CANCER DATA GENIE

CASE CONTEXT: PROJECT GENIE

The American Association of Cancer Research (AACR) recently released data from Project GENIE (Genomics Evidence Neoplasia Information Exchange) into the public domain. The goal of this project was to collect data about patients with cancer who have been seen at several major care centers around the country, and share it to promote research and patient awareness about cancer. This infographic describes some of these goals.

The GENIE data set includes both clinical data about patients (age, gender, ethnicity, race, cancer type) as well as data about the genetic mutations that were found in each patient's tumor. For this case study, we will be focused on only the clinical data about patients (a file "data_clinical.txt" which you can access by following the instructions listed under TASK ONE below).

Most patients don't have the ability to analyze or learn from these date, but they may have heard of Project GENIE and be interested in it. **Your goal is to put yourselves in the shoes of a recently diagnosed cancer patient and build an application to help patients learn from Project GENIE's clinical data file.**

PRE-WORK

Prior to the session, we recommend the following:
- Review the infographic to understand the goals of the GENIE project
- Register and create a Synapse account via the Sage Bionetworks website

TASK ONE: ANALYSIS (SETUP INSTRUCTIONS)

The Project GENIE data is hosted and available for public access on a website maintained by Sage Bionetworks here:
https://www.synapse.org/#!Synapse:syn7844527.

Navigate to the above link. The data file you want is the one called "data_clinical.txt". In order to download this file, you will need to register for a Synapse account (upper-right button on the website that says "Register") and follow the instructions. You can use your Stanford e-mail address. Let us know if you have any difficulty downloading the clinical data file.

Once you download the file spend some time understanding the data set. You should see 18,967 rows of data (one per patient, plus one header row), and the following columns of data:

Data element	Definition
AGE_AT_SEQ_REPORT	The patient's age at the time their tumor was sequenced (interpret this as *the patient's age at the time they were diagnosed with cancer*)
CENTER	The name of the cancer center where the patient received care
ETHNICITY	The patient's ethnicity (unique values: Non-Spanish/non-Hispanic, Spanish/Hispanic, Unknown)

ONCOTREE_CODE	The primary cancer diagnosis code based on the OncoTree ontology (http://cbioportal.org/oncotree)
PATIENT_ID	The unique, anonymized patient identifier for the GENIE project
PRIMARY_RACE	The primary race encoded for the patient (unique values: Asian, Black, Native American, Other, Unknown, White)
SAMPLE_ID	The unique, anonymized sample identifier for the GENIE project
SAMPLE_TYPE	The type of tumor sample that was available for the patient (unique values: Primary, Metastasis, Unspecified)
SEQ_ASSAY_ID	The institutional assay identifier to describe what type of genetic testing was performed on the patient's tumor
SEX	The patient's gender code (unique values: Female, Male)
CANCER_TYPE	The primary cancer diagnosis "main type" based on the OncoTree ontology (http://cbioportal.org/oncotree)
CANCER_TYPE_DETAILED	The primary cancer diagnosis label, based on the OncoTree ontology (http://cbioportal.org/oncotree). This data point provides more detailed information on the sub-type of cancer that each patient has.

TASK TWO: DESIGN (BRAINSTORM PRODUCT)

Often when patients are diagnosed with cancer, they become very interested in learning more about their diagnosis, finding others who have taken a similar journey, and seeking a general sense of community. We have heard quite often that patients are interested in finding out more about "patients like themselves."

Using the GENIE data set above, brainstorm ways in which patients might be able to discover patients with similar demographic characteristics (e.g., age, gender) and the cancer diagnoses they have faced. Help us understand why this might be helpful for patients.

DELIVERABLE
For this task, envision the best possible patient-facing interface that would enable patients to access the data from Project GENIE and ask important questions:
- Create a list of insights or questions your product would seek to answer for patients.
- Create a rough sketch/mock-up of your ideal product/interface that could help answer these questions for patients.

Design Questions

Step 1: Understanding the Users
Who are the users of this product?
How comfortable are these users with technology? What other forms of technology do they most likely use?

What do the users want?

Step 2: Designing the Product
What product are we going to build? (An app? A website? Software?)
How can we organize the data the customers want?
What would it look like to an user if they opened the product? (Consider drawing a wireframe of what your interface would look like.)
How can we store the data efficiently? (What ADTs/data structures should we use?)

Step 3: Testing the Product
How would you test this product (both technically and in terms of user experience, ease of use, etc.)?
To whom would you show this product in order to test it?
What types of user interviews would you conduct? What would you want to collect feedback on? What observations would you make about the way the users use the product?

Wireframe Example