TED UNIVERSITY

Faculty of Engineering

Department of Computer Engineering

CMPE 491 (Senior Project 1) – Project Specification Report

by

Yağız ÇİMEN – Efe TONTU – Mesut Nadir SEYFELİOĞLU

Supervisor: Assoc. Prof. Dr. Tansel DÖKEROĞLU

Table of Contents

1 Introduction

1.1 Introduction

As one of the most hyped terms in the market, Big Data also took the interest as senior projects. On the other hand, there is one more topic that is as popular as Big Data which is social media. This situation leads us to do a project that combines both these important and popular topics. Nowadays most of the software companies It is a very exciting area to work in because big data has lots of demands on the software market and social media has lots of users. Thus, we thought that it would be a very important thing in a couple of years. So, creating a big data system that predicts communities through social media data will change the eCommerce world and advertising work. The main reason for this change is, advertisement companies will know what kind of advertisement will be better for specific groups, and eCommerce companies can easily define their possible customer's thanks to these kinds of systems.

1.2 Description

As we know that commercials are very important for e-commerce and companies that work for commerce. Nowadays lots of companies prefer advanced, faster communication platforms like social media to advertise their products and reach out to more customers. We figure out if we can create a system which finds the communities in the social media and gives information to companies about these communities. For this purpose, the best thing to do is creating a system that helps these kinds of companies with big data systems. That’s why as a project group we decided to create a system for eCommerce and advertising companies. This system will search through hashtags, keywords, and comments through social media and it will find user’s communities by their social media sharing. This will help companies who want to advertise their products. By using this application company can easily find their communities that are related to their product. Additionally, these popular communities can be a great way to advertise their products through social media. The latter feature will be for companies who want to see their competitors' results so that they can build their marketing strategy by investigating their competitors' strategy.

1.3 Constraints

1.3.1 Time

We have a limited time to complete our project. This can affect the projects capabilities.

1.3.2 Cost

We are trying to make the project as free as possible. Accessing data which requires money can be challenging.

1.3.3 Human Resources

Being a 3 people team can be challenging. There is so much work to do so this can be a problem.

1.3.4 Health Concerns

We are doing this project most of the time on online meetings because of COVID-19. Because of that solving some problems can be hard to accomplish. Being in a working environment is much easier for problem solving.

1.3.5 Access to Data

Accessing personal social media account data can be a problem. We have to take consent to use that information.

1.4 Professional and Ethical Issues

1.4.1 Honesty

"Honesty, responsibility, respect and fairness are the values that drive ethical conduct for the project management profession," according to the Project Management Institute (PMI). Being honest towards group members is extremely important. It helps solving existing problems and confusion about the projects current state.

1.4.2 Fairness

Being fair is important because in a group project if someone feels like they are doing most of the job, it can cause arguments between group members. Making equal distribution of work is crucial for preventing such arguments.

1.4.3 Protection of Personal Data

While gathering data is important to a project, keeping them safe is more important. For protection of personal information, a good security system is very important to seek the benefits of the customer. A leakage of information from the company database is unacceptable.

1.4.4 Backstabbing

When it comes to group projects, then backstabbing is very common. Some members may steal the idea of the project and sell it to other companies. It is highly unethical.

1.4.5 Blaming Others

When in a group project, failure of one person is considered as a group failure. However, in those situations most people choose to blame the one person for the whole failure without thinking that being a group means that we are responsible for all our actions as a singular unit.

1.4.6 Biasness

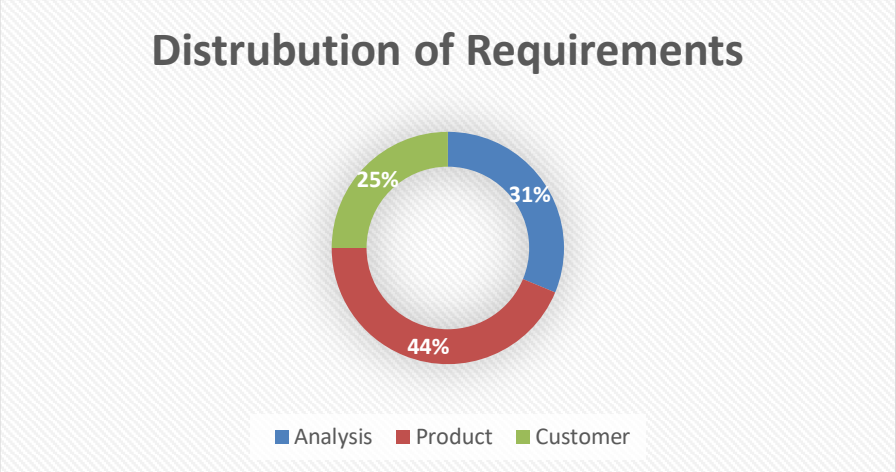
Biasness is common ethical problem. If in the making of the project someone makes a biased decision which makes the other group member disturbed, it can cause a lot of problems. In these situations, the greater good of the group members should be the focus of the decisions.

2 Requirements

For our Project, there are three main requirement groups to be considered. For the first phase, we have analysis requirements that include software technologies and other hardware components to run stable and correct analysis over big data. Analyzing big data cannot be done by traditional techs. There must be specific tools to compete in the process like the Hadoop environment etc. Analyzing big data requires some hardware specifications with it, like powerful computers and processors.

On the other hand, we need to clarify the production requirements. The plan for deployment is to convert to a mobile application in which companies and regular users can access and run algorithms better. For that purpose, we need to have an application that can work on many operating systems like IOS and Android. Production should be available 7/24, manage risks like information security, back-up, etc.

Third thing to be considered is customer requirements. we should have satisfied users. To achieve that goal, we are going to plan our requirements to reach maximum customer happiness. We aim to achieve a user-friendly application, with all the risk management and legal issues.



2.1 Analysis Requirements

2.1.1 Work Environment

2.1.1.1 Hadoop Environment

Hadoop HDFS is the system we are going to store our big data. HDFS gives us distributed file system which means storing data by separating computers. Managing big data is a problem so our project must be fast and reliable. In order to declare master and slave computers we started with developing a Hadoop environment.

2.1.1.2 Apache Spark

Spark is the most popular big data analysis tool nowadays. Apache Spark contains lots of libraries to make it easier. Big data analysis is a hard thing to do. Spark is much faster than Hadoop analysis system. As a requirement we need to add spark technology to our project

2.1.1.3 Java, Python Technologies

In out project we are going to need JAVA and PYTHON technologies to run all this system. Hadoop HDFS system works on java basically. Also spark libraries can be implemented both in java and python programming languages. Our aim is to implement binary classification and machine learning algorithms to achieve trustful analysis. Also, we want to show the performance comparison between java and python on algorithms.

2.1.1.4 Datasets

We have some big datasets to analyze. Our project is going to be on social media platforms for that reason we found our datasets from YouTube. We are going to analyze these data sets in order to find and detect the communities in social media.

2.1.2 Performance

Our project must be efficient and reliable. For that reason, we preferred to use Java python and spark technologies. For big data processing the kind of tools and technologies are proved that by their performance. On the other hand, our team going to compare the algorithm speeds on both java and python.

2.1.3 Reporting

As another requirement we should have reporting. Reporting is one of the most important things in progress. So, documentation the results and problems in order to make process more efficient is quite important.

2.1.4 Supporting Articles

To make our project we need some articles from real world. As a result of our search we have found some articles related to our project which has done by different institution's. You can find the links for article in References part.

2.1.5 Traceability

We should trace the events and their effect to our project in order to make it more efficient. We must know the work principles of each system and algorithm to design better products.

2.2 Product Requirements

2.2.1 Operating Systems

Our system must work on both ios and android operating systems. In future we can develop our project as a desktop application. In this case we should make it compellability to windows Linux and mac operating systems.

2.2.2 Maintenance

Our team should run performance and correctness check in a plan in order to balance the maintenance of our project. As a requirement we begin to plan our test dates for project.

2.2.3 Security

Information security is the most important thing for our project. Because we are working on bigdata and real world datas. For that reason, we cannot allow anyone to use our private data for another reasons. For both mobile and desktop applications we should run security checks on each operating system to have a secure application.

2.2.4 Real Time Estimation

Another important thing to be consider is to have an application that can real time estimations. As new datas arrive to our system we should analyze and update the system eventually without storing them for future progress.

2.2.5 Minimum hardware for Big Data Process

We are working on big data so we should have powerful computers and processors to complete operations like search update add delete…

2.2.6 Accessibility

Product must be accessible from everywhere. We cannot foresee the future. There might be customers with different equipment's. So as a requirement we need to add an accessibility from everywhere feature to our project.

2.2.7 Back-up and Restore

Backing-up and restoring big data is not an easy goal to achieve. But working on big data comes with responsibilities. We should arrange a storing system and restoring system to decrease customer unhappiness. Not just satisfaction but losing data can cause more problems for analysis.

2.3 Customer Requirements

2.3.1 User Friendly

Applying a user-friendly project can lead us to better performance. User friendly interfaces can decrease the distraction for customers and make them easy to develop filters to search communities.

2.3.2 Project Branding

We are going to use Social media platform to brand our project to world. As we are working on social media it is the best platform to make customers hear our voice.

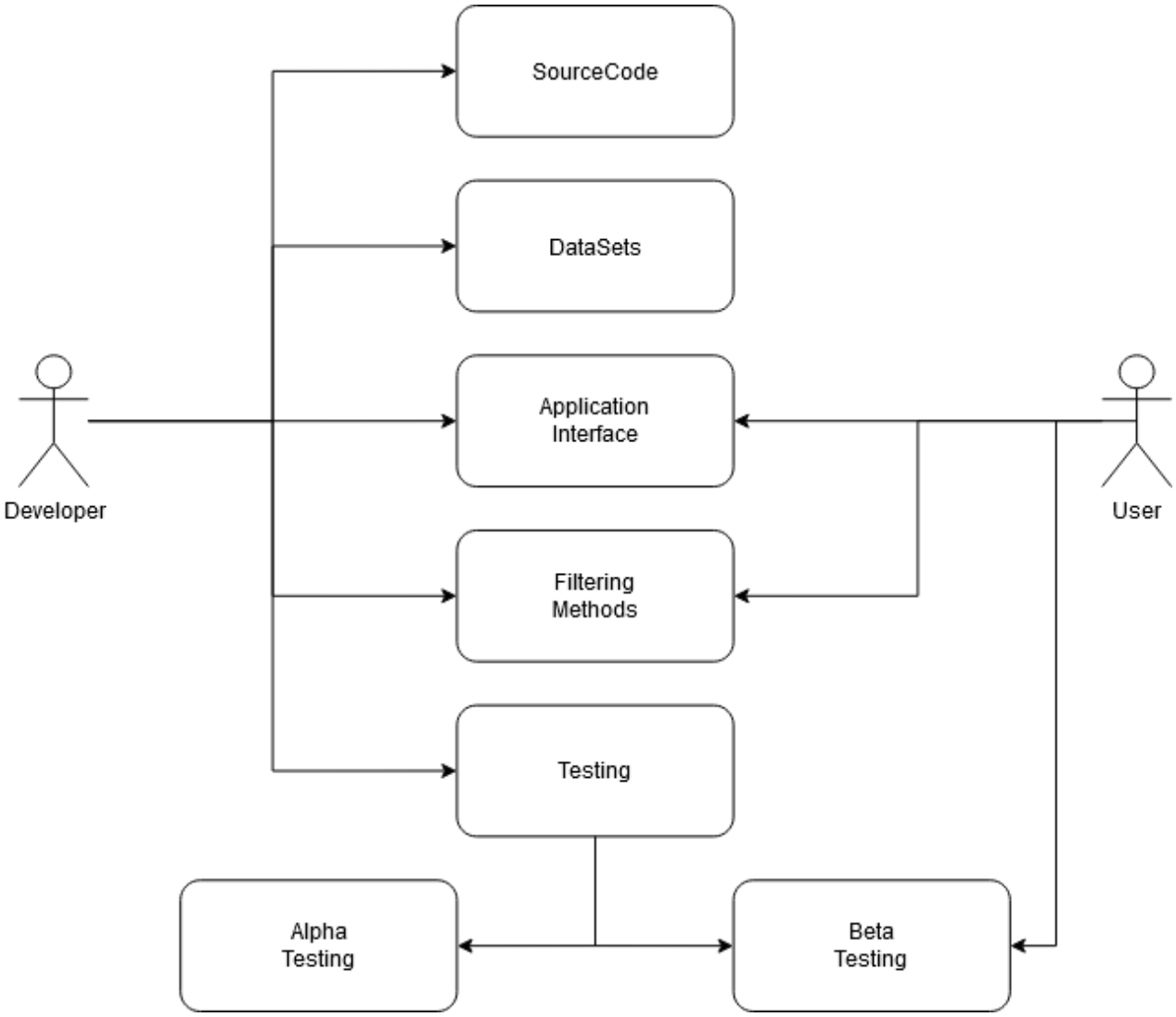
2.3.3 Project Risk Analysis

Risk analysis can occur in every phase of project. Testing implementing. Reporting… In each part we should consider the risks and report them as documents. As a requirement applying Risk analysis simultaneously while developing the project is quite important to have bug-free applications.

2.3.4 Legal and Compliance

As our team explained before in Ethical issues part, as a requirement we should consider the rules of legality to develop right project. We don’t want any of our customers to complain about the features of our project.

2.3.5 Use-Case Diagram



3 References

Reference to Ethical İssues

Reference to Constraints

Article 1

Article 2

Article 3

Article 4

Article 5

Article 6