Report on Sentiment Analysis of Product Reviews

Trinay Gangisetty, Pranit Katwe, Kanishka Ghodke

Abstract

This project investigates consumer preferences and opinions across 15 categories by using sentiment analysis on over 23,000 women's online clothing reviews. The study classifies sentiments as positive, neutral, or negative based on provided rating values using natural language processing and machine learning models such as Logistic Regression and Support Vector Machines. Logistic Regression has the highest accuracy (95%). The use of visualizations reveals information about sentiment distribution, positive word frequency, and age group patterns. The findings provide businesses with actionable insights to improve product quality and customer satisfaction in the e-commerce clothing domain.

Contents

1 Introduction

The analysis of product reviews has become an essential tool for businesses looking to better understand and cater to consumerism's diverse preferences and ever-changing landscape. Online platforms make it easy to access customer sentiments, opinions, and experiences in the digital age, where information is shared at a never-before-seen pace. By leveraging the potential of these user-generated reviews, companies can obtain priceless knowledge about the advantages and disadvantages of their offerings, allowing them to adjust, develop, and eventually raise customer satisfaction. By analyzing the reviews especially women's clothing reviews business owners can understand what their customers have to say about the fit, quality, and look of the product and customize the clothing line accordingly.

One of the best ways to analyze product reviews is to form sentiment analysis on those reviews. Opinion mining, another name for sentiment analysis, is a sophisticated natural language processing method that entails the methodical assessment of sentiments or opinions expressed within textual content. Finding and classifying the text's emotional tone as positive, negative, or neutral is the main goal. This analytical method makes use of statistical techniques, linguistic features, and machine learning algorithms to interpret the subjective information buried in language. Sentiment analysis has useful applications in many different domains, such as gauging customer satisfaction through product reviews, keeping an eye on social media sentiment, and analyzing comments from customer service encounters. There are various levels at which the analysis can be carried out: from assessments at the document level to more detailed evaluations at the sentence and aspect levels.

Sentiment analysis has several challenges, such as identifying irony or sarcasm, accounting for cultural differences in language use, and comprehending contextual subtleties. Sentiment analysis is aided by several tools and technologies, including lexicon-based techniques, machine-learning models, and libraries for natural language processing. Given the circumstances, sentiment analysis offers insightful information about the emotional aspects of textual data, enabling businesses to make decisions that are well-informed and influenced by the opinions of their target audience.

By using the sentiment analysis in the project, which examined e-commerce reviews of women's clothes, to learn essential information about the opinions and feelings of the customers about the products. After carefully reviewing all the reviews, it was possible to pinpoint common opinions about the women's clothing products that were being sold in the online store. According to the sentiment analysis, a sizable percentage of the reviews were positive and highlighted the customer's satisfaction with the product's quality, design, and overall shopping experience. Customers often complimented the quick delivery, attentive customer service, and variety of clothing options.

2 Methods

2.1 Data Description and Preprocessing

2.1.1 Data Description

This sentiment analysis project dives deep into the realm of women's online clothing reviews, specifically focusing on customer preferences and opinions across 15 distinct categories, ranging from dresses and sweaters to pants and activewear. This dataset boasts over 23,000 insightful reviews, meticulously categorized by product division (general, petite, intimate, etc.) and department (tops, bottoms, etc.).

Column names with Description:

Clothing ID: Unique Integer Categorical variable of product.

Age: Integer Variable of the reviewer's age.

Title: String variable for the title of the review.

Review Text: String variable for the body of the review.

Rating: Positive Integer Variable for the customer's product score from 1 being worst to 5 being best.

Recommended IND: Binary variable that provides information on whether the customer recommends the product where 1 is recommended and 0 is not recommended.

Positive Feedback Count: Positive Integer documenting the number of other customers who found the review helpful.

Division Name: Categorical name of product division such as General, Petite, and Intimates.

Department Name: Categorical name of product department such as Tops, Bottoms, Etc.

Class Name: Categorical name of product class names such as Dresses, Knits, Pants, etc.

This improved version strengthens the focus on specific research questions (customer preferences and brand perceptions within women's online clothing), quantifies data size, and provides clearer descriptions of variables while maintaining a concise and engaging tone.

2.1.2 Data Preprocessing

Data preprocessing constitutes a crucial step in preparing data for analysis and modeling. It involves transforming raw data into a comprehensible and consistent format by addressing issues such as missing values, duplicate entries, and inconsistencies.

Utilizing the "missing no" library from Python, Figure 1 visually identifies missing values within the dataset. As evident, the "Title" column exhibits the highest percentage of missing data, followed by the "Review Text" column. To address this, two viable approaches exist: imputing missing titles with a predefined value like "No Title" or employing a hyphen ("-") as a placeholder. For the "Review Text" column, given the low percentage of missing entries (3.5%), discarding rows containing missing text may be an acceptable solution. However, imputing "Not Given", or "Unknown" values could be considered for columns like "Division Name," "Department Name," and "Class Name" where missing values are sparse

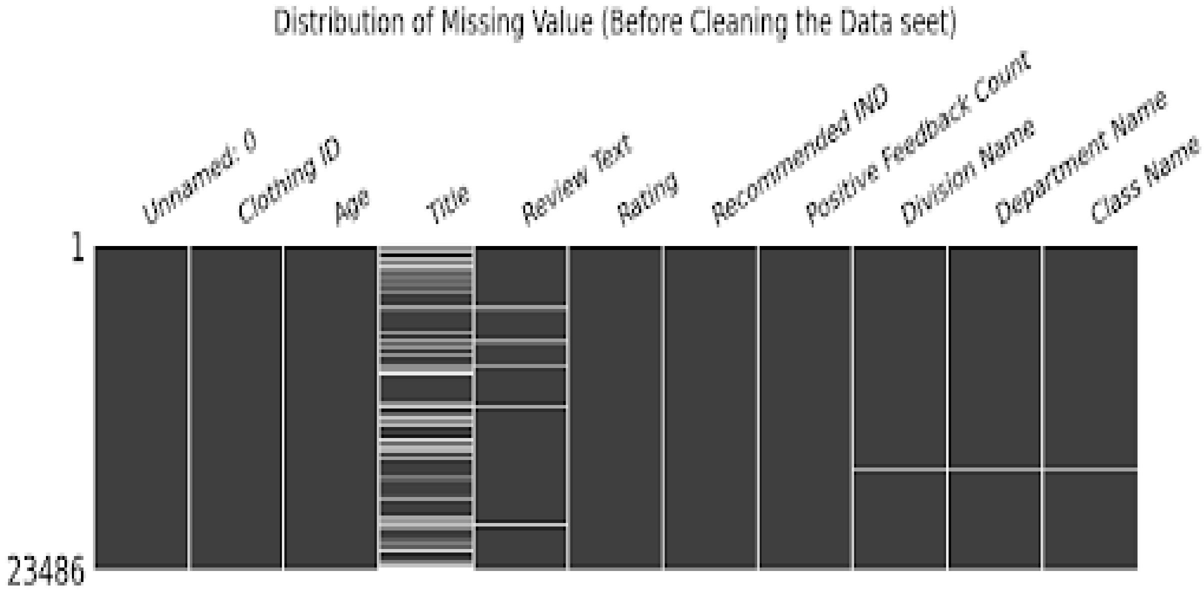

Figure 1: Missing Values

Finally, the "Unnamed:_0" column was dropped due to its lack of relevant information.

3 Exploratory Data Analysis

Exploratory data analysis is the most essential part of the entire project, and it is also the most underrated part. It is very possible to achieve excellent accuracy if and only if the dataset is cleaned well before it is fed into the model. Modeling will work to its complete extent only if the previous step is performed well. There are many sub-steps in EDA such as looking for the key features, analyzing the relationship between the columns, creating new features if required for the deeper analysis, detecting the target column, and analyzing how much each other column is contributing to the target column to be machine fed, extracting important insights from the visualizations created and dropping any further un-necessary columns or rows according to the insights acquired from the visualizations plotted, etc.

It is with this EDA alone; that many businesses are investing a lot in this part of the data analysis as this can lead to the discovery of wonderful and surprising insights that nobody would ever have known or expected before. The business could be conquered by getting this EDA right and implementing the measures required according to the insights obtained. Moreover, this is the most essential part of the subject of Data Mining which in turn refers to 'Mining insights from the data.'

In the next few lines of the paper, the insights of the project will be discussed with supporting visualizations for each of how each of them contributed to the overall project and what percentage of importance each of them contributed to the final project.

The above visualization shows the distribution of the target column **"Is_Recommended."**

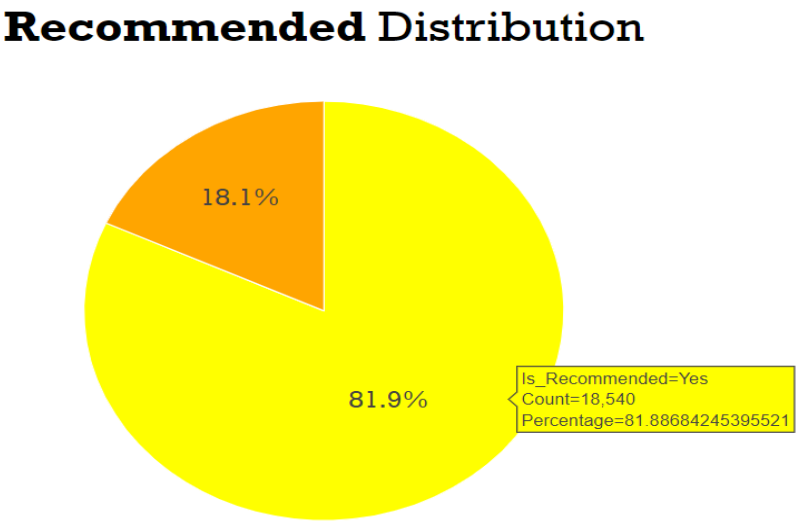

Figure 2: Classification of the target column

It can be said that the positive responses are 82% where are only 18% of the other. There is a huge bias in the data set which is expected in any of the review data sets. It is how the businesses run with more positive reviews. There are certain methods such as **"BRET"** or **"SMOTE"** analysis to deal with these biased data sets which will not affect the machine learning models to only perform well with the positive data and not the other. This is the simplest visualization yet the most powerful one in terms of the information it conveys to us.

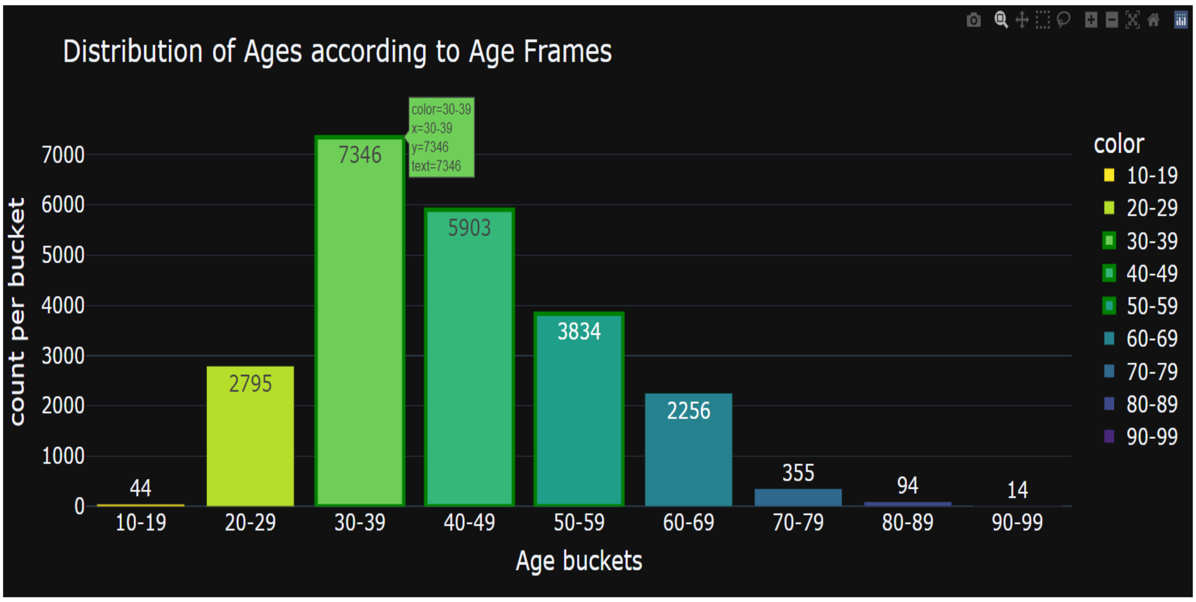

Figure 3: Distribution of people into age buckets

The entire population of the data set has been divided into certain age groups and the counts for each age group have been plotted to understand the distribution of the population. It can be seen that age bucket **30-49** contributes to the **60%** of the entire population. It can also be inferred that these two groups must be analyzed more deeply to train the model efficiently as they are the most active groups who are actively buying and reviewing the products. Also, inferring that there are few youths and elderly in the data set.

division_name	department_name	class_name
General	Bottoms	['Pants', 'Skirts', 'Jeans', 'Shorts', 'Casual...
	Dresses	['Dresses'] Categories (21, object): ['Blouses...
	Jackets	['Outerwear', 'Jackets'] Categories (21, objec...
	Tops	['Blouses', 'Sweaters', 'Knits', 'Fine gauge']...
	Trend	['Trend'] Categories (21, object): ['Blouses',...
General Petite	Bottoms	['Pants', 'Skirts', 'Jeans'] Categories (21, o...
	Dresses	['Dresses'] Categories (21, object): ['Blouses...
	Intimates	['Lounge'] Categories (21, object): ['Blouses'...
	Jackets	['Jackets', 'Outerwear'] Categories (21, objec...
	Tops	['Knits', 'Blouses', 'Fine gauge', 'Sweaters']...
	Trend	['Trend'] Categories (21, object): ['Blouses',...
Intimates	Intimates	['Intimates', 'Lounge', 'Sleep', 'Swim', 'Legw...
Not Given	Not Given	['Not Given'] Categories (21, object): ['Blous...

Figure 4: Pivot table for the distribution of classes

Sometimes by looking for any piping between the columns, they can be combined into a table such that they can be visualized further so that insights can be extracted from the table that is formed. We did try to do the same thing and finally were able to achieve it by combining the last three columns of the data into a single table. Initially, it was not possible to do so because the word **"Intimates"** has been spelled differently at various places and columns like **"Intimate,"** **"Intimate"**etc. This should also be one of the data cleaning methods or EDA methods that must be checked all the time.

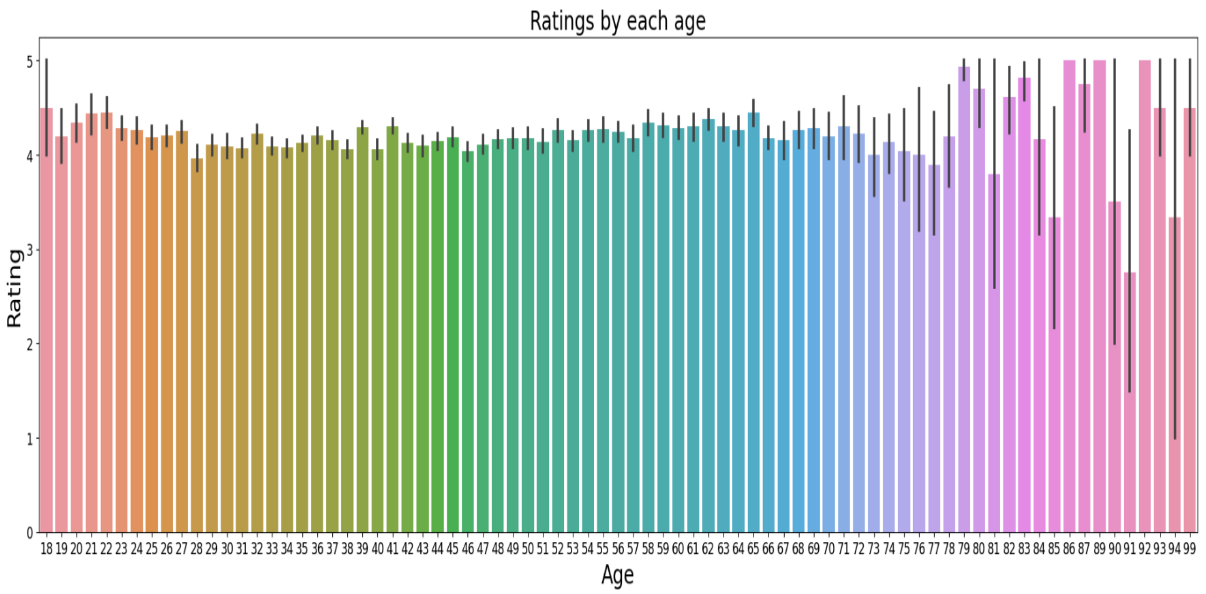

Figure 5: Ratings by Age with Error Bars

Fig 5 shows the distribution of ratings for each age and the proximity of error that each age group makes with the help of **"Error Bar"** plotted on the top of each histogram.

The error bar represents the entire group's possibility of error. The longer the length of error bar is the more dangerous the group would be to be considered for the final training of the data model. It can be seen from the visualization that the oldest age groups have the lengthiest error bars which make them unfit for the modeling because the reason might be their age factor which might fluctuate their ability to decide the worthiness of the product. So, these groups must be dropped before proceeding with further analysis. And there are very few elderly in the data set. So, it should not be a problem dropping them. The youngest also has an error bar with a certain length. This makes an especially important insight into age maturity.

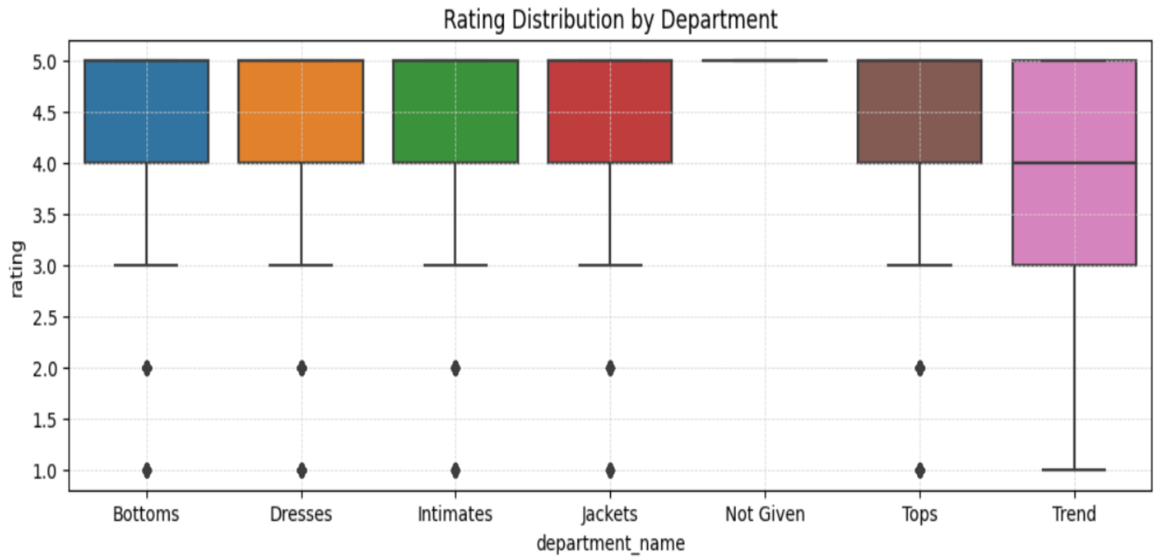

Figure 6: Distribution of Rating by Department

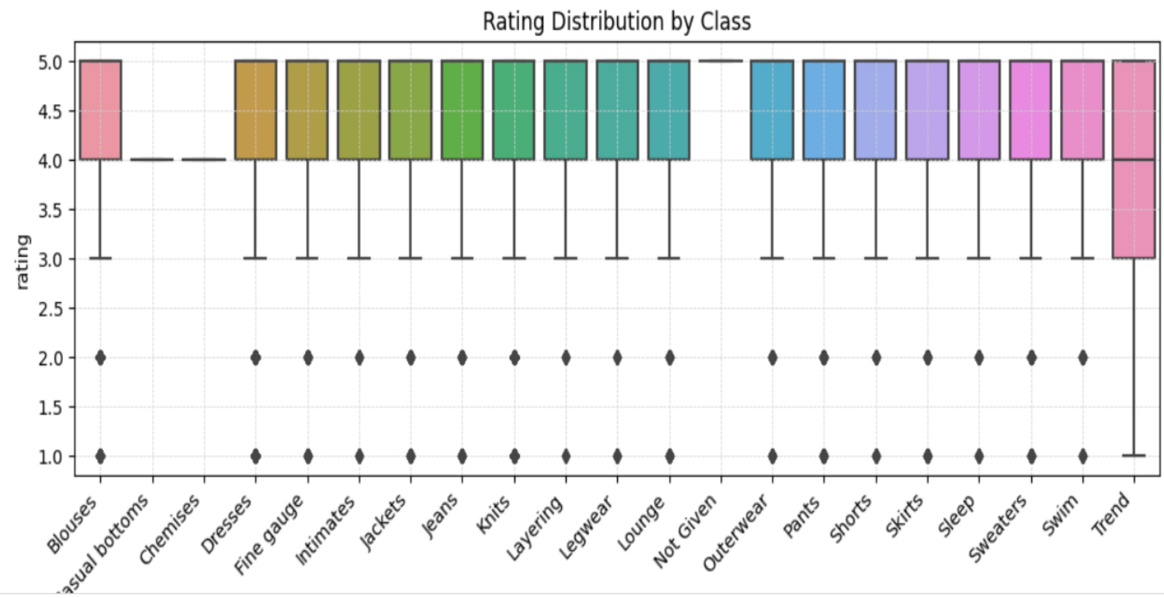

Figure 7: Distribution of Rating by Department

From the above two graphs, it can be seen that **"Trend"** is the only department or the class with the lowest ratings possible. Almost **25%** of the total reviews have ratings less than or equal to 3. For all the remaining categories there are very few ratings less than 3 which would be considered outliers. So **"Trend"** is the only department that must be analyzed further before making some important considerations or decisions. It can be seen in Fig 7 that two classes **"Casual Bottoms"** and **"Chemises"** have a line instead of a box plot. Further investigation of these classes shows that there are very few values or rows related to these two classes and all of those have the same rating which is 4. So, the box plot is truncated to a line.

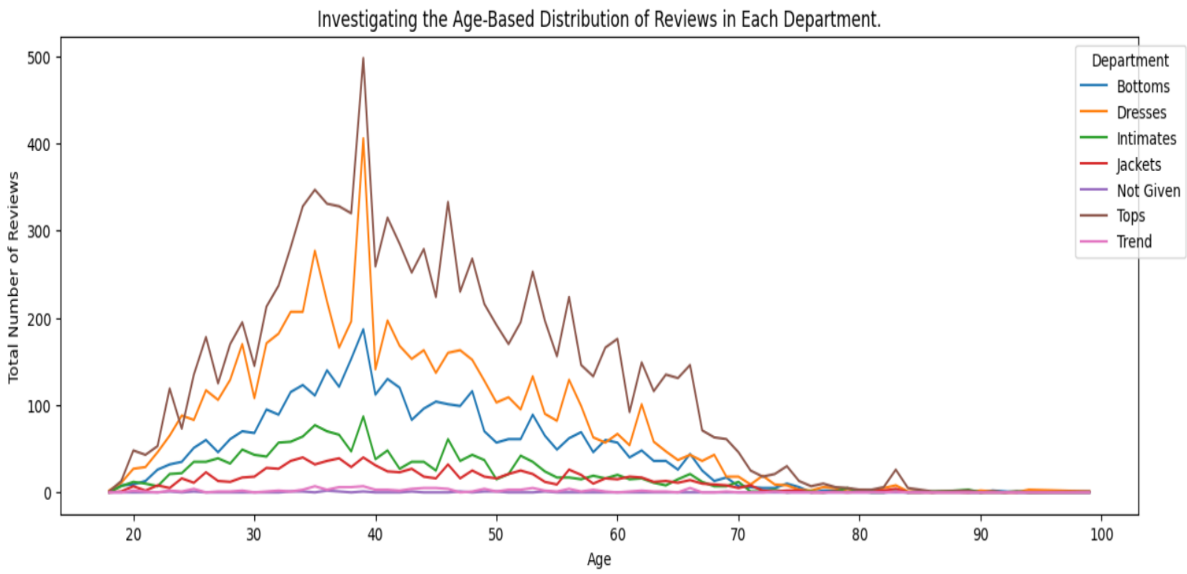

Figure 8: Age Distribution of Reviews in Each Department

Figure 8 depicts the distribution of reviews in each department for every single age in the data set. The **"Tops"** is the class with the highest number of reviews than any other class which is followed by **"Intimates."** The ages 39 and 40 are the two age groups that are giving the highest number of reviews.

Another crucial point that could be made from figures 3 and 8 is that ***"The recommendations being given is directly proportional to the products being bought in terms of age."***

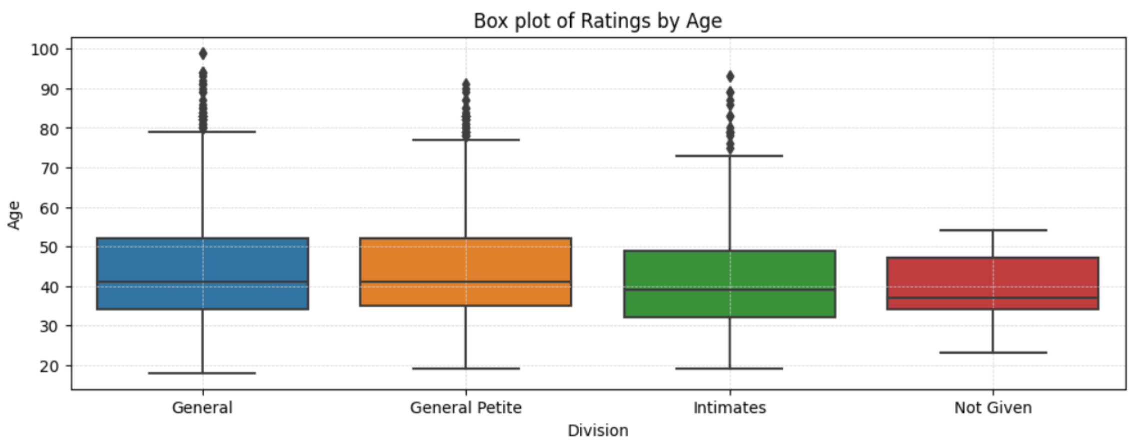

Figure 9: Box plot of Ratings by Age

Out of the three categories plotted above, **Intimates** is the category which has the lowest median age possible which is around 39 the entire Q3 lies below the age 50 and the highest possible outlier is 92. Whereas the other two categories have their median age above 40 and the Q3 above 50 and there is an outlier of age 99 in one of the classes. So, it can be concluded that **Intimates** is the only category that must targeted because of the target audience and the age group.

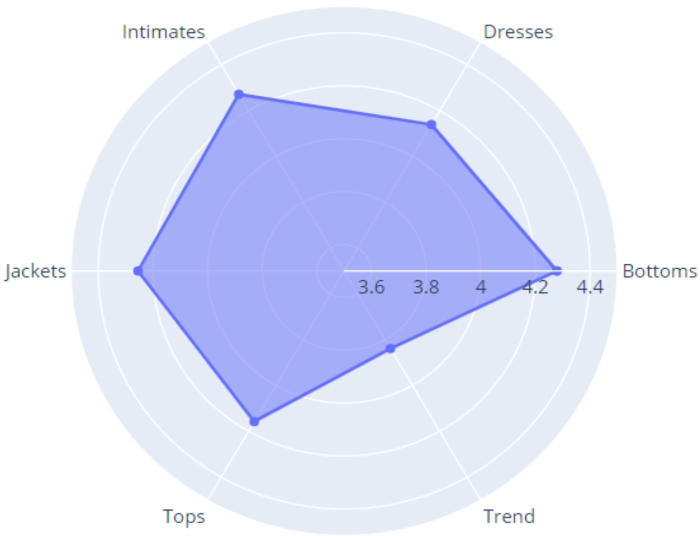

Figure 10: Radial chart for the mean ratings of departments

The above visualization depicts that the **Trend** is the only class with the lowest mean for the ratings of that class. This was the insight that was discussed in the previous visualizations and this visualization adds extra support and dimension to the above ones.

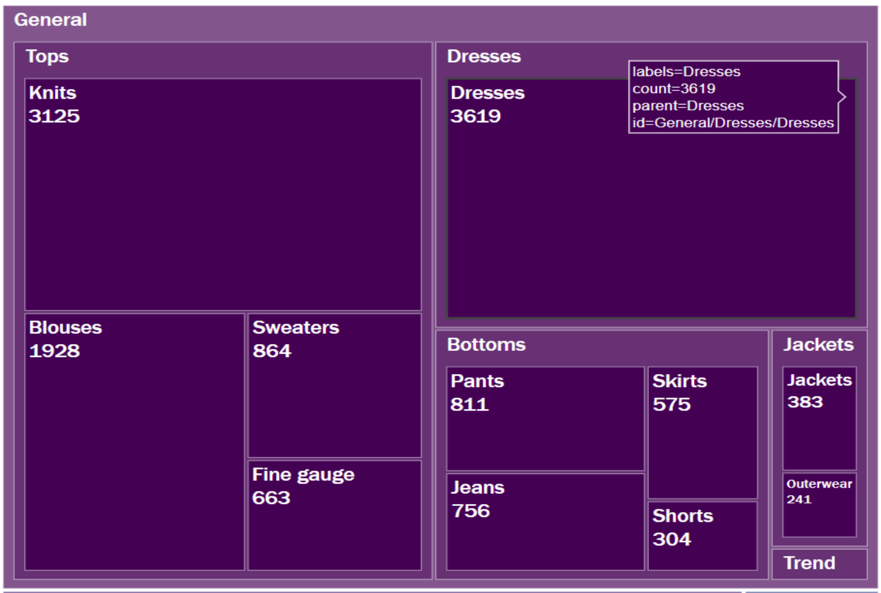

Figure 11: Tree Map for the piped columns

The above visualization is a Tree Map which shows the piping of the classes that are interlinked as a whole and what part of that class each of the items would occupy. It sizes the boxes according to the percentage of their distribution. This is one of the interesting visuals that can be used to understand the distribution of the piped classes.

So, many other important insights could be made from the data set by taking and grouping the other columns. There are an infinite number of possibilities that can be visualized.

4 Challenges

- **Bias:** One of the most challenging parts of the dataset that was reviewed was bias. There were around 80 percent positive reviews. Neural and negative reviews are similar in percentage which were around 10 each. To eliminate the bias an effective method that can be used in the future is data augmentation. Adding reviews that are more negative and neutral, will help get more accurate results.

- **Overfitting:** There were some of the models used on this data such as Neural Networks and Random Forest which provided an accuracy of 100% which is impossible for any real-time dataset. Dropout, weight decay, or early stopping can be used to prevent the overfitting of the dataset

- **Detecting Sentiments:** It is difficult for machine learning algorithms to detect sarcasm and figure-of-speech. Techniques like contextual embeddings and lexicons specifically designed to handle figure-of-speech can be improved.

5 Models and Classification

5.1 Text Processing in Review Text

Natural Language Processing (NLP) tackles the complex challenges of understanding human language. A crucial aspect of NLP involves handling stop words, familiar words with high frequency but low semantic value. To prepare the review text for further analysis, use NLTK's stop word list to eliminate these uninformative elements in the data.

Subsequently, feature extraction transforms the textual data into numeric features suitable for computational analysis. Here, Counter Vectorizers to convert each re to convert each review into a vector representation (numeric value), highlighting the frequency of each word within the text. This technique allows us to frequently use words, providing valuable insights into the sentiments and opinions expressed in the reviews.

5.2 Building a Sentiment Classifier

Given the absence of a dedicated sentiment column in the dataset, by implementing a sentiment classification based on the provided rating values. Reviews with ratings of 4 or higher will be classified as positive, ratings with 3 as neutral, and ratings of 2 or lower as negative. This method uses the implicit sentiment communicated by the rating scale to construct a sentiment label for each rating, allowing for additional analysis and modeling.

5.3 Splitting the Data

To prevent overfitting, a common challenge in machine learning is where a model learns the training data and performs poorly on unseen data, by splitting our dataset into training and testing sets. A 70/30 split is chosen here, balancing the need for sufficient training data with a robust testing set which can help for accurate model evaluation. This approach allows the model to learn effectively from most of the data while reserving a portion for unbiased assessment of its generalizability to new examples.

5.4 Model and Accuracy

Machine learning models are used to recognize the patterns in data or make predictions based on the given results. By using three models, and predicting the accuracy of the model which provides how accurate the sentiments are based on review texts and ratings.

5.4.1 Bernoulli NB

It is an interesting option for sentiment analysis which offers different strengths and weaknesses. Bernoulli NB is quick and training and predicting values for large datasets and real-time applications. By robustly performing on datasets where positive and negative examples are unevenly distributed which is quite common for analysing sentiments and it requires less hyperparameter tuning. One of the weaknesses of this model is the assumption of independence between features which might not hold in our case because ratings and review text are dependent on each other which has the highest correlation.

5.4.2 Logistic Regression

is a process of modeling the probability of a discrete outcome given an input variable. Logistic regression makes use of the sigmoid function which outputs a probability between 0 and 1. Logistic regression is simple to implement and computationally efficient, making it ideal for large datasets. Here by employing the logistic regression with word embeddings.

5.4.3 Support Vector Machines (SVM)

SVM is another powerful model for sentiment analysis that offers advantages as compared to Bernoulli NB and logistic regression. SVM can handle the complex relationships between features and sentiments by mapping data points to higher-dimensional space, making it suitable for nuanced sentimental analysis beyond simple positive and negative. Different kernel functions can be used to adapt SVM to various datatypes and relationships. Training on SVM can be slower because it is more resource-intensive than Logistic Regression ion and Bernoulli NB when it comes to large datasets.

6 Results

6.1 Accuracy of Machine Learning Models

After performing the machine learning models our results are generated. Bernoulli NB had an accuracy of 85%. Logistic Regression had an accuracy of 95%. Support Vector Machine (SVM) had an accuracy of 93%. Table 1, 2, and 3 shows the classification report of all these models which provides precision, recall, f1-scores, support, weighted average, and mean average.

Table 1: Classification Report for Bernoulli Naive Bayes

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Negative	0.83	0.55	0.66	1681
Neutral	0.60	0.50	0.55	1949
Positive	0.89	0.95	0.92	12218
Accuracy			0.86	15848
Macro Avg	0.78	0.67	0.71	15848
Weighted Avg	0.85	0.86	0.85	15848

Table 2: Classification Report for Logistic Regression

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Negative	0.93	0.88	0.90	1681
Neutral	0.91	0.77	0.83	1949
Positive	0.96	0.99	0.97	12218
Accuracy			0.95	15848
Macro Avg	0.93	0.88	0.90	15848
Weighted Avg	0.95	0.95	0.95	15848

Table 3: Classification Report for Support Vector Machine (SVM)

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Negative	0.93	0.82	0.87	1681
Neutral	0.96	0.66	0.78	1949
Positive	0.94	1.00	0.97	12218
Accuracy			0.94	15848
Macro Avg	0.94	0.83	0.87	15848
Weighted Avg	0.94	0.94	0.93	15848

6.2 Results Based on Visualization

From Figure 12, there are more than 17000+ positive reviews in this data while neutral and negatives are approximately 2500 each. A hypothesis based on this is that most of the reviews from this dataset are positive so there is some amount of bias in the data.

Figure 13 represents all the positive words that have the highest frequency in positive sentiments. One of the interesting patterns can be noticed that when women write reviews,

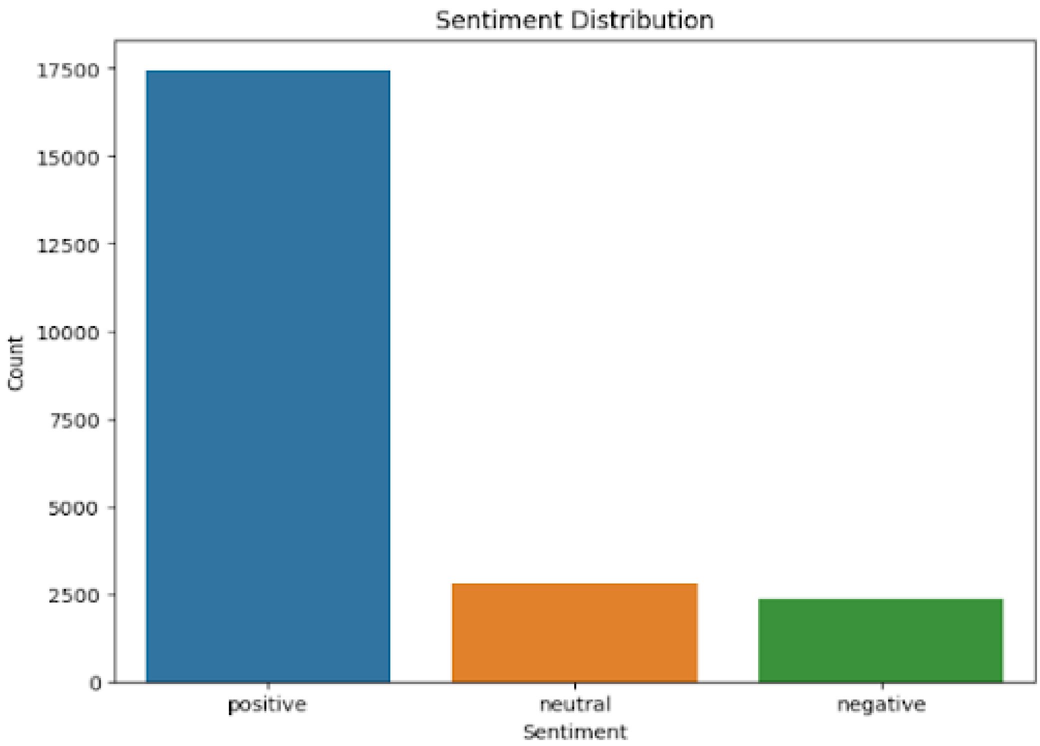

Figure 12: Bar Chart of Total Number of Sentiments

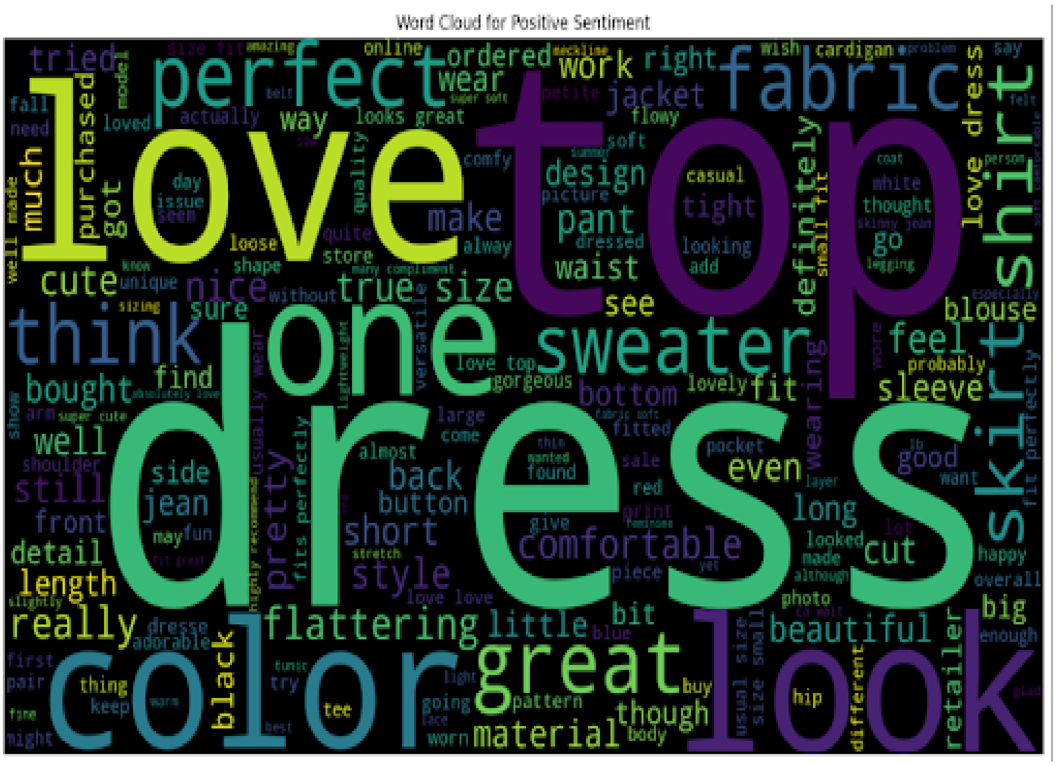

Figure 13: Word Cloud of Positive Sentiments

they write everything about the clothing which includes things such as fitting, comfort, material, design, etc.

Figure 14 provides the sentiment distribution based on the age group. This data shows

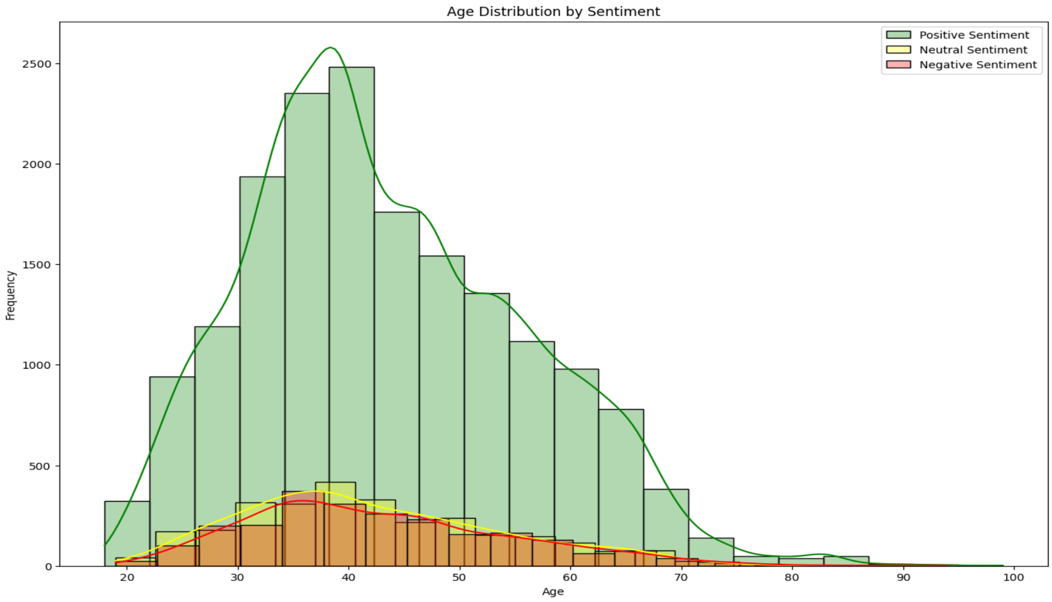

Figure 14: Age Distribution by Sentiments

robust evidence that women between the ages of 30 and 40 have more product reviews. A hypothesis can be made on this that women between the ages of 30 to 40 spend more on clothing in general than other age distribution groups.

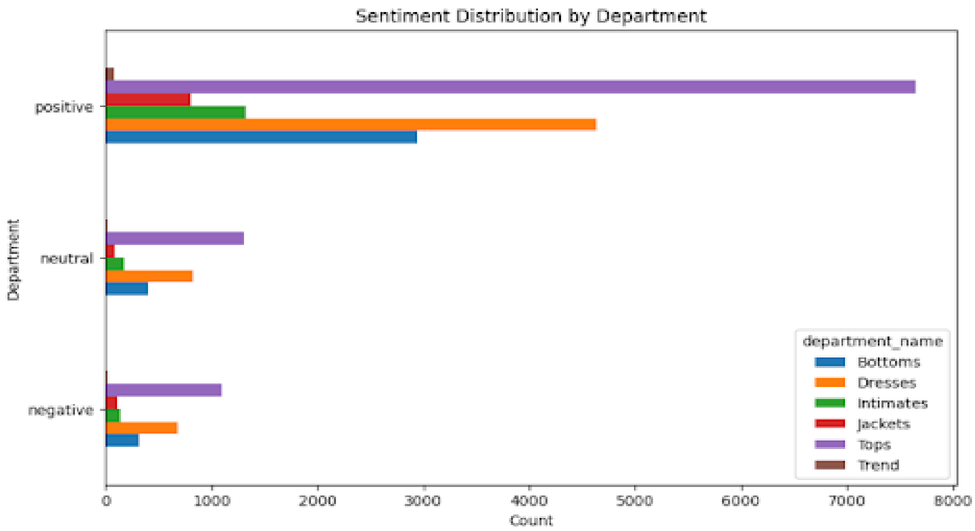

Figure 15: Sentiment Distribution by Department

Figure 15 provides insights into which department has more positive, negative, and neutral sentiments. This data is difficult to analyze using the figure only. Table 4 provides a tabular format of the data. Where the value of each sentiment by departments such as Bottoms, Tops, Dresses, etc. These insights can help some of the products in specific categories which has negative sentiments. Businesses and companies can benefit based on these insights. They can improve their products based on customer sentiments and improve product performance based on that.

Figure 16 provides sentiments based on the classes which include Dresses, Knits, Blouses, Pants, etc. These insights can help businesses and cloth manufacturers to provide validat-

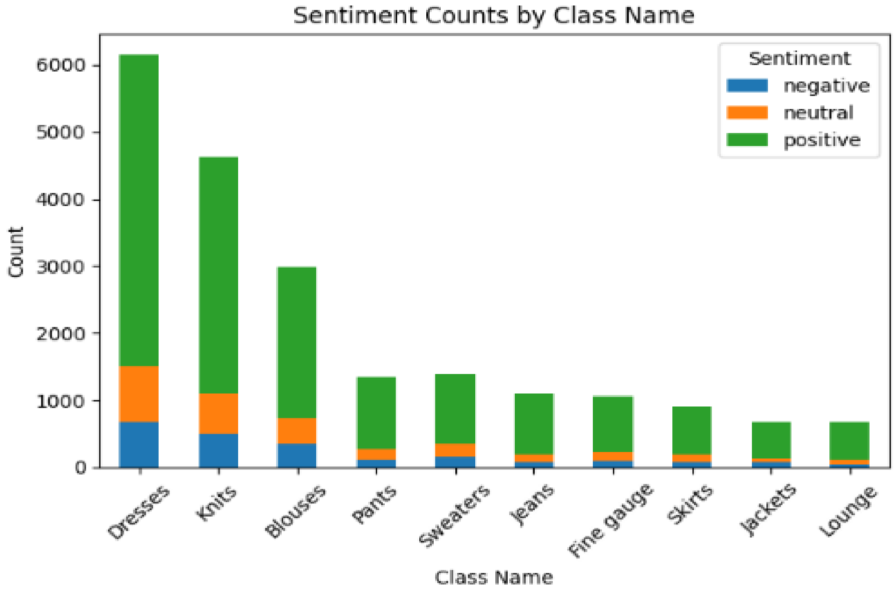

Figure 16: Sentiment Distribution by Class Name

Table 4: Sentiment Analysis by Department

Department	Sentiment	Count
Bottoms	Negative	317
	Neutral	407
	Positive	2938
Dresses	Negative	681
	Neutral	830
	Positive	4634
Intimates	Negative	147
	Neutral	177
	Positive	1329
Jackets	Negative	108
	Neutral	90
	Positive	804
Not Given	Negative	0
	Neutral	0
	Positive	13
Tops	Negative	1096
	Neutral	1300
	Positive	7652
Trend	Negative	21
	Neutral	19
	Positive	78

ing information about which products performing well in the market and which products are performing worst. Which can help them improve cloth quality of some of the clothing.

Figure 17 provides the review text length including stop words. The highest review text length is in negative sentiment which is 370 words. After doing more processing it can be

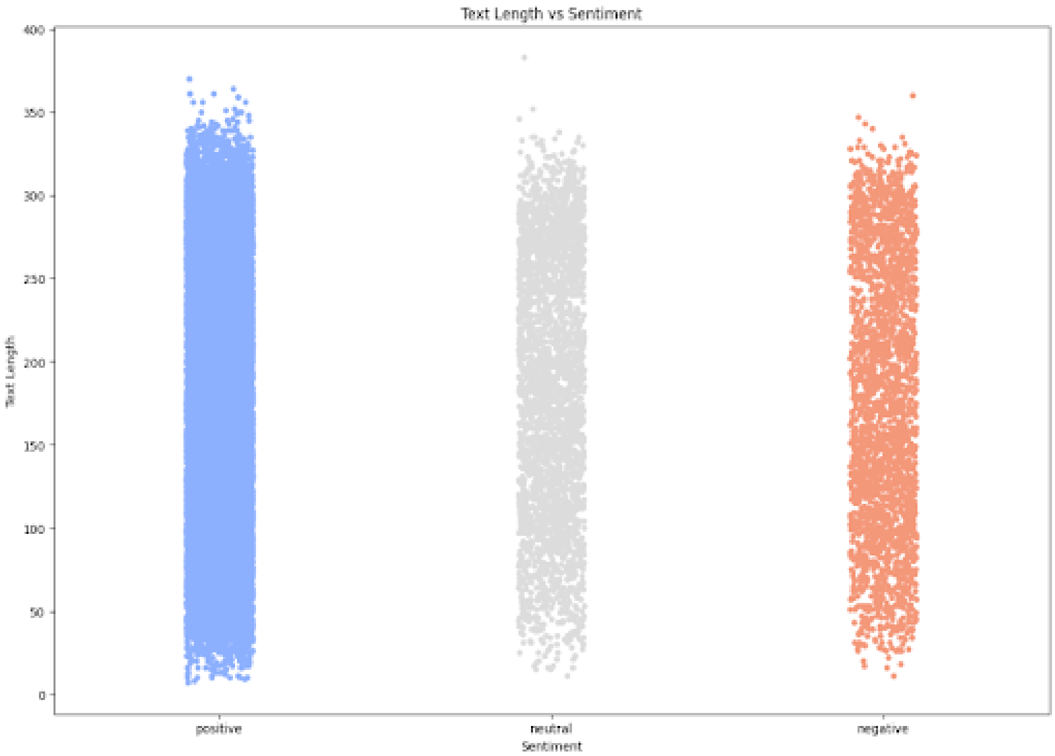

Figure 17: Sentiment Distribution by Class Name

seen in Table 5 where the average text length of neutral reviews is more than both positive and negative. The average length of neutral reviews is around 185 words. A hypothesis can be made based on these results that people like to write good and bad reviews which might be one of the reasons a neutral review's average number of words is more.

Table 5: Sentiment Scores

Sentiment	**Score**
Negative	179.458650
Neutral	185.168615
Positive	176.337861

7 Conclusion

It could be concluded that with the help of data analysis, businesses can make profits by making some strategic and informed decisions about whether to increase or decrease the price of the product, what inventory has to be stocked up and what not, which age of people should be targeted and what is the motive behind them that could be used to boost the sales and many more. Moreover, they can engage the audience by fixing certain budgets for the sake of Advertisements and promotional offers by analyzing association rules between the products and their purchase trends, etc. Many infinite possibilities can be inferred from this data analysis which could be used to generate profits running the business and competing in the centralized market.

8 Future work

Prospects for Sentiment Analysis of Women's Online Clothes Reviews in the Future:

- **More Complex Natural Language Processing (NLP) Methods:** Research in the future can concentrate on improving NLP methods for sentiment analysis through the use of more complex algorithms, like transformer models like GPT-4. By doing this, the system would be better able to comprehend and interpret the complex language used in reviews of women's clothing, thereby capturing nuanced feelings and context.

- **Multimodal Sentiment Analysis:** Combining sentiment analysis with image and video analysis may yield a more thorough understanding of consumer sentiment. This would include examining customer-posted images or videos in addition to text reviews to obtain a better understanding of how satisfied customers are with the clothing's perceived quality.

- **Fine-grained Sentiment Analysis:** By classifying sentiments according to more precise characteristics like fit, fabric quality, design, and customer service, sentiment analysis can be made even more granular. This will make it possible for companies to more precisely identify areas in need of improvement and raise customer satisfaction.

- **Contextual Understanding with Deep Learning:** Studies can examine the use of deep learning models created especially for this purpose. This would make it easier to understand the context in which particular feelings are expressed, lowering the possibility of misunderstandings and offering more precise information for business decision-making.

- **Real-time Sentiment Analysis:** Businesses could react quickly to customer feedback if they developed real-time sentiment analysis systems. This could entail putting in place streaming analytics and ongoing review monitoring, allowing businesses to promptly respond to complaints and take advantage of favorable sentiments.

- **Cross-language Sentiment Analysis:** By including reviews written in several languages, the analysis's scope can be expanded to serve a larger range of clients. For international e-commerce platforms, creating sentiment analysis models that can comprehend and evaluate sentiments expressed in various languages would be beneficial.

- **Bias mitigation and ethical considerations:** Future studies should concentrate on addressing moral dilemmas surrounding sentiment analysis, guaranteeing impartiality, and reducing biases. To do this, models that take cultural quirks into account must be created, and biases in the data must not be reinforced.

References

[1] T. K. Shivaprasad and J. Shetty, *Sentiment analysis of product reviews: A review*, In: *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Coimbatore, 2017, pp. 298-301, doi: 10.1109/ICICCT.2017.7975207.

[2] Fang, X., & Zhan, J. (2015). *Sentiment analysis using product review data. Journal of Big Data*, 2(1), 1-14. doi: 10.1186/s40537-015-0015-2.

[3] InData Labs. *Customer Review Sentiment Analysis to Unlock Customer Insights.* `https://indatalabs.com/blog/customer-review-sentiment-analysis`.

[4] Chauhan, C., & Sehgal, S. (2017). *Sentiment analysis on product reviews.* In: *2017 International Conference on Computing, Communication, and Automation (ICA)*, Greater Noida, India, pp. 26-31. doi: 10.1109/CCAA.2017.8229825.

[5] Pankaj, P., Pandey, Muskan, & Soni, N. (2019). *Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews.* In: *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Comicon)*, Faridabad, India, pp. 320-322. doi: 10.1109/COMITCon.2019.8862258.

[6] MonkeyLearn. *Sentiment Analysis & Machine Learning.* https://monkeylearn.com/blog/sentiment-analysis-machine-learning/.

[7] Hugging Face. *Getting Started with Sentiment Analysis using Python.* https://huggingface.co/blog/sentiment-analysis-python.

[8] GeeksforGeeks. *What is Sentiment Analysis?.* https://www.geeksforgeeks.org/what-is-sentiment-analysis/.

[9] TechTarget. *What Is Sentiment Analysis (Opinion Mining)? — Definition from TechTarget.* https://www.techtarget.com/searchbusinessanalytics/definition/opinion-mining-sentiment-mining.

[10] Coursera. *Demonstration: Case Study - Sentiment Analysis - Module 3: Classification of Text.* https://www.coursera.org/lecture/python-text-mining/demonstration-case-study-sentiment-analysis-MJ7g3.

9 Presentation

Presentation: Sentiment Analysis of Product Reviews (Presented in Class Tuesday, December 12th, 2023) Link of Presentation Embedded

10 GITHUB Repository Link

Link of GITHUB REPOSITORY Embedded