International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Amazon Reviews Sentiment Analysis, Segmentation, Classification and Prediction leveraging Multi-Class Multi-Output Classification

Ramana Nagavelli[1,] Ette Hari Krishna[2] and Dr. Katta Padmaja

Submitted: 25/09/2023 **Revised**: 16/11/2023 **Accepted**: 27/11/2023

Abstract: Most users provide their reviews on the assorted products on the Amazon website. The reviews provided by users are most often compact. Hence it becomes a loaded source for sentiment analysis. The Sentiment Analysis is a substantially employed method for locating and obtaining the appropriate polarity of text sources. This project centers on a contrastive study of machine learning techniques for classifying the emotions of the considered product reviews dataset into Positive polarity, Neutral polarity, and Negative polarity, segment into Product, Delivery, Packaging categories. This can be helpful for consumers who want to look at the reviews of products before purchase and for companies who wish to look at the public's reaction to their products. In this project, we correlate the performance of algorithms which support multi class multi output classifications, the accuracy of classifying the sentiment of an unknown review, insight on sentiment analysis, segmentation, and therefore comparison of the performance of the considered algorithms for the classification of the sentiments supported by several performance metrics.

Index-Terms - *Amazon Reviews; Segmentation; Opinion Mining; Sentiment Analysis; Machine Learning Classification*

1. Introduction

Introduction Innumerous new applications have been created as a result of latest developments in Natural Language Processing (NLP) and writing excavating, as well as improvements in information technologies [1]. Research on practical issues can be done to a great extent thanks to new techniques, which are usesin artificial intelligence, growing the volume and variability of documented data fashioned on the regular foundation. One of the most essential concepts in NLP and the text mining is text classification of bids. Associating pertinent material with labels that already exist is the definition of the text categorization problem. In this regards of variety category, dataset structure is decisive. Each text can be symbolized by a single label or several labels, depending on the obstruction's present-day state. This has resulted in changes.

Prevalent and moderately forthright text arrangement based on construction is known as binary classification. Each text is patented so that it can be identified by one of two stickers. Applications for binary text tagging include false news documentation [2], junk electronic-mail discovery [3], and junk mail assessment exposure [4], and composition verifications [5]. There are more than two tags in multi-class and multiple-output text cataloguing in contrast to binary classification. In mutually reproductions, each text is represented by a single tag, which is what binary text classification and multi-class text cataloguing have beenmutual. Multi-cast textual classification challenges include sentiment analysis [7], topic modelling [8], and synonym extraction [9].

Despite the fact that both binary and multi-class text categorization generate effective results, they fall short of especially in circumstances when people's opinions are required, such as in relation to items and proceedings, as the verbal we use and the terminologies we generate in these sceneries have more nuanced meanings. Even while it is conceivable, limiting textual expression to a single label frequently makes it unbearable to excerpt more exact data from edition. Multi-label data are therefore required in light of the advancements in information technologies in order to fulfil user expectations. The analysis of multi-label data makes use of text classification methods with several labels. In other arguments, each text in a multiple-taged text classification system may have just one marker or multiple labels. Despite the fact that both binary and multi-class text categorization generate effective results, they fall short of

Artificial intelligence applications as well as human behaviors, lives, and expectations have been impacted by developments in information technologies. The People have underway to utilize the online data setsdeeply to work, buy, have fun, and learn; in another way of work, people have underway to practice it in their

[1]*Assoc. Professor of CSE, UCE, Kakatiya University Kothagudem, Telangana, India. ramanauce.ku@kakatiya.ac.in*
[2]*Asst. Prof of ECE, UCE, Kakatiya University Kothagudem, Telangana, India. hari_etta@kakatiya.ac.in*
[3]*Assitant Professor CSE, University college of engineering, Kakatiya University. ajit2107@kakatiya.ac.in*

daily sequences. This is specificallyright in the supposedpublicarithmeticaltime induced by digitalizationin the COVID-19 pandemic circumstances. The best example of this influence is in electronic commerce (E-Commerce). According to recent data [10], overall 93.5% of social dependentworkers have made aconnectedoperation. In the US, 41% of consumersaccept one or more correspondences from Amazon every week; this figure increases to 50% for clienteles between the ages of 18-25 and upto 57 for those between the ages of 26 and 35 [10]. Approximately [11] that by 2040, e-commerce wills explanation for 95% of all connections. Consumers'online purchases are greatly influenced by product reviews; 55% of internet shoppers speak with groups and household when they were dissatisfied by a brand before business [10]. Additionally, 90% of shoppers examine before, gives the reviews by the visited customers placing an order. In this consequence, researchers have become interested in the volume of data that is gathered every day and how it affects customers. As a result, numerous studies on e-commerce customer reviews have been carried out.

According to [12], multi-tag analysis has not been considered and undertaken, and majority of researches conducted in the nonfictions are grounded on polarization analysis. The methodology we suggest here seeks to identify the numerous stickers included in appraisals, in contrast to the standard categorization strategies. Multi-Class, Multi-Output studies of this data are crucial because they allow for the discovery of insightful information that can benefit both those who purchase products and productions observing to enhance their customer relationship between the management via customer periodicals. Consider a scenario where someone is looking for a product that they need to acquire right now. Their main need during this process is for the merchant to send the items out right away. It would be beneficial for someone like that to read reviews by cataloguing them permitting to their standing. This prediction marks it decisive to contemplate both the sensitive and qualitative phases of consumer reviews. In this proposal inspired that, to categorize e-commerce consumer evaluations using a multiple-class multiple output technique for the aforementioned reasons as well as the shortage of comparable revisions on this issue in the literature work.

2. Scope and Objective of the Paper

A. Scope:

The initiative offers three key contributions to the study of e-commerce customer reviews for consumers.

- Aspect-based analysis of customer evaluations, as opposed to polarity-based analysis, is used to ascertain the specific thoughts of customers regarding items.
- Create a new multiple-label consumercriticism dataset for e-commerce. Researchers may be able to compare consumer behavior using this data collection. Which reviews mention the reviews' intentions, such as delivery, packing, and product reviews?
- Construct a multi-output classifier that aids in a better comprehension of the review text by doing a multi-label customer review analysis using a variety of algorithms and measurement approaches.

B. Objective:

- To determine the various output labels for review segmentation as product, packaging, delivery reviews.
- To classify the sentiment of the reviews into various sentiments such as Positive, Negative, Neutral Reviews.
- To develop a prediction system that can be used to classify an input customer review into various aspects as stated above to further understand the customer intensions on the reviews provided by them.

3. Literature Survey

A. Existing Systems

The regular evaluation of commons insolences, philosophies, and the sentiments towards the certain object is known as sentiment analysis or opinion mining [13]. Numerous researches have carried out customer review sentiment analysis. Polarity analysis has been the main topic of almost all of these works. By employing grid search and uni-gram for feature extraction, Muslim [14] aims to increase the Support Vector Machine (SVM) accurateness aimed at categorizing e-commerce consumer assessment datasets. They utilized datasets made up of Lazada and Amazon reviews that had been classified as good or negative. Their test findings demonstrated that the support vector machine (SVM) technique could be enhanced using unigram and grid search to increase the precision of Amazon reviews by 26.4% - 80.8% and Lazada reviews by 4.26% - 90.13%.

Proposal by Vanaja et al. [16] analyses sentiment at the aspect level for Amazon client review records. They observed examined the reviews' positivity, negativity, and neutrality. In their comparison investigation, they claimed to have institute 0.9023 correctness expending naive Bayes. Embedded based technical sentimentality investigation of e-commerce submission product assessments was given by Jabbar et al. [17]. They have designed a model for sentimentality investigation of the

Amazon review data they had gathered using SVM. Reviews were classified as auspicious or uncomplimentary. The sentimentalityenquiry of the analyses using SVM yielded an F1 score of 0.9354.Cutting-edge order to supplementary catalogue to the new review after gathering reviews from e-commerce websites.

Author Tripathi et al. [19] looked at the writing content of appraisals gathered from e-commerce websites with several effectiveness polls. They claimed that the 0.945 accuracy attained with a classifier using random forests.

A deep learning method for analysis sentimentality investigation was put out by Guan et al. [22]. They gathered Amazon criticisms are dividing the emotion into positive and negative categories. Their deep learning approach has a review sentiment analysis accuracy of 0.877.

A direct weighted multiple-classification model for e-commerce reviews was put out by Zhang et al. [12]. 10,000+ assessments from Amazon Criticism were used. For review sentiment, they employed multi-label categorization.

All the web sites which are direct weighted model have a recollection of 0.8 on average. A new-brand sentiment analysis prototypical terminal MBGCV was put out given by GU Et Al. [24]. 31,107 reviews that had been secret as good or bad were included in the revision. An accuracy of their suggested model in terms of review sentiment analysis was 0.94. The Turkish customer evaluations were subjected to LSTM network-based sentiment analysis by Bilen et al. [25]. For sentiment analysis, they made use of two separate datasets. They assimilated a brand-new quantity of more than 7000+ evaluations in order to undertake a sentiment analysis of Turkish customers' preferences. They classified the data they got as either positive or negative using an LSTM-based model, and they found that the accuracy for binary sentiment analysis was 0.905.

Bilen et al. [25] submitted the Turkish customer evaluations to LSTM network-based sentiment analysis. They used two distinct datasets for sentiment analysis. They acquired a brand-new corpus of more than 7,000 critiques in order to conduct a sentiment analysis of Turkish customers' preferences. They classified the data they got as either positive or negative using an LSTM-based model, and they found that the accuracy for binary sentiment analysis was 0.905.

Acikalin et al. [27] used BERT to do sentiment analysis on positive and negative customer movie and hotel reviews in Turkey. They stated that the highest result they came across during their investigation was 93.3%. Santur [28] classified Turkish e-commerce customer reviews using gated recurrent unit sentiment assessments as excellent, terrible, or neutral, and their greatest performance was an accuracy rate of 0.95 percent. Ozyurt et al. [29] successfully completed the aspect-based sentimentality evaluation of Turkish appraisals using LDA. They gathered 1292 user assessments, positioned them, and well-defined nine smartphone device elements. In these outcomes, they discovered an F-score assessment of 82.39%.

Bilen et al. [25] subjected the Turkish customer evaluations to LSTM network-based sentiment analysis. They used two distinct datasets for sentiment analysis. They acquired a brand-new corpus of more than 7000 critiques in order to conduct a sentiment analysis of Turkish customers' preferences. They classified the data they got as either positive or negative using an LSTM-based model, and they found that the accuracy for binary sentiment analysis was 0.905.Turkishconsumermovie and hotel reviews with positive and negative labels were subjected to sentiment analysis by Acikalin et al. [27] using BERT. They claimed that 93.3% was the highest outcome they discovered during their research.Santur [28] used Gated Recurrent Unit sentiment analysis to categorize Turkish e-commerce customer evaluations as good, bad, or neutral, and their best result was 0.95% of accuracy.Utilizing LDA,Ozyurt et al.[29] passed out the aspect-based sentimentalityexamination on Turkish appraisals. They gathered and poised 1292 user analysescollected and well-defined nine aspects for smartphones devices. They found an F-score evaluation of 82.39% in these consequences.

B. Drawbacks in Existing System

All of the classifiers or regressions discussed above have a single output and are based on either multiple labels or multiple classes as inputs. They rely on several types of tagged data to make predictions, which increase the requirement for big data sets. The majority of forecasts are based on small data sets with few reviews that don't truly advance the research.

C. Proposed System

Different insertingmethodologies, both are the frequency-based and prediction-based, as well as various classification methods are used throughout the project in our suggested methodology, as shown in below Fig. 1.

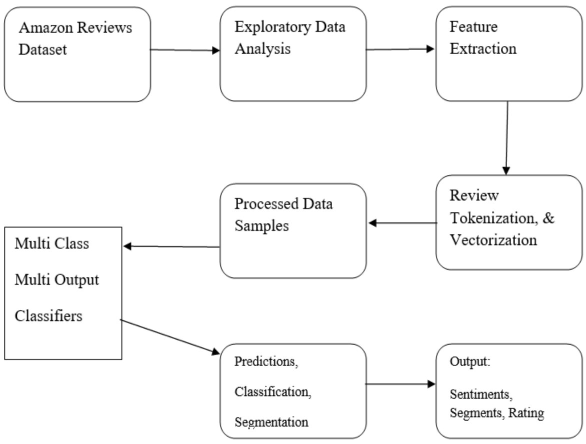

Fig. 1 Proposed System Graphical Illustration.

The methodologies Term Frequency-Inverse Document Frequency (TF-IDF) and prediction-based Global Vectors(GloVe) for Word Representation are the embedding techniques employed in this study. In the following step, Multiclass-Multioutput Classifiers such Random Forest (RF), Support Vector Classification (SVC), K-Nearest Neighbor, Decision Tree Classifier, and AdaBoost Classifier are utilized like given below.

1. We scrape amazon product review from website
2. Strainer dataset permitting to feature requirements and create a new dataset which has attributes according to analysis to be done.
3. Achieveuser reviews Pre-Processing on the dataset.
4. Split the data into two different training set data model and testing data set model.
5. Inspect the testing dataset using the classification technique after training the model with training data.
6. Finally, you will get results as accuracy metrics.

1. IV. System Architecture

A. Dataset

The following Table 1 shows the various data columns present in the dataset.

TABLE I

AMAZON CUSTOMER REVIEW DATA COLUMNS

Columns	Description
marketplace	The two-letter countrycode referees for the nation where the criticism was published.
Customer_Id	Customer identification is the Random identifier that can be used to aggregate reviews written by a single author.
Review_Id	The Unique ID assign to the review.
Product_Id	The unique Product ID the review pertains to.
Product_Parent	Random identifier that can be used to aggregate reviews for the same product, which customer viewed.
Product_Title	Title of the product, for search base.
Product_Category	Broad product type that can be used to group reviews (also used to group the dataset into coherent parts).
Star_Rating	The 1-5 star rating of the review.
Helpful_Votes	Number of helpful votes.
Total_Votes	Number of total votes for the review received.
Vine	Review was written as part of the Vine program.
Verified_Purchase	The review is on a verified purchase.
Verified_Purchase	The review is on a verified purchase.
Review_Headline	The title of the review.
Review_Body	The review text.

The statistics is a tag ('\t') encircled text file, without appraisals. The first line of each file is the header; No.1 line corresponds to unique record. The Fig. 2 below illustrations the column and row data of the statistics from the Jupiter laptop implementation.

B. Data Collection:

The dataset used consists of more than 10,000,000 reviews from amazon's books category available at https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10319090 entries, 0 to 10319089
Data columns (total 10 columns):
 #   Column             Dtype
---  ------             -----
 0   customer_id        int64
 1   product_id         object
 2   star_rating        int64
 3   helpful_votes      int64
 4   total_votes        int64
 5   vine               object
 6   verified_purchase  object
 7   review_headline    object
 8   review_body        object
 9   review_date        datetime64[ns]
dtypes: datetime64[ns](1), int64(4), object(5)
memory usage: 787.3+ MB
```

Fig. 2 Description of Amazon Reviews Data Set.

C. Data Pre Processing:

The dataset was pre-processed with several techniques as mentioned below to acquire the exact text needed for classification.

Three common data pre-processing steps are:

1. Structuring
2. Scrubbing
3. Sampler

The above steps can be performed using following techniques:

1. Take URLs and email addresses out of each and every sample because they don't offer anything useful.
2. Get rid of the punctuation; otherwise, your model won't realise that "good!" and "good" imply the same thing.
3. Lowercase every word, since you want the input text to be as general as possible and prevent words like "Good" at the beginning of a sentence from being read differently from "good" in another instance.
4. Get rid of "stop-words" – These are the furthermostpredominantarguments in a language, such "I," "have," "are," and so on.
5. Stemming and lemmatization: These two tasks, which both aim to extract the root words from each word in a phrase in the corpus data, are quite similar.
6. Transmute dataset (text) into numeric tensors — usually referred to as vectorization.

D. Sentiment Analysis:

In particular, sentimentalityinvestigation is one of the major capacities where NLP has been largely used. For organizations it is crucial to understand customer behavior and needs for the company's products and services. Generally, it is possible to classify the feedback of a customer about a product into positive, negative and neutral. Companies are able to assess the satisfaction of customers with their products or services by means of listening to customer feedback through product reviews. A sentiment analysis is a process by which positive and negative messages are detected in text. Businesses generally use it for the purpose of detecting emotions on social media, measuring brand reputation and understanding their customers.

Valence aware dictionary for sentiment reasoning (VADER) is popular rule-based sentiment analyzer. For the calculation of text sentiment, it uses a list of Lexie features.eg. The Confrontations that have been coded as good or bad based on their Semantic orientation. Vader sentimentality returns the possibility of a given input condemnation to be Positive, negative, and neutral.

Making use of VADER values and the rating values provided in the data set, a custom metrics is derived to perform sentiment analysis which is as shown in below flow chart Fig 3.

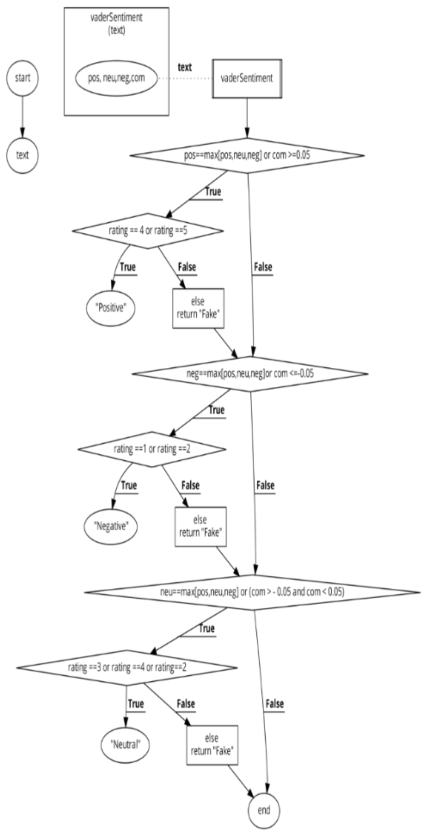

Fig. 3 Flow Diagram of Sentiment Analysis.

With the above flow chart, it’s clear that we can perform not only sentiment analysis in terms of positive, negative, neutral sentiments but we are also able to understand if the reviews are fake or wrongly marked with ratings.

Thus, we can eliminate the reviews which are mapped to fake as they may be fake reviews or ratings are false as they are not mapping to the sentiment of the reviews by the user. Thus, we drop these reviews there by reducing the reviews dataset size which helps in fast computing.

E. Segmentation:

To segment a review into product, delivery, packaging categories in the project, had to apply regular expressions to find the review text which contains delivery / packaging as tokens, if the text contains any of the categories in the text they will be labeled accordingly. Flow chart to represent this process is as described below fig 4.

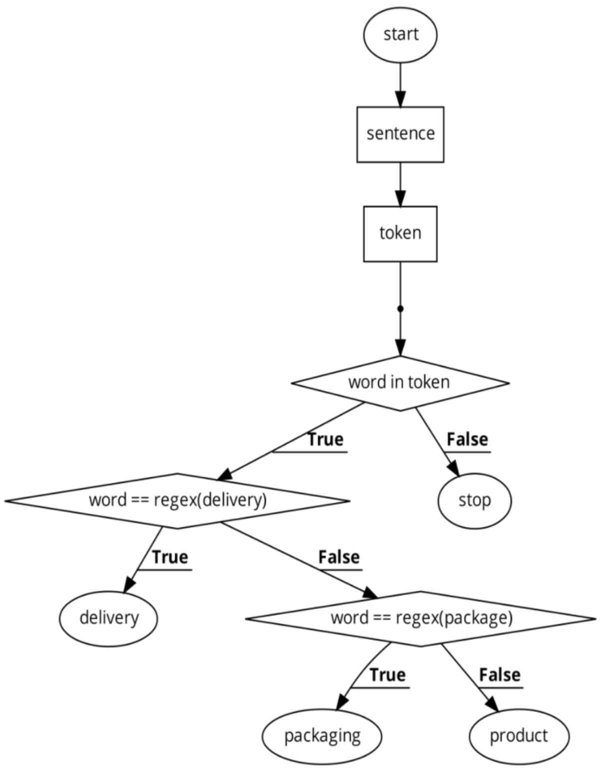

Fig. 4 Flow Chart Representing Review Segmentation into Products, Packaging and Delivery.

F. Classification:

Multi-output learning brings together a variety of learning issues from several academic fields to solve difficult decision-making in a variety of real-world applications. It is multidimensional in nature, and numerous outputs may have intricate relationships that structured inference is intended to manage. Depending on the nature of the ML problem, output values have a variety of kinds of data, for example, following list.

- 0/1-based binary output values can be relevant to multi-label classification problems.
- Nominal output values for multidimensional classification problems
- Ordinal output values for labeling ranking problems. • Real-valued output for multi-target regression problems.

A sort of mechanism teaching called multi-output organisation simultaneously anticipates several outputs. In a multi-output organisation, the model makes a prediction and then produces two or more outputs. In the case of other categorization categories, models often only forecast one result.

A classification assignment involving more than two classes is referred to as multi-class classification. For instance, you may be asked to categorise a collection of fruit photos that might include oranges, apples, or pears. Multiclass classification works under the premise that each sample is given one and only one label. For example, a fruit may only be an apple or a pear at any given time.

Each sample receives a set of target labels through the multi-label categorization process. Predicting characteristics of a data point that are not mutually exclusive, such as subjects that are pertinent for a paper, may be thought of as doing this. A work may simultaneously address religion, politics, finance, or education, or it may not address any of these topics at all.

With multi-output regression, a set of goal values is given to each sample. Several features, such as the direction and amplitude of the wind at a certain place, may be predicted for each data point in this way.

Multiclass-Multioutput Classifiers such as Random Forest (RF), Support Vector Classification (SVC), Decisions Tree Classifier, AdaBoost Classifier are used in this project.

4. Results

Decision Tree Classifier (Multi Output) results are as shown in fig 5:

- Accuracy score of Star Rating is: 0.46
- Accuracy score of Sentiment Class is: 0.77
- Accuracy score of Category Class is: 0.99

```
Column name: star_rating
classification_report:
               precision    recall  f1-score   support

         1.0    0.431655  0.295567  0.350877       609
         2.0    0.395105  0.422430  0.408311      1070
         3.0    0.328612  0.228346  0.269454       508
         4.0    0.451208  0.519755  0.483062      1797
         5.0    0.567037  0.567037  0.567037      1529

    accuracy                        0.462362      5513
   macro avg    0.434723  0.406627  0.415748      5513
weighted avg    0.458987  0.462362  0.457559      5513

Column name: sentiment_class
classification_report:
               precision    recall  f1-score   support

    Negative    0.607495  0.517212  0.558730      1191
     Neutral    0.796089  0.755467  0.775247      1509
    Positive    0.820020  0.894063  0.855442      2813

    accuracy                        0.774714      5513
   macro avg    0.741201  0.722248  0.729806      5513
weighted avg    0.767557  0.774714  0.769391      5513
```

Fig. 5 Decision Tree Classifier (Multi Output) Result.

Random Forest Classifier (Multi Output) results are as shown in fig 6:

- Accuracy score of Star Rating is: 0.46
- Accuracy score of Sentiment Class is: 0.77
- Accuracy score of Category Class is: 0.99

```
Column name: star_rating
classification_report:
               precision    recall  f1-score   support

         1.0    0.658683  0.176565  0.278481       623
         2.0    0.434160  0.441319  0.437710      1031
         3.0    0.430769  0.101449  0.164223       552
         4.0    0.443775  0.614230  0.515272      1799
         5.0    0.577473  0.642573  0.608286      1508

    accuracy                        0.488845      5513
   macro avg    0.508972  0.395227  0.400794      5513
weighted avg    0.501532  0.488845  0.464301      5513

Column name: sentiment_class
classification_report:
               precision    recall  f1-score   support

    Negative    0.865882  0.313725  0.460576      1173
     Neutral    0.789346  0.838046  0.812968      1556
    Positive    0.786088  0.970187  0.868489      2784

    accuracy                        0.793216      5513
   macro avg    0.813772  0.707320  0.714011      5513
weighted avg    0.803986  0.793216  0.766027      5513
```

Fig. 6 Random Forest Classifier (Multi Output) Result.

SVMClassifier (Multi-Output) results are as shown in fig 7:

```
Column name: star_rating
classification_report:
               precision    recall  f1-score   support

         1.0    0.655087  0.463158  0.542652       570
         2.0    0.541298  0.643295  0.587905      1141
         3.0    0.427350  0.103306  0.166389       484
         4.0    0.584836  0.685637  0.631238      1845
         5.0    0.706242  0.706721  0.706481      1473

    accuracy                        0.608380      5513
   macro avg    0.582963  0.520423  0.526933      5513
weighted avg    0.601700  0.608380  0.592404      5513

Column name: sentiment_class
classification_report:
               precision    recall  f1-score   support

    Negative    0.807163  0.745547  0.775132      1179
     Neutral    0.805521  0.826255  0.815756      1554
    Positive    0.940283  0.957194  0.948663      2780

    accuracy                        0.875023      5513
   macro avg    0.850989  0.842999  0.846517      5513
weighted avg    0.873827  0.875023  0.874088      5513

Column name: category_class
classification_report:
               precision    recall  f1-score   support

    delivery    0.995211  0.923556  0.958045      1125
   packaging    0.989064  0.922902  0.954839       882
     product    0.961327  0.999715  0.980145      3506

    accuracy                        0.971885      5513
   macro avg    0.981868  0.948724  0.964343      5513
weighted avg    0.972679  0.971885  0.971587      5513
```

Fig. 7 SVM Classifier (Multi Output) Result.

- Accuracy score of Star Rating is: 0.60
- Accuracy score of Sentiment Class is: 0.87
- Accuracy score of Category Class is: 0.97

ADA BOOST Classifier (Multi Output) results are as shown in fig 8:

```
Column name: star_rating
classification_report:
               precision    recall  f1-score   support

         1.0    0.566667  0.266458  0.362473       638
         2.0    0.456579  0.327668  0.381528      1059
         3.0    0.287489  0.636542  0.396088       509
         4.0    0.445987  0.412158  0.428406      1793
         5.0    0.591971  0.652576  0.620798      1514

    accuracy                        0.465808      5513
   macro avg    0.469738  0.459080  0.437859      5513
weighted avg    0.487444  0.465808  0.461623      5513

Column name: sentiment_class
classification_report:
               precision    recall  f1-score   support

    Negative    0.720270  0.457118  0.559286      1166
     Neutral    0.687564  0.845231  0.758288      1583
    Positive    0.878316  0.898336  0.888213      2764

    accuracy                        0.789770      5513
   macro avg    0.762050  0.733562  0.735263      5513
weighted avg    0.790117  0.789770  0.781339      5513

Column name: category_class
classification_report:
               precision    recall  f1-score   support

    delivery    0.719697  0.997375  0.836084      1143
   packaging    1.000000  0.477140  0.646032       853
     product    0.998580  1.000000  0.999290      3517

    accuracy                        0.918556      5513
   macro avg    0.906092  0.824838  0.827135      5513
weighted avg    0.940980  0.918556  0.910795      5513
```

Fig. 8 ADA BOOST Classifier (Multi Output) Result.

- Accuracy score of Star Rating is: 0.46

- Accuracy score of Sentiment Class is: 0.78
- Accuracy score of Category Class is: 0.91

Overall observing all the models and their performance metrics SVM performed better than all the other models.

5. Conclusion

The main goal of this study was to determine which machine learning algorithm methods performs better in the task of text classification. This was accomplished by using the Amazon reviews as data set. The classifiers were evaluated by comparing their accuracies in different cases of experiments.

The results from the study showed that in terms of accuracy the SVM approach achieves better results than the other approaches when the whole data set was used as training and testing data set.

The data set used for the model evaluation was down sampled which did had its impact on the model accuracy so need to work on complete dataset to achieve high accuracy and precision results.

More Multi Class Multi Output Classifiers can be employed to further increase the study on multi output classifiers and their implementation on amazon reviews dataset.

References

[1] Kowsari, K.; JafariMeimandi, K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text classification algorithms: A survey. Information 2019, 10, 150. [CrossRef]

[2] Rusli, A.; Young, J.C.; Iswari, N.M.S. Identifying fake news in Indonesian via supervised binary text classification. In Proceedings of the 2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), Bali, Indonesia, 7–8 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 86–90.

[3] Al-Rawashdeh, G.; Mamat, R.; Abd Rahim, N.H.B. Hybrid water cycle optimization algorithm with simulated annealing for spam e-mail detection. IEEE Access 2019, 7, 143721–143734. [CrossRef]

[4] Shehnepoor, S.; Salehi, M.; Farahbakhsh, R.; Crespi, N. NetSpam: A network-based spam detection framework for reviews in online social media. IEEE Trans. Inf. Forensics Secur. 2017, 12, 1585–1595. [CrossRef]

[5] Alterkavı, S.; Erbay, H. Novel authorship verification model for social media accounts compromised by a human. Multimed. Tools Appl. 2021, 80, 13575–13591. [CrossRef]

[6] Alterkavı, S.; Erbay, H. Design and Analysis of a Novel Authorship Verification Framework for Hijacked Social Media Accounts Compromised by a Human. Secur. Commun. Netw. 2021, 2021, 8869681. [CrossRef]

[7] Liu, S.; Cheng, X.; Li, F.; Li, F. TASC: Topic-adaptive sentiment classification on dynamic tweets. IEEE Trans. Knowl. Data Eng. 2014, 27, 1696–1709. [CrossRef]

[8] Esposito, F.; Corazza, A.; Cutugno, F. Topic Modelling with Word Embeddings. In Proceedings of the Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016), Napoli, Italy, 5–7 December 2016.

[9] Leeuwenberg, A.; Vela, M.; Dehdari, J.; van Genabith, J. A minimally supervised approach for synonym extraction with word embeddings. Prague Bull. Math. Linguist. 2016, 105, 111. [CrossRef]

[10] PForms. 68 Useful eCommerce Statistics You Must Know in 2022. Available online: https://wpforms.com/ecommercestatistics/ (accessed on 4 August 2022).

[11] Nasdaq. UK Online Shopping and E-Commerce Statistics for 2017. Available online: https://www.nasdaq.com/articles/ukonline-shopping-and-e-commerce-statistics-2017-2017-03-14 (accessed on 4 August 2022).

[12] Zhang, S.; Zhang, D.; Zhong, H.; Wang, G. A multiclassification model of sentiment for E-commerce reviews. IEEE Access 2020, 8, 189513–189526. [CrossRef]

[13] Medhat, W.; Hassan, A.; Korashy, H. Sentiment analysis algorithms and applications: A survey. Ain Shams Eng. J. 2014, 5, 1093–1113. [CrossRef]

[14] Muslim, M.A. Support vector machine (svm) optimization using grid search and unigram to improve e-commerce review accuracy. J. Soft Comput. Explor. 2020, 1, 8–15.

[15] Xu, F.; Pan, Z.; Xia, R. E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. Inf. Process. Manag. 2020, 57, 102221. [CrossRef]

[16] Vanaja, S.; Belwal, M. Aspect-level sentiment analysis on e-commerce data. In Proceedings of the 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 11 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1275–1279.

[17] Jabbar, J.; Urooj, I.; JunSheng, W.; Azeem, N. Real-time sentiment analysis on E-commerce application. In Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), Banff, AB, Canada, 9–11 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 391–396.

[18] Parveen, N.; Santhi, M.; Burra, L.R.; Pellakuri, V.; Pellakuri, H. Women's e-commerce clothing sentiment analysis by probabilistic model LDA using R-SPARK. Mater. Today Proc. 2021, in press. [CrossRef]

[19] Tripathi, P.; Singh, S.; Chhajer, P.; Trivedi, M.C.; Singh, V.K. Analysis and prediction of extent of helpfulness of reviews on E-commerce websites. Mater. Today Proc. 2020, 33, 4520–4525. [CrossRef]

[20] Kumar, K.S.; Desai, J.; Majumdar, J. Opinion mining and sentiment analysis on online customer review. In Proceedings of the 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, India, 15–17 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–4.

[21] Miyoshi, T.; Nakagami, Y. Sentiment classification of customer reviews on electric products. In Proceedings of the 2007 IEEE International Conference on Systems, Man and Cybernetics, Banff, AB, Canada, 5–8 October 2017; IEEE: Piscataway, NJ, USA, 2007; pp. 2028–2033.

[22] Guan, Z.; Chen, L.; Zhao, W.; Zheng, Y.; Tan, S.; Cai, D. Weakly-Supervised Deep Learning for Customer Review Sentiment Classification. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), New York, NY, USA, 9–15 July 2016; pp. 3719–3725.

[23] Shoja, B.M.; Tabrizi, N. Customer reviews analysis with deep neural networks for e-commerce recommender systems. IEEE Access 2019, 7, 119121–119130. [CrossRef]

[24] Gu, T.; Xu, G.; Luo, J. Sentiment analysis via deep multichannel neural networks with variational information bottleneck. IEEE Access 2020, 8, 121014–121021. [CrossRef]

[25] B˙Ilen, B.; Horasan, F. LSTM network based sentiment analysis for customer reviews. Politek. Derg. 2021. [CrossRef]

[26] Vural, A.G.; Cambazoglu, B.B.; Senkul, P.; Tokgoz, Z.O. A framework for sentiment analysis in turkish: Application to polarity detection of movie reviews in turkish. In Computer and Information Sciences III; Springer: Berlin/Heidelberg, Germany, 2013; pp. 437–445.

[27] Acikalin, U.U.; Bardak, B.; Kutlu, M. Turkish sentiment analysis using bert. In Proceedings of the 2020 28th Signal Processing and Communications Applications Conference (SIU), Gaziantep, Turkey, 5–7 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–4.

[28] Santur, Y. Sentiment analysis based on gated recurrent unit. In Proceedings of the 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey, 21–22 September 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5.

[29] Ozyurt, B.; Akcayol, M.A. A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. Expert Syst. Appl. 2021, 168, 114231. [CrossRef]

[30] Kadhim, A.I. Term weighting for feature extraction on Twitter: A comparison between BM25 and TF-IDF. In Proceedings of the 2019 International Conference on Advanced Science and Engineering (ICOASE), Zakho, Iraq, 2–4 April 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 124–128.