مقـــاله تحقیقی

تحلیل احساسات مرتبط با بیماری کرونا در وب‌سایت‌های خبری داخل ایران

سمیه کنعانی[1]،صادق سلیمانی[2*]

[1] دانشجوی کارشناسی ارشد مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه کردستان، سنندج، ایران

[2] استادیار گروه مهندسی کامپیوتر، دانشگاه کردستان، سنندج، ایران

* نشانی نویسنده مسئول:

دانشگاه کردستان، سنندج، ایران

نشانی الکترونیک: S.Sulaimany@uok.ac.ir

چکیده

زمینه: رشد اطلاعات در وب‌سایت‌های خبری و رسانه‌های اجتماعی، تجزیه و تحلیل احساسات را به یک زمینه سودمند برای یافتن نظرات عوام تبدیل کرده است. تحلیل احساس شامل ایجاد یک سامانه هوشمند برای جمع‌آوری و پی‌بردن به نظرات، احساسات و بازخوردهای انسان‌هاست. دست‌یابی به نگرش ایرانیان در شروع بیماری کرونا، با کمک الگوریتم‌های تحلیل احساس، هدف اصلی این پژوهش است.

روش کار: روش پژوهش از نوع محاسباتی و یک زیرشاخه جدیدتر از پردازش زبان طبیعی به نام تحلیل احساس است. داده‌ها از نظرات کاربران در دو وب‌سایت خبری پر مخاطب کشور، ایسنا و فارس نیوز، مرتبط با اخبار بیماری کووید-۱۹، در بازه زمانی اوایل اسفند ماه ۱۳۹۸ تا خرداد ماه ۱۳۹۹ جمع‌آوری شد. تعداد داده‌های جمع‌آوری شده در این پژوهش ۵۰۰ خبر می‌باشد که متناسب با اخبار نظرات کاربران نیز جمع‌آوری شد که تعداد نظرات کاربران نیز نزدیک به ۳۰۰ کاربر می‌باشد. برای تجزیه و تحلیل نظرات کاربران، از تکنیک‌ها و ابزارهای یادگیری ماشین مانند الگوریتم درخت تصمیم برای طبقه‌بندی احساسات مثبت و منفی استفاده شد. بدین صورت که از کلمات با بار احساسی منفی و مثبت برای تحلیل نظر کاربران استفاده شد. جمع‌آوری و پیش‌پردازش داده‌ها با استفاده از زبان برنامه‌نویسی پایتون وکتابخانه‌های موجود آن انجام گردید.

یافته ها: نگرش منفی نسبت به کووید-۱۹ در شروع همه‌گیری، در کاربران خبرگزاری فارس نیوز ۲۱ و در کاربران ایسنا ۱۷ درصد بود. یعنی نزدیک به ۸۰ درصد از کاربرانی که نظرات خود را منعکس کرده بودند، نگرش خوشبینانه و سهل‌گیرانه درباره بیماری داشتند.

نتیجه گیری: ممکن است یک علت شیوع و تلفات زیاد کووید-۱۹ در ایران در ابتدا، نگرش اشتباه و ناآگاهی عوام جامعه نسبت به آن باشد، چرا که برخلاف کشورهای دیگر جهان، مردم ایران در شروع همه‌گیری کووید-۱۹، نسبت به بیماری و مشکلات آن با کلمات دارای بار مثبت اعم از شادی، خوشبینی، سهل‌گیری و بی‌توجهی یاد کرده‌اند.

واژه های کلیدی: بیماری کرونا، ویروس کرونا، طبقه‌بندی احساسات، تحلیل احساسات، وب‌سایت‌های داخلی ایران

مقدمه

کرونا ویروس‌ها ۱خانواده بزرگی از ویروس‌ها هستند که برخی از آن‌ها باعث بیماری در انسان و حیواناتی دیگر مانند خفاش و شتر می‌شوند. ویروس‌های کرونا به‌طور معمول باعث بیماری‌های خفیف می‌شوند. بعضی اوقات، ویروس‌های کرونای حیوانات، به انسان‌ها نیز سرایت یافته و باعث بیماری‌های شدید مانند سندرم حاد تنفسی[۱] در سال ۲۰۰۲ و سندرم تنفسی مرس[۲] در سال ۲۰۱۲ شده‌اند (۱–۳). گونه جدید این ویروس، کرونا ویروس سندروم حاد تنفسی دو[۳]، با نام کووید–۱۹ که قبلا در انسان شناسایی نشده بود در دسامبر ۲۰۱۹ برای اولین بار در شهر ووهان چین شناسایی شد. شیوع این ویروس در چین گسترده‌تر بود و نمونه‌های مثبت یافت شده از این بیماری تا ۲۷ نوامبر ۲۰۲۰ در اکثر کشورهای دنیا گزارش شده است (۱،۴). طبق آمارهای موجود، ظهور این بیماری تقریبا در ۹۶ درصد افراد خفیف است و در ۴ درصد جدی و حیاتی می‌باشد(۵). در ایران، اولین مورد مثبت کووید–۱۹ در ۲۰ فوریه ۲۰۲۰ در شهر قم تایید شد (۶). طبق آمار جهانی، تا به امروز، ایران جزو ده کشور برتر از نظر میزان مرگ و میر ناشی از این ویروس در کل جهان بوده است (۵). تبعات منفی این ویروس به سلامتی انسان‌ها محدود نشده است بلکه علاوه بر شوک‌های سنگین اقتصادی به دنیا، سبب افزایش بیکاری و تغییر روند فعالیت‌های رایج مانند آموزش شده است (۷).

رسانه‌های اجتماعی به عنوان یک بستر اصلی در جمع‌آوری اطلاعات و یادگیری اجتماعی در طول بحران بهداشت عمومی نقش ایفا می‌کنند. رسانه‌های اجتماعی در واقع شیوه درک جامعه و واکنش به شیوع کووید–۱۹ را تغییر دادند و آگاهی مردم در رابطه با این ویروس ناشناخته را گسترش دادند. این رسانه‌ها به‌طور گسترده‌ای توسط متخصصان بهداشت عمومی برای نظارت بر سرایت بیماری و درک واکنش‌های عمومی در مورد مسائل بهداشت عمومی مورد استفاده قرار گرفته‌اند (۸). رشد اطلاعات در رسانه‌های اجتماعی و وب‌سایت‌های خبری، تجزیه و تحلیل احساسات را به یک زمینه مرتبط برای یافتن نظرات عوام تبدیل کرده است. تجزیه و تحلیل احساسات که به آن نظر کاوی[۴] نیز می‌گویند، شامل ایجاد یک سامانه برای جمع‌آوری و آزمودن نظرات انسان‌ها و احساسات و ارزیابی آن‌ها، درمورد ساخت یک محصول در پست‌های وبلاگ، نظرات، توییت‌ها و.... می‌باشد (۹). تحلیل احساسات و عقیده کاوی به‌معنی کشف و شناخت احساسات مثبت و یا منفی مردم از بین متون ارسالی آن‌ها، راجع به یک مسئله یا محصول است. این امر در بالاترین سطح شامل طبقه‌بندی نظرات یا عقاید در دسته‌های «مثبت»، «منفی» و یا «خنثی» است. وب‌سایت‌های خبری و شبکه‌های اجتماعی منبع مهمی برای داده‌های مربوط به احساس و رفتار طبیعی کاربران هستند. اطلاعات پروفایل‌ها، پیوندهای موجود در آن‌ها، دیدگاه‌های منتشر یا مبادله شده، منابعی غنی برای تحلیل‌گران شبکه به شمار می‌آیند تا به الگوهای فکری، رفتاری و شاخص‌های دیگر کنش‌گری کاربران، دست یابند. در واقع تحلیل احساسات ابزار بسیار مفیدی است زیرا به ما اجازه می‌دهد تا دیدگاه عموم را در رابطه با موضوعات خاص بدست آوریم. رایج‌ترین کاربردهای تحلیل احساس، به تجارت و فروش باز می‌گردد. به‌عنوان مثال، نظراتی در مورد محبوبیت یا کیفیت یک محصول می‌توانند برای خریداران بعدی و حتی فروشندگان راهگشا باشند (۱۰). حال اینکه در بین این نظرات نیز قطب‌های مثبت، منفی و خنثی وجود دارند.

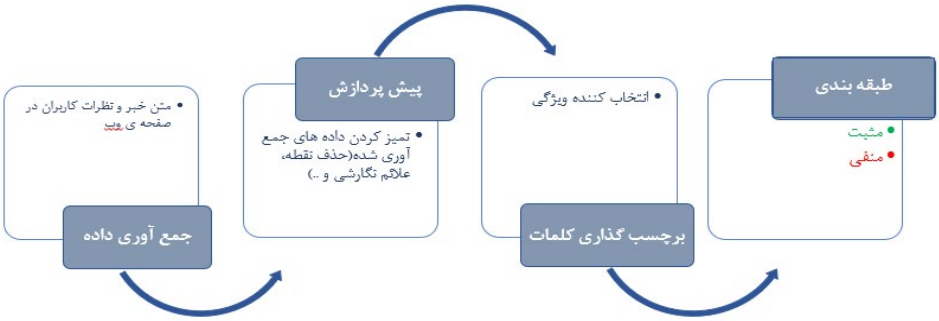
از تحلیل احساسات می‌توان در دیدگاه‌های مختلفی مانند سیاسی، اقتصادی و اجتماعی استفاده کرد. شاخص‌ترین کاربرد تحلیل احساسات در عرصه سیاست مربوط به انتخابات می‌باشد. که تیم‌های تبلیغاتی با استفاده از اطلاعات شبکه‌های اجتماعی می‌توانند رای افراد زیادی را بدست آورند. این افراد می‌توانند با استفاده از برچسب‌هایی مانند هشتگ نام یک کاندیدای ریاست جمهوری، میزان محبوبیت وی را در قیاس با رقیبان نشان دهند. آن‌ها با استفاده از این اطلاعات می‌توانند مناطق هدف به‌منظور تبلیغات بیشتر را تعیین کرده و همچنین متن سخنرانی را نیز براساس افکار ساکنان آن شهر تنظیم کنند. انتشار بیماری کرونا نیز درین روزها دستمایه سنجش و ارزیابی نگرش مردم به آن شده است. چرا که علاوه‌بر مخاطرات سلامت ناشی از همه‌گیری بیماری، اقتصاد دنیا نیز با بحران مواجه شده است. لذا با شناخت بهتر از احساس مردم نسبت به آن، می‌توان سیاستگذاری بهتری انجام داد. تجزیه و تحلیل احساسات را می‌توان به دو دسته تقسیم کرد (۱۱): دسته اول که به تجزیه و تحلیل واژگان موسوم است، هدف آن تعیین قطبیت یک سند از جهت معنایی کلمات یا عبارات موجود در اسناد است و دسته دوم که براساس یادگیری ماشین می‌باشد، شامل ساخت مدل‌هایی برای استخراج داده‌هایی از بین جملات، مثال‌ها و متن‌ها به‌منظور یافتن گرایش سند می‌باشد. تجزیه و تحلیل احساسات عمومی در هنگام شیوع یک بیماری، اطلاعات مفیدی برای اعمال پاسخ‌های مناسب بهداشت عمومی فراهم می‌کند. در طی شیوع بیماری کرونا با متوسل شدن کشورها به اعمال محدودیت‌ها و قرنطینه کردن، افراد برای بیان احساسات خود به شبکه‌های اجتماعی روی آورده‌اند و راهی برای آرام کردن خود پیدا کرده‌اند.

در ادامه مقاله‌هایی که در زمینه‌ی تحلیل احساسات ویروس کرونا منتشر شده‌اند بررسی خواهد شد. در این مقالات تجزیه و تحلیل احساسات بیماری کرونا در شبکه‌ی اجتماعی پرطرفدار توییتر، با استفاده از کلمات کلیدی مانندcovid-19, corona, coronavirus جمع‌آوری اطلاعات صورت گرفته‌است. در این مقالات با استفاده از کلمات کلیدی، هشتگ (#hashtag) پست‌ها و نظرات کاربران با استفاده از twitter API جمع‌آوری شده است. هدف از تحلیل و جمع‌آوری

۱. SARS
۲. MERS
۳. SARS-cov-2
۴. Opinion Mining

کردن این داده‌ها، بررسی میزان اهمیت دادن مردم درباره‌ی این بیماری جدید و همچنین تحلیل رویکرد دولت در مقابله با این بیماری می‌باشد. نتایج بدست آمده از این تحلیل‌ها، مثبت، منفی و خنثی می‌تواند به مردم و همچنین دولت در مقابله با این بیماری جدید، کمک کند. طبعا زمان انتشار همه این مقالات از سال ۲۰۲۰ به بعد و جدید است. شارما و همکاران (۱۲) نسبت به پژوهشی طراحی داشبوردی برای شناسایی اطلاعات غلط پیرامون ویروس کرونا در بستر شبکه اجتماعی توییتر اقدام کرده‌اند. لوین و همکاران (۱۳) در یک مطالعه با هدف بررسی روند جهانی چهار احساس ترس، عصبانیت، غم و شادی و روایت‌های اساسی این احساسات در طی بیماری همه‌گیر COVID-19, در شبکه اجتماعی توییتر با استفاده از کلمات کلیدی، "wuhan"،"nCov"، "corona"، ""covid" به پژوهش پرداخته‌اند. کاور و شارما (۱۴) نیز در مقاله‌ای احساسات مربوط به ویروس کرونا را در شبکه اجتماعی توییتر جمع‌آوری کرده‌اند. در این مقاله از زبان برنامه‌نویسی پایتون استفاده شده است. سپس با استفاده از کتابخانه NLTK پیش‌پردازش انجام شده و با استفاده از Textblob تحلیل داده‌ها انجام شده است. در پژوهشی دیگر ریسنانتویو و همکاران (۱۵) با جمع‌آوری داده‌های موجود در توییتر با عنوان"corona Virus" و انواع روش‌های یادگیری ماشین نظارت شده برای طبقه‌بندی این احساسات در دسته‌های مثبت، منفی و خنثی، به ارائه نتایج پرداخته‌اند. مطالعات مربوط به تحلیل احساسات مرتبط با کرونا تنها به موارد جهانی محدود نبوده است و به ازای کشورهای مختلف نیز به صورت مجزا، پژوهش صورت گرفته است. مطالعه‌ای در آمریکا (۱۶)، به بررسی نقش بازیگرهای کلیدی عمومی در شبکه‌های اجتماعی برای ویروس کرونا (COVID-19) می‌پردازد. نتیجه نشان می‌دهد شبکه‌های موضوع محور و ارتباطات شخص محور، نقش متفاوتی در شبکه‌های اجتماعی دارند. این مطالعه نشان می‌دهد که روسای جمهور، سازمان بهداشت جهانی (WHO) و دفاتر منطقه‌ای آن، مراکز کنترل بیماری‌ها و کانال‌های خبری اثر مهمی در اخبار COVID-19 بر مردم دارند. بازیگران اصلی مانند دونالد ترامپ، باراک اوباما و شبکه بی بی سی در مرکز چنین ارتباطاتی قرار دارند. پژوهش دیگری در چین (۱۷)، تحلیل احساسات مردم در رسانه‌ی اجتماعی محبوب چینی، سینا و ویبو را بررسی کرده است و دراین تحقیق از روش نظارت نشده BERT برای طبقه‌بندی احساسات استفاده کرده است. نمس و کیس (۱۸)، به دنبال افزایش دقت محاسبات از کلمات کلیدی "corona" و "coronavirus" در شبکه اجتماعی توییتر، داده‌ها را جمع‌آوری کرده‌اند و سپس با استفاده از پردازش زبان طبیعی و الگوریتم شبکه‌های عصبی نسبت به تجزیه و تحلیل احساسات اقدام کرده‌اند. دوبی (۱۹) نیز توییت‌های مربوط به دوازده کشور مختلف در شبکه اجتماعی توییتر را جمع‌آوری و پیش‌پردازش کرده است تا نحوه برخورد شهروندان کشورهای مختلف با اوضاع به وجود آمده را از منظر تحلیل احساس بررسی نماید. در کشور نپال نیز (۲۰) پژوهشی با استفاده از کلمات کلیدی "covid-19"و "corona virus" در شبکه اجتماعی توییتر انجام شده است که این مطالعه بر روی تحلیل احساسات توییت‌های رسانه‌های اجتماعی توییتر با استفاده از زبان برنامه‌نویسی پایتون با کتابخانه Tweepy و TextBlob متمرکز است. با توجه به عدم دسترسی به شبکه اجتماعی توییتر در ایران و در پی آن، عدم درج احساسات فراگیر جامعه ایرانی در چنین شبکه‌ای، براساس تجربه مولفین این پژوهش و بررسی‌های انجام شده، نظرات اخبار مربوط به وب‌سایت‌های خبری پربیننده داخل کشور، می‌تواند منبع خوبی برای انعکاس نظرات و احساسات کاربران و به نوعی جامعه عمومی ایرانیان نسبت به کرونا به حساب بیاید. در این مقاله به بررسی و تحلیل داده‌ها و اخبار مربوط به ویروس کرونا در دو وب‌سایت خبری پرطرفدار ایرانی ایسنا و فارس نیوز می‌پردازیم. در این مقاله از مدل یادگیری ماشین در تحلیل احساسات استفاده شده است و هدف اصلی، سنجش میزان خوشبینی یا بدبینی جامعه ایرانی تا تاریخ پیک اول شیوع کرونا بوده است تا شاید یکی از عوامل موثر برشیوع را بتوان از نظر پژوهشی شناسایی نمود.

شکل ۱. مراحل تحلیل احساسات کاربر در صفحه وب (منبع: یافته‌های پژوهش)



شکل ۲. نمونه‌ای از داده‌های جمع‌آوری شده وب‌سایت فارس نیوز

	title	body	category	pubDate	tags
0	...در شرایط کرونایی در مراسم تدفین بستگان شرکت کن	...به گزارش ایسنا ، اگر چه بزرگداشت درگذشتگا	سلامت	فروردین ۱۳۹۹ / ۰۴ ۰۷:۰۴	...کروناویروس, کووید-19, همه_یاهم_علیه_کرونا, مر]
1	...لزوم تعویق جراحی‌های غیرضروری دندانپزشکی تا مه	... به گزارش ایسنا، دکتر دالبند با تاکید بر اهمیت	سلامت	اسفند ۱۳۹۸ / ۱۳ ۰۵:۵۰	[کروناویروس, دندانپزشکی, همه_یاهم_علیه_کرونا]
2	...خواسته‌ای از مردم برای تعطیلی مراسم شادی و تر	...به گزارش ایسنا، دکتر کیانوش جهانپور، رییس مرکز	سلامت	اسفند ۱۳۹۸ / ۱۴ ۱۴:۳۶	[کروناویروس, همه_یاهم_علیه_کرونا]
3	... در بحران اخبار کرونا چگونه بر اضطراب خود علیه	... به گزارش ایسنا، با توجه به شرایط موجود به سبب	سلامت	اسفند ۱۳۹۸ / ۱۵ ۰۵:۳۰	[کروناویروس, اضطراب]
4	...توصیه تغذیه‌ای به بیماران تنفسی در روزهای شی ۶	...به گزارش ایسنا، دفتر بهبود تغذیه جامعه وزارت ب	سلامت	اسفند ۱۳۹۸ / ۲۱ ۰۸:۲۰	...وزارت بهداشت, کروناویروس, بیماری تنفسی, تغذیه]
...	...	...	...	...	...
401	...لیگ قهرمانان اروپا تک بازی شد/ لیسبون میزبان ف	...با اعلان یوفا در SID، به گزارش ایسنا به نقل از	جهان ورزش	خرداد ۱۳۹۹ / ۲۸ ۱۹:۳۸	[یوفا]
402	...دستگاه قضا با آتش افروزان منابع طبیعی برخورد م	...محمد قاسمی، دادستان عمومی و انقلاب قزوین در خص	قزوین	خرداد ۱۳۹۹ / ۲۸ ۱۹:۳۸	[استانی-اجتماعی, آتش سوزی مراتع]
403	...روحانی:دولت در کنار خانواده‌های معظم شهداء جان	...به گزارش ایسنا به نقل از پایگاه اطلاع‌رسانی ر	دولت	خرداد ۱۳۹۹ / ۲۸ ۲۰:۰۱	[رئیس جمهور]
404	یادداشت جابری انصاری درباره استاد نجیب مایل هروی	...به گزارش ایسنا متن یادداشت جابر انصاری که در ک	سیاست خارجی	خرداد ۱۳۹۹ / ۲۸ ۱۹:۰۹	[حسین جابری انصاری]
405	توقف استقلال مقابل رایکا با ۲ مصدوم	...به گزارش ایسنا، تیم فوتبال استقلالف امروز (چها	فوتبال، فوتسال	خرداد ۱۳۹۹ / ۲۸ ۲۰:۰۹	[تیم استقلال, تیم رایکا]

406 rows × 5 columns

روش کار

مراحل محاسباتی تحلیل احساس در شکل۱، آمده است. این مراحل، اصول کلی اجرای یک سیستم تحلیل‌گر است. پس از جمع‌آوری داده و تبدیل آن به قالب مناسب برای انجام محاسبات، با اجرای روش‌های یادگیری ماشین، کلمات مناسب برای تعیین احساس، شناسایی شده و در نهایت طبقه‌بندی نظرات بر اساس دارا بودن بار مثبت یا منفی، انجام می‌گیرد. در ادامه جزییات دقیق‌تری از هر مرحله بیان می‌گردد.

جمع‌آوری داده: نظر کاربر، معیار اصلی بهبود کیفیت خدمات ارایه شده و بهبود نتایج است. در نبود فعالیت عمده یا دسترسی به داده‌های شبکه‌های اجتماعی رایج جهانی در کشور، وب‌سایت‌های خبری پرطرفدار، از مهم‌ترین منابع برای جمع‌آوری داده می‌باشند. با افزایش استفاده از اینترنت، کاربران نظرات و دیدگاه‌های خود را تحت عنوان پیام‌های متنی در وب‌سایت‌ها ایجاد می‌کنند. این پیام‌ها گاهی نظرات متفاوتی را در مورد موضوعات مختلف بیان می‌کنند. این پیغام‌ها نیز به عنوان منبع داده برای طبقه‌بندی احساسات مورد استفاده قرار می‌گیرند. در این مقاله داده‌ها از اوایل شیوع ویروس کرونا در ایران یعنی در اسفند ۱۳۹۸ تا خرداد ماه ۱۳۹۹ در دو وب‌سایت خبری داخلی، ایسنا و فارس نیوز جمع‌آوری شده است. از زبان پایتون برای جمع‌آوری داده‌ها و تحلیل احساسات مورد استفاده قرار گرفته است. یکی از کتابخانه‌های معروف پایتون در خواندن و استخراج اطلاعات مفید از صفحات وب، کتابخانه "scrapy" است. کتابخانه scrapy یک چارچوب کامل برای خزش[۵] در وب با قابلیت دنبال کردن لینک‌های موجود در هر صفحه و استخراج تمام لینک‌های یافت شده است. در وب‌سایت‌های ایسنا و فارس نیوز به دلیل حجم بالای خبرها از یک فیلتر تاریخ استفاده شده است که اخبار را در بازه زمانی مطلوب مورد بررسی قرار دهد. تعداد داده‌های جمع‌آوری شده نزدیک به ۵۰۰ خبر و نظر است که الگوریتم‌های یادگیری ماشین برای تحلیل احساس بر روی آن‌ها پیاده‌سازی شده است. برای این مقاله، از هر خبر، متن، تاریخ، گروه و عنوان آن برای شناسایی مرتبط بودن، مورد بهره‌برداری قرار گرفته است.

پیش‌پردازش داده‌ها: در این گام براساس زبان برنامه‌نویسی پایتون و با استفاده از کتابخانه‌ی هضم[۶] که برای متون به زبان فارسی استفاده می‌شود، پیش‌پردازش توسط کتابخانه معروف NLTK بر روی داده‌ها انجام می‌شود. اسناد متنی از نوع غیرساختار یافته هستند و به همین دلیل نیاز است تا این نوع داده‌ها با عملیات مختلف (پیش‌پردازش) به داده‌های ساختاریافته و قابل محاسبات برای کامپیوتر تبدیل شوند. برای این کار از دو روش BoW[۷] و TF-IDF می‌توان استفاده کرد. BoW به زبان ساده تکرار کلمات مورد نظر در مستند را مد نظر دارد تا متن را برای پردازش یادگیری ماشین آماده کند در حالی که TF-IDF[۸] فراوانی نسبی کلمات در متن را نیز در نظر می‌گیرد.

انتخاب ویژگی: نظراتی که در یک شبکه اجتماعی و یا در یک فروشگاه اینترنتی راجع به یک محصول یا رویداد داده می‌شود ممکن است مثبت، منفی و یا خنثی باشد. این قطبیت به یک ویژگی یا جنبه از یک محصول یا موضوع نسبت داده می‌شود. برای مثال در متن جمله: استعمال دخانیات باعث افزایش خطر ابتلا به کرونا می‌شود. ویژگی یا جنبه‌های این موضوع «استعمال دخانیات» و «ابتلا به کرونا» است.

طبقه‌بندی: پردازش اصلی جهت تعیین قطبیت یک جمله یا یک

۵. crawling

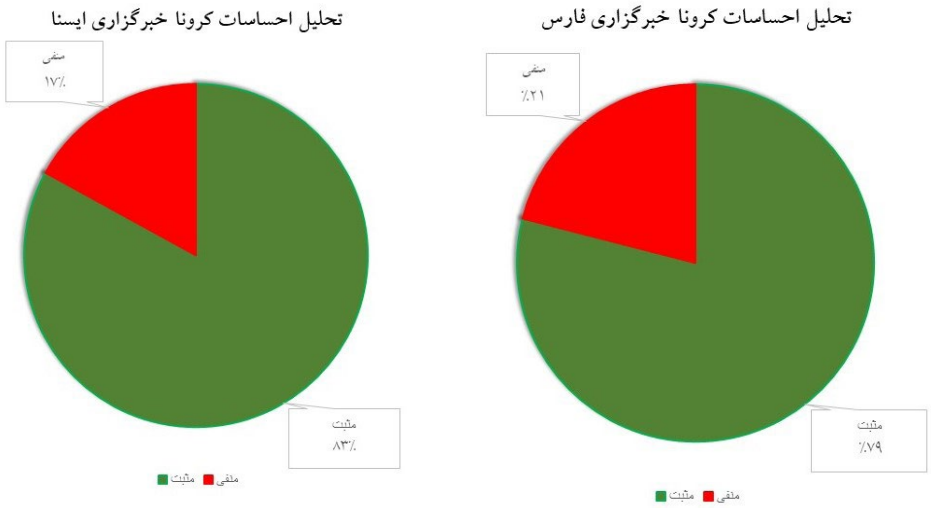
۶. https://www.sobhe.ir/hazm/
۷. Bag of words
۸. Term Frequency - Inverse Document Frequency

سند در این مرحله صورت می‌گیرد. طبقه‌بندی احساسات بر دو پایه اساسی استوار است (۹): واژه‌نامه و یادگیری ماشین. در طبقه‌بندی مبتنی‌بر واژگان که از فرهنگ لغت خاص استفاده می‌شود، روال بدین‌صورت می‌باشد که برای هر کلمه در یک جمله مانند اسامی، فعل‌ها، صفات و قیود موجود، در شبکه واژگان احساسی، صورت می‌گیرد و در صورتی که آن واژه در شبکه موجود باشد، برچسب احساسی آن استخراج می‌شود. هر واژه متناسب با قطبیتی که دارد «مثبت»، «منفی» و یا «خنثی» است. در این مقاله برای دریافت واژگان احساسی از وب‌سایت متن کاوی فارسی‌یار استفاده شده است. با توجه به اخبار و نظرات موجود در وب‌سایت‌های ایسنا و فارس‌نیوز، مردم بیشتر احساسات خود را با کلمات «مرگ»، «تدفین»، «خطر»، «استرس»، «ترس»، «بدون درمان»، «واگیر»، «ناراحت»، «شیوع»، «نگرانی»، «بحران»، «افسوس»، «وای»، «حیف»، «اضطراب» و «کمک» بیان می‌کردند. به همین دلیل از این کلمات با بار احساسی منفی در تحلیل احساس استفاده شده است و برای تحلیل احساسات مثبت نیز کلمات «شادی»، «واکسن» و «رعایت» مورد استفاده قرار گرفته است.

طبقه‌بندی احساسات به سه دسته تقسیم می‌شود. روش‌های یادگیری نیمه نظارتی، بدون نظارت و نظارت (۲۱ ، ۲۲). یادگیری نظارت‌شده فرایند یادگیری الگوریتم از مجموعه داده‌های آموزش برای دستیابی به خروجی‌های مطلوب است درحالیکه یادگیری بدون نظارت فرایند یادگیری بدون داشتن خروجی‌های مورد انتظار است. در یادگیری نیمه‌نظارتی به‌طور همزمان از داده‌های برچسب‌دار و بدون برچسب استفاده می‌شود تا بتوان دقت یادگیری را بهبود بخشید. رویکرد یادگیری ماشینی که برای تجزیه و تحلیل احساسات کاربرد دارد، بیشتر به‌طبقه‌بندی نظارت‌شده به‌طورکلی و تکنیک‌های طبقه‌بندی متن به‌طور خاص تعلق دارد. بنابراین، آن را «یادگیری نظارت‌شده» می‌نامند. در طبقه‌بندی مبتنی‌بر یادگیری ماشین، دو مجموعه از اسناد مورد نیاز است: آموزش و مجموعه آزمون. یک مجموعه آموزشی توسط یک طبقه‌بندی اتوماتیک برای یادگیری ویژگی‌های تمایز اسناد استفاده می‌شود و از یک مجموعه تست برای اعتبارسنجی عملکرد طبقه‌بندی خودکار استفاده می‌شود. در این مقاله ما از الگوریتم یادگیری ماشین درخت تصمیم برای تحلیل احساسات استفاده کرده‌ایم.

درخت تصمیم‌گیری یک روش رایج برای طبقه‌بندی است که جز الگوریتم‌های یادگیری ماشین نظارت شده می‌باشد. فرآیند یادگیری براساس قوانین ساده تصمیم‌گیری است که از ویژگی‌های داده بدست آمده‌اند (۲۳). این رویکرد از یک مدل درخت مانند برای نشان دادن تصمیم‌گیری استفاده می‌کند (۲۴). هر گره در درخت، یک ویژگی را نشان می‌دهد، هر شاخه تصمیم‌گیری، و هر برگ خروجی را نشان می‌دهد. در استفاده از درخت تصمیم‌گیری مهم است که مشخص شود کدام مشخصه‌ها می‌توانند انتخاب شوند، چه شرایطی برای تقسیم و تصمیم‌گیری و چه زمانی مناسب برای توقف است. در طی این فرآیند، تمام ویژگی‌ها در نظر گرفته می‌شوند و نقاط انشعاب مختلف امتحان و تست می‌شوند. برای شناسایی بهترین مسیر، حداقل هزینه انتخاب می‌شود. با هرس می‌توان عملکرد درخت را بیشتر بهبود بخشید. هرس فرایندی برای از بین بردن ویژگی‌هایی است که از اهمیت کمتری در جهت کاهش پیچیدگی درخت و در نتیجه افزایش قدرت پیش‌بینی آن با کاهش بیشتر انشعابات همراه است. کاور در پژوهشی از مدل درخت تصمیم‌گیری برای تحلیل واریانس عواطف انسانی استفاده کرده است (۲۵). از آنجا که انواع مختلف احساسات را می‌توان آنالیز و براساس درخت

شکل ۳. نتایج تحلیل احساس مرتبط با کرونا در نظرات کاربران خبرگزاری‌های ایسنا و فارس نیوز



شکل ۴. تحلیل احساسات مرتبط با کرونا در کشورهای مختلف



تصمیم‌گیری دسته‌بندی کرد، کاور از تجزیه و تحلیل طرد شدن برای کودکان با معلولیت مختلف برای شناسایی احساسات استفاده کرده بود. در این پژوهش برای پیاده‌سازی الگوریتم درخت تصمیم از زبان برنامه‌نویسی پایتون و کتابخانه‌های «Decision Tree» و «Sklearn» استفاده شد.

یافته ها

در این قسمت نتایج تحقیق در دو مرحله مورد بحث قرار می‌گیرد. در مرحله‌ی اول نتایج حاصل از تحلیل وب‌سایت‌های داخلی مورد نظر مورد بحث و بررسی قرار می‌گیرد. سپس در مرحله بعدی نتایج به‌دست آمده، با کشورهای دیگر مقایسه می‌شود. مطابق شکل ۳ با توجه به تحلیل‌های انجام شده در قسمت قبلی در دو وب‌سایت خبری داخلی ایسنا و فارس نیوز درصد نظرات مثبت و منفی نشان داده شده است. میزان دقت ارزیابی شده برای الگوریتم درخت تصمیم حدود ۹۰ درصد به دست آمده است.

در بازه زمانی اسفند ۹۸ تا خرداد ۹۹، در وب‌سایت فارس نیوز ۷۹ درصد نظرات مثبت و۲۱ درصد از آن‌ها منفی محاسبه شد. در حالی که در وب‌سایت ایسنا درصد نظرات مثبت ۸۳ درصد و نظرات منفی ۱۷ درصد به دست آمده است. بدین ترتیب به‌طور متوسط نزدیک به ۸۰ درصد از کسانی که درباره اخبار مربوط به کرونا در دو وبسایت ذکر شده، نظرشان را یادداشت کرده بودند، نسبت به این اخبار نظر مثبت یا خوش‌بین داشتند.

شکل ۴ احساسات منعکس شده درباره کرونا را در توییت‌های کشورهای مختلف نشان می‌دهد (۲۶). همان‌طور که در شکل قابل مشاهده است، پس از ایران، دومین احساس مثبت منعکس شده مربوط به نپال کشوری در آسیا در مجاورت هندوستان می‌باشد که ۸۵ درصد احساسات آن‌ها نسبت به کرونا مثبت بوده است و همچنین سومین احساس مثبت با توجه به پژوهش‌های انجام شده مربوط به کشور بلژیک در اروپاست که با ۶۳ درصد احساس مثبت و ۳۷ درصد احساس منفی، خوش‌بینی مردم آن کشور نسبت به کرونا را نشان می‌دهد.

برای آگاهی بیشتر از ارتباط احتمالی نگرش مثبت درباره کرونا با رعایت و مرگ و میر شهروندان کشورهای خوش‌بین، در جدول ۱، آمار فوت شدگان در بین کشورهایی که بیشترین احساس مثبت در نظرات مردم آن‌ها گزارش شده است، آمده است(۲۷).

جدول ۱. آمار جهانی بیماری کرونا تا هفتم نوامبر ۲۰۲۰

ردیف	کشورها	مبتلایان	فوت شدگان	درصد
۱	ایران	۶۵۴,۹۳۶	۳۶,۹۸۵	۵/۶۵
۲	استرالیا	۲۷,۶۳۳	۹۰۷	۳/۲۸
۳	بلژیک	۴۷۹,۲۴۶	۱۲,۵۲۰	۲/۶۱
۴	هند	۸,۴۱۱,۷۲۴	۱۲۴,۹۸۵	۱/۴۹
۵	نپال	۱۸۵,۹۷۴	۱,۰۵۲	۰/۵۶

با توجه به جدول ۱ کشور ایران در بین کشورهای نپال و بلژیک که دارای بیشترین احساس مثبت نسبت به بیماری کرونا هستند دارای بیشترین آمار فوتی است، اما همگرایی بین درصد فوت شدگان کشورهای خوش‌بین نسبت به کرونا مشاهده نشد.

کشورهای آلمان، فرانسه، سوئیس و ایالات متحده آمریکا کشورهایی هستند که تقریبا تعادل بین احساسات مثبت و منفی در بین آن‌ها وجود دارد. برعکس، چین توییت‌های با احساس منفی (۵۵ درصد) بیشتری نسبت به دیگر کشورها داشت و مرگ و میر در آن به تدریج روند ثابت نزولی به خود گرفت. در مجموع در شکل ۴، در بیشتر کشورهای مورد بررسی، شهروندان نظر مثبت نسبت به کرونا داشته‌اند.

بحث و نتیجه‌گیری

این پژوهش با هدف تحلیل احساسات و عواطف مردم ایران در طی بیماری همه‌گیر کووید-۱۹ انجام شد. در این مطالعه، مشخص شد علاوه بر ایرانیان، نپال، بلژیک، هند و استرالیا به ترتیب بیشترین احساس مثبت را در مورد کرونا داشتند و کشور چین نیز از بیشترین احساس منفی در مورد کرونا برخوردار بوده است. هر چند عوامل مختلفی در درصد مرگ و میر بالای ایران نسبت به سایر کشورهای جهان در بیماری کرونا دخیل هستند، اما یک علامت مهم کشف شده از این پژوهش می‌تواند، جدی نگرفتن و کم اهمیتی نسبت به بیماری کرونا در ابتدای ورود آن به ایران را نشان دهد. نسبت ۸۰ درصدی خوش‌بینی برای بیماری کرونا در کشور ایران، می‌تواند به موارد متعددی وابستگی داشته باشد که به نوبه خود از طریق پژوهش‌های دیگر، قابل بررسی است.

از موانع اصلی در تحلیل احساس مردم ایران نسبت به کرونا و افزایش قابلیت مقایسه آن با سایر کشورهای جهان، دسترسی به بستر اطلاعاتی رایج در فضای مجازی است. با توجه به عدم دسترسی عوام مردم در داخل ایران، به شبکه توییتر، شاید مقایسه نظرات افراد در وبسایت‌های داخلی، با نظرات مردم سایر کشورها در توییتر، همسان و متوازن نباشد. از سوی دیگر استفاده تعداد زیادی از مردم ایران از پیام‌رسان‌های خارجی فیلتر شده و اقبال کم به پیام‌رسان‌های داخلی، امکان پژوهش بر بسترهای رایج فناوری مورد استفاده عوام را کم می‌کند.

در زمینه‌ی تحلیل احساسات متون به زبان فارسی، محققان به دلیل پیچیدگی‌های خاص آن با چالش‌هایی روبه‌رو هستند که عبارتند از، بسیاری از حروف در زبان فارسی سه یا چهار فرم نوشتن دارد و تعداد استثنائات در کلمات زیاد است، وجود پسوند و پیشوندهای مختلف و همچنین بعضی از آواها که از زبان عربی وارد زبان فارسی شده‌اند، ممکن است در بعضی از مواقع نوشته و در برخی نیز نوشته نشوند مانند: «مساله، مسئله، مسأله». از دیگر مشکلات، استفاده کاربران از کلمات غیررسمی و عامیانه به‌جای نوشتن رسمی است. هم‌چنین ممکن است نوشته‌های خود را همان‌طور که صحبت می‌کنند تایپ کند. بهبود کارایی در تحلیل احساس در زبان فارسی سبب افزایش کیفیت نتایج خواهد شد. از میان مهم‌ترین مباحثی که در تحلیل احساسات خصوصا در زبان فارسی امکان پژوهش بیشتری دارد می‌توان این موارد را برشمرد: کار بر روی پیدا کردن ریشه کلمات در زبان‌فارسی همواره مبحثی پر از چالش است، تحقیق بیشتر در این زمینه باعث افزایش دقت انتخاب ویژگی و طبقه‌بندی خواهد شد. با گسترش و کاربردی‌تر شدن ابزارهای مدیریت کلان‌داده در دنیا، می‌توان بسیاری از روش‌های تشخیص احساسات در زبان فارسی را در ابزارهای خاص آن مانند: هدوپ و.... پیاده‌سازی نمود.

بررسی تغییرات نگرشی و احساسات مردم ایران نسبت به کرونا با گذر زمان، می‌تواند از پژوهش‌های ارزشمند دیگری باشد که در ادامه این تحقیق، قابل انجام است. به ویژه با شناخت بهتر از این بیماری و ابتلای بخش‌های بزرگتری از جامعه، ممکن است نگرش و احساس مردم ایران نیز نسبت به کرونا تعدیل یا تغییر کرده باشد. به عبارت دیگر با توجه به دسترسی داده‌های نظرات کاربران در تاریخ‌ها و بازه‌های زمانی مختلف، امکان تحلیل تغییرات نگرشی در این جامعه، سبب شناخت بهتر و دست‌آوردهای ارزشمندتر پژوهشی خواهد شد.

منابع

1. Mahase E. Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. BMJ.2020;368:m641.

2. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol. 2016;14(8):523-34.

3. Hilgenfeld R, Peiris M. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. Antiviral Res. 2013 Oct;100(1):286-95.

4. Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. Int J Antimicrob Agents. 2020 Feb 17:105924.

5. Worldometers. COVID-19 CORONAVIRUS OUTBREAK. World Meters. 2020. [cited 2020 March 14]. Available from: https://www.worldometers.info/coronavirus /

6. MOHME. Ministry of Health and Medical Education. Tehran, Iran. 2020. [cited 2020 March 14]. Available from: air.ir/ZwqjJao

7. TasnimNews(2020) اخبار اقتصادی -Available from: https://www.tasnimnews.com/.

8. 8. J. Pei, G. Yu, X. Tian and M. R. Donnelley, “A new method for early detection of mass concern about public health issues”, J. Risk Res., vol. 20, no. 4, pp. 516-532, Apr. 2017

9. B. Liu, Sentiment analysis and opinion mining. San Rafael: Morgan & Claypool, 2012.

10. L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, “A survey of sentiment analysis in social media,” Knowledge and Information Systems, pp. 1-47, 2018.

11. I. El Alaoui, Y. Gahi, R. Messoussi, Y. Chaabi, A. Todoskoff and A. Kobi, “A novel adaptable approach for sentiment analysis on big social data”, Journal of Big Data, vol. 5, no. 1, 2018.

12. K.Sharma, S.Seo, Ch.Meng, S.Rambhatla, Y.Liu.(October 23, 2020). COVID-19 ON SOCIAL MEDIA: ANALYZING MISINFORMATION IN TWITTER CONVERSATIONS. https://arxiv.org/pdf/2003.12309.pdf

13. M.O.Lwin, i.Lu, A.Sheldenkar, P.J.Schulz, W.Shin, R.Gupta, Y.Yang, (2020 Apr-Jun). Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. https://bit.ly/3cvEgqx

14. Ch.Kaur, A.Sharma. (16 March 2020). Twitter Sentiment Analysis on Coronavirus using Textblob https://bit.ly/2FZcF5o.

15. R.Risnantoyo, A. Nugroho, K. Mandara.(20 July 2020). SENTIMENT ANALYSIS ON CORONA VIRUS PANDEMIC USING MACHINE LEARNING ALGORITHM.Available from: https://ojs.uma.ac.id/index.php/jite/article/view/3798

16. S.Yum. (28 May 2020). Social Network Analysis for Coronavirus (COVID-19) in the United States. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/ssqu.12808

17. T. WANG, K.LU, K.P.CHOW, A.Q.ZHU. (August 7, 2020). COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model . https://ieeexplore.ieee.org/abstract/document/9151169

18. László.Nemes, A.kiss. (30 Jun 2020). Social media sentiment analysis based on COVID-19. https://bit.ly/36k8vjt

19. A.D.Dubey. (April 9, 2020). Twitter Sentiment Analysis during COVID-19 Outbreak. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3572023

20. B.P.Pokharel.(11 june 2020) Twitter Sentiment Analysis During Covid-19 Outbreak in Nepa. https://bit.ly/3i3qw7l.

21. S. Shayaa et al., “Sentiment Analysis of Big Data : Methods, Applications , and Open Challenges,” IEEE Access, vol. PP, no. c, p.1, 2018.

22. M. Tubishat, N. Idris, and M. A. M. Abushariah, “Implicit aspect extraction in sentiment analysis: Review , taxonomy , opportunities , and open challenges,” Inf. Process. Manag., vol. 54, no. 4, pp. 545–563, 2018.

23. S. Shayaa et al., “Sentiment Analysis of Big Data : Methods, Applications , and Open Challenges,” IEEE Access, vol. PP, no. c, p.1, 2018.

24. V. M. Pradhan, J. Vala, and P. Balani, “A Survey on Sentiment Analysis Algorithms for Opinion Mining,” Int. J. Comput. Appl., vol.133, vol. 133 no. 9, pp. 7-11, 2016.

25. J. Kaur, “Analysis and Indentifying Variation in Human Emotion Through Data Mining,” Int. J. Comput. Technol. Appl., vol. 3, no. 6, pp. 1963–1966, 2012

26. Dr. Akash D Dubey,” Twitter Sentiment Analysis during COVID19 Outbreak”, Jaipur, 2020.

27. Worldometers- COVID-19 CORONAVIRUS PANDEMIC (2020). Available from: https://www.worldometers.info/coronavirus/