

Kmetty Zoltán – Boros Krisztián – Dömsödi Lilla Bíborka –  
Knap Árpád – Mogyorósi Pálma

*Miniszterelnöki csata az online térben*

Bevezetés

A Fidesz–KDNP az ellenzéki összefogás pártjaihoz képest 16 százalékos előnnyel nyerte meg a 2022-es magyarországi választásokat. A nagy különbség nem volt látható a választás előtti nyilvános közvéleménykutatásokban, sőt, volt olyan választás előtti kutatás (Publicus), ami fej-fej melletti versenyt jósolt. Ha az összes nyilvánosan megjelent mérést nézzük, akkor megközelítőleg hét százalékos Fidesz előny volt látható (lásd M1 ábra a mellékletben), ami több mint 10 százalékkal kisebb, mint ami végül a valóságban realizálódott. Bár fontos és vizsgálendő kérdés, hogy miben hibáztak a hazai közvéleménykutatók, jelen tanulmányunk nem erről fog szólni, hanem arról, hogy az online térben volt-e jele annak, hogy ekkora különbség lesz a választás végén a tömbök között, és hogy volt-e olyan trendfordulat az online térben látható választói hangulatban, ami előrevetítette volna az eredményt.

A 2022-es választás két nagy szembenálló tömbről szólt és a tömbök élén álló miniszterelnök-jelöltekről. Orbán Viktor jelölti pozíciója három sikeres választás után nem volt kétséges, a kihívó személye azonban sokáig nem volt egyértelmű. Az ellenzéki összefogás pártjai egy előválasztáson döntöttek jelöltjükéről, amit Márki-Zay Péter nyert meg. Bár papíron pártlistákra lehetett szavazni országosan, a gyakorlatban a két jelölt vitte a kampányt. Ez nem meglepő, korábbi kutatások is arra mutattak rá, hogy a magyar politikai rendszer erősen vezetőközpontú. Tóka (2006) szerint nemzetközi és európai összehasonlításban a magyar polgárok politikai viselkedésének sajátossága a pártok vezető alakjaihoz való erős kötődés. Körösnéyi (2017) Weber plebisziter vezérdemokrácia elméletét ülteti át a 2010 utáni magyar rendszerre, melynek fontos összetevője a demokratikus legitimitáció, ami a vezér karizmájában, a vezér és a nép közötti kapcsolatban ragadható meg. Kiss (2015) szerint pedig napjainkra a pártelv gyengülni látszik, és a magyar politikai elitben kialakult egy általános vélekedés, miszerint a politikai sikerekhez elengedhetetlen egy karizmatikus vezető jelenléte. Az idézett kutatásokra építve, a tanulmányunkban a tömbök küzdelmét az online térben a két jelölt versengésével fogjuk vizsgálni.

A tanulmányunk fő kérdésselvetése, hogy a jelöltek népszerűségét, illetve népszerűségének esetleges változását lehetett-e látni az online térben a velük kapcsolatban megjelent cikkek alá érkező kommentek politikai színezetének és érzelmi töltetének vizsgálatával. Az online térben azokra az oldalakra fókuszálunk, amelyek politikai dimenziók mentén említették a jelölteket. Ilyenek a hírportálok, a politikai megmondó emberek oldalai, vagy akár a pártok és jelöltek oldalai. A vizsgálat azonban nem csak a kommentekre fókuszál. Írásunk töredékes lenne, ha nem mutatnánk be azt, hogy az online politikai térben milyen diskurzus vette körbe a jelölteket a cikkek<sup>44</sup> szintjén.

<sup>44</sup> Az egységes kezelés miatt tanulmányunkban következetesen cikknek hívjuk a Facebook posztokat is. Ezeknek a tartalmaknak a fő ismertető jege, hogy meg lehet őket osztani és lehet rájuk kommentelni.

Tanulmányunk céljait nem titkoltan azzal a szándékkal jelöltük ki, hogy megvizsgáljuk, mennyiben mutatnak más képet az online térben mérhető „preferenciák” (kommentek politikai színezete) a közvéleménykutatásokhoz képest, és alkalmasak-e ezek az adatok arra, hogy trendfordulatokat azonosítsunk a választói viselkedésben. Kutatásunk illeszkedik abba a hazai szakmai térben is felfutó tradícióba (Fokasz és munkatársai 2015; Bene 2018; Bene – Szabó 2019; Koltai – Stefkovics 2018, Szabó – Kmetty–Molnár 2021; Kmetty – Knap 2022), amely digitális viselkedési adatokra épülve (Kmetty 2018) próbál választ adni politikai-szociológia kérdésekre. Kutatásunk hangsúlyosan kvantitatív jellegű, de nem survey alapú, hanem az online térben megjelent, elsősorban szöveges adatok feldolgozására épít, különböző szövegelemzési és szövegbányászati megoldásokra támaszkodva (Németh – Katona – Kmetty 2020).

Írásunk első felében bemutatjuk azt, hogy az online adatokra épülve milyen sikeres és sikertelen nemzetközi (elsősorban amerikai) kísérletek voltak arra, hogy választási eredményeket becsüljenek meg. A szakirodalmi fejezet második része a hazai kutatási előzményekre fókuszál.

A kutatási megközelítések, adatok és módszerek bemutatása után két szinten vizsgáljuk meg a miniszterelnöki csata online lenyomatát – a narratív keretek mentén és a kommentek szintjén. A tanulmány utolsó, záró fejezetében pedig kontextualizáljuk a kapott eredményeinket és bemutatjuk a kutatási megközelítésünk korlátait.

## Szakirodalmi háttér

### Választási előrejelzés digitális lenyomatok alapján – nemzetközi kitekintés

A nemzetközi gyakorlatban már a 2000-es évek második felében megjelentek azok a próbálkozások, amelyek digitális adatok (elsősorban Twitter) alapján próbálták megbecsülni az egyes választások végeredményét. A legtöbb munka valamilyen szentiment elemzésre épített. A szentiment elemzés gyakorlatilag a tartalmak pozitív/negatív dimenziókba való elhelyezését jelenti. Tweetek esetén ez annak az eldöntése, hogy a jelöltet említő tweet pozitív vagy negatív érzelmi dimenzióban íródott-e.

A korai munkák kritikájaként kiemelhetjük Gayo-Avello és társai (2011) tanulmányát, amiben megkérdőjelezi a Twitter adatok predikciós erejét. A 2010-es amerikai kongresszusi választások tweetjeinek szentiment alapú elemzése alapján rámutatnak, hogy a korábbi tanulmányokban felmutatott modellek nem adnak megfelelő előrejelzést a választások eredményeire vonatkozóan. A szerzők mellett érvelnek, hogy nincs elegendő tudásunk a felhasználók demográfiai háttéréből adódó torzításról, míg a kérdőíves módszerek a megfelelő mintavétellel ki tudják küszöbölni ezt a problémát.

Ez a negatív eredmény nem vette el a szakma kedvét a próbálkozástól, a 2010-es években számos kutatás készült a témában. Jahanbakhsh és Moon (2014) Twitter bejegyzéseket használt fel a 2012-es amerikai elnökválasztás eredményének prediktálására, és előrevetítette a gépi tanuláson alapuló módszerek felhasználásának lehetséges jövőbeli potenciálját. A kifejlesztett szentiment-osztályozó eszköz által nyújtott szentiment-analízis és földrajzi címkék segítségével a kutatók sikeresen mérték be az elnökjelöltek népszerűségét az egyes

államokban. A Twitter szentimentek alapján Barack Obama volt vezető szerepben, ami összezsengett a választások végső kimenetelével (Jahanbakhsh – Moon 2014).

A Twitteren megjelenő, a választópolgárok által egyre növekvő mértékben generált adatmennyiség termékeny táptalajt nyújtott a módszer predikciós erejének tesztelésére, a következő amerikai elnökválasztás kapcsán is. A témában született elemzések a tweetek szentiment-analízisére alapozva vontak le következtetéseket a 2016-os amerikai elnökválasztás eredményeire vonatkozóan (Ramteke és mtsai 2016; Yaqub és mtsai 2017; Joyce és mtsai 2017; Caetano és mtsai 2018). A szentiment-analízis, valamint a felhasználói reakciók elemzése mellett a közösségi média lehetőséget adott a nyilvánosságban terjedő vélemények feltérképezésére is. Yaqub és munkatársai (2017) elemzése arra is rámutatott, hogy a győztes jelölt optimistább kampányüzenetet közvetített, és pozitívabb szentimentekkel rendelkezett az említésekben, mint a vesztes jelölt. A 2020-as elnökválasztás eredményei és a közösségi médiában kifejeződő hangulat összehasonlításából szintén az derült ki, hogy a kettő a legtöbb esetben egybeesik. Abban a négy államban is, amelyben a szentiment-eredmények nem egyeztek meg a választási eredménnyel, a hosszútávú trendek a győztes jelölt pozitív megítélésének növekedését, és a vesztes jelölt pozitív megítélésének visszaesését mutatták (Chaudhry és mtsai 2021).

A Twitter adatok választási predikciójára az említett elemzéseken túl többféle elméleti megközelítés, valamint módszertani megoldás született. Liu és társai (2021) kritizálták az eljárást, amelyben a választási eredményeket közvetlenül és kizárólag a Twitteren megállapítható hangulattal korrelálják. Ennek kiküszöbölésére, egy elméletileg megalapozottabb megközelítés érdekében, hagyományos, politikatudományban használatos modellekbe építették be a Twitter szentiment adatot, amelyben a választási eredményt a Twitter-hangulatból következő támogatási aránnyal és gazdasági növekedést jelző mutatókkal magyarázták.

Burnap és társai (2015) egy átfogó modell hiányát kifogásolták a választások előrejelzéséhez felhasznált közösségi média-adatok felhasználásában. Ezt a hiányt egy olyan előrejelzési modell felállításával igyekeztek betölteni, amely a szentiment-analízist és a pártok korábbi támogatottságát veszi figyelembe a parlamenti helyek eloszlásának valódi predikciója érdekében. A modellt a 2015-ös Egyesült Királyság-beli választások kimenetelének előrejelzésén sikeresen tesztelték. Wang és társai a 2017-es francia választásokat vizsgáló esettanulmányukban a semleges szentimentű tweetek figyelembevételével növelték a predikciós pontosságot (Wang és mtsai 2017). Ezeken túl, többek között Spanyolországban (Soler és mtsai 2012), Indiában (Singhal és munkatársai 2015) és Indonéziában (Budiharto és mtsai 2018) is felhasználták a politikai kampányra válaszként generálódó nagy mennyiségű Twitter adatokat a választások eredményeinek előrejelzésére.

Bár a kutatások egy része ígéretes volt az eredmények szempontjából, a törekvések egyes eredményét erősíti Koltai és Stefkovics (2018) tanulmányában készített metaelemzés, amelyben azt találták, hogy a 2008 és 2018 közötti időszakból, a nemzetközi szakirodalomból összegyűjtött 40 tanulmányból csupán 25-ben voltak a közösségi média adatok jó indikátorai a politikai preferenciáknak. A maradék 15 esetben csak gyenge, vagy egyáltalán nem megbízható eredmények születtek (Koltai – Stefkovics 2018: 112–114).

### Kedvelések és megosztások – hazai kutatási eredmények

A nemzetközi kutatások elsősorban Twitter adatokra építették a választási előrejelzéseiket. Magyarországon a Twitter elterjedtsége nagyon alacsony, ezért ez a közösségi média forrás nem alkalmas itthon ilyen jellegű kutatásokra. Ezzel szemben a Facebookot gyakorlatilag a teljes internetező lakosság használja<sup>45</sup>, ebből következően a hazai kutatások elsősorban ezt a közösségi média platformot vizsgálták választási előrejelzés szempontjából. A hazai kutatások abban is eltérnek a nemzetközi gyakorlattól, hogy alapvetően nem a tartalmak szentimentjére fókuszálnak, hanem inkább a kedvelési és megosztásai mintákra.

Koltai és Stefkovics (2018) pártok és vezetők Facebook kedveltségi adatait vetették össze közvéleménykutatási eredményekkel. Tanulmányuk szerint sem a pártok, sem a pártvezetők Facebook oldalának kedveltsége nem jó indikátora a pártok valódi támogatottságának. A szerzők szerint ez annak köszönhető, hogy ezek az értékek inkább statikusak, amelyek nem követik dinamikusan a pártpreferencia-változásokat. Az adott politikai szereplők (pártok és pártvezetők) bejegyzéseire érkezett kedvelések havi átlaga jobb közelítést adott a közvéleménykutatásokban mért támogatottsághoz. A 12 hónapból a kutatók 4-ben megegyező eredményre, további 3 esetben pedig hasonló eredményre jutottak a támogatottságot illetően, mint az adott időszakra vonatkozó közvéleménykutatások. A Facebook felhasználók demográfiai torzítását ellensúlyozni hivatott súlyozás nem javított az eredményeken. A kutatók a bevont indikátorok nagy fokú ingadozására hívják fel a figyelmet, ezzel utalva arra, hogy a Facebook adatok inkább hosszútávú, mintsem rövidtávú elemzések használatára alkalmasak, valamint a kvalitatív elemzési megközelítés, továbbá topikmodellezés és szentiment-elemzés szükségességét hangsúlyozzák a kvantitatív mutatók kiegészítéseként (Koltai – Stefkovics, 2018).

A magyar politikai tér közösségi médián való leképeződésének más megközelítését alkalmazza elemzésében Bene Márton (2018), amikor a felhasználók aktivitását vizsgálja a politikai szereplők közösségi média posztjaival összefüggésben. 183 egyéni választókerületi képviselő-jelölt Facebook-bejegyzéseinek vizsgálatánál azt találta, hogy az egyéni választókerületek induló jelöltjeinek Facebook bejegyzéseire érkező átlagos kedvelés- és kommentszámok nincsenek, míg az átlagos megosztásszámok összefüggésben vannak a választásokon elnyert támogatással. A szerző ezt a kétlépcsős hatás elméletével írja le, mely szerint az elnyert többlépcsős támogatások a képviselőjelöltek bejegyzéseit megosztó véleményvezéreken keresztül a nem stabil szavazókhoz eljutó üzeneteknek köszönhetőek.

### A médiatér fragmentáltsága

Az online térben a megosztások és a kommentek nem egy üres térben keletkeznek, a teret különböző szereplők alakítják, a média szakemberektől a politikai spin-doktoroktól át az algoritmusokig (Thorson-Wells 2016). Ezek a szereplők befolyásolják azt, hogy milyen narratívák jutnak el végül a választókig. A hírterjedést azonban a szereplők szándékolt vagy nem szándékolt intenciói mellett a médiatér fragmentáltsága is befolyásolja.

<sup>45</sup> Egy 2022 tavaszán végzett online reprezentatív kutatásunk alapján, az internetezők 87 százaléka használja heti rendszerességgel a Facebookot, míg a Twitter esetében ugyanez az arány 10 százalék.

A hazai médianyilvánosságot vizsgáló kutatások arra mutatnak rá, hogy a társadalomban jelenlévő magas fokú politikai polarizáltság ellenére (Kmetty – Tardos 2022) a magyar médiater politikailag közepesen fragmentált. Szabó és Bene (2016; 2019) három 2014-es belpolitikai eseményen keresztül, hálózatelemzés segítségével tárták fel a hazai médianyilvánosság szerkezetét. Noha a világnézeti szempontból eltérő csatornák egyértelműen más interpretációt közvetítettek közönségüknek, a kutatók nem tekintették a magyar nyilvánosságot erősen polarizáltnak. A média szerkezetében inkább centrum-periféria jellegzetességeket tártak fel. Az eredményeket a 2018-as választásokat megelőző időszak alatt online hírportálokat vizsgáló elemzés is megerősíti, amely alapján a centrum pozíciót elfoglaló médiumok anyagai a saját olvasótáborukon túlmenően is utánközlésre kerülnek. Szabó (2018, 401.) alapján az online médianyilvánosság szerkezetében „(...) *egy stabil, nagyfokú állandóságot mutató, a portálok figyelmét nagymértékben kisajátító és elsősorban egymásra hivatkozó centrum, és egy, a centrumhoz kapcsolódó, elsősorban rá hivatkozó periféria jelenik meg*”.

A médiumok ideológiai tömbösödéséből következően a magyar választópolgárok tájékozódása is politikai irányvonalak mentén szelektált. A magyar, Facebookon aktív választók tájékozódási környezetét feltáró kutatások összegzéséből kiderül, hogy a választók egy jelentős csoportja egyoldalú, homofil információs közegben létezik (Janky – Kmetty – Szabó 2019). A pártos tájékozódás főként az elsődlegesen televízióból tájékozódókra vonatkozatható, az online térben kisebb az elszigetelődés valószínűsége. A pártpreferenciákon keresztül tehát a tájékozódási források, és így a közéleti kérdésekben alkotott álláspontok összefüggenek. A médiabuborék mellett a hazai kutatások ugyanakkor azt is mutatják, hogy még a politikailag megosztott médiater ellenére is a magyarok többsége nyitott az eltérő nézőpontok meghallgatására, ha közéleti tájékozódásról van szó (Janky – Kmetty – Szabó 2019). Az ideológiailag meghatározott tömbök jelenléte miatt az elemzésünkben indokoltnak tekintjük a kormánypárti és kormánykritikus médiumok elkülönítését, a politikai dimenzió kiemelt elemzését (Kiss – Szabó 2019).

### Megközelítések, adatok és módszerek

A kutatásban felhasznált adatok legyűjtéséhez a Sentione szolgáltatását használtuk. A cég több mint 100 ezer weboldalt monitoroz a hazai online térben és hozzáférést nyújt az elmúlt 3 évben keletkezett publikus online tartalmakhoz. Az online térben keletkezett adatokat vizsgáló kutatásoknál az adatgyűjtés előre meghatározott kulcsszavak és/vagy weboldalak alapján történik, így csak azokat a tartalmakat gyűjtjük le, amik tartalmazzák a szóban forgó kulcsszót, vagy az adott weboldalon született a cikk. A kampányban megjelent teljes politikai tartalom mennyiség legyűjtése helyett egy sokkal kontúrosabb adatgyűjtési megközelítést választottunk. A két miniszterelnök-jelölt, Orbán Viktor és Márki-Zay Péter online megjelenéseit gyűjtöttük le. Azért is döntöttünk a jelöltekre fókuszáló gyűjtés mellett, mivel a kampány-perszonalizáció miatt a pártok helyett egyre inkább a politikusok kerülnek a fókuszba, illetve Magyarországon a politikai vezető személyek szerepe kiemelten fontos (Kiss 2015).

Elemzésünkhöz minden olyan tartalmat letöltöttünk a Sentione rendszeréből, amelyek említették Orbán Viktor vagy Márki-Zay Péter nevét<sup>46</sup>, illetve valamelyik politikus Face-

<sup>46</sup> Kulcsszólista: orbán\*, orbán viktor\*, márki-zay péter\*, márki-zay\*, „márki zay”, mzp\*

book oldalán született az előválasztás másnapja (2021. október 17. 00:00) és az országgyűlési választást megelőző nap (2022 április 3. 00:00) által keretezett időszakban. Az így létrejött adatbázis online újságcikkeket, fórumbejegyzéseket és ezekhez érkező hozzászólásokat, valamint publikus Facebook<sup>47</sup> posztokat és ezekre érkező kommenteket tartalmaz. A Facebook posztok esetében az adatbázisban látjuk az azokra érkezett reakciókat is (pl.: like, megosztás). Az oldalak között megtalálhatók híroldalak, politikai megmondó emberek oldalai, valamint pártok és politikusok oldalai is. Ha egy oldalnak volt „hagyományos” internetes megjelenése is és Facebook oldala is, mindkettőt bevettük az elemzésbe. Tehát például bekerült a Telex híroldal és a Telex Facebook oldala is. A tartalmi elemzés előtt előfeldolgoztuk a letöltött cikkeket<sup>48</sup>.

Az elemzésben alapvetően kormánypárti/kormánykritikus dimenzióban vizsgáltuk a tartalmakat. Ehhez szükségünk volt az oldalak ilyen szempontú csoportosítására. Ehhez már rendelkezünk egy saját kiinduló listával, amit kibővítettünk az összes olyan oldallal, ahol legalább 100 cikk megjelent, vagy megosztás szempontból a 100 legtöbb megosztást érte el az oldal<sup>49</sup>. Ha egy oldalról nem tudtuk eldönteni a politikai színezetét (például egy életmód oldal esetében), határon túli oldal (pl.: parameter.sk), hírgyűjtő oldal, vagy konspirációs teóriákat megosztó oldal volt, akkor kivettük az elemzésből. Az összes leggyűjtött tartalom 94 százaléka volt bekódolva általunk és a tartalmak 67 százaléka került be a kormánypárti vagy kormánykritikus csoportokba. Ez összesen 140 864 cikket jelent, 617 különböző oldalról, amelyeken belül csekély többségben vannak a kormánykritikus oldalak (54 százalék). A leggyűjtött cikkek megosztásainak 87 százalékáért „feleltek” a bekódolt kormánypárti/kormánykritikus oldalak.

A vizsgálati keretünk a hazai politikai online médiatér volt. A módszerünkben adódóan nem lehetünk abban biztosak, hogy a leggyűjtött tartalmak jól reprezentálják ezt a politikai médiatérrel. Ezért elvégeztük az elemzéseket egy olyan szűkített médiatérben is, amiben adatbázisunkban a megosztások alapján a legnagyobb 5-5 kormánypárti és kormánykritikus média szerepelt csak (Facebook oldalukkal együtt). Ezek kormánypárti oldalról a

<sup>47</sup> Az adatgyűjtés során bekerültek Youtube videók, tweetek, Instagram posztok és TikTok tartalmak is az adatbázisunkba. Mivel ezek aránya a teljes tartalom mennyiséghez képest elenyésző és feldolgozás szempontjából eltérnek a cikkektől és Facebook tartalmaktól, ezért ebben a tanulmányban ezeket nem elemeztük.

<sup>48</sup> Mivel a nyers szöveges adatok nehezen elemezhetőek, zajosak és sok irreleváns elemet tartalmaznak (pl. duplikációk és felesleges tartalmak), ezért az adatok a tartalmi elemzés előtt egy többlépcsős, automatizált, R és Python kódokban megírt előfeldolgozási folyamaton mentek keresztül. A tisztítási folyamatnak többek között része a duplikációk, a felesleges tartalmak és a speciális karakterek szűrése. Az előkészítés során a tisztított szöveget jól elemezhető, sztenderd egységekké alakítjuk. A folyamat során az emojiakat szöveggé alakítjuk, a DBpedia (Lehmann et al. 2015) alkalmazással azonosítjuk a névelemeket, majd az e-magyar vonatkozó moduljával lemmatizáljuk a tartalmakat (Zsibrita et al. 2013). A precízebb elemzés érdekében a gyakran ismétlődő szópárokat, az úgynevezett szignifikáns bigramokat, valamint a tagadásokat alsóvonás jellel vonjuk össze, így azok a továbbiakban egy tokenként jelennek meg. A technikai szűrés/előfeldolgozás után LDA topikmodell segítségével tartalmi szűrést is végeztünk a cikkeken és posztokon (Németh – Katona – Kmetty 2020). A topikmodell azonosítja a korpuszban jelenlévő témákat és ezek mentén csoportosítja a szövegeket. Az eredmények alapján kiszűrjük az olyan, a vizsgált témához szorosan nem kapcsolódó tartalmakat, mint a sporttal, vagy a koronavírusjárvánnyal és védőoltással kapcsolatos hírek és a hozzájuk tartozó kommentek

<sup>49</sup> A kódolás pontosságának ellenőrzésére, összevetettük a saját kódolásunkat egy Sik Endre által vezetett másik kutatócsoport politikai kódolásával. A mindkét kutatócsoport által bekódolt 173 kormánypárti oldal esetében 95%-os volt a kódolási egyezőség, a 144 kormánykritikus oldal esetében pedig 92%.

következők voltak: *Origo*, *Pesti Srácok*, 888, *Mandiner*, *M1 Híradó*. Kormánykritikus oldalról pedig a következő oldalak kerültek be: *Telex*, *HVG*, 444, *24.hu*, *ATV*.

A tanulmányunk fő kérdésfeltevése az, hogy lehetett-e látni az online politikai térben a jelöltek általános támogatottságát és a támogatottság időbeli változását. Ezt a kérdést a kommentek vizsgálatával igyekeztünk megválaszolni. A cikkek leválogatása és kormánypárti/kormánykritikus beazonosítása megteremtette a lehetőségét, hogy megvizsgáljuk milyen narratívákat használnak a két oldalon a jelöltek kapcsán és azt is, hogy a vizsgálandó kommentek körét kijelöljük. Ebbe a vizsgálati keretbe azok a kommentek kerültek be, amelyek kormánypárti vagy kormánykritikus cikk kapcsán „születtek”.

A „preferencia” mérésünk indikátorai a jelöltekkel kapcsolatos cikkek alatt megjelent kommentek érzelmi töltete és politikai színezete volt. Ez az a megközelítés, ami a leginkább párhuzamba állítható a Twitterre épülő előrejelző modellekkel. Mivel jelenleg magyar nyelvre nem áll rendelkezésre olyan modell, ami képes lenne érzelem és politikai színezet mentén bekezelni a kommenteket, ezért egy saját modellt fejlesztettünk ki. Ehhez egy úgynevezett felügyelt kategorizációt végeztünk. A felügyelt kategorizációs megközelítésnek az a lényege, hogy első lépésben bekódolunk a vizsgálandó kategóriák mentén egy nagyobb tartalom mennyiséget és ezt felhasználva gépi kódolással képesek vagyunk olyan tartalmakat is besorolni, amiket „kézzel” nem kategorizáltunk be. Az érzelmek elkülönítése két lépcsőben zajlott. Először 5 tágabb (pozitív, negatív, semleges, nincs érzelem, nem kódoljuk), majd 12 szűkebb (elégedettség, reménykedés, bánat, düh, elégedetlenség, félelem, gúnyolódás, zavartság, együttérzés, érdeklődés, nosztalgia, meglepődés) érzelmi kategória szerint lettek elkülönítve az egyes kommentek. A „nem kódoljuk” kategóriába azok a kommentek kerültek, melyek tartalmuk alapján alkalmatlanok voltak további felhasználásra (pl. linkek, üres kommentek).

A politikai töltet feltárására három lépcsőben alkalmaztunk klasszifikációs modelleket. Az első lépcsőben azt állapítottuk meg, hogy jelen van-e bármiféle politikai vonatkozású információ az egyes kommentekben. Ha volt a kommentben politika, akkor a következő lépcsőben aszerint kerültek besorolásra a tartalmak, hogy magyar vagy világpolitikai vonatkozásúak-e. Az utolsó, harmadik lépcsőben a belpolitikai kommenteket három szempontból soroltuk be: ellenzéki, kormánypárti, vagy egyéb politikai színezet szerint. Természetesen, e három kategória – ahogyan az érzelmek esetében sem – egymást nem feltétlenül kölcsönösen kizáró; viszont a jelen politikai diskurzus ismeretében plauzibilis ez a fajta elkülönítés. Az érzelmek kategorizációjához egy 15 ezer kommentből álló kiinduló adatbázist használtunk, a politikai kommentek esetében pedig 20 ezer bekezelített kommentünk volt. Minden kommentet két egymástól független kódoló kategorizált be egy előre kidolgozott kódolói utasítás alapján. Ahol a kódolók között nem volt egyetértés, ott a kutatás vezetői döntöttek. Ezeket a bekódolt kommenteket használtuk fel a teljes kommenthalmaz gépi klasszifikációjában<sup>50</sup>. A klasszifikációhoz használt modellek teljesítményeit a mellékletben található M2. és M3. táblák foglalják össze.

Elemzésünk két szintet tartalmaz. Első lépésben a cikkeket vizsgáljuk meg aszerint, hogy melyik jelölt szerepel bennük, mikor és hol jelent meg a tartalom. Különböző dimenziók mentén összevetjük a megjelent cikkek főbb narratív kereteit is. A szöveges adatok statisztikai elemzése a Benoit és munkatársai által kifejlesztett Quanteda csomag (Benoit és mtsai 2018) függvényeire támaszkodik. Ez az elemzés mutatja meg, hogy milyen keretben zajlott a politikusokkal kapcsolatos online diskurzus. Az elemzés második lépcsője

a konkrét kutatási kérdéseinkre próbál választ adni, miszerint látható volt-e az online politikai térben Orbán Viktor győzelme. Ehhez az első szinten azonosított cikkek alá érkező kommenteket vizsgáljuk, azon dimenziók mentén (érzelmeik, politikai töltet), amelyeket bemutatunk a korábbi bekezdésben.

Az adatelőkészítés és az elemzés során R és Python programnyelveket használtunk.

## Eredmények

### Cikkek szintje

A két miniszterelnök-jelölt körüli narratív keretek megtalálásához a velük kapcsolatban megjelent cikkeket vizsgáltuk meg. A cikkek egy része hírportálokon jelent meg, másrésze politikai megmondóemberek, pártok, vagy politikusok oldalán. A módszertani részben leírt szűrések után 140 864 db cikk maradt az adatbázisunkban.

Orbán Viktornek sokkal nagyobb volt a médiajelenléte a választási időszakban. A tartalmak 57 százalékában szerepelt Orbán Viktor, és mindössze 27 százaléka volt a cikkeknek és a posztoknak, amiben Márki-Zay neve említésre került. A tartalmak további 16 százalékában mind a két politikusra hivatkoztak. Ez a nagy különbség nem a Fidesz médiafölényével magyarázható. Ha szigorúan számosságot nézünk, a vizsgált média tartalmak 56 százaléka érkezett kormánykritikus oldalról és „csak” 44 százaléka kormánypárti oldalról. A megjelenések oldalak szerinti bontása egy nagyon érdekes képet rajzol ki. A kormánykritikus oldalon a jelöltekkel foglalkozó cikkek 69 százaléka csak Orbán Viktorral foglalkozott, és mindössze a cikkek 15 százaléka volt, amiben Márki-Zay Péter egyedül szerepelt. Ez gyakorlatilag a teljes kampány időszakra igaz volt, egyedül az előválasztás vége utáni napokban produkált Márki-Zay Péter Orbán Viktorhoz hasonló megjelenést a kormánykritikus médiában (1. ábra). A kampányfinanszírozásban még tovább nőtt az olló a két politikus között, Márki-Zay Péterről szinte alig lehetett olvasni a kormánykritikus médiában, szinte kizárólag Orbán Viktorral foglalkozó cikkek jelentek meg.

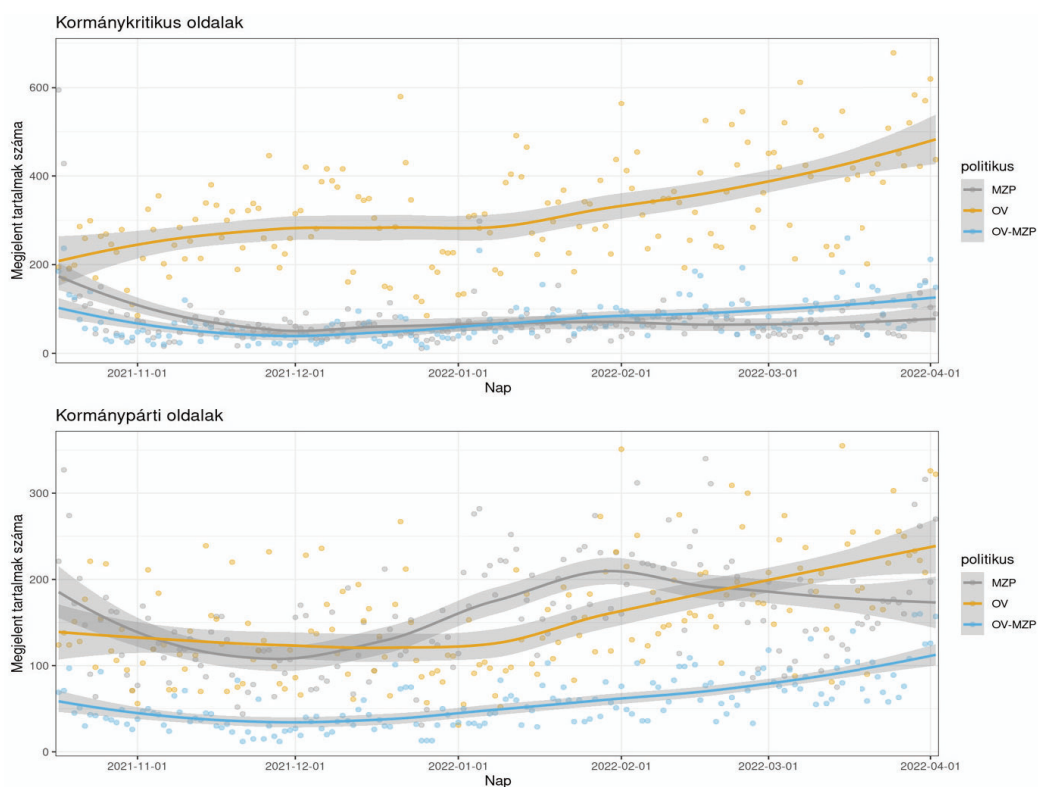
A kormánypárti média teljesen más képet mutatott. A cikkek 42 százaléka volt, amiben csak Orbán jelent meg és 43 százaléka, amiben csak az ellenzéki kihívója, azaz a megjelenési arány nagyon hasonlóan alakult. Időben is nagyon vegyes volt, hogy kivel foglalkoztak

<sup>50</sup> Mind az érzelmeik, mind a politikai vonatkozás elkülönítésére huBERT alapú gépi tanulási modellt alkalmaztunk. A BERT típusú modellek 2018-ban váltak népszerűvé főként a gyorsabb és effektív taníthatóságuk, valamint „state-of-the-art” teljesítményük miatt. A BERT alapú modellek – többek között – lehetővé teszik az úgynevezett transzfer tanulást, ami azt jelenti, hogy képesek átadni (transzferálni) a korábban tanult információkat egy újabb feladat elvégzéséhez (Azunre, 2021:5). Az általunk használt modell egy jellegzetes példája a transzfer tanulásnak. A huBERT modell egy hatalmas korpuszon lett betanítva a magyar nyelvre, majd ezt az előzetes tudást felhasználva tanítottuk be egy kisebb adatbázison a saját igényeink szerint. Tehát a modell „nyelvtudását” kiegészítettük egyik esetben érzelmi, másik esetben politikai tudással. A huBERT a BERT base modellarchitektúráján alapuló, a Magyar Webcorpus 2.0-n betanított transzformer alapú nyelvi modell (Nemeskey 2021, Kmetty 2022). A kutatásunkban használt modell annyiban tér el az eredeti huBERT-től, hogy a bemenő adatok klasszifikálására külső adatokat is felhasznál. A mi esetünkben ez azt jelenti, hogy a modell a kommentek klasszifikálásánál azt is figyelembe vette, hogy az adott tartalom milyen weboldaltól származik. Ez a plusz külső adatforrás főként a politikai klasszifikáció esetében jelenthet értékes információt, hiszen a mai magyar online diskurzusban jelentős a politikai hovatartozás szerinti buborékok kialakulása (Janky és munkatársai 2019).



inkább a kormánypárti lapok, a januári és februári időszakban például egyértelműen inkább Márki-Zay volt a fókuszban (1. ábra). Ha leszűkítjük az elemzést az 5-5 kiválasztott kormánypárti és kormánykritikus híroldalra, akkor ugyanezt az eredményt kapjuk. A kormánykritikus médiában végig túlsúlyban volt Orbán Viktor és a kampányfinisre ez a különbség tovább nőtt a megjelenésekben a jelöltek között. Ezzel szemben a kormánypárti médiában hasonló arányban kerültek be a jelöltekkel kapcsolatos cikkek. Márki-Zay Péterrel közvetlen az előválasztás után és a január-februári időszakban foglalkoztak többet a kormánypárti híroldalak.

1. ábra. A miniszterelnök-jelölteket említő cikkek időbeli megoszlása a kormánypárti médiában

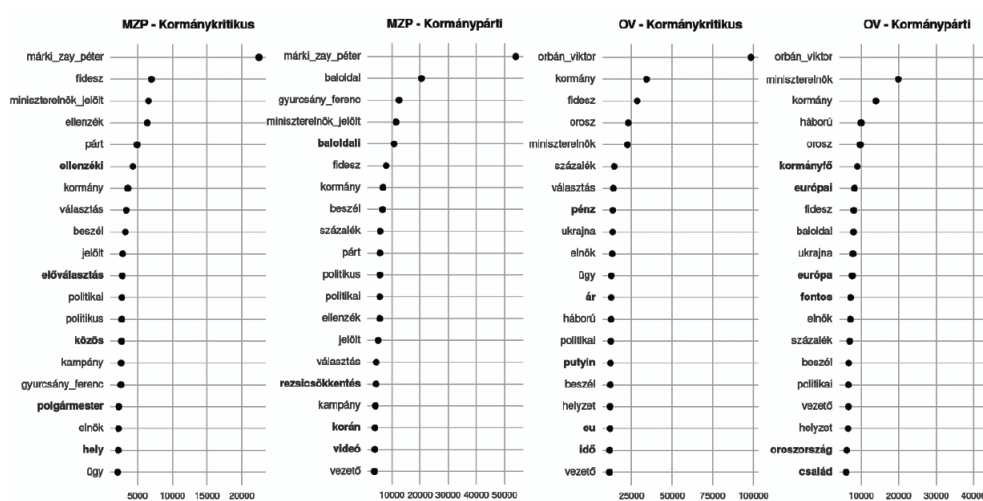


Természetesen a két politikus nagyon eltérő kontextusban jelent meg a kormánypárti és a kormánykritikus médiában. A 2. ábrán a miniszterelnök jelöltekkel kapcsolatos cikkekben a leggyakrabban említett szavakat jelöltük. Vastaggal szedtük azokat a szavakat, amelyek csak az adott jelölt és média oldal kapcsán jelentek meg. Márki-Zay kapcsán a kormánykritikus médiában általános kampánnyal kapcsolatos szavak dominálnak. Azt lehet látni, hogy nincs egy olyan téma sem, ami erősen „ragadt” volna rá. A kormánypárti médiában jobban kirajzolódnak a főbb üzenetek, elsősorban a „baloldal” és Gyurcsány Ferenc említése látványos, ami jelzi azt, hogy a kormánypárti média hangsúlyosan jelenítette meg Márki-Zay vélt politika pozícióját és igyekezett a jelöltet összekötni Gyurcsány Ferencsel. De témaspecifikus tartal-

## 150 Kampánykommunikáció a politika digitális tereiben

mak is felbukkannak, mint például a rezsicsökkentés (eltörlése). Orbán Viktor esetében nagyon más tartalmi fókuszot látunk, egyértelműen kirajzolódik a nagyon erős külpolitikai dimenzió, az Európai Unió, Oroszország és Ukrajna említésével. Érdekes eredmény annak fényében, hogy a kampány időszak hosszabb része tehető a háború kirobbanása előtti időszakra. Ez jelzi azt is, hogy ahogy a háború megindult, nagyon hangsúlyossá vált ez az elem a diskurzusban. Az öt-öt kiválasztott kormánypárti és kormánykritikus lapra leszűkített elemzés gyakorlatilag ugyanezt az eredményt hozta (lásd a mellékletben az M4 ábrát).

2. ábra. Leggyakoribb szavak a miniszterelnök-jelölteket említő cikkekben

*kormánypárti/kormánykritikus bontásban*

Természetesen a háború témája előkerült már a háború kitörése előtt is, akkor a cikkek hét százalékában találtunk ilyen jellegű utalást. Az orosz agresszió elindulása után ez az arány közel 50 százalékra ment fel a miniszterelnök-jelöltekkel kapcsolatos cikkekben, ami jelzi azt, hogy milyen jelentős mértékben dominálta a diskurzust a téma. A keretezés ráadásul nagyon eltérő volt a két politikus esetében, főleg a kormánypárti médiában. Mind Márki-Zay, mind Orbán kapcsán megjelenik hangsúlyosan a „belesodor”, „belesodródik” kifejezés a kormánypárti médiában, de míg Márki-Zay esetében ez például olyan kifejezésekkel párosul, mint a „fegyvert\_küld”, addig Orbán Viktor kapcsán a „kimarad” és a „béke” szavak a hangsúlyosak. A kormánykritikus médiában nincs ekkora különbség a két politikust említő cikkek szóképzésében, a főbb szavak alapján inkább a háború alakulása volt a középpontban, és nem feltétlenül a magyar pozíció a háború kapcsán. Ahogy előzetesen vártuk, a keretezés nagyon meghatározóan jelent meg a jelöltek kapcsán a magyar médiában, főleg a kormánypárti lapokban. Az elemzés következő lépésében azt vizsgáljuk meg, hogy a megjelent tartalmak alatt milyen komment klíma alakult ki.

### Kommentek szintje

Az elemzés második vizsgálati dimenziója a kommentek világa. A miniszterelnök-jelöltek valamelyikét említő politikai oldalon megjelent cikkekhez közel 6.6 millió komment érkezett. A Facebook dominanciáját a hazai online kommunikációban jól mutatja, hogy a kommentek több mint 89 százaléka érkezett Facebookon megjelent cikkek és posztok alá. A kommentek időbeli felfutása nagyon hasonló ahhoz, amit a cikkek esetében láttunk, tehát a kampányhajrához közeledve nőtt a kommentek száma is. A kommenteket két irányból vizsgáltuk meg: a hozzászólásokban megjelent szentimentek/érzelmek oldaláról és a kommentek politikai „preferenciája” alapján. Fontos már ezen a ponton megjegyezni, hogy az elemzési egységeink nem a kommentelők, hanem az egyes kommentek, előbbire GDPR okok miatt nem tudunk pontos számot mondani.

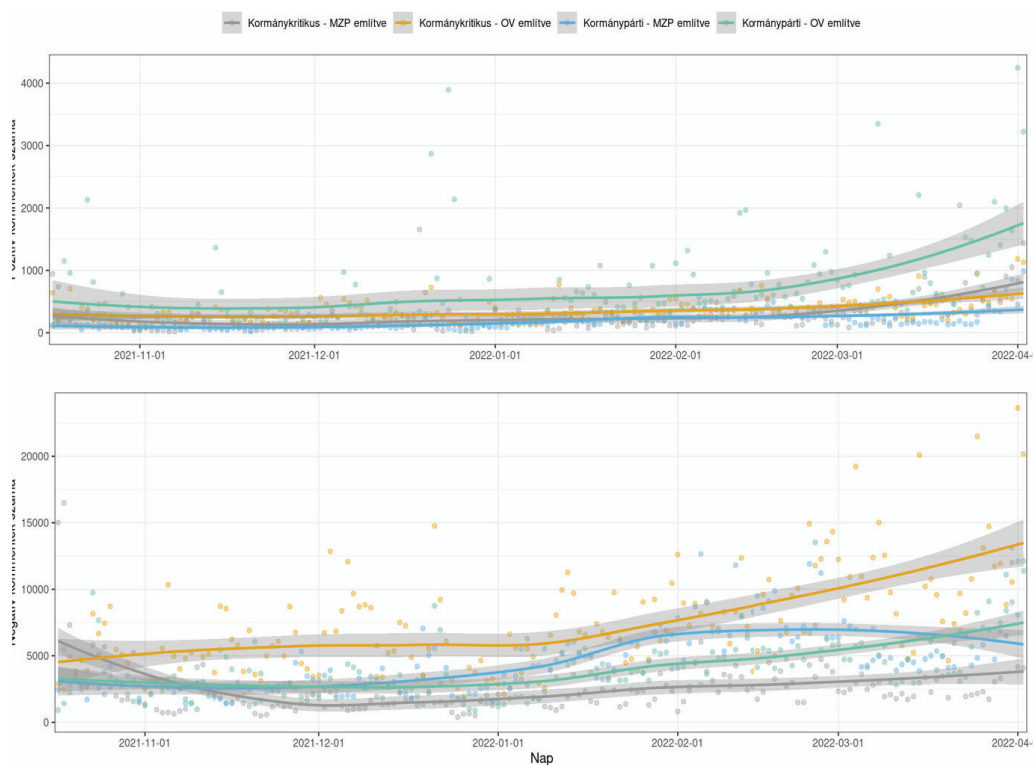
A kommentek természetesen nem mind tartalmaztak érzelmi reakciót, voltak érzelmi-  
leg nem bekegatorizálható kommentek is, például olyanok, amikben csak annyi szerepelt, hogy „oké”. Összesen a kommentek 62 százalékához rendeltünk hozzá valamilyen érzelmet. Egy komment akár több érzelmet is tartalmazhatott. A kommentek érzelmi töltete egyértelműen negatív volt. Az érzelmet tartalmazó kommentek 89 százaléka (!) tartalmazott negatív érzelmet, 10 százalékában azonosítottunk valamilyen semleges érzelmi kategóriát és mindössze hét százalékban jelent meg pozitív érzelmi töltet. Ha az eredményeket az összes kommentre vetítjük (tehát azokra is, melyekben nem találtunk érzelmet), akkor is egy nagyon magas 55 százalékos értéket kapunk a negatív kommentek arányára. A hazai komment kultúra negativitása nem új eredmény, korábbi kutatásink a trágárság nagyon magas (20-30 százalék közötti) elterjedtséget mutatták a hazai közéleti kommentekben (Szabó-Kmetty-Molnár 2021, Kmetty-Knap 2022). Az érzelmek közül egyértelműen a gúnyolódás dominált, az érzelmet tartalmazó kommentek 56 százalékában megjelent (lásd a mellékletben az M5 ábrát), de szintén gyakori volt az elégedetlenség és a düh megjelenése is. A sporadikusan megjelenő pozitív érzelmekből nem lehetett egy dominánsat sem megjelölni.

Az érzelmek időbeli felfutása érdekes képet mutat (3. ábra). A legtöbb pozitív komment a teljes időszakban a kormánypárti Orbán Viktorral foglalkozó cikkek alatt keletkezett. A kampányidőszak végére ez a különbség még látványosabb lett, nem függetlenül attól a tényről, hogy ebben az időszakban a kormánypárti média rengeteg cikkben foglalkozott Orbán Viktorral. De nem csak az abszolút számok nőttek meg, hanem cikkenként is volt egy csekély elmozdulás a pozitív kommentek növekedésének irányába. Szintén volt egy pozitív elmozdulás Márki-Zay Péter kapcsán született kormánykritikus kommentekben. Az abszolút növekedés úgy valósult meg, hogy közben a cikkek száma nem nőtt párhuzamosan.

De a kép nem lenne teljes, ha a negatív kommenteket nem állítanánk szembe a pozitív kommentekkel. Ebben szintén Orbán Viktor „nyert”. A kampány végén a kormánykritikus médiában is Orbán Viktor volt a fókuszban és ez magával hozta a napi negatív komment-mennyiség jelentős emelkedését is. Sőt, nagy mennyiségű negatív komment érkezett az Orbán Viktort említő kormánypárti cikkek alá is. Ez a nehezen értelmezhető tendencia, rámutat arra, hogy ezek az adatok magunkban nem adnak nekünk támpontot a kommentelők preferenciájának alakulásáról. A kommentszekcióban a hozzászólások sok esetben

## 152 Kampánykommunikáció a politika digitális tereiben

már nem is az eredeti cikkekre irányulnak, hanem vagy a másik jelölttel kezdenek el  
 3. ábra. Pozitív és negatív kommentek száma napi bontásban a jelölteket említő cikkeken,  
 a kormánypárti és a kormánykritikus médiában



foglalkozni, vagy egymást támadják a kommentelők (Szabó–Kmetty–Molnár 2021). A preferenciák megbecsléséhez a kommentek politikai színezete azonban talán közelebb tud minket vinni.

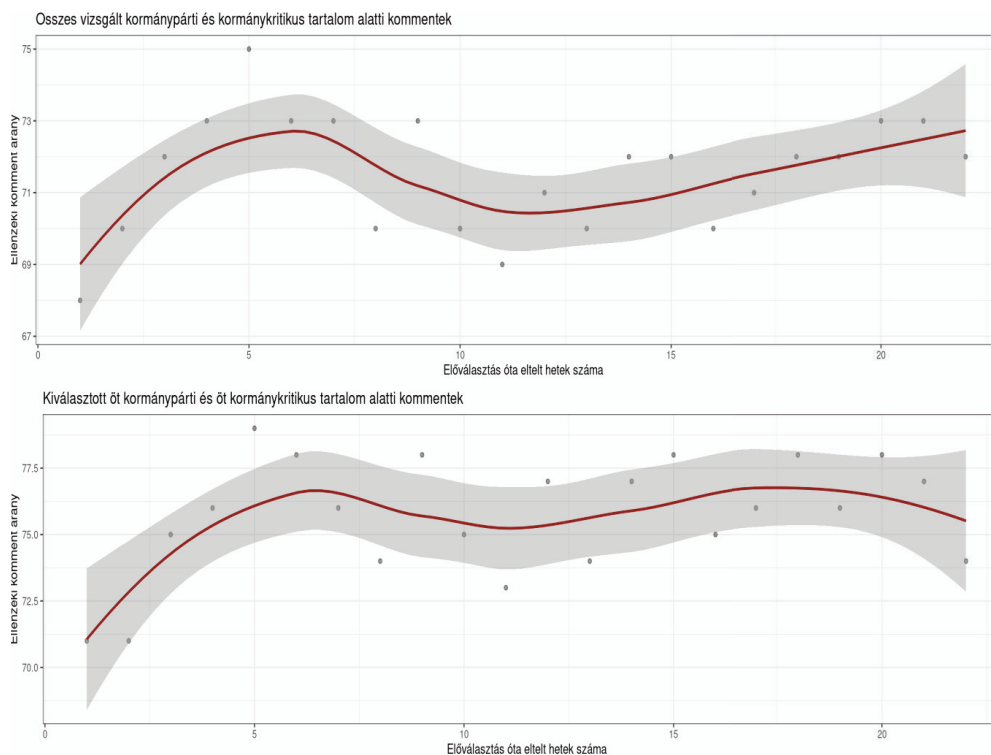
A kommentek 43 százalékáról volt megállapítható, hogy politikai jellegű, a politikai kommenteknek pedig 93 százaléka volt belpolitikai. A belpolitikai kommentek többsége (51 százalék) ellenzéki komment volt, a kormánypárti kommentek aránya pedig „mindössze” 21 százalék volt. A maradék 28 százaléknyi komment az egyéb kategóriába<sup>51</sup> került (utóbbi komment típust nem elemezzük most tovább). Ha csak a kommentek számát nézzük (M6 ábra a mellékletben), akkor a kampányfinishez közeledve, mind a kormánypárti mind az ellenzéki kommentek mennyisége nőtt. A növekedés azonban valamivel nagyobb volt az ellenzéki kommentek esetében. A kép azonban jóval tisztább, ha az ellenzéki/kormánypárti kommentek arányát vizsgáljuk. A nagyobb napi ingadozások kezelésére heti szintre aggregáltuk az adatokat. Az első hét az előválasztás lezárulta utáni

<sup>51</sup> Az „egyéb” kategóriába olyan kommentek kerültek, amik preferencia nélkül szólnak általános vagy szakpolitikai témákról.

hét, az utolsó (22.) hét pedig a választással ér véget. Az ellenzéki kommentek aránya szinte végig 71 százalék körül ingadozik, alig van olyan hét, amikor ehhez képest eltérne 2–3 százalékkal. A kampány vége felé a tendencia inkább felfelé mutat, azaz nemhogy csökkent volna az ellenzéki kommentek aránya, még nőtt is valamennyit. Egy regressziós modellben megvizsgáltuk, hogy volt-e összefüggés az ellenzéki kommentarány és a között, hogy a kommentek mekkora része íródott kormánypárti, és mekkora kormánykritikus média cikkei alatt, de a változók között szignifikáns összefüggést nem találtunk.

Az elemzésünket megismételtük a szűkített médiatérre is, amiben csak a kiválasztott 5-5 kormánypárti és kormánykritikus hírforrás szerepelt. Ezekben a médiatermékekben még magasabb volt az ellenzéki kommentek aránya. A tendencia másfelé mutatott, a kampányidőszak vége felé valamelyest csökkent az ellenzéki kommentek aránya – azonban ez a pár százalékos mozgás semmiképp sem tekinthető komoly átrendeződésnek.

4. ábra. Ellenzéki kommentek aránya heti bontásban



## Diszkusszió

Kutatásunk fő kérdése az volt, hogy lehetett-e látni az online térben azt, hogy a Fidesz ekkora aránnyal fogja megnyerni a 2022-es országgyűlési választásokat. Ennek a kérdésnek a megválaszolásához két szinten nyúltunk a rendelkezésre álló adatokhoz. Egyrésze

megvizsgáltuk a kampányidőszak média keretezését, másrészt a kommentek alapján kirajzolható politikai preferenciákat. Nem a teljes politikai spektrumot vizsgáltuk, hanem a hazai sajtóságokhoz igazodva a kampány középpontjában lévő két miniszterelnök-jelöltet, Orbán Viktort és Márki-Zay Pétert.

A kormánykritikus médiában Orbán Viktorhoz képest jóval kevesebbet szerepelt Márki-Zay Péter, ezzel szemben a kormánypárti médiában „kiegyensúlyozottabb” volt a jelöltek számszerű megjelenése, a kampány közepén Márki-Zay Péter jelent meg több cikkben, a kampány végén pedig Orbán Viktor. A kormánypárti és a kormánykritikus média nagyon eltérő narratívákban ábrázolta a jelölteket. Ez főleg a kormánypárti médiában volt szembevető, ahol a háború kapcsán Márki-Zay esetében hangsúlyosan jelent meg például a „fegyvert\_küld” szóösszetétel, ezzel szemben Orbán Viktor esetében a „béke” és a „kimaradás” szavak domináltak. Az elemzésünk arra is rámutatott, hogy a háború megváltoztatta a kampányt, vagy legalábbis a kampány narratív keretét. Az orosz támadás után a miniszterelnök-jelölteket említő cikkek felében megjelent a háborús tematika.

A keretek szerepét nem szabad sem alul-, sem felülértékelni. Bode (2016) kutatásai alapján a (közösségi) médiában megjelenő politikai információk csak részlegesen fedik le a politikai tájékozódás és meggyőzés csatornáit. A kommentek vizsgálata ráerősít erre az eredményre és azt sugallja, hogy az online térben a narratív keretek csak gyengén szabályozzák a választói reakciókat, utóbbiak egy ettől részben független logika mentén épülnek fel.

Ha csak az online kommentek alapján kellett volna megbecsülni a választás eredményét, akkor valószínűleg hasonlóan (vagy még inkább) elmértük volna az eredményt, mint a közvéleménykutatók. Bár a kampányfaisre egyre több pozitív komment jelent meg a kormánypárti média Orbán Viktort említő cikkei alatt, ugyanez elmondható a kormánykritikus média Márki-Zay Péterrel foglalkozó cikkei kapcsán is. Az még szembe tünőbb volt azonban, hogy a kormánykritikus médiában a kampány végén rengeteg negatív komment érkezett az Orbán Viktort említő cikkekre. A kommentek politikai színezete is hasonló képet mutatott.

Az ellenzéki kommentek a teljes kampányidőszak alatt jelenetős többletben voltak, és a kampány végén sem lehetett látni számottevő elmozdulást a kormánypárt irányába a kommentklímán keresztül. Az eredményeink egyértelműen arra utalnak, hogy az online térben megjelenő kommentek vizsgálata nem volt arra alkalmas, hogy a 2022-es választás kapcsán megmutassák a jelöltek támogatási mintáit. Ez egybevág Koltai és Stefkovics (2018) tanulmányának legfőbb üzenetével. A „miért nem működik” kérdésre legalább három különböző választ adhatunk. A három lehetséges magyarázat természetesen nem független egymástól, a mi esetünkben egyszerre akár több probléma is megjelenhetett, de ezt nem tudjuk ebben a tanulmányban ennél jobban szétválasztani.

1). A módszer lehetőségei szempontjából a legnegatívabb magyarázat az lehet, hogy egyszerűen ez a megoldás nem képes megmérni az emberek politikai preferenciáit. A kommentek esetében adatvédelmi okok miatt nem tudunk választói szintre aggregálni, ezért nem látjuk azt sem, hogy mekkora tömeg van a reakciók mögött, vannak-e olyanok, akik eltűnnek esetleg a jelöltek mögöl, vannak-e olyanok, akik megjelennek. A nemzetközi szakirodalomban hasonló célokra használt Twitter sokkal jobb ilyen szempontból, a felhasználók neveit látják a kutatók. Abban sem lehetünk biztosak, hogy egy negatív komment, vagy egy politikailag színezett komment jelenti-e egyáltalán egy jelölt támogatását vagy elutasítását.

2). A megközelítés kudarcának egy kevésbé negatív olvasata az lehet, hogy a módszer jó, csak a vizsgált adatbázissal van gond. Bizonyos szempontból a szöveges anyag túl tág volt, más szempontból pedig túl szűk. Szinte minden olyan nyilvános cikket és posztot bent hagytunk az elemzésben, amelyben valamely jelölt említésre került. Lehet érvelni amellett, hogy a cikkek egy részében csak mellékszálként említették a jelölteket és a cikk fókusza teljesen máshol volt, nem érintette a kampányt. Koltai és Stefkovics (2018) is felvetette a tanulmányukban, hogy érdemes lehet mondjuk topikmodellezéssel, vagy más eszközökkel szűkíteni a vizsgálati fókuszot. Ez az irány valóban megfontolandó lehet a későbbiekben. A túl tággal szemben a túl szűk keret is probléma bizonyos szempontból. A választott módszertan nyilvános oldalakon zajló kommunikációt tud vizsgálni. A nyilvános online térben főleg azok kommunikálnak, akik politikailag érdeklődőbbek és jellemzően erős pártszimpátiával rendelkeznek. A választáson azonban nem csak ők vesznek részt, hanem a csendes többség is. Ha képesek lennének beelátni a nem nyilvános tartalmakba is, akkor közelebb kerülhetnének ahhoz, hogy pontosabban feltárjuk az emberek viselkedés alapú szimpátiáját. Ebben segíthetne a különböző félig nyitott vagy teljesen zárt fórumok kommunikációjának a vizsgálata, vagy még inkább az emberek közötti közvetlen interakciók megfigyelése. Ezekben a tartalmakban azt is feltételezhetjük, hogy kevésbé dominálnak az aktuális média keretek, másképpen tematizálódik a közélet. Kutatói oldalról nézve, ilyen szempontból is előnyösebb a Twitter, ahol jellemzően látható a kommunikáció és nem csak a média által tematizált témákon belül folyik benne a diskurzus (Jungheer 2014). A Facebook alapú adatgyűjtés sem jelent feltétlen zsákutcát, de vélhetően érdemes más adatcsatornák felé fordulni. Ilyen szempontból érdekesek lehetnek azok az adatdonációs megközelítések, amikre már itthon is van példa (Breuer és munkatársai 2022; Kmetty 2021).

3). A harmadik lehetséges magyarázat abból indul ki, hogy a modell jól mér és az online teret használók véleményét pontosan képes megmutatni. A választásokon azonban olyanok is részt vesznek, akik vagy nincsenek is jelen az online térben, vagy jelen vannak, de csak passzív olvasóként és nem tartalom alakítóként. A Fidesz–KDNP kifejezetten jól szerepelt a 2022-es országgyűlési választásokon azokon a kisebb településeken, ahol a gazdasági helyzet nagyon rossz volt és az iskolázottság az átlag alatti. Korábbi kutatási eredmények és a 2022-es települési szavazati eredmények alapján feltételezhetjük azt, hogy a Fidesz jelentősen túl van reprezentálva az internetet nem, vagy csak korlátozottan használók között. Ha a kampány végén őket sikerült mozgósítani, azt nem mérhette be a megközelítésünk. Ha ez a feltételezés helyes, akkor akár azt is várhatjuk, hogy az internetpenetráció bővülésével és a lakosság digitális érettségének növekedésével egyre pontosabban lesz képes bemérni egy online megnyilatkozásokat vizsgáló modell a lakossági preferenciákat.

Az előrejelzési sikertelenség ellenére a tanulmányunk több új elemmel is tudta gazdagítani a tudásunkat. A választások kapcsán megjelenő média keretezés és az egyes médiumok pozíciója az online térben fontos elem a mindenkori politikai kommunikációs tudástárunkba. Ennél is fontosabbnak gondoljuk azonban azt, hogy rámutattunk arra, hogy az online klíma a 2022-es választási erőviszonyok kapcsán mennyire megtévesztő volt. Akik csak az online nyilvánosságból tájékozódtak, azt a következtetést vonhatták le, hogy egy nagyon szoros választási verseny lesz. Nem gondoljuk azt, hogy ne lenne fontos, hogy milyen vélemények és hangulati buborékok alakulnak ki a nyilvános online térben, de választást nem az online térben kell nyerni, hanem az urnáknál.

## Felhasznált irodalom

- Azunre, P., (2021). Transfer learning for natural language processing. Manning Publications Co, Shelter Island, NY.
- Bene, M. (2018). Post Shared, Vote Shared: Investigating the Link Between Facebook Performance and Electoral Success During the Hungarian General Election Campaign of 2014. *Journalism & Mass Communication Quarterly*, 95(2), 363–380. <https://doi.org/10.1177/1077699018763309>
- Bene M., – Szabó G. (2019). Bonded by Interactions: Polarising Factors and Integrative Capacities of the News Media in Hungary. *Javnost - The Public*, 26(3), 309–329. <https://doi.org/10.1080/13183222.2019.1639427>
- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. (2018) “quanteda: An R package for the quantitative analysis of textual data”. *Journal of Open Source Software*. 3(30), 774. <https://doi.org/10.21105/joss.00774>.
- Bode, L. (2016). Political news in the news feed: Learning politics from social media. *Mass communication and society*, 19(1), 24–48.
- Budiharto, W. – Meiliana, M. (2018). Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. *Journal of Big Data*, 5(1), 51. <https://doi.org/10.1186/s40537-018-0164-1>
- Burnap, P. – Gibson, R. – Sloan, L. – Southern, R. – Williams, M. (2015). 140 Characters to Victory?: Using Twitter to Predict the UK 2015 General Election. *Electoral Studies*. <https://doi.org/10.1016/j.electstud.2015.11.017>
- Breuer, J. – Kmetty, Z. – Haim, M. – Stier, S. (2022). User-centric approaches for collecting Facebook data in the ‘post-API age’: experiences from two studies and recommendations for future research. *Information, Communication & Society*, 1–20.
- Caetano, J. A. – Lima, H. S. – Santos, M. F. – Marques-Neto, H. T. (2018). Using sentiment analysis to define twitter political users’ classes and their homophily during the 2016 American presidential election. *Journal of Internet Services and Applications*, 9(1), 18. <https://doi.org/10.1186/s13174-018-0089-0>
- Chaudhry, H. N. – Javed, Y. – Kulsoom, F. – Mehmood, Z. – Khan, Z. I. – Shoaib, U. – Janjua, S. H. (2021). Sentiment Analysis of before and after Elections: Twitter Data of U.S. Election 2020. *Electronics*, 10(17), 2082. <https://doi.org/10.3390/electronics10172082>
- Fokasz Nikosz – Tóth Gergely – Micsinai István – Jelenfi Gábor – Előd Zoltán (2015). Kampány és valóságkonstrukció A 2010-es és 2014-es magyarországi választási kampányok összehasonlító elemzése a NOL és MNO oldalakon megjelent kampánytémák dinamikája alapján. *JEL-KÉP*, 3. <https://doi.org/10.20520/Jel-Kep.2015.3.25>
- Gayo-Avello, D. – Metaxas, P. – Mustafaraj, E. (2011). Limits of Electoral Predictions Using Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1), 490–493.
- Jahanbakhsh, K. – Moon, Y. (2014). *The Predictive Power of Social Media: On the Predictability of U.S. Presidential Elections using Twitter*. <https://doi.org/10.48550/ARXIV.1407.0622>



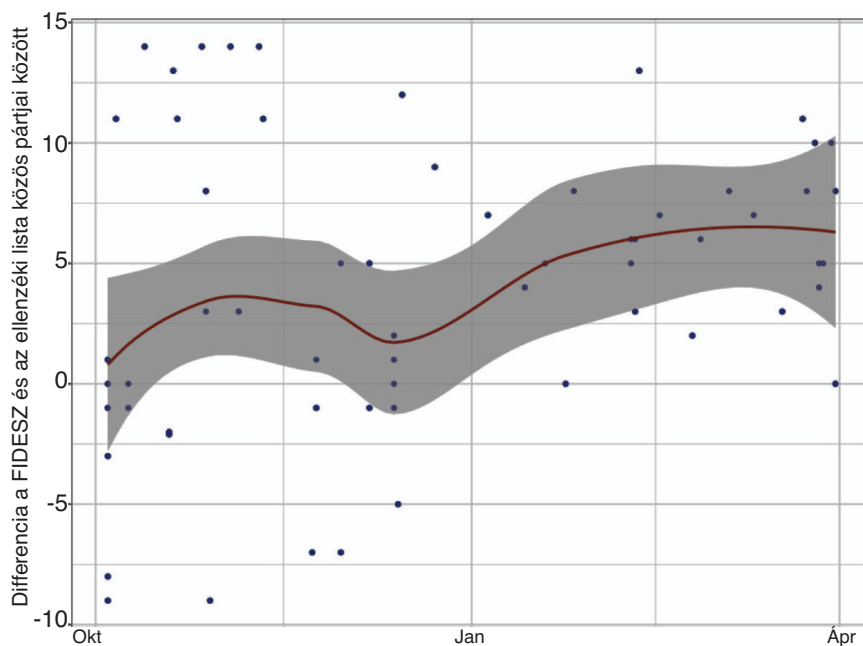
- Janky Béla – Kmetty Zoltán – Szabó Gabriella (2019). Mondd kire figyelsz, megmondom mit gondolsz! Politikai tájékozódás és véleményformálás a sokcsatornás kommunikáció korában. *Politikatudományi Szemle*, 28(2), 7–33. <https://doi.org/10.30718/POLTUD.HU.2019.2.7>
- Joyce, B. – Deng, J. (2017). Sentiment analysis of tweets for the 2016 US presidential election. *2017 IEEE MIT Undergraduate Research Technology Conference (URTC)*, 1–4. <https://doi.org/10.1109/URTC.2017.8284176>
- Jungherr, A. (2014). The logic of political coverage on Twitter: Temporal dynamics and content. *Journal of communication*, 64(2), 239–259.
- Kiss B. (2015). Prepossession and Indifference. Political leaders as integrative and disintegrative mechanisms in Hungary. *socio.hu*, 2015(4), 89–107. <https://doi.org/10.18030/socio.hu.2015.4.89>
- Kiss Balázs – Szabó Gabriella (2019). A médiafölényről. *Politikatudományi Szemle*, 28 (4), pp. 65–86
- Kmetty Zoltán (2018). A szociológia helye a Big Data paradigmában és a Big Data helye a szociológiában. *Magyar Tudomány*, 179(5), 683–692.
- Kmetty Zoltán (2021). Új utak a társadalom megismerésében. *Metszetek-Társadalomtudományi folyóirat*, 10(3), 6–26.
- Kmetty Zoltán (2022). Szóbeágyazási vektortérmodellek társadalomtudományi alkalmazása. *Statistikai Szemle*, 100(2), 105–136.
- Kmetty Zoltán – Knap Árpád. (2022): Trágárság mint érzelmi válasz a COVID-19-járvány idején. In Szabó Gabriella: Érzelmek és járványpolitizálás. Politikai érzelemmenedzserek és érzelmszabályozási ajánlataik Magyarországon a COVID-19 pandémia idején. Budapest, Magyarország : ELTE Eötvös Kiadó (2022) 199 p. p. 173
- Kmetty Z. – Tardos, R. (2022). Party Nexus Position Generator: New tool for the measurement of political homophily and political network diversity. *Social Networks*, 70, 112–125.
- Koltai Júlia – Stefkovics Ádém (2018). A big data lehetséges szerepe a pártpreferencia-becslésekben magyarországi pártok és politikusok Facebook-oldalainak adatai alapján. *Politikatudományi Szemle*, 27(2), 87–120. <https://doi.org/10.30718/POLTUD.HU.2018.2.87.120>.
- Körösényi András (2017). Weber és az Orbán-rezsim: Plebiszciter vezérdemokrácia Magyarországon. *Politikatudományi Szemle*, 26(4), 7–28.
- Lehmann, J. – Isele, R. – Jakob, M. – Jentsch, A. – Kontokostas, D. – Mendes, P. N –,Bizer, C. (2015). Dbpedia—a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic web*, 6(2), 167–195. <https://doi.org/10.3233/SW-140134>
- Liu, R., Yao, X. – Guo, C. – Wei, X. (2021). Can We Forecast Presidential Election Using Twitter Data? An Integrative Modelling Approach. *Annals of GIS*, 27(1), 43–56. <https://doi.org/10.1080/19475683.2020.1829704>
- Nemeskey D. M. (2021). Introducing huBERT. *XVII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2021)*, 3–14.
- Németh Renáta – Katona Eszter Rita – Kmetty Zoltán (2020). Az automatizált szövegelemzés perspektívája a társadalomtudományokban. *Szociológiai Szemle*, 30(1), 44–62. <https://doi.org/10.51624/SzocSzemle.2020.1.3>
- Ramteke, J. – Shah, S. – Godhia, D. – Shaikh, A. (2016). Election result prediction using Twitter sentiment analysis. *2016 International Conference on Inventive Computation Technologies (ICICT)*, 1–5. <https://doi.org/10.1109/INVENTIVE.2016.7823280>
- Singhal, K. – Agrawal, B. – Mittal, N. (2015). Modeling Indian General Elections: Sentiment Analysis of Political Twitter Data. *Information Systems Design and Intelligent Applications, Advances in Intelligent Systems and Computing*, 339, 469–477. [https://doi.org/10.1007/978-81-322-2250-7\\_46](https://doi.org/10.1007/978-81-322-2250-7_46)
- Soler, J. M. – Cuartero, F. – Roblizo, M. (2012). Twitter as a Tool for Predicting Elections Results. *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 1194–1200. <https://doi.org/10.1109/ASONAM.2012.206>

## 158 Kampánykommunikáció a politika digitális tereiben

- Szabó Gabriella – Bene Márton (2016). Széttöredezett vagy összekapcsolódó? *Politikatudományi Szemle*, 25(3), 33–58
- Szabó Gabriella (2018). Médiahálózatok 2018-ban. Az online portálok centruma és perifériái Magyarországon. In Böcskei Balázs – Szabó Andrea: *Várakozások és választások*. MTA TK Politikatudományi Intézet, Napvilág Kiadó 2018
- Szabo G. – Kmetty Z. – Molnar E. K. (2021). Politics and Incivility in the Online Comments: What is Beyond the Norm-Violation Approach?. *International Journal of Communication*, 15, 26.
- Tóka Gábor (2006). Vezérek passzív csodálói: A magyar választói magatartás nemzetközi összehasonlításban. In *Parlamenti választás 2006: Elemzések és adatok* (17–58).
- Thorson, K. – Wells, C. (2016). Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*, 26(3), 309-328.
- Wang, L. – Gan, J. Q. (2017). Prediction of the 2017 French election based on Twitter data analysis. *9th Computer Science and Electronic Engineering (CEECE)*, 2017, 89–93.
- Yaqub, U. – Chun, S. A. – Atluri, V. – Vaidya, J. (2017). Analysis of political discourse on twitter in the context of the 2016 US presidential elections. *Government Information Quarterly*, 34(4), 613–626. <https://doi.org/10.1016/j.giq.2017.11.001>
- Zsibrita J. – Vincze, V. – Farkas, R. (2013, September). magyarlanc: A tool for morphological and dependency parsing of hungarian. In *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*(pp. 763-771).

## Melléklet

*M1. A nyilvános közvéleménykutatásokban megjelent biztos szavazó pártválasztó csoportban a differencia a Fidesz–KDNP és az ellenzéki közös lista pártjai között.*



Forrás: Tóka Gábor, 2019-. "Választási közvélemény-kutatások Magyarországon. Excel adatbázis. URL: [kozvelemeny.wordpress.com](http://kozvelemeny.wordpress.com)

*M2. Szentimentek és érzelmek besorolási mutatói (HuBERT Model)*

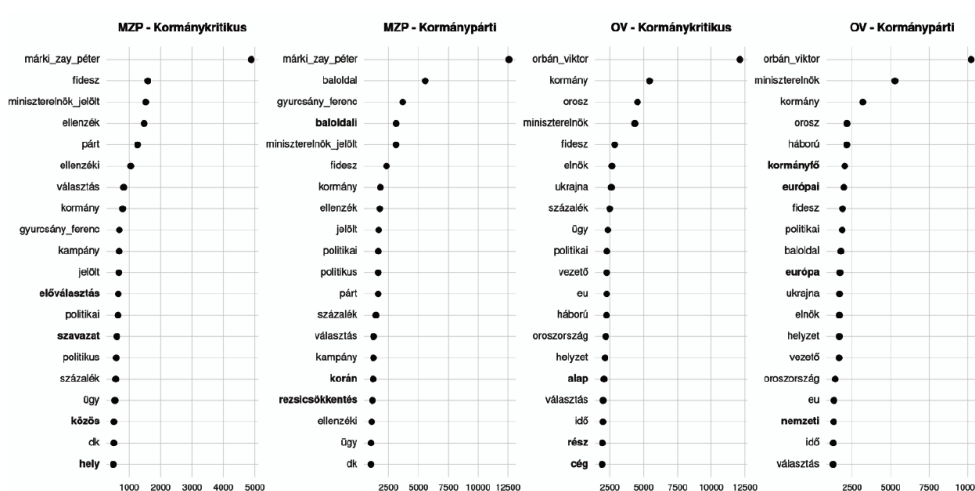
Érzelem	Accuracy	F1-score	Precision	Recall	MCC*
Pozitív	0.96	0.86	0.87	0.86	0.72
Negatív	0.81	0.81	0.81	0.81	0.62
Semleges	0.91	0.79	0.81	0.82	0.62
Nincs érzelem	0.82	0.79	0.78	0.79	0.56
Elégedettség	0.94	0.88	0.85	0.85	0.75
Reménykedés	0.98	0.85	0.82	0.89	0.71
Bánat	0.97	0.75	0.71	0.82	0.52
Düh	0.92	0.8	0.81	0.8	0.6
Elégedetlenség	0.84	0.76	0.74	0.77	0.51
Félelem	0.98	0.6	0.57	0.66	0.22
Gúnyolódás	0.83	0.81	0.8	0.81	0.61
Zavartság	0.95	0.75	0.73	0.78	0.5
Együttérzés	0.96	0.83	0.83	0.82	0.65
Érdeklődés	0.97	0.82	0.84	0.81	0.63
Nosztalgia	0.99	0.74	0.67	0.95	0.54
Meglepődés	0.99	0.73	0.7	0.76	0.46

\*Matthews Correlation Coefficient

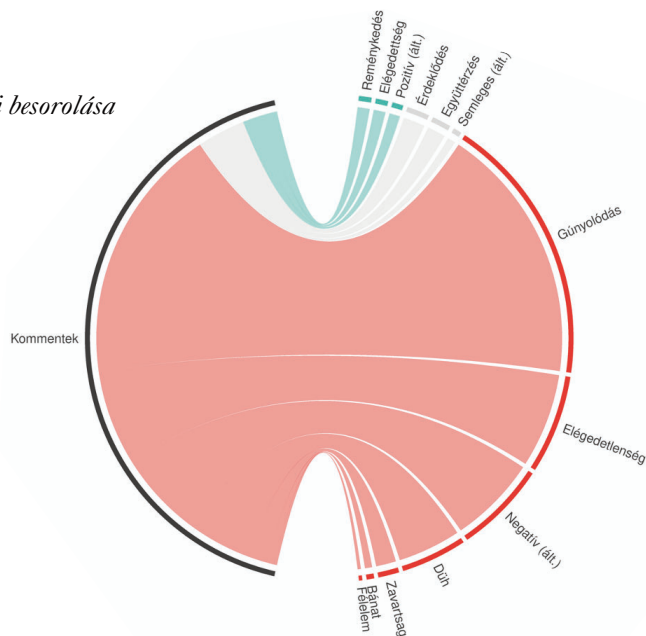
M3. Politikai töltet alapú besorolás mutatói (HuBERT Model)

Politikai	Accuracy	F1-score	Precision	Recall	MCC
Van – nincs	0.90	0.90	0.90	0.91	0.79
Hazai – nem hazai	0.97	0.64	0.70	0.61	0.63
Ellenzéki – kormánypárti – egyéb	0.76	0.61	0.60	0.60	-

M4. Leggyakoribb szavak a miniszterelnök-jelölteket említő cikkekben a kiválasztott öt-öt kormánypárti/kormánykritikus lapban



M. Kommentek érzelmi besorolása



*M5. Ellenzéki és kormánypárti kommentek napi száma*

