Sentiment Analysis for E-Commerce Product Reviews

Dr. Agbakwuru Alphonsus Onyekachi
Department of Computer Science,
Imo State University, Owerri.
alphosus1010@yahoo.com

Duru Esther Chidalu
Michael Okpara University of Agriculture Umudike,
Umuahia Abia State
ikisehestherjoan@gmail.com

DOI: 10.56201/ijemt.v9.no1.2023.pg18.32

ABSTRACT

Due to the rise of e-commerce, online reviews have become an important factor for customers when making purchasing decisions. Consequently, it is crucial for online retailers to comprehend customer views towards their products and services. To accomplish this, natural language processing software will be utilized to analyze the text of online reviews and classify them based on their sentiment. The resulting sentiment values will then be used to generate an overall score, which can be utilized to evaluate customer sentiment toward a specific product or service. After dividing the dataset into an 80:20 training and testing ratio, positive attitudes accounted for 50.56% of the total data, while negative sentiments accounted for 49.44%, with the machine learning algorithms being trained on the training dataset. To categorize the customer reviews dataset into positive and negative sentiments, a supervised machine learning algorithm called Random Forest classifier was utilized along with other evaluation metrics such as precision, recall, F1-score, and balanced accuracy. The model evaluation showed that the classifier's recall, precision, and f1-score were 98 %, 98%, and 98% respectively. Python programming language was used for this work, along with its libraries such as pandas, NumPy, Seaborn, and Matplotlib for data analysis and visualization, and Natural Language Tool Kit packages. This research will result in a sentiment analysis model that can accurately classify customer reviews based on sentiment and generate an overall score that represents customer sentiment toward a specific product or service.

KEYWORDS: *Sentiment analysis (SA), E-commerce, Machine learning, Decision Tree, Natural Language, Evaluation metrics.*

1. Introduction

Sentiment analysis is mainly concerned with the study of reviews or opinions. These reviews are in text formats. Each word of these reviews can be considered as a feature for analysis. Since sentiment analysis is concerned with the study of reviews, opinions on any topic and providing meaningful information, selecting a proper authentic set of reviews for processing is a challenging job. The reviews or comments provided by the people are mainly in the text format which is sometimes tough to understand and process. Hence, different mechanisms like stop word, numerical and special character removal, which do not play any active role in sentiment analysis of the texts and along with this all text are converted into either lower or upper case, to maintain uniformity during the analysis of the reviews. It is observed that sometimes the collection of all words becomes vast and it may contain words which may not affect the sentiment of the reviews. Thus, a feature selection mechanism needs to be adopted to select the best features out of all the features, which affect the sentiments of text and a proper preprocessing mechanism needs to be adopted to remove unwanted, confusing information for the data sets.

People's opinions can provide very useful information that can help in decision-making. As internet facilities have advanced, reviews ranking has grown to be a very difficult problem. Many companies and organizations allow customers to openly express and share their thoughts, ideas, and opinions in textual content. In order to identify thoughts and sentiments for an ever-growing volume of content and data online and offline, automatic sentiment classification of texts tries to address the issue. Organizations and stakeholders seeking information about products and services end up with a massive collection of people's opinions that is largely unorganized. It can be difficult to glean information from the opinion sources. Numerous methods have been tried, and conventional categorization techniques have produced results with a good level of accuracy.

Sentiment analysis is mainly concerned with the study of reviews or opinions. These reviews are in text formats. Each word of these reviews can be considered as a feature for analysis. Since sentiment analysis is concerned with the study of reviews, opinions on any topic and providing meaningful information, selecting a proper authentic set of reviews for processing is a challenging job. The reviews or comments provided by the people are mainly in the text format which is sometimes tough to understand and process. Hence, different mechanisms like stop word, numerical and special character removal, which do not play any active role in sentiment analysis of the texts and along with this all text are converted into either lower or upper case, to maintain uniformity during the analysis of the reviews. It is observed that sometimes the collection of all words becomes vast and it may contain words which may not affect the sentiment of the reviews. Thus, a feature selection mechanism needs to be adopted to select the best features out of all the features, which affect the sentiments of text and a proper preprocessing mechanism needs to be adopted to remove unwanted, confusing information for the data sets.

People's opinions can provide very useful information that can help in decision-making. As internet facilities have advanced, reviews ranking has grown to be a very difficult problem. Many companies and organizations allow customers to openly express and share their thoughts, ideas, and opinions in textual content. In order to identify thoughts and sentiments for an ever-growing volume of content and data online and offline, automatic sentiment classification of texts tries to address the issue. Organizations and stakeholders seeking information about products and services end up with a massive collection of people's opinions that is largely unorganized. It can be difficult to glean information from the opinion sources. Numerous methods have been tried, and conventional categorization techniques have produced results with a good level of accuracy. Modern deep learning models have, nevertheless, surpassed more established machine learning techniques. Despite that, based on the research done, there appear to be few studies examining sentiment analysis using pairwise ranking approaches. This thesis addresses the gap of available research done on sentiment analysis.

2. Literature review

Sentiment analysis is a subfield of Natural Language Processing (NLP) that analyzes people's opinions, sentiments, appraisals, attitudes and emotions through opinions expressed in written texts. Liu (2012), stated that sentiment analysis is often known as opinion mining and explained it as a natural language processing (NLP) method for identifying the polarity of sentiment in textual data.

Sentiment analysis uses computer methods to examine people's feelings and perspectives on specific issues. It examines how people feel and behave toward many things, including products, digital material, organizations, events, and other things. To produce useful information, sentiment analysis examines consumers' opinions and reviews of any company, product, and/or attribute Abinash (2017). Due of the vast amount of opinionated material on the web, sentiment analysis has become a particularly active topic of research. It has grown to be one of the most popular NLP research fields in recent years from both academia and industry; it has a wide range of applications but also poses many difficult research difficulties.

Terminology used in sentiment analysis

1. Data Mining: this refers to the process of identifying irregularities, trends, and connections within extensive data sets to anticipate outcomes.

2. Emotion: this is a multifaceted experience that encompasses awareness, physical sensations, and actions, reflecting the individual importance of an object, an event, or a situation.

3. Natural Language Processing (NLP): this is the ability of a computer program to comprehend human language in both written and spoken forms.

4. Opinion: is a perspective or conclusion formed about something that may not be founded on facts or information.

5. Sentiment Analysis: this is a technique in natural language processing, gauges individuals' tendencies to form opinions and is utilized to categorize data as positive, negative or impartial.

6. Textual: it pertains to the content of written or spoken language.

Sentiment analysis refers to the process of automatically extracting emotions from written text by analyzing unstructured information and creating a model to derive insights from it. It is also known as opinion mining and aims to identify subjective language and extract valuable information from textual data. The text is typically categorized as either positive or negative, with some cases being neutral. By utilizing various techniques and mechanisms, sentiment

analysis can provide valuable insights from large amounts of textual data, which can support decision making at operational, managerial, and strategic levels Bird et al,. (2009).
Although sentiment analysis and opinion mining appear to be the same thing; the term sentiment analysis was first used by Fernández-Gavilanes, Milagros, et al. (2016) and the term opinion mining was first employed by Hemalatha, (2014), there is a distinction: the former relates to discovering sentiment words and phrases that express feelings, whilst the latter refers to extracting and analyzing people's opinions for a certain entity.
For this study, both strategies are considered interchangeable. One's attitude toward a target thing is represented by the sentiment/opinion polarity, which can be positive, negative, or neutral.

According to Liu (2012), sentiment analysis is a tool that examines people's opinions, evaluations, attitudes, and emotions towards both tangible and intangible issues such as products, services, or topics. This area of research has gained significant popularity in natural language processing as it analyzes people's sentiments and attitudes towards various entities such as products, digital content, organizations, events, and issues. Hemalatha (2014) described sentiment analysis as a useful tool for identifying customer groups, assessing brand image, monitoring stock market trends, detecting patterns, and handling crisis situations.

Sentiment analysis is arguably the fastest growing areas of research in NLP. According to Bhargava and Rao(2018) sentiment analysis is an ongoing research area that helps determine customers' opinion or the situation of market on a particular entity. Sentiment analysis makes use of computational techniques to study peoples' emotions and opinions on given topics shared in form of text data. According to kaplan and Haenlein(2010) sentiment analysis is a tool in data mining that can overcome challenges like harnessing, analyzing and interpreting textual content since data is dispersed, disorganized, and fragmented by systematically extracting and analyzing online data without facing any time delays.

Sentiment analysis has a major impression on texts which hold some form of emotion, or dispositions. The concept of sentiment analysis is to study a group of text data to understand the opinion or sentiment expressed by the text. This is usually achieved by identifying the sentiment within the text(s) and representing them with a value that is positive or negative, known as polarity. From the sign of the polarity, the overall sentiment is often deduced as positive, neutral or negative.

Machine learning (ML) is a field of study that provides computers that opportunity to learn without being explicitly programmed Samuel (96). Machine learning is a computation algorithm which is built to emulate human intelligence by learning through experience Greener et al (2022).

Supervised machine learning technique is the most commonly used machine learning technique and it require a labelled dataset to train the learning algorithm. In this type of machine learning approach, both the training and testing data are labeled. Each text file of the dataset has a polarity value assigned to them either positive or negative or neutral. The training dataset is used by the system for training, and based on this information, the testing data is labeled O’Connor et al (2010).

3. Methodology

3.1 The 2 main approaches for doing sentiment analysis are:

- machine learning-based
- lexicon-based approach

3.1.1 Machine Learning Approach: Machine learning (ML) is a field of study that provides computers that opportunity to learn without being explicitly programmed. Machine learning is a computation algorithm which is built to emulate human intelligence by learning through experience.

Machine learning methods for sentiment classification are based on models that are calibrated to categorized data. This data is called training data. The data is trained accordingly, which can be applied to machine learning algorithms. The calibrated model or machine can then be used to categorize new data, much like a parameterized equation can be used to predict the value of the response variable in regression analysis. The training is based on features that have an effect on the data polarity, and they are chosen using feature selection methods.

Supervised Machine Learning

Supervised machine learning technique is the most commonly used machine learning technique and it require a labelled dataset to train the learning algorithm. In this type of machine learning approach, both the training and testing data are labeled. Each text file of the dataset has a polarity value assigned to them either positive or negative or neutral. The training dataset is used by the system for training, and based on this information, the testing data is labeled . As the testing dataset already has a label, both the labels are compared to obtain the final accuracy of the system.

Unsupervised Machine Learning

Unsupervised machine learning methods are not popular in sentiment classification problems due to not being easily applicable in non-review dataset. The learning algorithm uncovers pattern from unlabeled textual data. Examples are the clustering models such as k-means clustering ,Density-Based Spatial Clustering of Applications with Noise and fuzzy clustering. The pattern or structure aids in classifying the sentiment of a topic or subject. This approach is considered better than supervised models because it does not require training data, linguistics knowledge and saves time.

Semi-Supervised machine learning

Machine learning as the name suggests requires the model to be trained, lexicon based consists of a set of rules and does not require prior training, a third method although not used very often is a hybrid method that combines both lexicon and machine learning approaches, generally it yields better results than either method Table 3.1 will summarize the differences between the two approaches.

3.1.2 Lexicon-Based Approach

Lexical based approach depends on constructing a Lexicon, which is a "structure that keeps track of words and possibly information about them" where the words are referred to as "lexical items" (Charniak, 1996). Once the lexicon is constructed, the overall polarity of the text is then found by a possibly weighted count of those lexical items. The lexicon-based approach predicts the sentiments by using the lexical databases like SentiWordNet and WordNet. It obtains a score for each word in the sentence or document and annotates using the feature from the lexicon database that are present. It derives text polarity based on a set of words, each of which is annotated with the weight and extracts information that contributes to conclude overall sentiments to the text. Also, it is necessary to pre-process data before assigning the weight to the words. Lexicon dictionary or database contains the opinionated words that are classified with positive and negative word type, and the description of the word that occurs in current context. For each word in the document, it is assigned with numeric score, and average score is computed by summing up all the numeric scores and sentiment polarity is assigned to the document.
Lexicons can be created automatically or manually. Automatically built lexicons perform well across all domains (Taboada et al. 2011) because it uses general knowledge resources which have wider term coverage (Muhammad et al. 2013). While manually built lexicon approach may be preferred because it simulates the effect of contextual word or phrase. However, the latter approach is mostly domain specific, expensive and time consuming to build. There are several lexicons developed before the use of social media such as WordNet. WordNet is an online English lexical database developed by Miller (1995). These available lexicons tend to perform poorly when used to classify social media data due to the ambiguous and colloquial text present in social media data. Considering that previous studies have developed lexicons like Sentistrength to recognize the presence of ambiguity and colloquial words in social media data and were shown to have performed well in different domains. There are other popular lexicons namely, SentiWordNet, AFINN, SO-CAL, LIWC and NRC.

Differences between Lexicon-based and Machine Learning.

Criteria	Lexicon-based	Machine Learning
Domain	Independent	Dependent
Classification approach	Unsupervised	Supervised
Regular prior training	No	Yes
Adaptive learning	No	Yes
Time of result generation	Fast	Slow
Maintenance	Need maintenance of corpus	Do not require maintenance
Accuracy	Low	High then lexicon

Table 3.1: Comparison between lexicon and machine learning.

3.2 Sentiment Analysis Process

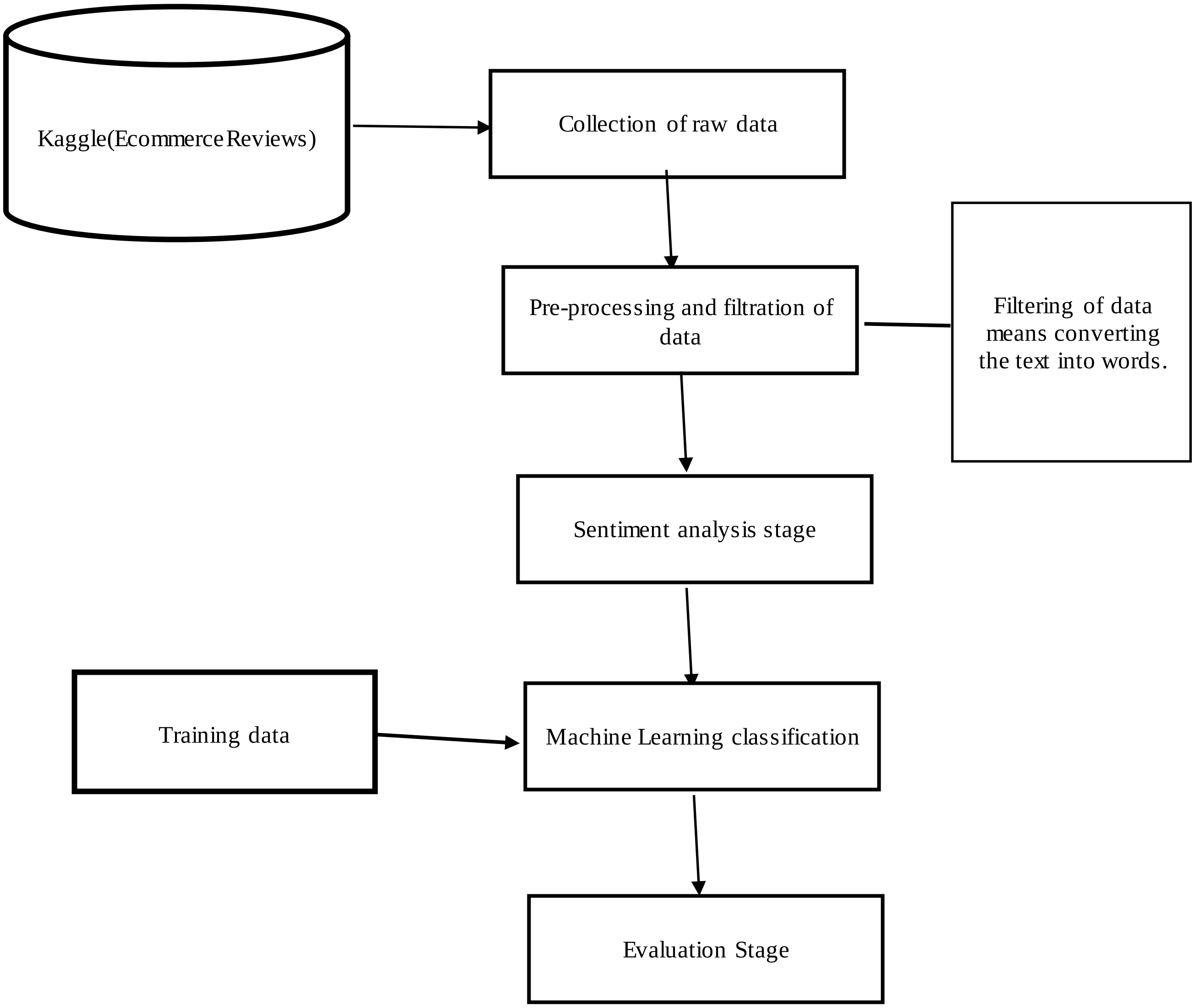

Fig 3.1 Ecommerce sentiment analysis workflow

The sentiment analysis processes are as follows:

- Collection of Raw data and filtration

Sentiment analysis harnesses the huge amount of content generated over the internet. Some data sources of the content are public forums, product review sites, blogs, discussion boards and social networking sites like Linkedin, Twitter and Facebook. The data is produced rapidly and is often bulky and unstructured. Because people will express their emotions differently with regards to choice of language and style of writing, some may prefer to use slang and emoticons to name a few. This means that manually analyzing this text is a tedious and an almost impossible task.

However, with sentiment analysis, we can make use of natural language processing techniques to filter the useful information for analysis and classify it accordingly.

- Data Pre-processing

The text preparation process begins by cleaning the textual data prior to the analysis stage. Normally text preparation involves identifying and removing non textual content from the data such as hash tags and hyperlinks. Also, some non-essential information that does not contribute to the reviewer's opinion such as name and location and date all these are removed, also any other content that is deemed irrelevant to the analysis is also removed

- Feature Extraction

Before the model can classify text, the text needs to be prepared so it can be read by a computer. Tokenization, lemmatization and stop word removal can be part of this process, similarly to rule-based approaches. In addition, text is transformed into numbers using a process called vectorization. These numeric representations are known as "features". A common way to do this is to use the bag of words or bag-of-ngrams methods. These vectorize text according to the number of times words appear.

In modern times, deep learning has revolutionized text vectorization techniques. The word2vec algorithm is one such example that employs a neural network model to learn word associations from vast amounts of text. This neural network model represents each unique word as a list of numbers or a vector. The significant advantage of this technique is that it assigns similar numeric representations to words with similar meanings, thereby aiding in sentiment analysis and improving accuracy.

- Training

The supervised Learning is a unique process of resolving classification problems as this enables for easiest way for future predictions of unknown data.

- Classification

Classification algorithms are used to predict the sentiment of a particular text. They are trained using pre-labelled training data. Classification models commonly use Naive Bayes, Logistic Regression, Support Vector Machines, Linear Regression, Decision Tree.

- Evaluation of sentiment classification

There are 4(four) common ways of evaluating the performance of a sentiment classification.

Accuracy = (TP+TN) / (TP+TN+FP+FN)

Precision = TP / (TP+FP)

Recall = TP / (TP+FN)

F1 = (2×Precision×Recall) / (Precision + Recall)

True Positives (TP): True positives refer to situations where the predicted class matches the actual class and both are true.
True Negatives (TN): true negatives occur when both the predicted and actual classes are false.
False Positives (FP): False positives are when the predicted class is true, but the actual class is false.
False Negatives (FN): False negatives are when the predicted class is false, but the actual class is true.

4. Results and Discussion.

The Dataset, Analysis and the outcome of the proposed system are presented in this section in context of performance evaluation.

- Dataset

In order to carry out the sentiment analysis for this research, reviews on some E-commerce products were used. This techniques has been tested on some other ecommerce dataset gotten from kaggle and other sites. This reviews were made by different people from different geographical areas around the world.
There is a total of 1676 reviews for the following products: Accucheck, Becadexamin, Evion, Neurobion, SevenseascodLiverOil, Shellac, Supradyn, shampoo. This reviews are further classified into informative and non-informative .Below are is a general statistics for the Informative (1) and Non informative (0) graphical representation. The dataset tends to go into data preprocessing and Exploratory Data analysis. The dataset were properly clean into to produce accurate results. There was a deduction of 21 reviews after cleaning which makes our new data a total of 1655.

	product	answer_option
3	Accucheck	fwegwrqdsdvwfg
82	Accucheck	qwerwetrjy
362	Neurobion	gehryetw
432	Shelcal	gehryetw
1449	Accucheck	ghyukuyujredfehrbv
1503	Accucheck	VVV. Good
1599	Accucheck	wqwasdbggn
1668	Accucheck	Ft GM

Fig. 4.1 Gibberish reviews

Out[41]:

	product	answer_option	label
0	Accucheck	Fast and accurate delivery	0
1	Accucheck	As usual it is genuine	0
2	Accucheck	Behavior of delivery boy is very bad. Delivery...	0
3	Accucheck	fwegwrqdsdvwfg	0
4	Accucheck	These strips were as per my requirment	0
...	...	...	...
1671	Accucheck	Ft GM	0
1672	Accucheck	I like	0
1673	Accucheck	Nice price with long expiry	0
1674	Accucheck	Price & Service	0
1675	Accucheck	Good discount	0

1676 rows × 3 columns

Fig. 4.2 product reviews

Out[42]:

label	0	1	All
product			
Accucheck	317	85	402
Becadexamin	53	27	80
Evion	89	33	122
Neurobion	286	137	423
SevenseascodLiverOil	60	22	82
Shelcal	262	126	388
Supradyn	50	23	73
shampoo	57	49	106
All	1174	502	1676

Fig4.3 Informative and non-informative overview of products reviews.

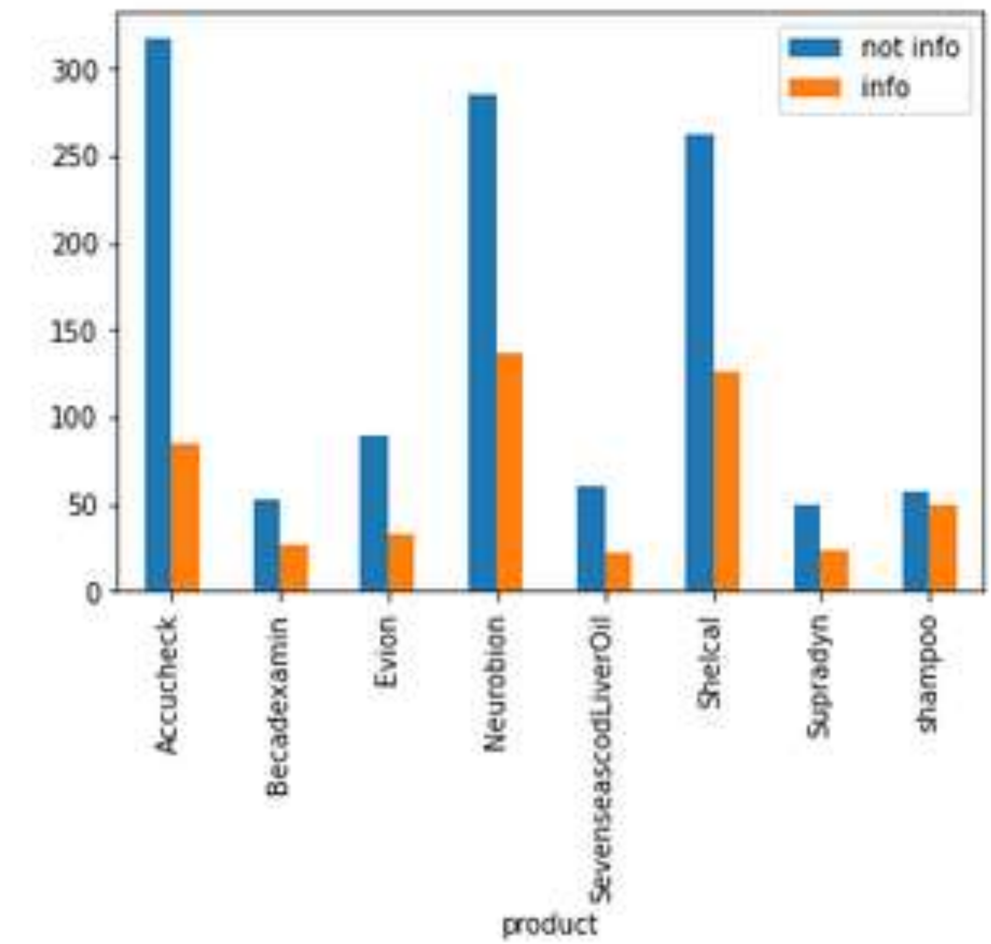

Fig.4.4. Geographical representation for informative and non informative reviews

Fig.4.5. Commonly used words.

count	1655.000000
mean	0.301511
std	0.459053
min	0.000000
25%	0.000000
50%	0.000000
75%	1.000000
max	1.000000

Fig.4.6 Statistical Overview for the dataset.

Performance Evaluation

```
Training Accuracy
 0.9963420883671341
Test Accuracy
 0.978724406851542
CLASSIFICATION REPORT
Training
               precision    recall  f1-score

           0       0.99      1.00      1.00
           1       1.00      0.99      1.00

    accuracy                           1.00
   macro avg       1.00      1.00      1.00
weighted avg       1.00      1.00      1.00

Test
               precision    recall  f1-score

           0       0.98      0.98      0.98
           1       0.98      0.98      0.98

    accuracy                           0.98
   macro avg       0.98      0.98      0.98
weighted avg       0.98      0.98      0.98
```

5 Challenges in sentiment analysis

1. **Data Availability:** The existing system of ecommerce sentiment analysis is heavily dependent on the availability of data. Without access to a sufficient amount of data, the effectiveness of the system is limited.
2. **Accuracy:** Although the existing system of ecommerce sentiment analysis can be fairly accurate, it is still easily prone to errors. The system is not able to accurately distinguish between different sentiments, leading to inaccurate results.
3. **Subjectivity:** The existing system of ecommerce sentiment analysis is based on subjective criteria and is highly dependent on the context of the data. This can lead to incorrect results if the context is not properly understood.
4. **Scalability**: The existing system of ecommerce sentiment analysis is not easily scalable. As more data is added to the system, it becomes increasingly difficult to accurately process and analyze the data.
5. **Cost**: The existing system of ecommerce sentiment analysis is generally expensive to implement and maintain. This is primarily due to the need for specialized hardware and software, as well as the cost of acquiring and processing data.

Application Areas

1. Automated product recommendation: Product recommendation systems can incorporate sentiment analysis to personalize their recommendations to customers. By analyzing customer reviews, sentiment analysis can identify products that customers with similar sentiment profiles may be interested in.
2. Customer segmentation: Sentiment analysis can help e-commerce websites segment customers into groups depending on their sentiment towards certain products or services. This can help e-commerce companies better understand their customers and tailor their service to meet the needs of each segment.
3. Customer service: Sentiment analysis can be used to identify customer complaints and proactively address them. This can help e-commerce companies improve customer service and ensure customer satisfaction.
4. Personalized marketing: By understanding customer sentiment, e-commerce companies can tailor their marketing messages to customers. This can help increase customer engagement and drive sales.
5. Product optimization: By analyzing customer reviews, e-commerce companies can identify areas of improvement for their products. This can help them make sure their products are up to customer standards.

Conclusion

The aim of this paper is to present a comprehensive analysis and comparison of different techniques used in opinion mining. The study covers machine learning and lexicon-based approaches, as well as cross-domain and cross-lingual methods, and various evaluation metrics. The findings reveal that machine learning methods, specifically logistic regression and decision tree classifier, have the highest accuracy and are considered the baseline learning methods. On the other hand, lexicon-based methods are useful in situations where little effort is required in human-labeled documents. The study also examines the impact of different features on classifiers, and it is concluded that cleaner data leads to more accurate results. The use of bigram model is found to provide better sentiment accuracy compared to other models. The paper suggests exploring the combination of machine learning and opinion lexicon methods to enhance the accuracy of sentiment classification.

Reference

Bhargava, M. G., & Rao, D. R. (2018). Sentimental analysis on social media data using R programming. International Journal of Engineering & Technology, 7(2.31), 80-84.

Bird, Steven, Ewan Klein, and Edward Loper. *Natural language processing with Python* 2009.

Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. Nature Reviews Molecular Cell Biology, 23(1), 40-55.

Kaplan and M. Haenlein. Users of the world, unite! the challenges and opportunities of social

Liu, B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language* media. Business Horizons, 53:59–68, 02 2010. doi: 10.1016/j.bushor.2009.09.003.

O'Connor, R. Balasubramanyan, R. Routledge, Bryan., and A. Smith, Noah. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. 2010.

Samuel, A. L. (1967). Some studies in machine learning using the game of checkers. II—Recent progress. IBM Journal of research and development, 11(6), 601-617. *Technologies*, 5(1):1–167, 2012.

Fernández-Gavilanes, Milagros, et al. "Unsupervised method for sentiment analysis in online texts." Expert Systems with Applications 58 (2016): 57-75.

Hemalatha, I., Dr GP Saradhi Varma, and A. Govardhan. "Case Study on Online Reviews Sentiment Analysis Using Machine Learning Algorithms." *International Journal of Innovative Research in Computer and Communication Engineering* 2.2 (2014):3182-3188.

"**Kaggle datasets," [Online]. Available: https://www.kaggle.com/karangadiya/fifa19.**