International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Sentiment Analysis on Consumers Behaviour Buying Computer Hardware

Author: Chandra Mohan Ray

University of Mumbai Institute of Distance & Open learning (IDOL),
Information Technology, University of Mumbai

Abstract:
In this paper we present a deep literature review on existing system for sentimental analysis. Basically, sentimental analysis (SA) is the measurement of preference of people's thoughts via natural language processing. This study leads to the behaviour pattern of consumer's thinking process and various aspects factors of buying. The main aim of sentiment analysis is to know the orientation of the sentiment described in script. In recent decades the researcher focuses on the study various algorithms for relevant research results of the sentiment analysis. This research paper provides a comprehensive overview of this field's latest update. In this review, some recent proposed improvements of algorithms and various SA applications are explored and briefly described. The aim of this paper is to provide knowledge about the different method related to sentimental analysis also how they are classified, what the applications of this analysis.

Keywords: - **Behaviour- Sentimental Analysis, Algorithm comparison, word cloud, Naïve Bayes, Social Media, python-Programming, Machine Learning, data collection,PandaML**

1. Introduction

Marketing concept start with the consumer needs and in behaviour of meeting those needs. Every action of a person in based on needs. The real problem is to learn what a customer takes into consideration when he chooses a particular brand. Such a study isconcerned with consumer behaviour.
Consumer/ buyer behaviour is that subset of human behaviour that is concerned with decisions and acts of individuals in purchasing and using products. Consumer behaviour is a subset of customers behaviour, is concerned with decisions that lead up to the act of purchase, consumer behaviour refers to the behaviour that consumers display in searching for purchasing, using evaluating and disposing of products and services that they expect will satisfy their needs. The study of consumer behaviour is

the study of how individuals make decisions to spend their available resources (time, money, effort) on consumption related items.

Our project observes the ongoing comments, reviews and some theoretical questions which includes the buying patterns and keeps an study tap on what they buy, why they buy it, when they buy it, where they buy it, how often they buy it, and how often they use it? Taking the product computer hardware. this research shows us the aspects on customer's opinion mining and how they individually see the market and what all factors they considered while buying the product , various mind factors like budget-oriented , feasible requirements and latest production are

embedded in the software based algorithm system, to gather various reviews about experiences of users on daily -based services we developed a portal, which sort of collects all the data and then data is trained and encoded through machine learning (ml), we have used here panda (machine learning) **Pandas** has been one

of the most popular and favourite data science tools **used** in Python programming language for data wrangling and analysis. ... And **Pandas** is seriously a game changer when it comes to cleaning, transforming, manipulating and analysing data. In simple terms, **Pandas** helps to clean the mess. Algorithm used for sampling of data is to be navies bayes, now **Naïve Bayes algorithm** is a supervised **learning algorithm**, which is based on **Bayes** theorem and used for solving classification problems. **Naïve Bayes Classifier** is one of the simple and most effective Classification **algorithms** which helps in building the fast **machine learning** models that can make quick predictions.

Sentiment analysis is the field of a study analysing people's opinions, sentiments and emotions towards entities such as products and services, organizations and a vast variety of topics. It uses natural language processing and data mining techniques to the problem of extracting opinions from the text. Sentiment analysis helps in tracking reviews of people about a particular object. This would prove useful in judging the success of a new product launched, how customers react to the product or service. Are they satisfied or they aren't?! We extracted feedbacks for a number of datasets & cleaned it. These comments are basically the reviews of people and can be bifurcated into four buying sentiments, Routine-response, Limiteddecisionmaking, Extensive decision making and Impulsive buying. Our project uses

Machine Learning algorithm techniques which are "Random forest" & "Naïve Bayes". These two are classification algorithms that classify the data into different categories and are a part of supervised machine learning. The purposes of selecting these algorithms are, they give better results for text classification. These two algorithms are ML (Machine Learning) algorithms. Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. Machine learning algorithms are often categorized as supervised or unsupervised. Supervised machine learning algorithms can apply what has been learned in the past to new data using labelled examples to predict future events.
Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

2.Method

The following is an overview of the project, as to what steps, processes were involved in execution of the project.

1. **Data collection through google form survey :** Data collection is the process of gathering and measuring information on targeted variables, here the targeted variables are the four consumer buying behaviour such as complex Buying Behaviour , Dissonance-reducing buying behaviour, Habitual Buying Behaviour, Variety seeking behaviour.
2. The collected Data is in the response form is in different text format with lots of special character, punctuation marks or some data I lower case or some data in upper case so we have to make the data in such a way so that while training data. machine can understand the exactly the way we pass the info machine can understand.

3. **Raw Data pre-processing:** Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. The data collected through a survey on google form .

- Cleaning.
- Transformation.
- Removing errors.
- Validation.
- Removing stop words.
- Removing white spaces.
- Removing hyperlinks if any.

As the raw data cannot be submitted to the algorithm so refining is required. The tweets are written by customers who at times unintentionally make mistakes while typing like grammatical errors, leave blank spaces, add hyperlinks if they have a complaint regarding a particular issue. Some even post screenshots along with the text. This makes it difficult for the system to understand the tweets as, our programming language can't understand read the sentiments from images. So, it becomes necessary to remove all such errors and make the data clean before applying the algorithms & extracting the sentiments.

Data Analysis: Data Analysis is the process of analysing data by applying algorithm. Here, we are applying two different algorithms for two different data set for analysing the sentiment of tweets. The two Renderforst classificsaation & Naïve Bayes Algorithm. These two are the ML (Machine Learning) algorithms which belong to the category of supervised machine learning. The data to be trained is called the testing or training data. The data which is already trained is called tested or trained data. The testing data has to match the trained data for better results.

4. **Data Modelling:** Data modelling is the process of documenting a complex software system design as an easily understood diagram. Over here we are using Iterative Incremental model for implementation of sentimental analysis. In an Iterative Incremental model, initially, a partial implementation of a total system is constructed so that it will be in a deliverable state. Increased functionality is added. Defects, if any, from the prior delivery are fixed and the working product is delivered. The process is repeated until the entire product development is completed. The repetitions of these processes are called iterations. At the end of every iteration, a product increment is delivered. Such a model helps us in understanding what changes or iterations needs to be made at every step & helps in better understanding for theproject.
5. **Desired Output:** The analysed data is then converted into statistical data which is represented in the form of graphical representation and sentimental word cloud, bar plots which is segregated into two sentiments positive & negative. In a word cloud words which are repeatedly found out in Google-Form are present in bigger fonts & words used less in smaller fonts. This indicated whether a repeated word is associated with a positive sentimentand a less repeated word is associated with a negative sentiment or vice-versa.
6. **Algorithm Comparison:** In this module we are comparing:

- The results amongst the two data sets.
- We compare the best algorithms amongst the two algorithms to see which oneis more efficient.
- We check the accuracy, precision and recall values of the two algorithms.

Software Requirement Specification:

Hardware Requirements:

- Computer/Laptop.
- Processor: I3 or Higher.

- RAM: 512 MB or Higher.
- Hard Disk: 16 GB or Higher.

Software Requirements:

- Operating System: windows XP /7 or higher, Linux or mac. ¢ Software: Anaconda Framework, Install Jupiter Notebook

Software Description:

The programming language we used is panda ml programming language Analysis

Architecture & Design is a systematic process that includes various phases. The phase involved in this section gives an overall understanding of the research which is considered for implementation.

The phases included in Architecture & Design are:

- Parametric Analysis

Parametric Analysis

The Parametric analysis defines the parameter used in the research. In our research we are using the algorithm to train the data to understand the sentiments of tweets, using the ML algorithms. The ML algorithms used are SVM & Naïve Bayes. To understand which algorithm is best suited for classification the parameters considered are:

- Accuracy
- Precision
- Recall

Now to understand each parameter Confusion Matrix is used, that gives the values for predefined formulas of parameters.

1.) Accuracy

Accuracy refers to the closeness of a measured value to a standard or known value. The accuracy plays an important role in understanding the algorithm as when the data is trained and tested with related testing data it shows how accurate the algorithm performed for the giving datasets.

Accuracy may vary according to the datasets and gives different values for different data given to the algorithms.

It also varies from algorithm to algorithm depending upon what type of algorithm is used.

2.) Precision

Precision is how close the measured values are to each other. It defines the preciseness of the data. When the data is tested the precision is used to check if the output resulted is close enough or not. The higher the distance between the data points the lower the preciseness ofthe algorithm.

Precision also varies from Algorithm and datasets given.

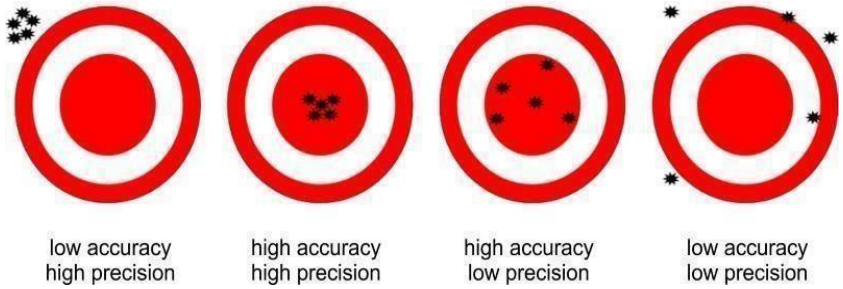
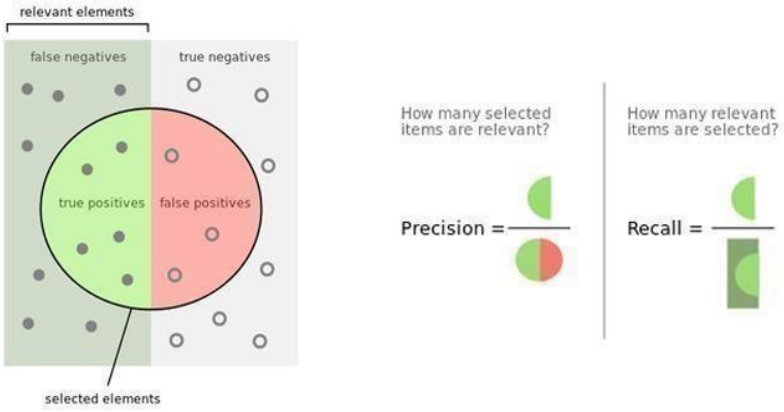


Figure 1: Accuracy vs Precision

3.) Recall

Recall is the fraction of the total amount of relevant instances that were actually retrieved. Both precision and recall are therefore based on an understanding and measure of relevance. Understanding the recall value is similarly important as it shows how much correct data hasbe retrieved.
The recall with help of precision shows the correctness of an algorithm, which can help in deciding the best suited algorithm among different datasets.



4.Results

The methodology of this research includes various implementation which are Sentimental Analysis, Word cloud Representation, SVM & Naïve Bayes Algorithm for training datasets.

The datasets included are the number of reviews &feedback which are extracted from surveyforms.

1.) Datasets & Data source

To conduct the research there are datasets which are different types of buying behaviour:

- Complex Buying Behaviour
- Dissonance Buying Behaviour
- Variety seeking Behaviour

These datasets are extracted from Google form using real time consumer buying behaviours by survey form.
47 responses are extracted and stored in CSV file to further processing.

2.) Sentimental Analysis

The sentimental analysis is process of finding the sentiments of user to understand what they are conveying and, in this research, it is done with the help of text.
The Sentimental Analysis Process includes various steps. These steps are shown with the help of diagram below.

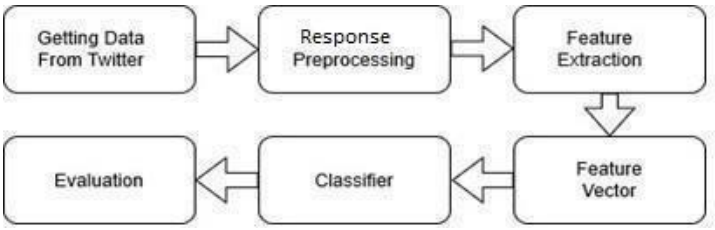


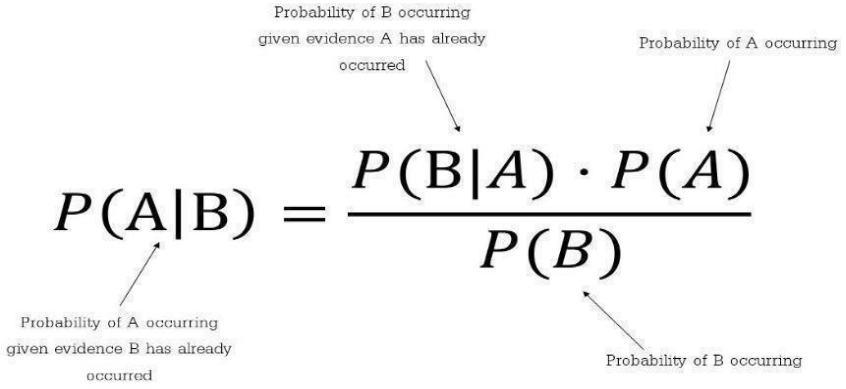
Figure 2: Sentimental Analysis Process The steps are explained below of above figure:

- **Getting Data of consumer responses: -** It is the first step that takes place which is gathering the responses from g form which used tokens and secret key developed by thefor-developer accounts.
- **Data Processing: -** After the responses has been extracted and stored in the form of CSV file. It is loaded and cleaned which is part of processing. Cleaning is the process of transforming the thought's into readable format by removing unnecessary words and steaming the words. This step is followed by Feature Extraction which is the third stepinvolved in analysis process.
- **Feature Extraction: -** It is the critical stage as what feature is needed for problem solving is defined in this stage. Feature can be anything varying from Text, Location, RT etc. These selected features are used to train the algorithm.
- **Feature Vector: -** The Vector is set of words combined together to be used in algorithms. Making a vector is like making a bag of words which are used for problemsolving technique.
- **Classifiers: -** Classifiers are the classification algorithms which are part of Machine Learning. In this research the two classifiers are Random Forest & Naïve Bayes on which the datasets are performed.
- **Evaluation: -** It is the last step of sentimental analysis which defines the understanding of the output results. The results in our case is the Accuracy of the algorithms on Uber & Ola.

These are the steps which are performed for sentimental analysis.

3.) Word cloud Formation

The word cloud is a graphical representation of the words which are most used in the data. The other definition is that the word cloud is an electronic representation of the word most used in survey forms which are highlighted by size. The bigger the size of word the more ithas appeared in the text.
It helps in understanding visually the content.



5. Discussion

These reviews can be studied for analysis of market trends. Sentiment Analysis is one of the important aspects of data mining where important data can be mined based on the positive or negative senses of the collected data. Here the source materials refer to opinions/reviews/comments.[5] The Sentiment found within comments, feedback or critiques provide useful indicators for many different purposes and can be categorized by polarity.[4] By polarity, we tend to find out if a review is overall a positive one or a negative one.

For example, 1) Positive Sentiment in the subjective sentence: “I loved the specification of device ”—This sentence is expressed positive sentiment about the System and we can decide that from the sentiment threshold value of the word
“loved”. So, the threshold value of the word “loved” has a positive numerical threshold value. Negative sentiment in subjective sentences: “This is not a trusted brand” defined sentence is expressed negative sentiment about Devices. The following example of the goggle form Reviews negative sentiment named “not a trusted” and we can decide that from the sentiment threshold value of a word is negative. So, the threshold value of the word “wasn’t” having a negative numerical threshold value. Traditional text mining concentrates on the analysis of facts whereas opinion mining deals with the attitudes.[6] Now, the use of sentiment analysis is growing ever since the use of it has created an impact on the commercial side. The Machine Learning algorithms used are “Naïve Bayes” and “RenderFrostClassifier These algorithms are used to calculate the accuracies, precisions (of positive and negative and neutral corpora) and recall values of (of positive, negative and neutral corpora). The difficulties in Sentiment Analysis are an opinion word that is treated as a positive side may be considered as negative in another situation. Also, the degree of positivity or negativity also has a great impact on the opinions. For example, “good” and “very good” cannot be treated the same. However, the latest text mining gives room for advanced analysis measuring the intensity of the word.[7]

Machine Learning is very popular which uses computer science and statistics. It is concerned with the development of techniques and methods which enable the computer to learn, devise and implement new algorithms. The development of algorithms that enables the machine to learn and perform a task on its own in a particular manner is called machine learning. Today we have many techniques and methodologies which are developed for machine learning tasks for various functions. RenderFrostClassifier is one such

machine learning algorithm. They are used for classification and regression.[1] In another term. RenderFrostClassifier is a classification and regression prediction tool used for prediction for businesses and many other fields. It hence uses machine learning techniques to maximize prediction for better results. A is, therefore, a machine learning algorithm that analyses data for classification. It also gives regression analysis according to the data sets. are used in text categorization, image classification, handwriting recognition, and all the prediction-based functions. It is also based for review collection. Support Vector classification aims is to understand a computationally efficient way of separating planes and learning how a 'good' separating hyperplane differentiates in a high dimensional feature. Hereby 'good' hyperplanes we understand that they are the ones optimizingthe general bounds.[3]

Naive Bayes classifiers classify the algorithm into strong, or naive, independence between attributes of various data points. Popular uses of naive Bayes classifiers include text analysis, medical diagnosis, review systems, etc. Hence this classifier is widely used for machine learning because of its simplicity. Naive Bayes is also known as simple/independence Bayes. For example, a fruit may be considered to be a mango if it is yellow in colour, round, and about 3.5 inches in diameter. These features depend on each other or upon the existence of the other features. All of these properties contribute to the probability that this fruit is or could be a mango. Hence, we call it 'Naive'.[4] Naive Bayes model is easy to understand. It has been helpful to understand large amounts of data sets. Simplicity is the key feature of Naive Bayes also it is known to outperform even highly sophisticated classification methods. Thus, to calculate accuracies of any data sets we can use these two machine learning algorithms.

6. Conclusion

Sentiment analysis can help companies identify customer pain points and areas for improvement in their products or services. By analyzing negative sentiment, companies can gain insights into the specific issues that are causing dissatisfaction among customers and take steps to address them.
Sentiment analysis can also help companies identify trends and patterns in consumers' preferences over time. For example, analyzing sentiment over a period of several years can help companies identify shifts in consumer demand for specific features or technologies, allowing them to adjust their product offerings accordingly.
One of the key benefits of sentiment analysis is that it provides a more objective and data-driven approach to understanding consumers' behavior. By analyzing sentiment from a variety of sources, researchers can gain a more comprehensive understanding of consumers' preferences and decision-making processes, rather than relying solely on anecdotal evidence or intuition.
While sentiment analysis is a powerful tool, it is important to remember that it is not a panacea. Like any research method, sentiment analysis has limitations and potential biases that must be carefully considered. For example, sentiment analysis algorithms may struggle to accurately interpret sarcasm or irony, and may be influenced by factors such as language and cultural differences.

In conclusion, sentiment analysis can be a valuable tool for companies and researchers looking to gain insights into consumers' behavior when purchasing computer hardware. By analyzing sentiment from a variety of sources, researchers can gain a more nuanced understanding of consumer preferences and decision-making processes, allowing them to develop more effective marketing and product development

strategies. However, it is important to use sentiment analysis in conjunction with other research methods and to carefully consider its limitations and potential biases.

Acknowledgment

We take much pride in presenting our Research Paper. During the development of our Research Paper, if we would like to mention the names of certain individuals, without whose assistance, our project would have been difficult undertaking indeed. We hereby pleased to have this opportunity to express our deep sense of gratitude for my Research Paper on **"Sentiment analysis on consumers behaviour buying computer hardware "**, further we are very thankful to , our internal project guide **"Prof Samer Saikh"** whose valuable guidance and suggestions helped us in accomplishing our project. Last but not the least, we would like to thank all our friends,
family members, non-teaching staff and colleagues for their support and individual help.

Reference:

[1] A.Pak and P. Paroubek. „Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320-1326

[2] R. Parikh and M. Movassate, "Sentiment Analysis of User- GeneratedTwitter Updates using Various Classi_cation Techniques",CS224N Final Report, 2009

[3] Go, R. Bhayani, L.Huang. "Twitter Sentiment ClassificationUsing Distant Supervision". Stanford University, Technical Paper,2009

[4] L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitterfrom Biased and Noisy Data". COLING 2010: Poster Volume,pp. 36-44.

[5] Bifet and E. Frank, "Sentiment Knowledge Discovery inTwitter Streaming Data", In Proceedings of the 13th InternationalConference on Discovery Science, Berlin, Germany: Springer,2010, pp. 1-15.

[6] Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data", In Proceedings of the ACL 2011Workshop on Languages in Social Media,2011 , pp. 30-38 [7] Dmitry Davidov, Ari Rappoport." Enhanced Sentiment Learning Using Twitter Hashtags and Smileys". Coling 2010: Poster Volumepages 241{249, Beijing, August 2010

[8] Po-Wei Liang, Bi-Ru Dai, "Opinion Mining on Social MediaData", IEEE 14th International Conference on Mobile Data Management,Milan, Italy, June 3 - 6, 2013, pp 91-96, ISBN: 978-1-494673-6068-5, http://doi.ieeecomputersociety.org/10.1109/MDM.