MIT WORLD PEACE UNIVERSITY

School of Computer Engineering and Technology

124, Paud Road, Kothrud, Pune 411038, Maharashtra- India

Web: www.mitwpu.edu.in

Prepared By

Swaroop Nayak(1032180202)

Kartik Bhutada(1032180229)

Amey Bhide(1032180301)

Divyang Bagla(1032180739)

Guided By

Prof. Avadhoot Joshi

Index

1. Introduction

Project statement

Sentiment Analysis and Summarization of Product Reviews using Deep Learning Techniques

Area

1. Natural Language Processing (NLP)
2. Machine Learning (ML)

Project Introduction and Aim

- ➔ As potential customers, people usually seek help from the online portals to gain knowledge on a particular product, and finally, decide if the purchase should be made or not.
- ➔ It takes several hours to read all the reviews, sometimes even leading to missing out the important ones, thus ending up making the wrong decision on purchasing the product.
- ➔ Each product has thousands of reviews each, and it is tedious for the customer to make a decision based on the varying user reviews.
- ➔ A more well defined and concise product review is proposed such that the user need not skim through all the reviews, thus saving their time and effort.
- ➔ As a solution to the ongoing problem that the customer experiences daily, automatic review summarization will be used to analyze the product reviews and convert them into a user-readable and in a more concise and precise format.

2. Literature Survey

Sr. No.	Paper Title	Publisher Name	Year of Publication	Technologies/ Algorithms used	Research Gaps
1.	Business Intelligence Visualization Using Deep Learning Based Sentiment Analysis on Amazon Review Data	IEEE	2021	Decision Tree, Logistic Regression, Stochastic Gradient Descent, Multinomial NB, SVM, BERT, and LSTM	Large model of BERT can be used to achieve high accuracy, there are various irregularities and ambiguities like sarcasm, dialect differences, metaphors, lack of context, homonyms, idioms, etc., in human language which impairs the ability of algorithms to acquire better efficiency.
2.*	Product Sentiment Analysis of Amazon reviews*	International Journal of Computer Science and IT (IJCSIT)	2021	BERT model, Bi-LSTM, NB, Random forest classification were used and compared.	Machine learning algorithms don't give good accuracy, sentimental analysis done on mobile reviews only, advanced feature extraction techniques can be used to achieve high accuracy.
3.	Sentiment Analysis of Ecommerce Product Review Data Based on Deep Learning	IEEE	2021	Convolutional Neural Network, BOW to perform Word Vector.	The model has been trained for the Chinese characters dataset and the implementation in English is yet to be done.
4.	Summarization and Prioritization of Amazon Reviews based on multi-level credibility attributes	IEEE	2021	LSTM, NLTK, TF-IDF	The paper discusses how we can find the credibility of reviews and summarizes the rest of the reviews based on positive and negative keywords.
5.	Text Summarization for Thai Food Reviews using	IEEE	2021	Simplified Sentiment Analysis, Multi-dimensiona	Processes only on 1 comment and summarizes it. Applied to Thai language dataset.

	Simplified Sentiment Analysis			l lexicon.	
6.	Classification of Shopify App User Reviews Using Novel Multi Text Features	IEEE	2020	**Feature engineering:-** TF-IDF, Chi, Chi-2, Bag of Words **ML Algorithms**: Random Forest, Adaboost Classifier, Logistic Regression	Only been experimented on a single dataset (the Shopify app dataset) and also does not discuss deep machine learning models on different text and categorical datasets for the purpose of user review classification.
7.	Natural Language Processing based Abstractive Text Summarization of Reviews	IEEE	2020	Preprocessing techniques like Lemmatization,removal of stop words etc., TF-IDF vectorization technique, LSTM model is used.	Text summarization done on food reviews of amazon dataset, Summarized reviews results are not upto the mark it can be improved.
8.*	Massive Multi-Document Summarization of Product Reviews with Weak Supervision*	ACM Chapter	2020	Weak Supervision, Clustering, Multi-Document Summarization.	The proposed work only gives the result in a single summarized manner and does not list the pros and cons of the products by classifying positive and negative reviews.
9.	A Unified Dual-view Model for Review Summarization and Sentiment Classification with Inconsistency Loss	ACM SIGIR	2020	Shared text encoder, summary decoder, Dual-view Sentiment Classification Module	Multi-document summarization is not been considered here, Results are not upto the mark.

10.	Summarizing Product Reviews Using NLP Based TextSummarization	IJSTR	2019	Seq2seq model for summarization, used LSTM layers for encoding and decoding modules during the summarization.	Only 1 comment can be processed at a time and multi-document summarization is not compatible, Other models can be used to achieve high accuracy.
11.	A Hierarchical End-to-End Model for Jointly Improving Text Summarization and Sentiment Classification	2018 International Joint Conference on Artificial Intelligence	2018	A hierarchical end-to end model, which consists of a summarization layer and a sentiment classification layer	Results are not satisfactory. Only 1 comment can be processed at a time and multi-document summarization is not compatible.
12.	Extractive Multi-document Summarization Using Multilayer Networks	Physical A Journal: Elsevier	2018	Extractive multi document summarization, multilayer networks, PageRank algorithm	Abstractive text summarization not explored and also sentiment analysis not looked into

3. Problem Statement

Project Scope

- ➔ Our opinion and purchasing decision-making are affected by the experience of others and their feedback about products. We always ask others about their opinion to get the benefit from their experience; hence, the importance of reviews has grown.
- ➔ Nowadays, a company can easily collect reviews from users via e-commerce platforms

and recommender systems, but it is difficult to read through all the wordy user reviews. Therefore, distilling salient information from user reviews is necessary.

- ➔ So there is a need to process the maximum amount of information in the least amount of time.
- ➔ It is critical to analyze such feedback due to the volume and redundancy. This work investigates an efficient way to analyze such feedback and solve the problems related to the classification and summarization of app reviews.

Project Limitations

- ➔ Multilanguage Reviews summarization is not compatible with our proposed system.

Project Objectives

- ➔ To provide a general overview of a product having tons of reviews available so as to take a quick glance with both pros and cons of any product.
- ➔ To provide a brief abstractive summary of a product with a large number of reviews available in order to take a quick glance at both the pros and cons of any product.
- ➔ We will tackle the above problem using deep learning techniques like LSTM-attention mechanism and/or BERT.
- ➔ We will have 2 phases: classification and summarization
- ➔ Processing techniques will be the same for summarization, we only have to perform processing once.

4. Project Requirements

Requirements Rationale

Requirement	**Rationale**
Results should be displayed in a reasonable amount of time (depending on the input data)	Issue: how long does it take before the user gets impatient Position 1: result is displayed in a reasonable amount of time Position 2: result is taking longer than expected time to generate result Argument: We don't want the user to get impatient Assumption: Any product which loses the attention of the user tends to go bust Decision 1: this requirement is mandatory and nonnegotiable Decision 2: generate results in a reasonable amount of time sacrificing.

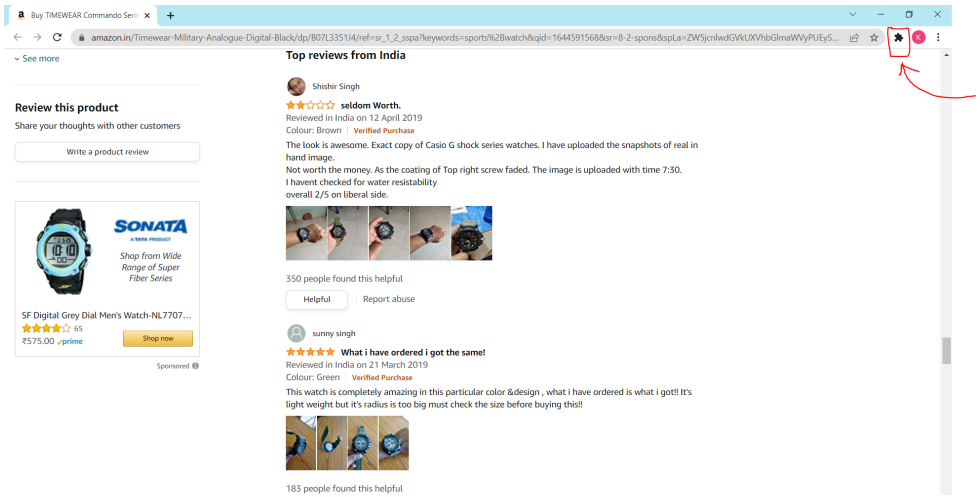
Risk Management

Probability	Risk	Mitigation
High	Time for result generation is high	Develop the product further focusing on the optimization of the algorithms used
Medium	Insufficient Data Input	Perform Data Augmentation to artificially increase the size of data input
Low	Similar product exists	Change the approach and methodology of solving the problem statement

Functional Specifications:

➔ External interfaces required

The external interfaces will be an icon in the browser. While using the browser, a user's click will trigger a panel containing a button.



➔ **Internal interfaces required**

- Sentiment Analysis Model
- Text Summarization Model
- Server which will take reviews from the client and pass on to Models.

➔ **Communication interfaces**

The extension and the server are the only two points of contact. JQueryAJAX will be implemented. It's used to send and receive requests. The protocol used will be HTTP.

➔ **Graphical User interface**

Chrome Extension's Graphical User Interface

➔ **Interactions**

- User clicks the "View Summary" button.
- System takes the web page and sends it to the cloud server.
- Web page is parsed, and the summarization algorithm takes its body part as input.
- The algorithm then sends the summary sentences to the extension. The extension shows the summary of the reviews in pros and cons form.

5. System Analysis Proposed Architecture/ high level design of the project

Design Consideration

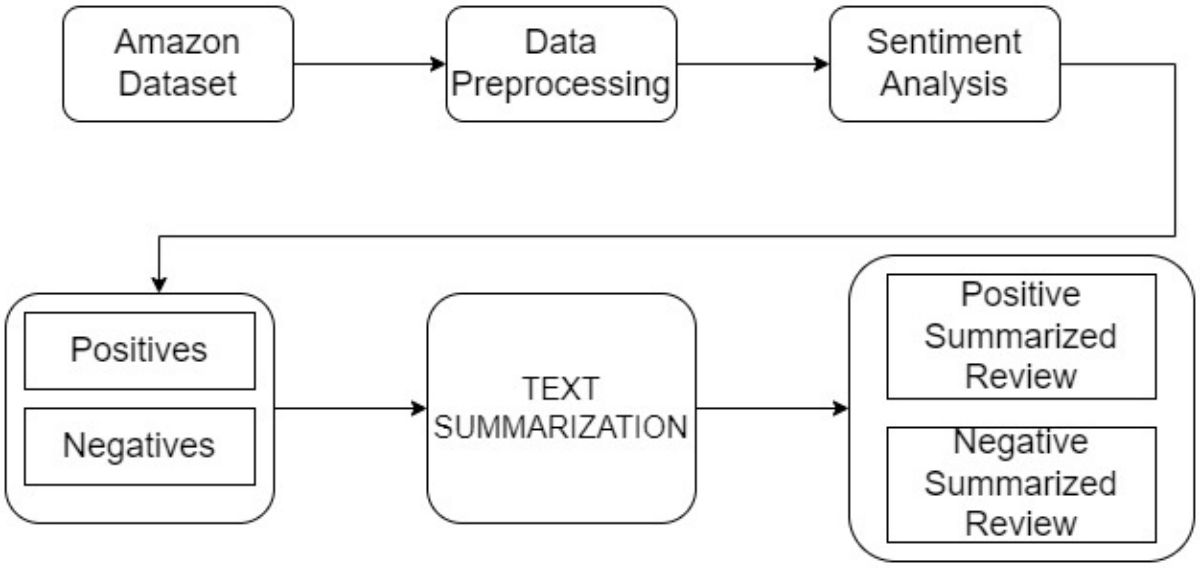
➔ Functionality

➔ Information Design

➔ Data Science - Extracting knowledge data and communicating it in a way that is not misleading.

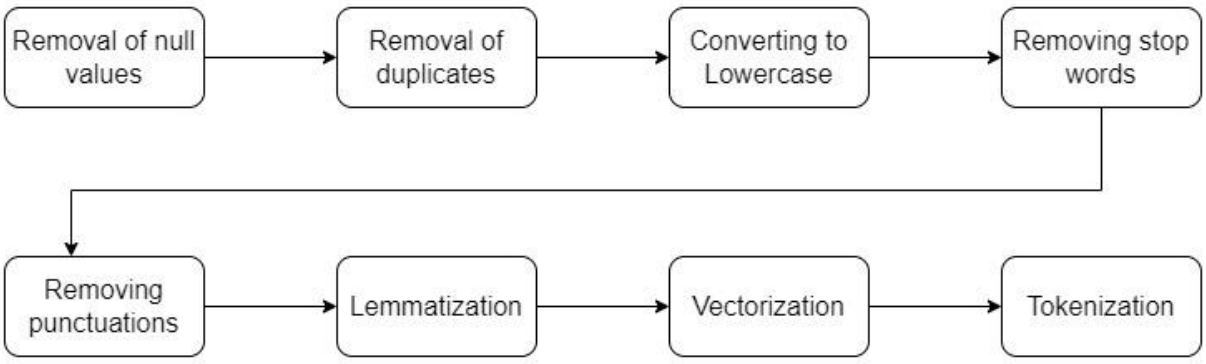
System Architecture

High Level Design

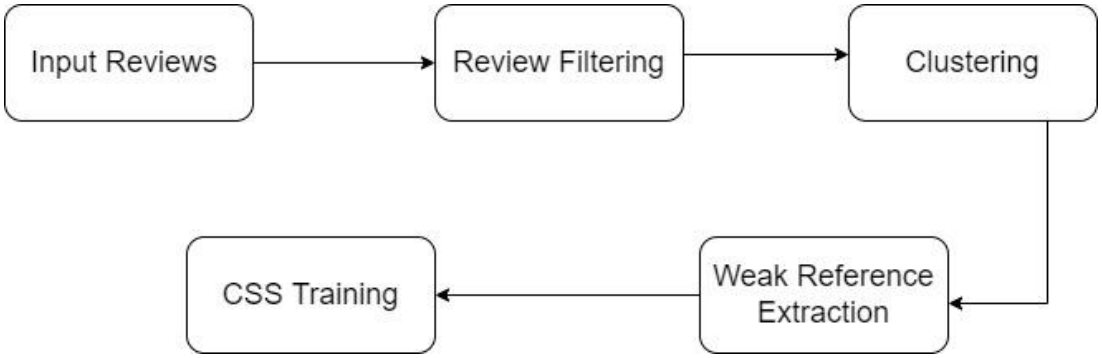


Low level Design

Low Level - Pre Processing



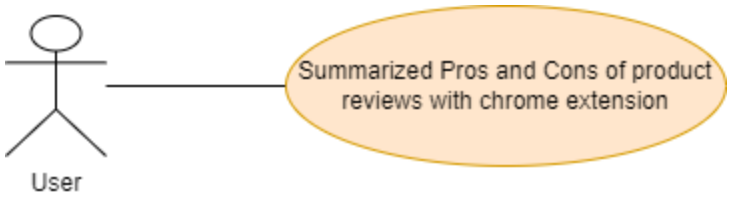
Low Level - Text Summarization



UML Diagrams/Agile Framework

Use Case Diagram

This viewpoint focuses on how different users interact with the system.

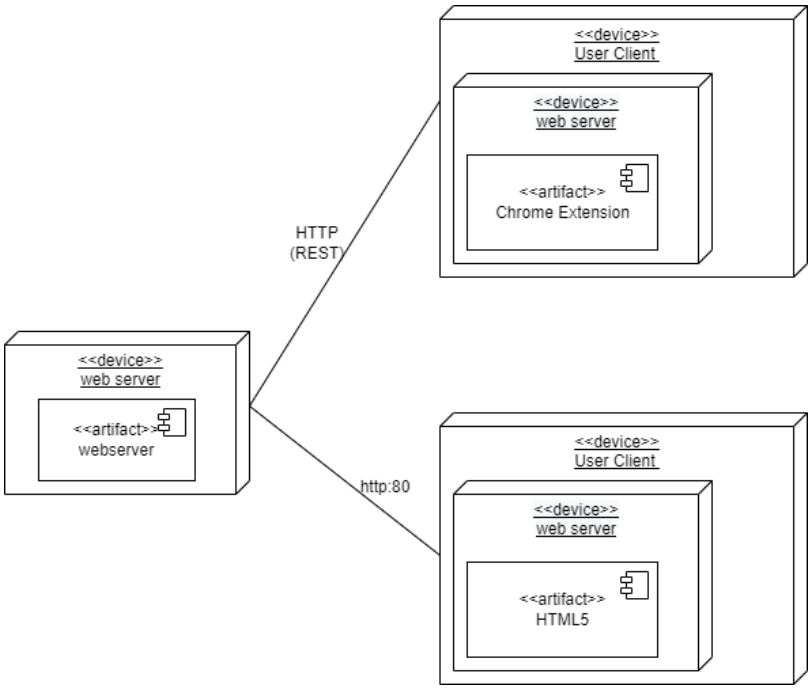


Use case Id	1
Use case	Summarize Product Reviews of a web page using Chrome Extension
Description	Getting summary of a ecommerce review webpage and showing it in the form of pros and cons without any selections on the webpage
Actor	User
Trigger	User clicks the “Summarize Reviews” button in the extension popup.
Primary Scenario	- User clicks the “View Summary” button. - System takes the web page and sends it to the cloud server. - Web page is parsed, and the summarization algorithm takes its body part as input. - The algorithm then sends the summary sentences to the extension. The extension shows the summary of the website.

Use case Id	2
Use case	Summarize Product Reviews of selected reviews using Chrome Extension
Description	Getting summary of a ecommerce review webpage and showing it in the form of pros and cons of selected reviews on the webpage
Actor	User
Trigger	- User selects some parts of the text on the webpage, by using the cursor - User clicks the “Summarize Reviews” button in the extension popup.
Primary Scenario	- User clicks the “View Summary” button. - System takes the sentences that the user selected. - These sentences are submitted to the cloud server via the extension. - These sentences are fed into a cloud-based algorithm, which generates summary sentences. - Summary sentences are then sent back to the extension

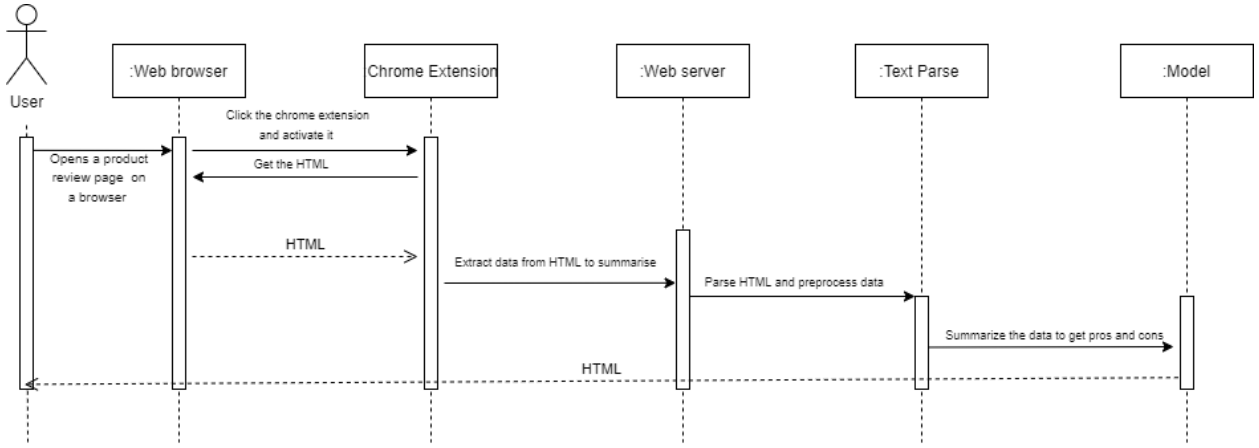
Deployment Diagram

This viewpoint focuses on the structure of the system and provides a top level view of the entire system from the perspective of each component. For this aim, the UML Deployment Diagram is provided that can be seen as below.

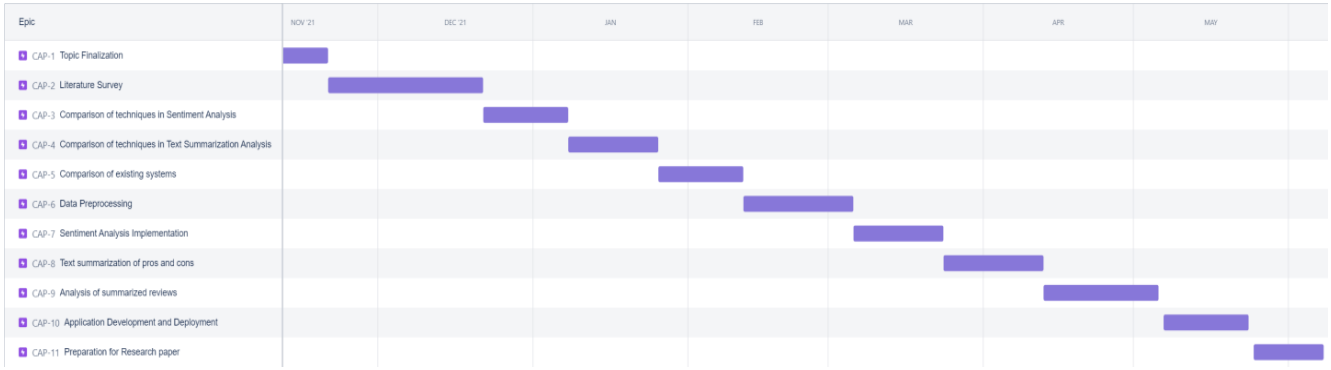


Sequence Diagram

In this view point the interaction among entities of the system will be visualized. UML Sequence Diagrams are used in order to provide representation of the interaction.



6. Project Plan



Link : [Timeline]

7. Conclusion

We have conducted an extensive literature review of multiple research papers. Along with that, we have also compared multiple methodologies by doing a comparative analysis and have

shortlisted our conclusions. The first phase of the project development has been started working on i.e. the data collection and preprocessing. From the initial phases of this project, we have a clear understanding of what we are trying to achieve and have set definitive goals and parameters to this project. We have also made a SDLC chart to help streamline our development process.

References

[1] Desai, Z., Anklesaria, K., & Balasubramaniam, H. (2021, July). Business Intelligence Visualization Using Deep Learning Based Sentiment Analysis on Amazon Review Data. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.

[2] AlQahtani, A. S. (2021). Product Sentiment Analysis for Amazon Reviews. *International Journal of Computer Science & Information Technology (IJCSIT) Vol*, *13*.

[3] L. Rong, Z. Weibai and H. Debo, "Sentiment Analysis of Ecommerce Product Review Data Based on Deep Learning," 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2021, pp. 65-68, doi: 10.1109/IMCEC51613.2021.9482223.

[4] Aishwarya, N., Bhuvana, L. S., & Kayarvizhy, N. (2021, August). Summarization and Prioritization of Amazon Reviews based on multi-level credibility attributes. In *2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)* (pp. 5-9). IEEE

[5] P. Porntrakoon, C. Moemeng and P. Santiprabhob, "Text Summarization for Thai Food Reviews using Simplified Sentiment Analysis," 2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2021, pp. 1-5, doi: 10.1109/JCSSE53117.2021.9493839.

[6] F. Rustam, A. Mehmood, M. Ahmad, S. Ullah, D. M. Khan and G. S. Choi, "Classification of Shopify App User Reviews Using Novel Multi Text Features," in IEEE Access, vol. 8, pp. 30234-30244, 2020, doi: 10.1109/ACCESS.2020.2972632.

[7] J. Shah, M. Sagathiya, K. Redij and V. Hole, "Natural Language Processing based Abstractive Text Summarization of Reviews," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 461-466, doi: 10.1109/ICESC48915.2020.9155759.

[8] Shapira, O., & Levy, R. (2020). Massive multi-document summarization of product reviews with weak supervision. *arXiv preprint arXiv:2007.11348*.

[9] Chan, H. P., Chen, W., & King, I. (2020, July). A unified dual-view model for review summarization and sentiment classification with inconsistency loss. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1191-1200).

[10] Ravali Boorugu, Dr. Gajula Ramesh, Dr. Karanam Madhavi (2019). Summarizing Product Reviews Using NLP Based Text Summarization. International Journal of Scientific & Technology Research

[11] Ma, S., Sun, X., Lin, J., & Ren, X. (2018). A hierarchical end-to-end model for jointly improving text summarization and sentiment classification. *arXiv preprint arXiv:1805.01089*.

[12] Tohalino, J. V., & Amancio, D. R. (2018). Extractive multi-document summarization using multilayer networks. *Physica A: Statistical Mechanics and its Applications*, *503*, 526-539.

Appendices

A. Base Paper(s)
    1. Product Sentiment Analysis of Amazon reviews
    2. Massive Multi-Document Summarization of Product Reviews with Weak Supervision

B. Plagiarism Report from any open source/proprietary source