Cora Sarzynski  
04/01/2022  
Homework 6

The data that I am using for this assignment is common flu symptoms determining if someone has the flu or not. The aspect I focused on was predicting if someone had the flu based solely on if they had a fever or not. I am arguing for Logistic Regression since, in my case, it had more correct predictions from my data. The first step in creating these classifiers is to split the data into training and testing data. The purpose of this is to allow the program to build a database from the training data to try and predict the outcome from the testing data. It is an integral step in classifying methods. Based on my data set, Logistic Regression was better at predicting the outcome of the test data when looking at the fever symptom and flu outcome. This is shown in the underlined numbers of the tables. Logistic Regression had more correct predictions of flu outcome based on fever than Naïve Bayes did. The difference between the models, and these differences make each useful, is that Naïve Bayes is based solely on probability and Logistic Regression is used when there only needs to be a solid “Yes” or “No”. Since my data was already a lot of “Yes” and “No” (0 and 1) data, it makes sense that the Logistic Regression model worked better on my data than the Naïve Bayes model. In conclusion, both models are useful when making predictions of test data based on training data, but there are times when one works better than the other. In my case, Logistic Regression came out on top since it had more correct predictions than Naïve Bayes did.

Naïve Bayes Table: Naïve Bayes Visualization:  
A picture containing diagram

Description automatically generated Chart, line chart, histogram

Description automatically generated



Logistic Regression Table: Logistic Regression Visual:  
A picture containing text

Description automatically generated Chart, line chart

Description automatically generated

