Aspect-Based Sentiment Analysis for Online Products

Sai Prakhyaath Ladella
Osmania University

Sindhuja Joginpally
Osmania University

K. Noah Ranjit
Osmania University

A number of resources are now available for gauging public opinion about businesses, products, services, brands, and events because of the advent of e-commerce. Customers may find it difficult to distinguish between legitimate assessments and biased or false reviews, which leads to worse goods or services. Aspect-based sentiment analysis is utilized in this study to develop a natural language processing model to examine customer reviews and comments to ascertain the sentiment expressed towards certain qualities or components of a good or service. A few procedures involved in this undertaking include aspect extraction, sentiment classification, training, feature extraction, and dataset preparation. The T5 model will be used to train the model using a dataset of customer reviews, and it will be evaluated using a variety of performance measures, including accuracy (84.62%), precision (84.58%), recall (84.62%), and F1-score (83.39%). The project's findings will help customers locate services that meet their needs, and they will also assist businesses in learning more about the advantages and disadvantages of their goods and services.

Keywords: T5 model, aspect-based sentiment analysis, aspect extraction

INTRODUCTION

Sentiment analysis uses natural language processing and machine learning to analyse textual data and extract subjective information, such as the emotional tone or attitude expressed in the text. To perform this analysis, the text data must first be pre-processed, tokenized, and feature extracted.

Sentiment lexicons, which contain lists of words and phrases associated with positive, negative, or neutral sentiment, are also used to classify the sentiment of the text data. The accuracy of the sentiment classifier depends on several factors, such as the quality and size of the training data, the features used, and the choice of algorithm. Sentiment analysis has wide applications in marketing, customer service, and social media monitoring, as it can be used to analyse customer feedback and gain insights into consumer behaviour. By analyzing this data, companies can improve customer satisfaction and make better business decisions.

Aspect Based Sentiment Analysis

The traditional approach to sentiment analysis focuses on identifying the overall sentiment in a piece of text, but it lacks insights into specific aspects or features being discussed. Aspect-based sentiment analysis (ABSA) addresses this limitation by analysing sentiment towards specific aspects or features of a product or service in customer reviews. ABSA offers detailed insights, trend monitoring, and competitor analysis. The ABSA process involves aspect identification and sentiment determination using techniques like dependency parsing and machine learning. Challenges include aspect extraction, context dependency, and the need for sufficient training data. ABSA is widely used in various industries and has seen advancements in deep learning models and cross-lingual analysis

Multilingual Aspect-Based Sentiment Analysis (MASA)

Multilingual aspect-based sentiment analysis (MASA) analyses sentiment towards specific aspects in customer reviews across multiple languages. Challenges include aspect extraction in complex grammar languages, language-specific sentiment classification, and limited training data for niche languages. Approaches include machine translation, developing language-specific models, and leveraging cross-lingual models. Ongoing research focuses on generating synthetic cross-lingual training data and standardizing sentiment analysis. MASA has valuable applications, helping businesses gain insights, identify improvements, and monitor sentiment in diverse markets.

Implicit Aspect-Based Sentiment Analysis (IASA)

Implicit aspect-based sentiment analysis (IASA) identifies sentiment towards aspects snot explicitly mentioned in the text. It infers sentiment from context using unsupervised or supervised learning techniques. IASA provides comprehensive analysis, aiding improvement areas and competitor analysis. Challenges include context identification, training data availability, and errors in complex contexts. Recent advancements have led to sophisticated IASA models, enhancing accuracy. ABSA, including MASA and IASA, remains crucial for understanding customer sentiment and driving business success.

LITERATURE SURVEY

In recent research on aspect-based sentiment analysis (ABSA), various aspects have been explored, including movie reviews, hotel reviews, and different approaches. Using ABSA techniques, researchers have focused on identifying sentiments expressed towards each aspect of a product or service. Here are some notable studies in the field:

Researchers have proposed various approaches for aspect-based sentiment analysis (ABSA). For example, Srinivasulu Reddy Uyyala (Sivakumar & Reddy, U. Srinivasulu, 2020) used LSTM with FL to classify consumer review sentences. Sangeet Srivastava (Ye, Yiran & Srivastava, Sangeet., 2019) introduced an ontology framework for feature-level sentiment analysis. Guangyao Pang (Pang, Guangyao & Lu, Keda & Zhu, Xiaoying & He, Jie & Mo, Zhiyi & Peng, Zizhen & Pu, Baoxing., 2019) developed a transformer encoder based on BERT for aspect-level sentiment analysis. Pimpa Cheewa (Pimpa Cheewa Prakobkit, n.d.) compared SVM and Naive Bayes algorithms, with SVM outperforming Naive Bayes. Peiman Barnaghi (Barnaghi, Peiman & Kontonatsios, Georgios & Bessis, Nik & Korkontzelos, Ioannis., 2018) used a convolutional neural network and word embedding layers for aspect extraction. Samuel Onalaja (Onalaja, Samuel; Romero, Eric; and Yun, Bosang, 2018) explored movie aspects' influence on sentiment in reviews.

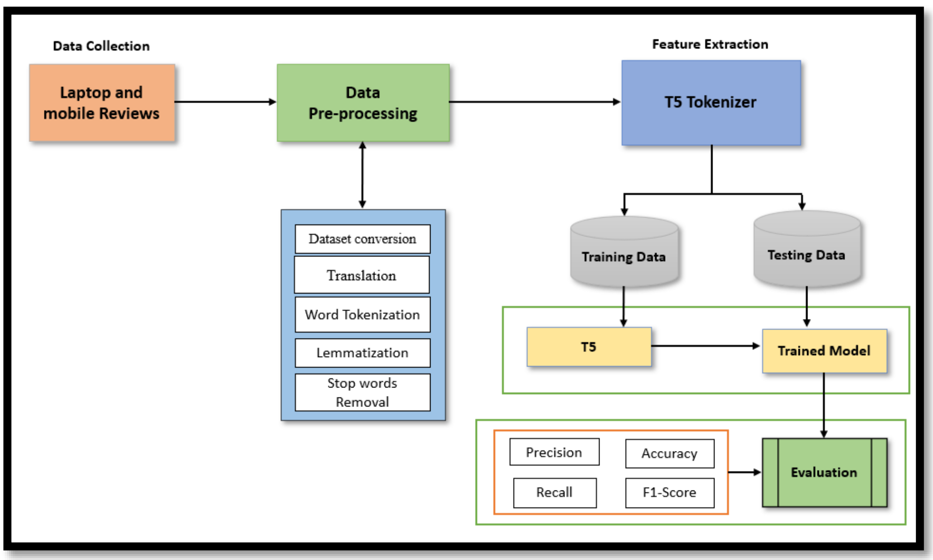
Sowmya B (Sowmya B; Dr. T. John Peter, 2018) provided an overview of ABSA techniques, discussing challenges and promising results of deep learning. Satarupa Guha (Guha, Satarupa & Joshi, Aditya & Varma, Vasudeva., 2018) described their SemEval 2015 Task 12 system and expressed interest in adapting it to other domains. Mickel Hoang (Hoang, Mickel, 2019) proposed a joint model for aspect and sentiment classification using BERT. Fredrik Olsson (Fredrik Olsson, Gustav Handmark, 2020) compared supervised models with an unsupervised dependency parser approach. Toma´s Brychc (Brychcín, Tomáš & Konkol, Michal & Steinberger, Josef., 2014) achieved above-average results using machine learning-based approaches in ABSA-Task. Yangyang Yu (Stanford, Yangyang Yu., 2014) explored class balancing techniques for ABSA. Abdulganiyu O. Harazeem (Abdulganiyu O. Harazeem, 2Umar Kabir, 3 Kakudi Habiba A, 2021) presented a novel aspect-based sentiment analysis model that emphasized the impact of aspects and emoticons on sentiment analysis. M. Ali Fauzi (Fauzi, Muhammad., 2018) employed the Word2Vec model for Indonesian sentiment classification, discussing the challenges of limited training data.

These studies showcase the advancements and potential in aspect-based sentiment analysis, with different models, techniques, and language-specific considerations.

METHODOLOGY

The methodology of this model follows certain steps, and it is depicted in the picture below, this is the model architecture.

FIGURE 1
MODEL ARCHITECTURE



DATA COLLECTION

Web Scraping

Web scraping is an automated process of extracting data from websites by fetching and parsing HTML code. It enables efficient collection of large amounts of data from diverse online sources for various purposes such as market research, data mining, and lead generation. However, it's important to respect website terms of service and robots.txt, implement rate limiting, and parse HTML carefully. Data usage and privacy must be prioritized, complying with applicable laws and ethical considerations. Responsible web scraping involves avoiding malicious activities, respecting website owners' rights, and seeking permission or legal advice when necessary.

Pre-Processing

Proper pre-processing of text data is crucial for accurate and understandable results in NLP. It removes noise like HTML tags, scripts, and advertisements from online texts. At the word level, irrelevant words that do not contribute to the overall meaning or sentiment are eliminated. This reduces dimensionality and improves classifier performance, speed, and real-time sentiment analysis. Pre-processing plays a pivotal role in cleaning and preparing text for classification in NLP, enhancing accuracy and efficiency.

Data Conversion

Converting emojis to text involves using the emoji library to transform emojis into their textual representations for consistent understanding in NLP models. If the text is not in English, a neural machine translator translates it into English for better comprehension. The translation supports multiple languages like French, German, Spanish, and Indian languages such as Hindi.

Removing numbers can be beneficial in certain cases, like sentiment analysis, where they carry little meaning. However, caution is needed for tasks like Name Entity Recognition or Part of Speech tagging, as removing numbers may affect accuracy. Regular expressions are used to identify and remove numbers. Lowercasing is a common pre-processing technique that treats words in different cases as the same entity, aiding in text featurization and reducing redundancy. It is performed before tokenization and vectorization.

Removing punctuation marks from text data standardizes word treatment by eliminating punctuation. The choice of which punctuation to exclude depends on the specific use case. In Python, the 'string.punctuation' module provides a range of punctuation symbols.

Word Tokenization

Word tokenization is a crucial step in NLP that breaks down text into individual words for analysis. It helps organize and process text data in applications like classification and sentiment analysis. Techniques like whitespace delimiters or advanced libraries like NLTK and spaCy are used for accurate tokenization. Customizing word token removal is important but requires caution. Emoticons or slang words may be removed in sentiment analysis, while machine translation may involve excluding or replacing specific words. However, removing word tokens should be carefully considered as it can result in losing valuable information or context, leading to inaccurate results. Word tokenization breaks text into manageable units for NLP analysis. It plays a vital role in various applications, but caution should be exercised when removing word tokens to avoid losing important information.

Lemmatization

Lemmatization is a technique in NLP that reduces words to their base form, known as the lemma. It helps group words with similar meanings and is widely used in text classification and sentiment analysis applications. Lemmatization normalizes text data by converting inflected or variant forms to their base form, enhancing analysis accuracy and efficiency.

The lemmatization process involves determining the word's part of speech and applying rules or algorithms to convert it to its base form. Libraries like NLTK and spaCy offer predefined rules and algorithms for lemmatization, which can be customized as per specific requirements. However, lemmatization is not flawless, and inaccuracies can occur due to imperfect part-of-speech identification and ambiguous base forms for certain words.

Removing Stopwords

Stop words are common words in a language often ignored in natural language processing and text analysis. They include articles, prepositions, conjunctions, and pronouns, carrying minimal meaning. Removing stop words is important to improve accuracy and efficiency in NLP tasks like text classification and sentiment analysis. This involves discarding these words before analysis, reducing data dimensionality. Libraries like NLTK and spaCy provide predefined lists of stop words, but customization is possible. However, it's essential to consider context and objectives, as removing stop words may not always be necessary or beneficial.

Different languages have different stop words, and some may not require their removal. Stop word removal is the final pre-processing step, refining text for further use.

FIGURE 2
THE SAMPLE OUTPUT OF PRE-PROCESSING STEPS

review	conversion	word_tokenize	lemmatize	Rev_sw	final_txt
Best Samsung phone 📱 in this price range.Good ...	best samsung phone mobile phone in this price ...	['best', 'samsung', 'phone', 'mobile', 'phone'...	['best', 'samsung', 'phone', 'mobile', 'phone'...	['best', 'samsung', 'phone', 'mobile', 'phone'...	best samsung phone mobile phone price range go...
Amazing smartphone. Camera , battery back up ,...	amazing smartphone camera battery back up disp...	['amazing', 'smartphone', 'camera', 'battery',...	['amazing', 'smartphone', 'camera', 'battery',...	['amazing', 'smartphone', 'camera', 'battery',...	amazing smartphone camera battery back display...
This smartphone is very good in camera and bat...	this smartphone is very good in camera and bat...	['this', 'smartphone', 'is', 'very', 'good', '...	['this', 'smartphone', 'is', 'very', 'good', '...	['smartphone', 'good', 'camera', 'battery', 'v...	smartphone good camera battery value money sam...
First I would like to thank you Flipkart for f...	first i would like to thank you flipkart for f...	['first', 'i', 'would', 'like', 'to', 'thank',...	['first', 'i', 'would', 'like', 'to', 'thank',...	['first', 'would', 'like', 'thank', 'flipkart'...	first would like thank flipkart fastest delive...
Pros:-Battery backup 🔋is simply awesome. I got...	pros battery backup batteryis simply awesome i...	['pros', 'battery', 'backup', 'batteryis', 'si...	['pro', 'battery', 'backup', 'batteryis', 'sim...	['pro', 'battery', 'backup', 'batteryis', 'sim...	pro battery backup batteryis simply awesome go...

Dataset Characteristics & Data Splitting

Around 8845 were scraped from various sources available on the internet and were compiled together to create a dataset for this project's training and testing purposes. Of those 8845 reviews, 6557 are positive and the remaining 2288 are negative.

When developing a ML project, one crucial decision is utilizing the existing data. One common technique is to split the data into a training set and a testing set. The training set is used for model development and parameter estimation, while the testing set is reserved for an unbiased assessment of the model's performance. The proportion of data allocated for testing depends on factors such as the size of the sample pool and the number of predictors. Choosing an appropriate split percentage that considers computational cost, representativeness of the datasets, and project objectives is important.

Our project divided the dataset into a 75% training set and a 25% testing set. The training set is used to build the machine learning algorithm, enabling the model to learn and adjust based on input data and expected outputs. The model's performance is evaluated using the testing set by making predictions on unlabelled data.

Feature Extraction

Feature extraction can be done manually by identifying and describing relevant features, or automatically using specialized algorithms or deep networks. Manual extraction requires domain knowledge and can involve methods tailored to specific data types. Automated extraction is efficient for quickly transitioning from raw data to machine learning algorithms. Deep networks have largely replaced feature extraction in image analysis. They automatically learn and extract features from raw data, reducing the need for manual or automated techniques. For feature extraction in our project, we have used T5 Tokenizer as well as FastText.

T5 Tokenizer

Google's AI research team developed a powerful language model for text processing. It breaks down text into tokens for language tasks. Its text-to-text framework handles diverse tasks without architectural changes, enabling transfer learning. T5 undergoes unsupervised pre-training on a large text corpus, predicting missing tokens and understanding language structure. Fine-tuning on specific tasks with labelled data follows. The model's flexibility and architecture make it state-of-the-art in NLP.

Fasttext

FastText, developed by Facebook's AI Research (FAIR) team, is a powerful framework for text classification and word embedding. It effectively handles out-of-vocabulary words and captures word morphology using character n-grams for feature extraction. FastText trains a shallow neural network by converting n-gram features to one-hot encoded vectors. The hidden layer captures complex relationships, and the output layer predicts class probabilities using learned word embeddings. Supervised learning algorithms optimize the weights based on the task, using negative sampling and stochastic gradient descent (SGD) during training. FastText's training process and feature extraction make it valuable in natural language processing tasks.

Training

A training model in machine learning uses input and output data to train an algorithm. Through iterative model fitting, the algorithm adjusts its output based on the provided sample output, aiming to improve accuracy. Supervised learning utilizes input and output values for training, while unsupervised learning focuses on discovering patterns without specific output references. Both methods play crucial roles in training models and enabling effective predictions and pattern identification. The accuracy of the training dataset is essential for achieving precision.

Training Arguments

Training arguments, or parameters, are vital in machine learning to optimize the learning process and determine the model's performance. They influence the model's behavior, effectiveness, and convergence. We can customize and optimize the learning process by adjusting these arguments to improve performance. Evaluation of the trained model is crucial to ensure its generalization and prevent overfitting. The learning rate, a key parameter, determines the weight updates during optimization, balancing speed and convergence. A learning rate scheduler dynamically adjusts the rate, improving convergence and preventing overfitting.

Effective logging captures important information for monitoring and analysis. Weight decay helps mitigate overfitting by encouraging smaller weights. The warmup ratio controls the gradual increase of the learning rate, preventing early overfitting. We enhance the learning process and achieve optimal results by tuning these parameters.

MODEL TRAINER

SEQ2SEQ Trainer

The Seq2Seq Trainer in Transformers simplifies training sequence-to-sequence models with options like distributed training, mixed precision training, and early stopping. It feeds input and output sequences iteratively, computes loss by comparing predicted and ground truth sequences, and updates the model.

The architecture comprises an encoder and a decoder, typically implemented with recurrent neural networks using LSTM or GRU cells. The encoder processes the input sequence, generating a fixed-length representation, and the decoder generates the output sequence. The Seq2Seq Trainer supports distributed training, mixed precision training, and early stopping, providing a versatile solution. The Seq2Seq architecture handles input and output sequences and is effective in machine translation, text summarization, and question answering tasks. It leverages recurrent neural networks and LSTM/GRU cells to capture sequence information.

EVALUATION METRICS

Evaluation metrics can help you assess your model’s performance, monitor your ML system in production, and control your model to fit your business needs. Our goal is to create and select a model which gives high accuracy on out-of-sample data. It’s crucial to use multiple evaluation metrics to evaluate your model because a model may perform well using one measurement from one evaluation metric while performing poorly using another measurement from another.

Accuracy

Accuracy of an algorithm is represented as the ratio of correctly classified predictions (TP+TN) to the total number of predictions (TP+TN+FP+FN).

Accuracy = (TP+TN) / (TP+TN+FP+FN) (1)

Precision

Precision is the ability of a classifier not to label an instance positive that is actually negative. For each class it is defined as the ratio of true positives to the sum of true and false positives.

- TP – True Positives
- FP – False Positives

Precision= TP / (TP+FP) (2)

Recall

Recall is the ability of a classifier to find all positive instances. For each class it is defined as the ratio of true positives to the sum of true positives and false negatives.

Recall = TP / (TP+FN) (3)

F1 Score

The F1 score is a weighted harmonic mean of precision and recall such that the best score is 1.0 and the worst is 0.0. Generally speaking, F1 scores are lower than accuracy measures as they embed precision and recall into their computation.

F1 Score= 2*(Recall∗precision) / (Recall+Precision) (4)

FIGURE 3
EVALUATION METRICS OF T5 MODEL

```
Train Metrics:
Precision: 0.8767536966891323
Recall: 0.8752327746741154
F1 Score: 0.8658560168367513
Accuracy: 0.8752327746741154

Test Metrics:
Precision: 0.8458520192979794
Recall: 0.8462697814619442
F1 Score: 0.8339886336329335
Accuracy: 0.8462697814619442
```

RESULTS AND ANALYSIS

The model is trained with the help of 8845 reviews that are available in e-commerce websites. The mobile reviews collected had both text and emojis. Then, the corpus is passed into T5 model, where the feature extraction is done with the help of T5 Tokenizer. Once the embeddings are done then the training starts. It is trained using Seq2Seq trainer. After the training both accuracy and loss are calculated.

T5 Model

FIGURE 4
ACCURACY GRAPH OF T5 MODEL

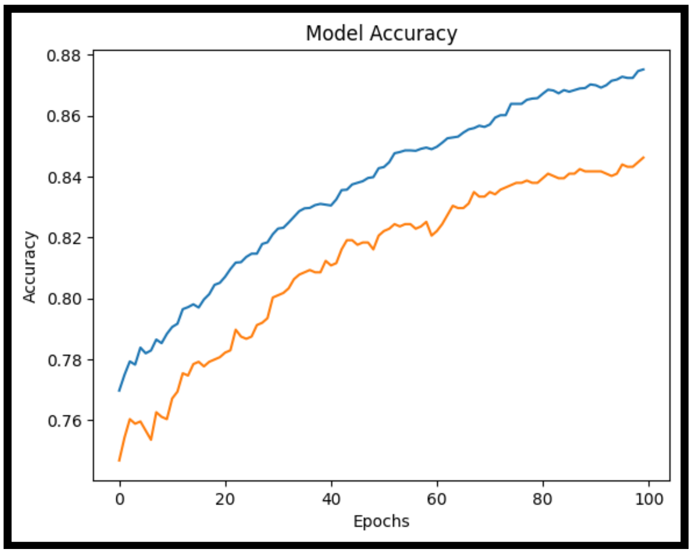
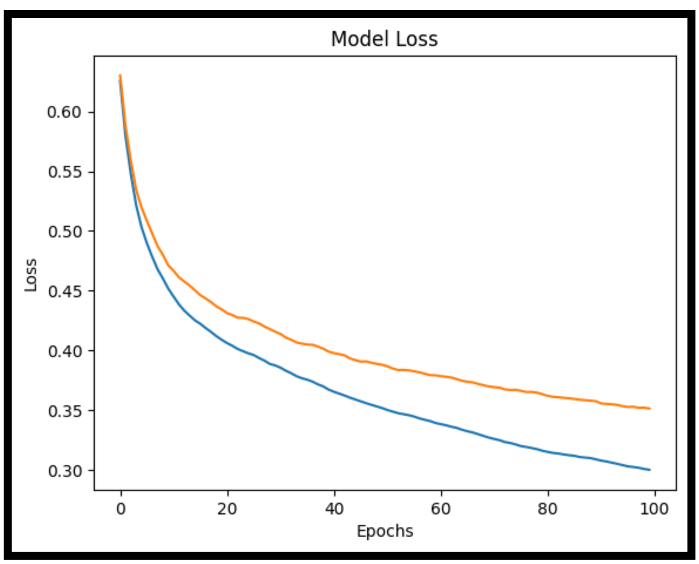


FIGURE 5
LOSS GRAPH OF T5 MODEL



FASTTEXT+BI-LSTM

FIGURE 6
ACCURACY GRAPH OF FASTTEXT+Bi-LSTM MODEL

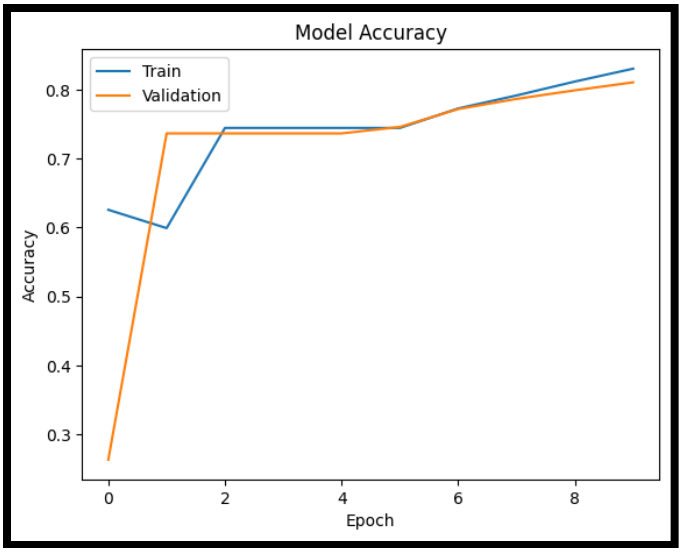
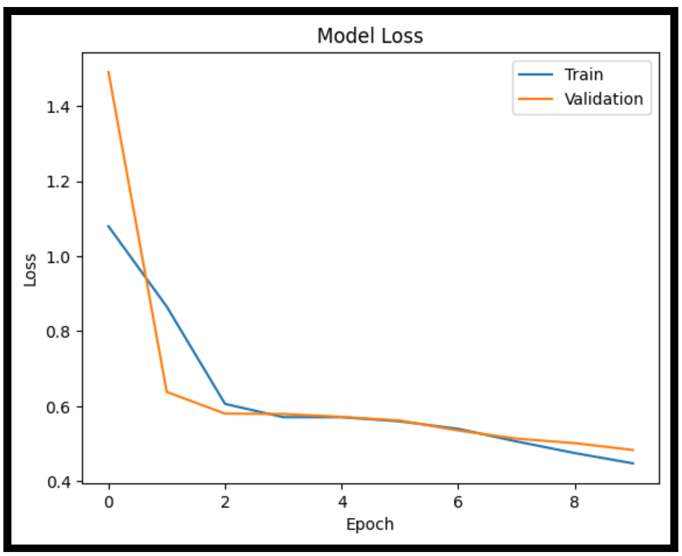


FIGURE 7
LOSS GRAPH OF FASTTEXT+Bi-LSTM MODEL



The models are then compared with each other to find the best performing model. Below represents the comparison of the T5 and FastText models concerning the evaluation parameters. The algorithms used are in combination with FastText are Bi-LSTM.

The proposed method of using T5 model for the ABSA has given satisfactory results. The accuracy of all these models is listed below. Moreover, it is observed that the T5 method results are slightly better than the FastText methods. The accuracy of T5 Model is 84.62% consistently. The second-best performing model is FastText + Bi-LSTM with an accuracy of around 81.06% but with some discrepancies. The accuracies of other models are not up to the expectation mark and are inconsistent.

CONCLUSION

The proposed work utilized various techniques, including T5, for text preprocessing, aspect extraction, and sentiment classification. T5, the ABSA system developed, was evaluated using performance metrics and compared to existing approaches. Results demonstrated its effectiveness in accurately identifying sentiments for different aspects.

One limitation was the system's computational requirements for processing emojis in sentences. Additionally, challenges arose in scrutinizing emojis directly for aspects, leading to cases where sentiments couldn't be displayed due to factors like spelling mistakes or extensive reviews. Future work aims to address these limitations and enhance the model's performance.

REFERENCES

Abdulganiyu, O.H., Kabir, U., & Kakudi, H.A. (2021). Emoticon aware aspect based sentiment analysis of online product review. *Dutse Journal of Pure and Applied Sciences (DUJOPAS)*, *7*(2a). https://doi.org/10.1007/978-3-319-75487-1_19

Akhtar, M., Ekbal, A., & Bhattacharyya, P. (2018). *Aspect based sentiment analysis: Category detection and sentiment classification for Hindi*. https://doi.org/10.1007/978-3-319-75487-1_19

Barnaghi, P., Kontonatsios, G., Bessis, N., & Korkontzelos, I. (2018). Aspect extraction from reviews using convolutional neural networks and embeddings. DOI: 10.1007/978-3-030-23281- 8_37.

Brychcín, T., Konkol, M., & Steinberger, J. (2014). UWB: Machine Learning Approach to Aspect-Based Sentiment Analysis. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Doi: 10.3115/v1/S14-2145.

Fauzi, M. (2018). Word2Vec model for sentiment analysis of product reviews in Indonesian language. *International Journal of Electrical and Computer Engineering (IJECE)*, *9*, 525. DOI:10.11591/ijece.v9i1.pp525-530

Guha, S., Joshi, A., & Varma, V. (2018). *SIEL: Aspect based sentiment analysis in reviews*. DOI:10.18653/v1/S15-2129.

Hoang, M. (2019). Aspect-based sentiment analysis using BERT. *Nordic Conference of Computational Linguistics*.

Olsson, F., & Handmark, G. (2020). *Sentiment analysis and aspect extraction for business intelligence*.

Onalaja, S., Romero, E., & Yun, B. (2018). Aspect-based sentiment analysis of movie reviews. *SMU Data Science Review*, *5*(3), Article 10.

Pang, G., Lu, K., Zhu, X., He, J., Mo, Z., Peng, Z., & Pu, B. (2019). Aspect-level sentiment analysis approach via BERT and aspect feature location model. *Wireless Communications and Mobile Computing*. DOI: 10.1155/2021/5534615.

Pimpa, C.P. (n.d.). *Comparing SVM and naïve bayes classifiers for online reviews based on aspect-based sentiment analysis*. Faculty of Business Administration Asia-Pacific International University.

Sivakumar, Reddy, U., & Srinivasulu, M. (2020). Aspect-based sentiment analysis of mobile phone reviews using LSTM and fuzzy logic. *International Journal of Data Science and Analytics*, *12.* DOI: 10.1007/s41060-021-00277-x.

Sowmya, B., & John, P.T. (2018). Aspect Based Sentiment Analysis: Approaches and Algorithms. *International Journal of Innovative Science and Research Technology (IJISRT)*, pp. 1668–1671 . ISSN-2456-2165. https://doi.org/10.5281/zenodo.7223089

Srivastava, S., & Yiran, Y. (2019). Aspect-based Sentiment Analysis on mobile phone reviews with LDA. *International Conference on Machine Learning Technologies*, pp. 101–105. DOI:10.1145/3340997.3341012

Stanford, Y.Y. (2016). *Aspect-based Sentiment Analysis on Hotel Reviews*.