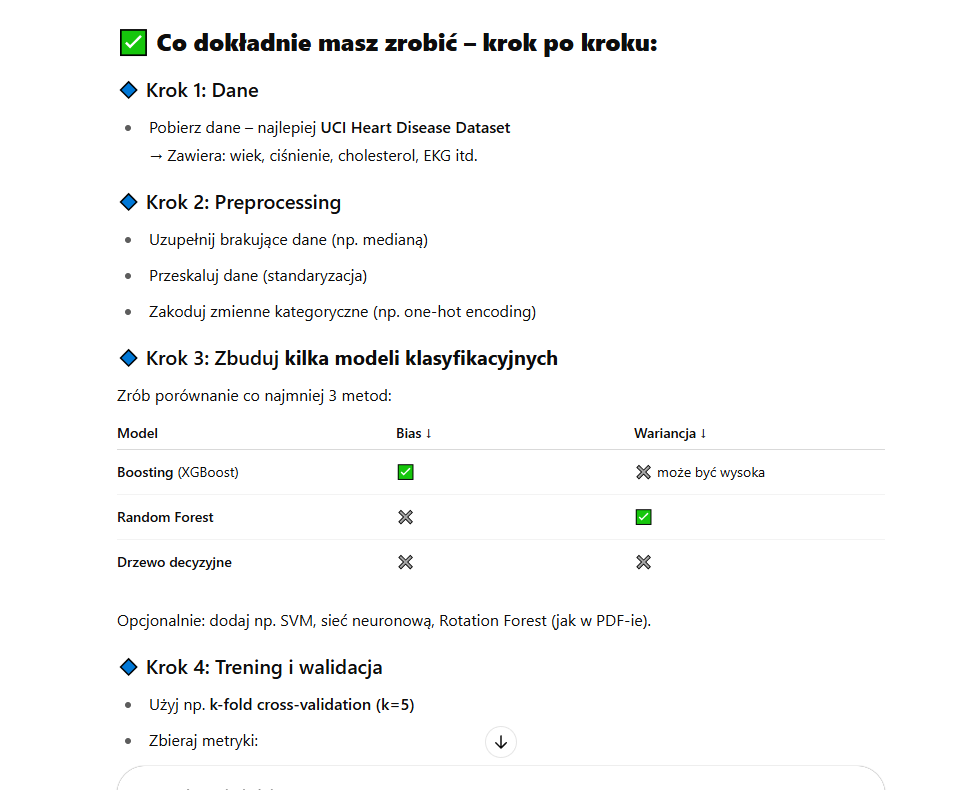
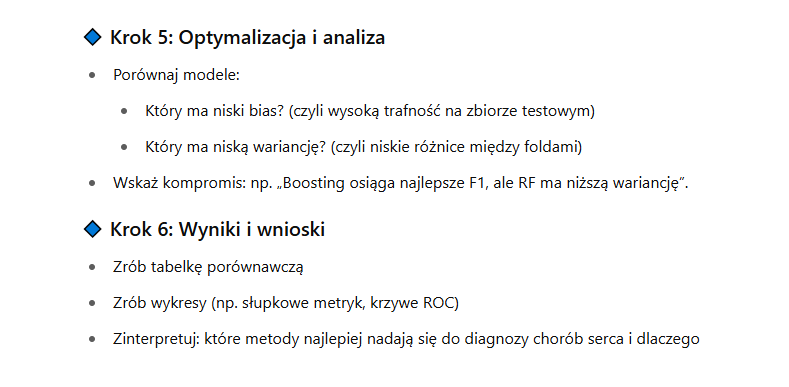
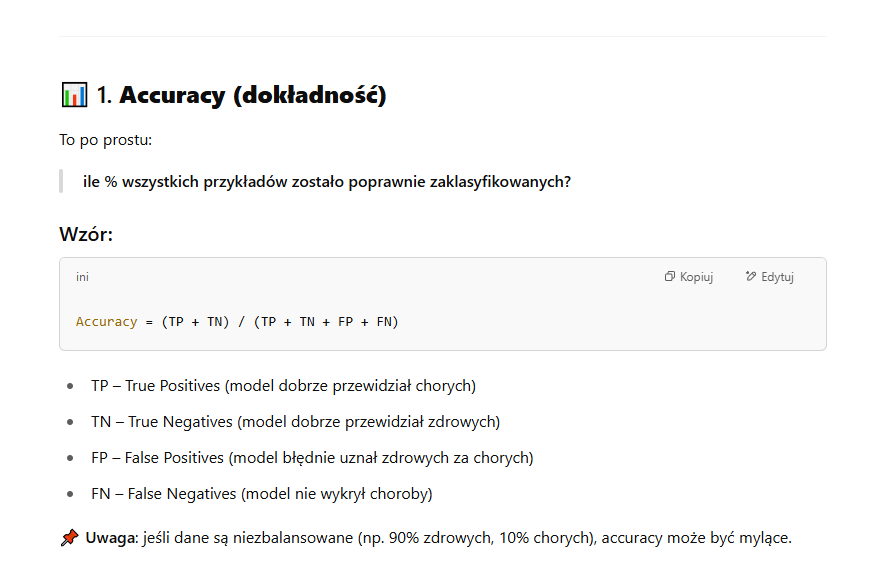
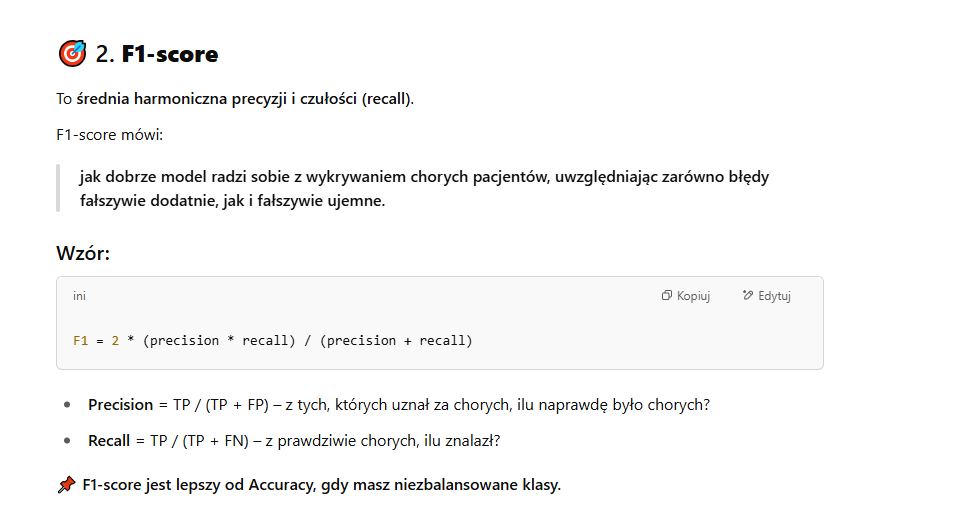
Ja robie dla klasyfikacji, można podobne dodac do klasteryzacji

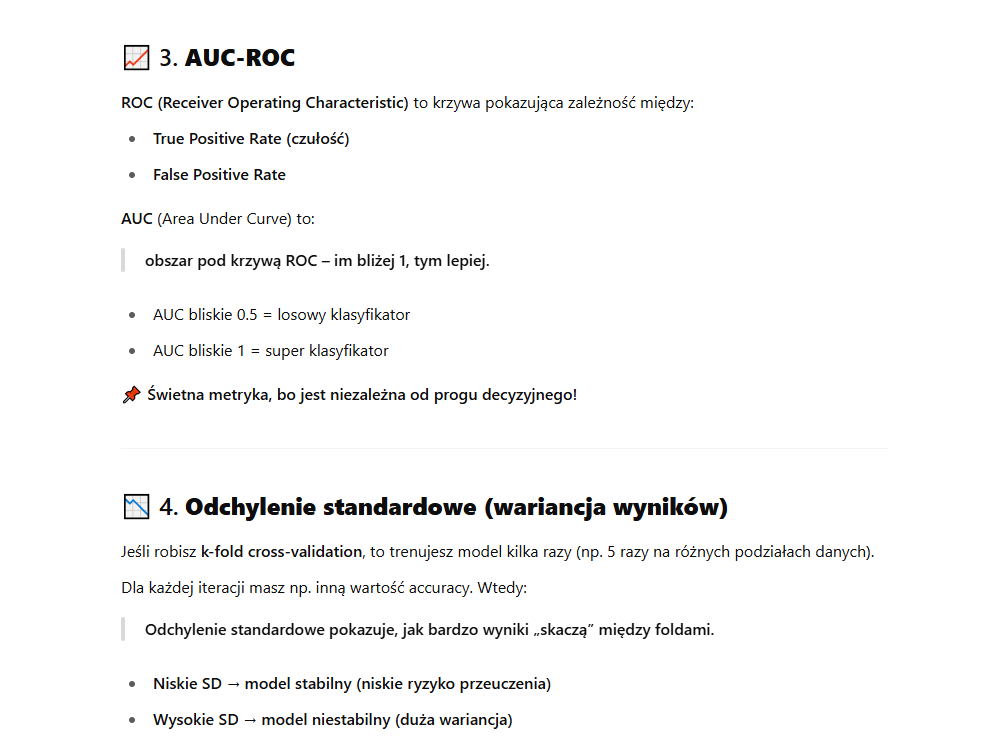
Klasyfikacja





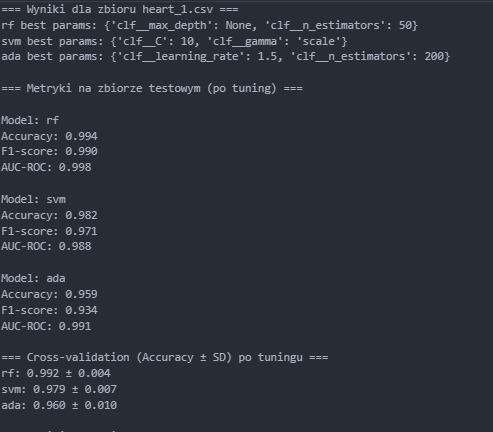






Do tego dodałem jeszcze Naive Bayes (nb)

Wyniki heart\_1 (gdzie jest 5000 rekordow):



**Wnioski:**

**Random Forest (RF)**

* Wyniki: Accuracy 0.994, F1 0.990, AUC 0.998
* Bias/Wariancja: Niski bias (bardzo dobrze dopasowuje się do danych), niska wariancja (mało różnic między CV wynikami).
* Wniosek: Bardzo silny model, dobrze radzi sobie z danymi, minimalne przetrenowanie (overfitting).

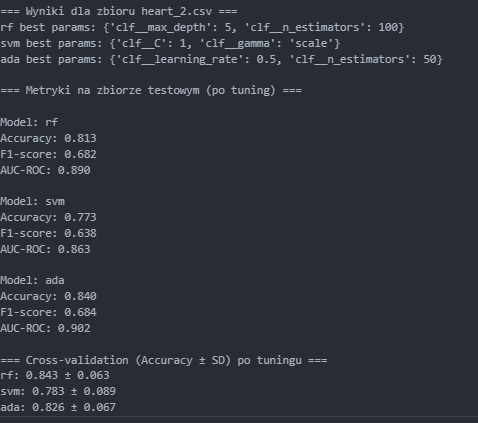
**SVM**

* **Wyniki:** Accuracy 0.982, F1 0.971, AUC 0.988
* **Bias/Wariancja:** Średni bias, niska wariancja (wyniki CV są stabilne).
* **Wniosek:** Model dobrze dopasowany, lekko bardziej restrykcyjny niż RF, ale nadal bardzo skuteczny.

**AdaBoost**

* **Wyniki:** Accuracy 0.959, F1 0.934, AUC 0.991
* **Bias/Wariancja:** Większy bias niż RF i SVM, umiarkowana wariancja (nieco większa rozpiętość wyników CV).
* **Wniosek:** Mniej elastyczny model, który jednak dobrze generalizuje, nieco bardziej konserwatywny.

Wyniki heart\_2 (gdzie jest tylko 300 rekordow):



**Wnioski:**

**Random Forest**

* **Wyniki:** Accuracy 0.813, F1 0.682, AUC 0.890
* **Bias/Wariancja:** Większy bias niż na heart\_1 (max\_depth=5 ogranicza złożoność), umiarkowana wariancja (większa od heart\_1).
* **Wniosek:** Model bardziej uproszczony, może niedopasowywać się do trudniejszych danych (wyższy bias), ale stabilniejszy.

**SVM**

* **Wyniki:** Accuracy 0.773, F1 0.638, AUC 0.863
* **Bias/Wariancja:** Wyższy bias (C=1, umiarkowana regularyzacja), stosunkowo duża wariancja (niestabilne wyniki CV).
* **Wniosek:** Model mniej skuteczny, prawdopodobnie zbyt prosty dla tego zbioru lub dane są bardziej złożone.

**AdaBoost**

* **Wyniki:** Accuracy 0.840, F1 0.684, AUC 0.902
* **Bias/Wariancja:** Najmniejszy bias spośród trzech, ale większa wariancja (rozpiętość wyników CV).
* **Wniosek:** AdaBoost radzi sobie najlepiej na trudniejszym zbiorze, bo adaptacyjnie poprawia słabe klasyfikatory, ale jest bardziej podatny na zmienność.