KARNAVATI UNIVERSITY

UNITEDWORLD SCHOOL OF COMPUTATIONAL INTELLIGENCE (USCI)

Summative Assessment (SA)

Submitted by
Ayush Jain
(Enrl. No.:
20210701017)

Course Code and Title: 21BSAI35E04-Natural Languague Processing

Project Title: Amazon food review analysis using vader

B.Sc. (Hons.) Computer Science / Data Science / AIML V Semester – July – Nov 2023

USCI

TABLE OF CONTENTS

1. Introduction

The modern digital era has witnessed an unprecedented surge in online retail, with e-commerce giants like Amazon playing a pivotal role in shaping consumer behavior. In the realm of e-commerce, customer reviews serve as a valuable source of information, influencing purchasing decisions and reflecting the overall sentiment towards products and services. This report delves into the intriguing domain of Amazon food reviews, employing advanced sentiment analysis techniques, specifically the VADER (Valence Aware Dictionary and sEntiment Reasoner) model and the Roberto model. Through this analysis, we aim to unravel the sentiments embedded within customer reviews, shedding light on the nuanced perceptions and experiences of consumers in the realm of food products on Amazon.

1.1 Project Description

The project centers around harnessing the power of sentiment analysis to comprehend the sentiments expressed in Amazon food reviews. With the colossal volume of reviews flooding online platforms, extracting meaningful insights manually becomes an arduous task. Therefore, the utilization of natural language processing (NLP) models becomes imperative. VADER, a pre-built sentiment analysis tool, and the Roberto model, a state-of-the-art transformer-based language model, are enlisted to decode the sentiment nuances present in the textual reviews. By leveraging these models, we aim to not only categorize reviews into positive, negative, or neutral sentiments but also to delve deeper into feature-level sentiment analysis. This involves identifying specific aspects within a review that contribute to the overall sentiment, providing a more granular understanding of consumer feedback.

1.2 Objectives

The primary objectives of this analysis are multifaceted. Firstly, we seek to provide a comprehensive overview of the sentiments prevailing in Amazon food reviews, offering insights into the overall satisfaction levels of consumers. Secondly, the utilization of the VADER model enables us to perform a sentiment analysis that is attuned to the inherent nuances of the English language, including slang, emoticons, and negations. Concurrently, the Roberto model, with its transformer architecture, allows for a more contextually aware understanding of the reviews. Thirdly, by comparing the outcomes of these two models, we aim to evaluate their respective strengths and weaknesses in deciphering sentiment in the context of food reviews.

In essence, this report endeavors to bridge the gap between unstructured textual data and actionable insights, empowering businesses, including Amazon and food industry stakeholders, to make informed decisions based on a nuanced understanding of customer sentiments. Through the lens of sentiment analysis, we embark on a journey to decode the language of consumer satisfaction and dissatisfaction, unraveling the intricacies of the ever-evolving landscape of e-commerce.

2. Literature Review

2.1 Amazon Food Review

The significance of customer reviews in the e-commerce ecosystem, particularly on platforms like Amazon, cannot be overstated. Amazon, as a global retail giant, hosts an extensive array of food products, ranging from gourmet delicacies to everyday essentials. The consumer-generated content in the form of reviews serves as a dynamic repository of insights into product quality, delivery experiences, and overall customer satisfaction. The diversity of products and the vastness of the Amazon marketplace make it an intriguing subject for analysis, as the reviews encapsulate the sentiments and preferences of a diverse consumer base.

The Amazon food review landscape is characterized by its sheer volume and diversity. Understanding the sentiments expressed in these reviews is crucial for businesses and sellers looking to enhance their offerings and customer experiences. Positive reviews can serve as powerful endorsements, attracting potential buyers, while negative reviews can highlight areas for improvement. This literature review sets the stage for our analysis by acknowledging the pivotal role Amazon food reviews play in shaping the online shopping experience.

2.2 Valence Aware Dictionary and Sentiment Reasoner (VADER)

VADER, a widely used sentiment analysis tool, is employed to discern the emotional tone of textual content. Developed specifically for social media text, VADER is adept at handling sentiment analysis in a manner that aligns with the nuances of informal communication. Its strength lies in its ability to recognize sentiment-laden words, account for negations, and comprehend the intensity of emotions. In the context of Amazon food reviews, the application of VADER allows for a nuanced understanding of consumer sentiments. By assigning polarity scores to individual words and considering their contextual usage, VADER provides a fine-grained analysis that goes beyond the binary classification of positive, negative, or neutral.

2.3 Roberto Model

The Roberto model represents a significant advancement in natural language processing, being based on transformer architecture. With the ability to capture contextual dependencies and long-range dependencies in language, the Roberto model surpasses traditional models in understanding the intricacies of textual data. In the domain of sentiment analysis, the Roberto model excels at discerning subtle nuances and capturing the semantic meaning of words and phrases within a given context. Its utilization in analyzing Amazon food reviews enables a more sophisticated examination of sentiments, delving into the subtle variations and connotations that may be missed by conventional models.

The Roberto model's transformer architecture facilitates a deeper comprehension of the relationships between words and phrases, allowing for a more accurate interpretation of sentiment within the context of food reviews. As we explore the interplay between VADER and the Roberto model in our analysis, we aim to unravel the strengths and limitations of each, ultimately contributing to the evolving landscape of sentiment analysis methodologies.

3. Methodology

3.1 Data Collection and Pre-processing

The cornerstone of our research lies in the comprehensive collection and meticulous pre-processing of the Amazon food reviews dataset. A diverse and representative sample of reviews is systematically gathered, encompassing various food categories and product types available on the Amazon platform. The aim is to ensure a broad spectrum of consumer experiences, capturing sentiments across different contexts and products.

Once the dataset is acquired, the pre-processing phase begins. This crucial step involves cleaning and organizing the data to prepare it for sentiment analysis. Irrelevant information, such as metadata and extraneous details, is removed. Missing or incomplete data is addressed to maintain the integrity of the dataset. Standardization procedures are applied to ensure consistency in the representation of textual content. Text normalization techniques, including the removal of stopwords and special characters, are employed to create a refined and uniform dataset, free from noise that might hinder accurate sentiment analysis.

3.2 VADER Model Implementation

The Valence Aware Dictionary and Sentiment Reasoner (VADER) model, chosen for its proficiency in handling informal language and nuanced expressions, is implemented to dissect the sentiment embedded in Amazon food reviews. VADER excels in its ability to assign polarity scores to individual words, considering factors like negation and intensity to compute an overall sentiment score for each review.

The VADER model is seamlessly integrated into the analysis pipeline, allowing it to process each review and categorize sentiments as positive, negative, or neutral. Its strength in deciphering the sentiment of user-generated content, often laden with slang and colloquial expressions, positions VADER as a valuable tool in unveiling the emotional tones hidden within the vast corpus of Amazon food reviews.

3.3 Roberto Model Implementation

Complementing the VADER model, the Roberto model is brought into play for a more nuanced and contextually aware sentiment analysis. Leveraging transformer architecture, the Roberto model is fine-tuned on the specific task of analyzing Amazon food reviews. The transformer architecture, with its attention mechanisms, enables the model to capture intricate relationships between words and phrases, providing a deeper understanding of context.

The implementation of the Roberto model introduces a layer of complexity, enhancing the analysis by capturing subtle variations in language and context that may be overlooked by traditional models. Its capacity to recognize semantic nuances positions the Roberto model as a potent tool for deciphering sentiment in the diverse and nuanced landscape of food reviews on Amazon.

3.4 Comparative Analysis

The heart of our methodology lies in the comparative analysis of the outputs generated by the VADER and Roberto models. This phase involves a meticulous examination of the strengths and

weaknesses of each model in the specific context of Amazon food reviews. Key metrics such as accuracy, precision, recall, and F1 score are employed to quantitatively assess the performance of the models.

The comparative analysis serves as a critical checkpoint, allowing us to discern how each model handles different facets of sentiment analysis. VADER's proficiency in recognizing sentiment-laden words and handling informal language is contrasted with the Roberto model's contextual awareness and semantic understanding. By evaluating and contrasting the outcomes, we gain a comprehensive understanding of the capabilities of each model and their suitability for the unique challenges posed by the diverse language in Amazon food reviews.

In summary, our methodology is a holistic approach that begins with robust data collection and pre-processing, followed by the implementation of two distinct yet complementary sentiment analysis models – VADER and Roberto. The comparative analysis provides a nuanced perspective, laying the foundation for the subsequent sections of the report where we delve into the sentiment analysis results, insights, and implications for both Amazon and the broader food industry.

4. Sentiment Analysis Results

4.1 Vader Model Results

In the realm of sentiment analysis, the Vader model unfolds a comprehensive narrative of the emotional undertones embedded within Amazon food reviews. Through the lens of Vader, we gain insights into the overall sentiment distribution across the dataset. The model categorizes reviews into positive, negative, or neutral sentiments, providing a bird's-eye view of customer satisfaction levels.

4.1.1 Feature-level Sentiment Analysis

Going beyond the binary classification, the Vader model facilitates a deeper understanding through feature-level sentiment analysis. This involves dissecting the reviews to identify specific features or aspects that contribute to the overall sentiment. For instance, the model can highlight whether positive sentiments are primarily associated with the taste of the food, prompt delivery, or packaging quality. This granular analysis empowers businesses with actionable insights, enabling them to pinpoint areas of strength and areas that may require improvement.

The feature-level sentiment analysis also reveals the intricacies of customer preferences and the factors that significantly influence their satisfaction. By understanding the nuanced components of positive and negative sentiments, stakeholders can tailor their strategies to align with consumer expectations, fostering a more responsive and customer-centric approach.

4.2 Roberto Model Results

Complementing the insights gleaned from the Vader model, the Roberto model introduces a layer of sophistication to the sentiment analysis of Amazon food reviews. The transformer architecture

empowers the Roberto model to capture contextual nuances and semantic intricacies, offering a richer understanding of the sentiments expressed by consumers.

4.2.1 Performance Metrics

Quantifying the efficacy of the Roberto model involves a thorough examination of performance metrics. Accuracy, precision, recall, and F1 score become instrumental in evaluating the model's ability to correctly classify sentiments. These metrics provide a quantitative assessment of the model's performance, shedding light on its strengths and potential areas for refinement.

The Roberto model's performance metrics are crucial in the comparative analysis, allowing us to juxtapose its outcomes with those of the Vader model. A balanced evaluation of precision, recall, and accuracy offers a comprehensive view of the model's proficiency in discerning sentiments within the intricate landscape of Amazon food reviews.

By scrutinizing the performance metrics, we aim to elucidate the nuanced contributions of the Roberto model to the sentiment analysis endeavor. This assessment goes beyond a simple comparison, delving into the unique strengths that transformer-based models bring to the table, especially in handling the complexity of language in diverse and subjective domains like food reviews.

In conclusion, the sentiment analysis results, both from the Vader and Roberto models, provide a holistic understanding of customer sentiments within the Amazon food review ecosystem. Feature-level analysis and performance metrics offer detailed insights into specific aspects of satisfaction and the effectiveness of each model. These findings serve as a foundation for the subsequent sections of the report, where we delve into actionable insights, implications, and potential strategies for enhancing the customer experience in the ever-evolving landscape of online food retail.

5. REFRENCES

1. https://docs.aws.amazon.com/AWSJavaScriptSDK/latest/AWS/ec2.html
2. https://medium.com/feedium/searching-on-medium-the-medium-search-function-3186cef4effd
3. https://aws.amazon.com/rekognition/
4. https://docs.aws.amazon.com/sns/

5. https://docs.aws.amazon.com/AWSJavaScriptSDK/latest/AWS/ec2.html
6. https://medium.com/feedium/searching-on-medium-the-medium-search-function-3186cef4effd
7. https://aws.amazon.com/rekognition/
8. https://docs.aws.amazon.com/sns/