Lifu Deng

Blanche Cohen

CS 1030-003

Apr 28, 2023

Final Project

Data Tools

For Data Tools the I learn that storing data sets can be text files, comma-separated files and the common tools are Spreadsheets and database. To interact with the database, as programmers and data analyst, they often use a query language. The most popular query language is SQL (Structed Query Language). In the Computing basic statistics, I have learned to use COUNTA, AVERAGE, MIN/MAX, SUM, COUNTAIF and similar IF functions from both Spreadsheets and SQL. For Summarizing you use pivot table and GROUP BY in SQL. In the Finding patterns in data sets I learned the spotting trend downward and upward or no trends by looking at the data sets. The data sets can be visualized, and Google trends is a site that visualizes the popularity of Google search terms over time.

Big Data

In this lesson I have learned where big data comes from and the ways we can use it. The big data is data coming from one very large source. Most of the time, big data is a collection of data from lots of little sources. The ways to use data are scientific research, digital libraries, medical records, and user-facing applications. For each different ways they gave examples like Data.gov, Internet Archive, MRI, CT scan, and Facebook.com etc. The challenge of big data is storage, processing and responsible use for unbelievable data that been generated every single day.

Bias In Machine Learning

From Machine Learning Algorithms I have learned three general approaches: reinforcement learning, unsupervised machine learning, supervised machine learning. An increasingly popular approach to supervised machine learning is the neural network. A neural network operates similarly to how we think brains work, with input flowing through many layers of “neurons” and eventually leading to an output. They also explained how to train a network, so now I have an idea about how training a network can help the neural network get better accuracy on classify an object. They have another name called “artificial intelligence” which is A.I., but if the biased training data feeds to A.I. then the algorithm is biased and that is more often than it’s not biased. I learned about different types of bias in machine learning such as bias in predictive algorithms, facial recognition, and language translation. Overall, this lesson is important to understand the possible bias in machine learning. Specifically, the A.I. age is already here, we should be aware of the biased information that A.I. generate could potentially misinform its users.

Unit Test

The experience with the Unit Test was good, because for each lesson the practice was well designed. Although the questions could be hard to understand sometimes, and I have got numerous wrong answers during these practices and the questions for the same practice could not be the same. But I think the essential concepts for these practices’ questions are the same for each lesson, and they will explain when you are wrong for your answer I think this design could help me better understand the concepts. So I would say the Unit Test served its purpose.