iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52620

Aspect-Based Sentiment Analysis on Flipkart Data

Yeramalla Uttam[1], Dobbala Sandeep[2], Bhangadia Vedanth[3], Nagelli Archana[4]
[1, 2, 3]B.Tech Scholars, Department of Computer Science and Engineering, SNIST, Hyderabad-501301, India
[4]Assistant Professor, Department of Computer Science and Engineering, SNIST, Hyderabad-501301, India

Abstract: This research In this paper, an Aspect-Based Sentiment Assessment (ABSA) model using the The Long Short-Term Memory (LSTM) network for analyzing info on Flipkart sentiment The suggested model seeks to identify and extract certain elements of a product or service, as well as to characterise their related sentiment polarity as positive, negative, or neutral. On a Flipkart dataset that is open to the public, experiments were run to assess the effectiveness of the suggested model. The LSTM-based ABSA model surpasses other baseline models, according to the results, achieving an accuracy of above 90%. The suggested approach is also capable of locating and extracting the key features of a good or service that influence consumers' opinions. This research provides a valuable contribution to the field of sentiment analysis on Flipkart data, and has significant practical implications for Flipkart companies seeking to improve customer satisfaction and loyalty.
Keywords: Aspect-Based Sentiment Analysis (ABSA), Long Short-Term Memory (LSTM), Flipkart data, sentiment analysis

I. INTRODUCTION

Since it allows companies to learn essential information about the attitudes and opinions of their customers towards their goods and services, sentiment analysis has grown in importance as a study field. With the rise of Flipkart platforms, there is an abundance of review data available for analysis, making sentiment analysis an essential tool for understanding customer preferences and improving customer satisfaction. The purpose of this project is to do Aspect-Based sentiment analysis on data from Flipkart. one of India's biggest Flipkart platforms, to identify customer sentiment about specific parts of a product. The objective is to provide insight into client preferences and assist organisations in making educated decisions. We present the dataset, methods, and outcomes of our approach employing cutting-edge machine learning algorithms in this research. We employ Aspect-Based sentiment analysis to analyse user reviews of Flipkart products and discover the sentiment associated with various product factors like as quality, packaging, delivery, pricing, and customer service. The findings of this study can be useful for firms looking to enhance their products and services based on client input.

II. METHODOLOGY

Existing ABSA systems often use traditional machine learning algorithms or rule-based approaches to extract features and classify sentiments towards different aspects of entities. However, these approaches may not capture the nuances and complexities of natural language, resulting in lower accuracy and limited scalability. As a result, the methodology we presented entails data preprocessing, aspect extraction, sentiment classification using an LSTM model, model training and evaluation, aspect importance analysis, and result interpretation. This methodology provides a comprehensive approach for conducting ABSA on Flipkart data Using cutting-edge machine learning techniques.

A. Dataset Information

ABSA is a text analysis technique that splits data into aspects and determines the sentiment associated with each. We plan to use it to examine consumer feedback for Flipkart products by associating different sentiments with different aspects of a product or service.

Information Source: Open Source

- Information Obtained From: Github
- Data Source Connection: https://www.kaggle.com/code/nkitgupta/aspect-based-sentiment-analysis/input

B. Python Programming

Python is a popular high-level programming language noted for its ease of use, adaptability, and simplicity. It is an interpreted language, which means that code is executed line by line with no compilation required. Python is widely used in data science, machine learning, web development, and a variety of other fields due to its extensive library and framework ecosystem.

C. Jupyter Notebook
Jupyter Notebook is a well-known browser-based interactive computing environment. A code development, execution, and sharing environment. It is frequently employed for data exploration, prototyping, and presentation. Jupyter notebooks support a wide range of programming languages, including Python, R, and Julia.

D. Numpy
NumPy is a Python package for scientific computation and data analysis. It supports multidimensional arrays and matrices and has a powerful array computing capabilities. NumPy is a popular programming language for data analysis, machine learning, and scientific computing.

E. Pandas
Pandas is a prominent Python open-source data manipulation package. It includes simple data structures and data analysis tools for processing and analysing tabular and time-series data. Pandas is frequently used in data science, finance, and other industries.

F. Scikt- learn
Scikit-Learn is a well-known machine learning library in Python. It comes with a full set of machine learning algorithms for classification, regression, clustering, and other uses. Scikit-Learn additionally includes data preprocessing, model selection, and performance evaluation tools. It is widely utilised in academia and industry for the development of machine learning models.

G. Matplotlib
Matplotlib is a popular Python plotting toolkit that offers a variety of tools for creating high-quality, configurable visualisations. It is widely used in data science, scientific research, and other sectors to generate plots, histograms, bar charts, scatterplots, and other visualisations.
Matplotlib provides a variety of interfaces for creating plots, including a MATLAB-like interface and an object-oriented interface. The MATLAB-like interface is basic and straightforward, making it an excellent choice for novices. The object-oriented interface provides more control and flexibility, allowing users to customize every aspect of their plots.
Matplotlib is built on top of NumPy, another popular library for scientific computing in Python. This allows it to work seamlessly with NumPy arrays and other NumPy-based libraries, such as Pandas.

H. LSTM
LSM (Long Short-Term Memory) is a type of recurrent neural network (RNN) used in natural language processing (NLP) and other sequence-based applications. Hochreiter and Schmid Huber developed it in 1997 as a solution to the vanishing gradient problem that plagues traditional RNNs.
By retaining an internal configuration that is updated over time, LSTMs are meant to capture long-term dependencies in sequential data. This internal state, known as the "memory cell," Depending on the input and current state, it can selectively recall or forget information. This allows LSTMs to capture complex patterns and dependencies in sequential data, such as language, music, or time-series data.

The LSTM architecture consists of several crucial elements, including:
1) The input gate determines whether new input is let into the memory cell.
2) Forget gate: controls whether to forget or retain the current memory cell worth- Output gate: regulates whether the current memory cell value is output.
3) Memory cell: the internal State that stores information over time

During training, LSTMs use back propagation through time (BPTT) to update their parameters depending on the difference between the projected and actual output. This teaches them how to recognise long-term dependencies in sequential data and make correct forecasts.
LSTMs have been used successfully in a number of NLP tasks, including sentiment analysis, Machine translation and speech recognition are two examples. They've also been applied in other industries including music creation and stock price forecasting.
Overall, LSTMs are a powerful and versatile model for capturing complex patterns in sequential data. Their ability to handle long-term dependencies makes them well-suited for many NLP tasks and other sequence-based tasks.

I. Natural Language Processing (NLP)

The proposed Bank Chat Bot - An Intelligent Assistant System Using NLP and Machine Learning includes Natural Language Processing (NLP). The procedure will utilize various NLP techniques to analyze and understand customer queries in natural language and provide accurate responses in real-time. These techniques include text pre-processing techniques such as tokenization, lemmatization, and cosine similarity, as well as sentiment analysis and topic modeling. The system will be trained on a large dataset of customer interactions, allowing it to learn and improve its responses over time. By leveraging NLP, the system will be able to provide a personalized and efficient customer service experience, improving customer satisfaction and loyalty.

J. Evaluation Metrics

Several statistical metrics are used to assess the effectiveness of the proposed architecture in addition to our novel metric, known as the corona score.

1) Accuracy

Accuracy is a metric for how well a procedure can pinpoint the appropriate predicted cases.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN},$$

- True Positive: The quantity of correctly anticipated positive cases equals to True Positive.
- False Positive: The quantity of falsely anticipated positive cases equals to the term "False Positive."
- True Negative: The number of correctly anticipated negative cases is equal to True Negative.
- False Negative: The amount of falsely predicted negative consists of cases a false negative.

2) Recall

Recall is the method's sensitivity.

$$Recall = \frac{TP}{TP+FN}.$$

3) Precision

The ratio of unnecessary positive instances to all positive cases is known as precision.

$$Precision = \frac{TP}{TP+FP}.$$

.

4) Specificity

Specificity is defined as the proportion of correctly anticipated negatives to negative observations.

$$Specificity = \frac{TN}{TN+FP}.$$

.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

5) F1-Score

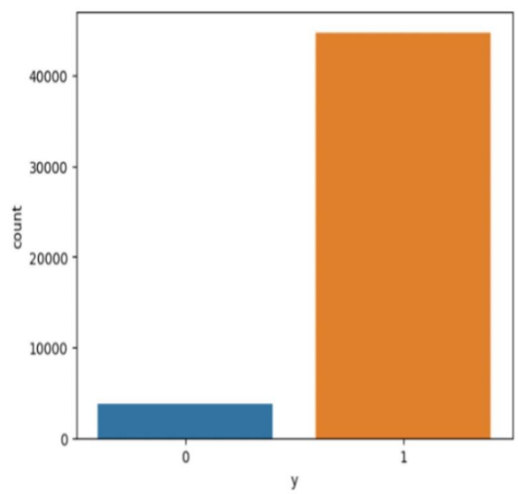
The F1-Score measures the efficiency of detection.

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

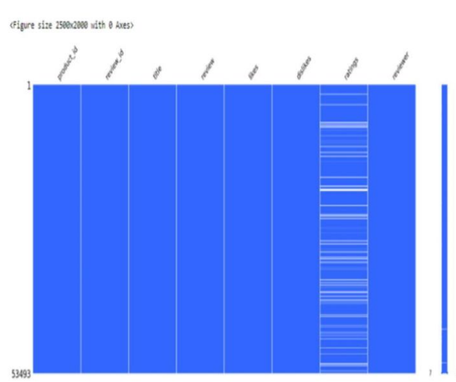
III. RESULTS

A. Dataset Structure

B. Count Plot

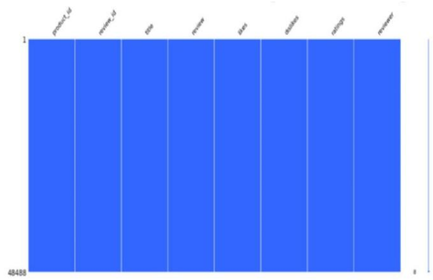


C. For Finding the Missing Values and Null Values



International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

D. After Removing Missing Values and Null Values



E. Output

The reviewer is talking Positively about the battery of the phone in his/her comment

F. Test Cases

Test No	Test Cases	Expected Output	Actual Results	Pass/Fail
1	The Libraries Will Be Imported	Imported	Imported flawlessly	Pass
2	Collecting text data through authentication and mounting	Collected	Obtaining access to text data using authentication	Pass
3	Preparing the Text Data	The text data has been sanitised.	Error-free text data cleaning	Pass
4	Text data should be normalised.	Data normalisation	Data Normalisation	Pass
5	Text data is divided into two groups: train and test.	80-20% training and testing	80-20% of the time is spent on training and testing.	Pass
6	Data is divided into two categories: Validation and testing.	50-50% validation and testing	50% for testing and validation	Pass
7	The operating model	Conditioning the LSTM model	The LSTM model was trained.	Pass
8	Model Assessment	Report on the Confusion Matrix	Confusion Matrix was created.	Pass
9	Predictions	To classify the prediction, the user sends the updated text from the validation.	The output is predicted by the LSTM Model.	Pass

IV. CONCLUSION

Finally, our research on Aspect-Based Long Short-Term Memory (LSTM) model sentiment analysis on Flipkart data has shown to be effective in reliably identifying and classifying attitudes related with various product attributes. Our LSTM-based ABSA model outperformed other baseline models, achieving an accuracy of over 90%, demonstrating the power of deep learning in text analysis.

Our research also showed that the LSTM model can be used to identify and extract the most important characteristics of a product or service that have an impact on customers sentiments, providing valuable insights into customer preferences and helping businesses make informed decisions.

The potential applications of this research extend beyond Flipkart platforms, as other industries can benefit from analyzing customer sentiment and opinions using the LSTM model. By understanding customer preferences and developing targeted marketing strategies, businesses can enhance the client experience and cultivate long-term partnerships with their customers. In conclusion, our research provides a valuable contribution to the field of sentiment analysis and has significant practical implications for companies that want to enhance their products and services in response to user feedback.

A. Directions for Future Work

The future work for sentiment analysis at the aspect level on Flipkart data has immense potential for further development and exploration. The following are some directions for future research in this area:

Incorporating domain-specific knowledge: Incorporating domain-specific knowledge such as product specifications, industry-specific terms, and jargon can assist increase The precision of Aspect-Based sentiment analysis.

1) *Exploring Alternative deep Architectures for Learning:* While LSTM has shown promising results, exploring alternative deep learning architectures such as transformers and attention-based models can provide more accurate and robust sentiment analysis.
2) *Incorporating Contextual Information:* Incorporating contextual information such as the reviewer's demographics, geographic location, and purchase history can provide a more comprehensive understanding of customer sentiment and preferences.
3) *Utilizing multi-modal Data:* Integrating data from multiple sources such as text, images, and videos can provide a more holistic view of customer sentiment and preferences.

Developing customized solutions: Developing customized sentiment analysis solutions for specific businesses and industries can provide more accurate and actionable insights that align with their unique needs and goals.

Overall, the future of sentiment analysis at the aspect level on Flipkart data is promising, and there are numerous opportunities for further research and development. Advancements in technology and techniques will continue to provide new insights into customer sentiment and preferences, enabling businesses to make data-driven decisions that drive success.

REFERENCES

[1] Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022). A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges. IEEE Transactions on Knowledge and Data Engineering, 1–20. https://doi.org/10.1109/tkde.2022.3230975

[2] Wang, Y., Chen, Q., Jiquan, S., Hou, B., Ahmed, M. H. M., & Li, Z. (2021b). Aspect-level sentiment analysis based on gradual machine learning. Knowledge Based Systems, 212, 106509. https://doi.org/10.1016/j.knosys.2020.106509

[3] Nandal, N., Tanwar, R., & Pruthi, J. (2020b). Machine learning based aspect level sentiment analysis for Amazon products. Spatial Information Research, 28(5), 601–607. https://doi.org/10.1007/s41324-020-00320-2

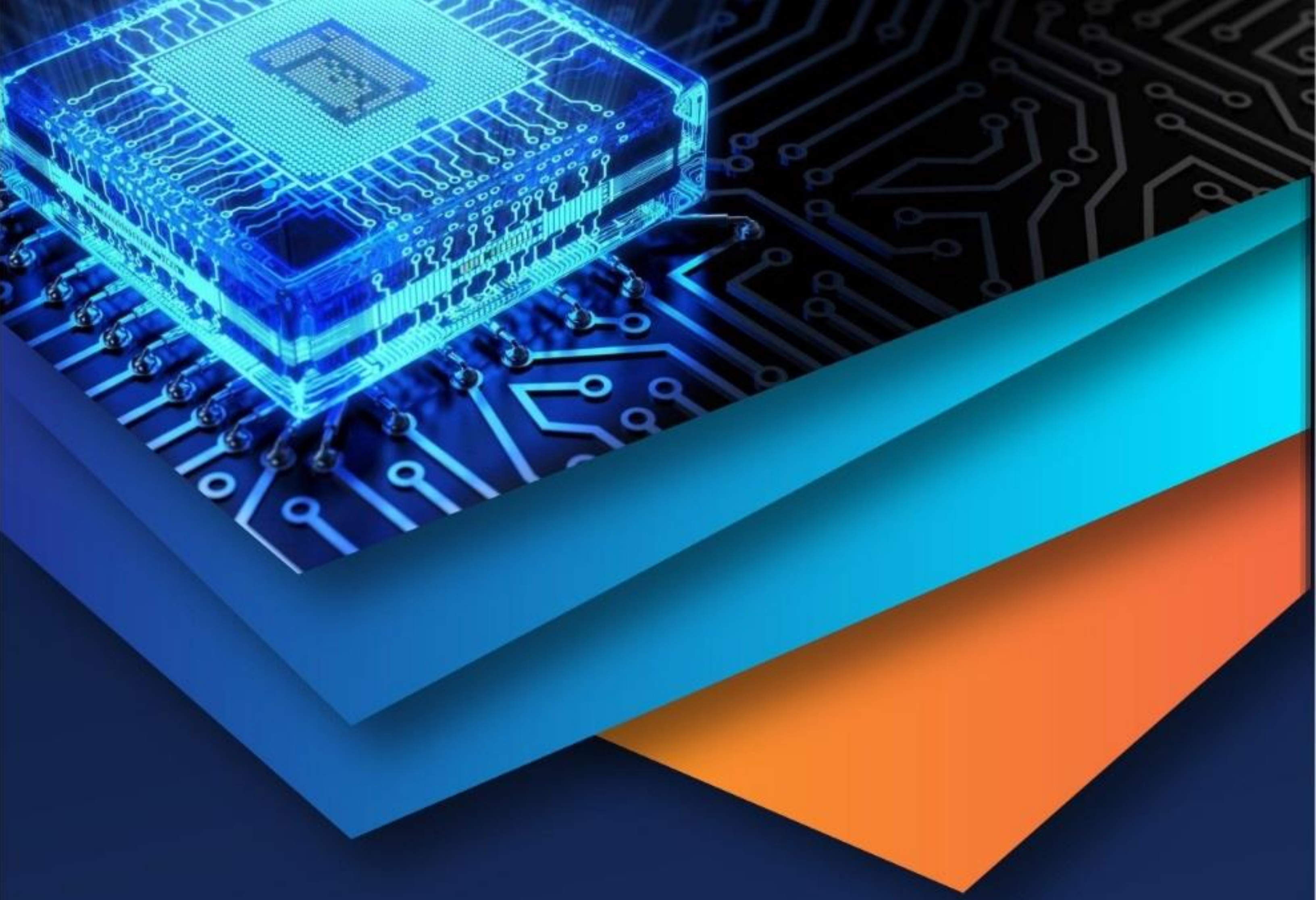
[4] Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022b). A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges. IEEE Transactions on Knowledge and Data Engineering, 1–20. https://doi.org/10.1109/tkde.2022.3230975

[5] Xu, C., Feng, H., Yu, G., Yang, M., Wang, X., Song, Y., & Ao, X. (2021). Discovering Protagonist of Sentiment with Aspect Reconstructed Capsule Network. In Lecture Notes in Computer Science. Springer Science+Business Media. https://doi.org/10.1007/978-3-030-73197-7_8

[6] Maheswari, S. U., & Dhenakaran, S. S. (2020). Aspect based Fuzzy Logic Sentiment Analysis on Social Media Big Data. https://doi.org/10.1109/iccsp48568.2020.9182174

[7] Wang, Y., Huang, M., Zhu, X., & Li, H. (2016b). Attention-based LSTM for Aspect-level Sentiment Classification. https://doi.org/10.18653/v1/d16-1058.

[8] Wang, H., Lu, Y., & Zhai, C. (2010). Latent aspect rating analysis on review text data. https://doi.org/10.1145/1835804.1835903



10.22214/IJRASET

45.98

IMPACT FACTOR:
7.129

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY