DOI: https://doi.org/10.53555/nncse.v2i3.481 **Publication URL: http://nnpub.org/index.php/CSE/article/view/481**

CREDIT CARD FRAUD PREDICTION SYSTEM

Mr. Alexander Mathew[1]*, Ms. Gayatri Borade[2], Mr. Ketan Bende[3], Ms. Neha Singh[4]
*[1234]*Information Technology, DBACER, Wanadongri, Nagpur, India*

Corresponding Author: -

Abstract: -

Identity crime is common, and pricey, and credit card fraud is a specific case of identity crime. The existing systems of known fraud matching and business rules have restrictions. To remove these negative aspects in real world, this paper proposes a data mining approach: Communal Detection (CD) and Spike Detection (SD). CD finds real social relationships to reduce the suspicion score, and is impervious to fake social relationships. This approach on a fixed set of attributes is whitelist-oriented. SD increases the suspicion score by finding discrepancies in duplicates. These data mining approaches can detect more types of attacks and removes the unnecessary attributes.

Keywords: - *Fraud Prediction, Communal Detection (CD), Spike Detection (CD)*

INTRODUCTION

The data mining consists of multiple algorithms for detection. Data mining algorithms are used in the online credit card application for fake detection. SD and CD algorithms are used in this system. These algorithms are used to detect the fraud and save the data in the database as original data or rejected database which is updated manually by the system. This system does not give a chance to defaulters in credit card application. Identity crime is defined as largely as feasible in this method. At one extreme, real identity theft refers to illegal use of innocent people's complete characteristics. These can be harder to obtain (although large volumes of some identity data are widely available) but easier to successfully apply. In actuality, identity crime can be committed with a mix of both synthetic and real identity details. Credit applications are Internet or paper-based forms with written requests by potential customers for smart cards, mortgage loans. Credit application deception is a specific case of fraud, involving fake identity fraud. As in identity crime, the credit application fraud has reached a critical mass of defaulters who are highly experienced, organized, and sophisticated. There are two types of duplicates: exact (or identical) Duplicates have the all same values; near (or approximate) duplicates have some same values, some identical values with a small different spelling, or both. In short, the new methods are based on White-listing and Detecting spikes of similar applications. White-listing uses existent common relationships on a fixed set of values. This lowers false positives by reducing some suspicion scores. Detecting spikes in duplicate, on a variable set of attributes. This increases true positives by adjusting suspicion scores appropriately.

OBJECTIVE OF THE WORK

Data mining is concerned with analysis of large volumes of data to discover without human intervention interesting regularities or relationships which in turn leads to better understanding of the underlying processes. The data mining consists of multiple algorithms, some algorithms uses for the detection of crime in smart card. Online credit card application uses these algorithms communal and spike detection uses to detect the multiple applicant and with the artificial intelligent CD and SD algorithm uses to make the fraudulent data in the black list.

EXISTING SYSTEM

There are non-data mining layers of defense to protect against credit application scam, each with its uniqueness. Business rule is the hundred-point physical identity check test which requires the applicant to provide sufficient point-weighted identity documents head to head. They must add up to at least 100 %, where a passport is worth 70 %. Another business rule is to contact (or investigate) the applicant over the web. The above two regulations are highly efficient, but human resource demanding. To depend less on HR, a ordinary rule is to compare an application's identity number, address, or any other details against external databases. This is suitable, but the public directories, and registers and completeness. Moreover scorecards for credit scoring can catch a small percentage of fraud which does not look creditworthy; but it also removes outlier applications which have a higher probability of being fake. The second existing defense is known fraud toning. Here, known frauds are complete applications which were confirmed to have the intent to defraud and usually periodically recorded into a blacklist. Consequently, the current applications are matched against the blacklist. This has the benefit and clarity of perception because patterns often replicate themselves. However, there are two main problems in using known frauds.

PROPOSED SYSTEM

In the proposed credit application fraud detection domain, the new layers will improve detection of fraudulent applications because the detection system can detect more types of attacks, better account for changing lawful behavior, and remove the unnecessary attributes. These new layers are not HR concentrated. They represent patterns in a score where the higher the score for an application, the higher the suspicion of fraud. In this way, only the highest scores require human intervention these two new layers, CD and SD, do not use peripheral databases, but only the credit application database. The two greatest challenges for the data miningbased layers of defense are adaptively and use of exceptional data. These confronts need to be addressed in order to reduce wrong outcomes. Adaptivity accounts for changing deceptive behavior, as the attempt to observe fraud changes its performance. But what is not clear, yet equally significant, is the need to also account for changing lawful behavior within a changing surrounding. In the credit application area, changing lawful behavior is exhibited by relationships and can be caused by other events. This means lawful behavior can be hard to tell apart from fake behavior, but it will be shown later that they are indeed distinguishable from each other. The detection system needs to implement caution with applications which reflect relationships. It also needs to consider certain external events. Quality data are highly popular for data mining and data quality can be improved through the real time removal of data errors (or noise). The detection system has to filter duplicates which have been reentered due to human error or for other causes. It also needs to ignore unnecessary attributes which have many issues.

METHODS

Communal Detection

Suppose there were two credit card applications that provided the same address, phone number, and
DOB, but one stated the applicant's name to be Alex Smith, and the other stated the applicant's name to be Plex Smith. These applications could be construed in three ways:

1) It is an impostor attempting to obtain multiple credit cards using near duplicated data
2) Or there are twins living in the same house, both are applying for a credit card.
3) Or it can be the same person applying twice, and there is a typo of one character in the first name.

With the CD layer, any two similar applications could be easily interpreted as (1) because this paper's detection methods use the similarity of the current application to all prior applications (not just known frauds) as the SS (suspicion score). However, for this particular situation, CD would also recognize these two applications as either (2) or (3) by lowering the SS due to the higher possibility that they are lawful. To account for lawful behavior and data errors, Communal Detection (CD) is the whitelist-oriented approach on a fixed set of aspects. The whitelist, a list of Communal and self relationships between applications, is critical because it reduces the scores of these lawful behaviors. Communal relationships are near duplicates which reflect the social relationships from tight familial bonds to casual acquaintances: family members, acquaintances and friends. The family member relationship can be further broken down into more detailed relationships such as spouses, father-son, cousins as well as uncle niece. Relationships highlight the same applicant as a result of legitimate behavior. Broadly speaking, the whitelist is created by ranking link-types between candidates by amount. The larger the amount for a link-type, the higher the probability of a mutual relationship.

Inputs
vi (current application)
W number of *vj* (moving window)
$^{S}Rx,link-type$ (link-types in current whitelist)
Tsimilarity (string similarity threshold)
Tattribute (attribute threshold)
η (exact duplicate filter)
α (exponential smoothing factor)
Tinput (input size threshold)
SoA (State-of-Alert)

Outputs
S(*vi*) (suspicion score)
Same or new parameter value
New whitelist

CD algorithm
Step 1: Multi-attribute link [match *vi* against *W* number of *vj* to determine if a single attribute exceeds *Tsimilarity*; and create multi-attribute links if near duplicates' similarity exceeds *Tattribute* or an exact duplicates' time difference exceeds η]
Step 2: Single-link score [calculate single-link score by matching Step 1's multi-attribute links against $^{S}Rx,link-type$]
Step 3: Single-link average previous score [calculate average previous scores from Step 1's linked previous applications]
Step 4: Multiple-links score [calculate *S*(*vi*) based on weighted average (using α) of Step 2's link scores and Step 3's average previous scores]
Step 5: Parameter's value change [determine same or new parameter value through SoA (for example, by comparing input size against *Tinput*) at end of *ux,y*]
Step 6: Whitelist change [determine new whitelist at end of *gx*]

Spike Detection
This part contrasts SD with CD; and presents the need for SD, in order to improve toughness. Before proceeding with a description of Spike Detection (SD), it is necessary to strengthen that CD finds real social relationships to reduce the SS, and is secure to artificial relationships. It is the whitelist related approach on a fixed set of characteristics. In comparison, SD finds spikes to increase the SS. Probe-resistance reduces the chances of impostor will discover characteristics used in the SD calculation. It is the characteristicoriented approach on a variable-size set of characteristics. SD cannot use a whitelist oriented approach because it was not designed to create multi-attribute links on a fixed-size set of attributes. CD has a limitation in its character threshold. CD must match at least three values for our dataset. With less than three corresponding values, our whitelist does not contain valid relationships because values, such as any given attribute and serial number, are not unique identifiers. The impostor can duplicate one or two important values which CD cannot detect. SD complements CD. The unnecessary characteristics are either too negligible where no patterns can be found, or too much where no values can be found. The unnecessary characteristics are continually filtered, only selected characteristics in the form of not-too-sparse and not too - intense characteristics are used for the SD. In this way, the revelation of the detection system to prying of characteristics is reduced because only one or two characteristics are adaptively selected. Suppose there was a bank's advertising campaign to give striking benefits for its new business titanium credit card. This will cause a spike in the number of lawful credit card applications by businessmen, which can be incorrectly interpreted by the system as an impostor attack. To account for the changing lawful behavior caused by other events, SD strengthens CD by providing character weights which reflect the degree of significance in characteristics. The characteristics are adaptive for CD in the sense that its characteristic weights are frequently determined. This addresses other events such as the entry of new society and exit of existing ones, and advertising campaigns of organizations which do not contain any patterns and are likely to cause three natural changes in characteristic weights. These changes are quantity drift where the overall quantity changes, population drift where the volume of both fraud and lawful classes fluctuate independent of one another, and idea drift which involves changing lawful characteristics that can become similar to fake characteristics. By changing characteristic weights, the detection system makes allowance for these other events. In general, SD trades off

effectiveness (degrades security because it has more anomalies without filtering out communal relationships and some data errors) for efficiency (improves computation speed because it does not match against the whitelist, and can calculate each characteristic in parallel on multiple workstations). In contrast, CD trades off efficiency for effectiveness.

Inputs
vi (current application)
W number of *vj* (moving window)
t (current step)
Tsimilarity (string similarity threshold)
θ (time difference filter)
α (exponential smoothing factor)
Outputs
S(*vi*) (suspicion score)
wk (attribute weight)

SD algorithm
Step 1: Single-step scaled counts [match *vi* against *W* number of *vj* to determine if a single value exceeds *Tsimilarity* and its time difference exceeds θ]
Step 2: Single-value spike detection [calculate current value's score based on weighted average (using α) of *t* Step 1's scaled matches]
Step 3: Multiple-values score [calculate *S*(*vi*) from Step 2's value scores and Step 4's *wk*]
Step 4: SD attributes selection [determine *wk* for SD at end of *gx*]
Step 5: CD attribute weights change [determine *wk* for CD at end of *gx*]

SYSTEM ARCHITECTURE

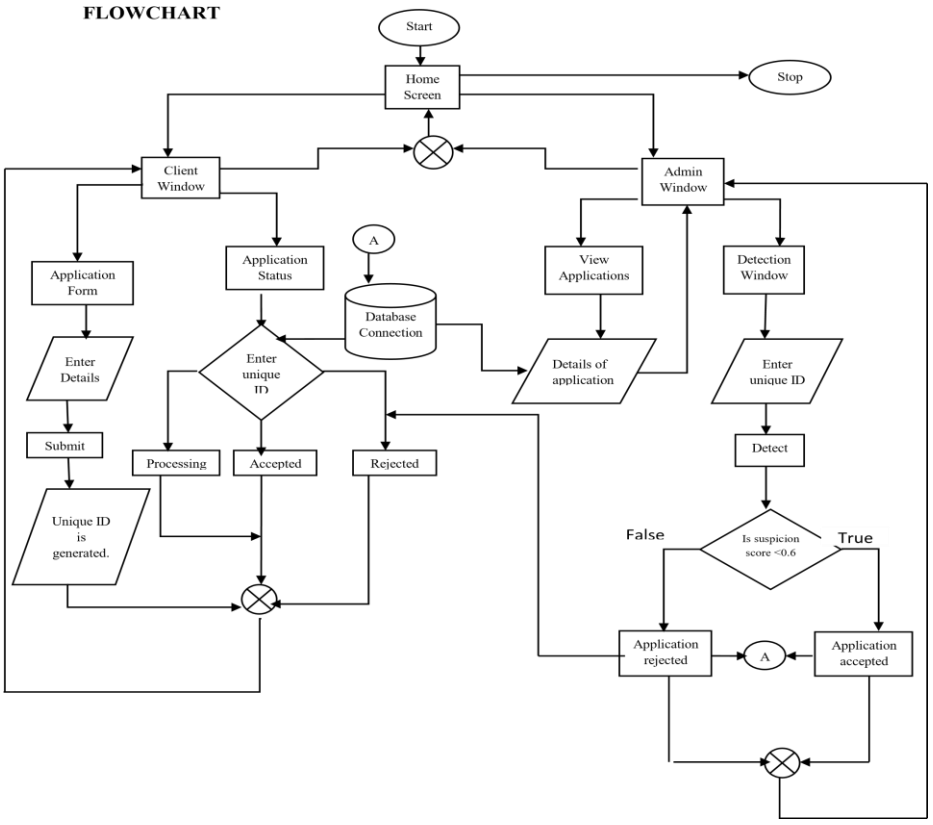
The architecture represents the overall structure of the system. The data is detected for the crime detection using the data mining algorithm communal detection and spike detection algorithm. These two algorithms combine together to remove the negative false and then proceeded to the proposed system algorithm. This algorithm retrieved and diagnosis the datum. If the data is fraud it is thrown into the black list database. If the data is original the data is stored in the database. The communal detection focused on attacks in the white list by fraudsters when they submit applications with synthetic relationship. The volume and ranks of the white list's real communal relationships change over time, to make the white list exercise caution with (more adaptive) changing legal behavior, the white list is continually being reconstructed. The spike detection is attribute oriented. It cannot be detected by fraud attribute will be updated regularly. The attributes used in spike detection will not be communal detection. By using the spike detection and communal detection detects the fraudsters in credit card application In addition to communal detection and spike detection we use case based reasoning algorithm to make this approach more efficient. CBR implements retrieval, diagnosis and resolution to make the data more secure. The CBR used to analyze and retrieval of data from the existing blacklist. The fraudulent datum is moved to the blacklist and the original datum is stored in the database.

CONCLUSION

The system detects fraud detection online credit card application. This system is used avoid the duplicates from the fraudsters while applying the credit card. Data mining algorithms are used this system. The existing algorithm communal detection and spike detection used to detect the multiple applicants. In proposed system combing with the existing algorithm spike detection and communal detection the algorithm is used to make the system more efficient and secure. The Algorithm is used to throw the fraudulent data in the blacklist and retrieve the datum from the blacklist database. The identity thief has limited time because innocent people can discover the fraud early and do the necessary tasks, and will rapidly use the same real identities at different places.

REFRENCES

[1]."*Resilient Identity Crime Detection*" by Clifton Phua, Kate Smith-Miles,Vincent Lee, and Ross Gayler.
[2]."*Adaptive Fraud Detection*" by Tom Fawcett, Foster Provost.
[3]."*A Taxonomy of Fraud and Fraud Detection Techniques*" by NaeimehLaleh, Mohammad AbdollahiAzgomi.
[4]."*Credit Card Fraud Prediction Uses Bayesian and Neural Networks*" by Sam Maes, Karl Tuyls, Bram Vanschoenwinkel


FLOWCHART
Start
Home Screen
Stop
Client Window
Admin Window
A
Application Form
Application Status
Database Connection
View Applications
Detection Window
Enter Details
Enter unique ID
Details of application
Enter unique ID
Submit
Processing
Accepted
Rejected
Detect
Unique ID is generated.
False
Is suspicion score <0.6
True
Application rejected
A
Application accepted