

Hírek, események News, events

Kísérleti statisztikák és innováció a Központi Statisztikai Hivatalban – Sajtóreggeli keretében fogadta a média képviselőit 2025. november 25-én dr. Kincses Áron, a KSH elnöke, Ádám Dénes és dr. Tóth Tamás, a hivatal elnökhelyettesei, dr. Kelemen Rita elnöki tanácsadó, Szabics András Zsolt és Jáger Viktor főosztályvezető, valamint Kollár Krisztina osztályvezető.

A Központi Statisztikai Hivatal elkötelezett a folyamatos innováció, a módszertani fejlesztések, valamint a legújabb adattudományi technológiák alkalmazása mellett. Ennek egyik kézzelfogható eredményét a 2024 elejétől közzétett kísérleti statisztikák jelentik, amelyek új megközelítésből mutatnak be jelenségeket, de a felmerülő felhasználói igények gyors kielégítésére is alkalmasak. Emellett a mesterséges intelligencia alapú fejlesztések (mint például a Virtuális Magyarország) volt még a témája annak a sajtóreggelinek, amelyet 2025. november 25-én rendezett a hivatal.

A sajtóreggeli szakmai előadásában Jáger Viktor, a KSH Módszertani és Innovációs Főosztályának vezetője ismertette a kísérleti statisztikák céljait, módszertani jellemzőit és innovatív megközelítéseit. Rámutatott, hogy a kísérleti statisztika új módszerekre, forrásokra és mérési megoldásokra épül, „a kísérleti jelző tehát az innováció jelenlétét jelzi” – hangsúlyozta a főosztályvezető. Előadásán keresztül bemutatta, miként teremtenek teret a fejlesztések és a modernizáció számára, valamint a résztvevők betekintést nyerhettek a KSH-ban zajló innovatív projektekbe, a mesterséges intelligenciát alkalmazó megoldásokba és a statisztikai működés modernizációjába.

A KSH az új adatkörök közzétételével lehetőséget teremt arra, hogy a felhasználók már a fejlesztés korai szakaszában megismerhessék

az adatokat, és visszajelzéseikkel hozzájárulhassanak azok továbbfejlesztéséhez.

A KSH a sajtóreggelit hagyományteremtő céllal negyedévente megrendezi.

A „most” előrejelzése: gépi tanulással a gazdaság nyomában – Eredményesen zárult a KSH *Főbb indikátorok előrejelzése (P022)* elnevezésű fejlesztése. Ez olyan módszer kidolgozását tűzte ki célul, amellyel megbízhatóan lehet előre becsülni kiemelt, rövid távú gazdasági mutatók értékeit, még azelőtt, hogy a hivatalos adatok megjelenjenek. Ezáltal a gazdaság aktuális helyzetére gyorsabban és pontosabban lehet következtetni.

A gazdasági döntéshozatalhoz naprakész és releváns adatokra van szükség, ezért elengedhetetlen a legkorszerűbb technikák, módszerek alkalmazása a statisztikai adatfeldolgozás és -elemzés területén. A projekt a gépi tanulás eszközeinek felhasználásával azt célozta meg, hogy három fontos, rövid távú gazdasági indikátor – az ipari termelési volumenindex (PVI), a termelőiár-index (PPI) és a szolgáltatások termelőiár-indexe (SPPI) – alakulását már a hivatalos statisztikai adatok közzététele előtt, megbízhatóan és valós időben lehessen becsülni. Az alkalmazott *nowcasting* módszer egy előrejelzési technika, amely képes a legfrissebb, de még nem teljes körű adatok felhasználásával időszerűbb és relevánsabb képet adni a gazdaság pillanatnyi helyzetéről, mint a hagyományos, késleltetett statisztikai eljárások. A célkitűzés tehát olyan előrejelző modell kialakítása volt, amely nemcsak gyors, hanem megbízható és pontos is, továbbá hatékonyan támogatja a gazdasági döntéseket. A fejlesztés szakmai alapját az Eurostat által szervezett *European Statistics Awards Nowcasting* versenyeken szerzett tapasztalatok adták. Ezek a nemzetközi megmérettetések lehetőséget nyújtottak különféle

módszertani megközelítések, modellek kipróbálására, komplex összefüggések kezelésére. A versenyek során kialakított munkafolyamatok – például az adatelemzés lépései, a segédváltozók kiválasztása és szűrése, valamint a változók közötti kapcsolatok (interakciók) kezelése – alapul szolgáltak a projekthez, a fejlesztési irányok kijelöléséhez, a megvalósított Machine Learning (ML-) modellek és nowcastingeljárások kidolgozásához.

A KSH eddig összesen hat alkalommal vett részt a *European Statistics Awards Nowcasting* versenyen (<https://statistics-awards.eu/>). Ezeket a megmérettetéseket az Eurostat azzal a céllal hirdeti meg, hogy ösztönözze új módszertanok és külső adatforrások feltárását, amelyek alkalmazhatók fontos gazdasági mutatók előrejelzésére. A versenyek során a KSH csapata egy alkalommal első helyezést ért el: az *Energetikai Nowcasting* – Oil kategóriában a hagyományos idősoros elemzéseket korszerű, erdőalapú becslési technikákkal ötvözték, ami különösen eredményesnek bizonyult.

A projekt 2024 elején elméleti alapozással indult, majd – a gépi tanulásra épülő nowcastingtechnika áttekintése után – a gyakorlati modellezés következett. Ennek keretében rekurzív neurális hálózatok (RNN-ek) kerültek implementálásra Python nyelven, amelyek kifejezetten alkalmasak idősoradatok feldolgozására és rövid távú előrejelzések készítésére. A projekt fő célja három STS-index – az ipari termelési volumenindex (PVI), a termelőiár-index (PPI) és a szolgáltatások termelőiár-indexe (SPPI) – előrejelzése volt. A modellezés során az egyik legnagyobb kihívást az jelentette, hogy nem álltak rendelkezésre olyan belső adatsorok, amelyek erős és időben elérhető összefüggést mutattak volna a vizsgált indexekkel, emiatt döntő szerepet kaptak a nyilvánosan elérhető internetes adatforrások, amelyekből úgynevezett proxyváltozókat – vagyis helyettesítő adatokat – lehetett azonosítani. Ezek olyan adatok, amelyek közvetett módon utalnak a célváltozók alakulására, bár esetenként csak gyenge korrelációval.

A PVI előrejelzésére sikerült egy jól működő, pontos előrejelzést adó modellt kialakítani, köszönhetően a rendelkezésre álló minőségi előrejelző változóknak, például a teherkilométer indexnek. Ugyanakkor a PPI- és az SPPI-indexek esetében nehezebb volt megfelelő minőségű proxykat találni, így ezek pontossága jelenleg még nem elég ahhoz, hogy éles környezetben is megbízhatóan használhatók legyenek, ezért további adatforrások bevonására van szükség a javításukhoz. Fontos megjegyezni, hogy a fejlesztés során – referencia céljából – benchmarkmodellek is készültek, amelyek kizárólag az idősor múltbeli adataira támaszkodnak, külső előrejelző változók nélkül. Ezek teljesítményéhez viszonyítva kerültek objektív értékelésre a gépi tanulási modellek eredményei.

A projekt egyik legnagyobb eredménye az automatizált pipeline (azaz automatizált adatfeldolgozási folyamatlánc) kialakítása, amely képes:

- a különféle adatforrásokból érkező információkat feldolgozni,
- gépi tanulási modelleket építeni (Gaussian, RNN, LSTM, *de-cision tree based ensemble methods*),
- és ezekkel előrejelzéseket készíteni.

A rendszer működése minimális emberi beavatkozást igényel: elegendő megadni az új proxyváltozókat, illetve kiválasztani, hogy milyen típusú modellt használjon.

Összességében a projekt eredményesnek tekinthető:

- Létrejött egy működőképes, gépi tanuláson alapuló automatizált előrejelző rendszer, amely már most képes pontosan becsülni a PVI rövid távú alakulását.
- A módszertan rugalmasan továbbfejleszhető: az adatbázisok frissítésével és pontosabb proxyváltozók bevonásával növelhető a modell pontossága, így más gazdasági mutatók előrejelzésére is alkalmas lehet.

A fejlesztés során számos értékes tapasztalat született, amelyek jó gyakorlatként jövőbeni projektek megvalósítása során is alkalmazhatók. Különösen fontos a Python-alapú,

strukturált munkafolyamatok elsajátítása, amely stabil alapot nyújt a további fejlesztési feladatokhoz. A projekt sikerességét jelentősen támogatta a módszertani fejlesztők és a vezetők közötti folyamatos, célorientált kommunikáció, valamint az, hogy a fejlesztők aktívan részt vettek a projektmenedzsmenttel kapcsolatos döntéshozatalban is. A hatékony időgazdálkodást elősegítette, hogy a fejlesztői csapaton kívüli egyeztetésekre kizárólag indokolt esetekben került sor.

Az előrejelzési modell úgy került kialakításra, hogy rugalmasan továbbfejleszthető és más szakterületeken is alkalmazható legyen. Magas szintű, könnyen használható programcsomag készült, amely lehetővé teszi, hogy – amennyiben az adatok a megfelelő formátumban rendelkezésre állnak – akár egyetlen függvény meghívásával előrejelzést lehessen készíteni bármely indikátorra. Ez a megközelítés lehetőséget ad arra, hogy más területek is gyorsan és hatékonyan tesztelhessék a modell működését, a becslések relevanciáját, valamint azok gyakorlati alkalmazhatóságát.

A fejlesztés során több mesterséges intelligencia (MI-) alapú – gépi tanulás

(Machine Learning, ML-) és mélytanulás (Deep Learning, DL-) – modell került tesztelésre. Ezek a modern módszerek értékes tudást adnak olyan statisztikák készítéséhez, amikor nagy mennyiségű adminisztratív adatot kell a becslésekhez felhasználni. Ilyen esetekben ugyanis olyan hatékony eljárásokra van szükség, amelyek képesek kezelni a nagyméretű adatállományokat, amire a hagyományos módszerek már nem mindig alkalmasak. A gépi tanulási eljárások viszont épp az ilyen helyzetekre nyújtanak hatékony megoldást.

A hivatal kiemelt célja, hogy adatfeldolgozási folyamatait a korszerű statisztikai módszerek és a mesterséges intelligencia kínálta lehetőségek integrálásával fokozatosan továbbfejlessze. Ennek köszönhetően gyorsabban és pontosabban tud naprakész, releváns információkat biztosítani, valamint hatékonyabb támogatást nyújtani az adatok mögött rejlő összefüggések feltárásához. E törekvés egyik fontos mérföldköve a most bemutatott *Főbb indikátorok előrejelzése* fejlesztés.