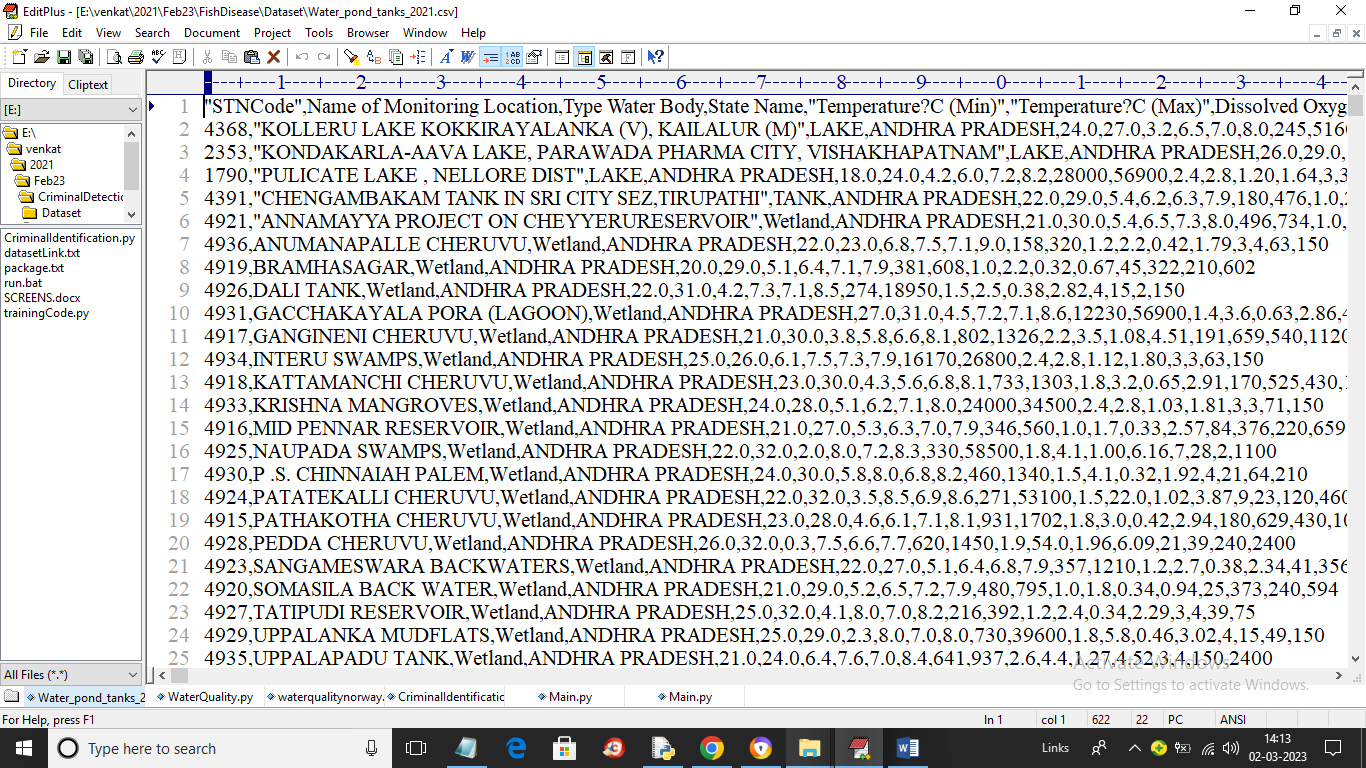
A Machine Learning Approach for Early Detection of Fish Diseases by Analysing Water Quality

In this paper author is analysing water quality by applying machine learning algorithm to predict fish diseases. If water contains high toxins or viruses then it will affect fish lungs which causes disease inside fish or fish may die. So we can collect water samples and then apply on machine learning model to predict water quality and if quality is not good then we can say fish is not healthy. In propose paper author has used water quality dataset from KAGGLE website and then trained with Gradient Boosting algorithm. This algorithm giving more than 95% accurate prediction accuracy on test data.

Dataset can be downloaded from below link

<https://www.kaggle.com/datasets/balabaskar/water-quality-data-india>

Below screen showing dataset details



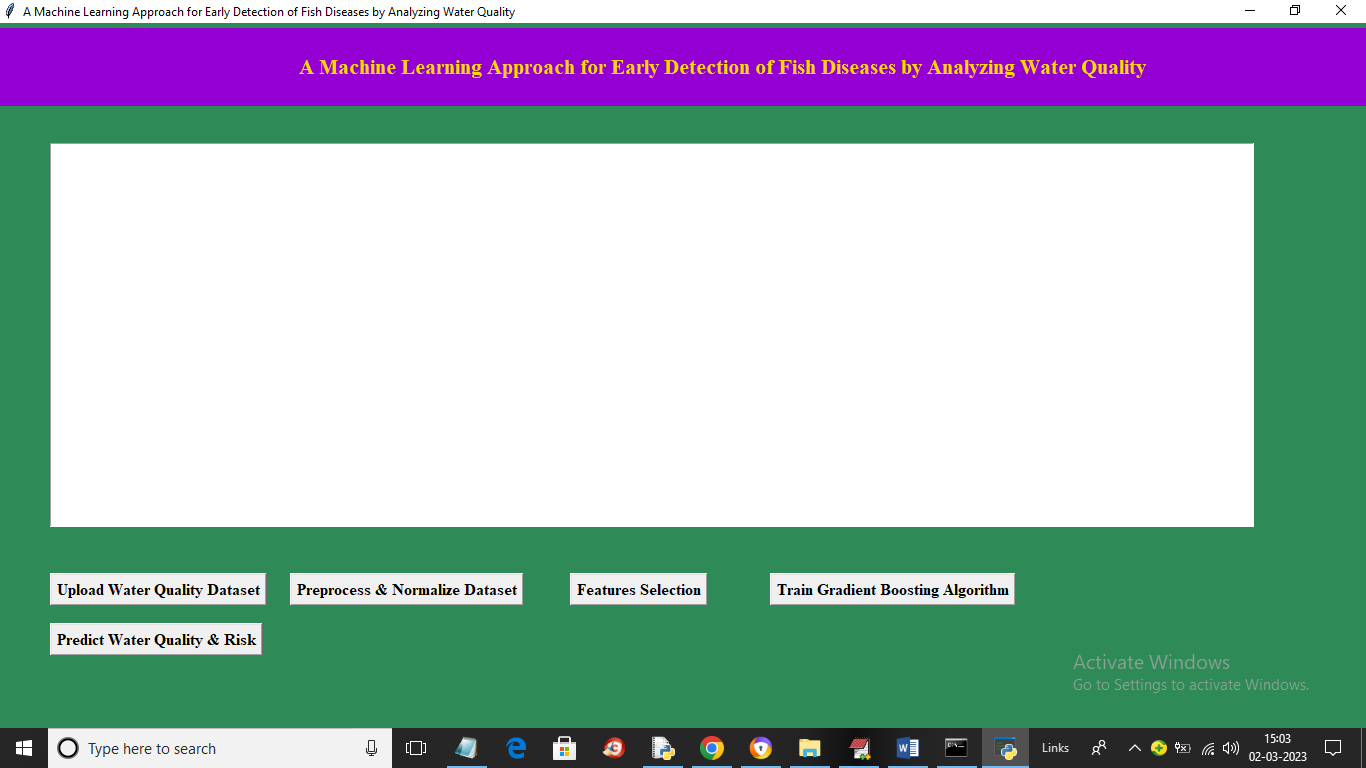
In above dataset screen first row contains dataset column names and remaining rows contains dataset values such as presence of chemical and ECOLI and other viruses and by using this dataset we will train Gradient Boosting algorithm for fish disease prediction.

To implement this project we have designed following modules

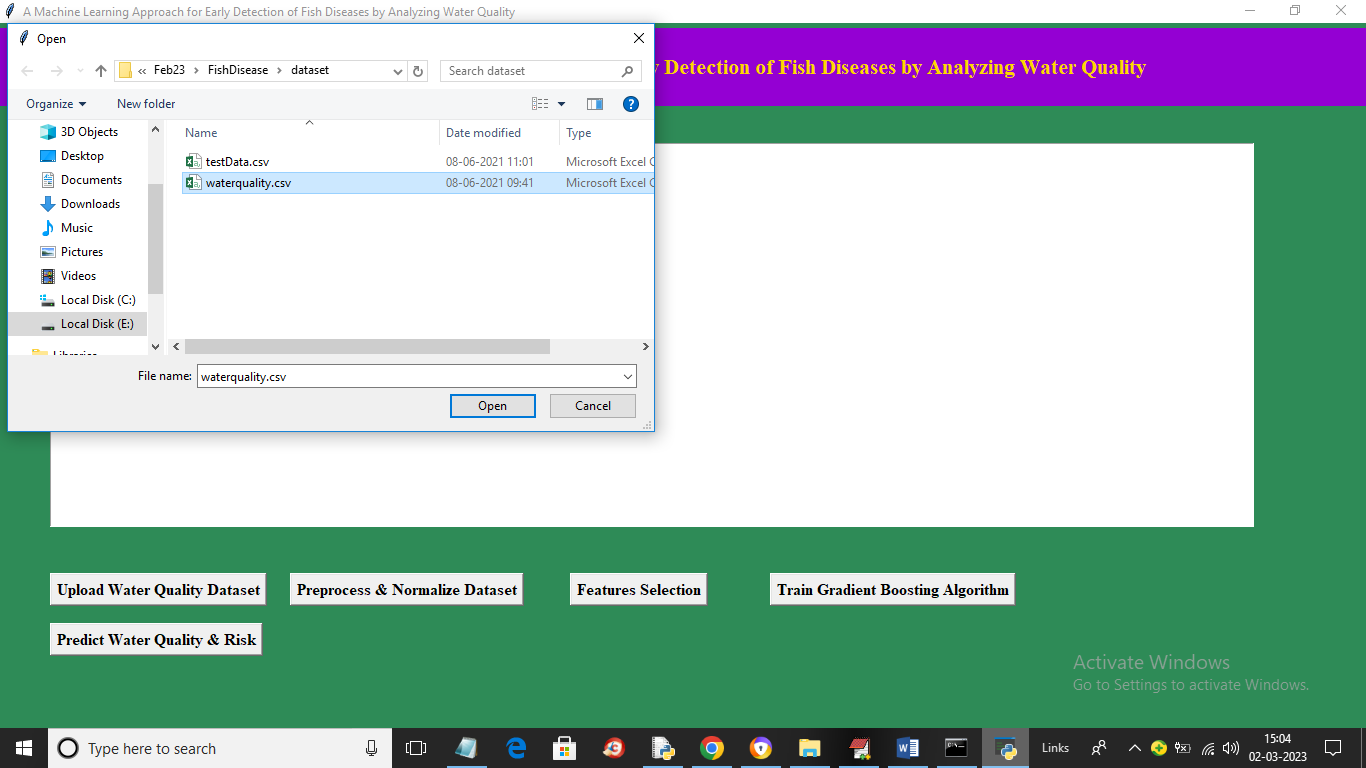
1. Upload Water Quality Dataset: using this module we will upload dataset to application
2. Preprocess & Normalize Dataset: using this module will convert all non-numeric data to numeric data and then normalize all values
3. Features Selection: using this module application will select X training features and Y class label and then split dataset into train and test where application using 80% dataset for training and 20% for testing
4. Train Gradient Boosting Algorithm: 80% training data will be input to Gradient Boosting algorithm to train a model and this model will be applied on test data to calculate prediction accuracy
5. Predict Fish Condition: using this module we will upload test data and then algorithm will predict weather fish is healthy or disease affected.

SCREEN SHOTS

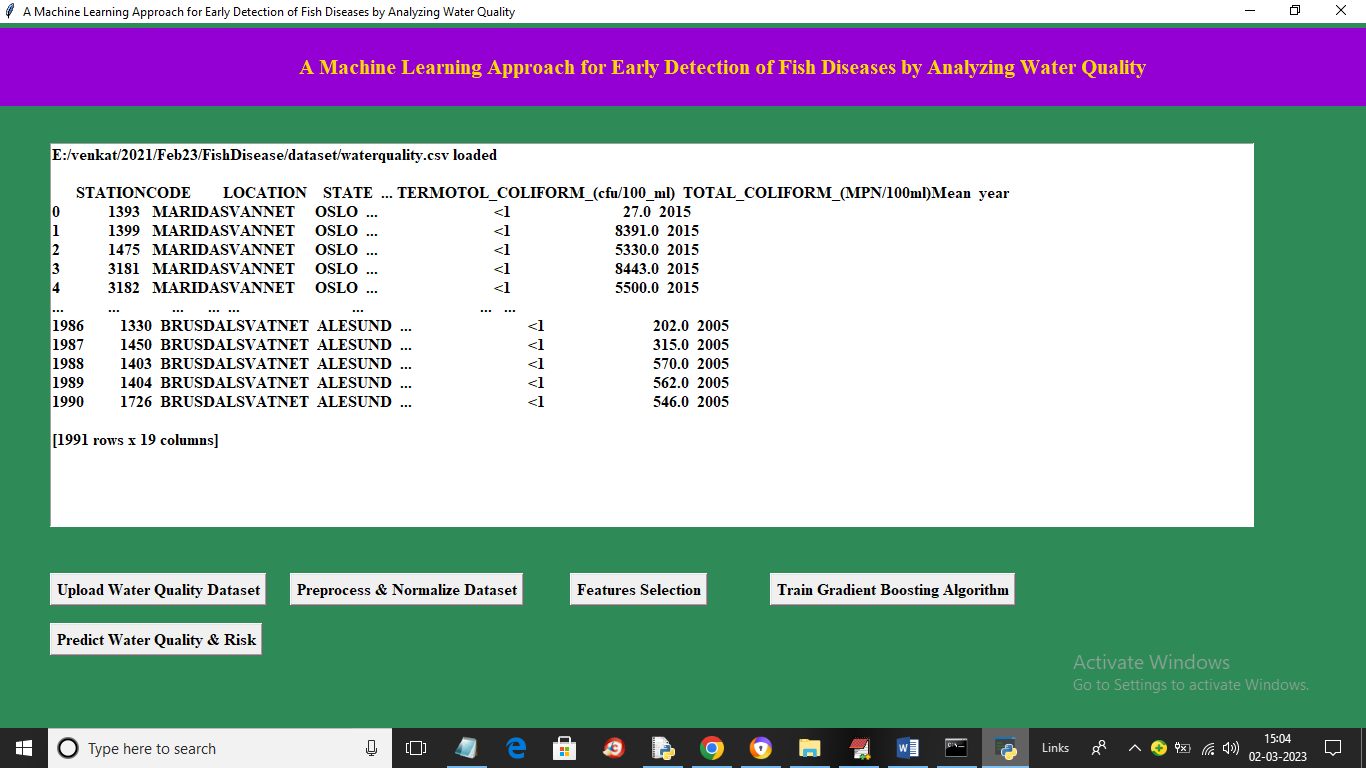
To run project double click on ‘run.bat’ file to get below screen



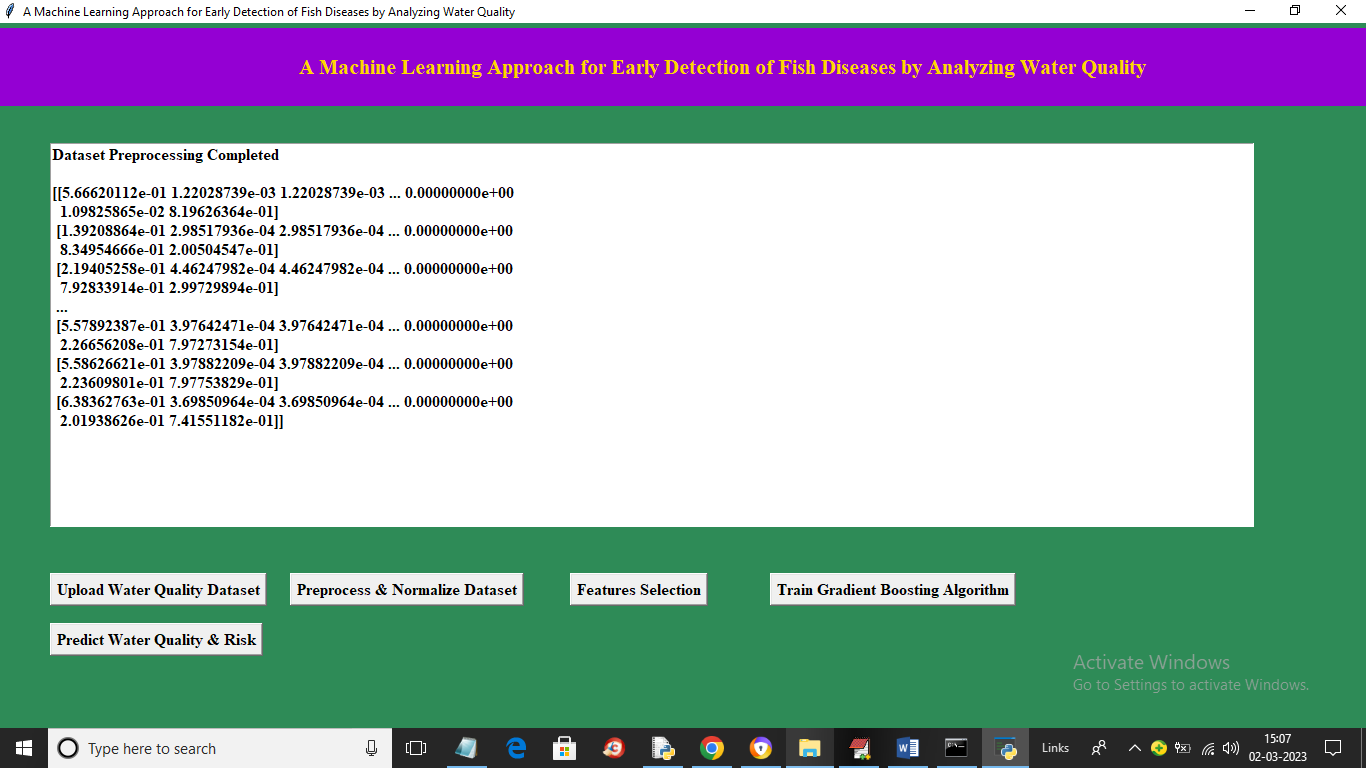
In above screen click on ‘Upload Water quality Dataset’ button to upload dataset and get below output



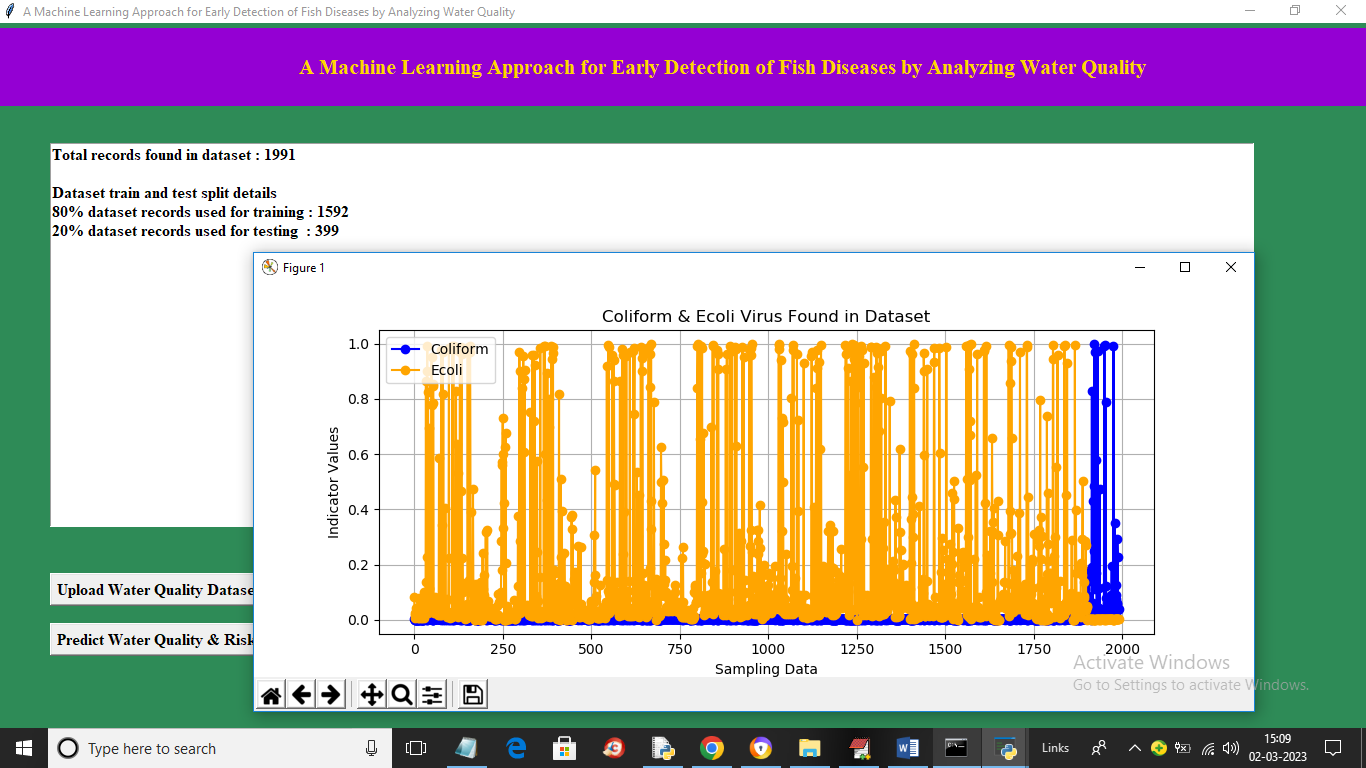
In above screen click on ‘Open’ button to load dataset and get below output



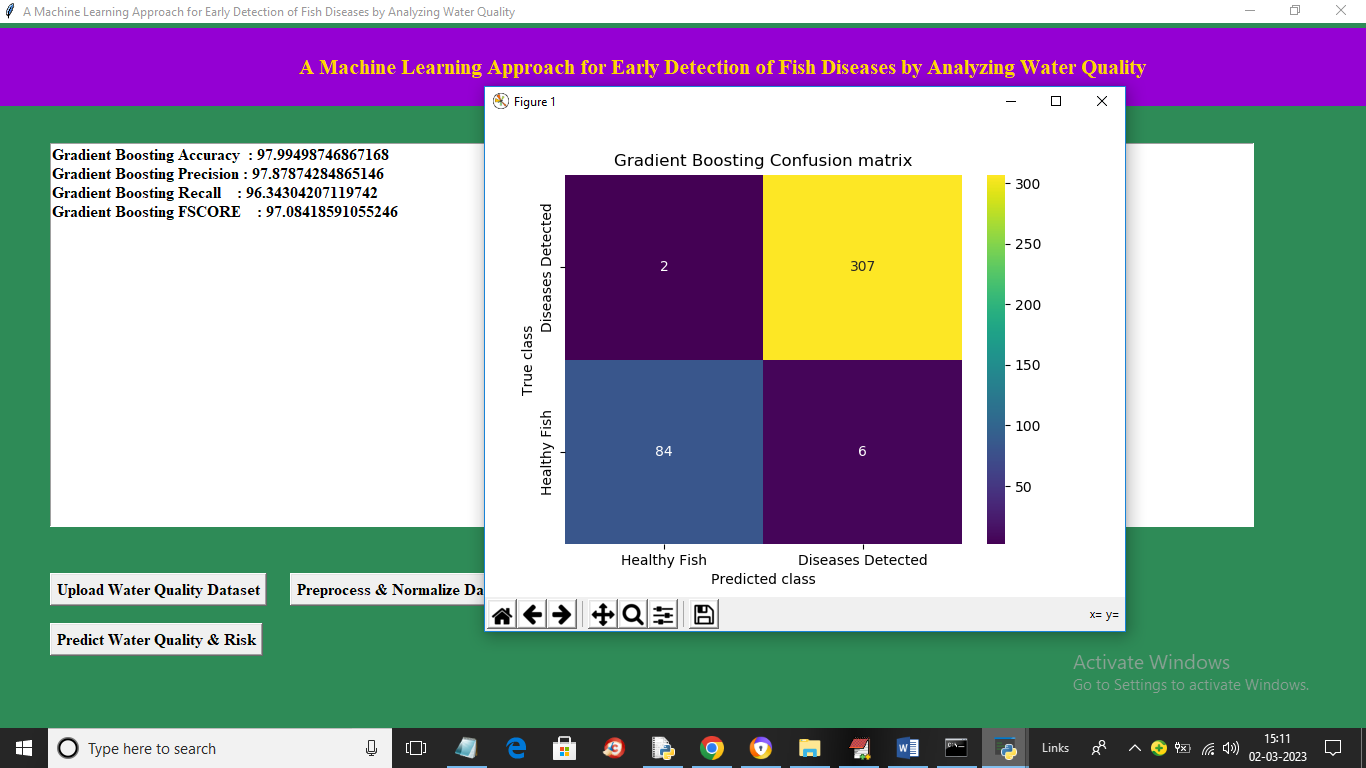
In above screen dataset loaded and we can see dataset contains numeric and non-numeric values and machine learning algorithms accept only numeric dataset so by applying label encoder class we can convert non-numeric data to numeric values so click on ‘Preprocess & Normalize Dataset’ button to get below output



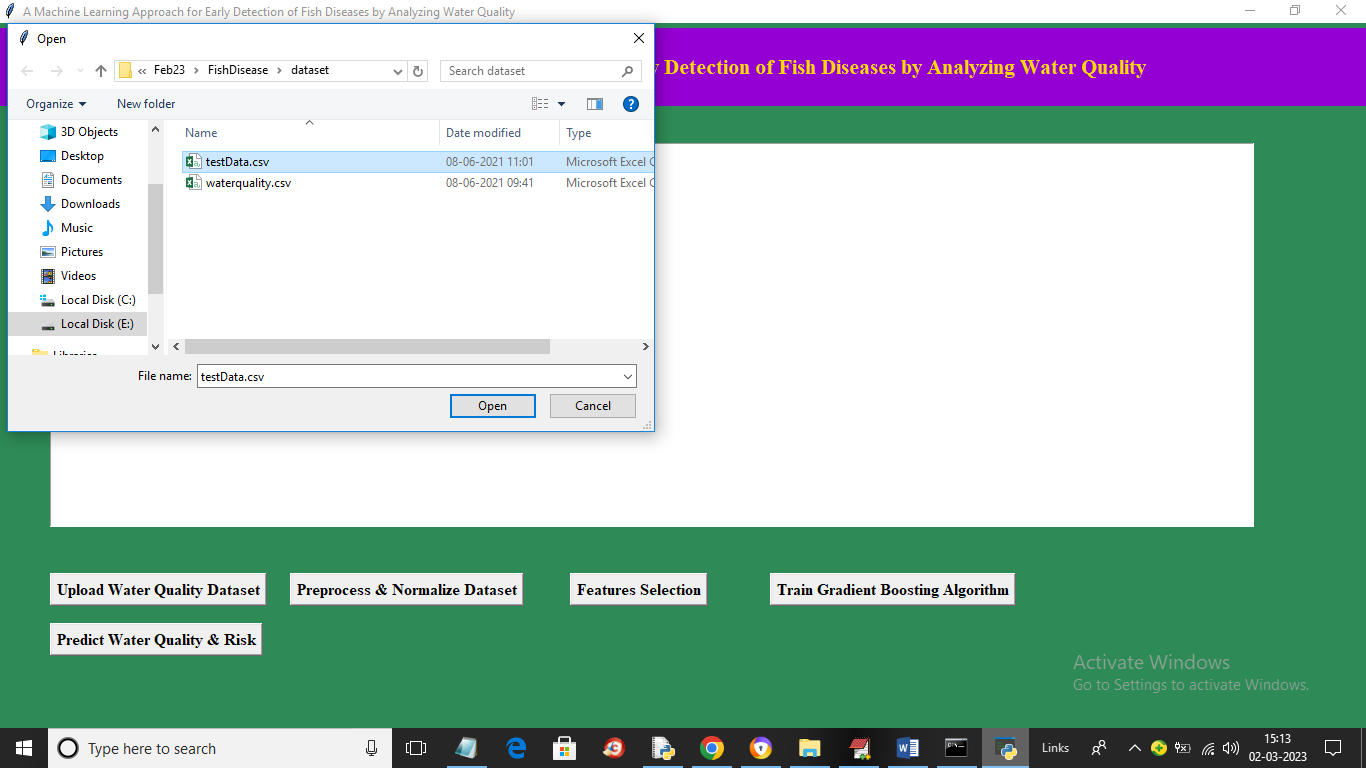
In above screen all values are converted to numeric format and now click on ‘Features Selection’ button to extract X and Y features from dataset and then split into train and test values and get below output



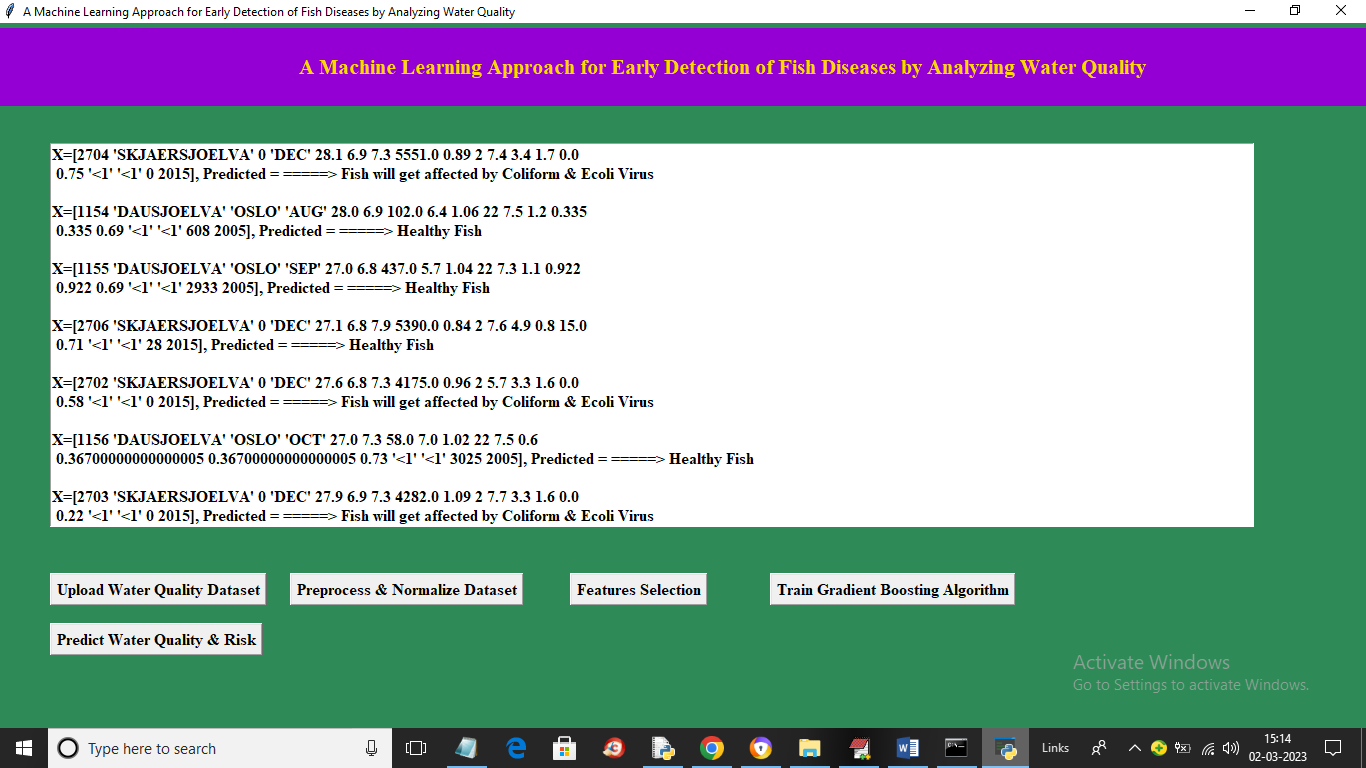
In above screen we can see dataset total values and then we can see training and testing dataset size and in graph x-axis represents number of records and y-axis represents presence quality of ‘Coliform and Ecoli’ virus where yellow line is for Ecoli and blue line for Coliform and now close above graph and then click on ‘Train Gradient Boosting Algorithm’ button to train algorithm and get below output



In above screen with Gradient Boosting we got 97% accuracy and in confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and blue colur boxes represents Incorrect prediction count which is 2 only and different colour boxes contains correct prediction count. Now close above graph and then click on ‘Predict Water Quality & Risk’ button to get below output



In above screen selecting and uploading ‘testData.csv’ file and then click on ‘Open’ button to load dataset and get below output



In above screen in square bracket we can see test data values and after arrow symbol we can see predicted values as healthy or disease affected fish