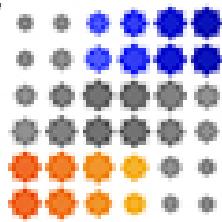


# آداده کاوی

dataacademy.ir



## داده کاوی و کاربردهای آن در مهندسی حمل و نقل

محمد رحمانی<sup>۱</sup>، وحید امینی طوسی<sup>۲</sup>، آلا نجار وظیفه دان<sup>۳</sup>

<sup>۱</sup> دانشجو دکتری برنامه ریزی توریسم، عضو هیئت علمی دانشگاه آزاد اسلامی همدان

Email mohammad.rahamani1@gmail.com

<sup>۲</sup> کارشناس کامپیوتر گرایش نرم افزار، کارشناس معاونت حمل و نقل و ترافیک شهرداری مشهد

Email:vahidamini@gmail.com

<sup>۳</sup> کارشناس شهرسازی

Email:alavazifedan@gmail.com

### چکیده

مدیریت و برنامه ریزی در فرایند حل مسائل و مشکلات پیچیده و مختلف ، شاید تنها راه حلی باشد که امروزه بدون داشتن اطلاعات کافی نه تنها سبب حل مشکل نمی گردد بلکه مشکلات جدیدی را نیز پیدا می آورد. تاریخچه مدیریت شهرها ، معمولاً آنکه از تصمیم گیری های عجولانه ای است که بانی تحمیل مشکلات افزونتری بوده اند . سیستم های پیچیده حمل و نقلی نیز از این قاعده مستثنی نمی باشند و برنامه ریزی های دراز مدت در این سیستم ها تنها راه بهبود آنها می باشند . در این راستا این مقاله سعی دارد تا به نوعی نقش فن آوری اطلاعات بخصوص بحث داده کاوی را در مدیریت سیستمهای حمل و نقلی روشن نماید . کاربرد داده کاوی در بخش حمل و نقل با توجه به حجم بالای اطلاعات ترافیکی و وجود انبوه داده ها غیر قابل انکار بوده و سبب می گردد تا با روش های ترکیب داده های مختلف از لطمات و صدمات فراوان وارد توسط فقدان مدیریت یکپارچه در امور حمل و نقل و ترافیک جلوگیری به عمل آورد .

### کلمات کلیدی

انباره داده، داده کاوی، سیستم اطلاعات جغرافیایی، سیستم موقعیت یاب جهانی

این فرآیند ما را در استخراج اطلاعات مفید از انبار داده های بسیار بزرگ یاری می نماید . اطلاعات استخراج شده می تواند در مدل های پیش بینی ، دسته بندی اطلاعات ، یافتن ارتباطات موجود ، تعریف روند داده ها و تصحیح مدل های موجود و استخراج خلاصه داده ها کاربرد داشته باشد . امروزه روش های بسیار متنوع داده کاوی گسترش یافته و به تنهایی دراستنتاج قوانین برای شبکه های عصبی و دسته بندی های مفهومی مورد استفاده قرار می گیرند و دامنه فعالیت آنان از فقط کار بر روی داده های چند بعدی فراتر رفته و داده های مالی را نیز شامل گشته است . مقالات جدید نشان می دهد که روش های داده کاوی دو هدف عمده را دنبال می کنند :

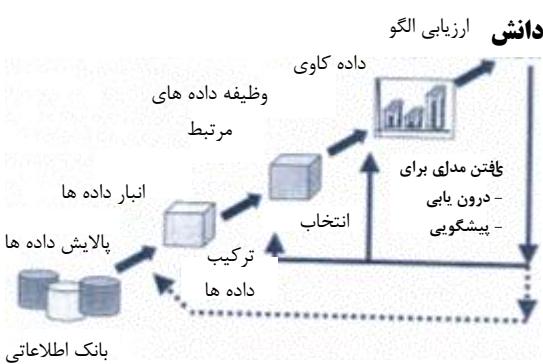
- دادن بیانش کلی از داده ها یا درون یابی
- پیشگویی

### ۱- مقدمه

به منظور اتخاذ تصمیمات موثر و مفید تر توسط تصمیم سازان و مدیران شهری ، استفاده از سامانه های بانک اطلاعاتی امری مسجل و میرهن می باشد و با توجه به گستردگی اطلاعات موجود در حوزه مدیریت شهری ، بانک های اطلاعاتی مختلف و گوناگونی وجود دارند و یک انبار داده مجموعه ای از این داده ها می باشد که از یک یا چند بانک اطلاعاتی جمع آوری شده که در گروه های مختلف اطلاعاتی تقسیم گردیده اند .

با توجه به رشد روزافزون تکنولوژی و علوم کامپیوتر برای استخراج داده ها از انبار داده ها ، ابزارهای گوناگونی معرفی شده اند ، یکی از این ابزارها داده کاوی می باشد . در واقع داده کاوی عبارت است از فرآیند کلی تشخیص الگوهای قابل فهم نهایی و معتبر در انبار داده ها

- ۵- انتخاب و معرفی داده های مرتبط
- ۶- انتخاب وظایف داده کاوی
- ۷- انتخاب ابزار داده کاوی (شبکه های عصبی ، الگوریتم ژنتیک ، درخت تصمیم ، روش نزدیکترین همسایگی ، معرفی اصول مربوطه و تجسم داده ها)
- ۸- تعریف ارتباطات ، کلاسها ، خوشه ها ، وابستگی ها ، الگوهای دنباله ای وغیره
- ۹- تفسیر نتایج
- ۱۰- تقویت دانش های کشف شده



شکل شماره (۱) : فرآیند داده کاوی و کشف دانش

## ۲ انبار داده

انباره داده مجموعه ای از داده های موضوع گرا ، مجتمع ، غیرفار و وابسته به زمان است که برای پشتیبانی از تصمیم گیریهای مدیریتی مورد استفاده قرار می گیرد. به عبارت دیگر یک انبار داده ، داده های دلخواه را از یک یا چند منبع جمع آوری کرده و آنها را به موضوعاتی با گروه های اطلاعاتی تبدیل می کند، سپس آنها را به همراه اطلاعات زمان و تاریخ برای پشتیبانی بهتر تصمیم گیریها ذخیره می کند، این سیستمهای دیده های متعددی از اطلاعات برای طیفهای مختلف کاربران فراهم می کنند، قدرت این مفهوم در آن است که به کاربران اجازه تحلیلها و پرس و جوهای گوناگون بر روی داده هایی می دهد که قبل از آن هیچ ارتباطی با هم نداشتند.

انبار داده به صورتی طراحی می شود که کاربران اطلاعات مورد نیاز خود را شناسایی کنند و بتوانند با استفاده از ابزارهای ساده ای به آنها دسترسی داشته باشند.

## ۲ خصوصیتهای اصلی انبار دادهها

یک انبار داده دارای چهار خصوصیت اصلی میباشد.

- موضوع گرایی: هر انبار داده ، داده های مرتبط با یک موضوع خاص را در خود نگاه میدارد و این دادهها را به منظور استخراج مفاهیم و نتایج خاصی به شکلی ویژه سازماندهی میکند. بدین ترتیب سرعت

درون گلخانه به ما در تشخیص روند و الگوی داده ها یاری نموده ، و پیش گویی مدلی قابل اعتماد براساس داده های ورودی به دست می دهد .

واضح است که موققیت روشهای داده کاوی بستگی زیادی به ذات داده ها دارد و ذات داده ها نیز بستگی زیادی به منبع ، کاربرد ، رفتار و مقیاس آنها دارد .

- منبع داده ها می تواند به صورت آنلاین یا آفلاین از یک سیستم استاتیک یا داینامیک باشد .

- کاربرد داده ها می تواند برای آنالیز ، طراحی و تشخیص باشد .

- رفتار داده ها می تواند گسسته یا پیوسته باشد .

- مقیاس داده ها می تواند کمی یا کیفی باشد .

کیفیت کار بر روی داده های کمی بستگی زیادی به تعداد داده های در دسترس دارد و کیفیت کار بر روی داده های کیفی نیازمند دانستن عوامل زیر است :

میزان پراکندگی یا چگالی داده ها ، میزان داده های ناخالص ، نماینده بودن نقاط در دسترس ، پراکندگی داده ها و تعداد داده های گمشده .

محققان تحقیقات وسیعی را در مباحث علمی داده کاوی بخصوص بحث درونیابی آغاز کرده اند ، حل مسایل رگرسیون ، طبقه بندی و

یافتن داده های گمشده از این دست می باشند که عموماً برای تشخیص الگوها ، کنترل و مانیتورینگ فرآیندها و تشخیص خطاهای غیره به کار می روند . در شاخه مهندسی حمل و نقل ، مطالعات مربوطه به حجم انبوهی از اطلاعات به شرح زیر منجر می گردد ،

اطلاعات مربوط به مدیریت ترافیک ، تصادفات ، وضعیت سواره روهای ، فهرست مسیرها و جاده های آینده ، فهرست چراغهای راهنمایی و رانندگی و علایم ، تعمیر و نگه داری پلهای عابر پیاده ، فهرست مشخصات جاده ها و غیره تمام اطلاعاتی هستند که تصمیم سازان بر اساس آنها به حل مسایل و مشکلات می پردازند . یکی دیگر از مشکلات موجود کار با داده های گسسته است ، توانایی تشخیص داده های در دسترس و تبدیل آن به فرمت مورد نیاز نیز از دیگر توانایی های داده کاوی است که در تصمیم سازی مدیران آنها را یاری می نماید .

قلمرو حمل و نقل به گونه ای است که فیلدهای متفاوتی از داده ها برای حل مسایل می باشند جمع آوری گردند و مطالعات جدید در زمینه داده کاوی افکارهای جدیدی را برای مهندسان و تصمیم سازان حمل و نقل باز می نمایند .

قدمهای اصلی در بخش داده کاوی به شکل زیر می باشند (شکل شماره (۱) ) :

۱- شناخت دامنه کاربرد داده ها

۲- جمع آوری بانک هدف

۳- پالایش داده ها و بررسی مجدد آنها

۴- انبار کردن داده ها

اما انبارهای داده برای پردازش OLAP طراحی شده اند.

- از لحاظ عملیات قابل اجرا بر روی آنها : عملیاتی که بر روی پایگاه دادهای صورت میگیرد، عموماً شامل عملیات به هنگام سازی است در حالی که عمل خواندن از انبار، عمله عملیات قابل اجرا بر روی انبارداده را تشکیل می دهد.
- از لحاظ مقدار داده: مقدار دادهای یک پایگاه داده در حدود چند مگابایت تا چند گیگابایت است در حالی که این مقدار در انبار داده در حدود چند گیگابایت تا چند ترابایت است.
- از لحاظ زمان پرس و جو: انبار داده با استفاده از دو تکنیک خلاصه سازی و سلسه مراتبی کردن فیلدها سرعت انجام پرس و جو را بهبود بخشیده است در حالیکه در پایگاه داده اینگونه نیست.

### ۳ داده کاوی

سازمانها معمولاً روزانه مقدار زیادی داده را در انجام عملیات تجاری خود تولید و جمع آوری می کنند. امروزه برای این پایگاه های داده شرکتی عجیب نیست که مقدار داده های آن در حد ترابایت باشد. با این حال علی رغم ثروت اطلاعاتی عظیم ذخیره شده حدس زده میشود که فقط ۷ درصد کل دادههایی که جمع آوری میشود مورد استفاده قرار میگیرد. بدین ترتیب مقدار قابل توجهی داده که بدون شک حاوی اطلاعات ارزشمند سازمانی است تا حد زیادی دست نخورده باقی میماند. در محیط تجاری عصر اطلاعات، که هر روز رقباتی تر میشود میتوان با استخراج اطلاعات از دادههای استفاده نشده به تصمیم گیریهای استراتژیک دست یافت. در طول تاریخ تحلیل دادهها از طریق رگرسیون و دیگر تکنیکهای آماری انجام شده است. برای استفاده از این تکنیکها ، لازم است که تحلیل گر مدلی خلق کند و فرآیند گردآوری دانش را سازمان دهد. اما امروزه این روش ها به تنها یک کافی نیستند و باید از روشها خودکار نیز استفاده کرد. داده کاوی عبارت است از فرآیند خودکار کشف دانش و اطلاعات از پایگاههای داده.

این فرآیند تکنیکهایی از هوش مصنوعی را بر روی مقادیر زیادی داده اعمال می کند تا روندها، الگوهای روابط مخفی را کشف کند. از این های داده کاوی برای کشف دانش یا اطلاعات از داده ها به کاربر اتنکنند، بلکه فرآیند پیشگویی واقعیتها را خودکار میسازند. این تکنولوژی نوظهور، اخیراً به طور فزایندهای در تحلیلهای مورد استفاده قرار میگیرد.

داده کاوی در پیاده سازی قابلیتهای خود از الگوریتمهای هوش مصنوعی، مثل شبکهای عصبی، بهره میگیرد، که میکوشند تقليید

جستجوها بسیار بالا خواهد بود. انبار داده برای پاسخگویی به پرسشهای مختلف در مورد یک موضوع خاص بهینه سازی میشود.

- مجتمع: در سیستمهای مختلف، دادهها ممکن است از جنبههای مختلفی با هم نامتنااسب باشند. مثلاً منابع داده در کشورهای مختلف با زمان و تاریخهای مختلف ذخیره شده اند و یا منابعی که از ماشین های مختلف هستند در اعداد، حروف و دیگر فیلدها متناسب با محدودیتهای ماشین، سیستم عامل و تطابق با استانداردهای مختلف متفاوت باشند. با توجه به مختلف بودن منابع انبار داده، قبل از ذخیره سازی آنها در انبار داده برای تأمین یکپارچگی از تکنیکهای پاکسازی دادهها و مجتمعسازی استفاده می شود. با توجه به این موضوع که دادهها زیر نظر مدیریت واحدی ذخیره سازی می شوند، داده های سازگاری خواهیم داشت که داشتن چنین سازگاری در سایر سیستمهای مثل سیستمهای شبکهای و توزیعی ناممکن است.

متغیر با زمان: دادهها در انبارداده برای تهییه اطلاعات تاریخی به کار میروند . هر ساختار کلیدی در انبار داده شامل عنصر زمان با همان مهر زمانی است. این مهر زمانی به عنوان کلیدی به سایر جداول عمل می کند. در سیستم انبار داده، دادهها هرگز به روز نمیشنوند، بلکه دادههای جدید و یا تغییر یافته با مهرهای زمانی جدید به انبار اضافه می شوند.

غیر فرار: دادههای انبارداده همیشه از لحاظ فیزیکی مجزا هستند، و هیچگاه نیازی به تغییر و به روز آوری نخواهند داشت. با توجه به این موضوع، انبارداده مثل پایگاه دادههای معمولی احتیاجی به پردازش تراکنش، بازگرداندن فرایند و مکانیزم کنترل تصادم ندارد علاوه بر این نیازی به ایجاد و دسترسی انحصاری به دادهها نخواهیم داشت.

### ۴ تفاوت انبار داده و پایگاه داده

وظیفه اصلی سیستم های پایگاه داده آنلاین، پشتیبانی از تراکنش های آنلاین و پردازش پرس و جو است. ولی انبارداده به کاربران یا تحلیلگر داده و تصمیم گیرنده ها، داده ارائه می کند. در زیر به موارد تفاوت انبار داده و پایگاه داده اشاره می شود.

- از لحاظ مدلها داده: پایگاههای داده برای مدل OLTP امکان پردازش تعداد زیادی تراکنش همروند، که اغلب حاوی رکوردهای اندکی هستند را دارد.

توان با استفاده از داده‌های تاریخی برای تولید یک مدل یا نمایی از یک گروه بر اساس ویژگیهای داده‌ها به کار برد. سپس میتوان از این مدل تعریف شده برای طبقه‌بندی مجموعه داده‌های جدید استفاده کرد. همچنین میتوان با تعیین نمایی که با آن سازگار است برای پیش‌بینیهای آتی از آن بهره گرفت.

برای مثال، برای طبقه‌بندی تخلفات و کلاهبرداری‌ها در صنعت و اعتبارات، با استفاده از قابلیت طبقه‌بندی داده کاوی، سیستم با استفاده از مجموعه‌های از پیش تعریف شده از داده‌ها، تعلیم می‌بیند. مجموعه داده‌های مورد استفاده در این نمونه باید هم شامل مجموعه هایی از داده‌های معتبر باشند و هم شامل مجموعه هایی از داده‌های جعلی. از آن جا که این داده‌ها از پیش تعریف شده هستند، سیستم پارامترهایی را می‌باید که میتوان از آنها برای تشخیص طبقه‌بندیهای متماز استفاده کرد. بعد از تعیین پارامترها سیستم از آنها برای طبقه‌بندیهای بعدی بهره خواهد گرفت.

در واقع سیستمهایی که بر اساس طبقه‌بندی داده کاوی میکنند، دو مجموعه ورودی دارند: یک مجموعه آموزشی که در آن داده‌هایی که به طور پیش فرض در دسته‌های مختلفی قرار دارند، همراه با ساختار دسته‌بندی خود وارد سیستم می‌شوند و سیستم بر اساس آنها به خود آموزش میدهد یا به عبارتی پارامترهای دسته‌بندی را برای خود مهیا می‌کند. دسته دیگر از ورودیهایی هستند که پس از مرحله آموزش و برای تعیین دسته وارد سیستم می‌شوند.

تکنیک‌هایی که برای داده کاوی که برای دسته بندی به کار می‌آیند عموماً شامل تکنیک‌های شبکه عصبی و درخت تصمیم گیری هستند.

یکی از متغیرهای مهم در قابلیت دسته بندی برآورد یا امتیاز دهنده است. هر جا طبقه‌بندی یک پاسخ دو تایی مثل بله و خیر ارائه کند، برآورد یک درجه بندی مثل پایین بالا و یا متوسط ارائه می‌کند. در اصل برآورد، چندین نما در امتداد مجموعه‌های از داده‌ها فراهم می‌آورد که نشاندهنده درجه تعلق یک نما به یک مجموعه است.

کاربرد قابلیت طبقه‌بندی در بازاریابی هدف، تصویب اعتبار و بررسی تقلب، است.

#### ۴-۱ الگوهای ترتیبی

قابلیتهای ترتیبی هم مانند قابلیتهای هم پیوندی این خاصیت را دارند که میتوانند وقایع را با هم مرتبط کنند. این کار در هم پیوندی سنتی یا تحلیل سبد بازار مجموعه‌های از مقولات را به عنوان مقولات پشت سر هم ارزیابی می‌کنند و از ابزارهایی مثل سریهای زمانی هم برای تعیین ترتیب بهره می‌برند. الگوهای ترتیبی علاوه بر آن، این قابلیت جدید را هم دارند که میتوانند فاصله زمانی بین دو واقعه را تخمین بزنند. برای مثال این قابلیت امکان نتیجه گیری‌هایی از قبیل اینکه "۸۰٪ افرادی که کامپیوتر می‌خرند ظرف مدت ۱ سال چاپگر هم خواهند خرید"، را مهیا مینماید. به این ترتیب شناسایی نوعی از خریدهای مقدماتی که پتانسیل خریدهای بعدی را در آینده تعیین می

از عملکرد مغز انسان برای محاسبات موازی ارائه کنند. داده کاوی توانسته است با استفاده از شبکه‌های عصبی و دیگر مفاهیم بر گرفته از هوش مصنوعی به نتایجی دست پیدا کند که حتی متخصصین این زمینه به آنها دست نیافتنه اند. این تکنیکها امکان تحلیل مقادیر بسیار بزرگتری از داده را در مقایسه با رو شهای سنتی فراهم می‌آورد. به علاوه داده کاوی کشف دانش از اطلاعات را در پی خواهد داشت و باعث پدید آمدن پیش‌بینیهایی می‌شود که فراتر از عملکرد متخصصین این حوزه است.<sup>[1]</sup>

اهداف داده کاوی عبارتند از:

- تفسیرگر: می‌تواند علت برخی از شرایط و پیشامدها را توضیح دهد، از قبیل چرا فروش جوراب سفید از سال گذشته کم نشده است؟
- تایید کننده: می‌تواند یک نظریه را تأیید یا رد کند.

#### ۴-۲ ابزارهای داده کاوی و قابلیتهای آن

به منظور تحقق اهداف داده کاوی که در بالا ذکر شد و در راستای نتایج مطالعات جامع، ابزارهای داده کاوی با استفاده از الگوریتم‌های درخت تصمیم، شبکه‌های عصبی و غیره علاوه بر بهبود فرایند استنتاج سبب افزایش سرعت چرخه طراحی نیز می‌گردد. تعدادی از این ابزارهای نرم افزاری به شرح زیر می‌باشد:

WizWhy, S-Plus, Darwin, Intelligent Miner, ModelQuest, NeuroShell, Enterprise Miner

در ادامه به برخی از قابلیتهای این ابزارها اشاره می‌شود:

#### ۴-۳ هم پیوندی

هم پیوندی قابلیتی برای یافتن روابط ناشناخته موجود در اطلاعات است. این روابط مواردی از قبیل اینکه حضور مجموعه‌های از مقولات اشاره به این دارند که مجموعه مقولات دیگری نیز احتمالاً وجود دارند را شامل می‌شود. این قابلیت اساساً روشی برای کشف مقولاتی است به هم میخورد. از آن با عنوان تحلیل سبد بازار یا گروه بندی خویشاوندی نیز یاد می‌شود.

برای مثال، گزار شهای هم پیوندی چنین شکلی دارند: "۸۰٪ درصد مشتریانی که کالای "آ" را خریداری نموده اند، کالای "ب" را نیز خریده اند. درصد خاص وقوع وقایع (مثلًا ۸۰٪ درصد این نمونه) را فاکتور اطمینان هم پیوندی "آ" و "ب" می‌نامند. همچنین ممکن است هم پیوندی‌های چند گانه وجود داشته باشد.

از کاربرد های هم پیوندی میتوان به برنامه ریزی موجودی، برنامه ریزی تبلیغاتی برای فروش و مراسلات و بازاریابی مستقیم اشاره نمود.

#### ۴-۴ طبقه‌بندی

طبقه‌بندی در واقع ارزشیابی ویژگیهای مجموعه‌های از داده‌ها و سپس اختصاص دادن آنها به مجموعه‌های از گروههای از پیش تعریف شده است. این متدالترین قابلیت داده کاوی می‌باشد. داده کاوی را می

در داده ها سبب کاهش چشمگیر هزینهها در طراحیهای ترافیک می گردد.

پاسخ به مباحث مختلف زیر ، برای حل مسایل و مشکلات مدیریت ترافیک توسط داده کاوی ضروری است :

- مباحث جدیدی که داده کاوی توانایی خودکار کردن آن فرآیندها را دارد .
- ساختار مناسب انبار دادهها برای پشتیبانی از مدیریت ترافیک .
- خط مشیهای جدید برای گسترش، نمایش و ارایه داده های ترافیکی و فیلمهای سه بعدی .

#### ۴ - نظارت بر رانندگان خواب آلو

روشهای داده کاوی با توجه به اطلاعات جمع آوری شده درباره خصوصیات رانندگان در خواب ، به طراحی سنسوری در اتاق راننده می پردازد که با آنالیز داده ها تعیین می کند که چه موقع رانندگان کامیون شیوه اوقاتی هستند که به خواب میروند و سپس به آنها هشدار داده و از تصادفات رانندگی جلوگیری به عمل می آورد .

۴ - آنالیز داده های موجود در تصادفات جاده ای هنگام طراحی شبکه خیابانی دادههای مرتبط با خطرات موجود ، جهت ایمنی جادهها جماعتی میگردد. تا در طراحی های آینده ما را راهنمایی نمایند . عوامل مختلف موثر در این داده ها گاهی سبب بروز اشتباهاتی در تصمیم گیری ها می گردد و برآورد قابل اعتمادی برای تعداد تصادفات در مکانهای مختلف مختلف جاده به دست نمی دهد . داده کاوی به کمک دادههایی که از شبکه خیابانی جماعتی شده است ، علیرغم تغییرات زیادی که در فراوانی تعداد تصادفات رخ می دهد ، نقاط حادثه خیز را مشخص می نماید . این روشهایی توانند با استفاده از فرآیندهای علت و معلولی عوامل مختلف محیطی و غیره را مشخص نمایند.

تکنیکهای دیگر داده کاوی در رابطه با مطالعات تصادفات که توسط سایر دانشمندان به کار برده شده است به شرح زیر میباشد :

- روش تصویری جدیدی که اینیمیشن تصادفات را در یک موقعیت مشخص و زمان خاص تولید میکند. این روش جهت یافتن نقاط حادثه خیز ما را یاری میسازد.
- اصول مرتبط که برای یافتن ارتباط میان تعداد جاده ها و انواع تصادفات متداول ما را کلی نماید.
- روش تکست کاوی و یافتن زیرگروهها ما را در تعیین انواع تصادفات متداول یاری مینماید.
- یافتن زیر گروهای پویا که سبب پر رنگ تر شدن مبحث کیفیت دادههای مشخص میگردد.

کنند، عملی می شود. در نتیجه از چنین تحلیلهایی به شدت در تبلیغات فروش استفاده میگردد .

#### ۳ - ۴ خوشه بندی

قابلیت خوشه بندی وظیفه تقسیم یک گروه ناهمجنس را در چندین زیر گروه بر عهده دارد. این فرایند یک تفاوت اساسی با طبقه بندی دارد. زیرا در این مدل هیچ گونه الگوی آموزشی نداریم. خوشه بندی به طور خودکار ویژگیهای متمایز کننده زیر گروهها را تعریف میکند و زیر گروهها را سازماندهی مینماید. و به عنوان نوعی قابلیت داده کاوی غیر مستقیم مطرح است.

این ابزارها پایگاه داده را بر اساس ویژگیهای داده ها به چندین بخش تقسیم می کنند و گروه هایی از رکوردها را به وجود میآورند که نمایانگر یا صاحب صفت خاصی هستند. الگوهای به دست آمده در ذات پایگاه داده نهادینه هستند و نشانگر بعضی اطلاعات غیرمنتظره و در عین حال ارزشمند شرکتی باشند.

مثالی از کاربردهای خوشه بندی در تقسیم بندی افرادی است که به پرسش نامهای پاسخ دادهاند. از این کار میتوان در تقسیم بندی مشتریانی که به پرسش نامهای پاسخ دادهاند در گروه هایی که اعضای آن بیشترین شباهت را با یکدیگر و بیشترین تفاوت را با اعضای سایر گروهها دارند، استفاده کرد. بعد از بخش بندی جمعیتی با استفاده از قابلیت خوشه بندی میتوان بر روی خوشهای مشخص شده تحلیل هم پیوندی انجام داد تا خریدهای به هم مرتبط یک گروه جمعیتی خاص شناخته شوند.

کاربرد خوشه بندی برای تعیین بهترین گروههای جامعه شناختی برای اهداف خاص بازاریابی است.

اغلب از خوشه بندی به عنوان اولین گام فرایندهای داده کاوی یاد میشود که قبل از سایر فرایندها برای شناسایی گروهی از رکوردهای مرتبط با هم که بعداً بتوانند نقطه آغاز تحلیلها باشند بر روی رکوردها اعمال می شود.

#### ۴ - کاربرد داده کاوی در مهندسی حمل و نقل

در زیر به عنوانهایی از مباحث مطرح در مهندسی حمل و نقل اشاره می گردد که استفاده از داده کاوی در آنها بسیار روش می باشد .

#### ۴ - مدیریت ترافیک

برای سالهای بسیاری محققان فقط بر روی مدیریت ترافیک و کنترل ترکم ترافیک تحقیق میکردند و سیستمهای و سنسورهای نصب شده در جاده ها میباشد توافقی دسته بندی و تفکیک وسایل نقلیه را می داشتند .

ولیکن امروزه جهت یافتن ارتباطات اساسی جریان ترافیک و تغییر آنها تحت شرایط مختلف محیط و جاده نیاز مبرم به داده کاوی و نرم افزارهای کاربردی دیگر حس می گردد ، یافتن این ارتباطات اساسی

روی نقشه تصحیح گردد . کاربرد داده کاوی در GPS تصحیح خودکار نقشه ها می باشد . این امر سبب می شود تا براحتی و سادگی خطوط مورد نظر را جا به جا نموده و همچنین اطلاع رسانی دقیق نیز داشت [5].

**جدول شماره ۱ : مشخصات داده ها در GIS و کاربرد آن در مهندسی حمل و نقل**

منطقی	فیزیکی	
تعریف منطقی	تسهیلات موجود	
مسیر	بزرگراه ها	ساختر
شبکه اصلی استانی	جاده ها	موجود
شبکه اصلی کشوری	تجهیزات	
شبکه خیابانی	تقاطع ها	
حدود مزدهای سیاسی		
ساختر داده ها	مقدار داده ها	
شبکه خیابانی	تعداد خطوط	ساختر
ارتباطات	نقاط	GIS
لینکها	پلی لاین ها	
نودها	پلی گون ها	
شبکه	خصوصیات	

#### ۴ داده های مخصوص از تصاویر ضبط شده جادهها

سیستم های ضبط تصاویر برای جمع آوری اطلاعات مربوط به سایل نقلیه، سواره روها و رعایت حق تقدیم ها می تواند استفاده شود. اطلاعات ویدیویی که به وسیله مسئولان بزرگراهها استفاده می شوند به فرمت آنالوگ در محلهای خاص انبار می گردد. این انبار شامل نوارها، فیلمها و دیسکهای لیزری می باشند. مهندسان آنها را در بانکهای اطلاعاتی متفاوتی نگه داری می نمایند. هدف از جمع آوری چنین اطلاعاتی به دست آوردن سریع، کم هزینه و کارآمد آن اطلاعات مخصوص در رابطه با مشخصات هندسی و اطلاعات مرتبط با خیابانها و جاده ها می باشد و کاربرد مهمتر آن، استفاده در طراحی های خاص است.

حال با توجه به اینکه می توان محل دقیق کلیه داده ها را در سیستم اطلاعات جغرافیایی نمایش داد، نیاز داریم که بتوانیم اطلاعات مورد نیاز را از تصاویر ضبط شده استخراج نماییم. در اینجا داده کاوی دارای پتانسیلهای مورد استفاده بسیاری در استخراج اطلاعات مهم از تصاویر ضبط شده است.

#### ۴ مدیریت رویه خیابان

سیستم مدیریت رویه خیابان که هم اکنون مورد استفاده قرار می گیرد ، از میان مقدار زیادی از اطلاعات استخراج گردیده است . آنالیز این دادهها برای بزرگراههای بسیاری انجام گردیده تا بتواند اساس تصمیمات مهم و اساسی در رابطه با رویه خیابان باشد .

اخیراً روشهای داده کاوی جهت برآورد نرخ قابلیت استفاده و تعمیر رویه خیابانها نیز به کار می روند و از آن به عنوان یک روش غیر مستقیم برای به دست آوردن طول عمر این رویهها استفاده می شود .

مطالعات داده کاوی اخیراً بر روی پارامترهایی به شرح زیر نیز به کار میروند:

- نرخ قابلیت استفاده و تعمیر خیابانهای وضع موجود (شیارها، کیفیت سواره ره، شرایط خیابان و نوع آسفالت)
- وضعیت پوششها (رویه اولیه و نوع رویه)

بعلاوه این روشهای میتوانند پارامترهایی شبیه درصد تعداد کامیون عبوری و وسایل نقلیه سبک را نیز در طراحیها لاحظ نموده و بر اساس آن طول عمر آسفالت و نوع آن را تعیین نماید. استحکام بتون استفاده شده و همچنین نوع آسفالت نیز در بررسی ها مورد استفاده قرار میگیرد. در انتها باید توجه داشت که پتانسیل های بسیاری در داده های سیستم مدیریت رویه خیابان برای استفاده داده کاوان وجود دارد.

#### ۵ سیستم اطلاعات جغرافیایی برای داده های حمل و نقل

در سیستم اطلاعات جغرافیایی مرتبط با حمل و نقل، سخت افزار، نرم افزار، اطلاعات جمعیتی و سازمانی و اصول موجود با هم ترکیب می گردد که این مهم منجر به تولید حجم وسیعی از اطلاعات می گردد . از کاربردهای سیستم اطلاعات جغرافیایی در شبکه های حمل و نقل ترکیب داده ها به شرح جدول شماره ۲ می باشد.

در رابطه با اطلاعات موجود در سیستم اطلاعات جغرافیایی بحث شده در بالا دیده می شود که داده کاوی می تواند یک ابزار عالی برای تشخیص این ارتباطات پیچیده بین داده های فیزیکی و منطقی در ساختار موجود و ساختار GIS ای باشد.[6]

#### ۶ داده های سیستم موقعیت یاب جهانی

امروزه روشهای خودکاری مانند سیستم موقعیت یاب جهانی به جزء لاینفک وسایل نقلیه جدید تبدیل شده اند . که می تواند جزئیات مسیر را بر روی نقشه مشخص نمایند. GPS توانایی درج مسیرهای مختلف حرکت ها را در جهات مختلف دارد که این امر سبب تولید حجم عظیمی از اطلاعات می گردد که می باشد به طور مرتبت بر

- میزان صدمه قفسه سینه
  - شکستگی استخوان چپ
  - شکستگی استخوان راست
- با استفاده از رکوردهای متغیرهای بالا و همچنین تعریف ۸ متغیر اول به عنوان متغیرهای وابسته اصول مرتبط با میزان شکستگی و صدمات به سر، میزان صدمات قفسه سینه و استخوان ران چپ و راست را با استفاده از نرم افزارهای داده کاوی موجود به دست آورد (WizWhy,WizRule). جزئیات مربوط به این نرم افزارها در منبع ۱۷ آمده است.

این نرم افزارها بعد از خوشنود داده ها ، برآوردهای مناسبی را برای پارامترهای موجود از قبیل، حداقل احتمال اصول اگر- آنگاه و حداقل تعداد رکوردهای یک اصل به دست می دهد و کاربر براحتی می تواند نیازهای خود را از نرم افزار در خواست نماید . همچنین اصول ناسالم موجود برای جلوگیری از خطاهای ممکن لیست می گردد .

احتمال در اصول اگر- آنگاه به شکل زیر تعریف می شود : "نسبت میان تعداد رکوردهایی که در آن هم شرایط و هم نتایج حفظ می شوند به تعداد رکوردهای که در آن شرایط حفظ شده و نتایج یا حفظ می شوند یا نمی شوند".

**سطح معنی داری** درجه اعتبار اصل به دست آمده را نشان می دهد . اگر سطح معنی داری ۱- باشد، احتمال خطا که کیفیت احتمال را می سنجد، نشان می دهد که اصل موجود به طور تصادفی در داده های تحت آنالیز وجود دارد.[3]

تعدادی از اصول اگر- آنگاه که به وسیله WizWhy تولید گردیده در زیر آمده است :

- اصل میزان صدمه به سر: اگر کارخانه سازنده شورلت باشد و وزن وسیله بین (۵/۶۱۹ و ۵/۸۲۰) پوند باشد و محافظ ماشین فقط کمربند باشد آنگاه میزان صدمه به سر بیش از ۹۰۳/۰ می باشد.
- احتمال اصل به دست آمده برابر یک است.
- اصل به دست آمده برای ده رکورد صادق است.
- سطح معنی داری خطا کمتر از ۱/۰۰۰۰۱ است.

- اصل میزان صدمه به قفسه سینه: اگر کارخانه سازنده ماشین ایسوزو باشد و محافظ ماشین فقط کمربند اینمی باشد آنگاه صدمات قفسه سینه بیش از ۴۸/۳۷ می باشد.
- احتمال اصل به دست آمده برابر یک است.
- اصل به دست آمده برای چهارده رکورد صادق است.
- سطح معنی داری کمتر از ۱/۰۰۰۰۱ است.

- اصل صدمه به استخوان ران چپ: اگر ساخت اتومبیل سال ۹۰ باشد و محافظ ماشین کمربند اینمی خودکار باشد

#### ۴ آنالیز داده های مرتبط با ناهمواریهای جاده

سنجدش ناهمواریهای جاده ما را در رسیدن به یک ایده کلی در رابطه با کیفیت سواره روهای میزان رضایتمندی کاربران آنها و هزینه عملکرد وسایل نقلیه یاری می نماید.

با توجه به همین مطلب مهندسان راهها و بزرگراهها ، بررسی ناهمواریهای دوره ای یا سیستماتیک ایجاد شده را به کمک سنجش ناهمواری ها انجام می دهند. در طی مطالعه داده های قابل اندازه گیری برای پارامترهای زیر جمع آوری می گردد:

- کد شناسه جاده
- نقاط مهم محلی
- ناهمواری جاده و غیره

ابزار داده کاوی برای تعیین، ارتباط میان موقعیت مکانی محل، میزان رضایت کاربران جاده و ناهمواری جاده مورد استفاده قرار می گیرد. در این بخش کاربردهای متفاوت داده کاوی در مباحث مختلف مهندسی حمل و نقل مورد بحث و بررسی قرار داده شد در بخش بعد ئیکی از کاربردهای داده کاوی در بررسی داده های تصادفات نشان داده خواهد شد .

#### ۵ بانک اطلاعاتی آزمایشی تصادفات

برای نمایش کاربردهای داده کاوی در شاخه مهندسی حمل و نقل، اطلاعات از روی انجام تصادفات آزمایشی جمع آوری شد . در این آزمایش اتومبیلها در حالت مصنوعی با راننده و یک مسافر در صندلی جلو با سرعت ۳۵ مایل در ساعت با دیوار برخورد می کردند.

مسئولان اینمی حمل و نقل بین المللی اطلاعات مربوط تأثیرات برخورد را در تصادفات در حالت مصنوعی جمع آوری نمودند. متغیرهای جمع آوری شده تعداد زیادی صدمه وارد به سر، شکستگی در قفسه سینه و شکستگی استخوان ران چپ و راست را نشان می دادند .[2]

همچنین متغیرهایی در رابطه با نوع و میزان خسارت وارد به ماشین تصادف کرده نیز در بانک داده ها وجود دارد. این بانک دارای ۳۵۲ رکورد با متغیرهای زیر است:

- کارخانه سازنده وسیله
- مدل وسیله
- سال ساخت وسیله
- تعداد درها
- وزن به پوند
- اندازه : متغیر فوق وسیله ها را به چند دسته تقسیم می کند(سبک ، میزی و نو)
- انواع محافظهای موجود در وسیله : (کمربند اینمی، کیسه هوا و ...)
- D/P : (صدمه وارد به راننده است یا مسافر)
- ناحیه صدمه دیده سر

کارخانه سازنده	مدل	سال	تعداد درها	وزن (پوند)	اندازه	محافظهای موجود	D\P	میزان صدمه به سر	میزان صدمه به سینه	میزان صدمه ران چپ	شکستگی استخوان ران چپ	شکستگی استخوان ران راست
Acura	Integra	۸۷	۲	۲۳۵۰	1t	Manual belts	Driver	۵۹۹	۳۵	۷۹۱	۲۶۲	
Audi	80	۸۹	۴	۲۷۹۰	comp	Manual belts	Driver	۶۰۰	۴۹	۱۶۸	۱۸۷۱	
BMW	325i	۹۰	۲	۲۸۶۲	comp	d airbag	Driver	۱۰۳۶	۵۶	۸۶۵		
Buick	Elect.Park Ave	۸۸	۴	۳۳۶۰	med	Manual belts	Driver	۱۴۶۷	۵۴	۷۱۲	۱۳۶۶	
Buick	Regal	۸۸	۲	۳۲۱۰	med	passive belts	Driver	۸۸۰	۵۰	۹۹۶	۶۴۲	
Cadillac	De Ville	۹۰	۴	۳۵۰۰	hev	d airbag	Driver	۴۲۳	۳۹	۵۴۱	۱۶۲۹	
Chevrolet	Astro	۸۸		۳۷۸۷	van	Manual belts	Driver	۱۶۰۳	۷۲	۱۵۷۲	۷۰۰	
Chrysler	Fifth Ave	۸۹	۴	۳۸۲۰	hev	d airbag	Driver	۷۸۶	۴۳	۱۱۳۲	۶۶۷	
Daihatsu	Charade	۸۸	۲	۱۸۲۰	mini	Manual belts	Driver	۷۶۸	۴۳	۵۷۴	۵۹۸	
Dodge	colt	۸۸	۴	۲۳۴۸	1t	Manual belts	Driver	۱۲۵۴	۵۳	۸۴۴	۴۶۱	
Eagle	Mededlion	۸۹	۴	۲۷۴۰	comp	Motorized belts	Driver	۷۴۵	۴۱	۱۷۲۱	۱۵۷۴	
Ford	Aerostar	۸۷		۳۰۱۳	van	Manual belts	Driver	۱۵۶۸	۴۹	۲۸۶	۵۹۰	
Geo	Metro	۸۹	۲	۱۵۹۰	mini	Manual belts	Driver	۹۵۱		۷۸۹	۴۵۸	
Honda	Accord	۸۷	۲	۲۴۴۰	1t	Passive belts	Driver	۷۶۹	۴۶	۸۶۳	۱۲۴۴	
Hyundai	Excel	۸۷	۴	۲۲۰۰	1t	Passive belts	Driver	۷۵۷	۵۴	۲۴۰۸	۱۱۸۷	
Infiniti	M-30	۹۰	۲	۳۳۷۰	med	d airbag	Driver	۴۶۶	۴۳	۹۳۶	۸۴۴	
Isuzu	Amigo	۹۰	۲	۲۹۰۰	mpv	Manual belts	Driver	۹۹۶	۵۶	۳۸۸	۵۰۱	
Jeep	Cherokee4x4	۸۹	۴	۳۲۸۴	mpv	Manual belts	Driver	۹۶۸	۶۹	۴۰۱	۵۴۰	
Lexus	ES250	۹۰	۴	۳۲۸۰	med	d airbag	Driver	۹۹۲	۵۵	۱۱۰۲	۱۰۱۲	

اصولی که به این شکل با استفاده از حجم عظیمی از اطلاعات با ابزار داده کاوی تولید می گردد می توانند به عنوان بخش صحیح و کاملی از یک سیستم تصمیم سازی یا پایه علمی متخصصان باشند.[4]

آنگاه میزان صدمه به استخوان ران چپ بیشتر از ۱۰۵۴/۰/۱ می باشد.

- احتمال اصل به دست آمده برابر ۹۱۷/۰ است.
- اصل به دست آمده برای یازده رکورد صادق است.
- سطح معنی داری کمتر از ۰/۰۱ است.

## ۶ برنامه های آینده محققین

در این مقاله علاوه بر بحث بر روی کاربردهای داده کاوی در حمل و نقل، چگونگی برخورد روشهای داده کاوی با داده های تصادفات وسیله نقلیه نیز نشان داده شد، اهداف عمیق تری دیگری نیز برای مطالعه در کاربردهای داده کاوی در بحث حمل و نقل وجود دارد که در زیر به تعدادی از آنها اشاره می گردد:

- بررسی روشهای کارآمدتر برای سر و کار داشتن با بانکهای داده بزرگ با ابعاد بالا
- روش غلبه بر مدلهای Over-Fitting
- مدیریت داده های شلوغ و گمشده
- لفتن ارتباطات پیچیده میان فیلدها
- پاکسازی الگوهای غیر قابل فهم

- اصل صدمه به استخوان ران راست: اگر تعداد درهای وسیله نقلیه ۴ تا باشد و محافظ ماشین کمربند اینمنی خودکار باشد و صدمه دیده راننده باشد، آنگاه میزان صدمه به استخوان راست بیشتر از ۷۴۰/۹۲ است.
- احتمال اصل به دست آمده برابر ۹۲۹/۰ است.
- اصل به دست آمده برای سیزده رکورد صادق است.
- سطح معنی داری کمتر از ۰/۰۰۱ است.

- طفتن بهترین اثرات متقابل کاربر با اطلاعات پیشین
- امکان یکپارچگی با سایر سیستمها

## ۷ نتیجه گیری

همانطور که بیان گردید داده کاوی به طور گسترده‌ای برای استخراج الگوهای جالب برای حجم زیادی از داده‌ها تعریف می‌گردد. روش‌هایی که برای داده کاوی استفاده می‌گردد از قواعد منظم بسیاری machine learning و بازیابی اطلاعات می‌باشد، این بدان معناست که برای هر کاربرد داده شده احتمالاً تعداد زیادی روش داده کاوی برای یافتن الگوهای مناسب وجود دارد که یک آرایه مغلوتوش از ابزارهای داده کاوی برای هر کاربرد می‌باشد که می‌بایست با دقیقیت فراوان یک انتخاب دقیق داشت. نتایج به دست آمده نشان می‌دهد که داده کاوی در بخش حمل و نقل نیز یک ابزار موثر می‌باشد. حال با توجه به عدم کاربرد آن در کشور ما امید است که با مقالاتی از این دست بتوان علاوه بر تکمیل بانکهای موجود و تشکیل انبارهای داده بزرگ از داده‌های موجود و جمع آوری اطلاعات مورد نیازی که هم اکنون در دسترس نمی‌باشد و با استفاده از این روشها تغییرات چشمگیری را در بحث حمل و نقل که امروزه در شهرهای بزرگ گریبانگیر همگان می‌باشد داشت.

## مراجع

- [۱] استالینگ، ولیام، **اصول طراحی و ویژگیهای داخلی سیستمهای عامل**، ترجمه صدیقی مشکن‌نای، محسن، پدرام، حسین، ویراسته بزنگوب، محمود، ویرایش سوم، اصفهان، نشر شیخ بهایی، بهار ۱۳۸۰.
- [۲] رئیسی دهکردی، معلولین و سد معبرهای معماری و ساختمانی ، سازمان حمل و نقل و ترافیک تهران، چاپ اول، ۱۳۷۶.
- [۳] سازمان ملل متحده، **دقیق در طراحی** ، ترجمه مهدی بابائی اهری، سازمان مشاور فنی و مهندسی شهر تهران، چاپ اول، پاییز ۱۳۷۳.
- [۴] مهندسین مشاور معمار و شهرساز آدک، **مطالعات مرحله اول ایجاد تسهیلات حرکت ویژه معلولین شهر مشهد**، ۱۳۸۳.
- [۵] مراجع مدیری، مهدی، **اشاره‌ای به سیستم‌های اطلاعات جغرافیایی GIS**. سازمان جغرافیایی نیروهای مسلح، ۱۳۸۴.
- [6] Malaga, Ross A., *Information Systems Technology*, p.65, Prentice Hall, 2005
- [7] Date, Chris J., *An Introduction to Database Systems*, p.36, O'reilly, 1995