

PREDICTIONS

- Question: Is a particular review helpful?
- Y is the helpful rating.
- Instead of predicting an exact rating, the data is split into helpful and unhelpful scores.
- Reviews with a helpful rating of over 85% are helpful.
- Reviews with a helpful rating of under 50% are not helpful.
- Leaving out the middle is justified because these reviews could go either way. User results may not be as accurate due to bias.





MACHINE LEARNING

.....

- Naive Bayes, Random Forests, Decisions Trees and Logistic Regression were all attempted to make predictions.
- Logistic Regression consistently delivered the best results, followed by Naive Bayes.
- Logistic Regression Cross-Validation had AUC means of over 90%.
- Confusion Matrix precision of unhelpful ratings were over 80%, and precision of helpful ratings were over 90%.
- Hyperparameter C tuned as 0.007742636826811269.