveal their intercorrelation and to demonstrate the dominant ones.

3.2 Proposed System

HR Analysts can make necessary arrangements to hire a specific group of specialists who take care of the analysis process of the employees so that he/she can be relieved from the burden of completing the task in person which further can be done using the cognos analytics.

The manager has to make sure the working environment is comfortable for the employees to work in and if they are physically and mentally fit for the job and

motivate them with job security, bonuses and increment according.

3.3 REQUIREMENTS ANALYSIS AND SPECIFICATIONS

The requirement engineering process of the feasibility study, requirements elicitation and analysis, requirements specifications, requirements validation, and requirements management. Requirements elicitation and analysis is an iterative process that can be represented as a spiral of activities, namely requirements discovery, requirements classification and organization, requirements negotiation, and requirements documentation. tasks

3.3.1 INPUT REQUIREMENTS

The basic input requirements include a stable internet connection with a suitable system, and the browser to work. On the other hand, we need a Dataset of employees.

3.3.2 OUTPUT REQUIREMENTS

The output requirements include a full-fledged computer system for doing interaction and the mail account signed up on any of the popular mail service platforms to send a notification regarding the vehicle lost.

3.3.3 FUNCTIONAL REQUIREMENTS

The functional requirements needed to implement this projects are the location data that needs to be sent and stored in the cloud for quick access and the stable electricity to run the server 24/7 and some pre-installed software to work on with the system.

3.4 SOFTWARE ENVIRONMENT

1. Operating System: Windows | Linux | Mac | any other stable operating system
2. Language used: Python
3. Tools: Colab.
4. DataBase: ExcelSheet.

3.5 SOFTWARE DESCRIPTION

Python:

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components. Python's simple, easy-to-learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms and can be freely distributed.

Often, programmers fall in love with Python because of the increased

productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. A source-level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a program is to add a few print statements to the source: the fast edit-test-debug cycle makes this simple approach very effective.

Colab:

Colaboratory is a data analysis tool that combines code, output, and descriptive text into one document (interactive notebook).

Colab provides GPU and is free. By using Google Colab, you can:

1. Build your analytics products quickly in a standardizedenvironment.
2. Facilitates popular DL libraries on the go such as PyTorch, andTensorFlow
3. Share code & results within your GoogleDrive
4. Save copies and create playground modes for knowledgesharing
5. Colab is runnable on the cloud or local server with Jupyter

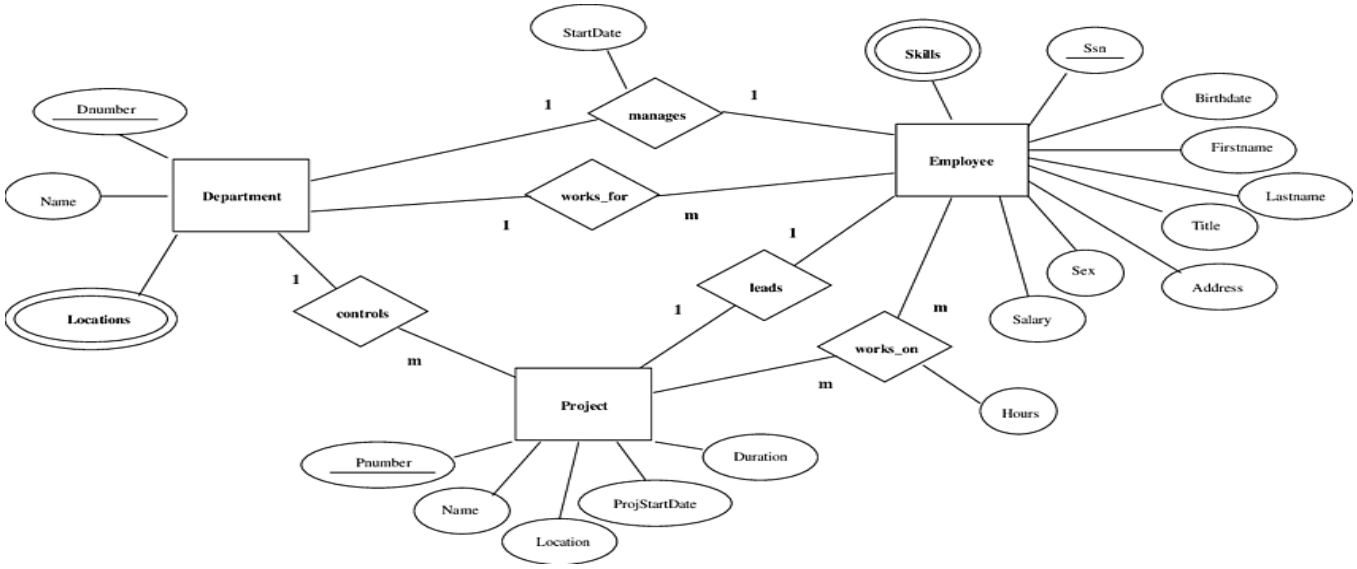
CHAPTER-4

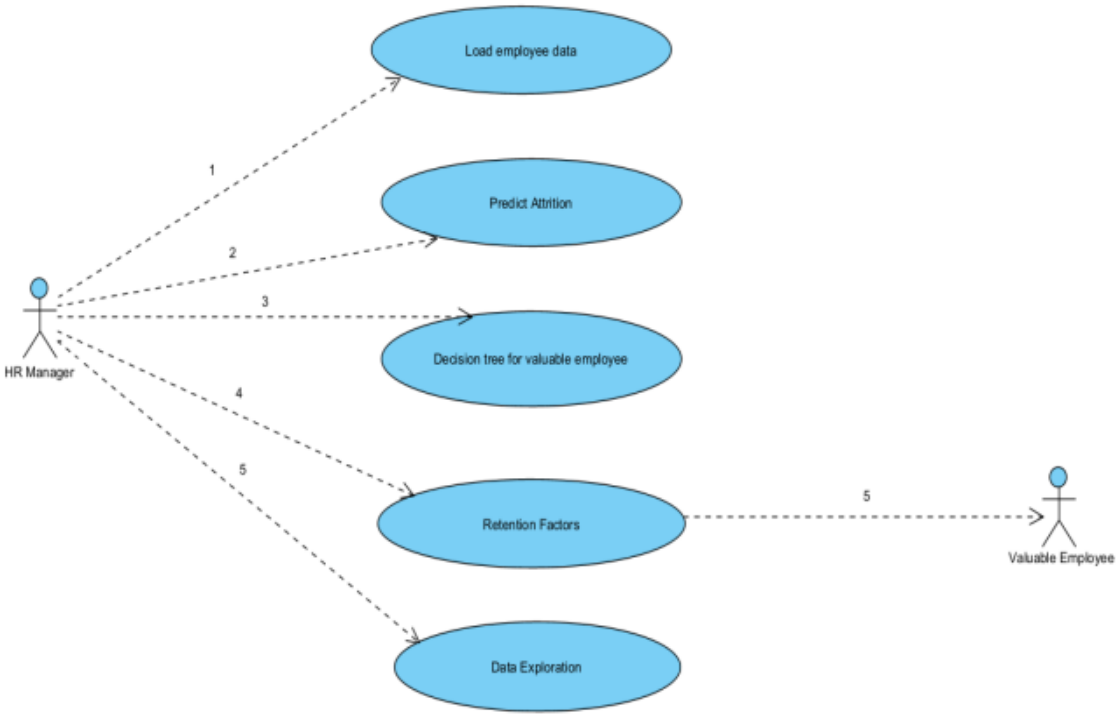
SYSTEM DESIGN

4. SYSTEM DESIGN

4.1 ER-DIAGRAM

ER Diagram stands for Entity Relationship Diagram, also known as ERD is a diagram that displays the relationship of entity sets stored in a database. In other words, ER diagrams help to explain the logical structure of databases. ER diagrams are created based on three basic concepts: entities, attributes, and relationships.




Load employee data
Predict Attrition
Decision tree for valuable employee
Retention Factors
Data Exploration
HR Manager
Valuable Employee
1
2
3
4
5
5

Test dataset and training dataset:

Separating data into test datasets and training datasets is an important part of evaluating data mining models. By this separation of total data set into two data sets we can minimize the effects of data inconsistency and better understand the characteristics of the model. The test data set contains all the required data for data prediction and training data set contains all irrelevant data. Here we have 788 records in test dataset and 682 records in training dataset. We apply data classification and data prediction on the test dataset of 788 records.

MODULE DESCRIPTION

1.Importing libraries that will be used throughout this program.

i. numpy
ii. pandas
iii. seaborn

2.Loading the data.

3.Storing the data into a dataframe.

4.Get the number of rows and columns in the data.

#There are 1470 rows of data or employees in the data set and 35 columns or data points on each employee.

5.Get the column data types.

```
Age                                 int64
Attrition                          object
BusinessTravel                     object
DailyRate                           int64
Department                         object
DistanceFromHome                    int64
Education                           int64
EducationField                     object
EmployeeCount                       int64
EmployeeNumber                      int64
EnvironmentSatisfaction             int64
Gender                             object
HourlyRate                          int64
JobInvolvement                      int64
JobLevel                            int64
JobRole                            object
JobSatisfaction                     int64
MaritalStatus                      object
MonthlyIncome                       int64
MonthlyRate                         int64
NumCompaniesWorked                  int64
Over18                             object
OverTime                           object
PercentSalaryHike                   int64
PerformanceRating                   int64
RelationshipSatisfaction            int64
StandardHours                       int64
StockOptionLevel                    int64
TotalWorkingYears                   int64
TrainingTimesLastYear               int64
WorkLifeBalance                     int64
YearsAtCompany                      int64
YearsInCurrentRole                  int64
YearsSinceLastPromotion             int64
YearsWithCurrManager                int64
```

6.Get a count of the number of empty values in each column.

```
Age                                 0
Attrition                           0
BusinessTravel                      0
DailyRate                           0
Department                          0
DistanceFromHome                    0
Education                           0
EducationField                      0
EmployeeCount                       0
EmployeeNumber                      0
EnvironmentSatisfaction             0
Gender                              0
HourlyRate                          0
JobInvolvement                      0
JobLevel                            0
JobRole                             0
JobSatisfaction                     0
MaritalStatus                       0
MonthlyIncome                       0
MonthlyRate                         0
NumCompaniesWorked                  0
Over18                              0
OverTime                            0
PercentSalaryHike                   0
PerformanceRating                   0
RelationshipSatisfaction            0
StandardHours                       0
StockOptionLevel                    0
TotalWorkingYears                   0
TrainingTimesLastYear               0
WorkLifeBalance                     0
YearsAtCompany                      0
YearsInCurrentRole                  0
YearsSinceLastPromotion             0
YearsWithCurrManager                0
```

Here there is no missing data since all of the columns are returning a value of 0. Let’s double check the data set for any missing values.

1. Double checking the data set for any missing values.

returned a value = False, indicates that there are no missing values

2. Viewing some basic statistics about the data like the percentile, maxmium, minimum etc.

9.Get a count of the number of employee attrition, the number of employees that stayed (no) and the number that left (yes) the company.

```
No     1233
Yes     237
Name: Attrition, dtype: int64
```

10.Now that we have the count, let's get a visual of it.

11.Show the number of employees that left and stayed at the company by age

Here the age with the highest count of employee attrition is age 29 & 31. The age with the highest retention is age 34 & 35.

12. Print all of the object data types and print their unique values.

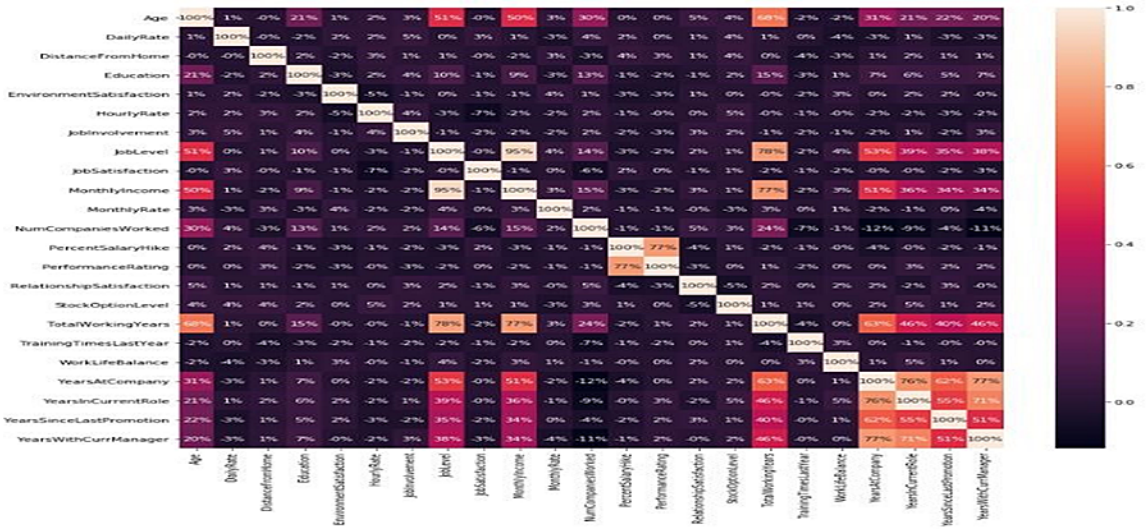
```
Attrition : ['Yes' 'No']
No      1233
Yes      237
Name: Attrition, dtype: int64
_____________________________________________________________________
BusinessTravel : ['Travel_Rarely' 'Travel_Frequently' 'Non-Travel']
Travel_Rarely        1043
Travel_Frequently     277
Non-Travel            150
Name: BusinessTravel, dtype: int64
_____________________________________________________________________
Department : ['Sales' 'Research & Development' 'Human Resources']
Research & Development    961
Sales                     446
Human Resources            63
Name: Department, dtype: int64
_____________________________________________________________________
EducationField : ['Life Sciences' 'Other' 'Medical' 'Marketing' 'Technical Degree'
 'Human Resources']
Life Sciences        606
Medical              464
Marketing            159
Technical Degree     132
Other                 82
Human Resources       27
Name: EducationField, dtype: int64
_____________________________________________________________________
Gender : ['Female' 'Male']
Male      882
Female    588
Name: Gender, dtype: int64
_____________________________________________________________________
JobRole : ['Sales Executive' 'Research Scientist' 'Laboratory Technician'
 'Manufacturing Director' 'Healthcare Representative' 'Manager'
 'Sales Representative' 'Research Director' 'Human Resources']
Sales Executive              326
Research Scientist           292
Laboratory Technician        259
Manufacturing Director       145
Healthcare Representative    131
Manager                      102
Sales Representative          83
Research Director             80
Human Resources               52
Name: JobRole, dtype: int64
_____________________________________________________________________
MaritalStatus : ['Single' 'Married' 'Divorced']
Married     673
Single      470
Divorced    327
```

13.Remove unneeded columns

14.Getting the correlation of the columns.

	Age	DailyRate	DistanceFromHome	Education	EnvironmentSatisfaction	HourlyRate
Age	1.000000	0.010661	-0.001686	0.208034	0.010146	0.024287
DailyRate	0.010661	1.000000	-0.004985	-0.016806	0.018355	0.023381
DistanceFromHome	-0.001686	-0.004985	1.000000	0.021042	-0.016075	0.031131
Education	0.208034	-0.016806	0.021042	1.000000	-0.027128	0.016775
EnvironmentSatisfaction	0.010146	0.018355	-0.016075	-0.027128	1.000000	-0.049857
HourlyRate	0.024287	0.023381	0.031131	0.016775	-0.049857	1.000000
JobInvolvement	0.029820	0.046135	0.008783	0.042438	-0.008278	0.042861
JobLevel	0.509604	0.002966	0.005303	0.101589	0.001212	-0.027853
JobSatisfaction	-0.004892	0.030571	-0.003669	-0.011296	-0.006784	-0.071335
MonthlyIncome	0.497855	0.007707	-0.017014	0.094961	-0.006259	-0.015794
MonthlyRate	0.028051	-0.032182	0.027473	-0.026084	0.037600	-0.015297
NumCompaniesWorked	0.299635	0.038153	-0.029251	0.126317	0.012594	0.022157
PercentSalaryHike	0.003634	0.022704	0.040235	-0.011111	-0.031701	-0.009062
PerformanceRating	0.001904	0.000473	0.027110	-0.024539	-0.029548	-0.002172
RelationshipSatisfaction	0.053535	0.007846	0.006557	-0.009118	0.007665	0.001330
StockOptionLevel	0.037510	0.042143	0.044872	0.018422	0.003432	0.050263
TotalWorkingYears	0.680381	0.014515	0.004628	0.148280	-0.002693	-0.002334
TrainingTimesLastYear	-0.019621	0.002453	-0.036942	-0.025100	-0.019359	-0.008548
WorkLifeBalance	-0.021490	-0.037848	-0.026556	0.009819	0.027627	-0.004607
YearsAtCompany	0.311309	-0.034055	0.009508	0.069114	0.001458	-0.019582
YearsInCurrentRole	0.212901	0.009932	0.018845	0.060236	0.018007	-0.024106
YearsSinceLastPromotion	0.216513	-0.033229	0.010029	0.054254	0.016194	-0.026716
YearsWithCurrManager	0.202089	-0.026363	0.014406	0.069065	-0.004999	-0.020123

15.Visualizing the correlation by using a heat map.



16.Spliting the dataset into 75% Training set and 25% Testing set.

17.Now by using the Random Forest Classifier to learn from the training data and see how accurate it is on that data.

18.Now get the accuracy of the model.

The model is about 97.9% accurate on the training data

19.Now Showing the confusion matrix and accuracy of the model.

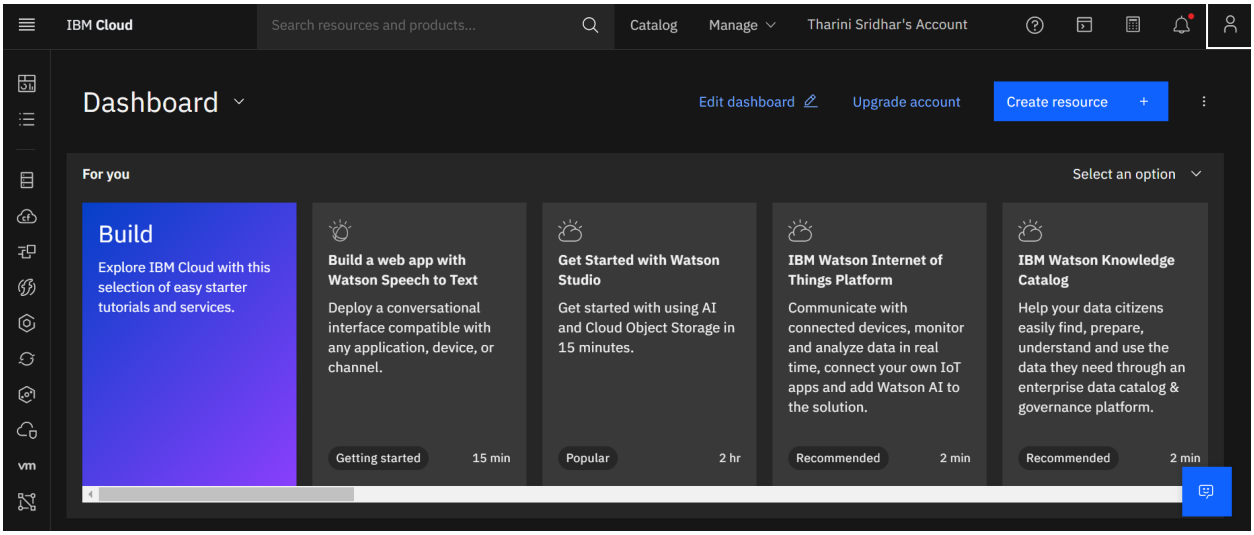
```
[[309   1]
 [ 49   9]]
Model Testing Accuracy = "0.8641304347826086!"
```

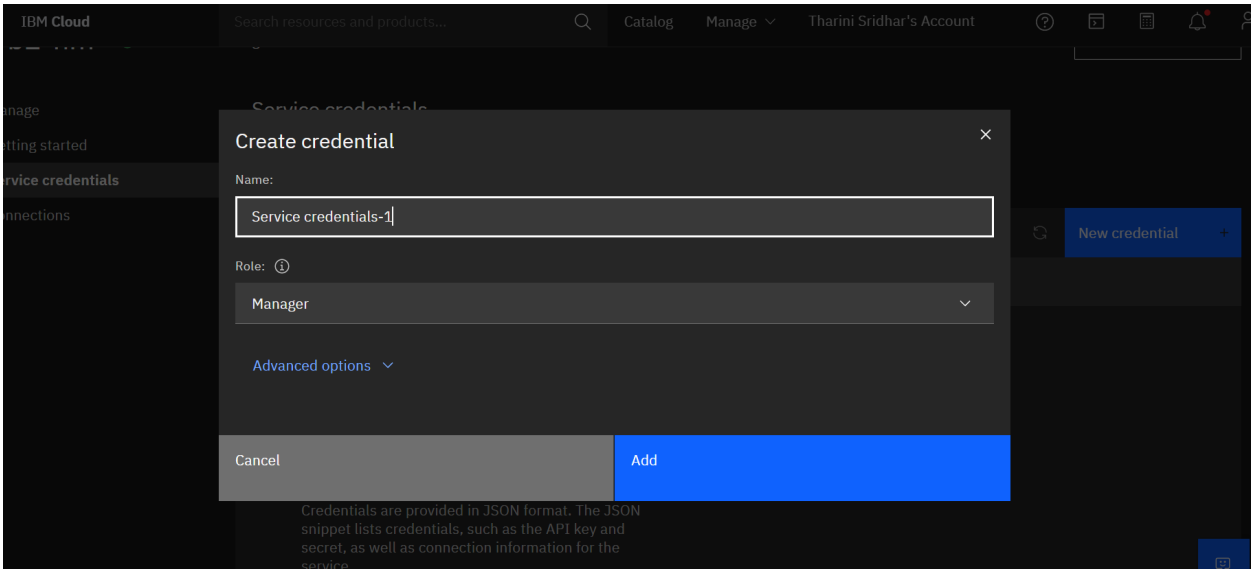
The model correctly identified 86.41% of the employees that left the company.

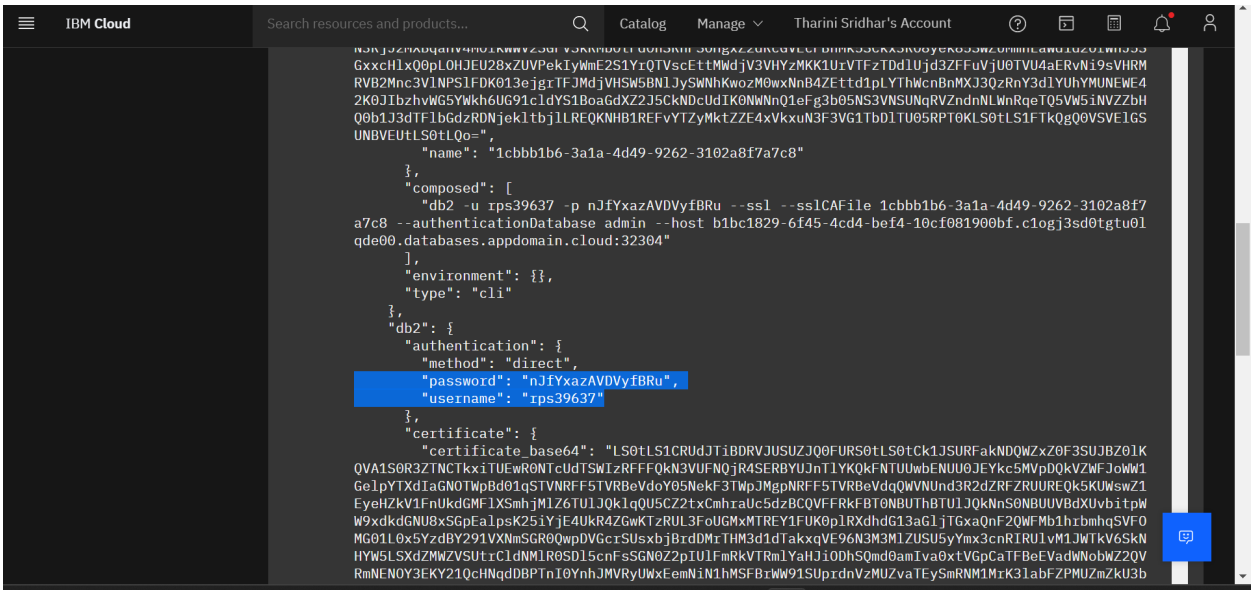
CHAPTER 6
SYSTEM IMPLEMENTATION

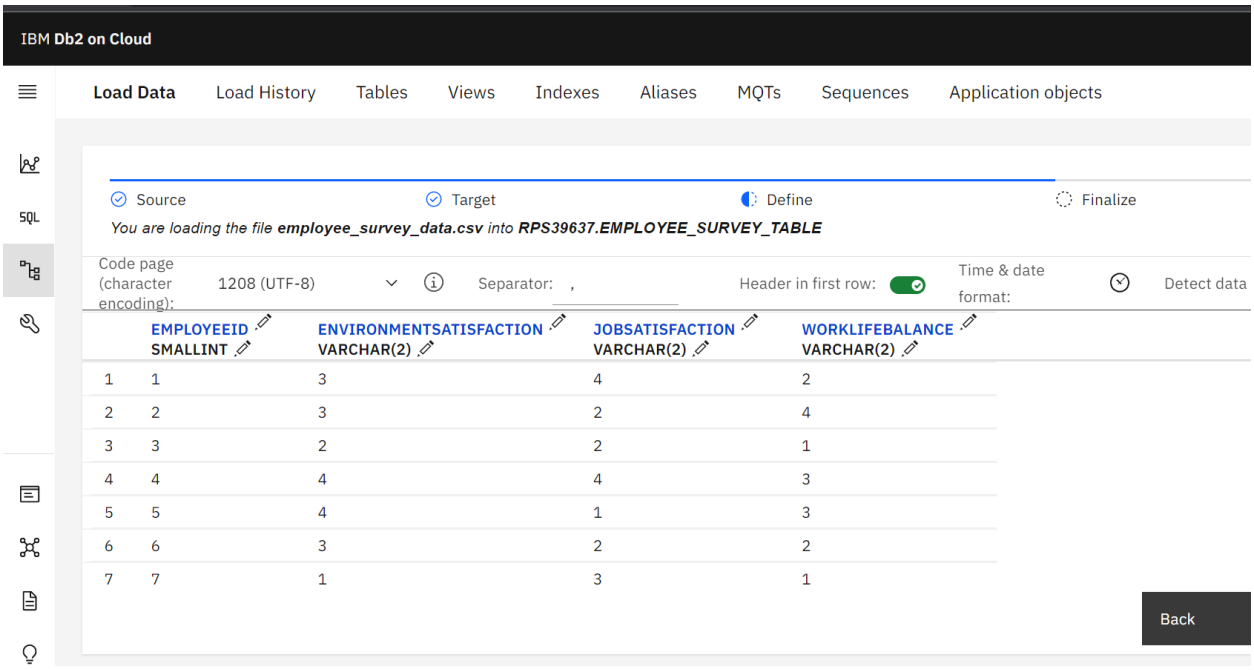
6.1 SAMPLE SCREENS

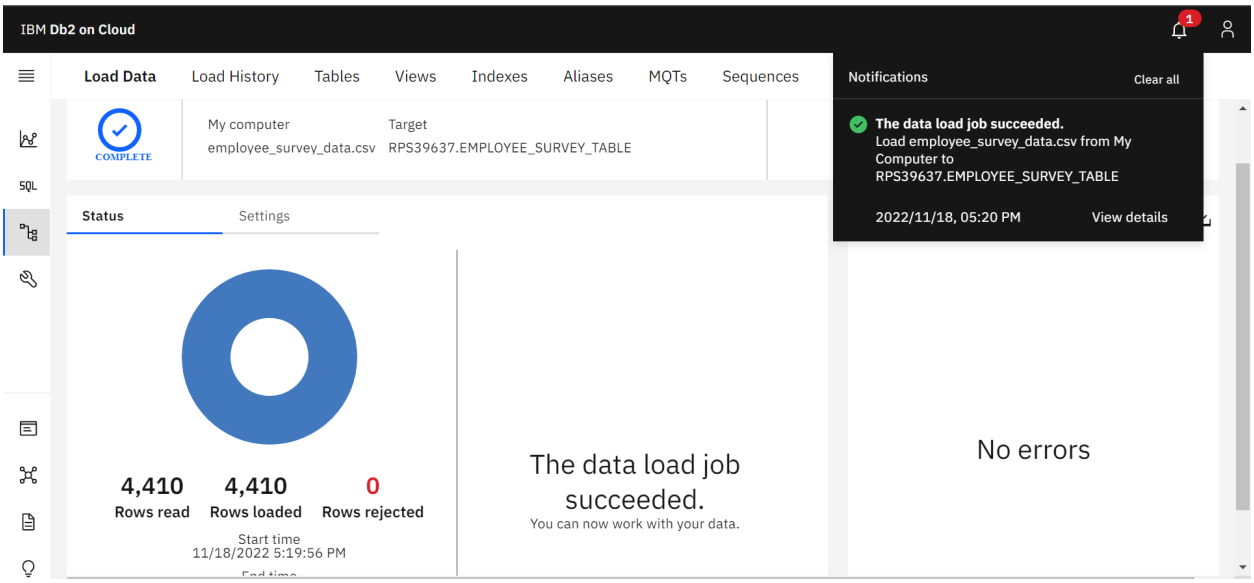
Coud Server Connection





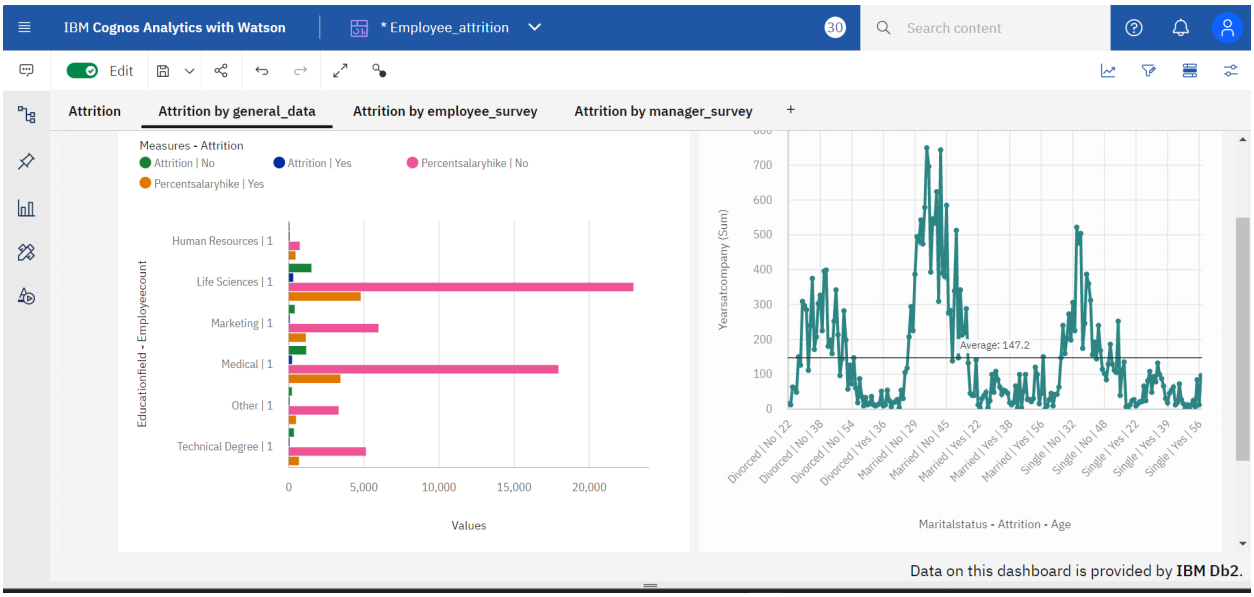

IBM Cloud
Search resources and products...
Catalog
Manage
Tharini Sridhar's Account
"password": "nJfYxazAVDVyfBRu",
"username": "rps39637"

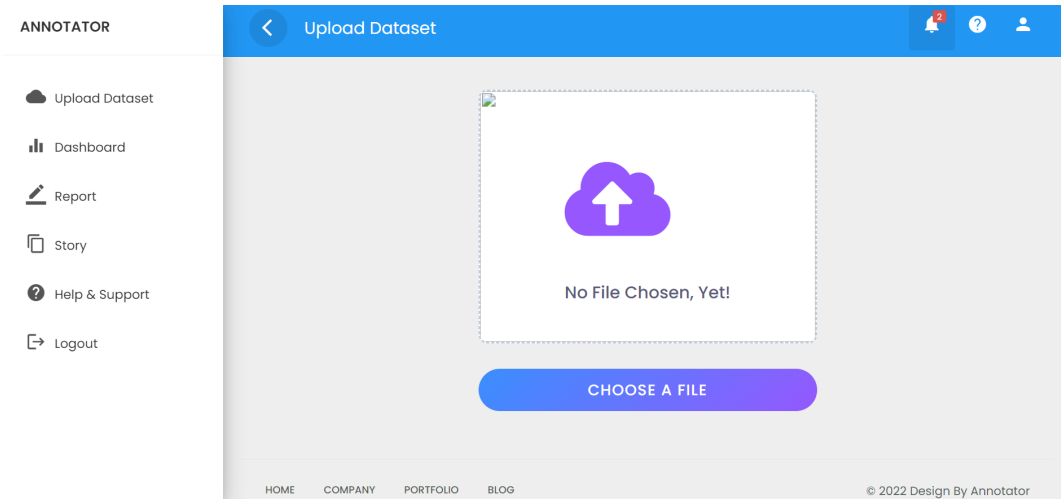

IBM Db2 on Cloud
Load Data
Load History
Tables
Views
Indexes
Aliases
MQTs
Sequences
Application objects
Source
Target
Define
Finalize
You are loading the file employee_survey_data.csv into RPS39637.EMPLOYEE_SURVEY_TABLE
Code page (character encoding):
1208 (UTF-8)
Separator: ,
Header in first row:
Time & date format:
Detect data
EMPLOYEEID SMALLINT
ENVIRONMENTSATISFACTION VARCHAR(2)
JOBSATISFACTION VARCHAR(2)
WORKLIFEBALANCE VARCHAR(2)
1 1 3 4 2
2 2 3 2 4
3 3 2 2 1
4 4 4 4 3
5 5 4 1 3
6 6 3 2 2
7 7 1 3 1
Back



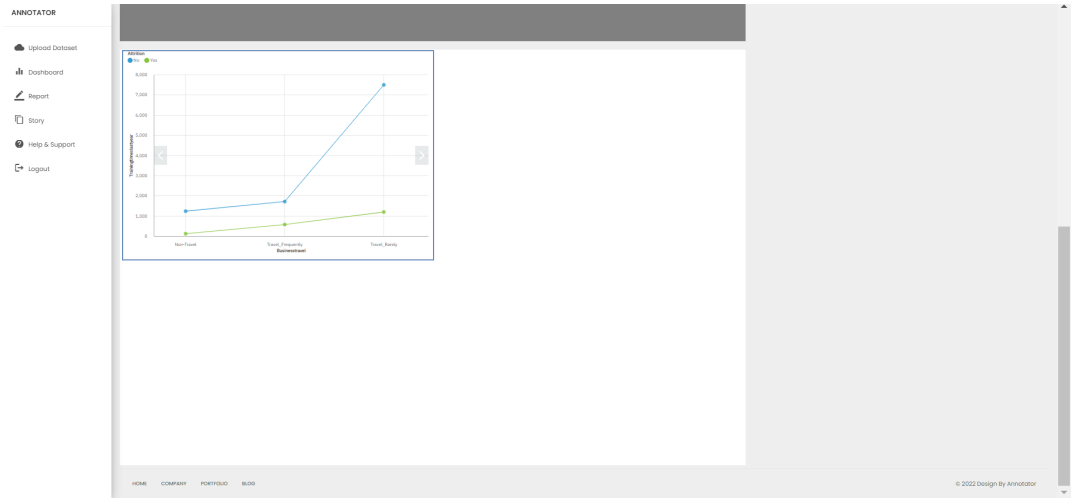
Output

Factors analysis

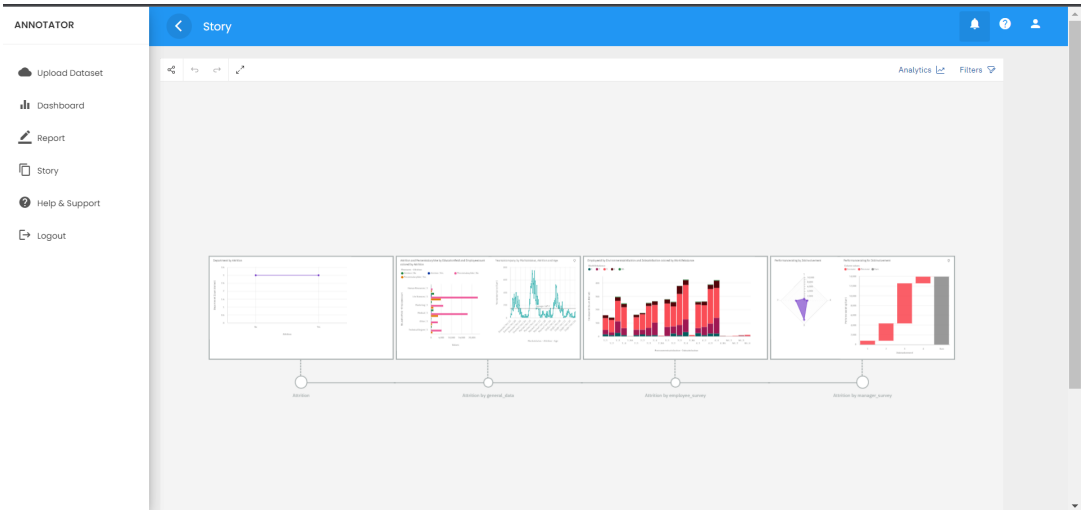




REPORT



STORY



6.2 SAMPLE CODING:

#ImportLibraries

import num py as np
import panda sa spd
import se born as sns

#Load the data from google.colab import files # Use to load data on Google Colab uploaded = files.upload() # Use to load data on Google Colab

#Store the data into the df variable

df = pd.read_csv('WA_Fn-UseC_-HR-Employee-Attrition.csv') df.head(7) *#Print the first 7 rows*

#Get the number of rows and number of columns in the data df.shape

#Get the column data types

df.dtypes

#Count the empty (NaN, NAN, na) values in each column df.isna().sum()

#Another check for any null / missing values

df.isnull().values.any()

#View some basic statistical details like percentile, mean, standard deviation etc. df.describe()

#Get a count of the number of employee attrition, the number of employees that stayed (no) and the number that left (yes)

df['Attrition'].value_counts()

```
#Visualize this count
sns.countplot(df['Attrition'])
#Show the number of employees
that left and stayed by
ageimportmatplotlib.pyplotasp
lt fig_dims = (12, 4)
fig, ax = plt.subplots(figsize=fig_dims)

#ax = axis
sns.countplot(x='Age', hue='Attrition', data = df,
palette="colorblind", ax = ax,
edgecolor=sns.color_palette("dark", n_colors = 1)); #Print
all of the object data types and their unique valuesfor
column in df.columns: if df[column].dtype == object:
print(str(column) + ' : ' + str(df[column].unique()))
print(df[column].value_counts())
print("_____
_____")

#Remove unneeded columns
#Remove the column EmployeeNumber
df = df.drop('EmployeeNumber', axis = 1) # A number assignment
#Remove the column StandardHours
df = df.drop('StandardHours', axis = 1) #Contains only value 80
#Remove the column EmployeeCount
df = df.drop('EmployeeCount', axis = 1) #Contains only the value 1
#Remove the column EmployeeCount
df = df.drop('Over18', axis = 1) #Contains only the value 'Yes'

#Get the correlation of the
columns df.corr()
```

```
#Visualize the correlation
plt.figure(figsize=(14,14)) #14in by
14in
sns.heatmap(df.corr(), annot=True, fmt='.0%')

#Transform non-numeric columns into
numerical
columnsfromsklearn.preprocessingimpo
rt LabelEncoder for column in df.columns:
if df[column].dtype == np.number:
continue
df[column] = LabelEncoder().fit_transform(df[column])

#Create a new column at the end of the dataframe that contains the same
value  df['Age_Years'] = df['Age']
#Remove the first column called age
df =
df.drop('A
ge', axis =
1) #Show
the
dataframe
df

#Split the data into independent 'X' and dependent 'Y' variables
    A.  = df.iloc[:, 1:df.shape[1]].values
    B.  = df.iloc[:, 0].values

# Split the dataset into 75% Training set and 25%
Testing setfromsklearn.model_selectionimport
train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.25,
random_state = 0)
```

#Use Random Forest Classification algorithm**fromsklearn.ensembleimpo**
rt RandomForestClassifier
forest = RandomForestClassifier(n_estimators = 10, criterion = 'entropy',
random_state = 0) forest.fit(X_train, Y_train)

#Get the accuracy on the training data
forest.score(X_train, Y_train)
#Show the confusion matrix and accuracy for the model on the test data
#Classification accuracy is the ratio of correct predictions to total predictions made.
fromsklearn.metricsimport
confusion_matrix cm =
confusion_matrix(Y_test,
forest.predict(X_test))
TN = cm[0][0]
TP = cm[1][1]
print('Model Testing Accuracy = "**{}**!"'.format((TP + TN) / (TP + TN +
FN + FP))) print()***# Print a new line***

importances = pd.DataFrame({'feature':df.iloc[:,
1:df.shape[1]].columns,'importance':np.round(forest.feature_importances_,3)})
#Note: The target column is at position 0
importances =
importances.sort_values('importance',ascending=**False**).set_index('feature')
importances

#Visualize
importances.plot.bar()

CHAPTER 7
RESULTS & DISCUSSION

7.1 PERFORMANCE ANALYSIS REPORT

7.1.1 EXISTING SYSTEM:

Deep learning algorithm has shown superiority over other machine learning algorithms in the prediction problem. Despite the imbalanced dataset, as reported in experiment 1, the system has shown high prediction accuracy over all other state-of-the-art methodologies. This is not only caused by using deep learning, but also due to proper adoption of preprocessing and selecting only the effective features. A synthetic balanced version of the dataset has been used and compared the same settings . Notice that the accuracy of KNN (K = 1) method , still better than our work due to the overfitting, as they admitted. Recall that measuring the prediction model using only train-test sets is not always fair. Therefore, cross-validation is used in experiment 3 to obtain a more realistic measurement. Since only the work has conducted cross-validation, we compared our work against it. The result shows that the accuracy of our model is much better than Linear SVM and KNN models..

7.1.2 PROPOSED SYSTEM:

The dataset is a good representative of the general workforce in today's organizations. The result from the random forest classifier justify that the

features chosen are causes that contribute to voluntary attrition.Intuitively, data points that are close to each other are likely to have the same outcome of attrition. This is the basis for choosing random forest algorithm. The random forest classifier has good accuracy values, the prediction accuracy using the original dataset is about 91%, whereas it is about 94% using a synthetic dataset. Instead of constructing a general model, it simply stores instances of the data and classifies by a majority vote of the classes.

CHAPTER 8
CONCLUSION AND FUTURE ENHANCEMENT

8.1 CONCLUSION

Employee attrition effects in financial, time and effort loss of organizations.
It is a big issue since a trained and experienced employee is difficult to substitute and its cost effective. We try to find to analyze the past and existing employees information to estimate the future attritioners and study the reasons of employee turnover. The results of this learning describe that data extraction algorithms can be utilized to construct reliable and accurate predictive methods for employee attrition. The issue of employee attrition identification is not just to depict attritioners from no attritioners .
By using the tentative data study and data extraction methods, we can depict the attrition probability for each one employee and provide them score to build the retention techniques.

8.2 **FUTURE ENHANCEMENT**

Predicting the reason that cause the employee to be attrite using psychological factors before attrition is highly effective than predicting the reason for the

employee’s attrition after attrition.

So in future, the project works on collecting employee’s psychological factors that may help the company to decide whether to retain the employee or not .

DEMO LINK:

https://drive.google.com/file/d/17IQvDyl91fzviBSUYViu94wPOPKae8Gz/view?usp=share_link

GITHUB LINK:

https://github.com/IBM-EPBL/IBM-Project-3937-1658672671