**CS560 - Knowledge Discovery and Management**

**Mid Term Report**

**Problem Statement:**

1. Selection of datasets from at least two different domains (e.g., twitter and Youtube)

2. Big data integration and aggregations models and algorithms

3. Predictive/recommendation models and algorithms

4. Mobile App/Web design

**Solution:**

Heart Disease Finder Idea :

The main idea of the project is to predict the presence/absence of heart disease with a given list of parameters from the User. Once the prediction is done, if the user is liable for a heart disease, the app tends to find all the possible heart disease and the respective services provided for the disease. The location of the nearest hospital is found and plotted in map, so that the user can go for a checkup. The cost range of different service providers close-by the location can be listed. The app redirects to certain preventive/dietary measures in case of heart diseases.

1. **Design**
2. **DataSet Description: (Health Domain)**

Heart Disease Prediction:

1. Age of the user  (Input Value)  
 2. Gender of the User (Selection -Male, Female)  
 3. Chest pain type (Selection)

* + Value 1: typical angina
  + Value 2: atypical angina
  + Value 3: non-anginal pain
  + Value 4: asymptomatic

4. Blood Pressure (Input Value)  
 5. Cholesterol in mg/dl (Input Value)  
 6. Blood sugar > 120 mg/dl  (Selection – Yes, No)  
 7. Electrocardiograph (Selection)

* + Value 0: normal
  + Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
  + Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria

8. Maximum Heart Rate (Input Value)  
 9. Exercise induced angina (Selection – Yes, No)  
 10. Oldpeak -ST depression induced by exercise relative to rest (Input Value)

11. Slope of the peak exercise ST segment

* + Value 1: UpSloping
  + Value 2: flat
  + Value 3: DownSloping

12. No of major vessels (0-3) colored by fluoroscopy(Input Value)   
13.Thalassemia(Selection)

* + 3 = normal;
  + 6 = fixed defect;
  + 7 = reversible defect

**DataSet 2: (Cost/Expense/Location by service providers)**

<https://data.cms.gov/Medicare/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3>

This dataset provides all the services provided by the provider(hospital), provider name, id, address, etc.

Features selected:

1. Heart disease services provided
2. Provider Address
3. City of service
4. State of service
5. Amount charged
6. **Integration model and algorithm**

From the machine learning of the first dataset, the output will be if the user is liable to get heat disease or not. If the user is more liable to get heart disease, the second dataset gives the possible diseases and the services available in each state and city with the cost. If the user selects a particular services, app plots the nearest service provider on google map to get treated. The 2nd dataset can be trained to give the cost of the service selected too.

1. **Classification/recommendation model and algorithm**

Used naive bayes classification algorithm.The model classifies if the user had chances of heart disease. It also finds the nearest hospital for the required service. The cost for the service can also be computed. The application also gives preventive measures (mash up) by showing the disease prevention control web pages.

1. **Selection of datasets**

DataSet1: <http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/>

The above data set has the prediction of heart disease with 13 different attributes from the user.

DataSet2**:** <http://www.phc4.org/dept/dc/docs/list35.pdf>

The above dataset gives the disease and services provided for the diseases. This dataset is used for integrating dataset1 and dataset3.

DataSet3: <https://data.cms.gov/Medicare/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3>

This dataset provides all the services provided by the provider (hospital), provider name, id, address, etc.

1. **Mobile App/Web design :**

The main idea of the project is to predict the presence/absence of heart disease with a given list of parameters from the User. Once the prediction is done, if the user is liable for a heart disease, the app tends to find all the possible heart disease and the respective services provided for the disease. The location of the nearest hospital is found and plotted in map, so that the user can go for a checkup. The cost range of different service providers close-by the location can be listed. The app redirects to certain preventive/dietary measures in case of heart diseases.

**2. Features Implemented**

1. **Classification algorithms :**

Used Naive bayes classification algorithm to analyse the data and predict the presence/absence of heart disease. The range of cost(high/low/medium) can also be done for the second dataset with the same classification algorithm.

1. **Solr indexing: Your own data services**

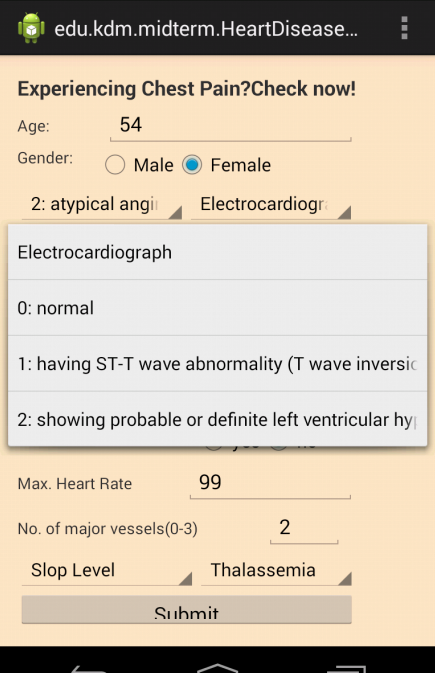
<http://134.193.136.127:8983/solr/collection1_shard1_replica1/select?q=id%3Amkdn9&wt=json&indent=true>

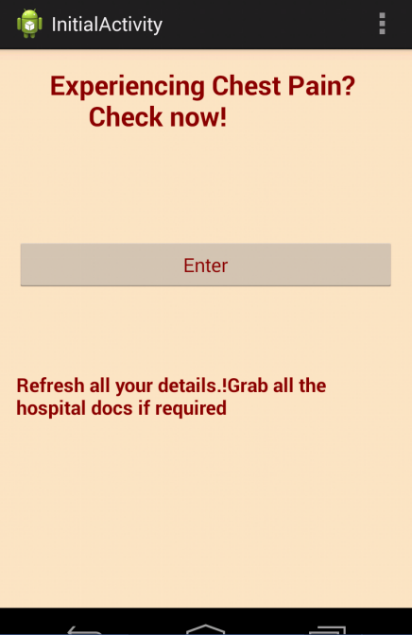
<http://134.193.136.127:8983/solr/collection1_shard1_replica1/select?q=id%3Amkdn10&wt=json&indent=true>

1. **Your own design**

The main idea of the project is to predict the presence/absence of heart disease with a given list of parameters from the User. Once the prediction is done, if the use is liable for a heart disease, the app tends to find all the possible heart disease and the respective services provided for the disease. The location of the nearest hospital is found and plotted in map , so that the user can go for a checkup. The cost range of different service providers close-by the location ca be listed. The app redirects to certain preventive/dietary measures which can be started in case of heart diseases.

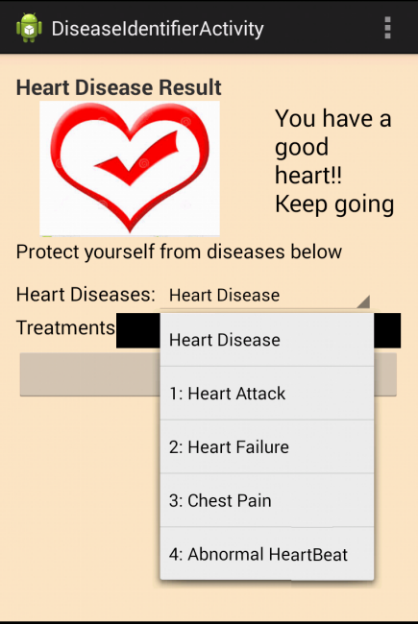
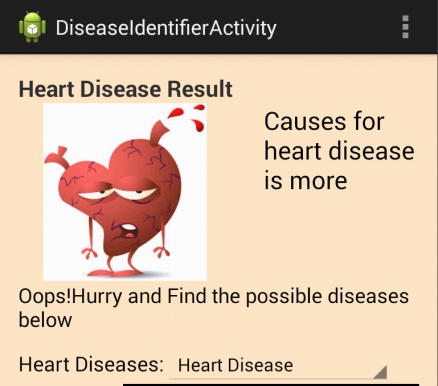
**3. Outputs: description with screenshots of the Features :**

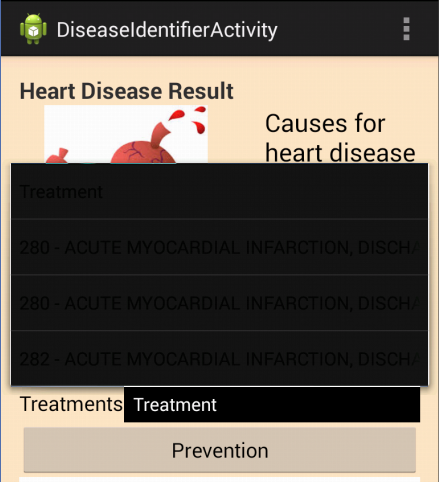
 First page of the application. User details retrieval



The User heart is fit and good.

The User heart has a possibility of heart disease

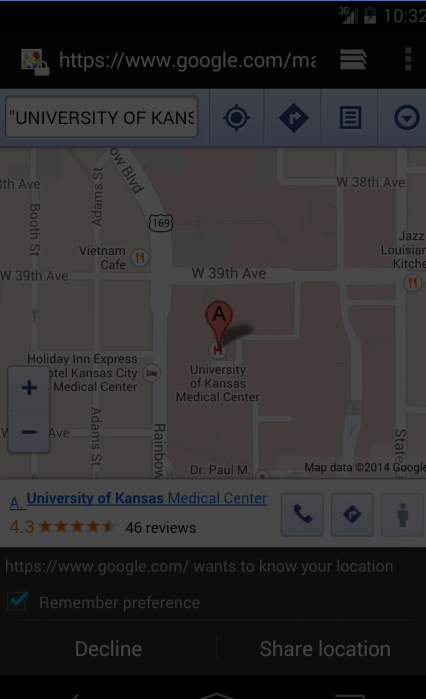


4 types of heart diseases found:

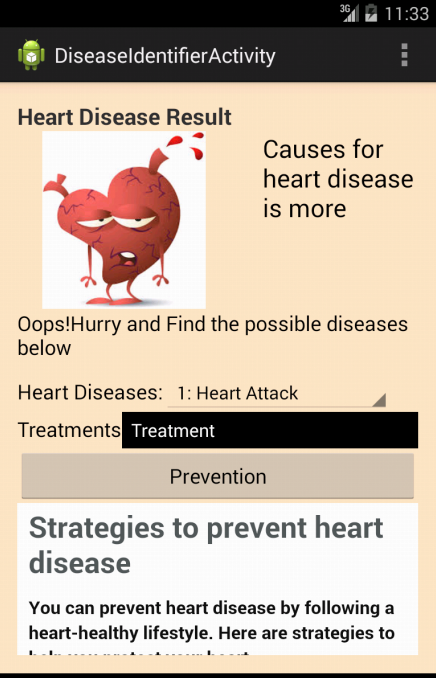
|  |
| --- |
| heart attack |
| heart failure |
| chest pain |
| Abnormal Heartbeat |

The respective services for each category of

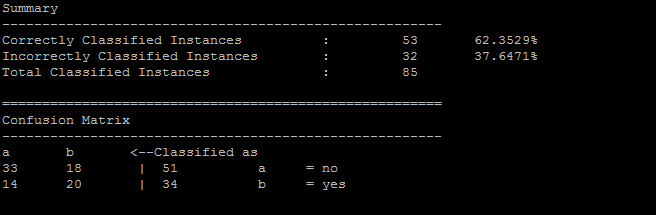
heart disease is found, with service provider. The cost can also be estimated.

Google map of the service provider: 

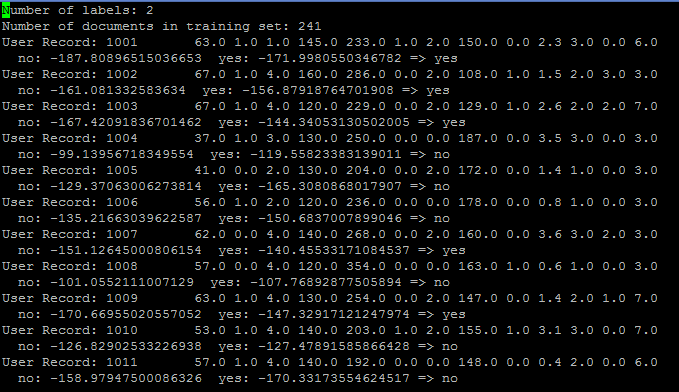
Preventive measures in a webview:



Machine learning on Data:



The analysis of heart-disease of the user.



**4. Your midterm Github URL :**

<https://github.com/CS560KDM-Assignments/MidTerm>

**5. References**

<http://archive.ics.uci.edu/ml/>

<http://data.gov>

<https://github.com/fredang/mahout-naive-bayes-example>

<http://chimpler.wordpress.com/2013/03/13/using-the-mahout-naive-bayes-classifier-to-automatically-classify-twitter-messages/>

<http://www.mayoclinic.org/diseases-conditions/heart-disease/in-depth/heart-disease-prevention/art-20046502>