

# **Patient Survival Prediction Using Machine Learning Models**

**Rana Demir 201401003**

**Zeynep Meriç Aşık 201410026**

All files are run on Python 3.8.5

The dataset that is worked on can be found as "dataset.csv" in the repository or can be downloaded from:  
<https://www.kaggle.com/datasets/mitishaagarwal/patient>

To be able to run all the files, run the command below in your preferred environment or terminal:

```
pip install -r requirements.txt
```

All the work that is done can be found as ipynb files date by date. The last version and the one containing the final models and work is "survivalprediction-18.12.2022.ipynb".

1) survivalprediction-18.12.2022.ipynb:

This file consists mainly of 2 main parts. First part is training machine learning models with default parameters. In this part preprocessing is done by traditional ways which includes examining the Pearson Correlation for numerical features and models like Naïve Bayes, LightGBM, XGBoost, Random Forest classifier are trained with their default parameters. Then, the prediction results are evaluated by accuracy, precision, recall, f1-score, and ROC-AUC metrics.

In the second part, preprocessing is done with a different way of feature selection that is calculating chi-squares and combining the compatible features to get higher results choosing the more beneficial features. In this process, Polychoric Correlation is also used to investigate the relation between categorical features. The dataset is also down sampled to increase the true positive results since the original dataset is highly imbalanced.

Afterwards, hyperparameter optimization is done by using Random Search algorithm and the best model is chosen using the best parameters. SHAP is applied to this model and five best features are used to train the model for the web app.

2) app.py:

To see the deployment of this model, app.py file is required. After going to the Streamlit website, this application can be deployed automatically. You can find the deployed version on this link:  
<https://meric2-patient-survival-prediction-yap470-app-rmgogf.streamlit.app/>

To use the app, the given parameters in the slider bar should be adjusted according to the patient results or a csv file should be uploaded in the format of the example.csv file. After filling the form, the prediction result can be seen in real-time.