# کاربرد داده‌کاوی در سازمان‌های پلیسی و قضایی به منظور شناسایی الگوهای جرم و کشف جرائم

پروانه کاظمی
کارشناس ارشد مهندسی فناوری اطلاعات، دانشگاه تربیت مدرس
Parvanehkazeni_pk@yahoo.com

جواد حسین پور
کارشناس کامپیوتر، معاونت فناوری اطلاعات و ارتباطات ناجا
تهران، ایران
javadhsn@yahoo.com

## 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 توسعهٔ سیستم‌های اطلاعاتی در سازمانها، هدف اصلی این تحقیق، مطالعه و بررسی روشی مبتنی بر داده کاوی است که با بکارگیری بانک‌های اطلاعاتی موجود و استفاده از ابزارها و الگوریتم‌های داده‌کاوی، بتوان داده‌های موجود را مدل و الگوهای جرم را شناسایی و کشف نمود تا بدین طریق پلیس بتواند وقوع جرم را پیش‌بینی کرده و با کنترل دقیق‌تر نیروها و آرایش نظامی آنان در منطقه جرم، از وقوع جرائم پیشگیری نماید.
از آنجایی‌که الگوریتم‌ها، تکنیک‌ها و روشهای داده‌کاوی یکی از ضرورت‌ها و محور اصلی مطالعات می‌باشد لذا در ابتدا به مفاهیم داده‌کاوی، تکنیک‌ها، ابزارها و چارچوب داده کاوی جرم پرداخته شده است. در قسمت بعد به مرور تجربیات و فعالیت‌های صورت‌گرفته در این حوزه به همراه روش‌ها و ابزارهای مربوط به تحلیل جرم در سازمان‌های پلیسی پرداخته شده است.

## واژگان کلیدی

داده کاوی جرم، الگوهای جرم، کشف جرم، ابزارهای داده کاوی جرم، فرآیند داده کاوی

## ۱- مقدمه

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

شاید بهترین مدل برای درک مقوله پیشگیری از وقوع جرم، دیدگاه کاهش فرصت ارتکاب جرم است. بکارگیری روش‌های پیشگیری از جرم با تئوری‌های کاهش فرصت ارتکاب جرم هماهنگ می‌باشد. 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 به کاهش جرائم در آن مکان منجر می‌گردد.

در این نوشتار تمرکز اصلی تحقیق، بر روی بکارگیری بانک‌های اطلاعاتی و داده کاوی جرم است که شانس وقوع جرم را از طریق پیش‌بینی‌های دقیق کاهش دهد. راهکارهای پیشنهادی در واقع کشف و شناسایی الگوهای جرم و همچنین شیوه و شگردهای مجرمان در میان گروه‌های بزه‌کار با استفاده از ابزارهای موجود در حوزه داده‌کاوی است تا بدین طریق بتوان با بهره‌گیری مناسب و بهینه افسران پلیس از وقوع جرم پیشگیری نمود.

## ۲- داده کاوی

تاکنون تلاش زیادی برای کشف دانش مخفی در مجموعه‌های داده به وسیله محققین در رشته های مختلف صرف شده است. در اواخر سال‌های ۱۹۸۰ اصطلاح جدید کشف دانش در پایگاه داده بکار برده شد و سریعاً از سوی متخصصین هوش مصنوعی و فراگیری ماشین برای پوشش فرآیند کلی استخراج دانش از پایگاه های داده از بدو تعریف مساله و اهداف آن، تا تحلیل‌های نهایی نتایج مورد استفاده قرار گرفت.

در این زمینه واژه "داده کاوی" به عنوان یک مرحله یا گام در فرآیند کشف دانش بکار برده شد. این تفسیر توسط فیاد در سال ۱۹۹۴ بدین شکل ارائه شد، بدین معنی که فرآیند کشف دانش شامل مجموعه از تکنیک‌های ریاضی و کامپیوتر براساس تحلیل داده‌های موجود در یک پایگاه داده بزرگ، برای یافتن یک راه حل براساس الگوهای کشف شده در داده‌ها و بکاربردن راه‌حل برای مساله تعریف شده می‌باشد. بعبارت دیگر کشف دانش در پایگاه داده فرآیند شناسایی درست، ساده، مفید، و نهایتاً الگوها و مدل‌های قابل فهم در داده‌ها است که داده کاوی، مرحله‌ای از فرایند کشف دانش است و شامل الگوریتم‌های مخصوص داده‌کاوی است، بطوریکه، تحت محدودیت‌های مؤثر محاسباتی قابل قبول، الگوها و یا مدل‌ها را در داده‌ها کشف می‌کند.

در اینجا منظور از الگوی مفید، مدلی در داده‌ها است که ارتباط میان یک زیر مجموعه از داده‌ها را توصیف می‌کند در حالیکه معتبر، ساده، قابل فهم و جدید است. در واقع داده‌کاوی یکی از مهمترین روش‌هایی است که به وسیله آن الگوهای مفید در درون داده‌ها با حداقل دخالت کاربران شناخته می‌شوند و اطلاعاتی را در اختیار کاربران و تحلیل‌گران قرار می‌دهند تا براساس آنها تصمیمات مهم و حیاتی در سازمان‌ها اتخاذ شود.

## ۳- فرآیند داده‌کاوی

طبق مدل CRISP-DM1 گامهای اصلی در فرآیند داده کاوی جهت کشف دانش عبارتند از:

۱- تعریف مساله
۲- ساختن پایگاه داده مربوط به داده کاوی
۳- جستجوی داده
۴- آماده ساختن داده برای مدل سازی
۵- ساختن مدل
۶- ارزیابی مدل
۷- ساخت مدل ونتایج

### ۳-۱- تعریف مساله

در ابتدای امر پیش‌زمینه کشف دانش، فهم درست داده و مساله است. بدون این فهم درست هیچ الگوریتمی صرف نظر از خبره بودن آن نمی‌تواند نتیجه مطمئنی حاصل نماید. همچنین قادر نخواهد بود مسائلی را که سعی در حل آن است تعریف کرده و داده را جهت کاوش آماده نموده و یا نتایج را به طور صحیح تفسیر نمائید. برای استفاده بهتر از داده کاوی باید یک بیان واضح از هدف وجود داشته باشد(Corporation , 1999).

### ۳-۲- ساختن یک پایگاه داده برای داده‌کاوی

این گام به همراه دو گام بعدی هسته آماده‌سازی داده را تشکیل می‌دهند. در مجموع این گام‌ها وقت و کار بیشتری از سایر گام‌ها می‌برند. ممکن است گام‌های تکراری در آماده‌سازی داده و ساختن مدل داشته باشد چرا که در هر مرحله ممکن است به نکته‌ای رسید که نیاز باشد داده را بهبود بخشید. این گام‌های آماده‌سازی داده می‌تواند ۵۰٪ تا ۹۰٪ وقت و کار از تمام فرآیند کشف دانش را به خود اختصاص دهد.

داده‌ای که می‌خواهد کاوش شود باید در یک پایگاه داده ذخیره شود. بر اساس مقدار داده، پیچیدگی داده و استفاده‌هایی که قرار است از آن شود یک فایل معمولی و یا یک SpreadSheet برای این کار کافی است. به احتمال زیاد نیاز است داده موجود در انباره داده را تغییر داد. به علاوه ممکن است فیلدهای جدیدی که از فیلدهای موجود محاسبه شده است را به انبار داده افزود.این یکی از دلایل استفاده از یک پایگاه داده جداگانه است.

دلیل دیگر برای این کار آن است که انبار داده‌های یکی شده ممکن است به آسانی انواع جستجوهایی را که برای فهم داده به آنها نیاز است انجام ندهد.

دلیل آخر اینکه ممکن است این داده را در یک سیستم مدیریت پایگاه داده به همراه یک طراحی فیزیکی متفاوت از انبار داده خود ذخیره نمود (Jeffery , 2004).

### ۳-۳- جستجوی داده

شناسایی مهمترین فیلدها در پیش‌بینی نتیجه و تعیین اینکه کدام یک از داده‌های بدست آمده مفید می‌باشد، هدف این مرحله است.

### ۳-۴- آماده‌سازی داده برای مدل سازی

این آخرین گام آماده‌سازی داده قبل از ساخت مدل‌است. چهار قسمت مهم در این مرحله وجود دارد:

- انتخاب متغیرها
- انتخاب سطرها

1 . Cross Industry Standard Process for Data Mining

- ساختن متغیرهای جدید
- تغییر شکل متغیرها

مهمترین مساله برای یادآوری در مورد ساخت مدل آن است که این کار یک فرآیند تکراری است. برای جستجو به مدل‌های جایگزین جهت یافتن سودمندترین آنها جهت حل مسائل نیاز است. آنچه که در جستجوی یک مدل مناسب یاد گرفته می‌شود می‌تواند به بازگشتن به عقب و انجام برخی تغییرات در داده مورد استفاده و حتی بهبود بیان مساله راهنمایی کند.

آماده‌سازی و آزمایش مدل داده کاوی احتیاج به این دارد که داده به حداقل دو گروه شکسته شود: یکی برای آماده کردن مدل و دیگری جهت تست مدل مربوطه(Corporation, 1999).

### ۳-۵- تائید اعتبارساده

پایه‌ای‌ترین روش تست داده تایید اعتبار ساده می‌باشد. برای انجام این کار چون درصدی از پایگاه داده را به عنوان یک تست پایگاه داده کنار بگذارید و به هر صورت از آن در برآورد و ساخت مدل استفاده ننمائید.

### ۳-۶- ارزیابی و تفسیر وتایید اعتبار مدل

بعد از ساخت یک مدل باید نتایج آن را ارزیابی نموده و همچنین اهمیت آن را نیز توضیح داد. معمولاً برای مسائل طبقه بندی یک ماتریس پیچیدگی ابزار مفیدی برای فهم نتایج می باشد.

## ۴- روش‌ها و الگوریتم‌های داده‌کاوی جرم

الگوریتم‌های داده‌کاوی جرم در حقیقت مکانیزمی جهت ایجاد مدل از درون بانک‌های اطلاعاتی مربوط به جرائم است. برای ایجاد یک مدل، در ابتدا یک الگوریتم باید مجموعه‌ای از داده‌ها را تحلیل و یک الگوی خاص و روندهای انجام کار را جستجو کند. سپس از نتایج این تحلیل استفاده کند تا پارامترهای مدل‌های استخراج را تعریف کند. مدل‌های استخراج که از یک الگوریتم حاصل می‌شوند می‌تواند انواع شکل‌های زیر را بگیرد:

- یک مجموعه از نقش‌ها که توصیف می‌کند چگونه بخش‌های جنایی در تجسس‌های حادثه‌ای گروه‌بندی می‌شوند.
- یک مجموعه از خوشه‌ها که توصیف می‌کند حوادث در یک پایگاه داده چگونه وابسته هستند.
- یک درخت تصمیم که پیش‌بینی می‌کند آیا تبهکاران ویژه این چنین جرمی را انجام خواهند داد.

انتخاب درست یک الگوریتم برای استفاده در یک حادثه ویژه، می‌تواند یک چالش باشد، زیرا ممکن است استفاده از الگوریتم‌های مختلف جهت اجرای همان کار لازم باشد. در واقع هر الگوریتم یک نتیجه مختلف را تولید می‌کند، همچنین الگوریتم می‌تواند بیش از یک نوع نتیجه را تولید کند.

در الگوریتم پیش‌بینی، از داده‌ها استفاده می‌شود تا رفتارها، الگوها و روندهای کار را پیش‌بینی کند.

*اکتشاف متوالی* برای تعیین الگوهای توالی در داده‌ها استفاده می‌شود. این ترتیب‌ها اغلب انواع فیلدهای داده را پیوند می‌دهند.

عمومیت دادن که توصیف یا خلاصه سازی نیز نامیده شده است، داده‌ها را به سمت زیرمجموعه‌هایی با توصیف مربوط به آنها استخراج می‌کند. عمومیت دادن یک روش داده کاوی نیست، این در واقع همان نتیجه حاصل از تکنیک داده کاوی است(Hsinchun Chen).

*الگوریتم وابستگی* در واقع وابستگی ارتباط بین ویژگیهای مختلف الگوریتم در یک پایگاه داده را پیدا می‌کند.

*الگوریتم کلاس بندی* یک یا چند متغیر گسسته را بر پایه ویژگیهای دیگر در پایگاه داده پیش‌بینی می‌کند.

درخت تصمیم، درختی است که هر شاخه آن یک انتخاب بین یک تعداد از پیشنهادها را نشان می‌دهد بطوریکه هر گره برگ، یک تصمیم را نمایش می‌دهد. درخت تصمیم بصورت عادی برای کسب اطلاعات به منظور تصمیم‌گیری استفاده می‌شود. در درخت تصمیم پیش‌بینی‌هایی مبنی بر برخی گرایشات به سمت یک نتیجه خاص تهیه می‌شود(Annabathula.R,2007 ).

*الگوریتم Naive Bayes*، این الگوریتم احتمال مشروط بین ستون ورودی و قابل پیش‌بینی را محاسبه می‌کند و فرض می‌کند که ستون‌ها مستقل هستند.

*رگرسیون خطی* یک ارتباط بین دو ستون متوالی تعین می‌کند.

*رگرسیون ترتیبی*؛ این متد کمک می‌کند تا پیش‌بینی‌هایی با پاسخ ترتیبی ساخته شود.

*الگوریتم سری‌های زمانی* یک الگوریتم رگرسیون است که می‌تواند تعداد مورد انتظار از جرائم را برای یک سال پیش‌بینی کند.

*الگوریتم شبکه عصبی* احتمالات را برای هر حالت ممکن از ویژگی‌های ورودی محاسبه می‌کند. الگوریتم شبکه عصبی برای تحلیل داده ورودی پیچیده مفید است.

## ۵- تکنیک‌های خوشه‌بندی

تکنیکهای خوشه‌بندی اقلام داده را به داخل کلاس‌ها با ویژگیهای مشابه، مطابق با حداکثر یا حداقل شباهت درون کلاسی (intaclass) طبقه بندی می‌کند مانند شناسایی مظنونین که جرایم را به شیوه‌های مشابه انجام می‌دهند یا تشخیص دادن گروه‌ها در میان دسته‌های مختلف جنایتکاران. این تکنیک‌ها مجموعه‌ای از پیش تعریف شده برای اختصاص اقلام ندارند.

برخی پژوهشگران از الگوریتم Concept space مبتنی بر آمار به منظور مربوط ساختن آبجکت‌های مختلف (مانند اشخاص، سازمان‌ها و خودروها) به صورت خودکار در رکوردهای جرم استفاده می‌کنند.

بکار گیری تکنیک‌های تحلیل پیوند در شناسایی تراکنش‌ها و شبکه جرائم مالی موثر بوده و داده محرمانه در اسناد بانک را استخراج می‌کند تا بدینوسیله جرم پولشویی و دیگر جرائم مالی را شناسایی و تحلیل کند.

Clustering Crime خوشه‌بندی جرائم می‌تواند به صورت خودکار یک بخش اصلی از تحلیل جرم را آنالیز کند اما به خاطر محاسبات بزرگ که به شدت مورد نیاز است محدود شده است.

Association rule mining خیلی اوقات مجموعه اقلام اتفاق افتاده در یک پایگاه داده را کشف می‌کند و الگوها را به عنوان یکسری قواعد نشان می‌دهد. این تکنیک‌ها در شناسایی نفوذ به شبکه بکار برده شده تا قواعد وابستگی را از سوابق تعاملات کاربران استخراج کند. محققیقین همچنین می‌توانند این تکنیکها را در شناسایی متجاوزان شبکه بکار ببرند تا به شناسایی حملات بالقوه به شبکه کمک کنند.

Sequential pattern mining مثلassociation rule mining، خیلی اوقات سلسله اتفاقات پی در پی روی یک مجموعه تراکنشها که در زمانهای مختلف اتفاق افتاده است را شناسایی می‌کند. درشناسایی نفوذ به شبکه، این رویکرد می‌تواند الگوهای نفوذ در میان داده ممهور به زمان را شناسایی کند.

آشکار سازی الگوهای پنهان، در تحلیل جرم مفید هستند، اما برای بدست آوردن نتایج معنی‌دار به داده گران‌بها و بطور عالی ساخت‌یافته نیاز داریم.

تشخیص انحراف (Deviation detection) از اقدامات ویژه در مطالعه داده که بطور محسوس از بقیه داده‌ها تفاوت دارد استفاده می‌کند. همچنین محققان می‌توانند از این تکنیک در کشف فریب، کشف نفوذ به شبکه و تحلیل دیگر جرائم استفاده کنند.

کلاس‌بندی خواص مشترک در میان موجودیت‌های جرم متفاوت را پیدا می‌کند و آنها را داخل کلاس‌های از پیش تعریف شده سازمان قرار می‌دهد.این تکنیک‌ها در شناسایی منابعی از ایمل‌های Spam بر مبنای الگوهای زبان شناختی و ساختار خصوصیات فرستنده استفاده شده است. این تکنیک‌ها اغلب بکار می‌رود تا گرایش به جرم را پیش‌بینی کند، کلاس‌بندی می‌تواند زمان مورد نیاز برای شناسایی موجودیت‌های جرم را کاهش بدهد. اگر چه آن تکنیک به یک نمونه کلاسه بندی از پیش تعریف شده نیازمند است، اما کلاس‌بندی به آموزش کامل و داده آزمایشی نیاز دارد زیرا درجه زیادی از داده‌های گم شده، دقت و صحت پیش‌بینی را محدود خواهد کرد.

تکنیهای String comparator مقایسه رشته به صورت دو به دو، فیلدهای رشته‌ای رکوردها ی پایگاه داده را مقایسه می‌کند و شباهت بین آنها را محاسبه می‌کند. این تکنیک‌ها می‌توانند اطلاعات فریبنده از قبیل نام، آدرس و شماره بیمه را در رکوردهای جنایی کشف کنند ماموران تحقیق می‌توانند از مقایسه کننده رشته در تحلیل داده متنی استفاده کنند اما آن تکنیک‌ها اغلب به محاسبات متمرکز نیاز دارند.

تحلیل شبکه اجتماعی (Social network analysis )، نقش‌ها و تعاملات در میان گره‌های یک شبکه مفهومی را ( conceptual network) توصیف می‌کند. ماموران تحقیق می‌توانند از این تکنیک در ساختن یک شبکه استفاده کنند که نقش‌های جنایی، جریانی از کالاهای قابل لمس وغیر قابل لمس، اطلاعات، و پیوستگی میان این موجودیت‌ها را شرح می‌دهد. تحلیل بیشتر می‌تواند نقشهای حیاتی، زیرگروه‌ها و آسیب‌پذیری‌ها درون شبکه را آشکار کند. این رویکرد، تصویری از شبکه‌های جنایی فراهم می‌کند اما اگر ماموران تحقیق خصوصیات و ویژگیهای اندکی را نگهداری کنند باز هم نمی‌توانند رهبران حقیقی شبکه را کشف کنند Chen, Chung, Jie,Gang (Wang, Chau, 2004).

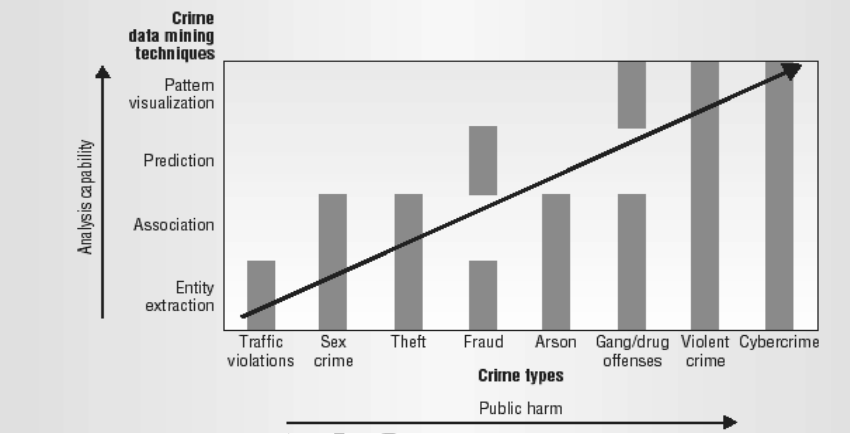
## ۶- چهارچوب داده‌کاوی جرم

تلاشهای زیادی برای استفاده از تکنیک‌های خودکار در تحلیل انواع مختلفی از جرم بکار رفته است.اما بدون یک چهارچوب واحد توصیفی نمی‌توان آنها را بکار برد. بطور خاص فهم روابط بین توانایی

تحلیل و ویژگیهای انواع جرم می‌تواند به ماموران تحقیق کمک کند تا به طور موثرتری از آن تکنیک‌ها استفاده کنند تا گرایشات و الگوها، مشکلات آدرس‌دهی نواحی و حتی پیش‌بینی جرائم را شناسایی کنند.

بر مبنای طبقه‌بندی پایگاه داده جرم سازمان پلیس Tuscon، که تقریباً شامل ۱،۳ میلیون مظنون و ثبت رکوردهای جنایی از ۱۹۷۰ تا کنون بود، چهارچوب کلی برای داده کاوی جرائم توسعه داده شد که در شکل (۱) میزان رشد آن نشان داده شده است.

این چهارچوب روابط بین تکنیک‌های داده‌کاری بکار رفته در تحلیل‌های هوشمند و جنایی، انواع جرایم لیست شده در شکل(۱) را نشان می‌دهد. محور عمودی تکنیک‌های داده کاوی را بر حسب توانایی تحلیل مرتب کرده است. در اینجا چهار دسته اصلی از تکنیک‌های داده‌کاوی جرم را شناسایی کرده‌است. استخراج موجودیت، پیوستگی، پیش‌بینی و تصویرسازی الگو. هر دسته یک مجموعه از تکنیک‌ها را برای استفاده در برخی از تحلیل جرم نشان می‌دهد. برای مثال ماموران تحقیق می‌توانند تکنیک‌های شبکه عصبی را در استخراج موجودیت و پیش‌بینی جرم بکار برند. تکنیک‌های خوشه‌بندی در پیوستگی جرائم و پیش‌بینی موثر هستند.




شکل(۱): روابط بین تکنیک‌های بکار رفته در جرائم و تحلیل هوشمند در سطح محلی، ملی و بین المللی

با نتیجه گرفتن قواعد وابستگی از تاریخچه فعل و انفعالاتی که کاربران دارد می‌توان در کشف نفوذ به شبکه استفاده نمود. محققان همچنین می‌توانند با استفاده از این تکنیک، شبکه خصوصیات متجاوزان را شناسایی کنند و حمله های بالقوه را به شبکه در آینده بررسی و کنترل نمایند.

تحلیل شبکه اجتماعی می‌تواند وابستگی جرم و تصویرسازی الگو جرم را تسهیل کند. ماموران تحقیق می‌توانند تکنیک‌های گوناگون مستقل یا ترکیبی را بکار برند تا مشکلات مربوط به تحلیل جرم خاص را مرتفع کنند.

آن چهارچوب قابلیت اجرا عمومی در تحلیل جرم و آنالیز هوشمند را دارد، زیرا تمام انواع جرائم اصلی مانند تکنیکهای سنتی و تکنیک‌های جدید داده کاوی هوشمند را در بر می‌گیرد (Chung, Jie,Gang Wang, Chau, 2004, Chen,).

حال با توضیح دادن چهارچوب داده‌کاوی جرم، سه مثال از مواردی که در پروژه Coplink استفاده شده است را توضیح می‌دهیم:

### ۶-۱- استخراج موجودیت مشخص (Named-entity extraction):

معمولاً پایگاه داده دادگستری جنایی تنها داده‌های ساخت‌یافته 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 بوسیله بکارگیری تکنیک‌های اتوماتیک مشکل است. بصورت تصادفی ۳۶ مورد از داروهای مواد مخدر از سازمان پلیس phonix انتخاب شد که نسبتاً زیاد بودند و با حروف کوچک نوشته شده بودند و شامل تعدادی اشتباهات املایی، تایپی و اشتباهات دستوری بودند. 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 بکار می‌برد، در اختیار گرفتند. به منظور پذیرش سیستم برای برنامه‌های کاربردی تحلیل جرم، آن را بهبود دادند تا پنج نوع موجودیت را شناسایی کند:" نام‌های اشخاص، آدرس‌ها، خودروها، نام‌های مربوط به مواد مخدر و خصوصیات فیزیکی". سپس یک کارآگاه مشاور بصورت دستی تمام موجودیت-هایی که به آن پنج نوع تعلق داشتند را شناسایی کرد. آن استخراج کننده بهبود داده شده، به خوبی شناسایی نام اشخاص (۷۴،۱ درصد) و مواد مخدر (۸۵،۴ درصد) را از مجموعه داده تستی انجام داد اما بهمان اندازه برای آدرس دهی(۵۹،۶ درصد) و خصوصیات اشخاص(۴۶،۸ درصد) مناسب نبود. فراخوانی‌دوباره برای همان اقلام به ترتیب ۷۳،۴، ۷۷،۹، ۵۱،۴ و ۴۷،۸ درصد بودند. در نتیجه نام خودرو تحلیل نشده بود زیرا تنها ۴ مورد از ۳۶ گزارش اتفاق افتاده مربوط به حوادث خودرو بود. این نتایج ابتدایی قابلیت و استعداد بالقوه مقادیر را در استفاده از تکنیک‌های استخراج موجودیت در داده کاوی جرم نشان داد، خصوصاً با ملاحظه اینکه گزارشات داستانی ارائه شده بسیار شلوغ تر از مقاله‌های خبری استفاده شده در ارزیابی‌های MUC-6 بودند. با اینکه تنها ۳۶ گزارش را در مطالعه خود آزمایش کرده‌اند اما بکارگیری داده آزمایشی و آموزشی را بطور نمونه در برنامه‌های کاربردی داده کاوی واقعی طرح ریزی کردند تا کل سیستم را ارزیابی کنند.

**۶-۲-اکتشاف خصوصیات فریبنده (detective – identity detection):**

مظنونین اغلب اسامی، تاریخ تولد یا آدرس‌ها را بصورت اشتباه به افسران پلیس می‌دهند. بنابراین ورودی‌های گوناگونی در پایگاه داده‌ها دارند که این موضوع تشخیص خصوصیات درست یک مظنون،گزارش حوادث گذشته‌ای که آن شخص درگیر بوده و یا گیر افتاده، را برای افسران مشکل می‌کند. وظیفه دوم داده‌کاوی کشف خودکار خصوصیات جنایی فریب‌آمیز در پایگاه داده سازمان پلیس TUSCON عنوان شد که شامل اطلاعاتی از قبیل نام، جنسیت، آدرس، شماره یا ID Number و توصیف بدنی (PHYSICAL DESCRIPTION) بود. کارآگاه مشاور بصورت دستی ۱۲۰ رکوردهای جنایی فریبنده را شناسایی کرد که ۴۴ متهم را از پایگاه داده بر مبنای خصوصیات جنایی فریبنده گیر انداخت. در بررسی موردی نام، تاریخ تولد، آدرس و شماره بیمه بازنشستگی همگانی را انتخاب کردند تا خصوصیات جنایی را نشان دهند و از دیگر فیلدهای کم معتبر صرف نظر کردند. متد مقایسه کننده رشته را بکار گرفتند تا مقادیر را در فیلدهای مشابه از هر جفت رکورد، مقایسه کند. مقایسه کننده، تشابه بین دو رشته را اندازه گرفت. مقدار های مشابه را بین صفر و یک نرمال کردند و یک وزن تشابه کلی را بین دو رکورد به عنوان اصل اقلیدوسی روی چهار فیلد انتخابی، حساب کردند.

یک متد اعتبار سنجی بسط یافته بکار گرفتند که از دو سوم داده‌ها برای آموزش و مابقی داده‌ها برای تست استفاده کردند.

نتایج بدست آمده نشان داد که داده کاوی جرم امکان پذیر و امید بخش است. خطاهای آزمایش که در نوع منفی کاذب اتفاق افتادند، در جائی بود که متهمان نا مرتبط، به عنوان شخصیت مربوط، سازمان‌دهی شده بودند، که ممکن بود معلول ارزش آستانه-ای بدست آمده از مرحله آموزش باشد. بنابراین برای گسترش یک فرآیند خودکار در تحقیقات آینده یک آستانه توافقی می‌تواند بیشتر مطلوب باشد. با این تکنیک افسران اجرای قانون می‌توانند رکوردهای هویت موجود را که مربوط به یک مظنون در پایگاه داده است را بازیابی کنند در حالیکه تکنیک‌های Exact match سنتی اغلب در مکان‌یابی آنها موفق نمی‌شود.

**۶-۳- تحلیل شبکه جنایی(criminal – network analysis):**

تبهکاران اغلب شبکه‌هایی را توسعه می‌دهند که از گروها و یا تیم‌هایی شکل گرفته که فعالیت غیر قانونی گوناگون را انجام می-دهند. وظیفه سوم داده کاوی عبارت است از شناسایی کردن زیر گروه‌ها و عضوهای کلیدی در این قبیل شبکه‌ها و سپس مطالعه الگوهای تعاملات آنها، تا استراتژی موثری را برای مختل کردن شبکه‌ها توسعه دهند. داده‌های مورد آزمایش از ۲۷۲ مورد خلاصه حادثه های سازمان پلیس TUCSON بود که با ۱۶۴ تا از جرائم مرتکب شده از ۱۹۸۵، تا انتهای می ۲۰۰۲ کار می‌کرد.

یک رویکرد Concept – Space در استخراج روابط جنایی از خلاصه های حوادث بکار گرفتند و یک شبکه احتمالی از مظنونین ایجاد کردند. وزن Co-Occurrence شدت روابط بین دو تبهکار را در حوادث یکسان، اندازه گرفت. دسته‌های سلسله مراتبی را بکار بردند تا شبکه را به داخل زیر گروه‌ها تفکیک کند و از رویکرد مدل‌سازی بلوک استفاده کردند تا الگوهای تعامل بین این زیرگروه-ها را شناسایی کنند. همچنین مرکزیت اقدامات را محاسبه کردند تا اعضای کلیدی در هر گروه از قبیل رهبران و مهره های کلیدی، را کشف کنند.

## ۷- ابزارهای داده‌کاوی موجود

طی سال‌های اخیر جریان سریعی از تمایل به ابزارهای داده‌کاوی در بازارهای نرم‌افزاری به وجود آمده است. بیشتر کاربران نرم‌افزارهای داده‌کاوی با تفکر استفاده تجاری از این نرم‌افزارها، خواهان استفاده از آن شده‌اند. نرم‌افزارهای داده‌کاوی معمولاً سه روش مختلف را برای استفاده از داده‌کاوی به کار می‌برند. ۱) اکتشاف ۲) استفاده از مدل‌های پیشگویی ۳) استفاده از آنالیز و تحلیل.

اکتشاف، فرآیند جستجو در داده‌هاست تا الگوهای مخفی موجود در داده‌ها را بدون هیچ ایدهٔ از پیش تعیین شده‌ای مشخص نماید.

در نرم‌افزارهای داده‌کاوی مبتنی بر مدل‌های پیشگویی، الگوهایی که از یک بانک داده کشف می‌شوند، برای پیش‌بینی آینده به کار می‌روند. مدل‌های پیش‌بینی به کاربران اجازه می‌دهند تا داده‌های نامشخص را به کار ببرد و این مقادیر نامشخص توسط نرم‌افزار کشف شود.

در مدل‌های تحلیلی نیز الگوهای یافت شده از داده‌ها برای تعیین مقادیر غیرعادی به کار می‌رود. برای تعیین مقادیر غیر عادی، ابتدا می‌بایست مقادیر عادی شناخته شود تا بر این اساس مقادیر غیرعادی و منحرف شناخته شوند.

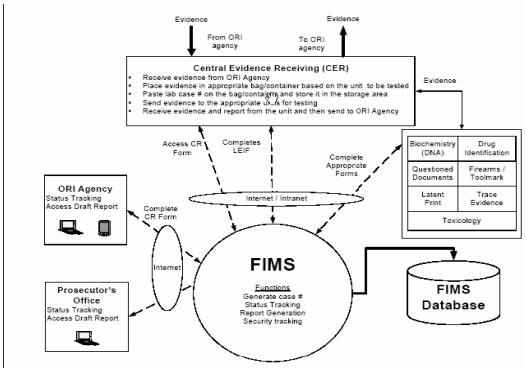
نرم‌افزارهای داده‌کاوی را می‌توان در حال حاضر در شش بخش کلی، ۱. دسته‌بندی داده‌ها، ۲. برآورد مقادیر نامشخص، ۳. پیش‌بینی مقادیر نامشخص، ۴. گروه‌بندی تقریبی داده‌ها، ۵. خوشه‌بندی داده‌ها و ۶. تشریح روابط بین داده‌ها تقسیم کرد.

### ۷-۱- سیستم مدیریت اطلاعات پلیس(FIMS)

این ابزار نرم‌افزاری (FIMS) توانسست به آزمایشگاه جنایی پلیس در ایالت ویرجینای غربی در شناسایی الگوهای جرم در بین داده‌های حوادث ذخیره شده در پایگاه داده FIMS (جرم، مظنون و اطلاعات قربانی)، کمک کند. این ابزار به WVSPFL و دیگرمراکز اجرای قانون در بهره‌برداری بهتر منابع و پیش‌بینی فعالیت جنایی ممکن، کمک کرد.

ویژگیهای کلیدی این ابزار عبارتند:

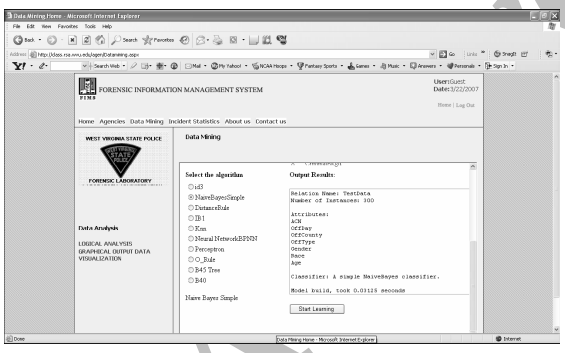
- خلق داده پویا
- شناسایی الگوی جرم با ۹ الگوریتم یادگیری
- تصویرسازی داده خروجی در یک نمونه برای فهم راحت تر
- ایجاد فایل خروجی در قالب XML که یک قالب انتقال داده است. معماری کلی از FIMS2.0 در شکل (۲) نشان داده شده است.




شکل (۲): معماری کلی از FIMS2.0

این ابزار به کاربر اجازه می‌دهد تا اطلاعات مرتبط را از پایگاه داده‌های مختلف بوسیله دستورات SQL جمع آوری کند.

کلاس صفحه اصلی نرم‌افزار شامل متدهایی هستند که با فرم کاربر تعامل دارند. شکل(۳) الگوریتم‌های استخراج سودمند داده کاوی و پنجره نتیجه را نشان می‌دهد.




شکل(۳): انتخاب متدهای داده کاوی

کاربر در این صفحه می‌تواند الگوریتم مناسب را انتخاب نماید، چنانچه کاربر نداند چه الگوریتمی را انتخاب کند سیستم به طور خودکار برحسب درجه دقت و کارایی از روی منبع، بهترین الگوریتم را انتخاب می‌کند، این می‌تواند بوسیله بررسی انتخاب خودکار انجام شود. (Annabathula 2007).

### ۷-۲- سیستم Crime Connect

یک سیستم مبتنی بر وب جهت اشتراک‌گذاری اطلاعات مربوط به جرم می‌باشد که به سازمان‌های قضایی و پلیسی کمک می‌کند تا اطلاعات خود مانند اشخاص مفقود، جرائم جنسی، ابلاغیه‌های رسمی و غیره را در دنیای واقعی به اشتراک بگذارند. این سیستم همچنین امکان جستجو در پایگاه داده جرم حرفه‌ای را برای افراد سازمان فراهم می‌کند تا بوسیله آن از سیستم‌های سایر حوزه‌های قضایی استفاده نموده و توانایی نیروهای پلیس را در حل جرائم افزایش دهند.

در واقع نرم‌افزار Crime Connect یک بستری را فراهم می‌کند تا حوزه‌های قضایی بسرعت و سهولت بتوانند اطلاعات مربوط به جرم را به داخل یک پایگاه داده امن وارد نموده و اطلاعات برخط همراه با داده سایر منابع را به اشتراک بگذارند.

هر آژانس عضو می‌تواند در سیستم Crime Connect داده‌های خود را جهت بهره‌برداری سایر آژانس‌ها به اشتراک گذارد، ارسال فایل کند، داده‌ها را تغییر و یا پاک کنند. این داده‌ها می‌تواند بصورت خودکار تجمیع شده و دستورات و گزارشات خلاصه را نمایش بدهند. همچنین محتوای دستورات و گزارشات را برای بازیابی توسط تمام افسران و واحدهای سیار اجرای قانون، قابل استفاده می‌سازد.

### ۷-۳- نرم‌افزار Crime Point Web

یک راه‌حل نرم‌افزاری مبتنی بر وب جهت تسهیل در به اشتراک‌گذاری اطلاعات، تحلیل و مدیریت اطلاعات برای نمایندگان اجرای قانون و امنیت عمومی است. ایده Crime Point تنها دادن مجوز جهت به اشتراک‌گذاری داده نیست بلکه اطلاعات حاصل. توسط ترکیب داده با تحلیلهای خودکار به راحتی قابل استفاده است، Crime Point Web می‌تواند هر دو اطلاعات، داده و تحلیل را هر کجا که بیشتر، نیاز است بگیرد.

### ۷-۴- سیستم AC

یک سیستم مفید و کاربر پسند است که از همه جنبه‌های داده-کاوی پشتیبانی می‌کند. این یک مجموعه جامع از ابزارها را جهت دسترسی، انتخاب، دست‌کاری و آماده کردن داده را فراهم می‌کند.

AC Knowledge Discovery Engine مبتنی بر جدیدترین تکنولوژیها و تکنیک‌های استنتاجی است که مدل‌های پیشگویی را بصورت خودکار می‌سازد. اینها در شکلی از درخت تصمیم نشان داده می‌شوند. هر فرد به راحتی می‌تواند مدل‌ها را تست و تایید کند و آنها را در حوزه‌هایی از قبیل پیش‌بینی،تقسیم بندی Segmentation، کلاس‌بندی Classification، و تخمین Estimation توسعه دهد. AC یک بسته (Tool Kit) جامع است که به هر فرد اجازه می‌دهد وظایف داده کاوی پیشرفته را اجرا کند و سیستم‌های پشتیبان تصمیم قدرتمند را، توسعه دهد.

### ۷-۵- نرم‌افزار داده کاوی CART

یک ابزار قدرتمند، مبنی بر درخت تصمیم بسیار جذاب است که براحتی، پایگاه‌های داده بزرگ و پیچیده را وارسی می‌کند و الگوها و وابستگی‌های مهم را جستجو و جداسازی می‌کند. این دانش استخراج شده جهت تولید مدل‌های پیشگویی easy-to-graps بکار گرفته می‌شود تا برای کاربردهایی از قبیل Profiling customers، Targeting direct mailing، کشف ارتباطات راه دور، سوءاستفاده از کارت‌های اعتباری و مدیریت ریسک‌های مالی بکار گرفته شود.

(Xindong,Quinlan,Ghosh,Yang,Motoda,McLachlan, Liu, Philip, Zhou, Steinbach, Hand, 2007).

این یک پردازش اولیه بسیار خوب و کامل نسبت به انواع دیگر تکنیک‌های تحلیل داده دارد. برای مثال خروجی CART (مقادیر پیش‌بینی شده) می‌تواند به عنوان یک ورودی استفاده شود تا صحت و درستی پیش‌بینی تکنیک‌های شبکه‌های عصبی و رگرسیون خطی را بهبود دهد.

### ۷-۶- برنامه BRAINCEL

BRAINCEL یک برنامه اکسل افزودنی است که به راحتی قابل استفاده است و پیش‌بینی‌ها را با قدرت شبکه های عصبی تسهیل می‌کند. دانش شبکه عصبی،علم ریاضیات و آمار برای استفاده از این برنامه نیاز نیست. کار با این برنامه تقریباً به عنوان ابزار رگرسیون استاندارد اکسل آسان، اما بسیار قدرتمندتر است.این یک گزینه BESTNET دارد که برنامه را هدایت می‌کند تا بهترین اندازه و قالب مدل ریاضی بعضی پدیده‌ها که رفتار عصبی دارند را پیدا کند. این برنامه، تعداد لایه‌ها و تعداد ماژول‌های پردازش در هر لایه را تغییر خواهد داد.

### ۷-۷- نرم‌افزار BrainMaker

نرم‌افزار شبکه عصبی BrainMaker به افراد اجازه می‌دهد تا ماشین متفکر آنها برای کسب و کار، پیش‌بینی بازاریابی و فروش، موجودی انبار، تعهد، کالای مصرفی که مقدار عرضه آن محدود باشد، قراردادهای آینده، پیش‌بینی آینده، شناسایی الگو، تشخیص عیب، پیش‌بینی نتایج و.. از آن استفاده کنند. این نرم‌افزار اجازه می‌دهد یادگیری شبکه را حفظ کرد و به راحتی یک شبکه‌ای را که بخوبی تست شده است را پیدا کرد. این ابزار از الگوریتم انتشار به عقب استفاده می‌کند و به صورت گرافیکی پیشرفت شبکه و چگونگی یادگیری آن را به خوبی نشان می‌دهد. این ابزار به تعیین میزان دقت و صحت کمک می‌کند.

### ۷-۸- نرم‌افزار CrimeStat III

یک برنامه آماری سه بعدی جهت تجزیه و تحلیل و شناسایی مکان-های حوادث جرم است که توسط Ned Levine & Associate تحت امتیاز 2002_IJ_CX_0007 مطابق استاندارد موسسه ملی دادگستری توسعه یافته است. این برنامه مبتنی بر Windows با فرم‌هایی مبتنی بر GIS است. این برنامه ابزارهای آماری تکمیلی را جهت کمک به نمایندگان اجرای قانون و محققان دادگستری جنایی در نگاشت جرم فراهم می‌کند. این سیستم توسط تعدادی از سازمانهای پلیس در اطراف شهر و همچنین دادگستری جنایی و دیگر محققان در حال بهره‌برداری است. ورودی‌های برنامه مکان-های حوادث(مکان‌های دزدی) در 'dbf'، 'shp'، ASCII و یا ODBC است. انواع آمارهای مقیاسی (سه بعدی) را محاسبه می-کند و نتایج را بصورت گرافیکی در Arc view، Map Info، Atlas GISIM *، Suffer برای ویندوزها و تحلیل سه‌بعدی Arc view نمایش می‌دهد (Annabathula, 2007).

## ۸- داده کاوی در برخی سازمان‌های پلیسی و قضایی

با توجه به اینکه در برخی کشورها از روش‌ها و تکنیک‌های داده-کاوی به منظور پیش‌بینی و پیشگیری از وقوع جرم استفاده کرده‌اند در این بخش به اختصار به برخی از آن تجربیات اشاره می‌شود

### ۸-۱- داده کاوی و تحلیل حوادث ویژه

در ویرجینای غربی حدوداً ۹۰۰ مرکز اجرای قانون وجود دارد که WVSPFL یکی از این مراکز در جنوب ویرجینای غربی است. این مرکز امکاناتی را جهت تحلیل نمونه‌های مربوط به حوادث ویژه فراهم کرده است. در واقع آزمایشگاه جنایی پلیس ایالت ویرجینای غربی از یک ابزار نرم‌افزاری به نام FIMS (سیستم مدیریت اطلاعات جنایی) استفاده کرده است که این ابزار با بکارگیری اصول آمار و الگوریتم‌های داده‌کاوی می‌تواند داده‌های جرم را تحلیل نماید و نتایج منطقی را بدست آورد. نرم‌افزار FIMS از سه قسمت تشکیل شده است که عبارتند از: آمار حوادث، داده کاوی و تصویرسازی داده. با کمک این ابزار، آمار حوادث به مراکز اجرای قانون ارائه و آنان را در تصمیم‌گیری مربوط به بررسی و کنترل حوادث، راهنمایی می‌کند.

برای ساخت یک مدل تحلیلی درست نیاز بود که داده‌ها را از جداول مختلف جمع‌آوری نمود. این ابزار به کاربر اجازه می‌دهد تا اطلاعات مرتبط را از پایگاه داده‌های مختلف بوسیله دستورات SQL جمع آوری کند.

همچنین این ابزار می‌تواند در شناسایی الگوهای جرم در بین داده حوادث ذخیره شده در پایگاه داده FIMS (جرم، مظنون و اطلاعات قربانی)، به آزمایشگاه جنایی پلیس در ایالت ویرجینای غربی کمک کند. این ابزار به WVSPFL و دیگرمراکز اجرای قانون در بهره برداری بهتر منابع و پیش‌بینی فعالیت جنایی ممکن کمک خواهد کرد.

### ۸-۲- داده‌کاوی و بررسی صحنه وقوع جرم

این بخش بررسی می‌کند چگونه تکنیک داده کاوی می‌تواند به کارایی ماموران تحقیق و بررسی صحنه جرم کمک کند. هوش قانونی (مانند اثر انگشت‌ها یا شناسایی DNA) به عنوان یک استاندارد تکنیک قانونی برای تحقیق و کشف یک طیف وسیع از انواع جرم اهمیت دارد. پلیس Nor thamptonshire تمام جرائم را درون یک پایگاه داده رابطه‌ای Oracle ثبت کرد. بخش پشتیبان علمی از سیستم کامپیوتر Trak-X در ثبت و مدیریت همه توابع استفاده کرد (Trak-x بصورت داخلی طراحی شده بود و توسط یک فروشنده نرم‌افزار توسعه یافته بود).

داده قانونی و جرم بین یکم ژانویه ۲۰۰۰ و نوزدهم ژوئیه ۲۰۰۵، برای این طرح آزمایشی استفاده شده بود. مجموعه داده‌ها ادغام شدند تا ۴۷،۷۳۰ رکورد فعالیت فردی مرتبط با صحنه‌های جرم را تولید کنند. دزدی از خانه، دزدی در ساختمانهای تجاری، سرقت و

دزدی از خودروها. این چهار نوع جرائم به چند دلیل زیر انتخاب شده بود:

- استعداد بالقوه جهت آزمایش تعداد زیادی از صحنه های جرم
- آنها جرائم کلیدی برای اغلب نیروهای پلیس و همچنین وزارت کشور UK هستند.
- به طور نمونه تکرارکننده جرائم"recidivist" هستند

اثر انگشتهای بدست آمده از صحنه جرم برای آزمایش به اداره انگشت‌نگاری ارسال شد و متخصصان نتایج را به یکی از ۴ گروه دسته بندی نمودند:

۱- نارسا (Insufficient): کیفیت هیچکدام از اثرانگشت ها به اندازه کافی خوب نیست یا به اندازه کافی عکس چاپی واقعی وجود ندارد، تا دسته بندی بشوند.

۲- حذف شده (Eliminated): اثر انگشتانی که می‌توان به افرادی که دسترسی قانونی به صحنه جرم دارند نسبت داد.

۳- تطبیق یافته (Matched): اثر انگشتانی که با یک شخصی که در پایگاه داده ملی (Ident1) ثبت شده است، مطابق می‌شوند.

۴- واریز نشده (Outstanding): اثر انگشتهایی که صلاحیت کافی برای طبقه‌بندی شدن را دارند اما با یک شخص ثبت شده در (Iden1) تطبیق داده نمی‌شوند.

سپس از یک الگوریتم خوشه‌بندی یادگیری غیر نظارتی K-means استفاده شد تا داده را مدل کند. در یادگیری غیر نظارتی یا خوشه‌بندی مربی صریح وجود ندارد و سیستم خوشه‌ها یا گروه‌بندی‌های طبیعی از الگوهای ورودی میسازد. یادگیری غیر نظارتی بطور نمونه با اشیائ ورودی (Input object) به عنوان یک مجموعه از متغیرهای تصادفی سر و کار دارد. (2006 و Bell )

یافته‌ها نشان داد که ماموران تحقیق می‌توانند مطابق توانایی-شان DNA و اثر انگشتها را از صحنه جرم جمع‌آوری کنند. هم چنین توانایی آنها را در پیش‌بینی را نشان داد، کدام یک از صحنه‌های جرم بهترین فرصت جمع‌آوری نمونه‌های قانونی را ارائه خواهد کرد در حالیکه با توانایی حقیقی آنها ارتباط ندارند.

### ۸-۳- داده‌کاوی و دوباره قربانی شدن

یک گروه تحقیقاتی متمرکز در دانشگاه ساندرلند (مرکز سیستمهای قابل تطبیق) از فرصت صنعتی بدست آمده استفاده کرد تا مزایای واقعی فناوری محاسبات پیشرفته در مناطقی که شامل نظارت موقعیت، کنترل هوشمند و اکتشاف دانش در پایگاه داده است، را بدست آورند. آن مرکز از سوی سازمان تجارت و صنعت مامور بود تا یک نرم‌افزار تحت عنوان SMART برای تصمیم‌گیران (SSDM[1]) آن سازمان پیاده سازی کند. در نتیجه سه تا از کاربردهای ضروری در مناطق زیر توسعه یافتند (Oatley , MacIntyre , EWART, and Mugambi).

۱- شناسایی داده‌های تکراری در پایگاه داده

۲- جستجو و تحقیق گروه‌های کسب و کار– این کار شامل پیش پردازش داده، انتقال و تفسیرداده می‌باشد

۳- تحقیق و بررسی تکرار قربانی در یک پایگاه داده جرم - این کار شامل پیش پردازش داده، انتقال، داده‌کاوی و تفسیر داده است.

از طرفی نیروهای پلیس به پدیده‌های تکرار قربانی علاقه‌مند بودند. در نتیجه در دو تا از آن پروژه‌ها ابزارهایی جهت استفاده مستقیم در فرآیندهای داده کاوی توسعه دادند. (OATLEY, 2002 )

مفهوم دوباره قربانی شدن در ابتدا توسط Sparks ذکر شد و بدین معناست که مکان‌هایی که در آن یک‌مرتبه 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 آماده‌سازی نرم‌افزار، Ewart، Inglis و Wilbert کم شدن فاصله زمانی بین دزدی پی درپی و دوباره قربانی شدن در یک ملک را ثابت کردند.

Peas نتیجه می‌گیرد که بهترین تحلیل به توجه همزمان به نحوه قربانی کردن، مکان و اطلاعات مرتکب جرم نیاز دارد. از این رو در بخش جزئیات با نیروی پلیس محلی کار می‌کند تا از عاملهای بالقوه بازجویی کند و با استفاده از مدل‌سازی کامپیوتر و داده کاوی کوشش کردند تا تعیین کنند آیا احتمال دوباره قربانی شدن کسی در یک جرم وجود دارد یا خیر؟ همچنین موضوعات خط سیر زمان را دنبال کردند و مشاهده کردند که چه زمانی احتمال تکرار جرم کم می‌شود. این مباحث به روشنی به سمت مدیریت منابع پلیس منشعب شد.

### ۸-۴- داده‌کاوی و حوادث تیر اندازی

همچنان که Colleen McCue به عنوان یک متخصص و یک افسر سابق پلیس VA (مدیر پروژه واحد تحلیل جرم در سازمان

1. SMART Software for Decision Makers

پلیس Richmond ،VA ) عنوان می نماید:"داده‌کاوی زمانی که در تحلیل جرم تاکتیکی بکار برده می شود یک ابزار اکتشاف دانش هست 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 می-کند"(McCue,2003).

علاوه بر ابزارهای نقشه برداری جرم و شبکه های عصبی و پیش‌بینی و تکنولوژی‌های طرح‌یابی، راه حل‌های تحلیل پیوند استفاده شدند تا دیدگاه با ارزشی را برای افسران فراهم آورند. برای مثال سازمان پلیس VA(Richmond) تحت دستور Colleen McCue، تکنیک-های داده‌کاوی و برنامه‌های کاربردی بسیاری را اجرا کرد. آن سازمان بین‌المللی ابزارهای SPSS و RTI را بکار برد تا حوادث تیراندازی اتفاقی را پیش‌بینی کند و حوادث تیراندازی شب عید سال نو ۲۰۰۳ در آن شهر ۴۷٪ بیش از سال قبل کاهش داد (Leon , 2005).

بکارگیری تحلیل‌های درخت تصمیم و تکنیک‌های آنها به افسران کمک کرد تا با سرعت بیشتری در برابر آن موقعیت در طول ۴۸ ساعت زمان بحرانی واکنش نشان دهد. تلاش دیگر در بیت لحم Bethlehem و پنسیلوانیا Pennsylvania، توسط William Pottenger از دانشگاه Lehigh ارائه شد و یک راه‌حلی را توسعه داد که D-Hotm نامیده شد. یک مخفف برای Distributed Higher-Order Text Mining. یک قسمت از سیستم آنها تبدیل خود کار داده متنی بدون ساخت به داخل یک پایگاه داده ساخت یافته را فراهم کرد که در سازمان پلیس بیت لحم در واحد تحقیقات آنها تست شد.

علاوه بر آن، شبکه یکپارچه فلوریدا برای سیستم بازیابی و تبادل داده (Finder) توسط دانشگاهی از فلوریدای مرکزی توسعه داده شد که قادر است بیش از ۱۲۰ آژانس را در داخل ایالت فلوریدا هماهنگ کند (Nichols.L.J, 2004).

### ۸-۵- داده‌کاوی و سرقت‌های مسلحانه

بیشتر کارشناسان توافق دارند زمانیکه سرقتهای مسلحانه همراه با تهاجم‌های خطرناک است، آنگاه خسارت وارد به قربانی وخیم‌تر است. بنابراین داده سرقت مسلحانه را در ریچموند؛ ویرجینا، در یک دزدی احتمالی آزمایش کردند تا خشونت‌های مسلحانه را محدود کنند. با استفاده از SPSS Clementine، یک مدل توسعه دادند که فاکتورهای وابسته به یک احتمال فزاینده از یک سرقت مربوط به حملات مسلحانه را مشخص کرد. سپس این نتایج را در یک نقشه برای استفاده توسط نیروهای گشت زنی و واحد تاکتیکی گسترش دادند.

نتایج نشان داد که مناطق با "ریسکهای بالا" مشابه هم هستند، اما دزدی‌های مسلحانه بصورت یکسان در همه جا توزیع شده نیست. در ناحیه‌های مشخص شده سرقت‌های مسلحانه لزوماً به یک درجه افزایش یافته وابسته نبود، بلکه آنها به یک احتمال فزاینده وابسته بودند. این یافته، اختلاف‌های دقیق بین بسامد نسبی از جرم و احتمال تهدیدات جدی را مشخص کرد. مناطق با ریسک بالا شامل، مناطق جغرافیایی کوچکتری در شهر می‌شدند. با تعین مناطق خطرخیز امکان افزایش گشت‌زنی و گسترش واحدهای تاکتیکی در یک زمان خیلی کوتاه به صورت هدفمند، فراهم گردید.

منابع امروزی استراتژی‌هایی را فراهم می‌کند، که تصمیم‌گیری را تسهیل می‌کنند و توانایی ما را بصورت موثر افزایش می‌دهند تا منابع کمیابی که برای آژانس‌های اجرای قانون ارزش زیادی را دارد گسترش دهیم (MCCUE.C, 2003).

### ۸-۶- داده‌کاوی و جرائم خشونت آمیز

بکارگیری فناوری هوش مصنوعی و مدل‌های دقیق می‌تواند توسعه یابد تا در پیش‌بینی اتفاقات آینده بکار گرفته شود. آگاهی و بینش در خصوص حوادثی که احتمالاً در آینده اتفاق خواهند افتاد یک فرصت منحصر بفرد و حرفه‌ای را به ماموران اجرای قانون می-دهد تا بصورت سریع واکنش نشان دهند. واضح است که نشان دادن تاکتیک‌های حرفه‌ای در مبارزه با جرایم از لحاظ اجتماعی بسیار مفید و موثر است.

در گذشته این تکنیک‌ها در مراکز تحقیقات دانشگاهی و بزرگترین آژانس‌های فدرال بصورت انحصاری بود، اما اکنون این ابزارها در محیط Desktop کامپیوترهای شخصی قابل دسترس هستند.

سازمان پلیس ویریجینا، از این فناوری به منظور کنترل‌های محلی استفاده کرد، و نظریه‌های خودش را با اداره دادستان کل ایالات متحده در بخش شرقی از ویریجینا تحت عنوان پروژه تحقیقاتی PSN و ارزیابی همکاران به اشتراک گذاشت. پروژه PSN به عنوان بخشی از برنامه ارزیابی وتحقیقات، از داده کاوی و تحلیل-های پیش‌گویانه در رسیدگی کردن به جرائم خشونت‌آمیز مربوط به تیر اندازی استفاده کرد. این رویکرد در واقع فرصت شرح کلی، درک و پیش‌بینی جرائم خشونت‌آمیز را فراهم کرد. در این پروژه از این

مدل‌ها استفاده شد تا حوادث آینده را پیش‌بینی کنند و با بکارگیری و صف آرایی افراد پلیس از جرائم خشونت‌آمیز، جلوگیری کنند(Nichols.L.J,2004).

### ۸-۷- داده‌کاوی و حملات تروریستی

پس از حمله تروریستی ۱۱ سپتامبر سازمانهای FBI،CIA و دیگر آژانسهای فدرال تصمیم گرفتند اطلاعات داخلی و خارجی مربوط به حوزه امنیت را جمع‌آوری کنند تا بتوانند از حملات تروریستی جلوگیری کنند. این تلاشها موجب ایجاد انگیزه در مقامهای محلی گردید تا به صورت دقیق‌تر جرائم قضایی حوزه خود را کنترل کنند.

چالش اصلی تمام مجریان قانون و سازمانهای گردآوری اطلاعات که با آن مواجه شدند دقت و موثر بودن میزان فزاینده تحلیل داده جرم است. بعنوان مثال حل نمودن توطئه‌های پیچیده اغلب مشکل هستند زیرا اطلاعات مظنونین می‌تواند از نظر جغرافیایی و گستردگی در دوره‌های زمانی طولانی متفاوت باشد. همچنین تشخیص جرائم مجازی می‌تواند سخت باشد زیرا ترافیک شبکه و تراکنش های برخط تکرار شونده مقدار زیادی داده تولید می‌کند که تنها یک بخش کوچکی از فعالیت‌های غیر قانونی را تشریح می‌کند.

داده کاوی در این حوزه برای ماموران تحقیق که فاقد آموزش در حوزه تحلیل داده و اکتشاف دانش از پایگاه داده می باشند به عنوان یک ابزار قدرتمند سریع و موثر مطرح است. کامپیوترها می‌توانند هزاران دستورالعمل را در چند ثانیه پردازش کنند و زمان زیادی را ذخیره کنند.

### ۸-۸- داده‌کاوی و سرقت از منازل

نیازهای قانونی باعث ایجاد تعهد برای پلیس UK و شرکاء محلی گردید تا در جهت بازبینی جرم و بی نظمی اقدام و با ایجاد خط مشی‌ها مبنی بر این بازبینی‌ها، یک وسیله مفید بسوی نگاشت و تحلیل از داده جرم فراهم کنند (McMullan & Radcliff ،۲۰۰۱).

لذا با سه نیروی پلیس UK، شامل Sunderland west، Cleveland، West Midland کار را شروع کردند.

از آنجا که داده West Midland police (wmp) بیش از همه جدید و بیش از همه رشد یافته است لذا از طریق یک همکاری اولیه که پروژه OVER نامیده شده کار شروع شد. پروژه Over در سال۲۰۰۰، پلیس west Midland را درگیر کرد و در مرکز سیستم‌های قابل تطبیق و گروهی از روانشناسان از دانشگاه Sunderland قرار گرفت. هدف اصلی کمک کردن به اداره و کنترل میزان جرم دزدی از منازل مسکونی (BDH[1]) با بکارگیری سیستم-های پشتیبانی تصمیم بود. پلیس به نرم‌افزاری نیاز داشت که به آنها در موارد زیر کمک کند:

۱- هدف‌یابی منابع برای خط مشی‌های کشف و جلوگیری کننده بطور بسیار موثر

۲- شناسایی کردن داده مهم تا در یک حادثه بتوان پرسنل را هدایت و کارآیی زمان را افزایش یابد

۳- فراهم کردن اطلاعات درباره طراحی سیستم‌ها که این داده سخت (قانونی) و داده نرم (اطلاعات صحنه جرم) و اطلاعات هوشمندی پلیس را یکی کند.

پروژه Over، توسعه‌ای از تکنیک‌های پشتیبانی تصمیم مبنی بر پیرامون پکپارچگی روانشناسی قانونی و داده کاوی، آمار و تکنیکهای عمومی علم کامپیوتر شد.

(OATLEY. G, EWART. B and ZELEZNIKOW. J ,2006).

تنوع تکنیک‌های داده کاوی استفاده شده در این پروژه (کلاس‌بندی و قواعد پیوستگی، شبکه عصبی، خوشه‌بندی، تحلیل-های بقاء و شبکه‌های Bayesian belief nets، دلایل موضوع محور، هستی شناسی و برنامه‌نویسی منطقی) پلیس را در کشف مرتکبین دزدی از منازل مسکونی، با یک نرخ اکتشاف ضعیف حمایت کرد.

### ۸-۹- داده‌کاوی و جرائم مجازی

اطلاعات مظنونین می‌تواند از نظر جغرافیایی و گستردگی دوره‌های زمانی طولانی، متفاوت باشد. همچنین کشف جرائم مجازی می‌تواند سخت باشد زیرا ترافیک شبکه شلوغ و تراکنش‌های درون خطی تکرار شونده مقدار زیادی داده تولید می‌کند، که تنها بخش کوچکی از این فعالیت‌های غیر قانونی را تشریح می‌کند. داده کاوی یک ابزار قدرتمند ارائه می‌دهد که ماموران تحقیق جنایی که ممکن است فاقد آموزش باشند را قادر می سازد به عنوان تحلیل-گران داده در اکتشاف پایگاه داده‌های بزرگ بسرعت و بطور موثر تحلیل نمایند. علاوه بر این هزینه‌های نصب و راه‌اندازی (کارکرد)

1 . burglary from a dwelling house

نرم‌افزار اغلب کمتر از استخدام و آموزش پرسنل است. کامپیوترها همچنین نسبت به ماموران تحقیق انسانی کمتر در معرض اشتباه هستند، مخصوصاً آنهایی که ساعتهای طولانی کار می‌کنند. محققان دانشگاه Arizona در همکاری با سازمان پلیس Tuscon و Phoeni بعد از ۱۹۹۷ مشغول به هدایت این موضوع هستند(OATLEY, EWART and ZELEZNIKOW ,2006).

## ۹- نتیجه‌گیری

جرائم یک ناهنجاری اجتماعی هستند و جوامع هزینه زیادی را به گونه‌های مختلف بابت آن می‌پردازند. مهمترین هدف‌های دولتها بکارگیری یک سیاست جنایی کارآمد، از بین بردن فرصت‌های ارتکاب جرم، پیشگیری وضعی از وقوع جرم و برخورد با مجرمین است. در این راستا سازمان‌های پلیسی از روش‌ها و تکنیک‌های مختلفی مانند (افزایش نیروهای پلیس، دوربینهای مدار بسته و...) استفاده کردند.

با توسعه و گسترش فناوری اطلاعات در سازمان‌ها و ایجاد بانک-های اطلاعاتی، داده‌کاوی نیز به عنوان یک ابزار نرم‌افزاری قدرتمند و به مراتب کم هزینه‌تر و کارآمدتر در اختیار سازمان‌های پلیس قرار گرفته است. با استفاده از داده کاوی می‌توان الگوهای جرم را شناسایی کرد تا جرائم را پیش‌بینی کرده و از وقوع آن پیشگیری نمود.

در دنیا ابزارهای مختلفی برای تحلیل داده‌های جرم به کار گرفته شده است که برخی از آنها عبارتنداز CART، Crime Connect، Crime Point Web، BRAINCEL، AC، BrainMaker، CrimeStat III.

اما برای استفاده از این ابزارها نیاز است که در یک چهارچوب معین فرآیندهای اصلی داده‌کاوی را طی نمود تا بتوان به مدل‌های دقیق‌تر برای جرایم و کشف الگوهای ناشناخته در بانکهای اطلاعاتی دست یافت. بدین ترتیب پلیس می‌تواند بصورت موثر و کارآمد نیروهای خود را در مناطق بکار گرفته و با اقتدار به اداره و کنترل حوزه استحفاظی خود بپردازد.

همانطور که در قسمت‌های قبل اشاره شد داده کاوی دارای قابلیت‌ها و امکاناتی می باشد که برخی از آنها عبارتند از:

۱- کشف ارتباط بین داده‌ها: از مهمترین قابلیت‌های داده کاوی شناسایی و کشف ارتباطات ناشناخته در بانک اطلاعاتی می باشد که به دست آوردن آن برای کاربران تقریباً غیر ممکن است.

۲- امکان تصویرسازی داده‌ها: به کمک GIS و تکنیکهای داده کاوی می توان نقاط حادثه خیز و خطرناک و همچنین توزیع جغرافیای جرم را برحسب سن، جنسیت، مکان و نوع جرم به صورت نمودار نمایش داد و استراتژی سازمان را تدوین یا در صورت لزوم تغییر داد.

۳- تحلیل حجم بسیار زیاد داده‌ها: به کمک تکنیک‌های داده کاوی می‌توان داده‌های فراوان موجود در بانک‌های اطلاعاتی را تحلیل نمود که این کار به صورت دستی و بدون کمک گرفتن از تکنیک‌های داده کاوی سخت و دشوار و زمان بر می‌باشد.

۴- و...

در کنار این قابلیتها نقدهایی به داده کاوی جرم وارد است که برخی از آنها عبارتند از:

۱- در صورت وجود داده‌های شلوغ و کثیف (پالایش نشده) در بانک اطلاعاتی، عملاً بهره‌برداری از ابزارهای داده کاوی بیهوده است.

۲- به منظور استفاده از الگوریتم‌های داده‌کاوی نیاز است که افراد آموزش‌های تخصصی مورد نیاز را در این حوزه دیده باشند.

۳- استفاده از تکنیکهای داده کاوی در دنیای واقعی منوط به انجام تستهای پیچیده در محیط آزمایشگاهی و مقایسه آن با نتایج استخراج شده به صورت دستی توسط کارشناسان خبره جهت تایید صحت و دقت پیش‌بینی تکنیک‌ها می باشد.

۴- از آنجایی که معمولاً یک الگوریتم به تنهایی برای کشف الگو کافی نمی‌باشد نیاز است تا بصورت ترکیبی از دو یا چند الگوریتم برای افزایش صحت و دقت پیش‌بینی مورد نیاز می باشد که این کار زمان بر است.

## مراجع


1- Annabathula.R (2007). " A WEB-BASED TOOL FOR ANALYSIS OF CRIME LABORATORY DATA"

2- Bell. C. (2006)." Concepts and possibilities in forensic intelligence "

3- Canhoto.A.I(2007)." Profiling behaviour: the social construction of categories in the detection of financial Crime"

4- Castro.V.E and Lee.I."Data Mining Techniques for Autonomous Exploration of Large Volumes of Geo-referenced Crime Data"

5- Chen. H, Chung. W , Jie. J,Gang Wang Yi Qin Xu.G.W , Chau. M." Crime Data Mining:A General Framework and Some Examples"

6- Corporation. T.C(1999) " Introduction to Data Mining and Knowledge Discovery "
7- Jeffery W. Seifer.(2004). " Analyst in information science and Technology Policy, Data Mining: An Overview"
8- Giles. C. Oatleya, Brian W. Ewart. (2003). " Crimes analysis software: 'pins in maps', clustering and Bayes net prediction"
9- Han.J, and Kamber.M.(2001)" Data Mining: Concepts and Techniques"
10- Havenstein H(2006) " Data Analysis May Help LAPD Fight Terrorism "
11- Hoffman. T. (2007). " CASE CRACKERS "
12- Keogh.E, Lonardi.S, Ratanamahatana. C.A.(2005)." Towards Parameter-Free Data Mining "
13- Lawlor.M. (2007)." Smart Companies Dig Data"
14- Li. S. T , Tsai1.F.C, Kuo.S.C, Cheng.Y.C." A Knowledge Discovery Approach to Supporting Crime Prevention"
15- MCCUE.C. (2003). "Data mining and crime analysis in the Richmond "
16- Oatley.G , MacIntyre.J , EWART.B, Mugambi.E." SMART Software for decision makers KDD experience "
17- 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 "
18- OHM. P.(2008)"The Olmsteadian Seizure Clause: The Fourth Amendment and The Seizure of Intangible Property Police Department"
19- REPORT (1998)."Data mining makes its mark at the scene of the crime"
20- Seifert.J.W.(2004) "Analyst in information science and Technology Policy, Data Mining: An Overview "
21- Xindong.W, Kumar.V, Quinlan.J.R, Ghosh.J. Yang.Q. Motoda.H, McLachlan.G.J, Ng.A, Liu.B, Philip S. Yu, Zhou.Z. Steinbach.M. Hand.D.J,(2007)." Top 10 algorithms in data mining"