Semantics of Data Mining Services in Cloud Computing

In this paper author is describing about various cloud services which can be used to access Machine Learning algorithms and in this paper author is composing services from various cloud such as Amazon, Google Cloud, Microsoft and others. This services allows us to deploy our machine learning algorithms in their cloud and then expose service access URL and by using that service URL we can access those machine learning algorithms.

In traditional applications we need to purchase all hardware and software in order to run or build any application but this cloud services provides those hardware and software resources virtually and we not require to purchase any such hardware or software requirement and by just paying some amount (pay as much you use) to cloud providers we can deploy our application on their cloud and then our customers can access application from that cloud.

This cloud services are using by almost all fields such as health care system to store their patient records in cloud, Banking system and online social media applications. Due to this growing popularity all cloud peoples are providing services to deploy machine learning algorithms which can be used to analyse data and then predict class label from new records.

All cloud services may provide different composition services to deploy and access machine learning algorithms and this different services may confuse users from getting proper cloud. So in this paper author is giving some functions using which we can call services form various cloud. Author has introduce some functions such as Publish services, input values to machine learning algorithm, output values from machine learning algorithms and getting machine learning performance metrics such as accuracy. This cloud services allow us to access some free services also.

Based on security, authentication and pricing application can choose best cloud services to deploy and access machine learning algorithms.

In this paper author has used Amazon, Google cloud services which are free to deploy and run machine learning algorithms so I designed my own dummy cloud which get trained random forest machine learning algorithm on diabetes dataset. After training random forest algorithm we can publish this services by start cloud server.

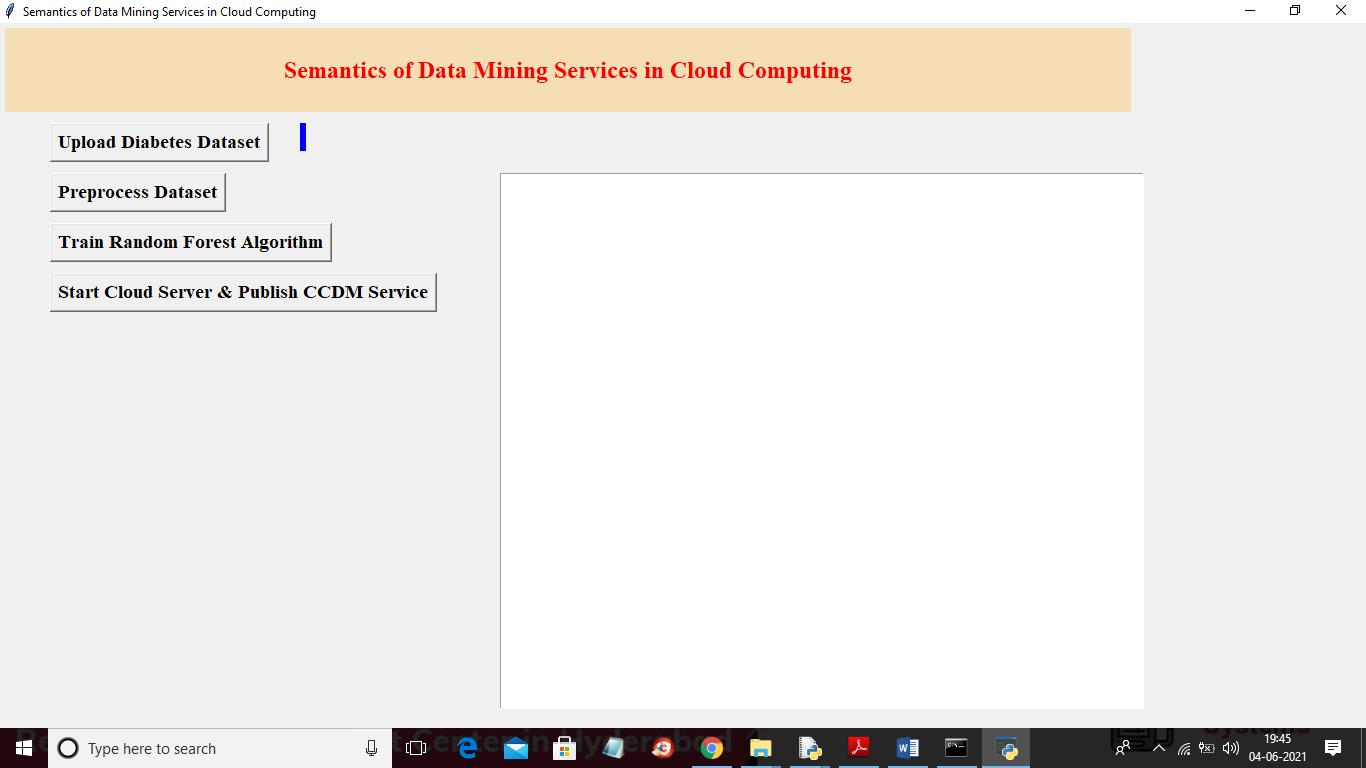
Any user from same or different system can upload test data from ‘User’ application and then this user application will connect to cloud machine learning service and then send input dataset to cloud and cloud will predict disease from test dataset.

To implement this project we have designed following modules

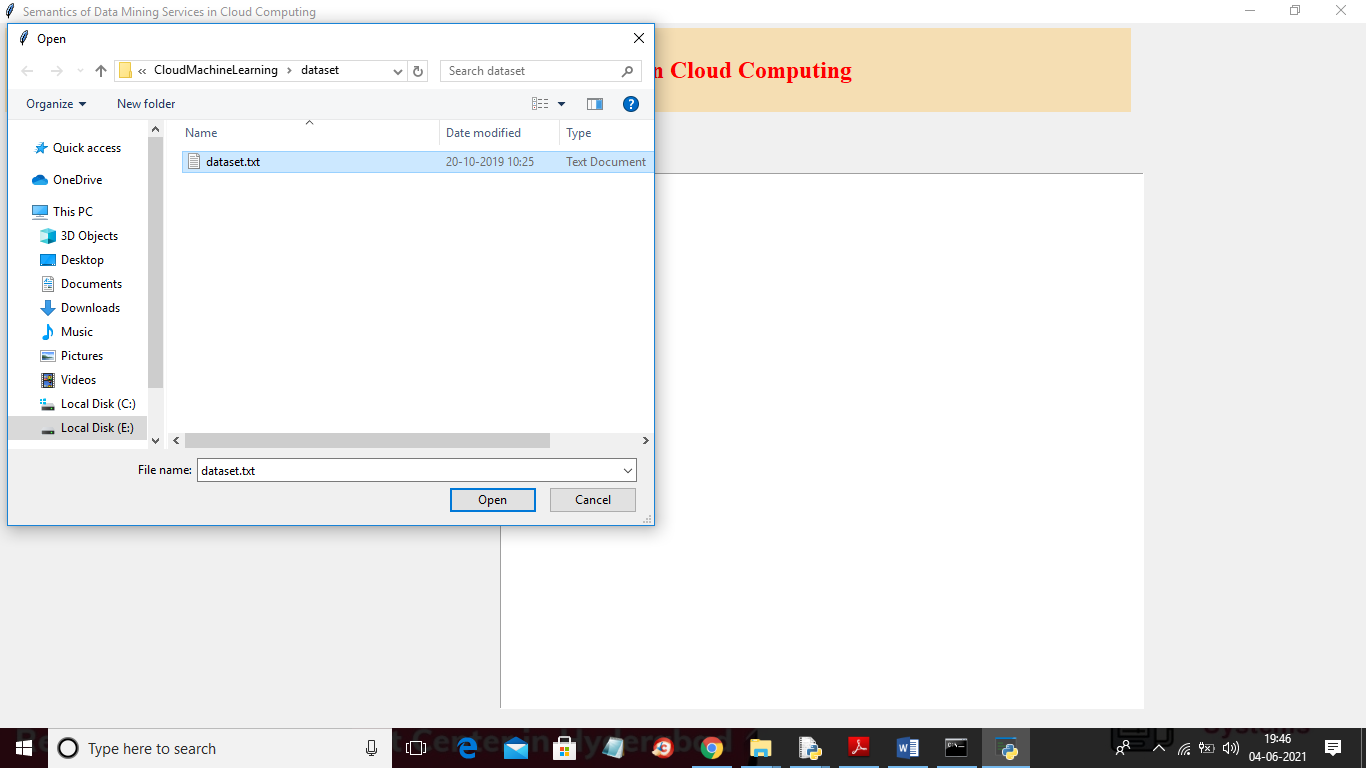
1. Cloud: this module will upload diabetes dataset and then train random forest algorithms and then publish cloud service. This service will accept input from user and then apply random forest to predict disease from test data.
2. User: this application will upload test data and then connect to cloud services and then send test data to cloud and cloud service will predict disease from test data.

SCREEN SHOTS

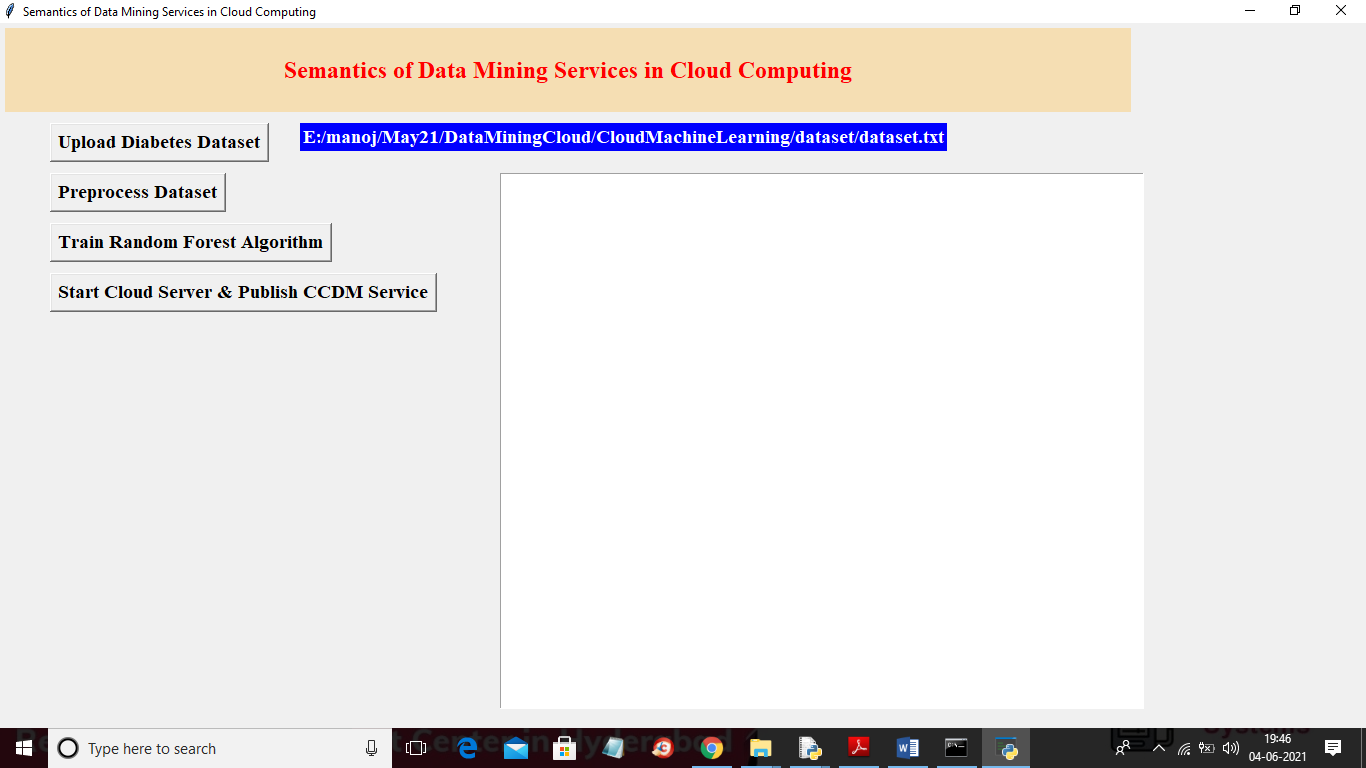
To run project double click on ‘run.bat’ file from ‘CloudMachineLearning’ folder to get below screen



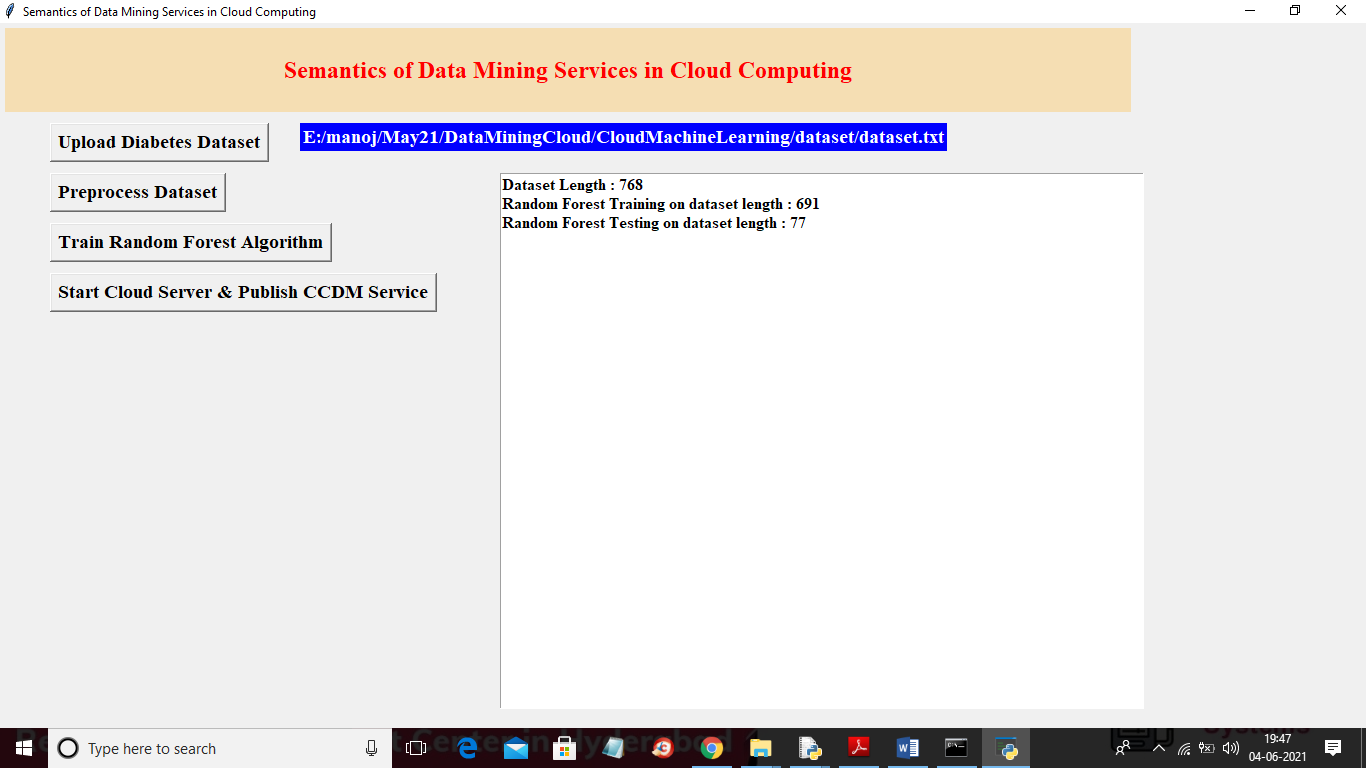
In above screen click on ‘Upload Diabetes Dataset’ button to upload dataset



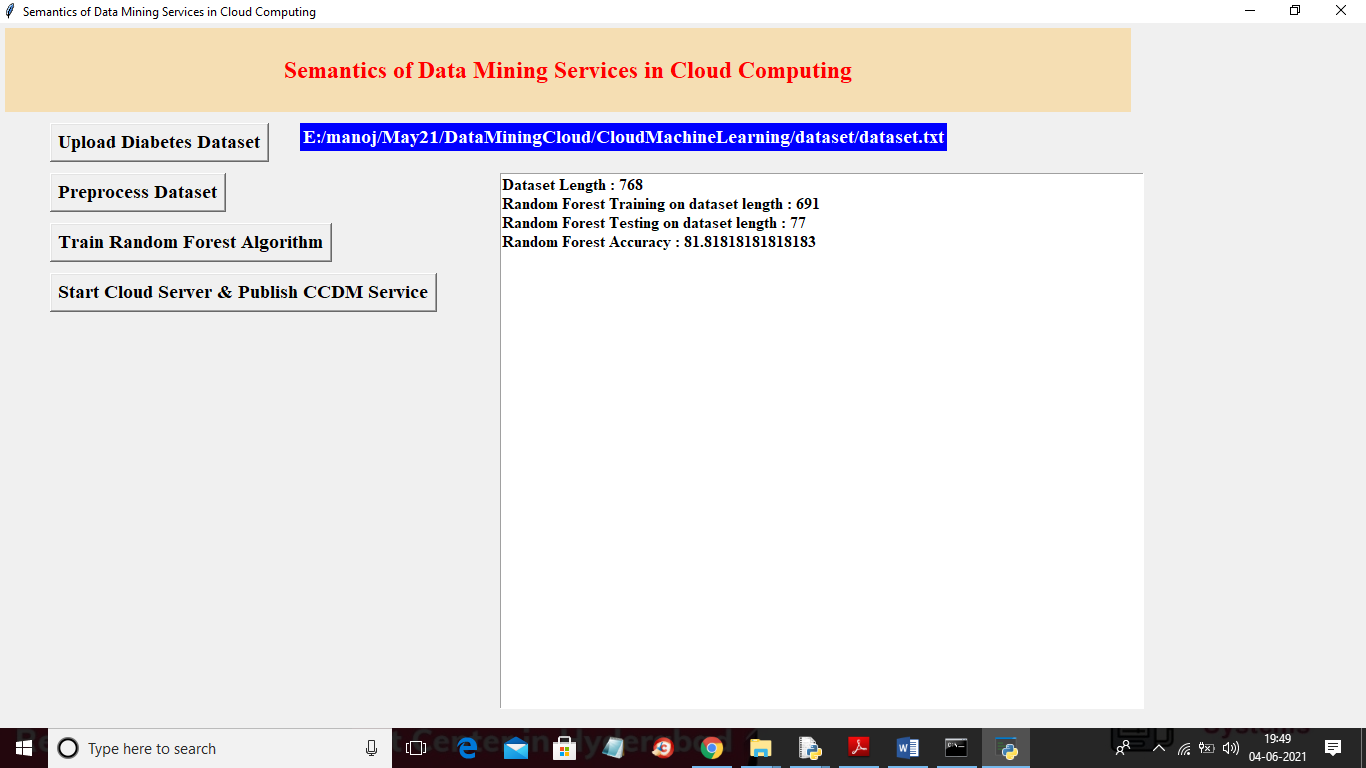
In above screen selecting and uploading ‘dataset.txt’ file and then click on ‘Open’ button to load dataset and to get below screen



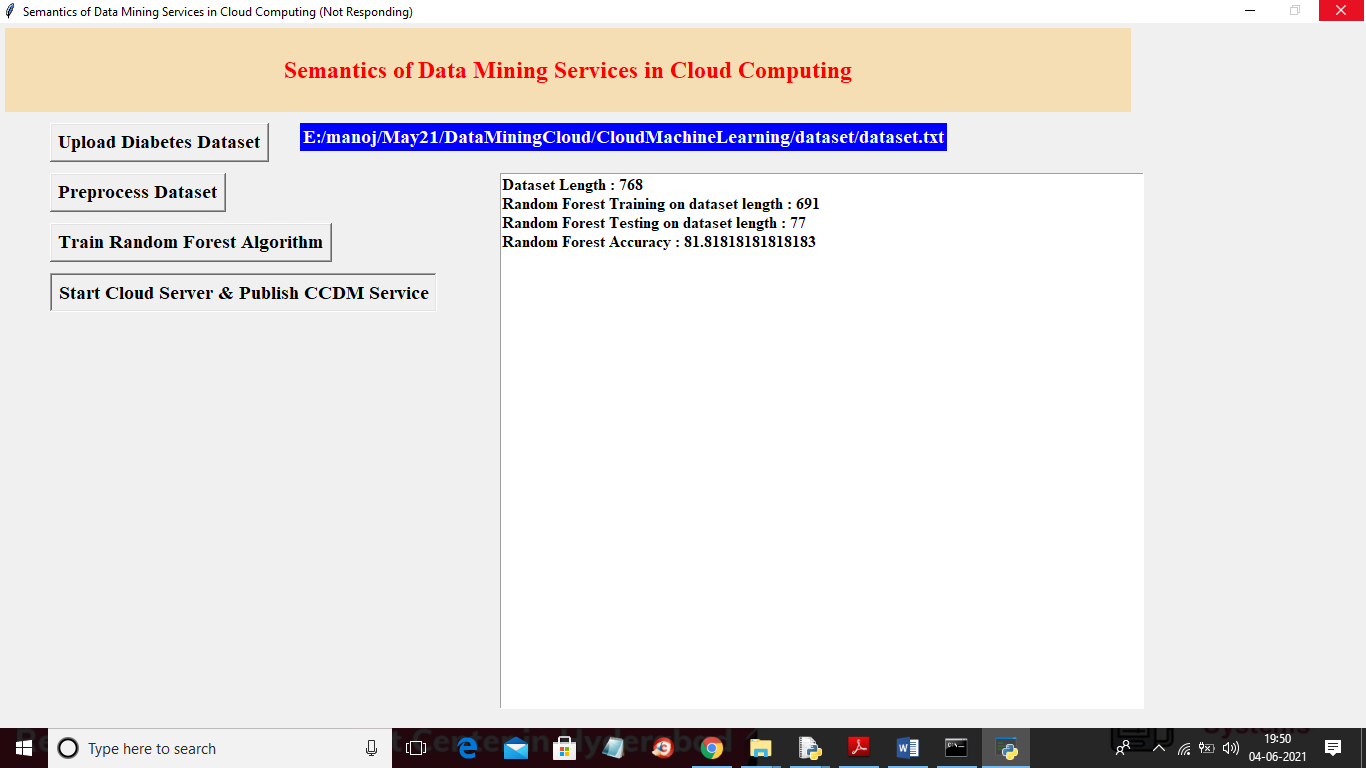
In above screen dataset loaded and now click on ‘Preprocess Dataset’ button to process dataset to remove empty values and then split dataset into train and test



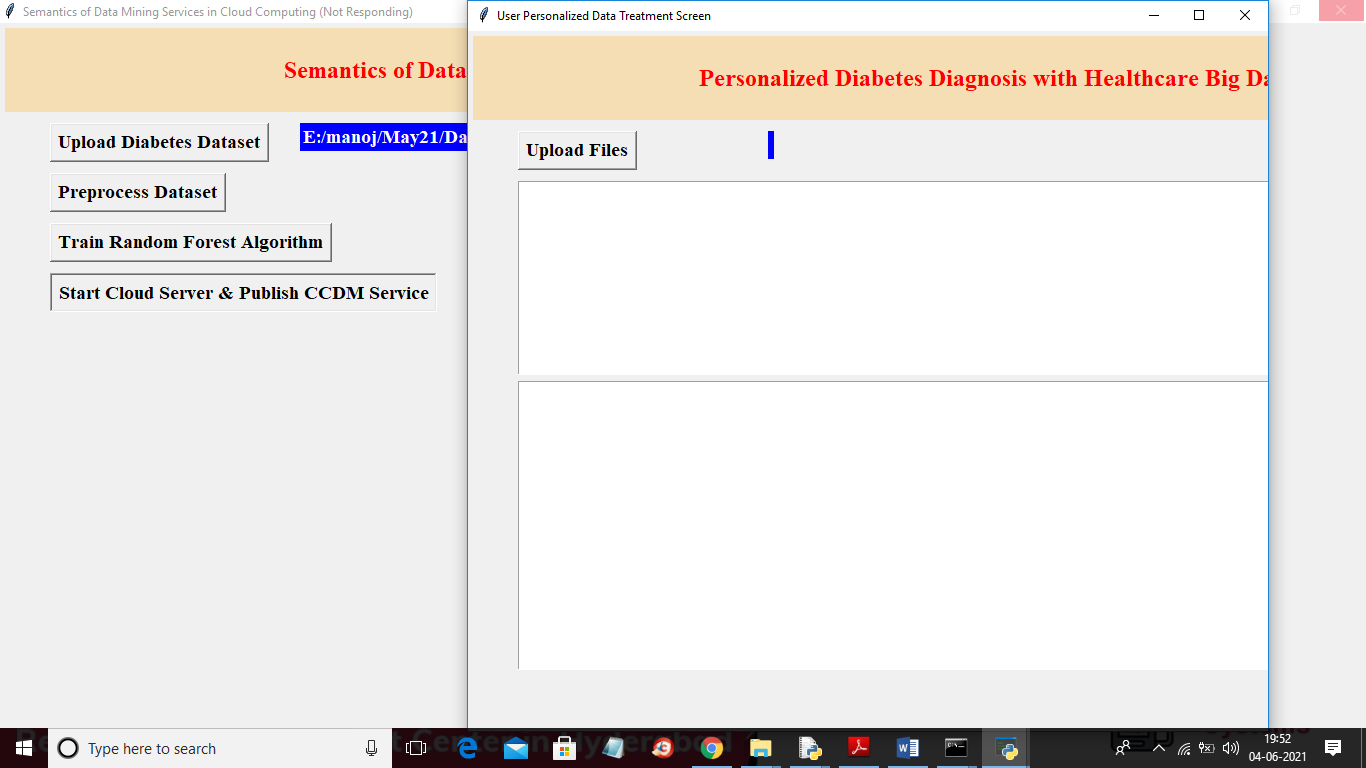
In above screen we can see dataset contains 768 records and then application using 691 (80%) records for training random forest and then using 77 (20%) records for testing random forest accuracy and now click on ‘Train Random Forest Algorithm’ button to train random forest



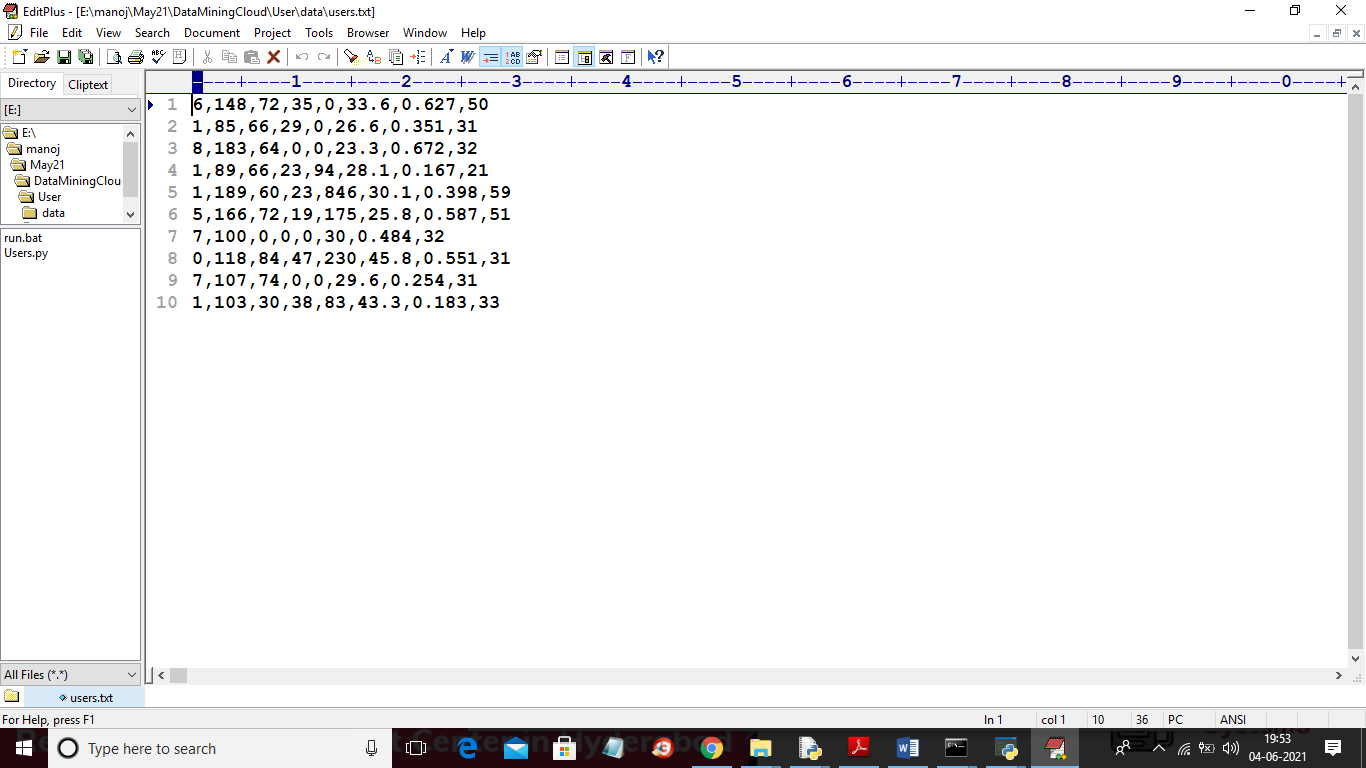
In above screen random forest is trained and we got its accuracy as 81% and now random forest Machine learning model is ready and now click on ‘Start Cloud Server & Publish CCDM Service’ button to start cloud server and to publish service



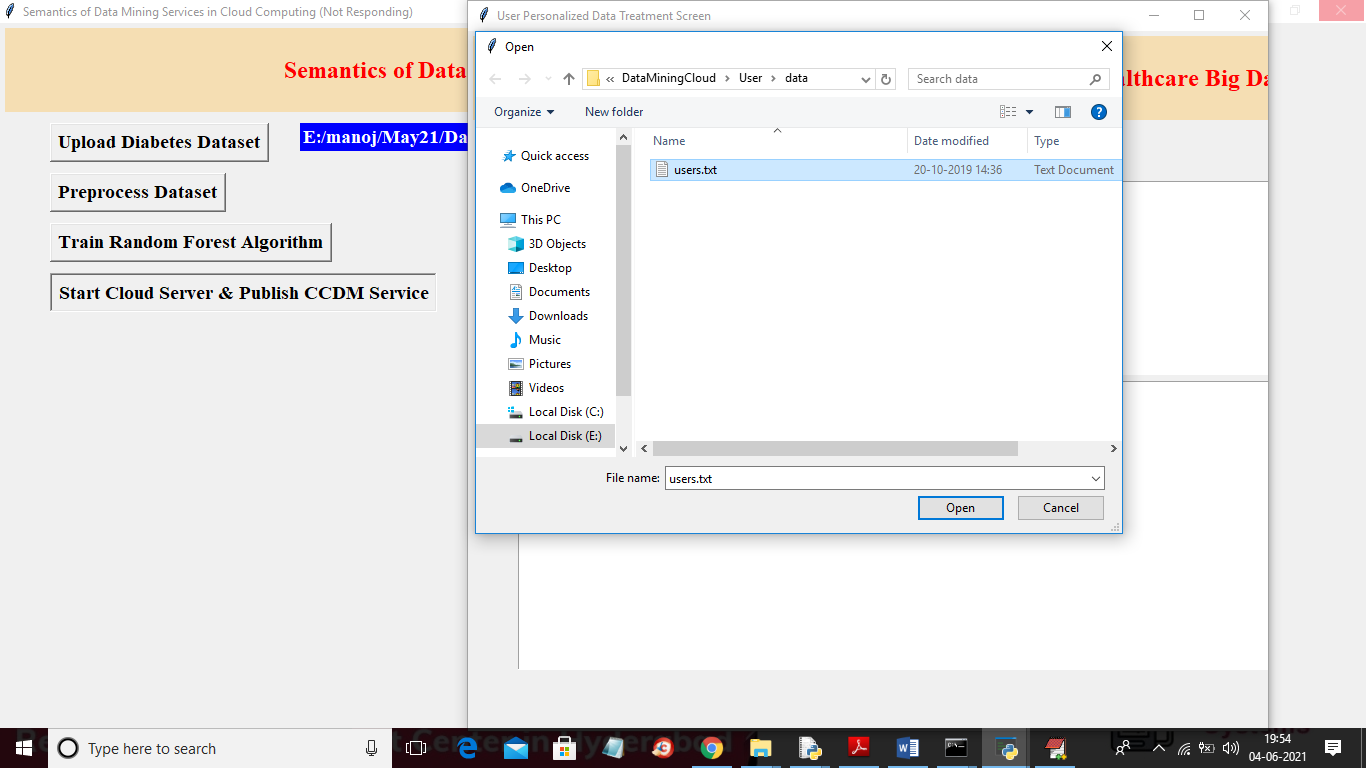
In above screen cloud server will start and runs in infinite loop to received request from users and now double click on ‘run.bat’ file from ‘Users’ folder to start user application



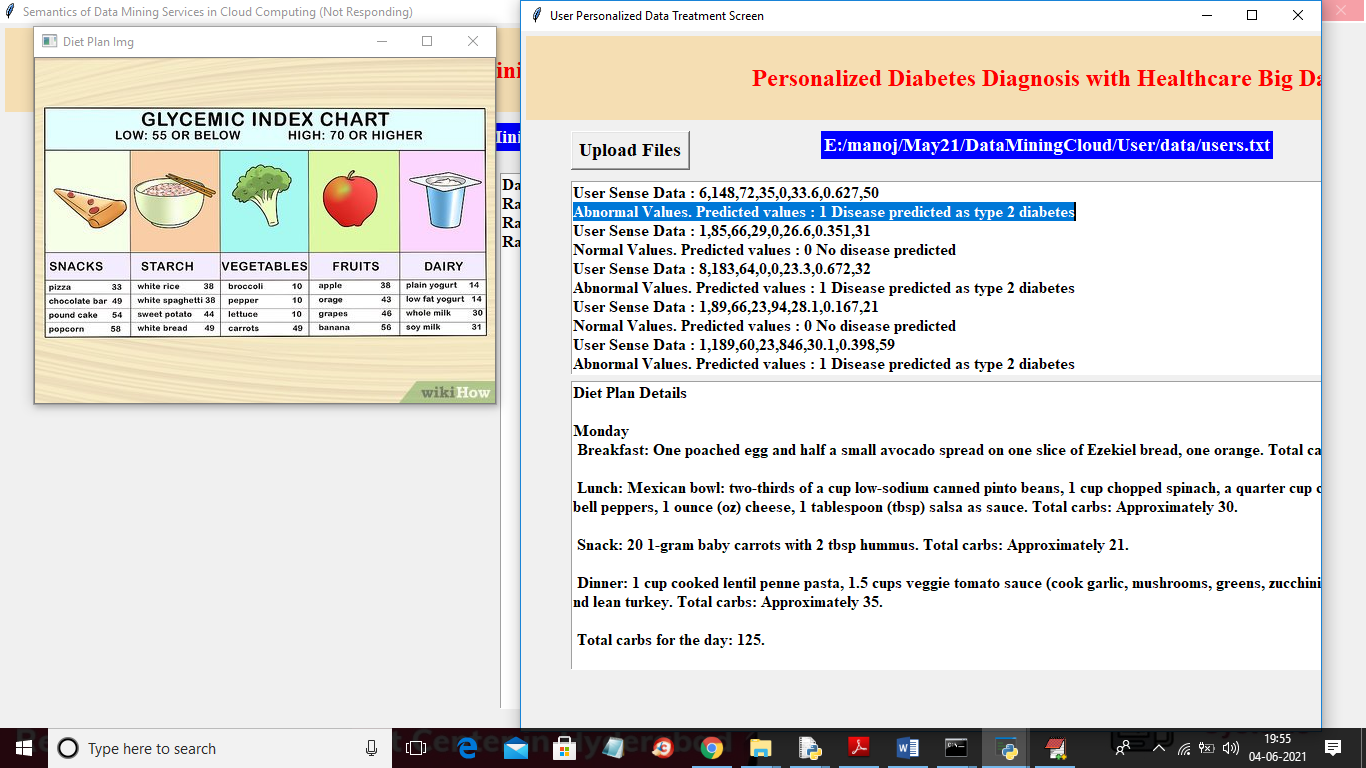
In above user screen click on ‘Upload Files’ button to upload test data and then user application will connect cloud service and the send request and cloud service will predict disease using random forest and send response back to user. Below is the test data used to send to cloud service



In above test dataset we have diabetes test values but don’t know presence of disease and cloud random forest will analyse above test data and give result



In above screen selecting and uploading ‘users.txt’ test data and then click on ‘Open’ button to load dataset and to get below result



In above screen in selected blue colour text we can see predicted disease details and this prediction done for each records and in second text area application suggesting diet for different stages of diabetes as type 1 or 2. In above image we can see food images for different diabetes type.

Similarly any cloud service provider can build such application and then publish URL and then user application can access that URL to get prediction from machine learning algorithms running on cloud. Client or user application can choose service URLS based on authentication, pricing and security. Here clouds are not free so we build our own cloud and if we have multiple clouds then we can choose service base on pricing.