

## Possibilities of Auto ML in Microsoft Replacing the Jobs of Data Scientists

Samyukta B\*, Arjun Sreekumar, Hari Varshan S R, Navaneeth P, Vaishnavi

Department of Integrated Data Science, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu, India

### ABSTRACT

Most of the people have confusion in the scope of Data Science because of the advancements in Artificial Intelligence (AI). Particularly, Auto ML (also known as Automated Machine Learning) has challenged the jobs of data scientists. This article aims on clearing the misconceptions about the subject.

**Keywords:** Data science; Artificial intelligence; Machine learning; Microsoft

### INTRODUCTION

The purpose of Auto ML is to make Machine Learning easier to work with and understand, and is accessible by automatically generating a data analysis. Therefore it greatly affects the work of data analysts and other data science jobs due to lesser manual work. However, Artificial Intelligence doesn't completely affect data science jobs. Data science is a vast subject covering numerous curricula involving data. And Machine Learning falls within the scope of Data Science and aids it in processing data. Data Analytics and Machine Learning are two sides of the same coin when it comes to Data Science. Data generation has been increasing rapidly over the years. So in this technical world where there is huge generation of data, the demand for data scientists and Machine Learning Engineers has been increasing and pursuing these jobs has now become trendy. Mastering the necessary skills for Data Science and Machine Learning will help secure a strong career in this era.

### DATA SCIENCE VS. AUTO ML

This 20th century has given a lot of improvements in the field of technology especially in artificial intelligence which shows a lot of promising developments in the field of technology. But this 20th century has also given birth to another field of technology called data science. In fact, only the job of data scientist is considered to be the sexiest in the 20th century. As mentioned before, improvements in Artificial Intelligence has been soaring [1]. Microsoft has worked on a project called Auto ML for years and come up with an Artificial intelligence to replace one of the works done by data scientists.

So, what exactly is the job of a data scientist and why is it considered to be the sexiest job?

- Will AI replace data scientists?
- Will all data scientists go unemployed?
- Is all the hype about data scientists, a lie?

To answer all these questions we will give a brief understanding of Data Science, AI, Auto ML and its relation with data science.

### Data science

The job done by a data scientist is wide and vast but for your understanding let us consider one of the applications of a data scientist. The recommendation algorithms you see in YouTube, Tik-Tok, Instagram and all such apps are the work done by data scientists. In a way, these recommendation algorithms steal your time and also offer right video to the right person. For e.g., let us say a person watches only videos about gaming, then his homepage will be loaded with videos related to games and stuff. In fact, these recommendations are the ones which increase the watch time of the users and keep the users in the app for a long time [2]. For a person who watches videos related to cooking, if his homepage is loaded with videos related to games he will absolutely close the app, but if his recommendations are also related to cooking he will click onto the next video and another video and so on. Personalized ads are also one of the applications of the data scientists.

**Corresponding authors:** SamyuktaB, Department of Integrated Data Science, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu, India, Email: samyukta299@gmail.com # All the authors have contributed equally to this paper.

**Received:** February 02, 2021; **Accepted:** February 16, 2021; **Published:** February 23, 2021

**Citation:** Samyukta B, Sreekumar A, Varshan SRH, Navaneeth P, Vaishnavi (2021) Possibilities of Auto ML in Microsoft Replacing the Jobs of Data Scientists. J Inform Tech Softw Eng. S3:001

**Copyright:** © 2021 Samyukta B, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

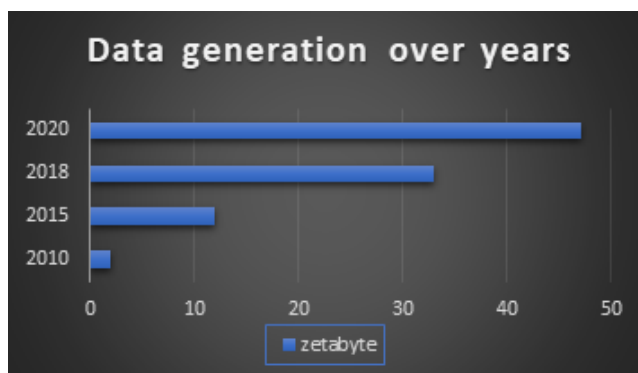
### Have you ever noticed that when you search for a product in goggle, the next day you find ads related to that product in all other apps?

Well, it is also one of the applications of data science and it is not an exploitation of your privacy. You can turn of such personalized ads and tracking of your browsing when you want to. Data science is literally used in all existing fields. It is used in space observations, medical sciences, airplane route planning, image recognition, etc. So the job of a data scientist is to process data in a nutshell. Technically, the main aim of a data scientist is to draw out business focused insight from the data available. It consists of data developing methods like recording, storing and analyzing of data, and extracting all the useful information required to use it to identify business opportunities.

Now for the second question,

### Why the job of a data scientist is considered to be the hottest in the 20th century?

One of the reasons is because, data scientists are expensive to hire and their job is also difficult because, the amount of data generated is hiking over the years.



**Figure 1:** Shown below is a graph of data generated over the years.

1 zettabyte=1 billion terabytes

In the year 2020, we have generated data for around 47 zettabytes. So it is 47 billion terabytes, and one terabyte is equal to sixty four 16 GB pen drives which we use every day. So imagine how much amount of data is generated. And this data generation is increasing rapidly over the years. The prediction for 2035's data generation is 2142 zettabytes. So the increasing generation in data and the demand for data scientists to process these data to create valuable information is more. That's why the job of data scientist is considered to be the sexiest in the 20th century (Figure 1).

## MACHINE LEARNING

The key process of Machine Learning is to make the learning algorithm to find patterns from the input data called training data. A set of new rules is generated by the learning data algorithm based on references from the data. The same learning algorithm can be used to generate different models by using different training data. For example, teaching the computer to translate languages and predicting the stock market can be done

by the same learning algorithm. The most common and important applications of Machine Learning are image recognition, speech recognition, traffic prediction, product recommendations, online fraud detection, virtual personal assistant, email spam and malware filtering, etc.

## ARTIFICIAL INTELLIGENCE

Artificial Intelligence, generally referred to as AI is about how machines are made to imitate the work of humans. It is about how I make them think and provide appropriate solutions to certain problems just like humans. Robots are just machines that can make the work of humans easier. Technically, AI is defined as the simulation of human intelligence in machines which are programmed to imitate the work of humans [3].

### Where is it used?

It is used almost everywhere from playing computer games to driving cars. In computer games, mainly in multiplayer games, when we don't have someone to play with, we see that there is a computer-player mode. This is an application of AI. In autonomous cars, AI is used to enable the cars to navigate through the traffic and handle complex situations. AI plays a huge role in self-driving cars. It can sync with traffic signals, break in case of emergencies, etc. Also, in the finance sector, like detecting unusual debit card usage, etc., AI helps in identifying fraudulent activities. AI is also used in the healthcare industry to detect the treatment and dosage of drugs for different patients. It is also used for surgical operations in the operating room. Today's world cannot function without the help of AI.

## ROLE OF MACHINE LEARNING IN DATA SCIENCE

Machine Learning plays a major role in Data Science. It helps in finding a pattern from the given data and also for making predictions. A pattern based on users' choice is predicted from the data available. Companies depend a lot on this pattern to provide the consumers with products which they require the most and give offers to increase their business sales [4]. Also, by making use of Machine Learning algorithms, future trends can be predicted. This is based on supervised learning. The machine is provided with a data set and is made to perform a specific task from the given data. The machine will learn the trends by itself by repeatedly performing its task.

## AUTO ML

Automated machine learning is the process of automating the process of applying machine learning to real-world problems. Auto ML covers the complete pipeline from raw dataset to deployable machine learning model. Auto ML is an artificial intelligence-based solution which was developed to challenge the application of Machine Learning. The goal of Auto ML is to make machine learning more accessible by automatically generating a data analysis pipeline that can include data pre-processing, feature selection, and feature engineering methods along with machine learning methods and parameter settings that are optimized for the given data. Auto ML's aim is to make machine learning more available by basically creating a pipeline for data analysis which can include data preprocessing, feature

selection, and feature engineering techniques, along with machine learning techniques and optimized parameter settings. For a machine learning specialist, each of these steps can be time consuming and can be crippling [5].

## HIGH DEMAND FOR DATA SCIENTISTS

Despite all the advancements in AI, the demand for data scientists is still more. By 2026, the demand for data scientists will drive by 27.6%. US called the job of a data scientist to be the number one for four consecutive years. People pursuing Data Science are highly paid. In fact, data science jobs are one of those highly paid jobs. Ever since Indian companies extended their analytic area, the need for trained data scientists has increased. The average salary of an entry level data scientist with less than a year experience is Rs.5,00,000 per year. An average data scientist earns around Rs.698,412. Data science can be made use of in almost every sector such as healthcare, business, e-commerce, banking, etc. People with knowledge in various backgrounds such as mathematics, statistics, analytics, analysis, data modeling, etc., get more opportunity in the industry [6].

Most popular data science careers

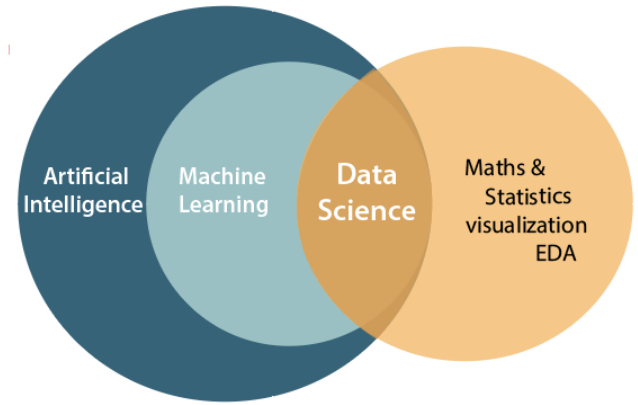
- Data Scientist
- Data Analyst
- Data Analytics Engineer
- Data Architect
- Machine Learning Engineer
- Applications Architect
- Business Intelligence Developer
- Quantitative Analyst
- Statistician

## RESOURCES AVAILABLE FOR LEARNING DATA SCIENCE

There are many resources for people to learn data science through both online and offline modes. Subscription based online learning platforms such as Udemy, LinkedIn learning, SKILLSHare, upGrad, etc., gives people the advantage to attend classes anywhere at their own pace. When compared to online classes, offline courses cost more. So opting for subscription based learning is better.

## WILL AUTO ML REPLACE DATA SCIENTISTS?

As told before, data science is all about converting raw data into information and getting insights and understandings about the data. While data science does this job, it depends on other traits such as machine learning, artificial intelligence, statistics, deep learning and also few others. In fact, data science itself is a collaboration of various categories of technologies (Figure 2).



**Figure 2:** Various categories of technologies.

80% of the data science jobs are all about data preparation. This data preparation is about making models with data which would predict a precise outcome of information. The model we refer to is a large table of organized data from which data insights can be taken. Machine Learning comes into place while creating the models for data science projects. Auto ML software such as DataRobot and H2O automate the repetitive work of creating models. We have to feed data for the software while it grinds through iteration of features, models and parameters. But still it is deficit when it comes to expert prediction and domain expertise, while it can replace the brute and repetitive work. And also the trick in using this Auto ML software is feeding in the right data which is entirely up to the human intelligence. Though Auto ML cannot replace human intelligence, it can replace a large amount of brute force done by data scientists. So the number of data scientists required for a specific project will absolutely drop, if we consider in a practical sense. And also the after work of a data scientist after data preparation also requires skilled data scientists [7].

## CONCLUSIONS

### Will Auto ML replace data scientists and machine learning engineers?

Yes it will, for data scientists who don't adapt to Auto ML software like Data Robot, and also for data scientists who have expertise in the repetitive work. On the other hand, Auto ML is a boon for those who get adapted to this software and a great future lies ahead of them. For aspiring data scientists and ML engineers, embracing changes and flexibility would land them in a good job.

### Will these technologies eradicate data science jobs in future?

To put it blindly, we can never say that until the time comes. The computer language, 'Java' is said to be replaced for a long time. But it is still used vastly. Unlike Java, data science is a still vast field and the probability of getting it replaced is less probable.

## REFERENCES

1. Wang D, Weisz JD, Muller M, Ram P, Geyer W, Dugan C, et al. Human-AI collaboration in data science: Exploring data scientists' perceptions of automated AI. *Proceedings of the ACM on Human-Computer Interaction*. 2019;3:1-24.
2. Jason Brownlee. *A tour of machine learning algorithms*. Machine Learning Algorithms. 2019.
3. Sharma R. Data scientist salary in India in 2021. *Data Science*. 2021.
4. Press G. *Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task*, Survey Says. 2016.
5. Kaur H, Nori H, Jenkins S, Caruana R, Wallach H, Wortman Vaughan J. Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. *Inproceedings of the 2020 chi conference on human factors in computing systems*. 2020;1-14.
6. Kim M, Zimmermann T, DeLine R, Begel A. The emerging role of data scientists on software development teams. *IEEE*. 2016;96-107.
7. Kaur H, Nori H, Jenkins S, Caruana R, Wallach H, Vaughan WJ. Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020;1-14.