MCIS6273 Data Mining (Prof. Maull) / Fall 2021 / HW2a

This assignment is worth up to 5 POINTS to your grade total if you complete it on time.

Points Possible	Due Date	Time Commitment (estimated)
5	Wednesday, October 15 @ Midnight	*up to* 3 hours

- **GRADING:** Grading will be aligned with the completeness of the objectives.
- **INDEPENDENT WORK:** Copying, cheating, plagiarism and academic dishonesty *are not tolerated* by University or course policy. Please see the syllabus for the full departmental and University statement on the academic code of honor.

OBJECTIVES

- Learn about the new data economy through Ocean Protocol and how Blockchain is fueling innovation in the data space

WHAT TO TURN IN

You are being encouraged to turn the assignment in using the provided Jupyter Notebook. To do so, make a directory in your Lab environment called `homework/hw2`. Put all of your files in that directory.

Then zip that directory, rename it with your name as the first part of the filename (e.g. `maull_hw2_files.tar.gz`), then download it to your local machine, then upload the `.tar.gz` to Blackboard.

If you do not know how to do this, please ask, or visit one of the many tutorials out there on the basics of using in Linux.

If you choose not to use the provided notebook, you will still need to turn in a `.ipynb` Jupyter Notebook and corresponding files according to the instructions in this homework.

ASSIGNMENT TASKS

(20%) Learn about the new data economy through Ocean Protocol and how Blockchain is fueling innovation in the data space

You have no doubt heard the phrase "Data is the new oil", which goes back to 2006, when British data scientist and mathematician Clive Humby first coined the phrase.

For an interesing chronology of the phrase over the past 15 years or so, please read the follow Medium.com post:

- "Data is the New Oil" — A Ludicrous Proposition by Michael Haupt

You can decide for yourself, after some contemplation, whether the "privatization" of data is right or wrong, and what one might do about it if they choose to.

Nonetheless, the parallels to oil, and the oil boom at the turn of the 20th century – which profoundly changed our relationship with energy and quite literally fueled innovation on a scale that humanity has never seen – are interesting. Data, however, is not a finite resource, and the "ownership" of data is quite different than that of oil (however controversial that may be). What should be obvious is that producing data and "enriching it" (adding "value") are two different things, and "value" is a term that requires a context.

Nonetheless, we are at an interesting point in the big data, data mining and data science conversation about what to do with and how to treat all the data now being produce, particularly as the kind of data that is becoming interesting to companies is increasingly personal (i.e. personal health data) and quite possibly being put to use in contexts we have no transparency or control over.

Blockchain technologies are rapidly become the *de facto* platforms for data-related transactions, and interesting work is being done to use these distributed digital ledgers as data exchange platforms. Even more work is being done to bring "data ownership" to a new level on digital marketplaces and the "Internet of Value" or IoV (see also this 2017 writeup on the IoV by Ripple) building the "new data economy".

You will be listening to a podcast on this subject involving the Ocean Protocol and its founder Bruce Pon. You will listen the almost 43 minute podcast from The Crypto Conversation Podcast on Brave New Coin.

- "Data is the new oil - Ocean Protocol is building the data economy" - October 28, 2020 Podcast with Bruce Pon;
- download the mp3 file directly.

§ Listen to the entire podcast and answer the questions below:

1. What was the original idea that Pon tried in 2013?
2. Who (or what) did Pon say was the "consumer" of Blockchain and big data? How does this relate to "data mining"?
3. Who does Pon say are the gatekeepers of the current "data economy"? How does Ocean Protocol address this?
4. How does Pon suggest data might be "securitized" over Ocean Protocol? Use the example he gives directly.
5. What are the four landscape users (or targets) of Ocean Protocol?
6. What example does he give of a liquid dataset that might have tremendous value in a variety of current and future contexts relating to our world?
7. What does Pon predict the crypto marketplace will look like in 10 years?
8. After consuming this podcast, please list your reaction to it in a short paragraph of at a minimum of three **complete** sentences. Address what you already know about crypto and the data economy as well as the things that you learned that were new knowledge. You may also address whether you agree or disagree with Pon's positions, especially how it relates to Ocean Protocol and cryptocurrencies, in general, becoming a great equalizer. You can also include any personal experiences you have with the technology.