GR-267 Churn Prediction

Abstract

Employee churn is a situation where people leave the organization voluntarily or involuntarily. It has become a serious problem in recent times. We have also seen that attrition rates in several industries are going high. So it is very much required to understand and analyze the reason behind attrition and why it is happening. We have to conduct an analysis to know what the factors affecting employee churn. It will create a huge impact on the organization if the attrition rate goes high. In order to resolve this issue we are trying to take up this issue and find the best solution for this.

Introduction

The act of leaving a company voluntarily or involuntarily is known as employee churn. It's become a serious problem lately. We also see high turnover rates in some industries. Therefore, it is very important to understand and analyze the reasons for wear and why they occur. Analysis needs to be done to know what factors influence employee turnover. A high turnover rate has a significant impact on an organization. To solve this problem, we take this issue and try to find the best solution.

Research Question(s)

We are building an efficient technique to build model which will predict whether candidate/Employee will accept the job offer or not. We are building classifier which will predict whether candidate will accept the job offer or not. It is basically a binary classifier. We are making use of machine learning algorithms like random forest and logistic regression which are powerful to make classification.

Materials and Methods

The act of leaving a company voluntarily or involuntarily is known as employee churn. It's become a serious problem lately. We also see high turnover rates in some industries. Therefore, it is very important to understand and analyze the reasons for wear and why they occur. Analysis needs to be done to know what 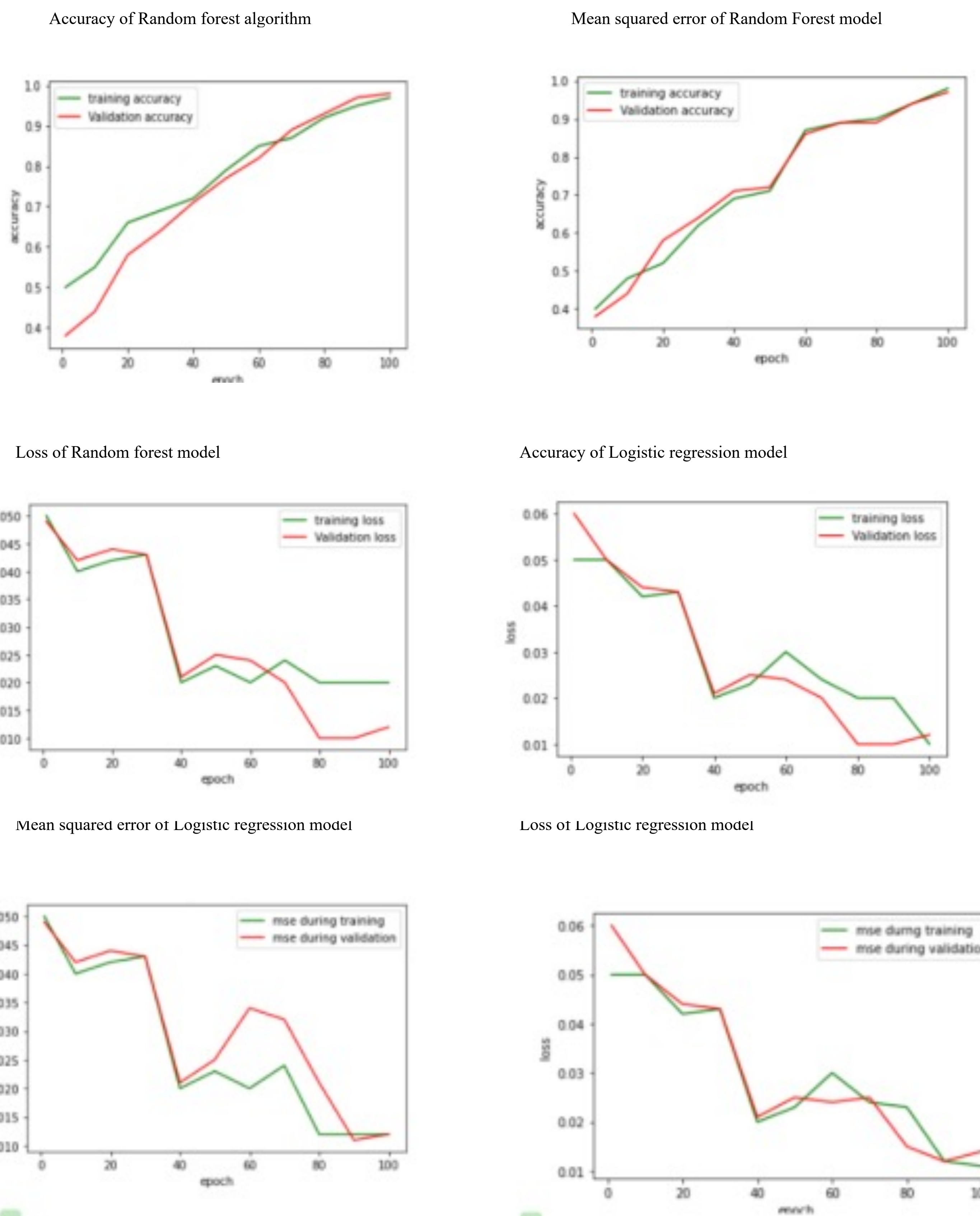factors influence employee turnover. A high turnover rate has a significant impact on an organization. To solve this problem, we take this issue and try to find the best solution.

We will be implementing ML algorithms like Random forest and Logistic regression. Random forest builds trees in parallel and we will be performing the bagging technique to find the average accuracy of all trees built. Learning is also done much better and training is done in parallel. These are the reasons why we have taken this ensemble model for our thesis. Logistic regression finds a sigmoid curve to classify the data into specific categories by finding the probability of occurrence of an event. As we are finding whether a candidate will be a churn or not, it will be best to apply logistic regression to find probability and make predictions for the same
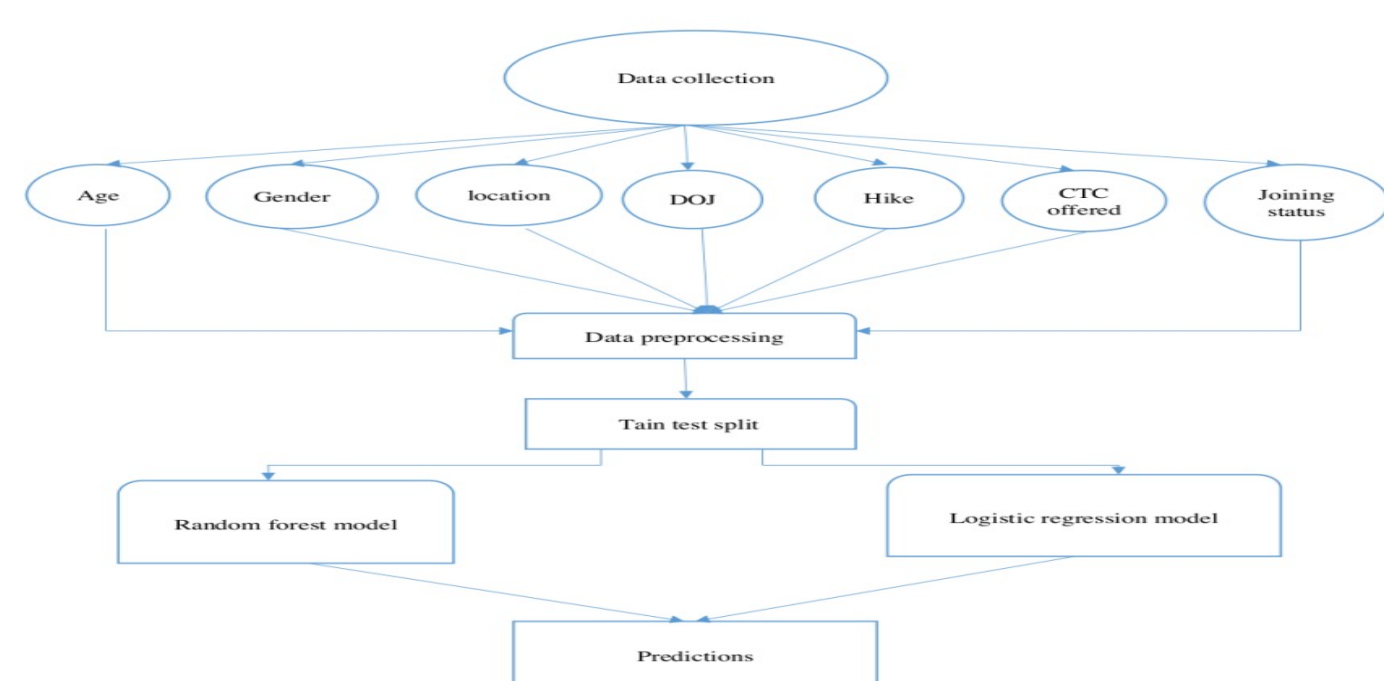
Results

We have built Random forest model with 97% accuracy and 98% precision. We have also got the confusion matrix with the help of which we have found out specificity and f1 score. F1 score of our model is 0.97 and specificity is 89%. Our second model is logistic regression model for which we have got 94% accuracy and 95.5% precision. . We have also got the confusion matrix with the help of which we have found out specificity and f1 score. F1 score of our model is 0.94 and specificity is 94%. Other results are given below.

Accuracy of Random forest algorithm

Mean squared error of Random Forest model

Loss of Random forest model

Accuracy of Logistic regression model

Mean squared error of Logistic regression model

Loss of Logistic regression model

Conclusions

In this project we collected Employee churn dataset. We have performed preprocessing, exploratory data analysis and visualization. We have made use of python libraries like pandas, matplotlib and sklearn libraries so far to implement this. We have built random forest and logistic regression model based on which we have come up with best model that is random forest model because of its evaluation metrics mentioned in results section.

Acknowledgments

As data modelling is not just about building model, we have written all the steps undertaken before building the model with the help of this block diagram shown in fig 1. Graphs showing accuracy and loss from the first epoch till end is also shown to properly display how the training has been done.
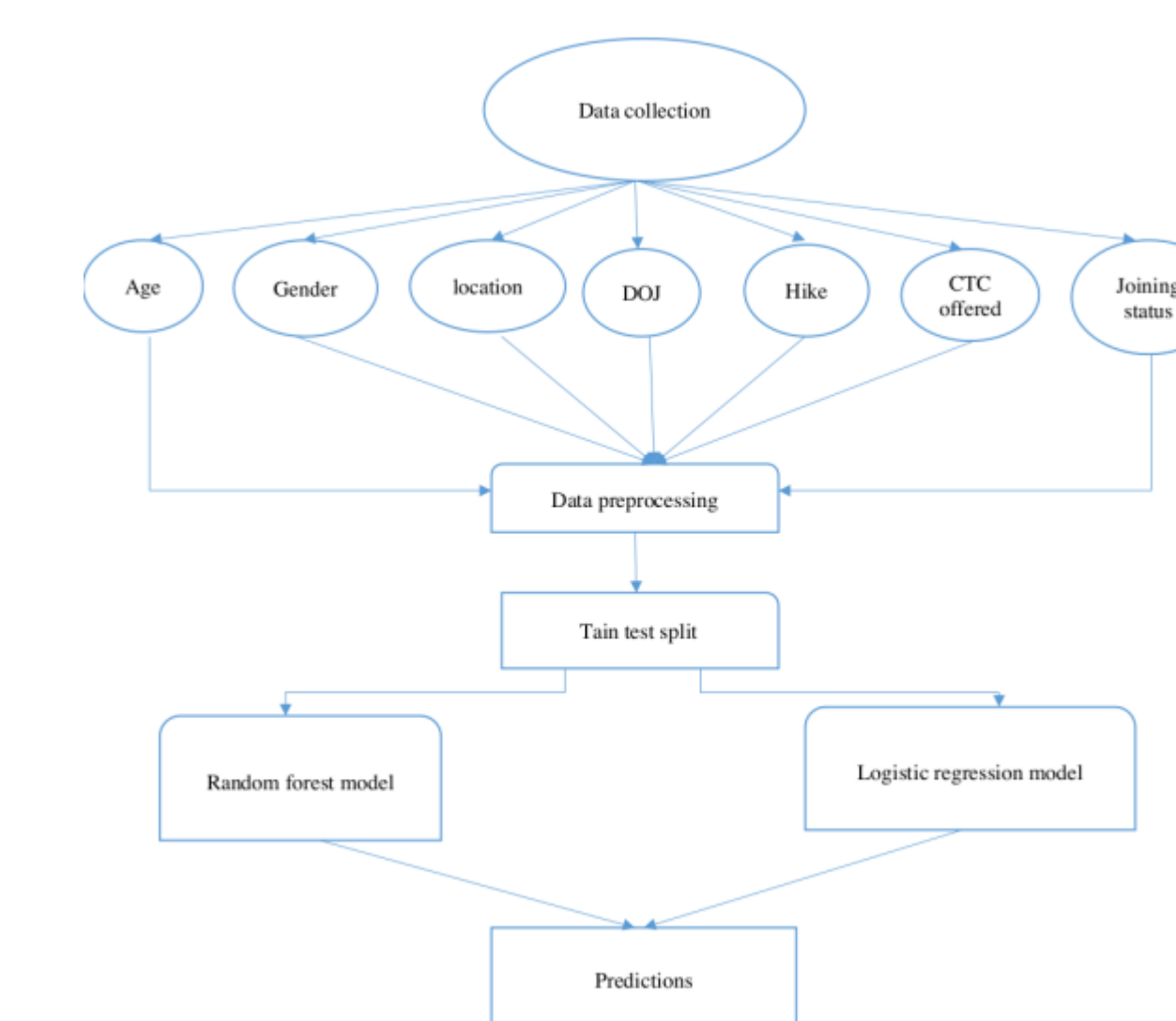
References

Ibrahim Onuralp Yiğit, Hamed Shourabizadeh, An Approach for Predicting Employee Churn Available at: https://www.researchgate.net/publication/320298197_An_Approach_for_Predicting_Employee_Churn_by_Using_Data_Mining

Andry Alamsyah, A Comparative Study of Employee Churn Prediction Model. Available at: https://www.academia.edu/37841192/A_Comparative_Study_of_Employee_Churn_Prediction_Model

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING

Author(s) : mdasari, nthota, dthota, rbasava, psugguna (@students.Kennesaw.edu)
Advisors(s): Dan Lo (dlo2@kennesaw.view.usg.edu)