

# Zadanie 4 - Raport

Jan Stusio

Maj 2024

## 1 Wstęp

Celem zadania jest zbadanie własności klasyfikatora SVM (z ang. Support Vector Machine, po pol. Maszyny wektorów nośnych) oraz drzewa decyzyjnego. Model SVM ma przewidywać wartości dyskretne – klasy, gdzie poszukujemy funkcji decyzyjnej, która najlepiej oddziela dane. Gdy zadanie jest liniowo nieseparowalne, stosuje się funkcję jądra, która mapuje dane do przestrzeni o wyższej wymiarowości. Drzewo decyzyjne jest modelem, który w każdym węźle dokonuje podziału na podzbiory, aż do osiągnięcia warunku stopu.

Klasyfikator wskazuje dla danego  $x$  najbardziej prawdopodobną wartość  $y$

## 2 Implementacja

Implementacja ładowania zbioru danych oraz treningu i ewaluacji przy wykorzystaniu walidacji krzyżowej (z założeniem 5 podziałów zbioru danych).

## 3 Badania

### 3.1 Badania wpływu parametrów w klasyfikatorze SVM:

1. Siła regularyzacji
2. Funkcja jądra
3. Liczba iteracji (bardzo niski parametr `tol` w funkcji `sklearn.svm.SVC`)

### 3.2 Badania wpływu parametrów w drzewie decyzyjnym:

1. Kryterium oceny
2. Technika podziału węzła
3. Maksymalna głębokość drzewa

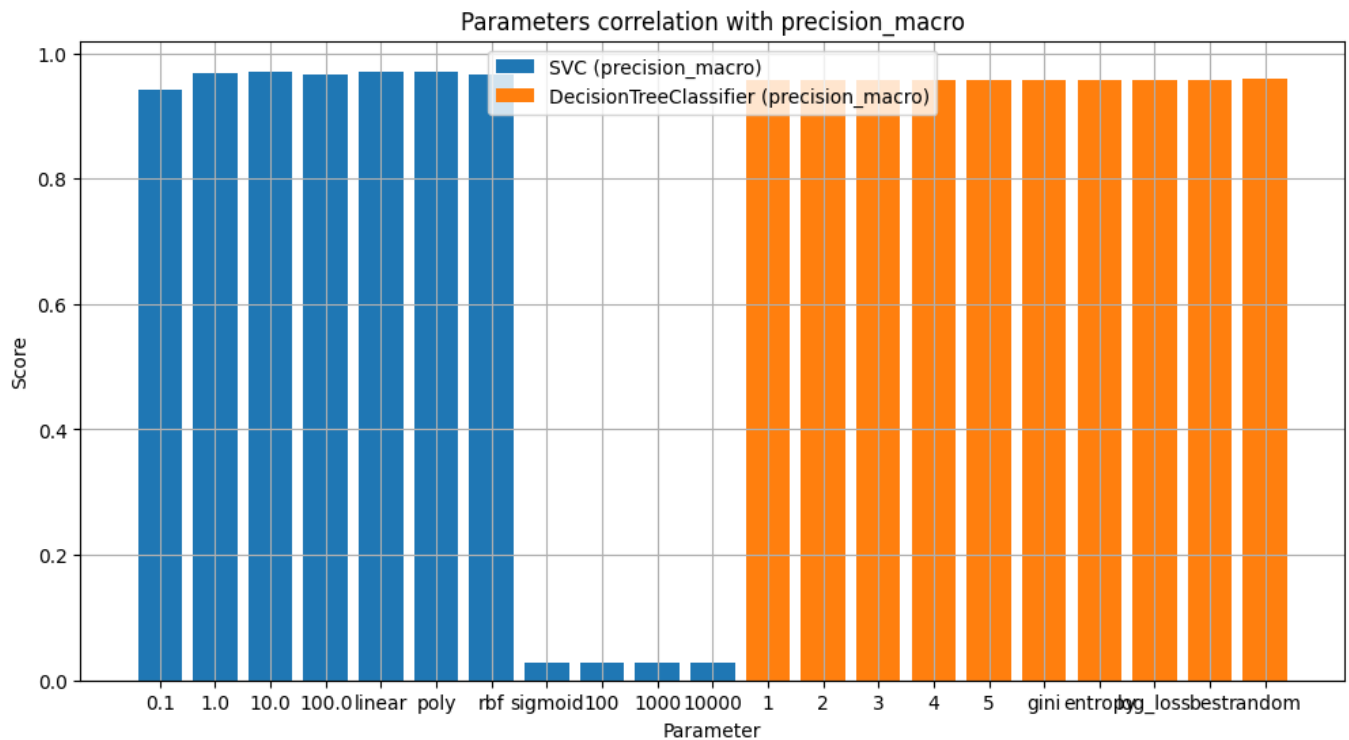
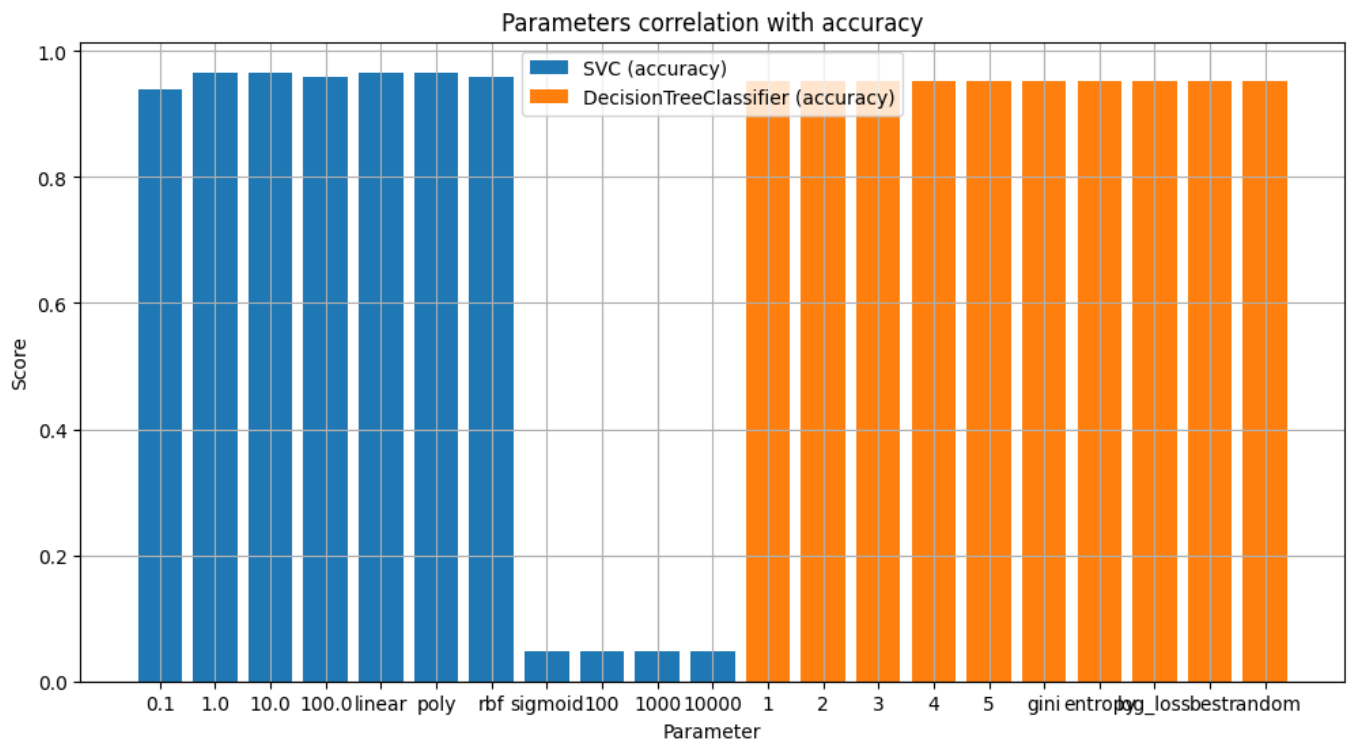
## 4 Wyniki eksperymentów

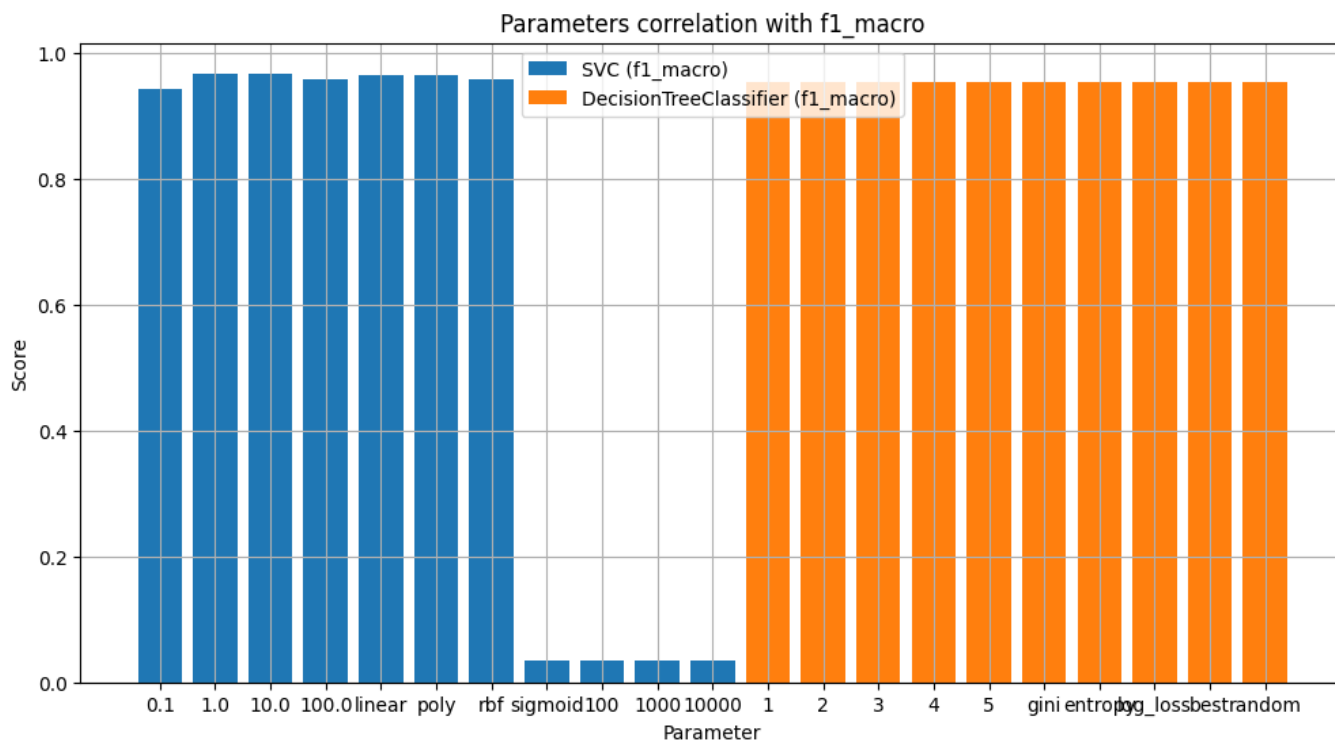
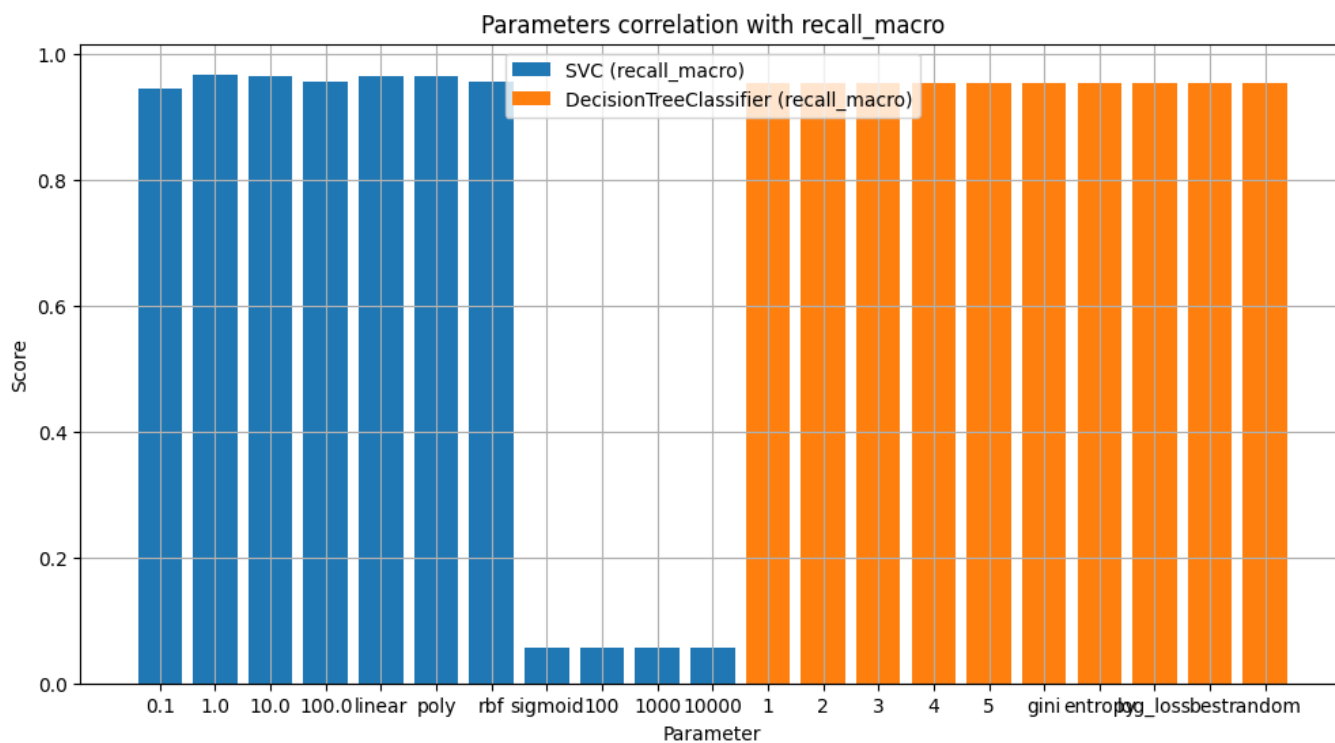
### 4.1 Tabela porównująca wyniki klasyfikatorów SVM i drzewa decyzyjnego.

Model	Parameter	Value	Metric	Score
SVC	C	0.100000	accuracy	0.940000
SVC	C	1.000000	accuracy	0.966667
SVC	C	10.000000	accuracy	0.966667
SVC	C	100.000000	accuracy	0.960000
SVC	C	0.100000	precision_macro	0.941717
SVC	C	1.000000	precision_macro	0.967222
SVC	C	10.000000	precision_macro	0.971453
SVC	C	100.000000	precision_macro	0.965527
SVC	C	0.100000	recall_macro	0.943990
SVC	C	1.000000	recall_macro	0.967222
SVC	C	10.000000	recall_macro	0.964921
SVC	C	100.000000	recall_macro	0.956402
SVC	C	0.100000	F1	0.940851
SVC	C	1.000000	F1	0.966981
SVC	C	10.000000	F1	0.965934
SVC	C	100.000000	F1	0.957527
SVC	kernel	linear	accuracy	0.966667
SVC	kernel	poly	accuracy	0.966667
SVC	kernel	rbf	accuracy	0.960000
SVC	kernel	sigmoid	accuracy	0.046667
SVC	kernel	linear	precision_macro	0.971082
SVC	kernel	poly	precision_macro	0.971082
SVC	kernel	rbf	precision_macro	0.965527
SVC	kernel	sigmoid	precision_macro	0.027103
SVC	kernel	linear	recall_macro	0.963810
SVC	kernel	poly	recall_macro	0.963810
SVC	kernel	rbf	recall_macro	0.956402
SVC	kernel	sigmoid	recall_macro	0.056270
SVC	kernel	linear	F1	0.964347
SVC	kernel	poly	F1	0.964347
SVC	kernel	rbf	F1	0.957527
SVC	kernel	sigmoid	F1	0.034511
SVC	iteration	100	accuracy	0.046667
SVC	iteration	1000	accuracy	0.046667
SVC	iteration	10000	accuracy	0.046667
SVC	iteration	100	precision_macro	0.027103
SVC	iteration	1000	precision_macro	0.027103
SVC	iteration	10000	precision_macro	0.027103
SVC	iteration	100	recall_macro	0.056270
SVC	iteration	1000	recall_macro	0.056270
SVC	iteration	10000	recall_macro	0.056270
SVC	iteration	100	F1	0.034511
SVC	iteration	1000	F1	0.034511
SVC	iteration	10000	F1	0.034511

Model	Parameter	Value	Metric	Score
DecisionTreeClassifier	depth	1	accuracy	0.953333
DecisionTreeClassifier	depth	2	accuracy	0.953333
DecisionTreeClassifier	depth	3	accuracy	0.953333
DecisionTreeClassifier	depth	4	accuracy	0.953333
DecisionTreeClassifier	depth	5	accuracy	0.953333
DecisionTreeClassifier	depth	1	precision_macro	0.958127
DecisionTreeClassifier	depth	2	precision_macro	0.958127
DecisionTreeClassifier	depth	3	precision_macro	0.958127
DecisionTreeClassifier	depth	4	precision_macro	0.958127
DecisionTreeClassifier	depth	5	precision_macro	0.958127
DecisionTreeClassifier	depth	1	recall_macro	0.953810
DecisionTreeClassifier	depth	2	recall_macro	0.953810
DecisionTreeClassifier	depth	3	recall_macro	0.953810
DecisionTreeClassifier	depth	4	recall_macro	0.953810
DecisionTreeClassifier	depth	5	recall_macro	0.953810
DecisionTreeClassifier	depth	1	F1	0.953044
DecisionTreeClassifier	depth	2	F1	0.953044
DecisionTreeClassifier	depth	3	F1	0.953044
DecisionTreeClassifier	depth	4	F1	0.953044
DecisionTreeClassifier	depth	5	F1	0.953044
DecisionTreeClassifier	criterion	gini	accuracy	0.953333
DecisionTreeClassifier	criterion	entropy	accuracy	0.953333
DecisionTreeClassifier	criterion	log_loss	accuracy	0.953333
DecisionTreeClassifier	criterion	gini	precision_macro	0.958127
DecisionTreeClassifier	criterion	entropy	precision_macro	0.958127
DecisionTreeClassifier	criterion	log_loss	precision_macro	0.958127
DecisionTreeClassifier	criterion	gini	recall_macro	0.953810
DecisionTreeClassifier	criterion	entropy	recall_macro	0.953810
DecisionTreeClassifier	criterion	log_loss	recall_macro	0.953810
DecisionTreeClassifier	criterion	gini	F1	0.953044
DecisionTreeClassifier	criterion	entropy	F1	0.953044
DecisionTreeClassifier	criterion	log_loss	F1	0.953044
DecisionTreeClassifier	splitter	best	accuracy	0.953333
DecisionTreeClassifier	splitter	random	accuracy	0.953333
DecisionTreeClassifier	splitter	best	precision_macro	0.958127
DecisionTreeClassifier	splitter	random	precision_macro	0.958860
DecisionTreeClassifier	splitter	best	recall_macro	0.953810
DecisionTreeClassifier	splitter	random	recall_macro	0.952193
DecisionTreeClassifier	splitter	best	F1	0.953044
DecisionTreeClassifier	splitter	random	F1	0.951866

## 4.2 Wykres przedstawiający wartości zależności parametrów o charakterze ciągłym i jakości klasyfikacji.





## 5 Wnioski

Przy badanych parametrach iteracje nie wpłynęły znacząco na jakość klasyfikacji. Technika podziału jądra sigmoidalnego w SVM okazała się najmniej skuteczna. Reszta wyników była zbliżona, co sugeruje, że parametry mogły być zbyt mało zróżnicowane, aby wykazać różnice w jakości klasyfikacji.