A DATA MINING PROCESS FOR PRODUCTS QUALITY REVIEW SYSTEM

First Author[1], Second Author[2]

[1] Aishwarya, Master of Computer Application BKIT-Bhalki
[2]Prof . Poojarani. N,, Master of Computer Application BKIT-Bhalki

Abstract - E-commerce platforms and retailers place a premium on customer feedback and reviews. Time and again, reviews have proved to be an important part of the final selection process.process. They are invaluable in assisting with the development of their item. and support. Feedback like this comes straight from customers and the More reviews are available for products with a larger user base. Simply said, there isn't enough time or resources to make it happen. time-consuming since one must go through each review and run an analysis. The use of NLP for this purpose is quite beneficial. Using Machine Learning techniques that provide thorough examination of Customer complaints and graphical representations of such complaints visualizations that help people draw conclusions about their company that can be put into action.

*Key Words***:** Data Mining, NLP, E-Commerce

INTRODUCTION

Phishing is a method used by hackers and attackers Nowadays, individuals are more interested in trading things on e-commerce websites instead of the offline market because of time efficiency and traditional property. In electronic commerce, product reviews are used on shopping sites to provide consumers a chance to rate and comment on things they have bought, directly on the product page. A common circumstance is to read through a review of the goods before purchasing. So, the reviews of the goods have swayed the consumer favorably or adversely. On the other hand, reading thousands of reviews is a superhuman achievement. In this era of expanding machine learning-based algorithms, it is rather time-consuming to sift through thousands of comments to discern a product where a review of a given category may be polarized to know its popularity among consumers worldwide.

Sentiment analysis is the component of text mining that aims to define the views, emotions, and attitudes contained in a text or set of texts. Online reviews are so important to businesses because they ultimately increase sales by giving the consumers the information, they need to make the decision to purchase the product. One very important factor in elevating the reputation, standard, and evaluation of an e-commerce store is product review. Product review is the most valuable resource accessible for Customer Feedback. This research intends to identify customers' positive and negative comments on various items and construct a supervised learning model to polarize a broad variety of evaluations. Last year, research on amazon reveals more than 88 percent [1] of internet customers trust reviews as much as a private referral. However, each internet goods with a great amount of favourable evaluations gives the item's credibility with a power comment. On the other hand, without evaluations, electronic gadgets or any other thing online puts prospective prospects in a condition of suspicion.

Sentiment analysis helps to examine these opinioned data and extract some crucial insights which will aid to other user to make choice. Because of time efficiency and traditional property, individuals are more interested in trading things on e-commerce websites rather than the offline market. Product reviews are used on shopping sites in electronic commerce to allow consumers to rate and comment on things they have bought directly on the product page. Before purchasing a product, it is normal practise to read a review on it. As a result, the product reviews have either influenced the buyer favourably or adversely. Reading hundreds of reviews, on the other hand, is an unnatural feat. In current day and age of expanding machine learning-based algorithms, it is time-consuming to sift through hundreds of comments to find a product where a review of a given category may be polarised to determine its popularity among customers globally.

Text mining sentiment analysis is the process of attempting to characterise the views, feelings, and attitudes expressed in a text or group of texts. Online reviews are critical for companies because they eventually improve sales by providing customers with the information they need to make a purchasing choice. Product reviews are a vital aspect in improving an e-commerce store's reputation, quality, and assessment. The most useful resource for Customer Feedback is product reviews.

The goal of this research is to categorise customers' positive and negative comments on different items and to develop a supervised learning model to polarise a broad variety of evaluations. According to Amazon study last year, more than 88 percent [1] of online customers trust reviews as much as a personal recommendation. Any online goods with a large number of positive evaluations, on the other hand, provides the item's validity with a powerful comment. On the other hand, the absence of reviews, technological devices, or any other item online causes prospective customers to be sceptical.

Sentiment analysis aids in the examination of opinionated data and the extraction of crucial insights that may be used to other applications.Data mining is concerned with data and the mining of patterns in data. Knowledge Discovery in Data-bases (KDD) is a non-trivial extraction of unique, implicit, and actionable knowledge from massive data sets that increasingly employs computer databases for successfully storing and retrieving information. [1] essentially facilitates data exploration, data analysis, and data visualisation in massive datasets. Data mining methods are the end result of research and product development, and they include artificial neural networks, decision trees, genetic algorithms, and many more. [2] is used in a variety of sectors, including manufacturing. (A paper on the data mining technique will be supplied) to make a choice.

Problem statement:

Every second, a massive amount of data is created, and it must be managed and processed. The processing of data and information provides access to knowledge, which is vital to any operation. This information may come from any step of the process, such as procurement, production, design, promotion, upkeep, or delivery. Knowledge Discovery is a significant output of datamining, which identifies hidden patterns in data. Manufacturing is the process of creating products for sale to consumers. Warehousing is a procedure that businesses employ to manage and direct their storage facilities, from the moment an item is received until it is sent. The efficiency of production relies heavily on good warehouse management.

SYSTEM ANALYSIS:

Existing System:

An Overview of Social Media Sentiment Analysis Methods - The authors of this review article, Shaikh Sayema Anwer and V.S. Karwande, employ a Naive Bayes classifier to analyse the sentiments of the data they collect and compress the sentences using natural language processing (NLP) without losing any relevant information.
An Overview of Methods for Analysing Twitter User OpinionsIn their research and assessment of current methods for opinion mining, Vishal A. Kharde and S.S. Sonawane found that machine learning and lexicon-based approaches yielded the highest accurate models, followed by SVM and Naive Bayes Classifier.

Proposed System

Data mining is part of an iterative process that includes:
1. Data selection is the process of determining whether data is accurate. The relevant data for the analysis is chosen and extracted from the data collection.
It also indicates the instrument that will be used to gather correct and authentic data.
2. Data cleaning, also known as data scrubbing, is the removal or replacement of noisy, undesired, incomplete, and irrelevant data from a collection.

3. Data transformation: also known as data consolidation. Following data preparation, the chosen data is processed, summarised, and normalised into a fitting representation suited for the mining technique [3].

4. Data integration: the combining of disparate data sources into a single common source [4].

5. Data mining: An important stage in which data is analysed from many viewpoints and usable information is extracted. These intelligent approaches are used to extract potentially relevant patterns.

6. Pattern evaluation: a strategy for identifying intriguing patterns that reflect knowledge and are discovered based on specified score measures with numerical significance.

7. information representation: the last process involves expressing the information gathered via mining.

It is attempting to make sense of the patterns discovered. This gathered information is graphically communicated to the customer. It is a necessary stage that employs a wide range of visualisation approaches to assist customers in quickly understanding and interpreting mining findings.

ARCHITECTURE
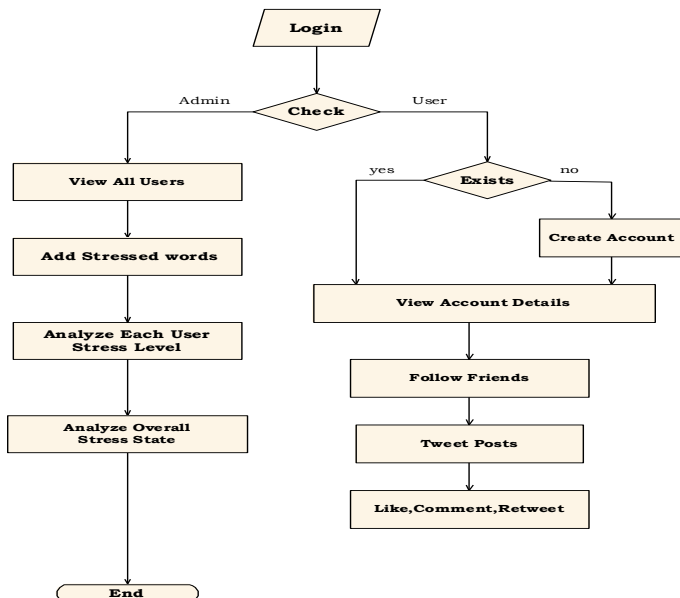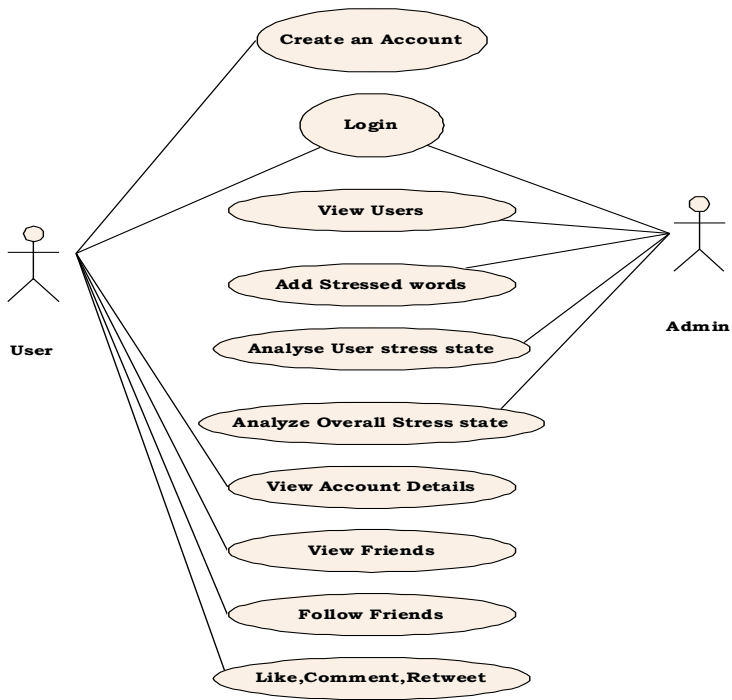
Results and Analysis:

Working:

Tokenization, stop word removal, and POS tagging are the three processes in pre-processing data.
1) Tokenization: This is the process of breaking a string sequence into distinct tokens such as keywords, symbols, phrases, and other items. Tokens may eventually be phrases, words, or even full sentences.
However, during the tokenization process, certain characters, such as punctuation marks, are eliminated. Tokens are used as input for a variety of functions, including parsing and text mining.
2) Removing stop words: Stop words are phrases that are not required for text mining in any division. These terms are frequently omitted in order to increase the accuracy of the evaluation. Stop words come in a variety of forms depending on the realm, language, and so on. In English, however, there are multiple stop words.
3) POS tagging: Part of speech tagging refers to the process of assigning one of the parts of speech to a particular word. It is often known as POS tagging. Nouns, verbs, adverbs, adjectives, pronouns, conjunctions, and their subcategories are all examples of parts of speech. A programme that performs this is known as a parts of speech tagger or POS tagger.

Conclusion:
The suggested system analyses customer evaluations and comments using Natural Language Processing and other machine learning classification methods. Many online retailers now use this strategy, since it has lately increased in popularity. The aforementioned system evaluates all comments and ideas submitted to see whether they are helpful or harmful. Logistic Regression was the most accurate of the three algorithms.

REFERENCES

[1] Ting Liu, Zhong Su, Honglei Guo, Yanyan Zhao, Che Wanxiang, and Che Wanxiang. Using Compressed Sentences in Aspect-Based Emotion Analysis. IEEE/ACM Transactions on Audio, Speech, and Language Processing 23, no. 12 (2015): 2111-2124

[2] Robert Schwartz, James Lin, David Zajic, Bonnie J. Dorr, and Richard Schwartz. Using Sentence Compression to Perform Document Summarization Tasks, "Multi-Candidate Reduction." 2BBN Technologies, 9861 Broken Land Parkway, Columbia, MD 21046; University of Maryland, College Park, Maryland, USA.
As cited by Trevor Cohn and Mirella Lapata

[3]. Word Reduction Is Not the Only Way to Condense a Sentence. Pages 137-144, 22nd Annual Meeting of the Association for Computational Linguistics (Coling 2008) Proceedings August of 2008 in Manchester.

[4] . Niloufar Salehi Dastjerdi, Seyed Hamid Ghorashi, Roliana Ibrahim, and Shirin Noekhah. IJCSI International Journal of Computer Science Issues, Volume 9, Number 4, Number 1, July 2012, ISSN (Online): 1694-0814, "A Frequent Pattern Mining Algorithm for Feature Extraction of Customer Reviews."
F.T. Martins,

[5] Dipanjan Das Andre. "A Review of Machine-Generated Summaries" Presentation at CMU's Language Technologies Institute on the 21st of November, 2007.
Credit goes to

[6] Kapil Thadani and Kathleen McKeown. Using Joint Structural Inference to Compact Sentences.New York, New York 10025, USA Computer Science Department Columbia University

[7] The authors are LuWang, Hema Raghavan, and Vittorio Castelli.A Sentence Compression-Based Framework for Query-Driven Document Summarization.T. J. Watson Research Centre, Yorktown Heights, New York 10598, USA; Cornell University, Ithaca, New York 14853, USA.