**Baku Engineering University**

**Computer Engineering**

**Lecturer: Khayyam Masiyev**

**Course Name: System analysis and Design**

**Elvin Aghammadzada (150105006)**

**Tural Sadigov (150105005)**

**Tural Ahmadov (150105002)**

**Elvin Bayramov (150105008)**

**FF**

**Contents**

[**1.Abstract** 2](#_heading=h.gjdgxs)

[**2. Introduction** 2](#_heading=h.30j0zll)

[**3. Preliminary Investigation Report** 3](#_heading=h.1fob9te)

[**3.1 Fishbone Diagram** 3](#_heading=h.3znysh7)

[**3.2 Project Scope and Constraints** 4](#_heading=h.tyjcwt)

[**3.3 Fact Findings** 5](#_heading=h.3dy6vkm)

[**3.3.1 Survey** 5](#_heading=h.1t3h5sf)

[**3.3.2 Interview** 6](#_heading=h.4d34og8)

[**3.4 Project Usability, Cost, Benefit and Schedule Data** 7](#_heading=h.2s8eyo1)

[**3.5 Evaluation of Feasibility** 8](#_heading=h.17dp8vu)

[**3.5.1 Operational Feasibility** 9](#_heading=h.3rdcrjn)

[**3.5.2 Technical Feasibility** 9](#_heading=h.26in1rg)

[**3.5.3 Economic Feasibility** 10](#_heading=h.lnxbz9)

[**3.5.4 Schedule Feasibility** 10](#_heading=h.35nkun2)

[**3.6 Project Request Form** 10](#_heading=h.44sinio)

[**3.6.1 Description of the project** 10](#_heading=h.2jxsxqh)

[**3.6.2 Requested Features** 11](#_heading=h.z337ya)

[**3.6.3 Main Objective of the Project** 11](#_heading=h.3j2qqm3)

[**3.7 Memo** 11](#_heading=h.1y810tw)

[**3.7.1 Overview of meeting:** 11](#_heading=h.4i7ojhp)

[**3.7.2 Problems and Result Solutions** 11](#_heading=h.2xcytpi)

[**3.7.3 Result of Meeting** 12](#_heading=h.1ci93xb)

[**3.8 Recommendations** 12](#_heading=h.3whwml4)

[**4. Requirement Analysis** 13](#_heading=h.2bn6wsx)

[**4.1 WBS** 13](#_heading=h.qsh70q)

[**4.2 Gantt Chart** 14](#_heading=h.3as4poj)

[**4.3 CPM** 14](#_heading=h.1pxezwc)

[**4.4 PERT Chart** 16](#_heading=h.49x2ik5)

[**4.5 List of Requirements** 16](#_heading=h.2p2csry)

[**5.**](#_heading=h.3o7alnk) **UML Diagrams** 19

[**5.1 Context Model** 19](#_heading=h.23ckvvd)

[**5.2 Activity Diagrams** 20](#_heading=h.ihv636)

[**5.2.1 User System** 20](#_heading=h.32hioqz)

[**5.2.2 Data Processing System** 21](#_heading=h.1hmsyys)

[**5.3 Use case diagrams** 22](#_heading=h.41mghml)

[**5.3.1 User System Use Case** 22](#_heading=h.2grqrue)

[**5.4 Sequence diagram** 23](#_heading=h.3fwokq0)

[**5.4.1 User System** 23](#_heading=h.1v1yuxt)

[**5.5 Class Diagram** 25](#_heading=h.4f1mdlm)

[**5.6 E-R Diagram** 26](#_heading=h.2u6wntf)

[**6.**](#_heading=h.19c6y18) **Conclusion** 26

[**7.**](#_heading=h.28h4qwu) **References** 27

**AI Based Review-Feedback Analyzer**

## **1.Abstract**

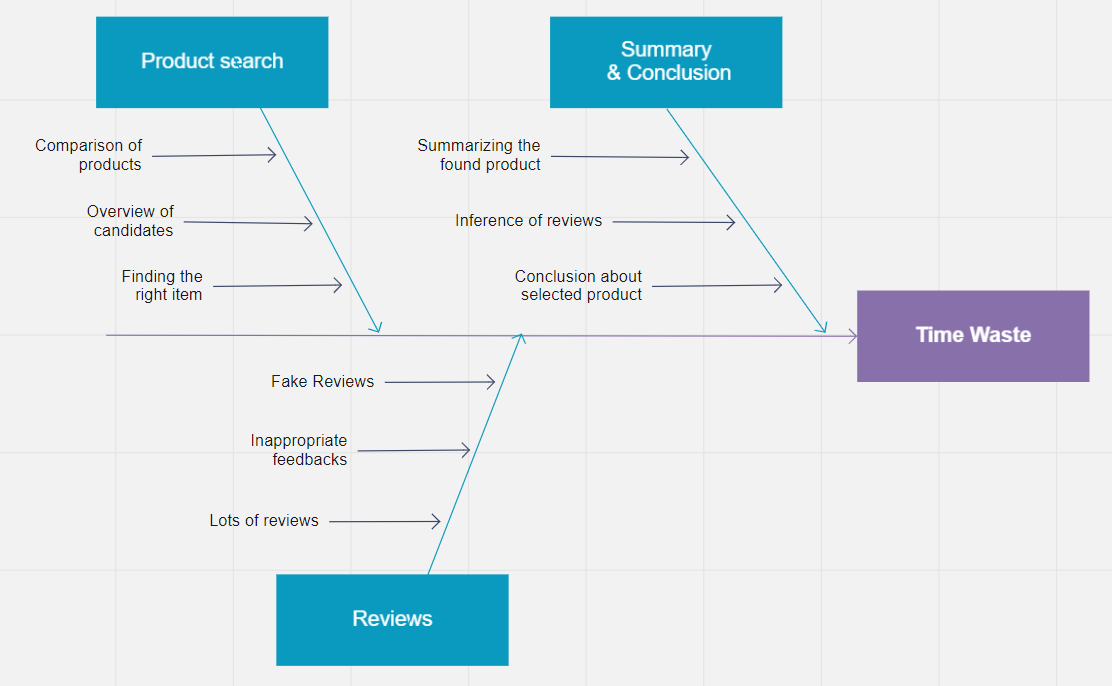
This document is about AI Based Review-Feedback Analyzer which aims to eliminate time waste of customers while reading reviews of products to buy. We will use an AI based software to make summaries of thousands of reviews by clearly pointing out merits and demerits of any product. Customers will be able to read the summaries of a myriad of reviews in a minute and choose the preferable product by putting less effort and spending less time on. In detail, all work done about this project is included here till now, which covers initial research, surveys and interviews. According to the results of them, list of requirements, relevant diagrams – WBS, Gantt Chart, CPM, PERT Chart and Context Model, Activity, Use Case and Sequence, Class and E-R Diagrams is composed.

# **2. Introduction**

Nowadays, thanks for the advances in the modern technology, especially in IT sector, more and more people are using the online shopping services namely, Amazon, AliExpress, Ebay. Even though it is a piece of cake to purchase an item by means of this kind of websites, there can be some cases in which things do not go as it was expected to be. Some people in these foregoing websites can fraud customers or tell a lie about a particular product. To prevent the possibility of these kind of unwanted actions, costumers tend to read the reviews of tens of thousands of products during their everyday life. Obviously, it is time-consuming and we all know how precious the time is in our today’s “busy world”. Even if a particular customer reads a review and feel satisfied, it can be fake one, and can mislead thousands of people to buy that product. In order to solve these issues, we need to eliminate the time waste which we spent on reading thousands of sentences. Moreover, we should also in some way figure out the fake reviews. A tool would definitely be beneficial, if it can differ fake and real reviews using special methods. Additionally, it would be absolutely time-saving if we had a tool which summarizes even tens of thousands of sentences. Since it is urgently needed in today’s society, people would even pay for it to save their precious time. Foregoing needed project is our project and its resultant tool in the end. With the help of the advances in the modern technology, namely Artificial Intelligence, our tool will summarize the texts of even thousands of words and give feedback according to it which would be short several sentences. Starting from identifying the fake reviews by the help of different methods, it will also store the corresponding processed reviews in the cloud database in order to retrieve faster whenever it is needed. Additionally, we also cared about companies to cooperate with. Thanks to the services of recommendation and advertising, other companies would be able to see their products prioritized in the search, or promoted on the left or right side of the website or application.

# **3.Preliminary Investigation Report**

## **3.1 Fishbone Diagram**

Customer’s time is extremely valuable so that it should not be spent on activities which can be done by automated systems such as our AI based review-feedback analyzer. The **problem** is the time waste and it arises when customer starts to search for wished item, reads reviews and feedbacks about product and after summarizing the content tries to conclude about which item to choose.

While searching for suitable product the customer faces 3 main time-consuming steps. Firstly, it’s about finding the appropriate items and making a list of them. Next, each of them should be analyzed and can be conducted a short overview of candidates. Finally, products will be compared based on those overviews to have some understandings.

Another stage is reading reviews which is exhausting. The problems encountered by customer can be divided into followings:

* Reading fake reviews – absolute waste of time
* Facing inappropriate feedbacks – can deceive you
* Lots of reviews – makes customer exhaust

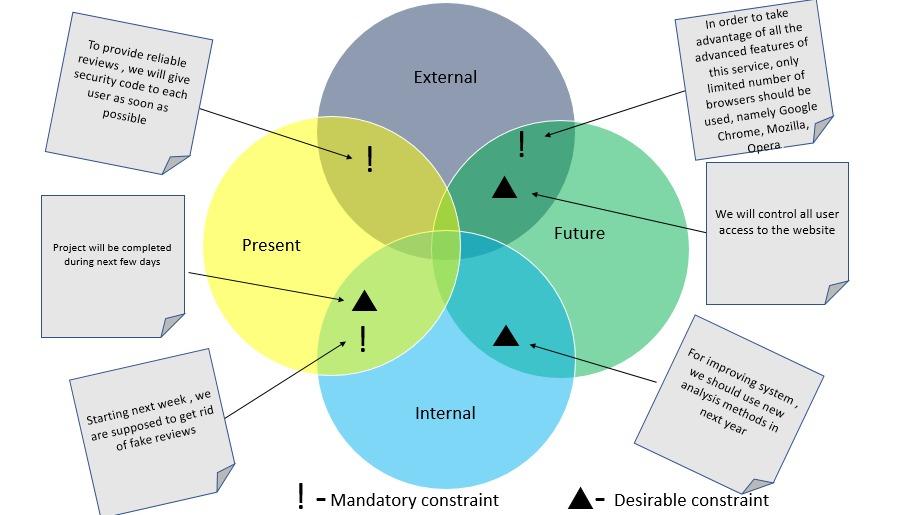
The factors listed above are just small pieces of the common **problem**.

Final part is making summaries of read reviews and feedbacks. Here, all reviews found for each item should be summarized. Then, collected summaries have to be inferenced. Lastly, it’s needed to create a conclusion about the selected product.

In conclusion to all of the problems told before, we can say that the problem itself creates an **opportunity** for us as a solver part and makes our product significant as it solves the listed issues and helps to save the most valuable thing – customer’s time.

## **3.2 Project Scope and Constraints**

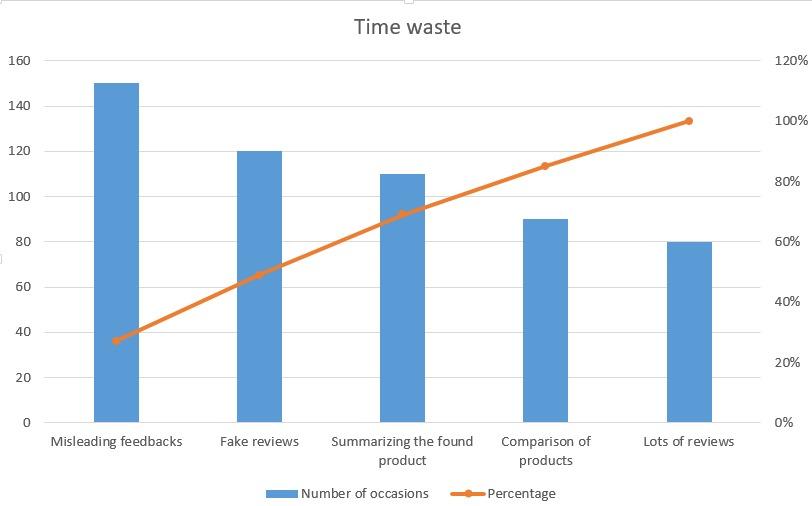
Scope of this project is mainly about designing database which stores processed summaries of reviews. Apart from that, it is also possible to retrieve needed information and data from thta database. In details, it will be much faster if the needed information is formerly processed and stored in database. However, it will obviously take some time to process raw reviews in order to retrieve summary in the end. Coming to the constraints of the project, first and foremost, project must be completed within budget tolerance. Additionally, project deliverables must be completed in the course of one year. Furthermore, in order to take advantage of all the advanced features of this service, only limited number of browsers should be used, namely Google Chrome, Mozilla, Opera.



## **3.3 Fact Findings**

As fact-finding technique, we have used methods of survey and interviews.

### **3.3.1 Survey**

Based on our recent survey, significantly large percentage of people said that, they definitely read reviews whenever they are about to purchase. Even though most people prefer to read others’ feedbacks about a particular product, our findings clearly show that there are bunches of fake or useless reviews that mislead desperate customers. Because of this unfortunate reality, our algorithms will firstly identify and subtract fake reviews from real ones, then start to process and summarize needed information which are based on real feedbacks.

According to the Pareto rule 80 % of the effects come from 20 % of the causes. When we apply the same rule to the project, it can be seen that most people complain about misleading feedbacks. However, it has only 20 % bad effect. On the other side, only 20 % of people mentioned lots of reviews as a major problem, but this problem has 80 percentage bad causes on consumers. It can be concluded that more reviews cause more problems and our project solve this major issue.

### **3.3.2 Interview**

**Interview Steps:**

**1.** Determining people to interview.

**2.** Asking for some needed information related to the projects.

**3.** Evaluation of answers of questions during interview and results.

**4.** Documentation process of interview.

**Interview Questions and Answers:**

1. *What’s the purpose of the project?*

* Purpose of our project is to create a tool which will have all – extension, application, and website which summarize reviews and feedbacks of thousands of customers in online shopping websites.

1. *What are the problems that can be faced in future?*

* As processed reviews will be stored in our database, our database can be overloaded due to the vast amount of stored information over time. We will try to solve it in future, but our main database will be the Google Cloud Service for now.

1. *Who are our competitors?*

* ReviewMeta, FakeSpot, SellerWit Review Analyzer tools and so on are our main competitors which are mostly focusing on figuring our fake reviews, not about using AI din implementation.

1. *What are our differences from other related projects?*

* Admittedly, most functions are the same, however, some significant differences such as implementing AI really stands out in our project.

1. *What there be any changes during the development of the project*

* It’s undeniable truth that every day more and more new technological advances are made and if any of them would improve our project, we will definitely utilize it during the implementation of project.

1. *How advertisement system will be provided?*

* Special parts, especially, left and right sides will be allocated for sponsored ads, namely in our website and time accuracy will be calculated according to the payment made by cooperative companies.

1. *How are you going to deal with financial issues?*

* Paid advertisement services will be provided to all cooperative companies to promote their products in our platform.

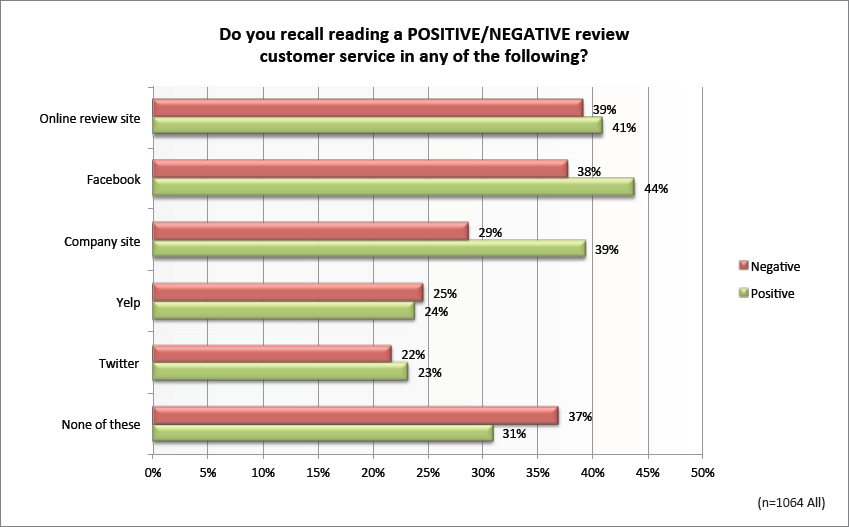
## **3.4 Project Usability, Cost, Benefit and Schedule Data**

Here we are explaining each concept one by one according to our project. Mainly, usability of service, cost, benefit of it and schedule in the end.

**Usability**

* **Thousands of people use this service around the globe** – Our service is popular around the world. Already existed review service like our project has been widely used especially over the last decade due to the rise in the popularity of online shopping.
* **More and more people** – According to the studies vast percentage of customers tends to use review services even though it can be time-consuming time-to-time.

1. Nearly 95% of shoppers read online reviews before making a purchase (Spiegel Research Center, 2017)
2. 94% of customers read online reviews (Fan and Fuel, 2016)
3. 93% of local consumers use reviews to determine if a local business is good or bad (Bright Local, 2017)
4. 72% of customers don't take action until they have read reviews (Testimonial Engine)



**Cost –** Finance section is certainly an essential one to consider. This includes expenditure that has to be spent on building the project, design of it and maintaining the service. Obviously, working developers should be paid and overall cost would be dependent on the number of employers. However, 1000 $ can be estimated for each developer, for example. Apart from that, payment for online cloud services should be taken into account while considering the cost of the project. They cost roughly 200 $ for each month. What’s more, additional fee will be paid in order to keep the service alive, such as to maintain the website.

**Benefit –** Question “what is the benefit and advantage of project?” is vitally important when running a new project. It can be easily noticed that thousands of people waste their precious time while reading the reviews for the particular product they need. These problems will profoundly be annihilated by means of our project. It is also mentioned in Fishbone diagram above in details.

**Schedule –** In the first stage, main service would be released to use customers, not including trivial services of project, such as recommendation system. Secondly, consumers would be able to utilize full advantages of the project to their benefit.

## **3.5 Evaluation of Feasibility**

With the help of Interview Questionnaire and taken surveys we identified the feasibilities. Here main thing is to determine system request that can be a great deal in the development stage.

### **3.5.1 Operational Feasibility**

Main goal of operational feasibility is to solve the problems using new suggested system. Our platform is considered to help online customers who are trying to find the right product by reading and analyzing huge number of reviews. Reading thousands of reviews can be time-consuming and users are not even guaranteed to find what they were looking for, eventually. To eliminate this kind of time loss, it’s advantageous to use this platform. On the other hand, if our platform confronts unexpected words some errors can occur. Besides this, results can be wrong while analyzing tricky sentences.

**SWOT**

* Saving customer’s time
* Accuracy of analysis is up to 85%-90%
* Big community
* Some unrelated ads may discourage potential customers

### **3.5.2 Technical Feasibility**

The most technically difficult part of the project is building a good recommendation system. At the initial stages of the project, we are obliged to use an open source dataset from Kaggle to train the model, since we don’t have any real user data. Quality of open source datasets are suspicious. They may contain plenty of noise and be biased. As the platform scales and attracts more and more users, our data scientist teams can improve the recommendation system by using higher quality data.

As our platform is considered to process huge amount of data, we need to use special analyze technics such as data mining, sentiment analysis, structure mining.

Firstly, we need to have all bunch of databases of reviews from online shopping stores. In order to retrieve target data, special APIs will be provided by co-operatives (AliExpress, Amazon, Ebay). On the next stage, the platform should convert retrieved raw data into summaries by means of data mining algorithms such as C 4.5, Support Vector Machines, etc.

Secondly, to systematically identify and extract subjective information such as reviewer’s feelings, thoughts, judgments, or assessments about a particular topic we are going to use sentiment analysis. Depending on the scale, two analysis types can be used: coarse-grained and fine-grained. Coarse-grained analysis allows for defining a sentiment on a document or sentence level. And with fine-grained analysis, you can extract a sentiment in each of the sentence parts.

Ultimately, another crucial tool we will be using through structuring of this project is Structure Mining (Web) which is mostly for enhancing marketing strategies and use of its functions in advertisement - more specifically, by means of structure mining, users of Review Analyzer – potential buyers or customers will get more appropriate product advertisement. It can be done in 2 ways which are retrieval of patterns from a hyperlink and mining the document structure. Even though most algorithms use first one, like Google’s Pagerank, we will mostly focus on the latter one as we need to read, process and, eventually, obtain summaries of web page or a document. More interestingly, usage of first – retrieval from hyperlinks will also lead us to not only summarize reviews of a particular product through one page, but also it will enable us to get summaries of reviews from similar pages which are about the same exact product.

**SWOT**

* Faster and more efficient solution
* Supports limited online stores, Dependence on external services
* Use of modern AI analysis tools
* External attacks on DB

### **3.5.3 Economic Feasibility**

Use of the service is free, but cost for the development of software should be taken into consideration. To process thousands of reviews and store their summary, we need to pay for services of cloud servers. According to predefined agreement with certain companies, APIs will be provided for free. There will be special ads that are sponsored by companies that provide API. Additionally, in case of purchase by the help of our platform, system charges seller companies fee to some percentage (1,2 %).

**Main costs of development:**

Engine Instance and GPU:

* Google Cloud Engine Instance - 8 GB RAM, 500 GB SSD, 4 cores CPU -105$ monthly
* Nvidia Tesla P4 – 1 GPU GDDR5- 8 GB – 302.0 USD monthly
* Developers and Engineers monthly wages:
* Developers - 4 x 2000 $
* Engineers – 2 x 1500 $
* Advertising costs – 500 $

**SWOT**

* Free
* Lack of marketing expertise
* Income from ads
* Not uniqueness of product – we have competitors

### **3.5.4 Schedule Feasibility**

Schedule feasibility is how much a due date for a system, plan, task or procedure is reasonable and feasible. Initially, free part the project will be released according to the budget. In cases of financial support, implementation of full project will be conducted immediately. To finish the project on time we are supposed to stick to the plan. Otherwise, other competitors can come with the idea.

**Project Timeline:**

* Start date: 15.06.2019
  + Development: 5 month
  + Testing: 2 month
* Finish date: 15.01.2020

**SWOT**

* Some project tasks can be implemented at the same time
* Realization may be delayed due to lack of financial support
* As we use Agile method the developer team can easily work exchanging tasks among each other
* Experienced staff is difficult to replace – sudden lost of them may lead to crisis

## **3.6 Project Request Form**

### **3.6.1 Description of the project**

According to a study on reviews in 2016, 91 percent of consumers regularly read online reviews which sometimes leads to the significant impact or change in purchasing decisions. Obviously, reviews play an important role, but in some cases reading reviews can take days or even several weeks until choosing the right product to purchase. To eliminate that kind of time waste, we will use an AI based software to make summaries of thousands of reviews by clearly pointing out the merits and demerits of any product. Customers will be able to read the summaries of a myriad of reviews in a minute and choose the preferable product by putting less effort and spending less time on.

### **3.6.2 Requested Features**

* **Browser extension –** any user can download and install it (for example, from Google Chrome Extension Store). On the online store page where the desired item is placed, by clicking the extension a user will receive summarized information formed by AI based analyzer ignoring the fake reviews. User can log in to save information about product he viewed.
* **Summarizing Service –** service should have ability to re-escalate vast amount of data and output the summary of it by the result of sentiment analyze.
* **Recommendation –** Based on summaries and users’ history of viewed products, they will get several appropriate recommendations.
* **Advertising –** Besides showing users relevant summaries of reviews for a particular product, users will see products that will be sponsored by some companies in order to get incomes.

### **3.6.3 Main Objective of the Project**

Main objective of this project is to eliminate the time waste that is spent on thousands of online shopping websites by millions of desperate buyers who are searching for the right product. Additionally, by figuring out fake reviews with the help of special tests, the application will prevent customers from being victims of scams and frauds who are trying to boost their incomes by making fake reviews.

## **3.7 Memo**

### **3.7.1 Overview of meeting:**

The scope of discussion is to make a tool that enables customers to easily find appropriate purchase through reading summaries without having need of reading all bunches of reviews related to corresponding product. We discussed about some issues that we face as groupmates, about problems that customers would face in future after the release of this product, and last, but not least, about the troubles that our partner company can face when using this software.

### **3.7.2 Problems and Result Solutions**

As a result, we reach conclusion about discussion (with group mates) which was mainly based on the sections below:

**Problems:**

* **Storage :**

As we are handling “big” data on our way to serve customers, we need to store processed data either in our servers as an offline database, or by the help of cloud server services which are offered by some people who share their servers to store data.

* **Solution :**

In the end, we reached a conclusion that, the latter case, using cloud-based servers, is much more advantegous in our today’s world, which will lead us to focus only on making software itself, but not worrying about how to store and manage stored information. Additonally, as service is offered by experts who are focused on storage of information, we will not have to be distracted by question of “how to manage processed data”. Therefore, we can easily focus on programming part of our work.

* **Tools and Algorithms**

We need a special and appropriate tools and algorithms for handling the overload of big data.

* **Solution :**

We have reached a conclusion that we have three principal tools that will help us to build our project which are Data Mining, Sentiment Analyze, Structure Mining. When it comes to algorithms, focusing mainly on algorithms such as C 4.5, Support Vector Machines, etc will be more beneficial.

* **Application Programming Interface (API)**

We need to have access to the reviews which are on the databases of corresponding shopping websites that we are dealing with. In order to get only needed information (reviews itself), we need to filter junk and useless data, and ,eventually, have exact information that we ,actually, need.

* **Solution :**

We will ask our cooperative companies to have access to all the reviews on the website about products. We will have that access by means of special APIs that are made by cooperative companies.

### **3.7.3 Result of Meeting**

We have reached a conclusion about some crucial points of our project which are storage, tools and algorithms, and, APIs. To sum up, as a storage, using Cloud-based Database services for storage, using tools such as Data Mining, Sentiment Analyze, Structure Mining and employing API to access to the exact information on corresponding website will be advantageous.

## **3.8 Recommendations**

According to the request, existing project will have some flaws and even delays at “peak times”. We hope that recommendations of our team about project will be taken into consideration and help us to get a much better product as an outcome of our collaboration. Firstly, we need to mention that using an Online Cloud-based Database System rather than servers would be much more efficient to employ in order to handle much information flood. It’s not only more efficient from the aspect of dealing with large bunches of information, but also, it’s definitely cost-effective when it comes to buying servers and finding place to keep them. Advantageously, using Cloud-based system enables you to “hire” servers of other people without needing to pay for actual servers, or worrying about finding a place to keep them. Surely, you need to pay for their services, though. Additionally, the word “software” should not only be constrained by only extension as mentioned in request form. It is also offered to have a website and an application form of this service which widens the usage of it.

# **4. Requirement Analysis**

In order to tackle the task effectively, we should employ following tools which are WBS, Gantt Chart, CPM and PERT Chart:

## **4.1 WBS**

WBS serves to give information about the project by dividing it into smaller parts. By the help of it, work has to be done is broken into smaller parts. By means of these structure principal partitions of project are obviously seen. Apart from that, it also helps to organize team work according to the manageable sections. Simply, it is the outline or map of the project. Understanding of major deliverables of the project is definitely advantageous before execution of the project.

Top level represents the final deliverable or the project itself. While going down on the levels from level 2 to level 4 in this particular WBS, work of different departments and employers can be seen with the divisions correspondingly.

## **4.2 Gantt Chart**

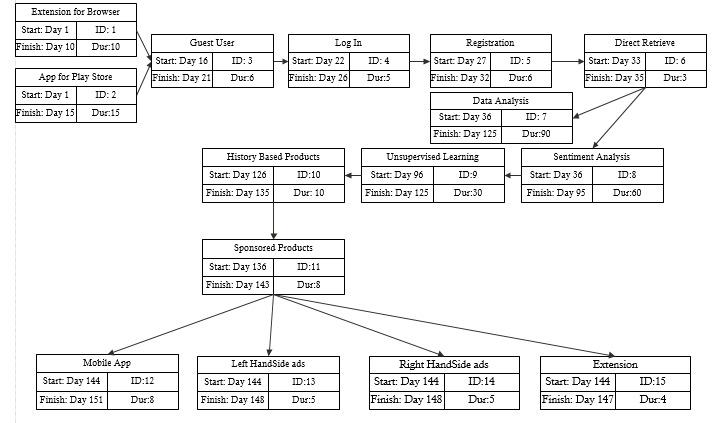
Gantt chart is powerful tool to illustrate activities while construction of the project against time and precedence. Obviously, it is the list of activities and their scale according to the time. Bar also reflects the time needed with the length of it, and position of it describes the precedence of given tasks. Simple, answers the questions of “what the activities are ?”, “when each activity starts and ends ?”, “how long is the time needed in order to execute these activities ?” , “how activities overlap and depend on each other ?” and in the end “when the project is started to be built and planned to be finished ?”.

## **4.3 CPM**

The critical path method (CPM) is a diagram for set of project activities. It shows step-by-step methodology for numerous complex activities which are interacted and dependent on one another. Since CPM is commonly used in conjunction with the program evaluation and review technique (PERT), we also included PERT below. A critical path is determined by identifying the longest stretch of dependent activities and measuring the time required to complete them from start to finish. Our CPM describes how activities in the project interact with one another in a complex manner. Apart from that, precedence activity of tasks is also shown. Direction of arrows shows the sequence of tasks to be performed.



## **4.4 PERT Charts**



PERT chart serves to give information about the project which is about schedule, organization and coordination of tasks within project. Letters o, r, p in this particular PERT show time period needed to execute tasks, such as the worst, the best, and average time. In the end, expected time is also calculated with the equation of (o + 4\*r + p) / 6.

## **4.5 List of Requirements**

We looked detailed into each requirement in this section of document.

1. **User Login –** Registered users will have username and password which leads to the system when logging in.
   * **Front- end** – User-friendly interface which contains all information about services and functions.
     + SublimeText will be used as the main development environment, since it has many opportunities to utilize and adequate libraries in order to create a website. (N.F.R.)
     + Open Source Libraries will be used to make project cost-effective. (F.R.)
     + In order to have a strong security, certificates will be used. (N.R.)
   * **Username and Password –** Authentication process should made when entering the information of username or it can also be email, and password.
     + Username should only be restricted by mails, since mails themselves are unique. (N.F.R)
     + Passwords can only be made with letters and numbers, excluding symbols. (F.R)
2. **User Registration –** New users will be added to the database of system whenever they get registered by entering valid information about themselves.

* **Personal Information** – Filling of some personal information is needed in order to register such as name, surname, age, nationality and so on.
* All shared personal information will be kept private and will not be shared with any third parties (N.F.)
* Since the system uses HTTPS protocols, it is fully secure and protected by any unauthorized use. (N.F.)
* **Confirmation Code** – In order to complete the registration form, confirmation code should be sent to either mail or telephone number.
* To complete registration form all users will be provided with unique confirmation code sent to their e-mail. (F.R)
* The confirmation code will have expiration time which is related to security issues. (F.R)

1. **Summarizing Functionality –** The link of the online store page (URL) where the desired item is placed, should be inputted to the web page form and only after that the user will receive summarized information formed by AI based analyzer ignoring fake reviews. User can log in to save information about product he viewed.

* **Connection** –Software should always be connected to the database in order to retrieve any needed information.
* Connection with the SQLite DB will be established locally at the initial stage of the development. (F.R.)
* **Local DB** – Result data of summarized reviews will be locally saved in local database.
  + - SQLite will be preferred as our offline database management system for the prototype version of the project. (N.R.)

1. **Extraction of fake reviews –** Counterfeit reviews should be extracted from the real set of reviews.

* **Detection of Fake structures –** By following special guidelines and methods, identification of unrealistic comments and feedback will be viable.
* Frequency of review – By calculation of the frequency of a particular reviews and repeated sentences, it will be possible to figure out fake ones. (F.R.)
* Lots of good reviews – Many good reviews for a particular user for a particular product seems to be doubtful. Due to this fact, those reviews will be considered as fake as well. (F.R.)

1. **Learning Capability –** According to the former practice, service should speed up its summarizing functionality and values will be more precisely calculated.

* **Sentiment Analysis -**Data retrieved from DB is indexed and saved for next stages. In further steps, that stored data will interpreted into a language which is more like machine language.
* Specifically as a sentiment analysis tool, TextBlob will be used in order to convert raw data into processed version. (F.R.)
* Meaning of the sentences will be marked as negative or positive according to the indexed words. (N.F.)
* **Unsupervised Learning** – With the help of the sentiment analysis tools, polarity of the reviews will be determined whether it is negative or positive. By time, this polarity will be calculated more precisely by the help of machine learning tools.
* No need to process same reviews with Sentiment Analysis again and again, since service learns by itself. (N.F.)
* As more and more reviews are processed through service’s “hand”, more the precise will be summaries of further reviews. (N.F.)

1. **History-based Promotion -** Based on summaries and users’ history of viewed products, they will get several appropriate recommendations.
   * **AI Recommendation –** Users will see products on his page according to our AI based Analyzer which processes user’s choice and shows corresponding ones
   * Sentiment analysis, Structure mining, Machine Learning will be our main tools in order to process users’ choice and show appropriate based on that. (F.R.)
   * During implementation of these analysis methods, Python libraries (Selenium, etc) would definitely make things easier for us. (F.R.)
2. **Ads Service–** Besides showing users relevant summaries of reviews for a particular product, users will see products that will be sponsored by some companies in order to get income.
   * **Fee** - There should be payment system for related companies who want to advertise their product on this platform and obviously, payment will be calculated according to the number of clicks.
     + Payment will be calculated according to the number of visits to the promoted products and websites. (F.R.)
   * **Ads –** Both right and left sides will be allocated for ads.

* According to the stats that more people tend to click ads on the right side, those on the right side will be pricey to hire. (N.R.)

1. **Display of summaries –** Summaries should be shown to the users after the process of initial extraction of fake reviews and sentiment analysis in the end.

* **Retrieve from database –** Processed summaries which will later be stored at our database will be retrieved in order to show them when requested.
* A Request will be sent to the database as a special query. Information related to the selected items will be retrieved if found. It then will be displayed in a user- friendly user interface. (F.R.)
* To boost the levels of readability of summaries, most crucial sentences will be signed as bold and underlined. In detail, some important key points will be promoted over other sentences (N.R)
* **Visualize retrieved data -** With the help of several data visualization types such as graphs and charts, related stats will be shown to users right after the sentiment analysis process.
* For visualization the retrieved data, Python libraries such as Leather, Seaborn, Pygal, Plotly, etc. will be used. (F.R.)

1. **Display of recommendations –** Users will see recommendations based of search history and seen products after logging in to the system. So that system can handle the search history of particular users.

* **History-Based Analysis -** According to the history of the previous analyses of a particular product, relevant suggestions will be offered to the users.
* A Special counter will calculate the frequency of particular clicks and, obviously, whenever they are clicked, relevant data will be written to the history of that particular user. (F.R.)
* According to the frequency of the clicks and their comparison with other products, some particular type of purchases will have precedence over others, and will be seen as recommendations (N.R)

# **UML Diagrams**

## **Context Model**

Project is divided into 6 principal portions which are Data Processing System, User System, Installation System, Summarising System, Recommendation System and Advertising System. All then will be explained in detail in the following.

## **5.2 Activity Diagrams**

UML is one of the tools which is advantageous to visualize and document software projects. Activity diagrams is about flowcharts which show the activity and workflow that happens in the system.

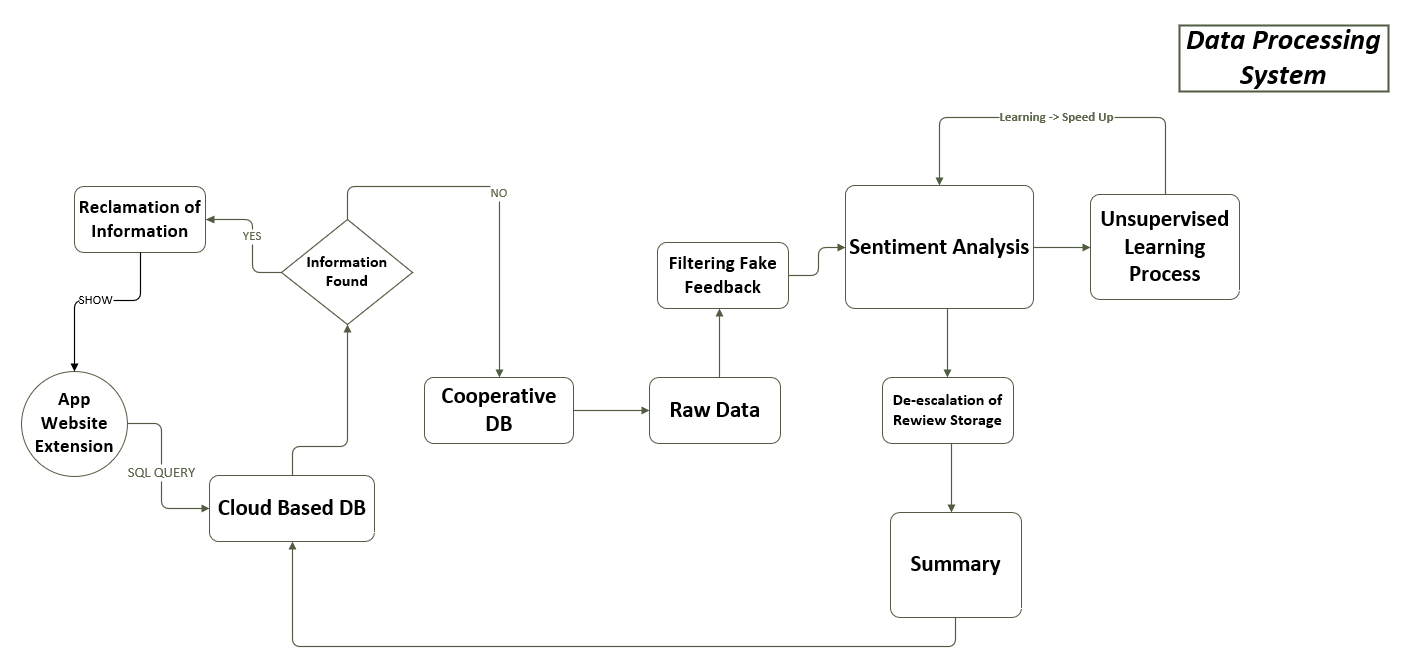
### **5.2.1 User System**

This activity diagram describes how the user is interacted with the system through GUI which has varying forms such as Browser Extension, Website and Application. Obviously, user initially sees First Page which gives options of Log In, Registration and Guest User. When choosing Log In or Registration section, relevant valid data should be filled in order to use the system with its all advantages. Otherwise, only guest option is left where users can only have the limited prieveleges to the system and its services.

### 

### **5.2.2 Data Processing System**

Starting from the Cooperative DB where the raw data is provided by, the process of filtering fake feedbacks is followed. Then, using one of the analyze methods which is Sentiment Analysis, reviews are marked according to the meaning, namely positive or negative. On the following stage, de-escalation of file happens, where data is decreased in volume while keeping the percentage of appreciation or deappreciation.



## **5.3 Use case diagrams**

Use case diagram give a simple overview about interaction between system and external actors. Additonally, it is about roles and how users exploit system through their roles.

### **5.3.1 User System Use Case**

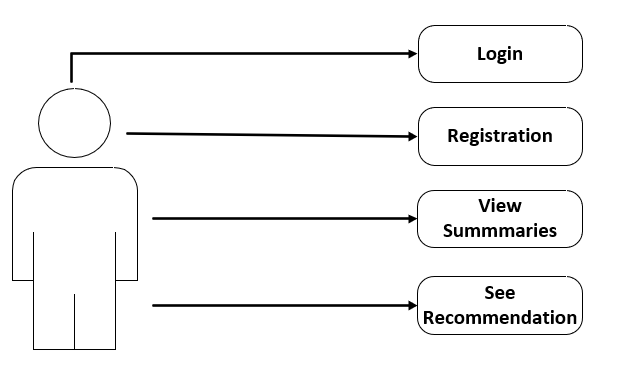
This use case is about user requirement which is written as Software in Preliminary Report.

**1. User Login –** Registered users will have username and password which leads to the system when logging in. (U.R.)

**2. User Registration –** New users will be added to the database of system whenever they get registered by entering valid information about themselves. (U.R.)

**9**. **Display of summaries –** Summaries should be shown to the users after the process of initial extraction of fake reviews and sentiment analysis in the end.

**10**. **Display of recommendations –** Users will see recommendations based of search history and seen products after logging in to the system. So that system can handle history of that particular users.



1. **In order to use service, log in is required.**
2. **Sign up is mandatory in case user does not have an account.**
3. **After signing up, user now has privilege to view the processed data and summaries.**

**4. User is recommended with some relevant products according to search history.**



## **5.4 Sequence diagram**

### **5.4.1 User System**

**System Requirements and Functional and Non-Functional Requirements:**

* 1. **Front- end** – User-friendly interface which contains all information about services and functions.

**1.1.1** IntelliJ IDE will be used as the main development environment, since it has many opportunities to utilize and adequate libraries in order to create an extension. (N.F.R.)

**1.1.2** Open Source Libraries will be used to make project cost-effective. (F.R.)

* 1. **Username and Password –** Authentication process should made when entering the information of username or it can also be email, and password.
     1. Username should only be restricted by mails, since mails themselves are unique. (N.F.R)
     2. Passwords can only be made with letters and numbers, excluding symbols. (F.R)

2.1 **Personal Information** – Filling of some personal information is needed in order to register such as name, surname, age, nationality and so on.

**2.1.1** All shared personal information will be kept private and will not be shared with any third parties (N.F.)

**2.1.2** Since the system uses HTTPS protocols, it is fully secure and protected by any unauthorized use. (N.F.)

**2.2** **Confirmation Code** – In order to complete the registration form, confirmation code should be sent to either mail or telephone number.

**2.2.1** To complete registration form all users will be provided with unique confirmation code sent to their e-mail. (F.R)

* + 1. The confirmation code will have expiration time which is related to security issues. (F.R)

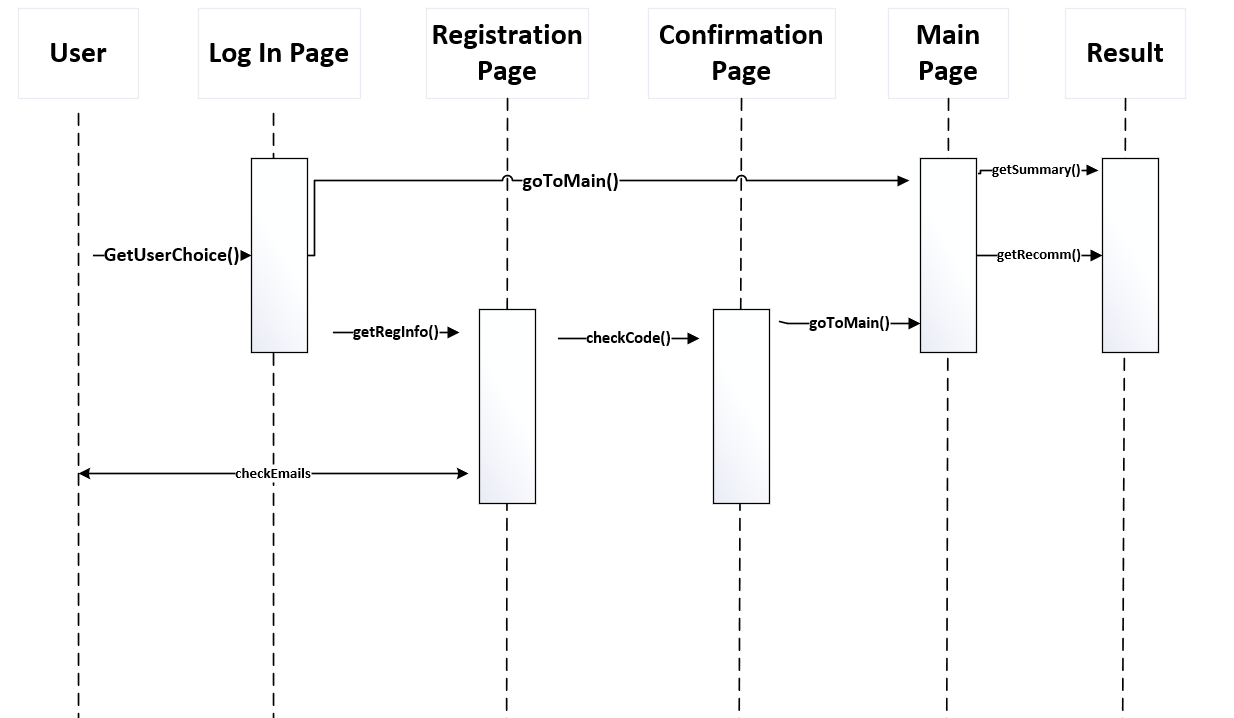
**9.1 Retrieve from Cloud-Based DB –** Processed summaries which will later be stored at our database will be retrieved in order to show them when requested.

* + 1. A Request will be sent to the database as a special query. Information related to the selected items will be retrieved if found. It then will be displayed in a user- friendly user interface. (F.R.)
    2. To boost the levels of readability of summaries, most crucial sentences will be signed as bold and underlined. In detail, some important key points will be promoted over other sentences (N.R)

**10.1 History-Based Analysis -** According to the history of the clicks on a particular purchase and appearance of them in search bars of software (App, Extension or Website), relevant suggestions will be offered to the users.

**10.1.1** A Special Counter will calculate the frequency of particular clicks and obviously whenever they are clicked, relevant data will be written to the history of that particular user. (F.R.)

**10.1.2** According to the frequency of the clicks and their comparison with other products, some particular type of purchase will have precedence over others, and will be seen as recommendations (N.R)





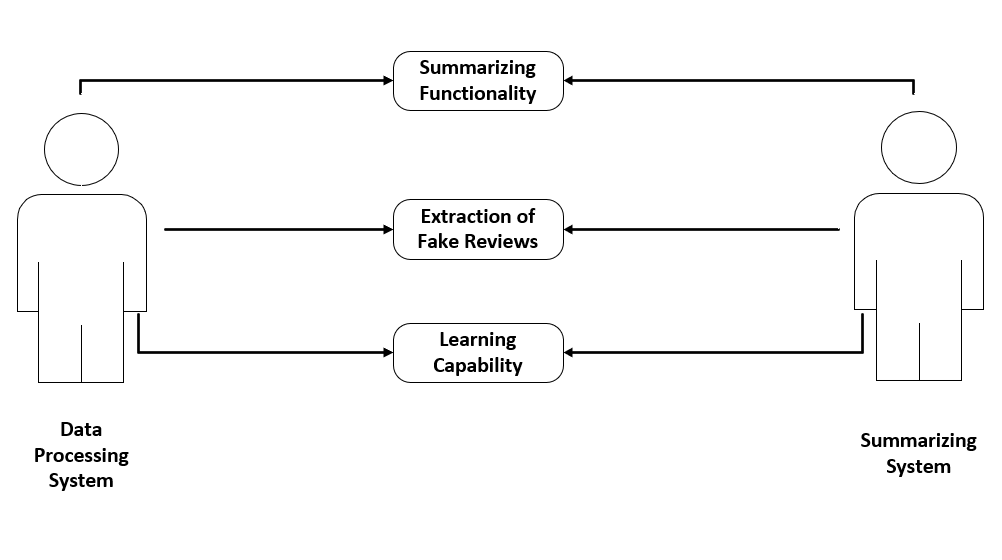
When coming to the first page, user sees 3 options which are using service as guest user, log in page and registration page. If registration option is chosen, main page will be seen after confirmation page where users fills personal information. With the checkCode() function, confirmation code is checked if valid or not. If it is valid, registration is complete and user can login and goto main page which is expressed with the goToMain() function.

* + 1. **Data Processing System and Summarizing System Use Case**

This Sequence Diagram comes from user requirements below:

**3. Summarizing Functionality –** any user can download and install it (for example, from Google Chrome Extension Store). On the online store page where the desired item is placed, by clicking the extension a user will receive summarized information formed by AI based analyzer ignoring the fake reviews. User can log in to save information about product he viewed.

**4. Extraction of fake reviews –** Counterfeit reviews should be extracted from the real set of reviews.

**5. Learning Capability –** According to the former practice, service should speed up its summarizing functionality.

In this use case diagram 2 Systems and their use cases are illustrated as Summarizing Functionality, Extraction of Fake Reviews and Learning Capability. While going the chart, we see **summarizing functionality** where reviews are re-escalated to a smaller storage after **fake reviews being extraction** initially. To add, on the background **learning capability** of system works where services perfects itself day by day by learning from the past experience.

* + 1. **Data Processing System and Summarizing System Sequence Diagram**

**3.1 Connection** –Software should always be connected to the database in order to retrieve any needed information.

* + 1. Back-up connection line will be kept in case of any emergency which can result in disconnection in the main connection line. (N.R.)
    2. To create connection between our database and cooperative companies’, special API will be provided. (F.R.)
  1. **Cloud-based DB** – Several factors can be mentioned about preferring cloud-based. First and foremost, we will be need to have any additional work force in order to manage the database. Second, it’s more modern and reliable in today’s world and we are guaranteed from any data lost. Additionally, it’s cost-effective as no need for spending money on offline servers.
     1. Google Cloud will be preferred as our cloud-based DB service. (N.R.)
     2. GPU Engine – Nvidia Tesla P4 will be used, since it is powerful engine for lower price. (F.R.)
  2. **Detection of Fake structures –** By following special guidelines and methods, identification of unrealistic comments and feedback will be viable.
     1. Frequency of review – By calculation of the frequency of a particular reviews and repeated sentences, it will be possible to figure out fake ones. (F.R.)
     2. Lots of good review – Many good reviews for a particular user for a particular product seems to be doubtful. Due to this fact, those reviews will be considered as fake as well. (F.R.)
  3. **Sentiment Analysis -**Data retrieved from DB is indexed and saved for next stages. In further steps, that stored data will interpreted into a language which is more like machine language.
     1. Specifically, C 4.5 and Support Vector Machines algorithms will be used in order to convert raw data into processed version. (F.R.)
     2. Meaning of the sentences will be marked as negative or positive according to the indexed words. (N.F.)
  4. **Unsupervised Learning** – Service trains itself by learning from the past experience on former processed reviews and that, in turn, lessens the processing period needed for further request on same products.
     1. No need to process same reviews with Sentiment Analysis again and again, since service learns by itself. (N.F.)
     2. As more and more reviews are processed through service’s “hand”, more the precise will be summaries of further reviews. (N.F.)



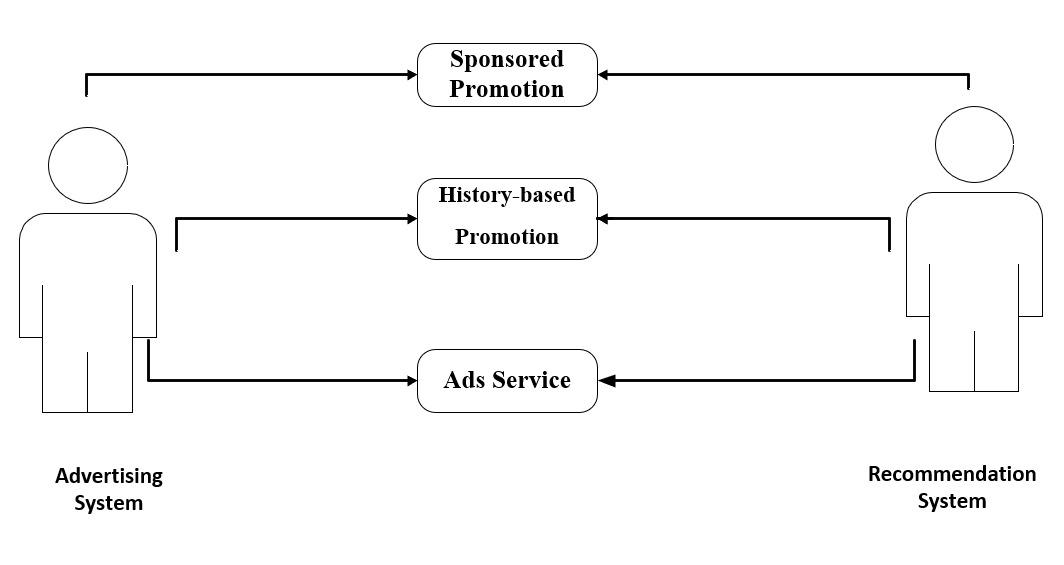
To increase accuracy of result we need to identify fake reviews from dataset and filter them. There are several stages that leads us to identify. Firstly, review dataset is tokenized to resolve errors such as incompleteness, inconsistency of data. Next stage is the process of reduction, where an initial set of raw variables is reduced to more manageable groups (features) for processing, while still accurately and completely describing the original data set. Then, content classification method is used for managing text and unstructured information by categorizing and clustering them. By labeling natural language texts with relevant categories from a predefined set, automatic document classification enables users to organize content quickly and efficiently. Additionally, categorized data set is analyzed using NLP tools to identify the objectivity and using the result of analysis returns us whether these reviews is spam or not.

* + 1. **Advertising and Recommendation Systems Use Case Diagram**

**6. Sponsored Promotion –** In order to financially support the project, particular products will be promoted according to the agreement of sponsorship.

**7. History-based Promotion -** Based on summaries and users’ history of viewed products, they will get several appropriate recommendations.

**8. Ads Service–** Besides showing users relevant summaries of reviews for a particular product, users will see products that will be sponsored by some companies in order to get income.



In these use case 2 systems which are advertising and recommendation systems illustrated in the chart. Here, sponsored promotion, history-based promotion and ads services are noted. In the sponsored products services, there will be paid advertising and according to the clicks to them, recommendaed products will also be modified. Then, in the history-based promotion, users see advertising according to their users, while getting recommendation as well. Thus, in the end, ds service works both according to the advertising system and recommendation system.

* + 1. **Sequence Diagram below come from the following system, functional and non-functional requirements**

**6.1 Sponsored Recommendation** – There will be sponsored products which users going to see. Cooperative companies will choose foregoing products.

* + 1. Sponsored products will be promoted according to the agreement with cooperative companies. (N.R.)
    2. By the help of provided API, corresponding sponsored products will be chosen and promoted above other others. (first in the search results) (F.R.)

**7.1 AI Recommendation –** Users will see products on his page according to our AI based Analyzer which processes user’s choice and shows corresponding ones

* + 1. Sentiment analysis, Structure mining, Machine Learning will be our main tools in order to process users’ choice and show appropriate based on that. (F.R.)
    2. During implementation of these analysis methods, Python libraries (Selenium, etc) would definitely make things easier for us. (F.R.)
  1. **Fee** - There should be monthly payment system for related companies who wants to advertise their product on this platform.
     1. Advertisements will be scheduled according to the amount of paid fee. More the payment is, more the time and priorities will be made. (N.R.)
     2. Payment will be calculated according to the number of visits to the promoted products and websites. (F.R.)
  2. **Ads –** Both right and left sides will be allocated for ads.
     1. According to the stats that more people tend to click ads on the right side, those on the right side will be pricey to hire. (N.R.)
     2. We will get additional cashback if customers visit and buy a particular product by the advertisement link on our website. (N.R.)



To provide our software with appropriate advertisements, firstly, a request for an ad is sent. Next, a query for a list of summary titles is executed with following respond as gettingTitles(). At the same time, user history is viewed for key words. Comparing new data coming from summaries user history sends information about viewed products to Ads block where all readings are analyzed for correct generation of most suitable ad. The chosen one will be returned to software as a respond.



## **5.5 Class Diagram**

Class Diagram is one of the tools that is used to explain how the project work. Here project is explained with the help of classes and their connections. Classes can be connected with help of 5 connection relationship, however in this project 3 types of relationships are used namely, association, composition, aggregation. Association relationships are used among classes of User, Ads and Recommended Products which means classes are connected in “any way”. Composition relationships are among classes of Summary, Viewed Products, Filtered Data, Review and Product. In detail, this connection means one part is not existed without the other one. It is illustrated with black filled quadrangle between classes. Aggregation connection are among classes which are dependent on one another which are classes named Viewed Products, Recommendation Products, Summary. Inside of classes there are 2 types of items which are attributes of related classes – parameters and variables, and class operations-methods. Plus and minus signs denote public and private attributes and operations correspondingly.



## **5.6 E-R Diagram**

Our entity relationship diagram shows the relationships of entity sets stored in our database. Entity set is a collection of entity and each entity has its own attributes. Attributes define the properties of entities. Connections between entities is illustrated with the help of diamonds, attributes with ovals, and entities with rectangles. Entities can be connected in various ways such as one to one, one to many, many to many. For instance, since each user can see many ads, connection between them is assigned as one to many. Additionally, one summary can be shown to many users and one user can see many summaries as well. Thus, connections are denoted as many to many. Furthermore, each product has one and unique summary and one summary belongs to one product and is unique with its unique product id.

# **Conclusion**

This document contains some crucial information about the project of AI based Review-Feedback Analyzer. Foregoing project is for processing tons of reviews and giving the user feedback in the form of several sentences. In this document, every single detail is included about the project starting from the simple description to the explanation of classes in the database in the end. During planning the work through the development of the project, we determined the list of requirements and our tasks. Apart from that Preliminary Investigation Report is made which contains all of them. Starting with the Fishbone Diagram which talks about the threats and problems which will be combatted, Project Scope and Constraints are also determined. In the initial stage before starting to build the project, several fact finding methods are used which are surveys and interviews in this particular document. They are also described visually with the help of Pareto Chart. Before coming to the feasibility study of the project, Project Usability, Cost, Benefit and Schedule Data is determined as well. Talking about the feasibility part, all sections are explained with the help of SWOT analysis starting from operational feasibility, to technical, economic and schedule. After all, Project Request Form is also included which explains the main user requests. Here, main objective of the project which is eliminating time waste on reading reviews is also explained. Next, in the Memo, problems and their solutions are written which are the results of the meeting. According to this consensus among group members, Recommendations about the project and user requests are also included. Next, requirement analysis comes, where Work Breakdown Structure of the project is profoundly explained and visualized. Gantt chart is inferred according to it and illustrated. Gantt Chart, CPM and Pert Chart show the precedence and time needed in order to conduct each partition of the tasks. On the next stage, all requirement is classified in the list of requirements section. According to these requirements, several UML diagrams included on the next pages. First-off, context model describes the partitions of the project dividing it into 6 principal portions which are Data Processing System, User System, Installation System, Summarizing System, Recommendation System and Advertising System. After that, Activity Diagrams are included explaining the flow of activities in the Users System and Data Processing System. According to these, Use Case diagrams are added where general interaction between systems and external factors are described. Apart from that, sequence diagrams are illustrated which shows the flow of activities while conducting the project. Next diagrams are about Class Diagrams where project is explained with the help of several classes and their interactions between them. Likewise, E-R diagram follows almost the same conception while having attributes in classes. As a result, all sections of the project are expressed clearly and one can easily understand any single detail about the project starting from the straightforward description of it to the complicated diagrams such as E-R diagram in the end. After implementation of all expresses tasks, efficient software product will emerge which eliminates the problem of time waste while reading on reviews when searching for products to buy on online platforms.

# **References**

AI Tools to be used

* <https://www.techworld.com/picture-gallery/apps-wearables/best-ai-machine-learning-tools-for-developers-3657996/>
* <https://www.altexsoft.com/blog/business/sentiment-analysis-types-tools-and-use-cases/> <https://selenium-python.readthedocs.io/>
* <https://textblob.readthedocs.io/en/dev/>

Project Management

* <https://www.projectsmart.co.uk/evolution-of-project-management.php>
* <https://www.pm4dev.com/resources/free-e-books/130-development-project-management/file.html>

Training of AI

* <http://ai.stanford.edu/~amaas/data/sentiment/>

Estimation of time in our Gantt chart

* <https://www.mindtools.com/pages/article/newPPM_01.htm>
* <https://www.quora.com/How-long-does-it-take-to-build-a-website>
* <https://blog.teamweek.com/2018/09/how-to-create-an-accurate-time-estimation/>
* <https://www.teamgantt.com/what-is-a-gantt-chart/how-to-make-a-gantt-chart>

***Systems Analysis and Design Tilley | Rossenblatt 11th Edition.pdf***