ZKG INTERNATIONAL
ISSN: 2366-1313

SENTIMENT ANALYSIS ON AMAZON PRODUCT REVIEWS USING MACHINE LEARNING MODELS

[1]KUNCHAM RAMYA, [2]PASUPULETI VENKATA HARISH, [3]VETURU HARI KRISHNA, [4]KOTTUM CHANDAN, [5]KURUVA PAVANKALYAN

[1]Assistant Professor, Dept. Of AI, ABR College of Engineering and Technology, Kanigiri

[2, 3, 4, 5] BTech Student, Dept. Of AI, ABR College of Engineering and Technology, Kanigiri

Abstract: *Online reviews have become an essential source of statistics for customers before making an informed purchasing decision. Early product reviews tend to have a disproportionate effect on subsequent product earnings. This paper takes the initiative to look at the first reviewers' behavioural traits with their published reviews on the real and international large e-commerce structures, namely Amazon, Flipkart, Myntra. They provide their feedback and then products are recommended for purchase and sale on that factor too. This Research is focusing to apply sentiment analysis to review the product of Amazon. In this paper, we use machine learning methods to perform this experiment. Sentiment Analysis is research in which consumers are conscious of a product reaction. A Kaggle of amazon product reviews gathers the data collection used for the experiment of the proposed work. Data set of review of customer has been considered in order to perform sentiment analysis. The proposed research is supposed to resolve the issues of previous research that were faced during sentiment analysis.*

Keywords: *Sentiment analysis, amazon product reviews, machine learning, Random forest algorithm.*

I. INTRODUCTION

In product reviews, sentiment analysis is becoming popular for textual mining. Research is also considering studies around computational linguistics. The research focuses on the relationship between Amazon's product feedbacks. The investigation is also considering the classification of products provided by consumers. The research considers traditional machine study algorithms such as Naive Bayes, SVM, KNN. Research has also looked at deep neural networks with recurrent neural networks (RNNs). The research aims to offer a better solution for diagnosing emotions [1].

In addition, problems, individuals, events, and concepts are considered along with their characteristics. The utility uses this type of concept, which is diverse. In the afterlife, any company should consider consumer opinion. Consumer sentiment is taken into account when designing products and offerings. Before using a program or buying a product, qualified consumers do not forget the thoughts and feelings of contemporary consumers. In addition, the researcher [2] uses this data for an in-deep analysis of industry dynamics and customer preferences. This form of the revision can lead to real predictions in the inventory market..

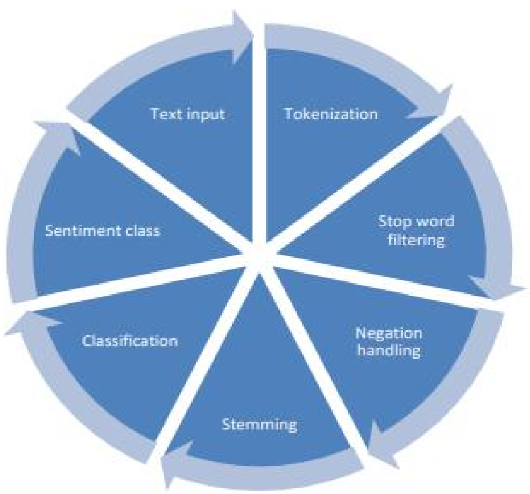


Fig.1 sentiment analysis approach for prediction

Despite this, finding and monitoring opinions is still beyond the average person's reach. Each site usually contains much-opinionated text in long posts on discussion forums and blogs, which is not always easy to understand.

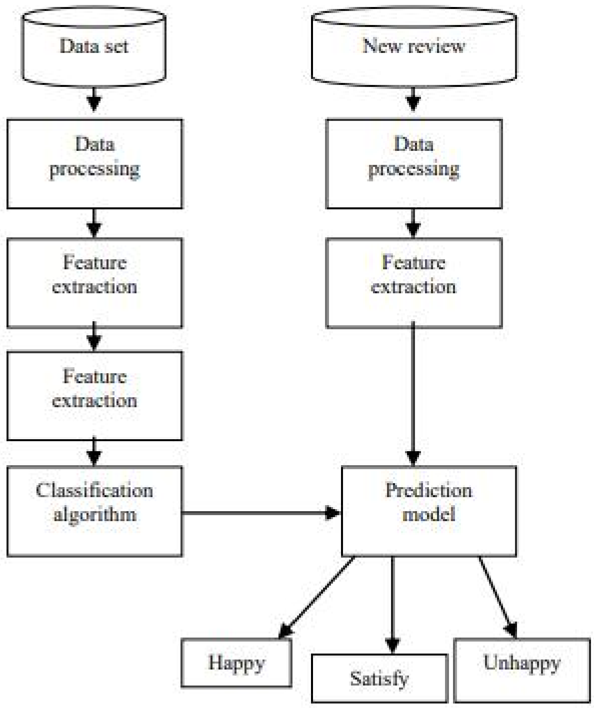


Fig.2 Prediction model for sentiment analysis

Amazon is undoubtedly one of the largest e-commerce companies distributed worldwide and offers a wide range of products such as books, medicines, fitness equipment, and many more. Our system will analyze Amazon product reviews, categorize textual content into comparative training, and extract consumer sentiments or emotions. For example, if a buyer writes a review, the product is amazing. Our gadget classifies it as positive. The main reason for this plan is to rate buyers'

ISSN: 2366-1313

reviews and get their emotions out. There are many types of algorithms in which knowledge is acquired. Our conscience is in the Naive Bayes Classifier and Support Vector Machine (SVM). To see which rank plays best in Amazon's overview (shared merchandise)[3].

II. REVIEW OF LITERATURE

In this part, we surveyed the latest works related to our curriculum. First, much research about e-commerce has done to improve the advertising focus and revenue margin of many organizations.

Ting Bai, et al.(2018) As long as the early auditor tends to assign a better diploma to a combined degree and (2) the early auditor tends to provide more helpful opinions. The product review score also indicates that the first reviewers' ratings and the assistance ratings obtained are likely to lead to the product's reputation. Given the public deployment process as a multiplayer opposition sport, we advocate for a full fringe inclusion model to predict an early baseline. The delight of one particular e-commerce dataset showed that our proposed devices outgrew a robust set of baselines [4].

Salganik et al. (2016) Successful songs, books, and movies tend to be more successful than average, indicating that "cool" alternatives are qualitatively great over "the rest,"; But specialists cannot wait for the products to be automated. We experimentally tried this paradox by developing a synthetic "ringtone market," where 14,341 people downloaded previously unknown songs with or without knowledge of alternatives to previous contributors. The rise of social electricity affects both high tolerances and unpredictability of compliance. Success becomes more beneficial. Moreover, it is determined in part exceptionally: the performance of great songs was in no way inferior, and the worst performance rarely happened correctly, But no other result is possible [5].

D. Imamori et al. (2015) we advocate a scoring technology that is currently creating debt on Twitter tailored to its potential for popularity. Early detection of last owed money helps trait forecasting, viral marketing, recommending people, etc. However, it is tough to evaluate new invoices because they have not yet received the recognition they deserve. We cannot monitor existing account valuation strategies based primarily on various hyperlinks or reputation. Our method detects first customers. First, that is customers who regularly discover new, desirable statistical resources before others.

ISSN: 2366-1313

Their system has found that the new invoices noticed with these super first users' help are promising, even though they don't have that many fans. To discover well first users, we respect the frequency spread of the link for each account, i.e., the diversity of cases where fans have copied the account view links. If the price is exorbitant, the archive should be one of the first customers to often find the right data sources before pursuing it. We are expanding a method for deducing links that can be created with the help of copy hyperlinks [6].

N. V. Nielsen et al. (2014) the most popular, no longer top-notch, categories of e-commerce are durables and laughter-related goods. Nearly half of international respondents to an online survey intend to buy clothes or make airline or hostel reservations using the internet over the next six months. On the other hand, the online market for grocery and other consumables is incredibly smaller. The purchase of weapons by their very nature and the enormous potential of that property limit the utility and purchasing method online. However, global target audiences tend to own the internet. Online purchase intent fees doubled in 3 years for 12 of the 22 categories measured. While consumer items will nonetheless lag behind non-consumables, the frequency of purchasing these products increases the appeal of e-commerce. Besides shopping, digital is increasingly becoming an essential search and sharing platform [7].

Maria Soledad Elli [8] did Sentiments from thinking about buyer reviews. They have analysed the results to enhance the model for a commercial enterprise. The author provided that the device provides better accuracy. Research has employed a wide variety of dialects. It acts as a classification. The mechanism is also helping the vector machine.

Callen Rain [9] has done extensive research in the field of herbal language processing. The research uses Naive Bayesian in addition to those who list the options. These mechanisms have been applied to the classification of proportional assessment. This diagnosis will be too big or too poor. The research uses deep learning neural networks. It has been noted that the neural community is well-known in sentiment diagnosis.

Ronan Collobert et al [10] has used convolutional networks. Semantic position tagging is designed to avoid excessive amounts of operational feature engineering.

III. DATASET AND FEATURES

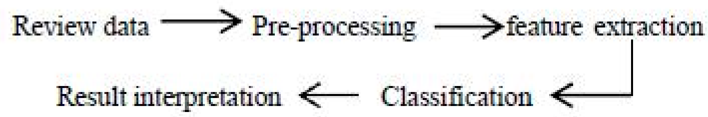
a) Dataset collection

ISSN: 2366-1313

Amazon is one of the best e-commerce websites where people buy the product they need and then rate it as 1 to 5 rated comments or techniques; there are many more ways of scaling. Kaggle datasets can be used to archive buyer reviews for Amazon products. The dataset contains product ID, classification, diagnostic text, and basic product information. The datasets are in CSV format. Typically, different sets of Amazon product review data are used to keep checks and balances on algorithms that are being implemented or finalized.

b) Accession

The access to input has been replaced with a forward positioning structure. Research data set gathered from Kaggle Amazon product reviews is collected from an e-commerce website. Each file is within the Comma Separated Values (CSV) registry setting. In the next step, the data is pre-handled, we remove the invalid values, and the data set is found to be very biased towards fine-tuning using the log viewing approach. Some more neutral and bad feedback is added to consolidate the dataset records. The previous phrases can be separated, and the directives, spaces, statistics, and special tokens are pre-processed in the next step. In case of deficiency, the content of the diagnostic test is changed, the previous words are deleted, and it is taken out. Feature extraction is performed in the next step via TF-IDF vectorization. After that, the dataset is divided into training and experimental statistics for teaching versions. Eventually, his fairs labeled the facts of preparation and experiments using logistic regression, navy bases, and random forest classifiers. We obtain accuracy and classification reports using exceptional algorithms.



Framework for Sentiment Analysis

IV. METHODOLOGY

In this paper, we took the initiative to study the first reviewers' behavioural characteristics through their comments posted on representative e-commerce platforms, for example, Amazon. Our goal is to make a useful analysis and make accurate forecasts with the first reviewers.

A. DATA PRE-PROCESSING

The data is already collected from the diagnostic data created by the user. This assessment is completed on products that can be provided with the help of Amazon. About thirty-five thousand points of information have been added through such facts. Each example includes type, product name, text review, and product classification. To better use the data, we first draw out two of the most important columns related to this work. These columns are classification and evaluation. Then, in examining the facts, we see the points of the facts that remain unqualified. When we remove this pattern, we usually have 34,627 data points left.

On the other hand, we have developed a classification assigned to define information. There have been five training sessions. A rating of 1 to 5 is given between these classes. Such instructions are unequal.

B. Feature Extraction

TF-IDF: An analytical metric that measures the period for a record or document. It is calculated by the matrix: the general appearance of a word and the inverse frequency of the word file in the whole file. Each phrase or expression has its TF and IDF classification. This can be improved. Therefore, the scale of TF * IDF is that which does not say the word and the sphere the other way.

The TF of a time period or expression is the number of times in a text that a time period or expression appears. A one-time IDF is an expensive process within a corpus of that name. It will always be included in the main search effects when the phrase expression/call / time period/commitment has a massive TF * IDF scale, so everyone can:

1. Avoid thinking about stop-words being used,
2. Find terms that are searched for the maximum number of times and lower competition successfully.

It is extracted function that is found from pre-processing from the cleaned text

C. CLASSIFICATION

We use the standard decision tree classifier for training review $D = \{X^n, Y^n\}$ where $X^n = \{x_1, x_2, x_3, \ldots\ldots . x_n\}$ refers to the features of n samples and $Y^n = \in \{0,1\}$ we use gini impurity as a standard function that measures department quality during the training process. The number of features to consider in order getting the best split is set like all the available features. Additionally, we set the minimum number of samples on nodes to 1. The minimum number of samples required to divide an internal node is set to 2.

Algorithm: Learning User embeddings

Input: $training\ instances\ \mathcal{T} = \{u \succ_p u' | u, u' \in \mathcal{U}\}$,

$products\ embeddings\ set\{v_p\}$,

$Learning\ rate\ \lambda$,

$Margin\ coefficient\ m$,

$embedding\ dimensions\ L$.

ISSN: 2366-1313

$\boldsymbol{Output}$: $user\ embedding\ \{v_u | \forall_u \in \mathcal{U}\}$

$\boldsymbol{Procedure}$:

1. $\boldsymbol{initialize}\ user\ embeddings$:
2. $v_u \leftarrow uniform\left(-\frac{6}{\sqrt{L}}, \frac{6}{\sqrt{L}}\right), \forall_u \in \mathcal{U}$
3. $v_u \leftarrow v_u / ||v_u||, \forall_u \in \mathcal{U}$
4. $\boldsymbol{loop}$
5. $sample\ a\ training\ instance\ \langle u \succ_p u' \rangle \in \mathcal{T}\ \boldsymbol{do}$
6. $update\ user\ embeddings$:

D. RANDOM FOREST ALGORITHM

Having supremacy on the top of a single decision tree concerning reliability and efficiency, the random forest classifier was chosen. It is a method of ensemble which is based on bulging. The categorizer works in this way: (as shown in figure 3) given D, the disposer first generates k bootstrap D specimens, with Di denoting each of the specimens. A Di has almost the same proportion of D rows that are selected with D-substitution. It demonstrates that a few authentic D rows may not be encompassed in Di, whereas other tuples may occur more than once, by sampling with replacement. Depending on each Di, the classifier will then create a decision tree. Consequently, a "forest" is generated consisting of k decision trees. Each tree returns its genre prognosis enumerating as a single ballot to identify an unknown tuple, X.

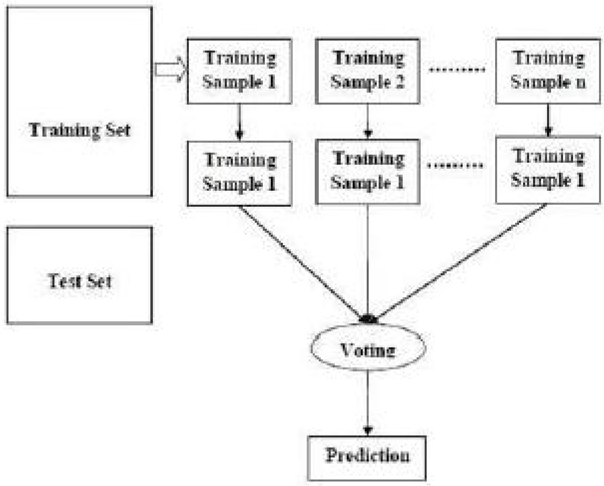


Fig.3 work flow of Random Forest algorithm

V. EVALUATION METRICS

In assessing the classification efficiency of a model, evaluating metrics plays an important role. The most widely used technique for this purpose is precision measurement. The efficiency of a sequencer on a given experimenting dataset is the proportion of those datasets that are properly grouped by the sequencer, and the accuracy measure of the text mining method is often not sufficient to provide proper decision or result, so some others metrics should be taken to assess the output of the classifier. There are

Accuracy: Accuracy is the ratio of the total number of instances of the correct prediction. Accuracy calculated as follows

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

ISSN: 2366-1313

Sensitivity: Sensitivity is used to determine the portion of the actual positive instances case classified adequately by the classifier. Sensitivity calculated as follows

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity: Specificity is used to know the ability of classifiers to identify incorrectly classified negative cases

$$Specificity = \frac{TN}{TN + FP}$$

Precision: Precision is an indicator that defines the true portion of the instances when predicted to be true. Precision calculated as follows-

$$Precision = \frac{TP}{TP + FP}$$

F1 Score: F1 Score is a harmonic mean of recall and precision. It must be one for good performance and zero for the bad performance of the classification algorithm. F1 score calculated as follows

Where, TP- True Positive, FP- False Positive, FN- False Negative, TN-True negative

VI. RESULTS AND DISCUSSION

After experimenting and priming the model with the dataset, their expected efficiency is decided and compared to expose which methodology is best categorizing reviews. As shown using table.3, The Random Forest model undergoes the finest predictive validity amidst the two models and Naïve Bayes has the inferior anticipating precision

Table.1 Classification report for Naïve Bayes

Sentiment score	Precision	Recall	F1 score
0	0.61	0.48	0.54
1	0.37	0.23	0.28
2	0.91	0.95	0.93

Table.2 Classification report for Random Forest

Sentiment score	Precision	Recall	F1 score
0	0.89	0.68	0.77
1	0.90	0.46	0.60
2	0.94	1.00	0.96

Table.3 Measured Accuracy

Method	Accuracy
Naïve Bayes	86.46
Random Forest	93.17

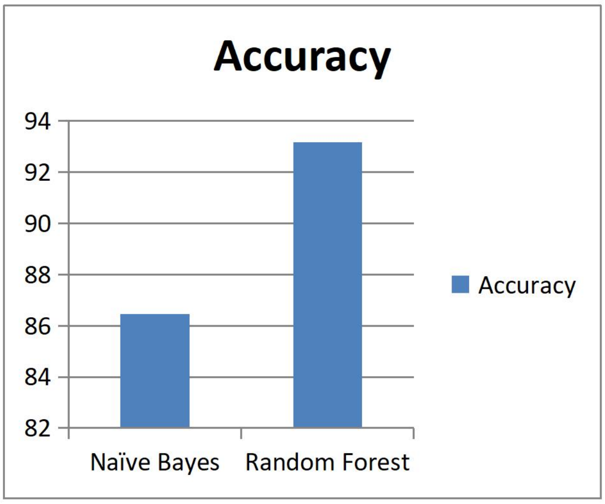


Fig.4 Accuracy comparison between two models

The conclusion among the two algorithms can be shown above in fig.4. It concludes that random forest to be more efficient. For the sentiment detection and analysis, a lot of high mathematical computations and algorithms have been found which can optimize the data more accurately as far as possible.

VII. CONCLUSION

A radical change from virtual platforms to digitalized platforms can be seen in a new age. The dependence of clients and consumers on online feedback has increased especially. Digital opinions have enhanced a forum for raising belief and shaping the trends of customer purchasing. By performing an opinion analysis of Amazon product checks and categorizing the opinions into optimistic, neutral, and negative feelings, our project aims to accomplish this. Four classification models were used to identify reviews after combining the data with some neutral and negative opinions. Out of these classifiers, Random Forest, with 93.17% accuracy, the predictive accuracy of Random Forest is found to be the highest.

REFERENCES

C. Yue and Y. Dong, 2020, "Collaborative attention neural network for multi-domain sentiment classification,".

[2] G. More, 2020, "Sentiment Analysis on Amazon Product Reviews with Stacked Neural Networks,", doi: 10.13140/RG.2.2.14746.67524.

[3] S. Thapa, and A. Aditya, 2020, "Feature Selection Based Twin-Support Vector Machine for the diagnosis of Parkinson's Disease", IEEE.

[4] S. Kundu and S. Chakraborti, 2020, "A comparative study of online consumer reviews of Apple iPhone across Amazon, Twitter and MouthShut platforms,", doi: 10.1007/s10660-020-09429-

[5] P. Pankaj, and N. Soni, 2019, "Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews,", pp. 320–322,

ISSN: 2366-1313

[6] S. Adhikari, and B. K. Shah, 2020, "Oversampling based Classifiers for Categorization of Radar Returns from the Ionosphere," IEEE, pp. 975-978.

[7] S. Hota and S. Pathak, 2018, "Knn classifier based approach for multi-class sentiment analysis of twitter data". pages 1372–1375.

[8] B. Liu and L. Zhang, "A Survey of Opinion Mining and Sentiment Analysis", pages 415–463.

[9] W. Zhao *et al.*, 2018, "Weakly-supervised deep embedding for product review sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 1, pp. 185–197.

[10] B. K. Shah, and A. Shroff, 2020, "Evaluation and Comparative Study of Edge Detection Techniques," vol. 22, no. 5, pp. 6–15.

[11] Prasadu Peddi (2015) "A machine learning method intended to predict a student's academic achievement", ISSN: 2366-1313, Vol 1, issue 2, pp:23-37.