

# Aplikacja identyfikująca język

Adam Machaj, Piotr Ksel

Aplikacja została zaimplementowana z wykorzystaniem języka python w wersji 3.7. Do projektu wykorzystaliśmy biblioteki pandas i sklearn do uczenia modelu. Aby zweryfikować czy nasz model dobrze sprawdza język wykorzystaliśmy Detect Language API.

Zbiór danych do uczenia modelu pozyskaliśmy ze strony kaggle: <https://www.kaggle.com/datasets/basilb2s/language-detection>. Posiada on 17 różnych języków.

	C1	C2
1	Text	Language
2	Nature, in the broadest sense, is the natural, physical, ...	English
3	"Nature" can refer to the phenomena of the physical world,...	English
4	The study of nature is a large, if not the only, part of s...	English
5	Although humans are part of nature, human activity is ofte...	English
6	[1] The word nature is borrowed from the Old French nature...	English
7	[2] In ancient philosophy, natura is mostly used as the La...	English
8	[3][4] The concept of nature as a whole, the physical uni...	English
9	During the advent of modern scientific method in the last ...	English
10	[5][6] With the Industrial revolution, nature increasingly...	English
11	However, a vitalist vision of nature, closer to the presoc...	English
12	[1] Within the various uses of the word today, "nature" of...	English
13	Nature can refer to the general realm of living plants and...	English
14	It is often taken to mean the "natural environment" or wil...	English
15	For example, manufactured objects and human interaction ge...	English
16	This more traditional concept of natural things that can s...	English
17	Depending on the particular context, the term "natural" mi...	English
18	[1] Earth is the only planet known to support life, and it...	English
19	Within the solar system, it is third closest to the sun; i...	English
20	Its most prominent climatic features are its two large pol...	English

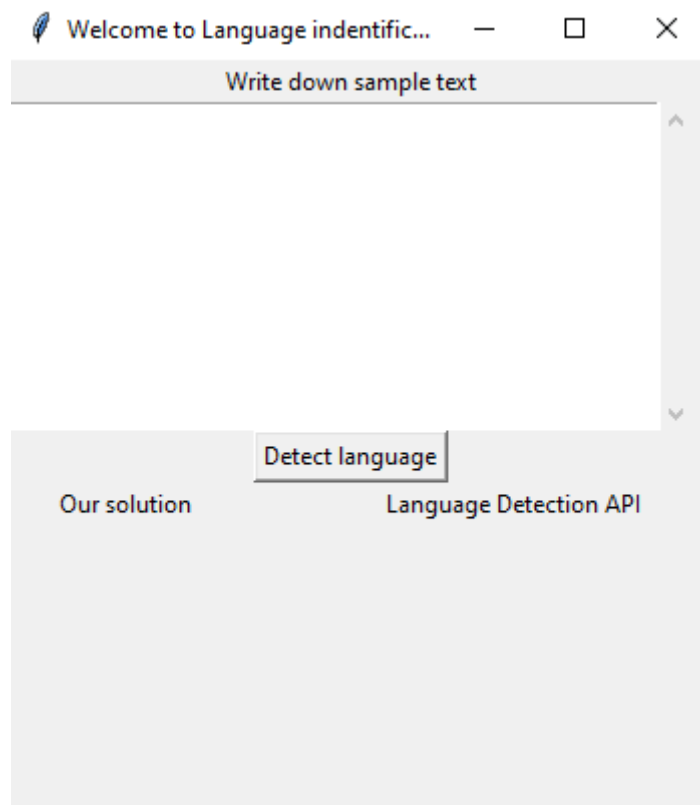
Detect Language API. jest darmowym API do rozpoznawania języków. Rozpoznaje 164 języki i jest bardziej dokładny od naszego modelu ze względu na bardziej dopracowany model. Dlatego wykorzystaliśmy je do zestawiania z naszymi wynikami.

Request zwraca dany model:

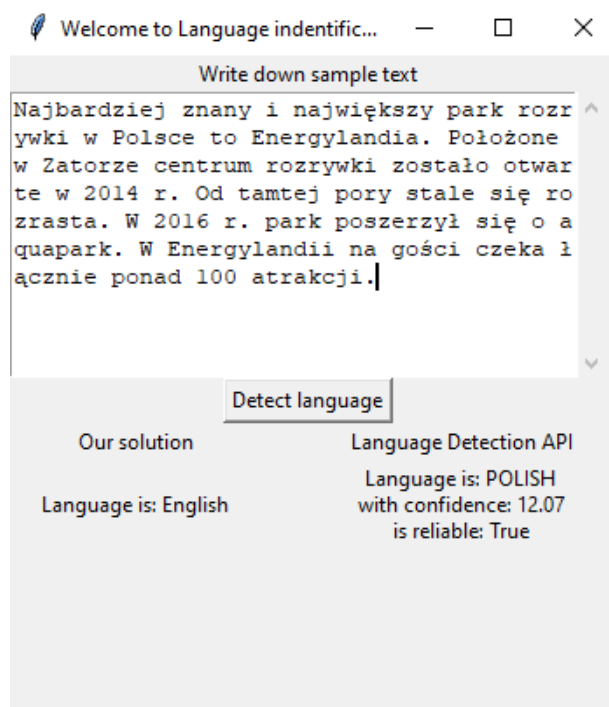
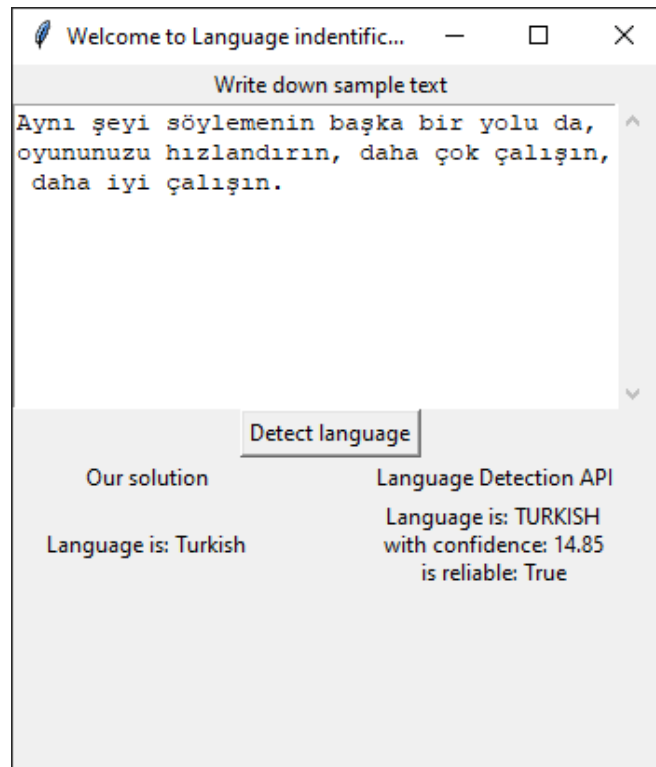
```
{"data":{"detections":[{"language":"es","isReliable":true,"confidence":10.24}]}}
```

Mówi on w jakim języku został napisany tekst oraz, czy identyfikacja jest wiarygodna.

GUI:



### Przykładowe wykorzystanie aplikacji:



Jak widać na powyższym screenie nasza aplikacja, przez brak próbek z niektórych języków np polskiego nie jest w stanie go zidentyfikować.