**CIND 820 Capstone Project Abstract**

Sang Won Baek

Student ID: 501144120

Supervisor: Not Assigned Yet

May 16, 2022

According to the World Health Organization, cardiovascular diseases cause an estimated 17.9 million deaths annually and is the lead cause of death globally (Kaptoge et al., 2019). By the year 2030, the UN sustainable development goals outlined a concerted effort to reduce premature mortality from non-communicable diseases by a third (Kaptoge et al., 2019). Towards achieving this goal, there must be a precedent set to drastically reduce incidences of cardiovascular diseases worldwide. Early detection of at-risk individuals and targeting interventions for modifiable risk factors leading to cardiovascular diseases are crucial steps towards disease prevention (Dahlöf, 2010). There are a multitude of modifiable risk factors related to cardiovascular diseases including smoking, stress, abnormal lipids, abdominal obesity, lack of physical activity, diabetes mellitus, and more (Dahlöf, 2010). Cohort studies have shown that only 2 to 7 percent of the general global population have no risk factors at all (Dahlöf, 2010). Moreover, within the population of at-risk individuals, over 70 percent have multiple risk factors (Dahlöf, 2010). Detection of at-risk patients are often a difficult task as most adverse cardiovascular events occur in people with modest and often unnoticed elevations in multiple risk factors, rather than an outlier single factor (Dahlöf, 2010). Therefore, the advancement of machine learning technologies may usher in an exciting age of cardiovascular disease detection. Though, the integration of machine learning frameworks in the field of medicine has been a slow process, recent years have been a rapidly changing landscape (Al’Aref et al., 2019). With the availability of increasing computational power, new machine learning frameworks, and big data, the machine learning community can help make monumental changes to the healthcare sector (Al’Aref et al., 2019).

Carrying on the inspiration of the machine learning community’s efforts in the healthcare sector, the aim of my capstone project is to utilize classification (supervised learning) to develop a prediction algorithm. The focus will be on developing both a decision tree-based algorithm and a logistic regression model. The efficacy of both methods will be compared to determine which algorithm is more effective in predicting heart disease events. The capstone project will also include an initial data exploration and visualization phase. The dataset I will be using is the “Heart Failure Prediction Dataset” found publicly on Kaggle.com (fedesoriano, 2021). The dataset contains 12 attributes and 918 observations (fedesoriano, 2021). The dataset is a compilation of 5 different public datasets (found on the UCI Machine Learning Repository) containing common features (fedesoriano, 2021). The project will be programmed in python using Jupyter notebook. Multiple libraries will be used throughout the project including numpy, pandas, matplotlib, seaborn, sklearn, and more.

Overall, through this capstone project I hope to further my experience in supervised learning methods and gain hands on experience using several high-profile python libraries. Moreover, the ultimate goal of the capstone project will be to develop a classification algorithm to predict heart disease events and to raise awareness on cardiovascular disease.

**References**

Al’Aref, S. J., Anchouche, K., Singh, G., Slomka, P. J., Kolli, K. K., Kumar, A., ... & Min, J. K. (2019). Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *European heart journal*, *40*(24), 1975-1986.

Dahlöf, B. (2010). Cardiovascular disease risk factors: Epidemiology and risk assessment.*The American Journal of Cardiology, 105*(1, Supplement), 3A-9A. doi:<https://doi.org/10.1016/j.amjcard.2009.10.007>

fedesoriano. (September 2021). Heart Failure Prediction Dataset. Retrieved [May 15, 2022] from <https://www.kaggle.com/fedesoriano/heart-failure-prediction>.

Kaptoge, S., Pennells, L., De Bacquer, D., Cooney, M. T., Kavousi, M., Stevens, G., ... & Di Angelantonio, E. (2019). World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *The Lancet Global Health*, *7*(10), e1332-e1345.