**Capstone Project synopsis**

**Smita Misra**

**Title:**

**Breast Cancer prediction through routine blood work**

**Executive Summary**

The goal of this study is to try and develop a prediction model to assess the potential of the routine blood work parameters as biomarker for prediction of breast cancer.

**Motivation**

Early detection of breast cancer is one strategy to tackle the disease and ensure a greater probability of having a better treatment outcome. Routine clinical breast exam is one way of screening for early detection. After the age of 40 mammography and ultrasound are used in addition to clinical breast exam as a measure to monitor any abnormalities in the breast. Often denser breast tissues are difficult to screen and often detection happens in the late stage. Disparity in the breast density is one reason associated with the race African American women have higher breast density. Despite widespread mammography, breast cancer remains the second leading cause of death in women, [killing about 40,000 every year](http://www.cancer.org/cancer/breastcancer/overviewguide/breast-cancer-overview-key-statistics) in the United States. [Any decline](http://www.cancer.org/cancer/news/report-breast-cancer-death-rates-decline-but-more-slowly-among-poor) in the breast cancer mortality rate is likely [the result of improved treatment](http://www.ncbi.nlm.nih.gov/pubmed/12681269). Thus, there is a need for better predictive models. There has been a plethora of information about association of obesity with breast cancer. The possible reasons that obesity is linked with cancer include: Increased levels of insulin and insulin growth factor-1 (IGF-1), which may help some cancers develop. Several candidates for biomarkers of breast cancer have been reported in the literature. In this study we aim to assess some of the predictive modeling techniques on the parameters assessed during routine blood examination.

**Data Question**

Our analysis will look into the parameters like Age, Body mass index, Glucose levels, Insulin levels, HOMA, Leptin, Adiponectin, Resistin and MCP1 to derive a predictive model for prediction of the breast cancer using logistic regression.

**HOMA** (Homeostatic model assessment) is a method for assessing β-cell function and insulin resistance (IR) from basal (fasting) glucose and insulin or C-peptide concentrations. **Leptin** is a hormone that helps regulate your body weight by controlling your appetite and energy level. This test is used to figure out how much body fat you have. Generally, the amount of leptin in your blood is related to the amount of fat tissue in your body. Leptin is released into your blood by your fat stores. **Adiponectin** is a hormone released by fat cells. It helps regulate tissue inflammation and responsiveness to insulin. **Resistin** also known as adipose tissue-specific secretory factor (ADSF) or C/EBP-epsilon-regulated myeloid-specific secreted cysteine-rich protein (XCP1) is a cysteine-rich adipose-derived peptide hormone that in humans is encoded by the RETN gene. Chemokines such as monocyte chemoattractant protein (**MCP-1**) are key agonists that attract macrophages to tumors to destroy the tumors it has been shown that lower levels of MCP1 lead to worst cancer outcomes.

**References**:

Patrício M, et al. Using Resistin, glucose, age and BMI to predict the presence of breast cancer BMC Cancer (2018) 18:29,

Crisóstomo J, et al. Hyperresistinemia and metabolic dysregulation: the close crosstalk in obese breast cancer. Endocrine. 2016;53(2):433-42.

Cole KD, He HJ, Wang L. Breast cancer biomarker measurements and standards. Proteomics Clin Appl. 2013;7(1–2):17–29

**Schedule (February 15 – March 10)**

1. Get the Data (Request made on 5/27/2018 and received the data)
2. Clean & Explore the Data (6/30/2018)
3. Analyze and build the model (7/17/2018)
4. Document/Pitch your project with a Presentation (7/20/2018)

**Data Sources**

The data from 116 patients with 10 features e will be used for this study. The study is documented on the UCI ML dataset (https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra) and data will be requested from the corresponding author of the study **Dr. Miguel Patricio**, Laboratory of Biostatistics and Medical Informatics and IBILI - Faculty of Medicine, University of Coimbra, Azinhaga Santa Comba, Celas, 3000-548 Coimbra, Portugal.

**Known Issues and Challenges**

Known issue at the moment is the size of the dataset which is small for predictive modeling.