تحلیل جرایم اخلاقی فضای سایبر با رویکرد داده‌کاوی

غلامرضا شاه‌محمدی[1]، مصطفی رجبی[2]

تاریخ دریافت:۹۷/۱۰/۰۵ تاریخ پذیرش:۹۷/۱۲/۰۳

چکیده

زمینه و هدف: جرایم سایبر در طیف جرایم نوظهوری هستند که همزمان با رشد فناوری‌های جدید و بیشتر مبتنی بر اینترنت به‌طور مداوم در حال تکامل و پیچیده‌تر شدن هستند. یکی از جرایم شایع در فضای مجازی، جرایم اخلاقی فضای سایبر است که با توجه به ارتباط آن با آبرو و حیثیت افراد، از اهمیت بسیار بالایی برخوردار است. هدف این پژوهش، تحلیل جرایم اخلاقی فضای سایبر با رویکرد داده‌کاوی است.

روش: این پژوهش از نظر هدف کاربردی و از نظر نوع، تحلیل محتوا است. جامعه آماری پژوهش، جرایم سایبری است. حجم نمونه، جرایم اخلاقی فضای سایبر در سال‌های ۱۳۸۹ تا ۱۳۹۴ است.

یافته‌ها و نتایج: بهترین مدل برای تحلیل و استخراج قوانین حاکم بر داده‌های جرایم اخلاقی فضای سایبر، مدل درخت تصمیم C5 است. نتایج پژوهش نشان داد مشخصه‌های تاثیرگذار بر وقوع جرایم اخلاقی فضای سایبر به ترتیب شغل، تحصیلات، جنسیت، سن، تاهل و نقش فرد است. در این پژوهش، قوانین حاکم بر جرم اخلاقی فضای سایبر، احصاء که با تحلیل این قوانین می‌توان راهکارهای پیشگیرانه ارائه داد.

کلیدواژه‌ها

جرایم اخلاقی فضای سایبر، تحلیل جرایم، داده‌کاوی، روش‌های پیش‌بینانه، مدل C5.

---

۱. دانشیار مهندسی کامپیوتر دانشگاه علوم انتظامی امین. (نویسنده مسئول). رایانامه:
Shah_mohammadi@yahoo.co.uk

۲. دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات

مقدمه

براساس گزارش‌های جهانی تا سال ۲۰۲۰ میلادی، ۵۰ میلیارد دستگاه متصل به اینترنت وجود خواهد داشت و ۱۵ میلیارد از این میزان را تلفن‌های همراه و تبلت‌ها تشکیل می‌دهند. بخشی دیگر از این آمارها حاکی از آن است که تا پایان این دهه، پهنای باند شبکه‌های تلفن همراه ۹۰ درصد از جمعیت کل جهان را تحت پوشش قرار خواهد داد (ارجمندی، ۱۳۹۳، ص ۲). این فضا ماحصل پیشرفت سریع اینترنت و رشد اطلاعات و همه‌گیرشدن آن است. پژوهشگران، اینترنت را منبع عظیم اطلاعات می‌دانند که با توجه به قابلیت‌های زیاد، جهان را به یک شهر کوچک تبدیل کرده است. ولی فضای سایبری با وجود مزایای زیاد و آثار مثبت فراوان، منشأ تهدیدهایی جدی برای کلیه افراد، سازمان‌ها و کشورهای جهان است (جلالی، ۱۳۹۱، ص ۷). فضای سایبری دارای ویژگی‌های خاصی مانند هزینه پایین ورود، گمنامی و سهولت انجام اعمال مختلف، جهانی و فرامرزی بودن، دسترسی دائم و آسان به اطلاعات، جذابیت و تنوع، وابسته‌نبودن به زمان و مکان خاص، سهولت و سرعت بالای تبادل اطلاعات است (شاه‌محمدی و اکباتانی، ۱۳۹۴، ص ۲۷) . این ویژگی‌ها علاوه بر این‌که سبب مهاجرت جرایم از فضای فیزیکی به فضای مجازی شده است، موجب بروز جرایم نوینی نیز شده است که قابل مقایسه با هیچ‌یک از جرایم موجود نبوده و از نظر دامنه تأثیر، بسیار خطرناک‌تر است (شاه‌محمدی و تاهو، ۱۳۹۳، ص ۱۰۰). در نتیجه ارتکاب جرایم در فضای مجازی به سرعت درحال افزایش است و آمار پلیس فتا[1] بیانگر این موضوع است (سایت پلیس فتا).

باتوجه به روند رو به رشد وقوع جرایم سایبری و مسؤلیت پلیس فتا در کشف و پیشگیری از جرایم سایبری، تحلیل جرایم سایبری به‌منظور بررسی علل وقوع این

۱. پلیس فضای تبادل اطلاعات نیروی انتظامی

جرایم و ارائه راهکارهای پیشگیرانه ضروری است. در این پژوهش، با تحلیل جرایم سایبری، ریشه‌ها و علل وقوع این جرایم بررسی، تا از این طریق بتوان راهکارهای پیشگیرانه ارائه داد. در واقع اگر پلیس فتا بداند چه افراد با چه ویژگی‌هایی از نظر سن، جنسیت، سطح تحصیلات، مرتکب یا قربانی جرایم اخلاقی سایبر می‌شوند، می‌تواند راهکارهای پیشگیرانه موثری برای آنها ارائه دهد.

برای تحلیل جرایم، روش‌های مختلفی ارائه شده است که روش داده‌کاوی یکی از رویکردهای موثر در تحلیل داده‌های جرایم است. در دهه اخیر از داده‌کاوی در سازمان‌های پلیسی خارج و داخل کشور نیز استفاده شده است. با وجود بانک اطلاعات جرایم سایبری در پلیس فتا و لزوم تحلیل جرایم سایبری، هدف این پژوهش تحلیل جرایم سایبری حوزه اخلاقی با استفاده از روش داده‌کاوی است تا از این طریق بتوان راهکارهای موثری برای پیشگیری از جرایم سایبری ارائه داد. در این پژوهش سعی می‌شود به این پرسش پاسخ داده شود که چگونه می‌توان با استفاده از تحلیل جرایم سایبری حوزه اخلاقی مبتنی بر داده‌کاوی، علل وقوع این جرایم را ریشه‌یابی کرد.

از نظر پیشینه می‌توان گفت داده‌کاوی دارای کاربرد وسیع در حوزه‌های مختلف ازجمله صنعتی، پزشکی، ارتباطات، کشاورزی، انرژی، علوم اجتماعی، فرهنگی، سیاسی، اقتصادی، بازرگانی، نظامی و آموزشی است، به‌گونه‌ای که امروزه مرز و محدودیتی برای کاربرد این دانش در نظر گرفته نمی‌شود. در حوزه داده‌کاوی جرایم کارهای خوبی انجام شده است که در ادامه به برخی از آن‌ها پرداخته می‌شود.

۱- فنون‌های داده‌کاوی در دسته‌بندی جرائم سایبری: داده‌کاوی و مدیریت جرایم سایبری یک برنامه جالب است که در آن نقش مهمی در مدیریت اطلاعات جرم‌شناختی ایفا می‌کند. در این پژوهش فنون داده‌کاوی برای تحلیل داده‌های وب از طریق کلاس‌بندی و خوشه‌بندی جرایم سایبری مورد استفاده قرار گرفت. در سناریوی

پیشنهادی، ویژگی‌ها و روابط را در صفحه وب استخراج و سناریوی جرم را بازسازی می‌کند (ساهو[۱]، ۲۰۱۷).

۲-استفاده از داده‌کاوی در جرایم سایبری: در سراسر جهان، افراد زیادی در حوزه‌های مختلف دسترسی به اینترنت دارند. هنگامی سرویس‌گرها خدماتی را در اختیار کاربران قرار می‌دهند، عملکردی وجود دارد که می‌توان فعالیت کاربران رادر فایل‌های شرح وقایع[۲] مشاهده کرد. این فایل‌ها شامل شرح مفصلی از فعالیت‌های کاربران در شبکه مانند آدرس IP، زمان ورود و خروج، رفتار کاربر و... را نشان می‌دهد. حملات مختلفی در اینترنت وجود دارد. تمرکز این پژوهش حملات DDOS با کمک فنون تشخیص الگو در داده‌کاوی است. DDOS یک حمله بسیار خطرناک است که به منابع اطلاعاتی یک سازمان آسیب می‌رساند، این حمله از طریق ارسال پیام‌های زیاد به‌سمت سرویس‌گر انجام می‌گیرد. آنان در پژوهش خود در مورد امنیت سایبر، جنایات سایبری، انواع آنها، تشخیص الگو، خوشه‌بندی و شناسایی حملات انکار سرویس از طریق داده‌کاوی بحث کرده‌اند (علی‌خان[۳]، کمار پرادهان[۴] و فاطیما[۵]، ۲۰۱۷، صص ۱-۳).

۳-کشف آزار و اذیت‌های سایبری متنی با استفاده از داده‌کاوی: امروزه شبکه‌های اجتماعی و استفاده از آن توسط عموم مردم باعث شده است که عده‌ای از کاربران ناآگاه مورد تهدید و تجاوز سایبری قرارگیرند که اساس این تهدیدها به اشتراک گذاشتن اطلاعات شخصی در محیط اینترنت توسط کاربران است. مفهوم قلدری اینترنتی و آزار و اذیت اینترنتی از سال ۱۹۹۱ و ۱۹۹۳ تعریف شده است. اغلب

---

۱. Sahu
۲. log
۳. Ali Khan
۴. Kumar Pradhan
۵. Fatima

تهدیدات اینترنتی از طریق شبکه‌های اجتماعی، پیام‌های متنی، اتاق‌های گفتگو، رایانامه و...؛ موضوعات مشترکی است که مورد سوء استفاده سایبری قرار می‌گیرند و براساس مشخصه‌های شخصی افراد نظیر ظاهر فیزیکی، نژاد و قومیت، جنسیت و هویت جنسی، هوش، پذیرش و طرد اجتماعی و سایر مولفه است؛ در این پژوهش یک نظرسنجی از کاربران انجام و شبیه‌سازی در شبکه‌های اجتماعی صورت گرفته است تا بتوان با استفاده از داده‌کاوی از کاربران مراقبت کرد (سینقال[۱]، ۲۰۱۳، صص ۵۶۷-۵۶۹).

٤-شناسایی رهبران شبکه‌های جنایی با استفاده از درخت پنهان شبکه‌های جنایی: کمال تاها[۲]، از درخت پنهان یک شبکه جنایی برای شناسایی رهبران آن استفاده کردند. آن‌ها سیستم تحلیل فانزیک به نام «ای.سی.ال فایندر»[۳] جهت شناسایی اعضای تاثیرگذار[۴] یک سازمان جنایتکار معرفی کردند. محققان جنایی معمولا به‌دنبال شناسایی اعضای تاثیرگذار سازمان‌های جنایتکار هستند، زیرا حذف آنها به احتمال زیاد مانع عملیات خرابکارانه این سازمان‌ها می‌شود. «ای.سی.ال فایندر»، ابتدا شبکه‌ای را براساس داده‌های ارتباط موبایل مرتبط با گزارش‌ها و حوادث جرایم سازمان جنایتکار، ایجاد می‌کند. سپس یک درخت حداقل در شبکه ایجاد می‌کند. این درخت، با تعیین رأس‌های مهم در شبکه اعضای تاثیرگذار را شناسایی می‌کند (تاها، ۲۰۱۷، ص ۴۴۵-۴۵۳).

٥- سیستم بصری برای جرم‌کاوی در فضاهای باز سراسر دانشکده: آیدین پاتریک شیا[۵] درخصوص موضوع ایمنی در دانشکده‌ها پژوهش کردند. وی دو هدف

۱.Singhal
۲. Kamal Taha
۳. ECLfinder
٤. influential
٥. Aidan Patrick Shea

عمده را در پژوهش‌های خود دنبال کردند. جمع‌آوری داده‌ها به‌صورت بی‌درنگ از پایگاه‌های دانشکده‌ها انجام شد؛ سپس، یک ابزار بصری که جنایات محوطه دانشگاه و اطراف آن را نشان دهد، ایجاد شد. آنان نقشه متحرک ماهواره‌ای ارائه کرده‌اند که نشان‌دهنده تکامل چالش جرم در محوطه دانشگاه و اطراف آن است (شیا، ۲۰۱۷، صص ۱-۳).

٦- تشخیص اطلاعات مشکوک در وب با استفاده از فنون داده‌کاوی جرایم: حسین‌خانی و همکاران، درخصوص استخراج اطلاعات مفید با استفاده از داده‌کاوی برای پیداکردن نقاط جرم‌خیز و پیش‌بینی روند جرم و جنایت پژوهش کردند. آنان با این فرض که یک صفحه وب مجرمانه می‌تواند به‌عنوان زنجیره‌ای از اقدامات با مجموعه‌ای از ویژگی‌های پنهان مورد توجه باشد، اطلاعات وب را از منظر حوادث تجزیه و تحلیل کردند. بیشتر مستندهای دیجیتالی از داده‌های متنی مانند رایانامه، صفحات وب و سیاهه‌های گفتگوی برخط جمع‌آوری شد. پژوهشگر از ابزارهای جستجو برای کشف و استخراج اطلاعات مفید از بین مدرک‌ها و مستندها استفاده کرد (حسین‌خانی و همکاران، ۲۰۱۴، ص ۳۲-۴۱).

۷- تحقیقات انجام‌شده در حوزه داده‌کاوی در نیروی انتظامی: در نیروی انتظامی نیز پژوهش‌هایی درخصوص داده‌کاوی در جرایم انجام شده است. به‌عنوان مثال کیوان‌پور و همکاران در پژوهش خود، عمل تطابق جرم را بر روی متغیرهای جرم غیرفضایی یعنی متغیرهای رفتاری جرم انجام دادند. برای مثال، در نرم‌افزار پلیس‌یار، متغیرهای جرم سرقت از منازل در چهار گروه نوع محل مورد سرقت، نحوه تعامل‌های مجرم با محیط سرقت، شیوه ورود و ابزار مورد استفاده مجرم دسته‌بندی شده‌اند (احمدوند و آخوندزاده، ۱۳۸۹، صص ۱۱-۲۲ و کیوان‌پور و همکاران ۱۳۸۸، صص ۹۸-۱۱۷).

پولادی، درخصوص کشف ارتباطات میان جرایم و مولفه‌های مختلفی مانند شغل، محل خدمت، تحصیلات، درجه و جایگاه پژوهش کرد. او با استفاده از فنون داده‌کاوی، الگوهای جرم را استخراج کرد تا از طریق بررسی علل و انگیزه انجام جرایم، بتواند راهکارهای پیشگیرانه را ارائه کند. پولادی در پژوهش خود از مدل‌های کلاس‌بندی درخت تصمیم C۵ ، QUEST ،CART ، CHAID و شبکه‌های عصبی استفاده کرد و با مقایسه الگوریتم‌های یادشده توسط نرم‌افزار داده‌کاوی، الگوریتم C۵ را به‌عنوان بهترین الگوریتم انتخاب کردند. وی در پژوهش خود قوانین منجر به وقوع جرایم کارکنان را ارائه کرد (پولادی، ۱۳۹۶).

شاه‌محمدی و عباسی در پژوهشی باعنوان بررسی عوامل موثر در تصادفات درون شهری با استفاده از روش‌های داده‌کاوی، عوامل موثر در تصادفات درون‌شهری شهر اصفهان را بررسی کرده‌اند. برای تحلیل اطلاعات از کشف قوانین انجمنی تکنیک سبد خرید و تولید درخت اشیاء مکرر FP-Growth استفاده کردند. نتایج نشان داد ویژگی های موثر در بروز تصادف به‌ترتیب ۱) عجله و شتاب، ۲) استفاده‌نکردن از کمربند و کلاه ایمنی، ۳) بی‌توجهی به جلو، ۴) نبود شانه راه، ۵) تجهیزنبودن خودرو به تجهیزات ایمنی هستند (شاه‌محمدی و عباسی، ۱۳۹۷، ص ۱۳۵-۱۶۲).

مبانی نظری

جرایم سایبری. واژه سایبر از یک لغت یونانی[۱] به‌معنی سکاندار[۲] یا راهنما مشتق شده است. در طی توسعه شبکه‌های رایانه به‌خصوص اینترنت، واژه‌های ترکیبی بسیاری از کلمه سایبر به‌وجود آمده است مانند: فضای سایبر[۳]، شهروند سایبر[۴]، پول سایبر[۱]،

۱.Kybernetes

۲.Sculls

۳.Cyberspace

۴.Cybercitizen

فرهنگ سایبر[۲]، تجارت سایبر[۳]، جرم سایبر[۴] و... . این واژه نخستین بار توسط ویلیام گیبسون نویسندهٔ داستان‌های علمی - تخیلی در کتاب نورومنسر[۵] بکار برده شد (شاه‌محمدی و تاهو، ۱۳۹۳، ص ۱۰۳).

جرایم سایبری یکی از پیش‌رونده‌ترین جرائمی است که با سرعت زیاد در حال پیشرفت است. این در حالی است که همگام با پیشرفت‌های علمی به‌ویژه در زمینه رایانه و اینترنت، عده‌ای برخلاف خدمت‌گزاران بشریت که به فکر استفاده‌های مثبت از فناوری‌ها هستند به فکر سوءاستفاده‌اند. فضای سایبر با ویژگی‌های خاص آن از جمله ارتباط سریع و آسان بین افراد و دسترسی به‌عنوان منبع اطلاعات باعث پیشرفت‌های بزرگ در روابط اقتصادی، اجتماعی، سیاسی و فرهنگی حاکم بر افراد شده است. همانند دنیای فیزیکی و مادی، در این فضا نیز افرادی یافت می‌شوند که بنا به نیات و انگیزه‌ها و اهداف خاص سعی دربرهم‌زدن نظم این اجتماع مجازی دارند. فضای مجازی به‌دلیل امنیت ناکافی و طبیعت مجازی، فرصت مناسبی را در اختیار افراد مجرم قرار می‌دهد. جرائم در این فضا نیز گاهی متفاوت و خاص فضای مجازی هستند. در فضای سایبر کشف جرم و پیگیری مجرم غالبا پیچیده‌تر از جرائم فضای فیزیکی است. نگرانی عمده در این فضای مجازی در رابطه با انتشار سریع اطلاعات، تخریب اطلاعات و سوء استفاده از آن است. انگیزه‌های جدید مجرمان باعث تنوع و بروز جرائم متنوعی در این فضا شده است و هر روز باید منتظر ظهور جرمی نو با شیوه‌ای جدید در این دنیای مجازی باشیم (بوربور، ۱۳۹۳، ص ۸-۱۳).

---

۱. Cyber cash

۲.Cyber culture

۳.Cyber business

۴.Cybercrime

۵.Neuromancer

داده‌کاوی: داده‌کاوی شامل استفاده از ابزارهای پیشرفته تحلیل داده به‌منظور کشف الگوهای معتبر، از قبل ناشناخته و روابط موجود در مجموعه داده‌های بزرگ است (اسماعیلی، ۱۳۹۴، ص ۱۵). داده‌کاوی ترکیبی از شیوه‌های یادگیری ماشین، تشخیص الگو، آمار، نظریه پایگاه داده و تلخیص و ارتباط بین مفاهیم و الگوهای جالب به‌صورت خودکار از پایگاه داده سازمان‌ها و شرکت‌های بزرگ است. هدف اصلی داده‌کاوی کمک به فرآیند تصمیم‌گیری از طریق استخراج دانش از داده‌ها است (الپایدین[۱]، ۲۰۱۰). داده‌کاوی فرایند کشف قوانین و دانش ناشناخته و مفید از انبوه داده‌ها و پایگاه داده است. انجام عملیات داده‌کاوی شامل (لین و همکاران[۲]، ۲۰۱۲): ۱) جداسازی داده مفید از داده بیگانه، ۲) یکپارچه‌سازی داده‌های مختلف تحت یک قالب واحد، ۳) انتخاب داده لازم از میان دیگر داده‌ها، ۴) انتقال داده به محیط داده‌کاوی برای کشف قوانین، ۵) ایجاد مدل‌ها و الگوهای مرتبط بوسیله روش‌های داده‌کاوی، ۶) ارزیابی مدل و الگوهای ایجادشده برای تشخیص مفیدبودن آنها و ۷) انتشار دانش استخراج‌شده به کاربران نهایی است.

روش‌شناسی‌های بسیاری برای پیاده‌سازی داده‌کاوی، مطرح است. از آنجا که روش‌شناسی کریسپ[۳] همواره به‌علت داشتن رویه واحد در کل طرح‌های داده‌کاوی و برخورداری از جامعیت و مقبولیت از سوی متخصصان داده‌کاوی یکی از روش‌های بسیار قوی در حوزه داده‌کاوی است و همچنین سازگاری آن با رویکرد پژوهش فعلی، از روند کلی مدل استاندارد کریسپ و گام‌های آن در اجرای طرح حاضر، استفاده می‌شود. کریسپ یک روش استاندارد داده‌کاوی است که در اواخر سال ۱۹۹۶، سه شرکت بزرگ «دایملر کرایسلر (بنز)»، «اس.پی.اس.اس» و «ان.سی.آر. آن» را ایجاد

۱.Alpaydin
۲.Lin et al.
۳.CRISP

کرده‌اند (چپمن[۱] و همکاران، ۲۰۰۰). این روش، یک مدل فرایندی برای داده‌کاوی ارائه می‌دهد که مروری بر چرخهٔ حیات هر طرح داده‌کاوی است. چرخهٔ حیات یک طرح داده‌کاوی، شامل شش مرحله است: گام‌های شناخت سیستم، شناخت داده‌ها، آماده‌سازی داده‌ها، مدل‌سازی، ارزیابی و توسعه مدل. هرکدام از این گام‌ها به زیربخش‌هایی تقسیم می‌شود (احمدی و وحیدبسط، ۱۳۸۹، ص ۱۶).

الف- شناخت کسب‌وکار: در گام اول، ابتدا به شناخت کسب‌وکار مورد نظر پرداخته می‌شود. سپس اهداف و عوامل موفقیت کلیدی آن تعیین شده و دوباره اهداف کسب‌وکار بازنگری می‌شود.

ب- شناخت داده‌ها: عبارت است از جمع‌آوری داده‌های اولیه، توصیف داده‌ها، بازرسی و بررسی داده‌ها و اعتبارسنجی کیفیت داده‌ها. کارایی داده‌کاوی به‌طور مستقیم با داده‌های مورد استفاده مرتبط است.

ج- آماده‌سازی داده‌ها: گام آماده‌سازی داده‌ها عبارت است از انتخاب داده‌ها، پاک‌سازی داده‌ها، آماده‌سازی داده‌ها جهت داده‌کاوی، مجتمع کردن آنها و قالب‌بندی داده‌ها و جمع‌آوری و محافظت از داده‌ها، گام بسیار مهمی است.

د- مدل‌سازی: در اولین قدم از مدل‌سازی باید روش مناسب را انتخاب کرد. انتخاب روش مناسب بسیار تعیین‌کننده است. با توجه به گستردگی و تنوع روش‌های مدل‌سازی و وجود الگوریتم‌های بسیار زیاد در هر روش مدل‌سازی، در این بخش با توجه به اهداف طرح، کیفیت و کمیت داده‌های در دسترس و قابل استفاده و همچنین اولویت‌های کسب‌وکار، سعی می‌شود مدل مربوطه انتخاب شود (شیرر[۲]، ۲۰۰۵، ص ۱۷).

---

۱. Chapman
۲. Shearer

ه- ارزیابی: نتایج ارزیابی باعث بهبود مدل شده و مدل را قابل استفاده می‌کند. مشاهدات و استنتاج‌های مجری می‌تواند به‌عنوان یک بخش از ارزیابی مورد استفاده قرار می‌گیرد (اسکندری و همکاران، ۱۳۸۸، ص ۵۰).

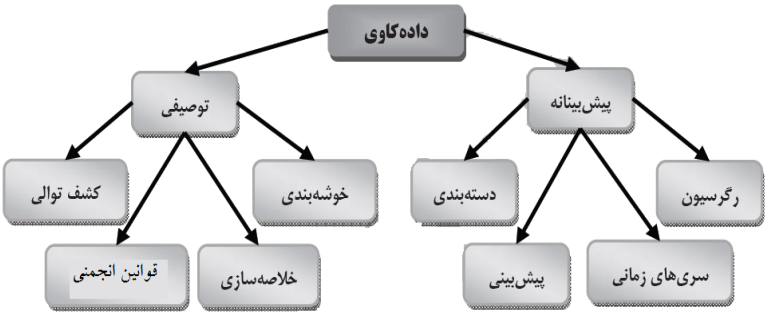
و- بکارگیری و پیاده‌سازی مدل: در این گام این طرح دوباره بررسی شده و مدون می‌شود. علاوه بر آن، نظارت و نگهداری پس از اتمام طرح نیز در این گام تهیه می‌شود (شیرر، ۲۰۰۵، ص ۲۲).

روش‌های داده‌کاوی: روش‌های داده‌کاوی مطابق شکل زیر به دو دسته کلی روش‌های پیش‌بینانه و روش‌های توصیفی تقسیم می‌شوند که با توجه به استفاده از روش‌های پیش‌بینانه در این پژوهش، به‌طور اجمالی این روش معرفی می‌شود.

داده‌کاوی پیش‌بینانه: این روش با استفاده از مجموعه داده‌ها، مدل‌هایی را برای توضیح سیستم تولید می‌کند که با استفاده از آنها می‌توان عملکرد متغیرهای مختلف را پیش‌بینی کرد. این روش دارای فرایند دو مرحله‌ای است. در مرحله اول یک مدل ساخته می‌شود که مجموعه‌ای از کلاس‌های داده‌ای یا مفاهیم را مشخص می‌کند. این مرحله را یادگیری[۱] می‌نامند. در این مرحله الگوریتم کلاس‌بندی، یک مدل را با تحلیل یک مجموعه آموزشی[۲] که شامل مجموعه‌ای از سطرهای پایگاه داده است، می‌سازد و برچسب کلاس‌های مربوط به این سطرها را مشخص می‌کند. مجموعه آموزشی به‌صورت تصادفی از پایگاه داده انتخاب می‌شود. از آنجایی که برچسب هر سطر آموزشی در این مرحله مشخص است، این مرحله یادگیری نظارت‌شده[۳] نامیده می‌شود. در مرحله دوم، از مدل ساخته‌شده برای کلاس‌بندی استفاده می‌شود. یادگیری می‌تواند برچسب کلاس هر سطر را پیش‌بینی کند. همچنین در این مرحله میزان دقت

۱.Learning
۲.Training Set
۳.Supervised Learning

کلاس‌بندی کننده تخمین زده می‌شود. برخی از الگوریتم‌های رایج در این فن شامل C۵، K نزدیکترین همسایه[۱]، شبکه عصبی و بیزین است که در ادامه بحث می‌شود (اسماعیلی، ۱۳۹۳، ص ۱۵).



شکل ۱. فنون و روش‌های یادگیری مدل در داده‌کاوی (دونهم، ۲۰۰۲)

از آنجا که مقادیر هدف پژوهش حاضر گسسته است، بنابراین در این پژوهش از فنون دسته‌بندی برای داده‌کاوی شامل ۱) شبکه بیزی؛ ۲) نزدیکترین همسایه؛ ۳) شبکه عصبی و ۴) درخت تصمیم استفاده شد.

درخت تصمیم: درخت تصمیم یکی از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی است و پیش‌بینی خود را در قالب یکسری قوانین اگر- آنگاه توضیح می‌دهد. روش کار در درخت تصمیم به این صورت است که یک گره ریشه در بالای آن کشیده شده و برگ‌های آن در پایین هستند. یک رکورد در گره ریشه وارد می‌شود و در این گره یک آزمون صورت می‌گیرد تا معلوم شود که این رکورد به کدامیک از گره‌های فرزند (شاخه پایین‌تر) می‌رود. معمولاً روش‌های مختلفی برای انتخاب این آزمون اولیه وجود دارد ولی هدف همه آنها یکی است. باید روشی را انتخاب کرد که

۱.KNN

بهترین جداسازی را در کلاس‌های هدف انجام دهد. این فرآیند آن قدر ادامه پیدا می‌کند تا رکورد به گره برگ برسد. تمام رکوردهایی که به یک برگ از درخت می‌رسند در یک کلاس قرار می‌گیرند. همچنین برای رسیدن از ریشه به یک برگ تنها یک راه وجود دارد و آن راه در واقع بیان قانونی است که برای دسته‌بندی رکوردها ایجاد شده است. ممکن است تعداد زیادی برگ وجود داشته باشد که همگی یک کلاس داشته باشند ولی هر برگ برای قرار گرفتن در دسته مورد نظر علت متفاوتی دارد. برای ایجاد درخت تصمیم الگوریتم‌های متفاوتی وجود دارند که معروف‌ترین آنها عبارت‌اند از: C۵، CART، QUEST و CHAID. شایان ذکر است که تمامی این روش‌ها برای دسته‌بندی، ساختار تقریباً مشابهی دارند و هدف همه آنها به‌دست آوردن درختی با کیفیت بالا و نرخ خطای کم برای دسته‌بندی داده‌ها است و بیشتر تفاوت‌ها در شیوه شاخه زدن و برش شاخه‌ها است.

ابزارهای داده‌کاوی: پس از انتخاب روش داده‌کاوی، باید در مورد ابزار داده‌کاوی تصمیم‌گیری شود. ابزارهای مختلفی برای داده‌کاوی معرفی شده است. در این پژوهش از IBM SPSS MODELER که نسخه کامل شده نرم‌افزار داده‌کاوی کلمنتاین است، استفاده شد. کلمنتاین امکان بهبود تصمیم‌گیری در عملیات کسب‌وکار از طریق توسعه سریع مدل‌های پیشگویانه را می‌دهد. این ابزار که برای مدل استاندارد-صنعتی کریسپ طراحی شده است، از کل فرآیند داده‌کاوی، از دریافت داده‌ها تا ارائه نتایج کسب‌وکار پشتیبانی می‌کند. کلمنتاین مجموعه متنوعی از روش‌های مدل‌سازی را ارائه می‌دهد. روش‌های قابل دسترس در این نرم‌افزار امکان استخراج اطلاعات جدید از داده‌های موجود را می‌دهد.[۱]

روش پژوهش

۱.http://www.Spss.com/clementine/CLEMENTINE ® ۱۲,۰ APPLICATIONS GUIDE

این پژوهش از نظر هدف کاربردی و از نظر نوع تحلیل محتوا است که با رویکرد داده‌کاوی انجام می‌شود. جامعه آماری پژوهش، جرایم سایبری است. حجم نمونه، جرایم اخلاقی فضای سایبر در سال‌های ۱۳۸۹ تا ۱۳۹۴ است. در ادامه داده‌های جمع‌آوری شده در بانک اطلاعاتی پلیس فتا نیروی انتظامی در حوزه جرایم اخلاقی، از سال‌های ۱۳۸۹ تا ۱۳۹۵ تحلیل و الگوهای وقوع جرم شناسایی شد. جهت داده‌کاوی جرایم اخلاقی فضای سایبر، ابتدا باید داده‌ها پیش پردازش شده و به شکل‌های مناسب برای استفاده در نرم‌افزارهای داده‌کاوی درآمد. در مراحل پیش پردازش داده‌ها، داده‌های دارای خطاها و داده‌های گم شده باید به‌نحو مناسب پاک‌سازی و اصلاح شد. فرایند پاک‌سازی داده، شامل تکمیل مقادیر مفقود، هموارسازی داده‌های مغشوش، شناسایی و حذف نقاط دورافتاده و برطرف کردن تناقض‌های موجود بین داده‌ها است. این پاک‌سازی، با حذف کل رکورد، حذف داده متناقض یا جایگزینی مقداری به‌جای این داده، امکان‌پذیر است. در مرحله بعد براساس فنون داده‌کاوی و درک و پویش داده‌ها ، الگوریتم و نرم‌افزار مناسب انتخاب و داده‌ها تجزیه و تحلیل شده و تحلیل‌های اولیه به‌دست آمد. درنهایت از بین قواعد به‌دست آمده، ارزیابی الگوریتم‌های بکار رفته، نتایج بدست آمده تفسیر شد. برخی از انجام مراحل آماده‌سازی داده‌ها در نرم‌افزار اکسل انجام شد.

فرایند پژوهش: با توجه به اینکه این پژوهش براساس روش‌شناسی کریسپ انجام شده است؛ گام‌های پژوهش حاضر براساس این روش‌شناسی انجام شد.

الف- استخراج داده‌های جرایم از سیستم جامع پلیس فتا. داده‌ها در این بانک اطلاعات در چهار دسته اطلاعات کلی، فردی، ادله دیجیتال و پرونده ثبت می‌شود که شامل فیلدهای گوناگون است ولی با توجه به هدف طرح، سن، جنسیت، مدرک تحصیلی، وضعیت تاهل، شغل مجرم و قربانی و سال وقوع جرم برای تحلیل بهتر جرایم

اخلاقی فضای سایبر با استفاده از داده‌کاوی از بین ویژگی‌های ثبت شده در پایگاه داده انتخاب شد.

ب- آمایش داده‌ها: داده‌ها در نرم‌افزار اکسل جمع‌آوری و با استفاده از این نرم‌افزار، داده‌ها محدودسازی و رکوردهای خاص بررسی شد تا قبل از بکارگیری داده‌ها در نرم‌افزار داده‌کاوی IBM SPSS MODELER، داده‌ها آماده‌سازی شوند؛ به‌طوری‌که بخش اعظمی از زمان این پژوهش صرف آماده‌سازی و پالایش داده‌ها شد. پس از آماده‌سازی داده‌ها، با توجه به این‌که داده‌ها در نرم‌افزار اکسل، تهیه و آماده‌سازی شد، داده‌ها وارد نرم‌افزار داده‌کاوی شدند. قبل از مدل‌سازی داده‌ها، برای درک بهتر از داده‌ها، نمایش گرافیکی هریک از ویژگی‌ها ارائه شد که در ادامه بیان می‌شود.

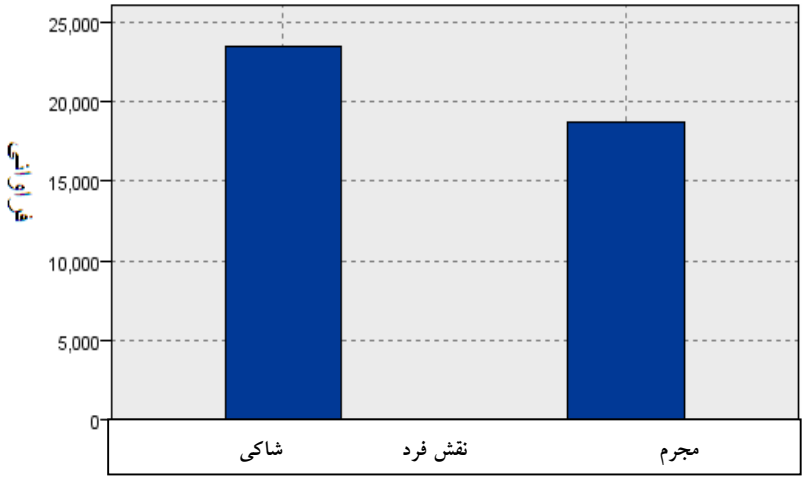
ج- انتخاب مدل و مدل‌سازی: در این پژوهش، با توجه به اینکه متغیر هدف پژوهش (موضوع جرم)، گسسته است؛ فنون و مدل‌های بکار رفته در این پژوهش، مدل‌های دسته‌بندی (روش با ناظر) است. در ادامه داده‌های پژوهش براساس الگوریتم‌های مدل‌های با ناظر، مدل‌سازی و بهترین مدل براساس دقت آن انتخاب می‌شود.

د- انجام داده‌کاوی: پس از اجرای مدل، عملیات داده‌کاوی بر روی داده‌ها انجام شده و نتیجه در محیط نرم‌افزار نشان داده می‌شود. با کلیک بر روی نتیجه داده‌کاوی، می‌توان فیلدهای تاثیرگذار در عملیات داده‌کاوی را به همراه قوانین خروجی استخراج کرد که در یافته‌های پژوهش ارائه می‌شود.

یافته‌های پژوهش

در این بخش یافته‌های حاصل از اجرای فرایند پژوهش در پنج بند ارائه می‌شود.

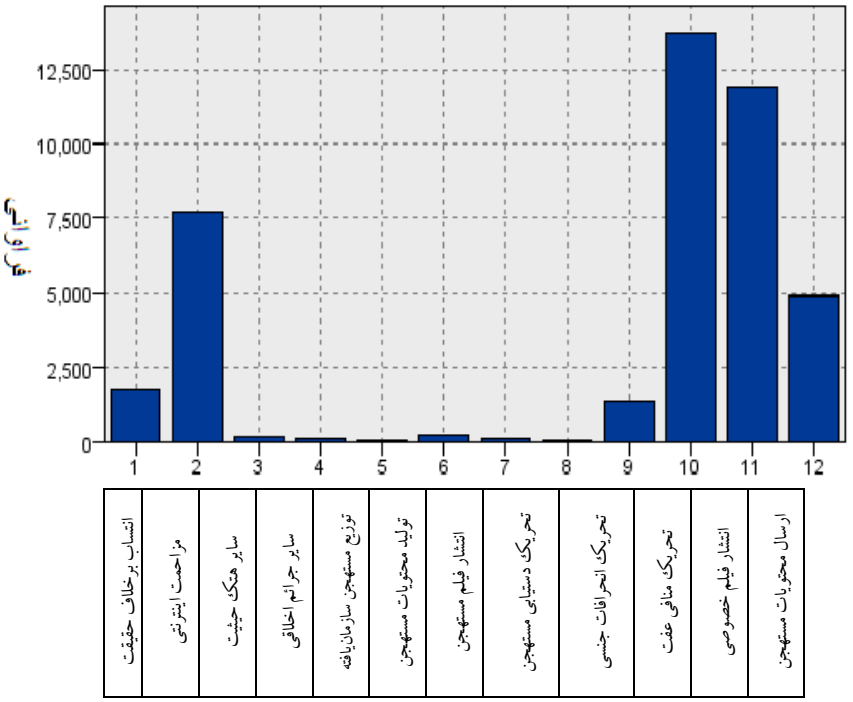
۱-نمایش گرافیکی ویژگی‌های داده‌های جرایم اخلاقی فضای سایبر. نمایش گرافیکی ویژگی‌های حاوی اطلاعات مفید و قابل بررسی است که به بخشی از این نمودارها در ادامه اشاره می‌شود. مطابق نمودار ۱، فراوانی نقش شاکی‌ها در بانک اطلاعاتی جرایم اخلاقی فضای سایبر بالاتر است که این امر بیانگر یک به چند بودن نسبت مجرمان به شاکیان در برخی از پرونده‌های جرایم اخلاقی است.



نمودار ۱. فراوانی نقش فرد

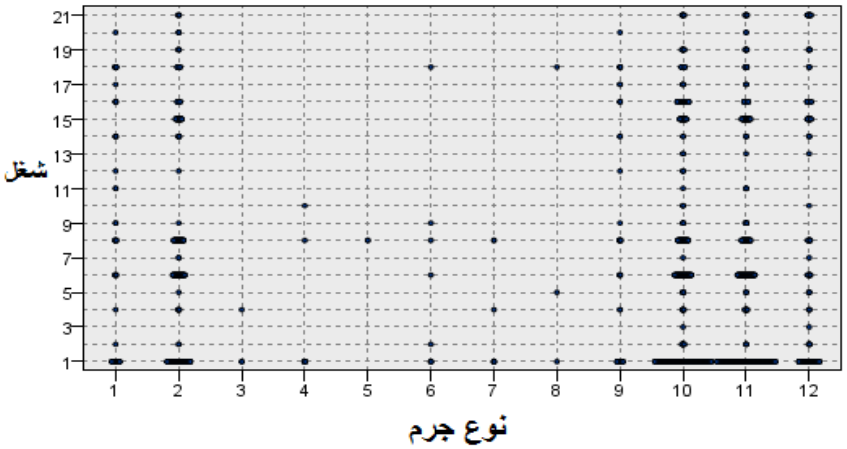
مطابق نمودار ۲، جرایم اخلاقی کد ۱۰ و ۱۱ (سایر هتک حیثیت و مزاحمت اینترنتی) دارای بالاترین فراوانی و کد ۵ و ۸ (تحریک دستیابی مستهجن و توزیع مستهجن سازمان‌یافته) دارای کمترین فراوانی هستند. همان‌طورکه در بخش‌های قبلی مطرح شد، در برخی از روش‌های داده‌کاوی، می‌توان داده‌های با فراوانی بسیار کم نسبت به کل داده‌ها (داده‌های نویز) را حذف کرد، ولی در این طرح به‌دلیل حساسیت نوع جرم، باید تمام جرایم تحلیل و ریشه‌یابی شود. به‌عنوان مثال جرمی مانند توزیع مستهجن

سازمان‌یافته، گرچه دارای فراوانی بسیار کم است ولی از ارزش بسیار بالـایی برخـوردار است؛ بنابراین از حذف این رکورد خودداری می‌شود.



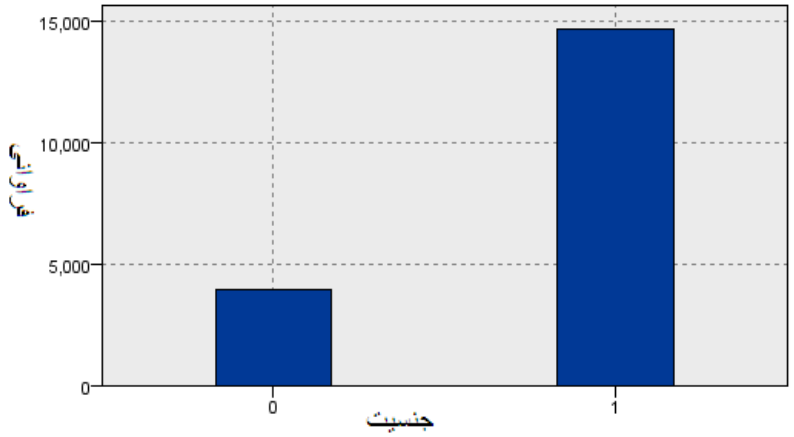
نمودار ۲. فراوانی موضوع‌های جرم

مطابق نمـودار ۳، افـراد بـا مشـاغل کـد ۱ و ۸ (آزاد، دانشـجو) در جـرایم کـد ۱۱ و ۱۲ (مزاحمت اینترنتی، انتساب برخلاف حقیقت) بیشترین فراوانی را دارند.



نمودار ۳. رابطه بین شغل فرد و موضوع جرم

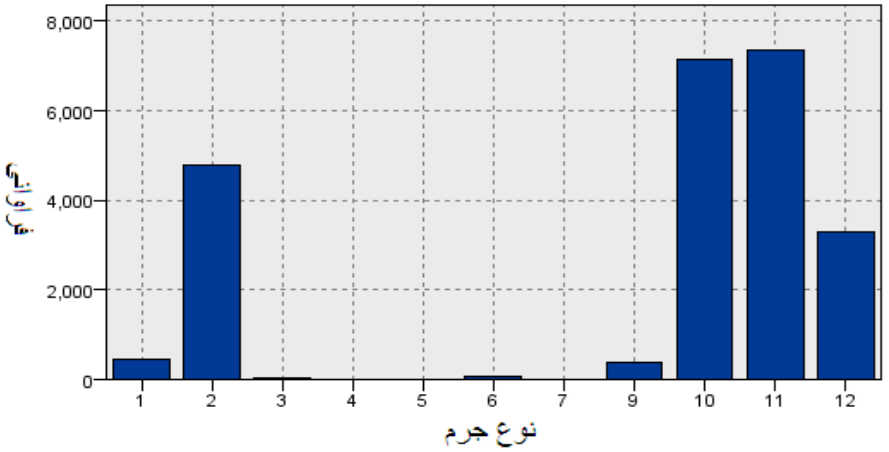
مطابق نمودار ۴، فراوانی مجرمان در مردان بیش از سه برابر زنان است.



نمودار ٤. فراوانی مجرمان جرایم سایبری براساس جنسیت

مطابق نمودار ۵، قربانیان جرایم کد ۱۰، ۱۱، ۲ و ۱۲ (انتشار فیلم خصوصی، سایر هتک حیثیت، مزاحمت اینترنتی، انتساب برخلاف حقیقت) دارای بالاترین فراوانی هستند. از طرفی در فراوانی موضوعات جرم در بین قربانیان، جرایم کد ۴، ۵، ۷ و ۸ (تحریک به

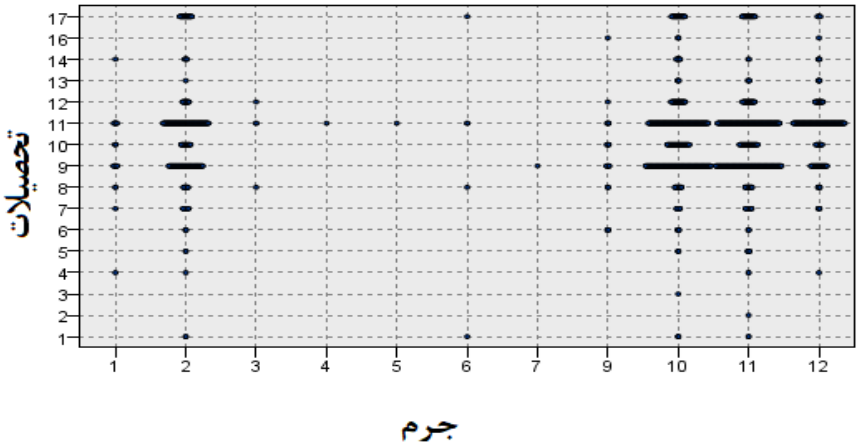
انحراف‌های جنسی و آموزش آن، تحریک به دستیابی به محتواهای مستهجن و آموزش آن، تولید محتواهای مستهجن به قصد معامله یا تجارت، تولید و توزیع محتواهای مستهجن به‌صورت سازمان‌یافته) دارای فراوانی بسیار پایین است؛ این مطلب بیانگر این است که در این موضوعات، کشف جرم به‌صورت پیش‌دستانه توسط پلیس با رصد فضای مجازی یا گزارش های بدون شکایت شاکی حقیقی، صورت می گیرد.



نمودار ۵. فراوانی موضوعات جرم در بین قربانیان جرایم سایبری

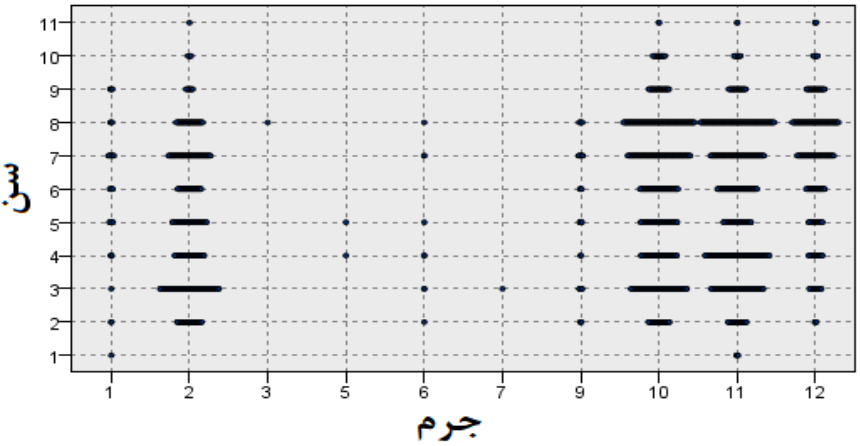
نکته حائز اهمیت در تحلیل نمودارهای حاضر این‌که، نمودارهای ارائه شده در این پژوهش را می‌توان از جنبه‌های مختلف بررسی و تحلیل کرد ولی به‌دلیل طولانی‌شدن مطالب، تمامی جوانب امر بررسی نشده و می‌توان از آن در پژوهش‌های آتی برای ریشه‌یابی جرایم به‌صورت مجزا استفاده کرد.

مطابق نمودار ۶، افراد با تحصیلات کارشناسی در کد جرم ۲، ۱۰، ۱۱ و ۱۲ (انتشار فیلم خصوصی و خانوادگی، سایر هتک حیثیت و نشر اکاذیب، مزاحمت اینترنتی، انتساب اعمالی برخلاف حقیقت) دارای بیشترین فراوانی (قربانی) هستند.



نمودار ٦. رابطه بین تحصیلات قربانیان و موضوع شکایت

مطابق نمودار ۷، افراد با سن ۴۱-۵۰ سال در کد جرم ۱۰ و ۱۱ (سایر هتک حیثیت و نشر اکاذیب، مزاحمت اینترنتی) و افراد با سن ۲۱-۲۵ سال در کد جرم ۲ (انتشار فیلم خصوصی و خانوادگی) دارای بیشترین فراوانی (قربانی) هستند.



نمودار ۷. رابطه بین سن قربانیان و موضوع شکایت

۲-ارزیابی مدل: در اجرای مدل بر روی داده‌های جرم، دو جدول ارائه می‌شود. جدول اول، جدول دقت مدل است که در آن میزان درستی در داده‌های آزمون، بیانگر دقت مدل و میزان اتکا به نتایج آن است. جدول دوم، جدول آشفتگی مدل است. جدول آشفتگی بیانگر این است که مدل در پیش‌بینی درست انواع متغیر هدف، چقدر موفق بوده است. درنهایت مدلی که دارای دقت بالا و ماتریس آشفتگی مناسب (پیش‌بینی دقیق تمام موضوعات جرایم اخلاقی فضای سایبر) باشد؛ قابل اتکا است.

در این پژوهش، درخت‌های تصمیم C۵، CHAID، CART، شبکه عصبی و نزدیکترین همسایه بر روی داده‌های جرایم اخلاقی سایبری اجرا شد که نتایج حاصل نشان داد درخت تصمیم C۵ بهترین دقت را در ارزیابی مدل دارا است. به‌دلیل محدودیت فضا، جداول مربوط به اجرای سایر روش‌ها ارائه نشده است.

مطابق جدول زیر، دقت درخت تصمیم C۵ (که مربوط به داده‌های آموزش است) ۶۲/۱۹ درصد است که دقت نسبتاً خوبی است. درخت یادشده ۳۲۰۱ کلاس‌بندی اشتباه برای داده‌های آزمون و ۱۲۵۸۳ کلاس‌بندی اشتباه برای داده‌های آموزش داشته است.

جدول ۱. دقت درخت *C۵*

بخش	آزمون		آموزش	
درست	۵۲۶۴	۶۲/۱۹ درصد	۲۱۱۵۷	۶۲/۷۱ درصد
نادرست	۳۲۰۱	۳۷/۸۱ درصد	۱۲۵۸۳	۳۷/۲۹ درصد
جمع	۸۴۶۵	۱۰۰ درصد	۳۳۷۴	۱۰۰ درصد

ماتریس آشفتگی درخت تصمیم مدل C۵ نشان می‌دهد که این درخت تمامی جرایم را پیش‌بینی کرد ولی در برخی از جرایم پیش‌بینی با دقت کمی همراه است. بنابراین بهترین مدل برای پیش‌بینی جرایم سایبری درخت تصمیم مدل C۵ است ولی برای اطمینان بیشتر به قوانین خروجی، همچنان دقت این مدل نیز پایین است و باید روشی را ارائه کرد که این مدل بتواند با دقت بالاتری داده‌ها را پیش‌بینی کرد.

۳- روش افزایش دقت مدل در داده‌کاوی داده‌های پیچیده: پس از اجرای داده‌کاوی در مرحله بعدی از فرایند کریسپ، باید مدل ارزیابی شود. مطابق نتایج ارزیابی مرحله قبل از آنجا که دقت درخت تصمیم C۵ پایین بود (۶۳ درصد)، درنتیجه برای افزایش دقت داده‌کاوی، دو مرحله تحلیل داده‌ها به‌شرح زیر انجام شد:

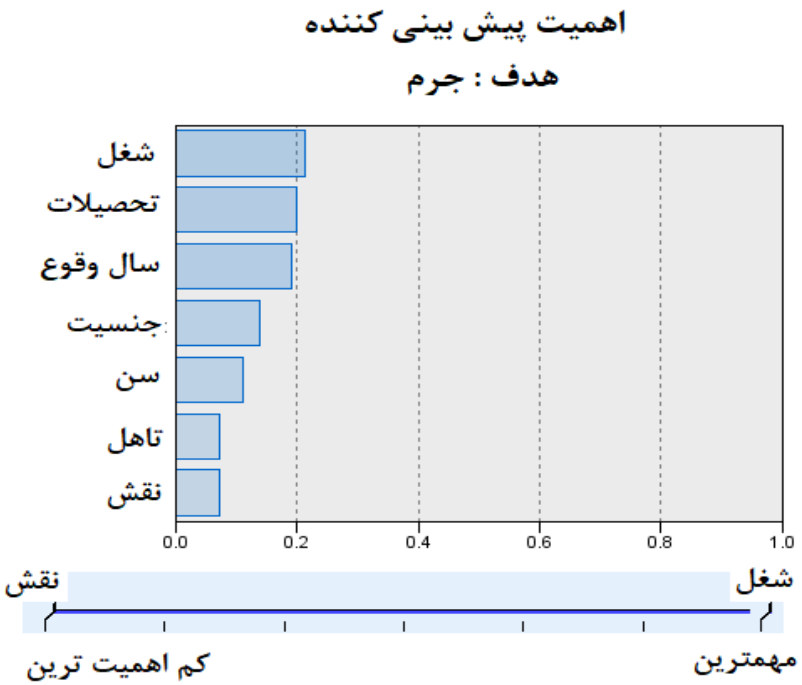
- درخت تصمیم C۵ به‌طور جداگانه بر روی داده‌های مجرم و قربانی اجرا و دقت حاصل بررسی و نتایج، بهینه نبود.
- در این مرحله، برای افزایش دقت مدل، داده‌کاوی بر روی انواع جرایم اخلاقی فضای سایبر به‌صورت جداگانه و متوازن اجرا شد که نتایج نشان داد دقت به مقدار قابل توجهی افزایش می‌یابد. به‌طور نمونه دقت برای داده‌های جرم ارسال محتواهای مستهجن و مبتذل مطابق جدول ۲، به مقدار ۸۷/۴۸ درصد افزایش یافت.

جدول ۲. دقت درخت C۵ بر روی داده‌های جرم ارسال محتواهای مستهجن و مبتذل

بخش	آزمون		آموزش	
درست	۶۹۲	۸۷/۴۸ درصد	۲۶۸۶	۸۸/۶۵ درصد
نادرست	۹۹	۱۲/۵۲ درصد	۳۴۴	۱۱/۳۵ درصد
جمع	۷۹۱	۱۰۰ درصد	۳۰۳۰	۱۰۰ درصد

٤-رتبه‌بندی مشخصه‌های ورودی تاثیرگذار بر نوع جرایم اخلاقی فضای سایبر: یکی دیگر از نتایج مهم حاصل از اجرای مدل بر روی داده‌ها، رتبه‌بندی مشخصه‌های فردی تاثیرگذار بر روی جرایم اخلاقی فضای سایبر است. در فرایند داده‌کاوی جرایم اخلاقی فضای سایبر، در اجرای هر مدل مشخصه‌های تاثیرگذار متفاوت بوده است. همان‌طور که در بخش‌های قبلی مشخص شد، مدل درخت تصمیم C۵ بهترین مدل برای داده‌های ثبت شده در جرایم اخلاقی فضای سایبر است و قوانین

آن قابل اطمینان است. بنابراین مشخصه‌های ورودی تاثیرگذار نیز با استفاده از این مدل، رتبه‌بندی می‌شود که در شکل زیر نشان داده است.



شکل ۳. مشخصه‌های تاثیرگذار بر روی داده‌های جرایم اخلاقی فضای سایبر

همان‌طورکه در شکل ۳ مشخص است، مشخصه‌های تاثیرگذار بر روی داده‌های پژوهش حاضر به‌ترتیب شغل، تحصیلات، جنسیت، سن، تاهل و نقش فرد است که بنا به آن، ویژگی شغل، بیشترین تاثیر را در جرایم اخلاقی فضای سایبر دارد.

۵- قوانین استخراجی از اجرای مدل: با استفاده از این قوانین می‌توان پیش‌بینی کرد که چه افرادی با چه ویژگی‌هایی مرتکب چه جرایمی می‌شوند یا در کدام جرم اخلاقی سایبری، قربانی می‌شوند و از این طریق، راهکارهای پیشگیرانه ازجمله آگاه‌سازی را طرح کرد.

بحث و نتیجه گیری

با توجه به رشد فناوری و افزایش استفاده از فضای سایبر و به‌دنبال آن رشد جرایم در این حوزه، انجام پژوهش‌های مفید و موثر در راستای پیشگیری از جرایم سایبری، ضروری است. مسئله تحلیل جرایم اخلاقی فضای سایبر و بررسی ریشه‌ها و علل وقوع این جرایم در این پژوهش، مطرح شد. هدف این پژوهش، تحلیل جرایم اخلاقی فضای سایبر بود تا از این طریق بتوان راهکارهای پیشگیرانه ارائه داد. در این پژوهش تلاش شد از طریق تحلیل جرایم و استخراج الگوهای جرم، مشخص شود چه افراد و با چه ویژگی‌هایی مرتکب چه نوع جرایم اخلاقی فضای سایبر (به‌ویژه در حوزه اخلاقی) می‌شوند و چه افرادی با چه ویژگی‌هایی، قربانی جرایم اخلاقی فضای سایبر می‌شوند.

برای تحلیل جرایم در این پژوهش از رویکرد داده‌کاوی که یکی از روش‌های قدرتمند در تحلیل داده‌های جرم است، استفاده شد. در ایران با توجه به ایجاد سامانه‌های اطلاعاتی در پلیس و ثبت اطلاعات مجرمان دراین سامانه‌ها، تلاش‌های متعددی به‌منظور تحلیل داده‌های جرایم مبتنی بر رویکرد داده‌کاوی انجام شده است. شاه‌محمدی و عباسی، مولفه‌های تاثیرگذار بر وقوع تخلفات رانندگی را مبتنی بر روش داده‌کاوی استخراج کرده‌اند. پولادی، ارتباط میان جرایم و ویژگی‌های شغلی کارکنان را مبتنی بر روش داده‌کاوی بررسی کرده است. بنا به بررسی‌های انجام شده در نیروی انتظامی تاکنون پژوهشی در مورد رابطه بین ویژگی‌های مجرمان و جرایم اخلاقی فضای سایبر انجام نشده است. بنابراین پژوهش حاضر در نوع خود منحصر بفرد است. در این پژوهش براساس روش‌شناسی کریسپ، شش گام داده‌کاوی بر روی داده‌های جرایم اخلاقی فضای سایبر اجرا شد. قبل از اجرای مدل‌های داده‌کاوی بر روی داده‌های جرایم، برای ارائه دید بهتر و دقیق‌تر درخصوص ویژگی داده‌های جرایم اخلاقی فضای سایبر، ابتدا نمایش گرافیکی از داده‌ها ارائه شد. در مدل‌سازی جرایم سایبری، اغلب

الگوریتم‌های درخت تصمیم، نزدیکترین همسایه، شبکه بیزی و شبکه عصبی بر روی داده‌های پژوهش اجرا شد که نتایج نشان داد الگوریتم C^{5} دارای بالاترین دقت است و به‌عنوان الگوریتم مناسب برای داده‌های جرایم اخلاقی فضای سایبر انتخاب شد.

در ادامه قوانین حاکم بر موضوعات جرایم که جزء پرسش اصلی پژوهش بوده است؛ استخراج شد. این قوانین حاوی یافته‌های بسیار غنی در راستای پیشگیری از جرایم اخلاقی فضای سایبر است. یافته‌های این پژوهش بیان‌گر این است که افراد با مشاغل آزاد، خانه‌دار و دانشجو، بیشترین جرایم اخلاقی فضای سایبر را مرتکب شده‌اند. همچنین افراد با مدارک دیپلم، کاردانی و کارشناسی بیشترین جرایم اخلاقی فضای سایبر را مرتکب شدند. از این دست قوانین، علاوه بر هوشیاری پلیس در مظنون‌یابی و کشف پیش‌دستانه، می‌توان از جنبه‌های آموزشی به خانواده‌ها یا آموزش در مدارس و دانشگاه‌ها استفاده کرد. بررسی اجمالی قوانین مربوط به قربانیان در این پژوهش بیانگر این است که افراد با تحصیلات دیپلم و کمتر، بیشتر قربانی این نوع جرم می‌شوند. این موضوع نشانگر لزوم افزایش سواد رسانه قربانیان جرایم فضای سایبر از طرق مختلف است. از کاربردهای جالب این قوانین در مخاطب‌شناسی جهت برگزاری جلسات آگاه‌سازی از قبیل طرح کوآ است. در این قبیل طرح‌ها، بسیار مهم است که بدانید طیف مورد بحث، قربانی چه نوع جرایمی هستند. با تحلیل صحیح و منطقی این قوانین، پلیس می‌تواند با دقت بیش از ۷۴ درصد (دقت مدل جهت داده‌کاوی جرایم اخلاقی فضای سایبر) پیش‌بینی کند چه افرادی با چه ویژگی‌هایی مستعد ارتکاب یا در معرض قربانی شدن چه نوع جرایم اخلاقی فضای سایبر هستند. از یافته‌های دیگر این پژوهش، ویژگی‌های تاثیرگذار افراد بر جرایم اخلاقی فضای سایبر است که به ترتیب شغل، تحصیلات، جنسیت، سن، تاهل و نقش فرد بوده است که بنا به آن ویژگی شغل بیشترین تاثیر را در این حوزه دارد.

در این پژوهش برای افزایش دقت مدل در داده‌های پیچیده جرایم اخلاقی فضای سایبر چند روش بررسی و اجرا شد؛ روش‌ها و مدل‌های مختلف داده‌کاوی پیش‌بینانه بر روی داده‌های حاضر آزمون شده است که الگوریتم C۵ بهترین روش بوده است؛ ولی دقت مدل بر روی داده‌های پژوهش حدود ۶۳ درصد بود که قابل اتکا و اطمینان نیست. با توجه به اینکه داده‌های این پژوهش مربوط به قربانیان و مجرمان جرایم اخلاقی فضای سایبر است؛ در این مرحله داده‌های قربانیان و مجرمان را از هم تفکیک و مدل C۵ بر روی داده‌های جدید اجرا شد؛ ولی دقت مدل افزایش قابل توجهی نداشته است.

پیشنهادها

۱. ارتقاء طرح‌های آگاه‌سازی مانند طرح کوآ از طریق تحلیل قوانین استخراجی و استفاده از نتایج پژوهش حاضر، در تهیه و ارائه محتوا، امکان‌پذیر است.

۲. از آنجا که برابر نتایج این پژوهش، افراد با مشاغل آزاد، خانه‌دار و دانشجو، بیشترین فراوانی ارتکاب جرایم اخلاقی فضای سایبر را دارند، استفاده از رسانه‌های مختلف از جمله صداوسیما و پایگاه‌ها با محتوای چندرسانه‌ای، امکان ارتقاء فرهنگ استفاده از اینترنت و فضای سایبر و اطلاع‌رسانی عواقب ارتکاب جرم در فضای سایبر، برای این‌گونه افراد، زمینه پیشگیری و کاهش جرایم اخلاقی فضای سایبر را فراهم می‌سازد.

۳. رصد دقیق و نظارت بر تولیدکنندگان محتوا، در جهت‌دهی مثبت به کاربران و کاهش جرایم سایبری تاثیرگذار است.

۴. ویژگی‌های تاثیرگذار بر وقوع جرایم اخلاقی فضای سایبر شامل شغل، تحصیلات، جنسیت، سن، تاهل و نقش فرد بیشترین تاثیر را در وقوع این نوع جرایم دارند. بنابراین برای پیشگیری از وقوع این نوع جرایم، علاوه بر رفع بیکاری، می‌توان با افزایش سطح آگاهی کاربران فضای سایبر، از وقوع این جرایم پیشگیری کرد.

۵. از آنجا که افراد با تحصیلات دیپلم و کمتر، بیشتر قربانی جرم اخلاقی فضای سایبر می‌شوند، باید به شیوه‌های مختلف از طریق رسانه‌های مختلف مانند صداوسیما و پایگاه‌های فضای مجازی آگاه‌سازی انجام شود.

منابع

- آذر، عادل؛ احمدی، پرویز و وحیدبسط، محمد. (۱۳۸۹)، طراحی مدل انتخاب نیروی انسانی با رویکرد داده‌کاوی، نشریه مدیریت فناوری اطلاعات، ۲(۴)، صص ۳-۲۲..
- ابراهیمی، مجیب؛ میرروشندل، سیدابوالقاسم و آقایی، جان احمد. (۱۳۹۴). جامعیت‌بخشی به مجموعه داده جرائم به‌منظور پیش‌بینی و شناسایی جرائم با استفاده از فنون داده‌کاوی، فصلنامه صنایع الکترونیک، ۶(۴)۴، صص ۶-۱۱.
- احمدوند، علی‌محمد و آخوندزاده، الهام. (۱۳۸۹). چارچوب کاربردی فنون داده‌کاوی در مدل‌سازی جرایم، دوماهنامه توسعه انسانی پلیس، ۷(۳۰)، صص ۱۱-۲۲.
- ارجمندی، اسماعیل. (۱۳۹۳). سیر تطور رسانه‌های مجازی، تهران همایش آینده‌نگاری فناوری اطلاعات.
- اسکندری، حمیدرضا؛ علیزاده، سمیه و کاظمی، پروانه. (۱۳۸۸)، کاربرد داده‌کاوی در شناسایی و کشف الگوهای پنهان جرم سرقت. فصلنامه نظم و امنیت انتظامی. ۴(۴)، صص ۳۵-۵۶.
- اسماعیلی، مهدی. (۱۳۹۴). داده‌کاوی مفاهیم و تکنیک‌ها. ویرایش سوم. تهران: نیاز دانش.
- بوربور، مسعود. (۱۳۹۳). تبیین شیوه‌های مبتنی بر فناوری اطلاعات برای پیشگیری از کلاهبرداری‌های مالی موجود در فضای مجازی. پایان‌نامه کارشناسی ارشد مهندسی فناوری اطلاعات دانشگاه علوم انتظامی امین.

- پولادی، رضا. (۱۳۹۶). داده‌کاوی ارتباط میان جرایم و ویژگی‌های شغلی کارکنان در نیروی انتظامی. پایان‌نامه کارشناسی ارشد مدیریت فناوری اطلاعات دانشگاه علوم انتظامی امین.
- جلالی، علی‌اکبر. (۱۳۹۱). رفتارشناسی مجرمان در فضای سایبر. فصلنامه کارآگاه، ۶(۲۱)، صص ۶-۲۵.
- ژیاوی، هان، میشلین، کمبر، ژان پی. (۱۳۹۳). داده‌کاوی (مفاهیم و تکنیک‌ها). (مهدی اسماعیلی، مترجم).
- سایت پلیس فتا.
- شاه‌محمدی، غلامرضا و اکباتانی، سمیه. (۱۳۹۴). پیشگیری مبتنی بر فناوری اطلاعات از آسیب‌های فضای مجازی. پژوهش‌نامه نظم و امنیت انتظامی. شماره ۲۹، صفحه ۲۷.
- شاه‌محمدی، غلامرضا و تاهو، منصور. (پاییز ۱۳۹۳). بررسی شیوه‌های پیشگیری از جرایم سایبری مبتنی بر فناوری اطلاعات. فصلنامه پژوهش‌های اطلاعاتی و جنایی. ۹(۳)، صص ۹۹-۱۲۰.
- شاه‌محمدی، غلامرضا و عباسی، سعید. (۱۳۹۷). بررسی عوامل موثر در تصادفات درون شهری با استفاده از روش‌های داده‌کاوی (مورد مطالعه: شهر اصفهان). پژوهشنامه جغرافیای انتظامی، ۶(۲۱)، صص ۱۳۵-۱۶۲.
- کیوان‌پور، محمدرضا؛ جاویده، مصطفی و ابراهیمی، محمدرضا. (۱۳۸۸). تحلیل رایانه‌ای جرم با بهره‌گیری از روش‌های هوش مصنوعی و داده‌کاوی کشف پیش‌دستانه جرم. فصلنامه کارگاه، ۲(۷)، صص ۹۸-۱۱۷.
- Alpaydin, E. (۲۰۱۰). Introduction to Machine Learning. London: The MIT Press Cambridge, pp ۳۵-۳۹.
- Chapman, P. & Clinton, J. & Kerber, R. & Khabaza, T. & Reinartz, T. & Shearer, C.& Wirth, R. (۲۰۰۰). CRISP-DM Step-Data Mining Guide, SPSS In, CRISPMWP.
- http:// iransignalman/data-mining/methodology-fig/semma
- http://www.spss.com/Clementine/

- Khan, M. A., Pradhan, S. K., & Fatima, H. (۲۰۱۷). Applying Data Mining Techniques in Cyber Crimes. In Anti-Cyber Crimes (ICACC), ۲nd Inter. Conf. on (pp. ۲۱۳-۲۱۶). IEEE.
- Sahu, N., & Darokar, S. (۲۰۱۷). Data Mining Techniques to Clustering Cyber Crime Data. International Education and Research Journal, ۳(۷).
- Shea, A. P. (۲۰۱۷). A Visual System for Mining Crime Mining across College Campuses. In Proceedings of the ۲۰۱۷ ACM Inter. Conf. on Management of Data (pp. ۱-۳). ACM.
- Shearer, C. (۲۰۰۵). The CRISP-DM Model: The New Blueprint for Data Mining. Journal of data Warehousing, ۵(٤), pp ۱۳-۲۲.
- Singhal, P., & Bansal, A. (۲۰۱۳). Improved Textual Cyberbullying Detection Using Data Mining. International Journal of Information and Computation Technology, ۳(٦), pp. ٥٦٩_٥٧٦.
- Taha, K., & Yoo, P. D. (۲۰۱۷). Using the Spanning Tree of a Criminal Network for Identifying Its Leaders. IEEE Transactions on Information Forensics and Security, ۱۲(۲), pp. ٤٤٥_٤٥٣.