

# Predictive Tagging

This document describes why predictive tagging is important

## What is tagged Data?

Tagged data is historical transaction data where all transactions which are known to be fraudulent are marked. Any unmarked transactions are assumed to be legal.

## Why is Tagged Data Important?

A tagged transaction set can be used to train a classifier. The classifier can then be used to score unseen transactions as to how similar they are to previous known fraud instances. The more consistently a set of training data is tagged, the better a classifier will perform, regardless of which learning algorithms it uses.

## Why is obtaining tagged data a problem?

Few banks keep accurate transaction level historical fraud data as their business processes are not designed to collect this information. Information is more likely to be available at the account level, for example:

- How many payments have been missed?
- Has the credit limit been breached?
- Has the account been blocked?

The payment history information will be used in conjunction with the customer profile and recent spending behaviour to assess the risk of default.

In addition, there may be a separate log of accounts maintained by a fraud department which contains a list of accounts containing confirmed fraud and a list of accounts under investigation. There will be some ambiguity as to what is and isn't fraud. For example, a card maybe reported lost in order to disguise a runaway debt problem or specific transactions may be contested. Accounts could be suspended and debts written off without a formal resolution as to the specifics of responsibility, especially if the losses are relatively minor.

In order to generate transaction level tagging, an 'autopsy' must be conducted for each confirmed fraud case. This is likely to occur many weeks after the first problem transactions occur. In many cases the information can be simply obtained from the card holder. There maybe a clear cut-off data after which all transactions are fraudulent or, in the case of cloned cards, there may be a mixture. In other cases the card holder may be unobtainable or the information supplied by them may be unreliable or untrusted.

In order for a training set to be generated for a classifier, it must be possible to transfer the tagging on to the original transaction data (containing the full set of accounts). If fraud cases are kept in a separate database then this may be problematic.

When confirmed fraud cases have been dealt with, a view must be taken on those cases for which fraud is suspected or which are still being contested. There may be more such cases than there are of actual confirmed fraud.

## What is the Ideal Tagged Data Set?

A classifier can only learn and generalise from the fraud (tagged) and negative (legal) examples given to it. If the untagged transactions contain many instances of fraud which have been missed for procedural reasons, this will have a major detrimental effect on performance. If it is desired that the classifier picks out the suspicious or undesirable transactions that a human expert would identify, then the more such transactions that are marked in the training data, the better. What we don't want, is lots of accounts containing suspect transactions or delinquent spending patterns muddying the legal transaction set.

The problems with existing training regimes are as follows:

- There is a long delay between fraud occurring and transaction data being ready for training. This has a twofold impact. First the system is always lagging behind both legal spending trends and fraud patterns. Secondly, there will always be a large number of as yet unidentified fraud cases mixed in with the legal data.
- Tagging the training data is a labour intensive and error prone process. System performance may strongly benefit from tagging suspected as well as confirmed fraud and from removing unresolved cases from the training set. This would be an additional overhead in data maintenance.

## Predictive Tagging

What a conventional system effectively does is try to estimate the probability that each transaction will turn out to be a confirmed fraud (which will ultimately be tagged and end up as training data).

Predictive tagging is a different approach to classifier training which estimates the probability that an account will be suspended or have the majority of its transactions refused in the near future. The training data consists of the list of accounts containing either confirmed or strongly suspected fraud together with all accounts which are significantly behind with repayments or which have been suspended. This information is fed into a process which automatically tags the data based on date information and the reply and settlement codes in the transaction set.

The end result is a system which gives early warning of problematic accounts without specific reference to fraud; that decision is left to the human operators. However, transaction level tagging of fraud can still be added into the system at any stage with a beneficial effect.