Data Analytics Capstone Topic Approval Form

The purpose of this document is to help you clearly explain your capstone topic, project scope, and timeline. Identify each of the following areas so you will have a complete and realistic overview of your project. Your course instructor cannot approve your project topic without this information.

Student Name: Patricia Walker

Student ID: 011821194

Capstone Project Name: Breast Cancer Diagonstics

Project Topic: Determining atrributes that impact the malignancy of breast cancer.

Research Question: Can the attributes that are visible with less invasive methods predict malignancy saving the time, cost and negative patient experience associated with more invasive methods. 80% of breast biopsies determine the area of concern to be benign.

> **Hypothesis:** A combination of visual attributes can be used to determine the if the mass is cancerous prior to biopsy.

Context: With 80% of biopsies resulting in a non-cancerous determination this can put a strain on resources which will be more pronounced in underserved communities, by better predicting from attributes that can be found from less invasive procedures resources could be preserved for the patients most likely to have cancer. Since the standard of care of breast cancer screening begins in the early 40s if not sooner for those with a family history determining an appropriate time to recommend biopsies vs over-recommending could help prevent procedure fatigue, which may help with getting the right care at the right time without a patient feeling like the doctors are just pushing tests.

Data: *Identify data you will need to collect that are relevant to the situation or question.* The data will need to include outcome as well as visual attributes

> *Data set: Breast cancer Wisconsin (diagnostic) dataset via scikit learn.*
>
> Availability: The dataset is made available within scikit learn as provided for use for academic purposes.

Data Gathering: Data will be gathered via built method within scikit learn

Data Analytics Tools and Techniques: I will use data cleansing to remove those attributes that are not considered visual to be detected with the less invasive methods, then use various supervised learning methods to determine the best and then fine tune the model.

> **Justification of Tools/Techniques:** Various supervised learning models will produce good results, by comparing various methods I will be able to determine the best model and then fine tune for the best f-score as this is a medical model and accurately determining positive cases will be important.

Application Type, if applicable (select one):

- ☐ mobile
- ☐ web
- ☒ stand-alone

Programming/Development Language(s), if applicable: Python

Operating System(s)/Platform(s), if applicable: Click here to enter text.

Database Management System, if applicable: Click here to enter text.

Project Outcomes: The project will be a machine supervised learning model that can be used to help determine after MRI the likilhood that a breast bioposy would result in a determination of cancerous. This prediction can help with patient discussion as well as determining the appropriateness of more advance testin.

Projected Project End Date: 11/30/2025

Sources: https://www.lubbockonline.com/story/business/2012/04/06/making-case-having-breast-mri-breast-biopsy/15164929007/

Human Subjects or Proprietary Information
Does your project involve the potential use of human subjects? (Y/N): N
Does your project involve the potential use of proprietary company information? (Y/N): N

STUDENT SIGNATURE

Patricia Walker

By signing and submitting this form, you acknowledge that any cost associated with the development and execution of your data analytics solution will be your (the student) responsibility.

TO BE COMPLETED BY AN INSTRUCTOR

The capstone topic is approved by an instructor.

COURSE INSTRUCTOR SIGNATURE:

Jim Ashe, Ph.D. Mathematics
COURSE INSTRUCTOR APPROVAL DATE:
Thursday, September 18, 2025

Project Compliance with IRB (Y/N): Y