Dalla Lana
School of Public Health

Datathon 2: Assessing Mortality Risk in Heart Failure Patients

Team 20: Hunter Pozzebon, Sarra Ali, Brenleigh Jebb

UNIVERSITY OF TORONTO
DALLA LANA SCHOOL OF PUBLIC HEALTH

1. Can the features predict mortality?
2. Are any features associated with the outcome of mortality?
3. Do the associated features predict mortality any better?

Dalla Lana School of Public Health

Data analysis process and main findings

Clean Data (No missing data, No outliers of concern)

⇩

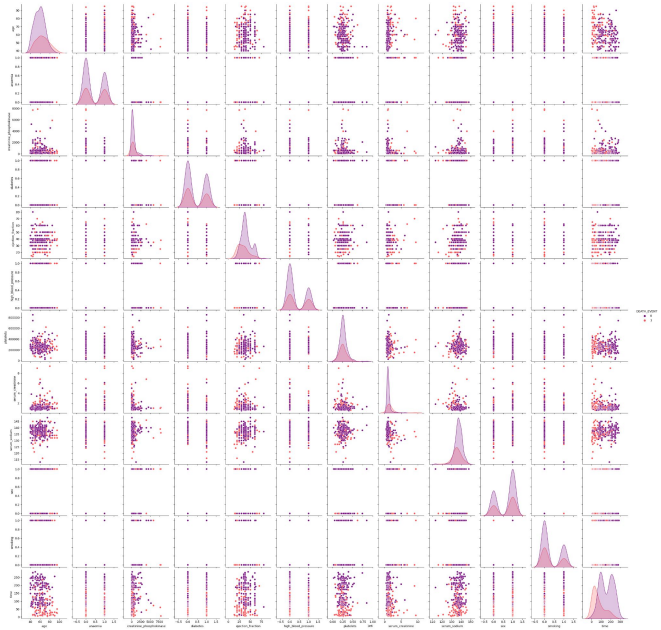
Pre-Processing (Data visualization showed no significant associations, further analysis needed)

⇩

Fit Logistic regression

⇩

Performance

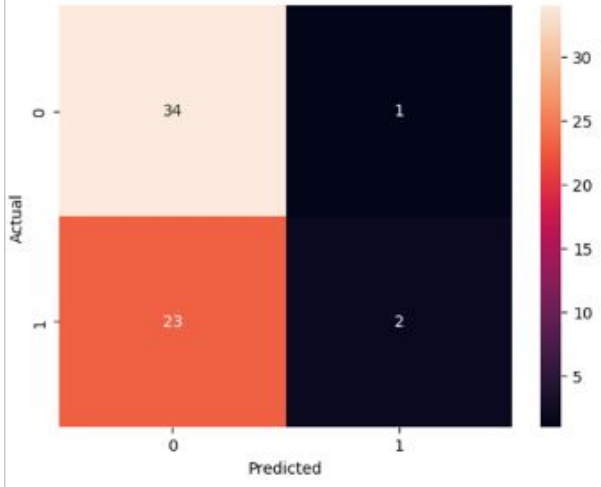


```
Optimization terminated successfully.
         Current function value: 0.492108
         Iterations 6
                           Logit Regression Results
==============================================================================
Dep. Variable:            DEATH_EVENT   No. Observations:                  299
Model:                          Logit   Df Residuals:                      287
Method:                          MLE   Df Model:                           11
Date:               Sun, 01 Oct 2023   Pseudo R-squ.:                  0.2160
Time:                       21:39:56   Log-Likelihood:                -147.14
converged:                      True   LL-Null:                       -187.67
Covariance Type:            nonrobust   LLR p-value:                 9.169e-13
============================================================================================
                               coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                        4.9643      4.601      1.079      0.281      -4.054      13.982
smoking                      0.1356      0.349      0.389      0.697      -0.548       0.819
age                          0.0557      0.013      4.241      0.000       0.030       0.081
anaemia                      0.4179      0.301      1.389      0.165      -0.172       1.008
creatinine_phosphokinase     0.0003      0.000      2.034      0.042    1.06e-05       0.001
diabetes                     0.1514      0.297      0.509      0.611      -0.431       0.734
ejection_fraction           -0.0703      0.015     -4.731      0.000      -0.099      -0.041
high_blood_pressure          0.4189      0.306      1.369      0.171      -0.181       1.019
platelets                -7.094e-07   1.62e-06     -0.439      0.661   -3.88e-06    2.46e-06
serum_creatinine             0.6619      0.173      3.817      0.000       0.322       1.002
serum_sodium                -0.0567      0.033     -1.698      0.090      -0.122       0.009
sex                         -0.3990      0.351     -1.137      0.255      -1.087       0.289
============================================================================================
```

```
Classification Report:
              precision    recall  f1-score   support

           0       0.60      0.97      0.74        35
           1       0.67      0.08      0.14        25

    accuracy                           0.60        60
   macro avg       0.63      0.53      0.44        60
weighted avg       0.63      0.60      0.49        60
```



Dalla Lana
School of Public Health

Model Improvements and Implications

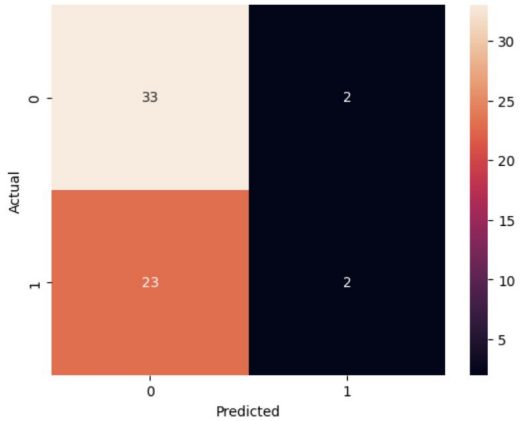
Additional models

⇩

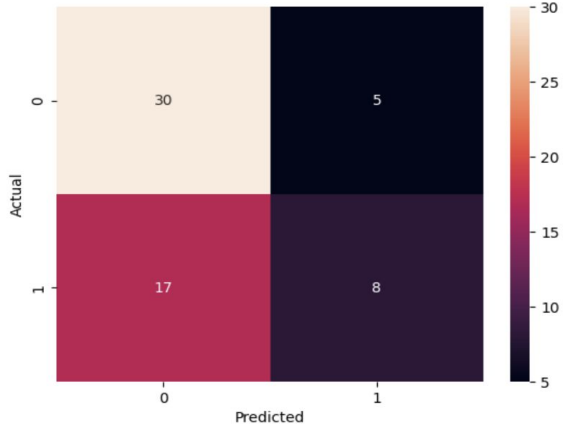
Performance

	coef	std err	z	P>|z|
const	4.9643	4.601	1.079	0.281
smoking	0.1356	0.349	0.389	0.697
age	0.0557	0.013	4.241	0.000
anaemia	0.4179	0.301	1.389	0.165
creatinine_phosphokinase	0.0003	0.000	2.034	0.042
diabetes	0.1514	0.297	0.509	0.611
ejection_fraction	-0.0703	0.015	-4.731	0.000
high_blood_pressure	0.4189	0.306	1.369	0.171
platelets	-7.094e-07	1.62e-06	-0.439	0.661
serum_creatinine	0.6619	0.173	3.817	0.000
serum_sodium	-0.0567	0.033	-1.698	0.090
sex	-0.3990	0.351	-1.137	0.255

Pseudo-R-squared = 0.085
Accuracy = 58%



Pseudo-R-squared = 0.192
Accuracy = 63%



Implications

Use with CAUTION!

UNIVERSITY OF TORONTO
DALLA LANA SCHOOL OF PUBLIC HEALTH