Geo-semantic profiling of brand-specific customer experience using citizen-generated social media comments

Ngele Emmanuel, Olubayo Adekanmbi, Anthony Soronnadi, David Akanji

ABSTRACT

A good customer experience is likely to influence a customer's decision to buy positively and equally a negative customer experience will most likely make a customer decide not to buy or go elsewhere. One negative customer experience is likely enough to make a customer leave a brand, and go to a competitor. To conduct this research, labeled Twitter data was utilized, and the Spacy library was employed to extract location information from tweets. The sentiment analysis of the tweets, categorizing them into positive, negative, and neutral sentiments, was accomplished using the Vader lexicon. The Vader lexicon, a valuable resource available in the Natural Language Toolkit (NLTK), provided a basis for sentiment evaluation.

Baobab extended abstracts submission

INTRODUCTION

The increasing use of social media platforms, customers are constantly sharing their experiences and opinions about brands. This data can be mined and analyzed to gain insights into customer behavior and preferences, which can inform brands about how customers perceive their brands and what to do to effect change that yields maximum results.

Geo-semantic profiling takes this a step further by incorporating geographic information into the analysis. By identifying the location of customers who are commenting on a brand, businesses can better understand how their customer experience may differ across different regions. This information can be used to tailor marketing efforts and improve the overall customer experience.

Citizen-generated social media comments are a valuable source of data for this type of analysis. By leveraging natural language processing techniques such as Name Entity Recognition and sentiment analysis businesses can extract meaningful insights from these comments and gain a deeper understanding of their customers.

Several studies have explored the use of geo-semantic profiling in analyzing social media comments. For example, a study by Kim et al. (2018) used this approach to analyze customer experiences at a luxury hotel chain.

Overall, geo-semantic profiling of brand-specific customer experience using citizen-generated social media comments is a powerful tool for businesses to better understand their customers and improve their pain points.

The research aims to analyze customer complaints directed at Access Bank using natural language processing (NLP) techniques to identify the most frequent issues and their sentiment. This analysis will help Access Bank improve their services and customer satisfaction.

METHODOLOGY

Data collection: The data used for this research work was collected from Twitter. Tweets containing the handle "@myaccessbank" and "@Accessbank" were targeted to gather customer complaints about Access Bank. Data was collected over a period of time using Twint open source tool to scrape Twitter platform resulting in a collection of tweets.

Data Annotation: The gathered data set was annotated using the NER annotation tool. This annotation was done so locations can easily be identified during training i.e location becomes an entity.

Training the NLP model: This research work used Spacy, a popular NLP library, to train a custom named entity recognition (NER) model to identify named entities in the tweets, in our case name of locations mentioned in the tweet. The model was trained on a labeled dataset of texts containing names of locations in Lagos Nigeria that were manually annotated to highlight named entities such as locations

Named entity recognition: A trained NER model was used to identify named entities in the collected tweets. The named entities were then stored in a CSV file.

Sentiment analysis: This research work used the Natural Language Toolkit's (NLTK) Vader sentiment analyzer to determine the sentiment of the tweets.

Geolocation of entities: The named entities were then geolocated using the geonames API to determine the latitude and longitude coordinates for each location mentioned in the tweets. These coordinates were stored in a separate CSV file.

Network Diagram: To get a view of what customers in our dataset are tweeting about and how each word relates to other words in the dataset we create a network diagram using a tool called

“Nocodefunction”. This tool takes a corpus of words and then converts it to a network diagram after some parameters have been set. The parameters include:

i) **Minimum word length**: removes words below a certain length, these words might not serve any interests in our dataset.

ii) **Minimum frequency for words:** This excludes words that have occurrences fewer than the words mentioned.

iii) Pointwise Mutual Correction(PMI): Co-occurrence based word association measures like PMI measures the strength of association between two words comparing the words pair’s corpus-level bigram frequency to some function of unigram frequencies of individual words (Damani, 2021).

RESULT

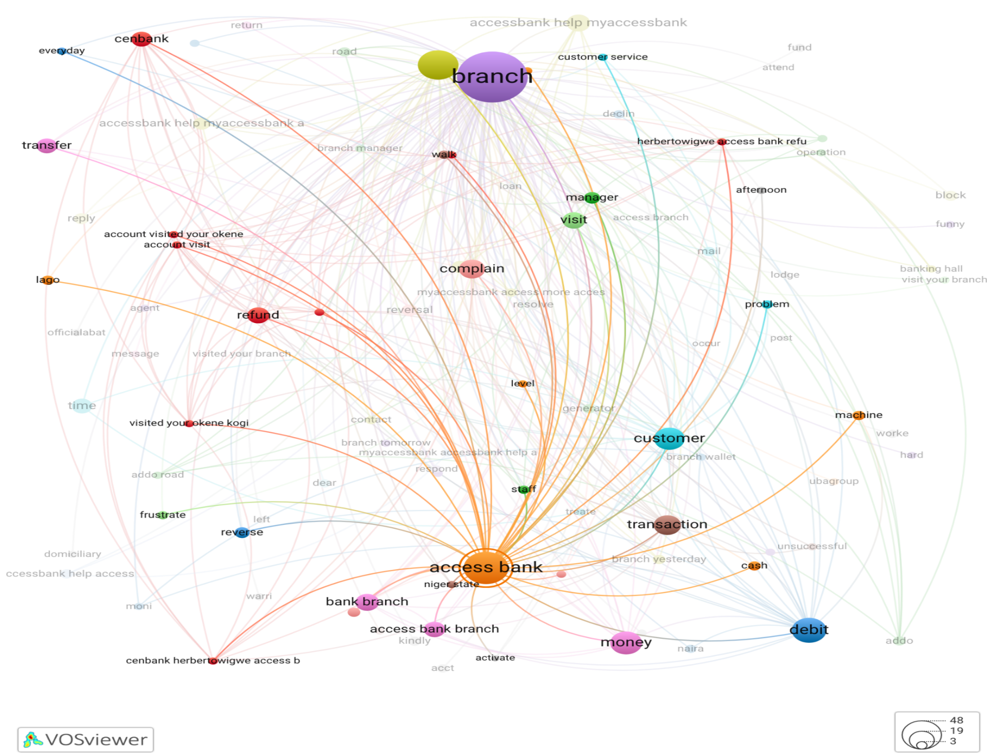

Figure 2

The network diagram presented above provides valuable insights into the discourse surrounding Access Bank's customer base. Through a comprehensive analysis of the dataset obtained, it is evident that customers predominantly engage in discussions related to transactions and debits, showcasing the key areas of interest and concern within the banking context.

The findings highlight the significance of these topics within the customer community, suggesting that transactional activities and debit-related matters hold substantial weight in shaping their overall banking experience. This research sheds light on the prevalent themes in customer conversations, emphasizing the relevance of such aspects in understanding customer needs, preferences, and potential pain points.

Figure 3

Figure 4

The results presented in Figure 3 and Figure 4 demonstrate the effectiveness of our NER model in recognizing and tagging specific locations mentioned by the customers in their tweets. This research outcome illustrates the capability of the model to identify relevant geographical references accurately, thereby contributing to the understanding of the spatial context within customer conversations.

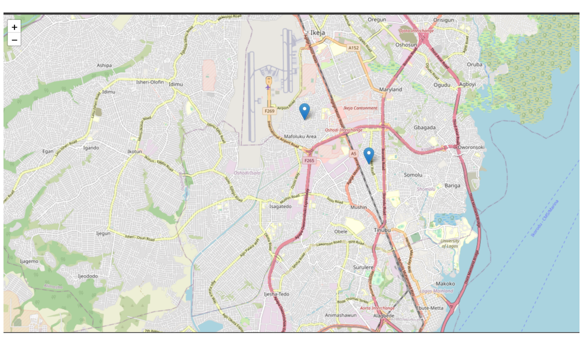

Figure 5

The visual representation presented in the figure above depicts a map illustrating the locations identified by the Named Entity Recognition (NER) model from the customer dataset. This research outcome demonstrates the successful application of the NER model in identifying the geographic information contained within customer interactions.

DISCUSSION

The NER model was trained to identify locations in Lagos only. The model had a very good accuracy, however it failed to identify certain locations, further research and findings revealed that one of the major causes of that shortcoming is the fact that the model was trained on an English dataset and some of the locations that were not identified were named in other languages.

CONCLUSION

In this study, we have investigated the potential of natural language processing techniques to create valuable tools for organizations to understand their customers' pain points and also identify the sources. Our findings suggest that these tools can help companies gain useful insights into their customers' preferences and improve their overall customer experience.

References

Kim, Y. J., & Kim, H. S. (2022). The Impact of Hotel Customer Experience on Customer Satisfaction through Online Reviews. Sustainability, 14(2), 848. https://doi.org/10.3390/su14020848

Oyewusi, W. F., Adekanmbi, O., Okoh, I., Onuigwe, V., Salami, M. I., Osakuade, O., Ibejih, S., & Musa, U. A. (2021). NaijaNER: Comprehensive Named Entity Recognition for 5 Nigerian Languages.

Damani, O. P. (2013). Improving Pointwise Mutual Information (PMI) by Incorporating Significant Co-occurrence. arXiv preprint arXiv:1307.0596v1 [cs.CL].