Business Applications of Machine Learning

Professor Kartik Hosanagar

Wharton
UNIVERSITY of PENNSYLVANIA

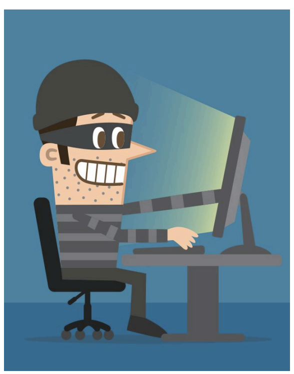
ML in Finance: Fraud Detection

- Intro to credit card fraud detection
- Supervised learning example
- Advantages and limitations of ML in fraud

Credit Card Fraud

“Several years ago, Avi…noticed a large, unusual transaction in a Las Vegas casino on his credit card. He hadn’t been in Las Vegas. He had only been there once a long time before… After an extensive conversation, his card provider reversed the transaction and replaced the card.”

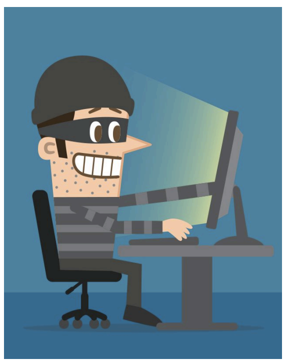
Fraud Occurs → Customer Impacted → Dispute Required → Card Replaced



Content/quotes from “Prediction Machines: The Simple Economics of Artificial Intelligence ” by Ajay Agrawal, Avi Goldfarb, and Joshua Gans

Fraud Detection with Machine Learning

Fraud Occurs



ML Detects

Customer Impact Avoided

Card Replaced

Early fraud detection with ML can help prevent fraud and save banks a lot of money

Wharton
UNIVERSITY of PENNSYLVANIA

Content/quotes from “A Human’s Guide to Machine Intelligence” by Kartik Hosanagar

Machine Learning Opportunity

- Fraud detection is really fraud prediction
  - Predicting whether a particular customer's purchase is fraudulent or not
  - The more accurately a business predicts fraud, the less false positives and false negatives they will have, thus lowering losses

- Both supervised and unsupervised learning are used in fraud

Supervised Learning

- Training occurs by using a large dataset with the details of individual transactions provided and with each transaction tagged as fraud or not. From this, the model learns the unique patterns of fraud.

Unsupervised Learning

- Anomaly detection can compare new transactions with prior ones to detect outliers. This can help identify fraud that doesn't necessarily fit a previously identified pattern (e.g. a new type of fraud).

Content/quotes from: https://stripe.com/radar/guide & https://www.fico.com/blogs/5-keys-using-ai-and-machine-learning-fraud-detection

Supervised Learning Example

- Requires input data for training (usually historical data of two types):
  1. “Properties that can be ‘read off’ a single credit card payment”
     - E.g.: country the card was issued in, IP address of payment, user’s email domain
  2. Behavioral data (provides “some of the most predictive signals”)
     - E.g.: Number of countries the card was used recently

Wharton
UNIVERSITY *of* PENNSYLVANIA

Content/quotes from: https://stripe.com/radar/guide

Supervised Learning Example (cont.)

- Requires input data for training (usually historical data of two types):
  1. "Properties that can be 'read off' a single credit card payment"
  2. Behavioral data (provides "some of the most predictive signals")

Sample Training Data*

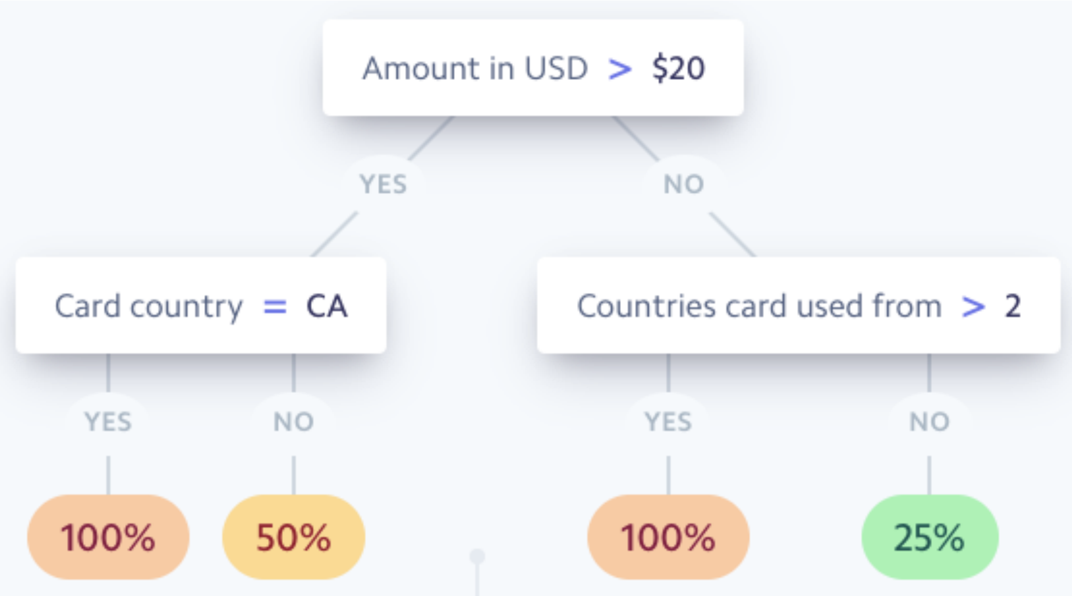
Amount in USD	Card country	Countries card used from (24h)	Fraud?
$10.00	US	1	No
$10.00	CA	2	No
$10.00	CA	1	No
$10.00	US	1	Yes
$30.00	US	1	Yes
$99.00	CA	1	Yes
$15.00	CA	3	Yes
$70.00	US	1	No

* This training data is very limited in order to provide a simple example. To build an accurate model you would need millions of rows as well as additional columns.

Content/quotes from: https://stripe.com/radar/guide

Supervised Learning Example (cont.)

- Produces an output model, such as the following decision tree
  - The tree answers: “of transactions in our data set with properties similar to the transaction we’re examining now, what fraction were actually fraudulent?”
  - “The machine learning part is concerned with the construction of the tree- what questions do we ask, in what order, to maximize the chances that we can distinguish between the two classes accurately?”



Sample Output *

*This decision tree is based on the same limited data from the previous slide.

Content/quotes from: https://stripe.com/radar/guide

Machine Learning Fraud Detection

- Supervised Learning: From this (very limited) data, the model would learn a unique pattern of fraud:
  - If >$20 & from Canada, **100% chance of fraud**
  - If <$20 & from >2 countries, **100% chance of fraud**
  - If >$20 & not from CA, or <$20 & from <2 countries, **not sure**
- Unsupervised Learning: Detecting transactions that appear like anomalies:
  - A transaction is for an exceptionally high amount + in a country where this person has not transacted before + In the past, foreign transactions were preceded by flight purchase to that country unlike this time = Anomaly

Content/quotes from: https://stripe.com/radar/guide

Advantages of ML for Fraud

Speed

Algorithms can quickly process a large volume of transactions.

This is important for fraud since a decision is needed in real time.

Scale

A challenge for humans, but algorithms improve as the amount of data increases.

Efficiency

Machine learning algorithms are better than humans at repetitive tasks.

Wharton
UNIVERSITY of PENNSYLVANIA

Content/quotes from: https://marutitech.com/machine-learning-fraud-detection/

Limitations of ML for Fraud

Transparency

Algorithms can’t always explain why someone was blocked.

- Hard to catch issues with the model if it isn’t well understood
- Also an ethical problem if biases go undetected (to be discussed in module 5)

Data Volume

Smaller companies may not have enough training data.

- Algorithm accuracy might be lower as a result

Wharton
University *of* Pennsylvania

Content/quotes from: https://marutitech.com/machine-learning-fraud-detection/

ML in Finance: Additional Applications

- Identity Verification & Authentication
- Loan & Insurance Underwriting
- Predicting Customer Churn
- Financial Forecasting
- Customer Experience
- Personal Finance

Identity Verification & Authentication

- ML can improve security through more than just detecting fraud patterns. It also provides new methods of improving identity verification.

Traditional Verification

- Passwords
- PIN numbers

➡

ML-Based Verification

- Biometric authentication using facial and voice recognition technologies

- One biometric use case would be when new accounts are opened and customers need to provide multiple forms of ID.
  - Customers could instead provide “selfies” or voice prints. Facial recognition and voice recognition technologies can be used to verify identity based on the images/audio provided.
  - ATMs in China are starting to use face recognition

Content/quotes from “Section 2: Known Applications of AI”

Loan & Insurance Underwriting

- ML can detect patterns between consumer data and loan or insurance outcomes, and use this to predict the outcomes of particular applicants.
  - E.g., supervised learning can be used by providing a training dataset with historical data on consumers & their lending/insurance results.
    - Consumer data: age, income, employment, etc.
    - Lending/insurance results: repaying loans on time vs. defaulting

Key Benefits & Limitations of Loan/Insurance Models

Benefits	Limitations
• Potential for "increasing loan volume & reducing risk…[by] using more diverse data as well as data with weaker signals." • Could reduce processing time	• Algorithm could be biased & could perpetuate historical discrimination. Companies need to make sure their algorithms don't discriminate (discussed further in module 4).

Content/quotes from: "Section 2: Known Applications of AI"
Additional Content from:: https://emerj.com/ai-sector-overviews/machine-learning-in-finance/

Predicting Customer Churn

- Banks want to retain customers/prevent churn & can apply ML to this goal.
- In much the same way as with fraud models, the customer data that banks have can be used to "create churn models based on customer attributes or features of those who did or did not churn for another competitor."

Key Benefits & Limitations of Churn Models

Benefits	Limitations
• Predictions from churn models are actionable b/c knowing in advance which customers might churn allows banks to make extra efforts to improve those customers' satisfaction.	• Predictions about who might churn don't necessarily provide insight into what is causing them to leave and how best to retain them.

Content/quotes from "Section 2: Known Applications of AI"

Three Additional Examples of ML in Finance

- **Customer Experience:**
  - Conversational AI platforms are being used to service customers via chat or over the phone to improve responsiveness and reduce costs.
- **Personal Finance:**
  - Budgeting apps that incorporate ML can be used for tracking spending. Algorithms can "analyze this data to identify spending patterns that customers might not be aware of & identify areas where they can save."
- **Financial Forecasting:**
  - For firms that use time-series modeling or other approaches to predict factors driving business growth, ML can be used to "add more information into their models, prune existing or legacy sources of information and increase the number of factors they are trying to predict."

Content/quotes for financial forecasting from "Section 2: Known Applications of AI"
Content/quotes for customer experience and personal finance section from https://www.alacriti.com/machine-learning-in-financial-services-potential-applications/

Wharton
UNIVERSITY of PENNSYLVANIA