

VDMML - zakres PoC

Wioletta Stobieniecka

Mariusz Dzieciątko

Łukasz Chojak

Mariusz Zychowicz



Obszar I

Proces modelowania

Obszar I proces modelowania

Wstępna propozycja zakresu:

- Wybór i implementacja 1 modelu predykcyjnego/klasyfikacyjnego
- Wykorzystanie Model Studio i porównanie modeli w pipeline analitycznym
- Warsztat – prezentacja cyklu analitycznego (rejestracja, retrenowanie)
- Rejestracja i porównanie z modelami Open Source (scikit-learn)
- Wykorzystanie API – automatyzacja procesu, m.in. budowy modelu

Dane:

Nie więcej niż 2 mln rekordów, około 20GB danych

Zasoby:

Potrzebne zasoby: RAM 196GB, HDD 200GB, 12-16 rdzeni

Obszar II

Analiza tekstu

Analiza opisów zdarzenia

Wstępna propozycja zakresu:

- Zdefiniowanie 3 kategorii/ miejsc zdarzenia
- Zdefiniowanie 10 encji pozyskiwanych z opisów dotyczących np. miejsca, pory, zakresu uszkodzeń
- Dla każdej z kategorii analiza anomalii
- Prezentacja wyników w postaci dashboardu/raportu

Dane:

Analiza opisów szkód i miejsca zdarzenia odbyłaby się na próbce szkód z jednego roku (2 mln rekordów) - czyli około 10GB danych

Zasoby:

Potrzebne zasoby: RAM 196GB, HDD 200GB, 10-12 rdzeni

Obszar III

Rozpoznawanie obrazów

Analiza obrazów szkód

Wstępna propozycja zakresu:

- Zdefiniowanie 10+ kategorii uszkodzeń
- Wytrenowanie modelu klasyfikacyjnego (2-3 SOTA architektury CNN)
- Rejestracja modelu w SAS Model Manager i publikacja w silniku CAS/MAS
- Prezentacja wyników w VI (jako dodatek do przykładowych alertów)

Dane:

Analiza rodzajów uszkodzeń odbyłaby się na próbce dostarczonej przez SAS (<1000 zdjęć)

Zasoby:

Potrzebne zasoby: RAM 196GB, HDD 200GB, 12-16 rdzeni



sas.com

Company Confidential – For Internal Use Only
Copyright © SAS Institute Inc. All rights reserved.

