**Abstract**

At present world, Breast cancer is a second main cause of cancer death in women after lung cancer. Breast cancer occurs when some breast cells begin to raise abnormally. It can arise in any portion of the Breast and it can be prevented if the treatment is started at the early stage of the Breast cancer. Breast cancer is a malignant tumor i.e. a collection of cancer cells arising from the cells of the breast Treatment of breast cancer relies on the cancer type and its stage (zero to fourth) and may include surgery, radiation, or chemotherapy. Mainly this paper focused on diagnosing the Breast cancer disease using various classification algorithm with the help of data mining tools. Data mining of the intelligent accumulated from previously disease detected patients opened up a new aspect of medical progression. In this paper, the capability of the classification of Naïve Bayes, Random Forest, Logistic Regression, Multilayer Perceptron, K-nearest neighbors in evaluating the Breast Cancer Disease dataset culled from UCI machine learning repository, was observed to predict the existence of Breast cancer. Data set has been explored in terms of Kappa Statistics, TP rate, FP Rate and precision.

**Introduction to Project**

Now a day, breast cancer is one of the burning issue all over the world. It is one of the major health problem for women. Globally the incidence of breast cancer is only second to that of Lung cancer. The disease represents the main cause of cancer death among women. Breast cancer is developed from breast tissue. Signs of breast cancer may include a breast lump, skin dimpling, fluid coming from the nipple, breast shape change, a newly inverted nipple, or a scaly patch of skin. Breast cancer typically attack postmenopausal women. Both genetic and ancestral factor play a role. About 5-10% of breast cancer are hereditary and occur in the patient with mutation BRCA1, BRCA2 genes [1]. Prolong estrogen exposure associated with early menarche, late menopause uses of hormone replacement therapy (HRT) has been associated with increased risk. Other risk factor includes Obesity Alcohol intake, nulliparity and late first pregnancy [2]. Breast cancer usually present as a Palpable mass with nipple discharge. Breast cancer may metastasis to bone (70%), lung (60%), liver (55%) and other organ [4].

**Functional Requirements**

**Admin:** we have to login with admin details

**Dataset upload:** data set uploading in a given manner

**Viewdataset:**data which we have given It is shown

**Classification:**there has been classified

**Evaluations:** evaluating the the things

**User** we have to login with our details

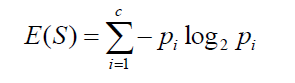
**Prediction:** predict the following substances

**Implementation**

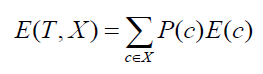
**Algorithm used in Project**

**J48 Decision Tree Algorithm**

**Step 1:** Calculate Entropy using the frequency table



**Step 2:** Split Entropy (T, X) is separated information of T on the value of the categorical attribute X.



E(T, P)=

E(T, N)=

**Step 3:**The resulting entropy is subtracted from the entropy before the split. The result is the Information Gain, or decrease in entropy.



if Gain(T, P)> Gain(T, N)

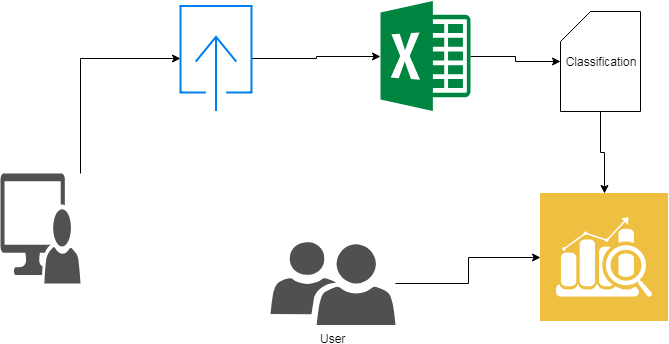
result is positive

else

result is negative

**Architecture diagram**

Project Architecture diagram



**SYSTEM CONFIGURATION:**

**Hardware requirements:**

Processer                     :           Any Update Processer

Ram                             :           Min 1 GB

Hard Disk                   :           Min 100 GB

**Software requirements:**

Operating System       :           Windows

Technology                 :           Python 3.6

IDE : PyCharm

UML : Star UML

DFD : DFD Drawer