



Identifying Psychological Disorders Based on Data in Virtual Environments Using Machine Learning

Maryam Abourehani Mohammadi¹, Mahshad Fadaei Moghadam Heydarabadi², Solmaz Zardary³, Soomaayeh Heysieattalab^{4*}

¹ Department of Psychology, Faculty of Education and Psychology, University of Tabriz, Tabriz, Iran

² Department of Psychology, Faculty of Education and Psychology, University of Tabriz, Tabriz, Iran

³ Department of Knowledge and Information Science, Faculty of Education and Psychology, University of Tabriz, Tabriz, Iran.

^{4*} Division of Cognitive Neuroscience, University of Tabriz, Tabriz, Iran. heysieattalab@gmail.com

Citation: Abourehani Mohammadi M, Fadaei Moghadam Heydarabadi M, Zardary S, Heysieattalab S. Identification Psychological Disorders Based on Data in Virtual Environments Using Machine Learning. *Journal of Cognitive Psychology*. 2020; 7 (4): 1-12. [Persian].

Keywords

Computational Psychology, Machine Learning, Social Media, Psychological Disorders

Abstract

Recently, research has been conducted on the use of social networks as a new platform for identifying people with mental disorders. In addition, because of the complexity of diagnosing psychological diseases using conventional methods, the use of machine learning for identifying these psychological diseases is increasing. The goal of this article was to systematically review the research conducted using social media data for predicting and diagnosing psychological disorders with the help of machine learning. Based on systematic review on the Prisma method, the aim of this article was achieved through searching the main keywords of diagnosis and the prediction of mental disorders combined with machine learning and social media data without considering the dates of their publications. Depression had the highest frequency among the final 20 selected articles with a predictive power of 42% and 87%, the lowest and the highest respectively. On the other hand, only 30% of studies used questionnaires for gathering data on social media and the most common approach for data collection was public posts on social media by the use of regular expressions. Twitter has also been used as the largest source of data collection in these sorts of studies. It seems that computational psychology based on machine learning methods could help to identify disorders at an appropriate time and select more effective treatments for mental disorders among the users of social media.

شناسایی اختلال‌های روان‌شناختی بر اساس داده‌های موجود در محیط‌های مجازی با استفاده از یادگیری ماشین

مریم ابوریحانی محمدی^۱، مهشاد فدایی مقدم حیدرآبادی^۲، سولماز زرداری^۳، سمیه حیثیت طلب^۴

۱. گروه علوم تربیتی و روانشناسی، دانشگاه تبریز، تبریز، ایران.

۲. گروه علوم تربیتی و روانشناسی، دانشگاه تبریز، تبریز، ایران

۳. گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روانشناسی، دانشگاه تبریز، تبریز

۴. نویسنده مسئول) گروه علوم اعصاب‌شناختی، دانشکده علوم تربیتی و روانشناسی، دانشگاه تبریز، تبریز.

heysieattalab@gmail.com

چکیده

اخیراً پژوهش‌هایی در راستای به‌کارگیری شبکه‌های اجتماعی به‌عنوان بستری جدید برای شناسایی افراد با اختلال‌های روانی صورت گرفته است. علاوه بر این به دلیل پیچیدگی تشخیص بیماری‌های روان‌شناختی با استفاده از روش‌های معمول، استفاده از یادگیری ماشین برای شناسایی این افراد رو به افزایش است. هدف این مقاله مرور پژوهش‌های انجام شده با استفاده از داده‌های شبکه‌های اجتماعی در پیش‌بینی و تشخیص اختلال‌های روان‌شناختی به کمک یادگیری ماشین است. در این پژوهش به روش مرور نظام‌مند مبتنی بر پریسما، از طریق جستجوی کلید واژه‌های اصلی تشخیص و پیش‌بینی اختلال‌های روان‌شناختی در ترکیب با واژه‌های یادگیری ماشین و فضای مجازی در پایگاه‌های اطلاعاتی تخصصی بدون در نظر گرفتن سال انتشار آن‌ها، یافته‌ها و اطلاعات مورد نظر جهت دستیابی به هدف پژوهش مورد واکاوی قرار گرفت. اختلال روان‌شناختی افسردگی در بین ۲۰ مقاله نهایی انتخاب شده بیشترین فراوانی با قدرت پیش‌بینی ۴۲ و ۸۷ درصد به ترتیب کمترین و بیشترین توان پیش‌بینی افسردگی را داشته است. از طرفی برای جمع‌آوری داده تنها ۳۰ درصد مطالعات از پرسشنامه فضای مجازی و بیشترین رویکرد، از پست‌های عمومی در شبکه‌های اجتماعی با عبارات منظم و نیز توییت‌ها به‌عنوان بیشترین منبع استفاده شده است. به نظر می‌رسد روانشناسی محاسباتی مبتنی بر روش‌های یادگیری ماشین بتواند به شناسایی و پیش‌بینی دقیق‌تر اختلال‌های روان‌شناختی کاربران فضاهای مجازی کمک نماید.

تاریخ دریافت

۱۳۹۸/۱۱/۲۲

تاریخ پذیرش نهایی

۱۳۹۹/۳/۱۴

واژگان کلیدی

روانشناسی محاسباتی،
یادگیری ماشین،
شبکه‌های اجتماعی،
اختلال‌های روان‌شناختی

مقدمه

بر روی نمونه‌های مشخصی از جمعیت تشخیص داده می‌شود. روش‌های غربال‌گری روانی، بین افرادی از جمعیت که احتمال دارد یک اختلال عاطفی یا روانی مهم داشته باشند و آن‌هایی که چنین اختلالی ندارند تمایز می‌گذارد و افراد مشکوک یا با یافته‌های مثبت را تحت آزمایش‌ها و معاینه‌های فشرده‌تر تشخیصی در مرحله دوم قرار می‌دهد (یعقوبی، ۱۳۸۷). با توجه به هزینه، زمان‌بر بودن و مشکلات دیگر پیش‌بینی و تشخیص اختلال‌های روانی به کمک چهار روش ذکر شده در مقیاس بزرگ و همچنین به دلیل گسترش استفاده از فضای مجازی، امروزه محققان تلاش‌هایی در جهت استفاده از داده‌های کلان به دست آمده از طریق فضای مجازی برای شناسایی و تشخیص اختلال‌های روان‌شناختی دارند. داده‌های کلان از نظر سطوح تنوع اطلاعاتی، از سطح مولکولی شامل ژنوم، پروتئوم و متابولیسم تا سطوح بالینی و همین‌طور اطلاعات فضای مجازی اهمیت دارند (پاسوس^۸، ۲۰۱۶). از این رو برای تحلیل داده‌های بزرگ روش‌های متفاوت یادگیری ماشین در سال‌های اخیر توسعه پیدا کرده است (لیبرنزاگاریا^۹، ۲۰۱۷). یادگیری ماشین^{۱۰} به توانایی یک سیستم برای تغییر رفتار خود بدون اینکه دقیقاً برای آن رفتار برنامه‌ریزی شده باشد اشاره دارد. یادگیری ماشین الگوریتم‌هایی را طراحی می‌کند که به کامپیوتر اجازه می‌دهند از داده‌ها یاد بگیرد. یاد گرفتن به معنای پیدا کردن طبقات آماری در مجموعه‌ای از داده‌ها است که به کامپیوترها اجازه می‌دهد تا شرایط جدید را با توجه به تجربه‌های قبلی، با تجزیه و تحلیل کردن و خودآموزی بررسی کنند (ورونسه^{۱۱}، ۲۰۱۳). در اصل فرایند آنالیز خودکار فضای مجازی با استفاده از یادگیری ماشین، با ساختن مدل‌های پیش‌بینی کننده به کمک ویژگی‌ها یا متغیرهای استخراج شده از طریق داده‌های فضای مجازی انجام می‌شود. برای مثال یک ویژگی مورد استفاده رایج شامل کدگذاری کردن زبان کاربر مثل میزان استفاده از هر کلمه، زمان پست گذاشتن و متغیرهای دیگر است که این متغیرها به‌عنوان متغیرهای مستقل در یک الگوریتم مشخص می‌شوند (به‌عنوان مثال الگوریتم‌هایی مانند

اختلال‌های روان‌شناختی یک الگوی رفتاری و روانی هستند که با نقص در کارکردهای زیستی، روان‌شناختی، ژنتیکی، فیزیولوژیکی و اجتماعی همراه هستند (کومی^۱، ۲۰۰؛ کاپلان^۲، ۱۹۹۶). سازمان بهداشت جهانی در سال ۲۰۰۱ اعلام کرد که تقریباً چهل و پنج میلیون نفر در سراسر دنیا از اختلال‌های روانی رنج می‌برند و از هر چهار نفر یک نفر در مرحله‌ای از زندگی خود علائم اختلال‌های روانی را تجربه می‌کند (یاسمی^۳، ۲۰۰۵). طبق یافته‌های آماری، بیماری‌های روانی پنجمین بیماری در جهان هستند (وایت فود^۴، ۲۰۱۳). همچنین هزینه اقتصادی ناشی از بیماری‌های روانی در سال ۲۰۱۰، ۲۰۵ تریلیون دلار تخمین زده شد و انتظار می‌رود این هزینه تا سال ۲۰۳۰ به دو برابر افزایش یابد (بلوم^۵، ۲۰۱۲). از طرفی افزایش اختلالات روان‌شناختی با کاهش کیفی سطح عملکردی فرد در زندگی همراه است (مساگنو و مولان گرت^۶، ۲۰۱۰). از این رو هدف اصلی برنامه جامع سلامت روان سازمان سلامت جهانی (۲۰۱۳-۲۰۲۰) تقویت سیستم اطلاعاتی سطوح سلامت روان افراد است تا بتواند با پیشگیری یا تشخیص زودهنگام این بیماری‌ها به سلامت عمومی کمک نماید (سازمان بهداشت جهانی، ۲۰۱۳؛ گلدبرگ^۷، ۱۹۷۲؛ یعقوبی، ۱۳۸۷). در این رابطه از دیرباز روش‌های مختلفی برای گردآوری اطلاعات سلامت روانی و میزان شیوع بیماری‌های روانی افراد وجود داشته که شامل ۴ روش متداول مصاحبه توسط روان‌پزشک، دستیاران پژوهشی، پزشکان عمومی و مطالعه‌های دو مرحله‌ای است. به‌عنوان مثال در مطالعه‌های دو مرحله‌ای، در مرحله اول از یک آزمون غربال‌گری و در مرحله دوم مصاحبه بالینی کامل توسط یک متخصص بالینی مجرب برای شناسایی سطوح سلامت روان استفاده می‌شود. در مرحله اول غربال‌گری اختلال‌های روانی شامل شناسایی بیماری‌های احتمالی روانی قبلاً شناسایی یا گزارش نشده که به کمک آزمون‌ها، معاینه‌ها و سایر روش‌های مناسب

¹ Cumbie

² Kaplan

³ Yasamy

⁴ Whiteford

⁵ Bloom

⁶ Mesagno, Mullane-Grant

⁷ Goldberg

⁸ Passos

⁹ Librenza-Garcia

¹⁰ Machine learning

¹¹ Veronese

کامپیوتر در مجله‌ها چاپ نمی‌شوند و بیشتر در کنفرانس‌ها بازنمون می‌یابند (شامیر^۸، ۲۰۱۰، ژوانگ^۹ و همکارانش، ۲۰۰۷؛ فرین و همکارانش^{۱۰}، ۲۰۱۰؛ قانع و کیومرثی، ۱۳۹۷). بنابراین اسناد ثبت اختراع و مقاله‌های ارائه شده در کنفرانس‌های مرتبط منابع مهمی محسوب می‌شوند (کالوو^{۱۱}، ۲۰۱۷). مقاله‌های مورد نظر با استفاده از ترکیب کلیدواژه‌هایی متمرکز بر پیش‌بینی اختلال‌های روانی و استفاده از فضای مجازی انتخاب شدند. برای جستجوی مقاله‌ها از کلیدواژه‌هایی از جمله "هوش مصنوعی"، "فضای مجازی"، "یادگیری ماشین"، "یادگیری عمیق"، "تویتر"، "فیس‌بوک"، "ردیت" و "افسردگی"، "اختلال‌های روانی"، "اسکیزوفرنی"، "دوقطبی"، "اختلال اضطرابی"، "خودکشی"، "اختلال اضطراب پس از سانحه"، "اختلال وسواس فکری-عملی" استفاده شد. لازم به یادآوری است که از مقاله‌های مرتبط با پیش‌بینی اختلال‌های روان‌شناختی با استفاده از هوش مصنوعی صرفاً بر مقاله‌هایی تمرکز گردید که در آن‌ها از هوش مصنوعی برای شناسایی اختلال‌های روان‌شناختی مبتنی بر داده‌های موجود در فضای مجازی استفاده شده باشد. همچنین برای جستجوی مقاله‌های مرتبط مناسب از بخش منابع مندرج در مقاله‌ها نیز استفاده گردید. فرایند بررسی و انتخاب مقالات به طور خلاصه در شکل یک آورده شده است.

رگرسیون خطی^۱ (نیر^۲، ۱۹۹۶) و ماشین بردار پشتیبان^۳ (کورتس^۴، ۱۹۹۵) و برای پیش‌بینی متغیر وابسته یک خروجی مورد علاقه (مثل سلامت روانی کاربر) با استفاده از الگوریتم در مجموعه آموزشی انتخاب شده آموزش داده می‌شوند و سپس بر روی بخش دیگر (بخش آزمون) ارزیابی می‌شود. همچنین برای پیش‌بینی از فرایند بیش‌برازش^۵ فرایند پیش‌بینی به‌عنوان یکی از نتایج ممکن گزارش می‌شود. در همین جهت محققان تعداد زیادی از مطالعات سلامت روانی را در بافت فضای مجازی بررسی کرده‌اند و الگوهای رفتاری افراد در فضای مجازی را با استرس، افسردگی، اضطراب، خودکشی و سایر اختلال‌های روانی مرتبط دانسته‌اند. هدف این مقاله مرور نظام‌مند بررسی پژوهش‌های انجام شده با استفاده از داده‌های فضای مجازی در پیش‌بینی و تشخیص اختلال‌های روان‌شناختی از طریق فنون یادگیری ماشینی است. همچنین در این مقاله مرور نظام‌مند به شناسایی رایج‌ترین ابزارها و منابع داده مورد استفاده و بیشترین اختلال‌هایی که با استفاده از فضای مجازی بررسی شده، پرداخته شده است.

روش

برای شناسایی بررسی‌های انجام شده و بازبینی یافته‌های گزارش شده، به اطلاعات موجود الکترونیکی و پایگاه اطلاعاتی ساینس‌دایرکت^۶ مراجعه شد و مقاله‌های مورد نیاز جمع‌آوری شد. سپس برای تکمیل فرایند جستجو و وجود فهرست‌بندی مجلات، کنفرانس‌ها و اسناد ثبت اختراع به گوگل اسکالر^۷ مراجعه شد. چرا که برخلاف حوزه سلامت روان سطح معنی‌داری از تحقیقات در علوم

¹ Linear regression

² Neter

³ Support-vector machines

⁴ Cortes

⁵ بیش‌برازش و یا Overfitting به این معنا است که الگوریتم فقط می‌تواند داده‌هایی را که درست یاد گرفته است درستی پیش‌بینی کند ولی اگر داده‌ای کمی از مجموعه آموزشی فاصله داشته باشد، الگوریتمی که Overfit شده باشد، نمی‌تواند به درستی پاسخی برای این داده‌های جدید پیدا کند و آن‌ها را با اشتباه زیادی طبقه‌بندی می‌کند.

⁶ ScienceDirect

⁷ Google Scholar

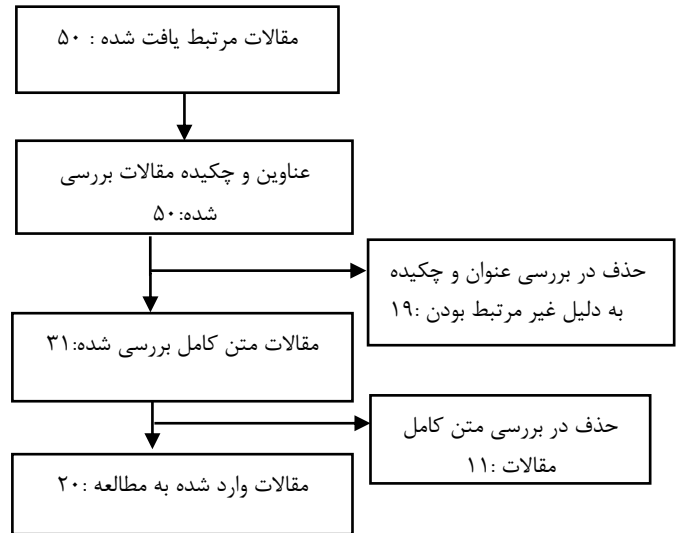
⁸ Shamir

⁹ Zhuang et.al.

¹⁰ Freyne et.al.

¹¹ Calvo

الف) اختلال‌های روان‌شناختی پیش‌بینی شده: در رابطه با اختلال‌های بررسی شده از بین مطالعه‌های صورت گرفته، افسردگی در ۱۰ مقاله، اختلال اضطراب پس از سانحه در ۴ مقاله، مطالعه‌های مرتبط با خودکشی در ۴ مورد، اختلال خوردن و بی‌اشتهایی عصبی در ۲ مطالعه، استرس، اضطراب، اضطراب و افسردگی مختلط، وسواس جبری، بیش‌فعالی، اختلال‌های شخصیت مرزی، افسردگی بعد از زایمان، اسکیزوفرنی، اختلال دوقطبی در یک مطالعه بررسی شده است. بنابراین همان‌طور که مشاهده می‌شود بیشترین مطالعه‌ها در مورد بیماری افسردگی انجام شده است که در این بین مدل یادگیری ماشینی استفاده شده با قدرت پیش‌بینی ۴۲ درصد دقت در بین مقاله‌های بررسی شده کمترین توان پیش‌بینی و با ۸۷ درصد دقت بیشترین توان پیش‌بینی افسردگی را است.



شکل ۱- فرایند بررسی و انتخاب مقالات (درخت تصمیم‌گیری)

یافته‌ها

از میان ۵۰ مقاله بازبینی شده و در دسترس، بعد از بررسی عنوان و چکیده مقاله‌ها با حذف مقالاتی که در آن‌ها از منبع داده‌ای غیر از فضای مجازی برای بررسی استفاده شده بود (مانند تصویرسازی تشدید مغناطیسی^۱)، حذف مقاله‌هایی که در آن‌ها به بررسی "اختلال یا یک مشکل روان‌شناختی" پرداخته نشده بود (مانند بررسی رضایت از زندگی) و همچنین حذف مقاله‌هایی که فقط به بررسی استفاده از فضای مجازی و اختلال‌های روان‌شناختی پرداخته شده بود و از یادگیری ماشین برای تشخیص و پیش‌بینی اختلالات شناختی استفاده نکرده بود صرف‌نظر شد. در نهایت ۲۰ مقاله انتخاب و بازبینی شدند. متغیرهای اصلی مورد بررسی شامل منبع داده‌های جمع‌آوری شده، اختلال‌های بررسی شده، روش جمع‌آوری و تحلیل داده، دقت، تعداد (کاربر، توییت یا پست) در تحقیق بودند. همه این موارد در جدول ۱ درج شده است. با توجه به این جدول به نظر می‌رسد بررسی‌های یاد شده از نظر اطلاعات گردآوری شده و روش دسته‌بندی کم و بیش به یکدیگر نزدیک باشند. در این راستا مطالعه‌های مرور شده به ترتیب سال انتشار در جدول قرار گرفته‌اند. در این رابطه مشخص شد که تمام مقالات بررسی شده در جدول بین سال‌های ۲۰۱۳ تا ۲۰۱۹ منتشر شده‌اند.

¹ Magnetic resonance imaging (MRI)

	SMO ¹⁵ = ۸۹/۵۰ CART= ۸۳/۱ ۸۱/۲J48= ۰/۷۷IB1=	term presence, term frequency, n-gram and part of speech		پیش‌بینی خودکشی در شبکه‌های اجتماعی		
۳۱۲۶ کاربر	اینستاگرام %۸۰= فیس‌بوک=۸۴%	STM ¹⁶ , Principle Component Analysis, WordNet, Data sets, FB-US ¹⁷ , IG_US ¹⁸ , FB_L ¹⁹ , IC_L ²⁰	اختلال‌های روانی مرتبط با فضای مجازی	مطالعه‌ای جامع بر روی شناسایی اختلال‌های روانی شبکه‌های اجتماعی از طریق شبکه‌های اجتماعی آنلاین	اینستاگرام و فیس‌بوک	Hong-Han Shuai, ۲۰۱۷
۳۷۸ کاربر	افسردگی: ۰/۸۷, اضطراب پس از سانحه: ۰/۸۹	RF ¹⁰ , LIWC, Time- series MT_Lab ¹¹ , Sentiment	افسردگی، اضطراب پس از سانحه	پیش‌بینی شروع و روند اختلال‌های روانی از طریق داده‌های توییت	توییت	Andrew G. Reece, ۲۰۱۷
۱۶۶ کاربر	%۷۰	pixel analysis, face detection MTurk ²² HSV ²³ , RF	افسردگی	عکس‌های اینستاگرام آشکارکننده‌ی نشانه‌های پیش‌بینی کننده‌ی افسردگی	اینستاگرام	Andrew G Reece, ۲۰۱۷
۱۴۰۲ کاربر	%۸۵	NB, MSNL ²⁴ , WDL ²⁵ , MDL ²⁶ , LIWC	افسردگی	شناسایی افسردگی از طریق شبکه‌های اجتماعی	توییت	Guangyao Shen, ۲۰۱۷
۱۱۴۵ کاربر	۰/۸۳=CNN ۰/۹۱=RNN	RNN ²⁷ , CNN ²⁸ , rando m trainable, SG ²⁹ , CBO W ³⁰ , CLPsych2015 ³¹ , Bell Let's Talk dataset	افسردگی، اضطراب پس از سانحه	یادگیری عمیق برای شناسایی افسردگی در کاربران توییت	توییت	Ahmed Husseini Orabi, ۲۰۱۸
۱۰۲۷ کاربر	-	BOW ³² model, UMLS ³³ , logistic regression, RF SVM, RNN	افسردگی، بی‌اشتهایی عصبی	شناسایی زود هنگام نشانه‌های بی‌اشتهایی عصبی و افسردگی در شبکه‌های اجتماعی با استفاده از چارچوب‌های مؤثر یادگیری ماشینی	ردیت	Sayanta Paul, ۲۰۱۸
کامنت ۷۱۴۶	بین ۶۰ تا ۸۰%	KNN ³⁴ , DT, SVM, LI WC, Ensemble Vote Classifier	افسردگی	شناسایی افسردگی در داده‌های شبکه اجتماعی با استفاده از فنون یادگیری ماشین	فیس‌بوک	Md Rafiqul Islam, ۲۰۱۸
۱۰۰ کاربر	%۸۵	NB, RF, gradient boosting Ensemble Vote Classifier	اضطراب و افسردگی مختلط	پیش‌بینی اختلال اضطراب و افسردگی مختلط از داده‌های فضای مجازی	توییت	Akshi Kumar, ۲۰۱۹

1. Principal component analysis, 2. Super Vector Machine, 3. Radial Basis Function Kernel, 4. Linguistic Inquiry and Word Count, 5. Linear discriminant analysis, 6. Dot NetNuck 7. Common Application Environment, 8. Term frequency-inverse document frequency, 9. Compact Language Detector, 10. Random Forests, 11. Material testing Lab Software, 12. Native Bayes classifier 13. Decision Trees, 14. Waikato Environment for Knowledge Analysis, 15. Sequential minimal optimization, 16. SNMD-based Tensor Model 17. User Study_ Facebook 18. User Study_ Instagram 19. Face Book large-scale datasets 20. Instagram large-scale datasets 21. social network mental disorders 22. Amazon's Mechanical Turk, 23. Hue, Saturation, and Value 24. Multiple Social Networking Learning, 25. Wasserstein Dictionary Learning, 26. Multimodal Depressive Dictionary Learning, 27. Recurrent neural network, 28. Convolutional neural network, 29. skip-gram, 30. Continuous Bag-of-words, 31. Computational Linguistics and Clinical Psychology, 32. Bag of words 33. Unified Medical Language System 34. k-nearest neighbors algorithm

به دو روش مشخص جمع‌آوری می‌شدند. جمع‌آوری داده به صورت مستقیم با رضایت شرکت‌کنندگان از طریق مصاحبه و ابزارهای جمع‌آوری الکترونیکی داده مانند فیس‌بوک. روش‌های استفاده شده برای جمع‌آوری داده،

(ب) رویکردهای کلی جمع‌آوری داده از کاربران فضای مجازی: پایگاه داده‌ای هر کدام از مقالات انتخاب شده به صورت مستقیم یا غیرمستقیم شامل داده‌های به دست آمده از طریق فضای مجازی است. داده‌های موجود

روان‌شناختی برای استفاده محققین است (گانتوکو^۸، ۲۰۱۷). در مقالات بررسی شده در این پژوهش لازم به ذکر است که تنها ۳۰٪ این مقالات از پرسشنامه استفاده نمودند و دو پرسشنامه افسردگی Beck's Depression Scale (BDI) و Center for Epidemiologic Studies Depression Scale (CES-D) بیشترین استفاده را در مطالعات مربوط به افسردگی داشته‌اند. به دلیل سنجش ذهن‌گرایانه (subjective)، امکان دستکاری بالا و اقتصادی نبودن از پرسشنامه از این مطالعات کمتر استفاده شده است (ندیم^۹، ۲۰۱۶). در بیشتر مقالات این پژوهش از شیوه‌ی دوم جمع‌آوری داده (پست‌های عمومی در شبکه‌های اجتماعی) استفاده شده و به بررسی عبارت باقاعده در پست‌های عمومی به اشتراک گذاشته شده در فضای مجازی پرداخته شده است. علاوه بر این مطالعه مروری ما نشان داد که ابزار LIWC^{۱۰} بیشترین استفاده را برای تحلیل متن در تحقیقات روان‌شناختی داشته است (در ۶۰ درصد مطالعه‌های بررسی شده در این پژوهش از این ابزار استفاده شده بود). این نرم‌افزار شمارش کلمات و جستجوهای زبان‌شناختی ابزاری روان‌سنجی مبتنی بر تحلیل متون زبان طبیعی است که درجه تعلق کلمات به طبقات گوناگون واژگانی را می‌سنجد (پنباکر و همکاران^{۱۱}، ۲۰۰۱). کاربران این ابزار می‌توانند فایل مورد نظر خود را انتخاب کنند و LIWC مشخصه‌های مورد نیاز و داده‌های به دست آمده از هر کدام از این مشخصه‌ها را استخراج می‌کند. البته این ابزار معایبی هم دارد: اولاً، یک نرم‌افزار تجاری است که به جز نسخه محدود وبی، برای استفاده کامل از آن کاربران باید آن را خریداری کنند. دوماً، تغییر دادن مشخصه‌های این ابزار کار راحتی نبوده و برای انجام آن محققان نیاز به مهارت‌های برنامه‌نویسی دارند (وانگکوبلاپ^{۱۲}، ۲۰۱۷). برای حل این مشکل ابزارهای جایگزین برای استخراج ویژگی‌ها وجود دارد. محدودیت این ابزارها مربوط به این مسئله است که هر کدام فقط می‌توانند چند ویژگی را استخراج کنند. به‌عنوان مثال وردنت^{۱۳} یک مجموعه واژگان دارای ارتباط

بسته به هدف مطالعه و پلتفرم مورد نظر متفاوت است. این روش‌ها شامل انتشار اطلاعات مربوط به پروژه در وبسایت‌های مرتبط مانند Amazon Mechanical Turk (ریس^۱، ۲۰۱۷) و سپس انتشار تکالیف در این پلتفرم‌ها برای افرادی که تمایل به شرکت در پروژه دارند. همچنین به‌عنوان بخشی از پرسشنامه شرکت‌کنندگان رضایت خود را برای جمع‌آوری داده‌ها از فضای مجازی متعلق به خود اعلام می‌کنند (وانگکوبلاپ، ۲۰۱۷). از نمونه مطالعه‌های بررسی شده در این مقاله که در آن‌ها از این روش برای جمع‌آوری داده استفاده شده می‌توان به تحقیقی اشاره کرد که در آن برای پیش‌بینی افسردگی بعد از زایمان از پرسشنامه‌ی سلامت بیمار (PHQ-9) استفاده شد. این پرسشنامه اطلاعات منتشره در طول دو هفته کاربر در فضای مجازی را بررسی می‌کند. در مقاله‌ای دیگر خودکشی کاربرها با استفاده از پرسشنامه غربال‌گر مقیاس خودکشی مرکز مطالعات همه‌گیرشناسی خودکشی^۲ از طریق پست‌های ۶ الی ۱۶ هفته اخیر توییت‌ر پیش‌بینی شد. دومین رویکرد برای جمع‌آوری داده، استفاده از پست‌های عمومی در شبکه‌های اجتماعی با استفاده از عبارات منظم^۳ است که در آن برای جستجوی پست‌های مربوطه مانند "من مبتلا به (عنوان مورد بررسی) تشخیص داده‌شده‌ام" استفاده می‌شود. برای جمع‌آوری داده‌های فضای مجازی، منبع داده نیازمند یک مکانیسم برای به دست آوردن داده‌های مورد نیاز خود هستند. به همین منظور تحقیقات مبتنی بر فیس‌بوک پایگاه داده‌ی کاربران را به‌وسیله توسعه ابزارهای سفارشی^۴ یا برنامه‌های کاربردی وبی^۵ متصل به رابط برنامه‌نویسی کاربری اند. برخی تحقیقات از رابط کاربری توییت‌ر، اینستاگرام و سینا ویبو^۶ (کیتور^۷، ۲۰۰۸) برای جمع‌آوری داده استفاده می‌کنند (وانگکوبلاپ، ۲۰۱۷). یکی دیگر از پایگاه‌های معتبر در مطالعه‌های این حوزه استفاده از داده‌های پایگاه my personality project است که شامل داده‌های شبکه اجتماعی و هم‌آزمون‌های

⁸ Guntuku

⁹ Nadeem

¹⁰ Linguistic Inquiry and Word Count

¹¹ Pennebaker

¹² Wongkoblapp

¹³ WordNet

¹ Reece

² Center for Epidermologic studies/CES-D

³ Regular Expressions

⁴ Custom tool

⁵ Web apps

⁶ Sina Weibo

⁷ Kittur

بحث و نتیجه‌گیری

مطالعه‌های قبلی نیز نشان می‌دهد بیشترین تعداد این نوع تحقیق‌ها بر افسردگی تمرکز داشته‌اند (گانتوکو^۵، ۲۰۱۷). افسردگی به شکل خفیف و موقت در پاسخ آدمی به فشارهای زندگی وجود دارد ولیکن تداوم طولانی مدت خلق افسرده منجر به اختلال افسردگی می‌شود (سادوک^۶ و همکاران، ۲۰۱۵) که تقریباً فقط نیمی از افراد افسرده توسط پزشکان شناسایی می‌شوند و از این میان فقط ۱۳ الی ۴۹ درصد آن‌ها حداقل درمان کافی را دریافت می‌کنند (وانگ^۷، ۲۰۰۵). یادگیری ماشین با آنالیز خودکار داده‌های فضای مجازی به صورت بالقوه روش‌هایی را برای تشخیص زودهنگام فراهم می‌کند. (چاندر^۸، ۲۰۱۷). از طرفی ارزیابی اختلال‌های روانی بر اساس شبکه‌های اجتماعی سؤال‌های اخلاقی بی‌شماری ایجاد نموده است. بیشترین نگرانی در مورد دسترسی به زندگی خصوصی افراد است که امکان استنتاج اطلاعات از طریق شبکه‌های اجتماعی در مورد وضعیت روانی افراد می‌تواند پیامدهای جدی در پی داشته باشد (مثلاً افزایش نرخ بیمه، عدم استخدام و غیره). همچنین به دلیل اینکه اختلال‌های روانی برچسب اجتماعی با بار منفی محسوب می‌شوند و ممکن است باعث تبعیض شوند، لازم است که به کاربران این اطمینان داده شود که از اطلاعات آن‌ها به خوبی حفاظت می‌شود (گونتاک^۹، ۲۰۱۷). به همین دلیل سازمان محققان اینترنت راهبردهایی برای محافظت از اطلاعات اشخاص پیشنهاد کرده است. برای مثال محققان کنوی و مک قواعد هنجاری را برای استفاده از اطلاعات عمومی توئیتر پیشنهاد کردند به این ترتیب که پاراگراف نویسی مجدد پست‌های جمع‌آوری شده، گرفتن رضایت از کاربران، مخفی کردن هویت افراد و حفاظت از اطلاعات جمع‌آوری شده است (کنوی، ۲۰۱۴؛ مک کی، ۲۰۱۳). همچنین ویلکنسون و همکاران پیشنهاد کرده‌اند که محققان نباید پست‌ها یا URL پیام‌ها را مستقیماً در مقاله‌ها نقل قول کنند، زیرا از این طرق می‌توان اشخاص را

معنایی بزرگ چندزبانه است که برای استخراج اجزای کلام با روابط مفهومی از متن و برای به دست آوردن معنای لغوی کلمات استفاده می‌شود (میلر^۱، ۱۹۹۵). Mallet یک ابزار مفید تحلیل زبان طبیعی برای طبقه‌بندی کردن مدارک، ایجاد کردن عنوان‌ها و اجرای برچسب زنی متوالی^۲ است که برای تحلیل و استخراج داده در این نوع مطالعات یعنی استخراج داده از فضای مجازی کاربرد دارد (مک کالوم^۳، ۲۰۰۲).

(ج) منبع داده: در تحقیق حاضر ۶۵ درصد مقاله‌ها از توئیتر به‌عنوان منبع داده‌ی خود استفاده کرده‌اند. استفاده از داده‌های عمومی (مانند توئیتر، فیس‌بوک، ردیت و سینا ویبو) فوایدی از جمله داشتن نمونه بزرگ‌تر و گردآوری داده به‌صورت سریع‌تر و کم‌هزینه‌تر نسبت به گردآوری داده از طریق مصاحبه‌ها دارد؛ هر چند ارزیابی‌های مبتنی بر مصاحبه سطح بالاتری از اعتبار برخوردارند. همچنین در تحقیق لین^۴ در سال ۲۰۱۴ در بین مقاله‌های بررسی شده توئیتر یک منبع پرکاربرد بین محققان بود. زیرا توئیتر دو راه متفاوت را برای دستیابی به داده فراهم می‌کند: جستجوی گذشته‌نگر و آینده‌نگر. در جستجوی گذشته‌نگر امکان دستیابی به تمام توئیتهای را با استفاده از جستجوی آینده‌نگر با قاعده فراهم می‌کند. درحالی که جستجوی آینده‌نگر به جستجوی همه توئیتهای مرتبطی که پست خواهند شد می‌پردازد. همچنین تفاوتی اساسی که بین تحقیقات انجام شده به‌وسیله فیس‌بوک در مقایسه با توئیتر و سینا ویبو وجود دارد این نکته مهم است که فیس‌بوک اجازه‌ی دسترسی به روابط دوستانه و تعامل بین کاربران را به محقق نمی‌دهد. به‌علاوه کاربران باید رضایت خود را برای اینکه یک برنامه اجازه‌ی استخراج داده‌های آن‌ها را داشته باشد اعلام کنند. در حالی که سایت‌های میکرو بلاگینگ مثل توئیتر امکان دسترسی به این نوع داده‌ها را برای محققین فراهم می‌کنند (وانگوبلاپ^۵، ۲۰۱۷).

⁵ Guntuku

⁶ Sadock

⁷ Wang

⁸ Chandra

¹ Miller

² Sequence labelling

³ McCallum

⁴ Huiji lin

روان، انجام بررسی‌های دوره‌ای نیازسنجی با هدف به‌روز کردن اولویت‌های ارتقای سلامت روان جامعه است و با در نظر گرفتن استفاده روزافزون مردم از شبکه‌های اجتماعی و گسترش روزافزون به‌کارگیری یادگیری ماشین در حوزه‌های مختلف می‌توان از این دو پتانسیل بالقوه در راستای شناسایی و پیش‌بینی دقیق‌تر اختلالات روان‌شناختی استفاده نمود.

شناسایی نمود. بنابراین ضروری است پژوهش‌هایی از این دست اطلاعات پروژه‌ی خود را به IRBS و یا کمیته‌های اخلاقی ارائه دهند تا صلاحیت لازم برای جمع‌آوری اطلاعات را کسب کنند (چودهوری، ۲۰۱۳). بنابراین محققان باید اطلاعات شخصی کاربران را محترم شمارند. به‌علاوه چگونگی اطلاع‌رسانی به کاربران در مورد اختلال‌های آن‌ها نیز باید مدنظر قرار گیرد.

بنظر می‌رسد با استناد به طرح جامع سلامت روان کشور یکی از اولین راهبردهای پیشگیری سطح نخست سلامت

منابع

- Birjali, M., Beni-Hssane, A., Erritali, M. (2017). Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks. *Procedia Computer Science*. 113: 65-72.
- Bloom, D.E., Cafiero, E.T., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L.R., Fathima, S., Feigl, A.B., Gaziano, T., Mowafi, M., Pandya, A., Prettnner, K., Rosenberg, L., Seligman, B., Stein, A.Z., & Weinstein, C. (2011). *The Global Economic Burden of Noncommunicable Diseases*. Geneva: World Economic Forum.
- Burnap, P., Colombo, G., Amery, R., Hodorog, A., Scourfield, J. (2017). Multi-class machine classification of suicide-related communication on Twitter. *Online Social Networks and Media*. 2: 32-44.
- Calvo, R.A., Milne, D.N., Hussain, M.S. (2017). Christensen H. Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering*. 23(5): 649-85.
- Conway, M. (2014). Ethical issues in using Twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature. *Journal of medical Internet research*. 16(12): e290.
- Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K. (2015). From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. 1-10.
- Coppersmith, G., Harman, C., Dredze, M. (2014). Measuring post-traumatic stress disorder in Twitter. In *Eighth international AAAI conference on weblogs and social media*
- Cortes, C., Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20 (3):273-97.
- Cumbie, S.A., Conley, V.M., Burman, M.E. (2004). Advanced practice nursing model for comprehensive care with chronic illness: model for promoting process engagement. *Advances in Nursing Science*. 27(1):70-80.
- De Choudhury, M., Counts, S., Horvitz, E.J., Hoff, A. (2014). Characterizing and predicting postpartum depression from shared Facebook data. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 626-638.
- De Choudhury, M., Gamon, M., Counts, S., Horvitz, E. (2013). Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*.
- Freyne, J., Coyle, L., Smyth, B., Cunningham, P. (2010). Relative status of journal and conference publications in computer science. *Communications of the ACM*. 53(11): 124-32.
- Ghane, M.R., Kumarsi, S. (2018). Four Decades of Scientific Activity in Iran from

- the Perspective of the Conference Proceedings, Highly Cited and Hot Papers, and Open Access Papers, in the Light of the Law of Economic, Social, and Cultural Development Plan of Iran. 34 (1): 27-56. [Persian]
- Goldberg, D.P. The Detection of Psychiatric Illness by Questionnaire. Maudsley Monograph 21 ed.
- Guntuku, S.C., Yaden, D.B., Kern, M.L., Ungar, L.H., Eichstaedt, J.C. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*. 18:43-9.
- Islam, M.R., Kabir, M.A., Ahmed, A., Kamal, A.R., Wang, H., Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. *Health information science and systems*. 6(1): 8.
- Kaplan, H., Sadouk, B. (1996). Abstract of clinical psychiatry. First edition, publication of Hayyan. 153-158.
- Kittur, A., Chi, E.H., Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 453-456
- Kumar, A., Sharma, A., Arora, A. Anxious Depression Prediction in Real-time Social Data. arXiv:1903.10222
- Kumar, M., Dredze, M., Coppersmith, G., De Choudhury, M. (2015). Detecting changes in suicide content manifested in social media following celebrity suicides. In *Proceedings of the 26th ACM conference on Hypertext & Social Media*, 85-94.
- Librenza-Garcia, D., Kotzian, B.J., Yang, J., Mwangi, B, Cao, B., Lima, L.N., Bermudez, M.B., Boeira, M.V., Kapczinski, F., Passos, I.C. (2017). The impact of machine learning techniques in the study of bipolar disorder: a systematic review. *Neuroscience & Biobehavioral Reviews*. 80 :538-54.
- Lin H, Jia J, Guo Q, Xue Y, Li Q, Huang J, Cai L, Feng L. (2014). User-level psychological stress detection from social media using deep neural network. In *Proceedings of the 22nd ACM international conference on Multimedia* 507-516.
- Mesagno C, Mullane-Grant T. (2010). A comparison of different performance Routines as possible choking interventions. *Journal of Applied Sport Psychology*. 22(3):343-60.
- McCallum, A.k. (2002). Mallet: A machine learning for language toolkit. <http://mallet.cs.umass.edu>.
- McKee, R. (2013). Ethical issues in using social media for health and health care research. *Health Policy*. 110(2-3): 298-301.
- Miller, G.A. (1995). WordNet: a lexical database for English. *Communications of the ACM*. 38(11): 39-41.
- Nadeem, M. (2016). Identifying depression on Twitter. ArXiv: preprint, 1607.073.
- Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W. (1996). *Applied linear statistical models*. Models. Chicago: Irwin;
- O'Dea, B., Wan, S., Batterham, P.J., Calcar, A.L., Paris, C. Christensen, H. (2015). Detecting suicidality on Twitter. *Internet Interventions*. 2(2), 183-191
- Orabi, A.H., Buddhitha, P., Orabi, M.H., Inkpen, D. (2018). Deep learning for depression detection of twitter users. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*. 88-97.
- Passos, I.C., Mwangi, B., Vieta, E., Berk, M., Kapczinski, F. (2016). Areas of controversy in neuroprogression in bipolar disorder. *Acta Psychiatrica Scandinavica*; 134(2):91-103.
- Paul, S., Jandhyala, S.K., Basu, T. (2018). Early Detection of Signs of Anorexia and Depression over Social Media using Effective Machine Learning Frameworks. In *CLEF (Working Notes)*.
- Pennebaker, J.W., Francis, M.E., Booth, R.J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Mahway: Lawrence Erlbaum Associates: .71: 2001.

- Reece, A.G., Danforth, C.M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*.6(1): 15.
- Reece, A.G., Reagan, A.J., Lix, K.L., Dodds, P.S., Danforth, C.M., Langer, E.J. (2017). Forecasting the onset and course of mental illness with Twitter data. *Scientific reports*. 7(1), 13006.
- Sadock, B. J., Sadock, V. A., Ruiz, P., & Kaplan, H. I. (2015). *Kaplan & Sadock's synopsis of psychiatry: Behavioral sciences, clinical psychiatry*. Philadelphia: Wolters Kluwer.
- Schwartz, H.A., Eichstaedt, J., Kern, M.L., Park, G., Sap, M., Stillwell, D., Kosinski, M., Ungar, L. (2014). Towards assessing changes in degree of depression through Facebook. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. 118-125.
- Shamir, L. (2010). The effect of conference proceedings on the scholarly communication in Computer Science and Engineering. *Scholarly and Research Communication*, 1(2).
- Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T.S., Zhu, W. (2017). Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution. In *IJCAI*, 3838-3844.
- Shuai, H.H., Shen, C.Y., Yang, D.N., Lan, Y.F., Lee, W.C., Philip, S.Y., Chen, M.S. (2017). A comprehensive study on social network mental disorders detection via online social media mining. *IEEE Transactions on Knowledge and Data Engineering*. 30 (7):1212-25.
- Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., & Ohsaki, H. (2015). Recognizing depression from twitter activity. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 3187-3196.
- Veronese, E., Castellani, U., Peruzzo, D., Bellani, M., Brambilla, P. (2013). Machine learning approaches: from theory to application in schizophrenia. *Computational and mathematical methods in medicine*, 2013.
- World Health Organization. *Mental health action plan 2013-2020*.
- Wang, P.S., Lane, M., Olfson, M., Pincus, H.A., Wells, K.B., Kessler, R.C. (2005). Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Archives of general psychiatry*. 62 (6): 629-40.
- Whiteford, H.A., Degenhardt, L., Rehm, J., Baxter, A.J., Ferrari, A.J., Erskine, H.E., Charlson, F.J., Norman, R.E., Flaxman, A.D., Johns, N., Burstein, R. (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study. *The Lancet*. 382 (9904): 1575-86.
- Wongkoblap, A., Vadillo, M.A., Curcin, V. (2017). Researching mental health disorders in the era of social media: systematic review. *Journal of medical Internet research*. 19(6): e228.
- Yagubi, H. (2008). Screening of Mental Disorders: Tests Position, determining of cut of -point and validation. *Quarterly Journal of mental health*. 1(1):39-51. [Persian].
- Yasamy, M.T., Sardarpour Goudarzi, S.H., Amin Esmaeeli, M., Mahdavi, N., Ebrahimpour, A. Bagheri yazdi, S.A. (2005). *Practical Mental health for general and family particitioner*. Tehran: Aramesh.
- Zhuang, Z., Elmacioglu, E., Lee, D., Giles, C.L. (2007). Measuring conference quality by mining program committee characteristics. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*. 225-234. ACM.