White Paper

# An Enterprise Architect's Guide to Machine Learning


## Content



*"We are in the early stages of a 10-year cycle which machine learning is morphing from a lab curiosity to a rich, pervasive technology value-add." - Phillip Harpur, Technology Analyst.*

Over 90% of the data in the world today has been created in the last two years alone. The current output of data is roughly 2.5 quintillion bytes a day. As a whole, 49.8% of the population has an internet connection. On average, the US alone spits out 2,657,700 gigabytes of Internet data every minute.[1]

Big data has the potential to make organizations stronger, smarter, and more productive by providing the basis to make impactful, data driven decisions. Organizations that let data fuel their business decisions are more successful.

Previously, business leaders proposed improvement plans and processes gathered from often untested hypotheses, intuition, gut feelings, and trial and error. Enterprise Architects of the past spent a great deal of time modeling extensive, rigid architecture models. These models took months to complete and were often created with aged data. More often than not, the situations that the architecture model addressed changed during planning and implementation. During this current age of rapid digital transformation, rigid architectures and obsolete data models have proven to be obsolete. Machine learning algorithms allow business leaders to remove the guesswork from important business decisions. Insights gained from machine learning tasks empower organizations to make smarter predictions by continually learning and adapting their models based on real-time data.

In order to bring measurable value to their firms, Enterprise Architects of

[1] http://www.iflscience.com/technology/how-much-data-does-the-world-generate-every-minute/


Tomorrow must understand, utilize, and evangelize the latest technologies driving the industry. These current trends include predictive analytics, deep learning, prescriptive analytics, and machine learning. All of these trends use current data to make predictions about unknown future events.

This white paper provides an overview of the approaches to machine learning, the steps to creating an algorithm, and how Enterprise Architects can use the insights generated from machine learning to plan future business decisions. This white paper also includes recommendations based on LeanIX's experience with machine learning and outlines steps to prepare your enterprise for machine learning.

## WHAT IS MACHINE LEARNING?

Machine learning is an application of artificial intelligence that provides systems with the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. The process of learning begins with observations of data, such as examples, direct experience, or instruction, in order to look for patterns in the data and make better decisions in the future based on the examples provided. The primary aim is to allow the computers to learn automatically without human intervention or assistance and adjust decisions accordingly.

## WHY USE MACHINE LEARNING IN YOUR ENTERPRISE

*"Any sufficiently advanced technology is indistinguishable from magic." - Clarke's Third Law*

Machine learning technology is not only relevant for internet giants, but also relevant for traditional firms like manufacturing, construction, insurance, finance, and any other industry.

For example, artificial intelligence offers innovative solutions in manufacturing like MAX - a cloud based predictive maintenance service for elevators. MAX collects data such as door movements, trips, car calls, and error codes. This data is sent to the cloud where unique algorithms analyze it for patterns and compute the equipment's operation and the estimated remaining lifetime of its components. Precise and predictive diagnostics are delivered to technicians in real time, indicating when and where intervention is required. [2]

Machine learning algorithms affect and benefit your life in many undetectable ways. They currently identify and eliminate spam from your inbox, select images for social media posts, curate content to appear in your social media timelines, monitor your credit score and help to prevent fraud.

At LeanIX, we assess the usage of our SaaS product and help customers to optimally benefit from our EA management solution. Amazon uses machine learning algorithms to show you trustworthy reviews, Paypal uses it for transactional fraud detection, hotels benefit from real time customer-specific strategic pricing, and top marketing firms use machine learning algorithms to gauge customer sentiments.

## THE CONNECTION BETWEEN ENTERPRISE ARCHITECTURE AND MACHINE LEARNING

*"The goal is to turn data into information, and information into insight." – Carly Fiorina, former CEO of Hewlett Packard.*

In the 2017 Economist Intelligence Unit report, Artificial Intelligence in the Real World, 75% of more than 200 business executives surveyed said artificial intelligence will be actively implemented in their companies within the next three years [3].

Our goal at LeanIX is to use current cutting-edge technologies to make Enterprise Architects' daily tasks easier. Presently, the line from machine learning to constructive EA practices can be ambiguous. However, Enterprise Architects will be required to plan and implement AI-like services in the future, therefore it is imperative to become familiar with the link between EA and AI.

[2] https://max.thyssenkrupp-elevator.com/assets/pdf/TK-Elevator-MAX-Brochure_EN.pdf
[3] https://www.eiuperspectives.economist.com/sites/default/files/Artificial_intelligence_in_the_real_world_1.pdf

For decades, simple reports of metrics (KPIs) produced enough data to make long-term decisions. Data derived from machine learning has the capacity to gather much deeper insights than simple KPIs.

For example - while selecting global standard applications or technology for their teams, Enterprise Architects can use machine learning algorithms to gather and test user generated data to choose the best application for the enterprise. A machine learning pipeline would query the number of logins on various applications, determine the length of time spent on said applications, gauge user satisfaction, and identify which information is relevant for the decision. Enterprise Architects will be able to tell which business applications provide the most value to their teams, which apps are redundant, and which apps go unused. Using user generated data, EAs can select one global standard app for each business capability, shut down the use of redundant apps, and cut IT costs.

## REFINING USER EXPERIENCE WITH MACHINE LEARNING

From machine learning tasks, your company can constantly improve your product or service. Identifying at-risk customers early and offering personalized content or customer care to educate them on best practices increases customer satisfaction and retention, and enhances their overall experience with your company.

The virtuous circle of artificial intelligence refers to a complex chain of events that reinforce themselves through a feedback loop. For EAs, the virtuous circle starts with the product. Users generate data while using the product. This data is fed into the machine learning algorithm. The results produced from the machine can be applied to improve the product (see figure 1).

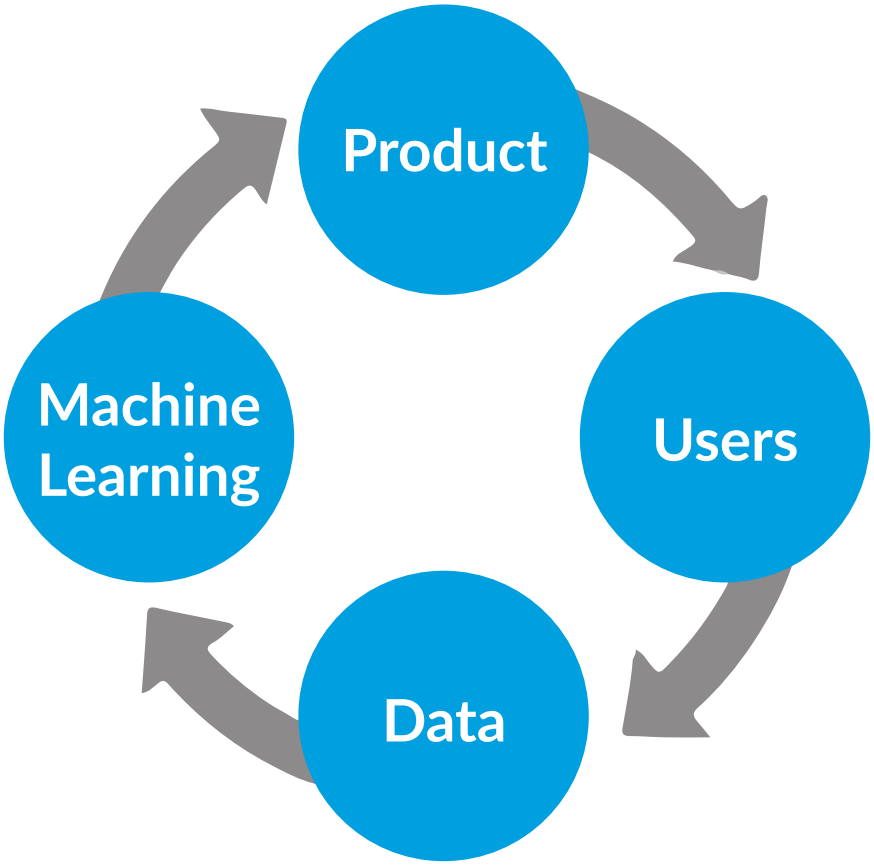


Figure 1: Virtuous circle of AI: User generated data helps to continuously improve your product.


## BROADLY, THERE ARE THREE TYPES OF MACHINE LEARNING

### 1. Reinforcement Learning

Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions.

### 2. Supervised Learning

This algorithm consists of a target or outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Using these sets of variables, we generate a function that maps inputs to desired outputs. The training process continues until the model achieves the desired level of accuracy on the training data.

### 3. Unsupervised Learning

As opposed to 1. and 2. this class if methods works without reinforcement or supervision. Most prominently are the algorithms for dimensionality reduction and segmentation or clustering.

In this white paper, we will focus on supervised learning. Supervised learning best represents the idea of artificial intelligence and is the more commonly used among the three.

## SUPERVISED LEARNING IN DEPTH

When performing supervised learning, think of yourself as the teacher and the machine learning algorithm as your student. You explain the topic to your student (i.e. you show the algorithm data and their correct classification), and the algorithm "learns". In the next step, you test your algorithm by showing it data to which you know the solution but your student is unaware. Depending on how your student performed, she must learn more data or you need to tune your learning method (e.g. tune the algorithm's parameters). As with most teachers, you probably do not have one single student, but many students - your students here are possibly different machine learning algorithms, with e.g. different underlying functions: linear, polynomial, etc.

If you wish to know which of your students is suited best for the task at hand, you need to evaluate them again (like in a final exam, or in this case a "validation").

Like in school, the students do not know the exact questions, but they do know which topic the final exam will be on, and the type of questions being asked. They are required to transfer their learned "knowledge" to unknown samples - which we sometimes call intelligence, hence the wording "artificial intelligence".

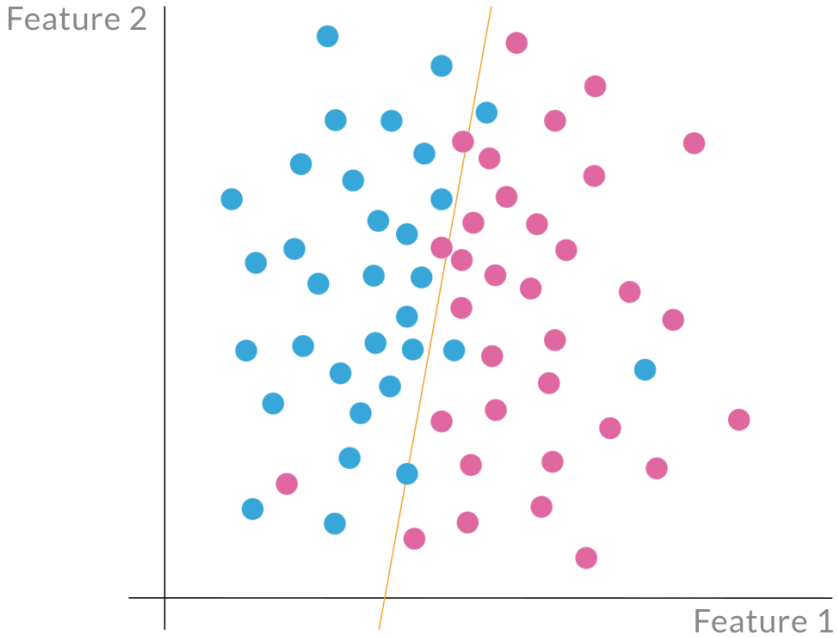


Figure 2: Clustered data after dimension reduction. The orange line shows a possible classifier.

## SaaS Example

In a simple example, every customer is represented as a sample of a high-dimensional feature space. In order to retain happy customers or users, you need to understand which customers/users are satisfied customers/users (or any other comparable variable). If you have a history with your customers, you can tell which customers are satisfied and which customers are not through various statistics including churn rate, unfavorable net promoter scores, length of use etc.

On that data, you can let your student/algorithm train and then have her identify the current customer satisfaction for each user/customer. There are various known methods for this classification task such as support vector machines, random forests, decision trees, k-nearest-neighbor algorithms etc. which are all known to produce reasonably good results for a task like this.
If the correct features (variables) are selected, customers that have a similar classification build clusters in the feature space (see fig. 2).

After acquiring data for your supervised learning tasks, select the correct features for the needs of your project, and choose a suitable machine learning algorithm. The machine learning algorithm will process the inputted data and produce results that display customer clusters grouped together based on common characteristics. Based on these clusters, one is able to make reliable predictions on future data that helps your company to steer business decisions in the right direction (see figure 3).

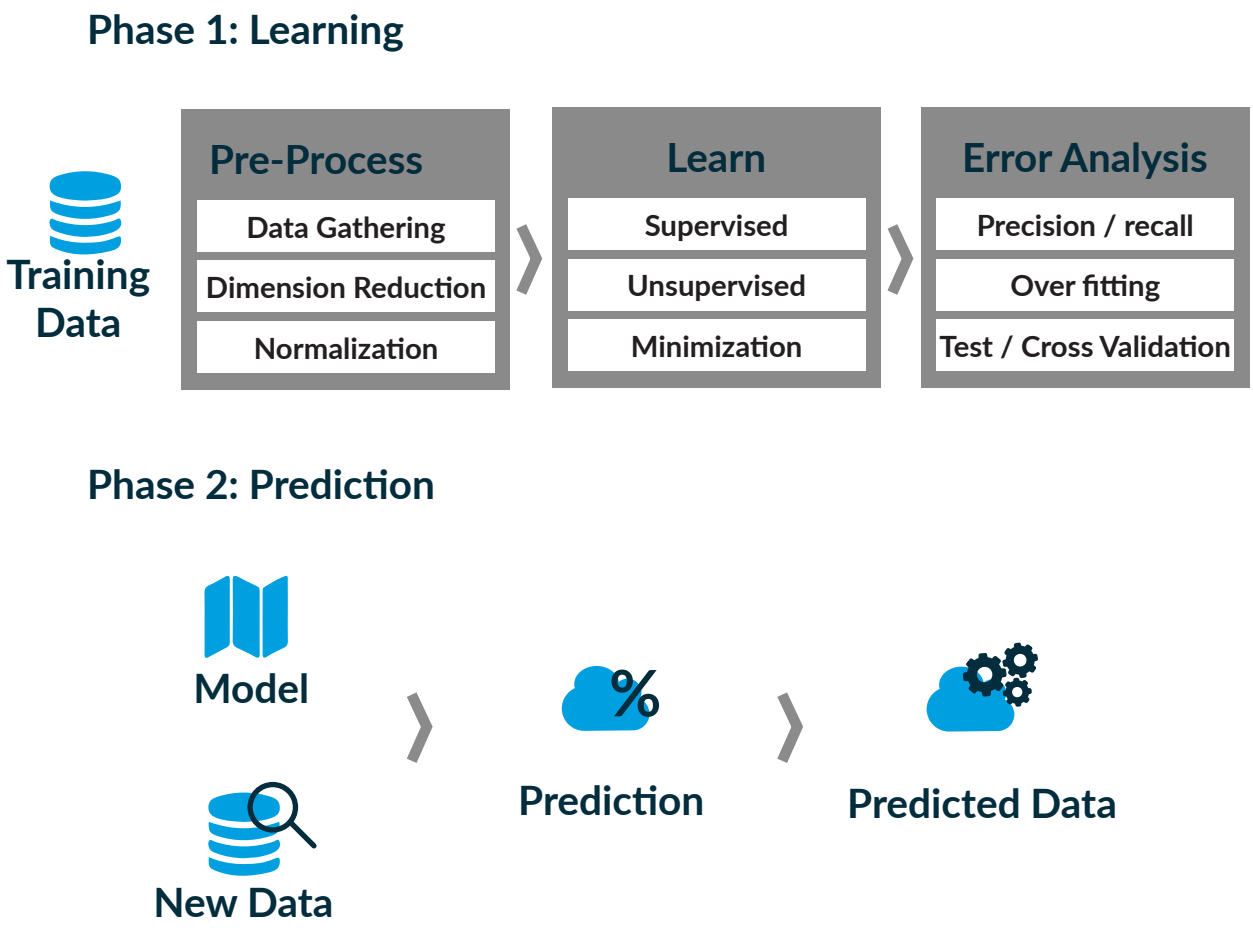


Figure 3: Machine Learning Process.


# HOW TO CREATE A MACHINE LEARNING ALGORITHM IN 6 STEPS

Although it is not particularly necessary for Enterprise Architects to become junior data scientists, Enterprise Architects looking to bring measurable change to their enterprises must have a general knowledge of trending subjects in order to consult teams on best practices. Below are the steps to creating a machine learning algorithm.

## 1. Gather the appropriate data

Determine strong variables that you would like to later query for, including log-in frequencies, amount of distinct users, amount of power users, time since the last contact, net promoter score from last feedback etc. Get creative here and think about your business. For example, if your company creates multimedia content for customers, think of incorporating vital statistics about the content, including word counts and post reach. If your company produces marmalade, include the history of the types of jam available, and the average amount purchased in one transaction.

## 2. Create interfaces between your connected systems that store your data

To extract meaningful value from large sets of data, your enterprise needs many tools and capabilities - analytics, algorithms, and big data processing capabilities. Consider the following: A microservice framework, cloud based servers, platform as a service (PaaS), and containerization. It is important to have access to the most up-to-date information from all relevant systems in order to extract the most meaningful features. You also need to establish a common key for your customer among all the systems. See lessons learned section for more information.

## 3. Start simple

A simple database management system will suffice for most projects in the beginning (e.g. Amazon RDS, PostgreSQL, or MySQL). You will require such a database and it should be independent of your production environment.

## 4. Prepare and transform the data

Most algorithms require the input variables to not be dependent, therefore you must transform your input data. Methods to do that would include Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) or Quadratic Discriminant Analysis (QDA). Eliminating dependent variables helps preserve the quality of your data, which is a prerequisite to many methods. This step will improve the accuracy, of your model for later stages.

If you are not into transforming your data, you can still use Random Forests, which do not require you to have uncorrelated inputs (but they tend to perform better on uncorrelated data).

## 5. Choose a suitable machine learning algorithm

Inform yourself about machine learning algorithms and which suits your challenge. Commonly used machine learning algorithms that can be applied to almost any data problem:

- Linear Regression
- Logistic Regression
- Decision Tree
- SVM
- Naive Bayes
- KNN
- K-Means
- Random Forest
- Dimensionality Reduction Algorithms
- Gradient Boost & Adaboost

There are numerous methods and algorithms to choose from. A word of advice: Start simple and get more complicated step by step.

## 6. Train, test, and re-evaluate the models

This includes dividing the data into three sets for training, testing, and validating. The training stage is


used to train the initial machine learning model. The testing stage is for evaluating the trained model: How does the model perform on data which is yet unknown to it? During the testing phase, it is important to calculate accuracy, precision, and recall.

Use a confusion matrix, or an error matrix, which is a specific table layout that allows visualization of the performance of a supervised learning algorithm.

Validation stage - If you have trained different models via different machine learning algorithms, you can pit them against each other by performing the same analysis of accuracy, precision, and recall on the validation set.

In order to run the accuracy tests, it is important to have sufficient data to analyze - a few hundred customers is the minimum.

## TOP MACHINE LEARNING LESSONS LEARNED FROM LEANIX

The following are the key lessons which we have learned from our experience with employing machine learning techniques:

### If using many services, especially microservices, begin by choosing a unique identifier that serves across all attributing platforms

Having one strong identifier to be used across different platforms will save your team from dealing with inconsistent and inaccessible data sources. LeanIX uses many different services in our daily business activities. If a strong identifier such as a unique customer ID is not designated, it is easy to lose track of customers across different services.

For example - If customer John Doe is identified as #23482558 in Zendesk, JDoe in Salesforce, and jonathanreginald@gmail.com in your email marketing software, his data will be lost, inaccessible, or too inconsistent to reference during machine learning tasks. Even worse: you will be required to spend a lot of time and resources cross-referencing and making your data consistent.

### Have a sharp focus on data quality

Data quality refers to the condition of a set of values of qualitative or quantitative variables. Data is considered “high quality” if it fits for its intended use in operations, decision making, and planning.

The International Organization for Standardization (ISO) sets the global standard for Data Quality and Enterprise Master Data. When setting up your database for machine learning projects, you should consider the 15 characteristics defined in the ISO 25012 requirements (see figure 4)

Questions to ask while setting up your machine learning software: Is the data accessible? Understandable? Complete? It is not necessary to achieve all 15 characteristics, but it is imperative to think about the quality of data while setting up your machine learning database. As an indirect result, your IT landscape will benefit from the improved standards of data quality as well.

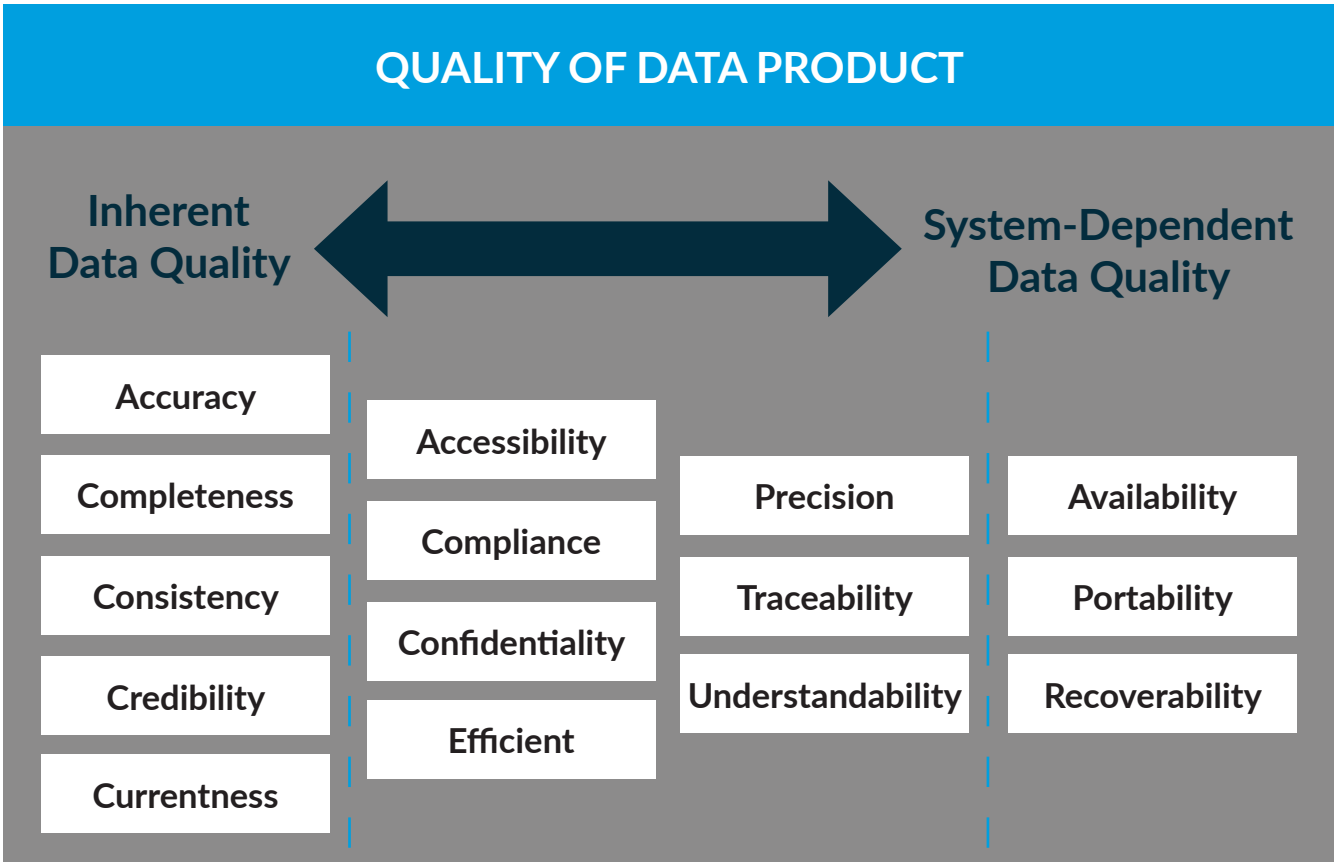


Figure 4: ISO25012 defines a general data quality model.

Figure 4 displays the ISO global standard for data quality. Inherent data quality refers to the degree to which quality characteristics of data have the inherent potential to satisfy stated and implied needs when data is used under specified conditions.

System dependent data quality refers to the degree to which data quality is reached and preserved within a computer system when data is used under specified conditions.

You can control the quality of your system dependent characteristics and you might require some changes in the paradigms of your company in order to have an impact on the data inherent quality characteristics.

## Bias in, bias out

To avoid self-fulfilling prophecies in machine learning, be aware of your biases when inputting the data and selecting variables. It is possible to input biased data without realizing it. Techniques for weighting existing data to compensate for unknown bias are randomization, bootstrap sampling, and including more data to the model. Keep in mind that some predictions may be wrong and in turn detrimental to your bottom line. Losing just 5% of your customer base due to systematic errors in your model can be costly.

## Correlation does not infer causation

Finding correlation between data samples does not necessarily infer dependency or causation. Two variables that do not correlate prove that there is no significant relationship data, but the converse is not necessarily true. Correlation is generally a measure of the strength of the relationship between two variables but does not directly infer causation.


A strong correlation between two data samples could mean four things:

1. A is caused by B.
2. B is caused by A.
3. A and B have a common cause.
4. The correlation is mere coincidence (if you believe this is the case, then getting more data will resolve your problem and the correlation should disappear. If it doesn't, then you should analyze critically).

Tip: It is important to think deeper when stumbling upon correlated data. If you find correlations as a result of your analysis think of what that could mean in terms of your company's business and analyze accordingly.

## Big data analytics require a lot of processing power

Big data is comparable to a voluminous, fast-moving and complex waterfall, and requires a specified supporting system. For some projects, you will have scripts that run for hours or days. To prepare for the exponential velocity of data, it is essential to separate your productive systems from your machine learning querying systems.

Tip: Refrain from using your productive systems to carry out any machine learning queries. Establish a separate cloud-based database for your data science and machine learning tasks, so that the queries and calculations do not impact the product or service that you provide to your customers.

For instance, LeanIX has a business intelligence (BI) and artificial intelligence (AI) server in the cloud which is dedicated to machine learning tasks. There our data scientists and analysts can do their work without affecting our SaaS solution.

## Manage expectations of machine learning and artificial intelligence

Popular culture and increasing hype leads us to believe exaggerated ideas about the potential of artificial intelligence and machine learning. The first referenced AI-based topic is usually the heavily hyped self-driving car popularized by Google's parent company, Alphabet. Although this car is a visionary attribute to the capacity of artificial intelligence, your firm most likely will not be using artificial intelligence in the same ways.

Trained machines perform a very specialized task and carry out that particular task well while failing to perform any other task at all. We are a long way from the same AI driving a car, playing chess, supporting an enterprise's customer support and analyzing your customer data.

Make sure your managers are aware of the limitations and have reasonable expectations of machine learning and artificial intelligence. Expert systems can be programmed to be 'better than humans' for very specialized tasks (e.g. deep learning), but business and enterprise stakeholders should not view machine learning technologies as a magical solution for every minute decision. Don't be the person/EA that promises a solution the machine cannot possibly keep.

## When choosing an algorithm: start simple

For most projects, a simple linear regression or classification will suffice. Avoid starting with complex algorithms like k-nearest neighbors with dynamic time warping on a custom feature space with a custom metric. The sheer technical sophistication of the method will blind you from the true nature of your data. Start with a simple algorithm and gradually progress to more complex algorithms. Strive for a robust and consistent solution. A few things are more frustrating than having a model that predicts very differently on slightly changed input data.

# HOW TO PREPARE YOUR COMPANY FOR MACHINE LEARNING

The line from machine learning to beneficial Enterprise Architects practices can be blurry. Our goal at LeanIX is to use current cutting-edge technologies to make Enterprise Architects' daily tasks easier.

Tech-savvy Enterprise Architects of tomorrow see the value in staying abreast of all of the current trends, and


at the moment, machine learning, artificial intelligence, and deep learning are hot topics. Having the basic knowledge of the possibilities of machine learning allows you to set the framework to benefit from the innovative technique.

Forrester predicts that corporate investment in artificial intelligence will triple in 2017, becoming a $100 billion market by 2025[4]. Preparing for machine learning requires foresight, clear business goals, retraining or hiring of new team members, and having a clear view of your IT landscape.

- First, decide how you want to use machine learning technologies: to automate decision processes, predict user behavior, strengthen customer interactions, improve marketing, define the price of a product in a fluctuating market, approve or deny credit applications, or detect customers with a high risk for churn etc.

- Later, assess the skill set and framework of your company and prepare it for AI. Corporations wishing to spearhead machine learning should place a sharp emphasis on math, statistics, and science. Consider hiring a data scientist - better yet a team of data scientists. They get better when challenging each other.

- Your programmers may be comfortable with other programming languages, but R and Python are the most used language for machine learning. R is the most popular platform for applied machine learning and Python is a general language with a high complexity-performance trade off and has a full suite of tools for productionizing machine learning.

- Lastly, create an AI strategy based on the needs of your enterprise. You may be required to change some paradigms in order to align your company with the benefits and risks of artificial intelligence.

## SUMMARY

During this period of rapid digital transformation, companies that do not innovate tend to lose their market share to eager start-ups entering the market. Merrill Lynch predicts that the global market for artificial intelligence and robotics will be just under $153 billion by 2020, and some industries will experience up to a 30% productivity increase through the use of those technologies alone. As machine learning algorithms mimic biological processes and continuously adapt to improve themselves over time, they provide accurate and effective data to increase productivity at your firm. Machine learning algorithms are a strong tool that may offer Enterprise Architects an abundance of opportunities to continually improve and strengthen business strategies

[4] https://go.forrester.com/wp-content/uploads/Forrester_Predictions_2017_-Artificial_Intelligence_Will_Drive_The_Insights_Revolution.pdf


## About LeanIX

LeanIX offers a Software-as-a-Service (SaaS) for Enterprise Architecture (EA), which enables organizations to take faster, data-driven decisions for their IT landscape. More than 80 leading brands such as adidas, DHL, Merck, Vodafone, and Zalando use the innovative solution worldwide. Users of LeanIX gain insights on how to organize and leverage their IT landscape to increase competitiveness and enable innovation going forward. LeanIX addresses the frequent problem that the required information about the IT landscape is missing, outdated, or difficult to analyze. Use cases include application rationalization, technology risk management, and the shift from monolithic architectures to microservices. LeanIX was founded in 2012 by Jörg Beyer and André Christ. The company's headquarter is in Bonn, Germany, with offices in Boston, Massachusetts, and Houston, Texas. A wide network of partners provides support in America, Europe, and Australia.

**LeanIX GmbH**
Bonn
Germany

Tel: +49 (0) 228 2862992-0
Email: info@leanix.net
www.leanix.net

[21] https://olui2.fs.ml.com/publish/content/application/pdf/GWMOL/RIC-Report-January-2016.pdf