Ahmed Youssef

Final Project Question Proposal

4/7/2019

After exploring many datasets, and evaluating possible solutions and relevance to societal need, we decided to dive into the topic of health, and specifically disease outbreaks. We narrowed down the general topic, and we choose a topic dealing with California immunization rates, in which we found a large dataset dealing with California Pertussis cases and rates over 2010 to 2014 years. Upon inspection on the dataset, we found it interesting what role did vaccination rates in every county play in infant Pertussis cases, as well as how did the reported exemption for a given school affect the outbreak, thus our question is; can we build a predictive model that predicts for a given year which counties in California are at a higher risk of Pertussis outbreaks based on previous years. This question would be of societal need as the Pertussis cases and rates over the years lead to infant hospitalization and unfortunately some deaths. In addition, if we can build a predictive model that uses previous years to predict upcoming year’s risk, then early intervention and increase in vaccination rates can help combat the issue and reduce outbreak cases. Furthermore, the model will be tested and validated on the 2014 year to help further improve the model and set possible modifications for future improvements.

Work Cited

[1] <https://www.kaggle.com/broach/california-kindergarten-immunization-rates#InfantData.csv>