

NYELVI MODELLEK A PÉNZÜGYI MODELLEZÉS SZOLGÁLATÁBAN

HTE Mesterséges Intelligencia
minikonferencia, 2024



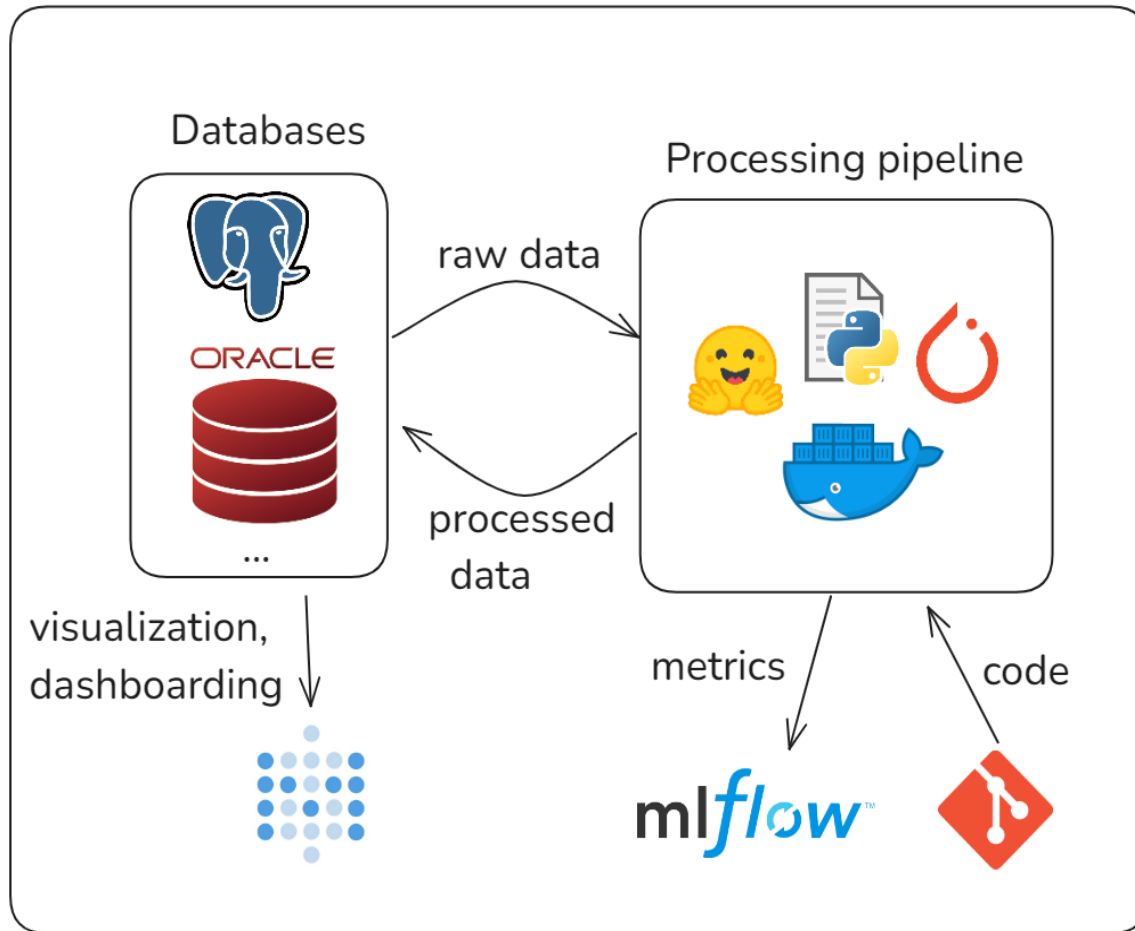
*100 éve Magyarország
gyarapodásáért*



Ónozó Livia

Digitalizációs technológia főosztályvezető

ML environment



Jól ismert technológiákra építünk

Népszerű, ipari használatban helytálló komponensek felhasználásával készült környezet.

Moduláris

A komponensek sztenderdizált interfészeken keresztül kommunikálnak és funkcionalitás mentén vannak elválasztva.

Újrafelhasználható

Több különböző projekt igényeit is képes kiszolgálni ugyanaz az infrastruktúra.



*100 éve Magyarország
gyarapodásáért*

1. FEJEZET

ONLINE PÉNZTÁRGÉP ADATOK AZ ÁRVÁLTÓZÁSOK BECSLÉSÉRE

Első körös adatátadás

- ~ 1 millió egyedi terméknév
- ~ 88 ezer vámtarifaszámmal rendelkezik (VTSZ)

Korlátozottan hozzáférhető

Adatgenerálás

- Szabály alapú
- Mélytanulás alapú (ChatGPT)

Cél:

Termékek kategorizálása

BIZONYLAT MÁSOLAT

LIDL MAGYARORSZÁG BT.
1037 BP. RÁDL ÁROK 6.
ÜZLET: 0316

1027 Budapest, Csalogány u. 43.
ADÓSZÁM: 21588017-2-44

	Ft
DESSZ. KEF. BAR-M	199 Ft
CSIRKE ÉDES-SAV(3)	699 Ft
DREHER 24 ALKOHOLM	209 Ft
MÁKOS GUBA	149 Ft
OLÍVÁS STANGLI	75 Ft

FIZETENDŐ -----
1 331

=====

BANKKÁRTYA 1 331

TID: HH331603 2021.07.15 09:24
MTD / OPTD - 0501547526 / 0000

VÁMTARIFASZÁM NÓMENKLATÚRA

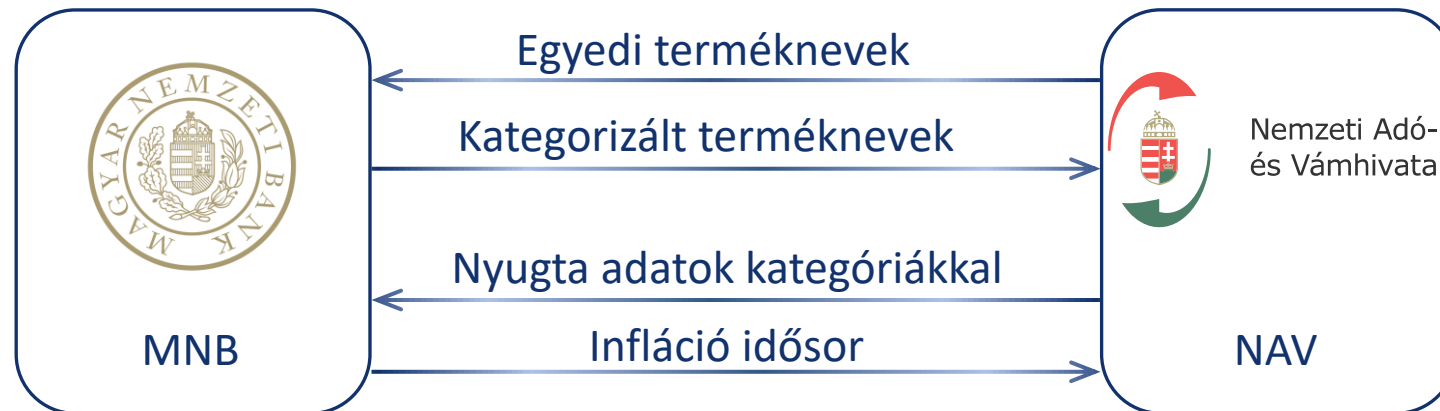


- Hierarchikus kategória rendszer
- COICOP (Classification of Individual Consumption by Purpose) osztályokba transzformálható
- Kiskereskedelmi egységek
- Infláció méréséhez:
mennyiségek + árak + COICOP osztályozás

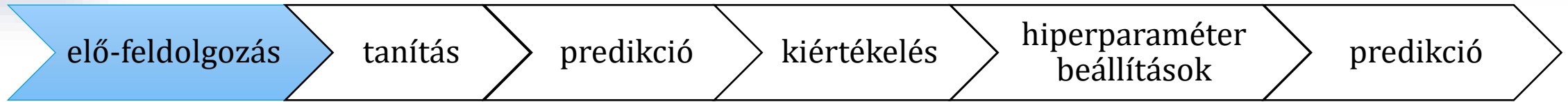
Árucsoport:	NÖVÉNYI TERMÉKEK
Árucsoport:	Gabonafélék
Szimulációs dátum:	2023.10.12
Taric Kód	Leírás
1001	Búza és kétszeres
1002	Rozs
1003	Árpa
1004	Zab
1005	Kukorica
1006	Rizs
1007	Cirokmag
1008	Hajdina, köles és kanárimag; más gabonaféle

[TARIC Web \(gov.hu\)](https://taric.web.gov.hu)

Két lépcsős adatátadás:



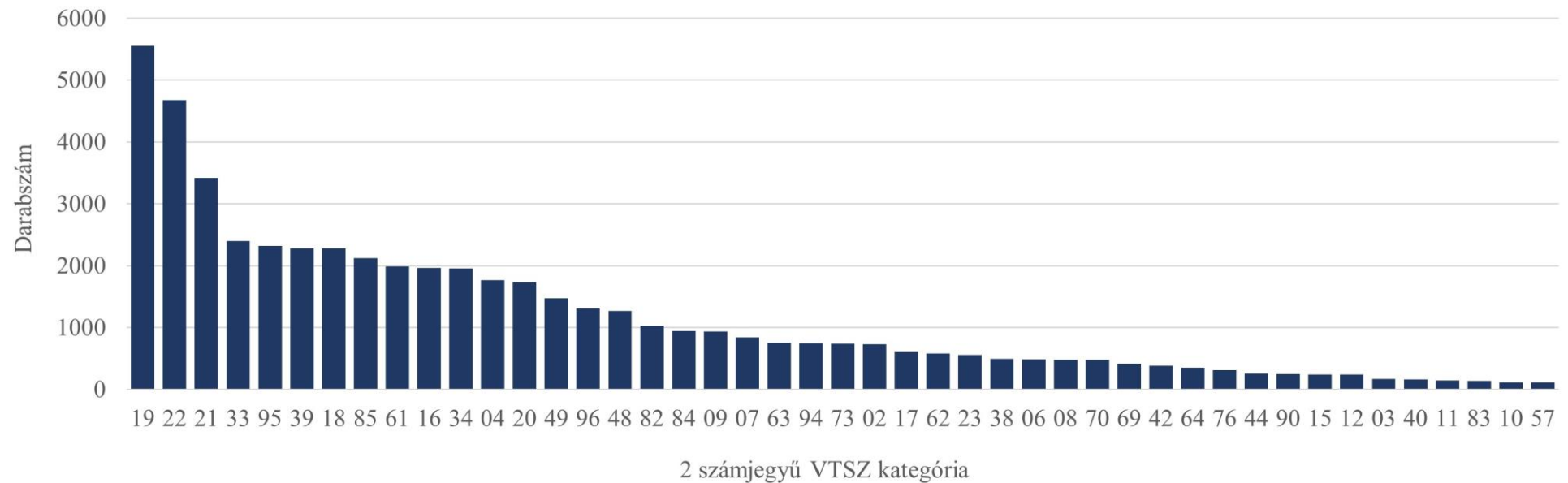
SOK KIS ELEMSZÁMÚ VTSZ 2 KATEGÓRIA



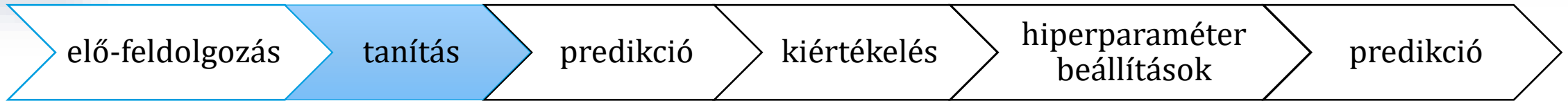
Adattisztítás szerepe

- Nyers adat minősége: hiányzó értékek, duplumok, zaj
- Ritka elemszámú kategóriák, adatgenerálás, más adatbázisok bevonása

Megfelelő tréning-validáció-teszt adat felbontás



KÉT TÍPUSÚ NYELVI MODELLEZÉS

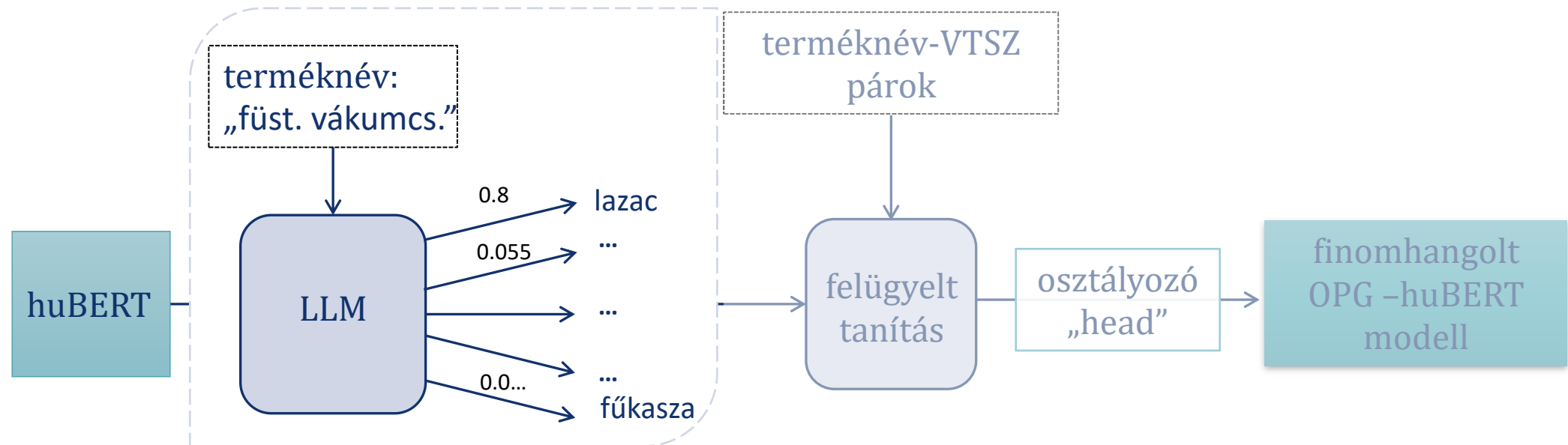


Előtanított nyelvi modell finomhangolása osztályozásra:

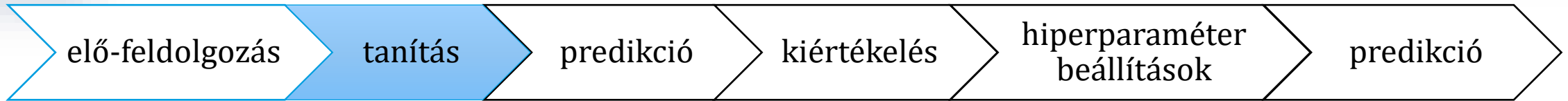
- termék-kategória párokon tanítunk (felügyelt)

Előtanított nyelvi modell tovább tanítása OPG adatokon: „saját” nyelvi modell

- termékneveken tanítunk (felügyelet nélküli)



KÉT TÍPUSÚ NYELVI MODELLEZÉS

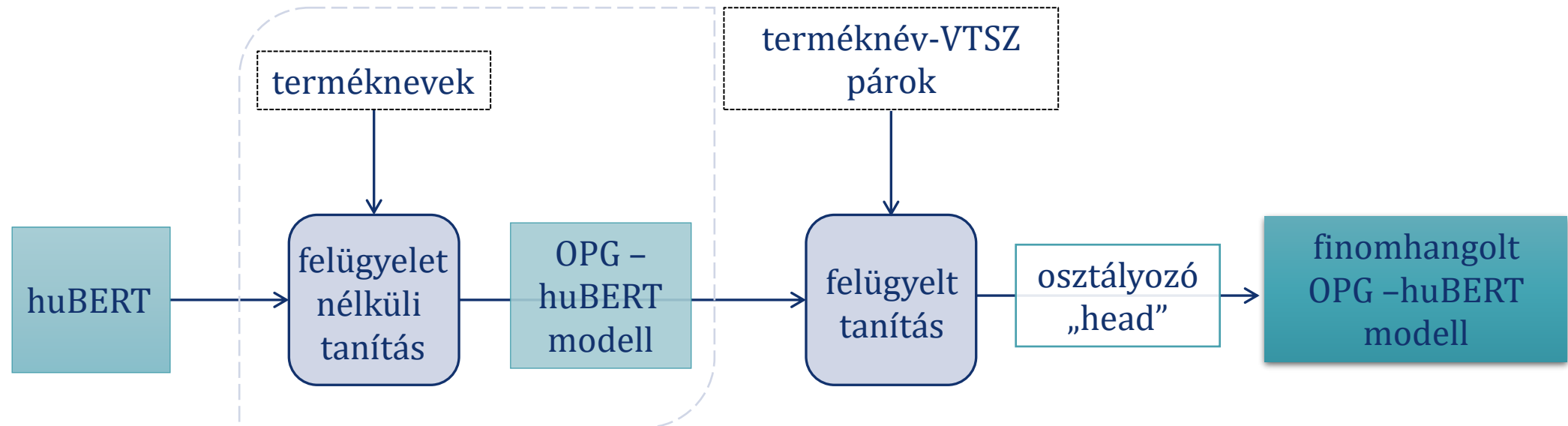


Előtanított nyelvi modell finomhangolása osztályozásra:

- termék-kategória párokon tanítunk

Előtanított nyelvi modell tovább tanítása OPG adatokon: „saját” nyelvi modell

- termékneveken tanítunk



TÉVESZTÉSI MÁTRIX



Predikció:

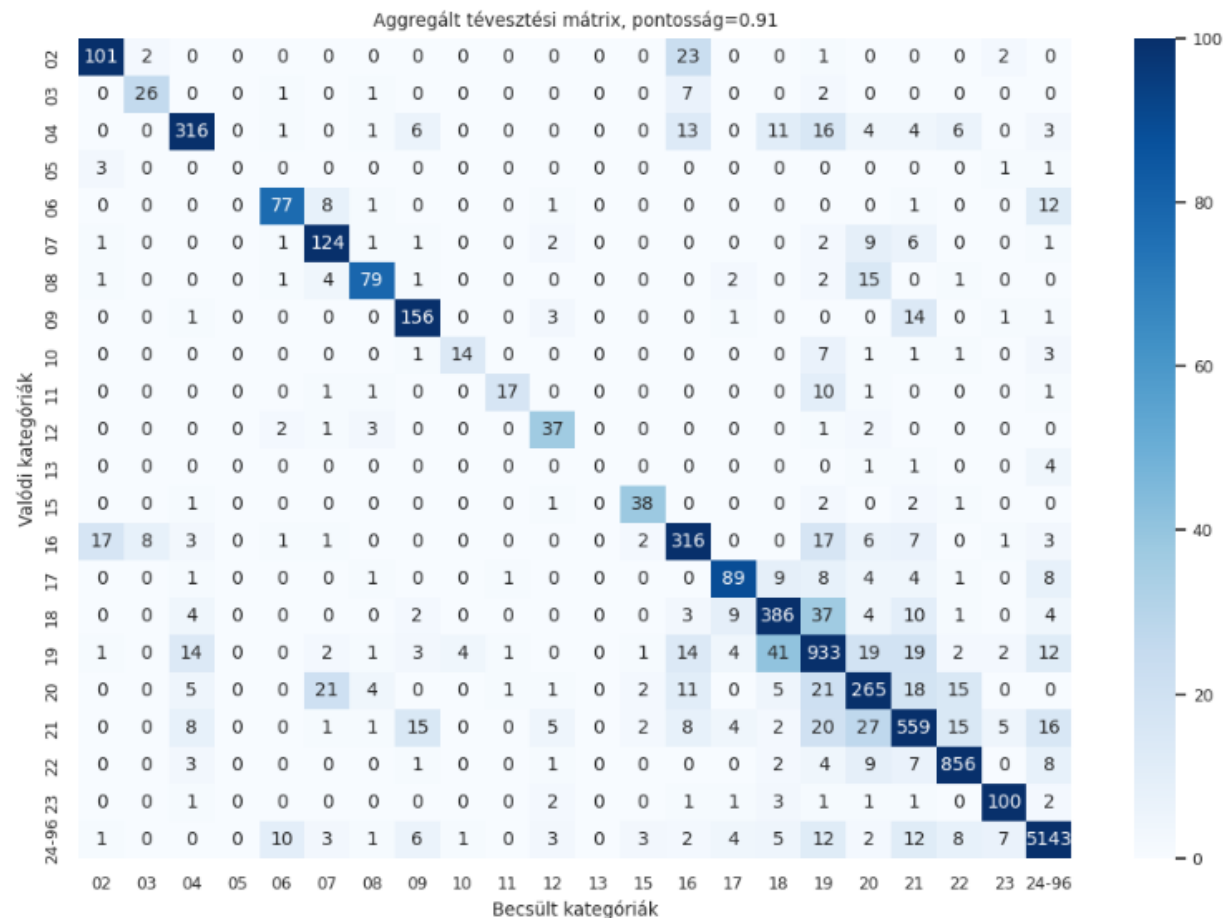
olyan termékek kategorizálása 2 és 4 számjegyű
vámtarifaszám kategóriákba, amiket még nem látott a modell

Kiértékelés: címkézett teszt adatokon

Accuracy, precision, recall, F1

Eredmények átlagolása:

a nem egyenletes eloszlású osztályok kezelése →
gyakorisággal súlyozott átlag



INTERAKTÍV DASHBOARD



elő-feldolgozás

tanítás

predikció

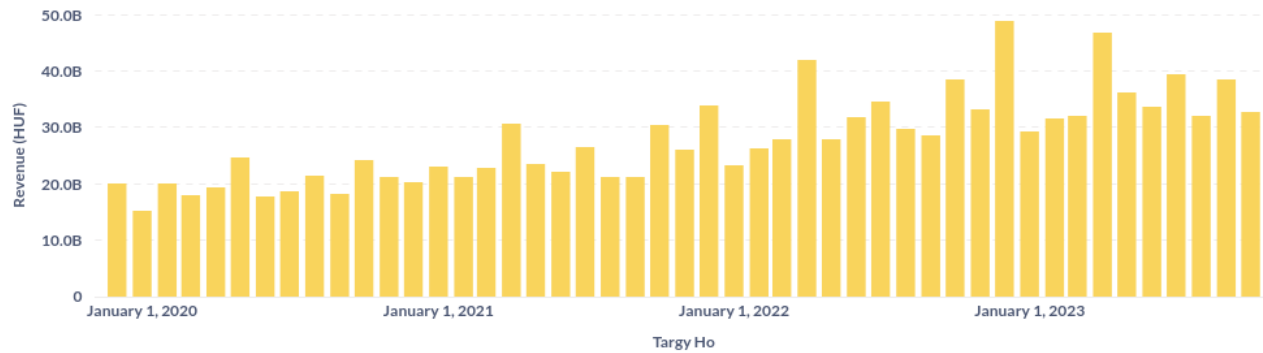
kiértékelés

hiperparaméter
beállítások

inferencia

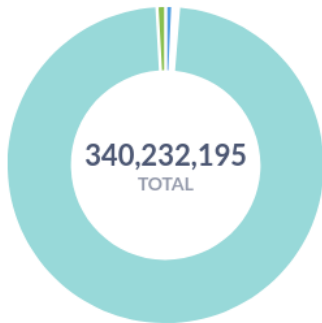
Macro Data

Monthly Revenue



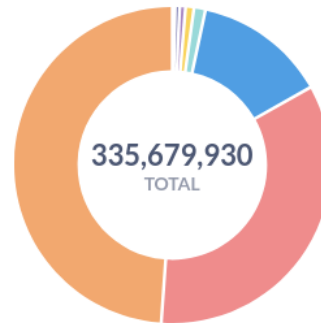
Bill types

- Visszaru bizonylat
- Sztornó bizonylat
- Speciális értékesítés bizo...
- Nyugta
- Egyszerűsített számla



Payment types

- Vegyes
- Készpénzes+Kártyás
- Kártyás+Egyéb
- Készpénzes+Egyéb
- Ismeretlen
- Egyéb
- Kártyás
- Készpénzes



VTSZ category *

04

VTSZ data

Tejtermékek; madártojás; természetes méz; másutt nem említett élelmiszeri alkalmas állati eredetű élelmiszer
Vtsz Desc lookup

Monthly revenue+trend+sa by VTSZ

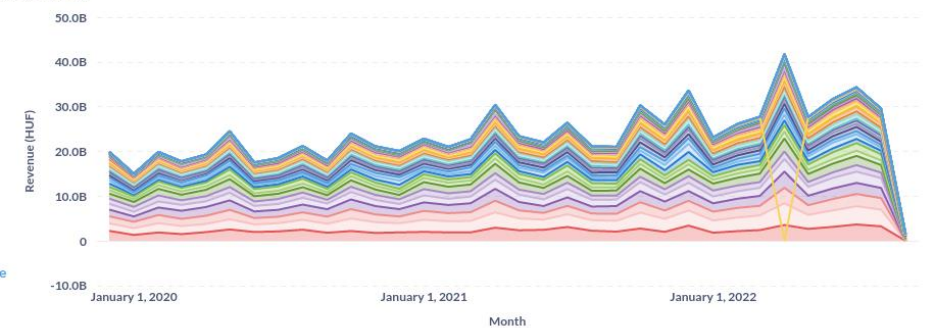
revenue revenue_trend revenue_sa



Monthly Revenue By Vtsz

- 67
- 56
- 66
- 30
- 29
- 65
- 71
- 83
- 91
- 36
- 31

And 49 more





*100 éve Magyarország
gyarapodásáért*

2. FEJEZET

TERMÉSZETES

NYELVFELDOLGOZÁS A GAZDASÁGI

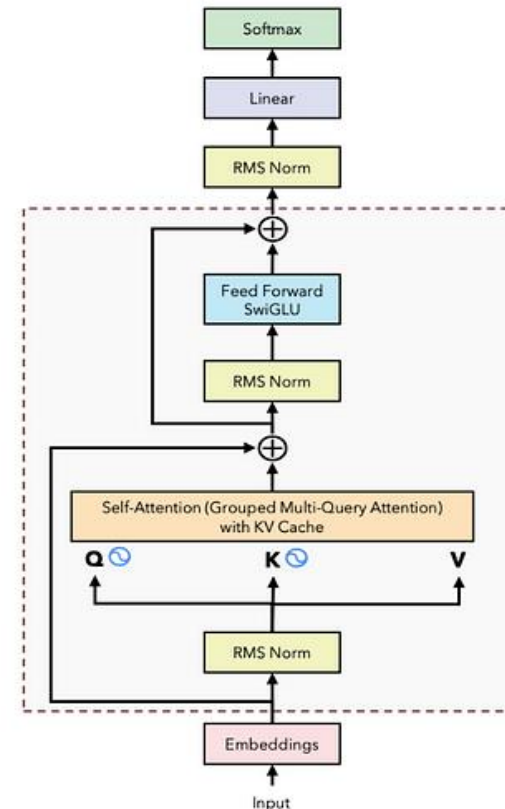
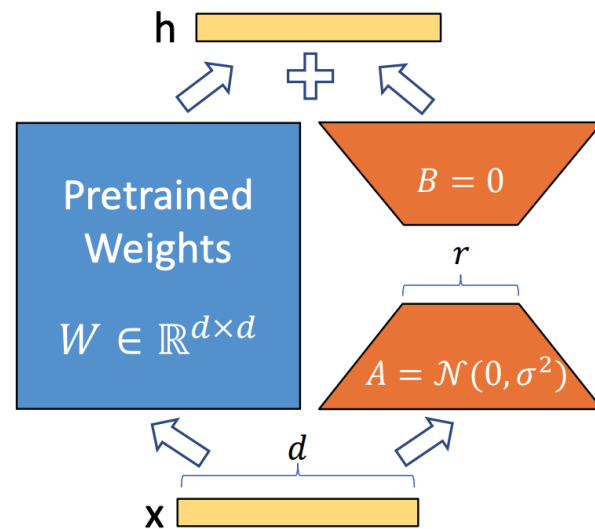
TELJESÍTMÉNY MÉRÉSÉRE

TRANSFORMER: LLAMA3 + QLORA

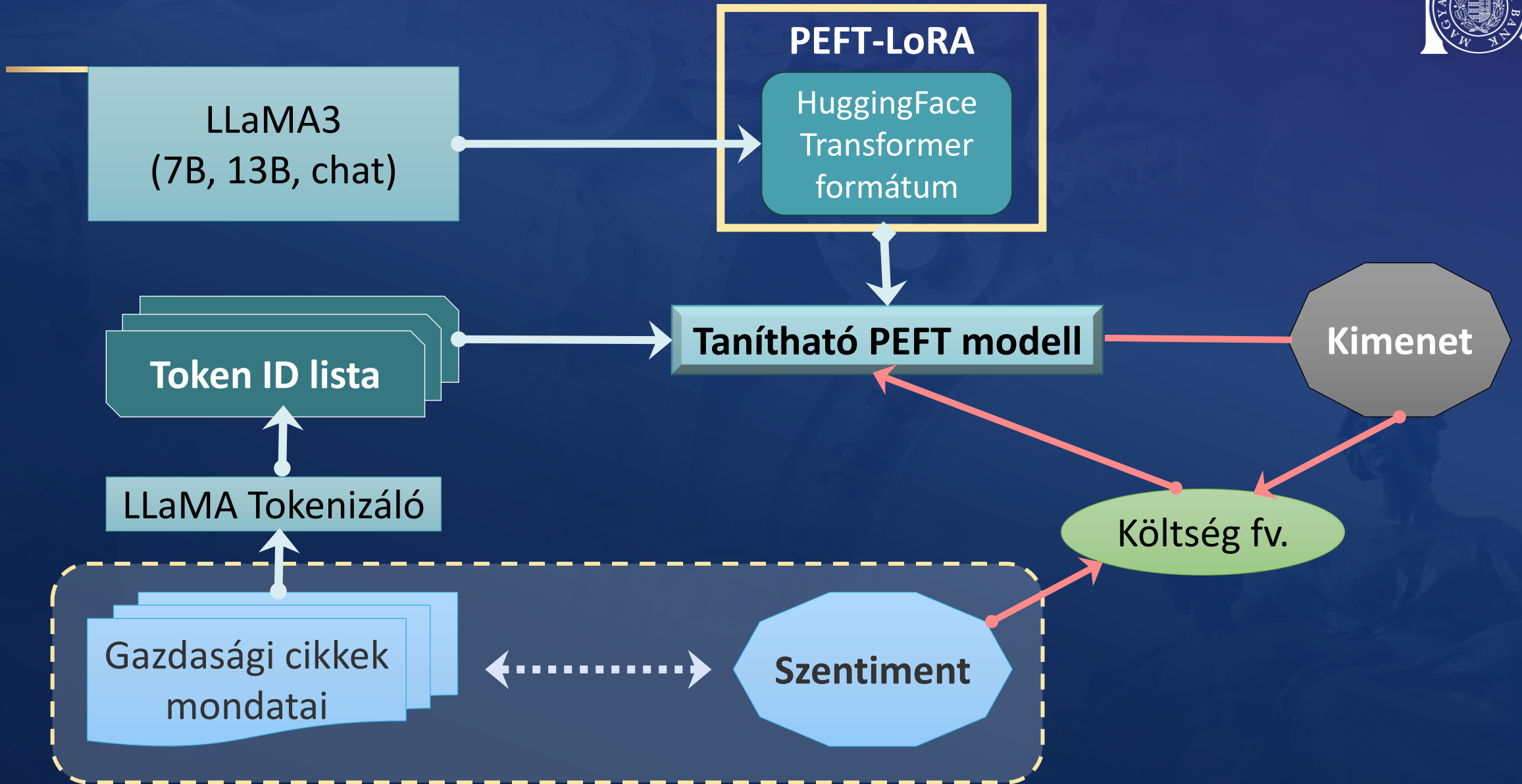


„ÉRTÉKELÉSÜK SZERINT A HEVES PIACI REAKCIÓ IS ENNEK VOLT BETUDHATÓ.”

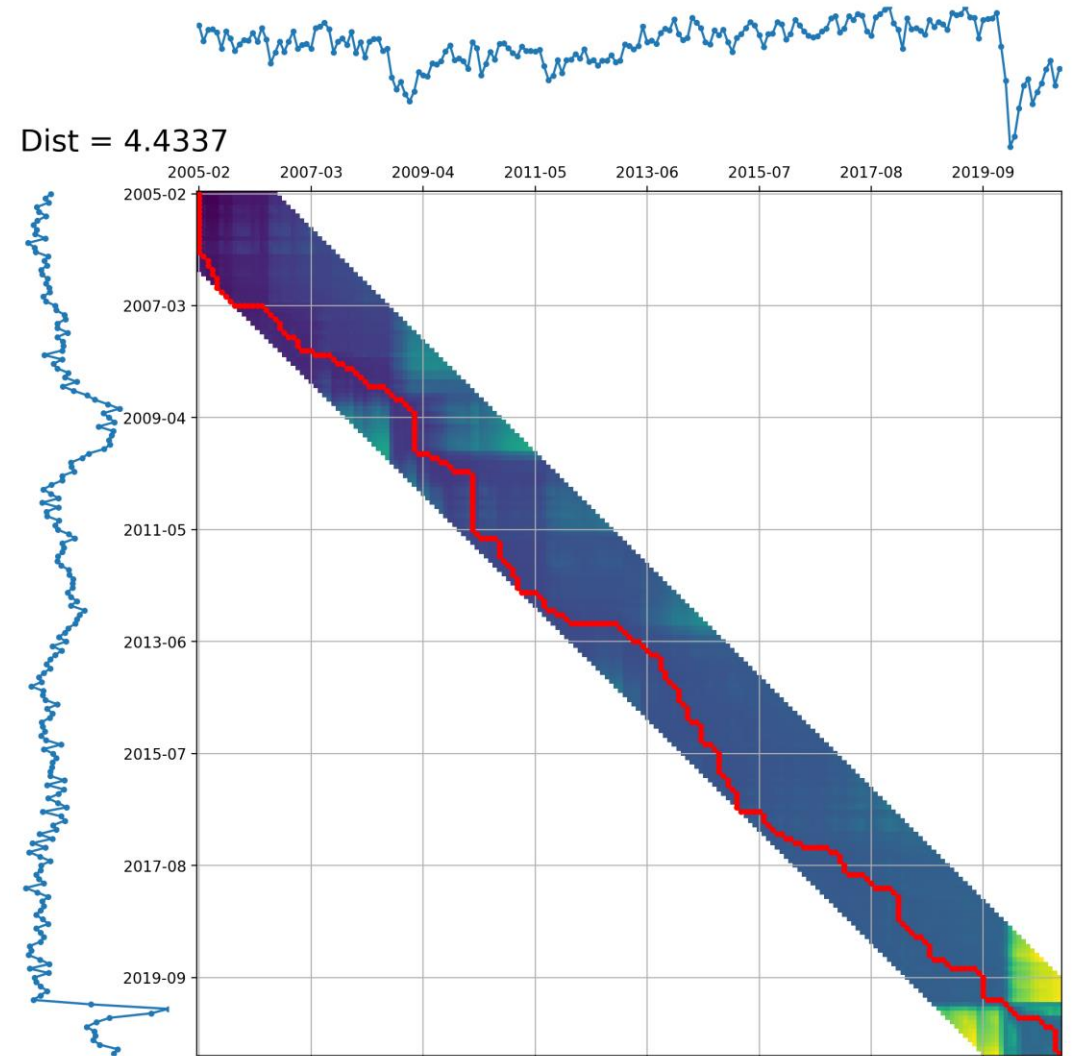
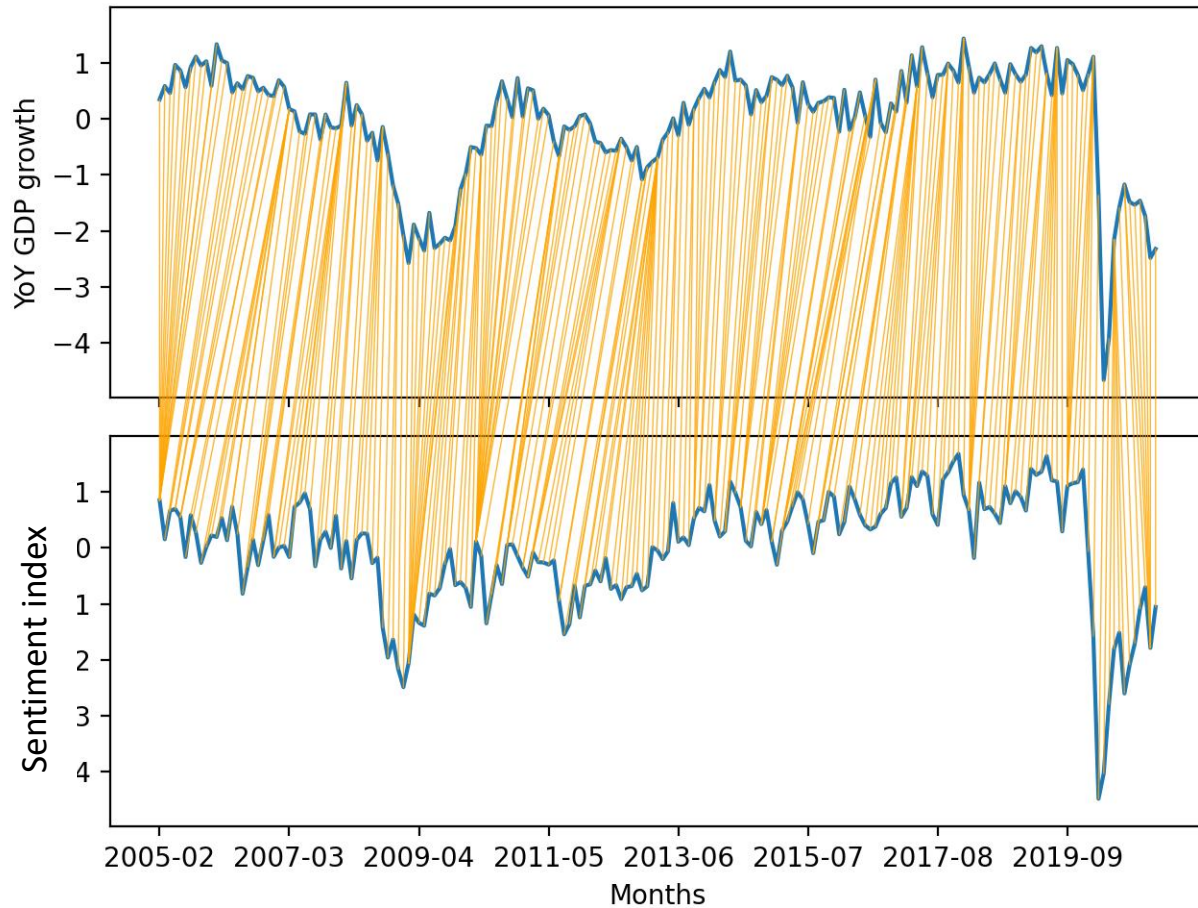
TOKEN_IDS': TENSOR([[1, 2715, 2163, ..., 3945, 2, ...]])

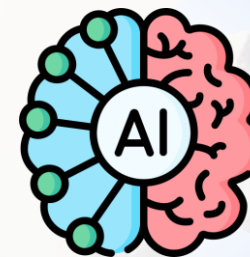
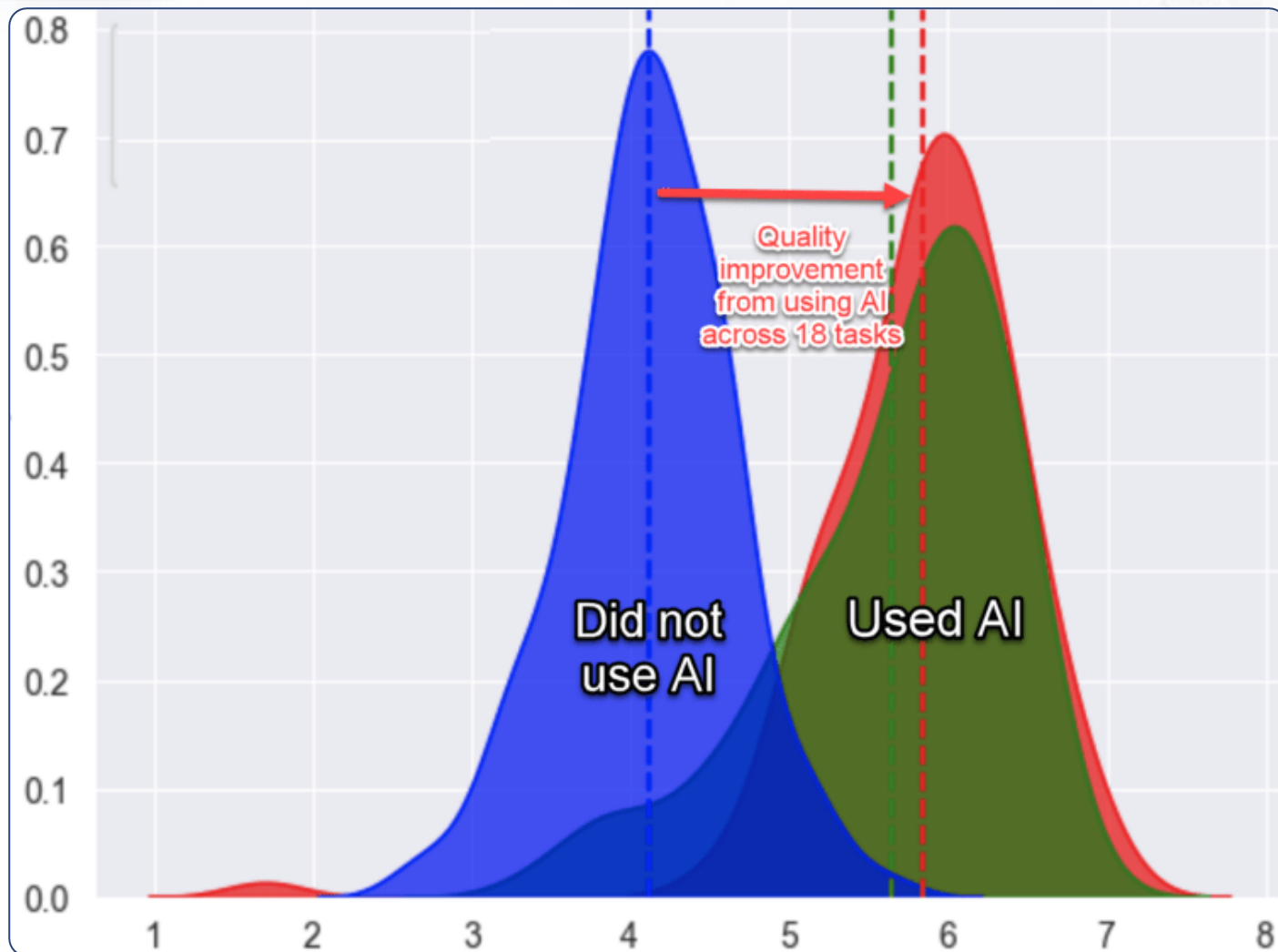


» NEGATÍV



$$DTW_q(x, x') = \min_{\pi \in \mathcal{A}(x, x')} \left(\sum_{(i, j) \in \pi} d(x_i, x'_j)^q \right)^{\frac{1}{q}}$$





Több feladatot

Gyorsabban

Jobb minőségben

KÖSZÖNÖM A MEGTISZTELŐ FIGYELMET!

onozol@mnb.hu



*100 éve Magyarország
gyarapodásáért*