



Wprowadzenie do Sztucznej Inteligencji

Marta Arendt
Maciej Mechliński
Michał Gąsecki
Stanisław Rachwał



Fundusze
Europejskie
Polska Cyfrowa



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego

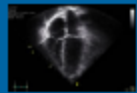
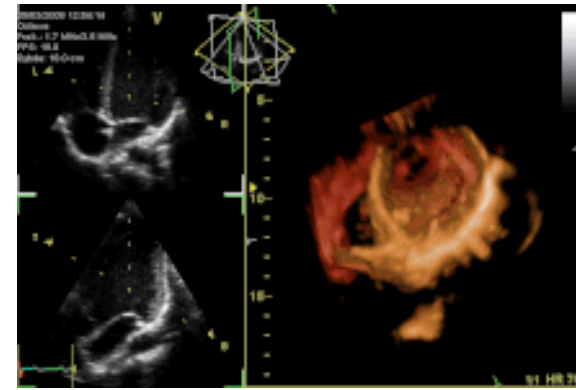


Projekt współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Rozwoju Regionalnego
Program Operacyjny Polska Cyfrowa na lata 2014-2020.

Oś priorytetowa nr 3 „Cyfrowe kompetencje społeczeństwa”, działanie nr 3.2 „Innowacyjne rozwiązania na rzecz aktywizacji cyfrowej”.

Tytuł projektu: „Akademia Innowacyjnych Zastosowań Technologii Cyfrowych (AI Tech)”.

Baza danych: "Echocardiogram"



Echocardiogram

Donated on 2/28/1989

Data for classifying if patients will survive for at least one year after a heart attack

Dataset Characteristics

Multivariate

Subject Area

Life

Associated Tasks

Classification

Attribute Type

Categorical, Integer, Real

Instances

132

Attributes

12

Standaryzacja danych

```
df_final = pd.get_dummies(df_dropped_ssa, columns = discrete, drop_first = True)
scaler = StandardScaler()
df_final[continuous] = scaler.fit_transform(df_final[continuous])
```

```
df_final.head(10)
```

	survival	still_alive	age_at_heart_attack	fractional_shortening	epss	lvdd	wall_motion_index	pericardial_effusion_1
10	-0.501329	1	2.158795	-0.710432	0.478508	-0.740287	0.716144	0
14	-1.160000	1	-0.049125	-0.812342	1.464339	1.204188	1.780111	0
16	-1.160000	1	0.981238	0.614404	-0.225657	-0.220109	0.394946	1
17	-1.160000	1	0.028741	-1.321895	1.041840	0.461077	1.117641	1
19	-1.125334	1	0.539654	0.206762	0.337675	0.708780	1.619512	1
20	-1.142667	1	0.981238	-0.506611	-0.084824	0.696395	0.364834	0
22	-1.160000	1	1.570017	0.308672	0.018406	1.526203	0.113898	0
23	-0.847998	1	1.275627	-0.302790	-1.774821	-0.220109	-0.889844	0
24	2.133356	0	0.245265	-0.098969	-0.943906	-1.669176	-0.669021	0
25	0.816013	0	-1.226682	1.022046	-0.788990	-1.210924	0.449148	0

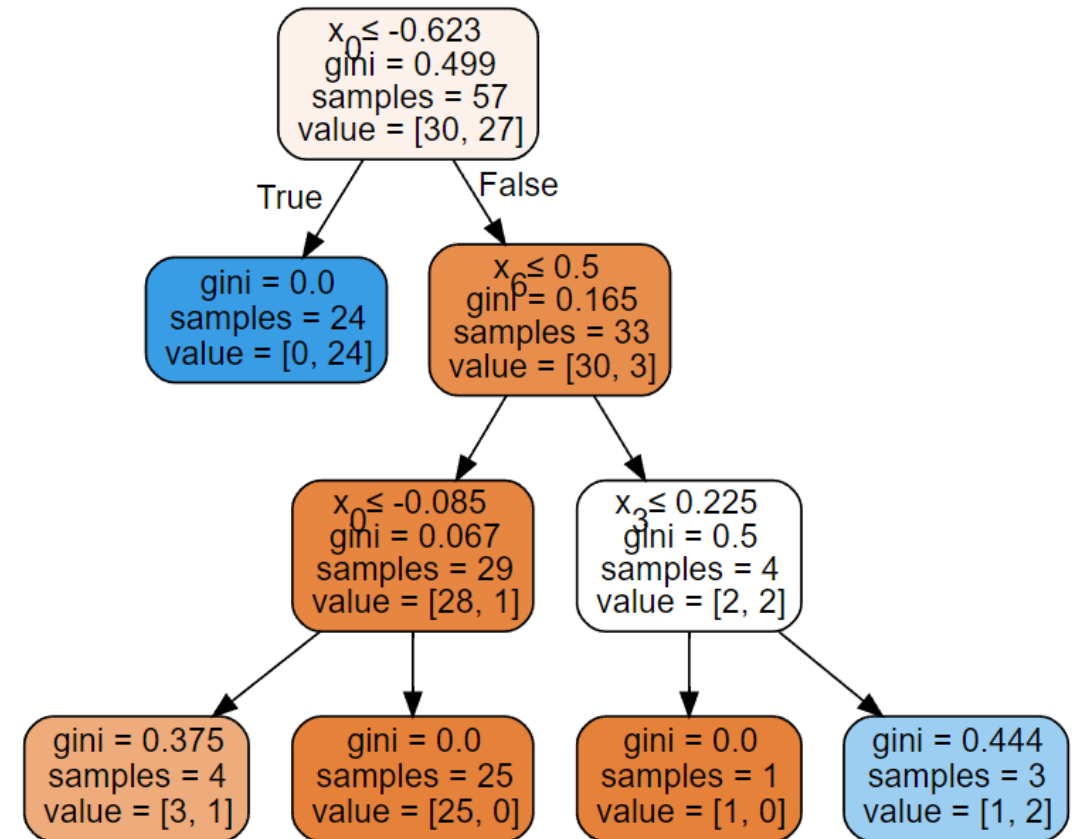
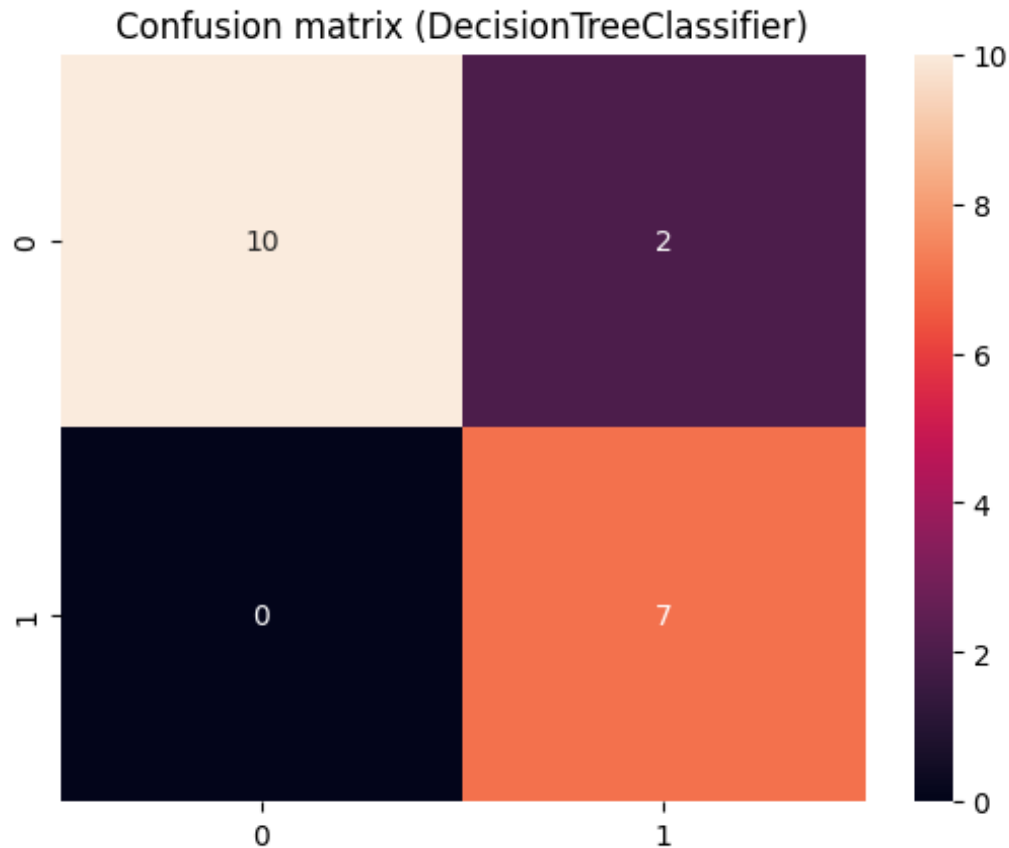
Podział na zbiory – dane ustandaryzowane

```
X = df_final.drop(['still_alive'], axis=1)
y = df_final['still_alive']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)
X_train.shape, X_test.shape
```

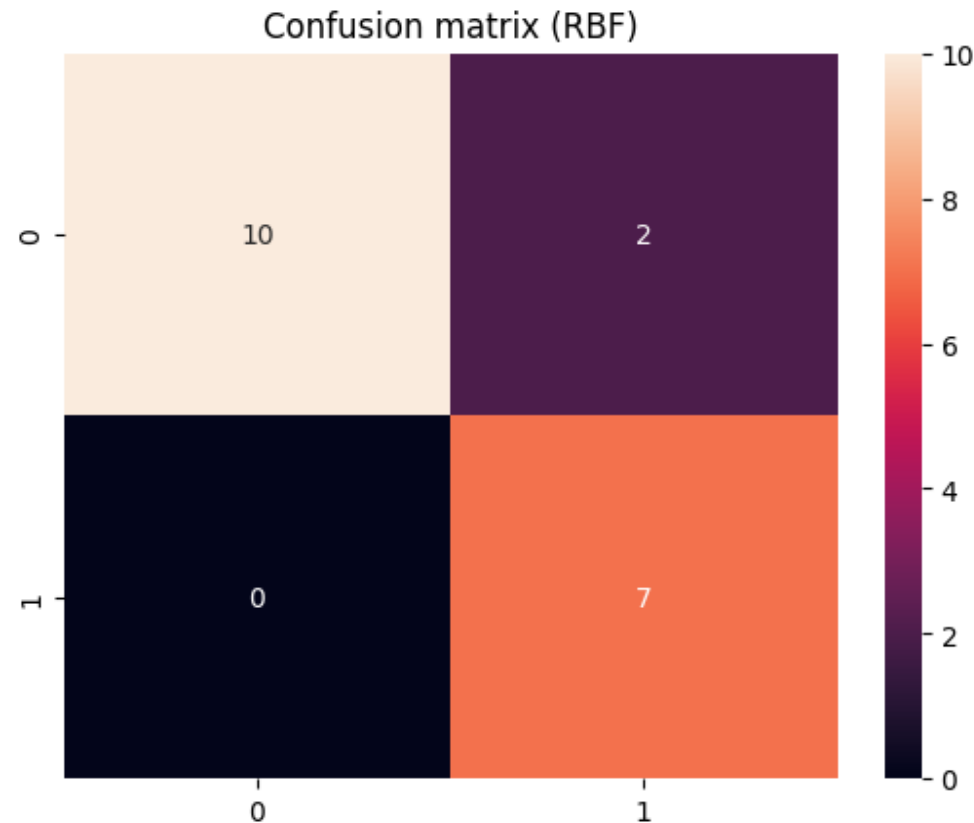
```
((57, 7), (19, 7))
```

Drzewa decyzyjne – DecisionTreeClassifier()



Accuracy: 0.9473684210526315

SVM – SVC(kernel='rbf', gamma=0.5, C=1)



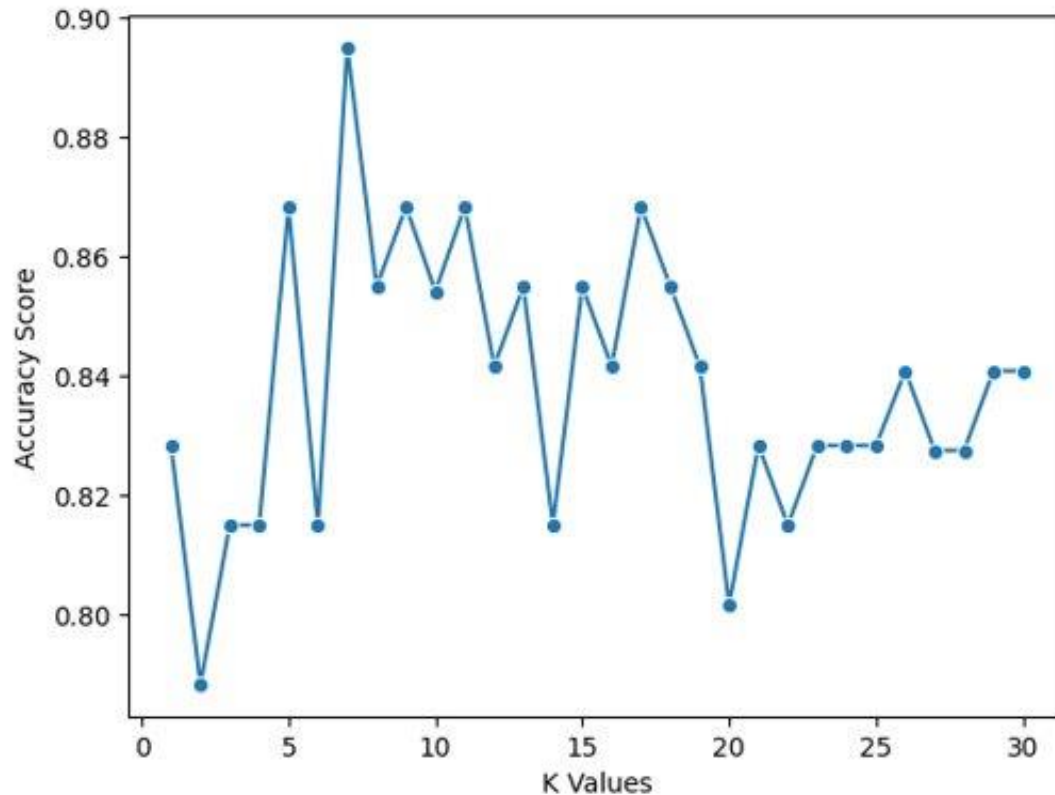
```
accuracy_score(y_test, y_pred)
```

```
0.8947368421052632
```

```
scores = cross_val_score(clf, X, y, cv=10)  
scores
```

```
array([1.         , 0.75        , 0.75        , 1.         , 1.         ,  
       0.875       , 0.85714286, 0.71428571, 0.71428571, 0.85714286])
```

KNN – kNeighborsClassifier()



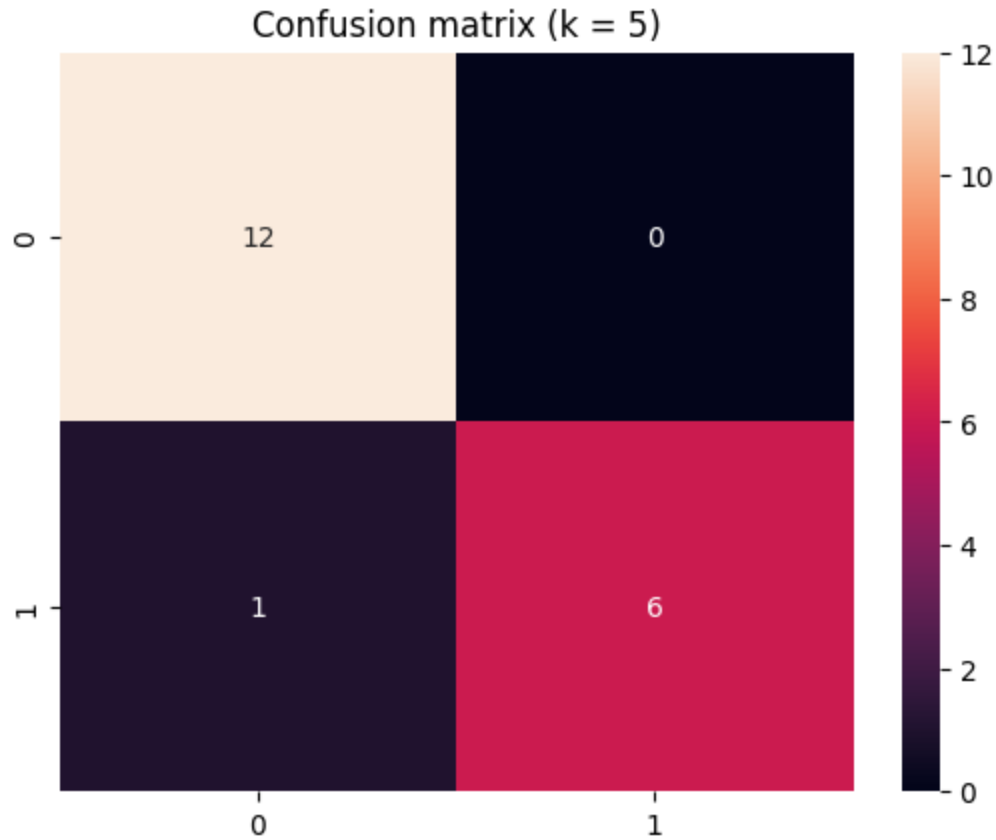
```
clf1 = KNeighborsClassifier(n_neighbors=5)
clf1.fit(X_train, y_train)
y_pred1 = clf1.predict(X_test)
```

```
clf2 = KNeighborsClassifier(n_neighbors=7)
clf2.fit(X_train, y_train)
y_pred2 = clf2.predict(X_test)
```

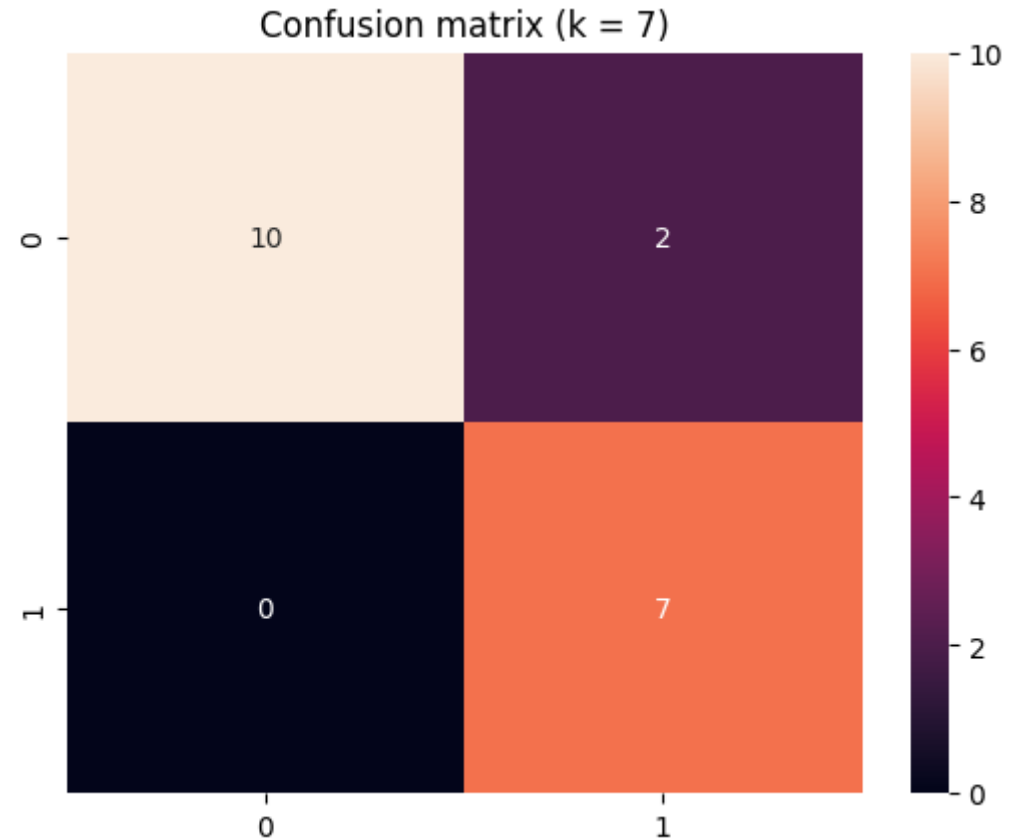
```
accuracy1 = accuracy_score(y_test, y_pred1)
accuracy2 = accuracy_score(y_test, y_pred2)
print("Accuracy k = 5:", accuracy1)
print("Accuracy k = 7:", accuracy2)
```

Accuracy k = 5: 0.8421052631578947
Accuracy k = 7: 0.8947368421052632

KNN – kNeighborsClassifier()



Accuracy k = 5: 0.8421052631578947



Accuracy k = 7: 0.8947368421052632

Podsumowanie statystyk

	Drzewa decyzyjne	Maszyna wektorów nośnych	Algorytm kNN	Literatura [1]	Literatura [2]
accuracy	94,73%	89,47%	89,47%	60%	76%
f1-score	92,3%	87,50%	87,50%	brak	brak

Wnioski

- Baza jest za mało zbalansowana, by dało się na niej wytrenować w pełni skuteczny model.
- Zaszła potrzeba mocnej modyfikacji danych, ze względu na dużą ilość pustych komórek w bazie.
- Bazę dodatkowo ograniczyło przeczyszczanie danych zgodnie z opisem bazy.
- Klasyfikator drzew decyzyjnych osiągnął lepszą dokładność niż algorytmy wykorzystane w [1] i [2] (kNN, opracowany przez autorów, ECL, itp.).
- Uzyskano lepsze wyniki dla klasyfikatora kNN niż w [1], prawdopodobnie ze względu na standaryzację danych

Bibliografia

- [1] Marc Sebban and Richard Nock and Stéphane Lallich. Stopping Criterion for Boosting-Based Data Reduction Techniques: from Binary to Multiclass Problem. Journal of Machine Learning Research, 3. 2002. (data dostępu: 01.06.2023 r.)
- [2] Federico Divina and Elena Marchiori. Handling Continuous Attributes in an Evolutionary Inductive Learner. Department of Computer Science Vrije Universiteit. (data dostępu: 01.06.2023 r.)
- [3] [www.kaggle.com%2Fcode%2Fadityamahimkar%2Fechocardiogram-eda-prediction-92&h=AT1c7cPreFbFGXrCMCoX4nOwlhtZGKxMqImfdobJnWcfGVTOuqa07z7ffx_sfwwlJr83Ckb_LgzLxJMj5FmPYpoYwMWI0fxWsPpBBRd_kLIJmnb3AYZl98RPm8dToQ](https://www.kaggle.com/code/fadityamahimkar/echocardiogram-eda-prediction-92&h=AT1c7cPreFbFGXrCMCoX4nOwlhtZGKxMqImfdobJnWcfGVTOuqa07z7ffx_sfwwlJr83Ckb_LgzLxJMj5FmPYpoYwMWI0fxWsPpBBRd_kLIJmnb3AYZl98RPm8dToQ) (data dostępu: 01.06.2023 r.)

Dziękujemy

Maciej Mechliński, Marta Arendt

Stanisław Rachwał, Michał Gąsecki



**Fundusze
Europejskie**
Polska Cyfrowa



**Rzeczpospolita
Polska**

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego



Projekt współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Rozwoju Regionalnego
Program Operacyjny Polska Cyfrowa na lata 2014-2020.

Oś priorytetowa nr 3 „Cyfrowe kompetencje społeczeństwa”, działanie nr 3.2 „Innowacyjne rozwiązania na rzecz aktywizacji cyfrowej”.

Tytuł projektu: „Akademia Innowacyjnych Zastosowań Technologii Cyfrowych (AI Tech)”.