



GRUDZIEŃ 2021

KLASYFIKACJA GATUNKÓW PTAKÓW NA PODSTAWIE ICH ODGŁOSÓW

PODSTAWY UCZENIA MASZYNOWEGO WYKORZYSTUJĄC BIBLIOTEKĘ
OPENSIMILE I OPTUNA W PYTHONIE

EMILIA STEFANOWSKA
[GITHUB.COM/EMILIACODES/BIRD-CLASSIFICATION](https://github.com/emiliacodes/bird-classification)

Wstęp

W niniejszym projekcie podjęto próbę sklasyfikowania różnych odgłosów ptaków występujących na terenie Polski do odpowiedniego gatunku. Problem rozpoznawania zwierząt, a w szczególności ptaków na podstawie próbek dźwiękowych jest obecny na komercyjnej platformie międzynarodowej. Na rynku istnieje już kilka aplikacji pozwalających na właśnie taką klasyfikację gatunkową (*Song Sleuth: Auto Bird Song ID*, *BirdNET*), jednak jeszcze wiele z nich zawężonych jest do wąskiego rejonu geograficznego lub mają ograniczoną skuteczność i wydajność. Zastosowanie algorytmów uczenia maszynowego w aplikacjach, które umożliwiają identyfikację ptaków na podstawie ich odgłosów pozwala na zbliżenie się do natury osobom, które mogą cieszyć się dostępem do informacji bez dużego nakładu wysiłku, a tym samym ma duży potencjał sprzedażowy i handlowy.

Do wykonania projektu wykorzystano bazę danych umieszczoną na portalu *kaggle.com*, w której zostały zamieszczone pliki audio odgłosów pięćdziesięciu gatunków ptaków wyselekcjonowane ze strony internetowej *xeno-canto.org* (link do bazy: <https://www.kaggle.com/monogenea/birdsongs-from-europe>). Baza danych zawiera 2150 plików w formacie mp3 (po 43 sygnały na gatunek) zawierających śpiewy i inne charakterystyczne odgłosy wydawane przez poszczególne ptaki. Długości plików wahają się między 30 sekundami, a ponad 33 minutami. Spośród tych sygnałów wybrano te pochodzące od pięciu gatunków ptaków, które były przedmiotem klasyfikacji w projekcie. Były to:

1. *Alauda arvensis* – skowronek zwyczajny;
2. *Crex crex* – derkacz;
3. *Bubo bubo* – puchacz zwyczajny;
4. *Glaucidium passerinum* – sóweczka zwyczajna;
5. *Loxia curvirostra* – krzyżodziób świerkowy;

Stworzona na nowo baza danych posiada zatem 215 obiektów, po 43 w każdej klasie. Długości plików były bardzo zróżnicowane (najkrótszy plik miał 30 sekund, najdłuższy – 9 minut 39 sekund), zdecydowano się zatem na wykorzystanie biblioteki *openSMILE*, która pozwoliła na ekstrakcję 88 parametrów z każdego pliku. Warto zaznaczyć, że biblioteka ta dla plików w formacie wave pozwala na uzyskanie aż 6373 cech sygnału. Mimo tego, założono że ekstrakcja 88 cech jest lepszym przygotowaniem danych ze względu na to, że praca na np. spektrogramach sygnałów wykreowałaby dużą stratę informacji wywołaną znaczną różnorodnością długości plików. Wykonano klasyfikację na wyekstrahowanych cechach z plików o różnych częstotliwościowych próbkowania, a następnie wykonano resampling wszystkich sygnałów do częstotliwości 44100 Hz, wyekstrahowano cechy ponownie i wykonano drugą klasyfikację.



Rysunek 1. Krzyżodziób świerkowy – samce.¹

¹ Zdjęcie autorstwa Elaine R. Wilson, <http://www.naturespicsonline.com/> CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=969811>

Przygotowanie i weryfikacja danych, preprocessing

W celu przygotowania danych przeprowadzono konieczne czynności. Wykonano opisaną wyżej ekstrakcję cech za pomocą biblioteki openSMILE zarówno na plikach o różnych częstotliwościach próbkowania jak i na sygnałach o częstotliwości znormalizowanej do 44100 Hz. Przygotowano etykiety dla otrzymanych obiektów, zgodnie z przynależnością plików do poszczególnych gatunków ptaków. Następnie wykonano podział na zbiory testowy i treningowy w proporcji 20/80 zachowując stratyfikację danych.

Klasyfikacja

W celu klasyfikacji z optymalizacją hiperparametrów wykorzystano dwie metody klasyfikacji – maszynę wektorów nośnych SVC oraz algorytm lasu losowego. Jako metody ewaluacji wykorzystano dokładność, F1, czułość, precyzję, a także obliczono macierze pomyłek.

W metodzie SVC w celu porównania wyników jako funkcję celu wykorzystano metrykę F1, a następnie dokładność. Za pomocą biblioteki *Optuna* wykonano optymalizację parametru C, jądro metody oraz stopień wielomianu, jeżeli zostało wybrane jądro wielomianowe. W każdym przypadku wykonano 1000 prób optymalizacji. Otrzymane w wyniku tego procesu parametry przedstawione zostały w tabeli 1.

Tabela 1. Wyniki optymalizacji hiperparametrów - metoda SVC.

Częstotliwość próbkowania	Funkcja celu	C	Jądro	Stopień wielomianu
Różnorodna	F1	1,9316	Sigmoidalne	Nie dotyczy
	Dokładność	1,9109	Sigmoidalne	Nie dotyczy
44100 Hz	F1	1,9186	Sigmoidalne	Nie dotyczy
	Dokładność	1,9116	Sigmoidalne	Nie dotyczy

Drugą wykorzystaną metodą klasyfikacji był algorytm lasu losowego. Jako funkcję celu wykorzystano metrykę F1 (macro). Wykorzystując bibliotekę *Optuna* wykonano optymalizację trzech parametrów – liczbę drzew w lesie losowym, maksymalną głębokość (ilość poziomów) drzewa oraz minimalną ilość obiektów w węźle, która umożliwi dalszy podział drzewa. Wykonano po 500 prób w procesach optymalizacji, a jej wyniki optymalizacji przedstawiono w tabeli 2.

Tabela 2. Wyniki optymalizacji - algorytm lasów losowych.

Częstotliwość próbkowania	Liczba drzew	Maksymalna głębokość	Minimalna liczba obiektów w węźle
Różnorodna	48	6	2
44100 Hz	40	6	6

Analiza wyników

Zbiór testowy liczył 43 obiekty. Uzyskane w wyniku klasyfikacji metryki (dokładność, F1, czułość, precyzja) z dokładnością do czwartego miejsca po przecinku przedstawiono w tabeli 3 i 4. Macierze pomyłek przedstawiono w tabelach 5-10.

Tabela 3. Ewaluacja klasyfikacji - metoda SVC.

Częstotliwość próbkowania	Funkcja celu	Dokładność	F1	Czułość	Precyzja
Różnorodna	F1	0,8605	0,8603	0,8605	0,8691
	Dokładność	0,8605	0,8603	0,8605	0,8691
44100 Hz	F1	0,8605	0,8603	0,8605	0,8691
	Dokładność	0,8605	0,8603	0,8605	0,8691

Tabela 4. Ewaluacja klasyfikacji - algorytm lasów losowych.

Częstotliwość próbkowania	Dokładność	F1	Czułość	Precyzja
Różnorodna	0,8837	0,8848	0,8837	0,8930
44100 Hz	0,8837	0,8814	0,8837	0,8948

Tabela 5. Macierz pomyłek. Metoda SVC (F1) – przed resamplingiem.

Gatunek oryginalny	Gatunek przewidywany				
	Skowronek	Derkacz	Puchacz	Sóweczka	Krzyżodziób
Skowronek	7	2	0	0	0
Derkacz	0	8	0	0	0
Puchacz	0	0	7	0	2
Sóweczka	0	0	1	8	0
Krzyżodziób	1	0	0	0	7

Tabela 6. Macierz pomyłek. Metoda SVC (F1) – po resamplingu.

Gatunek oryginalny	Gatunek przewidywany				
	Skowronek	Derkacz	Puchacz	Sóweczka	Krzyżodziób
Skowronek	7	2	0	0	0
Derkacz	0	8	0	0	0
Puchacz	0	0	7	0	2
Sóweczka	0	0	1	8	0
Krzyżodziób	1	0	0	0	7

Tabela 7. Macierz pomyłek. Metoda SVC (dokładność) – przed resamplingiem.

Gatunek oryginalny	Gatunek przewidywany				
	Skowronek	Derkacz	Puchacz	Sóweczka	Krzyżodziób
Skowronek	7	2	0	0	0
Derkacz	0	8	0	0	0
Puchacz	0	0	7	0	2
Sóweczka	0	0	1	8	0
Krzyżodziób	1	0	0	0	7

Tabela 8. Macierz pomyłek. Metoda SVC (dokładność) – po resamplingu.

Gatunek oryginalny	Gatunek przewidywany				
	Skowronek	Derkacz	Puchacz	Sóweczka	Krzyżodziób
Skowronek	7	2	0	0	0
Derkacz	0	8	0	0	0
Puchacz	0	0	7	0	2
Sóweczka	0	0	1	8	0
Krzyżodziób	1	0	0	0	7

Tabela 9. Macierz pomyłek. Algorytm lasów losowych – przed resamplingiem.

Gatunek oryginalny	Gatunek przewidywany				
	Skowronek	Derkacz	Puchacz	Sóweczka	Krzyżodziób
Skowronek	7	2	0	0	0
Derkacz	0	8	0	0	0
Puchacz	1	0	8	0	0
Sóweczka	0	0	1	8	0
Krzyżodziób	1	0	0	0	7

Tabela 10. Macierz pomyłek. Algorytm lasów losowych – po resamplingu.

Gatunek oryginalny	Gatunek przewidywany				
	Skowronek	Derkacz	Puchacz	Sóweczka	Krzyżodziób
Skowronek	6	2	1	0	0
Derkacz	0	8	0	0	0
Puchacz	0	0	9	0	0
Sóweczka	0	0	1	8	0
Krzyżodziób	1	0	0	0	7

Jak zostało zaprezentowane w powyższych tabelach, udało się uzyskać dużą skuteczność klasyfikacji za pomocą wykorzystanych metod. Na podstawie obliczonych metryk oraz przedstawionych w tabelach 5-8 macierzach pomyłek można stwierdzić, że uzyskano optymalne wyniki dla tej metody. Zarówno metryki jak i macierze są identyczne, niezależnie od resamplowania przed ekstrakcją cech oraz od wyboru funkcji celu przy optymalizacji hiperparametrów. Tabele 5-8 pozwalają stwierdzić, że przy klasyfikacji zbioru testowego algorytm błędnie sklasyfikował dwa sygnały pochodzące od skowronka (jako derkacza), jeden plik audio sóweczki (został uznany jako dźwięki puchacza) oraz jeden plik krzyżodzioba (jako skowronka). Pozostałe 39 sygnały zostały sklasyfikowane poprawnie.

Analiza metryk dla algorytmu lasów losowych pozwala na stwierdzenie, że resampling plików audio ma niewielkie znaczenie. Widzimy nieznaczne różnice między metryką F1 oraz precyzją (różnice na 3 miejscu znaczącym). Nie są to jednak na tyle duże różnice, aby stwierdzić że resampling ma decydujące znaczenie w klasyfikacji. Na podstawie porównania tabeli 3 i tabeli 4 (metryki obu metod) można wywnioskować, że algorytm lasów losowych poradził sobie nieznacznie lepiej od metody SVC. Dokładność sięga ponad 88%, podczas gdy maszyna wektorów losowych uzyskała wynik 86%. Pozostałe metryki również wykazują lepszą wydajność drugiej metody.

Podsumowanie

Celem projektu było stworzenie algorytmu uczenia maszynowego, pozwalającego na klasyfikację plików audio do odpowiedniego gatunku jednego z pięciu ptaków. Za pomocą biblioteki *openSMILE* wyekstrahowano 88 cech każdego z 215 sygnałów w oryginalnej częstotliwości próbkowania i po resamplingu do częstotliwości 44100 Hz. Na podstawie otrzymanych parametrów wykonano klasyfikację dwoma metodami – maszyną wektorów nośnych i algorytmem lasów losowych. Uzyskano satysfakcjonujące wyniki, gdzie dokładność (oraz inne pozostałe metryki) wynosiły kolejno około 86% i 88%. Jest to wynik bardzo dobry, pozwalający na prawidłową klasyfikację większości testowanych sygnałów.

W celu dalszego rozwoju projektu i być może uzyskania jeszcze lepszych wyników można spróbować wykorzystać inne klasyfikatory (np. k-najbliższych sąsiadów). Należałoby rozważyć opcję ekstrakcji analizowanych cech sygnałów w inny sposób, co być może wymagałoby konwersji plików do formatu wave. Pozwoliłoby to na wykorzystanie biblioteki *Librosa* i np. ekstrakcję melowo-częstotliwościowych współczynników cepstralnych (MFCC). Dodatkowo warto rozszerzyć algorytm o większą ilość klasyfikowanych gatunków ptaków, a tym samym stworzyć potencjalne wykorzystanie komercyjne.