Introduction

Search

Today's we have lots of data available in every field of study. With so many data we have access to any information, but large chunk of data has its own challenges. Due to large amount of data search for specific information has become challenging. To speed up search and retrieve the accurate data many search algorithms has developed.

Classification

Classification with numerical value is easy with so many states of art algorithm available like clustering, decision tree etc. When it come to text classification it become difficult to achieve reasonable performance. Classification is important part for to narrow down our search and can play important role to reduce search time.

Recommendation

Historically, in the past, we have been depended to another person for a good movie recommendation. With the advancement of technology and huge data now we can create an automated recommended system which can recommend a good movie specific to user choice. We don't have to rely to another person's opinion about which movie I might like, but this can be done automatically by using the recommender system and the best part of this system it will recommend a movie based on user interest. Movie recommendation will work according to user previous like, dislike and movie history and we can also integrate same user interest to make recommender engine more diverse.

Algorithm

Search

I used tf-idf VSM model for search the query from database. It follows bag-of-words model architecture.
Tf (term frequency) = number of times term appear in document / total number of terms
Idf = Inverse document frequency reduce impact of frequently occurring word
Idf is define as $\log(N/df_t)$
where N is number of documents
df_t = number of document terms appears

We created vector space model based on tf-idf values and calculate cosine similarity between document and query

Classification

Classification problem can be defined as "deciding class for unknown data based on its feature and learning from old known data and its classes". As an example, lets we have a dataset of spam and regular email. After looking into these data set, we decide some feature specific these data like sender, subject line, external link and domain. These features can act as a dimensionality of dataset.

Naïve Bayes Classification

Naïve Based classification assume each feature (in many literatures also refer as dimension are independent of each other). It is simple probability-based classifier. It uses Bayes probability theorem to solve classification problem.

Mathematical representation of Bayes Theorem
$P(A|B) = P(B|A) * P(A) / P(B)$
According to naïve bayes each feature is independent of each other
$P(y|x1,x2,.....xn) = P(x1|y)P(x2|y)..P(xn|y) P(y) /(P(x1)P(x2).....P(xn)$

Recommendation
Recommendation is being used to suggest its user a product which he is most likely to be interested. Recommendation engine is developed based on user profile and product he liked in past.
- Content based recommendation
- User Collaborative recommendation

Content based recommendation
In content-based recommendation user get recommendation based on his selection about product in past.

User Collaborative recommendation
In this approach first we calculate similarity between two users based on common item they liked. After calculation similarity we recommend product which is like by most similar user to another user.

Server
I used flask for server creation. It is easy to work upon.
- Refer http://flask.pocoo.org/docs/1.0/installation/ for installation and setup guide

Server hosting
PythonAnywhere can be used with python flask to add domain support for you online application and server creation. Using PythonAnywhere is very easy to add backend support for your application. PythonAnywhere provide free account with limited access. You just need to create a beginner account. Create a beginner account. Once account is created click on web app and create application with flask with your required python version. (I done it with python 3.6)

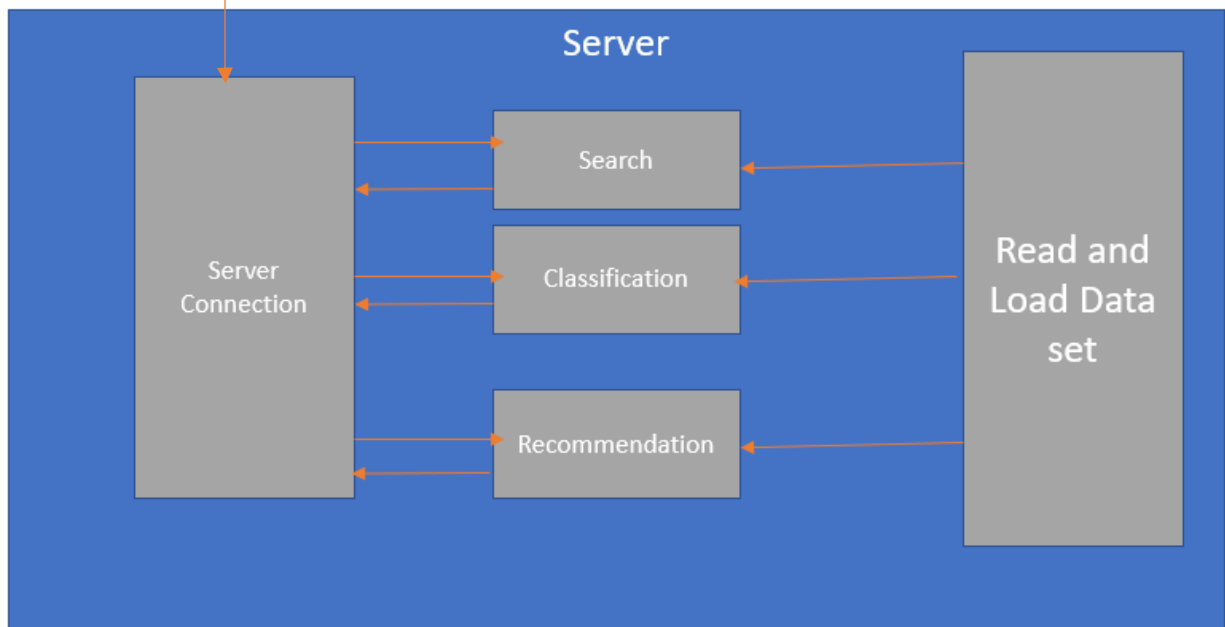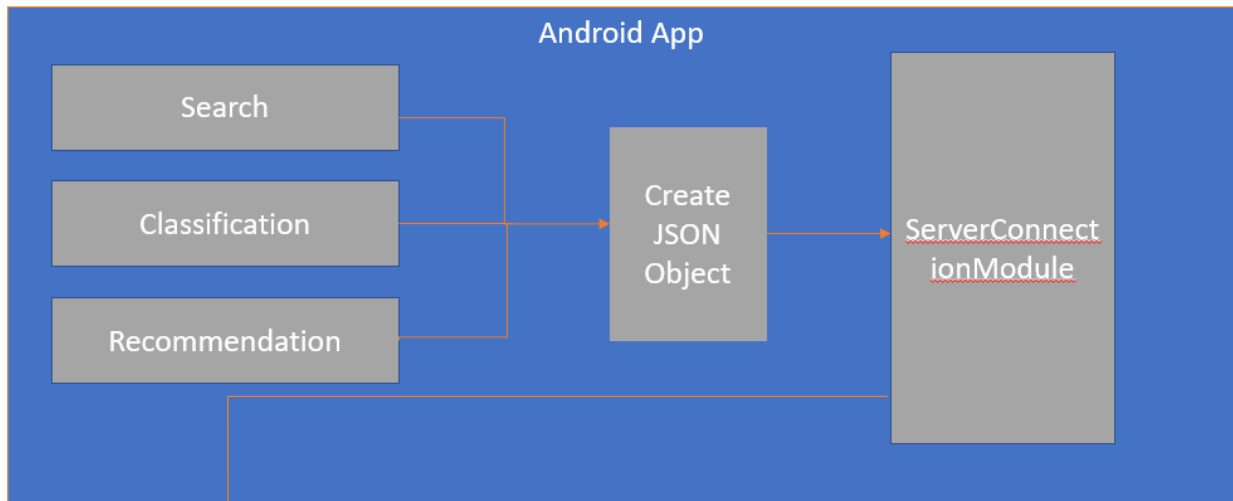Your default project working directory will be home/username/mysite
Add your all python file to your working directory
My server link: http://bhaskartrivedi.pythonanywhere.com/

Android Application
I have developed android application to communicate to server and show result of search, classification and recommendation to user.
I connect to server using java HTTPAsyncTask class. I send data using POST method in json object and receive result from server in the form of json object

Client to data subsets

Android App

| Search |
| Classification |
| Recommendation |

Create JSON Object

ServerConnectionModule

Server

Server Connection

| Search |
| Classification |
| Recommendation |

Read and Load Data set

Home screen

7:49 AM

ServerConnection

HOME SCREEN

SEARCH

CLASSIFICATION

RECOMMENDATION

Accept Query

Search Result

ServerConnection

Movie1

Toy Story

Led by Woody, Andy's toys live happily in his room until Andy's birthday brings Buzz Lightyear onto the scene. Afraid of losing his place in Andy's heart, Woody plots against Buzz. But when circumstances separate Buzz and Woody from their owner, the duo eventually learns to put aside their differences.

Movie2

Recommendation

7:50 AM ... ⊠

ServerConnection

Recommended Movies

Private Benjamin (1980)

Santa Clause, The (1994)

Muppet Treasure Island (

Crow, The (1994)

Babe (1995)