MACHINE LEARNING APPLICATION IN MEDICINE

Project Group ID: 9

Guided by Dr. Suganya E.

Harshil (20BCE10071)
Anubhav De (20BCE10085)
Ashish Kumar (20BCE10163)
Atharva Damle (20BCE10181)
Shubham Sangwan (20BCE10519)

ABSTRACT

In the modern world, humanity has witnessed success on a wide range of predictive tasks encompassing Artificial Intelligence. So much so that, machine learning techniques are garnering substantial interest from researchers and scientists throughout the world. The research paper focuses on one of the applications of this advancing technology that is in the field of medicine.

Multivariate Linear regression and Support Vector Machine have been used in order, to predict whether a person has heart disease or not by inputs such as pulse rate, blood sugar level, and many other features. In order to increase the accuracy of the model, preprocessing techniques like outlier removal, handling Null values, etc have been used. Initially, the model only allowed numeric inputs for prediction, it could be diversified by using Classification techniques like Logistic Regression, Naïve Bayes, K-nearest Neighbor, Decision Tree, SVM in order to classify the patients' disease, be it cardiac problems, cancer, viral, and many more.

NLP and ML are both parts of AI. NLP enables machines to not only read but interpret human language and voice.

INTRODUCTION

Heart Diseases are one of most prominent disease globally and is the cause for more than 18 million fatalities every year. World Health Organization (WHO) predicts that fatalities due to this disease will increase by upto 24.5 million in 2030. Heart diseases mainly involve diseases of the guts and blood vessels/artery. The ability to identify heart illnesses in a timely manner can aid in making judgments about patients, lowering their risks.

Early detection is critical in diseases like such and this is where Machine Learning Algorithms can be applied to a large extent.

This project presents a prediction model that uses entered symptoms in a webform to forecast whether a patient has a heart illness or not.

It predicts the patient's ailment based on the data or symptoms submitted into the online system and provides findings based on that data. The health industry can benefit from this technique by simply asking the user for symptoms and entering them into the system, and in a matter of seconds be able to diagnose the patient and advice the needed medical care.

MODULES AND METHODS

The prediction model made use of mainly two algorithms (Logistic Regression and KNN).

In statistics, the logistic model is a statistical model that models the probability of an event taking place by having the log-odds for the event be a linear combination of one or more independent variables. In regression analysis, logistic regression estimating the parameters of a logistic model.

In statistics, the K-nearest neighbors algorithm is a non-parametric supervised learning method used for classification and regression. The input consists of the k closest training examples in a dataset.

The Modules include a variety of designs that we have implemented in our system to predict disease using machine learning.

1. Entering Symptoms/Information: Once user successfully open the system then he/she has to enter the necessary inputs.

2. Disease Prediction: The predictive model predicts the disease of a person and gives output to the user in maybe tabular form.

Through calculations, it was determined that KNN algorithm performed better on the test variables and was hence used in this prediction model.

RESULTS

The developed Machine Learning model provided a platform to solve a large problem of administering heart diseases. This Machine Learning model made use of available Python based libraries and mathematical algorithms.

For linking the front-end part of our application with the model, we made use of the python-based framework Flask.

During the development process we found that KNN algorithm worked better and provided higher accuracy rate compared to Logistic Regression (as observed in figure 3&4)and hence was used in this model. The value of K in our model is 2 and thus, the classification is done based on the two nearest neighbours of a new data.

The model uses a train-test split of 80:20 out of a csv file containing test results of more that 1000 patients. With our project being dynamic in nature, the scope for future enhancements is vast .
(Using NLP we can develop a more user interactive website, User login/signup, Prediction of preventive measures, etc)

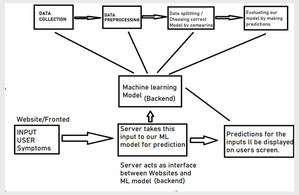

Figure 1. System Architecture

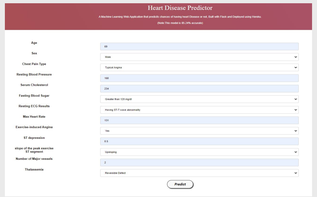

Figure 2. Website graphical user interface.

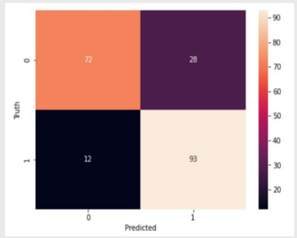

Figure 3. Accuracy of Logistic Regression

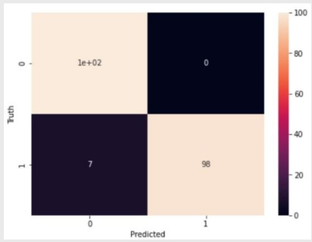

Figure 4. Accuracy of KNN Algorithm

MODELS AND EXPLANATION

Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome, something that can take two values such as true/false, yes/no and so on. The mathematics of logistic regression rely on the concept of "odds" of the event, which is the probability of an event occurring divided by the probability of an event not occurring.
Logistic regression is a transformation of a linear regression using sigmoid function. The vertical axis stands for the probability for a given classification and the horizontal axis is the value of x. It takes a linear combination of features and applies to them a nonlinear sigmoid function.
KNN (K-Nearest Neighbour) is a non-parametric algorithm that assumes the similarity between the new data and available data and puts the new case into the category that is most similar to the available categories. This means that when a new data appears, the prediction can be made by classifying that data into a well suited category through the KNN Algorithm.
The basic nearest neighbour algorithm makes classification predictions or regression predictions for an arbitrary instance where it identifies a training instance closest to the arbitrary instance. KNN algorithm expands this process by using a specified number K>=1 of the closest training instance.

CONCLUSIONS

Heart diseases continue to be a cause of concern for the population of this world and with vascular specialists being finite in number especially in developing countries, the ML model developed will help doctors administer such diseases at an early stage and save countless lives.

This developed model can also help people self-administer in case they have certain symptoms.

REFERENCES

- https://jpinfotech.org/prediction-of-heart-disease-using-machine-learning-algorithms/
- https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0244-x
- https://www.sciencedirect.com/science/article/abs/pii/S1386505614000190
- Santhana Krishnan.J, Geetha.S, "Prediction of Heart Disease Using Machine Learning Algorithms", IEEE 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), 2019.