Unlocking Insights from Amazon Beauty Product Reviews using Big Data Analytics

[redacted]

ABSTRACT

In the rapidly growing e-commerce industry, online customer reviews of products have become increasingly important as they influence a customer's purchasing decision. The opinions expressed in these reviews can be crucial in shaping a customer's perception of a product. This paper proposes a distributed sentiment analysis and recommendation system for Amazon beauty products. Natural language processing techniques were used to extract product reviews from Amazon's website and analyze them using sentiment analysis algorithms. Sentiment analysis was performed on customer reviews and comments using VADER Sentiment Analyzer. The sentiment scores are then aggregated and used to recommend products to users based on their preferences. The sentiment analysis output is then used to create a custom recommendation system using two different methods: a Content-based recommendation system and a collaborative filtering recommendation system. Experimental results show that the proposed system achieves high accuracy in sentiment analysis and generates relevant and personalized product recommendations. The system can potentially improve the overall customer experience on Amazon and increase sales of beauty products.

1 INTRODUCTION

Since the rise of social media usage, the e-commerce industry has increased. Customers worldwide are placing orders internationally and locally, allowing customers to give feedback on websites that anyone with the internet around the world can access. These reviews are crucial in the decision-making of other customers willing to purchase the products.

This user-generated content is, therefore, valuable as it determines the future growth of sales and profits as they provide essential information to new potential customers. Apart from helping people gauge the quality of the product, reviews are also helpful for businesses to analyze this feedback, allowing them to improve their product quality and overall experience further. Therefore, there is a high demand for automatically analyzing large datasets containing millions of customer reviews.

There are various ways to do so, sentiment analysis being the most popular way of analyzing the review text. This technique uses natural language processing (NLP) which helps determine the sentiment or emotion depicted in the text. This assists businesses in quickly determining the sentiments related to each review, allowing them to determine the trends and areas for improvement.

Our report provides insights to buyers and sellers; therefore, we also focus on recommendation systems and sentiment analysis, which is helpful for sellers and allows the customers to get personalized recommendations since the recommendation systems use preferences, behaviour and product information like ratings. This also helps businesses since it increases sales as only recommended products match customers' tastes. There are two broad categories of recommendation systems: the content-based approach, which checks product similarity, and collaborative filtering, which allows businesses to make personalized recommendations.

This report uses Amazon beauty product reviews for our analysis. The reason for choosing this was because the beauty industry has flourished ever since the rise of the online marketplace. Amazon is a leading e-commerce business which deals with diverse beauty products; therefore, it is highly relevant and helps us analyze market trends. Furthermore, the Amazon data sets have high amounts of data, which is ideal for distributed computing.

The main objective of this report is to use PySpark for scalable distributed computing and employ various distributed computing techniques and models for sentiment analysis and recommendations. The structure of the report is as follows.

- Section 1: Introduction
- Section 2: Related Work
- Section 3: Methodology
  - Various big distributed techniques and models were used.
- Section 4: Numerical Results
  - Discussing the findings related to Exploratory Data Analysis, Sentiment Analysis and Prediction models, and Recommendation systems.
- Section 5: Conclusion

Furthermore, we use various models and systems for predicting the sentiment of the reviews and creating recommendation systems for customers. We compare and contrast these models for prediction and recommendation and suggest which method is more effective in different scenarios so that companies like Amazon can benefit from this helpful information and use these models to devise marketing strategies and improve the customer experience by analyzing the reviews efficiently such as using Linear SVC with cross-validation and parameter tuning and making the recommendations more personalized by using ALS.

This paper will help Amazon devise better strategies for their beauty products in the future using past data. The problem we want to focus on is how to benefit both the consumer and producer side. For producers, i.e. Amazon, we discuss various insights in our Exploratory Data Analysis, such as which are the most loyal customers, by analyzing which customer posted the most reviews or which products were trending each year and how these trends have evolved. Furthermore, sentiment analysis would help Amazon decide which factors are essential for customers during decision-making. We will help Amazon decide which prediction models for reviews would give them the more accurate results, such as the Linear SVC with cross-validation. For customers, we help Amazon decide which customers should be rewarded to increase customer satisfaction. Additionally, we have used two recommendation models for customers and discussed which recommendation model would be suitable in which scenario so that customers can have an ideal experience and convenient shopping experience. Therefore, the problem discussed in this paper is to make buying and selling beauty products a better experience.

2 RELATED WORK

"Big Data Analysis: Recommendation System with Hadoop Framework" by Jai Prakash Verma et al. 2015 This paper proposes recommendation systems such as collaborative filtering and content-based filtering for a large amount of data available on the web in the form of ratings, reviews, opinions, complaints, remarks, feedback, and comments about any item (product, event, individual and services) using Hadoop Framework.

"Sentiment analysis using product review data" by Xing Fang 2015. This paper proposes a general process for sentiment polarity categorization is proposed detailed process descriptions using online product reviews data collected from Amazon.com. Experiments were for both sentence-level categorization and review-level categorization.

"Sentiment analysis for Amazon.com reviews" by Levent G¨uner et al. 2019 In this study, different machine learning algorithms are compared, trained and tested on a dataset (N = 60, 000) containing product reviews from Amazon.com which are randomly selected from a dataset available from Kaggle containing 4 million reviews. The performance of three different algorithms was compared: Multinomial Naive Bayes (MNB), Linear Support Vector Machine (LSVM) and Long short-term memory network (LSTM) [5].

"Sentiments Detection for Amazon Product Review" by Bickey Kumar et al. 2021. This paper proposes a sentiment analysis for the Amazon products review dataset using machine learning algorithms such as Logistic Regression, Naive Bayes, and Random Forest.

Furthermore, much work has been done on customer reviews in the past, and those projects are readily available on GitHub, such as "Sentiment Analysis in Amazon Customer Reviews" and "Sentiment Analysis using Pyspark on Multi Social Media Data". These PySpark Jupyter notebooks have served us as a starting point to initialize our project [3].

3 METHODOLOGY

- **Data Collection and Preprocessing:** We have used the Amazon Customer Review dataset available on Amazon Customer Reviews Dataset (AWS). Each product category has a different dataset; however, we used the beauty category for our analysis. The URLs for all the datasets are provided in the link below: Amazon Customer Review. This dataset is then unzipped and placed in Hadoop for further processing. The missing fields were removed for analysis using filtering, an in-built PySpark function.
- **Feature Extraction and Representation:** For extracting the relevant features from the text of the reviews, PySpark's machine learning library `MLlib` was used. Furthermore, to create a fixed-size feature vector, we used `HashingTF` transformer, which decreases the dimensionality of the text data. Each vector index represents the number of times the word in the document is repeated. Furthermore, we also use `Inverse Document Frequency` (IDF) estimator to assign weights according to the importance of terms in the text data. These techniques helped us extract essential information from the review text.
- **Sentiment Analysis** We have used `VADER` (Valence Aware Dictionary and sEntiment Reasoner) tool to assign a sentiment score to classify each review as negative or positive in the dataset. For further analysis, we tokenize the reviews and remove the stopwords. We also discuss a confusion matrix, evaluation metrics, `SQL` queries and `WordClouds`.
- **Model Selection and Training for Prediction:** After text preprocessing such as lowercasing, tokenization, lemmatization and punctuation removal. To create frequency `TF-IDF` (term frequency-inverse document) matrix, `CountVectorizer` and `IDF` are used from `MLlib`. Then `Naive Bayes`, `Logistic Regression`, `Linear Support Vector Machine (SVC)`, are used for machine learning algorithms.
- **Model Tuning and comparison**: For tuning the parameters for each of the machine learning algorithms stated above, cross-validation and grid search are used. After that, each model is compared with the other using `ROC` (Receiver Operating Characteristic) scores.
- **Recommendations** Multiple recommendation systems such as content-based, popularity-based and collaborative filtering models are used. Content-based recommendation model makes use of cosine similarity, popularity-based uses weighted average ratings, and collaborative filtering uses `ALS` (Alternating Least Squares) algorithm to perform recommendations

For quicker data processing, we have used PySpark's inbuilt functions for Exploratory Data Analysis (EDA) purposes, especially `SQL`.

4 NUMERICAL RESULTS

4.1 Exploratory Data Analysis

The column names of the dataset are marketplace, customer id, review id, product id, product parent, product title|product category, star rating, helpful votes, total votes, vine, verified purchase, review headline, review body, and review date. The dataset is a tab-separated value (TSV) file, making it easier to analyze product performance and customer satisfaction to analyze trends over time. The dataset is from 2000 to 2016 and the market place is only restricted to the US. The total number of data is 5,115,666 and after dropping the missing values, the count drops to 5,114,733. The URL for the dataset is https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_us_Beauty_v1_00.tsv.gz.

4.1.1 Product Reviews. Figure 1 below shows the total number of reviews for each star rating and that most reviews have a rating of 5, followed by a 4-star rating. Lower ratings have a low number of reviews indicating that most reviews of beauty products are favourable.

It is essential to check which products have the most reviews and which are trending and most popular over the years. The mostly highly reviewd product is the HSI Professional hair straightner.

Furthermore, Table 2 shows the summary statistics for reviews per product. Some key takeaways include high standard deviation which shows dispersion of number of reviews since some products

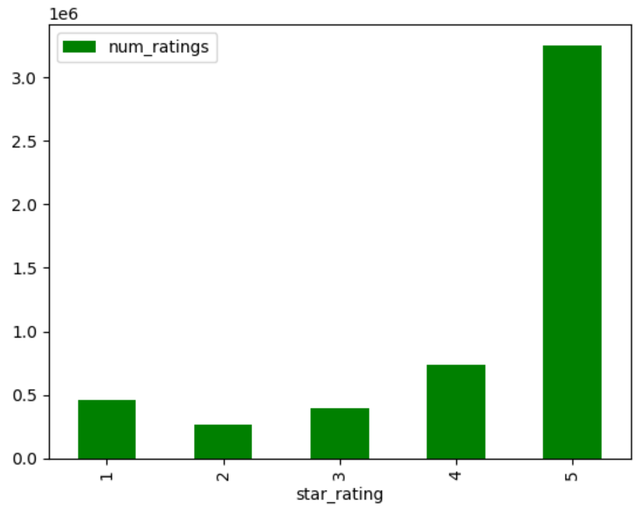

Figure 1: Number of Reviews per Rating

Table 1: Products with highest reviews

Product ID	Product Reviews
B001MA0QY2	15565
B000GLRREU	7272
B0049LUI9O	7166
B000JNQSIQ	5227
B003WR3QSG	5079

Table 2: Product reviews summary statistics

Summary	Product Reviews
count	588812
mean	8.69
stddev	53.1
min	1
25 percent	1
50 percent	2
75 percent	4
max	15565

only have one review while the others have more than thousand reviews. Around 75 percent of the products only have 4 or fewer reviews. A small mean with a relatively large standard deviation shows a that the distribution is skewed since only a small proportion of products have large number of reviews. Amazon should start promoting the products with low number of reviews for better consumer reach. Also, they can analyze the products with high number of reviews to determine what makes them so popular so that they can use those marketing strategies in these products.

Figure 2 shows that the total number of reviews increased rapidly over the years. Many factors might have influenced this, such as the rise of technology; customers could purchase from the online marketplace and submit their feedback. Furthermore, more people had access to smartphones, and social media usage and influence drastically increased during these years. This also shows that Amazon beauty products became more popular as years passed, i.e. beauty products gained more recognition.

Table 3: Customer summary statistics

Summary	Customer
count	2816349
mean	1.82
stddev	2.90
min	1
max	871

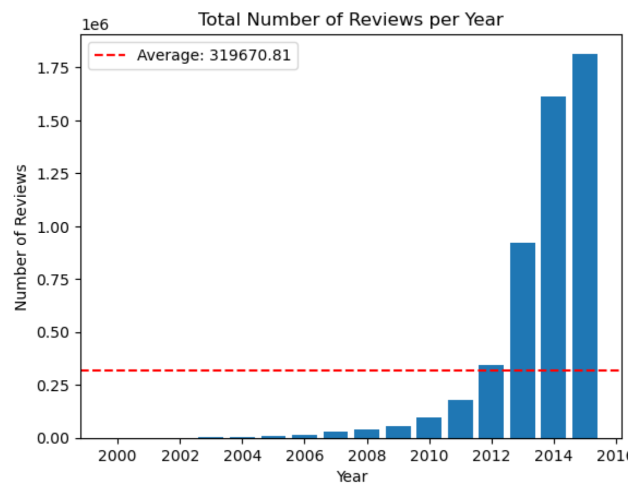

Figure 2: Number of reviews per year

4.1.2 Customer Analysis. Table 3 shows the summary statistics of the consumer base of Amazon beauty products. They have a considerable consumer base of 2,816,349; however, the average of reviews per customer could be higher. Amazon can use these insights to develop techniques encouraging customers to write valuable reviews. They can start by rewarding loyal customers.

One way to decide which consumers to reward for their contribution would be to compute which consumers have received the highest number of helpful votes from the audience reading those reviews, influencing their decision-making to purchase the product. Table 4 shows the customers who contributed the most.

4.1.3 Trends in beauty products over the years. In our Jupyter Notebook, we look at the most highly rated product for each year by calculating the average star ratings; however, it is not possible to choose only one product with the highest star rating since many products have 5.0 ratings, so the result would not be accurate. Therefore, we have chosen the most highly rated product, and if there are multiple products, we have chosen the product with the most

Table 4: Customer summary statistics

Customer ID	Helpful Votes
14425763	13364
18267561	10649
43552412	6478
51918881	5792
17538553	5751

reviews to determine the trend over the years.

The results show that it is evident that the popularity of the products changes every year as tastes and preferences fluctuate over time. This shift in product popularity is useful in devising market strategies. The feedback can also improve the product service in future years. Moreover, it is visible that product preferences have become more technological due to rapid advancements in technology. For example, in 2012, 2013, 2014 and 2015, people chose to opt for a more advanced hair straightener, "HSI Professional Glider," instead of purchasing traditional hair straighteners. Furthermore, this was the most highly reviewed product for consecutive years, from 2012 to 2015.

4.1.4 Time Series Analysis. Furthermore, we use the package `statsmodels` to decompose the time series data to check for any seasonality or trends observed over the years. The observed part displays how the average ratings change over the years in Figure 3. The trend graph depicts the pattern observed, which is increasing each year; therefore, more customers are satisfied with the quality of Amazon's beauty products. The trend graph eliminates the short-term non-meaningful fluctuations. There is no seasonality or unusual patterns; therefore, an upward trend is observed.

4.1.5 Verified vs non-verified reviews. There are around 4,230,196 reviews that have been verified, which comprises 82.69 per cent of the total reviews, while 885,256 reviews are unverified, which is around 17.31 per cent of the total reviews. Since most purchases are verified, this establishes trust and credibility for the general public. Therefore, customers can rely on these reviews since Amazon has confirmed the purchase. However, since almost one-quarter of the purchases are not verified, many reviews are not authentic. Thus, they might provide fake reviews, which can lead to customers buying the wrong products or Amazon losing sales. It is crucial to control these reviews since they can affect the customer's overall experience. In today's day and age, many stakeholders sponsor sites to write fake reviews [4].

4.2 Sentiment Analysis

Information is extracted from sentiment tokens and sentiment scores using the reviews. They are feature vectors used for sentiment categorization for training the classifiers, each entry of training data needs to be transformed to a vector that contains those features and it is imperative to control the dimensionality of these vectors [6].

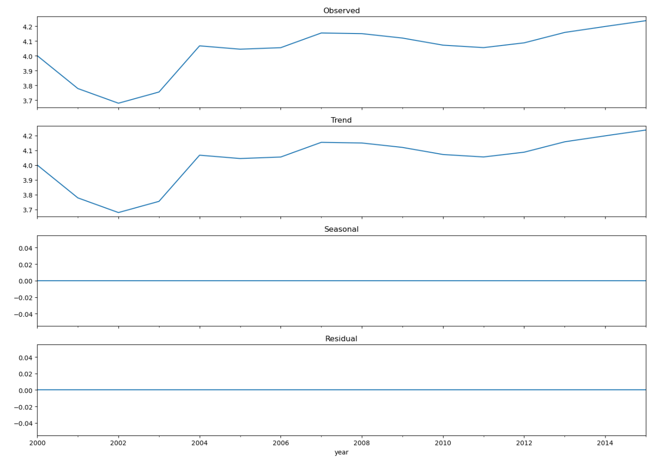

Figure 3: Seasonal Decomposition of Average Ratings per Year

Firstly, we analyse the distribution of the positive and negative reviews using the compound score of the VADER Sentiment Analyzer. The positive reviews are 3,995,612, which accounts for 78.12 per cent, and the negative reviews are 1,119,138, 21.88 per cent of the data set. This shows that most reviews are positive overall, which means that Amazon's beauty product customer satisfaction is generally remarkable.

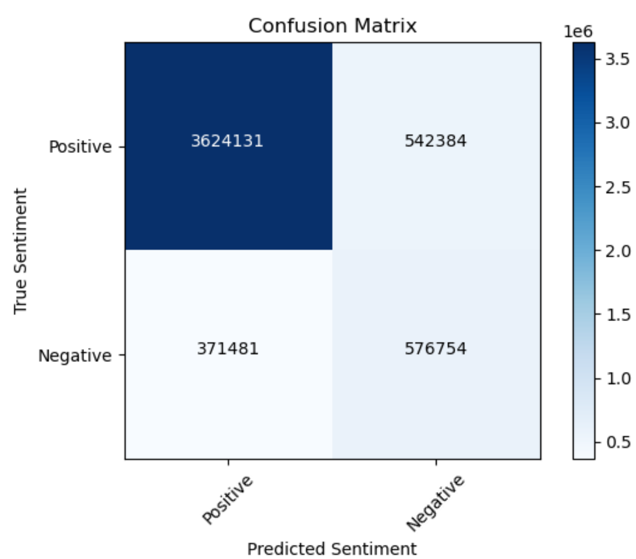

Figure 4: Confusion Matrix

4.2.1 Performance of VADER sentiment analyzer. Figure 4 represent the confusion matrix for performing the VADER sentiment analyzer. It shows the number of true positive, true negative, false positive and false negative sentiment analysis predictions conducted by this tool. It helps us understand which predictions are correct, which are true positives and true negatives, and the incorrect predictions

Table 5: Evaluation of `VADER Sentiment Analyzer`

Measure	Percentage
Accuracy	82.13
Precision	86.98
Recall	90.70
F1 Score	88.8

are false negatives and false positives.

The intensity of colours depicts the number of reviews in each category, and it is visible that there is a high number of true positives (3,624,131). Therefore, the sentiment analyzer performed well while predicting true positives. Boxes with the lightest colours are false positive and false negative. Therefore, these categories have the lowest amount of reviews which shows the good performance of this analyzer. The colour map is a good analysis method since it helps us visualize the performance efficiently. Since the colours of the top-left and bottom-right are darker than the other two cells, it is safe to assume that our analysis would be useful for Amazon.

For a more precise and numerical evaluation of our sentiment analysis model, we have computed the `accuracy, precision, recall and F1-score` based on the `confusion-matrix`. Table 5 above shows the high accuracy; therefore, this model correctly assigned the sentiment score of approximately 82.13 per cent of reviews. The 86.98 per cent precision shows that around 86.98 per cent of the reviews classified as positive by the model were positive. The recall of 90.70 per cent shows the actual positive reviews. We take F1 Score for a harmonic mean of precision and recall, which is useful for imbalanced data like ours (since there is a huge discrepancy between the number of positive and negative reviews). The high percentage of the F1 Score indicates the good performance of the sentiment analyzer prediction, which helps carry out further analysis.

4.2.2 Data Analysis. We employed SQL queries for our data analysis for sentiment analysis. First, we calculated the average sentiment for every product using the `GROUP BY` command in `SQL`. This will help Amazon to check the average sentiment score for any product they want to analyze.

Moreover, we use the review length to check the sentiment score for each length type. The classification is as follows:

- Very Short: reviews with less than 50 characters.
- Short: reviews with 50 to 99 characters.
- Medium: reviews with 100 to 499 characters.
- Long: reviews with 500 or more characters.

Long reviews have the highest sentiment analysis, and as the review length decreases, so does the average sentiment score, with concise reviews having the lowest average sentiment score; therefore, we can infer that review length is directly proportional to the average sentiment score as shown in Table 6.

Figure 5 shows that the average yearly sentiment score was high during 2020 and declined after that. From 2004 the average sentiment score increased and dropped in the later years. From 2012 to 2016, it reached its lowest score of 0.52, showing that Amazon beauty products got more negative reviews in the later years than in the initial years.

Table 6: Review Length vs Sentiment Score

Length	Average Sentiment Score	Number of Reviews
Long	0.71	627296
Medium	0.56	2886795
Short	0.47	625546
Very Short	0.43	975113

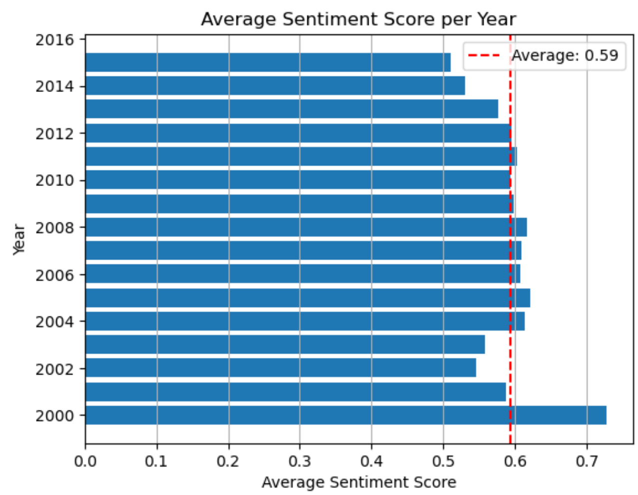

Figure 5: Average sentiment score per year

Table 7: Review Length vs Sentiment

Sentiment	Average Length
positive	234.81
neutral	280.64
negative	286.37

Furthermore, Table 7 shows that when the sentiment is positive or the sentiment score is high, then the average length is low; however, as the sentiment decreases or the review, i.e. the sentiment becomes neutral or negative, the average length of the review increases. Therefore, we can infer that longer reviews are more harmful than shorter reviews which are more favourable.

4.2.3 WordClouds. After tokenizing and splitting the review text, we separately created word clouds for positive and negative reviews to see the most frequently used words in both categories. Words like "product", "like", and "using" are common in 6 and 7. Surprisingly, 7 on Wordcloud have some positive words, meaning the negative reviews were not that harsh. Overall, positive words were used more than negative words, indicating that the customer experience has been above satisfactory. In positive reviews, "hair" and "skin" are used more frequently, which shows that mainly the beauty products reviewed were related to hair and skin. Other words such as "brush", "cream", "shampoo", and "soap" show that products related to skin and hair are the most popular.

Figure 6: WordCloud for Positive Reviews

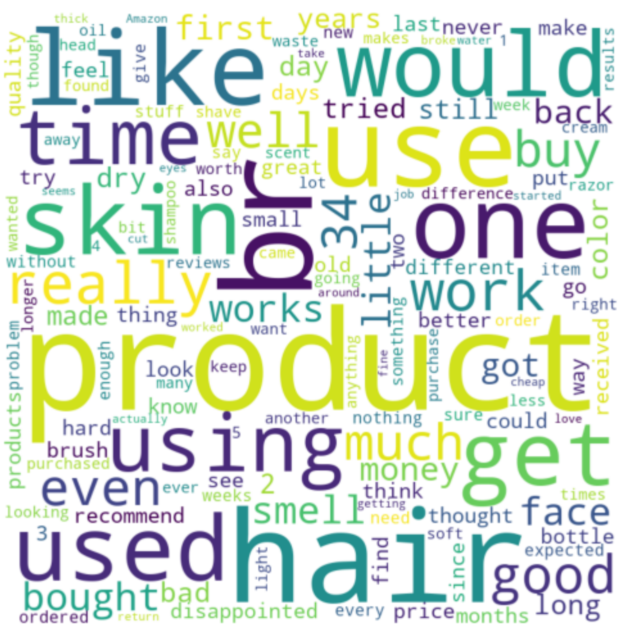

Figure 7: WordCloud for Negative Reviews

Also, the words "smell" and "scent" have been used in both negative and positive reviews frequently, which shows that how the product smells plays a massive role in customer decision-making. Figure 7 also includes "quality", "money", "broke", "dry", "price", "little", "small", and "old", showing that factors such as price, quality, quantity, expired products, and late order deliveries are the reasons why the products have been reviews negatively. On the other hand, Figure 6 has words like "easy", "natural", "clean", "pretty", "size", "colour", "smooth", and "soft", which shows that texture, quality, quantity and scent matter when purchasing the product. Moreover, customers prefer organic and clean products.

4.3 Prediction Models

In order to predict the sentiment of reviews, we have used various models such as Naive Bayes, Logistic Regression and Linear SVC using the `pyspark.ml.classification` library. We aim to select the best model using accuracy and ROC scores [1]. Much work has been done on this and this part makes use of various previous projects posted on GitHub, especially the "Sentiment Analysis in Amazon Customer Reviews" [7].

For Naive Bayes, we created a pipeline for feature extraction and Naive Bayes classifier training purposes. A similar pipeline was created for both Logistic Regression and Linear SVC. Each model has its advantage; Naive Bayes works the best with text data such as the review text. Logistic Regression gives interpretable results, while Linear SVC is the most effective for high-dimensional data.

After performing the machine learning algorithm for each model, cross-validation and grid search for fine-tuning the hyper-parameters for each model. This helps reduce the problem of over-fitting and provides more accurate predictions for new data by creating multiple subsets of the data and training the model on those. Grid search and parameter tuning provides a more accurate prediction and optimize the performance.

Figure 9 shows a minimal change in the accuracy of the cross-validation models. The accuracy for Naive Bayes and Linear SVC increases slightly. At the same time, Logistic Regression decreases, which shows that the model is robust and not influenced by the test and training data set. Therefore, these models work well with unseen data without cross-validation, and their performance is the same regardless of the choice of data, indicating that there was not a problem of over-fitting. The slight increase in the performance of the CV models might be due to hyper-parameter tuning using grid search, which optimizes the model to achieve better accuracy.

If we want to compare non-CV models with each other, we can see that Linear SVC performed the best again. The `Receiver Operating Characteristic` (ROC) evaluates the model performance by plotting the actual positive rate against the false positive rate. Therefore, Linear SVC has the highest ROC with the highest actual positive rate and the lowest false positive rate. It also has the highest accuracy. Therefore, if using cross-validation and tuning parameters is time-consuming or uses many resources, we suggest using Linear SVC. Otherwise, Linear SVC with cross-validation and hyper-parameters tuning should be considered for optimal results.

4.4 Recommendation Systems

In this section, we will discuss three recommendation systems to discuss their pros and cons and which one is more suitable in which scenario.

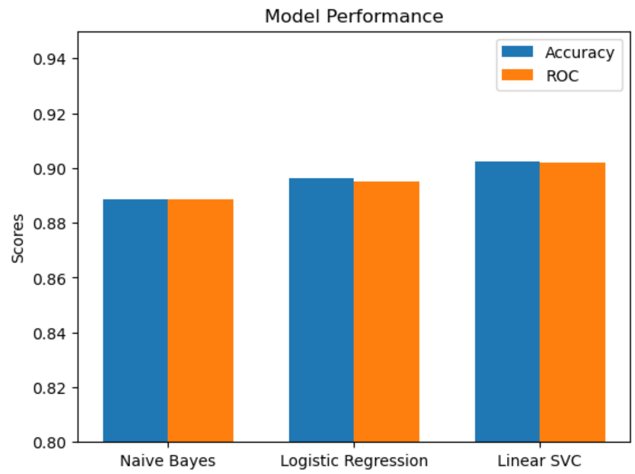

Figure 8: Model Performance of non-CV Models

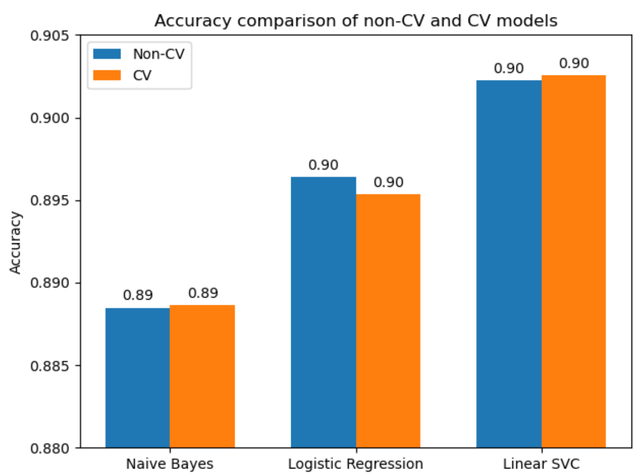

Figure 9: Accuracy comparison of non-CV and CV Models

4.4.1 Content-based recommendation system. This filtering method uses the product description and keywords i.e. reviews to compare which products are similar to each other for generating recommendations [4].

Cosine-similarity is calculated for term frequency-inverse document frequency (TF-IDF) vectors for each product which helps determine the similarity between the reviews for each product. Therefore, show similar reviews if the cosine similarity is high between the two products. Cosine similarity is used as a proxy for product similarity since Amazon can recommend those products to their customers if the reviews are similar. The cosine similarity ranges from -1 to 1. A value of 1 shows a high similarity between reviews, and -1 indicates that reviews are not similar. This is a highly adaptable recommendation system since the product can be changed by inputting a different product ID. So the customers get a highly personalized recommendation by just inputting the product ID. Furthermore, this recommendation system is highly efficient since we use sparse vectors like TF-IDF. However, for a more accurate recommendation system, we use a collaborative filtering model such as ALS in the next section.

Table 8: Customer summary statistics

Product ID	similarity
B00N30SE3E	0.9757761
B00B76ALVO	0.9543328
B004INQ65S	0.94657975
B0009XH6SC	0.93763614
B00BB8ZIRK	0.93135357
B007ZXR4I6	0.92609286
B000R8715M	0.9225896
B002LIXTFC	0.9220803
B00497BELY	0.919404
B008U5XS2O	0.9184702

Furthermore, we have used the HSI Hair straightener, which has the highest reviews and ratings for recommending products more products like it. This would help Amazon promote those products since they also have the potential to be as popular as the HSI Hair straightener. The Product IDs and their cosine similarities are provided in the Table 8.

4.4.2 Collaborative filtering recommendation system. For this method, we have used Alternating Least Squares (ALS) which is a matrix factorization algorithm. This is a more advanced method which gives personalized recommendations for all users. Then top 10 products are recommended for each user. The output can be found in the Jupyter notebook. For example the output looks like:

(31, [296429, 16.328068, 483865, 15.901779, 286438, 15.740509, 202136, 15.664841, 201398, 15.634038, 254428, 15.38573, 242095, 15.29113, 149967, 15.233497, 304100, 15.197423, 276113, 15.099527]).

This is the first row of the output where 31 represents the customer and first element in the tuple represents the product and the second part represents the ratings this user will give to this product which can be above 5 according to ALS. The parameters used for model training was rank to 10 which represents the dimensions in embedding space which is useful in model performance, maxIter to 10 shows the number of iterations, Also regParam to 0.1 to prevent over-fitting. Since we are setting coldStartStrategy to "drop", this model will not take into account new customers, therefore, it is not useful for making recommendations for new customers. The RMSE of this model is 1.76 which is high since the star ratings range from 1 to 5. Therefore, for future purposes it is recommended to change the parameters to achieve better accuracy.

5 CONCLUSION

In conclusion, this paper has presented a distributed sentiment analysis and recommendation system for Amazon beauty products that leverage machine learning models such as logistic regression, Naive Bayes, and Linear SVC to predict the sentiment of customer reviews.

Our analysis of the sentiment classification model has shown high accuracy, precision, recall, and F1-score, indicating its effectiveness in predicting sentiment. We also determined that factors like price, scent, size and quality play a hige role in purchasing products using wordclouds.

Furthermore, we have explored various recommendation systems, which are content-based, and collaborative filtering recommendation systems. The content-based recommendation system uses cosine similarity on TF-IDF vectors of product reviews to recommend similar products to customers. This method provides highly personalized recommendations by just inputting a product ID and is highly efficient. The collaborative filtering recommendation system, which uses the ALS algorithm, gives personalized recommendations for all users. However, this method only suits some customers and requires parameter optimization for better accuracy. Each recommendation system has advantages, and selecting the best system ultimately depends on the specific use case and available data. We suggest recommendation systems based on different scenario and Linear SVC with cross-validation as the best prediction model.

Our research highlights the importance of sentiment analysis and recommendation systems in the e-commerce industry, especially for online retail platforms like Amazon. By providing personalized recommendations based on customer sentiment, our proposed methods have the potential to improve the user experience and increase customer satisfaction. The sentiment analysis model employed in this study demonstrated a high level of accuracy, precision, recall, and F1-score, correctly assigning sentiment scores to approximately 82.13of reviews. The F1-score, useful for imbalanced data like ours, showed a high percentage indicating good performance. These results provide a solid foundation for further analysis and support the effectiveness of the sentiment analysis model.

Future research can extend this method to other product categories and explore additional features such as user demographics and social media data to enhance the accuracy of sentiment analysis and recommendation systems. However, like any other project, our project also has limitations. For example, this data is from 2000 to 2016, which means that it does not have the most recent data, which would mean that the results might not be highly applicable today. Furthermore, the marketplace is only restricted to the US; therefore, the analysis might not apply to other parts of the world [2].

REFERENCES

[1] Anshul Shroff Ashutosh Kumar Dixit Bickey Kumar Shah, Amar Kumar Jaiswal. 2021. Sentiments Detection for Amazon Product Review. *ICCCI* (2021). https://doi.org/10.1109/ICCCI50826.2021.9402414

[2] ChatGPT. [n. d.]. OpenAI. ([n. d.]). https://doi.org/

[3] Chait Hanya. 2020. Sentiment Analysis using Pyspark on Multi Social Media Data. *GitHub* (2020). https://doi.org/chaithanya21/Sentiment-Analysis-using-Pyspark-on-Multi-Social-Media-Data

[4] Atul Patel Jai Prakash Verma, Bankim Patel. 2015. Big Data Analysis: Recommendation System with Hadoop Framework. *IEEE International Conference on Computational Intelligence Communication Technology* (2015). https://doi.org/10.1109/CICT.2015.86

[5] Lamia Hadrich Belguith Sarsabene Hammi, Souha Mezghani. 2022. An Improved Hybrid Method for Sentiment Analysis. *Conference: An Improved Hybrid Method for Sentiment Analysis* (2022). https://doi.org/10.1109/INISTA55318.2022.9894143

[6] Justin Zhan Xing Fang. 2015. Sentiment analysis using product review data. *Journal of Big Data* (2015). https://doi.org/10.1186/s40537-015-0015-2

[7] Jesseq Zhen. 2018. Sentiment Analysis in Amazon Customer Reviews. *GitHub* (2018). https://doi.org/jesseqzhen/NLP_Sentiment_Analysis_-Spark-

[redacted]