International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Combating Seller Fraud in E-Commerce: Insights from a Multiple Instance Learning Approach

Gaddam Kavyasri

Lead-Technology, Synechron Technologies Pvt Ltd, Taramani, Chennai-600113, Tamilnadu, India.

Abstract

In recent days with the tremendous increase of e-commerce web sites, almost all users are showing their interest in online shopping rather than visiting the shops directly for those items. Opinion mining is one of the important factors to gather the user's feedback on several items from e-commerce web sites. In a recent survey conducted by a well-known print media e-commerce is growing faster and it is up over ninety five percent compared in the primitive days. As we know that all customers always find ease to buy things online without spending more and more time to visit several shops for purchasing their interested items.There were a lot of criminals who try to create fraud activities in online by placing fake products and reviews in illegal ways. These illegal ways are giving a huge loss for the genuine customers while selecting their interested product from online. There is no pro-active mechanism which can guarantee the customers to detect and identify the fraud based on user's individual reviews and opinions for the products. So in this proposed thesis we try to develop a proactive model like multiple instance learning approach to detect and identify the service providersfraud based on individual user reviews and opinions for the products. We try to launch a graph mechanism in order to show the opinion about every individual product based on several users' feedback. Our experimental and theoretical analysis shows that this model can probably distinguish between primitive e-commerce sites and future expected e-commerce sites and extensively decreases customer complaints based on this trustability graph.

Keywords: Proactive Model, Service Providers, Multiple Instance Learning Approach, Fraud Detection, Opinion Mining, E-Commerce

1. INTRODUCTION

In the present day's World Wide Web (www) increased a lot of users' attention towards electronic commerce, commonly referred to as e-commerce for purchasing a wide variety of products in online. There are many websites like eBay ,Amazon, Flipkart,Snapdeal , allow web users to shop for and sell products and services online[1],[2]. Although customers that shop is enjoying the benefits of online trading; at an equivalent time traitors/attackers also are taking advantage to accomplishing deceptive activities against candid parties to get dishonest profit. Since the emergence of the www as in [3], electronic commerce, which is usually called e-commerce as in [4] become more and more popular in recent years. No often can we now consider taking a stroll through the Supermarket before buying a mobile handset, but healthy online research which in some cases is consequently followed by a web purchase. The scenario isn't limited to mobiles alone; it even covers a good range of products like home appliances, consumer electronic goods, books, apparel, travelling packages, etc. and even the electronic content itself. With e-Commerce as in [4] then, you'll buy almost anything you would like for without actually touching the merchandise physically and inquiring the salesperson for a variety of times before placing the ultimate order. In existing online shopping business model sellers as in [5] sell their products or services at a pre-set price, where buyers can choose what product the best suits them which is of an excellent deal.

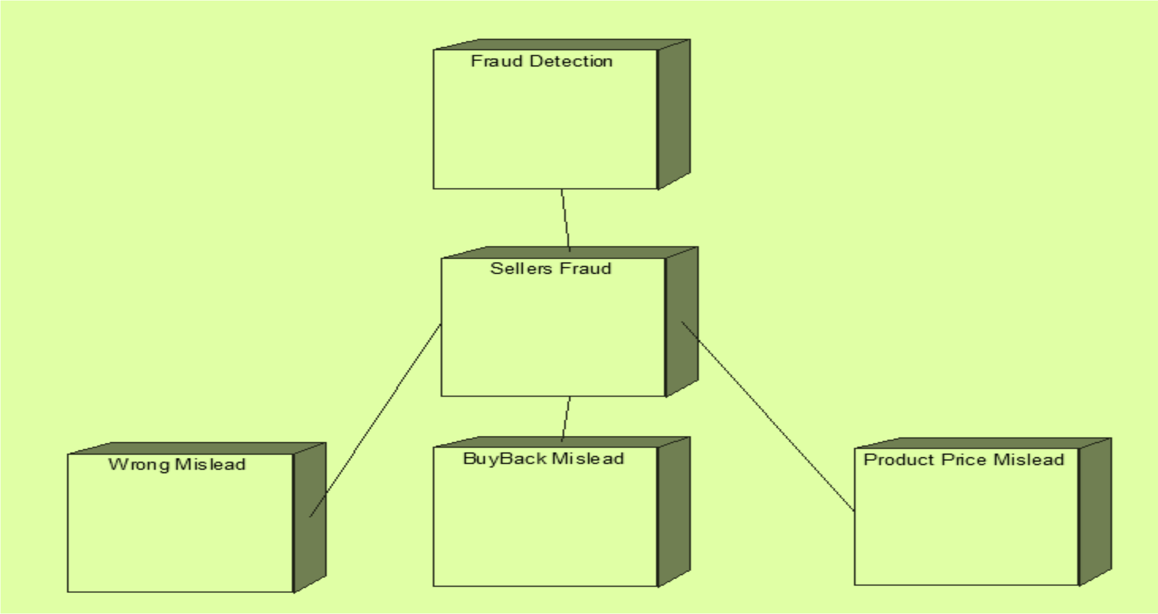


Figure. 1. Denote the Various Factors for Online Fraud Detection

In this section we will mainly discuss about the various factors for identifying the fraud in e-commerce web sites. In general the current e-commerce web sites try to identify the fraud service providers or sellers in reactive manner rather than in a pro-active manner. In this proposed thesis we try to find out the fraud in a pro-active manner based on the factors like placing wrong items or misleading the users with wrong products and another fraud is buy back option by attracting the users with various buy back features and finally the fraud which occur in e-commerce sites are product price mislead.

In this current thesis we try to extend these factors into a detailed manner for identifying fraud in a proactive manner [6].For extending the factors we try to analyse the application by studying hundreds and thousands of new customer’s complaints per day on the products. As this is mainly because of various levels of fraud that was done by the e-commerce websites. So we try to design an application where the fraud can be identified based on any of the levels like:

1) Not Delivered
2) Product Mismatch
3) Poor Service
4) Product Damaged

In this proposed thesis, the e-commerce user or customer can able to post the feedback or opinion for every purchased product in his account based on any of the 4 categories and once if the customer review is submitted to the system.These opinions are received to the admin in a detailed manner based on trustability graph and in turn the overall rating of the product will be automatically reduced based on the dynamic ratings given by the several customers.

2. BACKGROUND WORK

In this section, we try to define the backgroundwork that is carried out in order to propose this current application.

Preliminary Information:

Here we try to analyze the main factors that are required for identifying the fraud in online shopping or e-commerce web sites. They are as follows:

Initially the data pre-processing step need to be performed on the input data. Here the input data is nothing but a sample web site containing several products and opinions collected from e-customers. The pre-processing step can be split into two main phases-

1) Data Log Cleaning

The main objective of data log cleaning is to separate the desired records and noise records into two separate lists. Here we try to consider only the records which are desired for identifying the fraud sellers and avoid false records that may distort the results of the analysis.

2) User Authentication and Pattern Identification

In this phase the users will be authenticated by verifying their login details with registered database. Those who are valid can able to enter into their account and search for products which are available in that web site. If any user who don’t have valid login credentials cannot able to enter into the site and can search for products. This process is main step for calculating the count of positive opinions and negative opinions at the time of finding the feedback about the appropriate product.

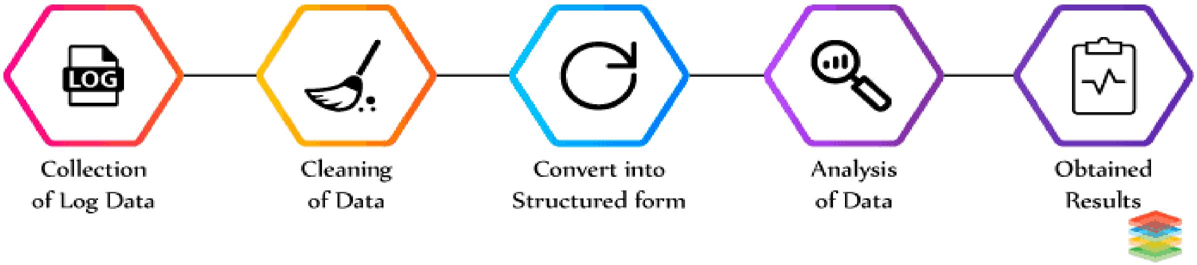


Figure. 2. Denote the Flow of Data Pre-Processing for Online Fraud Detection

From the above figure 2, we can clearly identify initially the system takes collection of log data from e-commerce web sites and try to clean the data into separate lists as we discussed in previous paragraph. Once the data cleaning is completed, now the data is converted into structured manner which is nothing but data is mainly divided into clustered manner. Here the clusters are nothing but 2 types of clusters, one is positive opinion cluster and another one is negative opinion cluster. So based on these two cluster data, the user can analyse the data records and obtain the result.

In this same stage , there is another topic like Pattern Identificationin detecting the most relevant and used parts of the website and the relationship existing between them, as well as identifying behavioural patterns related to the buying process[7]-[11]. This is as follows:

(a) Customer Usage and Navigational Pattern Identification:

In this stage we are going to find out all navigational patterns that illustrate customer's preferences while surfing the e-commerce web sites. Generally every individual customer has own interest in surfing the web site for various products he wish.Here we try to analyse all possible factors like which products are searched in first preference and which type of products are surfed in next preferences[12]-[18].

(b) Customer Behaviour Pattern To Identify the Type of Purchased Product :

In this stage we are going to find out the customer behaviour to identify which type of product he want to purchase from a set of products which are available. Regarding the buying process there are two specifications that we are interested in- First, user sessions showing interest in acquiring a specific product that corresponds to the events of adding a product to the cart and adding a product to the wish list. Here we need to identify one important factor like how many users are liked that appropriate product and also find out count of individual user who purchased the same product for how many times. Here the customer behaviour pattern will try to find out the individual product status in terms of positive count as well as negative count per individual customer[19]-[24].

3. OVER VIEW OF MULTIPLE INSTANCES LEARNING APPROACH

In this section we mainly try to demonstrate the overview of multiple instance learning approach for identifying the seller fraud based on customer'sbehaviour. Now let us discuss about that in detail as follows:

Motivation

Here we try to usemultiple instance learning approach for finding the fraudulent service providers in the E-commerce websites. For identifying the fraudulent sellers from a list of several sellers, we try to take a labeling method like bagged fashion as in [12] i.e. For each and every individual product a new labeling process starts, the system pick every seller in the queue and identify all his products in that current session[25]-[27]. If any product is found with any negative review or opinion posted by the customer, then he will try to separate that seller product in second cluster. If there are no single negative opinion for the appropriate products then they will be placed in first cluster.If any of the cases had been found to be fraud, then all the cases from this seller are labeled as fraud. In these types of scenarios they are to be handled by "multiple instance learning" as in [2].

Here the term multiple instance learning is nothing but

System 'S' will try to find out the instance of an product 'Y' based on number of observations or cases like 'K'.

Suppose for each seller **i** at time **t** there are $\mathbf{K_{it}}$ number of cases.

For all the K_{it} cases the labels should be identical, hence can be denoted as y_{it} .

For Pro-Active link function, a data variable for l-th case of selleri as z_{ilt} .

The multiple instance learning(MIL) model can be written as

$$y_{it} = 0 \text{iff} z_{ilt} < 0, \forall l = 1, \cdot \cdot \cdot , K_{it}$$

Otherwise

$$y_{it} = 1, \text{and } z_{ilt} \sim N\ (x'_{ilt}\alpha t, 1);$$

Here the trustability graph takes the input individually for all products ‘Y’ I.eYit based on time ‘t’.Here the samples are identified by the conditional posterior of βt and finally the process of sampling is shown identical based on the parameters like π(zt|yt, xt, βt) is different.

4. SEQUENCE TREE ALGORITHM FOR FINDING FREQUENT SEQUENCES IN E-COMMERCE SITES

In this section we try to find out the sequential tree algorithm in order to find out the frequent sequences in the log file. This is mainly having two stages like

a) Construction of the Sequence Tree based on reputation of Products and

b) Finding Most Frequent Item Sequences.

Input:

Initially this algorithm will takes input as sequential database containing various sequences pages visited by the user and also threshold [13]-[16].

Condition:

Find out the most frequent item in two sets like which are mostly supported by the customers as one set and which are having negative opinions as another set.

Output:

The frequent item sets will be identified from the log history collected from various customers opinions based on the min threshold values.

Procedure:

The procedure for sequence tree algorithm is divided into various stages as follows:

Stage 1:

Initially try to extract the input from sequence database which contains a set of products and their customer opinions. Here there will be a factor like ‘Key’, which has two parameters like:
Page name and Frequency of page.

Stage 2:

Map the items or products based on the descending order based on the Key Parameters with main attribute like frequency of the page.

Stage 3:

Try to construct a sequencetree with all the parameters like

Try to find out the top rated product from the database and mark that as root node. Initially the root node is termed as null

Now find out the next highest product and add that as branch node for that root node.

Verify whether this item is already repeated or it is the first one. If it is already available try to increment the counter variable by 1.

Else

A new branch is constructed and added to the previous branch.
Try to verify until all branches are added in a tree manner.
Exit the Procedure

Stage 4:

Construct a header table (HT) with a set of rows and columns.

Where a row contains a list of nodes or items present inside the sequence database.

Column contains the attributes that are present in that sequence database.
Here these parameters are mainly used in constructing the trustability preview graph for all individual products.

For each and every individual product, the customer will get the trustability chart separately and the main parameter for this trustability chart is frequency of the product.

Stage 5:

Start the Mining Process

```
For  each individual row in the HT
Find out the condition like
If (frequency < minimum threshold value)
Ignore that row from the HT
    Else
        Find the other nodes present in the  particular linked list
    If  (frequency < minimum support value)
        Ignore the total path
    Else
      Tree is traversed up till we reach the root node and store the details in the HT.
    EndFor;
    EndFor
```

Stage 6:

Result is displayed in terms of two separate lists

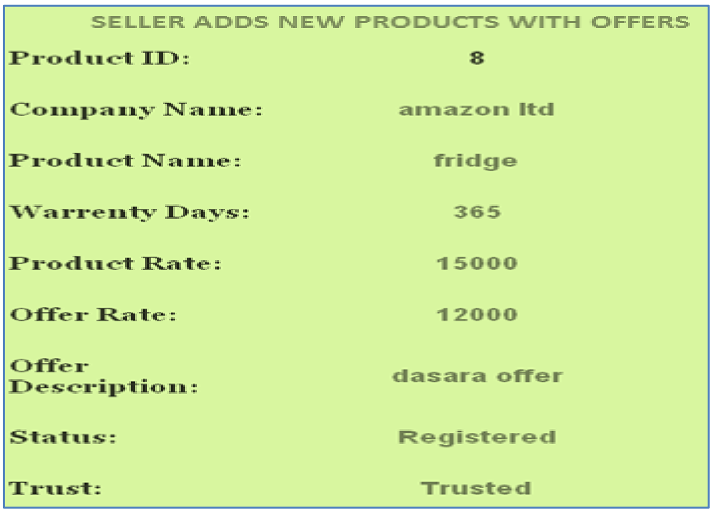
One is Finding the Most Frequent Items Supported by Max Customers

Another is Finding the Frequent Items Opposed by number of Customers

5. EXPERIMENTAL RESULTS

We have conducted experiment on a sample E-commerce web sites by gathering various products from a well-known shopping sites. Here we designed the E-commerce web site using Java Platform using JEE as development environment. We tested the application by registering multiple customers and try to login all customers and give comments for the products individually based on the pre-defined four categories. Here we try to use trustability graph for every individual graph which clearly denotes the status about that product in a pro-active manner.

Seller tries to add all new products in the Database

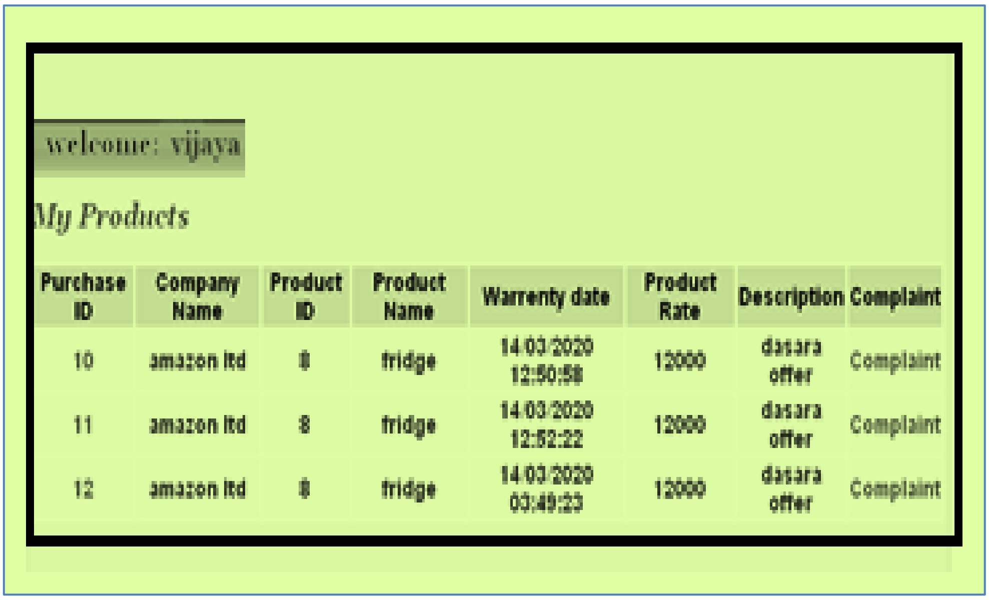


SELLER ADDS NEW PRODUCTS WITH OFFERS

Product ID:	8
Company Name:	amazon ltd
Product Name:	fridge
Warrenty Days:	365
Product Rate:	15000
Offer Rate:	12000
Offer Description:	dasara offer
Status:	Registered
Trust:	Trusted

From the above figure, we can clearly identify that seller will try to add a set of products into the database. Here the products will be added based on product description, name, warranty days and other fields along with offer details. Here we can see the final attribute like trusted which is given by the company admin.

Customer can view list of products purchased from the Product Database

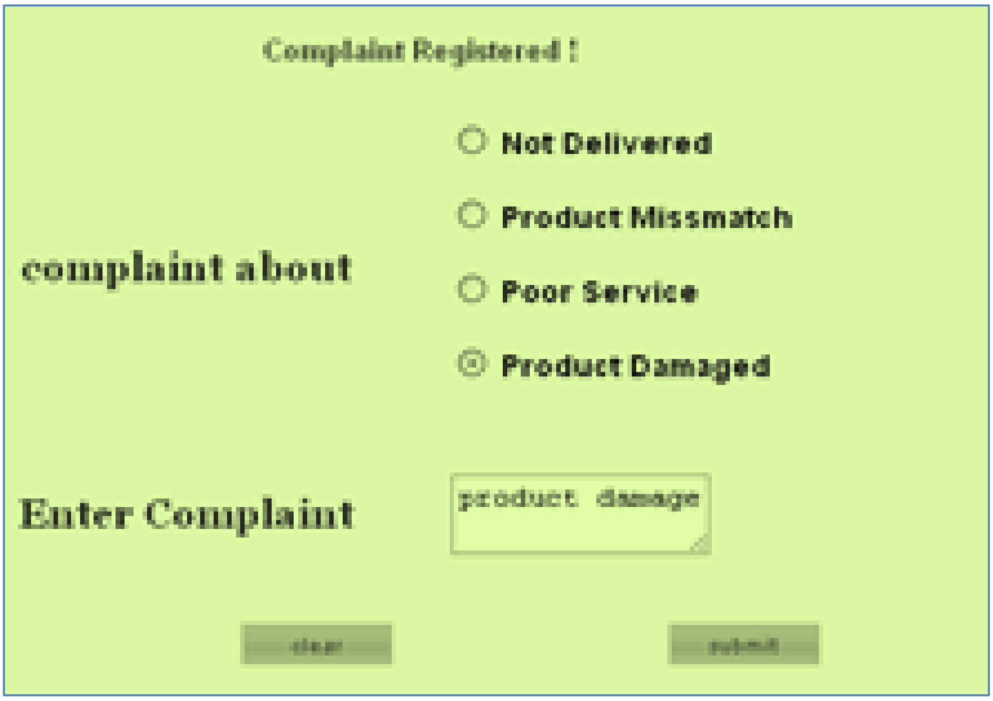


welcome: vijaya

My Products

Purchase ID	Company Name	Product ID	Product Name	Warrenty date	Product Rate	Description	Complaint
10	amazon ltd	8	fridge	14/03/2020 12:50:58	12000	dasara offer	Complaint
11	amazon ltd	8	fridge	14/03/2020 12:52:22	12000	dasara offer	Complaint
12	amazon ltd	8	fridge	14/03/2020 03:49:23	12000	dasara offer	Complaint

From the above figure we can able to find out all the products which are purchased by the customer. Here the complaint can be given if the customer really faces any problem with that purchased product. The complaint can be any one of the factor, which is seen below:



Complaint Registered !

complaint about

○ Not Delivered

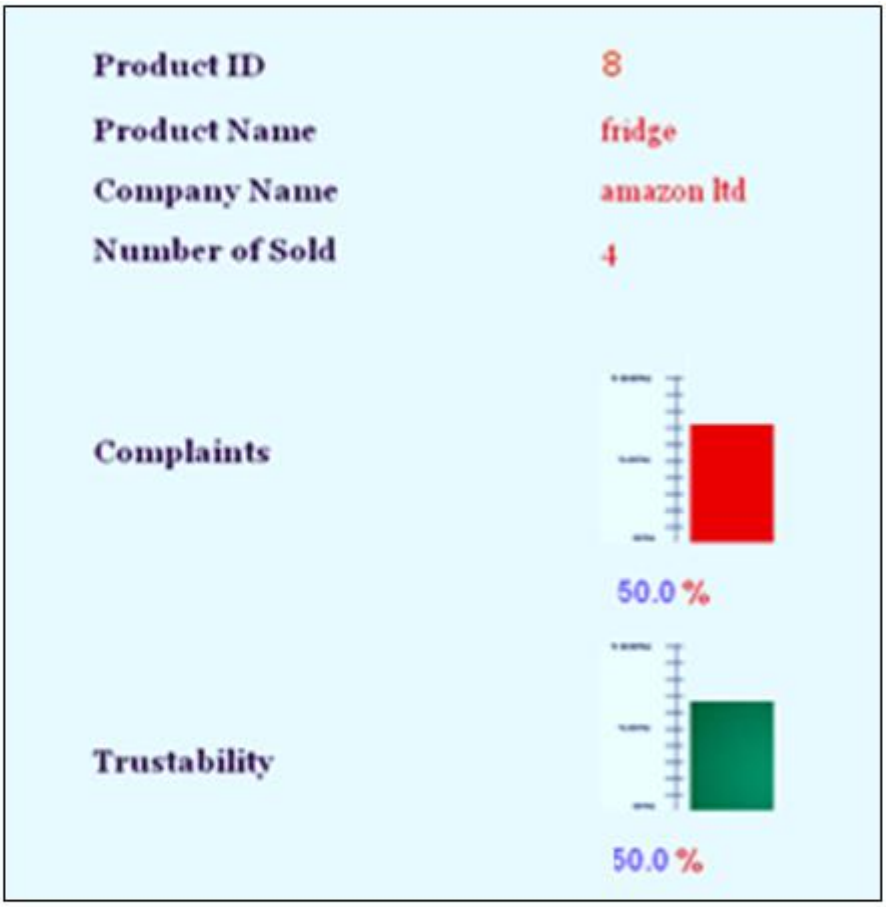
○ Product Missmatch

○ Poor Service

● Product Damaged

Enter Complaint: product damage

Customer can able to find out the Fraud Seller based on this Dynamic Opinions



Product ID	8
Product Name	fridge
Company Name	amazon ltd
Number of Sold	4
Complaints	50.0 %
Trustability	50.0 %

From the above window we can able to get a clear idea about the fraud seller with the help of customeropinions. Here we can see a product name like fridge under the company name Amazon ltd is having 50 percent trustability and remaining 50 percent it is having some different complaints posted by several customers. Hence by observing the trustability chart we can able to identify the fraud sellers or service providers and can distinguish accurately which product is genuine and which product is fake.

6. CONCLUSION

In this paper, we for the first time have developed a proactive model like multiple instance learning approach to detect and identify the service provider's fraud based on individual user reviews and opinions for the products. We try to launch a graph mechanism in order to show the opinion about every individual product based on several customer's feedback. By conducting various experiments on our current approach the theoretical and experimental evaluation has been validated and can probably distinguish between primitive e-commerce sites and future expected e-commerce sites and extensively decreases customer complaints based on this trustability graph. Here by implementing trustability graph we can able to identify the genuine suppliers and fake suppliers accurately based on graph prediction.

REFERENCES

[1]. Li, Y., yu Chen, C., and Wasserman, W. W. (2015). Deep feature selection: Theory and application to identify enhancers and promoters. JOURNAL OF COMPUTATIONAL BIOLOGY, pages 1-15.

[2]. Marbach, D., Scha_ter, T., Mattiussi, C., and Floreano, D. (2009). Generating realistic in silico gene networks for performance assessment of reverse engineering methods. Journal of Computational Biology, 16(2):229-239.

[3]. Montavon, G., Lapuschkin, S., Binder, A., Samek, W., and M• uller, K.-R. (2017). Explaining nonlinear classi_cation decisions with deep taylor decomposition. Pattern Recognition, 65:211-222.

[4]. R.S.M. Lakshmi Patibandla, B. Tarakeswara Rao, P. Sandhya Krishna, Venkata Rao Maddumala,(2020),"MEDICAL DATA CLUSTERING USING PARTICLE SWARM OPTIMIZATION METHOD", Journal of Critical Reviews, ISSN- 2394-5125, Vol 7, Issue 6, 2020, Page No. 363-367.

[5]. Patibandla R.S.M.L., Veeranjaneyulu N. (2018), "Survey on Clustering Algorithms for Unstructured Data". In: Bhateja V., CoelloCoello C., Satapathy S., Pattnaik P. (eds) Intelligent Engineering Informatics. Advances in Intelligent Systems and Computing, vol 695. Springer, Singapore

[6]. A.Naresh, R S M Lakshmi Patibandla, VidhyaLakshmi, M.MeghanaChowdary.(2020). "Unsupervised Text Classification for Heart Disease Using Machine Learning Methods", Test Engineering and Management, ISSN: 0193-4120 Page No. 11005 – 11016.

[7]. Tarakeswara Rao B., Patibandla R.S.M.L., Murty M.R. (2020) A Comparative Study on Effective Approaches for Unsupervised Statistical Machine Translation. In: Bhateja V., Satapathy S., Satori H. (eds) Embedded Systems and Artificial Intelligence. Advances in Intelligent Systems and Computing, vol 1076. Springer, Singapore.

[8]. RSM Lakshmi Patibandla. (2020), “Regularization of Graphs: Sentiment Classification”, Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural and Other Industries, John Wiley & Sons, pp:373.

[9]. Murugan, R., Paliwal, M., Patibandla, R.S.M.L., Shah, P., Balaga, T.R., Gurrammagari, D.R., Singaravelu, P., (...), Jhaveri, R. Amalgamation of Transfer Learning and Explainable AI for Internet of Medical Things (2024) Recent Advances in Computer Science and Communications, 17 (4), art. no. e191223224674, pp. 40-53. doi: 10.2174/0126662558285074231120063921

[10]. X. Xu, R. S. M. Lakshmi Patibandla, A. Arora, M. Al-Razgan, E. MahrousAwwad and V. OmolloNyangaresi, "An Adaptive Hybrid (1D-2D) Convolution-Based ShuffleNetV2 Mechanism for Irrigation Levels Prediction in Agricultural Fields With Smart IoTs," in IEEE Access, vol. 12, pp. 71901-71918, 2024, doi: 10.1109/ACCESS.2024.3384473.

[11]. S. Bhatnagar et al., "Efficient Logistics Solutions for E-Commerce Using Wireless Sensor Networks," in IEEE Transactions on Consumer Electronics, doi: 10.1109/TCE.2024.3375748.

[12]. Krishna, P. S., Reddy, U. J., Patibandla, R. L., &Khadherbhi, S. R. (2020). Identification of lung cancer stages using efficient machine learning framework. Journal of Critical Reviews, 7(6), 385-390.

[13]. BanavathuMounika, S. R. K., Maddumala, V. R., &Patibandla, R. L. (2020). Data distribution method with text extraction from big data. Journal of Critical Reviews, 7(6), 376-380. Patibandla, R. S. M., &Veeranjaneyulu, N. (2022). A SimRank based ensemble method for resolving challenges of partition clustering methods. Journal of Scientific & Industrial Research, 79(4), 323-327.

[14]. Patibandla, R.S.M.L., Rao, B.T., Rao, D.M., Ramakrishna Murthy, M. (2024). Reshaping the Future of Learning Disabilities in Higher Education with AI. In: Kaluri, R., Mahmud, M., Gadekallu, T.R., Rajput, D.S., Lakshmanna, K. (eds) Applied Assistive Technologies and Informatics for Students with Disabilities. Applied Intelligence and Informatics. Springer, Singapore. https://doi.org/10.1007/978-981-97-0914-4_2

[15]. Gadde, S., Rao, G. S., Veesam, V. S., Yarlagadda, M., &Patibandla, R. S. M. (2023). Secure Data Sharing in Cloud Computing: A Comprehensive Survey of Two-Factor Authentication and Cryptographic Solutions. Ingénierie des Systèmesd'Information, 28(6).

[16]. Patibandla, R. L., Rao, B. T., Murthy, M. R., &Bhuyan, H. K. (2024). Xai-based autoimmune disorders detection using transfer learning. In Machine Learning in Healthcare and Security (pp. 119-129). CRC Press.

[17]. Lakshman Narayana, V., Lakshmi Patibandla, R.S.M., Pavani, V., Radhika, P. (2024). OptimiertenaturinspirierteRechenalgorithmenzurErkennung von Lungenerkrankungen. In: Raza, K. (eds) Von der NaturinspirierteintelligenteDatenverarbeitungstechniken in der Bioinformatik. Springer, Singapore. https://doi.org/10.1007/978-981-99-7808-3_6

[18]. Narayana, V. L., M. Lakshmi Patibandla, R. S., Rao, B. T., &Gopi, A. P. (2022). Use of Machine Learning in Healthcare. Advanced Healthcare Systems: Empowering Physicians WithIoT-Enabled Technologies, 275-293. https://doi.org/10.1002/9781119769293.ch13

[19]. Lakshmi Patibandla, R.S.M., Yaswanth, A., Hussani, S.I. (2023). Water-Body Segmentation from Remote Sensing Satellite Images Utilizing Hierarchical and Contour-Based Multi-Scale Features. In: Marriwala, N., Tripathi, C., Jain, S., Kumar, D. (eds) Mobile Radio

Communications and 5G Networks. Lecture Notes in Networks and Systems, vol 588. Springer, Singapore. https://doi.org/10.1007/978-981-19-7982-8_21

[20]. Patibandla, R. L., Narayana, V. L., &Gopi, A. P. (2021). Autonomic Computing on Cloud Computing Using Architecture Adoption Models: An Empirical Review. Autonomic Computing in Cloud Resource Management in Industry 4.0, 195-212.

[21]. Lakshman Narayana, V., Rao, G.S., Gopi, A.P., Lakshmi Patibandla, R.S.M. (2022). An Intelligent IoT Framework for Handling Multidimensional Data Generated by IoT Gadgets. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. https://doi.org/10.1007/978-3-030-80928-7_9

[22]. Patibandla, R. L., Rao, B. T., Krishna, P. S., &Maddumala, V. R. (2020). Medical data clustering using particle swarm optimization method. Journal of Critical Reviews, 7(6), 363-367.

[23]. R. S. M. Lakshmi Patibandla, V. S. Srinivas, S. N. Mohanty and C. RanjanPattanaik, "Automatic Machine Learning: An Exploratory Review," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2021, pp. 1-9, doi: 10.1109/ICRITO51393.2021.9596483.

[24]. Narayana, V.L., Gopi, A.P., Patibandla, R.S.M. (2021). An Efficient Methodology for Avoiding Threats in Smart Homes with Low Power Consumption in IoT Environment Using Blockchain Technology. In: Choudhury, T., Khanna, A., Toe, T.T., Khurana, M., GiaNhu, N. (eds) Blockchain Applications in IoT Ecosystem. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-65691-1_16

[25]. Patibandla, R.S.M.L., Narayana, V.L., Mohanty, S.N. (2021). Need of Improving the Emotional Intelligence of Employees in an Organization for Better Outcomes. In: NandanMohanty, S. (eds) Decision Making And Problem Solving. Springer, Cham. https://doi.org/10.1007/978-3-030-66869-3_5

[26]. Lakshmi Patibandla R.S.M., Aienala, Lavanya, Alla, Hemasai Sri (2022). Rainfall Extrapolation through Machine Learning, Workshop on Advances in Computational Intelligence, its Concepts and Applications, ACI 2022, Volume 3283, ISSN: 1613-0073 ,pp: 307 – 315,