

Zjawisko under-fittingu oraz over-fittingu na przykładzie modeli opartych na metodzie drzew decyzyjnych

J48

Jak możemy zauważyć drzewo decyzyjne jest mocno rozbudowane. Mimo zwiększonego skomplikowania osiągnięta dokładność nie odbiega od wyników przy zastosowaniu mniej skomplikowanego modelu (96,05%).

```
petal_width_in_cm <= 0.6: Iris-setosa (50.0)
petal_width_in_cm > 0.6
|   petal_width_in_cm <= 1.7
|   |   petal_length_in_cm <= 4.9
|   |   |   petal_width_in_cm <= 1.5: Iris-versicolor (45.0)
|   |   |   petal_width_in_cm > 1.5
|   |   |   |   petal_width_in_cm <= 1.6: Iris-versicolor (2.0)
|   |   |   |   petal_width_in_cm > 1.6: Iris-virginica (1.0)
|   |   |   petal_length_in_cm > 4.9
|   |   |   |   petal_width_in_cm <= 1.5: Iris-virginica (3.0)
|   |   |   |   petal_width_in_cm > 1.5
|   |   |   |   |   sepal_length_in_cm <= 6.9: Iris-versicolor (2.0)
|   |   |   |   |   sepal_length_in_cm > 6.9: Iris-virginica (1.0)
|   |   petal_width_in_cm > 1.7
|   |   |   petal_length_in_cm <= 4.8
|   |   |   |   sepal_length_in_cm <= 5.9: Iris-versicolor (1.0)
|   |   |   |   sepal_length_in_cm > 5.9: Iris-virginica (2.0)
|   |   |   petal_length_in_cm > 4.8: Iris-virginica (43.0)
```

Number of Leaves : 10

Size of the tree : 19

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0325	
Root mean squared error	0.1623	
Relative absolute error	7.2887 %	
Root relative squared error	34.3344 %	
Total Number of Instances	51	

Decision Stump

Model przy zastosowaniu metody Decision Stamp budowany jest wyłącznie do 1 poziomu drzewa. Mocno ograniczone drzewo bazuje wyłącznie na jednej z dostępnych cech przy klasyfikacji, co może w znacznym stopniu przekładać się na dokładność modelu.

```
Decision Stump

Classifications

petal_length_in_cm <= 2.45 : Iris-setosa
petal_length_in_cm > 2.45 : Iris-versicolor
petal_length_in_cm is missing : Iris-setosa

Class distributions

petal_length_in_cm <= 2.45
Iris-setosa      Iris-versicolor Iris-virginica
1.0      0.0      0.0
petal_length_in_cm > 2.45
Iris-setosa      Iris-versicolor Iris-virginica
0.0      0.5      0.5
petal_length_in_cm is missing
Iris-setosa      Iris-versicolor Iris-virginica
0.3333333333333333      0.3333333333333333      0.3333333333333333

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      32      62.7451 %
Incorrectly Classified Instances    19      37.2549 %
Kappa statistic                     0.4507
Mean absolute error                 0.2357
Root mean squared error             0.3438
Relative absolute error             52.911 %
Root relative squared error         72.713 %
Total Number of Instances          51
```

Rysunek 2 Opis modelu Decision Stump