Disease Prediction Using Machine Learning

PRIYANKA J. PANCHAL[1], SHAEEZAH A. MHASKAR[2], TEJAL S. ZIMAN[3]

[1, 2, 3] *Finolex Academy of Management and Technology*

Abstract- In this digital world, most of the people are prone to diseases, due to lack of healthy food, proper sleep and daily exercise. It is very crucial to know if we are suffering from a disease, at an early stage rather than discovering it at a later stage. Hence disease prediction system plays an important role as it predicts the diseases based on symptoms. This disease prediction system uses Machine Learning algorithm named Random Forest. This system also suggests drugs that are most commonly used to cure the disease.

Indexed Terms- Diseases, Drugs, Machine Learning, Prediction, And Random Forest.

I. INTRODUCTION

With the advancement in technology, Machine Learning is becoming more popular and commonly used technology by industry experts for solving problems faced in real life. Machine Learning is the scientific study of algorithms and statistical models that computer use to perform a specific task without using explicit instructions, relying on patterns and inference instead. Machine Learning is also used by the healthcare industry to bring advancement in their techniques so that they can provide better services to their patients. The disease prediction system predicts diseases based on patient's symptoms and also some commonly prescribed medicines for a particular disease.

II. EXISTING SYSTEM

There are many prevalent systems used for disease prediction. The existing systems only predict the diseases. The various approaches used for predicting diseases is by using Machine Algorithms such as Naïve Bayes, Decision Tree, Random Forest, k-mean algorithm. Also, one of the approaches to build a disease prediction system is by using Big Data. Prediction using traditional disease risk model usually involves Machine Learning and supervised learning algorithm which uses training data with the labels for the training of the models.

III. PROPOSED SYSTEM

In the proposed system, a disease prediction model is built using a Machine Learning algorithm that is Random Forest Algorithm. Based on the symptoms that are input by the user, the disease is predicted and the drug that is most commonly prescribed by the doctor is suggested.

IV. RANDOM FOREST ALGORITHM

Random Forest is essentially a collection of Decision trees. In this algorithm, firstly the random samples are selected from a given dataset. Then a decision tree is constructed for every dataset. Then the prediction result is obtained from every dataset. Voting is performed for every predicted result. Finally the most voted prediction result is selected as final prediction result.

V. STEPS TO DEVELOP THE PROJECT

A. Analyzing the problem statement & requirements
Analyze the problem in terms of what we want to predict and what kind of observation data we have to make those predictions. Predictions are generally a label or a target answer; it may be a yes/no label (binary classification) or a category (multiclass classification) or a real number (regression).

B. Collect and clean the data
Identify what kind of historical data we have for prediction modeling, the next step is to collect the data from datasets or from any other data sources.

C. Prepare data for ML application
Transform the data in the form that the Machine Learning system can understand.

D. Prepare the Graphical User Interface (GUI) of the model

Graphical User Interface (GUI) is designed for taking input and displaying output. There are 5 input text boxes which consist of dropdown menu of symptoms and the user can select those one by one. Python Tkinter package is used for designing the GUI. On pressing the 'Result' button, the disease is predicted in the output field. Also, the drugs are described in the specified field.

E. Train the model

Before training the model, it is essential to split the data into training and evaluation sets, as we need to monitor how well a model generalizes to unseen data. Now, the algorithm will learn the pattern and mapping between the feature and the label.

F. Evaluate and improve model accuracy

Accuracy is a measure to know how well or bad a model is doing on an unseen validation set. Based on the current learning, evaluate the model on validation sets.

G. Test the model

Test the model on unknown data.
After the system starts working properly, the model is complete.

CONCLUSION

In this paper, algorithm used to predict the disease based on symptoms is discussed. Various symptoms are provided in the dropdown menu, out of which user selects any five of them and using algorithm the disease is predicted. The drugs that are commonly prescribed for a particular disease can also be suggested in this system.

The main aim is to predict the disease at the early stage and lead to early diagnosis. This system can also be used by doctors to avoid confusion while predicting the disease. This system can provide assistance to doctors.

ACKNOWLEDGMENT

REFERENCES

[1] Arthur Samuel, *Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming*. Artificial Intelligence in Design '96. Springer, Dordrecht.

[2] A. C. Jamgade, Prof. S. D. Zade (May 2019). Disease prediction using Machine Learning. *International Research Journal of Engineering and Technology (IRJET)*. [Online]. Available: https://www.irjet.net/archives/V6/i5/IRJET-V6I5977.pdf

[3] A. Borad (October 2018). How to develop Machine Learning applications for Business. *eInfochips- An Arrow Company* [Online]. Available:https://www.einfochips.com/blog/how-to-develop-machine-learning-applications-for-business/

[4] *Classification Algorithms – Random Forest.* Tutorialspoint [Online]. Available: https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm