



Közzététel: 2024. november 26.

A tanulmány címe:

**A felügyelt gépi tanulás alkalmazási lehetőségei szöveges adatokon.**

**A magyar országgyűlésben 1998–2018 között elhangzott beszédek elemzése**

Szerzők:

**BUDA JAKAB MÁTÉ**

az Eötvös Loránd Tudományegyetem Társadalomtudományi Kar Statisztika Tanszékének tanársegédje, kutatója, HUN-REN Rényi Alfréd Matematikai Kutatóintézet Mesterséges Intelligencia Osztályának tudományos segédmunkatársa

E-mail: [buda.jakab@tatk.elte.hu](mailto:buda.jakab@tatk.elte.hu)

**NÉMETH RENÁTA**

az Eötvös Loránd Tudományegyetem Társadalomtudományi Kar Statisztika Tanszékének tanszékvezető professzora, a Research Center for Computational Social Science társvezetője

E-mail: [nemeth.renata@tatk.elte.hu](mailto:nemeth.renata@tatk.elte.hu)

DOI: <https://doi.org/10.20311/stat2024.11.hu1087>

**Az alábbi feltételek érvényesek minden, a Központi Statisztikai Hivatal (a továbbiakban: KSH) Statisztikai Szemle c. folyóiratában (a továbbiakban: Folyóirat) megjelenő tanulmányra. Felhasználó a tanulmány vagy annak részei felhasználásával egyidejűleg tudomásul veszi a jelen dokumentumban foglalt felhasználási feltételeket, és azokat magára nézve kötelezőnek fogadja el. Tudomásul veszi, hogy a jelen feltételek megszegéséből eredő valamennyi kárért felelősséggel tartozik.**

1. A jogszabályi tartalom kivételével a tanulmányok a szerzői jogról szóló 1999. évi LXXVI. törvény (Sztj.) szerint szerzői műnek minősülnek. A szerzői jog jogosultja a KSH.
2. A KSH földrajzi és időbeli korlátozás nélküli, nem kizárólagos, nem átdadható, térítésmentes felhasználási jogot biztosít a Felhasználó részére a tanulmány vonatkozásában.
3. A felhasználási jog keretében a Felhasználó jogosult a tanulmány:
  - a) oktatási és kutatási célú felhasználására (nyilvánosságra hozatalára és továbbítására a 4. pontban foglalt kivétellel) a Folyóirat és a szerző(k) feltüntetésével;
  - b) tartalmáról összefoglaló készítésére az írott és az elektronikus médiában a Folyóirat és a szerző(k) feltüntetésével;
  - c) részletének idézésére – az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző(k) megnevezésével.
4. A Felhasználó nem jogosult a tanulmány továbbértékesítésére, haszonszerzési célú felhasználására. Ez a korlátozás nem érinti a tanulmány felhasználásával előállított, de az Sztj. szerint önálló szerzői műnek minősülő mű ilyen célú felhasználását.
5. A tanulmány átdolgozása, újra publikálása tilos.
6. A 3. a)–c) pontban foglaltak alapján a Folyóiratot és a szerző(ke)t az alábbiak szerint kell feltüntetni:  
„*Forrás: Statisztikai Szemle c. folyóirat 102. évfolyam 11. számában megjelent, **Buda Jakab Máté – Németh Renáta** által írt, **A felügyelt gépi tanulás alkalmazási lehetőségei szöveges adatokon. A magyar országgyűlésben 1998–2018 között elhangzott beszédek elemzése** című tanulmány (link csatolása)”*
7. A Folyóiratban megjelenő tanulmányok kutatói véleményeket tükröznek, amelyek nem feltétlenül esnek egybe a KSH vagy a szerzők által képviselt intézmények hivatalos álláspontjával.

Buda Jakab Máté – Németh Renáta

**A felügyelt gépi tanulás alkalmazási lehetőségei  
szöveges adatokon.  
A magyar országgyűlésben 1998–2018 között  
elhangzott beszédek elemzése\***

**The application of supervised machine learning  
on textual data.  
An analysis of speeches delivered  
in the Hungarian Parliament between 1998 and 2018**

Buda Jakab Máté, az Eötvös Loránd Tudományegyetem Társadalomtudományi Kar Statisztika Tanszékének tanársegédje, kutatója, HUN-REN Rényi Alfréd Matematikai Kutatóintézet Mesterséges Intelligencia Osztályának tudományos segédmunkatársa

E-mail: buda.jakab@tatk.elte.hu

Németh Renáta, az Eötvös Loránd Tudományegyetem Társadalomtudományi Kar Statisztika Tanszékének tanszékvezető professzora, a Research Center for Computational Social Science társvezetője

E-mail: nemeth.renata@tatk.elte.hu

Tanulmányunkban a magyar országgyűlésben 1998–2018 között elhangzott beszédeket elemezve a szöveges adatokon végzett felügyelt gépi tanulás alkalmazási lehetőségeit mutatjuk be a társadalomtudományokban. Megmutatjuk, hogy a politikai polarizáció nyelvi lenyomatának, vagyis bizonyos csoportok nyelvhasználata közötti távolságnak a mérésére használhatjuk a klasszifikációs modellek pontossági metrikáját, ugyanakkor kitérünk arra is, hogy kizárólag a modell teljesítményére való hagyatkozás önmagában kevés, érdemes a modellek működését alaposabban is megvizsgálni. A magyarországi polarizáció időbeli alakulásának tanulmányozásához n-gram-változókra épülő XGBoost klasszifikációs modelleket illesztünk az összes 1998 és 2018 közötti parlamenti ciklusra, külön-külön. A klasszifikációs feladat az volt, hogy megkülönböztessük a vizsgált időszak alatti két domináns párt (Fidesz és MSZP) képviselői által elmondott beszédeket. Bár az egymást követő parlamenti ciklusok során a klasszifikáció teljesítményének javulása nyilvánvaló – ami a nyelvhasználatban tükröződő polarizáció erősödésére utal –, e nyelvi változás sajátosságai további tisztázást igényelnek. A nyelvhasználatban tükröződő növekvő különbség még kézenfekvőbb, ha nemcsak a modellek pontosságát, hanem az általuk megjósolt valószínűségek eloszlását is megvizsgáljuk. Kísérletet teszünk a modellek magyarázatára is, és további lehetséges kutatási irányokat is vázolunk.

**Kulcsszavak:** természetes nyelvfeldolgozás, polarizáció, megmagyarázható gépi tanulás

\* Az NKFIH által támogatott kutatás (K-134428 azonosító alatt). A Kulturális és Innovációs Minisztérium ÚNKP-23-3 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.

In this study, we explore the potential of supervised machine learning for social research by analyzing speeches delivered in the Hungarian Parliament between 1998 and 2018. We show that the accuracy metric of the classification model can be used to measure the linguistic imprint of political polarization, i.e. the distance between certain groups. However, we point out that relying solely on the model's performance is not enough, it is worthwhile to examine the models in more detail. To study the evolution of polarization in Hungary over time, we trained XGBoost classification models with n-gram features for each parliamentary term separately within the study period. The aim was to distinguish speeches made by members of the two dominant parties in the timeframe of the study (Fidesz and MSZP). While model accuracy improves over time, indicating increased polarization, we argue that a comprehensive analysis requires examining not only overall performance but also the distribution of predicted probabilities and feature importance. The growing divergence in language use is even more evident when we look not only at the accuracy of the models but also at the distribution of the probabilities they predict. We also attempt to explain the models and outline further possible research directions.

Keywords: natural language processing, polarization, explainable machine learning

Az NLP (*natural language processing*, természetesnyelv-feldolgozás) izgalmas és a társadalomkutatás számára perspektivikus tudományterület, amelynek a célja az emberi nyelv elemzése és megértése számítástudományi technikák segítségével (Evans–Aceves, 2016; Ignatow–Mihalcea, 2017). Az elmúlt néhány évben az NLP a társadalomtudományokban is kezdett teret nyerni, a politológiától a közgazdaságtudományig (Ignatow–Mihalcea, 2017). A társterületekhez képest a szociológiában némi késéssel indult el a változás, de a 2010-es években már világos volt a trend (Németh, 2024). Az NLP társadalomkutatásban történő felhasználása tehát ehhez a közelmúltban elterjedt „szöveg mint adat” (*text as data*) (Gentzkow et al., 2019) megközelítéshez kapcsolódik, ahol a szöveg mint rendezett, jól strukturált, numerikus adatbázisba rendezett állomány adja a számítógépes algoritmusok inputját. A hagyományos szövegelemzéssel szemben itt a szöveg nem elolvasandó, megértendő tárgyként, hanem inkább az automatizált módszerek inputjaként jelenik meg anélkül, hogy azt ténylegesen bárki elolvasná.

## 1. Felügyelt gépi tanítás, predikciós teljesítmény

### 1.1. Felügyelt gépi tanulás – intuitív bevezető

A kortárs társadalomkutatásban használt két nagy NLP-módszercsalád a felügyelt és a felügyelet nélküli gépi tanulás. A mögöttük álló egzakt statisztikaelméletet *Vapnik (2000)* alaposan tárgyalja. A felügyelt és a felügyelet nélküli tanulás közötti különbség röviden azon alapszik, hogy már létező elmélet/meglevő háttérismeret empirikus megnyilvánulásait keressük (felügyelt tanulás, ahol a „felügyelet” maga a háttérelmélet), vagy induktív módon egy még nem vizsgált téma feltárása a cél (felügyelet nélküli tanulás). A dichotómia hasonlít a klasszikus szociológiai módszertan konfirmatív/exploratív különbségtételéhez.

A felügyelet nélküli módszerek nem igényelnek előzetes feltételezéseket vagy külső ismereteket, a modell maga tanulja meg az adatok szerkezetét mintázatok keresésével. Ezzel szemben a felügyelt módszerek előzetes címkézett adathalmazt igényelnek: jelen cikkünkben ezt a címkét a parlamenti beszédek előadójának párt-hovatartozása adja. A felügyelt tanulási modellek célja, hogy megtanulják, hogyan rendeljék ezeket a címkéket a szövegekhez (a példánál maradva: hogyan automatizálható a párt szerinti címkézés a beszédek szövegmintázatai alapján). A megtanulás itt azt jelenti, hogy (viszonylag kevés hiba mellett) az algoritmus maga is képes lesz címkéket rendelni még címkézetlen szövegekhez.

A modell teljesítményének értékeléséhez az adathalmazt két részre bontják: egy tanulóhalmazra, amelyet a modell tanítására használnak, és egy teszthalmazra, amelyet az értékelésre használnak. A teszthalmaz ideális esetben ugyanabból az eloszlásból származik, mint a tanulóhalmaz, de attól független. Amennyiben a teszthalmaz függetlensége teljesül, a teszthalmazon mért teljesítmény torzítatlan becslést ad arra, hogy egy ugyanebből az eloszlásból vett, korábban nem megfigyelt mintán hogyan teljesítene a modell. A tanítási folyamat során a modell paramétereit iteratív módon állítjuk be, hogy minimalizáljuk az előrejelzési hibákat a tanulóhalmazon. A pontosság egy gyakran használt mérőszám az osztályozómodellnek a teszthalmazon mért teljesítményének értékelésére, a helyesen osztályozott tesztpéldák és a tesztpéldák teljes számának arányaként számítják ki.

Az osztályozás (vagy klasszifikáció) a felügyelt tanulás egyik típusa, a bemeneti adatok előre meghatározott osztályokba vagy kategóriákba való besorolását jelenti (egy másik gyakran használt felügyelt feladat a regresszió, ahol magas mérésű szintű a kimenet). A predikció a gépi tanulásban azt a folyamatot jelenti, amikor egy már betanított modell segítségével új, címkézetlen bemeneti adatokhoz címkéket generálunk (például a beszéd előadójának párt-hovatartozása Fidesz vagy MSZP). A megjósolt valószínűség (0 és 1 között) a modell becslését jelenti arra

vonatkozóan, hogy az adatok milyen valószínűséggel tartoznak egy adott osztályba, ezeket diszkretizáljuk később, azaz alakítjuk címkékké. (Ebben az értelemben tehát sok klasszifikációs algoritmus valójában regresszióként közelít a feladathoz.) Tanulmányunkban a probléma bináris osztályozási feladatként van megfogalmazva, ahol csak a Fidesz és az MSZP képviselőire fókuszálunk. A modell célja annak a valószínűségének megbecslése, hogy egy beszédet a Fidesz által delegált képviselő adta-e elő; ha a beszéd előrejelzett valószínűsége 0,5 fölött van, akkor a modell úgy minősíti, hogy azt a Fidesz képviselője mondta, ellenkező esetben az MSZP-hez sorolódik a szöveg.

A gyakorlatban szöveges adatokon komplex modelleket alkalmaznak, komplex (például stiláris jegyeket reprezentáló vagy a szöveg mélyebb szemantikai viszonyait megragadó) magyarázóváltozókkal, a nagy mennyiségű (több ezer) magyarázóváltozó miatt dimenziócsökkentő megoldásokkal és a túlillesztést elkerülni hivatott megoldásokkal, de az algoritmus létrejöttének logikája ott is hasonló (bővebben: *Eisenstein, 2019*).

A felügyelt módszerek abban támogatják tehát a kutatókat, hogy (1) nagyobb szöveghalmazt dolgozhassanak fel, mint amennyire emberi kapacitás képes lenne. További (2) cél lehet, hogy a címkézést kiterjesszük egy már felcímkézett kisebb szöveghalmazról a teljes korpuszra. Ám a legfontosabb (inkább tudományos elvárásokhoz igazodó) felhasználási lehetőség az, hogy (3) megértsük az automatikus címkézés mögött álló szabályokat, egyszerű példát említve: lássuk, mely szóhasználat valószínűsíti leginkább az adott párthoz való tartozást. Ez utóbbi cél tulajdonképpen azoknak a tartalmi-társadalomkutatási interpretációra/magyarázatra lehetőséget adó jegyeknek a feltárását jelenti, amelyek az egyébként „fekete dobozként” működő kódoló algoritmus mögött állnak (*Molnar, 2020*). A felügyelt tanuláshoz több új továbbfejlesztése létezik, ahol az eredeti logika kissé módosul, így például az aktív tanulás esetében nem egy fix tanulómeghalmazunk van, hanem az algoritmus maga kéri menet közben konkrét, még címkézetlen, de a tanulásban fontosnak tűnő szövegek címkézését, a transzfer tanulás esetében pedig a címkézett adatoktól eltérő besorolási problémát kell a tanulási algoritmusnak megoldania (*Eisenstein, 2019*).

## 1.2. A predikciós teljesítmény mint a politikai polarizáció mértéke

*Németh (2023)* áttekintése szerint a politikai polarizációt vizsgáló, NLP-t használó, 2010 óta megjelent tanulmányok közül minden harmadik felügyelt gépi tanulást (osztályozást) alkalmaz a polarizáció vizsgálatára, ahol a cél a szöveg szerzőjének politikai hovatartozásának előrejelzése. Az osztályozás ebben az össze-

függésben olyan módszer, amely a szerző politikai álláspontját próbálja azonosítani az általa használt szavak alapján. Itt a gépi tanuló magas predikciós pontossága arra utal, hogy az egyik politikai oldal által használt nyelv egyrészt valamilyen szinten homogén, másrészt felismerhető mintázatok szerint különbözik a másik oldal által használt nyelvtől. Vagyis polarizációs metrika definiálható: minél inkább képes az osztályozómodell azonosítani a szerző álláspontját, annál nagyobb fokú a politikai polarizáció. Az áttekintés (*Németh, 2023*) egyik idetartozó kutatása *Bayram és szerzőtársai (2019)*, akik ezt a megközelítést követve elemezték az Egyesült Államok Kongresszusa képviselőházának felszólalásait. Az időben előre haladva egyre jobb hatékonyságú osztályozó algoritmusokat tudtak alkotni, amit úgy interpretáltak, hogy a politikai polarizáció egyre jobban detektálhatóvá vált a beszédek nyelvezetében.

Az alábbiakban saját esettanulmányként a magyar országgyűlésben 1998–2018 között elhangzott beszédeken végzünk elemzést, azzal a céllal, hogy felmérjük a politikai diskurzus e rétegében jelen lévő polarizáció mértékét és dinamikáját. Öt parlamenti ciklusra illesztettünk gépi tanulási modellt, csak a Fidesz és az MSZP szónokainak megkülönböztethetőségét vizsgálva. Az eredmények értelmezéséhez megmagyarázható gépi tanulási eszközöket használunk.

### 1.3. A predikciós modell fekete dobozának felnyitása

Ahogy korábban említettük, egy tanulási modell interpretációjakor/magyarázatakor a cél tulajdonképpen azoknak az (itt: szövegmintázattal kapcsolatos) tényezőknek a feltárását jelenti, amelyek az egyébként „fekete dobozként” működő osztályozó algoritmus mögött állnak. A mi esetünkben: ha jó hatékonysággal meg lehet különböztetni a Fidesz és az MSZP parlamenti felszólalóinak beszédét, milyen szövegmintázatok azok, amelyek ezt a különbségtételt lehetővé teszik?

A gépi tanulási algoritmusok fekete dobozként való kezelése jelentős kockázatokot rejthet magában, beleértve a modell tanításánál használt adatokban jelen lévő torzítások állandósulását, amely az eredmények torzulásához vezethet. Ez a felismerés ösztönözte a megmagyarázható mesterséges intelligencia területének kialakulását, amelynek célja olyan eszközök és technikák kifejlesztése, amelyek átláthatóságot biztosítanak ezen algoritmusok döntéshozatali folyamataiba.

Röviden érdemes itt kitérni az előrejelzés és magyarázat kettőségére (részletesen lásd *Németh, 2021*), hiszen míg az üzleti és alkalmazott kutatásokban inkább az előbbi a cél (spam-e-mail detektálása, képalkotó diagnosztikai eljárás outputjának besorolása rosszindulatú/jóindulatú kategóriába stb.), addig a társadalomtudományok számára az utóbbi fontos, ami az elméletalkotáshoz, a tudástermeléshez hozzá tud járulni. Az adattudós üzleti felhasználó olyan modellt épít, amelynek

csak az előrejelzési teljesítménye számít, és nem feltétlenül okoz számára gondot, ha a modell más szempontból fekete dobozként tekinthető.

A ma alkalmazott gépi tanulási algoritmusok interpretálása komoly kihívás. A klasszikus, egyszerű modellek, amilyenek a társadalomkutatásban is ismert regressziók, közvetlenül interpretálhatók: expliciten adott, hogy melyik magyarázóváltozó milyen előjellel, milyen súllyal, a többi változóval alkotott milyen interakciójával vesz részt egy predikció létrehozásában. Ezzel szemben a gépi tanulási modellek magyarázata alapvető nehézséget jelent: ezeket a modelleket éppen arra használják, hogy összetett mintákat találjanak az adatokban, amelyek gyakran meghaladják az emberi felfogóképességet, ezért magyarázatuk nem könnyű feladat, mert nincsen közvetlen, könnyen átlátható függvénykapcsolat az input és a predikció között. Többféle megközelítés létezik erre a problémára: a lokális magyarázatok az egyes döntések magyarázatára szolgálnak, míg a globális magyarázatok célja, hogy átfogó betekintést nyújtsanak a modell döntéshozatali folyamatába (ez általában egyszerűsített leírás, mivel ezek a modellek rendkívül összetettek). Vannak kísérletek arra vonatkozóan is, hogy önmagyarázó AI- és ML-algoritmusokat készítsenek (*Rudin–Radin, 2019*).

A modellek interpretálhatósága nemcsak azért fontos, mert az értelmezés segít eldönteni, hogy az előrejelzések értelmesek-e, hanem azért is, mert felfedhetjük, melyek azok a tényezők, amelyek a legnagyobb hatással vannak a modell döntéseire, amelyek csökkentik vagy növelik az osztályba sorolás valószínűségét, illetve, milyen hatásuk van az a priori fontosnak tartott tényezőknek. Az interpretálhatóságnak nem csupán a tudományos, de az üzleti alkalmazásokban is fontos, az utóbbi években felismert szerepe van, ugyanis a predikciók is kevésbé megbízhatóak, ha nem tudjuk, az üzleti környezet mely paramétereire épülnek – így az üzleti/társadalmi környezet megváltozása a predikció romlását hozhatja anélkül, hogy ennek tudatában lennénk.

*Németh (2023)* áttekintése a politikai polarizáció kutatásában is azonosítja ezt a problémát: eredményei szerint a polarizáció mérésére osztályozómodellt alkalmazó öt tanulmányból három csak a predikciós teljesítmény optimalizálására összpontosít, és nem tér ki az interpretációra, azaz arra, hogy milyen nyelvi jellemzők játszottak szerepet a szövegek politikai ideológia szerinti szétválasztásában. Pedig az interpretáció egyrészt a modell validálásában is megkerülhetetlen: akkor tekinthető validnak a tanuló, ha plauzibilis jellemzőket használ a besorolás során. Másrészt, ha megértjük, hogy a modell miért hozott egy bizonyos döntést, és megtaláljuk azokat a kifejezéseket, amelyek a leginkább utalnak például a konzervatív vagy a liberális álláspontokra, közelebb kerülünk a politikai polarizáció megértéséhez is.

## 2. Módszertan

### 2.1. Adatok

Kutatásunk fókuszát az 1998 és 2018 közötti időszakot átfogó hivatalos parlamenti vitákra szűkítettük, ahol kizárólag a két nagy párt (MSZP és Fidesz) valamelyikéhez kötődő képviselők által ebben az időszakban elmondott beszédeit elemezzük. Bár a vizsgált időszak alatt alapvető változások történtek, ez idő alatt két nagy párt létezett: a konzervatív Fidesz – Magyar Polgári Párt (2003-tól Fidesz – Magyar Polgári Szövetség) és a baloldali MSZP (Magyar Szocialista Párt). Bár ezeknek a pártoknak az ideológiai bal-jobb spektrumon elfoglalt helye ez idő alatt változott, ez számunkra nem lényeges, hiszen bennünket csak a pártok közötti különbség érdekel.

A vizsgált időszakban öt négyéves parlamenti ciklus volt. Az első ciklusban, 1998 és 2002 között a Fidesz volt a fő kormánypárt, koalícióban a Független Kisgazda-, Földmunkás- és Polgári Párttal (FKGP) és a Magyar Demokrata Fórummal (MDF). 2002-től 2006-ig a fő kormánypárt az MSZP volt, a kisebbik koalíciós partnerrel, a Szabad Demokraták Szövetségével (SZDSZ). A következő ciklusban is ez a felállás kormányzott (bár a koalíció 2008-ban felbomlott). 2010-ben a Fidesz és a Kereszténydemokrata Néppárt (KDNP) szövetsége kétharmados többséget szerzett a parlamentben, visszaállítva a Fidesz fő kormánypárti szerepét. Ez a tendencia a 2014-es parlamenti választásokon is folytatódott, hasonló eredménnyel. A két nagy párton tanított modellek előrejelzett valószínűsége alapján elemezzük még a következő kisebb pártokat is, ezek a szélsőjobboldali Magyar Igazság és Élet Pártja (MIÉP), a zöld-liberális Lehet Más a Politika (LMP), a szociálliberális Demokratikus Koalíció (DK), a liberális Együtt, a szélsőjobboldali Jobbik Magyarországért Mozgalom és a liberális-zöld Párbeszéd Magyarországért (PM).

A teljes beszédkorpusz 143 183 beszédből áll. A beszédeket a K-monitor (<http://parlament.k-monitor.hu/>) gyűjtötte össze a magyar parlament hivatalos honlapjáról (<https://www.parlament.hu/>). Kiszűrtük a tartalmilag irreleváns ügyrendi és az 51 szónál rövidebb beszédeket. A fennmaradó korpuszból párthovatározás és parlamenti ciklusok szerint kiegyensúlyozott véletlen mintát vettünk. Mivel a cél a modellek teljesítményének összehasonlítása a ciklusok között, a minta méretének minden ciklusra azonosnak kell lennie, ellenkező esetben a jobb teljesítmény a nagyobb tanulóhalmaz méretéből is eredhet, mivel egy nagyobb halmazban több információ van, amelyet a modell megtanulhat, és nem lenne lehetséges a tanulóhalmaz méretének hatását elválasztani a polarizáció változásának hatásától. A teszhalmaz pártonként és ciklusonként 500 beszédből áll, ami összesen 5000 beszédet jelent. A tanulóhalmaz pártonként és ciklusonként 4148, összesen 41 480 beszédből áll.



## 2.2. A szövegek előfeldolgozása

Az előfeldolgozás az NLP egyik alapvető lépése, célja a nyers szöveges adatok finomítása, hogy azok alkalmasak legyenek a gépi tanulási algoritmusok általi elemzésre. Itt a szöveg minden egyes szava egyedi jellemzőként („magyarázóváltozóként”) jelenik meg. Ezek a jellemzők megragadják az adott szavak jelenlétét vagy hiányát a szövegben, lehetővé téve a modell számára, hogy a beszédek alkotó szavak és a beszédekhez tartozó címkék közötti kapcsolatot megtanulja.

Az első előfeldolgozási lépés ebben a kutatásban a szöveg kisbetűvé alakítása volt, ezután eltávolítottuk a kevés szemantikai jelentést hordozó gyakori szavakat (ún. stopszavakat), mint például az „az”, az „és” vagy az „is”. Következő lépésként lemmatizálást végeztünk, amely a szavak alapformára, úgynevezett lemmaformára való redukcióját jelenti (toldalékok levágása). A névelemek azonosítása során azonos alakra hoztuk a szövegben található entitások, például személyek, szervezetek vagy helyszínek megnevezését. Ezek a lépések standardizálják a szavak előfordulásait, csökkentik az adatállomány redundanciáját és az adatok dimenzionalitását. A szignifikáns n-gramok létrehozásakor azonosítottuk és egyetlen kifejezéssé konvertáltuk a gyakran előforduló szószorozatokat. Ez a folyamat segít megragadni olyan értelmes kifejezéseket, amelyek konkrét szemantikai információt közvetíthetnek (például a szomszédos „millió” és „forint” szavakat „millió forint”-ra cseréltük). Ezenkívül a szövegekben előforduló honlapra hivatkozásokat egységesen egy <URL> tokenre változtattuk. Végül a beszédek leiratából eltávolítottuk a formális köszöntő (pl. „Tisztelt Ház”) és záró mondatokat („Köszönöm a figyelmet”), illetve a jegyző által beszűrt, a teremben elhangzó megjegyzéseket és viselkedéseket leíró részleteket („Taps a fideszes képviselők soraiból”).

## 2.3. Modellépítés

Ahogy említettük, NLP-t és gépi tanulási technikákat alkalmazunk a képviselők párhovatartozásának előrejelzésére, minden parlamenti ciklusra külön, kizárólag a felszólalásaik szövege alapján, és a polarizáció mértékét ennek az előrejelzésnek a pontosságaként operacionalizáljuk a teszthalmazokon. Ez a megközelítés megegyezik *Peterson és Spirlingével (2018)*, akik a brit parlamenti vitákat dokumentáló Hansard Recordot használták korpuszként.

A felépített modellek pontosságának időbeli mintázatának vizsgálata mellett elemeztük az előrejelzett valószínűségek párhovatartozás szerinti eloszlását is. Ha a pártok között valóban egyre nagyobb a szakadék, ezeknek az eloszlásoknak idővel egyre szélsőségesebbé kell válniuk. Például a 0,51-os és a 0,99-os előrejelzett valószínűség is a Fideszbe sorolt beszédre utal, de ha a pártok közötti szakadék

szélesedik, kevesebb beszéd fog 0,5 körüli előrejelzett valószínűséget kapni. Emellett megvizsgáltuk a két nagy párton tanított modellnek a többi pártra vonatkozó előrejelzett valószínűségeinek eloszlását is. Modelljeink érvényességét támaszthatja alá, ha ezen eloszlások és a két nagy párt eloszlásai közötti hasonlóság mértéke korrelál a pártok ideológiai és strukturális hasonlóságával. Például, ha egy párt a kormánykoalíció tagja, akkor azt várjuk, hogy e párt képviselői által elmondott beszédek előre jelzett valószínűségeinek eloszlása hasonlít a fő kormánypárthoz – és ez nagy valószínűséggel az ideológiai hasonlóságokra is elmondható.

Osztályozó algoritmusként XGBoost (*eXtreme Gradient Boosting*) -modelleket használtunk. Az XGBoost a strukturált adatok kezelésében való hatékonyságáról ismert osztályozó algoritmus, különösen alkalmas szöveges osztályozási feladatokra. Döntési fák sorozatát építi fel iteratív módon, ahol minden egyes új fa korrigálja az előzőek hibáit. Célunk nem a modellek pontosságának maximalizálása volt, hanem a különböző időszakokra illesztett modellek összehasonlíthatóságának biztosítása. A modellek a szöveget nem csupán szavanként, hanem szomszédos szócsoportonként dolgozták fel, azaz  $n$ -gram-változókat alkalmaztak. Az  $n$ -gramok  $n$  darab egybefüggő szóból álló szekvenciákat jelentenek a szövegben, ahol az  $n$  az unigramoktól (egyetlen szó) vagy bigramoktól (szópárok) a magasabb rendű  $n$ -gramokig terjedhet. A szavak sorozatainak rögzítésével az  $n$ -gramok kontextus és szintaktikai információt szolgáltatnak a modell számára, ami javítja a szöveg jelentésének és szerkezetének megértését. Jelen tanulmányban uni-, bi- és trigramváltozókat használtunk. A beszédek alkotó szavak gyakoriságát bevett megoldásként a TF-IDF (*term frequency-inverse document frequency*) -súlyok segítségével súlyoztuk, ezek a súlyok a szavakat a dokumentumban való gyakoriságuk alapján rangsorolják az összes dokumentumban való gyakoriságukhoz képest, ezáltal kiemelve azokat a szavakat, amelyek gyakoriak a dokumentumban, de ritkák a korpuszban. Minden egyes parlamenti ciklushoz külön modellt dolgoztunk ki.

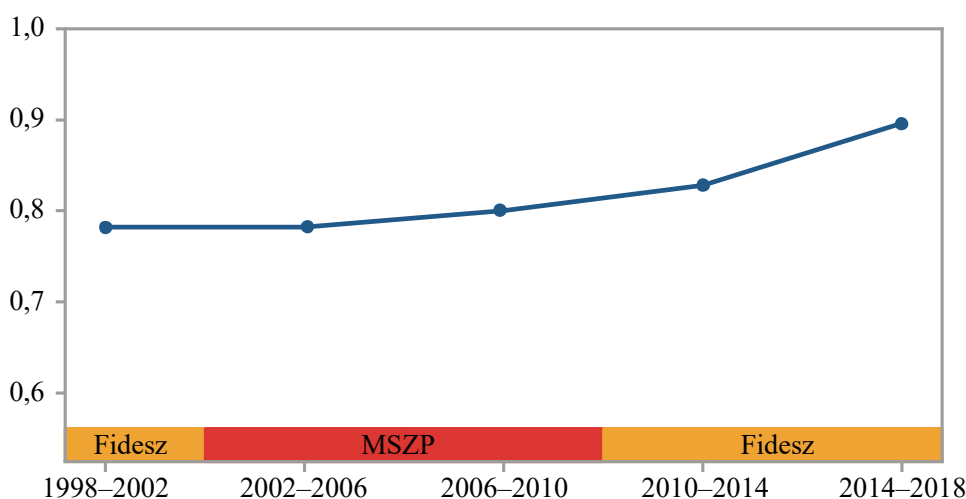
### 3. Eredmények

Első lépésként azt vizsgáltuk, hogy a parlamenti ciklusokra külön-külön illesztett modellek pontossága hogyan alakul az idő múlásával, illetve a vizsgált időszakban valóban tapasztalható-e a modellek teljesítményének javulása, mivel ez alátámasztaná azt az állítást, hogy a polarizáció növekedett. Az 1. ábra a modellek pontosságának alakulását mutatja. A pontosság valóban egyértelműen növekszik, az első

két ciklus között nincs jelentős különbség, de ezt követően a javulás határozottan felgyorsul.

1. ábra

**Az egyes parlamenti ciklusokra illesztett modellek pontosságának időbeli változása**  
*Change over time in the accuracy of models fitted to each parliamentary term*



Ezen a ponton arra következtethetünk, hogy mivel a modellek pontosságában monoton emelkedés tapasztalható, a két nagy párt közötti polarizáció is növekszik. Annak megértése, hogy pontosan hogyan különböztetik meg a modellek a beszédek, további vizsgálatot igényel. Ennek egyik módja a SHAP (*SHapley Additive exPlanations*) -értékek vizsgálata. A SHAP-értékek kiszámítása a gépi tanulásban a modellek magyarázatára alkalmazott módszer, ahol az egyes szövegjellemzőknek a predikcióhoz való hozzájárulásának megértése a cél. A SHAP-értékek kiszámításakor egy olyan helyettesítő modellt hozunk létre, amelynek viselkedése lokálisan megfelel az eredeti modellnek, és az egyes változók befolyásának összege kiadja az adott megfigyelésre adott predikció és a predikciók várható értékének különbségét. A modellek magyarázatát megkísérlő eljárásunk során tehát kiválasztottuk minden parlamenti ciklushoz az arra a ciklusra illesztett modell működése szempontjából legfontosabb szövegalkotó kifejezéseket, terminus technicusszal szólva tokeneket. A legfontosabb tokenek kiválasztását ötszörös kereszt-validálással végeztük, hogy csökkentsük a tanulóhalmazban jelentkező zaj hatását: a tanulómintát öt részre bontottuk, és öt modellt illesztettünk –mindig egy másik ötödöt használva – teszhalmazként, és a maradék 4/5-öt tanítóhalmazként. A legfontosabb tokeneknek pedig a szűrés során azokat tekintettük, amelyek legalább négy

modellben bekerültek azok közé az átlagos abszolút SHAP-értékük szerint legfontosabb tokenek közé, amelyek kombinált abszolút SHAP-értéke eléri az 1-et. (Ezt az 1-et a predikciók 0-1 skálájához kell mérni, az abszolút SHAP-értékek pedig a hatás nagyságát tekintik csak, a hatás irányát nem.) A kiválasztás eredménye a következő:

A 3. parlamenti ciklusban (1998–2002) öt n-gram bizonyult a legfontosabbnak: „képviselőtárs”, „fidesz\_magyar\_polgári\_párt”, „államtitkár\_úr”, „kormánypárti”, „mi”.

A 4. parlamenti ciklusban (2002–2006) öt változó volt a legmarkánsabb: „polgári kormány”, „természetes”, „ellenzéki képviselőtárs”, „államtitkár\_úr”, „szocialista”.

Az 5. parlamenti ciklusban (2006–2010) hat változó emelkedett ki: „államtitkár\_úr”, „természetes”, „ön”, „ember”, „miniszter\_úr”, „lesz”.

A 6. parlamenti ciklusban (2010–2014) három változót választottunk ki: „államtitkár\_úr”, „ön”, „fidesz”.

Végül a 7. parlamenti ciklusban (2014–2018) három változót választottunk ki: „államtitkár\_úr”, „magyar\_szocialista\_párt”, „fidesz”.

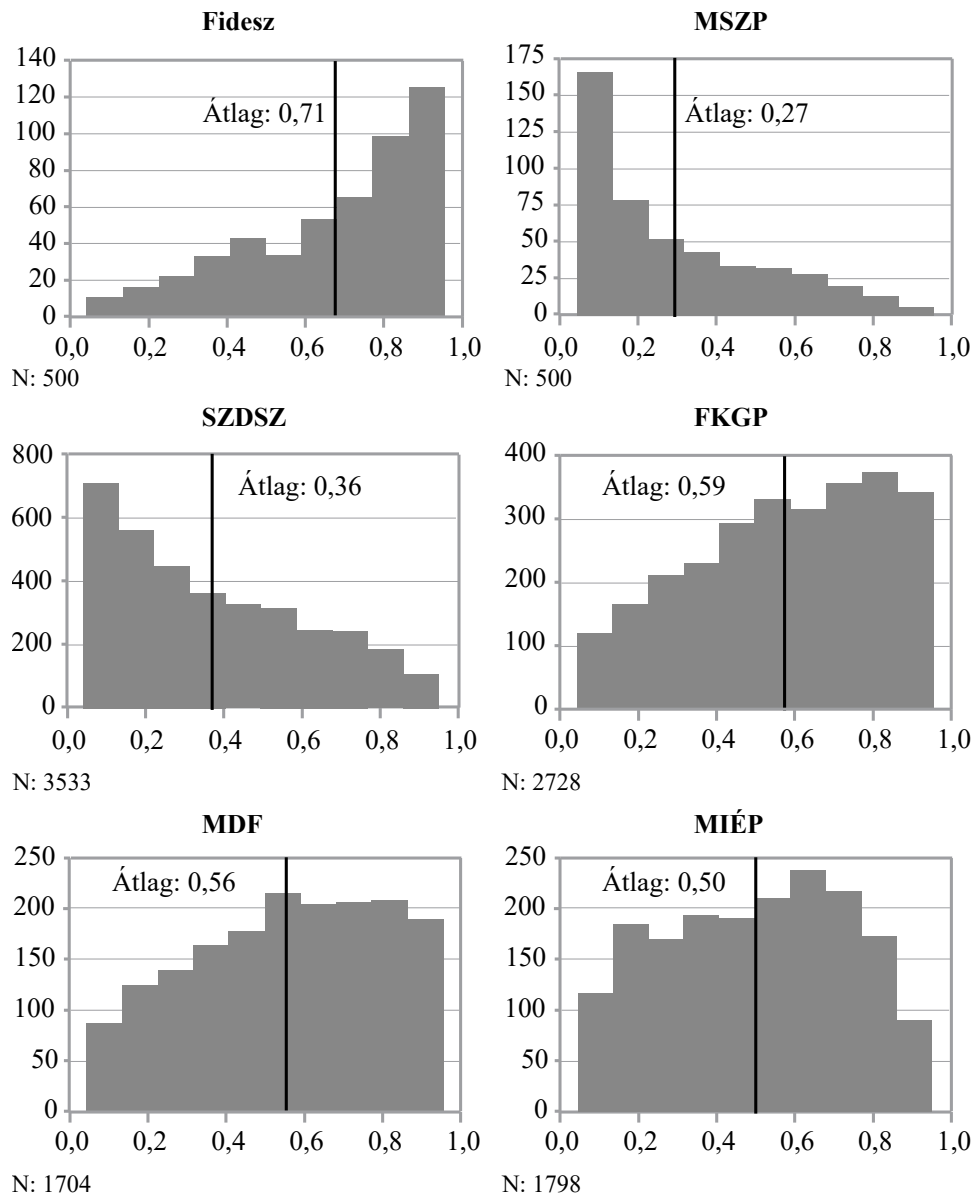
Ezek a kifejezések aláhúzzák, hogy nemcsak tartalmi (ideológiai), hanem eljárási tényezők is befolyásolják az osztályozó működését. A modellek felismerték például, hogy az ellenzéki képviselők gyakran érdeklődnek a kormány tevékenységéről, és az „államtitkár\_úr” formulával szólítják meg őket.

További érdekes következtetéseket vonhatunk le abból, hogy a modellek 0–1 közötti valószínűségi előrejelzései miként alakultak a vizsgált időszakban. A 2. ábra a 3. ciklusra illesztett modelltől származó becslések pártonkénti eloszlását mutatja, míg a 3. ábra ugyanezt szemlélteti a 7. ciklusra illesztett modell esetében. A modellek, ahogy fent rögzítettük, a Fideszhez tartozás valószínűségét becsülik az MSZP-vel szemben, és a két ábrán a Fideszhez, illetve az MSZP-hez tartozó beszédek becsült valószínűségeinek eloszlása azt mutatja, hogy ez a becslés jóval biztosabb a 7., mint a 3. ciklusban, vagyis a polarizáció erősödésével nemcsak a besorolás általános pontossága nő, hanem a párthovatartozás egyedi becslése is határozottabb (fideszes képviselő esetében 1-hez közeli, MSZP-s esetében 0-hoz közeli) lesz. Vagyis egyértelmű kontraszt figyelhető meg a 3. és a 7. ciklust összehasonlítva: a két vizsgált pártra vonatkozó előrejelzések az utóbbi ciklusban jelentősen pontosabbak és szélsőségesebbek. Ugyanakkor ez a határozott különbség csak az utolsó vizsgált ciklusban jelentkezik.

2. ábra

**A 3. ciklus teszhalmazán a 3. ciklusra betanított modell által  
becsült valószínűségek eloszlása pártoként**

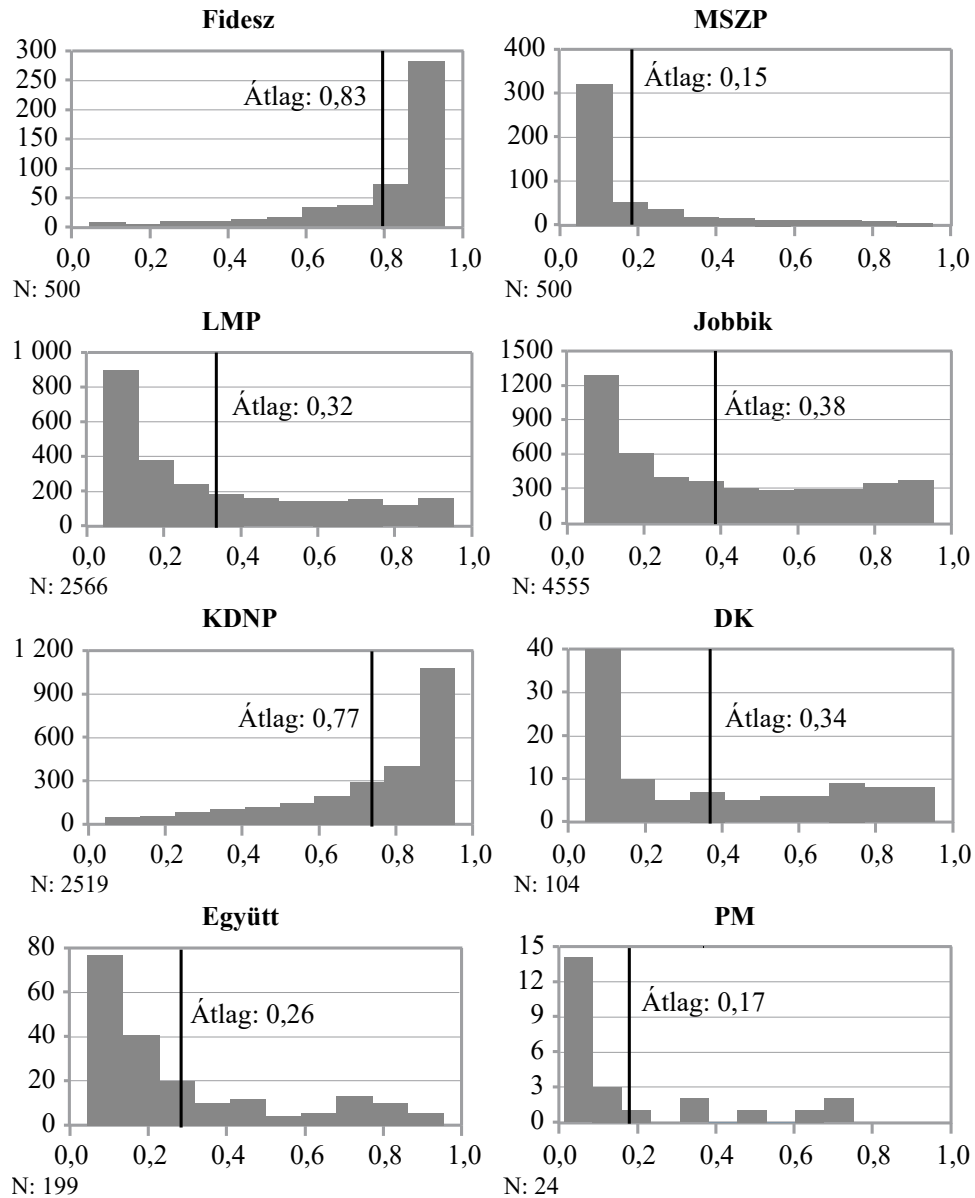
*The distribution of probabilities per party estimated  
by the model trained for Cycle 3 on the test set from Cycle 3*



3. ábra

**A 7. ciklus tesztalmezán a 7. ciklusra betanított modell által becsült  
valószínűségek eloszlása pártoként**

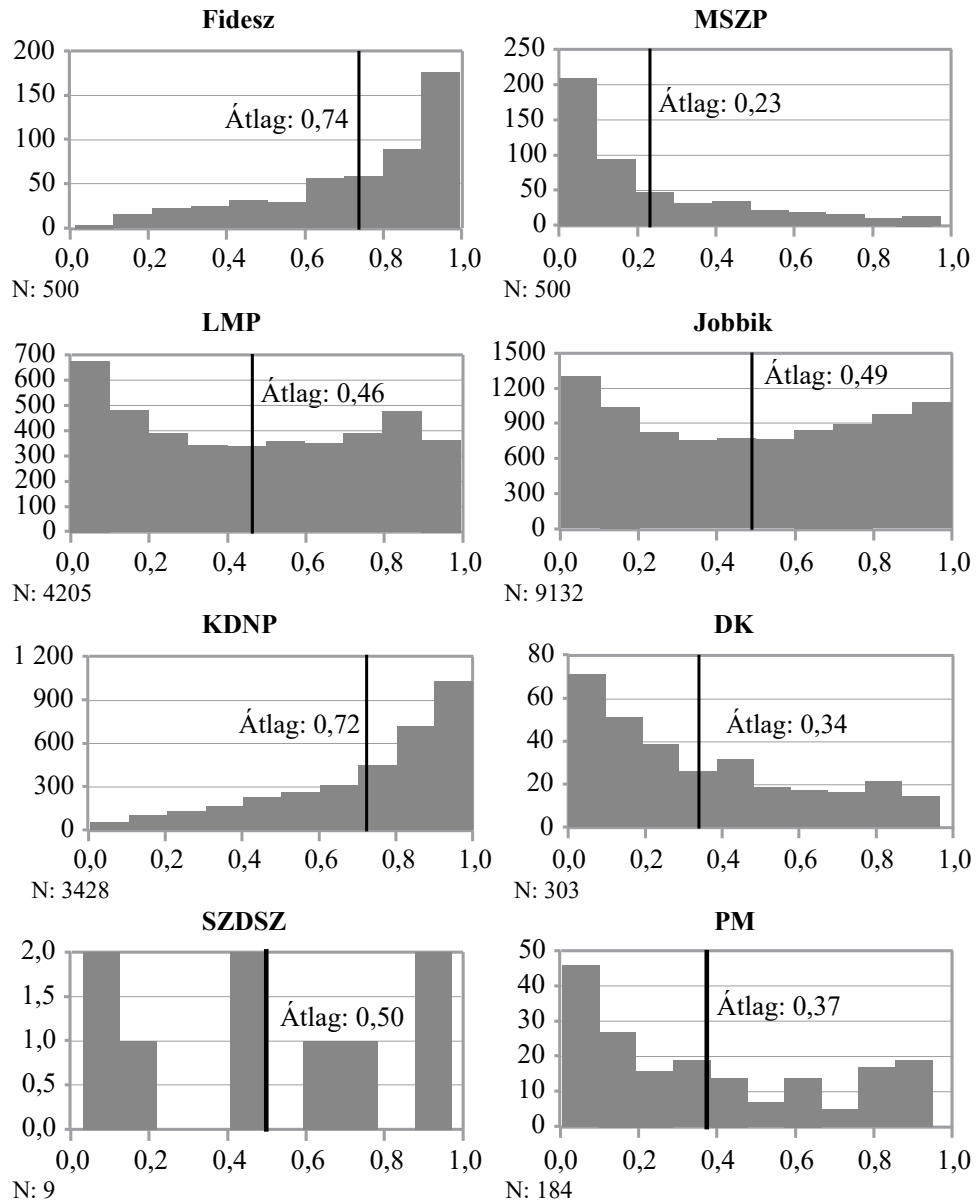
*The distribution of probabilities per party estimated  
by the model trained for Cycle 7 on the test set from Cycle 7*



4. ábra

**A 6. ciklus teszthalmazán a 6. ciklusra betanított modell által becsült  
valószínűségek eloszlása pártoként**

*The distribution of probabilities per party estimated  
by the model trained for Cycle 6 on the test set from Cycle 6*



A más pártok képviselőinek beszédjeihez rendelt becslések eloszlását vizsgálva is érdekes megfigyeléseket tehetünk. Megmutatkozik e pártoknak a Fidesztől, illetve az MSZP-től való távolsága is a polarizációs skálán: például a 3. ciklusban az MSZP-s becslésekhez az SZDSZ-esek vannak legközelebb eloszlásukat tekintve, a 7. ciklusban a fideszes becslésekhez a KDNP-s becslések. Az eredmények robusztusságát hangsúlyozza, hogy eszerint a modellek képesek megkülönböztetni, hogy koalíciós partnerként mely más pártok voltak hatalmon, és a becslések eloszlása az ideológiai spektrumról is mondhat valamit, hiszen a 3. ciklusban a liberális pártnak számító SZDSZ ellenzékben van, a hozzá tartozó előrejelzett valószínűségek átlaga az MSZP és a Fidesz között, de az összes jobboldali párt eredménye alatt van. Ezt támasztja alá az is, ha megvizsgáljuk, hogyan alakult ezeknek a becsléseknek az eloszlása a 6. és a 7. ciklus során. Míg a 6. ciklusban mind az LMP-s, mind a jobbkios képviselők beszédeire adott becslések átlaga közel volt az 50%-hoz, addig ugyanezek az átlagok a 7. ciklusban már jóval alacsonyabbak, ami azt mutatja, hogy ez a két párt a 2010-es kormányváltás utáni első ciklusban megpróbált egyenlő távolságot tartani a „hagyományos ellenzék” és a Fidesz között, míg az ezt követő, 2014-ben kezdődő ciklusra már erősebben ellenzéki pozíciót vettek fel.

#### 4. Összefoglalás

Kutatásunkban a felügyelt gépi tanulási algoritmusok társadalomkutatási felhasználhatóságát mutattuk be szöveges adatokon. Saját esettanulmányként a magyar országgyűlésben 1998–2018 között elhangzott beszédeken végeztünk elemzést, öt parlamenti ciklusra illesztettünk egyenként osztályozó (XGBoost-) modellt, csak a Fidesz és az MSZP szónokainak megkülönböztethetőségét vizsgálva. Az osztályozás ebben az összefüggésben olyan módszer, amely a szerző politikai álláspontját próbálja azonosítani az általa használt szavak alapján. Vagyis a politikai polarizációt osztályozási problémának tekintettük, ahol a gépi tanuló magas predikciós hatékonysága arra utal, hogy az egyik politikai oldal által használt nyelv egyrészt valamilyen szinten homogén, másrészt felismerhető mintázatok szerint különbözik a másik oldal által használt nyelvtől.

A vizsgált öt parlamenti ciklusban időben előrehaladva egyre jobb hatékonyságú predikciót mutató modelleket kaptunk (a pontosság 0,78-ról 0,9-re nőtt), vagyis a politikai polarizáció erősödő tendenciáját tapasztaltuk, azaz eredményünk



szerint a politikai polarizáció egyre jobban detektálhatóvá válik a parlamenti beszédek nyelvezetében. Fontos felhívni ugyanakkor a figyelmet arra, hogy már a vizsgált időszak elején magasnak mondható a kapott 0,78-os pontosság, vagyis már az 1998–2002 közötti időszakban is karakteres különbséget talált az automatizált elemző a két nagy párt beszédei között.

Az eredmények értelmezéséhez magyarázó gépi tanulási eszközöket használtunk. Az osztályozást leginkább befolyásoló kifejezések alapján megállapítható, hogy ez a növekvő polarizáció nemcsak tartalmi (ideológiai), hanem legalább részben eljárási/procedurális tényezőkben is megfogható (lásd: az ellenzéki képviselők gyakran érdeklődnek a kormány tevékenységéről). A két nagy párt kontrasztjára épített modell becsült valószínűségeinek eloszlását is vizsgáltuk, nemcsak a két nagy párt, hanem a kisebb pártok képviselőinek beszédére is. Utóbbi eredményeink szerint a modellek képesek megkülönböztetni, hogy koalíciós partnerként mely más pártok voltak hatalmon, illetve a becslések eloszlása az ideológiai spektrumról is képet adnak, ami robusztusságukat és érvényességüket is alátámasztja.

Az itt bemutatott eredmények csupán első lépését adják egy átfogó, hosszabb távú kutatásnak. A jövőben többféle megközelítésben tervezzük annak a kérdésnek a pontosabb megválaszolását, hogy milyen eszközökkel lehetne a pártok eltérő státuszából (kormánypárti/ellenzéki) adódó nyelvhasználati különbségeket szétválasztani a társadalomkutatási szempontból lényegesebb, ideológiai eltérésekből adódó tartalmi különbségektől. Tervezzük továbbá az előrejelzett valószínűségek alapján kiemelten polarizált/legkevésbé polarizált beszédek és képviselők azonosítását, osztályozását, kvalitatív vizsgálatát.

## 5. Diskusszió

Eredményeink reményeink szerint meggyőzően támasztják alá, hogy a gépi tanulás nemcsak a nagy korpuszok automatizált felcímkézését teszi lehetővé, de további inspiratív lehetőségei is vannak a társadalomkutatás számára. A tanulás imodell predikciós teljesítménye sok esetben önmagában is értelmezhetőséggel bír (lásd itt: polarizációmérték), míg a modellek fekete dobozának felnyitása, azaz interpretálásuk, vagy az előrejelzett valószínűségek vizsgálata új szakterületi tudást hozhat az adott probléma kapcsán.

Ezek az automatikus címkézéssel túlmutatató társadalomkutatási lehetőségek ritkán jelennek meg a tudományos nyilvánosságban. Így például *Macanovic (2022)*

az NLP társadalomkutatási lehetőségei kapcsán kizárólag a kézi címkézés gépiesítéséként írja le a módszert. Nem említi a modell interpretációjának lépését, aminek pedig igen fontos hozadéka lehet a társadalomkutató számára is, lásd a megkülönböztetni kívánt csoportok eltérő nyelvezetére, keretezésére, tematikájára vonatkozó potenciális, új információkat, amit egy ilyen interpretáció ad. Ennek oka talán az, hogy mivel maga az interpretáció/magyarázat fontossága is viszonylag újkeletű (az explainable AI/interpretable AI néhány éve meglevő törekvések), kevéssé ismertek még a gépi tanulási modellek értelmezésében rejlő társadalomkutatási lehetőségek is. Reményeink szerint tanulmányunk láthatóbbá teszi ezeket a lehetőségeket, és további kutatásokat inspirálhat ezen a területen.

## Irodalom

- Bayram, U. – Pestian, J. – Santel, D. – Minai, A. A. (2019): What's in a word? Detecting partisan affiliation from word use in congressional speeches. *International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, 14–19 July 2019, 1–8.
- Eisenstein, J. (2019): *Introduction to Natural Language Processing*. MIT Press, Cambridge.
- Evans, J. A. – Aceves, P. (2016): Machine Translation: Mining Text for Social Theory. *Annual Review of Sociology*, 42(1), 21–50.
- Gentzkow, M. – Kelly, B. – Taddy, M. (2019): Text as data. *Journal of Economic Literature*, 57(3), 535–574.
- Ignatow, G. – Mihalcea, R. F. (2017): *An Introduction to Text Mining: Research Design, Data Collection, and Analysis (1st edition.)*. SAGE Publications, Inc. Thousand Oaks.
- Macanovic, A. (2022): Text mining for social science –The state and the future of computational text analysis in sociology. *Social Science Research*, 108(102784).
- Molnar, C. (2020): *Interpretable Machine Learning*. Lulu.com, Morrisville, NC.
- Németh R. (2021): *Az okság alternatív fogalmi és módszertani megközelítései a szociológiában*. Savaria University Press, Szombathely.
- Németh, R. (2023): A scoping review on the use of natural language processing in research on political polarization: trends and research prospects. *Journal of computational social science*, 6(1), 289–313.
- Németh R. (2024): *Az automatizált szövegfeldolgozás szociológiai lehetőségei*. Savaria University Press, Szombathely.
- Peterson, A. – Spirling, A. (2018): Classification accuracy as a substantive quantity of interest: Measuring polarization in Westminster systems. *Political Analysis*, 26(1), 120–128.
- Rudin, C. – Radin, J. (2019): Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review*, 1(2), 10–1162.
- Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory* (2nd ed.). Springer-Verlag, New York.