# A REAL AND ACCURATE FAKE PRODUCT DETECTION SYSTEM AND GENERATE ORIGINAL REVIEWS USING DATA MINING MECHANISM

VVS  SASANK[1] ,Praveen S R Konduri[2], Prasanna Kumar Prathipati[3]

[1]*Assistant Professor,Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,Vaddeswaram,Guntur,AP,India*

[2]*Assistant Professor,Department of Computer Science and Engineering, Malla Reddy College of Engineering and Technology (MRCET), Hyderabad, Telagana, India*

[3]*Assistant Professor,Department of Computer Science and Engineering,Hyderabad Institute of Technology & management,Hyderabad,Telangana,India*

***Abstract:****Most of the people requires genuine information about the online product. Before spending their economy on particular product can analyze the various reviews in the website. In this scenario, they did not identify whether it may be fake or genuine. In general, some reports in the websites are good, company technical people itself add these for making the product famous. These people belong to media and social organization teams, they give reviews with a good rating by their own firm. Online purchasers did not identify the fake product because of this falsification in the reviews of the website. In this research,the SVM classification mechanism has been used for detect the fake reviews by using IP address. This implementation helpful for users find out the correct review of online product. In this accuracy is improved by 98.79%, F1 score increases by 10%.*

*Keywords: Fake reviews, data mining, online product, real time marketing.*

## 1.  Introduction:

As the general public of the people require audit approximately an object earlier than spending their coins at the object. So individuals pass over one of a kind audits inside the website online however those surveys are veritable or counterfeit isn't recognized via the consumer. In a few survey sites some tremendous audits are blanketed by using the item enterprise people itself a good way to make so that you can create bogus tremendous item surveys. They provide super audits for a few, diverse items fabricated via their personal firm. Client won't have the option to look if the audit is certifiable or counterfeit. To find out counterfeit survey inside the web page this "Phony Product Review Monitoring and Removal for Genuine Online Product Reviews Using Opinion Mining" framework is offered. This framework will discover counterfeit surveys made through posting counterfeit remarks approximately an item with the aid of distinguishing the IP address along audit posting designs. Client will login to the framework making use of his customer id and secret phrase and will see specific items and will supply survey approximately the object. To find out the audit is phony or actual, framework will find out the IP address of the consumer if the framework watch counterfeit survey ship by a similar IP Address numerous multiple instances it's going to train the administrator to expel that survey from the framework. This framework utilizes statistics mining gadget. This framework encourages the client to find out proper survey of the object.

In present days the usage of Internet and internet based marketing has gotten mainstream. A high-quality many objects and administrations are available in web primarily based showcasing that create huge degree of statistics. Consequently, it is difficult to discover the quality affordable administrations or objects perfect to the prerequisite. Clients straightforwardly take preference dependent on audits or conclusions that are composed via others depending on their encounters. Right now any individual can compose something, this

boost the quantity of phony audits. Different agencies are employing individuals to compose counterfeit tremendous audits approximately their administrations or items or out of line bad surveys approximately their opponents' administrations or gadgets. This process gives incorrect contribution to the brand new customers who wish to buy such things and therefore we need a framework to differentiate such phony audits and expel them. Right now have a look at exclusive directed, unaided and semi controlled statistics digging approaches for counterfeit audit identification depending on diverse highlights.

## 2. Literature survey:

As of late, the World Wide Web has greatly changed the technique for imparting the insights. Online audits are feedback, tweets, posts, conclusions on numerous on line tiers like survey destinations, news locales, net based totally enterprise destinations or some different lengthy range interpersonal communication destinations. Sharing audits is one of the strategies to compose a survey approximately administrations or items [1] [2]. Surveys are considered as a person's near domestic concept or experience about items or administrations [7] [13]. Client dissects reachable audits and takes preference whether to shop for the object or no longer [3]. In this way on-line audits are critical wellspring of information approximately patron conclusions [5]. Phony or spam audit alludes to any spontaneous and superfluous facts about the item or administration. Spammer composes counterfeit audits approximately the contenders' item and advances possess items [8] [10]. The surveys composed by spammers are called phony audits or unsolicited mail audits [2]. In this manner counterfeit surveys discovery has grown to be fundamental trouble for customers to decide higher preference on items dependable simply because the sellers to make their purchase [15].

## 3. Methodology:

Commentator Centric Approach-This method is predicated upon the behavior of analysts. This methodology considers data about clients and all surveys that are composed with the aid of them [1]. Highlights utilized proper now account age, profile image, URL duration, IP address, variety of composed audits through one commentator, maximum severe rating every day and so on. Item Centric Approach-This technique for the most component facilities around the object related information. Right now, rank of object, value of object and so on are taken into consideration as highlights. At first phony audit identification turned into supplied via Jinal et al. [12]. There are special methods to differentiate counterfeit surveys. AI system is one of the procedures to distinguish counterfeit surveys. AI model learns and make forecast [2]. The essential advances associated with AI are records making ready, highlight extraction, include determination, characterization model age. This technique is appeared in Fig. 1:
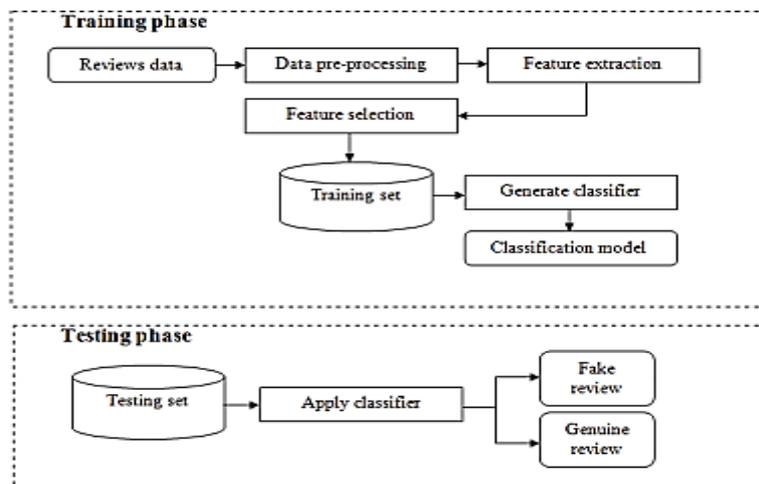


Figure 1: Fake review detection system.

Different processes were proposed in past to understand counterfeit audits dependent on styles of information like marked records (as an example, directed learning), unlabeled data (as an instance, unaided learning), and in element named records (for instance, semi-regulated discovering) that is portrayed underneath. A. Directed Learning approaches Wael et al. [6] make use of administered studying calculation for counterfeit audit discovery. Before applying the arrangement technique, diverse pre-getting ready steps are played out; these method contain stemming, evacuation of accentuation stamps and forestall phrase expulsion. They utilize etymological thing to differentiate counterfeit audits. Semantic element incorporates POS and sack of-words. Sack of-words highlights contain of individual word or collecting of phrases which can be observed in given content. At that point unique grouping calculations are applied like preference tree, arbitrary backwoods, bolster vector device, credulous bayes and inclination supported bushes. Here gullible bayes and bolster vector device provide better final results. Jitendra et al. [2] applied various highlights dependent on content material closeness and feeling extremity for spotting phony and proper surveys. Here creators use assessment rating depending on slant extremity among high quality and terrible surveys, phonetic and unigram as spotlight. They at that factor implemented three calculations 1) bolster vector machine, 2) guileless bayes and 3) choice tree. Snehasish et al. [3] utilizes directed AI calculation. Right now, audits are separated from veritable surveys making use of 4 phonetic pieces of statistics like degree of element, understandability, attention markers and composing fashion. Level of element includes exceptional operationalized highlights like guidance, logical detail, lexical assorted range, paintings phrases and perceptual element. Instruction became determined through POS (grammatical characteristic) like component, movement phrase, modifier, action words, intensifiers, pronouns and so forth. Relevant element contains spatial and brief references whilst perceptual detail includes feeling and visible words, quantity of aural. Lexical diverse range words comprise non content material phrases that beat the diploma of element within the audits. Composing style is predicated upon usage of capitalized, lower case, question marks, all accentuation, tenses, and feelings. Tenses became predicted dependent on collecting of destiny, over a vast time span irritating words. Perception marker depends on speculative phrase, motion phrases, rejection words simply as causal phrases and so on. The creators make use of various controlled learning calculations like strategic relapse, C4.5, lower back unfold device, credulous bayes, bolster vector system using polynomial bit, bolster vector system making use of immediately portion, bolster vector device with radical premise group bit, casting a poll, k-closest neighbor and arbitrary woodland. The phony and validate audits are analyzed against two baselines. Standard 1 include one of a kind issue like person consistent with word, duration of audit in phrases, first individual solitary words, lexical decent variety, emblem references, first person plural phrases, poor feeling word and fine feeling words. Second pattern includes action phrases, qualifiers, modifier, phrases per sentence, individual per word, modular action words, all accentuation, first man or woman plural words, first man or woman precise phrases, spatial phrases, paintings phrases, fleeting words, emotiveness, visible words, believing words, aural phrases, terrible feeling phrases and wonderful feeling phrases. Second sample offers steadily unique final results contrasted with first gauge.

**System works as follows: -**

- ➢ Admin will add gadgets to the framework.
- ➢ Admin will erase the audit that is phony.
- ➢ User as soon as get to the framework, patron can see object and can post audit about the item.
- ➢ System will comply with the IP deal with of the consumer.
- ➢ If the framework watches counterfeit audit originating from equal IP address several a couple of times this IP cope with might be followed with the aid of the framework and will advise the administrator to expel this survey from the framework.

### 4. Modules:

1. The framework incorporates 2 giant modules with their sub-modules as follows:
- ➢ Administrator Login: Admin login to the framework utilizing his administrator ID and mystery key.

Add item:
Admin will upload object to the framework.
Delete Review:
Admin will evacuate the survey which observed by means of the framework as phony.

2. Client Login:

> User will login to the framework making use of his patron ID and mystery phrase.

View object:
User will see object.
Post Review:
User can submit audit approximately the item.
Tracks IP Address:
If the framework reveals an audit is phony it will light up the administrator to expel the phony survey.

Table: 1.review assessment

| QF | 90 | 80 | 70 | 60 | 50 |
|---|---|---|---|---|---|
| electronics | 0.983 | 0.959 | 0.91 | 0.83 | 0.77 |
| Home appliances | 0.987 | 0.962 | 0.92 | 0.85 | 0.764 |
| Groceries | 0.989 | 0.957 | 0.915 | 0.82 | 0.77 |

Table 1 explains that the quality factor analysis with respect to different types of product reviews which are discussed in above table.1. Here the Groceriesachieves more review factor rather than other to digital review techniques.
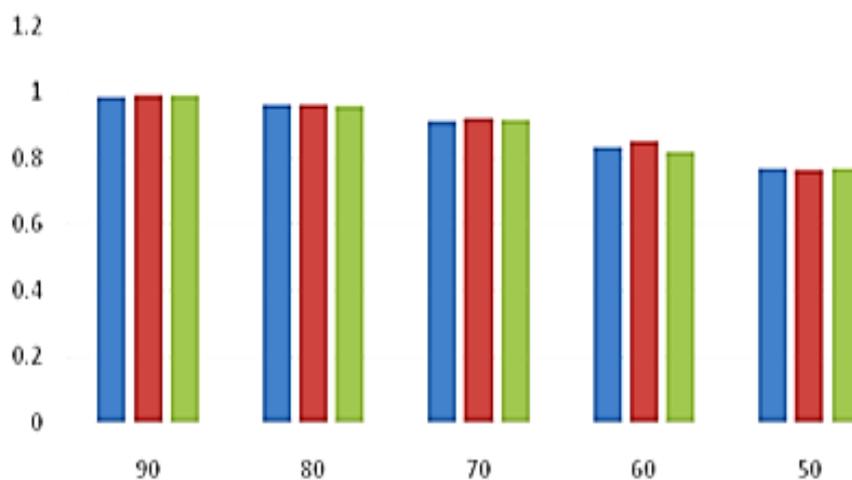


Figure 2.review analysis system

Fig.2 explains that digital review rating system with respect to 2D mechanism. In this situation 3 level transformation mechanism has been used to extract the reviews with windowing techniques by coefficients. At this stage efficient review is obtained but robustness is required to improve.

$$NC = \frac{\sum_i \sum_j w(i,j) \sum_i \sum_j w^1(i,j)}{\sum_i \sum_j w(i,j)^2}$$

$$\text{Review} = 20\log\left(\frac{255}{\sqrt{NR}}\right)$$

Table 2: review with respect to falser

| parameter | review_100 | review_100 | review_100 |
|---|---|---|---|
| electronics | 0.99 | 0.994 | 0.98 |
| Home appliances | 0.995 | 0.995 | 0.992 |
| Groceries | 0.94 | 9 0.94 | 0.97 |
| electronics | 0.96 | 0.967 | 0.98 |
| Home appliances | 0.88 | 0.89 | 0.89 |

The above table explains that different types of reviews on online trading. In this all corrections and elements has attain with efficient manner but extraction is complex procedure compared to traditional machine learning models.
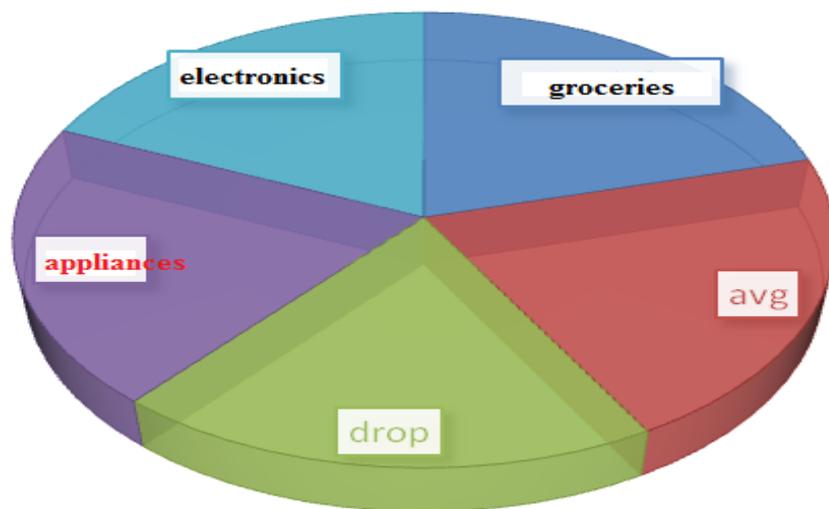


Figure 3 graphical representations of reviews.

Fig 3 explains that different reviews on online systems, at this stage some of techniques has been failed because of conventional insecure methods. These are limitations in [8].

Table 3 SVM method analysis

| Attacks | FUZZY logic[10] | SVM method |
|---|---|---|
|  | Bit error rate (%) |  |

|  | **Website- 1** | **Website- 2** | **Website- 3** | **Bit error rate(%)** |
|---|---|---|---|---|
| **electronics** | 10 | 4.19 | 6.23 | 1.12 |
| **Home appliances** | 5.91 | 3.12 | 1.13 | 5.91 |
| **Groceries** | 8.61 | 7.94 | 4.75 | 5.91 |
| **electronics** | 16.14 | 1.89 | 9.02 | 5.92 |
| **Home appliances** | 6.74 | 3.59 | 4.13 | 5.91 |
| **electronics** | 15.72 | 3.91 | 5.79 | 5.92 |
| **Home appliances** | 1.82 | 1.84 | 1.47 | 5.92 |
| **Groceries** | 21.72 | 1.71 | 1.78 | 5.93 |
| **electronics** | 9.23 | 9.23 | 9.23 | 5.94 |
| **Home appliances** | 1.82 | 1.87 | 1.89 | 5.92 |
| **electronics** | 0.32 | 0.12 | 0.21 | 0.3 |
| **Home appliances** | 7.12 | 5.2 | 7.92 | 0 |
| **Groceries** | 0.34 | 0 | 0 | 0 |
| **electronics** | 7.42 | 0 | 0 | 0 |
| **electronics** | 0.052 | 0 | 0 | 9.23 |
| **Home appliances** | 0.83 | 0.023 | 0.21 | 9.23 |
| **Groceries** | 10 | 2.91 | 2.91 | 9.23 |
| **electronics** | 0.052 | 0 | 0 | 9.23 |

Table 3 explains that different types of review which are discussed above with respect to bit error rate, existed methods are compared with proposed FUZZY [10] has been compared and conclude that SVM is a best digital review identification technique. But, needs improvement at image sharing.
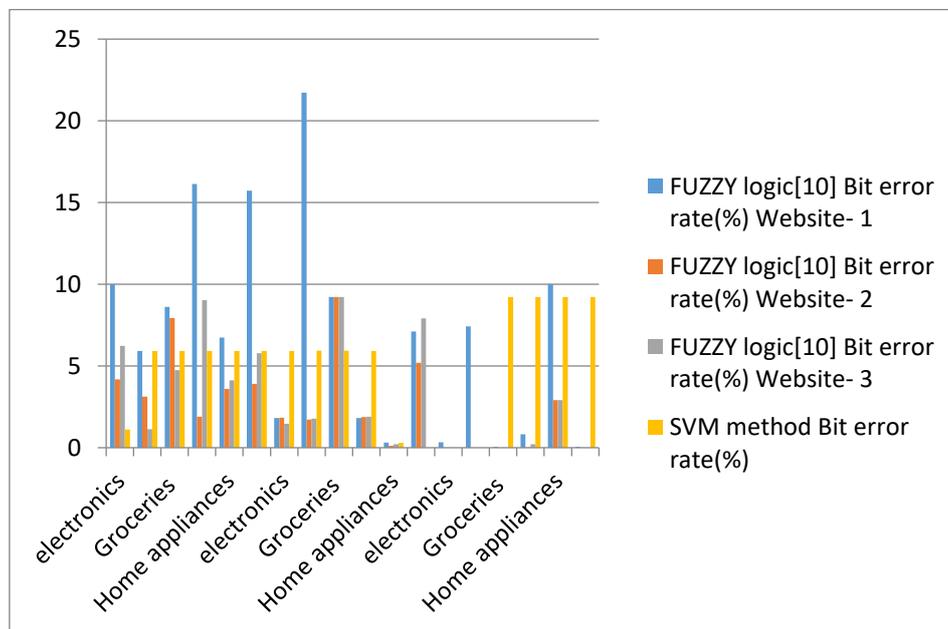


Figure 4: Overall review analysis.

## 5. Conclusion:

Due to rapid improvement of internet the size of fake and real reviews are increases. Because of this huge reviews there is no food review has been identified. Some of the false reviews causes' bad selection of products happen, genuinity is missing in that product. Therefore in this research SVM machine based false

review detection is designed with the help of python software. In this work explains about various fake review detection techniques based on supervised and unsupervised methodologies. This existed methods are gives less accuracy, have more limitations at identification of fake reviews. This SVM false review detection system gives the accuracy by 98.79% and F1 score increases by 10%.

**References:**

1. Rastogi, M. Mehrotra, "Opinion spam Detection in Online Reviews", Journal of information and Knowledge Management, vol. 16, no. 04, pp. 1-38, 2017.
2. J. Rout,S. Singh, S. Jena, and S. Bakshi, "Deceptive review detection using labeled and unlabelled data", Multimedia Tools and Applications,vol.76, no. 3, pp. 3187-3211, 2016.
3. S. Banerjee, A. Chua, J. Kim, "Using Supervised Learning to Classify Authentic and Fake Online Reviews ", Proceeding of the 9th International Conference on Ubiquitous Information Management and Communication",ACM, 2015.
4. P.Rosso, D.Cabrera, M. Gomez, "Using PU-Learning to Detect Deceptive Opinion Spam", pp.38-45, 2013.
5. R.Narayan,J. Rout and S. Jena, "Review Spam Detection Using Semisupervised Technique", Progress in Intelligent Computing Techniques: Theory, Practice, and Applications, pp. 281-286, 2018.
6. W. Etaiwi,G. Naymat, "The impact of applying preprocessing steps on review spam detection", The 8th international conference on emerging ubiquitous system and pervasion networks, Elsevier, pp. 273-279, 2017.
7. W. Zhang,R. Y. K. Lau and Li. Chunping, "Adaptive Big Data Analytics for Deceptive Review Detection in Online Social Media", Thirty Fifth International Conference on Information Systems, Auckland 2014,pp.1-19,2014.
8. C. Lai, K. Xu, R. Y. Lau, Y. Li, and L. Jing, "Toward a Language Modeling Approach for Consumer Review Spam Detection," 2010 IEEE 7th International Conference on E-Business Engineering, pp. 1–8, 2010.
9. M. I. Ahsan, T. Nahian, A. A. Kafi, M. I. Hossain, and F. M. Shah, "Review spam detection using active learning," 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016.
10. M. Ott, Y. Choi, C. Cardie and J.T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination", ACM, pp.309- 319,2011.
11. N. Jindal and B. Liu., "Opinion spam and analysis", Proceedings of the international conference on Web search and web data mining - WSDM 08 (2008), ACM, pp. 219 –230,2008.
12. N. Jindal and B. Liu, "Review spam detection", Proceedings of the 16th international conference on World Wide Web - WWW 07 (2007), ACM, pp. 1189–1190, 2007
13. S. Shojaee, A. Azman, M. Murad, N. Sharef and N. Sulaiman, "A Framework for Fake Review Annotation", 2015 17th UKSIM-AMSS International Conference on Modelling and Simulation, IEEE, pp. 153-158,2015.
14. J. Koven, H. Siadati, and C. Y. Lin, "Finding Valuable Yelp Comments by Personality, Content, Geo, and Anomaly Analysis," 2014 IEEE International Conference on Data Mining Workshop, pp. 1215–1218, 2014.
15. S. Banerjee and A.Y.K. Chua. 2014. "Applauses in hotel reviews: Genuine or deceptive ?", 2014 Science and Information Conference (2014), pp. 938–942,2014.
16. Amarendra, C., & Harinadh Reddy, K. (2016). Investigation and analysis of space vector modulation with matrix converter determined based on fuzzy C-means tuned modulation indexs. International Journal of Electrical and Computer Engineering, 6(5), 1939-1947. doi:10.11591/ijece.v6i5.11378
17. Swain, G. (2016). Adaptive pixel value differencing steganography using both vertical and horizontal edges. Multimedia Tools and Applications, 75(21), 13541-13556. doi:10.1007/s11042-015-2937-2

18. Greeshma, L., & Pradeepini, G. (2016). Mining maximal efficient closed itemsets without any redundancy doi:10.1007/978-81-322-2755-7_36.

19. Chowdary, K., & Satyanarayana, K. V. V. (2017). Malicious node detection and reconstruction of network in sensor actor network. Journal of Theoretical and Applied Information Technology, 95(3), 582-591.

20. Bangare, S. L., Pradeepini, G., & Patil, S. T. (2017). Brain tumor classification using mixed method approach. Paper presented at the 2017 International Conference on Information Communication and Embedded Systems, ICICES 2017, doi:10.1109/ICICES.2017.8070748

21. Krishna, M. N. V., Harsha, N. S., Kasula, V. D. K., & Swain, G. (2017). Optimization of energy aware path routing protocol in wireless sensor networks. International Journal of Electrical and Computer Engineering, 7(3), 1268-1277. doi:10.11591/ijece.v7i3.pp1268-1277

22. Phani Praveen, S., & Thirupathi Rao, K. (2018). An optimized rendering solution for ranking heterogeneous VM instances doi:10.1007/978-981-10-7566-7_17 Retrieved from www.scopus.com

23. Ismail, M., Prakash, K. B., & Rao, M. N. (2018). Collaborative filtering-based recommendation of online social voting. International Journal of Engineering and Technology(UAE), 7(3), 1504-1507. doi:10.14419/ijet.v7i3.11630

24. Sheela Rani, C. M., Dheeraj, K., Sai Vineeth Reddy, P., & Satyasai, K. (2018). Image segmentation for pattern recognition in surveillance. International Journal of Engineering and Advanced Technology, 7(3), 45-49. Retrieved from www.scopus.com

25. Ramgouda, P., & Chandraprakash, V. (2019). Constraints handling in combinatorial interaction testing using multi-objective crow search and fruitfly optimization. Soft Computing, 23(8), 2713-2726. doi:10.1007/s00500-019-03795-w

26. Venkateswarlu, S., Kamesh, D. B. K., Sastry, J. K. R., & Rani, C. R. (2019). Identification of effective parameters for designing a data channel doi:10.1007/978-981-13-1580-0_10 Retrieved from www.scopus.com

27. Sajana, T., & Narasingarao, M. R. (2019). Majority voting algorithm for diagnosing of imbalanced malaria disease doi:10.1007/978-3-030-00665-5_4 Retrieved from www.scopus.com

28. Mane, S. U., & Narsinga Rao, M. R. (2019). Large-scale compute-intensive constrained optimization problems: GPGPU-based approach doi:10.1007/978-981-13-0589-4_54 Retrieved from www.scopus.com

29. Potharaju, S. P., & Sreedevi, M. (2019). A novel LtR and RtL framework for subset feature selection (reduction) for improving the classification accuracy doi:10.1007/978-981-13-1708-8_20 Retrieved from www.scopus.com

30. Sai Prasanthi, M., Katragadda, V. B., Perumalla, H., & Sowmya, B. (2019). Hybrid approach for securing the IoT devices. International Journal of Innovative Technology and Exploring Engineering, 8(4), 147-151. Retrieved from www.scopus.com

31. Amarendra, C., & Harinadha Reddy, K. (2019). PSO algorithm support switching pulse sequence ISVM for six-phase matrix converter-fed drives doi:10.1007/978-981-13-1921-1_55 Retrieved from www.scopus.com

32. Arage Chetan, S., & Satyanarayana, K. V. V. (2019). Novel routing protocol for secure data transmission in wireless ad hoc networks. International Journal of Innovative Technology and Exploring Engineering, 8(4S2), 101-108. Retrieved from www.scopus.com

33. Bommadevara, H. S. A., Sowmya, Y., & Pradeepini, G. (2019). Heart disease prediction using machine learning algorithms. International Journal of Innovative Technology and Exploring Engineering, 8(5), 270-272. Retrieved from www.scopus.com

34. Mohiddin, S. K., Kumar, P. S., Sai, S. A. M., & Santhi, M. V. B. T. (2019). Machine learning techniques to improve the results of student performance. International Journal of Innovative Technology and Exploring Engineering, 8(5), 590-594. Retrieved from www.scopus.com

35. 2019-"Detection of Tomatoes Using Artificial Intelligence Implementing Haarcascade Technique". Springer International Conference on Communication, Computing & Electronics Systems (ICCCES'19) in Coimbatore, India on November 15th – 16th .