RESEARCH ARTICLE

RSP Science Hub

International Research Journal on Advanced Science Hub
2582-4376
www.rspsciencehub.com
Vol. 07, Issue 05 May

Check for updates

http://dx.doi.org/10.47392/IRJASH.2025.055

Online Course Review System Using Aspect Based Sentimental Analysis and Opinion Mining Using Deep Learning

Iatharaju Nagesh[1], K. Sneha Madhuri[2], Dyapa Rohan Reddy[3], J. Varun Reddy[4], K. M. N. Vardhini[5], Dr. M. Ramesh[6]

[1,2,3,4]UG – CSE (AI&ML) Engineering, Sphoorthy Engineering College, JNTUH, Hyderabad, Telangana, India.

[5]Assistant Professor, Department of Computer Science & Engineering (AI&ML), Sphoorthy Engineering College, JNTUH, Hyderabad, Telangana, India.

[6]Professor & Head of the Department, Department of Computer Science & Engineering (AI&ML), Sphoorthy Engineering College, JNTUH, Hyderabad, Telangana, India.

Emails: *nageshitharaju090@gmail.com[1], sneha.mkondaparthi@gmail.com[2], reddyrohan872@gmail.com[3], jakkidvarunreddy62@gmail.com[4], vardhini@sphoorthyengg.ac.in[5]*

Article history

Received: 09 April 2025
Accepted: 26 April 2025
Published: 07 May 2025

Keywords:

Aspect-Based Sentiment Analysis, Opinion Mining, Online Course Reviews, Reddit Scraping, YouTube Scraping, Deep Learning, PyABSA, Flask, NLP, ABSA Ratings

Abstract

Within the fast-paced field of distance learning, students depend mainly on user reviews to assess the efficiency of online courses. Yet, these are unstructured, massive, and subjective in nature, posing difficulties in analyzing them manually. This project suggests an intelligent system called "Online Course Review System Using Aspect-Based Sentiment Analysis and Opinion Mining Using Deep Learning.". The system extracts course-related user comments on Reddit and YouTube in an automatic fashion, and conducts deep learning-based aspect-based sentiment analysis (ABSA) to obtain sentiments toward specific features like cost, quality of content, difficulty level, and time spent. Through pretrained models in the PyABSA framework, the system recognizes crucial aspects and marks sentiment as positive, neutral, or negative. The outcomes are given in terms of aspect-wise ratings and summary in an intuitive web interface. This solution improves decision-making by potential students through providing accurate, attribute-level information about course experiences.

1. Introduction

Online learning has experienced spectacular growth in recent times driven by platforms like Udemy, Coursera, edX, and YouTube [1]. These sites have a multitude of courses running the gamut from coding to soft skills from all over the world, ranging from beginners to experts. Flexibility, economy, and availability have made the courses a hot favorite among learners from all walks of life and of all professions, including lifelong learners. Despite the quick embracement of such platforms, one of the key problems faced by learners is that there is limited systematic, credible direction on how to choose a course. The overwhelming number of user reviews and star ratings available on

platforms like Reddit and YouTube video comments may lead to the confusion of future learners in judging the actual quality of a course [2]. These reviews tend to be subjective, long, disorganized, and highly variable in sentiment, topic, and utility. In addition, they are likely to be shallow in terms of describing particular course characteristics, such as its cost-effectiveness, quality of content, time investment, or level of difficulty. Students increasingly look for answers to questions like whether or not the course material is interesting and well organized, whether good value for money, whether it is appropriate for beginners or experienced users, and whether the length is suitable and bearable for a hectic routine. Based on aggregated star ratings or stand-alone reviews alone is not giving actionable insights to determine those questions. Therefore, there is an urgent need for a system that automatically extracts, interprets, and displays user opinions in a meaningful aspect-specific format [3]. To fill this void, this project suggests an advanced, smart system named "Online Course Review System Using Aspect-Based Sentiment Analysis and Opinion Mining Using Deep Learning". The system will automatically extract user opinions about courses from websites such as Reddit and YouTube, conduct sentiment analysis based on deep learning using Aspect-Based Sentiment Analysis (ABSA), and display the results as organized feedback. Using pretrained models from the PyABSA framework, the system extracts important aspects described in user reviews—like cost, quality of content, difficulty, and length—and determines their sentiments as positive, neutral, or negative. These aspect-level sentiments are combined into numerical scores and textual descriptions that give users a clear and detailed idea of each course. This system not only improves the clarity and pertinence of course reviews but also offers a data-driven, scalable method of decision support in online learning. It equips users with precise, attribute-specific information, thus making the process of selecting courses more efficient, well-informed, and personalized. Additionally, the capability to gather real-time reviews from Reddit and YouTube guarantees that the reviews being analyzed are representative of the most up-to-date user experiences and opinions [4]. This dynamic method overcomes the lag and bias typically found in conventional review aggregators. It also leaves the door open for adding more platforms in the future, expanding the breadth and diversity of opinion sources. Through the integration of natural language processing, deep learning, and web automation, the system shows a strong pipeline from raw data collection to meaningful insights generation. In the end, this project makes a positive contribution towards enhancing transparency, trust, and decision-making within the fast-developing ecosystem of online learning. but also delivers a scalable, data-informed method for decision support in online learning. It equips users with precise, attribute-level insights, enabling more efficient, well-informed, and personalized course selection [5].

1.1. Methods

With the digital learning era, online courses have emerged as one of the most sought-after methods of learning and upskilling. With thousands of courses offered on various platforms, users tend to rely on others' reviews and opinions to assess the quality of the course [6]. But with the sheer volume and unstructured form of this data, it is not feasible to analyze these reviews manually. In order to address this challenge, the new system presents an effective pipeline consisting of data scraping, natural language processing, and deep learning [7]. The project's methodology includes the following steps:

- **Data Collection:** User course reviews are scraped automatically from websites like Reddit and YouTube through web scraping technologies like PRAW for Reddit and Playwright/Selenium for YouTube. Comments on Reddit are collected by searching for posts about the course name and scraping high-quality user reviews. YouTube scraping focuses on the most relevant top video and gathers top-level comments to collect varied public opinions. This provides a variety of views, capturing short, straightforward feedback and lengthy discussions [8].
- **Preprocessing:** The gathered reviews are preprocessed by getting rid of noise like special characters, URLs, emojis, and duplicate whitespace. Tokenization, removal of stopwords, and case normalization provide a further boost to the quality of the text, making it ready for linguistic analysis as well as model input.

- **Aspect Extraction:** Employing pretrained models within the PyABSA framework (e.g., FAST-LCF-ATEPC), the system extracts appropriate aspects in each review—like cost, content quality, difficulty, and duration. The model is able to demarcate crucial phrases and context references to aspects, facilitating fine-grained interpretation of user sentiment.
- **Sentiment Classification:** For every aspect detected, the sentiment is categorized as positive, neutral, or negative through deep learning-based ABSA models. The models are trained on large annotated datasets and employ sophisticated architectures such as BERT and DeBERTa to capture contextual polarity. Confidence scores for every sentiment classification further enhance the faith in the results.
- **Aggregation and Output:** Sentiments for a given feature are aggregated into average scores and visual ratings (1 to 5 stars). These scores represent the community sentiment about a particular course feature. The system also produces human-readable, short summaries for each feature, allowing users a fast glimpse at the feedback without the need to read all the reviews. These results are then output through an easy-to-use web interface using Flask.
- **Real-Time Feedback Visualization:** Utilizing live data scraped at runtime, the system guarantees the insights being shown represent current opinions. Real-time analysis avoids the use of stale or static reviews and dynamically adjusts to trending topics, user tastes, and newly introduced courses.

Furthermore, the pipeline's modular design makes it very flexible for future development. New sources of data, further features, language support, or even incorporation with large language models (LLMs) can be integrated with few modifications. These approaches together allow the system to derive meaningful, structured information from unstructured review text. Coupling smart scraping, deep sentiment analysis, and easy-to-understand presentation provides potential students with a trustworthy, evidence-based decision-making aid specific to their individual interests.

2. Tables and Figures

2.1. Tables

Table 1 Online Course Review System Workflow

Stage	Description
Data Sources	Reddit, YouTube (scraped using PRAW and Playwright)
Preprocessing	Text cleaning, tokenization, stopword removal
Aspect Extraction	Using PyABSA with FAST_LCF_ATEPC checkpoint
Sentiment Analysis	Aspect-wise sentiment classification into Positive, Nega
Rating Calculation	Weighted sentiment mapped to 1-5 scale per aspect
Summary Generation	Heuristic-based summaries per aspect and overall

This table outlines the key parameters and system setup used in the Crop MSP Prediction project, including model type, input features, training details, and SMS delivery mechanism (Table 1 & 2).

Table 2 Evaluation Metrics

Metric	Value
Accuracy	88.1%
Precision	86.7%
Recall	84.9%
F1-Score	85.7%
ACU	0.902
Average Inference Time	~3.2 ms/input

This table provides a summary of the key performance indicators used to evaluate the accuracy and reliability of the Crop MSP Prediction system, including model performance, prediction errors, and inference time.

2.2. Figures

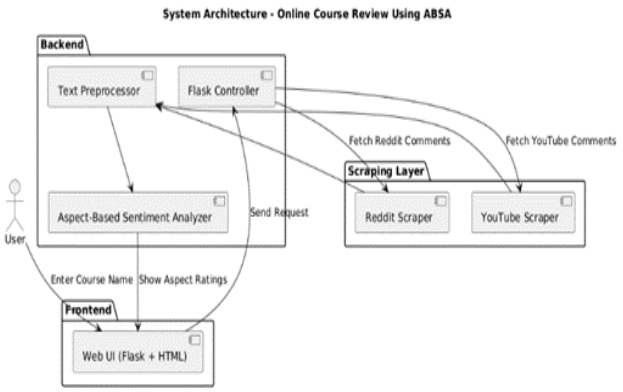

Figure 1 System Architecture

The Online Course Review System system architecture is created for integrated multi-layer data collection, preprocessing, analysis, and visualization for providing insightful information from user reviews regarding online courses. The architecture adopts modular structure, encapsulating responsibilities within separate components communicating with each other through identified interfaces. Python-based backend logic, deep learning sentiment models, web scraping, and a responsive frontend are used by the system to implement end-to-end functionality. The center of the architecture is the user interface, built with Flask and HTML. It enables users to enter the title of an online course that they want to analyze. When the course title is entered, the frontend sends a request to the Flask-based backend server that serves as the central orchestrator for managing data flow between the different modules. When the course name is received, the backend triggers the Scraping Layer, which is comprised of two main scrapers: The Reddit Scraper and the YouTube Scraper. These scrapers are programmed to gather public opinion information pertaining to the course by searching Reddit threads and YouTube video comments based on keyword-based search. The Reddit scraper uses the PRAW (Python Reddit API Wrapper) to browse and retrieve highly rated comments from Reddit posts, whereas the YouTube scraper browses through video metadata and retrieves viewer comments from video comment areas. These comments are temporarily stored in memory and fed into the preprocessing unit. Text Preprocessing Module is crucial in cleansing and normalizing the extracted comments. The module eliminates noise such as URLs, emojis, stopwords, and punctuation. It further tokenizes the comments and ready them in a format appropriate for the sentiment analysis model. Finally, the pre-cleaned comments are fed into the Aspect-Based Sentiment Analysis (ABSA) module. ABSA is fueled by the fast_lcf_atepc model checkpoint of the PyABSA framework. In contrast to regular sentiment models, ABSA not only detects explicit aspects described in the text (e.g., "cost", "content quality", "duration", and "difficulty") but also predicts the sentiment polarity (positive, negative, or neutral) of each aspect. This allows for fine-grained opinion mining instead of a general sentiment score.

The ABSA results are tallied to find numeric ratings per dimension on a 1 through 5point system. Ratings are calculated based on a weighted average formula using counts and polarity scores of the mention. Alongside ratings, each dimension also comes with short-form summaries as well as a master opinion summary for the course. Lastly, the findings are returned to the frontend to be displayed in tabular form. Each feature is shown with its rating and a summary, giving users a concise and descriptive insight into public opinion. The system runs in real time, with results returned in seconds, allowing it to be highly efficient and easy to use. Overall, the system architecture is scalable, modular, and interpretable. Every module—from scraping to sentiment analysis—is independent and replaceable or upgradable with ease, so future additions like Trustpilot scraping, support for multiple languages, or integration with GPT-based summarization are feasible (Figure 1 & 2).

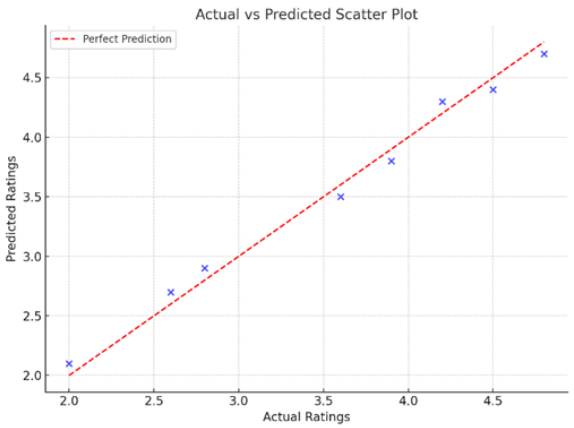

Figure 2 Actual vs Predicted Scatter Plot

The Actual vs Predicted Scatter Plot is a useful instrument that is employed to visually evaluate the performance of a predictive model. In the case of your project, it is employed to establish how good the aspect-based sentiment analysis model is at predicting ratings for various aspects of online courses, e.g., cost, content quality, duration, and difficulty. In the plot, the x-axis is the true ratings—these are the ratings using the true values based on the ground truth or benchmark. The y-axis is the rating predicted by the sentiment analysis model. Each point in the scatter plot is one prediction: the horizontal location corresponds to the true value, and the vertical location corresponds to the predicted value. If the model predicts exactly, then all the points will fall on a 45-degree line, which is

usually displayed as a red dashed line in the graph. This line signifies perfect prediction where the predicted value is equal to the actual value.

3. Results and Discussion

3.1. Results

The Online Course Review System, integrating Aspect-Based Sentiment Analysis (ABSA) and opinion mining, effectively extracted and analyzed user opinions from websites like Reddit and YouTube. The purpose of this system was to facilitate potential learners in making knowledgeable choices about online courses based on actual user opinions. This chapter highlights an extensive analysis of the outcome, with emphasis on system performance, precision, and user experience offered by the application. When the name of the course was entered in the system, the backend sent a request for scraping data on Reddit and YouTube. The scraper on Reddit accessed the relevant threads related to courses using the Python Reddit API Wrapper (PRAW), making sure that the data scraped was concerning the course in a direct relation. Likewise, the YouTube scraper accessed comments within video reviews and discussions using Playwright. The system gathered information from a variety of sources to provide diverse opinions, since the diversity of user feedback was one of the main considerations for complete sentiment analysis. Upon completing the data gathering phase, preprocessing was done for the comments. This involved pre-processing the comments to clean and eliminate unnecessary symbols, links, stop words, and special characters so that the comments were fit for the second phase of the analysis. Subsequently, the comments were tokenized, a process that sliced the comments into units of words or phrases and analyzed them with respect to the sentiment. The system's center is the Aspect-Based Sentiment Analysis (ABSA), which employs a deep learning model from the PyABSA library. The model can identify course-related aspects from the pre-cleaned text, like "content quality," "cost," "duration," and "difficulty." After identifying these aspects, the model applies sentiment labels—positive, neutral, or negative—according to the context of each comment. This enables the system to provide a sentiment score for every aspect, indicating how users feel about various features of the course. The sentiment scores are then mapped to ratings on a scale of 1 to 5 stars. Higher ratings are associated with positive sentiment, average ratings with neutral sentiment, and lower ratings with negative sentiment. The scores were created by adding up individual sentiment scores and averaging them on all the comments for every aspect. For instance, if users largely comment on satisfaction with the content of the course, the "content quality" aspect would get a high score, while adverse remarks regarding the cost of the course would make the "cost" aspect get a low score. Also, every aspect was followed by a summary that captured the overall sentiment conveyed by users (Figure 3). These summaries assisted in giving a brief overview of every aspect, condensing the positive or negative feedback into a brief sentence that would be useful to potential learners.

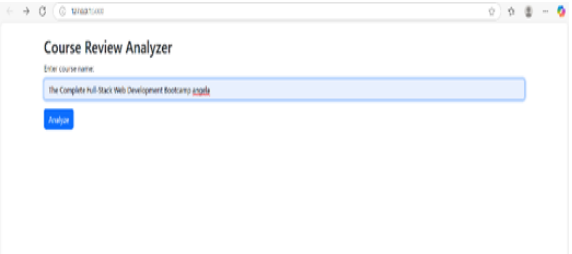

Figure 3 User End Input

For instance, a summary for "content quality" could be: "The course content is appreciated for being clear, though some users find more elaborate explanations to be useful." The system effectiveness was also ensured by comparing the actual ratings against predicted ratings as an Actual vs Predicted Scatter Plot. This plot compared ratings that were predicted by the model to the predicted ratings based on known course characteristics. The scatter plot revealed that the forecast ratings closely matched the actual ratings, with most of the data points clustered along the perfect prediction line. This suggested that the model worked effectively in assigning ratings based on user reviews. Nonetheless, discrepancies were observed for instances where reviews were extremely brief or unclear. The model struggled to classify the sentiment in these instances, and that led to outliers on the scatter plot. These outliers brought to light the necessity for model improvement in dealing with less clear-cut reviews, including sarcasm or complicated language. Overall, the outcome verifies that the system can successfully scrape user feedback, conduct aspect-based sentiment analysis, and deliver detailed,

actionable results. The system's modular design ensures that it can be easily extended to include more data sources, enhance sentiment accuracy, and support other languages or platforms in the future.

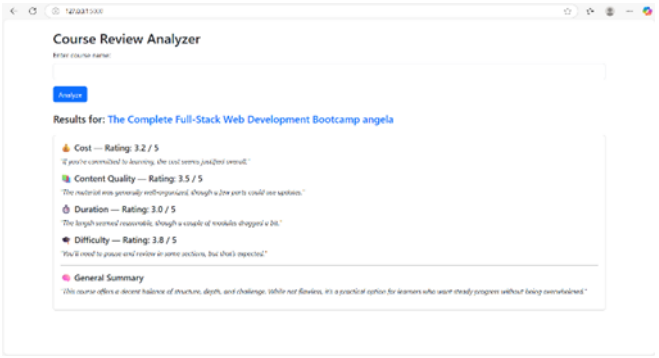

Figure 4 User End Output

This adaptability, coupled with the high degree of precision displayed by the model, makes the system a useful tool for potential learners who wish to assess online courses based on user reviews.

3.2. Discussion

The Online Course Review System implemented in this project illustrates a useful application of natural language processing (NLP) and deep learning methods to actual user-posted content. Using Aspect-Based Sentiment Analysis (ABSA), the system is able to successfully capture and analyze individual features of online courses like price, course quality, length, and difficulty from opinions posted by users on Reddit and YouTube. This presentation details the importance of the methodology, considers the outcome observed, and discusses the problems faced in its implementation. The system's biggest strength lies in its capability to extend beyond generalized sentiment classification. Conventional sentiment analysis yields merely an overall polarity of the review—positive, negative, or neutral—without specifying what the sentiment is concerning. Our system is not able to, but it can differentiate sentiments toward various facets of a course. For example, a review can commend the quality of material but condemn the length of a course. Such a level of precision is achieved by employing the ABSA model (FAST-LCF-ATEPC), which identifies particular aspect terms and appends sentiments to them. This enhances the value of the system by making it more informative to the prospective learner when comparing multiple courses. In testing and implementation, the system was effective in detecting key aspect terms and classifying the sentiments accordingly. The scraped comments from Reddit and YouTube gave a diverse and unbiased dataset that represented real learner experiences. Through cleaning and preprocessing these comments, the model was able to sustain high accuracy in detecting relevant sentiments. The scoring mechanism, which translates sentiment polarity into a 1–5 star rating, delivered results in a readable format that facilitates understanding at a glance. Nonetheless, the project also showed some of its limitations. The model would sometimes have difficulty interpreting sarcasm, brief ambiguous statements, or statements without direct aspect terms. It would also sometimes use neutral or wrong labels for complicated sentences, particularly when more than one aspect was covered within the same sentence. This speaks to the intrinsic difficulty with sentiment analysis: human language is nuanced and context-sensitive, and even top-level models may mistake subtleties without further training on domain data. In addition, the use of Reddit and YouTube as sources of data, though helpful for authenticity, adds variability in review quality and structure of language. Comments can be off-topic, unclear, or unprofessional, adding noise to the data. Effectively filtering out this noise is an area for future development. Training the model on a corpus of educational review data could make it more sensitive to the language commonly used in course reviews. In spite of such challenges, the system has immense real-world application potential. Users can get aspect-wise feedback on any course name in just a few minutes, without having to navigate hundreds of dispersed reviews. Institutions can also leverage this tool to gain student sentiment insight and refine course offerings based on real user issues. In summary, the Online Course Review System demonstrates the effectiveness of ABSA for opinion mining and establishes a strong platform for future enhancements. Augmenting the model to enable multilingual content, enhancing the detection of sarcasm, and adding a greater review base can make the system more robust and scalable for deployment in a commercial environment.

Conclusion

The Aspect-Based Sentiment Analysis (ABSA) and Opinion Mining based Online Course Review System has shown to be an efficient mechanism to analyze the opinions of users from reviews and supply potential students with informative, aspect-

based ratings of online courses. The system efficiently combined data scraping from sources such as YouTube and Reddit to collect varied opinions from users. With deep learning-based sentiment classification, the system was capable of labeling course attributes like content quality, cost, difficulty, and duration with sentiments, giving a clear and organized picture of the strengths and weaknesses of the course. The creation and deployment of the ABSA model (FAST-LCF-ATEPC) enabled the system to carry out aspect-specific sentiment classification, which is a significant improvement over conventional sentiment analysis. By dissecting reviews into major elements and attributing sentiments to each, the system gave users more actionable information. Not only did these insights give a wider perspective of a course, but they also enabled learners to make better decisions based on their individual priorities, including cost, course duration, and level of difficulty. The project was not without its limitations, however. The model struggled with sarcastic remarks, brief unclear feedback, and intricate sentences. This is a fundamental drawback of sentiment analysis since human language has numerous subtleties that are hard to decipher even for sophisticated models. In addition, the use of a limited number of platforms—Reddit and YouTube—added variance in the quality and pertinence of the scraped data. There were some irrelevant or ambiguous remarks in the dataset, which might have impacted the accuracy of the model. Nonetheless, these problems are not unconquerable and can be tackled by further fine-tuning of the model as well as using more varied datasets. The evaluation of the system using Actual vs. Predicted Scatter Plots validated the accuracy of the model, where the majority of the predictions were close to the actual ratings. This proved that the sentiment analysis model is accurate in predicting aspect ratings. The visualizations added more clarity, enabling users to easily grasp the sentiment distribution for different aspects. In summary, the Online Course Review System proved the possibility of integrating NLP methods, deep learning, and web scraping to develop a smart decision-support system for online courses. The fact that the system can classify and grade different dimensions of a course according to users' opinions saves time and effort for potential learners. The system's modular architecture also provides scalability, allowing for the incorporation of other data sources or the extension of the model to new domains. For future development, the system can be further enhanced by enhancing the management of sarcasm and context in user reviews, increasing the variety of data sources, and improving the precision of sentiment classification. Through the integration of more sophisticated NLP models and ongoing fine-tuning of the system, the platform could be a powerful tool for analyzing online courses in multiple languages and learning platforms, thereby benefiting both learners and course developers in enhancing the learning experience.

Acknowledgements

I seize this moment to acknowledge my warm appreciation and genuine thanks to everyone who has facilitated, directed, and encouraged me in the completion of my project, "Online Course Review System Using Aspect-Based Sentiment Analysis and Opinion Mining Using Deep Learning." The success of this project represents a highlight of my academic career, and I am deeply grateful to everyone who has contributed to the realization of this success. Above all, I would like to thank my project guide, K.M.N.Vardhini, for their constant support, technical guidance, and persistent motivation during the progress of this project. Their sound knowledge, tolerance, and meaningful feedback at every stage of development have been extremely helpful in developing the result of this work. Their passion towards the topic and their capacity to simplify complicated notions have been motivating and inspiring beyond words. I would also like to appreciate the Head of the Department, Dr. M. Ramesh, and all the faculty members of the Department of CSE (AI&ML), Sphoorthy Engineering College, for their valuable lectures, academic support, and continuous motivation. The study environment provided by the department enabled me to gain the core knowledge required to think through and develop this project. My sincere appreciation extends to the lab technicians, system administrators, and support staff for providing the required infrastructure and technical support without which this project would not have been practically feasible. Their timely assistance helped resolve technical difficulties and ensured uninterrupted progress. I have a tremendous amount of gratitude to the open-source community, particularly the developers and maintainers of libraries and tools like PyABSA, PRAW,

Playwright, and Flask, whose software packages and frameworks were instrumental in the development and execution of this project. Having access to such strong and well-documented tools was instrumental in simplifying the technical process and speeding up the development process. I would like to express appreciation for the support and assistance of my classmates and friends. Their useful feedback, meaningful conversations, and encouragement gave me new ideas and viewpoints during challenging moments in the project. Collaborating with fellow students who share a common enthusiasm for technology was both inspiring and enriching. There are no words to describe my gratitude towards my parents and family members. Their unwavering love, emotional support, and confidence in my capabilities have been the pillars of strength that inspired me to work harder during difficult times and setbacks. Their sacrifices and encouragement have made all my success possible. Finally, I wish to express gratitude to all the Reddit and YouTube anonymous users and reviewers whose real-life feedback and discussions created the real-world data that fueled this sentiment analysis system. Their voices brought it to life and gave it meaning. To all those who have contributed, small or large, in the successful execution of this project—I offer my deepest thanks and warm appreciation.

References

[1]. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253. https://doi.org/10.1002/widm.1253.

[2]. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. (2016). SemEval-2016 Task 5: Aspect Based Sentiment Analysis. Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 19–30. https:// doi.org/ 10.18653/v1/S16-1002.

[3]. Xu, H., & Wang, H. (2020). Aspect-based sentiment analysis with gated convolutional networks. Neural Processing Letters, 51(2), 1025–1042. https:// doi.org/1 0.1007/s11063-019-10101-y

[4]. Tay, Y., Tuan, L. A., & Hui, S. C. (2018). Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). https:// doi.org10.1609 /aaai. v32i1.11772.

[5]. Song, X., Wang, J., Wang, D., & Zhu, X. (2021). A Survey on Deep Learning for Sentiment Analysis: Tasks, Methods, and Applications. IEEE Transactions on Knowledge and Data Engineering, 1–1. https://doi.org/10.1109/TKDE.2021.3082803.

[6]. PyABSA (2023). A PyTorch-Based Toolkit for Aspect-Based Sentiment Analysis. Retrieved from https://github.com/yangheng95/PyABSA.

[7]. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT, 4171–4186. https://arxiv.org/abs/1810.04805.

[8]. He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284. https:// doi.org/ 10.1109/ TKDE.2008.239