

Predicting the Protest Movements in Cyberspace via Using Data Mining

Hosein Talebian¹

Received: 10, March, 2021

Accepted: 10, August, 2021

Abstract

Background and Aim: Today, various groups in society use cyberspace to publish calls for illegal gatherings, followed by protests. Therefore, the main purpose of this study is to predict protest flows in cyberspace based on data mining.

Method: The present study is applied in terms of purpose and descriptive-analytical in terms of method, in which the information data in documents, organizational documents as well as decentralized databases in police operational categories have been used. The statistical population of the study is all the information data of illegal gatherings in cyberspace that were used for sampling of data processing method and in full. In order to predict and prevent crime, using Rapid Miner software as an open source data mining tool in Java, we first identify the main components of illegal aggregation calls in cyberspace, the relationships between the components identified in the data, and the data in the data analysis.

Results: Among the various implemented models, the decision tree model with 91.39% accuracy provided the highest result for the data and predicted the occurrence of protest flows, and its most important achievements were designing a model for predicting illegal gatherings and determining the effective components in The occurrence of rallies is illegal.

Conclusion: The components of "thinking attitude of the publisher of the calls, the number of visits of the calls, the subfield of the calls and the number of members of the sources and targets had the most to the least impact on the occurrence of illegal rallies and protests; In other words, the ideological attitude of the publisher of the calls has a direct role in creating rallies, and the police can focus on it to predict the possibility of illegal rallies.

Keywords: Prediction, Call, Illegal Gatherings, Protest Streams, Cyberspace, Data Mining.

1. Assistant Professor of Crime Detection Department. Faculty of Intelligence and Criminal Investigation, Amin Police University. Email: h.talebian@yahoo.com

فصلنامه علمی پژوهش‌های مدیریت انتظامی

سال شانزدهم، شماره سوم، پاییز ۱۴۰۰

صص ۱۹۷-۱۷۳

پیش‌بینی جریان‌های اعتراضی در فضای مجازی مبتنی بر داده‌کاوی

حسین طالبیان^۱

تاریخ دریافت: ۱۳۹۹/۱۲/۲۰

تاریخ پذیرش: ۱۴۰۰/۰۵/۱۹

چکیده

زمینه و هدف: امروزه گروه‌های مختلفی در جامعه با استفاده از فضای مجازی اقدام به انتشار فراخوان‌های تجمعات غیرقانونی و به دنبال آن وقوع جریان‌های اعتراضی می‌کنند. از این رو هدف اصلی این پژوهش، پیش‌بینی جریان‌های اعتراضی در فضای مجازی مبتنی بر داده‌کاوی است.

روش: تحقیق حاضر از نظر هدف کاربردی و از لحاظ روش، توصیفی - تحلیلی است که در آن از داده‌های اطلاعاتی موجود در آساند، مدارک سازمانی و همچنین بانک‌های اطلاعاتی غیرمتمرکز در رده‌های عملیاتی پلیس استفاده شده است. جامعه آماری پژوهش، تمامی داده‌های اطلاعاتی فراخوان‌های تجمعات غیرقانونی در فضای مجازی است که برای نمونه‌گیری از روش پیش‌پردازش داده‌ای و به‌صورت تمام‌شمار استفاده شدند. به‌منظور پیش‌بینی و پیشگیری از ارتکاب جرائم، با استفاده از نرم‌افزار رپیدماینر به‌عنوان یک ابزار داده‌کاوی متن‌باز با زبان جاوا، ابتدا مؤلفه‌های اصلی فراخوان‌های تجمعات غیرقانونی در فضای مجازی شناسایی، روابط بین مؤلفه‌های شناسایی شده احصا، داده‌ها تحلیل و در نهایت مؤلفه‌های مؤثر در پیش‌بینی جریان‌های اعتراضی اولویت‌بندی شد.

یافته‌ها: از بین مدل‌های مختلف پیاده‌سازی شده، مدل درخت تصمیم با دقت ۹۱/۳۹ درصد، بالاترین نتیجه را برای داده‌های موردنظر و پیش‌بینی وقوع جریان‌های اعتراضی ارائه کرد و مهم‌ترین دستاوردهای آن، طراحی مدلی جهت پیش‌بینی تجمعات غیرقانونی و همچنین تعیین مؤلفه‌های تأثیرگذار در وقوع تجمعات غیرقانونی است. **نتیجه‌گیری:** مؤلفه‌های «نگرش فکری انتشاردهنده فراخوان‌ها، تعداد بازدید فراخوان‌ها، زیرعرصه فراخوان‌ها و تعداد اعضای منابع و نشانه‌ها» به ترتیب بیشترین تا کمترین تأثیرگذاری را بر وقوع تجمع‌ها و جریان‌های اعتراضی غیرقانونی داشته‌اند؛ به عبارت دیگر، نگرش فکری انتشاردهنده فراخوان‌ها در ایجاد تجمعات نقش مستقیم دارد و پلیس با تمرکز بر آن می‌تواند احتمال وقوع تجمعات غیرقانونی را پیش‌بینی کند.

کلیدواژه‌ها: پیش‌بینی، فراخوان، تجمعات غیرقانونی، جریان‌های اعتراضی، فضای مجازی، داده‌کاوی.

➤ استناد: طالبیان، حسین؛ (۱۴۰۰). پیش‌بینی جریان‌های اعتراضی در فضای مجازی مبتنی بر داده‌کاوی، ۱۶

(۳)، صص ۱۹۷-۱۷۳

DOI: dorl.net/dor/10.22034/PMSQ.2021.96919

مقدمه

بدون شک، گسترش استفاده از فناوری اطلاعات و انجام بسیاری از امور در فضای مجازی و در واقع رواج استفاده افراد و سازمان‌ها از اینترنت، گرایش و تمایل به انجام جرم در این محیط را افزایش داده است. امروزه، بسیاری از افرادی که در محیط‌های واقعی به دلایل متعدد از جمله شرم و حیا، مرتکب عمل مجرمانه‌ای نمی‌شدند، در فضای مجازی به دلیل ویژگی‌های خاص آن، به ارتکاب جرم تمایل پیدا کرده‌اند. مصداق بارز آن گروه‌های مختلف اجتماعی هستند که با استفاده از فضای رسانه‌ای مبتنی بر محیط سایبر و شبکه‌های اجتماعی، به انتشار فراخوان‌های تجمعات غیرقانونی با موضوع‌ها و گرایش‌های متنوع از جمله تجمع در پاسارگاد برای گرامیداشت سالروز کوروش و یا تجمعات متعدد در دی‌ماه سال ۹۶، اقدام کردند.

اغتشاشات دی‌ماه سال ۹۶ در شهرهای مختلف کشور شکل گرفت و بار امنیتی سنگینی به‌ویژه برای ناجا به‌وجود آورد. در این اغتشاشات بالغ بر ۶۰۰۰ مورد فراخوان و رویداد امنیتی از سوی عناصر داخلی، معاند، مخالف و سلطنت‌طلب به قصد ایجاد تجمع، اغتشاش و برهم زدن نظم و امنیت عمومی کشور در شبکه‌های اجتماعی و فضای مجازی انتشار یافت و به دنبال آن نیز جریان‌های اعتراضی پیوسته و مرتبط با انتشار این نوع فراخوان‌ها در فضای واقعی جامعه اجرا شد. (بولتن اطلاعاتی پلیس فتا ناجا، ۱۳۹۶).

هرچند فراخوان تجمعات غیرقانونی مستند به بند (ج) قانون جرائم رایانه‌ای با عنوان محتوا علیه امنیت و آسایش عمومی و بند یکم ماده ۴۹۸ ق.م.ا؛ همراه با تشکیل جمعیت، دسته، گروه در فضای مجازی با هدف برهم زدن امنیت کشور به‌عنوان جرم تلقی و برای آن مجازات تعیین شده است؛ اما بنا بر آمار موجود، تنها در بازه زمانی سال ۹۷ به تعداد ۴۹۰۰ مورد فراخوان در فضای مجازی ایجاد و به تبع آن تعداد ۱۲۰۰ مورد تجمع و جریان اعتراضی (تعداد نفرات در هر تجمع بالای ۱۰۰ نفر) در ابعاد جغرافیایی شهرستان، استان و حتی در سطح کشور روی داده است. البته یکی از مهم‌ترین نقش‌ها و وظایف پلیس و سیستم قضایی در این حوزه، از بین بردن فرصت‌های ارتکاب جرم،

پیشگیری وضعی از وقوع جرم و برخورد با مجرمان بود که این امر تا حدودی به درستی انجام شد. (بولتن اطلاعاتی پلیس فتا ناجا، ۱۳۹۸).

در این راستا، سازمان‌های پلیسی و جنایی به‌ویژه در دهه اخیر برای تحلیل و کشف جرائم و شناسایی روابط پیچیده میان جرم و مجرم، از ابزارها و فناوری‌های مختلفی بهره گرفته‌اند و در این بین، داده کاوی به‌عنوان ابزاری قدرتمند برای چنین سازمان‌هایی مطرح شده و در تحلیل داده‌های جرم نقش مؤثری را در کشف الگو و دانش از پایگاه داده جرم ایفا کرده است و در فرایند تصمیم‌گیری، کاهش فرصت ارتکاب و کنترل جرائم را در پی دارد و در نهایت می‌تواند روش‌های شناسایی جرائم را به مدل تبدیل و ارائه کند.

بر همین اساس، امروزه پلیس به‌منظور ایجاد و برقراری امنیت در فضای مجازی و به‌طور خاص برای جلوگیری از ایجاد تجمعات غیرقانونی در سطح جامعه در تلاش است تا با رصد فراخوان‌ها و جریان‌های اعتراضی در فضای مجازی، با استفاده از اطلاعات احصا شده و تحلیل آن‌ها به‌وسیله داده کاوی، ابتدا سایر یگان‌های اجرایی و عملیاتی را مطلع و سپس مجموعه خود را در جهت کنترل و مقابله با اغتشاشات آماده کند و از تبعات امنیتی این‌گونه تجمع‌ها و جریان‌های اعتراضی در فضای واقعی جامعه بکاهد.

در پژوهش حاضر، محقق قصد دارد با استفاده از ابزار داده کاوی، مدل و دانشی را از رابطه فراخوان‌های منتشره با جریان‌های اعتراضی سال‌های اخیر، استخراج و مؤلفه‌های مؤثر در فراخوان‌های منتشره بر وقوع جریان‌های اعتراضی را شناسایی کند تا پلیس بر اساس الگوی موصوف، به‌موقع از جریانی شدن فراخوان‌های تجمعات غیرقانونی در فضای واقعی جامعه مطلع شده و قبل از شکل‌گیری آن‌ها، اقدامات لازم را به عمل آورد. درواقع، این پژوهش به دنبال پاسخگویی به این سؤال اساسی است که چگونه می‌توان جریان‌های اعتراضی در فضای مجازی را با استفاده از داده کاوی پیش‌بینی کرد؟

پیشینه و مبانی نظری

از نظر پیشینه، می‌توان گفت داده‌کاوی دارای کاربرد وسیعی در حوزه‌های مختلف از جمله صنعتی، پزشکی، ارتباطات، کشاورزی، انرژی، علوم اجتماعی، فرهنگی، سیاسی، اقتصادی، بازرگانی، نظامی و آموزشی است؛ به گونه‌ای که امروزه مرز و محدودیتی برای کاربرد این دانش در نظر گرفته نمی‌شود. البته، در حوزه داده‌کاوی جرائم نیز پژوهش‌های خوبی انجام شده که در ادامه به برخی از آن‌ها اشاره می‌شود.

قاسم‌زاده (۱۳۹۸)، در تحقیق خود با عنوان «پیش‌بینی تجمعات غیرقانونی مبتنی بر داده‌کاوی» بر اساس راهکارهای مبتنی بر فناوری اطلاعات و با استفاده از نرم‌افزارهای داده‌کاوی تجزیه و تحلیل، مدل‌های مختلفی از روند تشکیل تجمعات غیرقانونی را بررسی و در پایان مدلی از روش پیش‌بینی تجمعات غیرقانونی پیشنهاد داده است.

شاه‌محمدی و رجبی (۱۳۹۷)، در پژوهش خود با عنوان «تحلیل جرائم سایبری با استفاده از داده‌کاوی» با هدف کشف الگوریتم مناسب برای داده‌های جرائم اخلاقی سایبری، شش گام داده‌کاوی متدولوژی کریسپ را بر روی داده‌های جرائم اخلاقی سایبری اجرا و سپس در مدل‌سازی جرائم با اعمال الگوریتم‌های درخت تصمیم^۱، نزدیک‌ترین همسایه^۲، شبکه بیزین^۳ و شبکه عصبی^۴ بر روی داده‌های پژوهش به این نتیجه دست یافت که الگوریتم C5 دارای بالاترین دقت برای تحلیل داده‌های جرائم اخلاقی سایبری است.

ابراهیمی و همکاران (۱۳۹۴)، در تحقیقی با عنوان «جامعیت بخشی به مجموعه داده‌ها به منظور پیش‌بینی و شناسایی جرائم با استفاده از تکنیک‌های داده‌کاوی» با هدف شناسایی الگوهای جرم، کشف و سپس پیشگیری، ماهیت پیچیده داده‌های بزه، بزهکاری و روابط نامحسوس میان داده‌ها را توسط ابزارها و الگوریتم‌های داده‌کاوی

-
1. Decision Tree
 2. k-nearest neighbors
 3. Bayesian network
 4. Artificial Neural Networks - ANN

بررسی و در نهایت به مدل پیش‌بینی ویژگی‌های جرائم ارتكابی در آینده و شناسایی جرائم مرتبط با آن‌ها دست یافته است.

مظاهری حسین آبادی و همکاران (۱۳۹۴)، در تحقیق خود با عنوان «شناسایی گره‌های قدرت در شبکه‌های اجتماعی به کمک داده کاوی» با هدف شناسایی گره‌های قدرت در شبکه اجتماعی تبیان از طریق کشف الگوهای پنهان در خصیصه‌های کاربران بر اساس روش استاندارد کریسپ^۱ به مهم‌ترین ویژگی‌های مرتبط در تعیین میزان قدرت یک کاربر در شبکه اجتماعی تبیان از جمله تحصیلات، سن و جنسیت رسیده است.

ربیعی (۱۳۹۲)، در تحقیقی با عنوان «تحلیل جرائم شبکه‌ای با استفاده از داده کاوی» با هدف کشف روشی مبتنی بر روش‌های داده کاوی جهت استخراج دانش از گزارش‌های متنی پلیس، با استخراج ویژگی‌های جرم، رفتار بازیگران یک شبکه را تحلیل و در آخر به راهکارهای شناسایی ارتباطات پنهان بین مجرمان دست یافته است.

جوشی^۲ (۲۰۱۸)، در تحقیقی با موضوع «داده کاوی برای جلوگیری از جرائم» با هدف پیش‌بینی و شناسایی الگوی انواع جرائم انجام شده در ایالت گجرات (یکی از ایالات هند) از تکنیک‌های داده کاوی استفاده کرده و به این نتیجه دست یافته که پلیس آگاهی گجرات می‌تواند سوابق جرائم خود را مورد تجزیه و تحلیل قرار دهد و بینش معناداری در تصمیم‌گیری برای حل پرونده‌های ثبت شده ارائه کند.

فارسى^۳ (۲۰۱۸)، در تحقیق خود با عنوان «داده کاوی جرم، تجزیه و تحلیل، تهدید و پیش‌بینی» با هدف احصای برخی از تکنیک‌های محاسباتی قادر به پیش‌بینی و پیش‌گیری از جرائم در پاکستان، تکنیک‌های تحلیل پیش‌بینی در چارچوب فضای مجازی را ارائه کرده است.

ساهو^۴ (۲۰۱۷)، در تحقیقی با عنوان «تکنیک‌های داده کاوی در دسته‌بندی جرائم سایبری» با هدف احصای فنون داده کاوی برای تحلیل داده‌های وب از طریق

1. Cross-industry standard process for data mining as CRISP-DM

2. Joshi

3. Farsi

4. Sahu

کلاس‌بندی و خوشه‌بندی جرائم سایبری ژاپن، به استخراج ویژگی‌ها و روابط در صفحه وب و بازسازی سناریوهای جرائم دست یافته است.

کمال طاها^۱ (۲۰۱۷)، در تحقیقی با عنوان «شناسایی رهبران شبکه‌های جنایی با استفاده از درخت پنهان» با بررسی داده‌های موجود در پرونده‌های جنایی پلیس ترکیه، ضمن ارائه الگوی شناسایی رهبران شبکه‌های جنایی به این نتیجه رسیده که سیستم تحلیل فارنزیک به نام ای.سی.آل. فایندر^۲ جهت شناسایی اعضای تأثیرگذار یک سازمان جنایتکار مؤثر و مفید است و می‌تواند در مجموعه پلیس بسیار کارآمد باشد.

داده‌کاوی. داده‌کاوی شامل استفاده از ابزارهای پیشرفته تحلیل داده به منظور کشف الگوهای معتبر، از قبل ناشناخته و روابط موجود در مجموعه داده‌های بزرگ است که از طریق ترکیبی از شیوه‌های یادگیری ماشینی، تشخیص الگو، آمار، تلخیص و ارتباط بین مفاهیم به طور خودکار و با هدف کمک به فرایند تصمیم‌گیری از طریق استخراج دانش از داده‌ها صورت می‌گیرد. به طور کلی، کاربرد فنون داده‌کاوی در شناخت جرائم در قالب دو دسته شامل شناسایی اقدامات قبل از وقوع جرائم و به منظور پیش‌بینی و پیشگیری از ارتکاب جرم و شناسایی اقدامات بعد از وقوع جرم به منظور بررسی و کشف مدارک و شواهد جرم صورت می‌گیرد؛ اما یکی از مهم‌ترین کارکردهای داده‌کاوی پیش‌بینی است که می‌تواند در پیش‌بینی جرائم و به تبع آن، پیشگیری و مبارزه با جرائم مفید واقع شود (پولادی، ۱۳۹۶، ۴۵).

اصولاً عملیات داده‌کاوی شامل جداسازی داده‌های مفید، یکپارچه‌سازی داده‌های مختلف، انتخاب داده لازم از میان دیگر داده‌ها، انتقال داده به محیط داده‌کاوی برای کشف قوانین، ایجاد مدل‌ها و الگوهای مرتبط به وسیله روش‌های داده‌کاوی، ارزیابی مدل و الگوهای ایجادشده برای تشخیص مفید بودن آن‌ها و در آخر انتشار دانش استخراج‌شده به کاربران نهایی است (شاه‌محمدی و رجبی، ۱۳۹۷، ۷۱).

البته در داده کاوی از روش‌های متعددی از جمله 'KDD، 'SEMMA و 'CRISP استفاده می‌شود، ولی روش CRISP از جامعیت و مقبولیت بیشتری برخوردار است. این روش، یک مدل فرایندی برای داده کاوی ارائه می‌دهد که مروری بر چرخه حیات هر پروژه داده کاوی بوده و شامل شش مرحله به شرح شکل (۱) است.



شکل ۱. مدل فرایندی داده کاوی

پیش بینی. پیش بینی، فرایند برآورد موقعیت‌های ناشناخته است. در واقع پیش بینی، امکان پیش گویی در مورد رویدادهای آینده را در اختیار می‌گذارد و می‌تواند تجارب گذشته را به پیش بینی حوادث آینده بدل سازد. در سال‌های اخیر، پیش بینی، به‌نوعی آینده‌نگری را در پی داشته و غالباً به همین منظور انجام می‌شود (آرمسترونگ، ۲۰۱۷، ۴).

1. Knowledge Discovery in Database (KDD)
2. Sample, Explore, Modify, Model, Assess (SEMMA) Knowledge
3. Cross Industry Standard Process for Data Mining (CRISP-DM)
4. Armstrong

فضای مجازی. واژه فضای مجازی برای نخستین بار توسط ویلیام گیسون^۱ در سال ۱۹۸۲ میلادی مورد استفاده قرار گرفت. او فضای مجازی را فضایی تخیلی می‌دانست که از اتصال رایانه‌هایی پدید آمده که تمامی انسان‌ها، ماشین‌ها و منابع اطلاعاتی در جهان را به هم متصل می‌کند. این معنا تقریباً مشابه معنایی است که امروزه از کاربرد لفظ فضای مجازی در نظر گرفته می‌شود. در واقع، امروزه فضای مجازی توصیفی برای تمامی انواع منابع اطلاعاتی موجود در شبکه‌های رایانه‌ای است که دارای ویژگی‌هایی اعم از هزینه کم، گمنامی، نامتقارن بودن در آسیب‌پذیری، جهانی و فرامرزی بودن، دسترسی دائم و آسان به آخرین اطلاعات، جذابیت و تنوع، عدم وابستگی به زمان و مکان خاص، چندرسانه‌ای بودن، سهولت تعامل و تبادل اطلاعات با دیگران و سرعت بالای تبادل اطلاعات و امکانات قابل توجه اینترنت برای افراد جامعه است. (رجبی، ۱۳۹۶، ۹).

جرائم رایانه‌ای

سوءاستفاده از رایانه‌ها اعم از هر رفتار غیرقانونی، غیراخلاقی یا غیرمجاز که مربوط به پردازش خودکار و انتقال داده باشد، به نوعی جرم رایانه‌ای است. در واقع «هر عمل مثبت غیرقانونی که رایانه در آن ابزار یا موضوع جرم باشد، جرم رایانه‌ای محسوب می‌شود». (قاسم‌زاده، ۱۳۹۸، ۱۴).

فراخوان تجمعات غیرقانونی. فراخوان در اصطلاح لغوی به معنای احضار و اعلان عمومی است و تجمع به معنی گردآمدن مردم از هر جا و یا به تعبیری دیگر به جمع شدن موقت گروهی از مردم که طی چند ساعت به بحث و تبادل نظر و تصمیم‌گیری درباره مسائل خاصی که ممکن است اجتماعی یا سیاسی و یا غیره باشد. (دهخدا، ۱۳۱۹). تجمعات غیرقانونی نیز به تلاش برای ایجاد و تشدید اختلاف میان صفوف ملت با استفاده از زمینه‌های متنوع فرهنگی، مذهبی و نژادی موجود در جامعه گفته می‌شود (عالی‌پور، ۱۳۹۶، ۳۶۱).

جریان اعتراضی. جریان را می‌توان مجموعه‌ای از افراد دارای مبانی و اصول فکری، هدف و آرمان مشترک دانست که در جهت رسیدن به آن هدف و آرمان به اقدامات و کنش‌های جهت‌دار و همسو به صورت سازمان‌یافته یا سازمان‌نیافته (موج‌های هدایت و کنترل ناشده جمعی) دست می‌زنند. جریان به این اعتبار، با جنبش اجتماعی، حزب، طبقه اجتماعی (در اندیشه مارکس)، قشر اجتماعی (که ملاک آن منزلت و قدرت است)، صنف و مفاهیمی از این سنخ قرابت می‌یابد (شرف‌الدین، ۱۳۹۵، ۸). اما در جنبش‌های اجتماعی بسیاری از مردم به طور فعال برای ایجاد دگرگونی یا ممانعت از برخی دگرگونی‌ها شرکت می‌کنند و به اصطلاح یک رفتار جمعی از خود بروز می‌دهند که تحت شرایط زودگذر، بدون ساختار و بی‌ثبات و به سبب فقدان آشکار هنجارهای مقبول اتفاق می‌افتد. این‌گونه رفتارها، عاطفی، خودانگیخته، اتفاقی و کمتر قابل پیش‌بینی‌اند. در واقع، رفتارهای جمعی اعتراض آمیز نیز در دسته رفتارهای گروهی شهروندان جای می‌گیرد که هیچانی‌اند، اغلب مخرب و ویرانگرند، تکانه‌ای هستند و با منافع مقامات حکومتی در تضادند. تنها نکته‌ای که آن‌ها را از هم متمایز می‌سازد، تعارض رفتارهای جمعی با منافع حکومتی است. (شریفی، ۱۳۹۸).

در شکل (۲) الگوی شماتیک اجرای داده‌کاوی آمده است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



شکل ۲. الگوی شماتیک اجرای داده‌کاوی

روش

پژوهش حاضر از نظر هدف، کاربردی و از لحاظ ماهیت، توصیفی-تحلیلی است که با استفاده از تکنیک تحلیل محتوا و از طریق داده‌کاوی انجام شده است. جامعه آماری پژوهش، تمامی داده‌های اطلاعاتی فراخوان‌های تجمعات غیرقانونی و جریان‌های اعتراضی احصا شده از فضای مجازی در بازه زمانی سال ۱۳۹۷ به تعداد ۴۹۰۰ را دربرمی‌گیرد که به صورت تمام‌شمار استفاده شده است. در این پژوهش از نرم‌افزار رپیدماینر^۱ به‌عنوان یک ابزار داده‌کاوی متن‌باز با زبان جاوا استفاده شده و به‌منظور پیش‌بینی و پیشگیری از ارتکاب جرائم، ابتدا مؤلفه‌های اصلی فراخوان‌های تجمعات غیرقانونی در فضای مجازی شناسایی، سپس روابط بین مؤلفه‌های شناسایی شده احصا، داده‌ها تحلیل و در آخر مؤلفه‌های مؤثر در پیش‌بینی جریان‌های اعتراضی اولویت‌بندی

و مدل شده‌اند. برای نیل به این اهداف، داده‌های موجود در بانک‌های اطلاعاتی پلیس فتا ناجا در حوزه رویدادها و جریان‌های فضای مجازی، مربوط به بازه زمانی سال ۱۳۹۷ جمع‌آوری و سپس با استفاده از الگوریتم کریسپ و الگوهای متداول آن، تحلیل شده و در نهایت، الگوهای وقوع جریان‌های اعتراضی شناسایی و طراحی شده‌اند. گفتنی است، فنون و روش‌های داده‌کاوی مطابق شکل زیر به دو دسته کلی شامل روش‌های پیش‌بینانه و روش‌های توصیفی تقسیم می‌شود (شکل ۳). در داده‌کاوی پیش‌بینانه با استفاده از مجموعه داده‌ها و بر اساس الگوریتم‌های مشخصی، مدل‌هایی را برای پیش‌بینی عملکرد متغیرهای مختلف ارائه می‌کند. برخی از الگوهای رایج در این تکنیک شامل درخت تصمیم^۱، نزدیک‌ترین همسایه^۲، شبکه بیزین^۳ و شبکه عصبی^۴ است (عباسی، ۱۳۹۵، ۱۰۶).



شکل ۳. فنون و روش‌های داده‌کاوی

در بین الگوهای موصوف، الگوی درخت تصمیم یکی از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی است که پیش‌بینی را در قالب برخی قوانین اگر - آنگاه توضیح

1. Decision Tree
2. k-nearest neighbors
3. Bayesian network
4. Artificial Neural Networks - ANN

می‌دهد. برای ایجاد الگوی درخت تصمیم، الگوریتم‌های متفاوتی وجود دارند که معروف‌ترین آن‌ها عبارت‌اند از: سی ۵^۱، کارت^۲، کوئست^۳ و چاید^۴ است. البته تمامی این الگوریتم‌ها برای دسته‌بندی ساختار مشابهی دارند و هدف همه آن‌ها، به دست آوردن درختی با کیفیت بالا و نرخ خطای کم برای دسته‌بندی داده‌هاست و تنها تفاوت آن‌ها در شیوه شاخه‌زدن و برش شاخه‌هاست. (رجبی، ۱۳۹۶، ۲۷-۲۶).

یافته‌ها

الف- شناخت و گردآوری داده‌های اطلاعاتی: در این پژوهش، نسبت به جمعیت و گردآوری داده‌های اطلاعاتی فراخوان‌های تجمعات غیرقانونی احصا شده از فضای مجازی که بخشی از آن در اسناد و مدارک سازمانی به همراه بازخوردهای میدانی آن‌ها و بخشی دیگر در بانک‌های اطلاعاتی غیرمتمرکز رده‌های عملیاتی پلیس فتا ناجا موجود بود، اقدام شد. تلاش محقق، منتج به تهیه و تدوین بانک اطلاعاتی جامع و متمرکزی به‌عنوان پیش‌نیاز اعمال الگوریتم‌های داده‌کاوی شد.

در ذیل چند نمونه از داده‌های اطلاعاتی موردنظر قبل از آماده‌سازی آورده شده است:

- گروه تلگرام (تعاونی مجاز وحدت بالای ۱۰۰ میلیون، ۳۶۳ عضو)، تجمع سپرده‌گذاران آرمان وحدت، زمان: چهارشنبه ۱ فروردین، ساعت ۲۰. مکان: خوزستان اهواز مقابل مؤسسه ملل شعبه کیان پارس.
- تلگرام (مشکلات بازار و بازارچه ۱۹۸۸۷ عضو)، اعتصاب مشکلات بازار و بازارچه/ زمان: چهارشنبه ۱۵ فروردین. مکان: تمام شهرهای مرزنشین.
- تویتر: بیانات رهبری در سخنرانی دیروز مشهد، واکنش‌های گسترده‌ای را در فضای مجازی به همراه داشته است. این اعتراضات با هشتک «دروغ‌گو نیستم» بیان شده است.

▪ سایت‌های خبری: جمعی از مردم و کشاورزان اصفهان در اعتراض به مدیریت غلط آب در ایران، نامه‌ای به دبیرکل سازمان ملل متحد، کمیسر عالی حقوق بشر سازمان ملل و اتحادیه بین‌المللی کشاورزان جوان تنظیم کرده و در حال جمع‌آوری امضا در فضای مجازی هستند.

▪ گروه واتس‌آپ (دیوان اهلنه، ۱۹۳ عضو): فراخوان تجمع جوانان عرب خوزستان با لباس عربی، زمان: پنجشنبه ۱۰ فروردین ۹۷ ساعت ۱۰ صبح. مکان: اهواز، چهارراه نادری.

▪ اینستاگرام (پیج Freedomforirani، فالوور ۵۴۱): فراخوان اعتراضات سراسری برای جلوگیری از فشار نیروهای امنیتی به مردم شریف خوزستان، زمان روزانه ساعت ۱۸. مکان سراسری میادین اصلی هر شهر.

همان‌طور که مشاهده می‌شود، داده‌های اطلاعاتی فوق بر اساس داده‌های اطلاعاتی ثبت‌شده در بانک‌های اطلاعاتی غیرمتمرکز مراکز عملیاتی پلیس فتا ناجا است که مؤلفه‌های مشترکی همچون استان، شهر، نشانی تجمع احتمالی، تاریخ تجمع، ساعت تجمع، عرصه، زیرعرصه، عنوان فراخوان، منبع انتشار فراخوان، بستر انتشار فراخوان در فضای مجازی، تعداد اعضای منبع انتشاردهنده، تعداد بازدید فراخوان، نگرش فکری انتشاردهنده یا انتشاردهندگان فراخوان‌ها وجود دارد که با متن کاوی تک‌تک فیلدهای مذکور، استخراج و به شکل داده‌های ساخت‌یافته جدولی تبدیل شدند.

ب- آماده‌سازی داده‌های اطلاعاتی: بخش اعظمی از زمان پژوهش صرف آماده‌سازی و پالایش داده‌های اولیه شد، چراکه در ابتدا می‌بایست نسبت به بررسی کلی داده‌ها و آماده‌سازی اولیه آن‌ها اقدام و سپس به‌عنوان داده‌های ورودی به نرم‌افزار داده‌کاوی از آن‌ها استفاده کرد. به همین علت پس از رفع نویز، پرکردن مقادیر گم‌شده در داده‌ها و حذف رکوردهای غیرضروری، داده‌های اطلاعاتی مناسب و یکنواخت در بانک اطلاعاتی موردنظر ایجاد و آماده‌سازی انجام شد. پس از بارگذاری داده‌ها در نرم‌افزار ریدماینر و قبل از مدل‌سازی داده‌ها، برای درک بهتر داده‌ها، بر اساس انتخاب هر یک از مؤلفه‌ها، فراوانی مقادیر آن‌ها به‌صورت نموداری و گرافیکی استخراج شد

(جدول ۱). همچنین در شکل (۴) فراوانی فراخوان‌ها بر اساس هر استان آمده است. بر اساس شکل ۴، استان‌های اصفهان، مازندران، خراسان رضوی و تهران بیشترین فراوانی فراخوان در کشور را دارند.

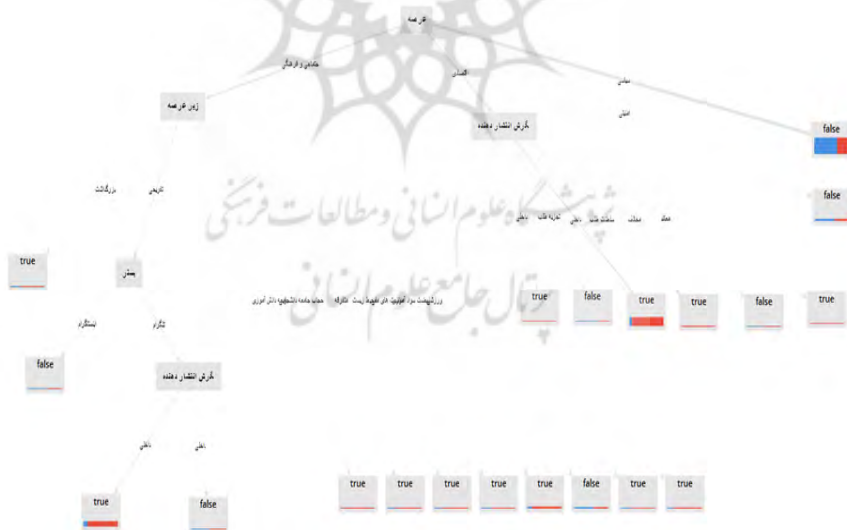
جدول ۱. آماده‌سازی داده‌ها

نام مؤلفه	مقادیر مؤلفه
استان	این مؤلفه بیانگر استان موردنظر برای برگزاری تجمع و جریان اعتراضی در فراخوان منتشره است. در داده‌های اطلاعاتی احصا شده، استان‌های ذیل به کثرت مدنظر کاربران فضای مجازی برای برگزاری تجمع و جریان اعتراضی طی فراخوان‌های منتشره بوده‌اند: اردبیل، اصفهان، چهارمحال و بختیاری، خراسان جنوبی، خراسان رضوی، زنجان، سمنان، تهران، فارس، قزوین، قم، کردستان، کرمان، مازندران، مرکزی و هرمزگان
شهر	این مؤلفه بیانگر شهر موردنظر برای برگزاری تجمع و جریان اعتراضی در فراخوان است: اصفهان، ابهر، ابوموسی، اراک، اردبیل، اصلاندوز، اقبالیه، الوند، پیربکران، آبیگ، آرادان، آلاشت، آمل، بابل، بابسر، بافت، بانه، بجنستان، بروجن، بزرگمهر، بستک، بشارگرد، بشرویه، بم، بندرعباس، بندرلنگه، بهارستان، بهشهر، بیجار، بیرجند، بيله سوار، پارس آباد، پارسیان، پاسارگاد، تایباد، تربت جام، تربت حیدریه، تنکابن، تهران، تیران، جاسک، جهرم، جویبار، جیرفت، چالوس، چناران، حاجی آباد، خاقانی، امیرکبیر، شهربابک، شهرضا، شهرکرد، شیراز، صدف، طبس، طرقله، عباس آباد، عنبرآباد، عنبران، فاریاب، فراهان، فرخ شهر، فردوس، فریدن، فریدونکنار، فسا، فلاورجان، فولادشهر، قائن، قروه، قزوین، قشم، قم، قهدریجان، قوچان، کاشمر، کازرون، کاشان، کامیاران، کرمان، کشکوئیه، کلاردشت، کمیجان، کهریزسنگ، کهنوج، کوثر، کوهناب، کوهپایه، کیاسر، گرمسار، گرمی، گلپایگان، گلشهر، گلوگاه، خانه اصفهان، خدابنده، خرمدره، خلخال، خمیر، خمین، خمینی شهر، خوراسگان، دامغان، درگز، دهگلان، دولت آباد، دیهوک، دیواندره، رابر، رامسر، رستم، کلا، رفسنجان، رودان، ریگان، زرنده، زرین شهر، زنجان، زهکوت، زیباشهر، زیرآب، زینیه، ساری، سامان، ساوه، سبزوار، سپاهان شهر، سرایان، سربیشه، سرعین، سروآباد، سقز، سمنان، سنندج، سه قلعه، سوادکوه زیرآب، سیرجان، تاکستان، قاین، قائم شهر، خوسف، داران
نشانی	این مؤلفه بیانگر مکان و نشانی موردنظر تجمع و جریان اعتراضی است؛ مانند تهران - مقابل مجلس
تاریخ	این مؤلفه بیانگر تاریخ ذکر شده در فراخوان برای ایجاد تجمع و جریان اعتراضی است؛ مانند یکشنبه ۱۹ فروردین
ساعت	این مؤلفه بیانگر ساعت برگزاری تجمع و جریان اعتراضی است که در فراخوان‌ها آمده است؛ مانند ساعت ۱۰ صبح
عرصه	این مؤلفه بیانگر عرصه‌های مختلف و حوزه‌های مختلف تجمع و جریان اعتراضی است که در

نام مؤلفه	مقادیر مؤلفه
	فراخوان‌ها آمده و به شرح ذیل است: سیاسی، اقتصادی، اجتماعی، فرهنگی و امنیتی
زیر عرصه	این مؤلفه، زیر عرصه‌های مختلف هر فراخوان، تجمع و یا جریان اعتراضی را بیان می‌کند، زیر عرصه‌ها در واقع زیرمجموعه هر عرصه و به شرح ذیل است: اعتراض به دولت، اعتراض به گرانی، احمدی‌نژادی‌ها، انقلابیون، بازاربان و اصناف، بزرگداشت، تفریحی، جامعه‌بازنشستگان، عامل خارجی، جامعه‌دانش آموزی، جامعه‌دانشجویی، جامعه فرهنگی، جامعه کارگری، جانبازان و ایثارگران، حجاب، درویش و اهل سنت، فعالان و زندانیان، قومیتی، سلامت، محیط‌زیست، مخالفان و معاندان، مسکن مهر، مناسبت‌های ملی، مؤسسات مالی، نهضت سوادآموزی، نیروهای مسلح، ورزشی و متفرقه
عنوان فراخوان	این مؤلفه بیانگر عناوین فراخوان‌ها و غالباً به‌عنوان تیتراژ فراخوان‌ها ذکر شده و در واقع «زیر عرصه» مخفف «عناوین فراخوان‌ها» نیز است.
منبع انتشار	این مؤلفه شامل نام کانال و گروهایی است که در فضای مجازی اقدام به انتشار فراخوان تجمع و جریان اعتراضی می‌کنند؛ مانند کانال ناسیونالیسم کوردی
بستر انتشار	این مؤلفه بیانگر نوع بستر انتشاردهنده فراخوان‌ها در فضای مجازی شامل انواع شبکه‌های اجتماعی و نرم‌افزارهای پیام‌رسان است؛ مانند تلگرام، اینستاگرام، فیس‌بوک، واتس‌آپ و وبسایت‌ها
تعداد اعضا	این مؤلفه بیانگر تعداد اعضا و تعداد دنبال‌کننده منبع انتشار است.
تعداد بازدید	این مؤلفه بیانگر تعداد بازدید هر فراخوان در فضای مجازی است.
نگرش انتشاردهنده	این مؤلفه بیانگر نگرش‌های فکری انتشاردهندگان فراخوان‌ها است؛ شامل داخلی، مخالف، معاند، سلطنت‌طلب، تجزیه‌طلب و فتنه
وضعیت وقوع تجمع و جریان اعتراضی	این مؤلفه بیانگر وضعیت وقوع تجمع و جریان اعتراضی است. به‌بیان‌دیگر آیا با انتشار این فراخوان، تجمع و جریان اعتراضی صورت گرفته است یا خیر؟ که به‌صورت صفر و یک نشان داده شده است، صفر به معنی عدم وقوع و یک به معنی وقوع است.

منبع انتشاردهنده، تعداد بازدید فراخوان، نگرش فکری انتشاردهندگان و وضعیت وقوع تجمع و جریان اعتراضی فراخوان‌ها که با متن کاوی، تک‌تک فیلدهای مذکور استخراج و به شکل داده‌های ساخت‌یافته جدولی تبدیل شدند.

۲. برای کشف ارتباط بین مؤلفه‌های شناسایی شده در جریان‌های اعتراضی، نسبت به انتخاب متدولوژی داده‌کاوی کریسپ و انجام گام‌های شش‌گانه آن بر روی داده‌های اطلاعاتی پژوهش اقدام شد. پس از انجام مراحل پیش‌پردازش داده‌ای و آماده‌سازی داده‌ها، الگوریتم‌های پیش‌بینانه درخت تصمیم بر روی داده‌های اطلاعاتی مذکور اعمال و درنهایت پس از بررسی نتایج داده‌کاوی، ارتباط تأثیرگذار برخی از مؤلفه‌های فراخوان‌های تجمعات غیرقانونی در وقوع جریان‌های اعتراضی شناسایی شد که عبارت‌اند از: نگرش فکری انتشاردهنده فراخوان‌ها، تعداد بازدید فراخوان‌ها، زیرعرصه فراخوان‌ها و تعداد اعضای منابع و نشانگاه‌ها که به ترتیب بیشترین تا کمترین تأثیرگذاری را بر وقوع تجمع‌ها و جریان‌های اعتراضی غیرقانونی داشته‌اند.



شکل ۶. درخت تصمیم

در شکل (۶) درخت تصمیم بعد از اجرای مدل در رییدماینر حاصل شده که می‌تواند یک تجمع و جریان اعتراضی را پیش‌بینی کند.

جدول ۲. فیلدهای تأثیرگذار در پیش‌بینی تجمعات و جریان‌های اعتراضی

Nodes	Importance	Importance	V4	V5
تعداد اعضا transformed	۰/۰۲۶	۰/۰۲۶۰	تعداد اعضا transformed	۰/۰۲۶۰
زیر عرصه transformed	۰/۱۳۳۶	۰/۱۳۳۶	زیر عرصه transformed	۰/۱۳۳۶
تعداد بازدید transformed	۰/۱۵۹۸	۰/۱۵۹۸	تعداد بازدید transformed	۰/۱۵۹۸
نگرش انتشاردهنده transformed	۰/۶۸۰۶	۰/۶۸۰۶	نگرش انتشاردهنده transformed	۰/۶۸۰۶

۳. به منظور تحلیل داده‌های اطلاعاتی پژوهش، پس از آماده‌سازی داده‌ها، با توجه به ماهیت کاربردی درخت تصمیم و الگوریتم‌های آن که از عملکرد مناسبی در دسته‌بندی، پیش‌بینی و بیان نتایج در قالب قوانین استخراجی داده‌کاوی برخوردارند، نسبت به ایجاد الگوریتم‌های پیش‌بینانه درخت تصمیم از جمله Decision Tree، ID3 و ترکیبی متا بر روی داده‌های اطلاعاتی پژوهش اقدام شد که نتایج میزان دقت الگوریتم‌های اعمال شده به شرح ذیل است.

جدول ۳. میزان دقت الگوریتم‌ها

ردیف	نام مدل داده‌کاوی	دقت برای داده‌های هدف
۱	مدل درخت تصمیم آیدی‌تری	۹۱/۳۹ درصد
۲	مدل ترکیبی متا	۸۷/۵۰ درصد
۳	مدل الگوریتم دیسیشن‌تری	۸۶/۴۱ درصد

۴. با توجه به نتایج داده‌کاوی فراخوان‌های تجمعات غیرقانونی، از بین الگوریتم‌های مختلفی که با بهره‌گیری از نرم‌افزار رییدماینر صورت پذیرفت، مدل درخت

تصمیم ID3 با دقت ۹۱/۳۹ درصد، بالاترین نتیجه را برای داده‌های موجود و پیش‌بینی وقوع جریان‌های اعتراضی به دست داد.

بحث و نتیجه‌گیری

هدف از اجرای این تحقیق، پیش‌بینی جریان‌های اعتراضی مبتنی بر داده‌کاوی رویدادهای آمیختی (فراخوان‌های تجمعات غیرقانونی) با استفاده از رصد فراخوان‌های موجود در فضای مجازی طی سال ۱۳۹۷ بود که این امر با به‌کارگیری بانک‌های اطلاعاتی غیرمتمرکز و استفاده از ابزارها و الگوریتم‌های داده‌کاوی مناسب انجام و توانست ماهیت پیچیده داده‌ها را مدل کرده و الگوی جرم را شناسایی و کشف کند. بدین طریق پلیس می‌تواند وقوع جرم را پیش‌بینی و با کنترل دقیق‌تر نیروها و آرایش نظامی آنان در منطقه هدف، از وقوع جرائم پیشگیری کند.

در این پژوهش، وضعیت انجام تجمع و وقوع جریان اعتراضی هر فراخوان به‌عنوان هدف داده‌کاوی مدنظر بود؛ بنابراین، بازخورد میدانی فراخوان‌هایی که دارای وقوع تجمع و جریان اعتراضی بوده‌اند از طریق بررسی آستاد، مدارک سازمانی و بانک‌های اطلاعاتی غیرمتمرکز رده‌های عملیاتی پلیس جمع‌آوری شد. در این مرحله، پس از احصای مؤلفه‌های اصلی بانک اطلاعاتی تهیه‌شده از فراخوان‌های تجمعات غیرقانونی که عبارت‌اند از: استان، شهر، نشانی تجمع، تاریخ تجمع، ساعت تجمع، عرصه، زیرعرصه، عنوان فراخوان، منبع انتشار فراخوان، بستر انتشار فراخوان در فضای مجازی، تعداد اعضای منبع انتشاردهنده، تعداد بازدید فراخوان، نگرش فکری انتشاردهندگان و وضعیت وقوع تجمع و جریان اعتراضی فراخوان‌ها، داده‌های اطلاعاتی پژوهش به شکل داده‌های ساخت‌یافته جدولی تبدیل شدند. در ادامه، برای کشف ارتباط بین مؤلفه‌های شناسایی شده فوق بر جریان‌های اعتراضی، نسبت به انتخاب متدولوژی داده‌کاوی کریسپ و انجام گام‌های شش‌گانه آن بر روی داده‌های اطلاعاتی پژوهش اعم از آشنایی با حوزه رصد فراخوان‌ها و جریان‌ات، شناخت داده‌های مربوطه، آماده‌سازی داده‌ها، مدل‌سازی، ارزیابی و پیاده‌سازی مدل پیش‌بینی‌کننده پژوهش اقدام شد. در مرحله پیش‌پردازش داده‌ای، پس از آماده‌سازی داده‌ها و بارگذاری در نرم‌افزار

رپیدمانیر، فراوانی تک‌تک مؤلفه‌های داده‌های اطلاعاتی به‌صورت نموداری و گرافیکی نمایش داده شد و در ادامه برای تحلیل داده‌ها به‌منظور پیش‌بینی جریان‌های اعتراضی، طی ایجاد الگوریتم‌های پیش‌بینانه درخت تصمیم از جمله Decision Tree، ID3 و ترکیبی متا بر روی داده‌های اطلاعاتی پژوهش و سپس ارزیابی به‌عمل آمده از میزان دقت الگوریتم‌های اعمال‌شده، ارتباط بین مؤلفه‌های فراخوان‌های تجمعات غیرقانونی در وقوع جریان‌های اعتراضی شناسایی شد. شایان ذکر است در بررسی به‌عمل آمده از نتایج داده‌کاوی، مؤلفه‌های تأثیرگذار بر روی داده‌های این پژوهش عبارت‌اند از: نگرش فکری انتشاردهنده فراخوان‌ها، تعداد بازدید فراخوان‌ها، زیرعرصه فراخوان‌ها و تعداد اعضاء منابع و نشانگاه‌ها که به ترتیب بیشترین تا کمترین تأثیرگذاری را بر روی انجام تجمع‌ها و جریان‌های اعتراضی غیرقانونی داشته‌اند.

همچنین با توجه به نتایج داده‌کاوی فراخوان‌های تجمعات غیرقانونی، از بین الگوریتم‌های مختلفی که بررسی شد، از مدل درخت تصمیم ساخته شده با الگوریتم ID3 با توجه به میزان بالای صحت مدل به میزان ۹۱/۳۹ درصد، می‌توان برای پیش‌بینی جریان‌های اعتراضی در فضای مجازی بر مبنای مؤلفه‌های اولویت‌دار شناسایی شده فراخوان‌های تجمعات غیرقانونی با احتمال وقوع آن در فضای حقیقی استفاده کرد.

به‌طورکلی، نتیجه این تحقیق می‌تواند تشخیص و پیش‌بینی جریان‌های اعتراضی در سطح کشور را ارتقا بخشد و احساس امنیت در بین شهروندان را افزایش دهد. به‌علاوه از طریق تشخیص و پیش‌بینی درست جریان‌های اعتراضی در سطح کشور می‌توان از هدر رفت و اعزام بعضاً غیرضروری نیروها و مأموران در پیشگیری از وقوع تجمعات و جریان‌های اعتراضی احتمالی و یا مقابله با آنان، جلوگیری و به‌صورت هدفمند اقدام کرد. بدین طریق پلیس می‌تواند وقوع جرم را پیش‌بینی و با کنترل دقیق‌تر نیروها و آرایش نظامی آنان در منطقه هدف، از وقوع جرائم بیشتر ممانعت کند.

پیشنهادها

«مؤلفه‌ها و مشخصه‌های مؤثر شناسایی شده فراخوان‌ها بر وقوع جریان‌های اعتراضی» و «اطلاعات بازخوردهای میدانی وقوع هر فراخوان و رویداد در فضای حقیقی» در بانک

اطلاعاتی فراخوان‌های تجمعات غیرقانونی پلیس فتا ناجا به صورت کامل ثبت و ضبط شده و بانک اطلاعاتی مذکور بر مبنای همین مؤلفه‌ها اصلاح شود. به علاوه دقت در ثبت اطلاعات درست و واقعی فراخوان‌ها، تا حد بسیار زیادی می‌تواند در تشخیص و پیش‌بینی جریان‌های اعتراضی در سطح کشور مؤثر واقع شود؛ بنابراین، به منظور کاهش نویز و خطا در بانک اطلاعاتی مربوطه، ضمن ایجاد یک بانک اطلاعاتی پشتیبان، آموزش‌های لازم تخصصی به کاربران ارائه شود و نظارت جدی‌تری بر ورود صحیح اطلاعات در بانک اطلاعاتی موصوف صورت پذیرد.

سپاسگزاری

پژوهشگر بر خود لازم می‌داند از مشارکت کنندگان در مصاحبه و سایر استادان برای همکاری‌های بی‌بدیل‌شان، تقدیر و تشکر کند.

منابع

ابراهیمی، مجیب؛ میرروشندل، سیدابوالقاسم؛ آقایی، جان‌احمد. (۱۳۹۴)، جامعیت بخشی به مجموعه داده جرائم به منظور پیش‌بینی و شناسایی جرائم با استفاده از فنون داده‌کاوی، فصلنامه صنایع الکترونیک، ۶(۴)، ۱۱-۶.

<https://www.sid.ir/fa/journal/ViewPaper.aspx?ID=249690>

پلیس فتا ناجا. (۱۳۹۶). بولتن اطلاعاتی پلیس فتا ناجا. پلیس فتا ناجا.

پلیس فتا ناجا. (۱۳۹۸). بولتن اطلاعاتی پلیس فتا ناجا. پلیس فتا ناجا.

پولادی، رضا. (۱۳۹۶). داده‌کاوی ارتباط میان جرائم و ویژگی‌های شغلی کارکنان در ناجا.

پایان‌نامه کارشناسی ارشد مدیریت فناوری اطلاعات، دانشگاه علوم انتظامی امین.

https://jpas.ihu.ac.ir/article_205038.html

دهخدا، علی‌اکبر. (۱۳۱۹). لغت‌نامه آنلاین دهخدا. <https://www.vajehyab.com>

ربیعی، مهدی. (۱۳۹۲). تحلیل جرائم شبکه‌ای با استفاده از داده‌کاوی. پایان‌نامه کارشناسی

ارشد مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد زنجان.

رجبی، مصطفی. (۱۳۹۶). تحلیل جرائم سایبری با استفاده از داده‌کاوی. پایان‌نامه کارشناسی

ارشد رشته اطلاعات، دانشگاه علوم انتظامی امین.

شاه‌محمدی، غلامرضا؛ رجیبی، مصطفی. (۱۳۹۷)، تحلیل جرائم اخلاقی فضای سایبر با رویکرد

داده‌کاوی. فصلنامه پژوهش‌های دانش انتظامی، ۲۰(۴)، ۶۳-۹۱.

http://pok.jrl.police.ir/article_20172.html

شرف‌الدین، سیدحسین. (۱۳۹۵). جریان‌شناسی اجتماعی با رویکردی جامعه‌شناختی. پایگاه

اطلاعرسانی دکتر سید حسین شرف‌الدین.

<http://pajuhesh.irc.ir/product/article/show/id/243881>

شریفی، رضا. (۱۳۹۸)، بررسی ابعاد روان‌شناختی خشونت‌گرایی جمعیت اغتشاشگر و

روش‌های تأثیرگذاری بر آن. فصلنامه مطالعات بسیج، ۱۲(۴۵)، ۶-۳۹.

<https://www.noormags.ir/view/fa/articlepage/552847>

عالی‌پور، حسن. (۱۳۹۶). مدیریت حقوقی اجتماعات اعتراض‌آمیز، فصلنامه مطالعات

راهبردی، ۱۰(۳۶)، ۳۴۹-۷۹. <https://www.magiran.com/paper/475056>

عباسی، سعید. (۱۳۹۵)، بررسی عوامل مؤثر بر تصادفات با استفاده از روش‌های داده‌کاوی.

پایان‌نامه کارشناسی ارشد رشته اطلاعات، دانشگاه علوم انتظامی امین.

قاسم‌زاده، حمیدرضا. (۱۳۹۸)، پیش‌بینی تجمعات غیرقانونی مبتنی بر داده‌کاوی. پایان‌نامه

کارشناسی ارشد رشته اطلاعات، دانشگاه علوم انتظامی امین.

مظاهری حسین‌آبادی، الهام؛ طالب‌لو، علیرضا؛ رضاییان، علی. (۱۳۹۴). شناسایی گره‌های

قدرت در شبکه‌های اجتماعی به کمک داده‌کاوی. پژوهش‌های مدیریت در ایران،

۲(۱۹)، ۱۶۱-۱۸۲.

<https://www.sid.ir/fa/journal/ViewPaper.aspx?id=273748>

Annabathula, R. (2007). A Web-based Tool for Analysis of Crime Laboratory Data. West Virginia University.

Armstrong, J. S. (2017). Principles of forecasting: a handbook for researchers and practitioners (Vol. 30). Springer Science & Business Media.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. (2010). CRISP-DM Step-Data Mining Guide, SPSS In CRISPMWP.

Dunham, M.H. (2012). Data Mining, Introductory and Advanced Topics. Prentice Hall.

- Farsi, M., Daneshkhah, A., Far, A. H., Chatrabgoun, O., Montasari, R. (2018). Crime data mining, threat analysis and prediction. In *Cyber Criminology* (pp. 183-202). Springer, Cham.
- Joshi, J. D. (2018). Data Mining for Prevention of Crimes. *International Conference on Human Interface and the Management of Information*, 705-717.
- Khan, M. A., Pradhan, S. K., Fatima, H. (2017). Applying Data Mining Techniques in Cyber Crimes. In *Anti-Cyber Crimes (ICACC), 2nd Inter. Conf. on* (pp. 213-216). IEEE.
- Shearer, C. (2000) The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, 5, 13-22.
- Sahu, N. S., Darokar, S. (2017). Data Mining Techniques to Clustering Cyber Crime Data. *International Education and Research Journal*, 3(7), 119-120.
<http://ierj.in/journal/index.php/ierj/article/view/1288/1238>
- Taha, K, Yoo, P. D. (2017). Using the Spanning Tree of a Criminal Network for Identifying Its Leaders. *IEEE Transactions on Information Forensics and Security*, 12(2), 445-453. DOI:10.1109/TIFS.2016.2622226.