International Journal of Research in Engineering, Science and Management
Volume-2, Issue-3, March-2019
www.ijresm.com | ISSN (Online): 2581-5792

Online In-Auction Fraud Detection using Naive Bayes Classifier

M. K. Santhosh[1], M. Sakthi Balan[2], R. S. Surya Prasaath Babu[3], M. Tharma Seelan[4]
[1,2,3,4]*Student, Department of Information Technology, K.L.N College of Engineering, Madurai, India*

***Abstract*: In this world of rising technologies, on-line frauds are speedily increasing with the increase in on-line searching era. It's been known within the past analysis that the shilling is main cause behind on-line auction frauds. Many researchers have projected varies strategies to counter the likelihood of shilling in on-line auction. In this paper we propose a hybrid model for improving the accuracy in predicting the online frauds from shilling. This is a applied mathematics model that generates the likelihood sequence supported the bids applied by users. The Naive Bayes algorithm has been applied to the data by considering totally different bidding habits of user. Using this model, the categorization of different bidding behaviors for various bids are done.**

***Keywords*: Auction fraud. Shill Bidding. HMM**

1. Introduction

In data processing, classification is one of the best techniques available to predict outcomes in data sets. Naïve Thomas Bayes Classifier, a conventional supervised classification technique, is one such classification techniques accustomed predict outcomes. This brings a brand new challenge in IT business of on-line frauds wherever a fraudster uses any sort of fraud theme that uses email, websites or the other web means that to gift a request for an add of cash to prospective victims, to conduct dishonest transactions or to other. Naive Bayes classifier with the assumption of Independence among predictors is a challenging task [7].

Online auction fraud is the kind of fraud that is principally caused by searching scams and is categorized into 3 parts: (1) Pre-Auction that's the initial phase wherever fraud happens by commercialism black market merchandise, merchandise at false worth etc. (2) In- Auction that includes purchase of things based mostly on bidding mechanisms. (3) Post-Auction that happens when the auction is over throughout payment or delivery of merchandise [7]. The on-line frauds have totally different ways that ways in that of execution based mostly on the bids patterns And winning policies like: (1) English Auction is an ascending worth auction in which bidders vie with one another by putting higher bids. (2) Dutch Auction that is drizzling worth auction within which auction starts at high worth and reaches to the minimum reserved worth. (3) 1st worth Sealed Auction within which all bidders bids at same time unaware of others bids. (4) Second worth Sealed Auction is similar to sealed bid sort however winner has to get second highest of winning bid [3].

In Naïve Mathematician Classifier, the idea that the possibilities of every attribute with relation to a category ar freelance of all alternative attribute values is found to be apt. This assumption is created to primarily change the calculation of chances. This assumption is called as Conditional Independence. Economy being the most valuable entity of any country want correct security from such kind of on-line frauds that lead to large loss of finance that will increase speedily as shown in figure two. The projected work is said to bar and detection of on-line in-auction fraud mistreatment Hidden mathematician Model. For the sake of simplicity, the approach uses most vital and lightest parameter i.e. no. of bids& bid values for detection and is organized into superimposed design that provides a standard platform for the implementation of projected approach. auction sorts and their fraud classification, give complete literature survey, the projected mechanism to notice and forestall on-line auction fraud, section four presents results and comparative analysis

2. Related work

Buyers and sellers can purchase and sell popular electronic products in an efficient way in online. In spite of popularity of online auctions, there are many fraudulent bidding or selling can occur during an auction [5]. Among auction frauds, shill bidding is one of the hardest types of auction fraud to be detected. The popularity of online auction is evidenced by the explosive growth of online auction sites with millions of users buying and selling goods from any parts of the world. This rapid growth of online auctions has gradually led to a respective increase in online frauds [6]. Many re-searchers have developed fraud detection and prevention methods. However, there are difficulties in using the past auction data to evaluate the effectiveness of these methods [4]. While using commercial data, it is not possible to accurately identify cases of fraud. Using synthetic data, the conclusions drawn may not extend to the real world [2].

3. Proposed work

This section delineate the projected work wherever we've mentioned our mechanism i.e. on-line Hybrid Model for Auction(OHMA)which is that the extension of antecedently projected model OHM [7] and the entire methodology used for the bar and detection of on-line in-auction fraud. first of all

OHMA is classified into 2 classes that is delineate below.

A. Introduction to naïve bayes

Naïve Bayes may be a set of Bayesian call theory. It's called naive because the formulation makes some naïve assumptions. Python's text-processing skills that get a divorce a document into a vector are used. This can be used to classify text. Classifies may put into human-readable form. It is a preferred classification methodology additionally to conditional independence, over fitting, and Bayesian methods. In the face of the simplicity of Naive mathematician, it will classify documents astonishingly well Instinctively a possible justification for the conditional independence assumption is that if the document is concerning politics, this is a good evidence of the kinds of other words found in the document. Naive Bayes may be a cheap classifier during this sense and has least storage and quick coaching, it's applied to time-storage crucial applications, like mechanically classifying websites into types and spam filtering. Considering a collection of objects, each of which belongs to a known class, and each of which has a known vector of variables, the aim is to create a rule that allows to apportion future objects to a category, given just the vectors of variables marking out the future objects. These issues area unit called ―supervised classification problem‖, area unit worldwide, and most of the strategies for constructing such rules are developed. It is terribly simple to determine, and no need any complicated repetitive parameter estimation schemes. This means it ought to be applied to very large knowledge sets. It is simple to interpret, so unskilled users in classifier technology can make out the reason for it is making the classification it makes. Finally, it usually will surprisingly well: it shouldn't be the most effective potential classifier in any explicit application, but it can usually be relied on to be robust and to do well.

B. HMM for detection of on-line in-auction fraud

The projected model is a 2-layered subject field model which at one layer stress on authentication of users whether or not sellers or consumers and at different layer identifies and forestall the fraud. In this model each user have to be compelled to register if they want to sell or get product, for this purpose there is an authentication part (explained in our previous analysis work [7]) if that's gone any user with success then solely the users are going to be ready to participate in any auction. more for detection; nature of bidder's biding habit is monitored mistreatment OAFDM and supported the results of the model more actions (like discarding the auction, announcement of winner) are going to be taken.

1) Prediction

Preventive measures are often simpler in on-line auctions than reactive measures. Wily traders sometimes exploit loopholes left in procedural rules to "attack" honest users and challenge system and mechanism designers. If the auction procedural rules embedded within the software system programs of on-line data technology applications ar airtight, fraud activities are often simply prevented, avoided and eliminated. As we all know a user might participate {in totally different in several numerous auctions at different times and therefore the bids applied by any user depends on the sort of item (for ex. Electronic, Groceries, household, cloths & accessories etc.) auction is related to. thus we tend to contemplate the transitions of various things in numerous auctions because the state transition in our model. the kind of every auction depends on the kind of item that is more associated to its merchandiser and therefore it forms the hidden states of HMM.

Now mistreatment the state and observations symbols we'd like to spot the likelihood matrices A, B, π to complete the tuple definition of HMM. These parameters can be determined within the initial part of HMM by analysis of initial observations mistreatment the Baum-Welch algorithmic program [7].

2) Generation of needed Observations

After this our next step is to dynamically verify the observation symbols corresponding to individual user participated in totally different auctions, for this purpose we tend to use a clump algorithmic program wherever clusters of users area unit created supported the auctions they have participated and no. of bids corresponding to AN individual user area unit calculable mistreatment K- means that clump algorithmic program [16]. The no. of clusters 'k' is same as the observation symbols and let c1, c2...cm be the calculated centroids for the clusters. Let b be the bid placed by user u in any auction, then the calculation of observation symbols is done as shown in equation (i) where arg_min returns the minimum price of deviation in current bid and its user's bidding habit:

$$o_b = \text{arg\_}min_i \mid \text{b-}\ c_i \mid \qquad (1)$$

3) User Profile Categorization

As mentioned earlier we've outlined 3 state symbols (l, m, h) and supported the clump applied for every user they're categorized into their bidding habits, namely, high- bidders (h), medium-bidder s(m), low-bidders (l) teams severally. Let pi be the proportion of the overall no. of bids corresponding to its cluster price ci. Then the bidding habit (BH) for several user is calculated as shown in equation (ii) wherever arg_max returns the most of no. of bids applied for every specific state symbol:

$$\text{BH(u)} = \text{arg\_}max_i \mid\ p_i \mid \qquad (2)$$

4) Preventive measures

Preventing in-auction fraud from happening is possibly the best solution, nonetheless, in some cases, when in-auction fraud cannot be prevented, approaches that can predict its occurrence can also reduce the risk to auction participants. Kauffman and Wood [7] examined how the fee structure on eBay may motivate shill bidding and first identified "reserve price

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-3, March-2019
www.ijresm.com | ISSN (Online): 2581-5792

shilling” based, in part, on their research into eBay auctions of rare coins in Apr 2001.They tested whether or not some questionable bidding behaviors square measure thanks to reserve worth shilling. According to the test results, they built an empirical profit model to predict reserve price shilling based on the seller’s previous behavior before the auction begins. In addition, researchers have tried to style bidding ways exploitation theory of games so as to assist honest users counteract scams [5]. Porter and Shoham proposed two equilibrium bidding strategies to counteract bid shading and false bids in sealed-bid auctions, namely first price sealed-bid auction and second price sealed-bid auction [5]. An equilibrium bidding strategy could be a Bayes-Nash equilibrium if the bidder’s expected gain is maximized once the bidding methods for all different bidders area unit mounted. Usually, the expected gain or utility function equals the product of the probability of winning and the difference between a winner’s highest willing-to-pay price and the actual winning price. The probability of winning can be estimated by the probability of cheating and the probability that a bidder’s highest willing-to-pay price is higher than that of a shill’s. Therefore, when knowing the possibility of cheating, equilibrium can be derived to counteract a shill and maximize a bidder’s expected gain.

4. Result

In the previous section, we've mentioned well concerning as shown in Table I severally with their several values of centroids. It is discovered that the bid 9,6,10,7 has been classified to cluster cl with center of mass price as eight &a total of four-hundredth ,bids 15,16,12,13 area unit classified to cluster cm with center of mass price as 14 a total of four-hundredth and bids eighteen, 19 have been classified to cluster ch with center of mass price as eighteen.5of hour of total bids wherever cl is low , cm is medium and ch id high bidding behaviors severally.

A. Output of K-Means algorithmic program

The projected algorithmic program. In this section we tend to have shown the initial experimental results (by scheming the user’s bidding behavior)of OHMA. To understand the result we've taken AN example that shows however the initial data is clustered and therefore the several bids area unit classified. In the method of identification of observation symbols for every user we tend to contemplate (say) the last ten auctions a user has participated and calculate the no. of bids a user has placed in every auction as shown in Table I and bid values of a participated.

Table 1
Auctions with no. of bids spent by a user in every auction

Auctions	1	2	3	4	5	6	7	8	9
No.of.bids	9	15	12	9	6	19	16	13	7

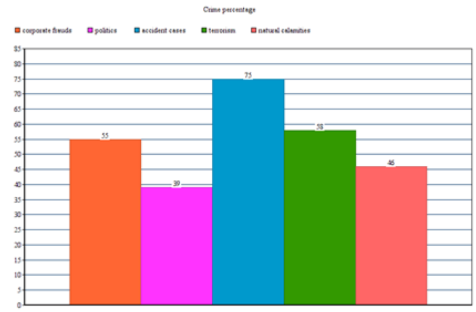

Fig. 1. Output of K-Means algorithmic program

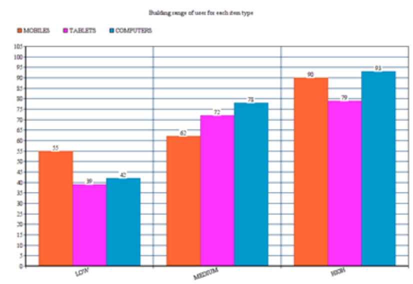

Fig. 2. Building range

Further mistreatment and equation (i) the observation symbols are often calculated for any bid placed by a user. After reading the initial data we tend to apply k-means clump algorithmic program on no. of bids and bid price and that we get the clusters if b = 177, then observation image V1 = has Ob = arg_min ((177-29), (177-150), (177-345)) Ob = arg_min (148, 27, 168)

Ob = twenty seven = cl = l (low), that is once calculated for every bid of user provides the observation sequence.

In this theoretical analysis of the projected model we tend to have known the bidding behavior of a user that is shown within the fig. 5and this can play an important role within the detection of on-line In-Auction fraud mistreatment HMM. equally we tend to area unit scheming this bidding behavior for every user taking part in auction and these calculated values area unit passed to build and train HMM model which is able to more result within the deviation of user from his traditional behavior (if any). ratio in the market of e-commerce and by reviewing varied on the market resolutions to the current downside we tend to found the want of an a good and efficient mechanism as a solution to this downside. In our work we tend to have conferred a mechanism for police investigation the shilling bids to stop the in-auction fraud. From a sensible purpose of read this model are often enforced in any auction platform for resolution the downside of shilling behavior of sellers. Later {we can |we'll |we are going to} show the enforced portal that can give a real time auction platform wherever users will register and can participate in auctions. mistreatment this we are going to generate the real time initial raw information needed for the detection and bar of on-line auction fraud and implementation of the Hidden mathematician Model into the designed portal.

5. Future enhancement

In this paper we tend to build on-line models for the auction fraud moderation and detection system designed for a significant Asian on-line auction web site. By empirical experiments on a real-word on-line auction fraud detection knowledge, we tend to show that our projected on-line integrity model framework, which mixes on-line feature choice, bounding coefficients from professional knowledge and multiple instance learning, will considerably improve over baselines and also the human-tuned model. Note that this on-line modeling framework may be simply extended to several alternative applications, like net spam detection, content improvement and then forth

6. Conclusion

This paper presented an overview of online in-auction fraud detection using naive bayes classifier

References

[1] Pouria Kaviani, Sunita Dho, "Short Survey on Naive Bayes Algorithm," in International Journal of Advance Engineering and Research Development Volume 4, Issue 11, November -2017.

[2] Fei Dong, Sol M. Shatz, Haiping Xu, Combating online in-auction fraud: Clues, techniques and challenges, Computer Science Review, Volume 3, Issue 4, 2009, Pages 245-258.

[3] D. H. Chau, S. Pandit, and C. Faloutsos, police investigation dishonest Personalities in Networks of on-line Auctioneers Principles and follow of information Discovery in info, 2006, pp. 103-114.

[4] Christodorescu, M., V. Ganapathy, J. Giffin, L. Kruger, S. Rubin, and H. Wang. "An auctioning name system supported anomaly detection." In Proceedings of the twelfth ACM conference on pc and communications security. Alexandria, VA: The Association of Computing Machinery. 2005.

[5] Ghasemi, Hamid-Reza, and Gholam-Reza Mohammadi. "Architecture-oriented approach for police investigation fraud in the on-line auction." In e- Commerce in Developing Countries: With Focus on e-Trust (ECDC),2014 eighth International Conference on, pp. 1-7. IEEE, 2014.

[6] Ankit Mundra, Nitin Rakesh, and S. P. Ghrera. "Empirical study of online hybrid model for web frauds bar and detection." In Human pc Interactions (ICHCI), 2013 International Conference on, pp. 1-7. IEEE, 2013.

[7] Priyanka gupta and Ankit Mundra "online In-Auction Fraud Detection Using Online Hybrid model" 2015 International Conference on, pp 1-7 IEEE, 2015.