

KÖZÖSSÉGI MÉDIA POSZTOK SIKERESSÉGI TÉNYEZŐINEK ELEMZÉSE ADATTUDOMÁNYI MÓDSZEREKKEL

Bíró-Szigeti Szilvia, PhD

egyetemi docens

Budapesti Műszaki és Gazdaságtudományi Egyetem
szigetesi.szilvia@gtk.bme.hu

Vlaszov Artúr

matematikus MSc szakos hallgató

Budapesti Műszaki és Gazdaságtudományi Egyetem
vlaszova@math.bme.hu

Vastag Emese

matematikus MSc szakos hallgató

Budapesti Műszaki és Gazdaságtudományi Egyetem
evastag@math.bme.hu

Molontay Roland PhD

egyetemi docens

Budapesti Műszaki és Gazdaságtudományi Egyetem
molontay@math.bme.hu

Virringer Dominika

matematikus MSc szakos hallgató

Budapesti Műszaki és Gazdaságtudományi Egyetem
virdom@math.bme.hu

DOI: 10.15170/ktk-fogyszimp-2023-08

ABSZTRAKT

A kutatás a Budapesti Műszaki és Gazdaságtudományi Egyetem Természettudományi Karán futó Témalabor című tantárgy hallgatóival és a HSDSLab adattudományi kutatócsoport együttműködésével valósult meg, ami több szálon kötődik a BME Gazdaság- és Társadalomtudományi Karához is. A vizsgálat fókuszába a Twitter TOP 100 hírességének közösségi média posztjait helyeztük azzal a céllal, hogy adattudományi módszerekkel beazonosítsuk azokat a személyes kommunikációs eszközöket és stratégiákat, amelyek a legnépszerűbb személyek számára nagy számú kedvelést eredményeznek, valamint amelyek a követőikből leginkább aktivitást váltanak ki. A kutatásunk olyan tartalomelemzést tartalmaz, mint például topic modelling (témaelemzés), sentiment analysis (érzelemelemzés) és engagement metrics (elkötelezettségi mutatók) vizsgálata. A tartalomelemzéssel és a kulcsfontosságú teljesítménymutatók vizsgálatával – például a kedvelések, megosztások és hozzászólások nyomon követésével – a kutatás betekintést kínál azokba a tényezőkbé, amelyek hozzájárulnak a hírességek által alkalmazott sikeres személyes márkaépítési tevékenységhez, a közösségi média kampányaik optimalizálásához és a közönségük elkötelezettségének növeléséhez. A hírességek online követői heterogén személyekből álló csoportok. Az eredményekből igyekszünk levonni arra is következtetéseket, hogy a legnépszerűbb és a legsikeresebb elemek (kampányok, hírek, közlemények) kommunikációs összetevőinek hatása, azok fogyasztásának aránya és időzítése vajon mit mutat meg abból, hogy mi jellemző a közösségi médiában a követők médiafogyasztására.

Kulcsszavak: adattudomány, közösségi média, médiafogyasztás, kommunikáció, viselkedésemelés

BEVEZETÉS

A hírességek között élesedő versenyben a közösségi média vette át a tevékenységüket hitelesítő szerepet, hiszen az új társadalmi gondolkodásmódra az a jellemző, hogy ami nem kerül a közösségi médián kipoztolásra, az nem is történik meg. Azonban nagyon nem mindegy, mit hogyan posztol a közösségi média embere, mert az őt követők reakciójának mennyisége és időzítése visszajelzést ad számára arról, vajon megfelelően végezte-e a munkáját vagy sem, elegendően észrevették és lereagálták-e a tevékenységét vagy sem, valamint arról is tájékozódhat, mi jellemző a követői közösségi médiafogyasztására.

A közösségi médiában a követő igen tágan értelmezhető fogalom, hiszen itt mind a drukkerék és az ellenfelek is megjelennek: gratulálnak vagy szitkozódnak, örülnek vagy háborognak, de mindenképpen kinyilvánítanak érzelmeiket. Követett is bárki lehet: dolgos vagy dologtalan (Veblen, 1975), értő vagy kontár, támogató vagy kritikus.

A mérés éppen ezért nem egyszerű, de az alábbiakban bemutatásra kerülő kutatással igyekszünk nagyobb rálátást nyerni erre a világra. Kutatásunk céljaként tűztük ki, hogy beazonosítsuk azokat a közösségi médiában hasznosítható, a posztolás eredményességét befolyásoló sikertényezőket, amelyeket a magánszemélyek használnak a weboldalon a közösségi médiában való megjelenésükhöz. A kutatásunkban a Twitter közösségi oldal 100 legnépszerűbb személyeinek nyilvánosan hozzáférhető posztjait töltöttük le 2018.01.01. és 2023.03.23. között. Ebből az ötéves időkeretből körülbelül 160.000 tweetet sikerült begyűjtenünk.

A kutatásunk ezen hírességek közösségi média platformokon alkalmazott személyes márkaépítési tevékenység sikerességi tényezőinek azonosítására összpontosított. A tartalomelemzéssel és a kulcsfontosságú teljesítménymutatók vizsgálatával, mint például a kedvelések, a megosztások és a hozzászólások nyomon követésével betekintést nyertünk azokba a tényezőkbé, amelyek hozzájárulnak a hírességek sikeres márkaépítési tevékenységéhez, a közösségi média kampányaik optimalizálásához és a közönségük elkötelezettségének növeléséhez. A kutatásunk olyan tartalomelemzéseket tartalmaz, mint például topic modelling (témamodellőzés), sentiment analysis (érzelemelemzés) és engagement metrics (elkötelezettségi mutatók) vizsgálata.

KUTATÁS

Az eredmények alkalmasak a legnépszerűbb és legsikeresebb elemek (kampányok, hírek, közlemények) marketing-kommunikációs összetevőinek, azok arányának és időzítésének azonosítására. A következőkben a kutatási eredményeinket mutatjuk be és keressük a választ arra, hogy mi jellemző a közösségi médiában a követők médiafogyasztására.

A közösségi média növekvő népszerűségével Saravanakumar és SuganthaLakshmi (2012) meghatározták azokat az elemeket, amelyeknek szerepelniük kell a magánszemélyek és kisvállalkozók számára készült közösségi média marketingjének keretrendszerében. Jenders és társai (2013) azt vizsgálták, hogyan válnak virálissá a tweetek, azaz hányan retweetelik őket. Általánosított lineáris modellel vizsgálták, hogy bizonyos faktorok – például a tweetek érzelmi tartalma - hogyan befolyásolják egy tweet virálissá válását. Mahdikhani (2022) munkájában a Covid-19 járvány különböző szakaszaiból származó tweeteket elemezte, ahol arra a következtetésre jutott, hogy az érzelmes tweetek hajlamosabbak népszerűvé válni. Egger és Yu tanulmánya (2022) négy témamodellező technika teljesítményét értékeli a Twitter-adatokon keresztül, de Instagram oldalak színelemzéssel is foglalkoztak (Yu & Egger, 2021). Az idézett szerzők munkái alapján úgy döntöttünk, hogy hasonló adattudományi módszerekkel mi is elemezzük a Twitter posztokat. Elsősorban olyan hírességekre kívántunk koncentrálni, akik a saját személyiségük és tevékenységük népszerűsítésére használják a Twitter közösségi oldalát.

2023. márciusi adatok alapján a Twitter 100 legtöbbet követett személyét választottuk ki. Mivel a Twitter nem biztosít hivatalos listát a legtöbbet követett fiókokról, ezt a listát különböző forrásokból (Twitter, 2023; Socialblade, 2023; Viralpitch, 2023) szereztük be. A legnépszerűbb fiókok listájának összeállításakor kizártuk a nem egyéni fiókokat, a főként nem angol nyelvű közönséggel rendelkezőket, valamint néhány olyan fiókot, amely különböző okokból az elmúlt években nem posztolt. A végleges listánk, amely 100 Twitter-fiókot tartalmaz, megtalálható a GitHub tárhelyünkön (GitHub tárhely, 2023). A TOP100 híresség fiókjának összes posztját a Python snsrape csomagjának segítségével töltöttük le.

A vizsgált 100 személy – a foglalkozásuk alapján – valamilyen tevékenységet folytató embereknek számítanak, de a társadalmi megítélés itt változó, hogy az dolgozó vagy dologtalan kategóriába tartozik-e (Csepeli, 2023). A 100 legnépszerűbb személy foglalkozásáról közzétett adatok vizsgálata alapján elmondható, hogy a legnagyobb arányban énekesek (39%) vannak jelen, őket színészek (14%), sportolók (11%), rapperek (7%), politikusok (6%) és komikusok (6%) követnek, majd a televíziós személyiségek (5%) következnek, végül a fennmaradó néhány százalékon média személyiségek (3%), vállalkozók (2%), DJ-k, YouTuber-ek (2%), modellek (1%) és írók (1%) osztozkodnak.

ELKÖTELEZETTSÉGI MUTATÓK ELEMZÉSÉNEK EREDMÉNYEI

A közösségi médiában a követők médiafogyasztásának vizsgálatához az elkötelezettségi mutatókat elemeztük, ami a közönség interakciójának tartalommal kapcsolatos mérését jelenti, például a kedvelések, a megosztások és a hozzászólások számát. Az eredmények rámutatnak arra, hogy a tartalom mennyire vonzó és hatékony a közönség számára. Először a kedvelések, a válaszok és a retweetek száma szerint normalizáltuk azt a követők számával.

Az eredeti posztokból létrehoztunk egy tisztított verziót az emoji-k és egyéb speciális karakterek és linkek nélkül. Az adatok 48,5%-a nem tartalmazott semmilyen típusú médiát, 28,8%-a képeket tartalmazott, 22,7%-a videókat tartalmazott, és csak 0,1%-a tartalmazott mind képeket és videókat. Egy adott felhasználó tweetjeit figyelembe véve azt a tweetet jelöltük meg népszerűnek, ha a kedvelések száma alapján a legjobb 5% közé tartozott, így végül a tweetek 5,3%-a népszerűnek számított. Mivel a tweetek hosszának maximális karakterkorlátja van, így nincsenek rendkívül hosszúak, a tweetek többsége jellemzően rövid. A népszerű tweetek között felülreprezentáltak a nagyon rövid és nagyon hosszú posztok. Feltételezésünk szerint a leghosszabb tweetek olyan fontos tartalommal rendelkeznek, amelyek hasznosak a

követők számára, míg a rövidebb tweetek könnyű és gyors információtartalmuk miatt lehetnek népszerűek.

Manapság a közösségi média posztok egy része tartalmaz emojiakat (28%) is, mivel ezek segítenek az üzenet érzelmi jellegének, esetleges átvitt értelmének közvetítésben. Az eredményekből kitűnik az, hogy egy tweet tartalmaz-e emojiakat vagy sem, annak nincs jelentős korrelációja a népszerűséggel.

A posztok többsége nem tartalmaz médiát. Ha összehasonlítjuk a képeket tartalmazó tweetek számát a videókat tartalmazó tweetek számával, mind az összes tweetet és a népszerű tweeteket figyelembe véve, a képeket tartalmazó tweetek aránya magasabb a népszerű tweetek között. Ez arra utal, hogy a felhasználók jobban szeretik a képeket tartalmazó posztokat, mint a videókat tartalmazó posztokat. A kedvelések átlagos számát illetően az összes tweetnél a csak képeket tartalmazó tweetek kapják átlagosan a legtöbb kedvelést, majd a médiát nem tartalmazó posztok következnek, és a videók kapják a legkevesebb kedvelést. Érdekes, hogy míg hétköznapokon posztolt tweetekben a fotók népszerűbbek a videóknál, a hétvégi posztokban ez a sorrend felcserélődik. Ez az eredmény rámutathat arra, hogy a követőknek több szabadidejük van hétvégén és emiatt több videót néznek meg, valamint így több kedvelést adnak a videóknak, mint a fotóknak.

A tweetek számának különbségét vizsgáltuk a hétköznapok és a hétvégék között. Az eredményekből arra lehet következtetni, hogy a hírességek, illetve a Twitter fiókjukat kezelő ügynökségek hajlamosabbak hétköznapokon posztolni. Az évek során egy enyhe csökkenő trend rajzolódik ki a tweetelésre fordított idő csökkenésében. Arra is kíváncsiak voltunk, hogy a hírességek mennyi posztot tesznek közzé óránként a hét különböző napjain, valamint a kedvelések átlagos száma miként alakul. A legnépszerűbb posztok azok lesznek, amelyek napközben 10.00 és 15.00 óra között kerülnek ki, csúcsidőnek pedig az ebédidőt tekinthetjük.

ÉRZELMI ELEMZÉS EREDMÉNYEI

Az érzelmi elemzés értékes betekintést nyújt a tweetekben kifejezett érzelmi hangulatba és véleményekbe. Az érzelmelemzés egy olyan módszer, amely a szöveges adatokban található érzelmeket és hangulatokat próbálja megérteni és osztályozni. A tweetek érzelmi mérésével, legyen az pozitív, negatív vagy semleges, megérthetjük, hogyan érzékelik és reagálnak a felhasználók a tartalomra. Ez az információ segít meghatározni a tweet népszerűségét befolyásoló tényezőket, trendek azonosítását, sőt még a virális vagy befolyásos tweetek előrejelzését is, lehetővé téve így a Twitter közönség jobb megértését és bevonását. Mindehhez a VADER (Valence Aware Dictionary and Sentiment Reasoner) (Hutto&Gilbert, 2014) eszközt alkalmaztuk, amelyet kifejezetten a közösségi médiában kifejezett érzelmek észlelésére terveztek. A VADER négy kategóriába képes besorolni az üzeneteket: pozitív, negatív, semleges és összetett.

Az évek során különböző érzelmekkel rendelkező tweetek többsége pozitív érzelmű volt, és míg a tweetek száma évről évre csökken, de a pozitív, semleges és negatív érzelmű tweetek aránya szinte változatlan maradt.

A bejegyzések számát havi bontásban is megvizsgáltuk, felfedezhető egy periodicitás az adatokban, ami azt jelenti, hogy a különböző hónapokban közzétett tweetek száma nem egyenletes. Vannak olyan hónapok, amikor több tweetet tesznek közzé: március, április, május, október és november. Másrészt vannak olyan hónapok, mint január, február, július, augusztus és szeptember, amelyek jelentősen kevesebb bejegyzést tartalmaznak.

A bejegyzések szófelhője alkalmas arra, hogy vizuális módon reprezentáljuk az érzelmeket közvetítő leggyakrabban használt szavakat vagy kifejezéseket. Ez a vizualizáció segít azonosítani a felhasználókkal rezonáló kulcsfontosságú témákat vagy fogalmakat, amelyek segítik megérteni, hogy mely aspektusok támogatják a közösségi média bejegyzések népszerűségét és elkötelezettségét. Fontos kiemelni, hogy ezek a szófelhők nem csak pozitív és negatív érzelmű szavakat tartalmaznak, hanem pozitív és negatív bejegyzésekből generálódnak. Tehát ha egy szó (függetlenül az érzelmi tartalmától) sok pozitív érzelmi tweetben szerepel, de sok negatív érzelmi tweetben is, akkor mindkét szófelhőben megjelenhet. Ezért fordul elő néhány szó mindkét helyen, például: “will”, “today”, “now”. A pozitív szófelhő legfontosabb szavai a “thank” és a “love”, a negatívon: “will”, “people”, “now”, “need” és “one”, ahol még néhány politikai kifejezés is található. Az átlagos kedvelések számát tekintve a semleges bejegyzések kapják a legtöbb kedvelést, a negatív bejegyzések a második legtöbb kedvelést kapja, és meglepő módon a pozitív bejegyzések átlagosan a legkevesebb kedvelést kapják.

TÉMAELEMZÉS EREDMÉNYEI

A témaelemzés egy olyan statisztikai modellezési technika, amelyet szöveges dokumentumok témáinak azonosítására és csoportosítására használnak. A témaelemzési technika azon feltételezésen alapul, hogy egy dokumentum különböző témákból áll össze, és minden témát olyan szavak jellemeznek, amelyek valószínűleg együtt jelennek meg. E technika segítségével minden tweet szövegéből összegyűjtöttük a legrelevánsabb szavakat, és ezek alapján megpróbáltunk olyan témanévvel találni, ami a legjobban összefoglalja ezeknek a szavaknak a jelentését. Ez alapján megvizsgáltuk, hogy vannak-e trendi témák a népszerű tweetek között.

Az eredményül kapott témáink: nemzeti identitás, ünneplés, hála, TV műsor, családi események, média csatornák, online vásárlás, pozitív tapasztalat, egyéb. Összességében ezek a témák viszonylag pontos leírást adnak a tweetekben rejlő alapvető tartalomról. Annak ellenére, hogy a „nemzeti identitás” volt a leggyakoribb téma (a tweetek 43%-a), ez csak 4-7%-ban vált népszerűvé. Az „ünneplés” téma esetében a népszerűség majdnem minden évben 7-13% között mozgott, és úgy tűnik, hogy ebben a témában közzétett tweeteknek sokkal nagyobb esélyük volt a népszerűsége. Ehhez hasonlóan a „média csatornák” téma tweetjei is nagyobb eséllyel váltak népszerűvé.

ÖSSZEGZÉS

A tanulmányunkban az elmúlt öt év 100 kiemelkedő híresség Twitter adatainak elemzését végeztük el. Célunk az volt, hogy azonosítsuk azokat a kommunikációs mintákat és stratégiákat, amelyek nagy számú kedvelést eredményeznek. Ennek eléréséhez kifejlesztettünk egy népszerűségi mutatót, és összehasonlítottuk a népszerű tweeteket a teljes adathalmazzal. Megvizsgáltuk a tweetek különböző aspektusait, beleértve a csatolt média típusát, az emoji jelenlétét és a tweetek hosszát. Emellett érzelem elemzést is végeztünk a VADER segítségével, hogy megértsük a tweetekben közvetített alapvető érzelmeket. A nem-negatív mátrix faktorizáció segítségével kilenc különböző témát fedeztünk fel az adatokban. Érdekes, hogy a legnépszerűbbek az ünneplés témájú tweetek voltak, amelyek hatalmas közönséget vonzottak és jelentős elkötelezettséget eredményeztek. A kommunikációs minták vizsgálatával, egy népszerűségi mutató alkalmazásával, a tweet attribútumainak elemzésével, érzelem elemzés végzésével és témák elemzésével értékes betekintést nyertünk a tweet népszerűségének dinamikájába.

VITAKÉRDÉSEK

1. Ha a személyek közötti versenyre tekintünk, akkor a közösségi média az egyéni megjelenés széles tárházát képes nyújtani, a kutatásunkban megjelenő eredmények pedig az egyént hozzá tudják segíteni a nagyobb láthatóságot és népszerűséget biztosító közösségi média jelenléthez. Vajon milyen tényezők jelezhetik azt, hogy a vizsgált személy közösségi médiában közzétett „dolgos” tevékenysége megfelelő és hiteles? Az egyén kultúrában domináns értékrend (Covey, 2016) mérése elegendő-e?
2. Ha rátekinünk a kutatott témákra távolabbról, akkor azt látjuk, hogy a népszerű személyek közötti verseny színtere napjainkban a virtuális tér, a verseny fokmérője pedig a követők és reakcióinak száma. A legnépszerűbb témák leginkább olyan hétköznapi tevékenységekhez és eseményekhez kapcsolhatók, amelyek az emberek mindennapjait teszik ki. A követők növekvő száma, valamint a követett témák rámutatnak-e arra, hogy az emberek a virtuális világba vezetik át a hétköznapi tevékenységeiket is, és egyre több időt töltenek ott, azaz távolodnak a reális élettől?

HIVATKOZÁSOK

- Covey, S. R. (2016), *Primary Greatness: The 12 Levels of Success*. Simon and Schuster, New York
- Csepeli, Gy. (2023), *Értékek ébresztése*. Kocsis Kiadó, Budapest
- Egger, R. & Yu, J. (2022), „A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts”, *Frontiers in Sociology*, 7, 7, 886498. <http://doi.org/10.3389/fsoc.2022.886498>.
- Github tárhely (2023), <https://github.com/v-mse/Identifying-brand-strategy-success-on-social-media>. Saját tárhely.
- Hutto, C. J. & Gilbert, E. (2014), VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Conference: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*. At: Ann Arbor, MI
- Jenders, M., Kasneci, G. & Naumann, F. (2013), Analyzing and predicting viral tweets. *WWW 2013 Companion - Proceedings of the 22nd International Conference on World Wide Web*, 657-664, <http://doi.org/10.1145/2487788.2488017>
- Mahdikhani, M. (2022), „Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of Covid-19 pandemic”, *International Journal of Information Management Data Insights*. 2 (1), 100053, <http://doi.org/10.1016/j.ijime.2021.100053>
- Saravanakumar, M. & SuganthaLakshmi, T. (2012), „Social Media Marketing”, *Life Science Journal*, 9 (4), 4444-4451.
- Socialblade (2023), <https://socialblade.com/twitter/top/100>
- Twitter (2023), <https://twitter.com/>
- Veblen, T. (1975), *A dologtalan osztály elmélete*. KJK, Budapest
- Viralpitch (2023), <https://viralpitch.co/topinfluencers/twitter/top-200-twitter-influencers/>
- Yu, J., & Egger, R. (2021), „Color and engagement in touristic Instagram pictures: A machine learning approach”, *Annals of Tourism Research*, 89 (2), 103204. <http://doi.org/10.1016/j.annals.2021.103204>