IMPLEMENT EFFECTIVE MANUFACTURING PROCESS ANALYTICS

INTRODUCTION

Low cost sensors and new wireless connectivity tools enable manufacturers to employ digital analytics more effectively than ever before. With the right tools, they can gather, cleanse, process, and visualize massive amounts of data from disparate sources that cover all phases of the product life cycle:

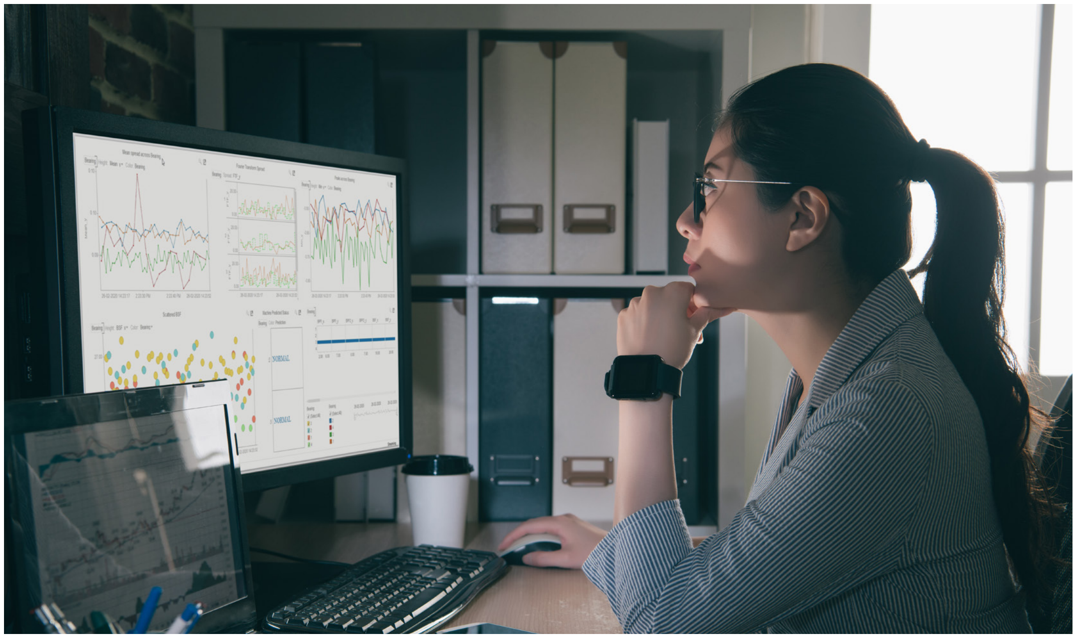
- Product and process design
- Assembly
- Material planning
- Quality control
- Scheduling
- Maintenance
- Fault detection
- Post-sale warranty claims

By extracting real value from their data, manufacturers can make accurate predictions about component life, replacement requirements, energy efficiency, utilization, and other factors that have direct impacts on production capacity, throughput, quality, sales, customer acceptance, and overall efficiency.

The manufacturing sector faces three key analytics challenges:

- Variety: Data is often trapped in organizational silos that make it hard to share between departments. It may be stored in incompatible formats and in a variety of databases, or even in PDFs or plain text. Without the right tools, sharing this data in useful ways is difficult and expensive, or even impossible in practical terms.
- Volume: The amount of data coming in from suppliers, distributors, customers, and other third parties as well as sensor networks (inside and outside the factory) is increasing constantly and can easily become overwhelming. The ability to handle all this data properly enables engineers and managers to extract useful, actionable information from it by focusing on outliers, spotting trends, and clusters.
- Velocity: Manufacturing supply chains change rapidly in structure and flow, and critical data often streams in on real-time message buses. A software infrastructure built for real-time operational environments is an absolute requirement in the current business environment.

Addressing these challenges successfully offers big payoffs in every dimension of the manufacturing business.

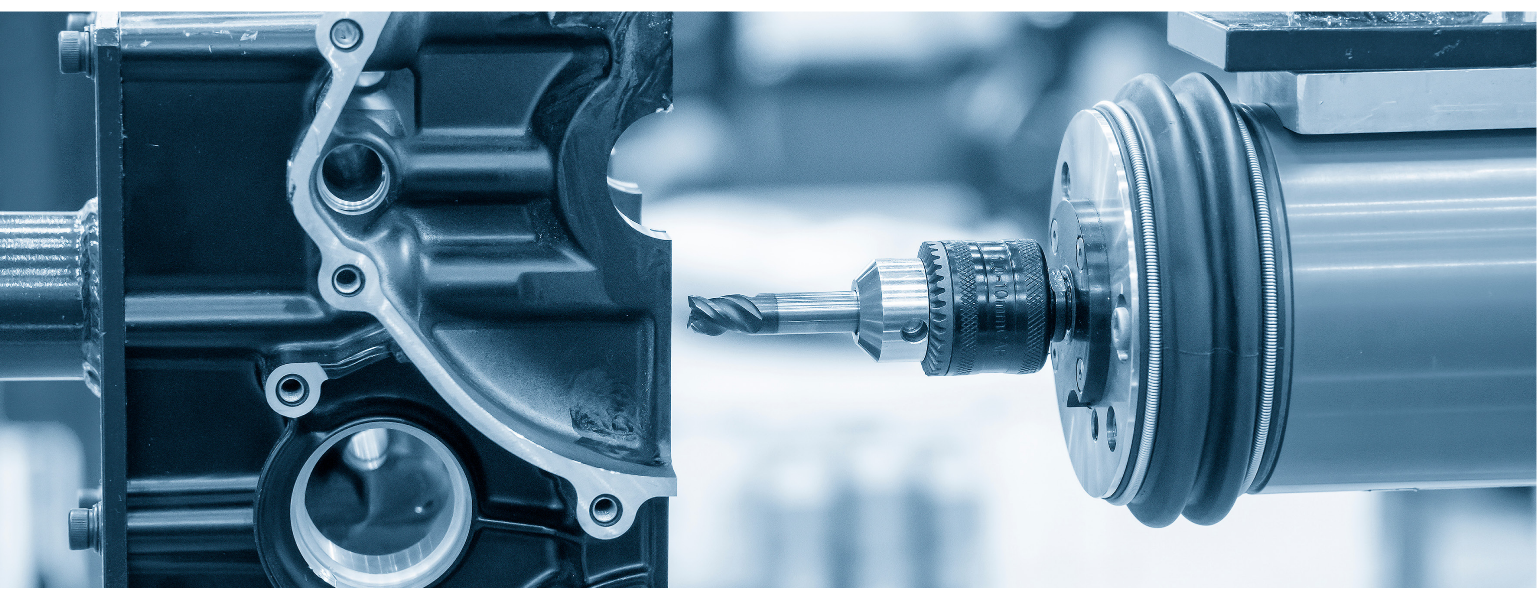


Altair has found that there are four key elements that must be in place for a firm to achieve the greatest possible gains from investments in analytics technology:

- Data-savvy engineering personnel
- Access to quality data
- Best-of-breed software
- Business processes and incentives that support data-driven decision making

Improve Manufacturing Productivity with Data Analytics

This guide explains some of the major challenges involved in applying data analytics to manufacturing processes and the benefits of developing optimized approaches to addressing those challenges.



DEVELOPING AN ANALYTICS STRATEGY

Manufacturers can increase productivity, reduce costs, improve quality, and operate more competitively by employing the right analytics tools and implementing them in smart ways.

Developing a successful strategy requires a thorough understanding of the capabilities and limitations of the chosen technologies, as well as the capabilities of the people within the organization.

Altair has found that there are four key elements that must be in place for a firm to achieve the greatest possible gains from investments in analytics technology:

- Data-savvy engineering personnel
- Access to quality data
- Best-of-breed software
- Business processes and incentives that support data-driven decision making

Your analytics strategy should begin with a clear-eyed assessment of the unique challenges confronting your firm, as well as the challenges that are likely to be common to all manufacturing operations. Next, investigate the data resources that may already be available (as opposed to accessible) for helping address those challenges. The stakeholders in the process must then come to agreement on the major outcomes the team should work towards, with defined milestones and deliverables. Finally, determine how the relevant analytics will be delivered to decision-makers and how the team will measure the success or failure of each initiative within the strategy.

The implementation process involves five distinct phases:

1. Business understanding
2. Data acquisition and understanding
3. Modeling
4. Deployment
5. User acceptance

This general approach is an adaptable, flexible way to think about a major analytics project within most manufacturing organizations.

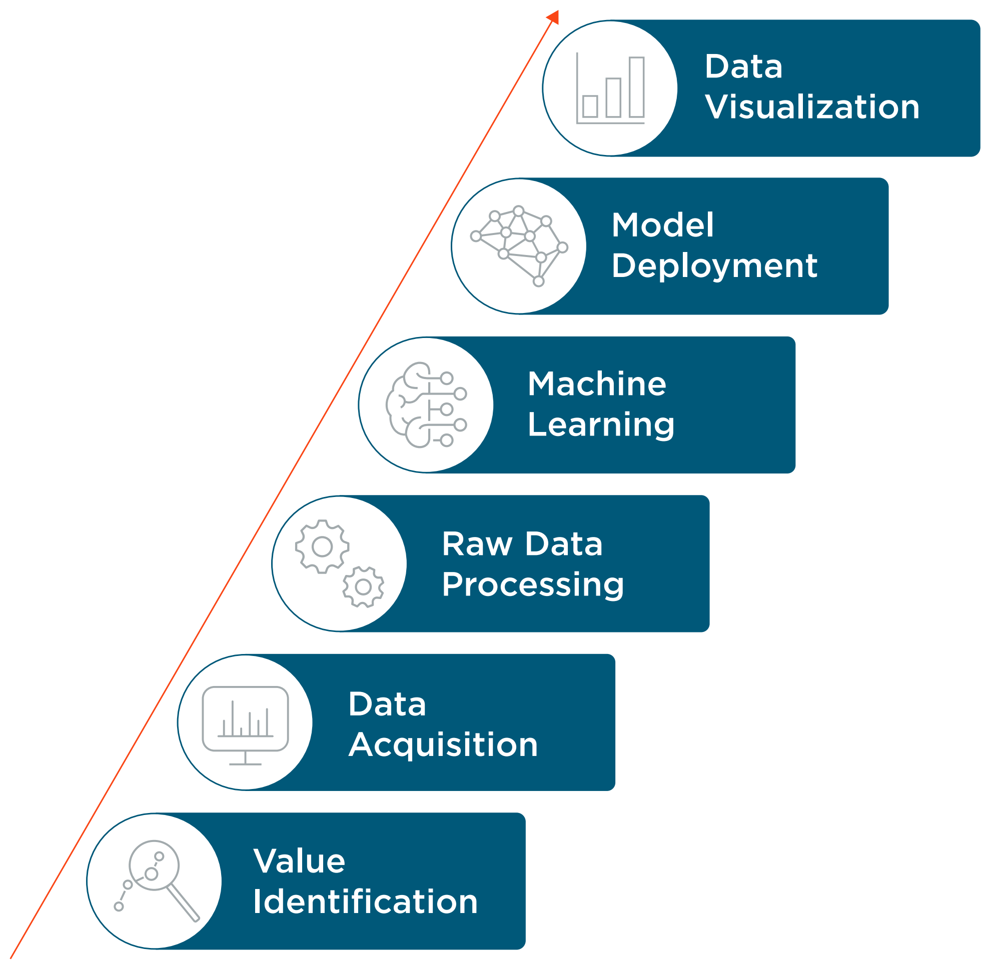
Insight doesn't come from what people can see on the surface of a data set. It comes from the hundreds or thousands of dimensions hidden in complex data.

The ability to analyze sensor data streaming in from production systems and correlate it with maintenance logs and other sources can support significant improvements in OEE and reduce maintenance costs.

THE DATA ANALYTICS WORKFLOW

Every manufacturing operation will have a different specific workflow, but this is an overview of the components that will make up a data analytics workflow for manufacturing use cases. Every step generates value.

- Identify value: Assess the impact of the planned measurements.
- Acquire the data: Often difficult since the useful data may reside in different data silos.
- Process: Turn the raw data into a useable, clean, consistent data set.
- Machine learning: Build and test models using the clean data generated previously.
- Deploy machine learning models: Put the models to use within the operation.
- Monitor: Process and visualize sensor and production data in real-time to look for anomalies, trends, outliers, relationships to help ensure smooth operations.



ABOUT DATA PREPARATION

As noted above, data preparation is often the most difficult process to implement successfully. It is critical to ensure that your data is clean, consistent, and accurate before using your data in a machine learning application or attempting to visualize it.

Sources for manufacturing data include a variety of possible sources, all of which may store data in different formats and nearly always generated and managed by completely separate systems. These include, ERP (Enterprise Resource Planning), MES (Manufacturing Execution System), and SCADA (Supervisory Control and Data Acquisition) systems, as well as data generated by PLCs (Programmable Logic Controllers) incorporated into production equipment and sensors.

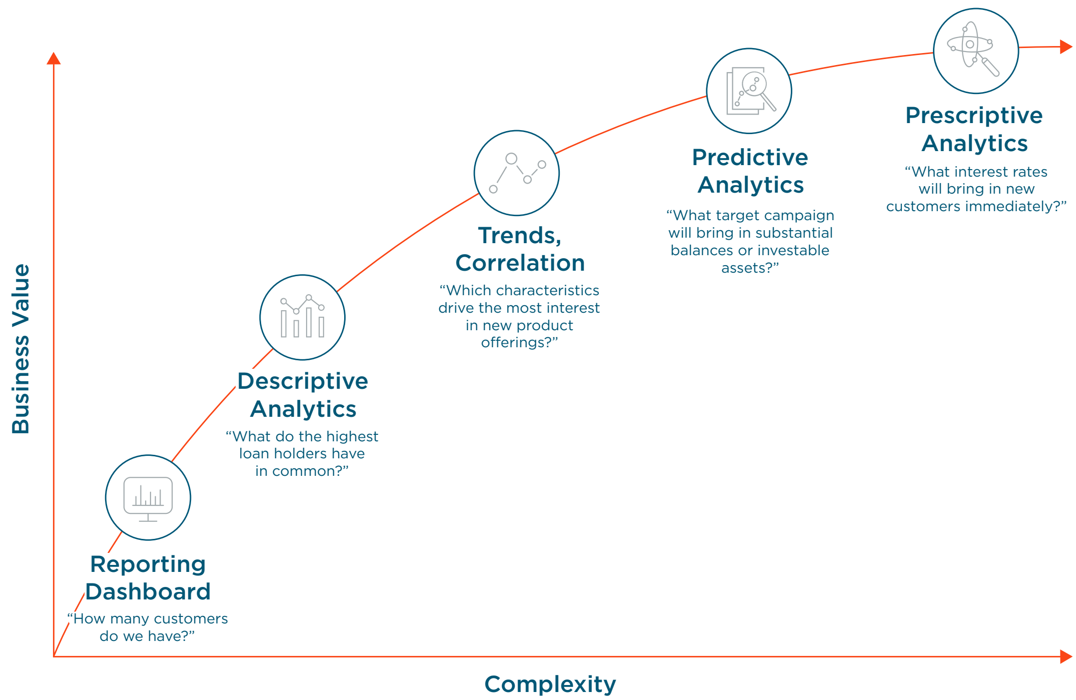
Sensor data typically requires special handling since it is usually real-time in nature and transported over message buses like MQTT. In many instances, this real-time data will require stream processing in order to apply statistical functions and calculations to it before storing it in a time series database. See below for more detail about stream processing (event processing) and time series data.

Building a clean set of validated training data is critical to achieving any degree of accuracy in machine learning models. Extracting data from all required systems, removing duplicates and errors, normalizing data when needed, and transforming it into a consistent format is usually impossible to accomplish without specialized tools. Even then, a solid data preparation process can often consume a larger time investment than building, testing, and running machine learning models. It is important not to underestimate the effort required or the value of performing the preparation process well in terms of the overall project outcome.

Selecting the appropriate data sources, figuring out how to best access the data, devising methods for combining them, and then ensuring that the output of the process is producing truly usable data requires insight into the business. Data preparation is not a task for people who have little understanding of the business context of the data. People who understand the operation and the business challenges being addressed must be intimately involved in addressing this vital step in the workflow.

ABOUT MACHINE LEARNING

Predictive analytics is the continuation of the analytical journey towards faster insight. Successful organizations overcome the complexities of data analytics to conclude when or how something may happen. They can also understand what will happen in the future based on changes in how data is measured. Put another way, predictive analytics is a defense against disruptive forces. To prevent, you must first predict. And if you predict accurately, you can address business problems before they become major issues. Predictive analytics can also identify useful answers to questions that have not yet been asked.



Implementation and proper use of predictive analytics and machine learning algorithms are critical elements in helping an organization respond effectively and quickly to a wide range of changing external factors.

Machine learning is defined by algorithms that improve themselves without relying on explicit programming to adapt to changing inputs and recommend appropriate actions. Machine learning algorithms learn automatically from observing hundreds of thousands, or even millions, of data points.

Predictive analytics and machine learning both help organizations develop insights from the hundreds or thousands of dimensions in data that people cannot see without computational assistance. Predictive analytics relies on a person's ability to understand and test the relationships between cause and effect; this is accomplished by refreshing the data inputs or updating the model design itself. Machine learning removes people from this step by recalibrating models autonomously after they are put into use.

Some data sets contain known target (or dependent) variables that users want to understand better; in these cases, the predictive analytics model will use other independent variables to predict how the target variables will respond to different inputs. Target variables in a dataset are often labeled as column and row headers. Supervised learning algorithms use target variables to classify datapoints and generate insight.

The unsupervised approach to predictive modeling is well suited to datasets containing unclassified and unlabeled records. In these cases, there is no target variable to measure (or predict) against. Unsupervised models identify hidden patterns and group records in clusters. An unsupervised learning algorithm acts on the dataset without human guidance and sorts the data into groups based on similarities between records. Deep learning (or neural network) models are common examples of unsupervised machine learning models. They emulate the human brain to perform nonlinear deductions. Deep learning systems generate decisions by learning from previous interactions or transactions to formulate conclusions and will adapt as patterns in their data inputs change.

ABOUT STREAMING ANALYTICS

Streaming analytics enables the management, monitoring, processing, and visualization of real-time data streams. In manufacturing applications, sensors are a typical source of real-time data which may be streamed over MQTT or similar message infrastructures.

Streaming analytics tools permit engineers and managers to analyze data as it becomes available. With the right data visualizations, they can identify outliers, clusters, trends, and anomalies in seconds. They can roll back time to understand the causal relationships that created the issue, devise a possible solution, and even backtest their solution using historical data before putting it to use.

A streaming analytics platform must be able to connect directly to the message buses in use within the firm without requiring middleware, apply statistical functions and calculations on-the-fly to millions of events every second. This platform must be able to output its own streams of real-time data so it can be captured in a specialized time series database and/or displayed in an analytics dashboard that combine the real-time data with historical data from many other sources.

As with every other step in the analytics workflow, providing business users, including manufacturing engineers, with the ability to connect to data sources, develop and test stream processing applications, and build their own user interfaces is critical to the overall success of an analytics initiative. This self-service capability vastly reduces the time-to-market for new applications, allows the people who understand the operation best to try new ideas without lengthy consultative processes involving a separate IT group, and modify existing analytics applications as needed based on changing conditions.

EMPOWER ENGINEERS WITH SELF-SERVICE DATA ANALYTICS

Specialized data science teams are often vital to the implementation and utilization of data analytics within a manufacturing organization. However, businesses quickly discover that depending on a separate group is slow, expensive, and can add confusion to the decision-making process. Building, managing, and interpreting all required data flows along with creating machine learning algorithms, streaming processing applications, and dashboards is a lengthly process. In order to add substantially to the overall productivity of the firm, data scientists must be able to work directly with the engineers and managers to fully understand the questions, problems, and decisions that must be addressed.

The right data analytics tools do not require a degree in data science to use effectively. In the most successful firms, engineers and managers who know the business are able to access data sources, build and test predictive models and event processing applications, and visualize the results on their own. This self-service approach fosters utilization of the tools, reduces the time-to-market for analytics-based business processes, and helps ensure that analytics projects are adding real business value. The data science team can focus its attention on developing and promulgating best practices, supporting efficient use of the tools, and digging in on the most difficult analytics challenges.

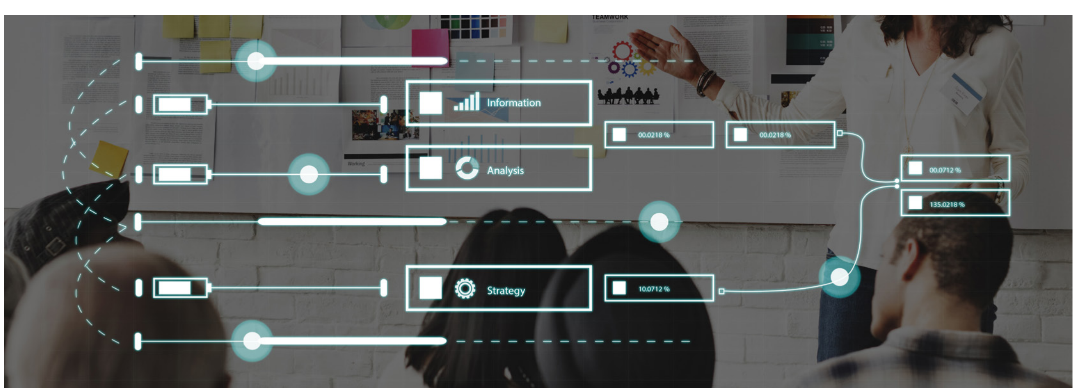
OPTIMIZE OVERALL EQUIPMENT EFFECTIVENESS (OEE)

High OEE levels enable firms to maximize the Return of Investment (ROI) for manufacturing equipment, improve output quality, increase competitiveness, reduce downtime and maintenance costs, and maximize productivity. The right data analytics tools support real-time monitoring of OEE. They also have the ability to dig deep into historical production, maintenance, and quality data to understand which approaches have made the biggest positive and negative impacts on past performance as well as the methods that have the greatest potential for increasing OEE.

Integrating data from multiple silos, including data streams from sensors and historical manufacturing data which may be stored in a variety of formats and systems, allows manufacturing engineers and managers to push improvements in performance, quality, and availability, the three primary Key Performance Indicators (KPIs) for OEE.

Data visualization software optimized for real-time data is especially useful in OEE applications. It enables managers to drill down to understand how well individual machines are operating currently as well back-trace potential causes of faults that have created OEE shortfalls in the past for any particular machine or group of machines. Combining real-time data visualization with stream processing and machine learning functions is even more powerful since data streaming in from Industrial Internet of Things (IIoT) and other sources on the plant floor can be compared on-the-fly with historical data using a large number of statistical techniques that can reveal problems that would otherwise be hidden.

Machine learning tools can efficiently identify issues and weak spots in production machines and proactively alert operators to potential causes of downtime while there is time to plan for necessary maintenance or component replacement. The ability to analyze sensor data streaming in from production systems and correlate it with maintenance logs and other sources can support significant improvements in OEE and reduce maintenance costs.



USE DATA ANALYTICS TO IMPROVE QUALITY

Great product quality wins new customers and keeps existing customers coming back. Consumers and business buyers have come to expect high build quality, long mean times before repair, and long product life.

Poor data in the supply chain is often the root cause of a product quality issue. Pricing, quoting, configuration, and delivery instruction errors can slow down production, result in cost-related margin reductions, and artificially create the need for component substitutions.

The right tools enable engineers, analysts, and manufacturing supervisors to aggregate and cleanse data from any source, ensure proper data governance throughout the analytics lifecycle, and develop consistent, accurate data that can feed machine learning, predictive analytics, and visualization systems which then produce reliable outputs.

Real-time data comes into play as well, since solving complex production problems often requires visibility into massive amounts of sensor data and historical information. The ability to access and analyze real-time data effectively enables faster, better informed decision-making to reduce production costs and quality issues driven by rework, returns, unscheduled machine downtime, and incorrect orders.

CASE STUDIES IN MANUFACTURING

The following are examples of common applications for data analytics within the manufacturing environment.

Tool Condition Monitoring (TCM)

Tool wear in metal cutting operations has a direct impact on the quality and accuracy of the finished surfaces. Various types of sensors can monitor how well a tool is performing, measure generated heat, speed, pressure, and other factors that, alone or in various combinations, signal that a tool is approaching the end of its life. Replacing a tool at the optimum time — while it is still performing up to spec but just before its degradation begins to cause damage, reduce output quality, or increase scrap rates — is highly desirable.

Machine learning and stream processing technology are natural fits for TCM analysis. The large amounts of data produced by sensors combined with human inspections of finished pieces can be used to train machine learning algorithms to identify the "sweet spot" and proactively alert operators when a tool is approaching time for replacement. Stream processing algorithms can also process all the sensor data being generated by any number of production machines, make on-the-fly comparisons with historical data, and amplify the accuracy of the machine learning algorithms.

Anomaly Detection in Production Systems

Identifying unusual behaviors or patterns in machine components using sensor data can prevent small glitches from creating major operational problems. In cases where large numbers of sensor feeds are involved, challenges emerge due to the sheer volume and velocity of data streaming off the equipment. In addition, meaningful analysis from the data is a nontrivial task, since slowing or shutting down production in order to examine a machine carefully should only be done when truly necessary. For these reasons, simple threshold-based alerting is normally unsuitable; it will generate too many false positives. More advanced methods can, however, easily be implemented and will flag potentially serious issues without reducing OEE.

Machine Failure Prediction

Machine learning technology leverages historical and real-time data from sensors mounted to production equipment as well as PLCs, SCADA, and other sources and can accurately flag potential failures of whole machines and/or critical components before they can cause downtime. Failures may be binary in nature; that is, either a failure occurred or not. Failures can also be multi-class and fall into several different categories, including reduced speed, throughput, or quality. Obviously, the more complex the machine (or system), the more machine learning models can help prevent failures that can impact productivity.

Root Cause Analysis

Root Cause Analysis (RCA) is critically important to the ongoing success of any manufactured product. Detecting design defects, raw material problems, build issues, and quality control shortfalls as early as possible fosters continuous product improvement, increased reliability and performance, and allows the company to maintain strong reputations for product brands.

RCA enables R&D personnel, purchasing agents, quality control inspectors, and the warranty team to find the fundamental cause of problems that may only become evident once products are in customers' hands. In addition to the benefits noted above, managing the RCA process properly will reduce the number and cost for warranty claims, improve profitability for the entire firm, and increase customer satisfaction.

Service Pack Optimization

Price optimization in any realm helps companies improve competitiveness, but when applied to service offers like extended warranties, service contracts, and out-of-warranty repairs, it can also condition customer responses to additional offers, mitigate production quality issues, and modulate demand to take into account the company's ability to deliver services. Optimizing service packs also helps convert what may be an unprofitable requirement for manufacturers into a significant income generator.

Offering the right service packs has a major impact on customer loyalty and brand reputation as well. Every manufactured product line will exhibit quality control issues, field failures, and premature end-of-life issues from time to time. Dealing with problems like this at prices that customers see as fair is critical, and likewise, offering warranty extensions and service contracts at attractive prices helps reduce customer worries about product longevity, particularly with new product introductions.

Manufacturers can also use data analytics to optimize the length and terms for original warranty offers by taking competition, product price points, customer preference survey results, and other factors into account.

Warranty Risk Profile Analysis

Most manufacturers must handle large numbers of warranty claims related to a variety of products and components. The volume of claims can easily run to millions per year for consumer goods manufacturers. It is critical to prioritize and understand which issues deserve high priority responses and detect patterns within the claims that indicate emerging quality or design problems that requiring immediate attention.

Warranty risk profile analysis, sometimes referred to as quality issue prioritization, is a vital part of any ongoing quality improvement process. The data from warranty claims, once cleansed and sorted, is one of the most valuable parts of the feedback loop that enables companies to improve the reliability and customer satisfaction for their products.

Expert Systems

Expert systems are computer programs that attempt to emulate a human expert, usually in a narrow field of expertise. Such advanced data analytics tools can augment or even replace the work of skilled engineers in the manufacturing environment, but with some important limitations. The best cases for applying expert emulation are those where the "rules of the game" are relatively easy for people to understand but difficult to write down or define in simple sets of rules. For example, the process of figuring out the optimal processes and order of work for sheet metal stamping is complex and typically requires the attention of manufacturing engineers with many years of experience. However, by applying machine learning techniques, an expert system can determine with a high degree of accuracy the best configuration of processes for a given sheet metal component in order to reduce waste, increase quality, and improve throughput.

Reduce Accounts Receivable Days Outstanding

Days Sales Outstanding (DSO) is a critical performance measure for many manufacturing operations, and anything that can reduce the firm's DSO will improve the bottom line. A common challenge involves the many different systems that a firm may use to manage its inventory, production processes, shipments, sales, and accounting. Reconciling data from such disparate systems, which were often implemented by different teams and different times with different objectives, is a nontrivial task. Analytics tools that can directly access the output and internal data resources of all relevant systems, cleanse that data, and transform it into governed, accurate, and useful information are therefore crucial. Machine learning and other data science algorithms can then be applied to anticipate potential slow payers and increasing DSO numbers before they can affect the business.

Price Optimization

Price has an immediate impact on demand, and finding the right price ensures that factory production is closely matched with demand. With the right statistical models and tools, manufacturers can determine how customers will respond to different prices through different channels and figure out the pricing models and price levels that will maximize operating profit, maintain market share, and fend off competition. Most markets are high competitive with constantly evolving customer requirements and interests, meaning that price optimization must be a continuous process.

Supply Chain Risk Management

Supply chains for most products today are long, unpredictable, and complicated, and involve hundreds or thousands of third parties. It presents one of the main areas of risk for high volume manufacturers. Firms can utilize wide varieties of data from government, shipping firms, suppliers, and companies with complementary offerings to prepare for supply chain delays and develop contingency plans. They must be able to clean and normalize all useful data and develop predictive models they can tweak on an ongoing basis as updated information becomes available.

Data analytics exposes the cost and quality for every outsourced component in your production life cycle. For example, managers can determine whether certain components are failing more frequently than expected or are not performing up to spec, and then use machine learning to help determine the best remedy.

Keeping pace with the continuously changing nature of the supply chain makes it essential to have a real-time view of order, delivery, and usage flows.

ALTAIR ADDRESSES DATA ANALYTICS CHALLENGES

Insight doesn’t come from what people can see on the surface of a data set. It comes from the hundreds or thousands of dimensions hidden in complex data. People need the right tools to easily access these hidden dimensions. Altair empowers business users to collaborate efficiently to access meaningful data, generate insight from this data, and share their finding throughout the enterprise.

Make fast, insightful decisions for your manufacturing operation using clean, comprehensive data and analytics.

Data Preparation

Altair’s data prep software enables businesspeople to build, discover, share, and collaborate on secure, governed, and trustworthy data sets and models. These tools can access, cleanse, and format data from a wide variety of sources (including Excel, CSV, PDF, TXT, JSON, XML, HTML, SQL databases, Big Data like Hadoop, and more) without any manual data entry or coding. Dozens of pre-built data preparation functions make combining disparate but related data sets easy to do quickly. This simple approach to data cleansing eliminates the need to code, script, or create pivot tables or vLookups in Excel.

Clients can deploy these tools on the desktop, with on-prem servers, or in the cloud.

Predictive Modeling and Machine Learning

Altair’s open, flexible predictive analytics platform is designed for data scientists and business analysts alike. Its industry-leading visual approach to analytic modeling enable data science teams to create high quality machine learning and AI models. Our collaborative approach to machine learning enables your data scientists and business users to minimize repetitive takes related to creating curated and governed data sets, share knowledge across the enterprise, and reuse steps within connected model workflows for faster analysis and sharing of insight. Altair's code-optional development environment enables data science teams to build models using combinations of SAS language, Python, R, and SQL code.

Stream Processing

Altair’s stream processing (also referred to as “event processing”) engine connects directly to a wide range of real-time streaming and historic time series data sources, including MQTT, Kafka, Solace, and many others. Users can build complex stream processing applications with a fully drag-and-drop interface, without writing any code. Applications may combine streaming data with historic data, calculate performance metrics using a wide variety of statistical and mathematical functions, aggregate, conflate, and compare data sets, and automatically highlight anomalies against user-defined thresholds.

Data Visualization

Altair's visual analytics platform is optimized for handling time-critical data, including data that may be changing with extreme rapidity. Business users can connect to data sources, build, and publish sophisticated real-time dashboards. The platform's filtering tools enable users to zoom in and out on the timeline, remove false positives from the screen, and focus on exceptions. Users can solve difficult problems quickly, understand complex relationships in seconds, and identify issues requiring further investigation with just a few clicks.

In order to make the firm more productive, data scientists must be able to work directly with the engineers and managers to fully understand the questions, problems, and decisions that must be addressed.

APPENDIX: DEFINING TERMS

A comprehensive glossary of data analytics terms is beyond the scope of this guide, but here are some key concepts that will be helpful to anyone looking into utilizing analytics to improve manufacturing operations.

Algorithm

In the context of data science, an algorithm is a series of repeatable steps that is often expressed in the form of Boolean logic. A simplistic example of this is: IF A=0 THEN SET B= 2. Data scientists can develop and implement algorithms using many different tools and methods, including coding in languages like Python or R, or within the framework of a machine leading software system. Some commonly used algorithms in data science include linear and logistic regression, decision tree, random forest, Naive Bayes, and KNN (K-Nearest Neighbors).

Artificial Intelligence (AI)

AI is something of a catch-all term, but at its core it is technology that can extract useful insights and identify patterns in large data sets, and often produce predictions based on that data. The concept — and perhaps the holy grail of data science — is to develop systems that can emulate human thought processes. In practice, however, the best uses of AI are as intelligent assistants that can, for example, troll through millions of purchase orders and bills of lading looking for examples of overbilling or delivery failures.

Big Data

Big Data is really a buzzword without a widely agreed definition, but put simply, it a set of data that is too large to fit on a single computer. Conventional tools like SQL databases and Excel are usually incapable of managing Big Data successfully and specialized tools are required. Most organizations have untapped Big Data resources that, when properly analyzed, can reveal useful insights into customer behavior, operations, and potential gains in production and delivery efficiency.

Code-Free

The ability to develop sophisticated software applications without writing and debugging programming code. Code-free tools are typically designed for use by businesspeople who wish to build systems to procure, analyze, and make predictions based on the data available to them.

Data Cleansing

The process of removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted. Cleansing is a critical step in the data preparation process, since “dirty” data will contaminate the outputs of machine learning or other analytics applications and can lead to incorrect conclusions and predictions.

Data Preparation

The process of gathering, combining, structuring, and organizing data so it can be used in machine learning, predictive analytics, visual analytics, and/or business intelligence (BI) applications. Data preparation systems must be able to connect to and collect data from a wide variety of disparate sources in a range of formats and data types, cleanse the data, normalize it if necessary, ensure that the process does not introduce errors of its own, and output the data in a consistent format usually consisting of rows and columns in a database.

Data Governance

A collection of practices and processes for data management in order to ensure that the ancestry of the data is known and understood and that only qualified people can alter or augment the data. The Data Governance Institute defines it as: "A system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods." Proper governance reduces the costs associated with managing large amounts of data, supports reproducible and accurate procedures around regulation and compliance actions, increases the value and utility of the data.

Data Visualization

Numeric data displayed in graphic form, including everything from simple line graphs and pie charts produced in a spreadsheet program to highly sophisticated interactive dashboards.

Data Modeling

In the context of data analytics, modeling involves building sets of algorithms that can make accurate predictions about future events based on historical data. Ideally, a model's predictions produce forecasts that businesspeople can use to make informed decisions.

Decision Tree

A decision tree is a graphical depiction of decisions, or nodes, in which every potential outcome is directed to a new branch of the tree. Decision tree-based algorithms are one of mostly used methods employed in machine learning systems; they offer high degrees of accuracy and stability and are easy for users and consumers of their output to understand.

Explainable AI

Simply put, this refers to systems that allow ordinary humans to understand how a set of algorithms in an AI system produced its outputs. To put it another way, explainable AI systems are the opposite of co-called "black box" systems that produce results that cannot be interpreted visually.

Explicit Programming

This is the type of computer programming most people are familiar with. Every instruction to the computer must be written out in a suitable computer language, and altered manually when new parameters need to be added or altered in the program. Enabling predictive models to making accurate predictions requires thousands of parameters, and therefore an explicit programming approach is generally not scalable for most practical applications in data science. Instead, most truly useful and effective predictive models make use of a combination of explicit programming and implicit programming (see machine learning below), which relies on the system itself to produce outputs that are not programmed step-by-step by a human.

Event Processing

Event processing — also called stream processing — involves taking action on a series of data points, with each data point being an "event" in a continuous stream of (usually real-time) data. Typical actions include aggregations, conflations, comparisons, and other types of calculations that may involve data from the real-time stream as well as historical data stored in a specialized time series database.

Machine Learning (ML)

Arthur Samuel coined the term "machine learning" and defined it as follows: Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed.

In other words, the algorithms in the system learn by themselves. ML systems are highly scalable and can adapt quickly as new variables come into play. They can predict outcomes without explicit human input.

A key concept of ML is the "training set". ML systems must be fed a historical data set for which the insights and patterns to be found are already known and confirmed using an existing technique. After "training" the ML model, the system can predict new results based on new data inputs.

Python

An open source object-oriented programming language. It is popular in the data science community because its user community has developed and published an extensive library of useful programming objects. Compared to many other languages, Python is also easy to learn and use.

R

An open-source language and environment optimized for statistical computing and analysis. Like Python, R is often used in data science applications. It is generally considered to be more difficult to learn than Python, but offers an excellent array of graphical and plotting capabilities.

Real-Time Data

Real-time data (also referred to as “streaming data”) is continuous with no beginning or end, and the individual data points in the stream of data may appear at regular or irregular intervals that are typically in the sub-second range. Nearly all operations produce some amount of real-time data from a variety of sources like sensors and transaction processing systems.

Real-time data streams are usually carried on a message bus (also called a message queue) capable of moving large amounts of data between distant points with extremely low latency. Examples of popular message buses include MQTT, Solace, and ActiveMQ.

SAS

The SAS programming language was created at North Carolina State University in the early 1970s to help analyze agricultural research data. The language of SAS remains a staple of the data science ecosystem.

Stream Processing

See “event processing” above. The terms “stream processing” and “event processing” are interchangeable.

Streaming Data

See “real-time data” above. The terms “real-time data” and “streaming data” are synonymous.

Time Series Data

This is essentially a recording a real-time data stream and is usually stored in a specialized time series database. Popular time series databases include InfluxDB, Prometheus, TimescaleDB, kdb+, Azure Time Series Insights, Amazon Timestream, and many others.

In the most successful firms, engineers and managers who know the business are able to access data sources, build and test predictive models and event processing applications, and visualize the results on their own.

RELEVANT CASE STUDIES AND CUSTOMER STORIES

- Expert emulation at Ford: https://www.altair.com/resource/driving-manufacturing-decisions
- Altair® Monarch® simplifies collection process for healthcare products manufacturer: https://www.altair.com/customer-story/earthlite
- Altair helps global industrial company improve quality and unit reduce costs: https://www.altair.com/resource/identifying-and-capturing-business
- Altair Monarch enables cost reductions, improved reporting accuracy, greater efficiency at global automotive parts manufacturer: https://www.altair.com/customer-story/a-leading-manufacturing-company-customer-story
- Data Analytics for High-Performance Motorsports: https://www.altair.com/resource/data-analytics-for-high-performance-motorsports
- Digital Twin for Sustainable Energy: https://www.altair.com/resource/digital-twin-for-sustainable-energy
- Expert emulation and crash optimization using machine learning: https://www.altair.com/resource/expert-emulation-in-crash-optimization-using-machine-learning