The Islamia University of Bahawalpur

**University College of Engineering & Technology**

**Department of Computer System Engineering**



**Solution**

COMP 03631 Data Mining

Complex Engineering Problem(CEP)

“Mobile Phone Recommendation System Using Cosine Similarity”

Semester: **6th**

**Supervised By:**

**Eng: Muhammad Bux Alvi**

**Submitted By:**

**M Ramzan Ali(19cs32)**

**Acknowledgments**

I would like to express our special thanks of gratitude to our teacher **Eng: Muhammad Bux Alvi** who gave me the golden opportunity to do this project on the topic “Mobile Phone Recommendation System Using Cosine Similarity” . This project helped me do literature study, learn data Extraction method from E-Commerce website like Daraz and perform exploratory data analysis (EDA), feature engineering, data representation, and transformation. Furthermore, it helped me to learn to create machine learning models and their optimization. The most significant part of the CEP was that I leant to think and act. Taking advantage of the opportunity, I would also like to thank my parents and friends who helped me in finalizing this project within a limited timeframe.

**Table of Contents**

1. Abstract **………………………………………………………………………………………..**4

problem Statement**...............................................................................................................**4

Introduction**..........................................................................................................................**4

1. Literature Review **.......................................................................................................................**5
2. Dataset Description**……………………………………………………………………….........5**

Dataset Attributes**……………………………………………………………………...….**5

Dataset**…………………………………………………………………………………….**6

1. Prediction Using Cosine Similarity**…………………………………………………………….**7
2. Code and Implementation **……………..……………………………………………………....**8

Data Scraper code**…………………………………………………………………………**8

Python Code**………………………………………………………………………………**8

1. Conclusion**………………………………………………………………………………….…**9
2. 6. References**………………………………………………………………………………..….**10

**Abstract**

In this report we present our projects for both Data Mining and Machine Learning. Using a massive data set of Daraz reviews we implemented Clustering, Cosine Similarity technique to propose suggestions in an e-commerce website. We developed a recommendation system with a collaborative filtering using ten different distance measures in order to find the best one.

**Problem Statement**

Many of the people stuck to purchase the Mobile phones of Good Specification due to high prices. We resolve this issue by using Unsupervised Learning technique. In this technique we use Clustering, Cosine Similarity which works on Products Features Similarity. Data set contains about 1k entries of different Mobiles. This recommendation engine will take as input data of a certain product and recommend approximately 10 similar products that resemble closely to the input data.

**Introduction**

Web scraping, also known as data mining, is the process of collecting large amounts of data from the web and then placing it in spreadsheets, databases or CSV file for future analysis, predictions and later use**.** The concept of scraping the web is not new, however, with modern programming languages enable us to build web scrapers that collects complex unstructured data and save this in a structured way. However, in addition to be a very complicated task, Web Scraping is resource and time-consuming, mainly when it is carried out manually.

Web pages are built using text-based mark-up languages (HTML and XHTML), and frequently contain a wealth of useful data in text form. However, most web pages are designed for human end-users and not for ease of automated use. Because of this, tool kits that scrape web content were created. A web scraper is an API or tool to extract data from a website. Companies like Amazon AWS and Google provide web scraping tools, services and public data available free of cost to end users. Newer forms of web scraping involve listening to data feeds from web servers. For example, JSON is commonly used as a transport storage mechanism between the client and the web server. Recently, companies have developed web scraping systems that rely on using techniques in DOM parsing, computer vision and natural language processing to simulate the human processing that occurs when viewing a web page to automatically extract useful information.

We have also tried to recommend items using clustering techniques on the relationship between theitems. As last attempt we tried to work on the text of the reviews representing them as TF IDF vectors and compare them in order to get suggestion based on similar opinions. Once we have implemented Cosine Similarity we tested them, and we evaluated their performance. At the end, based on that Daraz data, we have also figured out which of these techniques fits better for our purpose. Obviously Daraz contains a huge amount of reviews and ratings on its system. In fact, it is impossible that a single computer process all these data in a reasonable time. We Use Front-end is created in web using HTML, CSS and JS. Web Front-end is connected to python backed using eel. EEL is python library which allows communicating with JavaScript. For recommendation purpose I used cosine similarity in python The data is sent back to front-end using eel.

**Literature Review**

Web scraping is the process of extracting and creating a structured representation of data from a website. E-Commerce is known as electronic commerce, involves the trading of products or services over the internet using web-based applications. With the rapid development of the internet and the number of goods sold online, e-commerce plays an important role in the present trend. In the current business market, e-commerce is playing a role in services, retails, finance, telecommunications, and information technology services. E-commerce provides an easy way to search for any particular product efficiently using the web. It also deals with context-aware services which can be based on a provided context i.e. location, time of the day or user preferences. Today, the internet is the most widely used communication medium in modern enterprises. Many companies are redefining their business strategy to increase business output and productivity. Internet based business provides the opportunity in terms of a business hub where customers and partners can find their products and specific businesses, and also an interaction platform Nowadays, online business crosses the barriers of time and space compared with traditional physical offices. Big companies all over the world realize that e-commerce not only purchases and sells through the internet, but also increases the efficiency of competition with other giants in the global market. Due to this reason, data mining is sometimes referred to as knowledge discovery. Web mining is a data mining technique applied to the World Wide Web. There is a lot of information on the internet E-commerce websites are now one of the most important sources for buying various products. Many strategies have been developed to analyze the behavior of customers to attract more business and people’s participation. Since there are many e-commerce websites available, it is difficult for users to choose the best deal for the desired product in these websites. E-Commerce product comparison using web mining enables users to analyze prices and obtain desired products at the lowest price. Users can also choose multiple products that belong to the same category to compare their functionality. To get the best price from e-commerce sites, web crawlers and web scrapping technology are used to obtain detailed information

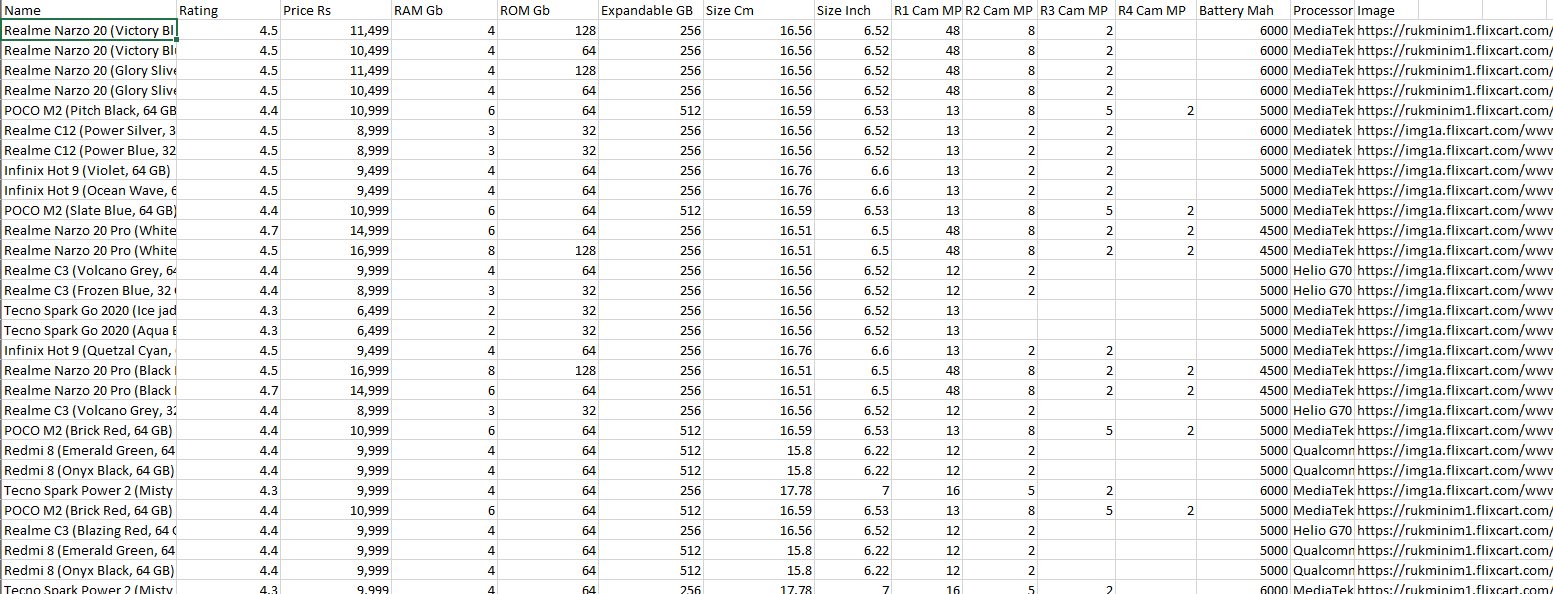
**Dataset Attributes**

We use the dataset named “Mobile Phones”, The attributes names are given below:

* Name
* Rating
* Price RS
* RAM GB
* ROM GB
* SIZE Inch
* BATTERY(Mah)
* Image

**Dataset**

The proposed system used a dataset which consists of over 963 mobile phone of various brands with their specifications, descriptions, user ratings on a 5-point scale and reviews. This information was scraped from the Daraz Web site for the corresponding mobiles as Daraz is one of the most popular sites and the data here is more organized when compared to other Web sites. Sample dataset is shown in Fig



**Figure 1**

**Prediction Using Cosine Similarity**

Cosine similarity is the measure of the similarity between two documents by counting the frequencies of similar terms and considering them as vectors. It is calculated by the cosine of the angle between two vectors and it ranges from 0 to +1 in the case of documents, because the frequencies of the terms cannot be negative. This measure suffers less from the curse of dimensionality than Euclidean distance. Document vectors tend to get long because there will be many words in a given document. So cosine similarity is a way to avoid the curse of dimensionality, which is measured by Eq. (1).

Cosine Similarity (|x. y|) = x. y /|| x ||- || y ||

Basically, it is the Euclidean Distance of the vector from origin. Similarly, || y ||| is the magnitude of vector y. If the cosine similarity value is closer to 1 then angle between the vectors is smaller and similarity between the vectors is greater.

After extracting the combined features for each mobile phone in the dataset, all the selected features are combined together. A count matrix was created from these new combined features. Now, cosine similarity was computed based on the count matrix. After this process, input was taken from the user. Based on the input given by the user, index of the mobile phone is retrieved from its name. A list of similar mobile phones was generated in descending order of similarity score, then names and details of first 10 similar mobile phones corresponding to the mobile phone the user had chosen were displayed. Then, the user can choose any of the mobile phone from the given top 10 recommendations and then a link to buy that mobile phone is given to the user.

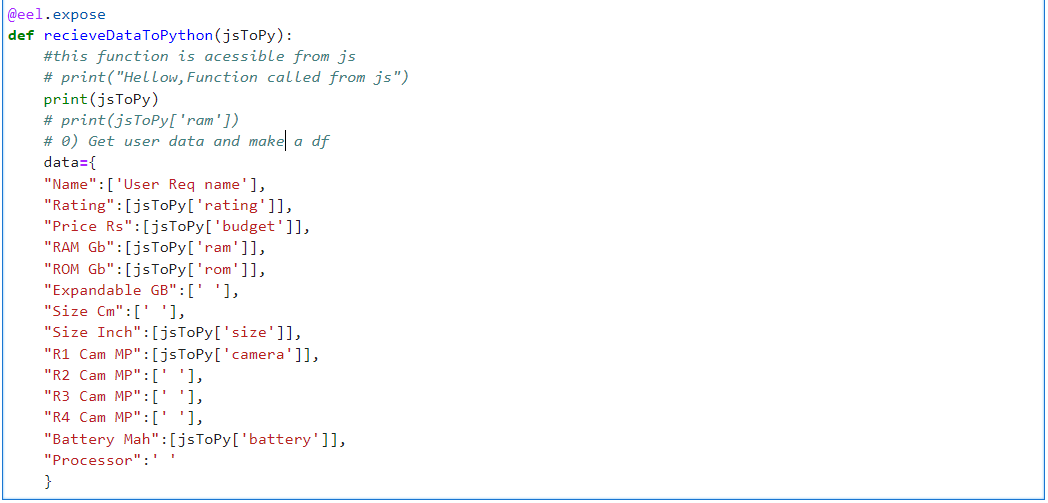
**Data Scraper code**

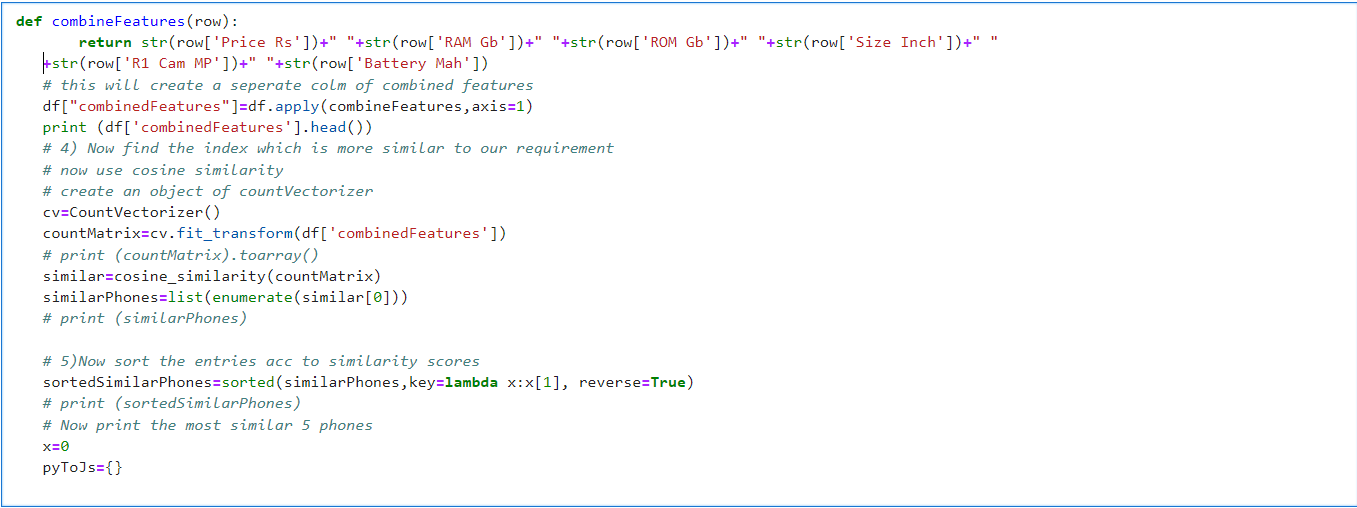
****

****

**Python Code**

****

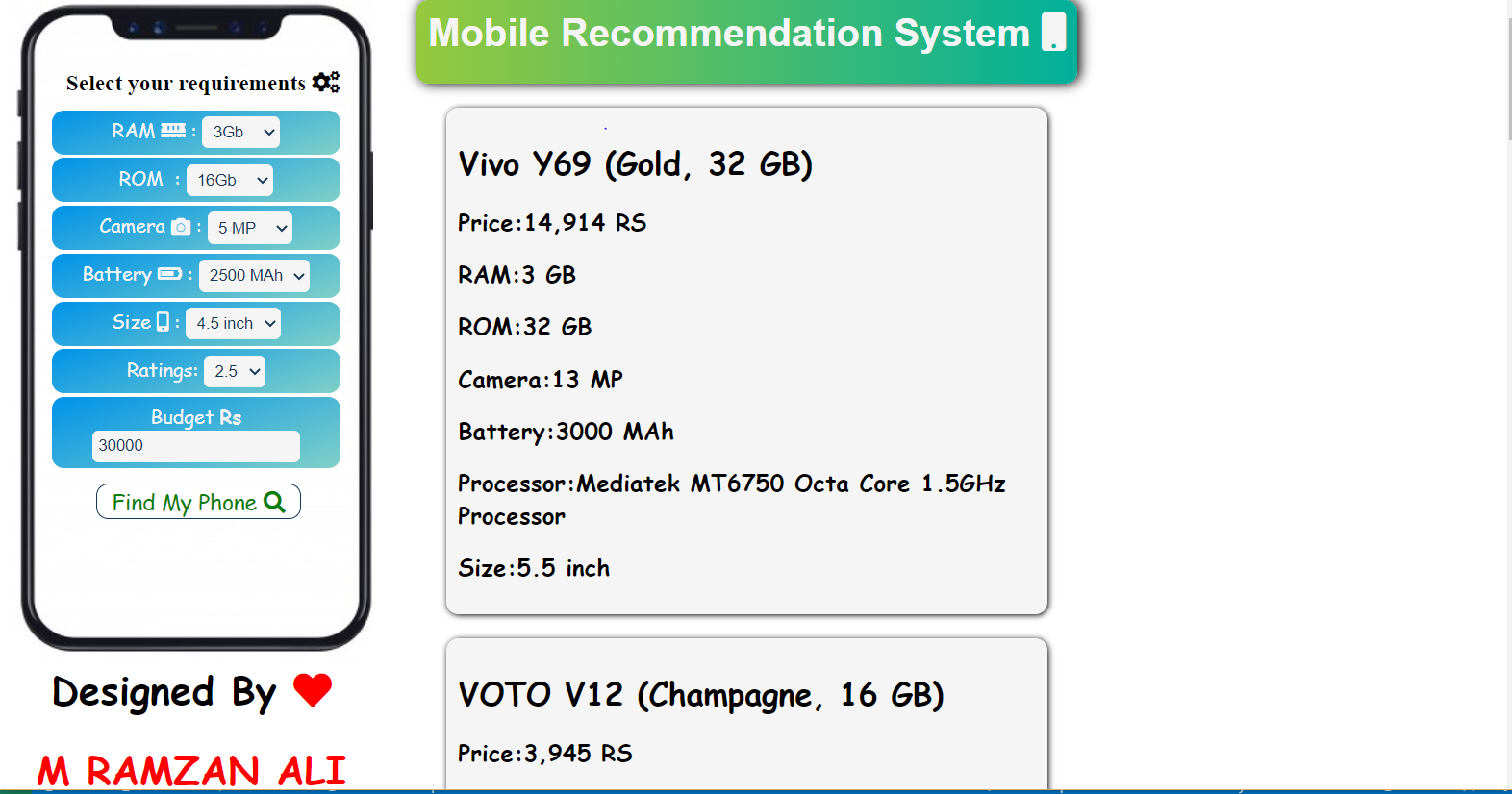
****





**Conclusion**

With the help of all the parameters, libraries and functions, it was made possible to analyze the Mobile Phones data. A system for mobile phone recommendation was proposed and its architecture, implementation details, and the results were discussed in this paper. The experimental analysis of the proposed system was done based on the and cosine similarity methods. The similarity indices were taken as reference for the recommendation. The proposed system could map each of the mobile phone to its similar mobile phones in the data set based on the filtering done by keeping a threshold for relevance. Proper and nearly accurate prediction of similar mobile phones based on all the input features and descriptions was achieved.  
  
The work can be further extended to be applied in real life where it can be implemented in the form of an application which does comparisons between multiple Web sites selling the same product and gives an in depth analysis of all the similar mobile phone to the input/searched mobile phone.  
This work can further be expanded to several recommendation systems of multiple different applications and also this work is confined to Daraz Mobile data and recommendations are done on this data. This is a GUI of Mobile Recommendation system. In this GUI I apply the cosine similarity on products Features. This will print 10 most similar product data on the screen.



**References**

[**https://github.com/theomkale/Mobile\_Recommendation\_System.git**](https://github.com/theomkale/Mobile_Recommendation_System.git)