

**Business Data Analytics Project Proposal**

Student’s full name: Veronika Kránicz

Student’s programme: MSc in IT for Business Data Analytics

**1. Data Set Selection**

The dataset that I have selected for this project is the “Heart Failure Prediction” dataset from Kaggle, available at: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>.

I chose this dataset because, in my current workplace, an IT consulting company, we had a client from the healthcare industry, and I encountered severe problems related to patient mortality risk prediction and other areas. Despite the availability of the relevant and huge sets of data, there was no model or any screening dashboards used. While looking for my Business Data Analytics Project dataset, this stood out, in terms of usability and accessibility.

Models like the one I aim to develop in my project could significantly benefit society by improving healthcare management and preventive care.

**2. Problem Definition**

The business problem revolves around the absence of a predictive model for mortality related to heart diseases, even though the necessary data is available. Additionally, healthcare management faces challenges in organizing regular screening tests. By implementing this model, a structured screening process could be built to identify high-risk patients early, enabling timely interventions and reducing mortality rates.

**3. Analytical Approach**

The project will follow a structured analytical approach just as in the previous semesters’ projects:

* **Data Preprocessing & Exploratory Data Analysis (EDA):** Handle missing values, perform feature selection, and visualize data distributions to identify key trends.
* **Feature Engineering:** Apply domain knowledge to create new predictive features.
* **Model Selection & Implementation:** Implement various classification models to determine the most effective approach.
* **Hyperparameter Tuning & Optimization:** Improve model performance using techniques such as grid search and cross-validation.
* **Model Validation & Evaluation:** Assess the models using performance metrics like accuracy, precision, recall, F1-score, and AUC-ROC.
* **Interpretability & Business Insights:** Use SHAP values and other interpretability tools to derive actionable insights from the model results.

**4. Expected Outcomes**

The expected outcome of this project is a predictive model for mortality risks related to heart diseases that can identify high-risk patients based on key clinical features. This will provide insights into the most significant factors contributing to heart failure risk, helping healthcare professionals make informed decisions.

Additionally, a well-documented GitHub repository will be created to contain all data preprocessing, model training, and evaluation steps.

Furthermore, business recommendations will be developed to guide healthcare providers on how predictive models can be integrated into their screening processes to improve patient care and reduce mortality rates.

**To sum up, this project aims to demonstrate how machine learning can be utilized to enhance healthcare decision-making, reduce mortality rates, and improve patient outcomes.**