A Decision Tree based Recommendation System for Tourists

In this paper author is implementing C4.5 decision tree algorithm with MRMR features selection to recommend travel areas to tourist by using dataset from past tourist experiences. All existing algorithms such as collaborative or content filtering algorithms uses current user past experience data to recommend him new locations. These algorithms will not work if this current user has no past experiences data.

To overcome from above problem author is asking to use C4.5 decision tree algorithms which take experiences of previous users and then build a model and if new user enter his requirements then decision tree will predict best location based on his given input. Decision tree don’t need new users past experience data.

To implement decision tree model we need to have dataset and this dataset sometime will have empty or garbage values and this values will put bad effect on decision tree model so we can remove such empty or garbage values by applying pre-process techniques.

Sometime to predict or build model no need to use all columns (attributes) values from dataset and this unnecessary attributes can be remove by apply features selection algorithms and here we are using MRMR features selection algorithms to remove unnecessary attributes to reduce execution time of building model and to increase system accuracy.

Below are the dataset columns or attributes taken from previous users to build model.

This data set is populated by crawling TripAdvisor.com. Reviews on destinations in 10 categories mentioned across East Europe are considered. Each traveller rating is mapped as Excellent (4), Very Good (3), Average (2), Poor (1), and Terrible (0) and average rating is used against each category per user.

Dataset columns and values

userid,art\_galleries,dance\_clubs,juice\_bars,restaurants,museums,resorts,parks\_picnic\_spots,beaches,theaters,religious\_institutions,location

Above are the column names and below are the column values

User 1,0.93,1.8,2.29,0.62,0.8,2.42,3.19,2.79,1.82,2.42,Amsterdam\_Heining\_2

User 2,1.02,2.2,2.66,0.64,1.42,3.18,3.21,2.63,1.86,2.32,Amsterdam\_Jachthaven\_ijbur

User 3,1.22,0.8,0.54,0.53,0.24,1.54,3.18,2.8,1.31,2.5,Amsterdam\_Bert\_Haanstra\_Kad

User 4,0.45,1.8,0.29,0.57,0.46,1.52,3.18,2.96,1.57,2.86,Amsterdam\_Ruigoord\_Ker

In above values first column is USER\_ID and second column is ART\_GALLERIES and third is DANCE CLUB etc and for each column user had given rating from 4 to 0 and 4 means Excellent service.

Now using above values we can build C4.5 decision tree and prediction will be done using below test values

'User 122',0.93,1.8,2.29,0.62,0.8,2.42,3.19,2.79,1.82,2.42,?

'User 222',1.02,2.2,2.66,0.64,1.42,3.18,3.21,2.63,1.86,2.32,?

'User 3222',1.22,0.8,0.54,0.53,0.24,1.54,3.18,2.8,1.31,2.5,?

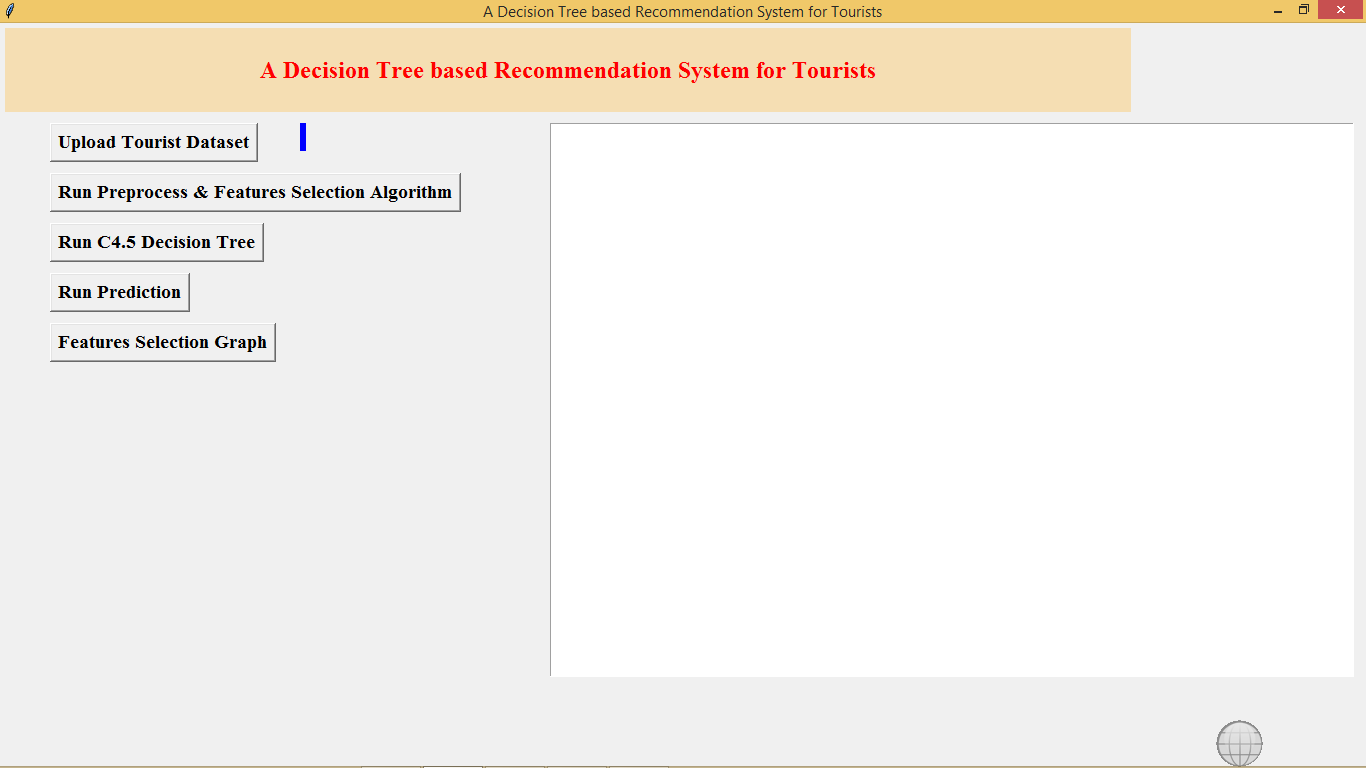
'User 4222',0.45,1.8,0.29,0.57,0.46,1.52,3.18,2.96,1.57,2.86,?

'User 522',0.51,1.2,1.18,0.57,1.54,2.02,3.18,2.78,1.18,2.54,?

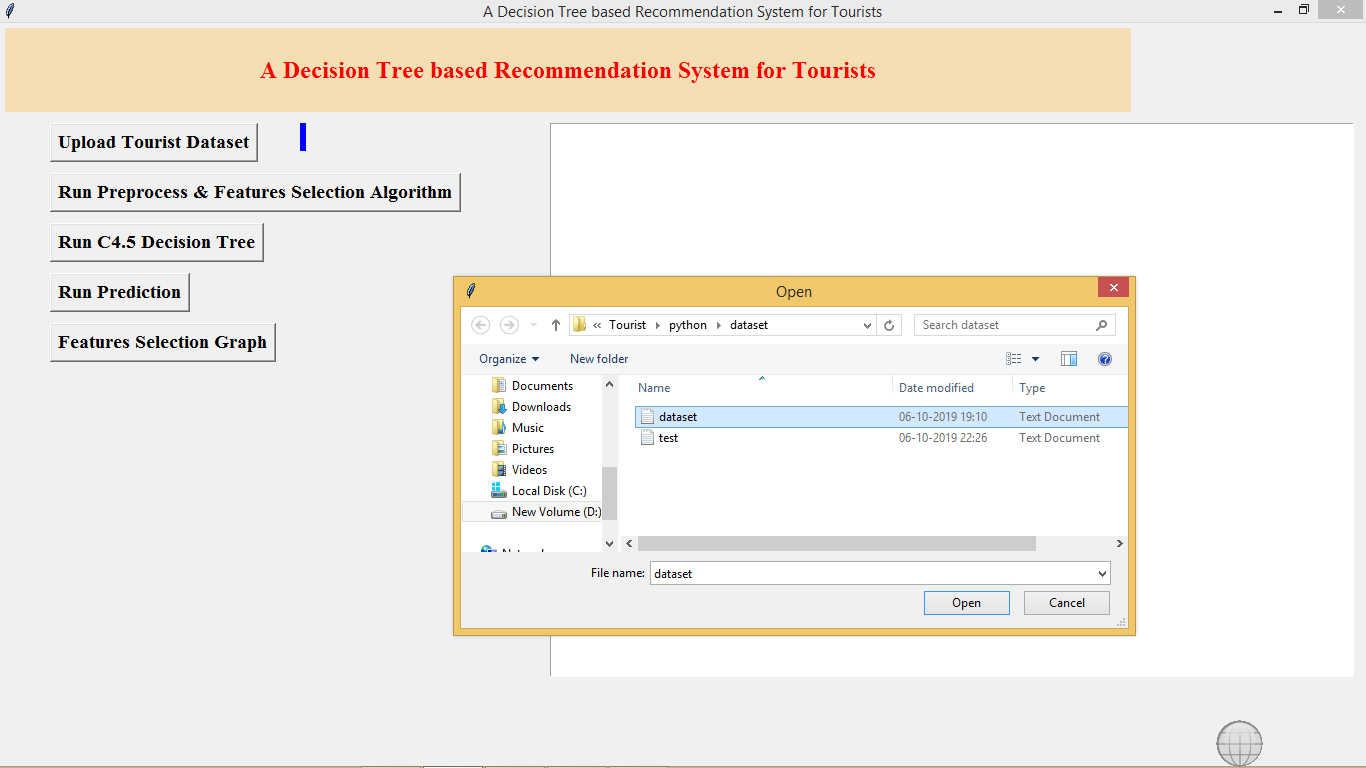
In above test values new user has given values to look for location which has above service rating but new user don’t know which location provides such services so he will put question mark and when we upload above test values to decision tree then it will take decision and predict best location and inform to user.

Screen shots

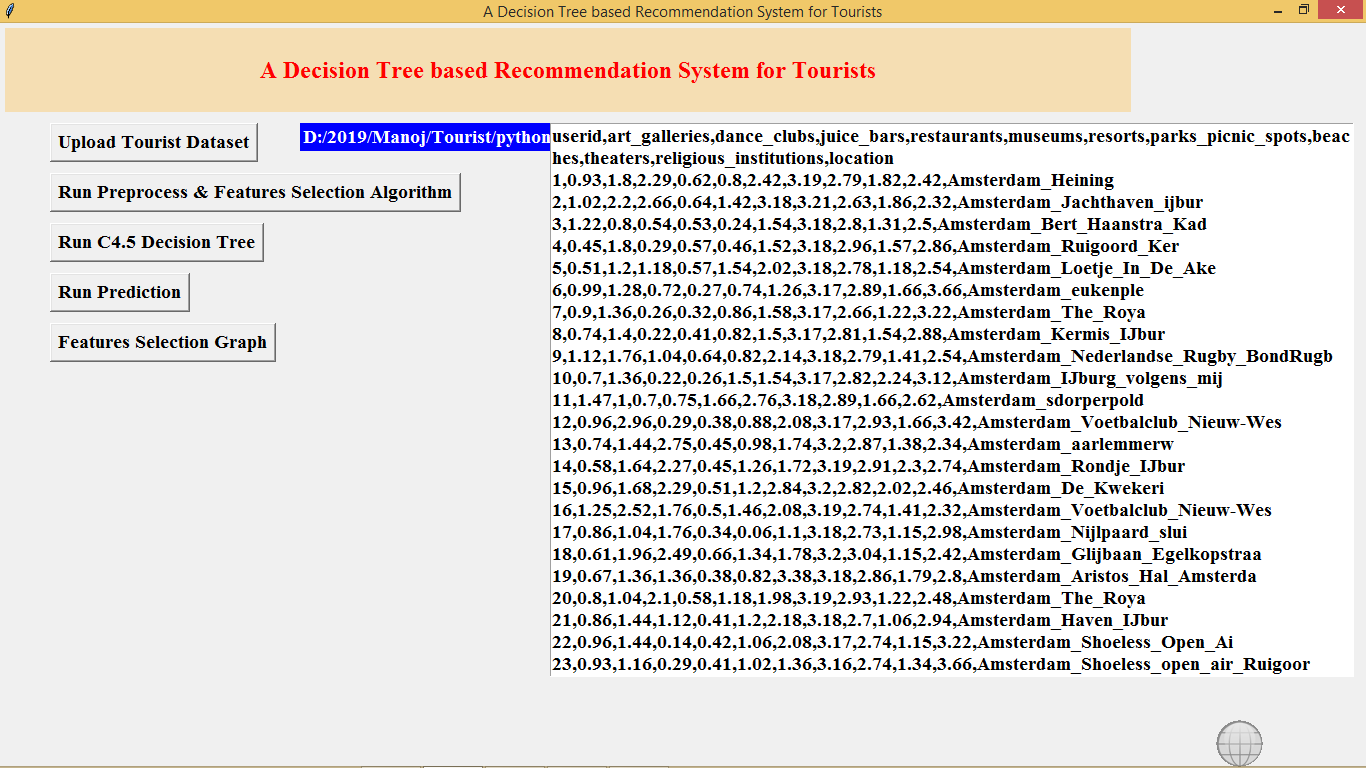
Double click on ‘run.bat’ file to get below screen



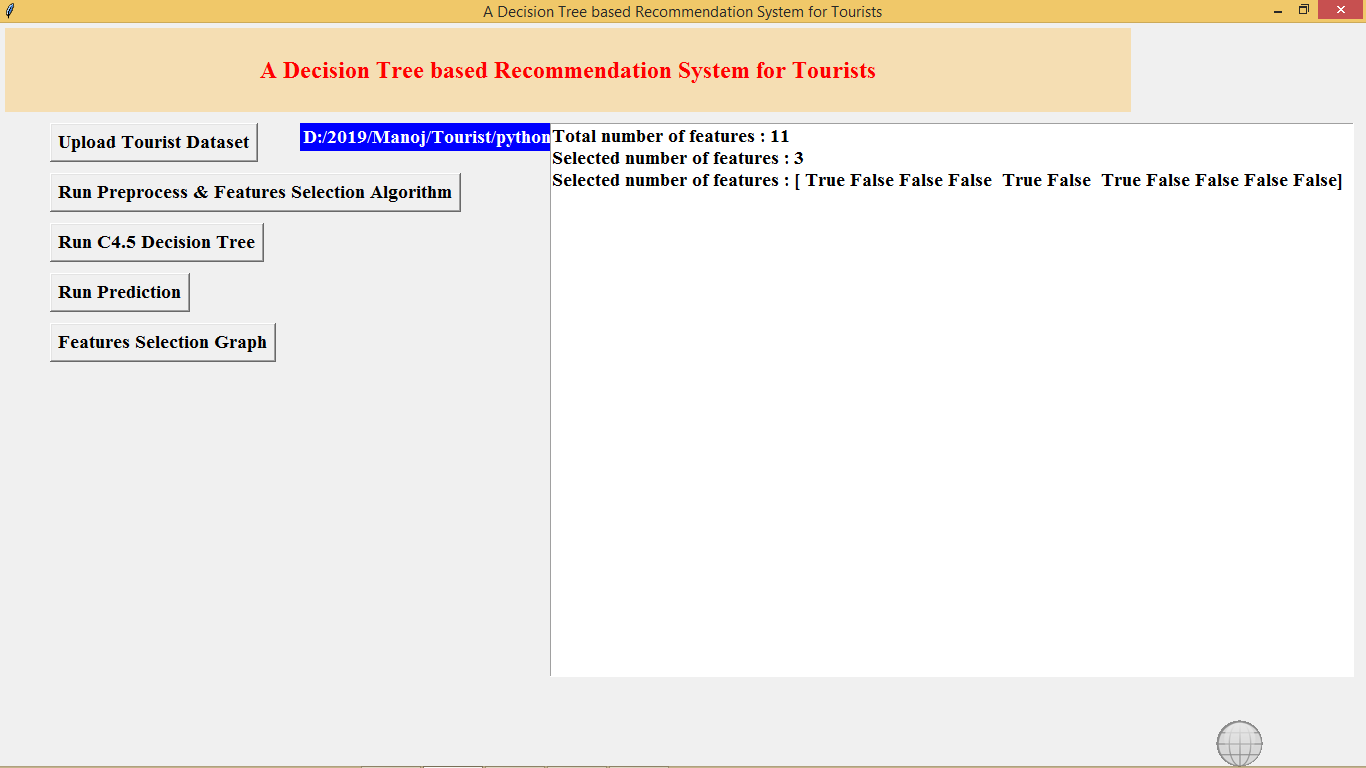
In above screen click on ‘Upload Tourist Dataset’ button and upload dataset file



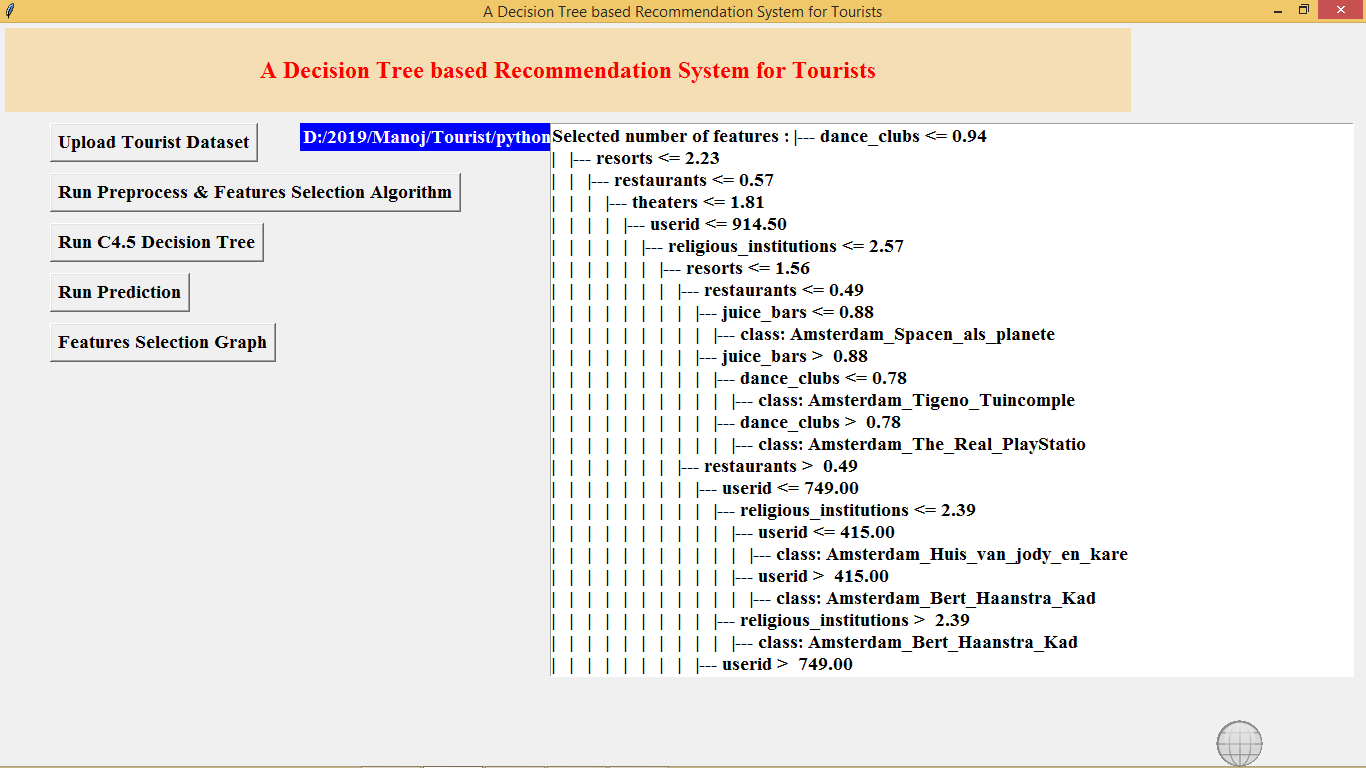
After file upload will get below screen with all dataset details



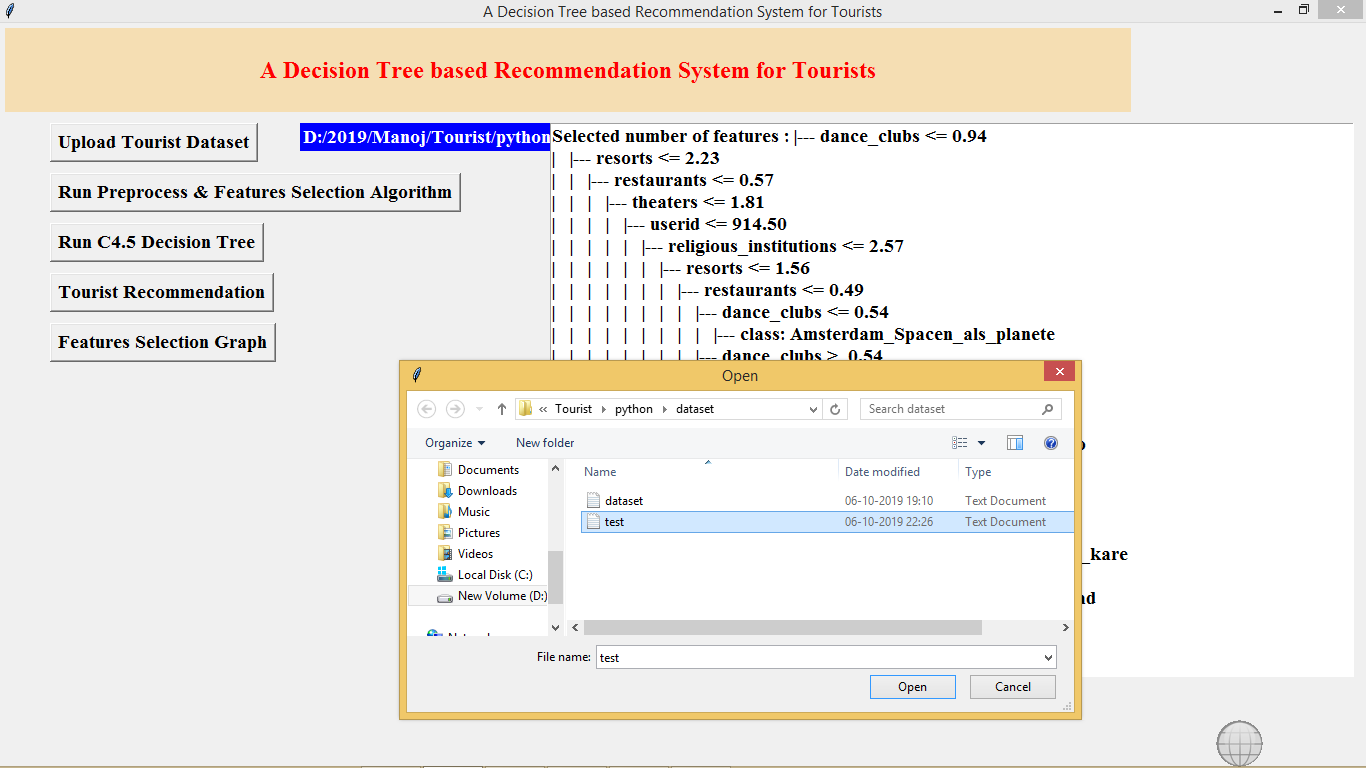
In above screen all users past experience dataset loaded and total 12 attributes are there in the dataset. Now click on ‘Run Preprocess & Feature Selection Algorithm’ button to remove empty values and reduce attributes size.



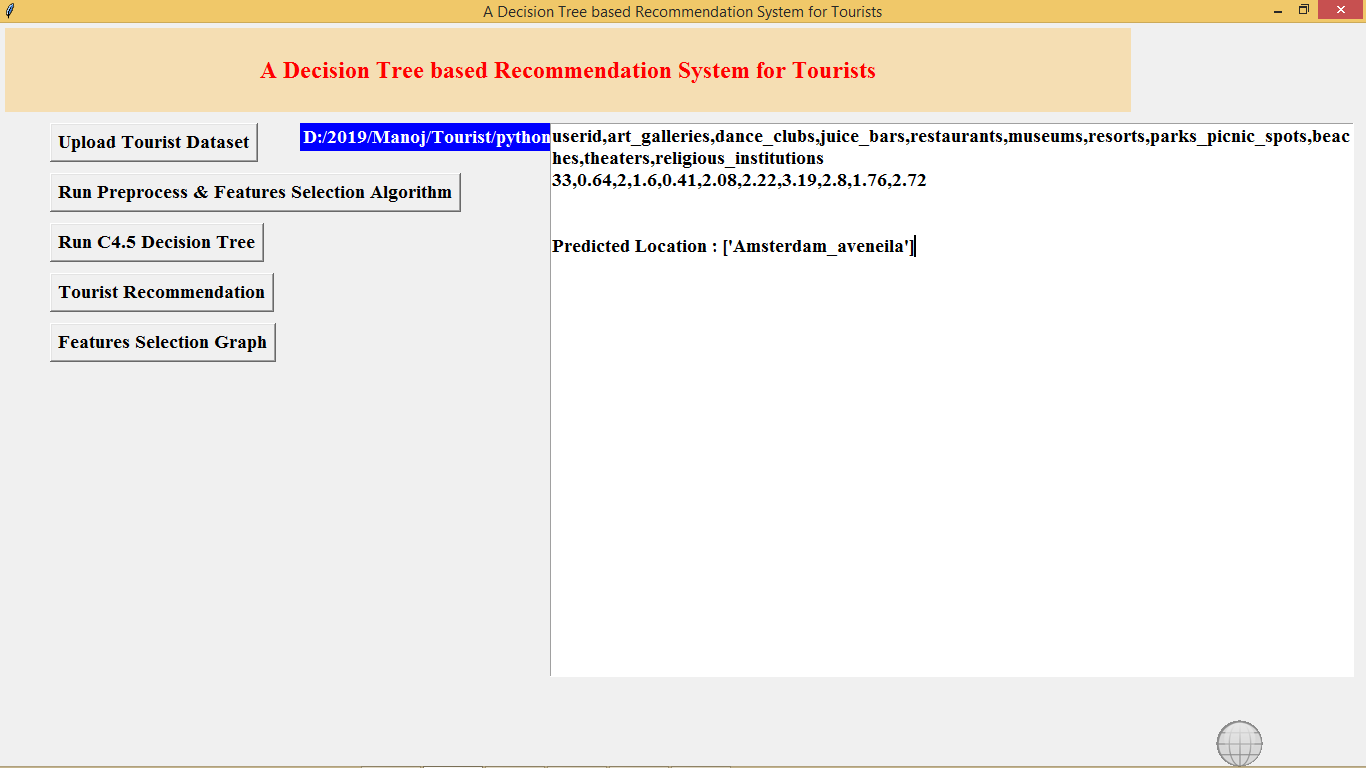
In above screen after applying MRMR features size reduces to 3 and only those attributes will be used whose column is TRUE and FALSE column will be ignore. Now click on ‘Generate C4.5 Decision Tree Model’ to build model



In above screen we can see using IF and ELSE statement decision tree has generated model. If > it will choose some decision if < it will choose some other decision. Now click on ‘Tourist Recommendation’ button to upload test file with no location name and application will predict it

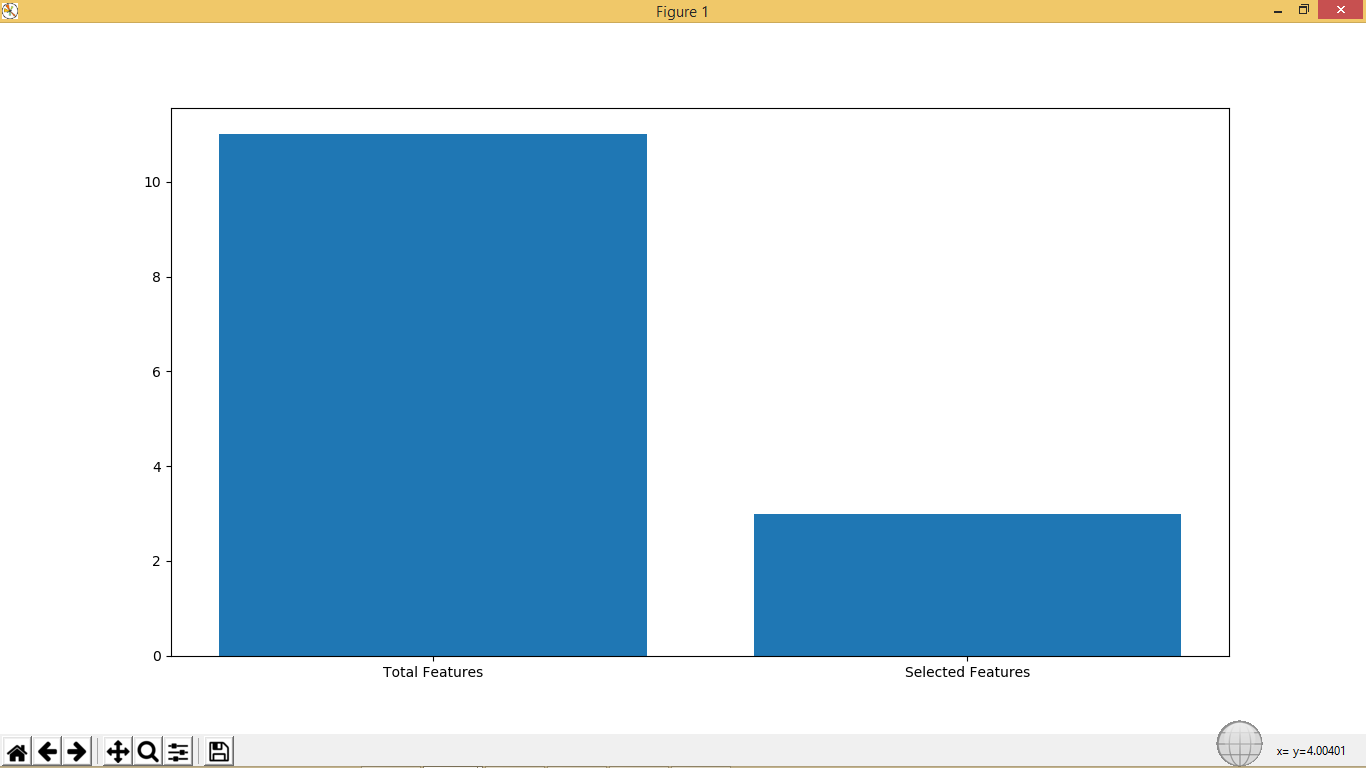


In above screen i am uploading test file now click open to get predicted or recommended location. In test file location name is not there application will give



In above screen after uploading test data we can see all values are there in test data but it not has location name and base on test values application predicted or recommend location name.

Now click on Features Selection Graph button to get below graph



In above graph x-axis contains total features and MRMR selected features and y-axis represents count of features and in above graph we can see after applying MRMR technique features size reduces to 3.