DATA SCIENCE IN INDUSTRY

Data Science Applications in An Industry

The role of Data Science Applications hasn't evolved overnight. Thanks to faster computing and cheaper storage, we can now predict outcomes in minutes, which could take several human hours to process.A Data Scientist gets home a whopping $124,000 a year and they owe it to the deficiency of skilled professionals in this field. This is the reason why Data Science Certifications are at an all-time high!Through this Article, we bring to you, applications that build upon the concepts of Data Science, exploring various domains such as the following:

- Fraud and Risk Detection
- Healthcare
- Internet Search
- Targeted Advertising
- Website Recommendations
- Advanced Image Recognition
- Gaming
- Augmented Reality
- Banking
- Finance
- Manufacturing
- Transport
- E-Commerce

Fraud and Risk Detection

The earliest applications of data science were in Finance. Companies were fed up with bad debts and losses every year. However, they had a lot of data which used to get collected during the initial paperwork while sanctioning loans. They decided to bring in data scientists to rescue them out of losses.Over the years, banking companies learned to divide and conquer data via customer profiling, past expenditures, and other essential variables to analyse the probabilities of risk and default. Moreover, it also helped them to push their banking products based on customer's purchasing power

HealthCare

The healthcare sector, especially, receives great benefits from data science applications.

- **Medical Image Analysis**

  Procedures such as detecting tumours, artery stenosis, organ delineation employ various methods and frameworks like MapReduce to find optimal parameters for tasks like lung texture classification. It applies machine learning methods, support vector machines (SVM), content-based medical image indexing, and wavelet analysis for solid texture classification.

DATA SCIENCE IN INDUSTRY

- **Genetics and Genomics**

Data Science applications also enable an advanced level of treatment personalization through research in genetics and genomics. The goal is to understand the impact of the DNA on our health and find individual biological connections between genetics, diseases, and drug response. Data science techniques allow integration of different kinds of data with genomic data in disease research, which provides a deeper understanding of genetic issues in reactions to drugs and diseases. As soon as we acquire reliable personal genome data, we will achieve a deeper understanding of human DNA. The advanced genetic risk prediction will be a major step towards more individual care.

- **Drug Development**

The drug discovery process is highly complicated and involves many disciplines. The greatest ideas are often bounded by billions of testing's, huge financial and time expenditure. On average, it takes twelve years to make an official submission.

Data science applications and machine learning algorithms simplify and shorten this process, adding a perspective to each step from the initial screening of drug compounds to the prediction of the success rate based on the biological factors. Such algorithms can forecast how the compound will act in the body using advanced mathematical modelling and simulations instead of the "lab experiments". The idea behind the computational drug discovery is to create computer model simulations as a biologically relevant network simplifying the prediction of future outcomes with high accuracy.

- **Virtual assistance for patients and customer support**

Optimization of the clinical process builds upon the concept that for many cases it is not actually necessary for patients to visit doctors in person. A mobile application can give a more effective solution by bringing the doctor to the patient instead.

The AI-powered mobile apps can provide basic healthcare support, usually as chatbots. You simply describe your symptoms, or ask questions, and then receive key information about your medical condition derived from a wide network linking symptoms to causes. Apps can remind you to take your medicine on time, and if necessary, assign an appointment with a doctor.

This approach promotes a healthy lifestyle by encouraging patients to make healthy decisions, saves their time waiting in line for an appointment, and allows doctors to focus on more critical cases

Internet Search

Now, this is probably the first thing that strikes your mind when you think about Data Science Applications.

When we speak of search, we think 'Google'. Right? But there are many other search engines like Yahoo, Bing, Ask, AOL, and so on. All these search engines (including Google) make use of data science algorithms to deliver the best result for our searched query in a fraction of seconds. Considering the fact that Google processes more than 20 petabytes of data every day.

DATA SCIENCE IN INDUSTRY

Had there been no data science, Google wouldn't have been the 'Google' we know today.

Targeted Advertising

If you thought Search would have been the biggest of all data science applications, here is a challenger – the entire digital marketing spectrum. Starting from the display banners on various websites to the digital billboards at the airports – almost all of them are decided by using data science algorithms.

This is the reason why digital ads have been able to get a lot higher CTR (Call-Through Rate) than traditional advertisements. They can be targeted based on a user's past behaviour. This is the reason why you might see ads of Data Science Training Programs while I see an ad of apparels in the same place at the same time.

Website Recommendations

Aren't we all used to the suggestions about similar products on Amazon? They not only help you find relevant products from billions of products available with them but also adds a lot to the user experience. A lot of companies have fervidly used this engine to promote their products in accordance with user's interest and relevance of information. Internet giants like Amazon, Twitter, Google Play, Netflix, Linked in, imdb and many more use this system to improve the user experience. The recommendations are made based on previous search results for a user.

Advanced Image Recognition

You upload your image with friends on Facebook and you start getting suggestions to tag your friends. This automatic tag suggestion feature uses a face recognition algorithm. In their latest update, Facebook has outlined the additional progress they've made in this area, making specific note of their advances in image recognition accuracy and capacity.

"We've witnessed massive advances in image classification (what is in the image?) as well as object detection (where are the objects?), but this is just the beginning of understanding the most relevant visual content of any image or video. Recently we've been designing techniques that identify and segment each and every object in an image, a key capability that will enable entirely new applications." In addition, Google provides you with the option to search for images by uploading them. It uses image recognition and provides related search results.

- Predict flight delay
- Decide which class of airplanes to buy
- Whether to directly land at the destination or take a halt in between (For example, A flight can have a direct route from New Delhi to New York. Alternatively, it can also choose to halt in any country.)
- Effectively driving customer loyalty programs Southwest Airlines, Alaska Airlines are among the top companies who've embraced data science to bring changes in their way of working.

You can get a better insight into it by referring to this video by our team, which vividly speaks of all the various fields conquered by Data Science Applications.

DATA SCIENCE IN INDUSTRY

Gaming

Games are now designed using machine learning algorithms which improve/upgrade themselves as the player moves up to a higher level. In motion gaming also, your opponent (computer) analyses your previous moves and accordingly shapes up its game. EA Sports, Zynga, Sony, Nintendo, Activision-Blizzard have led gaming experience to the next level using data science.

Augmented Reality

This is the final of the data science applications which seems most exciting in the future. Augmented reality.

Data Science and Virtual Reality do have a relationship, considering a VR headset contains computing knowledge, algorithms, and data to provide you with the best viewing experience. A very small step towards this is the high trending game of Pokémon GO. The ability to walk around things and look at Pokémon on walls, streets, things that aren't there. The creators of this game used the data from Ingress, the last app from the same company, to choose the locations of the Pokémon and gyms.

However, Data Science makes more sense once the VR economy becomes accessible in terms of pricing, and consumers use it often like other apps.

Though, not much has been revealed about them except the prototypes, and neither do we know when they would be available for a common man's disposal. Let's see what amazing data science applications the future holds for us! Though, not much has been revealed about them except the prototypes, and neither do we know when they would be available for a common man's disposal. Let's see what amazing data science applications the future holds for us!

Banking

Banking is one of the biggest applications of Data Science. Big Data and Data Science have enabled banks to keep up with the competition. With Data Science, banks can manage their resources efficiently, furthermore, banks can make smarter decisions through fraud detection, management of customer data, risk modelling, real-time predictive analytics, customer segmentation, etc.

Banks also assess the customer lifetime value that allows them to monitor the number of customers that they have. It provides them with several predictions that the business bank will derive through their customers. In case of fraud detection, banks allow the companies to detect frauds that involve a credit card, insurance, and accounting. Banks are also able to analyse investment patterns and cycles of customers and suggest several offers that suit you accordingly.

Furthermore, banks can risk modelling through data science through which they can assess their overall performance. With Data Science, banks can tailor personalized marketing that suits the needs of their clients. In real-time and predictive analytics, banks use machine learning algorithms to improve their analytics strategy. Furthermore, banks use real-time analytics to understand underlying problems that impede their performance.

DATA SCIENCE IN INDUSTRY

Finance

Data Science has played a key role in automating various financial tasks. Just like how banks have automated risk analytics, finance industries have also used data science for this task. Financial industries need to automate risk analytics to carry out strategic decisions for the company. Using machine learning, they identify, monitor, and prioritize the risks. These machine learning algorithms enhance cost efficiency and model sustainability through training on the massively available customer data. Similarly, financial institutions use machine learning for predictive analytics. It allows the companies to predict customer lifetime value and their stock market moves.

Data Science also plays a key role in algorithmic trading. Through rigorous analysis of data, financial institutions can make data-driven decisions. It is also playing an important role in making the customer experiences better for the users. Through extensive analysis of client experience and modification of preferences, financial institutions can create a personalized relationship with their customers.

This is further boosted by the real-time analytics of customers which increases personalization. Through various customer sentiment analysis techniques and machine learning algorithms, we can boost the social media interaction, boost their feedback and analyse customer reviews. Also, the additional machine learning techniques like natural language processing and data mining have contributed to the transformation of information for smarter governance that helps to increase the profitability of businesses.

Manufacturing

In the 21st century, Data Scientists are the new factory workers. That means that data scientists have acquired a key position in the manufacturing industries. Data Science is being extensively used in manufacturing industries for optimizing production, reducing costs and boosting the profits. Furthermore, with the addition of technologies like the Internet of Things (IoT), data science has enabled the companies to predict potential problems, monitor systems and analyse the continuous stream of data.

Furthermore, with data science, industries can monitor their energy costs and can also optimize their production hours. With a thorough analysis of customer reviews, data scientists can help the industries to make better decisions and improve the quality of their products. Another important aspect of data science in industries is Automation. With the help of historical and real-time data, industries can develop autonomous systems that are helpful in boosting the production of manufacturing lines. It has taken away the redundant jobs and introduced powerful machines that use machine learning technologies like reinforcement learning.

Transport

Another important application of data science is transport. In the transportation sector, Data Science is actively making its mark in making safer driving environments for the drivers. It is also playing a key role in optimizing vehicle performance and adding greater autonomy to the drivers. Furthermore, in the transport sector, Data Science has actively increased its manifold with the introduction of self-driving cars.

DATA SCIENCE IN INDUSTRY

Through extensive analysis of fuel consumption patterns, driver behaviour and active vehicle monitoring, data science has created a strong foothold in the transport industry. The self-driving cars are the most trending topic in the world today. With the introduction of autonomy to vehicles through reinforcement learning, vehicle manufacturers can create intelligent automobiles. Furthermore, industries can create better logistical routes with the help of data science. Using a variety of variables like consumer profile, location, economic indicators, and logistics, vendors can optimize delivery routes and provide a proper allocation of resources.

Also, various transportation companies like Uber are using data science for price optimization and providing better experiences to their customers. Using powerful predictive tools, they accurately predict the price based on parameters like a weather pattern, availability of transport, customers, etc.

E-Commerce

E-commerce and retail industries have been hugely benefited by data science. Some of the ways in which data science has transformed the e-commerce industries are-

- For identifying a potential customer base, data science is being heavily utilized
- Usage of predictive analytics for forecasting the goods and services.
- Data Science is also used for identifying styles of popular products and predicting their trends.
- With data science, companies are optimizing their pricing structures for their consumers.
- Data Science is also being heavily used in collaborative filtering, where it forms the backbone of advanced recommendation systems. Using this technique, the e-commerce platforms can provide insights to the customers based on their historical purchases and purchases made by people of the same style. This type of hybrid recommendation systems, consisting of both collaborative and content-based filtering are helping the industries to provide better services to their customers.
- Also, companies are making use of sentiment analysis to analyse the feedback provided by the customers. This makes use of natural language processing to analyse texts and online surveys. Fraud Detection, which is the central role of machine learning in industries, is tailored for finding fraud merchants and frauds in wire-transfers.

History of Industrial Revolution

Industry 1.0

- The first Industrial Revolution represented the period between the 1760s and around 1840.
- The First Industrial Revolution was the transition to new manufacturing processes using water and steam.
- Beneficial in terms of manufacturing a larger number of various goods and creating a better standard of living for some.
- The textile industry was transformed by industrialization, as was transportation.

Industry 2.0

- The beginning of 20th century marked the start of the second industrial revolution

DATA SCIENCE IN INDUSTRY

- The second industrial revolution picked up. Historians sometimes refer to this as “The Technological Revolution” occurring mainly in Britain, Germany, and America.
- During this time, new technological systems were introduced, most notably superior electrical technology which allowed for even greater production and more sophisticated machines.

Industry 3.0

- In the last few decades of the 20th century, the first computer era.
- Around 1970 the Third Industrial Revolution involved the use of electronics and IT (Information Technology) to further automation in production. Manufacturing and automation advanced considerably thanks to Internet access, connectivity, and renewable energy.
- Industry 3.0 introduced more automated systems onto the assembly line to perform human tasks, i.e., using Programmable Logic Controllers (PLC). Although automated systems were in place, they still relied on human input and intervention.

Industry 4.0

- The Internet and telecommunication industry in the 1990’s revolutionized the way we connected and exchanged information.
- The Fourth Industrial Revolution is the era of smart machines, storage systems and production facilities that can autonomously exchange information, trigger actions, and control each other without human intervention.
- This exchange of information is made possible with the Industrial Internet of things (IIoT) as we know it today. Key elements of Industry 4.0 include:
- Cyber-physical system — a mechanical device that is run by computer-based algorithms.
- The Internet of things (IoT) — interconnected networks of machine devices and vehicles embedded with computerized sensing, scanning, and monitoring capabilities.
- Cloud computing — offsite network hosting and data backup.
- Cognitive computing — technological platforms that employ artificial intelligence.
- Industry 4.0 starts to move towards Industry 5.0 when you begin to allow customers to customize what they want.

Industry 5.0

- Less than a decade has passed since talk of Industry 4.0 first surfaced in manufacturing circles, yet visionaries are already forecasting the next revolution — Industry 5.0. If the current revolution emphasizes the transformation of factories into IoT-enabled smart facilities that utilize cognitive computing and interconnect via cloud servers, Industry 5.0 is set to focus on the return of human hands and minds into the industrial framework.
- Industry 5.0 is the revolution in which man and machine reconcile and find ways to work together to improve the means and efficiency of production. Funny enough, the fifth revolution could already be underway among the companies that are just now adopting the principles of Industry 4.0. Even when manufacturers start using advanced technologies, they are not instantly firing vast swaths of their workforce and becoming entirely computerized.

DATA SCIENCE IN INDUSTRY

Green IT (green information technology)

Green IT (green information technology) is the practice of environmentally sustainable computing.

The goals of green computing are similar to green chemistry: reduce the use of hazardous materials, maximize energy efficiency during the product's lifetime, the recyclability or biodegradability of defunct products and factory waste. Green computing is important for all classes of systems, ranging from handheld systems to large-scale data centers.

Many corporate IT departments have green computing initiatives to reduce the environmental effect of their IT operations.

Approaches:

Product longevity:

LCA studies the environmental aspects and potential impacts throughout a product's life cycle (i.e. cradle-to-grave) from raw materials acquisition through production, use and disposal.

The biggest contribution to green computing usually is to prolong the equipment's lifetime. Another report from Gartner recommends "Look for product longevity, including upgradability and modularity." For instance, manufacturing a new PC makes a far bigger ecological footprint than manufacturing a new RAM module to upgrade an existing one.

Data center design:

Data center facilities are heavy consumers of energy, accounting for between 1.1% and 1.5% of the world's total energy use in 2010. The U.S. Department of Energy estimates that data center facilities consume up to 100 to 200 times more energy than standard office buildings.

The U.S. Department of Energy specifies five primary areas on which to focus energy efficient data center design best practices:

- Information technology (IT) systems
- Environmental conditions
- Air management
- Cooling systems
- Electrical systemS

According to a Greenpeace study, data centers represent 21% of the electricity consumed by the IT sector, which is about 382 billion kWh a year.

Software and deployment optimization:

DATA SCIENCE IN INDUSTRY

The efficiency of algorithms affects the amount of computer resources required for any given computing function and there are many efficiency trade-offs in writing programs. Algorithm changes, such as switching from a slow search algorithm to a fast search algorithm can reduce resource usage for a given task from substantial to close to zero.

In 2009, a study by a physicist at Harvard estimated that the average Google search released 7 grams of carbon dioxide (CO_2).

However, Google disputed this figure, arguing instead that a typical search produced only 0.2 grams of CO_2

Resource allocation

Algorithms can also be used to route data to data centers where electricity is less expensive. Researchers from MIT, Carnegie Mellon University, and Akamai have tested an energy allocation algorithm that successfully routes traffic to the location with the cheapest energy costs. The researchers project up to a 40 percent savings on energy costs if their proposed algorithm were to be deployed. However, this approach does not actually reduce the amount of energy being used; it reduces only the cost to the company using it.

Larger server centers are sometimes located where energy and land are inexpensive and readily available. Local availability of renewable energy, climate that allows outside air to be used for cooling, or locating them where the heat they produce may be used for other purposes could be factors in green siting decisions.

Approaches to actually reduce the energy consumption of network devices by proper network/device management techniques are surveyed in. The authors grouped the approaches into 4 main strategies, namely

1. Adaptive Link Rate (ALR),
2. Interface Proxying,
3. Energy Aware Infrastructure, and
4. Max Energy Aware Applications.

Virtualizing:

New virtual technologies, such as operating-system-level virtualization can also be used to reduce energy consumption. These technologies make a more efficient use of resources, thus reducing energy consumption by design. Also, the consolidation of virtualized technologies is more efficient than the one done in virtual machines, so more services can be deployed in the same physical machine, reducing the amount of hardware needed.

Cloud computing

DATA SCIENCE IN INDUSTRY

Cloud computing addresses two major ICT challenges related to Green computing – energy usage and resource consumption. Virtualization, dynamic provisioning environment, multi-tenancy, green data center approaches are enabling cloud computing to lower carbon emissions and energy usage up to a great extent. Large enterprises and small businesses can reduce their direct energy consumption and carbon emissions by up to 30% and 90% respectively by moving certain on-premises applications into the cloud.

Power management:

The Advanced Configuration and Power Interface (ACPI), an open industry standard, allows an operating system to directly control the power-saving aspects of its underlying hardware. This allows a system to automatically turn off components such as monitors and hard drives after set periods of inactivity. In addition, a system may hibernate, when most components (including the CPU and the system RAM) are turned off.

Storage:

Smaller form factor (e.g., 2.5 inch) hard disk drives often consume less power per gigabyte than physically larger drives. Unlike hard disk drives, solid-state drives store data in flash memory or DRAM. With no moving parts, power consumption may be reduced somewhat for low-capacity flash-based devices.

Display:

Unlike other display technologies, electronic paper does not use any power while displaying an image. CRT monitors typically use more power than LCD monitors. They also contain significant amounts of lead. LCD monitors typically use a cold-cathode fluorescent bulb to provide light for the display. Some newer displays use an array of light-emitting diodes (LEDs) in place of the fluorescent bulb, which reduces the amount of electricity used by the display. Fluorescent back-lights also contain mercury, whereas LED back-lights do not.

A light-on-dark color scheme, also called dark mode, is a color scheme that requires less energy to display on new display technologies, such as OLED. This positively impacts battery life and energy consumption. While an OLED will consume around 40% of the power of an LCD displaying an image that is primarily black, it can use more than three times as much power to display an image with a white background, such as a document or web site.

Materials recycling:

Recycling computing equipment can keep harmful materials such as lead, mercury, and hexavalent chromium out of landfills, and can also replace equipment that otherwise would need to be manufactured, saving further energy and emissions. Computer systems that have outlived their particular function can be repurposed, or donated to various charities and non-profit organizations

The recycling of old computers raises an important privacy issue. The old storage devices still hold private information, such as emails, passwords, and credit card numbers, which can be recovered simply

by someone's using software available freely on the Internet. Deletion of a file does not actually remove the file from the hard drive. Before recycling a computer, users should remove the hard drive

Telecommuting:

Teleconferencing and telepresence technologies are often implemented in green computing initiatives. The advantages are many; increased worker satisfaction, reduction of greenhouse gas emissions related to travel, and increased profit margins as a result of lower overhead costs for office space, heat, lighting, etc.

Telecommunication network devices energy indices

The information and communication technologies (ICTs) energy consumption, in the US and worldwide, has been estimated respectively at 9.4% and 5.3% of the total electricity produced. The energy consumption of ICTs is today significant even when compared with other industries. Some study tried to identify the key energy indices that allow a relevant comparison between different devices (network elements). This analysis was focused on how to optimize device and network consumption for carrier telecommunication by itself. The target was to allow an immediate perception of the relationship between the network technology and the environmental effect. These studies are at the start and the gap to fill in this sector is still huge and further research will be necessary.

Supercomputers:

The inaugural Green500 list was announced on November 15, 2007, at SC|07. As a complement to the TOP500, the unveiling of the Green500 ushered in a new era where supercomputers can be compared by performance-per-watt.[85] As of 2019, two Japanese supercomputers topped the Green500 energy efficiency ranking with performance exceeding 16 GFLOPS/watt, and two IBM AC922 systems followed with performance exceeding 15 GFLOPS/watt.

DATA SCIENCE IN INDUSTRY

Social media sites generating largest data

Google:

- 40,000 Google Web Searches Per Second.

Facebook:

- 500 Terabytes Per Day.
- 1.5 billion people are active on Facebook daily
- There are five new Facebook profiles created every second!
- More than 300 million photos get uploaded per day
- Every minute there are 510,000 comments posted and 293,000 statuses updated

Instagram:

- 400 million who are active every day
- Each day 95 million photos and videos are shared on Instagram
- 100 million people use the Instagram “stories” feature daily

YouTube:

- It is also the second-most popular search engine right after Google, racking up over a billion hours of views every day
- Every Second. It has been 5 seconds. 318,287 videos watched. Source. 45 hours of video uploaded. Source. 63,657 hours of video watched.

Twitter:

- 12 Terabytes Per Day.

Amazon:

- $258,751.90 in Sales Per Minute.

DATA SCIENCE IN INDUSTRY

Sigma Technologies:

Ben Runkle, CEO, Sigma Technologies, is trying to resolve a huge problem. The company is consistently losing long-time customers. He does not know why they are leaving, but he must do something fast. He is convinced that in order to reduce his churn, he must create new products and features, and consolidate existing technologies. He calls in his chief data scientist, Dr. Jessie Hughan.

She shows Runkle the most recent transcripts and finds something surprising:

".... Not sure how to export this; are you?"

"Where is the button that makes a new list?"

"Wait, do you even know where the slider is?"

"If I can't figure this out today, it's a real problem..."

It is clear that customers were having problems with the existing UI/UX, and weren't upset due to a lack of features. Runkle and Hughan organized a mass UI/UX overhaul and their sales have never been better.

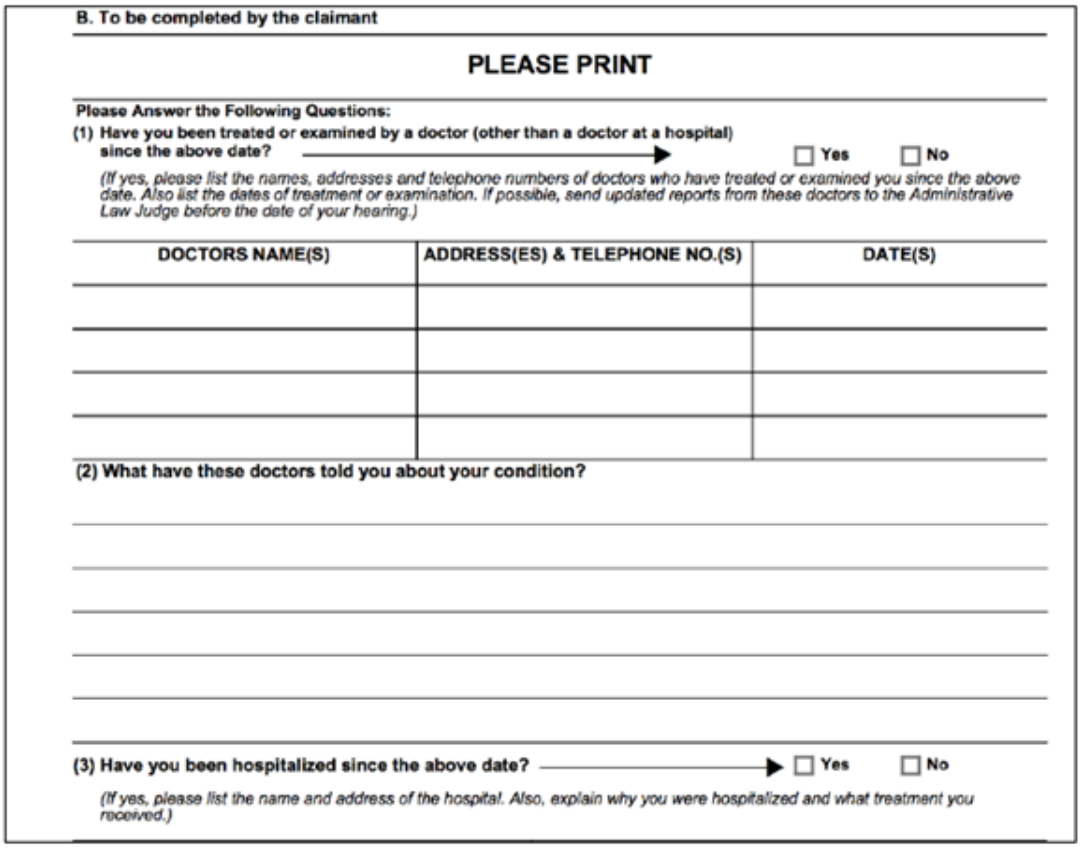
DATA SCIENCE IN INDUSTRY

Data Science case Study - Problem Solving statement:

Case study – automating government paper pushing

Social security claims are known to be a major hassle for both the agent reading it and for the person who wrote the claim. Some claims take over 2 years to get resolved in their entirety, and that's absurd!.

Let's look at what goes into a claim:



B. To be completed by the claimant

PLEASE PRINT

Please Answer the Following Questions:

(1) Have you been treated or examined by a doctor (other than a doctor at a hospital) since the above date? → ☐ Yes ☐ No

(If yes, please list the names, addresses and telephone numbers of doctors who have treated or examined you since the above date. Also list the dates of treatment or examination. If possible, send updated reports from these doctors to the Administrative Law Judge before the date of your hearing.)

DOCTORS NAME(S)	ADDRESS(ES) & TELEPHONE NO.(S)	DATE(S)

(2) What have these doctors told you about your condition?

(3) Have you been hospitalized since the above date? → ☐ Yes ☐ No

(If yes, please list the name and address of the hospital. Also, explain why you were hospitalized and what treatment you received.)

Not bad. It's mostly just text, though. Fill this in, then that, then this, and so on. You can see how it would be difficult for an agent to read these all day, form after form.

There must be a better way!

Well, there is. Elder Research Inc. parsed this unorganized data and was able to automate 20% of all disability social security forms. This means that a computer could look at 20% of these written forms and give its opinion on the approval.

Not only that, the third-party company that is hired to rate the approvals of the forms actually gave the machine-graded forms a higher grade than the human forms.

So, not only did the computer handle 20% of the load, it, on average, did better than a human.