

Sentiment Analysis of Corona-Related Tweets in Iran Using Deep Neural Network

**Mohammad Ehsan
Basiri**  *

Assistant Professor, Department of Computer
Engineering, Shahrekord University, Shahrekord,
Iran..

Shirin Habibi 

MSc., Computer Engineering, Shahrekord
University, Shahrekord, Iran.

Shahla Nemati 

Assistant Professor, Department of Computer
Engineering, Shahrekord University, Shahrekord,
Iran.

Abstract

With the spread of Covid-19 disease, quarantine, and social isolation, people are increasingly posting their opinions about the coronavirus on social networks such as Twitter. However, no study has yet been reported to analyze online opinions of individuals in order to understand their feelings about the Covid-19 epidemic in Iran. This study analyzes the emotions in the opinions of the Iranian people on the social network Twitter during the Corona crisis. For this purpose, a deep neural network model is presented. As there is no labeled dataset of Covid-19 tweets, the proposed model is first trained on the Stanford University Sentiment140 dataset, which contains 1.6 million tweets, and then used to classify the two classes of emotions contained in the collected corona-related tweets in Iran. The results show that the percentage of tweets with negative emotions is significantly higher than positive tweets. Also, the change in negative emotions of people in different months is proportional to the change in patient statistics.


Keywords: Corona Virus, Covid-19, Sentiment Analysis, Opinion Mining, Deep Neural Network.


* Corresponding Author: basiri@sku.ac.ir


How to Cite: Basiri, M, E., Habibi, S., Nemati, S., (2021). Sentiment Analysis of Corona-Related Tweets in Iran Using Deep Neural Network, *Journal of Business Intelligence Management Studies*, 10(37), 109-134.



تحلیل احساسات توئیت‌های مرتبط با کرونا در ایران با استفاده از شبکه عصبی عمیق

محمد احسان بصیری *  استادیار گروه مهندسی کامپیوتر دانشگاه شهرکرد، شهرکرد، ایران.

شیرین حبیبی  کارشناسی ارشد مهندسی کامپیوتر، دانشگاه شهرکرد، شهرکرد، ایران.

شهلا نعمتی  استادیار گروه مهندسی کامپیوتر دانشگاه شهرکرد، شهرکرد، ایران.

چکیده

با همه گیر شدن بیماری کووید-۱۹، قرنطینه شدن مردم و فاصله گذاری اجتماعی، افراد بیش از پیش نظرات خود درباره ویروس کرونا را در شبکه‌های اجتماعی مانند توئیتر منتشر می‌کنند. با این حال، هنوز مطالعه‌ای برای تحلیل نظرات برخط افراد به منظور درک احساسات آن‌ها در مورد همه‌گیری کووید-۱۹ در ایران گزارش نشده است. در این پژوهش به تحلیل احساسات موجود در نظرات مردم ایران در شبکه اجتماعی توئیتر در طول بحران کرونا پرداخته می‌شود. برای این منظور یک مدل شبکه عصبی عمیق ارائه می‌شود. با توجه به این که داده‌های برجسب گذاری شده از توئیت‌های مرتبط با کرونا در دسترس نیست، مدل پیشنهادی ابتدا روی مجموعه داده Sentiment140 دانشگاه استنفورد شامل یک میلیون و ششصد هزار توئیت آموزش داده شده، سپس برای طبقه‌بندی دو کلاسه‌ی احساسات موجود در توئیت‌های جمع‌آوری شده مرتبط با کرونا در ایران استفاده می‌شود. نتایج نشان می‌دهد درصد توئیت‌ها دارای احساسات منفی نسبت به توئیت‌های مثبت به شکل معنی‌داری بیشتر است. همچنین، تغییر احساسات منفی افراد در ماه‌های مختلف متناسب با تغییر در آمار بیماران می‌باشد.

کلیدواژه‌ها: ویروس کرونا، کووید-۱۹، تحلیل احساسات، نظرکاوی، شبکه عصبی عمیق.

مقدمه

گسترش ناگهانی بیماری کووید-۱۹ در سال گذشته با انتقال ویروس مشاهده شده در جمعیت حیوانات و سرایت آن به انسان، دانشمندان و سیاست‌گذاران را شگفت‌زده کرده و آن‌ها را تحت تأثیر قرار داد. بیماری کووید-۱۹ توسط سازمان بهداشت جهانی به‌عنوان یک بیماری همه‌گیر شناخته شده است که زندگی افراد را در سراسر جهان به خطر می‌اندازد (ولاولان و میر^۱، ۲۰۲۰). در برخی از مناطق موارد ابتلای کمی گزارش شده است در حالی که در برخی دیگر چند هزار نفر مبتلا وجود دارد و در مناطقی که دارای انتقال گسترده و غیرقابل کنترل هستند، میلیون‌ها نفر تاکنون مبتلا شده‌اند. برای به حداقل رساندن تعداد افراد آلوده، چندین استراتژی از جمله اقدامات فاصله‌گذاری اجتماعی، ممنوعیت سفر، قرنطینه و تعطیلی مشاغل اجرا شده است که به سرعت ساختار جوامع در سراسر جهان را تغییر می‌دهد (پرادهام و همکاران^۲، ۲۰۲۰). این وضعیت روزبه‌روز پیشرفت کرده است؛ بنابراین، انتشار سریع و به‌موقع اطلاعات اولویت اصلی کنترل و پیشگیری از بیماری است. علاوه بر این، تهدید بیماری به طرق مختلفی بر نحوه تفکر و رفتار افراد تأثیر می‌گذارد. برخی از تغییرات مشاهده‌شده طی ماه‌های گذشته اثرات قابل توجهی بر سلامت جامعه داشته و در نهایت تأثیر مهمی بر اقتصاد جهانی خواهد داشت. در مواجهه با این وضعیت بی‌سابقه، دولت‌ها راه‌حل‌های متفاوتی در زمینه بهداشت عمومی و اقتصادی ارائه داده‌اند (کیو و همکاران^۳، ۲۰۲۰). در نتیجه شاهد تغییر ساختار چشمگیری در نظام اقتصادی و اجتماعی هستیم که در گذشته تجارت و جامعه در آن به‌صورت سنتی فعالیت داشته‌اند.

با قرنطینه کردن مردم و فاصله‌گذاری اجتماعی، اکنون بحث‌های زیادی درباره ویروس کرونا به‌صورت برخط، در شبکه‌های اجتماعی مانند توئیتر وجود دارد. با توجه به مطالعاتی که درباره کاربردهای شبکه‌های اجتماعی برای تحلیل موضوعات مختلف مانند انتخابات و تغییرات قیمت در بورس انجام شده (فخری و حسین‌زاده ۱۳۹۶) تعداد

1. Velavan & Meyer

2. Pradhan et al.

3. Qiu et al.

فزاینده‌ای از مردم به بستر شبکه‌های اجتماعی اعتماد می‌کنند (کوثری لنگری و همکاران، ۱۳۹۸). با این حال، هنوز مطالعه‌ای برای درک احساس و تحلیل نظرات مردم در مورد همه‌گیری کووید-۱۹ گزارش نشده است. تحلیل احساسات نظرات کاربران یک کاربرد بین‌رشته‌ای است که با به‌کارگیری روش‌های هوشمند سعی در استخراج احساسات و تمایلات افراد از نوشته‌های آن‌ها دارد (عباسی و همکاران، ۱۳۹۶).

هدف از پژوهش حاضر، تحلیل احساسات بیان‌شده در توئیتهای مردم ایران برای تعیین چگونگی تغییر انتظارات، ادراکات و رفتارهای آن‌ها در طول بحران به‌مرورزمان است. برای این منظور در پژوهش حاضر، ابتدا یک مجموعه داده شامل توئیتهای مرتبط با ویروس کرونا در بازه زمانی دی ماه سال نود و هشت تا انتهای مرداد ماه سال نود و نه ایجاد شد. برای تحلیل احساسات و طبقه‌بندی توئیتهای مثبت و منفی همچنین مشخص کردن شدت میزان مثبت و منفی بودن نظرات یک مدل مبتنی بر شبکه عصبی عمیق ارائه شد. در این مدل برای استفاده بهتر از توالی کلمات در جمله و در نظر گرفتن زمینه متن از هر دو جهت (بصیری و همکاران^۱، ۲۰۲۰)، از واحدهای دوطرفه بازجریانی دروازه‌دار (BiGRU) استفاده می‌شود. سپس به این دلیل که توئیتهای موجود فاقد برچسب هستند، از یک مجموعه داده استاندارد در مقیاس بزرگ به نام Sentiment140 که متعلق به دانشگاه استنفورد است برای آموزش مدل پیشنهادی استفاده شد. در نهایت، توئیتهای مربوط به هر روز با استفاده از مدل پیشنهادی از نظر شدت مثبت یا منفی بودن احساسات بیان شده در آن‌ها دسته‌بندی شدند. علاوه بر این، داده‌های مربوط به آمار موارد جدید و مرگ‌ومیر ناشی از کرونا به همراه روند جستجوی کلمات مرتبط با کرونا در گوگل برای کشور ایران استخراج شد و در تکمیل نتایج حاصل از تحلیل احساسات با استفاده از مدل پیشنهادی به کار رفته شدند.

بر اساس مطالب ذکرشده، پژوهش حاضر با هدف پاسخ به سؤالات زیر در مورد ویروس کرونا در ایران انجام شده است:

• در جستجوی گوگل مربوط به ویروس کرونا چه ترکیباتی از کلمات بیشتر دیده شده‌اند؟

• آیا تفاوتی در شدت احساسات نشان داده شده در توئیت‌های مربوط به ویروس کرونا وجود دارد؟

• آیا بین تعداد موارد جدید مبتلا و مرگ‌ومیر با احساسات توئیت در ماه‌های مختلف ارتباط معنی‌داری وجود دارد؟

پاسخ به سؤال اول می‌تواند به کاربران و همچنین توسعه‌دهندگان نرم‌افزارهای مرتبط با تحلیل بیماری کرونا در جمع‌آوری بهتر داده‌های مرتبط کند. سؤال دوم در راستای مشخص کردن تأثیر شدت بیماری بر احساس کلی حاکم بر جامعه در زمان‌های مختلف مطرح شده است. با مشخص شدن تغییرات موجود در شدت احساسات روانشناسان می‌توانند راهکارهای مناسب‌تری برای بهبود کیفیت زندگی در هنگام شیوع بیماری‌های فراگیر پیشنهاد دهند. نهایتاً پاسخ به سومین سؤال می‌تواند تأثیر گسترش و تلفات بیماری بر احساسات حاکم بر فضای مجازی را مشخص نماید. این مسئله در کنترل بهتر احساسات و مدیریت ارائه‌ی آمارهای مربوط به بیماری می‌تواند مفید باشد.

نتایج این مطالعه به شناسایی عواملی که واکنش‌های عمومی به ویروس کرونا را در بستر فوریت‌های پزشکی عمومی که به سرعت در حال رشد است، کمک می‌کند. یافته‌های این مطالعه همچنین به راهنمایی مقامات بهداشت عمومی برای برقراری ارتباط مناسب با مردم و کمک به ارائه پاسخ‌های بهداشت عمومی برای کسانی که مستعد ابتلا به ویروس هستند کمک خواهد کرد.

نوآوری‌های اصلی این پژوهش به شرح زیر است:

• یک مدل شبکه عصبی عمیق مبتنی بر واحدهای دوطرفه بازجریانی دروازه‌دار و بردار از پیش آموزش‌دیده تعبیه کلمات برای تجزیه و تحلیل احساسات توئیت‌ها پیشنهاد می‌شود. این مدل در مجموعه داده‌های موجود در مقیاس بزرگ توئیت آموزش داده شده است.

- مجموعه داده‌های جدید توئیتر درباره توئیتهای مربوط به ویروس کرونا در ایران جمع‌آوری و تجزیه و تحلیل می‌شود.
 - علاوه بر مجموعه داده‌ی توئیتر، داده‌های روند گوگل برای کشور ایران برای یافتن الگوهای معنی‌دار مربوط به ویروس کرونا تجزیه و تحلیل شده است.
- در ادامه مقاله با مروری بر پژوهش‌های پیشین ادامه خواهد یافت. پس از آن داده‌های استخراج شده توصیف خواهند شد و در بخش بعدی مدل پیشنهادی معرفی خواهد شد. پس از آن تجزیه و تحلیل داده‌ها ارائه می‌شوند و در نهایت نتیجه‌گیری و پیشنهادها ارائه خواهد شد.

مبانی نظری و پیشینه پژوهش

طی ماه‌های گذشته، برخی از مطالعات اپیدمیولوژیک برای توصیف و کنترل بیماری کووید-۱۹ انجام شده است. در این قسمت به برخی از این مطالعات به صورت مختصر اشاره خواهد شد. مطالب این قسمت به دو بخش تقسیم شده است؛ پژوهش‌های اجتماعی-اقتصادی مرتبط و پژوهش‌های شبکه اجتماعی مرتبط.

پژوهش‌های اجتماعی-اقتصادی مرتبط

در یک مطالعه انجام شده توسط هولشو و همکاران^۱ (۲۰۲۰)، آن‌ها شناسایی، تشخیص، دوره بالینی و مدیریت اولین مورد از بیماری کووید-۱۹ که در ایالات متحده تأیید شده است را شرح داده‌اند. یافته‌های آن‌ها بر اهمیت مدیریت مستقیم پزشکان و مسئولان بهداشت عمومی در سطح محلی، ایالتی و فدرال و همچنین لزوم پخش سریع هر دو اطلاعات بهداشتی برای جلوگیری از شیوع بیشتر این بیماری و اطلاعات بالینی برای مراقبت از بیماران تأکید می‌کنند. این عفونت در حال تحول در پژوهش دیگری که توسط پرلمن^۲ (۲۰۲۰) منتشر شده برجسته شده است. در این پژوهش یکی از مباحث ذکر شده که

1. Holshue et al.

2. Perlman

ارزش مطالعه دارد، تأثیر اقدامات فاصله‌گذاری اجتماعی است. می‌دانیم که اقدامات فاصله‌گذاری اجتماعی در مقیاس دقیق، می‌تواند به کاهش قابل‌توجه تعداد موارد جدید کووید-۱۹ منجر شود. اگرچه، طیف وسیعی از روش‌های استفاده‌شده و میزان شدت استفاده از آن‌ها برای کسب اطلاعات در مورد اینکه دقیقاً چه کاری انجام می‌شود و طول مدت‌زمان آن، نیاز به بررسی بیشتری دارد.

نگرانی مهم دیگر در این مدت بی‌سابقه مقاومت اقتصاد در پاسخ به این بیماری همه‌گیر است. در مطالعه انجام‌شده توسط فیتزجرالد و همکاران^۱ (۲۰۲۰)، بیماری همه‌گیر کووید-۱۹ «بحران اقتصادی فوری» نامیده شده است. آن‌ها تحقیق کردند که بحران اقتصادی تا چه مدت و تا چه حد طولانی خواهد بود و در مورد بازگشایی محتاطانه زندگی اجتماعی و اقتصادی بحث کردند. همچنین، اثرات محدودیت‌های اعمال‌شده برای جمعیت‌ها برای جلوگیری از شیوع سریع ویروس از جمله قرنطینه‌ها، سفارش‌ها در منزل، تعطیلی مشاغل و ممنوعیت‌های مسافرتی که به رکود عظیم اقتصادی در جهان منجر شده است، مورد مطالعه قرار گرفت. داده‌ها برای تعیین کمیت این تأثیرات هنوز در دست بررسی است. کادنا و همکاران^۲ (۲۰۲۰) مطالعه‌ای در مورد چگونگی شروع مجدد اقتصادهای ملی در هنگام بحران ویروس کرونا انجام داده‌اند. آن‌ها اظهار داشتند که جهان باید هم در کنترل ویروس عمل کند و هم تأثیر منفی بر مشاغل شهروندان را به‌طور هم‌زمان کاهش دهد. پیشرفتی که در آن انجام می‌شود به شکل‌گیری بهبود اقتصادی کمک می‌کند. ایلانس و همکاران^۳ (۲۰۲۰) تأثیر ویروس کرونا بر آموزش عالی و نحوه پاسخ آن‌ها را بررسی کرده‌اند. آن‌ها یک مرکز عصبی یکپارچه پیشنهاد کردند که یک ساختار ساده، انعطاف‌پذیر و چند رشته‌ای است که برای تنظیم آن با شرایط سریع در حال تغییر است. این چهار نوع عمل شامل کشف، تصمیم‌گیری، طراحی و ارائه است. هدف اصلی مرکز عصبی یکپارچه پیشنهادی این موسسه این است که در پیشبرد حوادث کارآمد

1. FitzGerald

2. Cadena et al.

3. Illanes et al.

باشد و به موقع، هوشمندانه و تاکتیکی واکنش نشان دهد. کیو و همکاران^۱ (۲۰۲۰) نقش چندین عامل اقتصادی و اجتماعی را در تسهیل انتقال کووید-۱۹ در سطح محلی و بین‌شهری چین بررسی کردند. در این پژوهش از روش یادگیری ماشین برای انتخاب پارامترهای مهم استفاده شده است که به طور قوی پیش‌بینی انتقال ویروس در میان خصوصیات غنی آب‌وهوا را پیش‌بینی می‌کند.

پژوهش‌های شبکه اجتماعی مرتبط

امروزه روش‌های زیادی برای به اشتراک‌گذاری اطلاعات، نظرات و پیشنهادها توسط شبکه‌های اجتماعی استفاده می‌شود. بیش از ۲٫۹ میلیارد نفر به‌طور مکرر از شبکه‌های اجتماعی استفاده می‌کنند و بسیاری از آن‌ها برای مدت طولانی از این شبکه‌ها استفاده می‌کنند. در یک پژوهش اخیر شارما و همکاران^۲ (۲۰۲۰) ابزاری طراحی کردند تا اطلاعات غلط در گفتگوهای توئیتر را در دوره‌ی بی‌سابقه شیوع کرونا تجزیه و تحلیل کنند. این ابزار با پیشرفت اوضاع، به بحث در شبکه‌های اجتماعی پیرامون ویروس کرونا و کیفیت اطلاعات به اشتراک گذاشته شده می‌پردازد. آن‌ها همچنین یک روش تشخیص را برای یافتن مطالب گمراه‌کننده و فریب‌دهنده از جریان اطلاعات توئیتر طراحی کردند. در مطالعه‌ی دیگری، مرچنت و لوری^۳ (۲۰۲۰) نقش شبکه‌های اجتماعی را به‌عنوان ابزاری مهم در رسیدگی به بیماری همه‌گیر در حال وقوع در کنار تغییر ویژگی‌های آمادگی و پاسخ برای سوانح آینده موردبررسی قرار داده‌اند. آن‌ها تأکید کردند که باید از شبکه‌های اجتماعی برای پخش اطلاعات دقیق در مورد موضوعات مختلف مهم در هنگام بروز موارد همه‌گیری مانند زمان آزمایش دادن، انجام کار مناسب در مواجهه با بیماری و همچنین نحوه مراقبت در هنگام بیماری استفاده شود.

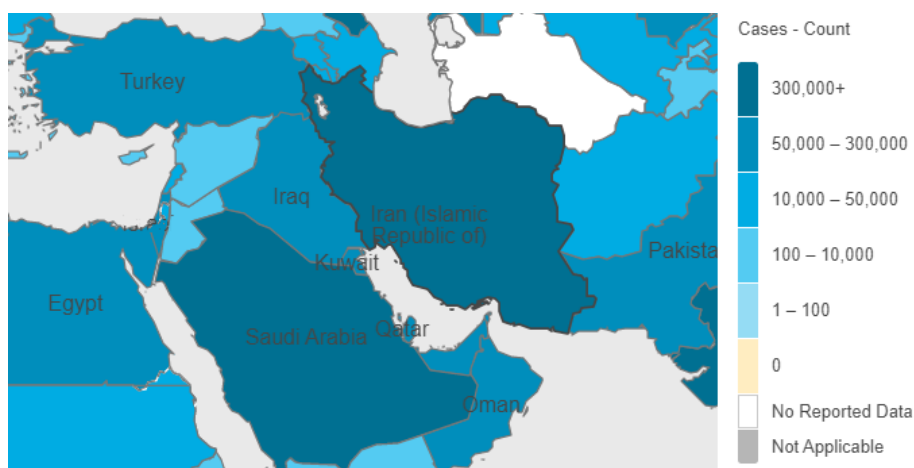
تجزیه و تحلیل احساسات با هدف یافتن نظر و نگرش افراد از دیدگاه آن‌ها در شبکه‌های اجتماعی نسبت به جنبه‌های مختلف محصولات و رویدادها انجام می‌شود

1. Qiu et al.

2. Sharma et al.

3. Merchant & Lurie

نظرات افرادی که از ایران توئیتهای خود را ارسال می کنند از جایگاه ویژه ای برخوردار است.



شکل ۱: نسبت تعداد مبتلایان به کووید-۱۹ در همسایگی ایران.

در این مطالعه، دو نوع داده یعنی روند گوگل و داده توئیتر استخراج شده است. از داده های روند گوگل برای تجزیه و تحلیل علاقه مردم برای به دست آوردن اطلاعات در مورد بیماری کووید-۱۹ با استفاده از جستجوی کلمات کلیدی مرتبط استفاده می شود. روند گوگل یک سرویس محبوب است که توسط گوگل برای به دست آوردن تعداد جستجوهای کاربران برای یک عبارت خاص در یک بازه زمانی معین ارائه شده است. این سرویس یک عدد را در محدوده [۰-۱۰۰] برمی گرداند که در آن ۰ کمترین تعداد جستجو را نشان می دهد و ۱۰۰ بیشترین تعداد جستجو برای کلمات کلیدی داده شده در چهارچوب زمانی ارائه شده است. این نوع داده ها برای کاربردهای مختلف از جمله پیش بینی بازار سهام (ژو و همکاران^۱، ۲۰۱۹)، کشف مباحث سیاسی در توئیتر (ریل و همکاران^۲، ۲۰۱۴) و پیش بینی سرعت انتشار ویروس کرونا (لین و همکاران^۳، ۲۰۲۰) استفاده شده است. برای

1. Xu et al.
2. Rill et al.
3. Lin et al.

پاسخ به سؤال اول پژوهش، تکرار نسبی عبارات مرتبط با کرونا در ایران که با استفاده از داده مربوط به روند گوگل به دست آمده در جدول ۱ نشان داده شده است.

جدول ۱: تعداد تکرار نسبی ده کلمه پرتکرار مربوط به کرونا در ایران^۱

عبارت	نسبت تکرار
ویروس کرونا	۱۰۰
کرونا ایران	۸۳
علائم کرونا	۷۹
کرونا در ایران	۶۴
آمار کرونا	۵۰
کرونا امروز	۳۸
واکسن کرونا	۳۵
اخبار کرونا	۳۱
بیماری کرونا	۲۷

از داده‌های توئیتر برای تجزیه و تحلیل تغییر احساسات در توئیتهای افراد در طول زمان استفاده می‌شود. در پژوهش حاضر بازه‌ی زمانی بهمن ماه نود و هشت تا انتهای آذر ماه نود و نه برای به دست آوردن داده‌های روند گوگل و توئیتر استفاده شده است. همچنین از کلمات کلیدی نشان داده شده در جدول ۲ و ترکیب آن‌ها با در نظر گرفتن علامت هشتگ در ابتدای کلمات برای جمع‌آوری توئیتهای استفاده شده است.

جدول ۲: لیست کلمات کلیدی مورد استفاده برای به دست آوردن توئیتهای.

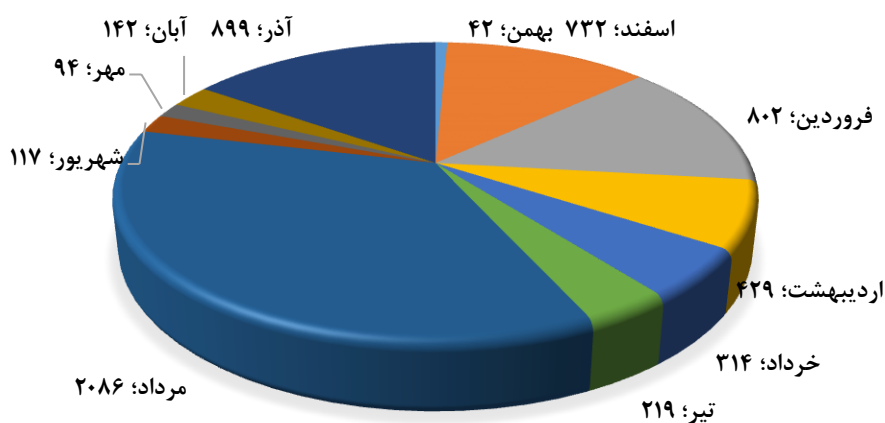
Corona	Coronavirus	Covid-19	Sars-Cov-2
Ncov	Sars-Cov-2	Ncov-۲۰۱۹	Covid-2019
Sars-Cov	Sars-Cov2	Sars-Covs	Mers-Cov

برای جمع‌آوری داده‌های مربوط به کرونا از کتابخانه‌ی ^۲twint در زبان برنامه‌نویسی پایتون استفاده شده است. پس از جمع‌آوری توئیتهای با استفاده از کلمات کلیدی اشاره شده،

1. <https://trends.google.com>

2. <https://github.com/twintproject/twint>

آن‌ها را به صورت ماهانه جدا کرده و مورد بررسی قرار دادیم. توزیع توئیت‌ها در ماه‌های مختلف در شکل ۲ نشان داده شده است. در کل تعداد ۵۸۷۶ توئیت مرتبط جمع‌آوری شد که به صورت متوسط هر توئیت شامل ۲۳/۲۳ کلمه ۱۷۳/۵۳ حرف می‌باشد.



شکل ۲: توزیع توئیت‌های مرتبط با کرونا در ایران از بهمن ماه نود و هشت تا آذر ماه نود و نه.

مروری بر یافته‌های مطالعات دیگر که مرتبط با مسئله و موضوع پژوهش باشد (پیشینه تجربی) و تبیین خلأ یا شکاف موجود درباره مسئله مورد پژوهش در راستای ضرورت انجام مطالعه حاضر.

مدل پیشنهادی

مدل پیشنهادی برای دسته‌بندی توئیت‌های مرتبط با کرونا در شکل ۳ نشان داده شده است. در این مدل از سه جزء اصلی استفاده شده است؛ بردار از پیش آموزش دیده، لایه‌ی بازجریانی^۱ و لایه‌ی انتخاب. دلیل استفاده از بردار تعبیه کلمات از پیش آموزش دیده در مدل پیشنهادی استخراج بهتر ویژگی‌های متنی از توئیت‌های کاربران است. این بردارها بر روی مجموعه داده‌های بسیار بزرگ آموزش دیده‌اند و به خوبی می‌توانند رابطه مفهومی

بین کلمات را مدل کنند (دولین و همکاران^۱، ۲۰۱۸). لایه‌ی بازجریانی به دلیل توانایی در مدل کردن زمینه^۲ و ارتباط بین کلمات در یک جمله در مدل پیشنهادی به کار گرفته شده است. نهایتاً لایه‌ی انتخاب به دلیل کاهش ابعاد و انتخاب مهم‌ترین بخش‌های بردار ویژگی‌ها استفاده شده است.

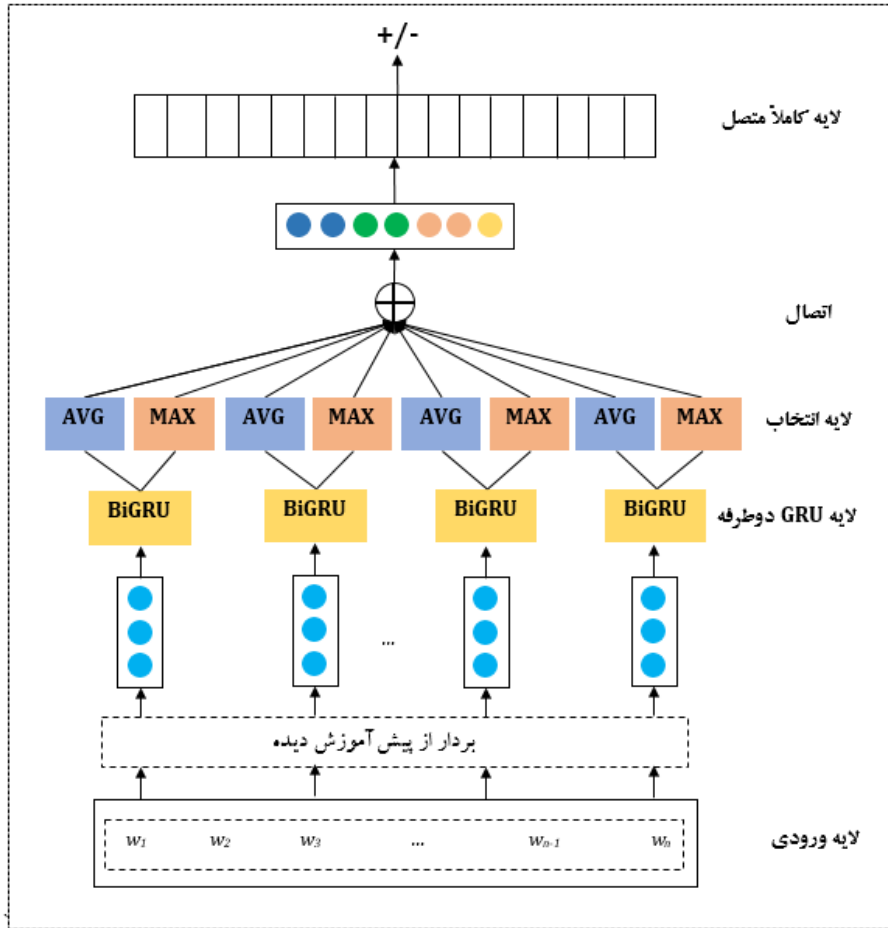
ورودی مدل یک توئیٹ (t_i) که مجموعه‌ای از کلمات (w_i) است، می‌باشد. در مرحله بعد از بردار تعبیه کلمات از پیش آموزش دیده^۳ BERT برای تبدیل کلمات به بردارهای عددی استفاده می‌شود. بردارهای عددی به دست آمده از مرحله قبل به عنوان ورودی لایه‌ی BiGRU استفاده می‌شوند. در این لایه از دولایه شامل واحدهای بازجریانی دروازه‌دار (GRU) در دو جهت ابتدا به انتها و انتها به ابتدای بردار ورودی استفاده می‌شود. ساختار داخلی GRU در شکل ۴ نشان داده شده است. هر واحد GRU شامل دو دروازه به روزرسانی (r) و بازنشانی (z) با روابط (۱) و (۲) است.

$$r_t = \delta(W_r h_{t-1} + U_r x_t + b_r) \quad (1)$$

$$z_t = \delta(W_z h_{t-1} + U_z x_t + b_z) \quad (2)$$

در روابط (۱) و (۲) نماد δ معرف تابع سیگموئید لاجیستیک است، W و U ماتریس وزن‌های سلول و b بردار بایاس می‌باشد. x_t ورودی و h_t وضعیت مخفی سلول هستند. برای محاسبه‌ی h_t از روابط (۳) و (۴) استفاده می‌شود. در این روابط $W \tilde{h}_t$ و $U \tilde{h}_t$ بردار وزن، \tilde{h}_t حالت مخفی کاندید، \tanh تابع تانژانت هایپربولیک و \odot ضرب درایه به درایه (هادامارد) بردارها است.

-
1. Devlin
 2. Context
 3. Pre-trained Word Embedding Vector



شکل ۳. مدل پیشنهادی برای تجزیه و تحلیل احساسات توئیت‌ها.

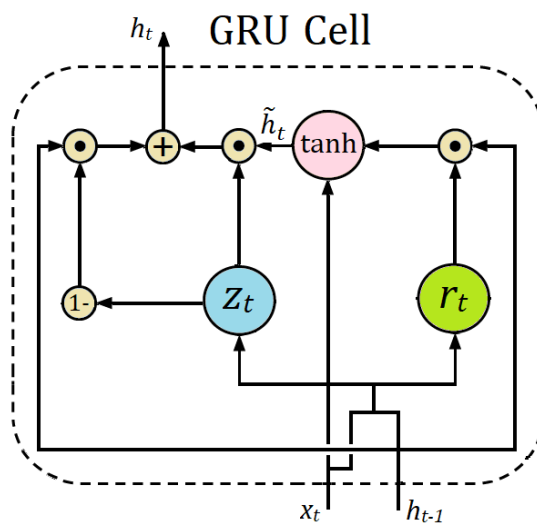
$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (3)$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}_t}(h_{t-1} \odot r_t) + U_{\tilde{h}_t}x_t) \quad (4)$$

خروجی لایه GRU به لایه انتخاب ارسال می‌شود که در آن به دو صورت انتخاب انجام می‌گیرد. در روش اول که در شکل ۳ با AVG نشان داده شده است، از انتخاب با استفاده از میان‌گیری^۱ ماتریس ورودی استفاده می‌شود و در روش دوم که در شکل ۳ با MAX

1. Global average pooling

نشان داده شده از انتخاب بیشینه^۱ ماتریس ورودی استفاده می شود. خروجی دو تابع انتخاب برای ایجاد یک ماتریس ورودی به لایه آخر به همدیگر متصل می شوند. لایه آخر یک لایه فشرده^۲ یا کاملاً متصل^۳ است که وظیفه یادگیری نهایی را بر عهده دارد. در نهایت دسته بندی با استفاده از تابع softmax انجام می گردد. مقادیر پارامترهای مدل پیشنهادی در جدول ۳ نشان داده شده است.



شکل ۴: ساختار داخلی یک سلول GRU

جدول ۳. هایپرپارامترهای استفاده شده در مدل پیشنهادی.

نام پارامتر	مقدار	توضیحات
Max features	۲۰۰۰۰	سایز واژگان لایه تعبیه شده.
Embedding dim	۳۰۰	ابعاد تعبیه شده.
Dropout	۰,۲	نرخ افت.
Number of cells	۸۰	تعداد سلولهای GRU در لایه BiGRU.
Activation function	Relu	تابع فعال سازی لایه کانولوشن.

1. Global maximum pooling
2. Dense
3. Fully connected

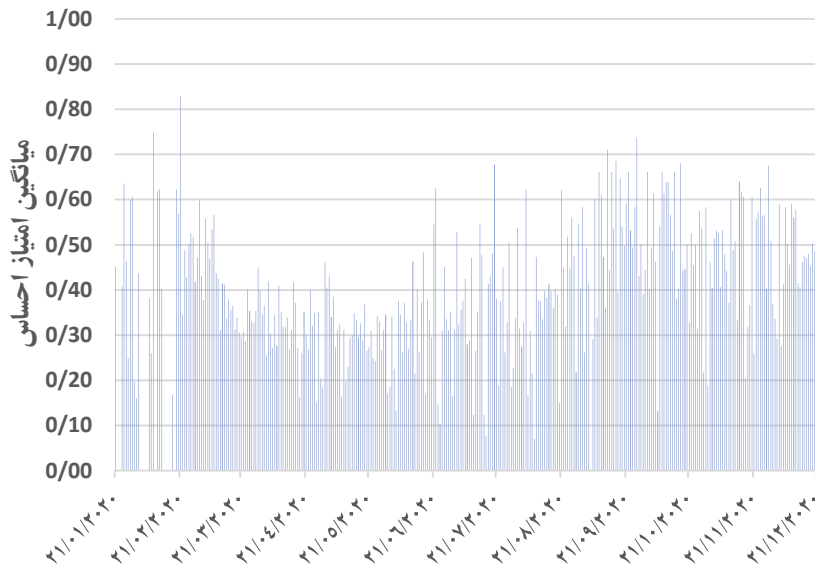
نام پارامتر	مقدار	توضیحات
Dense	۲	تعداد نرون‌ها در لایه خروجی.
Loss function	Categorical Cross Entropy	تابع اتلاف لایه خروجی.
Optimizer	Adam	الگوریتم بهینه‌سازی مدل.

تجزیه و تحلیل داده‌ها

همان‌طور که قبلاً گفته شد، تحلیل احساسات روشی برای دستیابی به افکار عمومی در مورد هر موضوعی است. در این پژوهش توئیت‌های جمع‌آوری شده که در بخش‌های قبل در مورد آن‌ها توضیح داده شد با استفاده از مدل پیشنهادی دسته‌بندی می‌شوند. به این دلیل که آموزش مدل پیشنهادی نیازمند داده‌های برجسب‌گذاری شده می‌باشد و داده‌های جمع‌آوری شده فاقد برجسب می‌باشند، از مجموعه داده Sentiment140 که دارای ۱۶۰۰۰۰۰ توئیت می‌باشد استفاده شد.

پاسخ به دومین سؤال پژوهش

برای پاسخ به دومین سؤال پژوهش که بیان می‌کند «آیا تفاوتی در شدت احساسات نشان داده شده در توئیت‌های مربوط به ویروس کرونا وجود دارد؟» پس از آموزش مدل، توئیت‌های هر روز بر اساس مقیاس صفر تا یک با استفاده از مدل پیشنهادی برجسب‌گذاری شدند. در این مقیاس، صفر نشان‌دهنده منفی کامل و یک نشان‌دهنده مثبت کامل بودن محتوای توئیت است. نتیجه‌ی برجسب‌گذاری انجام‌شده در شکل ۵ برای روزهای مختلف نشان داده شده است. در این شکل، میانگین امتیاز حاصل شده از تمام توئیت‌های هر روز برای آن روز در نظر گرفته شده است.

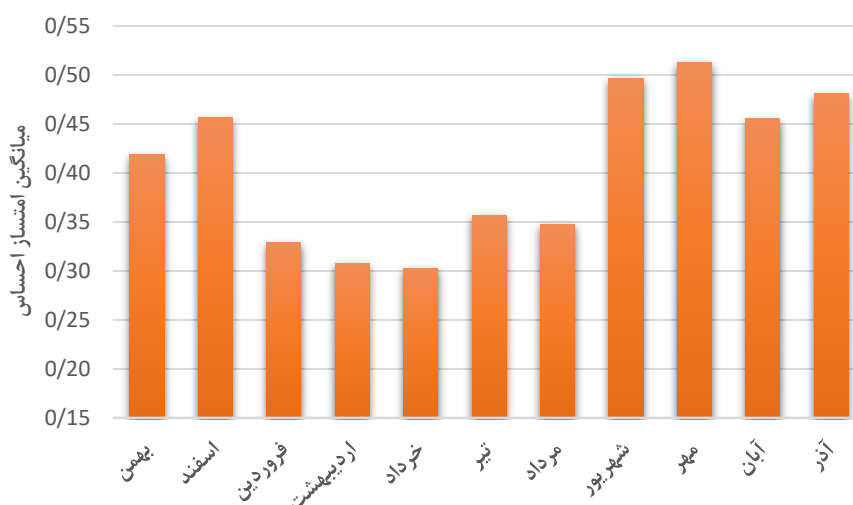


شکل ۵: شدت احساس در روزهای مختلف در بازه‌ی صفر (منفی کامل) تا یک (مثبت کامل).

همان‌گونه که در شکل ۵ نشان داده شده است به‌جز برخی از روزهای سال بیشتر روزها دارای توئیت‌هایی با احساس منفی (اعداد کمتر از ۰/۵) بوده‌اند. این مسئله نشان‌دهنده‌ی جو منفی کلی حاکم بر توئیت‌های مرتبط در می‌باشد؛ به‌عبارت‌دیگر، توئیت‌های ارسال‌شده را می‌توان بازتابی از عقاید نویسنده آن‌ها که تحت تأثیر اخبار و وقایع مربوط به بیماری کووید-۱۹ است تلقی کرد. برای بهتر مشخص شدن نتایج، در شکل ۶، میانگین شدت احساس توئیت‌ها به‌صورت ماهانه نشان داده شده است.

همان‌گونه که در شکل ۶ نشان داده شده است، کمترین میزان احساس منفی مربوط به سه ماه فروردین تا خرداد بوده و در اردیبهشت ماه میزان احساسات منفی از سایر ماه‌ها بیشتر بوده است. این می‌تواند به دلیل موج صعودی دوم بیماری باشد که جو ترس و ناامیدی بیشتری در افکار عمومی ایجاد کرده است. در توجیه کاهش شدت احساسات منفی در ماه‌های بعد می‌توان گفت به‌تدریج این روند برای افراد عادی شده و همین مسئله در توئیت‌های ماه‌های تیر تا شهریور منعکس شده است. پس از این دوره و با شروع موج

جدید کرونا شدت احساس منفی در آبان ماه بیشتر شده اما این افزایش شدت احساس منفی نسبت به موج اول کمتر است که این می‌تواند به دلیل گذشت زمان و کمتر شدن ترس کلی ناشی از بیماری باشد. نهایتاً در آذر ماه و با کاهش آمار مبتلایان دوباره از میزان احساسات منفی کاسته شده است.



شکل ۶: میانگین شدت احساس در ماه‌های مختلف در بازه‌ی صفر (منفی کامل) تا یک (مثبت کامل).

پاسخ به سومین سؤال پژوهش

برای پاسخ به سومین سؤال پژوهش که بیان می‌کند «آیا بین تعداد موارد جدید مبتلا و مرگ‌ومیر با احساسات توثیت در ماه‌های مختلف ارتباط معنی‌داری وجود دارد؟» و برای ریشه‌یابی تغییرات موجود در شکل‌های ۵ و ۶ می‌توان به نمودارهای موجود در شکل‌های ۷ و ۸ توجه کرد.



شکل ۷: تعداد بیماران جدید در ماه‌های مختلف سال از شروع کرونا در ایران.



شکل ۸: تعداد فوت‌شدگان جدید در ماه‌های مختلف سال از شروع کرونا در ایران.

همان‌گونه که در شکل‌های ۷ و ۸ مشخص است، در اردیبهشت ماه پس از کاهش چشمگیر آمار مبتلایان و کشته‌شدگان، موج دوم شروع شده که این می‌تواند باعث افزایش ناگهانی میزان احساسات منفی بیان‌شده در توئیت‌ها باشد. همچنین کاهش تعداد مبتلایان در خرداد ماه و ثابت شدن تقریبی آن در ماه‌های آتی می‌تواند باعث کاهش احساسات منفی بیان‌شده در توئیت‌ها در خرداد ماه و پس از آن ثابت ماندن آن در چهار ماه بعدی تا ابتدای آبان ماه در شکل ۶ باشد. همان‌گونه که در شکل‌های ۷ و ۸ نشان داده شده است، موج جدید بیماری و فوت‌شدگان از مهرماه آغاز شده که تأثیر آن در توئیت‌های آبان ماه

دیده می‌شود. البته افزایش شدت احساسات منفی در این موج نسبت به موج‌های قبلی کمتر است که می‌تواند به دلیل عادی شدن مسئله یا افزایش امیدواری افراد برای ساخت واکسن کرونا باشد.

برای بهتر مشخص شدن محتوای توییت‌های موجود در مجموعه داده استخراج شده، ابر کلمات احساسی مربوطه در شکل ۹ نشان داده شده است. برای به دست آوردن کلمات احساسی موجود در توییت‌ها از واژه‌نامه‌های احساسی مثبت و منفی استاندارد که توسط لئو و ژانگ^۱ (۲۰۱۲) ارائه شده، استفاده شده است. همان‌گونه که در شکل ۹ نشان داده شده است، کلمات مثبت و منفی زیادی در توییت‌ها وجود داشته‌اند که ده مورد پرتکرار از هر دسته در جدول ۴ آورده شده است. لازم به ذکر است که در پژوهش حاضر از روش مبتنی بر واژه‌نامه برای دسته‌بندی توییت‌ها استفاده نشده است بلکه با استفاده از مدل پیشنهادی که یک مدل با ناظر مبتنی بر شبکه عصبی عمیق است این کار انجام شده است.

همان‌گونه که در جدول ۴ نشان داده شده است، اگرچه از روش مبتنی بر واژه‌نامه (از پیش تعریف شده) در این پژوهش استفاده نشده است، کلمات مثبت و منفی به صورت معنی‌داری توسط مدل پیشنهادی شناسایی شدند. این مسئله تأییدکننده‌ی کارایی مدل برای تشخیص احساسات توییت‌ها می‌باشد.

نتیجه‌گیری و پیشنهادها

پژوهش حاضر به تحلیل احساسات موجود در توییت کاربران ایرانی در رابطه با کرونا می‌پردازد. این مسئله با توجه به شیوع گسترده بیماری کووید-۱۹ و تأثیر آن بر فضای احساسی حاکم بر جامعه در فضای مجازی می‌تواند مسئله‌ی مهمی باشد. به این دلیل که پژوهش حاضر اولین پژوهش در زمینه‌ی تحلیل احساسات داده‌های مرتبط با بیماری کووید-۱۹ است، توییت‌ها بر اساس واژگان کلیدی مرتبط با کرونا و در بازه زمانی دی ماه نود و هشت تا مرداد ماه ۹۹ جمع‌آوری شدند. بر اساس پروژه‌های پیشین بهترین نتایج گزارش شده در کاربردهای تحلیل احساسات توسط مدل‌های مبتنی بر یادگیری عمیق به دست آمده‌اند. لذا برای تحلیل احساسات موجود در این توییت‌ها یک مدل مبتنی بر شبکه عصبی عمیق ارائه شد که در آن برای نگاشت توییت‌ها به برداری از اعداد از بردار تعبیه کلمات از پیش آموزش دیده استفاده شد. همچنین از واحدهای دوطرفه بازجریانی دروازه‌دار برای مدل کردن توالی کلمات و به کارگیری زمینه متن‌ها از هر دو سمت استفاده شد. به این دلیل که مدل پیشنهادی یک مدل با ناظر می‌باشد و توییت‌های جمع‌آوری شده بدون برچسب هستند از مجموعه داده بزرگ 140 Sentiment که شامل ۱۶۰۰۰۰۰ توییت است برای آموزش استفاده شد. نتایج حاصل از اعمال مدل پیشنهادی بر روی مجموعه توییت جمع‌آوری شده در رابطه با کرونا نشان‌دهنده‌ی منفی بودن جو کلی حاکم بر توییت‌های در بازه‌ی مورد آزمایش می‌باشد. همچنین، نشان داده شد که با کاهش تعداد مبتلایان در برخی ماه‌ها میزان احساسات منفی موجود در توییت‌های مرتبط نیز کاهش یافته و با شروع موج دوم کرونا در ایران شدت احساسات منفی در توییت‌ها نیز افزایش یافته است. در مورد محتوای توییت‌ها نیز تحلیل کلمات موجود در آن‌ها نشان‌دهنده به کارگیری انواع مختلف کلمات احساسی مثبت و منفی است. این مسئله نشان می‌دهد که در صورتی که روش مبتنی بر واژه‌نامه که از روش‌های مرسوم تحلیل احساسات است به جای روش پیشنهادی مبتنی بر شبکه عصبی عمیق استفاده می‌شود، امکان گمراه شدن روش به خاطر وجود کلمات مثبت و منفی موجود در توییت‌ها وجود داشت. با توجه


به این که پژوهش حاضر اولین کار گزارش شده در زمینه تحلیل احساسات در مورد بیماری کرونا می باشد، محدودیت هایی در مورد این پژوهش وجود دارد. به عنوان مثال کمبود توثیقات های مربوط به کرونا به خصوص در ماه های اولیه شیوع این بیماری یکی از محدودیت های موجود است که با افزایش تعداد کاربرانی که در این زمینه نظرات خود را در شبکه های اجتماعی بیان می کنند این محدودیت در آینده برطرف خواهد شد. همچنین عدم دسترسی به مجموعه داده استاندارد در حوزه ی بیماری کرونا محدودیت دیگری برای پژوهش حاضر است که با به کارگیری روش های دستی یا روش پیشنهادی در این مقاله در پژوهش های آتی این محدودیت کمتر خواهد شد. به عنوان کارهای پیشنهادی آینده برای این پژوهش می توان از استفاده از سایر شبکه های اجتماعی و پیام رسانی های داخلی و خارجی برای جمع آوری داده ها در کنار به کارگیری روش های مبتنی بر یادگیری انتقالی در شبکه های عصبی عمیق نام برد. مورد اول در راستای کاهش اولین محدودیت اشاره شده و مورد دوم برای پاسخ به دومین محدودیت اشاره شده قابل پیشنهاد هستند. همچنین ترکیب مدل پیشنهادی با مدل های از پیش آموزش دیده می تواند کمک شایانی به رفع دومین محدودیت اشاره شده برای این پژوهش بنماید.


ORCID


Mohamma Ehsan Basiri

Shirin Habibi

Shahla Nemati

 <https://orcid.org/0000-0003-2893-3892>

 <https://orcid.org/>

 <http://orcid.org/0000-0003-2906-5871>

منابع

- فخری، پروین و حسین زاده، مهدی. (۱۳۹۶). آنالیز محتوای کمپین‌های انتخاباتی ۲۰۱۶ ریاست جمهوری ایالات متحده آمریکا در توئیتر. فصلنامه مطالعات مدیریت کسب و کار هوشمند، ۵(۲۰)، ۱۵۶-۱۲۱.
- کوثری لنگری، روح‌الله؛ سردار، سهیلا؛ امین موسوی، سید عبدالله و رادفر، رضا. (۱۳۹۸). مدلی برای انتشار داده‌های شبکه‌های اجتماعی برخط با حفظ حریم خصوصی. مطالعات مدیریت کسب و کار هوشمند، ۸(۲۹)، ۱۱۲-۸۶.
- عباسی، فاطمه؛ سهرابی، بابک؛ مانیان، امیر و خدیور، آمنه. (۱۳۹۶). ارائه مدلی جهت دسته‌بندی احساسات خریداران کتاب با استفاده از رویکرد ترکیبی. مطالعات مدیریت کسب و کار هوشمند، ۶(۲۱)، ۹۲-۶۵.

References

- Abdar, M., Basiri, M. E., Yin, J., Habibnezhad, M., Chi, G., Nemati, S., & Asadi, S. (2020). Energy choices in Alaska: Mining people's perception and attitudes from geotagged tweets. *Renewable and Sustainable Energy Reviews*, 124, 109781.
- Basiri, M. E., Abdar, M., Cifci, M. A., Nemati, S., & Acharya, U. R. (2020). A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques. *Knowledge-Based Systems*, 105949.
- Basiri, M. E., & Kabiri, A. (2017). Translation is not enough: comparing lexicon-based methods for sentiment analysis in Persian. In *2017 International Symposium on Computer Science and Software Engineering Conference (CSSE) IEEE*. Shiraz: IEEE.
- Basiri, M. E., & Kabiri, A. (2018). Words Are Important: Improving Sentiment Analysis in the Persian Language by Lexicon Refining. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 17(4), 26.
- Cadena, A., Child, F., Craven, M., Ferrari, F., Fine, D., Franco, J., & Wilson, M. (2020). No Title.
- Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(3), 102-107.
- Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (2017). *A practical guide to sentiment analysis*. Springer.
- Chauhan, U. A., Afzal, M. T., Shahid, A., Abdar, M., Basiri, M. E., & Zhou, X. (2020). A comprehensive analysis of adverb types for mining user sentiments on amazon product reviews. *World Wide Web*.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-

- training of deep bidirectional transformers for language understanding. *ArXiv Preprint ArXiv*,1810.04805.
- FitzGerald, A., Kwiatkowski, K., Singer, V., & Smit, S. (2020). An instant economic crisis: How deep and how long. *McKinsey Company*, 6.
- Fouladfar, F., Dehkordi, M. N., & Basiri, M. E. (2020). Predicting the Helpfulness Score of Product Reviews Using an Evidential Score Fusion Method. *IEEE Access*, 8, 82662–82687.
- Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, 40(16), 6266–6282.
- Haidong, S., Junsheng, C., Hongkai, J., Yu, Y., & Zhantao, W. (2020). Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing. *Knowledge-Based Systems*, 188, 105022.
- Hassan, S.-U., Aljohani, N. R., Idrees, N., Sarwar, R., Nawaz, R., Martínez-Cámara, E., ... Herrera, F. (2020). Predicting literature's early impact with sentiment analysis in Twitter. *Knowledge-Based Systems*, 192, 105383.
- Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., ... others. (2020). First case of 2019 novel coronavirus in the United States. *New England Journal of Medicine*.
- Illanes, P., Law, J., Mendy, A., Sanghvi, S., & Sarakatsannis, J. (2020). No Title.
- Lin, Y.-H., Liu, C.-H., & Chiu, Y.-C. (2020). Google searches for the keywords of “wash hands” predict the speed of national spread of COVID-19 outbreak among 21 countries. *Brain, Behavior, and Immunity*.
- Liu, B., & Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. In *Mining Text Data*, 415–463.
- Merchant, R. M., & Lurie, N. (2020). Social media and emergency preparedness in response to novel coronavirus. *Jama*.
- Nemati, S., Rohani, R., Basiri, M. E., Abdar, M., Yen, N. Y., & Makarenkov, V. (2019). A Hybrid Latent Space Data Fusion Method for Multimodal Emotion Recognition. *IEEE Access*, 7, 172948–172964.
- Nowak, J., Taspinar, A., & Scherer, R. (2017). LSTM recurrent neural networks for short text and sentiment classification. *International Conference on Artificial Intelligence and Soft Computing*, 553–562.
- Oueslati, O., Cambria, E., HajHmida, M. Ben, & Ounelli, H. (2020). A review of sentiment analysis research in Arabic language. *Future Generation Computer Systems*.
- Perlman, S. (2020). *Another decade, another coronavirus*. Mass Medical Soc.

- Pradhan, D., Biswasroy, P., Ghosh, G., Rath, G., & others. (2020). A review of current interventions for COVID-19 prevention. *Archives of Medical Research*.
- Qiu, Y., Chen, X., & Shi, W. (2020). Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. *Journal of Population Economics*, 1.
- Rill, S., Reinel, D., Scheidt, J., & Zicari, R. V. (2014). Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis. *Knowledge-Based Systems*, 69, 24–33.
- Ruz, G. A., Henríquez, P. A., & Mascareño, A. (2020). Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Future Generation Computer Systems*, 106, 92–104.
- Sharma, K., Seo, S., Meng, C., Rambhatla, S., & Liu, Y. (2020). Covid-19 on social media: Analyzing misinformation in twitter conversations. *ArXiv Preprint ArXiv:2003.12309*.
- Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. *Tropical Medicine & International Health*, 25(3), 278.
- Xu, Q., Bo, Z., Jiang, C., & Liu, Y. (2019). Does Google search index really help predicting stock market volatility? Evidence from a modified mixed data sampling model on volatility. *Knowledge-Based Systems*, 166, 170–185.
- Zhang, Z., Robinson, D., & Tepper, J. (2018). Detecting hate speech on twitter using a convolution-gru based deep neural network. *European Semantic Web Conference*, 745–760.
- Fakhri, P., Hosseinzadeh Zadeh, M. (2017). Content Analysis of the 2016 United States Presidential Election Campaigns on Twitter. *IT Management Studies*, 5(20), 121-156. [InPersian]
- Kosari Langari, R., Sardar, S., Amin Mousavi, S., & Radfar, R. (2019). A Model to Publish Online Social Networks Data with Privacy Preserving. *IT Management Studies*, 8(29), 87-112. [InPersian]
- Abbasi, F., Sohrabi, B., Manian, A., & Khadivar, A. (2017). A Model to Classify Book Buyers' Sentiments Using Ensemble Approach. *IT Management Studies*, 6(21), 65-92. [InPersian]

استناد به این مقاله: بصیری، محمد احسان، حبیبی، شیرین، نعمتی، شهلا. (۱۴۰۰). تحلیل احساسات توئیتهای مرتبط با کرونا در ایران با استفاده از شبکه عصبی عمیق، *مطالعات مدیریت کسب و کار هوشمند*، ۱۰(۳۷)، ۱۰۹-۱۳۴.
DOI: 10.22054/IMS.2021.54705.1799



Journal of Business Intelligence Management Studies is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License..