**How Data Science is saving lives**

Using Data Science techniques to understand and analyze the large data sets available today has a huge impact on human lives. It can provide targeted information to help healthcare professionals give the best treatment to patients, or help predict natural disasters so that people can prepare early, and much more besides. In healthcare, data scientists use predictive analytics developed from data mining, data modeling, statistics, and machine learning to find the best options for patients. This type of predictive analytics examines all known factors for a disease, including gene markers, associated conditions, and environmental factors. It then recommends appropriate tests, suitable trials, and any suggested treatments. Every individual physician has their own store of knowledge gained from their studies, interests, and experiences. Data science systems that use predictive analytics ensure that all physicians can also access the latest information about the disease, tests, and treatment plans, tailored to their specific patient. With this type of system, every physician has access to the same knowledge, and the best options can be consistently offered, improving patient outcomes. For example, a study by the Boston Consulting Group and AdvaMedDx, an industry association of medical diagnostics companies, examined the barriers to the adoption of potentially lifesaving diagnostic tests for patients with a specific cancer and a particular gene marker. The study discovered that the biggest factor in the patient being offered a specific test was the patient’s oncologist, who may or may not have known about the test and its relationship to the gene marker. By providing extra information through data science tools, physicians can be made aware of the most helpful tests and treatments for a specific patient. There are many opportunities to explore other ways to mine data, such as from electronic medical records for different types of medical research. Schools such as the NorthShore University HealthSystem in suburban Chicago, a leader in the implementation of Electronic Medical Records (EMR) systems, now offer guidance on data mining. It is the first healthcare provider in America to be awarded the highest level of EMR deployment for both inpatient and outpatient care. This remarkable effort has generated much-anonymized data available for innovative analytics research. Developing more sophisticated big data analytics capabilities helps healthcare organizations move from basic descriptive analytics towards predictive insights, thanks to data science. In the field of Disaster Preparedness, the ability to save lives using Data Science tools has been under development for many years. The use of predictive analytics tools is improving and providing new data analysis in a multitude of ways, alerting populations to danger faster than ever before. Large, high-quality data sets can be used to predict the occurrence of numerous types of natural disasters, which can be the difference between life and death for thousands of people. Earthquakes, hurricanes & tornados, floods, and volcanic eruptions can be predicted with the help of data science. Recent research at the University of Warwick in the UK used social media content such as photos and keywords to track the development of floods, hurricanes and other weather events. When added to the information recorded by scientists and weather stations, this type of data can be used to improve the predictions for localised weather events. Because the real benefit of this knowledge is so important, schools are starting to include this type of data science education in their curriculum. For instance, the University of Chicago Graham School offers a Master of Science course in Threat and Response Management. Data science tools enable organizations to analyse vast quantities of data from widely different sources, and present that information in a way that allows data scientists to gain new knowledge, in some cases, saving hundreds of lives.

**How Should Companies Get Started in Data Science?**

At the end of the day, for businesses, they know one thing, that if they are unable to measure something, they are unable to improve it. And if they are unable to measure their costs, they are unable to reduce them. If they're unable to measure their profits, they are unable to increase them. So the first thing a company has to do is to start recording information, start capturing data, data about costs. And the differentiate it by labor costs and material cost, the cost to how much it cost to sell one product and the total cost. And then you look at the revenue, where's your revenue coming from? Is 80% of your revenue coming from 20% of your customers? Or is it the other way around? So first thing first, start capturing data. Once you have data, then you can apply algorithms and analytics to it. So the first thing to do would be to capture data. If you're not capturing it, start capturing it. If you're capturing it, archive it. Do not overwrite on your old data thinking you don't need it anymore. Data never gets old. Data is always relevant, even if it's 100 years old, 200 years old. It is relevant to you and and your firm and your success. So keep data, capture it, archive it, make sure nothing goes to waste. Make sure there's a consistency. So someone 20 years later trying to understand, that data should be able to do so, so have proper documentation. Do it now. Put the best practices for data archiving in place the moment you start a business. And if you're already in business and you haven't done it, do it now. >> Start measuring things. Too many companies haven't measured things properly for a decade and, then they decide, they want data science. Data science inside a company is only going to be as valuable as the data collected. Garbage in, garbage out is a rule in any sort of analysis. >> If something is not measured, it's very difficult to improve it or to change it. So the very first step is measurement. If companies have existing data, then they should start looking at it and cleaning it. If they don't have existing data, then they need to start collecting it. >> I think to look for a team who love to work as a data scientist. >> The first stop is to have employees, that they are interested on data science. because if you don't have interest in your company, you will not have engagement. >> Companies should remember, that it's key to have a team. So it's not one data scientist, but a team of them, that each of them have strengths in different areas of data science.

**Applications of Data Science**

Data science and big data are making an undeniable impact on businesses, changing day-to-day operations, financial analytics, and especially interactions with customers. It's clear that businesses can gain enormous value from the insights data science can provide. But sometimes it's hard to see exactly how. So let's look at some examples. In this era of big data, almost everyone generates masses of data every day, often without being aware of it. This digital trace reveals the patterns of our online lives. If you have ever searched for or bought a product on a site like Amazon, you'll notice that it starts making recommendations related to your search. This type of system known as a recommendation engine is a common application of data science. Companies like Amazon, Netflix, and Spotify use algorithms to make specific recommendations derived from customer preferences and historical behavior. Personal assistants like Siri on Apple devices use data science to devise answers to the infinite number of questions end users may ask. Google watches your every move in the world, you're online shopping habits, and your social media. Then it analyzes that data to create recommendations for restaurants, bars, shops, and other attractions based on the data collected from your device and your current location. Wearable devices like Fitbits, Apple watches, and Android watches add information about your activity levels, sleep patterns, and heart rate to the data you generate. Now that we know how consumers generate data, let's take a look at how data science is impacting business. In 2011, McKinsey & Company said that data science was going to become the key basis of competition. Supporting new waves of productivity, growth, and innovation. In 2013, UPS announced that it was using data from customers, drivers, and vehicles, in a new route guidance system aimed to save time, money, and fuel. Initiatives like this support the statement that data science will fundamentally change the way businesses compete and operate. How does a firm gain a competitive advantage? Let's take Netflix as an example. Netflix collects and analyzes massive amounts of data from millions of users, including which shows people are watching at what time a day when people pause, rewind, and fast-forward, and which shows directors and actors they search for. Netflix can be confident that a show will be a hit before filming even begins by analyzing users preference for certain directors and acting talent, and discovering which combinations people enjoy. Add this to the success of earlier versions of a show and you have a hit. For example, Netflix knew many of its users had streamed to the work of David Fincher. They also knew that films featuring Robin Wright had always done well, and that the British version of House of Cards was very successful. Netflix knew that significant numbers of people who liked Fincher also liked Wright. All this information combined to suggest that buying the series would be a good investment for the company. They were right. It was a huge hit. Thanks to data science, Netflix knows what people want before they do.

**How Can Someone Become a Data Scientist?**

A real data scientist, the high-end data scientists, are mostly PhDs. They often come out of physics, out of statistics, they have to have a computer science background, they have to have a math background, they have to know about databases and statistics and probability and all that stuff. However, if you're coming into a data science team, I think the first skills you need is you need to know how to program, at least have some computational thinking, so having taken a programing course, you need to know some algebra, at least up to analytics, geometry, and hopefully some calculus, some basic probability, some basic statistics, I mean really have to understand the difference and different statistical distributions, and database. I mean, one of the easiest places to start is relational databases, which stores lots and lots of our data so people can first walk before they can run by at least understanding about computers and databases and how we store things and if you understand relational databases nowadays you can still, just with that understanding, use big data clusters as if they were just a big relational database. You don't have to really have understand the whole MapReduce programming model. But then, as you go further up in the field, then you have to know a lot of computer science theory and statistics, it's really, and probability, it's really the intersection of them that the high end data scientists, the PhD data scientists work with.

I do a lot of self-learning. I think everybody these days, I mean, I learned about Hadoop all by myself, I read some articles, I watched some videos, I thought, I played, although I'm a builder, I'm a tinkerer, so if I wanna figure out how to do something, I build it. I mean, my first HPC cluster I heard about this term a Beowulf cluster, I mean, yeah, what the hell's that? So I looked it up and said, oh, it's just a bunch of computers hooked together with a TCP/IP network, that's pretty easy, so we get a grant from Citi Bank and we built a five thing cluster and I said, oh, well, that's HPC. I said, I had one of the first HPC clusters at the university, it was tiny but a lot of our researchers loved it because they could run stuff 40 and 50 times faster. So I think one of the ways you learn things is you do them, you have to do them, and these online learning platforms especially now that we have things like IPython and Jupyter Notebooks and I guess Zeppelin means that you can actually go in and take some of these courses and you can do things right then and you can see them and feel them and play with them and, at that point, you know, you'll start to get your head around what is actually happening. Motivation is the key problem in all of these, is how to keep people motivated and I think the badge system that the, what was it, Big Data University has, is one of the ways is how do you get people to keep going through. But if they want to, they can. It's up to the individual to. So they have to understand what the goal is.

The place it can't sit is probably under the CIO, the Chief Information Officer. CIOs current chief information officers in many companies got there from an accounting background or a finance background, they're clueless. Sorry. But they really, it has to come out of the research side. So you'll find data scientists primarily in companies that have some research agenda, pharmaceuticals, finance, all of, any technology company. If you look at, we can't keep some of our PhD data scientists in our program, they are now at Facebook, they're at Linkedin, they're at Uber, they're at Lyft, because the demand out there for the PhD level data scientist is just unbelievable. They make large amounts of money and they're playing with problems that are really, really neat. How do you schedule the Uber cars? You have enormous amounts of data.

**Recruiting for Data Science**

When the companies are hiring people for a data science team, maybe a data scientist or an analyst, or a chief data scientist, the tendency would be to find the person who has all the skills, that they know the domain-specific knowledge. They're excellent in analyzing structured and unstructured data. And they're great at presenting and they've got great storytelling skills. So if you put all this together, you will realize you're looking for a unicorn. And your odds of finding a unicorn are pretty rare. I think what you need to do to is to see, given the pool of applicants you have, who has the most resonance with your firm's DNA. Because you can teach analytics skills, anyone can learn analytics skills if they dedicate time and effort to it. But what really matters is who's passionate about the kind of business that you do. Someone could be a great data scientist in the retail environment, but they may not be that excited about working in IT related firms or working with gigabytes of weblogs. But if someone is excited about those weblogs, if someone is excited about health-related data then they would be able to contribute to your productivity much more so. And I would say if I'm looking for someone, if I have to put together a data science team, I would first look for curiosity. Is that person curious about things not just for data science but anything like, are they curious about why this room is painted a certain way, why do the bookshelves have books, and what kinds of books? They have to have a certain degree of curiosity about everything that is in their vision, that they look at. The second thing is do they have a sense of humor because, you see, you have to have a lighthearted about it. If someone is too serious about it, they probably would take it too seriously, and would not be able to look at the lighter elements. The third thing I think, and I think the last thing that I would look for if I had to have a hierarchy, the last thing I would look for are technical skills. I would go through the social skills, curiosity, and sense of humor. The ability to tell a story. The ability to know that there is a story there. And then once all is there then I would say, well, can you do the technical side of it? And if there is some hope or some sign of some technical skills, I would take them because I can train them in whatever skills they need. But I cannot teach curiosity. I cannot teach storytelling. I cannot certainly, instill sense of humor in anyone. >> I think there's no hard and fast rule for hiring data scientists. I think it's going to be a case by case thing. I would say there has to be some sort of technical component, somebody should be able to work with and manipulate the data. They should be able to communicate what they find in the data. I find quite often nobody really cares about the r-square or the confidence interval. So you have to be able to introduce those things and explain something in a compelling way. And they also have to find somebody who is relatable, because data science, it been typically new means that the person in that role has to make relationships and they have to work across different departments. >> If these data scientist has a good mathematics and statistics background.>> They have to consider like problem solving abilities and analysis. The scientist needs to be good in analyzing problems.>> The persons they are hiring, they should love to play with data. And then they know how to play with the data visualization. They have analytical thinking.>> When a company is hiring anyone to work on a data science team, they need to think about what role that person is going to take. Before a company begins, they need to understand what they want out of their data science team. And then they need to hire to begin it. As they grow a data science team, they need to understand whether they need engineers, architects, designers to work on visualization. Or whether they just need more people who can multiply large matrices. >> From a skills point of view, let's focus on the technical skills and in that case, first thing would be what kind of a technical platform would you like to adopt? Let's say you want to work in a structured data environment and let's say you want to work in market research. Then the type of skills you need are slightly different than someone who would like to work in big data environments. If you want to work in the traditional market research data, structure data environment, your skills should be some statistical knowledge and some knowledge of basic statistical algorithms, maybe some machine learning algorithms. And these are the tools that you would like to develop. If you want to work in big data, then there's the other aspect of it and that is to be able to store data. So you start with the expertise in storing large amounts of data. And then you look into platforms that allow you to do that. The next step would be to be able to manipulate large amounts of data, and the final step would be to apply algorithms to those large sets of data. So it's a three-step process. But most likely it starts, most importantly, it starts with where you would like to be, in what field, in what domain. In terms of platforms, let's you want to be in the traditional predictive analytics environment, and you're not working with big data, then R or Stata, or Python would be your tools. If you're working mostly with unstructured data, then Python is most suitable than R. If you're working with big data, then Hadoop and Spark are the environments that you will be working with. So it all depends upon where you would like to be and what kind of work excites you and then you pick your tools. In addition to technical skills, the second aspect of the data science is to have the ability to communicate. The communication skills or presentation skills. I call them story telling skills, that is that you have your analysis done, now can you tell a great story from it? If you have a very large table, can you synthesize this and make it more appealing that when it goes on the screen, or is it part of a document that it just speaks? It sings the findings and the reader just gets it right there. So the ability to present your findings, either verbally, or in a presentation, or in a document. So those communication and presentation skills are equally important as the technical skills are. When you have a grading side, when you're presenting your results, imagine you're driving on a mountain and then there's a sharp turn. And you can't see what's beyond the turn. And then you make that turn and then suddenly, you see a tremendous valley in front of you. And this great sense of awe, that I didn't know that, right? So when you present your findings and you have this great finding and you communicate it well, this is what people feel because they were not expecting it. They were not aware of it, and then this great sense of happiness that now I know. And I didn't know this, now I know. And then it empowers them, it gives them ideas, what they can do with this knowledge, this new insight. It's a great sense of joy. And you are able as a data scientist, you are able to share with your clients because you enabled it.

**Careers in Data Science**

The emergence of Internet of things and advances in distributed computing have brought vast amounts of data and the technological capability to analyze it. Now that we can extract useful insights and new knowledge, we need to know how to shape that data to focus on what to do with it and what it can do for us. Enter data science. Companies like LinkedIn, Glassdoor, Indeed, and Dice track employment trends which show a career in data science moving up the list of most promising jobs to become number one since 2016. It remains one of the top three career choices for 2020. Dice noted that job postings are from companies in a wide variety of industries, not just tech. Global Industry Analysts Incorporated predicts that the data science platform market will grow by $314.8 billion US by 2025, driven by a compounded growth of 38.2%. McKinsey Global Institute warned of huge talent shortages for data and analytics by 2018. Forrester Research Analyst Brandon Purcell said, in January of 2019, the demand for data scientists will only grow as organizations increasingly rely on data-driven insights. We're now well into that period, and recruiters are finding it difficult to fill the growing need for talented data scientists. What motivates someone going into a data science? For one thing, data science applies to almost any discipline. So if you have the aptitude and desire to work with data, enjoy coding, have no problem learning math and statistics, and you are a good storyteller, then you can certainly enter a data science field and excel. For most people, this means acquiring additional tools and skills and continuously learning about new tools and techniques in the field. The women in data science initiative spearheaded by the Stanford Institute for Computational and Mathematical Computing have committed to inspire and educate data scientists worldwide, regardless of gender and to support women in the field. When you are seeking a career in data science, you need to make sure your skill set matches the role you are targeting. You can tailor your skill set to the specific area you want to enter, adding missing skills via one of the many excellent online training resources. Then you'll be prepared for a fascinating and rewarding career. So now it is time to move into this field, when there are such diverse choices available and education resources that make it a reality.

**High School Students and Data Science Careers**

Learn how to program. Learn some math. Take a course in probability. Learn a little bit of statistics. And then, play. Build something, write something. I mean, when I say build, programming and building systems, building things isn't just physical, right? You can build computer systems, statistical systems, whatever. But once you try to do something, then you'll know what tools you need, right? And you'll say, "Oh, oh my god, what? "There's this expression there, "what does an inner product mean? "What's that? "How do I, oh, okay, I can learn that." And then when they get to college, they will have a big jump on many of the other college students. And so when they get out of college, they'll have an even bigger jump, and then make a lot of money. And they'll be happy, too. This stuff is fun, right? It's fun. If you're in high school and you're considering data science, I would say get familiar with data bases, start learning SQL, start thinking about, you know, computer science, if that's interesting. If you have a computer science course in your school, you know, take it, and that's a good part of being a data scientist. Beyond that there are probably ways to foster your creativity, right, your curiosity. If you like detective games, that's kind of cool, right. And if you like treasure hunts or whatever, right, if you're into that stuff, I think you'll, and you get the opportunity to do that stuff, that will help you as a data scientist because it's a really a good way to kind of make sure that you can be curious as you go about your daily life. (Music) Encourage the curiosity, encourage the experimentation. It's kind of like science fairs, science fairs are great, they encourage that experimentation, that learning from, asking a question and answering it through a scientific method, but doing that with data sets rather than vinegar volcanoes. It's kind of the same thing, but learning from data and we're going through an election season right now, there's a lot of stuff in the news about polls and survey results and that's a great way to start a conversation and talk about how do the people who ran the polls, how do they know, how can they predict what's going to happen in the election. So that's another cool way to start a conversation about data science. (Music) I would say encourage the person who is interested in data science because to pursue to, because it's a great career and it is something that is definitely going to be in need in the future. It's one of those highly sought after knowledge professions that are really important to businesses around the world. So being a data scientist and being able to help companies as they grow and learn how do to things more efficiently or how to do things smarter, there will always be a need for people like that. And data scientists are those people. (Music) I would say that I understand what you're talking about because I was never a great mathematics student as well. And I think there's actually a bunch of data scientist, who are really successful and popular, who are in the same boat. You know there's kind of arithmetic and math in school is not necessarily everybody's best subject. But when you combine it with, you know these aren't just hypothetical numbers, these aren't just, problem statements that you have no connection to. When you have a connection to the problem, it suddenly becomes much easier to use math to help understand it, I found. And so you know, knowing the people who will benefit from the math that you do I think is really cool.