Data pre-processing

Sai Anirudh Chatakonda,
*School of Computing Sciences and Engineering, VIT Chennai, Tamilnadu, India 600127
Email: sai.anirudh2016@vitstudent.ac.in*

[Credit card default prediction]

Abstract—here this is data set of the credit card holders, which has the attributes like the cardholder's education job type, sex, marital status, income and depts.

Based on this we apply our machine learning algorithms to predict who are the defaults and who are not.

We are going to apply this on our test date set and check our predictions

1. Introduction

Here in this session we are taking the data set and pre-processing the data set like filling all the unknown values then converting all the values in the form of integers by giving the weights and plotting the graphs.

2. Database – Credit default prediction dataset

[Describe the database, its characteristics, if possible plot

This is the credit default prediction data set where there are attributes like education, income and credits based on that we predict will he be paying next month emi or not.

3. Experiments

In this we did the data pre-processing and fill all the missing data

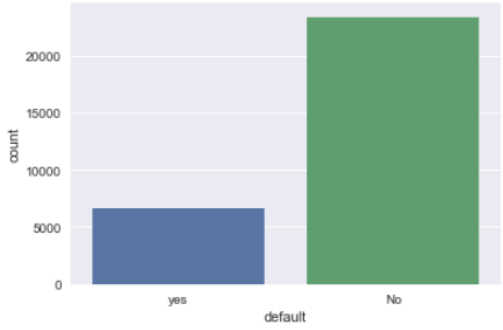
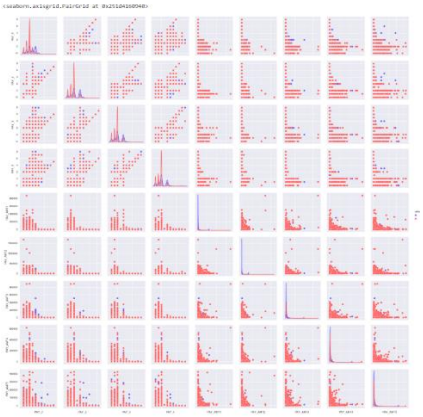


Figure 4. Three class classification performance

Here we can also find the from the graphs which are all the factors which are depend on predicting the defaulters



From this we can see on which it depend and we can take those factors into count and ignore those factors which does not make sence.

4. Conclusion

Here in this data set by applying the machine learning algorithms we can predict whether a given person will be able to pay the next month credit due or not.

References

[1] R. Fisher, "The use of multiple measurements in taxonomic problems," *Annual Eugenics*, vol. 7, no. Part 2, pp. 179–188, 1936.

[2] W. McCulloch and Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115–133, 1943.