

A Recommender System Based on the Analysis of Personality Traits in Telegram Social Network

Mohammad Javad Shayegan^{1*}, Mohadese Valizdeh²

^{1,2} Department of Computer Engineering, University of Science and Culture, Tehran, Iran

Received: 28 August 2023, Revised: 24 December 2023, Accepted: 26 December 2023

Paper type: Research

Abstract

Analysis of personality traits of individuals has always been one of the interesting research topics. In addition, achieving personality traits based on data obtained from individuals' behavior is a challenging issue. Most people spend most of their time on social media and may engage in behaviors that represent a character in cyberspace. There are many social networks today, one of which is the Telegram social network. Telegram also has a large audience in Iran and people use it to communicate, interact with others, educate, introduce products and so on. This research seeks to find out how a recommendation system can be built based on the personality traits of individuals. For this purpose, the personality of the users of a telegram group is identified using three algorithms, Cosine Similarity, MLP and Bayes, and finally, with the help of a recommending system, telegram channels tailored to each individual's personality are suggested to him. The research results show that this recommending system has attracted 65.42% of users' satisfaction.

Keywords: Users' Behavior, Recommender Systems, Social Networks, Telegram, Personality Analysis.

* Corresponding Author's email: shayegan@usc.ac.ir

یک سیستم توصیه‌گر بر اساس تحلیل ویژگی شخصیتی افراد در شبکه اجتماعی تلگرام

محمدجواد شایگان فرد^{۱*}، محدثه ولی‌زاده^۲

^۱ گروه مهندسی کامپیوتر، دانشگاه علم و فرهنگ، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۰۶/۰۶ تاریخ بازبینی: ۱۴۰۲/۱۰/۰۳ تاریخ پذیرش: ۱۴۰۲/۱۰/۰۵

نوع مقاله: پژوهشی

چکیده

تحلیل ویژگی‌های شخصیتی افراد همواره یکی از موضوعات جذاب پژوهشی بوده است. علاوه بر این، دستیابی به ویژگی‌های شخصیتی براساس داده‌هایی که از رفتار اشخاص به دست می‌آید، یک موضوع چالش برانگیز است. براساس پژوهش‌های انجام شده، اغلب مردم، بیشتر وقت خود را در شبکه‌های اجتماعی صرف می‌کنند و ممکن است در این شبکه‌های اجتماعی، رفتارهایی را از خود بروز دهند که نمایانگر یک شخصیت در فضای مجازی باشد. امروزه شبکه‌های اجتماعی بسیاری وجود دارند که یکی از آنها، شبکه اجتماعی تلگرام است. تلگرام در ایران نیز مخاطبان بسیاری دارد و افراد به منظور برقراری ارتباط، تعامل با دیگران، آموزش، معرفی محصولات و غیره از آن استفاده می‌کنند. این پژوهش به دنبال این موضوع هست که چگونه می‌توان یک سیستم توصیه‌گر را بر اساس ویژگی‌های شخصیتی افراد بنا نهاد. به این منظور، شخصیت کاربران یک گروه تلگرامی را با استفاده از سه الگوریتم Cosine Similarity، MLP و Bayes شناسایی شده و در نهایت با کمک یک سیستم توصیه‌گر، کانال‌های تلگرامی متناسب با شخصیت هر فرد، به او پیشنهاد می‌شود. نتایج حاصل از تحقیق نشان می‌دهد که این سیستم توصیه‌گر به طور میانگین ۶۵/۴۲ درصد رضایت کاربران را جلب کرده است.

کلیدواژگان: رفتار کاربران، سیستم‌های توصیه‌گر، شبکه‌های اجتماعی، تلگرام، تحلیل شخصیت.

* رایانامه نویسنده مسؤول: shayegan@usc.ac.ir

۱- مقدمه

رسانه‌های اجتماعی عبارتند از جایگاهی برای تعاملات اجتماعی میان مردم که در آن افراد اطلاعات و ایده‌هایی را در جوامع و شبکه‌های مجازی خلق می‌کنند، به اشتراک می‌گذارند و یا آنها را با یکدیگر تبادل می‌کنند. امروزه مردم بیشتر از پیش از رسانه‌های اجتماعی استفاده می‌کنند. اشتراک‌گذاری اطلاعات در شبکه‌های اجتماعی تا اندازه‌ای منعکس کننده شخصیت و رفتار کاربران می‌باشد [۱].

در ایران، شبکه‌های اجتماعی بسیاری در جذب مخاطب تاثیرگذار بوده اند که در این میان شبکه اجتماعی تلگرام بیشترین رشد را در جذب مخاطبان ایرانی داشته است. بر اساس آخرین آماري که توسط وزارت ارتباطات ایران منتشر شد، تلگرام بیش از ۴۵ میلیون کاربر فعال ایرانی دارد (از ۸۵ میلیون جمعیت ایران) که بالغ بر ۶۰ درصد از کاربران شبکه اجتماعی ایران را شامل می‌شود. همچنین مجموع تعداد پست‌های منتشر شده در تلگرام در یک روز به طور میانگین، ۲,۲۵۰,۰۰۰,۰۰۰ پست می‌باشد. تلگرام یک شبکه اجتماعی کاربردی است که علاوه بر امکان ارتباط peer to peer کاربران با یکدیگر، دو ویژگی کانال و گروه را نیز دارد. در کانال تلگرام، کاربرانی که در کانال عضو هستند، تنها یک مصرف کننده از محتوا هستند و فقط یک شخص (یا گاهی بیشتر) به نام ادمین کانال مسئولیت تهیه محتوا اعم از صوت، ویدیو، عکس، فیلم را دارد. اما در گروه‌های تلگرام، اعضای گروه با یکدیگر در تعامل و تبادل هستند و می‌توانند اطلاعات خود را در فضای گروه به اشتراک بگذارند. بنابراین به دلیل تبادل اطلاعات و عقاید اعضا در یک گروه تلگرامی، بستری جهت تحلیل و بررسی پیام کاربران گروه فراهم می‌شود که باعث می‌شود یک منبع داده‌ای غنی جهت تحلیل داده‌ها در این رسانه تولید شود [۲].

از طرف دیگر در سال‌های اخیر، تحلیل شخصیت و

رفتارشناسی در شبکه‌های اجتماعی مرکز توجه بسیاری از پژوهشگران بوده است. همچنین، با توجه به گستردگی شبکه‌های اجتماعی موبایلی به خصوص شبکه اجتماعی تلگرام، کاربران خیلی راحت‌تر از گذشته نظراتشان را در یک فضای تعاملی بیان می‌کنند که این امر می‌تواند زمینه‌ای برای تحلیل شخصیت کاربران از طریق پیام‌های تلگرامی باشد. در این تحقیق علاوه بر اینکه شخصیت افراد با توجه به پیام‌هایی که در گروه تلگرام درج می‌کنند مورد تجزیه و تحلیل قرار می‌گیرد، در مرحله بعد سیستم توصیه‌گری نیز توسعه می‌یابد تا کانال‌های مورد علاقه آن شخص را مطابق با تحلیل شخصیت و رفتارشناسی به او پیشنهاد دهد.

با توجه به اینکه تحلیل شخصیت و طراحی سیستم توصیه‌گر یک مطالعه بین‌رشته‌ای است، شکاف تحقیقی که وجود دارد این است که بسیاری از پژوهش‌های انجام شده معمولاً از یکی از روش‌های ماشینی یا علوم انسانی استفاده کرده اند. اما یکی از مزیت‌های مهم این تحقیق استفاده از یک روش تحقیق تلفیقی است که هم بر اساس روش پرسشنامه‌ای و هم روش ماشینی سعی کرده است کانال تلگرام را متناسب با شخصیت کاربران به آنها پیشنهاد بدهد. بنابراین استفاده از روش تلفیقی منجر به طراحی یک سیستم توصیه‌گری شده است که بیشتر مورد رضایت کاربر است. در ادامه پس از بررسی کارهای مرتبط، روش تحقیق و یافته‌های این پژوهش مورد بررسی قرار می‌گیرد.

۲- پیشینه پژوهش

در پژوهش‌های اخیر، تحقیقات گسترده‌ای در موضوع تحلیل شخصیت کاربران در شبکه‌های اجتماعی نظیر فیس‌بوک و توییتر صورت گرفته است. پیشینه این تحقیق را به ۳ دسته کلی می‌توان تقسیم کرد:

- تحلیل شخصیت کاربران در شبکه‌های اجتماعی مبتنی بر داده

شخصیت کاربران داشتند و مابقی صفات از دقت بسیار کمی در کلاس آزمایشی برخوردارند.

پنگ و همکارانش [۵] بر روی شبکه اجتماعی فیس‌بوک کار کرده‌اند و فیس‌بوک کاربران چینی را مورد بررسی قرار داده‌اند. پنگ در این پژوهش به دلیل دشواری قسمت بندی متون چینی از یک ماژول قطعه بندی به نام Jieba به زبان برنامه‌نویسی پایتون، متن‌ها و محتوای فیس‌بوک کاربران را قسمت بندی کرده و با استفاده از TF-IDF تعداد دفعات وقوع هر کلمه را به دست آورده است. سپس با توجه به این داده‌ها یک گروه یادگیری و یک گروه آزمایشی را تشکیل داده و تحلیل شخصیت را اعمال می‌کند. او از فایل نوشته‌های کاربران استفاده کرده و با استفاده از الگوریتم‌هایی مبنی بر «بیرون آوردن ویژگی‌ها» و «انتخاب ویژگی‌های مناسب» داده‌ها را پردازش کرده و یک فایل حاوی بردار مختص هر کاربر به دست آورده است. در نهایت با استفاده از الگوریتم یادگیری ماشین SVM توانسته شخصیت هر کاربر را پیش‌بینی کند.

سیدمن [۶] در کار پژوهشی خود در ابتدا دو عامل^۳ اصلی را به این شکل تعریف کرده است:

- ویژگی اکتسابی: شامل مشاهده استفاده دیگران از فیس‌بوک و تمایل به برقراری ارتباط و یادگیری از دیگران در یک شبکه اجتماعی.
- ویژگی رفتاری (بازتاب رفتار خود شخص): شامل به اشتراک‌گذاری یک حالت، یک متن یا یک تصویر در دیوار یا حساب کاربری شخص.

سیدمن پس از بررسی این دو عامل اصلی بر روی برخی از کاربران فیس‌بوک، پنج عامل شخصیتی را نیز در هر کدام از آنها بررسی کرده و نتیجه را در یک فایل مختص هر کاربر ثبت نمود. به طور مثال خانم‌هایی که از شاخص «پذیرا بودن» برخوردارند بیشتر تمایل به اشتراک

• تحلیل شخصیت کاربران در شبکه‌های اجتماعی مبتنی بر تصویر

• تحلیل شخصیت کاربران با استفاده از ردپاهای دیجیتال

تمرکز اصلی در این پژوهش روی تحلیل شخصیت کاربران در شبکه‌های اجتماعی مبتنی بر داده می‌باشد. بنابراین در ادامه سوابق مرتبط به شبکه‌های اجتماعی مبتنی بر داده که مرتبط با پژوهش پیش رو است، مورد بررسی قرار می‌گیرد.

۲-۱- تحلیل شخصیت کاربران در شبکه‌های اجتماعی مبتنی بر داده

در یک پژوهش بر روی زبان فارسی [۳] روند تحلیل شخصیت کاربران، در فضای شبکه اجتماعی تلگرام انجام شده است. در این تحقیق، تحلیل شخصیت کاربرانی که در یک گروه تلگرام حضور داشته و با یکدیگر درباره موضوعات مختلفی در حال گفتگو بوده‌اند، با کمک الگوریتم‌های Cosine Similarity، Bayes و MLP^۱ صورت گرفته است. براساس داده‌هایی که در این پژوهش آورده شده، نتایج به دست آمده از الگوریتم‌های MLP و شباهت کسینوسی، از صحت^۲ بهتری برخوردار است.

کری سولینگر [۴] بر روی محتوای فیس‌بوک کاربران، کار تحلیل شخصیتی را آغاز کرد و متغیرهایی از قبیل سال ورود کاربران در فیس‌بوک، تعداد دوستان و زنجیره دوستان نزدیک را در تحقیق خود نیز بررسی کرده بود. هم چنین پس از قطعه‌بندی احساسی هر کلمه با استفاده از ابزار متن باز پردازش زبان به نام NaturalLanguage Toolkit و به کارگیری ابزار وکا چند کلاس یادگیری به وجود آورده و در نهایت تحلیل شخصیت خود را بر روی کلاس‌های آزمایشی اعمال کرد. در آخر کار تنها دو صفت «پذیرا بودن» و «روان‌رنجوری» دقت بالایی در تحلیل

³ Factor

¹ Multi Layer Perceptron

² Accuracy

متن و محتوای توئیت هر کاربر را با استفاده از کتابخانه LIWC [۹] انجام داده و مجدداً بر روی متن‌ها و نوشته‌های کاربران پایگاه داده MRC [۱۰] را اعمال کرده است. با استفاده از کتابخانه LIWC 81، ویژگی مختلف متنی در پنج گروه دسته بندی شده است و MRC 150000 ویژگی زبانی و روانشناسی را مشخص کرده است. برای مرحله سوم یک تحلیل کلمه به کلمه و کاملاً دستی بر روی توئیت‌های کاربران انجام داده و امتیاز هر شخص را در پروفایل همان شخص مشخص نمود. در آخرین مرحله برای تحلیل هر یک از ویژگی‌های شخصیتی، از ابزار وکا^۲ و الگوریتم رگرسیون Zero R استفاده کرده است.

کاراناتسو و همکاران [۱۱]، رفتارها و ویژگی‌های شخصیتی کاربران در توئیت را از طریق پروفایل کاربران و توئیت‌های آنها بررسی کرده‌اند. در این تحقیق، ابتدا با طراحی یک روش مهندسی ویژگی‌های شخصیتی، طیف وسیعی از ویژگی کاربران استخراج شده و با استفاده از رگرسیون و جنگل تصادفی که از ویژگی‌های روانشناسی الهام گرفته است، رفتار و شخصیت کاربران پیش‌بینی شده است.

خسروی و عبدالحسینی [۱۲]، با دریافت اطلاعات از شبکه‌های اجتماعی و مقایسه آن با پرسشنامه‌ای که توسط افراد تکمیل می‌شود، صحت پاسخگویی به سوالات پرسشنامه را بررسی کرده و شخصیت کاربران را با استفاده از الگوریتم SVM و الگوریتم شبکه‌های عصبی ارزیابی کرده است.

اونی و همکاران، با استفاده از تکنیک‌های مبتنی بر الگوریتم یادگیری ماشین و یادگیری عمیق، پروفایل کاربران در فیسبوک و لینکدین را مورد بررسی قرار داده و علاوه بر این، به تجزیه و تحلیل متون نوشته شده توسط کاربر در این شبکه‌های اجتماعی پرداخته است. در این تحقیق راه‌حلهایی برای شناسایی مجرمان در شبکه‌های اجتماعی نیز ارائه شده است.

عکس‌های خود در فیس‌بوک دارند یا کاربرانی که از شاخص «روان‌رنجوری» برخوردارند بیشتر تمایل به برقراری ارتباط با دیگران دارند و در فیس‌بوک زنجیره ارتباطی بیشتری دارند.

در پژوهش سیدمن [۶] پیش از هر چیز کاربران آزمون نئو^۱ را داده و امتیازهای خود را دارند. سیدمن نیز این دو فاکتور اصلی را برای هر کاربر مشخص کرده و با استفاده از میانگین، انحراف معیار، واریانس و رگرسیون خطی تحلیل خود را انجام داد. در پایان مشخص می‌کند که تنها دو شاخص روان‌رنجوری و موافق بودن از دقت بالایی برخوردارند. لازم به ذکر است که سیدمن در پژوهش خود از هیچ الگوریتم خاصی استفاده نکرده و تنها با تکیه بر دو فاکتور اندازه‌گیری که خودش در ابتدای تحقیق معرفی کرده به نتایج تحلیل شخصیت دست یافت است.

تارا مارشال [۷] در پژوهش خود همانند سیدمن پیش رفته است با این تفاوت که فاکتور دیگری با عنوان «پاسخ به متن‌های اشتراک‌گذاری شده» برای ارزیابی و بررسی دقیق‌تر اضافه کرده است. او علاوه بر آزمون نئو پرسشنامه دیگری در زمینه‌هایی مثل: میزان فعالیت در فیس‌بوک، دلیل استفاده، علایق در فیس‌بوک، میزان ساعتی که هر شخص وقت می‌گذارد را در اختیار کاربران قرار داده و برای هر فرد میانگین، انحراف معیار و ضریب آلفای کرنباخ را به دست آورده و در نهایت در یک جدول مجدداً این ضرایب را با ۵ شاخص اصلی شخصیتی به دست آورده و نتایج را ثبت کرده است. نتایج نشان‌دهنده آن است که دقت مربوط به صفت‌های «روان‌رنجوری»، «دلپذیر بودن» و «موافق بودن» بالاتر از تحقیق سیدمن است.

جنیفر گلبک [۸] در پژوهش خود از کاربران شبکه اجتماعی توئیت استفاده کرده و روش‌های بیشتری را برای ارائه محکم نتایج به کار گرفته است. او در تحقیق خود پس از جمع‌آوری امتیازات کاربران از آزمون نئو، تحلیل

² Weka¹ NEO

۲-۲- انگیزه

تحلیل شخصیت کاربران در شبکه‌های اجتماعی یک موضوع جذاب بین رشته‌ای است. به ویژه اینکه یک شبکه اجتماعی کاربران زیاد با محتویات انبوه داشته باشد. محبوبیت تلگرام در بین کاربران ایرانی، شمار زیاد پست‌ها، تعداد زیاد اعضای آن، تعداد زیاد گروه‌ها و کانال‌های تلگرامی از دلایلی است که باعث شده تا پژوهش‌پیش رو به صورت خاص بر روی تلگرام انجام پذیرد. هدف این پژوهش این است که با به کارگیری پیام‌های کاربران یک گروه تلگرامی تحلیل شخصیت روی افراد گروه انجام شود و در نهایت با به کارگیری یک چارچوب توصیه‌گر کانال‌های تلگرامی متناسب با شخصیت به دست آمده افراد به آنها پیشنهاد شود. پیشنهاد چند کانال معدود در بین هزاران کانال موجود، که به شخصیت کاربران نزدیک باشد می‌تواند برای کاربران بسیار جذاب باشد.

۳- روش پژوهش

این پژوهش در دو فاز (۱) تحلیل شخصیت کاربران و (۲) ساخت یک سیستم توصیه‌گر به منظور پیشنهاد کانال‌های تلگرام براساس شخصیت کاربران انجام شده است. در ادامه، شکل ۱ یک نمای کلی از روند تحقیق را نشان می‌دهد.

۳-۱- فاز اول: تحلیل شخصیت کاربران

در این بخش، تحلیل شخصیت کاربران حاضر در یک گروه تلگرام با استفاده از دریافت پیام و پردازش آنها انجام می‌شود که در ادامه مراحل آن تشریح شده است.

۳-۱-۱- ساخت کانال تلگرام با عنوان کافه کاف

در ابتدای کار یک کانال با نام «کافه کاف» ساخته شد که این کانال بستری برای تحلیل، آموزش، بررسی چارت‌ها و مباحث مربوط به حوزه بورس و اقتصاد است. با قرار دادن مطالب خبری و آموزشی ارزنده در این زمینه، افراد زیادی در این کانال عضو شدند.

در جدول ۱ مقایسه‌ای اجمالی از برخی کارهای پیشین ارائه شده است. مطالعه کارهای پیشین نشان می‌دهد با وجود اینکه پژوهش‌های بسیاری در زمینه تحلیل شخصیت در شبکه‌های اجتماعی صورت گرفته است، اما در شبکه اجتماعی تلگرام به دلیل تفاوت در پلتفرم و ویژگی‌های خاصی که وجود دارد (وجود کانال و گروه)، پژوهش خاصی بر روی تحلیل شخصیت انجام نشده است. همچنین استفاده از روش تحقیق‌های تلفیقی ماشینی و علوم انسانی بر روی تحلیل شخصیت کمتر انجام شده است به ویژه اینکه جای یک سیستم توصیه‌گر نیز در کارهای قبلی خالی بوده است.

جدول ۱. مقایسه اجمالی از کارهای پیشین

| مرجع | بستر شبکه اجتماعی | روش تحقیق | محدودیت‌ها | سال |
|------|--------------------------------|---|--------------------------------------|------|
| [۱۳] | Sina weibo | الگوریتم‌های Bayes, Logistic regression | تمرکز روی زبان چینی | ۲۰۱۴ |
| [۴] | Facebook | ابزار Weka | تمرکز روی ابزار weka | ۲۰۱۴ |
| [۷] | Facebook | الگوریتم linear regression | کم بودن تعداد نفرات آزمایش | ۲۰۱۵ |
| [۱۴] | Facebook | Decision tree, Linear regression | کم بودن تعداد نفرات آزمایش | ۲۰۱۲ |
| [۸] | Twitter | LIWC, Weka, Zero(R) regression | انجام پژوهش تنها روی یک شبکه اجتماعی | ۲۰۱۱ |
| [۱۵] | Twitter | Linear regression | کم بودن تعداد نفرات آزمایش | ۲۰۱۶ |
| [۵] | Facebook | jeiba, weka | تمرکز روی زبان چینی | ۲۰۱۵ |
| [۱۶] | Twitter | Weka, Bayes | تمرکز روی یک شبکه اجتماعی | ۲۰۱۶ |
| [۱۷] | Instagram, Twitter, Foursquare | NUS-MSS dataset & MBTI algorithm | کم بودن تعداد | ۲۰۱۵ |
| [۱۸] | Twitter | link prediction & FriendTNS algorithm | تمرکز روی یک شبکه اجتماعی | ۲۰۱۴ |
| [۱۱] | Twitter | Regression & Random Forest | تمرکز صرفاً روی روش ماشینی | ۲۰۲۲ |
| [۱۲] | Instagram | SVM & Neural Network | بیشتر تمرکز روی شخصیت شناسی | ۲۰۲۳ |
| [۱۹] | Facebook, LinkedIn | Deep Learning | تمرکز روی روش ماشینی | ۲۰۲۳ |

در گام بعدی این ربات به عنوان عضوی از گروه «کافه کاف» و ادمین گروه، به گروه افزوده شد. ذکر این نکته مهم است که ربات حتماً می‌بایست ادمین گروه باشد تا بتواند کارهایی که از آن خواسته می‌شود را انجام دهد.

۳-۱-۳- دریافت پیام‌های کاربران گروه

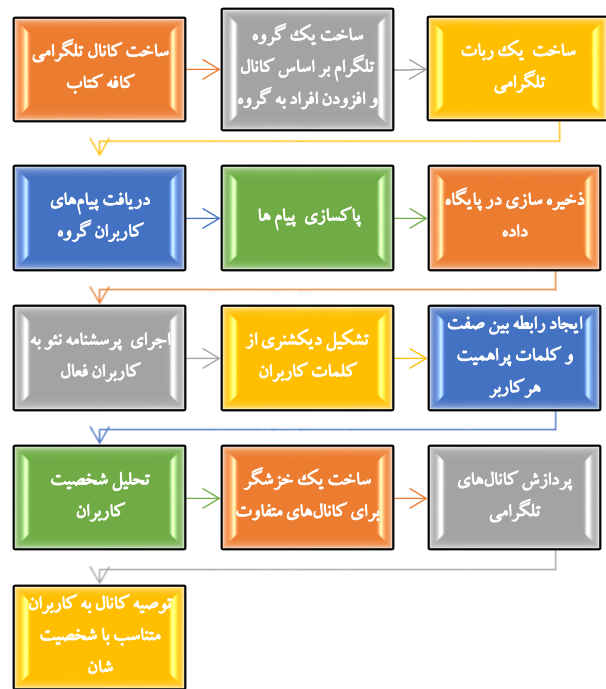
پس از افزودن ربات به گروه به عنوان ادمین، از ربات خواسته می‌شود تا به تمامی پیام‌های گروه اصطلاحاً گوش دهد و این پیام‌ها را در پایگاه داده SQL SERVER ذخیره کند. اما در این مرحله توجه به این نکته مهم است که در این پژوهش تنها پیام‌های متنی کاربران اهمیت دارد بنابراین در کد این ربات پیام‌هایی شامل: ویرایشی، بازی، صوت، شماره تلفن، آهنگ، زیرنویس، فیلم، لوکیشن، عکس و متن خالی رد شده و به پایگاه داده اضافه نمی‌شوند [۲۰].

۳-۱-۴- پاکسازی پیام‌ها

متن پیام دریافتی از کاربران توسط ربات می‌بایست تا جایی که امکان داشته باشد تمیز و عاری از هرگونه کاراکتر اضافی باشد. بنابراین در زبان برنامه‌نویسی C# متن پیام‌ها پاکسازی شد و هیچ کاراکتری اعم از: نقطه، ویرگول، پرانتز، شارپ، گیومه، مد، کروشه، حروف و اعداد انگلیسی در این پیام‌ها وجود ندارد. این مرحله با دقت و زمان زیادی انجام شده است که حتی می‌توان پیام‌های هر گروه یا هر کانال با موضوعات متنوع را با استفاده از این قطعه کدها با درصد تمیزی بالایی جهت پردازش‌های بعدی آماده نمود.

۳-۱-۵- ذخیره‌سازی در پایگاه داده

پس از دریافت پیام‌های صحیح توسط ربات، این پیام‌ها در پایگاه داده SQL SERVER ذخیره شده و جدولی با ستون‌های: نام شخص ارسال‌کننده پیام، پیام او و تاریخ و ساعت ارسال پیام ساخته می‌شود. بنابراین پیام‌های هر فرد با شناسه او نمایش داده می‌شود. از این پیام‌ها فایل



شکل ۱. مراحل کلی تحقیق

پس از گذشت چند ماه از افراد حاضر در کانال خواسته شد تا در گروهی با نام «گروه عمومی کافه کاف» عضو شوند. این امر به این دلیل بود تا افراد در یک فضای گفتار محور و تعاملی قرار بگیرند و با یکدیگر گفتگو داشته باشند. با ساخت گروه در واقع می‌توان از پیام‌های افراد جهت تحلیل شخصیت‌شان بهره‌مند شد. در این گروه علاوه بر گفتگو دربارهٔ مبحث بورس و اقتصاد راجع به موضوعات عمومی و اجتماعی-سیاسی نیز صحبت می‌شود بنابراین می‌توان تحلیل را صرفاً مختص مبحث اقتصادی ندانست.

۳-۱-۲- ساخت یک ربات تلگرامی

در این پژوهش با استفاده از زبان برنامه‌نویسی C# و کتابخانه telegrambot، یک ربات تلگرام نوشته شده است. ساخت ربات از طریق شناسه تلگرام BotFather انجام می‌شود، به این صورت که به این شناسه دستور ساخت ربات تلگرام با نام «قهوه‌چی» داده شد، سپس این شناسه ربات و یک api key را به عنوان خروجی برگرداند. این api key در واقع کلید دسترسی و ارتباط با ربات است که از طریق کدهای برنامه‌نویسی در back-end میسر می‌شود.

مشخص شدن صفت‌های کاربران، صفتی که برای هر کاربر بیشترین امتیاز را دارد به عنوان صفت شاخص آن کاربر ثبت می‌شود. در واقع هر کاربر با یک صفت برچسب می‌خورد که آن صفت بیشترین امتیاز را در بین ۵ صفت آزمون نئوی او داشته است.

۳-۱-۷- تشکیل دیکشنری از کلمات کاربران

پس از به دست آوردن پاسخ تست نئو برای هر شخص، پیام‌های این کاربران که در پایگاه داده ذخیره شده بود توسط زبان برنامه‌نویسی پایتون پردازش می‌شود. این پردازش به منظور ایجاد دیکشنری ای از کلماتی که کاربران در گروه استفاده کرده‌اند، صورت می‌گیرد. لازم به ذکر است که به دلیل استفاده زیاد از دیکشنری معروف LIWC در کارهای پیشین و عدم وجود دیکشنری مناسب در زبان فارسی، به نوعی تشکیل این دیکشنری به زبان فارسی در این بخش انجام شده است. به این صورت که ابتدا پیام‌های متنی هر فرد که با شناسه همان فرد مشخص شده است، توسط تابع Postagger که در درون خود از یک tokenizer استفاده می‌کند به شکل کلمه کلمه جدا می‌شود. (در این تابع حرف‌های اضافه، پرش سطری، فاصله و موارد اضافی حذف می‌شوند). بعد از این مراحل هر کاربر دو فایل به نام usermessage و profile usermessage منحصر به فرد خودش را دارد. که در فایل usermessage تمامی پیام‌ها که مختص خود اوست ذخیره شده و در فایل Profile بردار هر شخص قرار دارد که در ادامه توضیح داده شده است. هم‌چنین برای هر شخص برداری وجود دارد به نام dic که به شکل dic[item,value] لغت‌های موجود در پیام هر کاربر و تعداد تکرار آن لغت را در خود ذخیره می‌کند. بنابراین در ابتدای کار هر شخص دیکشنری از لغات به کار برده مخصوص به خود را دارد. در گام بعدی سعی شده تا کلمات مهم هر کاربر از این دیکشنری استخراج شود.

CSV به عنوان خروجی برای مراحل بعدی گرفته می‌شود. دلیل استفاده از فایل CSV این است که بتوان به راحتی ستون‌های جداول پایگاه داده را به شکل متن و جداشده مثلا با کمک «و» یا «،» در اختیار داشت و برای پردازش در مراحل بعدی استفاده نمود. در مرحله آخر با استفاده از دستور group by از پایگاه داده یک گزارش تهیه کرده تا مشخص شود چه کسانی در گروه تعداد بیشتری پیام داده‌اند و یا بیشتر صحبت کرده‌اند. پس از اجرای این دستور مشخص شد که ۸ نفر در این گروه بیشتر از سایر افراد صحبت کردند. اصطلاحا در این پژوهش به این کاربران، عنوان «کاربران فعال» تعلق می‌گیرد.

۳-۱-۶- اجرای پرسشنامه نئو توسط کاربران فعال

در مراحل قبلی، ضمن ذخیره پیام کاربران در پایگاه داده، بستری به منظور ادامه روند پژوهش نیز ایجاد شده است. حال می‌بایست آزمونی مبتنی بر تحلیل شخصیت از کاربران گرفته شود و از نتایج آن در این پژوهش استفاده گردد. با توجه به تحقیقات پیشین که در این حوزه انجام شده، یک پرسشنامه اینترنتی که مبتنی بر آزمون روانشناسی نئو (آزمون شخصیت شناسی ۵ عاملی) می‌باشد، ۵ ویژگی شخصیتی هر کاربر مشخص شده و به هر کاربر بر اساس این ۵ ویژگی امتیاز داده می‌شود [۲۱]. در واقع کاربران می‌بایست آزمون نئو را در کمال دقت و صحت انجام دهند تا در آخر به هر یک از آنها یک بردار اختصاص یابد. البته در تعداد کمتری از پژوهش‌ها [۱۶] محققین به جای استفاده از آزمون نئو از روش مایرز - بریگز (MBTI)^۱ استفاده کرده است [۲۲]. در این پژوهش، پس از مشخص شدن کاربران فعال، به آنها پرسشنامه نئو، داده شد. این پرسشنامه از ۶۰ سوال تشکیل شده است که در آخر برای هر فرد پنج صفت با نمره برای هر صفت داده می‌شود. بنابراین در این مرحله هر شخصی پنج صفت دارد که هر صفت او از ۶۰ امتیاز یک مقداری دارد. پس از

^۱ Myers-Briggs Type Indicator

روان‌رنجور مناسب و پراهمیت است. لازم به ذکر است که با این کار یک دیکشنری از کلمات مختلف که مربوط به صفت‌های متفاوت است در اختیار خواهیم داشت.

۳-۱-۹- تحلیل شخصیت کاربران

(۱) تحلیل شخصیت با استفاده از الگوریتم شباهت کسینوسی

به منظور پیدا کردن زوج مشابه در میان کاربران، لازم است تا از روش‌های تشخیص شباهت استفاده شود، لذا در این بخش از یکی از معروف‌ترین الگوریتم‌های تشخیص شباهت بین دو متغیر به نام الگوریتم [۲۴] شباهت کسینوسی، استفاده شده است. در این بخش دو لیست از داده‌های آزمایشی و داده‌های یادگیری موجود است. این دو متغیر به عنوان ورودی به تابع Cosin Similarity داده می‌شوند و تابع شباهت کسینوسی با استفاده از مقدار Cos زاویه که بین دو متغیر (دو بردار) وجود دارد، میزان شباهت را به عنوان خروجی باز می‌گرداند. در واقع خروجی این الگوریتم، کاربری را نشان می‌دهد که بیشترین شباهت به یکی از کاربران فعال (کاربران یادگیری) را دارد. به همین ترتیب تمامی کاربران مشابه به ازای هر کاربر موجود در لیست آزمایشی در لیست نهایی تحلیل شخصیت وارد می‌شوند.

(۲) تحلیل شخصیت با استفاده از یادگیری ماشین

در این پژوهش از کتابخانه SKLearn [۲۵] در زبان برنامه‌نویسی پایتون استفاده شده است و با هدف ایجاد کارایی بیشتر از دو الگوریتم MLP [۲۶] در شبکه‌های عصبی و الگوریتم بیز، استفاده شده است. برای این کار پس از اضافه شدن کتابخانه SKlearn، یک ماتریس کل که هر درایه آن حاصل ضرب دیکشنری هر کاربر (TF*IDF) با IDF کلمات است، تشکیل می‌شود. بنابراین ویژگی‌ها در اینجا IDF ها هستند که ستون‌های این ماتریس را شامل می‌شوند به صورت ((IDF0, IDF1, IDF2, ...)). سپس دو ماتریس train و

۳-۱-۸- ایجاد رابطه بین صفت و کلمات پر اهمیت هر کاربر در این مرحله پس از به دست آوردن دیکشنری برای هر کاربر در این گروه، یک دیکشنری سراسری به نام IDF نیز ساخته می‌شود که تمامی کلمات در تمام پیام‌ها به همراه تعداد تکرار آنها دارد. فرمول IDF به شکل زیر است:

$IDF = \log \frac{N}{\text{تعداد کلمات در محتوا}}$ = لگاریتم تقسیم تعداد کل محتوا بر محتوایی که شامل کلمه مورد نظر هستند.

$$idf(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|} \quad (1)$$

در این فرمول صورت نمایانگر کل سندها (در اینجا منظور تمام فایل‌های usermessage است) و مخرج نشان دهنده تعداد سندهایی است که کلمه X را در خود دارد.

در این پژوهش علاوه بر TF, IDF نیز محاسبه شده که برای هر نفر بیانگر تعداد تکرار یک کلمه خاص تقسیم بر تعداد کل کلمات آن فرد است. فرمول TF به شکل زیر است:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (2)$$

بعد از به دست آوردن این دو متغیر، لازم است که کلماتی از پیام‌های هر کاربر انتخاب شود که ارزش و اهمیت بیشتری داشته باشد. به عبارتی دیگر در این پژوهش کلماتی برای ما کاربردی خواهد بود که از درجه تکرار و اهمیت بالایی برخوردار باشند. بدین منظور لازم است ترکیب این دو متغیر یعنی TF * IDF [۲۳] کلمات هم محاسبه شود. در این پژوهش دیکشنری دیگری به منظور ذخیره مقدار TF*IDF هر کلمه نیز تشکیل می‌شود. بنابراین در این مرحله می‌توان گفت هر کاربر با توجه به صفتی که دارد (از طریق آزمون نئو) چه کلماتی را بیشتر در گروه به کار برده است. به طور مثال شخصی که روان‌رنجور است بیشتر از کلمات عصبانی، استرس، افسردگی، خستگی استفاده کرده است بنابراین می‌توان نتیجه گرفت که این دایره لغات برای تشخیص افراد

که این خزشگر وارد کانال‌های ذکر شده می‌شود و تمامی پست‌های موجود در آن کانال را دریافت می‌کند.

۳-۲-۲- پردازش کانال‌های تلگرامی

در نسخه دسکتاپ برنامه تلگرام از کانال‌های نامبرده یک پشتیبان گرفته شد که این پشتیبان به صورت یک صفحه وبسایت در مرورگر قابل رویت است. به این صورت که شامل یک div اصلی و چندین div تو در تو درون خودش است. وظیفه خزشگر دریافت div های کوچک (یا همان پست‌های ارسالی کانال‌ها) است. این پست‌ها درون پایگاه داده SQL SERVER با نام خود کانال ذخیره می‌شود. یعنی همانند افراد گروه، پایگاه داده شامل نام کانال، پست‌ها و تاریخ و ساعت است. در آخر هم فایل CSV این داده‌ها دریافت شده و با استفاده از زبان برنامه‌نویسی پایتون تمام مراحل که برای هر کاربر انجام شد برای هر کانال هم انجام می‌شود. در واقع هر کانال مانند یک فرد با شخصیت خودش دیده شده است. پس هر کانال با عنوان خودش دارای یک پروفایل است که درون خود بردار TF-IDF را داراست و می‌توان از همین بردار و استفاده از اجرای تابع شباهت کسینوسی بین کلمات پر اهمیت کانال و صفات، کانال‌های متناسب با هر صفت را به دست آورد [۲۴]. هم چنین دقیقاً طبق روندی که برای کاربران پیش رفت، از مکانیزم یادگیری ماشین و با استفاده از الگوریتم‌های بیز و MLP هم برای به دست آوردن شخصیت هر کانال استفاده شده که در ادامه نتایج آن مفصل‌تر بیان خواهد شد.

۳-۲-۳- توصیه کانال به کاربران متناسب با شخصیت شان با استفاده از خروجی تابع شباهت کسینوسی و الگوریتم‌های بیز و MLP می‌توان کانال‌هایی که متناسب با شخصیت کاربران هستند را به آنها پیشنهاد داد. در آخر کار از هر کاربر برای کانال‌هایی که به او پیشنهاد شده است پرسش می‌شود و نتایج و صحت نیز پیش‌بینی می‌شود.

test ساخته می‌شود به این صورت که ۳۳ درصد از ماتریس کل برای آزمایش و ۶۶ درصد از آن برای یادگیری برداشته می‌شود و همراه اینها لیست هدف نیز داده می‌شود که الگوریتم یادگیری ماشین بر اساس یادگیری می‌بایست ۳۳ درصد از کاربران آزمایشی را خودش تحلیل کند و نتیجه را در لیست هدف بازگرداند. لازم به ذکر است که لیست هدف هم به منظور دریافت نتیجه بهتر به دو روش ایجاد شده است:

- با استفاده از خروجی تابع شباهت کسینوسی
- با استفاده از پاسخ آزمون نئوی تمام کاربران (روش دستی).

پس از انجام الگوریتم نتیجه تحلیل شخصیت به صورت خروجی در لیست هدف محاسبه می‌شود.

۳-۲-۲- فاز دوم: ساخت یک سیستم توصیه‌گر برای پیشنهاد کانال به کاربران

در این بخش، پس از به دست آوردن شخصیت هر کاربر، کانال‌های مرتبط با شخصیت کاربران با استفاده از یک سیستم توصیه‌گر به آنها پیشنهاد می‌شود که در ادامه به تفصیل بیان شده است.

۳-۲-۱- ساخت یک خزشگر برای کانال‌ها

پس از به دست آوردن شخصیت هر کاربر، اکنون نوبت آن است که کانال‌هایی در زمینه‌های متفاوت و با توجه به شخصیت کاربران به آنها پیشنهاد شود. در این زمینه کانال‌هایی با عناوین: ارز دیجیتال، عصر بانک، banker، چرچیل(دارای مطالبی همچون سیگنال‌های بورسی، تحلیل بازار و بررسی جلسات بورسی و اقتصادی)، سینما، برنامه‌نویسان، دکتر سلامت، دوستی با خدا، فروغ فرخزاد، شعر سپید، حکایت و داستان، روانشناسی کودک، شیک پذیرایی کن، تک بیت ناب، ورزش ۳ و دنیای اقتصاد مورد ارزیابی قرار گرفت. بدین منظور یک خزشگر با زبان برنامه‌نویسی C# و کتابخانه html agility pack نوشته شد



شکل ۲. مراحل تحلیل ویژگی شخصیتی کاربران

۴- یافته‌ها

در این بخش می‌بایست تناسب بین نتایج این پژوهش و اهداف پژوهش مورد بررسی قرار گیرد و مشخص شود آیا نتایج حاصل از این روش‌ها، اهداف پژوهش را محقق کرده است یا خیر. در ادامه به ارزیابی روش و مقایسه نتایج این پژوهش با دیگر نتایج موجود پرداخته می‌شود.

۴-۱-۱- تحلیل آماری اولیه

در این پژوهش، پیام‌هایی که مورد بررسی قرار گرفته اند پیام‌های افرادی است که کاربران فعال در گروه هستند و بیشترین تعداد کلمات را به خود اختصاص داده اند، این کار به این منظور انجام شده است که اعتبار نتایج بیشتر شود و تعداد کلمات به کار رفته در دایره لغات این پژوهش افزایش یابد. هم چنین کانال‌هایی هم که جهت پیشنهاد به کاربران بررسی و ارزیابی شد به طور میانگین ۱۷۰۰۰ پیام و ۲۵۴۰۰۰ عضو داشته و از کانال‌های معتبر در زمینه‌های موضوعی مختلفی هستند.

در ادامه با توجه به نتایج تحلیل اولیه آماری این پژوهش، میزان درستی نتایج حاصل از این پژوهش به طور جداگانه در هر الگوریتم بررسی خواهد شد.

۴-۱-۲- دقت نتایج به دست آمده از الگوریتم‌ها

در این پژوهش از سه روش برای تحلیل شخصیت کاربران یک گروه تلگرامی استفاده شد. صحت حاصل از الگوریتم شباهت کسینوسی برابر با ۷۶/۲۵ درصد مشاهده شد و

۳-۲-۴- مراحل تصمیم‌گیری در سیستم توصیه‌گر

سیستم توصیه‌گر ارائه شده در این تحقیق، از یک فرآیند چندمرحله‌ای برای اتخاذ تصمیمات خود استفاده می‌کند. شکل ۲ روند تحلیل ویژگی شخصیت کاربر و در نهایت توصیه کانال به آنها رابه صورت مرحله‌ای نشان می‌دهد. این مراحل عبارتند از:

۱. جمع‌آوری داده‌های کاربر از گروه تلگرام: در این مرحله، اطلاعات و فعالیت‌های کاربران فعال گروه تلگرام جمع‌آوری می‌شوند. این موارد شامل پست‌ها، پیام‌ها، واکنش‌ها و سایر تعاملات درون گروه است.

۲. ارائه پرسشنامه نئو به کاربران: بعد از انتخاب کاربران فعال، به آنها تست روانشناسی نئو داده می‌شود و پاسخ این تست از کاربران دریافت خواهد شد.

۳. تحلیل داده‌ها برای شناسایی ویژگی‌های شخصیتی:

در این مرحله، داده‌های جمع‌آوری شده تحلیل می‌شوند تا ویژگی‌های شخصیتی کاربران شناسایی شوند. در واقع این فاز، تحلیل شخصیت افراد بر اساس مطالبی است که در گروه تلگرام منتشر می‌کنند.

۴. طبقه‌بندی کاربران به انواع مختلف شخصیتی: بر اساس نتایج تحلیل، کاربران به گروه‌های شخصیتی مختلف طبقه‌بندی می‌شوند. که این دسته‌بندی بر اساس مدل‌های روانشناسی معروف مانند مدل پنج عاملی شخصیت انجام می‌شود.

۵. توصیه کانال‌های تلگرامی بر اساس انواع شخصیتی شناسایی شده: در نهایت، بر اساس طبقه‌بندی شخصیتی کاربران، کانال‌های تلگرامی که با علایق و ویژگی‌های شخصیتی آنها مطابقت دارند، به آنها توصیه می‌شوند.

۶. دریافت بازخورد از کاربران: و در آخر، از کاربران فیدبک درستی این کار گرفته خواهد شد. همچنین برای سنجش صحت این الگوریتم، پاسخ تست نئوی کاربران با پاسخ دریافت شده از الگوریتم پیشنهادی نیز مقایسه می‌شود.

```

Users_result_analysis_Cosine_Similarity.txt - Notepad
File Edit Format View Help
Mohammad_Hassan_Torabi,39,29,40,46,51-----
R_Moafimadani,26,46,42,43,54-----
Behnam,28,32,34,44,38-----Abbas,36,35,40,52,44 s=4
M._Farham,23,42,42,46,50-----
A_K,33,38,39,40,47-----
Hadi_Alizadeh,29,40,36,38,45-----Abbas_Davari,28,44,39,27,37 s=5
SafaAli_Mozafarnejad,34,34,34,40,45-----Abbas,36,35,40,52,44 s=4
Reza_Sadoughi,45,45,46,46,46-----
reza,39,36,42,39,38-----
Somayeh,24,40,39,41,51-----
Arman,52,27,49,44,20-----
Mehdi,33,30,40,41,48-----Abbas,36,35,40,52,44 s=4
meyfi,43,27,38,37,37-----
Mercede,31,44,40,47,53-----Mr_Said_Zavarei,26,43,39,38,46 s=4
Vahid_m,34,42,36,32,49-----
Mr_Forex,31,40,34,53,42-----Mr_Said_Zavarei,26,43,39,38,46 s=4
Hamide,23,43,36,48,45-----Abbas_Davari,28,44,39,27,37 s=4
Hosein,35,34,38,45,52-----Abbas,36,35,40,52,44 s=3
Nabi_Aziznesuad,20,38,37,49,55-----
Arash,23,36,32,45,49-----Mr_Said_Zavarei,26,43,39,38,46 s=4
Mohammad_Mahdi,32,37,36,46,50-----
Vara,28,32,38,43,46-----Setmax,28,45,43,46,47 s=5
Mohammad_Sh,26,31,41,41,44-----
Pedram_Zahedi,34,41,43,37,42-----Abbas,36,35,40,52,44 s=4
Keyhan_Kamangar,26,33,35,38,44-----Abbas,36,35,40,52,44 s=4
mrr,21,36,42,45,51-----> Mr_Said_Zavarei,26,43,39,38,46 s=3
Majid,34,43,43,46,53-----Setmax,28,45,43,46,47 s=4
Mr_Haghini,33,35,36,43,50-----Mr_Said_Zavarei,26,43,39,38,46 s=4
AHF_XIII,26,41,38,39,44-----Mr_Said_Zavarei,26,43,39,38,46 s=5
61/16 /5 *100 = 76.25
    
```

شکل ۳. خروجی الگوریتم شباهت کسینوسی

برای این سیستم توصیه‌گر از الگوریتم شباهت کسینوسی برای به دست آوردن کانال‌های متناسب با شخصیت هر فرد استفاده شده است. برای به دست آوردن صحت نتایج حاصله، از کاربرانی که آن کانال به ایشان نگاشت شده پرسش شد که: «از بین کانال‌هایی که به شما پیشنهاد داده شد چه تعداد از این کانال‌ها متناسب با سلیقه شما است و شما به این کانال‌ها علاقه دارید؟» بنابراین درصد تعداد افرادی که راضی هستند میزان صحت این سیستم توصیه‌گر را نشان می‌دهد. نتایج به دست آمده نشان می‌دهد که ۶۵/۴۲ درصد از کاربران بازخورد مثبت به کانال‌ها نشان داده اند.

در شکل ۴ مقایسه‌ای بین سه الگوریتم به کار رفته در این پژوهش روی نمودارهای زیر قابل مشاهده است. لازم به ذکر است که در روش دوم (manual) دقت به دست آمده از الگوریتم بیز همانند روش اول ۸۲/۵ درصد مشاهده شده است و الگوریتم MLP نیز در روش دوم نتیجه‌ای نامطلوب از خود نشان داده است. هم چنین در این پژوهش از روش

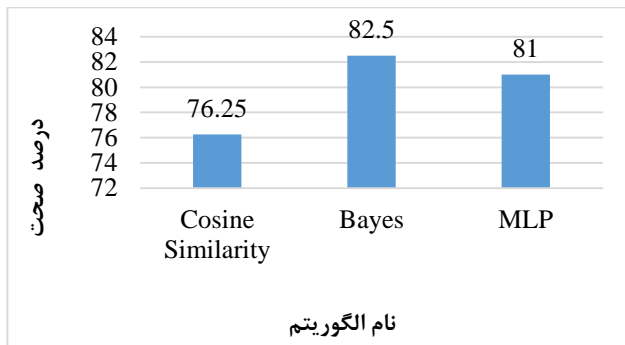
صحت نتایج به دست آمده از الگوریتم‌های بیز و MLP به ترتیب، ۸۲/۵ و ۸۱ درصد به دست آمد.

در شکل ۳ نمونه‌ای از خروجی مشاهده می‌شود. در این نمونه همانطور که قابل رویت است، به طور مثال شخصی به نام کاربری بهنام با استفاده از الگوریتم شباهت کسینوسی شبیه به شخصی به نام کاربری عباس شده است که نمرات آزمون نئوی هر یک به تفکیک نشان داده شده است. این خروجی نشان می‌دهد که الگوریتم با استفاده از مجموعه کلماتی که عباس در کار برده و شباهت بین کلمات او با بهنام، این دو نفر را در شخصیت شبیه به هم دانسته است. هم چنین در این شکل نمرات آزمون نئوی بهنام و عباس بسیار شبیه و از پنج نمره آزمون چهار مورد آن در یک بازه بوده و این خود نشان دهنده نتیجه مطلوب است. و یا به عنوان مثالی دیگر شخصی به نام کاربری mrr شبیه به شخصی به نام mr said zavarei شده است که همانطور که مشاهده می‌شود این دو نفر تنها سه صفت از پنج صفت آزمون نئوی‌شان در یک بازه بوده است. نکته قابل توجه در این شکل این است که با توجه به این که این خروجی توسط ماشین ارائه شده و کاملاً واقعی است، برخی افراد به دلیل عدم صحبت در گروه هیچ شباهتی برایشان نوشته نشده و مقابل اسامی آنها جای خالی دیده می‌شود.

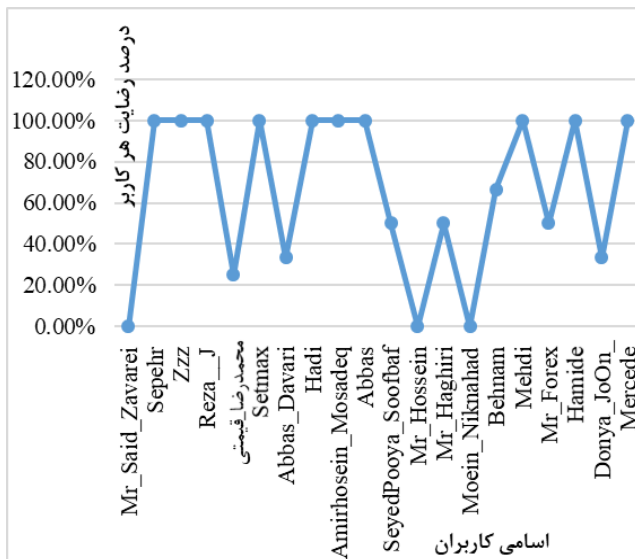
۴-۱-۳- دقت نتایج به دست آمده برای کانال‌های پیشنهاد شده

برای پردازش کانال‌های تلگرامی، از روش تحلیل شخصیت افراد گروه استفاده شده است. یعنی هر کانال با محتویات متفاوت به صورت اشخاص با شخصیت‌های متفاوت در نظر گرفته می‌شود که با به کارگیری این روش هر کانال با توجه به لغات و عباراتی که استفاده کرده به داده‌های یادگیری در این پژوهش نگاشت می‌شود.

دست آمده و زبان برنامه‌نویسی به کار رفته در پژوهش، به پژوهش [۵] مشابهت دارد با این تفاوت که در پژوهش [۵] از شبکه اجتماعی فیس‌بوک استفاده شده و زبان کاربران چینی بوده است.



شکل ۴. مقایسه درصد صحت الگوریتم‌های استفاده شده در تحلیل شخصیت



شکل ۵. میزان رضایت هر فرد از کانالی که به وی توصیه شده

جدول ۲. مقایسه کارهای پیشین انجام شده از نظر میزان صحت

| مرجع | صحت | بستر شبکه اجتماعی | الگوریتم و ابزار پیاده‌سازی در هر پژوهش |
|------|-----|-------------------|---|
| [۱۲] | ۷۰٪ | Sina weibo-Bayes | Logistic Regression |
| [۴] | ۷۰٪ | Facebook | Weka Tools, Bayes |
| [۷] | ۷۶٪ | Facebook | Regression |
| [۱۴] | ۵۴٪ | Facebook | Decision tree, Linear regression |
| [۸] | ۱۱٪ | Twitter | LIWC Dic, Weka Tools, Zero(R)Regression |
| [۱۵] | ۶۵٪ | Twitter | Linear regression |
| [۵] | ۶۰٪ | Facebook | Jeiba Tools, Weka Tools, Regression |
| [۱۶] | ٪۶۰ | Twitter | Weka Tools, Bayes |

سعی و خطا در جهت به دست آوردن نتایج مطلوب‌تر بسیار استفاده شده که همین روش موجب شد الگوریتم MLP نتایج مطلوبی را در اختیار نگذارد و مجدد از روش دیگری به نام الگوریتم بیز استفاده شود.

با توجه به شکل ۵، تنها سه نفر از بین این افراد جواب منفی به کانال‌های پیشنهادی داشتند و ده نفر از این افراد به طور کامل با کانال‌های پیشنهاد شده موافق بودند و دیگر افراد نیز از بین کانال‌های پیشنهادی فقط با برخی از آنها موافق بودند. که این نتایج با توجه به عدم وجود پیشینه کاری در این حوزه، مطلوب به نظر می‌رسد. همچنین ذکر این نکته قابل توجه است که اشخاصی که درصد رضایت‌مندی آنها صفر درصد بوده است، افرادی هستند که نتایج شان در آزمون نئو نیز در حالت متوسط (نه قوی نه ضعیف) از هر ویژگی شخصیتی قرار گرفته است. به طور مثال برای ویژگی با وجدان بودن نه خیلی افراد با وجدانی محسوب می‌شوند و نه خیلی افراد بی‌وجدان.

با توجه به این نکته مهم که این پژوهش از نظر شرایطی همچون زبان، نوع شبکه اجتماعی، مجموعه داده، نحوه پیاده‌سازی و روش تحقیق با مقالات پیشین متفاوت است، اما به منظور بیان نزدیکی نتایج این تحقیق به کارهای پیشین، مقایسه‌ای اجمالی با چند نمونه از کارهای سابق در قالب جدول ۲ ارائه شده است:

با توجه به تفاوت بستر شبکه‌های اجتماعی و الگوریتم‌های پیاده‌سازی در کارهای پیشین، میانگین درصد صحت نتایج حاصله حدود ۶۰ درصد می‌باشد. در میان این سوابق پژوهش [۸]، صحت بسیار اندکی در نتایج خود دارد که پژوهشگران این پژوهش صراحتاً این مورد را به عنوان ضعف اصلی کار خود بیان کرده‌اند. بنابراین با توجه به تفاوت‌های بسیاری که پژوهش پیش رو با پژوهش‌های پیشین در بستر شبکه اجتماعی، زبان کاربران، و الگوریتم پیاده‌سازی دارد؛ پژوهش پیش رو از نظر شباهت نتایج به

برای هر کلاس شخصیتی، سطح ترجیح برای هر نوع کانال (هنری، علمی، ورزشی، اجتماعی، خبری) با استفاده از کلمات «بالا»، «متوسط»، و «پایین» توصیف شده است.

۶- نتیجه‌گیری و پیشنهاد

هدف از انجام این پژوهش، تحلیل شخصیت کاربران به صورت خودکار و سپس ارائه یک سیستم توصیه‌گر برای پیشنهاد کانال‌های تلگرامی متنوع متناسب با شخصیت کاربران با استفاده از نتایج به دست آمده است. در این راستا به بررسی پژوهش‌های بسیاری در این حوزه پرداخته شد و در نهایت سیستمی با به‌کارگیری الگوریتم‌های شباهت کسینوسی، بیز و MLP طراحی شد که با استفاده از جمع‌آوری پیام‌های کاربران در یک گروه گفتگوی عمومی و نتایج آزمون نئوی هر کاربر، تحلیل شخصیت افراد گروه با درصد صحت ۷۶/۲۵ برای الگوریتم شباهت کسینوسی و ۸۲/۵ درصد برای روش بیز انجام شد. برای انجام فاز دوم، کانال‌های تلگرام نیز همانند کاربران مورد تحلیل و بررسی قرار گرفت و شخصیت هر کدام از آنها مشخص شد و با توجه به شباهت بین محتویات کانال‌ها و شخصیت افراد گروه، کانال‌های متناسب با شخصیت و سلیقه هر کاربر به او پیشنهاد شد. در ادامه ویژگی‌های این سیستم طراحی شده ذکر شده است:

- ایجاد دیکشنری و مجموعه داده نسبتاً خوب در زبان فارسی و در حوزه تحلیل ویژگی‌های شخصیتی به عنوان نقطه شروع (همانند دیکشنری LIWC)
- استفاده از شبکه اجتماعی تلگرام به عنوان یک بستر پژوهشی متفاوت و مبتنی بر پیام در این حوزه
- استفاده از سه الگوریتم متفاوت جهت به دست آوردن نتایج با دقت بالاتر

تحلیل کانال‌ها همانند افراد گروه با شخصیت‌های متفاوت بازخورد مطلوب کاربران از کانال‌های پیشنهادی و با درصد رضایتمندی ۶۵/۴۲ درصد و نهایتاً ارائه یک سیستم که به شکل موفقیت‌آمیزی می‌تواند بصورت اتوماتیک شخصیت

همچنین روش پیاده‌سازی و الگوریتم به کار رفته تا حدودی متفاوت است. اما در این پژوهش از شبکه اجتماعی تلگرام و زبان فارسی استفاده شده است و لازم به ذکر است که مجموعه داده در زبان فارسی جهت انجام این تحقیق بسیار اندک و ناقص است با این وجود در این پژوهش با تلاش بسیار سعی در جمع‌آوری داده‌های فارسی و به وجود آوردن دیکشنری لغات در زبان فارسی شده است. همچنین در این پژوهش از چند روش پیاده‌سازی و آزمون چند روش در جهت به دست آوردن نتیجه مطلوب‌تر و نزدیک‌تر به واقعیت استفاده شد. بنابراین گرچه برای مقایسه بین پژوهش‌ها باید مقیاس یکسانی جهت مقایسه وجود داشته باشد، اما این مقایسه به منظور نشان دادن نزدیک بودن نتایج به نتایج این پژوهش لحاظ شده است.

۵- تحلیل ارتباط میان کلاس‌های شخصیتی و توصیه‌های ارائه شده

در این تحقیق، نحوه تاثیرگذاری کلاس‌های شخصیتی مختلف بر انواع پیشنهاداتی که سیستم توصیه‌گر ارائه می‌دهد را بررسی می‌کنیم. برای مثال، کاربرانی که ویژگی‌های شخصیتی خلاقانه دارند ممکن است به کانال‌هایی با موضوعات نوآورانه یا هنری جذب شوند، در حالی که کاربران با ویژگی‌های تحلیلی ممکن است به کانال‌های علمی یا فناوری علاقه‌مند باشند. این ارتباطات از طریق تجزیه و تحلیل داده‌های رفتاری کاربران و بازخوردهای آنها به دست آمده و در جدول ۳ نمایش داده شده‌اند. در این جدول، ترجیحات احتمالی کاربران بر اساس کلاس‌های شخصیتی مختلف نشان داده شده است.

جدول ۳. ترجیح کاربران بر اساس کلاس‌های شخصیتی و توصیه‌های ارائه شده

| کلاس شخصیتی | ترجیح کانال‌های هنری | ترجیح کانال‌های علمی | ترجیح کانال‌های ورزشی | ترجیح کانال‌های اجتماعی | ترجیح کانال‌های خبری |
|-------------|----------------------|----------------------|-----------------------|-------------------------|----------------------|
| پایین | بالا | متوسط | پایین | متوسط | پایین |
| بالا | پایین | بالا | متوسط | پایین | بالا |
| متوسط | متوسط | بالا | بالا | بالا | متوسط |
| بالا | پایین | متوسط | پایین | بالا | بالا |

- [8] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from twitter," in 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing, 2011: IEEE, pp. 149-156.
- [9] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: LIWC 2001," Mahway: Lawrence Erlbaum Associates, vol. 71, no. 2001, p. 2001, 2001.
- [10] M. Coltheart, "The MRC psycholinguistic database," *The Quarterly Journal of Experimental Psychology Section A*, vol. 33, no. 4, pp. 497-505, 1981.
- [11] D. Karanatsiou, P. Sermpezis, D. Gruda, K. Kafetsios, I. Dimitriadis, and A. Vakali, "My tweets bring all the traits to the yard: Predicting personality and relational traits in Online Social Networks," *ACM Transactions on the Web (TWEB)*, vol. 16, no. 2, pp. 1-26, 2022.
- [12] A. Khosravi and H. Abdolhosseini, "Personality in social networks using thematic modeling of user feedback," *Soft Computing Journal*, 2023.
- [13] D. Wan, C. Zhang, M. Wu, and Z. An, "Personality prediction based on all characters of user social media information," in *Chinese National Conference on Social Media Processing*, 2014: Springer, pp. 220-230.
- [14] R. Wald, T. Khoshgoftaar, and C. Sumner, "Machine prediction of personality from Facebook profiles," in 2012 IEEE 13th International Conference on Information Reuse & Integration (Iri), 2012: IEEE, pp. 109-115.
- [15] L. Liu, D. Preotiuc-Pietro, Z. R. Samani, M. E. Moghaddam, and L. Ungar, "Analyzing personality through social media profile picture choice," in *Tenth international AAAI conference on web and social media*, 2016.
- [16] L. C. Lukito, A. Erwin, J. Purnama, and W. Danoekoesomo, "Social media user personality classification using computational linguistic," in 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE), 2016: IEEE, pp. 1-6.
- [17] S. Kedar and D. Bormane, "Automatic personality assessment: A systematic review," in 2015 International Conference on Information Processing (ICIP), 2015: IEEE, pp. 326-331.
- [18] R. Lambiotte and M. Kosinski, "Tracking the digital footprints of personality," *Proceedings of the IEEE*, vol. 102, no. 12, pp. 1934-1939, 2014.
- [19] S. Ouni, F. Fkih, and M. N. Omri, "A survey of machine learning-based author profiling from texts analysis in social networks," *Multimedia Tools and Applications*, vol. 82, no. 24, pp. 36653-36686, 2023/10/01 2023, doi: 10.1007/s11042-023-14711-8.
- [20] K. Biswas, P. Shivakumara, U. Pal, T. Chakraborti, T. Lu, and M. N. B. Ayub, "Fuzzy and genetic algorithm based approach for classification of personality traits oriented social media images," *Knowledge-Based Systems*, p. 108024, 2021.
- [21] T. A. Mooradian and J. B. Nezlek, "Comparing the NEO-FFI and Saucier's Mini-Markers as measures of the Big Five," *Personality and Individual Differences*, vol. 21, no. 2, pp. 213-215, 1996.
- [22] A. Furnham, "The big five versus the big four: the relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality," *Personality and Individual Differences*, vol. 21, no. 2, pp. 303-307, 1996.
- [23] S. Jabri, A. Dahbi, T. Gadi, and A. Bassir, "Ranking of text documents using TF-IDF weighting and association rules mining," in 2018 4th International Conference on Optimization and Applications (ICOA), 2018: IEEE, pp. 1-6.
- کاربران را شناسایی کرده و کانال‌های مورد علاقه وی را در میان هزاران کانال موجود به او پیشنهاد بدهد.
- مباحثی که در این پژوهش بیان شده است، مسائلی نوپا و جدید هستند که در ابعاد مختلف فرصت‌های زیادی برای تحقیق و بهبود آن فراهم است. یکی از فرصت‌های مهم برای ادامه روند پژوهش، غنی‌سازی مجموعه داده‌ها به زبان فارسی خواهد بود. با توجه به عدم ارائه پژوهش‌های زیاد در این حوزه و فقیر بودن مجموعه داده به زبان فارسی، با صرف وقت و تحقیقات زیاد و استفاده از دیکشنری ساخته شده در این پژوهش به عنوان کلید اولیه، می‌توان مجموعه داده‌ای غنی و مشابه با دیکشنری LIWC در زبان انگلیسی توسعه داد تا نتایج حاصل از تحلیل شخصیت با درصد بالایی از دقت و کارایی همراه شود. این موضوع قابل بسط در سایر زبانها هم خواهد بود. از دیگر فرصت‌های پیشنهادی می‌توان به استفاده ویژگی‌های شخصیتی افراد در حوزه‌های تجاری نیز نام برد.

مراجع

- [1] S. Han, H. Huang, and Y. Tang, "Knowledge of words: An interpretable approach for personality recognition from social media," *Knowledge-Based Systems*, vol. 194, p. 105550, 2020.
- [2] Y. Mehta, S. Fatehi, A. Kazameini, C. Stachl, E. Cambria, and S. Eetemadi, "Bottom-up and top-down: Predicting personality with psycholinguistic and language model features," in 2020 IEEE International Conference on Data Mining (ICDM), 2020: IEEE, pp. 1184-1189.
- [3] M. J. Shayegan and M. Valizadeh, "A Method for Identifying Personality Traits in Telegram," in 2022 8th International Conference on Web Research (ICWR), 2022: IEEE, pp. 88-93.
- [4] C. Solinger, L. Hirshfield, S. Hirshfield, R. Friendman, and C. Leper, "Beyond Facebook Personality Prediction," in *International Conference on Social Computing and Social Media*, 2014: Springer, pp. 486-493.
- [5] K.-H. Peng, L.-H. Liou, C.-S. Chang, and D.-S. Lee, "Predicting personality traits of Chinese users based on Facebook wall posts," in 2015 24th Wireless and Optical Communication Conference (WOCC), 2015: IEEE, pp. 9-14.
- [6] G. Seidman, "Self-presentation and belonging on Facebook: How personality influences social media use and motivations," *Personality and individual differences*, vol. 54, no. 3, pp. 402-407, 2013.
- [7] T. C. Marshall, K. Lefringhausen, and N. Ferenczi, "The Big Five, self-esteem, and narcissism as predictors of the topics people write about in Facebook status updates," *Personality and Individual Differences*, vol. 85, pp. 35-40, 2015.

- [26] M. W. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric environment*, vol. 32, no. 14-15, pp. 2627-2636, 1998.
- [24] B. Li and L. Han, "Distance weighted cosine similarity measure for text classification," in *International Conference on Intelligent Data Engineering and Automated Learning*, 2013: Springer, pp. 611-618.
- [25] J. Bergstra, D. Yamins, and D. D. Cox, "Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms," in *Proceedings of the 12th Python in science conference*, 2013: Citeseer, pp. 13-20.