AMAZON FOOD REVIEW

Sara Nil Acarsoy 19263   Defne Yazıcıoğlu 24098

The data used in this project is taken from https://www.kaggle.com/snap/amazon-fine-food-reviews and consists reviews of food from Amazon. The data has a 'Score' column, starting from 1 to 5 and 5 being the very positive review. Then we noticed that good scored reviews have longer text. Our hypothesis is that if a review is long it has a higher rating. Our prediction will be based on this.

The initial hypothesis testing shows that:

Then we search the most used words in highest and lowest scored reviews' texts separately while counting the word count. We made lists with positive and negative words. We searched their frequency in every review to prove our hypothesis. If a text has positive words in it we add 1 to its overall positive score and same thing for negative words.

According to these features we made a decision tree and knn analysis. Knn's accuracy was 0.84 and decision tree gives 0.85 accuracy. So for this case we would prefer decision tree because of its higher accuracy.