H2O for Real Time Fraud Detection

Organizations responsible for fraud prevention understand the power of analytics: very small differences in the ability to predict fraud can have a major impact on losses. There are a host of challenges that need to be addressed at the transaction, account and network-level to detect fraudulent behavior and suspicious activities. It is estimated that fraud costs at least $80 billion a year across all lines of insurance[1]. That is why companies at risk of fraud invest in machine learning as a preemptive approach to tackling fraud.

Companies across industries rely on H2O for scalable machine learning to detect fraud. The table below shows some examples of H2O users and the types of fraud they seek to prevent.

USER	APPLICATION
Payment Systems Company	Detect merchant fraud
Global Insurance Company	Predict fraudulent Workman's Compensation claims
Multinational Telecommunications Provider	Detect fraudulent SIM card activity
Leading U.S. Credit Card Issuer	Detect fraudulent card transactions

These companies turn to H2O because it is highly scalable and delivers superior performance; offers flexible deployment options; works seamlessly in a large scale data sets; and offers a simple interface.

Real Time Insight with Deep Learning

Fraudulent transactions are rare, but costly if they aren't detected. In the credit card business, for example, third-party fraud accounts for roughly 4 out of every 10,000 transactions. Modeling rare events is difficult, like finding a needle in a haystack. For best results, gather as much data as possible, and use the most advanced techniques available.

A U.S.-based payment systems company that handles billions of dollars in payments each month uses H2O Deep Learning for real time fraud detection. Working with a dataset of **160 million** records and 1,500 features, the company's data scientists use a test-and-learn approach to find the best-performing predictive model. H2O's distributed in-memory architecture enables them to run tests quickly and build the most accurate predictive models. The company estimates that a 1% reduction in fraud results in **$1 million savings per month**.

> **The company estimates that a 1% reduction in fraud results in $1 million savings per month.**
> – A U.S.-based payment systems customer

Deep Learning is a rapidly growing discipline that models high-level patterns in data as complex multi-layered networks. Because it is the most general way to model a problem, Deep Learning can solve the most challenging prediction problems.

Deep Learning applications use artificial neural networks (ANN) with multiple hidden layers, also called deep neural networks (DNN). Conventional neural networks date back to the 1950s, and are used in commercial fraud applications like FICO Falcon. However, they are very difficult to train and often do not outperform other machine learning techniques. Deep Learning, on the other hand, provides an optimization framework to outperform other methods.

DL Maxout 50x50x50, 1000 epochs
holdout AUC=0.9348
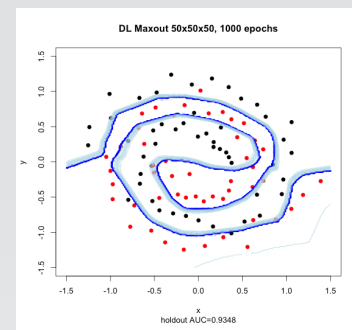
Leveraging Large Scale Data Sets

The insurance industry estimates that claims fraud amounts to $80 billion annually – in the United States alone. Big problems require big solutions; a global insurance company seeks to detect and prevent claims fraud in its Workman's Compensation business.

The existing process is entirely manual: professional claims examiners use judgment and experience to select suspicious claims for analysis. There is no substitute for a highly trained and experienced examiner, but a growing business needs a more automated approach. Moreover, examiners must pull information from a multitude of systems, a time-consuming process that contributes to the case backlog.

As a first step, the company has consolidated data from many different sources into a Hadoop data store. The data consists of a mix of structured and unstructured data, including handwritten notes from medical professionals.

Hadoop is an excellent low-cost way to store huge datasets, but it lacks a capability for sophisticated predictive analytics. Many of the company's data scientists use R for advanced analytics; but R by itself cannot scale to Hadoop-level data volumes, and extracting the data to an analytic server is time consuming. H2O solves this problem; it is co-located in the company's Hadoop cluster, so analysts can discover insights in the data without extracting it or taking samples. Data scientists interact with H2O with R; however, all of the work is performed in H2O where it is deployed, in the Hadoop cluster.

Rapid Model Deployment

Fraud perpetrators constantly change their methods, and fraud detection models tend to have a short usable life; fraud analytics must be agile. However, the models are highly complex, so programming them from scratch for deployment in a production system takes time and money. This is one of the key shortcomings of commercial fraud detection software.

When an analytics project is completed, H2O exports predictive models as Plain Old Java Objects (POJOs). POJOs can run anywhere in the organization that Java runs: in transaction processing systems, case management systems, network authorization systems, or wherever predictions are needed.

About H2O.ai

H2O is the #1 open source machine learning platform for smarter applications. H2O.ai is the Silicon Valley software company supporting and developing H2O. Leading insurance, healthcare and financial services companies are using H2O to make smarter predictions about churn, pricing, fraud and more. H2O.ai is fostering a grassroots movement of systems engineers, data scientists, data developers and predictive analysts to move machine learning forward. A rapidly growing community of H2O users is now active in more than 5000 organizations worldwide. H2O.ai is a Gartner Cool Vendor in Data Science for 2015.