

A decorative graphic consisting of blue circuit-like lines with small circles at the ends, extending horizontally from the left and right sides of the central text box.

Fake News Detection

FAKE NEWS

- Fake newsy – według www.gov.pl są to wpisy, wiadomości, całe kanały informacyjne, w których przekazywane dane okazują się nieprawdziwe lub przeinaczone. Celowo wprowadzają w błąd, szokują i budzą kontrowersje.

PRZYGOTOWANIE ZESTAWU DANYCH

Zbiór danych zawiera ponad 40 000 artykułów, każdy z nich ma przypisaną etykietę – „fake” lub „true”



Kroki przygotowania danych zawierały między innymi:

Konkatenacja tytułu i „body”

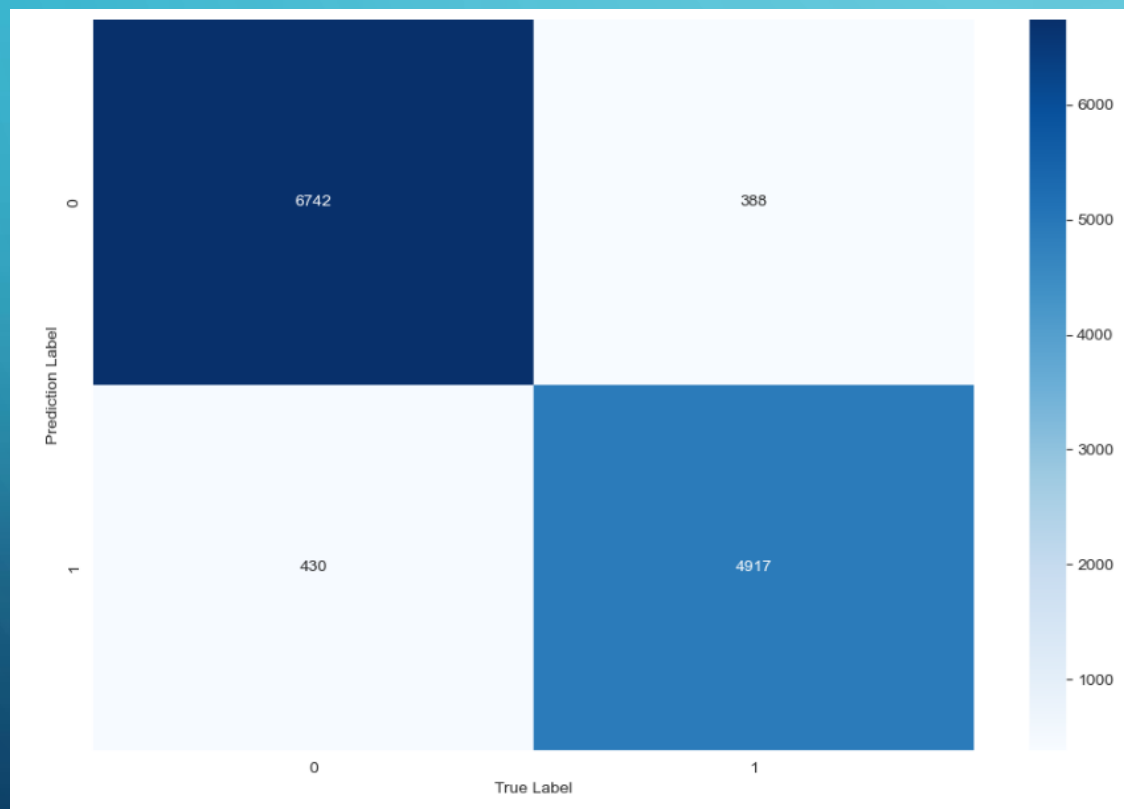
Usunięcie tzw. „stopwords”

Wyeliminowanie innych języków niż język angielski

Usunięcie linków, usunięcie znaków interpunkcyjnych

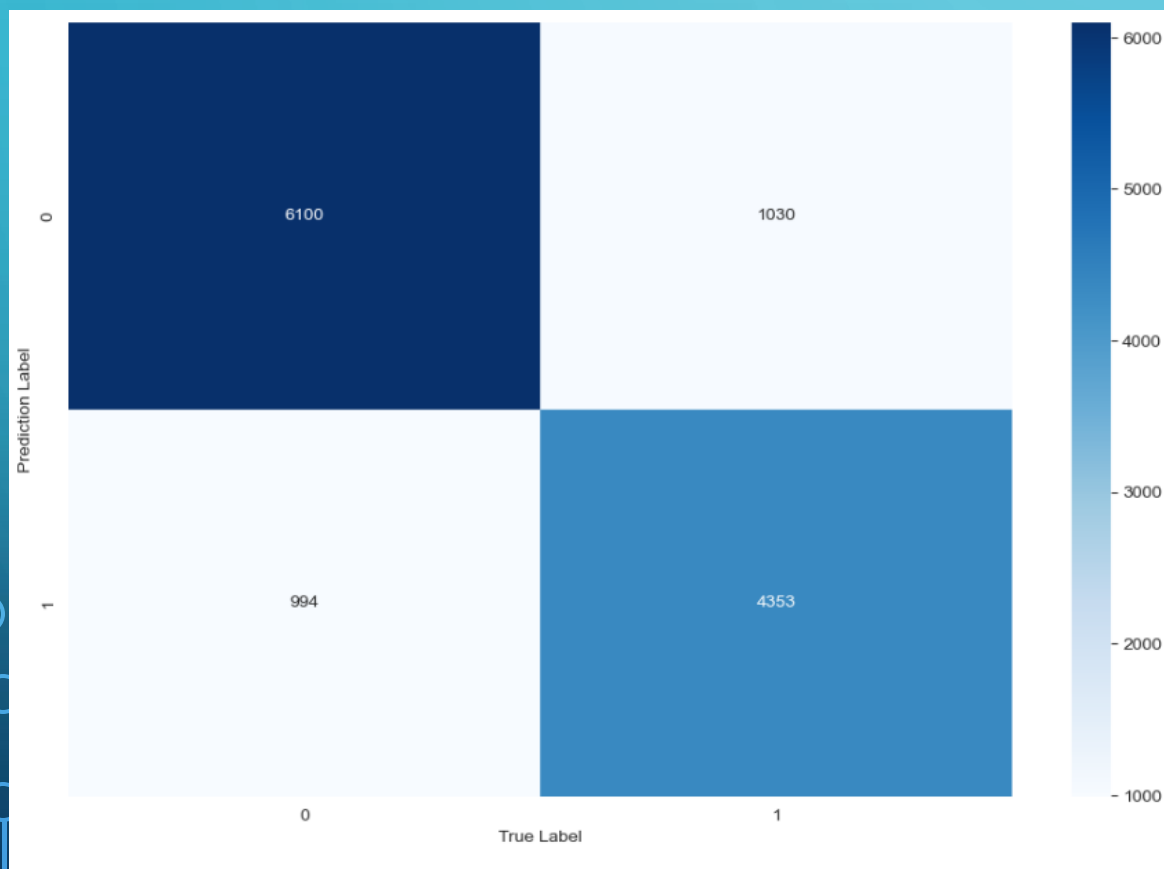
Zastosowanie algorytmu **TF-IDF**

WERYFIKACJA MODELI: DECISION TREE



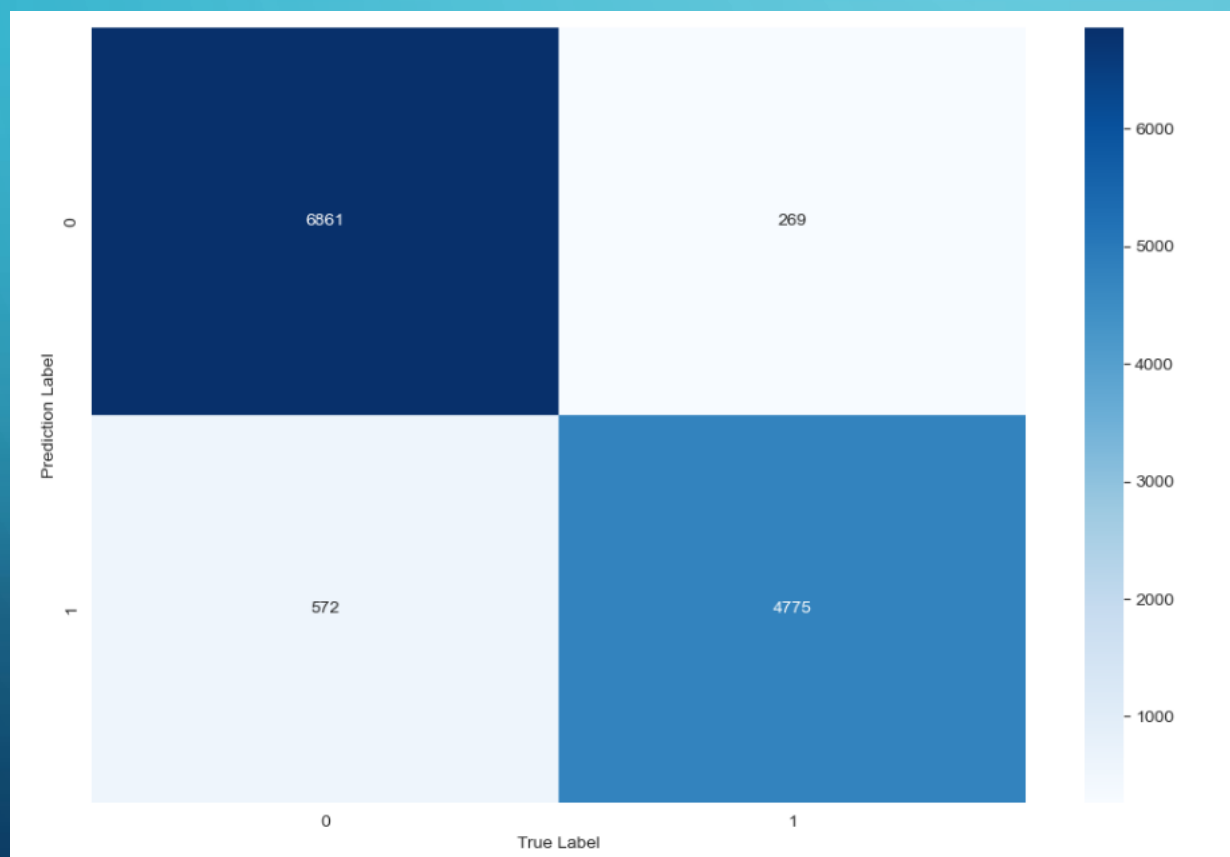
	precision	recall	f1-score	support
0	0.95	0.94	0.94	7172
1	0.92	0.93	0.92	5305
accuracy			0.93	12477
macro avg	0.93	0.93	0.93	12477
weighted avg	0.93	0.93	0.93	12477

WERYFIKACJA MODELI: DECISION TREE PO PCA



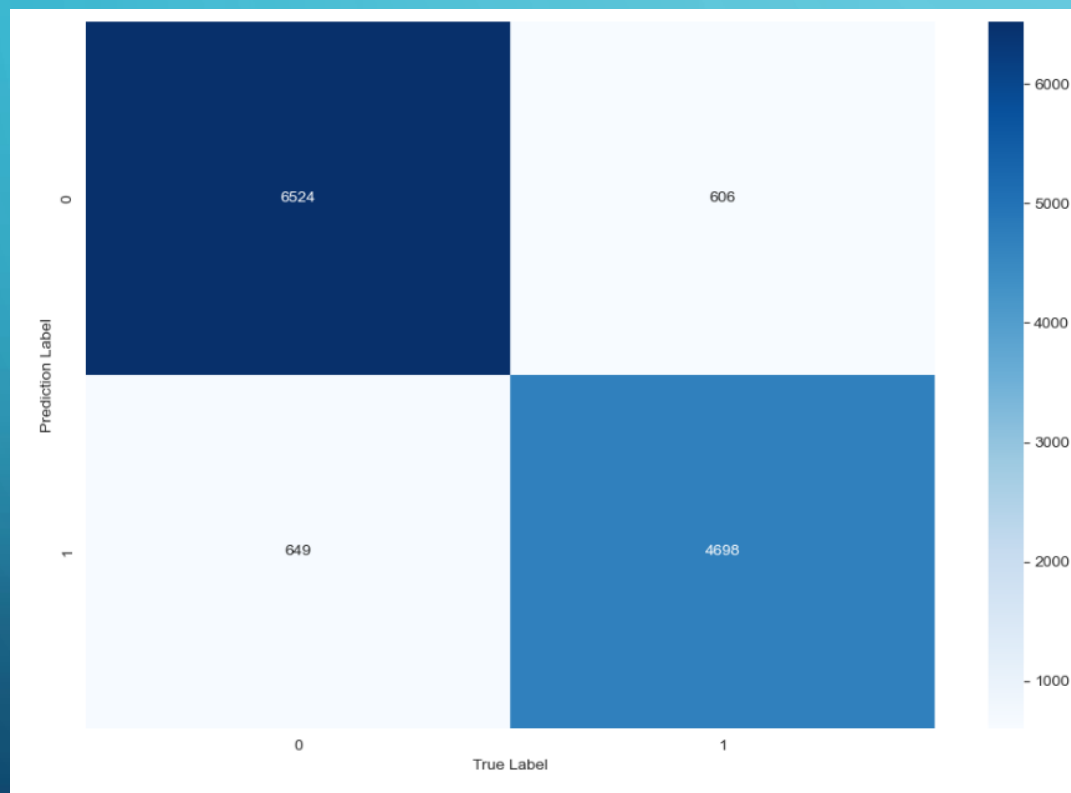
	precision	recall	f1-score	support
0	0.86	0.86	0.86	7094
1	0.81	0.81	0.81	5383
accuracy			0.84	12477
macro avg	0.83	0.83	0.83	12477
weighted avg	0.84	0.84	0.84	12477

WERYFIKACJA MODELI: RANDOM FOREST



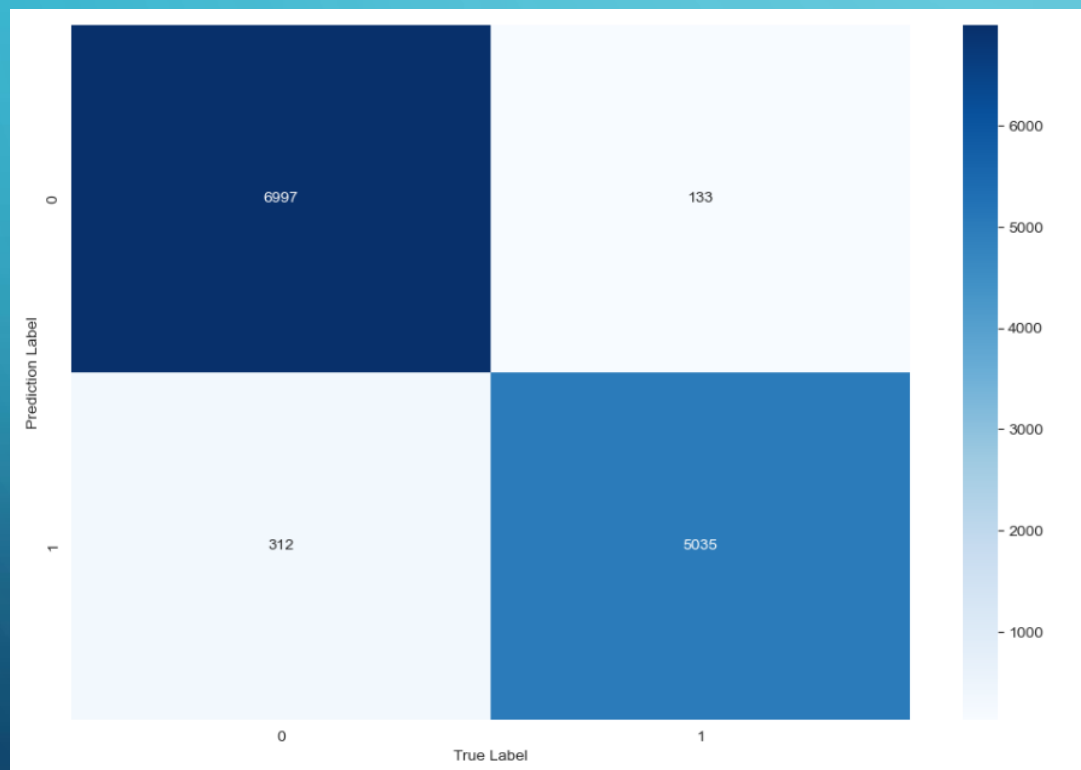
	precision	recall	f1-score	support
0	0.96	0.92	0.94	7433
1	0.89	0.95	0.92	5044
accuracy			0.93	12477
macro avg	0.93	0.93	0.93	12477
weighted avg	0.93	0.93	0.93	12477

WERYFIKACJA MODELI: RANDOM FOREST PO PCA



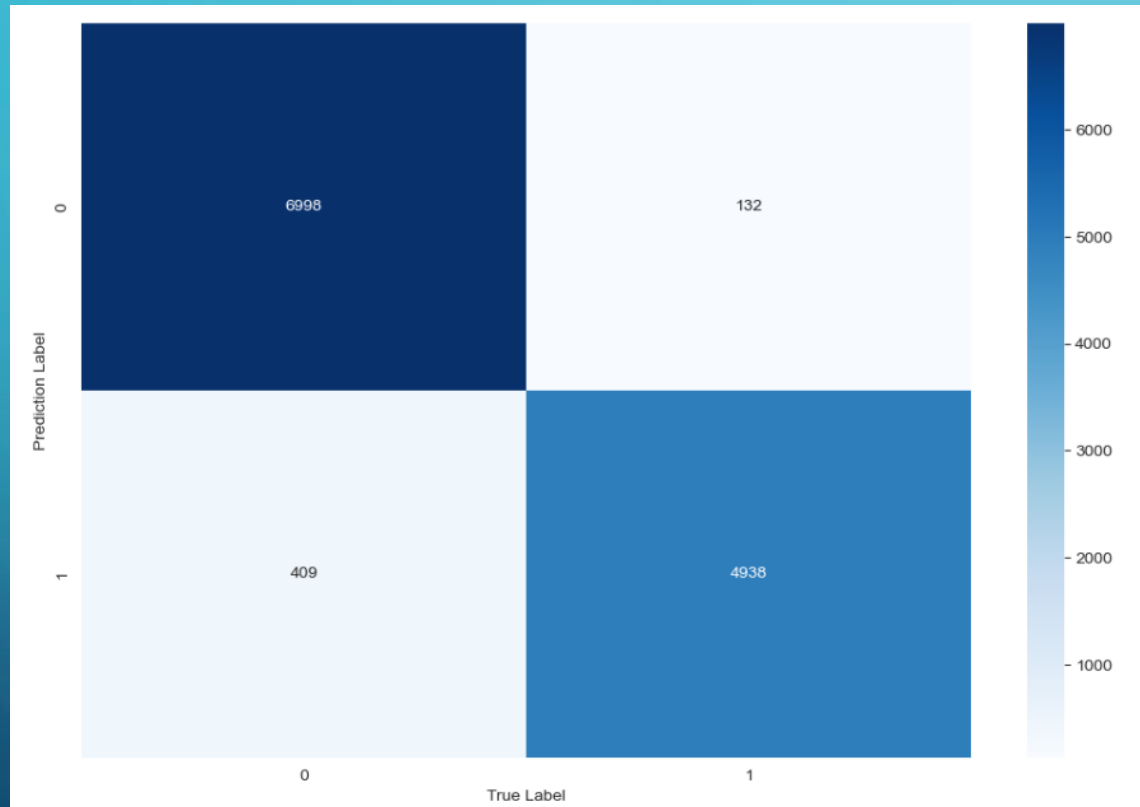
	precision	recall	f1-score	support
0	0.92	0.91	0.91	7173
1	0.88	0.89	0.88	5304
accuracy			0.90	12477
macro avg	0.90	0.90	0.90	12477
weighted avg	0.90	0.90	0.90	12477

WERYFIKACJA MODELI: XGBOOST



	precision	recall	f1-score	support
0	0.98	0.96	0.97	7309
1	0.94	0.97	0.96	5168
accuracy			0.96	12477
macro avg	0.96	0.97	0.96	12477
weighted avg	0.96	0.96	0.96	12477

OSTATECZNY MODEL = XGBOOST



	precision	recall	f1-score	support
0	0.98	0.94	0.96	7407
1	0.92	0.97	0.95	5070
accuracy			0.96	12477
macro avg	0.95	0.96	0.96	12477
weighted avg	0.96	0.96	0.96	12477