**SENTIMENT ANALYSIS ON AMAZON PRODUCT REVIEWS**

**ABSTRACT**

The world nowadays is becoming more digitalized. In this digitalized world e-commerce is taking the ascendancy by making products available within the reach of customers where the customer doesn’t have to go out of their house. As nowadays people are relying on online products, the importance of a review is getting higher. For selecting a product, a customer needs to go through thousands of reviews to understand a product. It is becoming necessary for companies to examine customer reviews on online platforms such as Amazon to understand better how customers rate their products and services. But in this prospering day of machine learning, going through thousands of reviews would be much easier if a model is used to polarize those reviews and learn from it. First, the reviews were transformed into vector representation using different techniques like Bag-of-words, Tf-Idf, and maximum entropy. Using supervised learning methods like SVM, LSTM model and Random Forest on a large-scale amazon dataset to polarize it and get satisfactory accuracy.

**Keywords**—Sentiment analysis, feature extraction, text classification, machine learning.

**INTRODUCTION**

People trade goods using various e-commerce websites as the world's business is nearly exclusively conducted online. Because of this, it's also typical practice to read product reviews before making a purchase. Additionally, consumers these days are more likely to purchase a product based on reviews. Thus, a crucial area of study these days is evaluating the data from those customer reviews to make the data more dynamic. In the era of ever-increasing machine learning algorithms, it might take a lot of time to sift through hundreds of reviews to comprehend a product. However, we can categorize reviews to determine how popular a product is with consumers worldwide.

The aim is to classify customer comments, both good and negative, regarding various products and develop a supervised learning model to polarize many evaluations. Over 88% of online customers trust reviews as much as personal recommendations, according to research conducted on Amazon last year. Any online product with many favorable reviews offers a compelling argument for the product's validity. On the other hand, books or any other online product without reviews makes prospective customers wary. In short, more reviews appear more credible. People respect other people's opinions and consent because the only method to learn about other people's opinions about a thing is through material reviews.

Consumer opinions gathered from their experiences with certain goods or subjects directly affect what future customers decide to buy. We classify our datasets using both an active learning strategy and a manual approach. Several classifiers are used in the active learning process to offer accuracy up until it reaches a desirable level. Once we had a result that satisfied us, we processed those tagged datasets. We took characteristics out of the processed dataset, and several classifiers used those features to classify the data. To get better accuracy, we used two different methodologies for feature extraction: the bag of words approach and the tf-idf & Chi square approach.

**MOTIVATION**

Customer reviews and ratings play a crucial role in understanding consumer sentiment towards products. However, interpreting reviews can be challenging, especially when reviewers assign ambiguous ratings like three stars. This ambiguity can lead to confusion for both customers and companies trying to gauge satisfaction levels. Therefore, there's a need for robust sentiment analysis techniques to accurately classify reviews and understand consumer sentiment. The motivation behind this project is to develop a sentiment analysis system tailored for Amazon product reviews, addressing the challenges of ambiguous ratings, and providing valuable insights for consumers and companies alike.

**Main Contributions & Objectives**

* Develop a supervised learning model for sentiment analysis on Amazon product reviews.
* Utilize feature extraction approaches to classify reviews as positive or negative.
* Compare and evaluate different machine learning algorithms for sentiment analysis, including Naïve Bayes, Support Vector Machine (SVM), and Random Forest.
* Achieve high accuracy and performance metrics, such as precision, recall, and F1 score, in sentiment classification.
* Investigate the limitations and challenges of sentiment analysis on large-scale review datasets.
* Explore the potential of ensemble learning methods, such as Random Forest, for improving sentiment analysis accuracy.
* Provide insights into the advantages and disadvantages of different machine learning algorithms for sentiment analysis tasks.

**RELATED WORK**

**Sentiment analysis in amazon reviews using probabilistic machine learning:**

It's encouraged for customers of the e-commerce platform Amazon to leave reviews for the goods they buy. Amazon doesn't really try to control or limit what can be said in these reviews. While the quantity of evaluations differs among goods, all of them offer easily accessible and copious amounts of data that may be analyzed with relative ease for a variety of purposes. The goal of this effort is to apply and expand on recent advances in sentiment analysis and natural language processing to data that has been obtained from Amazon.

A review can be classified as favorable or negative using naive Bayes and decision list classifiers. Supervised machine learning uses the number of stars a user assigns to a product as training data.

The goal is to investigate attitudes regarding items, both positive and negative, which is a minor portion of this larger issue. Sentiment analysis looks for textual characteristics that indicate the context of the text (positive, negative, objective, subjective, etc.) and builds algorithms to capitalize on these characteristics. Although the issue of categorizing text as positive or negative is not the only one, it does provide a reasonable foundation for future developments. Relatively lately, a lot of work has been done on sentiment analysis of content that contains personal opinions. Pang and Lee (2002) classified a sizable corpus of movie reviews using several machine learning algorithms. Despite not being the most effective strategy, Naive Bayes performed well when compared to the human-generated classification terms baseline. Similar research was conducted in 2002 by Yessenov and Misailovi, who collected comments on movie reviews from social networking sites in a somewhat more anonymous setting.

Overall, this experiment yielded very positive findings. Significantly more user-generated data was accurately tagged by the classifiers than the 50% random baseline. The majority of the newly evaluated features were comparatively unsuccessful, although this was probably because of how they were implemented in relation to the bag-of-words. The idea underlying their application is sound, and in subsequent studies, they could be used in tandem for greater effectiveness.

**Amazon reviews, business analytics with sentiment analysis:**

This research aims to analyze the consequences of sentiment analysis on over 2.7 million reviews for the business domain. Researchers at UCSD provided the Amazon product data set that we used for our project (McCauley et al, 2015). We merge two original data sets—one made up of customer reviews and the other of product information—to obtain perceptive business acumen and the overall image of the entire information we obtained.

To further our understanding of these perspectives and achieve our goals of detecting user emotions from reviews, gender based on names and reviews, and further fake reviews, we not only adapt Textblob and Genderizer, but also develop our classifier to assess the accuracy of the system. After that, we began analyzing specific viewpoints or accessories associated with well-known brands like HTC, Apple, and Nokia to further explore and analyze our intriguing results using various techniques.

Regarding the classification task, we utilize the Scikit-learn Python package to construct a Multinomial Nave Bayes (MNB) and a Support Vector Machine (SVM) classifier (Joachims, 1998; Wu et al., 2004). To determine the accuracy, we used 50% of the data to train both classifiers and the remaining 50% of the data to test them.

clients who are aggressive can be inferred from the contrasts in the frequency of phrases used by these businesses. For example, clients of Nokia may compare their products with those of Apple. Nevertheless, the word cloud examples we have also highlighted the drawbacks of these popular assessments. Good, outstanding, and excellent are examples of adjectives that don't really capture the kind of specifics about accessories that these brands can enhance. If we were to put ourselves in these firms' shoes, we would need to investigate more negative comments in order to improve their goods.

**Feature selection methods in sentiment analysis and sentiment classification:**

Opinion mining has access to a wealth of valuable information that has been generated by the rapid expansion of social networking sites, blogs, forums, and other online communities.   
These days, a lot of e-commerce websites have developed, and people prefer to purchase online because of the great deals, wide selection of products, etc. Prior to making a purchase, buyers can evaluate the product against other products and read reviews left by previous customer.  
Polarity classification is sentiment analysis's primary goal. Additionally, the reviews aid the manufacturers by revealing to them the shortcomings in their product, which might improve customer preference. Sentiment classification is a subfield of text classification that focuses on the viewpoint that a given topic conveys. Opinion mining, sentiment analysis, sentiment extraction, and affective rating are some other terms for sentiment analysis. Sentiment analysis is used to determine the review or comment's semantic orientation. Sentiment analysis consider both factual and subjective data.

The suggested approach is assessed for feature selection and sentiment classification using the publicly available Amazon dataset. Several processes are carried out, including data collecting using the publicly accessible Amazon dataset and preprocessing that filters the reviews by eliminating special characters and stop words that are not needed for further processing. Phrase level, single word, and multiword feature selection is done. Next comes vector creation, which creates a vector against the positive and negative class labels. Lastly, sentiment classification is carried out using the Naïve Bayes algorithm. It performs performance analysis for categorization.

Based on our experimental work, we have concluded that the Naïve Bayes algorithm produces superior results than the other two for phrase level feature extraction. The phrase level accuracy can be attributed to the suffix array method's implementation. The Naïve Bayes classifier performs well even when training on a smaller dataset and testing on a larger dataset. The primary benefit of employing Naïve Bayes is its simplicity in implementation. In this work, we have employed phrase level, single word, and multiword approaches to extract features. The Naïve Bayes algorithm is used to classify reviews, and the outcomes are examined.

**PROPOSED WORK**

We will go over the suggested methodology for sentiment analysis of product reviews on Amazon.com. The framework comprises multiple phases, such as preparing data, extracting features, training the model, and assessing the results. We will go over each step-in detail, emphasizing the methods, formulas, and strategies that were applied. We will also discuss the reasoning behind the selected course of action and how it resolves the issues mentioned in the problem description. We will also offer insights into the framework's implementation, including the software tools and programming languages that were employed.

To polarize a sizable amount of unlabeled product review data, we suggested a supervised learning approach. We presented our model, a supervised learning technique that combines two different feature extractor approaches. We reviewed the fundamental theory underlying the model, the methods we employed in our study, and the performance metric for the conducted experiment over a sizable amount of data. Additionally, we contrasted our outcome with a few other comparable product review studies.

Additionally, we read through a variety of research articles on sentiment analysis using text-based datasets. With the F1 measure, we were able to attain over 90% accuracy, over 90% precision, and over 90% recall. To get encouraging results, we experimented with several simulations utilizing cross validation, a different feature extraction procedure, and a training-testing ratio for comparing different amounts of data. Most of the time, tenfold increased accuracy, while support vector machines (SVM) produced the best classification outcomes. Because e-commerce sites have restrictions on what data they may share publicly, it is challenging to compile enormous amounts of gold standard datasets for this purpose. Furthermore, data scraping may provide challenges because insufficient data is available to be regarded as actual public evaluations of various products.

Classification by Random Forest Random forests are an ensemble learning technique for tasks like classification, regression, and other tasks. They work by building many decision trees during the training phase and producing a class that represents the mean prediction (regression) or the mode of the classes (classification) of the individual trees. Decision trees' tendency to overfit to their training set is corrected by random decision forests.   
One of the most widely used machine learning algorithms is the decision tree.   
Regression and classification issues are addressed by decision trees.

Decision Trees are a kind of Supervised Machine Learning in which the training data is continually divided based on a given parameter. In other words, you describe what the input and corresponding output in the data are. Decision nodes and leaves are the two entities that can be used to explain the tree. The choices or results at the end are the leaves. And the data is separated at the decision nodes.

Because LSTMs are made to solve the vanishing gradient issue that conventional recurrent neural networks (RNNs) face, they are superior at identifying long-term dependencies in sequential data, such as text. This ability is essential for sentiment analysis since it helps determine the sentiment of a sentence by taking preceding word context into account.

In natural language, sentences can differ greatly in length. LSTMs' capacity to preserve and update memory over time makes them suitable for handling variable-length sequences. Because of this, they work well in sentiment analysis jobs where the inputs can vary in duration.

Sentiment analysis frequently necessitates comprehending the sentiment conveyed in a series of words as opposed to a single word. By processing words one at a time while keeping information from prior words, LSTMs may successfully model sequential information and so capture the entire attitude portrayed in a sentence.

**DATA DESCRIPTION**

We'll give a thorough explanation of the dataset that was utilized to analyze sentiment in Amazon product reviews. We shall describe the features of the dataset, such as the quantity of reviews, types of products, and distributions of ratings. We will also go over the preprocessing procedures that were used on the dataset, including normalization, tokenization, and text cleaning. We will also describe any difficulties or problems that arose during the preprocessing of the data and how they were resolved. All things considered, this section will provide readers a thorough grasp of the dataset that was utilized for analysis and testing.

Tokenization is the act of breaking down a series of strings into individual tokens, which can be words, phrases, keywords, symbols, or other items. Singular words, phrases, or even entire sentences can be used as tokens. Certain characters are removed during the tokenization process, such as punctuation marks. Tokens are used as input in various processes, such as text mining and parsing.

Bag of Words: Used in information retrieval and natural language processing, bag of words is a technique for extracting features from simplified text or data. A text or document is represented in this approach as a bag (multiple set) of its words. Making a list of helpful terms is, thus, the essence of sentiment analysis's "bag of words." We have extracted our feature sets using a bag of words technique. Following preprocessing the dataset, we separated the various parts of speech using pos tagging. From there, we chose nouns and adjectives to employ in the creation of a bag of words. After that, we apply supervised learning to it to determine our findings and the most frequently used terms from the review dataset.

**RESULTS**

By polarizing a significant amount of unlabeled product review datasets using a supervised learning algorithm. We presented our model, a supervised learning technique that combines two different feature extractor approaches. We reviewed the fundamental theory underlying the model, the methods we employed in our study, and the performance metric for the conducted experiment over a sizable amount of data. We also contrasted our findings with those of a few other comparable product review studies. Additionally, we read through a variety of research articles on sentiment analysis using text-based datasets. With the F1 measure, we were able to attain over 90% accuracy, over 90% precision, and over 90% recall. To get encouraging results, we experimented with several simulations utilizing cross validation, a different feature extraction procedure, and a training-testing ratio for comparing different amounts of data. Most of the time, tenfold increased accuracy, while support vector machines (SVM) produced the best classification outcomes. Because e-commerce sites have restrictions on what data they may share publicly, it is challenging to compile enormous amounts of gold standard datasets for this purpose. Additionally, there may be issues with data scraping because there is insufficient data to treat as actual public reviews of various products. LSTM and Random Forest offer distinct advantages for sentiment analysis of Amazon product reviews. LSTM is a good fit for complicated datasets because it is good at capturing subtle linguistic subtleties and long-term dependencies. On the other hand, Random Forest works well with smaller datasets or when interpretability is important because it is resilient, interpretable, and less prone to overfitting.

**REFERENCES**

* Samha, Xu, Xia, Wong & Li “Opinion Annotation in Online Chinese Product Reviews.” In Proceedings of LREC conference, 2008.
* Nina Isabel Holleschovsky, “The social influence factor: Impact of online product review characteristics on consumer purchasing decisions”, 5 th IBA Bachelor Thesis Conference, Enschede, The Netherlands 2015
* Elli, Maria Soledad, and Yi-Fan Wang. "Amazon Reviews, business analytics with sentiment analysis." 2016
* Xu, Yun, Xinhui Wu, and Qinxia Wang. "Sentiment Analysis of Yelp‘s Ratings Based on Text Reviews." (2015).
* Rain, Callen. "Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning."Swarthmore College (2013).
* Bhatt, Aashutosh, et al. "Amazon Review Classification and Sentiment Analysis." International Journal of Computer Science and Information Technologies 6.6 (2015): 5107-5110.
* Chen, Weikang, Chihhung Lin, and Yi-Shu Tai."Text-Based Rating Predictions on Amazon Health & Personal Care Product Review." (2015)
* Shaikh, Tahura, and DeepaDeshpande. "Feature Selection Methods in Sentiment Analysis and Sentiment Classification of Amazon Product Reviews.” (2016)
* Nasr, Mona Mohamed, Essam Mohamed Shaaban, and Ahmed Mostafa Hafez. "Building Sentiment analysis Model using Graphlab." IJSER, 201