TIBCO

The Ultimate Guide to Anomaly Detection

Key industry solutions, methods, and autoencoder machine

Why Is Anomaly Detection Important?

Large volumes of data from business operations are generated daily. If used effectively, this data can help you make better decisions—and one way to use it for competitive advantage is through automated anomaly detection.

Typically, most of your data stream simply confirms that things are operating normally and so provides no new actionable information. However, detection that something has changed or is behaving in an abnormal manner may be an important insight leading to action—action that can stop a minor issue from becoming a widespread, time-consuming problem. This whitepaper covers the basics of anomaly detection, some use cases from our practice, and some key techniques.

What Are Anomalies?

Before we dive in, let's take a step back: What exactly are anomalies? An anomaly is an unexpected change or deviation from the expected pattern in a dataset. Therefore, anomaly detection is a way of detecting abnormal behavior. Note that anomalies aren't necessarily good or bad, but it's important to be alerted to any break in pattern to assess whether action needs to be taken.

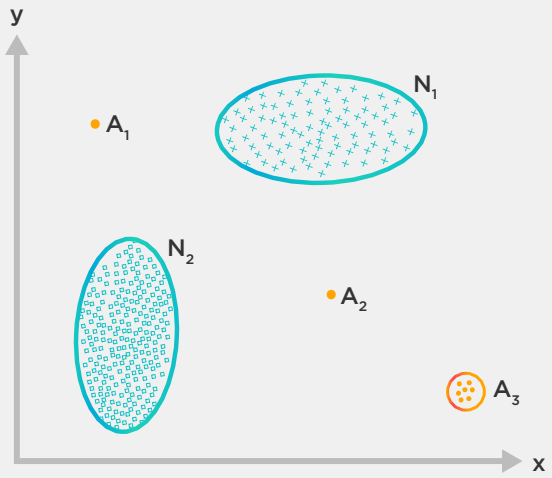

Figure 1. Anomalies (A1, A2, A3) in a two-dimensional dataset

What Is the Difference between Anomalies and Outliers?

There is much debate on this topic, and many people use the terms interchangeably. Synonyms for outliers in different application domains may include "discordant observations, exceptions, aberrations, surprises, peculiarities or contaminants."[1]

Data is generated by processes, and outliers are points with a low probability of occurrence within a given dataset generated by the baseline process. They are observation points distant from other observations within the normal population; however, they don't necessarily represent abnormal behavior or behavior that occurred because the process has changed. Outliers are generated by the baseline process but occur with lower probability, whereas anomalies are generated by different processes.

Applications of Anomaly Detection

Fighting Financial Crime and Insurance Fraud

Financial Crime

In the financial world, transactions worth trillions of dollars execute every minute. Identifying suspicious ones in real time can prevent large losses.

1 Chandola, Varun, Arindam Banerjee, and Vipin Kumar. Anomaly Detection: A Survey, ACM Computing Surveys. https://dl.acm.org/doi/10.1145/1541880.1541882

Over the last few years, leading financial companies have increasingly adopted big data analytics to identify abnormal transactions, clients, suppliers, or bad actors. Machine learning models are used extensively to detect anomalies in streaming data. By one estimate, Visa and Mastercard process more than 5,000 transactions per second.[2]

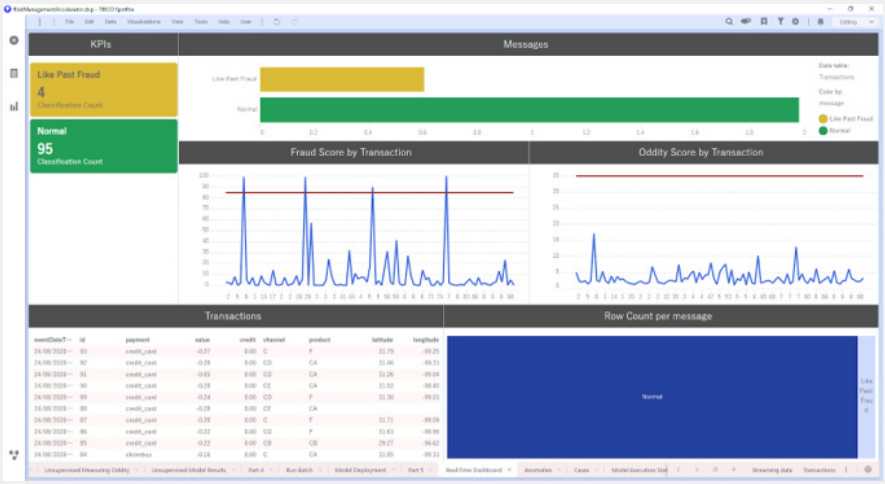

Figure 2. TIBCO Risk Management Accelerator, Real-time Fraud Detection Dashboard. Learn more and download the Accelerator.

Healthcare Insurance Fraud

Healthcare insurance fraud is a common occurrence. It is vital for insurance companies to identify fraudulent claims and ensure that no payout is made for them. The Economist recently published an article that estimated $272 Billion as the cost of US healthcare fraud and expenses involved in fighting it. Around 10% of annual Medicare & Medicaid spending is consumed by fraud. Companies are investing in big data analytics to build supervised, unsupervised, and semi-supervised models to predict insurance fraud.

Solutions for Financial Crime and Insurance Fraud

The TIBCO Risk Management Accelerator identifies potentially risky activities in a high-frequency event stream using machine learning. Supervised and/or unsupervised models can be built and hot deployed to the streaming event processing platform, where events are scored in real time. Alerts are then raised when potentially risky behavior is detected.

---

2 Vlastelica, Ryan. Why bitcoin won't displace Visa or Mastercard soon. Marketwatch, December 18, 2017. https://www.marketwatch.com/story/why-bitcoin-wont-displace-visaor-mastercard-soon-2017-12-15

The TIBCO Cloud Risk Investigation App identifies anomalous and suspicious transactions. It includes a case management framework that strengthens collaboration across the enterprise. With a centralized view of the investigation process, this highly customizable application provides an analytics-based framework with clear lines of accountability.

Cyberthreat Detection

Networked computers today are under constant threat of ransomware and other forms of cyberattack. System threats can be detected through analysis of computer log data using unsupervised learning models such as LSTM autoencoders to identify anomalies in the sequence of log events.

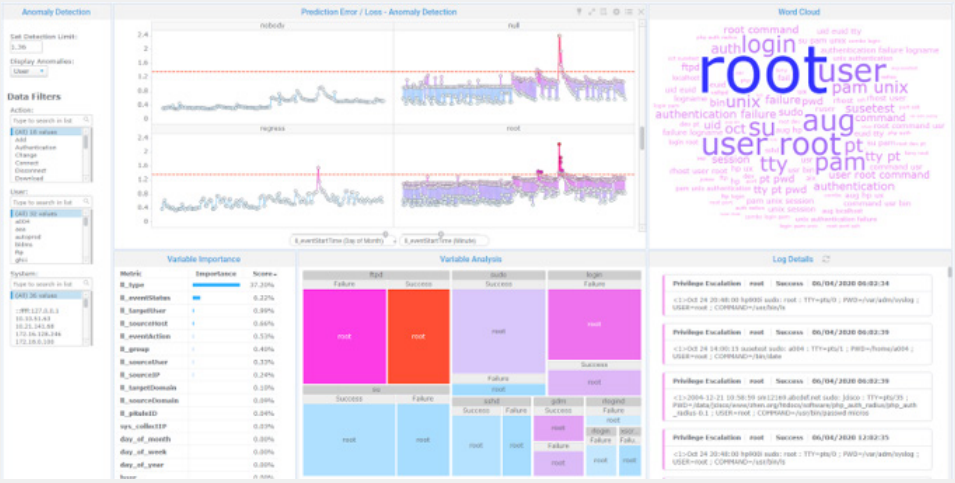

Figure 3. TIBCO Cyberthreat Detection Application dashboard

Maintaining Machine Health

The manufacturing, energy, telco, and transportation sectors all depend on machines. Detection of anomalies in machines is a key to maximizing uptime and performance, maintaining control of processes, and ensuring the quality of what they produce.

Preventing Machine Breakdowns with Connected Sensor Data

Many types of equipment, vehicles, and machines are now instrumented with sensors. Monitoring these sensor outputs can be crucial to detecting and preventing breakdowns and disruptions. Anomaly detection algorithms like autoencoders can be used to detect anomalous data patterns that may predict impending problems. When sensor time series traces exhibit repeating patterns, special techniques such Mueen's Algorithm for Similarity Search (MASS),[3] or the one used in our Sensor Anomaly Detection at the Edge[4] can be used.

Figure 4. Solution for Sensor Anomaly Detection and Root Cause Analysis Using TIBCO Analytics and Microsoft Cognitive Services.

Listening for Abnormalities in the Sounds Machines Make

Sounds can be used to assess machine health and prevent breakdowns. A good mechanic can tell whether your car is OK or not by listening to the sounds it makes. A really good one can even tell you what is wrong with it.

Abnormal sounds can be an indicator that a machine needs maintenance. The following image shows an example of an application that uses audio data from any device and learns to identify anomalous sounds made by machines.

3 Katz, David. An Introduction to Similarity Search, Matrix Profiles and MASS for Detection of Anomalies in Time Series. TIBCO Community, July 8, 2019. https://community.tibco.com/sites/default/files/matrix_profiles_and_mass_v2_0.pdf

4 Geiger, V. Anomaly Detection and Root Cause Analysis Using TIBCO Analytics and Microsoft Cognitive Services. TIBCO Community, September 17, 2021. https://community.tibco.com/wiki/anomaly-detection-and-root-cause-analysis-using-tibco-analytics-and-microsoft-cognitive

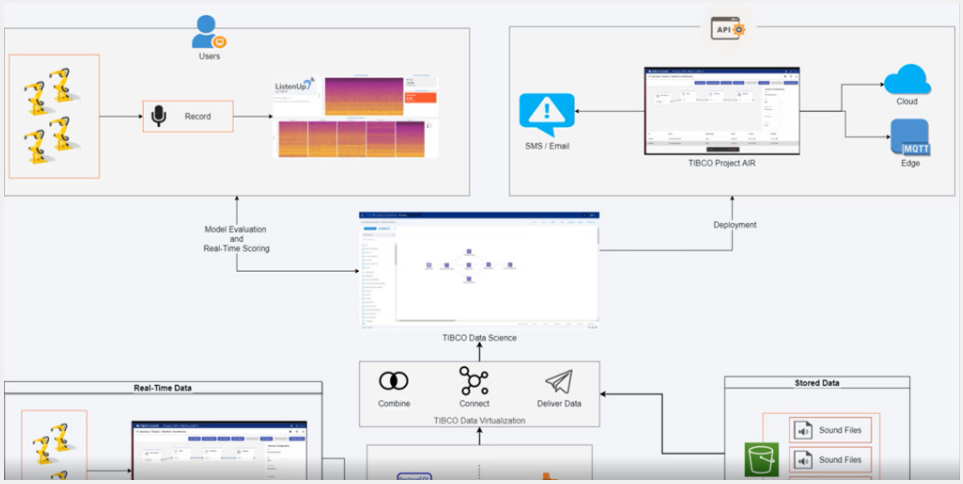

Figure 5. TIBCO Application for audio-based monitoring and anomaly detection for machines. View a demo of the application

Manufacturing: Maintaining Control of Processes and Ensuring Quality Product

In Manufacturing, manual processes used to find anomalies are laborious and reactive, and building machine-learning models for each part of the system is difficult. Therefore, some companies continuously monitor data on process results and manufactured components. As the models score new data, they can find any defects and anomalies quickly, and users can take preventative action.

Identifying Abnormal Product

Many manufactured products undergo some form of testing to determine suitability for use. Univariate and linear multivariate statistical process control methods can be used to detect anomalous product based on this data. However, with increasing component and system complexity, multivariate anomalies that also involve significant interactions and nonlinearities can be missed by these more traditional methods. These anomalies can be implicated in reliability and system failures. AI-based algorithms, such as autoencoders, can often be used to identify these complex anomalies. Once the anomalies are detected, their “fingerprints” can be generated so they can be classified and clustered, enabling investigation of the causes of the clusters. As new data streams in, it is scored in real time to identify new anomalies, assign them to clusters, and respond to mitigate potential problems.

Figure 6. Autoencoder application for detection of anomalous products or processes. Download this application and learn how to build good models with it.

Defects and Abnormalities in Images

Large amounts of raw image data are generated today by connected devices. People are very good at rapidly identifying abnormalities in images, but it is expensive and time-consuming for them to extract the critical information from large numbers of images; they often remain unprocessed, but AI algorithms are increasingly used to automate the process. Applications often involve some combination of unsupervised learning (where similar images are clustered), human verification that images contain abnormalities, and supervised learning to train models that automate the identification of abnormalities of interest. Examples include identification of cancer cells, manufacturing defects, and semiconductor wafermap spatial test and fail patterns.

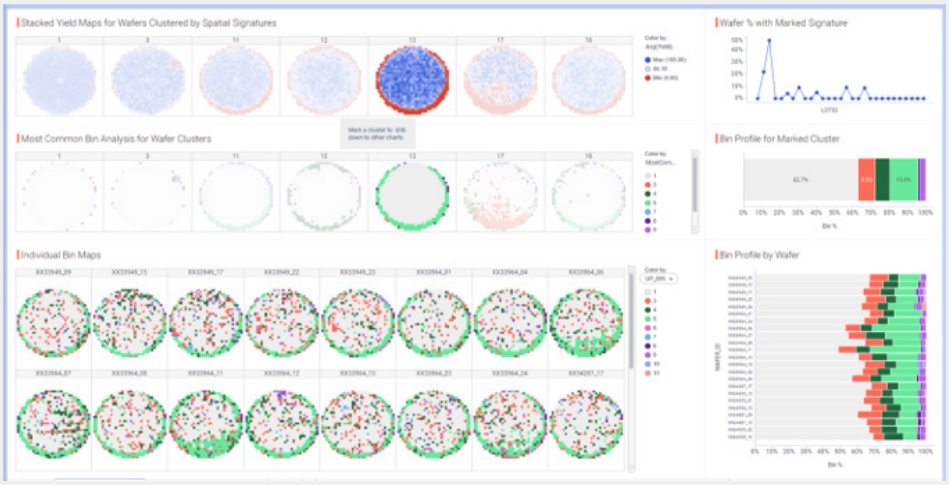

Figure 7. Application for identifying unique failure patterns seen on semiconductor wafermaps. Get more information.

Statistical Process Control

Control charts are widely used in manufacturing, energy, telco, technology, and many other sectors. They are an established form of anomaly detection used to monitor key metrics, detect deviations from the baseline, and generate automated alerts. TIBCO supports many types of Shewhart (univariate) and multivariate charts. Get more details about TIBCO SPC solutions.

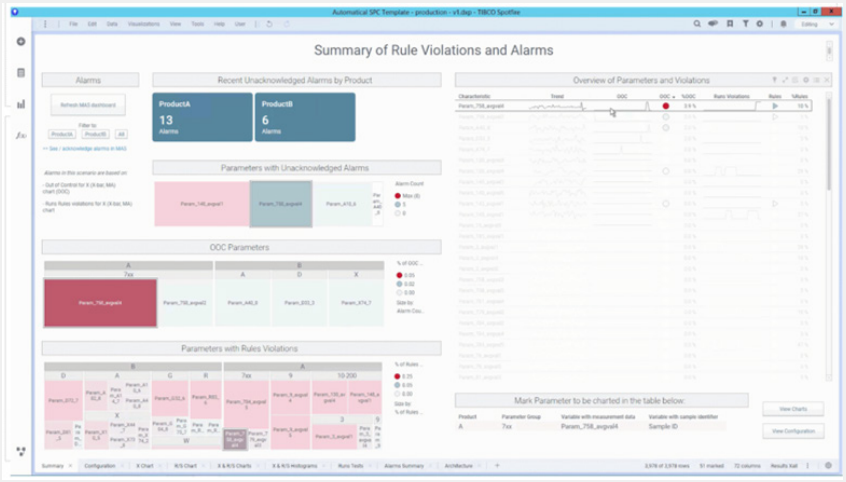

Figure 8. TIBCO Statistical Process Control solution. Get more information about this solution.

Other Examples

Beyond the use cases already described, there are many other examples of how anomaly detection can be applied across a wide variety of industries:

- Military surveillance: Image recognition
- Safety systems: Fault detection
- Hack protection: Anomalous network traffic detection
- Weather variables: Heat wave or cold snap detection
- MRI imaging: Alzheimer's or malignant tumor detection
- Spacecraft sensors: Faulty component detection
- Bank Financial Stress Testing

Methods for Anomaly Detection

Companies around the world have used various techniques to detect anomalous data. While the list below is not exhaustive, three anomaly detection techniques have been popular:

1 Visual Discovery: Anomaly detection can be accomplished through visual discovery. In this process, a team of data or business analysts visually monitor dashboards containing bar charts, scatter plots, statistical process control (SPC) graphs, and other visualizations to find unexpected behavior. This technique often requires prior business knowledge in the industry of operation and creative thinking to ensure the right visualizations are used to find targets. Because humans are inherently visual, we can quickly spot patterns in data. However, the downside is that we may not be able to visually inspect all of the data available in a reasonable timeframe.

2 Supervised Learning: Supervised learning is an improvement over visual discovery. In this technique, persons with business knowledge in a particular industry label a set of data points as normal or abnormal. A data scientist then uses this labeled data to build machine learning models that will be able to predict anomalies on unlabeled new data. Supervised learning is a great technique to use when you have known patterns in data that you would like to model. However, in a case where new patterns are continuously emerging (fraud, manufacturing), you may need to implement an unsupervised learning method.

3 Unsupervised Learning: Another technique that is very effective is unsupervised learning. In this technique, unlabeled data is used to build unsupervised machine learning models. These models are then used to predict new data. Because the model is tailored to fit normal data, any points that do not fit the model when you reconstruct the data are by definition, anomalous.

Examples of Unsupervised Learning Algorithms

Autoencoders

Unsupervised neural networks, or autoencoders, are used to replicate the input dataset, but only approximately; If the input is replicated exactly, the model cannot usefully be applied to new data. The approximation is made by restricting the number and size of hidden layers in a neural network. A reconstruction error is generated upon prediction. The reconstruction error is defined as the difference between the model output and the new input data. The higher the reconstruction error, the higher the possibility that the data point is an anomaly.

Clustering

In this technique, the algorithm attempts to classify each data point into one of many pre-defined clusters by minimizing the within-cluster variance. Models such as K-means clustering, K-nearest neighbors, and others, are used for this purpose. A K-means or a K-NN model serves the purpose effectively because it assigns a separate cluster for all data points that do not look similar to normal data. In general, this technique is most useful when the data is well separated into natural clusters, a requirement that should be checked.

One-class Support Vector Machine

A support vector machine defines a hyperplane that best divides a set of labeled data into two classes. For this purpose, the distance between the two nearest data points that lie on either side of the hyperplane is maximized. For anomaly detection, a one-class support vector machine is used to classify as anomalies those data points that lie much farther away than the rest of the data points. Similar to clustering, this is most effective when the points are well separated into natural clusters.

Time Series Techniques

Anomalies can also be detected through time-series analytics by building models that capture trends, seasonality, and levels in time series data. These models are then used along with new data to find anomalies. When the new data diverges too much from the model prediction, either an anomaly or a model failure is indicated.

Recent developments include the MASS techniques (Mueen's Algorithm for Similarity Search)[5] for fast scanning of time series for unusual subsequences (discords). In addition, related methods for change detection and for comparisons of multiple time series have been developed.

Autoencoders Explained

Autoencoders use unsupervised neural networks that are both similar to and different from a traditional feed-forward neural network. They are similar in that they use the same principles (for example, backpropagation) to build a model. They are different in that they do not use a labeled dataset containing a target variable for building the model. They use a training dataset and attempt to replicate the input dataset by restricting hidden layers/nodes.

The focus of this model is to learn and identify a function or an approximation of it that would allow it to predict an output that is similar to the input. The autoencoder achieves this by placing restrictions on the number of hidden units in the data. For example, if we have seven columns in a dataset (L1 in Figure 9) and only three hidden units (L2), the neural network is forced to learn a more restricted representation of the input. By limiting the hidden units, we can force the model to learn a pattern in the data, if one exists.

---

5 Ibid. Katz.

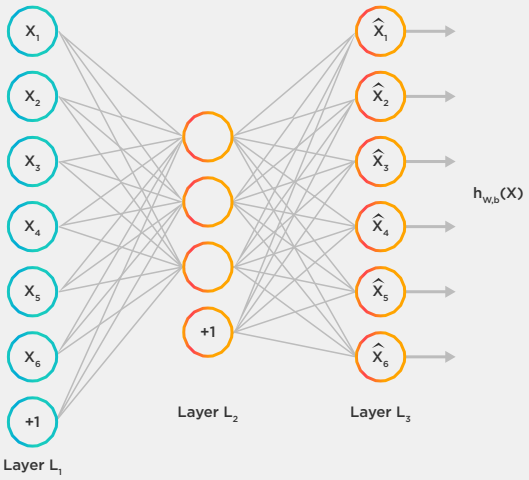

Figure 9. Example schematic of a single-layer sparse autoencoder.

Under-complete autoencoders are normally used for anomaly detection, for example in manufacturing or fraud detection. The autoencoder is built with constraints so that the input cannot be exactly reproduced, and the reconstruction error can be used for detecting anomalies.

Learn More about Anomaly Detection Solutions

Both anomaly detection and autoencoder machine learning models present rich opportunities for companies able to successfully implement them. Hopefully, this paper gives you a better understanding of the value such tactics provide and the possible applications across different industries.

Learn more about our anomaly detection solutions (applications, accelerators and data functions) and download them.

Learn more about how TIBCO supports anomaly detection at https://www.tibco.com/solutions/anomaly-detection

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