The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 10 Issue 1, 2022
pp: 122 - 126

Amazon Kindle Store Review Analysis Using Artificial Intelligence

[1]D.Likhitha, [2]J.Sahithi Priya, [3]G.Chaitanya Lakshmi, [4]B.Kalyani
[1,2,3,4] UG Student, Department of Computer Science Engineering,
Dr K V Subba Reddy College Of Engineering For Women, Kurnool, Andhra Pradesh, India

Abstract
The world that we see today is becoming increasingly digital. By making products accessible to customers without the need for them to leave their homes, e-commerce is gaining ground in this digitalized world. As more and more people rely on online products, reviews are becoming increasingly important. A customer must read thousands of reviews to understand a product before making a purchase. However, utilizing a model to polarize those reviews and learn from them would make going through thousands of reviews much simpler in this prosperous age of deep learning. On a massive Amazon dataset, we polarized it using deep learning techniques and achieved satisfactory accuracy. Reviews on Amazon aren't just about the product—they're also about the customer service. In this paper, we propose a system that categorizes customer reviews and then determines the sentiment of those reviews. This will make it easier for customers to make decisions if they can clearly distinguish between product reviews and service reviews. Product feature sentiment is also extracted using a rule. Additionally, we provide a visualization for our summary of the results.

1. Introduction

People trade goods through a variety of e-commerce websites due to the fact that the majority of commercial websites around the world are almost entirely online platforms. As a result, reviewing products prior to purchase is also common. Additionally, customers today are more likely to purchase a product based on reviews. Therefore, analyzing the customer reviews' data to make it more dynamic is a crucial field in today's world. In this day and age of increasingly sophisticated algorithms based on machine learning and deep learning, it takes quite a bit of time to read thousands of reviews to comprehend a product. However, we can polarize reviews based on a specific category to comprehend the product's popularity among buyers worldwide.

This paper aims to build a supervised learning model to polarize a large number of reviews and classify the positive and negative customer feedback about various products. Over 80% of online shoppers trust reviews as much as personal recommendations, according to a study conducted on Amazon last year. A strong indication of the item's legitimacy is provided by a large number of favorable online reviews. On the other hand, if there are no reviews on books or any other online product, potential customers will be skeptical. Simply put, more reviews appear to be more convincing. The only way to understand how others feel about a product is

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 10 Issue 1, 2022
pp: 122 - 126

through a review. People value the consent and experience of others. Users' opinions, gathered from their experiences with particular products or topics, have a direct impact on future customer purchase decisions. Similarly, negative reviews frequently result in reduced sales. The objective for those is to comprehend customer feedback and polarize appropriately based on a large amount of data. Similar projects have been carried out using the Amazon dataset. Deep learning is a branch of machine learning that is entirely based on artificial neural networks. Since neural networks are going to mimic the human brain, deep learning is also a kind of mimic of the human brain. In order to comprehend the polarized attitudes toward the product, opinion mining was used to examine a limited set of Amazon kindle product reviews. We don't have to explicitly program everything in deep learning. Deep learning is not a novel concept. Since a few years ago, it has been around. It's getting a lot of attention right now because we didn't have as much processing power or data in the past. Deep learning and machine learning emerged as a result of the exponential increase in processing power over the past two decades. Neurons are the formal definition of deep learning. A single neuron in the human brain is made up of approximately 100 billion neurons, and each neuron is connected to thousands of its neighbors.

This begs the question of how these neurons can be recreated in a computer. As a result, we construct an artificial structure with nodes or neurons that we refer to as an artificial neural net. In the hidden layer, there may be many neurons interconnected between the neurons responsible for the input value and the output value

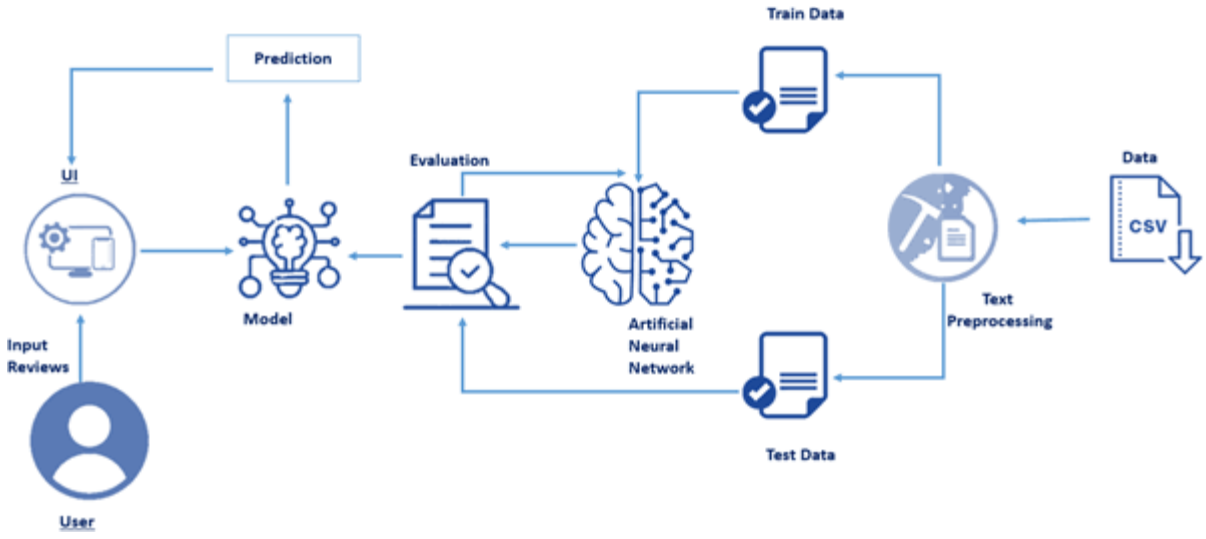

Fig.1 Machine Learning architecture

2. Literature Review

Tokenization preceding processing: It is the process of breaking up a string of words, phrases, keywords, symbols, and other elements into individual tokens. Words, phrases, or even entire sentences can be tokens. Some characters, like punctuation marks, are removed during

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 10 Issue 1, 2022
pp: 122 - 126

tokenization. The tokens are used as the input for various processes like text mining and parsing.

Stemming: The process of producing morphological variants of a root or base word is known as stemming. Stemming algorithms and programs are also commonly referred to as stemmers. The words "chocolate," "chocolatey," and "choco" are reduced by a stemming algorithm to the root word "chocolate," and the words "retrieval," "retrieved," and "retrieves" are reduced to the stem word "retrieve." In natural language processing, stemming is an important part of the pipelining process. Tokenized words are used as input for the stemmer.

Eliminating End Words: Stop words are items in a sentence that aren't necessary for text mining in any way. Therefore, in order to improve the analysis's accuracy, we typically disregard these words. Depending on the nation, language, etc., different stop words are used in various formats. There are a few in the English format.

LEMMATIZER: The process of grouping the various inflected forms of a word together so that they can be analyzed as a single item is called lemmatization. Similar to stemming, lemmatization provides context for the words. As a result, it ties together a single word with similar meanings

Fig.2 Sentiment Analysis

3. Proposed System

would be the second sample input, and "e" would be the output until we loop through the entire dataset. In order to make accurate predictions, we need to provide the model with as many examples as possible. The Data Augmentation (DA) Technique allows us to generate various versions of real datasets without actually collecting the data, thereby artificially increasing the size of training data. In order to improve performance on the classification task, the data need to be changed in order to keep the class categories.

Computer Vision and Natural Language Processing (NLP) employ a data augmentation strategy to deal with data scarcity and lack of diversity. Augmented images can be made fairly easily, but Natural Language Processing is more difficult due to the language's inherent complexities. Even if we were to substitute a synonym for every word, the context would still

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 10 Issue 1, 2022
pp: 122 - 126

be different. The model's performance is enhanced when the size of the training data is increased through data augmentation. The model will perform better the more data we have. The augmented data's distribution should not be too similar to the original or too different from it. Overfitting or poor performance may result as a result. Effective DA strategies should aim for a balance

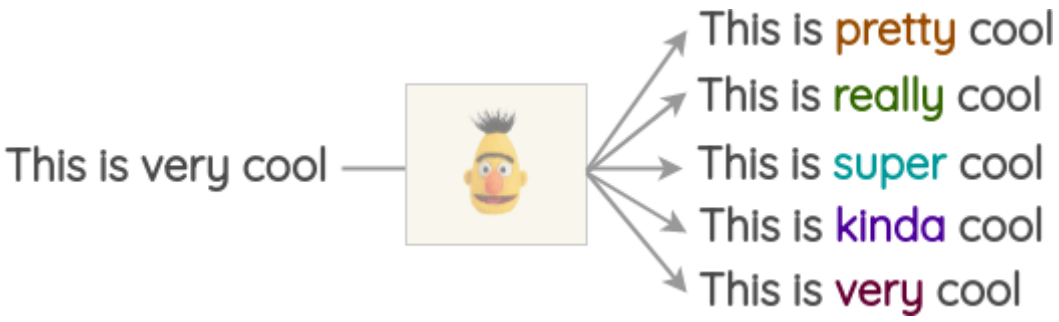

Fig.3 Variations of code

We are pretending that we are developing a system that can predict the sentiment of a textual book review as either positive or negative.This means that after the model is developed, we will need to make predictions on new textual reviews. This will require all of the same data preparation to be performed on those new reviews as is performed on the training data for the model. We will ensure that this constraint is built into the evaluation of our models by splitting the training and test datasets prior to any data preparation. This means that any knowledge in the data in the test set that could help us better prepare the data (e.g. the words used) are unavailable in the preparation of data used for training the model. That being said, we will use 80% train, 20% as a test of the data

Fig.4 Output Of The Project

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 10 Issue 1, 2022
pp: 122 - 126

5.Conclusion

It is impossible to make predictions using only raw text as input. As a result, we observed that the pre-processing step was crucial to the NLP process as a whole. Pre-processing the data requires careful consideration of the type of content that is contained within it in order to achieve better results, accuracy, and the ability of the machine to take all of the text as tokens. The ability to extract relevant features from the given data source is the most crucial aspect. In order to extract additional learning features and boost the predictive power of the models, this kind of data frequently serves as a useful complement. Additionally, the user can predict whether a comment will be positive or negative.

.

References

1. R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. Natural language processing (almost) from scratch. INTERNATIONAL JOURNAL OF CURRENT ENGINEERING AND SCIENTIFIC RESEARCH (IJCESR) ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-8, ISSUE-6, 2021 77 Journal of Machine Learning Research, 12(Aug):2493–2537, 2011.
2. M. S. Elli and Y.-F. Wang. Amazon reviews, business analytics with sentiment analysis.
3. C. Rain. Sentiment analysis in amazon reviews using probabilistic machine learning. Swarthmore College, 2013.
4. R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing, pages 1631–1642, 2013.
5. Y. Xu, X. Wu, and Q. Wang. Sentiment analysis of yelps ratings based on text reviews, 2015.
6. Bhatt, Aashutosh, et al. "Amazon Review Classification and Sentiment Analysis." International Journal of Computer Science and Information Technologies 6.6 (2015): 5107-5110. [7]Chen, Weikang, Chihhung Lin, and YiShu Tai."Text-Based Rating Predictions on Amazon Health & Personal Care Product Review." (2015)
7. Shaikh, Tahura, and DeepaDeshpande. "Feature Selection Methods in Sentiment Analysis and Sentiment Classification of Amazon Product Reviews.",(2016)
8. Nasr, Mona Mohamed, Essam Mohamed Shaaban, and Ahmed Mostafa Hafez. "Building Sentiment analysis Model using Graphlab." IJSER.,