

PROJECT PROPOSAL

You can find a sample proposal here!

What is the title of your project?	Predicting Breast Cancer from Genes
Background (Write about 200-300 words about what your project focuses on)	<p>Twelve percent of women in the United States will be diagnosed with breast cancer at some point in their lives [1]. Though somewhat common, breast cancer remains difficult to predict and prevent. Early diagnosis is one of the best methods for a full recovery [2]. However, for low-income countries, equipment and resources for detecting breast cancer are not readily available, which decreases the rate of survival. Using various models in machine learning, trends based on genetics and demographics can be used to create models that predict whether someone is likely to have breast cancer and their probability of survival. The data and demographics containing both breast cancer patients and control patients comes from a public dataset that was previously used by others for a different purpose. The genes were sequenced and analyzed through a process known as RNA-seq, which allows us to measure RNA transcripts that are transcribed. The transcribed genes are broken into small cDNA segments, upon which adaptors are attached, further allowing us to duplicate these segments through a polymerase chain reaction (PCR) and analyze accordingly [4]. When finding a relationship between variables, we will try using models such as logistic regression (determining the probability of the presence of breast cancer) and neural networks (using gene variables and random forests to find survival probability) to try to find various results that would help diagnose breast cancer in a faster time frame.</p>
What is the subject area you're covering? (E.g. AI and Healthcare, Data Science and...)	Machine Learning in Medical Diagnostics

What is the dataset you're using? Please either add the link or name what the dataset is called.	https://www.pnas.org/doi/10.1073/pnas.1908252116
If you're doing a research paper, what is the research question you're aiming to answer?	Can we accurately predict breast cancer from patients using their gene expression data?
What is the final project form that you and your mentor have chosen? <ul style="list-style-type: none"> - Presentation - Research Paper - Short Write up - Other (Please explain) 	Short Write Up / Research Paper if time allows.

Background:


- Why do we care about breast cancer (1-2 sentences)
- Why is breast cancer hard to predict (1-2 sentences)
- What data are we using (2-3 sentences)
 - Gene Expression intro video [here](#)
- What models will we use (1-2 sentences)

[1] <https://jamanetwork.com/journals/jama/article-abstract/2721183>

[2] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5715522/>

[3] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4127616/>

[4] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2949280/> +

 Transcriptomics 1: Introduction to RNA-Seq (bioinformatics analysis of gene and isoform ...