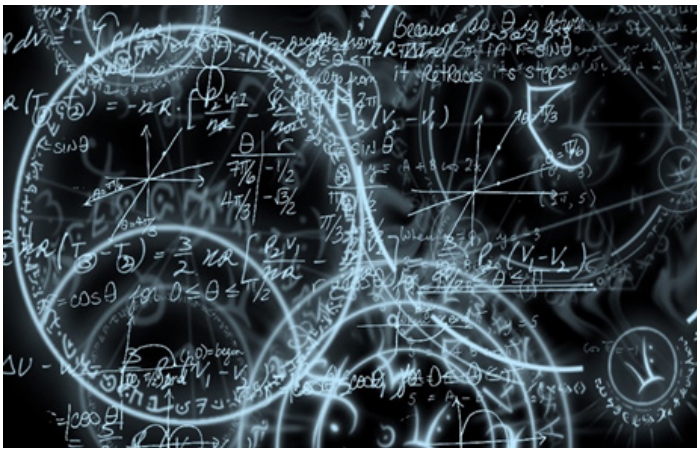
# MACHINE LEARNING

## FOR LEAD QUALIFICATION

Vijilent | 2016

# INTRODUCTION

Just the words, "machine learning" often brings to mind thoughts of the deviant self-learning computer, Hal, in the movie, 2001: A Space Odyssey or for a younger generation, cybernetic organisms, the Cyborgs, in the Terminator movies. The future is now and machine learning is no longer a fantasy. Fortunately, machines are not yet ready to take over the world, but they have infiltrated your office.



So what do you as a Lead Generation Manager need to know about machine learning? This guide will introduce you to machine learning, provide an example on why you as a Lead Gen Manager need to know about it and provides some suggestions on where to start using ML in your company. We'll leave you with the top 9 things a Lead Gen Manager should know about machine learning.

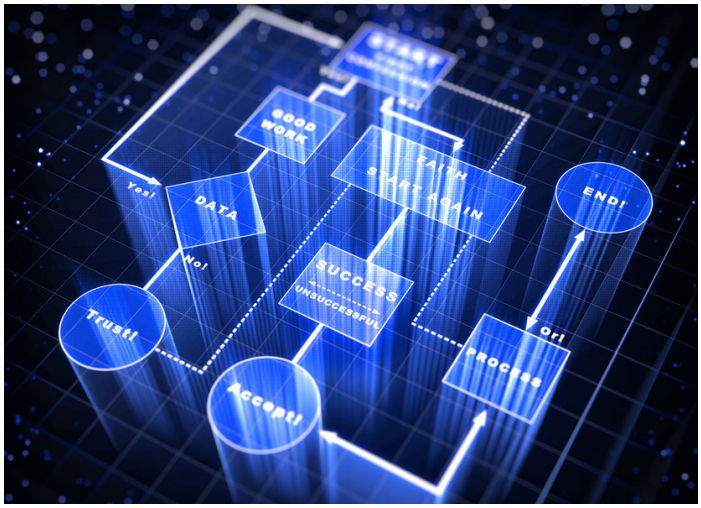
# MACHINE LEARNING DRIVES BUSINESS INTELLIGENCE

Big data, data mining, predictive analytics, and machine learning are concepts that are being used more often in business intelligence. These trends are based on the growing abundance of raw data generated, off-the-shelf programs for data management as well as the cheap computational power to crunch data gathered by companies. Googling the term 'machine learning' brings up articles about artificial intelligence, neural networks and deep learning so many companies feel it's not useful for them especially if they don't have data scientists/ statisticians on staff.

This guide shows that any size firm, with or without data scientists/ statisticians can use machine learning to build a deeper understanding of its current and prospective customers. If your firm isn't using ML techniques, you are missing a big opportunity for targeting the best customers for your firm cost effectively. Companies using these methods have reported 10X return on investment.

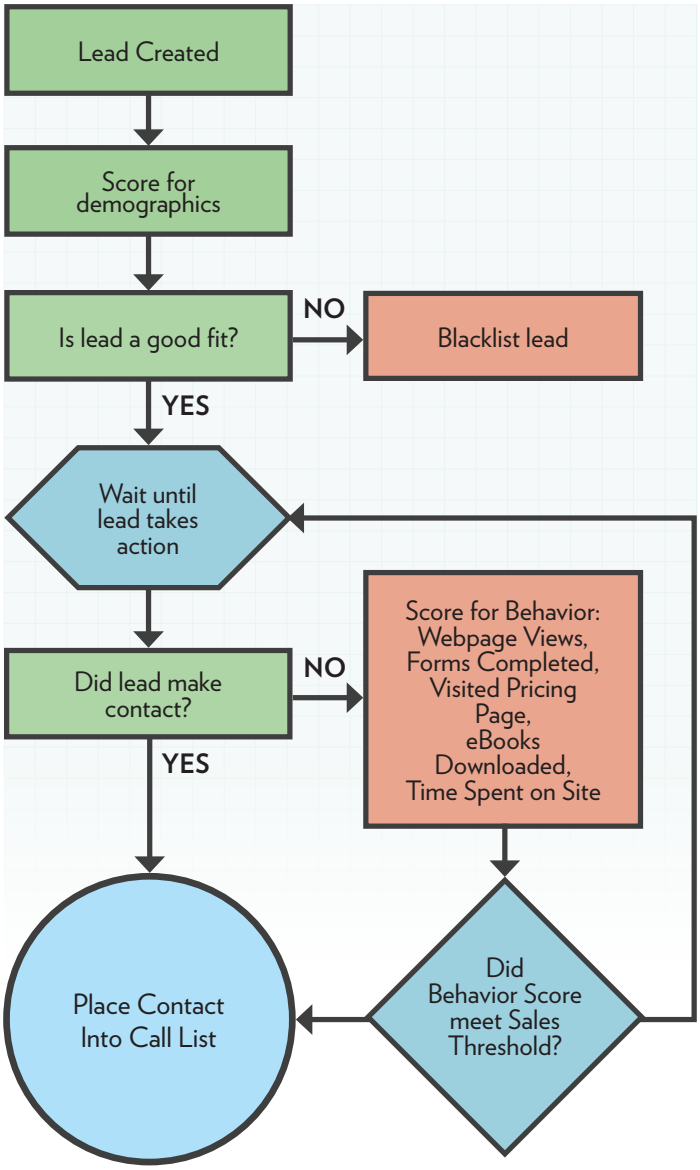
# MACHINE LEARNING DEFINED

Machine learning is a subfield of computer science that focuses on **pattern recognition** for making predictions. It's frequently used in spam filtering, search engines and face recognition. It has other names, such as statistical computing, data mining, optimization, and pattern recognition, but machine learning is the term most frequently used in business intelligence. ML is about constructing algorithms that provide insights that are difficult to observe without statistics. For lead generation it is most frequently used to identify ideal customers and predict probability of purchase intention.




# WHAT'S AN ALGORITHM?

An algorithm (\'al-gə-ˌri-thəm\) is a self-contained step-by-step set of operations to be performed. These ***rules-based instructions*** are followed in order to attain a specific goal. For example, in lead gen, you may have a set of guidelines you use when calling a new




prospect. Here's an example of an algorithm that you may follow. You would have built this type of algorithm/ flowchart based on the historical successes you have had in your lead generation process. You would have gathered input from various people within the organization to produce it. It would likely evolve over time as you gather additional insights or as the market changes over time.

Machine learning works in a similar way in that it takes the historical data you have archived in your CRM platform or similar data sets and evaluates it using statistical analyses. It too needs to be updated regularly when new data become available.

# WHY IS IT CALLED MACHINE LEARNING?

Machine learning algorithms for prediction are built on past examples of successes and failures rather than following strictly static algorithm instructions like the one displayed earlier. Identifying characteristics of your best customers to predict good candidates (success) or poor candidates (failure) is an example of building an algorithm using ML. the success/failure outcome is sometimes referred to as the 'target' variable – what is it you want to achieve – such as a 'yes' purchase decision, a 'yes' to a demo, 'yes' to a sales meeting, etc. The historical data you already own on successes and failures becomes the ***'training' set***.

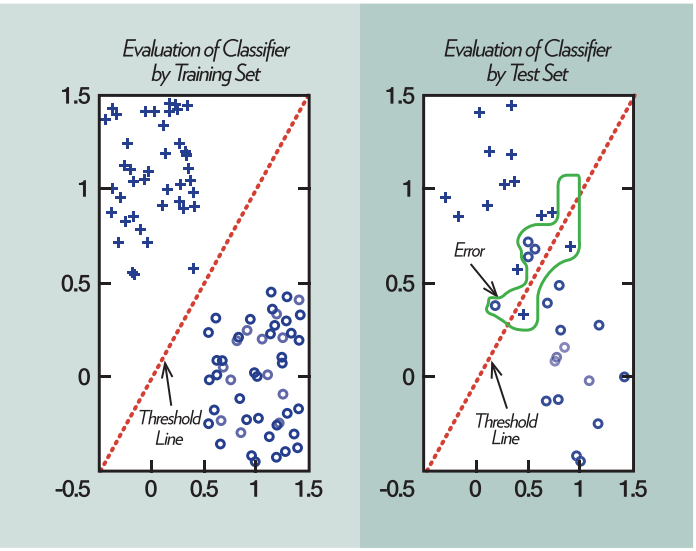
The "learning" comes in when the algorithm is used to accurately predict new, unseen examples/tasks after having learned from the training set. Given a new prospect, you can run the candidate through the ML algorithm to predict whether they are a good or poor candidate for your target goal based on your training set. Since your algorithm has 'learned' from these previous examples, it can provide a prediction on new prospects based on what it has learned about past prospects.

> *"It is impossible for humans to recognize the patterns in a megabyte of data, but it's easy for machine learning tools to find hidden insights."*

Machine learning is its most powerful with large amounts of data to train the algorithm. It is impossible for humans to recognize the patterns in a megabyte of data, but it's easy for machine learning tools to find hidden insights. Does the day of the week matter when you place a call or is the weather a factor? Maybe it matters for one group of customers but not another. When the Golden State Warriors lost the NBA Finals to the Cavaliers, you can bet Cleveland was a hot community for NBA memorabilia. Machine learning can capture these types of insights such as when is it best to reach out to prospective customers/clients.

# WHAT'S THE DIFFERENCE BETWEEN A TRAINING SET AND A TEST SET?

The historical data set from your CRM system are separated into a training set and a test set, usually in an 80/20 split. This split is typically accomplished using random assignment. This step is important as it allows you use the data to build your algorithm and also to evaluate the overall accuracy of the resulting predictive model.




If all the data is used for training, it is called overtraining the data set and greatly reduces the accuracy of prediction for new future data sets. The algorithm developed will only be accurate for that one particular set of data and is not useful for other data sets. The power of building ML algorithms is for predicting probability of completing a target goal for new customers so overtraining is bad. After the algorithm is built, the test set is used to determine how accurate it is in predicting the target outcomes of the data not used in developing the algorithm. It is very rare that the first model developed is the best model. A technique called a support vector machine may produce better results than a random forest technique (both of these are supervised learning techniques which will be discussed next). We would not know which one worked better on the data if we had used all the data in building the algorithm.

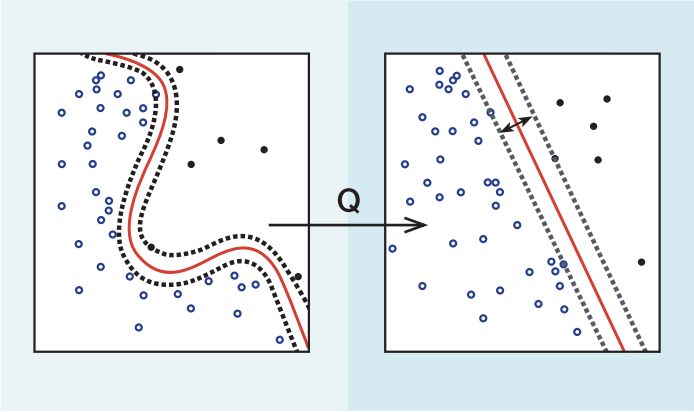
You will find that leveraging machine learning to make predictions is an iterative process that requires a lot of experimentation. The test set allows us to build the best models possible for the greatest predictive capabilities given the available data.

# MACHINE LEARNING TOOLS

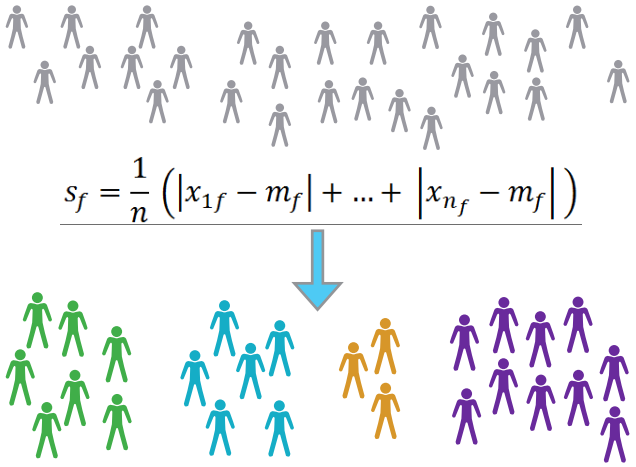
There are a number of different statistical tools a firm can use in machine learning. They are most often categorized as supervised learning vs. unsupervised learning.

***Supervised learning*** looks for patterns in a training set of data that has historical examples of both successes and failures of the phenomenon being observed. For example, a data set showing customers that have previously purchased memorabilia (successes) along with those that have not (failures) and any identifying characteristics of those customers and their purchases is a starting point for supervised learning. The input data will consist of variables that might possibly impact the outcome of the phenomenon.

A historical data set of NBA fans who have and have not purchased memorabilia could be used for indicating which variables are most useful in predicting purchase. For the example of the 2016 NBA finals, an employed, male Cavalier fan, aged 25-50, with children, living in Cleveland, who watches ESPN, and drives a SUV may be more likely to buy memorabilia after the final game than other fans. Popular methods of supervised learning include Support Vector Machines, Mutual Information, Naive Bayes, Logistic Regression, Decision Trees and Random Forests. Their principal drawback is that they are as good as what is taught and don't perform as well when there is more randomness in the data.



In ***unsupervised learning*** the data set will not have a record of successes and failures. Unsupervised learning also looks for hidden patterns but with unlabeled data. An example would be if we had a list of NBA fans and did not know if they had previously bought NBA memorabilia. Unsupervised learning would look for those characteristics that fit a particular group and would cluster them together. We would expect the Cavalier fans to more like each other than to a Warrior fan. The morning after the NBA final in 2016, we would target all fans that had similarities to Cavalier fans instead of Warrior fans.




Unsupervised learning encompasses many techniques that seek to summarize and explain the key features of the data. Approaches to unsupervised learning include: K-means clustering, mixture models, hierarchical clustering, anomaly detection, and neural networks. The limitation of unsupervised learning is that there is no error or reward indicators to evaluate how well the algorithm is performing. However, when the historical outcome data is not available, unsupervised learning is a possible solution for improving CPL.

# AN EXAMPLE: IDENTIFYING AN IDEAL CUSTOMER PROFILE

Ideal Customer Profiles (ICPs) are common tools created for reaching the best customers. An ICP gives a structured look at the demographics and psychographics of an individual, purchase intent timing, product features and content that matter most to them, and for B2B companies descriptors about the firm, such as revenues, # of employees, other products being used, etc.

Machine learning can help you create a profile of these customers. Supervised learning (if you have data with the target outcomes) and unsupervised learning can build an ICP based on data instead of intuition.

> *"The ICP can then be used for Lead Scoring, which prioritizes targeted accounts"*

Supervised learning methods can help to identify the importance of the different input variables on the outcome. For example, it could tell you if males are more likely to buy NBA memorabilia compared to females and if the number of children a man has makes a difference on purchase decisions. Unsupervised learning can identify hidden characteristics of an ideal consumer that you may not even be aware of.



The ICP can then be used for Lead Scoring, which prioritizes targeted accounts.

# GETTING STARTED

Whether you are attempting ML on your own or will be working with a vendor that offers ML services, you will need to work through the following steps.

## STEP 1: COLLECTING THE DATA

1. Determine your 'target' outcome. Is it to predict purchase timing? Is it determining an ideal consumer profile? Is it to determine when to place a prospective customer in a drip email marketing campaign?
2. Gather data set. The data is a set of input objects (variable that might impact the target variable) and corresponding outputs (target outcomes). It needs to be representative of your customers. Typically data sets will come from CRM systems
3. Clean the data set. This means removing any unnecessary data, working with missing values and formatting the data for entry into a statistical program.
4. Supplement the CRM data with external data for a more robust algorithm. See Vijilent's Lead Scoring Guidelines here or on our website for more information.

## STEP 2: DEVELOP THE ALGORITHM

1. Build an intuitive algorithm like the one shown earlier in this guide. This will help the person coding the ML to understand the problem at hand. The ML result are unlikely to match your initial intuition but that's the reason we do ML – to find the hidden patterns.
2. Work with the ML data scientist/programmer/statistician to determine the structure of the learning function and corresponding learning algorithm. For example, the data scientist may choose to use support vector machines or decision trees.
3. Run the learning algorithm on the gathered training set.
4. Evaluate the accuracy of the learned function. After parameter adjustment and learning, the performance of the resulting function should be measured on the test set.
5. Fail fast with the algorithms that perform poorly and iterate often.

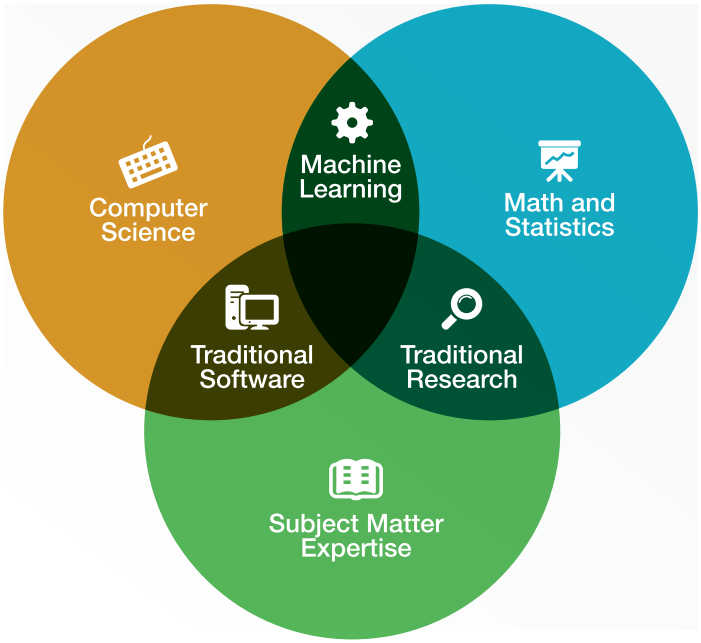
# DATA SCIENTISTS NOT REQUIRED (ALTHOUGH IT DOESN'T HURT).

As noted above, your company may have already used its intrinsic knowledge to develop an intuitive sales funnel algorithm or ideal consumer profile that the sales team follows. This is starting point if you want to think more about where ML may be able to help the company's business intelligence.

Once you have gained an understanding of how you might want to use ML, it's easy to start investigating different models though a variety of statistical packages that provide ML techniques. We list a few of them below in no particular order.

- IBM SPSS Modeler 30-day free trial, https://www.ibm.com/marketplace/cloud/spss-modeler/us/en-us
- SAS Rapid Predictive Modeler as a Microsoft Excel Add-in: http://www.sas.com/en_ae/software/how-to-buy.html
- BigML: Free trial at https://bigml.com
- Salford Predictive Modeler: 60-day free trial, https://www.salford-systems.com/products/spm
- Amazon-ML: free trial at https://aws.amazon.com/machine-learning/
- Google Prediction API: 6-mon free trial, https://cloud.google.com/prediction/

More and more MBAs are gaining an understanding of these packages, so a recent graduate or intern can get you started. If your firm provides the subject matter expertise, statisticians or computer scientists can be partners in your ML insights.




# TOP TEN THINGS TO KNOW ABOUT MACHINE LEARNING

1. Machine Learning is a powerful statistical tool to help predict the probability that a prospect is a good candidate for your product/service.
2. Your firm already has the data it needs to conduct Machine Learning. Find who in your organization is responsible for managing data and find out how you can work with it.
3. There is a ML mantra that "more data usually beats better algorithms". External data combined with internal data leads to better models. (Read our Lead Scoring white paper to learn more about external data).
4. Start with an 'intuition' algorithm based on your own knowledge supplemented with external data.
5. Show your ML algorithms to Sales, Marketing, Customer Service, anyone who works with the customer to see if it makes sense. ML is part science, part art.
6. Getting a good predictive model requires an iterative process and may require testing many different types of algorithms before finding the best one for your situation. Support 'failing fast.'
7. ML algorithms continually need to be updated, so review your algorithms at least quarter. Keep data up to date and use only what is relevant.
8. A common way of using ML is to build an Ideal Consumer Profile combined with Lead Scoring.
9. For best results, work with a company that specializes in machine learning, such as Vijilent.

# SUMMARY

This guide is meant as an introduction to help you and your team know how machine learning may be useful to your organization. Machine learning is a sophisticated science so this short guide cannot begin to discuss how to develop great models. Algorithm building is both an art and a science and great algorithms take time and expertise to develop correctly. There are many companies, including Vijilent that can help you develop ML algorithms to improve your business intelligence.

Look for companies that can also work to improve your predictive capabilities by assisting with data collection. Vendors can capture thousands of external data points such as website traffic, job listings, demographics, previous purchases and behavioral data that can be used to measure the level of intent to purchase. Vijilent provides insights into consumer behavior using DataPortraitsTM to supplement your existing data. For more on our DataPortraitsTM see our website.

# MACHINE LEARNING
## FOR LEAD QUALIFICATION

Vijilent

Vijilent is a SaaS enterprise working with companies to use data science to build business intelligence. The company provides uncanny people identification and insights to companies seeking to identify the ideal customer profile and build stronger and more valuable customer relationships. Vijilent's rich Data-Portraits™ include demographic, psychographic, and psycholinguistic information with personality (i.e. Big5) and customized behavior profiles about a segment-of-one (an individual or a firm). These insights are available in real-time and are powered by cutting-edge techniques: machine learning, natural language processing and sentiment analysis. The company focuses on delivering business intelligence to companies seeking to leverage data science in their lead generation programs.

WEBSITE TWITTER

1400 Leanne Court Raleigh, NC 27606 844-408-4545 email: rosanna.garcia@vijilent.com