Predicting Strep Throat in Children

By : Brady Pinter

About the Dataset

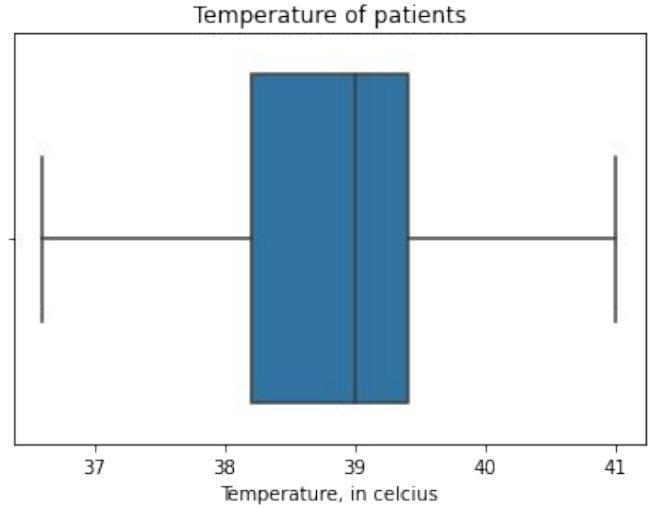
The data set I decided to us for my project was 676 cases of pharyngitis(inflammation of the throat) in children. This dataset included both quantitative and qualitative predictors, and gave measurements on a variety of questions asked to those who had pharyngitis. I decided to answer two major questions using this dataset.

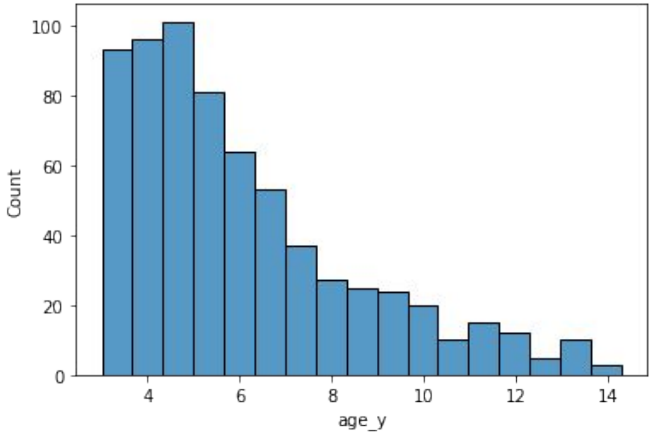
1. Can we predicting via KNN someone being radt (strep throat) positive? (Qualitative)
2. Can we using multiple linear regression to predict the age of a patient based on predictors like temperature, radt, and if they had pain when tested? (Quantitative)

Getting ideas

- To get some ideas with our dataset, here are two graphical visual representations I made of age and temperature of patients at time of measurement.
- We can tell that all our data comes from children, and that at the time of recording most had a fever. (Any temperature above 38 degrees celsius)

```
#Confirming no nans present
df_noNan = df.dropna()
```

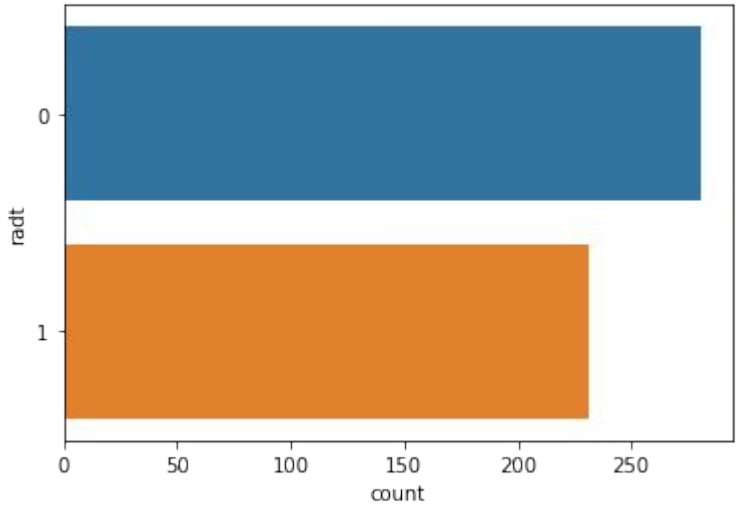




Qualitative - KNN

For the sake of time, I will show you one prediction I made. To get an idea what we will be trying to predict. (radt), here is a representation of the positives and negatives of test. You can notice that we have a fair representation of both.

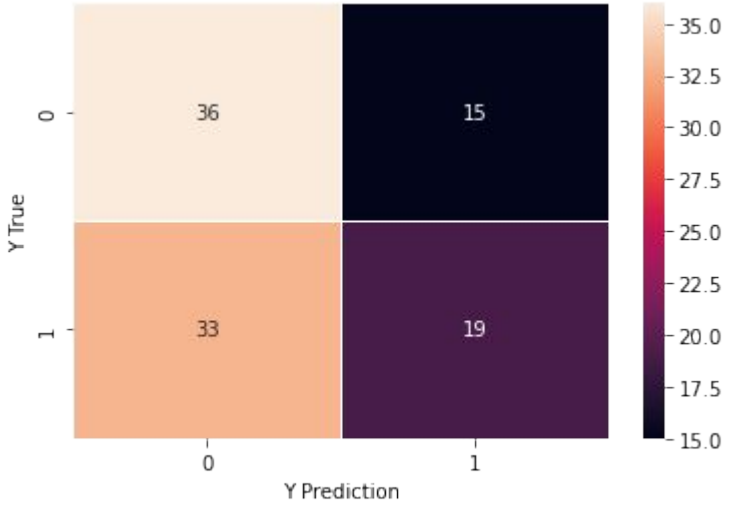
The goal of this KNN was to predict someone getting a true/negative radt based off all other predictors in the dataset.



Results

As shown by the results, our model was better at deciding on true negative than true positive. I find that our KNN model where k=3, that it is very bad a predicting true positive, only getting 19/52 correct.

The poorness of our model is also demonstrated by its accuracy score of 53%.



Accuracy Score = 0.5339805825242718

Thanks for Listening!