

Dokumentacja Projektu z przedmiotu Metody
Inteligencji Obliczeniowej
*Analiza SHAP istotności poszczególnych elementów
wektora wejściowego SSN*

20 czerwca 2025

Spis treści

1	Wstęp	2
2	Wykorzystane zbiory danych	2
3	Zbiór danych dotyczący sukcesu akademickiego uczniów	2
3.1	Preprocessing	2
3.2	Sieć neuronowa	2
3.3	Analiza SHAP	2
3.4	Ponowny trening po redukcji cech	2
3.5	Otrzymane wyniki	3
3.6	Wnioski	4
3.7	Podsumowanie	4
4	Zbiór danych dotyczący przewidywania upadłości polskich firm	4
4.1	Preprocessing	4
4.2	Sieć	4
4.3	Analiza SHAP	4
4.4	Ponowny trening po redukcji cech	4
4.5	Wyniki	4
4.6	Wnioski	4
4.7	Podsumowanie	4

1 Wstęp

Projekt obejmował tematykę analizy SHAP istotności poszczególnych elementów wektora wejściowego SSN. Postanowiliśmy wykonać cztery modele korzystające z czterech zbiorów danych i na nich przeprowadzić analizę.

2 Wykorzystane zbiory danych

Pierwszy zbiór dotyczył problemu przewidywania sukcesu akademickiego uczniów, natomiast drugi - przewidywania upadłości polskich firm.

3 Zbiór danych dotyczący sukcesu akademickiego uczniów

3.1 Preprocessing

W tym zbiorze należało przekształcić na początku etykiety na liczby, aby działał poprawnie w sieci neuronowej. Następnie podzieliliśmy zbiór na cechy oraz na zbiór uczący i testujący w proporcjach 80/20. Pamiętaliśmy o stratyfikacji, żeby dane były równomiernie podzielone oraz przeskalowaliśmy zbiór za pomocą StandardScaler.

3.2 Sieć neuronowa

Zbudowaliśmy model MLP; pierwsza warstwa ma 128 neuronów oraz funkcję aktywacji "relu". Warstwa ukryta posiada 64 neurony, a warstwa wyjściowa ma tyle neuronów, ile mamy klas, czyli 3. Została użyta funkcja aktywacji softmax. Zastosowaliśmy również dropout, żeby nie dopuścić do przeuczenia.

Wykonaliśmy kompilację modelu, a później trening, w którym liczba epok wynosiła 50. Następnie zrobiliśmy wykresy sprawdzające nasz model - zrobiliśmy historię treningu, sprawdziliśmy accuracy i macierz pomyłek.

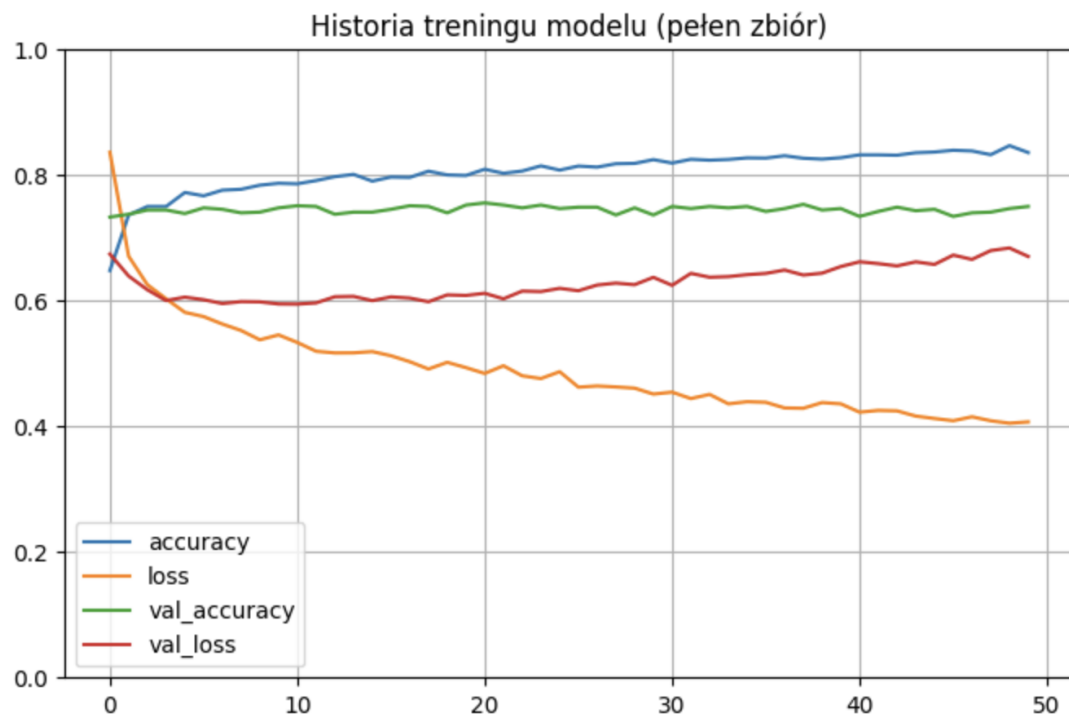
3.3 Analiza SHAP

Shap umożliwił nam zobaczenie, jak model wygląda w środku. Za pomocą KernelExplainer można było zobaczyć, co wpływa na predykcję konkretnych klas.

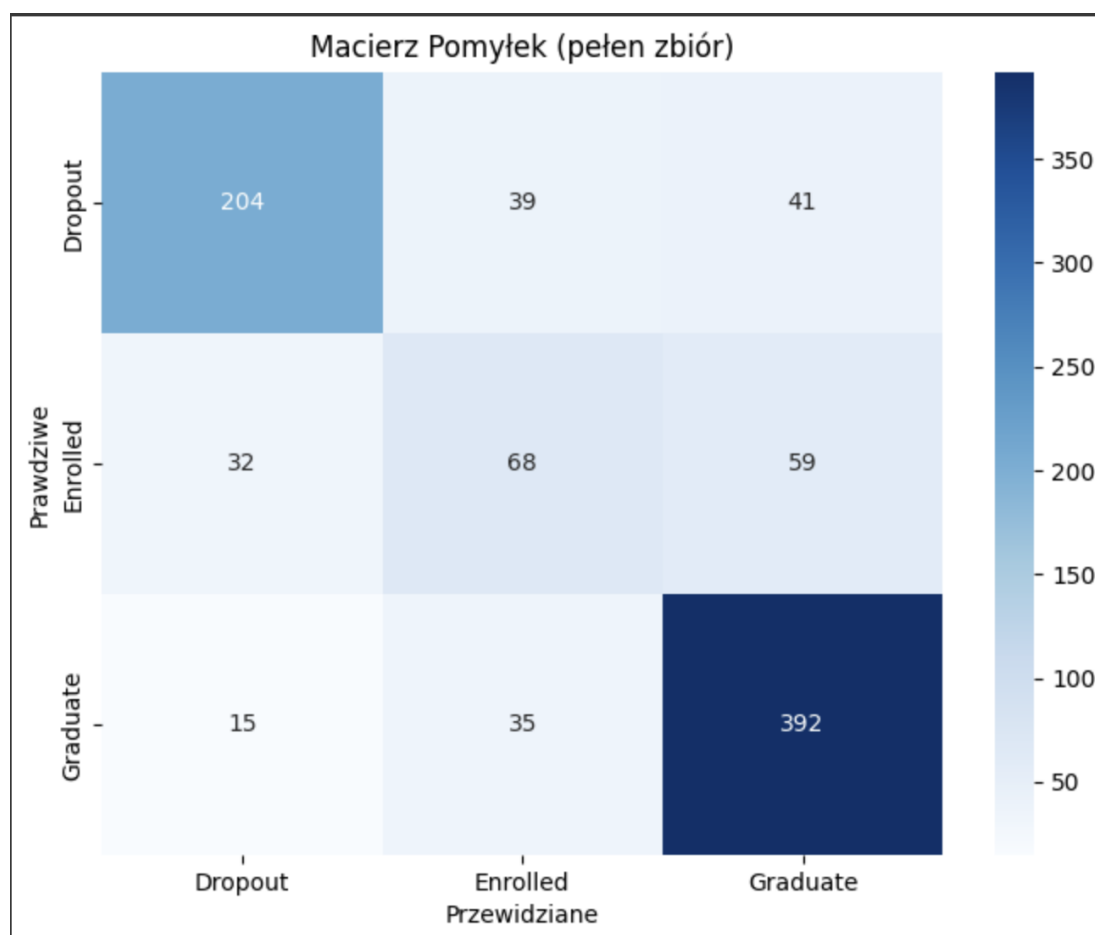
3.4 Ponowny trening po redukcji cech

Wybraliśmy 5 najmniej istotnych cech i usunęliśmy je. Zrobiliśmy ponowne skalowanie, ponieważ rozkład mógł ulec zmianie. Następnie wytrenowaliśmy model na zredukowanym zbiorze i porównaliśmy wyniki.

3.5 Otrzymane wyniki



	precision	recall	f1-score
Dropout	0.81	0.72	0.76
Enrolled	0.48	0.43	0.45
Graduate	0.80	0.89	0.84
accuracy			0.75
macro avg	0.70	0.68	0.68
weighted avg	0.74	0.75	0.75



3.6 Wnioski

3.7 Podsumowanie

4 Zbiór danych dotyczący przewidywania upadłości polskich firm

4.1 Preprocessing

4.2 Sieć

4.3 Analiza SHAP

4.4 Ponowny trening po redukcji cech

4.5 Wyniki

4.6 Wnioski

4.7 Podsumowanie