

Examining Hidden Meaning of E-commerce Platform

Rajesh Bose, Raktim Kumar Dey, Srabanti Chakraborty, Sandip Roy, Debabrata Sarddar

Abstract: *Different e-commerce companies try to maintain high importance for their customer satisfactions. It helps them to understand the performance of their products. Nowadays customers trust on the product reviews while shipping online. But it is a cumbersome task to handle millions of customer reviews within specific time period. Due to this problem consumers usually follow the set of reviews before taking decision for purchasing any products from online. Although, each consumer rates the product from 1 to 5 stars, these overall product rating judge products towards their customers satisfaction. Consumers also provide a text based summary as a review of their experiences and opinions about the products. Customer sentiment analysis is a method to process these customer reviews to summarize different products. In this manuscript, we analyzed the text summary of Amazon food products using NRC Emotion Lexicon to determine the overall responses of the products using eight emotions of the customers. Our result can be used to take further decision making for the future of the products.*

Keywords: *Amazon customer reviews, Latent Dirichlet Allocation (LDA), Opinion mining, Sentiment Analysis, Topic modeling, Word Cloud.*

I. INTRODUCTION

Product reviews by the customers who have purchased and used the product, can express their opinion towards the product. Usually customers rate the product from 1 to 5 stars, and write a text summary of their experiences about the product [1]. These text summaries become meaningful for analyzing product quality and also important part for the purchasing process [2].

Sentiment analysis is the contextual mining process of determining whether a given review summary is positive, negative, neutral, or mixed [3, 4]. Topic modeling is a process which can extract topics or themes from text summaries given by the customers [5]. Latent Dirichlet Allocation (LDA) is a probabilistic model that can be automatically grouped words based on different topics [6]. Amazon Comprehend is a LDA based learning model to identify the topic in a set of given reviews of the customers [7].

Revised Manuscript Received on October 10, 2019

* Correspondence Author

Dr. Rajesh Bose, Simplex Infrastructures Limited, Kolkata, India.
Email: bose.raj00028@gmail.

Raktim Kumar Dey, Simplex Infrastructures Limited, Mumbai, India.
Email: deyraktim@gmail.com.

Srabanti Chakraborty, Dept. of CST, EIEM, Kolkata, India. Email: srabanti2k@gmail.org.

Dr. Sandip Roy*, Dept. of CSE, Brainware University, Kolkata, India.
Email: sandiproy86@gmail.com.

Dr. Debabrata Sarddar, Dept. of CSE, University of Kalyani, Kalyani, India. Email: dsarddar1@gmail.com.

In this research, we examined using NRC Emotion Lexicon which is based on mainly English words and can be categorized by eight basic emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) and two sentiments (negative and positive) [8].

Moreover, word cloud is used to categorize words into eight emotions which can be visualized different font sizes or colors [9, 10].

Our paper is organized as follows; section 2 discuss about literature survey. Section 3 describe about our proposed algorithm. Section 4 deals with result analysis and section 5 draw the conclusion and the future direction.

II. LITERATURE SURVEY

Sentiment analysis is very crucial for e-commerce websites. There were various algorithms used for analyzing sentiment of the customers. In [11], authors worked on the reviews like Amazon, and used algorithm such as Logistic Regression, Naïve Bayes classifier and SentiWordNet Algorithm to classify the review as positive and negative review. They had also used quality metric parameters to measure the performance of those algorithms.

Rana et al. proposed hybrid rule-based approach for extracting and categorizing customer reviews using Google similarity distance association with particle swarm optimization (PSO) technique [12]. Singh [13] proposed an algorithm to analyze data from different e-commerce websites which can improve the business outcomes and make sure a very high level of customer satisfaction.

Rain proposed a probabilistic Machine Learning algorithm in [14]. He also mentioned that many users reflected on specific components of products when they reviewed on those products. In [15], authors used four machine learning classifiers examined and compared viz. Naïve Bayes, J48, BFTree and OneR. Their result shown that Naïve Bayes method is quite fast in learning whereas OneR generated more promising accuracy of 91.3% in precision, 92.34% in classified instances and 97% in F-measure. Fang et al. performed sentence-level and review-level categorization on online product reviews from Amazon.com [16]. Hassan et al. proposed a novel approach by removing unstructured data and classified using Naïve Bayes algorithm [17]. Study suggests that many researchers used to analyze the reviews of the customers in different e-commerce websites using different algorithms. Our research objective is to examine the reviews using NRC Emotion Lexicon and categorized the reviews into eight emotions and two sentiments [18].

III. APPROACH

A. Dataset

The Amazon Fine Food Reviews have 5, 68, 454 reviews [19, 20]. 3, 63, 122 reviews have a score of 5, 80, 655 reviews have a score of 4, 23, 640 reviews have a score of 3, 29, 769 reviews have a score 2, and 52, 268 reviews have a score of 1. Fig. 1 and Fig. 2 illustrate the steps for analyzing customer reviews.

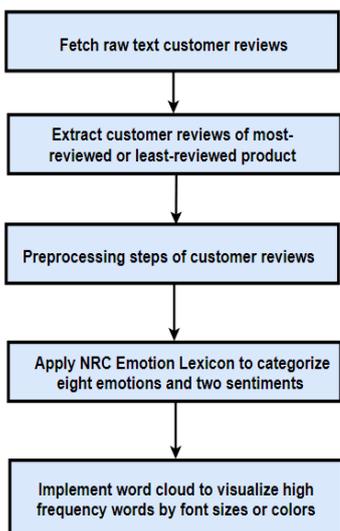


Fig. 1. Steps for review text summaries

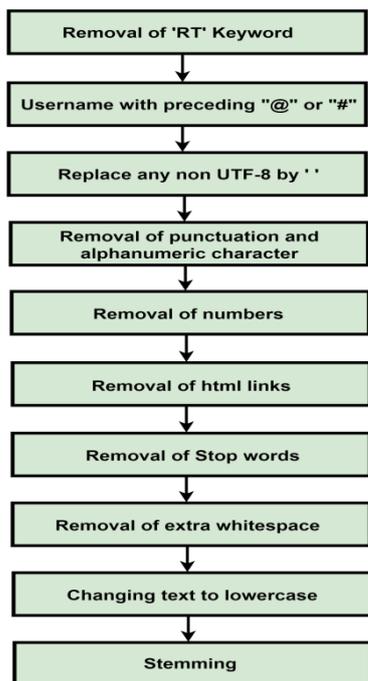


Fig. 2. Steps for review text summaries

B. Steps

At first, we imported the .csv dataset file into SQLite and applied SQL query to extract the reviews of different food products [21].

Then we performed data preprocessing method which is an important step in any text mining or opinion mining

process. Using the data preprocessing process, we eliminated all URLs, punctuations, symbols, numbers, and also replaced each words by their root words [22].

Lastly, we imported the “syuzhet”, “tm”, “wordcloud”, and “plotly” packages in R [23, 24]. We implemented the emotion analysis based on NRC Emotion Lexicon to categorize into eight basic emotions and two sentiments [25].

IV. RESULT AND ANALYSIS

In existing Amazon platform, users can review a product from 1 to 5, where low rate mean the specified product is less user-choice rather than the product with high rate. Moreover, when anyone wants to buy any product he/she can also check top positive reviews and top critical reviews of the customers along with star ratings in percentage [26]. There are different questions already answered which will be helpful for the customer to take decision for buying the product.

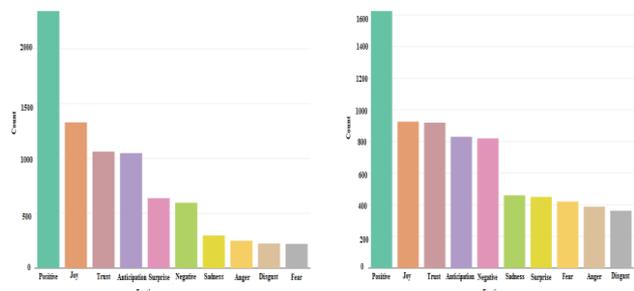
In our report, we analyze our results using NRC Emotion Lexicon and word cloud which is discussed below.

A. Sentiment Analysis

In our experiment, we analyzed mainly two types of products (e.g. most-reviewed or less-reviewed products). Fig. 3 a) – d) and Fig. 5 a) – d) show that the distribution of emotion of most or least numbers (except no review) of amazon food reviews respectively. Now, the results show that how most-reviewed products are getting more positive emotions than the less-reviewed products.

B. Word Cloud

Customer reviews are visualized by word cloud. Here we categorized word cloud by the eight emotions. The sizes of word fonts are proportional to their importance in the customer reviews. Word cloud can help the buyers quickly evaluate the quality of products and also identify the problem faced by the customers [27 - 29]. Fig. 4 a) – d) and Fig. 6 a) – d) illustrate that the word cloud of most or least numbers (except no review) of amazon food reviews respectively.



a)

b)

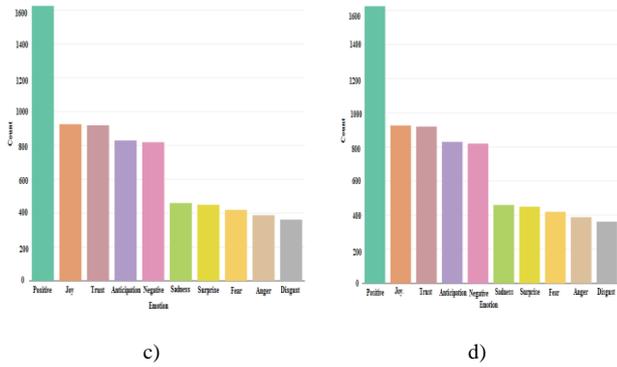


Fig. 3. Distribution of emotion of four most reviews Amazon food products

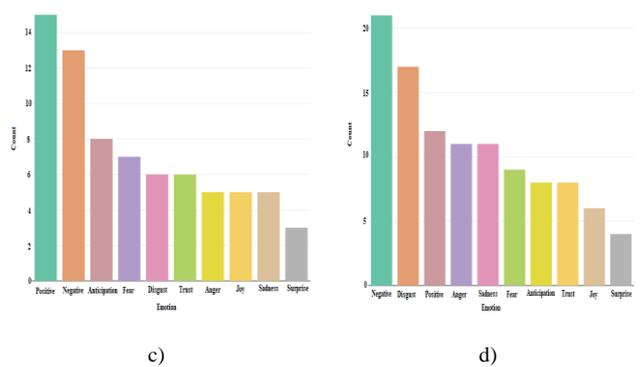


Fig. 5. Distribution of emotion of four least reviews Amazon food products

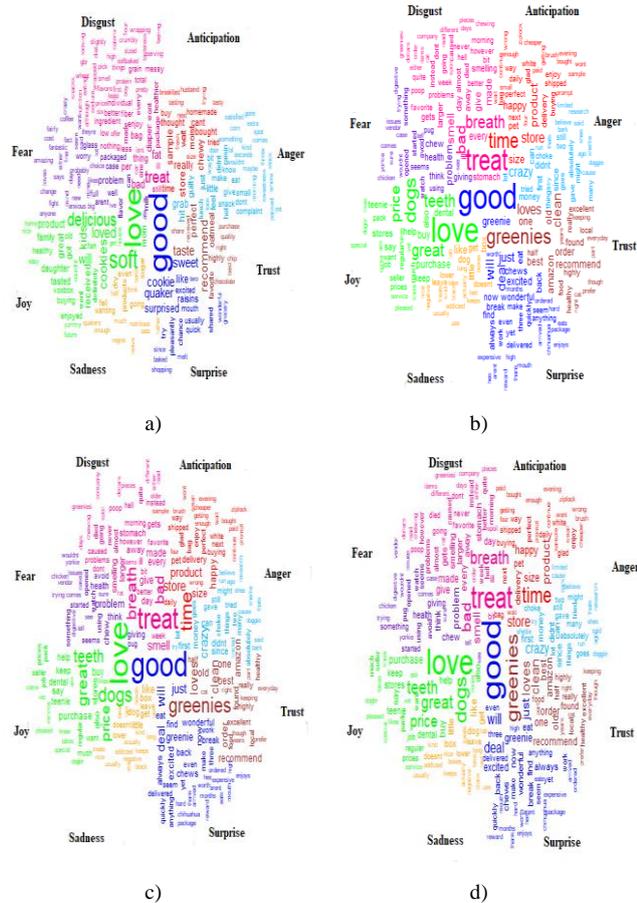


Fig. 4. Word cloud of four most reviews Amazon food products

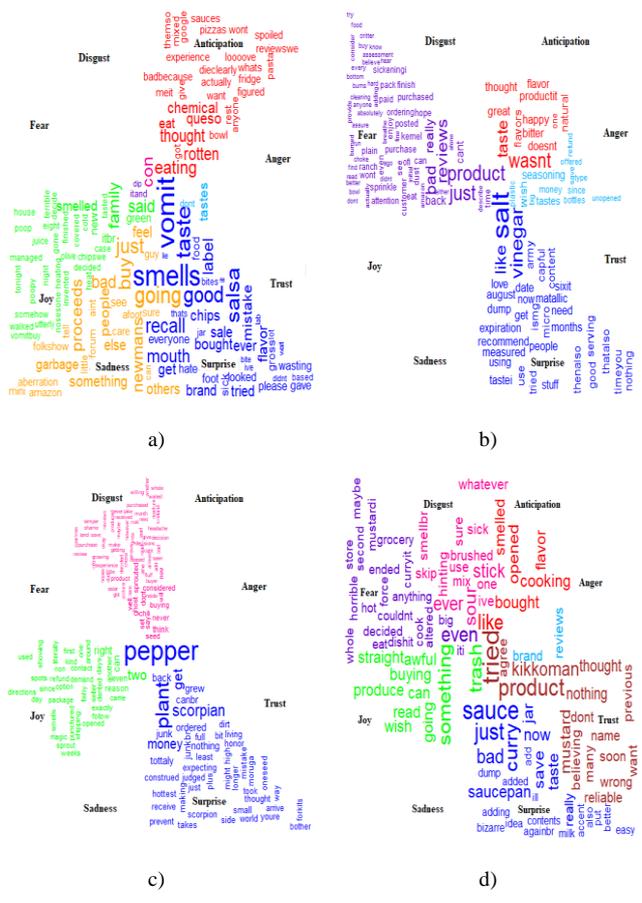
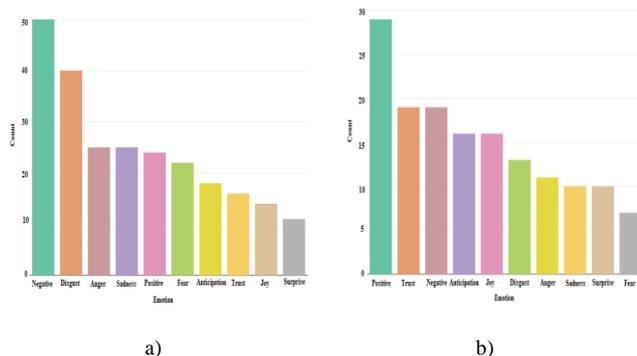


Fig. 6. Word cloud of four least reviews Amazon food products



V. CONCLUSION AND FUTURE WORKS

In this manuscript, we have proposed a novel approach for analyzing customer reviews using NRC Emotion Lexicon. Firstly, we categorized customer reviews into eight positive and negative emotions. Secondly, word clouds are a standard way to visually present the words in the text summary of the customer reviews. Typically the font size of a word cloud is proportional to its importance in the reviews.



In least-reviewed food, customers mostly commented by “vomit”, “smell”, “pepper”, and “taste” which identify the flavor of the foods whereas most-reviewed foods are mentioned by “good”, “love”, “greenies”, and “breath”.

In future, we will incorporate Latent Dirichlet Allocation (LDA) topic modeling technique which would help the customers to understand the quality of the products.

ACKNOWLEDGMENT

The authors sincerely acknowledge to the Simplex Infrastructures Ltd, Kolkata, India and Brainware University for providing data, infrastructure and support of other related facilities for successfully doing the research.

REFERENCES

1. M. Woolf. (2014, June). A Statistical Analysis of 1.2 Million Amazon Reviews. *Max Woolf's Blog*. Available: <https://minimaxir.com/2014/06/reviewing-reviews/>
2. Why Amazon Customer Reviews are Important for Sellers. *AMZINSIGHT*. Available: <https://www.amzinsight.com/amazon-product-review-importance/>
3. T. Escalona. (2018, January). Detect sentiment from customer reviews using Amazon Comprehend. *AWS Machine Learning Blog*. Available: <https://aws.amazon.com/blogs/machine-learning/detect-sentiment-from-customer-reviews-using-amazon-comprehend/>
4. Discover insights and relationships in text. *Amazon Comprehend*. Available: <https://aws.amazon.com/comprehend/>
5. K. Davenport. (2017, March). Topic Modeling Amazon Reviews. Available: <https://kldavenport.com/topic-modeling-amazon-reviews/>
6. D. M. Blei, A. Y. Ng, M. I. Jordan, “Latent Dirichlet Allocation,” *J. of Mac. Lear. Reas*, vol. 3, 2003, pp. 993 – 1022.
7. Topic Modeling. *Amazon Comprehend*. Available: <https://docs.aws.amazon.com/comprehend/latest/dg/topic-modeling.html>
8. S. Mohammad, P. Turney, “Crowdsourcing a Word-Emotion Association Lexicon,” *Computational Intelligence*, vol. 29, 2013, pp. 436 – 465.
9. WordCloud and Sentiment Analysis. Available: http://rstudio-pubs-static.s3.amazonaws.com/71296_3f3ee76e8ef34410a1635926f740c473.html
10. M. Hu, B. Liu, “Mining and Summarizing Customer Reviews,” *Proc. of the ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD-2004)*, 2004, pp. 1 - 10.
11. K. L. S. Kumar, J. Desai, J. Majumder, “Opinion mining and sentiment analysis on online customer review,” *ICIC*, 2016, pp. 1 - 4.
12. T. A. Rana, Y-N Cheah, “Hybrid rule-based approach for aspect extraction and categorization from customer reviews,” *9th Int. Conf. On IT in Asia (CITA)*, 2015, pp. 1 - 5.
13. P. K. Singh, A. Sachdeva, D. Mahajan, N. Pande, A. Sharma, “An approach towards feature specific opinion mining and sentimental analysis across e-commerce websites,” *5th Int. Conf. – Confluence The Next Generation Information Technology Summit*, 2014, pp. 329 – 335.
14. C. Rain, “Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning,” *Swarthmore College Computer Society*, 2013, pp. 1 – 7.
15. J. Singh, G. Singh, R. Singh, “Optimization of sentiment analysis using machine learning classifiers,” *Human-centric Comp. and Info. Sci.*, vol. 7, 2017, pp. 1 – 12.
16. X. Fang, J. Zhan, “Sentiment analysis using product review data,” *J. of Big Data*, vol. 2, June 2015, pp. 1 – 14.
17. M. K. Hassan, S. P. Shakthi, R. Sasikala, “Sentiment analysis of Amazon reviews using naïve bayes on laptop products with MongoDB and R,” *14th ICSET-2017*, vol. 263, 2017, pp. 1 – 10
18. R. Bose, R. K. Dey, S. Roy, Dr. D. Sarddar, “Analyzing Political Sentiment using Twitter Data,” *Proc. of ICTIS 2018*, 2018, pp. 1 - 9.
19. Amazon Fine Food Reviews-Analyze ~500, 000 food reviews from Amazon, Available: <https://www.kaggle.com/snap/amazon-fine-food-reviews>
20. J. McAuley, J. Leskovec, “From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews,” *WWW 2013*, 2013, pp. 1 - 11.
21. Import a CSV File Into an SQLite Table. *SQLITE TUTORIAL*. Available: <https://www.sqlitetutorial.net/sqlite-import-csv/>
22. Dr. S. Vijayarani, J. Ilamathi, M. Nithya, “Preprocessing Techniques for Text Mining – An Overview,” *J. of Com. Sci. & Comm. Net.*, vol. 5, 2015, pp. 7 – 16.
23. I. Feinerer, “Introduction to the tm Package Text Mining in R,” 2018, pp. 1 – 8.
24. M. Jockers. (2017, December). Introduction to the Syuzhet Package. Available: <https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html>
25. S. M. Mohammad, S. Kiritchenko, “Using Hashtags to Capture Fine Emotion Categories from Tweets,” *Sp. issue on Semantic Analy. in Soc. Med., Comp. Intell.*, 2013, pp. 1 – 22.
26. G. Lackermaier, D. Kailer, K. Kannaz, “Importance of Online Product Reviews from a Customer’s Perspective,” *Adv. in Economics and Business*, vol. 1, 2013, pp. 1 – 5.
27. S. Li, T.-S. Chua, “Document Visualization using Topic Clouds,” *NEXT Search Centre, National University of Singapore*, arXiv: 1702.01520v1 [cs.IR], 2017, pp. 1 – 5.
28. R. Bose, R. K. Dey, S. Roy, D. Sarddar, “Sentiment Analysis on Online Product Reviews,” *Inf. and Comm. Tech. for Sustainable Development. Adv. in Intelligent Sys. and Comp.* vol. 933, 2019, pp. 1 – 10.
29. D. Sarddar, R. K. Dey, R. Bose, S. Roy, “Topic Modeling as a Tool Gauge Political Sentiments from Twitter Feeds,” *Int. J. of Natural Comp. Research (IJNCR)*, vol. 9, 2019, pp. 1 – 22, to be published.

AUTHORS PROFILE



Dr. Rajesh Bose is an IT professional employed as Deputy Manager with Simplex Infrastructures Limited, Data Center, Kolkata. He has completed PhD from University of Kalyani, Kalyani, West Bengal. He graduated with a B.E. in Computer Science and Engineering from Biju Patnaik University of Technology (BPUT), Rourkela, Orissa, India in 2004. He went on to complete his degree in M.Tech. in mobile communication and networking from Maulana Abul Kalam Azad University of Technology, West Bengal (Formerly known as West Bengal University of Technology) - WBUT, India in 2007. He has also several global certifications under his belt. These are CCNA, CCNP-BCRAN, and CCA (Citrix Certified Administrator for Citrix Access Gateway 9 Enterprise Edition), CCA (Citrix Certified Administrator for Citrix Xen App 5 for Windows Server 2008). His research interests include cloud computing, IoT, wireless communication and networking. He has published more than 45 referred papers and 4 books in different journals and conferences. Mr. Bose also has experience in teaching at college level for a number of years prior to moving on to becoming a full time IT professional at a prominent construction company in India where he has been administering virtual platforms and systems at the company's data center.



Raktim Kumar Dey is an IT professional employed as Project Engineer with Simplex Infrastructures Limited, Mumbai. He received his degree in M.Tech. in Computer Science and Engineering from WBUT in 2018. He received his degree in B.E. in Computer Science and Engineering from BPUT in 2004. He also received degree in Masters in Information Management from Mumbai University in 2016. He has also two global certifications under his belt. These are SCJP (Sun Certified Programmer for the Java 2 Platform, Standard Edition 5.0), Oracle PL/SQL Developer Certified Associate. His research interests include sentiment analysis, data analysis, data science, data management, cloud computing and networking.





Ms. Srabanti Chakraborty is an Assistant Professor and Head of the Department of Computer Science & Technology of Elite Institute of Engineering & Management, Kolkata, West Bengal. She received her M. Tech. degree in Computer Science & Engineering in 2012 from Maulana Abul Kalam Azad University of Technology (Formerly known as West Bengal University of Technology), West Bengal and MCA degree in 2009 from Indira Gandhi National Open University (IGNOU). She has authored over 4 peer-reviewed papers in National and International journals and

conferences. Her main areas of research interests are Cloud Computing, Internet of Things (IoT) and Big Data.



Dr. Sandip Roy is an Associate Professor and HOD of Department of Computer Science & Engineering of Brainware University, Kolkata, West Bengal, India. He has awarded his PhD in Computer Science & Engineering from University of Kalyani, India in 2018. Mr. Roy received M.Tech. Degree in Computer Science & Engineering in 2011, and B.Tech. in Information Technology in 2008 from Maulana Abul Kalam Azad University of Technology, West Bengal (Formerly known as West Bengal University of Technology). He has authored over 30 papers in peer-reviewed journals, conferences, and is a recipient of the Best Paper Award from ICACEA in 2015. He has also authored of four books. His main areas of research interests are Data Science, Internet of Things, Cloud Computing, and Smart Technologies.



Dr. Debabrata Sarddar Assistant Professor of Department of Computer Science & Engineering of University of Kolkata, West Bengal, India. He has awarded his PhD in Computer Science & Engineering from University of Kalyani, India in 2018. Dr. Sarddar received his M.Tech. degree in Computer Science & Engineering from DAVV, Indore in 2006, and his B.Tech in Computer Science & Engineering from Regional Engineering College, Durgapur in 2001. He has authored over many papers in peer-reviewed journals, conferences. He has also published books. His research interest includes wireless and mobile system.