International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 07 | July -2017 www.irjet.net p-ISSN: 2395-0072

ONLINE FRAUD DETECTION- A REVIEW

Mini Singh Ahuja[1], Lovepreet Singh[2]

[1]*Assistant Professor, Dept. Of Computer Science & Engineering, GNDU RC Gurdaspur, Punjab, India*
[2]*Student Mtech (CSE), Dept. Of Computer Science & Engineering, GNDU RC Gurdaspur, Punjab, India*

Abstract - *Lack of time with community is indulging multiple users to participate in online social media for communication. Users can interact with each other through this peace of technology. Ecommerce websites also become popular since users does not have to visit actual stores. As user's increases, so do frauds. Detection of frauds is the prime objective of this literature. In order to accomplish this KNN and Euclidean distance mechanism is hybridized. Comparative analysis is present against KNN, Euclidean distance and hybrid approach. Results are expressed in terms of time consumption and number of fault detected.*

Key Words: Trust, Online Deception, Types of Deceptions, Techniques to Find Deception.

1. INTRODUCTION

1.1 THE CONCEPT OF TRUST

Trust is a difficult concept to understand since it is defined in legion of ways. There are reams of areas in which trust is required. In every area trust is defined conflictingly. There are various views associated with the trust. Certain elements appear to be common across many views. Modern societies will not be possible beyond trust. To build trust most researchers agree that experience is critical and information about past behavior, goals and reputation are required. Additionally, trust is commonly viewed as a dynamic process which must be building over time and is dependent on situational context. Some sociologists also believe that trust cannot exist in the absence of risk. For psychological perspective, trust research tends to focus on individual personality difference and development throughout life or interpersonal relationships. As with the sociological view, the body of trust research in psychological is fragmented and varied across situations and is multidimensional.

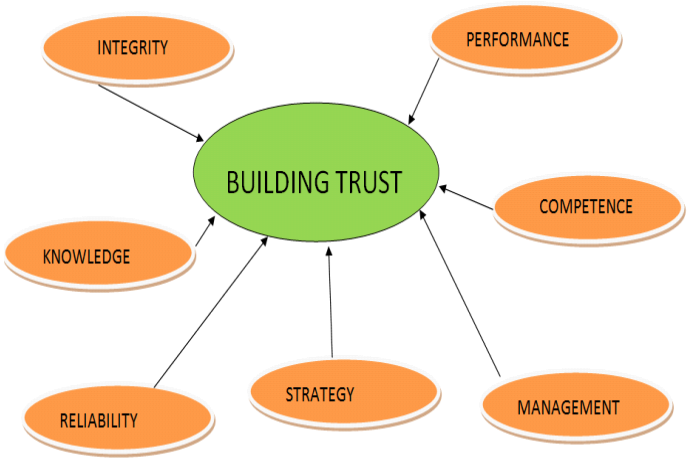

Figure 1 Shows Trust Building Factors

As with other disciplines, business related literature presents several views and definitions of trust. Trust may be defined as the willingness to depend on the exchanging partner in whom one has confidence. In the proposed paper we adopt this definition to create a model of trust. Trust in business to consumer (B2C) is established very differently than in business to business (B2B) e-commerce environment. In this context, the distinguishment is made between the hard and soft trust. Hard trust is associated with the issues of security and soft trust is associated with the privacy and quality of the service dimension. Issues associated with the soft trust cannot be easily resolved. The personal information which is gathered about the consumer is required to be secured. The trust of the consumer on vendor will be established if the information gather about the consumer is not shared among the unauthorized channels. Consumer need to trust that their Personal information will not be abused by the vendor or even sold to the highest direct marketing bidders. Concerns about the information practices will lead to consumers guarding their personal information or falsifying information, which deprives the online vendor of valuable information that could be used to tailor their services to individual customers. Trusting the quality of an online vendor's service may also be a challenge for

consumers. After all, the store down the block will likely be there tomorrow, but the store that exists in cyberspace is often not real in the customers concern. The other means of gathering trust is the use of recommender system. The system which satisfies the conditions or constraints must be promoted by the recommender system.

1.1.1 PARTIES INVOLVED IN THE ECOMMERCE TRUST

In the business to consumer model there are three main parties which are involved:

CONSUMER

This is the main party involved in the Trust Model. Consumer will require trust before interacting with the vendor. They have experience in the offline market but may not have experience in the online market. Individual customers will differ in the level of trust they need before performing the operations.

VENDORS

Vendors are those who require trust among the consumer in order to gather profit. The trust in the vendor will cause the consumer to buy goods from the corresponding vendor. They will have both physical and online presence. The vendors with the physical presence seem to be more successful than the vendor with the virtual presence.

REFEREE

Referees are third parties who provide independent recommendations on the trustworthiness of vendors. Trust referee may come in many forms, from individual Recommendations to privacy and security trust seals to media representatives. In particular, this paper focuses on trust as a mediating factor in building online consumer trust.

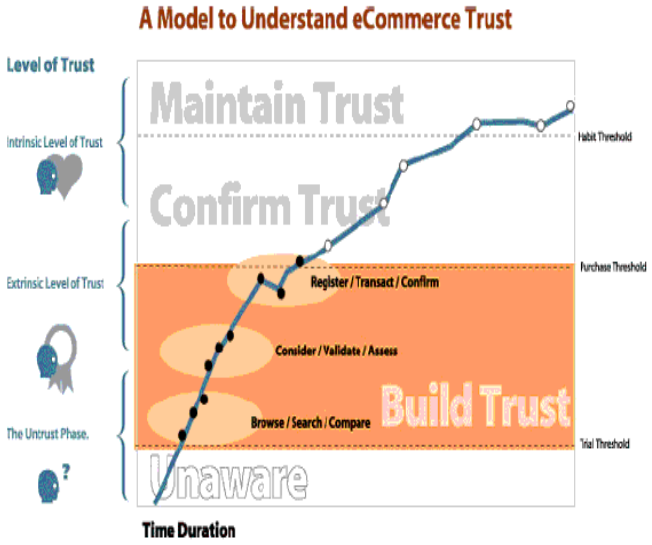

Figure 2 Shows a Model to Understand Ecommerce Trust.

2. ONLINE DECEPTION (FRAUD)

[1, 4, 5]The deception will cause falsifying information to be transmitted from source to the destination. Deception can be at the smaller scale or at the large level. Deception can cause damages both at physical and mental level. With the advent of the technology large number of users is using the internet. There are number of social networking sites which user use in order to interact with each other. They share their thoughts, feelings, experience etc. some information which they share may be sensitive in nature. Also there is some private information which is presented over the social media. Deception which takes place over the internet is under the category of Online Deception.

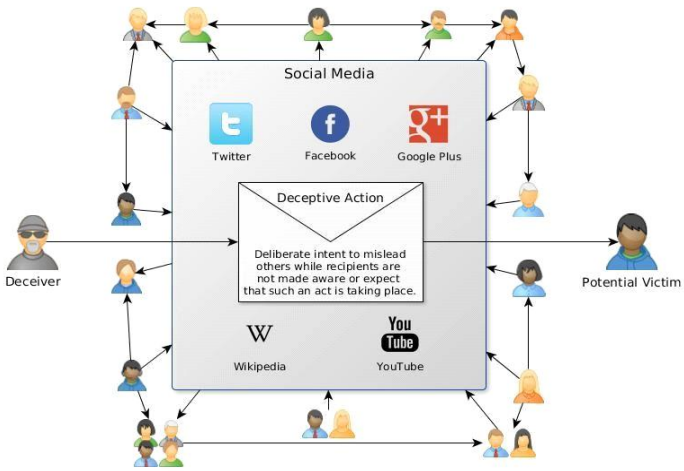

Figure 3 Entities involved in online social media.

3. VARIOUS DECEPTIONS (FRAUD)

Criminals use various numbers of ways to commit fraud and their imaginations in finding new ways to gain. Some of the most commonly known types of fraud include:

1. **Check Fraud**- Check fraud occurs when a person pays for something with a check knowing that there is not enough money in the account to cover the cost, or when an individual forges a check stolen from someone else.
2. **Internet sales**- Internet frauds are becoming more prevalent since the world relies heavily on technology. This type of fraud involves selling fake or counterfeit items, or taking payment with no intention to ship or deliver the item.
3. **Website misdirectio**n- This occurs when hackers mimic reputable companies such as Amazon, eBay, or PayPal, redirecting consumers to another website where they enter their credit card information. The criminal then use this information to make personal purchases.
4. **Charities fraud**- For many years, criminals have taken advantage of the fact that Americans generously give to worthy causes. Criminals solicit people to make donations to various causes that do not actually exist.
5. **Work-from-home scams-** Working from home sounds like a dream to many people, so it is not surprising that a number of Americans fall for this type of fraud each year. Criminals promise income to people who sign up for their fictitious work-from-home job, often requiring that money be paid up front with the promise of a big payoff in a short amount of time.
6. **Pyramid schemes**- These work-from-home type schemes promise the consumer large returns on their investment if they are able to recruit others to their network. People often pay money up front to buy into the business, or to buy a sales package of some type, and only make money if they get a large number of people to join beneath them.
7. **Identity theft-** One of the most commonly perpetrated types of fraud in this age, identity theft robs victims of their money, credit rating, and personal identity. Fraudsters obtain credit card, bank account, and other personal information, using them for personal gain.
8. **Credit card fraud-** In today's climate of electronic money, credit card fraud has become a prevalent crime. By obtaining people's credit card information, through a variety of means, the perpetrator can quickly make a large amount of purchases before the consumer even realizes what is happening. Credit card fraud is classified as identity theft, identity assumption, or a fraud spree, depending on the specifics of the crime.
9. **Debt elimination-** Many Americas find themselves deeply in debt, making it easy for criminals to offer them an opportunity to climb out from under a mountain of bills. Fake companies produce ads and other solicitations promising to help eliminate every type of debt, from credit card bills to taxes, for a partial payment up front. The victim fronts the payment as well as their credit card information, getting nothing in return, and often having their information sold to other fraudsters.
10. **Insurance fraud-** Insurance fraud is committed every day in the U.S., by people who otherwise would not consider them to be criminals. False or inflated insurance claims for automobile damages, health care expenses, and homeowners or renters insurance are considered to be insurance fraud, and may be charged as felonies, depending on the circumstances and amount of the fraud.
11. **Unusual Behavior-** The perpetrator will often display unusual behavior, that when taken as a whole is a strong indicator of fraud. The fraudster may not ever take a vacation or call in sick in fear of being caught. He or she may not assign out work even when overloaded. Other symptoms may be changes in behavior such as increased drinking, smoking, defensiveness, and unusual irritability and suspiciousness.
12. **Complaints-** Frequently tips or complaints will be received which indicate that a fraudulent action is going on. Complaints have been known to be some of the best sources of fraud and should be taken seriously. Although all too often, the motives of the complainant may be suspect, the allegations usually have merit that warrant further investigation.
13. **Stale Items in Reconciliations-** In bank reconciliations, deposits or checks not included in the reconciliation could be indicative of theft. Missing deposits could mean the perpetrator absconded with the funds; missing checks could indicate one made out to a bogus payee.
14. **Excessive Voids-** Voided sales slips could mean that the sale was rung up, the payment diverted to the use of the perpetrator, and the sales slip subsequently voided to cover the theft.
15. **Missing Documents-** Documents which are unable to be located can be a red flag for fraud. Although it is expected that some documents will be misplaced, the auditor should look for explanations as to why the documents are missing, and what steps were taken to locate the requested items. All too often, the auditors will select an alternate item or allow the audited to

select an alternate without determining whether or not a problem exists.

16. **Excessive Credit Memos-** Similar to excessive voids, this technique can be used to cover the theft of cash. A credit memo to a phony customer is written out, and the cash is taken to make total cash balance.
17. **Common Names and Addresses for Refunds-** Sales employees frequently make bogus refunds to customers for merchandise. The address shown for the refund is then made to the employee's address, or to the address of a friend or co-worker.
18. **Increasing Reconciling Items-** Stolen deposits, or bogus checks written, are frequently not removed, or covered, from the reconciliation. Hence, over a period of time, the reconciling items tend to increase.
19. **General Ledger Out-of-Balance**- When funds, merchandise, or assets are stolen and not covered by a fictitious entry, and the general ledger will be out of balance. An inventory of the merchandise or cash is needed to confirm the existence of the missing assets.
20. **Adjustments to Receivables or Payables-** In cases where customer payments are misappropriated, adjustments to receivables can be made to cover the shortage. Where payables are adjusted, the perpetrator can use a phony billing scheme to convert cash to his or her own use.
21. **Excess Purchases-** Excess purchases can be used to cover fraud in two ways:
    i. Fictitious payees are used to convert funds.
    ii. Excessive purchases may indicate a possible payoff of purchasing agent.
22. **Duplicate Payments-** Duplicate payments are sometimes converted to the use of an employee. The employee may notice the duplicate payment, and then he or she may prepare a phony endorsement of the check.
23. **Ghost Employees-** Ghost employee schemes are frequently uncovered when an auditor, fraud examiner, or other individual distributes paychecks to employees. Missing or otherwise unaccounted for employees could indicate the existence of a ghost employee scheme.
24. **Employee Expense Accounts-** Employees frequently conceal fraud in their individual expense account reimbursements. These reimbursements should be scrutinized for reasonableness and trends, especially in the area of cash transactions on the expense account.
25. **Inventory Shortages-** Normal shrinkage over a period of time can be computed through historical analysis. Excessive shrinkage could explain a host of fraudulent activity, from embezzlement to theft of inventory.
26. **Increased Scrap**-In the manufacturing process, an increased amount of scrap could indicate a scheme to steal and resell this material. Scrap is a favorite target of embezzlers because it is usually subject to less scrutiny than regular inventory.
27. **Large Payments to Individuals-** Excessively large payments to individuals may indicate instances of fraudulent disbursements
28. **Employee Overtime-** Employees being paid for overtime hours not worked by altering time sheets before or after management approval.
29. **Write-off of Accounts Receivable-** Comparing the write-off of receivables by customers may lead to information indicating that the employee has absconded with customer payments.
30. **Post Office Boxes as Shipping Addresses-** In instances where merchandise is shipped to a post office box, this may indicate that an employee is shipping to a bogus purchaser.

4. TECHNIQUES TO DETECT FRAUD

Different Techniques used for fraud detection fall into two primary classes: statistical techniques and artificial intelligence.

Examples of statistical data analysis techniques are:

Data preprocessing techniques for detection, validation, error correction, and filling up of missing or incorrect data. Calculation of various statistical parameters such as averages, quintiles, performance metrics, probability distributions, and so on.

For example, the averages may include average length of call, average number of calls per month and average delays in bill payment. Models and probability distributions of various business activities either in terms of various parameters or probability distributions, Computing user profiles, Time-series analysis of time-dependent data. Clustering and classification to find patterns and associations among groups of data.

Matching algorithms to detect anomalies in the behavior of transactions or users as compared to previously known models and profiles. Techniques are also needed to eliminate false alarms, estimate risks, and predict future of current transactions or users. Some forensic accountants specialize in forensic analytics which is the procurement and analysis of electronic data to reconstruct, detect, or otherwise support a claim of financial fraud. The main steps in forensic analytics are (a) data collection, (b) data preparation, (c) data analysis, and (d) reporting.

For example, forensic analytics may be used to review an employee's purchasing card activity to assess whether any of the purchases were diverted or divertible for personal use. Forensic analytics might be used to review the invoicing activity for a vendor to identify fictitious vendors, and these techniques might also be used by a franchisor to detect fraudulent or erroneous sales reports by the franchisee in a franchising environment.
Fraud management is a knowledge-intensive activity.

The main AI techniques used for fraud management include:

Data mining to classify, cluster, and segment the data and automatically find associations and rules in the data that may signify interesting patterns, including those related to fraud.
Expert systems to encode expertise for detecting fraud in the form of rules.

Pattern recognition to detect approximate classes, clusters, or patterns of suspicious behavior either automatically (unsupervised) or to match given inputs.

Machine learning techniques to automatically identify characteristics of fraud.

Neural networks that can learn suspicious patterns from samples and used later to detect them. Other techniques such as link analysis, Bayesian networks, decision theory, land sequence matching are also used for fraud detection.
Machine learning and data mining. Main articles: Machine learning and Data mining Early data analysis techniques were oriented toward extracting quantitative and statistical data characteristics. These techniques facilitate useful data interpretations and can help to get better insights into the processes behind the data. Although the traditional data analysis techniques can indirectly lead us to knowledge, it is still created by human analysts.
To go beyond, a data analysis system has to be equipped with a substantial amount of background knowledge, and be able to perform reasoning tasks involving that knowledge and the data provided. In effort to meet this goal, researchers have turned to ideas from the machine learning field. This is a natural source of ideas, since the machine learning task can be described as turning background knowledge and examples (input) into knowledge (output).

If data mining results in discovering meaningful patterns, data turns into information. Information or patterns that are novel, valid and potentially useful are not merely information, but knowledge. One speaks of discovering knowledge, before hidden in the huge amount of data, but now revealed.

4. LITERATURE SURVEY

[8]Late years, more multi-view information is broadly utilized as a part of numerous true applications. This sort of information (for example, picture information) are high dimensional and gotten from diverse element extractors, which speaks to particular points of view of the information. Step by step instructions to group such information productively is a challenge. In this paper, we propose a novel multi-see bunching system, called Re-weighted Discriminatively Embedded Means (RDEKM), for this undertaking. The proposed technique is a multi-view minimum outright lingering model which initiates strength to proficiently mitigate the impact of anomalies and figures it out measurement lessening amid multi-see bunching. In particular, the proposed model is an unsupervised streamlining plan which uses Iterative Re-weighted Least Squares to settle least absolute lingering and adaptively controls the circulation of different weights in a re-weighted way just in light of its own low-dimensional subspaces and a typical grouping marker grid. Moreover, hypothetical investigation (counting optimality also, union investigation) and the improvement calculation are additionally exhibited. Contrasted with a few best in class multi-see grouping strategies, the proposed technique significantly progresses the precision of the grouping comes about on generally utilized benchmark datasets, which exhibits the prevalence of the proposed work.

[9] In this concise, we propose a low-multifaceted nature compositional execution of the K-implies based bunching calculation utilized broadly in portable wellbeing observing applications for unsupervised and administered learning. This has been tended to by the utilization of a 2-D Coordinate Rotation Digital Computer-based low-many-sided quality motor for figuring the n-dimensional Euclidean separation required amid bunching. The proposed grouping motor was combined utilizing the TSMC 130-nm innovation library, and a place and course was performed taking after which the center territory and power were evaluated as 0.36 mm2 furthermore, 9.21 maw at 100 MHz, individually, making the plan pertinent for low-control constant operations inside a sensor hub.

[10] This paper proposes a novel bunching method actualized in the quaternion area for subjective arrangement of E-nose information. The proposed system is in numerous courses like the well known k-implies bunching calculation. In any case, calculations completed in the quaternion space have yielded better class

distinctness and higher group legitimacy. A pool of conceivable bunch focuses was made by subjecting each beginning focus to a settled pivot in the quaternion space. The test tests were then contrasted and each of the focuses in the pool what's more, doled out to a fitting focus utilizing least Euclidean separate measure. The developing groups have been assessed occasionally for their minimization and interclass partition utilizing the Davis–Bouldin (DB) record. The arrangement of groups having least DB record was picked as an ideal one. It was watched that utilizing the proposed system the opposite DB list remains fundamentally higher with progressive emphases inferring a steady execution on the group legitimacy front. Moreover, groups framed utilizing quaternion variable based math have been seen to have a littler DB list. At long last, when contrasted and the conventional k-implies calculation, the proposed method performed fundamentally better as far as rate grouping of unlabeled specimens.

[11] In this letter, a straightforward and successful unsupervised approach in light of the consolidated contrast picture and k-implies grouping is proposed for the engineered opening radar (SAR) picture change recognition assignment. In the first place, we utilize a standout amongst the most mainstream demising strategies, the probabilistic-fix based calculation, for dot clamor decrease of the two multi temporal SAR pictures, and the subtraction administrator and the log proportion administrator are connected to produce two sorts of basic change maps. At that point, the mean channel and the middle channel are utilized to the two change maps, separately, where the mean channel concentrates on rolling out the improvement outline and the neighborhood, and the middle channel is utilized to save the edge data. Second, a straightforward blend system which utilizes the maps gotten by the mean channel and the middle channel is proposed to produce a superior change delineate. At last, the k-implies bunching calculation with k = 2is utilized to bunch it into two classes, changed region and unaltered range. Neighborhood consistency and edge data of the distinction picture are considered in this technique. Exploratory outcomes acquired on four genuine SAR picture informational indexes affirm the viability of the proposed approach.

[12]Extortion recognition is one of the greatest difficulties in the telecom industry. Usually utilized methodologies, for example, govern sets, anomaly recognition, and order, have high computational taken a toll, so they don't function admirably on mass information regarding precision also, speed. Furthermore, those calculations are bad at recognizing new misrepresentation designs. In this paper, we propose a UIS (United Smart Scoring) calculation for extortion recognition which has three benefits. To begin with, it has brought down computational many-sided quality. We utilize Manhattan remove rather than Euclidean separation to measure likeness between extortion tests and ordinaries. Second, new misrepresentation examples can be identified viably by joint misrepresentation likelihood. At long last, UIS can create and refresh constant scores, which recognize early-time misrepresentation and limits monetary misfortunes. Incorporated trials on genuine datasets of the telecom industry exhibit that UIS is continuous, viable, furthermore, hearty in various circumstances.

[13] Energy provided by the power utility does not reach to the purchaser end all in all. A segment of vitality is lost in the conveyance framework in light of Technical and Nontechnical misfortunes. The goal of this paper is to distinguish the Non-specialized misfortunes by checking client unpredictable utilization profiles in power circulation framework with the help of information mining strategies. As an initial step fluffy CMeansbunching is performed to gathering clients of same utilization designs. At that point fluffy based order method connected with help of fluffy enrollment work what's more, the separations of group focuses are measured by Euclidean separation, and the separations are standardized and requested with unitary file score, the most noteworthy score speaks to fraudsters. The approach was tried on genuine information, demonstrating great execution in undertakings of misrepresentation and estimation deformity location contrasting and burglary record of Power Distribution Company.

[14] As online business deals keep on growing; the related online extortion remains an alluring wellspring of income for fraudsters. These false exercises force an extensive budgetary misfortune to dealers, making on the web misrepresentation identification a need. The issue of extortion location is worried with not just catching the false exercises, additionally catching them as fast as could reasonably be expected. This auspiciousness is critical to diminish monetary misfortunes. In this examination, a profiling strategy has been proposed for charge card extortion identification. The attention is on misrepresentation cases which can't be distinguished at the exchange level. In the proposed strategy the examples characteristic in the time arrangement of totaled day by day sums spent on an individual charge card account has been extricated. These examples have been utilized to abbreviate the time between when a misrepresentation happens and when it is at long last identified, which

brought about timelier extortion identification, enhanced discovery rate and less budgetary misfortune.

[15] The characterization issue is intermittent in the unique circumstance of regulated learning. A characterization issue is a class of computational undertaking in which marks must be doled out to protest cases utilizing data obtained from named examples of a similar sort of articles. At the point when these items contain time delicate information, uncommon characterization strategies could be utilized to exploit of the innate additional data. To the extent this paper is concerned, the time delicate information are successions of qualities that speak to the deliberate vitality utilization of private customers in a given month. Customary classifiers don't take transient elements into record, translating them as a progression of random static data. The proposed technique is to create strategies for characterization to be connected in a genuine time series issue that some way or another consider the time arrangement as being a similar esteem being over and again measured. Two new methodologies are proposed to manage this issue: the first is a Hybrid classifier that utilizations bunching, DTW (Dynamic Time Warp) and Euclidean separation to mark a given example. The second is a Weighted Bend Comparison Algorithm that makes utilization profiles and contrasts them and the obscure example to characterize it.

REFERENCES

1. Tsikerdekis M, Zeadally S. Online Deception in Social Media. Commun ACM. 2014;57(9):72–80.
2. Vishwanath A. Diffusion of deception in social media: Social contagion effects and its antecedents. Inf Syst Front. 2014 Jun;17(6):1353–67.
3. Tsikerdekis M, Zeadally S. Detecting and Preventing Online Identity Deception in Social Networking Services. IEEE Internet Comput. 2015 May;19(3):41–9.
4. Guang-xing WEI. Team Cooperation Incorporating Inequity Aversion and Social Norms. 2009;(2003):1174–81.
5. Tsikerdekis M, Zeadally S. Multiple Account Identity Deception Detection in Social Media Using Nonverbal Behavior. IEEE Trans Inf Forensics Secur. 2014 Aug;9(8):1311–21.
6. Tsikerdekis M, Zeadally S. Online deception in social media. Commun ACM. 2014 Sep;57(9):72–80.
7. Oneto L, Bisio F, Cambria E, Anguita D. Statistical Learning Theory and ELM for Big Social Data Analysis.
8. Xu J, Han J, Nie F. Discriminatively Embedded K-Means for Multi-view Clustering. Cvpr. 2016;7149(c):5356–64.
9. Adapa B, Biswas D, Bhardwaj S, Raghuraman S, Acharyya A, Maharatna K. Coordinate Rotation-Based Low Complexity K-Means Clustering Architecture. IEEE Trans Very Large Scale Integr Syst. 2017;PP(99):1–5.
10. Kumar R, Dwivedi R. Quaternion Domain k-Means Clustering for Improved Real Time Classification of E-Nose Data. IEEE Sens J. 2016;16(1):177–84.
11. Zheng Y, Zhang X, Hou B, Liu G. Using Combined Difference Image and-Means Clustering for SAR Image Change Detection. Geosci Remote Sens Lett IEEE. 2014;11(3):691–5.
12. Niu K, Jiao H. A Real-Time Fraud Detection Algorithm Based on Intelligent Scoring for the Telecom Industry. 2016;1:1–4.
13. Fabris F, Margoto LR, Varejão FM. Novel approaches for detecting frauds in energy consumption. NSS 2009 - Netw Syst Secur. 2009;546–51.
14. Seyedhossein L, Hashemi MR. Mining information from credit card time series for timelier fraud detection. 2010 5th Int Symp Telecommun IST 2010. 2010;619–24.
15. Das S, Nandeshwar VJ, Phadke GS. Discrimination of adulteration orange juice by Linear Discriminant Analysis (LDA). 2015 IEEE Int WIE Conf Electr Comput Eng WIECON-ECE 2015. 2016;39–42.