

Analýza ekonomických údajov s využitím platformy Microsoft Azure

Autor: Mgr. Roderik Antol · Školiteľ: doc. RNDr. Eugen Ružický, CSc.



PANEURÓPSKA VYSOKÁ ŠKOLA

Fakulta informatiky

Motivácia

Objem dát každoročne rastie a zvyšuje nároky na informačné systémy. Analytické riešenia však bývajú budované po častiach:

- nástroje self-service BI sa používajú izolovane, bez integrácie dátových procesov a automatizácie,
- cloudové dátové riešenia zostávajú v literatúre prevažne na konceptuálnej úrovni, bez end-to-end implementácie,
- analýza sentimentu je spracovaná oddelene od kvantitatívnych dát.

Chýba tak ucelený, prakticky overený rámec, ktorý tieto časti spojí do jedného celku.

Ciele práce

- **Navrhnuť a implementovať dátový tok** na spracovanie štruktúrovaných dát verejných financií v prostredí Microsoft Azure.
- **Vytvoriť interaktívny BI dashboard** v nástroji Power BI pre exploratívnu analýzu fiškálnych ukazovateľov naprieč krajinami.
- **Zakomponovať analýzu sentimentu** textových dát ako doplnkovú analytickú dimenziu ku kvantitatívnym ukazovateľom.
- **Demonštrovať použiteľnosť** navrhnutého rámca na prípadovej štúdii verejných financií krajín V4 a Rakúska.

Prípadová štúdia

Rámec je overený na verejných financiách krajín V4 a Rakúska. Deficit, dlh, dane či výdavky štátu sa dotýkajú každodenného života a riešenie ich sprístupňuje v jednom automaticky aktualizovanom dashboarde.

Navrhnuté riešenie

ZDROJOVÉ SYSTÉMY

Eurostat – 10 datasetov cez REST API (SDMX 3.0)
Európska komisia – ekonomické prognózy (scraping webu)
Európska centrálna banka – tlačové správy (RSS)



ORCHESTRÁCIA A VÝPOČTY

Azure Data Factory – parametrizované pipeline,
aktivity Lookup / ForEach / Switch, Data Flow, monitoring
Azure Batch – vlastné Python skripty pre text a LLM



DÁTOVÁ VRSTVA – ADLS Gen2 (medailónová architektúra)

BRONZE

surové dáta,
historizácia
TSV / JSONL

SILVER

čistenie, unpivot,
dátová kvalita,
Parquet

GOLD

biznis logika,
validácie, dáta
pre Power BI



PREZENTAČNÁ VRSTVA – Microsoft Power BI

viacstranový interaktívny dashboard nad zlatou vrstvou;
transformačná logika je oddelená od vizualizačnej

Analýza sentimentu pomocou LLM

- Modely **gpt-4o-mini** a **GPT-5.5** namiesto lexikónových metód – lepšie zachytia kontext formálnych inštitucionálnych textov.
- Texty sa normalizujú a delia na segmenty (chunks); model vracia štruktúrovaný JSON so skóre v škále -2 až +2, mierou istoty a odôvodnením.
- Agregácia: vážený priemer segmentov podľa istoty modelu → skóre dokumentu → priemer za mesiac / rok.

Kľúčové výsledky

- **Funkčné end-to-end riešenie** – 10 datasetov Eurostatu, ekonomické prognózy a tlačové správy sa spracúvajú automaticky jedným parametrizovaným tokom na mesačnej báze.
- **Interaktívny Power BI dashboard** pokrývajúcí fiškálnu pozíciu, príjmy a výdavky, zahraničný obchod, štruktúru daní aj výdavkov (COFOG) a sentiment.
- **Vplyv pandémie:** deficit klesol až na -16,4 % HDP (PL), -15,3 % (HU), -9,2 % (ČR) a -8,7 % (SR). Nárast spôsobili zvýšené výdavky, nie prepád príjmov.
- **Zahraničný obchod SR:** 79,4 % exportu smeruje do EÚ; automobily tvoria 31,7 % a elektrotechnika ~18 % celkového vývozu.
- **Dane a výdavky:** daňové zaťaženie SR 35,8 % HDP oproti 43,7 % v Rakúsku (2024); na sociálnu ochranu ide 36,4 % výdavkov SR (ČR 30,7 %, rok 2023).
- **Sentiment prognóz EK:** prepád na -0,61 v roku 2020, dlhodobý priemer 0,1. Sentiment reagoval na udalosti skôr než samotné fiškálne ukazovatele.

Prínos a využitie

- Rozšíriteľný rámec – nový štruktúrovaný zdroj stačí pridať do konfiguračného JSON súboru.
- Sentimentová analýza nemusí byť izolovanou úlohou, ale súčasťou dátového toku.
- Postup je prenositeľný aj na iné domény – verejnú správu či komerčnú sféru.
- Bezpečnosť bez správy hesiel: spravované identity a role (IAM) medzi službami Azure.