Prognosis using Data Analytics: The Proposed System
Prof. Indu Dokare¹, Neha Bhagchandani², Jayesh Ahuja³, Manish Manghwani⁴, Bhavesh Mohinani⁵
Department of Computer Engineering

Abstract:
In past 5-10 years adapting with westernized lifestyle, fatal diseases such as Heart Disease, Lung Cancer had become the key reason of death in India. Today, Hospitals store patient records so that when patients comes back for treatment, hospital staff should know that from which disease patient was suffering previously and how their treatment should be taken forward. This way, there is huge amount of data getting generated that could be used efficiently for prediction of a disease of a new patient based on prediction analysis on the old patients. Hence, we present a prediction system to predict the disease and hence find ways to prevent it by checking patient’s lab reports.

Keywords: Data Analytics, Disease Prediction, Healthcare, Machine Learning

I. INTRODUCTION
A very concerning issue in the healthcare organizations is the unavailability of resources with required quality and time-delay issues. The clinical conditions are very poor. Any ambiguity regarding tests if results in increased number of tests, adds to the cost as well as increase the workload of hospitals [3]. The doctors may encounter many patients in his/her experience but it is very difficult to remember all the cases [4]. Current database system in hospital contains huge amount of data which is rarely used. This gives rise to the need of an intelligent system for doctor’s assistance. Now, since the records are being stored in the hospitals, this experience in the form of trained model could be made useful to doctors by designing an application that assists the doctors while the new patients comes. The focus of the proposed system is to make an application system assisting the doctors to predict the probability of occurrence of a disease for a new patient based on the lab reports of the patients.[3]

II. LITERATURE SURVEY
[1] Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction by Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni :

It provides a good comparison between the machine learning algorithms used for the prediction of the heart disease. The outcome of predictive data mining technique reveals that Decision Tree & Bayesian classification performs well as compared to other algorithms like KNN, Neural Networks, Classification based on clustering, etc. Decision Tree is a popular classifier which is simple and easy to implement. It requires no domain knowledge or parameter setting and can handle high dimensional data. The results obtained from Decision Trees are easier to read and interpret.

[2] A Survey on Data-Mining Technologies for Prediction and Diagnosis of Diabetes by Dr.B.L. Shivakumar and S.A. Iby:

This paper provides a good overview on the work that has been done on the prediction of the diabetes. Mechanical life style and obesity have contributed to the sudden rise of type 2 diabetes worldwide. It is evident that the occurrence of diabetes is having a strong relation with
• Diseases like Wheeze, Edema, Oral diseases
• Female pregnant
• Increase of age.

III. METHODOLOGIES
Machine learning lies at the intersection of computer science, engineering and statistics and often appears in other disciplines. Machine learning is categorised in to supervised learning and unsupervised learning. The first category, supervised learning is training the model when it is already known what is to be predicted. The supervised learning can be done using classification or regression analysis. An algorithm which comes under supervised learning is Naive Bayes, Decision tree, k-Nearest Neighbours, etc. The second category under machine learning is unsupervised learning. In unsupervised learning we do not know what value is to be predicted, i.e. we do not have label in unsupervised learning. The similar data-items are grouped together to form clusters. This is known as clustering. The unsupervised learning could be used when we have to find a statistical value corresponding to the data. This is known as density estimation. Algorithms which come under unsupervised learning are k-means, DBSCAN. The proposed system falls under supervised learning. [10]

Some of the algorithms of supervised learning:

Naive Bayes:
Naive Bayes classifier is a probabilistic classifier. This classifier is based on the assumptions that each attribute of the dataset is not dependent to any other attribute. Naive Bayes is a very simple classifier. [6]

Decision tree:
Decision tree is a tree-like structure. It is an inverted tree. Decision tree recursively splits the attributes based on their values. This is known as recursive partitioning. The
classification is done to predict the value of one attribute based on values of other attributes. [7]

IV. PROPOSED WORK

The proposed system contains following modules:

1. Collection of data: Data will be collected in the form of lab reports (for eg. Heart Disease)

2. Data preprocessing: Often data in Real world are incomplete, noisy or inconsistent. Preprocessing may consist of following steps:

   I. Data Cleaning
   II. Data Integration
   III. Data Transformation
   IV. Data Reduction
   V. Data Discretization

   Data is analysed and preprocessed by the above listed methods as per the requirement.

3. Training Module: In training module, Prediction model is generated by using supervised algorithms such as Decision Tree, Naive Bayes, etc. Based on the efficiency of algorithm, final Prediction module is generated.

4. Prediction model: It is used to predict the outcome of a disease for every new lab report data.

5. Visualisation module: The outcome of Prediction Module is visualised as in the form of probability of occurrence of a disease.

V. CONCLUSION

The focus is to design a predictive model for disease detection using Machine Learning techniques from data sets that is capable of enhancing the reliability of disease diagnosis. Machine Learning Algorithms has the potential to transform and revolutionize the way healthcare systems use technologies to gain valuable insight from the data repositories. In the future it is sure there will be widespread use of data mining algorithms across the different areas of healthcare industry.

VI. REFERENCES


[4]. https://www.elsevier.com/connect/seven-ways-predictive-analytics-can-improve-healthcare


[7]. http://docs.rapidminer.com/studio/operators/modeling/predictive/trees/parallel_decision_tree.html

