Chance Pickett (011004180)

DASC 1223 – Intro to Data Science (Dr. Schubert)

The Role of Data Science in Today's World – My View

So far, I've learned programming in R and Python, focusing on their applications in data analysis, particularly in creating graphs. Through guest speakers, I've gained insights into the daily activities of a data scientist and their significance in the modern world. The importance of data scientists stems from the need of businesses to manage and analyze large volumes of data. Today, the role of data science is to identify patterns within extensive datasets and provide answers to complex questions across various industries.

In my programming class, my professor showed us how he used data science to analyze the spread of covid in China. He used R language to find where the different strains of covid originated from and how fast they were spreading. Using data science allowed him to answer questions such as: Where did covid originate from? What caused this epidemic? What areas are more vulnerable than others? This is a great example of the application of data science in bioinformatics. Bioinformatics involves biological data that can be analyzed to make predictions about diseases, medicinal treatments, and other biological processes. In Success in Data Studies, I analyzed hospital data on Sepsis patients. At the time, I did not know how to use Python or R, so I used Excel to find patterns in the patient death rates in multiple hospitals. Using simple data science techniques, I was able to find meaningful insights into sepsis care.

Another role of data science in today's world is data integrity and ethical considerations. Earlier this year we did a project on Theranos and Elizabeth Holmes. Theranos was secretive about its internal data and operations, which prevented scientists from validating their technology. In data science, it's important to validate and reproduce results with accuracy. Because Theranos didn't allow independent verification of their test results, Elizabeth Holmes was found guilty on fraud charges. Not only did Theranos not share their data, but they also manipulated their data that they shared to investors. For example, they used misleading data in demonstrations to investors by selecting specific data to show favorable results, which is highly unethical. Also, Theranos misled patients and doctors about the accuracy of their blood tests, which is a violation of their rights. Ethical data science practices require being honest about the capabilities and limitations of your technology.

Last year, we did a project on airline data to find insights into flight arrivals and departures. This was my first hands on data science EDA, so the basics of cleaning and organizing the data were some of the hardest steps. If dealing with a very large excel sheet is difficult for a team of 5, I can't imagine what a company like Apple or Amazon must deal with. The amount of data I give amazon is amazing. Just by purchasing an item on Amazon, I am trusting them with my credit card, my name and address, and my phone number. They also probably store all the metadata like tracking my package, the price, the details of the actual item, the seller's data, and everything in between. For a company ensure that this data stays secured, organized, accurate, and clean, is honestly a miracle. With so much data, they are going to need large physical data warehouses with people who can create sophisticated database management systems. So, a large role of data science today is being able to ensure that the data tat comes from companies like these can be understood and well managed. Garbage in, garbage out.

With ChatGPT being so accessible now, AI and machine learning are very important today. While applying to internships for a data science role, I noticed that many companies don't sell a product, but instead sell their AI expertise and tools to help other companies with business problems. From what I understand, AI and machine learning involves integrating knowledge from various fields like data science, linguistics, and statistics. Returning to my earlier comment on having enough storage for large amounts of data, I listened to a podcast a few months ago that spoke about how AI models like ChatGPT are getting more advanced quicker because they price to buy the store the data that teaches the models are getting cheaper. This is interesting because data science is so generally used that it can be applied in many ways to this event. It could be used to ask why the price of storage is getting cheaper, where is the best place to store the data, but also more specific questions like how much will they model improve if we buy x amount more storage, or what data exactly does the model need to learn this or that.

To conclude, I've been able to form my own opinion on the role of data science in today's world by having hands on experience with programming languages that are commonly used by data scientists, create exploratory data analyses on modern datasets, and learning about real world situations like Theranos, and my professor's project on COVID. The role of data science is for companies to answer questions about their customers, their finances, and other operations. Companies like Amazon must deal with tons of data that needs to be organized, cleaned, and secured. They do this by creating complex databases in large data warehouses. The role of data science today is to answer questions about mankind's health and lives to make the world a better place. Data scientists use real world data such as

hospital visits to track disease such as COVID. Today, we have to deal with war and disease just like our ancestors, proving that data science will forever play a vital role in our lives.