

صلى الله عليه وسلم



بسمه تعالی

تاییدیه اعضای هیات داوران حاضر در جلسه دفاع از پایان نامه

آقای احمد ابراهیمی پایان نامه ۶ واحدی خود را با عنوان ارائه مدلی برای تحلیل و کشف جرم با بهره گیری از روش های داده کاوی - مورد کاوی در بانک داده های بجرمین موجود در آگاهی ناجا در تاریخ ۱۳۸۹/۸/۵ ارائه کردند. اعضای هیات داوران نسخه نهایی این پایان نامه را از نظر فرم و محتوا تایید کرده و پذیرش آنرا برای تکمیل درجه کارشناسی ارشد مهندسی صنایع - مهندسی فناوری اطلاعات سیستمهای اطلاعاتی پیشنهاد می کنند.

عضو هیات داوران	نام و نام خانوادگی	رتبه علمی	امضا
استاد راهنما	دکتر رضا سمیع زاده	استادیار	
استاد مشاور	دکتر عباس آسوشه	استادیار	غایب
استاد ناظر	دکتر سید کمال چهارسوقی	دانشیار	
استاد ناظر	دکتر فرامک زندی	استادیار	
مدیر گروه (یا نماینده گروه تخصصی)	دکتر سید کمال چهارسوقی	دانشیار	

آیین نامه چاپ پایان نامه (رساله) های دانشجویان دانشگاه تربیت مدرس

نظر به اینکه چاپ و انتشار پایان نامه (رساله) های تحصیلی دانشجویان دانشگاه تربیت مدرس ، مبین بخشی از فعالیتهای علمی - پژوهشی دانشگاه است بنابراین به منظور آگاهی و رعایت حقوق دانشگاه، دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می شوند:

ماده ۱: در صورت اقدام به چاپ پایان نامه (رساله) ی خود، مراتب را قبلاً به طور کتبی به «دفتر نشر آثار علمی» دانشگاه اطلاع دهد.

ماده ۲: در صفحه سوم کتاب (پس از برگ شناسنامه) عبارت ذیل را چاپ کند:

«کتاب حاضر، حاصل پایان نامه کارشناسی ارشد / رساله دکتری نگارنده در رشته فناوری اطلاعات است که در سال ۱۳۸۹ در دانشکده فنی و مهندسی دانشگاه تربیت مدرس به راهنمایی / جناب آقای دکتر رضا سمیع زاده ، مشاوره جناب آقای دکتر عباس آسوشه از آن دفاع شده است.»

ماده ۳: به منظور جبران بخشی از هزینه های انتشارات دانشگاه، تعداد یک درصد شمارگان کتاب (در هر نوبت چاپ) را به «دفتر نشر آثار علمی» دانشگاه اهدا کند. دانشگاه می تواند مازاد نیاز خود را به نفع مرکز نشر در معرض فروش قرار دهد.

ماده ۴: در صورت عدم رعایت ماده ۳، ۵۰٪ بهای شمارگان چاپ شده را به عنوان خسارت به دانشگاه تربیت مدرس، تأدیه کند.

ماده ۵: دانشجو تعهد و قبول می کند در صورت خودداری از پرداخت بهای خسارت، دانشگاه می تواند خسارت مذکور را از طریق مراجع قضایی مطالبه و وصول کند؛ به علاوه به دانشگاه حق می دهد به منظور استیفای حقوق خود، از طریق دادگاه، معادل وجه مذکور در ماده ۴ را از محل توقیف کتابهای عرضه شده نگارنده برای فروش، تامین نماید.

ماده ۶: اینجانب احمد ابراهیمی دانشجوی رشته فناوری اطلاعات مقطع کارشناسی ارشد تعهد فوق و ضمانت اجرایی آن را قبول کرده، به آن ملتزم می شوم.

نام و نام خانوادگی:

تاریخ و امضا:



دستورالعمل حق مالکیت مادی و معنوی در مورد نتایج پژوهشهای علمی

دانشگاه تربیت مدرس

مقدمه: با عنایت به سیاست‌های پژوهشی دانشگاه در راستای تحقق عدالت و کرامت انسانها که لازمه شکوفایی علمی و فنی است و رعایت حقوق مادی و معنوی دانشگاه و پژوهشگران، لازم است اعضای هیات علمی، دانشجویان، دانش‌آموختگان و دیگر همکاران طرح، در مورد نتایج پژوهشهای علمی که تحت عناوین پایان‌نامه، رساله و طرحهای تحقیقاتی که با هماهنگی دانشگاه انجام شده است، موارد ذیل را رعایت نمایند:

ماده ۱- حقوق مادی و معنوی پایان‌نامه‌ها / رساله‌های مصوب دانشگاه متعلق به دانشگاه است و هرگونه بهره‌برداری از آن باید با ذکر نام دانشگاه و رعایت آیین‌نامه‌ها و دستورالعمل‌های مصوب دانشگاه باشد.

ماده ۲- انتشار مقاله یا مقالات مستخرج از پایان‌نامه / رساله به صورت چاپ در نشریات علمی و یا ارائه در مجامع علمی باید به نام دانشگاه بوده و استاد راهنما مسئول مکاتبات مقاله باشند.

تبصره: در مقالاتی که پس از دانش‌آموختگی بصورت ترکیبی از اطلاعات جدید و نتایج حاصل از پایان‌نامه / رساله نیز منتشر می‌شود نیز باید نام دانشگاه درج شود.

ماده ۳- انتشار کتاب حاصل از نتایج پایان‌نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با مجوز کتبی صادره از طریق حوزه پژوهشی دانشگاه و بر اساس آئین‌نامه‌های مصوب انجام می‌شود.

ماده ۴- ثبت اختراع و تدوین دانش فنی و یا ارائه در جشنواره‌های ملی، منطقه‌ای و بین‌المللی که حاصل نتایج مستخرج از پایان‌نامه / رساله و تمامی طرح‌های تحقیقاتی دانشگاه باید با هماهنگی استاد راهنما یا مجری طرح از طریق حوزه پژوهشی دانشگاه انجام گیرد.

ماده ۵- این دستورالعمل در ۵ ماده و یک تبصره در تاریخ ۱۳۸۴/۴/۲۵ در شورای پژوهشی دانشگاه به تصویب رسیده و از تاریخ تصویب لازم‌الاجرا است و هرگونه تخلف از مفاد این دستورالعمل، از طریق مراجع قانونی قابل پیگیری خواهد بود.

احمد ابراهیمی

تاریخ و امضا





دانشگاه تربیت مدرس

دانشکده فنی و مهندسی

پایان نامه برای دریافت درجه کارشناسی ارشد

رشته صنایع گرایش مهندسی فناوری اطلاعات

ارائه مدلی برای تحلیل و کشف جرم با بهره گیری از روش های داده کاوی

(مورد کاوی در بانک داده های مجرمین موجود در آگاهی ناجا)

نگارنده

احمد ابراهیمی

استاد راهنما : دکتر رضا سمیع زاده

استاد مشاور : دکتر عباس آسوشه

آبان ماه ۱۳۸۹

تقدیم به پدر و مادرم برای همه محبت‌هایشان

همسرم برای همراهی و بردباریش

و دختر عزیزم که امید آینده من است

تشکر و قدردانی

« من لم یشکر المخلوق لم یشکر الخالق »

از تمامی عزیزانی که در این تحقیق مرا یاری نمودند و بدون تردید بدون حمایت های آنها این

تحقیق شکل نمی گرفت، صمیمانه تشکر و قدردانی می نمایم.

از زحمات بی شائبه استاد راهنمای ارجمندم جناب آقای **دکتر رضا سمیع زاده** و همچنین

استاد محترم مشاور جناب آقای **دکتر عباس آسوشه** که در مدت انجام این پروژه با رهنمود های

ارزشمندشان به بنده کمک فراوانی نمودند، همچنین جناب آقای مهندس سهلانی به جهت همراهی و

راهنمایی ایشان صمیمانه تقدیر و تشکر می نمایم.

چکیده

حجم انبوه داده‌ها و اطلاعات مرتبط با جرایم در ادارات 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 کارشناسان صحنه جرم با تجربه به کار گرفته می‌شوند، اولاً) نیازمند صرف زمان و هزینه انسانی بسیار زیادی باشند. ثانیاً) به دلیل میزان دخالت بالای عامل انسانی در تصمیم‌گیری‌ها، قادر به لحاظ کردن تمام فاکتورهای تاثیرگذار در یک جرم و ارتباطات بین آنها نبوده و باعث کند نمودن مراحل کشف جرم می‌شوند. چنین شرایطی، ضرورت به کارگیری یک روش نظام‌مند جهت کشف و تحلیل جرایم را بیش از پیش نمایان می‌سازد.

تهیه و ارائه یک الگوی برای سیستم گزارشات پلیس جهت تحلیل و کشف جرم، بر اساس تکنیک‌های داده‌کاوی می‌تواند راهگشای دو چالش فوق باشد. برای این منظور در این تحقیق سعی شده است برای تحلیل و کشف جرم از طریق تشخیص شباهت جرایم با یکدیگر، بر اساس الگوریتم‌های داده‌کاوی دو مرحله عمده ارائه گردد.

۱- ابتدا باید معیارهای تشابه را پیدا نمود که بدین منظور می‌توان از روش‌های معمول در انتخاب ویژگی همچون خوشه‌بندی فازی با متد GKClust و خوشه‌بندی با K-means استفاده نمود.

۲- سپس باید با استفاده از تکنیک‌هایی چون KNN^۱ جهت یافتن دسته‌هایی از جرایم و یا مجرمین با معیارهای تشابه نزدیک به جرم رخ داده فعلی استفاده کرد. برای تشخیص معیارهای تشابه، وزن‌دهی ویژگی‌های مورد استفاده، بسیار مؤثر می‌باشد. برای دستیابی به وزن‌دهی مناسب الگوریتم PCA^۲ کارا به نظر می‌رسد.

کلید واژه‌ها: داده‌کاوی - تحلیل جرم - تطبیق جرم - کشف جرم

^۱ K-nearest neighbour and memory-based reasoning (MBR)
^۲ Principal Component Analysis

فهرست مطالب

صفحه	عنوان
ج	فهرست جدولها
ح	فهرست شکلها
۱	فصل اول
۱	معرفی و کلیات
۲	۱-۱. مقدمه
۳	۲-۱. اهمیت موضوع
۶	۳-۱. انگیزه پژوهشگر
۶	۴-۱. سوالات تحقیق
۶	۵-۱. اهداف تحقیق
۷	۶-۱. تاریخچه موضوع
۸	۷-۱. مساله تحقیق
۸	۸-۱. محدودیت های تحقیق
۹	۹-۱. روش گردآوری اطلاعات
۹	۱۰-۱. تعاریف و اصلاحات
۹	۱۱-۱. خلاصه و نتیجه
۱۱	فصل دوم
۱۱	جرم شناسی و تحلیل جرم
۱۱	اصول و مفاهیم اولیه
۱۲	۱-۲. مقدمه
۱۴	۲-۲. تعریف جرم
۱۵	۳-۲. دسته بندی انواع جرم
۱۶	۴-۲. انواع متغی های جرم
۱۸	۵-۲. تاریخچه جرم شناسی
۲۰	۶-۲. تحلیلی جرم
۲۱	۷-۲. تفاوت جرم شناسی و تحلیلی جرم
۲۳	۸-۲. مزایای تحلیلی جرم
۲۴	۹-۲. خلاصه و نتیجه
۲۵	فصل سوم
۲۵	داده کاوی و تکرک های آن

۲۵	در تحلیلی و کشف جرم
۲۶	۱-۳. مقدمه
۲۶	۲-۳. کاربردهای داده کاوی
۲۷	۳-۳. ارتباط داده کاوی با سای علوم
۲۷	۴-۳. تفاوت داده کاوی و آنالیز آماری
۲۹	۵-۳. مراحل داده کاوی
...۲.۹.	۱-۵-۳. جمع آوری و پاک سازی داده ها
...۳.۱.	۲-۵-۳. آماده سازی داده ها
...۳.۱.	۳-۵-۳. پاکسازی داده ها
...۳.۲.	۴-۵-۳. نرمال سازی داده ها
...۳.۳.	۵-۵-۳. انتخاب داده ها
...۴.۱.	۶-۵-۳. یکپارچه سازی داده ها:
...۴.۱.	۷-۵-۳. تبدیل داده ها:
...۴.۱.	۸-۵-۳. کاهش در داده ها:
...۴.۲.	۹-۵-۳. ارزیابی الگوها:
...۴.۲.	۱۰-۵-۳. بازنمایی دانش :
۴۳	۶-۳. تکررک های داده کاوی
۴۴	۷-۳. خوشه بندی
...۴۵.	۱-۷-۳. خوشه بندی فازی
۴۸	۸-۳. مرور تکررک های خوشه بندی
...۴۸.	۱-۸-۳. روش های تفکیکی
۴۹	۱-۱-۸-۳. الگوریتم K-means
...۵.۱.	۲-۸-۳. الگوریتم های سلسله مراتبی
...۵.۲.	۳-۸-۳. روش های متکی بر چگالی
۵۲	۱-۳-۸-۳. الگوریتم k نزدیکترین همسایه
۵۵	۲-۳-۸-۳. محاسبه آنالیز اجزای اصلی

..۵۷. ۴-۸-۳. روش‌های متکی بر مدل
..۵۷. ۵-۸-۳. روش‌های متکی بر گرید
۵۸ ۹-۳. داده کاوی از دیدگاه تحلیل جرم
۶۰ ۱۰-۳. خلاصه و نتیجه
۶۲ فصل چهارم
۶۲ ارائه مدل و پیاده سازی
۶۳ ۱-۴. مقدمه
۶۳ ۲-۴. روش تحقیق
۶۴ ۳-۴. محیط پیاده سازی
۶۴ ۴-۴. روش پیشنهادی
..۶۷. ۱-۴-۴. مرحله اول شامل تهیه داده از بانک های مجرمین سرقت های مسلحانه
..۶۸. ۲-۴-۴. مرحله دوم آماده سازی داد ها
..۶۹. ۳-۴-۴. خواندن داده های آماده شده
..۷۰. ۴-۴-۴. خواندن داده های جرم اتفاق افتاده فعلی
..۷۲. ۵-۴-۴. نرمال سازی داده ها
..۷۲. ۶-۴-۴. اجرای الگوریتم فازی روی داد ها برای خوشه بندی نرم
..۷۳. ۷-۴-۴. اجرای الگوریتم k-means برای خوشه بندی
..۷۶. ۸-۴-۴. اجرای آنالیز مولفه اصلی
..۷۷. ۹-۴-۴. اجرای عمل تست و Training
..۷۷. ۱۰-۴-۴. اجرای الگوریتم K نزدیکی همسایگی
۷۸ ۵-۴. ارائه پیشنهاد
۷۸ ۶-۴. ارزیابی کیفی مدل ارائه شده
۸۱ ۷-۴. خلاصه و نتیجه گیری
۸۲ فصل پنجم
۸۲ نتیجه گیری
۸۳ ۱-۵. مقدمه
۸۳ ۲-۵. مروری بر فصول پیشین و خلاصه کارهای انجام شده
۸۴ ۳-۵. دستاوردهای تحقیق

۴-۵	ملاحظات تحقیق	۸۵
۵-۵	زمینه های تحقیقاتی آتی	۸۵
۵-۶	نتیجه گیری	۸۶
۹۰	مراجع:	۹۰

فهرست جدولها

- جدول ۱-۲ خلاصه ای از تحقیقات انجام شده در زمینه کشف جرم ۱۳
- جدول ۲-۲ متغیرهای جرم در حیطه ی تحلیل جرم ۱۷
- جدول ۳-۲ متغیرهای مجهول جرم در حیطه ی تحلیل جرم ۱۷
- جدول ۴-۲ متغیرهای جرم در حیطه ی اطلاعات شخصی و سوابق مجرم ۱۷
- جدول ۱-۳ تفاوت های کلی روش های آماری و دیگر روش های داده کاوی ۲۸
- جدول ۲-۳ اطلاعات احصاء شده از بانک اطلاعاتی مجرمین سرقت های مسلحانه (به همراه دو رکورد نمونه) ۳۰
- جدول ۳-۳ مشخصه های متغیرها بر اساس جرم اتفاق افتاده ۳۵
- جدول ۱-۴ بدست آمده از گزارشات جرائم سرقت های مسلحانه آگاه ی (رکورد نمونه) ۶۷
- جدول ۲-۴ نمونه ای از داده های آمده شده ۶۹
- جدول ۳-۴ نمونه جرم رخ داده برای تست سربستم ۷۱
- جدول ۴-۴ مقادیر داده ها را قبل از نرمال سازی ۷۲
- جدول ۵-۴ مقادیر داده ها را بعد از نرمال سازی ۷۲
- جدول ۶-۴ نتایج به دست آمده بعد از اجرای الگوریتم فازی روی داد ها ۷۳
- جدول ۷-۴ خوشه های بدست آمده بعد از اجرای الگوریتم K-MEANS ۷۴
- جدول ۸-۴ مراکز خوشه های بدست آمده بعد از اجرای الگوریتم K-MEANS ۷۵
- جدول ۹-۴ دیتای اصلی قبل از اعمال PCA ۷۶
- جدول ۱۰-۴ دیتای اصلی بعد از اعمال PCA ۷۶
- جدول ۱۱-۴ رهن خوشه های بدست آمده از داده های سال ۸۹ ۷۹
- جدول ۱۲-۴ مراکز خوشه بدست آمده از داده های سال ۸۹ ۷۹
- جدول ۱۳-۴ نمونه جرم رخ داده سال ۸۹ ۸۰

فهرست شکلها

شکل ۳-۱	مراحل کشف دانش از داده ها	۴۳
شکل ۳-۲	مفهوم خوشه بندی	۴۵
شکل ۳-۳	نمونه ای از خوشه بندی داد ها	۴۷
شکل ۳-۴	نمونه ای از خوشه بندی فازی داده ها با متد GKCLUST	۴۷
شکل ۳-۵	شکل به دست آمده از الگوریتم K-MEANS	۵۰
شکل ۳-۶	روش خوشه بندی مبتنی بر چگالی	۵۳
شکل ۳-۷	نتیجه حاصل از الگوریتم KNN	۵۳
شکل ۳-۸	نتیجه حاصل از عملکرد الگوریتم KNN	۵۴
شکل ۳-۹	بردار های وئیکلی الگوریتم PCA	۵۶
شکل ۴-۱	نمایی از نرم افزار تهیه شده	۶۵
شکل ۴-۲	روش پیشنهادی برای کشف جرم	۶۶
شکل ۴-۳	معرفی داده های آماده به سیستم	۷۰
شکل ۴-۴	معرفی داد های جرم اتفاق افتاده	۷۱
شکل ۴-۵	عملکرد سیستم را بعد از اجرای الگوریتم K-MEANS	۷۴
شکل ۴-۶	نتیجه حاصل از اجرای مولفه اصلی قبل و بعد از نرمال سازی	۷۶
شکل ۴-۷	نتیجه اجرای الگوریتم K نزدیکی همسایگی روی داده ها	۷۸

فصل اول

معرفی و کلیات

۱-۱. مقدمه

بدون تردید شرایط زندگی اجتماعی بشر، رویارویی با پدیده‌ای به نام جرم را غیر قابل اجتناب می‌سازد، و انسان همواره نیازمند دانش تحلیل جرم خواهد بود. تحلیل جرم عبارت است از به‌کارگیری یک شیوه نظام‌مند جهت شناسایی، کشف و پیش‌بینی جرایم.

ورودی یک سیستم تحلیل جرم، داده‌ها و اطلاعات منتسب به متغیرهای جرم و خروجی آن، پاسخ به پرسش‌های تحلیلی، استخراج دانش و نهایتاً مصور سازی نتایج می‌باشد.

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. حجم انبوه داده‌ها و اطلاعات مرتبط با جرایم در ادارات 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 کارشناسان صحنه جرم با تجربه به‌کار گرفته می‌شوند، (اولاً) نیازمند صرف زمان و هزینه انسانی بسیار زیادی باشند. ثانیاً) به دلیل میزان دخالت بالای عامل انسانی در تصمیم‌گیری‌ها، قادر به لحاظ کردن تمام فاکتورهای تاثیرگذار در یک جرم و ارتباطات بین آنها نباشند. چنین شرایطی، ضرورت به کارگیری یک روش نظام‌مند جهت کشف و حل جرایم را بیش از پیش نمایان می‌سازد. تکنیک‌های داده‌کاوی می‌تواند راهگشای دو چالش فوق باشد.

با توجه به نوظهور بودن بحث به کارگیری تکنیک‌های داده‌کاوی در زمینه تحلیل جرم و همچنین گستره وسیع تحقیقی که در صورت تلفیق دو دانش داده‌کاوی و جرم‌شناسی حاصل می‌گردد، نیاز به انجام مطالعات میدانی در این زمینه را ضروری می‌نماید. از این رو امروزه در محافل علمی، به کارگیری تکنیک‌های داده‌کاوی با محدودیت تحلیل جرم، اهمیت ویژه‌ای پیدا نموده است.

در ادامه به بیان برخی از مفاهیم اصولی تحلیل جرم و همچنین مفاهیم اولیه داده کاوی و برخی تکنیک‌های آن، نظیر خوشه‌بندی پرداخته شده است، پس از آن، نحوه‌ی استفاده از تکنیک‌های داده کاوی همچون خوشه‌بندی جهت استخراج الگوریتم رفتار کاوی مجرمین^۳ و نهایتاً تطابق جرایم با مجرمین^۴، به شیوه‌ای کاربردی مورد بررسی قرار گرفته است.

۱-۲. اهمیت موضوع

در اوایل دهه ۱۹۶۰ اولین ایده‌های ایجاد بانک‌های اطلاعاتی مطرح شد که در آن زمان بیشتر به صورت سیستم‌های مبتنی بر فایل بوده است. تا در اواخر دهه ۱۹۷۰ و اوایل دهه ۱۹۸۰ سیستم‌های مدیریت پایگاه داده^۵ مطرح شدند و به سرعت به محبوبیت رسیدند. در بین انواع سیستم‌های بانک اطلاعاتی که در آن زمان مطرح شدند مانند پایگاه‌های داده سلسله‌مراتبی^۶ و یا شبکه‌ای^۷، پایگاه داده رابطه‌ای^۸ به سرعت جای خود را در سیستم‌های بانک اطلاعاتی باز کرد از جمله دلایل محبوبیت بانک‌های اطلاعاتی رابطه‌ای می‌توان به وجود تئوری‌های ریاضی قوی در آن اشاره کرد. (jiawei han & micheline kamber-(2006.))

استفاده از بانک‌های اطلاعاتی تحولی شگرف در سیستم‌های مدیریت اطلاعات^۹ (MIS) بوجود آورد و به کاربران اجازه داد تا بتوانند با سرعت و دقت بالایی به جواب پرسش‌های خود در سیستم دست یابند. البته این پرسش‌ها به صورت سؤال‌اتی با جزئیات بالا و سطح انتزاع پایین بود زیرا در این سیستم‌ها روی داده‌ها پردازش‌های تحلیلی صورت نمی‌گرفت و داده‌ها به صورت خام در پرس و جوها مورد استفاده قرار می‌گرفتند به این نوع از سؤال و جواب‌ها در سیستم‌های اطلاعاتی، پردازش آنلاین تراکنش (OLTP)^{۱۰} گفته می‌شود که در اکثر سیستم‌های اطلاعاتی امروزه موجود می‌باشند.

Classification^۳
Matching^۴
Data Base Management System^۵
Hierarchical^۶
Network^۷
Relational^۸
Management Information System^۹
On Line Transaction Processing^{۱۰}

به طور کلی می توان گفت که رویکرد سیستم های اطلاعاتی در اوایل دهه ۱۹۷۰ به صورت سیستم های مبتنی بر داده ^{۱۱} بوده است . ولی کم کم با گسترش سیستم های پایگاه داده ای و افزایش حجم داده های ذخیره شده در این سیستم ها ، نیاز به ابزارهایی حس می شد که بتوان بوسیله آنها داده های ذخیره شده را خلاصه تر کرد و با اجرای پردازش های تحلیلی بر روی داده ها اطلاعاتی در اختیار کاربران قرار داد که دارای سطح انتزاع بالاتری باشند تا بتوان از آن اطلاعات برای تصمیم گیری های استراتژیک استفاده کرد زیرا با استفاده از پرسش های ساده در SQL و ابزارهای گوناگون گزارش گیری معمولی ، تنها می توان اطلاعات ساده و اولیه و خامی در اختیار کاربران قرار داد ، اما وقتی که حجم داده ها بسیار بالا باشد ، کاربران هر چند زبر دست و با تجربه باشند نمی توانند الگوهای مفید را در میان حجم انبوه داده ها تشخیص دهند و یا اگر قادر به این کار هم باشند ، هزینه عملیات از نظر نیروی انسانی و مالی مقرون به صرفه نیست . به عنوان مثال، بسیاری از پایگاه داده ها چنان گسترش یافته اند که شامل چندصد میلیون تا چندصد میلیارد رکورد ثبت شده هستند و امکان تحلیل و استخراج اطلاعات با روش های معمول آماری از دل انبوه داده ها مستلزم چند روز کار با رایانه های موجود است. برای بر طرف کردن این مشکل نیاز به پردازش های تحلیلی بر روی داده های خام می باشد از این روی در اواخر دهه ۱۹۸۰ سیستم های انبار داده ^{۱۲} مطرح شدند که ابزاری برای پردازش هایی تحلیل آنلاین (OLAP) ^{۱۳} بر روی داده های چند بعدی به صورت محاوره ای در اختیار می گذاشتند و باعث ساده تر شده مراحل عمومی سازی داده ها ^{۱۴} به صورت موثری می گردیدند. (G.C. Oatley, B.W. Ewart, J.-2005)

تا بدین جا با گذر از سیستم های مبتنی بر فایل به سیستم های دارای بانک اطلاعاتی برخورد شد بعد از آن با توجه به نیاز سیستم های پشتیبان تصمیم (تصمیم ساز مدیریتی) به اطلاعات و نتایج کلی تر و با سطح انتزاع بالاتر ، سیستم های انباره داده پدیدار شدند ولی این سیستم ها به صورت

^{۱۱} Data Driven System
^{۱۲} Data Warehouse
^{۱۳} On Line Analytical Processing
^{۱۴} Generalization

محاوره‌ای می‌باشند و کاربران معمولاً فرضیه‌ای را مطرح می‌کنند و سپس بر اساس گزارشات مشاهده شده به اثبات یا رد فرضیه می‌پردازند، در حالی که امروزه نیاز به روشهایی است که اصطلاحاً به کشف دانش بپردازند یعنی با کمترین دخالت کاربر و به صورت خودکار الگوها و رابطه‌های منطقی را بیان نمایند.

داده کاوی یکی از مهمترین این روشها است که به وسیله آن الگوهای مفید در داده ها با حداقل دخالت کاربران شناخته می‌شوند و اطلاعاتی را در اختیار کاربران و تحلیل گران قرار می دهند تا براساس آنها تصمیمات مهم و حیاتی در سازمانها اتخاذ شوند. هم اکنون در هر کشور، برای سازمانها و شرکت های مختلف در امور گوناگونی مانند بازرگانی، پرسنلی، آموزشی، پایگاه داده ها ایجاد یا خریداری شده است، به طوری که این پایگاه داده ها برای مدیران، برنامه ریزان و پژوهشگران جهت تصمیم گیری های راهبردی، تهیه گزارش های مختلف، توصیف وضعیت جاری خود و ... می - تواند مفید باشد.

داده کاوی عبارت است از فرآیند استخراج اطلاعات معتبر، از پیش ناشناخته، قابل فهم و قابل اعتماد از پایگاه داده های بزرگ و استفاده از آن در تصمیم گیری در فعالیتهای تجاری مهم. اصطلاح داده کاوی به فرآیند نیم خودکار تجزیه و تحلیل پایگاه داده های بزرگ به منظور یافتن الگوهای مفید اطلاق می شود.

داده کاوی یعنی جستجو در یک پایگاه داده ها برای یافتن الگوهایی میان داده ها. داده کاوی یعنی استخراج دانش کلان، قابل استناد و جدید از پایگاه داده های بزرگ. داده کاوی یعنی تجزیه و تحلیل مجموعه داده های قابل مشاهده برای یافتن روابط مطمئن بین داده ها.

همانگونه که در تعاریف گوناگون داده کاوی مشاهده می شود، تقریباً در تمامی تعاریف به مفاهیمی چون استخراج دانش، تحلیل و یافتن الگوی بین داده ها اشاره شده است.

۱-۳. انگیزه پژوهشگر

از آنجایی که امروزه اهمیت دانش و تکنولوژی داده کاوی به حدی است که رد پای آن به طور مستقیم و غیرمستقیم در اکثر علوم مهندسی به چشم می خورد و با توجه به جد بودن استفاده از این دانش در زمینه کشف علمی و تحلیل جرایم، کاربردی به نظر می رسد. داده کاوی را می توان استخراج دانش های نامحسوس و پنهان موجود در انبوه داده ها تعریف نمود. با توجه به این تعریف و همچنین چالش های مطرح شده، ضرورت استفاده از تکنیک های داده کاوی در تحلیل علمی جرایم نمایان می گردد.

در اکثر پژوهش های علمی جرم شناسی و تحلیل جرم در دهه اخیر، رد پای استفاده از تکنیک های داده کاوی به وضوح قابل مشاهده است. اولین گام در تلفیق دو دانش داده کاوی و تحلیل جرم، شناسایی برخی از نیازمندی های کلیدی موجود در حوزه جرم شناسی است که توسط قابلیت های داده کاوی قابل برطرف شدن هستند. انگیزه محقق در این جا استفاده از روش های داده کاوی در کشف جرم می باشد که بتوان با پیشنهاداتی، وضعیت کشف جرم را بهبود بخشید. کشف جرم منجر به کسب اطلاعات بیشتر در مورد مجرمین و بالابردن سرعت کشف می شود.

۱-۴- سوالات تحقیق

با توجه به مطالب ذکر شده سوال تحقیق به شرح زیر می باشد:

آیا می توان سیستم گزارشات جرائم پلیس را با استفاده از داده کاوی ارتقاء داد ؟

چه پیشنهاداتی برای ارتقاء سیستم گزارشات جرائم می توان ارائه داد ؟

آیا روش های داده کاوی می تواند در کشف جرم کمک نماید ؟

۱-۵. اهداف تحقیق

در تحقیقات قبلی با استفاده از برخی تکنیک های داده کاوی - مانند شبکه های عصبی و

درخت های تصمیم و همچنین الگوریتم های دیگر به نتایجی رسیده اند ؛ در این تحقیق در ابتدا