Medi-App

Ali Rizvi. Alwin Baby. Kelly Holtzman. Mansi Patel.

December 8, 2020

Investigation

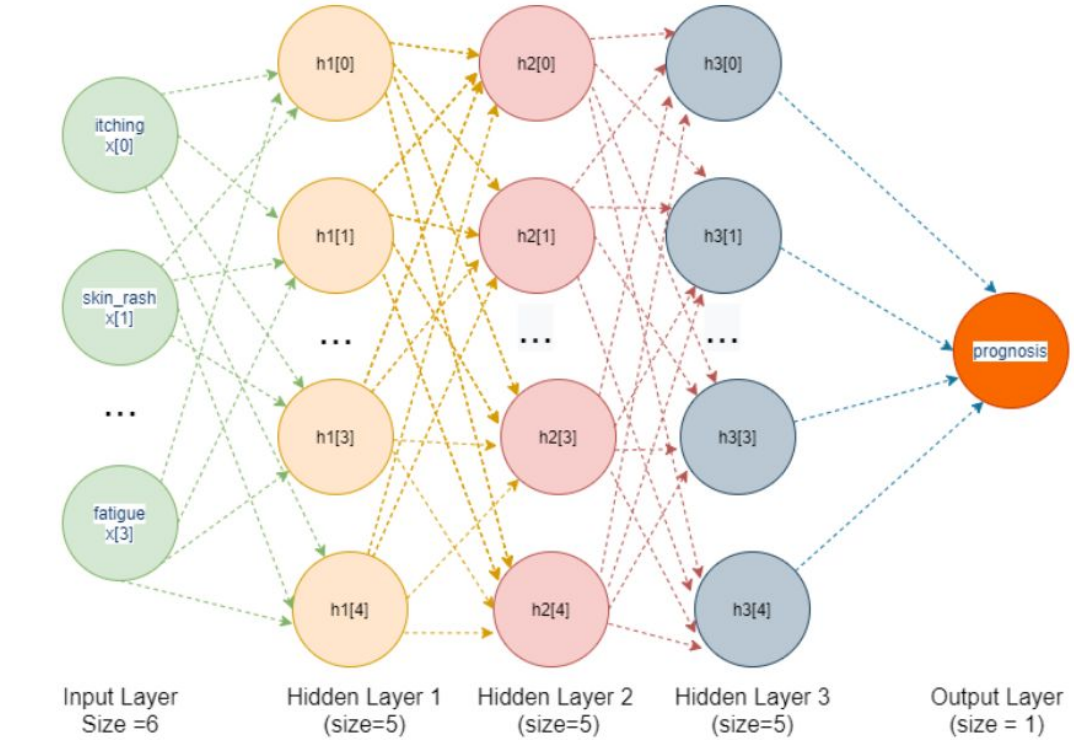
Problem Description:

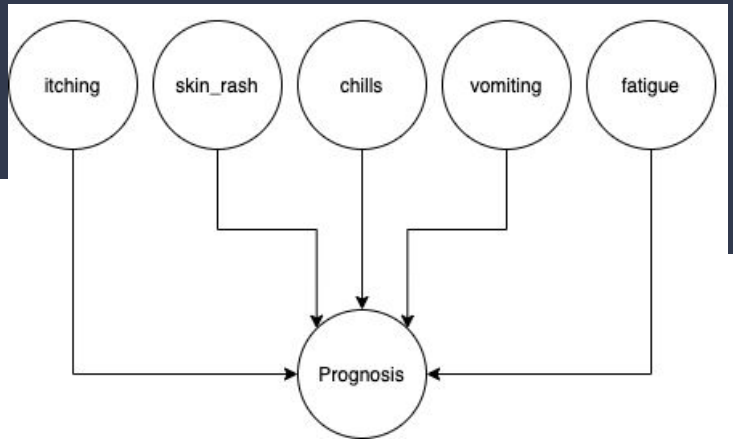
To prevent the potential situation where someone ends up convincing themselves they have an unlikely illness searching online, our team created an application for self-diagnosis of illnesses that would provide legitimate suggestions for the mystery illness without the user needing to research alone.

Objectives:

1. Predict illness based on the user's input symptoms without any complications.
2. Reliably predict the illness from the given data and user input and should not give random result for the same input model.
3. It should be desirable to use over manual investigation of illness for the non-software or medical person.

Implementations





Symptom I Probability			
itching	0.137804878	not_itching	0.862195122
skin_rash	0.1597560976	not_skin_rash	0.8402439024
chills	0.162195122	not_chills	0.837804878
vomiting	0.3890243902	not_vomiting	0.6109756098
fatigue	0.3926829268	not_fatigue	0.6073170732

ANN using `sk-learn` packages

Bayesian Network using `pomegranate`

Criterion for Comparison	Preferred Algorithm Given Criteria	Bayesian Network (Using `pomegranate` package)	ANN (Using `sk-learn` packages)
Full runtime (See Appendix A)	Bayes's Net, faster time	0.1044455678 seconds (from samples)	3.2411391276 seconds
Runtime to generate model (See Appendix A)	Bayes's Net, faster time	0.0436754752 seconds (from samples)	2.685259133 se
Understandability	Bayes's Net, because of it's simple, graphical nature	Bayes's Nets are comparatively simple to understand as the graphical format is straightforward and the idea that the distribution tables are tied to nodes made it easier to translate to a program	ANN was comp confusing: choc many layers, nc iterations would needed was not straightforward if our implemer even the optima for these feature
Implementation Complexity	Bayes's Net, because examples online could be used almost directly for our purposes despite the difficulty looking at deeper package methods	Fairly simple to start with using the online examples, however the `pomegranate` package is not very well defined and the input/output provided by package objects and their functions was difficult to understand and use	Finding the right package modules for training was difficult because no online examples were using the latest python version. `sk-learn` had very good documentation once the right modules were identified
Dataset Complexity	ANN, because the structure of the ANN can be taken directly from dataset header data	It is difficult to prepare the dataset manually for the algorithm, we must know the format of data and program for it. It is possible for the package to figure out it's own pattern but it is usually not correct, hence we have to explicitly define the graph nodes and edges	Have to explicitly know the structure of the ANN in order to program it from a known dataset. Similar to Bayes's Net
Accuracy (of package used)	ANN, because it is easier in our case to obtain more data than it is to obtain known, accurate data	Depending on data used, if the dataset is accurate the Bayes's Net's predictive accuracy will be on par at least	More data means more accuracy: the number of data attributes is directly proportional to accuracy

Bayes's Net is the preferred solution
considering criteria for comparison

Results

References

Images in slides made by Medi-App Team, references below for spoken presentation

Anonymous (Nirma University) (2020). Disease Prediction Using Machine Learning (Version 1) [Data set]. Kaggle. https://www.kaggle.com/kaushil268/disease-prediction-using-machine-learning

Das, Balaram. (2008). Generating Conditional Probabilities for Bayesian Networks: Easing the Knowledge Acquisition Problem. https://arxiv.org/abs/cs/0411034

Schreiber, J. M. (2018). *pomegranate* (Version 0.13.2). Schreiber, J. M.. https://pomegranate.readthedocs.io/en/latest/