Colleen Lemak
Professor Sprint
CPSC 222
14 October 2021

project_part3

1. What table could be populated from this data source?
    a. In Table #1, I could produce a table of screen time in hours, number of times I pick up my phone, and the amount of notifications I get based on the day of the week.
2. What is at least one other table of data you could combine with this one to form a larger and more informative dataset?
    a. In Table #2, I could combine the previous table with the weather statistics based on the day of the week according to the date, so I can see correlation between phone statistics and temperature and weather in general on that day.
3. For each table in your dataset:
        a. What is an instance? What is the universe of instances?
        i. An instance is a labeled row of data.
        ii. A universe of instances is the scope of what data you choose to observe.
        iii. For Table #1, the instances are screen time in hours, number of pick-ups as an integer, and the amount of notifications as an integer, and the universe of instances is my screen time data within the last few months.
        iv. For Table #2, the instances would be temperature in degrees, and true or false values in the instance labeled “sunny weather”. The universe of instances would be the weather application, including the last few months, matching up with Table #1.
        b. What are the attributes? For each attribute:
            1. Attributes include the date and day of the week, and number of homework hours done that day

2. Is it categorical/discrete or continuous?
    a. Date is discrete, and the number of hours is continuous
    ii. What is its scale of measurement (e.g. nominal, ordinal, interval, ratio)
        a. Date is ordinal (ordered and spaced consistently), and number of homework hours is ratio (cannot be lower than zero).

c. Is there a key?
    1. The key could be date in both Table #1 and Table #2 as they share this attribute which is distinct to the data provided.

d. Is there an attribute (or attributes) that would logically serve as a class for supervised learning? Meaning would it be logical/interesting to predict this attribute based on the other attributes? What would be the value of predicting this attribute?
    1. From this data, you could predict the attribute involving screen time application usage breakdown, so for example, if you open Instagram, you are more likely to have higher screen time than if you open Facebook. This would be helpful for the user to predict their phone usage, and it would provide useful information for companies.

4. What would be a common key you could use to identify instances across your tables?
    a. We could use default indexing like 0, 1, 2…, as distinct identifiers, as these are labeling each row so they are clearly different and usable.