Busting Financial Crime with TIBCO

TIBCO Connected Intelligence

The TIBCO Connected Intelligence platform gives you all the intelligence and insight you need, plus the ability to act on that intelligence in real time. TIBCO interconnects your systems, data, people, and processes—and lets you augment human intelligence by processing and presenting data for making faster data-driven decisions.

What if you could use just one platform to detect all types of major financial crimes? One platform to handle the analytical tasks of fraud detection, including:

- Data processing and aggregation
- Data visualization
- Statistical/mathematical/machine learning modeling
- Batch/real-time scoring

One platform that could successfully reduce complex and time-consuming fraud investigations by combining extremely different domains of knowledge including Business, Economics, Finance, and Law. A platform that can cover payments, credit card transactions, and know your customer (KYC) processes, as well as similar use cases like Anti Money Laundering (AML), trade surveillance, and crimes such as insurance claims fraud.

As criminals grow ever more sophisticated, regulations more stringent, and customers more intolerant of cumbersome security procedures, financial institutions face increasing business and technical challenges. We will examine several scenarios:

Criminal activities

Criminal techniques are becoming more sophisticated, for example, mule accounting (using an often coerced third party to collude in the payment chain), hibernation fraud (conducting regular transactions during an initial period of scrutiny, and when that period finishes, running up huge debt), under the radar transactions and front running (many small trades, and in front of the broader market).

Data growth

The growing number of customer contact channels generates huge volumes of data points that financial institutions have to continuously track. Can banks connect the dots to confirm if a transaction is fraudulent?

Social and network data

Because fraud is usually not something an individual commits alone, nor happens in isolation, the use of networked data in fraud detection is important. Behavioral and social network analytics offer new insights into the propagation of fraud, as patterns exist across groups of loosely connected people like an infection from one to another.

False positives

Rules-based detection systems are often slow to update fraud patterns and only apply a coarse-grained filter to transactions. The result is an excess of legitimate transactions unnecessarily flagged for investigation (false positives).

Inconvenienced customers/customer churn

Investigations can take too long, leading to customer dissatisfaction and churn. Investigators typically don't have immediate access to all the appropriate data needed to make a prompt decision, which inconveniences customers when their payments and /or credit cards are frozen.

Regulatory demands

Sophisticated machine learning (ML) models can be difficult to explain to demonstrate compliance, so institutions may be forced to use less sophisticated (but explainable) techniques that generate more false positives. This means they lose out on modeling results, meaningful visualizations, and the benefits of state-of-art technology (grid computing or Spark frameworks) to parallelize model parameters and fine-tuning.

In this whitepaper we explore how TIBCO's Connected Intelligence platform addresses these common issues.

TIBCO Connected Intelligence for Financial Crime

TIBCO's approach to fighting financial crime relies on six components for crime detection: data integration and feature engineering, data visualization, machine learning, model explainability, model deployment, and advanced case management.

TIBCO Data Virtualization

With less physical overhead and cost than a traditional data warehouse or ETL solution, TIBCO Data Virtualization software orchestrates access to multiple and varied data sources. IT delivers the datasets and foundational IT-curated data services for many analytics solutions. Users are empowered because they can easily search for and select from a self-service directory of virtualized business data and then use their favorite analytics tools to obtain results.

MDM for Data Accuracy

TIBCO EBX data management software makes it easier to manage your data assets, with rich capabilities for master, reference, and metadata, including simple setup of multiple lenses and user hierarchies. Custom applications and traditional purpose-built master data management (MDM) solutions are hard to change, while EBX software is flexible and agile. EBX software uses a unique what-you-model-is-what-you-get design, generating fully configurable applications on the fly. These capabilities eliminate the need for long, costly, and endless development projects. And EBX software includes all the enterprise-class capabilities you need to create data-driven applications. Data stewardship, workflow, data quality, and data integration are built right in. Get more information at https://www.tibco.com/products/tibco-ebx-software

Data Integration and Feature Engineering

The first challenge is gathering the data. You need to combine the representative data from disparate systems into a format that can be used for feature engineering, visual analytics, and modeling. To accomplish this, data can be sourced from a vast range of databases, applications, and unstructured sources through TIBCO Data Virtualization and TIBCO integration applications. It can then be used to both build models and provide runtime evaluation of deployed models.

Data is the root of model accuracy. Data used in visual analytics and modeling needs to be representative of the financial crime under consideration. More data is not necessarily helpful. Unsupervised learning models look for outlier transactions over the set of raw data attributes and derived features. Supervised models require transactions to be labeled as fraud or non-fraud; or some other measure related to the crime. Labeled data should be correctly categorized or the resulting model won't accurately capture the fraud pattern. Labeled data is often in short supply, so it is the combination of supervised and unsupervised learning that is most effective in predicting anomalous transactions.

Data interconnect solutions from TIBCO provide scalable data integration. You can access disparate data sources (cloud, mainframe, and streaming data such as from Internet of Things devices such as POS machines and mobile terminals, and engineer features from the raw data via transforms and calculations, including natural language processing and reference data inferencing. TIBCO takes data preparation seriously, providing functionality for multiple roles including data engineers, data scientists, and business analysts. Functionality includes convenient data profiling, missing data imputation, standardization, matching, enrichment, and active data-quality assessment.

Data Visualization

The second challenge is combining data into effective formats for holistic views. TIBCO Spotfire visual analytics allows users to combine data into interactive reports or views that maximize clarity and insight. Visual analytics across all data types is fast and easy to set up and configure to spot the fire in the data. The visual palette is rich, including scatter plots, bar charts, heat maps, tree maps, density plots, parallel coordinate plots, correlation matrices, tables and sparklines, time series, spatial, and geoanalytics, word clouds for unstructured data, and network charts for social connection. These visual analytics are easy to adorn, color, and shape with lines, curves, smoothers, clusters, and symbols. This brings immediate understanding and generates questions (and answers) that would not be possible from static charts or lists.

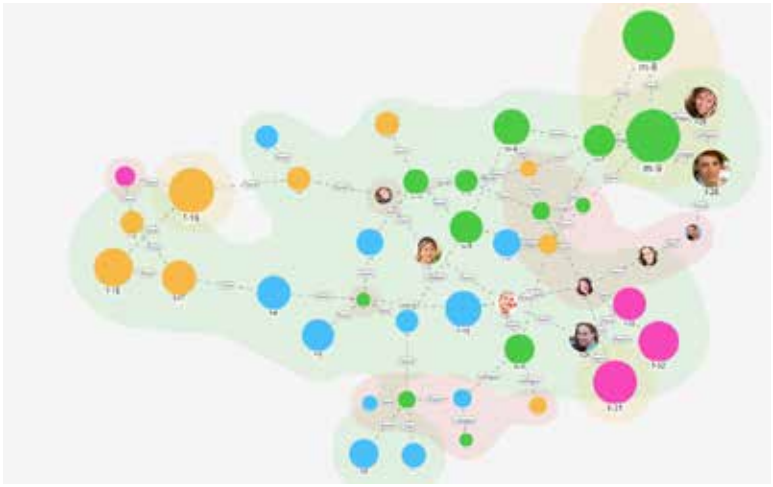

Figure 1. Spotfire network chart displaying social connections among customers.

Visualization also augments business intelligence by delivering AI-driven analytics, including interactive map charts for demographic/geographic correlation, and brush-linking with drill-down to details on-demand, contributing to context. These features help analysts easily identify patterns and trends in complex and messy datasets faster than previously possible. Having identified data of interest, specific queries can be launched from within Spotfire software to pull additional data for further analysis.

Anti-fraud analysts can interrogate data from any angle using dynamic filtering. By applying keyword mining, they can see the complete relationship between accounts, keywords, time, and patterns of activity. This speeds data analysis and supports more accurate and confident decisions across the organization.

Machine Learning

The next challenge is to take advantage of Artificial Intelligence. Machine learning (ML) models use historic data to identify anomalies indicating risky or abnormal behavior from transactions, clients, suppliers, or other market participants. The process uses two types of models:

- **A** — Supervised learning algorithms provide models for describing known patterns of fraud. These models are trained on data including some known and labeled fraudulent transactions.
- **B** — Unsupervised learning algorithms provide models for separating anomalies from usual transactions. These are trained on a collection of transactions that don't include any labeling.

Trained algorithms are then applied, either in real time or batch, to current data. Transactions that are identified fraudulent can be rejected and investigated. Setting of the threshold for investigation involves a trade-off between identifying true fraudulent transactions and minimizing any false positives. The threshold takes into account the cost of missing a true positive and the cost of investigating a false positive.

Machine learning is more than model training. To have good models that make accurate predictions, you need good data. This means there is an increasing emphasis on data aggregation, transformation, and feature engineering. Moreover, increasingly large and complex datasets have made it difficult to gauge the optimal specifications for applying data transformation and cleansing, for feature engineering, choosing the best algorithm, and for optimizing the parameterization of an algorithm.

TIBCO Data Science–Team Studio software is an enterprise analytics data science platform that uses a web browser interface to build data science workflows. The platform allows data scientists and business users to collaborate on projects using advanced analytics that leverage scalable in-database and in-cluster distributed processes. Data scientists, analysts, and data engineers can build machine learning workflows with drag-and-drop operators and notebooks, while leveraging the power of big data platforms. The collaboration interface allows the analytics team to share insights and data with the rest of the organization, driving action for the business. The time spent in manual trial and error of feature engineering and ML modeling can be enormous, and a number of predefined features within Team Studio aims to make preparation of data and ML algorithms simpler and faster, so that time can be better spent focusing on an operational business solution. TIBCO Data Science software provides support for the following:

Figure 2. The social media styled collaboration panel enables compliance officer and data scientist to communicate freely within Team Studio.

Data Preparation

Data preparation is the foundation and most important step in a complete data science process. Most organizations have easy access to a lot of often low-quality data, and a big part of the work is spent transforming and cleaning this data for use in feature engineering and model fitting. Within Team Studio, operators enable joining, transforming, and aggregating data held in disparate tables, and where values might be recorded at inconvenient granularities that need to be wrangled for further analysis. Data may be incomplete, incorrect, or inconsistent: Using various Team Studio operators, it is possible to impute missing observations, remove non informative transactions, and develop features based on simple rules and operations.

Feature Engineering

The fidelity and accuracy of a model is largely dependent upon feature engineering. It is important to evaluate which features are driving the accuracy of the model. Removing features that are redundant improves model accuracy. For some models, certain transformations should be applied before model estimation. For example, variable standardization is required in lasso regression to reduce the influence of magnitude of each variable, and when hot encoding or impact encoding is needed for GBM models. While it can be difficult for any software to completely automate the feature engineering process, the domain expertise of a data scientist paired with the autoML technology in TIBCO Data Science software can significantly reduce the time for creating an optimal model.

Model Selection

For any given dataset, it can be difficult to choose the most highly performant algorithm for a use case. Manually testing all of the potential algorithms available to the data scientist or compliance officer is not only time consuming, but may introduce algorithm bias. AutoML in TIBCO Data Science software (Figure 3) generates features and automatically searches across the hyperparameters of the algorithms in the Team Studio library of algorithms to select optimal models using a codified set of data science best practices.

Hyperparameter Optimization

Hyperparameter optimization involves choosing the optimal parameters for a given algorithm. For example, for a decision tree: What is the optimal number of trees and what is their depth? What are the splitting criteria? Traditional approaches to hyperparameter optimization involve going through a large “for loop,” testing virtually every possible combination of parameters, and selecting the one that proves most accurate. This approach is computationally expensive and wasteful. The

autoML feature of Team Studio includes an intelligent approach to hyperparameter optimization that efficiently identifies the optimal set of parameters and adjusts them accordingly.

Figure 3. Auto feature engineering/AutoML workflow in TIBCO Data Science sofware.

Model Explainability

The ability to explain a model is of increasing importance as they become more sophisticated and regulations more stringent. Model explainability includes understanding model behavior in localized regions and combinations of the predictors. For example, does the model have training data support for different segments of customers and transactions; and what is the model accuracy in such local regions? Models can be assessed with ROC curves, confusion matrices, and other summaries that indicate sensitivity to recall true positives and specificity / precision to classify true negatives. Variable importance plots display the variables most valuable in making predictions. These are helpful in explaining how the model works on different segments of customers and transactions.

Model Deployment

Because fraudsters are constantly changing their ways, models need to be managed, deployed, monitored, and refined so performance can be optimized and updated.

The TIBCO Connected Intelligence platform provides efficient work-group collaboration and version control for compliance purposes. It enables a collaborative team of statisticians/data scientists, architects, business analysts, model auditors, and validation testers to rapidly combine, deploy, and maintain algorithms and metadata stored in a version-controlled centralized repository.

Processes for model rebasing, validation, deployment, approval, and retirement lifecycles can be scheduled on a regular basis or triggered based on key events such as decaying of prediction accuracy or new data providing improved predictor coverage. Information on where each model is in the lifecycle, how old the model is, who developed the model, and who is using it for what application can also be logged and securely stored.

TIBCO Spotfire

TIBCO Spotfire data visualization and analytics software delivers a complete set of analytics to empower casual business users, analysts, and data scientists to identify and present critical insights for faster and better decision-making, without requiring IT intervention. It includes AI-powered insight discovery and natural language search, real-time streaming analytics, location and predictive analytics, and embedded data science, making it a powerful tool for fraud detection, risk avoidance, and opportunity identification and response.

TIBCO Data Science – Team Studio

TIBCO Data Science – Team Studio is a modern, web-native data science collaboration platform that offers drag-and-drop and Python notebook interfaces for data access, feature engineering, machine learning, and workspace/model management. Team Studio is optimized for scalable big data distributed calculations within the data source – without moving data anywhere.

Team Studio also provides data scientists and the broader analytics community — including data engineers, business professionals, analysts, IT staff, developers, and executives — the ability to solve a myriad of business challenges.

This information also supports compliance and helps business leaders understand how model origin, use, and performance is evolving.

Each part of the TIBCO platform can work independently or as a whole. Models produced in TIBCO Spotfire visual analytics and TIBCO Data Science software can be managed and deployed to the TIBCO Streaming runtime without any system downtime.

Supervised Learning Algorithms

Supervised learning model training requires data with confirmed fraudulent and non-fraudulent cases. Decision trees, random forests, gradient boosting machines, neural networks, support vector machines, and logistic regression are examples of supervised learning algorithms available in TIBCO Data Science software. Models are fit to the data so as to maximize recall of true positives (sensitivity) while maintaining specificity in classifying true negatives and minimizing false positives.

TIBCO Data Science software 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 workflow to include new features from new data. Variable Importance analysis can then be applied to these new features to understand their contribution to explain probable fraudulent transactions. The user is also free to include a number of algorithms or just a preferred one.

Results in Figure 4, including any quality tests performed, are published back to the Spotfire system. In the example the user only needs to know that the best model is arguably the one with the highest area under the curve (ACU) as shown in the table.

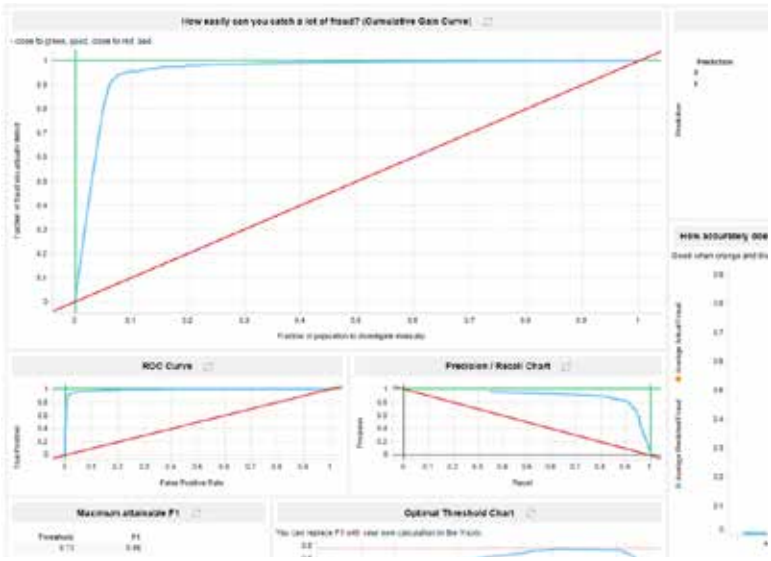

Figure 4. Example of results from quality tests.

TIBCO Data Science – TERR

The TIBCO Enterprise Runtime for R (TERR) system provides a fast and scalable advanced analytics environment based on the popular R language. In addition to broad R package compatibility, TERR delivers superior performance and memory management for running scripts and large datasets.

TERR is:

- Embedded in the Spotfire solution, with no additional installation needed, and in third-party products and custom applications
- Embedded in the TIBCO Streaming analytics product for real-time predictive model scoring
- Deployable in big data environments (such as Apache Hadoop and Spark) and in grids (via TIBCO GridServer software) for highly scalable advanced analysis on big data
- Free to individual R users from tap.tibco.com

TIBCO Data Science workflows and algorithms can be called from Spotfire software through point-click interfaces via the Spotfire data function construct. Data functions enable citizen data scientists and business analysts to simply map input data and output data to TIBCO Data Science software. The analysis runs asynchronously in the background, updating the Spotfire analysis and notifying the Spotfire user. Data functions are easily created and shared among the Spotfire and TIBCO Data Science user community. Business analysts can use them without needing to write code, while leveraging the smarts and protections afforded by data scientists. As such, business users are empowered to make better decisions and create fraud models without being exposed to underlying complexity.

Unsupervised Learning Algorithms

Using only supervised models is not enough. There will be cases where you don't have any historic knowledge about fraudulent transactions, for example, from a new channel or new fraud types. Even when you do have historic knowledge, it can never be 100% certain whether all previous fraud cases were correctly identified. Worse still, fraudsters are creative, and if one strategy did not work, they will try to find new ones that will leave a different fraud signature and thus look completely different to the model.

Unsupervised models do not need prior knowledge regarding which transactions are fraudulent and which are not. Without a target variable directing the analysis, unsupervised algorithms aim for separate anomalies (or oddities) from the normal bulk of the transaction data. In addition to standard unsupervised methods such as K means clustering and self-organizing maps, many supervised models have an unsupervised companion method, for example, deep learning autoencoders or single class support vector machines. When applied to financial crime data, these methods allow for profiling "normal" operations and spotting anomalies.

With high dimensional multivariate financial transactions, an additional trick is to use Principal Component Analysis (PCA) to reduce dimensionality. The combination of PCA and K means works well in practice. Figure 5 shows an example of such output. 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 information is not the result of any human assumption, and is derived directly from the pattern of the transactions. Transactions that are unusual beyond an agreed threshold are worthy of investigation.

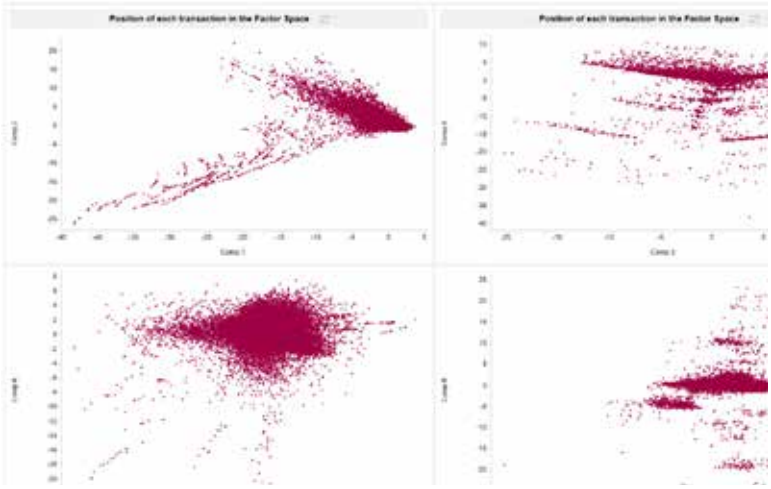

Figure 5. PCA and cluster analysis for anomaly detection in financial transactions.

Autoencoders are unsupervised neural networks that can be used to replicate high-dimension transactions with a lower-dimension approximation by restricting the number of hidden layers and nodes. This is like a non-linear PCA analysis. For example, if a transaction has 200 features, we may represent these with a single layer neural network with say 20 nodes. The predictions from this reduced dimensional space that approximate the observations well are not anomalies. Conversely observations that aren't well-predicted by the lower dimension autoencoder are potential anomalies. We use the reconstruction error from the autoencoder to reflect the transaction oddity, or degree of anomaly. The higher the reconstruction error, the higher the possibility that the transaction is an anomaly.

TIBCO Data Science software includes the autoencoder analysis via TensorFlow in a Python notebook, as a Python data function, and via R/TERR with the H2O autoencoder implementation. Templates are available for implementing this analysis from Spotfire and TIBCO Data Science software. TIBCO has also made a base version of the TensorFlow autoencoder available on the Amazon AWS Marketplace. TIBCO Data Science–Team Studio also offers an integration with SageMaker in which the Python notebook autoencoder can be run as a parallelized compute for big data. This is a seamless operation from TIBCO Data Science software, where the AWS cluster is automatically configured from the TIBCO Data Science environment.

Closed Loop Continuous Learning

An identified anomaly does not necessarily imply a financial crime has occurred. Rather, the identified anomaly requires further analysis and verification. Typically, identified anomalies are written to a database and case-managed to resolution. The transaction can then be labeled as fraud or not, which provides an additional, verified transaction for the supervised algorithm to use. This back and forth between the supervised and unsupervised model is part of the TIBCO Data Science approach. We call this closed-loop continuous learning.

Good Features Make Good Models

Any predictive model is only as good as the features included. One challenge companies often face is identifying which characteristics to focus on to identify fraudulent events. Good fraud features are those that are highly correlated with unusual behavior. Features can be continuously enhanced by applying enhancements from third party sources, for example, IP address/MAC number or transformations to an existing dataset e.g. binning, ratio operations, power transformations.

Some relevant features for identifying fraud in card payments include:

- Value of transaction compared to the value of historical transactions with vendors having that merchant category code
- Geographical location of the vendor
- Internal risk scoring of that vendor

If you are considering the platform for a use case like AML, some relevant features may include:

- 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.
- Multiple transactions over a short period of time that exceed a total threshold.

Many of these features are gathered, calculated, and monitored with systems such as Actimize, but are often calculated as individual measures and not mathematically optimized. ML models built within the TIBCO Data Science platform can combine all features optimally.

In different scenarios, different sets of features may be used for finding fraud; For example, medical claims for an insurance policy that covers emergencies may exhibit fraud when providers declare routine procedures as emergencies. The incentive for this type of fraudulent behavior is specific to the terms of the insurance policy. 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 prescribed medicine
- Time between emergencies per patient and per family
- Medical assessment of the symptoms

When data features have been crystallized, Spotfire software can collect these straight from the relevant datasource, identify how those transactions behave, and give a visual representation. For example, on the left-hand chart of Figure 6, it's easy to spot odd behavior.

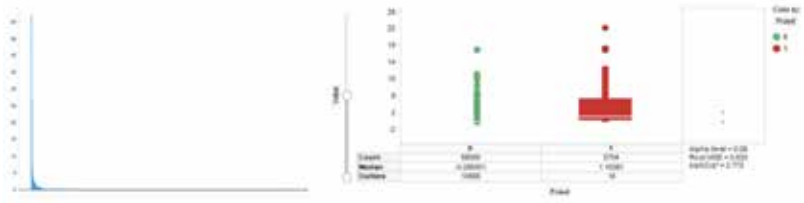

Figure 6. Visual analysis of related features with no past knowledge about which were fraudulent. The left visualization is deceivingly powerful, plotting the number of credit card transactions from all users in the last 24 hours. It shows that the majority of people have very stable behavior (around 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.

Unstructured data may need to be preprocessed, for example medical assessment documents might need to be digitized. Required information like illness_name, first_notified, first_occurence, first_treatment, and prescribed_medication could be extracted from the medical documentation using text analytics that can then be used to enrich the features available. The fraud prevention analyst can use various natural language

processing (NLP) techniques to highlight keywords from the unstructured data and the underlying distribution of claims frequency. An investigation of a few cases may 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, it can inform the search for fraud revealing features. On the right-hand chart of Figure 6, a zoomed-in box plot shows that transactions with higher value have been more likely 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, or texts/ calls between certain key suspects.

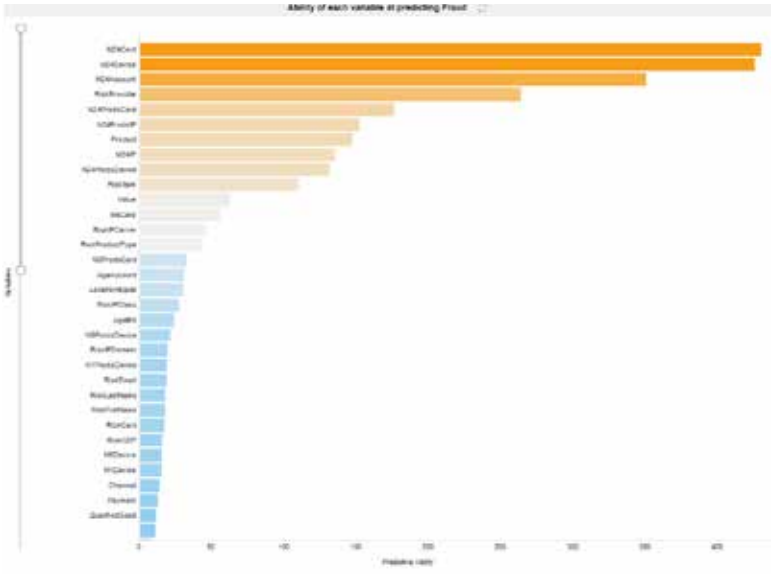

Figure 7. Visualizing the contribution of different predictors used in a model.

Not every feature needs to be visualized. In big datasets 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. These features can then be visualized individually. Figure 7 shows which features have a higher contribution to the model (the longer the bar, the more important the feature). Although Figure 7 showed that "Value" (eleventh in rank) is important, other features are more powerful at distinguishing fraudulent transactions.

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

Streaming Analytics

TIBCO's streaming analytics environment enables the capture, aggregation, and analysis of real-time and historical data of any variety, volume, and velocity. This builds contextual awareness and allows preemptive actions and interventions. TIBCO Streaming technologies enable:

- Understanding of historical patterns and dynamic event sequences
- Monitoring all event streams to anticipate filtering for certain sources or qualities, correlating in real time, and detecting meaningful patterns
- Acting 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
- Rapidly capturing, analyzing, and acting on any complex combination of events in real time

TIBCO BusinessEvents

TIBCO Streaming

TIBCO Spotfire X

1. **Validation of the model's quality.** Cross-validation or hold-out validation can be used to assess how the results of a statistical analysis generalize to an independent dataset.
2. **Correlation is not causation.** It is necessary to ask questions that lead to a better understanding of the reality being predicted. Interaction between features can shed additional light on this.
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 a large predictive ability. This knowledge, when shared with the business, can 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 applied to current data, in batch or real time.

So You Have Your Models, Now What?

After business users have tested the ability of a TIBCO Data Science model to efficiently spot potentially fraudulent or risky transactions, they can use Spotfire analytics as an interface to run various what-if scenarios to inform how the model may be most efficiently deployed. Models allow combining all incoming transactions into two states: anomalous and normal. Spotfire what-if analysis enables users to set adequate thresholds for distinguishing these states, and Spotfire templates allow users to set thresholds that balance the expected number of alerts with the capacity of available investigators.

How TIBCO Deploys Models in Real Time

Because a model is a summary of historic data, it can be as light as a rules table, or as complex as an ML model, and live beyond the data. Once the user is satisfied with the quality of the model and the respective thresholds, Spotfire analytics can push that model to TIBCO Streaming analytics to monitor new transactions as they occur.

Business Process Management

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

- A model-driven environment to speed and simplify process design, shielding implementation complexity with a fast, collaborative, and iterative approach
- The ability to 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
- The ability to 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
- A native integration foundation for true business digitization, allowing your data, people, processes, systems, and things to be easily brought together to support all your business initiatives

TIBCO BPM Enterprise

TIBCO Cloud Live Apps for case management

Case Management & Reducing Investigation Time

Alerts often take too long to investigate. TIBCO Streaming analytics helps address this by calculating features and scoring transactions for their probability of fraud and risk in real time. The TIBCO Financial Fraud Accelerator and the TIBCO Risk Management Accelerator are free fast-start templates available on the TIBCO Community that implement this flow, applying models to the transaction streams in real time, separating those that exceed the respective thresholds, and for each of these alerts, opening a new case in the case management system. The accelerators also build a rich context for each alert from any number of data sources using tools like TIBCO EBX data management and TIBCO Data Virtualization software that connect contextual data to the new case and model data in the flow. Investigators may then receive notifications of potential fraudulent transactions as cases are queued to the investigation team. TIBCO Streaming analytics, can also process and visualize flow statistics in real time.

Figures 8 through 11 provide examples of automatically generated cases, where all the relevant data has been pulled into consolidated views to enable investigators to make fast decisions. This includes the ID of the transaction, a link to the respective investigative model, the scores for probability of it being like past fraud, its degree of oddity, the respective thresholds, and the model versions that generated the scores. The case also includes the relevant context regarding specific transactions—including data gathered from other related data sources. The investigator can swiftly review the attached information, source more information as required, and make judgements on transaction validity.

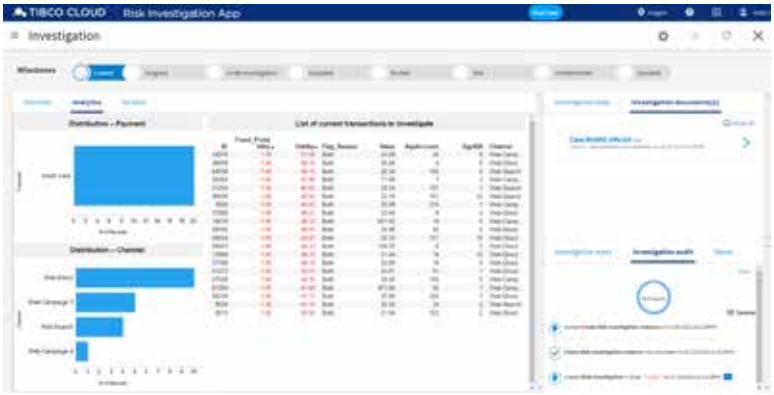

Figure 8. Investigative template that can be sent to an investigator in real time containing the entire context of a risky transaction.

Financial Fraud / Risk Management Accelerators and Cloud Starters

TIBCO accelerators are free, fast-start templates and design pattern examples.

The Fraud and Risk Management accelerators use Spotfire software to guide data professionals through developing supervised and unsupervised models that detect probability of fraud and anomalous transactions from a known dataset. Models may be developed using TIBCO Data Science software and hot deployed to TIBCO Streaming analytics for evaluation at runtime. When a transaction is scored, it will either pass or be flagged as probable fraud or anomaly. When this occurs, the streaming analytics platform raises an alert and creates a case in the TIBCO BPM Enterprise system or in a TIBCO Cloud Live Apps low-code integration platform app to facilitate investigation of potential fraud.

The accelerator has embedded model quality tracking built into the process management, from which users can be warned of the potential need to revise the model. TIBCO provides a lot of flexibility regarding how to incorporate model quality tracking into a financial crime fighting solution. For example, models can be retrained if quality of prediction is perceived to be diminishing. Using Team Studio, the new model can be tested against the previous one and updated if deemed better based on a number of model quality measures. The business can be notified automatically for model approval, and all comments can be stored securely for audit and compliance purposes.

Figure 9. Example of detection report providing a drill down to relevant context.

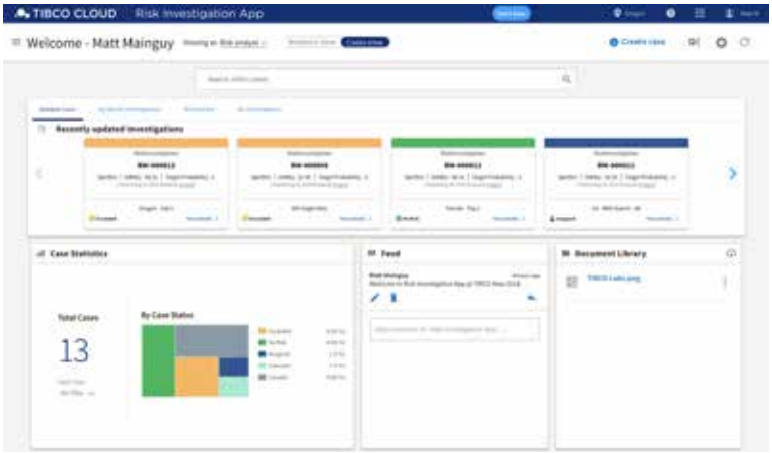

Figure 10. Example detection report.

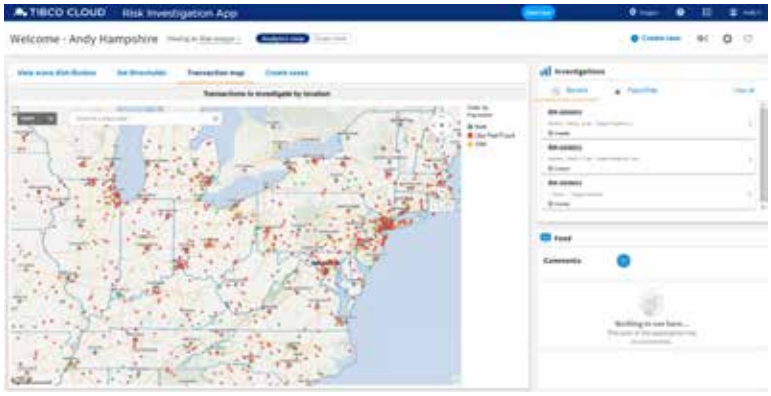

Figure 11. Example detection report.

Self-learning Abilities

One important advantage of TIBCO's approach 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 used in the next version 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.

Deep Feature Synthesis

Current systems and techniques can exhibit a large number of false positives that carry misclassification costs. False positive reduction can be performed by building better-targeted features that take into account profile, terminal, geolocation, and other information. These features are hand-built by predictive modelers and trained data scientists who are constrained by their tools. Deep Feature Synthesis (DFS) is a procedure to automatically generate derived features from processing the interaction and relationships between collections of datasets, figuring out joins, aggregations, roll-ups, time series windows, etc. These derived features comprise the "context" of a transaction that would otherwise need to be built by a specialist.

Injecting DFS into the modeling workflow enables the creation of an enhanced fraud detection solution that reduces false positives by taking into account localized context for all transactions. For example, how does the transaction size compare to that of the typical transaction in the last few days? What's the mean duration between transactions? What's the mean location change of transactions in general for this user over the last six hours?

Transparency

One of the most important advantages of using TIBCO's platform is that it is not an opaque black box, and in fact, a lot of information can be made available to business users depending on their skill sets. The whole platform can be modularized, meaning you can pick and choose the components that best suit your current IT landscape.

Better models often come from better features. Business users can try new features from an easy-to-use dashboard and visually test their relevance (as we do in Figure 6). These features feed straight into the supervised model, which produces a clear measurement of their worth in detecting past fraud (as in Figure 7).

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 software. In TIBCO's Financial Fraud Accelerator, business users can apply these from easy to use dashboards and keep focused on the business at hand. However, because the algorithms are entirely open, your own data scientists can not only customize them, but replace them at will with their favorite approaches.

Finally, the investigation case management app provides an audit trail of exactly what each investigator does with each case and allows them to log the reasons for the decisions made alongside all relevant information. This capability enables regulatory compliance in use cases where it's needed as well as a better understanding of the investigation process.

Conclusion

Understanding risk and opportunity in real time is increasingly critical, and most organizations already hold the data to make it possible; But all too often, it is difficult to reach or wrangle quickly enough. The overall 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 transactional information to be able to react precisely. TIBCO brings many analytic solutions together 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 platform for multiple diverse use cases, including credit card fraud, trade surveillance, anti-money laundering (AML), and medical fraud. It provides:

- Monitoring of all transactions in one auditable, repeatable, self-learning process
- Increasing customer satisfaction because it impacts only those exposed to real risk
- Increased investigative team productivity by only calling attention to risky transactions that can be investigated with better contextual data for optimal decision-making
- Advanced mathematics at your fingertips, transparently
- Real domain experts the ability to apply their knowledge of specific business operations, without the need for a degree in advanced math or computer science
- 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. You can expedite the investigation process so staff across your organization can evaluate potentially risky transactions and make the right decisions quickly. Advanced analytics has been used for years; now is the time to move to the next level.

Contact us to learn more about TIBCO Data Science software and TIBCO Connected Intelligence.

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

TIBCO Software Inc. unlocks the potential of real-time data for making faster, smarter decisions. Our Connected Intelligence platform seamlessly connects any application or data source; intelligently unifies data for greater access, trust, and control; and confidently predicts outcomes in real time and at scale. Learn how solutions to our customers' most critical business challenges are made possible by TIBCO at www.tibco.com.

18Dec2020