Catch them in the Act

Fraud Detection in Real-time

Seshika Fernando
Technical Lead, WSO2
seshika@wso2.com
$4 Trillion in Global Fraud Losses
That’s 5% of Global GDP
Efficient Fraud Detection

Domain Expertise

Streaming Analytics

Batch Analytics

Predictive Analytics
Complex Event Processing

Notify if there is a 10% increase in overall trading activity AND the average price of commodities has fallen 2% in the last 4 hours
Many ways

• Generic Rules
• Fraud Scoring
• Markov Models
• Machine Learning
Domain Expertise → Generic Rules
Typical Credit Card Fraudster

- Use stolen cards
- Buy Expensive stuff
- In Large Quantities
- Very quickly
- At odd hours
- Ship to many places
- Get rejected often
Moving Averages

from TransactionStream#window.time(60 min)
select itemNo, avg(qty) as avg, stdev(qty) as stdev
group by itemNo
update AvgTbl as a
on itemNo == a.itemNo;

from TransactionStream
[itemNo== a.itemNo and qty > (a.avg + 2*a.stdev) in AvgTbl as a]
select *
insert into FraudStream;
from e1 = TransactionStream ->

  e2 = TransactionStream[e1.cardNo == e2.cardNo] <2:>

within 5 min

select e1.cardNo, e1.txnID, e2[0].txnID, e2[1].txnID

insert into FraudStream
The False Positive Trap

- So what if I buy Expensive stuff
- And why can’t I buy a lot
- Very Quickly
- At odd hours
- Ship to many places

Rich guy
Gift giver
Impulse Shopper
Night owl
Many girlfriends?

Blocking genuine customers could be counter productive and costly
How to avoid False Positives

- Use combinations of rules
- Give weights to each rule
- Single number that reflects many fraud indicators
- Use a threshold to reject transactions

• You just bought a Diamond Ring?
• You bought 20 Diamond Rings, in 15 minutes at 3am from an IP address in Nigeria?
How to score

Score =

0.001 * itemPrice
+ 0.1 * itemQuantity
+ 2.5 * isFreeEmail
+ 5 * riskyCountry
+ 8 * suspiciousIPRange
+ 5 * suspiciousUsername
+ 3 * highTransactionVelocity
Are we safe?
Markov Models

- Model randomly changing systems
- Detect rare activity sequences using
  - Classification
  - Probability Calculation
  - Metric Calculation
Markov Models: Classification

Each transaction is classified under the following three qualities and expressed as a 3 letter token, e.g., HNN

- Amount spent: Low, Normal and High
- Whether the transaction includes high price ticket item: Normal and High
- Time elapsed since the last transaction: Large, Normal and Small
Markov Models: Probability Matrix

- Create a State Transition Probability Matrix

<table>
<thead>
<tr>
<th></th>
<th>LNL</th>
<th>LNH</th>
<th>LNS</th>
<th>LHL</th>
<th>HHL</th>
<th>HHS</th>
<th>HNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LNL</td>
<td>0.976788</td>
<td>0.542152</td>
<td>0.20706</td>
<td>0.095459</td>
<td>0.007166</td>
<td>0.569172</td>
<td>0.335481</td>
</tr>
<tr>
<td>LNH</td>
<td>0.806876</td>
<td>0.609425</td>
<td>0.188628</td>
<td>0.651126</td>
<td>0.113801</td>
<td>0.630711</td>
<td>0.099825</td>
</tr>
<tr>
<td>LNS</td>
<td>0.07419</td>
<td>0.83973</td>
<td>0.951471</td>
<td>0.156532</td>
<td>0.12045</td>
<td>0.201713</td>
<td>0.970792</td>
</tr>
<tr>
<td>LHL</td>
<td>0.452885</td>
<td>0.634071</td>
<td>0.328956</td>
<td>0.786087</td>
<td>0.676753</td>
<td>0.063064</td>
<td>0.225353</td>
</tr>
<tr>
<td>HHL</td>
<td>0.386206</td>
<td>0.255719</td>
<td>0.451524</td>
<td>0.469597</td>
<td>0.810013</td>
<td>0.444638</td>
<td>0.612242</td>
</tr>
<tr>
<td>HHS</td>
<td>0.204606</td>
<td>0.832722</td>
<td>0.043194</td>
<td>0.459342</td>
<td>0.960486</td>
<td>0.796382</td>
<td>0.34544</td>
</tr>
<tr>
<td>HNS</td>
<td>0.757737</td>
<td>0.371359</td>
<td>0.326846</td>
<td>0.970243</td>
<td>0.771326</td>
<td>0.015835</td>
<td>0.574333</td>
</tr>
</tbody>
</table>
Markov Models: Probability Comparison

- Compare the probabilities of incoming transaction sequences with thresholds and flag fraud as appropriate
- Can use direct probabilities or more complex metrics
  - Miss Rate Metric
  - Miss Probability Metric
  - Entropy Reduction Metric
- Update Markov Probability table with incoming transactions
Markov Models for Fraud Detection
Learn from Data

• Apply **Predictive** Analysis on **Batch** Data and provide Classifiers to **Streaming** Analytics
Dig Deeper using Big Data

• Provide access to historical data to dig deeper

• Make querying and filtering easy and intuitive

• Provide useful visualizations to isolate incidents and unearth connections
Visualize
Visualize