

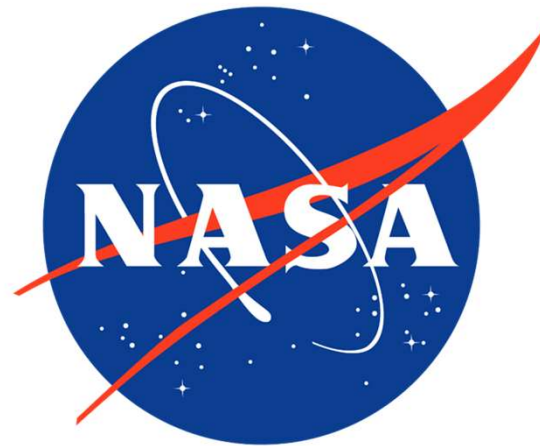


Armageddon!

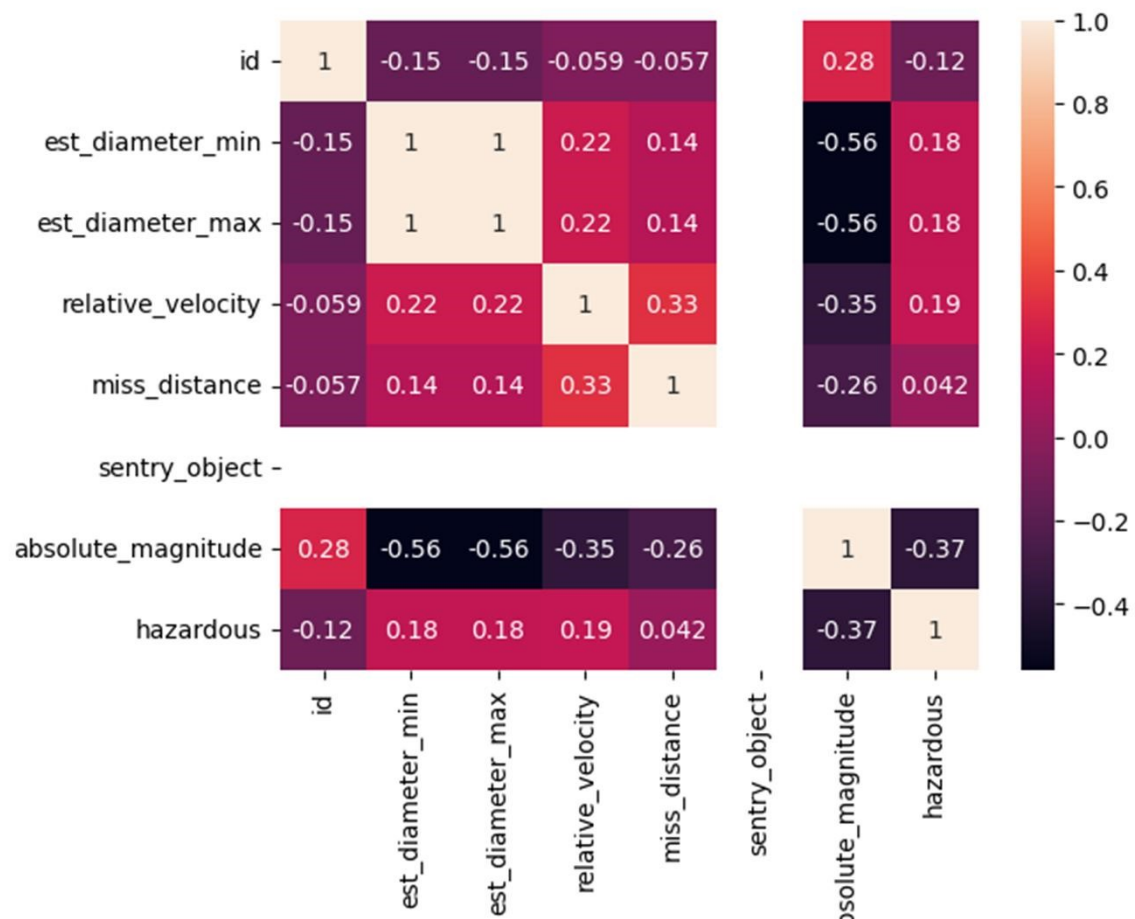
Wykrywanie asteroid potencjalnie
zagrożających Ziemi

Temat projektu

- dane NASA o asteroidach zaobserwowanych między 2001 a 2022 rokiem
- Cel ML: wykrywanie asteroid potencjalnie zagrażających Ziemi
- zastosowanie biznesowe: ratowanie świata

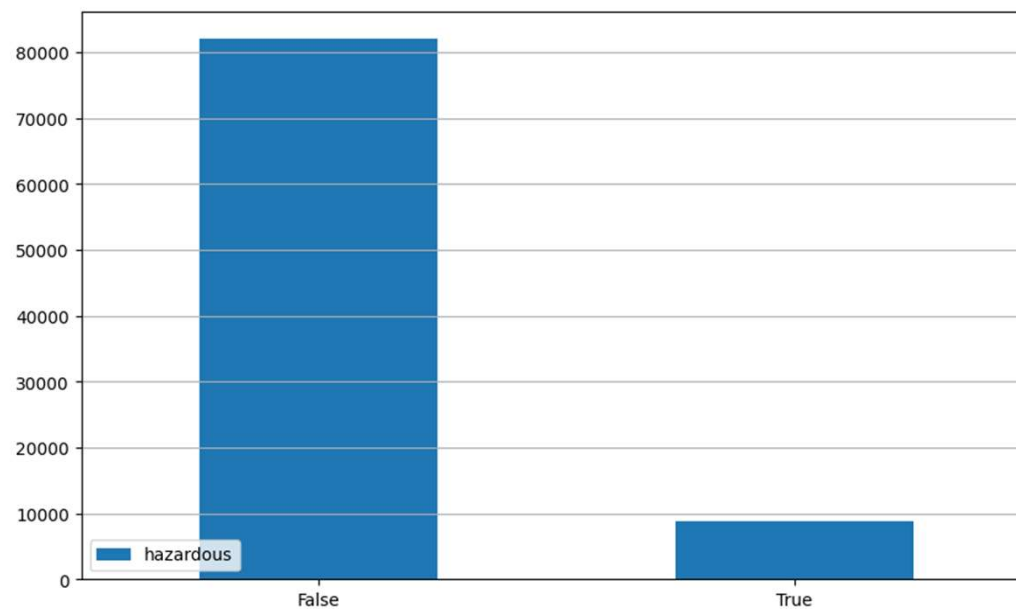


Dane



Przygotowanie danych do modelowania:

- standaryzacja
- równoważenie klas

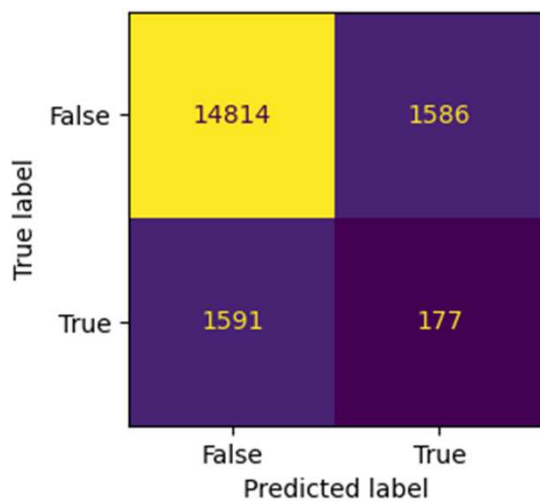


Modelowanie

- klasyfikacja
- modele:
 - dummy classifier,
 - regresja logistyczna,
 - drzewo decyzyjne,
 - las losowy,
 - xgboost
- metryki:
 - accuracy,
 - precision,
 - recall,
 - f1 score
- confusion matrix

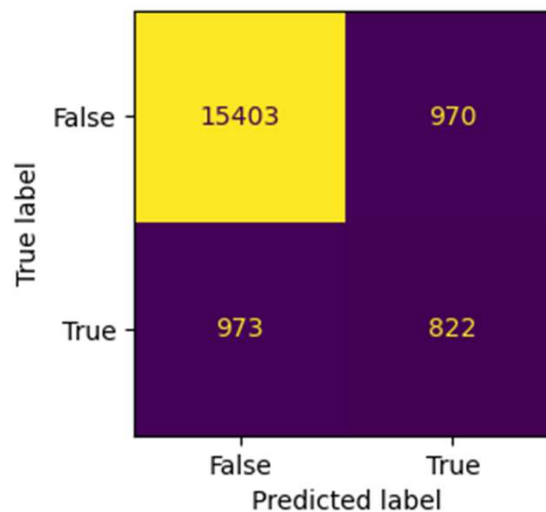
Dummy Classifier

Accuracy: 0.83
Precision: 0.10
Recall: 0.10
F1 score: 0.10



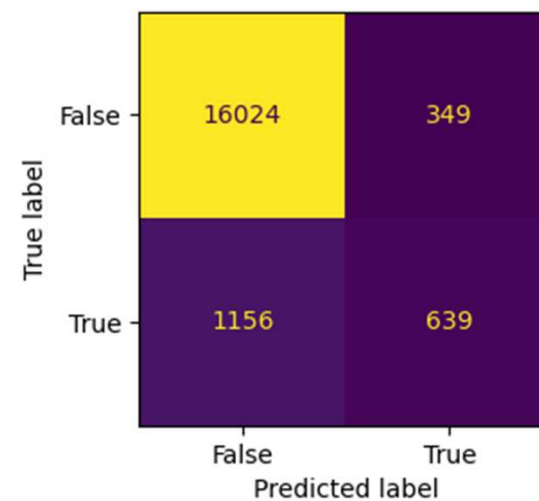
Drzewo decyzyjne

Accuracy: 0.89
Precision: 0.46
Recall: 0.46
F1 score: 0.46



Las losowy

Accuracy: 0.92
Precision: 0.65
Recall: 0.36
F1 score: 0.46



Regresja logistyczna

Accuracy: 0.79
Precision: 0.31
Recall: 0.93
F1 score: 0.47

True label	False	True
	False 12678	True 3695
Predicted label	False 122	True 1673

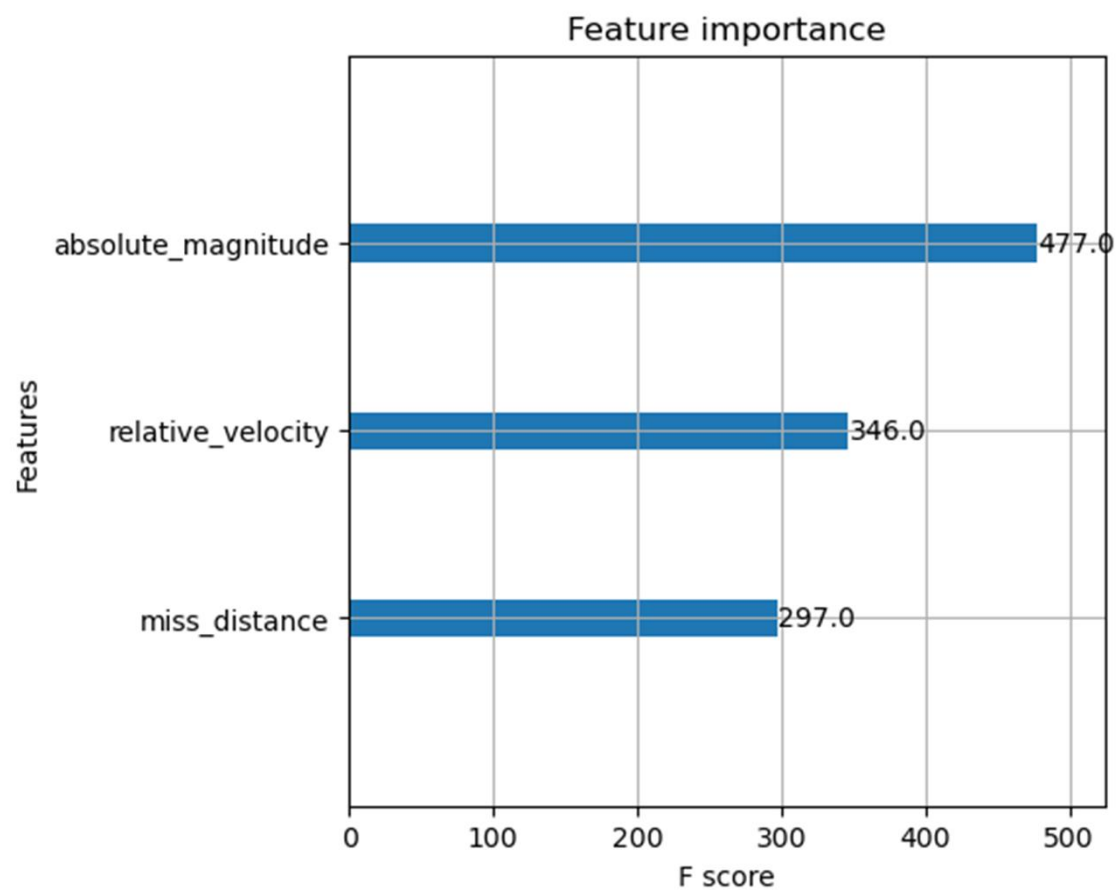
XGBoost

Accuracy: 0.92
Precision: 0.71
Recall: 0.24
F1 score: 0.36

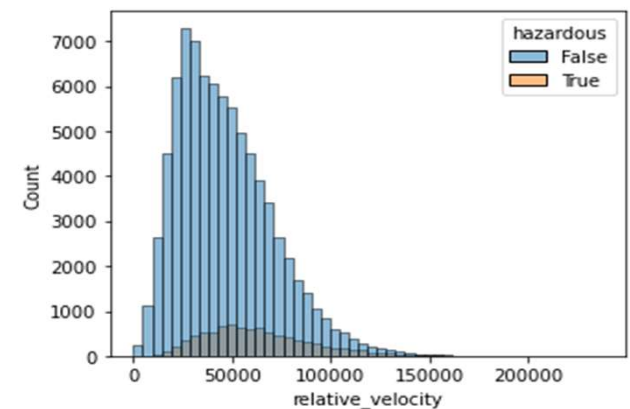
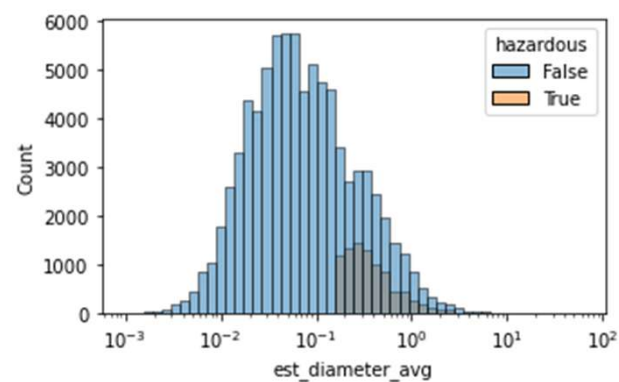
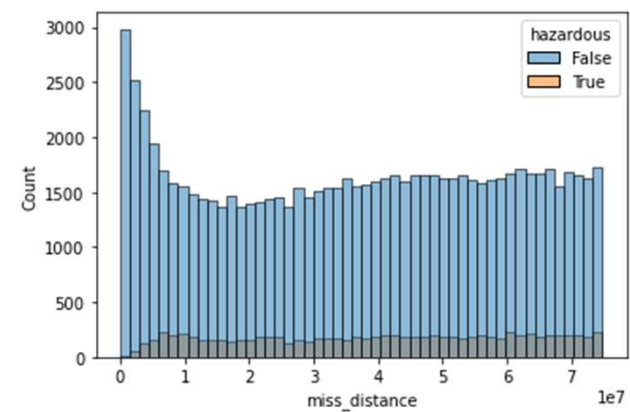
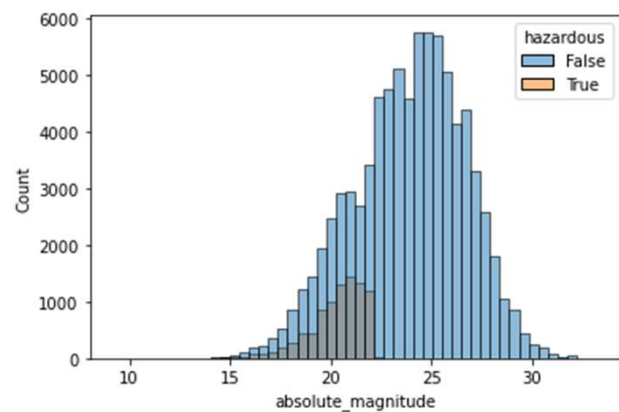
True label	False	True
	False 15752	True 648
Predicted label	False 1701	True 67

Praca z modelem

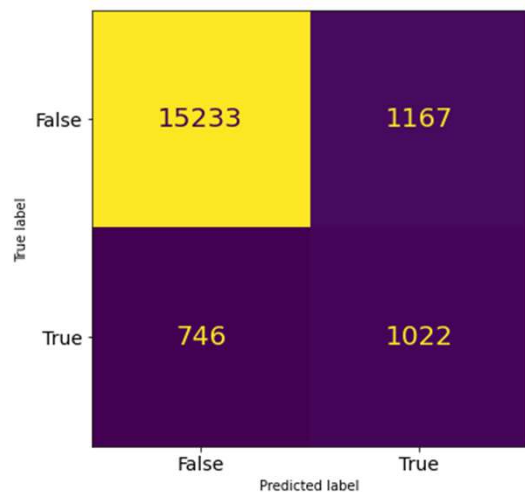
- dobór parametrów za pomocą hyperopt



Rozkłady danych



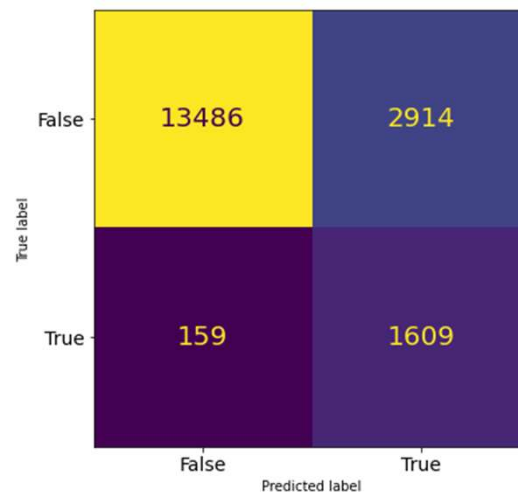
Rekomendacije doboru modelu XGBoost



Model 1:

$P \geq 0.3$

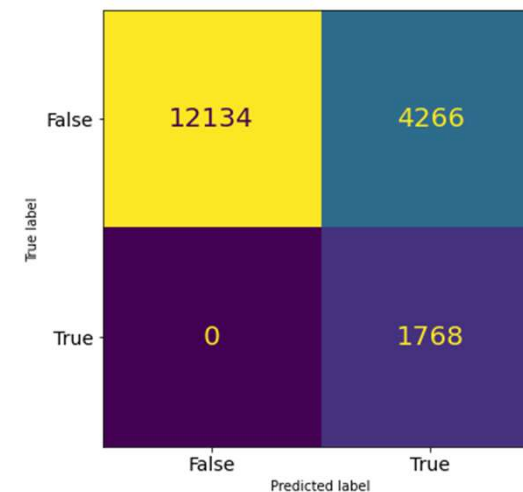
Accuracy: 0.90
Precision: 0.47
Recall: 0.58
F1 score: 0.52



Model 2:

$P \geq 0.2$

Accuracy: 0.83
Precision: 0.36
Recall: 0.91
F1 score: 0.51



Model 3:

$P \geq 0.01$

Accuracy: 0.77
Precision: 0.29
Recall: 1.0
F1 score: 0.45

Wnioski

- największy wpływ na wynik ma zmienna absolute magnitude (czyli prawdopodobnie m.in. średnica asteroidy),
- stosunkowo niski wpływ odległości asteroidy od Ziemi, prawdopodobnie związany z gorszym wykrywaniem małych asteroid znajdujących się bardzo daleko,
- XGBoost dał początkowo słabe wyniki, ale najlepsze możliwości ich poprawy,
- ostateczny wybór modelu zależy od tego, jak ważne jest wykrycie każdej niebezpiecznej asteroidy i jak bardzo ograniczają nas koszty