



دانشگاه کاشان
University of Kashan

مجله محاسبات نرم

SOFT COMPUTING JOURNAL

تارنمای مجله: sci.kashanu.ac.ir



شخصیت‌شناسی در شبکه‌های اجتماعی با استفاده از مدل‌سازی موضوعی نظرات کاربران

آرش خسروی^۱، استادیار، حمیدرضا عبدالحسینی^۲، دانشجوی کارشناسی ارشد

^۱ دانشکده مهندسی، مرکز آموزش عالی محلات، محلات، ایران.

^۲ گروه مهندسی کامپیوتر، دانشگاه شهاب دانش، قم، ایران.

اطلاعات مقاله

چکیده

تاریخچه مقاله:

دریافت ۲۴ مهر ماه ۱۴۰۰

پذیرش ۱۶ اسفند ماه ۱۴۰۱

کلمات کلیدی:

شبکه اجتماعی

شخصیت‌شناسی

داده‌کاوی

مدل‌سازی موضوعی

شبکه عصبی عمیق

SVM

شخصیت فرد مجموعه‌ای از خصوصیات و واکنش‌های فرد است که موجب ایجاد رفتار فرد در موقعیت‌های گوناگون می‌شود. به طور معمول این رفتارها در موقعیت‌های مشابه تکرار می‌شوند. داشتن اطلاعات در مورد شخصیت فرد، این امکان را به ما می‌دهد که واکنش فرد را در موقعیت‌های مشابه پیش‌بینی کنیم. از طرفی با توجه به اینکه افراد در شبکه‌های اجتماعی مجازی نظرات مستقیمی بر نظرات خود احساس نمی‌کنند، استفاده از این شبکه‌ها جهت جمع‌آوری اطلاعات و تحلیل شخصیت افراد بسیار قابل توجه و منطقی است. در این مقاله، با دریافت نظرات کاربران از شبکه‌های اجتماعی و مقایسه آن با پرسشنامه پر شده توسط افراد، میزان درستی پاسخ‌گویی به سوال‌های پرسشنامه را محک می‌زنیم و به گونه‌ای دقیق‌تر به تحلیل شخصیت کاربران می‌پردازیم. با پردازش نظرات کاربران توسط روش‌های هوشمند متن‌کاوی و استفاده از یادگیری ماشین تلاش شده است که شخصیت کاربران پیش‌بینی شود. در نهایت، در هر چهار روش از روش‌های الگوریتم SVM و همچنین روش شبکه عصبی عمیق، معیارهای درستی، دقت و بازخوانی به نسبت بسیار خوبی بالا بودند (بین ۰/۷۵ تا ۰/۹۹) و این حاکی از مطابقت بالای نظرات جمع‌آوری شده از طریق پرسشنامه با پست‌های منتشر شده در فضای مجازی همان اشخاص می‌باشد.

© ۱۴۰۱ نویسندگان. مقاله با دسترسی آزاد تحت مجوز CC-BY

۱. مقدمه

پیش‌بینی واکنش وی در موقعیت‌های مشابه را فراهم می‌کند. شناخت شخصیت افراد می‌تواند برای شناخت نیازهای بالقوه افراد مفید باشد و به سیستم‌های هوشمند این امکان را بدهد که بتوانند خود را با نیازهای اشخاص تطبیق دهند. در کل، شناخت شخصیت افراد در زمینه‌های بسیاری می‌تواند مفید باشد، از آن جمله می‌توان به سیستم‌های بازاریابی، فناوری‌های توصیه‌گر^۱، آموزش الکترونیکی^۲، تجارت الکترونیک^۳، سیستم‌های بهداشت

شخصیت فرد را می‌توان به عنوان مجموعه‌ای از خصوصیات توصیف کرد که رفتار فرد را در موقعیت‌های گوناگون می‌سازد که به طور معمول در موقعیت‌های مشابه تکرار می‌شوند. از این رو، آگاهی و داشتن اطلاعات در مورد شخصیت یک فرد، امکان

✦ نوع مقاله: پژوهشی

* نویسنده مسئول

پست(های) الکترونیک: khosravi.280@gmail.com (خسروی)

hr.engineer.hr@gmail.com (عبدالحسینی)

¹ Assistive

² E-Learning

³ E-Commerce

شبکه‌ها برای گردآوری اطلاعات و تحلیل شخصیت افراد بسیار قابل توجه و منطقی‌تر باشد.

در گذشته تحقیق‌های شناخت شخصیت، الگو و چهارچوب قابل قبولی برای توصیف ساختار و معرفی شخصیت نداشتند و در خصوص ویژگی‌های شخصیت مورد تحقیق، بین محقق‌ها اختلاف نظرهایی وجود داشت. اما ارائه مدل «پنجگانه بزرگ» [۳]، تحولی در دانش شخصیت‌شناسی پدید آورد و بسیاری از روانشناسان این مدل را یکی از بهترین و باثبات‌ترین مقیاس‌های شناخت شخصیت می‌دانند. پنج ویژگی بزرگ شخصیت شامل موارد زیر می‌باشد:

- ۱) پذیرا بودن در تجربه^۴: کنجکاوی، هوشمندی و تخیل
- ۲) باوجدان^۵: مسئول، سازمان‌یافته و دارای پشتکار. افراد باوجدان شدیداً قابل اعتماد هستند.
- ۳) برون‌گرا^۶: اجتماعی، دوستانه، شجاع و پرانرژی. افراد برون‌گرا از موقعیت‌های اجتماعی الهام می‌گیرند.
- ۴) سازگاری^۷: همکاری و کمک‌کننده. اینها کسانی هستند که به طور کلی خوش‌بین بوده و به دیگران اعتماد می‌کنند.
- ۵) روان‌رنجوری^۸: مضطرب، نامطمئن و حساس. اینها افرادی عصبی، اخمو و ناراحت هستند و به راحتی در معرض احساسات منفی قرار می‌گیرند.

در این تحقیق، ابتدا شخصیت هر داوطلب از طریق پرسشنامه شخصیت تحلیل شده و در ادامه با دریافت اطلاعات مرتبط با وی در شبکه‌های اجتماعی مانند متن‌هایی که خود می‌نویسند و یا متن‌هایی که از دیگران به اشتراک می‌گذارد و همچنین متن‌ها و پست‌هایی که لایک می‌کنند، شخصیت وی تحلیل و یک مدل شخصیت‌شناسی ساخته می‌شود. سپس متن به اشتراک گذاشته شده به وسیله وی توسط یادگیری ماشین مدل‌سازی می‌گردد. بعد از آموزش ماشین و بر مبنای مدل ساخته شده، می‌توان شخصیت افراد را به واسطه متون اشتراک گذاشته شده آنها پیش‌بینی کرد. در تحقیقات گذشته، شخصیت‌شناسی بر اساس

و سلامت^۱، سیستم‌های پیشنهاددهنده^۲ و موارد دیگر اشاره کرد [۱]. روش معمول برای به‌دست آوردن شخصیت کاربر، پرکردن پرسشنامه توسط خود شخص است. اما از طرفی هم نیاز به شناخت شخصیت کاربر به صورت کاملاً آزاد و بدون نظارت مستقیم نیز اهمیت دارد. فرض معمول این است که کاربران با شخصیت‌های مشابه الگوهای رفتاری یکسانی هنگام برقراری ارتباط از طریق شبکه‌های اجتماعی مجازی نشان می‌دهند و از این الگوها می‌توان به منظور پیش‌بینی شخصیت آنها استفاده کرد [۲]. در فرآیند ایجاد پروفایل‌های شبکه‌های اجتماعی، کاربران با آنچه که به اشتراک می‌گذارند و نحوه بیان مطالب، شخصیت خود را نشان می‌دهند. بسیاری از ویژگی‌های شخصیت کاربر از طریق شرح ویژگی‌ها و بیان شخصی خود، به‌روزرسانی مطالب، عکس‌ها و غیره به دست می‌آید [۳]. همچنین در انتخاب‌های برخط مشتریان در فروشگاه‌های مجازی [۴]، شناخت علایق و پیش‌بینی سلیقه خریداران توسط رفتارشناسی آنها برای بازاریابان بسیار مفید است. یکی دیگر از کاربردهای شخصیت‌شناسی، مربوط به خشونت سایبری^۳ است. می‌توان نشان داد که ویژگی‌های شخصیتی چقدر با خشونت سایبری ارتباط دارد. لذا برای پیشگیری از چنین خشونت‌هایی و همچنین جرایم مختلف سایبری، فراهم کردن الگوریتم‌های یادگیری ماشین که بتواند شخصیت مخاطبین را در شبکه‌های اجتماعی پیش‌بینی کند، بسیار مفید خواهد بود [۵].

امروزه با توجه به گسترش روزافزون استفاده از اینترنت، برنامه‌ها و شبکه‌های مجازی توسط کاربران، انسان‌ها در فضای مجازی زندگی می‌کنند و وقت قابل توجهی را جهت مشاهده و به اشتراک گذاری مطالب در این فضا می‌گذارند و این موقعیت و پتانسیل مناسبی برای جمع‌آوری داده‌های بهتر و دقیق‌تر در مورد اشخاص در این بستر است. از طرف دیگر، با توجه به اینکه افراد در شبکه‌های مجازی نظارت مستقیمی بر ارائه نظرات خود احساس نمی‌کنند، صداقت بیشتری در بیان خود داشته و همین امر سبب می‌شود استفاده از ماهیت و بستر این

⁴ Openness to Experience

⁵ Conscientiousness

⁶ Extroversion

⁷ Agreeableness

⁸ Neuroticism

¹ Health

² Recommender Systems

³ Cyber-violence

دیگر، این مطالعات پیشرفت قابل توجهی در زمینه ایجاد خودکار پروفایل شخصی کاربران دارند، با این حال بیشتر آنها فقط از یک منبع و یا یک روش واحد استفاده کرده‌اند که این امر سبب می‌شود چنین پروفایل‌هایی عملکرد مطلوبی نداشته باشند. با توجه به اینکه اکثر کاربران شبکه‌های اجتماعی در زندگی روزانه خود از بیش از یک شبکه اجتماعی استفاده می‌کنند، پس منطقی است برای ایجاد پروفایل شخصی کاربران از بیش از یک منبع استفاده شود.

دقت کنید که در مدل‌سازی موضوعی، فرض می‌کنیم که متون وارد شده به سیستم از چند موضوع نامشخص ساخته شده و می‌بایست این موضوعات نامعلوم پیدا شوند. هر موضوع به صورت جداگانه یک توزیع احتمال نامعلوم روی واژه‌ها و هر متن یک توزیع احتمالی روی موضوع‌ها می‌باشد.

بویتر [۶]، چارچوب پیشنهاد دهنده محصول مبتنی بر شخصیت (PBPR)^۲ را بر اساس نظریه شخصیت بنا نهاده است که ادعا می‌کند شخصیت کاربر به طور قابل توجهی بر بسیاری موارد مهم تاثیر می‌گذارد. این چارچوب برای تجزیه و تحلیل داده‌های رسانه‌های اجتماعی به منظور پیش‌بینی شخصیت کاربر و سپس ترجیحات محصول بر اساس شخصیت فرد بسیار کاربردی است. چارچوب توصیه‌کننده مبتنی بر شخصیت شامل سه موتور است که عبارتند از بازبایی ویژگی‌های برخط شبکه اجتماعی مربوط به شخصیت، پیش‌بینی شخصیت کاربر و توصیه محصول. این مدل بر مبنای داده‌های شخصی خود گزارش شده از شبکه‌های اجتماعی، مورد تحقیق قرار گرفته است و برای متغیرهای دیگر از جمله سن، جنسیت، شخصیت و وضعیت سلامتی می‌بایست دوباره تکرار شود.

اورتیگوستا و همکاران [۲]، با هدف استنباط شخصیت از تجزیه و تحلیل تعاملات کاربر در شبکه‌های اجتماعی، یک برنامه را روی فیس‌بوک توسعه داده‌اند که برای جمع‌آوری اطلاعات در مورد ویژگی‌های شخصیتی بیش از ۲۰,۰۰۰ کاربر، همراه با تعاملات آنها استفاده شده است. بر اساس داده‌های جمع‌آوری شده، طبقه‌بندی‌کننده‌های خودکار با استفاده از تکنیک‌های

متن فارسی کمتر مورد توجه بوده است و این تحقیق تلاش دارد که به محققین این حوزه کمک کند. نوآوری تحقیق جاری در چند جنبه قابل بیان است. اول اینکه این یک تحقیق بین رشته‌ای است (روانشناسی و هوش مصنوعی) و سعی می‌کند شخصیت افراد را از جملات و الفاظی که در شبکه اجتماعی به اشتراک می‌گذارند، پیش‌بینی کند. با اینکه این کار در زبان انگلیسی بر مبنای شخصیت‌شناسی پنجگانه بزرگ انجام شده، اما در زبان فارسی با رویکرد یادگیری ماشین کمتر مورد توجه بوده و در علم روانشناسی و پزشکی مطالعه هر مورد یا مجموعه‌ای از موارد با توجه به شرایط فرهنگی و عوامل محیطی می‌تواند یک پژوهش نو باشد. در اینجا استفاده از عبارات فارسی، جمع‌آوری پرسشنامه اشخاص و به علاوه متنی که به اشتراک گذاشته‌اند و سپس پیش‌پردازش آنها برای یادگیری ماشین و پیش‌بینی بر اساس یادگیری ماشین، پژوهشی زمان‌بر و پرهزینه بوده است. لذا تولید دیتاستی که بتواند در این حوزه مورد استفاده قرار گیرد از نوآوری‌های دیگر این تحقیق است. واضح است که عبارات و کلماتی که برای هر شخصیت استفاده شده است در زبان فارسی و با توجه به شرایط فرهنگی و محیطی متفاوت از پژوهش‌های غیرفارسی می‌باشد. دوم، استفاده از هسته‌های مختلف SVM^۱ به همراه یادگیری عمیق و مقایسه آنها می‌تواند به عنوان یک گزارش علمی برای پژوهشگران این حوزه مفید باشد. سوم، نوشتن خزشگر وب و جمع‌آوری متن‌های نوشته شده توسط هر فرد در اینستاگرام و پیش‌پردازش آنها از کارهای مقدماتی برای جمع‌آوری داده بوده و همزمان پرسشنامه شخصیت‌شناسی افراد هم جمع‌آوری شده است. در ادامه این مقاله کارهای پیشین، مدل پیشنهادی تحقیق و تحلیل نتایج خواهد آمد.

۲. کارهای پیشین

درک تفاوت‌های فردی در بحث شبکه‌های اجتماعی، موضوعی مهم در روانشناسی است. هر چند که نظریه‌های مفیدی وجود دارد که این تفاوت‌ها را بیان می‌کنند، اما به نظر می‌رسد که این نظریه‌ها در حوزه‌های مختلف پژوهش متفاوت هستند. از طرف

^۲ Personality-based Product Recommender

^۱ Support Vector Machine

دیوار کاربر ارسال گردد و در صورت نیاز به جمع‌آوری داده‌های جدید، دسترسی به اطلاعات کاربر حتی زمانی که کاربر به شبکه متصل نیست انجام شود و ایمیل به کاربر ارسال گردد. هنگامی که کاربر آزمون را تمام می‌کند و قبل از ارسال نتایج، برنامه پیشنهاد می‌کند که کاربر دوستان خود را برای پیوستن به این کمپین دعوت کند و در نهایت، کاربر باید تنظیمات حریم خصوصی خود را پیکربندی کند.

ابعاد شخصیت افراد را نیز می‌توان با بررسی تعامل کاربران در شبکه‌های اجتماعی برخط بررسی کرد. سه مدل مشهور از شخصیت‌شناسی شامل، مدل سه‌بعدی عامل آیزنک (که به عنوان مدل PEN^۳ معروف است و شامل روان‌گرایی، برون‌گرایی و روان‌رنجوری است)، مدل پنجگانه بزرگ^۴، و مدل پنجگانه جایگزین^۵ می‌باشند. سوری و همکاران در [۱] برای تحلیل شخصیت افراد از روش هوشمند ترکیبی AdaBoost^۶ استفاده کردند. آنها پس از تنظیم پارامترها برای هر یک از متغیرهای هدف، به منظور پیدا کردن طبقه‌بندی مناسب برای هر پنج عامل، روش‌های طبقه‌بندی مختلف را برای هر پنج شخصیت مختلف بر روی داده‌های آزمایش مشابه اجرا نمودند. برای یافتن شخصیت فرد در مدل بزرگ پنجگانه، ابتدا متغیرهای پروفایل فیس‌بوک کاربر وارد مدل شده تا بتوان بر اساس آنها مدل را آموزش داد. از سوی دیگر، فرد پرسشنامه NEO^۷ (پنج شخصیت) را پر کرده و برای هر عامل شخصیت، بر اساس نمره به دست آمده از پرسشنامه شخصیت، برحسب اختصاص یافته است که به عنوان متغیر هدف وارد مدل می‌شود. حال با کمک مدل‌سازی و طبقه‌بندی افراد، می‌توان شخصیت کاربران را در شبکه‌های اجتماعی بدون هیچ‌گونه سابقه‌ای از آنها و یا حتی پر کردن سوالات روانشناختی پیش‌بینی نمود.

وی و همکاران در [۸] یک چارچوب گروه اطلاعاتی ناهمگن (HIE)^۸ پیشنهاد می‌کنند تا ویژگی‌های شخصیت کاربران را با

مختلف یادگیری ماشین، با هدف جستجوی الگوهای تعاملی که اطلاعاتی در مورد ویژگی‌های شخصیتی کاربران فراهم می‌کند، آموزش داده شدند. این طبقه‌بندی‌کننده‌ها می‌توانند شخصیت کاربر را از پارامترهای مربوط به تعاملات کاربر همانند تعداد دوستان یا تعداد پست‌های دیواری پیش‌بینی کنند. نتایج نشان می‌دهد طبقه‌بندی‌کننده‌ها از سطح بالایی از دقت برخوردار هستند و این روش پیشنهادی را به روشی مطمئن برای پیش‌بینی شخصیت کاربر تبدیل می‌کند.

در مقاله [۷] نیز از مدل «پنجگانه بزرگ» ابعاد شخصیت استفاده شده است و پیش‌بینی شخصیت به وسیله برنامه توییت‌ر پیگیری شده است. ابتدا یک برنامه توییت‌ر ایجاد شده و سپس تعداد ۴۵ سوال به کاربران داده و توییت‌های ایشان جمع‌آوری شده است. تعداد طرفداران، تراکم شبکه اجتماعی، تعداد اشاره‌ها^۱، تعداد پاسخ‌ها و هشتک‌ها و پیوندها و ... عناوینی هستند که جمع‌آوری شده و مورد بررسی قرار گرفته‌اند. تجزیه و تحلیل این متن‌ها با تحلیل رگرسیون و در نرم‌افزار Weka، به پیش‌بینی شخصیت منجر می‌شود.

آدامپولوس و همکاران [۴]، برای نخستین بار به بررسی تاثیر صفات شخصیتی و شباهت شخصیت پنهان در بازاریابی دهان به دهان (WOM)^۲ پرداخته و نشان دادند ویژگی‌های شخصیت پنهان از کاربران برخط، اثربخشی WOM در سیستم‌عامل‌های رسانه‌های اجتماعی را تشدید می‌کند. در واقع آنها نشان دادند WOM نقش مهمی در شکل دادن رفتار و ترجیحات مصرف‌کنندگان ایفا می‌کند.

همانطور که بیان شد، فرض اصلی این است که از کاربران با شخصیت مشابه انتظار می‌رود که الگوهای رفتاری معمول را هنگام برقراری ارتباط از طریق شبکه‌های اجتماعی مجازی نشان دهند و این الگوها را می‌توان به منظور پیش‌بینی گرایش شخصیت کاربر استفاده کرد. گولبک و همکاران [۳]، در شبکه اجتماعی فیس‌بوک برای هر شخص یک پرسشنامه را ارسال می‌کنند و از وی می‌خواهند تا اجازه دهد نتایج آزمون روی

³ Psychoticism, Extraversion, and Neuroticism

⁴ Big Five

⁵ Alternative Five

⁶ Adaptive Boosting

⁷ Neuroticism (N) Extraversion (E) Openness (O)

⁸ Heterogeneous Information Ensemble framework

¹ Mentions

² Word of mouth

هر ویژگی شخصیتی مدل پنجگانه بزرگ استفاده کردند. یافته‌های آنها ارتباط معنی‌داری بین وضعیت شخصیت افراد و قصد خرابکاری آنها را نشان می‌دهد. این نتیجه می‌تواند شاخص خوبی برای حساسیت کاربران برخط به خشونت سایبری باشد.

مورنو-آرمنداریز و همکاران در [۱۱] مدلی مبتنی بر شبکه‌های عصبی عمیق را برای پیش‌بینی شخصیت ظاهری ارائه داده‌اند. این مدل می‌تواند ویژگی‌های شخصیتی را با استفاده از مدل پنج عاملی (پنج بزرگ) از یک تصویر پرتره، مشخص کند. در این روش، مجموعه جدیدی با ویژگی‌های شخصیتی مرتبط با آنها از یک منبع موجود فیلم با برچسب مقایسه دوتایی برای اطمینان از سازگاری استخراج شده است. آنها چندین مدل با استفاده از شبکه‌های عصبی برای استخراج خودکار ویژگی‌هایی از یک پرتره که شاخص‌های ویژگی‌های شخصیتی هستند، پیشنهاد کرده‌اند. این مدل‌ها خصوصیات را برای هر یک از عوامل پنج بزرگ به یک کلاس باینری طبقه‌بندی می‌کنند که شامل باز بودن به تجربه (O)، وظیفه‌شناسی (C)، برون‌گرایی (E)، موافقت (A) و روان‌رنجوری (N) هستند. علاوه بر این، روش آنها با کدگذاری ویژگی و آموزش یادگیری برای غنی‌سازی نمایش تصاویر با پرتره‌های بدون برچسب اضافی آزمایش شده و به دقت ۶۵/۸۶٪ دست یافته است. به طور متوسط دقت پنج عامل به صورت $O=0.61/48$ ، $C=0.69/56$ ، $E=0.73/23$ ، $A=0.60/68$ و $N=0.64/35$ حاصل شده است. در مقایسه با قضاوت انسان (با میانگین دقت ۵۶/۶۶٪)، مدل در چهار عامل از پنج عامل، عملکرد بالاتر و دقت بالاتری به دست آورده است.

اتکین و همکاران [۱۲]، به مدل پنج عاملی شخصیت و ساختار علائم روانشناختی در نوجوانان پرداخته‌اند. این مطالعه با هدف کشف ساختار فاکتورهای شایع‌ترین علائم روانشناختی در نوجوانی و کشف ارتباطات بین عوامل روانی آسیب شناختی با

ادغام اطلاعات ناهمگن از جمله استفاده از زبان آنها، نماد، شکلک، آواتار و الگوهای واکنش و همچنین طرفداران و توییت‌های منتشر شده نشان دهد. این مقاله تقسیم‌بندی سه طبقه را برای طبقه‌بندی مثبت، عصبی و منفی برای هر بعد در پنج شخصیت پذیرفته است و توسط چند فرمول، کاربران مثبت را از کاربران منفی تشخیص داده و کاربران عصبی را نادیده می‌گیرد.

گو و همکاران در [۹] یک مدل شخصیت شخصی را پیشنهاد می‌دهند که متشکل از شخصیت، ویژگی‌های تاثیر، رفتار و شناخت و روابط بین این ویژگی‌ها است.

سنگوپتا و همکاران در [۱۰] یک سیستم مقیاس‌پذیر در زمان واقعی را برای تجزیه و تحلیل احساسات داده‌های تویتر ارائه می‌دهند. تحقیق آنها توییت‌های کاربران را در زمان واقعی جمع‌آوری نموده و بنابراین مبنایی برای شناسایی هر توییت به صورت مثبت یا منفی بر اساس ذهنیت کاربر را فراهم کرده و بدین ترتیب تجزیه و تحلیل زمان واقعی کاربران در مورد یک موضوع خاص را ارائه می‌دهد. طراحی و پیاده‌سازی کار آنها در Flask و Celery انجام شده^۱ که شامل وظایف استخراج و طبقه‌بندی است. این سیستم با توجه به میزان و اندازه داده‌های ورودی مقیاس‌پذیر بوده و مزایای آن از نظر مقیاس‌پذیری، عملکرد و دقت طبقه‌بندی به صورت تجربی هم مورد ارزیابی قرار گرفته است.

خشونت سایبری یک مشکل جدی در تحقیقات بهداشت الکترونیکی است و تمرکز آن بر روی شناسایی رفتارهای مضر از محتوای برخط تولید شده توسط کاربر به منظور جلوگیری و محافظت از قربانیان است. زرنوف و اییک در [۵] نشان داده‌اند که پنج ویژگی شخصیتی چقدر با رفتار خشونت‌آمیز یک فرد می‌تواند ارتباط داشته باشد. آنها از مجموعه‌ای از الگوریتم‌های یادگیری ماشین با ویژگی‌های مربوط به واژگان استفاده شده در

² Openness

³ Conscientiousness

⁴ Extraversion

⁵ Agreeableness

⁶ Neuroticism

^۱ چارچوب Flask، یک چهارچوب نوشته شده با زبان پایتون برای برنامه‌نویسی وب و Celery چهارچوبی جهت اجرای برنامه‌ها به صورت موازی روی سرور می‌باشد و ترکیب این دو چارچوب به منظور پیاده‌سازی وبگاه‌های با تعداد کاربران موازی بالا استفاده می‌شود.

بیشتر به استراتژی‌های متمرکز بر احساسات، پذیرش و انکار، فعالیت در راهبردهای فعال و قاطع و جامعه‌پذیری به منظور دستیابی به حمایت و مشاوره عاطفی کمک می‌کند. علاوه بر این احساسات تکانشی به نوعی مصرف مواد را انکار می‌کند. در حالی که پرخاشگری-خصومت به تمرکز و تخلیه احساسات، برای جلوگیری از کنار آمدن با این مساله کمک می‌کند.

ویژگی‌های شخصیتی مدل پنج عامل و شکست‌های شناختی ذهنی موضوعی است که ستین و همکاران در [۱۶] به آن پرداخته‌اند. تحقیق آنها به چگونگی ارتباط صفات و جنبه‌های شخصیتی پنج عامل با شکست شناختی پرداخته و نشان می‌دهد روان‌رنجوری بالاتر با نارسایی‌های شناختی بیشتری همراه است در حالی که وظیفه‌شناسی و موافقت با عدم موفقیت‌های کمتر، کنترل ویژگی‌های شخصیتی را به دنبال دارد. آنها همچنین نشان دادند که تجزیه و تحلیل چهره، تصویر دقیق‌تری از چگونگی ارتباط صفات با شکست شناختی ارائه می‌دهد.

۳. شبکه عصبی عمیق

شبکه عصبی عمیق یا همان یادگیری عمیق، یک زیرشاخه از یادگیری ماشین است. شبکه عصبی عمیق یک ساختار سلسله مراتبی بر مبنای مجموعه‌ای از الگوریتم‌ها است که مفاهیم انتزاعی سطح بالا در داده‌های متفاوت را به صورت یک مدل ساده‌تر بیان می‌کند. این فرآیند مدل‌سازی با استفاده از یک گراف عمیق انجام می‌شود که دارای چند لایه پردازشی است که هر لایه از چندین لایه تبدیلات خطی و غیرخطی تشکیل شده است و مدل‌سازی مفاهیم انتزاعی مذکور با استفاده از این لایه‌ها انجام می‌گردد. به زبان دیگر، پایه شبکه عصبی عمیق، یادگیری نمایش دانش و ویژگی‌ها در لایه‌های مدل است. یک نمونه آموزشی می‌تواند به گونه‌های متفاوت مدل‌سازی شود، به عنوان مثال، می‌توان به یک بردار ریاضی پر شده از مقادیر به ازای هر پیکسل و در دید کلی‌تر به شکل یک مجموعه از زیرشکل‌های کوچک‌تر اشاره کرد. برخی از این روش‌های مدل‌سازی سبب ساده شدن فرآیند یادگیری ماشین می‌شوند [۱۷]. در یادگیری شبکه عصبی عمیق امید به جایگزینی استخراج این ویژگی‌های

مدل پنج عاملی شخصیت و عامل عمومی شخصیت (GFP)^۱ انجام شده است. نمونه‌ای از ۸۳۵ نوجوان گزارش‌های شخصی و آسیب‌شناسی روانی را تکمیل کردند. نتایج این مطالعه نشان می‌دهد که روش پیشنهادی دوعاملی، به اندازه کافی ساختار آسیب‌شناسی روانی را در نوجوانی به تصویر می‌کشد. این ساختار توسط ارتباطات صفات شخصیتی با هر عامل ایجاد شده پشتیبانی می‌شود.

طی سال‌های اخیر، مطالعات متعددی رابطه بین آسیب‌شناسی استفاده از اینترنت (PIU)^۲ و پنج ویژگی بزرگ شخصیتی را بررسی کرده است. یک از این مطالعات مربوط به آلونسو و رومرو [۱۳] است. نتایج نشان می‌دهد که این تحقیق می‌تواند سطح بالایی از روان‌رنجوری و برون‌گرایی و سطح پایین موافقت، گشودگی و وظیفه‌شناسی PIU را پیش‌بینی کند. این مطالعه همچنین تعیین جنبه‌های خاص پنج بزرگ را که با افزایش خطر PIU مرتبط هستند، امکان‌پذیر می‌کند. نتایج حاکی از دخیل بودن مکانیسم‌های شخصی متفاوت در ابعاد مختلف تشکیل PIU است.

دهه‌ها است که محققان به تفاوت‌های جنسیتی بین فرهنگی در شخصیت علاقه‌مند بوده‌اند. مورفی و همکاران [۱۴]، به مقایسه بین‌المللی تفاوت‌های جنسیتی در مدل پنج عاملی شخصیت (تحقیقی در ۱۰۵ کشور) پرداخته‌اند. این تحقیق در مورد مدل پنج‌گانه شخصیت بزرگ (FFM)^۳ متمرکز بر تخمین تفاوت بین زن و مرد در ابعاد مختلف شخصیت پرداخته است و با استفاده از یک نمونه بزرگ بین‌المللی از داده‌های شخصیتی و روش‌های پیشرفته تحلیلی، تخمین‌های دقیقی از تفاوت‌های جنسیتی بین کشورهای مختلف در شخصیت را نشان می‌دهد.

گوما و همکاران [۱۵]، با هدف تعیین ابعاد شخصیتی که توسط پرسشنامه شخصیت زاکرمن-کوهلمن اندازه‌گیری شده است، بهترین راهکار مقابله‌ای را توضیح داده‌اند. علاوه بر این، برای روشن شدن موقعیت این استراتژی‌ها در چارچوب گسترده‌تر ساختار شخصیت به تحقیق پرداخته‌اند. روان‌رنجوری-اضطراب

¹ The General Factor of Personality

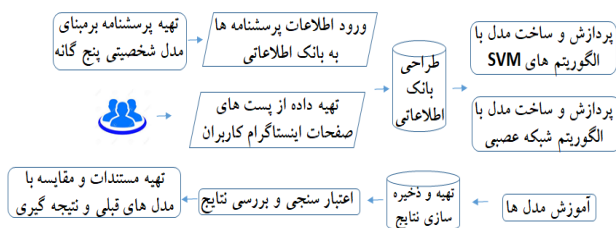
² Pathological Internet use

³ The five-factor model of personality

زمینه شخصیت‌شناسی از دقت بالاتری برخوردار بوده [۹] و کاربرد وسیع‌تری در این حوزه داشته است [۱۰]، [۱۱].

۴. مدل پیشنهادی

مدل‌های شخصیت‌شناسی مختلفی توسط روانشناسان پیشنهاد شده است که در این تحقیق با توجه به استفاده گسترده از مدل شخصیت‌شناسی پنج‌عاملی، این مدل استفاده شده است [۳]. برای شناخت شخصیت افراد بر اساس این مدل هر فرد باید پرسشنامه استاندارد NEO (پنج شخصیت) را پر کند. در این پرسشنامه پنج عامل اصلی شخصیت شامل روان‌نژندی (پایداری هیجانی)، برون‌گرایی یا درون‌گرایی، اشتیاق به تجارب تازه، توافق‌پذیری (سازگاری) و مسئولیت‌پذیری مورد تحقیق و بررسی قرار می‌گیرند. در پرسشنامه مذکور از فرد درخواست می‌شود تا سن، جنسیت و نام کاربری اینستاگرام و مدت زمان حضور در این برنامه را یادداشت نماید. سپس توسط حساب کاربری ایجاد شده در اینستاگرام، تمام حساب‌های کاربری موجود در پرسشنامه‌ها دنبال می‌شود تا بتوان به صفحه شخصی وی جهت کاوش دسترسی پیدا کرد. پرسشنامه شخصیت‌شناسی کلید استاندارد دارد که با استفاده از آن می‌توان پاسخ هر پرسشنامه را تحلیل کرد و شخصیت هر فرد را براساس پنج عامل بررسی نمود. روند کلی روش پیشنهادی در شکل (۱) ذکر شده است.



شکل (۱): مدل روش پیشنهادی

در این تحقیق از یک طرف از طریق تحلیل کلید پرسشنامه استاندارد شخصیت، شخصیت هر فرد تحلیل شده و از طرف دیگر برای هر فرد با توجه به حساب اینستاگرام وی کلیه متونی که به اشتراک گذاشته است، استخراج می‌گردد. هدف آن است که بینیم بین شخصیت وی و متونی که به اشتراک می‌گذارد چه

تصویر به دست بشر با روش‌های کاملاً خودکار بدون نظارت و نیمه‌نظارتی وجود دارد. این مدل‌ها نظیر شبکه عصبی عمیق، شبکه عصبی پیچیده، شبکه باور عمیق پیشرفت‌های خوبی را در حوزه‌های پردازش زبان‌های طبیعی و پردازش تصویر ایجاد کرده‌اند [۱۸].

یکی از روش‌هایی که در حال حاضر به صورت گسترده و بسیار کاربردی برای حل مساله‌های دسته‌بندی^۱ مورد استفاده قرار می‌گیرد، روش ماشین بردار پشتیبان (SVM) است [۱۹]. شاید به گونه‌ای بتوان محبوبیت و کاربردهای بسیار وسیع کنونی روش ماشین بردار پشتیبان را با محبوبیت شبکه‌های عصبی در دهه گذشته مقایسه کرد. مبنای کار SVM دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کند خطی را انتخاب کند که حاشیه اطمینان بیشتری داشته باشد. حل معادله پیدا کردن خط بهینه برای داده‌ها به وسیله روش‌های QP^۲ انجام می‌شود که این روش‌ها شناخته شده در حل مساله‌های محدودیت‌دار هستند [۲۰]. این روش‌ها از نوع Two-norm و با حاشیه نرم می‌باشند. از آنجا که لاگرانژ Two-norm تابع جریمه سخت‌گیرانه‌تری دارد، می‌تواند دقت بالاتری فراهم کند. ماشین بردار پشتیبانی یکی از روش‌هایی یادگیری بانظارت^۳ است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند.

الگوریتم آموزش SVM مدلی را ایجاد می‌کند که نمونه‌های جدید را به یک دسته یا دسته دیگر اختصاص می‌دهد و آن را به یک طبقه‌بندی کننده خطی باینری غیرممکن تبدیل می‌کند. یک مدل SVM نشان‌دهنده نمونه‌ها به عنوان نقاط نقشه‌برداری شده در فضا است به گونه‌ای که نمونه‌های دسته‌های جداگانه با یک شکاف روشن (که تا حد ممکن گسترده است)، تقسیم می‌شوند. نمونه‌های جدید سپس در همان فضا ترسیم شده و متعلق آنها به یک دسته بر اساس شکاف بین آنها پیش‌بینی می‌شود [۲۱].

هرچند استفاده از شبکه عصبی عمیق در سال‌های اخیر مورد توجه قرار گرفته است [۱۱]، اما یادگیری SVM در مقایسه با سایر الگوریتم‌های یادگیری ماشین طبق تحقیقات پیشین در

^۱ Classification

^۲ Quadratic Programming

^۳ Supervised Learning

عبارات، استخراج ریشه کلمات و حذف ایست واژه‌ها (Stop Words) روی داده‌های ورودی اعمال می‌شود. در این فرآیند از SQL Server به عنوان پایگاه داده استفاده شده و عملیات پیش‌پردازش توسط برنامه‌ای به زبان پایتون انجام شده است. در قدم بعد تعدادی از بیشترین کلمات پرتکرار هر کاربر را که در بانک موجود است، پیدا کرده و تعداد تکرار آن را محاسبه می‌نماییم. به این ترتیب هر کلمه پرکاربرد تبدیل به یک عدد می‌شود. برای این کار، از معیار TF-IDF¹ جهت تشخیص کلمات پرکاربرد استفاده شده و تعداد کلمات برای هر شخص به ۲۵۰ کلمه محدود شده است. همانطور که در شکل (۲) نشان داده شده است، این تعداد تکرار در فیلد Features قرار می‌گیرد. در قدم بعد با توجه به شخصیت هر کاربر و داشتن کلمات پرکاربرد وی می‌توانیم به ماشین آموزش دهیم، به این صورت که برای آموزش شبکه عصبی و الگوریتم SVM، با توجه به ویژگی‌ها که همان کلمات پرکاربرد کاربر هستند و شخصیت تحلیل شده کاربر که همان برجسب و متغیر Target می‌باشد، یادگیری انجام می‌شود. به هر حال، در این تحقیق از هر دو روش یادگیری ماشین استفاده شده که نتایج پیش‌بینی در ستون Predict جهت محاسبه معیارهای ارزیابی قرار گرفته است. با توجه به کاربرد زیاد روش SVM در این پژوهش، از هسته‌های مختلف این روش استفاده شده و نتایج آنها با هم مقایسه شده است تا بتواند برای سایر پژوهشگران مفید باشد. استفاده از یادگیری ماشین با روش‌های فوق در زبان فارسی برای شخصیت‌شناسی انجام نشده و پژوهش جاری می‌تواند برای محققین این حوزه مورد توجه باشد. با توجه به الگوریتم SVM این تحقیق می‌تواند نشان دهد که انتخاب هسته مناسب و متناسب حوزه کاربرد در این الگوریتم می‌تواند در بهبود نتایج موثر باشد. در این تحقیق و طبق شکل (۱)، پس از پیش‌پردازش داده‌ها و ذخیره‌سازی آنها، دو مدل یادگیری ماشین با SVM و شبکه عصبی عمیق ساخته می‌شود و یادگیری و آزمون مدل‌ها انجام می‌پذیرد. سپس نتایج با معیارهای اعتبارسنجی بررسی شده و دقت و صحت یادگیری مشخص می‌شود.

ارتباطی است و آیا می‌توان با تحلیل متونی که هر فرد در شبکه اجتماعی به اشتراک می‌گذارد پی به شخصیت وی برد. این کار برای اولین بار در زبان فارسی انجام شده است که می‌تواند نوآوری این تحقیق باشد و جمع‌آوری اطلاعات بسیار زمان‌بر و هزینه‌بر بوده است. جامعه پژوهشی تعداد ۲۵۰ نفر از دانشجویان دانشگاه‌های استان قم و واحدهای مشاوره‌ای سطح استان قم بودند. پرسشنامه مذکور توسط دانشجویان و مراجعان این واحدهای مشاوره‌ای و تحقیقاتی پر شده است، به گونه‌ای که تفکیکی بین دانشجویان و مراجعان مراکز مشاوره وجود نداشته و همگی در یک سطح قرار دارند. از میان پرسشنامه‌های پر شده تعداد ۷۴ پرسشنامه، به علت ناقص بودن شامل عدم ذکر حساب کاربری، پر نکردن بیش از ۱۵ سوال پرسشنامه و یا مخدوش بودن، حذف شدند. بدین ترتیب تعداد ۱۷۶ پرسشنامه که به طور کامل و صحیح تکمیل شده است، مورد قبول واقع گردید تا در مراحل بعدی مورد بررسی بیشتر قرار گیرند. تعداد کل کاربرانی که در برگه پرسشنامه خود عنوان حساب کاربری خود را ذکر کرده بودند، ۱۳۵ نفر، تعداد کل پست‌های اینستاگرام دنبال شده توسط حساب‌های کاربری، ۳۲۱۳ عدد و همچنین تعداد پست‌های مرتبط با کاربران ۱۶۳۵ پست بود. پست‌های منتشر شده توسط کاربران، به وسیله برنامه‌ای با زبان پایتون استخراج شده (خزشگر) و در پایگاه داده ذخیره گردید. این داده‌ها مجموعه‌ای از کلمات هستند که در برنامه مورد نظر به صورت کلمه به کلمه درآمده‌اند که می‌بایست از نظر ساختاری و محتوایی پالایش شوند. بدین منظور این کلمات توسط پایگاه داده مجزایی از نظر کلمات اضافی و افعال و حروف ربط غربال شده و کلمات اضافی آن حذف گردید.

برای رسیدن به هدف تحقیق نیاز به یادگیری ماشین وجود دارد. شخصیت هر فرد با توجه به مدل پنج عاملی، به عنوان برجسب هر رکورد در نظر گرفته شده و هر کدام از عباراتی که در شبکه اینستاگرام به اشتراک گذاشته شده است، یک ویژگی در آن رکورد به حساب می‌آید. پیش‌پردازش متن، تعداد عبارات (که همان ویژگی‌ها هستند) را به حداقل رسانده و متن را پاکسازی می‌کند. لذا پیش‌پردازش‌های استخراج کلمات، نرمال کردن

¹ Term Frequency-Inverse Document Frequency

گرفته شده است که اگر این مقدار کمتر از ۲۵۰ کلمه باشد هیچ کلمه‌ای انتخاب نمی‌شود و در صورتی که تعداد کلمات درج شده بیشتر از ۲۵۰ باشد، حداکثر ۲۵۰ کلمه ابتدایی انتخاب خواهند شد. حداکثر اندازه واژگان نیز در این مدل ۲۰ کاراکتر در نظر گرفته شده است. در لایه اول امبدینگ مقادیر حداکثر ماتریس، اندازه واژگان و اندازه ورودی دریافت و خروجی آن یک ماتریس 20×250 جهت انتخاب ویژگی‌ها خواهد بود.

در لایه دوم ماتریس سه بعدی یک فیلتر با چهار متغیر اندازه واژگان، اندازه هسته‌ها، فاصله اعمال فیلتر برابر از بالا و پایین و چپ و راست و انتخاب نوع تابع فعال‌ساز (relu) می‌باشد. از آنجا که اندازه فیلتر انتخابی ترکیب ماتریس در این مدل ۲۰ در نظر گرفته شده است، اندازه فیلتر نیز 20×20 تنظیم شده است. خروجی لایه دوم ماتریس دوبعدی 20×250 شامل مجموعه‌ای از ماتریس‌های ترکیب شده 1×1 به ازای ماتریس‌های 20 در 20 خواهد بود. در لایه سوم نرمال‌سازی مقادیر ماتریس لایه دوم با اندازه ۲ انجام می‌شود. که خروجی آن ماتریس دوبعدی نرمال شده با اندازه 20×125 از ماتریس ورودی لایه دوم خواهد بود. در لایه چهارم ماتریس دوبعدی لایه سوم به عنوان ورودی دریافت و خروجی آن یک ماتریس تک بعدی یا تخت با اندازه 250×1 خواهد بود. در لایه پنجم دو پارامتر اندازه خروجی با مقدار 560 و تابع فعال‌سازی با مقدار sigmoid بوده که خروجی آن یک ماتریس تک‌بعدی با اندازه 560 خواهد بود. لایه ششم تکرار لایه پنجم با اندازه 15 خواهد بود. این 15 خروجی همان برچسب‌ها می‌باشند که شامل 5 عامل شخصیتی است که هر کدام سه درجه شدت دارند که در مجموع 15 حالت مختلف شخصیتی را می‌سازند. در انتها مقادیر و تنظیمات فوق با متغیرهای نوع محاسبه خطا با مقدار دودویی کراس انتروپی و بهینه‌سازی با مقدار Adam^۳ و معیار خروجی دقت پیش‌بینی با مقدار Accuracy کامپایل می‌شود. خروجی همه مراحل و

است. منظور از مقدار واقعی، مقدار استاندارد یا رفرنسی که ما به آن اعتماد داریم [۲۳]. پارامتر مهم دیگری به نام «معیار اف» وجود دارد که برای ارزیابی عملکرد دسته‌ها بسیار مورد استفاده قرار می‌گیرد و از ترکیب دو پارامتر دقت و بازخوانی به دست می‌آید و به صورت زیر تعریف می‌شود:

$$Fscore = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

۵.۲. تجزیه و تحلیل یافته‌ها

اطلاعاتی که در مرحله یادگیری مدل تولید می‌شود، می‌بایست در مرحله ارزیابی مورد تحلیل قرار گیرد تا بتوان ارزش آن را مشخص کرد و به دنبال آن میزان کارآمد بودن الگوریتم مدل را نیز تعیین نمود. این معیارها را می‌توان برای داده‌های آموزشی در مرحله یادگیری و هم برای رکوردهای آزمایشی در مرحله ارزیابی محاسبه کرد. تعداد کاربران با ویژگی «باوجدان بودن» ۱۷۴ نفر می‌باشند و ۲ نفر دارای این ویژگی نیستند. همچنین تعداد کاربران با ویژگی «روان‌رنجوری» ۱۷ نفر هستند و ۱۵۹ نفر عاری از این خصیصه می‌باشند. تمامی کاربرانی که به پرسشنامه مذکور پاسخ داده‌اند دارای ویژگی «پذیرا بودن در تجربه» هستند. ویژگی «پاسخ‌گو بودن» صفتی است که بر مبنای اطلاعات حاصل از این پرسشنامه، ۱۷۴ نفر دارای آن خصیصه هستند و تنها ۲ نفر دارای این ویژگی نیستند. در مورد ویژگی «درون‌گرایی - برون‌گرایی»، ۱۸ نفر از افرادی که پرسشنامه را پر نموده‌اند افراد درون‌گرا هستند و مابقی افراد برون‌گرا می‌باشند. در این تحقیق از الگوریتم‌های SVM با هسته‌های مختلف و شبکه عمیق استفاده شده است. شبکه عمیق در این تحقیق ۵ لایه دارد. مقدار حداکثر ستون‌های لایه اول ۱۷۵۰۰ در نظر گرفته شده است که شبکه‌ای شامل یک ماتریس بزرگ دو بعدی می‌باشد. این عدد حاصل ضرب سطر و ستون‌های کلمات انتخابی و مقادیر مرتبط با آنها بوده و به طور کلی محدودیت کلمات انتخابی را تعیین و از سرریز شدن مقادیر ماتریس جلوگیری می‌کند. حداکثر تعداد کلمات اخذ شده از هر کاربر ۲۵۰ در نظر

^۲ The Rectified Linear Activation Function

^۳ Adam یک الگوریتم بهینه‌سازی است که می‌تواند به جای روش کلاسیک گرادینان نزولی تصادفی برای به‌روزرسانی وزن‌های شبکه بر اساس داده‌های آموزشی استفاده شود.

^۱ F-Score

کامپایل فوق ۱۵ شاخص شخصیت خواهد بود نتایج ارزیابی روش ها روی آنها در جدول (۱) نمایش داده شده است. همانطور که نتایج نشان می‌دهند، الگوریتم SVM با هسته شعاعی و چندجمله‌ای دقت و بازخوانی بالاتری نسبت به سایر الگوریتم‌های مورد مقایسه دارد.

جدول (۱): مقایسه نتایج

عنوان آزمون	۱	۲	۳	۴
SVM با هسته سیگموئید	۰/۸۴۰۰	۰/۸۲۳۵	۰/۸۵۷۱	۰/۷۲۴۱
SVM و در هسته چند جمله ای	۰/۹۹۰۱	۰/۹۸۰۴	۰/۹۸۹۸	۰/۹۸۰۴
SVM و هسته خطی	۰/۹۸۰۰	۰/۹۶۰۸	۰/۹۸۹۹	۰/۹۶۰۸
SVM و در پایه تابعی شعاعی	۰/۹۹۰۱	۰/۹۸۰۴	۰/۹۹۹۸	۰/۹۸۰۴
شبکه عصبی عمیق	۰/۸۵۷۱	۰/۸۲۳۵	۰/۸۹۳۶	۰/۷۵۰۰

۶. نتیجه‌گیری و کارهای آینده

صفات شخصیتی به عنوان صفاتی درونی، پایدار، سلسله مراتبی ساختار یافته اساسی تنظیم شده توسط عوامل بیولوژیکی مانند ژن‌ها و ساختار مغز تعریف می‌شود. این صفات در تمام طول عمر و از طریق موقعیت‌های مختلف کاملاً پایدار باقی می‌ماند و به همین دلیل شخصیت کاربر یک نقطه شروع خوب برای پیش‌بینی رفتار وی است. شبکه‌های اجتماعی مکانی هستند که کاربران در آن با به اشتراک‌گذاری علاقه‌مندی‌های خود و اعلام نظرات در مورد مساله‌های مختلف، خود را به دیگران معرفی می‌کنند و جزئیات شخصیت و بینش خود را در زندگی عرضه می‌کنند. کاوش این مطالب و تجزیه و تحلیل آنها می‌تواند اطلاعات خوبی از شخصیت فرد به ما بدهد. برای این منظور، در این تحقیق از شبکه اجتماعی اینستاگرام و ۱۶۳۵ پست متنی مربوط به ۱۳۵ نفر استفاده شده است. این افراد پیش از این پرسشنامه شخصیت‌شناسی NEO را پر کرده‌اند و شخصیت آنها

مراجع

تحلیل شده است. بنابراین در این تحقیق داده‌های لازم جهت یادگیری ماشین در مرحله اول جمع‌آوری و پیش‌پردازش شده است و در مرحله بعد با استفاده از الگوریتم‌های یادگیری ماشین SVM (با هسته‌های مختلف) و شبکه عصبی عمیق، مدل یادگیری ماشین برای این مجموعه داده ایجاد شد. ارزیابی مدل ایجاد شده نشان داد، روش پیشنهادی می‌تواند با دقت بالای ۷۵ درصد شخصیت افراد را بر اساس پست‌هایی که در شبکه‌های اجتماعی به اشتراک می‌گذارند، پیش‌بینی کند. این تحقیق می‌تواند برای متخصصین حوزه روانشناسی که می‌خواهند شخصیت افراد را به صورت غیرمستقیم شناسایی کنند، مفید باشد. با توجه به اینکه این تحقیق با پست‌های زبان فارسی انجام شده است، مجموعه داده فراهم شده می‌تواند برای سایر محققین حوزه داده کاوی مورد استفاده قرار گیرد. برای کارهای آینده پیشنهاد می‌شود که از تعداد افراد بیشتر و پست‌های بیشتری استفاده شود و همچنین اطلاعاتی مثل میزان لایک کردن‌ها، تصاویر و فیلم‌های اشتراکی نیز مورد تحلیل قرار گیرد. به عنوان کار آینده دیگر، استفاده از طیف مختلف شخصیتی مثل استفاده از شخصیت‌های خاص و افرادی که دارای زمینه درون‌گرایی شدید یا اختلالات شخصیتی دوقطبی یا افرادی که دارای حالت روحی ناپایدار هستند، نیز می‌تواند با نظر روانشناسان به مجموعه داده‌ها اضافه شود تا تنوع داده‌ها افزایش یافته و میزان دقت مدل یادگیری ماشین بالا رود. همچنین استفاده از مدل‌های شخصیت‌شناسی جدیدتر می‌تواند بر میزان موثر بودن شخصیت‌شناسی بیفزاید.

تعارض منافع: نویسندگان اعلام می‌کنند که هیچ تعارض منافی ندارند.

[1] A. Sourı, S. Hosseınpour, and A.M. Rahmani, "Personality classification based on profiles of social networks' users and the five-factor model of personality," Hum. Centric Comput. Inf. Sci., vol. 8,

p. 24, 2018, doi: 10.1186/s13673-018-0147-4.

[2] A. Ortigosa, R.M. Carro, and J.I. Quiroga, "Predicting user personality by mining social interactions in Facebook," J. Comput. Syst. Sci., vol. 80, no. 1, pp.

- 57-71, 2014, doi: 10.1016/j.jcss.2013.03.008.
- [3] J. Golbeck, C. Robles, and K. Turner, "Predicting personality with social media", In *Int. Conf. Hum. Factors Comput. Syst. (CHI), Extended Abstracts Volume*, Vancouver, BC, Canada, 2011, pp. 253-262, doi: 10.1145/1979742.1979614.
- [4] P. Adamopoulos, A. Ghose, and V. Todri, "The Impact of User Personality Traits on Word of Mouth: Text-Mining Social Media Platforms," *Inf. Syst. Res.*, vol. 29, no. 3, pp. 612-640, 2018, doi: 10.1287/isre.2017.0768.
- [5] R. Zarnoufi and M. Abik, "Big Five Personality Traits and Ensemble Machine Learning to Detect Cyber-Violence in Social Media," in *Learning and Analytics in Intelligent Systems*, vol. 7. Springer, Cham., 2020, doi: 10.1007/978-3-030-36778-7_21.
- [6] R. Buettner, "Predicting user behavior in electronic markets based on personality-mining in large online social networks - A personality-based product recommender framework," *Electron. Mark.*, vol. 27, no. 3, pp. 247-265, 2017, doi: 10.1007/s12525-016-0228-z.
- [7] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting Personality from Twitter," in *3rd Int. Conf. Priv. Secur. Risk Trust, and 3rd Int. Conf. Soc. Comput.*, Boston, MA, USA, 2011, pp. 149-156, doi: 10.1109/PASSAT/SocialCom.2011.33.
- [8] H. Wei, F. Zhang, N.J. Yuan, C. Cao, H. Fu, X. Xie, and W.Y. Ma, "Beyond the words: Predicting user personality from heterogeneous information", in *Proc. 10th ACM Int. conf. Web Search Data Mining*, 2017, pp. 305-314, doi: 10.1145/3018661.301871.
- [9] A. Guo, J. Ma, S. Tan, and G. Sun, "From affect, behavior, and cognition to personality: an integrated personal character model for individual-like intelligent artifacts," *World Wide Web*, vol. 23, no. 2, pp. 1217-1239, 2020, doi: 10.1007/s11280-019-00713-w.
- [10] A. Sengupta and A. Ghosh, "Mining social network data for predictive personality modelling by employing machine learning techniques", in *Comput. Adv. Commun. Circuits Syst., Lecture Notes in Electrical Engineering*, vol 575, Springer, Singapore, 2020, doi: 10.1007/978-981-13-8687-9_11.
- [11] M.A. Moreno-Armendariz, C. A. Duchanoy Martínez, H. Calvo, and M. Moreno-Sotelo, "Estimation of Personality Traits From Portrait Pictures Using the Five-Factor Model," in *IEEE Access*, vol. 8, pp. 201649-201665, 2020, doi: 10.1109/ACCESS.2020.3034639.
- [12] P. Etkin, L. Mezquita, F.J. Lopez-Fernandez, G. Ortet, and M.I. Ibanez, "Five Factor model of personality and structure of psychopathological symptoms in adolescents," *Pers. Individ. Differ.*, vol. 163, no. 1, p. 110063, 2020, doi: 10.1016/j.paid.2020.110063.
- [13] C. Alonso and E. Romero, "Study of the Domains and Facets of the Five-Factor Model of Personality in Problematic Internet Use in Adolescents," *Int. J. Ment. Health Addiction*, vol. 18, pp. 293-304, 2020, doi: 10.1007/s11469-018-9960-2.
- [14] S.A. Murphy, P.A. Fisher, C. and Robie, "International comparison of gender differences in the five-factor model of personality: An investigation across 105 countries," *J. Res. Pers.*, vol. 90, p. 104047, 2021, doi: 10.1016/j.jrp.2020.104047.
- [15] M. Goma-i-Freixanet, Y.M. Ortega, and A. Arnau, "The location of coping strategies within the Alternative Five Factor Model of personality," *New Ideas Psychol.*, vol. 60, p. 100834, 2021, doi: 10.1016/j.newideapsych.2020.100834.
- [16] A.R. Sutin, D. Aschwanden, Y. Stephan, and A. Terracciano, "Five Factor Model personality traits and subjective cognitive failures," *Pers. Individ. Differ.*, vol. 155, p. 109741, 2020, doi: 10.1016/j.paid.2019.109741.
- [17] Z. Wang, V. Joo, C. Tong and D. Chan, "Issues of Social Data Analytics with a New Method for Sentiment Analysis of Social Media Data," in *IEEE 6th Int. Conf. Cloud Comput. Technol. Sci.*, Singapore, 2014, pp. 899-904, doi: 10.1109/CloudCom.2014.40.
- [18] A.K. Nassirtoussi, S. Aghabozorgi, T.Y. Wah, and D.C.L. Ngo, "Text mining for market prediction: A systematic review," *Expert Syst. Appl.*, vol. 41, no. 16, pp. 7653-7670, 2014, doi: 10.1016/j.eswa.2014.06.009.
- [19] S. Razzaghzadeh, P. Norouzi Kivi, and B. Panahi, "A hybrid algorithm based on Gossip architecture using SVM for task scheduling in cloud computing," *Soft Comput. J.*, vol. 9, no. 2, pp. 84-93, 2021, doi: 10.22052/scj.2021.242822.0 [In Persian].
- [20] Z. Roozbahani, M. Yari, and R. Ghiasi, "Developing a Filter-Wrapper Feature Selection Method and its Application in Dimension Reduction of Gen Expression," *Soft Comput. J.*, vol. 6, no. 2, pp. 48-59, 2018, dor: 20.1001.1.23223707.1396.6.2.4.8 [In Persian].
- [21] B.E. Boser, I.M. Guyon, and V.N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Ann. Worksh. Comput. Learn. Theor.*, Pittsburgh, PA, USA, 1992, pp. 144-152, doi: 10.1145/130385.130401.
- [22] A. Khosravi, H. Abdulmaleki, and M. Fayazi, "Predicting the academic status of admitted applicants based on educational and admission data using data mining techniques," *Soft Comput. J.*, vol. 9, no. 2, pp. 94-113, 2021, doi: 10.22052/scj.2021.242837.0 [In Persian].
- [23] J.C. Waters, "Accuracy and precision in quantitative fluorescence microscopy," *J. Cell. Biol.*, vol. 185, no. 7, pp. 1135-1148, 2009, doi: 10.1083/jcb.200903097.