



A Novel Approach of Detecting Frauds in Ecommerce Sites by Hybridizing Knn and Euclidean Distance Mechanism

Mini Singh Ahuja
Assistant Professor
GNDU RC Gurdaspu
Punjab, India

Lovepreet Singh
Student M.Tech (CSE)
GNDU RC Gurdaspu
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 piece 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.

Keywords- KNN, Euclidean Distance, Hybridization of KNN and Euclidean distance.

I. INTRODUCTION

The trust and associated parameters are basic in the success of the system. The shopping websites in order to pick up benefit and draw in user enhance these parameters. Trust count can involve hectic computations and it is not a direct metric. Direct metric have arguments which are visible and can be calculated directly. The indirect measures, for example, quality are relatively hard to calculate. The considered parameters in the proposed systems utilize both of these measures. [1-3]The trust management is done as search engine improvement. The search engine streamlining is required in practically every area of ecommerce. Building Square with through which Ecommerce website finish their branches. The online shopping websites increase their users usually on the premise of user reviews and certainty. The user reviews are analyzed on scale of 1 to 5. These are otherwise called star evaluations. Higher the evaluations more successful will be website. There are certain truths about the user reviews.

- 70% of customers counsel reviews or evaluations before making a last purchase
- 63% of consumers are more likely to purchase from a site in the event that it has item evaluations and reviews.
- 67% of consumers read 6 reviews or less before they feel they can believe a business enough to make a purchase.
- As numerous as 79% of consumers trust item reviews as much as a personal recommendation.
- 80% of consumers have changed their psyche about purchases based on negative data they have discovered online.
- On average, 75 percent of reviews posted on review websites are positive, and 71 percent of consumers agreed that reviews make them more comfortable that they are buying the correct item or service.

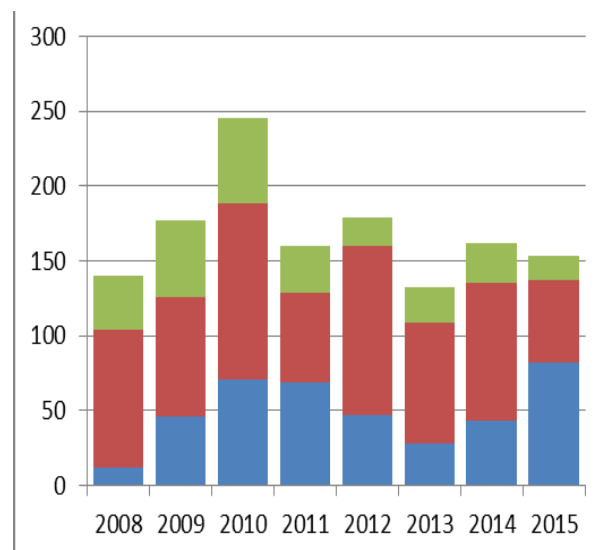
The second parameter in lane for success of online shopping sites is certainty. [6]– [15] This parameter is hard to construct and consume time to fabricate. As the sites grew older and for the duration of the life spam if their reviews are positive then certainty will construct. Building certainty will be accomplished just if website performance is more than 90% every one of the times in operation.

At the point when part of the fundamental memory gets liberated, the Medium-term scheduling is a piece of the working framework takes a gander at the rundown of suspend prepared procedures, chooses which one is to be swapped in (contingent upon need, memory and different assets required, and so on). This scheduler works in close conjunction with the long haul scheduler. It will play out the swapping-in capacity among the swapped-out procedures. Medium-term scheduler executes somewhat all the more much of the time.

The information about the user reviews and certainty is collected from UCI dataset. The online dataset provide the data about the year 2010-12.

The certainty parameter additionally taken into consideration however this parameter is hard to calculate and hence is used less frequently as compared to user reviews.

On the premise of certainty Amazon is considered best possible one. The information is collected in 2014 and represented on the premise of outline. The approach considered for evaluation includes KNN, Euclidean Distance and Hybridization of KNN and Euclidean distance.



ADVANTAGES AND DISADVANTAGES OF KNN, EUCLIDEAN DISTANCE AND HYBRID APPROACH.

i. **ADVANTAGES OF KNN APPROACH**

- Powerful to boisterous preparing information (especially on the off chance that we use Inverse Square of weighted distance as the "distance").
- Effective if the preparation information is large(5,6).

ii. **DISADVANTAGES OF KNN APPROACH**

- Need to determine value of parameter K(number of nearest neighbors)
- Distance based learning is not clear what type of distance to use and which attribute to use to produce the best results. Might we use every one of the attributes or certain attributes as it were?
- Calculation cost is quite high because we need to compute distance of each query instance to all preparation samples. Some indexing may reduce this computational cost.

iii. **ADVANTAGES OF EUCLIDEAN DISTANCE**

- It is reportedly faster than most other means of determining correlation [3].
- It compares the relationship between genuine appraisals. This means the Euclidean distance is a reasonable measure of how comparative evaluations are for specific preferences or items [7-9].

iv. **DISADVANTAGES OF EUCLIDEAN DISTANCE**

- Where there is a high noise-to-flag proportion and negative spikes, any correlation is hard to establish. The Euclidean distance method likewise suffers in such cases.
- Preferences are listed by rank instead of as per genuine values.
- It can't determine the correlation between user profiles who have comparative trends in tastes, however different evaluations for some of the same items. A method like the Pearson correlation would give a sign of how comparative a set of preferences are, regardless of vacillations in individual evaluations.

v. **ADVANTAGES OF PROPOSED APPROACH**

- Powerful to boisterous preparing information (especially on the off chance that we use Inverse Square of weighted distance as the "distance").
- Effective if the preparation information is large.
- It is reportedly faster than most other means of determining correlation [3].
- It compares the relationship between genuine appraisals. This means the KNN + Euclidean distance is a reasonable measure of how comparative evaluations are for specific preferences or items.
- Need to determine value of parameter K(number of nearest neighbors)
- Distance based learning is clear with hybrid approach.
- Calculation cost is quite high because we need to compute distance of each query instance to all preparation samples. Some indexing may reduce this computational cost.
- Where there is a high noise-to-flag proportion and negative spikes, any correlation is hard to establish.

The KNN+Euclidean distance method does not suffers in such cases.

- KNN+Euclidean distance measure the correlation between quantitative, nonstop variables. It is suitable for ordinal information, where preferences are listed by rank instead of as per genuine values.
- It can determine the correlation between user profiles who have comparative trends in tastes, however different evaluations for some of the same items.

II. RELATED WORK

There is a work which is done toward increasing the trust inside the ecommerce. The papers we have analyzed are enlisted here. [10] Trust is keys for the success of online business. The case of hotel reservation is considered in this case. The hotel reservation will work properly if the booking is as indicated by the desired guidelines as purposed by the consumer. Otherwise trust breach will happen. [11] Trust in the suggested paper is accomplished with the help of breaking down the culture. Some users are not expecting change and hence can't be forced to believe a system which is alien to them. In such environment gathering trust will be very troublesome. Such circumstances are analyzed in the purposed paper. [12] Trust is enhanced by the use of this technique. The concept of user feedback is considered in this case. The feedback if positive then the trust will be enhanced and vendor can earn benefit. The trust can likewise be breached if for some online exchanging the feedback become negative [13] trust concept is considered in this case. The trust will be expressed by the use of the vendor consumer relationship. The relationship must be solid. The concept of coupling and cohesion will be introduced in this case. [14] The review of online consumer trust will be considered in this case. The online consumer trust is the new concept where user purchases the merchandise online if the online shopping gateway has the trust. The trust will decide the benefit earn by the vendor. [15] Online trust building mechanism is considered. The online word is with the end goal that some consumer apprehensive from it. Because of this reason confide in comes into existence. The trust figure is imperative the success of online web entry. Increase in trust will cause more activity toward the online ecommerce whereas breach in the trust will cause lose. These circumstances are analyzed in this paper.

From every one of the papers we have analyzed that very few or nothing is done towards the increase in trust by the use of parameters like quality component. A trust web model based on a distributed search calculation and a network of trusted broker that can authorize a trusted channel through which end to end executing parties deal basically directly and chance free with each other is proposed in this paper.

From the analyzed paper we get to realize that parameters like certainty and user reviews are considered in mellow form and not in useful manner. To increase the profitability of online shopping websites these parameters will assume basic role.

III. RESULTS AND PERFORMANCE ANALYSIS

The result which comes from KNN and Euclidean Distance doesn't detect much fraud as compared to the results which

comes from the Hybrid Approach of KNN and Euclidean Distance.

Table 1: Showing Fraud Detection

Population Analyzed	KNN	Euclidean	KNN+ Euclidean
100	7	6	15
200	14	12	26
300	21	18	39
400	28	24	52
500	35	30	65
600	42	36	78
700	49	42	91
800	56	48	104
900	63	54	117
1000	70	60	140

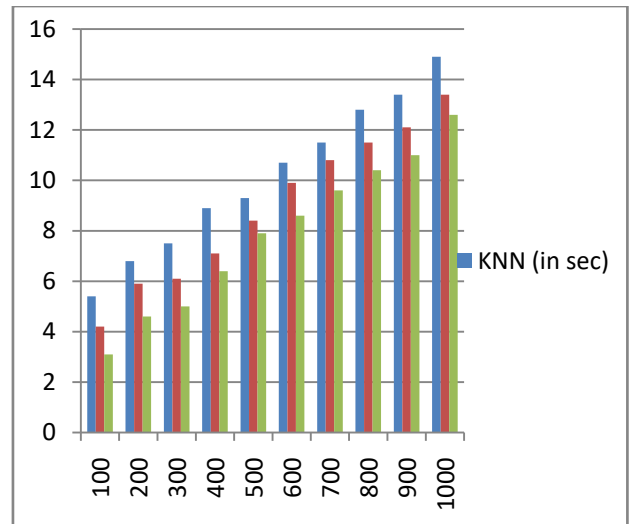


Figure 2: Showing time consumption

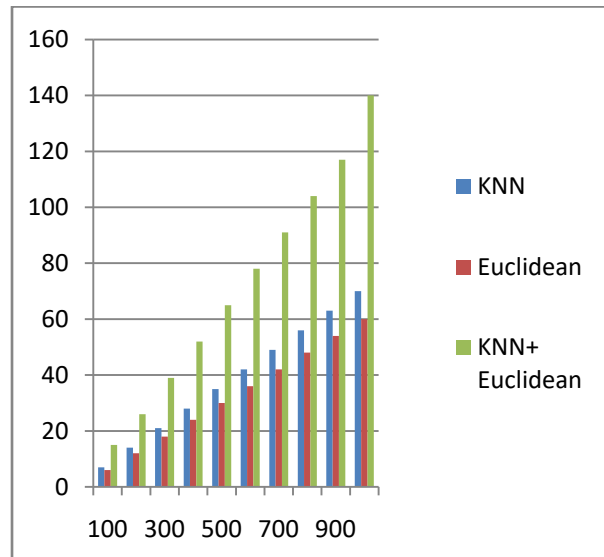


Fig 1: Showing the Fraud Detection in the different population situation

TIME CONSUMPTION:

Time Consumption is the fact that means the total time it takes to complete the execution of the process. The Time Consumption is listed as under:

Table 2: Showing time consumption

Population Analyzed	KNN (in sec)	Euclidean (in sec)	KNN+ Euclidean
100	5.4	4.2	3.1
200	6.8	5.9	4.6
300	7.5	6.1	5.0
400	8.9	7.1	6.4
500	9.3	8.4	7.9
600	10.7	9.9	8.6
700	11.5	10.8	9.6
800	12.8	11.5	10.4
900	13.4	12.1	11.0
1000	14.9	13.4	12.6

IV. CONCLUSION AND FUTURE WORK

The trust is the variable which has numerous definitions. Trust is established in case frauds are limited within the system. Fraud detection procedure must be in place for the detection of frauds. KNN and Euclidean distance mechanisms are common approaches used to detect Frauds but approach consumes time. In order rectify the problem, hybrid approach of KNN and Euclidean distance approach is hybridized and better result is obtained.

V. REFERENCES

1. Wang H. Review of studies on online consumer trust. In: 2010 Second International Conference on Computational Intelligence and Natural Computing [Internet]. IEEE; 2010 [cited 2016 Feb 15]. p. 97–100. Available from: <http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=5643778>
2. Yuan Y, Wu X, Lu Y, editors. Trustworthy Computing and Services [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013 [cited 2016 Mar 13]. (Communications in Computer and Information Science; vol. 320). Available from: <http://www.springerlink.com/index/10.1007/978-3-642-35795-4>
3. Pradhan S, Gay V, Nepal S. Improving dental care recommendation systems using trust and social networks. In: 2014 IEEE International Conference on Communications (ICC) [Internet]. IEEE; 2014 [cited 2016 Feb 9]. p. 4264–9. Available from: <http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6883990>
4. Buzzi MC, Buzzi M, Leporini B, Akhter F. User Trust in eCommerce Services: Perception via Screen Reader. In: 2009 International Conference on New Trends in Information and Service Science [Internet]. IEEE; 2009 [cited 2016 Feb 15]. p. 1166–71. Available from: <http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=5260447>
5. Begum S, Chakraborty D, Sarkar R. Data Classification Using Feature Selection and kNN Machine Learning Approach. 2015 Int Conf Comput Intell Commun Networks [Internet]. 2015;811–4. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7546208>

- [6]. 6. Yan Z, Xu C. Combining KNN algorithm and other classifiers. Proc 9th IEEE Int Conf Cogn Informatics, ICCI 2010. 2010;(1):800–5.
- [7]. 7. Sinwar D, Kaushik R. Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering. Int J Res Appl Sci Eng Technol. 2014;2(5):270–4.
- [8]. 8. Nguyen HK, Song J Bin, Han Z. Distributed Demand Side Management with Energy Storage in Smart Grid. IEEE Trans Parallel Distrib Syst [Internet]. 2015 Dec 1 [cited 2016 May 3];26(12):3346–57. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6963474>
- [9]. 9. Wu K-L, Yang M-S. Alternative c-means clustering algorithms. Pattern Recognit [Internet]. 2002;35(10):2267–78. Available from: <http://www.sciencedirect.com/science/article/pii/S0031320301001972> (http://ac.els-cdn.com/S0031320301001972/1-s2.0-S0031320301001972-main.pdf?_tid=9de9cf82-db5f-11e4-b393-00000aacb360&acdnat=1428216624_f37260d02e89aa3079901fd9391f4c0b)
- [10].10. Patel T, Kale MO. A Secured Approach to Credit Card Fraud Detection Using Hidden Markov Model. 2014;3(5):1576–83.
- [11].11. Ganji VR, Mannem SNP. Credit card fraud detection using anti-k nearest neighbor algorithm. Int J Comput Sci Eng. 2012;4(06):1035–9.
- [12].12. Tripathi KK, Pavaskar MA. Survey on Credit Card Fraud Detection Methods. Int J Emerg Technol Adv Eng. 2012;2(11):721.
- [13].13. Abdulla N, Rakendu R, Varghese SM. A Hybrid Approach to Detect Credit Card Fraud. 2015;5(11):304–14.
- [14].14. Dheepa V, Dhanapal R. Analysis of Credit Card Fraud Detection Methods. International J Recent Trends Eng. 2009;2(3):126–8.
- [15].15. Brennan P. A comprehensive survey of methods for overcoming the class imbalance problem in fraud detection. Thesis. 2012;(June):1–107.