JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY
NOIDA

MINOR PROJECT 2018-2019

REPORT

ONLINE COURSE RECOMMENDER

SUBMITTED TO:

- DR PRAKASH KUMAR

SUBMITTED BY:

- YASHWANT WARDHAN (16103011) (B8)
- PRASHANT RATHI (16103021) (B8)
- AYUSH RAJ (16103281) (B8)
- VIDISHA NAINWAL (16103321) (B8)

ACKNOWLEDGEMENT

We have taken efforts in this project, Online Course Recommender. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

We are highly indebted to Dr. Prakash Kumar for his guidance and constant supervision as well as for providing necessary information regarding the project and also for his support in completing the project.

We would like to express our gratitude towards our parents and friends for their kind co-operation and encouragement which help us in completion of this project.

We would like to express our special gratitude and thanks to college persons for giving us such attention and time.

Our thanks and appreciations also go to our colleague in developing the project and people who have willingly helped us out with their abilities.

TABLE OF CONTENTS

INTRODUCTION

The project aims to fulfil United Nations Development Programme's one of the sustainable goals i. e. **Quality Education**. It also to provide equal access to affordable vocational training, to eliminate gender and wealth disparities, and achieve universal access to a quality higher education.

We aim to achieve a quality education platform for the real comparison and effective course recommendation according to user's interests and web based data activity. Basically we are implementing web data crawling and applying clustering and recommendation algorithms on the data secured. In this way proper and advanced quality reaches to the last node of modern engineering and various stream enthusiasts and bachelors as well as masters.

GENERAL FEATURES

- Comparing various courses from different websites.
- Recommended courses on the basis of user's work history and activity
- Recommendation based on ratings and comments
- Allows to rate and review
- Advanced front end development and user interface

LANGUAGES AND CONCEPTS USED

- **Front end** of the website using **HTML, CSS, Bootstrap**
- **Back-end** of the website using **Node.js, SQL**
- **Machine Learning** using **Naïve-Bayes** Algorithm in Python
- **Web Crawling and Scraping** using **Scrapy** library in Python

FRONT-END OF THE WEBSITE

The front-end of our project is mainly based on HTML, CSS and Bootstrap. We have made sure to provide a good user interface on our website. It is easy to use and understand. Front-end is basically what the user see on his/her screen. It is mainly the combination of many webpages put into a sequential and sensible manner.

HTML

Hypertext Mark-up Language (**HTML**) is the standard mark-up language for creating web pages and web applications. With Cascading Style Sheets (CSS) and JavaScript, it forms a triad of cornerstone technologies for the World Wide Web.

- HTML is the standard mark-up language for creating Web pages.
- HTML stands for Hyper Text Mark-up Language
- HTML describes the structure of Web pages using mark-up
- HTML elements are the building blocks of HTML pages
- HTML elements are represented by tags
- HTML tags label pieces of content such as "heading", "paragraph", "table", and so on
- Browsers do not display the HTML tags, but use them to render the content of the page

CSS

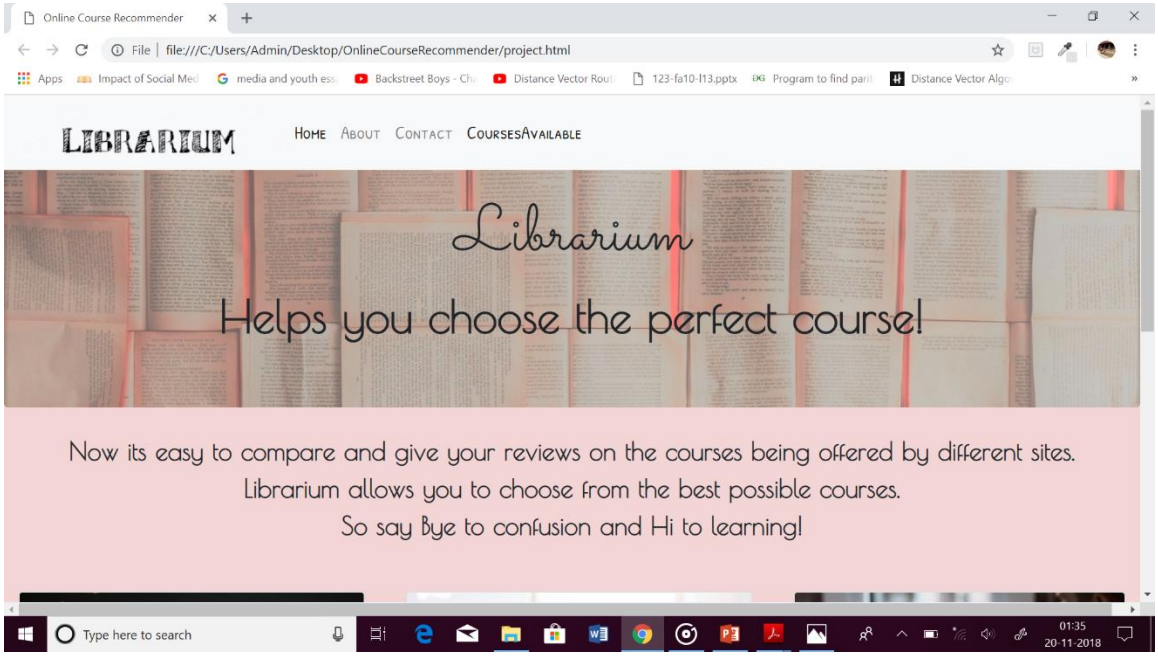
Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a mark-up language like HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.

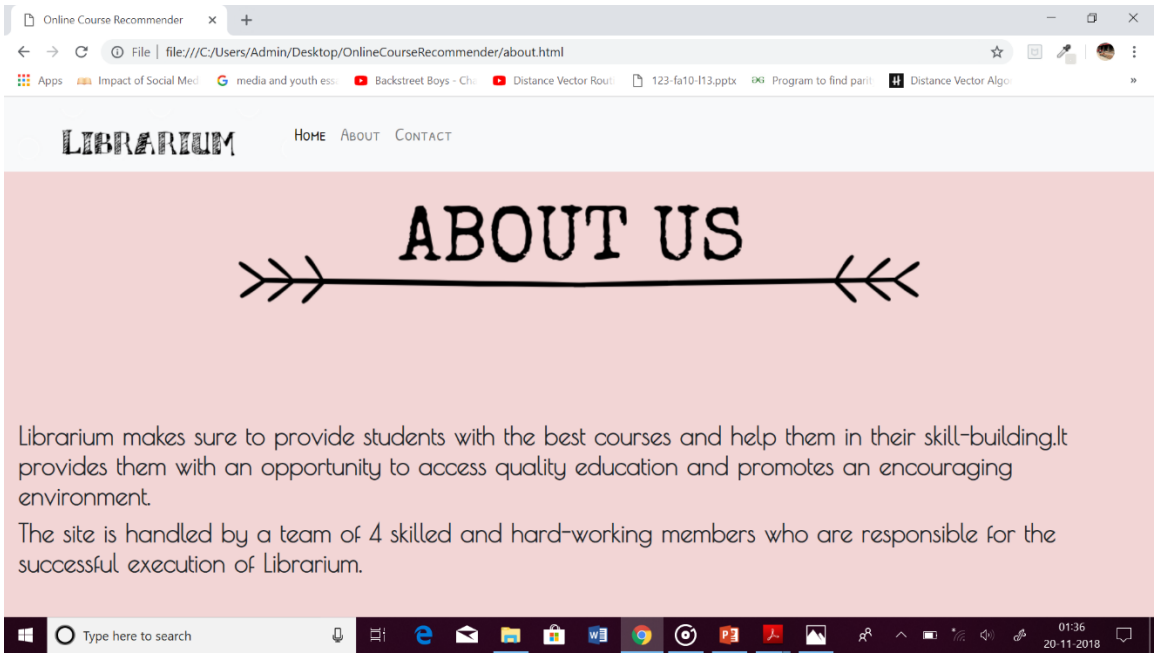
CSS is designed to enable the separation of presentation and content, including layout, colours, and fonts. This separation can improve content accessibility, provide more flexibility and control in the specification of presentation characteristics, enable multiple web pages to share formatting by specifying the relevant CSS in a separate .css file, and reduce complexity and repetition in the structural content.

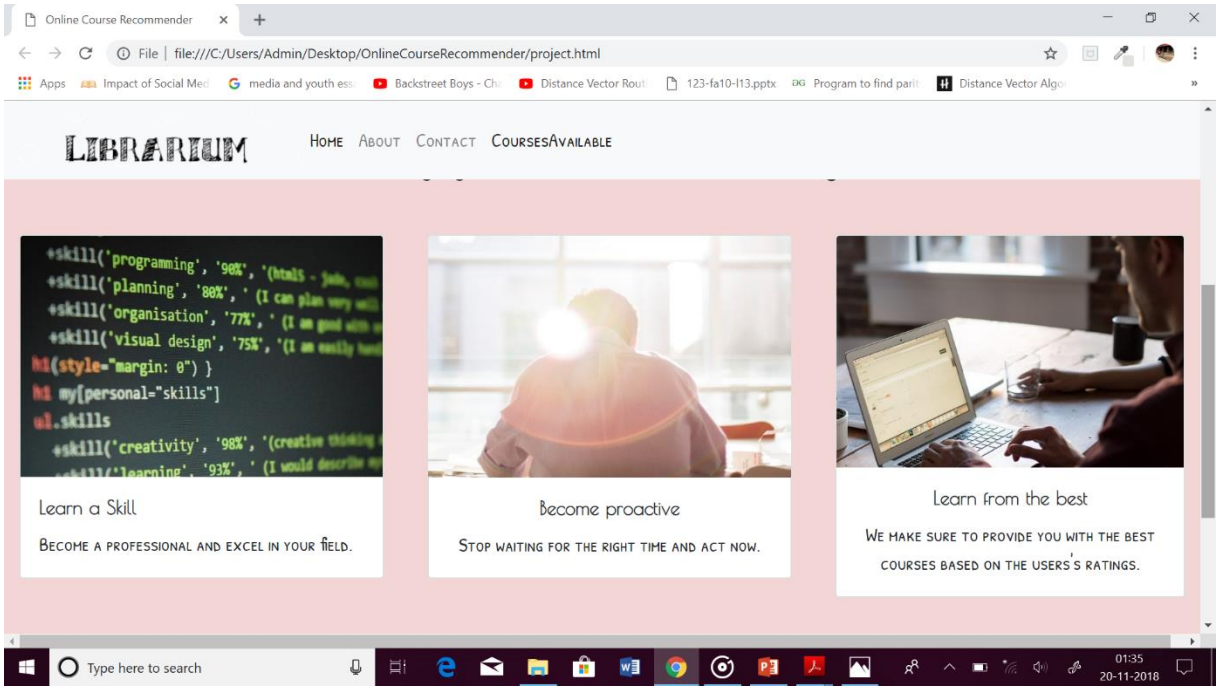
BOOTSTRAP

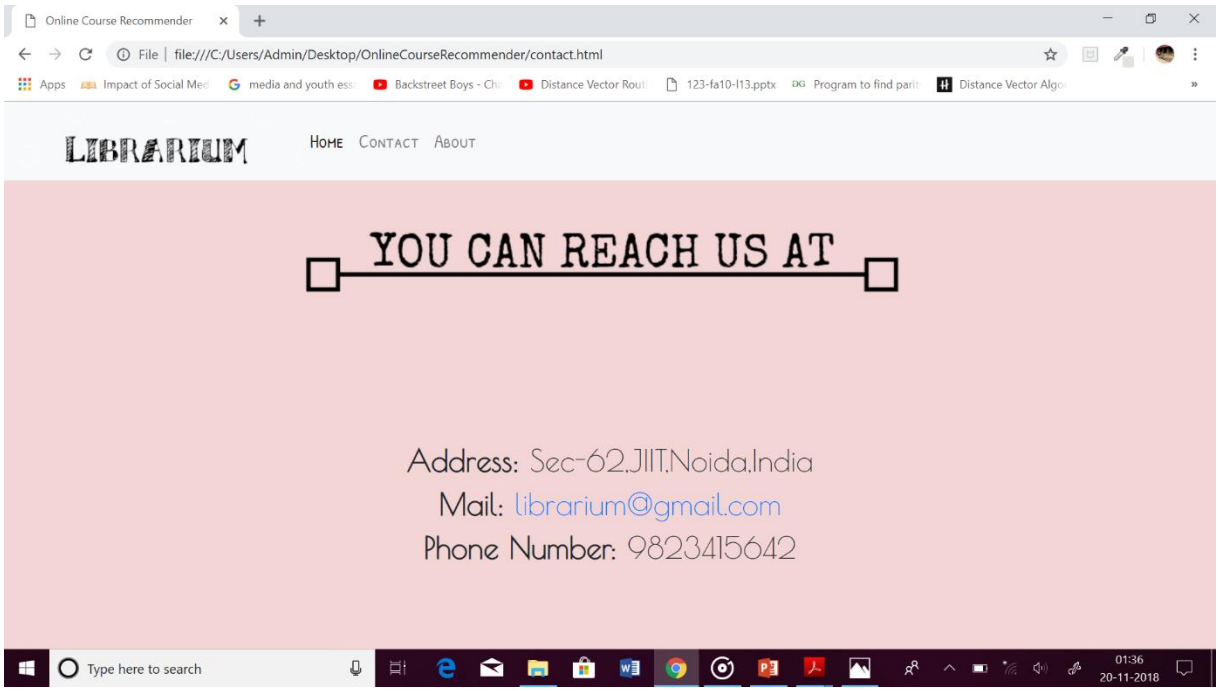
Bootstrap is a free and open-source front-end framework for designing websites and web applications. It contains HTML- and CSS-based design templates for typography, forms, buttons, navigation and other interface components, as well as optional JavaScript extensions. Unlike many earlier web frameworks, it concerns itself with front-end development only.

SCREENSHOTS






Online Course Recommender
file:///C:/Users/Admin/Desktop/OnlineCourseRecommender/project.html
LIBRARIUM
HOME ABOUT CONTACT COURSESAVAILABLE
Learn a Skill
BECOME A PROFESSIONAL AND EXCEL IN YOUR FIELD.
Become proactive
STOP WAITING FOR THE RIGHT TIME AND ACT NOW.
Learn from the best
WE MAKE SURE TO PROVIDE YOU WITH THE BEST COURSES BASED ON THE USERS'S RATINGS.
Type here to search
01:35
20-11-2018


Online Course Recommender
file:///C:/Users/Admin/Desktop/OnlineCourseRecommender/contact.html
LIBRARIUM
HOME CONTACT ABOUT
YOU CAN REACH US AT
Address: Sec-62,JIIT,Noida,India
Mail: librarium@gmail.com
Phone Number: 9823415642
Type here to search
01:36
20-11-2018

BACK-END OF THE WEBSITE

The back-end part of our project is based on Node.js and SQL. The user is free to post comments and reviews on the courses mentioned in the website. This is done in order to compare the courses. This user input is taken through node.js and stored in the tables in the database. The data is fetched using SQL.

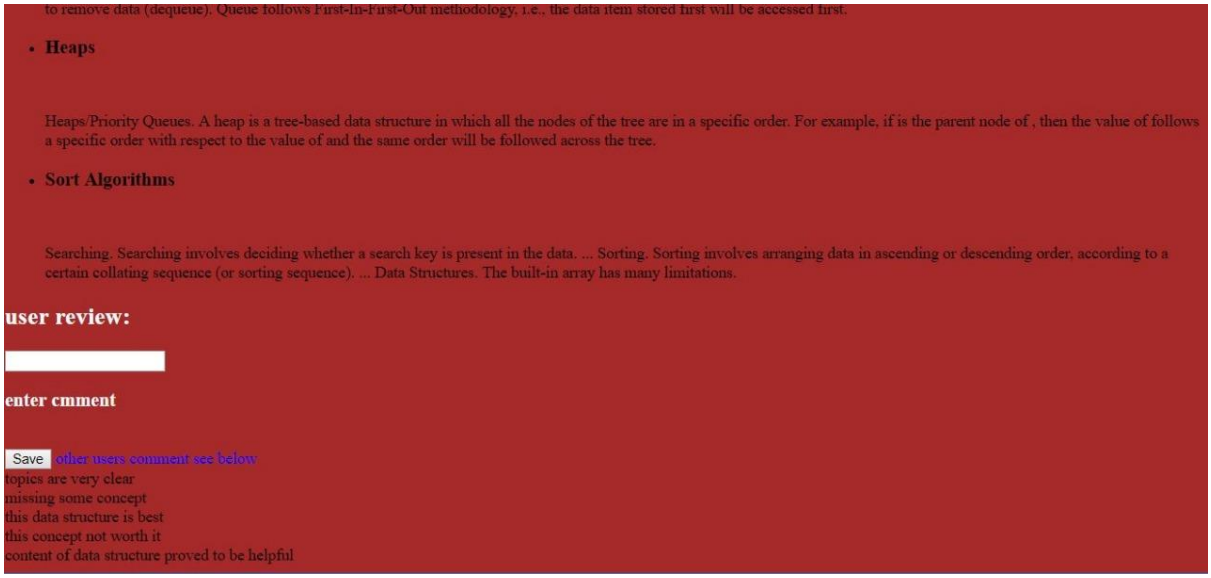
NODE.JS

Node.js is an open-source, cross-platform JavaScript run-time environment that executes JavaScript code outside of a browser. Typically, JavaScript is used primarily for client-side scripting, in which scripts written in JavaScript are embedded in a webpage's HTML and run client-side by a JavaScript engine in the user's web browser. Node.js lets developers use JavaScript to write Command Line tools and for server-side scripting—running scripts server-side to produce dynamic web page content before the page is sent to the user's web browser. Consequently, Node.js represents a "JavaScript everywhere" paradigm, unifying web application development around a single programming language, rather than different languages for server side and client side scripts.

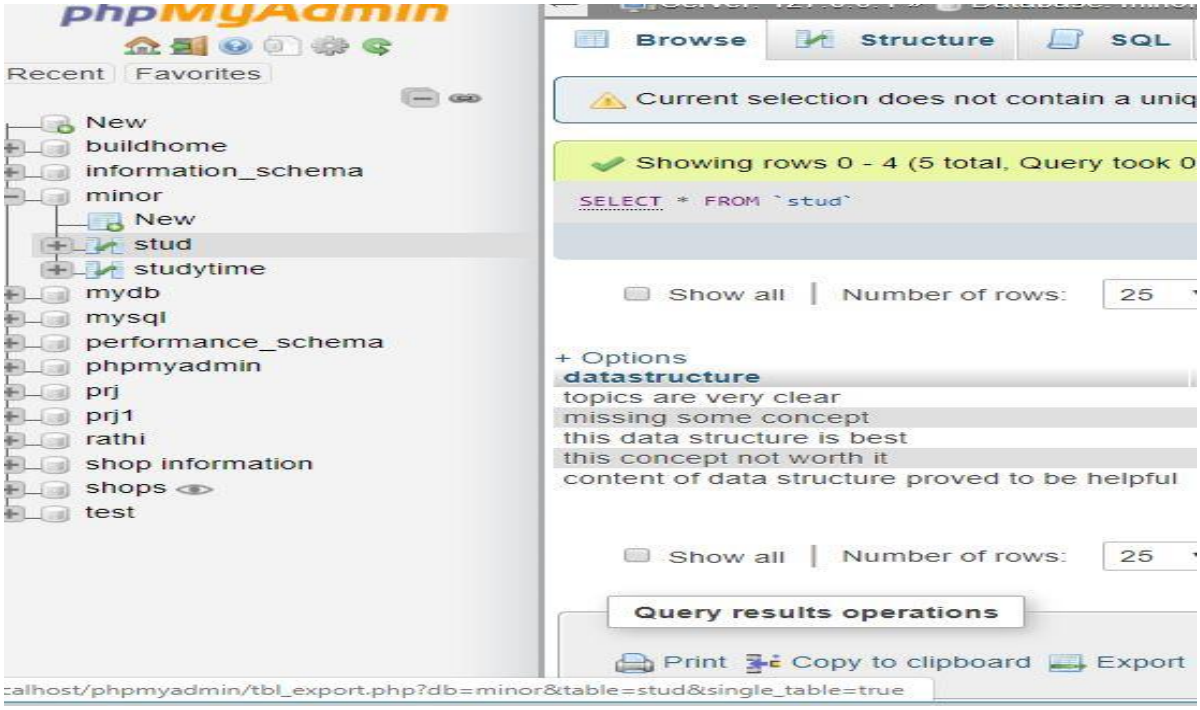
SQL

SQL ("sequel"; Structured Query Language) is a domain-specific language used in programming and designed for managing data held in a relational database management system (RDBMS), or for stream processing in a relational data stream management system (RDSMS). It is particularly useful in handling structured data where there are relations between different entities/variables of the data. SQL offers two main advantages over older read/write APIs like ISAM or VSAM: first, it introduced the concept of accessing many records with one single command; and second, it eliminates the need to specify how to reach a record, e.g. with or without an index.

SCREENSHOTS



User input page on the website

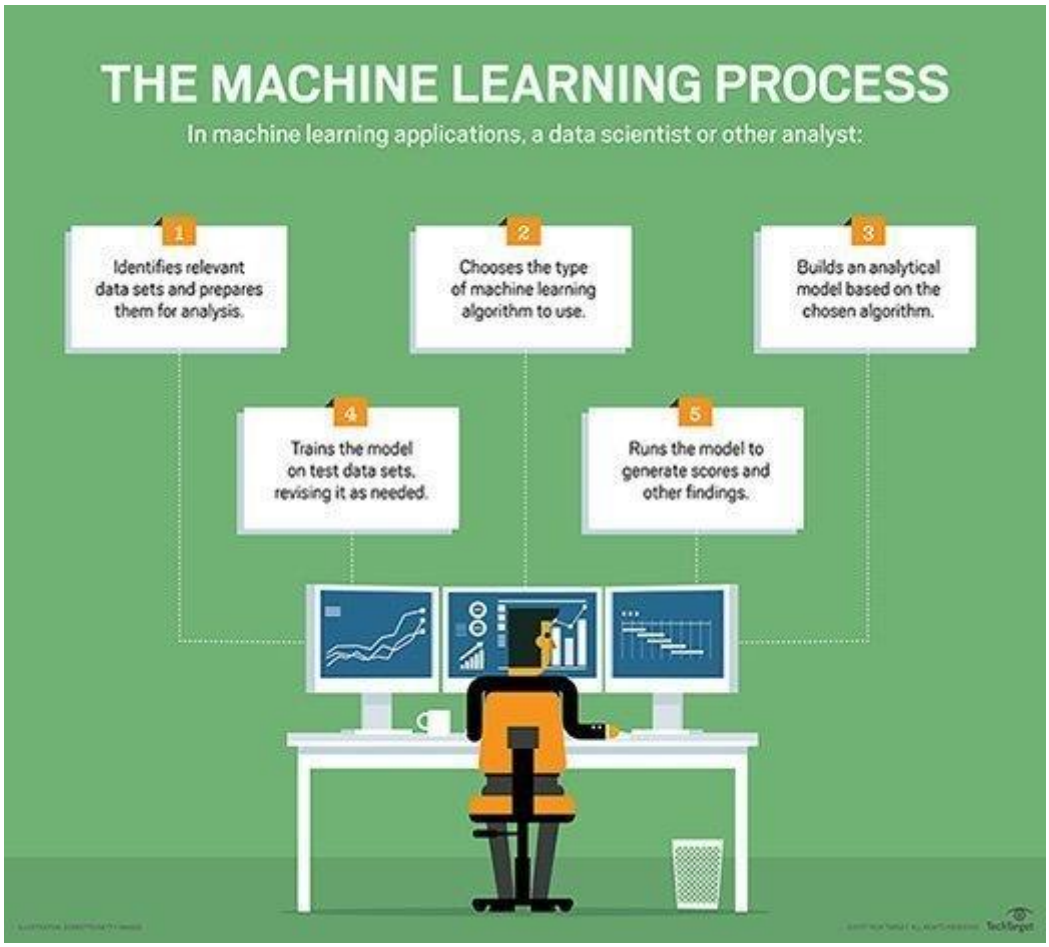


Database

MACHINE LEARNING

ML is used to train our website to distinguish between positive and negative comments and reviews registered by the user. The machine is first trained with a huge data set and then when user data comes, it marks positive comment with +1 while -1 is used to mark negative comments. Based on this, two courses are compared and displayed on the website.

Machine learning (ML) is a category of algorithm that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic premise of machine learning is to build algorithms that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available.



NAÏVE-BAYES ALGORITHM

Naive Bayes is a classification algorithm for binary (two-class) and multi-class classification problems. The technique is easiest to understand when described using binary or categorical input values.

The representation for naive Bayes is probabilities.

A list of probabilities are stored to file for a learned naive Bayes model. This includes:

Class Probabilities: The probabilities of each class in the training dataset.

Conditional Probabilities: The conditional probabilities of each input value given each class value.

SCREENSHOTS

Dataset to train our machine

test - Notepad

File Edit Format View Help

```
explaination very bad,-1
good explanation,1
its perfect,1
machine learning not explained clearly,-1
data structure explained perfect,1
conceptually very clear,1
not explained in good manner,-1
can not make concept clear,-1
algorithm explained perfect,1
topic not worth it,-1
love this tutorial,1
missing some concept,-1
understandable tutorial,1
```

Testing Dataset for ML

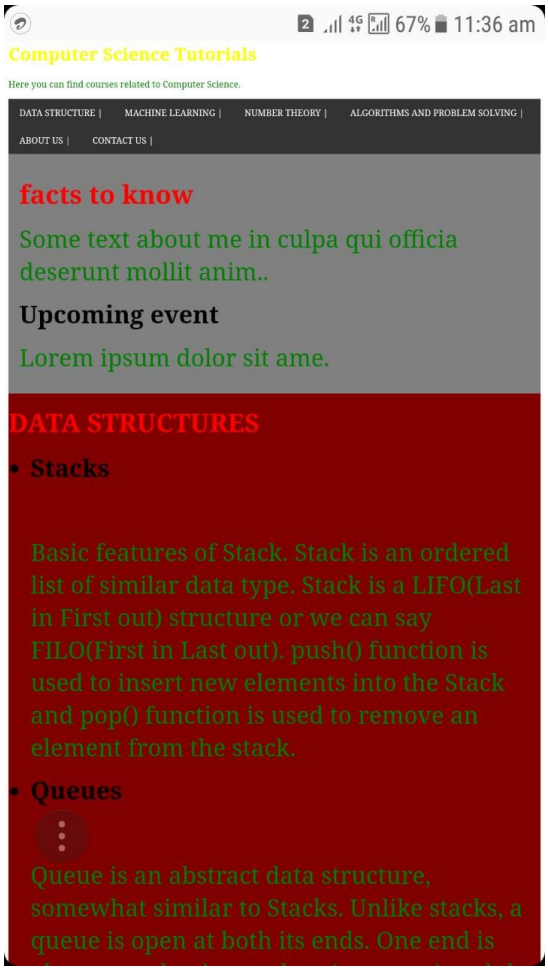
Python 3.7.0 Shell

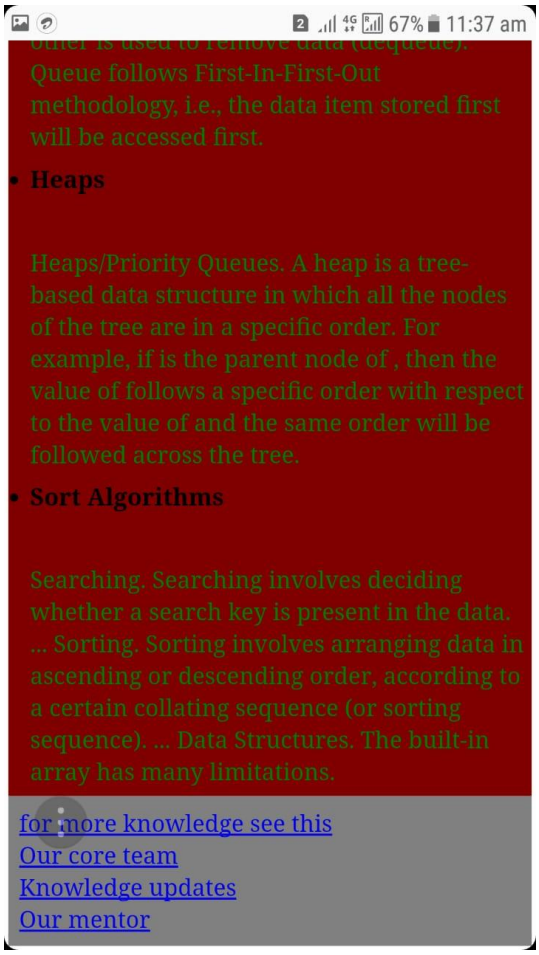
File Edit Shell Debug Options Window Help

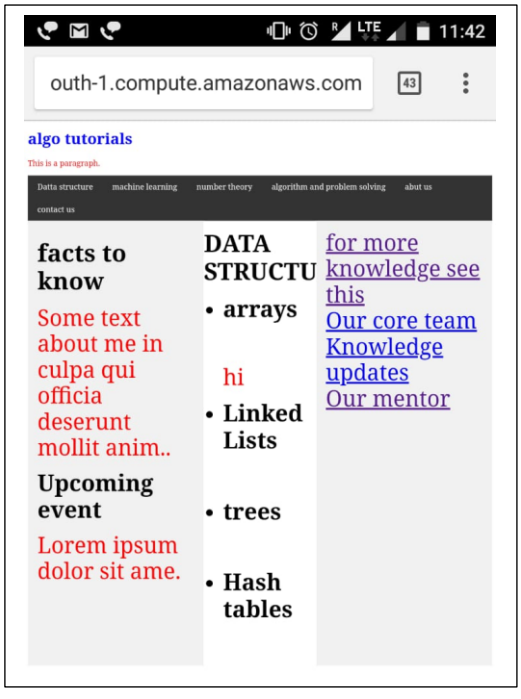
```
Python 3.7.0 (v3.7.0:1bf9cc5093, Jun 27 2018, 04:06:47) [MSC v.1914 32 bit (Intel)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
========== RESTART: C:\Users\prashant rathi\Desktop\minor\mnr.py ==========
Negative text sample: this concept not worth it bad example not a decent way to explain data structure algorithm topics ar
Positive text sample: great explanation of topic this data structure is best explanation is perfect conceptually very clea
Review: great explanation of topic
Negative prediction: 3.072363299832322e-08
Positive prediction: 3.364199283538042e-07
[-1, 1, 1, -1, 1, 1, -1, -1, 1, -1, 1, -1, 1]
>>>
```

Output for the test dataset

WEBSITES USED FOR MACHINE LEARING, WEB CRAWLING AND WEB SCRAPING







WEB SCRAPING AND CRAWLING

The library used goes to different websites (here, the two self-deployed websites) and fetches the data from those websites and store in a database. Crawling is basically going to website and scanning the website and hyperlinks and fetching the data to store in the database. Scrapping is done to fetch a particular i.e the required data.

Web Crawling is the process of locating information on World Wide Web(WWW), indexing all the words in a document, adding them to a database, then following all hyper links and indexes and adds that information also to the database. Web Crawling crawls html, content on pages, style sheets, meta data, images etc.

Web scraping is the process of automatically requesting a web document and collecting limited information from it rather then all data. Strictly speaking, to do web scraping, you have to do some degree of web crawling to move around the websites.

SCRAPY LIBRARY

An open source and collaborative framework for extracting the data you need from websites in a fast, simple, yet extensible way.

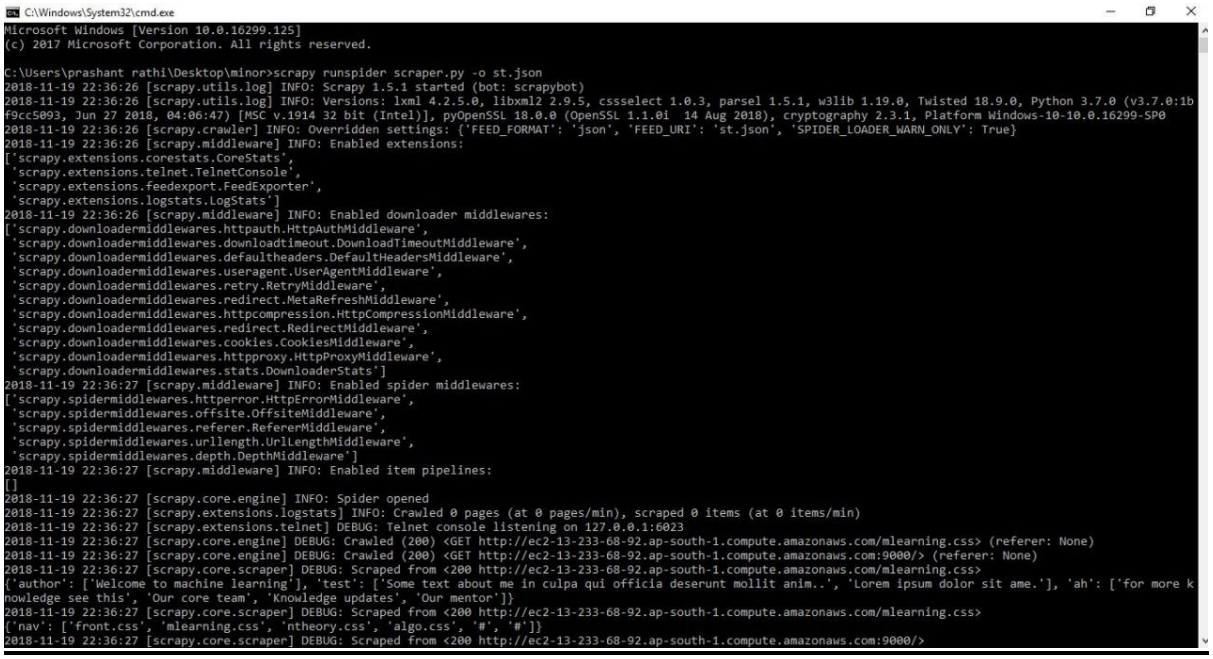
In this, spiders are created to meet the required functions. Spiders are classes that you define and that Scrapy uses to scrape information from a website (or a group of websites). They must subclass scrapy. Spider and define the initial requests to make, optionally how to follow links in the pages, and how to parse the downloaded page content to extract data.

Our Spider subclasses **scrapy.Spider** and defines some attributes and methods:

- **name**: identifies the Spider. It must be unique within a project, that is, you can't set the same name for different Spiders.
- **start_requests()**: must return an iterable of Requests (you can return a list of requests or write a generator function) which the Spider will begin to crawl from. Subsequent requests will be generated successively from these initial requests.
- **parse()**: a method that will be called to handle the response downloaded for each of the requests made. The response parameter is an instance of **TextResponse** that holds the page content and has further helpful methods to handle it.

  The **parse()** method usually parses the response, extracting the scraped data as dicts and also finding new URLs to follow and creating new requests (**Request**) from them.

SCREENSHOTS



Scrapy Running