

## Providing a Framework for Assessing and Evaluating Network Data Studies in the Fight Against Social Anomalies


Mohammad Khalili<sup>1</sup>


Hamzehali

Nourmohammadi<sup>2\*</sup>

Nader Naghshineh<sup>3</sup>

 1. Ph.D in Knowledge and Information Science, Department of Knowledge and Information Science, University of Tehran, Tehran, Iran.  
Email: Mohammadxhalili45@gmail.com

 2. Associate professor, Department of Knowledge and Information Science, Shahed University, Tehran, Iran. (Corresponding Author)

 3. Assistant professor, Department of Knowledge and Information Science, University of Tehran, Tehran, Iran.  
Email: nnaghsh@ut.ac.ir

Email: [nourmohammadi@Shahed.ac.ir](mailto:nourmohammadi@Shahed.ac.ir)

### Abstract

**Purpose:** In today's world, every society is struggling with the challenges of crime and its control. Committing a crime represents one of the most significant social harms, necessitating the involvement of police and judicial institutions for effective management. Assessing, evaluating, and identifying crime patterns, as well as detecting and preventing criminal activities, have been the focal points of judicial and law enforcement agencies since ancient times. Consequently, crime prevention is consistently prioritized over crime detection after it occurs. Various methods have been introduced for crime detection, including a wide range of innovative computer techniques. Data mining is regarded as one of the most effective tools for data and information analysis in the field of crime detection. Many effective parameters are available for analyzing crime through data mining methods. Data mining serves as a valuable tool for examining crime data warehouses, helping to extract hidden knowledge within them. The application of data mining techniques, along with various machine learning methods, can yield significant benefits in identifying, predicting, and preventing crime in any society. Diagnosing, predicting, and preventing crime through data mining represents a progressive approach supported by statistical methods, psychology, artificial intelligence, criminology, machine learning, and database technologies. Consequently, the primary objective of this research is to establish a framework for analyzing digital network data in the battle against social anomalies, particularly the crime of theft. In this research, materials related to the crime of theft have been categorized into three areas: crime identification, prediction, and prevention. Additionally, the application of data mining methods has been explored within these domains.

**Methodology:** This research is applied in terms of its purpose and was conducted using documentary methods, content analysis, and data mining. The statistical population for this study consists of information related to theft crimes recorded by law enforcement and police organizations in 2019. Initially, the relevant data were collected in the form of documents and subsequently selected using the content analysis method to

Date of Reception:  
24/10/2023

Date of Acceptation:  
07/04/2024



Mohammad Khalili <sup>1</sup>

Hamzehali  
Nourmohammadi <sup>2\*</sup>

Nader Naghshineh <sup>3</sup>

Date of Reception:  
24/10/2023

Date of Acceptation:  
07/04/2024



facilitate further analysis. Patterns of theft crimes were identified based on various factors, including the type of crime (such as home and car theft), entry location, entry method, search techniques, and residential area. This study utilized digital network data analysis tools and methods, specifically data mining, for classification and validation purposes. Additionally, clustering techniques, such as k-means, were employed to identify different forms of theft crimes. Classification algorithms, including neural networks, Bayesian rules, Bayesian navigation, and support vector machines, were used to predict theft crimes. The primary data analysis tool utilized in this research was Excel software.

**Findings:** The findings indicate that the season in which a theft occurs positively correlates with the month of the crime. Additionally, the method used to enter the crime scene shows a positive correlation with the method used to exit the scene. Furthermore, the accuracy of the Bayesian model in predicting and detecting the type of crime is 0.412. The model demonstrates the highest prediction accuracy for home thefts at 73%, while the lowest prediction accuracy for thefts from private locations is 27%. The results are expressed as a percentage. Additionally, the accuracy of each technique employed has been compared. The ROC findings indicate that the accuracy of the Bayesian methods and the Multilayer Perceptron (MLP) neural network, as well as the support vector machine, in predicting theft from public places is higher than for other types of theft. Conversely, in predicting home theft, these methods demonstrate lower accuracy compared to other thefts. Furthermore, the prediction accuracy of the support vector machine method, at approximately 91%, surpasses that of Bayesian methods (around 73%) and neural networks (about 90%) in predicting theft from public places. Moreover, the ROC chart for the support vector machine indicates that it is 7% less accurate in predicting pickpocketing compared to other thefts.

**Conclusion:** The results indicated that the Bayesian rules method is the most effective for detecting and predicting patterns of theft crimes. It demonstrates higher accuracy compared to other methods. Specifically, for predicting thefts from public places, the support vector method is recommended. As a result of the findings from data mining techniques, it is recommended that police and judicial organizations utilize a combination of data mining and artificial intelligence methods to enhance the detection and identification of thieves, particularly those with varying criminal histories related to theft. This approach aims to improve the accuracy of the information obtained about these individuals. Furthermore, based on the data mining techniques discussed in this research, an expert intelligent system can be developed to predict the likelihood of future criminal attempts by offenders.

**Keywords:** Social anomalies, Crime of theft, Data mining, Struggle, Network analysis.

# ارائه چارچوب برای سنجش و ارزیابی مطالعات داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی

محمد خلیلی<sup>۱</sup>حمزه علی نور محمدی<sup>\*۲</sup>نادر نقشینه<sup>۳</sup>

۱. دانش‌آموخته دکتری تخصصی گروه علم اطلاعات و دانش‌شناسی دانشگاه تهران، ایران.

Email: Mohammadkhalili45@gmail.com

۲. دانشیار گروه علم اطلاعات و دانش‌شناسی دانشگاه شاهد، تهران، ایران. (نویسنده مسئول)

۳. استادیار، دانشکده مدیریت، گروه علوم اطلاعات و دانش‌شناسی دانشگاه تهران، ایران.

Email: nnaghsh@ut.ac.ir

Email: nourmohammadi@Shahed.ac.ir

## چکیده

**هدف:** هدف پژوهش پیش‌رو ارائه چارچوبی برای سنجش و ارزیابی مطالعات داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی است.

**روش‌شناسی:** این پژوهش از نظر هدف کاربردی است و با استفاده از روش اسنادی، تحلیل محتوا و داده‌کاوی انجام شده است. جامعه آماری این پژوهش شامل اطلاعات مرتبط به جرم سرقت ثبت‌شده در سازمان‌های انتظامی و پلیسی در سال ۱۳۹۹ است. برای تحلیل اطلاعات از روش‌های تحلیل داده‌های شبکه‌های دیجیتال به خوشه‌بندی (کای میانگین)، شناسایی انواع جرم سرقت و دسته‌بندی (شبکه‌ها شبکه‌های عصبی، قوانین بیزی، ناو بیزی و ماشین بردار پشتیبان) به‌منظور پیش‌بینی جرم سرقت استفاده شده است.

**یافته‌ها:** یافته‌های پژوهش نشان داد دقت مدل بیزی در پیش‌بینی و تشخیص نوع جرم برابر با ۰.۴۱۲ درصد است و بیشترین درصد پیش‌بینی برای جرم سرقت از خانه برابر با ۷۳ درصد و کمترین درصد پیش‌بینی سرقت از اماکن خصوصی برابر با ۲۷ درصد است. همچنین میزان دقت هریک از روش‌های انجام‌شده باهم مقایسه شده‌اند. یافته‌های Roc نشان داد دقت روش قوانین بیزین، شبکه‌های عصبی (MPL) و بردار ماشین پشتیبان در پیش‌بینی جرم سرقت از اماکن عمومی نسبت به سایر سرقت‌ها بیشتر و در موارد مشابه در پیش‌بینی سرقت از منزل دارای دقت کمتری نسبت به سایر سرقت‌ها است. همچنین دقت پیش‌بینی روش بردار ماشین پشتیبان (حدود ۹۱ درصد) در پیش‌بینی سرقت از اماکن عمومی بهتر از قوانین بیزین (حدود ۷۳ درصد) و شبکه‌ها - شبکه‌های عصبی (حدود ۹۰ درصد) است. علاوه بر این نمودار Roc در روش بردار ماشین پشتیبان نشان داد در پیش‌بینی جرم جیب‌بری نسبت به سایر سرقت‌ها دقت کمتری (۷ درصد) دارد.

**نتیجه‌گیری:** نتایج این پژوهش نشان داد که ارزیابی و سنجش با استفاده از روش قوانین بیزین مناسب‌ترین روش در تشخیص و پیش‌بینی یک الگوی مناسب از جرم سرقت (نوع جرم) است و نسبت به سایر روش‌ها دقت بیشتری دارد؛ اما به‌طور خاص، برای پیش‌بینی سرقت از اماکن عمومی، استفاده از ارزیابی و سنجش به‌وسیله روش بردار پشتیبان توصیه می‌شود.

**واژگان کلیدی:** ناهنجاری‌های اجتماعی، جرم سرقت، داده‌کاوی، مبارزه، تحلیل شبکه.

صفحه ۱۸۰-۱۶۳

دریافت: ۱۴۰۲/۰۸/۰۲

پذیرش: ۱۴۰۳/۰۱/۱۹



## مقدمه و بیان مسئله

امروزه هر جامعه‌ای با مشکلات جرم و چگونگی کنترل آن دست‌وپنجه نرم می‌کند. ارتکاب جرم یکی از مهم‌ترین آسیب‌های اجتماعی است و کنترل آن نیازمند تلاش و فعالیت گسترده نهادهای پلیسی و قضایی است. کارآمدی دولت‌ها در برخورد با جرائم و تأمین امنیت و آسایش جامعه مستلزم اختصاص دادن بودجه زیادی برای نیروهای انتظامی و قضایی است؛ بنابراین، هدف این پژوهش ارائه چارچوب تحلیل ابر داده‌های شبکه‌های دیجیتال برای کمک به پلیس و نیروهای انتظامی است. کارایی دولت‌ها تنها با توسل به روش‌های کیفری پس از وقوع جرم احراز نمی‌شود، بلکه آن‌ها وظیفه‌دارند با در پیش گرفتن روش‌هایی قبل از وقوع خطا، در از بین بردن شرایط تحقق آن تلاش کنند. در دنیای اطلاعاتی امروز روش‌های گوناگونی برای تسریع کشف و کم کردن میزان وقوع جرائم موردبررسی و آزمایش قرار گرفته‌اند که از این میان، روش‌های علمی و دارای هوشمندی کشف جرم، به دلیل برخورداری از پشتوانه علمی و دانش ریاضی در حوزه جرم‌شناسی نوین کاربرد فراوانی دارد (کیوان‌پور و همکاران، ۱۳۸۸). وضعیت حال مسئله بی‌شک، شرایط اجتماعی بشر، مواجه‌شدن با نمادی به اسم جرم را غیرقابل دوری می‌سازد و انسان پیوسته نیازمند دانش، روا داشتن، جرم است. گزارش جرم عبارت است از به‌کارگیری طریقه‌ای ترتیب‌مند به منظور شناسایی، پیش‌بینی و کشف جرائم. ورودی یک سامانه تحلیل جرم، داده‌ها و اطلاعات منسوب به متغیرهای جرم است و خروجی آن بازتاب به استفسارهای تحلیلی، استخراج دانش و در فرجام مصورسازی برآیندها است (کاظمی و حسین پور، ۱۳۸۸).

جرم‌ها، بی‌قاعدگی‌های اجتماعی هستند که جامعه‌ها مخارج زیادی را به شیوه‌های گوناگون بابت آن پرداخت می‌کنند. مهم‌ترین آماج‌های دولت‌ها به کار بست یک تدبیر جنایی بهینه، حذف کردن مجال‌های اقدام جرم، جلوگیری از رویداد جرم و تلاقی با مجرمان است. نهادهای انتظامی از روش‌های مختلفی مانند (ازدیاد نیروهای انتظامی، دوربین‌های مدار بسته و ...) برای مقابله با جرم استفاده می‌کنند. با بسط و توسعه فناوری اطلاعات در تشکیلات و ایجاد بانک‌های اطلاعاتی، داده‌کاوی نیز به دیباچه یک ادوات نرم‌افزاری قدرتمند و به درجات کم‌هزینه‌تر و کارآمدتر در گزینش سازمان‌های انتظامی قرار گرفته است. با کمک داده‌کاوی می‌توان طرح‌های جرم را شناخت تا بتوان جرائم را پیش‌بینی کرد و از وقوع آن جلوگیری نمود (کاظمی و همکاران، ۱۳۸۸). از این رو پژوهش به دنبال پاسخ به این پرسش اساسی است که چارچوب تحلیل ابر داده‌های شبکه‌های دیجیتال در مبارزه با ناهنجاری‌های اجتماعی چگونه است؟

## پرسش‌های پژوهش

۱. ارزیابی آماده‌سازی داده برای تحلیل محتوا چگونه انجام می‌شود؟
۲. آیا بین جرائم و متغیرهای جرم‌شناسی رابطه وجود دارد؟
۳. تحلیل خوشه‌های حاصل از الگوریتم کای میانگین چگونه است؟
۴. تحلیل داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی (جرم سرقت) از نظر الگوریتم ناویز چگونه است؟
۵. تحلیل داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی (جرم سرقت) از نظر الگوریتم قوانین بیزین چگونه است؟
۶. تحلیل داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی (جرم سرقت) از نظر الگوریتم شبکه عصبی (MPL) چگونه است؟

۷. تحلیل داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی (جرم سرقت) از نظر بردار ماشین پشتیبان چگونه است؟

## چارچوب نظری

در این پژوهش کاربرد روش‌های داده‌کاوی (تحلیل داده‌های شبکه دیجیتال) در جرم‌شناسی مورد بررسی قرار گرفته است. همچنین مطالب مرتبط با جرم‌شناسی در اشکال شناسایی (شناخت)، پیش‌بینی (عاقبت‌اندیشی) و پیشگیری (جلوگیری) جرم طبقه‌بندی‌شده و استفاده از شیوه‌های تحلیل داده‌های شبکه دیجیتال در این سه حوزه بررسی خواهد شد. بر اساس پژوهش‌های انجام‌شده برای پیش‌بینی مبادرت جرم، دست‌اندرکاران و تعداد اقلام تأثیرگذار در آن، از روش‌های پیش‌بینی بیشتر از سایر ابزارهای داده‌کاوی استفاده شده است. از بین الگوریتم‌های پیش‌بینی، مدل‌های رگرسیون حجم زیادتری از این پژوهش‌ها را به خود اختصاص داده‌اند. در مواردی نیز ترکیبی از روش‌های خوشه‌بندی و پیش‌بینی دیده می‌شود. روش‌های دیگر داده‌کاوی نیز می‌توانند در این قسمت کارساز شوند (آخوندزاده و احمدوند، ۱۳۸۹). طراحی سیستم‌های پشتیبان تصمیم مبتنی بر دانش کشف‌شده توسط روش‌های داده‌کاوی می‌تواند در این زمینه مثمر ثمر واقع شود (Karlis & Meligkotsidou, 2007; Murtagh et al., 2009; Murtagh et al., 2009). کارلیس و همکاران (D'Alessio & Stolzenberg, 2010; Liu & Brown, 2003) و مورتاق و همکاران (Murtagh et al., 2009) با استفاده از روش‌های خوشه‌بندی به شناسایی جرائم پرداخته‌اند. دالیو و همکاران (D'Alessio & Stolzenberg, 2010) با استفاده از روش‌های پیش‌بینی رگرسیون به پیش‌بینی جرائم؛ و لیو و همکاران (Liu & Brown, 2003) با استفاده از روش فازی؛ و آتلی و همکاران (Oatley & Ewart, 2003) با استفاده از ترکیب روش‌های رگرسیونی و شبکه‌ها - شبکه‌های عصبی و بیزین به پیشگیری جرائم پرداخته‌اند.

از زمان‌های گذشته شناسایی الگوهای جرم، کشف جرائم و پیشگیری از آن مورد توجه نهادهای قضایی و پلیسی بوده به‌گونه‌ای که همواره پیشگیری از جرم بسیار پراهمیت‌تر از شناسایی جرم پس از وقوع آن است. تاکنون روش‌های مختلفی برای شناسایی جرم معرفی شده که روش‌های نوین رایانه‌ای طیف گسترده‌ای از آن را شامل می‌شود. داده‌کاوی به‌عنوان یکی از کارآمدترین ادوات تحلیل داده و اطلاعات، مورد توجه بسیاری از پژوهشگران قرار گرفته است. پارامترهای متعدد مؤثر در جستار تحلیل جرم و جنایت، استفاده از روش‌های طرح‌شده در دانش‌های گوناگون را طلب می‌کند. روش‌های داده‌کاوی به‌عنوان ادوات و آزمون‌های انبار داده به استحصال دانش پنهان در درون آن‌ها می‌پردازد. استفاده از شیوه‌های داده‌کاوی و پیرو آن روش‌های یادگیری رایانه در شناسایی، پیش‌بینی و پیشگیری جرم و جنایت می‌تواند نتیجه نافع به همراه داشته باشد. تشخیص، پیش‌بینی و پیشگیری از رخداد جرائم با داده‌کاوی یک آرمان جدید است که به‌وسیله روش‌های آماری، روان‌شناسی، هوش مصنوعی، جرم‌شناسی، یادگیری رایانه و فناوری‌های پایگاه داده تحقیق‌یافته است.

تجزیه و تحلیل داده‌های حجیم و پیش‌بینی نتایج آن‌ها بر اساس یافته‌ها و ارزیابی برنامه‌های کاربردی در حوزه‌های مبارزه با جرم و جنایت، کسب‌وکار، امور مالی، بهداشت و درمان و غیره صورت می‌گیرد. ابر داده نیز به‌طور مشابه راهش را در زمینه مبارزه با جرم و جنایت پیدا کرده است. ابر داده به‌طور فزاینده‌ای از حفظ نظم و آرامش هر روزه و تلاش‌های امنیتی سایبری خبر می‌دهد. هر نوع اجرای قانون، از روش‌های ابر داده برای اشاره دقیق به نقاط بالقوه جرم

ارائه چارچوب برای سنجش و ارزیابی مطالعات داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی

و جنایت یا پیش‌بینی خانه‌هایی که احتمال حمله مسلحانه به آن‌ها می‌رود، استفاده می‌کند. حملات سایبری از همه نوع در حال افزایش است. در حوزه جرم و جنایت، داده‌های انبوه علاوه بر کاهش جنایات خشونت‌آمیز و دزدی، به تشخیص مجرمان یقه‌سفید که در جرائم مالی از جمله تقلب‌های بیمه، تجارت‌های خودی، پول‌شویی و تقلب در مراقبت‌های بهداشتی و درمانی نیز درگیر هستند، کمک می‌کنند. حملات سایبری از همه نوع در حال گسترش است و یکی از راه‌های مقابله با این هجوم استفاده از ابر داده برای آگاهی از حملات سایبری و یافتن نقاط آسیب‌پذیر برای دفاع در برابر آن است.

ارائه چارچوب تحلیل ابر داده‌ها در مبارزه با آسیب‌های اجتماعی در حقیقت نیازمند تحلیل حجم زیادی از داده‌ها است که مقیاس آن‌ها بیشتر از معیاری است که با نرم‌افزارهای متداول یا شیوه‌های سنتی بتوان آن‌ها را در شبکه‌های دیجیتال در یک زمان مطلوب اخذ، نکته بین‌سازی، گردانش و پردازش کرد و تحلیل و دسته‌بندی آن‌ها به‌وسیله یک دسته از عوامل ساختاری و انباشت داده‌های شایع همواره مورد چالش است. نهادهای امنیتی با تراکم اطلاعات گوناگون در شبکه‌های دیجیتال همچون خبرگزاری‌ها، ایمیل‌ها، عکس‌ها، ویدیوها، گزارش‌ها و اطلاعات روزانه، نوشته‌ها و اطلاعات شبکه‌های اجتماعی، داده‌های تلفن‌های همراه از جمله موقعیت‌یاب‌ها و پیامک‌ها، پردازش زبان طبیعی و نرم‌افزارهایی که ذخیره داده انجام می‌دهند، تحقیقات تراکنش‌های مالی، جست‌وجوی‌های رایانه‌ای، تحلیل‌های تقلب‌های اقتصادی، مالی و تجاری، جمع‌آوری داده‌ها و اطلاعات مربوط به مناطق جرم و جنایت خیز محدوده شهری و نقاط حساس کشور، داده‌های تحلیل امنیت سرورهای رایانه‌ای مهم در مبارزه با حملات سایبری و محتوای وب‌گاه‌ها روبرو هستند که باید به‌طور هوشمندانه‌ای رده‌بندی شود تا به‌سرعت در مبارزه با آسیب‌های اجتماعی قابل بررسی و دستیابی باشند. مفهوم «اندازه» در داده‌های حجیم به شکل پیوسته در حال دگرسانی است و باگذشت زمان کلان‌تر می‌شود. داده‌های حجیم گروهی<sup>۱</sup> از روش‌هایی است که نیازمند قالب نوینی از یکپارچگی‌اند تا بتوانند ارزش‌های سودمندی را که در مجموعه‌های پیچیده، وسیع، بزرگ و متنوع داده، نهان شده‌اند، پدیدار سازند. با توجه به این رویکرد و رشد روزافزون داده‌ها و ضرورت بهره‌برداری و گدازش آن‌ها، به‌کارگیری مجموعه قواعد شکل‌گیری داده‌های بزرگ با استفاده از نرم‌افزارهای داده‌کاوی از جمله نایم در شبکه‌ها - شبکه‌های دیجیتال از اهمیت ویژه‌ای برخوردار است. از این‌رو تحقیق حاضر به دنبال پاسخ به این پرسش بنیادی است که چارچوب تحلیل ابر داده‌های شبکه‌ها - شبکه‌های دیجیتال در مبارزه با ناهنجاری‌های اجتماعی چگونه است؟ به‌بیان‌دیگر با استفاده از روش‌های داده‌کاوی - روش‌های پیش‌بینی (شبکه‌ها - شبکه‌های عصبی، ماشین بردار پشتیبان، قوانین بیزین و ناو بیزی) به پیش‌بینی الگوی جرم سرقت پرداخته خواهد شد.

## پیشینه پژوهش

اسکندری و همکاران (۱۳۹۰) مطالعه‌ای با عنوان کاربرد داده‌کاوی در شناسایی و کشف الگوهای پنهان جرم سرقت انجام دادند. هدف اصلی این تحقیق توسعه و ارائه یک مدل داده‌کاوی با بهره‌گیری از بانک‌های اطلاعاتی جرائم موجود و استفاده از ابزارها و الگوریتم‌های داده‌کاوی بوده است. در پژوهش دیگری حاجی‌زاده و عسکری (۱۳۹۶) به بازشناسی موارد شبهه‌زا به پول‌شویی بر اساس قانون مبارزه با پول‌شویی مصوب سال ۱۳۸۶ و دستورالعمل‌های ده‌گانه بانک مرکزی برای مبارزه با پول‌شویی پرداخت و نشان داد که داده‌های مورداستفاده به‌صورت

اتفاقی پدیدآوری شده‌اند و تعدادی داده شبهه زا، به درون داده‌ها وارد شده است، درصد کشف درست داده‌های شبهه زا در این مقاله ۷۴ درصد اعلام شده است. آگاروال و همکاران (Agarwal et al. 2014) در مقاله خود روشی برای خوشه‌بندی جرائم بر اساس الگوریتم کامینز ارائه کرده‌اند. ابزار مورد استفاده در تحقیق نرم‌افزار راپید ماینز بوده است. آن‌ها در این مقاله فقط اقدام به خوشه‌بندی اقلام داده‌ای جرائم بر اساس مجموعه داده‌های متفاوت در سال‌های مختلف و صرفاً برای جرم قتل انجام دادند. خلیلی و نوایی فیجانی (۱۳۹۹) کاربرد ابر داده در مبارزه با جرم و جنایت را با ارائه راه‌حل‌های استنادی بررسی کرده‌اند و به نتایج قابل قبولی هم دست یافتند.

در خارج از ایران، دالسیو (D'Alessio & Stolzenberg, 2010) به‌طور کلی کاربرد شیوه‌های داده‌کاوی در شناخت جرائم را در دودسته اقدامات در نظر گرفت. اولین دسته شامل اقداماتی می‌شوند که قبل از وقوع جرائم به‌منظور پیش‌بینی و پیشگیری از ارتکاب جرم انجام می‌گیرند و دسته دوم پیرامون اقدامات انجام‌شده بعد از وقوع جرم به‌منظور بررسی و کشف مدارک و شواهد جرم پس از وقوع آن است. لیو و براون (Liu & Brown, 2003) در پژوهشی به شناسایی و پیش‌بینی جرائم و پیشگیری از آن با جداسازی کاربرد روش‌های داده‌کاوی پرداختند؛ اما هیچ تفاهمی در مورد چگونگی کنترل مردم به‌عنوان عاملان جرائم و منبع اطلاعاتی نیافتند. در پژوهش دیگری اتلی و اوارت (Oatley & Ewart, 2003) با گزینش روش بهینه بر اساس نوع جرم و جنایت نشان داد که هیچ اتفاق نظری در مورد استفاده از یک شیوه ویژه مدنظر نیست و در نهایت تشریح برآیندها به دیدگاه آمارشناسان، جامعه‌شناسان و محققان جرم و جنایت بستگی دارد.

با بررسی تحقیقات انجام‌شده در این زمینه می‌توان نتیجه گرفت در داخل کشور تحقیقات چندانی در این موضوع انجام‌نشده و کمبود پژوهش‌ها به‌وضوح دیده می‌شود و پژوهشگران کشورمان در این موضوعات آن‌چنان‌که باید به تحقیق و نوآوری نپرداخته‌اند. در مقابل در خارج از ایران پژوهش‌های متعدد و کاملی صورت گرفته ولی هیچ تفاهمی بین آن‌ها در مورد انتخاب یک موضوع و روش وجود ندارد. این پژوهش بنا به‌ضرورت و اهمیت تحقیق در زمینه جرم سرقت در داخل کشور انجام شده و هدف آن بهبود وضعیت مطالعه (با توجه به پیشینه محدود) و ارائه یک چارچوب مشخص است. در این زمینه همچنین نگرش به موضوعات دیگر مبارزه با ناهنجاری‌ها می‌تواند به پیشرفت تحقیقات در داخل کشور کمک کند.

جدول ۱. چارچوب کاربرد روش‌های داده‌کاوی در مدل‌سازی جرم و جنایت

حوزه‌های کاربرد	شیوه‌های مورد استفاده	مرجع
شناسایی جرائم	خوشه‌بندی	کارلیس و ملیگکوتیستو (Karlis & Meligkotsidou, 2007)
	خوشه‌بندی	مورتاق و همکاران (Murtagh et al., 2009)
	خوشه‌بندی باینری	مون و همکاران (Moon et al., 2010)
پیش‌بینی جرائم	پیش‌بینی - رگرسیون	کوراپاچوقلو و اردوغان (Çorapçioğlu & Erdoğan 2004)
	پیش‌بینی - رگرسیون	مون و همکاران (Moon et al., 2010)
	پیش‌بینی	لیو و براون (Liu & Brown, 2003)
	پیش‌بینی - رگرسیون	دالسیو و استولزنبرگ (D'Alessio, & Stolzenberg, 2010)
	پیش‌بینی مبتنی بر نقاط جرم خیز	لیو و براون (Liu & Brown, 2003)
	پیش‌بینی - سری‌های زمانی	دیدمن (Deadman, 2003)

ادامه جدول ۱. چارچوب کاربرد روش‌های داده‌کاوی در مدل‌سازی جرم و جنایت

حوزه‌های کاربرد	شیوه‌های مورد استفاده	مرجع
پیش‌بینی جرائم	پیش‌بینی - رگرسیون	دیدمن (Deadman, 2003)
	خوشه‌بندی - پیش‌بینی	ایکس یوای و همکاران (Xue & Brown, 2006)
	خوشه‌بندی - قوانین انجمنی	مالاتی و سنتوش (Malathi, & Santhosh, 2011)
پیش‌گیری جرائم	قوانین انجمنی فازی	دالیسو و همکارانش (D'Alessio, & Stolzenberg, 2010)
	فازی سام	لی و همکاران (Li et al., 2010)
پیش‌گیری جرائم	ترکیبی از روش‌های رگرسیون، شبکه‌ها - شبکه‌های عصبی، شبکه بیزین	آتلی و اوارت (Oatley & Ewart, 2003)
	رگرسیون لجستیک	دالیسو و استولزبرگ (D'Alessio, & Stolzenberg, 2010)

لی خاک و همکاران (Li et al., 2010) یک الگوریتم شناسایی موارد مشکوک به پول‌شویی، بنانهاده شده، بر داده‌کاوی و شبکه‌های عصبی ارائه و نشان دادند الگوریتم ارائه‌شده در مبارزه با پول‌شویی بانک سرمایه‌گذاری بین‌المللی ایرلند آزمون شده و نتایج به‌دست‌آمده هم قابل قبول بوده است. الگوریتم ارائه‌شده برای خوشه‌بندی تراکنش‌ها از روش کای - میانگین استفاده کرده و خروجی به‌دست‌آمده به‌سوی آموزش یک شبکه عصبی از نوع پس انتشار به‌عنوان آموزش‌دهنده موارد معمول و دارای شک به‌کاررفته است. فووا و همکاران (Phua et al., 2010) در یک مقاله مروری روش‌های داده‌کاوی به‌کاررفته برای کشف انواع کلاهبرداری مالی را ارائه و نشان دادند دو کمبود کلی در قلمرو بازشناسی کشف کلاهبرداری بر پایه روش‌های داده‌کاوی وجود دارد. نخستین کمبود، نداشتن داده واقعی عمومی برای انجام سنجش است و دومین عدم انتشار شیوه‌های موفق کاربردی در این حوزه است. علاوه بر این لوپزروچاس و آکسلسون (Lopez-Rojas & Axelsson, 2012) در راهکار تشخیص پول‌شویی خود از الگوریتم‌های یادگیری ماشین شامل «درخت تصمیم» و «قواعد تصمیم» استفاده کرده‌اند و دلیل آن را فهم راحت‌تر خروجی این دسته الگوریتم‌ها برای انسان بیان کرده‌اند. در آزمایش ایده این الگوریتم از داده‌های مصنوعی استفاده شده است. بهره‌گیری از داده‌های غیرواقعی همواره با این خطر روبروست که این داده‌ها بیان صحیحی از دنیای واقعی ارائه ندهند و یا ممکن است برآیند حاصله رو به بیراهه باشد اما از جهتی می‌توان به‌وسیله آن‌ها مقدماتی آماده کرد که شاید داده‌های واقعی آن‌ها هیچ‌گاه فراهم نکنند.

### روش‌شناسی پژوهش

این تحقیق از نوع هدف کاربردی و از نظر روش اسنادی، تحلیل محتوا و داده‌کاوی است و روش کار هم به این صورت است که به‌وسیله داده‌ها به مدل موردبحث برسیم. از آنجاکه CRISP-DM یک استاندارد جهانی است و در پروژه‌های داده‌کاوی مورد استفاده قرار می‌گیرد؛ بنابراین قالب اجرایی تحقیق بر همین اساس مورد واکاوی قرار گرفته است. در این تحقیق مجموعه داده‌ها از داده‌های واقعی سازمان پلیس تهیه شده است. جامعه آماری این پژوهش متشکل از داده‌های مربوط به ۵۰۰۰ نفر سارق است که از ۱۳۹۸-۱۳۹۹ مرتکب جرم شده و اطلاعات مربوط به جرم آنان در سیستم ثبت شده است. این اطلاعات مربوط به افرادی است که هم از خانه (ویلا، آپارتمانی و غیره)، هم از خودرو، هم از اماکن خصوصی و عمومی (بانک‌ها، سازمان‌ها و غیره)، هم به صورت شخصی (جیب‌بری، کیف‌قاپی و



غیره) دزدی به صورت مسلحانه و غیرمسلحانه انجام داده‌اند.

از این رو اطلاعات پنج نمونه سارق جمع آوری شده است. گفتنی است اگر فردی طی این مدت دو بار سرقت کرده باشد، فقط یک بار آن توسط سیستم ثبت و وارد مجموعه داده‌های تحقیق شده است. به عبارت دیگر، جامعه آماری (و به تبع آن نمونه آماری) هر سارق دارای یک رکورد در پایگاه داده‌ها است. برای تحلیل محتوا از نرم‌افزار نایم<sup>۱</sup> استفاده شده است.

## یافته‌های پژوهش

### پاسخ به پرسش اول پژوهش. آماده‌سازی داده برای تحلیل محتوا چگونه انجام می‌شود؟

در مجموعه عملکردهای مهیاسازی برای پالایش و پیش‌پردازش داده‌ها، دو عملیات نرم‌سازی (اعمال تغییرات در شکل داده‌ها)، کاهش داده‌ها بر روی پایگاه داده رابطه‌ای انجام شده است. برای انجام الگوریتم درون پایگاه داده بعد از مرحله پیش‌پردازش برخی از رکوردها حذف، برخی اضافه و برخی نیز تغییر یافته‌اند. همچنین برخی متغیرها (خصیصه‌های سارقان)، متغیرهای تأثیرگذاری در رفتار سارقان نبودند. این قسمت‌ها در پایگاه داده‌ها محتوی اطلاعات هویتی شامل (نام، نام خانوادگی، نام پدر، شماره شناسنامه، شماره گواهینامه)، یگان اعتراف‌گیرنده، یگان دستگیر کننده، نوع قرار صادره و نام قاضی است. این رکوردها از نظر مفهومی ارتباطی به رفتار سارق نداشتند و می‌بایست از پایگاه داده حذف شوند. پس از اجرای مرحله آماده‌سازی داده‌ها، مشخصه‌های زیر به شرح جدول ۲ برای مدل‌سازی استفاده شده است:

جدول ۲. خصایص مورد استفاده برای ورودی الگوریتم

ردیف	عنوان	نام حوزه
۱	نوع جرم	CRIME_TYPE_ID
۲	محل وقوع جرم	OCCER_CRIME_ID
۳	جنسیت مجرم	SEX_ID
۴	میزان تحصیلات مجرم	EDUCATION_ID
۵	وضعیت تأهل مجرم	MARRIAGE_STATUS_ID
۶	ماه وقوع جرم	OCCER_MONTH_CRIME_ID
۷	سن مجرم	AGE_ID
۸	فصل وقوع جرم	OCCER_FASLE_CRIME_ID
۹	شغل	JOB_ID
۱۰	ساعت وقوع جرم	TIME_ID
۱۱	مکان ورود (روش ورود)	INTER_PLACE_ID
۱۲	مکان خروج (روش خروج)	OUT_PLACE_ID
۱۳	نوع خانه مورد سرقت گرفته	TYPE_HOME_ID



جدول ۳. هشت خوشه حاصل از پیاده‌سازی الگوریتم کای میانگین

نام خوشه	تعداد رکود اطلاعاتی	توصیف خوشه
۱	۵۷۸	مجرمان در وقت ظهر و بعدازظهر از اماکن خصوصی و عمومی ماشین سرقت می‌کنند. (میانگین سن ۳۲، زمان رخداد: ساعت ۱۶ عصر، نوع جرم: اماکن خصوصی، مکان وقوع: خیابان، جنسیت مجرم: مرد، میزان تحصیلات: زیر دیپلم، وضعیت تأهل: متأهل، ماه وقوع: اردیبهشت در فصل بهار)
۲	۲۰۳	(میانگین سن ۳۰، زمان رخداد: ساعت ۲۰ شب، نوع جرم: سرقت از خانه، مکان وقوع: خانه، جنسیت مجرم: مرد، میزان تحصیلات: زیر دیپلم، وضعیت تأهل: متأهل، ماه وقوع: خردادماه در فصل بهار)
۳	۳۶۴	(میانگین سن ۳۲، زمان وقوع جرم: ساعت ۶ صبح، نوع جرم: خانه، مکان وقوع: خانه، جنسیت مجرم: مرد، میزان تحصیلات: زیر دیپلم، وضعیت تأهل: متأهل، ماه وقوع: خردادماه در فصل بهار)
۴	۱۰۵۵	(میانگین سن ۳۱، زمان رخداد: ساعت ۱ و نیم بامداد، نوع جرم: خانه، مکان وقوع: خیابان، جنسیت مجرم: مرد، میزان تحصیلات: زیر دیپلم، وضعیت تأهل: متأهل، ماه وقوع: اردیبهشت در فصل بهار)
۵	۴۶۱	(میانگین سن ۳۱، زمان رخداد: ساعت ۳ صبح، نوع جرم: اماکن عمومی، مکان وقوع: خیابان، جنسیت مجرم: مرد، میزان تحصیلات: زیر دیپلم، وضعیت تأهل: متأهل، ماه وقوع: اردیبهشت در فصل بهار)
۶	۶۹۱	(میانگین سن ۳۰، زمان رخداد: ساعت ۲۲ شب، نوع جرم: اماکن خصوصی، مکان وقوع: خیابان، جنسیت مجرم: مرد، میزان تحصیلات: زیر دیپلم، وضعیت تأهل: متأهل، ماه وقوع: اردیبهشت در فصل بهار)
۷	۳۹۴	(میانگین سن ۳۱، زمان رخداد: ساعت ۱۲ ظهر، نوع جرم: اماکن عمومی، مکان وقوع: خیابان، جنسیت مجرم: مرد، میزان تحصیلات: زیر دیپلم، وضعیت تأهل: متأهل، ماه وقوع: اردیبهشت در فصل بهار)
۸	۱۲۵۴	(میانگین سن ۲۹، زمان رخداد: ساعت ۸ صبح، نوع جرم: اماکن عمومی، مکان وقوع: خیابان، جنسیت مجرم: مرد، میزان تحصیلات: زیر دیپلم، وضعیت تأهل: متأهل، ماه وقوع: اردیبهشت در فصل بهار)

در این بخش با استفاده از الگوریتم‌های داده‌کاوی به دسته‌بندی متغیرهای جرم پرداخته و دقت هریک از مدل‌ها و الگوریتم در پیش‌بینی متغیرها جرم (نوع جرم به‌عنوان متغیر وابسته) محاسبه می‌شود.

### پاسخ به پرسش چهارم پژوهش. تحلیل داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی (جرم سرقت) از نظر الگوریتم ناویز چگونه است؟

اطلاعات جدول‌های ۴ و ۵ نشان می‌دهند بیشترین درصد پیش‌بینی برای جرم سرقت از خانه برابر با ۷۳ درصد و کمترین درصد پیش‌بینی سرقت از اماکن خصوصی برابر با ۲۷ درصد است.

جدول ۴. ماتریس کانفوژن (Confusion) برای مشخصه‌های نوع جرم از نظر الگوریتم ناویز

نوع جرم	سرقت از خودرو	سرقت اماکن خصوصی	سرقت اماکن عمومی	سرقت از شخص	سرقت از خانه	جیب‌بری
سرقت از خودرو	۷۶	۱۰	۶۳	۱۲	۶	۲۴
سرقت اماکن خصوصی	۱۹	۴۹	۸۶	۱۴	۲۴	۳۷
سرقت اماکن عمومی	۴۶	۵۲	۱۵۱	۱۸	۳۴	۳۰
سرقت از شخص	۳۷	۳۰	۸۷	۳۰	۲۳	۳۱
سرقت از خانه	۱۶	۳۱	۴۶	۲	۲۴۸	۱۵
جیب‌بری	۲۳	۹	۴۵	۹	۳	۶۴

جدول ۵. ماتریس کانفوژن (Confusion) درصد و حساسیت پیش‌بینی برای مشخصه‌های نوع جرم با دقت ۰.۴۱۲ از نظر الگوریتم ناویز

نوع جرم	True Positive	False Positive	True Negative	False Negative	درصد پیش‌بینی	حساسیت	اختصاصی
سرقت از خودرو	۷۶	۱۴۱	۱۱۶۸	۱۱۵	۰.۳۵	۰.۳۹۷	۰.۸۹۲
سرقت اماکن خصوصی	۴۹	۱۳۲	۱۱۳۹	۱۸۰	۰.۲۷۱	۰.۲۱۳	۰.۸۹۶
سرقت اماکن عمومی	۱۵۱	۳۲۷	۸۴۲	۱۸۰	۰.۳۱۶	۰.۴۵۶	۰.۷۲۰
سرقت از شخص	۳۰	۵۵	۱۲۰۷	۲۰۸	۰.۳۵۳	۰.۱۲۶	۰.۹۵۶
سرقت از خانه	۲۴۸	۹۰	۱۰۵۲	۱۱۰	۰.۷۳۴	۰.۶۹۲	۰.۹۲۱
جیب‌بری	۶۴	۱۳۷	۱۲۱۰	۸۹	۰.۳۱۸	۰.۴۱۸	۰.۸۹۸

پاسخ به پرسش پنجم پژوهش. تحلیل داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی (جرم سرقت) از نظر الگوریتم قوانین بیزین چگونه است؟

بر اساس داده‌های جدول ۶ و ۷ دقت مدل قوانین بیزین در پیش‌بینی و تشخیص نوع جرم برابر با ۰.۵۱۳ است. داده‌ها نشان می‌دهند که با استفاده از ماتریس فیوژن، بیشترین درصد پیش‌بینی برای جیب‌بری برابر با ۱۰۰ درصد و کمترین درصد پیش‌بینی سرقت از اماکن خصوصی برابر با ۳۴ درصد است.

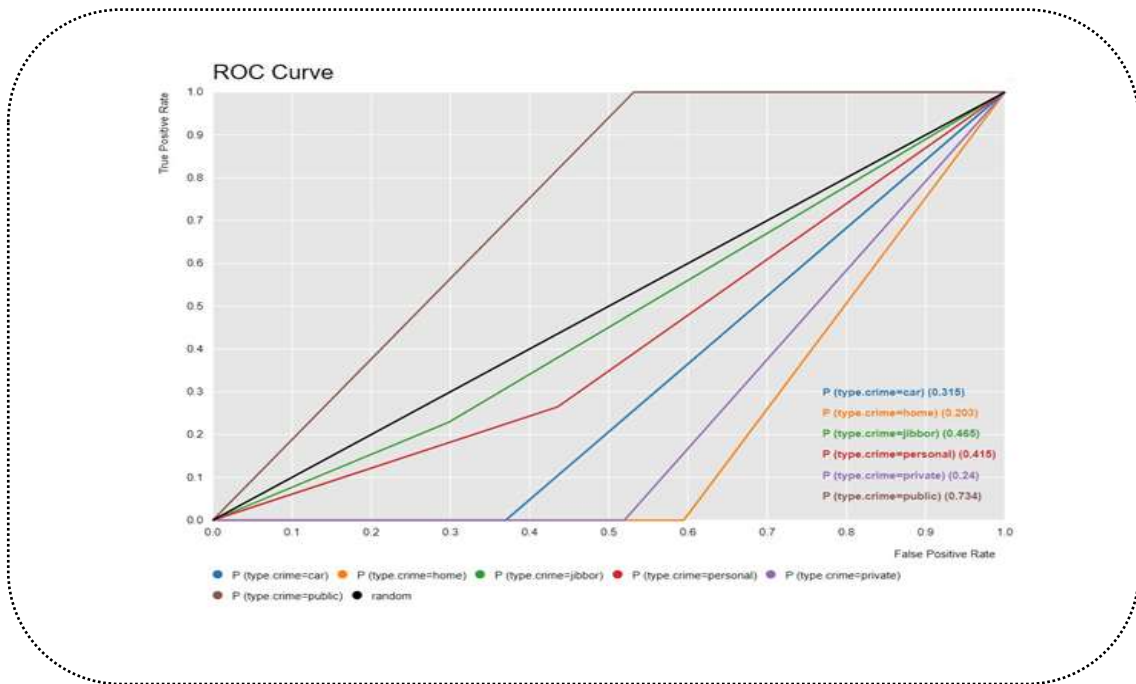
جدول ۶. ماتریس کانفوژن (Confusion) برای مشخصه‌های نوع جرم از نظر الگوریتم قوانین بیزین

نوع جرم	سرقت از خودرو	سرقت اماکن خصوصی	سرقت اماکن عمومی	سرقت از شخص	سرقت از خانه	جیب‌بری
سرقت از خودرو	۱۰۳	۷۳	۰	۰	۰	۰
سرقت اماکن خصوصی	۰	۲۳۷	۰	۰	۰	۰
سرقت اماکن عمومی	۳۱۳۲	۱۲۶	۱۲۵	۰	۶۲	۰
سرقت از شخص	۲۳	۹۲	۲۳	۶۹	۳۶	۰
سرقت از خانه	۴۴	۱۲۱	۰	۰	۱۷۹	۰
جیب‌بری	۷	۴۲	۲۲	۷	۲۱	۰

جدول ۷. ماتریس کانفوژن (Confusion) درصد و حساسیت پیش‌بینی برای مشخصه‌های نوع جرم با دقت ۰.۵۱۳ از نظر الگوریتم قوانین بیزین

نوع جرم	True Positive	False Positive	True Negative	False Negative	درصد پیش‌بینی	حساسیت	اختصاصی
سرقت از خودرو	۱۰۳	۱۰۵	۱۲۱۹	۷۳	۰.۴۹۵	۰.۵۸۹	۰.۹۲۱
سرقت اماکن خصوصی	۲۳۷	۴۵۴	۸۰۹	۰	۰.۳۴۳	۱	۰.۶۴۱
سرقت اماکن عمومی	۱۲۵	۴۵	۱۱۱۱	۲۱۹	۰.۷۳۵	۰.۳۶۳	۰.۹۶۱
سرقت از شخص	۶۹	۷	۱۲۵۰	۱۷۴	۰.۹۰۸	۰.۲۸۴	۰.۹۹۴
سرقت از خانه	۱۷۹	۱۱۹	۱۰۳۷	۱۶۵	۰.۶۰۱	۰.۵۲	۰.۸۹۷
جیب‌بری	۵۷	۰	۱۳۴۴	۹۹	۱۰۰	۰.۳۶۵	۱۰۰

اطلاعات منحنی در شکل ۲ میزان دقت و عملکرد قوانین بیزین را در تشخیص و پیش‌بینی انواع جرم سرقت (سرقت از منزل، خودرو و غیره) نشان می‌دهد. مشاهده می‌شود میزان عملکرد (دقت) این روش در پیش‌بینی دزدی از اماکن عمومی ۰.۷۳ و دقت (عملکرد) آن در پیش‌بینی دزدی از منزل (۰.۲۰۳) است. از این رو می‌توان گفت شبکه‌های عصبی سرقت از اماکن عمومی را نسبت به سایر سرقت‌ها بهتر و با دقت بیشتری پیش‌بینی می‌کنند (دقت این روش در پیش‌بینی تقریبی این نوع جرم سرقت ۷۳ درصد است) و جرم دزدی از منزل را نسبت به سایر سرقت‌ها بدتر و با دقت کمتری پیش‌بینی می‌کند (دقت آن در پیش‌بینی جرم سرقت از منزل فقط ۲۰ درصد است).



شکل ۲. منحنی ROC (receiver operating characteristic curve) برای عملکرد قوانین بیزین در پیش‌بینی انواع جرم

پاسخ به پرسش ششم پژوهش. تحلیل داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی (جرم سرقت) از نظر الگوریتم شبکه عصبی (MPL) چگونه است؟

اطلاعات جدول ۸ و ۹ نشان می‌دهند دقت مدل شبکه عصبی در پیش‌بینی و تشخیص نوع جرم برابر با ۰.۴۲ است. همچنین بیشترین درصد پیش‌بینی برای سرقت از خانه با ۷۷ درصد و کمترین درصد پیش‌بینی سرقت از شخص برابر با ۲۰ درصد است.

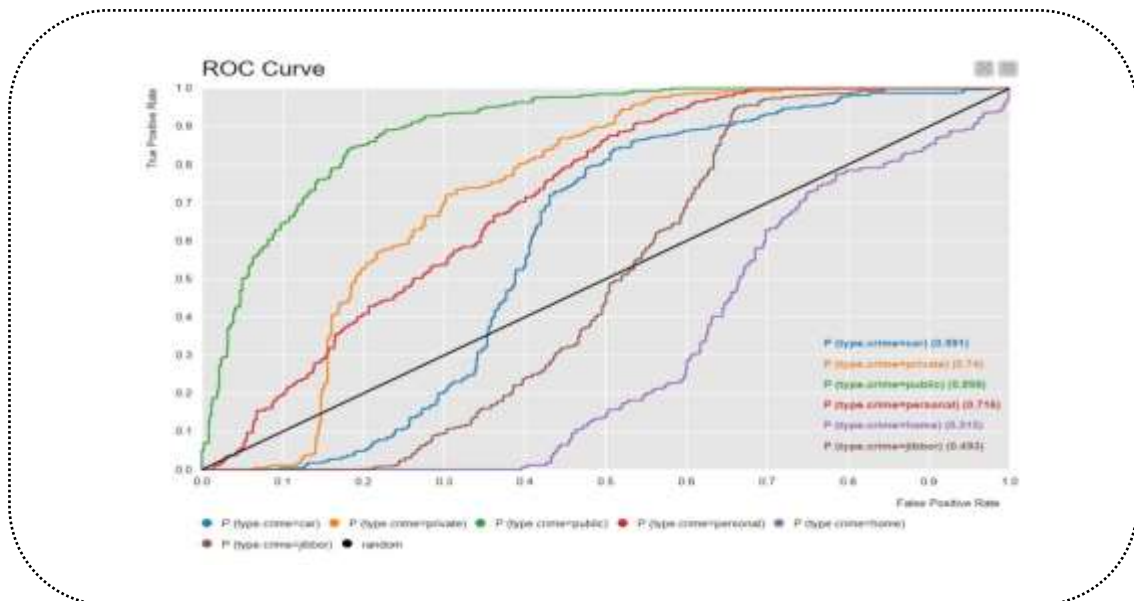
جدول ۸. ماتریس کانفوزن برای مشخصه‌های نوع جرم از نظر الگوریتم شبکه عصبی (MPL)

نوع جرم	سرقت از خودرو	سرقت اماکن خصوصی	سرقت اماکن عمومی	سرقت از شخص	سرقت از خانه	جیب‌بری
سرقت از خودرو	۶۱	۲۰	۷۷	۰	۱۱	۱۹
سرقت اماکن خصوصی	۱۶	۳۰	۱۱۷	۷	۱۲	۱۹
سرقت اماکن عمومی	۳۵	۳۱	۲۰۳	۱۱	۲۹	۴۴
سرقت از شخص	۲۶	۲۰	۱۳۱	۷	۱۸	۲۴
سرقت از خانه	۱۲	۹	۷۶	۷	۲۵۹	۱۸
جیب‌بری	۱۳	۵	۵۵	۲	۶	۷۰

جدول ۹. ماتریس کانفوژن (Confusion) درصد و حساسیت پیش‌بینی برای مشخصه‌های نوع جرم با دقت ۰.۴۲ از نظر الگوریتم شبکه عصبی (MPL)

نوع جرم	True Positive	False Positive	True Negative	False negative	درصد پیش‌بینی	حساسیت	اختصاصی
سرقت از خودرو	۶۱	۱۰۲	۱۲۱۰	۱۲۷	۰.۳۷۴	۰.۳۲۴	۰.۹۲۲
سرقت اماکن خصوصی	۳۰	۸۵	۱۲۱۴	۱۷۱	۰.۲۶۱	۰.۱۴۹	۰.۹۳۵
سرقت اماکن عمومی	۲۰۳	۴۵۶	۶۹۱	۱۵۰	۰.۳۰۸	۰.۵۷۵	۰.۶۰۲
سرقت از شخص	۷	۲۷	۱۲۴۷	۲۱۹	۰.۲۰۶	۰.۰۳۱	۰.۹۷۹
سرقت از خانه	۲۵۹	۷۶	۱۰۴۳	۱۲۲	۰.۷۷۳	۰.۶۸	۰.۹۳۳
جیب‌بری	۷۰	۱۲۴	۱۲۲۵	۸۱	۰.۳۶۱	۰.۴۶۴	۰.۹۰۸

اطلاعات منحنی شکل ۳ میزان دقت و عملکرد شبکه‌ها - شبکه‌های عصبی (MPL) را در تشخیص و پیش‌بینی انواع جرم سرقت (سرقت از منزل، خودرو و غیره) نشان می‌دهد. مشاهده می‌شود میزان عملکرد (دقت) این روش در پیش‌بینی دزدی از اماکن عمومی ۰.۸۹۹ و دقت (عملکرد) آن در پیش‌بینی دزدی از منزل ۰.۳۱۳ است. از این رو می‌توان گفت شبکه‌های عصبی سرقت از اماکن عمومی را نسبت به سایر سرقت‌ها بهتر و با دقت بیشتری پیش‌بینی می‌کنند (دقت این روش در پیش‌بینی این نوع جرم سرقت حدوداً ۹۰ درصد است) و جرم دزدی از منزل را نسبت به سایر سرقت‌ها بدتر و با دقت کمتری پیش‌بینی می‌کنند (دقت آن در پیش‌بینی جرم سرقت از منزل فقط ۳۴ درصد است).



شکل ۳. نمودار ROC (receiver operating characteristic curve) برای عملکرد شبکه‌ها شبکه‌های عصبی در پیش‌بینی انواع جرم سرقت از نظر الگوریتم شبکه عصبی (MPL)

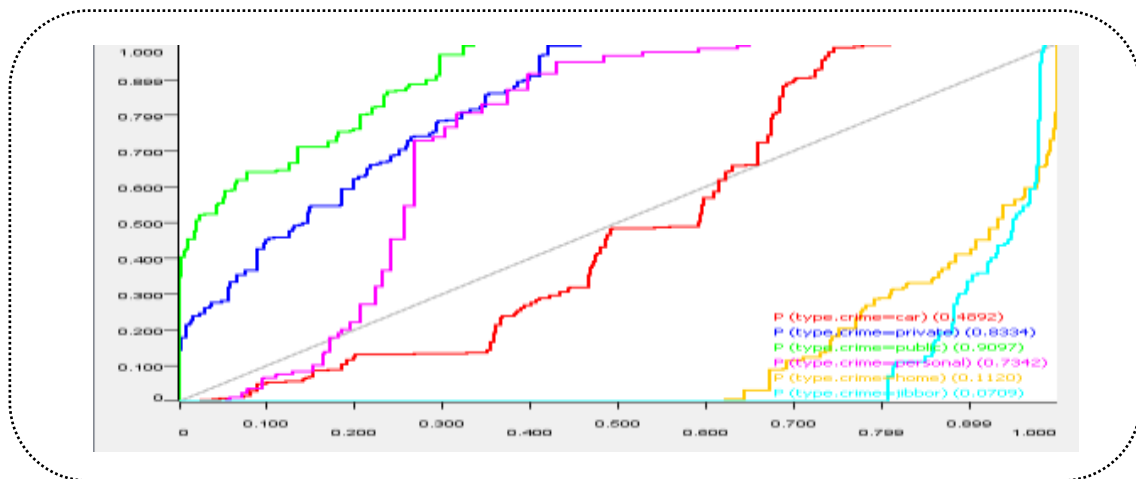
## پاسخ به پرسش هفتم پژوهش. تحلیل داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی (جرم سرقت) از نظر بردار ماشین پشتیبان چگونه است؟

جدول ۱۰ نشان می‌دهد دقت مدل بردار ماشین پشتیبان در پیش‌بینی و تشخیص نوع جرم برابر با ۰.۱۸۷ است. بر اساس داده‌ها بیشترین درصد پیش‌بینی برای سرقت از اماکن عمومی ۲۹ درصد و کمترین درصد پیش‌بینی سرقت از خودرو برابر با صفر درصد است.

جدول ۱۰. ماتریس کانفوزن (Confusion) درصد و حساسیت پیش‌بینی برای مشخصه‌های نوع جرم با دقت ۰.۱۸۷ از نظر بردار ماشین

نوع جرم	True Positive	False Positive	True Negative	False Negative	درصد پیش‌بینی	حساسیت	اختصاصی
سرقت از خودرو	۰	۱۷۶	۱۳۲۴	۰	۰	-	۰.۸۸۳
سرقت اماکن خصوصی	۴۰	۱۹۷	۱۰۴۰	۲۲۳	۰.۱۶۹	۰.۱۵۲	۰.۸۴۱
سرقت اماکن عمومی	۱۰۰	۲۴۴	۸۱۴	۳۴۲	۰.۲۹۱	۰.۲۲۶	۰.۷۶۹
سرقت از شخص	۸۴	۲۶۰	۸۴۸	۳۰۸	۰.۲۴۴	۰.۲۱۴	۰.۷۶۵
سرقت از خانه	۱۹	۲۲۴	۱۱۷۷	۸۰	۰.۰۷۸	۰.۱۹۲	۰.۸۴
جیب‌بری	۳۷	۱۱۹	۱۰۷۷	۲۶۷	۰.۲۳۷	۰.۱۲۲	۰.۹۰۱

اطلاعات منحنی شکل ۴ میزان دقت و عملکرد بردار ماشین پشتیبان را در تشخیص و پیش‌بینی انواع جرم سرقت (سرقت از منزل، خودرو و غیره) نشان می‌دهد. مشاهده می‌شود میزان عملکرد (دقت) این روش در پیش‌بینی دزدی از اماکن عمومی ۰.۹۰۹ و دقت (عملکرد) آن در پیش‌بینی جرم جیب‌بری (۰.۰۷) است. از این رو می‌توان گفت شبکه‌ها - شبکه‌های عصبی سرقت از اماکن عمومی را نسبت به سایر سرقت‌ها بهتر و با دقت بیشتری پیش‌بینی می‌کنند (دقت این روش در حدود پیش‌بینی این نوع جرم سرقت ۸۳ درصد است) و جرم جیب‌بری را نسبت به سایر سرقت‌ها بدتر و با دقت کمتری پیش‌بینی می‌کند (دقت آن در پیش‌بینی جرم سرقت از منزل فقط ۷ درصد است).



شکل ۴. منحنی ROC(receiver operating characteristic curve) برای عملکرد بردار ماشین پشتیبان در پیش‌بینی انواع جرم



## بحث و نتیجه گیری

در این پژوهش با استفاده از روش های خوشه بندی (مشخصاً کای میانگین) و الگوریتم های دسته بندی از قبیل قوانین بیزین، ناو بیزی، ماشین بردار پشتیبان و شبکه های عصبی به پیش بینی و تشخیص یک الگویی از جرم سرقت پرداخته شد. نتایج خوشه بندی حاصل از کای میانگین نشان داد که مشخصه ها و متغیرهای جرم سرقت در هشت خوشه قرار دارند که تمامی خوشه ها در اغلب مشخصه های جرم سرقت به جز ساعت وقوع و سن مجرم یکسان بودند. خوشه اول مرتبط به سارقان خودرو، خوشه های ۲، ۳ و ۴ مرتبط به سارقان خانه (منزل)، خوشه های ۵، ۷ و ۸ نیز مرتبط به سارقان اماکن عمومی و خوشه ۶ مرتبط به سارقان اماکن خصوصی است.

مقایسه نتایج این پژوهش با نتایج پژوهش اسکندری و همکاران (۱۳۹۰) نشان داد که خصایص (متغیرهای) جرم از قبیل نوع جرم انتخابی، جنسیت افراد، گروه سنی، شغل افراد، ساعت ارتکاب جرم، سال و روز ارتکاب جرم و روش ارتکاب جرم یکسان است و تنها وجه تمایز افراد متعلق به دو خوشه وضعیت تأهل سارق و ماه وقوع جرم است. در تحقیق اسکندری و همکاران (۱۳۹۰) خوشه بندی، مجرمان بر اساس نوع جرم انتخابی با پژوهش حاضر متفاوت است؛ چراکه در تحقیق حاضر نوع جرم به پنج نوع سرقت از خانه - سرقت از خودرو، سرقت از اماکن خصوصی، سرقت از اماکن خصوصی و سرقت از شخص (جیب بری و کیف قاپی) محدود می شود و به طور خاص به سرقت های زیرمجموعه هریک از سرقت ها اشاره نشده است.

نتایج الگوریتم های دسته بندی داده کاوی نشان داد دقت شبکه های عصبی، ماشین بردار پشتیبان، ناو بیزی و قوانین بیزین به ترتیب برابر با ۰.۴۲، ۰.۱۸۷، ۰.۴۱۲ و ۰.۵۱۳ است. از این رو می توان گفت در تشخیص و پیش بینی یک الگوی مناسب از جرم سرقت (نوع جرم) روش قوانین بیزین مناسب ترین روش و دارای دقت بالاتری نسبت به سایرین است. همچنین نمودار ROC نشان داد دقت روش قوانین بیزین و شبکه ها - شبکه های عصبی (MPL) و بردار ماشین پشتیبان در پیش بینی جرم سرقت از اماکن عمومی نسبت به سایر سرقت ها بیشتر و در موارد مشابه پیش بینی سرقت از منزل دارای دقت کمتری نسبت به سایر سرقت ها است. البته دقت پیش بینی روش بردار ماشین پشتیبان (حدود ۹۱ درصد) در پیش بینی سرقت از اماکن عمومی بهتر از قوانین بیزین (حدود ۷۳ درصد) و شبکه های عصبی (حدود ۹۰ درصد) است. همچنین نمودار ROC برای روش بردار ماشین پشتیبان نشان داد که در پیش بینی جرم جیب بری نسبت به سایر سرقت ها دارای دقت کمتری است (۷ درصد).

مقایسه نتایج این پژوهش با پژوهش اسکندری و همکاران و استفاده از روش بیزنت، میزان دقت و درستی در تشخیص برخی از ویژگی های مجرمان از قبیل سن مجرم، سال وقوع، ساعت وقوع جرم، ماه وقوع جرم، وضعیت تأهل مجرم و غیره نشانگر میزان دقت و درستی این الگوریتم (بین ۷۵ الی ۸۰ درصد) برای کلیه ویژگی های مجرمان است که این میزان دقت قابل قبول است. از سویی پراکندگی میزان دقت الگوریتم در پیش بینی کلیه ویژگی ها کم است و می توان از این الگوریتم هم زمان برای تشخیص و پیش بینی ویژگی های مورد اشاره استفاده کرد؛ این در حالی است که در پژوهش حاضر، به عنوان مثال پراکندگی میزان دقت روش شبکه های عصبی در پیش بینی انواع سرقت بین ۳۴ الی ۸۹ درصد است. به بیان بهتر می توان گفت که برای هر یک از سرقت ها باید از یک الگوریتم یا روش داده کاوی استفاده کرد.

به استناد نتایج روش های علم سنجی و مشخصاً داده کاوی در مورد کشف جرائم به منظور تصمیم گیری های بهتر و دقیق تر در سازمان های پلیسی و قضایی پیشنهاد می شود از روش های ترکیبی داده کاوی و هوش مصنوعی در تشخیص

ارائه چارچوب برای سنجش و ارزیابی مطالعات داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی

و شناسایی سارقانی که بیشترین و کمترین سابقه را در ارتکاب جرم سرقت دارند، استفاده نمایند تا دقت بیشتری در ارائه اطلاعات صورت گیرد. همچنین می‌توان یک سیستم هوشمند خبره طراحی کرد تا بتوان کیفیت مبادرت جرم را برای مجرمین در آینده پیش‌بینی کرد. در این تحقیق محدودیت‌هایی وجود داشت که اولین و مهم‌ترین آن مجموعه داده‌های مرتبط به متغیرهای جرم بود که برخی از متغیرهای جرم‌شناسی به دلیل ناقص بودن داده‌های آن در این تحقیق مورد بررسی و تحلیل قرار نگرفته‌اند. افزون بر این نتایج نشان داد جرائم ناهنجاری‌های اجتماعی هستند که ابعاد وسیعی دارند و کشف آن‌ها بسیار دشوار است. از این رو با گسترش فناوری اطلاعات و استفاده از داده‌کاوی و شیوه‌های علم‌سنجی در نهادهای دولتی و خصوصی و احداث بانک‌های اطلاعاتی، استفاده از داده‌کاوی به‌عنوان ابزار نرم‌افزاری توانا و کم‌هزینه و بهتر می‌تواند به کشف، شناسایی، پیش‌بینی و پیشگیری از الگوهای جرم و وقوع آن کمک کند.

### پیشنهاد‌های اجرایی پژوهش

- پیشنهاد می‌شود از یافته‌های این پژوهش برای رشد و توسعه علم و سیاست‌گذاری صحیح در مبارزه با آسیب‌های اجتماعی استفاده شود تا بتوانیم:
- از مزایای تجزیه و تحلیل پیش‌بینانه داده‌ها در شبکه‌ها جهت پیشگیری از ناهنجاری‌های اجتماعی استفاده شود.
  - برای کاهش ناهنجاری‌های اجتماعی و کمک به نیروهای انتظامی بهره‌جست.
  - انجام پیگیری به‌موقع، دقیق و قانونی برای کمک به سیاست‌گذاران کشور و افزایش ایمنی و امنیت دسترسی به اطلاعات در مورد ناهنجاری‌های اجتماعی به‌منظور بهینه‌سازی شبکه‌ها.
  - آسیب‌شناسی پژوهش‌های این حوزه برای کاهش هزینه‌های عملیاتی و انجام تحقیقات کاربردی توسط پژوهشگران و دانشمندان در این حوزه.

### پیشنهاد برای پژوهش‌های آتی

- دانشمندان و محققان این رشته با استفاده از روش‌های داده‌کاوی برای پیشگیری و مبارزه با ناهنجاری‌های اجتماعی اقدام کنند.
- در صورت تمایل به پژوهش مجدد در مورد ناهنجاری‌های اجتماعی پیشنهاد می‌شود تحقیق در موضوعات بیشتر و مناطق جغرافیایی بزرگ‌تری صورت گیرد و تفاوت‌های اجتماعی و فرهنگی در آن مورد توجه واقع شود.
- شاخص‌های خوشه‌بندی داده‌کاوی به شکل جدا از هم مورد بررسی قرار گیرند.
- ارتباط میان سرقت و عملکرد اجتماع نیز مورد مطالعه قرار گیرد.
- بررسی تأثیر ابعاد ارزش ثروت اشخاص و بیماری‌های روانی، فقر و نابرابری‌های اجتماعی بر عملکرد داده‌کاوی و مبارزه با آسیب‌های اجتماعی (جرم سرقت و غیره) از دیگر پیشنهادها برای پژوهش‌های آتی است.

### تقدیر و تشکر

مقاله حاضر برگرفته از رساله دکتری با عنوان «ارائه چارچوب برای سنجش و ارزیابی مطالعات داده‌های شبکه‌ای در مبارزه با ناهنجاری‌های اجتماعی» در پردیس بین‌المللی کیش است.

## فهرست منابع

- احمدوند، ع.، و آخوندزاده، ا. (۱۳۸۹). چارچوب کاربردی تکنیک‌های داده‌کاوی در مدل‌سازی جرائم. *توسعه سازمانی پلیس*، ۷(۳۰)، ۱۱-۲۲. [http://pod.jrl.police.ir/article\\_9364.html](http://pod.jrl.police.ir/article_9364.html)
- اسکندری، ح.، علیزاده، س.، و کاظمی، پ. (۱۳۹۰). کاربرد داده‌کاوی در شناسایی و کشف الگوهای پنهان جرم سرقت. *پژوهش‌نامه نظم و امنیت انتظامی*، ۴(۱۶)، ۳۵-۵۶. [http://osra.jrl.police.ir/article\\_9745.html](http://osra.jrl.police.ir/article_9745.html)
- حاجی ده آبادی، ا.، و عسگری م. (۱۳۹۸). ترک فعل در جنایات غیر عمدی: از معیار تا مصداق. *فصلنامه پژوهش حقوق کیفری*، ۸(۲۹)، ۲۳۳-۲۵۹. <https://doi.org/10.22054/jclr.2019.37121.1790>
- خلیلی، م.، و نوایی فیجانی، آ. (۱۳۹۵). کاربرد ابر داده در مبارزه با جرم و جنایت [مقاله کنفرانسی]. *همایش ملی پژوهش‌های نوین در علوم و فناوری*. <https://civilica.com/doc/531474/certificate/print>
- کیوان‌پور، م.، جاویده، م.، پهلوان زاده، ا.، و ابراهیمی، م. (۱۳۸۸). تحلیل و کشف جرم با بهره‌گیری از روش‌های داده‌کاوی [مقاله کنفرانسی]. *دومین کنفرانس داده‌کاوی ایران، دانشگاه امیرکبیر*. <https://civilica.com/doc/70469>
- کاظمی، پ.، و حسین‌پور، ج. (۱۳۸۸). کاربرد داده‌کاوی در سازمان‌های پلیسی و قضایی به‌منظور شناسایی الگوهای جرم و کشف جرائم. *فصلنامه علمی کارآگاه*، ۳(۸)، ۳۲-۶۳. [http://det.jrl.police.ir/article\\_10620.html](http://det.jrl.police.ir/article_10620.html)
- Agarwal, J., Nagpal, R., & Sehgal, R. (2014). Reliability Of Component Based Software System Using Soft Computing Techniques – A Review. *International Journal of Computer Applications*, 94(2), 12-16. <https://www.ijcaonline.org/archives/volume94/number2/16314-5548/>
- Ahmadvand, A., & Akhonzadeh, E. (2011). Applied Framework for Data Mining Techniques in Crime Modeling. *Police Organizational Development*, 7(30), 11-22. [http://pod.jrl.police.ir/article\\_9364.html?lang=en](http://pod.jrl.police.ir/article_9364.html?lang=en) [In Persian].
- Çorapçıoğlu, A., & Erdoğan, S. (2004). A cross-sectional study on expression of anger and factors associated with criminal recidivism in prisoners with prior offences. *Forensic science international*, 140(2-3), 167-174. <https://doi.org/10.1016/j.forsciint.2003.11.019>
- D'Alessio, S. J., & Stolzenberg, L. (2010). Do Cities Influence Co-Offending? *Journal of Criminal Justice*, 38(4), 711-719. <https://doi.org/10.1016/j.jcrimjus.2010.04.045>
- Deadman, D. (2003). Forecasting Residential Burglary. *International Journal of Forecasting*, 19(4), 567-578. [https://doi.org/10.1016/S0169-2070\(03\)00091-8](https://doi.org/10.1016/S0169-2070(03)00091-8)
- Eskandari, E., Alizadeh, S., & Kazemi, P. (2011). Crime analysis using data mining techniques. *Quarterly of order and security guards*, 4(16), 35-56. [http://osra.jrl.police.ir/article\\_9745.html?lang=en](http://osra.jrl.police.ir/article_9745.html?lang=en) [In Persian].
- Hajidehabadi, A., & Asgari Morovat, A. (2020). Manslaughter by Omission: from Criterion to Instance. *Journal of Criminal Law Research*, 8(29), 233-259. <https://doi.org/10.22054/jclr.2019.37121.1790> [In Persian].
- Karlis, D., & Meligkotsidou, L. (2007). Finite Mixtures of Multivariate Poisson Distributions with Application. *Journal of Statistical Planning and Inference*, 137(6), 1942-1960. <https://doi.org/10.1016/j.jspi.2006.07.001>

- Kazemi, P., & Hosseinpour, J. (2008). The use of data mining in police and judicial organizations to identify crime patterns and detect crimes. *Detective Journal*, 3(8) 32-63. [http://det.jrl.police.ir/article\\_10620.html](http://det.jrl.police.ir/article_10620.html) [In Persian].
- Keyvanpour, M., Javideh, M., Pahlavanzadh, A., & Ebrahimi, M. (2008). Analysis and detection of crime using data mining methods [Conference presentation]. *The second data mining conference of Iran*, Amir Kabir University. <https://civilica.com/doc/70469> [In Persian].
- Khalili, M., & Navaei Fijani, A. (2016). Application of big data to fight crime [Conference presentation]. National conference of new researches in science and technology. <https://civilica.com/doc/531474/certificate/print> [In Persian].
- Li, S. T., Kuo, S. C., & Tsai, F. C. (2010). An intelligent decision-support model using FSOM and rule extraction for crime prevention. *Expert Systems with Applications*, 37(10), 7108-7119. <https://doi.org/10.1016/j.eswa.2010.03.004>
- Liu, H., & Brown, D. E. (2003). Criminal incident prediction using a point-pattern-based density model. *International journal of forecasting*, 19(4), 603-622. [https://doi.org/10.1016/S0169-2070\(03\)00094-3](https://doi.org/10.1016/S0169-2070(03)00094-3)
- Lopez-Rojas, E. A., & Axelsson, S. (2012). Multi agent based simulation (mabs) of financial transactions for anti money laundering (aml) [Conference Paper]. In *Nordic Conference on Secure IT Systems*. Blekinge Institute of Technology. <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A834702&dsid=-1975>
- Malathi, A., & Santhosh, B. (2011). An Enhanced Algorithm to Predict a Future Crime using Data Mining. *International Journal of Computer Applications*, 21, <https://www.ijcaonline.org/volume21/number1/pxc3873335.pdf>
- Mande, U., Srinivas, Y., & Murthy, J. V. R. (2012). An intelligent analysis of crime data using data mining & auto correlation models. *Int J Eng Res Appl (IJERA)*, 2(4), 149-153. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=96b814e73cb7140de589555a53f26d0eafdc7a7e>
- Moon, B., McCluskey, J. D., & McCluskey, C. P. (2010). A General Theory of Crime and Computer Crime: An Empirical Test. *Journal Criminal Justice*, 38(4), 767-772. <https://doi.org/10.1016/j.jcrimjus.2010.05.003>
- Murtagh, F., Ganz, A., & McKie, S. (2009). The Structure of Narrative: The Case of Film Scripts. *Pattern Recognition*, 42, 302 – 312. <https://doi.org/10.1016/j.patcog.2008.05.026>
- Oatley, G. C., & Ewart, B. W. (2003). Crimes Analysis Software: 'Pins in Maps', Clustering and Bayes Net Prediction. *Expert Systems with Applications*, 25(4), 569-588. [https://doi.org/10.1016/S0957-4174\(03\)00097-6](https://doi.org/10.1016/S0957-4174(03)00097-6)
- Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *ArXiv preprint arXiv:1009.6119*. <https://arxiv.org/pdf/1009.6119>
- Xue, Y., & Brown, D. E. (2006). Spatial Analysis with Preference Specification of Latent Decision Makers for Criminal Event Prediction. *Decision Support Systems*, 41(3), 560- 573. <https://doi.org/10.1016/j.dss.2004.06.007>