Method to Facilitate E-Commerce Buying Power by Using Machine Learning Techniques

Junzhi Liu *

Donghua University, Shanghai, China

* Corresponding Author Email: 200720616@mail.dhu.edu.cn

Abstract. The incremental internet usage triggers the rising of e-commerce, a burgeoning shopping mode. Unlike other papers which focus primarily on the technical construction of a sentiment classification model, this paper combines machine learning techniques with business strategies. It aims to determine how sentiment analysis facilitates businesses' improvement of offerings on e-commerce platforms, increasing customers' buying power. First, the paper defines consumer sentiment analysis, summarizes the methods different scholars used when classifying sentiment on aspect level, and points out how sentiment analysis is valuable to both businesses and customers. Second, the paper describes an e-commerce notebook, which covers how sentiment analysis can be carried out using data from Olist online retailing store in Brazil. Naïve Bayes and Logistic Regression are utilized when implementing sentiment classification. Finally, according to the word cloud for positive and negative words in reviews, the paper gives some coming-up suggestions for tackling with the most frequently appeared complaint - the delivery time. Businesses can decompose the supply chain into six sub-systems, and adopt computer vision and GIS system in the packaging management system and delivery management system respectively to squeeze the delivery time.

Keywords: Sentiment analysis; machine learning; e-commerce; natural language processing.

1. Introduction

The boom of e-commerce has changed both the supply and demand side in the global market. On the one hand, after shopping online, consumers are accustomed to sharing comments on logistics, sales services, and the characteristics of the products, which can be referred to by other potential consumers. On the other hand, businesses should either enhance the quality of products and services or make adjustment to future marketing strategies and the construction of customer recommendation system in response to those comments that contain opinions, sentiments and attitudes towards the objects. Sentiment analysis plays an integral role in catering to the needs from both sides.

2. Consumer Sentiment Analysis: Definition, Techniques, and Meaning

2.1. What is consumer sentiment analysis?

Consumer sentiment analysis, also known as opinion mining, is a process of determining whether a particular section of online reviews (such as those of products or services, tweets, online surveys, etc.) is neutral, positive, or negative [1]. In addition, through the use of terms such as opinion, tone, and context, consumer sentiment analysis can be utilized to specifically detect the emotions, urgencies and interests of the customers [1,2]. In terms of the fine granularity of the analysis, consumer sentiment analysis can be categorized into three groups: document level, sentence level and aspect level. The process of document level classification extracts sentiment from the entire review and classifies an opinion based on the overall sentiment of the opinion holder. The goal is to determine the polarity of the whole review. The process of sentence level classification basically includes two steps. The first step is to classify whether the sentence is objective or subjective, and the second step is to do sentiment classification of subjective sentences and divide them into two classes: positive and negative. Aspect level classification identifies and extracts object features that have been commented on by the opinion holder and determines the polarity of the certain aspects of an entity. Feature synonyms are grouped, and a feature-based summary of multiple reviews is produced [3].

2.2. How to do consumer sentiment analysis?

This paper focuses mainly on aspect-based sentiment analysis, which can be divided into two tasks: aspect extraction and aspect sentiment classification. The second task can be accomplished by the use of sentiment lexicons, such as the General Inquirer, Linguistic Inquiry and Word Count (LIWC) and MPQA Subjective Cues Lexicon, or other semi-supervised lexicon that bootstrap a complete lexicon through learning, leveraging merely few labeled examples or a few hand-built patterns [3]. Figure 1 below demonstrates the pipeline of finding sentiment for aspects.

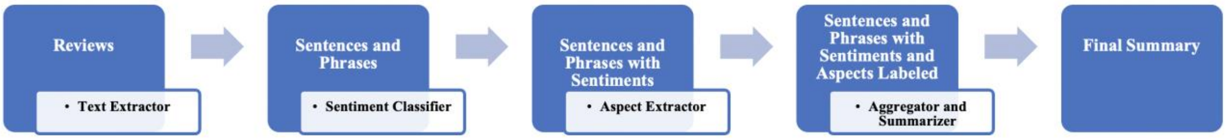

Fig 1. Pipeline of finding sentiment for aspects

But how to filter sentiment efficiently from unstructured texts? Researchers have made contributions to seeking various novel methods in different fields. Jochen Hartmann et al. compared the performance of ten approaches (five lexicon-based, five machine learning algorithms) by using 41 social media datasets spanning social media platforms (e.g., Twitter, Facebook, IMDb, Yelp), three sample sizes, and languages. They concluded that RF and NB perform the best in correctly detecting human instinct [4]. RF, in particular, consistently outperforms for three-class sentiment, NB for small sample sizes. SVM never outperforms the other methods. When compared with machine learning, all lexicon-based approaches, particularly LIWC, perform poorly. They proposed that considering NB and RF can benefit marketing research.

Naresh gathered Twitter data of an airline and created a dataset that contain 1200 client tweets by applying API [5]. 560 tweets were classified as positive, 362 as negative, and 278 as neutral. Among the three classifiers Naresh adopted, DT excels in accuracy, precision, recall and f1-score, with all four scores around 80%, compared with SVM and DT. The accuracy rate of KNN and SVM is only 67.0% and 68.0% respectively. The precision rate of the two models is 70.5% and 69% respectively. The recall rate is 69.3% and 68.1%, and the f1-score is 67.9% and 68.7%. Notwithstanding, some other researchers gave credit to SVM. Demircan, Seller, Abut et al did sentiment analysis on more than 250,000 reviews of products from the hepsiburada.com e-commerce site, and found that SVM and RF are feasible approaches to determine the polarity of the product reviews compared with DT, LR and KNN [6].

Haque, Saber & Shah selected 48,500 Amazon product reviews of three categories: electronics, cell phone and accessories and musical instrument [7]. Active learning was used to label the dataset as positive (ratings $\geq$ 4) or negative (ratings < 3). 3-star ratings were discarded because of neutral meaning. After going through the data pre-processing stage (including tokenization, stopwords removal and POS tagging) and feature extraction stage (in which Bag of Words, TF-IDF and Chi Square were applied), the study assessed the performance of Linear SVM, NB, SGD, RF, LR and DT in predicting the sentiments of the reviews. In general, accuracy, f1-score, precision, and recall of those classifiers are around 90%. Logistic regression did not perform well, while Linear SVM provided the best classifying results, with 10-fold accuracy reaching 94% and other three indicators higher than 97%.

H.J. Alantari, I.S. Currim, Y. Deng et al collected data across nine product types from five review sites in the two settings of review sites (Tripadvisor.com, Amazon, drugs.com and Goodreads) and social media (airline's twitter) [8]. In the setting of review sites, consumers express sentiments both in scores and comments; In the latter setting, users often deliver unformal texts and emojis. The study found that several neural network-based learning models, including models that are not trained beforehand (FastText, CNN-LSTM, and BiLSTM) and pre-trained models (BERT, XLNet, RoBERTa, DistilBERT, ALBERT), outperform other machine learning methods, such as KNN, NB, SVM, LR, and RF that have been applied by marketing scholars and others such as Ordered LR, DT, XGBoost and AdaBoost that have not, in terms of the mean and standard deviation of accuracy.

2.3. Why consumer sentiment analysis?

Through consumer sentiment analysis, firms detect the polarity of the text sentiment, gage brand reputation of competitors and gain insight into the actual demands of customers [6]. Businesses can therefore improve their offerings and design new products by predicting the fashion trend, marketers can therefore reach out to those who require extra care, and the e-commerce platform that connects B with C can optimize the searching engines and the recommendation system based on the sentiment of the reviews and the preferences of various customer segmentations, which eventually elevate customer satisfaction and sales [9]. While for customers, the risks of buying something with unexpectedly inferior quality are reduced, since the reviews left by others facilitate new buyers to make the right decision.

3. Application of Consumer Sentiment Analysis in a Brazilian E-Commerce Case

This section describes the notebook from Thiago Panini [10] who first went through an exploratory data analysis on Brazilian online purchasing. Then Thiago implemented sentiment analysis on customer reviews to make a text classification using NLP tools.

3.1. Reading the Data

3.1.1 An Overview from the Data

The Brazilian ecommerce public dataset used by Thiago Panini gathered orders made at Olist Store. Olist is one of the largest online retailing platforms in Brazil. The dataset contained information on 100,000 orders placed at various Brazilian marketplaces between 2016 and 2018. Key features of the dataset include item description, price, order status, payment, freight value, review scores and comments.

3.2. Natural Language Processing

3.2.1 Data Understanding

In total, there are 41,753 comments which can be utilized for training the sentiment analysis model. However, in the first place, the comment input should be transformed into a vector that a Machine Learning model can interpret [10]. As shown in Table 1.

Table 1. A Sample of the Dataset (41753, 2)

	score	comment
0	5	Recebi bem antes do prazo estipulado. (I received it well before the stipulated time.)
1	5	Parabéns lojas lannister adorei comprar pela I…(Congratulations Lannister stores I loved shopping for I…)
2	4	aparelho eficiente. no site a marca do aparelh…(efficient device. on the website the brand of the device…)
3	4	Mas um pouco ,travando...pelo valor ta Boa.\r\n (But a little slowing down...for the price it's good.)
4	5	Vendedor confiável, produto ok e entrega antes…(Reliable seller, ok product and delivery before…)

3.2.2 Regular Expressions

In the pre-processing stage, Thiago Panini applied RegEx to delete the HTML tags, break lines and other special characters. a) Break Line and Carriage Return: Filter \r and \n. b) Sites and Hyperlinks: Define another function re_hiperlinks (text_list) and apply RegEx to it. Replace the hyperlink with the word ‘link’. c) Dates, Money, Numbers: Apply RegEx to dates, money and numbers in the text. d) Negation: When removing negative words, such as não (not), the meaning of the phrases and sentences may be altered. Therefore, Thiago Panini replaced them with “negação” ("negative" in English). e) Special Characters: Delete emojis and punctuation marks in the same way. f) Additional Whitespaces: Eliminate unnecessary whitespaces to clean the text.

3.2.3 Stopwords

Referring to the corpus in NLTK, remove all Portuguese stopwords to further clean the data.

3.2.4 Stemming

Define a function "stemming_process(text, stemmer=RSLPStemmer())" to apply the stemming process to the comments.

3.2.5 Feature Extraction

After the preprocessing of data, Thiago Panini used approaches like Bag of Words, TF-IDF and Word2Vec to implement feature extraction. To make the analysis easier, Thiago defined a new function that takes a text and a vectorizer object, and applied the feature extraction process to the text. a) CountVectorizer: On the Bag of Words approach, Thiago created a dictionary vocabulary with all Portuguese affix. Thiago indexed each word, which is extracted from reviews, into a vector that represents the occurrence (1) or absence (0) of each word. This is a method for converting a text into a frequency vector using word dictionary (a literal bag of words) [10]. The (41753, 300) matrix that contains corpus features was created using the 300 most common words with at least 80% frequency and occurs in at least 7 text strings in the corpus. b) TF-IDF: However, the Bag of Words Model merely gives hints about the occurrence by designating the same weight to each word. Whereas calculating TF-IDF (Term Frequency and Inverse Document Frequency) helps assess the significance of a word, which increases proportionally to the number of times a word appears, but scales down when the inverse document frequency approaches 0. Following the two formulas, the TF-IDF method can be utilized within the scikit-learn library.

3.2.6 Labeling Data

To train a supervised machine learning model, by adding a review score column, Thiago labeled the comments with 0 (negative comments with review score ranging from 0 to 3) and 1 (positive comments with review score ranging from 4 to 5). As shown in Figure 2.

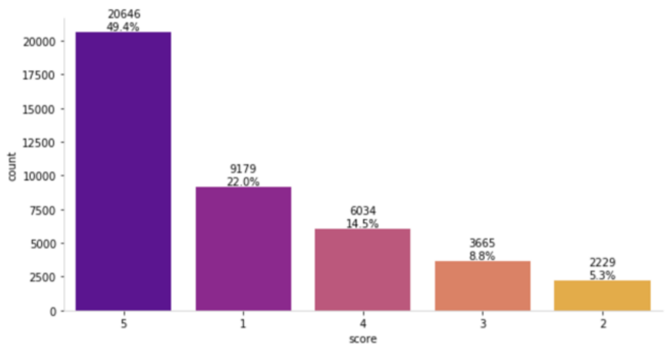

Fig 2. Distribution of review scores

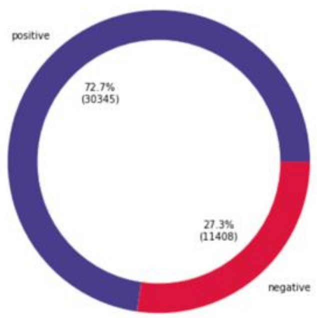

Fig 3. Ratios of positive and negative comments

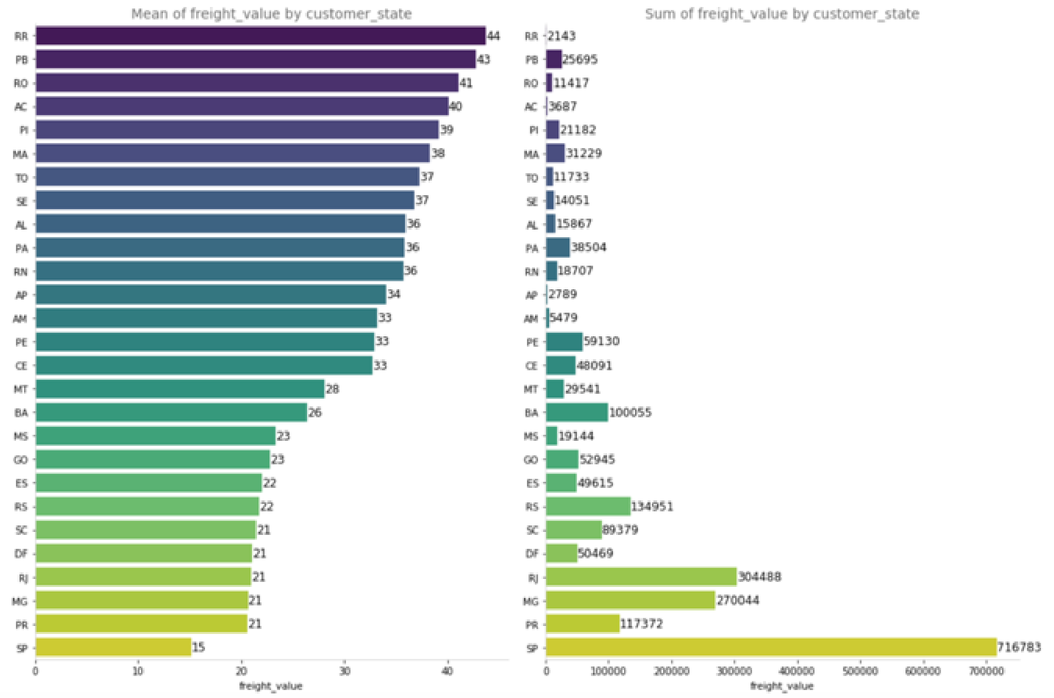

Fig 4. Distribution of sum and mean freight value within states in Brazil

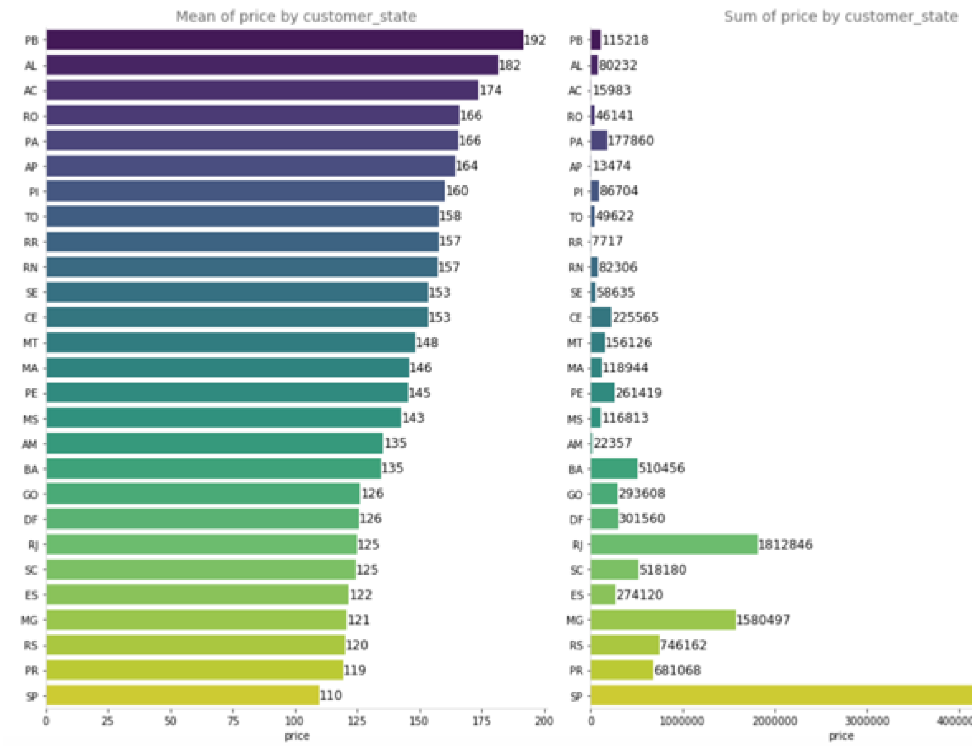

Fig 5. Distribution of sum and mean price within states in Brazil

Indicating from Figure 2 and Figure 3 above, nearly half of all commenters gave five stars, with 22% marking only one star. Generally, three-quarters of all commenters are satisfied with the online stores. The most frequently appeared positive comment is “delivery on time”. But the majority of negative comments are also related with the timeliness, reading “haven’t received yet”. The second highest complaint is about the number of products received. The divergence of opinions on the delivery time within the two classes may lie in that the freight each customer pays varies in different states. For instance, as shown in Figure 4 and Figure 5, SP (São Paulo) ranks the first in both the sum of price and the sum of freight value, but the lowest in mean value of the two indicators. In the contrary, people in Roraima (RR) pay more on freight per order. This results from the fact that the more remote a state is from the metropolis, the fewer orders will be made and the more expensive it is to transport the goods. Hence the freight is higher, and it is more likely that the goods will not arrive timely.

3.3. Sentiment Classification

After building the pipeline of applying regex transformers and extracting features from the corpus using TF-IDF, Thiago trained an algorithm that classifies the sentiment of a text string by adopting Logistics Regression and Naïve Bayes. As shown in Table 2. Generally, Logistic Regression performs better in accuracy (acc=$\frac{TP+TN}{TP+TN+FN+FP}$=0.88), recall rate (R=$\frac{TP}{TP+FN}$=0.92), f1-score (=$\frac{2\times P\times R}{P+R}$=0.92) and the total running time, while Naïve Bayes has slightly higher precision rate (P=$\frac{TP}{TP+FP}$=0.94).

Table 2. Comparison between Logistics Regression and Naïve Bayes

	model	approach	acc	precision	recall	f1	total_time
0	Logistic Regression	Train 5 K-folds	**0.884**	0.9216	**0.9189**	**0.9202**	**5.452**
1	Logistic Regression	Test	**0.8834**	0.9219	**0.9161**	**0.919**	**0.034**
2	Naïve Bayes	Train 5 K-folds	0.834	**0.9354**	0.8292	0.8791	7.281
3	Naïve Bayes	Test	0.8342	**0.9345**	0.8284	0.8782	0.084

3.4. Final Implementation

The last step is to define the transformers for preparing the text input. Simulate three reviews to feed the sentiment_analysis function. For example, the result of "Awful product! I no longer shop at this store because the delivery was late and it was so expensive" is 97% negative. The result of "It's fantastic, and it exceeded my expectations. I bought it at a low price. Wonderful" is 99% positive. The result of "I'm not sure if I liked this product. The price was low, but the quality was poor. If you're lucky, it's worthwhile" is 70% positive. According to the word cloud for positive and negative words on the dataset, "Produto (product)", "antes (before)" "prazo (deadline)", "chegou (has arrived)", "entrega (delivery)", "qualidate (quality)", "bom (good)" are the words most frequently occurred in the positive comments, indicating that the Olist store in Brazil performed well in the quality of the products, but there was controversy over delivery service.

3.5. Coming-Up Strategies

Sentiment analysis uncovers potential problems of the online store. Dwindling delivery time should be prioritized. It can be inferred from the dataset that the average delivery time was 9 days from January 2017 to December 2018, and 3.2 days were required to process the orders before dispatch. Orders that could not arrive timely accounted for more than 8% of total orders. In peak months such as November, December, February and March, the delivery time rose 30% to approximately 11-13 days in 2018 compared with other months. During that period, orders that could not be dispatched on time increased by 15-21%, including 18 orders with sales value greater than 1000 [11]. Figure 6 shows average order processing time and average total delivery time in different cities. To tackle with the problem, first, decompose it into smaller tasks. Figure 7 below demonstrates the entire flow chart of supply chain management, which involves six sub-systems: Inventory Management, Purchase Management, Sales Order Management, Packaging Management, Delivery Management and Financial Management. Delay in delivery time may stem from either lack of inventory, redundant order processing procedures, or transportation problems. For businesses concerned with shortage in inventory, the Olist platform should contact them proactively before online shopping festivals each year as for whether they have already made orders based on sales last year. Furthermore, in the packaging management cycle, Olist can introduce the automatic sorting system to squeeze the order processing time. The scientific core of the automatic sorting system is the use of computer vision, which is a branch of artificial intelligence research that aims to assist in making the correct decision about the description of objects and scenes in an image [12]. Thanks to computer vision, computers and systems can take actions or make recommendations based on the meaningful information derived from digital images, videos and other visual inputs. Popular applications of computer vision include 3D map plotting, face recognition, self-driving vehicles, motion detection and so on. In warehouses, Staff only scan the bar code and enter the description of products into the sorting system. With the help of computer vision and other deep learning techniques, the sorting robots then automatically select the goods according to the weight, material, color and other information of the products. Thus, by “feeling” the ambient world, the sorting robots navigate itself and deliver the goods to designated places within short period. In addition, the sorting robots gives feedback about the number of in-stock goods and warns about supply shortage.

In the delivery management cycle, Olist can utilize GIS system. On one hand, businesses obtain technical support for vehicle dispatching, location planning for distribution outlets, process monitoring and the optimization of the distribution route. On the other hand, customers are provided with visualized information considering the estimated delivery time and the real-time location of the products in-transit. According to the location, customers are enabled to either pick the products on their own or wait for delivery [13]. In peak seasons, presale products with significant discounts may divert delivery pressure to some extent.

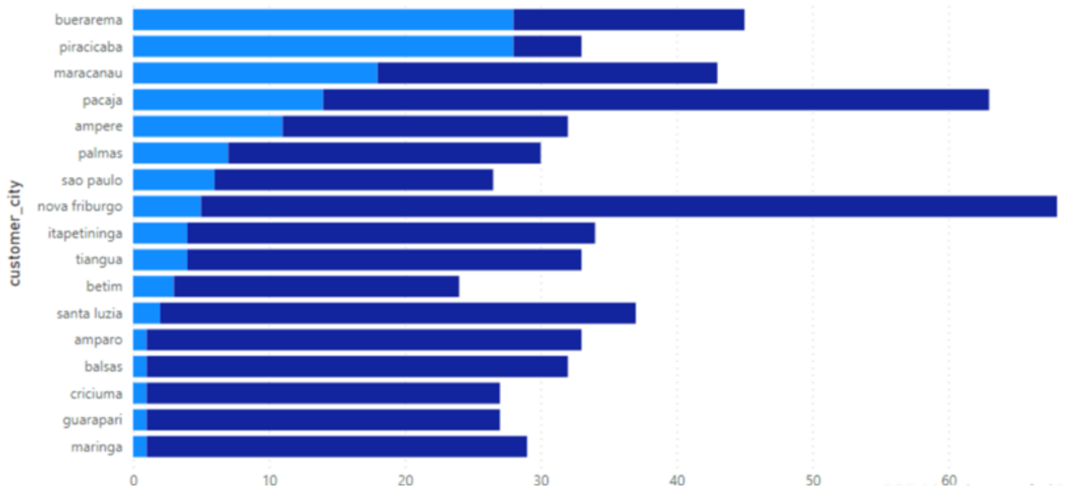

Fig 6. Average order processing time and Average total delivery time in different cities

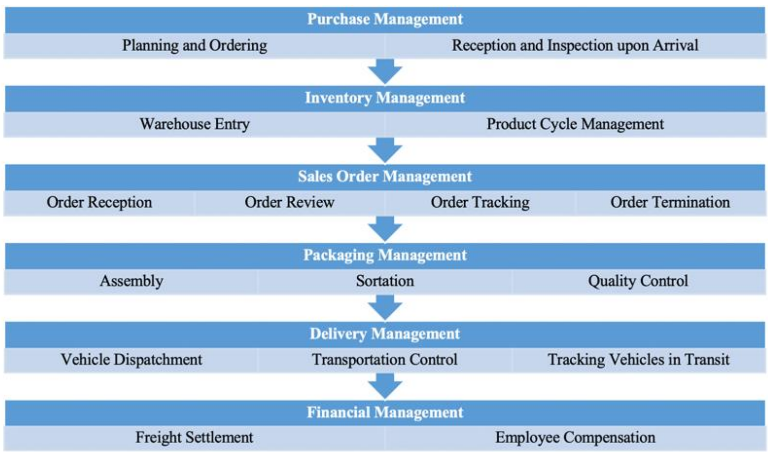

Fig 7. Flow chart of supply chain management

4. Conclusion

The paper discusses how sentiment analysis facilitates e-commerce buying power, aided by machine learning techniques. Sentiment analysis enables businesses to understand customers’ aggregate opinions and attitudes towards certain products by distinguishing the polarity of the reviews, and also helps customers make the right decision. Aspect extraction and sentiment classification of each aspect are the two steps that everyone will go through when implementing sentiment analysis. In recent years, researchers have been searching for the optimal classifier for aspect sentiment. In general, lexicon-based approaches, are inferior to machine learning algorithms. Also, some scholars found that neural network-based learning models outperform other machine learning methods. To put the methodology into practice, the paper describes an e-commerce notebook from Kaggle, which covers how sentiment analysis can be carried out using e-commerce data from Olist store. First, use RegEx to delete the HTML tags, break lines, numbers, negation, other special characters and additional white spaces. Second, delete the stopword and apply the stemming process. Third, implement feature extraction using CountVectorizer and TF-IDF. Fourth, label the data as 0 or 1, and plot the n-grams of the comments on the two classes. Fifth, do sentiment classification using Naïve Bayes and Logistic Regression, and analyze the word cloud for positive and negative words in reviews. It is obvious that Logistic Regression performs better in this case, with precision, accuracy, recall rate and f1-score all around 90%. Eventually, the paper gives some coming-up suggestions for handling the most frequently appeared complaint - the delivery time. Businesses can decompose the supply chain into six sections, and adopt computer vision and GIS system to squeeze the delivery time. It is worth noting that for businesses, the main goal of doing sentiment analysis is to diagnose problems in products or services, and solve them. The former is closely related with cutting-edge algorithms, while the latter is all about business.

References

[1] Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. Computer science review, 41, 100413.

[2] Basmmi A B M N, Abd Halim S, Saadon N A. Comparison of web services for sentiment analysis in social networking sites[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2020, 884(1): 012063.

[3] Katrekar A, AVP B D A. An introduction to sentiment analysis[J]. GlobalLogic Inc, 2005.

[4] Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. International Journal of Research in Marketing, 36(1), 20-38.

[5] Naresh, A. (2021). Recommender system for sentiment analysis using machine learning models. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(10), 583–588.

[6] Demircan, M., Seller, A., Abut, F., & Akay, M.F. (2021). Developing Turkish sentiment analysis models using machine learning and e-commerce sata. International Journal of Coginitive Computing in Engineering, 2, 202-207.

[7] Haque, T. U., Saber, N. N., & Shah, F. M. (2018, May). Sentiment analysis on large scale Amazon product reviews. In 2018 IEEE international conference on innovative research and development (ICIRD) (pp. 1-6). IEEE.

[8] Alantari, H. J., Currim, I. S., Deng, Y., & Singh, S. (2022). An empirical comparison of machine learning methods for text-based sentiment analysis of online consumer reviews. International Journal of Research in Marketing, 39(1), 1-19.

[9] Vyas, V., & Uma, V. (2019). Approaches to sentiment analysis on product reviews. In Sentiment Analysis and Knowledge Discovery in Contemporary Business (pp. 15-30). IGI global.

[10] Panini, Thiago. (2021). E-Commerce Sentiment Analysis: EDA + Viz + NLP. https://www.kaggle.com/code/thiagopanini/e-commerce-sentiment-analysis-eda-viz-nlp/notebook.

[11] Jameslei23. (2021). Olist Retailing Data Analysis. https://blog.csdn.net/jameslei23/article/details/120365008.

[12] Moon, S., Kim, M. Y., & Iacobucci, D. (2021). Content analysis of fake consumer reviews by survey-based text categorization. International Journal of Research in Marketing, 38(2), 343-364.

[13] Zhao Rong, Guo. (2023). Application research on smart distribution system based on smart logistics. China Storage and Transport Magazine (01),114-115. doi: 10.16301/j.cnki.cn12-1204/f.2023.01.061.