Indellient
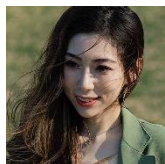

May 20, 2020

Meina Zhou, Lead Data Scientist

How to Use Machine Learning to Detect Fraud

Machine Learning has gained increasing popularity over the past several years. More and more companies have started to use Data Science in order to improve their business strategies and tactics. Sample applications of machine learning models include prediction, image/speech recognition, recommendation engine, and fraud detection.

As the Lead Data Scientist at Indellient, I have helped many clients mitigate risk by taking control of their data. Today we are going to focus on the applications of Machine Learning in Fraud Detection.

What is Machine Learning?

By definition, machine learning is the process of teaching the computer to learn data patterns and improve the prediction automatically from experience (when more has

been fed into the system).

There are three key components for building machine learning algorithms:
A machine learning (ML) algorithm is a process or set of procedures that help a model adapt to the data given an optimizing objective.

1. **Clear business objectives & goals**: what do you want to optimize using the machine learning model? What are the key business metrics for your model?
2. **Good quality sample data**: the sample data must contain accurate information about the historical events (such as sales, usage, client information).
3. **Choice of algorithm**: there are wide ranges of machine learning models, including supervised learning, unsupervised learning, and semi-supervised learning. It is vital to choose the algorithm that best fits the business goal and sample data.

How to Detect Fraud using Machine Learning Algorithms?

Now that we have introduced the basics of Machine Learning, let's talk about how we can use machine learning to detect fraud.
To build a fraud detection algorithm, we first need to identify and understand the fraud scenarios for the business:

- What kind of fraud are we focusing on: consumer fraud, financial statement fraud, expense fraud, or other?

- How large is the impact of fraud? What is the expected improvement?

Then, we will start to collect sample data based on the scenarios that we have identified. One of the biggest challenges for fraud detection is the imbalance dataset. Imbalance dataset means that only a small percentage of consumers have fraudulent intentions. Hence, it is important to gather as many fraud cases as possible. The accuracy of the machine learning algorithms improves as more data has been fed into the system.
Lastly, we will feed the collected sample data into our choice of machine learning algorithms and finetune the models to make the best predictions. We will compare the model performance based on the key business metrics and select the best performing algorithm.

Let the Experts Handle It

Indellient is a Software Development Company that specializes in Data Analytics, Cloud Development Application, DevOps Services, and Business Process Management.

Find out more: https://www.indellient.com/blog/how-to-use-machine-learning-to-detect-fraud/