Strategic data analysis model in social networks

Rasoul Lotfi[1]
Kamran Feizi[2]
Sayyed Mohammad Khatami Firoozabadi[3]
Mohammadreza Taghva[4]

Type of article: Research article extracted from doctoral dissertation

Received:2021/10/18 Accepted: 2021/12/6

NAJA Strategic Studies Quarterly/Vol.7/NO.1/(serial23)Spring 2022∗83-104

DOR: 20.1001.1.25381946./ssj.2022.98600

Abstract

Today, the Internet, with the creation and development of global communication platforms, has been able to play a fundamental and strategic role in governing cyberspace, and one of these spaces that has attracted a large audience is social networks. These networks have created an environment where users can overcome the limitations and barriers of their real life space and take full advantage of the opportunity provided for their interaction, communication and message transmission; Thus, the ability to analyze social networks on a large scale is a specific technical capability that each organization must implement in accordance with its mission. In the field of police actions, in order for the police management structure to be able to assess the impact of its actions on public opinion, it must be made public; therefore, in this study, in order to show how to analyze the strategic data of social networks, the pattern of data collection and processing has been drawn and in order to turn the mental abstract pattern into a tangible practical pattern, using content analysis and text analysis to examine the views of Iranian users on Twitter. Has been. In total, after deleting additional tweets, 39,000 samples were used for text analysis. In order to analyze the tweets, the samples were done based on the users' point of view. The duration of information analysis was performed in a period of 3 years (1397-1400) and during the description, data mining structure and text analysis were mentioned to understand the differences and similarities of analytical models.

Keywords: social network, twitter, text mining, feta police, strategic analysis

1. PhD Student in Information Technology Management, Faculty of Management, Allameh Tabatabaei University, Tehran, Iran (Corresponding Author), iamlotfi@yahoo.com
2. Professor, Department of Industrial Management, Faculty of Management, Allameh Tabatabai University, Tehran, Iran
3. Professor, Department of Industrial Management, Faculty of Management, Allameh Tabatabai University, Tehran, Iran
4. Professor, Department of Industrial Management, Faculty of Management, Allameh Tabatabai University, Tehran, Iran

الگوی تحلیل داده‌های راهبردی در شبکه‌های اجتماعی

رسول لطفی[1]
کامران فیضی[2]
سیدمحمد خاتمی فیروزآبادی[3]
محمدرضا تقوا[4]

نوع مقاله: مقاله پژوهشی مستخرج از رساله دکتری

تاریخ دریافت: ۱۴۰۰/۷/۲۶ تاریخ پذیرش نهایی: ۱۴۰۰/۹/۱۵

فصلنامه مطالعات راهبردی ناجا/سال هفتم/شماره۱(پیاپی ۲۳)-بهار ۱۴۰۱* ۸۳-۱۰۴

DOR: 20.1001.1.25381946./ssj.2022.98600

چکیده

امروزه، اینترنت با ایجاد و توسعه پلتفرم‌های ارتباطی جهانی توانسته‌است نقش اساسی و راهبردی را در حکمرانی فضای مجازی بازی کند و یکی از این فضاها که مخاطبان زیادی را جذب نموده‌است، شبکه‌های اجتماعی است. این شبکه‌ها، فضایی ایجاد کرده‌اند که کاربران می‌توانند محدودیت‌ها و موانع فضای واقعی زندگی خود را جبران کرده و از فرصتی که برای تعاملات، ارتباطات و انتقال پیام آنها فراهم شده‌است، به‌خوبی بهره ببرند؛ از این رو، توانایی تحلیل شبکه‌های اجتماعی در مقیاس وسیع، یک توانمندی خاص فنی است که هر سازمان به‌تناسب مأموریت خود باید آن را پیاده‌سازی کند. در حوزه اقدامات پلیس نیز برای آنکه ساختار مدیریتی پلیس بتواند تأثیر اقدامات خود را بر افکار عمومی بسنجد، باید از نظر عموم جامعه آگاه شود؛ بنابراین، در این پژوهش، به‌منظور نشان‌دادن چگونگی تحلیل داده‌های راهبردی شبکه‌های اجتماعی، الگوی جمع‌آوری و پردازش داده‌ها ترسیم شده‌است و به‌منظور تبدیل الگوی انتزاعی ذهنی به الگوی کاربردی ملموس، با استفاده از تحلیل‌محتوا و متن‌کاوی به بررسی دیدگاه کاربران ایرانی در توییتر نسبت‌به پلیس فتا پرداخته شده‌است. درمجموع، پس از حذف توییت‌های اضافی، تعداد ۳۹۰۰۰ نمونه برای متن‌کاوی استفاده شدند. به‌منظور تحلیل توییت‌ها، نمونه‌ها بر اساس دیدگاه کاربران انجام پذیرفت. مدت زمان تحلیل‌اطلاعات در یک بازه زمانی ۳ ساله(۱۳۹۷ تا ۱۴۰۰) انجام شد و در طی تشریح مطالب، به ساختار داده‌کاوی و تحلیل‌متن اشاره شد تا تمایز و تشابه الگوهای تحلیلی درک شود.

واژگان کلیدی: شبکه اجتماعی، توییتر، متن‌کاوی، پلیس فتا، تحلیل راهبردی

---

۱. دانشجوی دکتری مدیریت فناوری اطلاعات، دانشکده مدیریت، دانشگاه علامه طباطبایی(ره)، تهران، ایران(نویسنده مسئول)، iamlotfi@yahoo.com

۲. استاد گروه مدیریت صنعتی، دانشکده مدیریت، دانشگاه علامه طباطبایی(ره)، تهران، ایران
۳. استاد گروه مدیریت صنعتی، دانشکده مدیریت، دانشگاه علامه طباطبایی(ره)، تهران، ایران
۴. استاد گروه مدیریت صنعتی، دانشکده مدیریت، دانشگاه علامه طباطبایی(ره)، تهران، ایران

مقدمه

امروزه، اینترنت با ایجاد و توسعه پلتفرم‌های ارتباطی جهانی توانسته‌است نقش اساسی و راهبردی را در حکمرانی فضای مجازی بازی کند؛ بر این اســاس، روشــن اســت که نمی‌توان نقش فناوری اطلاعات را در تمامی عرصه‌های مرتبط بشری نادیده گرفت. یکی از این فضاها که مخاطبان زیادی را جذب نموده‌اســت، شبکه‌های اجتماعی است. شبکه‌های اجتماعی مجازی اغلب سرویس‌های مبتنی‌بر وب هستند و سرویس‌های آنلاین شامل پلتفرم‌ها یا سایت‌هایی هستند که افراد از طریق آنها نظرها و علاقه‌مندی‌های خود را بیان کرده و با دیگران به اشتراک می‌گذارند. آندریس کاپلان و مایکل هایلین[1] رسانه اجتماعی را به‌مثابه یک دسته ابزار مبتنی‌بر اینترنت تعریف می‌کنند که بر بنیاد ایدئولوژیک و فناوری وب ۲ اســتوار هســتند و به کاربر امــکان تولید محتوا و مبادله آن را می‌دهند(کاپلان و هایلین، ۲۰۱۰) این شــبکه‌ها، فضایی ایجــاد کرده‌اند که کاربران می‌توانند محدودیت‌هــا و موانع فضای واقعــی زندگی خود را جبران کــرده و از فرصتی که برای تعاملات، ارتباطات و انتقال پیام آنها فراهم شده‌است، به‌خوبی بهره ببرند(اومالی۲۰۱۴[2]). از جمله کارکردهای شبکه‌های اجتماعی، نقش بازخوردی آنهاست؛ به‌گونه‌ای که افراد، سازمان‌ها و دولت‌ها می‌توانند بازخورد اقدامات خود را در این شبکه‌ها جستجو کنند و از آن برای جمع‌آوری پردازش و تحلیل راهبردی داده‌ها اســتفاده کنند؛ چراکه ســرویس‌های مبتنی‌بر شبکه که بتوانند داده‌های کلان[3] از نظرات آرا و کنش و واکنش‌های کاربران را دریافت نمایند، در دســترس اســت و سرویس‌های تحلیل‌داده که بتواند با استفاده از نرم‌افزارهای تحلیلی و هوش مصنوعی[4] آنها را تحلیل کرده و موجب خروجی تصمیم‌ساز شوند، قابل‌اجرا و پیاده‌سازی است. در حال حاضر، شبکه‌های اجتماعی بخش جدایی‌ناپذیر جامعه انسانی شده‌است و مردم در این شبکه‌ها عقاید خود را درخصوص موضوعات مختلف بیان می‌کنند؛ به‌عنوان مثال، روز پنجشنبه ۴ اوت ۲۰۱۱، مارک دوگان توسط یک افسر پلیس در تاتنهام کشته شد، صبح روز ششم، محتوای رسانه‌های اجتماعی خصومت فزاینده‌ای را نشان می‌دادند که تهدیدهای صریح بر ضدّ پلیس بود. از روز هفتم، اطلاعات رسانه‌های اجتماعی نشــان‌دهنده گســترش احتمالی اعتراضات بود، اختلال در سایر نقاط لندن و سپس انگلستان را درگیرکرد. در این رخ‌داد، داده‌های زیادی منتشــر شــد که می‌توانست درخصوص اطلاعات غلط درحال‌انتشــار و افراد پشــت آن با هویت مجعول به پلیس کمک کند. بعدها پلیس اذعان کرد که

1. Kaplan & Haenlein
2. Umali
3. Big Data
4. AI(Artificial Intelligence)

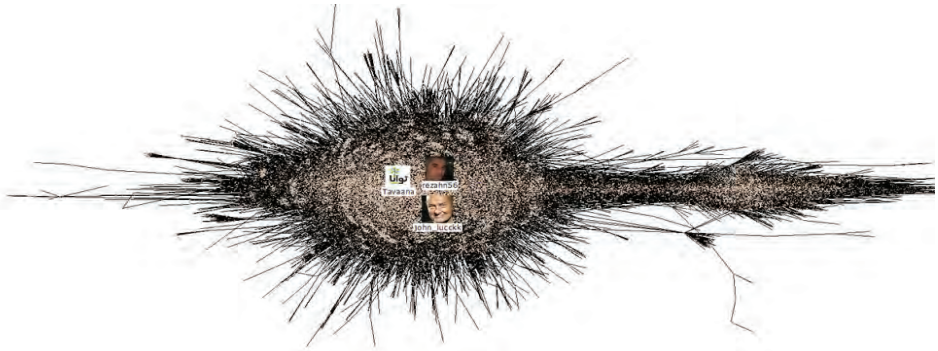
به ابزار و دانش کافی برای جمع‌آوری اطلاعات از شـــبکه‌های اجتماعی مسلط نبوده‌است(اوماند و همکاران، ۲۰۱۲). براین اساس، در حوزه اقدامات پلیس، برای آنکه ساختار مدیریتی پلیس بتواند تأثیر اقداماتش را بر افکار عمومی بسنجد، باید از نظر عموم جامعه آگاه شود. توانایی تحلیل داده‌های شبکه‌های اجتماعی در مقیاس وسیع یک توانمندی است که سازمان‌های بزرگ به‌تناسب مأموریت خود باید از آن بهره‌مند باشن؛ بنابراین، این پژوهش در تلاش است تا الگویی برای تحلیل داده‌های راهبردی در شبکه‌های اجتماعی ارائه نماید.

ادبیات نظری پژوهش

شبکه‌های عصبی مصنوعی[1]

شـــبکه‌های عصبی، نسل جدید روش‌های داده‌کاوی به‌شمار می‌آیند که در دو دهه اخیر، توسعه زیادی یافته‌اند. از این روش‌ها هم می‌توان برای کشف و استخراج دانش از پایگاه داده‌ها و هم برای ایجاد الگوهای پیش‌بینی اســـتفاده کرد. شبکه‌های عصبی به‌دلیل توسعه‌های اخیرشان، به‌عنوان یک ابزار داده‌کاوی بســـیار متداول شده‌اند. این شبکه‌ها ابزارهایی هستند که در موارد پیش‌بینی، رده‌بندی، خوشـــه‌بندی و تخمین کاربرد دارند. همچنین، استفاده‌های صنعتی این ابزار نیز بسیار فراوان است که از آنها می‌توان به تشخیص سلسله‌های زمانی، تشخیص شرایط اجتماعی و تحولات آن، تشخیص خوشـــه‌های افراد باارزش(هر ارزشی مثبت یا منفی) و شناسایی تخلف‌هایی که در بیمه صورت می‌گیرد، اشـــاره نمود. تولید گراف‌هایی با حجم، جهت، گســـتردگی و یا پراکندگی و همچنین، اراده اجتماعی در یک موضوع و یا تشـــریح شـــبکه‌های جرایم و تروریسم یکی دیگر از کاربردهای شـــبکه‌های عصبی اســـت. گراف‌هایی که درنتیجه عملکرد شبکه‌های عصبی تولید می‌شوند را می‌توان یکی از بهترین نمونه‌های آن دانست. از نرم‌افزارهایی که با استفاده از شبکه‌های عصبی از داده‌های ارتباطی، گراف‌های واضح و قابل‌بررسی را تولید می‌کنند، می‌توان نرم‌افزار متن باز گِفی[2] و نرم‌افزار دولتی I۲[3] با دسترسی محدود را نام برد.

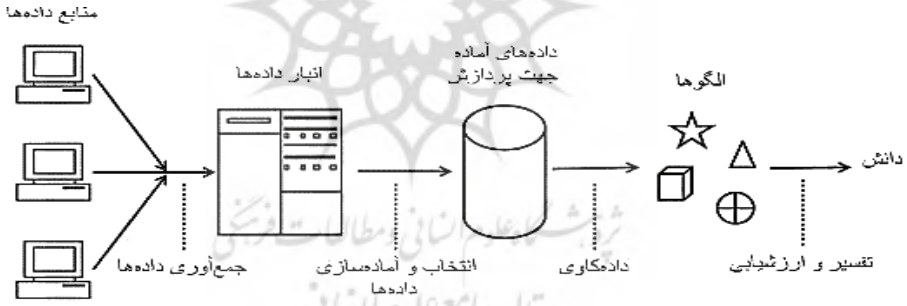
1. Artificial Neural Networks
2. Gephi Network Analysis
3. IBM Intelligence Analysis



شکل ۱. خروجی تحلیل روابط شبکه‌های تروریستی در توییتر با استفاده از نرم‌افزار گِفی

تمایز روش‌های آماری و داده‌کاوی

به‌عنوان یک قانون کلی، فرض روش‌های آماری بر این اساس است که توزیع داده‌ها مشخص است و در بیشـــتر موارد، فرض بر این اســـت که توزیع نرمال است و درنهایت، درستی یا نادرستی نتایج نهایی، به درست بودن فرض اولیه وابسته است(شکل ۲).



شکل ۲. فرآیند داده‌کاوی

در مقابل روش‌های یادگیری، یادگیری ماشین از هیچ فرضی در مورد داده‌ها استفاده نمی‌کند و همین باعث تفاوت‌هایی بین این دو روش می‌شود(مانیلا[1] و همکاران، ۲۰۱۱). در جدول شماره ۱ تفاوت‌های روش آنالیز آماری و داده‌کاوی مشاهده می‌شود.

1. Manila

روش	آنالیز آماری	داده کاوی
فرضیه	آمار شناسان همیشه با یک فرضیه شروع به کار می کنند.	به فرضیه احتیاجی ندارد.
نوع داده ها	آنها از داده های عددی استفاده می کنند.	ابزارهای داده کاوی از انواع مختلف داده، نه تنها عددی می توانند استفاده کنند.
ایجاد روابط	آمارشناسان باید رابطه هایی را ایجاد کنند که به فرضیه آنها مربوط است.	الگوریتمهای داده کاوی به طور اتوماتیک روابط را ایجاد می کنند.
صحت داده ها	آنها می توانند داده های نابجا و نادرست را در طول آنالیز مشخص کنند.	داده کاوی به داده های صحیح و درست نیاز دارد.
قابلیت تفسیر	آنها می توانند نتایج کار خود را تفسیر و برای مدیران بیان کنند.	نتایج داده کاوی نسبتا پیچیده می باشد و نیاز به متخصصانی جهت بیان آنها به مدیران دارد.

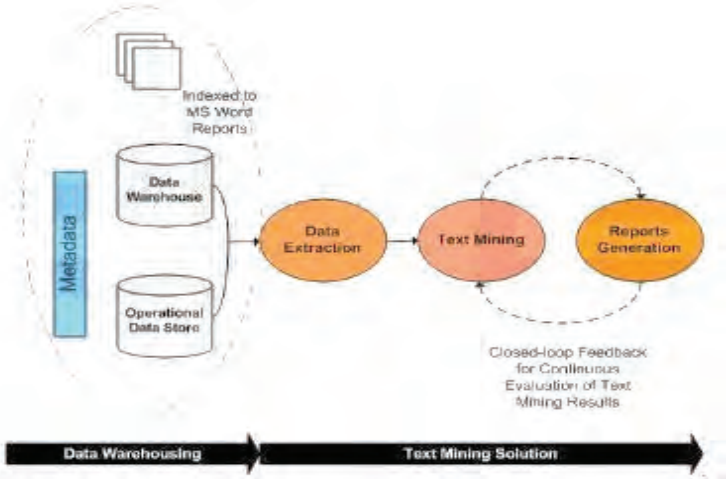
جدول ۱. تفاوت‌های روش آنالیز آماری و داده‌کاوی(حسین‌زاده، ۱۳۹۲)

تحلیل‌محتوا

درخصوص تحلیل‌محتوا تعریف‌های مختلفی وجود دارد؛ اما یکی از تعریف‌های جامع، تعریف برنارد برلسون[1] است که پس از گذشت سال‌ها، هنوز بسیاری از پژوهشگران به آن استناد می‌کنند. برلسون، تحلیل‌محتوا را یک شیوه تحقیقی تعریف کرده‌است که برای تشریح عینی، منظم و کمّیِ محتوای آشکار پیام‌های ارتباطی به کار می‌رود(ضغیمی و همکاران، ۱۳۹۷). به عبارت دیگر، برنارد برلسون تحلیل‌محتوا را روش مطالعه و تجزیه‌وتحلیل ارتباط به شیوه‌های نظام‌مند، عینی، کمّی و آشکار برای اندازه‌گیری متغیرها توصیف می‌کند؛ اما در کل، تحلیل‌محتوا به فرآیند تحلیل‌متن برای استخراج یا کشف اطلاعات و واقعیت‌های معتبر، جدید و از پیش‌ناشناخته، پنهان، مفید و قابل‌درک از داده‌های ساخت‌نیافته و نیمه ساخت‌یافته به‌صورت خودکار گفته می‌شود.

در شکل ۳، معماری مفهومی فرآیند تحلیل‌محتوا شامل مرحله انبار داده، استخراج ویژگی و الگوسازی توصیف شده‌است. در بخش انتهایی(تولید گزارش که توسط کاربر سیستم اعمال می‌شود)، می‌توان ایرادها را بهبود بخشید یا نظرات متخصص حوزه را اعمال کرد.

1. Berelson



شکل ۳. معماری مفهومی تحلیل‌محتوا

شباهت‌ها و تفاوت‌های داده‌کاوی و تحلیل‌محتوا

متن‌کاوی برخی از رهیافت‌های خود را از پژوهش‌های داده‌کاوی به دست آورده‌است؛ بنابراین، شباهت‌های زیادی بین معماری داده‌کاوی و تحلیل‌محتوا وجود دارد. در هر دو روش، رویه‌هایی مبنی‌بر پیش‌پردازش، الگوریتم‌های کشف الگو و مؤلفه‌های لایه بازنمایی وجود دارد(مانند ابزارهای بازنمایی بصری برای بهبود نتیجه‌ها)؛ همچنین، تحلیل‌محتوا بسیاری از انواع الگوها را در عملیات اکتشاف دانش هسته‌ای که ابتدا در تحقیقات داده‌کاوی معرفی شده‌اند، اتخاذ می‌کند. در تفاوت داده‌کاوی و تحلیل‌محتوا نیز می‌توان مطرح کرد که داده‌کاوی شامل روش‌هایی برای اکتشافی از انواع داده‌ها مانند داده‌های عددی، چندرسانه‌ای و... است(شیخ و شامبیاتی، ۱۳۹۴).

این امکان وجود دارد که تمام انواع داده با استفاده از الگوریتم‌های عمومی داده‌کاوی پردازش شوند و تحلیل‌محتوا به‌عنوان زیرمجموعه‌ای از داده‌کاوی عمومی تلقی شود. تفاوت اصلی داده‌کاوی و تحلیل‌محتوا، در نوع داده و الگوریتم‌های مربوط[1] به آنهاست. همان‌طورکه درجدول شماره ۲ مشاهده می‌شود، داده‌کاوی بر داده‌های ساختاریافته تمرکز دارد؛ این درحالی است که در تحلیل‌محتوا بر روی اطلاعات غیرساخت‌یافته یا نیمه‌ساخت‌یافته پردازش صورت می‌گیرد.

1. Barbour & Ploug

	داده‌کاوی	متن‌کاوی
موضوع موردبررسی	داده‌های عددی (ساختاریافته)	متون (غیر ساخت‌یافته و نیمه ساخت‌یافته)
ساختار موضوع	بانک‌های اطلاعاتی رابطه‌ای	متون فاقد شکل
هدف	توصیف، پیش‌بینی و...	بازیابی اطلاعات، دسته‌بندی محتوا، مقایسه متون و...
روش‌ها	یادگیری ماشین، درخت تصمیم‌گیری، شبکه‌های عصبی، رگرسیون و...	الگوریتم‌های خوشه‌بندی، طبقه‌بندی، زبان‌شناسی، هستان‌شناسی و...
دوره توسعه	پیاده‌سازی وسیع از سال ۱۹۹۴	پیاده‌سازی وسیع از سال ۲۰۰۰

جدول ۲. تفاوت تحلیل‌محتوا و داده‌کاوی

روش‌شناسی پژوهش

قلمرو مکانی این پژوهش جهت ارائه الگوی تحلیل داده‌های راهبردی شبکه‌های اجتماعی، توییتر بوده و جامعه آن، کاربران ایرانی توییتر است. روش گردآوری، پیمایش شبکه اجتماعی بوده و سعی شده‌است تا با استفاده از ابزارهای موجود، نسبت‌به دریافت توییت‌های کاربران اقدام شود. در این تحقیق، از طریق طراحی یک برنامه کاربردی سعی شده‌است تا داده‌های هدف از شبکه اجتماعی توییتر دریافت شود؛ بدین‌صورت که داده‌های دریافت‌شده پس از خلاصه‌سازی، مضمون‌بندی(کد) شده و پژوهشگر با مرور متون، معانی را کدبندی و استخراج کرده و پس از آن نسبت‌به کمی‌سازی اقدام شد. در مرحله بعد، شمارگان کلمات به‌دست آمد و با رویکردی جدید گرافی از شبکه کلمات حاصل شد و در کنار آن، به تحلیل محتوای کیفی پرداخته شد. تحلیل محتوای کیفی را می‌توان روش تحقیقی برای تفسیر ذهنی محتوای داده‌های متنی از طریق فرآیندهای طبقه‌بندی نظام‌مند، کدبندی و تم‌سازی یا طراحی الگوهای شناخته‌شده دانست(هسیه و شانون[1]، ۲۰۰۵). تحلیل‌محتوا نه‌تنها به خلاصه‌سازی متن اصلی کمک می‌کند؛ بلکه نگرش‌ها و ادراک صاحب پیام را نیز منتقل می‌نماید(کریستوفر[2] و همکاران، ۲۰۱۳).

1. Hsieh & Shannon
2. Christopher

چارچوب عملی و کارکرد پژوهش

در این پژوهش، نشان داده می‌شود که فرآیند پیاده‌سازی تحلیل‌محتواهای راهبردی چگونه است. در ابتدا، با استفاده از سرویس‌های نرم‌افزاری منابع باز[1] و با استفاده از API توییتر[2](ای.پی.ای که به برنامه‌های دیگر اجازه استفاده از داده‌های خود را می‌دهد) شبکه محتوایی هدف ـ که توییتر است ـ بررسی شده و با تعریف کلیدواژه‌های مرتبط با جستجو[3] تعیین محل ذخیره‌سازی[4] جمع‌آوری داده‌های مرتبط آغاز می‌شود.

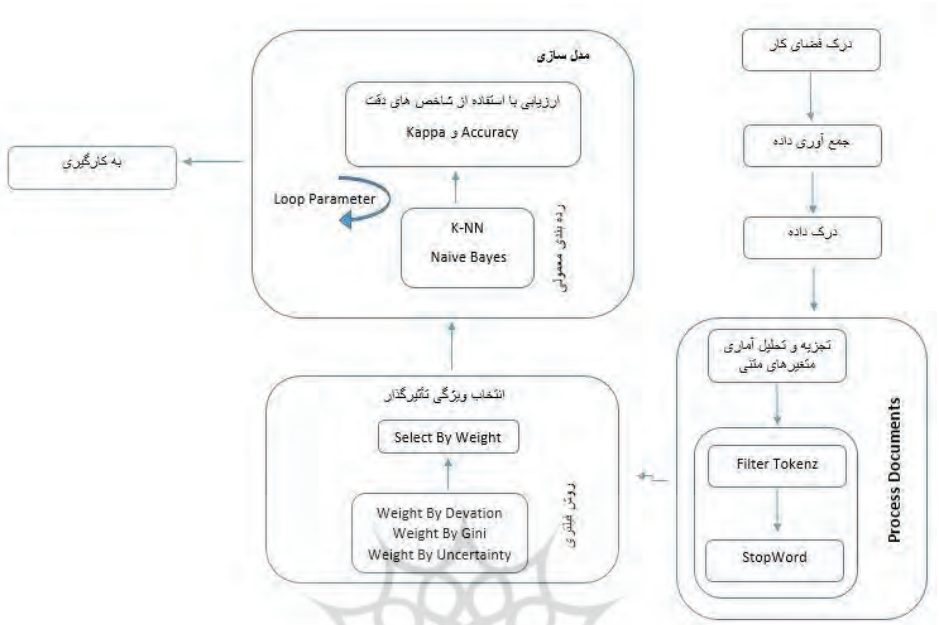
۱. به‌کارگیری الگو

دانش حاصل‌شده به سازماندهی نیاز دارد و باید به‌شکلی ارائه شود که بهره‌وران بتوانند از آن استفاده کنند. بسته به ملزومات کار، فاز به‌کارگیری می‌تواند به‌سادگی ایجاد یک گزارش یا به پیچیدگی اجرای یک فرآیند قابل‌تکرار داده‌کاوی باشد. با توجه به اینکه این پژوهش، یک پژوهش کاربردی است، نتایج می‌تواند به‌صورت یک سیستم سایه‌ای، به‌موازات کار کارشناسان کاربرد داشته باشد؛ به این معنا که نتایج طبقه‌بندی و ارزیابی سیستم داده‌کاوی، بر ارزیابی کارشناسان تصمیم‌ساز در سازمان تأثیرگذار باشد.

۲. درک مسئله کاری

این مرحله، ابتدا بر درک اهداف و ملزومات طرح از دیدگاه سازمان متمرکز می‌شود. هدف داده‌کاوی در این پژوهش، "پیش‌بینی‌کننده" و نوع آن، "طبقه‌بندی" است؛ بدین‌گونه که با توجه به بررسی محتوای کاربران شبکه‌های اجتماعی، نظرات آنها را در طیف‌های مختلف جمع‌آوری، تحلیل و دسته‌بندی می‌نماید.

1. Open source
2. Application Programming Interface
3. Keyword
4. Storage



شکل ۴. الگوی کلی فرآیند تحلیل داده‌های شبکه‌های اجتماعی

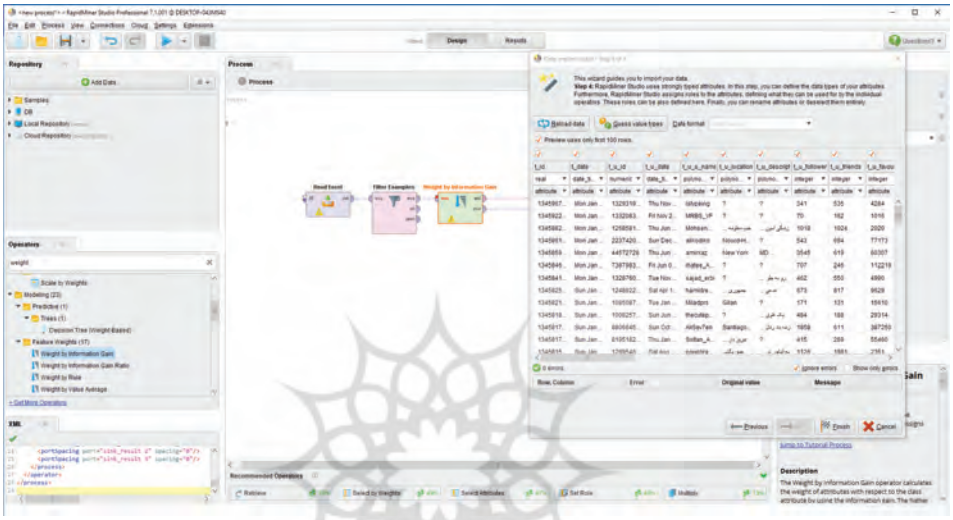
۳. درک داده‌ها

این مرحله با جمع‌آوری اولیه داده‌ها شروع می‌شود و به توصیف داده‌ها و تعیین کیفیت آنها می‌پردازد. متن(کامنت)ها به‌صورت انسانی کدبندی شده و هر کاربر با توجه به دیدگاهش برچسب(لیبل) زده می‌شد. در این پژوهش، دسته‌بندی‌ها به‌صورت زیر تعریف شدند:

* فکاهی؛
* مطالبه‌ای؛
* خبری؛
* انتقادی؛
* اعتراضی؛
* تمجید؛
* نامرتبط.

پس از دریافت داده‌ها با کلیدواژه موردنظر(که در این نمونه، پلیس فتا مدّ نظر قرار گرفته و تمامی متن‌هایی که پلیس فتا در آن درج شده‌است، جمع‌آوری شده‌است) کاربران و متن‌های

آنها در دسته‌بندی‌های فوق تقسیم می‌شوند. این مرحله، درواقع، "آماده‌سازی داده" است. مرحله آماده‌سازی داده‌ها، شامل کلیه فعالیت‌هایی است که برای ساختن مجموعه داده‌های نهایی(داده‌هایی که برای الگوسازی آماده شده‌اند) از داده‌های خام اولیه به کار می‌رود. هرچه کیفیت این آماده‌سازی بهتر باشد، الگوسازی نیز بهتر خواهد بود(مجبی‌منش، ۱۳۹۰).



شکل ۵. ورود داده‌ها به نرم‌افزار رپید ماینر

وظیفه آماده‌سازی داده‌ها در چند دوره انجام می‌گیرد و هیچ ترتیب از پیش تعریف‌شده‌ای ندارد. این وظیفه شامل انتخاب جداول[۱]، رکوردها[۲]، خصیصه‌ها[۳] و نیز انتقال[۴] و پاک‌سازی داده[۵] برای الگوسازی است. در این مرحله، داده‌هایی که در مرحله قبل دریافت شدند، استخراج و ثبت شدند. سپس داده‌ها در یک پایگاه داده جامع و یکپارچه(البته پالایش‌نشده) به‌منزله یک پایگاه داده رابطه‌ای قرار گرفتند. سپس داده‌ها پالایش شده و ساختار موردنظر برای الگوسازی روی آنها اعمال شد. برای پاک‌سازی و پیش‌پردازش داده‌ها، دو عملیات کاهش داده[۶] و اعمال تغییرات[۷] در شکل داده‌ها، روی پایگاه داده رابطه‌ای[۸] صورت گرفت.

1. Tables
2. Records
3. Attributes
4. Transferring
5. Data Cleaning
6. Reduction
7. Change assigning
8. Relational Database

۴. الگوسازی

در این مرحله، انواع روش‌های الگوسازی انتخاب شده و به کار گرفته می‌شوند. درمجموع، برای یک نوع مسئله داده‌کاوی، چندین روش وجود دارد. برخی از روش‌ها نیازمند شکل ویژه‌ای از داده‌ها هستند؛ بنابراین، به‌منظور آماده‌سازی داده برای اعمال روش‌ها در پژوهش حاضر، با استفاده از نرم‌افزار رپیدماینر[1] الگوسازی داده‌ها، به‌طورجداگانه روی هر الگوریتم طبقه‌بندی انجام گرفت و الگوریتمی که بالاترین صحت را داشت، مبنای الگوسازی و استخراج دانش از آن قرار گرفت. هدف از این کار، استخراج دانش با توجه به داده‌های آموزشی از الگوریتم انتخابی است.

۵. ارزیابی نتایج

در این مرحله از طرح، الگویی که از دیدگاه تحلیل‌داده، کیفیت بالایی دارد، ساخته شده‌است. پیش از اقدام برای به‌کارگیری الگو، باید الگو به‌طورکلی ارزیابی شده[2] و گام‌های اجرایی برای تطابق با اهداف سازمانی مرور شوند. در اینجا، با توجه به مقایسه‌های صورت‌گرفته، روشی که از بقیه پیش‌بینی دقیق‌تری را انجام می‌دهد، برای استفاده بهره‌وران ارائه می‌شود.

۶. ورود داده و پیش‌پردازش

داده‌های آماده‌شده در نرم‌افزار رپیدماینر وارد شده و در مراحل ابتدایی، تعیین می‌شود که کدام ویژگی‌های داده از اهداف داده‌پردازی موردنظر محسوب می‌شوند. انتخاب ویژگی را می‌توان به‌عنوان فرآیند شناسایی ویژگی‌های مرتبط و حذف ویژگی‌های غیرمرتبط و تکراری با هدف مشاهده زیرمجموعه‌ای از ویژگی‌ها که مسئله را به‌خوبی و با حداقل کاهش درجه کارایی تشریح می‌کنند، تعریف کرد. این کار مزایای گوناگونی دارد؛ از جمله:

الف. بهبود کارایی الگوریتم‌های یادگیری ماشین؛

ب. درک داده، کسب دانش درباره فرآیند و کمک به بصری‌سازی آن؛

ج . کاهش داده کلی، محدود کردن نیازمندی‌ها ذخیره‌سازی و احتمالا کمک به کاهش هزینه‌ها؛

د. کاهش مجموعه ویژگی‌ها، ذخیره‌سازی منابع در دور بعدی گردآوری داده یا در طول بهره‌برداری؛

هـ . سادگی و قابلیت استفاده از الگوهای ساده‌تر و کسب سرعت.

در کنار همه دلایل پیشین، در سناریوهای "تحلیل کلان‌داده"، انتخاب ویژگی نقشی اساسی ایفا می‌کند. با توجه به حجم داده‌های سه‌ساله، برای بالابردن دقت الگوها، از انواع الگوریتم‌های کاهش

1. RapidMiner
2. Evaluation

داده[1] که یک ویژگی تکنیکی است، برای کاهش داده‌ها در فرآیند داده‌کاوی استفاده می‌شود. در کاهش داده‌ها حجم داده‌ها را کاهش می‌دهند تا بتوان از آنها برای تجزیه‌وتحلیل کارآمدتر استفاده کرد؛ چراکه مجموعه داده ممکن است دارای تعداد زیادی ویژگی باشد(felici & others,2006).

۷. یادگیری نظارت‌شده

در یادگیری نظارت‌شده[2] کار با ایمپورت‌کردن مجموعه داده‌های شامل ویژگی‌های خصیصه‌های آموزش[3] خصیصه‌های هدف[4] آغاز می‌شود. الگوریتم یادگیری نظارت‌شده رابطه بین مثال‌های آموزش و متغیرهای هدف مختص آنها را به دست می‌آورد و آن رابطه یادگرفته‌شده را برای دسته‌بندی ورودی‌های کاملا جدید(بدون هدف‌ها) مورد استفاده قرار می‌دهد. برای نمایش اینکه یادگیری نظارت‌شده چگونه کار می‌کند، یک مثال از پیش‌بینی نمرات دانش‌آموزان بر پایه ساعات مطالعه آنها ارائه می‌شود:

از منظر ریاضی، $Y = f(X)+ C$:

که در آن:

- F رابطه بین نمرات و تعداد ساعاتی است که مدیران راهبردی به‌منظور آماده‌شدن برای امتحانات به مطالعه می‌پردازند؛
- X ورودی است(تعداد ساعاتی که دانش‌آموز خود را آماده می‌کند)؛
- Y خروجی است(نمراتی که دانش‌آموزان در آزمون کسب کرده‌اند)؛
- C یک خطای تصادفی است.

هدف نهایی یادگیری نظارت‌شده، پیش‌بینی Y با حداکثر دقت برای ورودی جدید داده‌شده X است. چندین راه برای پیاده‌سازی یادگیری نظارت‌شده وجود دارد که برخی از متداول‌ترین رویکردها در ادامه موردبررسی قرار می‌گیرند.

بر پایه مجموعه داده موجود، مسئله یادگیری ماشین در دو نوع دسته‌بندی[5] و رگرسیون[6] قرار می‌گیرد. اگر داده‌های موجود دارای مقادیر ورودی(آموزش) و خروجی(هدف) باشند، مسئله از نوع دسته‌بندی است. اگر مجموعه داده دارای مقادیر عددی پیوسته[7] بدون هرگونه برچسب هدفی

1. Feature subset selection
2. Supervised Training
3. Training attributes
4. Target attributes
5. Classification
6. Regression
7. Continuous numerical values

باشد، مسئله از نوع رگرسیون محسوب می‌شود.

۸. دسته‌بندی

در این پژوهش، داده‌های توییتر به‌منظور پیش‌بینی اینکه کاربران درخصوص اقدامات پلیس فتا چه گفته‌اند، مدّ نظر قرار گرفته‌است؛ با این هدف که نتیجه این تحقیق در دسته‌بندی می‌تواند برای مدیران راهبردی کارگشا باشد. دسته‌بندی[1] یک مسئله پیش‌بینی[2] است که برچسب‌های[3] کلاس دسته‌ای را که گسسته یا بدون‌ترتیب هستند، پیش‌بینی می‌کند. این یک فرآیند دو مرحله‌ای است که شامل مرحله یادگیری و دسته‌بندی می‌شود.

روش‌های دسته‌بندی و انتخاب بهترین آنها

الگوریتم‌های متداول داده‌کاوی برای تحلیل‌داده استفاده می‌شود و بسته به نوع داده‌ها و اهداف مدّ نظر و بر اساس منابع علمی و تجربی دانشمندان علم داده[4]، می‌توان الگوریتم موردنظر را در نرم‌افزار به کار گرفت تا به نتایج دقیق‌تری حاصل شود. در اینجا، به برخی از این الگوریتم‌ها و ویژگی‌های آنها پرداخته می‌شود.

الف. نزدیک‌ترین همسایگی(KNN[5]):

روشی از دسته‌بندی است و با وجود سادگی، عملکرد خوبی را برای مجموعه‌های آموزشی بزرگ ارائه می‌دهد که بر اساســی‌ترین فرض نهفته در پیش‌بینی‌ها تکیه می‌کند؛ اینکه مشــاهدات با ویژگی‌های مشابه نتایج مشابهی خواهند داشت. روش KNN یک مقدار پیش‌بینی‌شده را برای یک مشاهده جدید بر اساس تعدد یا میانگین در مجموعه آموزشی[6] تعیین می‌کند؛ به این معنا که با مقداری از داده با الگوریتم KNN به سیستم یاد می‌دهیم که داده‌های دسته‌بندی‌نشده جدید را با توجه به خصیصه‌های آنها(همسایگی) در کنار هم دسته‌بندی نماید. با توجه به حجم نامحدودی از داده‌ها، هر مشــاهده‌ای دارای "همسایه"های زیادی خواهد بود که با توجه به همه ویژگی‌های اندازه‌گیری‌شده به‌طور خودکار نزدیک هستند(Joshua S. Richman,2013).

ب. درخت تصمیم(Decision Tree)

منظور از "درخت تصمیم" الگوهای پشتیبانی تصمیم هستند که الگوها را با استفاده از یک سلسله

1. Classification
2. Prediction
3. Label
4. Data Science
5. K- Nearest Neighborhood
6. Training dataset

قوانین کاملاً مشـــخص[1] طبقه‌بندی می‌کنند. آنها نمودارهای درختی[2] هستند که در آنها هر گره شـــاخه[3] یک دسته بین تعدادی از دسته‌ها را نشان می‌دهد و هر گره برگ[4] نشان‌دهنده نتیجه‌ای از انتخاب‌های این دسته‌ها است(Ranganathan,2013).

ج. نایو بیز(Naïve Bayes)

در مرحله یادگیری، الگوی دسته‌بندی، دسته‌بند را با تحلیل مجموعه داده آموزش می‌سازد. در مرحله دسته‌بندی، برچسب‌های کلاس برای داده‌های موجود پیش‌بینی می‌شوند. تاپل‌های مجموعه داده و برچســـب‌های کلاس مربوط درحال‌تحلیل، به دو دسته "مجموعه آموزش"[5]و "مجموعه آزمون" تقسیم می‌شوند. مجموعه داده‌آزمون به‌منظور تخمین پیش‌بینی دسته‌بندی استفاده می‌شود. دقت دســـته‌بندی درصدی از آزمون است که توسط دسته‌بند به درستی دسته‌بندی شده‌اند. به‌منظور کسب دقت بیشتر، بهترین راه آزمون الگوریتم‌های گوناگون و آزمودن شاخص‌های مختلف در هر الگوریتم است. بهترین حالت با اعتبارسنجی متقابل[6] قابل‌انتخاب است. برای انتخاب یک الگوریتم خوب برای مسئله، شاخص‌هایی مانند دقت، زمان آموزش، خطی‌بودن[7] تعداد شاخص‌ها و شرایط خاص باید برای الگوریتم‌های متفاوت درنظر گرفته شود.

طرح عملی تحلیل داده‌های راهبردی

در این تحقیق، جملات انتشاریافته توسط کاربران و نظرات آنها نسبت‌به پلیس فتا دسته‌بندی شد. داده‌ها پس از متن‌کاوی به الگویی برای کل داده‌های سه‌ساله تبدیل شد که دقت‌های به‌دست‌آمده با سه الگوریتم پیش‌گفته، در جدول‌های زیر قابل‌مشاهده است.

1. Well-Defined
2. Tree-like graphs
3. Branch node
4. leaf node
5. Training set
6. Cross-validation
7. Linearity

دقت با الگوریتم k نزدیک‌ترین همسایگی

accuracy: 92.41% +/- 1.60% (mikro: 92.41%)

	true فکاهی	true مطالبه	true خبری	true نا مرتبط	true انتقادی	true اعتراضی	true تمجید	class precision
pred. فکاهی	159	62	19	54	2	22	14	47.89%
pred. مطالبه	0	700	6	3	0	0	0	98.73%
pred. خبری	0	4	646	1	0	0	3	98.78%
pred. نا مرتبط	4	9	5	402	0	0	0	95.71%
pred. انتقادی	0	0	0	1	14	4	0	73.68%
pred. اعتراضی	1	2	0	2	6	599	0	98.20%
pred. تمجید	3	0	4	0	0	0	293	97.67%
class recall	95.21%	90.09%	95.00%	86.83%	63.64%	95.84%	94.52%	

دقت با الگوریتم درخت تصمیم

accuracy: 38.04% +/- 10.15% (mikro: 38.04%)

	true فکاهی	true مطالبه	true خبری	true نا مرتبط	true انتقادی	true اعتراضی	true تمجید	class precision
pred. فکاهی	75	0	0	2	0	0	24	74.26%
pred. مطالبه	92	777	676	461	12	348	266	29.52%
pred. خبری	0	0	3	0	0	0	0	100.00%
pred. نا مرتبط	0	0	0	0	0	0	0	0.00%
pred. انتقادی	0	0	0	0	6	0	0	100.00%
pred. اعتراضی	0	0	0	0	4	277	0	98.58%
pred. تمجید	0	0	1	0	0	0	20	95.24%
class recall	44.91%	100.00%	0.44%	0.00%	27.27%	44.32%	6.45%	

دقت با الگوریتم نیوبیز

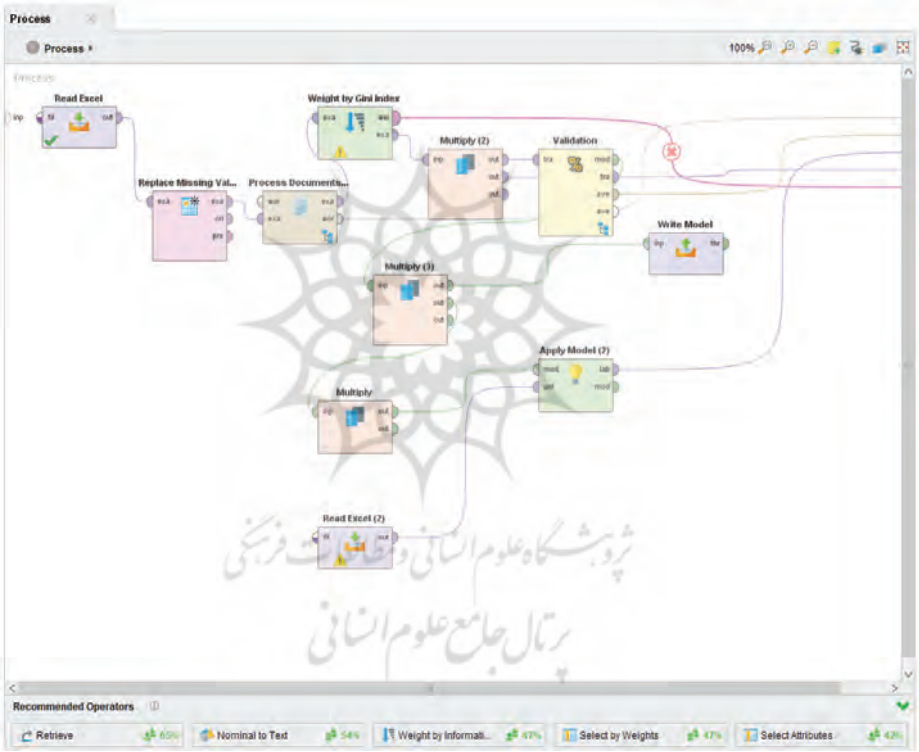
accuracy: 93.23% +/- 1.13% (mikro: 93.23%)

	true فکاهی	true مطالبه	true خبری	true نا مرتبط	true انتقادی	true اعتراضی	true تمجید	class precision
pred. فکاهی	155	3	0	5	0	0	6	91.72%
pred. مطالبه	7	733	10	20	0	9	4	93.61%
pred. خبری	1	12	656	9	2	4	2	95.63%
pred. نا مرتبط	0	16	2	411	0	1	4	94.70%
pred. انتقادی	2	0	0	7	20	42	0	28.17%
pred. اعتراضی	1	13	1	7	0	569	0	96.28%
pred. تمجید	1	0	11	4	0	0	294	94.84%
class recall	92.81%	94.34%	96.47%	88.77%	90.91%	91.04%	94.84%	

جدول ۳. دقت‌های به‌دست‌آمده در تحلیل داده‌های توییتر با کلیدواژه فتا

نوع الگوریتم	دقت
Naive Bayes	۹۳.۲
K-nn	۹۲.۴
Decision Tree	۳۸.۰۴

دســـته‌بندی توییت‌ها به‌معنای برچســـب‌زدن یک مفهوم به هر توییت و جای‌گذاری توییت‌های مشـــابه در یک دسته مشابه اســـت. به‌عبارتی، هر توییتی با توجه به دارابودن کلمات و واحدهای معنایی مشابه مستتر در آن، به یک دسته خاص اختصاص داده می‌شود. تمام توییت‌های مربوط به پلیس فتا، پس از خوانش اولیه در پنج دسته و مقوله کلی شامل خبری، انتقادی، اعتراضی، مطالبات و فکاهی، بر روی تعداد ۳۹۹۴۵۶ توییت الگوهای یادگیری ماشین با سه الگوریتم پیش‌گفته[1] برابر شـــکل زیر در نرم‌افزار رپیدماینر محاســـبه گردید. در گام بعدی، به کمک دو اپراتور log و loop parameter دقت الگو با تغییر شاخص‌های مختلف بررسی گردید.



سلسله‌زمانی در داده‌کاوی

ارزش و اهمیت مجموعه داده‌ها و انواع مختلف داده، به‌دلیل کاربردهای فراوان بر کســـی پوشیده نیســـت. یکی از مهم‌ترین و پرکاربردترین انواع داده، داده‌های سلسله‌زمانی[2] اســـت. مجموعه‌های

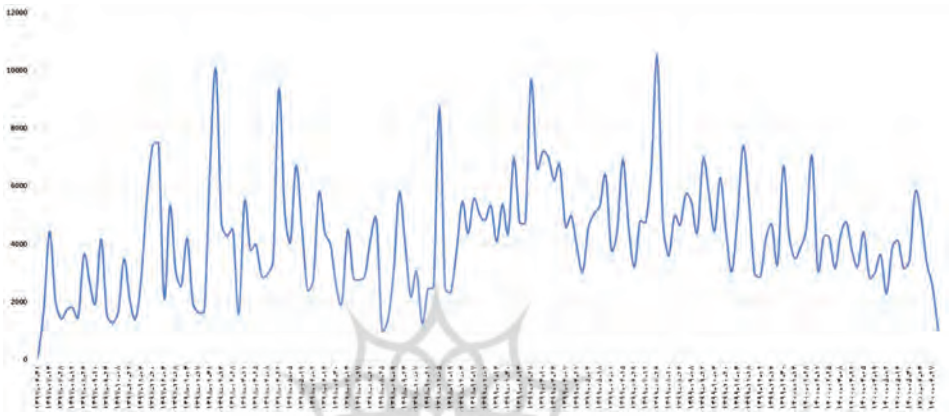
1. Knn,Naive bayse, Decision Tree
2. Time Series

داده‌ای وجـود دارد که در آنها ویژگی هدف وابسـته به زمان اسـت؛ زیرا بـا دنباله‌ای از دوره‌های متوالی در طول بعد زمان مرتبط اسـت. در چنین شرایطی، گفته می‌شود که مقادیر متغیر هدف نشـان‌دهنده یک سلسه‌زمانی اسـت. از این نوع داده، برای پیش‌بینی اسـتفاده می‌شود. درواقع، سلسـله‌زمانی به کشـف الگوی رفتار داده‌ها در گذشـته و پیش‌بینی رفتار آینده از روی این نوع داده اطلاق می‌شود(Brockwell,۲۰۱۰). این داده‌ها در داده‌کاوی برای پیش‌بینی، بسیار با اهمیت هستند. یک سلسله‌زمانی مجموعه‌ای از مشاهدات درباره یک متغیر است که در نقاط گسسته‌ای از زمان که معمولا فاصله‌های مسـاوی دارند، اندازه‌گیری شـده و بر حسب زمان، مرتب شده‌اند؛ بنابراین، یک سلسله‌زمانی از مشاهده یک پدیده در طول زمان به‌دست می‌آید. سلسله‌زمانی یک کلاس مهم از داده‌های زمانی است و مجموعه مشاهداتی است که به‌ترتیب زمان انجام شده‌است. داده‌های سلسله‌زمانی دارای ویژگی‌هایی مثل حجم بالای داده، ابعاد بالا و به‌روزرسانی مداوم است؛ علاوه‌بر این، داده‌های سلسله‌زمانی با ماهیت عددی و پیوسته مشخص می‌شوند و به‌عنوان یک کل درنظرگرفته می‌شود. افزایش استفاده از داده‌های سلسله‌زمانی، تلاش‌های تحقیق و توسعه زیادی را در زمینه داده‌کاوی آغاز کرده‌است. تولید حجم بالای داده‌ها در طول زمان، برای داده‌کاوی بسیار اهمیت دارد. اسـتخراج الگوها و اطلاعات مفید از دل این حجم اطلاعات، اهمیت سلسله‌زمانی را نشـان می‌دهد(Rojas,2019). اگر یک بردار را درنظر بگیریم و این بردار نشان‌دهنده زمان باشد، می‌توان سلسه‌زمانی را این‌گونه تعریف نمود:

$$X(t), \qquad t = 0,1,2,3, \dots$$

بنابر معادله فوق، زمان می‌تواند زمان شروع یک پدیده یا زمان ثبت اولین اطلاعات مربوط‌به حوزه پژوهش باشد؛ بنابراین، یک متغیر تصادفی است و مقدار را در لحظه‌ای نشان می‌دهد.

تحلیل سلسله‌های‌زمانی یک ابزار راهبردی است Y با این تفسیر که تحلیل داده‌ها کمک می‌کنند تا تصمیم‌گیران اقداماتی را که از منظر کاربران قابل‌توجه بوده و یا دارای اهمیت فراوان بوده‌است، درک کنند. این نوع فهم از داده‌ها کمک می‌کند تا خطاهای سامانه و سازمان کاهش یابد و اقدامات خود را با اطلاع‌رسانی قبلی و پیوست رسانه‌ای، از گزند شایعات و هجوم اخبار منفی مصون دارد.



شکل ۵. سلسله‌زمانی توجه کاربران به پلیس فتا در توییتر

در شکل بالا، نقاط اوج، نشان‌دهنده اقدام فتا در موضوعات مختلف و نقاط نزول، زمان‌های عدم‌توجه به فتا بوده‌است. این نمودار زمانی کمک می‌کند تا برنامه‌ریزی رسانه‌ای به‌گونه‌ای صورت پذیرد که کاربران هدف ـ که مردم جامعه هستند ـ مدام در معرض داده‌های منتشره از فتا باشند. مدیران می‌توانند با ترکیب این زمان‌ها با وقایع دیگر، مفاهیم پنهان را کشف و اقدامات موثر را تقویت و یا اقدامات مضر هویت اجتماعی پلیس را کاهش دهند.

ابر کلمات

ابر کلمات[1] یک روش تجسم است که مجموعه‌ای از برچسب‌های مربوط‌به یک مورد خاص را خلاصه می‌کند. در ابر کلمات، متخصصین تحلیل‌داده می‌توانند به‌صورت بصری از رشد یک کلیدواژه و هم‌جواری آن با سایر کلمات یک پدیده را تحلیل کنند، بیشترین کلمات تکرارشده درشت‌تر به‌نظر می‌رسند و کلمات کنار هم می‌توانند معنای خاصی داشته باشند (Rokne & Alhajj, 2014).

1. Tag Cloud

شـــکل ۵. ابر کلمات حاصل از تحلیل تکرار متن‌های تولیدشـــده در توییت کاربران درخصوص پلیس فتا

نتیجه‌گیری

تحلیل داده‌های راهبردی نقشـــی موثر را در خدمات ســـازمانی و شکل‌گیری هویت سازمانی ایفا می‌کند. بسته به هدف تحلیل، داده‌ها به‌خاطر دارا بودن اصالت می‌توانند تصمیم‌گیری‌ها را به‌سمت عاری از اشـــتباه ســـوق دهند؛ به همین منظور، لازم است که مدیران راهبردی درخصوص تحلیل داده‌ها و نقش آن در پیشبرد اهداف سازمانی از درک درستی برخوردار باشند. پیاده‌سازی روش‌های جمع‌آوری و تحلیل داده‌ها برای تمامی ابعاد سازمانی امری الزامی است که بی‌توجهی به آن موجب خدشه‌دار شدن هویت و اقدامات مفید سازمان خواهد شد.

درحقیقت، مهم‌ترین موضعی که می‌تواند یک سازمان را از لحاظ هویتی(محبوبیت و یا تنفر) نشان دهد، موضوع جهت‌گیری افکار عمومی نسبت‌به اقدامات آن است. در این پژوهش، به‌منظور ارائه الگوی تحلیل داده‌های راهبردی شبکه‌های اجتماعی، الگوی جمع‌آوری و پردازش داده‌ها ترسیم گردید و به‌منظور تبدیل الگوی انتزاعی ذهنی به الگوی عملی ملموس، با استفاده از تحلیل‌محتوا و متن‌کاوی به بررسی دیدگاه کاربران ایرانی در توییتر نسبت‌به یکی از پلیس‌های فرماندهی انتظامی

جمهوری‌اسلامی‌ایران پرداخته شد. در این مسیر، طی تشریح موضوع، به ساختار داده‌کاوی و تحلیل‌متن اشاره شد تا تمایز و تشابه الگوهای تحلیلی درک شود.

فهرست منابع

منابع فارسی

- ضیغمی و همکاران(۱۳۹۷)، "تحلیل محتوا"، نشریه پرستاری ایران، سال بیست‌ویکم، شماره۵۳، صص ۷۶ ـ ۸۸
- کریپندورف، کلاوس(۱۳۸۳)، تحلیل محتوا: مبانی روش‌شناسی، ترجمه هوشنگ نایبی، تهران، همزه
- ادوارد والتز(۱۳۸۶)، عملیات و اصول جنگ اطلاعات، ترجمه غلامعلی جانگداز، تهران، انتشارات دانشکده امام محمد باقر(ع)
- محبی منش، امید(۱۳۹۰)، "بررسی تأثیر مدیریت کیفیت زنجیره تامین(SCQM) بر رضایت مشتریان با استفاده از مدل‌سازی معادلات ساختاری"، پایان‌نامه کارشناسی ارشد، تهران، دانشکده مدیریت دانشگاه تهران

منابع انگلیسی

- Chianese, A. & Piccialli, F. , 2016. International workshop on Data Mining of Iot Systems(DaMIS): A service oriented framework for analysing social network activities
- Christopher, A. A. , & alias Balamurugan, S. A. (2013, March). Data mining approaches for aircraft accidents prediction: An empirical study on Turkey Airline In Emerging Trends in Computing, Communication and Nanotechnology(ICE-CCN), 2013 International Conference on(pp. 739-745). IEEE
- Kaplan, A. M. , & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. Business horizons, 53(1), 59-68
- Joshua S. Richman, , 2011, Pages 397-408, Multivariate Neighborhood Sample Entropy: Methods in Enzymology,Volume 487 A Method for Data Reduction and Prediction of Complex Data

- Joo Chuan Tong, Shoba Ranganathan, Computational T cell vaccine design in Computer-Aided Vaccine Design, 2013
- Triantaphyllou, E. and G. Felici(Eds.), Data Mining and Knowledge Discovery Approaches based on Rule Induction Techniques, Massive Computing Series, Springer, Heidelberg, Germany, pp. 227–252, 2005
- P. J. Brockwell, Time Series Analysis ,International Encyclopedia of Education(Third Edition), 2010
- Rojas, F. , Herrera, L. J. , Pomares, H. , Rojas, I. (Eds.) Theory and -Applications of Time Series Analysis,Springer, Selected Contributions from ITISE 2019Valenzuela, 2019
- Reda Alhajj, Jon Rokne , Tag Clouds In book: Encyclopedia of Social Network Analysis and Mining Edition: 1,Springer New York,2014