

## Compare and contrast three approaches to processing data needed for your project

Your preliminary project proposals roughly fall into two categories:

- A) Projects that aim to *detect a person's behavioral or emotional state based on user-generated data* (e.g., using a combination of sensor-collected and self-reported data)
- B) Projects that aim to *use a dataset to create representations of personal health data* in order to provide insight to the user about what the data mean.

This activity is designed to help you identify the data that are important for your project and understand approaches to working with these data. Do these parts *in order* and in your project group, to maximize effectiveness.

**Part One: Describe your data.** Follow the steps below and document the work that you completed for each step according to the instructions given.

1. Pick **one** of the categories listed above (A or B) based on its relevance to your project proposal. Specify the category that you chose.
2. Examine the data set that you plan to collect (for Category A above) **or** the pre-existing data set that you have identified (for Category B above).
3. List the data types included in the data set that are relevant to your project goals and provide a one-sentence description for each. For data types that are closely related, put the data items into a group (but be sure to address individual data types in that group). For example, you might have an "IMU" group, which could include accelerometer data capturing acceleration in three dimensions (x, y, and z) and gyroscope data capturing angular rotation around three dimensions (x, y, and z). If this were all the inertial data that you planned to collect for your project, this group would have six data types listed with an explanation for each.
4. Now, go back through the list and explain *why, in detail*, each data type you listed is relevant to your project goals. For example, to measure movement of a user, it might be necessary to use angular rotation data to determine rotation of your accelerometer--since gravity provides a constant acceleration in a certain direction, understanding tilt of the sensors relative to the direction of gravity will make your measurements more meaningful.
5. If you chose Category A, take one more pass through your list and explain *how, in detail* you plan to acquire the data types you listed. For example, if you need angular rotation, will you use a MEMS gyroscope or something else? Will you acquire sensors and if so, what particular sensors will you use? How is the transfer of data handled and what is the sampling rate, precision, and accuracy of the

data provided by the sensor? In what format will the data be for processing (e.g., is there an API available, csv data export options, etc.?) If you chose Category B, explain *how, in detail* you plan to *represent the data* in a user interface. How will you structure and organize copious amounts of data of a particular type to allow the user to find relevant data points (e.g., for example, for multiple hours of continuous video, how will you allow the user to preview this data? Will you show screenshots as thumbnails, captured automatically at regular intervals, allow visual browsing of the set using interactive slider widgets, etc.)?

6. Now, take a final pass through your list and specify any *technical challenges you foresee* in working with each of the data types you listed. For example, if you are working with an existing data set, are there any missing data or data in an incorrect format? Which data are affected and to what extent? Is there heterogeneity in the sampling technique? If you are working with data that you expect the user to generate, what scenarios could lead to difficulties interpreting the data? Do you need to make calls to an API service that might limit the frequency of such calls?

## **Part Two: Discuss approaches to "working with your data" to get meaningful results.**

This part of the homework will depend on which general category your project falls in, A or B. You will identify and consider three approaches to working with your data set.

Follow the steps below and document the work that you completed for each step according to the instructions given.

1. Identify *three approaches to data processing* for your data set in order to get a meaningful "signal" from it.

For Category A, if you need to determine specific types of movement (e.g., sitting versus standing) and plan to use inertial sensors, how will you process data from these sensors to automatically classify these states? If you are reading physiological measurements, how will you determine variability in these data? Will you cross-reference additional data types to help contextualize elevated or depressed values? Will you use modeling approaches or heuristics? What data features are you considering for any planned experiments?

For Category B, will you create interaction or presentation techniques to allow the user to browse large data sets? How will these be designed and developed? For example, will you summarize some of the data or collapse data into "episodes" that could be browsed in a user interface? If so, how will you go about creating automated summaries or grouping episodes? Will you employ clustering approaches to group data? Which ones and why do you expect them to work?

2. Describe each approach that you identified (two to four paragraphs for each should be enough to capture the important details of *how* the approach works).
3. Now, discuss the strengths and limitations of each approach that you described in a separate section. Will one approach be computationally expensive? Will it require several training stages that are impractical in the proposed usage scenario?
4. Finally, revisit any technical challenges that you listed for Part One 6. For each, discuss any plans you have to mitigate the challenges you identified.

If you are finding that this activity is MORE valuable than what you had already specified for P2, you can turn in the results of this activity as your P2 if you choose to.