Disease Prediction using Machine Learning

# L.N.V .Mukesh1,Masineni Krishna Sai2, S.Rakesh3, S.V Tharun Kumar Reddy4, Dr.R.Venkatesan5

*1*

----------------------------------------------------------------------\*\*\*---------------------------------------------------------------------

**ABSTRACT**

Disease Prediction system is based on the disease of the user on the basis of the symptoms that user provides as an input to the system. If the patient isn't in any danger and the user merely wants to know what kind of ailment he or she has had. It is a system that gives the user suggestions and methods on how to keep their health system in good shape, as well as a way to find out if they have a sickness utilising this forecast. Decision tree Classifier and Naïve Bayes Classifier calculates the probability of the disease. For the prevention and treatment of illness, accurate and timely investigation of any health-related problem is critical. The development of a medical diagnosis system based on machine learning (ML) algorithms for disease prediction can aid in a more accurate diagnosis than the current method. With the growth of big data in the biomedical and health-care communities, accurate medical data analysis promotes early disease identification and patient care.

***Key Words*:** Machine Learning, Symptoms based disease prediction, Python.

1. **INTRODUCTION**

Medicine and healthcare are some of the most crucial parts of the economy and human life. There is a tremendous amount of change in the world we are living in now and the world that existed a few weeks back. Everything has turned gruesome and divergent. In this situation, where everything has turned virtual , the doctors and nurses are putting up maximum efforts to save people’s lives even if they have to danger their own. There are also some remote villages which lack medical facilities. Virtual doctors are board-certified doctors who choose to practice online via video and phone appointments, rather than in-person appointments but this is not possible in the case of emergency. Machines are always considered better than humans as, without any human error, they can perform tasks more efficiently and with a consistent level of accuracy. A disease predictor can be called a virtual doctor, which can predict the disease of any patient without any human error.

We're using machine learning to keep track of all of the hospital's data regarding the disease. Machine learning technology assists doctors to make better decisions about patient diagnoses and treatment options by allowing them to construct models swiftly evaluate data and deliver outcomes faster. The most prominent example of machine learning in the medical industry is healthcare.

According to research,70 percent of people in India suffer from general disease, and 25 percent die as result of early ignorance.

The main reason for developing this project is so that a user can sit at their convenience and have check-up. The UI is designed in such a simple way that everyone can easily operate on it and can have a check-up.

**2.LITERATURE SURVEY**

According to Tom Mitchell, "a computer programme is said to learn from experience with experience." Machine learning is a combination of correlations and relationships, and the majority of machine learning algorithms exist to uncover and/or exploit interactions between datasets. When Machine Learning Algorithms are able to detect specific connections, the model can either use these links to forecast future observations or generalise the data to highlight intriguing patterns. Regression, Linear Regression, Logistic Regression, Naive Bayes Classifier, Bayes theory, KNN (K-Nearest Neighbor Classifier), Decision Tress, Entropy, ID3, SVM (Support Vector Machines), K-means Algorithm, Random Forest, and others are some of the methods used in Machine Learning.

Arthur Samuel created the term "machine learning" in 1959. Machine learning entails the research and development of algorithms that can learn from and predict data. Computational statistics, which similarly focuses on making predictions using computers, is closely connected to (and frequently overlaps with) machine learning. It is closely linked to mathematical optimization, which provides the discipline with tools, theory, and application domains. Machine learning and data mining are sometimes confused, however the latter subject, known as unsupervised learning, focuses more on exploratory data processing.

Machine learning is a way for developing complicated models and algorithms that lend themselves to prediction in the field of data analytics; in commercial use, this is known as predictive analytics. Through learning from previous linkages and trends in the data, these analytical models enable researchers, data scientists, engineers, and analysts to "create dependable, repeatable judgments and results" and find "hidden insights."

**3.METHODOLOGY**

From an open-source dataset, an excel sheet was created where we listed down all the symptoms for the respective diseases. After which depending on the diseases were specified as a part of the dataset. We listed down around 43 diseases with more

than 100 unique symptoms in all. The symptoms of an individual were used as input to various machine learning algorithms.

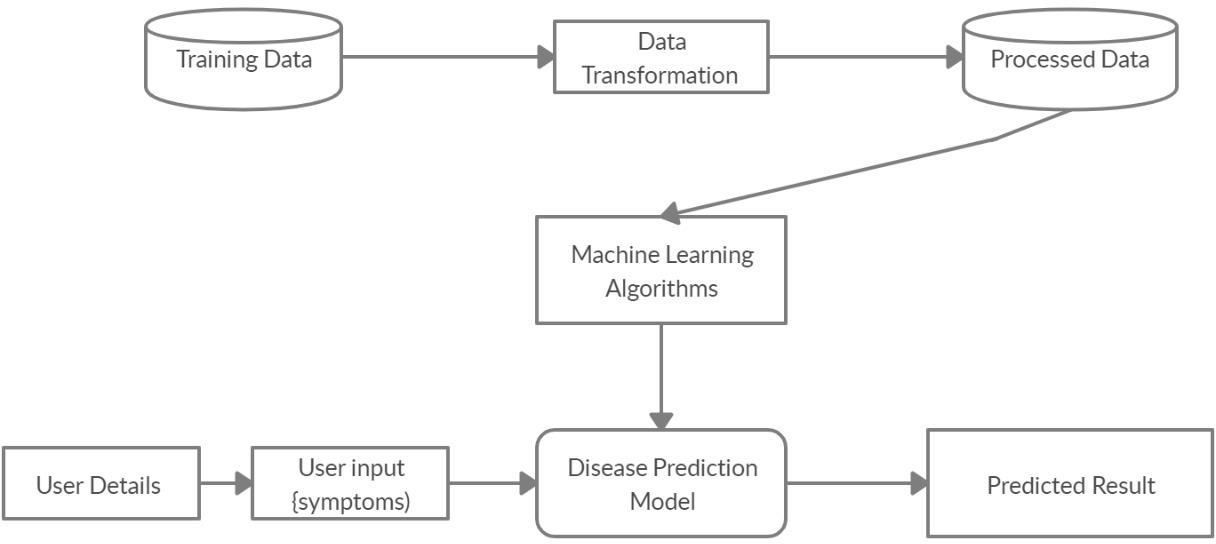


Figure 1:- System Architecture

User Input: When designing the algorithm, we assumed that the client would have a clear idea about the signs he was seeing. The produced prediction takes into account 95 manifestations, during which the customer might allow the indications from his preparation to be used as input.

Data Processing: Information pre-processing is a method of data mining that transforms raw data or encrypts it again to build a structure that can be decrypted with the use of calculations.

Disease Prediction models: The entire system is designed in such a way that it can predict diseases using three algorithms: Decision Tree and Naïve Bayes classifier models, so that a predictive analysis study can be proposed at the end of the research by examining the speed, efficiency, and performance of the various algorithms for the input dataset.

Predicted output: When a framework is built with the preparation set and validated calculations, standard datasets are created, and when the client indications are provided as an input to the algorithm, and the side effects are composed in accordance with the standard dataset created, arrangements are made and the high probability infection is predicted.

**Tkinter**

Tkinter is a Python binding for the Tk graphical user interface toolkit. It is Python's de facto standard GUI and the primary Python interface to the Tk GUI toolkit. Tkinter is standard with Python installations on Linux, Microsoft Windows, and Mac OS X. Tkinter is derived from the Tk interface. Fredrik Lundh is the author of Tkinter. Tkinter is open-source software distributed under the Python license.

Tkinter is constructed as a Python wrapper around a complete Tool Command Language (TCL) interpreter contained in the Python interpreter, similar to most other current Tk bindings. Tkinter calls are converted into Tcl commands, which are then fed into this embedded interpreter, allowing Python and TCL to

coexist in a single application. The Frame widget in Tkinter is the basic organisational element for complicated layouts. A frame is a rectangular region in which additional widgets can be placed. A parent-child relationship is generated when any widget is created. If you put a text label inside a frame, for example, the frame is the label's parent.

Python provides a variety of GUI development possibilities (Graphical User Interface). Tkinter is the most commonly used of all the GUI techniques. It's a standard Python interface to the Python-supplied Tk GUI toolkit. The fastest and easiest approach to construct GUI applications is with Python with tkinter.

**4.IMPLEMENTATION**

The Disease forecast framework is executed utilizing the two information mining calculations for example Decision tree classifier and Naive Bayes Theorem. The portrayal and implementation of the calculations are provided beneath.

**a)Decision Tree**

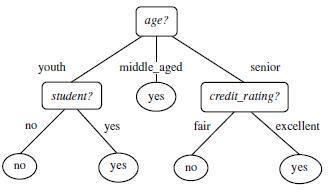
The decision tree learning algorithm is similar to a decision tree algorithm in that it maps input about an object to the item's output. The tree models with finite output classes are referred to as Trees of classification The leaves of these tree structures display class. Labels and branches depict the relationship between attributes.as a result of the system's class labels.

A decision Node and a Leaf Node are the most important parts of a tree.

Decision Node: There are at least two branches in a decision node. Every one of the manifestations is treated as a decision node in our analysis.

Leaf Node: Makes up the order, indicating that the decision can

come from any of the branches. As a result, each illness represents a leaf node.

 Figure 2:- Decision tree sample

**b)Naïve Bayes Classifier**

The Naive Bayes algorithm represents the supervised machine learning method of classification. It employs a probabilistic model in which the outcomes/outputs are assigned probabilities..

It's utilised to solve challenges that are both analytical and predictive. The Bayes algorithm is resistant to noise in input dataset.



Figure 3. Naïve Bayes Probability

The probability of hypothesis h given the data d is P(c|x). The posterior probability is the term for this. The chance of data d being true if hypothesis h is true is P (x|c). P(c) is the chance that hypothesis h is correct (regardless of the data). This is referred to as the h prior probability. The probability of the data is P(x) (regardless of the hypothesis.

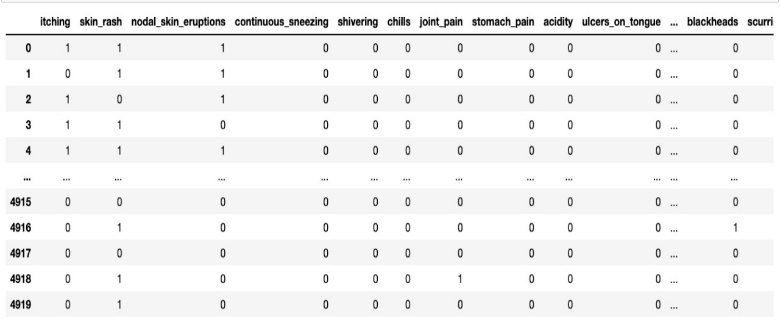


Figure 4. Data Set

**5.RESULTS**

For the supplied input dataset, various machine learning models were applied to investigate disease prediction. For the prediction, we used two machine learning models. The accuracy varied depending on the size of our dataset, i.e. for the training set, it was small and huge. In comparison to the other ML algorithms, it showed to be the best accurate model due to this variation. We collected raw data and separated it into categories based on symptoms.

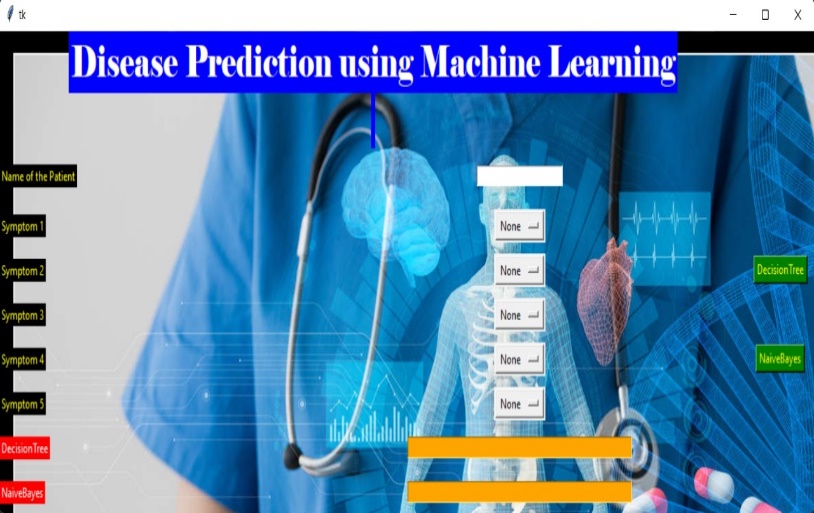
****

Figure 5. GUI Output

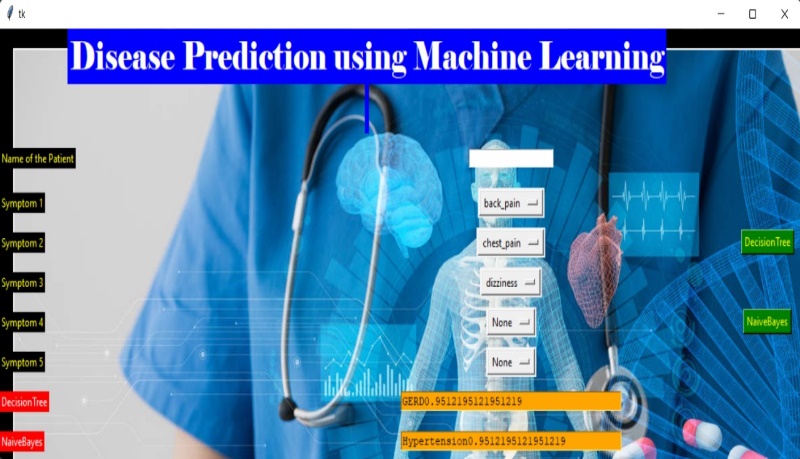
****

Figure 6. Predicted Output

The above figures represents the GUI output and the predicted output from the user inputs. Accuracy can be shown in the output that is calculated from the machine learning algorithms.

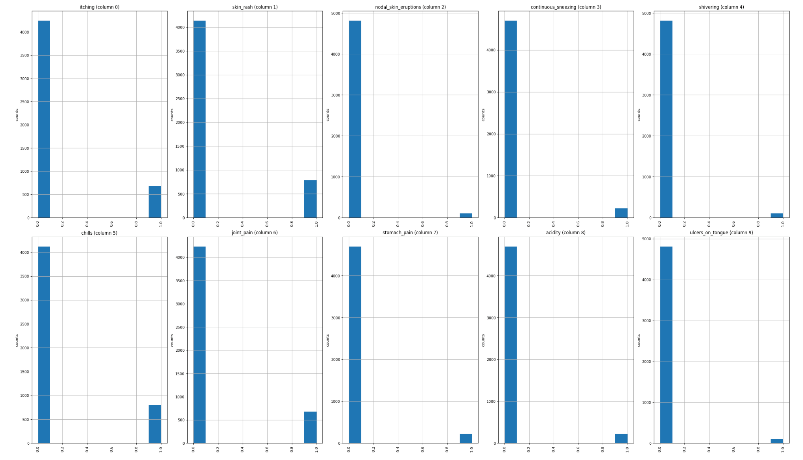


Figure 7. Data Analysis Graph

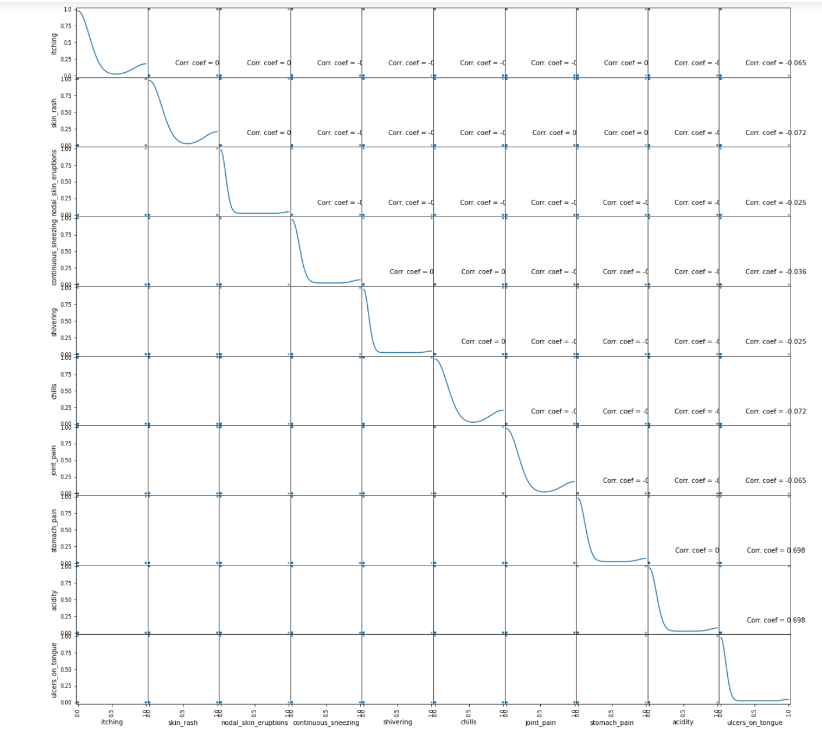


Figure 8. Correlation Graph

|  |  |
| --- | --- |
| Algorithm | Accuracy |
| Decision Tree | 0.95121 |
| Naïve Bayes | 0.95121 |

Table 1. Accuracy

**6. CONCLUSION AND FUTURE SCOPE**

Finally, based on the documented expansion of (ML) Machine Learning techniques and approaches in the clinical field, it is highly likely that frameworks and systems have emerged that have enabled improved data exploration through the fundamental and direct use of Machine Learning (ML) models. This research presents a close examination of three models' execution of a clinical record, with each model achieving a 95 percent accuracy score. Finally, the paper's accuracy value and disorder lattice are explored. Because of the accessibility of massive data created and stored by cutting-edge innovation, artificial consciousness will play a far larger role in information research in the future.

The focus of future work will be on providing medical assistance and suitable medication to patients as soon as feasible in order to construct the best infrastructure and the quickest and easiest way possible in the medical sectors.

**ACKNOWLEDGMENT**

This work are often utilized in Medical field and on basic ground level in all parts of world. Thus strengthening the Medical and Healthcare Industries. Authors are thankful to Faculty Dr. R. Venkatesan Associate Professor , Karunya

University.

**REFERENCES**

[1] Disease Prediction and Doctor Recommendation System by [www.irjet.net](http://www.irjet.net/).

[2] Disease Prediction Based on Prior Knowledge by the

Journal [www.hcup-us.ahrq.gov/nisoverview.js](http://www.hcup-us.ahrq.gov/nisoverview.js).

[3] Disease Prediction Using Machine Learning by the International Research Journal of Engineering and Technology (IRJET).

[4] Machine Learning: A multistrategy approach by Tom M.Mitchell.

[5] Arthur Samuel : Wikipedia

[6] [https:/en.wikipedia.org/machine\_learning\_algorithms](https://www.wikipedia.com/machine-learning-algorithms)

[7] A. N. Repaka, S. D. Ravikanti and R. G. Franklin, "Design And Implementing Heart Disease Prediction Using Naives Bayesian," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 292-297, doi: 10.1109/ICOEI.2019.8862604.

[8] <https://www.researchgate.net/publication/325116774_disease_prediction_usin>[g\_machine\_learning\_techniques](https://www.researchgate.net/publication/325116774_disease_prediction_using_machine_learning_techniques).

[9] <https://wiki.python.org/TkInter>

[10] A.S. Monto, S. Gravenstein, M. Elliott, M. Colopy,

J. Schweinle, Clinical signs and symptoms predicting in-

uenza infection, Archives of internal medicine 160(21),

3243 (2000).