An Ecommerce Module for Anaylzing Qualitiative Product Reviews:

Demonstrating using a Women's Clothing Reviews Dataset from an Amazon Business

***Abstract*—** This project explores the relationship between customer reviews and product recommendations, using a sample dataset from a Woman's Clothing business on Amazon. Through sentiment analysis, this project seeks to quantify the impact of textual feedback on consumer satisfaction and preferences. The research is pivotal for businesses looking to tailor their offerings more effectively to consumer needs. The study builds on previous research in sentiment analysis and consumer behavior, utilizing a methodology that combines data cleaning, machine learning, sentiment polarity analysis, and exploratory data analysis.

Key observations from this project include: **1.** A strong correlation between reviews and recommendations, with the TF-IDF model for SVM predicting this relationship with an accuracy rate of 85%. **2.** Discrepancies between numerical ratings and textual sentiment in approximately 700 reviews, suggesting the importance of textual analysis alongside numerical ratings. **3**. Sentiment polarity analysis across demographics, revealing insights into popular and best-reviewed products, as well as common themes in positive and negative reviews. **4.** Topic modeling on reviews of top-selling items, provide further insights into customer discussions and preferences.

The study concludes that the combination of TF-IDF and SVM delivers the best results in sentiment analysis, with an 85% accuracy rate. Therefore, this analysis module helps identify areas for product improvement, which ultimately means increased customer satisfaction. These findings can inform marketing and product development strategies, with implications for Amazon and merchants in terms of display optimization, inventory adjustment, and after-sales service improvement. This model can be used on other related businesses' to help better understand their customers' feedback.

I. Introduction

E-commerce has significantly altered the way consumers shop, which means customers review products directly online. These reviews are a critical for the business' future of making their products better. Customers provide feedback and insights into their consumer satisfaction. Business' do not need to guess how if their customers are enjoying their products, they can simply look at the products review. However, the challenge comes when business analysts do not know how to interpret or analyse text reviews. In our project, we researched how understanding the intricate relationship between customer reviews and their subsequent recommendations of products are valuable insights that need to be analyzed. The significance of this study lies in its innovative approach to quantifying sentiment analysis and the correlation of textual feedback with product endorsement. This research serves as a project, exploring how sentiment analysis can be leveraged to decode customer attitudes towards products, which is crucial for businesses aiming to tailor their offerings to meet consumer needs more effectively. The potential benefits of this study extend to various stakeholders, including consumers, who may receive better-targeted products, and companies, which can enhance product appeal and customer satisfaction. Textual free- text reviews are great for analyzing specific feedback from customers, however how do businesses with a high volume of textual reviews analyze, interpret, and use the feedback to improve their products or services? This research exercises various textual analysis models to determine the best method for gaining insights into the perceptions and satisfaction levels of their target customers, while factoring in their shared opinions. This research uses a classification analysis to assess the correlation between the strength of review texts and the likelihood of product recommendations, alongside, performing a sentiment analysis to identify the words associated with positive, neutral, and negative sentiments in the reviews. This kind of analysis enables a deeper understand of the language and emotions expressed by customers in their feedback. This methodological research can be replicated in any company receiving free text, textual reviews on a product or service to better understand their customer's opinions/reviews. Which ultimately, provides business' with data that can influence marketing strategies or back-up physical product changes. Ethically, the project adheres to respecting user privacy and data protection standards, while socially, it aims to enrich consumer choice and satisfaction.

II. Rational / Motivation

Majority of ecommerce businesses fear allowing free-text review because of how much time is required to quantify, compare and analysis those reviews. We wanted to create a model to analysis free-text opinions, reviews and feedback that could allow businesses to easily examine reviews and enable businesses to connect their market strategies to the results of customer feedback data.

III. Research Question

Our research seeks to answer the question: How do qualitative product reviews, particularly in terms of sentiment and language used, influence the likelihood of product recommendations by consumers? Specifically, we aim to examine the intensity of sentiments expressed in customer reviews and investigate whether these sentiments correlate with the reviewers' willingness to recommend the products to others. This research explores various sentiment textual review data analyzing models to produce a recommendation for the most effective model for this kind of analysis. What kinds of data analysis can be used to organize and better understand textural consumer reviews in the E-Commerce industry? What modes are most effective?

IV. Literature Review

McKinsey & Company published a report that says "most of the people check the ratings and the reviews before buying a product or signing up for a service and for that reason sentiment analysis has played a crucial role in planning long-term strategies for some of the world's biggest companies and helped increase their revenue." [1] There are various benefits for business' capitalizing on sentiment analysis techniques such as, improving marketing strategies, making effective decisions about the products, getting to know your customer better, mitigating risks of production quality or social media backlash based on content presented by the business, and ultimately improve customer service. [1]

Sentiment analysis and National Language Processing (NLP) in the context of e-commerce has been explored by various researchers, who emphasized its importance in understanding consumer behavior online, especially when it comes to reviews [2]. There are various pieces of literature that outline sentiment analysis and machine learning as a methodology to analyzing the opinions that emerged out of the COVID-19 crisis, such as online learning [3]. Shanghao and co-authors examined various textual social media posts using sentiment analysis to expose their emotions and topic modeling to understand the reasons why people were against online learning. [3] This research project explores this same combination of sentiment analysis and topic modeling to analyze customer reviews, to better understand the relationship between understand this sample business' dataset's customers qualitative product reviews and their product recommendation. The other aspect of our research deals with customer satisfaction and loyalty [4]. By performing these kinds of textual analysis methods on a daily basis, the business can present real-time emotional changes happening to their customers, quickly identify satisfied customers, monitor multiple customers at once, identify what words are triggers for positive and negative emotions, and potentially cutting down customer churn. [4]

Our research suggests that combining qualitative and quantitative data, such as textual reviews and demographic information can inform patterns in customer behaviour to help better target consumers [5]. This project recognizes the value in combining the demographic information associated with the positive and negative reviews. This kind of data would provide the business with more context as to who is happy or frustrated with the products. While not all demographic information about customers will correlate to a specific kind of textual review or recommendation, this kind of analysis will provide a further specificity of customer identification. The idea of combining qualitative and quantitative data is not new, but the methodology of using sentiment analysis and topic modeling to easily see these relationships are in a timely matter is what this project is presenting as unique.

V. DATASET

The dataset used for this analysis was *Women's E-Commerce Review*. The dataset had 23486 rows and 10 feature variables. The data description follows, RN (Row Number), Clothing_ID (product idneitfication number), Age (Age of the Reviewer), Title (Title of the Review), Review_Text (Review Text), Rating (Rating for the Product as given by the reviewer, Recommended_IND (whether the reviewer would recommend the product to others, Postive_Feedback_Count (how many people found the reviews helpful), Division_Name (classification of the product at division level), Department_Name (classification of the product at the department level), Class-Name (classification of the product at category level). [6,7,8,9].

VI. THEORETICAL FRAMEWORK / METHOGOLOGY

Our research problem is based on the increasing use of textual data in E-Commerce industry but the limited tools that exist for analysis. We are focused on establishing a framework of Sentiment Analysis and implementing the results for different functions within organizations. For example, a review from the women's clothing amazon dataset reads, "Love this dress! Its sooo pretty." How can companies' asses this positive review? Our methods combine textual data with different quantitative categories like reviewer's age, gender, product's category, rating etc. to understand if there is any prevalent pattern or correlation among the categories.

Our methodology is rooted in sentiment analysis, aiming to establish a framework that not only assesses the sentiments expressed in product reviews but also implements these insights across different organizational functions. We propose a comprehensive approach involving:

1. Implement Stop Words Removal, Regex expressions, Unicode, and Lemmatization for Data Cleaning and Preprocessing.

2. Determine the Sentiment Labels based on the Ratings given by the reviewer. We took '<3' as 'Negative', '3' as 'Neutral', and '>3 as Positive'.

3. We then performed Feature Engineering using Word2Vec Model and TF-IDF Models.

4. Check if the Textual Reviews matched the Recommendation (Recommend_IND) of the reviewer. Used SVM Model and Multinomial Naive Bayes Model (MNB). Applied features generated from Word2Vec and TF-IDF to both; SVM as well as MNB Model.

5. Next we determined Sentiment Polarity using TextBlob Model. From the model we gauged Sentiment of the reviews. The results are bounded between (-1,1) range. We used the following logic to determine the Qualitative measure of the polarity.

Positive: (>0.2 to 1)

Neutral: (between -0.2 to +0.2)

Negative: (Less than -0.2)

Using the Results from the Sentiment Polarity Analysis we compare the Human generated Sentiment Labels in Step 2 and Compared with the results generated here. Based on the differences observed and as per the business objective the Marketing/Customer Service team could further investigate what led to differences in the tone of the Reviews and the value of the Ratings. This would help in boosting the Social Standings of the Organizations.

6. We also perform Exploratory Data Analysis by combining Average Sentiment Polarity Values across Age, Clothing Category (Class_Name) and Products (Clothing_ID). This analysis would help us understand a lot of different aspects regarding product strategy. We would get to understand which Age Range of customers would buy which Category of Products and how they Rate different Categories and Products. Depending on the results the relevant teams would determine future strategies. On which products they would focus and which Age Range they would target.

7. We also performed Parts of Speech Tagging (POS Tagging) to generate the most common Verbs (Actions) and Adjectives (Descriptions) used for Positive and Negative Reviews.

8. We performed Topic Modeling (a machine learning technique) that can group similar words and phrases in textual documents, such a free-text reviews into categorical clusters to organize the textual data. Topic Modeling clusters patterns and reoccurring words to create a quick glance report at what customers are saying. We compared the Topic Modeling results with POS tagging. Comparing helps understand the theme of the reviews. Topic Modeling aims to understand the theme behind the reviews of products.

By integrating these steps, our framework aims to provide actionable insights into consumer sentiment and behavior, thereby guiding marketing and product development strategies.

VII. Observations

1. Correlation between Reviews and Recommendations: Out of the 4 models, our analysis revealed that the TF-IDF model for SVM outperformed other models, successfully predicting the correlation between reviews and recommendations with an accuracy rate of 85%.

TF-IDF report for SVM Classification is:

	precision	recall	f1-score	support
0	0.66	0.39	0.49	1247
1	0.87	0.96	0.91	5546
accuracy			0.85	6793
macro avg	0.77	0.67	0.70	6793
weighted avg	0.83	0.85	0.83	6793

Figure 1: Classification Performance Metrics for SVM Using TF-IDF on Customer Review Data

This indicates a strong relationship between the sentiment expressed in reviews and the likelihood of a product being recommended by customers.

2. Discrepancies between Ratings and Textual Sentiment: Upon comparing sentiment labels and polarity, we discovered that approximately 700 reviews exhibited discrepancies between their numerical ratings and the sentiment conveyed in the text. This finding suggests that numerical ratings alone may not fully capture the nuances of customer sentiment, highlighting the importance of conducting a deeper textual analysis.

	precision	recall	f1-score	support
Positive	0.87	0.75	0.80	17448
Neutral	0.21	0.55	0.30	2823
Negative	0.67	0.04	0.08	2370
accuracy			0.65	22641
macro avg	0.58	0.45	0.39	22641
weighted avg	0.77	0.65	0.67	22641

Figure 2: Sentiment Polarity (Text Blob Model) Performance

3. Sentiment Polarity across Demographics: Through our exploratory data analysis, we generated plots that provided insights into the most popular and best-reviewed products. For instance, dresses, knits, and blouses emerged as the top three selling clothing items.

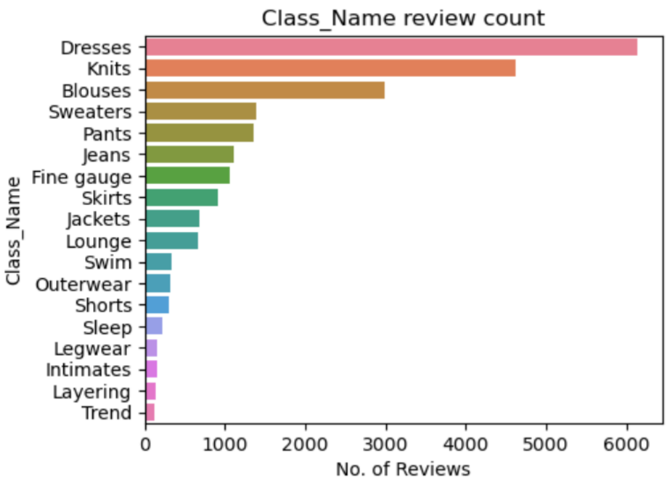

Figure 3: Number of Reviews Per Clothing Category

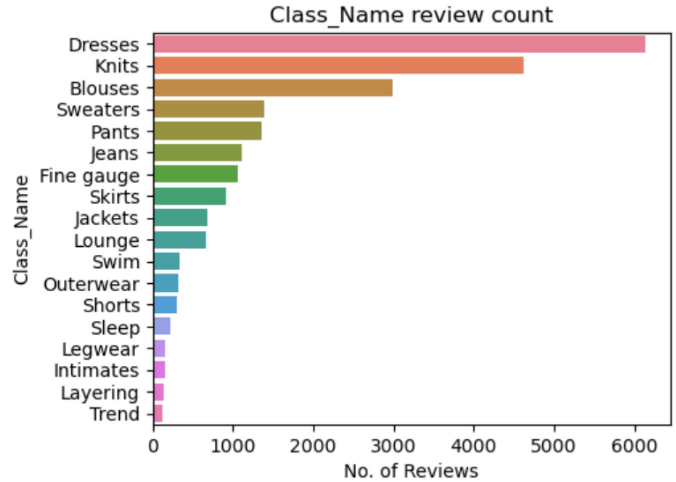

Figure 4: Number of Reviews Per Clothing Category

Plot Sentiment Polarity Across Age Groups, Product Categories, and Product Identifiers. This would help us understand which are the best selling products and also the best reviewed products across the spectrum. The Marketing/Strategy team can then proceed further. Specifically, clothing item 1078 was identified as the highest seller, with an average positivity rating of 0.23 to 0.28 across different age groups(Appendix D). While blouses received similarly high ratings, their sales were significantly lower compared to dresses, indicating a potential area for increased marketing focus.

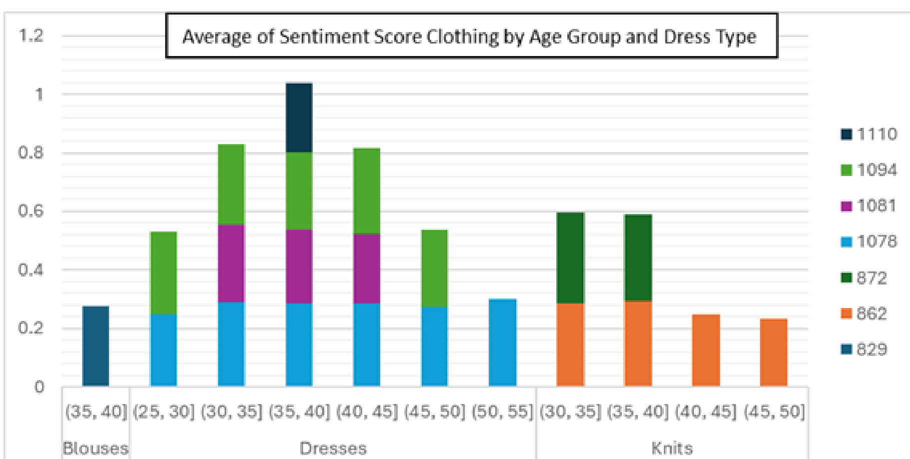

Figure 5: Average of Sentiment Score Clothing by Age Group and Dress Type

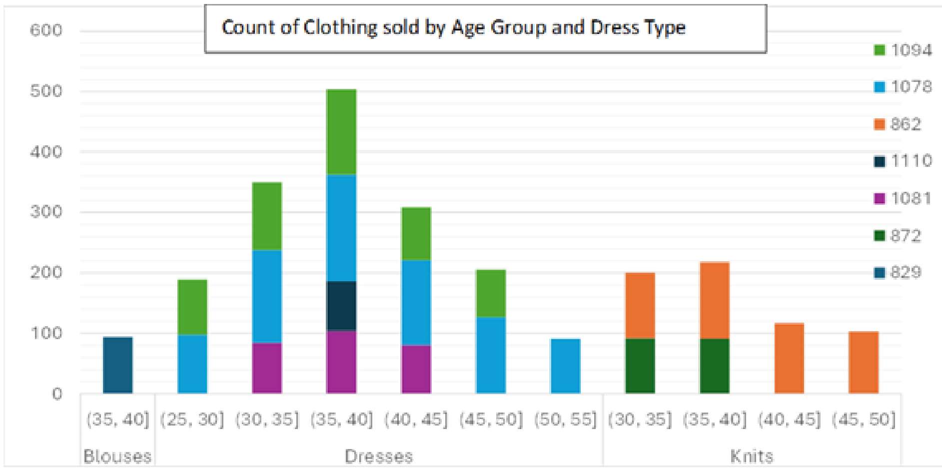

Figure 6: Count of Clothing sold by Age and Dress Type

4. The word clouds generated from our analysis provided a visual representation of the most frequently discussed topics in positive and negative reviews. The word strength is based on frequency. The most common themes in positively reviewed products included attributes such as "fit," "fabric," and "look," while negatively reviewed products often mentioned issues related to "size" and "quality." This information is invaluable for understanding customer preferences and addressing common concerns.

Figure 7: Plot Word Clouds for most used words (Verbs and Adjectives) for Positive and Negative Reviews

5. The topic modeling conducted on reviews of dresses and knits (the top two selling items) provided further insights into customer discussions. For dresses, the predominant topics related to the fit, fabric, look, and length of the garments. For knits, customers frequently discussed the size, fit, and wear of the tops. These findings can guide the marketing team in crafting messages that resonate with customers' preferences and experiences.

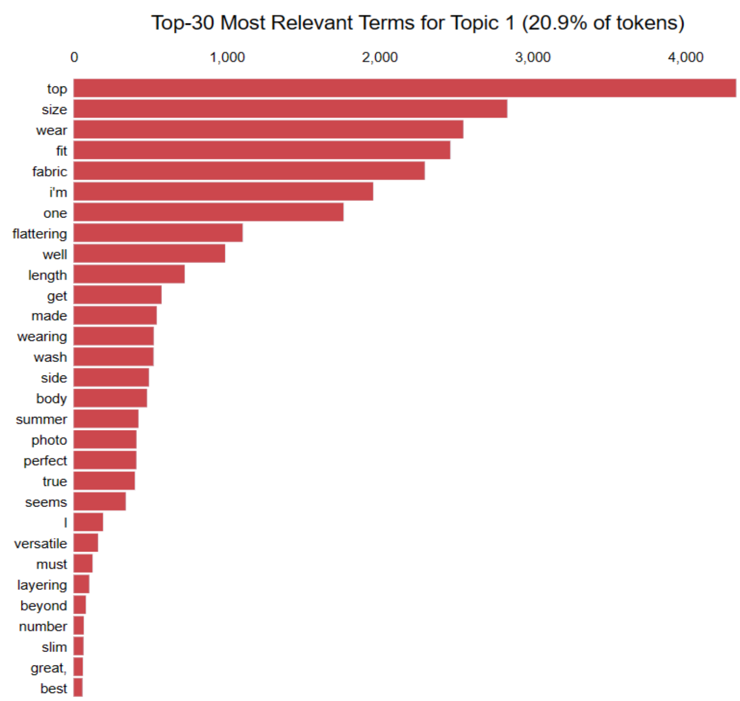

Figure 8: Top 30 Most Relevant Terms for Topic 1

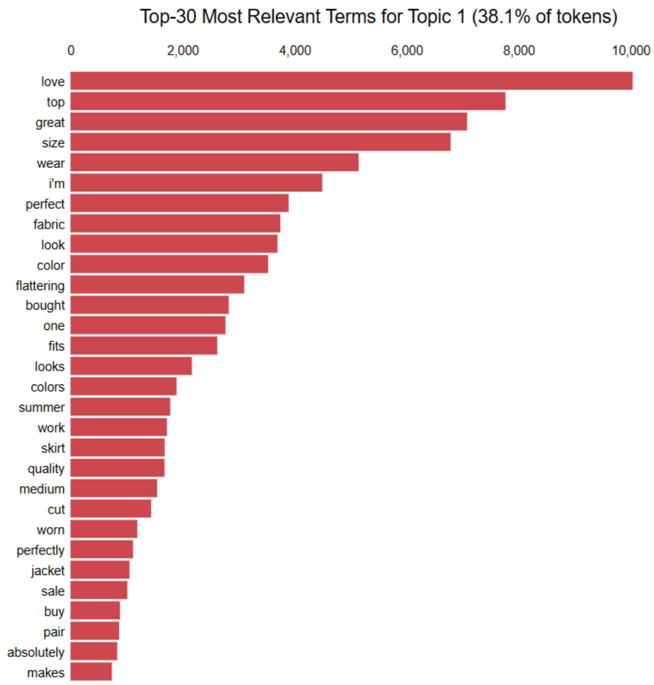

Figure 9: Top 30 Most Relevant Terms for Topic 1

VIII. BUSINESS ADOPTION

This kind of Topic Modeling machine learning technique allows businesses to save money on hiring employees to analysis textural reviews. Topic Modeling is a quick way to see the businesses reviews to adjust the market strategies or products themselves. Topic Modeling is an unsupervised method to review a business' textual reviews and would be especially effective for ecommerce businesses operating on Amazon.

IX. CONCLUSIONS

Recommendation and Polarity Analysis produced the Sentiment of Reviews. TF-IDF and SVM combination delivered the best results (85% Accuracy). This kind of analysis module helped us identify target areas of improvement for products. Word clouds and LDA plots help organize what customers paint points for certain clothing products. This kind of data analysis allows us to understand that the largest customer base between ages 35-40 years old prefers productions with IDs 1110, 1094, 1081, and 1078. This is a strong framework for conducting a meaningful Sentiment Analysis for how to analyze textual customer review.

For e.g. people in the age range of 40-45 rated Category_ID 872 (Knits) much higher (0.31) vs Category_ID 862 (0.22). However sales for 872 are lower, so that is a target area to focus on. The word clouds and the LDA plots help understand that customers loved the Fit, Fabric, Look of Dresses and Knits, and hence as a Company those should be our prime targets. It also helped us understand that our largest customer base is aged between 35-40 years and their preferred products are IDs 1110, 1094, 1081, and 1078. We believe that we have established a strong framework for conducting meaningful Sentiment Analysis and how to leverage the findings into action points.

X. IMPLICATIONS

Based on the results of our analysis, there are many meaningful takeaways for Amazon and merchants on Amazon.

For Amazon, displaying reviews that are highly correlated with positive emotions facilitates customer trust and informed purchase decisions. For products where a high number of purchase reviews are available, Amazon can optimize the display structure so as to facilitate users' purchase decisions. Secondly, different products are preferred by different age groups. Amazon can customize the user interface to display the products that are most likely to be purchased and have high sentiment scores for users of different age groups. Third, reviews have a monitoring effect. Amazon could impose stricter quality controls or certifications on women's apparel in terms of fit, fabrics, and help build consumer trust. Finally, after-sales service is strongly associated with reviews. Reviews with a greater proportion of negative sentiment are more likely to contact the Amazon after-sales team for support. Training after-sales service teams using insights from sentiment analysis can quickly build Q&A lists and better target customer service, leading to improved service quality and efficiency and increased customer experience.

For merchants, they can refer to reviews and buyer insights when designing new products or improving existing ones. For example, for knits and dresses, attention should be paid to the most highly rated features, including fit and fabric. Additionally, adjust inventory based on sentiment and sales data for garments, especially those with ID's 1110, 1094, 1081 and 1078, to avoid understocking issues. Finally, considering the pricing strategy, merchants can decide whether they need to make a moderate premium for products with high favorable reviews and positive sentiment.

In summary, this analysis provides a foundation for understanding the complex relationship between text feedback and product success. Stakeholders should continue to collect and analyze this type of data on a regular basis to adapt to changing consumer trends and maintain a competitive advantage.

XI. Limitations

1. Sentiment analysis and natural language processing (NLP) techniques rely on algorithms and models to interpret and analyze textual data.

Therefore, nuances including sarcasm, idioms, and cultural references are challenging for these techniques to fully capture. If the reviews come from an international audience, the analysis might not fully capture cultural nuances or idiomatic expressions that affect sentiment interpretation.

2. The preprocessing steps, such as stop words removal and lemmatization, may not be optimal for all types of textual data and can affect the analysis results.

3. The study relies on data from a single source – Amazon's women's clothing reviews. This can limit the generalizability of the findings to other e-commerce platforms, product categories, or demographic groups.

4. The dataset may not be representative of all Amazon users or purchasers of women's clothing. There may be inherent biases, such as more engaged or dissatisfied customers being more likely to leave reviews.

XII. Acknowledgements

We thank Professor Elio Velazquez for his insights and guidance during the creation of this research project. We also would like to thank the Faculty of Science and Institute for Data Science at Carleton for arranging Data Day to present our research in poster format.

XIII. References

[1] “Improve Your Business with Sentiment Analysis.” *Improve Your Business with Sentiment Analysis*, www.axisgroup.com/data-industry-insights-blog/improve-your-business-sentiment-analysis. Accessed 14 Apr. 2024.

[2] Mehtab, Almas M. (2021, January 13) Sentiment Analysis in E-Commerce.FederatedAI.

[3] Li, Shanghao, et al. “Sentiment Analysis and Topic Modeling Regarding Online Classes on the Reddit Platform: Educators versus Learners.” *Applied Sciences*, vol. 13, no. 4, 2023, pp. 2250-, https://doi.org/10.3390/app13042250.

[4] Parish, Kate. 5 Ways to Use Sentiment Analysis to Improve Customer Experience. Shortstack.

[5] Couzin, Gradiva., and Jennifer. Grappone. *Five Stars : Putting Online Reviews to Work for Your Business*. 1st edition, Wiley, 2014, pp. 6-27.

[6] Doomlab, 2021 https://github.com/doomlab/statsofdoom-files/tree/master/nlp/Lectures%202019/06.classify

[7] Doomlab, 2021 https://github.com/doomlab/statsofdoom-files/tree/master/nlp/Lectures%202019/07.sentiment

[8] Parish, Kate. 5 Ways to Use Sentiment Analysis to Improve Customer Experience. Shortstack.