International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

Breast Cancer Prediction using multiple models based on Machine Learning and Deep Learning

Kaveri Bhusare[1],
Department of Computer Engineering,
Wagholi, Pune

Pratiksha Deshmukh[2]
Department of Computer Engineering,
Wagholi, Pune

Bhakti Bodkhe[3]
Department of Computer Engineering
Wagholi, Pune

Abstract: Now a day, machine learning algorithm is one of the important field in Healthcare sectors to provide higher accuracy of Diseases prediction and deep study of medical field data. Authors are used to various models of machine learning to predict the higher accuracy of patient diseases and medical test such as breast cancer prediction with higher Accuracy. We are studied various machine learning models. In this paper we are used and predict the breast cancer patient datasets from using CSV files. Machine learning models are very good techniques to predict diseases. In this paper machine learning classifier, Nearest neighbor training accuracy, SVM, SVM classifier, Gaussian naïve bayes, Decision tree and Random forest Algorithm these machine learning model we can use and analyze the overall performance of each model and predict best accuracy. The performance of all various machine learning models can be analyzed on Breast cancer datasets, after analyze of various machine learning models that gives better accuracy on breast cancer datasets.

Keywords: Machine Learning, Scikit-learn, supervised learning, SVM, DT, Random Forest Flask, bagging, boosting, etc.

INTRODUCTION

Breast cancer is second most dangerous cancer in the world, cancer that identified in breast called as Breast Cancer. There are many various type of cancer in the world. In the term of death ratio breast cancer is the second most dangerous cancer in world. In the human body when cell does not grow or out of control then starts cancer in human body. In breast cancer tumor usually form of cancer cells that out of control and felt as lump. Breast cancer is one of the cancers responsible for death in the world. According to NCRP of India, Thirteen thousand Indians die every day due to cancer. As per the report of world health organization (WHO), death rate of cancer in India is 79 per 100,000, while the number of death percentage in India is 6%. According to National Institute of Cancer Prevention and Research (NICPR), an overall 2.5 million people identified due to cancer in India and more than 1,157,294 new cancer patients are identified in India every year. As per the reports of 2018, 7, 48,821 people died due to cancer (out of that 4, 13,519 men and 3, 71,302 women) in India. According to NICPR report of projected cancer patients with cancer in India male cancer rate was 679,421(94.1 per 100, 000) and overall females was (712, 758 per 100, 000) for the year 2020.

Breast cancer is second most common type of cancer in women, according to reports 14% women are registered due to Breast Cancer in India. In India diagnose a breast cancer patient in every four minutes in both rural and urban area of India. As per the reports of Breast cancer 1, 62,468 are registered cases in India out of those 87,090 patients' deaths. As per the reports 14.5 lakh cases are expected and it will increase to 17.3 lakh patients in 2020.

Machine Learning algorithm has played an important role in the field of medical, healthcare sectors to predict disease with higher accuracy on large amount of datasets, large amount of data can be generates day by day in healthcare sectors. In the terms of patients several diseases reports and huge amount of information database however healthcare sectors, medical field these datasets can be widely distributed and used machine learning algorithm is a process of that mainly predicts the patient's dieses and report with higher accuracy. In this paper we review the related survey of various classifications, Naïve Bayes, nearest neighbor training accuracy, Decision Tree, SVM, SVM (RBF classifier), Random forest etc. Accuracy is the measured by using various parameters of machine learning algorithm. Hence we used python anaconda tools for machine learning model datasets. Anaconda Jupiter Notebook is very popular tools for ml models, we have used

anaconda Jupiter notebook interface for determining the accuracy with the use of training set testing data.

In the world near about 12% women are identified due to Breast Cancer, as increase the number of Breast Cancer patient in world day by day, when breast cancer identified in any person body, due to that death chances of that persons are very high. Here we have using machine learning algorithm in breast cancer prediction. Doctors are able to find out 70 to 80% patient disease is identified due to breast cancer and remaining are unable to predict. So here we Introduce Machine learning algorithm, we can solve this problem with higher Accuracy, we provide machine learning algorithm to solve this problem with higher Accuracy, we provide 100% information of patient to machine learning models and ML models can predict patient has disease or not due to breast cancer with higher Accuracy 95 to 99%. Machine learning model predict the best accuracy result of breast cancer patients, model can predict patient disease due to breast cancer or not, now for prediction purpose we are using machine learning classifier, Nearest Neighbour, SVM classifier, Gaussian Naïve Bayes, Algorithm, fuzzy clustering and classification and feature selection.

II- LITERATURE SURVEY

Many Researcher are using different techniques of machine learning on different datasets, predict and analyzing report. Discuss multiple machine learning model and multilayer perception on WEKA tool. This algorithm can use various datasets for evaluation and prediction purpose Author evaluates and analyze accuracy using different performance parameters.

[1][Kashishara shakil,et-al,2012] to disease diagnosis purpose discuss many data mining approaches used on WEKA tool. 10 crore fold validation used for testing datasets and analyze report. To compared classified dataset by different data mining algorithm and find out better approached from them**. [2][Peter Adebayo Idowu,2020]** developed a model for prediction and cancer disease using multiple machine learning model. Performance of model using different performance parameter with 10 fold cross validation, it is provide accuracy 96% and 97% respectively **[3][** S. **Muthuselvan,2016]** in this paper use different classification approaches to evaluate and analyze high risk patient. Knn and decision tree are used for disease prediction and calculating accuracy **[4]**

[Yingming Sun, 2018] in this paper discussed the risk of breast cancer and importance of machine learning model for prediction. Identify the disease in human body. These issues are analyzed by using machine learning model and calculating accuracy for breast cancer disease.

III – BACKGROUND

Breast Cancer Classification Breast cancer classification divides carcinoma into categories depending on how they have spread or if they have spread at all. Classification algorithms predict one or more discrete variables, supported the opposite attributes within the dataset. data processing software is required to run the classification algorithms. the aim of classification is to pick the simplest treatment. Classification is vital because it allows scientists to spot, group, and properly name organisms via a uniform system. Classification and clustering are two widely used methods in data processing. Clustering methods aim to extract information from a knowledge set to get groups or clusters and describe the info set itself. Classification, also referred to as supervised learning in machine learning, aims to classify unknown situations supported learning existing patterns and categories from the info set and subsequently predict future situations. The training set, which is employed to create the classifying structure, and therefore the test set, which tends to assess the classifier, are commonly mentioned in classification tasks Classification may be a quite complex optimisation problem. Many ML techniques are applied by researchers in solving this classification problem. The most famous algorithm that is used for breast cancer classification or prediction is an artificial neural network, random forest, support vector machine, etc. Scientists strive to seek out the simplest algorithm to realise the foremost accurate classification result, however, data of variable quality also will influence the classification result. Further, the rarity of knowledge will influence the number of algorithm applications also. If the carcinoma is found early, there are more treatment options and a far better chance for survival. Women whose carcinoma is detected at an early stage have a 93 percent or higher survival rate within the first five years. Getting checked regularly can put your mind comfortable. Finding cancer early can also save your life.Anaconda Jupiter notebook is used here to implementing the machine learning algorithm for predicting the disease in huge amount of data.

A. Jupiter Notebook

The aim of Jupiter notebook is to deliver complete group of analysis and implementation of machine learning algorithm for exclusively industrialist, research scholars and educationalist. Jupiter notebook tool is help to compare the various type of machine learning, artificial, data mining techniques. Jupiter notebook is very easy to use; understand because the tool was developed in

simple manner programming interface. Jupiter notebook supports the comma separated value (CSV) file format, .txt, .xls format. All of the above format can be downloaded from different web application or URL's from any web page, and also provided excel format can be converted into a csv format using Jupiter notebook. Primary goal of Jupiter notebook is to provide a Graphical user interface to users contains different applications; 'Explorer' is the first application is helps to open or retrieve dataset from the local storage or any other web. The second application is to experimental comparison of machine learning algorithm which also exists in Jupiter notebook. Hypothetically unlimited data stream in work incrementally.

B. Machine Learning:

Machine is subset of Artificial intelligence that provides functionality to learn automatically and improve the performance of without explicit program, machine learning algorithm can perform various application and improve the performance of the problem, it mainly focus on performance for that purpose it takes huge amount of data and use it and learn from themselves.

IV- PROPOSED SYSTEM

To solve this problem we are using different machine learning algorithm model and compare the result of algorithm in term of prediction of disease for that purpose we are use first machine learning model supervised learning classifier and predict the best accuracy, we have load the CSV files to our ML model and data are displayed using sklearn library on output panel, for prediction purpose. We have select some parameters and we provide a numeric type of data to our model in two dimensional Array format and then target way check patient tumor has been malignant or benign malignant, based on that value decides or calculated result patient cause or not due to breast cancer. In this case provide binary value for patient disease, malignant means patient cause breast cancer (0) and benign means patient does not have breast cancer (1), we provide large amount of dataset in machine learning model some data having malignant and some having a benign. Expert can select the cells of tumor and extract that into different features and store the value of patient, also store benign features.

We collect dataset information and visualize it using different parameters, now we can check how many are malignant tumor and how many are benign tumor, those patient value reduce or equal to one. Also calculated the heat map, calculate the relation between one to other features in datasets we can easily find out and distribute datasets and check the disease of patient

A. Import Library

Importing library is the initial step in Jupiter notebook before exploring the datasets from local storage device. The required files are installing using following Jupiter notebook command

```
import pandas as pd
import numpy as np
```

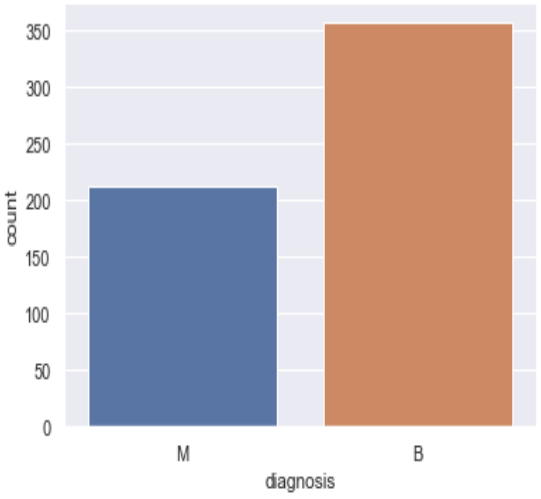
From the above command line installing numpy and pandas, numpy helps to numerical and complex calculation in the dataset and pandas help to analysis of datasets using different parameters.

B. Data Exploration

Data Exploration is the second step of the Jupiter notebook applications. We have convert file in appropriate file format called as CSV. After execution of command we can store data on local storage. We can explore this dataset using CSV file format in Jupiter notebook for prediction.

C. Preprocessing the data

One of the most important processes of data called as Data- Preprocessing. Encoding categorical data values in to numerical format for analysis, to remove noise and inconstancies in to data, before train datasets in preprocessing data must be needful.



1. System Architecture:

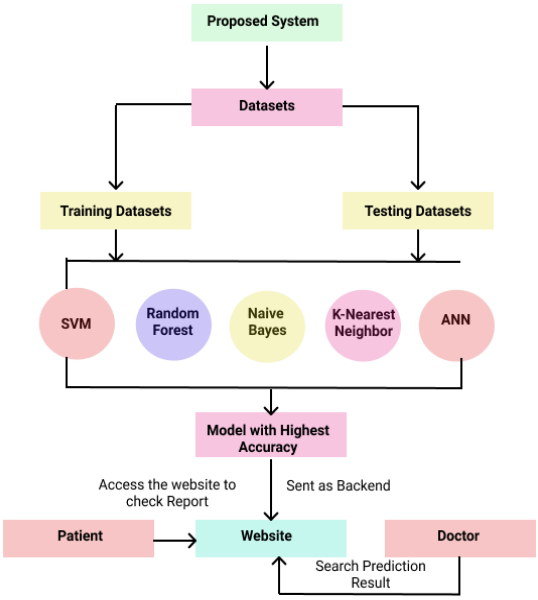


Figure: architecture of Fake news Detection

V- CONCLUSION

Breast cancer if found at an early stage will help save lives of thousands of women or even men. These projects help the real world patients and doctors to gather as much information as they can. The research on nine papers has helped us gather the data for the project proposed by us. By using machine learning algorithms we will be able to classify and predict the cancer into being or malignant. Machine learning algorithms can be used for medical oriented research, it advances the system, reduces human errors and lowers manual mistakes.

VI- BIBLIOGRAPHY

[1] De Magalhães, João Pedro. "How ageing processes influence cancer." Nature Reviews Cancer 13.5 (2013): 357.
[2] Yixuan Li, Zixuan Chen, Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction, Applied and Computational Mathematics. Vol. 7, No. 4, 2018, pp. 212-216. doi: 10.11648/j.acm.20180704.15
[3] Aloraini, Adel. "Different machine learning algorithms for breast cancer diagnosis." International Journal of Artificial Intelligence & Applications 3.6 (2012): 21.
[4] Gupta, P., and P. S.. “analysis of Machine earning echniques for Breast Cancer Prediction”. International Journal of Engineering and Computer Science, Vol. 7, no. 05, May 2018, pp. 23891-5, http://www.ijecs.in/index.php/ijecs/article/view/4071
[5] Asri, Hiba, et al. "Using machine learning algorithms for breast cancer risk prediction and diagnosis." Procedia Computer Science 83 (2016): 1064-1069.
[6] Sultana, Jabeen, Abdul Khader Jilani, & . .. "Predicting Breast Cancer Using Logistic Regression and Multi-Class Classifiers." International Journal of Engineering & Technology [Online], 7.4.20 (2018): 22-26. Web. 30 Nov. 2019
[7] Kriti Jain et al. “Breast Cancer Diagnosis Using Machine earning echniques”, International Journal of Innovative Science, Engineering & Technology, Vol. 5, Issue 5, May 2018.
[8] Brown, Gavin. Diversity in neural network ensembles. Diss. University of Birmingham, 2004.
[9] Zheng, Bichen, Sang Won Yoon, and Sarah S. Lam. "Breast cancer diagnosis based on feature extraction using a hybrid of Kmeans and support vector machine algorithms." Expert Systems with Applications 41.4 (2014): 1476-1482.
[10] Puneet Yadav et al. “Diagnosis of Breast Cancer using Decision ree Models and SVM”, International Research Journal of Engineering and Technology, Vol. 5, Issue 3, Mar 2018. [11] Medjahed, Seyyid Ahmed, Tamazouzt Ait Saadi, and Abdelkader Benyettou. "Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules." International Journal of Computer Applications 62.1 (2013).
[12] Chaurasia, Vikas, Saurabh Pal, and B. B. Tiwari. "Prediction of benign and malignant breast cancer using data mining techniques." Journal of Algorithms & Computational Technology 12.2 (2018): 119-126.
[13] Williams, Kehinde ladipo et al. “Breast Cancer Risk Prediction Using Data Mining Classification echniques.” (2015).
[14] Syakur, M. A., et al. "Integration k-means clustering method and elbow method for identification of the best customer profile cluster." IOP Conference Series: Materials Science and Engineering. Vol. 336. No. 1. IOP Publishing, 2018.
.