An Opinion Mining Based Pretrained Model Analysis Depending on Multiple Thinking Patterns

Abstract—**People love to express their opinions online, and the use of social media for this purpose has skyrocketed in the internet age. Customers also frequently share their first-hand knowledge of the goods or services they have used. Reviews can be either favorable or bad, and both have an impact on people's decisions and businesses. Therefore, it is essential to predict people's opinions in order to preserve the credibility of the online review system, for which traditional machine learning could have a significant impact. Pretrained models based on transformers, such as BERT, RoBERTa, and DistilBERT, may be effective at classifying people's opinions. In this work, we explored the behavioural pattern of the models based on different types of datasets. At the end of this work we were able to find that the models were pretrained with biased datasets.**

Index Terms—**EDA, tokenization, BERT, RoBERTa, DistilBERT**

I. Introduction

Nowadays, people communicate their views and beliefs differently thanks to the internet. Today, millions of people share their daily thoughts and emotions on social networking sites like Facebook, Twitter, Google Plus, and others. We have access to interactive media through online communities, where members can use forums to educate and persuade others. There is a wealth of sentiment-rich data generated through tweets, status updates, blog posts, comments, reviews, and other forms of the social media material. Social media also gives businesses an opportunity by giving them a platform to interact with their clients for marketing purposes. When making judgments, people mainly rely on user-generated content from the internet. For instance, before choosing whether to buy a product or use a service, a person will research it online and discuss it on social media. Too much user-generated content is being produced to be processed by the average user. Numerous different sentiment analysis techniques are in use because this must be automated.

Sentiment analysis is a technique for examining text data to determine its intention. It is contextual text mining that recognizes and extracts subjective information from the source material. It assists businesses in understanding the social sentiment of their brands, products, and services while keeping an eye on online discussions [1]. It is often referred to as opinion mining or emotion AI. In other words, it involves looking through online writing to determine whether it has a positive, negative, or neutral emotional tone. To estimate overall sentiment, the objective is to automatically identify and classify opinions stated in the text.

The fundamental goals of textual information retrieval strategies are to process, search for, or examine the factual material that is already there. Even if facts have an objective component, some other literary contents exhibit subjective traits. Sentiment Analysis's fundamental components—opinions, sentiments, assessments, attitudes, and emotions—are primarily represented by these contents (SA). In large part because of the enormous expansion in the amount of information available online from sources like blogs and social networks, it presents many challenging chances to design new applications. For instance, by using SA and taking into account factors such as positive or negative attitudes about the goods, recommendations of items proposed by a recommendation system can be predicted. Figure 1 shows the whole diagram of our proposed model. We have taken a total of five datasets on which we have used models to identify the performance based on opinion mining. Whenever people work on datasets there is a very high chance to have a few unnecessary data which causes a bad impact on the results. So, in starting we have to preprocess our datasets through EDA (Exploratory Data Analysis) which helps to remove unnecessary values from the datasets. Unnecessary values refer to null values and avoid those words which are not related to emotions. After data analysis, data can be tokenized by dividing up different types of text input into tokens. As our dataset includes punctuation and other special characters so they are deleted from the input data during tokenization because they don't affect the analysis's correctness. We are using tokenization because these tokens aid in context comprehension or model development for NLP (Natural language processing) [2]. By examining the word order in the text, tokenization aids in comprehending the text's meaning. Then we are using three models named -

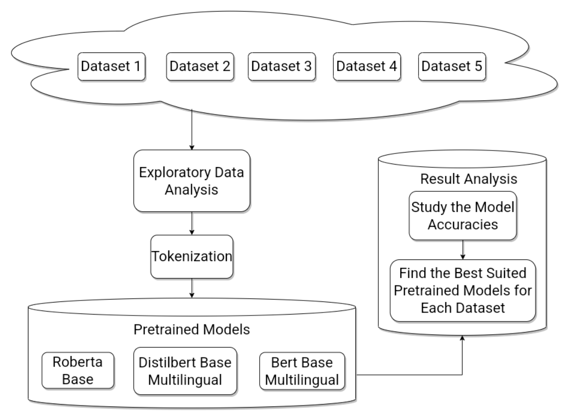

Fig. 1. System Block Diagram

Roberta, Distilbert, and BERT. In the end, we have analyzed with graphs and accuracy which model gives the most perfect opinion mining to review.
Our motivation is that future customers' purchasing decisions are influenced by the reviews posted on internet platforms, and this has a financial impact on firms. People enjoy leaving reviews and like to express their opinions on social media. Building a trustworthy online review system can often be challenging when it comes to text classification. Finding the behaviour of the models based on the types of the datasets can help us to identify the proper and correct model which we can use to analyze the opinions of the customers more accurately and this is our target.

II. LITERATURE REVIEW

The development of sentiment analysis systems is fraught with difficulties. It is first important to determine the text's content. Due to the complexity of language, which contains many semantic nuances not found in other types of data, this is not a simple task. Second, it is necessary to categorize feelings in some way in order to ascertain their orientation. There are various approaches to solving this issue.

In the existing literature, De Kok, S., et al [3] worked on the amount of material available on the World Wide Web there is now more interest in sentiment analysis, which aims to ascertain people's opinions on a subject. Instead of analyzing the sentiment of the entire text, aspect-based sentiment analysis focuses on the sentiment of the text's mentioned components inside a specific segment based on restaurant reviews. They employ a review-based algorithm and a sentence aggregation technique for this objective. Our models perform much better at classification when using ontology as their knowledge basis. Additionally, compared to the sentence aggregation method, the review-based approach makes predictions that are more accurate.

In another literature, Bansal, B., et al [4] analyze Twitter sentiment analysis which is a simple and affordable method for contemporary election predictions and real-time election monitoring. They have offered Hybrid Topic Based Sentiment Analysis (HTBSA), which aims to capture word relationships and co-occurrences in brief tweets. Using the Biterm Subject Model (BTM), they first extract latent themes from a large corpus of short texts and then learn the sentiments associated with each topic from pre-existing lexical resources. Geo tagging is used for keyword phrases that are not just related to elections.

Hasan, A., et al [5] have performed sentiment, subjectivity analysis, or polarity calculations to analyze the opinions or text that are present on various social media sites. The contribution of this research is the use of a hybrid strategy that uses a sentiment analyzer with machine learning to address these issues. Additionally, this study compares sentiment analysis methods used to analyze political opinions using supervised machine-learning techniques like Naive Bayes and support vector machines (SVM).

K. Ravi, et al [6], this study reviews sentiment analysis which illustrates opinions expressed in more than 100 articles published in the previous ten years addressing the crucial tasks, methodologies, and sentiment analysis applications. This study, which examines material that was published between 2002 and 2015, is divided into smaller tasks that must be completed, as well as applications of sentiment analysis and machine learning. Along with a summary table of 116 articles, the document also discusses unresolved concerns.

Dictionary classifications based on semantic orientation are used in unsupervised semantics-based techniques to categorize various word kinds [7]. Finding from the literature that they have worked on one type of particular dataset from biased/unbiased/semi-biased rather than performing on every type. In order to demonstrate which behavior dataset is appropriate for our classifier and to determine accuracy, we tried to work on every type of behavior dataset simultaneously.

III. SYSTEM IMPLEMENTATION

A. Details of dataset

We have taken a few datasets based on reviews from consumers. In our daily life, we use to buy clothes, electronic devices or apps which have become a regular part of our life and consumers give reviews based on their satisfaction or disappointment. The datasets we have chosen are named - amazon preprocessed kindle book reviews [8], amazon women's dresses reviews [9], apple iPhone SE reviews [10], and Spotify reviews [11]. Each of these datasets includes features that measure peoples' ratings, reviews, comments, and other feedback. Attributes declare the satisfaction of the consumers based on using materials like- Kindle, Women's dress, iPhone, and Spotify. People put ratings out of five stars and give comments or reviews based on how satisfied they are with utilizing them. These datasets are surveys which is helpful for the sellers to judge their product and service. In the beginning, we have to preprocess every dataset to clean the data and prepare it for a machine learning model, which also improves the model's accuracy and effectiveness.

B. Exploratory Data Analysis (EDA)

EDA's major goal is to encourage data analysis before making any assumptions. It can assist in finding glaring errors, better understanding data patterns, spotting outliers or unusual occurrences, and discovering intriguing relationships between the variables. Exploratory Data Analysis (EDA) [12] is an approach for data analysis that employs a variety of techniques -

1. null value drop from datasets;
2. counting of five-star ratings and Pie Chart based on ratings;
3. Counting words and Average use of words in each review tag;
4. comment length frequency

Null value drop: From every dataset we have planned to ignore data consisting of null values. In the EDA process first, we have dropped out those columns which consist of a null value because they are not useful for further procedures. For null values; the model sometimes does not give good accuracy.

TABLE I
CHANGES IN THE DATASETS AFTER DROPPING THE NULL VALUES

Dataset	Before check null values	After dropping null values
amazon preprocessed kindle book review	12000	12000
amazon women dresses reviews	23486	19662
apple iPhone	9713	9713
spotify reviews	9713	9713

After using the dropna() method to drop null values from the datasets, from Table I we could see that every dataset remained the same except “amazon women dresses reviews” which consist of 19662 after deleting null values.

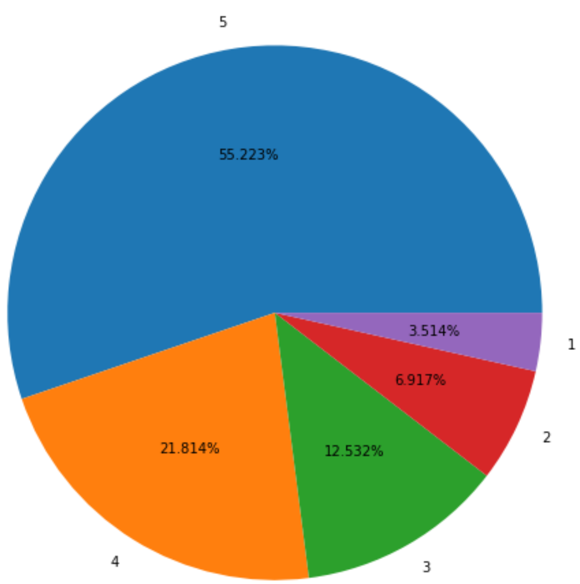

Fig. 2. Counting ratings as percentage

Counting of five ratings: For good service sellers put a rating system on which they could understand their service and product. People who use iPhone, kindle, Spotify and Women’s dress would like to give reviews through five stars rating system based on using them. Next step, we counted the rating values from 1 star to 5 stars individually out of 100%. We have also shown the pie chart in Figure 2 which consists of counting one to five stars out of 100% from the reviewers. For example, in Figure 2 we have shown just the dataset “amazon women dresses reviews”. The rest of the datasets have proceeded in the same way.

Behavior of datasets: We checked our datasets if they are biased,semi-biased, or unbiased. This is one of the main parts of our model and motivation to work in every type of dataset. Figure 3 makes anyone understand that it is biased and Table II also shows the behavior of all datasets we have taken for this model which are also given for the “amazon women dresses reviews” dataset only.

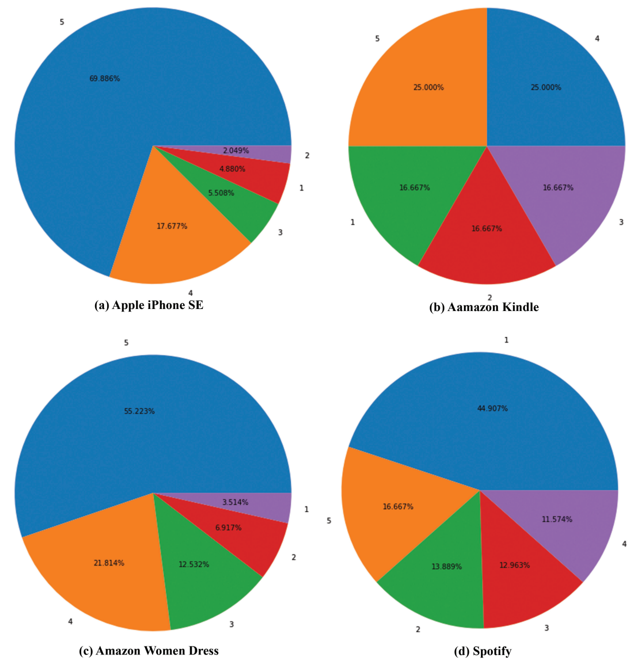

Fig. 3. Biasness of Datases

Figure 3 shows us the visual representation of the biasness of the datasets. Moreover, we have put the types of the biasnesses in Table II. Here, the Amazone kindle dataset looks so much even which usually can not be found in the real life. The reason behind this, may be this dataset is built unbiased intentionally.

TABLE II
TYPE OF DATASET

Dataset	Dataset type
amazon preprocessed kindle book review	Unbiased
amazon women dresses reviews	Biased
apple iPhone	Biased
spotify reviews	Unbiased

Table II also refers to the main idea/concept to use our models in different types of datasets to check the workflow rather than working on just one type of dataset.

Counting words and Average use of words in each review tag: From the Review attribute we have counted those words from sentences that are not necessary or counted as important words to understand the review. We have calculated the total words from the datasets. In Figure 4 we have shown the dataset "amazon women dresses reviews". Then we have calculated the average words for each tag in Table III based on ratings.

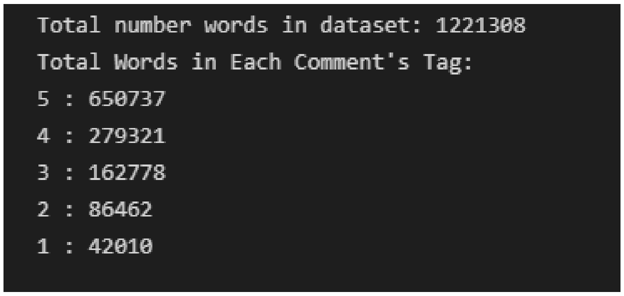

Fig. 4. Calculating total words in the dataset

TABLE III
AVERAGE WORDS IN EACH COMMENT'S TAG

Ratings	Average Words
5	60
4	65
3	66
2	64
1	61

For example, in the "amazon women dresses reviews" dataset a review is - "Love it; the bit of stretch in the denim makes it less stiff than traditional denim." From this sentence, words like - Loved, stretch, makes, less, stiff, than, traditional, denim have been counted. So, in Table III for five ratings people have given reviews that consist of 60 words on average after cutting out other unnecessary words which do not relate to emotions which we could see many count of words from Figure 4 . Moreover, from the rest of the reviews, the same words in a line we have also reduced to calculate the average word from each tag. Our goal is to shrink the sentence as much as we can in order to achieve success to get good accuracy. The model operates more readily with simpler inputs. We have also used tokenization, later on, to make it more feasible in the Sentiment Analysis.

Calculating word frequency: Under sentiment analysis, we have also calculated word frequency to avoid those words which are unnecessary or not related to showing emotions.

From Table IV we could see that those words are either prepositions or pronoun words that do not mean emotions. As we are working based on emotions, we ought to ignore these words. Because of this, we ignored these emotionless terms prior to the use of classifiers.

Comment length frequency:

We suggest a new family of algorithms that utilize relevant patterns in the input data to automatically discover patterns that may be used to enhance frequency predictions. The suggested algorithms combine machine learning's advantages with the formal guarantees made possible by algorithm theory

TABLE IV
AVERAGE WORDS IN EACH COMMENT'S TAG

Highest using Words	Counting word value
i	67392
and	44531
a	43431
it	38174
is	33411
to	21750
this	18556
in	18149
but	14338
on	12332
for	12197
of	12036
with	11345
was	10873
so	10207
my	9770
not	8297
that	8247
I	7913

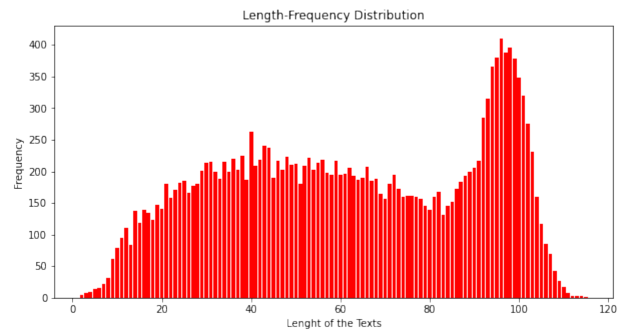

Fig. 5. Length of text based on frequency

[13]. In Figure 5 the graph shows that the maximum review length is 99 and the highest frequency is near 400. This plot shows the relation pattern between words and sentence length of reviews.

C. Sentiment Analysis

Sentiment analysis' main goal is to determine whether customers like or dislike a product based on the review data they have provided [13]. Sentiment analysis is text-centered, yet there are real difficulties in determining the precise sentiment of the statement. The stages of sentiment analysis Tokenization are included as well. A user can tokenize data by grouping various input text data into tokens. Individual words, keywords, numbers, and punctuation make up tokens. Punctuation and other special characters in the input data are removed during tokenization since they are ignored to increase analysis accuracy. Tokens serve as the input data for each step in the construction of the classification model.

D. Model Specification

Pretrained Bert Base: A transformer architectural paradigm called BERT (Bidirectional Encoder Representations from Transformers) is used for NLP. Its pre-training data comes from a sizable, unlabeled text corpus that includes both Wikipedia and a corpus of books. BERT is used in two different ways [14]. One is Masked Language Prediction, which involves making a small number of the input text's words before feeding them into the BERT model to predict the masked words. In order to forecast the masked word, the BERT model analyzes the context of the words that come before and after the masked word. Next Sentence Prediction is the next strategy. Advance model performance than with older techniques. a capacity for processing more text and language. A simple method for utilizing trained models (transfer learning). Metrics can be adjusted and used right away. Because the model is updated constantly, its accuracy is exceptional. This is possible with effective fine-tuning training.

Pretrained Distilbert: Distilled Bert is superior to BERT in terms of weight, price, size, and speed. DistilBERT employs better performance than the BERT base model [15]. This paradigm mitigates a few other BERT shortcomings, such as word piece embedding and fixed input length size issues, in addition to having low computational and resource requirements. With random initialization, this model uses a cosine embedding loss, masked language modeling loss, and student loss. Before training the model, the input text must be tokenized, transformed into token IDs, and padded. DistilBERT was chosen as one of the models to be used in the classifier because it has a minimum resource need, which is in line with the goal of this research. The use of large pre-trained models in the transfer learning approach has gained popularity recently. These large-scale models improve performance significantly, but they frequently have millions of parameters. Recent research on trained models indicates that training larger models can still produce better results for downstream tasks.

Pretrained Roberta: a robustly optimized approach that outperforms Google's self-supervised Bidirectional Encoder Representations from Transformers (BERT) approach for pre-training natural language processing (NLP) systems. BERT is a ground-breaking approach that, rather than using a language corpus that has been labeled particularly for a certain task, obtained state-of-the-art results on a variety of NLP tasks while depending on unannotated material taken from the internet. Since then, the method has gained popularity as a final task architecture as well as a baseline for NLP research. The collaborative character of AI research is further highlighted by BERT. Because of Google's open release, it was possible to replicate BERT's performance and identify areas for improvement. Modern results on the widely recognized NLP benchmark, General Language Understanding Evaluation, are produced by this optimized approach, RoBERTa (GLUE) [16].

During every model, we checked for the five datasets. After tokenization and showing the dataset types we used these models to find our final step of accuracy. We have tried to convert the reviews/comments to positive or negative signs. We have referred to them as numbers such as - Negative = 0 Neutral = 1 Positive = 2 The main reason for taking negative, positive, and neutral reviews as numbers is to fit the reviews of customers converting from characters to numeric numbers. Numeric numbers will help to find the final result of accuracy. Then with the converted training dataset, we also trained the dataset to find the accuracy.

IV. RESULT ANALYSIS

We collected the accuracy for all the three classifiers for all datasets and presented them in Table V. From the table, it can be seen that the accuracy depends largely both on the model used and on the datasets fed. Furthermore, it's also known that the datasets ”apple iPhone SE reviews” and ”amazon women dress” are highly biased towards the rating of 5. This can be seen from Figure 3. From the table V, it can be seen that the biased datasets are giving higher accuracy than the unbiased data. This can be caused from various factors like, the pre-tained model can be also biased, or, the dataset had enough of each type of input to judge correctly, or, the nature of the dataset that was used as a test. This can be seen strongly in the dataset of ”spotify”. Because it's the least biased, all the accuracy found related to this data was more comparable. As a result, it can be said that the pre-tained models were trained with the biased data. Again it can be proven if we notice to the result of the ”amazon kindle book review”. This dataset is accurately unbiased and as a result, the accuracy of the model with this dataset is the lowest compared to others.

TABLE V
FINDING ACCURACY USING CLASSIFIERS

Models	Datasets	Accuracy (%)
BERT	amazon preprocessed kindle book review	60
BERT	amazon women dresses reviews	80
BERT	apple iPhone SE reviews	88
BERT	spotify reviews	72
DISTILBERT	amazon preprocessed kindle book review	73
DISTILBERT	amazon women dresses reviews	80
DISTILBERT	apple iPhone SE reviews	79
DISTILBERT	spotify reviews	76
RoBERTa	amazon preprocessed kindle book review	71
RoBERTa	amazon women dresses reviews	83
RoBERTa	apple iPhone SE reviews	90
RoBERTa	spotify reviews	79

Biased dataset works better for the classifiers comparable to unbiased and Semi- biased datasets. For biased dataset all classfiers crossed 79% accuracy. From Table V RoBERTa gives the best accuracy of 90% and even the BERT classifier also giving the second best accuracy of 88%.

So, from Figure 3 table II and Table V we could go through our motivation we wanted to do in our whole project to find out the behaviour of the models based on the biased and unbiased datasets. Before entering to our result EDA preprocess helpes us to fit the datasets in such a pattern that we were able remove all those unnecessary words which are not needed or applicable to work this model.

V. CONCLUSION

To conclude Future customers' purchasing decisions are influenced by the reviews posted on internet platforms, and this has a financial impact on firms. Recognizing phony reviews is essential because spam reviews that are generated with a specific goal might mislead customers. Systems for detecting false reviews have been built by researchers using machine learning and neural networks in a rigorous manner. However, the execution of feature engineering for conventional machine learning algorithms calls for domain expertise and huge labeled datasets. To create a review classifier, we used transfer learning and the transformer-based pre-trained models BERT, RoBERTa, and DistilBERT. In order to build a classifier that can identify fraudulent reviews broadly while utilizing the fewest amount of computational resources possible, these models are trained using 10% and 50% of the Yelp dataset. SA is carried out after pre-processing (EDA) to analyze and assess datasets. The pre-trained models are then adjusted for both data samples. Accuracy and assessment metrics are taken into consideration while evaluating the performance of all models. RoBERTa attains a 69% accuracy rate.

However, the datasets we have got from online it seems that there are few data which are previously modified. The data are not authentic enough to make this model perfect. In the result, our classifiers performed with under 60% accuracy because of modifying data.

Nowadays, unsupervised machine learning is more useful because clustering helps to remove unnecessary data from a dataset. Mean or median is also another process to handle those unnecessary nun values which are more effective rather than removing null values from the dataset. We want to use more classifiers like - ALBERT and ELECTRA which might give us better results. Our next primary goal will be to build a predictor model which will help organizations to make a decision about making or expanding products or not in their business based on a few people's reviews.

REFERENCES

[1] S. Gupta, 'Sentiment analysis: Concept, analysis and applications', Towards Data Science, 07-Jan-2018. [Online]. Available: https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17. [Accessed: 08-Dec-2022].

[2] Nadkarni, P. M., Ohno-Machado, L., Chapman, W. W. (2011). Natural language processing: an introduction. Journal of the American Medical Informatics Association, 18(5), 544-551.

[3] De Kok, S., Punt, L., Van Den Puttelaar, R., Ranta, K., Schouten, K., Frasincar, F.: Reviewlevel aspect-based sentiment analysis using an ontology. Proc. ACM Symp. Appl. Comput. 315–322 (2018). https://doi.org/10.1145/3167132.3167163.

[4] Bansal, B., Srivastava, S.: On predicting elections with hybrid topic based sentiment analysis of tweets. Procedia Comput. Sci. 135, 346–353 (2018).https://doi.org/10.1016/j.procs.2018.08.183.

[5] Hasan, A., Moin, S., Karim, A., Shamshirband, S.: Machine Learning-Based Sentiment Analysis for Twitter Accounts. Math. Comput. Appl. 23, 11 (2018). https://doi.org/10.3390/mca23010011.

[6] K. Ravi and V. Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, KnowledgeBased Systems 89 (2015), 14–46.

[7] W. Medhat, A. Hassan and H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams Engineering Journal 5(4) (2014), 1093–1113

[8] Akhtar, N., Zubair, N., Kumar, A., Ahmad, T.: Aspect based Sentiment Oriented Summarization of Hotel Reviews. Procedia Comput. Sci. 115, 563–571 (2017). https://doi.org/10.1016/j.procs.2017.09.115.

[9] Amazon reviews on Women dresses (23K Datapoints). (2022, June 14). Kaggle. https://www.kaggle.com/datasets/surajjha101/myntra-reviews-on-women-dresses-comprehensive

[10] Apple iPhone SE reviews ratings. (2021, September 16). Kaggle. https://www.kaggle.com/datasets/kmldas/apple-iphone-se-reviews-ratings

[11] Spotify App Reviews. (2022, July 10). Kaggle. https://www.kaggle.com/datasets/mfaaris/spotify-app-reviews-2022

[12] 1.1.1. What is EDA? (n.d.). https://www.itl.nist.gov/div898/handbook/eda/section1/eda11.htm

[13] Hsu, C. Y., Indyk, P., Katabi, D., Vakilian, A. (2019, January). Learning-Based Frequency Estimation Algorithms. In International Conference on Learning Representations.

[14] Geetha, M. P., Renuka, D. K. (2021). Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. International Journal of Intelligent Networks, 2, 64-69. https://www.sciencedirect.com/science/article/pii/S2666603021000129

[15] Gupta, P., Gandhi, S., Chakravarthi, B. R. (2021, December). Leveraging transfer learning techniques-bert, roberta, albert and distilbert for fake review detection. In Forum for Information Retrieval Evaluation (pp. 75-82).

[16] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., amp; Stoyanov, V. (2019, July 26). Roberta: A robustly optimized Bert pretraining approach. arXiv.org. Retrieved December 16, 2022, from https://arxiv.org/abs/1907.11692