**CONTENTS**

1.Title of the project

1.1 Introduction

1.2 Objectives of Research

1.3 Problem statement

2.Review of Literature

3.Data Collection

4.Methodology

4.1 Exploratory Data Analysis

4.1.1Figures and Tables

4.2 Statistical techniques and data visualization

4.3 Data Modeling Using Supervised ML techniques

5.Finding and Suggestions

6.Conclusion

**Company Turnover Prediction**

**1.1-Introduction**

This post presents a reference implementation of a company turnover analysis project that is built by using Python Script-Learn library. In this article, we introduce Logistic Regression, Random Forest, and Support Vector Machine. We also measure the accuracy of models that are built by using Machine Learning, and we assess directions for further development.The companies invest time and money in training those employees, not just this but there are training programs within the companies for their existing employees as well. The aim of these programs is to increase the effectiveness of the companies.

**1.2 Objectives of Research**

* The objective of the proposed study wants to shows that there is a relationship/impactbetween organization culture, pay scale, evaluation by fair standard and tension and is directly or indirectly related with company turnover.
* The objective of this study is to know the organizational culture that effect on companyturnover.
* Through this study we know the pay scale and company satisfaction that prevail inmarket.
* Through this study we know how organization gives benefit to company and how theyevaluate the company performance.
* This study also identifies the employee mental level satisfaction and tension that causescompany turnover.

**1.3: Problem Statement:**

The aim of the present report is to study factors like salary, superior – subordinate relationship, growth opportunities, facilities, policies and procedures, recognition, appreciation, suggestions, co- workers by which ithelps to know the Attrition level in the organizations and factors relating to retain them. This study also helps to find out where the organizations are lagging in retaining.

**2. Review of literature:**

Company turnover can be interpreted as a leak or departure of intellectual capital from the organizations.

Turnover is an accounting concept that calculates how quickly a business conducts its operations. Most often, turnover is used to understand how quickly a company collects the amount which has to be spent by a company how fast the company sells its inventory.

**3. Data Collection**

Data will be collected from secondary resources employees, manager, reference material, current affairs, resources, in-service training, community participation and others. We collect a data from interview. The data will be collected company report, employees who work within the company and company culture and infrastructure. Some test are conducted through this we collect the data.

**4. Methodology**

Brief Description of Algorithms Used:

*Decision Tree Classifier:*

Decision Tree is a supervised machine learning algorithm used to solve classification problems. The main objective of using Decision Tree in this research work is the prediction of target class using decision rule taken from prior data. It uses nodes and internodes for the prediction and classification. Root nodes classify the instances with different features. Root nodes can have two or more branches while the leaf nodes represent classification. In every stage, Decision tree chooses each node by evaluating the highest information gain among all the attributes.

*Random Forest:*

Random Forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because it’s simplicity and the fact that it can be used for both classification and regression tasks. In this post, you are going to learn, how the random forest algorithm works and several other important things about it.

*K-Nearest Neighbors:*

K-Nearest Neighbors (KNN) is one of the simplest algorithms used in Machine learning for regression and classification problem. KNN algorithms use a data and classify new data points based on a similarity measures (e.g. distance function). Classification is done by a majority vote to its neighbors. The data is assigned to the class which has the most nearest neighbors. As you increase the number of nearest neighbors, the value of k, accuracy might increase.

*Accuracy Measures:*

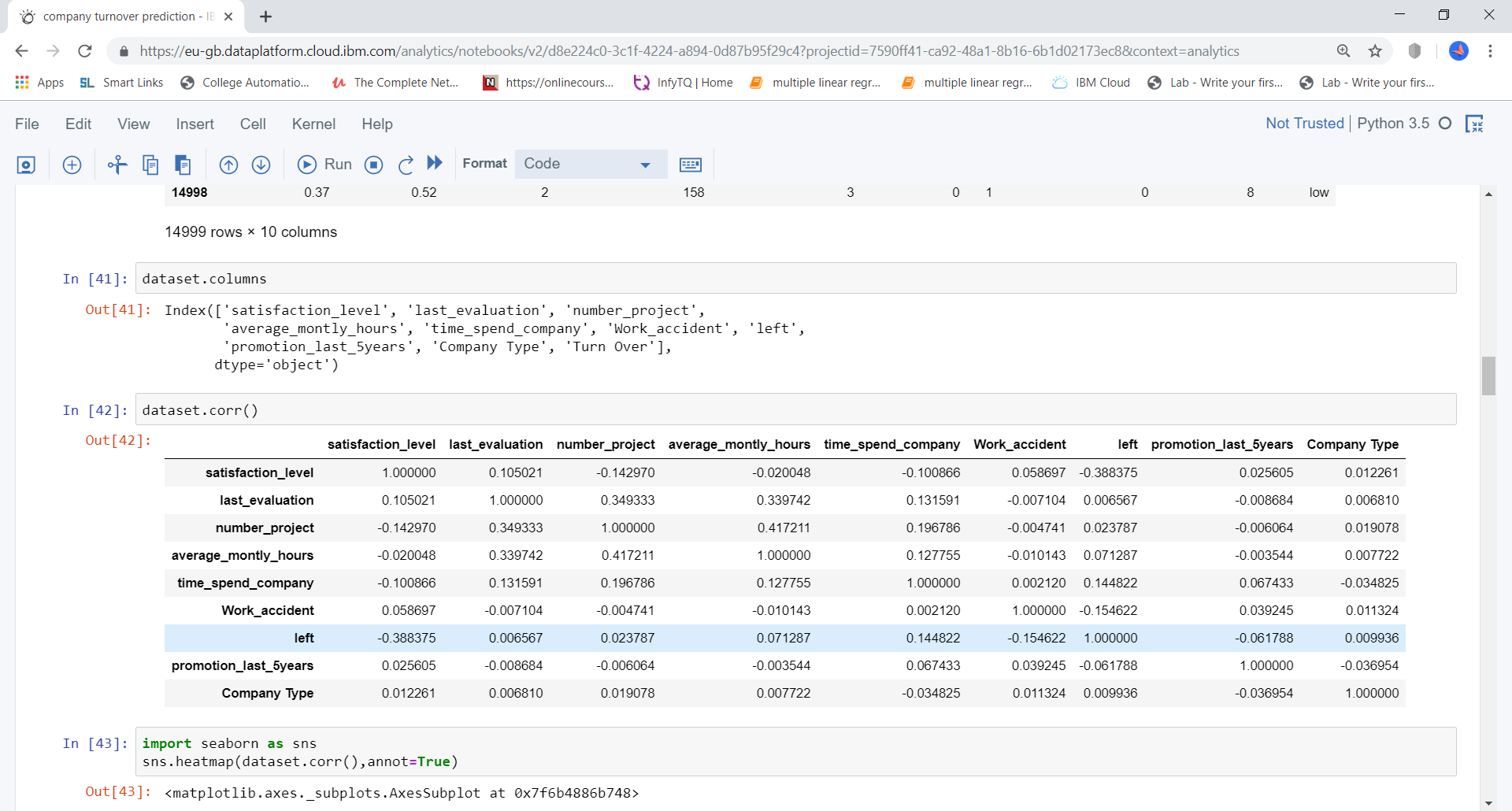
Decision Tree, Random Forest, KNN algorithms are used in this research work. Experiments are performed using internal cross-validation 10-folds. Accuracy, confusion matrix and ROC (Receiver Operating Curve)measures are used for the classification of this work.

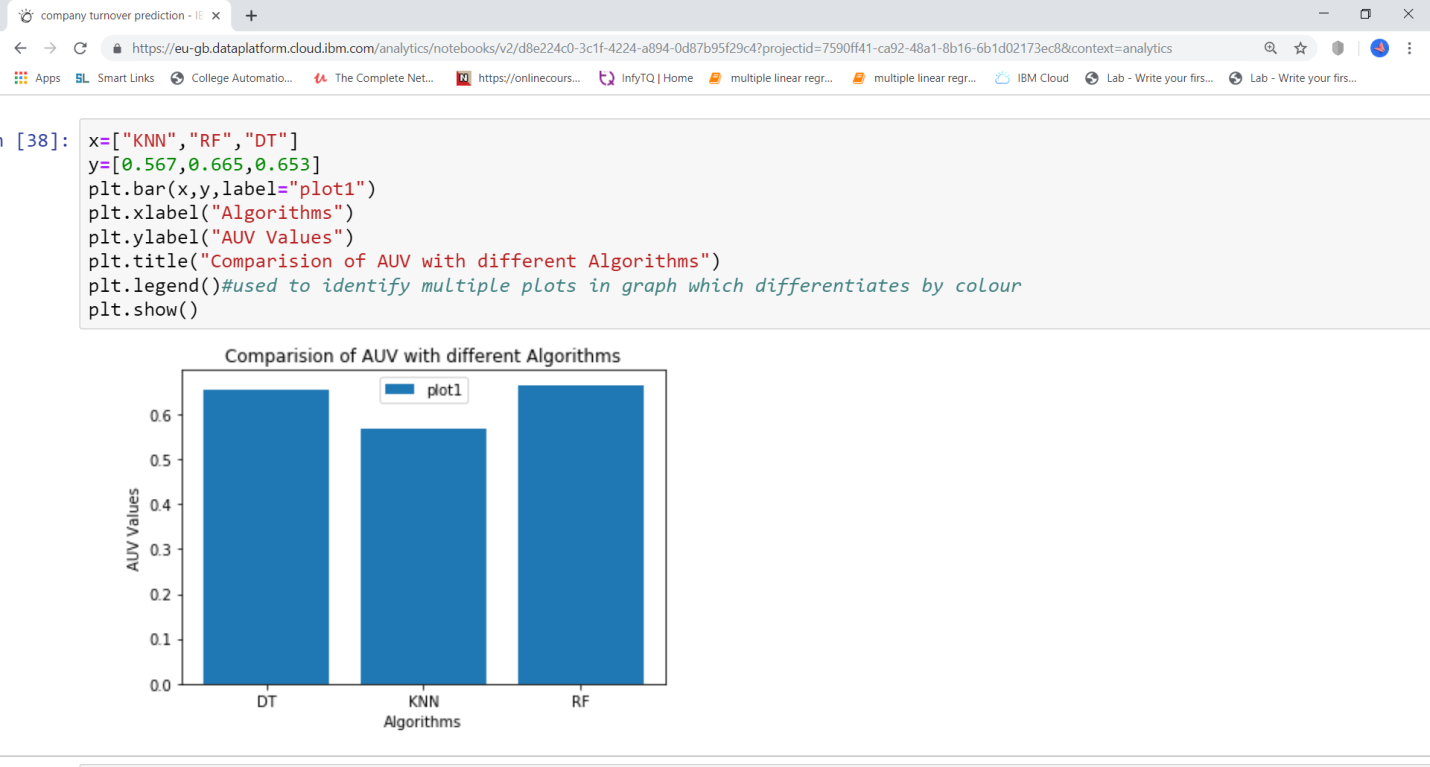
**4.1 Exploratory Data Analysis**

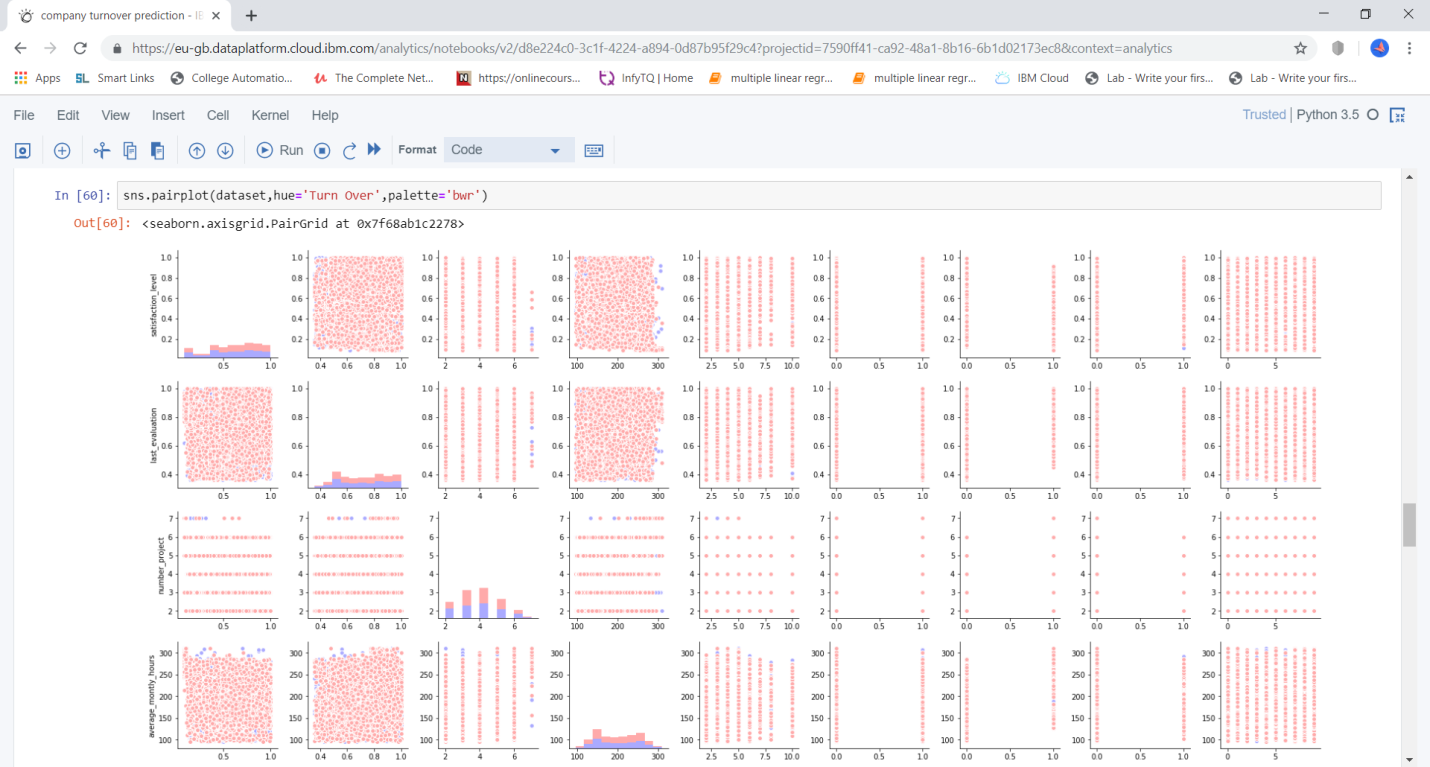
Descriptive statistics will be used to describe the company turnover. Exploratory multi-level statistical data analysis will be used to find out if there are any associations between the variousadditional organization culture, pay scale, evaluation by fair and tension and employee turnover.Descriptive statistics use to first of all we summarize the data by getting five figure summary,check the data from errors, summarize the data examining and comparing frequency distributionand check the normality of data. We analysis the data to make conclusion

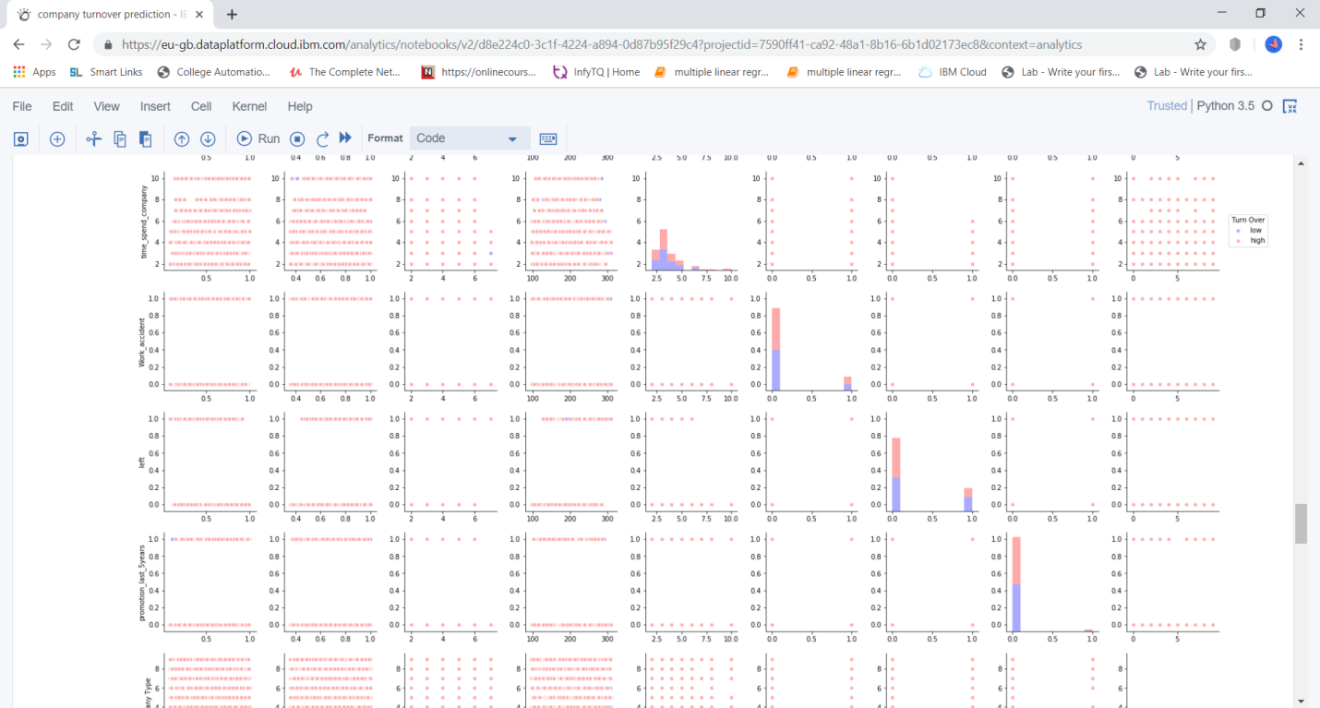
**4.1.1 Figures and Tables**

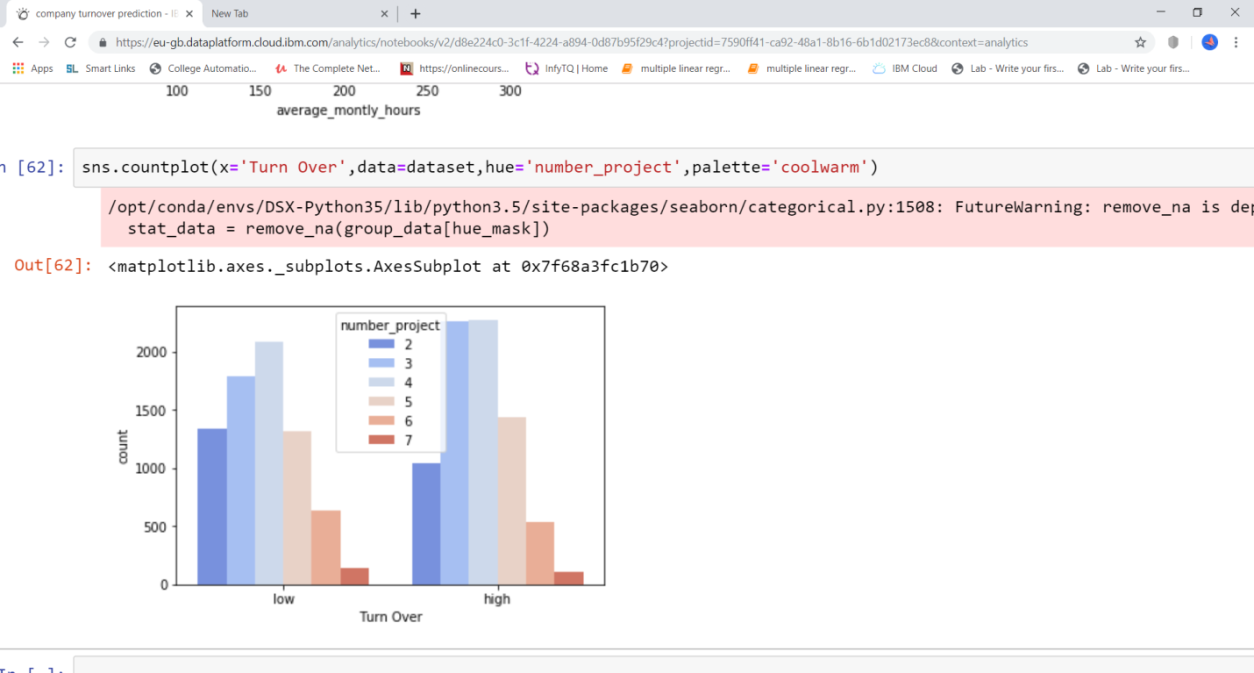
By importing matplotlib.pyplot library we have drawn graphs to demonstrate the AUC-ROC

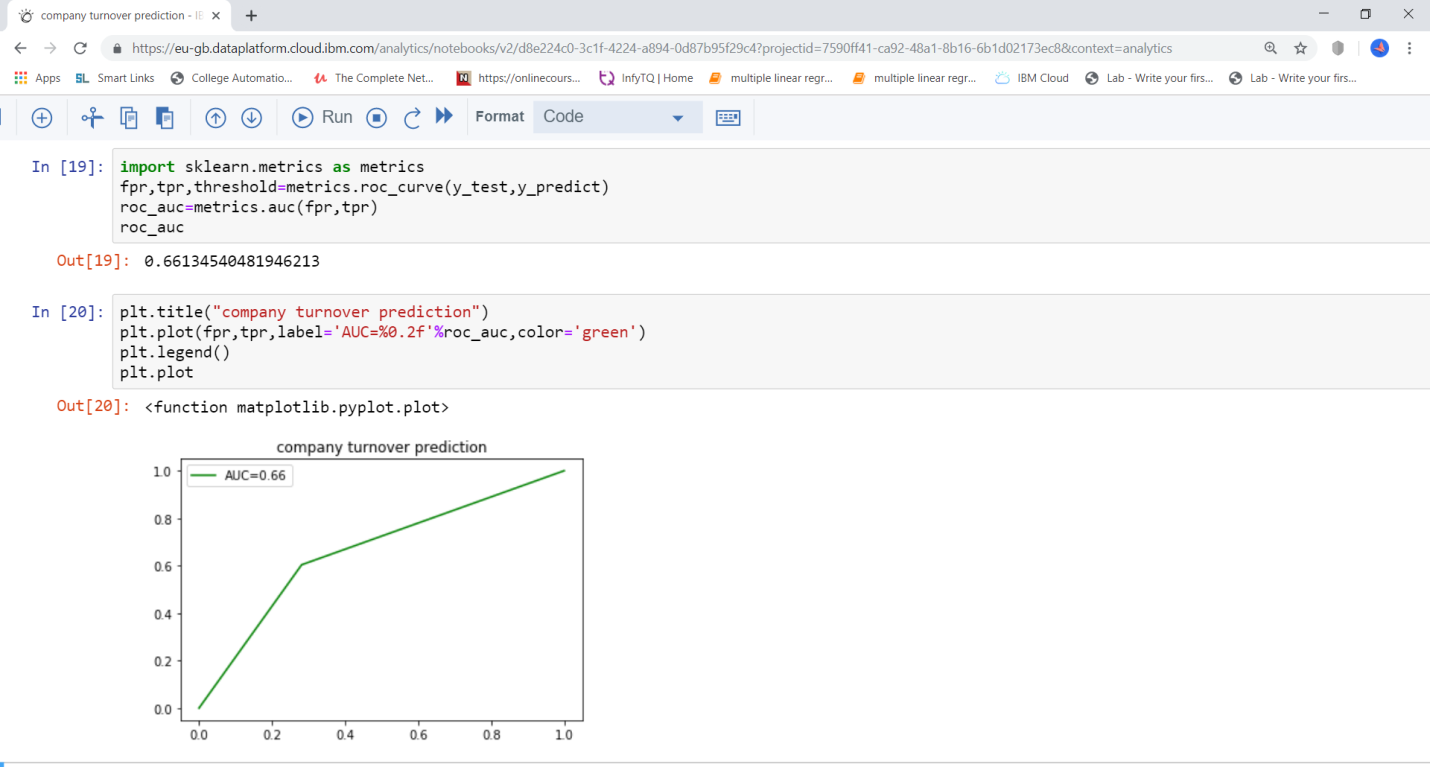


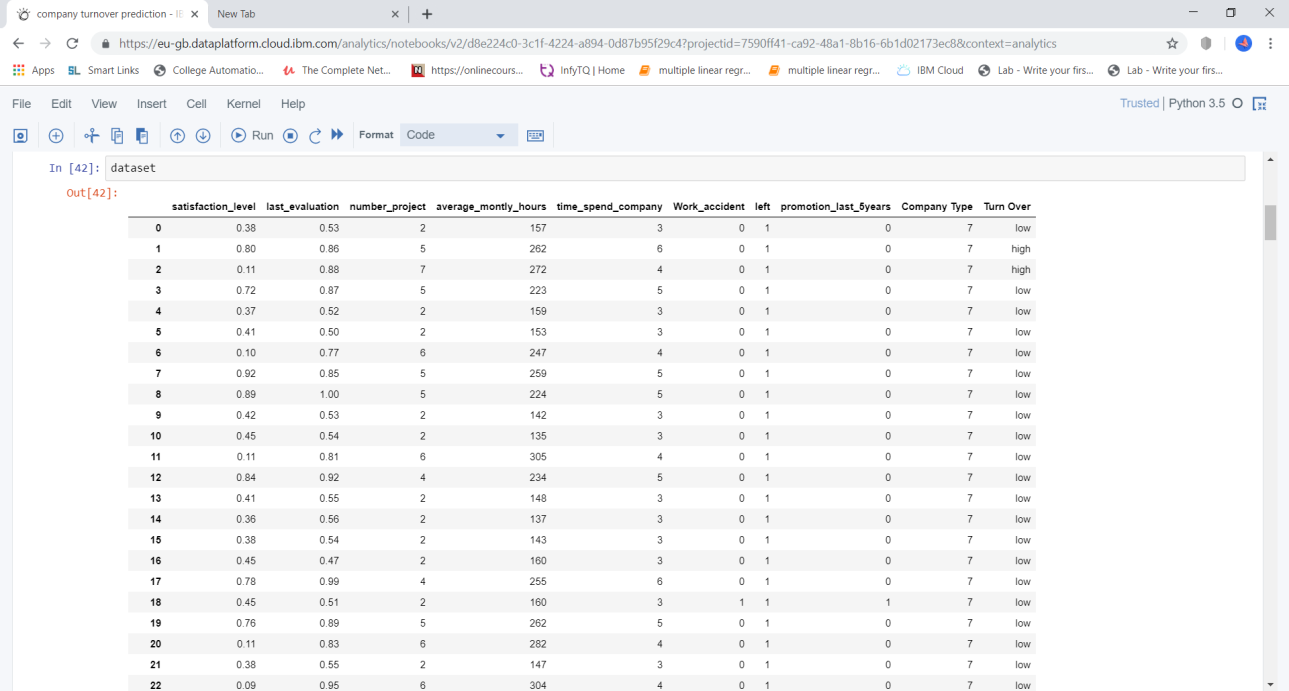


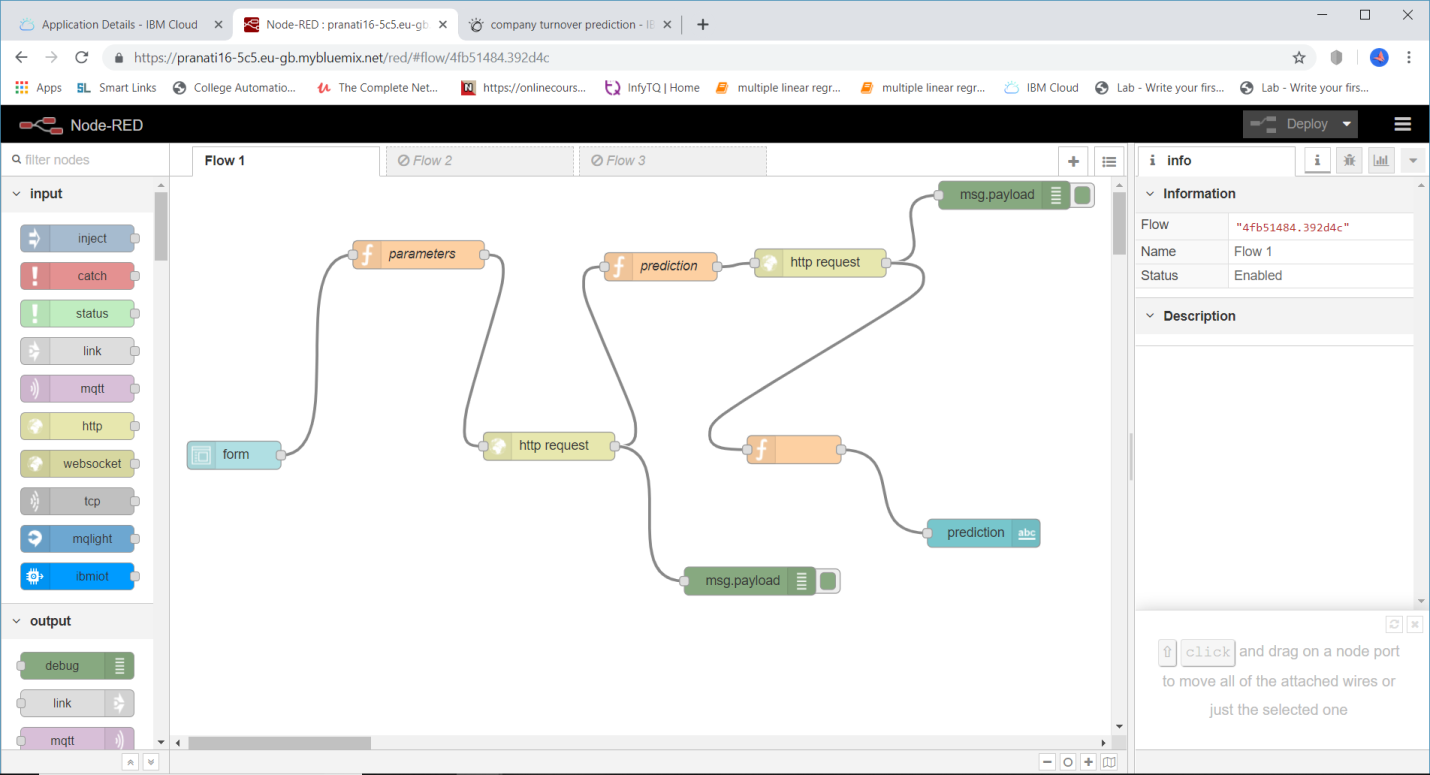


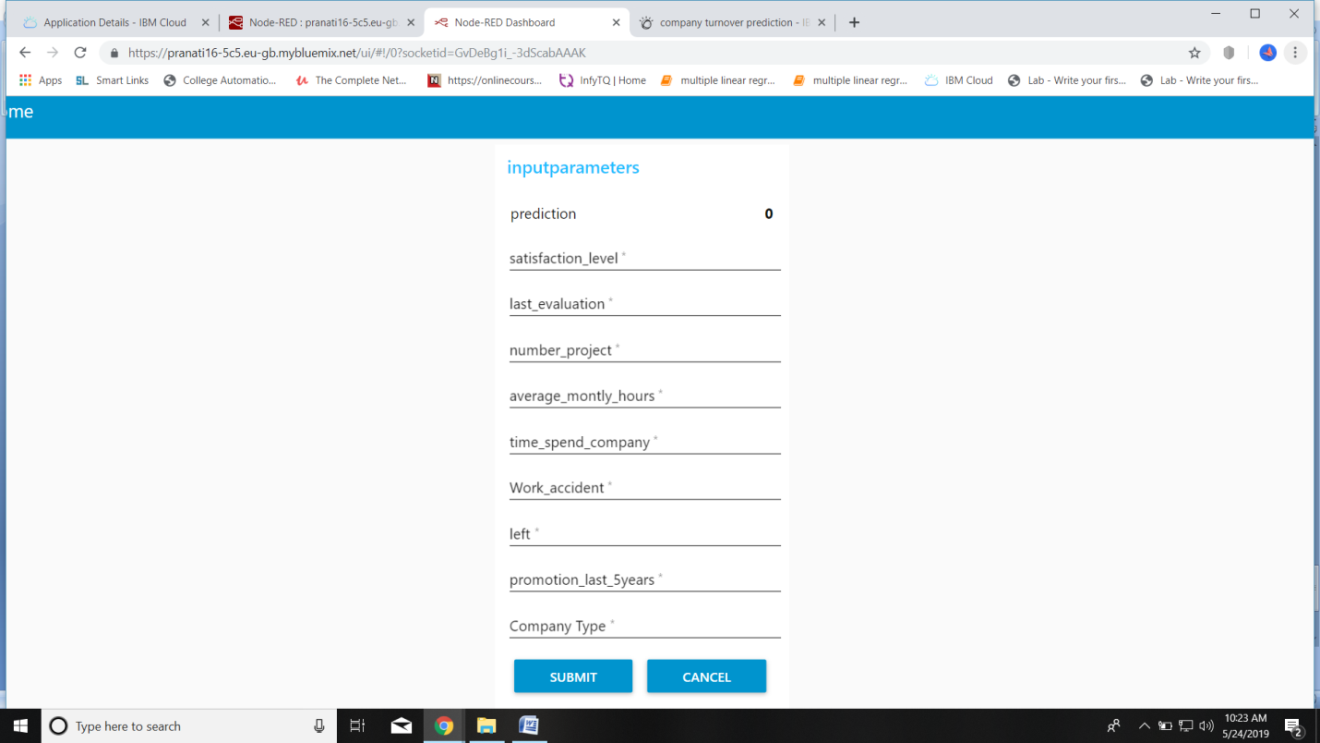


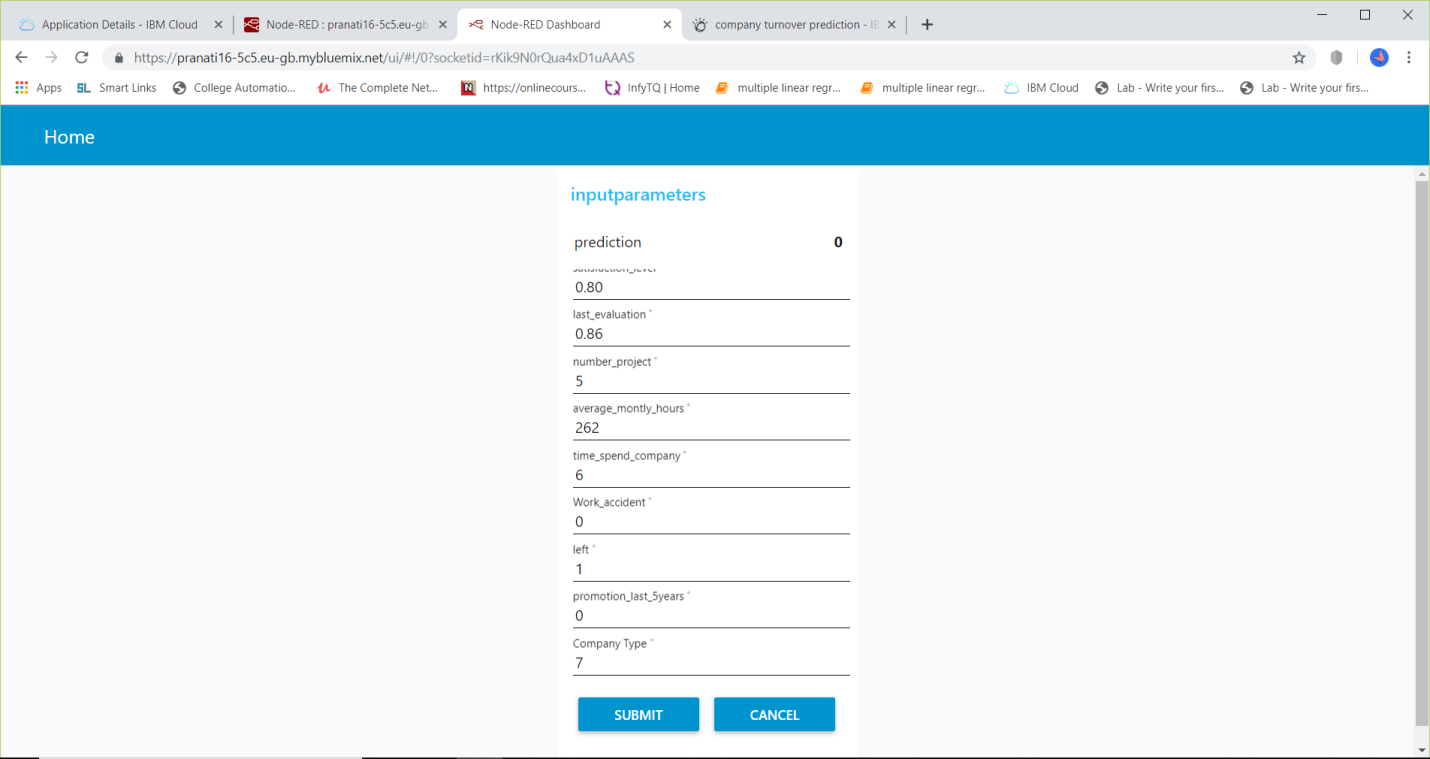


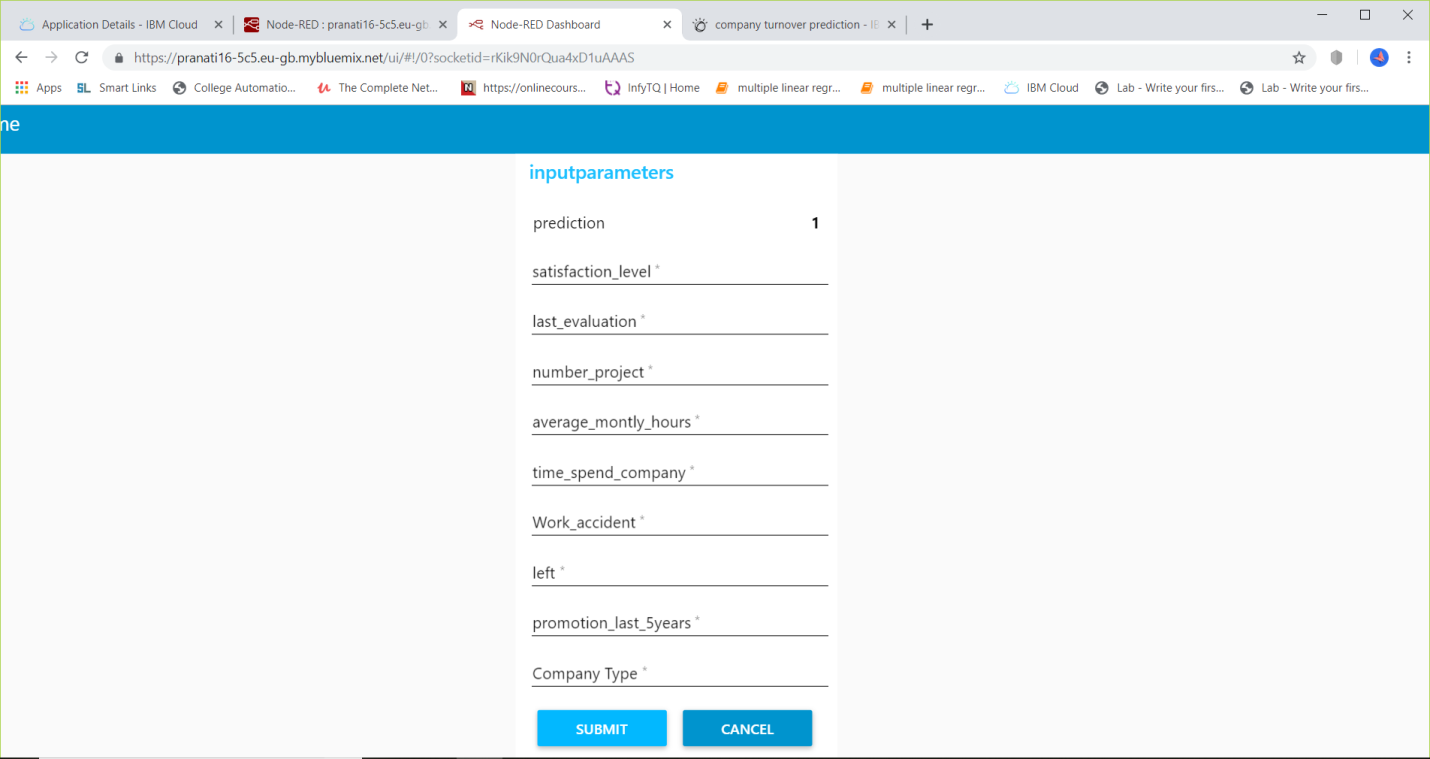








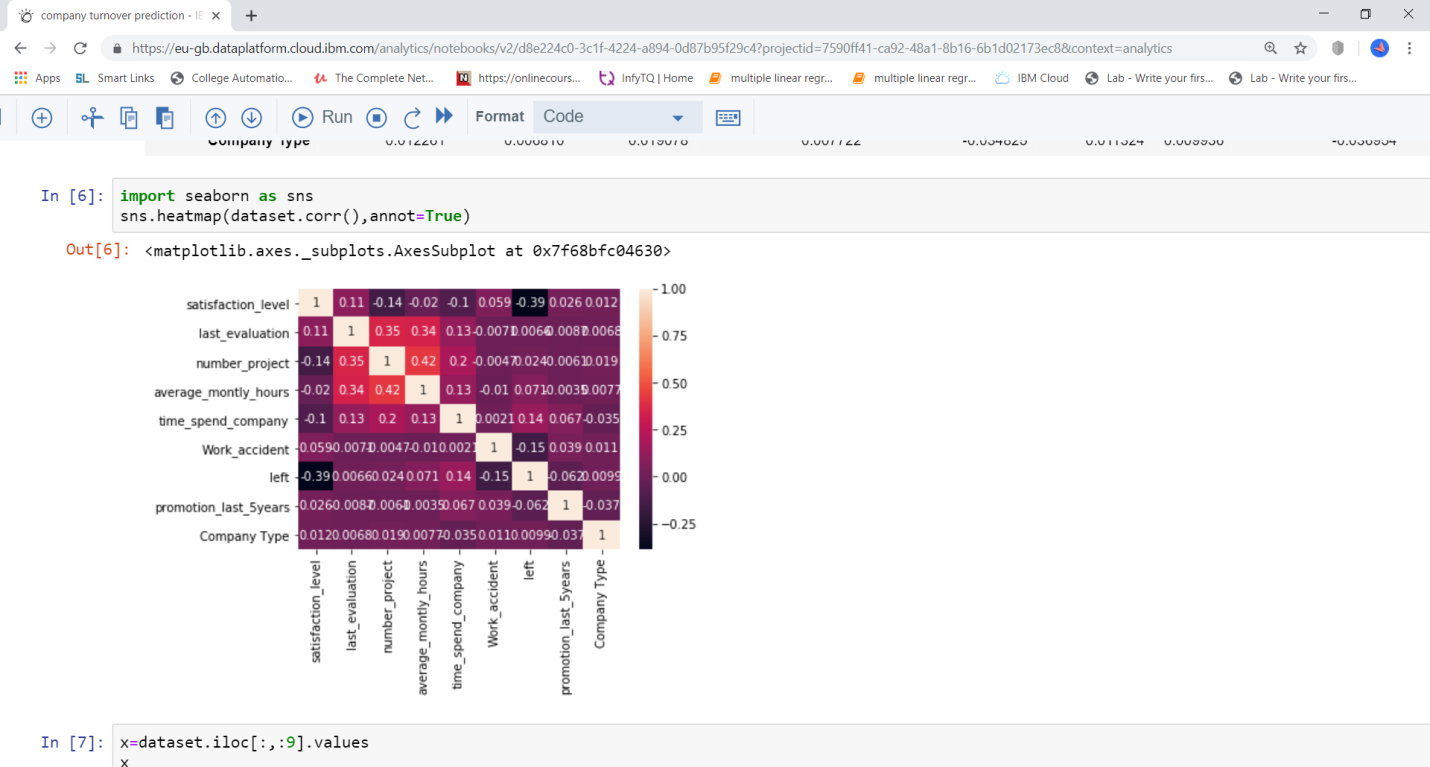




4**.2 Statistical techniques and data visualization**

By importing matplotlib**.**pyplot library we have drawn graphs to demonstrate AUC-ROC curves and by using bar graphs we have visualized the percentage levels of different techniques.

And we have used the co-relation function to demonstrate the impact of every factor on each other.

*Correlation-graph:*

**4.3 Data Modelling**

In our model we have used Random Forest Classifier Algorithm.Random Forest is a supervised learning ensemble algorithm. Ensemble algorithms are those which combine more than one algorithms of same or different kind for classifying objects. The ‘forest’ that Random Forest Classifier builds, is an ensemble of Decision Trees, most of the time trained with the ‘bagging’ method. The general idea of the bagging method is that a combination of learning models increases the overall result .Random Forest does not overfit . We can run as many trees as you want. It is fast.

Random forest classifier creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test object.Random Forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model.

The Random Forest creation pseudocode:

1. Randomly select “**K**” features from total “**m**” features where **k << m**
2. Among the “**K**” features, calculate the node “**d**” using the best split point
3. Split the node into **daughter nodes** using the **best split**
4. Repeat the **a to c** steps until “l” number of nodes has been reached
5. Build forest by repeating steps **a to d** for “n” number times to create **“n” number of trees**.

**5. References**

https://www.clarifai.com/

https://niki.ai/

https://lyrebird.ai/

https://www.cortica.com/

https://towardsdatascience.com/

https://www.kaggle.com/  
**6.Conclusion**

Finally the turnover prediction of the dataset is calculated by using algorithms like Random Forest,KNN algorithm and Decision tree .The accuracy value is also predicted by these algorithms, the Auc-Roc curve is also have been Plotted .Future research could examine reasons for termination of company, as this would allow for any seasonal fluctuations in employment. Development of a more specific data mining tool which would address such factors as the existence of support systems among employees, changes in organizational strategies and inconsistencies in job expectations based on initial job which would provide valuable data for companies to retain their employees.