

Jak osiągać cele statystyczne ze sztuczną inteligencją AI – wybrane doświadczenia

Zmiany. Usprawnienia. Eksperymenty. Innowacje

Jolanta Szutkowska

Naczelnik Wydziału Rozwoju i Innowacji

Departament Handlu i Usług

Agenda

- 1. Wyzwania w obszarze statystyki handlu i usług**
- 2. Czym jest AI i w jaki sposób może pomóc?**
- 3. Zalety i wady zastosowania AI przy danych skanowanych w pomiarze inflacji CPI/HICP**
- 4. Zalety i wady zastosowania AI przy danych skrapowanych w pomiarze inflacji CPI/HICP**
- 5. Dlaczego myślenie krytyczne jest tak ważne w pracy z AI?**
- 6. Podsumowanie czyli jak osiągać cele statystyczne z AI i unikać pułapek**

Wyzwania w obszarze statystyki handlu i usług

1. **Turbulentna gospodarka (pandemia, agresja Rosji na Ukrainę, zmiany klimatyczne, klęski żywiołowe, inne nietypowe zjawiska)**
2. **Gospodarka cyfrowa, platformy cyfrowe - handel e-commerce, zrównowagony rozwój, gospodarka obiegu zamkniętego, transformacja energetyczna, globalizacja**
3. **Postęp technologiczny, przemysł 4.0, rozwój AI, big data w tym dane skanowane, dane skrapowane**
4. **Rosnące potrzeby interesariuszy statystyki publicznej i ich wpływ na produkcję statystyczną (ekosystem danych, dane otwarte, dane prywatne, rosnący popyt na dane terminowe, bardziej szczegółowe - w różnych przekrojach, dobrej jakości)**

Wyzwania w obszarze statystyki handlu i usług dok.

5. Wykorzystanie statystyki wieloźródłowej do pomiaru wskaźnika cen towarów i usług konsumpcyjnych CPI czyli popularnej miary inflacji

6. Wyzwania wobec nowych metod i sposobów pracy - modernizowanie systemu zbierania, przetwarzania i opracowywania tych danych

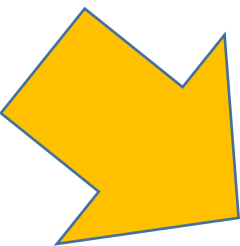
7. Wyzwania metodologiczne związane m.in. ze stosowaniem metod multilateralnych, tworzeniem nowych wag, sposobów agregacji i kalkulacji wskaźników cen

8. Rozwój statystyki eksperymentalnej

Czym jest AI i w jaki sposób może pomóc?

**ROZPORZĄDZENIE PARLAMENTU EUROPEJSKIEGO I RADY (UE)
2024/1689 z dnia 13 czerwca 2024 r.**

**w sprawie ustanowienia zharmonizowanych przepisów
dotyczących sztucznej inteligencji oraz zmiany
rozporządzeń.....**



„system AI” oznacza system maszynowy, który został zaprojektowany do działania z różnym poziomem autonomii po jego wdrożeniu oraz który może wykazywać zdolność adaptacji po jego wdrożeniu, a także który – na potrzeby wyraźnych lub dorozumianych celów – wnioskuje, jak generować na podstawie otrzymanych danych wejściowych wyniki, takie jak predykcje, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne

Czym jest AI i w jaki sposób może pomóc? cd.

- **Ogromny, dynamiczny, różnorodny wolumen nieustrukturyzowanych danych skanowanych i skrapowanych**
- **Dostęp do danych w czasie rzeczywistym – economy based on data versus economy driven by data**
- **Kwestie interoperacyjności i dostosowania alternatywnych źródeł danych do celów statystycznych**
- **Ocena jakości danych na wejściu i danych na wyjściu, tajemnica statystyczna, tajemnica biznesowa, handlowa, zarządzanie ryzykiem, zasada partnerskiej współpracy sektora publicznego i prywatnego oparta na etyce i fair play**

Czym jest AI i w jaki sposób może pomóc? dok.

- **Relacja: statystyk versus AI :** 1.) inspirowanie i wspieranie; 2.) zastępowanie w pracach powtarzalnych, monotonicznych i uciążliwych, 3) kontrola prac AI i zapobieganie w tworzeniu się obszarów do automatycznego, niezależnego działania bez nadzoru statystyka,
- **Przykłady zastosowań:** a) analizy danych; b) tworzenie klastrów dla danych tzw. grup homogenicznych według zadanych kryteriów z wykorzystaniem machine learning i text mining; c) walidacja danych; d) integracja danych z różnych źródeł dla rozwoju statystyk wieloźródłowych w tym przetwarzanie big data i sprawdzanie ich interoperacyjności; e) kalkulacja nowych wskaźników statystycznych; f) tworzenie projektów publikacji statystycznych i wizualizacji; g) zasilanie baz danych po ich walidacji; h) zwiększanie optymalizacji i efektywności działania, redukcja obciążenia respondentów i kosztów badań statystycznych

Zalety i wady zastosowania AI przy danych skanowanych w pomiarze inflacji CPI/HICP

Przez dane skanowane (ang. scanner data) należy rozumieć szczegółowe dane o dobrach konsumpcyjnych uzyskane dzięki skanowaniu ich kodów kreskowych w punktach sprzedaży

Zalety 😊

- ☐ Dostęp w czasie rzeczywistym do ogromnego wolumenu różnorodnych danych o produkcie m.in. kod punktu sprzedaży, etykieta opisu produktu, jednostka miary
- ☐ Lepsze rozpoznanie sytuacji w danym segmencie rynku
- ☐ Wzbogacanie metodologii badania pod kątem np. wyboru formuł indeksów cenowych czy systemu wag

Wady ☹️

- ☐ Dynamiczny charakter zbiorów – konieczność stosowania szybkich algorytmów i dodatkowych serwerów, tworzenie nowych zbiorów do uczenia maszynowego
- ☐ Odrębne oprogramowanie dla modułu do kalkulacji wskaźników cen
- ☐ Dylematy - filtrowanie danych a reprezentatywność próby, ujęcie produktów sezonowych

Zalety i wady zastosowania AI przy danych skrapowanych w pomiarze inflacji CPI/HICP

Skrapowanie danych (*Web scraping*= zdrapywanie lub skrobanie stron - to zautomatyzowany proces pozyskiwania danych ze stron internetowych, najczęściej przy wykorzystaniu specjalnego programu czy skryptu (zwanego skraperem), który jest w stanie naśladować zachowanie człowieka przeglądającego stronę internetową.

Zalety 😊

- ☐ Duża ilość danych dostępnych w danym momencie
- ☐ Dane oparte na faktach, eliminowanie „fake newsów”
- ☐ Doskonalenie kwalifikacji pracowników m.in. dostosowanie platformy i nadzór nad zmianami w strukturze stron internetowych

Wady ☹️

- ☐ Dane skrapowane często nieujednoliczone, nieuporządkowane, niewystandaryzowane i niekompletne
- ☐ Nadmierne lub niedostateczne pokrycie produktów i ceny ofertowe zamiast cen transakcyjnych
- ☐ Duża rotacja i krótkotrwałość dostępu produktów

Zalety i wady zastosowania AI przy danych skrapowanych w pomiarze inflacji CPI/HICP dok.

Zalety 😊

- ☐ Doskonalenie metodologii badania m.in. metody integracji i agregacji danych, usuwanie niespójności metodologicznych i jakościowych
- ☐ Doskonalenie metod multilateralnych przy kalkulacji wskaźników cen
- ☐ Wdrażanie netetykiety

Wady 😞

- ☐ Przerwy w dostępie do Internetu, błędy w działaniu systemów IT, wirusowe oprogramowanie
- ☐ Luki w szeregach czasowych danych, przeprogramowania skryptów, ograniczenia w porównywaniu danych liczonych na bazie źródeł tradycyjnych a tych opartych na źródłach internetowych
- ☐ Kwestie prawne versus rozwiązania metodyczno-organizacyjne

Dlaczego myślenie krytyczne jest tak ważne w pracy z AI?

- Weryfikacja i kontrola prac AI – sprawdzenie działania algorytmu, identyfikacja znaczących związków przyczynowo-skutkowych (korelacji), eliminowanie „fake newsów” i dezinformacji
- Eliminowanie rozwiązań AI, których się nie rozumie
- Traktowanie AI jako wsparcie a nie wyparcie czynnika ludzkiego
- Wykorzystywanie AI do optymalizacji procesów i poprawy efektywności pracy w zakresie interoperacyjności, harmonizacji, standaryzacji w tworzeniu tzw. ramki danych oraz jako źródło inspiracji
- Zarządzanie ryzykiem w pracy z AI

Dla eliminowania pułapek AI kluczowe są:

KRYTYCZNE MYŚLENIE

+

KOMPETENCJE+ WIEDZA+ WIZJA

Podsumowanie czyli jak osiągać cele statystyczne z AI i unikać pułapek

- **Decydująca rola czynnika ludzkiego w opracowaniu danych z udziałem innowacyjnych technik wykorzystujących sztuczną inteligencję – zachowanie wymogów metodologicznych jak i jakościowych (wytłumaczalność/interpretacja, dokładność, powtarzalność/odtwarzalność, terminowość, opłacalność, niezawodność/ stabilność)**
- **Krytyczne myślenie statystyków - przewycięzanie pułapek-halucynacji AI i wpływu bańki informacyjnej, nieumiejętnego formułowania promptów i niewystarczającego wytrenowania modeli: sprawdzanie wyliczeń AI z danymi z innych źródeł, wykonywanie powtórnych wyliczeń, rezygnacja z kalkulacji danych w oparciu o AI, których się nie rozumie i które trudno wytłumaczyć**

Podsumowanie czyli jak osiągać cele statystyczne z AI i unikać pułapek cd.

- **Korzystanie z innowacyjnych technik AI**
 - **stosowanie metod i technik uczenia maszynowego oraz text mining przy klasyfikowaniu produktów i rozpoznawaniu tekstu przy skanowaniu i skrapowaniu danych**
 - **szybkie do wykonania algorytmy i korzystanie z dodatkowych serwerów przy filtrowaniu danych, dopasowywaniu, klasyfikowaniu czy obliczaniu indeksów cen**
 - **tworzenie autorskiego pakietu w środowisku R – odrębne oprogramowanie do obliczania wskaźników cen**
- **Wykorzystywanie dobrych praktyk z innych krajów**

Podsumowanie czyli jak osiągać cele statystyczne z AI i unikać pułapek cd.

- **Wyzwania organizacyjno-metodologiczne i prawne:**
 - **potrzeba efektywnej, partnerskiej współpracy z gestorem danych (zawieranie porozumień, umów, zagwarantowanie długofalowego i terminowego dostępu do danych, informowanie i uzgodnienia dotyczące nagłej, częstej zmiany kodowania i etykietowania produktów z uwagi na konieczność przygotowania nowych zbiorów uczących dla metod uczenia maszynowego w celu poprawnego mapowania produktów i zapewnienia ich porównywalności oraz spójności, stworzenie kodeksu etycznego, kodeksu dobrych praktyk)**

Podsumowanie czyli jak osiągać cele statystyczne z AI i unikać pułapek cd.

- **Wyzwania organizacyjno-metodologiczne i prawne dok. :**
 - **optymalizacja procesów i próby (niedostateczna reprezentatywność versus nadreprezentatywność),**
 - **efektywność i skalowalność opracowywanych narzędzi i systemów,**
 - **zmiany w metodach agregacji danych niezbędne przy integracji danych ze źródeł tradycyjnych ze źródłami typu big data,**
 - **odchodzenie od indeksów bilateralnych w kierunku indeksów multilateralnych.**

Podsumowanie czyli jak osiągać cele statystyczne z AI i unikać pułapek cd.

Dwa limeryki o AI w statystyce publicznej

*Ma statystyk orzech twardy,
Bo AI rozdaje karty
I z big data robi Kosmos.
Cóż, niełatwo jest i prosto
Choć jest learning i text mining.*

*Ciągle wabi, czar swój wkręca
Ta sztuczna inteligencja.
Daje w statystyce popis
I eksperymentów opis
A I TAK CÓŻ..... JEST BEZ SERCA*

Podsumowanie czyli jak osiągać cele statystyczne z AI i unikać pułapek dok.

Pomocna literatura:

- ❖ Prezentacja „Artificial Intelligence – a buzz, must or bust for official Statistics”, Mariana Kotzeva, Director General, Eurostat, IV Congress of Polish Statistics dostępna pod linkiem:
<https://kongres2024.stat.gov.pl/>
- ❖ Opracowanie „Nowoczesne technologie i nowe źródła danych w pomiarze inflacji” GUS, Warszawa 2022 dostępne pod linkiem:
<https://stat.gov.pl/obszary-tematyczne/inne-opracowania/inne-opracowania-zbiorcze/nowoczesne-technologie-i-nowe-zrodla-danych-w-pomiarze-inflacji,47,1.html>
- ❖ Artykuł Arzu Taghiyeva „The role of Artificial Intelligence when generating official statistical data. Rola sztucznej inteligencji w generowaniu oficjalnych danych statystycznych”, Wiadomości Statystyczne. The Polish Statistician, 2024, 69(10), 75–88, link:
<https://doi.org/10.59139/ws.2024.10.5>

***Ze StaTYsTYką na TY
Możesz być nawet i TY!
A z AI w StaTYsTYce...?
Cóż - to nowe przeżycie 😊***

Dziękuję za uwagę