## **Current Status**

A dataset containing reviews of Amazon products given by the consumers has been selected for the recommendation system. The CSV file contains review text, ratings, product information, reviewer information, and other things.

Amazon uses a cloud computing platform called Amazon Web Services (AWS). It includes platform as a service (PaaS), infrastructure as a service (IaaS), packaged software as a service (SaaS) offerings. Its services can be separated into multiple categories like compute, Data management, networking, etc. according to the customer’s needs. Amazon has spread its data centers across availability zones in regions around the world. A region is a collection of AZs which contains multiple data centers. Amazon stores data as S3 objects in S3 buckets that give scalable object storage to backup data and helps to keep the data organized. For AWS users, PostgreSQL, Oracle, SQL Server, high-performance database “Amazon Aurora”, etc. are relational database management systems. Through Amazon DynamoDB, AWS offers NoSQL databases too. AWS provides administrators that can manage and keep track of cloud resource configurations, monitors resource and applications health. AWS also provides cloud security services and tools that help to measure potential security risks. AWS uses a variety of AI models for voice and text technology, text-to-speech translation, and image and facial analysis, etc. (Gillis, 2021).

## **Requirements**

The dataset was collected to make a recommendation system for the consumers by analyzing their previous reviews and ratings of the Amazone products. This system will recommend products to the consumers that they might like. The recommended products vary with every consumer because all consumers don’t have the same taste.

To build the system, first of all, the right datasets of the customer activities need to be collected. The amount of data will decide how good will be the recommendation system. The type of data that is being dealt with will decide the type of storage that needed to be implemented which might include a NoSQL or a SQL or some type of object storage. Then extraction of relevant information by filtering the data to make required and accurate recommendations is done using algorithms. There are different types of algorithms used for making a recommendation system which are shown below.

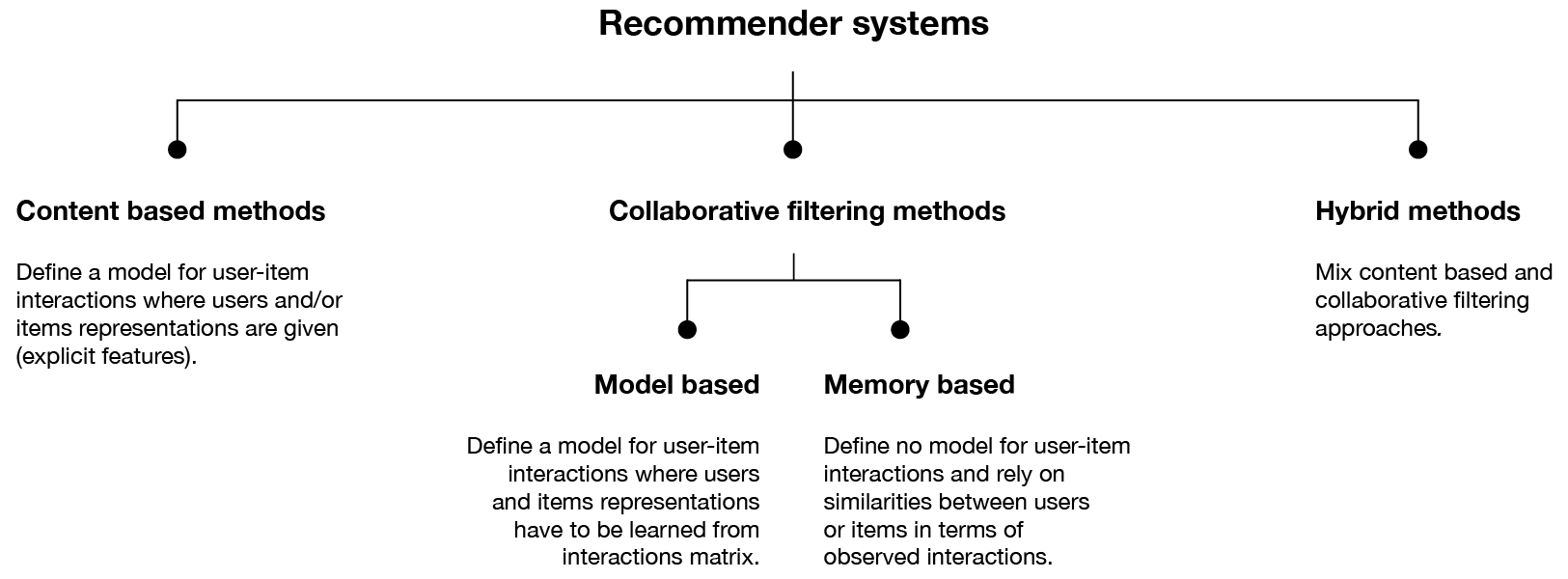


Figure 1: Algorithms used for making a recommendation system (Rocca, 2019).

Python will be the programming language used for making the system by using one of the above methods. Python provides its libraries that will help to make the system more easily. Matrix factorization will be also used while making the system (Sharma, 2018). Then a server will be required to run the system. The server might be cloud-based or traditional or hybrid.

## **Quality of Service (QoS)**

QoS is the technology that controls data traffic to minimize jitter, packet loss, and latency on the network of service and this has great impacts on the overall performance of the service. QoS provides priorities for specific kinds of data on the network which helps managing network resources. Bandwidth, latency, jitter, packet loss, mean opinion score are the parameters that measure how good is the QoS quantitatively  (Lutkevich, 2021).

Since the recommendation system uses lots of data from the user’s previous activities and recommends different types of items to different users, it requires high bandwidth, very low-latency networks for the better functioning of the system. The system will recommend items with more accuracy to a user if it can access a high amount of data of that user. There should be more focus on the data QoS because the system deals with lots of user’s data and newly generated data. There should not be any packet loss and a very little amount of jittering while retrieving the data for a better and accurate recommendation for each user.

## **Infrastructure**

A cloud-based storage system will be better for the recommendation system because it will not only save the cost but also provide a high level of accessibility and additional security layers, high-speed performance, better data back up and restore, more reliability, mobility, availability, and scalability, pay-per-use, unlimited storage capacity, fast and effective virtualization, etc. (n.d., 2021). Cloud-based storage will be very fast enough to recommend a different types of items to users according to the change in user activities.

For the smooth and better functioning of the system, very high-speed processors like Intel® Xeon® Scalable Processor, Intel® FPGAs, etc. (n.d., 2021), high-speed rams with storage from 128 GB to 4 TB (Chambille, 2017) will be required. Very high-speed SSD with enough storage will be used for computing purposes. For the data storage purpose, cheaper disks with a high storage capacity, reliability, and performance, a low failure rate will be required.

## **Data Collection**

The dataset was collected from “data.world” website which is an open-source world's largest collaborative data community from where people can share, analyze, and discover lots of data. This dataset gives consumer's reviews on the Amazone products. It contains 1597 rows, 27 columns and also an extra column “index”. This column was added to the dataset to give indexing for each product because this will be used while developing the recommendation system.

Dataset Link: <https://data.world/datafiniti/consumer-reviews-of-amazon-products/workspace/file?filename=7817_1.csv>

Two screenshots of the CSV file is shown below.

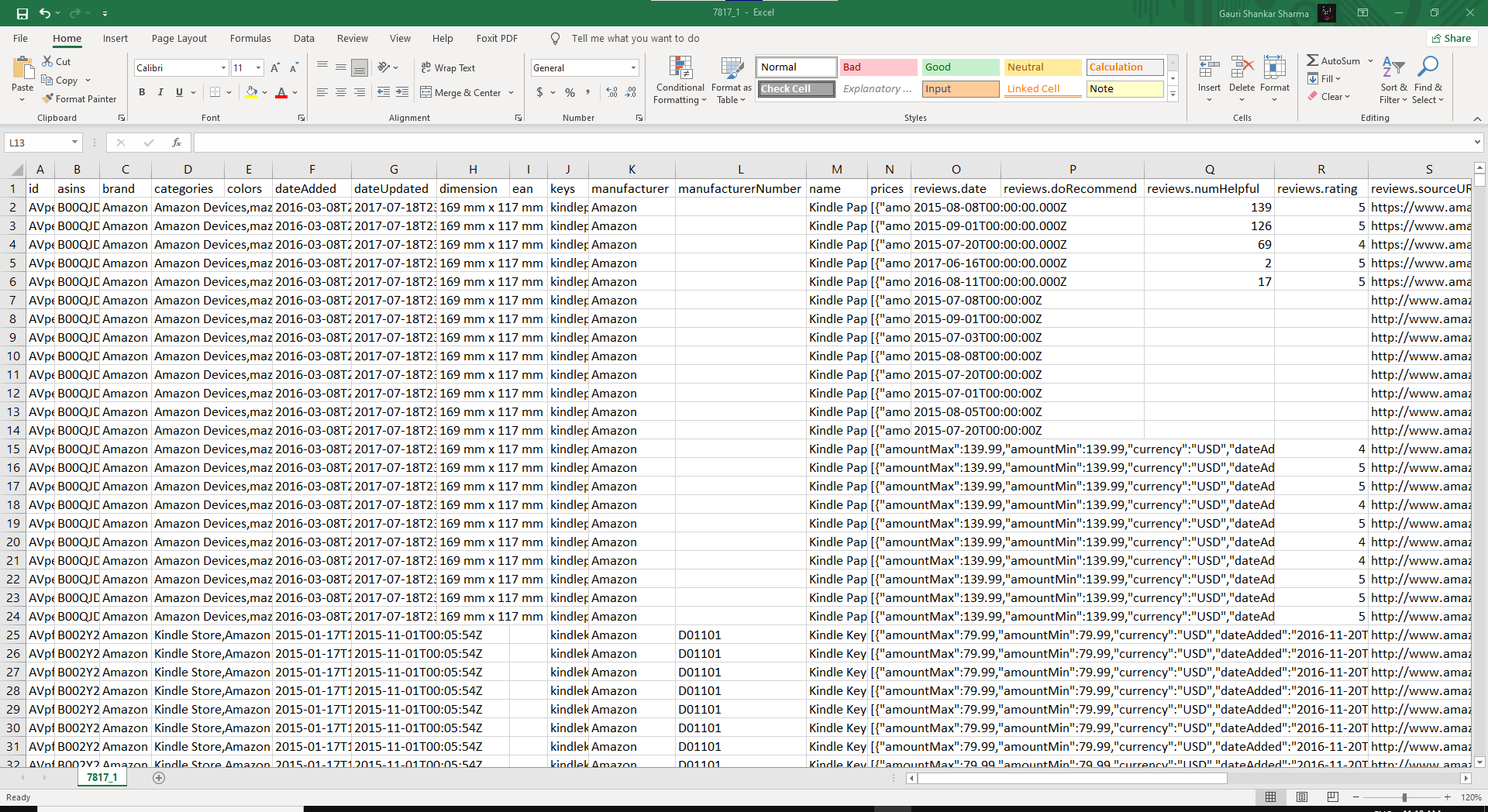


Figure 2: Consumer's reviews on Amazone products.

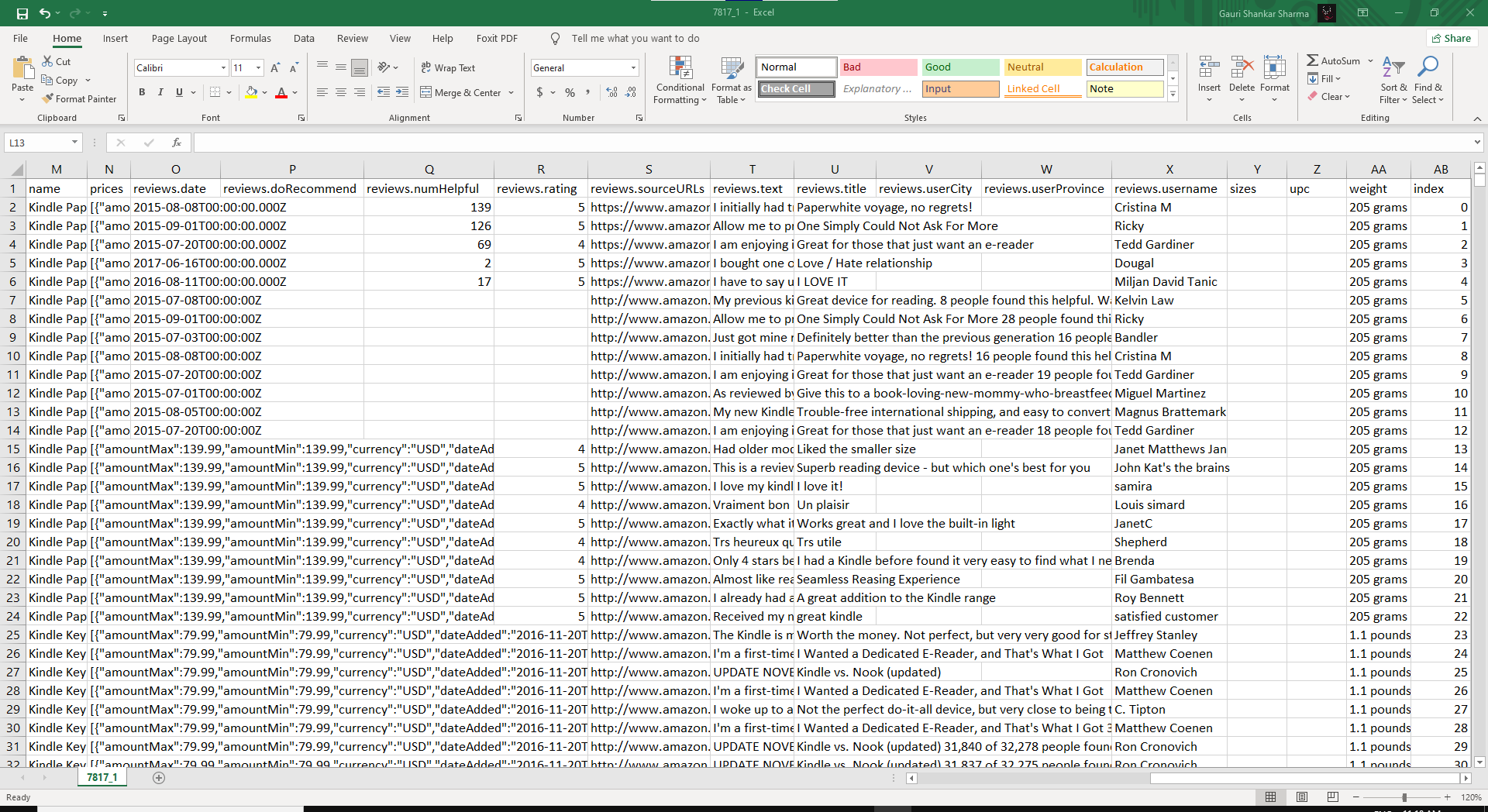


Figure 3: Consumer's reviews on Amazone products.

## **Data Use**

Content-based, collaborative, and hybrid are the methods used in the recommendation systems which are also shown in figure 1 above.

In this recommendation system, content based method will be used. This method recommends similar items based on the metadata, such as descriptions, details, reviews text, reviews title, etc. of a particular item. The idea behind this recommendation method is that if a person likes an item, then the person also might like items similar to the item. For example, youtube recommends new videos to users based on their history (Sharma, 2020).

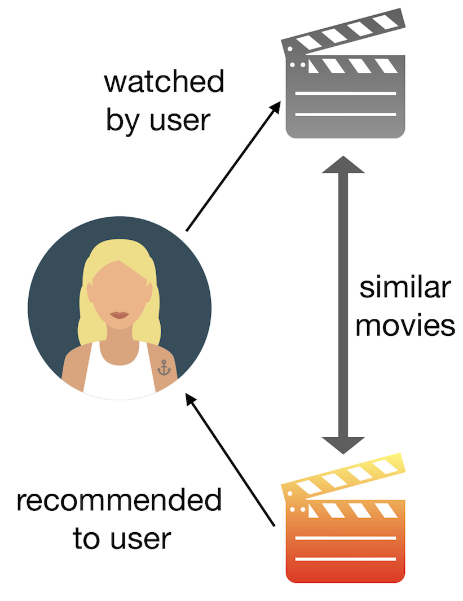


Figure 4: Content based recommendation (Grimaldi, 2018).

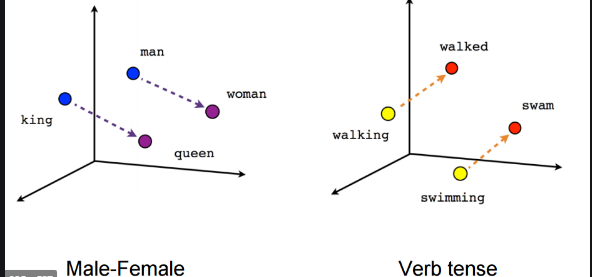
To design the system, word to vector conversion will be done for the selected columns. 

Figure 5: Vector

The objects containing both a mangnitude and a direction is called vetcoter. Vectorizing a word will help to draw valuable information from the plotted data.The TF-IDF vectors will be computed for the selected columns. TF refers to Term Frequency which gives how often an item occurs in a document. IDF refers to Inverse Document frequency which gives how important a term is. The TF-IDF gives an matrix and this matrix will be used to compute similarity score. The consine similarity will be used for this. This gives a numeric quantity which idenfies the similarities between two items. For TfidfVectorizer and cosine similarity, “sklearn” library offered by python will be used (n.d., 2020).

### **Objectives**

* Get the required dataset
* Access the required item and it’s metadata
* For all item, compute pairewise similarities scores based on selected columns
* Acessing the items that matches with the similarities scores of the main item
* Recommend correct items to correct users based on the similarities scores

## **Recommendation**

While Recommending products, “reviews.text” and “reviews.title” columns of the dataset were taken as the product features to compute the similarity score.

### **Source Code**

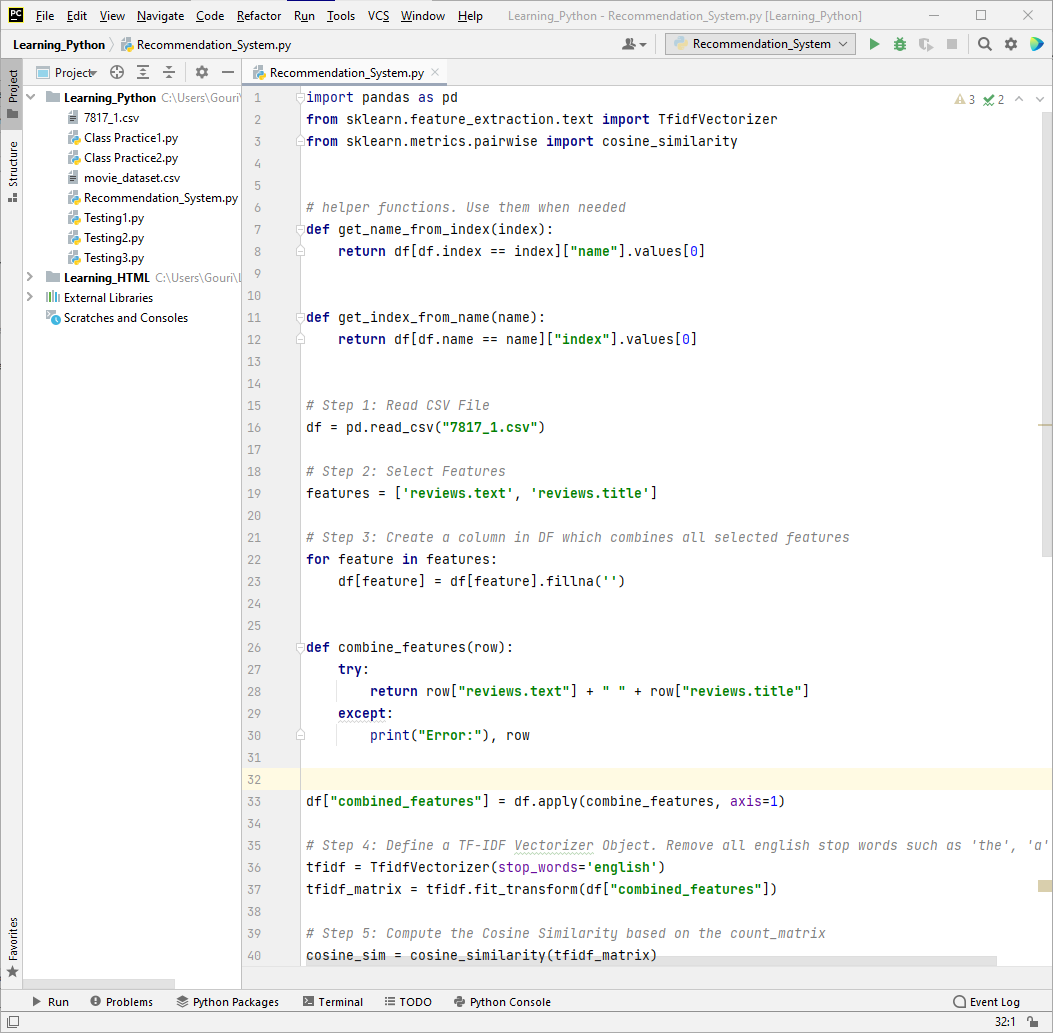


Figure 6: Content based recommendation system source code.

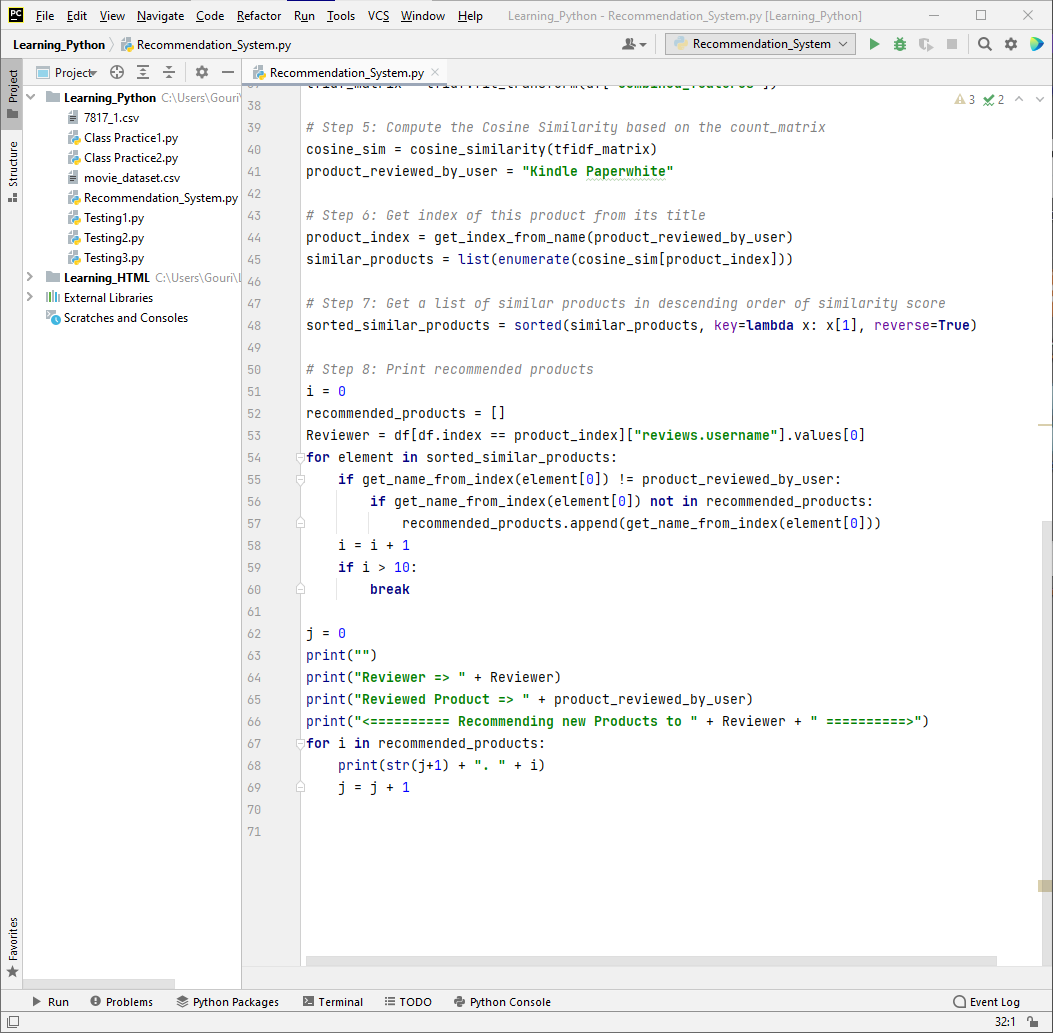


Figure 7:Content based recommendation system source code.

### **Recommending products to consumers**

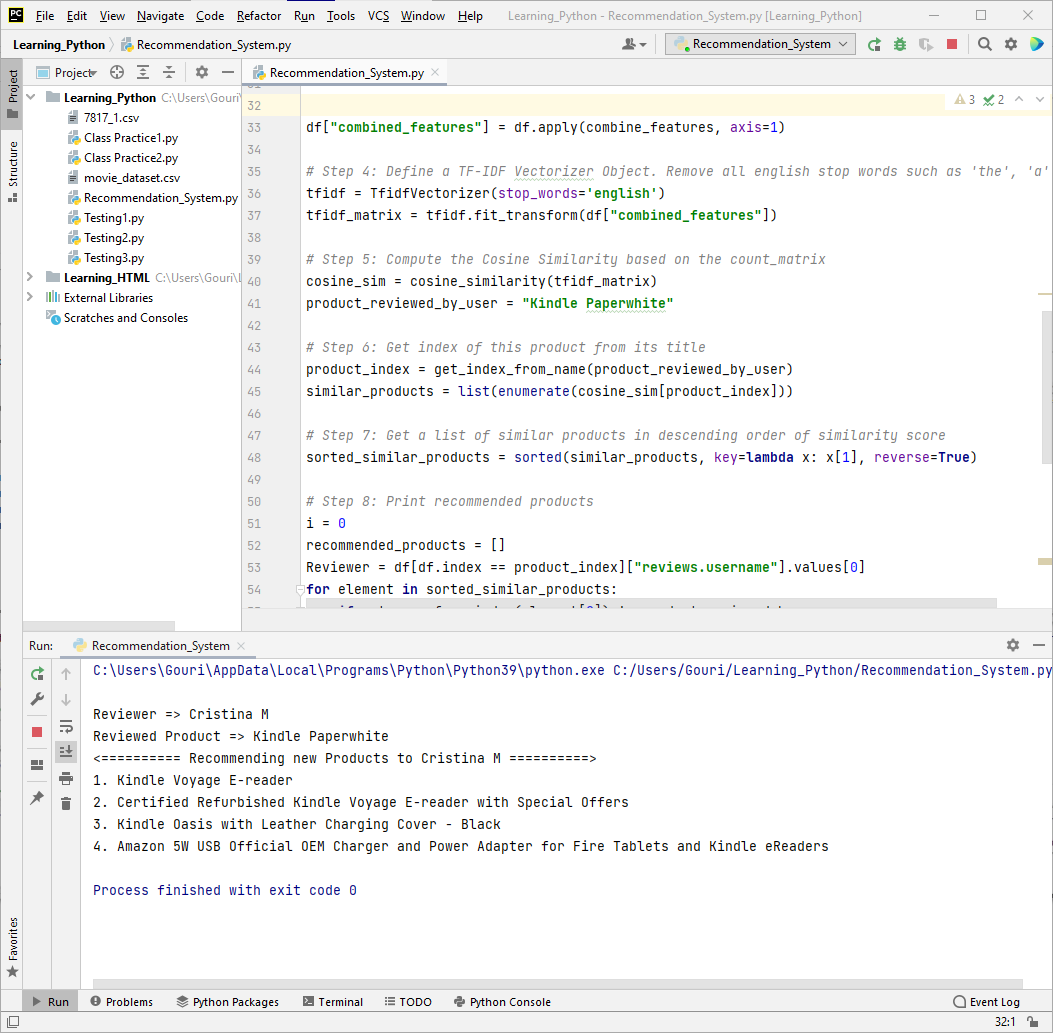


Figure 8: Recommending other products to “Cristina M” who reviewed product "Kindle PaperWhite".

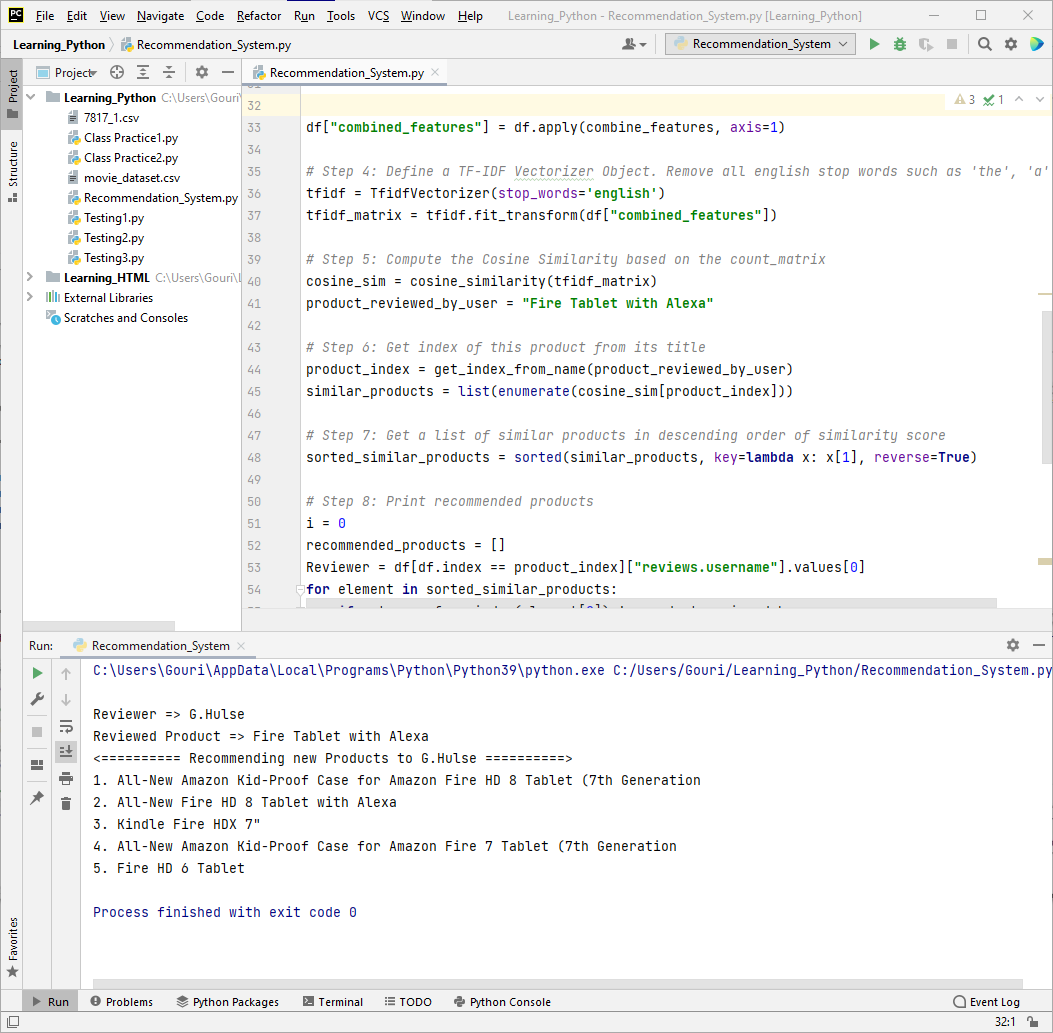


Figure 9: Recommending other products to “G.Hulse” who reviewed product "Fire Tablet with Alexa".

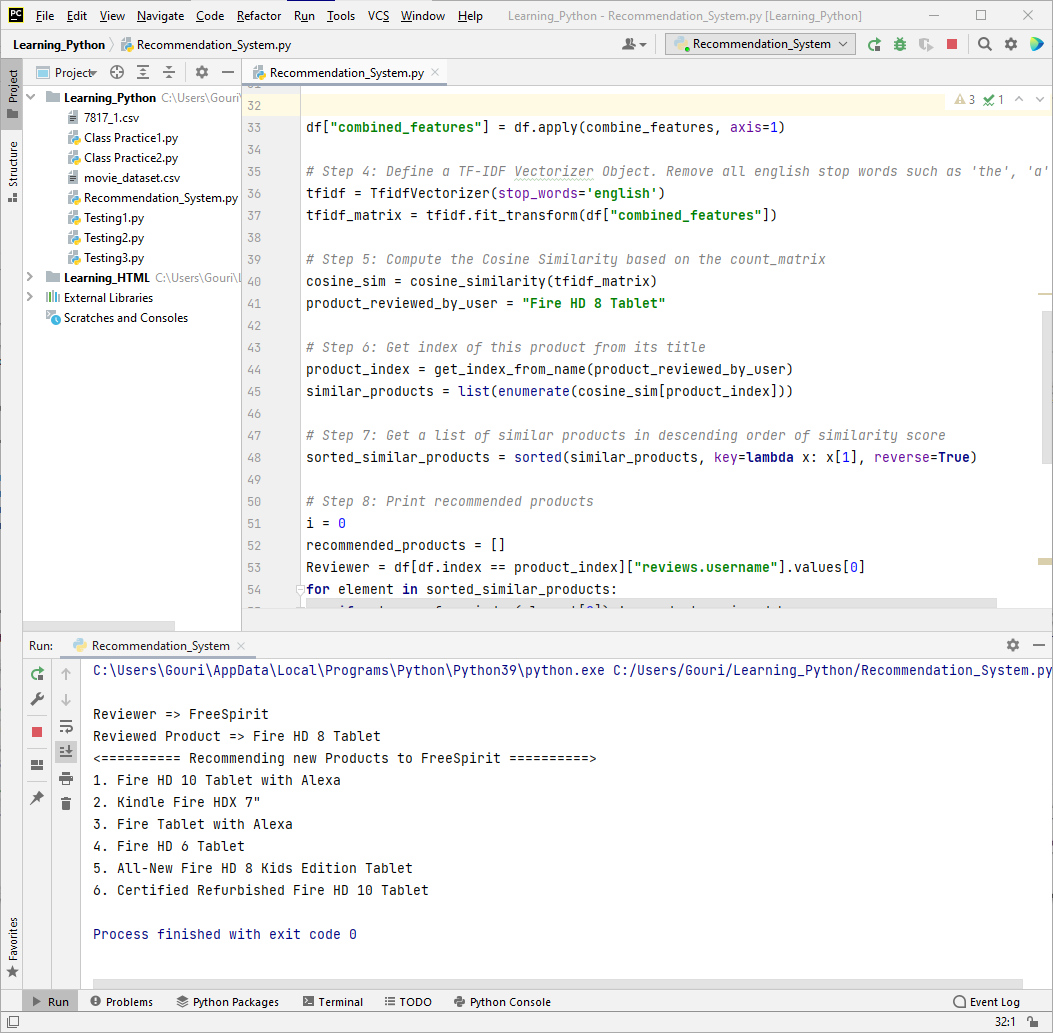


Figure 10: Recommending other products to “FreeSpirit” who reviewed product "Fire HD 8 Tablet".

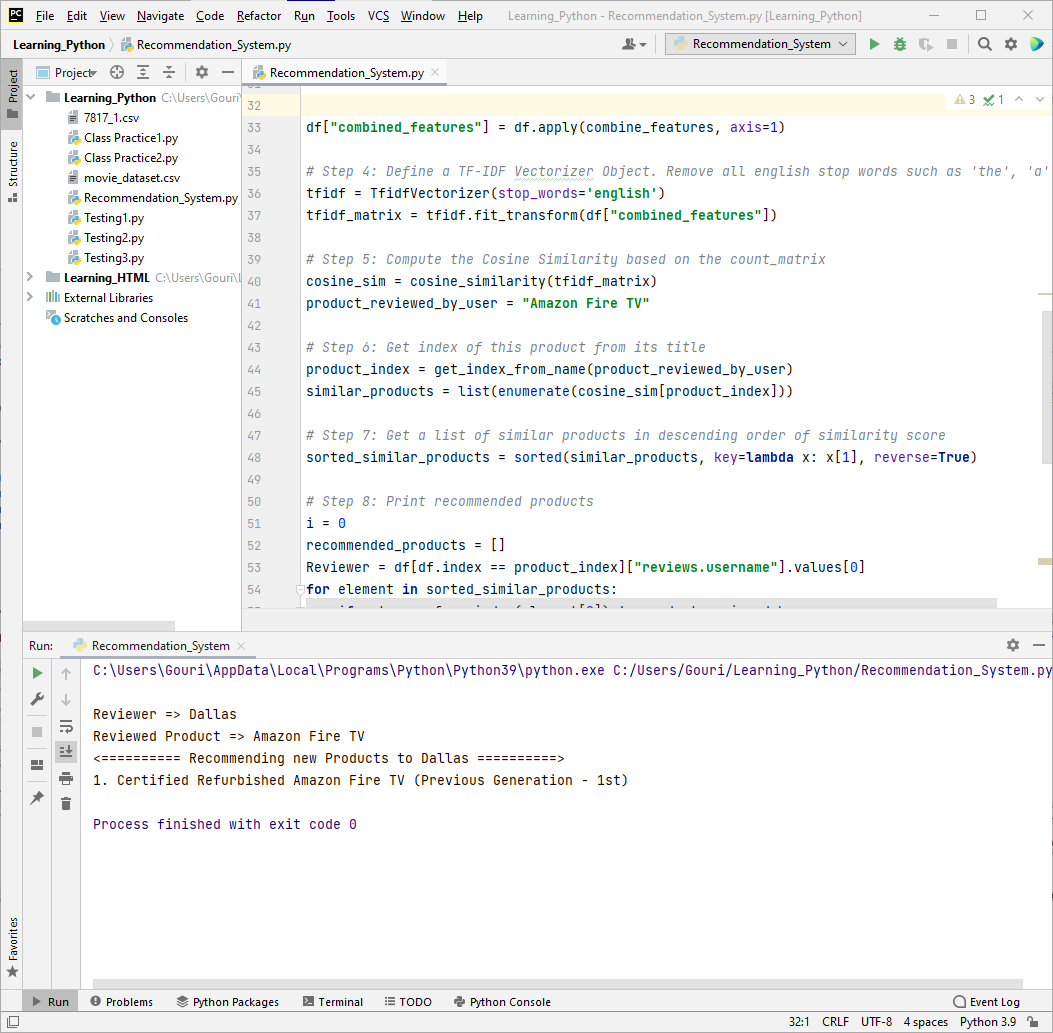


Figure 11: Recommending other products to “Dallas” who reviewed product "Amazon Fire TV".

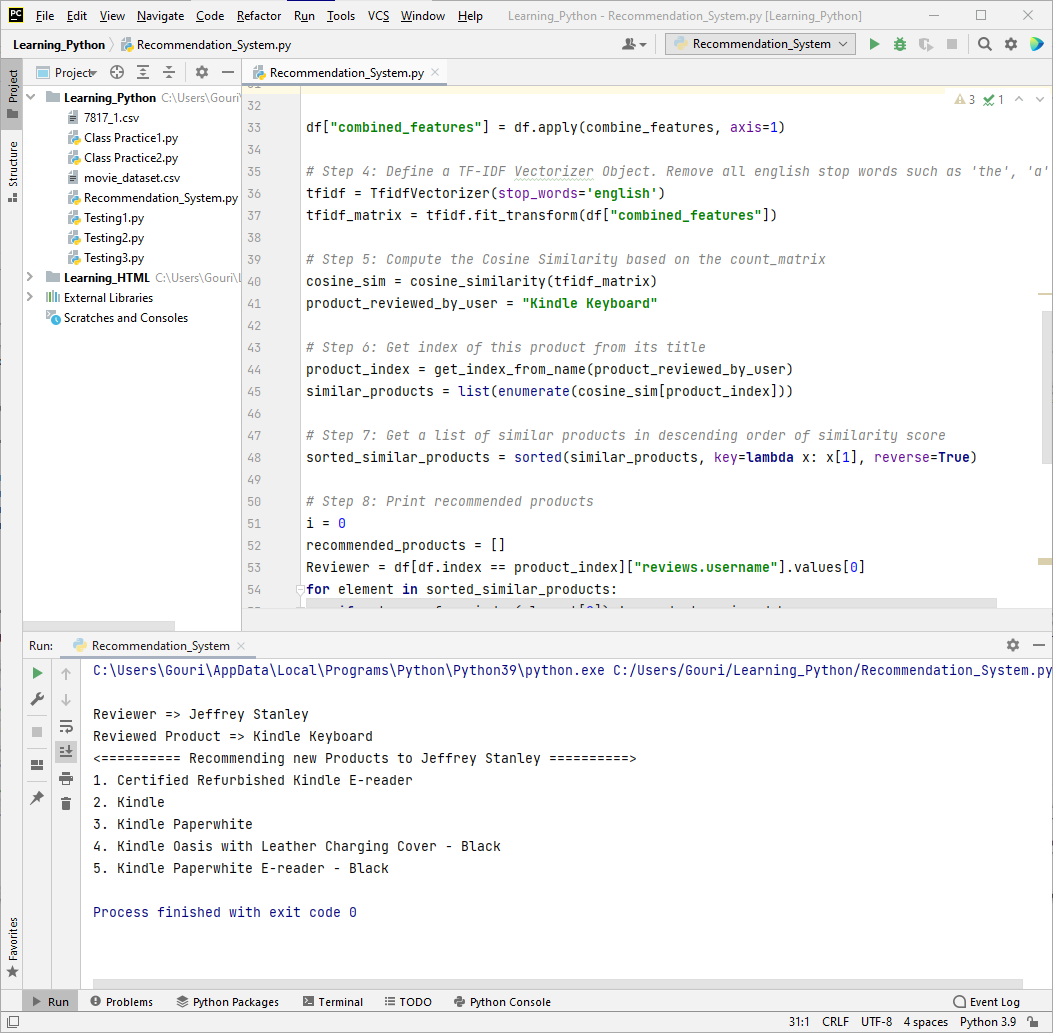


Figure 12: Recommending other products to “Jeffrey Stanley” who reviewed product "Kindle Keyboard".