As data science tools become more sophisticated and capable of performing tasks that are usually delegated to data scientists, the future of data scientists becomes uncertain. Currently, the responsibilities of the data scientist include collecting data, processing data, exploring data, identifying problems, creating solutions with data, deploying models, and being up to date with all the newest tools, languages, and more (Yildirim, 2020). Many of these responsibilities can be automated, which has led data scientists such as Saxena (2021) to argue that there will be a lack of data science jobs in the future due to automation.

Although I agree with Saxena (2021) that there will be a shortage of data science jobs in the future, I also would like to argue that the role of a data scientist in the future will not be what Yildirim (2020) says it is today. Przybyla (2021) makes a similar argument, stating that while many of the processes that data scientists do can be automated, that is not where the value of a data scientist lies. Automated machine learning (AutoML) cannot understand the problems of a business, cannot know what data sources to look for, and cannot understand what to do with the results it has created. The human part of the data scientist, the curiosity, the insights, and the ability to research and explore data is the true value of a data scientist. In addition, if the current responsibilities of data scientists become automated then data scientists will have much to look forward to in the future. According to 76% of data scientists, the least enjoyable part of the job is data preparation, which also consumes around 80% of their time (Press, 2016). In the future, these tasks will likely be automated and allow data scientists to use their knowledge and time to tackle the more human parts of the data science methodology.

The responsibilities of the data scientist are changing, allowing for data scientists to use their knowledge and expertise for more than just data preparation. If the 80% of the time data scientists use for data preparation (Press, 2016) becomes automated, data scientists will have more time to identify problems, create solutions, and perhaps even specialize into more advanced data science topics, such as genomic data science. However, if so much of a data scientist’s current workload is automated, there will not be as high of a demand for data scientists as there is now, leading to Saxena’s (2021) predicted shortage in data science jobs due to automation. The role of the data scientist is evolving, with their responsibilities becoming more about the value they as a person can provide.

Reference List:

Press, G. (2016) Cleaning Big Data: Most Time Consuming, Least Enjoyable Data Science Task, Survey Says. Available from: https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#54b742616f63 [Accessed 15 May 2023]

Przybyla, M. (2021) 5 Examples Where Data Scientists Can't Be Automated. Available from: https://towardsdatascience.com/5-examples-where-data-scientists-cant-be-automated-c3d82c518d37 [Accessed 10 May 2023]

Saxena, P. (2021) There Will Be a Shortage of Data Science Jobs in the Next 5 Years? Available from: https://towardsdatascience.com/there-will-be-a-shortage-of-data-science-jobs-in-the-next-5-years-9f783737ed23 [Accessed 7 May 2023].

Yildirim, S. (2020) The Dark Side of the Sexiest Job of the 21st Century. Available from: https://towardsdatascience.com/the-dark-side-of-the-sexiest-job-of-the-21st-century-fd9c46bf4cae [Accessed 10 May 2023]