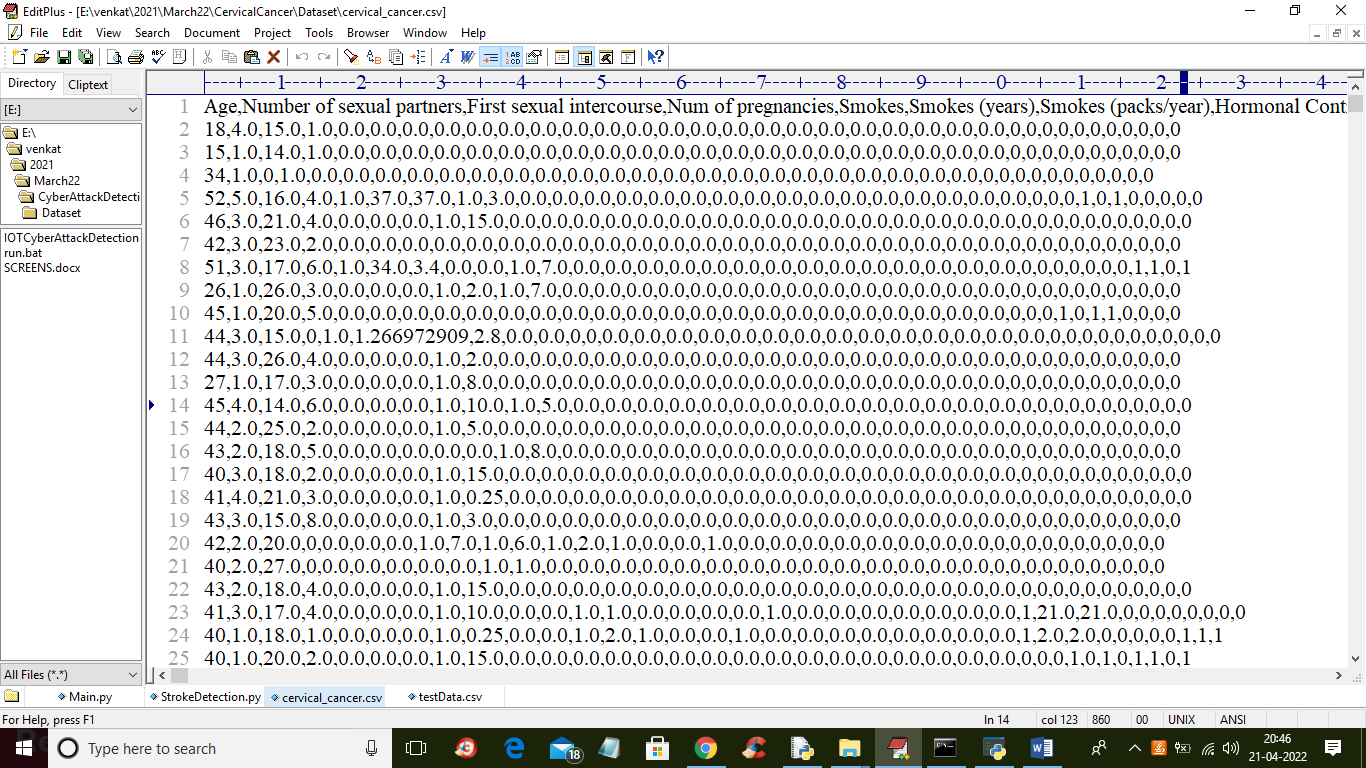
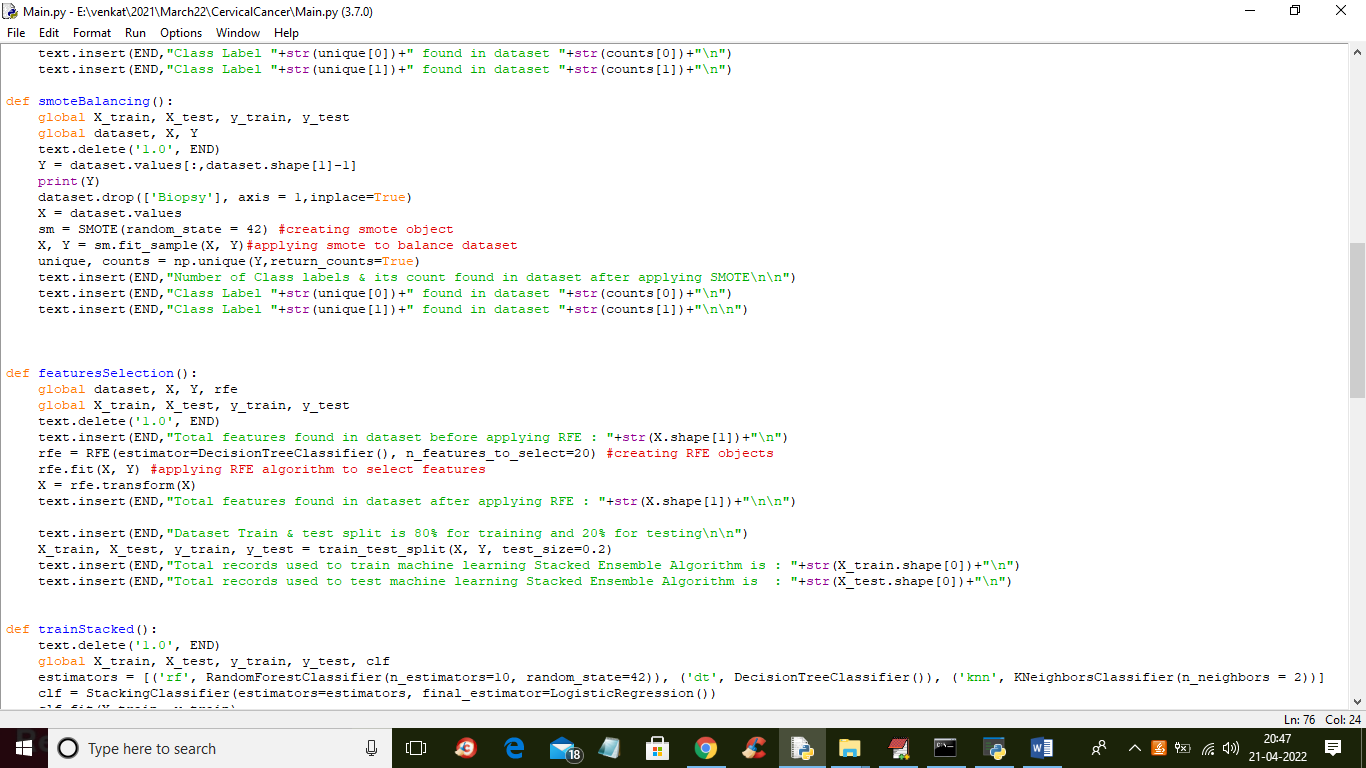
Optimised stacked ensemble techniques in the prediction of cervical cancer using SMOTE and RFERF

In this project we are predicting cervical cancer by using stacked ensemble machine learning algorithm which is a combination of stacked classifier such as Random Forest, Decision Tree and KNN. To train stacked ensemble algorithm we have used cervical cancer dataset which 900 records and in that only 53 records belongs to CANCEL class and remaining 847 records are normal and this dataset is highly imbalanced and to balance this dataset we have used SMOTE algorithm and this dataset contains 32 attributes and all attributes are not important so we applied Recursive Features Selection (RFE) algorithm to select important features from dataset. Belo screen showing dataset details.



In above dataset screen first row represents dataset column names and remaining attributes contains dataset values. We will used above dataset to train stacked algorithm.

In below screen we are showing code with comments on SMOTE and RFE usage



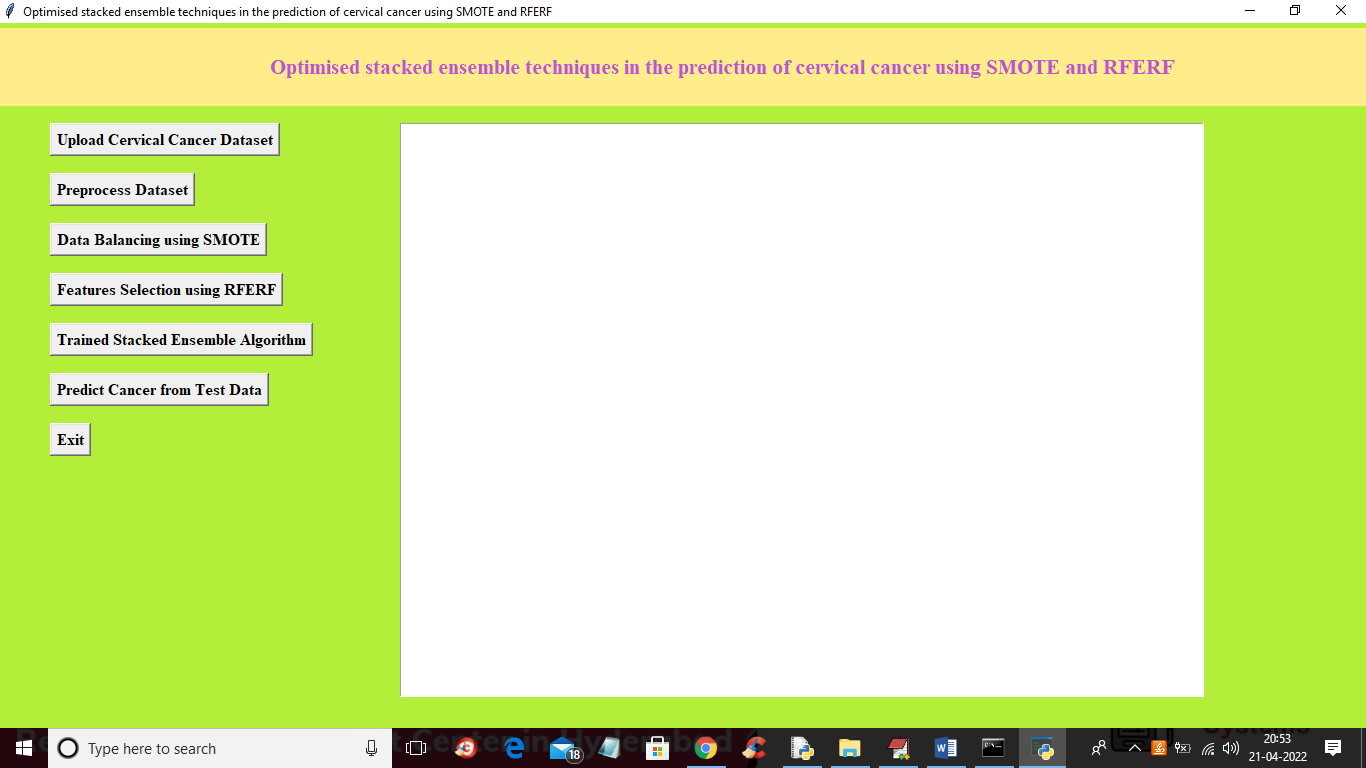
In above screen read red colour comments to know about SMOTE and RFE features selection.

To implement this project we have designed following modules

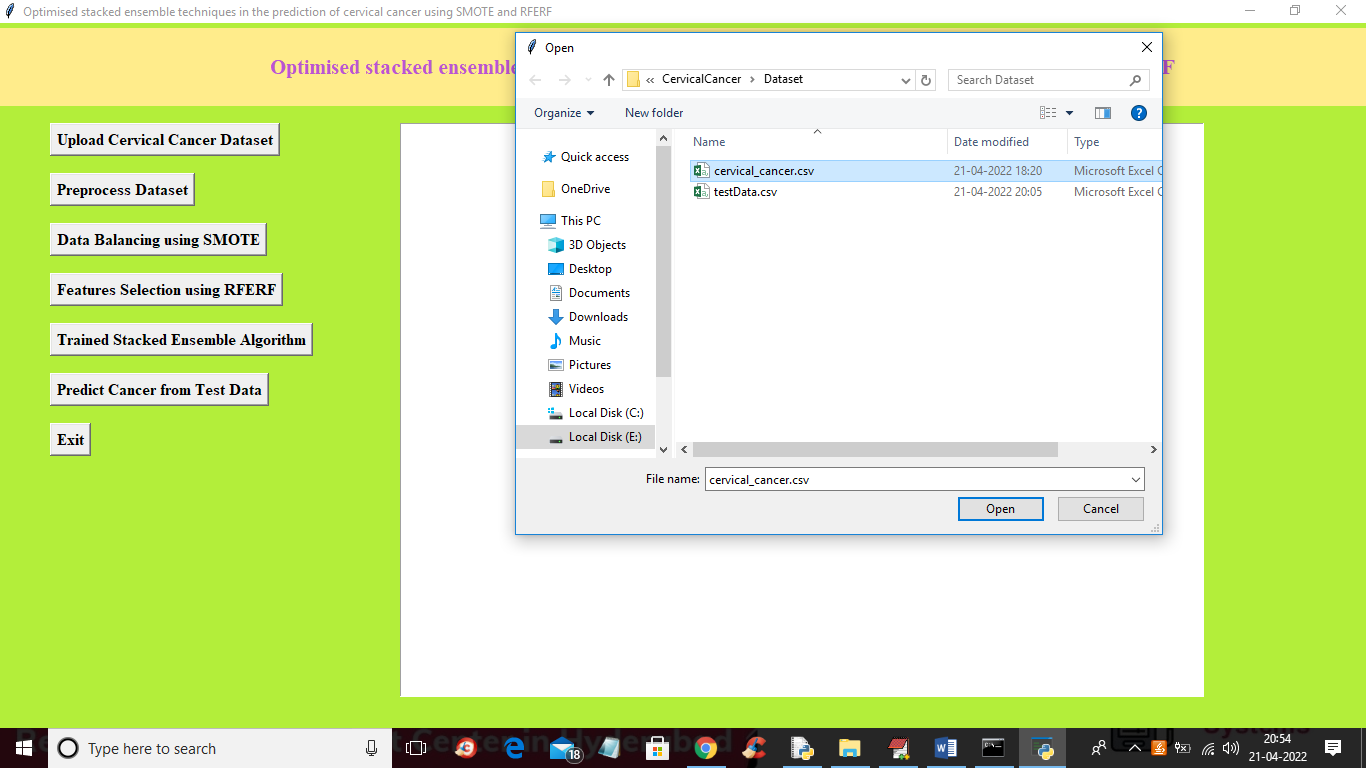
1. Upload Cervical Cancer Dataset: using this module we will upload dataset to application and then find total number of dataset records, columns and normal and cancer count records
2. Preprocess Dataset: dataset often contains missing values so by using this module we have cleaned the dataset by replacing missing values with 0 and then normalized all records and then displaying total number of records for each class before applying SMOTE
3. Data Balancing using SMOTE: using this module we will apply SMOTE to balanced dataset
4. Features Selection using RFERF: using this module we will apply RFE to select important features from dataset and then split dataset into train and test where 80% dataset used for training and 20% dataset size used for testing
5. Trained Stacked Ensemble Algorithm: using this module we will trained stacked ensemble model and then trained model will be applied on 20% test data to calculate prediction accuracy
6. Predict Cancer from Test Data: using this module we will upload test data and then stacked ensemble will predict weather test data is normal or contains cervical cancer.

SCREEN SHOTS

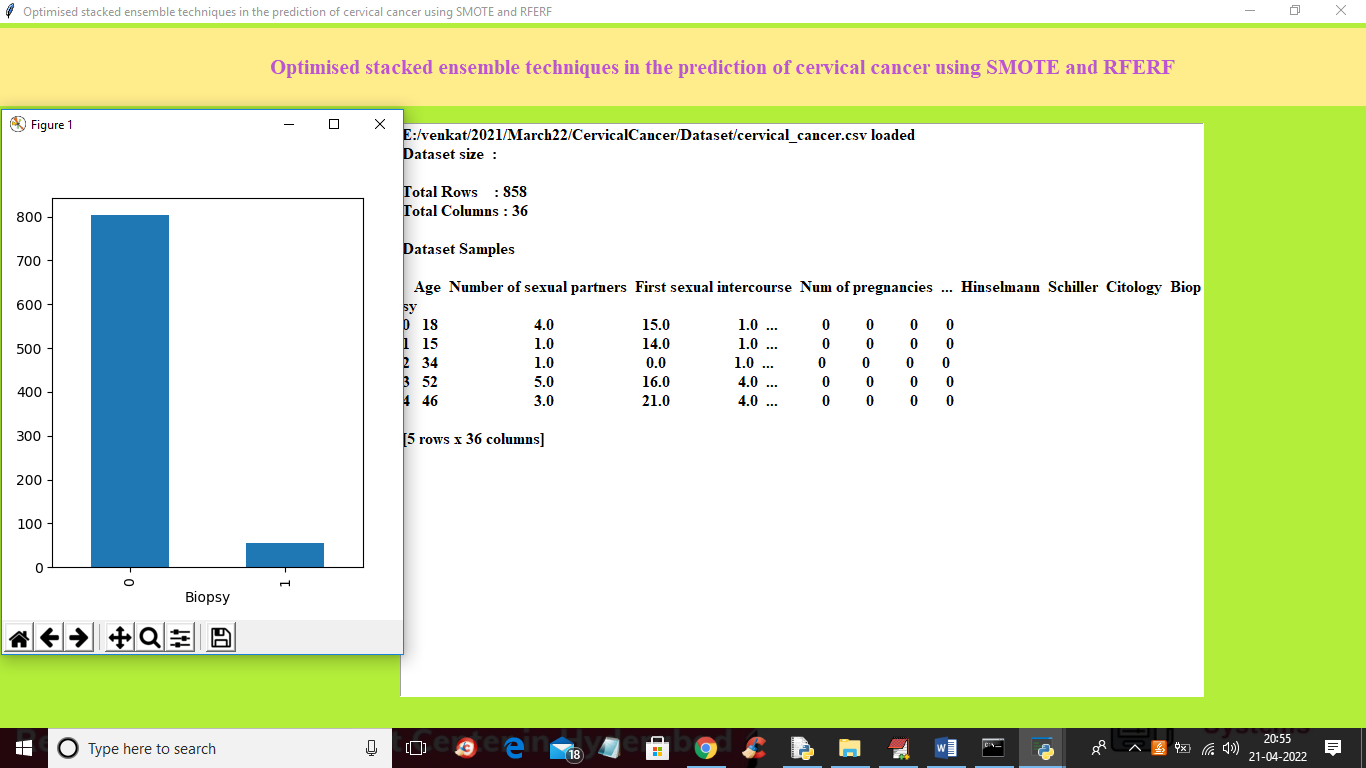
To run project double click on ‘run.bat’ file to get below screen



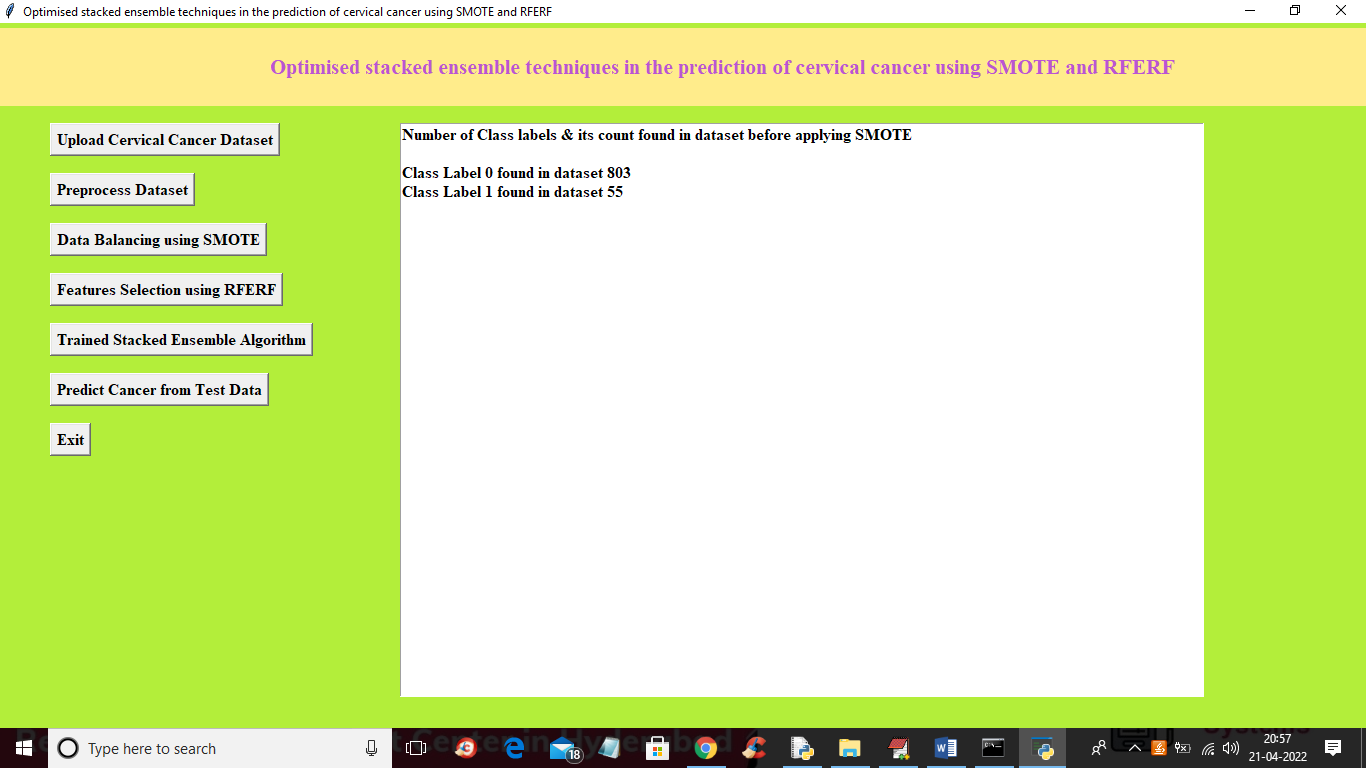
In above screen click on ‘Upload Cervical Cancer Dataset’ button to upload dataset and to get below screen



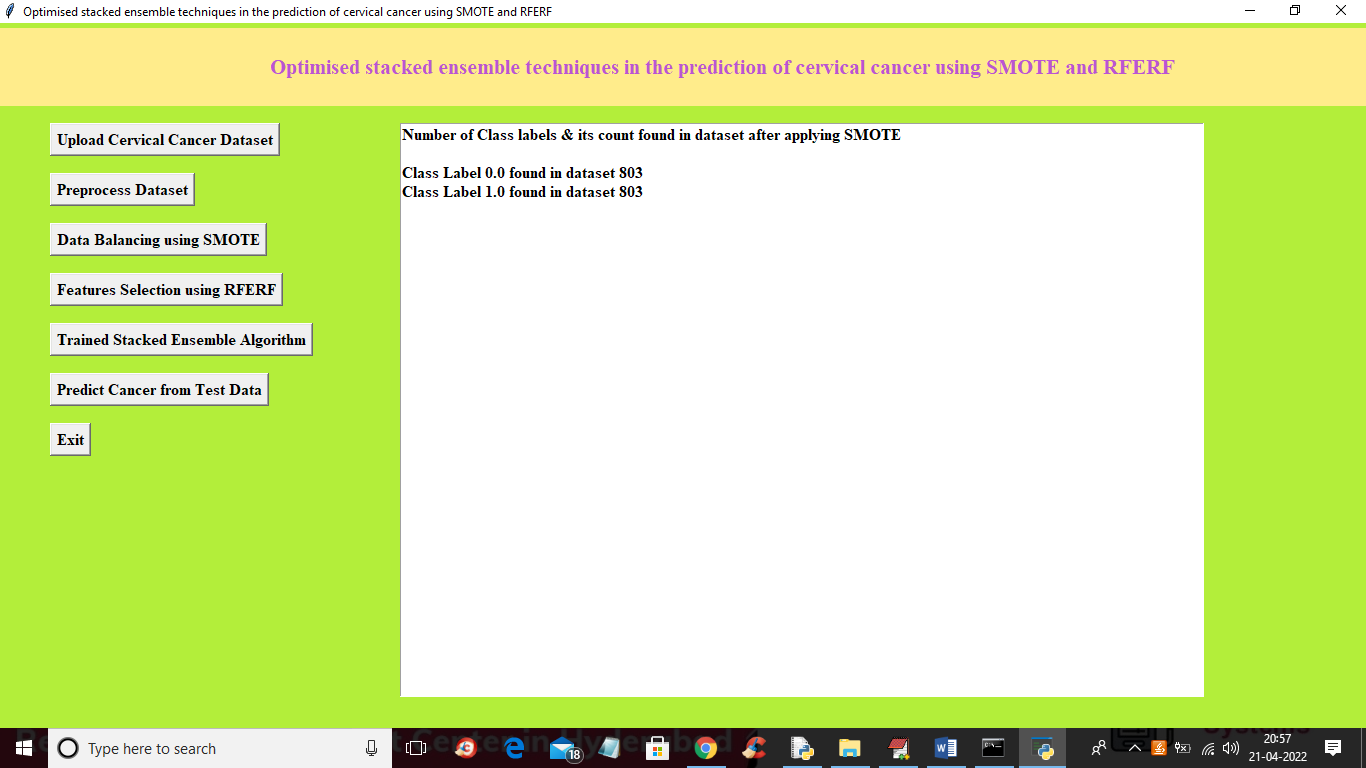
In above screen selecting and uploading dataset file and then click on ‘Open’ button to load dataset and to get below screen



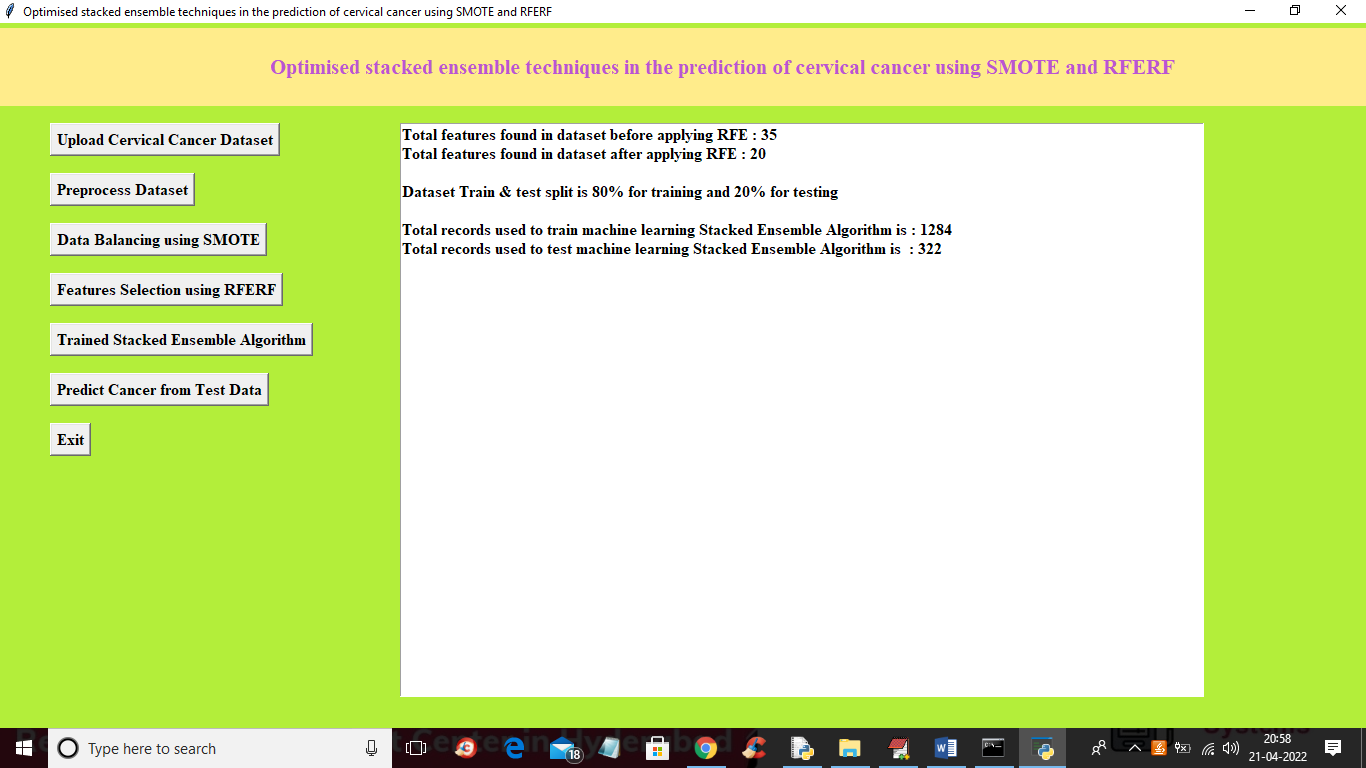
In above screen we can see dataset loaded and in graph x-axis represents class label 0 (normal) and 1 (cervical cancer) and y-axis represents number of records and we can see dataset is highly imbalanced as ‘0’ contains more than 800 records and ‘1’ class label contains only 53 records. Now close above graph and then click on ‘Preprocess Dataset" button to remove missing values and normalize dataset



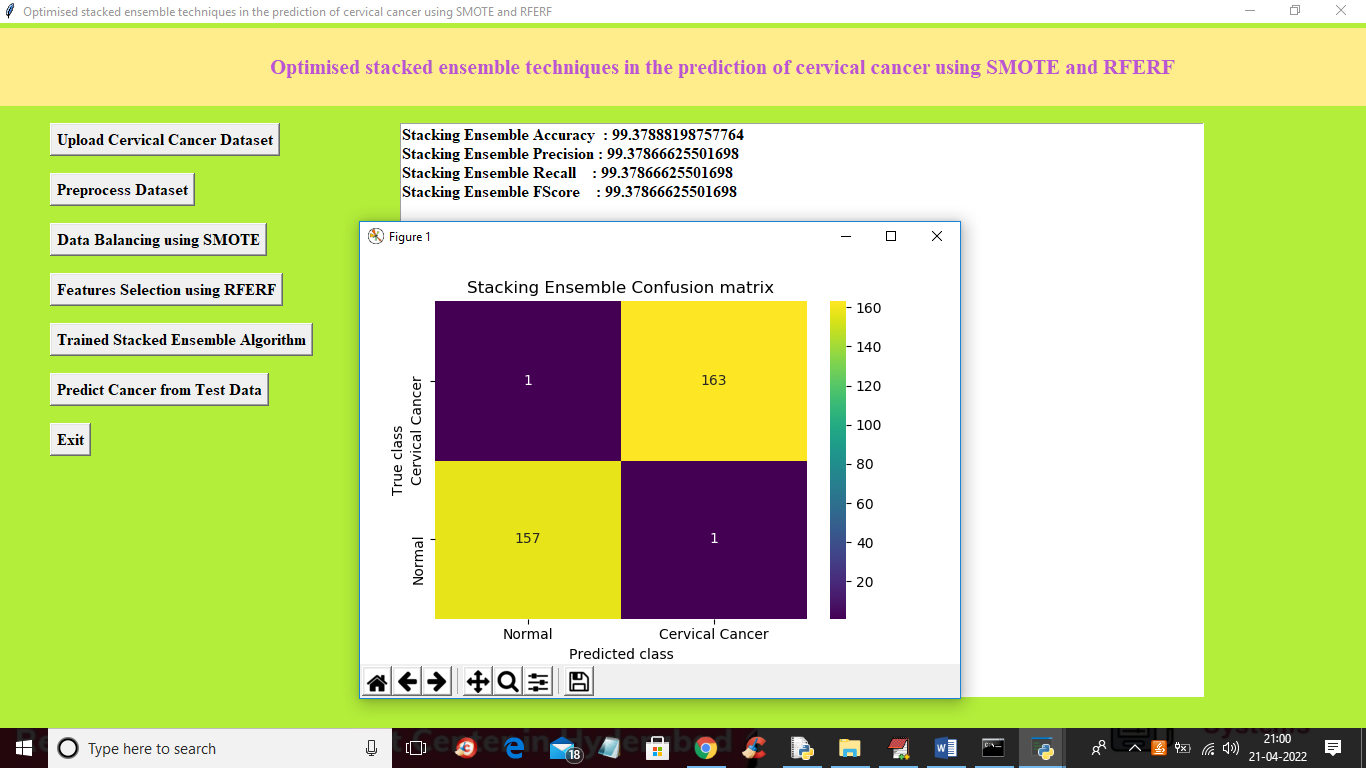
In above screen dataset is processed and class 0 contains 805 records and 1 contains 55 records so click on ‘Data Balancing using SMOTE’ button to balanced dataset and get below output



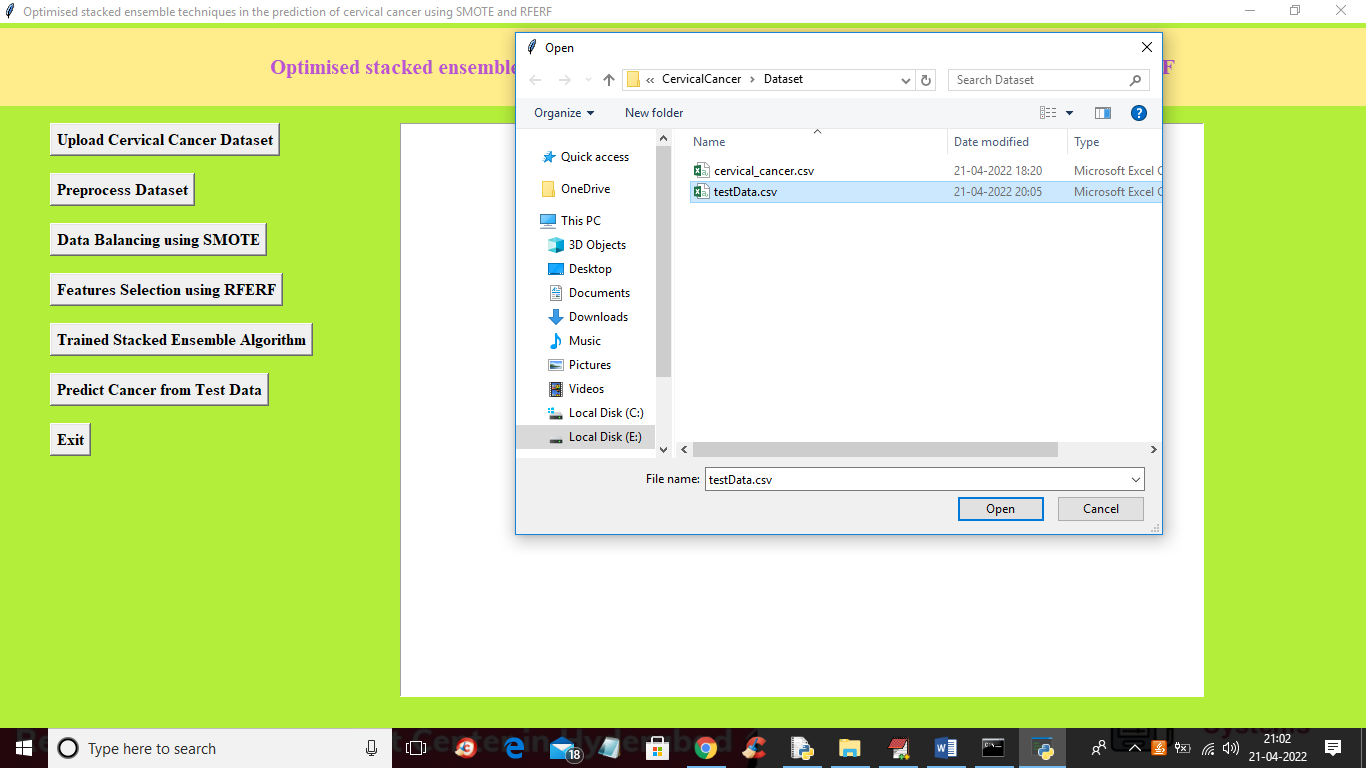
In above screen after applying smote both classes contains 803 records so dataset is balanced and now click on ‘Features Selection using RFERF’ button to get below output



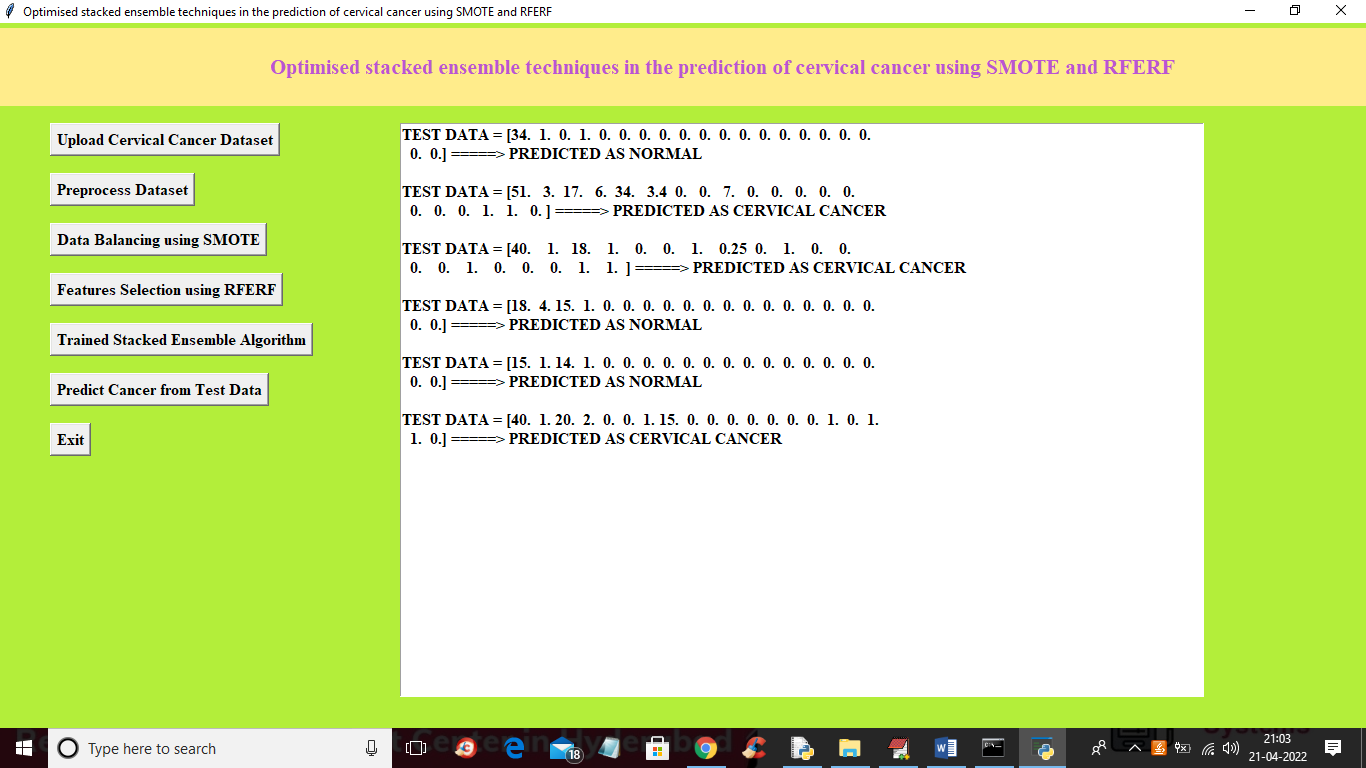
In above screen in first 2 lines we can see dataset contain 35 attributes and after applying RFE, attributes size reduced to 20 and then we can see dataset train and test split details. Now dataset is ready and now click on ‘Trained Stacked Ensemble Algorithm’ button to train algorithm and get below output



In above screen with stacked ensemble algorithm we got 99.37% accuracy and in confusion matrix graph we can see x-axis represents predicted classes and y-axis represents TRUE classes and only 1 class in wrongly predicted and remaining are the correctly predicted. Now close above graph and then click on ‘Predict Cancer from Test Data’ button to upload test file and get below output



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



In above screen in square bracket we can see test data and then after arrow symbol we can see predicted result as NORMAL or CERVICAL CANCER