فصلنامهٔ پژوهش‌های اطلاعاتی و جنایی، سال پانزدهم
شمارهٔ اول، بهار ۱۳۹۹

پیش‌بینی تجمعات غیرقانونی مبتنی بر داده‌کاوی رویدادهای امنیتی

نوع مقاله: پژوهشی

تاریخ دریافت: ۹۸/۰۹/۲۰ تاریخ پذیرش: ۹۸/۱۲/۱۰

از صفحهٔ ۹۷ تا ۱۱۸

سعید بختیاری[۱]، حمیدرضا قاسم‌زاده[۲]

چکیده

زمینه و هدف: امروزه قابلیت‌های شبکه‌های اجتماعی، امکان انتشار فراخوان‌های متنوع را فراهم آورده است که برخی از این فراخوان‌ها منجر به تجمعات غیرقانونی و اغتشاش می‌شود.. این پژوهش، با استفاده از فناوری داده‌کاوی رویدادهای امنیتی، با هدف تحلیل و بررسی تجمعات غیرقانونی و رابطهٔ آن با فراخوان‌ها و رویدادهای امنیتی انتشاریافته در شبکه‌های اجتماعی در جهت پیش‌بینی تجمعات غیرقانونی انجام یافته است.

روش: تحقیق حاضر به روش کمی و بر روی داده‌های توصیفی انجام شده است. جامعهٔ آماری مشتمل بر بانک اطلاعاتی فراخوان‌های سال‌های ۹۵ و ۹۶ بوده است و براساس این بانک، بیش از ۴۰۰۰ رویداد و فراخوان در ۱۶ استان و ۱۷۸ شهر کشور ثبت شده که با استفاده از نرم‌افزارهای داده‌کاوی، مدل‌های مختلفی از روند تشکیل تجمعات غیرقانونی مورد بررسی قرار گرفت.

یافته‌ها: از بین الگوریتم‌های مختلف پیش‌بینانهٔ داده‌کاوی به‌کار رفته در این تحقیق، الگوریتم C5 با دقت ۹۱/۷۲ درصد، بهترین نتیجه را برای پیش‌بینی تجمعات غیرقانونی ارائه داد که از جمله مهم‌ترین دستاوردهای تحقیق، می‌توان به ارائه مدلی جهت پیش‌بینی تجمعات غیرقانونی و همچنین تعیین مؤلفه‌های تأثیرگذار رویداد امنیتی در به وقوع پیوستن تجمعات غیرقانونی اشاره کرد.

نتیجه‌گیری: مؤلفهٔ «نگرش فکری انتشاردهندهٔ فراخوان‌ها»، بالاترین امتیاز و نشانهٔ انجام تجمعات غیرقانونی براساس فراخوان‌های انتشاریافته در شبکه‌های اجتماعی را دارا است. به عبارت دیگر، نگرش فکری انتشاردهندهٔ فراخوان‌ها در ایجاد تجمعات نقش مستقیم دارد و پلیس با تمرکز بر روی نگرش فکری انتشاردهندهٔ فراخوان‌ها می‌تواند احتمال وقوع تجمعات غیرقانونی را بررسی کند تا از هدررفت و اعزام نیروها و مأموران در زمان مناسب برای جلوگیری از انجام تجمعات احتمالی فراخوان‌ها پیشگیری و جلوگیری به عمل آورد.

کلید واژه‌ها: شبکه‌های اجتماعی، فراخوان، تجمعات غیرقانونی، داده‌کاوی، رویدادهای امنیتی.

استناد: بختیاری، سعید و قاسم‌زاده، حمیدرضا (بهار ۱۳۹۹). پیش‌بینی تجمعات غیرقانونی مبتنی بر داده‌کاوی رویدادهای امنیتی. *فصلنامهٔ پژوهش‌های اطلاعاتی و جنایی*. ۱۵(۵۷)، صص ۱۱۸-۹۷.

۱. استادیار امنیت شبکه و اطلاعات دانشگاه علوم انتظامی امین، نویسندهٔ مسئول: saeid.bakhtiarii@chmail.ir
۲. کارشناس ارشد مهندسی فناوری اطلاعات دانشگاه علوم انتظامی امین، rezapalgosh@chmail.ir

مقدمه

امروزه رسانه‌های جدیدی از جمله شبکه‌های اجتماعی و نرم‌افزارهای موبایل‌پایه از بستر فضای مجازی و سایبر به وجود آمدند و در سطح جامعه گسترش یافته‌اند. گسترش این رسانه‌ها به‌واسطهٔ تجهیزات هوشمند و همراهی است که در اختیار مردم قرار دارد. این بسترها علاوه بر ایجاد فرصت‌های مختلف مانند اطلاع‌رسانی سریع اخبار و رویدادها، یافتن دوستان قدیمی، انتقال فایل‌ها، تصاویر و ویدئوها و کمک به بخش‌های مختلف اقتصادی مانند گسترش خدمات دولت الکترونیک، تبلیغات و تجارت الکترونیک، تهدیداتی نیز بر علیه امنیت و آسایش عمومی به همراه دارد. از انواع تهدیداتی که در این بستر رونق پیدا کرده است، موضوع انتشار فراخوان‌های مختلف با موضوعات فرهنگی، اجتماعی، سیاسی، امنیتی، صنفی و اقتصادی است.

امروزه گروه‌های مختلف با استفاده از این فرصت‌ها و تهدیدات، اقدام به انتشار فراخوان می‌کنند (خانیکی و جهرمی بصیریان، ۱۳۹۲). این گروه‌ها از طیف‌های مختلف از داخل و خارج می‌باشند؛ گروه‌هایی که از خارج کشور اقدام به انتشار فراخوان تجمعات غیرقانونی می‌کنند، اغلب معاند و مخالف نظام جمهوری اسلامی ایران می‌باشند و عمدتاً فراخوان‌های سیاسی- امنیتی را به قصد برهم زدن نظم و امنیت عمومی و به چالش کشاندن نظام و حاکمیت انجام می‌دهند و از برخی مسائل ملی‌گرایانه، قومیتی و مشکلات معیشتی در جهت تحریک کاربران فضای مجازی در داخل کشور جهت ایجاد تجمعات غیرقانونی سوءاستفاده می‌کنند. از نمونه‌های این فراخوان‌ها می‌توان به فراخوان برای تجمع در پاسارگاد و روز کوروش در ۷ آبان هر سال نام برد. این طیف همیشه مترصد کوچک‌ترین مسائل جهت رسیدن به اهداف شوم خود می‌باشند و سعی می‌کنند مسائل هرچند کوچک را با استفاده از گسترش فضای سایبر و با تهییج کردن مردم به چالش بزرگ امنیتی - اجتماعی تبدیل کنند. برای نمونه، اغتشاشات دی‌ماه سال ۱۳۹۶ در شهرهای مختلف ایران را می‌توان نام برد. در این اغتشاشات، بالغ بر ۲۰۰۰ فراخوان و رویداد امنیتی از گروه‌های داخلی، معاند، مخالف و سلطنت‌طلب به قصد ایجاد تجمع و اغتشاش و برهم زدن نظم و امنیت کشور در شبکه‌های اجتماعی تلگرام و فضای مجازی انتشار دادند. دستهٔ دیگر، گروه‌های داخلی هستند که اغلب با نیت مطالبات اقتصادی، فرهنگی و اجتماعی اقدام به انتشار فراخوان‌های امنیتی می‌کنند. این گروه‌ها ممکن است کارگران ناراضی کارخانه‌ها و بنگاه‌های اقتصادی باشند یا جوانان جویای هیجان که به

دنبال ایجاد تجمعات دورهمی، تجمع و مراسم آب‌بازی مختلط هستند یا گروه‌های حامی حیوانات و محیط‌زیست که از این فضا برای رسیدن به اهداف خود استفاده می‌کنند.

با توجه به اهمیت موضوع ذکرشده، قانون‌گذاران را بر آن داشته تا قوانینی را در جهت جلوگیری و مقابله و پیشگیری با این تهدیدات وضع کنند. در ایران، با این‌گونه فراخوان‌ها برای تجمعات غیرقانونی برابر قانون جرائم رایانه‌ای در بند (ج)، تحت عنوان محتوا علیه امنیت و آسایش عمومی (سطر ۱ - با تشکیل جمعیت، دسته، گروه در فضای مجازی (سایبر) با هدف برهم زدن امنیت کشور (مادهٔ ۴۹۸ ق.م.ا)) به عنوان جرم برخورد می‌شود و برای آن مجازات تعیین شده است و نقش و وظیفهٔ پلیس و سیستم قضایی، پیشگیری از این قبیل جرائم است و آن‌ها را بر آن داشته تا نسبت به شناسایی این قبیل فراخوان‌ها و انتشاردهندگان آن و همچنین نسبت به روش‌های پیشگیری و پیش‌بینی تجمعات غیرقانونی در راستای وظایف ذاتی خود، اقدام کنند. مسئله پژوهش حاضر این است که چه فراخوان‌هایی و با چه مشخصاتی منجر به تجمع خواهند شد؟ با توجه به گستردگی و حجم انبوه اطلاعات در نهادهای امنیتی، بانک‌های اطلاعاتی مناسبی جهت اجرای پروژه‌های داده‌کاوی وجود دارد تا با استفاده از روش‌های داده‌کاوی در شناسایی، پیش‌بینی و پیشگیری از تجمعات غیرقانونی در کشور، بتوان گام‌های مفید و مؤثری برداشت.

با به‌کارگیری بانک اطلاعاتی موجود در حوزهٔ رویدادهای امنیتی در فضای سایبری و استفاده از ابزارها و الگوریتم‌های داده‌کاوی، می‌توان الگوهای رفتاری تجمعات غیرقانونی را شناسایی و کشف کرد تا به این طریق، نیروهای امنیتی و انتظامی بتوانند وقوع تجمعات را پیش‌بینی کرده و با کنترل دقیق‌تر در حوزهٔ سایبری، اشراف بیشتری در این حوزه داشته باشند. از طرفی، محققان معتقدند عدم‌پردازش دقیق اطلاعات انبوه در بانک‌های اطلاعاتی سازمان‌ها و عدم کشف روابط پنهان اطلاعات، با توجه به حجم انبوه آن، باعث تضعیف مدیران سطح بالا در تصمیم‌گیری مؤثر می‌شود. داده‌کاوی به عنوان یکی از مراحل اصلی کشف دانش از پایگاه دادهٔ بزرگ در این امر بسیار مهم است؛ به‌طوری که نپرداختن به آن در درازمدت می‌تواند برای هر سازمانی آسیب‌زا باشد و موجب هدر رفت نیروها خواهد بود. در تحقیق حاضر، پژوهشگران قصد دارند با استفاده از داده‌کاوی، مدل و دانشی را از فراخوان‌های سال‌های اخیر استخراج کنند تا پیش‌بینی نمایند فراخوان‌ها با چه مشخصاتی منجر به ایجاد تجمع غیرقانونی خواهند شد و از این

طریق بتوان از ایجاد تجمعات غیرقانونی جلوگیری کرد و با استفاده از مدل احصاء شده، از هدر رفت و اعزام نیروهای امنیتی برای فراخوان‌هایی که منجر به تجمع نمی‌شوند جلوگیری شود. در همین راستا، مسئله اصلی این تحقیق در پاسخ به این سؤال اساسی است که چگونه می‌توان تجمعات غیرقانونی را با استفاده از داده‌کاوی رویدادهای امنیتی (فراخوان‌های تجمعات غیرقانونی) منتشرشده در فضای مجازی پیش‌بینی نمود؟ لذا، هدف اصلی این پژوهش، پیش‌بینی تجمعات غیرقانونی مبتنی بر داده‌کاوی رویدادهای امنیتی (فراخوان‌های تجمعات غیرقانونی) است که برای دست یافتن به این مهم، می‌بایست به اهداف فرعی زیر دست یافت:

- کشف مؤلفه‌های تأثیرگذار فراخوان‌ها در ایجاد تجمعات غیرقانونی؛
- بررسی روش‌های داده‌کاوی و انتخاب روش مناسب برای داده‌کاوی رویدادهای امنیتی برای پیش‌بینی تجمعات غیرقانونی؛
- احصاء ارتباط میان مؤلفه‌های تأثیرگذار در فراخوان‌هایی که منجر به تجمعات غیرقانونی می‌شوند.

پیش از ورود به مباحث نظری، می‌بایست برخی از مفاهیم و اصطلاحات مانند رویدادهای امنیتی، فراخوان، عرصه و زیر عرصه، مدنظر گرفته شود.

- منظور از رویداد، اتفاقی است که به میزان خارق‌العاده‌ای توجه رسانه‌ها را جلب می‌کند. این توجه عموماً دامنهٔ بین‌المللی دارد و از مرزهای بین اخبار عامه‌پسند و رویداد سیاسی می‌گذرد و معمولاً نقطهٔ مرجعی در تخیل فرهنگی و تاریخی پس از آن بر جا می‌گذارد. رویداد امنیتی نیز به کلیهٔ اخبار، فراخوان‌های تجمع و اعتصابات در زمینهٔ گوناگون سیاسی، اقتصادی، فرهنگی و غیره و وقایع و رخدادهایی که جنبهٔ امنیتی - انتظامی داشته باشد و می‌تواند منجر به برهم زدن نظم و امنیت عمومی جامعه شود، گفته می‌شود.
- اعلان عمومی یا خصوصی به صورت آشکار یا پنهان به‌منظور دعوت به اقدامی گروهی جهت دستیابی به اهداف مشخص و از پیش تعیین‌شده از جمله حمایت، تجمع، تحصن، اعتراض و غیره صورت می‌پذیرد.
- به زمینه‌های فعالیت افراد با توجه به مقاصد و نوع نگرش آن‌ها عرصه گویند که به بخش‌های مختلف فرهنگی، سیاسی، امنیتی و سایر زمینه‌ها اطلاق می‌شود. هر عرصهٔ فعالیت به چندین نوع فعالیت مختلف تقسیم می‌شود که در اصطلاح زیرعرصه گویند،

مانند موضوعاتی همچون بانک‌ها و مؤسسه‌های مالی، زیرعرصه‌ای از عرصهٔ اقتصادی خواهد بود.

با توجه به موضوع تحقیق، موضوعات مشابه به علل گوناگون مانند محرمانگی اطلاعات و تحقیقاتی که انجام گرفته، در دسترس عموم قرار نگرفته است، اما در سطح بین‌الملل و کتابخانه‌های مرجع، منابع و مقالات مرتبط با تحقیق مورد مطالعه قرار گرفت که در ادامه به برخی از مهم‌ترین آن‌ها اشاره می‌شود.

ابراهیمی، میرروشندل و آقایی (۱۳۹۴) در تحقیقی با عنوان «جامعیت‌بخشی به مجموعه دادهٔ جرائم به منظور پیش‌بینی و شناسایی جرائم با استفاده از تکنیک‌های داده‌کاوی» اشاره کرده‌اند که اطلاعات مظنونان ممکن است از نظر جغرافیایی و گستردگی دوره‌های زمانی طولانی، متفاوت باشد. همچنین، کشف جرائم مجازی ممکن است مشکل باشد؛ زیرا ترافیک شلوغ شبکه و تراکنش‌های درون‌خطی تکرارشونده، مقدار زیادی داده تولید می‌کند که تنها بخش کوچکی از این فعالیت‌های غیرقانونی را تشریح می‌کند. داده‌کاوی ابزاری قدرتمند ارائه می‌دهد که مأموران تحقیق جنایی که ممکن است فاقد آموزش باشند را قادر می‌سازد به عنوان تحلیل‌گران داده در اکتشاف پایگاه داده‌های بزرگ به‌سرعت و به‌طور مؤثر تحلیل کنند. علاوه بر این، هزینه‌های نصب و راه‌اندازی (کارکرد) نرم‌افزار اغلب کمتر از استخدام و آموزش کارکنان است. همچنین، رایانه‌ها نسبت به نیروی انسانی، مخصوصاً کسانی که ساعت‌های طولانی کار می‌کنند، کمتر در معرض اشتباه هستند. کیوان‌پور، جاویده و ابراهیمی (۱۳۸۸) نیز در مقالهٔ خود با عنوان «تحلیل رایانه‌ای جرم با بهره‌گیری از روش‌های هوش مصنوعی و داده‌کاوی در کشف پیش‌دستانهٔ جرم»، عمل تطابق جرم را بر روی متغیرهای جرم غیرفضایی یعنی متغیرهای رفتاری جرم انجام دادند. برای مثال، در نرم‌افزار پلیس یار که توسط آن‌ها طراحی و پیاده‌سازی شده، متغیرهای جرم سرقت از منازل در چهار گروه نوع محل مورد سرقت، نحوهٔ تعاملات مجرم با محیط سرقت، شیوهٔ ورود و ابزار مورد استفادهٔ مجرم دسته‌بندی شده‌اند.

محققان، کاربرد تکنیک‌های داده‌کاوی در مدل‌سازی جرائم را مبتنی بر چارچوب معرفی کردند. به‌طور کلی، در این چارچوب می‌توان یک دسته‌بندی کاربردی از کارهای انجام شده در زمینهٔ شناسایی، پیش‌بینی و پیشگیری جرائم به تفکیک کاربرد آن مشاهده کرد. از آنجایی که تمرکز این تحقیق بر روی به‌کارگیری علم داده‌کاوی در حوزهٔ

رویدادهای امنیتی است، لذا به تحقیقات صورت گرفته در حوزهٔ جرائم با استفاده از داده‌کاوی اشاره می‌شود. کارلیس و ملیکوسیدو[1] (۲۰۰۷) و مورتاق، گنز و مسکی[2] (۲۰۰۹) با استفاده از روش خوشه‌بندی داده‌کاوی و ماند و سرینیواس[3] (۲۰۱۲) با استفاده از روش خوشه‌بندی باینری در جهت به‌کارگیری علوم داده‌کاوی در حوزهٔ شناسایی جرائم تلاش کرده‌اند. در حوزهٔ پیش‌بینی جرائم، کراپسیوگلو و اردوگا[4] (۲۰۰۴)، مون و مسکلوسکی[5] (۲۰۱۲)، دالسیو و ستولزنبرگ[6] (۲۰۱۰)، دیدمن[7] (۲۰۰۳) و فریلیچ و پریدمور[8] (۲۰۰۷) با استفاده از روش رگرسیون و ایکسو و برون[9] (۲۰۰۶)، هادجیدج، دبای، لونیس، ایکبال، سزپورر و بنردجم[10] (۲۰۰۹) و بابو[11] (۲۰۱۱) با استفاده از روش خوشه‌بندی و بوکزاک و جیفورد[12] (۲۰۱۰) با استفاده از قوانین انجمنی فازی اقدام به داده‌کاوی کرده‌اند. همچنین، لی، کو و تیسای[13] (۲۰۱۰) با روش فازی و سام و اُتلی و ایوارت[14] (۲۰۰۳) با ترکیبی از فنون رگرسیون، شبکهٔ عصبی و شبکهٔ بیزین اقدام به داده‌کاوی در حوزهٔ پیشگیری جرائم کرده‌اند.

تظاهرات عمومی، تحریم‌ها و دیگر اشکال اعتراضی، از اشکال طبیعی در دستیابی به تغییر سیاسی برای شهروندان در دموکراسی‌های بالغ است. همین‌طور، در کشورهایی که موج سوم دموکراتیزه شدن را در دههٔ ۱۹۷۰ تجربه کردند، عملکرد رژیم‌های دموکراتیک منجر به کاهش رفتار اعتراضی شده است (اینگلهارت و گاتربرگ[15]، ۲۰۰۲). با این حال، کشورهای اسپانیا، یونان و پرتغال، شاهد افزایش فعالیت اعتراضی در چند سال گذشته بوده‌اند. فعالیت‌های اعتراض‌آمیز در شیلی افزایش یافته است و سال ۲۰۱۱، «زمستان شیلی» به عنوان گروه‌های وابسته به قبیله نام‌گذاری شده است. این اعتراضات نسبت به

1.Meligkotsidou
2.Murtagh, Ganz & McKie
3.Mande & Srinivas
4.Corapcıoglu & Erdogan
5.Moon & McCluskey
6.D'Alessio & Stolzenberg
7.Deadman
8.Freilich & Pridemore
9.Xue & Brown
10.Hadjidj, Debbabi, Lounis, Iqbal, Szporer & Benredjem
11.Baboo
12.Buczak & Gifford
13.Li, Kuo & Tsai
14.Oatley & Ewart
15.Inglehart & Catterberg

شروع به تقاضای تغییرات عمده در آموزش، محیط‌زیست، انرژی و سیاست انجام گرفته است. محصلان دبیرستانی و دانشگاهی، بیش‌ترین میزان این اعتراضات را داشته‌اند (والنزولا و همکاران، ۲۰۱۲). والنزولا، آریاگا و چرمن[1] (۲۰۱۲) در تحقیقی با عنوان «رفتار اعتراضی جوانان بر پایهٔ رسانه‌های شبکه‌های اجتماعی: مورد مطالعاتی شیلی» و با هدفِ یافتن ارتباط بین استفاده از رسانه‌های اجتماعی و اعتراض جوانان و همچنین مکانیسم‌های میانجیگری و تعدیل‌کننده، این رابطه را با استفاده از داده‌های نظرسنجی جمع‌آوری‌شده در شیلی در سال ۲۰۱۰ بررسی کرد. نتایج تحقیق ایشان نشان می‌دهد که استفاده از فیس‌بوک حتی پس از در نظر گرفتن نارضایتی‌های سیاسی، منابع مادی و روانی، ارزش‌ها و استفاده از رسانه‌های خبری به‌طور قابل‌توجهی با فعالیت اعتراضی همراه بوده است. چن[2] و همکارانش (۲۰۰۴) نیز تحقیقی با عنوان «داده‌کاوی جرم؛ یک چارچوب کلی و چند نمونه» با هدف تجزیه و تحلیل تکنیک‌های داده‌کاوی از طریق شبکه‌های رسانه‌ای اجتماعی بین سال‌های ۲۰۰۳ تا ۲۰۱۵ انجام دادند. در این تحقیق، استراتژی‌های تحقیق مبتنی بر معیارهای ۶۶ مقاله به عنوان منبع بررسی استفاده شده است. پس از بررسی دقیق این مقالات، دریافتند که ۱۹ تکنیک داده‌کاوی با داده‌های رسانه‌های اجتماعی برای رفع نه هدف تحقیق مختلف در شش حوزهٔ صنعتی و خدمات مختلف استفاده شده است. با این حال، برنامه‌های کاربردی داده‌کاوی در رسانه‌های اجتماعی هنوز خام بوده و نیاز به تلاش بیشتر توسط دانشگاه‌ها و صنعت است و پیشنهاد شده است تحقیقات بیشتری توسط دانشگاه‌ها و صنعت انجام شود؛ زیرا مطالعات انجام‌شده تاکنون به اندازهٔ کافی کامل نبوده است.

ساهو و داروکار[3] (۲۰۱۷) پژوهشی با موضوع «تکنیک‌های داده‌کاوی در دسته‌بندی جرائم سایبری»، انجام داده‌اند. داده‌کاوی و مدیریت جرائم سایبری یک برنامهٔ پرکاربرد است که نقش مهمی در مدیریت اطلاعات جرم‌شناختی ایفا می‌کند. بنابراین، تحقیقات جنایی سایبری نقش بسیار مهمی در نظام پلیس هر کشوری دارد. در این تحقیق، تکنیک‌های داده‌کاوی برای تحلیل داده‌های وب مورد استفاده قرار گرفت و مطالعهٔ دقیق در مورد طبقه‌بندی و خوشه‌بندی ارائه شده است. کلاسه‌بندی، فرآیند طبقه‌بندی نوع

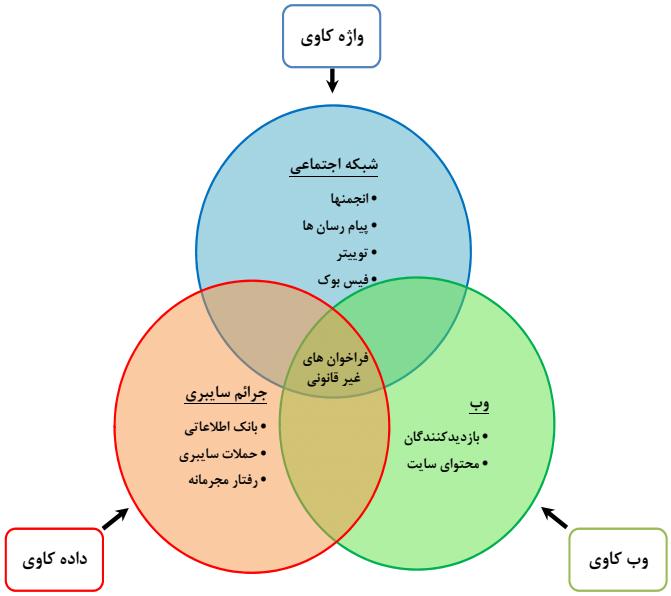
1.Valenzuela, Arriagada & Scherman
2.Chen
3.Sahu & Darokar

جرم و خوشه‌بندی، فرآیند ترکیب داده‌های داده به گروه‌ها است. ساختار سناریو این بود که ویژگی‌ها و روابط را در صفحهٔ وب استخراج و سناریوی جرم‌یابی را بازسازی کند. خان، رادهان و فاطیما[1] (۲۰۱۷) نیز در تحقیقی با عنوان «اعمال تکنیک‌های داده‌کاوی در جرائم سایبری»، تمرکز خود را بر روی حملات انکار خدمات (DoS) با کمک تکنیک‌های تشخیص الگو در داده‌کاوی قرار دادند. در سراسر جهان، افراد زیادی در حوزه‌های مختلف دسترسی به اینترنت دارند. هنگامی سرورها خدماتی را در اختیار کاربران قرار می‌دهند، عملکردی وجود دارد که می‌توان فعالیت کاربران را در فایل‌های log مشاهده کرد. این فایل‌ها شرح مفصلی از فعالیت‌های کاربران را که در شبکه رخ می‌دهد مانند آدرس IP، زمان ورود و خروج، رفتار کاربر و غیره است. حملات مختلفی در اینترنت وجود دارد و تمرکز در این تحقیق، حملات DDOS با کمک تکنیک‌های تشخیص الگو در داده‌کاوی است و از طریق آن، حملهٔ DDOS شناسایی می‌شود. DDOS یک حملهٔ بسیار خطرناک است که به منابع اطلاعاتی یک سازمان آسیب می‌رساند و از طریق ارسال پیام‌های زیاد به سمت سرور انجام می‌گیرد. با استفاده از روش پیشنهادی خان و همکاران، می‌توان انکار حملهٔ سرویس را به‌راحتی در حمله DoS تشخیص داد. سینگهال[2] (۲۰۱۳) نیز تحقیق دیگری با عنوان «کشف آزار و اذیت‌های سایبری با استفاده از داده‌کاوی» انجام داد. در این تحقیق یک نظرسنجی از کاربران انجام و شبیه‌سازی در شبکه‌های اجتماعی صورت گرفته است تا بتوان با استفاده از داده‌کاوی از کاربران مراقبت کرد.

انجام عملیات واکاوی داده‌ها به سه بخش داده‌کاوی، واژه‌کاوی و وب‌کاوی انجام می‌پذیرد. همان‌طور که در پیشینه گفته شد، برخی از پژوهش‌ها در بخش شبکه‌های اجتماعی انجام پذیرفته، با کمک واژه‌کاوی به نتایجی که مورد نظرشان بود دست یافتند، اما نوع دیگر واکاوی داده‌ها در محیط وب انجام می‌پذیرد که از روشی تحت عنوان وب‌کاوی استفاده می‌شود. در این روش، اطلاعات موردنظر از محتوی سایت‌ها و کاربران این محیط بهره‌برداری می‌شود، اما داده‌های تجمیع‌شده در پایگاه داده‌های مختلف را که با حجم بالا به صورت متمرکز جمع‌آوری شده با استفاده از روشی تحت عنوان داده‌کاوی

1.Khan, Pradhan & Fatima
2.Singhal

تحلیل و بهره‌برداری می‌کند. در این پژوهش، از تلفیق این سه بخش استفاده شده است. در بخش شبکه‌های اجتماعی، از اطلاعات و فراخوان‌های رصد شده توسط کارشناسان رصد پلیس فتا استفاده شده و همچنین از فراخوان‌های رصد شدهٔ وب‌سایت‌ها که در بانک اطلاعاتی تجمیع شده است، بهره‌برداری شده و سپس با استفاده از روش داده‌کاوی بر روی بانک اطلاعاتی موجود که حاصل رصد کارشناسان این بخش در طول سال‌های ۹۵ و ۹۶ است، تجزیه و تحلیل انجام می‌شود تا به اهداف اصلی و فرعی این پژوهش دست یافت.



نمودار ۱ – مدل مفهومی پیش‌بینی فراخوان‌های غیرقانونی در فضای مجازی

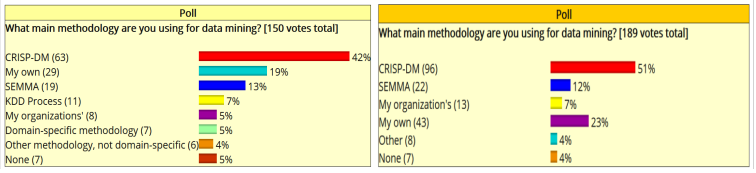
روش‌شناسی تحقیق

روش‌های مختلفی برای پیاده‌سازی و اجرای پروژه‌های داده‌کاوی وجود دارد که در این تحقیق، سه روش معروف در این حوزه بررسی شد. یکی از روش‌های بسیار قوی در حوزهٔ داده‌کاوی، روش کریسپ دی‌ام[1] است. این روش به علت اینکه یک رویهٔ همسان در کل پروژه‌های داده‌کاوی ترسیم می‌کند و از جامعیت و مقبولیت نیز برخوردار است، بسیار

1.Crisp-Dm

موردقبول متخصصان این رشته است. از مهم‌ترین موارد انتخاب این متدولوژی برای تحقیق حاضر می‌توان به موارد زیر اشاره کرد:

- خبرگان داده‌کاوی، اغلب از متدولوژی CRISP-DM بهره می‌برند. بنا به گزارش سایت KDnuggets که از پیشگامان مبحث داده‌کاوی به حساب می‌آید، در نظرسنجی‌های خود از خبرگان در سال‌های ۲۰۰۲، ۲۰۰۴ و ۲۰۰۷ در مورد متدولوژی مورد استفاده در اجرای فرآیند داده‌کاوی در کسب‌وکار خود، به نتایج ارائه شده در شکل ۱ رسیده است. همان‌طور که از نتایج برمی‌آید، متدولوژی CRISP-DM همواره در این سال‌ها پیشتاز بوده است و در کنار آن متدولوژی سما[1] قرار گرفته است.



ب) روش‌های داده‌کاوی ۲۰۰۷ الف) روش‌های داده‌کاوی ۲۰۰۲

شکل ۱ – نمودار نظرسنجی از خبرگان در خصوص نوع متدولوژی مورد استفاده در داده‌کاوی

- اگرچه سایر متدولوژی‌ها، دارای عناصر ضروری هر پروژهٔ داده‌کاوی هستند، اما تنها مربوط به مسائل آماری، مدل‌سازی و ابزار تغییر داده‌ها از روند داده‌کاوی است. این متدولوژی‌ها، فاقد برخی از بخش‌های اساسی از هر پروژهٔ سیستم‌های اطلاعاتی از جمله تجزیه و تحلیل، طراحی و مرحلهٔ پیاده‌سازی است. اما مهم‌تر اینکه نقش سازمان و سهامداران را در طول پروژه در نظر نمی‌گیرد و داده‌کاوی را به عنوان بخش جدایی‌ناپذیر در یک چشم‌انداز سیستم نمی‌بیند (آزودو و سنتوس[2]، ۲۰۰۸).

- با توجه به مقبولیت، سادگی و جامع بودن این روش و همچنین سازگاری آن با رویکرد تحقیق فعلی، از روند کلی مدل استاندارد CRISP-DM که در نرم‌افزار SPSS MODELER18 پایه‌گذاری شده است و گام‌های آن برای اجرای پروژG حاضر استفاده شده است.

تکنیک‌ها و روش‌های یادگیری مدل در داده‌کاوی فراخوان‌ها: همان‌طور که در بالا بحث شد، روش‌های داده‌کاوی به دو دستهٔ کلی تقسیم می‌شوند؛ داده‌کاوی پیش‌بینی کننده (با ناظر) و داده‌کاوی توصیفی (بی ناظر). در تحقیق حاضر، متغیر هدف (وضعیت

1.SEMMA
2.Azevedo and Sentos

انجام تجمع غیرقانونی)، از نوع گسسته است و با توجه به موضوع از روش‌های مدل‌سازی پیش‌بینانه باید استفاده شود. بر این اساس، تکنیک‌های دسته‌بندی برای داده‌کاوی استفاده شده است. در این تحقیق از الگوریتم‌های درخت تصمیم (CHAID,Quest,CART,Random Tree, Tree-AS,C5)، نزدیک‌ترین همسایه، شبکهٔ بیزین و شبکهٔ عصبی برای داده‌کاوی به عنوان روش‌های پیشگیرانه استفاده شد. براساس گام‌های شش‌گانهٔ متدلوژی CRISP-DM، بانک اطلاعاتی داده‌ها و فراخوان‌ها بررسی می‌شود. در نهایت، نتایج داده‌کاوی تحلیل و الگوریتم بهینه برای داده‌کاوی داده‌های حوزهٔ فراخوان‌ها و رویدادهای سایبری به همراه قوانین استخراجی بیان می‌شود.

یافته‌های تحقیق

براساس گام‌های شش‌گانهٔ CRISP-DM، مراحل متدولوژی که به ترتیب شناخت کسب‌وکار، شناخت داده، آماده‌سازی داده، مدل‌سازی داده، ارزیابی و به‌کارگیری و پیاده‌سازی مدل است، در خصوص موضوع تحقیق ارائه می‌شود.

شناخت داده‌ها: براساس بانک اطلاعاتی، رویدادهای امنیتی رصد شده در سال‌های ۹۵ و ۹۶ در بانک اطلاعات براساس مؤلفه‌های زیر ثبت می‌شود.

جدول ۱ – لیست مؤلفه‌های فراخوان‌ها

استان	شهر	عرصه	زیرعرصه	عنوان	آدرس	تاریخ	ساعت	بستر	نگرش	تعداد اعضاء	بازدید	وضعیت

اطلاعات کلی شامل استان، شهر، آدرس تجمع احتمالی، تاریخ تجمع، ساعت تجمع، عرصه، زیرعرصه، عنوان فراخوان، منبع انتشار فراخوان، بستر انتشار فراخوان در فضای سایبر، تعداد اعضای منبع انتشاردهنده، تعداد بازدید فراخوان در فضای سایبر، نگرش فکری انتشاردهنده یا انتشار دهنگان فراخوان‌ها و در نهایت وضعیت تجمع که آیا تجمعی با انتشار هر فراخوان صورت گرفته است یا خیر که به عنوان هدف این داده‌کاوی مورد بررسی و تحلیل قرار می‌گیرد. بنابراین، از بین فیلدها و ویژگی‌های ذخیره‌شده در بانک اطلاعاتی پلیس فتا ناجا، فیلدهای فوق انتخاب شدند.

استخراج و آماده‌سازی داده‌ها: علیرغم اینکه داده‌های مثبت و مفیدی در بانک اطلاعات ثبت شده بود، اما به دلیل اینکه داده‌های موردنیاز این تحقیق نبوده است،

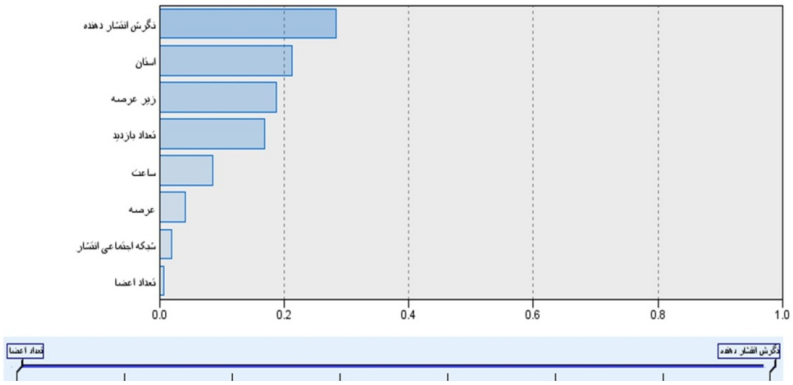
داده‌های این تحقیق به‌صورت انتخابی با نظر خبرگان تهیه شد. سپس با استفاده از نرم‌افزار اکسل، داده‌ها محدودسازی و رکوردهای خاص بررسی شد تا قبل از به‌کارگیری داده‌ها در نرم‌افزار داده‌کاوی IBM SPSS MODELER، داده‌ها آماده‌سازی شوند، به‌طوری که بخش اعظمی از زمان این تحقیق صرف آماده‌سازی و پالایش داده‌ها شد.

در داده‌های اولیه، برخی از فیلدها در رکوردها مفقود شده یا دارای اطلاعات نویزی بوده یا برخی از رکوردها غیرضروری بوده است که به صورت دستی به روش‌های زیر اصلاح شد:

- رکوردهایی که تعداد ویژگی مفقود شدهٔ آن زیاد بود، به‌طور کامل حذف شده است. به عنوان مثال، در بانک اطلاعاتی، وضعیت تجمع یک فراخوان ثبت نبود و کل رکورد حذف شد؛
- پر کردن مقادیر مفقود شده با محاسبهٔ ارتباط بین نمونه‌ها و ویژگی‌ها؛
- کردن مقادیر مفقودشده با محاسبهٔ میانگین همان ویژگی؛
- پر کردن مقادیر مفقودشده با مقداری که بیش‌ترین تکرار را داشته است؛
- جایگزین کردن مقدار مناسب در داده‌های نویزی.

مدل‌سازی: همان‌طور که قبلاً اشاره شد، با توجه به اینکه متغیر هدف تحقیق (وقوع یا عدم وقوع تجمع) گسسته است، تکنیک‌ها و مدل‌های بکار رفته در این تحقیق، مدل‌های دسته‌بندی (روش با ناظر) است. در ادامه، داده‌ها براساس الگوریتم‌های درخت تصمیم (CHAID,Quest,CART,Random Tree, Tree-AS,C5)، نزدیک‌ترین همسایه، شبکهٔ بیزین و شبکهٔ عصبی مدل‌سازی می‌شود.

جهت مدل‌سازی در روش‌های با ناظر، ابتدا داده‌ها به دو گروه آموزشی و آزمون تقسیم می‌شود. پس از تعیین داده‌های آموزشی و آزمون، در مرحلهٔ بعد، مدلی را که عملیات داده‌کاوی با آن انجام می‌شود (به عنوان مثال، مدل C5.0) انتخاب می‌شود. در مرحله بعد، داده‌های ورودی و خروجی را تعیین و سپس مدل اجرا می‌گردد. پس از اجرای مدل، عملیات داده‌کاوی بر روی داده‌ها انجام شده و نتیجه در محیط نرم‌افزار نشان داده می‌شود. با کلیک بر روی نتیجه داده‌کاوی، می‌توان فیلدهای تأثیرگذار در عملیات داده‌کاوی را به همراه قوانین خروجی مشاهده نمود که در شکل ۲ نشان داده می‌شود.



شکل ۲ – نتایج حاصل از اجرای داده‌کاوی

پس از اجرای داده‌کاوی، در مرحلهٔ بعدی از فرآیند CRISP، مدل بایستی ارزیابی شود. با ارزیابی مدل‌های اجراشده که در ادامه ارائه می‌شود، در بهترین حالت، ۹۱/۷۲ درصد که مربوط به مدل C5.0 است و در بدترین حالت مربوط به مدل شبکهٔ عصبی بوده با دقت ۳۳/۱۷ درصد است.

اجرای مدل‌های مختلف بر روی مؤلفه‌های فراخوان‌ها: با انجام داده‌کاوی حاصل از بانک اطلاعاتی داده‌های ۱۶ استان و ۱۷۸ شهر کشور در خصوص فراخوان‌های تجمعات غیرقانونی به شرح جدول ۲ برای مدل‌های مختلف داده‌کاوی با دقت‌های مختلف به دست آمده است.

جدول ۲ – دقت مدل‌های داده‌کاوی انجام‌شده

ردیف	نام مدل داده‌کاوی	دقت برای داده‌های هدف	دقت برای داده‌های آزمون
۱	مدل شبکهٔ بیز	۷۸/۶۹ درصد	۷۷/۶ درصد
۲	مدل شبکهٔ عصبی	۶۵/۹ درصد	۳۳/۱۷ درصد
۳	مدل نزدیک‌ترین همسایه KNN	۸۵/۲۷ درصد	۸۲/۶۲ درصد
۴	مدل درخت تصمیم CHAID	۸۷/۹۱ درصد	۸۵/۸ درصد
۵	مدل درخت تصمیم QUEST	۸۶/۱۶ درصد	۸۴/۸۲ درصد
۶	مدل درخت تصادفی random trea	۹۰/۱۷ درصد	۸۶/۵۴ درصد
۷	مدل درخت trea as	۸۴/۱ درصد	۸۲/۸۶ درصد
۸	مدل داده‌کاوی C5	۹۱/۷۲ درصد	۸۸/۹۸ درصد

با توجه به نتایج به دست آمده از داده‌کاوی فراخوان‌ها، مدل‌های C5 و درخت تصادفی random trea بالاترین دقت را در برابر سایر مدل‌های انجام یافته به دست آمده است که به عنوان نتیجهٔ داده‌کاوی، دقت بسیار بالایی به حساب می‌آید.

رتبه‌بندی مشخصه‌های ورودی تأثیرگذار فراخوان‌ها بر وقوع تجمعات غیرقانونی: یکی دیگر از نتایج مهمی که از اجرای مدل بر روی داده‌ها می‌توان گرفت،

رتبه‌بندی مشخصه‌های ورودی تأثیرگذار بر روی وقوع تجمعات غیرقانونی ناشی از انتشار فراخوان‌ها و رویدادهای امنیتی است. در فرایند داده‌کاوی فراخوان‌ها، در اجرای هر مدل، مشخصه‌های تأثیرگذار متفاوت بوده است. همان‌طور که مشخص شد، مدل درخت تصمیم C5 بهترین مدل برای داده‌های ثبت‌شده در بانک اطلاعاتی فراخوان‌ها و رویدادهای امنیتی است. با عنایت به این موضوع، قوانین استخراجی در این تحقیق از این مدل منتج شد. بنابراین، مشخصه‌های ورودی تأثیرگذار نیز از این مدل رتبه‌بندی می‌شود. برای مشخصه‌های تأثیرگذار نیز در هر یک از مدل‌های داده‌کاوی به نسبت متفاوت است. جدول مشخصه‌های تأثیرگذار فراخوان‌های تجمعات غیرقانونی به شرح جدول ۳ است.

جدول ۳ – رتبه‌بندی مشخصه‌ها براساس مدل‌های مختلف داده‌کاوی

ردیف	نام مدل داده‌کاوی	رتبهٔ ۱	رتبهٔ ۲	رتبهٔ ۳	رتبهٔ ۴	رتبهٔ ۵	رتبهٔ ۶	رتبهٔ ۷
۱	مدل شبکه بیز	نگرش	بازدید	زیرعرصه	ساعت	بستر	اعضا	شهر
۲	مدل شبکه عصبی	بستر	زیرعرصه	اعضا	بازدید	استان	تاریخ	منبع
۳	مدل نزدیک‌ترین همسایه KNN	-	-	-	-	-	-	-
۴	مدل درخت تصمیم CHAID	نگرش	استان	ساعت	زیرعرصه	شهر	عرصه	اعضاء
۵	مدل درخت تصمیم QUEST	زیرعرصه	نگرش	استان	عرصه	ساعت	شهر	بازدید
۶	مدل درخت تصادفی random trea	ساعت	عرصه	نگرش	شهر	اعضا	بستر	بازدید
۷	مدل درخت trea as	زیرعرصه	شهر	نگرش	ساعت	بازدید	اعضاء	استان
۸	مدل داده‌کاوی C5	نگرش	بازدید	شهر	زیرعرصه	استان	اعضاء	ساعت

با یک وزن‌دهی ساده به مؤلفه‌ها و مشخصه‌های فراخوان‌ها در مدل‌های مختلف داده‌کاوی، می‌توان رتبه‌بندی این مؤلفه‌ها را از مجموع مدل‌های بررسی‌شده به دست آورد که با انجام این عمل، رتبه‌بندی مشخصه‌های فراخوان‌های تجمعات غیرقانونی به ترتیب مطابق، جدول ۳ محاسبه شد. هرچند مشخص است مؤلفهٔ وضعیت وقوع تجمع که به عنوان هدف داده‌کاوی بوده است، به عنوان مهم‌ترین و تأثیرگذارترین مؤلفه از بین مؤلفه‌های دیگر هر فراخوان به شمار می‌رود. این مؤلفه در فرایند داده‌کاوی به عنوان مؤلفهٔ خروجی و هدف موردنظر بوده است، اما از بین مؤلفه‌های ورودی فرایند داده‌کاوی، با انجام محاسبه از جدول ۳، این رتبه‌بندی مطابق جدول ۴ به دست آمد.

جدول ۴ – جمع‌بندی و رتبه‌بندی مؤلفه‌های ورودی فرایند داده‌کاوی برحسب مجموع مدل‌های بررسی‌شده

رتبهٔ مشخصه	رتبهٔ ۱	رتبهٔ ۲	رتبهٔ ۳	رتبهٔ ۴	رتبهٔ ۵	رتبهٔ ۶	رتبهٔ ۷	رتبهٔ ۸	رتبهٔ ۹
مؤلفه	نگرش	زیرعرصه	ساعت	شهر	بازدید	استان	اعضاء	بستر	عرصه
امتیاز	۳۷	۳۶	۲۴	۲۱	۲۱	۱۸	۱۵	۱۲	۱۲

مؤلفه‌ها و رتبهٔ تأثیرگذاری آن‌ها در فرایند داده‌کاوی برای مدل C5 که بالاترین دقت را در داده‌کاوی فراخوان‌های تجمعات غیرقانونی مدنظر بوده است، مطابق جدول ۵ به دست آمده است.

جدول ۵ – رتبه‌بندی مؤلفه‌های ورودی تأثیرگذار در فرایند داده‌کاوی برحسب مدل با بالاترین دقت

رتبه مشخصه	رتبهٔ ۱	رتبهٔ ۲	رتبهٔ ۳	رتبهٔ ۴	رتبهٔ ۵	رتبهٔ ۶	رتبهٔ ۷	رتبهٔ ۸	رتبهٔ ۹
مدل C5	نگرش	بازدید	شهر	زیرعرصه	استان	اعضاء	ساعت	بستر	عرصه

همان‌طور که در شکل ۲ نمایش داده شده، مشخصه‌های تأثیرگذار بر روی داده‌های تحقیق حاضر به ترتیب نگرش فکری انتشاردهندهٔ فراخوان‌ها، تعداد بازدید فراخوان، استان محل تجمعات غیرقانونی، عرصه، زیرعرصه فراخوان‌ها، شهر محل وقوع تجمع و هدف فراخوان، ساعت وقوع تجمع مربوط به فراخوان و تعداد اعضای گروه و کانال انتشاردهنده، بیش‌ترین تأثیر را در وقوع تجمعات غیرقانونی ناشی از انتشار رویدادهای امنیتی و فراخوان‌ها دارد.

ارزیابی مدل: جهت مدل‌سازی در روش‌های با ناظر، ابتدا داده‌ها به دو گروه آموزشی و آزمون تقسیم می‌شود. در این تقسیم‌بندی، ۳۰ درصد داده‌ها آزمون و ۷۰ درصد آموزشی در نظر گرفته می‌شود. پس از ارزیابی مدل، نتیجه به صورت ماتریس آشفتگی نمودار می‌شود. ماتریس آشفتگی درخت تصمیم مدل C5 به شرح جدول ۶ است و نشان می‌دهد که این درخت تمامی فراخوان‌ها را پیش‌بینی کرده است. با توجه به موارد بالا و نتایج حاصله که در جدول ۶ آمده است:

- پیش‌بینی تجمعات برای کلیهٔ داده‌ها انجام گرفته است و ستون داده‌های پیش‌بینی نشده، خالی و صفر است؛
- برای داده‌های هدف، از مجموع ۹۸۰ مورد دادهٔ تجمعات انجام نپذیرفته، ۸۶۲ مورد را درست و ۱۱۸ مورد را غلط پیش‌بینی کرده است؛
- برای داده‌های هدف، از مجموع ۲۱۶۰ مورد تجمعات انجام شده، ۲۰۱۸ مورد را درست و ۱۴۲ مورد را غلط پیش‌بینی کرده است؛
- دقت پیش‌بینی مدل با ۹۱/۷۲ درصد دقت موردقبول است.

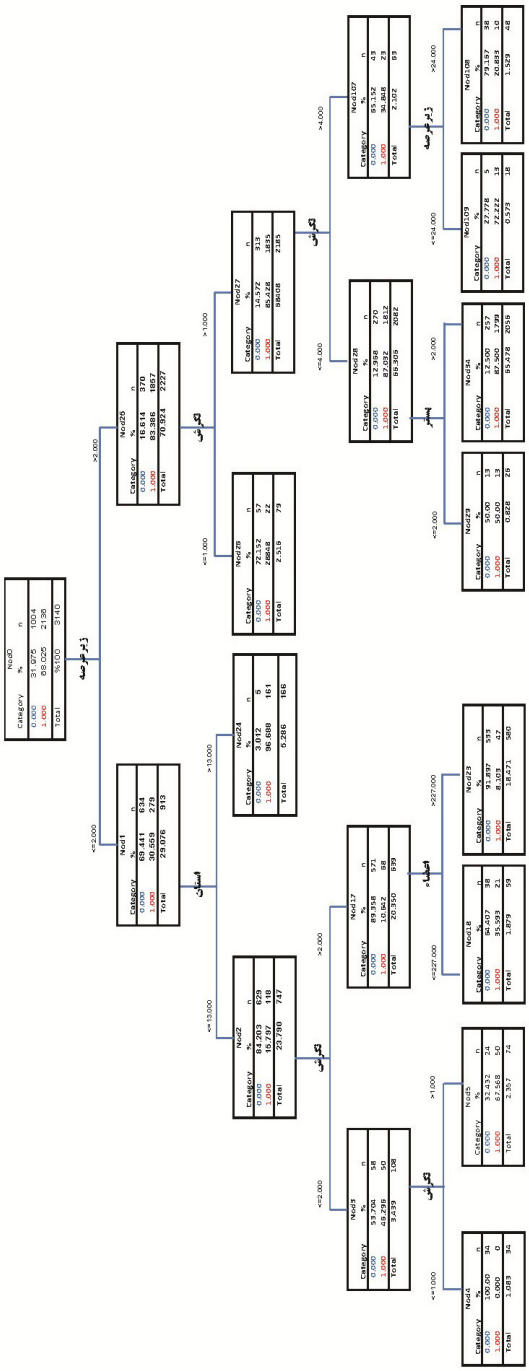
جدول ۶ – ماتریس آشفتگی درخت C5

خالی	۱/۰۰۰۰	۰/۰۰۰۰	داده‌های آزمون (تست)
۰	۴۹	۲۱۵	۰/۰۰۰۰۰
۰	۵۱۲	۴۱	۱/۰۰۰۰
خالی	۱/۰۰۰۰	۰/۰۰۰۰	داده‌های آموزش
۰	۱۴۲	۸۶۲	۰/۰۰۰۰۰
۰	۲۰۱۸	۱۱۸	۱/۰۰۰۰

قوانین استخراجی از اجرای مدل: با توجه به اینکه مدل درخت تصمیم C5، بر روی موضوعات فراخوان‌ها به‌طور جداگانه اجرا شد، بنابراین قوانین استخراجی نیز از اجرای مدل بر روی‌داده‌های فراخوان‌ها به تفکیک موضوعات و مؤلفه‌های فراخوان‌ها است. نکتهٔ حائز اهمیت در اینجا این است که با توجه به اینکه هدف این پروژه، کشف روابط بین خصوصیات فراخوان‌ها در ایجاد تجمعات غیرقانونی است، لذا قوانین کشف‌شده، قوانینی مربوط به نتایج یک و صفر شدن متغیر خروجی در هر یک از موضوعات فراخوان‌های تجمعات غیرقانونی (نگرش انتشاردهنده، استان، شهر، عرصه، زیرعرصه، تعداد اعضاء، تعداد بازدید و ساعت) است. قوانین وقوع تجمعات غیرقانونی بر حسب انتشار فراخوان‌ها و مؤلفه‌های آن‌ها به صورت شبه‌کد در جدول ۷ نمایش داده شده است. همچنین، بخشی از ساختار درختی مدل ارائه‌شدهٔ وقوع تجمعات غیرقانونی برحسب انتشار فراخوان‌ها و مؤلفه‌های آن‌ها در شکل ۳ نمایش داده شده است.

جدول ۷ – رول‌های مدل ارائه‌شده

Rules for 1 contains 12 rule(s)
1- Rule 1 for 1.0
If (state > 1 & sub_arena > 2 & seen > 224 & seen <= 255) then 1.000
2- Rule 2 for 1.0
If (state > 14) then 1.000
3- …
⋮
12- Rule 12 for 1.0
If (city > 57 & city <= 153 & clock = 900 & sub_arena > 2 & attitude > 1) then 1.000
Rules for 0 - contains 19 rule(s)
13- Rule 1 for 0.0
If (state <= 6 and clock = 1700 and sub arena <= 2 and seen > 139) then 0.000
14- …
⋮
31- Rule 19 for 0.0
If (state > 5 & city > 71) then 0.000
Default: 1
	State = "استان"
	City = "شهر"
	Arena = "عرصه"
	Sub_arena = "زیر عرصه"
	Date = "تاریخ"
	Clock = "ساعت"
	Follower = "تعداد اعضاء"
	Seen = "تعداد بازدید"
	OTT = "بستر انتشاردهنده"
	Attitude = "نگرش انتشاردهنده"
	Status = "وضعیت وقوع تجمع"



شکل ۳ – نمای درخت تصمیم C5.0

بحث و نتیجه‌گیری

هدف از اجرای این تحقیق، پیش‌بینی تجمعات غیرقانونی مبتنی بر داده‌کاوی رویدادهای امنیتی (فراخوان‌های تجمعات غیرقانونی) بوده است و این داده‌کاوی بر روی بانک اطلاعاتی فراخوان‌های رصد شده از فضای مجازی در طی سال‌های ۱۳۹۵ و ۱۳۹۶ که توسط کارشناسان رصد پلیس فتا جمع‌آوری شده بود، صورت پذیرفت. با توجه به اینکه وضعیت انجام تجمع هر فراخوان به عنوان هدف داده‌کاوی در نظر گرفته شد، بنابراین در این بین، فراخوان‌هایی که دارای بازخورد میدانی بودند و نتیجهٔ انتشار فراخوان و انجام تجمع در بانک اطلاعاتی ثبت شده بود، به عنوان دادهٔ ورودی قابل بهره‌برداری به تعداد ۳۱۴۰ مورد مورد استفاده قرار گرفت که مربوط به ۱۶ استان و ۱۷۸ شهر بوده است و با بهره‌گیری از نرم‌افزار IBM SPSS MODELR18 و با بارگذاری داده‌های بانک اطلاعاتی یاد شده در این نرم‌افزار، با متدولوژی CRISP-DM و با الگوریتم‌های مختلف داده‌کاوی با قابلیت پیش‌بینی اجرا شد. وضعیت انجام تجمعات غیرقانونی برحسب مؤلفه‌های مختلف فراخوان‌ها بررسی شد. پس از بررسی‌های به عمل آمده، مؤلفه‌های تأثیرگذار بر روی داده‌های تحقیق حاضر به ترتیب نگرش فکری انتشاردهندهٔ فراخوان‌ها، تعداد بازدید فراخوان، استان محل تجمعات غیرقانونی، عرصه، زیرعرصهٔ فراخوان‌ها، شهر محل وقوع تجمع و هدف فراخوان، ساعت وقوع تجمع مربوط به فراخوان و تعداد اعضای گروه و کانال انتشاردهنده بوده‌اند که به ترتیب، بیش‌ترین تا کمترین تأثیرگذاری را بر روی انجام تجمعات غیرقانونی داشته‌اند. از بین الگوریتم‌های مختلفی که بررسی شد، مدل C5 با دقت ۹۱/۷۲ درصد، بالاترین نتیجه را برای داده‌های موردنظر و انجام تجمعات غیرقانونی داده است.

نتایج این تحقیق با تحقیق والنزولا و همکاران (۲۰۱۲) کاملاً همسو است؛ زیرا در تحقیقات ایشان اثبات شد که بین نارضایتی‌های عمومی و سوءاستفاده از شبکه‌های اجتماعی جهت برگزاری تجمعات غیرقانونی ارتباط مستقیمی وجود دارد. همچنین، این تحقیق با به‌کارگیری تجربیات تحقیقات چن و همکاران (۲۰۰۴) و ساهو (۲۰۱۷) در حوزهٔ به‌کارگیری علوم داده‌کاوی صورت پذیرفته است که همسویی این تحقیق با تحقیقات گذشته را نشان می‌دهد.

با توجه به مدل به دست آمده، تأثیرگذارترین مؤلفهٔ فراخوان‌ها برای پیش‌بینی تجمعات غیرقانونی، نگرش و دیدگاه انتشاردهندگان فراخوان است که باید در بانک

اطلاعاتی فراخوان‌های رصد شده به صورت کامل ثبت شود و پیشنهاد می‌شود بانک اطلاعاتی از فراخوان‌ها بر مبنای این مؤلفه‌ها (مشخصه‌های تأثیرگذار) تهیه شود و برای تکمیل این تحقیق توسط آیندگان و بهره‌برداری کامل‌تر از اطلاعات به دست آمده از این پژوهش، بانک اطلاعاتی از کلیهٔ فراخوان‌های تجمعات غیرقانونی کشور که در فضای مجازی انتشار می‌یابند، به همراه اطلاعات بازخوردهای میدانی هر فراخوان در بانک اطلاعاتی مربوطه به صورت کامل ثبت و ضبط شود. با توجه به تأثیرگذاری مؤلفه‌های مورد بررسی در این تحقیق، دقت در ثبت اطلاعات واقعی فراخوان‌ها بسیار می‌تواند در تشخیص و پیش‌بینی درست انجام تجمعات غیرقانونی در سطح کشور و بالا بردن احساس امنیت در بین شهروندان مؤثر باشد. همچنین، از هدر رفت و اعزام نیروها و مأموران برای جلوگیری از انجام تجمعات احتمالی فراخوان‌ها ممانعت به عمل خواهد آمد. بنابراین، به عنوان پیشنهاد کاربردی و پژوهشی، می‌توان به موارد زیر اشاره کرد:

- می‌توان با جمع‌آوری پایگاه داده‌ای کامل از تمامی فراخوان‌های منتشرشده در کلیهٔ مناطق کشور، به پایگاه داده‌ای جامع در زمینهٔ فراخوان‌ها دست یافت. از مزایای این کار، کامل بودن بانک اطلاعاتی فراخوان‌ها در سطح کشور است و همین‌طور با این کار می‌توان تأثیر ویژگی‌های فراخوان‌ها بر تجمعات غیرقانونی کلیهٔ مناطق کشور را مورد بررسی قرار داد.
- می توان با اندیشیدن تدابیری، محدودیت پایگاه دادهٔ موجود در زمینه‌های رویدادهای امنیتی، فراخوان‌ها و اعتصابات را از بین برد.
- می‌توان از الگوریتم‌های دیگر داده‌کاوی در این زمینه استفاده کرد.
- می‌توان با ترکیب الگوریتم‌هایی که درجهٔ درستی بالایی دارند، به نتایجی بهتر دست یافت.

منابع

منابع فارسی

- ابراهیمی، مجیب؛ میرروشندل، سیدابوالقاسم و آقایی، جان احمد (۱۳۹۴). جامعیت بخشی به مجموعه دادهٔ جرائم به منظور پیش‌بینی و شناسایی جرائم با استفاده از فنون داده‌کاوی. فصلنامهٔ صنایع الکترونیک. ۶(۴)۴، صص ۱۱-۶.
- خانیکی، هادی و بصیریان جهرمی، حسین (تابستان ۱۳۹۲). کنشگری و قدرت در

شبکه‌های اجتماعی مجازی. *فصلنامهٔ علوم اجتماعی*. ۲ (۶۱)، صص۴۵-۸۰. بازیابی از:
http://qjss.atu.ac.ir/article_9797.html
- کیوانپور، محمدرضا؛ جاویده، مصطفی و ابراهیمی، محمدرضا (۱۳۸۸). تحلیل رایانه‌ای جرم با بهره‌گیری از روش‌های هوش مصنوعی و داده‌کاوی کشف پیش‌دستانهٔ جرم. *فصلنامهٔ کارگاه*. ۲(۷)، صص۱۱۷-۹۸. بازیابی از:
http://journals.police.ir/article_10614.html

منابع انگلیسی

- Azevedo, A. I. R. L., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. IADS-DM. Retrieved from: https://recipp.ipp.pt/handle/10400.22/136
- Baboo, S. S. (2011). An enhanced algorithm to predict a future crime using data mining. International Journal of Computer Applications, 975, 8887. Retrieved from: https://pdfs.semanticscholar.org/195a/247055cd1be24a4f27c607fc8c6a75a64f2f.pdf
- Buczak, A. L. & Gifford, C. M. (2010). Fuzzy association rule mining for community crime pattern discovery. In ACM SIGKDD Workshop on Intelligence and Security Informatics (pp. 1-10). https://doi.org/10.1145/1938606.1938608
- Chen, H. Chung, W. Xu, J. J. Wang, G. Qin, Y. & Chau, M. (2004). Crime data mining: a general framework and some examples. computer, 37(4), 50-56 .DOI: 10.1109/MC.2004.1297301
- D'Alessio, S. J. & Stolzenberg, L. (2010). Do cities influence co-offending? Journal of Criminal Justice, 38(4), 711-719. DOI: https://doi.org/10.1016/j.jcrimjus.2010.04.045
–Dayan, D. & Katz, E. (1994). Media events: harvard university press. ISBN: 0674559568, 9780674559561
–Deadman, D. (2003). Forecasting residential burglary. International journal of forecasting, 19(4), 567-578 .DOI: https://doi.org/10.1016/S0169-2070(03)00091-8
–Freilich, J. D. & Pridemore, W. A. (2007). Politics, culture, and political crime: Covariates of abortion clinic attacks in the United States. Journal of Criminal Justice, 35(3), 323-336. DOI: https://doi.org/10.1016/j.jcrimjus.2007.03.008
- G.C. Oatley, J. Z. B.W. Ewart. (2005). Matching and Predicting Crimes. 19-32. DOI: https://doi.org/10.1007/1-84628-103-2_2
- Hadjidj, R. Debbabi, M. Lounis, H. Iqbal, F. Szporer, A. & Benredjem, D. (2009). Towards an integrated e-mail forensic analysis framework. digital

investigation, 5(3), 124-137. DOI: https://doi.org/10.1016/j.diin.2009.01.004
- Inglehart, R. & Catterberg, G. (2002). Trends in political action: The developmental trend and the post-honeymoon decline. international Journal of comparative Sociology, 43(3-5), 300-316. DOI: https://doi.org/10.1177/002071520204300305
- Khan, M. A. Pradhan, S. K. & Fatima, H. (2017). Applying Data Mining techniques in Cyber Crimes. Paper presented at the Anti-Cyber Crimes (ICACC), 2017 2nd International Conference on. DOI: 10.1109/Anti-Cybercrime.2017.7905293
- Li, S.-T., Kuo, S.-C. & Tsai, F.-C. (2010). An intelligent decision-support model using FSOM and rule extraction for crime prevention. Expert Systems with Applications, 37(10), 7108-7119 .DOI: https://doi.org/10.1016/j.eswa.2010.03.004
- Mande, U., Srinivas, Y. & Murthy, J. (2012). An intelligent analysis of crime data using data mining & auto correlation models. Int J Eng Res Appl (IJERA), 2(4), 149-153 .Retrieved from: https://b2n.ir/180186
- Moon, B., McCluskey, J. D. & McCluskey, C. P. (2010). A general theory of crime and computer crime: An empirical test. Journal of Criminal Justice, 38(4), 767-772. DOI: https://doi.org/10.1016/j.jcrimjus.2010.05.003
- Murtagh, F., Ganz, A. & McKie, S. (2009). The structure of narrative: the case of film scripts. Pattern Recognition, 42(2), 302-312. DOI: https://doi.org/10.1016/j.patcog.2008.05.026
- Oatley, G. C. & Ewart, B. W. (2003). Crimes analysis software:'pins in maps', clustering and Bayes net prediction. Expert Systems with Applications, 25(4), 569-588. DOI: https://doi.org/10.1016/S0957-4174(03)00097-6
- Sahu, N. & Darokar, S. (2017). Data Mining Techniques to Clustering Cyber Crime Data. International Education and Research Journal, 3(7). Retrieved from: http://ierj.in/journal/index.php/ierj/article/view/1288
- Singhal, P. & Bansal, A. (2013). Improved textual cyberbullying detection using data mining. International Journal of Information and Computation Technology, 3(6), 569-576. Retrieved from: https://pdfs.semanticscholar.org/aa57/e82fce1b367ba5959023552bfe702c91d506.pdf
- Valenzuela, S., Arriagada, A. & Scherman, A. (2012). The social media basis of youth protest behavior: The case of Chile. Journal of Communication, 62(2), 299-314. DOI: https://doi.org/10.1111/j.1460-2466.2012.01635.x
- Xue, Y. & Brown, D. E. (2006). Spatial analysis with preference specification of latent decision makers for criminal event prediction. Decision support systems, 41(3), 560-573. DOI: https://doi.org/10.1016/j.dss.2004.06.007