≡

3PILLAR
GLOBAL

Home (https://www.3pillarglobal.com/) / Insights (https://www.3pillarglobal.com/insights) / Credit Card Fraud Detection – An Insight Into Machine Learning and Data Science

CREDIT CARD FRAUD DETECTION – AN INSIGHT INTO MACHINE LEARNING AND DATA SCIENCE

✉

(/conta

The importance of Machine Learning and Data Science cannot be overstated. If you are interested in studying past trends and training machines to learn with time how to define scenarios, identify and label events, or predict a value in the present or future, data science is of the essence. It is essential to study the underlying data and model it by selecting an appropriate algorithm to approach any such use case. The various control parameters of the algorithm need to be tweaked to fit the data set. As a result, the developed application improves and becomes more efficient in solving the problem.

In this blog, we have attempted to illustrate the modeling of a data set using a machine learning paradigm classification, with Credit Card Fraud Detection being the base. Classification is a machine learning paradigm that involves deriving a function that will separate data into categories, or classes, characterized by a training set of data containing observations (instances) whose category membership is known. This function is then used in identifying in which of the categories a new observation belongs.

How do you spot 492 fake credit card transactions out of 284K+? Start by reading this! (https://twitter.com/intent/tweet?url=https%3A%2F%2Fwww.3pillarglobal.com%2Finsights%2Fcredit-card-fraud-detection&text=How%20do%20you%20spot%20492%20fake%20credit%20card%20transactions%20out%20of%20284K%2B%3F%20Start%20by%20reading%20this%21&via=3pillarglobal&related=3pillarglobal)

Click To Tweet (https://twitter.com/intent/tweet?url=https%3A%2F%2Fwww.3pillarglobal.com%2Finsights%2Fcredit-card-fraud-detection&text=How%20do%20you%20spot%20492%20fake%20credit%20card%20transactions%20out%20of%20284K%2B%3F

Problem Statement:

The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be fraud. This model is then used to identify whether a new transaction is fraudulent or not. Our aim here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications.

Data Set Analysis:

This problem has been picked from Kaggle (https://www.kaggle.com/dalpozz/creditcardfraud).

Observations

1. The data set is highly skewed, consisting of 492 frauds in a total of 284,807 observations. This resulted in only 0.172% fraud cases. This skewed set is justified by the low number of fraudulent transactions.
2. The dataset consists of numerical values from the 28 'Principal Component Analysis (PCA)' transformed features, namely V1 to V28. Furthermore, there is no metadata about the original features provided, so pre-analysis or feature study could not be done.
3. The 'Time' and 'Amount' features are not transformed data.
4. There is no missing value in the dataset.

Inferences drawn:

1. Owing to such imbalance in data, an algorithm that does not do any feature analysis and predicts all the transactions as non-frauds will also achieve an accuracy of 99.828%. Therefore, accuracy is not a correct measure of efficiency in our case. We need some other standard of correctness while classifying transactions as fraud or non-fraud.
2. The 'Time' feature does not indicate the actual time of the transaction and is more of a list of the data in chronological order. So we assume that the 'Time' feature has little or no significance in classifying a fraud transaction. Therefore, we eliminate this column from further analysis.

Theory:

Credit Card Fraud Detection is a typical example of classification. In this process, we have focused more on analyzing the feature modeling and possible business use cases of the algorithm's output than on the algorithm itself. We used the implementation of Binomial Logistic Regression Algorithm in the 'ROCR' package on the PCA transformed Credit Card Fraud data.

Some Definitions:

The following are essential definitions – in the current problem's context – needed to understand the approaches mentioned later:

- True Positive: The fraud cases that the model predicted as 'fraud.'
- False Positive: The non-fraud cases that the model predicted as 'fraud.'
- True Negative: The non-fraud cases that the model predicted as 'non-fraud.'
- False Negative: The fraud cases that the model predicted as 'non-fraud.'
- Threshold Cutoff Probability: Probability at which the true positive ratio and true negatives ratio are both highest. It can be noted that this probability is minimal, which is reasonable as the probability of frauds is low.
- Accuracy: The measure of correct predictions made by the model – that is, the ratio of fraud transactions classified as fraud and non-fraud classified as non-fraud to the total transactions in the test data.
- Sensitivity: Sensitivity, or True Positive Rate, or Recall, is the ratio of correctly identified fraud cases to total fraud cases.
- Specificity: Specificity, or True Negative Rate, is the ratio of correctly identified non-fraud cases to total non-fraud cases.
- Precision: Precision is the ratio of correctly predicted fraud cases to total predicted fraud cases.

Incorrect Measures of Efficiency of a Data Model:

Let's look at the various measures of efficiency that fail at analyzing the correctness of the underlying data model.

1. Total/Net Accuracy: One approach to gauge the compute model's correctness is to use Accuracy as the deciding parameter. But, as stated earlier, in a highly skewed data set like this, we know that even if we predict all values as non-fraudulent, we'll have only 492 wrong predictions out of 284,807 in total. So, the accuracy is excellent, but it still doesn't solve our problem as we want to identify as many fraud cases as possible. So, we can't use accuracy as a deciding factor here.
2. Confusion Matrix: Merely tabulating the confusion matrix will not provide a clear understanding of the performance of the data. This is because the total number of fraud cases is much less, and variation in the confusion matrix will be so small that it will be equivalent to a justified error in a balanced dataset (probably even less!). So, this measure is also ruled out.

This first part of three-part blog series provides an insight into the analysis of the data and the pitfalls in handling a skewed data set. In the subsequent posts, we shall try to fit a model to this data set, analyze the results and look into the various measures of efficiency that can be resorted to as the metric for defining the utility (correctness) of the modeling. So, stay tuned for our next blog for continued work on Credit Card Fraud Detection.

(https://www.3pillarglobal.com/author/akanksha-chaturvedi)

AKANKSHA CHATURVEDI
SOFTWARE ENGINEER

Akanksha Chaturvedi is a Software Engineer at the 3Pillar Noida office in India. She has experience in Java technology, AngularJS, and iOS application development in Swift. She is very passionate about the design and analysis of computer algorithms, and has a keen interest in how those algorithms solve real world problems. She earned a Bachelor of Engineering in Computer Engineering from Netaji Subhas Institute of Technology (NSIT), Delhi. In her leisure time, she loves to travel, play badminton, watch tennis, and read novels.

[(https://www.3pillarglobal.com/author/skhandelwal)](https://www.3pillarglobal.com/author/skhandelwal)
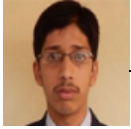
SPARSH KHANDELWAL

SOFTWARE ENGINEER

Sparsh Khandelwal is a Software Engineer at 3Pillar Global. He has experience in .NET technology and iOS application development in Objective C. Sparsh is a graduate from the renowned Indian Institute of Technology (IIT), Delhi. He is an avid reader and loves to spend his time reading novels.

21 RESPONSES TO "CREDIT CARD FRAUD DETECTION – AN INSIGHT INTO MACHINE LEARNING AND DATA SCIENCE"

Sanjana Dude on February 1st, 2018 - 10:44pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123361)

Where can I find the blog articles in continuation to this topic?

Reply

> *Srini* on April 4th, 2018 - 5:10am (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123493)
>
> Can you please help me to access the rest of this blog for credit card fraud detection?
> Thanks much in advance
>
> Regards
> Srini
>
> Reply

> *Siddharth Agarwal* on May 4th, 2018 - 6:24pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123551)
>
> Work in fraud detection
>
> Reply

> *jawad* on May 23rd, 2018 - 3:40pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123581)
>
> Where can I find the blog articles in continuation to this topic?
>
> Reply

> *sgds* on May 24th, 2018 - 4:51pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123586)
>
> good
>
> Reply

> *Vinay Kandala* on May 25th, 2018 - 12:05am (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123587)
>
> Thanks
>
> Reply

Ankush Dhiman on July 2nd, 2018 - 2:00pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123655)

Where can i find next part?

Reply

Mushtaq on July 26th, 2018 - 7:27am (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123736)

where can i find the remaining part of CREDIT CARD FRAUD DETECTION

Reply

Srini on April 3rd, 2018 - 12:33pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123492)

Excellent explanation. Need help to go through the rest of the blog please.

Reply

Iswarya on May 21st, 2018 - 11:22am (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123574)

Need articles on credit cards fraud detectiom

Reply

raj on August 22nd, 2018 - 4:57am (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123829)

Interesting read

Reply

Phil on August 26th, 2018 - 4:54pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123841)

Nice first entry

Reply

ram on September 3rd, 2018 - 5:00am (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123856)

Where can I find the blog articles in continuation to this topic?

Reply

VINEET KAPOOR on October 31st, 2018 - 1:52pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-123956)

Nice Article and would like to know about other parts, I am a data engineer by profession

Reply

Sai Diksha Narahari on December 8th, 2018 - 1:15pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-124061)

Very nice explanation. Can I get access to continuation of the article "credit-card-fraud-detection"..

Reply

Alka Singh on December 18th, 2018 - 11:09pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-124084)

It's very comprehensive and insightful.

Reply

kushagra shroff on January 17th, 2019 - 7:00am (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-124126)

🙂

Reply

Vinitha on January 28th, 2019 - 8:36pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-124157)

I would like to understand the data pre processing steps in the data and the feature selection

Reply

Vinitha on January 28th, 2019 - 8:36pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-124158)

Would like to.get a deep understanding in the data preprocessing and feature selection of the data

Reply

suresh on January 31st, 2019 - 12:30pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-124168)

please guide me to other parts

Reply

Vinitha Shree on February 3rd, 2019 - 1:39pm (https://www.3pillarglobal.com/insights/credit-card-fraud-detection#comment-124172)

Would like to understand the data analysis that happens in creditcard data

Reply

LEAVE A REPLY

Name (required)

Your Name*

Mail (required)

Your E-Mail*

Comment

Your Comment here...

I'm not a robot

reCAPTCHA
Privacy - Terms

SUBMIT

Learn More
Articles (https://www.3pillarglobal.com/news/articles)
Awards (https://www.3pillarglobal.com/news/awards)
Events (https://www.3pillarglobal.com/news/events)
Press Releases (https://www.3pillarglobal.com/news/press-release)
Blogs (https://www.3pillarglobal.com/insights/blog-posts)
Podcasts (https://www.3pillarglobal.com/insights/podcast)
Videos (https://www.3pillarglobal.com/insights/videos)

LET'S WORK TOGETHER (/CONTACT)

CONTACT
(HTTPS://WWW.3PILLARGLOBAL.COM/CONTACT)

PRIVACY POLICY
(HTTPS://WWW.3PILLARGLOBAL.COM/PRIVACY-POLICY)

TERMS AND CONDITIONS
(HTTPS://WWW.3PILLARGLOBAL.COM/TERMS-AND-CONDITIONS)

✉ LET'S TALK (/CONTACT?UTM_SOURCE=3PILLAR_WEBSITE_CONTACT_BUTTON&UTM_MEDIUM=3PILLAR_WEBSITE&UTM_CAMPAIGN=3PILLAR_CONTACT_BUTTON)

in
(HTTPS://WWW.LINKEDIN.COM/COMPANY/3PILLAR-GLOBAL)

(HTTPS://TWITTER.COM/3PILLARGLOBAL)

(HTTPS://INSTAGRAM.COM/USER34...)

(HTTPS://WWW.YOUTUBE.COM/C/3PILLARGLOBAL)

f
(HTTPS://FACEBOOK.COM/3PILLARGLOBAL)

(HTTPS://INSTAGRAM.COM/3PI...)

/