

1 Midterm Project Proposal

For my project, I am choosing the first option of applying existing algorithms to a new problem and potentially new data. Specifically, I would like to examine medical data and attempt to find patterns in the data as well as build predictive models to see if it is possible to predict which symptoms and lifestyle choices might be connected to the likelihood that a patient is diagnosed with a certain condition. First of all, I will start off by utilizing the Naive Bayes Classifier to examine a dataset. I'm interested in using the Naive Bayes algorithm to examine trends in data and provide a recommendation. Specifically, I have found examples of this algorithm being used to examine symptoms and measurements of patients and whether they were subsequently diagnosed with breast cancer[1], so I want to attempt to do something similar. I found a data set detailing the symptoms presented by and lifestyle details of a variety of patients as well as whether they had a history of heart attack or not, so I think I would do a project examining this data set and seeing whether I can find patterns in their demographic data, their symptoms, and whether they have a history of heart attack. This is a comparatively small data set, so in broadening the scope of this project I hope to apply what I have learned on a larger scale to data sets with more data points, and possibly examine a wider range of health conditions not just limited to physical health but also including mental health, and examining verbal data from patients in addition to physical attributes, as well as testing more potential models to see

which is best at predicting.

2 References

References

- [1] https://shiring.github.io/machine_learning/2017/03/31/webinar_code