

Detect Credit Card Fraud with Machine Learning

Anna Helna¹ and Laiby Thomas²

Student, CSE, Naipunnya Institute of Management and Information Technology, Thrissur, India¹

Assistant Professor, CSE, Naipunnya Institute of Management and Information Technology, Thrissur, India²

Abstract: *This Credit card fraud is becoming a common problem today. Digital world made people to rely on credit cards and debit cards and it in turn became a threat. Machine learning is a new approach to find a solution for this problem. It can automatically detect fraud characteristics. By using a person's historical data, the user's pattern and behavior can be analyzed and can determine whether the transaction is fraudulent or not. There are many approaches are there for detecting the credit card fraud. Random Forest Algorithm, K-Nearest Neighbored and K-Means Clustering are among them. In this paper, various techniques analyzed and compared for finding the best solution.*

Keywords: Machine Learning, Credit Card, Fraud Detection, Learning Techniques, etc.

I. INTRODUCTION

The people in the twenty first century live in a digitalized world that comprises with the knowledge of computer for the ease of living. The money transactions became digitalized and there came the use of credit cards. Credit card is a electronic payment system that are used for the non-cash transactions. The bank or financial institution issue credit card to a consumer to facilitate payment to a merchant of goods and services [1]. Credit score card typically refers to a card this is assigned to the purchaser (cardholder), commonly permitting them to purchase items and services inside credit restrict or withdraw coins in advance. Credit card presents the cardholder and gain of time, i. e., it presents time for their clients to pay off later in a prescribed time, by using sporting it to the subsequent billing cycle. [2]

A credit card is considered fraudulent when another person uses your credit card for you without your authorization. Fraudulent steal the credit card PIN code or account details to perform one of the unauthorized transactions without robbery of the original physical card. Using credit card fraud detection, we could find out if new transactions are fraudulent or bona fide [3].

In 2017, there had been 1,579 facts breaches and nearly 179 million statistics amongst which credit card frauds have been the most common shape with 133,1/2reviews, then employment or tax-related fraud with 82,051 reports, telephone frauds with 55,1/2reports followed through bank frauds with 50,517 reviews from the statics released via FTC. [2]

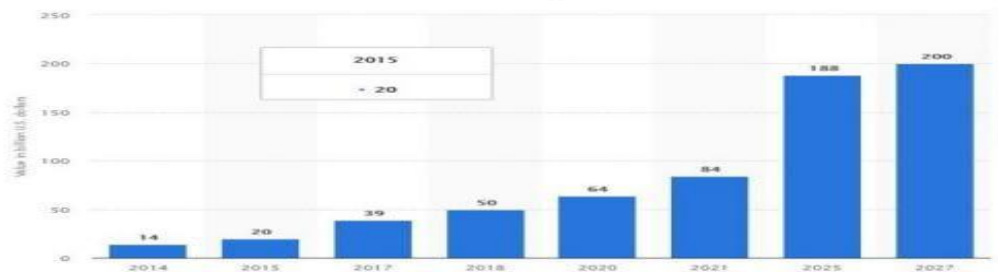


Figure 1: Growth of E-Commerce Sites [3]

The fraud this is devoted may involve the card which include a credit score card or debit card. In this, the cardboard itself acts as a fraudulent source within the transaction. The reason of committing the crime may be to achieve the goods without paying cash or to gain the unauthorized fund. Credit score cards are a nice goal for

fraud. The motive is that in a completely quick time a variety of money can be earned without taking many risks or even the crime will take many weeks to be detected [3].

As the usage of the internet these days [fig.1] may be very a good deal growing there may be many possibilities for the fraudsters to commit the credit card frauds. The principles fraud instances which are in those of the e-commerce websites. In the gift generation, human beings are displaying a good deal hobby in getting matters on-line instead of going and purchasing them, and due to this, the growth of the e-trade websites is growing and thereby there's a huge hazard of credit card fraud. With the intention to keep away from such credit score card frauds, we want to discover the first-rate set of rules that reduces credit card frauds [3].

II. METHODS FOR DETECTING FRAUD

Methods for credit card fraud detection with quite a few studies strategies and numerous fraud detection strategies with a special interest in the neural networks, facts mining, and allocated statistics mining. Many different techniques are used to locate such credit score card fraud. Whilst accomplished the literature survey on various strategies of credit card fraud detection, we able to conclude that to hit upon credit score card fraud there are numerous other methods in machine getting to know itself.

A. Random Forest Algorithm

Random forest algorithm is one of the broadly used supervised learning algorithms. This could be used for each regression and class purposes. However, this set of rules is especially used for classification problems. Normally, a forest is made from bushes and in addition, the Random Forest algorithm creates the selection trees on the pattern facts and receives the prediction from every of the pattern facts. Then Random Forest of rules is better than decision trees due to the fact it reduces the over-becoming by using averaging the end result. [3]

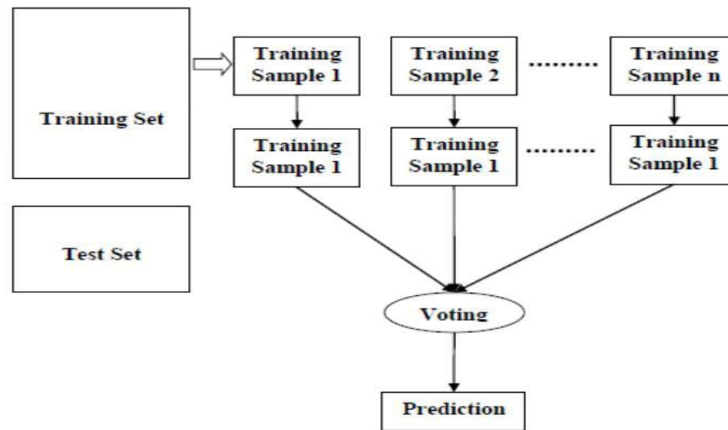


Figure 2: Random Forest Algorithm

Steps of Random Forest Algorithm

- Assume the Kaggle praise card extortion dataset that is prepared and arbitrarily select a portion of the example information.
- Utilizing the arbitrarily made example information presently makes the Decision trees that are utilized to characterize the cases into the misrepresentation and non-extortion cases.
- The Decision trees are framed by dividing the hubs; the hubs which have the most noteworthy data gain make it as the root hub and characterize the misrepresentation and non-extortion cases.
- Presently the larger part vote is performed and the Decision trees might bring about 0 as result which incorporates that these are the non-extortion cases.
- At last, we track down the exactness, accuracy, review, and F1 – score for both the extortion and non-misrepresentation cases.

B. K-Nearest Neighbour

A basic, simple to-carry out managed AI strategy that utilizes ordered information to create a work that gives a reasonable result when given extra unlabelled information. Both characterization and relapse issues can be addressed with the K-Nearest Neighbours’ (KNN) calculation, which is fast and clear to apply. Utilizes marked information to show a capacity that produces an OK presentation for new information. In the KNN calculation, the similarity between the new case and the case and the cases that are currently arranged is determined. When the new case is set in a class that is most similar to the gathering. In a comparable to design, KNN arranges every single open datum and orders new focuses relying upon how comparative they are. This depicts whenever new information arises, it is simply a question of fitting a k-N characterization plan to it. The calculation gets altogether slower as indicators/autonomous factors increment. As displayed in the figure underneath: [1]

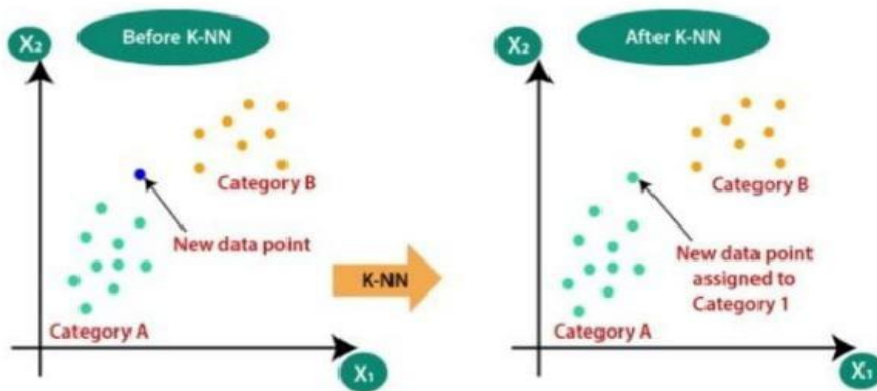


Figure 3: General Structure Working of the KNN[1]

C. K-Means Clustering

As a result of its effortlessness and adequacy, it is the most broadly utilized unaided learning technique. By computing the mean distance between main informative elements, this strategy assigns focuses to gatherings. It then, at that point, refreshes this cycle to work on the exactness of it classifies over the long haul.

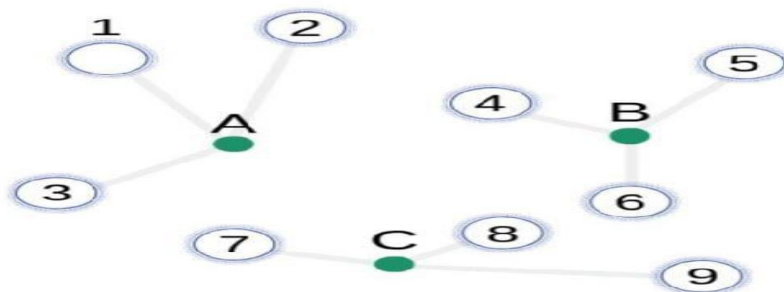


Figure 4: General structure working of the K-MC[1]

III. TECHNIQUES COMPARISON

Random Forest Algorithm

Advantage

- Decreases over fitting in choice trees and assists with working on the exactness
- It is adaptable to both order and relapse issues
- It functions admirably with both all out and nonstop qualities
- It mechanizes missing qualities present in the information
- Normalizing of information isn’t needed as it utilizes a standard based methodology

Disadvantage

- It requires a lot of computational power as well as assets as it fabricates various trees to consolidate their results.
- Additionally, requires a lot of time for preparing as it consolidates a ton of choice trees to decide the class.
- Because of the group of choice trees, it additionally endures interpretability and neglects to decide the meaning of every factor.

K- Nearest Neighbour

Advantage

- KNN is called Lazy Learner (instance-based learning). It advances nothing in the preparation period. There is no preparation period. It stores the preparation dataset and gains from it just at the hour of making constant forecasts.
- New information can be added without affecting the calculation execution or precision
- KNN Algorithm is extremely simple to carry out. You really want just two information

Disadvantage

- Execution issue with huge informational collection: the time expected to compute the distance between the new point and each current focus is immense. This then, at that point, corrupts the presentation of the calculation.
- Doesn't function admirably with high aspects/features: the KNN calculation doesn't function admirably with huge no of features layered information in light of the fact that with enormous number of features, it becomes challenging for the calculation to ascertain the distance in each feature.
- Worth of K: it is truly critical to figure out what worth to appoint to K. with various K you come by various outcomes

K-Means Clustering

Advantage

- Rehashed method.
- Effective and quick
- Chips away at ordered advanced information.

Disadvantage

- Load of repeats.
- Need to choose you have K worth.

Should comprehend the instance of your information well.

IV. DISCUSSION OF COMPARISON

Through this paper find the advantages and disadvantages of each algorithm. Each algorithms advantage and disadvantages are different each other. That shows the one is more effective to detect fraudulent transaction. A number of demanding situations are related to credit score card detection, namely fraudulent behavior profile is dynamic, this is fraudulent transactions generally tend to seem like valid ones; credit score card transaction datasets are hardly ever to be had and particularly imbalanced (or skewed); most appropriate feature (variables) choice for the models; appropriate metric to assess overall performance of techniques on skewed credit score card fraud data. Credit card fraud detection overall performance is substantially suffering from kind of sampling approach used, choice of variables and detection technique(s) used.

V. CONCLUSION

Credit card fraud is become a threat to the people in this digitalized world. For detecting credit card fraud, we, Machine Learning Algorithm is used. Random Forest and K-Nearest Neighbour and K-Means Cluster are different approaches that are used for this. These algorithms help the credit card companies to identify the fraudulent transactions more accurately within a short time and low cost. All this technique has its own advantages and disadvantages. Based on the company need one can choose any one of these approaches. This kind of comparative study will allow the people to build a hybrid approach most accurate for fraudulent credit card transaction detection.

REFERENCES

- [1] Survey Paper On Credit Card Fraud Detection Techniques. Fayyomi, Aisha Mohammad, Eleyan, Derar and Eleyan, Amina. 09, SEPTEMBER 2021, INTERNATIONAL JOURNAL OF SCIENTIFIC&TECHNOLOGY RESEARCH, Vol. 10, pp. 72-79.
- [2] Credit Card Fraud Detection Using Machine Learning Alorithms. Dornadula, Vaishnavi Nath and S, Geetha. Chennai,India : s.n., 2019, INTERNATIONAL CONFERENCE ONRECENT IN ADVANCED COMPUTING, pp. 632-641.
- [3] Credit Card Fraud Detection Using Machine Learning. Sailusha, Ruttala, et al. 2020, International Conference on Intelligent Computer and Computer System, pp. 1264-1270.
- [4] K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint. <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>.
- [5] <https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning>
- [6] <https://www.yadsmic.com/post/k-nearest-neighbors-advantages-and-disadvantages>
- [7] <https://www.mygreatlearning.com/blog/random-forest-algorithm>
- [8] https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.html