International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1156-1160 www.ijaem.net ISSN: 2395-5252

Customer Satisfaction improvement using product reviews through Natural Language Processing and IDIC framework

Kolipyaka Rajesh[1], Nalla Greeshma[2]
[1]*Design Engineer, Cyient Limited, Hyderabad, Telangana*
[2]*Student,Institute of Aeronautical Engineering College,Hyderabad, Telangana.*

Submitted: 10-03-2021 Revised: 30-03-2021 Accepted: 01-04-2021

ABSTRACT:The current world is inclining more towards the digital world due to the data revolution. In recent years, well-established e-commerce platforms have evolved to a stage where reviews on their products and services by customer's inputs help in calibrating public mood. The reviews written by users in an e-commerce platform give different opinions on a product or a service. Online reviews are psychologically influencing the product growth, new product development, understanding user tastes and varied marketing strategies. This paper aims to undertake a stepwise methodology of analysing sentiment of real-time product review data from an e-commerce shopping site. The reviews are extracted through web-scraping techniques and a set of NLTK supervised learning principles for classification of customer reviews. The vast impact created by product reviews in businesses using IDIC framework and Machine Learning techniques has been studied in this paper.

KEYWORDS: Sentiment Analysis, Natural Language Processing, IDIC Framework, E-Commerce, Customer reviews, Visualization, Machine Learning, Product Strategy.

I. INTRODUCTION

Before the 21st century, offline review sources like word of mouth, surveys and public opinion are used to make any product or services decision making. More than 80% of the internet population spends around 7 hours a day online, looking for e-shopping to buy goods and services [1]. A recent study by Google LLC reveals nearly 67% of consumers are being influenced by online reviews [2]. According to Pew Research Centre, in general, an organization conducts surveys and market research to find the pulse of the public [3]. But in the 21st century where data is more valuable than money, everyday enormous data is being generated by the user through web shopping reviews in different e-commerce platforms. The main sources of reviews are from forum discussions, blogs, social media platforms like Twitter, Facebook and reviews posted at merchant websites.Customer reviews are crucial to retailers and product manufacturers to understand the product performance and service experience; this in turn drives in making key customer-oriented business decisions, marketing strategies and friendlier customer support. As technology gets sophisticated, the ways to interpret and process the user's opinion has seen great change, right from enquiring from family and friends to automatically identifying sentiment of a review i.e., positive, negative, or neutral to decide the purchase impact on a product. Sentiment analysis is a type of text analysis which can use text mining techniques to understand the bias of the review context. In this paper, web scraping techniques are used to extract the review data from an e-commerce website. Second, apply a sentiment analysis on review data to generate classification of customer reviews (positive/negative). Finally, the review data analysed for effective business decision making and improved customer support using IDIC framework. Due to the exponential increase in customer review data in e-commerce platforms and social media sites, it is important for developing an unsupervised sentiment analysis model to determine the sentiment of the reviews [4]. Hence, there is a necessity for a novel approach to summarize the customer reviews into a simple understandable format. More Customer-Centric businesses are using the latest technologies and business research to extract deeper insights from e-commerce platforms and social media comments to pitch for new product ideas, improvements and solutions to achieve competitive advantages in a business.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1156-1160 www.ijaem.net ISSN: 2395-5252

II. LITERATURE REVIEW

1) Review: A review is a user opinion based on his or her purchase or service experience. It's the most important part of the business development. Firms adapt numerous ways to collect genuine reviews for a product or service to develop their products in the market.

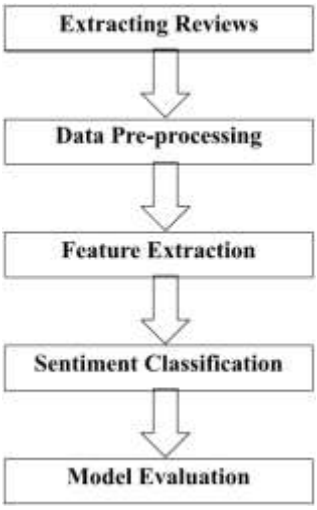
2) Customer Relationship Management: Customer relationship management (CRM) refers to the principles, practices, and guidelines that an organization follows when interacting with its customers [5]. The user reviews will help firms to understand the customer intent and improve quality decision making and customer service.

3) Sentiment: Sentiment analysis is extremely important because it helps businesses quickly understand the overall opinions of their customers on a product or service. By sorting the sentiment (positive/negative) of the reviews automatically, it helps in faster and more accurate decision making for business development and customer experience.

III. ANALYSIS APPROACH

It's estimated that 90% of the world's data is unstructured. Huge volumes of unstructured data are generated every day in the form of emails, support tickets, chats, social media conversations, surveys, articles, documents, videos, photos, etc. Therefore, it is tedious to analyse sentiment for every review.

In current paper, five steps are proposed for sentiment classification is discussed.



Proposed Methodology

1. Extracting Reviews: Data from an e-commerce site is collected by using web scraping python libraries like requests, beautiful soup is used to extract the reviews of customers from an e-commerce site to a CSV (Comma Separated Value) file [6]. Thus, the customer data is collected and used for sentiment analysis.

	Unnamed: 0	Review_data
0	0	Awesome
1	1	Just wow!
2	2	Wonderful
3	3	Best book to be successful in every sphere of ...
4	4	Mind-blowing purchase
...	...	...

Review Data CSV file

Text visualization using word clouds:

Word cloud: Word cloud is also called 'Tag Cloud'. It's a cloud of words for visualization of text data. The text data is visualized depending on the importance and the frequency of the word. Matplotlib library in python is used in the generation of a word cloud. Below is the word cloud of our customer review data

Word Cloud from Review data

2. Data Pre-processing: Data Preprocessing is the most important step in performing sentiment analysis. Most of the text extracted from the internet is in unstructured form i.e., in a heterogeneous format. So, data cleaning and processing is required which plays a major role for analyzing the sentiment analysis of customer reviews. The data preprocessing steps for building the NLP model include:

I) Tokenization: Tokenization is separating the data into smaller units called 'Tokens'. Breaking of sentences into individual words is called "Word tokenization" and the breaking of words into individual characters is called "Character tokenization". Before converting the words to tokens, the reviews of the data set are all converted into lower case format. For Example, Consider the sentence: "come, let's do research on sentiment analysis." Now, the tokenizationof this sentence is: ['come', ',', 'let', "'s", 'do', 'research', 'on', 'sentiment', 'analysis', '.']

II) Removing special characters: Special characters like #, $, %, @, *, &, ~, (,), ", >, < add no value in understanding the meaning of the sentence.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1156-1160 www.ijaem.net ISSN: 2395-5252

Thus, removing such noisy special characters from the reviews dataset is necessary.

III) Removing stop words: Stop words do not give any positive or negative impact on the sentence. Some examples of stop words are 'while', 'of', 'at', 'by', 'for', 'with', 'I', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves'.

IV) Stemming/lemmatization: Sometimes the same words may appear in different tenses like “plays”, “played” and “playing”; these all-tense forms express the same meaning of the word “play”. Hence, stemming cuts the starting and ending of a word considering the prefix and suffix of the words. In lemmatization, grammar and root of the word is considered.

Stemming operation

3. Feature Extraction

Bag of Words: Bag of words (BOW) is a method used in natural language processing which creates a vocabulary of all the unique words occurring in all the documents in the training set. It extracts features from the text documents that are used to train the machine learning model [10]. It represents the word count of a sentence and tells us whether the word is present in the document or not and also does not consider the order of the words.

Vectorizing the Text: In this step, the word tokens are represented in a machine-readable format as vectors. The Count Vectorizer constructs a vector with the counts of the words of each review. This process of transforming the text tokens into vectors is called Vectorization. By importing the Count Vectorizer from sklearn.feature_extraction.text library sparse feature vectors are created from the word tokens.

4. Sentiment Classification: In this implementation, a logistic regression classifier is used on the dataset for evaluation of sentiments.

Logistic Regression: Logistic regression is a simple machine learning model that predicts the probability of an outcome that can only have two values (such as pass/fail, win/lose) [8]. Logistic regression produces a logistic curve (S-shaped curve) whose values lie between 0 and 1, but never exactly at those limits.

Logistic Regression equation:

$$p = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p)}}$$

Where p = dependent variable to be predicted
b0= y-intercept
b1, b2, b3, bp = coefficient of independent variables
x1, x2, x3, xp = independent variables.

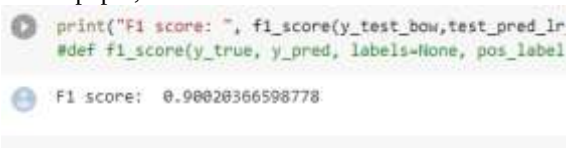
After the bag of words model, the train and test data are split, the train data is used for training (build up) the model and test data are used for testing the model and validating it by predicting against the test set. Here, 80% of the data set is used for training the machine learning model and the remaining 20% is used as test data. The train and test data are fit into the logistic regression model for predicting the output of the sentiment classification.

5. Model Evaluation: There are many metrics to evaluate the performance of a model; the reliable metrics are F-Score and the Accuracy method. By implementing an accuracy method false negative and false positives are not considered. Therefore, the F-score method is adopted in this sentiment analysis. F-score is a measure of model accuracy on a dataset. It’s more reliable in evaluation of classification models like positive and negative [6].

The formula to evaluate F-score is the harmonic mean of the precision and recall. A perfect model has an F-score of 1.

F1= 2/((1/recall) *(1/precision))
F1= tp/(tp+0.5(Fp+Fn))

In this paper, the evaluated F-score is 0.90203.



IV. BUSINESS CONTEXT- IDIC FRAMEWORK

By understanding the importance of negative sentiments helps to develop more appealing branding techniques and marketing strategies [11]. The sentiments of the reviews are taken as a base for further understanding the business context and help to understand the pulse of the customers.The framework has been adapted from one of the CRM principles like IDIC. Here, attempts are made to resolve customer issues by IDIC framework.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 3 Mar. 2021, pp: 1156-1160 www.ijaem.net ISSN: 2395-5252

IDIC framework for customer satisfaction improvement using online reviews.

1.Identify: The negative reviews based on the sentiment analysis model are identified and analyzed for customer issues and their needs. The potential issues have to be identified.
2.Differentiate: Differentiate the reviews based on service, product quality, customer experience and sorted according to priority like high, medium and low. This differentiation helps to prepare strategies that meet customer expectations for strong customer relations [12].
3.Interact: Customer interactions help in understanding customer expectations on products and services. The interaction, based on customer issues provides a chance to collect feedback, opinions and improvements. This helps to develop customer-oriented products and services and improved customer relationship [12].
4.Customize: After identifying, differentiating and interacting based on customer reviews, ideas can be drawn for new strategies and improved product and service customizations.

V. CONCLUSION

This work presents a novel collaborative sentiment analysis of online customer reviews by using Natural Language Processing techniques on e-commerce sites. To show how the system works in practice, a suitable case study is explained, i.e., the product review data is collected using web scraping techniques from e-commerce site and appropriate NLP techniques are applied to find sentiment of the reviews. The sentiments of these reviews are further processed for IDIC Framework implementation in order to improve customer satisfaction and gain product quality and marketing competency. The review data can also be extracted from many social media sites, travel websites, job boards and various other online or offline platforms. This model can be further improved to better analyze sentiment of the reviews to track customer sentiment and emotions over time. To determine which customer segment feels stronger about your brand and also to apply IDIC framework tos further understand the customer spending habits like what they buy, how much they spend.

REFERENCES

[1]. Anderson, M; Tsao, T-C; and Levin, M., 1998, "Adaptive Lift Control for a Camless Electrohydraulic Valvetrain," SAE Paper No. 98102
[2]. https://www.webhostingsecretrevealed.net/blog/ecommerce/online-shopping-ecommerce-and-internet-statistics-2020-you-should-know/
[3]. https://www.millenniumlocalmarketing.com/a-new-study-shows-that-67-of-consumers-are-influenced-by-online-reviews/
[4]. https://www.pewresearch.org/internet/2016/12/19/online-reviews/
[5]. Bagheri, A, Saraee, MH and deJong,FRathore, Ilavarasan&Dwivedi, 2016"Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews"
[6]. Customer relationship management:https://en.wikipedia.org/wiki/Customer_relationship_management#:~:text=Customer%20relationship%20management%20(CRM)%20is,study%20large%20amounts%20of%20information.&text=CRM%20may%20be%20used%20with%20past%2C%20present%20or%20potential%20customers.
[7]. N Greeshma, C Raghavendra, K Rajendra Prasad; (IJITEE) ISSN: 2278-3075, Volume-9 Issue-1, November 2019."Visualization of Optimal Product Pricing using E-Commerce Data"
[8]. G.Vinodhini,RM.Chandrasekaran (IJARCSSE) Volume 2, Issue 6, June 2012."Sentiment Analysis and Opinion Mining: A Survey"
[9]. AtulKamat, SnehalChavan,NeilBamb,HiralAthwani, Prof. Shital A. Hande."Natural Language

Processing on Hospitals: Sentimental Analysis and Feature Extraction."

[10]. SoumyaGeorge ,Shibily Joseph. Feature (IOSR-JCE) Volume 16, Issue 1, Jan 2014. "Text Classification by Augmenting Bag of Words (BOW) Representation with Co-occurrence."

[11]. https://dzone.com/articles/applications-of-sentiment-analysis-in-business-1

[12]. Thakur, S. and Chetty, P. (2019). The IDIC model of Customer Relationship Management (CRM). [online] Project Guru. Availableat:https://www.projectguru.in/customer-relationship-management/ [Accessed 20 Mar.2021].