ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

Use of AI in Fraud Prevention

Vaibhav Gharad, Devendra Gadekar

M.E, Department of Computer Science, JSPM's imperial college of Engineering Wagholi, Pune, India

Ph.D, Department of Computer Science, JSPM's imperial college of Engineering Wagholi, Pune, India

ABSTRACT: Frauds in finance industry is growing risk with the mode of the payments being increased every day. It is required for industry to put measures in place to address this risk for keeping their reputation and for government compliance. Currently finance industry does this manually, which requires effort and workforce like there have to be fraud manager, which look into every record manually and takes action it. Applying simple rules on transactions to pass or failed may lead customer in many false positive cases. The vital observation of out paper is that can rules developed based on the deep learning and machine learning can solve this problem in efficient way. The contribution of work is to apply machine algorithms in the form of AI rules that can be used for preventing fraudulent transaction automatically. We argue that, for the fraud detection domain, fraud catching rate (True Positive rate) and false alarm rate (False Positive rate) are better metrics than the overall accuracy when evaluating the learned fraud classifiers. Specifically, we use a rule learning program to uncover indicators of fraudulent behavior from a large database. These indicators are used to create profilers, which then serve as features to a system that combines evidence from multiple profilers to generate high confidence alarms. Experiments indicate that this automatic approach performs nearly as well as the best hand-tuned methods for detecting fraud. Currently most of this work gets done manually or by auto rule system which have constant rule in place. With proposed system this can be automated i.e. system will save manual efforts as well as provide better accuracy over single rule based system.

KEYWORDS: Machine learning, Artificial Intelligence fraud detection, financial transactions, supervised learning

I. INTRODUCTION

Every financial institution developed their own custom fraud prevention system, which is targeted to their own asses. In current world, banks have reached to conclusion that a unified, global approach is required which will share periodic information with each other to prevent global attacks. This information helps building a global fraud detection solution. However sharing data between the systems can bring some risks such as:

1) Integrity loss of customer data
2) Size of the data grows rapidly.
3) Certain data cannot be passed to other system.
4) Infrastructure issues

Many institutions address their risks either with fraud risk manager, which is manual, or by applying auto rules. With manual efforts, there is cost associated with it and cause delay in response based on the acceptance. With auto rules like reject transaction from specific location causes false positive cases for e.g. if location for transaction is Nigeria then reject transaction.

If (Location==Nigeria) == FRAUD

However, with such a rule there is risk involved as if valid customer is in travel to Nigeria customer may be flagged as fraud a case called false positive. As there are pitfall to above both approaches, we are proposing an approach of using AI rules that crated using machine learning, in this experiment we will be using supervised learning algorithm to train our dataset over the period. Meta-learning (Chan Stolfo 1993) [11]is a general strategy that provides the way of learning how different datasets can be combine to come up with meaningful result by applying results from separately learned classifiers and model.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

Our goal is to compute a classifier that predicts whether a transaction is fraud or non-fraud. To achieve this will consider the following steps in our experiment.

1) Data Sampling: In real world fraud, prevention system a huge volume of data would be required to be stored, this data will grow periodically. However, for our experiment work we are only considering 100000 records.

2) Data Sanitization: As data have multiple dimensions and not all of them would be necessary for our algorithm, so the first thing that we have done is removed fields. Based on the observation it can state that by removing redundant data dimension it does not affect to quality of the data.

3) Data Accuracy: The overall accuracy is important, for our data we know a dummy algorithm can always classify every transaction as non-fraud and still better rate for overall accuracy. In real world for detecting fraud and reduce down the false positive scenarios, to achieve better result we need most accurate data. A lower the rate for catching rate for fraudulent transaction means more large number of fraud activity would happen. If we raise more alarm then that means, there are more false positive. As in industry rate of blocking legitimate customer would cause less harm that that of failing to prevent fraudulent transaction. We have given priority to our classifiers using first the fraud-catching rate over false positive cases.

4) Data combination: By adding more information to certain data, it can be become very useful to algorithm for e.g. adding address information to user data along with the email would be helpful to determine customer more accurately. We have added more parameters to improve overall quality of the data.

5) Classification metrics: Classifiers can be generated using different algorithm and training data set all these classifiers stand candidates for base classifiers for Meta learning Integrating all of them incurs a high overhead and makes the final classifier hierarchy very complex.

Selecting a few classifiers to be integrated would reduce overhead and complexity, but potentially decrease the overall accuracy. Before we examine the tradeoff, we first examine how we select a few classifiers for integration. In addition to True Positive and False

Positive rates, the following evaluation metrics are used in our experiments:

a) Diversity (Chan 1996): using entropy, it measures how differently the classifiers predict on the same instances [11]

b) Coverage (Brodley and Lane 1996): it measures the fraction of instances for which at least one of the classifiers produces the correct prediction.[7]

c) Correlated Error (Ali and Pazzani 1996): measures the fraction of instances for which a pair of baseclassifiers makes the same incorrect prediction. [5]

Motivation: The system developed using the proposed approach will be able to automatically detect fraudulent transaction in better way than existing manual or rule based approach.

The system will be providing following advantages:

- No manual effort will required i.e. a fraud manager.
- Cost of the operation will be reduced.
- Better accuracy over existing process.

II. PROPOSED METHODOLOGY

In proposed methodology we argue that effective use of machine learning modules like supervised learning over large data set can be useful for fraud detection. A set of rules can be created over each set of data. Each set of rule will be responsible for validating information at individual level and determine score. Combination of individual score will determine if transaction is valid or fraud. To undertake this project work to developed solution with Java agents and machine learning algorithms (Stolfo et al. 1997) to create AI rules rules [11]. Though there are many rules that can be created on dataset and applied we will be using following rules. Considering scope of the work.

1) Valid domain: Check if the email address of the user is from the domain which is marked as fraudulent in domain database. There are external systems like Email age that can be used. In our experiment we have set of domains which can be marked as fraud domains.

2) Past history of the transaction: If currently using card is being flagged as stolen or loss card database then most likely transaction which is happening is not legitimate transaction.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

3) Location of the transaction: There are certain location which from where rate of fraudulent transaction is more. If transaction is getting originated from that location and usual transaction happening for current user is different than most likely its fraudulent transaction.

4) Device fingerprint: During user registration capture device details in user record also log the same for user's transaction. If device fingerprint is different than transaction can be fraud one.

AI rules can be formed based on above assumption and the data. As shown in Fig 1 these rules will be effectively applied on incoming transaction. If any of the rules gets kicked off then score of the transaction will be less. An individual system owner can define score that it can allow transaction to get through. The lesser the allowed score chances of transaction fraud will be more i.e. it is advised to have high score to let transaction process.

Advantages of Proposed System:

1) No human efforts are required
2) Reduction in cost
3) Lesser number of false positive scenarios
4) Self learning system
5) Automated and salable system

A. Architecture

The Fig.1 shows the proposed system architecture.

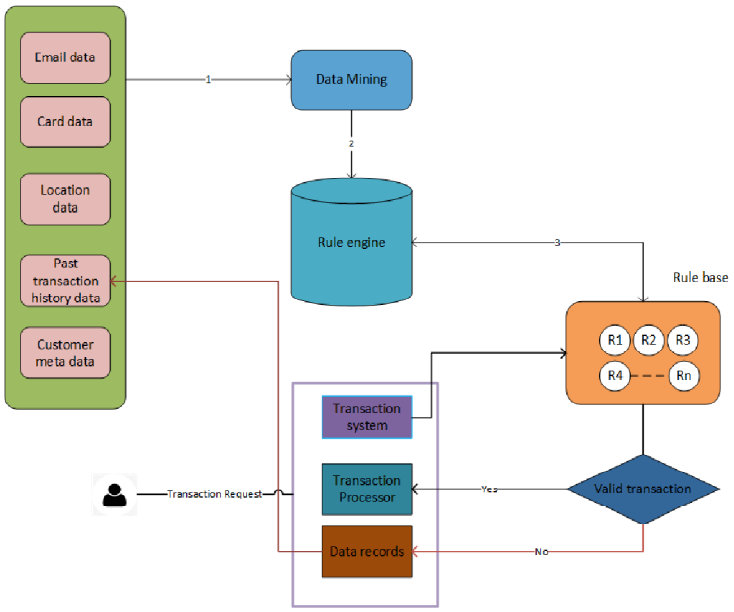

Figure1. System Architecture

B. Algorithms

Class-combiner.:

Select the best base classifiers for meta-learning using the class-combiner (Brodley and Lane 1996) strategy according to the following different. [7]

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

For each record:
(a) Classifiers with the highest True Positive rate (or fraud catching rate)
(b) Classifiers with the highest diversity rate:
(c) Classifiers with the least correlated error rate;
(d) Classifiers with the highest score

Naive Bayes algorithm
Steps:
1) P(c—x): posterior probability of class (C, target) given predictor(x,attributes). This represents the probability of c being true, provided x is true.
2) P(c): is the prior probability of class. This is the observed probability of class out of all the observations.
3) P(x—c): is the likelihood which is the probability of predictor-given class. This represents the probability of x being true, provided x is true.
4) P(x): is the prior probability of predictor. This is the observed probability of predictor out of all the observations.

III. RESULT AND DISCUSSIONS

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 8 GB memory, Windows 7, MySQL 5.1 back-end database and Jdk 1.8. with file size of 100000 transactions. The application is dynamic web application used tool for design code in Eclipseand execute on Tomcat server. Some functions used in thealgorithm are provided by list of jars like Hadoop and spark etc.
We have obtained a large database, 851018 records, of credit card transactions from keggle (Paysim synthetic dataset of mobile money transactions. All these records processed by system to generate result. As it required more efficient hardware to process large data set we have used 100000 records to perform comparison.
Result comparison is done in following area:
1) Accuracy: Accuracy is the most important performance measure which is nothing but the ratio of correctly predicted fraudulent transactions out of total transactions. Accuracy of the auto rule system cannot be increased as it is based on condition on other hand accuracy of the rule based system can be easily increased across the time as supervised model will train itself on the data.
Accuracy TP+TN/TP+FP+FN+TN
During manual effort it is quite obvious that person will be validating all parameters and then approve transaction rate of accuracy will be highest. However by using AI rules human interface can be mimic by system.
2) Efficiency: Efficiency of the system can be measured total number of fraudulent transactions captured correctly with the time and by the cost. As with auto rule we there is tradeoff for accuracy. Auto rule based system provides constant performance up to certain number of record but there is performance hit as number of transaction increased.However rule based system guarantees constant time performance as these rules parallel to each other during operation.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

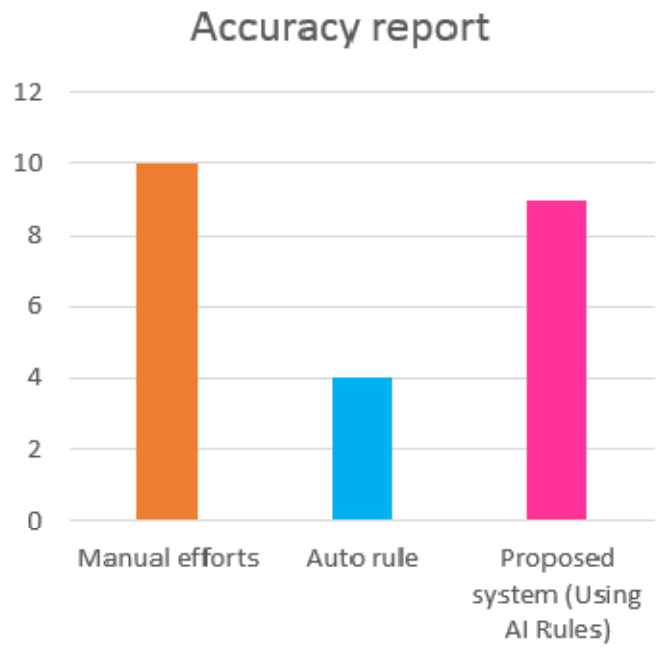

Figure2. Accuracy of detecting fraudulent transactions

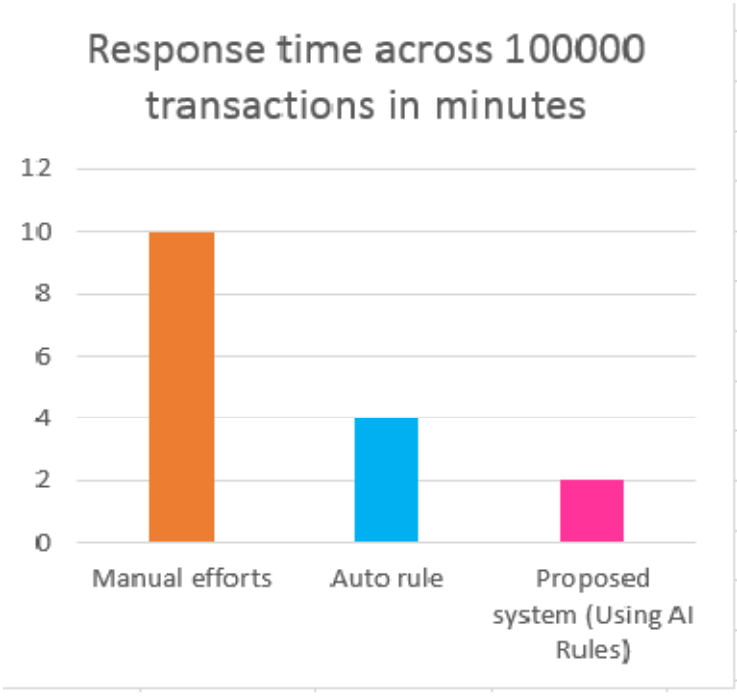

Figure3. Accuracy of detecting fraudulent transactions

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

TABLE I Comparison of the performance of Auto rule process VS AI based system across 1000 transactions

	Auto rule	AI Rule Based
False positive	5	1
Accuracy	60	99
Operation Cost	100	78
Execution Time (sec.)	300	110

From TABLE 1 comparison it can see that system developed with AI rule is more effective in all parameters compared to the existing system.

IV. DISCUSSION AND FUTURE WORK

With the help of few rules that is based on small subset of data it is observed that fraud detection efficiency of the system increases to very large extend. If data from different financial institution gets collected and collaborated with metadata of customer then an effective fraud prevention system can be developed. Machine learning modules like supervised and unsupervised learning along with classification algorithms can easily use to create rule base system. In industry there are many companies working toward implementing fraud prevention solution. The work proposed can be contributed to their goal. In current work we only applied few rules. There can be many dimensions to data which is useful for strengthen system for fraud detection some of them are as follows.

1) Time of transaction.
2) Pattern of transaction.
3) Frequency for transactions.
4) Maximum transaction value per month.
5) Minimum transaction value per month.
6) Frequency of spending outside zip.
7) Average number of transactions for the IP
8) Possible history of the IP address of transactions.
9) Type of transactions for the IP.
10) Spending in the same zip as customer address
11) Variation in addresses.

V. CONCLUSION

This experiment tested several machine learning algorithms as well as meta-learning strategies on real world data. The experiments reported here that effective use of AI rules formed using machine learning can improve efficiency, result and reduces overall efforts.AI rule based system is created with combination of rules across diff rent data dimension of the user. This system is very effective for fraud prevention in finance industry. A combination of the rules increase chances of catching fraud cases and reduces false positive scenarios. AI Rule base system will train itself across the period as dataset will grow and it will have more data to harvest its rules. AI rule based system provides flexibility of deciding level of security that can be implemented to system for fraud prevention.

REFERENCES

[1] Quinlan, J. R. 1993. C4.5: programs for machine learning. San Mated, CA: Morgan Kaufmann
[2] Breiman, L.; Friedman, J. H.; Olshen, R. A.; and Stone, C. J. 1984. Classification and Regression Trees. Belmont, CA: Wadsworth.
[3] Cohen, W. 1995. Fast effective rule induction. In Proc. 12th Intl. Conf. Machine Learning, 115-123.
[4] Clark, P., and Niblett, T. 1989. The CN2 induction algorithm. Machine Learning 3:261-285.
[5] K. Ali and M. Pazzani (1996), Error Reduction through Learning Multiple Descriptions, in Machine Learning, 24, 173-202

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 4, April 2019

[6] L. Breiman, J. H. Friedman, R. A. Olshen, and C.J. Stone (1984), Classification and Regression Trees, Wadsworth, Belmont, CA, 1984
[7] C. Brodley and T. Lane (1996), Creating and Exploiting Coverage and Diversity, Working Notes AAAI-96.
[8] William W. Cohen (1995), Fast Effective Rule Induction, in Machine Learning: Proceeding of the Twelfth International Conference, Lake Taho, California, 1995. Morgan Kaufmann
[9] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth (1996), The KDD Process for Extracting Useful Knowledge from Data, in Communications of the ACM, 39(11): 27-34, November 1996
[10] J.R. Quinlan (1986), Induction of Decision Trees, Machine Learning, 1: 81-106
[11] Philip K. Chan and Salvatore J. Stolfo (1997) Metrics for Analyzing the Integration of Multiple Learned Classifiers (submitted to ML97).Available at: www. Cs. Columbia, edu/-sal/JAM/PROJECT
[12] Chan, P., and Stolfo, S. 1993. Meta-learning for multi strategy and parallel learning. In Proc. Second Intl. Work. Multi strategy Learning, 150-165.
[13] Workshop Integrating Multiple Learned Models (pp. 8- 14)
[14] Toward Scalable Learning with Non-uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection.
[15] Philip K. Chan and Salvatore J. Stolfo (1993), Toward Parallel and Distributed Learning by Meta-learning, in Working Notes AAAI Work. Knowledge Discovery in Databases (pp. 227-240)
[16] Email age system
[17] Cybersource.
[18] Experian.
[19] LexisNexis.