

مقاله پژوهشی: شناسایی زود هنگام سیگنال‌های ناآرامی در رسانه‌های اجتماعی (مطالعه موردی اعتراضات سال ۹۷ ایران در تلگرام)

۲۰.۱۰۰۱.۱.۳۳۲۹۲۵۳۸.۱۴۰۱.۱۲.۴۵.۱.۴

هومن اهرامی^۱، احمد جعفرنژاد چقوشی^۲، مسعود اسدپور^۳

تاریخ پذیرش: ۱۴۰۰/۰۵/۱۵

تاریخ دریافت: ۱۳۹۹/۰۵/۰۹

چکیده

ایران از زمان فراگیری شبکه دیتای تلفن همراه تا امروز بیش از ۳۵ میلیون کاربر پیام‌رسان‌های اجتماعی دارد. فراگیری رسانه‌های اجتماعی (در ایران با ضریب نفوذ ۴۹٪) در بین عموم مردم و کارکردهای متنوع آن‌ها می‌تواند نقطه شروع استفاده از سنجش اجتماعی باشد؛ در این حالت کاربران مشابه یک حسگر، داده‌هایی برای واکافت فراهم می‌کنند؛ یکی از کاربردهای سنجش اجتماعی مدیریت بحران است. از آنجایی که شناسایی زود هنگام رویدادها و بحران‌های سیاسی، اقتصادی و اجتماعی از اهمیت بالایی برای مدیریت جامعه برخوردار است؛ روشی که بتواند با استفاده از واکافت کلان داده‌ها به این هدف بپردازد؛ به‌عنوان یک سیستم پدافندی می‌تواند از هزینه‌های انسانی، اقتصادی و اجتماعی بحران‌ها و رخدادها کاسته و به‌عنوان یک ابزار برای بالابردن آگاهی موقعیتی و ارتقا امنیت ملی مورد بهره‌برداری قرار گیرد. مجموعه داده پژوهش حاضر بر اساس خزش و متن‌کاوی خوشونت کلامی در یک میلیون و چهارصد هزار کانال عمومی فارسی پیام‌رسان تلگرام در سال ۹۷ جمع‌آوری شده و پس از پالایش بر اساس سری زمانی میانگین متحرک نمایی مدل شده است؛ برای شناسایی سیگنال‌های بحران در این مدل از اسیلاتور تعقیب‌گر روند مومنتوم (که بیشتر در تحلیل‌های مالی از آن استفاده می‌شود) و میانگین متحرک همگرایی واگرایی (مکدی) تحلیل شده است؛ از این واکافت در علوم اجتماعی محاسباتی نخستین بار است که در جهت پیش‌بینی بحران‌های امنیتی و حوادث سیاسی و ایجاد امکان اشراف حاکمیتی بر آن استفاده شده است. براساس یافته‌های پژوهش حداقل ۶ اعتراض اجتماعی کشور در سال ۹۷ پیش از رخداد قابل شناسایی و مدیریت بوده است؛ همچنین سامانه‌ای که بتواند از چنین واکافت‌هایی در کلان داده‌های رسانه‌های اجتماعی به‌صورت بلادرنگ استفاده کند، کارایی لازم برای اعلان هشدار زود هنگام و اندازه‌گیری مخاطرات سیاسی و امنیتی جامعه را خواهد داشت.

کلیدواژه‌ها: شناسایی زود هنگام، سیگنال بحران، سری زمانی، مکدی، رسانه‌های اجتماعی

۱. دانشجوی دکتری مدیریت فناوری اطلاعات، دانشگاه آزاد واحد تهران مرکز

<https://scholar.google.com/citations?user=HzMAYoAAAAJ&hl=en>

۲. استاد تمام و عضو هیئت علمی دانشگاه تهران، (نویسنده مسئول)

<https://scholar.google.com/citations?user=JTt0qOIAAAAJ&hl=en>

۳. استادیار و عضو هیئت علمی دانشگاه تهران

<https://scholar.google.com/citations?user=MKwwcvIAAAAJ&hl=en>

مقدمه و بیان مسئله

با فراگیری تلفن‌های همراه هوشمند و گسترش دسترسی به اینترنت، رسانه‌های اجتماعی طی دهه گذشته به محرکی مهم برای تولید، انتشار و کسب اطلاعات در حوزه‌های گوناگون مانند کسب‌وکار، سیاست و افکار عمومی، علوم اقتصادی، مدیریت بحران و سلامت تبدیل شده است؛ یکی از دلایل محبوبیت رسانه‌های اجتماعی، فرصت ایجاد و یا تبادل بلادرنگ پیام‌های عمومی با هزینه‌های بسیار پایین و در همه جاست. رشد عظیم استفاده از رسانه‌های اجتماعی منجر به انباشت داده‌های بیشتر شده است که «کلان داده‌های رسانه‌های اجتماعی»^۱ را شکل می‌دهند. رشد استفاده از رسانه‌های اجتماعی فرصت‌های جدیدی را برای تجزیه و تحلیل چندین جنبه و الگوهای ارتباطات فراهم می‌کند؛ برای مثال داده‌های رسانه‌های اجتماعی را می‌توان تجزیه و تحلیل کرد تا بیش‌های مربوط به مسائل، روندها، بازیگران تأثیرگذار و سایر اطلاعات را به دست آورد؛ (این موضوع در مطالعه نظام‌مند فیلیپ لارنس^۳ و دیگران بر روی ۱۰۶ مقاله این حوزه گزارش شده است).

در طی چند سال گذشته، رسانه اجتماعی امکان ساخت و تبادل محتوای تولید شده توسط کاربر^۴ (UGC) را در مقیاس وسیعی فراهم کرده و نحوه ارتباط افراد در جوامع را تحت تأثیر قرار داده‌اند؛ با توجه به دلایل ذکر شده رسانه‌های اجتماعی می‌توانند نقطه شروع انجام واکف‌هایی^۵ باشند که بر استفاده از کاربران به‌عنوان تولیدکنندگان اطلاعات تمرکز می‌کند؛ این نوع مطالعات در ادبیات نظری متأخر را سنجش اجتماعی^۶ می‌نامند. (دوانگ^۷ و دیگران، ۲۰۱۹) سنجش اجتماعی مبتنی بر این ایده است که جوامع یا گروه‌های مختلف مجموعه‌ای از اطلاعات را فراهم می‌کنند که مشابه با اطلاعاتی است که

^۱ Social media

^۲ Social media big data

^۳ Phillips, Lawrence

^۴ User generated content

^۵ Analysis

^۶ Social sensing

^۷ D. Wang

می‌توان از یک حسگر دریافت کرد؛ این حجم از اطلاعات، دانشی پیچیده و کامل از یک یا چند موضوع خاص تولید می‌کند؛ بنابراین می‌توان به کاربران رسانه‌های اجتماعی به چشم حسگرهای واقعی که اطلاعات در چند موضوع خاص را فراهم می‌کنند، نگاه کرد و آن کاربران، مبنای مطالعات مختلفی از جمله واکافت‌های اجتماعی باشند. توییت‌ها که از جمله برجسته‌ترین پلتفرم‌های رسانه اجتماعی است؛ یکی از مهم‌ترین سکوهاى مطالعه سنجش اجتماعى محسوب مى‌شود. داده‌های تولید شده توسط میلیون‌ها انسان در این رسانه اجتماعی می‌تواند به روش‌های اتوماتیک یا به‌واسطه سایت‌های مختلف مستقیماً در دسترس قرار بگیرند. مطالعات بسیاری (گزارش شده در لارنس و دیگران، ۲۰۱۷) در زمینه تحلیل داده توییت‌ها مانند ردیابی رخدادهاى همچون زلزله، تحلیل احساسات توییت‌ها و تحلیل هشتگ‌های برتر توییت‌ها، انجام شده است.

علی‌رغم همه پژوهش‌های صورت پذیرفته با رویکرد گذشته‌نگر یا در-لحظه، تاکنون در خصوص ردیابی سیگنال‌هایی که می‌توانند وقایع آینده یا روندهای آتی را شناسایی کنند پژوهشی گزارش نشده است. ردیابی سیگنال از این جهت اهمیت دارد که ذی‌نفعان می‌توانند سیگنال موجود در داده‌ها را شناسایی کرده و علت به‌وجود آمدن آن را بررسی نمایند؛ در صورتی که یک سیگنال منفی باشد، اقدامات متقابلی برای مواجهه و تدبیر در آن خصوص در دستور کار قرار گیرد؛ از این رو تحلیل داده‌های رسانه‌های اجتماعی برای حاکمیت‌ها از منظر پیشگیری از اتفاقات منفی، بسیار حائز اهمیت است. خاصه آنکه شناسایی آن سیگنال‌ها به مؤلفه‌ها، معیارها و سنجه‌های کیفی (فرهنگی، اجتماعی، سیاسی) وابسته باشد؛ به‌عنوان مثال در علوم سیاسی، بالاخص بحران‌های سیاسی تلاش‌هایی در یافتن سنجه و مدل صورت پذیرفته است که از مهم‌ترین آن‌ها می‌توان به شاخص محرومیت نسبی آتدرابرت‌گر اشاره نمود که سنجه‌هایی کلان‌نگر دارد.

۱. سیگنال در اینجا عبارتست از یک تغییر ناگهانی نسبت به مجموعه داده‌های گذشته

۲ Relative deprivation

۳ Ted Robert Gurr

یکی از این سیگنال‌های مهم که می‌توانند روندهای آتی را نشان دهند؛ سیگنال‌های بحران در جامعه است. بر اساس پژوهش شجاع مؤدب و دیگران (۱۳۹۷: ۱۷۹-۲۰۴)، ظهور بحران‌های سیاسی - امنیتی تحت‌تأثیر عوامل روانی و اجتماعی است. ایجاد این عوامل باعث ظهور بحران و رفع آن‌ها باعث کنترل بحران می‌شود؛ به عبارت دیگر چرخه ظهور بحران‌ها تحت‌تأثیر چرخه تقویت و تضعیف این عوامل روانی و اجتماعی است. امروزه عوامل روانی و اجتماعی که باعث ایجاد بحران‌های سیاسی - امنیتی می‌شوند تنها تحت‌تأثیر فضای فیزیکی و ویژگی‌های آن نیستند و فضای سایبر و ویژگی‌های آن نیز بر این عوامل تأثیر می‌گذارند و این تأثیر می‌تواند باعث تشدید بحران شده و یا آن را تضعیف کند؛ بنابراین به‌طور خلاصه انتقال عوامل موجب بحران به فضای سایبر و اثرپذیری از آن و ورود مجدد آن‌ها در قالبی تغییر یافته به فضای فیزیکی بر شکل بحران و نتایج آن تأثیرات قابل‌توجهی دارد. درک چگونگی این تحولات به متولیان امنیت ملی کمک می‌کند تا اثر تشدید و یا تضعیف‌کننده آن‌ها را پیش‌بینی کرده و اقدامات مدیریتی و کنترلی مناسبی را در این خصوص اتخاذ کنند.

هدف مقاله حاضر، شناسایی زودهنگام سیگنال‌های بحران در رسانه‌های اجتماعی است به‌نحوی که بتوان پیش از رخداد بحران در فضای حقیقی برای آن تمهیدی اندیشید؛ بدین ترتیب می‌توان از واکافت داده در جهت افزایش آگاهی موقعیتی نسبت به شکل‌گیری، رخداد پدیده و مدیریت بحران بهره گرفت؛ از سوی دیگر در صورت بی‌توجهی به این سیگنال‌ها، عملاً مسئولان مواجهه با پدیده باید به‌جای مدیریت ریسک و کاهش مخاطرات، به مدیریت بحران و کاهش هزینه‌های آن روی بیاورند. شناسایی زودهنگام در این پژوهش بر بستر محتوای کانال‌های عمومی تلگرام صورت پذیرفته است. سکوی تلگرام از آن جهت انتخاب شده است که اولاً در میان کاربران رسانه‌های اجتماعی فارسی زبان دارای ضریب نفوذ بالایی است و ثانیاً تاکنون پژوهشی با رویکرد علوم اجتماعی محاسباتی^۲ در حوزه امنیت ملی بر روی آن صورت نگرفته است.

۱ Situational awareness

۲ computational social science

نخستین گام در کسب آگاهی موقعیتی که هدف این مقاله است؛ شناسایی سریع و دقیق سیگنال‌های بحران در جهت شناسایی منشأ آن است. بر اساس مطالعه سهرابی و مسلمی (۱۳۹۴: ۹-۳۲)، شاخص ثبات در شبکه‌های اجتماعی مجازی بیشترین اهمیت را در بین شاخص‌ها برای امنیت ملی جمهوری اسلامی ایران داشته است؛ به عبارت دیگر بهره‌گیری از شبکه‌های اجتماعی و ارزیابی دائم آن‌ها در جهت ثبات داخلی جامعه و امنیت پایدار است که این مهم با چنین واکافت‌هایی میسر خواهد شد. در مطالعه عباسی و شریعت (۱۳۹۷: ۱۵۰-۱۲۷) نیز نتیجه‌گیری شده است که از ظرفیت رسانه‌های اجتماعی می‌توان در کنترل بحران‌های امنیتی بهره برد که مقدمه آن شناسایی زودهنگام منشأ و روند شکل‌گیری این بحران‌ها در رسانه‌های اجتماعی است. از آنجاکه در تشخیص بحران علاوه بر جنبه‌های سرعت، همه‌گیری، انگیزه‌مند بودن، سازمان و رهبری، باید جلوه‌های بروز بحران (مانند اعتصاب، تظاهرات، شورش، خرابکاری، اغتشاش و ...) نیز مد نظر قرار گیرد، در این مطالعه تمرکز شناسایی بر اعتراضات اصناف و اقشار مختلف و بازنمایی آنها در فضای مجازی در سال ۹۷ بوده است که پس از خروج آمریکا از برجام (۱۸ اردیبهشت)، افزایش چشمگیر قیمت دلار (اوایل مهر) و اعمال مجدد تحریم‌های آمریکا (۱۳ آبان) رخ داده‌اند. حاصل پژوهش حاضر، مویده این نکته است که بهره‌گیری از ابزارهای واکافت کلان‌داده رسانه‌های اجتماعی در قالب روشی نظام‌مند برای دست‌یافتن به بینش و تدبیر پیش‌دستانه و پیشگیری از وقوع بحران در جامعه مؤثر است.

۱. مبانی نظری

تشخیص موضوعات و روند محتوا و مسئله شناسایی خودکار و زودهنگام آن‌ها با توجه به حجم گسترده‌ای از موجودیت‌های اینترنتی از جمله داده‌های شبکه‌های اجتماعی، اخبار، دانشنامه‌ها، کتابخانه‌های دیجیتالی و... از لحاظ عملی بسیار مهم است؛ از آنجایی که داده‌های شبکه‌های اجتماعی بلادرنگ ایجاد می‌شوند، زمینه را برای رویکردهای تحقیقاتی فراهم می‌کنند که قادر به پیش‌بینی پدیده‌ها مانند شیوع بیماری، شرایط اضطراری و

تغییرات نظرات و ذائقه کاربران است. همچنین نتایج مطالعات پیشین نشان می‌دهد که از نیمه دوم دهه ۱۹۹۰ میلادی به بعد، فضای مجازی با قابلیت‌هایی نظیر اطلاع‌رسانی گسترده، شبکه‌سازی و هویت‌سازی، تأثیر قابل توجهی بر موفقیت و بسیج اعتراض‌ها در سطح جهانی داشته است و به نظر می‌رسد امروزه فضای مجازی در تشویق و سازماندهی اعتراض‌های گسترده ملی، منطقه‌ای و جهانی، چه به صورت برخط در فضای مجازی و چه به صورت واقعی، در جوامع مختلف بسیار مؤثر بوده است و با اطلاع‌رسانی سریع درباره مسائل حاد جامعه، انتقال نمادهای اعتراضی و تحریک احساسات، به کنش‌های اعتراضی گسترده دامن زده است. به گفته سردارنیا (۱۳۸۷)، در نتیجه در مطالعه بحران و اعتراضات، فضای مجازی (خصوصاً رسانه‌های اجتماعی به عنوان ابزار ارتباطی مؤثر عصر حاضر) عامل تعیین‌کننده‌ای محسوب می‌شود.

بیشتر پژوهش‌های منتشر شده در یک دهه اخیر با همین رویکرد و با هدف یافتن مدل‌های ریاضی و استفاده از روش‌های ماشینی‌سازی برای تشخیص روندها و موضوعات در داده‌های شبکه‌های اجتماعی است. به جهت سهولت پژوهش غالباً شبکه اجتماعی توییتر و در برخی موارد فیسبوک (به جهت آرایه API مورد نیاز پژوهشگران) مورد مذاقه واقع شده است و کمتر پژوهشی خارج از این دو شبکه اجتماعی به بررسی مسائل علمی پرداخته‌اند؛ همان‌طور که در جدول (۱) مشاهده می‌شود از ۲۰ پژوهش مرتبط با این مقاله نوزده پژوهش در بستر توییتر و یک پژوهش در بستر فیسبوک انجام شده است. محوریت این مطالعات حول مفهوم انفورماتیک بحران^۱ و با هدف مواجهه با بلایای طبیعی و هشدار زودهنگام در این خصوص است. بحث شناسایی موضوعات و روندها نیز شاخص‌ها و روش‌های متعددی دارد که پژوهش‌های اخیر در صدد یافتن کاراترین آن‌هاست؛ به عنوان مثال ف. نظیر و دیگران در شناسایی سیگنال در رسانه‌های اجتماعی با استفاده از حجم

۱. Crisis informatics

علی‌رغم آنکه که واژه crisis اغلب به همه اشکال بحران، خصوصاً بحران‌های انسان پدید اشاره دارد و Disaster به بلایای طبیعی؛ این ترکیب در ادبیات پژوهشی صرفاً در مفهوم انفورماتیک بلایای طبیعی به کار رفته است.

توییت‌ها، هشتگ و تحلیل احساسات و با بهره‌گیری از سه الگوریتم شامل الگوریتم آستانه حرکت متوسط، الگوریتم گوسین و الگوریتم هیبریدی به دنبال ردیابی سیگنال در حجم توییت‌ها، احساسات توییت‌ها و هشتگ‌های برتر بر روی مجموعه داده‌های آرامکو سعودی و بریتیش پترولیوم آمریکایی‌اند که بهترین روش را بیابند و یا رضا مصطفی در پژوهش تشخیص زودهنگام سخنان بحث‌برانگیز به زبان اردو در رسانه‌های اجتماعی با مدل‌های ماشین بردار پشتیبان آیز ساده^۲ و رگرسیون لجستیک^۳ به دنبال یافتن کاراترین روش در تشخیص مباحث بحث‌برانگیز در رسانه‌های اجتماعی برای کاربردهای مختلفی نظیر معرفی محتوا و استخراج رویدادهای بحث‌برانگیز است. دیمتریوس میلیوریس در تشخیص روند و انتشار اطلاعات در شبکه‌های اجتماعی پویا با روش کیسه کلمات یا n-gram روش جدیدی را برای تشخیص موضوع، طبقه‌بندی و سنجش روند در متون کوتاه ارائه کرده است؛ همه این پژوهش‌ها بر نحوه شناسایی و یافتن روند و موضوعات کلی تأکید دارند و کمتر پژوهش‌هایی به تبیین موضوعات خاص (خصوصاً حوزه علوم انسانی) مانند بحران‌های سیاسی، اجتماعی و... پرداخته‌اند؛ به‌عنوان مثال جی بورگرهلدفردر الگوهای ساختاری و دینامیکی در شبکه‌های اجتماعی آنلاین: جنبش پانزدهم مه اسپانیا به‌عنوان یک مطالعه موردی که در خصوص موضوع خاص جنبش سیاسی می‌پردازد؛ تنها به رابطه شبکه‌های اجتماعی و جنبش اجتماعی اسپانیا (به‌عنوان یک مکانیسم سازماندهی اساسی) می‌پردازد و توسعه مدل‌هایی برای شناسایی و ارائه این‌گونه مدل‌ها کمتر در پژوهش‌های منتشر شده ملاحظه می‌شود. در ادامه جدول مرور ادبیات مرتبط با این پژوهش ارائه شده است.

جدول (۱) - مرور ادبیات موضوع

ردیف	شماره منبع	سال انتشار	موضوعات اصلی
۱	۲۷	۲۰۱۷	شناسایی مباحث بر بحث زبان اردو در تویتر
۲	۴۰	۲۰۱۷	شناسایی زودهنگام کمپین‌های ترویجی در تویتر
۳	۸	۲۰۱۸	تشخیص زودهنگام مشکلات ذاتی عملکرد ضعیف پروژه
۴	۲۸	۲۰۱۹	شناسایی موضوعات مهم در رسانه‌های اجتماعی با استفاده از حجم توییت‌ها، هشنگ و تحلیل احساسات
۵	۲۱	۲۰۱۹	شناسایی نفرت‌پراکنی برخط در تویتر، فیس‌بوک و فروم‌ها
۶	۴۲	۲۰۱۹	پیشنهاد یک سیستم هشدار سریع با هدف شناسایی موارد احتمال خودکشی در هنگ‌کنگ
۷	۲۹	۲۰۲۰	پیشنهاد یک سیستم هشدار زودهنگام برای شناسایی و پیش‌بینی تلاطم بازار سهام
۸	۴۵	۲۰۲۰	طراحی سیستم‌های هشدار زودهنگام مردم‌محور برای مواجهه با بلایای طبیعی در کشور چین
۹	۳۴	۲۰۲۰	سیستم‌های هشدار زودهنگام برای نجات جان افراد، کاهش آسیب‌ها در حوادث فاجعه‌بار
۱۰	۴۱	۲۰۲۰	ارایه مدل هشدار زودهنگام بحران‌های سیستم توزیع انرژی

۲. روش شناسی تحقیق

گام نخست در این پژوهش داده‌کاوی است. داده‌کاوی یا «کشف دانش از داده»، فرایند استخراج اطلاعات و دانش از داده‌های موجود در پایگاه‌داده یا انبار داده است که امروزه به‌واسطه گسترش روزافزون اینترنت و ظهور کلان داده‌ها کاربرد بسیاری در اکثر علوم دارد. تفاوت اصلی داده‌کاوی و علم آمار، در حجم داده‌های مورد تحلیل، روش مدل‌سازی داده‌ها و استفاده از هوش مصنوعی است. داده‌کاوی شاخه توسعه‌یافته علم آمار

است. داده‌کاوی یک روش حل مسئله مبتنی بر داده‌های موجود است (الدرزان و دیگران، ۲۰۰۹). بر اساس متدولوژی استاندارد جهانی کریسپ (CRISP-DM) شامل فرایندهای فهم کسب‌وکار، فهم داده، آماده‌سازی داده، مدل‌سازی، ارزیابی، توسعه‌است؛ در این متدولوژی داده‌کاوی (از داده‌های خام خزش شده تا شکل‌دهی به دانش جدید) شامل فرایندهای بازگشتی جمع‌آوری و درو داده، تصفیه داده، یکپارچه‌سازی داده، انتخاب داده، تبدیل داده، داده‌کاوی، ارزیابی الگو، آرایه دانش است (شوکلایس دیتیر، ۲۰۱۹) که بر اساس نیاز پژوهشی این مقاله، فرایندهای ذیل صورت پذیرفته است:

درو داده. درو داده یا جمع‌آوری داده از اینترنت بر اساس روش‌های مختلفی صورت می‌پذیرد که ما در این مقاله از خزشگر^۷ (کراولر)^۸ نوشته شده توسط آزمایشگاه شبکه‌های اجتماعی دانشگاه تهران^۹ استفاده نموده‌ایم. خزشگر تلگرام هم مشابه یک خزشگر وب، یک بات^{۱۰} است که با جستجو در میان کانال‌های عمومی (برای رعایت اخلاق در پژوهش،

۱ Elder, John

۲ Business Understanding

۳ Data Understanding

۴ Data Preparation

۵ Modeling

۶ Evaluation

۷ Deployment

۸ Data harvesting

۹ Data Cleaning

۱ Data Integration

۱ Data Selection

۱ Data Transformation

۱ Data Mining

۱ Pattern Evaluation

۱ Knowledge Representation

۱ Schewe, Klaus. Dieter

۱ Crawler/Spider

۱۸. خزشگرهای وب، نرم افزار یا اسکریپت‌های خودکاری هستند که پیوندهای اینترنتی را با یک الگوریتم مشخص و هدفمند دنبال می‌کنند و با فراخوانی وب سایت‌های مختلف، محتوای موجود آن‌ها را برای پایگاه داده‌های موتورهای جستجو و یا سایر هدف‌هایی که برنامه‌نویسان تعیین کرده‌اند ایندکس می‌نمایند.

۱ <http://sociallab.ir/>

۲ Bot

این مهم در کانال‌های خصوصی و یا گروه‌های تلگرامی صورت نمی‌پذیرد) کانال‌های جدید و محتوای جدید را پیدا کرده و در پایگاه داده خود ذخیره می‌کند.

تصفیه داده. در این فرایند پس از مجتمع کردن داده‌ها، نویز (نوفه) از مجموعه دیتاست حذف و تدابیری برای داده‌های گم شده تمهید می‌شود.

نرمال‌سازی داده. یک‌دست‌سازی و هم‌بازه کردن داده‌ها در این مرحله صورت می‌پذیرد.

یکپارچه‌سازی داده‌ها. در اغلب مسائل داده‌کاوی، داده‌ها از منابع داده گوناگون باید به یک‌باره مورد تحلیل قرار بگیرند که در این پژوهش به علت تک‌منبع بودن (تلگرام) از آن بهره‌گیری نشده است.

انتخاب داده. در فرایند انتخاب داده، باید داده‌های مرتبط با تحلیل انتخاب شده و از مجموعه داده برای انجام تحلیل‌ها بازیابی شوند؛ این انتخاب بر اساس لکسیکون تولید شده (که بعداً تشریح خواهد شد) و بر اساس میزان تولید و بازدید پست‌های تلگرامی صورت پذیرفته است.

تبدیل داده. تبدیل داده یک روش تثبیت داده نیز هست؛ در این فاز داده‌های انتخاب شده به فرم دیگری تبدیل می‌شوند؛ این کار به ساده‌تر شدن، بهبود صحت و دقت فرایند کاوش کمک می‌کند.

با توجه به این‌که هدف از این پژوهش شناسایی زودهنگام سیگنال‌های بحران‌های سیاسی^۴ در رسانه‌های اجتماعی است، ابتدا با روش متن‌کاوی، به‌عنوان یکی از کاربردهای

۱ Dataset

۲ Missing Values

۳ lexicon

۴. این بحران‌ها اشکال گوناگونی دارد که پس از مطالعه و بررسی به‌واسطه احصا تگ‌های مربوطه و حذف هم‌پوشانی‌ها در محتوای فارسی رسانه‌های اجتماعی با دسته‌بندی ذیل در این پژوهش لحاظ شده است: اعتصاب (Strike)، تظاهرات (Demonstration)، هم‌دیداری (Meeting)، شورش (Revolt)، جنبش (Movement)، قیام (Upheaval)، هرج‌ومرج (Chaos)، آشوب (Riot)، خرابکاری (Sabotage)، فعالیت‌زیرزمینی (Underground)، اغتشاش (Disturbance)، تهییج (Agitation)، ناآرامی (Restlessness)، خشم (Furious)،

داده‌کاوی و بر اساس ساخت یک لکسیکون از واژه‌های مرتبط با خشونت کلامی^۱ (کلاس‌بندی شده مشابه^۲ EMTerms و به‌عنوان سیگنال‌های یک بحران سیاسی (ایتمینکوآ و دیگران، ۲۰۱۵) داده‌ها جمع‌آوری شده است؛ این دیتا ست از خزش در کانال‌های عمومی تلگرام (بر اساس خزش صورت‌گرفته در آزمایشگاه شبکه‌های اجتماعی دانشگاه تهران از ۱۳۹۵۹۷۷ کانال عمومی در زمان پژوهش) در طول شش ماه صورت پذیرفته است و پست‌های مرتبط استخراج و با پالایش داده در نهایت میزان تولید محتوا و بازدید از آن‌ها در این بازه احصا گردید؛ از آنجاکه این شش ماه یک اسلات زمانی (یک محدوده زمانی درست قبل از آنکه موضوع به محتوای پرتفردار تبدیل شود) برای شناسایی سیگنال‌ها محسوب می‌شوند، بهترین روش برای مدل‌سازی و یافتن الگوها، تحلیل سری‌های زمانی است. سری زمانی دنباله‌ای از داده‌ها است که در یک محدوده زمانی مشخص جمع‌آوری شده‌اند؛ این داده‌ها تغییراتی که پدیده در طول زمان دچار آن شده است را نشان می‌دهند؛ بنابراین می‌توانیم این مقادارها را یک بردار وابسته به زمان بدانیم؛ در این حالت اگر X یک بردار باشد، سری زمانی را می‌توان به‌صورت زیر نشان داد که در آن t ، بیانگر زمان و X نیز یک متغیر تصادفی است.

$$X(t), t=0, 1, 2, \dots$$

طبق این تعریف زمان $t=0$ نیز قابل‌تعریف است؛ این لحظه می‌تواند زمان تولد یک پدیده یا هنگامی باشد که اولین اطلاعات در آن لحظه ثبت شده است؛ به‌این ترتیب $X(t)$ متغیر تصادفی X را در زمان t نشان می‌دهد. مقادارهای مشاهده شده این متغیر تصادفی دارای ترتیبی هستند که زمان وقوع هر داده را نشان می‌دهند؛ در این داده‌ها اغلب بین بازه‌های زمانی همبستگی وجود دارد و مقادیر حال وابسته به مقادیر گذشته می‌باشد؛ به‌عبارت‌دیگر می‌توان با استفاده از مقادیر گذشته مقادیر آینده را پیش‌بینی نمود. اگر یک

اضطراب (Anxiety)، تنش اجتماعی (Social Tension)، تشن (Convulsion)، بلوا (Riot, Calamity)، طغیان (Rebellion) و انقلاب (Revolution)

^۱ Verbal abuse
^۲ Crisis Lexicon
^۳ I. Temnikova

سری زمانی را کاملاً بتوان پیش‌بینی کرد، آن را سری زمانی غیرتصادفی می‌گویند، اما معمولاً نمی‌توان یک داده را به‌طور کامل پیش‌بینی نمود.

الگوی رفتار یا مدل تغییرات یک‌سری زمانی را می‌توان به چهار مؤلفهٔ روند،^۱ تناوب،^۲ تغییرات فصلی^۳ و تغییرات نامنظم^۴ تفکیک کرد. روند (تمایل سری زمانی به افزایش، کاهش یا حتی ثابت بودن، روند را تشکیل می‌دهد)، تناوب (تغییرات یکسان و تکراری در مقاطع میان‌مدت، تناوب در سری زمانی نامیده می‌شود)، تغییرات فصلی (در سری زمانی، تغییراتی که در دوره‌ای کوتاه‌تر از یک تناوب به‌صورت تکراری رخ می‌دهد، به تغییرات فصلی معروف است؛ همان‌طور که مشخص است دورهٔ تکرار تغییرات فصلی کوتاه‌تر از دورهٔ تکرار برای تغییرات تناوبی است) و تغییرات نامنظم (این‌گونه تغییرات بر اثر عوامل تصادفی و غیرقابل پیش‌بینی ایجاد می‌شوند)؛ هدف اصلی از تجزیه و تحلیل سری زمانی معمولاً پیش‌بینی مقادیر آینده می‌باشد. سه هدف دیگر از بررسی سری‌های زمانی که معمولاً نسبت به پیش‌بینی کمتر مورد توجه قرار می‌گیرد، توصیف^۵ شریح^۶ و کنترل^۷ می‌باشد. هنگامی امکان پیش‌بینی برای سری زمانی وجود دارد که به‌صورت ایستا^۸ تبدیل شده باشد؛ منظور از یک سری زمانی ایستا، دنباله‌ای از مقادیر وابسته به زمان است که میانگین و واریانس آن به زمان وابسته نباشند. (در سری زمانی ایستا، قوانین حاکم بر تغییرات مقادیرها، وابسته به زمان نیست) (مهاراج^۹ و دیگران، ۲۰۱۹)

برای ایستاکردن داده‌ها می‌توان از دو روش تفاضل‌گیری و تجزیهٔ کلاسیکی به اجزای تشکیل‌دهندهٔ مؤلفهٔ روند، مؤلفهٔ فصلی و مؤلفهٔ باقیمانده‌ها استفاده کرد. در روش تفاضل‌گیری هر مشاهده را از مشاهدهٔ قبلی کم می‌کنند. لازم به ذکر است در هر بار

۱) Trend

۲) Cyclic

۳) Seasonal

۴) Irregular

۵) Explanation

۶) Description

۷) Control

۸) Stationary

۹) Maharaj

تفاضل‌گیری یکی از تعداد بازه‌های زمانی کاسته می‌شود؛ اگر از یک تابع ایستا تفاضل‌گیری انجام شود، داده حاصل نیز ایستا است. (تفاضل‌گیری غیرضروری علاوه بر این‌که تعداد بازه‌های مورد بررسی را کم می‌کند، باعث پیچیده شدن داده‌ها و همبستگی کاذب بین داده‌ها می‌شود.)

بعد از شناسایی و حذف مؤلفه‌های اصلی سری زمانی، می‌توان عمل پیش‌بینی را انجام دهیم؛ یعنی مدل ریاضی برای ارتباط بین مقادیرهای سری زمانی را پیدا کنیم. دو مدل پیش‌بینی سری زمانی به نام میانگین متحرک^۱ و اوتورگرسیون^۲ وجود دارد. میانگین متحرک یک روش هموارسازی داده‌ها است؛ البته این عملگر، مؤلفه روند را برای سری زمانی برآورد می‌کند تا در زمان پیش‌بینی از آن استفاده شود.

از تجزیه و تحلیل سری زمانی معمولاً برای پیش‌بینی مقادیر آینده در علوم پزشکی و اقتصاد بیشتر استفاده می‌شود، توصیف داده‌ها معمولاً با استفاده از رسم نمودار صورت می‌گیرد؛ با استفاده از رسم نمودار می‌توان توصیف ساده‌ای از خواص داده‌ها، مانند صعودی یا نزولی بودن روند داده‌ها، وجود الگوی فصلی و وجود تغییر ناگهانی یا آرام در روند سری‌های زمانی انجام داد؛ دومین هدف از تجزیه و تحلیل سری زمانی تشریح است که از تغییرات در یک سری برای بیان تغییرات در سری دیگر استفاده می‌شود؛ سومین هدف کنترل می‌باشد که رابطه بسیار نزدیک با پیش‌بینی دارد؛ استفاده در جهت دادن هشدار قبل از طغیان یک کمیت مورد مطالعه می‌باشد که بیشتر در تحلیل تکنیکال^۳ و در بازارهای مالی استفاده می‌شود و ابزارهای مورد استفاده نیز متناسب با نیازهای این حوزه توسعه یافته‌اند. (حسن‌زاده و دیگران، ۹۴)

جهت پردازش سری‌های زمانی لازم است که مدلی (بر اساس الگویی که در سال ۱۹۷۰ توسط باکس و جنکینز^۴ ابداع شد و به سرعت نیز تکامل یافته است) داشته باشیم تا

^۱ moving average

^۲ Autoregressive

^۳ Technical analysis

^۴ Box & Jenkins

بتوانیم عملیاتی چون پیش‌بینی آینده سری و یا دسته‌بندی آن را بر اساس آن مدل انجام دهیم؛ با توجه به ماهیت داده‌های این پژوهش، مدل پیش‌بینی سری زمانی میانگین متحرک‌نمایی انتخاب گردید. در ادامه به بررسی روش محاسبات میانگین متحرک‌نمایی و مدل‌سازی ریاضی آن می‌پردازیم.

فرمول کلی میانگین متحرک‌نمایی عبارت است از:

$$EWMA(t) = a * x(t) + (1-a) * EWMA(t-1)$$

$EWMA(t) = t$ میانگین متحرک در زمان t

$a = 1$ و 0 بین مقدار پارامتر مختلط

$X(t) = t$ مقدار سیگنال x در زمان t

میانگین متحرک‌نمایی نوعی میانگین متحرک است که وزن و اهمیت بیشتری به جدیدترین داده‌ها می‌دهد. میانگین متحرک‌نمایی به‌عنوان میانگین متحرک وزنی‌نمایی شناخته می‌شود. میانگین متحرک‌نمایی در مقایسه با میانگین متحرک ساده که وزن برابری به همه مشاهدات در دوره زمانی می‌دهد نسبت به تغییرات داده‌های جدید واکنش قابل توجه‌تری نشان می‌دهد؛ در این مدل سری زمانی $X(t)$ به‌صورت زیر نوشته می‌شود:

$$X(t) = \theta_0 + Z(t) + \theta_1 Z(t-1) + \dots + \theta_q Z(t-q) \quad X(t) = \theta_0 + Z(t) + \theta_1 Z(t-1) + \dots + \theta_q Z(t-q)$$

که در آن $Z(t)$ همان خطای تصادفی با میانگین صفر و واریانس ثابت است؛ به این مدل میانگین متحرک مرتبه q گفته و به‌صورت $MA(q)$ نمایش داده می‌شود؛ در این مدل مقدار سری زمانی در زمان t یک ترکیب خطی از خطاها تا زمان t است. با توجه به تعریف مدل و صفر بودن میانگین $Z(t)$ ، مشخص است که میانگین سری زمانی ایستا برابر است با θ_0 .

۱. Exponential moving average

۲. انواع میانگین متحرک عبارتند از: ساده یا Simple، نمایی یا Exponential، مثلثی یا Triangular، تغییر یا Variable، وزنی یا Linear Weighted.

استفاده از این مدل باید با در نظر گرفتن یک شرط صورت پذیرد؛ شرطی که در این مدل باید رعایت شود؛ آن است که می‌بایست قدر مطلق ضرایب مدل کوچک‌تر از یک باشند. مدل میانگین متحرک درجه ۱ به صورت زیر نوشته می‌شود.

$$X(t) = Z(t) + \theta_1 Z(t-1)$$

اگر میزان خطا را برای زمان t به صورت $X(t) - Z(t)$ نشان دهیم، خواهیم داشت.

$$X(t) - Z(t) = \theta_1 Z(t-1)$$

به این ترتیب می‌توان به وسیله کمیته‌سازی مقدار مربعات خطا، پارامتر مدل را برآورد کرد. مجموع مربعات خطا در این مدل به صورت زیر است (ترنسی سی میل، ۲۰۱۹)

$$MSE = \sum_{t=1}^n (X(t) - Z(t))^2 = \sum_{t=1}^n (\theta_1 Z(t-1))^2$$

در نهایت برای تحلیل داده‌های محاسبه شده از روش فوق برای اولین بار از اندیکاتور $MACD^2$ استفاده شده است که در یافته‌های پژوهش به آن خواهیم پرداخت. برای روایی و اعتبارسنجی پژوهش نیز از آزمایش خارج از نمونه^۳ با توجه به در دسترس بودن داده‌های کل سال ۹۷ بهره‌گیری شده است.

۳. تجزیه و تحلیل یافته‌ها

در مدل‌سازی داده‌های سری زمانی بازنمایی رویدادهای سیاسی سال ۹۷ در پلتفرم تلگرام از روش میانگین متحرک‌نمایی استفاده شده است؛ از آنجایی که میانگین متحرک‌نمایی به‌عنوان یکی از روش‌های اصلی هموارسازی داده^۴ سریع‌تر نسبت به تغییرات در مدل واکنش نشان دهد، مبنای مناسبی برای شناسایی روندها در داده‌های استخراج شده

^۱ Terence C. Mills

^۲ Moving Average Convergence Divergence

^۳ Out. of. sample testing

^۴ Smoothing Methods

است؛ یکی از روش‌های تحلیل این مدل اندیکاتور میانگین متحرک همگرایی - واگرایی^۱ که به اختصار MACD (مکدی) خوانده می‌شود که یک اندیکاتور مومنتوم تعقیب‌گر روند است و در تحلیل تکنیکال، رابطه بین دو میانگین متحرک در نمودار قیمت اوراق بهادار را نشان می‌دهد (مورفی، ۱۳۹۹).

اندیکاتور مکدی توسط جرال د اپل محقق آمریکایی در اواخر سال ۱۹۷۰ طراحی شد؛ مکدی در خانواده نوسان‌گرها قرار می‌گیرد و در سال‌های اخیر یکی از پر استفاده‌ترین اندیکاتورها در بین تحلیلگران بازارهای مالی است؛ همان‌طور که از نام مکدی پیداست یکی از کاربردهای آن بررسی واگرایی و همگرایی می‌باشد که برای تحلیلگران بازارهای مالی حائز اهمیت می‌باشد. از ویژگی‌های مثبت مکدی ارائه سیگنال واگرایی به بهترین نحو ممکن است. اگر قیاسی بین خانواده نوسانگرها داشته باشیم متوجه خواهیم شد که مکدی در تشخیص ضعف روند بسیار هوشمند عمل می‌کند و به ادعای جرال د اپل بهترین اندیکاتور جهت بررسی قدرت روند می‌باشد؛ این اندیکاتور شامل دو میانگین متحرک‌نمایی به نام‌های «خط مکدی» و «خط سیگنال» می‌باشد که در محاسبات میانگین مکدی در بازارهای مالی به قیمت‌های نزدیک‌تر به روزهای آخر معاملاتی وزن بیشتری داده شده است. خط مکدی ویژگی تند و پرنوسانی دارد ولی خط سیگنال بسیار کند و کم‌تلاطم است که دلیل این تمایز، متفاوت بودن میانگین‌گیری در خط مکدی و خط سیگنال می‌باشد. جزء سوم مکدی هیستوگرام است که در قالب خطوط عمودی اختلاف خط مکدی و خط سیگنال را نمایش می‌دهند (مورفی، ۱۳۹۹).

^۱ Moving Average Convergence Divergence
^۲ Trend. following momentum indicator

اندیکاتورهای تحلیل سری های زمانی در بورس	روند (Trend)	(Average Directional Movement Index) میانگین جهت دار
		(Bollinger Bands) باند بولینگر
		(Envelopes) پوشاننده
		(Ichimoku Kinko Hyo) ایچیمیکو
		(Moving Average) میانگین متحرک
		(Parabolic SAR) پارابولیک سار
		(Standard Deviation) انحراف معیار
		(Average True Range) میانگین محدوده واقعی
		(Bears Power) قدرت خرس ها
		(Bulls Power) قدرت گاو ها
		(Commodity Channel Index) شاخص کانال کالا
		(DeMarker) دیمارکر
	(Force Index) شاخص قدرت	
	(MACD) میانگین متحرک همگرایی و واگرایی	
	(Momentum) اندازه حرکت	
	(Moving Average of Oscillator) میانگین متحرک نوسانگر	
	(Relative Strength Index) شاخص قدرت نسبی	
	(Stochastic Oscillator) نوسان ساز تصادفی	
	(Williams' Percent Range) دامنه درصد ویلیامز	
	(Accumulation/Distribution) تراکم/توزیع	
	(Money Flow Index) شاخص جریان نقدینگی	
	(On Balance Volume) حجم تعادلی	
	(Volumes) حجم معامله	
	(Accelerator Oscillator) آشکار کننده شتاب حرکت قیمت	
	(Alligator) تمساح	
	(Awesome Oscillator) نوسانگر مهیب	
	(Fractals) فراکتال	
	(Gator Oscillator) مکمل نوسان نمای تمساح	
(Market Facilitation Index) شاخص تسهیل بازار		

جدول (۲) - اندیکاتورهای تحلیل سری های زمانی بورس

خط مکدی به عنوان یک خط پرنوسان و سریع؛ همواره با خط سیگنال در حرکت است که گاهی اوقات در تقاطع یکدیگر نیز قرار می گیرند که در این حالت و در شرایطی خاص سیگنال های مهمی را به تحلیلگران ارائه می دهند؛ از طرفی نیز حرکات خط مکدی با توجه به پیروی از نمودار قیمت، تشکیل پیوت هایی را می دهد که در بررسی هماهنگی یا

pivot.)

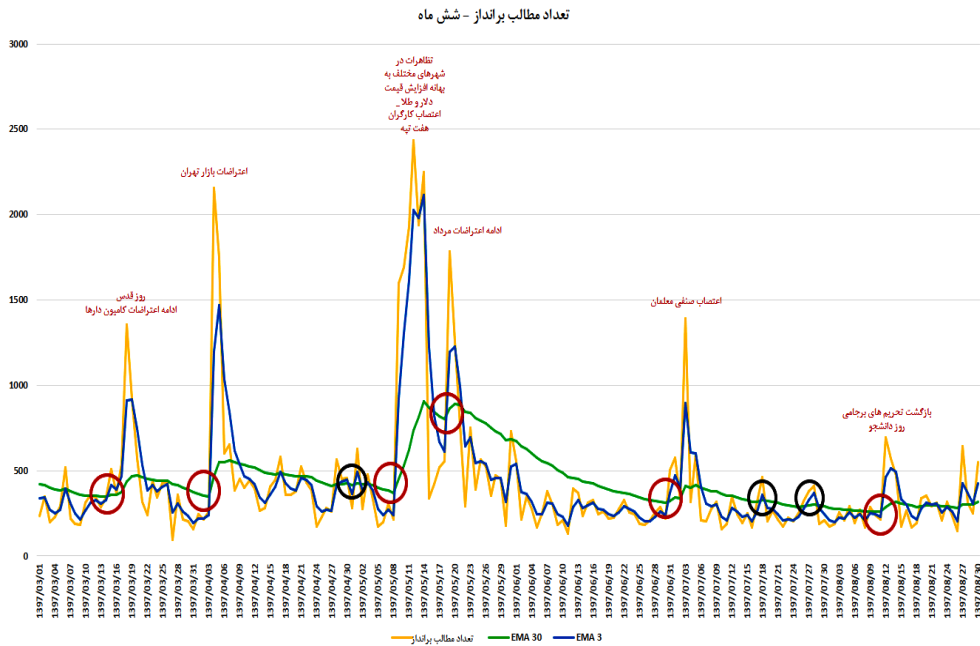
پیوت در تحلیل تکنیکال به نقاط مهم و محوری گفته می شود که با توجه به آنها از ابزار های تحلیل تکنیکال مانند خط روند، الگو ها و ... استفاده می شود.

ناهماهنگی رفتار نمودار قیمت و خط مکدی سیگنال‌های مهم دیگری نیز صادر می‌شود. در محاسبه خط مکدی دو میانگین متحرک از نوع EMA یا نمایی، میانگین نمایی ۱۲ دوره‌ای قیمت و میانگین نمایی ۲۶ دوره‌ای قیمت را تفاضل گیری می‌نمایند. (تفاضل میانگین ۱۲ دوره‌ای از میانگین ۲۶ دوره‌ای خط مکدی را نمایش می‌دهد)

همان‌طور که اشاره شد سیگنال مکدی از تقاطع خط مکدی و خط سیگنال صادر می‌شود؛ زمانی که خط مکدی (خط سبز) از ناحیه زیر صفر، خط سیگنال (خطر قرمز) را از پایین به بالا قطع می‌کند سیگنال خرید صادر می‌شود؛ همچنین زمانی که خط مکدی (خط سبز) از ناحیه بالای صفر، خط سیگنال (خطر قرمز) را از بالا به بالا قطع کند؛ در آن صورت سیگنال فروش صادر خواهد شد. باید توجه داشت که تقاطع خط مکدی و خط سیگنال در نواحی اشاره شده و با ویژگی‌های فوق به‌عنوان سیگنال محسوب می‌شوند و عدم وقوع هریک از موارد فوق سیگنال مکدی را نقض می‌کند. در تحلیل از طریق هیستوگرام نیز زمانی که فاصله بین دو خط در حال بیشتر یا کمتر شدن (بازتر یا بسته شدن) باشد؛ اگر هیستوگرام در موقعیت فوقانی خط صفر مخصوص خود قرار گرفته باشد؛ اما در عین حال، به سمت پایین شروع به کم شدن کند، نشانه ضعیف‌شدن روند صعودی است؛ اما برعکس این حالت برای زمانی است که هیستوگرام در موقعیت پایینی خط صفر مخصوص خود قرار گرفته باشد و به سمت بالا یا خط صفر، شروع به افزایش می‌کند که در این موقع، روند صعودی در حال ضعیف‌شدن است؛ زمانی که هیستوگرام از خط صفر مخصوص خود عبور کرده باشد، اختار خرید و فروش ارسال می‌گردد و هر چرخشی که در هیستوگرام مکدی رخ دهد؛ به معنای هشدار زودهنگام مربوط به ضعیف‌شدن روند فعلی است (گرایمز، ۲۰۱۲).

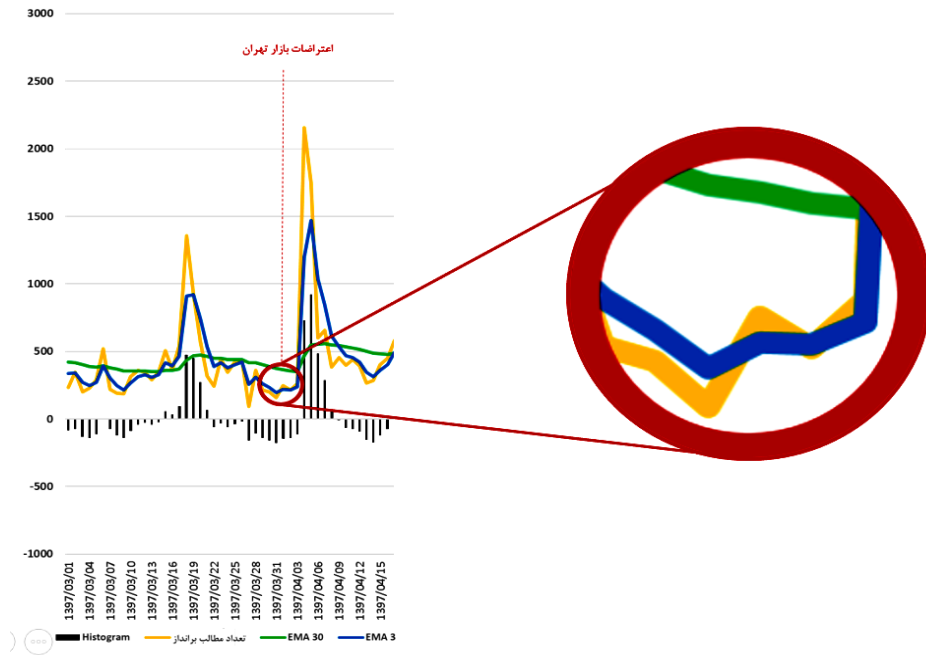


در این پژوهش برای نخستین بار از این تکنیک برای شناسایی سیگنال‌های بحران در پلتفرم تلگرام استفاده شده است. معمولاً مکدی از تفریق میانگین متحرک نمای ۲۶ دوره‌ای از میانگین متحرک نمای ۱۲ دوره‌ای به دست می‌آید؛ نتیجه این محاسبه خط مکدی است که خط سیگنال نامیده می‌شود که می‌تواند نشانگر سیگنال‌های خرید و فروش باشد. معامله‌گران معمولاً اوراق بهادار را هنگامی خریداری می‌کنند که مکدی از خط سیگنال خود عبور کرده و بالاتر رفته باشد و اوراق بهادار را هنگامی می‌فروشند که مکدی از خط سیگنال عبور کرده و پایین‌تر رفته باشد؛ از آنجایی که در بحران‌ها و رویدادهای سیاسی تأثیر عوامل طبیعی (نارضایتی مردم، خشم، ترس و...) و غیرطبیعی (جریان‌های سازماندهی شده معارضان حکومت و یا موافقان آن) خود را در میزان تولید محتوای رسانه‌های اجتماعی نشان می‌دهد؛ در این پژوهش با مدل‌سازی و واکافت میزان تولید محتوای مربوط به خشونت کلامی با اندیکاتور مکدی در دوره ۳ و ۳۰ روزه (این واکافت برای دوره متداول ۲ و ۲۹ روز نیز امتحان شد که دقت نتایج دوره ۳ و ۳۰ بهتر بوده است) در داده‌های سال ۹۷ تلگرام، با هدف امکان پیش‌بینی رویدادها مورد بررسی قرار گرفته است؛ همان‌گونه در شکل شماره (۱) مشاهده می‌شود



شکل شماره (۱) - روندنمای مطالب برانداز در سکوی تلگرام

بازنمایی تمامی رویدادهای سیاسی سال ۹۷ در پیام‌رسان تلگرام با این شیوه؛ پیش از رخداد قابل‌شناسایی زودهنگام است؛ به‌عنوان مثال مطابق استراتژی مکدی زمانی‌که بین میانگین متحرک ۳ روز و ۳۰ روز میزان تولید محتوای پست‌های خوشونت کلامی دچار صعود شود؛ می‌توان سیگنال هشدار در نظر گرفت (شکل ۲) - اعتراضات بازار). به-عبارتی نتایج پژوهش نشانگر آن است که اندیکاتور مکدی در مدل سری زمانی میانگین متحرک‌نمایی می‌تواند به‌عنوان یک روش شناسایی زودهنگام سیگنال‌های بحران (و به‌عنوان هشدار) می‌تواند به کار رود.



شکل (۲) - واکافت مکدی مدل سری زمانی میانگین متحرک مربوط به اعتراضات بازار

در این اسلات داده، خطای منفی کاذب^۱ وجود نداشته است؛ اما سه مورد خطای مثبت کاذب^۲ در مجموع ۹ رویداد مشاهده می شود که دقت قابل قبولی محسوب می شود.

نوآوری پژوهش

بر اساس مطالعه ادبیات پژوهش تقریباً اغلب پژوهش ها با تمرکز بر شبکه اجتماعی توییتر به علت امکان API آن برای پژوهشگران، محدودیت داده ای توییتر (تعداد کاراکتر پیام) و واکافت ساده تر نسبت به سایر شبکه های اجتماعی استفاده شده است؛ این پژوهش برای اولین بار و با توجه به فراگیری پیام رسان تلگرام در بین آحاد جامعه ایرانی (توییتر بیشتر قشر نخبگانی و رسانه ای جامعه را معطوف به خود نموده است) و علی رغم همه

^۱ false negative error

^۲ false positive error

محدودیت‌ها (خزش در تلگرام توسط ربات‌های خزشگر با محدودیت‌های جدی مواجهه است) صورت پذیرفته است. مهم‌ترین نوآوری پژوهش؛ تلفیق علوم اجتماعی محاسباتی و مدل‌های ریاضی با بهره‌گیری از اندیکاتورهای تحلیل سری زمانی است که پیش از این سابقه ندارد. تحلیل سری‌های زمانی بیشتر در علوم اقتصادی و سلامت مورد استفاده قرار گرفته و در علوم اجتماعی کمتر بدان پرداخته شده است؛ به‌عنوان مثال در بررسی پایگاه داده‌های مقالات فارسی دانشگاهی جهاد دانشگاهی sid.ir از مجموع ۲۴۴ مقاله با موضوع سری زمانی ۱۳ عنوان در حوزه پزشکی ۸۹ عنوان در حوزه کشاورزی و منابع طبیعی، ۴۸ عنوان در حوزه فنی و مهندسی، ۲۳ عنوان در حوزه علوم پایه و ۷۱ عنوان در حوزه علوم انسانی‌اند که بررسی نویسندگان این مقاله بر روی چکیده مقالات حوزه علوم انسانی نشان داد که موضوعات پرداخته شده در مقالات عموماً ذیل سه دسته جغرافیا، اقتصاد و حسابداری قرار دارند. در مجموعه پایگاه داده‌های پایان‌نامه‌های پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) با موضوع سری زمانی در مقطع کارشناسی ارشد و دکترا مجموعاً ۴۳۲۲ پایان‌نامه ثبت شده است که بررسی فراوانی کلیدواژه‌های چکیده‌های این پایان‌نامه‌ها نشانگر آن بود که کلیدواژه بورس اوراق بهادار تهران ۱۴۲ مورد، سری زمانی (انگلیسی) ۱۳۸، سری زمانی (علم فیزیک - انگلیسی) ۱۲۸ مورد، سری زمانی (فیزیک - فارسی) ۱۱۸ مورد، سری زمانی (بررسی و توسعه مفهوم سری زمانی) ۱۰۹ مورد و شبکه عصبی (فنی و مهندسی) ۹۰ مورد کلیدواژه‌های اصلی پایان‌نامه‌ها و رساله‌های صورت پذیرفته است؛ همان‌گونه که مشاهده می‌شود بر اساس این کلیدواژه‌ها بیشتر موضوعات مورد پژوهش (غیر از بحث سلامت و پزشکی که در این پایگاه داده ایندکس نشده است). با محوریت موضوع علوم اقتصادی، جغرافیا، آمار، فیزیک و علوم مهندسی است. نتایج حاصله از پژوهش، نشانگر افق‌های جدید با توسعه این روش‌ها در پیش‌بینی رخدادهای اقتصادی، اجتماعی، سیاسی و امنیتی با استفاده از واکنش‌های ریاضی بر روی کلان‌داده‌هاست.

۴. نتیجه گیری

از آنجاکه شناسایی زودهنگام رویدادها و بحران‌های سیاسی، اقتصادی و اجتماعی از اهمیت بالایی برای حاکمیت‌ها در مدیریت جامعه برخوردار است؛ روشی که بتواند با استفاده از واکافت کلان داده‌ها به این هدف پردازد در قالب یک سیستم می‌تواند به‌عنوان یک پدافند سایبری از هزینه‌های انسانی، اقتصادی، سیاسی و اجتماعی بحران‌ها و رخدادها کاسته و موجب بالا بردن آگاهی موقعیتی شده و از آن می‌توان در جهت تأمین امنیت ملی پایدار بهره جست.

اگرچه در این پژوهش با مدل‌سازی ریاضی و مطالعه گذشته‌نگر داده‌ها به شناسایی سریع ظهور بحران در رسانه‌های اجتماعی پرکاربر پرداخته شد؛ اما یافته‌های پژوهش مؤید آن است که در صورت وجود سامانه‌ای که بتواند از چنین واکافت‌هایی در کلان‌داده‌ها به‌صورت بلادرنگ استفاده کند؛ کارایی لازم را برای اعلان هشدار سریع و اندازه‌گیری مخاطرات به‌صورت برخط خواهد داشت.

در این پژوهش نیز شناسایی زودهنگام اعتراضات سال ۹۷ و بازتاب آن در فضای مجازی به‌عنوان سیگنال‌های بحران مدنظر بوده است. این اعتراضات عموماً بر مسائل و دغدغه‌های اقتصادی متمرکز بوده و سعی داشتند توزیع منابع دولتی را تغییر دهند. به‌عنوان نمونه در اردیبهشت‌ماه کامیون‌داران خواستار افزایش دستمزدها شدند و پس از آن، اعتصاب کردند. بازار تهران نیز در ۴ تیر به‌دلیل نگرانی از تأثیر تحریم‌ها و کاهش ارزش ریال تعطیل شد. در ۵ مرداد ماه نیز اعتراضاتی در شهرهای ایران از جمله تهران، اراک، اصفهان، کرج و شیراز به‌دلیل افزایش چشمگیر قیمت‌ها صورت گرفت.

بر اساس مدل‌سازی سری زمانی بر روی داده‌های خزش شده یک سال کانال‌های عمومی تلگرام (به‌عنوان پیام‌رسان پرکاربر جامعه ایرانی) در این پژوهش، اندیکاتور مکدی (اندیکاتور مومنتوم تعقیب‌گر روند میانگین متحرک همگرایی و اگرایی که پیش‌تر در حوزه اقتصاد مورد بهره‌برداری قرار می‌گرفته)، توانست در مدل بازنمایی اعتراض‌های سال ۱۳۹۷ در تلگرام با دقت قابل قبولی به‌عنوان یکی از ابزارهای پیش‌بینی بحران‌های سیاسی برای

اولین بار مطرح گردد و نشان داد که حداقل شش اعتراض سیاسی کشور در سال ۹۷ (از طریق واکافت صورت پذیرفته در این پژوهش بر کلان داده‌ها) با دقت قابل قبولی پیش از رخداد قابل‌شناسایی و مدیریت بوده است.

پیشنهاد

از آنجایی که بهره‌گیری از طیف متنوعی از ابزارهای تحلیل سری‌های زمانی در علوم محاسباتی اجتماعی در افزایش دقت و اثربخشی مدل مؤثر است؛ پیشنهاد می‌شود از سایر ابزارها (اندیکاتورها و اسیلاتورها) در توسعه این پژوهش بهره گرفته شود. (در این پژوهش از یکی از این ابزارها استفاده شده است)؛ به‌عنوان مثال بررسی مدل‌سازی میانگین متحرک‌نمایی و استفاده از اندیکاتورها و اسیلاتورهای پیشرفته مختلف و واکافت مربوطه مانند ابر ایچیموکو^۲ یا الگوهای نمودارهای شمعی^۳ نیز که بر مبنای واکافت میانگین متحرک در بازارهای مالی و بورس نیز مورد می‌تواند در داده‌های مشابه، جهت ارزیابی دقت و قابلیت شناسایی روندهای سیاسی یا اجتماعی مورد استفاده قرار می‌گیرند. همچنین توسعه این‌گونه ابزارها (با رویکرد ایجاد یک ابزار اختصاصی) به‌موازات توسعه پردازش^۴ و فهم زبان طبیعی^۵، یادگیری ماشینی^۶ و توجه به روش‌های متن‌کاوی برای این عرصه علمی میان‌رشته‌ای بسیار ضروری است.

^۱ oscillator

^۲ Ichimoku Cloud

^۳ candlestick chart

^۴ Natural language processing

^۵ Natural language understanding

^۶ Machine learning

فهرست منابع و مآخذ

الف. منابع فارسی

- جان مورفی (۱۳۹۹)، تحلیل تکنیکال در بازارهای سرمایه، ترجمه، کامیار فراهانی فرد، رضا قاسمیان لنگرودی، تهران: چالش
- سهرابی، محمد؛ مسلمی، سیاوش (۱۳۹۴)، راهبردهای شبکه‌های اجتماعی مجازی برای کنترل بحران امنیتی، فصلنامه علمی امنیت ملی ۳۲-۹، ۵(۱۵)
- شجاع مؤدب، حمیدرضا؛ حسینی تاش، سید علی؛ آقایی، محسن؛ امیری مقدم، رضا (۱۳۹۹)؛ ارائه مدل تحول در چرخه‌های ظهور بحران‌های سیاسی - امنیتی، متأثر از فضای سایبر، فصلنامه علمی امنیت ملی ۲۰۴-۱۷۹، ۱۰(۳۵).
- عباسی، حسن؛ شریعت، جهانگیر (۱۳۹۷)، کارکرد سیاسی شبکه‌های اجتماعی مجازی، فصلنامه علمی امنیت ملی ۱۵۰-۱۲۷، ۸(۳۰).

ب. منابع انگلیسی

- Chang-Won Kim, Wi Sung Yoo, Hyunsu Lim, Ilhan Yu, Hunhee Cho, Kyung-In Kang, (2018) Early-warning performance monitoring system (EPMS) using the business information of a project, International Journal of Project Management, Volume 36, Issue 5, Pages 730-743.
- D. Wang, B. K. Szymanski, T. Abdelzaher, H. Ji and L. Kaplan, "The Age of Social Sensing," in *Computer*, vol. 52, no. 1, pp. 36-45, Jan. 2019, doi: 10.1109/MC.2018.2890173.
- Kruspe, Anna & Kersten, Jens & Stein, Benno & Klan, Friederike. (2018). Classification of Incident-related Tweets: Tackling Imbalanced Training Data using Hybrid CNNs and Translation-based Data Augmentation.
- MacAvaney S, Yao H-R, Yang E, Russell K, Goharian N, Frieder O (2019) Hate speech detection: Challenges and solutions. *PLoS ONE* 14(8): e0221152.
- Maharaj, E. A., D'Urso, P., & Caiado, J. (2019). Time series clustering and classification. CRC Press
- Molaei, M. Kargari, M. SheikhMohammady, A. Akramizadeh (2019) Deterrence Model in Cyberspace Based on Bayesian Belief Attack Graph by using Risk Creating Payoff Function, *Journal of Electronical & Cyber Defence*, Vol. ۷, No. 1, P25-38
- Murphy, J. J., & Murphy, J. J. (1999). Technical analysis of the financial markets: A comprehensive guide to trading methods and applications. New York: New York Institute of Finance.
- Mustafa, Raza & Nawaz, M. Saqib & Ferzund, Javed & Lali, Muhammad Ikram & Shahzad, Basit & Fournier Viger, Philippe. (2017). Early Detection of Controversial Urdu Speeches from Social Media. *Data Science and Pattern Recognition*. 1. 26-42.

- Nazir, F., Ghazanfar, M.A., Maqsood, M. et al. Social media signal detection using tweets volume, hashtag, and sentiment analysis. *Multimed Tools Appl* 78, 3553–3586 (2019).
- Peiwan Wang, Lu Zong, Ye Ma, An (2020) integrated early warning system for stock market turbulence, *Expert Systems with Applications*, Volume 153.
- Phillips, Lawrence & Dowling, Chase & Shaffer, Kyle & Hodas, Nathan & Volkova, Svitlana. (2017). Using Social Media to Predict the Future: A Systematic Literature Review.
- Schewe, Klaus-Dieter & Singh, Neeraj. (2019). Model and Data Engineering 9th International Conference, MEDI 2019, Toulouse, France, October 28-31
- Sofyan Sufri, Febi Dwirahmadi, Dung Phung, Shannon Rutherford, (2020) A systematic review of Community Engagement (CE) in Disaster Early Warning Systems (EWSs), *Progress in Disaster Science*, Volume 5.
- Stieglitz, S., Bunker, D., Mirbabaie, M., & Ehnis, C. (2017a). Sense-Making in Social Media During Extreme Events. *Journal of Contingencies and Crisis Management (JCCM)*.
- Terence C. Mills. (2019). *Applied Time Series Analysis*, Academic Press, Pages 1-12
- Varol, O., Ferrara, E., Menczer, F. et al. (2017) Early detection of promoted campaigns on social media. *EPJ Data Sci.* 6, 13.
- Wu Jiekang, Wu Zhijiang, Mao Xiaoming, Wu Fan, Tang Huiling, Chen Lingming (2020), Risk early warning method for distribution system with sources-networks-loads-vehicles based on fuzzy C-mean clustering, *Electric Power Systems Research*, Volume 180.
- Yi Chai, Hao Luo, Qingpeng Zhang, Qijin Cheng, Carrie S.M. Lui, Paul S.F. Yip, Developing an early warning system of suicide using Google Trends and media reporting, *Journal of Affective Disorders*, Volume 255, 2019, Pages 41-49.
- Zhang, X., Zhong, Q., Zhang, R., & Zhang, M. (2020). People-centered early warning systems in China: A bibliometric analysis of policy documents. *International journal of disaster risk reduction: IJDRR*, 51, 101877.