AiSECLAB

ROZWIĄZANIA Z ZAKRESU SZTUCZNEJ INTELIGENCJI OPRACOWANE W POLSCE NA TLE ROZWIĄZAŃ ŚWIATOWYCH GIGANTÓW

Marcin Piołun-Noyszewski, AiSECLAB, CEO

AGENDA

- O nas
- O naszych wdrożeniach systemów SI
- Propozycje zastosowania SI w wymiarze sprawiedliwości
- Czym różnimy się od innych?

O NAS

- Start w 2016
- Polska firma
- Polski kapitał
- Polskie rozwiązania
- Polski zespół
- Pełny proces R&D i wytwórczy w Polsce
- Pełna własność modeli SI w rękach polskiej firmy
- Współpraca i osiągnięcia z Free Construction
- MPN:
  - Lider Informatyki za system dla Sądu Najwyższego RP
  - Deloitte Forbes Technology Fast 50 (18. w kat. sprzedaży bezpieczeństwa)

DZIAŁANIA NAUKOWE

Wnioski patentowe

- Automatyczny dobór algorytmu przewidywania trendów
- Przewidywanie na podstawie danych pochodzących z niejednorodnych okresów

Współpraca naukowa

- Wojskowa Akademia Techniczna
- Instytut Maszyn Matematycznych
- Uniwersytet Warszawski

Największe konferencje i publikacje naukowe

- Polskiego Towarzystwa Statystycznego
- Secure NASK, Gigacon
- Taksonomia - Prace Naukowe Uniwersytetu we Wrocławiu, Wydawnictwo Uniwersytetu Łódzkiego


POWER OF ATTORNEY TO PROSECUTE APPLICATIONS BEFORE THE USPTO


ASSIGNMENT OF INVENTION

KLIENCI AISECLAB I FREE CONSTRUCTION

> 1 mln
użytkowników na świecie

> 27 krajów
zasięgu terytorialnego

> 60
największych korporacji, urzędów i instytucji

> 12
rozwiązań branżowych

KLIENCI PRODUKTÓW AIS & FC

- The Courts Service of Ireland (Irlandia)
- Federalne Ministerstwo Obrony Niemiec (Niemcy)
- Departament Obrony (Australia)
- Raytheon Systems Ltd (Wielka Brytania)
- Rolls Royce Marine Electrical Systems (Wielka Brytania)
- Staffordshire Police (Wielka Brytania)
- Vulco SA (USA)
- Ashland Inc (USA)
- Banque Raiffeisen (Luxemburg)
- Canal de Isabel II (Hiszpania)
- Colliers Jardine Nowa Zelandia (Nowa Zelandia)
- Staffordshire Police (Wielka Brytania)
- Ministerstvo Financii SR (Słowacja)
- Australian Bureau of Statistics (Australia)
- Hyundai Cars (Wielka Brytania)
- Formula One Management Ltd (Wielka Brytania)
- Mattel Group (Holandia/Polska)
- Total Fina Elf Gas & Power Ltd (UK)
- Total Gas & Power Ltd (UK)
- Baltimore Aircoil International (Belgia)
- Royal Australian Navy (Australia)
- Grupa COMP (Polska)
- NASK (Polska)
- Policja Polska (Polska)
- ... wielu innych

PRZYKŁADY NSZYCH SYSTEMÓW SZTUCZNEJ INTELIGENCJI

SYSTEM SI W NASK OSE B3

OGÓLNOPOLSKA SIEĆ EDUKACYJNA B3

Tabela 5 Maksymalne skalowanie Systemu - po skorzystaniu z prawa opcji

Województwo	Regionalny Węzeł Bezpieczeństwa	Przepustowość Systemu B3 [Mbps]*	Liczba zapytań [http/HTTPS na sekundę]
MAZOWIECKIE	WAW	48 000	93 750
ŚLĄSKIE	KAT	28 800	56 250
MAŁOPOLSKIE	KRA	15 360	30 000
WIELKOPOLSKIE	POZ	19 200	37 500
PODKARPACKIE	RZE	11 520	22 500
LUBELSKIE	LUB	11 520	22 500
DOLNOŚLĄSKIE	WRO	11 520	22 500
ŁÓDZKIE	LOD	11 520	22 500
POMORSKIE	GDA	11 520	22 500
KUJAWSKO-POMORSKIE	TOR	11 520	22 500
ZACHODNIOPOMORSKIE	SZC	11 520	22 500
WARMIŃSKO-MAZURSKIE	OLS	7 680	15 000
ŚWIĘTOKRZYSKIE	KIE	7 680	15 000
PODLASKIE	BIA	7 680	15 000
OPOLSKIE	OPO	7 680	15 000
LUBUSKIE	ZGO	7 680	15 000

OSE B3 – PODSTAWY FORMALNE

- Ustawa o Ogólnopolskiej Sieci Edukacyjnej (OSE) 2017
- Uruchomienie w 2021
- Część spójnego systemu B1, B2 i B4 zaprojektowanego przez NASK
- B3 skupia się na analizie korzystającej ze sztucznej inteligencji i ML
- Realizacja B3: Enigma SOI, Free Construction i AiSECLAB

OSE B3 – MODELE ML/DL

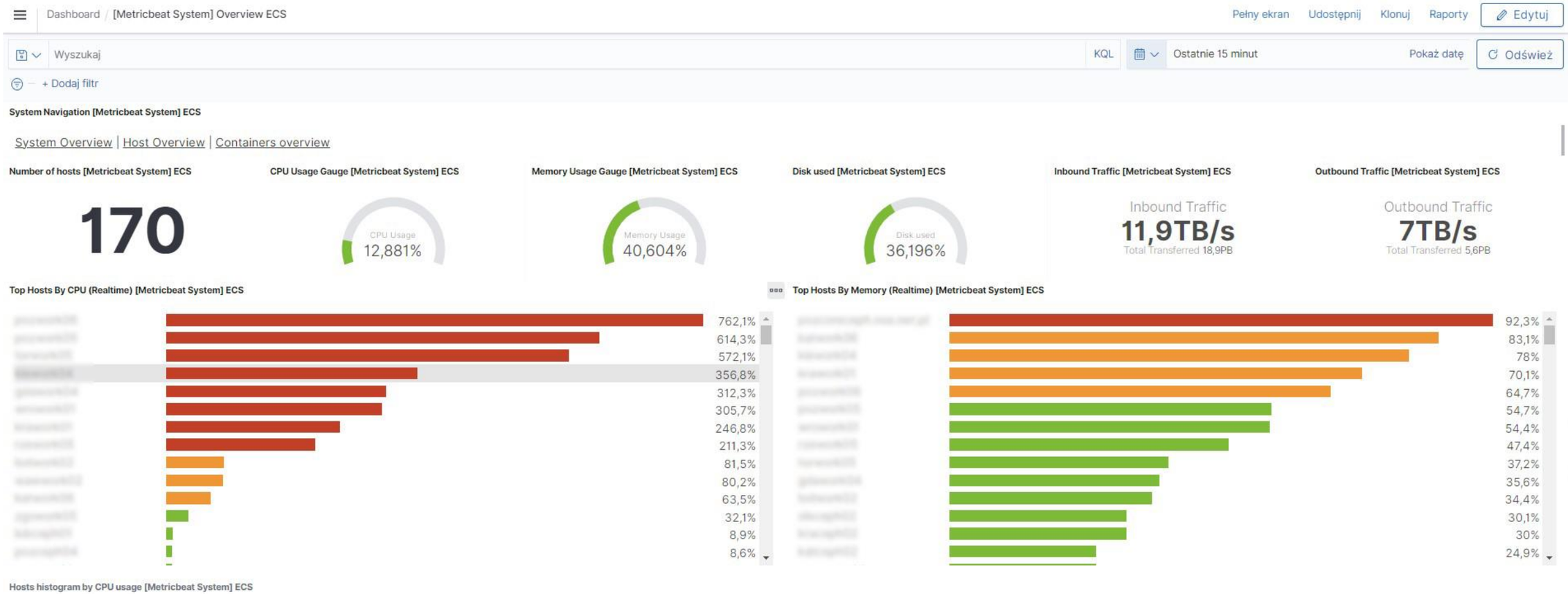
- Modele DL/ML z kategorii Computer Vision, Natural Language Processing
- Miara F1 oscyluje w granicach 90% - 98%

- narkotyki
- broń
- pornografia
- nagość
- cyberprzemoc
- autoagresja
- gore
- sexting
- sextorsion
- childgrooming
- mowa nienawiści
- przemoc
- materials adult
- tytoń
- alkohol
- niebezpieczne treści
- hazard

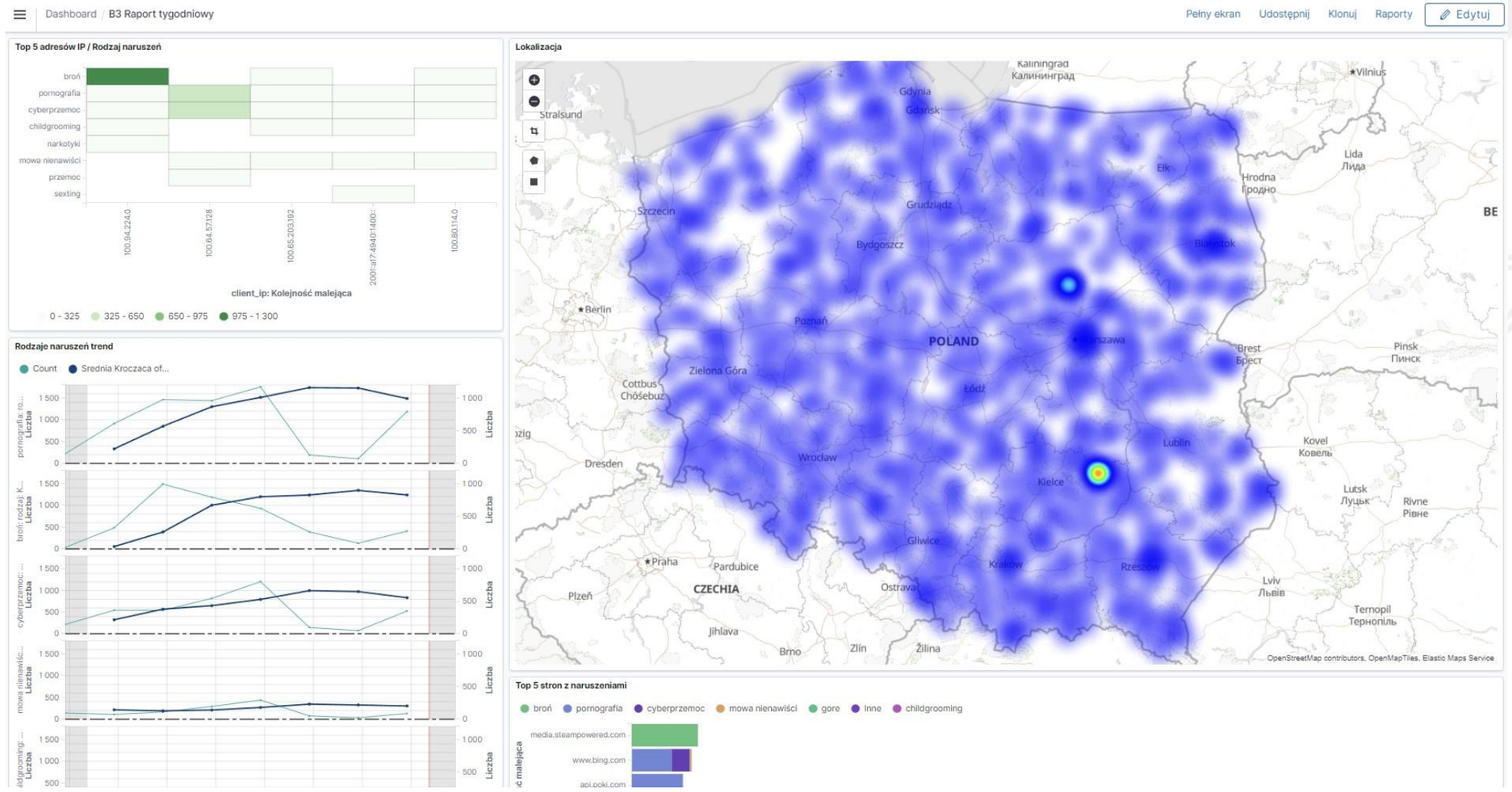
OSE B3 – SKALA SYSTEMU

- Objęte ponad 20 000 szkół i innych placówek
- 16 regionów
- ponad 500 komponentów (serwerów) procesujących dane
- ponad 1 600 serwisów
- chwilowe obciążenia do ponad kilku TB/s,
- około 80 milionów obsługiwanych każdej doby żądań analizy SI/ML/DL:
  - ponad 12 milionów obrazów
  - 1.5 miliona filmów
  - 40 milionów tekstów
  - 20 milionów innych materiałów

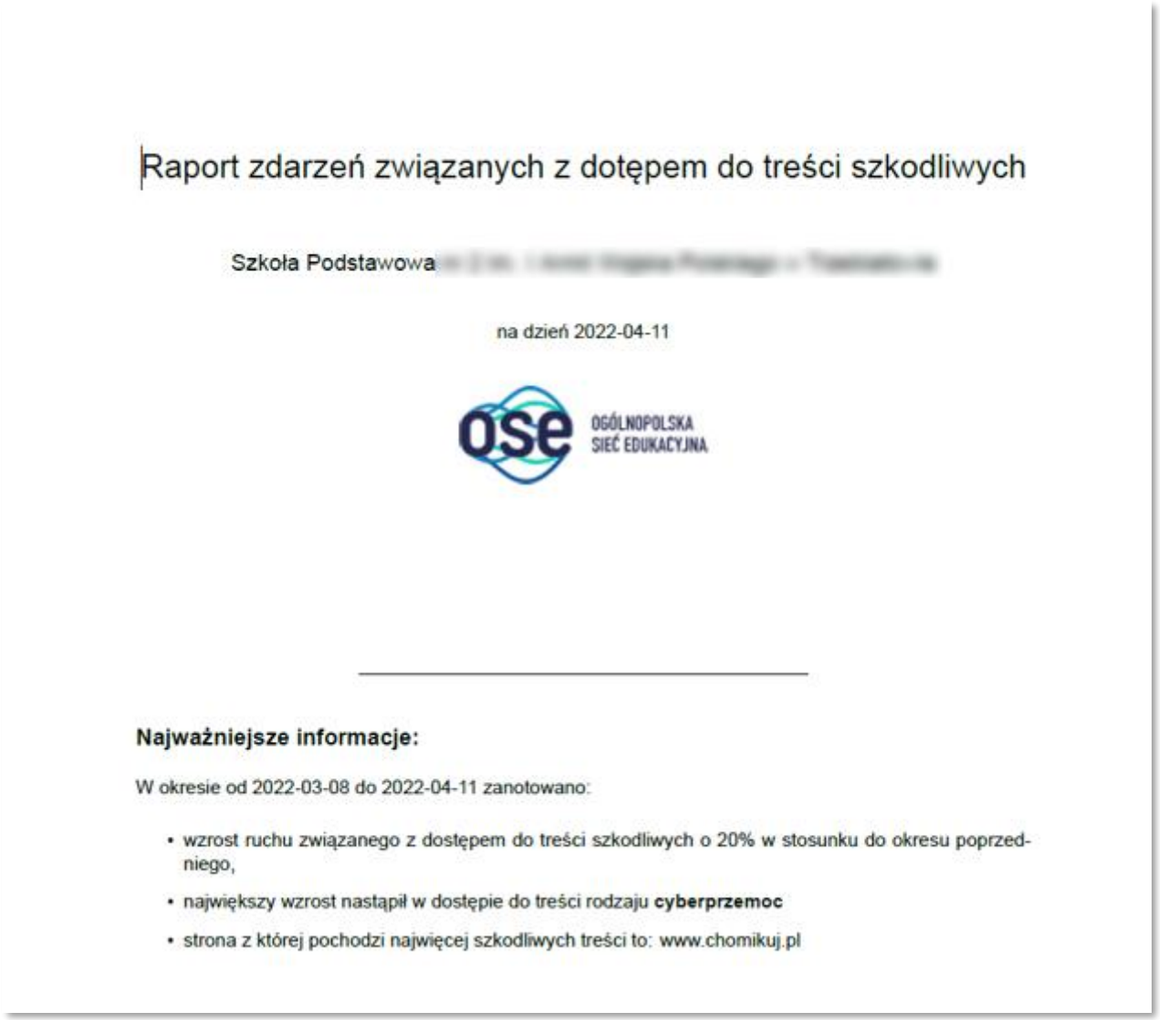
OSE B3 – CHWILOWE OBCIĄŻENIE



WIZUALIZACJE OSE B3



RAPORTY – STRONA TYTUŁOWA



Raport zdarzeń związanych z dotępem do treści szkodliwych

Szkoła Podstawowa

na dzień 2022-04-11

OGÓLNOPOLSKA SIEĆ EDUKACYJNA

Najważniejsze informacje:

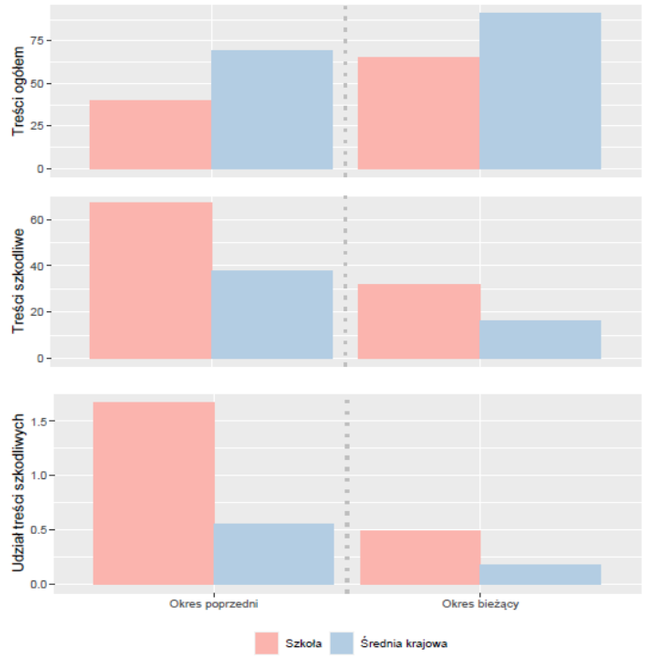
W okresie od 2022-03-08 do 2022-04-11 zanotowano:

- wzrost ruchu związanego z dostępem do treści szkodliwych o 20% w stosunku do okresu poprzedniego,
- największy wzrost nastąpił w dostępie do treści rodzaju **cyberprzemoc**
- strona z której pochodzi najwięcej szkodliwych treści to: www.chomikuj.pl

RAPORTY - ZAWARTOŚĆ

Okres od 2022-03-08 do 2022-04-11

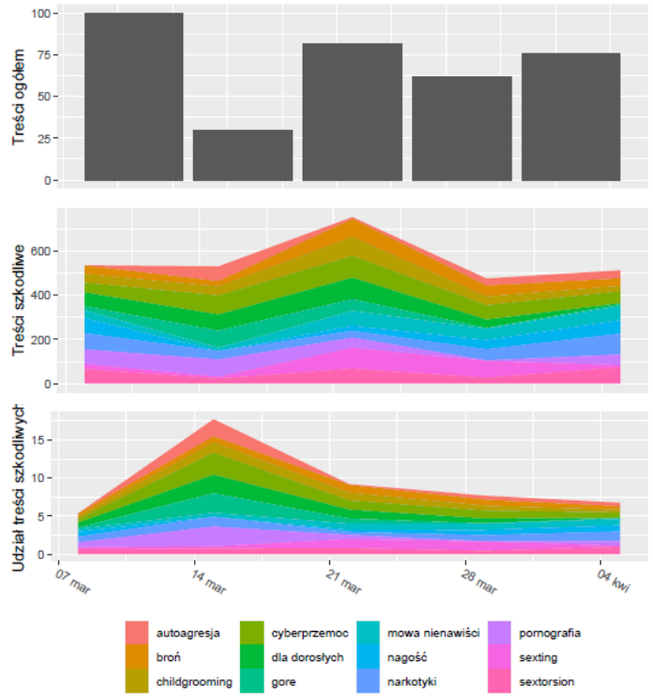
Analiza liczby zdażeń związanych z dostępem do treści szkodliwych i ich udział w ruchu danego typu szkoły na tle kraju w raportowanym okresie.



Rekomendacje:

- brak rekomendacji

Analizowany ruch i rodzaje potencjalnych naruszeń w szkole w kolejnych dniach.



Podsumowanie:

Liczba potencjalnych naruszeń rodzaju **pornografia, cyberprzemoc** jest większa niż liczba potencjalnych naruszeń z poprzedniego okresu. Proponowane działania:

- zapytaj nauczycieli czy zaobserwowali dziwną aktywność wśród uczniów
- sprawdź czy ktoś niepowołany nie korzystał z komputerów

Strony z których pochodziło najwięcej szkodliwych treści.

rodzaj	strona	liczba
broń		
	youtube.com	72
	onet.pl	74
	gazeta.pl	61
pornografia		
	test.pl	65
	youporn.com	44
	pornhub.com	47

An tSeirbhís Chúirteanna
Courts Service

POLICJA

SYSTEM SI W COURTS SERVICE OF IRELAND I W POLICJI POLSKIEJ

COURTS SERVICE OF IRELAND

- Administracja i zarządzanie sądami w Irlandii
- Zarządzanie sądami
- Wspieranie sędziów
- Dostarczanie opinii publicznej informacji o systemie sądów
- Zapewnianie budynków sądowych i obiektów dla użytkowników sądów
- ok. 1 500 użytkowników

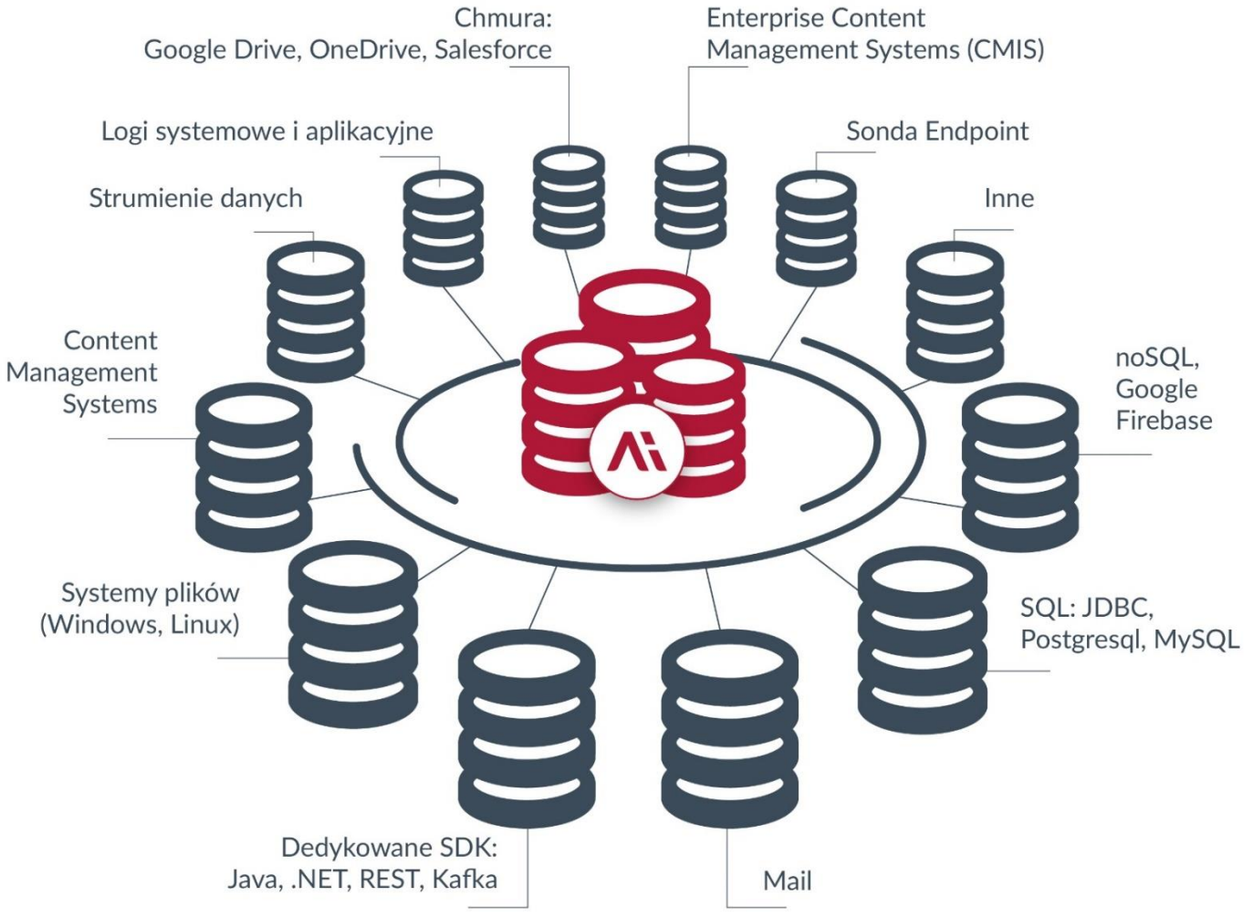
POLICJA POLSKA

- Formacja służąca społeczeństwu
- Przeznaczona do ochrony bezpieczeństwa ludzi
- Utrzymywanie bezpieczeństwa i porządku publicznego
- > 100 000 użytkowników

ROZWIĄZANIE – SECURE MAIL INTELLIGENCE!

- Secure Mail Intelligence!
- Analiza treści poczty elektronicznej
- Zapewnienie bezpieczeństwa poczty elektronicznej
- Ochrona przed niekontrolowanym wyciekiem treści
- Rozliczalność i audyt użytkowników i administratorów

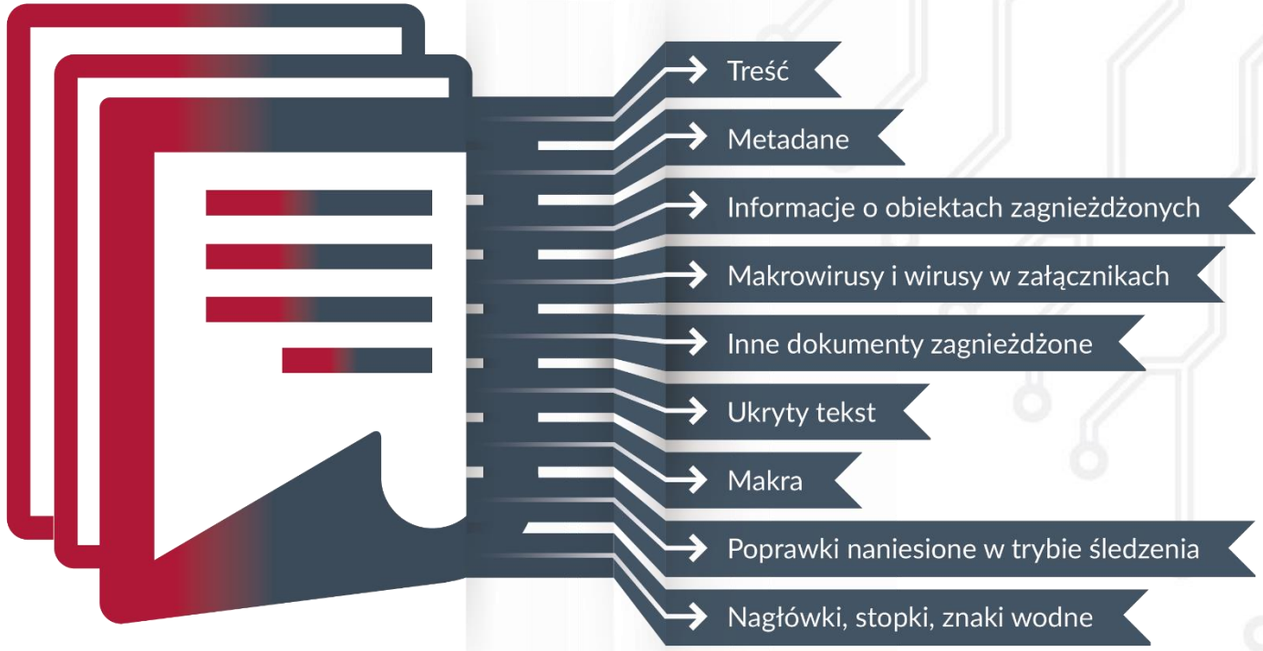
ŹRÓDŁA DANYCH



PRZEDMIOT ANALIZY

- Dokumenty
- Rekordy
- Strumienie danych

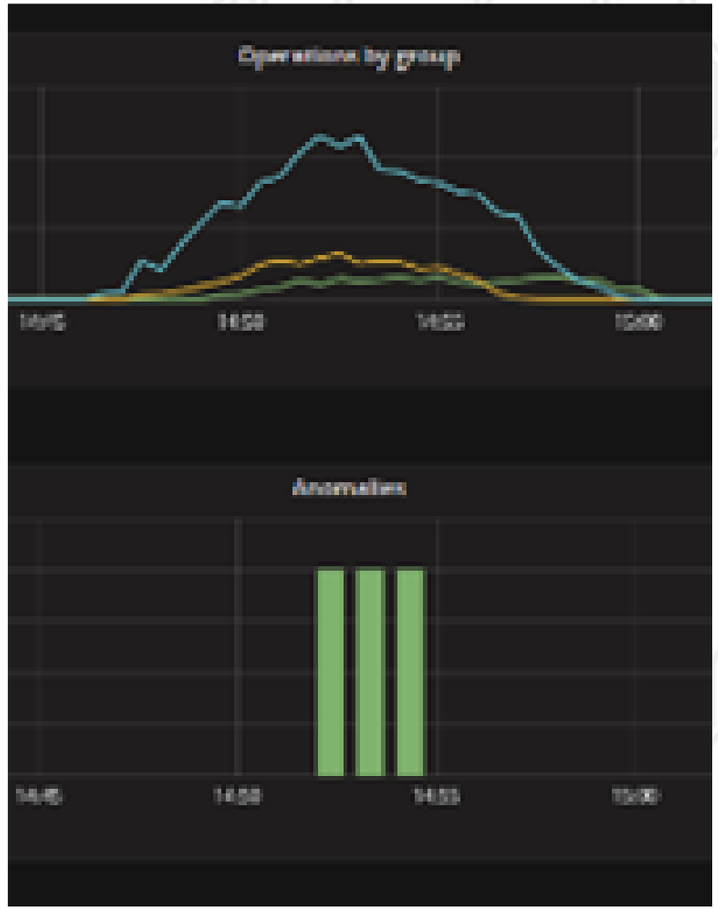
- Dane o systemie
- Informacje o użytkowniku



WYKRYWANIE ANOMALII, DLP

- Analiza w ramach dozwolonego użytku
- Zbyt częste, zbyt duże odczyty plików
- Treści odbiegające od innych pracowników działu
- Logowania do systemów w niestandardowych godzinach
- Transakcje finansowe odbiegające od specyfiki danego klienta
- Ataki DDOS

- Porównania na tle:
  - grup
  - okresów
  - regionów
  - serwerów



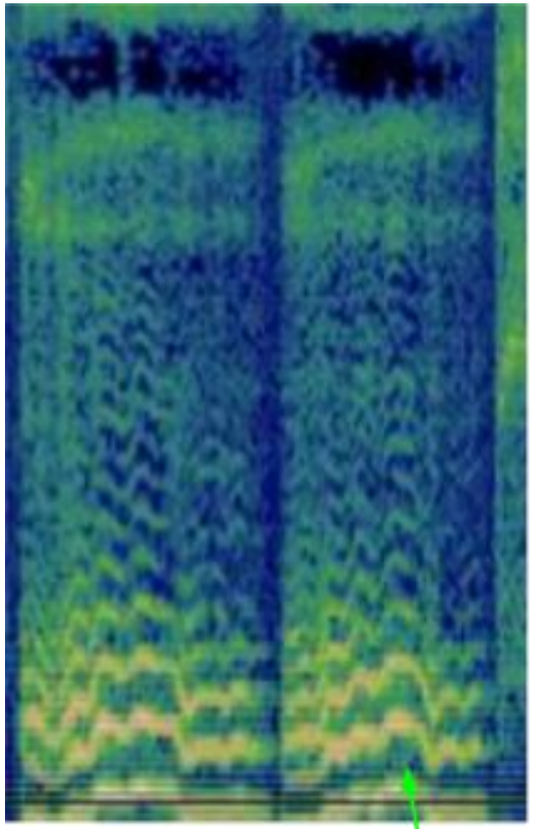
KLASYFIKACJA OBRAZÓW

- Computer Vision
- Rozpoznawanie obiektów i klasyfikacja obrazów
- Wykrywanie stempli, podpisów, tablic rejestracyjnych
- Wykrywanie broni i narkotyków, terrorystów

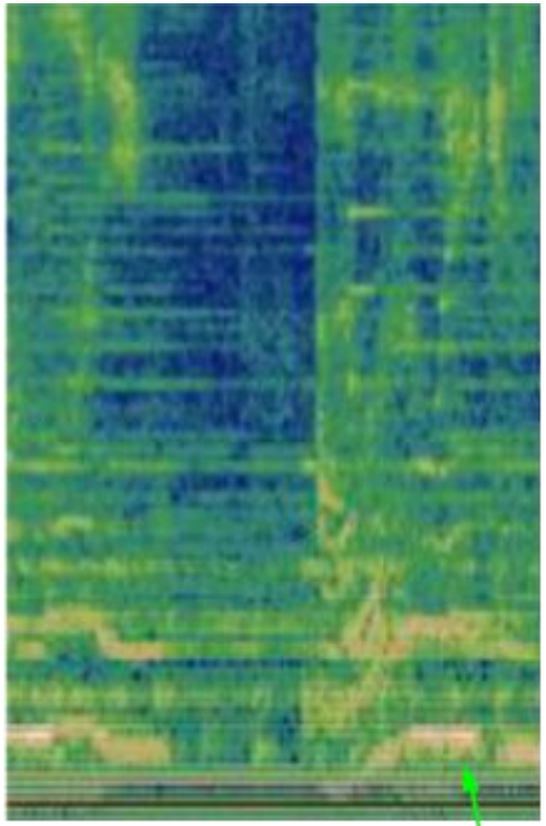
ANALIZA TREŚCI

- Własne słowniki z ponad 25 000 słów
  - polskie
  - angielskie
  - francuskie
- Natural Language Processing
- Odkrywanie znaczenia, podmiotu, struktury zdań
- Wordnet, sieci bayesowskie, fastnet, lucene
- Wektorowe modele sztucznej inteligencji, LTSM
- Mechanizmy uwagi
- Wykrywanie danych osobowych i wrażliwych

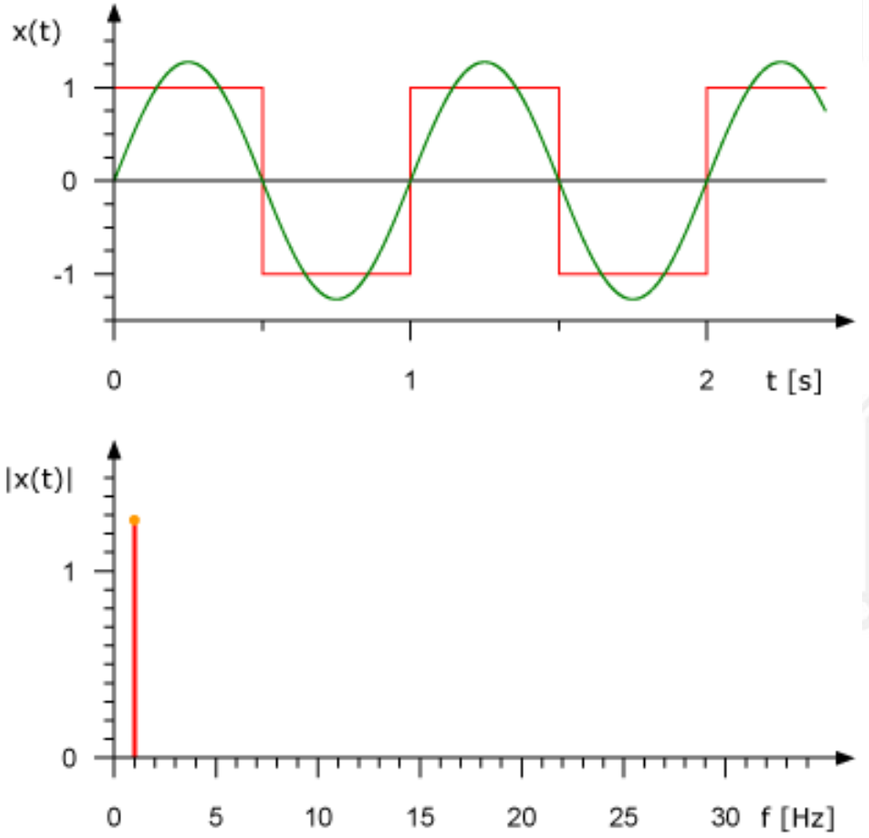
KLASYFIKACJA I WYKRYWANIE DŹWIĘKÓW



Humming



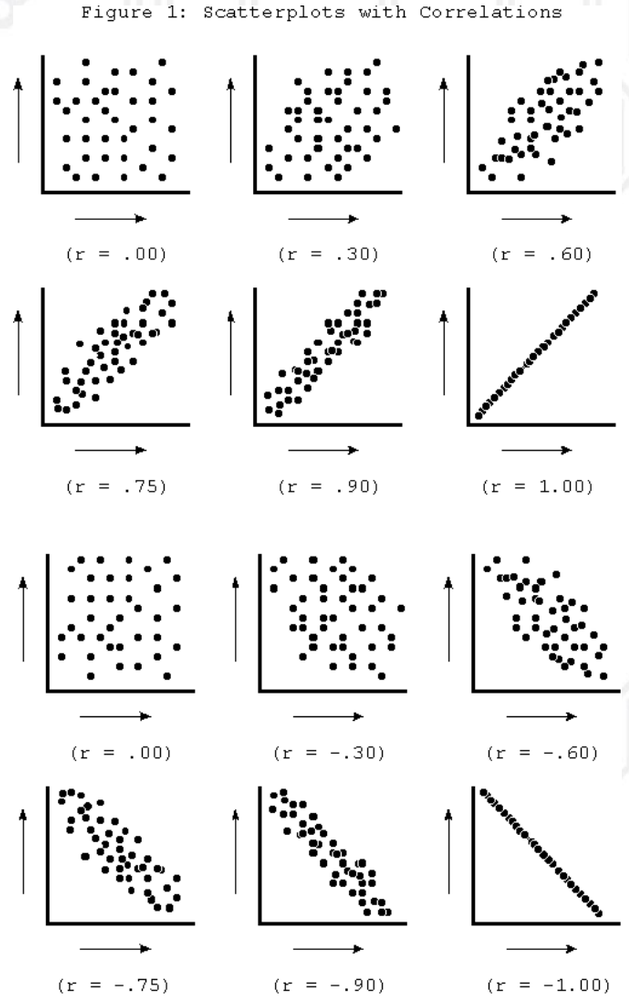
Studio Recording



WYKRYWANIE KORELACJI I UKRYTYCH ZWIĄZKÓW

Wykrywanie zależności pomiędzy np.:

- godziną logowania do systemu a liczbą wytworzonych dokumentów
- wiekiem a zestawem używanych słów
- działem a liczbą zapisów na nieznanych nośnikach



SYSTEMY SZTUCZNEJ INTELIGENCJI DLA WYMIARU SPRAWIEDLIWOŚCI

SDE – PROFILOWANIE OSADZONYCH

- Dane: zdarzenia z dozoru stacjonarnego
- Segmentacja, klasyfikacja, korelacje

- Zagrożenie życia i zdrowia
- Anomalie
- Przewidywanie ucieczki

SDE – WSPOMAGANIE REGUŁ ZBLIŻANIA SIĘ

- Dane: zdarzenia z dozoru mobilnego
- Przygotowanie do zamachu
- Kto do kogo się zbliża?
- Eliminowanie fałszywych alarmów

SDE – PLANOWANIE WIZYT

- Planowanie i optymalizowanie tras
- Przekształcanie incydentów w zgłoszenia do właściwych organów

TRANSKRYPCJA

- Rozmowy telefoniczne
- Monitoring wizyjny (CCTV)
- Posiedzenia

- Analiza NLP zapisu tekstowego
- Analiza sentymentów wypowiedzi

ANALIZA MONITORINGU WIZYJNEGO (CCTV)

- Analiza strumieni wideo
- Wykrywanie sekwencji, obiektów

- Bójki
- Rozróby
- Zgromadzenia, tłum
- Wandalizm
- Próby samobójcze
- Broń

AISECLAB

MODELE GENERATYWNE

- Podobne do GPT
- Wyuczone na konkretnych obszarach wiedzy
    - prawo, regulaminy
    - instrukcje
    - dokumentacja

ANALIZA ZAWARTOŚCI NOŚNIKÓW

- Skanowanie
- Wykrywanie
- Indeksowanie
- Wyszukiwanie

- Przemoc
- Terroryzm
- Szpiegostwo
- Child Sexual Abuse Material (CSAM)

WYKRYWANIE DEEP FAKE

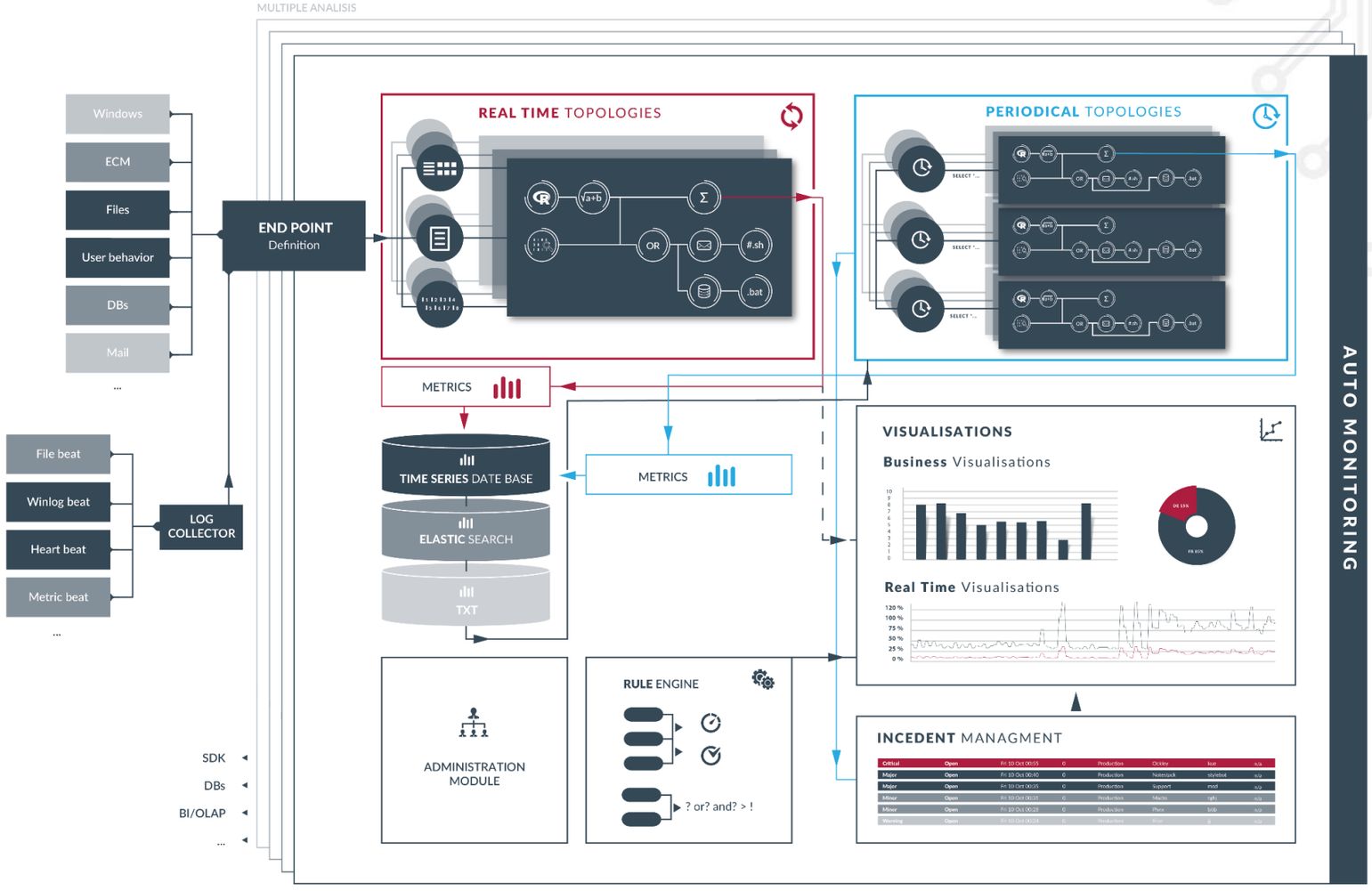
- W wideo
- W audio

O NASZYCH SYSTEMACH SI/ML/DL

CECHY AIPLATFORM

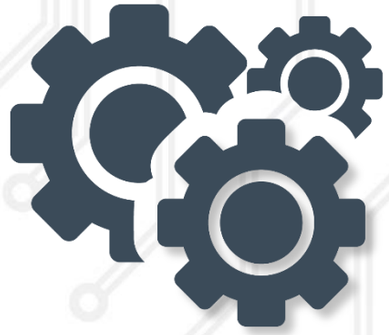
- System rozproszony, dowolnie skalowany horyzontalnie i wertykalnie
- Łatwo rozbudowywalny
- Stabilny, audytowalny i rozliczalny, auto-monitorujący
- Wykorzystujący moc tysięcy CPU i GPU
- Wygodnie zarządzany
- Łatwo aktualizowany
- Szybki
  - wnioskowanie na CPU kilkunastoma modelami ok. 120 ms
  - na GPU od kilku do kilkudziesięciu ms
  - na tanich akceleratorach ok. 50% czasu CPU

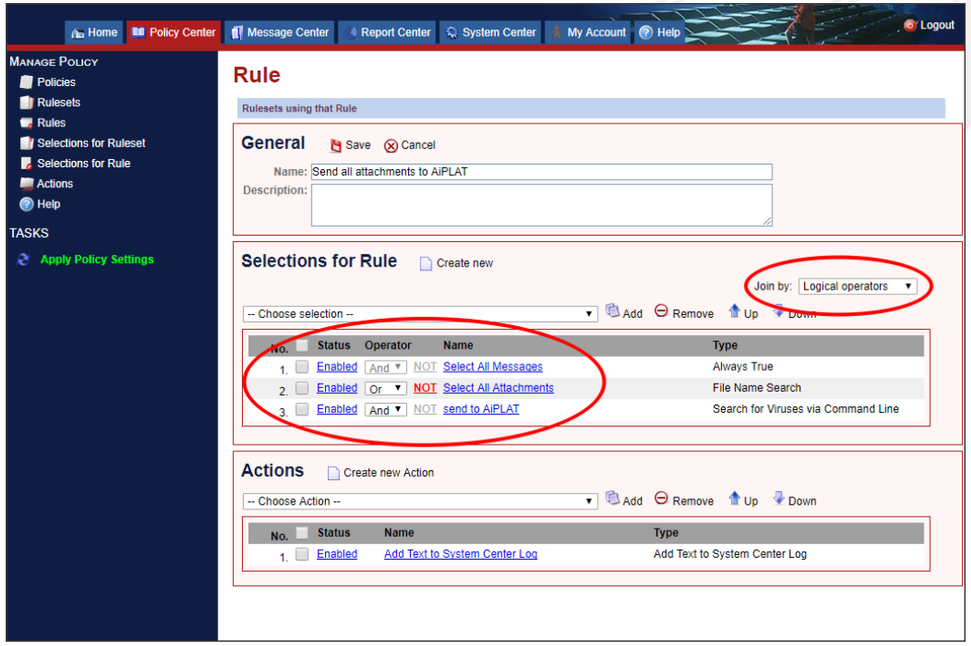
ARCHITEKTURA



SILNIK REGUŁ

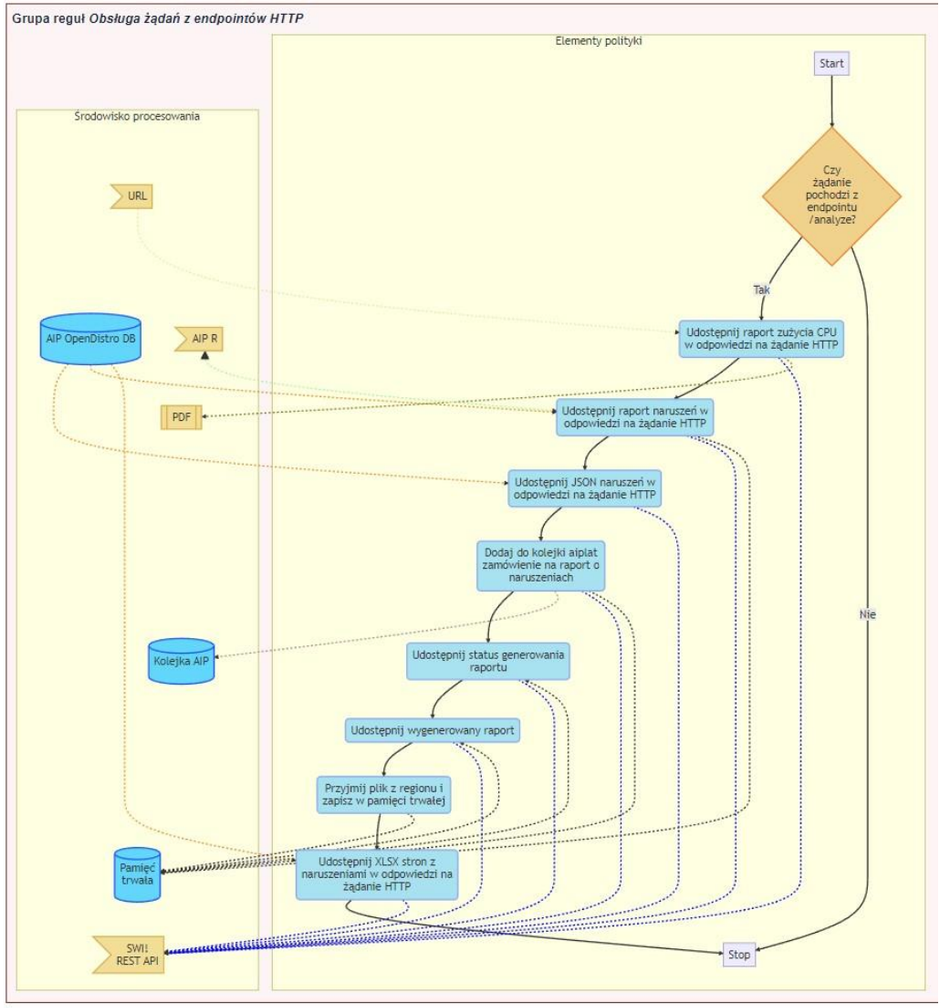
- Reguły o charakterze przyczynowo skutkowym
- Językiem są:
  - selekcje
  - akcje
  - reguły

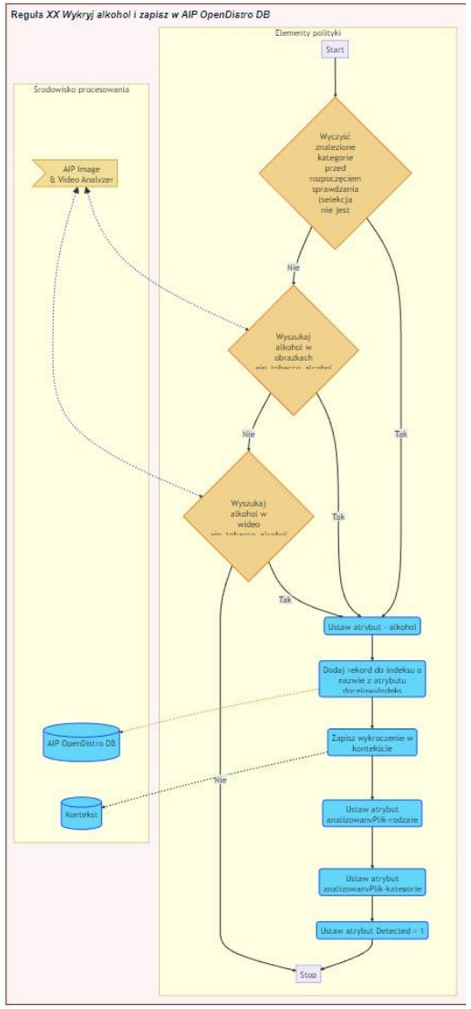




- Operatory logiczne
- Uwzględnianie wag poszczególnych zdarzeń
- Własne selekcje, akcje, reguły
- SDK / API dostępowe
- Analiza kosztu wykonania operacji
- Wygodny interfejs graficzny

WIZUALIZACJE REGUŁ

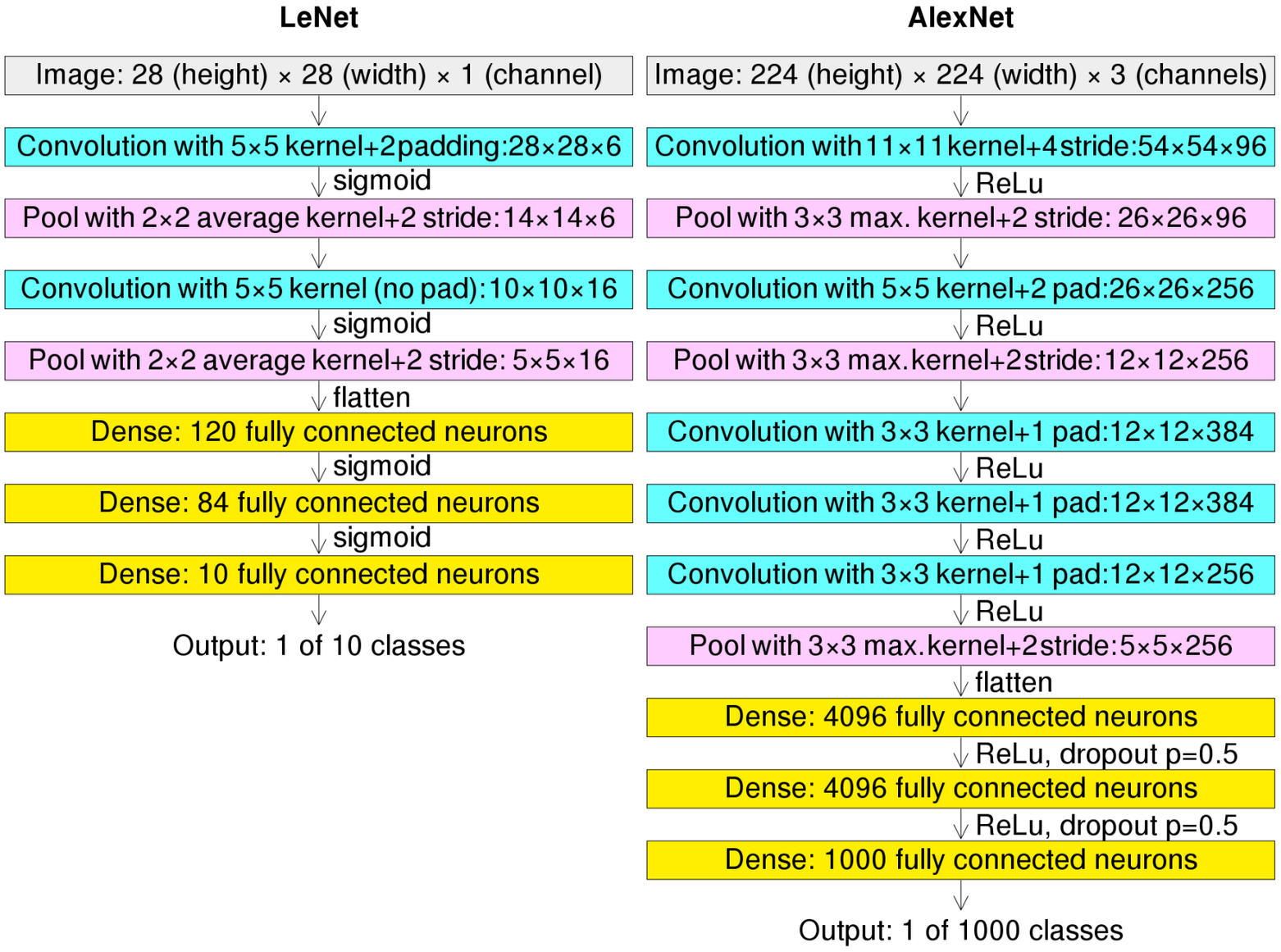




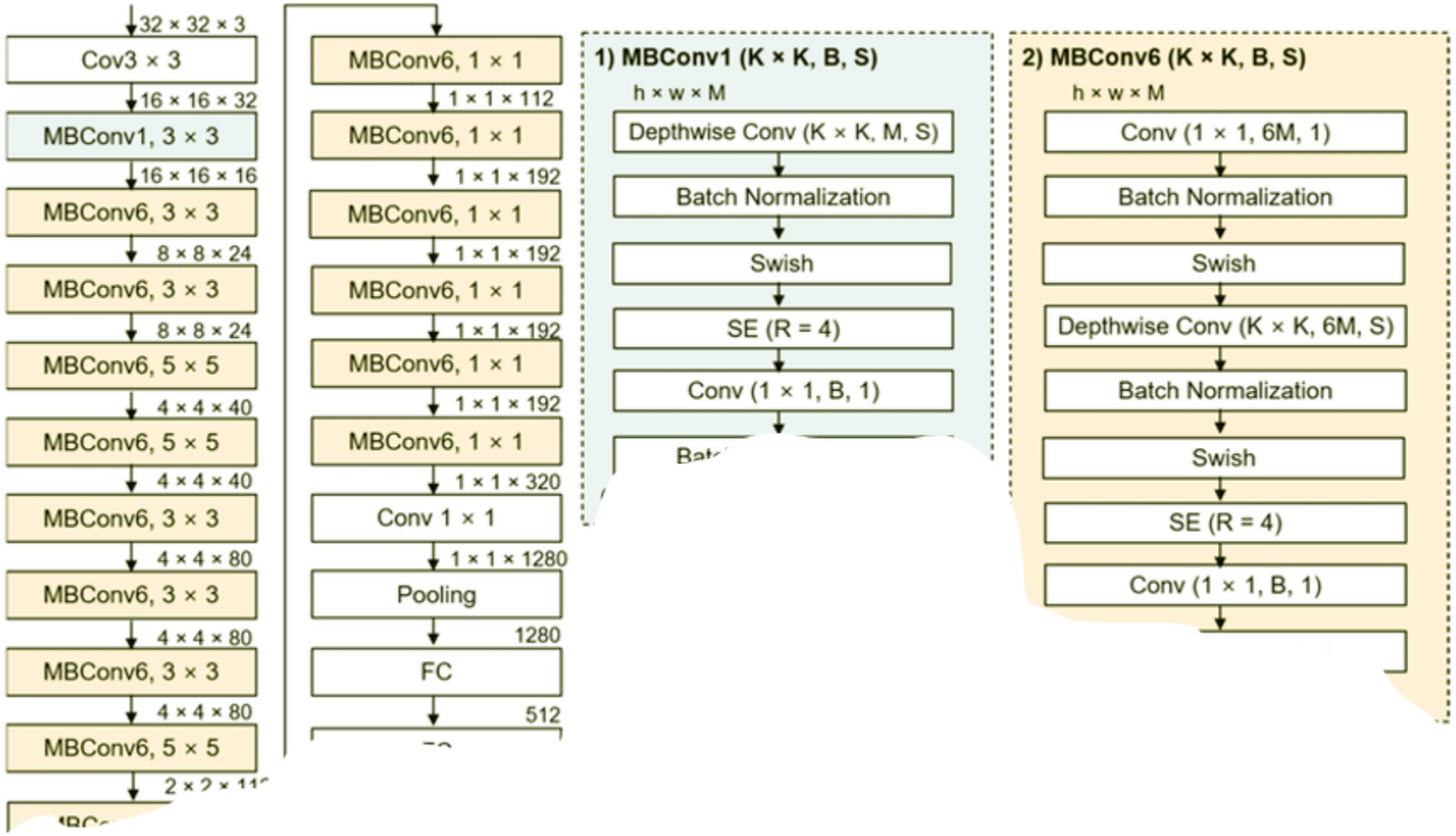
NAJWAŻNIEJSZE KOMPONENTY

- Serwer uruchamiania i synchronizacji modeli SI/ML/DL (Model Server)
- Moduł tworzenia i edycji reguł
- Rozproszona baza danych noSQL
- System tworzenia wizualizacji i raportów
- Repozytorium aktualizacji
- Moduł audytu i kontroli
- Graficzna konsola zarządzająca
- Feeds Service (przeszukuje podane źródła danych)
- Media Browser

PRZYKŁADOWE SIECI DL

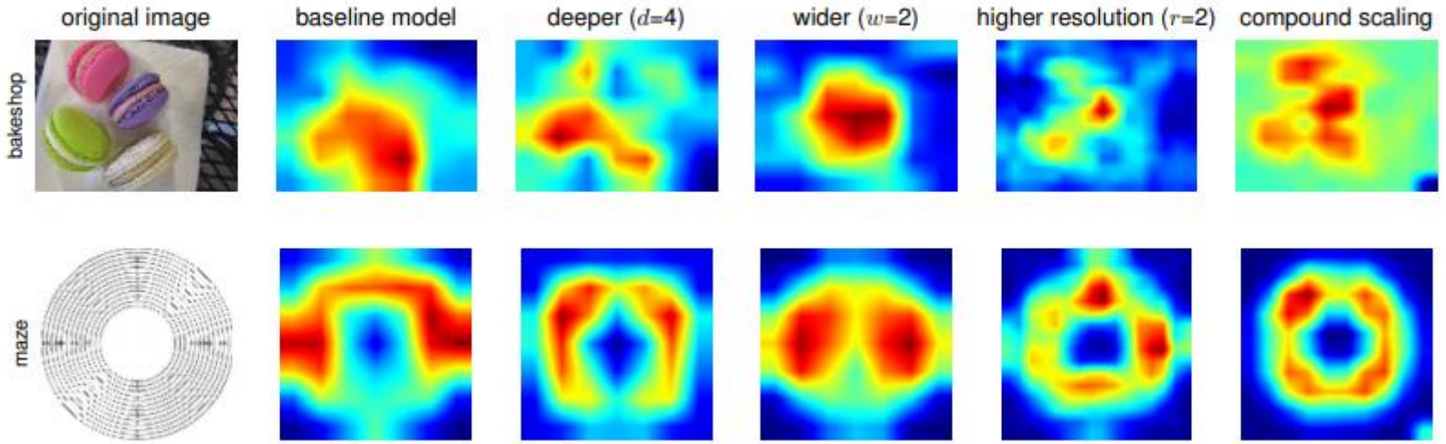


FRAGMENT NASZEJ SIECI COMPUTER VISION



SIECI KONWOLUCYJNE

- warstwy splotowe
- kernele - filtry
- przesuwanie kernela / filtra



PROCES BUDOWANIA SIECI DL/ML

- Określenie celu i miary sukcesu
- Wybór reprezentatywnych danych (próba losowa, odzwierciedlająca rzeczywistość, kwotowanie)
- Oczyszczenie i przygotowanie danych (puste, odstające, redukcja wymiarów, zmiana skali)
- Wczytanie danych
- Normalizacja danych
- Podział na zestawy uczące, testowe i walidacyjne
- Stworzenie architektury (warstwy, parametry-neurony, algorytmy)
- Dobór meta-parametrów, regularyzacja, optymalizacja
- Uczenie modelu
- Walidowanie modelu
- Udoskonalanie modelu

CIĄGŁE UDOSKONALANIE MODELI

Cześć,

Z mojej analizy false-positive'ów na produkcji z października i listopada wynika, że

Model aip_porn:
dobrze by było zaktualizować na produkcji do wersji 10 a 10-kę douczyć setami zawierającymi normale:
1. gołe noworodki/bobasy
2. ludzi w obcisłych spodniach z lycrą

Model aip_tobacco_alcohol:
douczyć setami zawierającymi normale:
1. mydło w płynie,żele do kąpieli,
2. perfumy we flakonach,
3. oliwa w butelkach
4. keczupy, sosy w buteleczkach,
5. dymy, mgły i obłoczki bez ludzkich twarzy

Model aip_wn:
zaktualizować na prod do wersji 28 a wersję 28 douczyć setami zawierającymi normale:
1. piłka ręczna do cięcia
2. kolorowe landrynki
3. kajdanki

Model autoagresja_gore:
zaktualizować na prod. do wersji 10
douczać wg mnie w tej chwili nie ma potrzeby

Jeżeli można prosić Ewę albo Zbyszka o pomoc w szybkim zebraniu datasetów w ww. kategoriach, to bnardzo proszę.

Chyba, że Andrzej powie, że któryś z nich jest niepotrzebny albo niemożliwy do nauczenia.

Dobrze by było to wszystko do jutra wieczór zebrać, co jest realne.

tego niestety nie dorzuce ....

MY NA TLE INNYCH ROZWIĄZAŃ

- Nasze, własne, opracowane przez nas modele z pełnymi prawami autorskimi
- Nasz własny, stabilny, sprawdzony system uruchomienia i koordynacji
- Całe know-how u nas
- Możliwa instalacja wyspowa, w chmurze publicznej, prywatnej i hybrydowej
- Pełna odpowiedzialność i ciągłość umów SLA
- Brak zależności od zewnętrznych dostawców i chmur
- Żadne dane, w tym wrażliwe, nie muszą być wysyłane poza organizację

MY NA TLE INNYCH ROZWIĄZAŃ

- Elastyczność, szybkie dostosowanie
- Szybkie tworzenie nowych, dedykowanych modeli
- Szybkie wdrożenie
- Elastyczne licencjonowanie (nie musi być zależne od woluminu danych)
- Ceny dostosowane do rynku

THANK YOU

DZIĘKUJĘ ZA UWAGĘ

www.aiseclab.com

AISECLAB