Project Proposal

**Team Members:**  Daniel Carpenter, Zachary Knepp & Sonaxy Mohanty

**Project Title:** Breast Cancer Diagnosis Prediction

**Description of the problem context:** Breast cancer is a disease in which cells in the breast grow out of control. It’s important to understand that most breast lumps are benign and not cancer (malignant). Any breast lump or change needs to be checked by a health care professional to find out if it is benign or malignant (cancer) and if it might affect your future cancer risk. If doctors find an area of concern on a screening test, or if an individual has symptoms that could mean breast cancer, then they may refer for a fine needle aspiration biopsy (FNA). In an FNA, the doctor uses a very thin, hollow needle attached to a syringe to withdraw (aspirate) a small amount of breast tissue or fluid from a suspicious area.

The information that the data set contains are computed features from a digitized image of a FNA of a breast mass. They describe characteristics of the cell nuclei present in the image. The objective is to predict whether the breast lump is benign or malignant based on these computed real valued features for each cell nucleus.

**Type of the problem:** Supervised learning. Methods including logistic regression, support vector machines and maybe decision tree can be investigated.

**Initial thoughts on techniques that might be used:** Preprocess the missing values with a valid numerical imputation method. Construct and validate the learning models that would predict the breast cancer lump correctly using performance measures.

**Hyperlink of the problem description:** <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>