ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Review Sentiment Analysis for Amazon Products

[1] Dr. B. Sai Venkata Krishna, [2]L. Ruchitha,

[1]Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.
[2] MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Abstract—

Since the majority of people now utilize things purchased online, we thought it would be interesting to assess how people feel about these products in this technologically advanced era via evaluations. They provide their opinions, and based on that, things are suggested for sale or purchase. Customers are allowed to leave product reviews on several of the major e-commerce sites, including Amazon, Flipkart, Myntra, and many more. Before purchasing a product, consumers will do thorough research to have a better grasp of the product and how it functions. The product in question will be really basic and will be categorized as either positive, neutral, or negative according to the interpretation. To conduct this experiment, we may make use of machine learning techniques. Research using sentiment analysis involves customers who are aware of how a product makes them feel. We gathered this data from a Kaggle search of Amazon product reviews. To achieve the highest level of accuracy in feedback classification, we use a variety of methods including Logistic Regression, Naive Bayes, and Random Forest. We found the Random forest machine-learning algorithm to be the most accurate out of all the methods tested. emotional state. This field of research has grown in popularity in the modern era of the internet. However, there are a few downsides to this massive amounts of input. To start, not a single one of these reviews has ever promised or even implied that the product is authentic. This is due to the fact that imposter users may also provide false views and remarks. Secondly, there is a lack of accessibility to reviewers and internet reviews in general. The demise of these retail portals is sometimes caused by this. Use sentiment analysis to evaluate the product's structure, learn what consumers like and don't like, see how our products stack up against the competition, get a feel for what's new in the market, and save a bunch of time compared to manual conversion. To help users quickly decide whether to purchase a product, the primary goal is to categorize evaluations as either favorable or negative. Still, a number

Keywords—Machine learning, Classification, Linear regression, Naïve Bayes, Random Forest.

INTRODUCTION

Everything is moving online these days due to the prevalence of technology and digitalization. Online shopping has replaced traditional methods of purchasing food, clothing, and household goods, as well as technology devices. As a result, e-commerce platforms have grown significantly. On these sites, you may find a variety of items from various manufacturers. It will be quite challenging to choose a dependable and practical product because of this. To acquire a product that will be of benefit to them, consumers read evaluations to get a feel for the product, learn more about it, and then make a buying decision. One of the first things a customer does when purchasing anything online is look at customer reviews. The opinions and experiences of other users are more reliable. Reviews are the sole determinant of whether a customer continues with a purchase or returns it. The significance of reviews may be shown thus [8]. However, it's not practical to go through hundreds of evaluations every time someone considers purchasing a product. Since this is the case, it is wise to glean some relevant information from these evaluations. Machine learning is useful in this situation. The advent of AI and machine learning has changed the game completely [16–17]. Machine learning's applications have been extensively studied in domains such as healthcare analytics, business, sentiment analysis, and others [18–19]. To learn how people feel or what they think about a product or service, one may use a computer approach called sentiment analysis [14]. In theory, it's a rating system that highlights the fact that each review conveys a neutral, negative, or favorable attitude. In this article, we will talk about a few different ways.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

RELATED WORK

Citing a 2002 study, Pang, Lee, and Vaithyanathan proposed classifying film reviews by emotion using ML approaches [5]. The researchers examined the use of the Naïve Bayes, Max Entropy, and Help Vector Machine models for sentiment analysis on data bigrams and monographs. According to their experiment, the most effective combination of SVM and unigram function extraction produced the greatest results. An accuracy rate of 82.9% was obtained. In their 2004 article, Mulle and Collier finalized the sentiment classification of jewelry and footwear product criticism [4]. The hybrid approaches of Support vector machine, Naïve Bayes, logistic regression, and decision trees were compared with the feature extraction methods based on Lemmas and Osgood theory. With an accuracy rate of 86.6%, the support vector machine outperforms the competition. In their 2017 paper, Elmurngi and Gherbi suggested a method to identify fabricated movie reviews. A comparison was made between Naive Bayes and SMV's, decision books, and knit performance on two datasets: one with stop words and one without [11]. SVM comes out on top with a precision of 81.75% and 81.35% in the two situations, respectively. In 2018, Bijoyan Das and Sarit Chakraborty did an experiment using SVM, TF-IDF, and the Next Word Negation in conjunction with an Amazon Product Opinion dataset. They achieved an accuracy rate of 89% [10]. Comparative research on sentiment analysis, morphological-based methods, and machine learning in film reviews has been conducted by

The 2018 edition included Bhavitha, Rodrigues, and Chiplunkar. For the SentiWordNet approach, they have achieved 74% production, and for the SVM method, they have achieved 86.40%. In the 2018 IEEE paper, a supervised learning approach was suggested for the purpose of polarizing several untagged product opinion datasets, such as Tanjim, nudrat, and Faisal. It combines two types of extractor methods and is a monitored learning approach [12]. A recovery rate of 90%, together with F1 measurement precision, allowed them to attain an accuracy level over 90% [8]. In a 2017 paper, Ceenia Singla, Sukhchandan Randhawa, and Sushma Jain conducted an emotional analysis of reviews for mobile phones, categorizing them as either good or negative. None other than Decision Tree, Naïve Bayes, and Support Vector Machine were used as classifiers. According to research, SVM has the highest prediction accuracy at 81% [13]. In this study, we will examine several approaches of using Sentiment Analysis and evaluate their accuracy. Methods for Emotional Intelligence In sentiment analysis, two primary methods are used. It combines a word-book method with a machine learning one. For text categorization and lexicon-based approaches, machine learning algorithms depend on standard symbolic methods. Methods and tactics for learning from text, such as case-based learning, root-based learning, and analogy, allow for its classification. The two main schools of thought in machine learning, supervised and unsupervised, may be broadly classified. Part A: A Machine Learning Strategy Machine learning is seen as a crucial branch of AI that operates by training a machine to understand via the use of code. The method employs morphological and expressive characteristics. We have various documentations for priming and categorization, and it studies sentiment analysis as a problem of periodic text classification. The prototype is told to predict the grade for the most recent case. Decision trees, neural tree networks, Naïve Bayes, logistic regression, and Support Vector Machine are among the classifiers that are often used. Both supervised and unsupervised learning are used to create these classifiers. (1) Method of Supervised Learning Among the many machine learning methods that make use of a labeled dataset, supervised learning stands out. Along with the anticipated results, these coaching records also include some input data. Next, new instances are categorized using classifiers that are based on machine learning. Numerous approaches have been developed and recorded; this section describes a few of them. 2) A technique for unsupervised learning When compared to unsupervised learning, supervised learning is more practical, but it requires a collection of labeled training data that isn't always there. No labelled data is necessary for these learning algorithms. A tiny fraction of the data is labeled, while a big portion of the learning data is untagged, in low-supervised learning. On the other hand, input training devices in non-monitored learning are not receiving any communication on expected yield values. Collection analysis and expectation-maximization algorithms are a few instances of unsupervised learning. B. Method based on a dictionary One key component of the Lexicon-Based method is the ability to identify bias. A large number of recognized and precompiled words representing viewpoints make up a lexicon [1]. This perspective lexicon may be used for textual analysis. A manual approach to opinion, a corpus-based technique, and a dictionary-based method are the three main components of the lexicon-based methodology [2]. This works best when combined with the other two techniques.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Approach	Methodologies used
Machine Learning	1)Support Vector Machine (SVM) 2)Decision Trees 3)Neural Networks 4)Naïve Bayes 5)Logistic Regression
Wordbook Approach (lexicon)	1)Dictionary-based approach 2)Manual opinion approach 3) Corpus-based approach

Table.1 Methodologies used

FRAMEWORK

A. Features and dataset A great example of an e-commerce website is Amazon, where customers may purchase the things they need and then provide feedback in the form of comments or ratings (from 1 to 5). Using datasets from Kaggle, which include a record of customer reviews for Amazon purchases, is one option. There is basic product information, a rating, and review content in the collection. Datasets are saved as a CSV file. The algorithms that are applied or run are typically checked and balanced using two additional datasets of Amazon product reviews. Part B. Acquire The transmitted substructure follows the access. First, reviews of products sold on an online store's Kaggle page are used to compile the exploratory data [6]. The CSV file format specifies that each entry should be presented in a separate column. Data is pre-treated in the next phase; we eliminate After removing the null values, a data visualization approach revealed that the dataset was heavily skewed towards positive reviews. To rectify this, more neutral and negative reviews were added to the dataset. The next phase involves pre-processing records in order to separate specific tokens, whitespace, diacritical markings, figures, and stop words. Review material is stemmed, transformed, and stop words are eliminated in lower case. Following this, TF-IDF vectorization is used to extract features. Priming and experimental data are then extracted from the dataset in order to train the model. Using Logistic Regression, Naïve Bayes, and Random Forest Classifier, the priming and experimental data is finally categorized. Various algorithms provide us with reports on accuracy and classification. Section V. Methods Section A: Data Gathering and Visualization You may find customer feedback on Amazon goods in the Amazon product reviews database. We extract four key characteristics—ID, review text, review rating—from the dataset, and these features are crucial for further processing. After visualizing the data, it becomes clear that there is a significant bias towards favorable ratings. To counteract this, neutral and negative evaluations were included in the dataset. To determine which terms are most often used in the review text, exploratory data analysis is used [15].

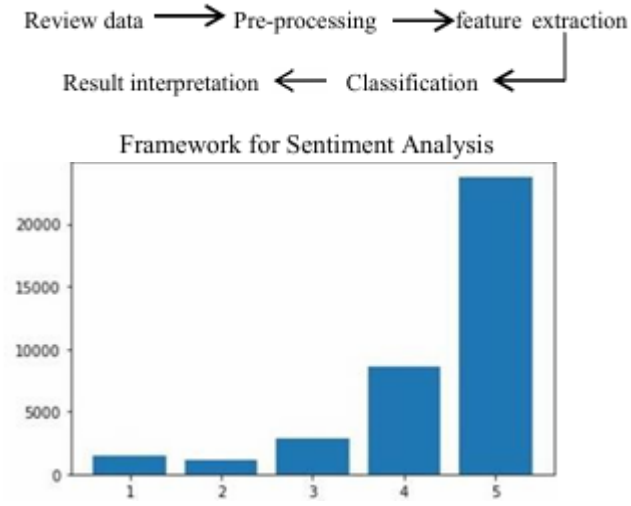

fig.1 Number of Reviews Vs Rating Visualization.

B. Initial Steps The process of tokenization involves breaking a string of words into its component parts, such as names, keywords, expressions, and tokens. Tokens might be anything from single words to whole sentences. Through the tokenization process, some characters are discarded, including vowels, punctuation marks, and countless others. As input values, tokens are used for separate courses of action like text mining and parsing. Change the case of a word from uppercase to lowercase (TREE to tree). To make everything more consistent and point in the same direction, we've lowercased every word. Terminology for Cleaning: "Stop words" are expressive keywords that have no business being part of text analytics. Therefore, such words are often disregarded in order to improve the study's efficiency. Stop words come in a wide variety of forms, each reflecting the unique culture, language, and history of a particular nation. We need to remove several stop words from the English format. Normalizing a term by tracing it back to its origin is known as stemming. The process of assigning sentiment scores involves ranking several items. Reviews with ratings of 1 or 2

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

get a score of 0, reviews with ratings of 3 get a score of 1, and reviews with ratings of 4 or 5 get a score of 2. Zero points: pessimism Level 1: indifferent attitude Level 2: optimistic outlook C. Extracting Features The term frequency-IDF (TF-IDF) is a statistical measure for determining a word's relative importance in a given record. Two metrics are used to calculate this: the number of times a word occurs in the document and the inverse document frequency of word. The TF and IDF scores are unique for each word and phrase. It may be fine-tuned so that the TF*IDF measure is greatest for very improbable words and minimum for very likely words [9]. The term frequency (TF) is the frequency with which a phrase or expression occurs in a text. What happens to the value in the corpus that bears the name of a word is called its IDF. Searches including expression/name/term/figure words with a high TF*IDF metric will always include it, allowing anybody to: 1. Forget about stop-words. 2. Locate phrases that are searched the most and effectively reduce competition. This function is retrieved from the cleaned text during the preparation phase. D. Sorting by Type Opinions may be categorized into three groups according to their emotional intensity: positive, neutral, and negative. While 20% is utilized for experiments, 80% is used for priming the model. 1) Identifying A number of classifiers were engaged. The Random Forest classifier, Naive Bayes, and Logistic regression are used. Calculated Relapse predicts the likelihood of a categorical dependent variable in logistic regression. A parallel variable that is coded as yes or no is included into the subordinate. Comparing the huge test estimate with calculated relapse yields better results. Figure 2 shows that the computed work may be a sigmoid function, which accepts any real input x and returns a value between 0 and 1.

$$\square(t) = \frac{1}{1 + e^{-t}}$$

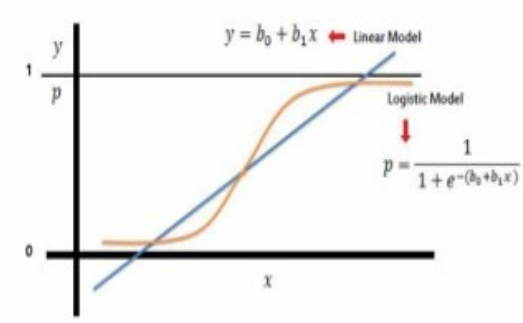

fig.2 Logistic Regression Graphical Representation Naïve Bayes Method:

According to [3], Naïve Bayes is the probabilistic classification approach. Everything is subject to the Bayes theorem. Two distinct versions of naïve Bayes are available for text. Both Bernoulli naïve Bayes and multi-nomial naïve Bayes are used in this context. Each feature value is counted in the Multinomial naïve Bayes technique, when data simply follows a multinomial distribution. Each feature is binary and the data is distributed according to a multivariate distribution in Bernoulli naïve Bayes. The Bayes rule determines the conditional probability of occurrence X because evidence Y is computed by the Finding feeling. You may write it as: P(X/Y) = [P(X) P(Y/X)] / P(Y) Random forest method: The random forest classifier was selected because it was the most efficient and reliable option among the single decision tree [7]. It is an ensemble technique that relies on bulging. Here is how the classifier operates: (as seen in image 3) The disposer begins by creating k bootstrap D specimens from the provided D, where Di represents each specimen. Using D-substitution, almost all of a Di's rows are chosen. It shows that certain real D rows may not be in Di and that some tuples could appear several times when sampling with replacement. The classifier will then construct a decision tree based on each Di. This results in the creation of a "forest" with k decision trees. In order to find a tuple X, the genre prediction of each tree is returned as a single vote.

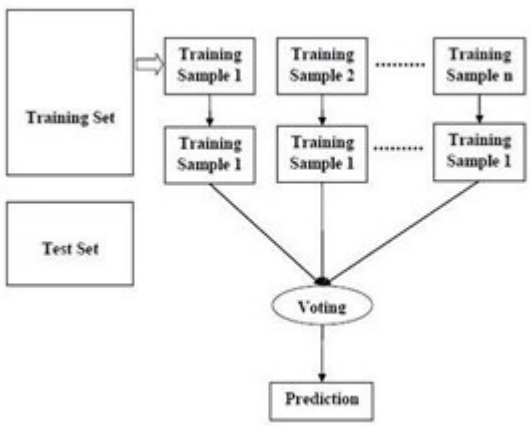

fig.3 Working of random forest

EVALUATING METRICS

Metric evaluation is crucial for determining a model's categorization efficiency. In this regard, precise measuring is by far the most used method. Since the accuracy metric of the text mining method is

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

frequently insufficient to deliver appropriate decisions or results, additional metrics should be used to evaluate the classifier's performance. A sequencer's efficiency on a particular experimental dataset is the proportion of datasets that are correctly grouped by the sequencer. Memory, accuracy, and F-measurement are three additional important and often utilized markers [20]. Before we can talk about the different processes, there are a few terms you should learn. True Positive, abbreviated TP, is the percentage of correctly anticipated positive occurrences in a given sample. The fraction of positive cases that were incorrectly categorized is known as FP (False Positive). FN, short for "False Negative," is a statistical measure of the percentage of false negatives. The True Negative (TN) statistic is applied to the percentage of correctly categorized negative cases in the sample.

Accuracy: The classifier's efficiency is measured by the number of correct return records. Low false positives are an indicator of extreme accuracy, while large false positives are an indicator of near-perfect accuracy. It (P) is the fraction of positively classed occurrences that were correctly identified out of all the positive instances that were positively identified. Another way to look at it is as follows: TP divided by TP plus FP equals P. Memory: It quantifies the sensitivity of a classifier, or the amount of optimistic data it produces. Recalling too much leads to fewer false negatives. The recall is the percentage of occurrences that were properly categorized relative to the total predicted cases. R = TP / TP+FN is one way to express this. An F-measure, which is the calculated harmonic mean of accuracy and recall, is one of the single metrics that may be obtained by combining precision and recall. F = 2P*R / P+R is one possible way to explain it. Efficiency or accuracy measures the degree to which the classifier makes the correct prediction. The percentage of correct forecasts as a percentage of all guesses is called accuracy. Correct prediction divided by total data points is the accuracy.

Sentiment score	precision	recall	F1 score
0	0.79	0.63	0.70
1	0.62	0.35	0.45
2	0.93	0.98	0.96

Table.2 Classification report of Logistic regression

Sentiment score	precision	recall	F1 score
0	0.81	0.25	0.38
1	0.54	0.01	0.02
2	0.87	1.00	0.93

Table.3 Classification report of Multinomial Naive 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.4 Classification report of Bernoulli Naive Bayes

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.5 Classification report of Random forest

Method	Accuracy
Logistic regression	90.88
Multinomial Naive Bayes	87.09
Bernoulli Naive Bayes	86.46
Random Forest	93.17

Table.6 measured Accuracy

To find out which approach is superior for classifying reviews, we test and prime the model using the dataset, then compare their predicted efficiency. With the help of the table, it is shown. Naïve Bayes has the lowest anticipating accuracy,

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

whereas the Random Forest model has the best predictive validity out of the three models. By comparison, Bernoulli Naïve Bayes and multinomial Naïve Bayes are less efficient than logistic regression. Among all, Bernoulli Naïve Bayes is the one with the lowest accuracy.

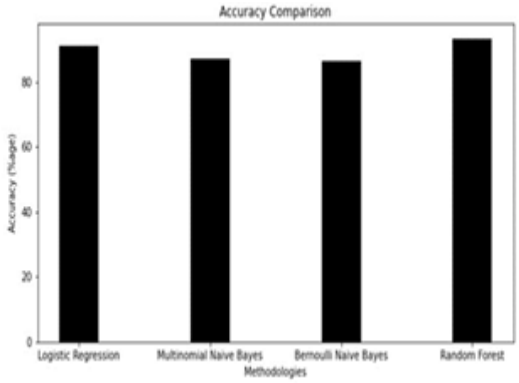

fig.4 Comparison of accuracy for different methods

Used Figure 4 shows the result after considering all of the algorithms. The results show that random forest performs better. There are a plethora of techniques and high-level mathematical calculations that can optimize the data for sentiment recognition and analysis to the best of our abilities.

CONCLUSION AND FUTURE WORK

In this modern era, we are seeing a dramatic shift from virtual to digital platforms. Customers and clients rely more and more on reviews posted online. An online platform for increasing belief and influencing consumer purchase patterns has been boosted by digital views. To achieve this goal, our project will analyze customer reviews on Amazon and classify them as positive, neutral, or negative. After merging the data with a few neutral and unfavorable thoughts, four classification models were used to detect reviews. Random Forest has the best prediction accuracy (93.17 percent) compared to Multinomial and Bernoulli Naïve Bayes, Logistic Regression, and the other classifiers. Future iterations of the work could include opinion-based product grade estimation. Since the product's grade and the reviewer's emotional investment do not always align, this would provide buyers with a trustworthy rating. More consistent results are possible if we combine our dataset with an equal amount of positive, negative, and neutral comments. Electronic commerce assiduity will benefit much from the chosen work augmentation as it will increase consumer trust and loyalty. Hyperparameters and natural language processing allow us to improve our data even further. The data will be optimized and the calculation result will be improved with this kind of method use.

REFERENCES

[1]. [1] C. Yue, H. Cao, G. Xu, and Y. Dong, “Collaborative attention neural network for multi-domain sentiment classification,” Appl. Intell., 2020, doi: 10.1007/s10489-020-02021-7. G. More, “Sentiment Analysis on Amazon Product Reviews with Stacked Neural Networks,” no. October, 2020, doi: 10.13140/RG.2.2.14746.67524.

[2]. S. Thapa, S. Adhikari, A. Ghimire, and A. Aditya, “Feature Selection Based Twin-Support Vector Machine for the diagnosis of Parkinson’s Disease”, in 2020 8th R10 Humanitarian Technology Conference (R10-HTC), 2020: IEEE

[3]. S. Kundu and S. Chakraborti, “A comparative study of online consumer reviews of Apple iPhone across Amazon, Twitter and MouthShut platforms,” Electron. Commer. Res., no. 0123456789, 2020, doi: 10.1007/s10660-020-09429-w.

A. Ejaz, Z. Turabee, M. Rahim, and S. Khoja, “Opinion mining approaches on Amazon product reviews: A comparative study,” 2017 Int. Conf. Inf. Commun. Technol. ICICT 2017, vol. 2017- December, pp. 173–179, 2018, 10.1109/ICICT.2017.8320185. doi:

[4]. P. Pankaj, P. Pandey, M. Muskan, and N. Soni, “Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews,” Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019, pp. 320–322, 2019, doi: 10.1109/COMITCon.2019.8862258.

[5]. S. Adhikari, S. Thapa, and B. K. Shah, "Oversampling based Classifiers for Categorization of Radar Returns from the Ionosphere," in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020: IEEE, pp. 975-978, doi: 10.1109/ICESC48915.2020.9155833.

[6]. H. B. Wang, Y. Xue, X. Zhen, and X. Tu, “Domain Specific Learning for Sentiment Classification and Activity Recognition,” IEEE Access, vol. 6, pp. 53611–53619, 2018, doi: 10.1109/ACCESS.2018.2871349.

[7]. K. R. Jerripothula, A. Rai, K. Garg, and Y. S. Rautela, "Feature- Level Rating System Using Customer Reviews and Review Votes," IEEE Trans. Comput. Soc. Syst., vol. 7, no. 5, pp. 1210– 1219, 2020, doi: 10.1109/TCSS.2020.3010807.

[8]. J. Zhang, D. Chen, and M. Lu, "Combining sentiment analysis with a fuzzy kano model for product aspect preference recommendation," IEEE Access, vol. 6, pp. 59163–59172, 2018, doi: 10.1109/ACCESS.2018.2875026.

[9]. Benlahbib and E. H. Nfaoui, "Aggregating customer review attributes for online reputation generation," IEEE Access, vol. 8, pp. 96550–96564, 2020, doi: 10.1109/ACCESS.2020.2996805.

[10]. V. Gupta, A. Aggarwal, and T. Chakraborty, "Detecting and characterizing extremist reviewer groups in online product reviews," arXiv, vol. 7, no. 3, pp. 741–750, 2020.

[11]. Y. Zhou and S. Yang, "Roles of Review Numerical and Textual Characteristics on Review Helpfulness Across Three Different Types of Reviews," IEEE Access, vol. 7, pp. 27769–27780, 2019, doi: 10.1109/ACCESS.2019.2901472.

[12]. W. Zhao et al., "Weakly-supervised deep embedding for product review sentiment analysis," IEEE Trans. Knowl. Data Eng., vol. 30, no. 1, pp. 185–197, 2018, doi: 10.1109/TKDE.2017.2756658.

[13]. B. K. Shah, V. Kedia, R. Raut, S. Ansari, and A. Shroff, "Evaluation and Comparative Study of Edge Detection Techniques," vol. 22, no. 5, pp. 6–15, 2020, doi: 10.9790/0661- 2205030615.

[14]. S. Thapa, P. Singh, D. K. Jain, N. Bharill, A. Gupta, and M. Prasad, "Data-Driven Approach based on Feature Selection Technique for Early Diagnosis of Alzheimer's Disease", in 2020 International Joint Conference on Neural Networks (IJCNN), 2020: IEEE, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9207359.

[15]. S. Thapa, S. Adhikari, U. Naseem, P. Singh, G. Bharathy, and M. Prasad, "Detecting Alzheimer's Disease by Exploiting Linguistic Information from Nepali Transcript," in International Conference on Neural Information Processing, 2020: Springer, pp. 176-184.

[16]. A. Ghimire, S. Thapa, A. K. Jha, S. Adhikari, and A. Kumar, "Accelerating Business Growth with Big Data and Artificial Intelligence", in 2020 Fourth International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020: IEEE, pp. 441-448, doi: 10.1109/I-SMAC49090.2020.9243318.

[17]. A. Ghimire, S. Thapa, A. K. Jha, A. Kumar, A. Kumar, and S. Adhikari, "AI and IoT Solutions for Tackling COVID-19 Pandemic", in 2020 International Conference on Electronics, Communication and Aerospace Technology, 2020: IEEE