

DATA SCIENCE FOR BUSINESS (DATA ANALYTICS)

PREDICTION OF MOVIE SUCCESS USING RANDOM FOREST A MACHINE LEARNING ALGORITHMS AND THEIR COMPARISON

IV. DISCUSSION AND CONCLUSION

Two open-source tools like R-code and Python are used to analyze movie data. The purpose of the study is the same on two instruments. The main aim is to conduct predictive and exploratory analyses. For evaluating the pattern of the data set exploratory analysis is used. It includes descriptive information on each of the dataset variables. The study shows that the dataset initially had 5000 records and 28 features (column). Both numerical and categorical data are available. There is also NA and incomplete data in the data collection. It is complex to create a precise model without pre-processing the data. We have prepared the model with a random forest model during this study. Movie quality according to the IMDb rating is graded as standard, good, and poor. The Random Classifier forest constructs a model for movie quality prediction. For both classification and regression, the random forest classifier is used. The two methods for the analysis of data are somewhat similar, both packages and the library are used to carry out different data manipulations. In Python, a continuous variable like IMDb Score has been used for this. The number of movie critics is closely linked to the number of movie Facebook that is common. No other variables are associated with the length. There's a great link between several voters and the number of ratings, as well as gross and movie Facebook likes.