

کاربرد داده کاوی در شناسایی و کشف الگوهای پنهان جرم سرقت

حمیدرضا اسکندری^۱

سمیه علیزاده^۲

پروانه کاظمی^۳

تاریخ پذیرش: ۹۱/۴/۲۹

تاریخ وصول: ۹۱/۲/۸

چکیده

پیشگیری از جرم همواره یکی از موضوعات اساسی و مهم در زندگی بشری بوده که در طول تاریخ به شیوه‌های مختلف اعمال گردیده است. با توجه به گسترش فناوری‌های اطلاعات و ارتباط و راه‌اندازی سیستم‌های جامع اطلاعاتی در نیروی انتظامی و ثبت اطلاعات مجرمان در بانک‌های اطلاعاتی، کاربرد تکنیک‌های داده کاوی و کشف دانش به منظور تحلیل و ریشه‌یابی جرایم از جمله سرقت یکی از ضرورت‌های پلیس ایران و دستگاه‌های قضایی می‌باشد. هدف اصلی این تحقیق توسعه و ارائه یک مدل داده کاوی است که با به کارگیری بانک‌های اطلاعاتی جرایم موجود و استفاده از ابزارها و الگوریتم‌های داده کاوی بتوان الگوهای جرم را شناسایی و کشف نمود تا به این طریق پلیس بتواند وقوع جرم را پیش بینی کرده و با کنترل دقیق‌تر نیروها و آرایش نظامی آنان در منطقه جرم، از وقوع جرایم پیشگیری نماید. در مدل ارائه شده از دو تکنیک قوانین تلازمی با الگوریتم احتمال پیشین و خوشه‌بندی با الگوریتم کای میانگین برای استخراج الگوها از یک پایگاه داده با بیش از صد هزار رکورد سرقت استفاده شده است.

واژگان کلیدی: داده کاوی، قوانین تلازمی، الگوهای جرم، فرایند داده کاوی.

۱- عضو هیئت علمی بخش مهندسی صنایع، دانشگاه تربیت مدرس

۲- عضو هیئت علمی دانشکده مهندسی صنایع، گروه فناوری اطلاعات، دانشگاه خواجه نصیرالدین طوسی

۳- کارشناس ارشد مهندسی فناوری اطلاعات، دانشگاه تربیت مدرس



مقدمه

جوامع، چندین سال است که با معضل جرم و چگونگی کنترل آن درگیر بوده‌اند که تمرکز و خط‌مشی اصلی آنان در کنترل جرم، تنها افزایش و حرفه‌ای‌تر کردن سیستم پلیسی و قضایی بوده است. عملکرد دولت‌ها نیز در برخورد با جرم، اختصاص حجم بیشتری از پول به سازمان‌های قضایی و پلیسی بوده که باعث افزایش سریع بودجه و کارکنان در آن مقطع گردیده است.

ایجاد امنیت و آرامش در جامعه تنها با توسل به شیوه‌های کیفری پس از وقوع جرم محقق نمی‌شود، بلکه دولت وظیفه دارد با در پیش گرفتن راه کارهایی، قبل از وقوع بزه، در از بین بردن شرایط تحقق آن تلاش نماید.

شاید بهترین مدل برای درک مقوله پیشگیری از وقوع جرم، دیدگاه کاهش فرصت ارتکاب جرم در جامعه است. تئوریسین‌های این مدل پیشنهاد دارند که رفتارهای تبهکارانه ناشی از وجود فرصتی مناسب جهت ارتکاب جرم در یک مکان و زمان خاص می‌باشد و از بین بردن یا کاهش این فرصت‌ها باید به کاهش جرایم در آن مکان منجر گردد. در این مدل، روشن‌ترین دیدگاه این است که چگونه فرصت، مهم‌ترین عامل در ارتکاب و یا پیشگیری از وقوع جرم است. این مدل پیشنهاد می‌کند که جرم بیشتر در مناطقی روی می‌دهد که مجرمان به دلیل وجود هدف و قربانی مناسب، تحریک شده و نگرهبانی جهت پیشگیری از ارتکاب جرم وجود ندارد.

امروزه فناوری اطلاعات با توجه به فاکتورهای اصلی و متغیرهای اساسی در فرایند تصمیم‌گیری با ارائه سیستم‌های مبتنی بر تئوری‌های فوق، می‌تواند نقش مؤثری در سازمان‌های قضایی و پلیسی ایفا نماید. در بین فناوری‌های معرفی شده داده کاوی به عنوان یک ابزار قدرتمند در تحلیل داده‌های جرم نقش مؤثری در کشف الگو و دانش از پایگاه داده جرم در راستای فرایند تصمیم‌گیری به منظور کاهش فرصت ارتکاب جرم و کنترل جرایم دارد.



بدون تردید شرایط اجتماعی بشر، رویارویی با پدیده‌ای به نام جرم را اجتناب ناپذیر ساخته و انسان همواره نیازمند دانش تحلیل جرم است. تحلیل جرم عبارت است از به کارگیری یک شیوه نظام‌مند جهت شناسایی، کشف و پیش‌بینی جرایم (بیابانی، ۱۳۸۶). در ادامه به بیان برخی از مفاهیم اولیه داده کاوی و برخی تکنیک‌های آن و همچنین اصول اولیه تحلیل جرم پرداخته شده است. سپس به برخی تجربیات جهانی در زمینه داده کاوی جرم پرداخته و پس از آن نحوه استفاده از تکنیک‌های داده کاوی همچون قوانین تلازمی^۱ و خوشه‌بندی جهت استخراج الگوهای رفتاری مجرمان به شیوه‌ای کاربردی مورد بررسی قرار گرفته است.

قوانین تلازمی در داده کاوی

کشف دانش در پایگاه داده فرایند شناسایی درست، ساده، مفید و در نهایت الگوها و مدل‌های قابل فهم در داده‌ها است که داده کاوی مرحله‌ای از فرایند کشف دانش است و شامل الگوریتم‌های مخصوص به خود است، به طوری که، تحت محدودیت‌های مؤثر محاسباتی قابل قبول، الگوها و مدل‌ها را در داده‌ها کشف می‌کند. در اینجا منظور از الگوی مفید، مدلی در داده‌ها است که ارتباط میان زیر مجموعه‌ای از داده‌ها را توصیف می‌کند در حالی که معتبر، ساده، قابل فهم و جدید است. در واقع داده کاوی یکی از مهم‌ترین روش‌هایی است که به وسیله آن الگوهای مفید در درون داده‌ها با حداقل دخالت کاربران شناخته و اطلاعاتی را در اختیار کاربران و تحلیل‌گران قرار می‌دهند تا بر اساس آنها تصمیمات مهم و حیاتی در سازمان‌ها اتخاذ شود.

قوانین تلازمی یکی از مهم‌ترین شیوه‌های داده کاوی است. شاید این روش را بتوان به عنوان رایج‌ترین شکل روش‌های کشف الگو در سیستم‌های یادگیری غیرنظارتی نام برد. این شیوه داده کاوی، بیشترین شباهت را به رفتار مردم در هنگام شروع فراگیری داده کاوی دارد، یعنی «جست و جوی طلا درون پایگاه داده‌ای بسیار بزرگ». طلا در این مورد یعنی قاعده‌ای که چیزی را درباره پایگاه داده می‌گوید که تاکنون قابل درک نبود و اکنون که

کشف شده بسیار مورد توجه می باشد. این روش‌ها، تمامی الگوهای ممکن دلخواه را درون پایگاه داده پیدا می کنند. قوانین کشف شده از یک سو می تواند نقطه قوت باشد زیرا هیچ چیزی نیست که کاویده نشده باشد، از طرفی دیگر یک نقطه ضعف است زیرا کاربر با انبوهی از قوانین روبرو خواهد بود که تحلیل کارایی آنها امری زمان‌بر و پرهزینه می باشد. معیارهایی مانند ضریب پشتیبان و ضریب اطمینان برای کشف قوانین با کیفیت بیشتر استفاده می شود. نمای یک قاعده تلازمی کننده به شکل زیر است (Han and Kamber, ۲۰۰۱).

$$A \Rightarrow B \text{ [ضریب اطمینان، ضریب پشتیبان]}$$

که در آن AB ، A و B زیرمجموعه‌ای از مجموعه اقلام پرتکرار^۱ هستند.

ضریب پشتیبان^۲: درصد مجموعه تراکنش‌هایی است که شامل هر دو زیرمجموعه A و B باشد.

ضریب اطمینان^۳: نسبت وابستگی یک زیرمجموعه اقلام پرتکرار را به زیرمجموعه اقلام پرتکرار دیگر بیان می کند و با استفاده از رابطه زیر محاسبه می شود:

$$(A) \text{ ضریب پشتیبان} / (A, B) \text{ ضریب پشتیبان} = (A \Rightarrow B) \text{ ضریب اطمینان}$$

الگوریتم کای میانگین^۴

الگوریتم کای میانگین از جمله ساده‌ترین و رایج‌ترین الگوریتم‌هایی است که برای خوشه‌بندی مجموعه داده‌های بزرگ به کار می‌رود. این الگوریتم که از معیار مربع خطا برای خوشه‌بندی استفاده می‌کند با یک تقسیم‌بندی اولیه و تصادفی شروع به تخصیص دادن نمونه‌ها به خوشه‌ها می‌کند. بر اساس معیار شباهت^۵ بین نمونه‌ها و خوشه‌ها تا زمانی

1. Item Set
2. Support
3. Confidence
4. K-Means
5. Similarity criterion



که معیار همگرایی برآورده شود فرایند تخصیص نمونه‌ها و محاسبه میانگین فاصله را تکرار می‌کند. طبیعی است در صورتی که نتوان هیچ نمونه‌ای را از خوشه‌ای جدا کرد و به خوشه‌ای دیگر تخصیص داد به گونه‌ای که مربع خطا کاهش یابد، الگوریتم متوقف می‌شود.

ارزیابی کیفیت خوشه‌بندی

تاکنون معیارهای زیادی برای اعتبار سنجی جواب خوشه‌بندی ارائه شده‌اند. میلیگان و همکاران و پال و بیسواس چندین معیار برای ارزیابی کیفیت جواب روش‌های مختلف خوشه‌بندی پیشنهاد کرده‌اند. در اینجا از شاخص دیویس بولدین که یکی از مهم‌ترین شاخص‌های ارزیابی نتایج خوشه‌بندی به شمار می‌رود برای ارزیابی نتایج جواب‌های خوشه‌بندی استفاده می‌شود.

شاخص دیویس - بولدین به شکل زیر تعریف می‌شود:

$$DB_{nc} = \frac{1}{n_c} \sum_{i=1}^{n_c} R_i ,$$

$$R_i = \max_{i=1, \dots, n_c, i \neq j} R_{ij} ,$$

در تعریف بالا روشن است که DB_{nc} میانگین شباهت بین هر خوشه $C(i=1, \dots, n)$ و شبیه‌ترین خوشه به آن می‌باشد. مطلوبست که برای هر خوشه کمترین شباهت ممکن را با دیگر خوشه‌ها داشته باشد. بنابراین خوشه‌بندی که DB_{nc} را کمینه کند، از دید این شاخص خوشه‌بندی بهینه است.

شاخص DB_{nc} هم در معادلاتش به طور مستقیم از تعداد خوشه‌ها استفاده نمی‌کند. اما همانند شاخص دان می‌توان با ترسیم تابع DB_{nc} برای تعداد خوشه‌های مختلف، تعداد بهینه خوشه‌ها را در محل کمینه شدن DB_{nc} پیدا نمود (Han and Kamber, 2001).



کاربرد داده‌کاوی در جرایم

حجم انبوه داده‌ها و اطلاعات مرتبط با جرایم در اداره پلیس از یک سو و پیچیدگی ارتباطات میان این جرایم از سوی دیگر، موجب می‌گردد تا روش‌های سنتی تحلیل جرم که غالباً توسط کارآگاهان زبده و با تجربه و بازرسان صحنه جرم به کار گرفته می‌شوند، اولاً نیازمند صرف زمان و هزینه انسانی بسیار زیادی باشند و ثانیاً به دلیل میزان دخالت بالای عامل انسانی در تصمیم‌گیری‌ها قادر به لحاظ کردن تمام فاکتورهای تأثیرگذار در یک جرم و ارتباطات بین آنها نباشند. چنین شرایطی ضرورت به کارگیری یک روش نظام‌مند مبتنی بر فناوری اطلاعات جهت کشف و حل جرایم را بیش از پیش نمایان می‌سازد. ماهیت پیچیده داده‌های مرتبط با جرم و بزه‌کاری و روابط نامحسوس میان این داده‌ها موجب مقبولیت روز افزون استفاده از دانش داده‌کاوی در میان جرم‌شناسان و تحلیل‌گران جرم شده است. در واقع دانش حاصل از اعمال روش‌های داده‌کاوی در حوزه تحلیل جرم بستر مناسبی را برای پشتیبانی اطلاعاتی فرماندهان و مدیران به منظور انجام فعالیت‌های آتی پلیس فراهم می‌آورد (بیابانی، ۱۳۸۶). همچنان که کولین مک کیو^۱ به عنوان مدیر پروژه واحد تحلیل جرم در سازمان پلیس ویرجینیا عنوان می‌نماید: «داده‌کاوی زمانی که در تحلیل جرم تاکتیکی به کار برده می‌شود یک ابزار اکتشاف دانش است که می‌تواند مجموعه داده‌های جامع را با سرعت بررسی کند و یک آرایه بیکران از متغیرها تهیه کند، که این موضوع به مراتب برتر است از آنچه که یک تحلیل‌گر به تنهایی یا حتی یک گروه تحلیلی یا گروه رزمی مشترک با دقت و درستی بررسی می‌کند» (۲۰۰۲ Oatley et al).

در برخی کشورها از روش‌ها و تکنیک‌های داده‌کاوی به منظور پیش‌بینی و پیشگیری از وقوع جرم استفاده کرده‌اند که در ادامه به اختصار بیان می‌گردد. مون و همکاران (۲۰۱۰) از رگرسیون برای پیش‌بینی جرایم رایانه‌ای استفاده کردند. براساس نتایج به دست آمده، میزان ساعات استفاده از رایانه و عضویت در گروه‌ها و شبکه‌های اینترنتی میزان جرایم رایانه‌ای را افزایش داده و به عنوان متغیرهای اصلی پیش‌بینی‌کننده میزان

1. Colleen McCue



جرایم معرفی شدند. چونگ و همکاران (۲۰۰۶) نیز با معرفی داده کاوی به عنوان یکی از کاراترین ابزارها در جرایم رایانه‌ای، مروری بر کاربرد تکنیک‌های داده‌کاوی در این حوزه نمودند. در این تحقیق یک مطالعه موردی در تایوان مورد بررسی قرار گرفته و در مورد جرایم رایانه‌ای و مشکلات مربوطه بحث شده است. در نهایت هم پیشنهاداتی برای مقابله با جرایم رایانه‌ای ارائه شده است.

لی و همکاران (۲۰۱۰) یک مدل پشتیبان تصمیم بر اساس تکنیک فازی SOM برای تشخیص و تحلیل الگوها و روندهای موجود در وقوع جرایم ارائه نمودند. این مدل در داده‌های مربوط به پلیس بین‌الملل تایوان پیاده سازی شده است. نتایج به دست آمده برای مدیران نیروی پلیس در تدوین راهبردهای جلوگیری و پیشگیری از جرم و جنایت مفید واقع شده است.

آزمایشگاه پلیس ایالت ویرجینای غربی از یک ابزار نرم‌افزاری به نام (سیستم مدیریت اطلاعات جنایی)^۱ استفاده کرده است که این ابزار با به کارگیری اصول آمار و الگوریتم‌های داده کاوی می‌تواند داده‌های جرم را تحلیل نماید و نتایج منطقی را به دست آورد. پلیس نورثامپتونشر^۲ از داده‌های قانونی (اثرانگشت‌ها یا DNA) و جرم برای این طرح آزمایشی استفاده کرد. یافته‌ها نشان داد که مأموران تحقیق می‌توانند مطابق توانایی‌شان نسبت به جمع آوری DNA و اثر انگشت‌ها از صحنه جرم اقدام کنند. هم چنین توانایی آنها را در پیش-بینی اینکه کدام یک از صحنه‌های جرم بهترین فرصت جمع‌آوری نمونه‌های قانونی را ارائه خواهد کرد، افزایش داد در حالی که بدون آن و با توانایی فردی آنها ممکن نبود (Annabathula, ۲۰۰۷).

یک گروه تحقیقاتی متمرکز در دانشگاه ساندرلند (مرکز سیستم‌های قابل تطبیق) مأمور شد تا یک نرم افزار تحت عنوان «نرم افزار هوشمند برای تصمیم گیران»^۳ آن سازمان پیاده‌سازی کند در نتیجه در آن پروژه‌ها ابزارهایی جهت استفاده مستقیم از فرایندهای داده کاوی توسعه دادند. نیروهای پلیس به بررسی پدیده‌های تکرار قربانی علاقه‌مند بودند.

1. FIMS
2. Northamptonshire
3. Smart software for decision makers

مفهوم دوباره قربانی شدن در ابتدا توسط اسپارکس^۱ بیان شد و به این معناست که مکان‌هایی که در آن یک مرتبه جرم اتفاق افتاده باشد به طور نامتناسب احتمال دارد که دوباره آن جرم اتفاق بیفتد. برای مثال احتمال دوباره دزدیده شدن بعد از ۲۸ روز از اولین دزدی بود که می‌بایست تا سطح طبیعی، (پس از ۶ ماه) کاهش یابد. پس از آماده سازی نرم‌افزارهای Ewart، Inglis و Wilbert کم شدن فاصله زمانی بین دزدی‌های پی در پی و دوباره قربانی شدن در یک ملک را ثابت کردند (Oatley et al, ۲۰۰۶).

پس از حمله تروریستی ۱۱ سپتامبر سازمان‌های سیا، اف بی آی و دیگر آژانس‌های فدرال تصمیم گرفتند اطلاعات داخلی و خارجی مربوط به حوزه امنیت را جمع آوری کنند تا بتوانند از حملات تروریستی جلوگیری کنند. این تلاش‌ها موجب ایجاد انگیزه در مقام‌های محلی گردید تا به صورت دقیق‌تر جرایم قضایی حوزه خود را کنترل کنند. چالش اصلی تمام مجریان قانون و سازمان‌های گردآوری اطلاعات که با آن مواجه شدند دقت و مؤثر بودن میزان فزاینده تحلیل داده جرم است. به عنوان مثال حل نمودن توطئه‌های پیچیده اغلب مشکل هستند زیرا اطلاعات مظنونان می‌تواند از نظر جغرافیایی و گستردگی در دوره‌های زمانی طولانی متفاوت باشد. همچنین تشخیص جرایم مجازی می‌تواند سخت باشد (Annabathula, ۲۰۰۷).

پروژه اور^۲ در سال ۲۰۰۰، پلیس میدلند غربی را درگیر کرد هدف اصلی از این سیستم کمک به اداره و کنترل میزان جرم دزدی از منازل مسکونی (BDH)^۳ با بکارگیری سیستم‌های پشتیبان تصمیم بود (Annabathula, ۲۰۰۷).

پروژه کاپلینک^۴ توسط محققان دانشگاه آریزونا در همکاری با سازمان پلیس توسکان و فوتنی بعد از سال ۱۹۹۷ اجرا گردید. هدف از اجرای این پروژه ارائه یک چارچوب کلی برای شناسایی و کشف انواع جرایم به کمک تکنیک‌های داده کاوی جرم بود. هر دسته یک مجموعه از تکنیک‌ها را برای استفاده در تحلیل برخی از جرایم نشان می‌دهد.

1. Sparks

2. Over

3. Burglary from a dwelling house

4. <http://ai.bpa.arizona.edu/coplinc>



آنان اعتقاد دارند که چارچوب آنها قابلیت اجرا عمومی را در تحلیل جرم و آنالیز هوشمند دارد، زیرا تمام انواع جرایم اصلی مانند تکنیک‌های سنتی و تکنیک‌های جدید داده‌کاوی هوشمند را در برمی‌گیرد (Chung et al, ۲۰۰۴).

معمولاً پایگاه داده دادگستری جنایی تنها داده‌های ساخت یافته که در فیلدهای از پیش تعریف شده گنجانیده شده‌اند، را ثبت می‌کند. اولین وظیفه داده‌کاوی استخراج موجودیت‌های مشخص از گزارش‌های تشریحی (داستانی) پلیس است که تحلیل آنها به وسیله به کارگیری تکنیک‌های اتوماتیک مشکل است. سازمان پلیس فینکس^۱ امریکا یک نسخه اصلاح شده از سیستم استخراج کننده موجودیت AI که یک پردازش سه مرحله‌ای در شناسایی نام‌های اشخاص، مکان‌ها و سازمان‌ها در یک سند به کار می‌برد، در اختیار گرفتند. مضمونان اغلب اسامی، تاریخ تولد یا آدرس‌ها را به صورت اشتباه به افسران پلیس می‌دهند. بنابراین ورودی‌های گوناگونی در پایگاه داده‌ها دارند که این موضوع تشخیص اطلاعات درست یک مضمون، گزارش حوادث گذشته‌ای که آن شخص در آن درگیر شده و یا گیر افتاده، را برای افسران مشکل می‌کند. افسران اجرای قانون سازمان پلیس توسکان توانستند رکوردهای هویت موجود را که مربوط به یک مضمون در پایگاه داده است را با تکنیک اکتشاف خصوصیات فریبنده^۲ بازیابی کنند (Annabathula, ۲۰۰۷).

تبهکاران اغلب شبکه‌هایی را توسعه می‌دهند که از گروه‌ها و یا تیم‌هایی تشکیل می‌شوند که فعالیت غیر قانونی گوناگون را انجام می‌دهند. سازمان پلیس توسان امریکا سعی نمود با گسترش استفاده از تکنیک‌های داده‌کاوی در شناسایی زیرگروه‌ها و عضوهای کلیدی در این قبیل شبکه‌ها و سپس مطالعه الگوهای تعاملات، استراتژی مؤثری را برای مختل کردن شبکه‌ها توسعه دهد (Seifert, ۲۰۰۴).

روش شناسایی تحقیق

در این بخش روش جدیدی ارائه شده است که کمک می‌کند تا از داده‌های مربوطه به دانش مورد نظر برسیم. از آنجایی که CRISP-DM^۳ یک استاندارد جهانی است که در

1. phonix

2. detective – identity detection

3. Cross Industry Standard Process for Data Mining

پروژه‌های داده کاوی مورد استفاده قرار گرفته است بنابراین ساختار اجرایی تحقیق بر همین اساس مورد بررسی قرار گرفته که در این بخش، به صورت عملیات تجزیه و تحلیل که بر روی مجموعه داده‌های واقعی سازمان پلیس انجام شده است، بیان می‌گردند. مدل جدید ارائه شده براساس استاندارد CRISP-DM مطابق شکل ۱ ارائه می‌شود.

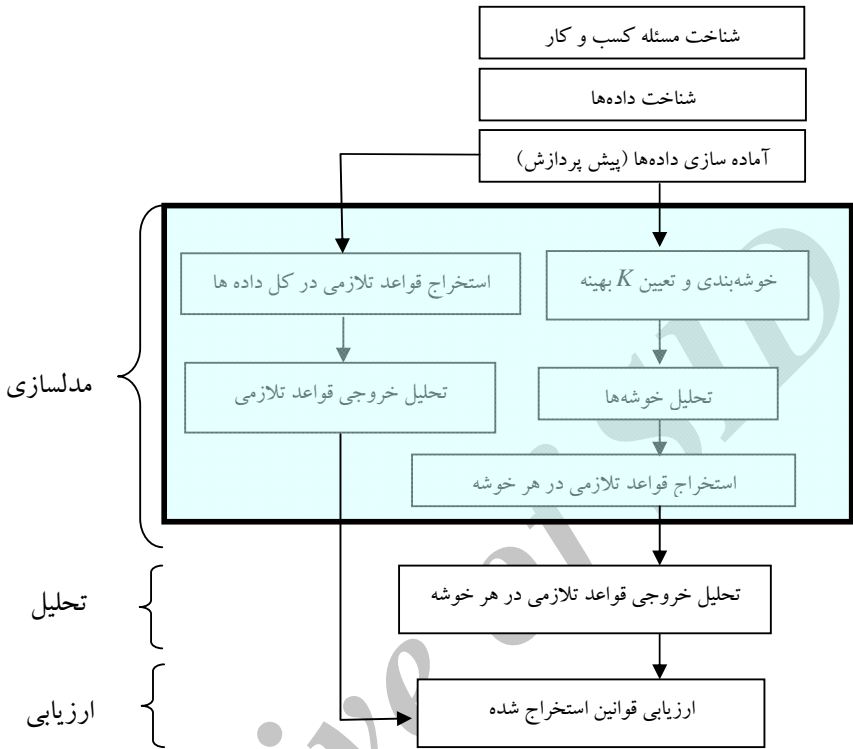
شناخت مسئله کسب و کار: در این فاز از فرایند، ابتدا اهداف اصلی کسب و کار تعیین گردید که اصلی‌ترین هدف کسب و کار (پلیس آگاهی تهران بزرگ) در این مقاله، پیدا کردن روابط پنهان در پایگاه داده‌های مربوطه به منظور شناسایی و کشف قواعد و الگوهای جرم سرقت است. این کار براساس کشف الگوهای پنهان موجود بین خصیصه‌های افراد سارق و پرونده‌های سرقت می‌باشد، که اطلاعات مربوط به آنها در سیستم نگهداری شده است.

شناخت داده‌ها^۱: جامعه آماری تحقیق متشکل از داده‌های مربوط به ۱۰۲۲۷۳ نفر سارق می‌باشد که در طی سال‌های ۱۳۸۲ تا اواخر سال ۱۳۸۷ مرتکب جرم سرقت شده و در سیستم، اطلاعات مربوط به جرم آنان ثبت شده است.

آماده‌سازی داده‌ها^۲: در فرایند آماده‌سازی، باید برای پاکسازی و پیش پردازش داده‌ها، دو عملیات مهم کاهش داده‌ها^۳ و اعمال تغییرات در شکل داده‌ها بر روی پایگاه داده رابطه‌ای صورت گیرد. در این مرحله، عملیات کاهش داده‌ها (حذف داده‌ها) به صورت دستی در پایگاه داده، به دو شکل زیر صورت گرفت:

▪ حذف رکورد (سارق) به دلیل ناقص بودن اطلاعات سارقان (صلاحیت ورود به مدل نهایی را ندارد. زیرا متغیرهای اساسی را در بر نمی‌گیرد مانند اینکه سارق مرتکب چه نوع جرمی شده است و در کدام دسته از انواع سرقت قرار دارد).

1. Data understanding
2. Data preparation
3. Data reduction



شکل ۱. مدل داده کاوی ارائه شده به منظور شناسایی قوانین پنهان جرم سرقت

حذف فیلد (خصیصه سارق یا پرونده سرقت)

برای اعمال الگوریتم بر روی پایگاه داده بعد از مرحله پیش پردازش برخی فیلدها حذف و برخی از فیلدها اضافه و برخی فیلدها نیز تغییر یافته است. برخی متغیرها (خصیصه‌های سارقان)، متغیرهای تأثیرگذاری در رفتار سارقان نبودند. این فیلدها در پایگاه داده تهران بزرگ شامل:

اطلاعات هویتی (نام، نام خانوادگی، نام پدر، شماره پرونده، شماره شناسنامه و شماره گواهینامه) یگان اعتراف گیرنده، یگان دستگیر کننده، نوع قرار صادره و نام قاضی است. این فیلدها عملاً و از نظر مفهومی ارتباطی به رفتار سارق نداشتند و می‌بایست از پایگاه داده حذف شوند.

پس از انجام مرحله آماده سازی داده، چهارده مشخصه به شرح جدول ۱ برای مدل سازی استفاده شد:

جدول ۱. خصایص مورد استفاده به عنوان ورودی الگوریتم

FIELD NAME	عنوان	ردیف
FK_CRIME_TYPE_ID	نوع جرم	۱
OCCURRENCE_SITUATION_ID	حالت وقوع	۲
FK_TECHNIQUE_TYPE_ID	شگرد متهم	۳
FK_OCCURRENCE_PLACE_ID	مکان وقوع جرم	۴
FK_SEXUALITY_ID	جنسیت	۵
FK_EDUCATION_ID	میزان تحصیلات	۶
FK_MARRIAGE_STATUS_ID	وضعیت تأهل	۷
OCCURRENCE_MONTH	ماه وقوع جرم	۸
AGE_ID	گروه سنی	۹
OCCURRENCE_FASL	فصل وقوع	۱۰
OCCURRENCE_NIMSAL	نیمسال وقوع	۱۱
Job	شغل	۱۲
OCCURRENCE_TIME	زمان وقوع جرم	۱۳
OCCASION_OCCURRENCE_CRIME	مناسبت وقوع جرم	۱۴

مدل سازی: در این مرحله، روش های مختلف مدل سازی، انتخاب و به کار گرفته می شوند. نوعاً برای یک نوع مسئله داده کاوی، چندین روش وجود دارد. برای ساخت مدل لازم است که ابتدا روش مدل سازی انتخاب شود که در این تحقیق از دو عملیات قوانین تلازمی و خوشه بندی استفاده شده است. در این تحقیق از دو تکنیک خوشه بندی و قوانین تلازمی استفاده شده است در مدل اولیه ابتدا تمامی داده ها به هشت خوشه تبدیل شدند خوشه های خروجی تحلیل گشت سپس به منظور استخراج قوانین در هر خوشه، عملیات احتمال پیشین بر روی خوشه ها اجرا گردید. نتایج حاصل از مدل اولیه، خوشه بندی تمامی داده ها و سپس اجرای عملیات احتمال پیشین بر روی هر خوشه می باشد.



مدل ثانویه با مقدار پشتیبان و ضریب اطمینان اولیه ساخته شد. پس از آن که قوانین توسط الگوریتم احتمال پیشین استخراج شد نتایج حاصل از آن، به اطلاع خبرگان پلیس آگاهی تهران بزرگ رسید. این افراد از بین قوانین استخراجی آنهایی که برای سازمان مفید بودند را مشخص کردند.

مرحله سوم مقایسه نتایج خروجی مدل اولیه با خروجی مدل ثانویه می‌باشد. داده‌ها با استفاده از روش کای میانگین خوشه‌بندی شده و نتایج زیر به دست آمد. برای ارزیابی حالت‌های مختلف خوشه‌بندی و انتخاب بهترین حالت از شاخص دیویس-بولدین استفاده شده و تعداد هشت خوشه به عنوان بهترین تعداد خوشه‌ها برای این داده به دست آمده است.

تحلیل نتایج

تحلیل خوشه‌های حاصل از الگوریتم کای میانگین

هشت خوشه حاصل از پیاده سازی الگوریتم کای میانگین مطابق جدول ۲ می‌باشد. با بررسی جدول ۲، خوشه یک و خوشه چهار در تمامی خصایص از قبیل نوع جرم انتخابی، جنسیت افراد، گروه سنی، شغل افراد، ساعت ارتکاب جرم، سال و مناسبت روز ارتکاب جرم و تکنیک ارتکاب جرم یکسان هستند. تنها وجه تمایز افراد متعلق به این دو خوشه وضعیت تأهل سارقان و ماه وقوع جرم در خوشه یک و چهار است. خوشه دو معرف سارقان مردی است که در دی ماه سال ۱۳۸۴ مرتکب سرقت شده‌اند. خوشه پنج را می‌توان جیب‌برها و کیف‌قاپ‌هایی که در معابر عمومی به شکل گروهی سرقت می‌کنند، نامید. خوشه هشت مختص سارقان مجرد زورگیر است که به عنف، یعنی تهدید و ارباب افراد یا گاهاً ایجاد ضرب و جرح مبادرت به سرقت کردند. با توجه به دو خوشه سه و شش به وضوح می‌توان دریافت تنها تفاوت بین این دو خوشه در ساعت وقوع جرم و وضعیت تأهل سارقان است. خوشه هفت شامل سارقان مرد که تحت پوشش مأمور آب و برق یا نیروی بسیج از اماکن دولتی مثل بانک، بیمارستان، امامزاده و بقعه متبرکه مرتکب سرقت شده‌اند.



جدول ۲. هشت خوشه‌های حاصل از پیاده سازی الگوریتم کای میانگین

نام خوشه	تعداد رکورد اطلاعاتی	توصیف هر خوشه
خوشه ۱	۲۰۹/۲۰۸	سارقان مجرد اماکن خصوصی که در فروردین‌ماه از اماکن خصوصی مثل کارگاه و انبار و ... سرقت می‌کنند
خوشه ۲	۸/۰۲۸	سارقان مرد که در دی ماه سال ۸۴ مرتکب سرقت شده‌اند
خوشه ۳	۸/۴۱۵	سارقان منزل متأهل که بین ساعات ۶۰۰ صبح تا ۱۲۰۰ ظهر سرقت می‌کنند
خوشه ۴	۱۷/۰۵۵	سارقان متأهل اماکن خصوصی که در اردیبهشت ماه سرقت می‌کنند.
خوشه ۵	۱۳/۰۵۶	جیب‌برها و کیف‌قاپ‌هایی که در معابر عمومی با شکل گروهی سرقت می‌کنند.
خوشه ۶	۱۲/۶۵۷	سارقان منزل مجردی که بین ساعات ۱۲۰۰ تا ۱۶۰۰ سرقت می‌کنند.
خوشه ۷	۷/۱۵۲	سارقان اماکن دولتی
خوشه ۸	۶/۷۰۲	سارقان مجرد زورگیر که به عنف یعنی تهدید و ارباب افراد یا گاهاً ایجاد ضرب و جرح میدارند به سرقت کرده‌اند

ترکیب دو الگوریتم خوشه‌بندی و قوانین تلازمی

در این قسمت به بررسی و تأثیر ترکیب دو الگوریتم خوشه‌بندی و قوانین تلازمی در بهبود نتایج خروجی با داده‌های موجود پرداخته می‌شود. برای این منظور در ابتدا تنها از یک الگوریتم احتمال پیشین^۱ که تکنیک قوانین تلازمی در داده کاوی می‌باشد داده‌های خود را مدل نموده و به بررسی قوانین خروجی، صحت و اعتبار قوانین پرداخته می‌شود که در بند ۶، نمونه‌ای از قوانین تأیید شده توسط خبرگان پلیس آگاهی تهران بزرگ پرداخته خواهد شد. جدول زیر قوانین استخراج شده با تعیین حداقل ضریب اطمینان و ضریب پشتیبان ده را نمایش می‌دهد. با توجه به داده‌های این جدول ضریب اطمینان در ۲۶ قاعده از قوانین استخراجی کمتر از ۵۰ درصد است و این به آن معناست که قوانین استخراج شده از این روش از اعتبار خیلی بالایی برخوردار نیستند. به همین دلیل در گام بعد به دنبال روش تلفیقی دیگری هستیم که قوانین معتبرتری را ارائه دهند. در این روش تلفیقی جدید از دو روش خوشه بندی و قوانین تلازمی استفاده می‌کنیم.



جدول ۳. تعداد و ضریب اطمینان قوانین حاصل از الگوریتم احتمال پیشین

تعداد قوانین به دست آمده	تعداد قوانین با ضریب اطمینان بین ۷۰ تا ۱۰۰ درصد	تعداد قوانین با ضریب اطمینان بین ۵۰ تا ۷۰ درصد	تعداد قوانین با ضریب اطمینان کمتر از ۵۰ درصد
۳۱	۲	۳	۲۶

طبق شرایط اولیه، همان داده‌های آزمایش اول را با یکی از تکنیک‌های خوشه بندی به نام K میانگین که در مرحله قبل به صورت کامل بیان و هشت خوشه حاصل از آن تحلیل شد به عنوان ورودی برای الگوریتم احتمال پیشین در نظر گرفته می‌شود یعنی به صورت جداگانه تک تک خوشه‌ها با الگوریتم احتمال پیشین اجرا می‌گردد و قواعد خروجی هر خوشه تحلیل و نتایج آن با خروجی حاصل از اجرای الگوریتم احتمال پیشین به تنهایی مقایسه می‌گردد. جدول زیر نتایج حاصل از ترکیب دو الگوریتم احتمال پیشین و K میانگین را نمایش می‌دهد.

با مقایسه دو شاخص ضریب اطمینان و پشتیبان در دو حالت قبل در می‌یابیم در حالت اول ضرایب اطمینان و پشتیبان در قواعد استخراجی پایین است و این برعکس حالت دوم است یعنی در قواعد حاصل از ترکیب دو الگوریتم K میانگین و احتمال پیشین درصد اطمینان و مقدار پشتیبان بسیار بالاتر از حالت اول می‌باشد. بدیهی است نتایج حاصل منطقی است زیرا در حالت اول به دلیل ناهمگن بودن رکوردهای اطلاعاتی درصد پشتیبان که تعداد تراکنش‌های موجود در بانک اطلاعاتی که شامل تمام فیلدهای پیش شرط و نتیجه نسبت به کل رکوردهای اطلاعاتی است پایین می‌باشد، همچنین درصد اطمینان که تعداد وقوع قسمت پیش شرط و نتیجه بر تعداد وقوع قسمت پیش شرط در رکوردهای اطلاعاتی است که در نتیجه درصد اطمینان در این حالت کمتر است ولی در حالت دوم استفاده از الگوریتم خوشه‌بندی باعث بخش‌بندی داده‌ها در خوشه‌هایی می‌شود که شباهت رکوردهای اطلاعاتی درون خوشه‌ها بسیار زیاد و تفاوت بین خوشه‌ها حداکثر می‌باشد پس داده‌های درون خوشه متجانس تر هستند که باعث می‌شود درصد اطمینان و درصد پشتیبان در حالت دوم بیشتر شود.

جدول ۴. تعداد و ضریب اطمینان قوانین حاصل از اجرای الگوریتم احتمال پیشین بر روی هر یک از

خوشه‌ها

شماره خوشه	تعداد قوانین به دست آمده	تعداد قوانین با ضریب اطمینان بین ۷۰ تا ۱۰۰ درصد	تعداد قوانین با ضریب اطمینان بین ۵۰ تا ۷۰ درصد	تعداد قوانین با ضریب اطمینان کمتر از ۵۰ درصد
۱	۱۶۵	۳۹	۵	۱۲۱
۲	۴۷	۲۹	۵	۱۳
۳	۶۶	۴۱	--	۲۵
۴	۱۵۱	۲۲	--	۱۲۹
۵	۲۰۱	۱۳۰	۱۱	۶۰
۶	۷۷	۳۰	۱۰	۳۷
۷	۶۵	۳۰	۶	۲۹
۸	۱۰۱	۲۹	۵۲	۲۰

ارزیابی نتایج: در این مرحله از پروژه، مدلی که از دیدگاه تحلیل داده، کیفیت بالایی دارد ساخته می‌شود. قوانین استخراج شده از الگوریتم به عنوان دانش استخراج شده از تحقیق تلقی می‌گردد. سپس با توجه به نتایج استخراج شده، قوانینی که از نظر افراد خبره معتبر و مفید بود از الگوریتم‌ها استخراج و مورد بررسی بیشتری واقع شد. در نهایت فرایند صورت گرفته برای اطمینان از صحت عملیات انجام شده مرور می‌گردد.

به کارگیری مدل^۱: ایجاد مدل، عموماً پایان پروژه نیست. حتی اگر هدف مدل، ارتقا دانش از داده‌ها باشد؛ آنگاه دانش حاصل شده نیازمند این است که سازماندهی شود و به شکلی ارائه گردد که بهره‌برداران بتوانند از آن استفاده نمایند. بسته به ملزومات کار، فاز به کارگیری می‌تواند به سادگی ایجاد یک گزارش یا به پیچیدگی اجرای یک فرایند قابل تکرار داده‌کاوی باشد. در بسیاری از موارد، این بهره‌برداران هستند که گام‌های بکارگیری را انجام می‌هند نه تحلیل‌گران داده.



قوانین تأیید شده (قوانین تلازمی)

در این بخش پس از به دست آوردن نتایج حاصل از الگوریتم احتمال پیشین، کلیه قوانین کشف شده مورد بررسی قرار گرفت. در ادامه چند نمونه از نتایج آن بررسی خواهد شد.

تأثیر شکل بنای ساختمان بر ارتکاب سرقت منزل

در سرقت از اماکن، شناسایی و انتخاب محل یکی از مهم‌ترین و دقیق‌ترین کارها برای سارق است.

جدول ۵. تأثیر شکل بنای ساختمان بر ارتکاب سرقت منزل

قانون	شاخص
منزل ویلایی ← سرقت منزل	٪ ۹۱.۶۴ = ضریب اطمینان
	٪ ۱۸.۹۹ = ضریب پشتیبان
منزل آپارتمانی ← سرقت منزل	٪ ۱۱.۶۷ = ضریب اطمینان
	٪ ۰.۵۳ = ضریب پشتیبان

الف. قاعده حاصل از جدول فوق

در جدول فوق در قانون اول ضریب اطمینان ۹۱/۶۴ درصدی به آن معنی است که ۹۱/۶۴ سرقت منزل در منازل ویلایی بوده است و در قانون دوم که ضریب اطمینان ۱۱/۶۷ درصد است نشان می‌دهد که ۱۱/۶۷ درصد سرقت منزل در منازل آپارتمانی بوده است. آنچه از جدول فوق قابل درک است این است که خانه‌های ویلایی نسبت به آپارتمان‌ها بیشتر در معرض تعدی و تجاوز هستند.

ب. نظر خبرگان در خصوص تأییدیه قاعده فوق

طبق نظر کارشناسان و خبرگان سرقت منزل معمولاً سارقان خانه‌های جنوبی را برای سرقت بر می‌گزینند، زیرا بلافاصله پس از رفع مانع وارد منزل می‌شوند. حال آنکه در منازل شمالی وضع ورودی و نفوذ به منزل برای سارقان ناشناخته بوده و خطرات بیشتری برای او دارد.

سرقت از آپارتمان‌ها دارای ویژگی‌های خاص خود است. به علت نزدیک بودن به یکدیگر کمتر مورد دستبرد سارقان حرفه‌ای قرار می‌گیرند. این گونه اماکن بیشتر مورد دستبرد سارقان معتاد و به اصطلاح «دله دزد» واقع می‌شوند که حتی از بردن جفتی کفش یا لباسی از روی بند ابایی ندارند.

تأثیر جنسیت بر روی ارتکاب جرم جیب‌بری

جدول ۶. تأثیر جنسیت بر روی ارتکاب جرم جیب‌بری

شاخص	قانون
۳۳٪ = ضریب پشتیبان / ۴۳٪ = ضریب اطمینان	زن ← جیب‌بری
۹۶٪ = ضریب پشتیبان / ۲۶٪ = ضریب اطمینان	مرد ← جیب‌بری

الف. قاعده حاصل از جدول فوق

ضریب اطمینان قانون اول جدول فوق ۴۳/۹۸ درصد است و این به آن معنی است که در داده‌های موجود ۴۳/۹۸ درصد زنان مجرم جیب‌بری هستند. و این نشان می‌دهد که سارقان جیب‌بری بر اکثر زنان هستند. ضریب اطمینان قانون دوم جدول فوق ۲۶/۵۷ درصد است و این به آن معنی است که در داده‌های موجود ۲۶/۵۷ درصد مردان مجرم جیب‌بری هستند.

آنچه از جدول ۶ قابل مشاهده است احتمال وقوع سرقت جیب‌بری توسط زنان بیش از ۱/۵ برابر مردان است.

ب. نظر خبرگان در خصوص تأییدیه قاعده فوق

سارقان جیب‌بری اغلب زنان هستند و طعمه‌های آنها هم معمولاً زنانی هستند که در فروشگاه‌ها مشغول خرید هستند یا در صف‌های مختلف ایستاده‌اند. از طرفی به دلیل ضعیف بودن عملکرد زنان پلیس در ایران، احتمال دست‌گیری و به دام افتادن جیب‌برهای زن بسیار کم است. این قانون برای کارشناسان پلیس آگاهی تهران بزرگ از قبل مشخص شده نبود اما چون دانشی در این قانون وجود داشت و مطلب مهمی را به نمایش می‌گذاشت، توسط آنان به صورت جدی ارزیابی و مورد تأیید قرار گرفت.



تأثیر فصل بر ارتکاب جرم سرقت قطعات و محتویات خودرو

جدول ۷. تأثیر فصل بر ارتکاب جرم سرقت قطعات و محتویات خودرو

شاخص	قانون
۱۸.۹۵٪ = ضریب اطمینان	بهار ← سرقت قطعات و محتویات خودرو
۲۷.۷۱٪ = ضریب پشتیبان	
۱۷.۶۳٪ = ضریب اطمینان	زمستان ← سرقت قطعات و محتویات خودرو
۲۹.۹۳٪ = ضریب پشتیبان	

الف. قاعده حاصل از جدول فوق

در قانون اول نشان می‌دهد که در ۱۸/۹۵ درصد موارد تخلف در بهار سرقت قطعات و محتویات خودرو است. در قانون دوم نیز نشان می‌دهد که در ۱۷/۶۳ درصد موارد تخلف در زمستان سرقت قطعات و محتویات خودرو است. آنچه از جدول ۷ به دست می‌آید احتمال وقوع سرقت محتویات داخل خودرو در فصل بهار نسبت به فصل زمستان بیشتر است.

ب. نظر خبرگان در خصوص تأییدیه قاعده فوق

اهداف سارقان خودرو و محتویات داخل خودرو که عموماً خوشگذرانی است معمولاً بیشتر در فصول خوش آب و هوا مثل بهار و تابستان اقدام به سرقت خودرو و محتویات داخل آن (ضبط و پخش، چنجر و...) می‌نمایند. بنا به نظر کارشناسان و افراد خبره اداره مبارزه با سرقت خودرو اغلب خودروهای سرقتی در شمال کشور کشف می‌شوند که سارقین خودرو که معمولاً جوان هستند پس از ارضای امیال نفسانی خود آنها را در آن مناطق رها می‌نمایند.

قواعد استخراج شده از الگوریتم احتمال پیشین در خصوص سرقت به عنف

جدول ۸. قواعد استخراج شده از الگوریتم احتمال پیشین در خصوص سرقت به عنف

شاخص	قانون
۴۴.۵۴٪= ضریب اطمینان	تهدید با اسلحه سرد ← سرقت به عنف
۱۱.۵۸٪= ضریب پشتیبان	

الف. قاعده حاصل از جدول فوق

طبق جدول فوق با احتمال ۴۴/۵۴ درصد سارقان به عنف از تکنیک تهدید با اسلحه سرد استفاده می کنند.

ب. نظر خبرگان در خصوص تأییدیه قاعده فوق

سارقان حرفه‌ای همیشه سلاح سرد یا گرم برای مبادا همراه خود دارند و نهایت سعی آنان حتی الامکان عدم استفاده از آن است و علت آن هم این است که هدف و مقصد آنان سرقت وسایل است نه درگیری و نزاع و فقط در مقام دفاع از این گونه وسایل استفاده می کنند.

نتیجه گیری

در این تحقیق با استفاده از داده کاوی و الگوریتم‌های آن سعی شده است تا تحلیل‌های دقیقی بر روی داده‌های جرم موجود در پایگاه داده‌ها انجام شود و با استفاده از الگوهای پنهان موجود در داده‌ها مدیران مربوطه را در اخذ تصمیم‌ها یاری نمایند. در این مقاله با استفاده از دو روش قوانین تلازمی و روش خوشه بندی سعی شده است تا این قبیل الگوهای مورد نیاز مدیران کشف و در اختیار آنها قرار گیرد. به عنوان مثال برخی از الگوهای استخراج شده نشان می دهند که خانه‌های ویلایی نسبت به آپارتمان‌ها بیشتر در معرض تعدی و تجاوز هستند و یا برخی از این قواعد نشان می دهند که احتمال وقوع سرقت جیب‌بری توسط زنان بیش از ۱/۵ برابر مردان است. همچنین این الگوها نشان



می‌دهند احتمال وقوع سرقت محتویات داخل خودرو در فصل بهار نسبت به فصل زمستان بیشتر است.

در این مقاله چهار قاعده مفید از مجموع پنجاه قاعده که توسط خبرگان تفسیر و تأیید گردید بیان شد. البته آنچه واضح و مبرهن است این است که داده کاوی جرم به مراتب وسیع تر و گسترده تر از آن است که بتوان به صورت جامع در قالب یک مقاله ارائه نمود. موضوع امنیت یک موضوع ملی و گاهی منطقه‌ای و جهانی است از این رو بهره برداران داده کاوی جرم عموماً داعیه داران امنیت (قوه قضائیه، وزارت اطلاعات، شورای امنیت ملی، نیروی انتظامی و دیگر سازمان‌های ذیربط) می‌باشند. بسته به نوع جرایم، داده کاوی جرم می‌تواند در زمینه‌های مختلف زیر مورد بحث واقع شود.

پیشنهادها

محقق با مطالعاتی که در حوزه ادبیات موضوع تحقیق انجام داده، پیشنهادهایی را جهت انجام تحقیق‌های آتی ارائه نموده است که در ادامه بیان می‌گردند. تمامی این تحقیق‌ها به مدیران سازمان‌های مربوطه کمک می‌کنند تا با استفاده از دانش استخراج شده از داده‌های تاریخی سازمان تصمیم‌های درست و دقیق‌تری اتخاذ نمایند:

- ۱- استفاده از روش‌ها و مدل‌های ترکیبی داده کاوی و هوش مصنوعی به منظور شناسایی سارقانی که بیشترین سابقه را در ارتکاب جرم دارند؛
- ۲- طراحی و ایجاد یک سیستم هوشمند خبره که با استفاده از تکنیک‌های داده کاوی وضعیت ارتکاب جرم را برای مجرمان در آینده پیش بینی کند؛
- ۳- استفاده از تکنیک خوشه‌بندی بر روی قوانین انجمنی حاصل از پایگاه داده‌ها به منظور کاهش در تعداد قوانین تلازمی و هم‌خوشه کردن قوانین استخراجی مشابه در پایگاه داده جرم و بررسی تفاوت در جابه‌جایی ترکیب این دو روش داده کاوی (خوشه بندی و قوانین انجمنی)؛

۳- شناسایی اماکن جرم نیز از حیث ارتکاب به جرایم مختلف و مناطق آلوده از حیث معامله اموال مسروقه با کمک spatial mining؛



۴- استفاده از داده‌کاوی در کشف ارتباطات پنهان بین سارقان، مالخران و اموال

مسروقه.

منابع فارسی

بیابانی، غلامحسین. (۱۳۸۶). شیوه‌ها و شگردهای مجرمانه. انتشارات کارآگاه. جلد ۱.

منابع لاتین

- Annabathula, R., 2007. *A web-based tool for analysis of crime laboratory data*. Master of Science Thesis, West Virginia University.
- Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y. and Chau, M. 2004. 'Crime Data Mining: A General Framework and Some Examples', *Computer: Security and Privacy in an Online World*, Vol. 37 (4), pp. 50 – 56.
- Chung, W., Chen, H., Ch., Chang, W. & Chou, SH. (2006), Fighting Cybercrime: A Review and the Taiwan Experience, *Decision Support Systems*, Vol. 41, pp. 669–682.
- Han, J. and Kamber, M. 2001. 'Data mining: Concepts and Techniques', The Morgan Kaufmann Series in Data Management Systems.
- Li, S.T., F.C. Tsai, S.C. Kuo, and Y.C. Cheng. 2006. "A Knowledge Discovery Approach to Supporting Crime Prevention." In Proceedings of JCIS 2006.
- Li, SH.T., Kuo, SH.CH. & Tsai, F.CH. (2010), An Intelligent Decision-Support Model Using FSOM and Rule Extraction for Crime Prevention, *Expert Systems with Applications*, Vol. 37, pp. 7108–7119.
- McCue, Colleen 2003. "Data Mining and Crime Analysis in the Richmond Police Department." SPSS Executive Report. Online. http://www.spss.com/registration/premium/consol056.cfm?WP_ID=132.
- Moon, B., McCluskey, J.B. & McCluskey, C.P. (2010), General Theory of Crime and Computer Crime: An Empirical Test, *Journal of Criminal Justice*, Vol. 38, pp. 767–772.
- Oatley, G.C., MacIntyre, J., Ewart, B.W., & Mugambi, E., 2002. SMART Software for Decision Makers KDD Experience. *Knowledge Based Systems*. Vol. 15, pp. 323-333.
- Oatley, G., Ewart, B and Zeleznikow, J., 2006. Decision support systems for police: Lessons from the application of data mining techniques to "soft" forensic vidence. *Artificial Intelligence and Law*. Vol 14, pp. 35-100.
- Seifert, J.W., 2004. "Data Mining and Homeland Security: An Overview". CRS Report for Congress.