Using Data Analysis for Detecting Credit Card Fraud

(By Min Win 01/ 05/2021)

Companies today are employing analytical techniques for the early detection of credit card frauds, a key factor in mitigating fraud damage. The most common type of credit card fraud does not involve the physical stealing of the card, but that of credit card credentials, which are then used for online purchases.

Imagine that you have been hired as a Data Analyst to work in the Credit Card Division of a bank. And your first assignment is to join your team in using data analysis for the early detection and mitigation of credit card fraud.

In order to prescribe a way forward, that is, suggest what should be done in order for fraud to get detected early on, you need to understand what a fraudulent transaction looks like. And for that you need to start by looking at historical data.

Here is a sample data set that captures the credit card transaction details for a few users.

IP Address	User ID	Account Number	Age	Shipping Address	Transaction Date	Transaction Time	Transaction Value	Product Category	Units Purchased
3.56.123.0	johnp	25671147	32	1542, Orchid Lane, WA 98706, US	15-5-20	15:00:05	$121.58	Clothing	1
3.56.123.0	johnp	25671147	32	1542, Orchid Lane, WA 98706, US	10-6-20	10:23:10	$79.23	Electronics	2
3.56.123.0	johnp	25671147	32	1542, Orchid Lane, WA 98706, US	1-6-20	07:12:45		Home Décor	1
1.186.52.7	johnp	25671147	32	In-store	3-6-20	01:11:10	$2,009.99	Electronics	10
	johnp	25671147	32	In-store	2020-06-03	01:15:12	$4,131.00	Electronics	15
1.186.52.7	johnp	25671147	32	P.O. Box 1049	03-06-2020	01:22:24	$3,010.50	Tools	20
1.58.167.2	davidg	51422789	47	90 Robinson Blvd, Alberta, 97602, Canada	15 May 2020	17:02:08	$234.20	Furniture	1
1.58.167.2	davidg	51422789	47	90 Robinson Blvd, Alberta, 97602, Canada	18 May 2020	19:12:45	$141.00	Kithcen Supplies	3
	davidg	51422789	47	90 Robinson Blvd, Alberta, 97602, Canada	01 June 2020	17:34:15	$157.25	Car Spares	2
1.58.167.2	davidg	51422789	47	90 Robinson Blvd, Alberta, 97602, Canada	13 June 2020	18:02:10	$59.99	Kithcen Supplies	1
172.165.10.1	ellend	11568528		P.O. Box 1322	07 June 2020	15:53:12	$99.99	Clothing	1
172.165.10.1	ellend	11568528		P.O. Box 1322	08 June 2020	17:15:30	$53.15	Beauty	1
1.167.255.10	ellend	11568528		P.O. Box 5401	02 July 2020	00:05:10	$4,895.00	Laptop	1

Problem-solving approach

- Check list: IP address, account num, date, product category, shipping address, transaction value, transaction time
- Irrelevant data: age

Findings

- Errors /Missing data:

  Row 2 & 3: Transaction date in reverse order

  Row 3: Missing transaction value (johnp)

  Row 5: Missing IP address, transaction (johnp)

  Row 9: Missing ip (David)

  Row 11-13: Missing age (ellend)

Wrong formats

Row 5 & 6: Date format 2020-06-03

Anomalies

johnp

- Change in Frequencies, Order Significantly higher, Bulk orders - Row 4, 5, 6
- Change in IP: 3.56.123.0 (US) - 1.186.52.7(India)
- Address Change: johnp (Row 6)

ellenD

- Row 12: **172.165.10.1** Mismatched IP address **1.167.255.10** (From US to Taiwan)
- Mismatched PO Box, Transaction Time (midnight)
- Trasaction value too high but only one item.

davidg data seems to be normal. (Except, IP address (1.58.167.2) from China. Shipping address to Canada)

Q & A

1. List at least data points that are required for the analysis and detection of a credit card fraud.
   - IP address
   - Shipping Address
   - Transaction Date
   - Transaction Value
   - Units Purchased for same customer (Name and Account Number)
2. Identify errors/issues that could impact the accuracy of your findings, based on a data table provided.
   - Missing IP Addresses (Ip Address Column) in row 5 and 9
   - Transaction Date column has different date format especially in row 5 and 6
   - Transaction Value missing in row 3
3. Identify anomalies, or unexpected behaviors, that would lead you to believe the transaction may be suspect, based on a data table provided.

   For customer johnp,

   - Change in Frequencies went up in row 4, 5 and 6.
   - Orders went significantly higher with bulk amounts.
   - Shipping address changed in row 5 and 6

   For ellend,

   - In row 12, IP Address mismatched,
   - Shipping Adress (PO Box) mismatched
   - Order significantly higher
   - Row 12: **172.165.10.1** Mismatched IP address **1.167.255.10** (From US to Taiwan)
   - Mismatached PO Box, Transaction Time (midnight) Trasaction value too high but only one item.

Other anomalies found:

- Row 4, 5, 6 (Change in Frequencies, Order Significantly higher, Bulk orders)
- Change in ip: 3.56.123.0 (US) - 1.186.52.7(India)
- Row 6: Address Change (Johnp)

4. Briefly explain your key take-away from the provided data visualization chart.

   For johnp, Transaction Values in Transaction 3, 4 and 5 are significantly higher than normal spending.

   For ellend, Transaction Values in Transaction 3 significantly higher than normal spending.

   For davidg, Transaction Values are normal.

5. Identify the type of analysis that you are performing when you are analyzing historical credit card data to understand what a fraudulent transaction looks like.

   Descriptive techniques of analysis was used to analyze historical credit card data because this technique help you gain an understanding of what happened, include the identification of patterns and anomalies in data.