

Társas támogatás megjelenése egy depresszió és szorongás témájú online fórumon¹

Máté Fanni²

<https://doi.org/10.51624/SzocSzemle.2021.1.3>

Beérkezés: 2020. 11. 23.

Átdolgozott változat beérkezése: 2021. 04. 23.

Elfogadás: 2021. 04. 28.

Összefoglaló: Az online segítő fórumok, közösségek jellemző csatornáit a társas támogatásnak. Ezek a platformok a depresszióval, szorongással küzdők számára különösen jelentős segítséget nyújthatnak. Kutatásom célja kettős: egyfelől feltárni a társas támogatás mintázatait egy depresszió és szorongás témájú fórumon, másfelől pedig előzetes eredményekkel szolgálni a társas támogatás természetesnyelv-feldolgozással történő osztályozásához. Kutatásom egyedisége abban rejlik, hogy először a teljes használt adatbázis hagyományos kvalitatív elemzését végeztem el, s az így előálló adatbázist használtam az automatizált osztályozáshoz. A kvalitatív elemzés tapasztalatai a modellek definiálásához, azok működésének és teljesítményének megítéléséhez alapos ismeretekkel szolgálnak. Ezen ismeretek az automatikus szövegosztályozás szociológiai témában történő használhatóságának feltérképezésekor elemi információkat jelentenek. A vizsgált fórumon átlagosan ötből négy hozzászólás köthető a társas támogatáshoz. A segítő hozzászólásokban legjellemzőbb az információs (59,9 százalék) és az érzelmi támogatás (44,7 százalék). A társas támogatás tipizálását végző legjobb osztályozómodellek a hozzászólások közel 80 százalékát sorolták be a megfelelő kategóriába. Eredményeim alapján elmondható, hogy a vizsgált fórumon a társas támogatás típusának automatizált felismerésére alkalmazott modellek pontossága bizakodásra ad okot.

Kulcsszavak: depresszió, társas támogatás, online fórum, természetesnyelv-feldolgozás, NLP

Bevezetés

Legyen szó hírek olvasásáról, hivatalos tájékoztatásról vagy egyéni problémákhoz kapcsolódó segítségkérésről, az információszerzés elsődleges forrása napjainkban az internet. A személyes problémákkal kapcsolatos információgyűjtés és segítségkérés gyakran online közösségekben történik: az internetezők ezeken a helyeken keresnek választ a kérdéseikre, vagy gyűjtenek információt az adott témában jártasabb, tapasztaltabb, azonban sokszor szintén laikus hozzászólóktól (Hanley et al. 2019, Prescott et al. 2017, White–Dorman 2001). Rengeteg témában léteznek ilyen csoportok, a névnyugondozási tanácsokra alakult Facebook-csoporttól a gyereknevelési tapaszta-

1 A kutatás az Információs és Technológiai Minisztérium ÚNKP-19-3 kódszámú Új Nemzeti Kiválóság Programjának szakmai támogatásával készült. A szerző köszönetet mond Németh Renátának a segítségéért.

2 Eötvös Loránd Tudományegyetem, Társadalomtudományi Kar, email: matefanni429@gmail.com

tokat megosztókon át az egészséggel kapcsolatos anonim fórumokig szinte az élet összes területét lefedve. Így van ez a mentális egészség kérdésében is (Coulson 2005, White–Dorman 2001), amelynél az online fórumok anonimitása a stigmatizálástól való félelemérzetet, szégyenérzetet is csökkenti (Barak et al. 2008, Huang et al. 2014, Rogers–Pilgrim 2005). Az ilyen online közösségek térben és időben határtalanok, hiszen egy feltett kérdésre a világ bármely más pontjáról érkezhet válasz, viszont a válaszok gyakran késleltetve, nem rögtön érkeznek. Az online fórumok segítségével tehát személyes kapcsolathálónkon túl, fizikai helyzetünk és lakóhelyünk által nem korlátozva hozzánk hasonló helyzetben lévő embereket érhetünk el. Amellett, hogy a sorstársakkal való megismerkedés, velük való társalgás már önmagában támogatást jelenthet (Hanley et al. 2019), az online fórumok további előnye, hogy nagyobb energiabefektetés, tehát a lakásból való kimozdulás nélkül, viszonylag olcsón elérhetőek – hiszen pusztán internet-hozzáférésre van szükség a csatlakozáshoz.

Az online támogató fórumok használatának motivációja jellemzően a társas támogatás iránti vagy annak nyújtására irányuló igény (Hanley et al. 2019, Walther–Boyd 2002). A társas támogatás szerepe pedig – ahogyan más negatív élethelyzetben is – a depresszióval és szorongással élők számára különösen fontos (Gariépy et al. 2016). Mivel a jövőben a depresszió és szorongás még szélesebb elterjedése várható (World Health Organization 2017), számíthatunk arra is, hogy az érintettek közül egyre többen veszik majd igénybe az online fórumokat, közösségeket. Ráadásul egy jelenlegi világvárványhoz hasonló helyzetben szintén megnövekedhet az igény a mentális egészséghez kapcsolódó online segítség iránt, ugyanis az önkéntes vagy kötelező karantén a depressziót is előidézheti vagy fokozhatja (Lazáry 2020), az offline lehetőségek elérhetősége is korlátozott. Azonban már ma is elérhetőek a fiatalabb korosztályok szokásaihoz igazodó, telefonra letölthető, mentális egészséget támogató alkalmazások,³ amelyek lényegében online fórumként funkcionálnak, vagy a mesterséges intelligenciát hívják segítségül beszélgetőpartnernek. A technológiában rejlő lehetőségekre és azok széles körű használatára pedig a terület szakértőjének is reagálnia kell majd a kezelések során (Hanley et al. 2019, White–Dorman 2001). A fenti okok miatt az ezen fórumokból származó szövegek jelentős információval szolgálhatnak a depresszió és szorongás megélésének vizsgálatához, diagnosztizálásához (Karmen et al. 2015) vagy akár kezelésük támogatásához. Ezek feldolgozása azonban az adatok mennyisége miatt emberi erővel kivitelezhetetlen, így a gépi tanulást, azon belül is a természetesnyelv-feldolgozást (Natural Language Processing, NLP) kell szolgálatunkba állítanunk. A gépi tanulás ugyanakkor azzal is kecsegtet, hogy olyan mintázatokat, azonosságokat fedez fel az általa feldolgozott szövegretegben, amelyet a humán erő nem (lenne) képes, és a kódolást konzisztens módon végzi el a teljes adatbázisra (Molnár 2016).

3 Ilyen például a TalkLife vagy a Wysa.

Jelen kutatásom célja kettős: vizsgálom a társas támogatás megjelenését egy angol nyelvű online fórumon, másfelől pedig feltárom, mennyire különíthetők el természetes nyelv-feldolgozás segítségével a társas támogatás különböző kategóriái a vizsgált fórum hozzászólásaiban. A kutatás azzal az igénnyel született, hogy megalapozzon egy nagyobb volumenű, több fórumra kiterjedő kutatást, melynek célja a társas támogatási típusok azonosítása fórumbejegyzések szövegeiben annak érdekében, hogy mesterséges intelligenciaalapú fejlesztésekkel lehessen támogatni a fórumhasználók megküzdését. Annak ellenére, hogy jelen tanulmány angol nyelvű szövegeket vizsgál, tapasztalatai a hazai kutatók számára is hasznosak lehetnek. Az elemzés módszertani eredményei hozzájárulnak a hazai szociológiában még nem rutinszerűen használt (Németh et al. 2020) természetesnyelv-feldolgozás módszerének megismertetéséhez és hasznosíthatóságának megítéléséhez, továbbá ahhoz, hogy láthatóvá váljanak a módszerben rejlő lehetőségek, vagy éppen korlátok. A bemutatott módszerek pedig magyar nyelvű kutatásokban is alkalmazhatók, azzal a kitételrel, hogy a nyelvek különbözősége miatt részben eltérő eszközök igénybevételére van szükség.

A tanulmány első részében a kutatáshoz kapcsolódó főbb fogalmakat tisztázom. Ez után a kutatási kérdések megválaszolására alkalmazott módszereket, majd eredményeimet mutatom be. Az írás végén összefoglalom a kutatásom főbb eredményeit és tanulságait, illetve felvázolom a jövőbeli kutatási lehetőségeket.

Elméleti háttér

Társas támogatás

Társas támogatás alatt azokat a mások (nem szakértők) által nyújtott tevékenységeket értjük, amelyek segítenek egy negatív élethelyzetben lévő személynek az azzal való megküzdésben (Cutrona–Suhr 1994). A társas támogatás tipizálása nem teljesen egységes: egyes szakirodalmakban négy (House et al. 1988) másokban pedig öt kategóriát használnak (Cutrona–Suhr 1994, Vayreda–Antaki 2009) amely a korábbi négy kategóriát a kapcsolati hálóra épülő támogatással egészíti ki. A kutatás során én az újabb, öt kategóriát tartalmazó besorolást használtam, mivel a tudományos eredmények alapján a kapcsolatok megléte vagy hiánya depresszió esetén fontos tényező lehet (Grav et al. 2012, Kaplan et al. 1977). Az öt főkategória a következő: információátadás (*informational*), kézzelfogható segítség (*tangible*),⁴ érzelmi támogatás (*emotional*), megbecsülés (*esteem*) és kapcsolatokra, kapcsolati hálóra épülő támogatás (*network*) (Cutrona–Suhr 1994, Vayreda–Antaki 2009).

Az egyes kategóriákat részletesen Cutrona–Suhr (1992) tanulmánya alapján mutatom be, mely több más kutatás alapját is képezi (Braithwaite et al. 1999, Coulson 2005,

4 Ez többít instrumentális segítségként is említik a szakirodalomban (Gariépy et al. 2016), míg mások (Semmer et al. 2008) az instrumentális kategóriát az érzelmi kategória ellenpontjaként fogalmazzák meg. Utóbbihoz tartozik a megbecsülés vagy törődés, míg előbbihez az információ nyújtása vagy a kézzelfogható segítség adása (Semmer et al. 2008).

Huang et al. 2014). Ez a kategorizálás eredetileg természetesen offline társas támogatásra fogalmazódott meg, de online támogatást vizsgáló kutatásokban is bevett felosztás (Coulson 2005, Huang et al. 2014). Így rámutathatunk arra is, hogy mely területek lehetnek azok, amelyek az online kommunikációban kevésbé jelen(het)nek meg a személyes támogatáshoz képest.

A kutatás során használt osztályozás alapjaként a Cutrona és Suhr (1992) által alkotott *Social Support Behavior Code* (1. táblázat) szolgált. Az első kategóriába, az *információ átadásába* tartozik, ha információkkal, tanácsokkal, ötletekkel segítjük az érintettet, de az is, ha segítünk átértékelni a helyzetét, vagy szakértőt ajánlunk számára. A *kézzelfogható segítségbe* tartoznak azok a formák, amelyek „konkrét” segítségek, például kölcsön nyújtása vagy feladatok átvállalása, de az is idetartozik, ha előre felajánljuk a segítségünket, amennyiben arra szüksége lesz az érintettnek. Az *érzelmi támogatás* a legbővebb kategória, amelynek része a fizikai kontaktus, például ölelés (amely online helyzetben legfeljebb a szándék szintjén valósulhat meg), a negatív élethelyzetben lévő személlyel való kapcsolat vagy az iránta érzett szeretet kifejezése, ahogyan a sajnálat vagy együttérzés kifejezése is. Szintén idetartozik az érintett figyelmes meghallgatása, bátorítása vagy az adott személyért történő imádkozás is. A *megbecsülés* típusú társas támogatáshoz a bókolást, az érintettel való egyetértést és a kialakult helyzetben játszott felelősségének enyhítését sorolhatjuk. Végül a *kapcsolatok, kapcsolati hálóra épülő* kategóriához a támogató jelenlét, illetve az érintett új emberekkel történő megismertetése köthető, továbbá az is, ha emlékeztetik az érintettet arra, hogy nincs egyedül, és mások is az övéhez hasonló problémával küzdenek.

1. táblázat: A társas támogatás 5 fő- és 23 altípusa

Információ	
Tanács	Ötleteket, lehetséges tevékenységeket javasol.
Ajánlat	Egyéb, külső segítséget ajánl.
A helyzet átértékelése	A segítő újradefiniálja, átértékeli a helyzetet.
Tanítás	Olyan tudást, híreket, információkat oszt meg, és olyan képességekről beszél, amik segítik kezelni a helyzetet
Kézelfogható segítség	
Kölcsön	Felajánlja, hogy kölcsön ad valamit.
Konkrét feladat	A negatív élethelyezethez közvetlenül kapcsolódó feladat elvégzését ajánlja fel.
Közvetett feladat	Felajánlja, hogy átvállalja az érintett egyéb, a negatív élethelyezethez nem kapcsolódó feladatainak egy részét.
Aktív jelenlét	Csatlakozik az érintetthez a stressz csökkentését szolgáló tevékenységben.
Hajlandóság	Potenciális segítséget ajánl fel, felajánlja, hogy segít, ha kell.
Érzelmi	
Kapcsolat	Hangsúlyozza a kapcsolatuk fontosságát és a szeretetet.
Fizikai támogatás	Fizikai kontaktust (ölelést, puszit, kézfogást stb.) nyújt.
Titoktartás	Kifejezi, hogy bizalmasan kezeli az érintett helyzetét.
Szimpatia	Sajnálatot, együttérzést fejez ki.
Figyelem	Figyelmes megjegyzéseket tesz, amikor a másik beszél.
Megértés/empátia	Kifejezi a megértését, vagy olyan saját történetet mond el, ami azt jelenti, hogy átérzi az érintett helyzetét.
Bátorítás	Reményt és bizalmat nyújt az érintettnek.
Ima	Imádkozik az érintettért.
Mebecsülés	
Bók	Kiemeli az érintett előnyeit vagy hangsúlyozza a képességeit.
Érvényesítés	Kifejezi, hogy egyetért az érintett nézőpontjával.
Felelősség enyhítése	Csökkenti az érintett büntudatát.
Kapcsolatok, kapcsolati háló	
Hozzáférés	Felajánlja, hogy új emberekkel ismerteti meg az érintettet.
Jelenlét	Felajánlja a társaságát az érintettnek.
Társaság	Emlékezteti az érintettet támogató közösségekre, olyanokra, akik az érintetthez hasonló helyzetben vannak.

Depresszió, szorongás és társas támogatás

Kutatásomban a társas támogatás megjelenését egy depresszió és szorongás témájú fórumon vizsgálom. A depresszió az Amerikai Pszichiátriai Társaság (APA) 2013-ban kiadott kézikönyve (*Diagnostic and Statistical Manual of Mental Disorders: DSM-5TM, 5th ed.*) szerint akkor áll fenn, ha az alábbi tünetek közül legalább öt legalább két héten keresztül egyszerre megfigyelhető:

- lehangoltság
- érdeklődés, örömeire való képesség elvesztése
- étvágyban vagy testsúlyban bekövetkező nagymértékű növekedés, vagy csökkenés
- megnövekedett vagy csökkent alvási igény, álmatlanság
- pszichomotoros agitáció (nyugtalanosság) vagy retardáció (lassulás)
- fáradtság, energiahány
- csökkent önértékelés, indokolatlan büntudat érzése
- gondolkodásra való képtelenség, döntésképtelenség, koncentrációhiány
- halálra vonatkozó gondolatok gyakorisága, öngyilkossági készletés.

Az első két tünet kitüntetett, közülük legalább az egyiknek mindenképpen teljesülnie kell a diagnózishoz (APA 2013). A szorongás egyfelől túlzott félelmet, feszültséget jelent, másfelől az olyan szorongásos zavarokat jelöl, mint például a generalizált szorongás, a pánikzavar, a különböző fóbiák vagy a poszttraumás stressz szindróma.

A szorongás és a depresszió sok esetben együtt jár (Hirschfeld 2001), és több kutatásban is együtt kezelik őket. A fentiek miatt saját kutatásom során én is együtt kezelem a két fogalmat, a vizsgált fórum mindkét témát felöleli.

A társas támogatás és depresszió együttes vizsgálata már Durkheimnél is megjelent,⁵ aki úgy találta, hogy a társas kapcsolatok, az erős kötések hiánya kapcsolatban áll az öngyilkossággal (1951), mely – főként az egoista típusú öngyilkosság esetén – közvetetten a társas támogatás meglétéhez vagy hiányához köthető. Brown és munkatársai azt találták (2002), hogy bizalmi kapcsolattal és munkahellyel nem rendelkező, gyermekeikkel otthon lévő nők nagyobb eséllyel lettek depressziósak stresszes élethelyzetben, mint más nők. Comer és társai (2013) a depresszió szociokulturális megközelítése kapcsán arról írnak, hogy a depresszió okai a személyközi kapcsolatokban keresendők, például a gyenge szociális képességek miatti csökkenő interakciók, alacsony szintű támogatást nyújtó házasságok vagy társadalmi izoláció. A depresszió szociológiai megközelítésében tehát az izoláció és a társadalmi kapcsolatok hiánya potenciális okként jelentkezik.⁶

A depresszió és a társas támogatás kapcsolatát vizsgáló kutatásokból készített metaelemzést Gariépy munkatársaival (2016) a társas támogatás három típusára (érzelmi, információs, instrumentális/kézzelfogható) fókuszálva, amelyben ezek lehetséges hatásait vizsgálták. Összesen száz kutatás eredményei szerint a 18 és 27 éves kor közöttiek esetén a legtöbb esetben szignifikáns, a depresszió csökkenésének irányába ható kapcsolatot elsősorban az érzelmi támogatás, másodsorban az instrumentális, harmadsorban pedig az információt nyújtó segítség mutatott a depresszióval. A szerzők ezek alapján arra a megállapításra jutnak, hogy a társas támogatás megléte segíthet a depresszió csökkenésében.

Társas támogatás az online közösségekben

A virtuális tér, az internet a társas támogatásban is új csatornákat nyitott, térben és időben korlátlanul elérhetővé téve azokat. Az online tér anonimitása is az offline segítségkérés helyett az online csoportok felé terelheti az egyént, különösen ha olyan esetben kell segítséget kérnie, amely stigmatizálással járna, vagy kellemetlenséget okozna számára (Hanley et al. 2019, Prescott et al. 2017, White–Dorman 2001).

Az elmúlt időszakban egyre több tanulmány foglalkozik az online közösségekben, online fórumokon elérhető társas támogatással. Neil Coulson (2005) irritábilis bél szindrómával élők online fórumát vizsgálta az általam is alkalmazott öt kategória

5 Durkheim elmélete annak ellenére is megkerülhetetlen bázisát képezi az illetén vizsgálatoknak, hogy módszereit gyakran kritikával illetik (Selvin 1958, van Poppel–Day 1996).

6 Ez természetesen csak egy aspektusa a depresszió szociológiai megközelítésének. További irodalmak a témában például: Busfield 2000, Píkó 2002, Rogers–Pilgrim 2005.

alapján. Kutatásában, amelyben 572 üzenetet vizsgált deduktív tematikus elemzéssel, arra jutott, hogy az általa vizsgált fórumon az öt közül a leggyakoribb kategória az információcélú társas támogatás. Ennél tágabb nézőpontból vizsgálta a témát Deetjen és Powell (2016), akik többféle egészségi témával foglalkozó fórumot – ideértve a mentális egészséggel foglalkozó fórumokat is – vizsgálták az alapján, hogy az érzelmi vagy az információs támogatás gyakoribb-e az ott folyó diskurzusokban. Automatikus szövegelemzést, azon belül bayesi klasszifikációt használva arra az eredményre jutottak, hogy a mentális egészséggel foglalkozó fórumokon az érzelmi támogatást nyújtó hozzászólások dominálnak, míg hosszú ideig fennálló, elsősorban fizikai jellegű problémák esetében az információs típusú támogatás a jellemzőbb. Kutatásomhoz a leginkább hasonló Prescott és munkatársainak (2017) Nagy-Britanniában végzett kutatása, amelyben fiatalok (11-25 évesek) online kommunikációját vizsgálták a társas támogatás megjelenése szempontjából mentális egészséggel foglalkozó fórumokon. Eredményeik szerint a fiatalok mind információs, mind emocionális támogatásra számíthatnak a vizsgált fórumokon. E kutatás újdonsága az volt, hogy az elemzés során megkülönböztették a célzott (*directed*) és nem célzott (*non-directed*) segítséget: míg utóbbi típusnál személyes tapasztalatait, történeteit osztja meg valaki, a célzott segítségnél tanácsot (is) ad (Prescott et al. 2017).

A korábbi kutatások alapján tehát elmondhatjuk, hogy az online fórumok a vizsgált öt kategória közül jellemzően információt és érzelmi támogatást nyújtanak. Kérdés azonban, hogy ez mennyire felel meg a fórumokat felkeresők igényeinek. Hiszen ha a fórumokat kifejezetten támogatás kérése céljából használják (Hanley et al. 2019, Walther–Boyd 2002), akkor a fórumok sikerességének fokmérője is lehet az, hogy mennyire képesek a felhasználói igényeket kielégíteni. Ezt a gondolatmenetet erősíti Cutrona és Suhr (1994) azon megállapítása, hogy a társas támogatás akkor éri el igazán stresszcsökkentő hatását, ha olyan jellegű segítség érkezik, amelyre a negatív helyzetben lévőknek szüksége van. Például egy olyan esetben, amikor az érintett kontrollálni tudja a helyzetet, a megbecsüléshez kötődő segítség; míg egy kontrollálhatatlan helyzetben a kézzelfogható segítség csökkentheti igazán hatékonyan a stresszt. Egy későbbi munkájában Cutrona munkatársaival (2007) a korábitól részben eltérő szempontok mentén fogalmazza meg a „kereslet” és a „kínálat” találkozását: eszerint akkor beszélhetünk a vágyott és a kapott támogatás egyezőségéről (*matching*), ha olyan típusú támogatás érkezik, amelyet a negatív élethelyzetben lévő igényel. Szintén ezt a definíciót használja Vayreda és Antaki (2009), akik új fórumhasználók kezdőposztjaiban megjelenő igényeket és az azokra érkező válaszokat elemezték. Az általuk vizsgált fórumnak helyet adó honlapra deklaráltan a medikális megközelítés jellemző, vagyis a fórumon zajló beszélgetés, az abban részt vevők, sőt, a moderátor is betegségeként, medikális úton kezelhető problémaként tekint a bipoláris zavarra. Eredményeik szerint, ha a fórum új használója ezt az értelmezést nem utasítja el, a fórumszokások szerint számíthat válaszra. A válaszok azonban sok esetben nem

találkoznak a posztot író igényével, ugyanis elsősorban közvetlen, kéretlen tanácsot kap, akkor is, ha annál általánosabb támogatást vagy információt kért volna.

A kutatás módszertana

Kutatásomban tehát egy depresszió témájú online fórumon vizsgálom, hogy milyen jellegű társas támogatás jelenik meg, és hogy osztályozhatók-e automatikusan az egyes hozzászólások a társas támogatás előre meghatározott típusaiba. E két cél az alábbi kutatási kérdésekkel ragadható meg:

- 1) Mennyire gyakori az adott fórum hozzászólásaiban a társas támogatás megjelenése, és azon belül melyik fő- és alkategóriák fordulnak elő leggyakrabban?
- 2) Milyen mértékben tárható fel automatikus szövegosztályozási módszerekkel az, hogy egy-egy hozzászólás melyik társas támogatási kategóriába tartozik?

Az első kérdés megválaszolásához a fórum hozzászólásainak kvalitatív tartalom-elemzését végzem el. Második kérdésemhez pedig automatizált kvantitatív szöveg-elemzéssel mintázatokot keresek a vizsgált fórum hozzászólásaiban. Ez utóbbi célja, hogy felderítő kutatásként szolgáljon a társas támogatást tartalmazó hozzászólások automatikus szövegelemzéssel történő osztályozását végző további kutatásokhoz, programokhoz. A teljes kutatás pedig esettanulmányként szolgál fórumhasználók célzott segítségét támogató programok későbbi kidolgozásához. A konkrét eredményeim általánosíthatósága valamelyest korlátozott, hiszen az öt típus konkrét arányai a különböző fókuszú fórumok esetében eltérhetnek. Azon eredményeim azonban, amelyek az általam alkalmazott automatizált kvantitatív klasszifikációs modellek hatékonyságát mutatják az ilyen témájú és nyelvezetű hozzászólások esetén, fontos és általánosan alkalmazható viszonyítási pontot jelenthetnek a későbbi kutatásokban.

Használt adatbázis

Kutatásom adatbázisát egy, az RC²S² kutatócsoport⁷ keretében végzett, depresszió témájú angol nyelvű fórumokhoz kapcsolódó kutatás adatbázisának egy részhalmaza adja. Ezen adatokat a SentiOne bocsátotta rendelkezésünkre, a GDPR szabályainak megfelelően, regisztráció nélkül olvasható nyilvános fórumokról. A kutatásomban nem vizsgálom a hozzászólások szerzőit, és a szövegeket is csak magyarra fordítva idézem, így is biztosítva az anonimitást.⁸

Az eredeti adatbázist adó fórumokat a „*depression forum*” és „*depression online*” kifejezésekre történő kereséssel állítottuk össze, azt modellezve, ahogy valaki ilyen témájú fórumokat kereshet. Jelen kutatásomban az így létrejött adatbázisból a legtöbb

7 rc2s2.eu A depresszióval foglalkozó kutatást végezték: Németh Renáta, Sik Domonkos, Katona Eszter és a szerző.

8 Mivel a fórumon jellemzően nicknévvel szerepelnek a felhasználók, anonimitás alatt elsősorban azt értem, hogy ne lehessen összekapcsolni egy felhasználó összes hozzászólását, vagyis ne lehessen egyfajta profilt alkotni a szerzőkről. Különösen akkor indokolt ez a védelem, ha a felhasználó személyes információkat is megosztott magáról a nicknévben, adatlapján vagy a hozzászólásaiban.

hozzászólást tartalmazó *healthunlocked.com* oldal egyik, depresszióhoz és szorongáshoz kapcsolódó fórumának (*Anxiety and Depression Support*) beszélgetésfolyamait (vagyis *threadjeit*) vizsgálom. Beszélgetésfolyam alatt azokat a tematikus al fórumokat értem, amelyeket külön címmel indít el egy-egy fórumhasználó. A fórumot a leírása szerint az *Anxiety and Depression Association of America* indította el „szorongással és depresszióval érintettek számára, hogy olyanokkal beszélhessenek, akik igazán megértik őket”. Kutatásomban a hozzászólásokban megjelenő társas támogatás típusainak kvalitatív elemzése (mely a kvantitatív szakasz alapját is képezi) miatt szükséges területi korlátok okán az al fórumnak csak a 2018-as évben született hozzászólásait elemeztem. Ezáltal az elemzett adatok mennyisége feldolgozható mértékű volt a kutatás keretein belül, és az esetleges szezonális hatások (McCarthy et al. 2002) sem torzítják az adatokat. A kiinduló adatbázis 2358 hozzászólást tartalmazott, amelyek tisztítása⁹ után összesen 2336 hozzászólást osztályoztam. A hozzászólásokat az al fórumokon belül időbeli sorrendbe állítva elemeztem, így egy-egy beszélgetésen belül a hozzászólásokat az eredeti kontextusukban értelmezve követtem végig.

Alkalmazott módszerek

A kvalitatív elemzés során először öt kategóriába soroltam a hozzászólásokat az alapján, hogy köthető-e a hozzászólás társas támogatáshoz. Az öt kategória a következő volt: (1) *nem jelenik meg társas támogatás a hozzászólásban*; a hozzászólás írója valamilyen típusú (2) *társas támogatást kér*; a hozzászólás írója valamilyen típusú (3) *társas támogatást nyújt*; a (4) *támogatást megköszönő hozzászólás*; vagy pedig (5) *irreleváns hozzászólás*. Ez utóbbi kategóriába tartoztak a csak linkeket, hirdetéseket, vagy csak köszönést tartalmazó hozzászólások és egy, a fórum használatára vonatkozó (kezdő) hozzászólás. Ezt követően a társas támogatást tartalmazó (azt kérő vagy azt nyújtó) hozzászólásokat besoroltam az öt fő kategória 23, fent ismertetett alkategóriáiba. Gyakran egy hozzászóláson belül többféle társas támogatás fordult elő, ilyen esetben mindegyik kategóriát feltüntettem az adott hozzászólásnál, nem téve különbséget a kategóriák között (például azok súlyában vagy az előfordulásuk sorrendjében).

Annak feltárására pedig, hogy automatikus osztályozással besorolhatók-e a vizsgált hozzászólások a társas támogatás valamelyik kategóriájába, természetesnyelv-feldolgozást alkalmaztam, amely a természetesen előforduló (beszélt és írott) nyelvek számítástudományi módszerekkel történő feldolgozását végzi többféle alkalmazási terület számára oly módon, hogy az az emberi feldolgozáshoz hasonló minőségű legyen (Liddy 2019). Az osztályozást többféle felügyelt algoritmussal végeztem. A felügyelt algoritmusok működésének lényege, hogy az osztályba sorolást a gép egy előzetesen humán kódolók által osztályozott adatbázison „tanulja meg”, majd az újabb adatokon ez alapján végzi el az automatizált besorolást. Ezen módszereknél tehát a gépi besorolás nem pusztán az adatok hasonlóságából-különbségeiből indul ki, ha-

9 A tisztításra az automatikus adatgyűjtésből származó hibák miatt volt szükség, így eltávolítottam az ismétlődő hozzászólásokat, és összekapcsoltam a több egységre töredezett hozzászólásokat.

nem ismeri azok „valós” kategóriáját is (Németh–Koltai 2020, Sebők et al. 2016, Tikk 2007). Az előzetes besorolást tartalmazó adatbázist jelen esetben az első kutatási kérdés megválaszolásához alkalmazott besorolás képezi.

Alkalmazott modellek

Háromféle modellt alkalmaztam az automatikus osztályozáshoz: naiv Bayes¹⁰-osztályozót, SVM-et (*Support Vector Machine*)¹¹ és logisztikus regressziót. A naiv Bayes-algoritmus a Bayes-tételre alapuló osztályozó, amely a tanuló adatbázis alapján kiszámítja egy hozzászólás szavaira az egyes osztályokba tartozás valószínűségét, majd a hozzászólást abba az osztályba sorolja, amelynél ezen valószínűségek szorzata maximális. A modell nevének naiv előtagja arra utal, hogy a valószínűségek szorzata a szavak függetlenségét feltételezi, amely a legtöbb esetben sérül (Tan et al. 2012, Tikk 2007) (hiszen egy szöveget alkotó szavak jellemzően nem egymástól függetlenül tűnnek fel a szövegben).

Szövegek osztályozása során a logisztikus regresszió „független változói” a szöveg egyes szavai, esetleg egyéb, általunk megadott jellemzői. A modell a tanuló adatbázisban minden kategória esetében együtthatókat rendel a szavakhoz (egyéb tényezőkhöz), melyeket a teljes adatbázis szavaira (egyéb tényezőire) alkalmazva végzi el az ismeretlen szövegek osztályozását (Hosmer Jr. et al. 2013). Ennek előnye, hogy a modell működésébe is bepillantást enged, láthatjuk, melyek voltak a legfontosabb szavak az egyes kategóriák melletti döntés meghozásakor.

Az SVM algoritmus alapvetően lineáris döntési határt alkalmaz, azonban nem egy egyenes (több dimenzióban [hiper]sík) jelöli ki a döntési határt az egyes kategóriák elkülönítésére, hanem kettő, amelyek a lehető legszélesebb tartományt foglalják magukban. A sík a megkülönböztetendő csoportok két adatpontja¹² között található a legoptimálisabb helyzetben. Ahhoz, hogy egyszerre több csoportot is osztályozhassunk általa, kétféle megközelítést alkalmazhatunk. Az egyik megközelítés esetén a mintát egy-egy kitüntetett csoportra és annak komplementerére bontjuk. Ebben az esetben annyi bináris osztályozást kell elvégeznünk, ahányféle csoportunk van. A másik megoldás esetén páronként vesszük sorra a lehetőségeket, és mindig csak e két csoport közötti besorolást végezzük el. Ezt követően összegezzük az egyes osztályozás eredményeit, és a többségi elv alapján döntünk. Az általam használt algoritmus ez utóbbit követte (Aggarwal 2018).

10 Egyéb írásmódja külföldi és magyar szakirodalmakban: Naive Bayes, Naïve Bayes.

11 Tikk Domonkos (2007) könyvében részben magyarul, szupportvektorgépként hivatkozik rá, azonban azóta a magyar szaknyelvben is elterjedt a support vector machine kifejezés használata, így én is ezt a formát használom.

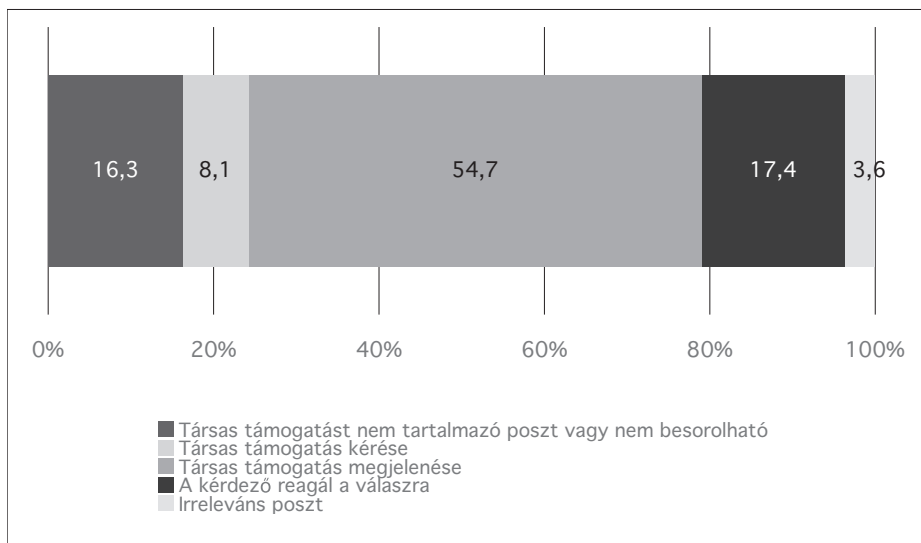
12 Ezek az adatpontok maguk a „tartóvektorok” (support vectors).

Eredmények

A társas támogatás megjelenése

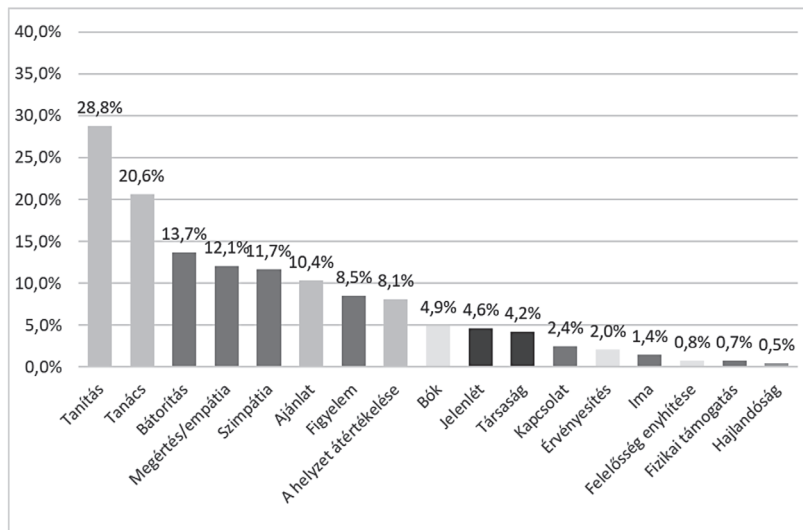
Ahhoz tehát, hogy feltárjam, milyen típusú társas támogatás jelenik meg az adott fórum hozzászólásaiban, kvalitatív tartalomelemzést végeztem. Az alábbi ábrán látható, hogy a fórumon előforduló hozzászólások jelentős hányada valamilyen formában a társas támogatáshoz kapcsolódik. A hozzászólások 80,1 százaléka, vagyis 5-ből 4 hozzászólás vagy társas támogatás iránti igény kifejezését vagy annak nyújtását, vagy pedig ez utóbbi megköszönését tartalmazza. Figyelemre méltó, hogy a vizsgált hozzászólások valamivel több mint fele társas támogatás nyújtását tartalmazza.

1. ábra: Társas támogatás előfordulása a hozzászólások között (N=2336)



A társas támogatást nyújtó hozzászólásokat (N=1277) ezt követően besoroltam valamelyik támogatási kategóriába. Az elemzés során nem az öt főkategória, hanem a 23 alkategória megjelenését vizsgáltam a hozzászólásokban, mivel azok könnyebben azonosíthatók, és egyben beazonosítják magát a főkategóriát is. Két rövid hozzászólás esetében nem volt eldönthető az alkategória, így azt a két hozzászólást végül kihagytam az elemzésből.

2. ábra: Alkategóriák előfordulási gyakorisága a társas támogatást nyújtó hozzászólásokban (N=1275)



A maga 28,8 százalékos megjelenésével egyértelműen a leggyakoribb¹³ a *tanítás* típusú támogatás, amikor valamilyen információval, tudással segítik egymást a fórumhasználok. Tanítás típusú hozzászólásra jellegzetes példa a következő két hozzászólás.

„Nagyon hasonló történetünk van a Zolofittal. És én IMÁDOM a klonopint. Először elálmosíthat, szóval nem szedheted napközben.”¹⁴

„De ha úgy érzed, a Prozac még olyanabbá tesz vagy növeli a problémákat vagy ront a helyzeten ahelyett, hogy javítana a kezdeti fázisában, habozás nélkül hívd fel az orvosodat és mondd el neki, és kérdezd meg, hogy lehetne-e, hogy alacsonyabb dózison, 20 mg-on tartson helyette [...] de újra, 2 nap nagyon kevés, hogy meg lehessen mondani. Ugyanígy, ne hagyj abba soha semmilyen gyógyszert az orvosod jóváhagyása nélkül, mert az is okozhat még rosszabb problémákat is.”

Ezek a hozzászólások jellemzően gyógyszerekkel kapcsolatos saját tapasztalatot, információt tartalmaznak, azon belül is a hatásosságukra, mellékhatásaikra vagy a gyógyszerek szedésének, elhagyásának módszereire vonatkoznak. Ez utóbbival kapcsolatban gyakran megjelent, hogy idő kell a gyógyszereknek, amíg hatnak, ne keseredjen el a használó a gyógyszer elkezdése utáni napokban. Meg kell említeni, hogy a

¹³ Mivel egy hozzászólásnál minden olyan kategóriát feltüntettem, ami abban előfordult, így az itt közölt százalékos arányok azt mutatják, hogy a társas támogatást nyújtó hozzászólások mekkora arányában jelent meg az adott kategória.

¹⁴ A fordítások során megőriztem az eredeti hozzászólások írásmódját (például nagybetűk, írásjelek használata).

gyógyszerszedésre vonatkozó információk – amennyiben a szedésre vonatkozó konkrét javaslatot adnak – szakértői tudás nélkül veszélyesek lehetnek, így kockázatokat rejthet, ha ilyen döntéseket csak laikus fórumok alapján hoznak meg az érintettek. A következő idézet jól példázza az ilyen típusú hozzászólásokat.

„Ne csináld, a xanax erősen addiktív lehet és ha más anyagokkal kevered, halálos lehet. Én egyik gyógyszert sem ajánlom a benzo-k közül.”

Ez a jelenség az online (mentális) egészséggel foglalkozó fórumok használatának veszélyeihez tartozik, a vizsgált fórumon azonban sok esetben megjelent, hogy orvos nélkül ilyen kérdésekben ne döntsön senki, mivel a fórumhasználók csak saját tapasztalataikat tudják egymásnak átadni, egy gyógyszer hatása viszont egyénenként változó lehet. A hozzászólók jellemzően el is tanácsolták egymást a gyógyszerek hirtelen történő abahagyásától („cold turkey”) vagy meggondolatlan cseréjétől, ahogyan a második idézetben is olvasható. Fontos megemlíteni, hogy a szakértők bevonására tett javaslatok (például orvos, pszichiáter, pszichológus felkeresése) nem ebbe a kategóriába, hanem az *ajánlatba* tartoznak (amely a hozzászólások 10,4 százalékában fordult elő).

A hozzászólások több mint ötödében (20,6 százalék) fordult elő konkrét, praktikus tanács, amely általánosabban a szorongás, depresszió kezelésére vonatkozott. Ilyen javaslat volt a naplóírás, sport, meditáció vagy hasznos gondolati sémák alkalmazása. Idetartoztak azok a hozzászólások is, amelyekben konfliktushelyzet esetén a családtagokkal való beszélgetést javasolták. Az alábbi két hozzászólás jól illusztrálja az altípusba tartozó hozzászólásokat.

„Én azt mondom, figyelj magad folyamatosan. Emellett aludj eleget, egyél megfelelően és mozogj. A mozgás sokat segíthet”

„Esetleg gondolj arra, hogy az ezzel kapcsolatos gondolataid irracionálisak és mondd magadnak, hogy egészséges vagy. A negatív gondolatokat le lehet győzni pozitív és igaz gondolatok ismétlésével”

A korábbi két kategóriához képest jóval ritkábban, átlagosan 100-ból 12-14 hozzászólásban nyújtanak bátorítást, fejeznek ki megértést vagy szimpátiát a fórumhasználók. Bátorítás esetén jellemzően egy megkezdett, pozitív változásokat előidéző tevékenység folytatására biztatták egymást a fórumhasználók, vagy arra, hogy ne adják fel, később jóra fordulnak a dolgok. A megértést tartalmazó hozzászólásokban jellemzően annak kifejezése jelenik meg, hogy a segítséget nyújtó átérzi, tapasztalta azt a helyzetet, amelyet az eredeti hozzászólás írója megfogalmazott. A szimpátia kategóriájába tartozó hozzászólásokban olyan mondatokat találunk, amelyek sajnálatot fejeznek ki, vagy éppen reményt azzal kapcsolatban, hogy a korábbihoz képest már javult a másik személy helyzete. Ezekre a kategóriákra is jellemző, hogy egy hozzászólásban

egyszerre több kategória is megjelenik, ahogyan a példamondatokon is látszik. Az elsőben a bátorítást és megértést, míg a másodikban a szimpátiát és imát találjuk meg:

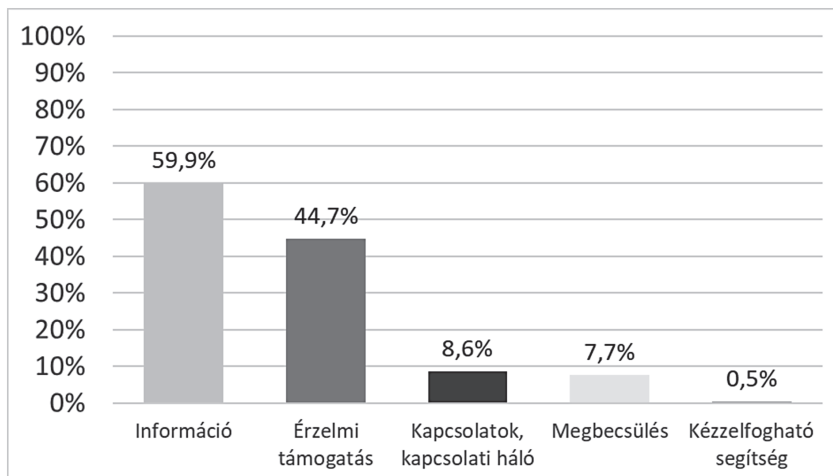
„Csak azt akarom neked mondani, hogy 2015-ben összeomlottam úgy, hogy két gyereket nevelek, és én mondom, a dolgok jobbra fordulnak majd és túl fogsz jutni ezen!!”

„Nagyon sajnálom, hogy ilyen nehéz időszakon mész most keresztül. [...] Imádkozom, hogy találj egy megoldást, ami segít fizikálisan és érzelmileg is jobban érezni magad.”

Mivel egy hozzászólás több kategóriába is tartozhat, érdemes megnézni az alkategóriák együttes előfordulását is (függelék, 1. táblázat). Jellemzően az azonos főkategóriába tartozó alkategóriák jelennek meg egyszerre egy hozzászólásban, illetve a gyakoriságoknak megfelelően a bátorítás és megértés fordul elő gyakran az információnyújtásba tartozó tanács, tanítás és ajánlat kategóriákkal. Emellett a bátorítás és bók kategóriák fordulnak elő még viszonylag gyakran (13 esetben) egyszerre, amely azt mutatja, a hozzászólót vélt előnyeit, képességeit kiemelve biztatták. Az érzelmi jellegű támogatási típusok mellett megjelenik még több esetben a jelenlét kifejezése, vagyis a szimpátiát, megértést vagy bátorítást kifejező hozzászólásokban a támogatók egyúttal arról is biztosítják az érintettet, hogy elérhető, ha szükség van rájuk (például privát üzenet formájában).

Az alkategóriákat ezt követően összevontam főkategóriákká. Ahogyan az alkategóriákat mutató ábráról is szembetűnő: az információs és az érzelmi típusú támogatás fordult elő leggyakrabban a hozzászólásokban. A kapcsolatokhoz vagy a megbecsüléshez tartozó kategóriák elenyésző arányban jelennek meg a hozzászólásokban. A fórum virtuális voltából adódóan nem meglepő, hogy a kézzelfogható segítség szinte egyáltalán nincs jelen. Ebbe a kategóriájába csak azok a hozzászólások kerültek, amelyek jövőbeni, általános segítséget ajánlottak fel.

3. ábra: Fő kategóriák előfordulási gyakorisága a társas támogatást nyújtó hozzászólásokban (N=1275)



A társas támogatást nyújtó hozzászólásokban tehát az információ típusú támogatás van túlnyomó többségben (59,9 százalék), de gyakori az érzelmi támogatás nyújtása is, mely a hozzászólások 44,7 százalékában található meg. Az emberi kapcsolatokon alapuló támogatási formát tartalmazó hozzászólások jellemzően magára a fórumközösségre utalnak, arra, hogy ott vannak egymásnak a fórum használói, de idetartoznak azok a hozzászólások is, amelyben azt ajánlja fel valaki, hogyha az illetőnek szüksége van rá, bármikor írhat neki. A megbecsülés fő kategóriáján belül elsősorban a büszkeség kifejezése jelenik meg, illetve az, ha az érintett előnyeit, képességeit emelik ki. Ez utóbbi esetben érdemes szem előtt tartani, hogy a fórumhasználók egymást csak a fórumhozzászólások vagy üzenetek és az azokban (saját nézőpontból) leírt történetek alapján tudják megítélni, holott a megbecsülés kategóriához a másik személy alapos ismerete szükséges. A kapcsolatok kategória pedig közösen elérhető társaságot feltételez, amely az ország vagy akár világ másik pontján élő személlyel – hacsak nem online társaságról van szó – nem áll fenn. Ez egyben azt is jelenti, hogy egy online fórumon a megfigyelt-nél nagyságrendekkel nagyobb előfordulás nem is feltétlenül várható.

Automatikus kategorizálás

Az osztályozás automatizálhatóságának elemzése során természetesnyelv-feldolgozással vizsgáltam,¹⁵ hogy a különböző társas támogatási formák elkülöníthetők-e egymástól gépi tanulással

A természetesnyelv-feldolgozás során a szövegeket bizonyos előfeldolgozási lépéseknek kell alávetni, melyek a probléma jellegétől függenek. Ilyen lépés a szótövezés (lemma-

15 Az elemzést a saját magam által Python v3.7. programnyelven írt kód segítségével végeztem különböző csomagok felhasználásával.

tizálás), azaz a különböző formában szereplő (például toldalékolt) szavak azonos alakra hozása (Tikk 2007), így például a „has”-ból „have” lesz. A szövegben gyakran előforduló szókapcsolatok (két szó esetén bigramok, három szó esetén trigramok) azonosítása lehetővé teszi, hogy a számítógép a feldolgozás során együtt kezelje a több szóból álló kifejezéseket, elkülönítve az akár többletjelentéssel bíró kifejezést az azt alkotó egyedi szavaktól. További lépés lehet, hogy az ún. stopszavakat eltávolítjuk a szövegből, „koncentráltabbá” téve a szövegekörpuszt. A stopszavak olyan gyakori szavak, amelyek önálló jelentéssel nem bírnak, ezért a szövegek egymástól való elkülönítésében, klasszifikációjában sem játszanak jelentős szerepet. Ilyen szavak a határozószók, névmások és prepozíciók.

Ahhoz, hogy a szöveget a számítógép számára értelmezhetővé tegyük, a tisztított szövegeket számokká, számvektorokká kell alakítani. Ennek két gyakran alkalmazott módja a tf-idf vektorizálás és a count vektorizálás. A tf-idf figyelembe veszi az egyes szavak dokumentumokon (jelen esetben hozzászólásokon) belüli gyakoriságát és azt is, hogy a dokumentumok mekkora arányában fordul elő az adott szó (Tikk 2007). A count vektorizálás ezzel szemben csak azt veszi figyelembe, hányszor fordult elő egy szó egy szövegben, vagy bizonyos esetekben csak azt, hogy előfordult-e egyáltalán.

A feldolgozott szövegen futtatott osztályozómodell teljesítményének megítéléséhez egy olyan adatbázison kell azt tesztelni, amelyben a kategóriák már ismertek az egyes szövegekre, de azok nem azonosak a tanító adatbázissal, amelyen a modell tanult. Ennek – főként kis méretű adatbázis esetén – egyik módja a keresztvalidálás. Keresztvalidálás során a tanító adatokat k (például öt vagy tíz) egyenlő, diszjunkt részre bontjuk véletlenszerűen, ahol $k-1$ rész szolgál tanító adatbázisként, 1 egység pedig tesztadatként. A tanítást és tesztelést k -szor végezzük el, mindig egy-egy egységet hagyva tesztalumnak. Ekkor a modell teljesítményét a k futtatás átlagos eredményeként írjuk le, megalapozottabb képet kapva a teljesítményről, mintha csak egyszer futtattuk volna le a modellt.

A modell teljesítményének értékelésére használt gyakori mérőszám a pontosság (*accuracy*). Figyelembe kell venni, hogy a pontosság azonban ilyen egyenlőtlen osztályeloszlás esetén jobb eredmény felé torzíthat, mivel már egy egyszerű valószínűség-alapú becsléssel is viszonylag jó eredményt mutathat.¹⁶ Ezért érdemes vizsgálni a *Cohen-kappa* (Landis–Koch 1977) értékét, amely a véletlenszerű osztályozáshoz képest mutatja a teljesítményt. A felidézés (*recall*) és precizitás (*precision*),¹⁷ valamint e kettő harmonikus átlagaként számolt ún. *F1-score* érték kategóriánként vizsgálja az osztályozás jószágát. A felidézés azt adja meg, hogy az adott kategóriába tartozó hozzászólások mekkora arányát sorolta a modell valóban abba a kategóriába, a precizitás pedig azt, hogy az adott kategóriába sorolt hozzászólások mekkora arányát kellett valóban abba a kategóriába sorolni.

16 Ha például az egyik kategóriába 100, a másikba 900 eset tartozik, akkor az összes esetre a domináns kategóriát becslve is elérhető 90 százalékos pontosság.

17 A *precision* magyar szakirodalomban jellemzően pontossággként fordítják (Sebők et al. 2016, Tikk 2007), én azonban az *accuracy* szótól való megkülönböztetés érdekében a *precizitás* fogalmát használom, akárcsak a Bevezetés az adatbányászatba című könyv (Tan et al. 2012) fordítója. Tikk az *accuracy* szót szabotosságnak (2007), míg Sebők és társai (2016) hitelességnek fordítják.

Az adatbázis előkészítése

A teljes adatbázis 2336 hozzászólást tartalmazott, a legfeljebb három szóból álló hozzászólások¹⁸ eltávolítása után a természetesnyelv-feldolgozáshoz szükséges előfeldolgozásnak vetettem alá a szöveget. Első lépésként kitöröltem minden olyan karaktert, amely nem betű, szám vagy szóköz, majd lemmatizáltam a szöveget, és azonosítottam a szövegben előforduló bigramokat. Ezt követően eltávolítottam a társas támogatást (nyújtását vagy kérését) nem tartalmazó hozzászólásokat, így 1408 hozzászólás maradt az adatbázisban. Mivel a kézzelfogható segítség elenyésző számban jelent meg a fórumban (4 eset), azokat az eseteket kizártam az automatikus osztályozási folyamatból. Ugyanígy tettem azokkal a támogatást kérő hozzászólásokkal, amelyekben nem volt egyértelműen azonosítható a támogatási típus (8 eset).

Egyosztályos kategorizáció

Az adatbázist továbbá átalakítottam úgy, hogy egy-egy hozzászóláshoz csak egy fő-kategória tartozzon, egyszerűbbé téve az osztályozási problémát a modell számára. Azokban az esetekben, amikor egy hozzászólás több támogatási fő-kategóriát is tartalmazott, az lett a kiválasztott fő-kategória, amelyből a hozzászólás több alkategóriát tartalmazott. Azokat a hozzászólásokat (192 eset), ahol az alkategóriák között a relatív többségi elv alapján nem lehetett dönteni, kitöröltem az adatbázisból.¹⁹ Ezen szűrések végrehajtása után 1204 hozzászólás maradt. Ebben az adatbázisban 751 hozzászólás tartozott az információs, 373 az érzelmi, 43 a kapcsolatokra, kapcsolati hálóra és 37 az önbecsülésre vonatkozó társas támogatási típusba, a csoportok eloszlása tehát meglehetősen egyenetlen volt. Mivel a támogatást kérő hozzászólások jellegükben eltérnek a támogatást nyújtó hozzászólásoktól (hosszabbak, szerteágzóbb témákat, így változatosabb szövegeket tartalmaznak), és eltérő mintakeresési feladat a támogatás nyújtásának és arra irányuló kérésnek a felismerése, a kutatás végső szakaszában²⁰ csak a támogatást nyújtó hozzászólásokra építettem a modelleket (N=1030).

Mivel az önbecsülés (37 esettel) és a kapcsolatok, kapcsolati háló típusú társas támogatás (36 esettel) meglehetősen ritka a vizsgált adatbázisban, a tanulási szakaszban is kevés adat áll rendelkezésre a mintázatok felismeréséhez. Ennek következményeként a modellek ezekbe a kategóriákba nem sorolnak eseteket, tehát nem használhatóak ezek felismerésére, ellenben az osztályozás jóságának megítéléséhez pontatlan eredményeket kapunk. Így tehát azokat a hozzászólásokat, amelyek ezekbe a kategóriákba tartoztak, kizártam az elemzésből. Ezt a döntést erősíti, hogy korábbi

18 Az ennél rövidebb, 14 hozzászólás jellemzően csak születésnapjü köszöntést vagy gyógyszernevet tartalmazott, így inkább csak zajként jelentek volna meg az osztályozásban.

19 Ez ugyan viszonylag nagy mértékű adatvesztés, viszont korábbi tapasztalataink alapján (Németh, Sík, Máté, 2020) a modell teljesítménye jobb, ha kevesebb, de biztosabban azonosítható adattal dolgozik, mint ha több, de bizonytalanabb esettel. Ilyenkor fennáll annak a lehetősége, hogy új, ilyen módon „nem kipreparált” adatok besorolása során nem működik megfelelően az algoritmus, elsődleges célom azonban az volt, hogy feltérképezzem a lehetséges legjobb körülmények között a modellek általános teljesítményét.

20 A támogatást kérő és nyújtó hozzászólásokra épített legjobb modellek eredményei a függelékben megtalálhatók.

kutatások is elsősorban az információs és érzelmi támogatás fontosságát emelték ki az online segítő fórumok kapcsán (Braithwaite et al. 1999, Coulson 2005), így a kutatói érdeklődés fókuszában is sokszor ezek a kategóriák állnak. Az így előállt adatfájl 602 információs típusú támogatást nyújtó és 355 érzelmi típusú támogatást nyújtó hozzászólás képezi.

A szövegek feldolgozásának változtatása modellben

A fentiek alapján előállított szövegeket háromféle formában alkalmaztam a modellezés során: (1) a stopszavakat eltávolítva; (2) benne hagyva a szövegben; illetve (3) a stopszavakat ugyan meghagyva, de a túl rövid, csak két karakterből álló szavakat kitörölve a szövegből. Ez utóbbi lépés a prepozíciók, rövidítések és a „be” létige törölését szolgálta, amelyek a hozzászólások témájának meghatározását jelentősen nem befolyásolják (hiszen mindegyik csoportba tartozó szövegben gyakran előfordulnak), ellenben a modell működését lassítják. A számítások során ötszörös keresztvalidálást alkalmaztam, vagyis véletlenszerűen kiválasztott 765 hozzászóláson tanult a modell, és a 192 hozzászóláson tesztelte a tanulás eredményességét.

A modellek jóságának eredményeit a különböző mérőszámok alapján a 2–4. táblázat tartalmazza.

2. táblázat: A naiv Bayes-osztályozó teljesítménye különböző módon feldolgozott szövegek esetén (N=192)

	Naiv Bayes			
	Pontosság	Precizitás	F1-score	Cohen-kappa
Stopszavazva	0,696	0,791	0,620	0,22
Stopszavazás nélkül	0,670	0,784	0,571	0,13
Stopszavazás nélkül, hosszabb szavakkal	0,670	0,784	0,570	0,13

3. táblázat: A logisztikus regressziós osztályozó teljesítménye különböző módon feldolgozott szövegek esetén (N=192)

	Logisztikus regresszió			
	Pontosság	Precizitás	F1-score	Cohen-kappa
Stopszavazva	0,733	0,770	0,691	0,34
Stopszavazás nélkül	0,783	0,814	0,759	0,48
Stopszavazás nélkül, hosszabb szavakkal	0,757	0,788	0,727	0,41

4. táblázat: Az SVM osztályozó teljesítménye különböző módon feldolgozott szövegek esetén (N=192)

	SVM			
	Pontosság	Precizitás	F1-score	Cohen-kappa
Stopszavazva	0,726	0,785	0,676	0,32
Stopszavazás nélkül	0,775	0,807	0,750	0,46
Stopszavazás nélkül, hosszabb szavakkal	0,752	0,793	0,719	0,40

A fenti eredmények alapján egyfelől látható, hogy összességében a logisztikus regresszió nyújtja a legjobb teljesítményt mindegyik mérőszám alapján, a naiv Bayes pedig a legrosszabbat, amely a véletlen besoroláshoz képest nem nyújtott érdemben jobb teljesítményt (Cohen-kappa értékek 0,2 alattiak). A szövegek tisztítása szempontjából mindegyik mérőszám azt mutatja, hogy az SVM és a logisztikus regressziós modell akkor a legjobb, ha a legteljesebb, legkevésbé módosított szövegre alkalmazzuk, míg a naiv Bayes a stopszavazott szövegen mutatta a legjobb teljesítményt. A Cohen-kappa érték értelmezésének irányadó műve (Landis–Koch 1977) alapján azt mondhatjuk, a legjobb logisztikus regressziós és SVM modellek csak mérsékelten nyújtanak jobb teljesítményt, mint a véletlenszerű besorolás.

Ahhoz, hogy lássuk, miként működött a gépi tanulás, milyen szavak alapján „döntött” a gép, megvizsgáltam a legjobb, stopszavazás nélküli logisztikus regressziós modell regressziós együtthatóit. Az információs társas támogatás kategóriájánál legmagasabb együtthatóval rendelkező húsz szó között olyan, jelentéssel bíró szavakat találunk, mint a „try” (próbál), „help” (segíteni, segítség), „taking” (szedés), „side_effect” (mellékhatás), „should” (kellene), „doctor”, „mg”. E szavak egyértelműen felrajzolják, hogy ebben a kategóriában tanácsok, azon belül is elsősorban a gyógyszereszedéssel kapcsolatos tapasztalatok, javaslatok fogalmazódnak meg.

Az érzelmi társas támogatáshoz kapcsolódó fontos szavak elsősorban a szimpátia kifejezéséhez és a bátorításhoz tartoznak, amely alkategóriák a kategórián belül is gyakoriak voltak az adatbázisban. Ilyen például a „hope” (remény, remélni), „good_luck” (sok szerencsét), „luck” (szerencse), „sorry” (sajnálni), „wish” (kívánni, kívánság). Érdemes kiemelni, hogy ebben a kategóriában a második legmagasabb együtthatóval az „im” (vagyis az „I am”, „én, én vagyok” összevont alakja), az ötödik legmagasabbal pedig a „you” (te, ti) szerepel. A stopszavazott vagy a rövid szavakat nem tartalmazó szövegben ezek a kifejezések nem szerepelnének, de úgy tűnik, az érzelmi támogatás felismerésében jelentős szerepet játszanak, ami eredhet a „sajnálalak téged / sajnálom a helyzetedet” kifejezések kategórián belüli gyakori előfordulásából (erre utal a szintén szereplő „so” [ebben a kontextusban nagyon, annyira] is), például:

„Sajnálom, hogy így érzel a munkáddal kapcsolatban.”

A legfontosabb szavakban a „happy” mellett a „birthday” (születésnap) is szerepel. Ennek oka: az egyik alfórum arról szólt, hogy az illető 64 éves lett, és így számos köszöntést kapott.

A fenti elemzések során legjobban teljesítő logisztikus regressziós modell esetén készítettem még három, komplexebb modellt is, amelyekbe a szövegeken túl más független változókat, vagyis feature-öket is bevontam. Az első modellben csak a hozzászólás előfeldolgozáson átesett szövege szerepelt (ugyanúgy, ahogy a modelleket összehasonlító elemzésben is), a második modellben a hozzászólás szövege mellett az alfórum címét is bevontam a modellbe, a harmadik esetben pedig az előbbieket mellett azt is beépítettem, hogy előfordul-e a hozzászólásban gyógyszernév – amely információ főként az információs támogatás kategóriájának felismerését segíti. Az alfórum nevét az abban megfogalmazott tématerület, a fórum eredeti céljának ismerete miatt építettem be a modellbe. Az alfórum témája az információs jellegű támogatási típus felismerésében például a gyógyszerek használatával kapcsolatos szavak (például „Szedett valaki Xanaxot korábban?”) jelenléte miatt segíthet; az érzelmi típusú támogatás kategóriába való besorolást a „need” (szükség) vagy „help” (segítség) szavak mozdíthatják elő. Az alfórum nevét nem a hozzászólásoknál használt tf-idf-alapú vektorizálással, hanem előfordulás-alapú vektorizálással alakítottam át, ahol a vektor értékei csak azt mutatják, hányszor fordult elő az adott szó az adott szövegben. Ez a fajta vektorizációs eljárás a tf-idf módszerhez képest abban különbözik, hogy nem veszi figyelembe, hogy a dokumentumok mekkora arányában fordul elő az adott szó, és a szöveg hosszát sem. Mivel a fórumcímek beépítésének célja csak a fórum témájának meghatározása volt, elegendő volt az az információ, hogy milyen szavak és hányszor fordultak elő a címben, nem volt szükség annak az információnak a beépítésére, hogy adott szavak a többi fórumcímben hányszor fordultak elő. Továbbá egy-egy fórumcím meglehetősen rövid, csupán néhány szóból áll, így szintén nem volt szükséges a szöveg hosszával arányosítani az előfordulást, mivel nem feltételeztem összefüggést a szó jelenléte és a szöveg hossza között.

5. táblázat: Az alkalmazott logisztikus regressziós modellek teljesítménye

„Független változó” módosításai	Csak a hozzászólás szövege		A hozzászólás szövege és az alfórum címe		Szöveg, alfórum név, gyógyszer előfordulás	
	Pontosság	Cohen-kappa	Pontosság	Cohen-kappa	Pontosság	Cohen-kappa
A szöveg módosításai						
Stopszavak nélkül	0,733	0,34	0,762	0,46	0,767	0,48
Stopszavakkal együtt	0,783	0,48	0,760	0,46	0,771	0,49
Stopszavakkal együtt, de rövid szavak nélkül	0,757	0,41	0,764	0,47	0,768	0,48

A fenti táblázatban láthatók a háromféle szövegmódosítást és az osztályozást segítő információk kombinációiból előálló logisztikus regressziós modellek pontosság- és a Cohen-kappa értékei. A pontosságot és a Cohen-kappát együttesen figyelembe véve megállapítható, hogy a legjobb modell továbbra is a legegyszerűbb, amely semmilyen más tényezőt nem vesz figyelembe az osztályozáskor, csak a stopszavakat tartalmazó szöveget. Mindazonáltal a legmegbízhatóbb teljesítményt az alfórum nevét és a gyógyszer előfordulását is figyelembe vevő modell mutatja, ami eredhet a rendelkezésre álló információk mennyiségéből.

Többosztályos besorolás

Mivel egy hozzászólás több kategóriába is tartozhat, a fent ismertetett, csak főkategóriát becsülő osztályozó túlegyszerűsíti a problémát. Így megvizsgáltam azt az osztályozási feladatot is, hogy az adott kategória előfordul-e a hozzászólásban, vagy sem. Ehhez az osztályozási kérdéshez tehát nem kellett eltávolítani olyan eseteket, amelyeknél nem lehetett egyértelműen meghatározni egy kategóriát. Szintén az adatok között maradhatnak, és így további információval szolgálhatnak a két kisebb (kapcsolatok, kapcsolati háló és megbecsülés) kategóriából származó esetek. Ezt a modellt is a társas támogatást nyújtó hozzászólások besorolására használtam, így az adatfájl végül 1219 esetet tartalmazott. A szövegek vektorizációját szintén tf-idf vektorizációval végeztem, és szintén három formában építettem be a szövegeket a modellekbe. A modellek validálása itt is ötszörös keresztvalidációval történt (a tanulóadat nagysága 975 eset, a tesztadat pedig 244). A modelleket mind a négy kategória esetében megvizsgáltam, de a két kisebb kategóriát ekkor sem ismerték fel a modellek. A modellek teljesítményét az alábbi táblázatok tartalmazzák.

6. táblázat: Naiv Bayes-modell az információs és érzelmi támogatási kategóriák előfordulására (N=244)

Naiv Bayes								
	Információ típusú társ. tám.				Érzelmi típusú társ. tám.			
	Pont.	Precizitás	F1-score	Cohen-kappa	Pont.	Precizitás	F1-score	Cohen-kappa
Stopsz.	0,686	0,788	0,608	0,21	0,669	0,698	0,635	0,28
Stopsz. nélkül	0,660	0,780	0,558	0,13	0,659	0,722	0,605	0,25
Stopsz. nélkül, hossz.	0,663	0,782	0,565	0,14	0,662	0,715	0,613	0,26

7. táblázat: Logisztikus regressziós modell az információs és érzelmi támogatási kategóriák előfordulására (N=244)

Logisztikus regresszió								
	Információ típusú társ. tám.				Érzelmi típusú társ. tám.			
	Pont.	Precizitás	F1-score	Cohen-kappa	Pont.	Precizitás	F1-score	Cohen-kappa
Stopsz.	0,744	0,778	0,712	0,39	0,689	0,699	0,673	0,34
Stopsz. nélkül	0,772	0,794	0,752	0,47	0,717	0,724	0,706	0,40
Stopsz. nélkül, hossz.	0,758	0,782	0,735	0,43	0,701	0,708	0,689	0,37

8. táblázat: SVM modell az információs és érzelmi támogatási kategóriák előfordulására (N=244)

SVM								
	Információ típusú társ. tám.				Érzelmi típusú társ. tám.			
	Pont.	Precizitás	F1-score	Cohen-kappa	Pont.	Precizitás	F1-score	Cohen-kappa
Stopsz.	0,733	0,775	0,695	0,36	0,685	0,704	0,661	0,32
Stopsz. nélkül	0,778	0,801	0,760	0,49	0,715	0,726	0,701	0,39
Stopsz. nélkül, hossz.	0,759	0,786	0,735	0,43	0,696	0,711	0,678	0,35

A modellek teljesítménye hasonló az egyosztályos problémához, csak néhány százados eltérést találhatunk, szisztematikus eltérés egyik irányba sem figyelhető meg, kivéve a naiv Bayes-modellnél, ami rosszabb eredményeket mutat ennél az osztályozási feladatnál. A két társas támogatási típus eredményei azonban jelentősen eltérnek egymástól, az információs támogatást magabiztosabban sorolja be a modell, mint az érzelmi támogatásba tartozó eseteket. Ennek oka lehet az érzelmi támogatás kisebb esetszáma, de az is, hogy ebben a kategóriában egymástól jobban elkülönülő témák jelennek meg.

A stopszavazás nélküli logisztikus regressziós modell együttthatóit vizsgálva az egyosztályos modellhez hasonló szavakat találhatunk. Az információs típusú támogatásnál a „try”, „side_effect”, „medication”, „should” szavak mellett a húszt legmagasabb együttthatójú szavak között szerepel a „Prozac” antidepresszáns neve is. Az érzelmi kategóriánál a „hope”, „you” és „im” a három legmagasabb együttthatóval rendelkező szó, a „sorry”, „good_luck” mellett a „good” (jó), „sending” (küldés, például ölelést), „glad” (örülni), „feel_better” (jobban érezni [magát]), „hug” (ölelés) és „best” (legjobb) kifejezésekkel egészült ki a legmagasabb együttthatójú szavak listája. Az új szavak a szimpátia mellett a bátorítás alkategóriákhoz sorolhatók, amelyek megerősítik azt, hogy a szimpátia kategória fontos szerepet játszik az osztályozásban annak ellenére, hogy a három főkategória közül a legritkább – igaz, csak két százalékponttal ritkább, mint a bátorítás kategória.

Összefoglalás

Kvalitatív kutatási kérdésemre, miszerint „Mennyire gyakori az adott fórum hozzászólásaiban a társas támogatás megjelenése, és azon belül melyik fő- és alkategóriák fordulnak elő leggyakrabban?” az a válasz adható, hogy a fórum hozzászólásainak több mint fele, 54,7 százaléka nyújt valamilyen típusú társas támogatást. A két domináns típus az információs és érzelmi jellegű támogatás, amelyek a hasonló élethelyzetben lévők információval, tanáccsal történő támogatását fejezik ki, illetve megértést, szimpátiát és bátorítást. Jellemző továbbá az is, hogy egy hozzászólás többféle kategóriát tartalmaz. A kvalitatív elemzésem eredményei megegyeznek Prescott és munkatársai (2017) következtetéseivel, miszerint a mentális egészséggel kapcsolatos fórumokon elsősorban információra számíthatnak a felhasználók, azonban eltérnek Deetjen és Powell (2016) eredményeitől, akik azt találták, hogy mentális egészséggel kapcsolatos témákban az érzelmi támogatás dominál az információ típusú támogatással szemben. Az általam elemzett fórumon az alkategóriákat vizsgálva látható, hogy elsősorban tudásátadás és konkrét tanácsok nyújtása jelenik meg, az érzelmi jellegű támogatás pedig főként a hozzászólást író helyzetének megértését, sajnálat kifejezését vagy bátorítást jelent a vizsgált fórumon.

Az automatizált klasszifikációs modellezéssel kapcsolatban kutatási kérdésem az volt: „Milyen mértékben tárható fel automatikus szövegosztályozási módszerekkel az, hogy egy-egy hozzászólás melyik társas támogatási kategóriába tartozik?” A kérdés megválaszolása érdekében kétféle osztályozási problémát vizsgáltam, háromféle algoritmussal, háromféleképpen feldolgozott szövegen. Kezdeti eredményeim nyilvánvalóvá tették, hogy a teljes szövegen belül kis számosságban és alacsony arányban előforduló társas támogatási kategóriák (megbecsülés és kapcsolatok, kapcsolati háló) felismerésére a modellek nem alkalmazhatók. Ahhoz, hogy a szövegekben alacsony gyakoriságú kategóriák is felismerhetővé váljanak a modell számára, nagyságrendekkel több adatot tartalmazó tanító adatbázis szükséges. Az adatbázis növelése egyben a humán erőforrás növelését is maga után vonná. Kódolók tömegeinek elérésére többféle ún. crowdsourcing platform érhető el, például az Amazon Mechanical Turk és a CrowdFlower, amelyeket nyelvészeti kutatásokban széles körben használnak (Sabou et al. 2014), megválaszolandó kérdésként merül fel azonban, hogy mennyire alkalmasak speciális előzetes tudást igénylő (szociológiai) kutatásokhoz.

Amikor a modelleknek a hozzászólásokat csak egy osztályba kellett besorolniuk az információs és érzelmi típusú támogatás közül, az alkalmazott logisztikus regresszió és SVM modell 70 százalék feletti pontosságot és a véletlenszerű besoroláshoz képest legalább 30 százalékkal jobb teljesítményt mutatott. Ettől a jellemzően gyengébb teljesítményt nyújtó naiv Bayes algoritmus maradt el, 67 százalékos pontossággal és a véletlen besorolásnál minimálisan jobb teljesítménnyel. Az alkalmazott modelleknél 2-5 századnyi különbséget eredményeztek a különbözően feldolgozott szövegek, a legjobb teljesítmény az eredeti szöveg legteljesebb megőrzésekor figyelhető meg logisztikus regresszió és SVM esetén, mely rámutat a szövegek feldolgozására vonatkozó dönté-

sek fontosságára. Azonban, ha nem csak a szöveg, hanem egyéb információk (alfórum címe és gyógyszernév előfordulása a szövegben) is segítették a modell döntését, robusztusabb eredményeket kaptam, a szövegek feldolgozásának különbségei nem eredményeztek tényleges eltérést a modellek teljesítményében. Ugyanakkor a modellek jobb teljesítményt nyújtottak, mint a stopszavak vagy rövid szavak nélküli modellek. Az egyosztályos megkötés hátránya, hogy csupán részben teljesíti az elvárásokat, amennyiben csak a dominánsabb témát ismeri fel a hozzászólásban, illetve a domináns témával nem rendelkező hozzászólások eltávolításával csökken az adatmennyiség.

Amikor az osztályozómodelleknek csak azt kell felismerniük, hogy előfordul-e az információs vagy érzelmi jellegű támogatás a hozzászólásban, nincs szükség a csak egy kategóriába be nem sorolható adatok eltávolítására és előzetes megkötésekre sem. Emellett jobban modellezi azt a feladatot, pontosabban illeszkedik arra, hogy a hozzászólásokban megjelenő társas támogatási típusokat azonosítsuk. Ezen modell eredményei is az egyosztályos modell eredményeihez hasonlóak, de annál néhány századdal gyengébbek voltak, ami eredhet pontosan abból, hogy nem csak az egyértelműen egy kategóriába sorolható hozzászólások alkották az adatbázist. Az információs típusú társas támogatás emellett magabiztosabban felismerhető a modell által, mint az érzelmi típus, bármelyik modellspecifikációt is vizsgáljuk.

A vizsgált együttthatók alapján a logisztikus regressziós modell működése mindkét kérdés esetében érvényesnek mondható: valóban azok a szavak bírnak jelentős súllyal az osztályozási döntésben, amelyek a szövegek kvalitatív elemzése során tapasztalt témákhoz kapcsolódnak. Összességében elmondható, hogy a társas támogatási kategóriák automatizált felismerésének eredményei bizakodásra adnak okot, a modellek ismeretlen hozzászólásokra való megbízható alkalmazhatóságához azonban a vizsgált modellek teljesítményében még további fejlesztések szükségesek.

A kutatás egyik korlátja, hogy csak egy személy (a cikk írója) osztályozására alapszik; megbízhatóbb képet kapnánk akkor, ha többen kategorizálták volna a hozzászólásokat, és ezen kategorizációkat együttesen vettük volna figyelembe az osztályozás során.

A kutatás célja az volt, hogy kiindulópontként szolgáljon egy későbbi, nagyobb léptékű kutatáshoz, amelyben olyan technikákat tervezünk kidolgozni, amelyekkel nagy megbízhatósággal azonosíthatóvá válnak a társas támogatási típusok fórumszövegekben. Ahhoz azonban, hogy valóban hatékony és megbízható eredményekhez jussunk, egyfelől több adat szükséges, amelyben az egyes, főleg ritkán előforduló kategóriák (kapcsolatok és megbecsülés) számossága is megfelelő; másfelől pedig olyan adat, amely többféle fórumról származó hozzászólásokat tartalmaz, hogy az adott fórumot használók nyelvezetétől kevésbé függjenek a modellek. Mind a kvalitatív, mind a kvantitatív elemzés eredményei alapján érdemes lehet egy olyan osztályozó algoritmus tesztelése, ahol az elsődleges osztályozandó egységet nem a hozzászólások, hanem azok egyes mondatai képzik, s azok összességéből ered a hozzászólás kategóriájának meghatározása.

Social Support on an Online Forum for Depression and Anxiety

Abstract: Nowadays, online communities are typical sources of social support, which is a considerable help especially for those suffering from depression or anxiety. The aim of my research is to investigate the patterns of social support on an online depression and anxiety forum and to serve as an exploratory research of Natural Language Processing usage to classify comments into the categories of social support. The uniqueness of my research is the quantitative text analysis based on a complete qualitative analysis of the whole dataset. The conclusions of the qualitative analysis provide profound information for model definition, and for their evaluation. This knowledge is important for the investigation the potential of automatic text analysis in sociology. On average, four out of five comments are related to social support on the examined forum. Informational support appears in 59.9 percent of the supportive comments, while emotional support appears in 44.7 percent. The applied models' accuracies are nearly 80 percent, which means that they classified the vast majority of comments into the right category. The results show that there is a potential in building reliable models in order to classify the comments into the previously defined categories of social support.

Keywords: depression, social support, online forum, Natural Language Processing, NLP

Irodalom

- Aggarwal, C. C. (2018): *Machine Learning for Text* (1st ed.). Springer Publishing Company, Incorporated. Cham. <https://doi.org/10.1007/978-3-319-73531-3>
- Barak, A. – Boniel-Nissim, M. – Suler, J. (2008): Fostering empowerment in online support groups. *Computers in Human Behavior*, 24(5), 1867–1883. <https://doi.org/10.1016/j.chb.2008.02.004>
- Braithwaite, D. O. – Waldron, V. R. – Finn, J. (1999): Communication of Social Support in Computer-Mediated Groups for People With Disabilities. *Health Communication*, 11(2), 123–151. https://doi.org/10.1207/s15327027hc1102_2
- Brown, G. (2002): Social Roles, Context and Evolution in the Origins of Depression. *Journal of Health and Social Behavior*, 43, 255–276. <https://doi.org/10.2307/3090203>
- Busfield, J. (2000): Introduction: Rethinking the sociology of mental health. *Sociology of Health & Illness*, 22(5), 543–558. <https://doi.org/10.1111/1467-9566.00219>
- Comer, R. J. (2013): *Abnormal Psychology*. Worth Publishers. New York, New York
- Coulson, N. S. (2005): Receiving Social Support Online: An Analysis of a Computer-Mediated Support Group for Individuals Living with Irritable Bowel Syndrome. *CyberPsychology & Behavior*, 8(6), 580–584. <https://doi.org/10.1089/cpb.2005.8.580>
- Cutrona, C. E. – Shaffer, P. A. – Wesner, K. A. – Gardner, K. A. (2007): Optimally matching support and perceived spousal sensitivity. *Journal of Family Psychology*, 21(4), 754–758. <https://doi.org/10.1037/0893-3200.21.4.754>
- Cutrona, C. E. – Suhr, J. A. (1992): Controllability of stressful events and satisfaction with spouse support behaviors. *Communication Research*, 19(2), 154–174. <https://doi.org/10.1177/009365092019002002>
- Cutrona, C. E. – Suhr, J. A. (1994): Social support communication in the context of marriage: An analysis of couples' supportive interactions. In Burleson, B. R., Albrecht, T. L., &

- Sarason, I. G. *Communication of social support: Messages, interactions, relationships, and community*. (pp. 113–135). Sage Publications, Inc. Thousand Oaks, California
- Deetjen, U. – Powell, J. A. (2016): Informational and emotional elements in online support groups: A Bayesian approach to large-scale content analysis. *Journal of the American Medical Informatics Association*, 23(3), 508–513.
<https://doi.org/10.1093/jamia/ocv190>
- Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed.* (2013): American Psychiatric Publishing, Inc. <https://doi.org/10.1176/appi.books.9780890425596>
- Durkheim, E. (1951): *Suicide: A study in sociology* [1897]. Translated by JA Spaulding and G. Simpson. Glencoe, Illinois: The Free Press, 1951.
- Gáriépy, G. – Honkaniemi, H. – Quesnel-Vallée, A. (2016): Social support and protection from depression: Systematic review of current findings in Western countries. *British Journal of Psychiatry*, 209(4), 284–293. Cambridge Core.
<https://doi.org/10.1192/bjp.bp.115.169094>
- Grav, S. – Hellzèn, O. – Romild, U. – Stordal, E. (2012): Association between social support and depression in the general population: The HUNT study, a cross-sectional survey. *Journal of Clinical Nursing*, 21(1–2), 111–120.
<https://doi.org/10.1111/j.1365-2702.2011.03868.x>
- Hanley, T. – Prescott, J. – Gomez, K. U. (2019): A systematic review exploring how young people use online forums for support around mental health issues. *Journal of Mental Health*, 28(5), 566–576.
<https://doi.org/10.1080/09638237.2019.1630725>
- Hirschfeld, R. M. A. (2001): The Comorbidity of Major Depression and Anxiety Disorders: Recognition and Management in Primary Care. *Primary Care Companion to The Journal of Clinical Psychiatry*, 3(6), 244–254.
<https://doi.org/10.4088/pcc.v03n0609>
- Hosmer Jr., D. W. – Lemeshow, S. – Sturdivant, R. X. (2013): *Applied logistic regression* (Vol. 398). John Wiley & Sons. Hoboken, New Jersey
<https://doi.org/10.1002/9781118548387>
- House, J. S. – Umberson, D. – Landis, K. R. (1988): Structures and Processes of Social Support. *Annual Review of Sociology*, 14(1), 293–318.
<https://doi.org/10.1146/annurev.so.14.080188.001453>
- Huang, K.-Y. – Chengalur-Smith, I. – Ran, W. (2014): Not Just for Support: Companionship Activities in Healthcare Virtual Support Communities. *Communications of the Association for Information Systems*, 34, 561–594.
<https://doi.org/10.17705/1CAIS.03429>
- Kaplan, B. H. – Cassel, J. C. – Gore, S. (1977): Social support and health. *Medical Care*, 15(5), 47–58. <https://doi.org/10.1097/00005650-197705001-00006>
- Karmen, C. – Hsiung, R. C. – Wetter, T. (2015): Screening Internet forum participants for depression symptoms by assembling and enhancing multiple NLP methods. *Computer Methods and Programs in Biomedicine*, 120(1), 27–36.

- <https://doi.org/10.1016/j.cmpb.2015.03.008>
- Landis, J. R. – Koch, G. G. (1977): The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lazáry J. (2020): A COVID–19-pandémia akut hatásai a pszichés állapotra a járvány első szakaszában. *Neuropsychopharmacologica Hungarica*, 22(4), 172–177.
- Liddy, E. (2001): Natural Language Processing. In Kent, A., Lancour, H. (eds.) *Encyclopedia of Library and Information Science*, 2nd Ed. NY: Marcel Decker, Inc
- McCarthy, E. – Tarrier, N. – Gregg, L. (2002): The nature and timing of seasonal affective symptoms and the influence of self-esteem and social support: A longitudinal prospective study. *Psychological Medicine*, 32(8), 1425–1434. <https://doi.org/10.1017/s0033291702006621>
- Molnár, C. (2016): Szöveggé kódolás a gyakorlatban: Kézi, géppel támogatott és gépi megoldások. In Sebők M. (szerk.): *Kvantitatív szövegelemzés és szövegbányászat a politikatudományban*. Budapest, L'Harmattan, 24–36. https://poltextlab.tk.hu/uploads/files/Kvantitativ_szovegelemzes_keszpdf.pdf (utolsó megtekintés: 2021. április 20.)
- Németh, R. – Katona, E. R. – Kmetty, Z. (2020): Az automatizált szövegelemzés perspektívája a társadalomtudományokban. *Szociológiai Szemle*, 30(1), 44–62. https://szociologia.hu/dynamic/44_62_oldal.pdf (utolsó megtekintés: 2021. április 22.)
- Németh, R. – Koltai, J. (2020): The potential of automated text analytics in social knowledge building. In: Rudas T., Péli G. (eds.) *Pathways Between Social Science and Computational Social Science – Theories, Methods and Interpretations*. Springer. Cham
- Németh, R. – Sik, D. – Máté, F. (2020): Machine Learning of Concepts Hard Even for Humans: The Case of Online Depression Forums. *International Journal of Qualitative Methods*. <https://doi.org/10.1177/1609406920949338>
- Pikó, B. (2002): A depresszió társadalom-lélektana: Kórtünet vagy kortünet? *GYEMSZI – MOB. Valóság*, 55(7), 53–62.
- Prescott, J. – Hanley, T. – Ujhelyi, K. (2017): Peer Communication in Online Mental Health Forums for Young People: Directional and Nondirectional Support. *JMIR Mental Health*, 4(3), e29–e29. PubMed. <https://doi.org/10.2196/mental.6921>
- Rogers, A. – Pilgrim, D. (2005): *A sociology of mental health and illness* (5th ed.). Open University Press. Berkshire England, <https://doi.org/10.1002/cbm.413>
- Sabou, M. – Bontcheva, K. – Derczynski, L. – Scharl, A. (2014): Corpus Annotation through Crowdsourcing: Towards Best Practice Guidelines. *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14)*. http://www.lrec-conf.org/proceedings/lrec2014/pdf/497_Paper.pdf (utolsó megtekintés: 2021. április 23.)
- Sebők, M. – Kubik, B. – Molnár, C. – M. Balázs, Á. – Vancso, A. – Zorigt, B. – Zágoni, B. – Szabó, G. (2016): *Kvantitatív szövegelemzés és szövegbányászat a politikatudományban*. Budapest, L'Harmattan

- https://poltextlab.tk.hu/uploads/files/Kvantitativ_szovegelemzes_keszpdf.pdf
(utolsó megtekintés: 2021. április 20.)
- Selvin, H. C. (1958): Durkheim's suicide and problems of empirical research. *American Journal of Sociology*, 63(6), 607–619. <https://doi.org/10.1086/222356>
- Semmer, N. K. – Elfering, A. – Jacobshagen, N. – Perrot, T. – Beehr, T. A. – Boos, N. (2008): The emotional meaning of instrumental social support. *International Journal of Stress Management*, 15(3), 235–251. <https://doi.org/10.1037/1072-5245.15.3.235>
- Tan, P.-N. – Kumar, V. – Steinbach, M. (2012): *Bevezetés az adatbányászatba—Alapvetés*. Panem Kft. https://regi.tankonyvtar.hu/hu/tartalom/tamop425/0046_adatbanyaszat/adatok.html
(utolsó megtekintés: 2021. április 20.)
- Tikk, D. (szerk.). (2007): *Szövegbányászat*. Budapest, Typotex
- van Poppel, F. – Day, L. H. (1996): A Test of Durkheim's Theory of Suicide—Without Committing the 'Ecological Fallacy'. *American Sociological Review*, 61(3), 500–507. <https://doi.org/10.2307/2096361>
- Vayreda, A. – Antaki, C. (2009): Social Support and Unsolicited Advice in a Bipolar Disorder Online Forum. *Qualitative Health Research*, 19(7), 931–942. <https://doi.org/10.1177/1049732309338952>
- Walther, J. – Boyd, S. (2002): *Attraction to Computer-Mediated Social Support*. Cresskill, NJ: Hampton press.
- White, M. – Dorman, S. M. (2001): Receiving social support online: Implications for health education. *Health Education Research*, 16(6), 693–707. <https://doi.org/10.1093/her/16.6.693>
- World Health Organization (2017): *Depression and other common mental disorders: Global health estimates*. World Health Organization. Geneva <https://apps.who.int/iris/handle/10665/254610> (utolsó megtekintés: 2021. április 18.)

Függelék

1. táblázat: A 23 társas támogatási kategória együttes előfordulása

	Információ				Kézseif.		Érzelmi						Önbecsülés				Kapcsolatok		Össze- sen
	Tanács	Ajánlat	A helyz. át- érték.	Tanítás	Hajland.	Kapcs.	Fiz. tám.	Szimp.	Figy.	Meg- ért.	Bát.	Ima	Bók	Érvény.	Fel- enyh.	Jelen- lét	Társ.		
Információ	21	12	39	0	2	0	14	6	18	27	3	9	2	5	8	11	263		
	21	2	25	2	0	0	7	4	10	13	1	1	0	1	4	7	132		
A helyzet át- értékelése	12	2	4	0	0	0	5	1	4	11	0	4	0	3	1	3	103		
Tanítás	39	25	4	1	0	0	6	5	16	19	1	4	3	1	5	6	367		
Hajlandóság	0	2	0	1	0	0	1	0	1	1	0	0	0	0	2	0	6		
Kapcsolat	2	0	0	0	0	0	1	2	0	3	1	4	1	0	1	0	31		
Fizikai támog- atás	0	0	0	1	0	0	1	0	1	1	0	0	0	0	2	1	9		
Szimpátia	14	7	5	6	1	1	9	9	13	8	0	2	2	2	11	7	149		
Figyelem	6	4	1	5	0	2	0	9	5	4	1	1	2	1	3	0	109		
Megértés/ empátia	18	10	4	16	1	0	13	5	25	25	0	4	2	0	9	8	154		
Bátorítás	27	13	11	19	1	3	8	4	25	5	13	4	1	11	6	175			
Ima	3	1	0	1	0	1	0	0	0	5	2	0	0	2	0	18			
Bók	9	1	4	4	0	4	2	1	4	13	2	1	0	2	4	63			
Érvényesítés	2	0	0	3	0	1	0	2	2	4	0	1	0	0	1	0	26		
Felhasználó enyhítése	5	1	3	1	0	0	2	1	0	1	0	0	0	0	1	1	10		
Jelenlét	8	4	1	5	2	1	11	3	9	11	2	2	1	1	3	59			
Társaság	11	7	3	6	0	0	7	0	8	6	0	4	0	1	3	54			
Összesen	263	132	103	367	6	31	149	109	154	175	18	63	26	10	59	54	1275		

2. táblázat: A társas támogatást kérő és azt nyújtó hozzászólásokat tartalmazó adatokra épülő modellek teljesítménye

2.1. táblázat: A naiv Bayes osztályozó teljesítménye különböző módon feldolgozott szövegek esetén

	Pontosság	Precizitás	F1-score	Cohen-kappa
Stopszavazva	0,648	0,701	0,532	0,09
Stopszavazás nélkül	0,637	0,694	0,508	0,04
Stopszavazás nélkül, hosszabb szavakkal	0,639	0,704	0,512	0,05

2.2. táblázat: A logisztikus regresszió osztályozó teljesítménye különböző módon feldolgozott szövegek esetén

	Pontosság	Precizitás	F1-score	Cohen-kappa
Stopszavazva	0,709	0,675	0,655	0,32
Stopszavazás nélkül	0,732	0,687	0,684	0,38
Stopszavazás nélkül, hosszabb szavakkal	0,726	0,684	0,678	0,36

2.3. táblázat: Az SVM osztályozó teljesítménye különböző módon feldolgozott szövegek esetén

	Pontosság	Precizitás	F1-score	Cohen-kappa
Stopszavazva	0,688	0,685	0,607	0,22
Stopszavazás nélkül	0,710	0,681	0,648	0,30
Stopszavazás nélkül, hosszabb szavakkal	0,695	0,678	0,624	0,25