

# AutoKeras

Tomasz Kurpiński, Jan Kruszewski, Damian Skowroński

May 2022

## 1 Wstęp

W poniższym artykule omówione zostaną efekty pracy naszej grupy podczas projektu w ramach przedmiotu warsztaty badawcze. Postawionym przed nami zadaniem było zapoznanie się z wybranym przez nas frameworkiem AutoML, stworzenie funkcji przygotowującej dowolny zbiór danych opisujących binarny problem predykcyjny, która następnie tworzy dla niego model. Na koniec stworzoną funkcję mieliśmy przetestować na zbiorach dotyczących binarnej klasyfikacji z benchmarku. W projekcie będziemy odnosić się do badań właśnie przedstawionych w artykule benchmarku, próbując stworzyć swego rodzaju benchmark dla pominiętego w badaniu pakietu AutoKeras.

## 2 Opis frameworku

W tym ustępie, krótko opiszemy framework AutoKeras.

### 2.1 Twórcy

AutoKeras to framework AutoML oparty na pakiecie Keras, rozwijany przez grupę osób związanych z uniwersytetem Texas AM. Głównym założeniem twórców było stworzenie pakietu, który da możliwość korzystania z uczenia maszynowego osobom nawet nie mającym dużego doświadczenia w tej dziedzinie. Pakiet posiada obszerną i czytelną dokumentację.

### 2.2 Co wyróżnia AutoKeras?

Parę kwestii wyróżnia AutoKeras w porównaniu z innymi frameworkami AutoML. Nie potrzebne jest wykorzystywanie usług w chmurze, które nie są darmowe i nie są przystępne dla każdego, który chce używać machine learningu. AutoKeras zapewnia również bezpieczeństwo i prywatność danych. Warto dodać również, że AutoKeras skupia się na problemach deep learning w przeciwieństwie do frameworków takich jak SMAC, TPOT, Auto-WEKA czy Auto-Sklearn, które skupiają się na shallow models.

## 2.3 API

Interfiers AutoKerasa został stworzony na podobieństwo sklearna. Budowę modelu można przeprowadzić w 3 liniach kodu za pomocą konstruktora i metod `fit` oraz `predict`. Warto zaznaczyć, że posiada on dwa poziomy 'task-level' dla użytkowników o mniejszej znajomości systemu oraz 'search-level' dla zaawansowanych użytkowników, którzy sami kontrolują preprocessing oraz architekturę sieci neuronowej. Generalnie AutoKeras udostępnia nam następujące modele: `ImageClassifier`, `ImageRegressor`, `TextClassifier`, `TextRegressor`, `StructuredDataClassifier`, `StructuredDataRegressor` i `AutoModel`. Framework posiada również wiele ułatwień dla użytkownika.

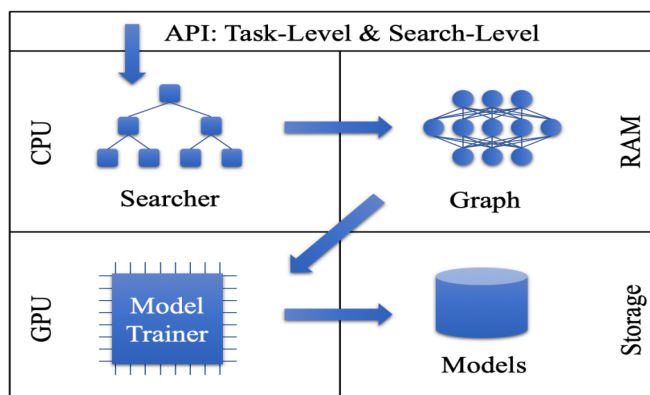
## 2.4 Wykorzystanie zasobów przez AutoKeras

AutoKeras z założenia nastawiony jest na pełne wykorzystanie możliwości CPU, GPU i pamięci RAM, w której zapisywane są tylko aktualnie ważne informacje. Reszta danych zapisywana jest w innych miejscach takich jak dysk twardy.

## 2.5 Podział AutoKerasa na moduły

Framework jest podzielony na 4 główne moduły.

- Moduł `Searcher`, odpowiedzialny jest za wyszukiwanie architektury neuronowej. Używa on optymalizacji bayesowskiej oraz procesu gaussowskiego. Algorytmy szukające korzystają z CPU.
- W module `Trainer` trenowane są sieci neuronowe.
- `Graph` to moduł przetwarzający wykresy obliczeniowe, które kontroluje `Searcher` w celu morfizmu sieci.
- Ostatni moduł to `Storage`. Odpowiedzialny jest za zapisywanie gotowych modeli. Ze względu na ich rozmiar nie są zapisywane w pamięci RAM.



### 3 Preprocessing

Niezbędny preprocessing pakiet wykonuje za nas przed trenowaniem. Stąd nasza funkcja nie posiada żadnych dodatkowych fragmentu kodu poświęconych tej czynności. AutoKeras wykonuje za nas podstawową wektoryzację, czyszczenie danych czy normalizację, będąca ważnym krokiem dla sieci neuronowych.

### 4 Wyniki ewaluacji

Udało nam się sprawdzić działanie naszej funkcji dla 20 binarnych zbiorów z benchamarka. Wyniki dla metryki accuracy oscylowały wokół wartości 0.725. Najlepiej model poradził sobie ze zbiorami jasmine, mfeat-factors oraz shuttle. Dla każdego z tych zbiorów model osiągnął wynik równy około 0.97. Zdecydowanie najgorszy wynik mogłoby się zdawać mamy dla zbioru sylvine, jednak mamy tu do czynienia z klasyfikacją binarną, więc wystarczy zamienić etykiety by otrzymać bardzo dobry wynik.