INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 11, November 2023, pp : 170-173

DESIGN OF NLP TECHNIQUES FOR TEXT SUMMARIZATION AND CATEGORIZATION

Pawar Santosh Sanjay[1], Abinash Kumar[2]

[1]PG Scholar, CSD, Dr. APJ Abdul Kalam University Indore, M.P., India.

[2]Assistant Professor, CSD, Dr. APJ Abdul Kalam University Indore, M.P., India.

ABSTRACT

The Amazon and Flipkart datasets have been used in a number of tests. It is applied to a number of products on two e-commerce websites named Products of all kinds are described on Amazon and Flipkart. The. The PCSA system gathers review data for 10 categories, including mobile devices such as laptops, phones, cameras, air conditioners (AC), routers, TVs, books, and articles

comprising apparel, kitchenware, furnishings, and transportable goods. We've tested this system on 37,344 evaluations. Numerous observations were made based on the experimental findings. Both positive and negative ratings were most prevalent in the "Mobile Phone" category on Flipkart and Amazon. Senti WordNet and the logistic regression classifiers produced the best ratings of 4.24 and 4.51 for products sold on Amazon and Flipkart, respectively. Product star ratings on Flipkart were anticipated by the PCSA method to be greater than those on Amazon. When compared to Amazon, Flipkart received greater ratings for all classification algorithms. First, the login and registration information is set. The user is required to supply the set of links, the number of links, and the categorization technique name. The training phase and the testing phase are the two stages of this system. The training step uses input from known comments from sources such as Flipkart and Amazon. After that, they undergo preprocessing and are organized. These comments provide the segmented sentences and elements that are taken from them. These functions are used to train all five classifiers. After that, the system asks the user to choose the classifier, which separates the comments into groups that are mutually exclusive.

Keywords. Senti WordNet, PCSA system, PCSA system, e-commerce websites, indexing algorithm

1. INTRODUCTION

The retail market industry has expanded in recent years to include online product sales as well as the ability for customers to offer their insightful opinions, recommendations, and suggestions. Within a sizable text-based review collection, the opinion summary and classification methods extract and identify a variety of viewpoints regarding various online-available products. Over time, numerous automated opinion classification systems have advanced in this direction, making it one of the most challenging fields in natural language processing. Numerous methods of this type have been created and put into use for the purpose of classifying and summarizing reviews and text about online goods. Numerous data sources and websites, like Amazon, Flipkart, Snapdeal, and others, are used to sell goods online. An notion, viewpoint, or state of mind is referred to as sentiment or opinion, particularly when it is predominantly based on emotion rather than logic. Customers' opinions are represented by their sentiments, which might be neutral or positive, negative, or like. One can voice an opinion about something in its entirety or about any one of its characteristics. An individual's opinions are their thoughts, assessments, judgments, and convictions on a certain topic.

Opinion mining is the process of locating and extracting specific information from source sources using computational linguistics, text analysis, and language processing. Sentiment analysis takes into account people's opinions about particular things. Sentiment analysis is the algorithmic handling of subjectivity, sentiment, and views in text. The specifics of several classification methods for comment analysis utilizing NLP principles are shown in this section. The classification is provided in Section 1.3.1. technique strategy for comment analyzing. They consist of strategies based on qualities, sources, behavior, and language as well as learning and sentiment. Each of these methods is further divided into smaller categories. Segment 1.3.2 outlines five of the best supervised learning methods, including Naïve Bayes, Logistic K-Nearest Neighbor, Random Forest, SentiWordNet, and regression. The final portion 1.3.3 presents the ideas of classification and comment summarizing, and then the two-way relationship.

2. OBJECTIVE OF THE WORK

The research project's objective is to develop an automatic comment analyzer.The goal of this research project is to develop an automated system for classifying and analyzing comments that can efficiently ascertain the polarity of user feedback gathered from the Flipkart and Amazon data domains. The vast amount of reviews ought to be handled using this system. It should use five of the best supervised learning classifiers—NB, LR, SentiWordNet, RF, and KNN—to

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 11, November 2023, pp : 170-173

classify the comments into positive, negative, and neutral categories. use of clustering techniques to enhance web crawlers' efficiency.

This research project's aim is restricted to gathering English-language reviews exclusively from the Flipkart and Amazon websites. Secondly, emoticons and similar images will not be processed by it. Stated differently, its capabilities will be limited to processing and classifying the text. Thirdly, the evaluation of the items will be the main priority. Thus, reviews that are based on services or discussion forums are outside the purview of this suggested task.

3. METHODOLOGY

The proposed Product Comment Summarizer and Analyzer (PCSA) systemdesign is a generic, robust and fast system, which classifies online English comments collected from Amazon and Flipkart shopping websites using five different supervised learning classification techniques. These techniques are Naïve Bayes, logistic regression, SentiWordNet, random forest and K-Nearest Neighbor. The PCSA system is designed in the training and testing phases. During the system training, the user login registration form and credentials data base is set. It accepts multiples product URLs (Uniform Resource Locators) either from Amazon or from Flipkart websites. It preprocesses the comments, segments the sentences, extracts their features, summarizes them and lastly classifies them using any one classification techniques. The system is trained with all five classification techniques. It is designed to classify the comments of multiple products through any one technique at a time. During the system testing, this proposed PCSA system is tested for a different unknown set of product comments collected from both websites. It goes through all these steps one by one and the chosen trained classifier categorizes these comments efficiently

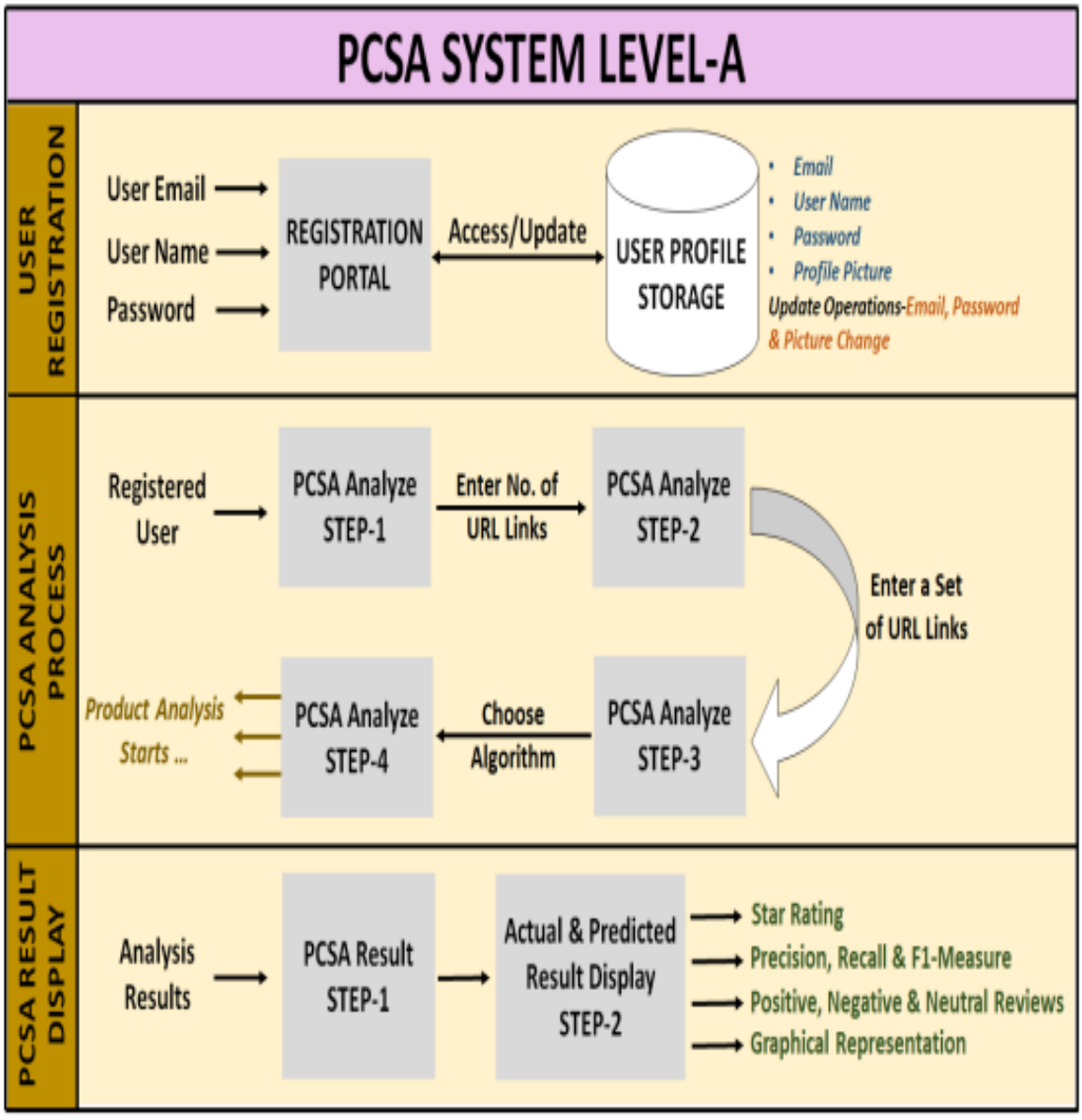

Figure 1: PCSA system level-A: Detailed PCSA system Design

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 11, November 2023, pp : 170-173

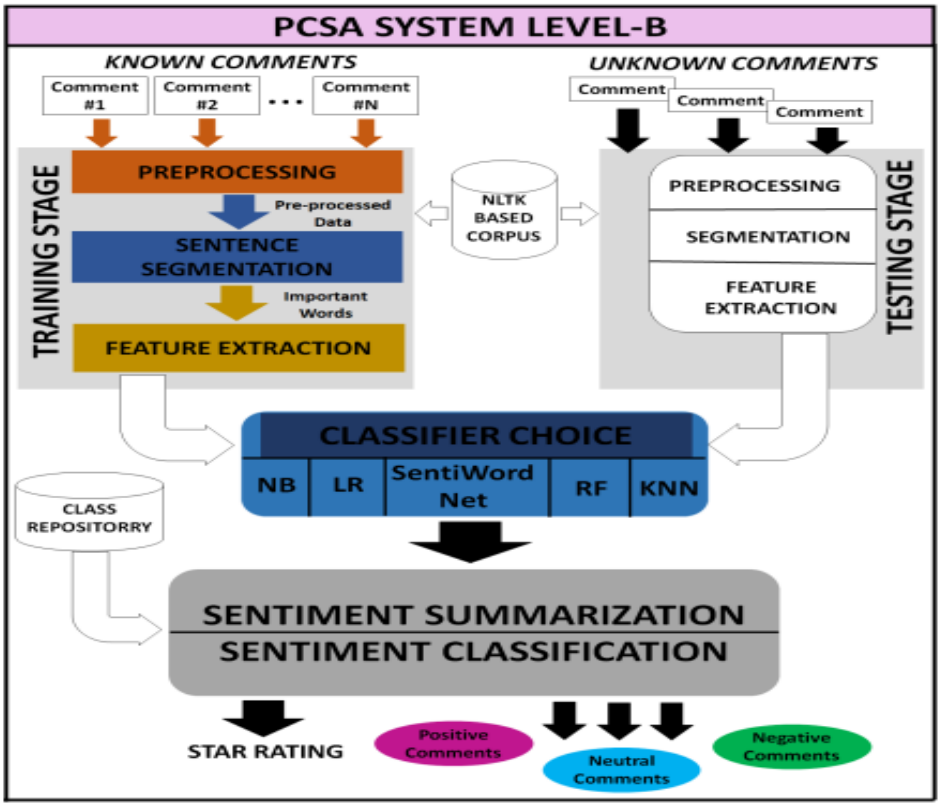

Figure 2: PCSA system level-B: Detailed PCSA system Design

4. RESULT

The web crawling, preprocessing, feature extraction, and classification processes for the various products on various e-commerce websites were all incorporated in the proposed PCSA system. For implementation, it made use of the usual Amazon and Flipkart datasets. It conducted the tests on English text comments devoid of any emojis, smileys, or images. Numerous observational data are examined based on the PCSA system's experimental outcomes. The technique was tested using a total of 47958 training reviews from Flipkart and 140882 training reviews from the Amazon website. 21413 for Amazon testing, 162295 for all of Amazon, 15931 for Flipkart testing, 63889 for all of Flipkart, 188840 for system training, 37344 for system testing, and 226184 for overall reviews were among the total reviews that were used. According to additional data, it received 906 neutral, 8054 negative, and 98105 favorable evaluations for Amazon products. Additionally, it received 620 neutral, 8329 negative, and 70706 good ratings for Flipkart products. This led to the discovery of the final totals for the following reviews: 168811 positive, 16383 negative, and 1526 neutral. Each of the five classifiers completed a total of 107065 and 79655 reviews for Flipkart and Amazon, respectively. It was noted that there is no rating difference for the "TV" and "laptop" categories on Flipkart or Amazon. The greatest variations discovered were in the "Camera" and "Clothes and Wearables" categories on Amazon, with a difference of -0.4 and +0.18, respectively. The greatest disparities observed were for "Camera" and +0.54 and -0.94.

5. CONCLUSION

The user registration phase of the proposed Product Comment summarizer and analyzer (PCSA) system required the user to provide some information on Visit the site for PCSA. Thus, the system initially obtained the product evaluations from the two websites, Flipkart and Amazon; separated and handled them; gathered their characteristics, condensed them, and categorized them into positive, After classifying the products into negative and neutral categories, the ratings were given. The Five supervised learning algorithms were used for classification, and they were Random forest, SentiWordNet, logistic regression, Naïve Bayes, and K-nearest adjoint. The outcomes were shown on the portal by the PCSA system. The PCSA system's successful implementation and a number of experimental outcomes have been demonstrated. In order to conduct these trials, a sizable amount of product category reviews are gathered from for-profit websites like Flipkart and Amazon.One-quarter of the product subcategories were used for the testing phase and the remaining three-quarters were used for the training phase. These outcomes pertain to classification of reviews, synthesis and rating, and category reviews. The study of the experimental results yielded numerous observations. The maximum positive and maximum negative values come first. Reviews were gathered for

the "Mobile Phone" category on Flipkart and Amazon.com. On Amazon and Flipkart, respectively, the "Mobile Phone" and "Router" categories yielded the highest number of indifferent ratings. The second is that, for Amazon and Flipkart products, SentiWordNet and logistic regression classifiers yielded the best ratings of 4.24 and 4.51, respectively. The third finding is that the PCSA system anticipated that products sold on Flipkart would have better star ratings than those found on Amazon. When compared to Amazon, Flipkart received greater ratings for all classification algorithms.

6. REFERENCES

[1] M. H. Dunham, Data Mining: Introductory and Advanced Topics, Pearson Education. J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques, 3rd edition, Waltham, Massachusetts: Morgan Kaufmann Publishers.

[2] N. P. Padhy, Artificial Intelligence and Intelligent Systems, 3rd Edition, Oxford, New York, Oxford University Press.

[3] Singh, V., & Dubey, S. K. (2014). “Opinion Mining and Analysis: A Literature Review," 5th International Conference-Confluence: The Next Generation Information Technology Summit, IEEE Press, pp. 232-239.

[4] Unknown Author, (2011). “Opinion Mining and Sentiment Analysis,” Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Data-Centric Systems and Applications, Springer-Verlag Berlin Heidelberg, pp. 459-526.

[5] Ezhilarasan, M., Govindasamy, V., Akila, V., & Vadivelan, K. (2019). “Sentiment Analysis on Product Review: A Survey,” International Conference on Computation of Power, Energy, Information and Communication, IEEE Press, pp. 180-192.

[6] Jadhav, H. B., & Jadhav, A. B. (2020). “Systematic Approach Towards Sentiment Analysis in Online Review‟s,” In: Pandian A., Senjyu T., Islam S., Wang H. (eds) Proceeding of the International Conference on Computer Networks, Big Data and IoT, Lecture Notes on Data Engineering and Communications Technologies, Springer, Cham, Vol. 31, pp. 358-369.

[7] N. Arunachalam, Sneka, S. J., & G. MadhuMathi, (2017). “A Survey on Text Classification Techniques for Sentiment Polarity Detection,” International Conference on Innovations in Power and Advanced Computing Technologies, IEEE Press, pp.1-5.

[8] ChandraKala, S., & Sindhu, C. (2012). “Opinion Mining and Sentiment Classification: A Survey,” ICTACT Journal on Soft Computing, Vol. 3, Issue 1, pp. 420-427.

[9] Rahul, Raj, V. & Monika, (2019). “Sentiment Analysis on Product Reviews” International Conference on Computing, Communication, and Intelligent Systems, IEEE Press, pp. 5-9.

[10] Chen, H., & Zimbra, D. (2010). “AI and Opinion Mining,” IEEE Intelligent Systems: Trends & Controversies, IEEE Computer Society, pp. 74-80.

[11] Nassr, Z., Sael, N., & Benabbou, F. (2020). “Machine Learning for Sentiment Analysis: A Survey,” In: Ben Ahmed M., Boudhir A., Santos D., El Aroussi M., Karas İ. (eds) Innovations in Smart Cities Applications Edition 3, Lecture Notes in Intelligent Transportation and Infrastructure, Springer, Cham, pp. 63-72.

[12] Singh, R. K., Sachan, M.K., & Patel, R.B. (2020). “360 Degree View of CrossDomain Opinion Classification: A Survey,” Artificial Intelligence Review, pp. 1-122.

[13] Esuli, A., & Sebastiani, F. (2006). “SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining,” Proceedings of the Fifth International Conference on Language Resources and Evaluation, European Language Resources Association, pp. 417-422.

[14] Baccianella, S., Esuli, A., & Sebastiani, F. (2010). “SENTIWORDNET3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining,” Proceedings of the Seventh International Conference on Language Resources and Evaluation, European Language Resources Association, pp. 2200-2204.

[15] Binali, H., Potdar, V., & Wu, C. (2009). “A State of the Art Opinion Mining and its Application Domains,” International Conference on Industrial Technology, IEEE Press, pp. 1-6.