INFORMATION SYSTEMS EDUCATION JOURNAL

Volume 19, No. 1
February 2021
ISSN: 1545-679X

Teaching Cases - Special Issue

In this issue:

iscap
Information Systems & Computing Academic Professionals

The **Information Systems Education Journal** (ISEDJ) is a double-blind peer-reviewed academic journal published by **ISCAP** (Information Systems and Computing Academic Professionals). Publishing frequency is six times per year. The first year of publication was 2003.

ISEDJ is published online (https://isedj.org). Our sister publication, the Proceedings of EDSIGCON (http://proc.iscap.info) features all papers, panels, workshops, and presentations from the conference.

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the EDSIGCON conference. At that point papers are divided into award papers (top 15%), other journal papers (top 30%), unsettled papers, and non-journal papers. The unsettled papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in the ISEDJ journal. Currently the target acceptance rate for the journal is under 40%.

Information Systems Education Journal is pleased to be listed in the Cabell's Directory of Publishing Opportunities in Educational Technology and Library Science, in both the electronic and printed editions. Questions should be addressed to the editor at editor@isedj.org or the publisher at publisher@isedj.org. Special thanks to members of ISCAP/EDSIG who perform the editorial and review processes for ISEDJ.

INFORMATION SYSTEMS EDUCATION JOURNAL

Teaching Case

Can you Predict the Money Laundering Cases?

Richard V. McCarthy
Richard.mccarthy@quinnipiac.edu

Wendy Ceccucci
wendy.ceccucci@quinnipiac.edu

Computer Information Systems
Quinnipiac University
Hamden, CT, 06518, USA

Mary McCarthy
Mary.mccarthy@ccsu.edu
Accounting Department
Central Connecticut State University
New Britain, CT 06050, USA

Nirmalkumar Sugurmar
Nirmalkumar.Sugumar@peoples.com
People's United Bank
Bridgeport, CT 06604, USA

Abstract

This case is designed to be used in business analytics courses; particularly those that emphasize predictive analytics. Students are given background information on money laundering and data from People's United Bank, a regional bank in the northeast United States. The students must develop their hypothesis, analyze the data, develop and optimize predictive models, and then score the models. Students are challenged to develop a better baseline model than what is currently being used by People's United Bank.

Keywords: Anti-money laundering, business analytics, predictive analytics, SAS Enterprise Miner

1. INTRODUCTION

This case is designed to be used in a business analytics course with a focus on the development and subsequent optimization of predictive models. Six of the nine steps in the Predictive Analytics Process Model (see Figure 1) (McCarthy, McCarthy, Ceccucci, Halawi, 2019) are reinforced through this case. The case begins with the development of a hypothesis that supports the business problem that is described herein. Data is provided with the case. The data must then be analyzed and manipulated for analysis. Predictive models are then developed and subsequently optimized to determine which model provides the best fit.

The case provides an opportunity to build multiple predictive models using a variety of tools (e.g., SAS, R, Python) and to optimize those models. Once a best fit model has been selected, it can then be implemented. One of the requirements for selection of the best fit model is to produce a model that exceeds the benchmark within this case. A scoring data set is provided to analyze how the model supports ongoing production needs and to emphasize that the purpose of predictive models is to apply them to future business activities/decisions.

Students are provided the background of the business problem, a data set for building and testing predictive models and a data set to score the best model. The business problem is to develop an optimized predictive model to determine which cases must be investigated for potential money laundering. Next, the details of the business case are presented; beginning with a definition of money laundering and why it is a critical issue for the banking industry.

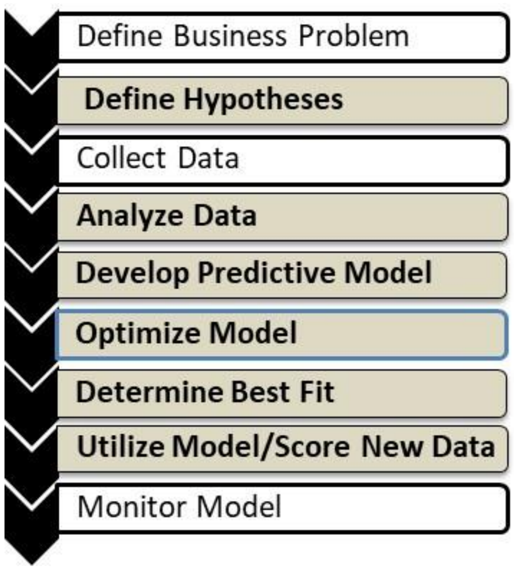

Figure 1 Predictive Analytics Process Model (McCarthy, et. al., 2019)

Money Laundering

Money laundering is the illegal process of concealing the origins of money obtained illegally by passing it through a complex sequence of banking transfers or commercial transactions. An attempt to make large sums of money obtained through illegal activities to look legitimate through banking transactions (Dreyer, 2011). The term money-laundering comes from this process of taking 'dirty' money (i.e., money from illegitimate, criminal activities) and transforming it into 'clean' money (i.e., money that appears legitimate and cannot be traced back to the criminal activity). Money laundering is estimated to be a one to two trillion-dollar problem and may represent two to five percent of the gross domestic product of the entire world (Ruce, 2011).

Dreyer (2011) describes three layers of money laundering:

1. Placement – moving the funds from the criminal activity that generated them.
2. Layering – using complex financial transactions to disguise the funds.
3. Integration – making the money available for subsequent use.

Initially, money laundering was considered to be a tool used by drug dealers and racketeers. More recently, it is also viewed as a tool used by terrorists to finance illicit activities.

Bank Secrecy Act

The Currency and Foreign Transaction Reporting Act (1970) placed a requirement on banks in the United States to work with the U.S. government to investigate money laundering (this act is more commonly known as the Bank Secrecy Act (BSA)). This act placed a requirement on banks to report transaction activity that the government considers useful in monitoring criminal and tax matters (Ruce, 2011). The major tool to perform this monitoring was the creation of the *Suspicious Activity Report* (SAR). A SAR must be filed when a bank has knowledge of, or suspects that, a financial transaction has occurred as a result of funds earned through illegal activities. This report is one of the primary tools used to combat money laundering. There are millions of transactions that occur on a daily basis. The vast majority of these transactions are legitimate. However, while money laundering involves a relatively small number of transactions compared to the total number their consequences can be severe. Under reporting of suspicious activity could therefore lead to an increase in money laundering activity. Conversely, over reporting can result in the investigation of to many legitimate transactions making it more difficult to focus on those that should be scrutinized (Meltzer, 1991). It is important therefore to balance both of these issues and still meet the requirements of the Bank Secrecy Act. A report must be filed when:

1. The transaction is designed to evade the requirements of the BSA or

2. The transaction has no apparent lawful purpose and the bank has no knowledge of a legitimate business reason for why the customer would engage in the activity resulting from the transaction (Ruce, 2011).

The Bank Secrecy Act required the reporting of cash transactions that exceeded $5,000. The threshold was subsequently changed to $10,000 in 1984. This resulted in attempts to circumvent this requirement by structuring multiple transactions that were each below the $10,000 threshold. This caused requirements to be put in place to identify attempts to subvert this threshold. To address this problem, in 1986, the Money Laundering Control Act added to this requirement by criminalizing money laundering. The act defines *specified unlawful activities* (SUA's). These include attempts to conceal the source, control or ownership of funds (Salinger, 2013).

As a result of the September 11, 2001 terrorist attacks on the United States, the *PATRIOT Act of 2001* was passed to strengthen the penalties for terrorist acts that occur either domestically or abroad. This included money laundering and terrorist financing. It enhanced the due diligence requirements and SAR reporting requirements (Sensenbrenner, 2001).

2. BUSINESS PROBLEM

To comply with the requirements of the Banking Secrecy Act and subsequent legislation, banks must have a system in place to identify suspicious activity that has the potential to involve money laundering or terrorist financing. The volume of legitimate banking transactions that they process on a daily basis makes it unrealistic to manually evaluate every transaction. Therefore, they have systems and controls in place to flag suspicious activity for further investigation. It is helpful to have develop a probability for each transaction that indicates the likelihood that a transaction is suspicious for prioritization purposes (Mehmet and Buchholtz, 2014). The system generates alerts using automated rules. The alerts need to be reviewed and investigated by anti-money laundering (AML) analysts. The goal of the alert system (AML System) is to come as close as possible to identifying only those specific transactions that involve these illegal activities. If too many false alerts are generated, it consumes too much time on the part of the analysts to investigate each one. If too few alerts are generated then there is the potential to miss transactions that support illegal activity. The better the model for evaluating alerts, the closer the system gets to its optimal performance.

When an analyst reviews an alert, they either create a case for further investigation because they suspect fraud or they close the alert. Cases requiring further investigation are sent to an operations team for review. Upon completion of the review, cases that still require further investigation require the filing of a Suspicious Activity Report (SAR). Alerts are reviewed based upon priority (i.e., the higher the probability of suspicious activity).

3. ANALYZE DATA

The data for this case was provided by People's United Bank and consists of 38,515 transactions that were a sample of production transactions from October 2014 through September 2015. This data set was used to produce their model. The data set and accompanying documentation serves as the baseline to develop a predictive model that improves upon this baseline.

People's United Bank is a diversified financial services company with more than $60 billion in assets. Founded in 1842, People's United Bank is a community-based, regional Northeast bank with more than 6500 employees, over 400 retail locations in Connecticut, New York, Massachusetts, Vermont, New Hampshire and Maine. There are over 148 branch locations co-located within Stop & Shop grocery stores in Connecticut and New York. They offer full-service, extended hour retail banking and commercial banking and as well as wealth management services.

Two data sets were created from the dataset provided. The first data set is used to develop and train a predictive model to determine if a case requires further investigation for potential money laundering (i.e., the case should result in a SAR). The first data set contains 35,000 observations (records). The data dictionary is presented in Appendix A.

The second dataset used for scoring, contains the remaining 3,515 observations. This dataset is used to test the best fit model and determine differences in cases selected for further review between the best fit model and the baseline model.

The datasets are provided in an Excel, CSV or SAS file format providing flexibility for a variety of analytic tool use.

4. DEVELOPMENT AND OPTIMIZATION OF A BEST FIT MODEL

The baseline model was developed by People's United Bank by analyzing six iterations of predictive models. The best fit model was chosen using the Kolmogorov-Smirnov (KS) and Receiver Operating Characteristic (ROC) statistics. The best fit model was developed using 70% of the data to build the model and 30% to validate the model. The best fit model for this baseline was a backward regression model. The statistics for the model are presented in Table 1.

Statistic	Result
KS	44.13
ROC	0.7923
R-Square	0.1687

Table 1. Baseline Model Statistics.

Regression is one commonly used technique when developing predictive models, however many other techniques are available in the data analyst's toolbox. The goal is to develop a model that provides a better fit. Even a modest improvement can be significant as it will result in better control over the review of cases evaluating potentially illegal activities. Any predictive technique may be utilized, this is not limited to only using regression models.

5. FINAL REPORT

The best fit model enables a bank to identify and detect potentially illegal activity more accurately and quickly. It provides a more precise review of transactions that require investigation and can reduce the investigators workload.

In your final report, you should discuss the following:

1. Determine Hypotheses:
   What were the hypotheses that you tested? If any variables were excluded, discuss why they were removed from the subsequent analysis.

2. Analyze Data:
   Which variable(s) contained missing values and how were they treated? Why was the technique chosen appropriate? Which variable(s) contained outliers? How did you identify and handle them? What variable(s), if any, were skewed? How did you handle them? What partition size(s) were utilized? It is appropriate to replicate the 70/30 split that the baseline utilized; however, if a more optimal partition was utilized that did not result in over-training the model then discuss the results.

3. Predictive Model:
   For each predictive model technique, discuss the technique and the properties that resulted in the best fit model? Compare the results of the selection statistics and discuss which model produced the best fit. Present the results of all of the selection statistics. Compare the misclassification rate of each model type that was evaluated and discuss the difference in the results of both Type I and Type II errors. Recall that a type I error occurs when a true null hypothesis is rejected, and a Type II error occurs when you fail to reject a null hypothesis when it is really false. How do these errors impact investigators?

4. Scored Results:
   Which cases resulted in the five highest probabilities for generating a SAR? Were there any cases that were not previously flagged by the baseline model that were flagged by your best fit model?

9. REFERENCES

Dreyer, T., (2011, December), Anti-Money Laundering, *Credit Union Management*, 40-42.

H.R.3162 - Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT ACT) Act of 2001, https://www.congress.gov/bill/107th-congress/house-bill/3162

McCarthy, R., McCarthy, M., Ceccucci, W., Halawi, L., (2019), *Applying Predictive Analytics: Finding Value in Data*, Springer Publications.

Mehmet, M., Buchholtz, M., (2014), Money Laundering Detection Framework to Link Disparate and Evolving Schemes, *Journal of Digital Forensics, Security, and Law, 8*(3), 41-70.

Meltzer, P., (1991). "Keeping Drug Money from reaching the Wash Cycle: A Guide to the Banking Secrecy Act", *Banking Law Journal, 108*(3), 230-255.

Ruce, P., (2011). "The Bank Secrecy Act: Considerations for Continuing Banking Relationships after the Filing of a Suspicious Activity Report", *Quinnipiac Law Review, 30*(1), 43-68.

Salinger, L.M., (2013), *Encyclopedia of White Collar and Corporate Crime*, Sage Reference, Thousand Oaks, CA., 1(2), p83-84.

Sensenbrenner,J.,(2001), (https://www.congress.gov/bill/107th-congress/house-bill/3162

.

Appendix A – Data Dictionary

Variable	Type	Note
ID	Unique ID	Unique identifier for alerts generated by the anti-money laundering (AML) system
Wires_mult	Derived	The number of wire transfers with more than $10K If an alert has 1 wire transfer more than $10K then wires_mult will be 0, If the same alert has 3 wire transfers with $10K transfer for each wire, then wires_mult will be 2.
Wires_size	Derived	The total wire transfer involved in the AML alert in proportion to the size of the total amount. =Sum (# of wire transfers)/10,000; rounded down to the nearest integer. e.g., if an AML alert has 25,000 wire transfers, wires_size will be 25,000/10,000 =2
Max_crr	Direct	Maximum cash reserve ratio score e.g.: ID 123 has 5 scores available in scoring table – the largest value is used.
Num_tran_alert	Derived	Number of distinct transactions involved in a specific AML alert
Num_acct_alert	Derived	Number of distinct accounts involved in a specific AML alert
Num_related	Derived	Number of related transactions – for that specific AML alert
Num_tran_type	Derived	Number of different transaction types involved in the AML alert
Num_tran_bin	Derived	Number of trigger transactions are grouped in bins by the following order: If num_tran_alert ≤ 5 ='1' 6 – 15 ='2' 16 - 25 = '3' 26 – 50 = '4' 51 -100 = '5' ≥101 = '6'
Num_acct_bin	Derived	Number of accounts involved in the AML alert are grouped by the following order. If num_acct_alert ≤ 1 ='1' 2 – 2 ='2' 3 - 3 = '3' 4 – 4 = '4' ≥5 = '5'
Trig_amt_bin	Derived	Trigger amount stratified in ranges – If the amount < 15000 ='1' 15,000 – 50,000 ='2' 50,000 – 100,000 = '3' 100,000 – 250,000 = '4' 250,000 - 500,000 = '5' 500,000 – 1,000,000 = '6' 1,000,000 – 2,000,000 = '7' >2,000,000 = '8'
Rela_amt_bin	Derived	Related transaction amount stratified in ranges – If the related amount

		< 15000 =’1’ 15,000 – 50,000 =’2’ 50,000 – 100,000 = ‘3’ 100,000 – 250,000 = ‘4’ 250,000 - 500,000 = ‘5’ 500,000 – 1,000,000 = ‘6’ 1,000,000 – 2,000,000 = ‘7’ >2,000,000 = ‘8’
Scen_Cat_code	Derived	If an AML alert belongs to a certain scenario then it is set to 1 else it is set to 0. This is computed by analyzing the last 12 months of AML alerts.
Num_rela-bin	Derived	Number of related transactions are grouped in bins by the following order. value numfmt = ‘1’ < 5 =’1’ 6 – 15 =’2’ 16 - 25 = ‘3’ 26 – 50 = ‘4’ 51 -100 = ‘5’ 101 > = ‘6’
Num_trigger	Derived	Number of trigger transactions for that specific AML alert
Prod_ind	**Target**	Binary, 1- indicates the alert is a productive alert- Productive is defined as that particular alert for further investigation for AML related activities; 2- 0 – indicates the alert is not a productive alert