

1.EMPLOYEE ATTRITION PREDICTION

1.1 INTRODUCTION

Artificial Intelligence is the simulation of human intelligence processes by machine, especially computer systems. These processes include learning, reasoning and self-correction. Particular application of AI includes expert system, speech recognition and machine vision. Popular AI cloud offering includes Amazon AI services, IBM Watson Assistant, Microsoft cognitive services and Google AI services.

Examples of AI technology:

1. Machine Learning
2. Natural language processing (NLP)
3. Robotic
4. Self-driving cars

Machine Learning

Machine is the part of Artificial intelligence that allows computer system to learn from examples, data and experience. Through enabling computers to perform specific tasks intelligently, machine learning systems can carry out complex processes by learning from data, rather than following pre-programmed rules. It includes algorithmic interpretability, robustness, privacy, fairness, inference of causality, human-machine interaction, and security

Python for Machine Learning

Machine learning focuses on the development of Computer Programs that can change when exposed to new data. In this Project we'll implement a simple machine learning algorithm using python. Python is an interpreted, object- oriented, high level programming language with dynamic semantics. Python is widely used for data analytics. Almost all cloud platforms offer support and often new features becomes available in Python first.

Anaconda

Directly from the platform and without involving DevOps, data scientists can develop and deploy AI and machine learning models rapidly into production. Anaconda provides the tools needed to easily:

- Collect data from files, databases, and data lakes
- Manage environments with Conda
- Share, collaborate on, and reproduce projects

1.2 OBJECTIVES OF RESEARCH:

This project is done on Human Resource Management topic in the area of “Attrition Management”.The most challenging job for any manager is to retain their employees.This report seeks to utilise primary research in order to offer a greater understating of the complex issues raised by staff attrition.Specific aims were to identify trends in leaver’s behaviour and the reasons employee change jobs/organisation and identify the employers “perspective

on employees” reasons for leaving . Among all employee related problems, employee attrition is one of the key problem in the today's scenario despite the changes in the external environment. Attrition is said to be gradual reduction in number of employees through resignation, death and retirement. The other name given for Attrition is attrition. When a well-trained and well-adapted employee leaves the organization for any of the reason, it creates an empty space in an organization that is there occurs a vacuum in the organization. It creates a great difficulty for a Human resource personnel to fill the gap that has occurred. Modern Human resource managers is taking various steps to reduce the employee attrition rate and it has been a pivotal challenge for today's Managers. Many of the employees may also tend to leave the job for various undisclosed factors such as lack of job security, lack of career advancement, desire for change in new opportunities, anticipating higher pay, problems with supervisors and few other personal reasons. This study helps in knowing why attrition occurs, reasons for employee attrition, challenges faced by managers in retaining employees and also suggest some measures in retaining employees.

1.3 PROBLEM STATEMENT

Employee turnover is one of major problems faced by organizations and often looked at as an opportunity to cut costs associated with it. This paper will showcase the application of data mining methods to predict employee attrition and to determine key factors that might contribute in attrition. Four different classification models are presented in this paper. Models performance is evaluated and compared to determine the best classification model.

2. REVIEW OF LITERATURE

Employee attrition refers to the gradual loss of employees over time. Most literature on employee attrition categorizes it as either voluntary or involuntary. Involuntary attrition is thought of as the mistake of the employee, and refers to the organization firing the employee for various reasons. Voluntary attrition is when the employee leaves the organization by his own will. This paper focuses on attrition of an employee. A meta-analytic review of voluntary attrition found that the strongest predictors of voluntary attrition included age, pay, and job satisfaction. Other studies showed that several other features, such as working conditions, job satisfaction, and growth potential also contributed to voluntary attrition . Organizations try to prevent employee attrition by using machine learning algorithms to predict the risk of an employee leaving, and then take pro-active steps for preventing such an incident.

(Ho et al., 2010) emphasized the relevance of push and pull factors while understanding employee attrition in the Malaysian service industry. Work stress, improper work-life balance, and poor relations with co-workers are some of the factors which comes under the umbrella of push factors and promotion opportunities, better compensations, desire for higher studies, and interesting work are some of the pull factors.

(Deepa and Stella, 2012), in their exploratory study on “Employee turnover in the IT industry with special reference to Chennai city”, highlighted a number of factors which contribute to employee turnover. Demographics, Organization’s performance, Organization’s culture (in terms of its reward system, leadership, shared goals etc.), Employees personal traits (like desire to learn, change in personal life, new job offer etc.), Job characteristics, Unrealistic expectations are the factors that are perceived to enhance employee turnover in the IT industry.

(Joy P. and Radhakrishnan 2012) in their study on marketing executives of financial product companies identified six factors which affects attrition. 58.77 percent of variance was jointly explained by all the six factors. Work specific attrition, human resource policy of the organization, boss behaviour with sub-ordinates, prevailing economic forces, fellow employee influence and opportunities in the society are the six factors identified in this study. A positive relationship among all the factors was supported by the correlation matrix.

(Zahra, et al., 2013) have highlighted the significance of commercialization of education which leads to faculty turnover. As the number of Universities are increasing there is a paradigm shift which is increasing the staff turnover because of abundant opportunities.

(Vinit et al., 2013) highlighted that factors like appraisal, openness, training, and flexibility act as key influencers for employee turnover. Thus these are the parameters on which an employee thinks before deciding to leave an organization.

(Saleem and Affandi, 2014) investigated the impact which Human Resources practices have on the employee attrition. Fairness of rewards and growth opportunities were considered to be the key factors influencing the employees’ decision to leave an organization. III.

3.DATA COLLECTION

We used the data from the website “kaggle.com”. We have data about past and current employees in a spreadsheet on desktop. It has various data points on our employees. but most interested in whether they’re still with company or whether they’ve gone to work somewhere else. And we want to understand how this relates to workforce attrition.

3.1 Primary Data Collection Method

The primary data source is the most important aspect of any research as it is the initial point of starting the dissertation. Hence it requires to be genuine and must provide accurate data with evidence. In this research, the primary data was collected by the study on various research papers and case studies on employee attrition, predicting employee attrition and Factors influencing retention. The weighting for each

factor influencing the attrition was given after the survey was carried among the HR Managers and Professionals. The detailed information about the survey is given in “kaggle.com”.

3.2 Secondary Data Collection Method

Secondary data was used in my research for implementing the research technically. Thus, whatever was studied theoretically, needed to be implemented practically wherein I referred to Technical thesis and other research Journals for transforming business logic to technical implementation. Here, I studied about the Machine Learning Algorithms and how they can be implemented for predicting employee attrition and implementing decision tree for categorizing the valuable employee from ordinary one. I also studied about the previous implementations done by IBM Watson for workforce analytics and Talent Management tool (IBM Watson, 2018) and how I can improvise those tools and make considerable advancements in those.

4. METHODOLOGY

4.1 Exploratory Data Analysis

4.1.1 Figures

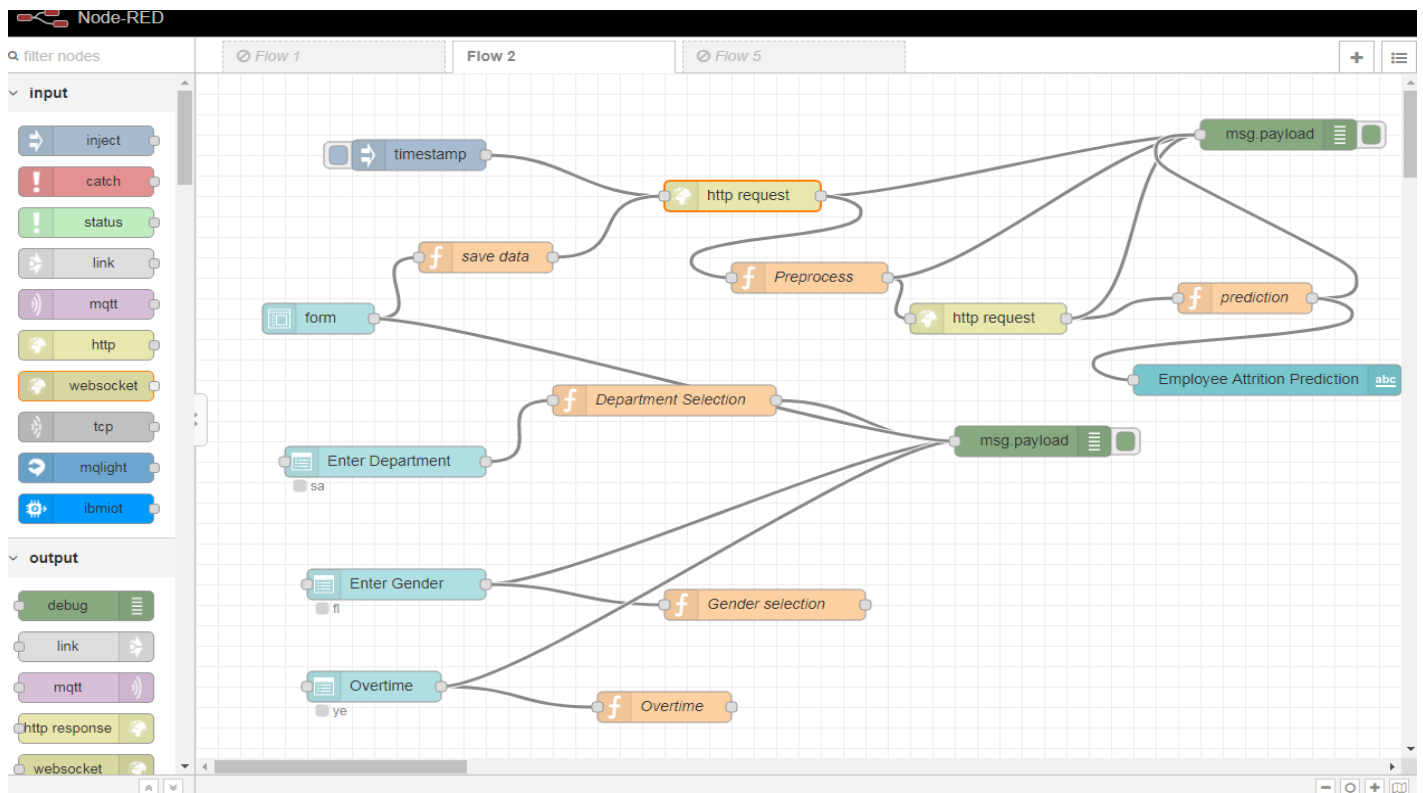


Fig 4.1.1 Node-RED flow

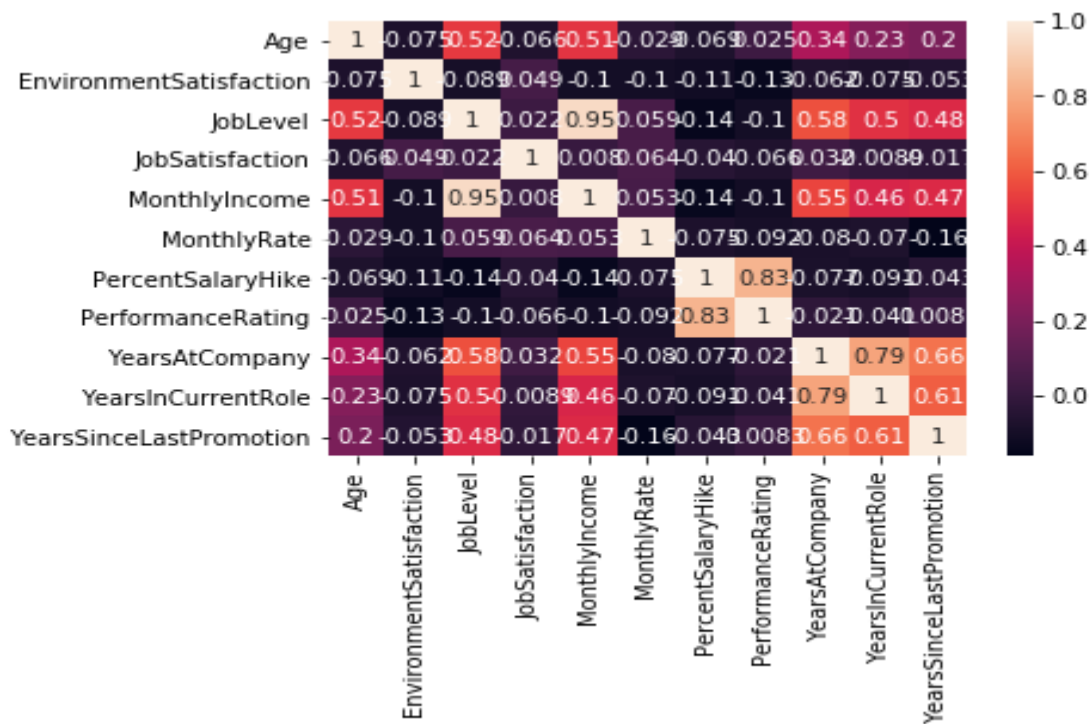


Fig 4.1.2 Corelation

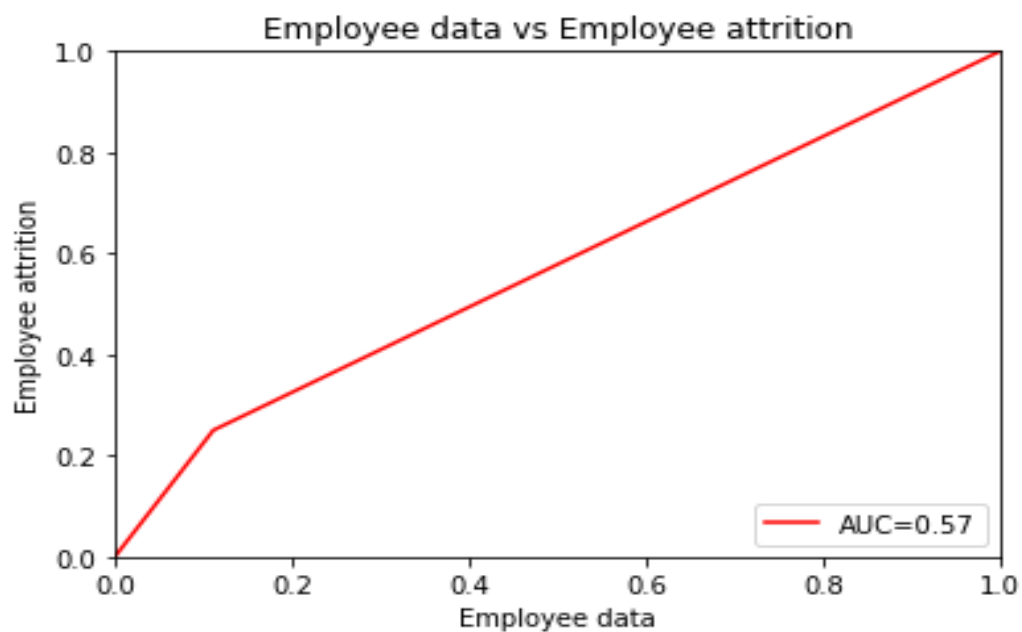


Fig 4.1.3 roc_auc

Enter Age *

Enter EnvironmentSatisfaction *

Enter JobLevel *

EnterJobSatisfaction *

Enter MonthlyIncome *

Enter MonthlyRate *

Enter PercentSalaryHike *

Enter PerformanceRating *

Enter YearsAtCompany *

Enter YearsInCurrentRole *

Enter YearsSinceLastPromotion *

SUBMIT

CANCEL

Employee Attrition
Prediction

**leaving the
company**

Fig 4.1.4 User Interface

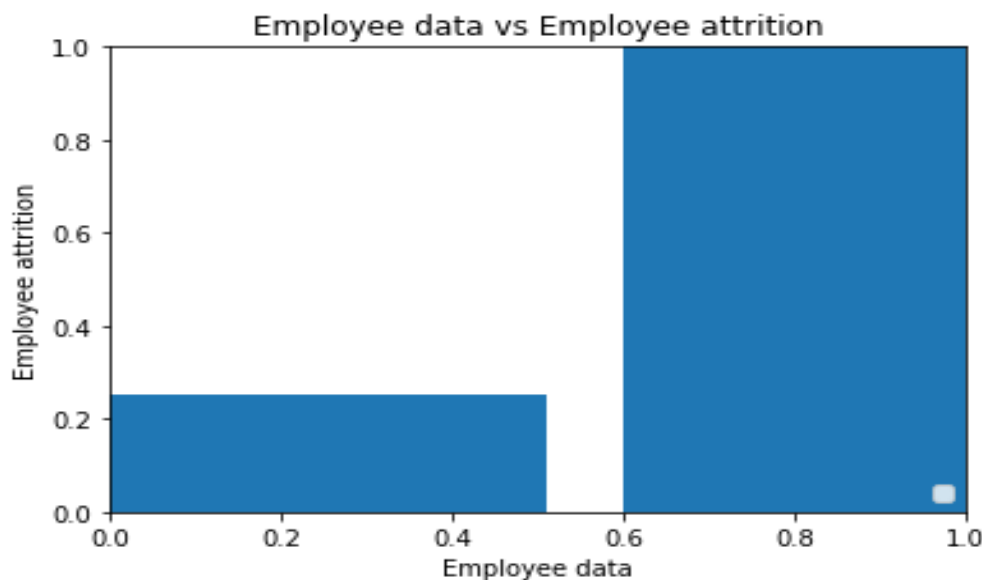


Fig 4.1.5 Bar chart

4.2 Data Modelling

We have our final dataset. If you won't have the true values how would you know that the predictions are correct. Now you will realize that, how important the training data phase is. We train the model in a way that it can predict (almost) correct results. In this dataset, we don't have any missing values for Attrition, we will split the data into train and test. We will train the model on training data and predict the results on test data. For this particular exercise we will use k-nearest neighbors (KNN). Before jumping into code, let's get a little background about the RF classifier. The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems .

Implementation

Import libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import types
import pandas as pd
from botocore.client import Config
import ibm_boto3
def __iter__(self): return 0
# @hidden_cell
```

```
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share your notebook.
client_3032b2439ebd429882097d9e71584e90 = ibm_boto3.client(service_name='s3',
    ibm_api_key_id='0L9JKR0MK73oxI6LktSKykXyQupRDp069_Ls4KwY16ur',
    ibm_auth_endpoint="https://iam.bluemix.net/oidc/token",
    config=Config(signature_version='oauth'),
    endpoint_url='https://s3.eu-geo.objectstorage.service.networklayer.com')
body=client_3032b2439ebd429882097d9e71584e90.get_object(Bucket='employeeattritionprediction-
donotdelete-pr-m3c6rd3rqpexwo',Key='employee_attrition.csv')['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType( __iter__, body )
dataset = pd.read_csv(body)
dataset.head()
dataset.isnull().any()
x=dataset.iloc[:,0:14].values
y=dataset.iloc[:,14:].values
```

Label Encoding

```
from sklearn.preprocessing import LabelEncoder
lb=LabelEncoder()
x[:,1]=lb.fit_transform(x[:,1])
lb1=LabelEncoder()
x[:,3]=lb1.fit_transform(x[:,3])
lb2=LabelEncoder()
x[:,8]=lb2.fit_transform(x[:,8])
lb1.classes_
y[:,0]=lb.fit_transform(y[:,0])
```

Splitting data

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
y_train=y_train.astype('int')
from sklearn.pipeline import Pipeline
pipeline=Pipeline([('scalar',sc),('kn',knn)])
model=pipeline.fit(x_train,y_train)
```

KNN

```
from sklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier(n_neighbors=5,p=2)
```



```

knn.fit(x_train,y_train)
y_pred=knn.predict(x_test)
y_pred
y_test
y_test=y_test.astype('int')
from sklearn.metrics import accuracy_score
print("Accuracy score:",accuracy_score(y_test,y_pred)*100,"%")
from sklearn.metrics import confusion_matrix
pd.DataFrame(confusion_matrix(y_test,y_pred))
import sklearn.metrics as metrics
fpr,tpr,threshold=metrics.roc_curve(y_test,y_pred)
roc_auc=metrics.auc(fpr,tpr)
roc_auc
plt.plot(fpr,tpr,'b',label='AUC=%0.2f' % roc_auc,color="red")
plt.legend(loc='lower right')
plt.xlim([0,1])
plt.ylim([0,1])
plt.xlabel("Employee data")
plt.ylabel("Employee attrition")
plt.title("Employee data vs Employee attrition")
plt.show()
!pip install watson-machine-learning-client --upgrade
from watson_machine_learning_client import WatsonMachineLearningAPIClient
wml_credentials={"url":"https://eu-gb.ml.cloud.ibm.com",
                 "access_key":"814edUwidBPao0FzUbjtNRYGqaTsRu042w19Y8JLTCoi",
                 "username":"88632294-f105-416c-8803-b3c5a1e3c1b9",
                 "password":"abb8f456-133f-4bca-91e1-01eed99c6952",
                 "instance_id":"de1c3184-06fa-4d87-9d99-2cc85c30b5b1"}
client=WatsonMachineLearningAPIClient(wml_credentials)
model_props={client.repository.ModelMetaNames.AUTHOR_NAME:"Durga",
client.repository.ModelMetaNames.AUTHOR_EMAIL:"durgalakshmi2016@gmail.com",
client.repository.ModelMetaNames.NAME:"employee attrition prediction"}
model_s=client.repository.store_model(model,meta_props=model_props)
client.repository.list()
published_model_uid=client.repository.get_model_uid(model_s)
published_model_uid

```

```
d=client.deployments.create(published_model_uid,name="employee attrition prediction")  
scoring_endpoint=client.deployments.get_scoring_url(d)  
scoring_endpoint
```

5. REFERENCES

- [1]Kaggle.com.(n.d.).Retrievedfrom<https://www.kaggle.com/pavansubhasht/ibm-hr-analyticsattrition-dataset>
- [2] Dancho, M.(2017) HR Analytics using Machine Learning.
- [3] Kuhn,M.(2017).The caret Package. Retrieved from Github: <http://topepo.github.io/caret/index.html>
- [4] Max Kuhn, K. J. (n.d.). Applied Predictive Modeling. Springer.
- [5] Neff, D. L. (1993). Predicting and managing turnover in human service agencies: a case study of an organization in crisis.
- [6] Wiener, A. L. (2002). Classification and Regression by randomForest.
- [7] Bhalla, D. (n.d.). Retrieved from Listen Data: <http://www.listendata.com/2015/07/gbm-boosted-models-tuningparameters.html>
- [8] Brownlee, J. (2016). Retrieved from Machinelearningmastery: <https://machinelearningmastery.com/tune-machine-learningalgorithms-in-r/>

6. CONCLUSION

Employee turnover has huge implications n organizations and is a non-value add cost. Predictive modeling can help HR departments predict employee attrition. This paper analyzesfour classification models to predict employee attrition based on 35 distinct employee attributes. . The KNN classifier has good ROC-AUC and accuracy values. Instead of constructing a general model, it simply stores instances of the data and classifies by a majority vote of the classes of the nearest neighbors.With ROC as the key performance metric, the k-nearest neighbors (KNN) algorithm was the best model . The Random Forest model was the worst performing model while Logistic Regression provided consistent yet average results. The attribute importance plots for each model point out the key factors involved in employee attrition. The classification models discussed in this paper if scaled and tailored can be applied in real life. Other more advanced models like neural networks and adaptive boosting can be applied to this problem.