The relationship between the presence of positive keywords in reviews and the overall rating of products on Amazon

Dongyeon Kang, Sunwoo Kim

*College of Information Science and Technology*

*Pennsylvania State University*

dkk5326@psu.edu

szk6076@psu.edu

*Abstract*— This study investigates the relationship between the presence of positive keywords in online product reviews and the overall rating of products on Amazon. With the proliferation of e-commerce platforms, understanding the impact of consumer sentiment on product ratings has become increasingly important for businesses. The research question addressed in this study is: What is the relationship between the presence of positive keywords in reviews and the overall rating of products on Amazon? The study employs text mining techniques, sentiment analysis, and statistical analysis to explore this relationship. By identifying prevalent themes in reviews, scoring keywords based on sentiment, and analyzing the relationship between positive keywords and ratings, this research aims to provide insights into consumer behavior and product perception on online platforms.

*Keywords*— Online reviews, Sentiment analysis, Product ratings, E-commerce platforms, Consumer behavior, Text mining

1. INTRODUCTION

As the world increasingly pivots to the digital sphere, the role of customer reviews on e-commerce platforms like Amazon has transcended beyond mere opinions to become critical factors influencing market trends and consumer behavior. This study probes into the quantifiable impact of positive keywords in customer reviews on the overall product ratings, spotlighting the transformative effect of consumer sentiment in the domain of online shopping.

With the surge in digital marketplace activity, customer reviews have attained a multifaceted significance. They are not only feedback for the sellers but also serve as guideposts for potential buyers in a densely populated market of choices. Positive reviews, characterized by affirming keywords, contribute more than accolades; they have the potential to elevate a product’s prominence in a fiercely competitive landscape. At the intersection of data analytics and consumer psychology, this research adopts a systematic methodology to dissect the influence of positive sentiments in customer reviews on product ratings.

Embarking on this analytical journey, our study recognizes online purchasing as a mainstream habit, making the dissection of customer reviews an indispensable tool for deciphering the language of the digital economy. It is based on the premise that the nuances in reviews reflect consumer perceptions that, once decoded and quantified, can unveil significant market patterns.

Leveraging the extensive repository of user-generated content on Amazon, this paper harnesses sentiment analysis, statistical modeling, and data visualization to explore the nexus between positive review keywords and product ratings. By analyzing real user feedback, we aim to elucidate the mysterious sway of review sentiments on product standing.

The paper is methodically structured to commence with an exhaustive review of related works, providing the scholarly framework for our exploration. Subsequently, we delve into the methodological intricacies, illuminating the process of data curation and sentiment analysis that fortifies our model. The study culminates with a thorough examination of the findings, leading to a conclusion that encapsulates our insights and contemplates future research pathways within consumer analytics.

This investigation is a deep dive into the psyche of the consumer, interpreted through the lens of data analysis to deliver a wider comprehension of its influence on business success. It is a foray into the essence of consumer thought, articulated into valuable knowledge for vendors and consumers in the ever-evolving arena of online commerce.

II. Related Works

The burgeoning field of consumer behavior analysis within the domain of e-commerce has witnessed the advent of novel techniques for extracting and utilizing consumer insights. Traditional approaches such as surveys and focus groups have long served as the pillars for discerning consumer perceptions, enabling brands to visualize these insights through perceptual maps (Ries & Trout, 2013; Kotler & Keller, 2018)​​. However, inherent limitations of these methods, including questions about the validity of consumer responses and the risks of drawing conclusions from potentially uninformed perspectives, have catalyzed the search for more robust and authentic analytical methods (Bas¸, 2015)​​.

This quest for precision has led to the embracement of opinion mining and sentiment analysis—a paradigmatic shift powered by text mining advancements that leverage consumer-generated content on digital platforms (Cambria et al., 2010; Hardeniya et al., 2016)​​. The digital era has drastically altered the dynamics of customer feedback, with e-commerce reviews emerging as a critical resource for brand positioning strategies (Lackermair et al., 2013; Filieri & Mariani, 2021)​​. By sifting through vast quantities of such consumer-generated data, opinion mining presents a methodological transformation that provides a more nuanced and contemporaneous representation of brand positioning in the consumer's psyche.

Several studies have underscored the efficacy of machine learning and lexicon-based methods in opinion mining (Taboada et al., 2011; Rajput et al., 2016)​​. These methods have demonstrated their potency in capturing the sentiment of consumer reviews, with applications spanning various sectors from technology products to political opinion (Lee et al., 2015; Kızılkaya, 2018)​​. This research presents an innovative fusion of opinion mining with perceptual mapping, employing lexicon-based sentiment analysis to translate consumer reviews into quantifiable data that can be represented visually through perceptual maps and radar charts (Ahuja et al., 2019; Yun-tao et al., 2005)​​.

The findings of this approach have illustrated its distinctive value. By effectively converting qualitative data from textual reviews into quantitative scores, the study by Yılmaz and Altunay (2023) has successfully positioned brands on perceptual maps, revealing the comparative stance of brands as perceived by consumers in terms of various product attributes. This methodology extends beyond mere sentiment polarity, delving into the sentiment scores assigned to each product feature, thereby offering a granular analysis of consumer perceptions (Saglam et al., 2019)​​.

Future explorations in this realm could delve deeper into the demographic nuances of reviewers, analyze sentiment across multiple languages, and possibly integrate real-time updates to consumer perceptions, thus continually refining the accuracy and relevance of brand positioning strategies (Daabes & Kharbat, 2017; Chang et al., 2020)​​. The promise of this analytical fusion in providing strategic marketing intelligence underscores its potential as a cornerstone in the evolution of brand positioning research.

III. Model Implementation and Methods

The model implementation and methods section of our research paper is underpinned by an extensive analysis of a dataset containing 1000 ratings and reviews sourced from Amazon. This dataset serves as the foundational basis for our investigation into the correlation between the presence of positive keywords in reviews and the overall rating of products. Leveraging this dataset, our methodology employs a systematic approach to unraveling the intricate dynamics between consumer sentiments and product ratings in the context of online retail platforms. Through rigorous data preprocessing techniques, including text mining methodologies such as Term Frequency-Inverse Document Frequency (TF\*IDF) and Latent Dirichlet Allocation (LDA), ath aim to uncover pivotal insights into consumer behavior and preferences. Furthermore, the dataset is meticulously partitioned into training and testing sets to facilitate subsequent model selection and evaluation processes. By meticulously executing these methodologies, our research aims to figure out the underlying mechanisms driving consumer decision-making processes in the realm of e-commerce.

1. Dataset

The dataset comprises over 1,000 Amazon product ratings and reviews sourced from the official Amazon website. It includes various features such as product ID, name, category, discounted and actual prices, discount percentage, ratings, number of ratings, product description, user ID, username, review ID, review title, review content, image link, and product link. This dataset serves as a valuable resource for research in the field of e-commerce analytics, sentiment analysis, recommendation systems, and customer behavior analysis. For this research, we will primarily focus on specific features such as product ratings, review content, and possibly discount information. By analyzing the relationship between the presence of positive keywords in reviews and overall product ratings, we aim to gain insights into consumer sentiment and its impact on purchasing decisions in the e-commerce domain.

1. Preprocessing

The data preprocessing phase is the unsung hero of our analytical endeavor, transforming raw, unstructured text into a distilled form ripe for insight extraction. Upon procuring the dataset from a reputable web source, we initiated a scrupulous cleaning regimen. This initial stage of purification was instrumental, excising textual anomalies such as punctuation and special characters. To ensure a level analytical playing field, we applied a blanket of lowercase uniformity across the dataset, thereby stripping away any capitalization biases that could skew our interpretation.

The march towards clarity continued as we engaged in the meticulous removal of stopwords. This critical excision not only quelled the noise within the corpus but also homed in on the most sentimentally potent elements of language, those with the power to reflect genuine customer sentiment. By casting aside these linguistic filler items, we enhanced the dataset's expressiveness, facilitating a laser-focused sentiment analysis to follow.

Our methodological rigor persisted unabated. With tokenization, we deconstructed complex textual structures into individual words, the basic units of meaning. In pursuit of uniformity and precision, lemmatization was next to be applied. It methodically streamlined these tokens to their base or root forms, stripping away inflectional endings and conflating variants into a singular representation. The result? A refined and standardized dataset that forms the bedrock of subsequent analysis.

Moreover, in tandem with these text mining techniques, we integrated the TF\*IDF statistical measure to discern the importance of terms within the corpus. This approach not only prioritized features with significant analytical value but also shaded our perspective on the keywords that shape customer sentiment. Concurrently, Latent Dirichlet Allocation (LDA) for topic modeling unveiled prevalent themes, thus illuminating hidden patterns and sentiments laced within the reviews.

Finally, we filtered out the static of reviews brimming with solely positive sentiments to maintain a balanced analytical lens. This critical filtration step ensured that our exploration captured a diverse range of emotional expressions, setting the stage for a comprehensive assessment of the nuanced interplay between keyword sentiment and overall product ratings. Through this intricate dance of data preparation, our study stands on a foundation of clean, consistent, and contextually enriched data, poised for the rigorous application of sentiment analysis and model selection that followed.

1. Sentiment Analysis and Keyword Scoring

Entering the realm of sentiment analysis, our approach was twofold: identify emotionally significant keywords and quantify their impact. Tools like VADER sentiment analysis granted us a quantitative measure of the sentiment valence for keywords such as 'good,' 'great,' and 'easy.' These terms stood out, not merely for their frequency but for the emotional resonance they held within customer reviews.

1. Mapping Keywords to Product Ratings

In bridging the lexical with the empirical, we meticulously mapped these sentiment-rich keywords to their occurrence in reviews and the resultant product ratings. This stage was pivotal, revealing an unexpected disconnect: keywords with substantial sentiment scores did not uniformly translate to high product ratings. Thus, we uncovered a more complex interplay at play—a multifaceted relationship between sentiment, keyword frequency, and customer satisfaction, hinting at deeper influences on consumer decision-making.

1. Model Selection

In this research on predicting product ratings based on the presence of positive keywords in reviews, we have chosen to employ Support Vector Machine (SVM) and Random Forest models as the supervised learning classification algorithms. SVM is a robust and widely used algorithm known for its effectiveness in handling high-dimensional data and nonlinear relationships. It works by identifying the optimal hyperplane that separates different classes in the feature space, making it suitable for capturing complex relationships between input features and target labels. SVM's ability to generalize well to unseen data makes it a strong candidate for accurately predicting product ratings based on the presence of positive keywords in reviews. On the other hand, Random Forest is an ensemble learning algorithm that constructs multiple decision trees and aggregates their predictions. It is particularly well-suited for handling noisy data and large datasets, making it an ideal choice for this research. By leveraging the strengths of both SVM and Random Forest, we aim to develop robust models that can effectively capture the nuances of consumer sentiments expressed in reviews and accurately predict product ratings.

1. Support Vector Machine

SVM is a powerful supervised learning algorithm chosen for this research on predicting product ratings based on the presence of positive keywords in reviews. SVM excels in dealing with high-dimensional data, making it well-suited for analyzing the diverse set of features extracted from product reviews. By identifying the optimal hyperplane that separates different classes in the feature space, SVM can effectively capture the complex relationships between the presence of positive keywords and product ratings. Its ability to generalize unseen data ensures robust performance in predicting ratings for new products based on their reviews, thus providing valuable insights for consumers making purchasing decisions.

1. Random Forest

Random Forest is another key algorithm selected for our research due to its ability to handle noisy data and large datasets, making it suitable for analyzing the extensive collection of product reviews available in our dataset. As an ensemble learning method, Random Forest constructs multiple decision trees and combines their predictions to generate accurate results. This approach allows Random Forest to capture the varied sentiments expressed in reviews and make reliable predictions about product ratings. By leveraging the ensemble of decision trees, Random Forest can effectively identify important features and interactions between positive keywords and product ratings, providing valuable insights into consumer preferences and behaviors in the e-commerce domain.

IV. Performance Evaluation Metrics

In our research paper, we will adapt the performance evaluation metrics based on the context and objectives of our study. Given that our research focuses on analyzing the relationship between the presence of positive keywords in reviews and product ratings on Amazon, our performance evaluation metrics will be tailored to assess the effectiveness of our machine learning models in predicting product ratings accurately. We can modify the performance evaluation metrics as follows:

1. Accuracy:

Accuracy measures the proportion of correctly predicted product ratings out of the total product ratings in our dataset. It is essential in our research as it provides an overall indication of how well our machine learning models are performing in predicting product ratings based on the presence of positive keywords in reviews. Evaluating accuracy is crucial because it gives us insight into the overall effectiveness of our models. A high accuracy score indicates that our models are successfully capturing the relationship between positive keywords and product ratings, which is essential for making reliable prediction.

1. Precision:

Precision measures the proportion of true positive predictions among all positive predictions made by the model. In our research, precision reflects the ability of our models to accurately identify positive product ratings based on the presence of positive keywords in reviews. Precision is important because it helps us understand the reliability of our models in identifying positive product ratings. A high precision score indicates that our models are making fewer false positive predictions, which is crucial for ensuring the credibility of our predictions.

1. Recall:

Recall measures the proportion of true positive predictions among all actual positive product ratings in the dataset. In the context of our research, recall assesses the ability of our models to capture all positive product ratings accurately. Recall is important because it indicates how well our models are capturing positive product ratings. A high recall score suggests that our models are effectively identifying positive ratings, which is essential for ensuring comprehensive coverage of positive sentiments in the dataset.

1. F1 Score:

The F1 score is the harmonic mean of precision and recall, providing a balance between the two metrics. It evaluates the overall performance of our models in predicting product ratings based on positive keywords in reviews. The F1 score is important as it provides a comprehensive assessment of our models' performance, considering both precision and recall. It helps us understand the trade-off between making accurate positive predictions and capturing all positive ratings. A high F1 score indicates a good balance between precision and recall, leading to reliable predictions of product ratings based on positive keywords in reviews.

In this research, the confusion matrix will serve as a pivotal tool for evaluating the effectiveness of our machine learning models in predicting product ratings based on the presence of positive keywords in reviews. By analyzing the confusion matrix, we can assess how well each model performs in classifying instances of positive ratings, distinguishing between true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). This breakdown allows us to gain insights into the models' strengths and weaknesses, particularly in identifying positive sentiments accurately. Additionally, we will leverage performance evaluation metrics described to quantify the performance of each model. These metrics are crucial for ensuring that our models effectively capture the relationship between positive keywords in reviews and overall product ratings on Amazon, thereby guiding our analysis and prediction tasks effectively.

1. Visual Analysis and Insight Integration

As a vital component of our research, we delved into visual analytics to unveil patterns and relationships not immediately apparent in raw data. Here are the insights derived from each image:

A graph with blue dots

Description automatically generated

The scatter plot presented herein illustrates an intricate narrative contrary to the conventional hypothesis positing a direct correlation between the frequency of positive lexicon within product reviews and the resulting product ratings. The distribution of individual data points serves to refute the premise that an increased prevalence of affirming language unequivocally equates to heightened customer satisfaction. Instead, a discerning analysis reveals a sophisticated spread of ratings, affirming the thesis that customer verdicts are predicated upon a complex amalgam of factors.

A graph with colorful squares

Description automatically generated

The bar chart rendered here encapsulates the essence of consumer sentiment as manifested through the frequency of specific positive terms in feedback provided by customers. The prominent recurrence of the term 'good' within this visual depiction calls into question the depth and genuine sentiment it purportedly conveys. This observation challenges us to contemplate whether its pervasive presence may, in fact, attenuate its sentiment value, prompting a more nuanced exploration into the intricacies of linguistic expressions of sentiment in consumer feedback.

A graph with blue dots

Description automatically generated

The scatter plot juxtaposes the aggregated sentiment scores—derived from the combined metrics of term frequency and sentiment valence—against the mean product ratings to examine the underlying association. Notably, certain terms endowed with substantial sentiment scores, such as 'good', do not exhibit a consistent correlation with superior product ratings. This counterintuitive revelation impels a pivot towards a more multifaceted sentiment analysis, transcending the simplistic dichotomy of positive term frequency.

The visual analyses transcend their role as mere illustrations, operating instead as pivotal empirical evidence that informs the methodological compass of this research. They refine our analytical lens and bolster the empirical underpinnings of the study's conclusions, ensuring that the insights garnered from this investigation are grounded in quantitative substantiation.

V. Conclusion and Future Work

Our incisive exploration of the relationship between the prevalence of positive keywords in Amazon product reviews and their corresponding product ratings has unraveled a complex narrative. Through a blend of advanced text mining and rigorous sentiment analysis, our study has disrupted the conventional narrative that equates the abundance of positive language with heightened customer satisfaction. The counterintuitive findings suggest a more intricate web of factors that contribute to product evaluations, transcending beyond the straightforward presence of affirmative language.

The visual and computational analyses have offered a wealth of empirical evidence, illustrating that the frequency of positive terms and their sentiment scores do not uniformly predict product ratings. These outcomes indicate a sophisticated landscape where consumers' perceptions are shaped by diverse considerations, hinting at the rich tapestry of individual experiences and expectations.

Diving deeper, our analytical journey through the world of consumer feedback has revealed that terms with substantial sentiment scores do not necessarily align with the highest product ratings. This observation underscores an essential insight: the expressions of customer satisfaction are complex and layered, eluding simple quantification by sentiment metrics alone. Therefore, it stands to reason that businesses and e-commerce platforms should seek to understand the contextual nuances and the multifaceted nature of customer reviews.

As we contemplate the road ahead, the next phase of research beckons with the promise of new methodologies and broader horizons. We propose a future research agenda that includes:

1. **Contextual Language Analysis**: Investigating the impact of context on sentiment in reviews. Future studies could utilize context-aware sentiment analysis algorithms to capture the subtle connotations of consumer language.
2. **Temporal Dynamics**: Monitoring how sentiment and keyword significance fluctuate over time, particularly in response to market trends or product life cycles.
3. **Linguistic Features Impact**: Assessing how negations, intensifiers, and emoticons alter the emotional tone of reviews and their influence on consumer decision-making.
4. **Technological Advancements**: Leveraging deep learning and natural language processing to unravel the complex emotional fabric woven into customer feedback. Techniques such as sentiment neuron networks could be pivotal in advancing this line of inquiry.
5. **Cross-Cultural and Multilingual Analysis**: Broadening the dataset to include multilingual reviews and diverse product categories, thus enhancing our understanding of the global consumer sentiment landscape.

This comprehensive research has not only contributed a substantial chapter to the annals of academic knowledge but also delineated actionable strategies for e-commerce stakeholders. The insights herein should inspire businesses to venture beyond the facile analysis of positive term frequency and to resonate authentically with the refined sentiment expressions of their customers.

In conclusion, this study acts as a clarion call for an integrated approach that marries advanced analytical techniques with deep consumer behavioral insights to navigate the complexities of the digital marketplace. As we chart the course for future inquiry, we invite fellow researchers and industry practitioners to join this quest for knowledge—a quest that seeks to decode consumer sentiment and leverage it as a beacon for product success in the ever-evolving tapestry of e-commerce.

Reference

[1] B. Lu, B. Ma, D. Cheng, and J. Yang, “An investigation on impact of online review keywords on consumers’ product consideration of clothing,” MDPI, https://www.mdpi.com/0718-1876/18/1/11 (accessed Mar. 24, 2024).

[2] Author links open overlay panelYi Han et al., “Analysis of sentiment expressions for user-centered design,” Expert Systems with Applications, https://www.sciencedirect.com/science/article/pii/S0957417421000452 (accessed Mar. 24, 2024).

[3] P. SONG and Y. LIU, “An XGBOOST algorithm for predicting purchasing behaviour on ...,” An XGBoost Algorithm for Predicting Purchasing Behaviour on E-Commerce Platforms, https://hrcak.srce.hr/file/355628 (accessed Mar. 24, 2024).

[4] D. A. Sanchez‑Loor and W. Chang, “Experimental study of the efects of structural assurance, personal experiences, and product reviews on repurchase behavior in e‑commerce platforms,” Sign in to your account, https://link-springer-com.ezaccess.libraries.psu.edu/article/10.1007/s10660-021-09525-5 (accessed Mar. 24, 2024).

[5] V.  Deshpande and P. K. Pendem, “Logistics Performance, Ratings, and Its Impact on Customer Purchasing Behavior and Sales in E-Commerce Platforms,” Sign in to your account, https://pubsonline-informs-org.ezaccess.libraries.psu.edu/doi/10.1287/msom.2021.1045 (accessed Mar. 24, 2024).

[7] Author links open overlay panelMustafa Kemal Yılmaz a, a, b, and AbstractPurposeBrand managers can determine the ideas, “Marketing insight from consumer reviews: Creating brand position through opinion mining approach,” Telematics and Informatics Reports, https://www.sciencedirect.com/science/article/pii/S2772503023000543 (accessed Mar. 24, 2024).