



Department of Computer Science & Engineering

Course Title – Software Engineering

Course Code – CSE 322

Section - A₁

NLP based Laptop Recommendation System

Final Report

Date of Submission – April 17, 2022

Submitted by:

Md. Hasibur Rahman (19101009)
Tanmoy Mazumder (19101013)
Shawan Das (19101020)

Submitted to:

Dr. Shah Murtaza Rashid Al Masud
Associate Professor,
University of Asia Pacific

1. Introduction

The pandemic has forced the world to go on lockdowns multiple times in the past 20 months. Each time students and a large number of office workers had to shift to their tasks from home and this created an unprecedented surge in Computer sales, especially Laptops since laptops are more mobile compared to their desktop counterparts and they come with a pre-built video calling specs.

Problem Statement

As the number of sales of laptops increased, so did the number of complaints from the users that they had been ripped off at the store for not understanding the specifications. This has become more and more common as there are multiple things to be considered while buying a laptop and an average person has very little to no idea of these specifications, let alone which specs would be the perfect choice for their needs.

This problem can be tackled by searching online for reviews, as user reviews reflect a true image of how the product really is. However, this method falls short because of one problem: there are literally thousands of reviews online about hundreds of such devices.

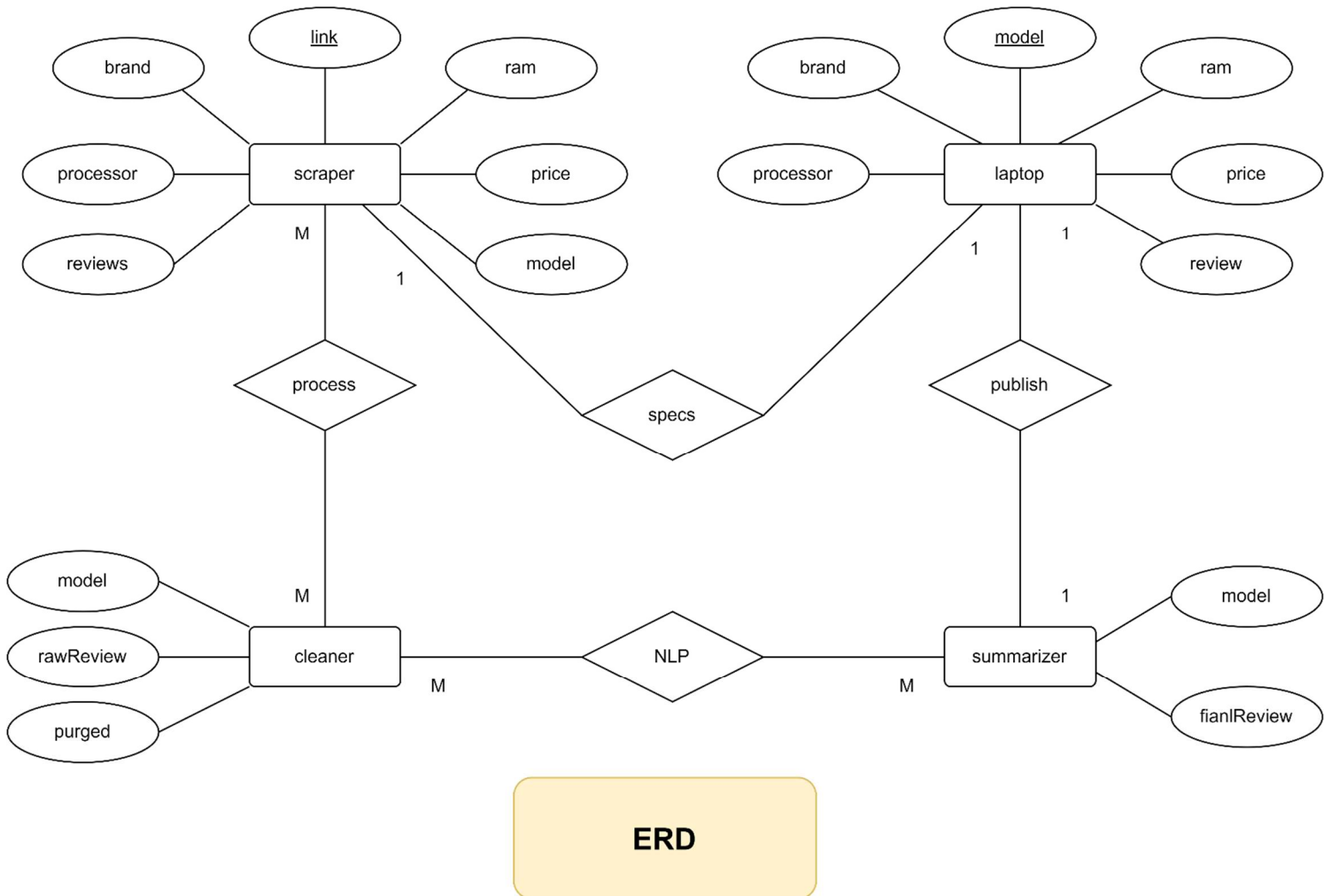
2. Proposal

This project tries to overcome this certain problem using NLP (Natural Language Processing) and some Machine learning algorithms used on data collected through text-mining from online review sites. The core goals of the project:

- Collecting, cleaning and analyzing user review data from the internet
- Building a system with the most suited NLP model(BERT) capable of summarizing the human-written reviews
- Building an arsenal of well-summarized reviews for a user to get the perfect suggestion from.

3. Design

Entity Relationship Diagram



Description:

There are 4 entities: Scraper, Cleaner, Summarizer, Laptop.

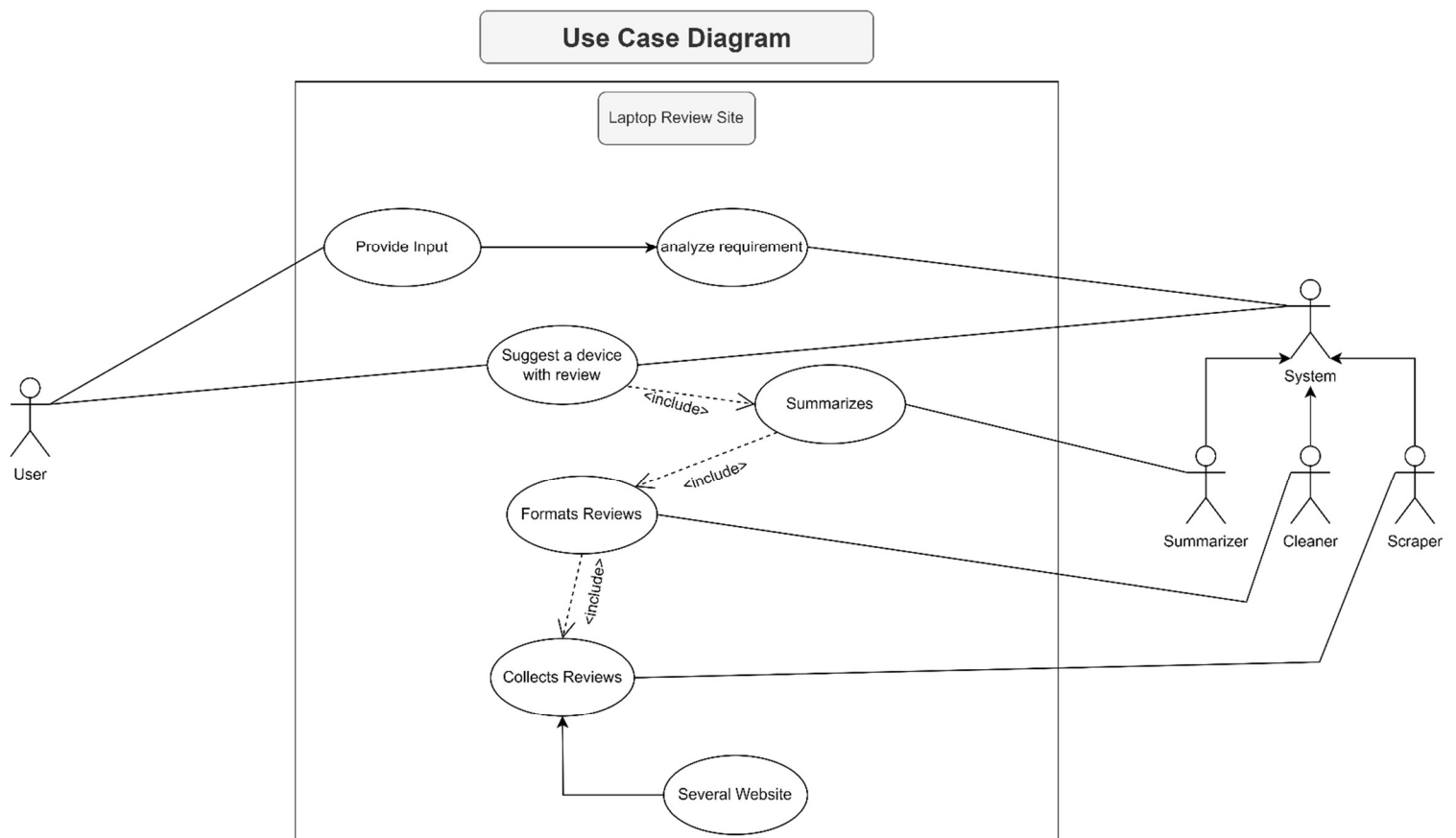
- I. Scraper: The Scraper goes over each of the laptop's review pages and collects information such as the model, reviews, price, brand, ram, and processor.

- II. Cleaner: The cleaner processes the raw reviews collected by the scraper, purges the collected data of any unwanted portions that are hindrances to the NLP model's functionality, and creates clean and usable versions of each review.
- III. Summarizer: The summarizer comprises of the NLP model and the final review for the given laptop.
- IV. Laptop: This is the general entity for each of the laptops in the database. It holds the basic info from the scraper and the final review from the summarizer.

ERD Relations:

Entity A	Entity B	Relation	Type
Scraper	Cleaner	Process	Many to many
Cleaner	Summarizer	NLP model	Many to many
Summarizer	Laptop	Publish	One to one
Laptop	Scraper	Specs	One to one

Use Case Diagram



Description:

There are 2 actors: The User and The System(admin). They interact through the website.

The user provides input about what type of laptop they are looking for and then waits for the output. The input would then be analyzed by the system and it would provide suggestion by matching the best fit available from the summarized reviews.

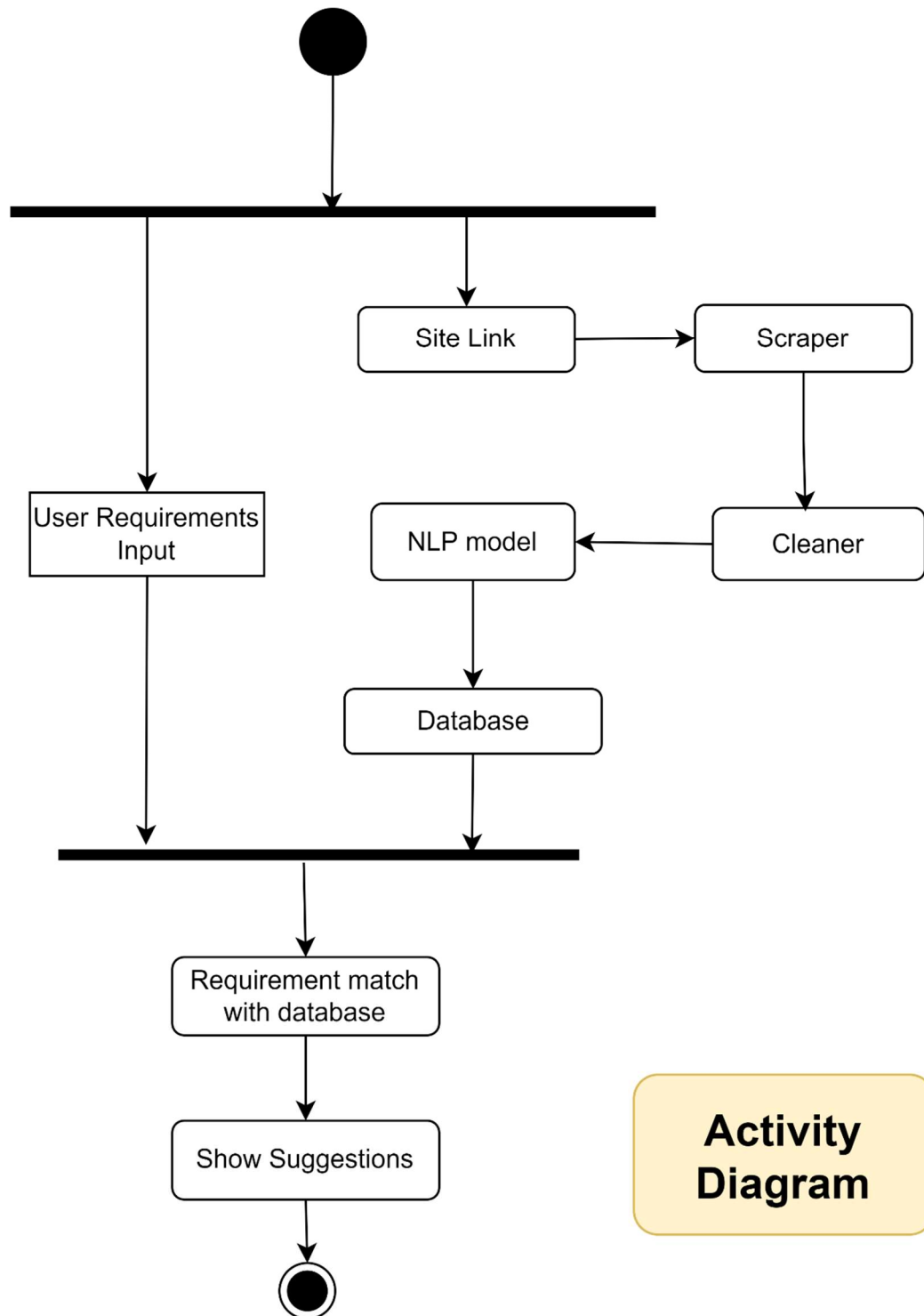
These summarized reviews are the products of the subsystem Summarizer. The summarizer uses the multiple cleaned reviews provided to it by the Cleaner subsystem.

The Cleaner subsystem is responsible for creating clean-formatted reviews from the raw reviews collected for each laptop.

The raw reviews are collected by the Scraper subsystem. The Scraper runs on the multiple links of review sites.

Following this process chain, the output suggestion would be created and provided to the user.

Activity Diagram



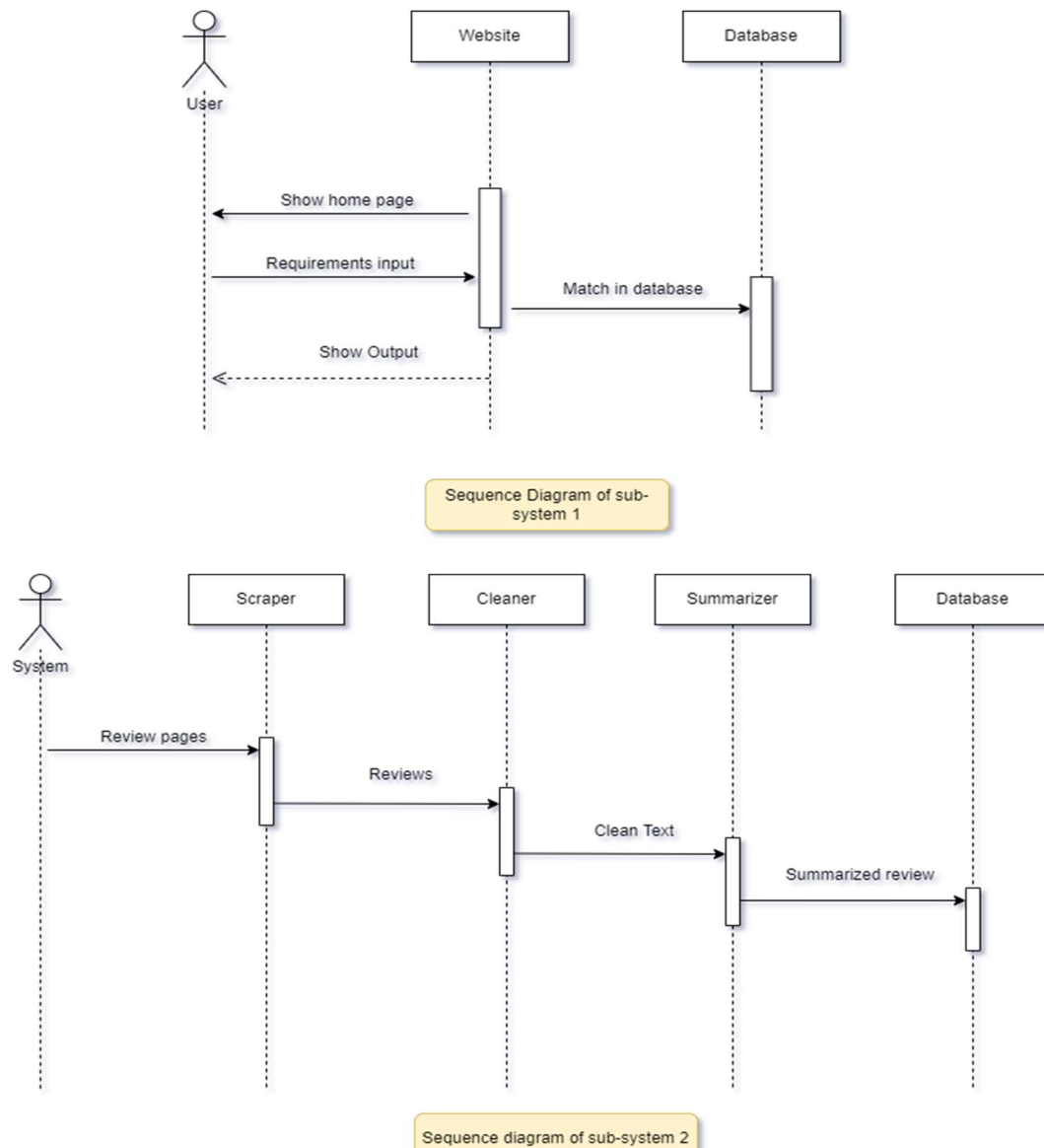
**Activity
Diagram**

Description:

Each day, at a certain point in time, the system will run its scraper, cleaner, and NLP model's summarizer in sequence to update the database.

Whenever a user uses the system and inputs their requirements, the system would run the request with the requirements parameters and match it with available laptop models and their review from the latest updated database. Once the system produces a result, it would be the output.

Sequence Diagram



Description:

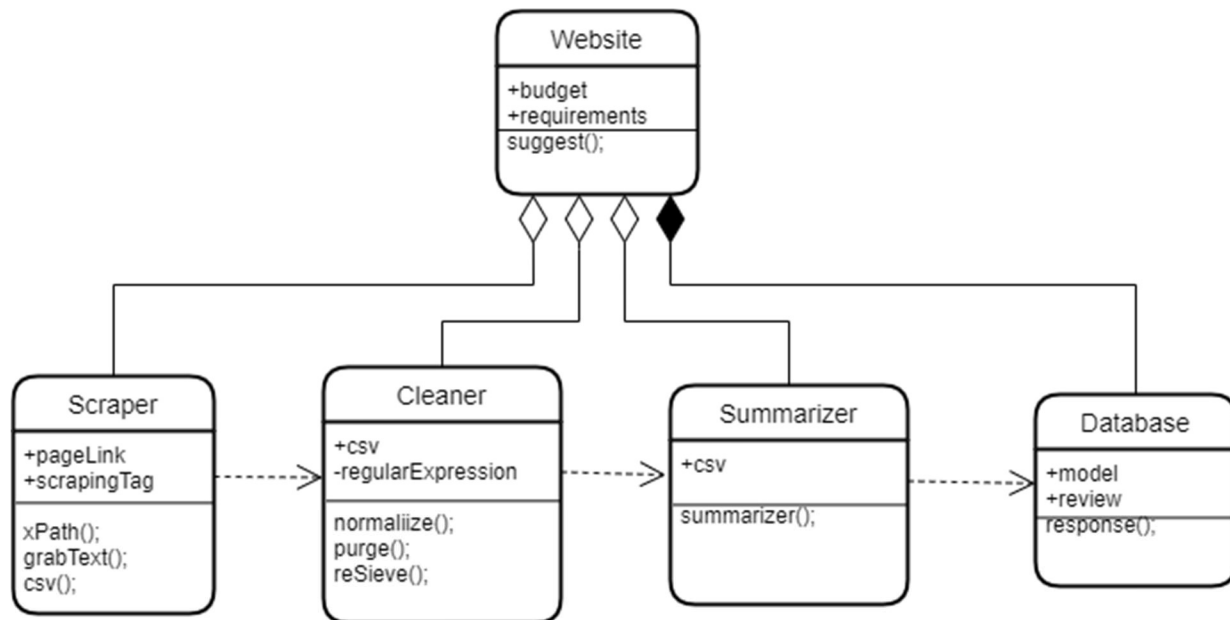
Our system needs two sub-system to work. For sequence diagram subsystem 1 it works like this:

When user will visit our website, **Website** lifeline will show home page. Our activation occurrence will stay at homepage until the user leave or give a requirement input for his desired product. If the user provides the request the requirement input our **Database** lifeline will match the input with user given data and activation occurrence will remain until it finds the output. In the end our website will provide the output to the customer.

On our sub-system 2 our system will act like an actor. It will collect the review from online pages and **scraper** lifeline will go over each of the laptop's review and collect information, activation occurrence will remain until it goes to next lifeline which is Cleaner. **Cleaner** will clean the raw reviews collected by scraper, purges the collected data of any unwanted portions that are hindrances to the NLP models functionality, and creates clean and usable versions of each review, occurrence will remain until it goes to next lifeline which is Summarizer. The **Summarizer** comprises of the NLP model and the final review for the given laptop. Occurrence will remain until it goes to next lifeline which is Database. **Database** will store the summarized review data in it.

When user wants to see reviews it will provide reviews from Database from Subsystem 2 to subsystem 1.

Class Diagram



Class Diagram

Description:

There are 5 classes: Website, Scraper, Cleaner, Summarize, Database. Where Scraper, Cleaner, summarizer is sub-part of Website class and Database is an important class for Website.

Scraper has 2 attributes and 3 operations. “pageLink” takes all necessary information, comments, review about laptops from different online platforms. “scrapingTag” collects all specification. “xpath()” is a part of selenium which used is for collecting data from websites. “grabText()” collects all important data from the page. “csv()” is used for collects all previous tasks from “Scraper” class.

Cleaner has 2 attributes where “csv” is public and “regularExpression” is private attribute. ‘csv’ will come from “Scraper’s” class ‘csv’. ‘regularExpression’ will collect only necessary phrases. “normalizer()” will convert all texts in common specific format. “purge()” will eliminate all useless data or text from recently filtered data and “reSieve()” is used for removing unnecessary word and filter all purged data.

Summarizer has only one attributes “csv” which have all filtered data. “summarizer()” will re-arrange and organize all data in proper way and create a single summary including all necessary information about laptops.

Database is the store house of all laptop information which will be upgrade every day. It has 2 attributes “model”, “review” which contains a summary review for a specific model laptop. And display it to the user.

4. Conclusion

Throughout the semester we have designed our system based on what we have learnt in the class. The diagrams that we have learnt have been useful in laying out a proper planning for building the eventual system. Some of the diagrams have proven to be challenging in adaption for this particular system as it required them to address the whole system as culmination of two subsystems. Going forward, the lessons of this course would prove to be immensely helpful in completing the project in the coming semesters, and also for other projects as well.

References:

Sentiment analysis is the analysis of customer's opinions, expressions, likes and dislikes towards products, organizations or services. These papers primarily work on this.

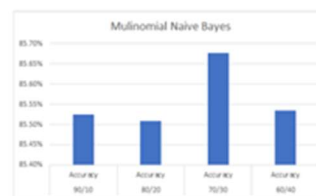
A. Product Recommendation using Sentiment Analysis of Reviews: A Random Forest Approach

Authors: Gayatri Khanvilkar, Prof. Deepali Vora

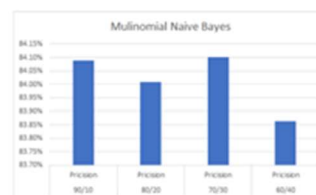
Ø Tries multiple approaches

Ø Based on various products

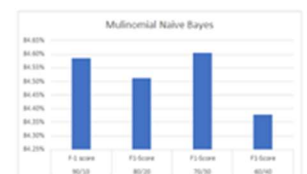
Ø Classifies the content based on polarity



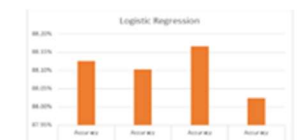
Graph 1(a) Accuracy of Multinomial NB



Graph 1(b) Precision of Multinomial NB



Graph 1(d) Recall of Multinomial NB



Graph 2(a) Accuracy of Logistic Regression



B. Data Analysis: Opinion Mining and Sentiment Analysis of Opinionated Unstructured Data

Authors: Harshi Garg, Niranjan Lal

- Ø Provides a more structurally defined approach
- Ø Goes deeper into the topic of sentiment analysis
- Ø Discusses both approach of Machine learning and Lexicon-based

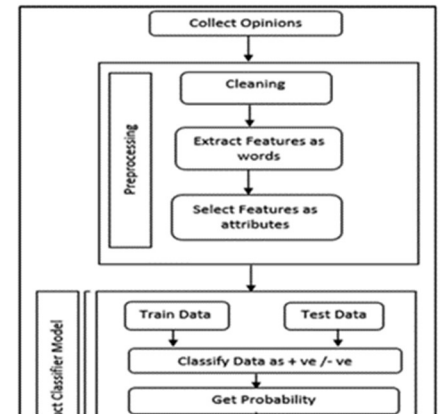


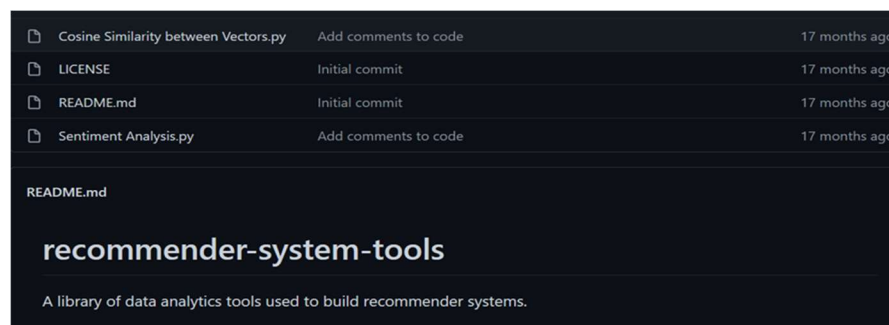
Table 5: Results for the impact of the brand on polarity review.

Brand	% of reviews	
	Positive	Negative
Samsung	79.94	20.06
Apple	77.3	22.7
Nokia	78.01	21.99
BlackBerry	74.3	25.7
Asus	77.41	22.59
LG	77.2	22.8
Sony	79.86	20.14
ZTE	82.9	17.1

D. Tell Me What You Want: A Recommender System Based on Customer Preferences and Product Reviews

Author: Josh Barua

This is an abandoned work found on GitHub that can be used as a base for one portion of the project.



Complex Engineering Problem & Mapping:

❖ How Ps are addressed through the project:

Ps	Attribute	How Ks are addressed through the project	COs	POs
P1	Depth of Knowledge Requirement	Our project requires Machine learning (K2), rigorous study of existing projects(K8), surveys & reviews from users, monitoring new products (K3, K4), knowledge of web-development(K6), data collection & analysis(K5).	CO1 CO2 CO3 CO7	PO1 PO2 PO3 PO5
P3	Dept of Analysis Requirement	Users' requirements, device feather's data, users demand etc.	CO4 CO7	PO2 P12
P4	Familiarity of Issues	We need to Analyze & implement various programming knowledge and skills to fix many issues	CO7	PO3 PO9
P7	Interdependence	Our project involves interdependent components such as requirement analysis, designing, back-end, front-end, software testing, dataset etc.	CO2 CO7	PO2 PO10

❖ **How As are addressed through the project:**

Ps	Attribute	Ps are addressed through the project
A1	Range of Resources	Device's data is collected. Web framework is an important resource for this project. It engages diverse resources including various designing tools.
A2	Level of Interaction	Successful interaction between Web framework and NLP
A5	Familiarity	Our project deals with Internet users.

❖ **How COs are addressed through the Project:**

Ps	CO Statements	Corresponding POS
CO1	Identifying a real-life problem that can be transmitted to an engineering or computing solution through design, development and validation.	PO4 PO10 PO12
CO2	Identify, formulate and analyze a real world compels engineering problem based on requirement	PO2 PO3
CO3	Design/Develop a working solution on a complex software-intensive system and verify and validate the solution using industrial state of the practice, that indicates a high-quality software-intensive system	PO1 PO5 PO11
CO4	Use a modern/popular IDE to test complex software-intensive systems.	PO7
CO7	Work as a team and fulfill individual responsibility	PO9

CO	CO Statements:	Appendix-1	Appendix-2
CO1	Apply the Engineering knowledge to provide a working solution on a complex engineering problem and submit a mapping.	1	1/Apply
CO2	Identify, formulate, and analyze a real-world complex engineering problem based on requirement analysis.	2	1/Analyze
CO3	Design/Develop a working solution on a complex software intensive system and verify and validate the solution using industrial state of the practice, that indicates a high-quality software-intensive system	3	1/Apply
CO4	Use a modern/popular IDE to test complex software-intensive systems.	5	1/Apply
CO5	Identify societal, health, safety, legal and cultural issues related to the project.	6	1/Analyze
CO6	Practice concepts of professional ethics, confidentiality, industrial standards.	8	3/Valuing
CO7	Work as a team and fulfill individual responsibility.	9	1/Apply
CO8	Communicate effectively through presentation and write effective reports and documentations on the project.	10	1/Apply
CO9	Apply project management principles using Version Control System, and appraise project operating cost, financial risk analysis for complex software intensive systems.	11	1/Apply
CO10	Recognize the need for, and have the preparation and ability to engage in independent and life-long learning for art of project management, distributed and collaborative software development and risk analysis for developing complex software-intensive systems.	12	1/Apply

No.	PO	Differentiating Characteristic
1	Engineering Knowledge	Breadth and depth of education and type of knowledge, both theoretical and practical
2	Problem Analysis	Complexity of analysis
3	Design/ development of solutions	Breadth and uniqueness of engineering problems i.e., the extent to which problems are original and to which solutions have previously been identified or codified
4	Investigation	Breadth and depth of investigation and experimentation
5	Modern Tool Usage	Level of understanding of the appropriateness of the tool
6	The Engineer and Society	Level of knowledge and responsibility
7	Environment and Sustainability	Type of solutions.
8	Ethics	Understanding and level of practice
9	Individual and Team work	Role in and diversity of team
10	Communication	Level of communication according to type of activities performed
11	Project Management and Finance	Level of management required for differing types of activity
12	Lifelong learning	Preparation for and depth of Continuing learning.