

The Societal and Transformational Impacts of Data Science

Akhil Songa^{*1}, Sailesh Edara², Sai Tanmai Raavi³, Sai Vasavi Harsha Vardhan Gupta Somisetty⁴,

Rahul Bolineni⁵, Sri Teja Kumar Reddy Tetali⁶

^{1,2,3,4,5,6}Student, Department of Computer Science Engineering,
Gitam University, Visakhapatnam,
Andhra Pradesh, India

Abstract: Companies have understood the wonders that can be done with data, and with the Harvard Business Review describing Data Scientist as the Sexiest Job of the 21st Century, Data science is now more than a buzzword. This paper discusses the various stages by which data science is implemented and describes some interesting business case studies on how large organizations like Apple, Facebook, Google, Tesla, Virgin Hyperloop, Netflix, and many more have benefited from Data Science. Data Science can be regarded as the most powerful tool to provide customer satisfaction, thereby guiding companies to hit the jackpot. For instance, a BioPharmaceutical Organization, AstraZeneca, developed fast and effective healthcare methodologies. Walmart made use of data science by which they could predict products that can have high demand in the future resulting in customer convenience/satisfaction. Virgin Hyperloop One could simulate a fast and cost-efficient transport system. Netflix, Spotify, and Amazon Prime could deliver the most efficient user recommendation system for music and shows, thereby increasing their consumer base, and the list is endless.

Keywords: Artificial Intelligence, Data Science, Data-driven approach, Deep learning, Machine Learning, Rule-based approach

I. INTRODUCTION

Data can be defined as information or knowledge of any kind that is stored for further analysis and analytics. In short, any information that is stored for processing and generating insights can be referred to as data. Earlier, all of this data was recorded physically in journals, newspapers, textbooks without any intervention of technology. With the advancement in latest technology, more processing power, cheap storage methods, and ease of access to the internet, the average data being stored has increased exponentially. For instance, a Forbes article published in 2015 stated that the amount of data produced in the past two years is more than the entire history of the human race. It also said that, as of 2015, less than 0.5% of the entire data has been analyzed and used.

The main question lies in how the data that is being generated every second is being used. Here comes the role of a data scientist. A data scientist is a person who is capable of making use of such vast volumes of data and transforming them into useful information. Technically, Data Science can be defined as the study of structured or unstructured data by using various statistical and computational tools in progressive stages and hence derive insights that are useful for businesses to make accurate

decisions or for some other cause.

In data science, one or many of the available Machine learning or deep learning algorithms are used to develop computer models which train on huge data sets and are made ready to predict output for real-life examples. This was not the case earlier. A traditional artificial agent was said to be intelligent if it could access a knowledge base consisting of pre-existing rules provided by the developer and act on as per the instructions in the Knowledge Base. These rules can be thought of as simple If-else conditional statements instructing the machine what to do if it encounters a certain situation. This technique can be called a rule-based approach as the intelligent agent can only follow the rules set by the developer and perform certain actions. It is evident that this approach is more static as there is no scope for the machine to think on its own.

In contrast, with the evolution of Data mining and Data science, the computer models now created are completely dynamic in their working. It means that these models can act according to their previous experiences and act according to the situation. Such models can update their knowledge base if they encounter any new situation and also remember the action it performed. This way, if the same situation is encountered in the future, it can efficiently perform a suitable action. Such models are called Data-driven models as data totally drives their functioning that it takes as input. This paper discusses the various stages followed in retrieving useful information from raw data. Also, various case studies relating to how large-scale organizations like Apple, Google, Netflix, Spotify, and others have benefited from the use of Data Science are discussed in the paper.

II. STAGES

A. Business Understanding

Business Understanding is the preliminary stage for implementing Data Science. Here, all the stakeholders involved in the process gather to define the problem to be solved, discuss the objectives of the project to be carried out and sketch out a draft solution for the problem. After this stage has been completed, every stakeholder is clear about the project objectives and how to achieve them.

B. Analytical Approach

The role of a data scientist becomes prominent in this stage. Here, based on the outcome to be achieved, the data scientist transforms the entire problem into a

statistical orientation, helping the data scientist decide what model to develop. For suppose, the project outcome requires predicting a binary output like a Yes or No, True or False. The Decision Tree is best suited. Different models work efficiently over different kinds of data.

C. Data collection

Data collection refers to the process of gathering information from a variety of sources that help to train a computational model that is capable of performing either descriptive or predictive analysis on real-life scenarios.

Data is prominently collected in two different ways

Primary data collection: Primary data collection is done when working on a new or unique problem. Here, there is either minimum or no data available for that problem. Surveys, Interviews, Group discussions, emails, online forums are the major sources for collecting such data.

Secondary data collection – Unlike Primary data collection, Secondary data collection is done when working on a problem from an existing data source. Kaggle, Gap-minder, Government Census, Magazines, Books, Journals, and News are a few secondary data collection sources.

D. Data Understanding and Processing

This stage involves the process of transforming raw data which is incomplete, noisy, and inconsistent into more useful, efficient, or well-formed data.

Following tasks are done during this stage:

1) Data cleaning: The data cleaning process removes unwanted information, removes noise, filters outliers, and handles any missing values in the data collected. Missing values can be avoided by ignoring the data row, explicitly filling the missing values, using a global constant in place of missing values, or using the mean of the entire data. Imperfections like noise and outliers are handled using the Binning technique.

2) Data integration: It involves compiling data from different sources like databases, flat files, multidimensional databases, and other sources into a single repository of data.

3) Data transformation: The transformation of data into alternate forms thereby making it suitable for model development is called data transformation. For instance, in most cases, strings in the data set like country names are assigned with numbers, normalize the data to change the shape of the distribution of data, scale the data in order to change the range of the data. This process generally involves techniques like Smoothing, Aggregation, Generalization, Normalization

4) Data discretization: Data discretization is the method of transforming continuous data into discrete data. In some situations working with discrete data is better than continuous data. In such situations, data discretization is applied. Binning is one of the Data Discretization techniques.

5) Feature selection: Feature selection is used to find the most appropriate subset of attributes from a huge number of attributes. Feature selection helps the model to train faster as the number of unnecessary or duplicated features decreases, and also it helps to remove unnecessary attributes that act as noise.

E. Modelling

An ML model is a mathematical function formed by finding patterns in a large, diverse dataset that can be used to predict outcomes of future events. This stage involves developing computational models, which can be either a Machine Learning (ML) or a Deep Learning (DL) model. Generally, data scientists develop models for descriptive analytics or predictive analytics. Descriptive analysis is used to analyze, understand past data and help the data scientist gain insights. In contrast, predictive analysis is used in those scenarios where it is required to predict what would happen in the future. Irrespective of whether a descriptive or a predictive model is developed, the following steps are involved

- i. Model selection
- ii. Model training
- iii. Model Evaluation

F. Deployment

Deployment is the most important and last stage in data science, this is a method where a machine learning model is integrated into a production environment for real-time prediction. These machine learning models are deployed using Django, Flask, Streamlit frameworks in platforms like Heroku, PythonAnywhere, Algorithmia, Google Cloud Platform, etc.

III. CASE STUDIES

A. Netflix

Netflix was founded in the year 1997 by Reed Hastings and Marc Randolph in Scotts Valley, California. It is an American-based over-the-top(OTT) content platform and production company [1].

Some of the giant companies like Facebook, Twitter, amazon prime, and Netflix became successful because of the data-driven approaches they use to provide the most profitable outcomes for the customer by utilizing the data powerfully. Most people do not know about the hard work and success story of Netflix, how it is raised from a DVD rental service to one of the most successful online streaming services.

At first, when Netflix was launched in 1997, there was a huge competition going around with blockbusters. The rise of the internet in early 2000 fuelled the success of Netflix. Netflix started to provide content to their customers digitally with the help of the internet. The craze for Netflix began when they started to adopt data-driven approaches to satisfy their customer's needs.

More than 80 percent of the viewers watched content is based on the personalized recommendations that are given for viewers [2]. Netflix's recommendation system strives to suggest to the viewers the best content which they enjoy a lot. The Netflix recommendation system estimates the likelihood of which contains a particular user will watch based on the number of factors like viewing history, how past content titles are rated, genre, actors, year of release, categories, time and date a viewer watches a show, how long a viewer watches the show, in which device viewer is watching the show, the scenes that are watched repeatedly and the list goes on [3].

Netflix is not only an OTT platform; it is also a production company using data science to solve many production problems. By using data science, Netflix helps its production team decide where and when it is best to shoot a movie. For this, Netflix takes the factors like actors and crew availability, venue and budget, scene requirements for the production like weather, day shoot or night shoot, location availability, and current risks at the location. Netflix strives to provide the best decision for the production team, enabling them to manage their expenses and reduce unnecessary costs[4]. Netflix is also available in 190 countries, so their main challenge is managing the network traffic because a long buffer time will vex the viewers [5]. Netflix started to use data science to make sure that the streaming content for viewers is smooth with minimal buffers. Netflix uses real-time data to decide when and what content to cache in regional servers, thereby ensuring faster load times when there is high demand for particular content.

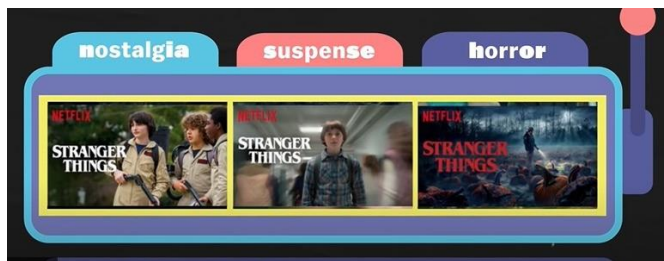


Figure1. Different thumbnails based on genre for the same movie

According to the study of Netflix, a typical viewer spends 1.8 seconds considering each title, and also, Netflix believes that if it does not capture the viewer's attention within 90 seconds, the viewer may lose interest and move to another activity. Upon extensive research, Netflix found that the thumbnails were the deciding factor to grab a user's attention. Netflix already has information regarding every user's interests and preferences, it could easily tailor thumbnails specific to a user's taste. To make this possible, Netflix uses a selection process called Aesthetic Visual Analysis (AVA). This algorithm searches an entire Netflix video for the best possible thumbnail frames. Out of these frames, thumbnails that probably suit a user's taste are displayed on the dashboard. For instance, consider an hour-long "Stranger Things"; it contains nearly 86000 video frames. The AVA analyses all these 86000 frames to find the best possible thumbnails that could be put up for display. Different viewers regularly see changes in thumbnails of the content based on their engagement with previous content. For the same content, the viewer who is a fan of comedy sees the thumbnail depicting a comedy scene from the movie, whereas a viewer who is a fan of horror sees the thumbnail depicting a horror scene from the movie. Also, based on the location of the viewer, the thumbnails may be different. The figure shows the popular thumbnails for the same content in different regions. The figure shows the different thumbnails for the same content based on the viewer's taste [6].

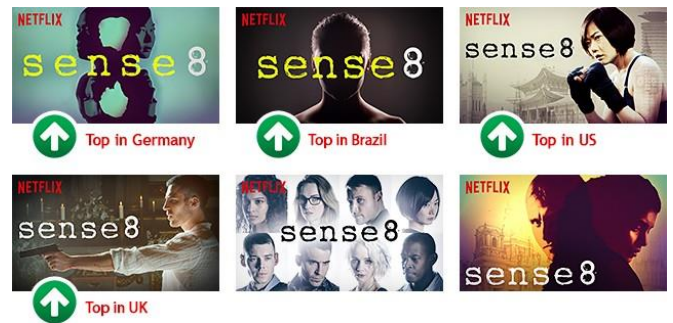


Figure2. Different thumbnails based on the location for the same movie

B. Spotify

Spotify is an audio streaming company founded in 2006. Its headquarters is located in Stockholm, Sweden. As of 2020, Spotify generated nearly 8 billion euros in revenue. Spotify has become the world's most popular streaming service provider with 365 million monthly active users, over 70 million tracks, 2.9 million podcasts, and over 4 billion playlists available in 178 markets and 165 million premium subscribers. Spotify services can be used everywhere, be it Linux, Macintosh, Windows, Android, IOS, or even smart speakers[7,8]. Spotify has faced innumerable challenges during its journey from small music streaming company to the world's largest audio streaming service provider. Their biggest problem was to tackle their competition and increase their consumer base. So they started their research in user music recommendations. Spotify uses the technique popularly known as "Collaborative Filtering". This is one of the Initial and basic approaches to provide music recommendations to users.

Spotify filters out the music based on the liking of a similar group of users to recommend songs, playlists to a new particular user whose taste is similar to the group. This is done by collaborative filtering [9].

As collaborative filtering only works on limited members and is inefficient in recommending newly released music to users, Spotify started to use NLP along with Collaborative Filtering to enhance their recommendation engine. The NLP model understands what is trending in the musical industry by gathering resources from social media, news, blogs, and many more. Based on insights from these sources and the filtering done in collaboration, new songs and albums are recommended to a user. These models can be biased and completely recommend a wrong song to a user if the content from the sources is fake. So, as an advancement to the present audio model, Spotify developed a Convolutional Neural Network fed with technical specifications of an audio file like the Beats Per Minute (BPM), loudness, amplitude, and many more. Now, Spotify recommends a specific song or a playlist to a user only if it can pass the Collaborative Filtering, NLP audio model, and finally the Convolutional Neural Network [10,11].

Following are the various types of recommendations provided by Spotify to its users

1) **Daily Mix:** Selectively tailored playlist refreshed daily based on user taste.

2) **This is:** Curated by music experts, "This is" is a playlist of the most popular songs of an Artist. If anyone would like to switch over to a new artist, they could do it by listening to such playlists.

3) **Recommended Songs:** An endless list of songs that match the tastes and likes of a particular user.

4) **Weekly mix:** This playlist is similar to the "Daily Mix", but the difference is that these playlists update themselves and present new music to the user every week. It updates every Monday

It seems extremely easy to perform the above tasks, but by iteratively collecting data, finding patterns, and with many machine learning engineers, artificial intelligence experts, and data scientists working in collaboration, Spotify is able to deliver such excellent features and continuously stay a step ahead of their counterparts [12].

C. Apple

Apple is an American company founded by Steve Jobs, Ronald Wayne, and Steve Wozniak in 1976.

Back then, Apple was famous for making beautiful, user-friendly computers, but now it is mainly famous for its (privacy, security) smartphones, followed by its computers, Tablets, earphones, and many others.

As generations pass by, Mobile phones are becoming one of the necessities for human beings to survive. Traditionally mobile phones were invented purely for communication purposes, but today they have developed beyond the imagination of Alexander Graham Bell. Most tasks like gaining Knowledge, workout tracking, sleep tracking, alarm, social media, video calling, playing games and much more can be done on today's smartphone. In simple words, a smartphone records our digital lives. With the advancements in technology, these digital records are at a higher risk of being accessed by the unauthorized. A lock that a person himself could access is the need of the hour, and companies like Apple are the best examples for providing such security. Apple always strives to provide privacy to its users irrespective of costs and complications in developing such a system. In this regard, Apple introduced FaceID technology in 2017 through their iPhone X. The pre-existing Touch-ID was discontinued because Apple claims that the chances of a random face unlocking are 1 in 1,000,000. However, by using TouchID, the possibility of unlocking is 1 in 50,000, clearly stating that FaceID is much more secure than TouchID [13].

FaceID technology was developed in-house by Apple for their iPhone lineup starting from 2017 with a motive to increase the privacy of their users. Here, they used special depth sensors called Flood illuminator, Infrared Camera, ambient light sensor and Dot projector. The dot projector projects 30,000 infrared points on the user's face. The reflection of those infrared dots creates a detailed 3D map

of the face and feeds it to Convolutional Neural Networks working to provide efficient authentication. The TrueDepth camera system, neural network, and Bionic chips work together to identify the person [14]. This can adapt to changes in admin's appearance, like Wearing cosmetic makeup, growing facial hair, sunglasses, hats, scarves and even contact lens. This is only possible due to Powerful processors and the implementation of machine learning and deep learning, which cannot be possible by hard programming or any other method that exists now [15].



Figure3. Dot projection on face

Apple makes the most of machine learning to convert handwritten text into digital format. This could be possible with the help of Apple Pencil. Whenever a user writes with an Apple pencil on an iPad, it automatically recognizes the text a user intends to write. Apple gathered a lot of data regarding different handwritings having different strokes, curves and styles from people worldwide and trained their machine learning model. This technology has become so prominent and brought convenience among users as they need not type every time they wish to fill a form, browse through the internet or write an Email. Wherever there is a possibility of inserting text on the device, people can take out their pencil and jot it down on the device as if they are writing it on paper [16].

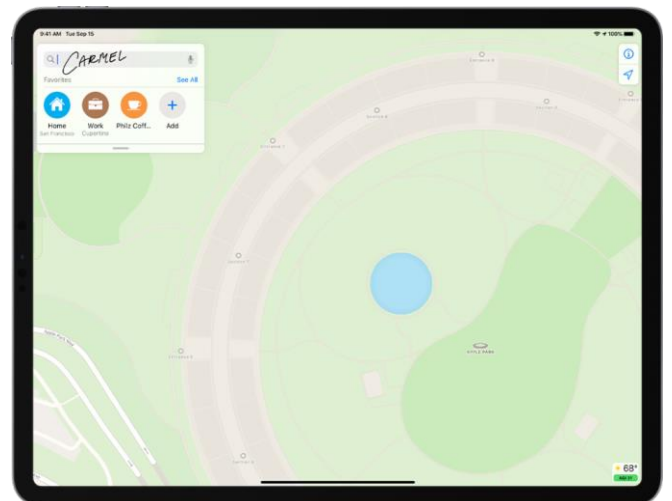


Figure4. Filling the text field by using apple pencil



Figure5. Apple's face recognition in images

Apple uses data science for an interesting use case known as Image Clustering. Generally Image clustering is a technique followed by computer models to group images by recognizing faces in a particular image. Apple's machine learning models perform these tasks efficiently by first detecting the face and then a person's upper body in a given photo. The face and upper body detections are then sent to machine learning models that further analyze the image. Upon satisfying a pre-defined confidence measure, the model clusters the photo into a group containing images of that particular person. The key highlight in the case of Apple's image clustering algorithm is that it can cluster an image even if a person's face is not visible in the image.

Apple uses machine learning for sound recognition features that notify a user when there is a specific sound that is played around them like a fire alarm, baby's cry, running water[17].

Apple also uses machine learning for screen recognition(Apple reads all the components present on a screen) for visually impaired users, virtual assistant(Siri) and voice control(user can control phone, write and edit text with voice over commands).

Apple watch makes extensive use of machine learning for continuously monitoring and detecting abnormal heart rates that could potentially signal an emergency like cardiac arrest and notify the user well before. Fall detection is another such scenario. According to WHO, Every year, an estimated 684,000 individuals die from falls globally. Out of these deaths, over 80% were from low and middle-income countries. The most significant percentage of fatal falls were recorded in Adults who are older than 60 years. Also, over

37.3 million falls occur each year that are severe enough to require medical attention. Whenever a person falls, trips or slips, the Apple watch takes charge immediately and does an SOS emergency call.

Another exciting use case of data science in the Apple watch is the hand wash timer. The world health organization states that a typical hand wash should last for at least twenty seconds for better safety from microorganisms. Apple took this precaution seriously and

developed a feature in their Apple watch that could automatically start a twenty-second timer when the user starts to wash their hands. This feature proved to be beneficial during the pandemic and helped people to stay healthy.

D. Google

Google is a California based Technological company founded in 1998 by Larry Page and Sergey Brin. As of 2020, Google's total revenue is \$182,527 million. Google is popular for its search engine and various other platforms that provide user-specific content. Following are some of the use cases of how Google made use of data science:

1) Pixel Visual Core: Pixel Visual Core(PVC) is a co-processor for smartphones designed only for Pixel phones by Google to eradicate the competition in Mobile Photography. PVC features a dedicated 64-bit cortex - A52 processor, 512 MB LPDDR4 RAM, one PCIe, eight IPU cores, MIPI, which are held only for one single purpose, i.e., for simulating a better image. This technology first appeared in the Pixel 2 and Pixel 2XL series and is available in Pixel 2 and Pixel 3[18]. It is widely accepted that if someone is looking for a smartphone having priority as the camera, the Pixel phone is the only option.

"So I'll just say straight up PIXEL 3 takes the best photos of any smartphone camera better than iPhone XS, Red Hydrogen, Samsung, Lg phone better than Pixel 2" - MKBHD.

Google uses the PVC to run machine learning and neural network algorithms to enhance the phone. Generally, the PVC inputs lots of pixel data from the camera and compares it with each other to give the best possible looking image. Whereas phones like Apple iPhone 8 Plus use dual cameras (one for picture and the other as depth finder) to create a shallow depth of view effect. But the same is achieved by google using a single camera setup along with Convolutional Neural Network(CNN) that has been trained on almost a million pictures of faces and bodies to produce a Depth effect. Later, to increase the speed of processing power, they made Pixel Neural Core which is the successor of Pixel visual core, first introduced on Google Pixel 4. Due to these capabilities, Pixels stands out in the smartphones category concerning photo camera performance [19].



Figure6. Pixel 2 vs Iphone 8 plus camera comparison in 100 and 5 Lux

2) Google assistant: It is one of the “smartest” and the most efficient virtual assistants currently being used. These applications make use of Natural Language Processing (NLP) to understand the meaning of the voice or text of the user interacting with it. Based on the user information, the assistant can periodically understand the behavior, likes, dislikes, interests, and many traits of a particular person. With this, the assistant gains the ability to provide the user with the information he likes. Also, developers at Google have named this application aptly as it performs all kinds of errands that a personal assistant does. For instance, setting reminders, alarms, and astonishingly, booking service on a particular date via a phone call. It means that the Google Assistant is capable of making phone calls, conversing with the recipient over the phone in real-time, understanding the nuances of the conversation, and booking a reservation for a saloon or a restaurant and whatnot. Apart from this, it can draft Emails, send messages and do everything one would expect a virtual assistant to do. Google Assistant can be assumed like a close companion of our's, helping out to better perform our tasks, but in the end, it is typical software. Its capability to provide such user-related content is due to data science and rigorous training it receives from the massive volumes of data [20].

Following are the applications developed by Google that widely make use of Machine learning and Data Science:

i) Google Augmented Reality Microscope

ii) Google Translate

iii) Youtube

Recommendation

iv) Google Photos

v) Google AdSense

vi) Gmail

Google is a pioneer in impacting the lives of a huge population with its products in different domains. Consider Google Augmented Reality Microscope, for instance. It is one of the most sophisticated instruments in the field of healthcare and medical research. This is a unique solution to speed up the time taken by pathologists to analyze tissue samples in their laboratory. Google's Augmented Reality Microscope lends a helping hand to pathologists by analyzing tissue samples in real-time and highlighting any important findings in the specimen. As of now, it could easily predict cancer cells that cause Breast Cancer and Prostate Cancer [21]. Google translate is another such case. Apart from just simple text from one language to another, it uses AI and Machine Learning for efficient translation of text provided as input either through keyboard, voice or camera. Google also uses data science and machine learning on its video streaming platform Youtube. Here, based on the preferences of a user, Youtube recommends videos to its users. As youtube integrates their videos with advertisements, they can earn revenue if a potential customer clicks on an advertisement.

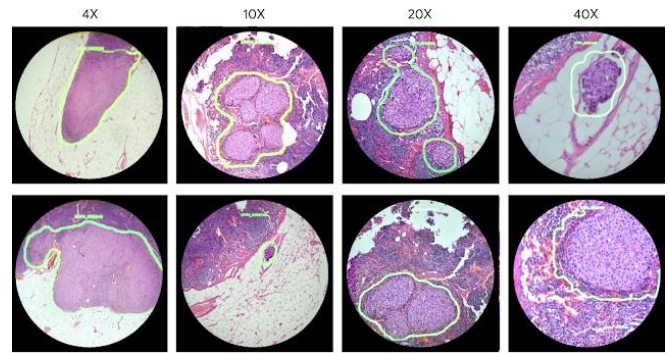


Figure7. Specimen sample under Google Augmented Microscope

E. Facebook

Facebook remains the world's largest social network by a significant margin. With 2.89 billion monthly active users, everyone uses it to stay in touch with their friends, read news, create social events, etc. For an end-user, any of the services provided by Facebook can be used with absolutely no cost. With annual revenue of 86 billion USD, Facebook must have some source of income rather than what the general public uses for connecting and interacting in the social world.

Most of the revenue generated by Facebook comes from those businesses that would like to find customers who have a high chance of buying their product. These businesses pay Facebook to advertise their products. Having all sorts of data regarding nearly 3 billion users, Facebook displays those advertisements on the feed of those users who are likely to be interested in that product or service. As of Quarter 3 of 2020, there are 10 million active advertisers who pay Facebook to put their advertisements with the hope to find their potential customers [22].

Facebook is not the only medium for advertising. There are many modes like Television, Newspaper, Billboard, and others. Then how is Facebook able to attract businesses to advertise on their platform, even if there are multiple options for a business. The cost factor is the key factor attracting small and medium scale businesses to switch to digital marketing rather than traditional ways. The huge reduction in cost is due to the concept of Pay-Per-Click (PPC). It means that organizations like Facebook charge the business organization only if a user clicks on the advertisement displayed on their feed. In traditional marketing, massive sums of money are to be spent even for a commercial irrespective of whether a potential customer views it. This is because the competition is as such, and businesses are spending to their fullest to attract customers to use their service or product. Start-ups and small organizations can not afford such huge sums just for advertisements, and with many such companies starting up, platforms like Facebook serve the need with minimal cost.

It can be seen that the majority of revenue generated for Facebook is through advertisements, so how does it accurately target an advertisement to a particular user. Facebook uses a Data science based tool FB Learner Flow which is often referred to as the “AI Backbone of Facebook”. This tool is designed to generate thousands of

machine learning models whose task is to personalize news feeds, provide targeted advertisements, highlight trending topics, and many more using the humongous data they collect in real-time. This tool can proactively analyze user data like demographics, interests, nature of work, likes, comments, and many more. The FBLeaRner flow is the goto tool for Facebook to put in the right advertisement in front of the right user whose chances of going through the advertisement is almost guaranteed.

Apart from targeted advertisements and personalized feed, Facebook works on the following verticals by using data science:

1) Deep Face: (Best in-class Face recognition algorithm) DeepFace is a Face recognition algorithm built by Facebook to recognize human faces in a given image better. DeepFace generates a representation for a given face using a nine-layer deep neural network. The DeepFace model has been trained with over four million labeled facial images which was the largest, in 2014. Facebook popularly used this technology to introduce something called Auto-Tagging, wherein a particular person was automatically identified and tagged by Facebook until there were litigations on the grounds of user privacy. Their research claimed that this algorithm could recognize the faces of a person with an accuracy of 97.35%, which is almost at par with the human capacity to recognize faces [23].

2) Deep Text: DeepText is another technology developed in-house by Facebook that allows a machine to understand the text's original meaning rather than just searching for keywords in it. For instance, consider writing something like "I need to book a ride". When this sentence is provided to the DeepText Algorithm, it understands that the user wants to book a ride and provides quick options for the user to book a cab. Similarly, if something like "I have just completed my ride" is provided to the text, it can intelligently understand that the user has completed the ride and does not provide any option to book a ride. As an idea, this seems to be very simple, but when it comes to training the machine to derive the meaning on its own, it is a huge task involving multiple layers of neural networks. Similar features are recently being implemented in Google Messages and Gmail [24].

F. Virgin Hyperloop

The hyperloop is referred to as the fifth mode of transportation that plans to transport passengers and commercial goods from one place to another in no time and with less cost and almost zero emissions of greenhouse gases. This concept was first developed in the year 2013 and is now being tested rigorously by Virgin Hyperloop. Take a look at how Data Science is making a pivotal role in shaping the future of Transportation technology [25].

When Virgin Hyperloop One was at the initial stages of development, their R&D team was able to utilize data that was in Megabytes and draw insights from it. As the project progressed and scaled up, Gigabytes of raw data were being piled up at their end. They had to team up with

a data analytics corporate capable of handling large datasets and derive insights from them. Databricks were their choice. Databricks was founded in 2013 by the original creators of Apache Spark™, Delta Lake, and MLflow, the leading software products used for data analytics across the industry.

The data that was regularly collected for analysis were

1) DevLoop Test Track runs. DevLoop is a 500m full-scale hyperloop track developed to run real-time simulations.

2) Socio-economic data to model a nearly perfect price, thereby optimizing operational costs. The main motto of the data team from Databricks working for the Virgin Hyperloop One project was to develop the nearly ideal product that maintains industry standards, ensures safety, and minimizes costs to offer the service to the customers at a comparatively low price.

The Hyperloop data team used the analytics provided by Databricks to run various simulations that helped them predict passenger demand based on multiple features like source and destination, travel time. These simulations were developed on "MLFlow", an open-source platform capable of developing end-to-end machine learning models and integrating with Apache Spark to work with Big Data.

Following are a few key achievements of the data science team of Databricks working on the Virgin Hyperloop One project

1) They reduced the predicted operational costs by 70% even before making their service available to the public in the market.

2) Data processing time was reduced by 95%. For instance, earlier for the Virgin Hyperloop One team, any data analytics project used to take around six months to complete. However, now they complete and start a new project within a week.

3) Data processing time was reduced from hours to minutes.

Virgin Hyperloop highly believes in data science and data visualization as it provides them with the ability to run thousands of simulations quickly within a reasonable time and derive huge insights from them. These insights enabled the entire team of Virgin Hyperloop to make quick and accurate decisions and subsequently bring the vision of the fifth mode of transportation into a reality at the earliest [26].

G. AstraZeneca

AstraZeneca is a BioPharmaceutical company founded in the year 1999, located at Cambridge Biomedical Campus, England. AstraZeneca is popularly known for developing

the Covid-19 vaccine in collaboration with OXFORD university, sold under Covi-shield and Vaxzevria [27].

Apart from the Covid-19 vaccine, AstraZeneca developed a variety of medicines and vaccines for various diseases that fall under the category of Cardiovascular, Renal, and Metabolic, Oncology, Neurology, Respiration, and Gastrointestinal [28].

Apart from being an Active Pharmaceutical Ingredient (API), AstraZeneca believes data and technology are the key ingredients that help them to develop medicines quickly and efficiently. In order to harness the power of vast data being recorded at their end, they have collaborated with institutes like The Cambridge Centre for AI in Medicine, Mila, Schrödinger, MIT MLPDS, MELODY, BenevolentAI, and AI Sweden, which work closely with them to discover and develop medicines using AI [29].

AstraZeneca is one of the pioneers in Pharmaceuticals, and hence there are various projects on which they work simultaneously. But traditionally, pathologists manually analyze each and every tissue sample of the research they make every week, which is obviously a time-consuming task. In order to pace up this process, they have trained and implemented an AI system to assist pathologists in analyzing samples accurately and identify various biomarkers and patterns that are potentially useful to develop a drug. Integration with AI saved nearly 30% of analysis time for AstraZeneca.

In order to develop a new medicine, Randomised Clinical Trials (RCT) is the go-to choice for the pharma industry. Contrastingly, data showed that this method is the most expensive and becomes complex over time. With the advancements in data science, the concept of Electronic Health Records (EHR) is widely adopted. This methodology rapidly improves clinical trials and increases the efficiency of their research. AstraZeneca started implementing Federated EHR technology to refine and replace clinical trials. They assigned AI and machine learning tools to extract more data, analyze, interpret and report on the safety and effectiveness of the clinical trial data [30].

For instance, the AI system developed and being used by AstraZeneca was trained in such a way to detect a biomarker, called PD-L1 in tumor and immune cells, which has the potential to help to detect immunotherapy-based treatment decisions for bladder cancer. AstraZeneca's AI system looks at thousands of images from tissue samples, methodically checking each one for PD-L1. It saves pathologists time and effort to identify biomarkers in a specimen [31].

One of the significant use cases of Data Science in AstraZeneca is developing Knowledge Graphs. A computer model simulates these relationships into a beautiful knowledge graph using vast amounts of scientific data representing relationships between various

biological components like genes, proteins, diseases, compounds, etc. Here, all these components are represented as nodes in a graph-like visualization. Researchers at AstraZeneca use these knowledge graphs to further explore and understand various relationships of a certain disease with other compounds, thereby helping them add to the pre-existing intuition to disease and get new insights that may be potentially useful to develop a drug. For example, in collaboration with BenevolentAI, AstraZeneca has developed an ML platform that helped predict the existence of a novel target in cellular mechanisms that causes chronic kidney disease. Following are the key achievements in the field of medicine and healthcare by using data science:

1. A reduced attrition rate of target identification by nearly 10%
2. Due to Image analysis by AI, the time taken to analyze tissue samples was reduced by 30%
3. Nearly 50% of the clinical trials process is automated.

With the help of data science, it is increasingly becoming easy to design smarter trials and reinforce better discoveries, ultimately developing better drugs and best treatment techniques [32].

H. Royal Dutch Shell

The Royal Dutch Shell is a Dutch-British multinational oil and gas company that was founded in 1907. Its headquarters is located in The Hague, Netherlands [33]. In 2013 the revenue of Royal Dutch Shell was equivalent to 84% of the Dutch national GDP [34].

Shell has its presence in 70 countries and offers nearly 44,000 services around the world. They manufacture an astonishing volume of 3.7 million barrels of oil every day. As of 31st December 2019, Shell holds reserves worth over 11.1 billion barrels of oil equivalent.

Royal Dutch Shell was formed from the merger of two companies, a Netherlands petroleum company named "Royal Dutch Petroleum" and a UK based transport and trading company named "Shell". Royal Dutch Shell has operations ranging from extracting petroleum from fossil fuels to transporting it around the world. In order to reduce costs pertaining to all such tasks, Shell has adopted the data-driven approach.

With a rapidly growing population and the continuous need for energy, the demand for petroleum has increased a lot. Petroleum is a non-renewable resource, and the sources that produce petroleum are dwindling. As time passes, companies need to drill deep in order to find fossil fuels. This will lead to the inevitable rise in fuel costs as the companies invest more money to drill deep.

The exploration of these hydrocarbons needs a significant investment of manpower, equipment, and energy. Given the high expense of drilling a conventional deep-water oil well, which may cost \$100 million or more, it is essential that drilling takes place in the optimum locations that provide the hydrocarbons and best profit.

Generally, Sensors are used in order to find new fossil resources. Sensors are inserted into the earth surface to record the low-frequency seismic waves generated by tectonic activity under the earth. Sensors will record these seismic waves differently based on the material they are travelling through, i.e., waves that travel through solid rock will record differently, compared with the waves that travel through liquids or gaseous material.

Based on the recorded data from sensors, the likely location of hydrocarbon deposits is identified. Records have proven that this is a hit and miss process, i.e., when they drill, sometimes they find the hydrocarbon and sometimes they do not. In situations where fossil fuel is not found, the cost for the test drills exceeded the income generated from hydrocarbons found in other locations. Shell started to use a data-driven approach in order to overcome losses caused by test drills. A data-driven approach can solve this problem by comparing the record taken from the current drill with the records of the past drill.

The more the current drill resembles the profiles of previous locations from the past data where substantial resources have been discovered, the more likely it is that a full-scale drilling programme will be profitable. The model does not directly match the current drill data with the past data; it considers different parameters, does much computation and tries to predict the likeliness of the hydrocarbons to be found. Along with predicting the likely location to drill, Shell also uses data science to keep track of their equipment's operation and condition.

The equipment the shell uses for drilling and while extracting fuels undergo abnormal conditions, so proper maintenance of the equipment is a must for a long time. If the shell spends money for maintenance even though it is not required, the expenses will increase, so Shell needs to know the correct time for maintenance to save both their money and the equipment. Sensors are planted on the equipment and constantly captures the data about the functioning of the equipment. The recorded data from sensors and the past data model predicts the performance and the likelihood of failure. This enables more efficient routine maintenance, decreasing overheads even further. Shell drills in hundreds of places, and it has thousands of equipment all over the world. Data science helps the shell by predicting the possibilities of hydrocarbons and when equipment needs service or maintenance. If the shell tries to manage these properly, then it can save millions of dollars every year [35,36].

I. Tesla

Tesla is an American electric car manufacturing company based in California, United States, founded in 2003 by Martin Eberhard and Marc Tarpenning. Elon Musk started his CEO journey in Tesla from 2008 as he made the majority of the funding from the early stage. Currently Tesla production lineup includes electric cars, battery energy storage, solar panels and solar roof tile. Tesla is well known for their electric car, famous for its auto pilot abilities and long range. Though the current auto pilot is sophisticated enough, Tesla's main aim is to make fully automated cars.

To make the auto pilot fully functional, Tesla makes use of 2 AI enabled chips. These circuits perform separate assessments on what the best decision is, given the current situation of the car. to guide the car accordingly, which results in safer and confident auto pilot experience. The main disadvantage of such architecture is that using two powerful chips drains battery rapidly which indirectly decreases the range of the car. To tackle this problem, Tesla has started to optimise their AI chips. Initially, Tesla collaborated with NVIDIA to manufacture and optimise AI chips; later on they collaborated with Samsung [37,38].

After 14 months of research and improvement in the working of the chips, they are optimised to run at 2GHZ and perform 36 Trillion operations per second, which is 21 times more powerful than the previous one [39].

A Tesla car is equipped with 8 cameras with a maximum visible range of 250 meters and 12 ultrasonic sensors [40]. By using data from these sensors and cameras, Deep neural networks analyze raw camera footage to provide depth prediction, semantic segmentation and object detection. Another feature called the "Bird eye view" has View Parsing Network which takes video from all the 8 cameras to create a 3D model of the car surroundings for better understanding this network is trained from most complicated and different situations. To achieve this type of network, Tesla has implemented 48 neural networks which work spontaneously. In order to train this neural network, it is estimated that 70,000 hours of GPU training is required which is almost 8 years [41].

Earlier, they were implementing the "Round Robin" methodology. So they adopted the "pool of workers" strategy, resulting in 1000 tensors at a time. This is not possible on a standard architecture so Tesla adopted HydraNets architecture where it is a shared backbone [42]. The neural networks are trained using Pytorch wherein each image of 1280 x 960 dimension is provided to the neural network, Dilated convolution is used for deep semantic segmentation where the backbone is modified ResNet50. From all this hard work, Tesla cars in auto pilot can change lanes, turn in specific directions, avoid obstacles, predict whether a forward car will stop and even avoid collisions from blind spots. By all these Tesla has become the most selling electric car in the world [43].

J. Amazon

Amazon is an E-commerce company founded by Jeff Bezos in July 5th 1994 in Washington USA. Amazon uses data science for its services like product search in amazon website, Amazon Alexa, Amazon prime, Supply chain forecasting, and many more.

Recently, Amazon developed a unique, out-of-box solution for physical supermarkets. It is popularly known as Amazon Go. It is the most advanced convenience store in the world. The 15000 square foot store is completely automated using Computer vision, Deep learning, sensor fusion and is equipped with numerous cameras to cover every angle similar to self driving cars. When a buyer enters the store, by using CV the algorithm can identify the person and differentiate them using face recognition, height, user biometrics and past purchase history. With

sensor fusion the algorithm can understand whether the person picked the item or kept it back on the shelf, and add that item to the "Smart cart" which calculates and displays the total money to be paid by the customer. Then, the system sends the invoice to the customer through Email and checkout on the go without the hassle of waiting to pay. This allows customers to have a "Grab and Go" experience. This algorithm can even predict what item a user is going to pick, by analysing the user's position in the mall, direction, and past purchase history [44,45].

k. Walmart

Walmart is one of the biggest retailers in the world; it is an American company and has nearly 20,000 retail marts in 28 different countries. It was founded by Sam Walton on 2 July 1962, Rogers, Arkansas, United States. It is no surprise that a company of this size has massive transactional data. For Every hour, Walmart stores around 2.5 Petabytes of their customers' transactional data. One petabyte is equal to $1e+6$ Gigabyte or 1000 Terabyte, approximately equal to 20 million filing cabinets worth of text. Walmart outsmarted many of the good's shortage problems based on the transactional data, mainly during hurricanes. Before Hurricane Frances was on its way to hit Florida, Walmart started analyzing their past customer transaction data recorded during the hurricane named "Hurricane Charley," which hit Florida a few months before Hurricane Frances. By doing so, they were able to satisfy the customers' needs during Hurricane Frances. After analyzing the transaction data recorded during "Hurricane Charley," Walmart found out that Strawberry and Pop-Tarts sales increased seven times during the hurricane. By mining such data, Walmart stocked up all its local marts with the products that are very much needed for people during the disaster. Also, they found out that beer was the top-selling item during the pre-hurricane time. Extra supplies of these were dispatched to stores before Hurricane Frances hit, and they were sold in a jiffy. After this, Walmart decided to scale up its analytics and big data department to record huge profits by mining all the stored customer transaction data to find new trends and patterns [46].

Once the sales of one particular product of Walmart started declining, Walmart used transaction data to find out the reason; after analysis, they found out that the sales had declined due to the pricing error for the product. Once they rectified the error, then the sales recovered within days. By using data science techniques, Walmart can take down the time for spotting a problem and solving it from 2-3 weeks to around 20 minutes. If Walmart didn't estimate a good price for products, then they may lose their customers. They need data science in order to compete with the other competitors; Walmart constantly updates their prices based on the demand and supply. One of Walmart's senior Statistical Analysts, Naveen Peddamail, said, "If you can't get insights until you've analyzed your sales for a week or a month, then you've lost sales within that time."

In 2019 Walmart tested its new technology that can be capable of predicting when shelves get empty and when

to stock up the shelves on one of the Walmart stores in new york city. The store contains over 30,000 items, and the area of the store is 50,000 square feet. Walmart called this store IRL or Intelligent Retail Lab. Through an amazing network of sensors, cameras, and computers, IRL is able to gather information about what's going on inside the store and keep track of every product on the shelves. The main work of the data science algorithm is to determine whether an item is going out of stock by keeping track of all the items on the shelf. When a perishable item is not bought for a long time, the algorithm will then notify Walmart to take measures to maintain its freshness. Another advantage of stores implementing IRL is that the employees need not track the inventory present in the store manually; thereby, they can spend more time enhancing customer relationships. Walmart believes that these types of stores make customers happier because of the availability of goods at every moment [47].

IV. CONCLUSION

Data is the new oil, and businesses are trying to make the most of it. Accumulation of large volumes of data led to the field of data science that deals with what data is to be collected, how the data is to be collected, how to train the model, how to find interesting patterns, and many more. Large Multinational Corporations are spending their fortunes to analyze their consumers, predict market demand and trends beforehand, and take calculated risks. The quality of insights being delivered by the analysts and the data scientists makes an organization stand out in the tough competition. Various case studies, as discussed above, state how advantageous data science was to scale up their business. With businesses analyzing every possible pattern in the data being recorded and striving hard to satisfy and retain their consumer base, data science is helping to improve the quality of a service or a product which indirectly develops the quality of the entire society.

REFERENCES

- [1] Kesharwani S. Disruptive Technology Relocate a well-established Product and Engender Benchmark in Industry. *Glob J Enterp Inf Syst*. 2021;13(1):1–6.
- [2] Nast C. This is how Netflix's top-secret recommendation system works. *Wired UK* [Internet]. [cited 2021 Aug 30]; Available from: <https://www.wired.co.uk/article/how-do-netflixs-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like>
- [3] How Netflix's Recommendations System Works [Internet]. Help Center. [cited 2021 Aug 30]. Available from: <https://help.netflix.com/en/node/100639>
- [4] Blog NT. Data Science and the Art of Producing Entertainment at Netflix [Internet]. Medium. 2018 [cited 2021 Aug 30]. Available from: <https://netflixtechblog.com/studio-production-data-science-646ee2cc21a1>
- [5] Where is Netflix available? [Internet]. Help Center. [cited 2021 Aug 30]. Available from: <https://help.netflix.com/en/node/14164>
- [6] About Netflix - The Power of a Picture [Internet]. About Netflix. [cited 2021 Aug 30]. Available from: <https://about.netflix.com/>
<https://about.netflix.com/en/news/the-power-of-a-picture>
- [7] Spotify — Company Info [Internet]. Spoti fy. [cited 2021 Aug 30]. Available from: <https://newsroom.spotify.com/company-info/>
- [8] For Your Ears Only: Personalizing Spotify Home with

- Machine Learning [Internet]. Spotify Engineering. 2020 [cited 2021 Aug 30]. Available from: <https://engineering.atspotify.com/2020/01/16/for-your-ears-only-personalizing-spotify-home-with-machine-learning/>
- [9] Pérez-Marcos J, Batista VL. Recommender system based on collaborative filtering for spotify's users. In Springer; 2017. p. 214–20.
- [10] How AI helps Spotify win in the music streaming world [Internet]. Outside Insight. 2018 [cited 2021 Aug 30]. Available from: <https://outsideinsight.com/insights/how-ai-helps-spotify-win-in-the-music-streaming-world/>
- [11] Recommending music on Spotify with deep learning [Internet]. Sander Dieleman. [cited 2021 Aug 30]. Available from: <https://benanne.github.io/2014/08/05/spotify-cnns.html>
- [12] Our Spotify Cheat Sheet: 4 Ways to Find Your Next Favorite Song [Internet]. Spotify. 2018 [cited 2021 Aug 30]. Available from: <https://newsroom.spotify.com/2018-11-02/our-spotify-cheat-sheet-4-ways-to-find-your-next-favorite-song/>
- [13] About Face ID advanced technology [Internet]. Apple Support. [cited 2021 Aug 30]. Available from: <https://support.apple.com/en-in/HT208108>
- [14] An On-device Deep Neural Network for Face Detection [Internet]. Apple Machine Learning Research. [cited 2021 Aug 30]. Available from: <https://machinelearning.apple.com/research/face-detection>
- [15] Pocket-lint. What is Apple Face ID and how does it work? [Internet]. 2021 [cited 2021 Aug 30]. Available from: <https://www.pocket-lint.com/phones/news/apple/142207-what-is-apple-face-id-and-how-does-it-work>
- [16] Apple Pencil and Scribble - User Interaction - iOS - Human Interface Guidelines - Apple Developer [Internet]. [cited 2021 Aug 30]. Available from: <https://developer.apple.com/design/human-interface-guidelines/ios/user-interaction/apple-pencil-and-scribble/>
- [17] Accessibility - Hearing [Internet]. Apple (India). [cited 2021 Aug 30]. Available from: <https://www.apple.com/in/accessibility/hearing/>
- [18] Projects [Internet]. opensource.google. [cited 2021 Aug 30]. Available from: <https://opensource.google/projects/>
- [19] Portrait mode on the Pixel 2 and Pixel 2 XL smartphones [Internet]. Google AI Blog. [cited 2021 Aug 30]. Available from: <http://ai.googleblog.com/2017/10/portrait-mode-on-pixel-2-and-pixel-2-xl.html>
- [20] Google Assistant guide: Make the most of your virtual assistant [Internet]. Android Authority. 2021 [cited 2021 Aug 30]. Available from: <https://www.androidauthority.com/google-assistant-838138/>
- [21] An Augmented Reality Microscope for Cancer Detection [Internet]. Google AI Blog. [cited 2021 Aug 30]. Available from: <http://ai.googleblog.com/2018/04/an-augmented-reality-microscope.html>
- [22] Facebook: revenue and net income [Internet]. Statista. [cited 2021 Aug 30]. Available from: <https://www.statista.com/statistics/277229/facebooks-annual-revenue-and-net-income/>
- [23] Taigman Y, Yang M, Ranzato M, Wolf L. Deepface: Closing the gap to human-level performance in face verification. In 2014. p. 1701–8.
- [24] Introducing DeepText: Facebook's text understanding engine [Internet]. Facebook Engineering. 2016 [cited 2021 Aug 30]. Available from: <https://engineering.fb.com/2016/06/01/core-data/introducing-deep-text-facebook-s-text-understanding-engine/>
- [25] Nøland JK. Prospects and challenges of the hyperloop transportation system: A systematic technology review. IEEE Access. 2021;
- [26] Customer Story: Virgin Hyperloop [Internet]. Databricks. [cited 2021 Aug 30]. Available from: <https://databricks.com/customers/virgin-hyperloop>
- [27] AstraZeneca and Oxford University announce landmark agreement for COVID-19 vaccine [Internet]. [cited 2021 Aug 30]. Available from: <https://www.astrazeneca.com/media-centre/press-releases/2020/astrazeneca-and-oxford-university-announce-landmark-agreement-for-covid-19-vaccine.html>
- [28] Medicines - Our focus areas - AstraZeneca [Internet]. [cited 2021 Aug 30]. Available from: <https://www.astrazeneca.com/our-therapy-areas/medicines.html>
- [29] Data Science & Artificial Intelligence: Unlocking new science insights [Internet]. [cited 2021 Aug 30]. Available from: <https://www.astrazeneca.com/r-d/data-science-and-ai.html#pushingthebound>
- [30] Data Science & Artificial Intelligence: Unlocking new science insights [Internet]. [cited 2021 Aug 30]. Available from: <https://www.astrazeneca.com/r-d/data-science-and-ai.html>
- [31] AstraZeneca, BenevolentAI home in on computer-generated target for chronic kidney disease drugs [Internet]. FierceBiotech. [cited 2021 Aug 30]. Available from: <https://www.fiercebiotech.com/medtech/astrazeneca-benevolentai-home-in-on-computer-generated-target-for-chronic-kidney-disease-drugs>
- [32] A New Era in Target Discovery: Collaborating with AstraZeneca on CKD and IPF | BenevolentAI [Internet]. [cited 2021 Aug 30]. Available from: <https://www.benevolent.com/news/a-new-era-in-target-discovery-collaborating-with-astrazeneca-on-ckd-and-ipf>
- [33] Garavini G. The Rise and Fall of OPEC in the Twentieth Century. Oxford University Press; 2019.
- [34] Shell, Glencore, and Other Multinationals Dominate Their Home Economies. Bloomberg.com [Internet]. 2013 Apr 5 [cited 2021 Aug 30]. Available from: <https://www.bloomberg.com/news/articles/2013-04-04/shell-glencore-and-other-multinationals-dominate-their-home-economies>
- [35] Marr B. Big Data In Big Oil: How Shell Uses Analytics To Drive Business Success [Internet]. Forbes [cited 2021 Aug 30]. Available from: <https://www.forbes.com/sites/bernardmarr/2015/05/26/big-data-in-big-oil-how-shell-uses-analytics-to-drive-business-success/>
- [36] Artificial Intelligence [Internet]. [cited 2021 Aug 30]. Available from: <https://www.shell.com/energy-and-innovation/digitalisation/digital-technologies/shell-ai.html>
- [37] Eliot L. Tesla's AI Chips Are Rolling Out, But They Aren't A Self-Driving Panacea [Internet]. Forbes. [cited 2021 Aug 30]. Available from: <https://www.forbes.com/sites/lanceeliot/2019/09/29/teslas-ai-chips-are-rolling-out-but-they-arent-a-self-driving-panacea/>
- [38] Csongor R. Tesla Raises the Bar for Self-Driving Carmakers [Internet]. The Official NVIDIA Blog. 2019 [cited 2021 Aug 30]. Available from: <https://blogs.nvidia.com/blog/2019/04/23/tesla-self-driving/>
- [39] Shankland S. Take a close-up look at Tesla's self-driving car computer and its two AI brains [Internet]. CNET. [cited 2021 Aug 30]. Available from: <https://www.cnet.com/tech/computing/meet-tesla-self-driving-car-computer-and-its-two-ai-brains/>
- [40] Autopilot [Internet]. [cited 2021 Aug 30]. Available from: <https://www.tesla.com/autopilot>
- [41] Artificial Intelligence & Autopilot [Internet]. Tesla. [cited 2021 Aug 30]. Available from: <https://www.tesla.com/AI>
- [42] McCaw B. The Batch: Tesla Parts the Curtain, Detecting Dangerous Bugs, Mapping Disaster Zones, Detecting Humans from Wi-Fi, Toward Trustworthy AI [Internet]. [cited 2021 Aug 30]. Available from: <https://blog.deeplearning.ai/blog/the-batch-tesla-parts-the-curtain-detecting-dangerous-bugs-mapping-disaster-zones-detecting-humans-from-wi-fi-toward-trustworthy-ai>
- [43] Learn Self-Driving Cars, Computer Vision, and cutting-edge Artificial Intelligence [Internet]. Think Autonomous. [cited 2021 Aug 30]. Available from: <https://www.thinkautonomous.ai/blog/>

- [44] Amazon.com: Amazon Go Store: Amazon Go [Internet]. [cited 2021 Aug 31]. Available from: <https://www.amazon.com/b?ie=UTF8&node=20931384011>
- [45] Wankhede K, Wukkadada B, Nadar V. Just walk-out technology and its challenges: A case of Amazon Go. In 2018 International Conference on Inventive Research in Computing Applications (ICIRCA) 2018 Jul 11 (pp. 254-257) IEEE.
- [46] Provost F, Fawcett T. Data science and its relationship to big data and data-driven decision making. Big data. 2013 Mar 1;1(1):51-9. [47]
- [47] Walmart's New Intelligent Retail Lab Shows a Glimpse into the Future of Retail, IRL [Internet]. Corporate - US. [cited 2021 Aug 31]. Available from <https://corporate.walmart.com/newsroom/2019/04/25/walmarts-new-intelligent-retail-lab-shows-a-glimpse-into-the-future-of-retail-irl>