Richard Albright

12 April 2019

ISyE 6501 Introduction to Analytics Modeling Course Project

This project should be done individually.

The web sites https://www.sas.com/en_us/customers.html, https://www-03.ibm.com/software/businesscasestudies/us/en/corp, and https://www.informs.org/Impact/O.R.-Analytics-Success-Stories (among others) contain brief overviews of some major Analytics success stories. In this course project, your job is to think carefully about what analytics models and data might have been required.

1. (1) Browse the short overviews of the projects. Read a bunch of them – they're really interesting. But don't try to read them all unless you have a lot of spare time; there are lots!
2. (2) Pick a project for which you think at least three different Analytics models might have been combined to create the solution.
3. (3) Think carefully and critically about what models might be used to create the solution, how they would be combined, what specific data might be needed to use the models, how it might be collected, and how often it might need to be refreshed and the models re-run. **DO NOT find a description online (or elsewhere) of what the company or organization actually did.** I want this project to be about your ideas, not about reading what someone else did.
4. (4) Write a short report describing your answers to (3).

Answer:

The Following case study was chosen for Analysis:

https://www.sas.com/en_us/customers/landsbankinn.html

Landsbankinn is a bank in Iceland with approximately 120,000 accounts and has a 39% market share of that country's banking business. It is on the list of Global systemically important banks. The bank chose SAS to help implement the Iceland's Anti Money Laundering regulations.

In order to implement an effective Anti-Money Laundering (AML) policy. The bank must identify customers who are at high risk for money laundering and monitor their transaction activity for any suspicious activity. The current state of AML monitoring prior to using analytics lead to a large amount of transactions false identified positives (transactions identified as money laundering, that were not) and as false negatives (transactions not identified as money laundering that were. This is a complex analytics problem that requires a few different models to correctly identify money laundering activity.

The first step is to identify customers that have been positively identified as money launderers, then determine what other customers the bank has that could also be potentially laundering money.

- Given:
    - Customer account
    - Account type
    - Customer id
    - Type of customer (business, personal)
    - Primary business activity of customer.
    - Account opening documents of customer
    - Customer credit rating
    - Country of domicile of customer account
    - account aggregate transaction statistics with counterparty IDs domiciled in countries on the FATF blacklist (monthly pct. of transaction counts, monthly pct. of value).
    - account aggregate transaction statistics with counterparty IDs who are known money launderers (monthly pct. of transaction counts, monthly pct. of value).
    - account aggregate transaction statistics with counterparty IDs that are NOT in countries on the FATF blacklist (monthly pct. of transaction counts, monthly pct. of value).
    - Customer is a known/not known money launderer (cluster centers)
- Use:
    - K-means clustering (k=2)
- To:
    - Segment similar customers into potential money launderers based upon their similarity to known money launderers.

The next step is to properly identify accounts that are high risk for conducting money laundering activities. Any customers that have banking transaction activity associated with countries on the Financial Action Task Force (FATF) blacklist should be considered high risk. The FATF blacklist is a list of countries that have poor banking standards and/or are consider state sponsors of terrorism (the current list of countries can be found at http://www.fatf-gafi.org/countries/#high-risk). This should be considered the initial model to broadly identify high risk customers. Granted, not all customers identified as money launderers have accounts in a FATF blacklisted country, but it should help speed up identification for those customers and counterparties that are located in those countries. Use the same inputs that were used for the K-means Cluster Model and include output category from the model as well.

- Given:
    - Customer account
    - Account type
    - Customer id
    - Type of customer (business, personal)
    - Primary business activity of customer.
    - Account opening documents of customer
    - Customer credit rating
    - Country of domicile of customer account
    - Potential Money Launderer (category from K-means cluster model)

  - account aggregate transaction statistics with counterparty IDs domiciled in countries on the FATF blacklist (monthly pct. of transaction counts, monthly pct. of value).
  - account aggregate transaction statistics with counterparty IDs who are known money launderers (monthly pct. of transaction counts, monthly pct. of value).
  - account aggregate transaction statistics with counterparty IDs that are NOT in countries on the FATF blacklist (monthly pct. of transaction counts, monthly pct. of value).
- Use:
  - Logistic Regression
- To:
  - Create risk score (the logistic probability value) for customers who are at high risk for money laundering.

Once a risk score is assigned to clients, a risk score of counterparty IDs should also be established.

- Given:
  - Counterparty ID (these can also be existing customers)
  - Counterparty type (business, personal)
  - Primary business activity of counterparty.
  - Country of domicile of counterparty account
  - account aggregate transaction statistics with customers (monthly pct. of transactions counts, monthly pct. of value).
  - account aggregate transaction statistics with customers (monthly pct. of transactions counts, monthly pct. of value).
- Use:
  - Logistic Regression
- To:
  - Create risk score for counterparties who are at high risk for money laundering.

Once the risk score customers and counterparties for money laundering are computed. The accounts need cross referenced amongst each other. We are in essence are creating a something similar to the web page rank algorithm, but for probabilities that a customer has links to a money launderer. Two graph models of markov chain probabilities will be created, one for pct. of count of transactions involving the counterparty, one of pct. of value of involving the counterparty.

- Given:
  - Customer IDs
  - Counterparty IDs
  - Customer/Counterparty Risk Score as a weight on the node.
  - The pct. of transactions (counts and value)
- Use:
  - Network (Graph) Models
- To:
  - Identify networks of potential money laundering associations

Considering the size of the graph model is going to be extremely large, the Louvain Modularity algorithm should be applied to the graph model to find higher level relationships between communities to determine what communities are potential money launderers.

- Given:
  - Network Graph Model
- Use:
  - Louvain Modularity
- To:
  - Identify networks of potential money laundering associations

Now that potential relations amongst known and suspected money launderers are identified. The transactions will need analyzed for their relationships among communities of money launderers.

- Given:
  - Customer Transaction
  - Customer Risk Score
  - Counterparty Risk Score
  - Louvain Modularity
  - Other Graph metrics
- Use:
  - K-means clustering
- To:
  - Flag transaction as potential money laundering transaction.

The model has now identified a potential money laundering transaction, that transaction should now be researched and determine if it is a false positive. Conversely, if it is brought to the bank's attention from an outside source (a counterparty bank) that a transaction is a money laundering transaction, and it was not flagged internally, it should be identified as a false negative.

- Given:
  - Inputs to transactions identified as false positives and false negatives
  - Inputs to transactions identified as true positives and true negatives
- Use:
  - Random Forest
- To:
  - Reduce the number of False Positives and False Negatives

The Anti-Money Laundering Model as a whole then becomes a very iterative process. We would take our transactions that were flagged as potential money laundering transactions and input them into the Random Forest Model. If the transaction is identified by the model as still potentially being a true positive, the bank would then review the transaction. If it is identified as a money laundering transaction, and it is a customers'/counterparty's first transaction flagged, they would be flagged for input to rebuild the model from scratch all over again. The risk scores would be adjusted, the all the subsequent models that flagged those transactions would be adjusted. Considering Landsbankinn went from over 1000 false positives a day to approximately 100 a day over many months indicates that this model was likely rerun daily. Slowly, over the course of many months of iterations, the bank reduced the amount of transactions needing reviewed on a daily basis.