Text

Description automatically generated with low confidence

Goal: To create a Machine Learning model to predict review score. We converted review score to binary variable by using the following code:

Graphical user interface, text, application

Description automatically generated

As our target variable is binary, we choose classification model. We created 6 different models and compared their accuracy. Over all we got pretty good accuracy >85% for all the models we tested but Random forest performed the best for our data:

1. Linear Logistic Regression: We got accuracy of 88%
2. K-NN model: We were hoping to receive higher accuracy with K-NN but we got lower accuracy of 86%
3. Kernel SVM:Non Linear: We got accuracy of 87%
4. Decision Tree: We got accuracy of 95%
5. Ensemble: Randon Forest: As this combines various decision tree models, it gave us the best value of accuracy of 96%
6. Artificial Neural Network: We were hoping to get the highest accuracy with this one but it gave us accuracy of 87%

Here is a graph that shows the comparison of all the models we tested:

Chart, bar chart

Description automatically generated