POLSKO-JAPOŃSKA AKADEMIA TECHNIK KOMPUTEROWYCH

Sebastian Szczepański S29805 Warszawa, 21.01.2024r.

1. Cel: Umożliwienie predykcji zgonu na podstawie parametrów biometrycznych i trybu życia pacjenta.

2. Pozyskanie danych:

Heart Failure Prediction

Dataset · 4y ago · by Larxel

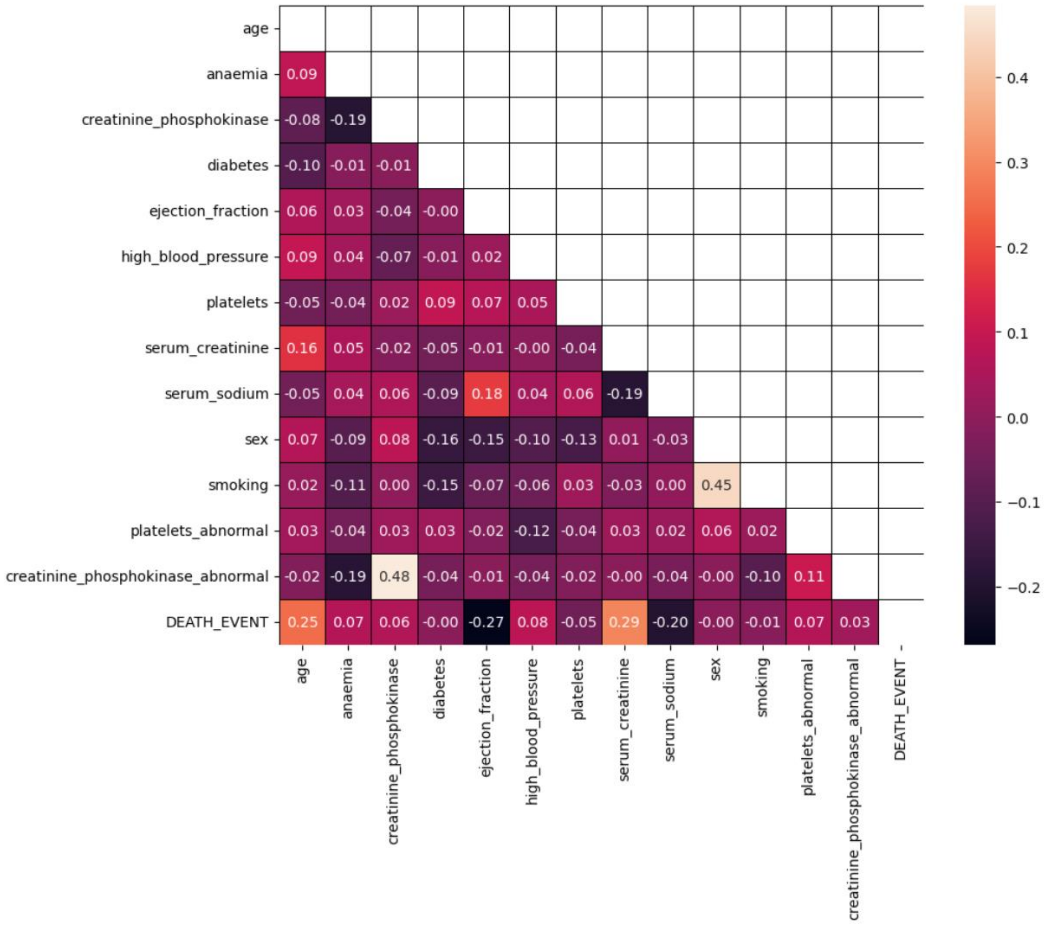
12 clinical features por predicting death events.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 299 entries, 0 to 298
Data columns (total 13 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   age                       299 non-null    float64
 1   anaemia                   299 non-null    int64
 2   creatinine_phosphokinase  299 non-null    int64
 3   diabetes                  299 non-null    int64
 4   ejection_fraction         299 non-null    int64
 5   high_blood_pressure       299 non-null    int64
 6   platelets                 299 non-null    float64
 7   serum_creatinine          299 non-null    float64
 8   serum_sodium              299 non-null    int64
 9   sex                       299 non-null    int64
 10  smoking                   299 non-null    int64
 11  time                      299 non-null    int64
 12  DEATH_EVENT               299 non-null    int64
dtypes: float64(3), int64(10)
memory usage: 30.5 KB
```

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
0	75.0	0	582	0	20	1	265000.00	1.9	130	1	0	4	1
1	55.0	0	7861	0	38	0	263358.03	1.1	136	1	0	6	1
2	65.0	0	146	0	20	0	162000.00	1.3	129	1	1	7	1
3	50.0	1	111	0	20	0	210000.00	1.9	137	1	0	7	1
4	65.0	1	160	1	20	0	327000.00	2.7	116	0	0	8	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
294	62.0	0	61	1	38	1	155000.00	1.1	143	1	1	270	0
295	55.0	0	1820	0	38	0	270000.00	1.2	139	0	0	271	0
296	45.0	0	2060	1	60	0	742000.00	0.8	138	0	0	278	0
297	45.0	0	2413	0	38	0	140000.00	1.4	140	1	1	280	0
298	50.0	0	196	0	45	0	395000.00	1.6	136	1	1	285	0

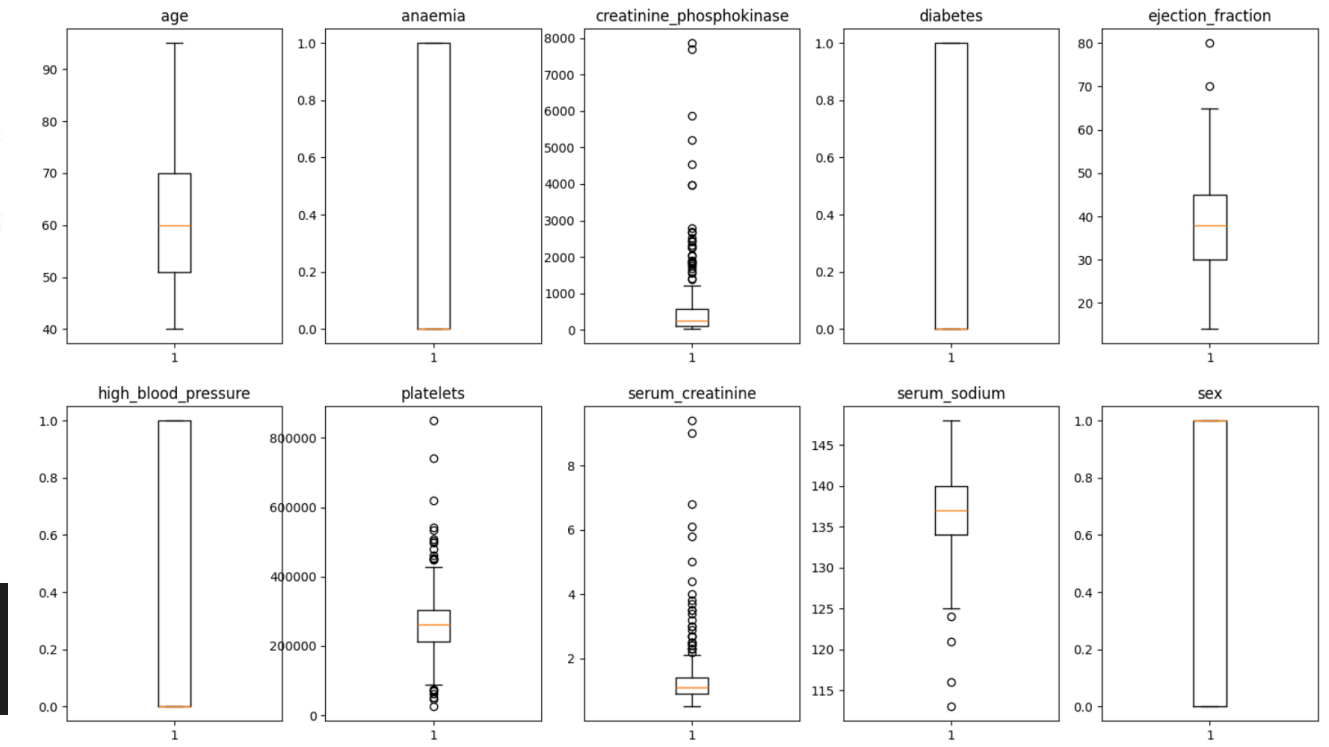
299 rows × 13 columns

3&4. Wstępna ocena danych i przygotowanie danych:



	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	platelets_abnormal	creatinine_phosphokinase_abnormal
0	75.0	0	582	0	20	1	265000.00	1.9	130	1	0	0	1
1	55.0	0	7861	0	38	0	263358.03	1.1	136	1	0	0	1
2	65.0	0	146	0	20	0	162000.00	1.3	129	1	1	0	0
3	50.0	1	111	0	20	0	210000.00	1.9	137	1	0	0	0
4	65.0	1	160	1	20	0	327000.00	2.7	116	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
294	62.0	0	61	1	38	1	155000.00	1.1	143	1	1	0	0
295	55.0	0	1820	0	38	0	270000.00	1.2	139	0	0	0	1
296	45.0	0	2060	1	60	0	742000.00	0.8	138	0	0	1	1
297	45.0	0	2413	0	38	0	140000.00	1.4	140	1	1	1	1
298	50.0	0	196	0	45	0	395000.00	1.6	136	1	1	0	0

299 rows × 13 columns




```
DEATH_EVENT
0    203
1     96
Name: count, dtype: int64
```

5&6&7. Modelowanie, ewaluacja modelu i wdrożenie:

	Random Forest Classifier	Decision Tree Classifier	Logistic Regression
Hiperparametry:	max_leaf_nodes=10, random_state=42	ccp_alpha=0.08965, max_depth=2, max_leaf_nodes=3, random_state=42	random_state=42
Cross Validation Score – średnia z 4 grup:	**73%**	**73%**	**73%**
Dokładność:	75%	67%	**77%**
Precyzja:	75%	48%	**78%**
Pełność:	32%	**58%**	37%
F miara:	44%	**52%**	50%