PROJECT REPORT - ACM WORKSHOP 2022

Analyzing Company related complaint tweets using ABSA to determine its shortcomings

Purvi Chaurasia Nikita Shrivastava

Under guidance of - Prof. Rishabh Kaushal, Prof. Nisha Rathee, Prof. Kalpana Yadav and Prof. Ankita

1. Abstract

With the expanse of social media expanding day by day, social media is becoming the go-to place for customers to review their experiences with customer based companies. More so, FMCG and online delivery companies are becoming more and more aware of the power of social media in maintaining their company's image and most importantly the ability of social media to reflect the common grievances of customers on which the company might want to work on. Thus, this project aims to provide an analytical and easily understandable method to provide companies and even customers the common grievances and sentiments expressed on social media with respect to various aspects of the company. The project uses tools of sentiment analysis to analyze the general tone of customers on social media when reviewing the services and products of a given company. Further, the project uses ABSA to specifically analyze the aspects upon which customers have specifically reviewed negatively on, upon which the company needs to improve, and the aspects the customers have reviewed positively or neutrally on, on which the company can build further specialization and a user-base. This project has made use of LSA based approaches and compared it with traditional sentiment analysis approaches like BERT.

2. Introduction

With the growing number of companies catering to the needs of the public, it has become the need of the hour to come up with more refined approaches towards customer review analysis. In order for a company to grow, there needs to be a positive and interactive customer-company relationship which can only exist if organizations are responsive to the needs of their customers. If such analysis were to exist, organizations can be held accountable for their shortcomings and can be expected to perform better.

In today's world, information through mediums like Facebook, Instagram, Twitter and other social media platforms has become the key resource for organizations to analyze customer opinions, reviews and complaints. Studies have shown that Twitter is the main platform for companies to interact with customers and their grievances. [8] Using this as the basis, we have used our LSA model to classify and analyze customer grievances of Myntra and Flipkart by collecting tweets via their twitter handles. Using the Local Semantic Aggregation (LSA) model [1] can help fine-tune the analysis process by giving organizations targeted clusters of their shortcomings as well as achievements.

Aspect based sentiment analysis (ABSA) is a more refined approach towards the traditional sentiment analysis. However sentiment analysis fails to take into account the inter-relational polarities within a sentence. For example, if we have the sentence -

"The delivery of food was very bad but the taste was good."

This sentence will give a negative polarity in traditional sentiment analysis models. However, as we can see, the sentence is neither positive nor negative, rather it depends on the 'aspect' that we are looking at. In this scenario, the sentence with respect to food is positive and with delivery it is negative. Therefore in order to successfully identify and extract aspects with their respective sentiments we use the LSA+DeBERTa model.

Initially, sentiment analysis was done using the BERT-BASE models [12], however due to low training and higher requirement of computational resources, more and more sophisticated models have been developed. The ABSA model that we have used in our project functions on the principle that sentiment dependency usually exists in clusters. Taking this as the basis, it covers two major concepts of '*Sentiment Coherency*' and *'Sentiment Clustering'* [1] which have been explained in detail later.

In order to justify using this particular model we have also compared the results of our ABSA model with the traditional BERT model. This is done using the infamous 4 datasets - Laptop 14 and Restaurant14, Restaurant15 from [9] and Restaurant16 from [10]. We chose these particular datasets for comparisons as they have been used for ABSA models in the past as well and would fit both models satisfactorily.

Following the comparative analysis of LSA and BERT we moved to the main aim of our project. We collected tweets about Myntra and Flipkart, keeping the size of the dataset varied in order to see any changes that it might have on our model's functioning.

On careful observation, we found that a larger dataset, like that of Myntra's (2500 tweets) needed K-means clustering to be performed on its final outcome. This was because similar meaning aspects were

not grouped together and hence their net impact could not be judged without clubbing them together and viewing them as one.
For example - In a large dataset there were many variations of ‘delivery’, such as - ‘delivered’, ‘deliver’, ‘delivering’. All these aspects point to the same underlying meaning of having a problem with the delivery and hence need to be treated as one.

K-means clustering helped us find the unique aspects which had a negative sentiment towards the company. We then clubbed the frequencies of each of these aspects to find the contribution of each shortcoming to the company’s portfolio. This was done using simple string and dictionary operations.

Finally, in order to provide a conclusive and easily understandable view to the users, we have presented our findings for both the companies in the form of a pie chart. The pie chart clearly shows the impact that each aspect/shortcoming has, as derived from the tweets of the company.

3. Related Works

Citation No.	Year	Title and Publication	Core Idea of the Paper	Approaches Used	Experimental Data
[1]	2021	Yang, Heng, et al. "Back to Reality: Leveraging Pattern-driven Modeling to Enable Affordable Sentiment Dependency Learning. " - *arXiv preprint arXiv:2110.08604 (2021).*	Defines the adjacent aspects with the similar sentiment as the sentiment groups in a sentence, found that most sentiment dependency occurs between neighboring aspects. Proposed the sentiment patterns (SP) to guide the model dependency learning. Introduced the LSA procedure to focus on learning the sentiment dependency in the sentiment group.	LSA+ deBERTa	Laptop 14, Restaurant 14 datasets from SemEval-2014 Task 4 (Pontiki et al. 2014), Restaurant 15, Restaurant 16 datasets from SemEval-2015 task12(Pontiki et al. 2015) and SemEval-2016 task5(Pontiki et al. 2016)
[2]	2019	Liu, Yinhan, et al. "Roberta: A robustly optimized BERT pretraining approach." - *arXiv preprint arXiv:1907.11692* (2019).	Achieved state-of-the-art results on GLUE, RACE and SQuAD, a set of important BERT design choices and training strategies to increase task performance,	BERT, Dynamic masking, Model Input Format and Next Sentence Prediction	BOOKCORPUS (Zhu et al., 2015) plus English WIKIPEDIA, CC-NEWS, which we collected from the English portion of the CommonCrawl News dataset, OPENWEBTEXT (Gokaslan and Cohen, 2019), an open-source recreation of the WebText, STORIES, a dataset introduced in Trinh and Le (2018) containing a subset of CommonCrawl data
[3]	2016	Wang, Yequan, et al. "Attention-based LSTM for aspect-level sentiment classification." - *Proceedings of the 2016 conference on*	Attention-based LSTM for Aspect-level Sentiment Classification	Attention-based Long Short-Term memory,	SemEval 2014 Task 4 2 (Pontiki et al., 2014)

		empirical methods in natural language processing. 2016.			
[4]	2019	Song, Youwei, et al. "Attentional encoder network for targeted sentiment classification." - *arXiv preprint arXiv:1902.09314* (2019).	Designed an attentional encoder network to draw the hidden states and semantic interactions between target and context words, Use pre trained BERT to enhance model performance	Attentional Encoder Network (AEN), BERT	SemEval 2014 Task 4 2 (Pontiki et al., 2014) dataset, ACL 14 Twitter dataset gathered by Dong et al. (2014).
[5]	2021	Dai, Junqi, et al. "Does syntax matter? a strong baseline for aspect-based sentiment analysis with roberta." - *arXiv preprint arXiv:2104.04986* (2021).	Showed that induced tree from fine tuned RoBERTa (FT-RoBERTa) outperforms the parser-provided tree, three tree-based ALSC models	Inducing Tree Structure from PTMs, roBERTa, Perturbed Masking, ALSC	Six benchmark datasets. Three of them, namely, Rest14, Laptop14, and Twitter, are English datasets. Rest14 and Laptop14 are from SemEval 2014 task 4 (Pontiki et al., 2014), containing sentiment reviews from restaurant and laptop domains. Twitter is from Dong et al. (2014)
[6]	2017	Vaswani, Ashish, et al. "Attention is all you need." - *Advances in neural information processing systems* 30 (2017).	Proposed the neural architecture for Transformers, Encoder, Decoder and Attention	Positional Encoding, BERT	WMT 2014 English-German dataset
[7]	2017	Bing, Peng Chen Zhongqian Sun Lidong, and Wei Yang. "Recurrent Attention Network on Memory for Aspect Sentiment Analysis."	Proposed neural network architecture to find opinion targeted sentiment	Multi Attention mechanism, LSTM	Two are from SemEval2014, i.e. reviews of restaurants and laptops; a twitter dataset, for testing its performance on social media data; and a Chinese news comment dataset, for testing its language sensitivity
[8]	2011	Burton, Suzan, and Alena Soboleva. "Interactive or	Analyzed and compared use of Twitter in 12 accounts held by six	Consumer Marketing	

		reactive? Marketing with Twitter." *Journal of consumer marketing* (2011).	organizations in the USA and Australia, drawing on existing models of interactive communications		
[9]	2015	Pontiki, Maria, et al. "Semeval-2015 task 12: Aspect based sentiment analysis." *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*. 2015.	Provided manually annotated reviews in three domains (restaurants, laptops and hotels)	Dataset collection and tagging for ABSA	
[10]	2016	Pontiki, Maria, et al. "Semeval-2016 task 5: Aspect based sentiment analysis." *International workshop on semantic evaluation*. 2016.	Provided manually annotated reviews in three domains (restaurants, laptops and hotels)	Dataset tagging and labeling for ABSA	
[11]	2018	Blohm, Matthias, et al. "Comparing attention-based convolutional and recurrent neural networks: Success and limitations in machine reading comprehension." *arXiv preprint arXiv:1808.08744* (2018).	Layed out the behavioral differences between convolutional and recurrent neural networks, with the limitations of each	CNN, RNN, Hierarchical Attention-based Compare-Aggregate Model	MovieQA dataset
[12]	2019	Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).	Laid the groundwork for BERT encoders		BooksCorpus (800M words) (Zhu et al., 2015) and English Wikipedia (2,500M words)

4. Dataset

4.1. Dataset Collection for tweets of Myntra and Flipkart

Dataset collection is a key part of the training of our models as it affects the overall quality and accuracy of the results delivered. The Datasets of this particular project have been largely divided into 3 portions.
The first is the Dataset collected for well known large scale Indian FMCG companies that was used to test the model in its deployment incorporating both roBERTa [2] and LSA [1].
The second is the dataset used to train roBERTa [2], the sentiment analysis machine learning model of our choice. This dataset is key in providing us a model with a wide and diverse scope in identifying different sentiments from the text at hand.
And the third and final dataset is the one used to train the Local Sentiment Aggregating model. This dataset is used to train the model to help identify the different aspects of a given product/service/company which might be present in a given sentence and the sentiment expressed regarding them.

While collecting the data with respect to the FMCG companies, we had to keep in mind certain key parameters which had to be considered in order to produce a dataset representative of the grievances expressed by the customers availing services from these companies. These key parameters were uniquely identified for each company, consciously, and were used to extract data from social media platforms like Twitter. As per various studies on the nature of the use of Twitter by various companies [8], most companies use the platform to interact with customers and address their grievances and complaints. With the growing use of twitter among the digitally equipped population of India, a large percentage of today's global market, we decided to use the social media platform as the primary source for the data extraction regarding the customer complaints of a specific company as it most accurately represents the modern day internet-based diaspora of the country.
Thus, we collected 2500 tweets from twitter using the Twitter module of the Social Network Scrape (snscrape) library in python, while using popular parameters commonly identified in customer complaints for said companies to filter through the complaints using the advanced search feature of the website.
The same is used to collect 500 tweets for Flipkart as well. The varied sizes are purposefully chosen in order to observe the impact of a smaller dataset on the model and its findings.

4.2. Preprocessing of Data

Upon obtaining the required tweets, we preprocess the tweets by removing unnecessary characters such as @, #, $,% etc. ,dropping duplicates and single word tweets and tokenizing the text in the tweets, we obtain a dataframe containing the clean data which will be used for the testing and training in the future.

4.3 Training and Testing Data

In the past, the BERT [6] dataset was mostly used to perform NLP related tasks such as sentiment analysis and text classification. However, the BERT [6] model was later found to be severely undertrained. Thus, roBERTa [2] came to be, a retraining of BERT [6] with improved training procedure and more data and computing power. The roBERTa [2] model is a transformers model pre-trained on a large corpus of English data in a self- supervised fashion. This implies that the training dataset used to train the roBERTa [2] model is made of raw texts only, with no human labeling them in any way. The model was additionally pre-trained on the Masked language modeling in which 15% of the input data sentence is masked and fed into the model and the model is trained such as to predict the masked words as the output. The primary datasets, however, that have been cited to have been used in the pre-training of the model, are BookCorpus and text from the English Wikipedia. Thus, roBERTa is trained to have 4-5% higher training time than BERT and to be 2-20% higher in efficiency than its predecessor as well.

The LSA model [1] is used to deduce the sentiment dependance in a given sentiment cluster while identifying the relevant aspects present in the sentence. The model used 4 public datasets for pre-training which are divided into sentiment cluster and sentiment classifications to provide the best outputs and predictions on the nature of the sentiment dependency related to the aspect in a given text. The 4 public datasets used for the pre-training of the model are the Laptop 14, Restaurant 14 from the SemEval-2014 Task 4 [9] , the Restaurant 15 [9] and Restaurant 16 [10] dataset. These datasets are key for comprehensively evaluating the performance of the LSA process that we are using in this project.

	Positive		Negetive		Neutral	
Datasets	Train	Test	Train	Test	Train	Test
Laptop 14	994	341	870	128	464	169
Rest14	2164	728	807	196	637	196
Rest15	909	326	256	180	36	34
Rest16	1240	468	437	117	69	30

Statistics of the 4 datasets with respect to the different aspects [1]

5. Proposed Approach

5.1. Flowchart

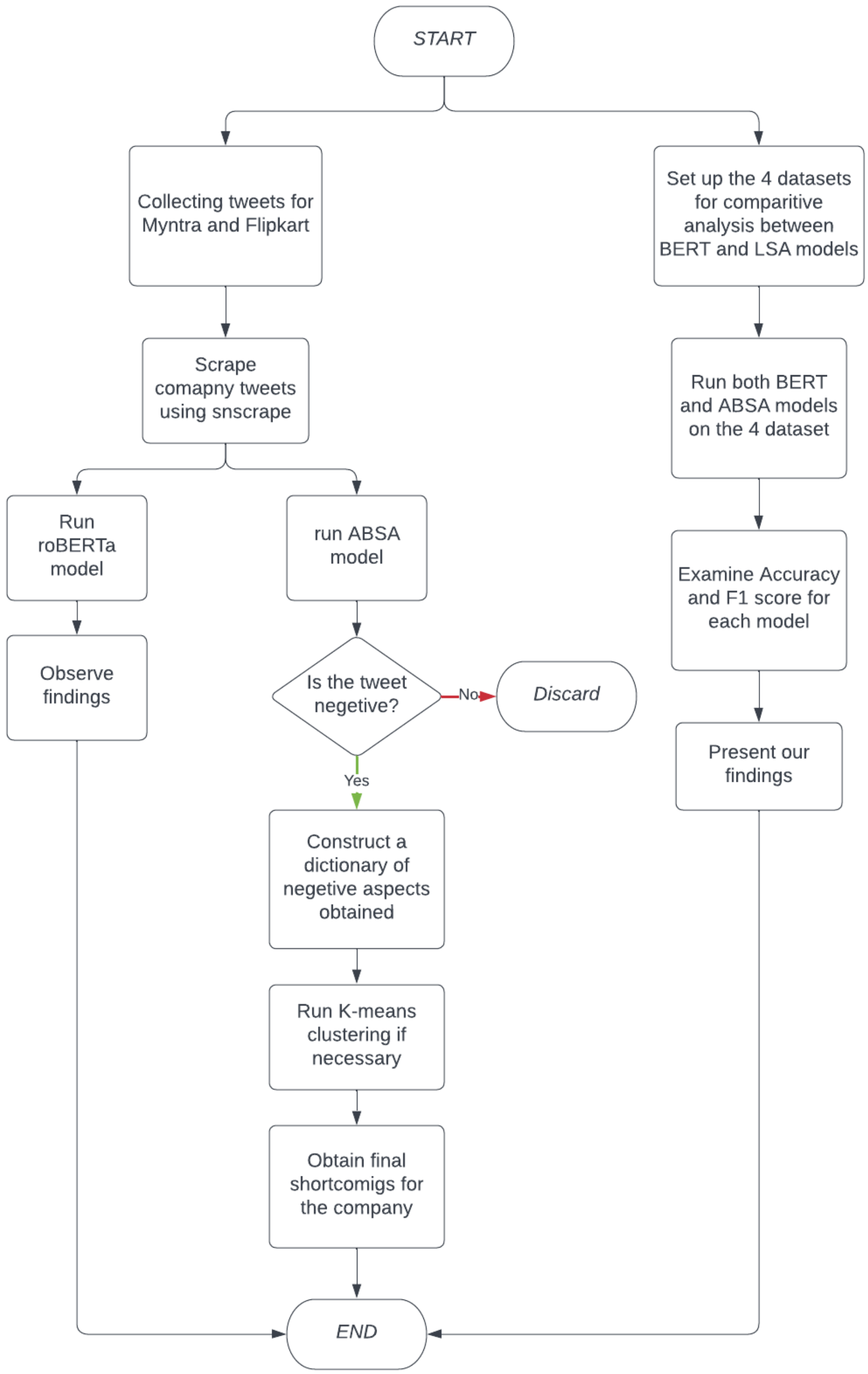

Fig 4.1. Flowchart explaining the working of the entire model using BERT and roBERTa

5.2. roBERTa for Sentiment Analysis

Why did we use roBERTa instead of BERT ?

There is a simple explanation. Both models are derived from the transformers neural architecture. The transformers model was a groundbreaking discovery because it transformed the way we approach and model NLP models. Before transformers, NLP was driven by 'Recurrent Neural Networks' (RNN) [7]. Although effective, RNNs were inefficient in processing long paragraphs and large datasets [7, 11] as it functioned on sequential processing of words in order to capture their order. Transformers, on the other hand, function on parallel processing of words making it an extremely efficient and fast model to train even for larger datasets. This change was brought due to 2 individual concepts -

1. Positional Encoding [6]
   In RNN sequential processing was done to retrieve the original order of words. Positional encoding striked the need for it by storing information about the order in the data itself rather than structure of the network. This enabled it to run on larger datasets and hence the model was trained to interpret these encodings.

2. Self Attention [6]
   This is the key concept that made transformers an efficient network for other models. It helped understand the individual meaning of words in context to the words around it. This helped decipher the positional encodings with respect to each word.

Though the architecture of BERT was state-of-the-art, its later usage found that its model was severely undertrained. While BERT was trained on 16 GB, roBERTa is trained on 160 GB data [2] making it a supercedingly more effective model. The underlying framework for both is the same however due to a larger training dataset, roBERTa is practically more efficient.
Hence the roBERTa model was used to run a sentiment analysis model on our company tweets dataset. Its findings revealed something crucial. Though it enabled us to look at a generic view that people have towards a company, it didn't give us the full picture. We still didn't have information regarding 'why' the company is facing backlash on its customer front. Therefore now our need was to not only gather general sentiment but analyze the reason behind that sentiment.

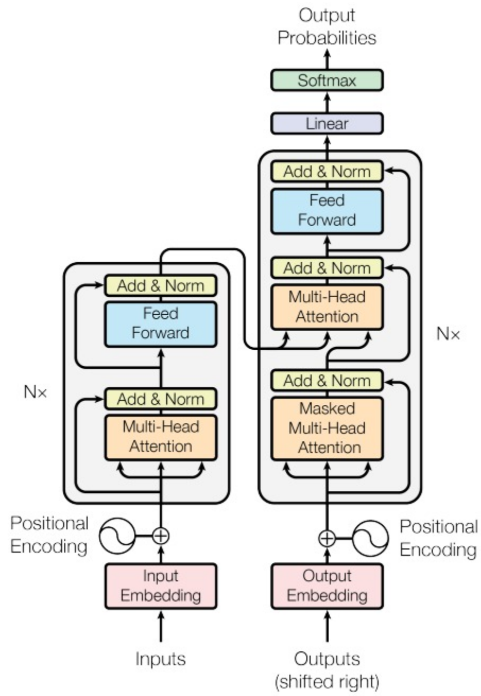

Fig 5.2.1. The Transformer Model Architecture [6]

5.3. Aspect Based Sentiment Analysis

The ABSA model is based on two key output parameters, the Aspect and the Polarity associated with the aspect. It has been found that in the ABSC (Aspect Based Sentiment Classification), there are not only different aspects with varying sentiment duality, but also, that these sentiment polarities between each aspect may or may not be dependent on each other and even still, they might be contradictory as well.

This observation thus forces the model to consider the dependence between various aspects instead of conventionally considering the aspects and their polarities as independent units.

For eg. *'The Food and delivery was good but the app was extremely hard to use, the quality was very good as well.'*

In the above sentence, the aspects Food and delivery's sentiment polarity are dependent on each other and cannot be mutually excluded from each other. Thus we can say that the both of them possess a local coherency between each other. However, the aspect app's sentiment polarity does not hold a dependency to any other aspect and can be taken as an independent unit.

Thus, in order to accommodate this observation, the ABSA model proposes the adoption of two techniques, namely the Sentiment Clustering (**SP1**) [1] and Sentiment Coherency (**SP2**) [1]. Users tend to group aspects with similar polarities together. In the example given above, the aspects 'Food' and

‘Delivery’ have a positive sentiment and are hence grouped together while the app has negative polarity associated with it and is written later. This is termed as ‘Sentiment Clustering’ [1] and it can be used to assign common polarities to grouped aspects. While aspects with similar polarity can exist in the same sentiment cluster, it can also exist outside of it. Referring to the same example above, ‘Food’ and ‘Delivery’ have a positive polarity but so does ‘Quality’. Therefore we can say. ‘Food’ and ‘Delivery’ have local coherency [1] while all three of them (Food, Delivery and Quality) have a global coherency [1]. Using both these techniques covers all correlations that can exist within these aspects.

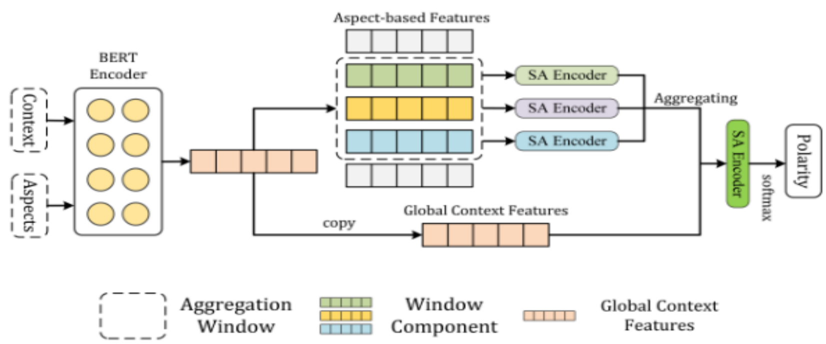

Fig 5.3.1. Main Framework for Local Sentiment Aggregation Mechanism [1]

The aggregation window for ABSA is dependent on the usage of the Aspect-Oriented Context Features. An Aggregation window, to put it simply, is a set of functions that perform a calculation over a given record of data which is called a window [1]. It is primarily constructed by using ELCF features of aspects next to each other i.e, the kth aspect to the left and right of the aspect under consideration, in our model, the hyperparameter k=1 has been taken into account citing the relatively less complex natures of the sentiments and general independence in the Sentiment duality of the respective aspect. As the model is also taking sentiment coherency into account, Local Context is a key feature of this parameter. The calculation of local context can be divided into the token distance-based method (**LSA-T**) [1] and syntax distance-based method(**LSA-S**) [1].

Padding is essential as Neural Networks require input data to be of the same shape and size. Therefore, before processing the data into our aggregation window we need to ‘fit’ them into a certain frame for the model to run effectively. Thus, in ABSA the padding of the aggregation window is done using the aspect related features. The huge advantage of ABSA is that while padding the model with the aspect based features, the model does not get degenerated (as would be the case otherwise due to the change during the normalization of the data during padding) as the padded components have the same sentiment data which maintain SP1 and SP2 while modeling the sentiment groups.

Finally in the output layer, the global context feature and the features learned from local sentiment aggregation are combined due to the loss of the context feature during the ELCF [1] feature extraction to predict the sentiment polarities.

5.4. Deployment for tweets of Myntra and Flipkart

Once we had trained and tested the ABSA and roBERTa model, it was time to implement it on the dataset of tweets that we had initially collected for Myntra and Flipkart.

We started out with implementing the ABSA model on Myntra's tweet dataset. This dataset had around 2500 tweets and upon running the model we came across various aspects. However, due to the large dataset, we had certain aspects that had the same underlying meaning but differ structurally and hence were counted as separate. This wasn't giving us a clear overview of the sectors that Myntra users were facing problems in. We therefore decided to run K-means clustering on the obtained data in order to group those aspects together which had similar basis. This gave us a clearer view of the shortcomings of Myntra and the results are covered in the next section.

Flipkart's dataset consisted of 500 tweets out of which around 250 tweets were negative in nature and were the ones majorly considered. We wanted to study the impact of a smaller dataset and determine if smaller datasets can also be used for our project. However, the result was not as elaborate as with the dataset of Myntra. It also didn't require us to run K-means clustering on the obtained data as the aspects were unique and non-repetitive in nature.

Therefore in this way we were able to achieve 2 things - Do a comparative analysis between LSA and BERT based models and Use the ABSA model along with clustering in order to determine the shortcomings of various companies through the tweets.

9. Experimental Design and Results

9.1. Comparative Analysis

The aspect based sentiment analysis model has been used along with clustering to detect shortcomings in a company based on the tweets made about it. Before moving forward to its implementation we drew a comparative analysis between LSA based approach and BERT based approach to determine the model with a better accuracy and F1 score.

In order to do this, we've used 4 datasets - the Laptop 14 and Restaurant 14 datasets from SemEval-2014 Task4(Pontiki et al. 2014), the Restaurant 15, Restaurant 16 datasets from SemEval-2015 task12 (Pontiki et al. 2015) and SemEval- 2016 task 5 (Pontiki et al. 2016), respectively.

Here are the statistics for the datasets that we have used for determining the accuracy of the ABSA model in comparison with the BERT-BASE approach.

	Positive		Negetive		Neutral	
Datasets	Train	Test	Train	Test	Train	Test
Laptop 14	994	341	870	128	464	169
Rest14	2164	728	807	196	637	196
Rest15	909	326	256	180	36	34
Rest16	1240	468	437	117	69	30

Fig 9.1.1. Distribution of Positive and Negative tweets across all 4 datasets [1]

Both the models have been initially trained for 5 rounds each and their average performance has been used as an indicator of their accuracy. BERT-BASE [12] is the model that laid the groundwork for the initial working of BERT.
On running the all 4 datasets on BERT-BASE and LSA our findings were as follows:

Model	**Laptop 14**		**Rest 14**		**Rest 15**		**Rest 16**	
	Acc	*F1*	*Acc*	*F1*	*Acc*	*F1*	*Acc*	*F1*
BERT-BASE	79.73	75.5	82.74	73.73	82.16	64.96	89.43	74.2
LSA	**81.35**	**78.35**	**87.14**	**81.04**	**84.81**	**72.21**	**92.20**	**79.50**

Fig 9.1.2. Accuracy and F1 scores for BERT and LSA models across the 4 datasets [1]

From this comparative analysis we can deduce that LSA performs with greater accuracy than BERT in all 4 datasets making it a better model to prefer for aspect-sentiment detection.
Compared to BERT, the LSA based model outperforms it by an accuracy of approximately 2% across all 4 datasets.

9.2. Implementation for Myntra tweets

Initially 2500 tweets were collected for the analysis of Myntra. We have used the ABSA model in order to get aspects with their respective sentiments from the tweets. However we were not entirely satisfied by its outcome alone as this didn't allow us to gauge how much of an impact each of these aspects were having.

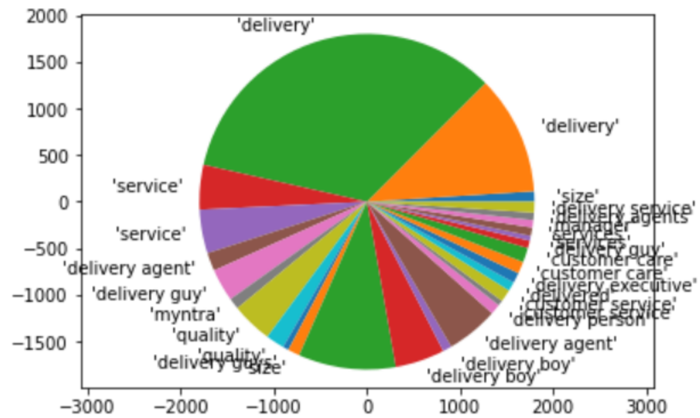

Fig 9.2.1. Pie-chart showcasing various shortcomings of Myntra without K-means

For eg- Since tweets are free handed, there's no uniformity in their expression. Therefore, 'deliver', 'delivery', 'delivered', etc. would be considered as separate entities by ABSA. This posed a problem as all these aspects are pointing to a common shortcoming.

In order to group these similar meaning aspects together, we used the K-Means Clustering algorithm and added the frequency of each to the shortcoming cluster they belonged to.

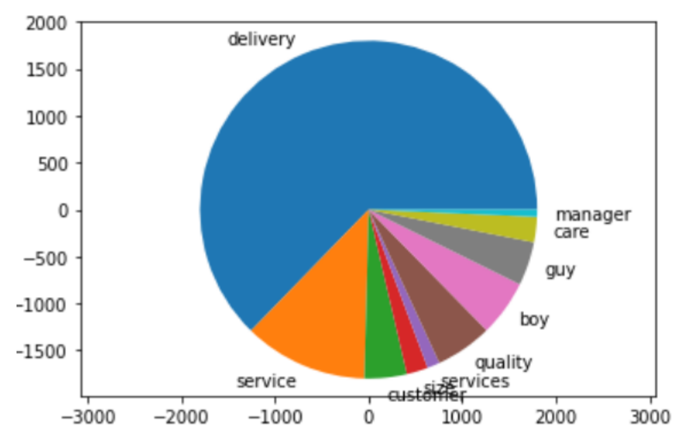

Fig 9.2.2. Pie-chart showcasing impact of various shortcoming after K-means

As we can observe, clustering gives us a much clearer view as to what areas Myntra is lacking in. The key shortcomings of myntra are - Delivery with 1660 complaints, followed by service with 315 complaint tweets.

9.3. Implementation for Flipkart tweets

We wanted to examine the impact of a smaller set of tweets about a company and see what outcome we will get. Therefore for flipkart we collected 500 tweets as an experimental approach.

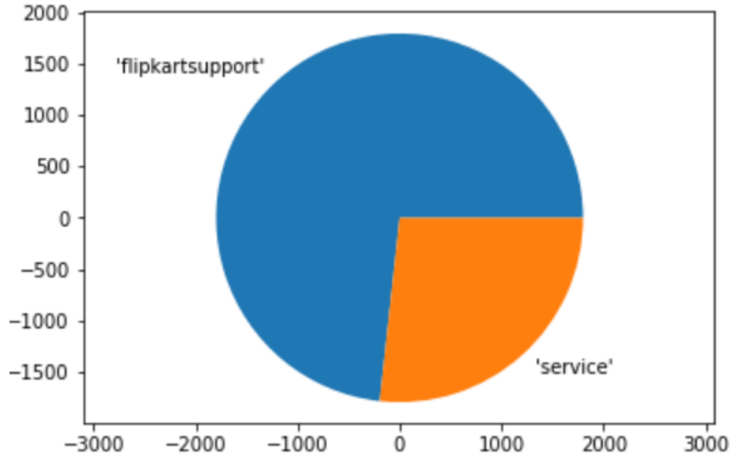

Fig 9.3.1. Pie-chart showcasing impact of various shortcomings for Flipkart

The number of pre-dominant aspects for a small dataset were minimal - limited to only 'flipkart support' which had 74 complaint tweets and the 'service' having 27 complaint tweets. The rest tweets were neutral in nature and are hence not accounted for in this pie chart.

K-means was not required for this particular dataset because of the uniqueness of each aspect and the frequency associated with it.

Therefore from this we can conclude that, in order to get a deeper understanding of the company's shortcomings through its tweets we need to provide a large dataset, so that a larger number of unique aspects can be identified. However in large datasets, we would also have to run k-means in order to group similar aspects and sum up their frequencies for a more conclusive analysis.

10. Limitations

While the results of this project do showcase the ability and effectiveness of the work done in recognizing the general sentiment reflected by the customers on social media platforms and the aspects on which negative sentiment have been reflected, there are certain limitations that have been taken into consideration as follows:

1. While ABSA is considered to be the most effective model in generating aspect based sentiment analysis with great accuracy, there are certain limitations that were observed in the model during its execution in this project.

    a. In ABSA, the local context features and the global context features are extracted from the input and the local context features are fed into the aggregation window. In the output layer, the global context feature is superimposed onto the final output of the aggregation window and the result is generated. This superimposition of the global context often leads to misleading results as the global context isn't taken into consideration in the aggregation window.

    Eg. The food was bad but the desserts were good, overall a nice experience at this restaurant.

    Here, the food will be an aspect with negative sentiment and desserts will be an aspect with positive sentiment but "nice experience" does not represent the global positive sentiment as per the model.

    b. In the ABSA model, when measuring the sentiment correlated to a given aspect, k=1 i.e only the immediate letter left and right to the aspect is taken into consideration. This value cannot be changed as it is a property of the model. Thus, if the k value was slightly larger, it would add more context to the sentiment analysis of the model.

2. After performing exploratory data analysis on the results obtained by ABSA, it was observed that the aspects with similar meaning and even spellings were considered different by the model.

Eg. "Delivery" and "delivery " were considered to be different aspects
"Delivery boy" and "Delivery agent" were considered to be different aspects

While this problem has been attempted to be solved using clustering, words which are synonyms but have vastly different spellings such as "online portal" and "user interface" still are not grouped in the same cluster which results in certain data points being missed from consideration.

11. Conclusion and Future Work

Through the research and analysis done in this project we deduce that the customer experience and grievances can be objectively quantified into measurable values that can hold meaning to not just the company in understanding the services and products at which they lack in or need improvement in but also to fellow customers who wish to make informed decisions regarding the products and services they wish to invest in or buy in the future. The aspect analysis portion of this paper also provides a fresh glimpse into the use of sentiments towards services and products in the FMCG sector and how customer reviews or complaints can be used to benefit the companies into improving their customer services.

While the models used in this project are highly robust and maintain a great degree of accuracy, they do pose a few minute limitations.
For eg, dismay or disappointment may not always be presented with the usage of negative words such as ‘bad’, ‘worst’, ‘horrible’ etc. They may sometimes also be expressed in the imperative such as ‘I want my money back’, or in a satirical form such as, ‘The food was so good that even my dog didn’t want to eat it’. Such forms of complaints and reviews are not always identified accurately and sometimes even go unreported due to the relative lack of conventional word formations which these models have been trained in.

While the scope of this project has been limited to the conventional analysis of sentiments as has been done in the past, it has come to our notice during the collection of data from twitter, that the newer generations often represent their dismay and even admiration in various forms of satires and other digital mediums such as memes, videos , gifs etc. Thus, in order to broaden the scope of this project, in the future we look to do more work in the field of satire selection and the direct impact and usage of satire and other digital mediums in the customer reviews and complaints on twitter.

12. References

[1] *Yang, Heng, et al., 2021, "Back to Reality: Leveraging Pattern-driven Modeling to Enable Affordable Sentiment Dependency Learning. "*

[2] *Liu, Yinhan, et al., 2019, "Roberta: A robustly optimized BERT pretraining approach."*

[3] *Wang, Yequan, et al., 2016, "Attention-based LSTM for aspect-level sentiment classification." -Proceedings of the 2016 conference on empirical methods in natural language processing. 2016.*

[4] *Song, Youwei, et al., 2019, "Attentional encoder network for targeted sentiment classification."*

[5] *Dai, Junqi, et al. 2021, "Does syntax matter? a strong baseline for aspect-based sentiment analysis with roberta."*

[6] *Vaswani, Ashish, et al., 2017, "Attention is all you need." -Advances in neural information processing systems 30 (2017).*

[7] *Bing, P. C. Zhongqian Sun Lidong, W. Yang., 2017, "Recurrent Attention Network on Memory for Aspect Sentiment Analysis."*

[8] *Burton, Suzan, A. Soboleva.,2011, "Interactive or reactive? Marketing with Twitter." Journal of consumer marketing (2011).*

[9] *Pontiki, Maria, et al., 2015, "Semeval-2015 task 12: Aspect based sentiment analysis."*

[10] *Pontiki, Maria, et al., 2016, "Semeval-2016 task 5: Aspect based sentiment analysis." International workshop on semantic evaluation*

[11] *Blohm, Matthias, et al., 2018, "Comparing attention-based convolutional and recurrent neural networks: Success and limitations in machine reading comprehension."*

[12] *Devlin, Jacob, et al., 2019, "Bert: Pre-training of deep bidirectional transformers for language understanding."*