**IT WORKSHOP PROJECT**

**SYNOPSIS**

**GROUP 7**

Submitted To-

Ms. Garima Jaiswal

Group Members-

023- Charu Singh

030- Jhanvee Khola

051- Princy Singhal

052- Khushi Punia

B.Tech CSAI - 2nd Semester

****

**Disease Prediction and Detection System**

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | |  |

Abstract

|  |  |
| --- | --- |
| * Due to modern lifestyle, diseases are increasing at an alarming rate. Our lifestyle, work type and food habits are the various factors which influence our health causing heart diseases and other health issues like Hypertension, Diabetes etc. In emerging countries like India, the urban population faces the wrath of it with 5.8 million deaths per year. * Data science and AI systems can be used for prediction and detection of such lethal diseases which will in turn help to take early actions against them. * Data mining technique is one of the most challenging and leading research areas in healthcare due to the high importance of valuable data. The recent blooming in the data mining approaches has provided a solid platform for various applications in the healthcare field through a variety of algorithm approaches that have been utilized for disease prediction. | |
| ***Introduction*** |  |
|  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | |  | |
| In our project the summarization of the different algorithms of data mining are used in the field of medical prediction. | The main focus is on using different algorithms and combination of several target attributes for different types of disease prediction using data mining i.e. The process of identifying commercially useful patterns or relationships in databases or other computer repositories through the use of advanced statistical tools. It is defined as sifting through very large amounts of data for useful information.  Various datasets are used to predict/detect whether a patient is likely to get a disease based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relevant information about the patient and this data is then acted upon with important techniques like categorization and preprocessing to extract necessary information. | |
|  | | |  | |

|  |
| --- |
| ***Here are the main steps involved in the making of the prediction and detection system: -***   1. *Data collection: Datasets with numerous attributes and patient data, containing valuable information regarding a disease, are collected through authentic sources. These datasets give a number of variables along with a target condition of having or not having a particular disease. The data is investigated using ML explainability tools and techniques.*     *2.  Data Analysis and Cleaning: This is in general looking at the data to*  *figure out what's going on. Inspect the data: Check whether there is*  *any missing data, irrelevant data and do a clean up.*     1. *Data Visualization: Plotting various functions in the dataset to visualise and understand the data better.*      1. *Feature selection: Selecting necessary features for classification or constructing new attributes from the given set of attributes to help the mining process.* 2. *Searching for trends, relations & correlations: Finding common trends or relations between different attributes. Combining statistical analysis, ML & database technology to extract hidden patterns and relationships from*   *Databases.*     1. *Drawing final inference: Using the extracted and filtered information, predict whether the patient can be identified to be having the mentioned disease or not.* 2. *Performance Analysis: Analysing and then enhancing the accuracy of the trained model.*    Literature Survey/ Work done so farResearch Papers and Surveys done so far-  * ”Disease prediction using principal of component analysis” in  2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC). * ”Disease Prediction in Data Mining Technique – A Survey” by S.Vijiyarani and S.Sudha * A Survey on Disease Prediction by Machine Learning over Big Data from Healthcare Communities   International Journals-   * ”International Journal of Computer Applications & Information Technology” * [2017 International Conference On Smart Technologies For Smart Nation (SmartTechCo](https://ieeexplore.ieee.org/xpl/conhome/8356464/proceeding)n) * IOSR Journal of Engineering (IOSRJEN) * International Journal of Computer Trends and Technology (IJCTT) * International Research Journal of Engineering and Technology (IRJET)) |

# Applications made:

* A smartphone app called Healthians can provide smart reports from user inputs and helps you reach required treatment goals.
* Smart Health Care App: An Android App to detect disease based on symptoms.

##### Objective

* The aim of our project is to predict the diseases among the trained dataset using classification algorithms. We seek to promote effective andtimely decision making by timely prediction of disease.
* Sole purpose of the project is to minimise the attributes and enhance the precisionsof algorithms. Through Collected data and Medical data mining we wish to integrate it to form a hospital information system. This new technology should enable the innovation of trends and predictive patterns in data, the creation and testing of hypotheses and generation of insight-provoking visualizations.
* The concept of data mining helps the end users to extract useful information from large databases. Machine learning algorithms can also be helpful in providing vital statistics, real-time data, and advanced analytics in terms of the patient’s disease, lab test results, blood pressure, family history, clinical trial data, and more to doctors.

Scope of the work

* The implementation of artificial intelligence provides benefits in early detection by being able to pinpoint any risk alerts a patient may have. This alert will allow the practitioners to get patients to a MRI/CT scan sooner for a disease evaluation.
* It has become important to discover hidden patterns and relationships from medical databases. In classical clinical diagnosis, it requires lots of tests which could complicate the disease prediction. Hence the data mining techniques can help medical expertise to make the decision about the disease using computer aided decision support systems.
* In a study, the early detection alert provided 87.6% accuracy in a diagnosis and prognosis evaluation. That said, the practitioners will be able to implement treatment sooner and predict whether the patient will have a higher possibility of future disease similar to that. Previously, it was challenging for healthcare professionals to collect and analyse the huge volume of data for effective predictions and treatments since there were no technologies or tools available. Now, with machine learning, it’s been relatively easy, as big data technologies are mature enough for wide-scale adoption.
* For example, machine learning can be used in 48-hour post stroke patients gaining a perdition accuracy of 70% whether the patient may have another stroke or not. Many people's lives are cut short due to cancer. However, due to the age of big data we are able to combat this malicious disease.
* Also, patients with Liver disease have been continuously increasing because of excessive consumption of alcohol, inhalation of harmful gases, intake of contaminated food, pickles and drugs. So, patient datasets can be used to evaluate prediction algorithms in an effort to reduce burden on doctors.
* The model can be expanded so that it can use image recognition to detect plant disease in crop and thus help agriculture. This would help in prevention of lose of harvest due to plant diseases caused by pests.
* The system allows collection, maintenance and analysis of large amount of data which eases the process of research on these diseases and eases work of medical researchers.

### **Proposed methodology**

* Searching for the diseases : Choosing the datasets for those diseases which have life threatening characteristics and finding the efficiency of the proposed method in fruitful manner and some relativity is there between these datasets.
* Collecting datasets: Collecting datasets regarding patient information for diseases such as diabetes, coronary heart disease and breast cancer data as input. The data is collected from authentic sources and is selected so that it has the appropriate attributes with accurate patient data.

* Data Preprocessing: Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format. These data are loaded and checked and missing values are converted to null. True and false value is replaced to 1 and 0, respectively.

* Data Visualization: The data is plotted graphically to have a better understanding of the data.

# Algorithms used :-

1. Naive Bayes classification algorithm: Here Bayes theorem is used for classification purposes and to assume that classification is predictor independent. Naive Bayes model is compatible for very large datasets to build and for further analysis. This model is a very simple and sophisticated classification method, and it performed well even in complicated scenarios.
2. Random forest algorithm: The random forest (RF) is a hierarchical collection of tree structured base classifiers. Text data usually has many numbers of dimensions. The dataset contains a large number of irrelevant attributes. Only a few important attributes are informative for classifier model. RF algorithm uses a simple predetermined probability to select the most important relevant attribute. It has calculated the results accuracy, and this algorithm gives the three classes separately using a confusion matrix.

* Performance Analysis: To find the efficiency of the proposed method, the trained data are compared separately with the proposed algorithms and also checked the performance of test data. The proposed method is also applicable for testing the real-time disease data for classification and to identify whether the patient is affected by the particular disease or not.

|  |
| --- |
| THANK YOU. |