**employees PERFORMANCE rating Using ML**

**Introduction:**

Software challenge managers compare actual final touch of activities in opposition to the development reviews filled by way of project contributors to identify substantial deviations from the estimated schedules and manage software program assignment risks. However, quantitative measurements are restricted due to the layout of mission documents, which are mostly herbal languages. The proposed approach can recognize entailment from textual content successfully and outperform other textual entailment approaches. The textual entailment approach to venture monitoring and manipulate, which not only reduces the assignment value and human’s effort however also provides a foundation for challenge managers to qualitatively evaluate the overall performance of every challenge member. The employee rating in the company is taken and analysed with different type of attributes. The employee ratings are provided based on the project reviews, client feedback and client satisfaction level about the project.

**Keywords:**

Decision tree classifier, Dataset, productivity score, Algorithms, pre-processing, Data cleaning.

.

**Motivation:**

Once configured, the software continuously monitors any data trails that a corporation can make available for each of its employees in the background. The system learns the normal workflow for different workers by using machine learning algorithms. The software assigns each employee a "productivity score" and it has determined their regular pattern of behaviour. Even if employees in a corporation perform various tasks, their ratings can still be compared. A productivity score also shows whether employee effort boosts or depletes productivity.

The project evaluation can be tested with the machine learning algorithm prediction results. Since the random forest algorithm will be used analysis , the accuracy of the algorithm result will be helpful to evaluate the results. The accuracy score of the algorithm in the project helps to evaluate the dataset.

**Significance:**

The main significance of the project is the employee rating productivity improvement will be the python based application which contributes to imporve the production of the software company. It will be helpful to the software company to imporve the skills of the employee based on the project requirements and also to improve the company productivity level.

**Objective:**

The objective of the project is to improve the productivity of the company by anlyzin the employee ratings. In this work , the dataset containing the employee rating will be taken into consideration. The pre-processing will be applied in to the dataset and the noisy and null value data will be removed from the dataset. After the data will be analyzed and visualized for further processing. The machine learning algorithm will be choosed to make the analysis.[1][2]

The application will be develoepd with Google Colab Python Tool as the project can be directly executed in any type computer systems with internet connection. There is no need of any specific softwares to be installed in the user system. The Colab Tool helps to develop and run the application directly inside the cloud server where the Python libray files are installed. The machine learning algorithm libraries are buil inside the Colab.

**Features:**

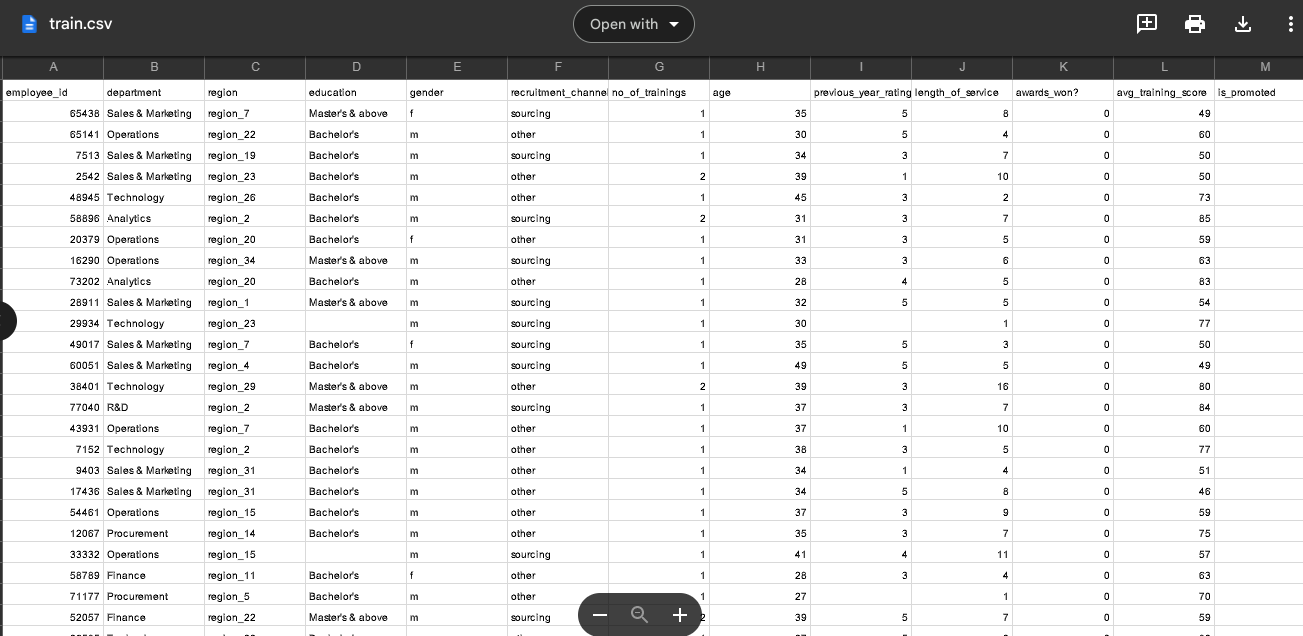
There are different advances engaged with building a machine learning project yet not every one of the means are compulsory to use in a solitary undertaking, and everything relies upon the information. In this work, will construct a machine learning on employee performance project and find out about the machine learning project lifecycle.[2][3] Despite the fact that it seems to be a straightforward task or simply fostering a model, the dataset we have is uproarious and needs loads of element designing, and preprocessing.

This project endeavors to foster a commonsense comprehension of the positive and negative employee experiences because of man-made consciousness reception and the production of techno stress. It disentangles the human resource improvement related difficulties with the performance in the software company.

**Dataset:**

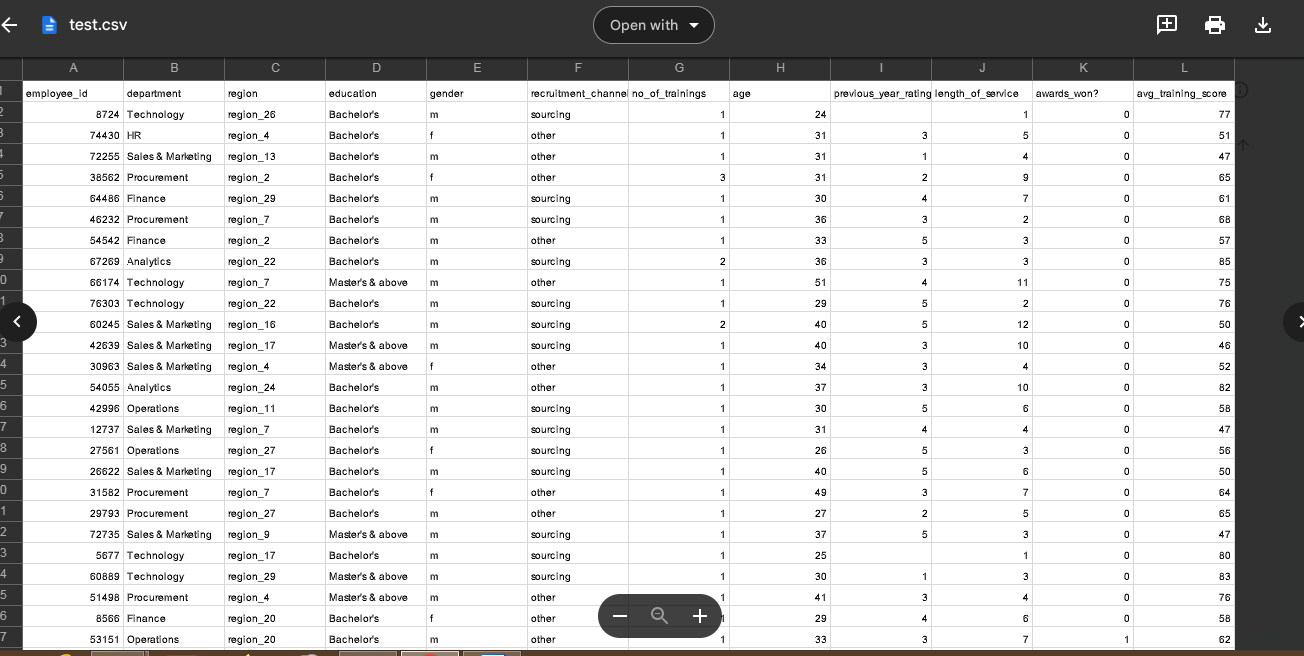
Most of the columns in a dataset are noisy and contain lots of information. But with feature engineering do, will get more good results. The first step is to import the libraries and load data. After that will take a basic understanding of data like its shape, sample, is there are any NULL values present in the dataset. Understanding the data is an important step for prediction or any machine learning project. It is good that there are no NULL values. And we need little changes in weight and Raw column to convert them to numeric by removing the unit written after value.

train.csv



The above dataset is the training set for the employee performance.

Test.csv



The testing dataset is used for the evaluation of the machine learning algorithm.

**Detailed Design of Features:**

This dataset contain the fields needed for the analyzing of the performance of the employee in the software company. Exploratory examination is a cycle to investigate and comprehend the information and information relationship in a total profundity with the goal that it makes highlight designing and demonstrating steps smooth and smoothed out for expectation.Exploratory examination assists with validating our presumptions or misleading.[5]

**Analysis of employee Performance :**

It will begin from the principal segment and investigate every section and comprehend what influence it makes on the objective segment. At the necessary step, we will likewise perform preprocessing and include designing undertakings. The point in acting top to bottom exploratory examination is to get ready and clean information for better machine learning demonstrating to accomplish elite execution and summed up models. So it should begin with breaking down and setting up the dataset for expectation.

**Modules:**

1) Dataset collection

2) Data cleaning

3) Exploratory Data Analysis

4) Machine learning Modeling

5) Report

1) Dataset collection:

The information about the employee rating with different types of attributes with employee performance data are collected from the software company . The feedback of client, project managers are taken to the dataset.

2) Data Cleaning:

The large dataset contains more noisy and improper data which have to be pre-processed to produce the quality dataset for further pruning. The data is cleaned and processed with initial stage of removing the null values.

3) Exploratory Data Analysis

Exploratory analysis is a process to explore and understand the data and data relationship in a complete depth so that it makes feature engineering and machine learning modeling steps smooth and streamlined for prediction. EDA involves Univarate, Bivariate, or Multivariate analysis. EDA helps to prove our assumptions true or false. In other words, it helps to perform hypothesis testing.

4) Machine learning Modeling

Machine learning modeling helps to find the best algorithm with the best hyper parameters to achieve maximum accuracy. The dataset is split into 2 variants. 70% of records are taken as training data and used to train the machine learning algorithm. The remaining 30% of dataset is applied to testing which helps to predict the process.

5) Report:

The Data is visualized based on the output of the machine learning algorithm and the data is mapped with different types of graphs to analyze and visualize the exact data to the user for the prediction of employee performnce. Marplot libraries are implemented to map the results based on the user requirements.

**Load Packages:**

First step have to import the necessary packages to the application. Importing all the Required Libraries:

\* Import Numpy, Pandas, Matplot, and Seaborn for Data Analysis and Visualizations

\* import ipywidgets, Sweetviz, ppscore for Exploratory Data Analysis

\* Import Sklearn, Imblearn for Machine Learning Modelling

# for mathematical operations

import numpy as np

# for dataframe operations

import pandas as pd

# for data visualizations

import seaborn as sns

import matplotlib.pyplot as plt

# for machine learning

import sklearn

import imblearn

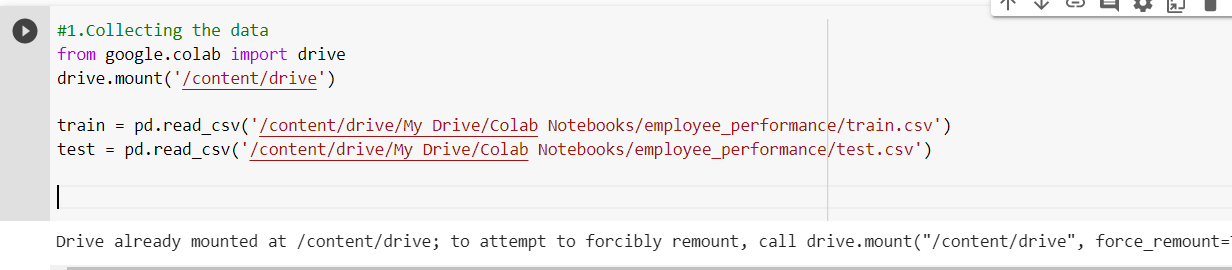
# setting up the size of the figures

plt.rcParams['figure.figsize'] = (16, 5)

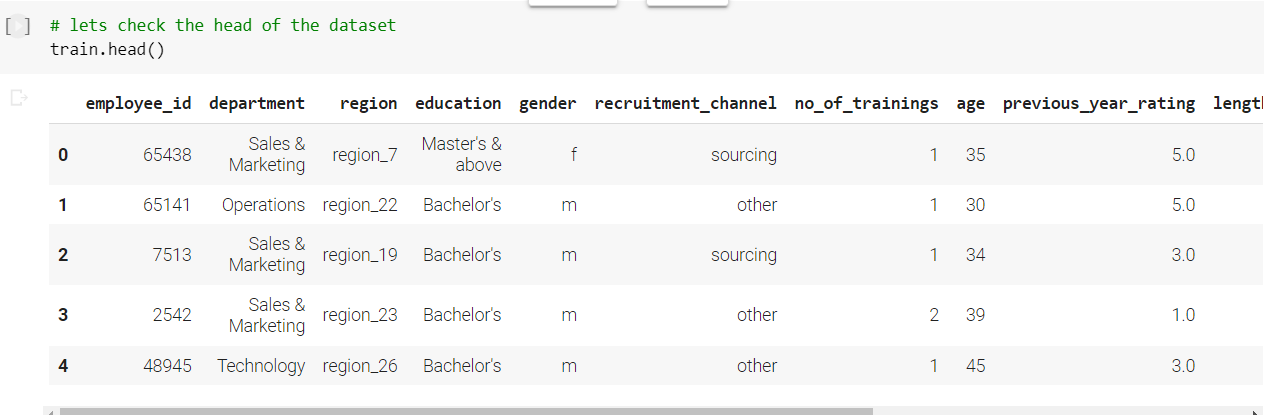
# setting up the style of the plot

plt.style.use('fivethirtyeight')

Next the dataset would be connected from the Google colab. Initially the dataset is uploaded into the Google colab folder. Then the python file should connect to the path from the Google colab folder.

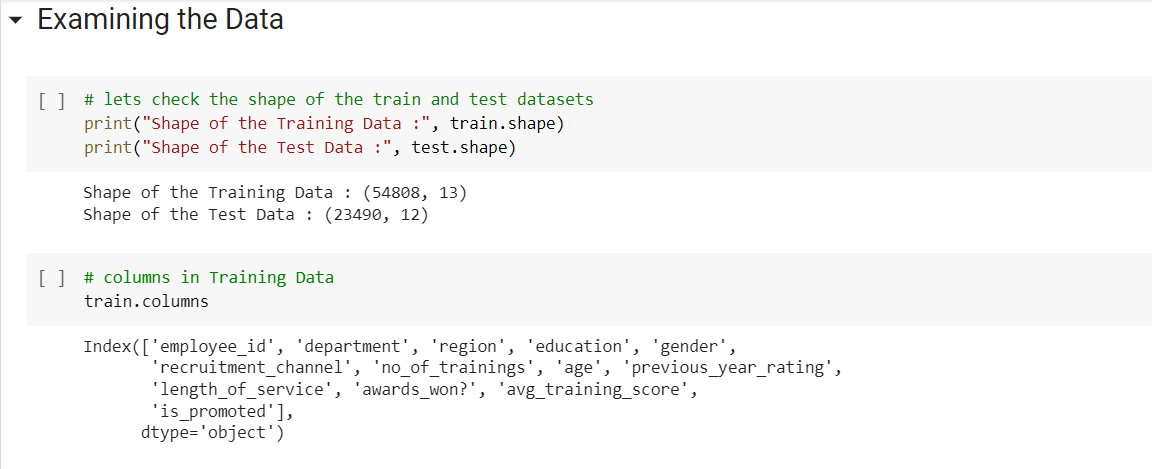


The information has an extremely straightforward design with elements. Each column is related with an interesting employee details.

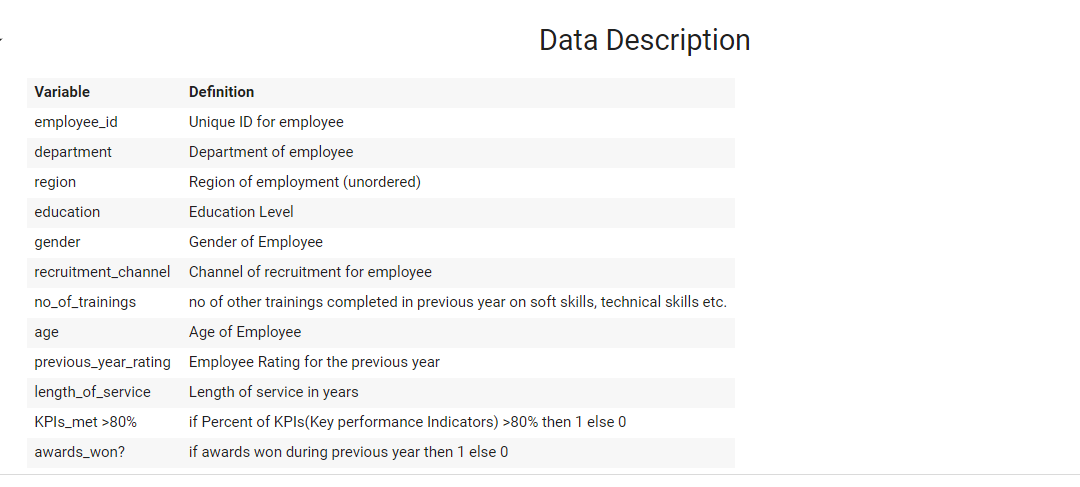


The dictionary shows the records displayed with head values of first 5 records from the dataset. The label has been set with different label description features of:

* Employee\_id
* Department
* Recruitment\_channel
* Previous\_year\_rating
* Avg\_training\_score
* Is\_promoted
* No\_of\_trainings
* age
* length of service
* awards\_won



A Dissipate plot is utilized when both the sections are mathematical and it responds to our inquiry in a superior manner.



The data description shows all the attributes of the employee dataset of both training and test data.

Descriptive Statistics:

* 1)Descriptive Statistics for the Numerical Columns for Numerical Columns we check for stats such as Max, Min, Mean, count, standard deviation, 25 percentile, 50 percentile, and 75 percentile.
* 2)Then check for the Descriptive Statistics for Categorical Columns for Categorical Columns we check for stats such as count, frequency, top, and unique elements.

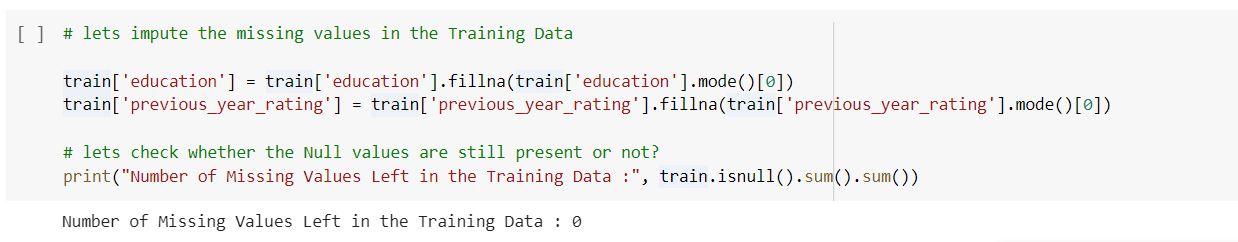
**Implementation:**

The dataset once loaded into the python colab, the initial pre-processing is done to remove the noisy data. Have to clean the dataset as this information may be fragmented and it can't be sent straightforwardly to the model. So will make a capability of cleaning which does the accompanying system to clean the information and returns the cleaned words:

a) Eliminate numbers, Alphanumeric words for example words which contain the two letter sets and numbers for example hello123 .

Pre-processing:

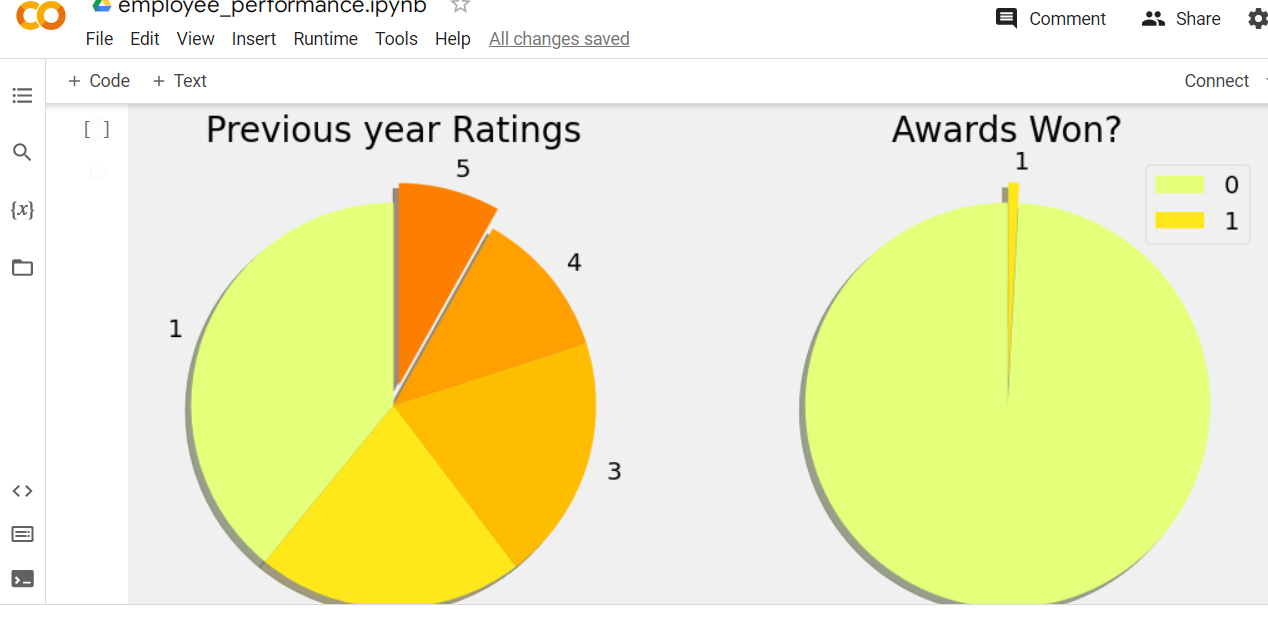
The noisy data, empty values in the cell are pre-processed. The columns which are not needed for the evaluation of the model also removed using th drop function in the python



* There are so many types of Missing Values such as
  + Missing values at Random
  + Missing values at not Random
  + Missing Values at Completely Random

Data Visualization:

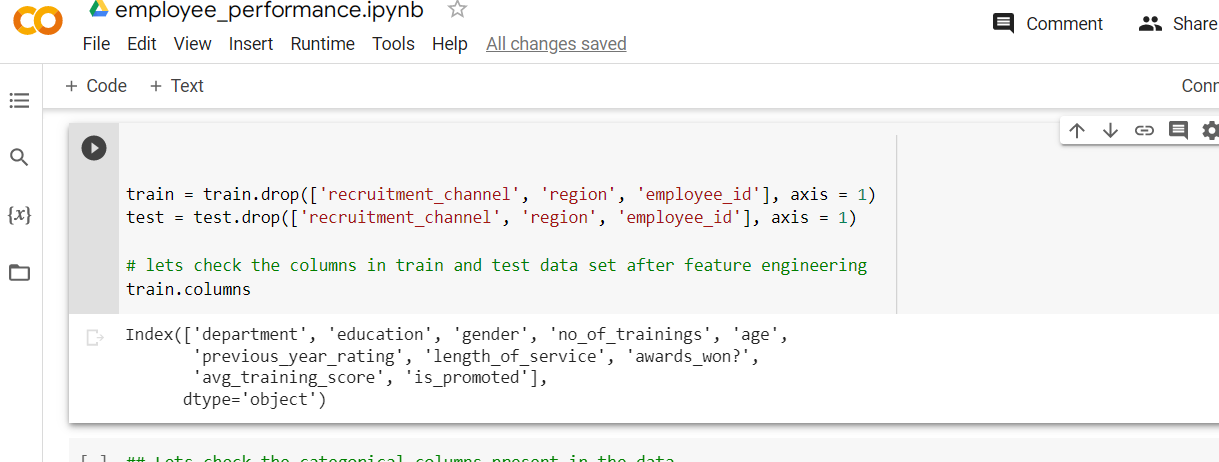
The employee deatils with there ratings.



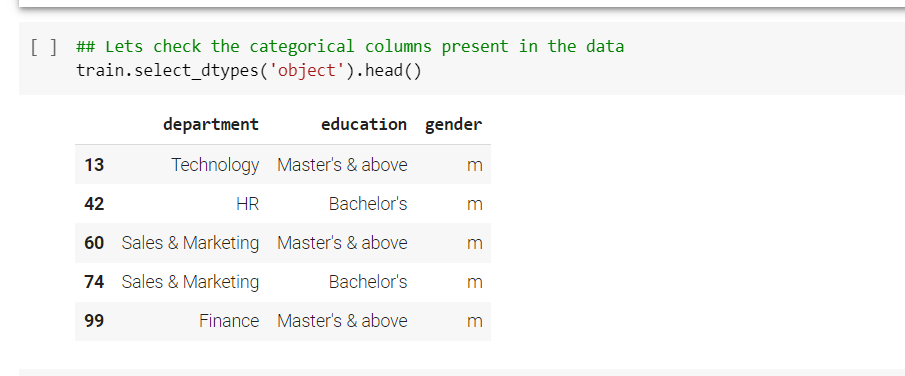
The previous year ratings and the awars won are taken into graphical report for the visualization.

**Implementation of Machine Learning models:**

The dataset is divided into training and testing data. The two type of dataset is passed into the different machine learning algorithm models and the accuracy levels were found

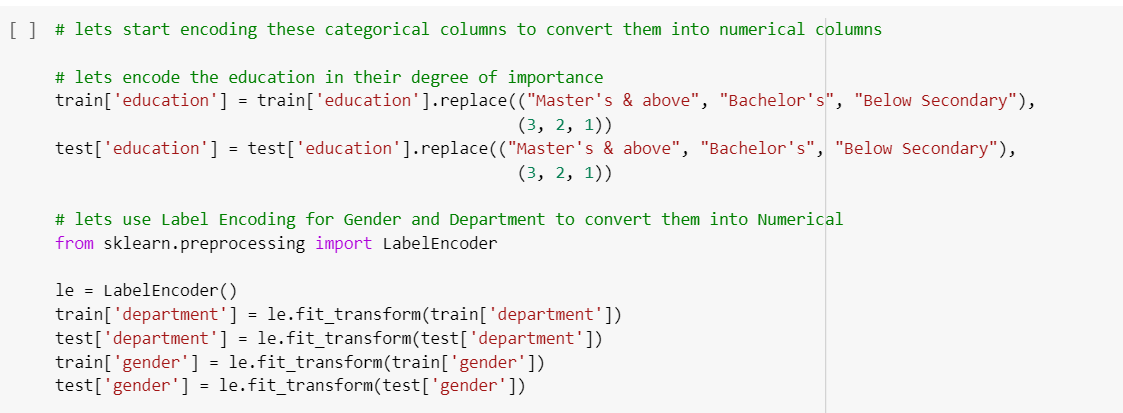


The dataset containing of test.csv data are given for testing and train.csv records are applied for the training.



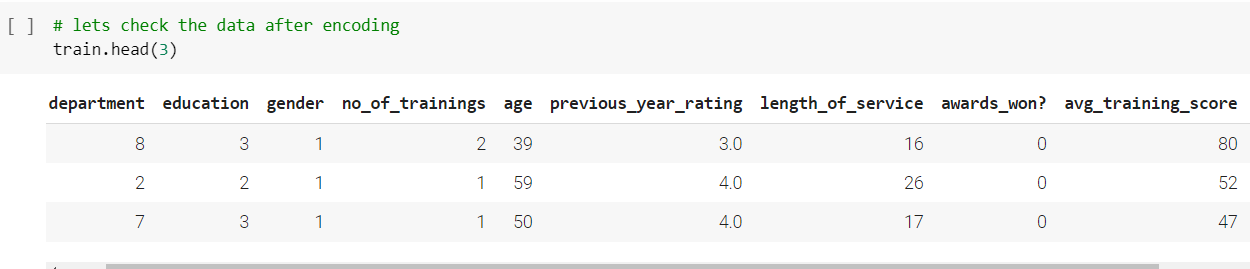
The relevant columns are displayed with the data type which is applied in to the application.

Encoding categorical column to numeric:

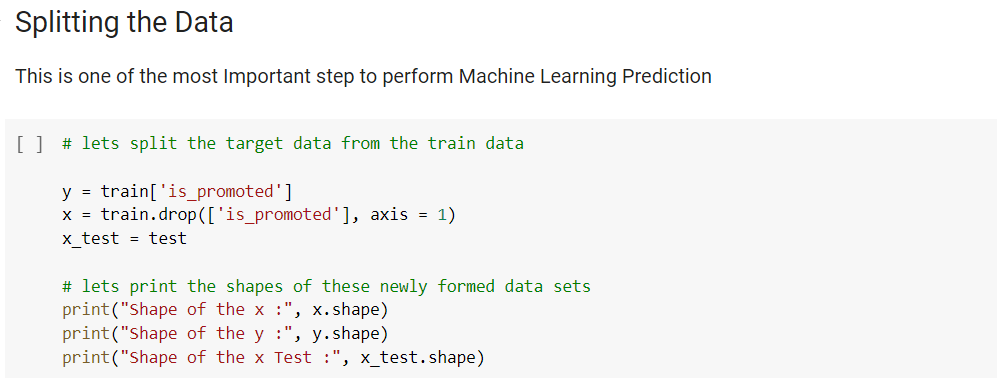


 The categorical columns  are converted to convert  into numerical columns.

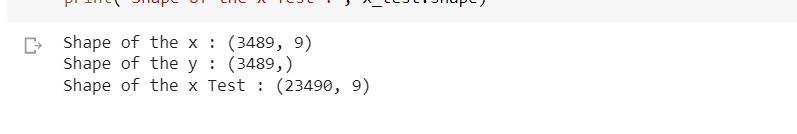
Results of theEncoding categorical column to numeric**:**



The category columns are converted into numeric columns.



Result of splittting data:



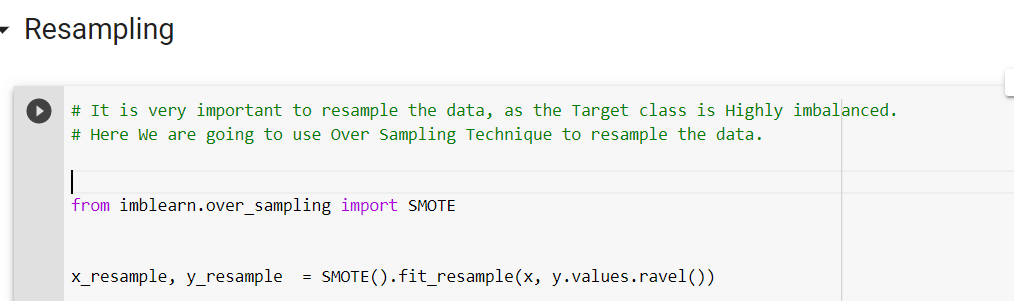
The x have 3489 records with 9 columns and the y have 3489 records.

**Model Evaluation:**

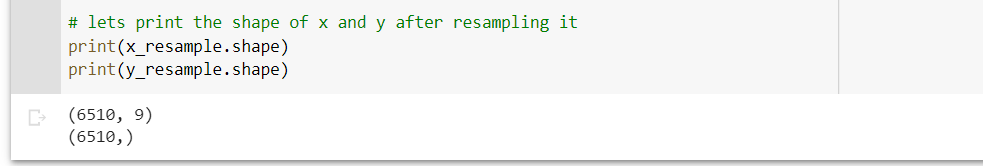
**Resampling**

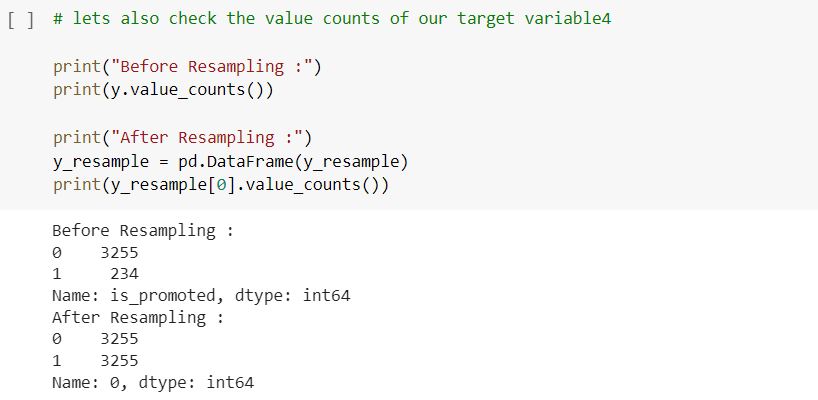
Resampling is the method that consists of drawing repeated samples from the original data samples. The method of Resampling is a nonparametric method of statistical inference.

The prediction variables are selected for the purpose of the model evaluation with the dataset.

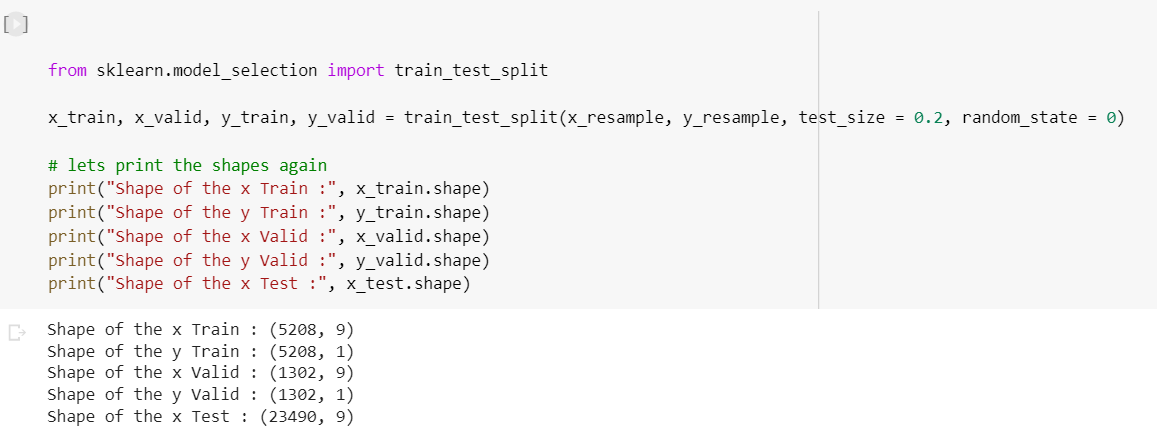


Results of resampling:

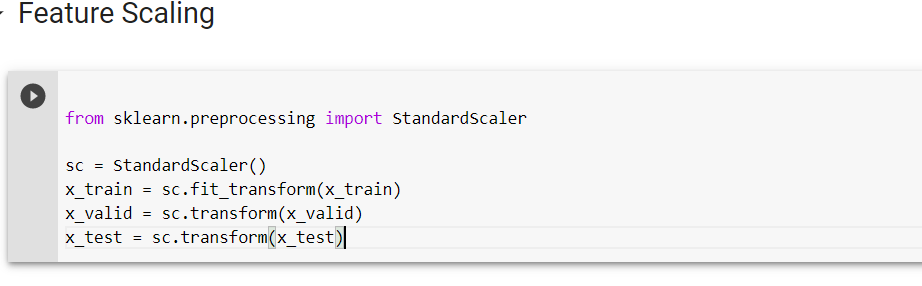




Evaluating the model:



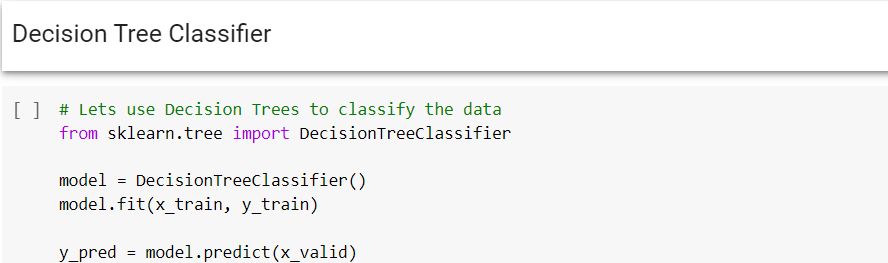
The training and test data are split to pass into the model for machine learning algorithm.



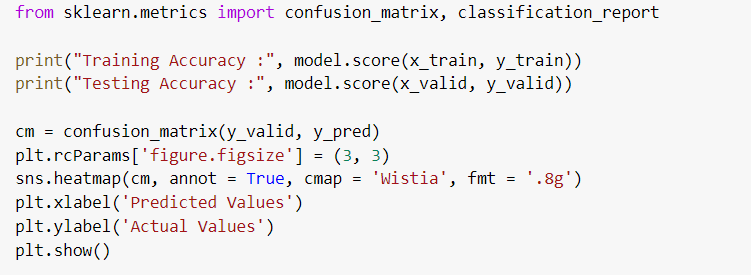
Feature scaling is a method used to normalize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data preprocessing.

**Machine Learning Predictive Modelling**

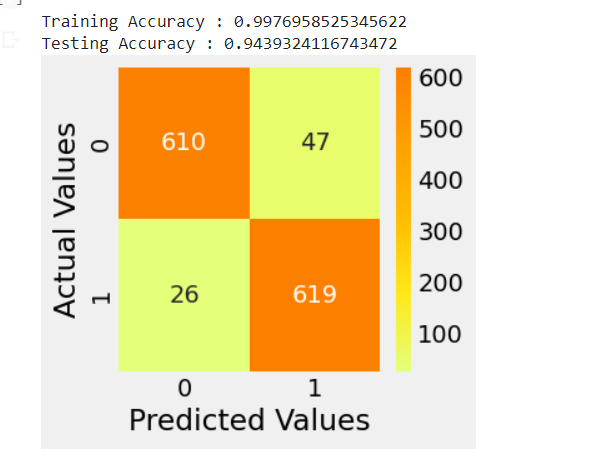
The Decision Tree classifier is applied for the employee perfromacne rating system.



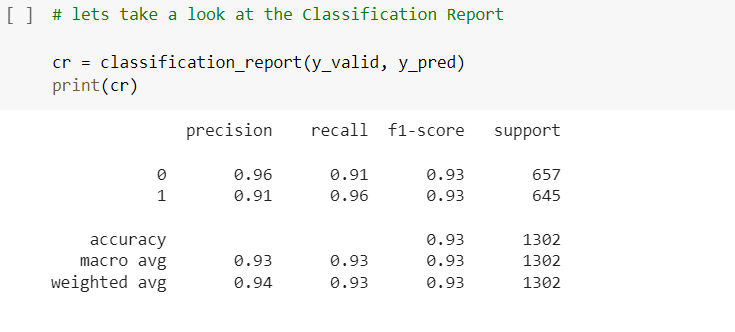
The algorithm is executed and the accuracy is calculated.



Results:

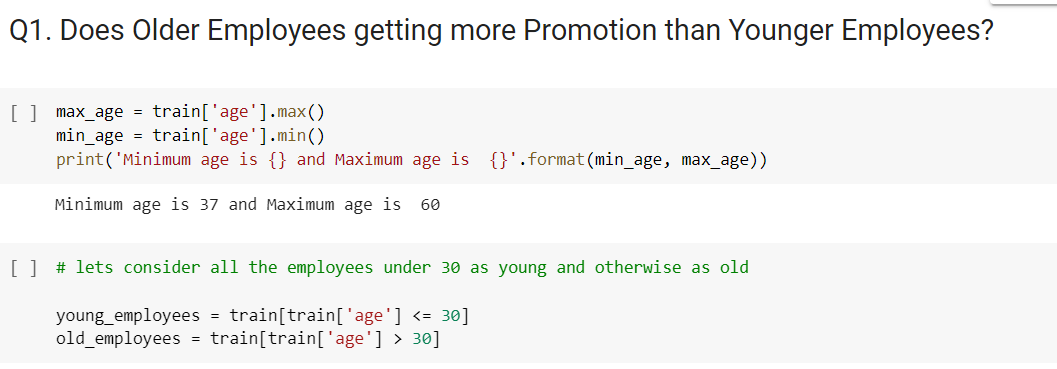


Classification report:



The classification report displays the accuracy, preicion, recall, f1-score and support.

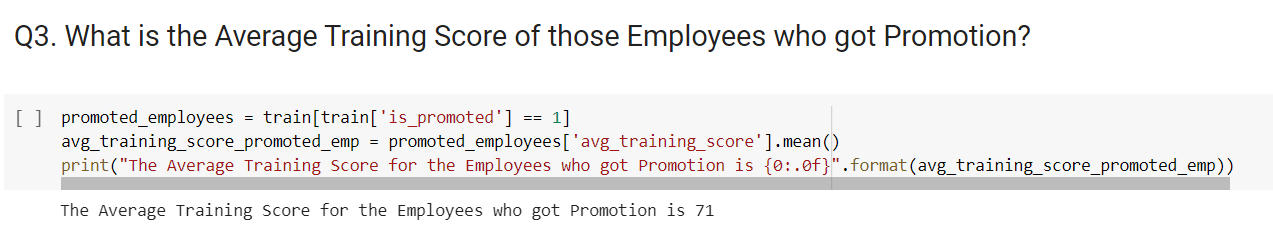
Results of Employe get promotion:







The analysis shows the empoyees won the awards is promoted or not in the softwre company.



The results shows the avaerage training score for the employee who got promotion is 71.

Work completed:

• Description

The employee performance dataset containing the information of the employee performance details report has been analyzed with different machine learning algorithms and the results are compared with the graphical reports. The logistic regression algorithm provides the better results.

• Responsibility (Task, Person)

Preprocessing – Divya Chilukuri

Analysis –

Model evaluation -

• Contributions (members/percentage)

Preprocessing –

Analysis – Divya Chilukuri

Model evaluation -

**Work to be completed**:

• Description

The future enhancement of large dataset can be planned and the accuracy will be verified so that the employe ratings performacne prediction accuracy will be higher.

• Issues/Concerns

Predicting the exact employe ratings performacne based on the dataset is one of the major challenges in the findings and it can be improved by considering large dataset.

**References/Bibliography:**

[1] Kevser Şahinbaş, " Employee Promotion Prediction by using Machine Learning Algorithms for Imbalanced Datase ", *n September 14,2022 at from IEEE Xplore©2022 IEEE*.

[2] Jung-gu Choi1\*, Inhwan Ko2\*, Jeongjae Kim1, Yeseul Jeon3, and Sanghoon Han1,2, " Machine Learning Framework for Multi-level Classification of Company Revenue", *IEEE* 2021.

[3] P. Choudhury, C. Foroughi and B. Larson, "Work- from- anywhere: The productivity effects of geographic flexibility", *Strategic Management Journal*, vol. 42, no. 4, pp. 655-683, 2021.

[4] Richard Josep, Shreyas Udupa and Sanket Jangale, "Employee Attrition Using Machine Learning And Depression Analysis ", *978-0-7381-1327-2/21/$31.00 ©2021 IEEE*.

[5] B. Becker, B. Gerhart, “The Impact of Human Resource Management on Organizational Performance: Progress and Prospects,” Academy of Management Journal, vol. 39, no. 4, pp.779–801, Aug. 1996.

[6] D. E. Guest, “Human resource management and performance: a review and research agenda,” International Journal of Human Resource Management, vol. 8, no. 3, pp. 263–276, Jun. 1997.

[7] J. Pfeffer, “Seven Practices of Successful Organizations,” California Management Review, vol. 40, no. 2, pp. 96–124, Winter, 1998.

[8] J. E. Delery, D. H. Doty, “Modes of Theorizing in Strategic Human Resource Management: Tests of Universalistic, Contingency, and Configurational Performance Predictions,” Academy of Management Journal, vol. 39, no. 4, pp. 802–835, Aug. 1996.

[9] J. B. Arthur, “Effects of Human Resource Systems on Manufacturing Performance and Turnover,” Academy of Management Journal, vol. 37, no. 3, pp. 670–687, Jun. 1994.

[10] M. A. Youndt, S. A. Snell, J. W. Dean, Jr., D. P. Lepak, “Human Resource Management, Manufacturing Strategy, and Firm Performance,” Academy of Management Journal, vol. 39, no. 4, pp. 836–866, Aug. 1996.

[11] J. P. MacDuffie, “Human Resource Bundles and Manufacturing Performance: Organizational Logic and Flexible Production Systems in the World Auto Industry,” Industrial and Labor Relations Review, vol. 48, no. 2, pp. 197–221, Jan. 1995.

[12] M. A. Huselid, “The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance,” Academy of Management Journal, vol. 38, no. 3, pp. 635–672, Jun. 1995.

[13] <https://colab.research.google.com/>

[14] <https://www.tutorialspoint.com/google_colab/what_is_google_colab.htm>

[15] <https://www.codingforentrepreneurs.com/courses/python-google-colab-sheets-drive/>

[16] V. Veloso, R. Santos, and P. Cortez “Predicting Employee Performance Using Machine Learning Techniques: A Case Study in a Software Company" IEEE 2017

[17] M. U. Khan, S. A. Hussain, and M. F. Mughal, “Performance Prediction of Employees Using Machine Learning Techniques" IEEE 2019 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)

[18] P. R. Deshpande, A. D. Shinde, and S. P. Sonavane, "Predicting Employee Performance in Software Organizations using Machine Learning Techniques" IEEE 2020

[19] M. U. Khan, S. A. Hussain, and M. F. Mughal, "Performance Prediction of Employees Using Machine Learning Techniques" IEEE 2019

[20] M. Bhatti, A. Asif, and A. Alvi, "Predicting Employee Performance Using Machine Learning Techniques" IEEE 2017