ASSIGNMENT 2

GROUP 12

Members

Aagam Jain	20BCM001
Atharva Jagdale	20BCM007
Gautam Suryavanshi	20BCM023
Mudit Maheshwari	20BCM051
Shriya Ballal	20BCM076

OCTOBER 25, 2022

SPMQA

CSI0704

SRS

By,

Mudit Maheshwari

20BCM051

Problem Statement

Summary of lecture and Recommending Resources based on lecture.

SRS DOCUMENT

INTRODUCTION

1.1 Purpose

A summary is a short overview of the main points of a text. The purpose of providing a summary of lecture is to quickly give the reader or listener an idea of what this material is saying. You may find it helpful to create summaries of your own work, but more often, you will create summaries of material by other authors, such as articles, plays, films, lectures, stories, or presentations.

When you ask yourself, after reading an article (and maybe even reading it two or three times), "What was that article about?" and you end up jotting down–from memory, without returning to the original article to use its language or phrases–three things that stood out as the author's main points, you are summarizing.

Recommendations are arguably the most important part of the analysis phase—this is where you'll suggest specific interventions or strategies to address the issues and constraints identified in the assessment. Recommendations should directly respond to key findings arrived at through data collection and analysis.

1.2 Intended Audience and Use

A novelist has just finished their masterpiece (Video or Audio), but they want to get the reader's attention to read it thoroughly. They create a summary of that to establish the main storyline while leaving out specific details to appetize the reader to explore further.

A student is asked to describe what happened in a Lecture. They listened over the source material and decide that a summary is a concise way to present the information. They give the main ideas of the document to present the information to the class in an easy-to-understand manner.

A lesson writer has just completed their research on how to write a summary. They have written a detailed lesson on the process, but need to make sure that students understand the important points. They decide to include a bulleted list of the important takeaways from the lesson to concisely explain the principal information.

A student or a writer, who watched a video and read the summary, now want to read more about that particular topic or want more material for the reference, need not to search differently for the books as recommendation are already being provided there.

People with hearing loss and deafness can still read and understand the summary of the lecture provided to them and can explore more through provided recommendations

1.3 Scope

Summarizing a lecture, or distilling its essential concepts into a paragraph or two, is a useful study tool as well as good writing practice. A summary has two aims:

(1) to reproduce the overarching ideas in a text, identifying the general concepts that run through the entire piece, and

(2) to express these overarching ideas using precise, specific language.

When you summarize, you cannot rely on the language the author has used to develop his or her points, and you must find a way to give an overview of these points without your own sentences becoming too general. You must also make decisions about which concepts to leave in and which to omit, taking into consideration your purposes in summarizing and also your view of what is important in this lecture.

Recommendation systems are widely used in several different domains for the recommendation of articles, music, movies, and even people. Portals such as Amazon and Submarino use recommendation systems to suggest products to their customers. Meanwhile, social networks such as LinkedIn and Facebook use them to suggest new contacts.

To accomplish that, the most used techniques employed in recommendation systems are The collaborative filtering and content-based systems. The collaborative filtering does not take into account the type of items, nor their attributes. It takes exclusively into account the expressed opinion about the other items in order to make recommendations. Meanwhile, content-based filtering uses the knowledge it has of the items and their attributes to make recommendations.

These techniques perform well, but they employ cluster solutions to solve the scalability problem and be able to process high chunks of data.

Overall Description

2.1 User Needs

The summary is a great technique to prepare a specific topic, the ideal is to make the summary for each different point that a topic has and then one together.

Some of the needs for preparing the summary are:

1. **To have better synthesis capacity**

You can summarize with your own words and synthesize a larger text. In this way you can differentiate the fundamental from the secondary, something essential to be able to study only the most important.

2. **Will not have to reread the entire topic**

Something that is very important when studying a topic is getting involved in learning so that you don't have to reread the whole topic over and over again. For this, it is necessary to understand it and synthesize the information in the form of diagrams and summaries.

3. **Will learn the concepts better**

Once the steps are done and you make a summary with your words, you will realize what you know and what you have to review. In addition, the concepts that you already know will have been more effectively fixed in your mind.

4. **Will know how to better structure the contents**

Another fundamental aspect when making the summaries is that you will learn to structure the contents with a coherent and logical order, something fundamental for learning and for your mind to better integrate the knowledge

5. **Will have greater fluency**

When a summary is made, being written with your own words will improve your vocabulary and therefore you will feel much more confident when it comes to explaining what you have learned with your words, both orally and in writing

A recommendation system uses the process of information filtering to predict the products a user will like and accordingly rate the products based on users' preferences. A recommender system easily highlights the most relevant products to the users and ensures faster conversion.

Some Practical Use Cases of Recommendation Systems

1. **eCommerce Recommendations**

   eCommerce is by far the commonest and most frequently encountered use case of recommendation systems in action.
   Amazon was a pioneer in introducing this change back in 2012 by making use of item-item collaborative filtering to recommend products to the buyers. The result? A resounding 29% uplift in sales in comparison to the performance in the previous quarter! Soon enough, the recommendation engine contributed to 35% of purchases made on the platform, which was bound to impact the bottom line of the eCommerce giant.

2. **Media Recommendations**

   Most media streaming service providers employ a relational understanding of the type of content consumed by the user to suggest fresh content accordingly. Additionally, the self-learning and self-training aspect of AI in recommendation engines improves relevancy to maintain high levels of engagement while preventing customer churn.

3. **Video Games and Store**

   Gaming platforms like Steam, Xbox Games Store, PlayStation Store are already well-known for their excellent recommender engines that suggest games based on the player's gaming history, browsing history, and purchase history. As such, someone who has an interest in battle royale games like Fortnite will recommend games like PUBG, Apex Legends, and CoD rather than MMORPGs like WoW.

4. **Location-Based Recommendations**

   Gaming platforms like Steam, Xbox Games Store, PlayStation Store are already well-known for their excellent recommender engines that suggest games based on the player's gaming history, browsing history, and purchase history. As such, someone who has an interest in battle royale games like Fortnite will recommend games like PUBG, Apex Legends, and CoD rather than MMORPGs like WoW.

2.2 Models Used

- Convert Audio File into Text: Whisper OpenAI
- Summary of an Audio File: BART Model
- Keyword Extraction from Transcribe Text: Rake Model
- Comparing the keywords that we fetch from transcribe text

- With combined column of youtube dataset: Sentence Transformer
- Finding Similarity between sourceText and CombineText: Cosine Similarity

2.3 Assumptions and Dependencies

Assumptions:

- The lectures whose summary is to be extracted from the video/Audio lecture is assumed to be in English Language only.
- The lecture is assumed to be in Audio Format for extraction of the summary. Text lectures are not included for producing the summary as of now

Dependencies:

- We are fully dependent on the dataset we are having with us, as after finding summary from the audio lecture we find the keywords and produce the recommendations. Recommendations include the book and video related to specific keywords. So we must have the huge dataset of keywords and their related books and the videos.

2.4 Conclusion and Future Works

We made a book recommendation method, namely probabilistic relational matrix factorization (PRMF). For a specific recommendation task, the proposed approach jointly learns keywords from the summary and the users searches, to improve the recommendation accuracy. Empirical results on real datasets demonstrate the effectiveness of PRMF, in comparison with strong baseline algorithms. The future work will focus on the following directions. First, we would like to develop more efficient optimization algorithms for PRMF using the parallel optimization framework. Second, we are also interested in extending PRMF to solve Top-N item recommendation problems with user's implicit feedback. The potential directions are adopting logistic matrix factorization or ranking metrics optimization methods to model user's implicit feedback.

System Features and Requirements

3.1 Functional Requirements

- **Use Case :** Speech to Text of lectures

Primary Actor : User

Pre-Condition : Audio in mp3 format

Main Scenario :

1) User is required to upload the audio file in the drop section
2) The audio file gets converted into text

Alternate Scenario :

1) The file uploaded is not in mp3 format
2) Network issue.

- **Use Case :** Summary of lectures

Primary Actor : User

Pre-Condition : Audio in mp3 format

Main Scenario :

1)User is required to upload the audio file in the drop section

2)The audio file gets converted into a Useful Summary along with keywords.

Alternate Scenario :

3) The file uploaded is not in mp3 format
4) Network issue.

- **Use Case :** Recommendation based on Lecture

Primary Actor : User

Pre-Condition : Audio in mp3 format

Main Scenario :

1)User is required to upload the audio file in the drop section

2)The audio file gets converted into Keywords and based on that YouTube Videos and Books are recommended from the dataset

Alternate Scenario :

5) The file uploaded is not in mp3 format
6) Network issue.

3.2 Non-functional Requirements

1) System with a browser and internet connectivity
2)Atleast 4gb RAM for running the application

3) Microphone access for Speech to text conversion functionality

3.3 System Issues

- **Cold Start**

  The issue with Cold Start is more prevalent in recommendation systems. The problem occurs at the start of the system. A recommender system usually compares the user profile with some reference characteristics. These characteristics can be based on information or on the user's social environment. In content-based approach, the system should be able to match the characteristics of an item to relevant characteristics in the user's profile. In order to do that, a model with sufficient detailed information on the user, including his/her tastes and preferences must first be built. Because of the Cold Start issue, items not previously assessed would be ignored in the collaborative filtering approach.

- **GraySheep**

  If a user has rare tastes, the recommendation may not be accurate, as there are no "close neighbours". This problem is called gray sheep

- **Early Rater**

  When a new item emerges, it cannot be recommended to a user before a person assesses it. This issue is clearly identified in collaborative filtering. When a new item with no user assessment or recommendations is inserted, it cannot be recommended. In content-based filtering, knowing the contents of an item is enough to enable a recommendation to a user.

- **Sparse Evaluations**

When there are few users and many items, the evaluations may become sparse and it becomes difficult to find similar users

- **Super-Specializations**

  Only items that are similar to those previously evaluated by the user will be recommended. Exploring new item categories is not possible

- **Serendipity**

  This is related to the lack of surprise in the recommendation. Products that are not related to the user's profile may never be recommended

- **Scalability**

  When the quantity of users, items and evaluations is too large, the system that executes real-time calculations of the relations among users may provide a very long response time and may need computer resources that are not available. This is a common problem in both approaches.

3.4. Future Extensions:

a. The software is limited to recommendations for lectures regarding Technical department of Computer Science Subject. In future we can work on dataset to include various other subjects of other domain.

RISK ANALYSIS

By,

Shriya Ballal

20BCM076

The risks that can be faced during or after project completion are:

1. **Technological Risks**
   These are problems/issues that could be faced on the technological front of the project. We have identified the types of technological risks that are more probable in our project. These are:
   - Component Compatibility
     For our software, we have used a lot of models and the whole system runs on compatibility. Therefore every component of the software should be compatible with the user system for proper working of the same.
   - Obsolescence
     Both the user system as well as our Product should be using the latest versions of the respective software available, as opposed to the older obsolete versions, for optimum working.
   - Design Updates
     The routine updates that our software has, has to be compatible. Also, the UX and UI of the system has to be excellent for maximization of user satisfaction.
   - Data Security
     The user data should be secured and not publicized.
   - Information Privacy
     The information that our systems provide should not be shared/leaked anywhere else.

2. **Resource Risks**
   There are often situations wherein the resources required for the software development or the resources that the said software provides as the output is creating the problem. These risks can be:

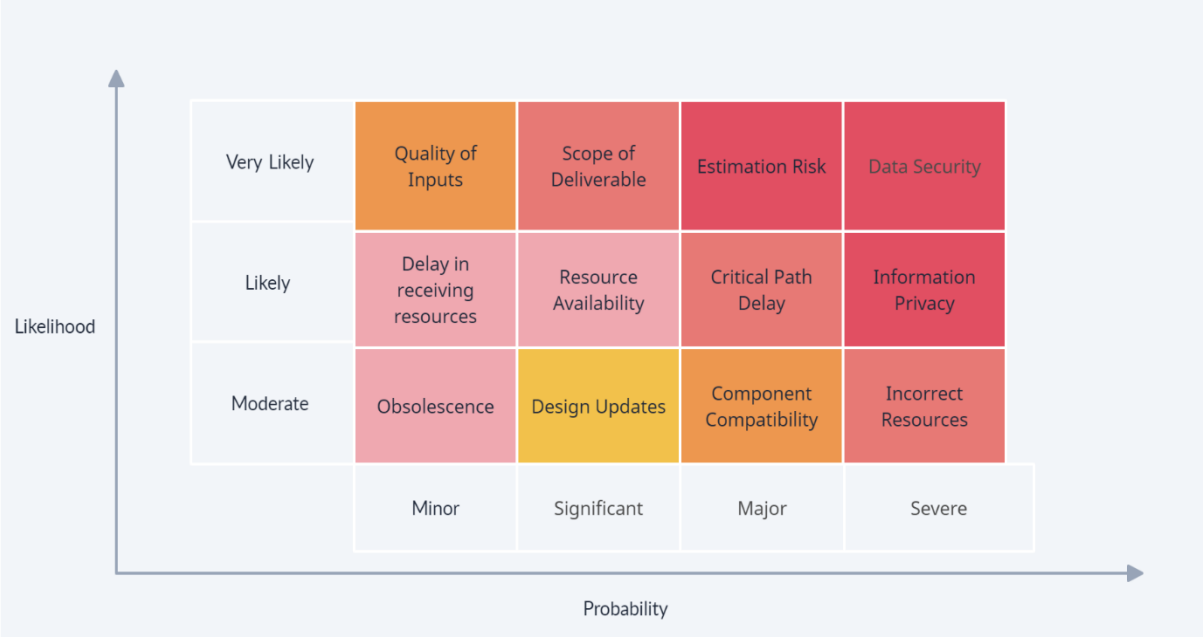
- Resource availability
  The appropriate resources needed for the project must be available at the correct time for the project to go on as scheduled. At the flip side, if the user does not find the resource he/she was looking for, then the whole point of the software would be null.
- Delay in receiving resources
  A delay in resources would setback the project by valuable time which would upset the budget and the schedule of the whole project.
- Incorrect resources
  Receiving incorrect resources would be disastrous to a software project. Continuing with the incorrect resource would waste valuable time, money and effort. In any case, this would lead to an upset schedule and budget which would be irreversible.

3. **<u>Schedule Risk</u>**
   The wrong schedule affects the project development and delivery. These risks are mainly indicates to running behind time as a result project development doesn't progress timely and it directly impacts to delivery of project. Finally if schedule risks are not managed properly it gives rise to project failure and at last it affect to organization/company economy very badly. These can be:
   - Estimation risk
     If the estimations based on which the project schedule is prepared doesn't play out the way they were supposed to, the entire project could get jeopardized.
   - Critical path delay
     If tasks along the critical path are delayed, the timeline of the entire project will be impacted.

4. **Deliverable Risks**:
   These are the type of risks that threaten the project completion based on the scope and further development of the same.
   - Quality of Inputs
     The inputs used for project development should be reliable and should cover the entire spectrum. The inputs shouldn't be biased, for an optimum project.
   - Scope of Deliverable
     The scope of the entire project has to be thoroughly understood. The project would be developed incorrectly otherwise.

TASK NAME	PRIORITY
Technology risk	
Component compatability	High
Obsolescence	Very Low
Design updates	Low
Information Privacy	Very High
Data Security	Very High
Resource risk	
Resource availability	Medium
Delay in receiving resource	Medium
Incorrect resources	High
Schedule risk	
Estimation risk	High
Critical path delay	Medium
Deliverable risk	
Quality of inputs	Low
Scope of deliverable	Medium

Assignment-2

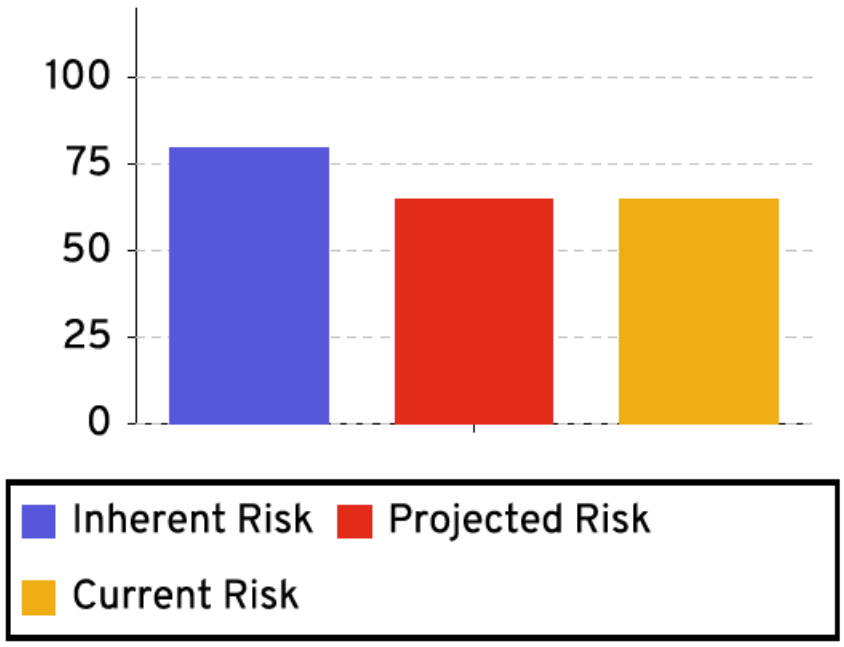


Assignment-2

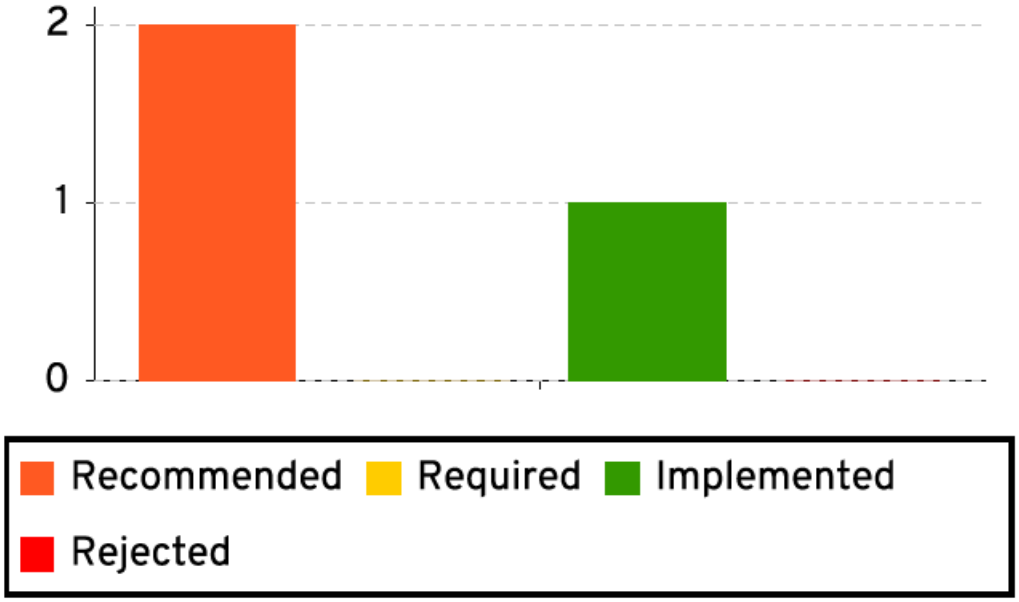
Countermeasures

	Components	Countermeasure	Tags	Priority
☐	General Product Threats			
☐		Backup		
☐		CyberSec		
☐		Standby resources		

Current Risk Summary



Countermeasures Summary



Appendix A: Countermeasure Details

This appendix shows all of the countermeasures mitigating the threats found in the project.

Component: General Product Threats

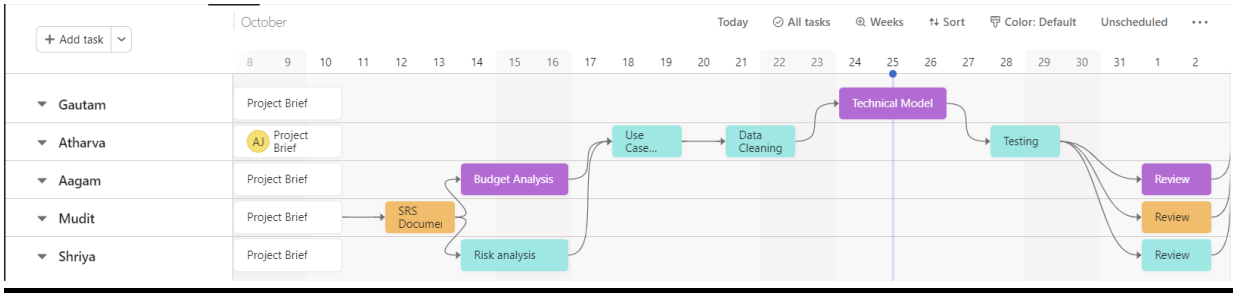
Id	Name	Description	State	Result
backup	Backup	Keep certain number of backups always ready who are somewhat looped with the project	Recommended	Not tested
standby_resources	Standby resources	keep some resources on standby or avail it on lease	Recommended	Not tested
cybersec	CyberSec	Keep a team of cybersec experts to protect the systems	Implemented	Not tested

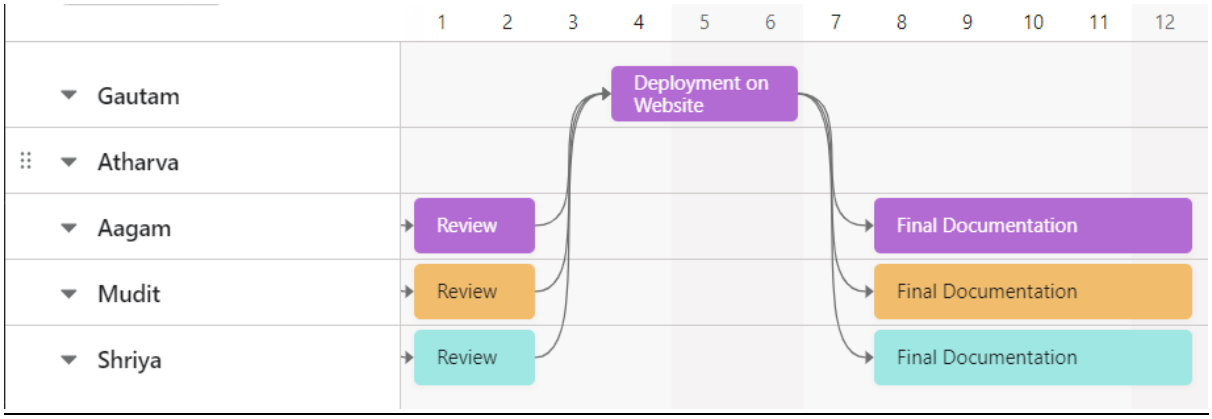
---

Assignment-2

TIMELINE

By Atharva Jagdale 20BCM007





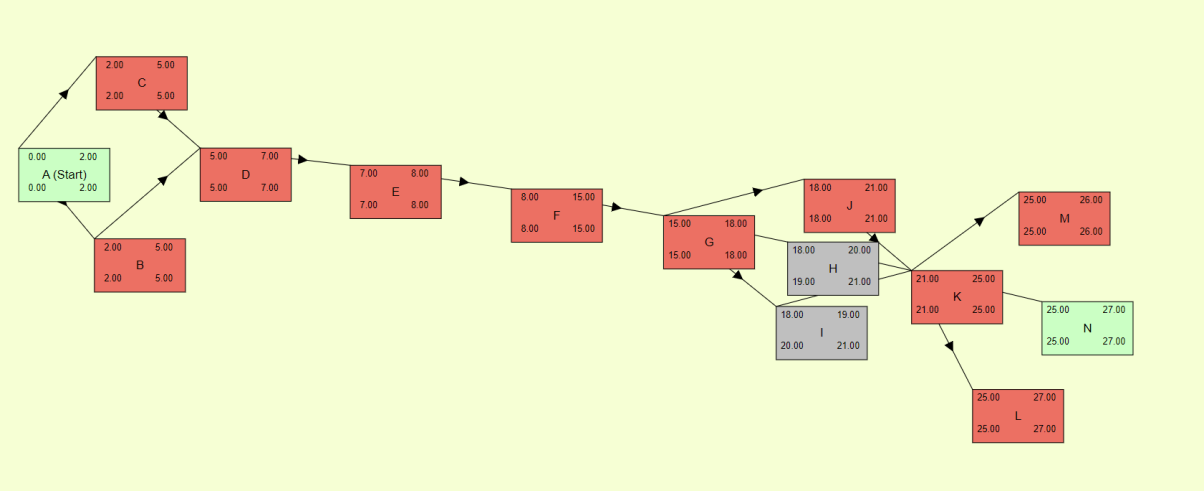
CPM/PERT

- **By Atharva Jagdale 20BCM007**

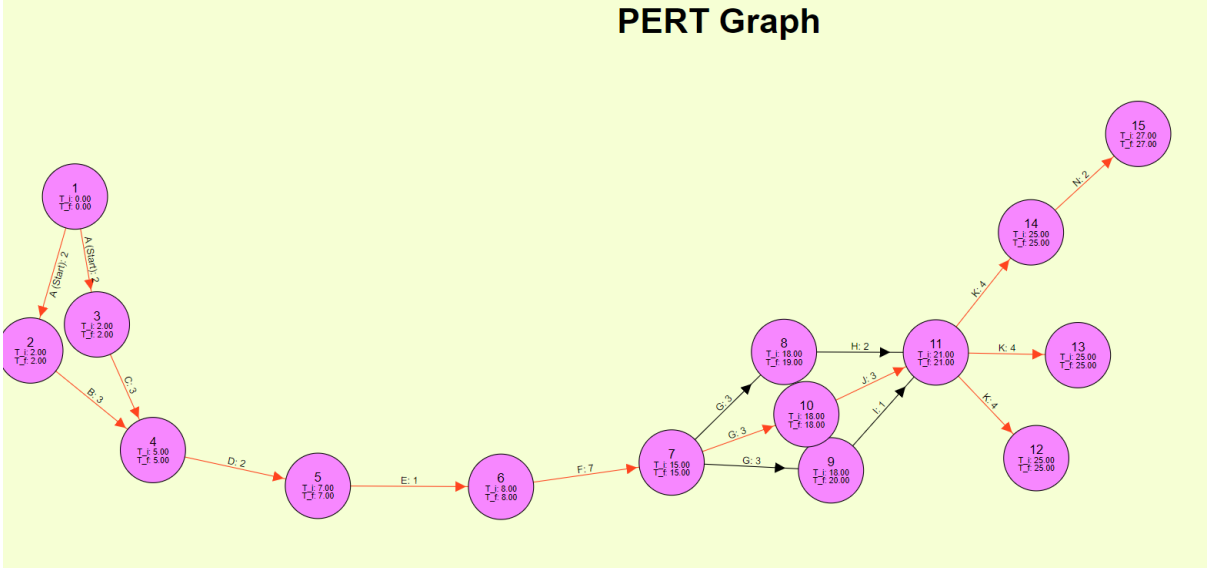
Activity	Predecessor Activities	Duration
A (start)		2
B	A	3
C	A	3
D	B,C	2
E	D	1
F	E	7
G	F	3
H	G	2
I	G	1
J	G	3
K	H,I,J	4
L	K	2
M	K	1
N (end)	K	2

Assignment-2

CPM



PERT



Budget

-By Aagam Jain 20BCM001

Company name: SpeakWorld

Email address: Info@SpeakWorld.com

Overview: We at SpeakWorld try to help you by making a summary of any vocal notes. Our software will take any speech, message, or video as input and then we here at SpeakWorld convert it, produce a document with only important portions, and give out a summary.

We also have the feature of providing recommendations regarding the summary which we will provide through our system. The recommendation system works in a way that it will provide you the link to webpages or YouTube videos by taking keywords from the summary document.

Income: We have various ways to generate income from the software

1. **We will make a website that will generate money from Advertising.**
2. **We will provide the system to schools, colleges, and different Institutions.**

3. **We will have top-up income from the Recommendation feature which we provide if needed.**

Description	Budget	Final Cost
Advertising Money	10$*50hours = 500$	10$*52hours = 520$
Direct System	250$(one time)	240$(one time)
Recommendation System	25$(Subscription)	25$(Subscription)
Total	775$	785$

Profit Or Loss Total: The total amount that we have gained is 785$ on the selling at minimum cost and the total amount that we have to decide at first while budgeting was 775$

Therefore, at last, we have earned 10$ of the cost we have decided.

Expenses: Majorly we have 2 expenses first one being the development of the Technology and the second being the Management and Maintenance team.

Description	Budget	Final Cost
Technology Developer	6 pm 1pm = 250$ Total = 1500$	6 pm 1pm = 250$ Total = 1500$
Management and Maintenance	3 pm (management) 1pm=150$ Total = 450$ Maintenance	3 pm (management) 1pm=150$ Total = 450$ Maintenance

	100$(per year)	100$(per year)
Total	2050$	2050$

Costs overrun: The basically have the work very nice that we have use the budget money accordingly and the final cost match our expected cost.

Analysis: We have used the Three-Point Estimate method, this method is basically using three different calculations to estimate a budget: the most expensive scenario, the most cost-effective scenario and the scenario that's most likely to occur. By using these figures, you can roughly estimate there required budget, but also assess some of the risks involved in the project.

Also, to analyze the we have provided the value in the way we want we have used Earned value analysis, this tells that the accuracy of a project's budget during the execution of the project. It consists of constantly comparing the costs for each phase of the project with the ones that were previously estimated, typically by using Three-Point Estimate.

TESTING

- **By Atharva Jagdale 20BCM007**

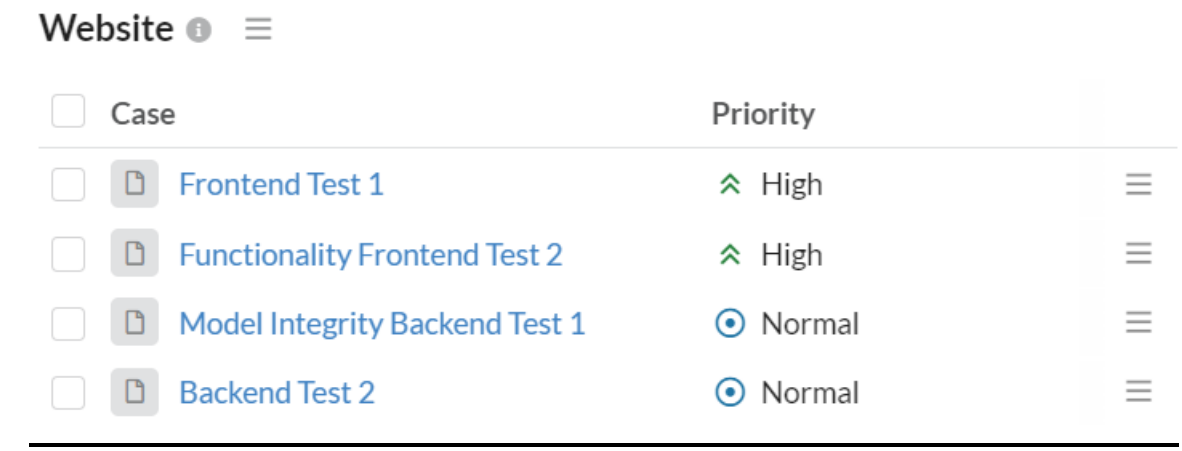
Data

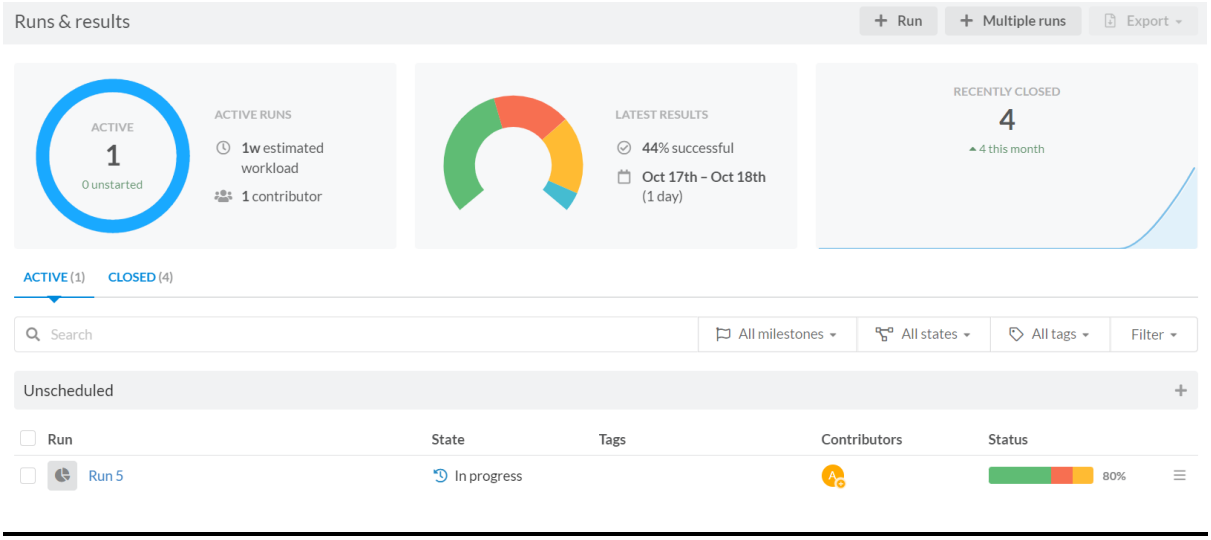
Case	Priority
Data Test 1	Normal
Data Test 2	High
Data Test 3	Low
Data Test 4	High

ML Model

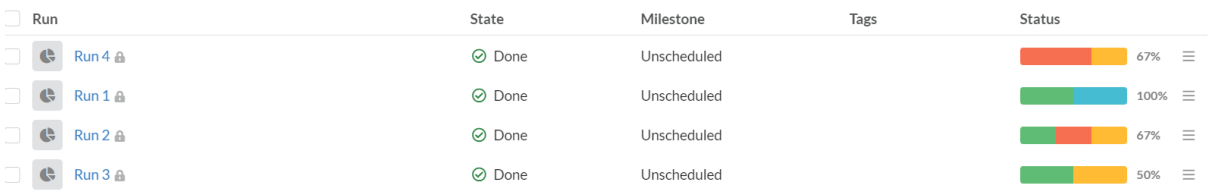
Case	Priority
Bart Model Test 1	High
Rake Model Test 2	Low
Sentence Transformer Model Test 3	High
Cosine Similarity Model Test 4	Low

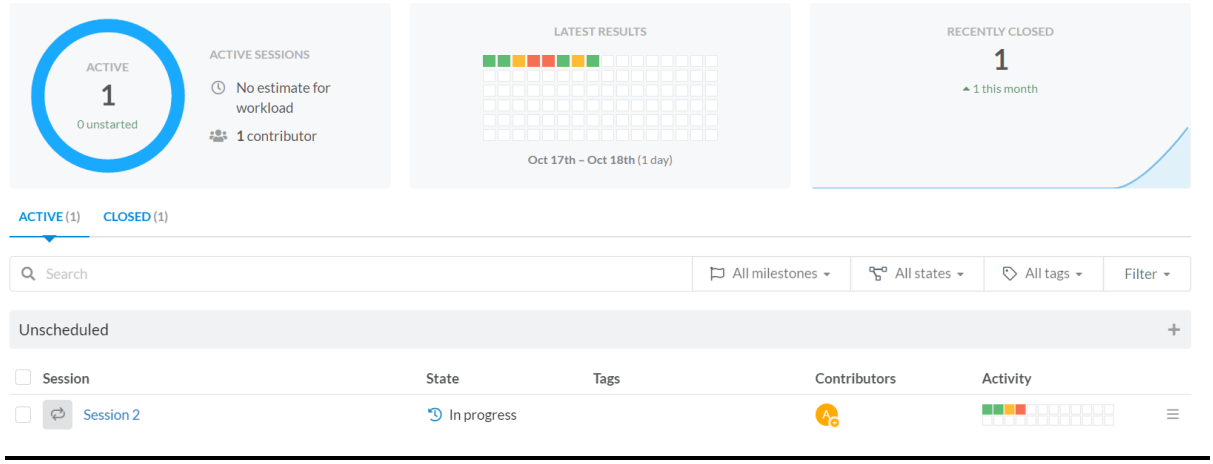
Assignment-2

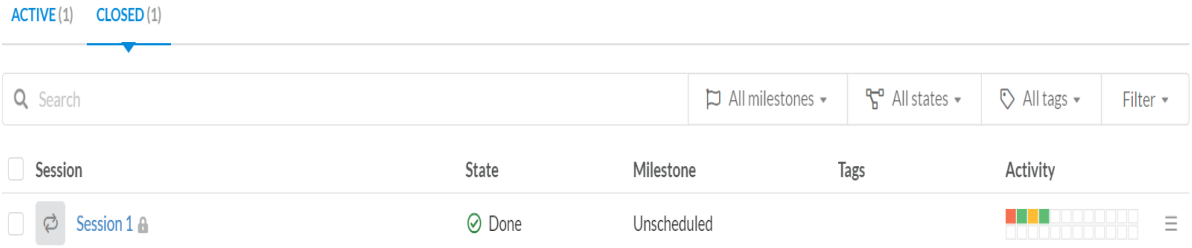




Assignment-2







Assignment-2

ABOUT

Active
Created Oct 17th

Contributors

TIMELINE

10/18/2022

- A Added Session 2
- A Closed Run 4
- A Added Run 5

10/17/2022

- A Closed Session 1
- A Added Session 1
- A Added Run 4
- A Closed Run 1
- A Closed Run 2
- A Closed Run 3
- A Added Run 3
- A Added Run 2
- A Added Run 1

Project Review After Completion By The Management

- **The project was completed within timeframe decided at the beginning**
- **The product post-production and testing met with required Functionalities**
- **The cost of the project was slightly higher than the budget pre decided.**
- **Extra Functionality like Speech to text was introduced in-built to the website smoothly with unexpected new Customer Requirements**
- **The Integrity of the data and the model was maintained throughout.**

- **Stakeholders were satisfied with the overall Project**

- **Resources Handling can be improved in future Projects**

Assignment-2

Effort Estimation

	ENTER ESTIMATION VARIABLES, BELOW
PROJECT MANAGEMENT EFFORT %	30%
PROJECT MANAGEMENT OFFICE EFFORT %	40%
PROJECT CONTROL OFFICE EFFORT %	30%
TOTAL SHOULD BE 100% ---------->	100%
WORK DAY LENGTH IN HOURS	8
ESTIMATED START DATE	10-01-2022
COMPUTED ESTIMATED END DATE	10-25-2022
ESTIMATED PROJECT DURATION IN WEEKS	3.4

PHASE	ESTIMATED HOURS	ACTUAL HOURS
Business Requirements	10	11
Functional Specifications	40	35
Detailed Design	7	10
Code and Unit Test	5	8
System Testing	12	14
User Acceptance Testing	4	6
Project Manager	7	5
Project Control Office	8	6
Project Management Office	4	12

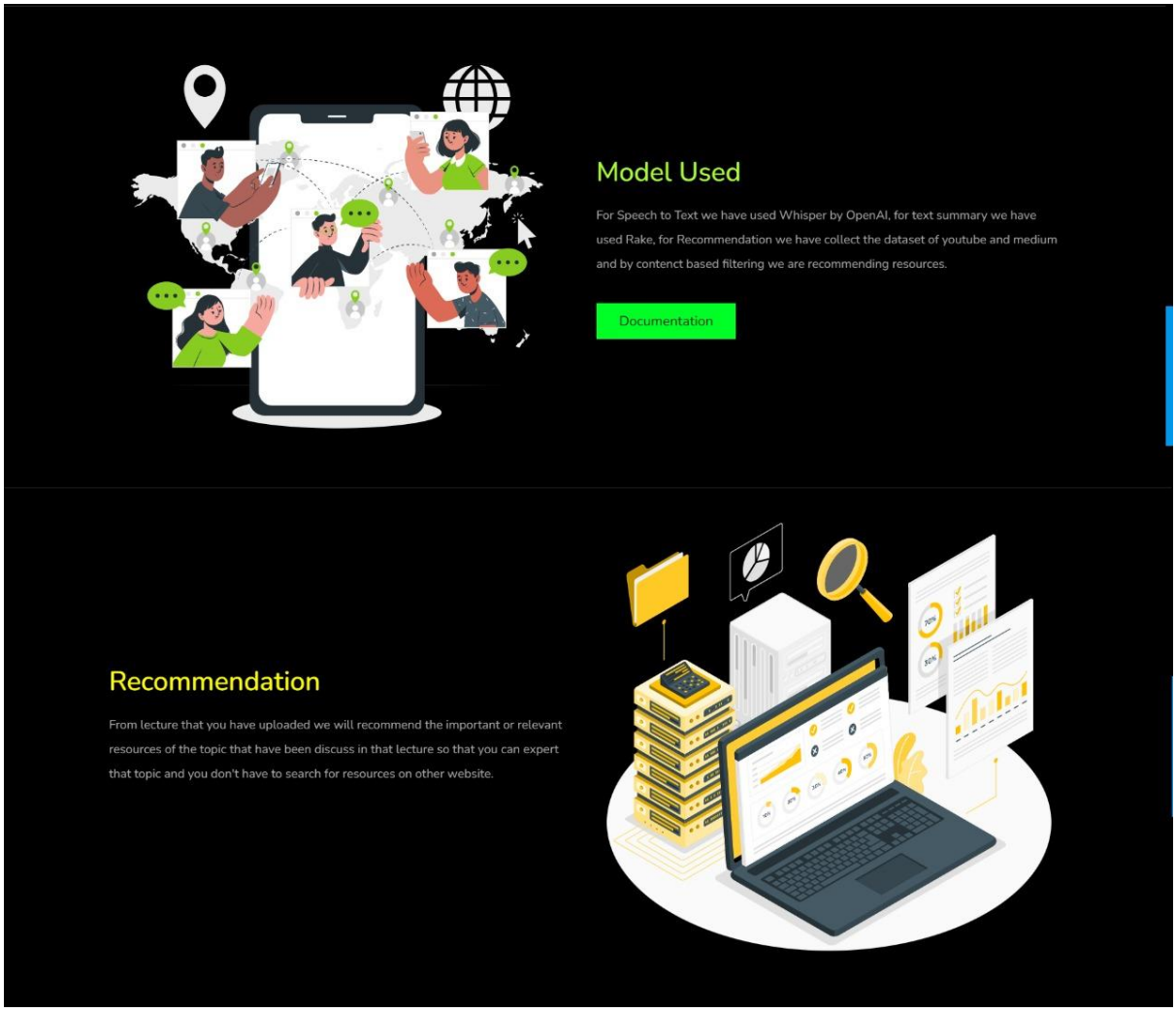
PHASE ACTIVITY	STANDARD WORK EFFORT %	PHASE TEAM SIZE	COMPUTED WORK EFFORT HOURS	COMPUTED TASK DURATION IN WEEKS	COMPUTED AVERAGE RESOURCE HOURLY COST	ESTIMATED COST	COMPUTED TASK DATE OF COMPLETION	COMPUTED WORK EFFORT IN DAYS	COMPUTED TASK DURATION IN DAYS
Business Requirements	15%	2	11	0	₹ 25,750.00	₹ 2,83,250.00	10-01-2022	1.4	0.7
Functional Specifications	10%	4	35	0	₹ 2,500.00	₹ 87,500.00	10-03-2022	4.4	1.1
Detailed Design	30%	3	10	0	₹ 10,791.67	₹ 1,07,916.67	10-04-2022	1.3	0.4
Code and Unit Test	10%	1	8	0	₹ 44,950.00	₹ 3,59,600.00	10-05-2022	1.0	1.0
System Testing	10%	2	14	0	₹ 7,300.00	₹ 1,02,200.00	10-06-2022	1.8	0.9
User Acceptance Testing	25%	2	6	0	₹ -	₹ -	10-07-2022	0.8	0.4
TOTAL SHOULD BE 100% ---------->	100%		84	1		₹ 9,40,466.67		10.5	4.4
Project Manager	30%		5	1	₹ 2,300.00	₹ 11,500.00		0.6	4.4
Project Control Office	40%		6	1	₹ 2,100.00	₹ 12,600.00		0.8	4.4
Project Management Office	30%		12	1	₹ 2,500.00	₹ 30,000.00		1.5	4.4
	PROJECT MANAGEMENT		23			₹ -			
	BASELINE PROJECT HOURS ESTIMATE		107			₹ 9,40,466.67			

ENTER RISK CONTINGENCY	5%	₹ 47,023.33
BASELINE PROJECT COST ESTIMATE		₹ 9,87,490.00
PROJECT MINIMUM ESTIMATE	0%	₹ 9,87,490.00
PROJECT MAXIMUM ESTIMATE	0%	₹ 9,87,490.00

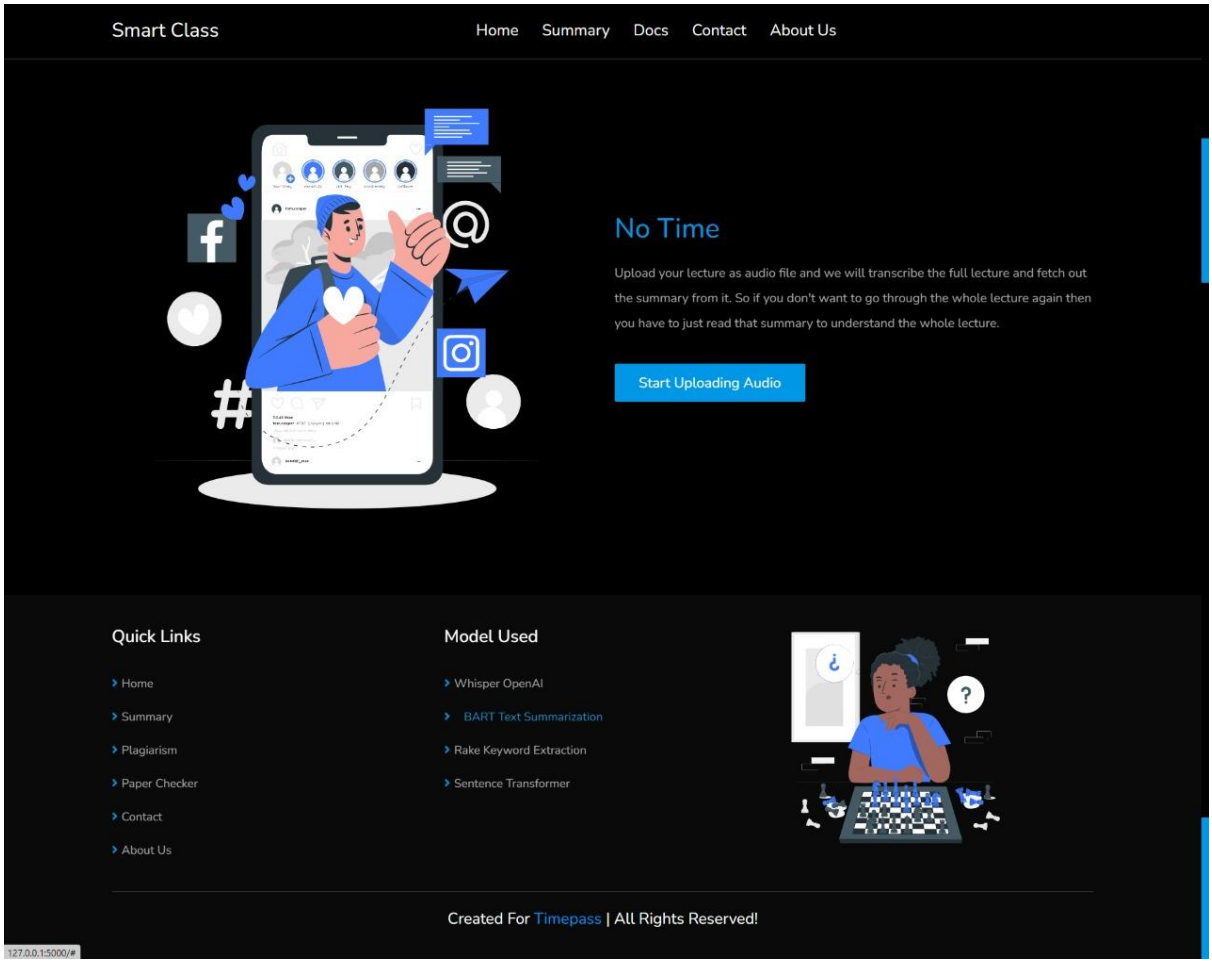
Website

- **By Gautam Suryavanshi 20BCM023**

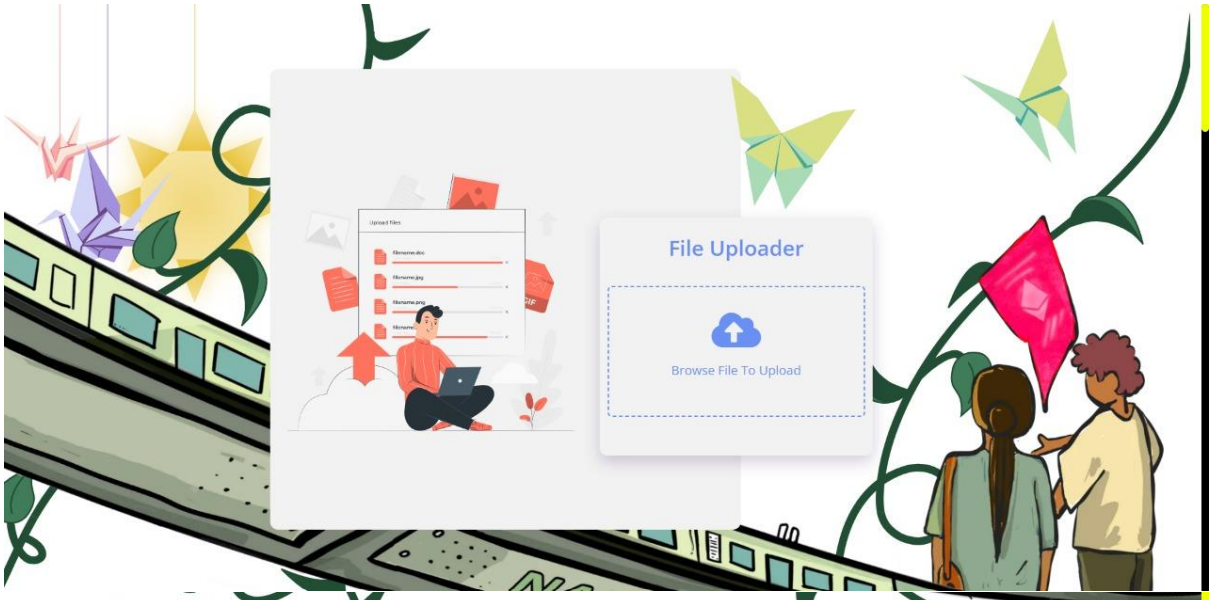
Assignment-2

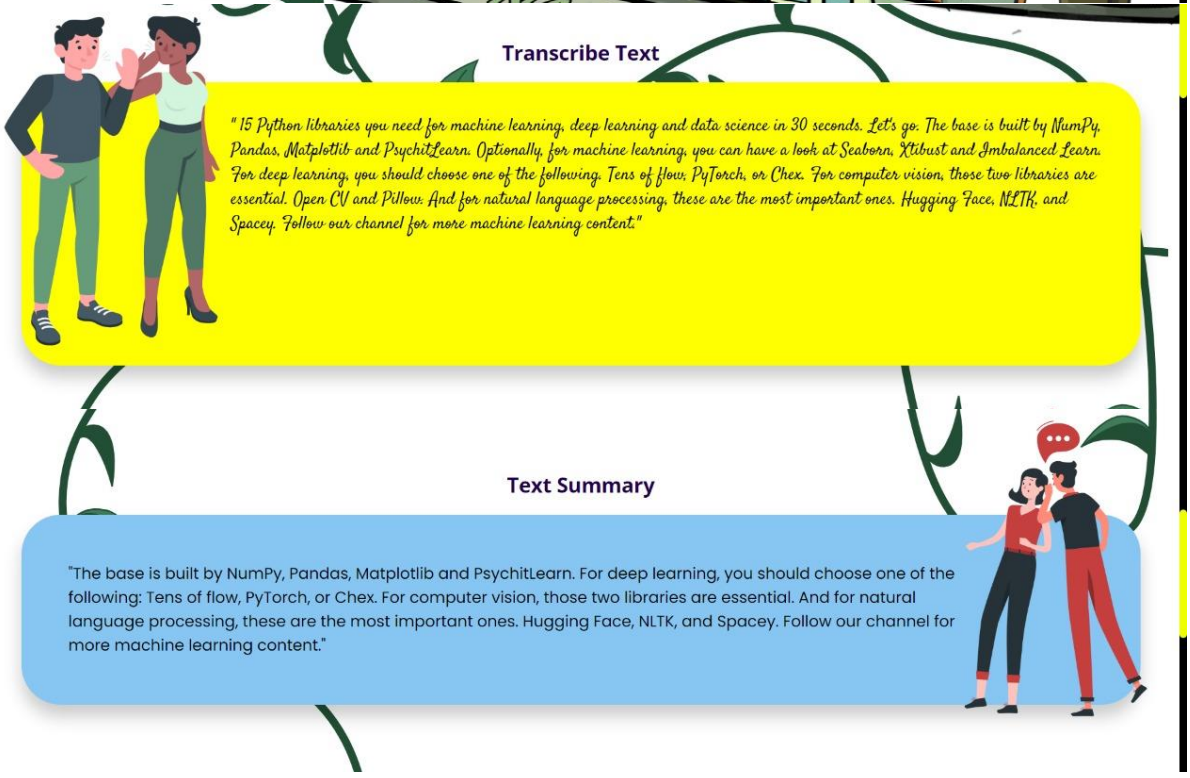


Assignment-2



Implementation





Assignment-2

Recommender

Youtube Recommendation

A Python Project to Solve Your Debt Problems...

Applying Software Engineering Principles To Your Data Science Tasks In Pytho...

Automating Your Data Science Tasks In Python (importing CSV files to databas...

How I Learned Data Science (resources to get a job) in 2021...

Most Common Coding Mistakes on Data Science Interviews...

Book Recommendation

Intelligent Projects Using Python: 9 real-world AI projects leveraging machine learning and deep learning with TensorFlow and Keras...

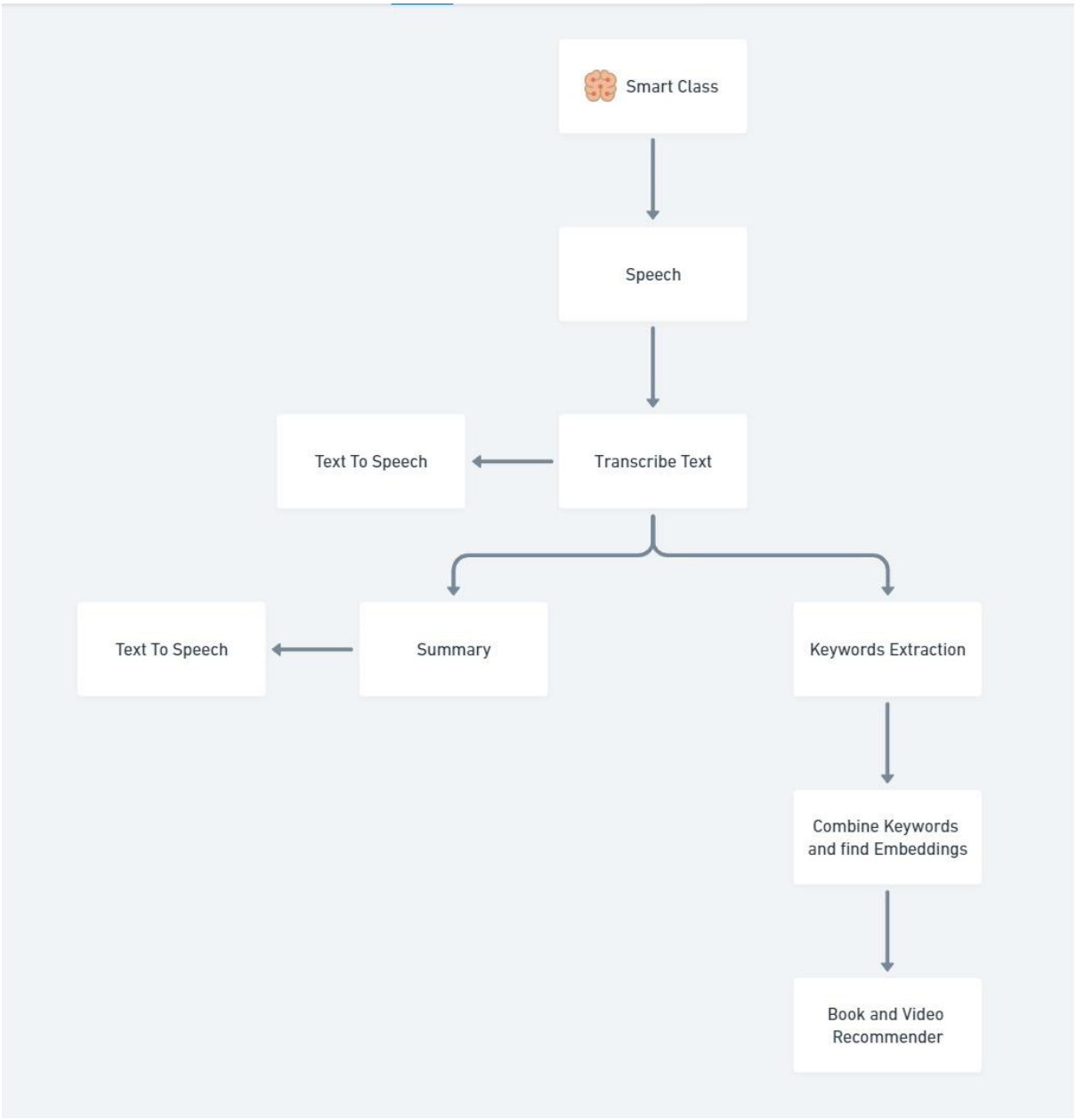
Introduction to Deep Learning...

Grokking Deep Learning...

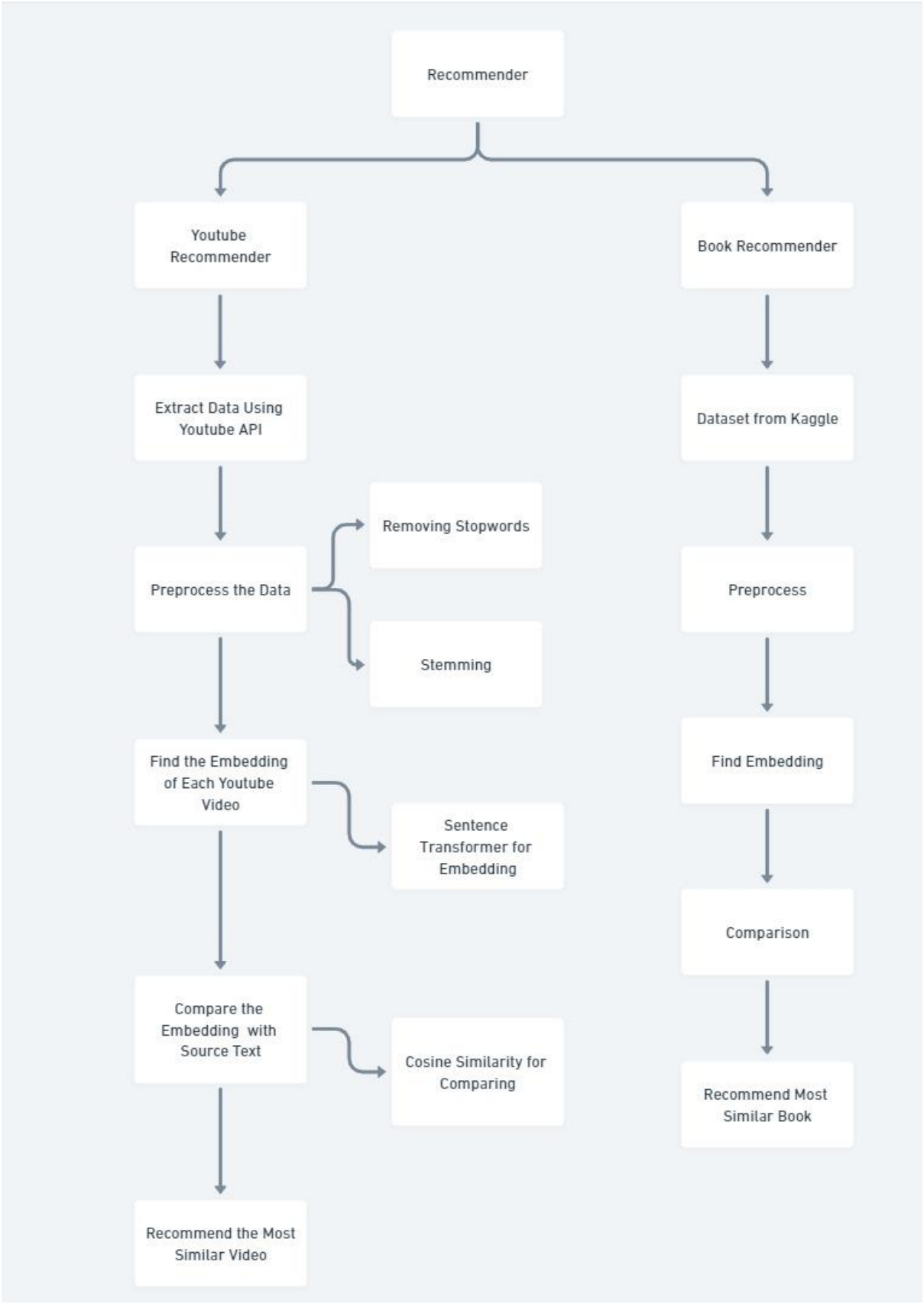
Software Engineering, 10th Edition...

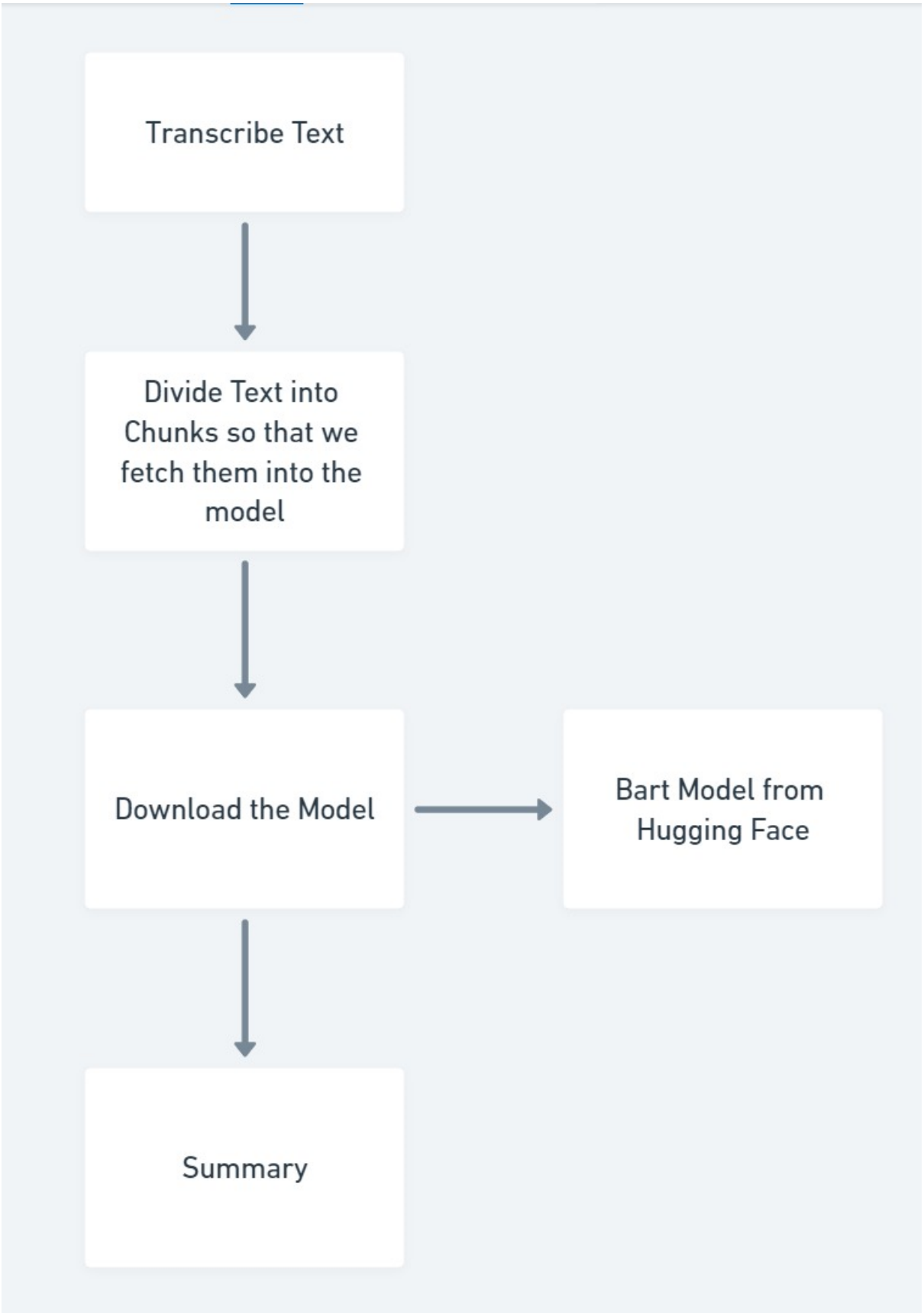
Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning...

Assignment-2



Assignment-2




Transcribe Text
Divide Text into Chunks so that we fetch them into the model
Download the Model
Bart Model from Hugging Face
Summary

Assignment-2