J.Tejaswi 39110395 EmployeePromotionPrediction

**Table of Contents**

**INTRODUCTION**

OVERVIEW

PURPOSE

**LITERATURE ANALYSIS**

EXISTING PROBLEM

PROPOSED SOLUTION

**THEORITICAL ANALYSIS**

BLOCK DIAGRAM

HARDWARE/SOFTWARE DESIGNING

**EXPERIMENTALINVESTIGATION**

**FLOWCHART**

**RESULT**

**ADVANTAGES & DISADVANTAGES**

**APPLICATIONS**

**CONCLUSION**

**REFERENCES**

**1.INTRODUCTION**

**OVERVIEW** : The problem of employee turnover has shot to prominence in organizations because of its negative impacts on issues ranging from work place morale and productivity, to disruptions in project continuity and to long term growth strategies. One way organizations deal with this problem is by predicting the risk of attrition of employees using machine learning techniques thus giving organizations leaders and Human Resources (HR) the foresight to take pro-active action for retention or plan for succession. However, the machine learning techniques historically used to solve this problem fail to account for the noise in the data in most HR Information Systems (HRIS). Most organizations have not prioritized investments in efficient HRIS solutions that would capture an employee’s data during his/her tenure. One of the major factors is the limited understanding of benefits and cost. It is still difficult to measure the return of investment in HRIS . This leads to noise in the data, which in turn attenuates the generalization capability of these algorithms PURPOSE: Employee Promotion plays a big role in Employee Satisfaction. It aids in employee engagement, boosts morale, reduces absenteeism, and ultimately in productivity. Employee Promotion is also a helpful tool for reducing attrition retention •

**PURPOSE** :Employee Promotion plays a big role in Employee Satisfaction. It aids in employee engagement,boosts morale,reduces absenteeism, and ultimately in productivity.

Employee Promotion is also a helpful tool for reducing attrition retention. Attrition has been a dominant problem in all companies. By practicing proper performance appraisals and employee promotion, this problem can be tackled.

**2. LITERATURE SURVEY**

**EXISTING PROBLEM**: To overcome this problem, organizations are now taking support via machine learning techniques to predict the employee turnover. With high precision in prediction, organizations can take necessary actions at due course of time for retention or succession of employees. Most of the data comes from basic HR based database systems, which are not highly efficient in prediction and modeling and these models are not very accurate in data models and cannot assist the organizations to take successful decisions.

**PROPOSED SOLUTION :** We can train our data on different algorithms. For this project weare applying four classification algorithms. The best model is saved based on its performance. To evaluate the performance confusion matrix and classification report is used.

Decision Tree Model: A function named decisionTree is created and train and test data are passed as the parameters. Inside the function, DecisionTreeClassifier algorithm is initialized and training data is passed to the model with .fit() function. Test data is predicted with .predict() function and saved in new variable. For evaluating the model, confusion matrix and classification report is done.

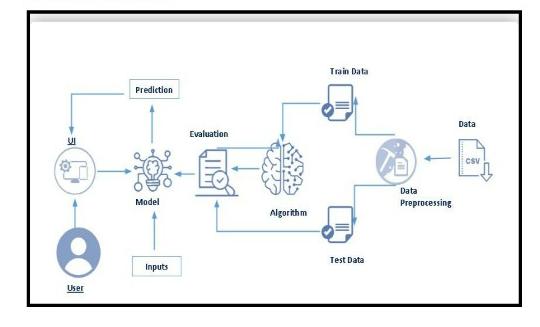
### **Random Forest Model****:** A function named randomForest is created and train and test data are passed as the parameters. Inside the function, RandomForestClassifier algorithm is initialized and training data is passed to the model with .fit() function. Test data is predicted with .predict() function and saved in new variable. For evaluating the model, confusion matrix and classification report is done.

### KNN Model: A function named KNN is created and train and test data are passed as the parameters. Inside the function, KNeighborsClassifier algorithm is initialized and training data is passed to the model with .fit() function. Test data is predicted with .predict() function and saved in new variable. For evaluating the model, confusion matrix and classification report is done.

Xgboost Model: function named xgboost is created and train and test data are passed as the parameters. Inside the function, GradientBoostingClassifier algorithm is initialized and training data is passed to the model with .fit() function. Test data is predicted with .predict() function and saved in new variable. For evaluating the model, confusion matrix and classification report is done.

**3.THEORITICAL ANALYSIS**

**Block Diagram:**



**Hardware Requirement: System :**

Pentium i3 Processor.

Hard Disk : 500 GB

Monitor : 15’’ LED

Input Devices : Keyboard, Mouse Ram : 2 GB

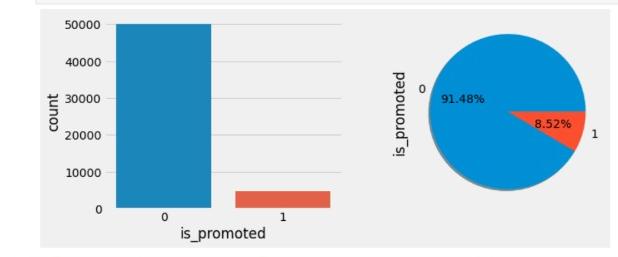
**Software Requirement :**

Operating system : Windows 10.

Coding Language : Python

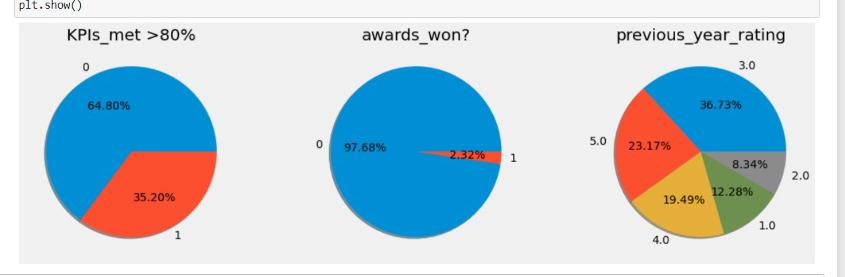
**4.EXPERIMENTAL INVESTIGATION**

**Univariate Analysis**

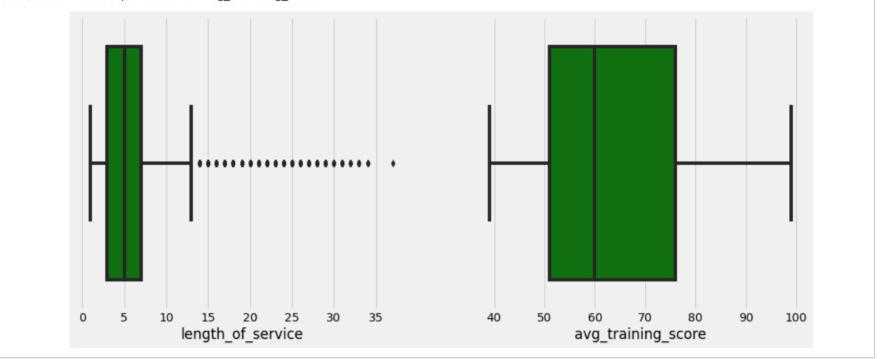


univariate analysis is understanding the data with single feature. Here we have displayed two different graphs such as pie plot, box plot and count plot.

* Count plot and pie plot are used on the target variable. From the below image, we identified our data is imbalanced. 91% of the employees are not promoted. To get better model performance, imbalanced data should be converted to balanced data. Handling imbalanced data will be discussed on data pre processing.



* A pie plot is used on value counts() of the required features. From the below graph, we get a clear understanding that 97.68% of employees have not won any awards. Around 65% of employees have KPIs > 80%. More than 75% of employees have a previous year rating > 3.0. Instead of pie plot count plot can also be used.



* Box plot is used on the length of service and average training score feature. Length of services feature has more outliers. The model should not be built without handling the outliers. Here, outliers are handled by the capping method. Capping will be discussed on data pre-processing.

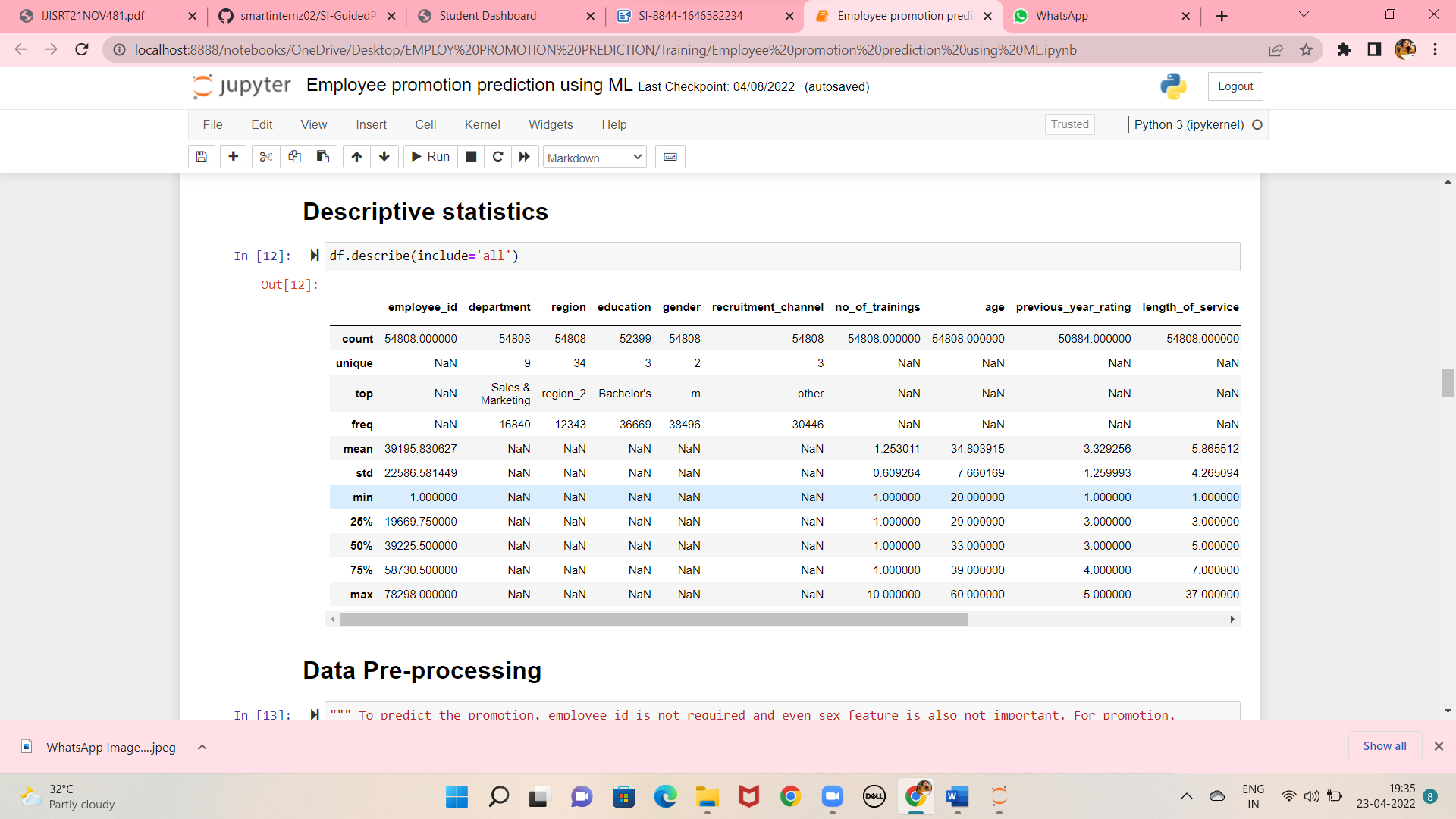
**Multivariate Analaysis**

multivariate analysis is to find the relation between multiple features. Here we have used barplot from seaborn package.

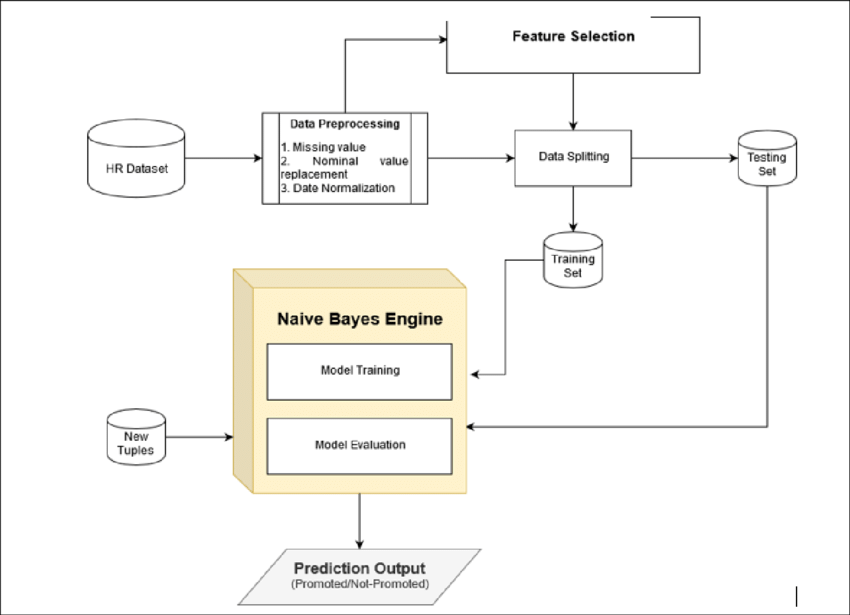
* Three features are passed as parameters for barplot(). A clear pattern is understandable from the below plot. Employees with an average training score greater than 95 and a previous year rating greater than 3 got promotions (100%).



### **Descriptive Analysis** **:** Descriptive analysis is to study the basic features of data with the statistical process. Here pandas has a worthy function called describe. With this describe function we can understand the unique, top and frequent values of categorical features. And we can find mean, std, min, max and percentile values of continuous features.



**5.FLOW CHART:**

flow chart of the employee promotion prediction

**6.RESULT:**

For this project,the dataset we will be working with is Employee promotion

prediction data set which consists of 54808 employee information and 14

attributes .

The 14 attributes are:

employee\_id

departement

region

education

gender

recuirement\_channel

no\_of\_trainings

**7.Advantages**

The more you promote your brand, the more will the customers know about you and your company and the more will they be interested in your products. Promotion can be done even by giving out free samples which work wonders for customers

**Disadavantages**

No new or fresh ideas are brought into the organisation. The job advertised may require skills not currently available within the organisation. Promotion of an internal employee could cause resentment amongst other employees, who may feel they deserve the post more than the promoted employee.

Testing Model

In this module we test the trained machine learning model using the test dataset Quality assurance is required to make sure that the software system works according to the requirements. Were all the features implemented as agreed? Does the program behave as expected? All the parameters that you test the program against should be stated in the technical specification document. Moreover, software testing has the power to point out all the defects and flaws during development. You don’t want your clients to encounter bugs after the software is released and come to you waving their fists. Different kinds of testing allow us to catch bugs that are visible only during runtime.

increasing customer traffic.

The more you promote your brand, the more will

the customers know about you and your company and the

more will they be interested in your products.

Promotion can be done even by giving out free samples

which work wonders for customers.

**TESTING MODEL**

In this module we test the trained machine learning model using the test dataset

Quality assurance is required to ensure that the software system works accordingto the requirements. Were all the features implemented as agreed? Does the

program behave as expected? All the parameters that you test the program

against should be stated in the technical specification document.

Moreover, software testing has the power to point out all the defects and flaws

during development. You don’t want your clients to encounter bugs after the

software is released and come to you waving their fists. Kinds of testing allow us

to catch bugs that are visible only during runtime.

increasing customer traffic.

The more you promote your brand, the more will the customers know about you

and your company and the more will they be interested in your products.

Promotion can be done even by giving out free samples which work wonders for

customers

Predictive models are only as good as the data from which they are built, so good

data collection practices are crucial to developing high-performing models.

The data need to be error-free (garbage in, garbage out) and contain relevant

information for the task at hand. For example, a loan default model would not

benefit from tiger population sizes but could benefit from gas prices over time.

In this module, we collect the employee promotion prediction data from kaggle

dataset archives. This dataset contains the information of divorce in previous

years.

**8.APPLICATIONS:**

Employee promotion is is important for employees who work in a

company.Generally, based on the performance promotion is given. No

organizations will promote their employees by gender, region, and recruitment

channel.

**9.CONCLUSION:**

Our application is able to predict the promotion of a employee.The importance of

predicting employee promotion in organizations and the application of machine

learning in building models was presented in this paper.

The key challenge of noise in the data from HRIS that compromises the accuracy

of these predictive models was also highlighted. Data from the HRIS of a global

retailer was used to compare the XGBoost classifier against, six other supervised

class that had been historically used to build turnover models.

The results of this research demonstrate that the XGBoost classifier is a superior

algorithm in terms of significantly higher accuracy, relatively low runtime and

efficient memory utilization for predicting turnover.

its regularization makes it a robust technique capable of handling the noise in the

data from HRIS, as compared to the other classifiers, thus overcoming the key

challenge in this domain.

Because of these reasons it is recommended to use XGBoost for accurately

predicting employee turnover, thus enabling to take actions for retention or

succession of employees.

**11.REFERENCES:**

[1] S. Jahan, “Human Resources Information System (HRIS): A Theoretical

Perspective”, Journal of Human Resource and Sustainability Studies, Vol.2 No.2,

Article ID:46129, 2014.

[2] M. Stoval and N. Bontis, “Voluntary turnover: Knowledge management –

Friend or foe?”, Journal of Intellectual Capital, 3(3), 303-322, 2002.

[3] J. L. Cotton and J. M. Tuttle, “Employee turnover: A meta-analysis and review

with implications for research”, Academy of management Review, 11(1), 55-70,

1986.

[4] L. M. Finkelstein, K. M. Ryanand E.B. King, “What do the young (old) people

imagine me? Content and accuracy of age-based meta stereotypes”, European

Journal of Work and Organizational Psychology, 22(6), 633-657, 2013.

[5] B. Holtom, T. Mitchell, T. Lee, and M. Eberly, “Turnover and retention research:

A glance at the past, a closer review of the present, and a venture into the future”,

Academy of Management Annals, 2: 231-274, 2008