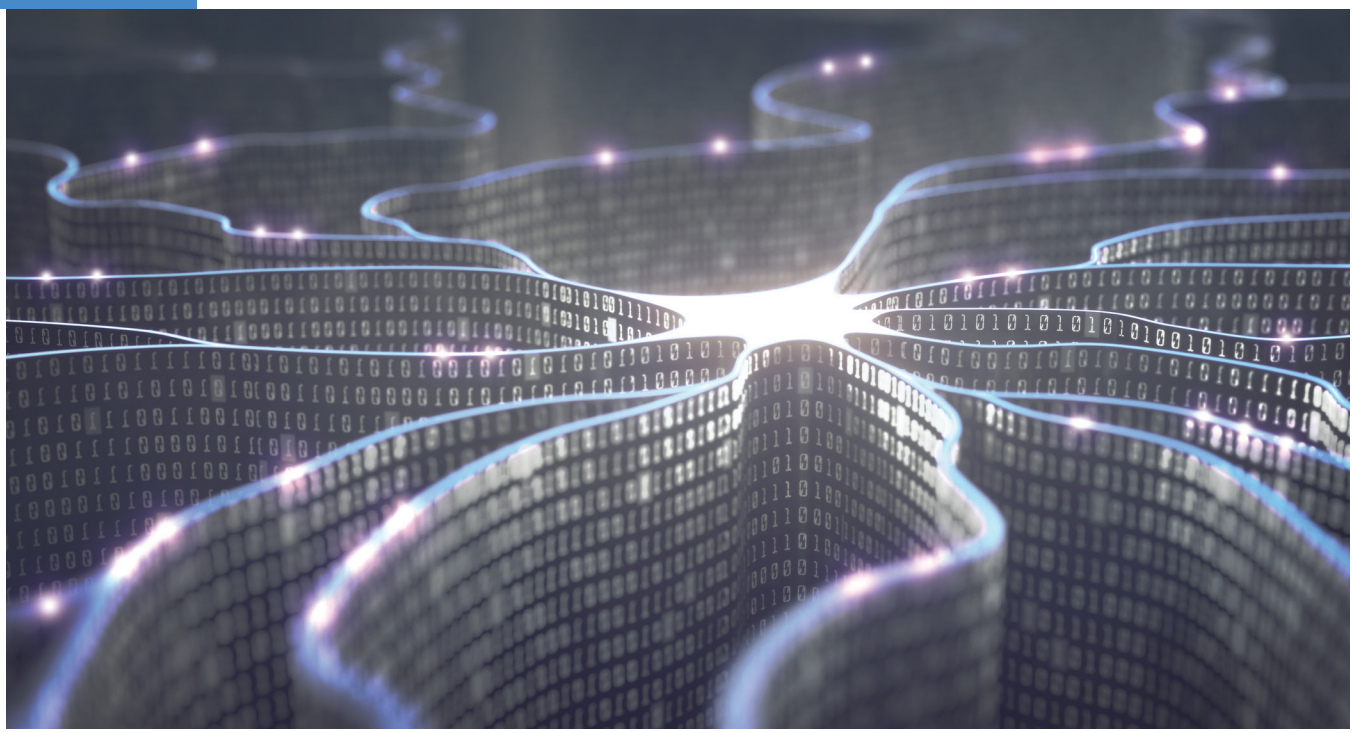
TIBCO®



Busting Financial Crime with TIBCO

— Ana Costa e Silva, PhD
Senior Data Scientist, TIBCO Software

What if you could use just one financial crime fighting solution that would empower your business users to improve handling of financial crimes such as anti-money laundering (AML), credit card fraud, trade surveillance, or medical fraud? Current financial crime fighting systems have a number of disadvantages, including:

- They flag too many false positives that cause investigators to focus on the wrong cases.
- They involve manual procedures that result in investigations taking too long to complete.
- They tend to be a "black box" requiring expensive consultancy to keep ahead of the fraudsters' ever-changing tactics.

In the following pages we explore how TIBCO's financial crime solution addresses these disadvantages.

PRODUCING MORE RELEVANT ALERTS

TIBCO's approach to fighting financial crime places machine learning at the center of the crime detection system. Machine learning models use historic data to learn how to spot risky or abnormal behavior exhibited by transactions, clients, suppliers, or other players. It uses two types of models:

Supervised learning algorithms, which tell us how similar to past fraud a new transaction is.

Unsupervised learning algorithms, which tell us how odd a new transaction seems when compared to past transactions.

TIBCO SPOTFIRE

TIBCO Spotfire® data visualization and analytics software delivers the most complete set of analytics to empower every individual to develop and depict critical insights for faster, better decision-making. With Spotfire, organizations can seize new business opportunities and avoid risks with unmatched speed and flexibility. Using interactive dashboards, visualizations, and predictive and event-driven analytics, users can develop insights immediately on any device. Spotfire is an enterprise-class analytics platform that helps both business and technical users quickly explore data to develop actionable insights, without requiring IT intervention.

TERR

TIBCO® Enterprise Runtime for R (TERR) provides a fast and powerful enterprise-grade platform on which to run a huge variety of advanced analytics based on the popular R language. In addition to broad R package compatibility, TERR delivers superior performance and memory management for running existing scripts and large datasets. TERR is:

- Embedded in TIBCO Spotfire® Desktop and TIBCO Spotfire® Analyst, with no additional installation needed, and in third-party products and custom applications
- Integrated into the TIBCO platform, including TIBCO StreamBase® and TIBCO BusinessEvents® streaming analytics for real-time predictive model scoring
- Deployable in big data environments (such as Apache® Hadoop® and Spark®) and in grids (via TIBCO GridServer®) for fast, advanced analysis of big data
- Free to individual R users from tap.tibco.com

The first model guarantees accuracy, the second the ability to adapt to changing realities.

These models learn from history and then these learnings are applied to the present, either in real time or in batch, by simply scoring current data against the models. Transactions that are found to be fraud-like or odd beyond a certain threshold will be manually investigated. The setting of the threshold is a business decision supported by what-if analyses in TIBCO Spotfire®.

The following examples explain how the solution can be trained to detect different types of crime.

SUPERVISED LEARNING ALGORITHMS

Modeling with supervised learning algorithms involves obtaining data of confirmed fraudulent and non-fraudulent cases. For example, for a list of transactions monitored for AML, one column contains a value of "1" when past transactions were fraudulent and a value of "0" when non-fraudulent. Decision trees, random forests, neural networks, support vector machines, and logistic regression are all examples of supervised learning algorithms. Given everything else known about each transaction, they generate optimal ways (models) of separating the 1s and 0s to obtain true positives with a minimum of false positives.

A model is a summary of a pattern in the historic data and therefore a much smaller representation of the original data. For example, the decision tree in Figure 1 could have been learned from millions of lines of historic data.

```
If value of the transaction > 40K then
    If proportion of value over average balance is >85% then
        Probability of a transaction being Fraud = 80%,
    else Probability of a transaction being Fraud = 30%
else Probability of a transaction being Fraud = 10%.
```

Figure 1 Example of the result of a decision tree algorithm on a big data set.

TIBCO Spotfire is a visual analytics software tool that allows you to run advanced statistical models. An easy-to-use Spotfire template can guide the business end-user through the steps of building and testing model types, even if the user has no deep knowledge of statistics or data science. The user just needs to understand the business.

Figure 2 below shows an example of such a template. On the left hand side under "Response" the user chooses the variable he wants to model. As an example for credit card fraud, the variable would contain the value 0 for all non fraudulent transactions and 1 for the fraudulent ones. Under "Predictors," the user chooses which features, or columns, of the data he wants to use to build the model. Then under "Models," the user has the option to try different types of models.

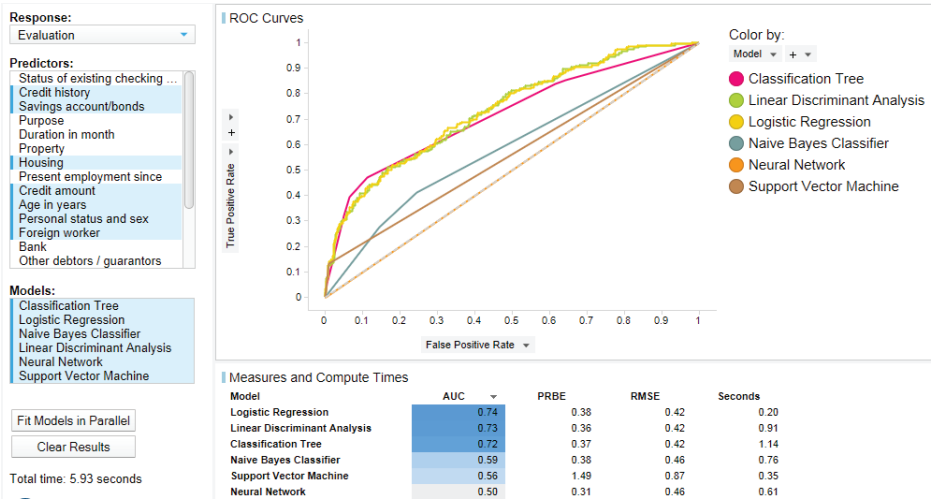


Figure 2 Example of a Spotfire template where business users can train and test different supervised models.

Spotfire can automatically detect the presence of new columns in a dataset, and add them to the list of Predictors, which means the user can quickly and dynamically adapt a Spotfire template to include new features. The user is also free to include a number of algorithms or just one. When the user presses the button "Fit models in parallel," Spotfire calls out to its statistics engine to run the relevant calculations. Results, including any quality tests performed, are published back to Spotfire. In the example in Figure 2, the user only needs to know that the best model is arguably the one with highest Area Under the Curve (ACU) as shown in the table.

The Spotfire data function carries out all of this work in the background typically needs to be developed by a data scientist. Data functions are calculations using your preferred statistical scripting language or workflow tool and are designed for collaboration. Once the data function is created, it can be easily shared. Any analyst can use data functions without needing to know any coding. All appropriate business users are empowered to make better decisions, including creating fraud models, without being exposed to unnecessary complexity. Spotfire supports data functions in different statistical engines, such as TIBCO® Enterprise Runtime for R (TERR), which is embedded in Spotfire, as well as in open-source R, SAS, Matlab, KNIME, and Lavastorm.

UNSUPERVISED LEARNING ALGORITHMS

Using only supervised models is not enough. In fact, some companies may be starting a financial crime fighting unit without historic knowledge about which transactions were fraudulent and which were not. Even when you do have historic knowledge, it is never certain whether all past fraud cases were correctly identified. And even when this certainty is fairly high, fraudsters are creative, and if one strategy did not work, they will try a new one that will leave a different fraud trail behind.

TIBCO recommends using a combination of supervised and unsupervised models, and the TIBCO solution accommodates this.

Unsupervised models have no requirements for prior knowledge about which transactions were fraudulent and which were not. Without a goal variable per se, this type of algorithm aims to capture what is "normal" in the data and which different types of normal there are. Clustering algorithms and self-organizing maps are examples of this type of model. When applied to financial crime data, these methods allow for profiling normal operations and spotting unusual ones. Unusual does not mean criminal, it means warranting human verification.

Figure 3 below contains an example of a Spotfire template that, with minimal training, a business user can use to develop an unsupervised model. In this case, a well-established matrix operation called Principal Component Analysis (PCA) is used to represent all transactions. The relevant components are shown on the axes, where normal transactions appear close to the origin of the chart (0,0 point) and abnormal ones farther from that point. The distance of any new transaction to this origin is a measure of its oddity. This relevant information is not the result of any human assumption, but is derived directly from the pattern drawn by the whole history of transactions. Transactions that are unusual beyond an agreed threshold should be investigated.

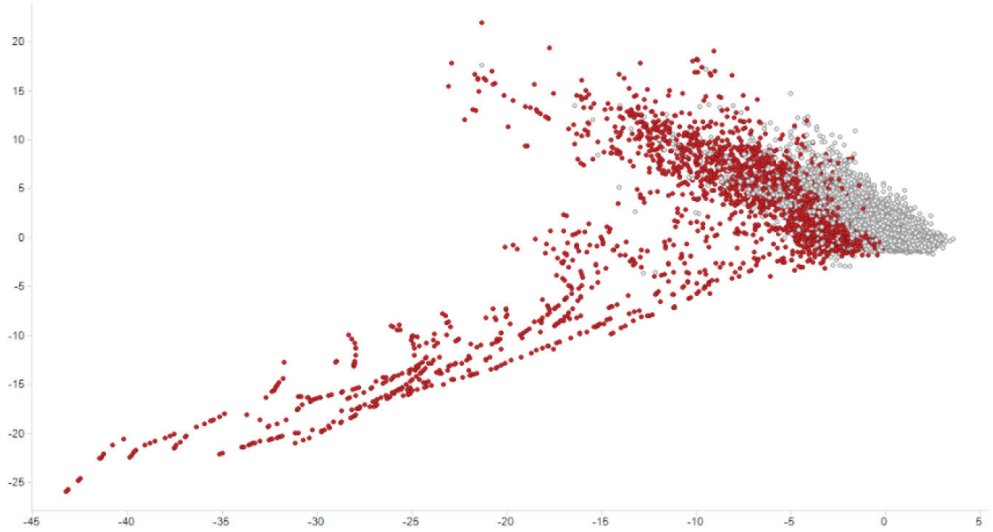


Figure 3 Example of a Spotfire template in which business users can develop an unsupervised model.

GOOD FEATURES MAKE GOOD MODELS

Any predictive model is as good as the features put into it. One challenge companies often face is identifying which characteristics to focus on to identify fraudulent events. Good fraud features are those that allow spotting unusual behavior. Often external business consultants are hired to suggest such features, but it's been our experience that the best features are already intuitively known by the experts in the firm. Consultants often gather business knowledge from internal experts and translate this knowledge into quantifiable features that can be extracted from their databases using SQL. For example, in AML, some relevant features are:

- Total amount of cash withdrawn. Unusually high values warrant an investigation.
- Value of withdrawal as a proportion of the account owner's average balance. High value withdrawals in relation to the average account balance are worth checking.
- The amount of time between the withdrawal and a previous deposit of a similar amount.

- The value of the withdrawal as a proportion of the value of the previous deposit.
- The value of the withdrawal as a proportion of the mean withdrawal of customers who share similar characteristics (gender, age, income, etc.) or of companies in the same economic sector, size, and region.
- Whether the account owner has a family relationship with one of the bank staff, etc.

Many of these features are gathered and monitored with systems such as Actimize, but are treated with individual, and not mathematically optimized, thresholds. Mathematical models can combine all features optimally.
A different set of features would be used for finding fraud in medical insurance claims. Imagine an insurance policy that covers 100% of all emergency claims. An obvious fraud for this policy is to declare routine procedures as emergency. The incentive for this potential fraudulent behavior is specific to the set-up of this particular insurance policy. This is why your business people, who understand the terms of your different policies, will know best what would be unusual and potentially fraudulent behavior. A conversation with the database administrators is all they need in order to derive the optimal SQL that helps grasp the relevant features. Relevant features for this policy might be:

- Total number and proportion of emergencies by doctor / clinic / patient
- Time between an emergency appointment and the purchase of the prescribed medicine
- Time between emergencies per patient and per family

When features have been crystallized into SQL, Spotfire can collect this data straight from the relevant datasource and visually portray how transactions behave. For example, on the left hand chart of Figure 4, it's easy to spot odd behavior. In Spotfire you can select the people or transactions that display the oddest behavior and list them. A quick investigation of a few cases will provide a better feel for the usefulness of the features in detecting criminal activity.

If historic data already contains the information on which transactions were fraudulent or non-fraudulent, this knowledge can inform the search for fraud revealing features. On the right hand chart of Figure 4, a zoomed-in box plot shows that transactions with higher value have also been more likely to be fraudulent.

Network charts provide another rich source of insight to identify people who have a big impact on the network as a whole, for example, people or organizations receiving a large amount of cash deposits from many different people.

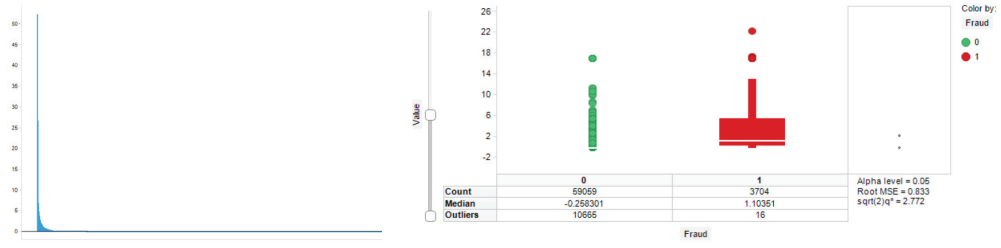


Figure 4 Visual analysis of AML related features with no past knowledge about which were fraudulent. The left visualization is deceivingly powerful and plots the number of credit card transactions from all users in the last 24 hours. It shows that the majority of people have very stable behavior (around 0), with just a few users showing unusual values that would merit investigation. The right shows a distribution of selected variables by status showing that higher value transactions are more often fraudulent.

STREAMING ANALYTICS

TIBCO's streaming analytics solutions allow you to capture, aggregate, and analyze real-time and historical data of any variety, volume, and velocity to gain contextual awareness and act preemptively. Our technologies supply the ability to:

- Understand historical patterns and dynamic event sequences
- Anticipate by monitoring all event streams, filtering for certain sources or qualities, correlating in real time, and detecting meaningful patterns
- Act by first testing measures of event significance, then setting business rules that drive action, which could include creating new events to be fed back into the system for discovery
- While capabilities vary based on goals and requirements, our best-in-class technologies rapidly capture, analyze, and act on any complex combination of events in real time:

TIBCO BusinessEvents®
TIBCO StreamBase®

One does not necessarily need to visualize every feature, however. Especially for big data containing many features, it may be impossible to visualize them one by one. The results of the supervised model can guide the search for the most relevant features for spotting fraudulent activity. We should then visualize these individually. Figure 5 shows which features have a higher contribution to the model (the longer the bar, the more important the feature). Although Figure 4 showed that "Value" (eighth in rank) is important, other features are more powerful at distinguishing fraudulent transactions.

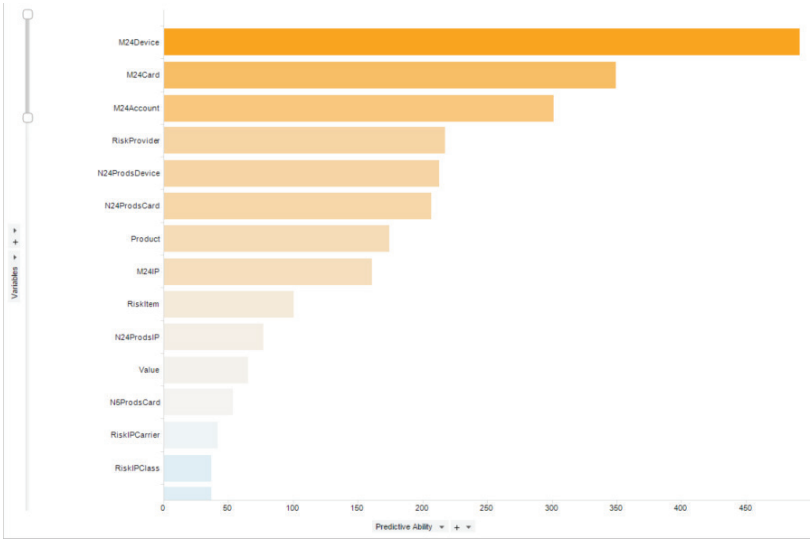


Figure 5 Visualizing the contribution of different predictors used in a model.

A combination of visualizing the ranking of the features as well as the detail of the individual features is important for a number of reasons:

1. **Validation of the model's quality.** Maybe your best feature is so good because it is part of the answer and should therefore be excluded. For example, if you inadvertently included the total value of fraud that was stopped as a predictor, that will obviously (and erroneously) appear as the best predictor.
2. **Correlation is not causation.** It is necessary to ask questions that lead to a better understanding of the reality being predicted.
3. **Validation of the data's quality.** Were you expecting a different feature to have more power than what is showing? Perhaps there are data quality issues causing a lack of relevance, or maybe outliers introduced a bias. These quality issues can be quickly spotted in a visualization.
4. **Surprising top features.** Sometimes predictors expected to be irrelevant turn out to have huge predictive ability. This knowledge, when shared with the business, will inevitably lead to better decisions.
5. **Inspiration for new features.** Sometimes the most informative features are the reason to delve into new related information as a source of other rich features.
6. **Computational efficiency.** Features with very low predictive power should be removed from the model as long as the prediction accuracy on the test dataset stays high. This ensures a more lightweight model with a higher degree of freedom, better interpretability, and potentially faster calculations when applying it to current data, in batch or real time.

BUSINESS PROCESS MANAGEMENT

Your processes should conform to your business requirements, not to your system capabilities. TIBCO's completely model-driven business process platform provides the complete spectrum of business process styles as well as unprecedented scalability and performance to handle all of your business process needs:

- Use a model-driven environment to speed and simplify process design, shielding implementation complexity with a fast, collaborative, and iterative approach
- Work with any process style: human and system integration processes, human workflows, dynamic and event-driven processes, case management, to-do lists, or approval processes in a single platform
- Alleviate IT involvement in day-to-day changes. Business users can adjust and change their operations immediately to take advantage of opportunities or avoid threats
- Rely on a native integration foundation for true business digitalization, allowing your data, people, processes, systems and things to be easily and seamlessly brought together to support all your business initiatives

TIBCO ActiveMatrix® BPM

HAVE YOUR MODELS, NOW WHAT?

Once business users have created and tested the new models that efficiently spot potentially fraudulent transactions; They can use Spotfire as an interface to manage how the models will be deployed. The models allow boiling all incoming predictors into two measurements: fraudulence and abnormality. It is now time to set adequate thresholds for them. Embedded what-if analysis in the Spotfire template allows setting thresholds for the two metrics and balancing the expected number of alerts with the size of your investigative team. Notice that these two metrics are machine-learning optimized to combine all incoming predictors in the best possible way.

HOW TIBCO DEPLOYS MODELS IN REAL TIME

Because a model is a summary of historic data, it can be as light as an equation and live beyond the data. Once the user is satisfied with the quality of the model and the respective thresholds, a press of a button in the Spotfire template is all that's required to send that model to the real-time event processing engine that will monitor transactions as they occur. TERR, TIBCO's statistics engine, has excellent integration with all of TIBCO's streaming analytics products to support this automated capability.

REDUCING INVESTIGATION TIME

How does TIBCO propose dealing with the second biggest flaw in current fraud detection systems: the fact that each alert takes too long to investigate?

TIBCO's streaming analytics software sits elegantly in the background. It seamlessly receives models updated from Spotfire and uses these to score in real-time every single transaction for its probability of fraud and abnormality. In TIBCO's Fraud Accelerator, a TIBCO StreamBase® workflow receives the model(s) and threshold(s) from Spotfire. It keeps track of model versioning, applies the model to the streams of transactions in real time, separates those that exceed the respective thresholds, and for each of these alerts creates a new case in TIBCO's Business Process Management tool (BPM). StreamBase also collates the context of each alert from any number of data sources using Spotfire and sends an email to investigators warning them a new potentially fraudulent transaction has been spotted. It also sends all output data to TIBCO® Live Datamart and TIBCO LiveView™, which allows visualizing the flow in real time.

Figure 6 provides an example of an automatically generated email report. It includes the ID of the transaction, a link to the respective investigative template, the scores for probability of it being like past fraud and for its degree of oddity, the respective thresholds, and the model versions that generated the scores. It also includes two links:

1. To a new Spotfire instance, shown in Figure 7, which contains all the relevant context regarding the specific transaction—including data gathered from available data sources about all people or companies involved. The interactive nature of Spotfire enables users to complete a swift but solid investigation from any choice of web browser for computer or mobile device. In this case, the network chart shows that the sender has had transactions with two more people who had received AML alerts in the past.

FINANCIAL FRAUD ACCELERATOR

TIBCO accelerators are provided as fast start templates and design pattern examples.

The Fraud Accelerator uses Spotfire Analyst to guide data professionals through developing both supervised and unsupervised models to detect probability of fraud and transaction abnormality from a known dataset. Models are developed using the R programing language and then hot deployed to TIBCO StreamBase for evaluation at runtime using TIBCO Enterprise Runtime for R. When a transaction is scored it will either pass, or be flagged as probable fraud, odd transaction, or both. When this occurs the streaming analytics platform raises an alert and creates a case in ActiveMatrix BPM that facilitates the investigation of the potential fraud.

TIBCO Fraud Accelerator

2 To the case that has been created in TIBCO's Business Process Management (BPM) tool. BPM can serve as investigators' front-end, such that all actions taken on each and every alert are auditable any time in the future. BPM allows investigators to log in, see the cases that have been assigned to them, open the respective investigation, consult the context via BPM's embedded Spotfire instance (Figure 6B), and take action such as escalating the case, considering it fraud or non-fraud, or informing the relevant financial authority.
The embedded Spotfire instance also allows zooming out of each case and managing the pipeline of cases, identifying bottlenecks or procedural steps that should be restructured or automated, spotting users who follow inconsistent procedures or investigate in a biased manner, tracking model quality and suggesting model revision, though this can actually be done in real time as well.

In the Accelerator we have embedded model quality tracking into BPM, from which users can be warned of the need to manually revise the model. In reality, TIBCO provides a lot of flexibility regarding how to incorporate model quality tracking into a financial crime fighting solution. For example, tracking can be done in real time and model revision can be triggered automatically, if quality is perceived to be diminishing, using TIBCO Spotfire® Automation Services; the new model can be tested against the previous one and updated if deemed better; users can be notified automatically.

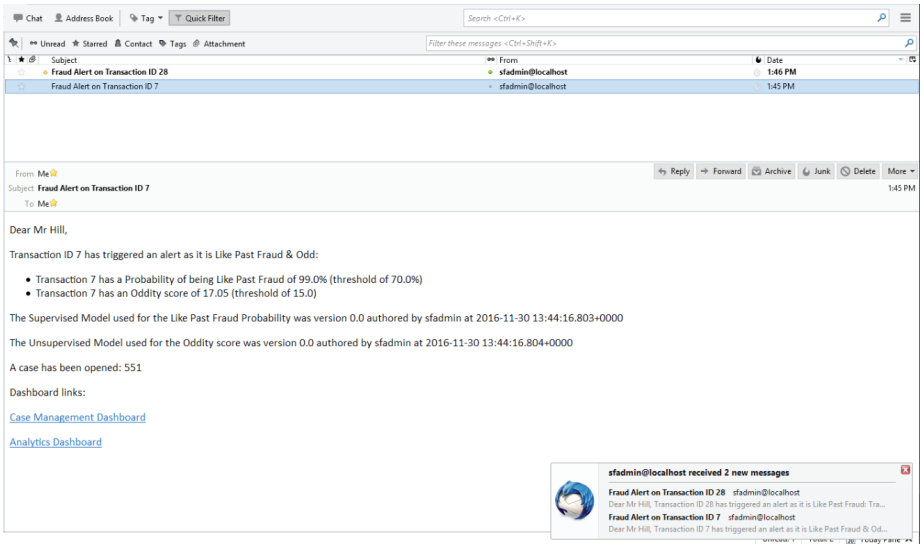


Figure 6 Example of a Spotfire email report including a visualization providing relevant context, the ID of the transaction, and a link to the respective investigative template (Figure 7).

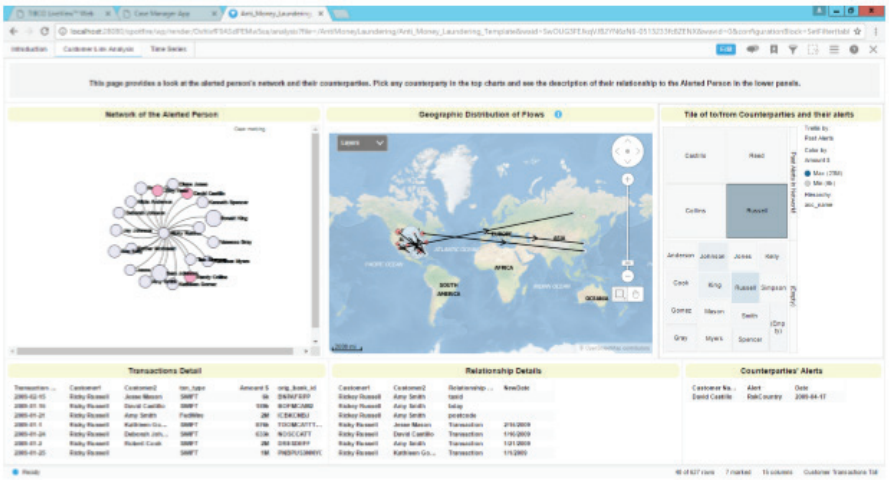


Figure 7 Investigative template that can be sent to an investigator in real time containing the entire context of a risky transaction.

SELF-LEARNING ABILITIES

One important advantage of TIBCO's solution is that, in time, organizations become better at spotting financial crime. First, by gathering better features; second, as the system generates alerts, transactions are investigated and classified as either a true or false fraud alert. This knowledge is fed back into the system and informs the next round of the supervised model. So naturally, the more often the system runs (with the latest information available), the better it gets at finding future fraud that looks like past fraud. This is what the term "machine-learning" actually means, a machine that is capable of learning on its own.

TRANSPARENCY

Another important advantage of TIBCO's solution is that it is not a black box and in fact all important parts are open to the business. For example, better models often come from better features. Business users can try new features in Spotfire from an easy-to-use dashboard and visually test their relevance (as we do in Figure 4). These features feed straight into the supervised model, which produces a clear measurement of their worth in detecting past fraud (as per Figure 5).

Another aspect of transparency is that the algorithms that create the supervised model, the unsupervised model, and the actual scoring of transactions are all open and can be consulted from Spotfire. In TIBCO's Fraud Accelerator, we used algorithms written in TERR. Your business users can apply these from easy to use dashboards and keep focused on the business at hand, not on the nuts and bolts of the technology or maths. However, because the algorithms are entirely open, your own data scientists can not only customize them but also replace them at will with their favourite approaches.

CONCLUSION

Understanding risk and opportunity in real-time is critical and many organizations already hold the data that makes this possible. However, all too often it is difficult to reach or wrangle quickly enough. The goal is to make better use of your data to build up better defenses and reduce fraud and financial crime losses. Too much effort is spent managing relentless fraud attempts without the incomparable speed and insight delivered by self-learning analytics.

The scale of today's problem means that you need easy access to constantly updating transaction information to be able to react precisely. TIBCO brings together different analytic solutions in a flexible, integrated platform that can be controlled by business users. It delivers continuously self-learning models fed by historic and real-time information, giving you a smooth user interface for intelligently improved customer experiences.

TIBCO proposes one modular financial crime fighting solution for anti-money laundering (AML), credit card fraud, trade surveillance, medical fraud and other financial crime, which:

- Monitors all transactions in one auditable, repeatable, self-learning process
- Increases customer satisfaction because it limits involvement to only those potentially exposed to real risk
- Increases investigative team productivity by only calling their attention to risky transactions that can be investigated with instantly provided context for optimal decision-making
- Puts advanced mathematics at your fingertips, transparently
- Enables real domain experts to apply their knowledge of specific business operations, without the need for a degree in advanced math or computer science
- Provides all this via easy-to-use dashboards built for business users

With machine learning at its heart, TIBCO's fraud-prevention platform enables you to monitor transactions as they occur and easily generate views of accurate, real-time information within the context of any suspicious transactions. This means you can expedite the investigation process so staff across your organisation can evaluate potentially risky transactions and make the right decisions quickly. Advanced analytics have been used for years; now is the time to move to the next level.

Learn more about the TIBCO Insight Platform by contacting us or by visiting www.tibco.com and spotfire.tibco.com.

TIBCO®

Global Headquarters
3307 Hillview Avenue
Palo Alto, CA 94304
+1 650-846-1000 TEL
+1 800-420-8450
+1 650-846-1005 FAX
www.tibco.com

05/12/17