

# MAKE MACHINE LEARNING WORK FOR YOU: AN EXECUTIVE'S GUIDE



Knowledge Studio

WOE Helper - Binning Editor

Ac Status

Category	Bad	Good	Total	WCE
Farming-fishing	76 7.14%	988 92.86%	1,064 13.34%	0
Handlers-cleaners	193 17.95%	1,191 88.05%	1,384 17.35%	-0.855
Adm-clerical	315 22.71%	1,192 77.29%	1,507 17.39%	-0.958
Machine-op-inspct	385 28.27%	991 71.73%	1,376 17.1%	-0.992
Craft-repair	1354 17%	1387 17%	2741 17%	-2.234
Transport-moving	833 10%	7485 89%	8318 10%	-1.485
Protective-serv	1974 25%	5992 75%	7966 100%	-0.206
Armed-Forces	1974 25%	5992 75%	7966 100%	-0.206

WOE: -0.992

Bin	Bad	Good	Bad	Good	%	WCE	
Other-service	27	806	3.2%	96.8%	833	10.4%	-2.2313
T...	76	988	7.1%	92.9%	1,064	13.3%	-1.40005
Adm-clerical	193	1,191	13.9%	86.1%	1,384	17.3%	-0.85496
Craft-repair	315	1,072	22.7%	77.3%	1,387	17.4%	-0.9591
Protective-serv	385	951	28.8%	71.2%	1,336	16.7%	-0.9602
Armed-Forces	385	951	28.8%	71.2%	1,336	16.7%	-0.9602
Total	1,887	6,081	23.8%	76.2%	7,978	100.0%	

Initial Entropy: 0.11833 Final Entropy: 0.80534 Initial IV: 0.05054 Final IV: 0.05054

Optimization:  Monotonic  No Pure Nodes Entropy Equal Size: 0.3

Update Gini View WCE

Cancel Back Next

# INTRODUCTION

Have you ever wondered why some of your online payment card transactions are processed in an instant, while others involve requests for further action? Typically, that might mean supplying a one-time password (OTP) or PIN. Occasionally, your purchase might even be blocked.

In almost every case, the answer will likely involve machine learning (ML). That's because payment card issuers now use this technology to spot transactions that might be the work of fraudsters or criminals. When their ML systems identify a potential risk, those systems can help decide the best course of action, like putting a stop on the transaction or asking for proof that the customer is genuine.

Protecting consumers and enterprises involved in online transactions is just one example of how ML influences our daily lives. In fact, the list of use cases is already long, diverse – and growing fast. The reason is clear – ML is a game-changing tool that enables organizations to make better decisions faster. What's more, ML is highly effective at balancing conflicting objectives. In the case of online transactions, the aim of the card issuer is not simply to tackle fraud. That could be achieved by making the requirement to provide an OTP (also known as two-factor authentication) standard with every payment. In practice, however, enterprises must balance fraud prevention with the need to maximize sales and ensure a positive customer experience. That means keeping the process fast and frictionless for as many genuine customers as possible, while reliably identifying suspicious transactions.

ML helps banks and other financial enterprises handle conflicting objectives when making decisions on loan or mortgage applications. In this highly competitive market, lenders use ML to quickly distinguish between applicants who are likely to meet their repayment schedules, and those who pose a substantial default risk.

There are numerous applications for ML in the insurance sector as well. When a consumer uses an online tool to search for the lowest cost auto insurance, for example, insurers typically use ML models to generate competitive quotes on the fly using historical data on competitor pricing and other parameters.

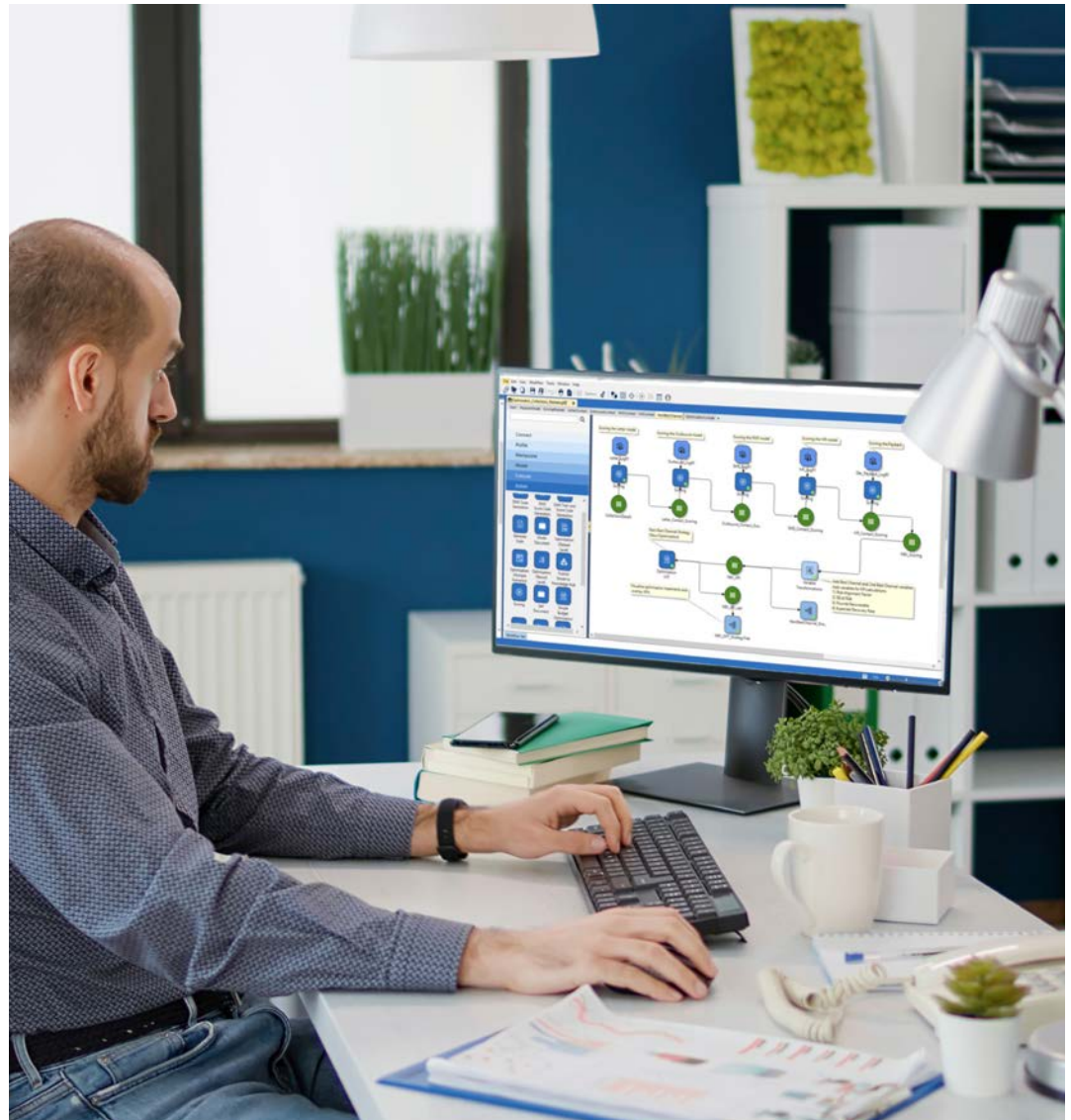
In industries with physical assets and products like automotive and aerospace, ML is a perfect tool to help manufacturers make accurate predictions about when to replace key components in production machinery, forecast defects and other quality control issues, and predict optimal times to order components and materials.

ML is also valued by businesses because it can quickly group together data with similar characteristics. This explains why ads in your social media feed seem to have an uncanny ability to match your interests, and how your favorite streaming service generates personalized suggestions for the next series to binge.

Given the breadth and depth of potential use cases, one thing is clear - more and more people will find themselves working in environments where ML plays a critical role. And thanks to the emergence of low-code and no-code software, ML is no longer the exclusive preserve of programmers, data scientists, and people who paid attention in math class. More of us can - and will - be involved in developing and deploying practical ML solutions.

**Fair For You (FFY) is a non-bank financial institution that provides loans to low-income families. By switching to an ML-enabled, data driven approach for real-time credit scoring, FFY increased its loan application acceptance rate from 20% to 27% within just 120 days, while reducing the number of customers who went into arrears within the first two months by 28%.**

[Read the full story here.](#)



Simply put, ML uses algorithms to create a model based on sample data (known as training data) to make predictions or decisions. ML systems learn from data without having to be programmed explicitly.

This eGuide provides a brief introduction to the world of ML. What is it? What are the keys to success? What roles can ML play within an organization, and what benefits should it deliver? Hopefully, it'll also encourage you to take a deeper dive. At the end, you'll find links to a comprehensive library of ML videos, case studies, and more.

[02 / Introduction](#)

[06 / What is ML?](#)

[09 / Why Use ML?](#)

[09 / Where Should an ML project start?](#)

[11 / Data: The Vital Ingredient](#)

[14 / Predictive Analytics - What Happens Next?](#)

[15 / Picture This: The Importance of Decision Trees](#)

[16 / Unsupervised Learning - What Do We Have In Common?](#)

[17 / Learning on the Job](#)

[18 / Prescriptive Analytics - What Should We Do Now?](#)

[19 / ML in Action](#)

[20 / Building Better ML Solutions](#)

[21 / The Democratization of ML](#)

[23 / Glossary](#)

[26 / Resources](#)

**According to Fortune Business Insights, the global market for machine learning is expected to be worth \$209.91 billion in 2029, up from \$15.44 billion in 2021.**

Machine Learning (ML) Market Size, Share and COVID-19 Impact Analysis

Fortune Business Insights

March 2022

<https://www.fortunebusinessinsights.com/machine-learning-market-102226>

# WHAT IS ML?

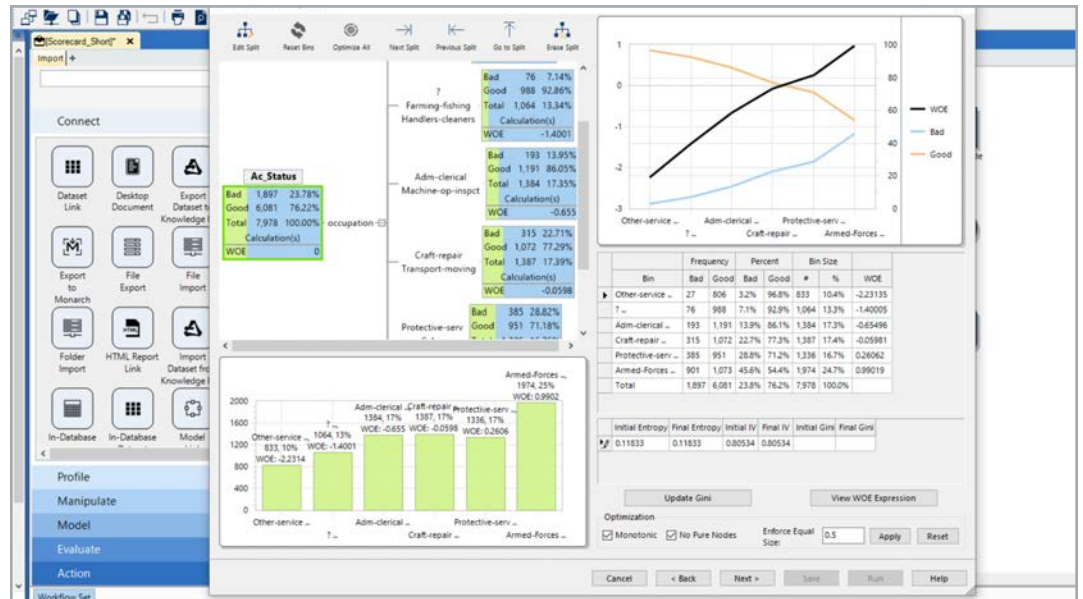
There's certainly no shortage of hype surrounding ML, with industry analysts predicting dramatic investment growth in the coming years. However, while ML has certainly secured buzzword status, explanations as to what it is, and what it can deliver for businesses are somewhat vague.

**ML is often used in conjunction with an equally hot topic – artificial intelligence (AI). In fact, some people use the two terms interchangeably. But while there are no hard and fast rules as to where one ends and the other begins, ML is generally regarded as a subset of AI.**

**In other words, ML is how a computer system develops its AI – the ability to simulate the reasoning that humans employ instinctively when they gather new information and make decisions based on it.**

In practice, AI is a nebulous concept. In comparison, ML is much easier to nail down. So, let's start with a simple definition: ML uses algorithms to models based on sample data (known as training data) to make predictions or decisions. Crucially, and as the name suggests, ML systems learn from data and produce new outputs when presented with new data.

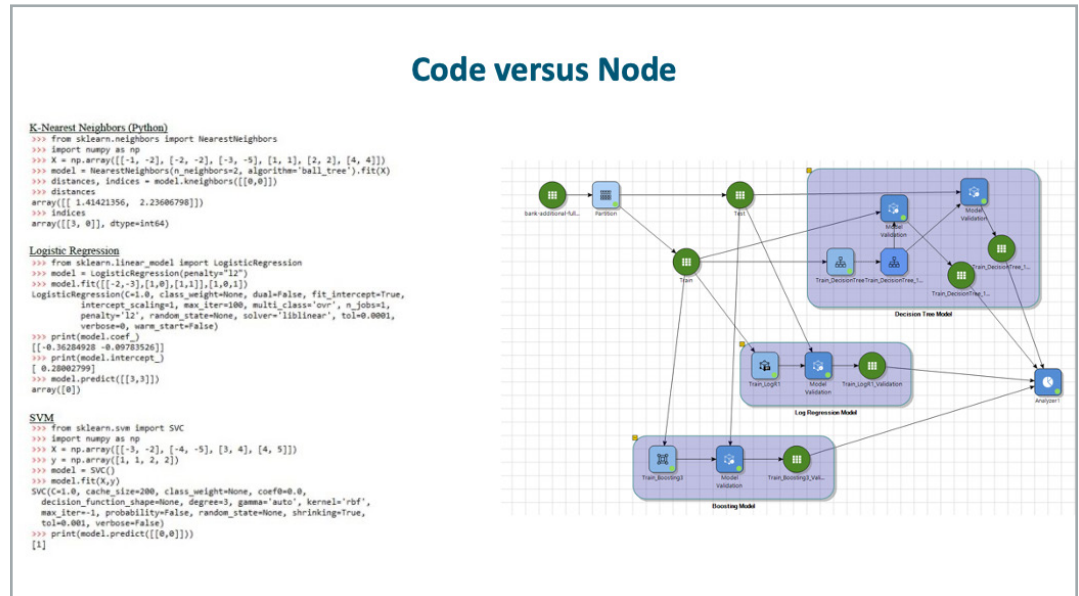
Algorithms are the building blocks of ML. So, here's another definition: an algorithm is a set of step-by-step instructions for a computer, designed to fulfill a specific task.



Decision trees are easy to understand and quick to build, making them one of the most useful ML models. We explore the role of decision trees in more detail later in this eGuide. However, in this example, the decision tree shows the key attributes that will determine whether a customer is likely to stay with a bank or leave it. Banks use systems like these to predict customer churn rates.

In the context of ML, data scientists have traditionally developed and implemented these models using a range of different methods. Notably, these include coding in R, Python, and the SAS language. However, state-of-the-art software systems are democratizing and simplifying the creation of algorithm-based models. As a result, ML is becoming accessible to a far wider array of potential users.

## Code versus Node



The image above illustrates how the latest software is transforming ML. The highlighted boxes on the right are three separate predictive models, and each model is using a wizard-driven algorithm in addition to other nodes in a simple, visual representation of the traditional algorithm coding on the left.

ML is not a new concept. Despite all the current interest and excitement, the technology dates back as far as 1959, when the term “machine learning” was first coined by Arthur Samuel, an IBM employee. Even in the commercial domain, ML has deep roots. Deployment first started to gather momentum in the 1980s when the financial sector pioneered the use of ML for credit scoring.



# WHY USE ML?

There are compelling reasons why investment in ML is on a steep upward curve. When designed and implemented correctly, ML can radically improve the accuracy, speed, and efficiency of decision-making in a diverse range of business settings. In many cases, ML has the potential to fully automate the decision-making process.

Successfully harnessing this technology can make organizations more productive, efficient, and profitable. ML also enables enterprises to monetize an often-neglected resource: the vast amounts of data they already generate from their normal day-to-day activities.

## Where should a ML project start?

Organizations looking to take advantage of ML should always start with this fundamental question: what outcome do we want to predict?

It may sound obvious, but one of the more common mistakes that companies can make when they decide to invest in ML is simply failing to define their objectives clearly enough.

**“The first and most important step in any machine learning project is to determine what exactly you want to predict.”**

Sam Mahalingam, CTO, Altair

To go back to our example of online shopping, the payment card issuer is using ML to predict whether the customer trying to complete a purchase is genuine or a fraudster. The card issuer is then using ML to help decide the most appropriate course of action.

Similarly, in our example of a financial institution offering mortgages or loans, the lender is using ML to predict whether an applicant is likely to default. On top of helping the lender decide whether to offer a loan, they might also use ML to determine the loan's value, length, and interest rate.

In the manufacturing sector, this ability to predict outcomes is proving just as valuable. For example, organizations can use ML to predict if production machinery is likely to break down – and it can then help decide what action the team should take. As a result, teams can target maintenance activity far more accurately, preventing sudden machine failure and subsequent disruption to the production process. At the same time, the organization may be able to eliminate some routine maintenance activities if ML predicts shutting down a machine for scheduled maintenance will have little or no impact on reliability.

Across an array of commercial and industrial sectors, there are numerous examples of ML transforming the efficiency and effectiveness of business processes. ML is delivering clear, quantifiable benefits; it's reducing costs, increasing sales, and improving profit margins. But in many respects, we're still scratching the surface of what it can do. As modern software makes ML more accessible and as more companies appreciate the benefits of ML, we can look forward to some truly revolutionary advances. For example, in the not-too-distant future, ML technology could predict if you're likely to get sick long before either you or your medical practitioner have detected any symptoms. What's more, ML could then help doctors identify the best preventative treatment, fully personalized to your diagnosis.

**Mabe is a leading designer and manufacturer of home appliances sold under the GE Appliances name and its own name. Having recently launched a high-end, connected washing machine, the company is using ML to analyze data generated by 20 different sensors. Providing real-time insight into product performance and failure prediction, ML is informing future product development, and boosting the efficiency and effectiveness of service support, thereby improving reliability and enhancing the customer experience.**

[Read the full story here.](#)

# DATA: THE VITAL INGREDIENT

If the first question that needs to be asked is, What are we trying to predict? The second is, Do we have the data?

Historical data – more commonly known as training data – fuels all ML models. Indeed, the whole concept of ML is based on the idea that patterns and relationships in historical data will continue in the future, enabling accurate predictions about future outcomes.

A good example is the way that commercial lenders, like banks, draw on patterns in historical data when assessing whether an applicant is likely to repay a loan or default on it. Typically, that will involve a wealth of different information, including (but not limited to) credit score, employment, income, and home ownership status.

What lenders are looking for is the impact that this information has on an applicant's ability to repay a loan. In practice, it's the sort of complex calculation that brings out the best in ML. However, here's an extremely simplified example to illustrate the concept: if the data shows that homeowners with stable employment and an annual income of over \$100,000 rarely default on a loan of up to \$10,000, it might be reasonable to predict that this will be the case in the future. As a result, lenders can offer loans of this value to applicants who match the criteria with confidence.

## The Difference Between Dependent and Independent Variables

In the world of predictive models, a dataset consists of variables (columns) and observations (records). Independent variables are inputs to the model and dependent variables are what the model is being trained to predict.

Predictive ML solutions must identify the independent variables with the most predictive power. As the name suggests, these are the variables that have the greatest influence on the dependent variable – the outcome that we want to predict.

It's important to note that while many ML use dependent variables, not all of them do. For example, unsupervised ML models that group and categorize movies based on your previous "likes" don't seek to predict dependent variables.

**“Many organizations have not fully democratized their data, but data must be accessible to monetize it.”**

Mark Do Couto  
Senior Vice President, Data Analytics  
Altair

### We're Going to Need Bigger Data

To consistently deliver accurate predictions, ML needs heaps of data. This data also needs to be as accurate and complete as possible – quantity and quality must go hand in hand. It is also important that the people responsible for developing ML models have a thorough understanding of their data. In other words, an understanding of the domain in which ML is being applied is just as important as “pure” data science or programming skills.

If real historical data is unavailable, data science teams can use other techniques to create suitable datasets, including data augmentation, sampling, and [engineering simulations](#).

To be of any use in making accurate predictions, the data being used to train a supervised ML model must have known outcomes. For example, does the historical data we have on a loan applicant with a \$100,000 salary show whether their loans were paid back on time or if the applicants defaulted? Similarly, does our data show whether a piece of equipment that recorded an abnormally high oil temperature broke down before its next scheduled service or continued operating safely and efficiently? If the data doesn't provide these crucial pieces of information, then it's worthless to a predictive ML model.

### We Didn't See That Coming

To be useful, the data also needs to reflect the current environment. This might mean that historic data no longer offers the same predictive power. The best recent example of this phenomenon is the COVID-19 pandemic. In the loans and mortgages sector, the large numbers of people who suddenly lost their income or were put on furlough added new variables to the decision-making process – it also diminished existing variables' predictive power.

### Keeping It Clean

Often, the necessary data will be drawn from many diverse sources and in many different formats, which can be a major challenge for any ML user. Reflecting this, data preparation is often a project's most time-consuming aspect. However, modern software tools that automate and speed this process are now readily available. These solutions can take data from multiple formats and cleanse and prepare them for analysis.

**When creating credit scoring models in the consumer finance sector, our experience with clients indicates that usually 70-90% of total project time is dedicated to data preparation.**

### ML's First Home Run?

Let's consider a particularly well-known example of good data.

Based on the [bestselling book by Michael Lewis](#), the [2011 film "Moneyball"](#) tells the true story of how an underachieving, cash-strapped baseball team, the Oakland Athletics, used data analytics to create a team capable of reaching the Major League playoffs.

The story is also a great example of conflicting objectives. The Athletics didn't simply need to identify players that could help them to reach the playoffs. The management also needed to identify players they could sign to lower salaries than those often paid by the league's more lucrative teams.

Fortunately, the Athletics were able to draw on a wealth of good data with known outcomes. They could identify teams that had previously reached the playoffs and from there could mine individual performance data for players on successful teams, as well as players who didn't make it to the playoffs. By using ML to analyze this mountain of data, the Athletics identified the variables with the most predictive power.

In this case, those variables included a player's ability to reach base. For the Athletics and their data-driven approach, that was the lightbulb moment. Historically, baseball players were recruited on their ability to hit home runs or for high averages. As a result, such players commanded the highest salaries. By focusing on players good at getting on base rather than on other metrics, the Athletics built a team capable of reaching the playoffs. Just as important, they did it with one of the league's lowest salary budgets.



# PREDICTIVE ANALYTICS – WHAT HAPPENS NEXT?

In the introduction, we highlighted two key strengths of ML. One is the ability to predict outcomes, and the other is to group data with similar characteristics.

Reflecting on these capabilities, two of the most important types of ML models are known as supervised and unsupervised learning models. Supervised models are used to predict outcomes and have a dependent variable. Unsupervised models are used to group together data and don't have a dependent variable.

Most of the examples we've highlighted so far are based on supervised models. That includes fraud prevention in online transactions, loan and mortgage lending decisions, predictive maintenance for production equipment, and identifying baseball players who can take a team to the playoffs. Typically, supervised models use datasets that contain labeled records. As the name suggests, that means data that is clearly labeled or classified, such as a loan applicant's status or a baseball player's on-base percentage.

## **Project the Future: Time Series Forecasting**

Making forecasts based on time series data is different than classical predictive analysis, which relies on static sets of historical data. Time series data can be difficult to work with due to its highly dynamic nature.

Time series data sets have three things in common: the data that arrives is almost always recorded as a new entry, the data arrives in time order, and time is the primary axis. Tools are now available, however, that make working with time series data easy and allow organizations to use it effectively in forecasting applications.

Time series forecasting is useful in a range of applications, including sales demand predictions to inform production planning for a manufacturer, call volume predictions to anticipate personnel requirements in telemarketing operations, inventory and stock analysis, supply chain optimization, and more.

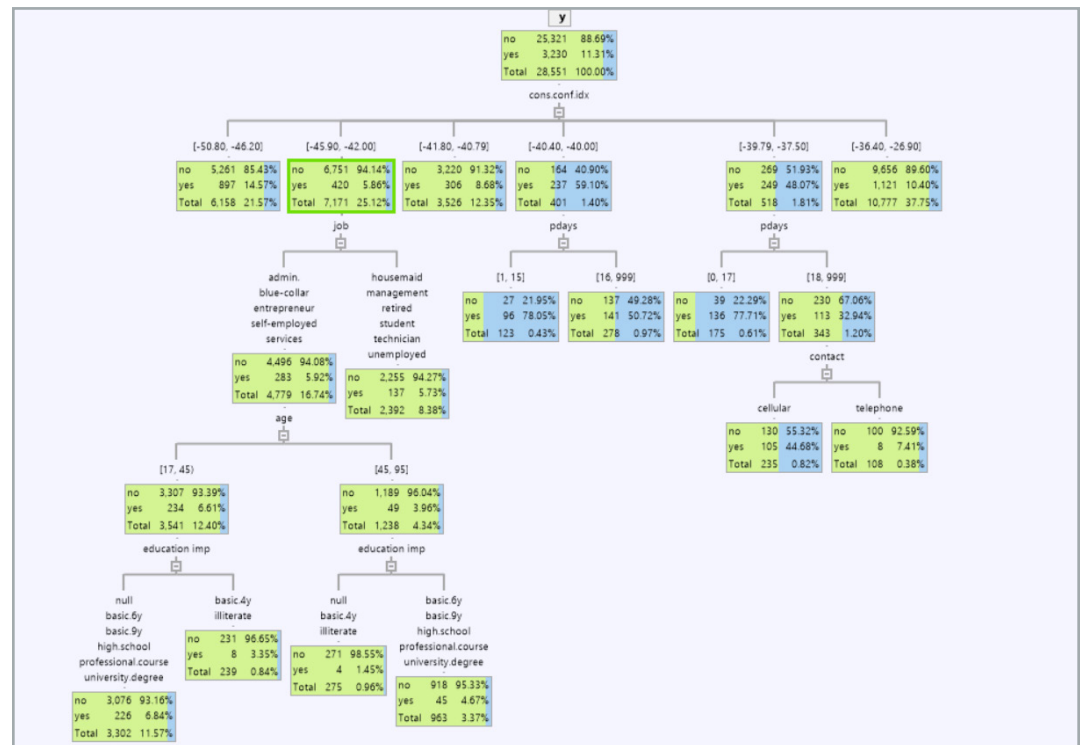
**“Prediction is very difficult,  
especially if it's about the future!”**

Physicist Neils Bohr

# PICTURE THIS: THE IMPORTANCE OF DECISION TREES

There are many different techniques for designing and explaining a supervised ML model. However, probably the most popular option – and certainly the easiest to understand – is the decision tree. In fact, decision trees are popular because they're so easy to understand, particularly when implemented using software tools that offer good visualization capabilities. They're an essential component supporting the idea of “explainable AI”; people with no ML background or training can understand the process a decision tree uses to generate an outcome.

A decision tree is just an illustration of how a specific ML model operates. It's a graphical depiction of decisions (also known as nodes), in which every potential outcome is directed to a new branch of the tree.



This decision tree informs and illustrates the best potential prospects to target in a marketing campaign. In other words, which types of consumers are most likely to respond?

# UNSUPERVISED LEARNING – WHAT DO WE HAVE IN COMMON?

In contrast to supervised learning models, unsupervised learning is the term used when ML is employed to group data with similar characteristics. Supervised learning is all about making predictions, while unsupervised learning is all about finding patterns. In contrast to supervised learning, there's no dependent variable. The unsupervised approach is well-suited to datasets that contain unlabeled and unclassified records.

Examples of unsupervised learning include the program suggestions generated by streaming services. An unsupervised algorithm groups series and movies with similar characteristics together automatically; viewers watching one show in a group will see recommendations about other, similar shows. The more programs a viewer watches, the more refined the personalization process becomes. Done correctly, that should lead to better viewing figures and improved subscriber retention rates for the streaming service.

Natural Language Processing (NLP) algorithms depend on unsupervised learning models to parse grammar, associate correct meanings with similar words, and trace relationships between words and phrases.

The same approach is applied to social media advertising. As the platform gathers data on the things that a user is interested in, ML can identify products and services that share similar characteristics. The platform can then tailor the user's advertising to reflect this, which can lead to more sales or enquiries for the advertiser.



# LEARNING ON THE JOB

As we've seen, there are important distinctions between supervised and unsupervised ML models. However, they share that vital characteristic of all ML models: the ability to recognize patterns in data and learn from them, much like the way people do. However, ML can process much larger volumes of data; it can also recognize data patterns a person might miss.

This learning process can be complex, but think about it like this: people learn from their senses. We can't teach a machine to learn how to feel, taste, or smell, but we can teach the machine with historical data. Once we train the machine, it can go out into the world and predict the future. In that way, it takes cues from raw data much like how we take cues from our senses (which are also a form of raw data).

Supervised learning models use historical data where the outcome is known, to the point where it can predict outcomes based on new data that it hasn't seen before. By their nature, unsupervised models can group and categorize new data just as easily as old data.

For the data scientist, this flexibility is crucial. For example, new variables may come into play, bringing with them new data that will require training new models, finding new patterns, and delivering new insights.

# PRESCRIPTIVE ANALYTICS - WHAT SHOULD WE DO NOW?

While predictive analytics is used to answer the question of what's going to happen, prescriptive analytics answers the question of what action(s) an organization should take to achieve its objectives. Prescriptive analytics is about guiding, or even automating, the decision-making process.

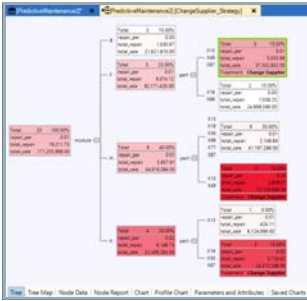
As we've explained above, ML-based decision making is especially powerful in situations with conflicting objectives. For example: the desire to maximize the value of the loan book versus the need to minimize loan defaults; the desire to recruit players that will help the team get to the playoffs versus the need to stay within a fixed budget; the desire to prevent equipment downtime versus the need to avoid disrupting normal operations.



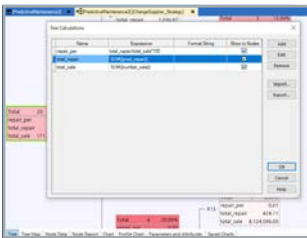
# ML IN ACTION

Putting ML models in production involves complexities and risks that can sink a project or even create negative value. It's critical to have technical and human infrastructure in place that can put models into production smoothly and accommodate updates and reconfigurations without slowing down the business. To deliver value, organizations must operationalize their ML models in secure, governed, and scalable production environments. This process is fundamentally different than typical software deployment processes.

We've highlighted several ML use cases already, but the breadth and depth of ML deployment is much wider and continues to grow. Here are some of the principal industries and areas where ML is playing an increasingly significant role:



This strategy tree is part of an ML model that predicts key component failures in plastics manufacturing equipment and identifies the most critical parts suppliers. It automatically filters data to expose the top 20 parts with the highest failure rates, and those components with the highest repair rates are flagged as “change suppliers.”



This strategy tree is part of the same ML model and uses three key performance indicators: Total Sales, Total Repairs, and Repair Percentage. With this information, engineers in the plant take action to prevent unscheduled downtime.

## Asset Management

- Digital asset marketing: assets at risk, channel marketing, buyer prediction, seller prediction
- Natural Language Processing (NLP): transcript segmentation
- Workplace investment: transfer out

## Finance and Banking

- Risk analytics: scorecards, fraud detection, debt recovery, regulatory risk
- Customer analytics: acquisition, retention, next best action, sentiment analysis, micro-segmentation.

## Insurance

- Risk analytics: technical and competitive pricing, claims management, application and claim fraud, cyber exposure, regulatory reporting
- Customer analytics: next best action, cross-sell and up-sell, sentiment analysis, micro-segmentation

## Telecommunications

- Risk analytics: scorecards, affordability, fraud detection, debt recovery, pricing
- Customer analytics: retention, next best action, sentiment analysis, micro-segmentation

## Manufacturing

- Tool condition monitoring, anomaly detection in product systems, machine failure prediction, root cause analysis, service pack optimization, warranty risk profile analysis, price optimization, supply chain risk management, product performance analysis, customer service and support optimization

## Sales and Marketing

- Customer segmentation, customer behavior analysis, recommendation engines, campaign optimization

## Healthcare

- Resource planning and utilization, medical research, patient risk identification, outcome prediction, infectious disease planning and prediction, behavior adjustment

# BUILDING BETTER ML SOLUTIONS

In the years ahead, ML will deliver profound, widespread benefits. But the technology isn't a magic bullet – it, like anything, has its limits. Perhaps most notably, this includes the need for data of sufficient quality in sufficient quantity.

## **The Importance of Ethics and Accountability**

There are also wider concerns surrounding the ethical and responsible use of ML. Not all the hype surrounding ML and its near-relation AI has been positive. Some ML- and AI-powered systems, including facial recognition, have been accused of inherent racial and gender bias. In the social media sphere, people have raised concerns about the “echo chamber effect,” where algorithms relentlessly point users toward accounts that reinforce, rather than challenge, existing viewpoints. Perhaps most alarmingly, some observers even predict that “runaway” AI and ML will eventually create systems that are beyond human understanding and control.

Whatever the merits of these debates, there's no doubt ML should always operate within a framework that prioritizes responsibility, transparency, and accountability. If ML is being used to predict outcomes and make decisions automatically, it's vital that the people acting on these results can understand and explain how they were reached. What must be avoided is the so-called “black box” approach, where systems produce results that can't easily be interpreted.

To achieve these goals, ML developers and designers can rely on industry-recognized standards. Equally, the popularity of decision trees and scorecards is being driven by the ease with which they can be understood – not only by data scientists, but non-specialists too. What's more, modern software tools are now making the creation and visualization of decision trees fast and straightforward.

Again, this accessibility extends well beyond traditional ML experts. The latest software is also automating the process of document creation, which is integral to strategies and policies for responsible ML.

Serba Dinamik is a multi-billion-dollar supplier of power generation solutions for the oil and gas industry. The company uses ML to implement an advanced predictive maintenance system for turbines that power critical systems on offshore platforms. The solution provides the firm's engineers with on-demand visibility of turbine performance and any future variances, anomalies, and outliers in the sensor data that might predict impending component failure. The system also supports optimized decision-making regarding maintenance scheduling, minimizing the impact on rig operations.

[Read the full story here.](#)

#### **The Democratization of ML**

In fact, responsible ML is going hand in hand with democratization. By offering an ever-wider array of employees the opportunity to explore, design, develop, and deploy ML solutions, modern software enables greater accountability and explainability.

There are also compelling commercial benefits to democratization. Data scientists are in high demand and short supply. However, empowering a new cohort of “citizen data scientists” is about more than saving on the wage bill. At the same time, automating and accelerating the more routine and time-consuming elements of an ML project gives data scientists more freedom to focus on areas where they can add more value. Notably, this includes the opportunity to retain and reuse legacy code (including the languages of SAS, Python, and R) to speed the development journey and subsequent return on investment.

## QuantHub estimates that there is a shortfall of 250,000 data scientists in the US job market.

[Read More](#)

In the media, there’s sometimes a tendency to frame ML and AI in the world of science fiction, but it’s an increasingly accessible and practical tool that supports better and faster decision-making. Ultimately, the goal of these technologies – like any technology – is to make a better, safer, greener future where people can live, play, and work smarter so they can enjoy better lives.

# GLOSSARY

## Algorithm

In the context of data science, an algorithm is a series of repeatable steps or instructions. Data scientists can develop and implement algorithms using many different tools and methods, including coding, or within an ML software system.

## Artificial Intelligence (AI)

AI is something of a catch-all term, but at its core is a technology that can extract useful insights and identify patterns in large data sets and often produce predictions based on that data. The concept ultimately aims to develop systems that simulate and emulate human thought processes. ML is generally regarded as a subset of AI, and the means by which a computer system achieves the goal of artificial intelligence.

## Big Data

Another buzzword without a widely agreed definition. Put simply, a set of data that is too large to fit on a single computer.

## Low-Code/No-Code

The ability to develop sophisticated software applications without writing and debugging programming code.

## Data Cleansing

The process of removing or modifying data that is incorrect, incomplete, irrelevant, duplicated or improperly formatted. A crucial step in the data preparation process, as dirty data can lead to incorrect conclusions and predictions in ML models.

## Data Preparation

The process of gathering, combining, structuring, and organizing data so it can be used in ML.

## Data Visualization

Numeric data that is displayed in graphic form.

## Data Modeling

In the context of data analytics, modeling involves building sets of algorithms that can make accurate predictions on future events, based on historical data.

## Decision Tree

A graphical depiction of decisions (or nodes) in which each potential outcome is directed to a new branch of the tree.

## Dependent Variable

The predicted outcome that an ML model is designed to identify. Also known as the target variable.

## Explainable AI

Systems that enable users to understand how an AI or ML system produced its outputs.

**Independent Variable**

Data that is used to predict an outcome (dependent variable)

**Machine Learning (ML)**

ML involves building algorithms to create a model based on sample data (known as training data) to make predictions or decisions.

**Predictive Analytics**

ML that is used to predict outcomes.

**Predictive Power**

A measure of the influence that an independent variable has on the dependent variable.

**Prescriptive Analytics**

A type of ML model that is used to guide or automate decision making.

**Supervised Learning**

Also known as predictive analytics, supervised learning typically uses labeled or classified data to predict an outcome (the dependent variable).

**Target Variable**

The predicted outcome that an ML model is designed to identify. Also known as the dependent variable.

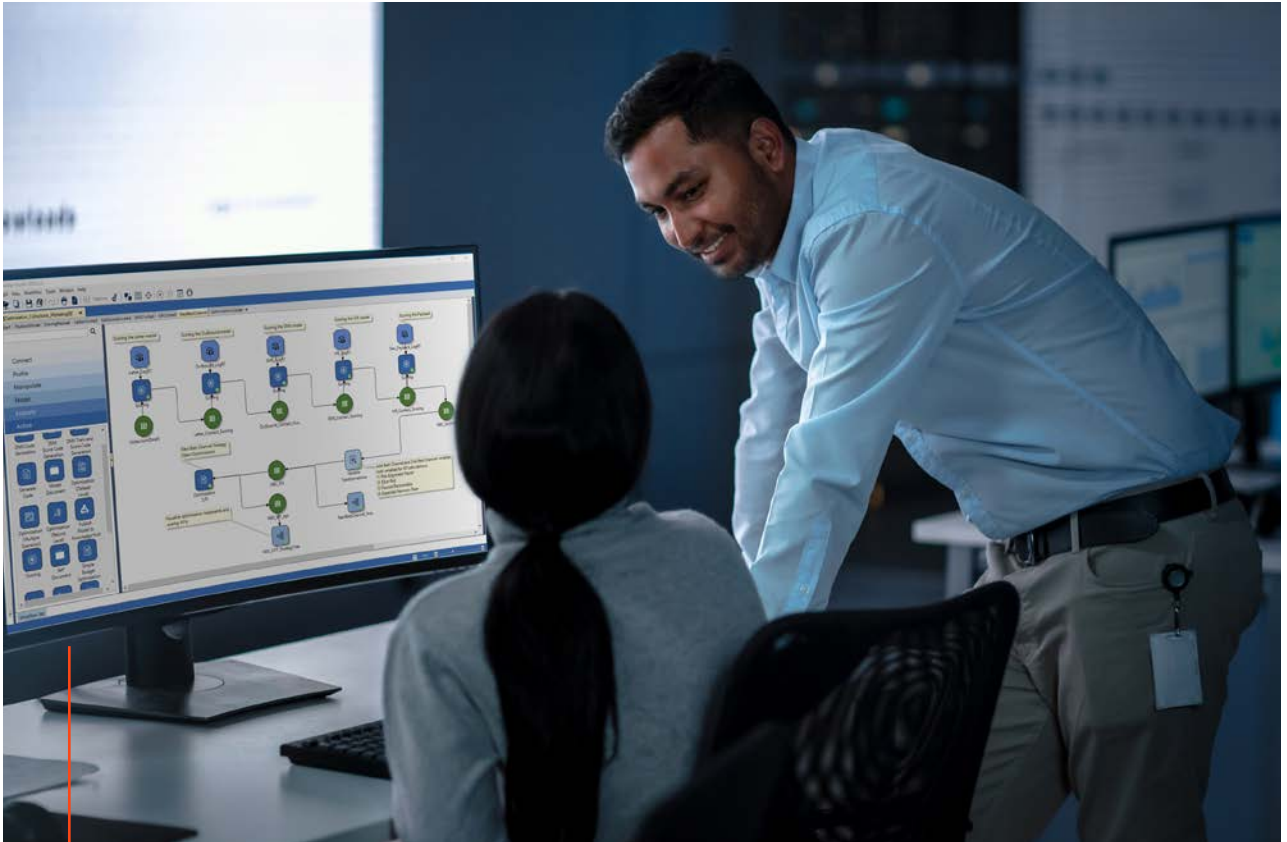
**Training Data**

The historical data used to train an ML model to predict outcomes or identify patterns.

**Unsupervised Learning**

The term used when ML is employed to group together data with similar characteristics (rather than predict outcomes). Typically uses unlabeled or unclassified data. In contrast to supervised learning, it does not have a dependent variable.





Thanks to the emergence of low-code/no-code software, ML is no longer the solely the preserve of programmers and data scientists with specialized training. More of us can – and will – be involved in the development and deployment of practical ML solutions.

# RESOURCES

[Real-Time Credit Scoring: Reduce Approval Times, Increase Loan Numbers, Improve Borrower Experience](#)

[Analytics for Heavy Equipment](#)

[Connected Products Deliver Big ROI](#)

[Altair® Knowledge Studio® Speeds Stamping Process Selection, Increasing First-time-through Rates](#)

[A Smart Solution for a Circular Economy](#)

[Explore how an alternative SAS language environment can enable customers to maximize their infrastructure and software investments](#)

[AMD's Pervasive AI Strategy - Past, Present, and Future](#)

[Panel Discussion: Fraud, Risk and Regulatory Management - The Way Forward](#)

[Intel® oneAPI AI Analytics Toolkit](#)

[Google Partner Presentation: Neural Architecture Search \(NAS\)](#)

[Artificial Intelligence Driven Product Development - How Mahindra Enabled Artificial Intelligence for Product Design Insights](#)

[CTO Insights: How Edge-AI Impacts Enterprise Transformation and Business Outcomes](#)

[AI for Unstructured Data in Banking](#)

[Prodrive - Using Data Analytics to Enhance Race Performance](#)

[Reimagine Finance Transformation to Achieve Operational Excellence](#)

[Thrive with an AI-First Strategy - Today and Tomorrow](#)

[Journey to AI](#)

[Human Centered AI for the Enterprise](#)

[New Enterprise AI Solution: Accelerate Innovation and Foster a Data-Driven Culture](#)

[AI-Powered Product Design](#)

[Democratizing an Industry through AI, by The Electric Storage Company](#)

[The Demand for AI in the Public Sector](#)

[The Future of AI in Retail](#)

---

Altair is a global leader in computational science and artificial intelligence (AI) that provides software and cloud solutions in simulation, high-performance computing (HPC), data analytics, and AI. Altair enables organizations across all industries to compete more effectively and drive smarter decisions in an increasingly connected world – all while creating a greener, more sustainable future.

To learn more, please visit [altair.com/data-potential](https://altair.com/data-potential)