BHAGALPUR COLLEGE OF ENGINEERING
SABOUR, BHAGALPUR, BIHAR -813210.
(Department of Science and Technology, Government of Bihar)

A mini project on

CREDIT CARD FRAUD DETECTION

Under the guidance of

Mintu Singh

Department of Computer Science & Engineering

Submitted By:-

Name	**Reg. No.**
Rahul Sharma	16105108026
Nitesh Kumar	16105108033
Shivam Kumar	16105108018
Kundan Kumar	16105108010

BHAGALPUR COLLEGE OF ENGINEERING
SABOUR, BHAGALPUR, BIHAR -813210.
(Department of Science and Technology, Government of Bihar)

BONAFIDE CERTIFICATE

This is to certify that this project report entitled "**Credit Card Fraud Detection**" submitted to **Bhagalpur College of Engineering, Bhagalpur** is a Bonafide record of work done by

"**Rahul Sharma**"	Reg. No. 16105108026
"**Nitesh Kumar**"	Reg. no. 16105108033
"**Shivam Kumar**"	Reg. No. 16105108018
"**Kundan Kumar**"	Reg. No. 16105108010

under my supervision during 7th semester (2016 batch).

Mintu Singh
(Assistant Professor)
Department of Computer Science & Engineering

Prof. Mani Kant Mandal
(Head of Department)
Department of Computer Science & Engineering

Bhagalpur College of Engineering,
Sabour, Bhagalpur – 813210, Bihar, India.

DECLARATION BY AUTHOR

This is to declare that this report has been written by me. No part of the report is plagiarized from other sources. All information included from other sources have been duly acknowledged. I aver that if any part of the report is found to be plagiarized, I shall take full responsibility for it.

"**Rahul Sharma**"	Reg. No. 16105108026
"**Nitesh Kumar**"	Reg. no. 16105108033
"**Shivam Kumar**"	Reg. No. 16105108018
"**Kundan Kumar**"	Reg. No. 16105108010

ACKNOWLEDGEMENTS

It is a pleasant duty to express our sincere gratitude to Aryabhatta Knowledge University, Patna for giving us such a wonderful opportunity to complete our project.

We ascribe countless number of thanks to Hon'ble principal, ***Prof. Achintya*** *and TEQIP team, Bhagalpur College of Engineering, Bhagalpur.*

We would like to acknowledge our guide mentor ***Mintu Singh****, who helped us and guided us throughout the course of completion of this project on "Credit Card Fraud Detection".*

Lastly, and most importantly, we thank our parents for their love, support and motivation throughout our life.

We also place on record, our sense of gratitude to one and all who, directly or indirectly, have lent their helping hand in this venture.

"**Rahul Sharma**"	Reg. No. 16105108026
"**Nitesh Kumar**"	Reg. no. 16105108033
"**Shivam Kumar**"	Reg. No. 16105108018
"**Kundan Kumar**"	Reg. No. 16105108010

LITERATURE SURVEY

Throughout the financial sector, machine learning algorithms are being developed to detect fraudulent transactions. In this project, that is exactly what we are going to be doing as well. Using a dataset of nearly 28,500 credit card transactions and multiple unsupervised anomaly detection algorithms, we are going to identify transactions with a high probability of being credit card fraud. In this project, we will build and deploy the following two machine learning algorithms:

- Local Outlier Factor (LOF)
- Isolation Forest Algorithm

1. Local Outlier Factor (LOF)

An outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism.

For any positive integer k, the k-distance of object p, denoted as k-distance (p), is defined as the distance d (p,o) between p and an object o ∈ D such that:
(i) for at least k objects o'∈D\{p} it holds that d(p,o') ≤ d(p,o), and (ii) for at most k-1 objects o'∈D\{p} it holds that d(p,o') < d(p,o).

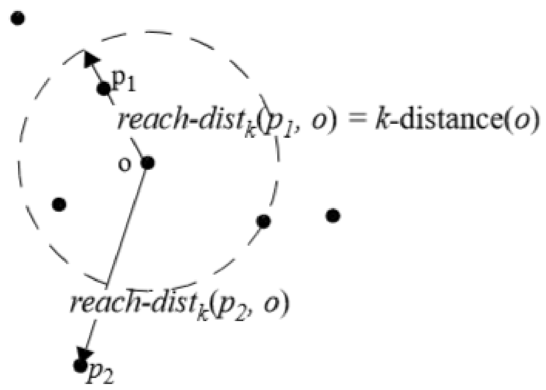


Figure 2: ***reach-dist***(p_1,o) **and** ***reach-dist***(p_2,o)**, for** k**=4**

PROPERTIES OF LOCAL OUTLIERS

Theorem 1: Let p be an object from the database D, and $1 \leq MinPts \leq |D|$.
Then, it is the case that

$$\frac{direct_{min}(p)}{indirect_{max}(p)} \leq LOF(p) \leq \frac{direct_{max}(p)}{indirect_{min}(p)}$$

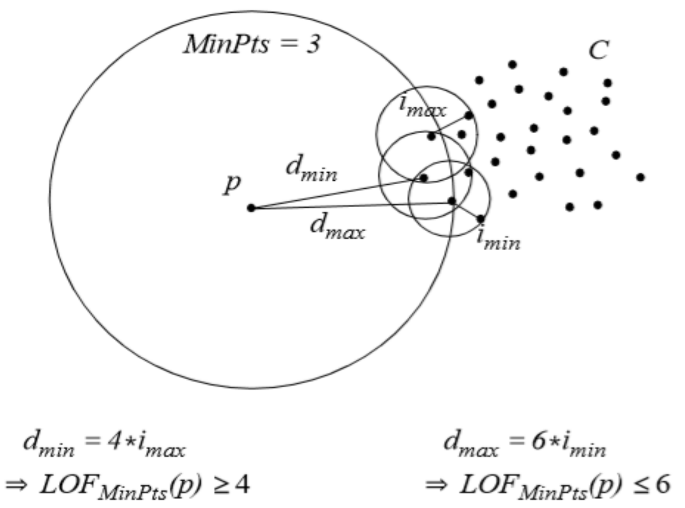


Figure 3: Illustration of theorem 1

Theorem 2: Let p be an object from the database D, $1 \leq MinPts \leq |D|$, and C1, C2,..., Cn be a partition of NminPts(p), i.e. NminPts(p)=C1 ∪ C2 ∪ ... ∪ Cn ∪ {p} with Ci ∩ Cj = ∅ Ci ≠∅ for 1 ≤ I,j ≤ n, I ≠ j.

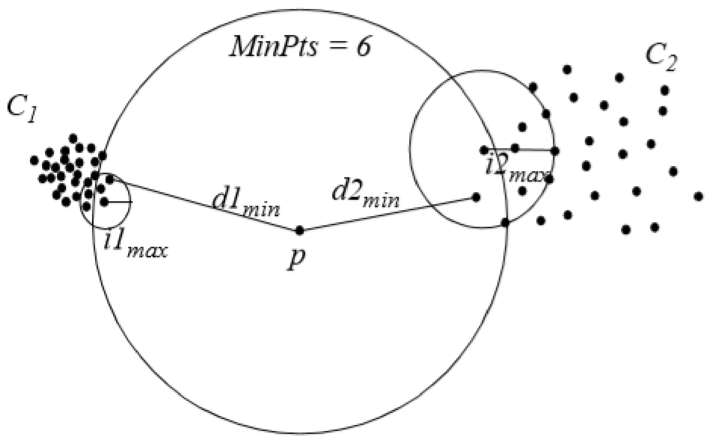


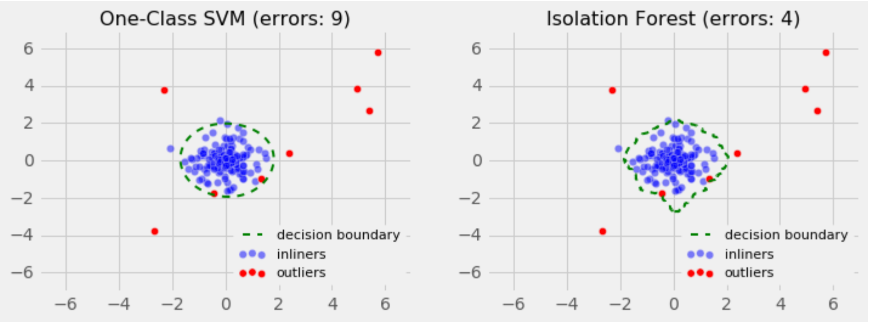
Figure 6: Illustration of theorem 2

2. Isolation Forest Algorithm

The Isolation Forest algorithm isolates observations by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature. The logic argument goes: isolating anomaly observations is easier because only a few conditions are needed to separate those cases from the normal observations. On the other hand, isolating normal observations require more conditions. Therefore, an anomaly score can be calculated as the number of conditions required to separate a given observation.

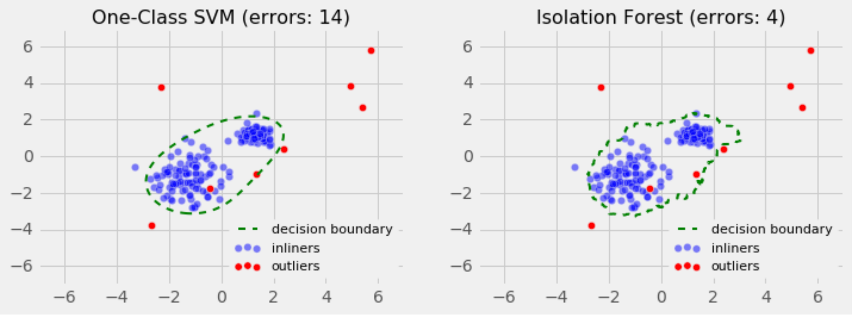
Comparison of Isolation Forest and One-Class Support Vector Machines

The Isolation Forest algorithm shows strong promise, and I tried to estimate its performance against the well-known One-Class Support Vector Machine outlier detection algorithm. First, I compare the two algorithms when the normal observations behave normally and belong to a single group.



As observed, the Isolation Forest algorithm detections have fewer errors as it did not construct a parametric representation of the search space.

Lastly, I evaluated the performance using a slightly more complex case when the observation is grouped in two uneven clusters.



The Takeaway

The method of using Isolation Forests for anomaly detection in the online fraud prevention field is still relatively new. It’s no secret that detecting fraud, phishing and malware has become more challenging as cyber criminals become more sophisticated. We should be using the most advanced tools and methods to prevent current and future fraud.

Advanced outlier detection methods such as Isolation Forests are imperative for companies looking to reduce fraud because this method detects anomalies purely based on the concept of isolation without employing any distance or density measure — fundamentally different from all existing methods.

As a result, Isolation Forests are able to exploit subsampling to achieve a low linear time-complexity and a small memory-requirement, and to deal with the effects of swamping and masking effectively. This gives us better tools to improve our detection rates and react faster to new fraud attacks.

REFERENCES AND BIBLIOGRAPHY

1. http://www.dbs.ifi.lmu.de/Publikationen/Papers/LOF.pdf
2. https://blog.easysol.net/using-isolation-forests-anamoly-detection/
3. https://www.kaggle.com/mlg-ulb/creditcardfraud