IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 8, Issue 12, July 2020)

Credit Card Fraud Detection using Bayesian Optimized K-Nearest Neighbors

Deepika Kanungo
deepikakanungo21@gmail.com

Lokesh Parashar
lokesh23324@gmail.com

Abstract – **Credit card fraud is one of the most important problems that financial institutions are currently facing. Although the technology has allowed to increase the security in the credit cards with the use of PIN keys, the introduction of chips in the cards, the use of additional keys such as tokens and improvements in the regulation of its use is also a necessity for banks, to act preventively against this crime. To act preventively, it is necessary to monitor in real time the operations that are carried out and have the ability to react in a timely manner against any doubtful operation that is performed. This paper presents an implementation of automatic credit card fraud detection system using Bayesian Optimized K-Nearest Neighbors on Kaggle dataset. The selection of proper attributes for reducing the training overhead and claiming higher accuracy for the fraud detection using soft computing. Performance evaluation is achieved using confusion matrix plot with accuracy, sensitivity and precision values.**

Keywords – **Artificial Intelligence, Bayesian Optimization, CIA, Data Mining, FBI, KDD, KNN, Machine Learning.**

I. INTRODUCTION

After the attacks of September 11, 2001, agencies such as the CIA and the FBI increased their intelligence blocks with one main purpose: to find information related to terrorist groups. On the other hand, the most used techniques in this process are:

- Geographical-visual techniques for hot zone detection [1].
- Standard Deviation Ellipses, by means of which groups of facts identified by means of clustering techniques can be delimited.

In addition, there are statistical analysis packages for criminal information, which work on GIS. Some of them are: Spatial and Temporal Analysis of Crime, CompStat and CrimeStat [2]. Meanwhile, techniques such as Concept Space have been implemented by the Artificial Intelligence Laboratory of the University of Arizona, in Tucson, to extract relationships between police information and thus detect possible bands or suspects.

Concept Space relied on the use of data mining and, specifically, on Hierarchical Clustering [3]. It should be noted that the resources offered by these types of techniques have borne fruit; between 1985 and 2002, the United States Government detected 16 key members of large criminal organizations [4].

There are basically two ways of acting of the people who commit this type of crimes: on the one hand the obtaining of the physical card as such and on the other the recording of the data of the magnetic stripe for later use, either through a new card or using the data in purchases made through the Internet.

In the first case, in which the criminals obtain the physical card, one way of obtaining it discreetly in order to commit their crime is as follows:

- In the slot, where the card must be inserted, a new slot is placed that will take a stop so that the card, when inserted, does not reach the cashier. In this way, the card has been caught, as shown in figure 1.

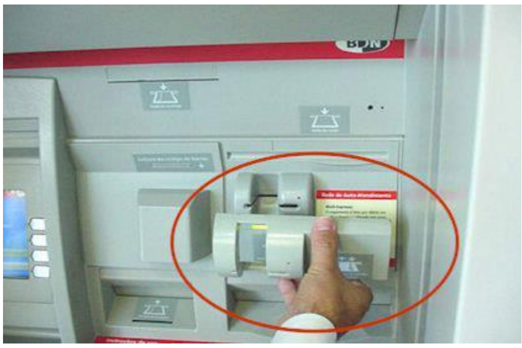

Figure 1: Altering the slot of an ATM

- Taking advantage of the fact, one of the criminals will approach the card user and tell him that the same thing has happened to him, and that he must dial a number of numbers and to finish his personal key, that the offender He will be watching and memorizing.

- The next step, once the cardholder has left, confident that they will solve your problem, is that a second offender (complicit in the first) approaches the cashier and remove the card, with what already have the card and the personal key.

Other frequent ways of acting are to obtain the card data and then record it in another to be able to operate with it. There are a multitude of readers / recorders of magnetic strips on the market, which make this task easier for criminals.

In all cases it is not necessary to make a physical or physical copy of the credit card to carry out a fraudulent use of it, you can make purchases through the Internet, that is, through electro trade only using the card number and expiration date. This way of buying with the credit card is one of the most widespread.

Information Exploitation (Data Mining) is the process by which understandable and useful knowledge - previously unknown - is extracted from databases, in various formats and automatically. Then, Information Exploitation poses two challenges: working with large databases and applying techniques that automatically convert these data into knowledge [5].

Likewise, Data mining is a fundamental element for a broader technique whose objective is to discover knowledge in large databases (Knowledge Discovery in Databases —KDD) [6] [7].

The further development of the use of Information Exploitation in activities related to systems auditing has to do with the detection of intruders in telecommunications networks. Even in the scientific literature there are antecedents linked to the location of fraud using data mining [8].

This text refers to a specific case of fraud associated with credit cards and commonly known as the card cloning, a circumstance that represents a risk for clients attached to a bank.

Humans, having cognitive ability, develop a series of behaviours that can be defined as pattern depending on certain situations. In turn, the moment in which a crime is committed is no exception; a group of psychologists determined that there are patterns of behaviour associated with factors such as location, time of day and temperature. Such information, managed through data mining, allows us to develop a predictive model of ideal situations - scenarios - where a crime could happen. For the cited example, three scenarios are identified that are identified with the aforementioned sponsors: bicycle theft, firearm theft and wallet theft [9].

Consequently, the development of this predictive tool generates a positive impact on society since it allows - to the forces of public order - to have faster reaction times and thus avoid being delayed by reaching the scenes of crime. However, it can also generate a negative impact if a citizen is wrongly prejudged due to poor system documentation (falsification of public documents, for example) [9].

The manual and technical review of fraud prevention does not detect some of the most prevalent patterns such as the use of a credit card several times, in multiple locations (physical or digital) and in a short time [10].

This paper develops a framework for automatic credit card fraud detection using Bayesian optimized K-Nearest Neighbors classifier of Kaggle dataset. Rest of paper is organized as follows. Section 2 presents the proposed methodology followed by the simulation and results in section 3 and finally the conclusion and future aspects are detailed in section 4.

II. PROPOSED METHODOLOGY

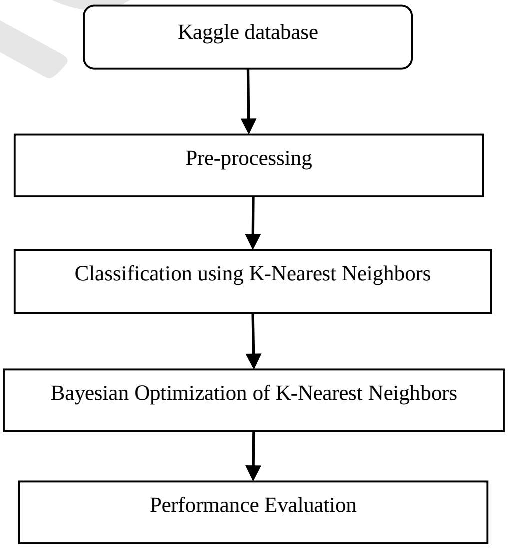
Kaggle database
Pre-processing
Classification using K-Nearest Neighbors
Bayesian Optimization of K-Nearest Neighbors
Performance Evaluation

Figure 2: Flow diagram for proposed credit card fraud detection system

A. Dataset Description

The Kaggle datasets contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions [11].

B. *Pre-Processing*

- Create a response variable in a categorical form where:
  1= fraudulent activity
  0 = non-fraudulent or normal activity
- Remove Time attribute from categorical class

C. *Classification Algorithms*

1) K-Nearest Neighbor (KNN)

The K-nearest neighbor algorithm is simple, but who can give interesting results if the data range is large enough. It's about a classification method widely used in many fields and is also found among the top 10 data mining algorithms [12]. Typically, houses that are close to each other have similar characteristics. We can group them and give them a classification. The algorithm uses this same logic to try to group the elements that are close to each other. KNN is an example of instance-based learning. It operates in situations where each instance can be defined by a vector of n dimensions, where n is the number of attributes used to describe each instance and the classifications are discrete values. The training data is stored and, when a new instance is found, it will then be compared to the training data to find its nearest neighbors.

The nearest neighbors are those who are closest following the Euclidean distance. The distance between two elements $A = \langle a_1, \ldots, a_n \rangle$ and $B = \langle b_1, \ldots, b_n \rangle$ is calculated as follows:

$$d = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \qquad (1)$$

Sorted by the k neighbors closest to the new instance, the classification assigned to it will be the class with the highest occurrence among them.

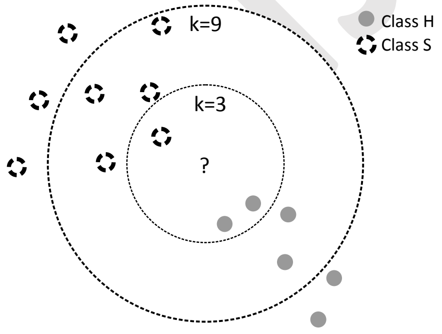

Figure 3: Operation of a simple KNN with k = 3 and k = 9

Here is the pseudo-code representing the algorithm:

Pseudo-Code for KNN

```
Requires 3 parameters: a set of examples X, a given x and k $\epsilon\{1, \ldots k\}$
For each example $x_i \epsilon X$
Calculate the distance between $x_i$ and $x$: $\delta(x_i, x)$
End for
For $j\epsilon\{1, \ldots k\}$ do
    $KNN(j) \leftarrow \arg\min \delta(x_i, x) i \in 1, \ldots n$
            $\delta(x_i, x) \leftarrow +\infty$
End for
Determine the class of $x$ from the class of examples whose number is stored in the KNN.
```

2) Classification by using Bayesian Optimized KNN

First of all, three parameters are to be taken into consideration: the sample data, the number of nearest neighbors to select (k), and the point we want to evaluate (x). Subsequently, for each element of the sample, we evaluate the distance between reference point X and point x; of the set of learning and we check if the distance between them is less than one of the distances contained in the list of nearest neighbors. If so, the point is added to the list. If the number of items in the list is more significant than k, the last value is simply removed from the list. The algorithm itself is not very complicated and can give a good result with brute force if sampling is not too big. However, since we are talking about data mining, the number of individuals to be evaluated is often very big, that's why an optimization algorithm is needed.

The main idea of Bayesian Optimization (BO) is to construct a surrogate probabilistic model sequentially to try to infer the objective function. Iteratively, new observations are made, and the model is updated, reducing its uncertainty allows working with a known and cheaper model, which is used to construct a utility function that determines the next point to evaluate. The different steps of the BO methodology are described below.

First, the apriori model must be chosen over the possible space of functions. For this, different parametric approaches can be used, such as Beta-Bernoulli Bandit or Linear Models (Generalized), or non-parametric models such as t-Student Processes or Gaussian processes [13].

Then repeatedly until a particular stopping criterion [14]:

The prior and the likelihood of the observations so far are combined to obtain a posterior distribution. This is done using Bayes' theorem, hence the origin of the name.

Recall Bayes' theorem. Let A and B be two events such that the conditional probability $P(B|A)$ is known, then the probability $P(A|B)$ is given by:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad (2)$$

Where $P(A)$ is the a priori probability, $P(B|A)$ is the probability of event B conditional on the occurrence of event A, and $P(A|B)$ is the posterior probability. Then, a particular utility function is maximized on the a posteriori model to determine the next point to evaluate and the new observation is collected to repeat until the stop criterion.

Since the KNN approach uses a discretization technique for the continuous parameter, therefore it results in less accurate results with the data loss. This proposed work discusses the algorithm that can tune KNN parameter.

The algorithms that are proposed in this paper are related to optimum value for K. The parameters are i) weight, C; and ii) kernel function. The weight represents the trade-off between specific misclassifying points and correctly classifying others, while the kernel is used to instantaneously tune KNN parameter and select the feature subset.

BO-KNN Algorithm

```
Input: n, m, C, γ, and termination criterion
Output: Optimal value for KNN parameter
Begin
Initialize n solutions
call KNN algorithm to evaluate n solutions
T = Sort ($k_1, \dots, k_n$)
while number of iteration ≠10 do
for i = 1 to m do
select k according to its weight
sample selected k
store newly generated solutions
call KNN algorithm to evaluate newly generated solutions
end
T = Best (Sort $k_1, \dots, k_{n+m}$), n)
end
End
```

In the algorithm, n is the size of solution archive, m is the number of models that are used to generate solutions, C is the regularization or soft margin parameter, γ is the kernel function parameter called the margin or the width parameter, and finally, the termination conditions for the best values for KNN parameter (k).

III. Simulation And Results

Here, TP=174, TN=142122, FP=72 and FN=35

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{174+142122}{178+142122+72+35} = 99.9\%$$
$$Precision = \frac{TP}{TP+FP} = \frac{174}{174+72} = 70.73\%$$
$$Recall\ or\ Sensitivity = \frac{TP}{TP+FN} = \frac{174}{174+35} = 83.3\%$$

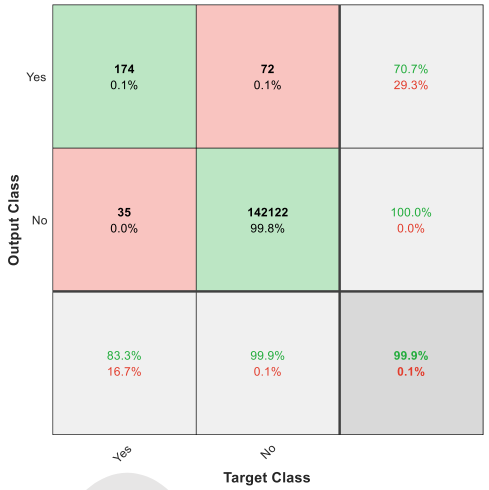

Figure 4: Confusion matrix plot for proposed credit card fraud detection using Bayesian optimized KNN

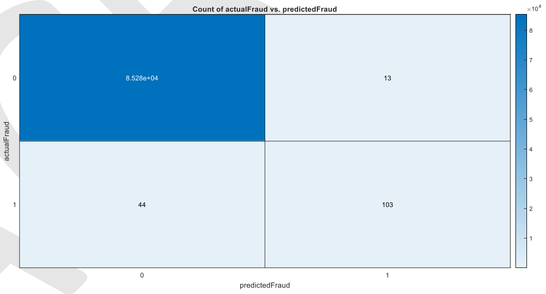

Figure 5: Count of actual fraud vs. predicted fraud

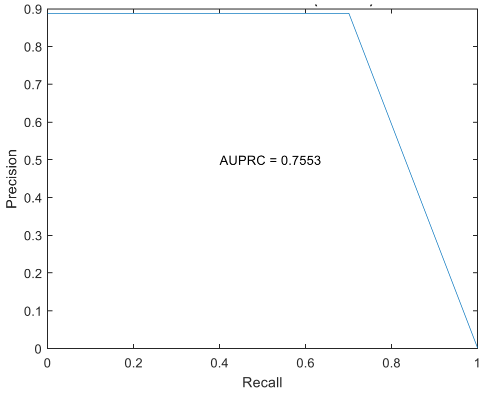

Figure 6: Precision-Recall curve

IV. Conclusion

This paper has examined the performance of Bayesian optimized k-nearest neighbors classifier. The Kaggle dataset on credit card transactions is used in this paper. This work achieves maximum accuracy of 99.9%.

Although the proposed method obtains good results on small set data, there are still some problems such as imbalanced data. Our future work will focus on solving these problems. The proposed algorithm itself should be improved. For example, the voting mechanism assumes that each of base classifiers has equal weight, but some of them may be more important than others. Therefore, we also try to make some improvement for this algorithm.

REFERENCE

[1] Eck, John, Spencer Chainey, James Cameron, and Ronald Wilson. "Mapping crime: Understanding hotspots." (2005): 1-71.

[2] Block, Carolyn Rebecca, and Icjia Senior. "Illinois criminal justice information authority." (1985).

[3] Levine, Ned. "CrimeStat III: a spatial statistics program for the analysis of crime incident locations (version 3.0)." *Houston (TX): Ned Levine & Associates/Washington, DC: National Institute of Justice* (2004).

[4] Chen, Hsinchun, Wingyan Chung, Jennifer Jie Xu, Gang Wang, Yi Qin, and Michael Chau. "Crime data mining: a general framework and some examples." *computer* 4 (2004): 50-56.

[5] Britos, Paola, Oscar Dieste, and Ramón García-Martínez. "Requirements elicitation in data mining for business intelligence projects." In *IFIP World Computer Congress, TC 8*, pp. 139-150. Springer, Boston, MA, 2008.

[6] Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases." *AI magazine* 17, no. 3 (1996): 37-37.

[7] Britos, Paola, Hernan Grosser, Dario Rodríguez, and Ramon Garcia-Martinez. "Detecting Unusual Changes of Users Consumption." In *IFIP International Conference on Artificial Intelligence in Theory and Practice*, pp. 297-306. Springer, Boston, MA, 2008.

[8] Gunderson, L. F. "Using data mining and judgment analysis to construct a predictive model of crime." In *IEEE International Conference on Systems, Man and Cybernetics*, vol. 7, pp. 5-pp. IEEE, 2002.

[9] Brown, Donald E., and Rosemary B. Oxford. "Data mining time series with applications to crime analysis." In *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat. No. 01CH37236)*, vol. 3, pp. 1453-1458. IEEE, 2001.

[10] Yee, Ong Shu, Saravanan Sagadevan, and Nurul Hashimah Ahamed Hassain Malim. "Credit card fraud detection using machine learning as data mining technique." *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 10, no. 1-4 (2018): 23-27.

[11] Credit Card Fraud Detection Dataset. Online available at: https://www.kaggle.com/mlg-ulb/creditcardfraud

[12] Gazalba, Ikbal, and Nurul Gayatri Indah Reza. "Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification." In 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), pp. 294-298. IEEE, 2017.

[13] Acuna, Daniel, and Paul Schrater. "Bayesian modeling of human sequential decision-making on the multi-armed bandit problem." In Proceedings of the 30th annual conference of the cognitive science society, vol. 100, pp. 200-300. Washington, DC: Cognitive Science Society, 2008.

[14] Pelikan, Martin, David E. Goldberg, and Erick Cantú-Paz. "BOA: The Bayesian optimization algorithm." In Proceedings of the genetic and evolutionary computation conference GECCO-99, vol. 1, pp. 525-532. 1999.