

Cikkismertetés: Az egészségműveltség osztályozása mesterséges intelligencia használatával

Article review: Classifying health literacy with machine learning

Ismertető: Osváth Máttyás ✉
Debreceni Egyetem, Általános Orvostudományi Kar, Egészségtudományok Doktori Iskola, Debrecen

Ismertetett cikk: Balyan R, Crossley SA, Brown W, et al. Using natural language processing and machine learning to classify health literacy from secure messages: The ECLIPPSE study. PLoS ONE, 2019;14(2):1–17. doi: 10.1371/journal.pone.0212488

Beküldve: 2020. 09. 03.
doi: 10.24365/ef.v61i4.632

Kulcsszavak: egészségműveltség; NLP; természetes nyelvfeldolgozás; gépi tanulás; nyelvelemzés

Keywords: health literacy; NLP; natural language processing; machine learning; linguistic analysis

ELMÉLETI ÁTTEKINTÉS

A krónikus betegségek (pl. cukorbetegség) kezelése összetett feladat, és gyakori kommunikációt igényel a páciens és az egészségügyi dolgozók között.

A hiányos egészségműveltség csökkenti az ellátás eredményességét, ami növeli az állapotromlás, a szövődmények és a halálozás esélyét. A nem megfelelő kommunikáció és csökkent terápiahűség is hozzájárulhat – az egyébként elkerülhető – szenvedéshez és az egészségügyi költségek növekedéséhez.

A betegeknek szóló online portálok, a köznyelven betegportálok olyan biztonságos weboldalak, amelyek a páciensek és az egészségügyi dolgozók felvehetik egymással a kapcsolatot, üzenet válthatnak, valamint alkalmasak személyes egész-

ségügyi információk (pl. kórházi látogatások időpontjainak, receptek, számlák és teszteredmények) tárolására.

Az elektronikus egészségügyi információkkal rendelkező, online betegportálok elősegítik a kommunikációt a páciens és az egészségügyi személyzet közötti üzenetváltások által, támogatást nyújtva a kórházi kezelések közötti időszakban is. A betegportálok használata javíthatja a terápiahűséget, az orvos-beteg kommunikációt és az ellátás igénybevételét.

Az online kommunikációra a beteg egészségműveltségének mértéke hatással van. Az utóbbi 5 évben a hiányos egészségműveltséggel rendelkező páciensek között 65%-os növekedés volt megfigyelhető a betegportálok használatában, összehasonlítva a megfelelő egészségműveltségű páciensekkel, ahol a növekedés 41% volt.

Az egészségműveltség szintjét mérő jelenlegi kérdőívek sokszor önbevalláson alapulnak, és időigényesek lehetnek. Azonban a korábbi kutatások által alátámasztott természetes nyelvfeldolgozás (*natural language processing*, továbbiakban NLP) - a számítógépes nyelvészetet és gépi tanulást felhasználva – orvosi szakterületen is bizonyítottan alkalmazható és hatékony eszköz lehet, viszont egészségműveltség mérésére még nem alkalmazták.

Az interjúk, kérdőívek használata helyett az automatizált szövegfeldolgozás hatékony lehet a különböző egészségműveltséggel rendelkező személyek azonosításában, a személyre szabott intervenció és kommunikáció kialakításában nagyobb populációk esetén is.

A kutatás célja a cukorbetegség egészségműveltségük alapján hiányos és megfelelő kategóriákba sorolása számítógépes NLP segítségével, és a besorolás összehasonlítása a kérdőívvel felmért egészségműveltséggel, illetve összevetése a betegek szocio-demográfiai jellemzőivel és kezelési eredményeivel. Amennyiben sikerül az automatizálható NLP-vel azonosítani a hiányos egészségműveltségű betegeket, akkor a kezelőorvosnak való visszajelzés javíthatja a kommunikációt, és ezáltal az ellátás eredményességét.

MÓDSZER

Az adatokat a Kaiser Permanente Northern California (KPNC) Diabetes Registry adatbázisa szolgáltatja, ami több mint 150 000 etnikailag különböző páciens és 9000 orvos 1 millió üzenetét tartalmazza. A páciensek egy csoportja ($n = 14\,357$) egy korábbi kutatás során kitöltött, önbevalláson alapuló, egészségműveltséget mérő kérdőívet, aminek változóit ez a kutatás is felhasználta. Az üzenetek közül ($n = 1\,050\,577$) a nem háziorvosnak és a nem angol nyelven írt üzeneteket kizárták.

Az üzenetek száma 2-től 205-ig (átlagosan 39,9 üzenet), valamint az egy beteg által írt szavak száma 1-től 16 469-ig terjedt (átlagosan 2 059 szó), azonban az 50 szónál kisebb terjedelmű üzeneteket eltávolították. Az elemzésre felhasznált, megtisztított vizsgálati anyag 6 941 páciensből és 283 216 üzenetből állt.

A kutatásban alkalmazott NLP eszközöket olyan nyelvi jellegzetességek mérésére fejlesztették ki, ami kapcsolatba hozható az egészségműveltséggel (pl. szövegkomplexitás, -kohézió, szintaktikai összetettség, akadémiai és orvosi szavak, szóismétlés).

A szocio-demográfiai változók (pl. kor, nem) és kezelési eredmények (kórházi felvétel, elégtelen gyógyszeradagolás, hipoglikémia) mellett 185 nyelvi jellemzőt nyertek ki az egészségműveltségi profil kialakítására.

A nyelvi változók közül kiemelhető a lexikális diverzitás (a szavak változatossága), valamint határozószavak, melléknevek és funkciószavak (pl. személyes névmás) gyakorisága, amelyek használata magasabb egészségműveltségű személyekre jellemzőbb.

Az egészségműveltséget mérő kérdőív állításaiból három, diszkrét (hiányos és megfelelő) függő változót alkottak, amiket a profilok besorolására használtak fel.

Az elemzést felügyelt gépi tanuláson alapuló modellekkel végezték a 185 nyelvi változón és a diszkrét változóvá alakított, egészségműveltséget mérő indexeken. Az osztályozást a pontosság, érzékenység, specificitás, C-statisztika (ROC-görbe), pozitív és negatív prediktív érték jóságai mutatókkal mérték.

A nyelvi jellemzők alapján kategóriába sorolt egészségműveltséget összehasonlították az önbevalláson alapuló egészségműveltséggel, szocio-demográfiai státusszal és kezelési eredményekkel.

EREDMÉNYEK

A hiányos és megfelelő egészségműveltségi profilok kategorizálása során a jóságai mutatók megfelelőek voltak (szenzitivitás = 75,49; C-statisztika = 0,74).

A szövegjellemzők vizsgálata során a hiányos egészségműveltségű csoportba sorolt páciensekre a kisebb szószám és lexikai diverzitás (pl. több szóismétlés), valamint a kevésbé szofisztikált kifejezőmód (pl. konkrétság) volt jellemző; továbbá több negatív és kevesebb pozitív érzelmet kifejező szót

használtak, és gyakrabban fogalmaztak harmadik személyben, mint első személyben.

A hiányos egészségműveltségű kategóriába sorolt személyek életkora átlagosan 1-3 évvel volt több a mintában, 70,8-76,1% volt nem-fehér, és 84,7-88,7% rendelkezett középfokú végzettséggel, összehasonlítva a megfelelő egészségműveltségű csoportba sorolt személyekkel, akik között 59,9-63,5% volt nem-fehér, és 93,4-95% rendelkezett középfokú végzettséggel. Ezek az eredmények hasonlóak a kérdőíves módszerrel nyertekhez.

A hiányos egészségműveltségűek osztályába sorolt személyek alacsonyabb terápiahűséggel ($p < 0,047$) és magasabb hipoglikémiás szinttel ($p < 0,001$), valamint jelentősen több komorbid kondícióval rendelkeztek a megfelelő egészségműveltségű csoporthoz viszonyítva.

Az évenkénti járó- és fekvőbeteg, valamint a sürgősségi ellátás igénybevételének száma is többször fordult elő ($p < 0,001$) a hiányos egészségműveltségűként kategorizált betegek körében.

DISZKUSSZIÓ

A korábbi kutatásokkal összhangban a modell a páciensek egyharmadát hiányos egészségműveltségűnek osztályozta.

TANULSÁGOK A HAZAI SZAKEMBEREK SZÁMÁRA

A hiányos egészségműveltség akadály a páciens és az ellátó közötti kommunikációban, és az optimális kezelési eredmények elérésében. Ennek elkerülésére az egészségügyi rendszerekben az előrejelző modellek alkalmazása növekszik, amely révén javulás érhető el populációs és individuális szinten, javítva a kommunikációt és intervenciót. A páciensek üzeneteiből kinyert szövegjellemzők alkalmasak lehetnek a különböző egészségműveltséggel rendelkező személyek megkülönböztetésére, ezáltal feltehetően kevesebb kérdőív és felmérés kitöltésére lesz szükség, vagyis korszerűbb módszerrel az időigényes és költséges adatgyűjtés jelentős részben kiváltható.

A különböző nyelvi jellemzők (pl. kevésbé szofisztikált kifejezések, kevesebb pozitív szó) és gépi tanulási modellek elfogadható szinten jósolták be az önbeműveltségen alapuló egészségműveltséget, amelyre építve kialakítható az eltérő egészségműveltségű személyek megkülönböztetésére alkalmas nyelvi profil.

A kutatás egyik korlátja, hogy csak azokat a pácienseket vizsgálták, akik a háziorvosuknak küldtek üzenetet, azonban a betegek egyharmadára nem jellemző az üzenetváltás általi kapcsolatfelvétel.

Az egészségműveltség mérése kihívást jelenthet egyéni és populációs szinten, ezért egy automatizált szövegfeldolgozási módszer a nagy valószínűséggel hiányos egészségműveltséggel rendelkező személyek azonosításával hozzájárulhat az orvosi kommunikáció javításához, a krónikus betegségek szövődésének, előfordulásának és halálzásának csökkentéséhez.

Az orvos ezáltal figyelmeztetést kaphat arra, hogy adott betegnek nehézséget jelenthet a szóbeli és/vagy írásos tájékoztatások értelmezése, azaz az ilyen páciensek utánkövetésére nagyobb szükség lehet az orvosi tanácsok és a gyógyszeres kezelések betartása érdekében.

Ez az egyik legelső vizsgálat a nagy valószínűséggel hiányos egészségműveltséggel rendelkező és a kezelési eredmények szempontjából kockázatos személyek NLP általi azonosítására.