## Introduction

Every day we deal with unlabeled data. And every document has certain words or meaning which helps to convey certain meaning for text. So, while analyzing document clustering can play vital role. So, I am going to build topic identification with clustering using k-means algorithm. Machine learning has many methods like supervised, unsupervised and reinforcement learning method. And clustering falls under unsupervised learning method because in this method we don’t have specify computer where to look it and what to look after. By analyzing certain perimeters, it classifies data with certain command or instructions. Topic identification will help people to find out the keywords or topic which will help them to know about their contents. First my system will analyze contents, preprocess it, vectorize it and train, then it will detect its topic and flask it will used to handle in on server and client side will be implemented by using angular. So, there are many parts like clustering, server handling and client-side handling. It will take knowledge of different area to get work done. To build this system there are thing that needs to be learned in limited time. In our 2nd year we learn about artificial intelligence, but it was only surface, and we did not touch coding part in python. This makes it hard to create build because I have to learn many things by myself. In this project development, my knowledge and experience of JavaScript framework angular is only thing that I don’t have to start from bottom. But for server-side development by using flask and clustering using k-means, I have to learn from ground up. So, I have to learn and implement parallelly for development of the system. And there are also diverse ways to getting work done. Therefore, I have to see, test and analyze different way to get work done.

Although, I have clear road map and willingness to walk but human needs proper management, discipline and attitude to get task done. Building this system and its implementation requires required more dedication and serious planning. To get his done, there are so much things to learn, trying out new thing and implementing. And I have to manage time, maintain momentum by constant motivation and stay focus on my task. And I also have to adapt changes that will occur during project. Since, I am doing what I’ve never done because so many things I may feel down and may my estimation will not work out that way I’ve plan. And there’s where optimism and communication with supervisor will play vital role for guiding and showing way. So, there are not just technological aspect but also psychological and philosophical aspect for development because where there is human involvement, there’s also human problem arises. For example, I am procrastinator, but I am optimistic. So, this thing can also affect timeline and quality of the project, so I have to keep in mind about these things also and find way to overcome my weakness and minimize their effect on my project. So, in this reflective report, I am going to write about my personal experience about this project.

## Self-management

Every task needs certain discipline and way to get things done. So, to also get this task done we need to implement rules, plans and stick with them. Due to condition that arises thing could be different but to get task done, there are certain discipline or self-control that we have to follow to build that product that we want to create. And above all, I also have to know my own skills and compare with complexity of the project. And after realizing that gap between I have to fill it with all my effort. First estimation is necessary because there are much more factors which can affect the quality of the project. So, not just technical and academic but I’ve to also work on my personal aspect like self-motivation.

### Self-motivation

Our quality of the work depends on our quality of our thoughts and moods. So, mode swing, personal problem or other thing can affect work in invisible way. So, for this I have to constantly keep myself motivated. First, I was excited about this project and I knew it’s going to be hard for me. And when you try to do thing that you haven’t done then we might fail. And I was aware that I am going to face this problem. and also, it’s easy to get distracted in resources and misinformation given in internet. Also, other common problem such as procrastination, human biasedness, overthinking and fear of failing. Without overcoming these problems and mentally being ready for challenges it’s hard to produce quality project.

I believe everything that we feel is governed by our conscious and unconscious reasons. For example, when we fail it’s common to feel I am totally wrong. But I always come across that it’s because lack of understanding that fail us. So, I start with something else or review my work. And since, I love to do myself but there are sometimes thing gets out of my reach. And then when my supervisor Dr. Swati Sah helps with insight and understanding. I am the person who loves to do own my own but whenever I ask for ask to my supervisor, she always tried her best to provide me knowledge, insights, help in best possible way and I am thankful for everything. And procrastination is also another main problem for me. I admit that I am procrastinator. It’s common for me to think too much but do nothing at all. So, to overcome I’ve made routines that could help to stick with plans and do something. And also, I have noticed that noon and evening is most productive hours for me. So, I try to utilize those hours to make them as useful as possible.

### Time Management

A step at a time is needed to produce this work. Without proper time management and discipline to follow it, it can bring chaos to project. So, time management is the key to the project. Since there’s also other project of SPPM. So, I have to divide my priorities and manage my time. I have to manage time to learn and implement it on project also. To work on project, I have allocated two to three hours a day. But sometimes it takes less or more depending on the nature of the task I am trying to solve.

College decided to open for lecture for 3 days a week. For those 3 days’ time can be different but usually I work on my project at night and learn during days. These are most productive hours for me. To overcome procrastination, I try to do something everyday because even little work, reading or learning helps to create a forward momentum.

Time management was going well but then world pandemic COVID-19 outbreak occurred and then I have to get back to my home time. And due to limited time for electricity and slow internet connection, I have to manage time according to the time when electricity present and my battery backup work. Since, I have to working on different project it’s hard, but I guess where there’s will there’s way.

### Changes for feedback

While creating a project there are many feedbacks that I got from supervisor, teacher and friends. Since, everyone can give feedback, but I am the filter in this process. I’ve got to pick what is best for my project. Since, I have clear aim and objective but method and ways to achieve them can vary. So, if there are way from which I can improve my project by reducing complexities without compromising quality then I am open to feedback. First, I had plan to only create model but after talking with previous years students, I decided to also build interface for this project. There are also many feedbacks from my supervisor like which quality of paper should I have to read and what should be the possible feature in my interface.

### Changes from experience

Our map of executing our project is based on what we know. But when I try learning new thing, different realities revels themselves. First there were many myths, or my assumption like for client side I should also use python because I was using it for server side, but it was different than I expected. So, from new experience new possibilities revels themselves. And I am ready to implement them if they make sense and improve my systems.

## 3. Communication & Quality Report

### 3.1 Communication with supervisor

My supervisor for this project is Dr. Swati Sah. I had to report her once a week. In that session, we used to discuss about my progress, problems that I faced and what I is my plan for the next week. In those discuss, he would show me path where I was stuck. Since, she was one of the experts in Artificial intelligence, she helped me a lot in finding good quality of research papers and what should be do for next week. It would usually happen on Tuesday after lectures. Someday, I would be working late and sometimes I did miss supervisor meeting at last she provided me with all help. These communication and help made me more optimistic and gave me feeling that I am not alone in this project. But since due to COVID-19 outbreak I am unable to meet her directly, but we are in touch over internet.

### 3.2 Analysis of report

Report is the one of the main parts of project. This main purpose of our project as per university is how to do it because in professional world there are many report and research that we might need to deal. So, report should be well structured and well presented. To make it structured, I’ve used given framework from university with some modification for authenticity.

## 4. Reflection

### 4.1 Initialization of the project

At the beginning of the project, I had only theoretical knowledge. Our 6th semester started on February. And my first step was to learn about how I can implement it practically. First, thing I need was to system setup. For this I have installed anaconda to run Jupyter Notebook. And as I began to explore new packages and installed them with the help of package managers. And for server side I install flask and for client side I installed npm and node for angular. So, I knew what technology that I need to use. After setting up my environment, it was easy for me to learn about technologies.

### 4.2 Analysis of level of challenge

There are certain challenges while working on this project. For example, learning flask, clustering, revising python learning about algorithms and mathematics for machine learning. Many things were for me and there are certain levels of challenges that I needed to tackle to do my project. First, I start with revising python and learning basic about mathematics. And after lockdown due to COVID-19 new challenge arise and it affected other aspect of life also like less meeting with supervisor. Since, I have to return my hometown, I had to face electricity problem and slow internet access also.

### 4.3 Analysis of Required Skill

As with nature of challenges there were also many required skills, I need to tackle them. First skill that I need to learn was python with more depth because in previous I’ve only used python for basic thing but for flask and machine learning, I had to deal with packages and new things. Also due to discontinuation I’ve forgot many syntaxes of python. And then since our course has loosely touched mathematical and algorithmic aspect of programming, so I had to learn them for learning machine learning as well as implementing them in project. I started with linear algebra, then probability and statistics. With their knowledge, it helps me to understand many aspects of clustering. And then, many topics of document clustering were also complex in itself and I’ve came across many term and related technologies. And also, flask is something that I’ve never working so, with right tutorial and resource, it was very helpful. And for machine learning and clustering, I’ve enrolled to many online courses and many of them help but most of them were about machine learning and only few things about document clustering. Online articles were very helpful and as well as documentation were also helpful.

### 4.4 Progress

First, I had less idea about how thing will turn. For example, my initial or plan for the project is given below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Task Name** | **Start Date** | **End Date** | **Duration** | **Status** |
| **Project** | 3-Feb-20 | 10-May-20 | 97 | incomplete |
| Development | 3-Feb-20 | 10-May-20 | 97 | incomplete |
| System setup | 3-Feb-20 | 10-Feb-20 | 7 | incomplete |
| Resource gathering | 8-Feb-20 | 15-Feb-20 | 7 | incomplete |
| create model and learn | 15-Feb-20 | 4-Mar-20 | 18 | incomplete |
| Test And validate | 4-Mar-20 | 22-Mar-20 | 18 | incomplete |
| Deploy in backend | 26-Mar-20 | 9-Apr-20 | 14 | incomplete |
| Create web interface | 9-Apr-20 | 17-Apr-20 | 8 | incomplete |
| Integrate backend and frontend | 17-Apr-20 | 24-Apr-20 | 7 | incomplete |
| Testing | 24-Apr-20 | 30-Apr-20 | 6 | incomplete |
| Evaluation | 30-Apr-20 | 7-May-20 | 7 | incomplete |

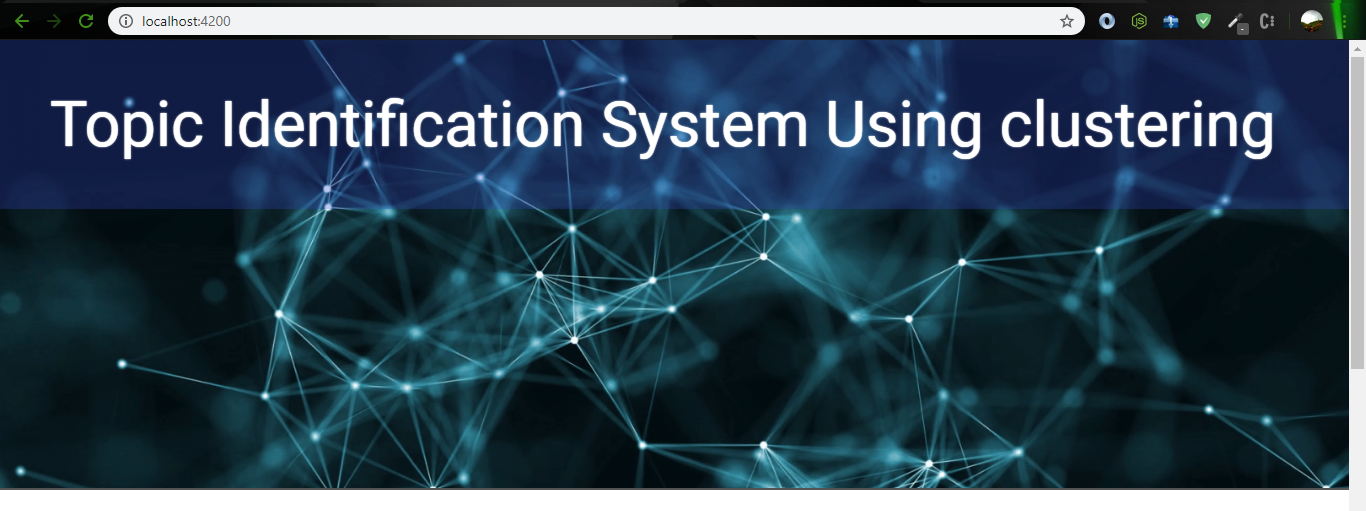
There were many things that I underestimated and something that I overestimated. For example, I did take them in linear way, for example training model before backend and creating interface and there’s no time for learning. And after some clarity I found out that I can do them parallelly like setting up environment and routes for them. There are many things which are done like setup, resource gathering, deploying backend and creating fronted. But there are some issues with creating modal. And I found out that Testing and validate doesn’t fit here because it’s unsupervised method of machine learning.

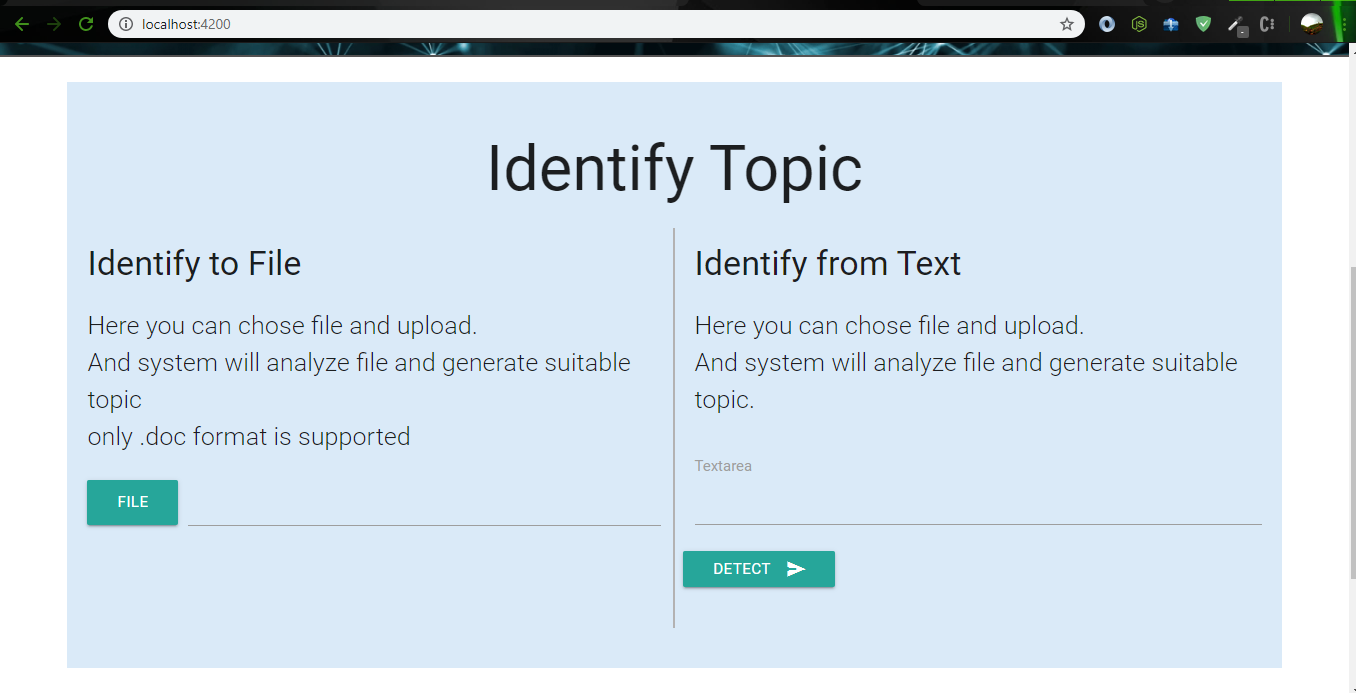
So, I’ve made new project plan table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Task Name** | **Start Date** | **End Date** | **Duration** | **Status** |
| **Project** | 3-Feb-20 | 24-May-20 | 111 | In progress |
| Development | 3-Feb-20 | 24-May-20 | 111 | In progress |
| System setup | 3-Feb-20 | 10-Feb-20 | 7 | complete |
| Resource gathering | 8-Feb-20 | 15-Feb-20 | 7 | complete |
| Extracting data | 15-feb-20 | 25-feb-20 | 10 | complete |
| Preprocessing data | 25-feb-20 | 9-march-20 | 13 | complete |
| Vector space matrix | 16-march-20 | 23-march-20 | 7 | complete |
| Deploy in backend | 15-feb-20 | 28-Apr-20 | 73 | In progress |
| Create web interface | 15-feb-20 | 28-Apr-20 | 73 | In progress |
| Applying clustering to text documents | 4-Apr-20 | 28-Apr-20 | 24 | In progress |
| Integrate backend and frontend | 28-Apr-20 | 3-May-20 | 5 | incomplete |
| Testing | 3-May-20 | 10-May-20 | 7 | incomplete |
| Evaluation | 10-May-20 | 15-May-20 | 5 | incomplete |

From above we can see that many things are different from what I originally planned. And there are also many things that I’ve changes. And second one is more detailed.

Some images of my work: \_





### 4.4 Future Plan

Some from above table we can clearly see that there is so much work to do. Some of them are in progress like creating server and client sider interface because they depend on clustering. And after clustering, I’ll implement them on server side and from client-side client will access and analyze text to find topics of the document. Then my next plan will be to test and evaluation my system.

### 4.4 Room for improvement

There are also room for improvement like on client side, I can also provide other facilities for clients like good interface and give user to upload file or paste documents.

## 5. Detailed thesis contents

### 5.1 Introduction

In this section contained the introduction of the project. Here thing such as background, purpose of study, Aim, objective and significance of the project were detailly explained. In this section main thing about project such as our aim and objective were discussed because it the first and main part of the thesis. If there are fuzzy concept about aim and objective, then our whole research will be fuzzy and even right action can lead to wrong outcome.

### 5.2 Literature review

Since, research is all about reviewing that search which are already done and finding the connection between them. So, this is one of the most important part of research. In this part our assumptions are testing, we have analyzed other’s work critically get down to conclusion without any biasness. And this helps to understand our system better and also helps us to find what the status of our related subject.

### 5.2.1 Technical Background of Topic identification using clustering

This provides the all the technical aspect of the project. In this section. I talked about artificial intelligence, clustering and its techniques, classification and about the document clustering. These terms are the basis thing about my research and also can be consider as the technical background of the system.

### 5.2.2 Relevant Work done in Topic identification using clustering

This part contain all the work which were done in past by other research. This part helped me get idea about how can I implement and what are the challenges that I can face while implementing in into the system.

### 5.3 Data Analysis & market Research

This part of the research deals with the data which were collected from surgery and feedbacks. This help to understand what really people want and what improvement that can done to make product which is need of the real people.

### 5.4 Project Plan

In this section I’ve planned on my assumption and knowledge. This part contained table and gant chart where I’ve created data and days for my project.

### 5.5 Planning Artifacts

This part of the thesis contains the designing plan of my artefacts such as it’s requirements, system design which includes use case diagram, activity diagram, sequence diagram, Low-fi, Mid-fi and testing Evaluation Strategy.

## 6. Conclusion

It is great learning experience and there are many things that I’ve learned form this. First, it sounded like impossible but after from hard work, help, self-motivation and constant working, it is finally become more achievable and amazing learning experience. At last it boosted me with new confidence and self-empowerment.