Uczenie maszynowe - jak zacząć

Patryk Miziuła
Data Community Bydgoszcz i Toruń
27.02.2019



- 1. Kim jestem
- 2. Poprzednie wystąpienie
- 3. Środowisko
- 4. Biblioteki
- 5. Linki



Kim jestem



Kim jestem



- **✗** VI L.O. Bydgoszcz
- **X** doktorat z matematyki w IMPAN
- X data scientist w deepsense.ai



Poprzednie wystąpienie



Poprzednie wystąpienie

Bydgoszcz JUG meetup #18: http://y2u.be/Qb67Lfq1igM

- 🗴 popyt na data scientistów
- X dokonania algorytmów
- 🗶 zasada działania algorytmów





- x uczenie przez wzmacnianie
- **x** open source
- drugi najsilniejszy silnik
 szachowy świata
 (49,5 50,5 w finale 23.02.19)



Środowisko







- 🗴 🛮 dane mają mniej niż 1 GB
- × nie używasz sieci neuronowych
- x nie stosujesz skomplikowanych technik walidacyjnych ani wyszukiwania hiperparametrów



NIE

✗ GPU/TPU/Xeon Phi/...

x klaster obliczeniowy

🗶 chmurka



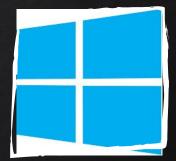
System operacyjny











nie ma się czego bać



Język programowania





https://www.kaggle.com/stardust0/python-vs-r





Jupyter Notebook ≤ Jupyter Lab



Biblioteki



Podstawowe biblioteki

NumPy - obliczenia na wektorach i macierzach



- SciPy narzędzia statystyczne
- pandas praca z ramkami danych 🛗
- pandas-profiling automatyczne raporty 🟳
- seaborn wizualizacje 📈
- scikit-learn podstawowe modele uczenia maszynowego





Najlepsze modele

dane tabelaryczne



gradient boosting decision trees



XGBoost, LightGBM



Najlepsze modele

obraz, dźwięk, język naturalny 🔯 🎵 🚞



sieci neuronowe 🐒



Keras

TensorFlow





Linki



github.com/buus2/dcbit

- 🗶 kod dema LightGBM
- 🗶 link do dema Obcy kontra Predator (Keras i PyTorch) 🤍
- 🗴 🛮 link do linków do ciekawych miejsc 🔍



Dziękuję za uwagę!

Pytania?

linkedin.com/in/patryk-miziuła buus@wp.pl