

# Data Analytics Approaches for Breast Cancer Survivability: Comparison of Data Mining Methods

Eyyub Y. Kibis\*<sup>1</sup>

Faculty: İ. Esra Büyükahtakın, PhD<sup>1</sup>, Ali Dag, PhD<sup>2</sup>

<sup>1</sup> *Wichita State University, Department of Industrial, Systems & Manufacturing Engineering*

<sup>2</sup> *University of South Dakota, Beacom School of Business*

In the early stages of breast cancer, inefficient treatment methods, as well as the patient's health condition may impact the patient's lifetime expectancy. In this study, given a set of explanatory variables that include the patient's demographics, health condition, and cancer treatment regimen, our objective is to investigate the performance of four different machine learning methods including an artificial neural network, classification and regression tree, logistic regression, and Bayesian belief network. We utilize these four methods with a ten-fold cross validation to predict the ten-year survivability of a breast cancer patient after initial diagnosis. The results of each method are compared with respect to accuracy, sensitivity, specificity, and area-under-the-curve (AUC) metrics. We observe that the logistic regression method shows better performance compared to the others with respect to the AUC metric. In all prediction models, the stage of the cancer is the most important predictor of breast cancer survivability.