e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:05/May-2024 Impact Factor- 7.868 www.irjmets.com

AUTOMATIC ANALYSIS OF PRODUCT REVIEW USING COMPUTER VISON TECHNIQUES

Dhirendra Kumar Ojha[*1], Ashish Kumar Sharma[*2], Anand Prakash Singh[*3], Raunak Singh[*4], Mr. Alok[*5]

[*1,2,3,4]Student, Department of CSE, Greater Noida Institute of Technology, Greater Noida, U.P, India.

[*5]Assistant Prof, Department of CSE, Greater Noida Institute of Technology, Greater Noida, U.P, India.

DOI : https://www.doi.org/10.56726/IRJMETS57278

ABSTRACT

In today's competitive market, understanding client feedback is critical for businesses looking to improve their products and services. This study investigates the novel application of computer vision techniques to automatically interpret product reviews, with a focus on the visual content given by customers. Our study uses advanced image processing techniques to glean useful insights from photographs that accompany product reviews. Using image recognition tools, we detect significant features, objects, and emotional cues in these photographs. This computerized analysis enables businesses to acquire a better grasp of client feelings and preferences than textual reviews.

Our approach entails training models to recognize specific traits and expressions, allowing for a thorough examination of the visual elements in client feedback. The insights gained from this study can assist organizations in making informed, data-driven decisions to improve product quality, increase customer satisfaction, and better customize their offers to fit consumer wants. The use of computer vision in product review analysis marks a huge step forward in understanding consumer behaviour, providing a more nuanced view of how things are perceived in the market. This study highlights the potential of visual content analysis to drive innovation and create a customer-centric approach to product development and marketing tactics.

Keywords: Computer Vision, Microstructure, Machine learning, Deep Learning.

I. INTRODUCTION

To grow any business intelligently in a competitive market, identification of potential customer should be done timely. Online consumer reviews have become a critical resource for buyers evaluating products on e-commerce platforms. Reviews provide experiential information from existing users that complements product specifications and descriptions from sellers. They enable buyers to make better-informed purchasing decisions. However, the enormous volume of reviews available for popular products makes comprehensive manual analysis impossible. For instance, on Amazon, the iPhone 12 has received over 70,000 customer reviews. Manually reading through such a high quantity of reviews with long textual feedback and images of varying quality is simply infeasible for consumers. This has driven significant interest in developing intelligent systems that can automate the analysis of textual and visual content in online consumer reviews using advanced computer vision techniques.

Computer vision refers to the automated extraction of meaningful information from digital images and videos. Deep learning approaches like convolutional neural networks (CNNs) have revolutionized modern computer vision. Computer vision techniques have diverse applications for automated analysis of visual content in product reviews – from detecting product objects in images to predicting image aesthetics. Automated analysis aims to distill millions of consumer reviews into useful summary insights and metrics for enabling better-informed purchase decisions.

In this comprehensive survey, we provide a holistic overview of the current state-of-the-art in enabling automated analysis of textual and visual content in online consumer reviews using computer vision techniques. The key contributions of this survey are as follows:

Discussion of commonly used datasets and evaluation metrics for developing and assessing techniques.

Review of traditional computer vision techniques like SIFT, HOG, and modern deep CNN architectures commonly applied for automated review analysis.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:05/May-2024 Impact Factor- 7.868 www.irjmets.com

Overview of core applications including review summarization, highlight extraction, defect detection, retrieval systems, fake review identification, and design feedback mining.

The exhaustive study presented through this survey will enable readers to develop an in-depth perspective on the current capabilities and limitations of computer vision techniques for automating the analysis of online consumer reviews. It provides a holistic landscape view of this fascinating area of research combining computer vision, machine learning, and consumer review analytics. The survey highlights crucial developments, discusses practical examples, identifies challenges, and most importantly, lays out pathways for pushing forward progress on this immensely useful application at the intersection of computer vision and textual/visual review analysis[4].

II. METHODOLOGY

Method The methodology for automated analysis of product reviews using computer vision techniques encompasses several key steps aimed at extracting meaningful insights from textual and visual data. Initially, the process involves importing and preprocessing the data to ensure its readiness for analysis. Once prepared, the focus shifts to the implementation of computer vision algorithms for review analysis.

The primary approach employed in this methodology involves leveraging computer vision techniques to extract relevant features from product review images. By utilizing methods such as convolutional neural networks (CNNs), the system can automatically identify and classify objects, sentiments, and other pertinent information depicted in the images. In order for the system to interpret variables such as product appearance, usage scenarios, and overall customer happiness based on visual cues, it is imperative that this phase be completed in order to comprehend the visual content of the evaluations [1].

The approach incorporates natural language processing (NLP) techniques to assess the reviews' written content after visual features are extracted. NLP algorithms are employed to parse the text, identify key phrases, sentiments, and themes, and extract actionable insights from the reviews. By combining computer vision with NLP, the system can comprehensively analyze both the visual and textual aspects of product reviews, providing a holistic understanding of customer feedback [2].

To facilitate the analysis process, the methodology adopts clustering algorithms, such as K-means and DBSCAN, to group similar reviews based on their visual and textual characteristics. Clustering helps identify patterns, trends, and common sentiments among the reviews, enabling the system to categorize them into distinct clusters representing different aspects of product performance, customer preferences, and satisfaction levels [3][7].

Additionally, the methodology incorporates machine learning models, particularly decision tree classifiers and random forest classifiers, to further analyze and classify the reviews based on predefined criteria. These models enable the system to predict customer sentiments, identify important features influencing product perceptions, and make recommendations for product improvements or marketing strategies [6][11].

To validate the effectiveness of the analysis, the methodology employs cross-validation techniques, ensuring robustness and reliability in the evaluation process. By splitting the data into multiple subsets and evaluating the performance of the models across different validation sets, the system can assess its predictive accuracy, sensitivity to data variations, and potential for generalization to unseen data [5][10].

Throughout the analysis, the methodology emphasizes the interpretability of results and the actionable insights derived from the review data. The system facilitates informed decision-making, optimizes product offers, and improves customer experiences for enterprises by clarifying certain patterns, trends, and client preferences [12].

Sentiment Classification Algorithms:

• **Naïve Bayesian classifier**: This algorithm utilizes conditional probability to predict the class of a new data point based on observed features. Using the Bayes theorem, it determines the posterior probability of each class given the observed characteristics given a training dataset represented by feature vectors and multiple classes. In figure 2(a), the class with the highest posterior probability is allocated to the new data point, resulting in a simple yet efficient method of sentiment categorization.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:05/May-2024 Impact Factor- 7.868 www.irjmets.com

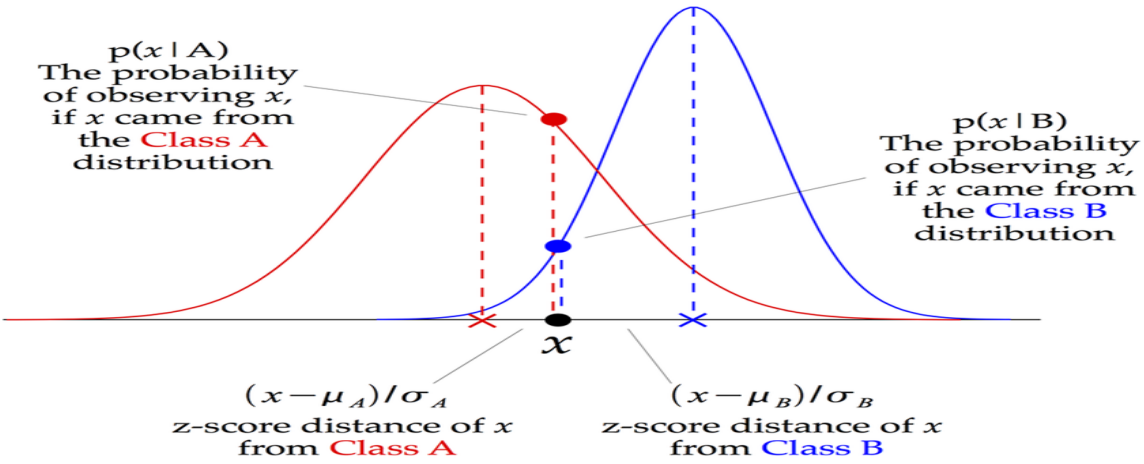

Figure 2(a). Naïve Bayesian classification

• **Random forest classifier**: Chosen for its superior performance, the random forest algorithm employs ensemble learning through bagging. It creates multiple decision trees by bootstrapping from the original dataset and aggregates their predictions to make a final classification. Each decision tree contributes to the classification process by casting a vote, with the majority vote determining the final class assignment. This ensemble approach enhances robustness and generalization capabilities, in figure 2(b).

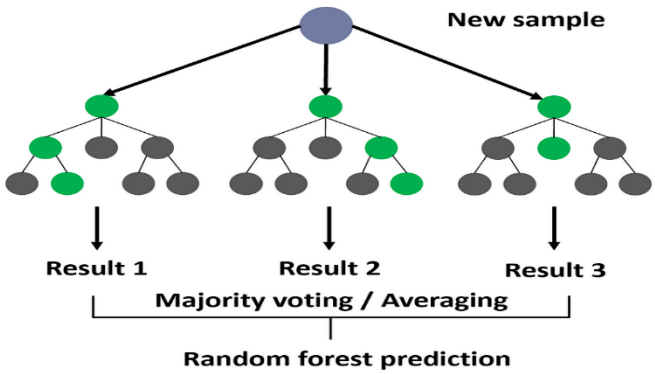

Figure 2(b). Random Forest Prediction Analysis

• **Support vector machine (SVM):** A flexible technique that can be used for both linear and nonlinear classification challenges. SVM looks for the best hyperplane to maximize the margin between classes in data that is linearly separable. SVM maps nonlinear data into a higher-dimensional space where linear separation is feasible by using kernel functions. SVM can recognize complicated correlations in the data thanks to a popular choice of the Gaussian Radial Basis Function (RBF) kernel. Because of its adaptability, SVM can be used for a variety of classification applications, including sentiment analysis.

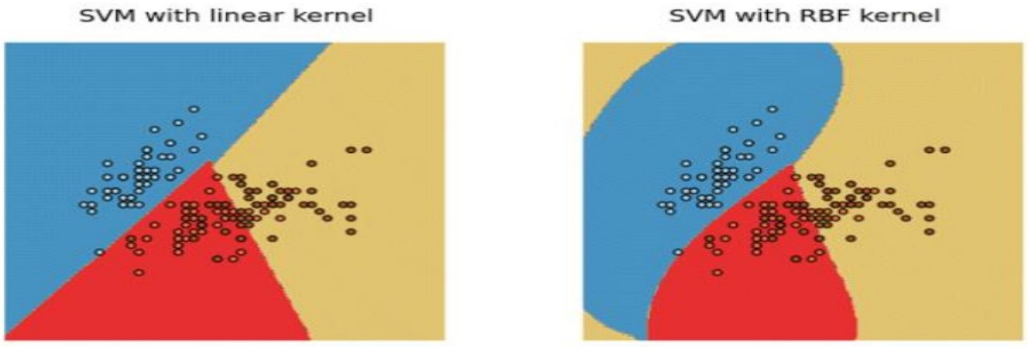

Figure 2(c) Support Vactor Machine analysis

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:05/May-2024 Impact Factor- 7.868 www.irjmets.com

• RFM ANALYSIS

The abbreviation of RFM means recency, frequency, monetary. This RFM analysis may be a marketing technique want to determine quantitatively which customers are the simplest ones by examining how recently a customer/people has purchased (recency), how often people purchase (frequency), and how much the customer spends (monetary). Let's see how this algorithm works, the substance of RFM analysis is to divide guests into groups grounded on how lately guests made their last purchase, how frequently people buy effects, and thus the average value of their orders. Each of these criteria, assign guests to one of three groups, which are assigned as figures.

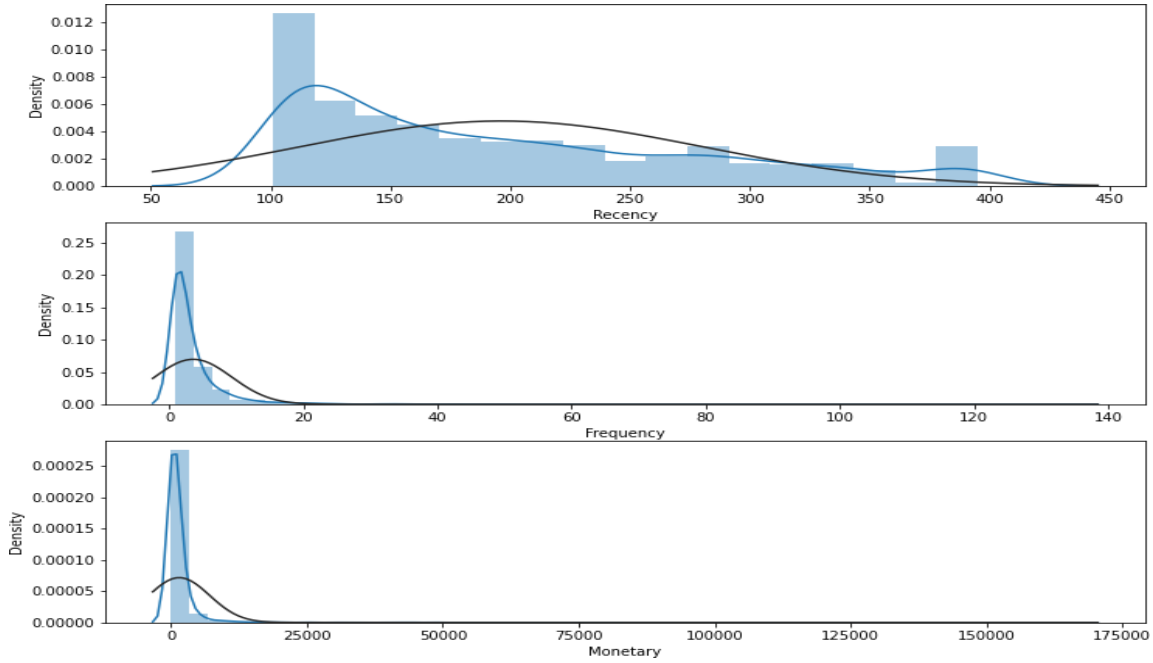

Figure 2(d). RFM Distribution Graph.

"Frequency is calculated as the outside of the number of purchases by the client in the last 12 months as shown in figure 2(d). Monetary is calculated as the loftiest value of all purchases by the client expressed as a multiple of some standard value" [16].

III. ABOUT THE DATASET

The first step in the research process is to collect the data needed for the classifiers' training and testing. The information is obtained from SNAP data sets because Amazon does not provide an API that functions similarly to Twitter and permits users to retrieve reviews. One review per line made up the JSON file that was downloaded. The file was changed to Comma Separated Values (CSV) format since Python handles this type of file more readily. In the data collection, there are 252000 reviews of different beauty products. Each review includes the nine features listed below.

Table 1: Every product's features

Feature	Description
ReviewerId	Id of the user
ReviewerName	Name of the user
"stu"	Product Id
Helpful	Part of the reader who thought the review was beneficial
Review text	The review's text
Overall	Product rating
Summary	Review the synopsis.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:05/May-2024 Impact Factor- 7.868 www.irjmets.com

An example JSON file review is shown below.:

{" ReviewerID":" ADFEE214424",

" ReviewerName":" Alpha2121 Customer",

" stu":" DFDEHD2",

" Helpful": [1, 1],

"ReviewText": "I needed certain things, and this product satisfied them. It felt light on my skin, effortlessly smoothed over, and free of oil. After applying the concealer, you'll need to use a press power appropriate for your skin tone. Really happy with the outcome.",

" Overall": 5.0,

" Summary":" Upper Lip Shadow Magic Stick",

}

IV. RESULTS AND DISCUSSION

The This section presents the study's findings. The percentage of the testing data set that the model correctly classifies is shown by the accuracy value. The accuracy of two different machine learning techniques on two sets of experiments is shown in Tables 2, 3, and 4.

FIRST EXPERIMENT

The results of the first experiment are shown in Table 2, which compares the accuracy of SVM and Naive Bayes on the complete set of data, including summaries and reviews.

Table 2: The accuracy of the machine learning methods on the whole data set

	Naïve Bayes	SVM
On reviews	80.16%	81.02%
On ummaries	82.72%	84.20%

In both scenarios, the accuracy of both algorithms was over 90%; however, SVM produced superior outcomes. When applied to the summaries, the Naïve Bayes method yielded better results.

SECOND EXPERIMENT

The findings from the second experiment, which was conducted twice on the reviews and summaries, are displayed in Tables 3 and 4. In this experiment, the top ten goods with the most views were used to train and test the classifiers.

Table 3: Using the reviews as input, the machine learning techniques' accuracy on ten products

ProductId	Naïve Bayes	SVM
DIDKFDSFO1	86.79%	89.63%
DIDKFDSFO2	86.79%	89.05%
DIDKFDSFO3	86.79%	84.23%
DIDKFDSFO4	86.79%	84.02%
DIDKFDSFO5	86.79%	89.01%
DIDKFDSFO6	86.79%	83.19%
DIDKFDSFO7	86.79%	86.35%
DIDKFDSFO8	86.79%	87.06%
DIDKFDSFO9	86.79%	83.12%
DIDKFDSFO10	86.79%	89.45%

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:05/May-2024 Impact Factor- 7.868 www.irjmets.com

Table 4: The machine learning techniques' accuracy on ten products utilizing the summary

ProductId	Naïve Bayes	SVM
DIDKFDSFO1	87.18%	87.93%
DIDKFDSFO2	89.01%	89.20%
DIDKFDSFO3	89.03%	89.13%
DIDKFDSFO4	89.00%	89.03%
DIDKFDSFO5	85.07%	87.10%
DIDKFDSFO6	88.01%	88.23%
DIDKFDSFO7	87.98%	88.03%
DIDKFDSFO8	88.91%	88.92%
DIDKFDSFO9	82.75%	83.04%
DIDKFDSFO10	80.68%	82.23%

The Naïve Bayes technique outperformed the SVM in both circumstances in the second set of studies, where there were far less reviews than in the first. Still, the accuracy, which ranges from 80% to 90%, is fairly good.)

V. CONCLUSION

This study addresses a critical requirement in the competitive landscape of e-commerce enterprises by delving into the field of automated analysis of product reviews using computer vision techniques. Through the application of cutting-edge techniques including association rules, RFM analysis, and clustering with k-means, as well as the use of DBSCAN in conjunction with the better approach, we have attempted to effectively address the problem of finding new clients. Based on focused client identification and purchasing habits, this integrated strategy helps marketing managers make strategic and lucrative decisions by illuminating customer opinions on products.

Furthermore, by solving the cold start issue that arises with new client encounters, this research makes a substantial contribution to raising sales profitability. Furthermore, our model offers a thorough grasp of client preferences and behaviors by combining demographic information—such as age, gender, and income—with customer purchasing patterns. This integrated method promises to give firms a competitive edge and improve sales profitability because it can be implemented directly.

VI. REFERENCES

[1] Balmeet Kaur, Pankaj Kumar Sharma, “International Journal of Innovative Technology and Exploring Engineering (IJITEE)”, ISSN: 2278-3075, Volume-8 Issue-6S, April 2019 .

[2] Chinedu Pascal Ezenkwu, Simeon Ozuomba, Constance kalu, Electrical/Electronics & Computer Engineering Department, University of Uyo, Uyo,Akwa Ibom State, Nigeria,” (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.10, 2015”.

[3] J. du Toit, R. Davimes, A. Mohamed, K. Patel,and J. M. Nye, Eskom, Brackenfell, South Africa,”Journal of Advances in Information Technology Vol. 7, No. 2, May 2016”.

[4] Hasan Ziafat , Majid Shakeri,” Hasan Ziafat Int. Journal of Engineering Research andApplications “,ISSN : 2248-9622, Vol. 4, Issue 9(Version 3), September 2014, pp.70-79 .

[5] Ion Smeureanu1, Gheorghe Ruxanda2, Laura Maria Badea3,” Journal of Business Economics and Management”, ISSN 1611-1699 print/ISSN2029-4433 online, 2013 Volume 14(5): 923–939.

[6] Juni Nurma Sari 1,2 , Lukito Edi Nugroho1,Ridi Ferdiana 1 ,P. Insap Santosa 1 1 Department of Electrical Engineering and Information Technology, University of Gadjah Mada, Jogjakarta, Indonesia 2 Department of InformaticsTechnology, Polytechnic Caltex Riau, Pekanbaru, Indonesia,” Review on Customer Segmentation Technique on Ecommerce”.

[7] A. Joy Christy a , A. Umamakeswari a , L. Priyatharsini b , A. Neyaa, “RFM ranking – An effective approach to customer segmentation”

[8] Kishana R. Kashwan, Member IACSIT, and C.M. Velu," International Journal of Computer Theory and Engineering, Vol. 5, No. 6, December 2013".

[9] YongSeog Kim, W Nick Street, Gary J Russell and FilippoMenczer (2017) "Customertargeting: A neural network approach guided by genetic algorithms", Management Science,vol. 51, no. 2, pp. 264-276.

[10] P. Kotler and K. L. Keller (2016) Marketing management, Pearson EductationInc, pp.268-269, 2016.

[11] Ş.Ozan,"A Case Studon Customer Segmentation by using Machine Learning Methods," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018, pp. 1-6, doi: 10.1109/IDAP.2018.8620892.

[12] Derrick S Boone and Michelle Roehm (2016) "Customer segmentation using artificial neural networks", International journalofresearchinmarketing, vol.19, no.3, pp.287- 301.Wierzchoń, S. T. and Kłopotek, M. A. (2018) Modern algorithms of cluster

[13] Wierzchoń, S. T. and Kłopotek, M. A. (2018) Modern algorithms of cluster analysis.

[14] Springer M. J. A. Berry and G. S. Linoff (2000). Mastering Data Mining. New York: Wiley.

[15] I. S. Dhillon and D. M. Modha, "Concept decompositions for large sparse text data using Clustering," Machine Learning, vol. 42, issue 1, pp. 143-175, 2001.

[16] Hosseini, S.M.S., Maleki, A., Gholamian, M.R.: Cluster analysis using data mining Approach to develop CRM methodology to assess the customer loyalty. Expert Syst. Appl. 37(7), 5259–5264 (2010)

[17] Subramanya, K. B., Somani, A.: Enhanced feature mining and classifier models to predict Customer churn for an E-retailer. In 2017 7th International Conference on Cloud Computing, Data Science and Engineering-Confluence, pp. 531–536. IEEE, January 2017.