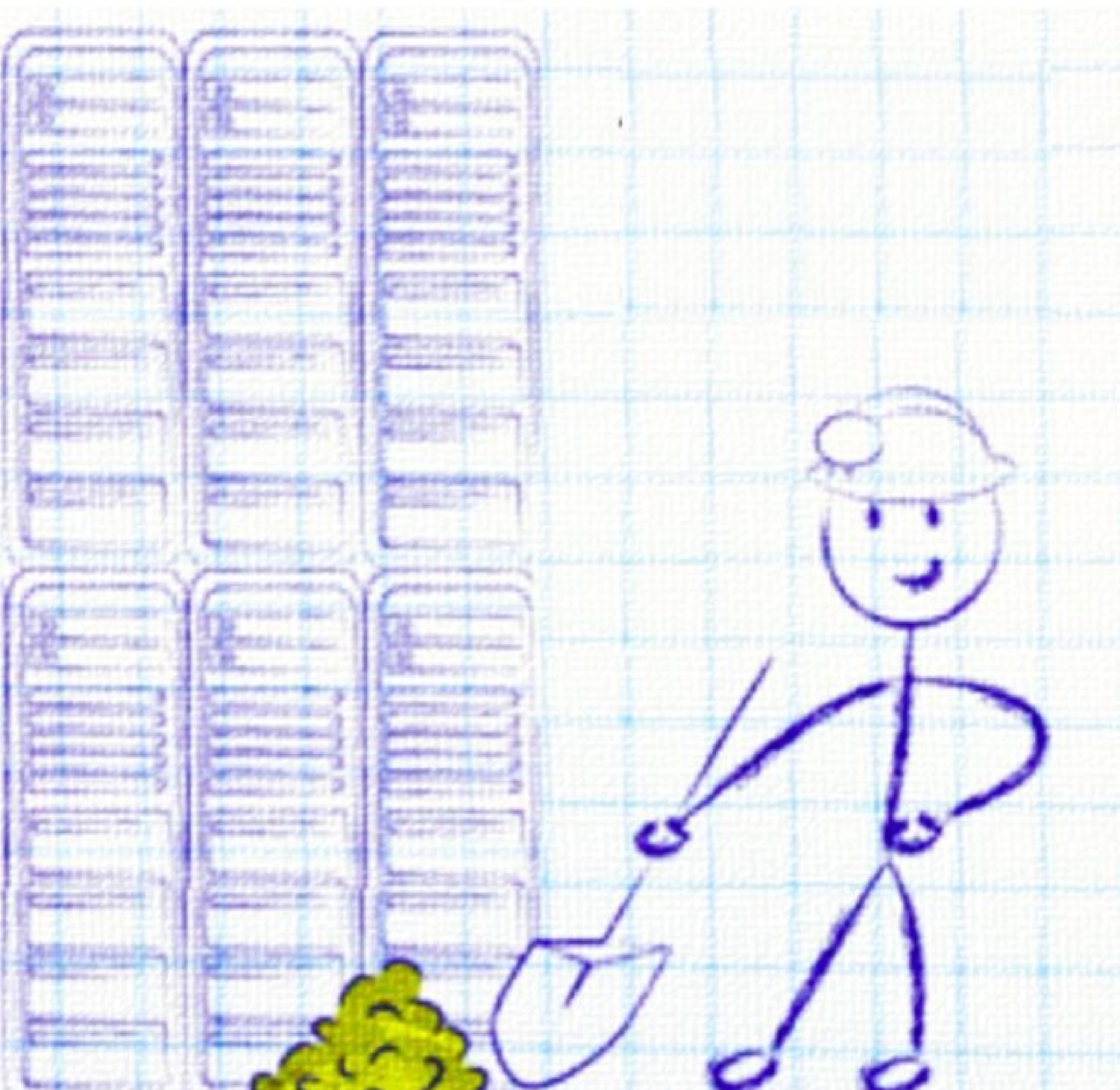
Machine Learning

For Absolute Beginners



Oliver Theobald

Machine Learning for Absolute Beginners

Oliver Theobald

First Edition

Contents Page

Introduction

It's a Friday night at home and you've just ordered a pizza from Joe's Pizzeria to be delivered to your house. The squeaky voice teen over the phone tells you that your pizza will arrive within 30 minutes.

But after hanging up the phone, you receive a message from your girlfriend (or boyfriend) asking if she/he can come over tonight.

Your girlfriend doesn't have a car, so you will have to drive over to her house and pick her up. While of course you want her to come over, you also don't want to wait until after the pizza has been delivered before you collect her - as the pizza will just sit there and get cold. You also don't want to pick her up after eating your pizza because then you'll miss the football game live on TV.

You need to make a quick decision. The first question you need to ask yourself, is do you have enough time to pick up your girlfriend before the pizza arrives?

Remember that the pizza is estimated to arrive within 30 minutes. If you leave now, you should be back within 30-40 minutes. As you know the route to your girlfriend's house, you can safely predict the journey time with a high degree of accuracy.

But just as you're about to walk out the door you realize there's another variable you haven't considered. You realize that what you also need to predict, in addition to the journey time to pick up your girlfriend, is the timing of the pizza being delivered. This too is something you have less control over.

Joe's Pizza is a popular pizzeria, and tonight also happens to be a Friday night. There's thus a range of factors that could affect your pizza delivery, including how many other people are ordering pizza, and the navigation ability of the delivery guy.

These two variables both have the potential to delay the delivery time of your pizza. However, this is your first time ordering a pizza on a Friday night. Perhaps unaware to you, Joe's Pizza has more delivery staff on call on Friday than say on a normal weeknight.

There are three potential methods to tackle this problem:

The first option is to apply existing knowledge. However you have no previous experience of ordering a pizza on a Friday night. Unfortunately there's also no app to calculate the average wait time on a Friday night for a pizza delivery in your area.

The second option is to ask someone else. You have exhausted this option already. The teenager on the other end of the phone at Joe's Pizzeria has already told you that your pizza will arrive "within 30 minutes".

The third option is to apply statistical modelling.

Given you've picked up this model on machine learning, let's go with the third option.

You think back to your previous experiences of ordering home delivery from Joe's Pizzeria. You then apply this information to predict the likelihood of the pizza arriving at your house on time. If the expected time of delivery exceeds 30 minutes then you can justify your decision to collect your girlfriend and return home in time for the delivery guy to arrive with your pizza.

Let's assume you have previously ordered pizza on 8 occasions, and the delivery time was late by greater than 10 minutes on four occasions. This means that the pizza arrived on time, or was early to arrive 50 percent of the time. This also means that there is roughly a 50% chance that the pizza delivery will be late again tonight.

Your mental decision-making progress is not comfortable with anything less than 70% (that the pizza delivery will be late). You thus remain at home to receive the pizza and make up an excuse not to see your girlfriend tonight.

Using existing data to base your decision is known as the empirical method. The concept of empirical data-backed decision-making is integral to what is known as machine learning.

Machine learning concentrates on prediction based on already known properties learned from the data.

In this example of the pizza delivery, we only considered the attribute of "frequency," the frequency of previous late deliveries. Machine learning models though consider at least two factors.

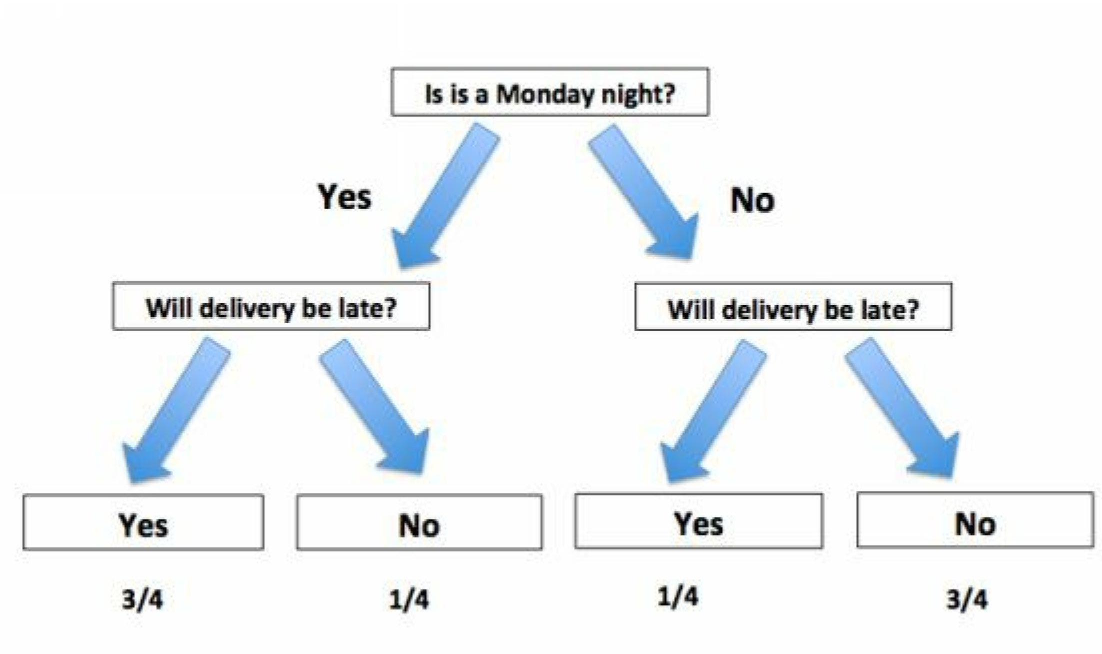
One factor is the result you wish to predict, known as the dependent variable. In this example, the dependent variable is whether the pizza delivery will be significantly late (more than 10 minutes). The second factor is the independent variable, which again predicts whether the pizza will be late but on a different independent variable. Day of the week, for example, could be an independent variable.

It could be a case that in the past, when the pizza was delivered on a Monday night the delivery time qualified as 'late'. This could be explained by the fact that Joe's Pizza has less delivery drivers on call on Monday nights.

Based on your previous experience, and not withstanding the three late deliveries that occurred on Monday night, pizza deliveries from Joe's Pizzeria typically arrive within the estimated time period.

This being the case, you could establish a model to simulate the probability that the pizza will arrive late based on whether or not it is a 'Monday night'.

A decision tree can be used to map out this particular example.



We now see that under this modelling there is only a 25% chance of the pizza delivery being late.

The process is relatively simple when considering a single independent variable. It does however become more complicated to calculate once a second or third independent variable are added to the equation.

Let's now add 'rain' as a third variable that could affect the pizza delivery time. A rainy night could of course slow down the delivery time due to safety precautions and extra traffic on the road.

This new variable is then added to the decision-making process. The new model now includes two independent variables in addition to one dependent variable.

We now need to predict the number of minutes the pizza will be late based on the level of rain (light = 2 minutes, moderate = 5 minutes, heavy = 15 minutes) and the day of the week. The predictions produced by this model will give us an idea on how late the pizza will be on any given day of the week. In this case though, a decision tree is of very little use as it can only predict discrete values (yes/no).

However, with the help of machine learning techniques you can apply the method of linear regression to predict the result.

It's now time to sit down at your computer. For the sake of the story let's forget the fact that your girlfriend is waiting for you to reply to her message.

Let's also turn our attention to discuss machines learning.

For decades, machines operated on the basis of responding to user commands. In other words, the computer would perform a task as a result of the user directly entering a command.

But as you may know, that has all changed.

The manner in which computers are now able to mimic human thinking to process information is rapidly exceeding human capabilities in everything from chess to picking the winner of a song contest. This leads us into the realm of artificial intelligence and machine learning.

In the modern age of machine learning, computers do not strictly need to receive an 'input command' to perform a task, but rather 'input data'. From the input of data they are able to form their own decisions and take actions virtually as a human would – but of course within the confines set by the machine's operator.

In machine learning, a computer creates a model to analyze the scenario based on existing data (experiences). The model in this case is predicting whether the pizza delivery will be late in future cases.

From here the computer treats the data very similar to normal human thinking. But given it is a machine, it can consider many more scenarios and execute far more complicated calculations to solve complex problems.

This is the element that excites data scientists and machine learning engineers the most. The ability to solve complex problems never before attempted. This is also perhaps one reason why you have picked up this book, to gain an introduction to machine learning, and techniques such as linear regression.

In the following sections we will first dive in and consider machine learning from an aerial view and discern the relationship between our topic and the larger field of data science.

Overview of Data Science

The Evolution of Data Science and the Information Age

Data science is a broad umbrella term that encompasses a number of disciplines and concepts including big data, artificial intelligence (AI), data mining and machine learning.

The discipline of studying large volumes of data, known as 'data science', is relatively new and has grown hand-in-hand with the development and wide adoption of computers. Prior to computers, data was calculated and processed by hand under the umbrella of 'statistics' or what we might now refer to as 'classical statistics'.

Baseball batting averages, for example, existed well before the advent of computers. Anyone with a pencil, notepad and basic arithmetic skills could calculate Babe Ruth's batting average over a season with the aid of classical statistics.

The process of calculating a batting average involved the dedication of time to collect and review batting sheets, and the application of addition and division.

The key point to make about classical statistics is that you don't strictly need a computer to work the data and draw new insight. As you're working with small data sets it is possible even for pre-university students to conduct statistics.

Indeed statistics are still taught in schools today, and as they have been for centuries. There are also advanced levels of classical statistics, but the data sets remains consistent - in that they are manageable for us as human beings to process.

But what if I wanted to calculate numbers (data) at a higher velocity (frequency), higher volume and higher value? What if I wanted to conduct calculations on my heart beat? Calculations not just on my heart beat, but also how my heartbeat reacts to temperature fluctuations and calories I consume. This is not something I can calculate in my head or even on paper for that matter. Nor would it be practical to collect such data.

This is where the information age and the advent of computers have radically transformed the subject of statistics. Modern computing technology now provides the infrastructure to collect, store and draw insight from massive amounts of data.

Artificial Intelligence

Artificial Intelligence, or AI as we also like to call it, has also been developing over the same period. It was first coined over sixty years when American computer scientist John McCarthy introduced the term during the 2nd Dartmouth Conference in 1956.

AI was originally described as a way for manufactured devices to emulate or even exceed the capabilities of humans to perform mental tasks.

AI today upholds a similar definition, anchored on enabling machines to think and operate similar to the human brain. AI essentially operates by analyzing behavior to solve problems and make decisions within various situations.

It’s interesting to note that the term AI is slightly controversial, in that it tends to confuse or intimidate those uninitiated to data and computer science.

IBM, for example, have gone to great lengths to disguise AI as ‘cognitive thinking’ so as not to intimate the average observer.

As part of a project my startup worked on with IBM Australia, we were featured in a video series exploring the possibilities of ‘Cognitive Thinking’ in Asia.

When we asked IBM why we had to say ‘cognitive thinking’ instead of ‘artificial intelligence’ or ‘AI’, their public relations team explained why based on their research. IBM was worried that the average person on the street would associate AI with robo-terminators eventually seeking out to kill everyone.

The portrayal of machines in movies hasn’t helped the plight of ‘AI’. An addition, as many have rightly pointed out, man has always found diametrical ways to cause great harm from new technology.

The other problem with ‘AI’ is that there’s a false illusion on parts of the Internet that AI and machine learning can be used interchangeably. This though is just poor reporting in the media or ignorance by the guy or girl on the social media team of big P.R companies.

Both are popular buzzwords but this is not how a trained data scientist perceives the two terms.

Within the very broad field of data science there are various disciplines that are used to manage and analyze big data. These disciplines include data mining, big data analytics, artificial intelligence and machine learning.

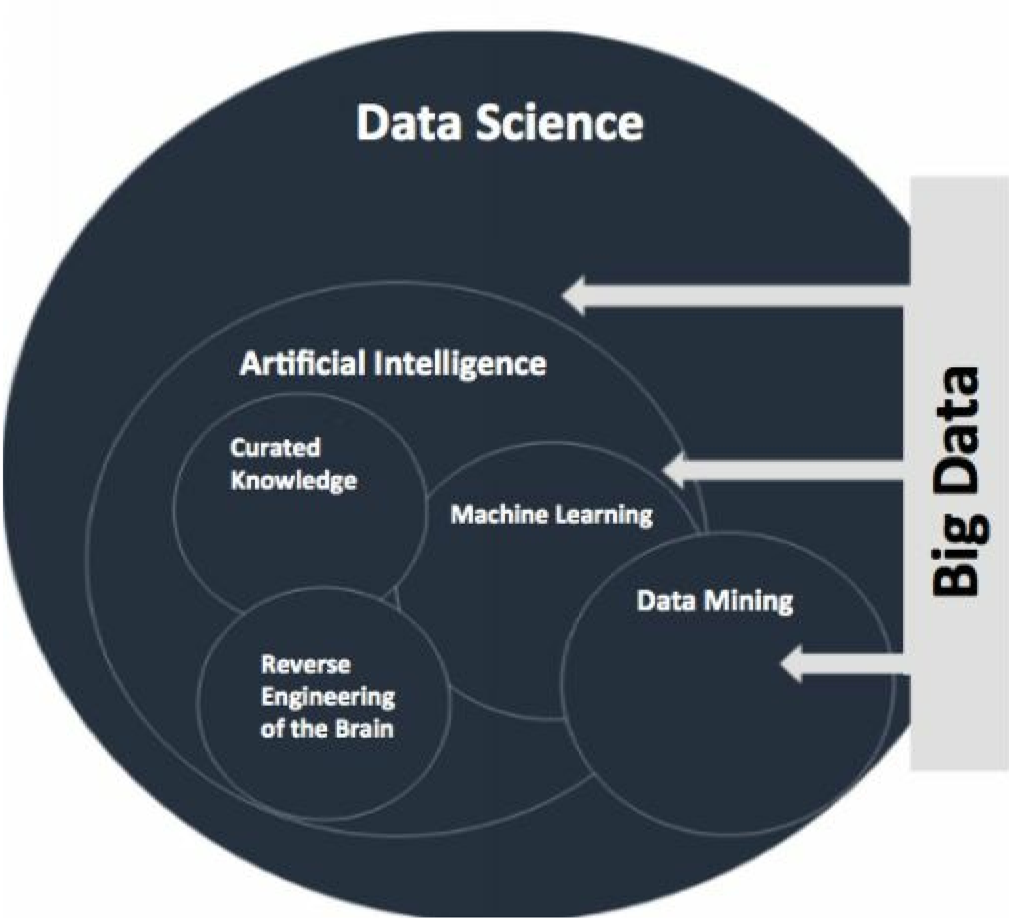
Big data analytics is an independent discipline that processes big data with the used of advanced algorithms based on a starting hypothesis.

An example of a big data analytics’ hypothesis could be: A relationship between the ambience (measured in decibels) at Manchester United home games played at Old Trafford and the likelihood of the home team coming from behind to win.

The next popular discipline within data science is data mining. Data mining involves applying advanced algorithms to unearth previously unknown relationships, patterns and regularities from a very large data set. Data mining is therefore similar to big data analytics but is different in that it doesn't have a starting hypothesis.

Much like prospecting for gold during a 19th Century Gold Rush, data mining begins without a clear future outcome. In fact, you don't even know what you are mining for! It could be gold, but it could just as equally be silver or oil that you stumble upon.

Lastly, artificial intelligence is a grouping of several techniques including machine learning. Machine learning overlaps with data mining because the machine learning's self-learning algorithms can also be applied to data mining in order to uncover previously undiscovered relationships.



Source: Inovancetech 2014

Evolution of Machine Learning

Machine learning algorithms have existed for virtually two decades but only in recent times has computing power and data storage caught up to make machine learning so widely available.

Computers for a long time were inept at mimicking human-specific tasks, such as reading tasks, translating, writing, video recognition and identifying objects. However, with advances in computing power, machines have now exceeded human capabilities at identifying patterns found in very large data sets.

Machine learning focuses on developing algorithms that can learn from the data and make subsequent predictions. For example, when you type in to Google "machine learning", it pops up with a list of search results.

But over time certain results on page one will receive fewer clicks than others. For example, perhaps result three receives fewer clicks than result four. Google’s machine learning based algorithm will recognize that users are ignoring result three and that entry will thereby begin drop in ranking.

Machine learning can also be applied independently or be applied to data mining on top of other data mining techniques.

The following chapters will walk you through the definitions and unique characteristics of other terms related to data science and machine learning.

Big Data

What is “big data”?

Big data is used to describe a data set, which due to its value, variety and velocity defies conventional ways of processing. Big data is therefore reliant on technology to be managed and analyzed.

In other words, big data is a collection of data that would be virtually impossible for a human to make sense of without the help of a computer.

Big data does not have an exact definition in size or how many rows and columns it would take to house such a large data set. But data sets are becoming increasingly bigger as we find new ways to efficiently collect and store data at low cost.

It’s also important to note that not all data is big data. Let’s use an example to illustrate the difference between “data” and “big data”.

First, imagine we want to know the total number of coffees sold by Starbucks over one business day in one suburb in the U.S. Total sales can be calculated on the back of a napkin by recording the total number of sales of each store within that suburb, and totalling those numbers using simple addition. This however – as you may have guessed by the mention of a *napkin* – is not considered ‘big data’.

Simple calculations such as total revenue, total profits and total assets have been recorded for millennia with the aid of pen and paper. Other rudimentary calculation tools such as abacuses in China have been used with equal success.

Nor does Starbucks dwarf the size of companies in existence prior to the computer age. The British Empire is a notable example of a highly organised and massive organization that could calculate income generated across a multitude of far-flung geographical territories without the aid of computers.

Therefore, what today defines big data is the power to process <u>very larger sets of data to unearth information never seen before with the aid of computers.</u>

So what then can the luxury brand Louis Vuitton learn today from big data that they couldn’t 50 years ago?

We can assume that profits, sales revenue, expenses and wage outlays are recorded with virtually the same precision today as they were 50 years ago. But what about other observations? How does, for example, staff demographics impact total sales?

Let’s say we want to know how age, company experience and the gender of Louis Vuitton service staff impacts a customer’s purchasing decision?

This is where technology and computers come into the frame. Digital equipment, including staff fingerprint check-in systems, customer relationship management systems (to manage details about sales and staff members), and payment systems can be all linked into one ecosystem.

The data is then stored in a database management system on a physical server or a distributed computing storage platform such as Hadoop, within a series of interconnecting tables that can be retrieved for instant access, or analyzed at a later date.

Big data analytics or data mining can then be applied to clean up and analyse the data to analyze or uncover interesting variables and gain insight from the trove of information collected.

Other business examples are plentiful. Starbucks now chooses store locations based on big data reports that factor in nearby location check-ins on social media, including Foursquare, Twitter and Facebook.

Netflix invested in a whole TV series based on a direct relationship they extracted via big data analytics and data mining. Netflix identified that:

- Users who watched the David Fincher directed movie *The Social Network* typically watched from beginning to end.
- The British version of “*House of Cards*” was well watched.
- Those who watched the British version “*House of Cards*” also enjoyed watching films featuring Kevin Spacey, and/or films directed by David Fincher.

These three synergies equated to a potential audience large enough in size to warrant purchasing the broadcasting rights to the well-acclaimed American TV series *House of Cards*.

Another big data business example is Juwai.com. Juwai is an Australian founded online real estate platform that lists overseas properties to its user base of Chinese investors.

This online real estate platform is now leveraging their access to big data to feed hedge fund managers and investment bankers.

Based on the data they can collect regarding what users search for on their portal, Juwai can collect data early in the purchasing decision cycle and synthesise search queries in rapid time through cloud computing and a super computer called a Vulcan (only five in the world).

The online behaviour they can capture from users on the site can then be packaged and commercialised to pinpoint future real estate patterns based on exact locations.

As an example, Juwai explained to me that a major trend over the last 12 months has been a surge in interest in Japanese real estate. A historically low Yen and growing exposure to Japan through Chinese tourism is leading to strong demand for Japanese properties from China, and this has been driving Chinese-language search queries for Japanese properties on their portal.

With Juwai's data 6-12 months ahead of the purchasing cycle, investment firms can stock up on urban hotspots in Japan and properties in close proximity to universities (which are a traditional magnet for Chinese investment money).

However, it's important to remember that big data is not a technique or process in itself. It is a noun to describe a lot of data.

Also, you don't necessarily have to have troves of data to conduct machine learning and data mining. Both machine learning and data mining techniques can be applied to a modest source of data found on an Excel spread sheet.

However in order to find valuable insight, big data provides a rich new source of data to extract value from, which would not be possible from a smaller data set.

Machine Learning

Machine learning, as we've touched upon already, is a discipline of data science that applies statistical methods to improve performance based on previous experience or detect new patterns in massive amounts of data.

A very important aspect of machine learning is the usage of self-improving algorithms. Just as humans learn from previous experience and trial and error to form decisions, so too do self-improving algorithms.

Not only can machine learning think and learn like us, but its more effective too. Humans are simply not predisposed to be as reliable and proficient at repetitive tasks to the same standard of computers in handling data. In addition, the size, complexity, and speed in which big data can be generated exceed our limited human capabilities.

Imagine the following data pattern:
1: [0, 0]
2: [3, 6]
3: [6, 12]
4: [9, 18]
5: [12, ?]

As humans it's pretty easy for us to see the pattern here. As the second number in each row is twice as large as the subsequent number to its left in the brackets, we can comfortably predict that the unknown number in brackets on row five will be '24'. In this scenario we hardly need the aid of a computer to predict the unknown number.

However, what if each row was composed of much larger numbers with decimal points running into double digits and with a far less clear relationship between each value? This would make it extremely difficult and near impossible for anyone to process and predict in quick time.

This task however is not daunting to a machine.

Machines can take on the mundane task of attempting numerous possibilities to isolate large segments of data in order to solve the problem at hand, as well as collecting, storing and visualizing the data.
Machine learning therefore frees up our time to focus on improving the results or other business matters.

But how can we program a computer to calculate something we don't even know how to calculate ourselves?

This is an important aspect of machine learning. If properly configured, machine learning algorithms are capable of learning and recognising new patterns within a matter of minutes.

But machine learning naturally doesn't just start by itself. As with any machine or automated production line, there needs to be a human to program and supervise the automated process. This is where data scientists and data professionals come into the picture.

The role of data scientists is to configure the equipment (including servers, operating systems and databases) and architecture (how the equipment interacts with each other) as well as programming algorithms using various mathematical operations.

You can think of programming a computer like training a guide dog. Though specialized training the dog is taught how to respond in various situations. For example, the dog is taught to heel at a red light or to safely lead its master around certain obstacles.

If the dog has been properly trained then the trainer is no longer required and the dog will be able to apply his/her training to various unsupervised situations.

This example draws on a situational scenario but what if you want to program a computer to take on more complex tasks such as image recognition.

How do you teach a computer to recognise the physical difference between various animals? Again this requires a lot of human input.

However, rather than programming the computer to respond to a fixed possibility, such as a navigating an obstacle on the path or responding to a red light, the data scientist will need to approach this method differently.

The data scientist cannot program the computer to recognise animals based on a human description (i.e. four legs, long tail and long neck), as this would induce a high rate of failure. This is because there are numerous animals with similar characteristics, such as wallabies and kangaroos. Solving such complex tasks has long been the limitation of computers and traditional computer science programming.

Instead the data scientist needs to program the computer to identify animals based on socializing examples the same way you teach a child.

A young child cannot recognise a 'goat' accurately based on a description of its key features. An animal with four legs, white fur and a short neck could of course be confused with various other animals.

So rather than playing a guessing game with a child, it's more effective to showcase what a goat looks like by showing the child toy goats, images of goats or even real-life goats in a paddock.

Image recognition in machine learning is much the same, except teaching is managed via images and programming language.

For example, we can display various images to the computer, which are labelled as the subject matter, ie. 'goat'. Then the same way a child learns, the machine draws on these samples to identify the specific features of the subject.

At work I even read an example of Chinese company that had developed machine learning algorithms to detect illicit video content and pornography. Now to answer what you are probably thinking… yes, the computers would have been fed a high volume of pornographic material in order to develop such advanced video recognition capabilities!

Whether its recognizing animals, human faces or illicit adult material, the machine can apply examples to write its own program to provide the capability to recognize and identify subjects. This eliminates the need for humans to explain in detail the characteristics of each subject and dramatically mitigates the chance of failure.

Once both the architecture and algorithms have been successfully configured, machine learning can take place. The computer can then begin to implement algorithms and models to classify, predict and cluster data in order to draw new insights.

Data Mining

Data mining, as mentioned, is a data science discipline that aims to unearth previously unknown relationships, patterns and regularities from large data sets, and does not start with a set hypothesis.

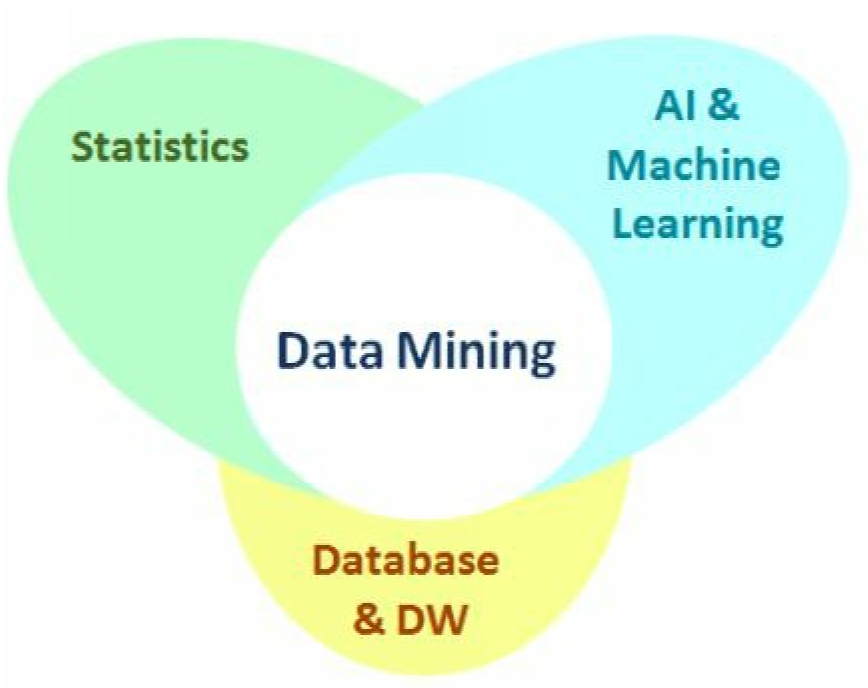
A key point to remember regarding data mining is that it only applies to situations where you are seeking to find patterns and regularities within the data set that are yet to be seen.

Given that data mining does not begin with an exact hypothesis as an initial starting point, a myriad of data sorting techniques are applied, including text retrieval, clustering, sequence analysis and association analysis.

A big question for people new to data science is: What's the difference between 'data mining' and 'machine learning'?

First, we know that both disciplines fall under the broad umbrella of data science, and computer science as well for that matter. Machine learning also falls within the field of artificial intelligence due to its ability to mimic human learning processes from the application of trial and error.

There is however a correlation between the two. In some cases, data mining utilizes the same algorithms applied to machine learning in order to interpret data. Popular algorithms such as k-means clustering, dimensions reduction algorithms and linear regression are used in both data mining and machine learning.



Given the close interconnectivity between data mining and machine learning, it is important to understand both disciplines.

At a very abstract level, both are concerned with analyzing data and extracting valuable insights.

Whereas machine learning uses algorithms to improve with experience at a given task, data mining focuses on analyzing data to discover previously unseen patterns or properties and applies a more broad range of algorithms.

Machine learning concentrates on studying and reproducing specifically known knowledge, whereas data mining is exploratory and searches for unknown knowledge.

Machine learning algorithms though can be used within data mining to identify patterns. A machine learning algorithm such as k-means, for example, could be applied to determine if any clusters exist in the data. K-means is an algorithm that learns from known structures within the data.

Machine Learning Tools

There are several important underlying technologies that provide the infrastructure for machine learning.

Infrastructure is technology that allows data to be collected, stored and processed. Data infrastructure includes both traditional hardware and virtual resources.

Traditional hardware is physically stored on premise in the form of computer servers. Virtual resources are provided through cloud computing from major cloud providers including Amazon and Microsoft.

Similar to the way you consume and pay for your electricity, gas, water and traditional utilities, cloud computing offers you full control to consume compute resources on-demand. As a user you can simply rent compute resources from a cloud provider in the form of virtual machines.

In business, government and virtually all sectors, traditional hardware is rapidly being replaced by cloud infrastructure. Rather than procure their own private and physical hardware to house and process data, companies can pay a monthly, pay-as-you-go or upfront fee to access advanced technology offered by cloud providers. This means that companies only pay for what they need and use.

By using data infrastructure services available on the cloud, companies can avoid the expensive upfront cost of provisioning traditional hardware as well as the expensive cost to maintain and later upgrade the equipment.

Cloud technology also frees up data scientists to focus on data management and machine learning rather than configuring and maintaining the hardware. Updates and data backups can be made automatically on the cloud.

Data services, including database storage and analytics are available on the cloud through vendors such as Amazon, IBM and Google.

The affordability of cloud technology has led to an increase in demand from companies to conduct data science programs in order to solve business problems. Meanwhile, this has led to greater demand for data scientists and machine learning professionals to manage such programs.

As with any hardware you also have software. Machine learning software typically falls into two camps. There are text-based interfaces, which rely on programming languages and written commands – a black screen with a lot of code.

The advantage of text-based interfaces is that they're easy to share, transplant and replicate.

Then there are graphical interfaces that incorporate menus and widgets, which you can interact with to process the data and create data visualisation.

The advantage of a graphic interface is that it offers an intuitive workspace and you can drag widgets to manage your data operations. It also allows you to see your data results visually.

Machine Learning Case Studies

Online Advertising

An easy-to-digest example of machine learning is online advertisements. Ever wondered how Facebook, YouTube or Amazon can see into your brain and know what you want? This is where machine learning meets precision marketing.

The process relies on pooling together data collected from millions of online users and applying self-learning algorithms to retrieve user insight from the data

YouTube for example processes your previous online activities and applies an algorithm to populate ads within your browser whenever you visit YouTube. Through pooling data from various sources such as Google search queries, YouTube is able to know what you like.

The ads displayed to you should also be different to your colleague or classmate sitting next to you, as it is based on the unique data collected of each user. YouTube also doesn't know what each user likes until they apply machine learning techniques to learn from the data and draw out insight.

Still, not all companies and websites have this ability to tailor ads to each user. I recently had a New Zealand friend share a print-screen of a funny news story to a private Facebook group. While his intention was to stir up the Aussies in the group, it quickly backfired. On the right hand side of the webpage that he'd print-screened were ads for 'Viagra' and 'single men living nearby.'

While I don't know my New Zealand friend well enough to determine if the Viagra ad was related to his online activities, the second advertisement appeared unrelated.

The ads were not populated by a machine learning algorithm linked to his online viewing activity, but instead linked to the demographic of site viewers (predominately male) and keywords found in the text, which included references to homosexuality.

The ads on my friend's webpage were just as likely to appear as the same on my webpage if I was to visit the website. This hardly qualifies as machine learning. It was a very simple algorithm based on a data set no larger than the contents of that webpage. Not that I took the liberty of explaining this to the group!

Applying machine learning to extract personal preferences demands a much more sophisticated process.

Google, Facebook, Amazon and YouTube for example not only collect data on your online activity but on the browsing habits of millions of users. To process all this data, they need extensive infrastructure to collect, store, sort and export the information.

These big tech giants can then sell that data onto other companies or share with subsidiary companies. Other websites can develop their own precision marketing ad campaigns based on the data.

A major pull factor behind Google's acquisition of YouTube was indeed the access and synchronization of data flow. Google knew they could make YouTube ads more effective by leveraging their access to users' Google search habits.

It is also now possible for Facebook, Ebay and other online sites to know where you do your offline shopping. The free Wi-Fi available in shopping centres is able to track your exact where-a-bouts and record relevant information such as how long you stood in the golf shop or the Apple store.

Those data points are collected, packaged and sold to third parties before the likes of Ebay and Amazon process that data and then feed that data into their advertisement display algorithms to your browser.

This degree of surveillance rings of a Hollywood blockbuster about a future era in human history staring Bruce Willis - except it is happening today right at this moment!

E-commerce companies are not alone in leveraging machine learning for their own commercial benefit. Police departments and even sports teams are processing big data with machine learning to gain unprecedented and scarily accurate predictions never seen before.

Google's Machine Learning

The world of search engine optimization is changing and machine learning is firmly behind the new face of SEO.

As virtually everyone (outside of Mainland China and North Korea) with access to the Internet can use Google to search online, Google's new machine learning SEO technology is an easy to digest example of machine learning.

Prior to the integration of machine learning into search engine algorithms, Google focused their search efforts around strings of letters.

Google indexed millions of web pages each day to track their content for strings of letters. This included strings of letters in the webpage title, website menu, body text, meta tags, image descriptions and so forth.

With all these strings of letters and combinations on record, Google could match results based on the string of letters you entered into the search bar. If you typed in: "Donald Trump," the search engine would then go away and look for strings of letters in the following order:

D-O-N-A-L-D T-R-U-M-P

While there are various factors that influence SEO rankings, including backlinks and page speed, string letter matching has always been a major part of Google's SEO efforts. Webpages that contained the exact string of letters entered by the user would thereby feature prominently in the search results.

However, if you were to jumble up the letter sequence in any significant way, such as R-O-N-A-L-D D-R-U-M, the results would differ dramatically.

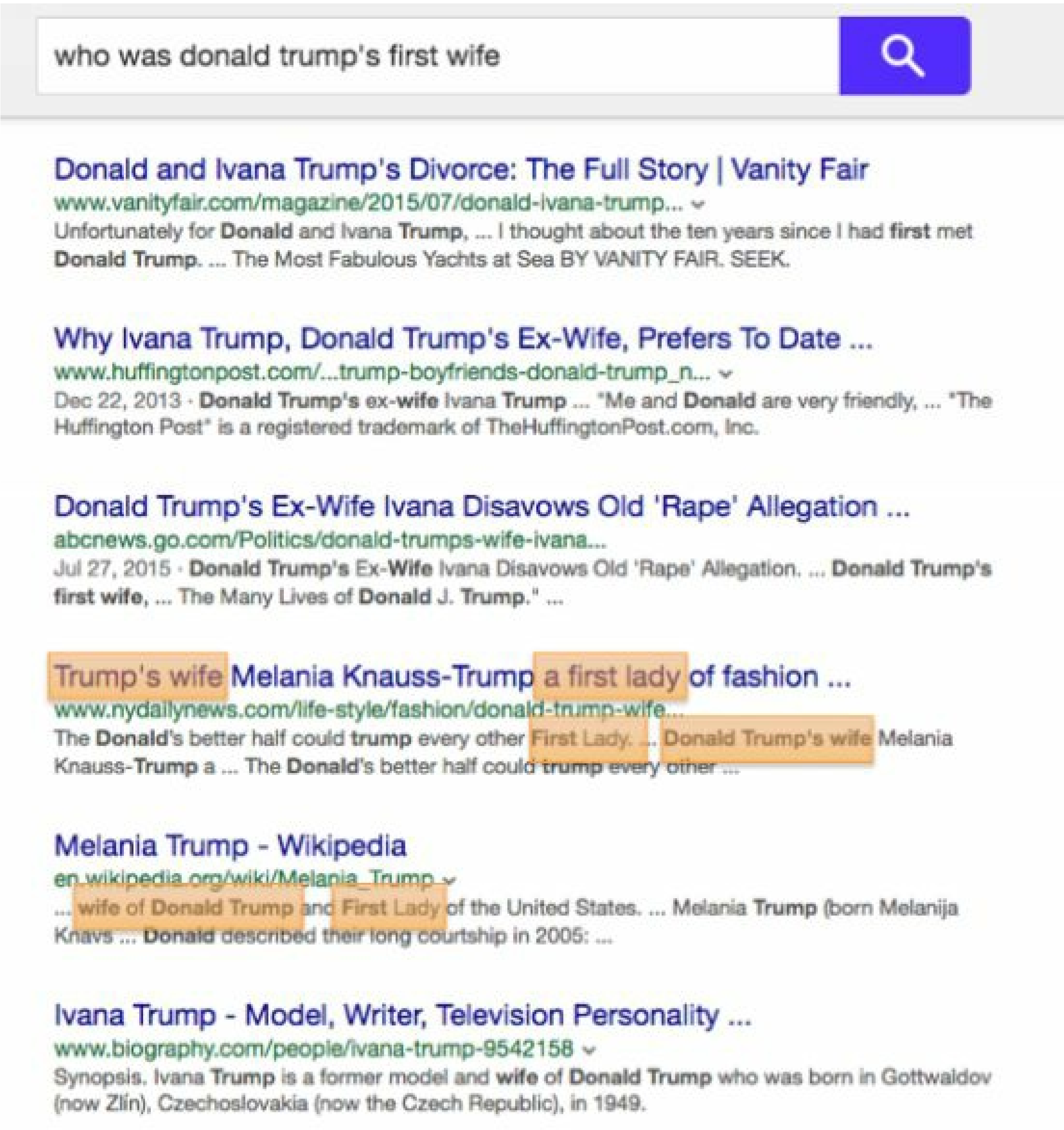
But Google's new algorithm – backed by machine learning – looks at "Donald Trump" not as a string of letters but as an actual person. A person who has a defined age, a defined job profile, a list of relatives and so forth.

Google can thereby decipher information without only relying on matching strings of letters.

For instance, say you search: "Who is Donald Trump's first wife?"

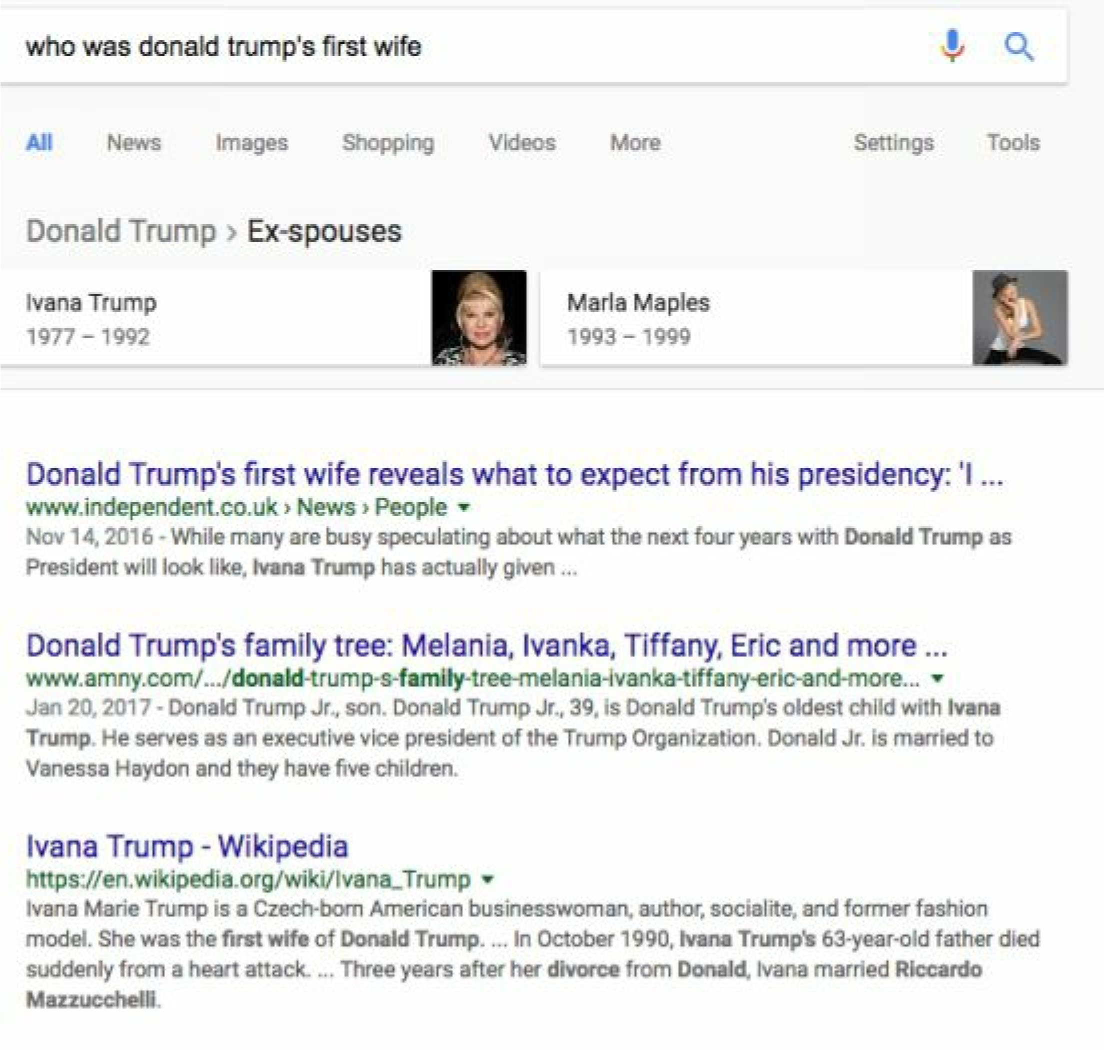
Prior to machine learning, Google would search its online repository for webpages containing those six keywords. However, the accuracy of search results could be variable.

The search engine, for example, may find an overwhelming number of web pages with keywords mentioning “Donald Trump’s wife” “Melania Trump” as the “First Lady” of the U.S. Google could thereby be tricked into featuring an article regarding Melania Trump within the first page of search results. The same still happens if you search on Yahoo today.



Source: Yahoo Search

Google though is much smarter thanks to the invisible hand of machine learning. Google is able to decipher words not strictly as strings of letters but things. Google knows Donald Trump is a person, and Google knows who his first wife is. It can then processes this information in rapid time to display information regarding Donald Trump’s first marriage to Ivana Trump.



Source: Google

But what's even more exciting is Google's new ability to understand interconnected search queries. For example, say you follow up the next Google search with the question: "Who was his second wife?"

Again, prior to machine learning, Google would search its online repository for webpages containing those exact keywords. But Google would not be able to connect your first search query with the second search query.

Machine learning though changes the way we search. Now, given that Google already knows our first search query was regarding, "Donald Trump", it can thereby decipher the second search query with less specific information provided.

For example, you could follow up by asking: "Who is his wife?"
And BANG, Google will come back with results regarding Donald Trump's wife – based on self-learning algorithms.

Google's new line of learning and thinking is very similar to human behavior and which is why Google's new technology falls within the field of machine learning.

Machine Learning Techniques

Introduction

Machine learning algorithms can be split into different classes of algorithms, including supervised, unsupervised and reinforced.

Supervised

Supervised algorithms refer to learning guided by human observations and feedback with known outcomes.

For instance, supposes you want the machine to separate email into spam and non-spam messages.

In a supervised learning environment, you already have information that you can feed the machine to describe what type of email should belong to which category. The machine therefore knows that there are two labels available in which to sort the incoming data (emails).

Or to predict who will win a basketball game, you could create a model to analyze games over the last three years. The games could be analyzed by total number of points scored and total number of points scored against in order to predict who will win the next game.

This data could then be applied to a model of classification. Once the data has been classified and plotted on a data plot we can then apply regression to predict who will win based on the average of previous performances. The final result would then supply an answer based on overall points.

As with the first example, we have instructed the machine which categories to analyze (points for, and points against). The data is therefore already pre-tagged.

Supervised algorithms, with tags applied, include Linear Regression, Logistic Regression, Neural Networks, and Support Vector Machine algorithms

Unsupervised

In the case of an unsupervised learning environment, there is no such integrated feedback or use of tags. Instead the machine learning algorithm must rely exclusively on clustering separate data and modify its algorithm to respond to its initial findings - all without the external feedback of humans.

Clustering algorithms are a popular example of unsupervised learning. Clustering group together data points which are discovered to possess similar features.

For example, if you cluster data points based on the weight and height of 13-year old high school students, you are likely to find that two clusters will emerge from the data. One large cluster will be male and the other large cluster will be female. This is because girls and boys tend to have separate commonalities in regards to physical measurement.

The advantage of applying unsupervised algorithms is that it enables you to discover patterns within the data that you may not have been aware existed – such as the presence of two different sexes.

Clustering can then provide the springboard to conduct further analysis after particular groups have been discovered.

Unsupervised algorithms, without tags, include clustering algorithms and descending dimension algorithms.

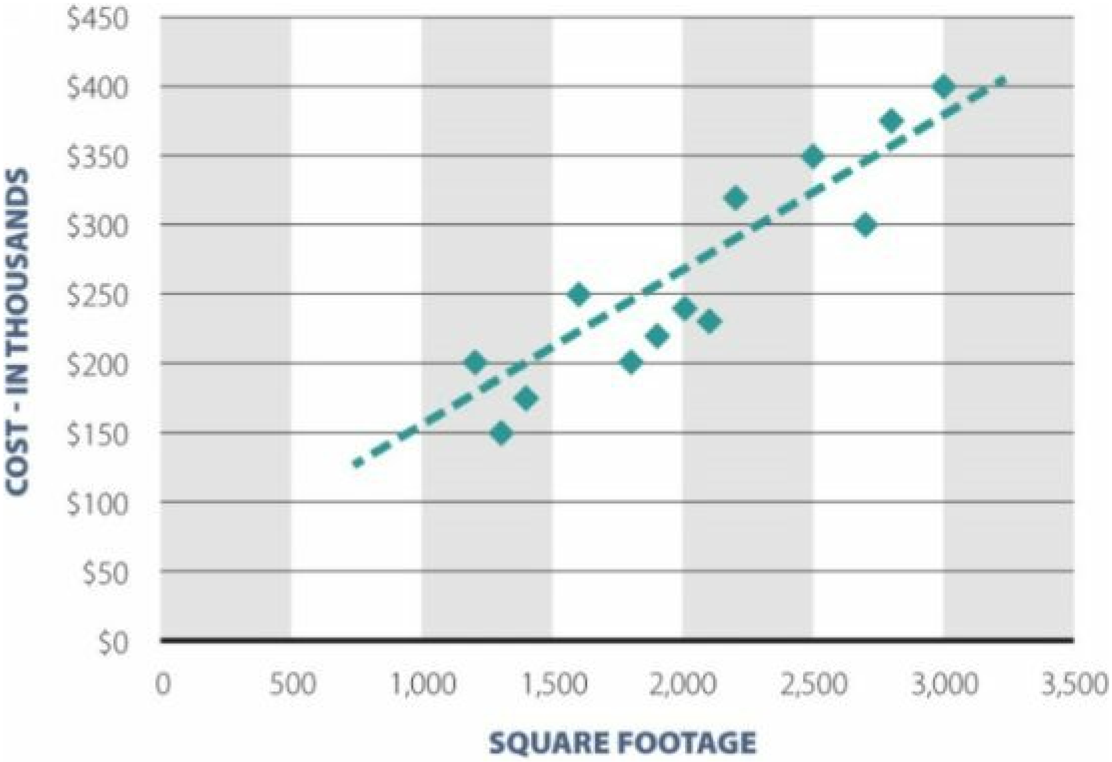
Regression

Regression is an important aspect of machine learning. Regression is important as it provides the base for other more advanced machine learning algorithms (including neural networks and recommendation algorithms), and it relatively easy to understand.

So what is regression?

Regression is a statistical measure that takes a group of random variables and seeks to determine a mathematical relationship between them. Expressed differently, regression calculates numerous variables to predict an outcome or score.

A simple and practical way to understand regression is to consider the scatter plot below:



The two quantitative variables you see above are house cost and square footage. House value is measured on the vertical axis (Y), and square footage is expressed along the horizontal axis (x). Each dot (data point) represents one paired measurement of both 'square footage' and 'house cost'. As you can see, there are numerous data points representing numerous houses within one particular suburb.

To apply regression to this example, we simply draw a straight line through the data points as seen above.

But how do we know where to draw the straight line? There any many ways we could split the data points with the regression line, but the goal is to draw a straight line that best fits all the points on the graph, with the minimum distance possible from each point to the regression line.

This means that if you were to draw a vertical line from the regression line to every data point on the graph, the distance of each point would equate to the smallest possible distance of any potential regression line.

As you can see also, the regression line is straight. This is a case of linear regression. If the line were not straight, it would be known as non-linear regression, but we will get to that in a moment.

Another important feature of regression is **slope**. The slope can be simply calculated via referencing the regression line. As one variable (X or Y) increases, you can expect the other variable will increase to the average value denoted on the regression line. The slope is therefore very useful for forming predictions.

What's more, the closer the data points are to the regression line, the more accurate your prediction will be. If there is a greater degree of deviation in the distance between the data points and your regression line then the less accurate your slope will be in its predictive ability.

Do note that this particular example applies to a bell-curve, where the data points are generally moving from left-to-right in an ascending fashion. The same linear regression approach does not apply to all data scenarios. In other cases you will need to use other regression techniques – beyond just linear.

Regression is used in a range of disciples including data mining, finance, business and investing. In investment and finance, regression is used to value assets and understand the relationship with variables such as exchange rates and commodity prices.

In business, regression can help to predict sales for a company based on a range of variables such as weather temperatures, social media mentions, previous sales, GDP growth and inbound tourists.

Specifically, regression is applied to determine the strength of a relationship between one dependent variable (typically represented at Y) and other changing variables (known also as independent variables).

There are various types of regression, including linear regression, multiple linear regression and non-linear regression methods, which are more complicated.

Linear regression
Linear regression uses one independent variable to predict the outcome of the dependent variable, or (represented as Y).

Multiple regression
Multiple regression uses two or more independent variables to predict the outcome of the dependent variable (represented as Y).

Regression can be calculated as follows:

Linear Regression: $Y = a + bX + u$

Multiple Regression: $Y = a + b1X1 + b2X2 + b3X3 + ... + btXt + u$

Where:

Y = is the variable you are predicting (the dependent variable)
X = is the variable you are using to predict the Y value (independent variable)
a = is the intercept

b = is the slope
u = the regression residual

In the case of linear regression, the relationship is denoted in the form of a straight line that best approximates the individual data points.

In the case of multiple regression, the separate variables are differentiated via numbers with subscript.

Non-linear regression
Non-linear regression modelling is similar in that it seeks to track a particular response from a set of variables on the graph. However, non-linear models are somewhat more complicated to develop.

Non-linear models are created through a series of approximations (iterations), typically based on a system of trial-and-error. The Gauss-Newton method and the Levenberg-Marquardt method are popular non-linear regression modelling techniques.

Logistic regression
Logistic regression and linear regression are similar in nature but different in regards to the problems they solve. Linear regression addresses numerical problems and forms numerical predictions (in numbers). Whereas, logistic regression is used within classification algorithms to predict discrete classes.

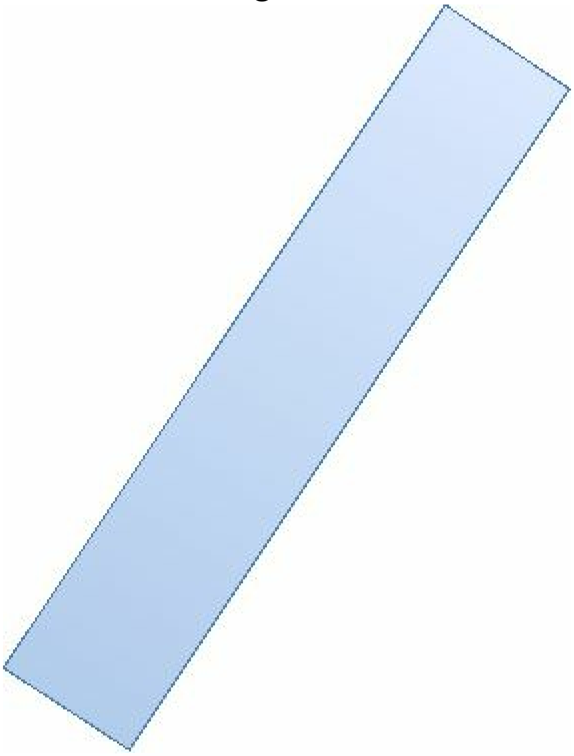
Logistic regression, for example, is often used in fraud detection or to identify spam email.

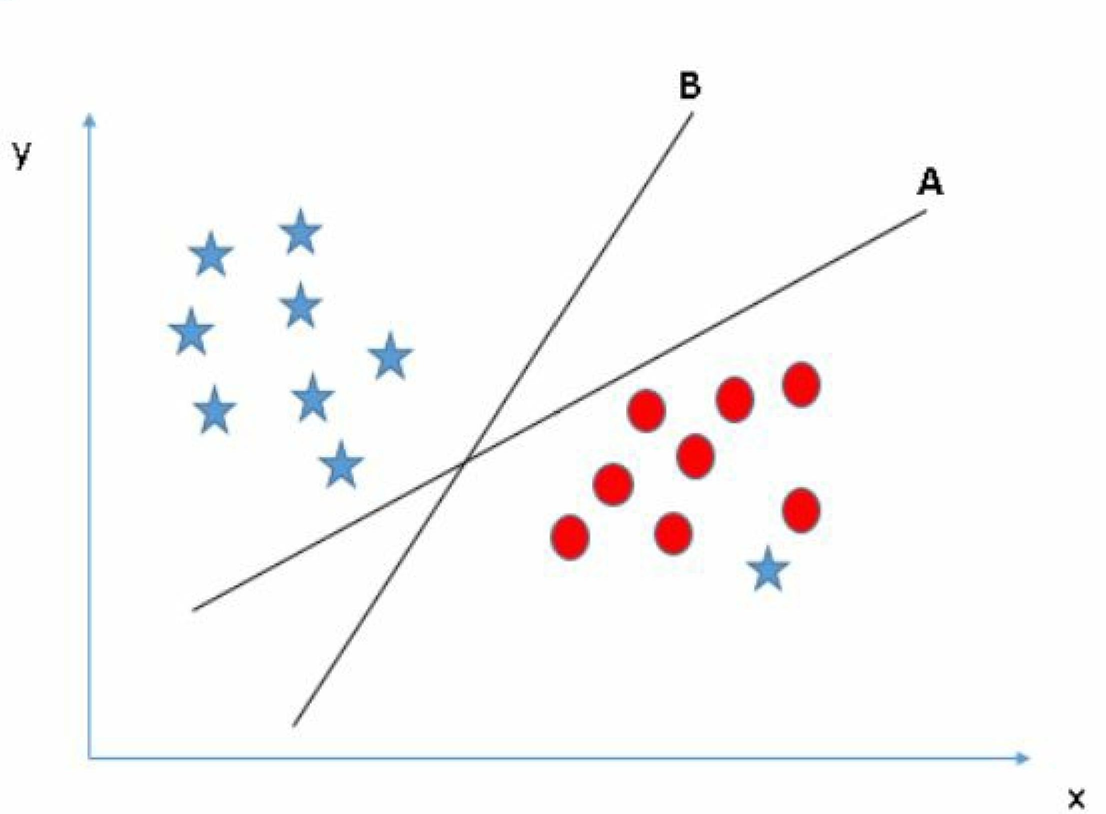
In practical usage, logistic regression is applied similarly to linear regression. Logistic regression however adds a Sigmoid function to compute the result and converts numerical result into a number of probability between 0 and 1.

A value of 0 represents no chance of occurring, and 1 represents a certain chance of occurring. The degree of probability for values located between 0 and 1 can be calculated according to how close they rest to 0 (impossible) to 1 (certain possibility). The value 0.75 for example would be considered a probable possibility, or expressed as a 75% chance.

Support Vector Machine Algorithms

Support vector machine (SVM) algorithms are an advanced progression from logistic regression algorithms, as we have just explored. SVM algorithms are essentially logistic regression algorithms with stricter set conditions. To that end, SVM algorithms are better at drawing classification boundary lines.





Let's see what this looks like in practice. Above on the plane are data points that are linearly separable. A logistic regression algorithm, as we know, will split the two groups of data points with a straight line that minimizes the distance between all points.

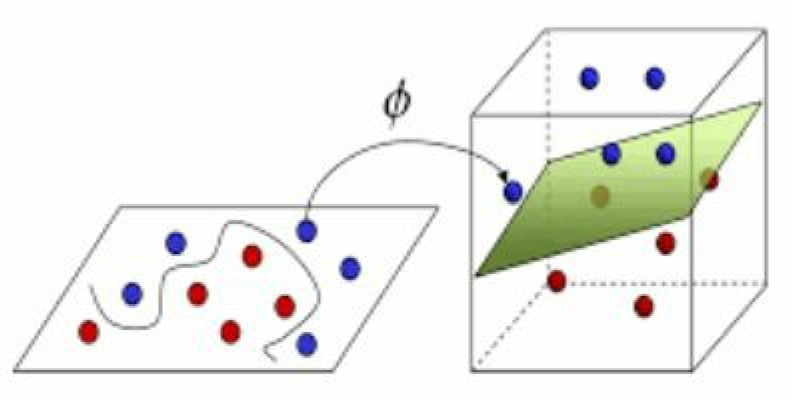
In the picture above you can see that Line A (logistic regression hyperplane) is positioned snuggly between points from both groups.

As you can also see, line B (SVM hyperplane) is also separating the two groups but from a position with maximum space between itself and the two groups of data points.

You will also notice that within the image is a light blue area that denotes Margin. Margin is the distance between the hyperplane and the nearest point, multiplied by two. An SVM hyperplane should be located in the middle of the Margin.

If however the data is not linearly separable, then it is possible to apply what is known as a Kernel Trick. When combined with SVM, the Kernel trick can map data from low-dimensional to high-dimensional.

Transitioning from a two dimensional to a third dimensional space allows you to use a linear plane to achieve similar result to split the data but within a 3-D space.



Source: ieeexplore

Artificial Neural Networks - Deep Learning

Deep learning is a popular area within data science today.

Deep learning became widely popular in 2012 when tech companies started to show off what they were able to achieve through sophisticated layer analysis, including image classification and speech recognition.

Deep learning is also just a sexy term for Artificial Neural Networks (ANN), which have been around for over forty years.

Artificial Neural Networks (ANN), also known as Neural Networks, are one of the most widely used algorithms within the field of machine learning. Neural networks are commonly used in visual and audio recognition.

ANN emphasizes on analyzing data in many layers, and was inspired by the human brain, which can visually process objects through layers of neurons.

ANN is typically presented in the form of interconnected neurons that interact with each other. Each connection has numeric weight that can be altered and is based on experience.

Much like building a human pyramid or a house of cards, the layers or neurons are stacked on top of each other starting with a broad base.

The bottom layer consists of raw data such as text, images or sound, which are divided into what we called neurons. Within each neuron is a collection of data. Each neuron then sends information up to the layer of neurons above. As the information ascends it becomes less abstract and more specific, and the more we can learn from the data from each layer.

A simple neural network can be divided into input, hidden, and output layers. Data is first received by the input layer, and this first layer detects broad features. The hidden layer/s then analyze and processes that data, and through the passing of each layer with less neurons (which diminish in number at each layer) the data becomes clearer, based on previous computations. The final result is shown as the output layer.

The middle layers are considered hidden layers, because like human sight we are unable to naturally break down objects into layered vision.

For example, if you see four lines in the shape of a square you will visually recognize those four lines as a square. You will not see the lines as four independent objects with no relationship to each other.

ANN works much the same way in that it breaks data into layers and examines the hidden layers we wouldn't naturally recognise from the onset.

This is how a cat, for instance, would visually process a square. The brain would follow a step-by-step process, where each polyline (of which there are four in the case of a square) is processed by a single neuron.

Each polyline then merges into two straight lines, and then the two straight lines merge into a single square. Via staged neuron processed, the brain can see the square.

Four decades ago neural networks were only two layers deep. This was because it was computationally unfeasible to develop and analyze deeper networks. Naturally, with the development of technology it is possible to easily analyze ten or more layers, or even over 100 layers.

Most modern algorithms, including decision trees and naive bayes are considered shallow algorithms, as they do not analyze information via numerous layers as ANN can.

Clustering Algorithms

Algorithms that are able to identify tags from training the data are known as unsupervised algorithms, whereas algorithms that are used to train data with set tags are known as supervised algorithms. Popular unsupervised algorithms are clustering algorithms.

Simply put, a clustering algorithm computes the distance between groupings and divides data points into multiple groups based on their relational distance to one another. Clustering differs from classification.

Unlike classification, which starts with predefined labels reflected in the database table, clustering creates its own labels after clustering the data set. Analysis by clustering can be used in various scenarios such as pattern recognition, image processing and market research.

For example, clustering can be applied to uncover customers that share similar purchasing behaviour. By understanding a particular cluster of customer purchasing preferences you can then form decisions on which products you can recommend to the group based on their commonalities. You can do this by offering them the same promotions via email or click ad banners on your website.

Descending Dimension Algorithms

A descending dimension algorithm is another category of unsupervised algorithm that effectively reduces data from high-dimensional to low-dimensional.

Dimensions are the number of features characterizing the data. For instance, hotel prices may have four features: room length, room width, number of rooms and floor level (view).

Given the existence of four features, the hotel room would be expressed on a four dimensional (4D) data graph. However, there is an opportunity to remove redundant information and reduce the number of dimensions to three by combining 'room length' and 'room width' to be expressed as 'room area.'

Applying a descending dimension algorithm will thereby enable you to compress the 3D data graph into a 2D data graph.

Another advantage of this algorithm is visualization and convenience. Understandably, it's much easier to work and communicate information on a 2D plane rather than a 4D data graph.

Descending dimension algorithms are commonly used to compress data and improve the efficiency of other machine learning algorithms. A popular algorithm in this category is Principal Component Analysis (PCA).

Association Analysis

Association analysis algorithms are commonly used by e-commerce websites and retailers to analyze transactional data and identify specific items that are commonly purchased together. This insight allows e-commerce sites and retailers to strategically showcase and recommend products to customers based on common purchase combinations and thereby increase purchasing.

Association algorithms fall into two primary categories:

1. **Content-based**

Content-based algorithms recommend items to a user based on items similar to their purchase. For example, an e-commerce store offering charcoal to customers before they checkout purchasing a home BQQ set. As long as items are properly tagged, these algorithms can be highly effective.

2. **User-based**

Used-based algorithms recommend items to a user based on the on items purchased by other users with shared interests. For example, if fans of hard metal music who enjoy listening to Song A also enjoy listening to Song B, and Soundify determines that you fit the same user category of heavy metal enthusiast, Soundify will recommend you listen to Song B after listening to Song A.

The first step in association analysis is to construct **frequent itemsets** (X). Frequent itemsets means a combination of items that regularly appear together, or have an affinity for each other. The combination could be one item with another single item. Alternatively, the combination could be two or more items with one or more other items. From here you can calculate an index number called support (SUPP) that indicates how often these items appear together.

Please note that in practice, “support” and “itemset” are commonly expressed as “SUPP” and “X”.

Support can be calculated by dividing X by T, where X is how often the itemset appears in the data and T is your total number of transactions. For example, if E only features once in five transactions, then the support will be 1 / 5 = 0.2.

However in order to save time and to allow you to focus on items with higher support, you can set a minimum level known as **minimal support** or **minsup**. Applying minsup will allow you to ignore low level cases of support.

The other step in association analysis is **rule generation**. Rule generation is a collection of if/then statements, in which you calculate what is known as confidence. Confidence is a metric similar to conditional probability.

IE, Onions + Bread Buns > Hamburger Meat

Numerous models can be applied to conduct association analysis. Below is a list of the most common algorithms:
- Apriori
- Eclat (equivalence class transformations)
- FP-growth (frequent pattern)
- RElim (recursive elimination)
- SaM (split and merge)
- JIM (Jaccard itemset mining)

The most common algorithm is Apriori. Apriori is applied to calculate support for itemsets one item at a time. It thereby finds the support of one item (how common is that item in the dataset) and determines whether there is support for that item. If the support happens to be less than the designated minimum support amount (minsup) that you have set, the item will be ignored.

Apriori will then move on to the next item and evaluate the minsup value and determine whether it should hold on to the item or ignore it and move on.

After the algorithm has completed all single-item evaluations, it will transition to processing two-item itemsets. The same minsup criteria is applied to gather items that meet the minsup value. As you can probably guess, it then proceeds to analyse three-item combinations and so on.

The downside of the Apriori method is that the computation time can be slow, demanding on computation resources, and can grow exponentially in time and resources at each round of analysis. This approach can thus be inefficient in processing large data sets.

The most popular alternative is Eclat. Eclat again calculates support for a single itemset but should the minsup value be successfully reached, it will then proceed directly to adding an additional item (now a two-item itemsets).

This is different to Apriori, which would move to process the next single item, and process all single items first. Eclat on the other hand will seek to add as many items to the original single item as it can, until it fails to reach the set minsup.

This approach is faster and less intensive in regards to computation and memory but the itemsets produced are long and difficult to manipulate.

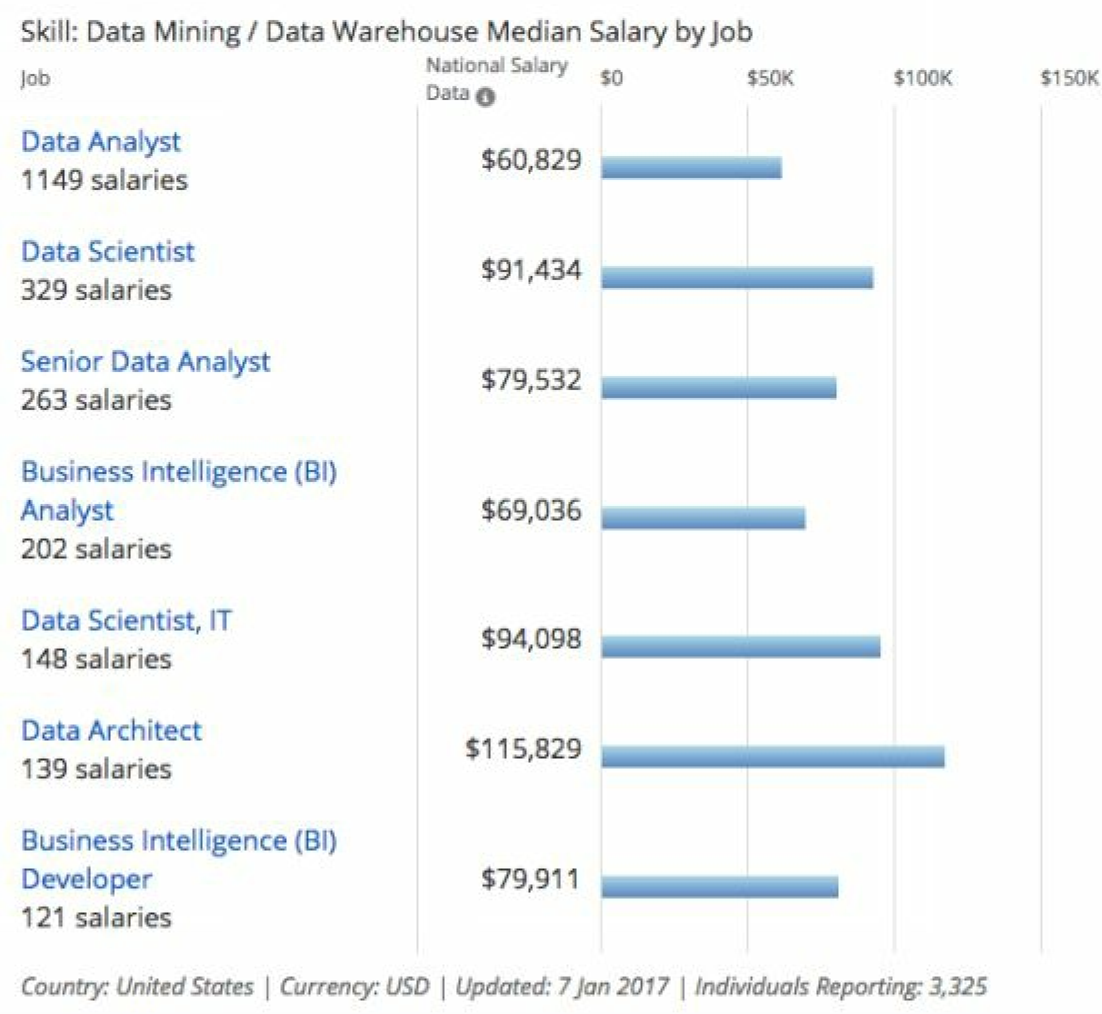
As a data scientist you thus need to form a decision on which algorithm to apply and factor in the trade-off in using various algorithms.

Where to From Here

Career Opportunities in Machine Learning

It is natural to associate 'big data,' 'artificial intelligence' and 'machine learning skills' with success and a big pay packet.

Six figure incomes are relatively standard for data science professionals in places like the U.S and this is by no means the top bracket of industry talent.



Source: [Payscale]

Given the speed at which artificial intelligence is taking over almost every industry, many more jobs are going to become redundant.

This book is by no means claiming that you have to learn data science to keep your job, and that everyone needs to understand machine learning programming to have a job in the future. Data science programs and machine learning teams are ultimately just one contingent of a business, government department or sporting club.

However the rate at which artificial intelligence is integrating into all aspects of organizational activities - from marketing to human resources - means that now has never been a more important time to be studying machine learning.

As a CEO, an online marketing professional, a politician, a professional coach or a decision maker in your organization it’s extremely important to understand machine learning. This includes the various processes, resources, advantages and limitations of machine learning.

Career opportunities in machine learning are both expanding and becoming more lucrative at the same time. Due to current shortages in qualified professionals and the escalating demand for experts to manage and mine data the outlook for machine learning professionals is bright.

To continue your path to working in machine learning you will need both a strong passion for the field of study and dedication to educate yourself on the various facets of data science.

There are various channels in which you can start to train yourself in the field. Identifying a university degree, an online degree program or online curriculum are common entry points.

Along the way it is important to seek out mentors who you can turn to for advice on both technical machine learning questions but also on career options and trajectories.

A mentor could be a professor, colleague, or even someone you don’t yet know. If you are looking to meet data scientists with more industry specific experience it is recommended that you attend industry conferences or smaller offline events held locally. You could decide to attend either as a participant or as a volunteer. Volunteering may in fact offer you more access to certain experts and save admission fees at the same time.

Linkedin and Twitter are terrific online resources to identify professionals in the field or access leading industry voices. When reaching out to established professionals you may receive resistance or a lack of response depending on whom you are contacting.

One way to overcome this potential problem is to offer your services in lieu of mentoring. For example, if you have experience and expertise in managing a Wordpress website you could offer your time to build or manage an existing website for the person you are seeking to form a relationship with.

Other services you can offer are proof reading books, papers and blogs, or interning at their particular company or institute. Sometimes its better to start your search for mentors locally as that will open more opportunities to meet in person, to find local internship and job opportunities. This also conveys more initial trust than say emailing someone across the other side of the world.

Interviewing experts is one of the most effective ways to access one-on-one attention with an industry expert. This is because it is an opportunity for the interviewee to reach a much larger audience with their ideas and opinions. In addition, you get to choose your questions and ask your own selfish questions after the recording.

You can look for local tech media news outlets, university media groups, or even start your own podcast series or industry blog channel. Bear in mind that developing ongoing content via a podcast series entails a sizeable time commitment to prepare, record, edit and market. The project though can bear fruit as you produce more episodes.

Quora is an easy-to-access resource to ask questions and seek advice from a community who are naturally very helpful. However, do keep in mind that Quora responses tend to be influenced by self-interest and if you ask for a book recommendation you will undoubtedly attract responses from people recommending their own book!

However, there is still a wealth of non-biased information available on Quora, you just need to use your own judgement to discern high value information from a sales pitch.

In regards to specific careers in data science, popular job titles include Data Scientist, Business Intelligence Architect, Business Analytics Specialist and Machine Learning Specialist.

Data Scientist

National U.S Average Salary: $63,632 - $138,782

Data science is a broad term, and a data scientist is an equally general job title. As a generalist, the key role of a data scientist is to collect as much relevant data as possible to conduct analysis on past performances to attempt to predict the future.

Compared to other more specialized jobs in data science, there are less entry requirements to finding employment as a data scientist. Reasonable training in computer science or statistics should be sufficient to find an entry-level work position.

A postgraduate degree, such as a Master's degree in Data Science, would be of advantage but not strictly required.

Another key competency to becoming a successful data scientist is strong communications skills. You need to be able to competently present findings to decisions makers.

Data scientists also have promising potential to grow into leadership positions within a company given their knowledge of the company’s performance metrics.

Business Intelligence Architect

National U.S Average Salary: $78,556 - $140,165
Bonus: $1,994 - $19,928
Profit Sharing: $-0.50 - $22,510
Total Pay: $80,303 - $152,210

A Business Intelligence Architect or ‘BI’ is responsible for collecting, managing and processing corporate data, as well as communicating and providing actionable information to decision leaders within the company.

Business Intelligence Architect positions are generally offered to experienced data science professionals, and not as an entry-level position. A Business Intelligence Architect will most often work above a technical team or as a senior member of the team.

Their main responsibility is to plan and execute a system to maximize the full value of their company's data assets. The architectural aspect of this role – and hence the name – is to design a system that can pool together relevant data from numerous stand-alone data collection points.

The next aspect of the job is to synthesize the data through various processing systems to produce meaningful insights. The final part is to then effectively communicate those insights to decision makers.

Machine Learning Scientist/Engineer

National U.S Average Salary: $65,436 - $163,091

Machine Learning Scientists (or Engineers) are responsible for programming computers to learn on their own. Given the inherent complexities of programming a computer how to think, this job title is well paid but entails higher requirements.

To work as a Machine Learning Scientist it is important that you are not only creative, organized and have a high attention to detail, but be well trained. Technical requirements include expertise in programming languages such as Python, C++, Java and R.

As Machine Learning Scientists are often working on cloud computing infrastructure you will also need to be familiar with cloud technology and distributed computing software such as Hadoop.

Sound training in statistics, probability and math skills are other essential credentials.

Business Analytics Specialist

National U.S Average Salary: $65,115 - $128,800

A business analytics specialist straddles both the business and technical aspects of data mining to implement a strategy set by the company's BI architecture. If a company does not have the resources to hire a BI architect and implement a customized architecture, then a business analytics specialist will depend on third party software products to integrate business analytics capabilities into the company.

Degrees & Certifications

Recommended Degrees in the U.S:

Southern Methodist University, Dallas, Texas
Online Master of Science in Data Science
Available online over 20 months. Ranked a Top National University by US News.

Syracuse University, Syracuse, New York
Online Master of Science in Business Analytics
Available online. GMAT waivers available.

Syracuse University, Syracuse, New York
Online Master of Science in Information Management
Available online. GRE waivers available.

American University, Washington DC
Online Master of Science in Analytics
Available online. No GMAT/GRE required to apply.

Villanova University, Villanova, Pennsylvania,
Online Master of Science in Analytics
Available online.

Purdue University, West Lafayette, Indiana
Master of Science in Business Analytics and Information Management
Full-time 12-month program. Eduniversal ranks Krannert's Management Information Systems field of study #4 in North America.

University of California-Berkeley, Berkeley California
Available Online. #1 ranked public university by US News

Final Word

Now has never been a better time to dive into data science and learn machine learning.

Despite the rigorous training required, machine learning can bring immense personal rewards financially, and help to solve business and global problems.

This book I hope has also helped to ease you into the field of data science and translate machine learning theory into layman's terms.

I hope you enjoyed this book and I wish you all the best with your future career in machine learning.

Many thanks,

Oliver Theobald

Table of Contents