

تحلیل مکانی فاکتورهای مؤثر بر تصادفات جاده‌ای با استفاده از تکنیک‌های داده کاوی (Data Mining) و سیستم اطلاعات جغرافیایی (GIS)

(مطالعه موردی آزاد راه قزوین - زنجان)

مهدی کیخایی^۱

چکیده

تلفات و جراحات ناشی از تصادفات جاده‌ای و ترافیکی یکی از بزرگ‌ترین موضوعات چالش برانگیز در زندگی افراد جامعه و کشورهاست. در این پژوهش از الگوریتم‌های داده‌کاوی شامل: کلاسترینگ، قوانین انجمنی، درخت تصمیم CHAID و طبقه‌کننده Naïve Bayes به همراه تحلیل‌های GIS عوامل مکانی تاثیرگذار، به منظور کشف الگوها و پیش‌بینی فاکتورهای مؤثر بر شدت تصادفات جاده‌ای در آزادراه قزوین به زنجان استفاده گردید. در این مطالعه تعداد ۹۸۶ داده تصادف ثبت شده توسط پلیس راهور در دوره زمانی ۲۰۱۲ تا ۲۰۱۴ بکار گرفته شد. الگوریتم خوشه‌بندی X-Means به منظور خوشه‌بندی داده‌ها و همگن نمودن ویژگی‌های گوناگون تصادفات استفاده و داده‌ها در ۶ دسته جاسازی گردیدند. در ادامه با استفاده از الگوریتم‌های درخت تصمیم CHAID و Naïve Bayes Classifier، عملیات پیش‌بینی شدت تصادفات جاده‌ای و نیز فاکتورهای مهم و تاثیرگذار بر آن مشخص گردید. الگوریتم CHAID Decision Tree با دقت ۸۴٫۸٪ نسبت به Naïve Bayes با دقت ۸۴٫۶٪، وضعیت نسبتاً مناسبتری را در مدل‌سازی و پیش‌بینی داده‌های آزمون داشته است که منحنی ROC نیز صحت آن را نشان می‌دهد. همچنین مهمترین پیش‌بینی‌کننده‌ها و عوامل تاثیرگذار بر شدت تصادفات توسط الگوریتم CHAID DT به ترتیب کاهش اهمیت شامل: عوامل نوع خودرو (۰/۲۵)، کاربری‌های حاشیه‌ای جاده (۰/۱۶)، شرایط توپوگرافیکی محل (۰/۱۳)، عامل انسانی تصادف (۰/۱)، ماه وقوع حادثه (۰/۱)، شرایط آب و هوایی (۰/۰۸)، شرایط روشنایی و ساعت وقوع تصادف (۰/۰۷)، هندسه جاده (۰/۰۵) و شرایط سطح جاده (۰/۰۵) استخراج گردید.

واژگان کلیدی: تصادفات جاده‌ای، داده کاوی، خوشه‌بندی، قوانین انجمنی، درخت تصمیم، Naïve

Bayes Classifier

۱. (*نویسنده مسئول): مدرس گروه جغرافیا، دانشگاه امام علی (ع)، تهران، ایران.

مقدمه

تصادفات جاده‌ای موضوع اساسی در زندگی امروزه به شمار می‌روند (کاشیپ و چاندرا^۱، ۲۰۱۶: ۲۲۵). بر اساس گزارش‌های سازمان سلامت جهانی در حدود ۱,۲۵ میلیون از افراد جهان به‌طور سالانه در تصادفات جان خود را از دست می‌دهند و در حدود ۲۰ تا ۵۰ میلیون نیز دچار صدمات یا معلولیت‌های ناشی از آن می‌شوند. کشور ایران از نظر سوانح ترافیکی یکی از کشورهای دارای بیشترین موارد تصادف و مرگ و میر از طرف سازمان جهانی سلامت معرفی گردیده است. بر اساس گزارش سال ۲۰۱۵ این سازمان در سال ۲۰۱۴-۲۰۱۲ به ازای هر ۱۰۰۰۰۰ نفر نرخ تخمین تلفات جاده‌ای در ایران ۳۲,۱ نفر اعلام گردیده است (شکل ۱). اگر چه روند نرخ تصادفات در سالیان اخیر نزولی بوده است، در رده‌بندی جهانی در وضعیت نامناسبی قرار دارد (سازمان بهداشت جهانی^۲، ۲۰۱۵).

شناسایی موقعیت‌های مکانی که بیشترین تصادفات در آنجا رخ می‌دهد و مشخص کردن ویژگی‌های مختلف مرتبط با تصادفات ما را در شناخت و درک بهتر شرایط مختلف محل وقوع یاری می‌رساند. اگر چه تلفات جاده‌ای در طول سال‌ها منجر به مرگ و میر شده‌اند، بیشتر این سوانح قابل پیش‌بینی و پیشگیری می‌باشند (کومار و درگا^۳، ۲۰۱۶: ۱۰).

کشف دانش در پایگاه داده به فرایند غیر بدیهی شناسایی الگوهای معتبر، تازه، مفید و در نهایت قابل درک در مجموعه داده‌ها و توصیف آن به روش مختصر و معنی‌دار تعریف می‌گردد؛ به عبارتی دانش استخراج اطلاعات با ارزش از پایگاه داده‌های بزرگ است (هن^۴ و همکاران، ۲۰۱۱: ۸، آکومولافه^۵ و همکاران، ۲۰۱۲: ۳۰، جایابابو^۶، ۲۰۱۷: ۵۶۲).

^۱ Kashyap & Chandra

^۲ World Health Organization

^۳ Kumar & Durga

^۴ Han

^۵ Akomolafe

^۶ Jayababu

بنابر تحقیقات مجمع جهانی راه سهم تجمعی عامل انسان در وقوع تصادفات ۹۳٪، عامل وسیله نقلیه ۱۳٪ و عامل محیط راه ۳۴٪ است. عامل انسانی به تنهایی تقریباً ۷۰٪ از علل حوادث را در بر می‌گیرد؛ لذا در بیشتر تحقیقات انجام گرفته به این عامل توجه می‌گردد (دستورالعمل امنیت جاده‌ها^۱، ۲۰۱۳: ۳۵). توسعه روش‌های جدید برای جلوگیری از وقوع سوانح ترافیکی یا کاهش شدت آن در راه‌های برون شهری یکی از اهداف اصلی مطالعات ایمنی راه است (عفتی^۲ و همکاران، ۲۰۱۴: ۹).

با توجه به اینکه نقش عوامل مکانی نیز از فاکتورهای بسیار مهم و تاثیرگذار در این مقوله است، کشف ویژگی‌های مکانی خاص و تلفیق آن با سایر عوامل می‌تواند در بهبود ایمنی و کاهش سوانح در راه‌های برون شهری تأثیر به سزایی داشته باشد (عفتی و همکاران، ۲۰۱۴: ۱۱). دانش حاصل می‌تواند پشتیبانی بیشتری به‌منظور تصمیم‌گیری بهتر به‌منظور کمک به مردم در حمل و نقل هوشمند، امنیت و سلامت عمومی و پیش‌بینی نتایج آینده ایفا نماید (لی^۳ و همکاران، ۲۰۱۶: ۹۰).

چندین روش و تکنیک داده کاوی از قبیل کلاسترینگ، طبقه‌بندی و قوانین انجمنی به‌منظور تحلیل داده‌های تصادفات و ارائه راهکارهایی برای کاهش آن‌ها استفاده می‌گردد (سنمیگوئل^۴ و همکاران، ۲۰۱۵: ۵۱، مارتین^۵ و همکاران، ۲۰۱۴: ۱۷، تیواری^۶، ۲۰۱۷: ۱۶). الگوریتم قوانین انجمنی یک روش‌شناسی شناخته شده برای شناسایی ارتباطات قابل توجه بین داده‌های یک پایگاه داده است (لی و شرستا و هو^۷، ۲۰۱۷: ۱۲). آنالیز کلاسترینگ به‌عنوان یکی از تکنیک‌های داده کاوی است که برای دستیابی به اهداف مختلف استفاده

^۱ PIARC Road Safety Manual

^۲ Efatti

^۳ Li

^۴ Sanmiquel

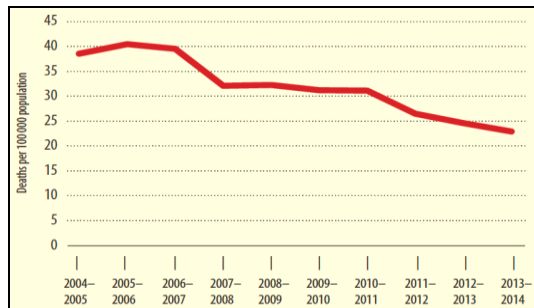
^۵ Martin

^۶ Tiwari

^۷ Li & Shrestha & Hu

می‌شود و می‌تواند به حذف برخی ناهمگونی‌ها در داده‌ها و قرار داده‌های همگون‌تر در کنارهم کمک نماید (کومار و توشنیوال^۱، ۲۰۱۵: ۱۳).

هدف از این پژوهش استفاده از الگوریتم‌های داده کاوی شامل کلاسترینگ، قوانین انجمنی، درخت تصمیم CHAID (Chi-squared Automatic Interaction Detector) و طبقه‌کننده Naïve Bayes به منظور کشف و شناسایی مکان‌های حادثه‌خیز، تعیین فاکتورهای مرتبط و مؤثر بر شدت تصادفات جاده‌ای و پیش‌بینی این سوانح در اتوبان قزوین- زنجان با در نظر گرفتن عوامل مکانی و غیر مکانی مرتبط با حادثه است. شناسایی حوزه مکانی و چگونگی تأثیر فاکتورهای مکانی در قطعات راه در کنار سایر عوامل در این مطالعه مورد بررسی قرار می‌گیرد. نتایج این مطالعه می‌تواند کارشناسان پلیس و متخصصان ایمنی راه را در گرفتن تصمیمات لازم به منظور کاهش تصادفات و سوانح جاده‌ای یاری رساند.



شکل ۱: نمودار روند تصادفات در ایران (WHO, 2015)

Source: Forensic Medicine Organization and National Statistical Center.

مبانی نظری و پیشینه

تصادفات حوادث غیرقابل پیش‌بینی می‌باشند که بررسی آن‌ها نیازمند شناخت فاکتورهای مختلفی است. هزینه‌های تلفات انسانی ناشی از این سوانح، تأثیر سوء زیادی بر جامعه دارد (کومار و توشنیوال^۲، ۲۰۱۵: ۱۴). لذا محققان زیادی به تعیین شناسایی عوامل تأثیرگذار بر شدت تصادفات جاده‌ای پرداخته‌اند.

^۱ Kumar & Toshniwal

^۲ Kumar & Toshniwal

اداشوا^۱ و همکاران در سال ۲۰۱۶ در تحقیق خود اثرات هندسه جاده و شرایط دیگر در تصادفات جاده‌ای در کشور اسپانیا در طول سه سال را به‌منظور طبقه‌بندی داده‌های مربوطه با استفاده از الگوریتم داده کاوی Random Forest بررسی نموده‌اند. هدف اصلی این مطالعه، تحلیل اثر هندسه جاده بر شدت تصادفات بوده و نتیجه‌گیری می‌کند که جاده‌های با عرض کم و فاقد شانه راه و همپنین شیب‌های تند از علل اصلی شدت تصادفات بوده است.

پتی^۲ و همکاران در سال ۲۰۱۷ در پژوهشی که در کشور ایتالیا انجام داده است از الگوریتم‌های معروف CHAID Decision Tree و Bayesian Network به‌منظور پیش‌بینی شدت تصادفات دوچرخه سواران استفاده نموده است. تعداد ۴۹۶۲۱ تصادف جاده‌ای در این پژوهش مورد بررسی قرار گرفته و عواملی از قبیل نوع برخورد، علائم حاشیه‌ای راه، سن و جنس دوچرخه سوار، فاکتورهای محیطی و عوامل مکانی و زمانی مدنظر قرار داده شده است. نتایج نشان می‌دهد جاده‌های روستایی که فاقد مسیر مشخص جهت عبور و مرور دوچرخه سواران و نیز موتورسواران می‌باشد یکی از علل مهم بروز سوانح جاده‌ای بوده و در بین فاکتورهای مؤثر نوع جاده، نوع برخورد و سن دوچرخه سواران در بین مهم‌ترین عوامل در شدت سوانح استخراج گردیده است.

کومار و توشنیوال در سال ۲۰۱۶ در تحقیق خود در کشور هند تعداد ۱۱۵۷۴ تصادف بین سال‌های ۲۰۰۹-۲۰۱۴ را به‌منظور تعیین الگوهای مختلف سوانح جاده‌ای بررسی نموده‌اند. در این بررسی از تکنیک‌های کلاسترینگ (k-mods) به‌منظور دسته‌بندی داده‌های تصادفات استفاده شده و سپس برای شناسایی شرایط مختلفی که با وقوع یک حادثه همراه است از الگوریتم قوانین انجمنی بهره گرفته شده است. با ترکیب کلاسترینگ و کاوش قوانین انجمنی، ارتباطات عوامل مختلف که در بین داده‌ها مستتر می‌باشد به راحتی آشکار می‌گردد.

¹ Dadashova

² Pati

سین و کاشیاپ^۱ در سال ۲۰۱۶ تحقیق خود را در استان Uttar pradegh کشور هند که از نرخ بالایی در تصادفات جاده‌ای برخوردار است معطوف ساخته است. در این پژوهش فناوری داده کاوی برای ارتباط بین ویژگی‌های ثبت شده جاده و شدت تصادفات و مجموعه‌ای از قوانینی که توسط مراکز دولتی هند برای بهبود ایمنی استفاده می‌کنند بکار گرفته شده و از الگوریتم طبقه‌بندی Naïve Bayes classifier استفاده شده که صحت دسته‌بندی با میزان ۸۹٫۴۵٪ حاصل گردیده است.

تیواری در سال ۲۰۱۷ در پژوهش خود به‌منظور آنالیز داده‌های تصادفات از تکنیک‌های داده کاوی شامل SOM، K-modes، Naïve Bayes و Decision Tree استفاده کرد و نتیجه گرفت که بهترین طبقه‌بندی پس از عملیات بخش‌بندی^۲ داده‌ها با روش‌های خوشه‌بندی حاصل می‌گردد. به‌منظور بررسی این موضوع تعداد ۱۳۰۶۲ تصادف بین سال‌های ۲۰۱۵-۲۰۱۱ استفاده و عوامل مهمی از قبیل تعداد وسایل نقلیه، زمان تصادف، سطح راه، شرایط جوی، شرایط روشنایی، جنسیت و سن رانندگان و روز و ماه وقوع تصادفات در نظر گرفته شد.

داده‌ها، ابزار و روش تحقیق

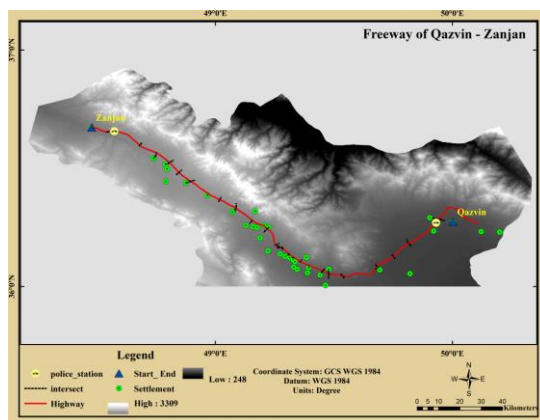
بزرگراه مورد مطالعه

آزاد راه مورد مطالعه در این پژوهش آزاد راه قزوین به زنجان است که قسمتی از آزاد راه اتصالی شهر تهران به تبریز محسوب می‌گردد. محدوده ارتفاعی این آزاد راه کوهستانی بین ۱۲۶۰ تا ۱۷۸۰ متر متغیر است. این محور یک طرفه به طول ۱۶۵ کیلومتر از مختصات جغرافیایی $36^{\circ} 16' N$ و $49^{\circ} 56' E$ در شهر قزوین شروع و به مختصات جغرافیایی $36^{\circ} 39' N$ و $48^{\circ} 33' E$ در شهر زنجان ختم می‌گردد و از راه‌های اصلی برون شهری پرخطر محسوب می‌گردد. شکل ۲ نقشه ارتفاعی این منطقه را نشان می‌دهد.

^۱ Kashyap & Singh

^۲ segmentation

تحلیل مکانی فاکتورهای مؤثر بر تصادفات جاده‌ای با استفاده از تکنیک‌های داده کاوی ... / ۵۵



شکل ۲: منطقه مورد مطالعه (بزرگراه قزوین - زنجان)

داده‌ها و ابزار تحقیق

در این پژوهش از تعداد ۹۸۶ تصادف ثبت شده در سال‌های ۲۰۱۴-۲۰۱۲ در این آزاد راه که توسط پلیس راه برداشت گردیده است استفاده گردید. این داده‌ها شامل اطلاعاتی از قبیل: شدت و پیامد تصادف، شرایط روشنایی، شرایط آب و هوایی، نوع توپوگرافی، نوع رویه سطح راه، نوع کاربری‌های حاشیه راه، شرایط سطح راه، عوامل انسانی تصادف، نوع خودرو، سن و جنسیت راننده، موقعیت تصادف، ساعت، روز و ماه تصادف است. جدول ۱ آمارهای مربوطه را نشان می‌دهد. به منظور انجام تحلیل‌های دقیق‌تر، داده‌های مکانی تصادفات شامل: طرح هندسی راه، کاربری‌ها، اماکن حاشیه‌ای راه، مراکز جمعیتی و تقاطع‌ها نیز استفاده گردیده‌اند.

در این تحقیق از نرم افزارهای Rapid Miner 7.6، IBM SPSS Modeler 14.2 و Arc GIS 10.5 استفاده گردیده است.

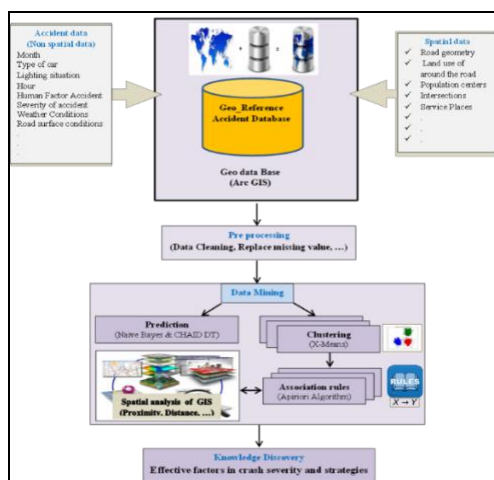
روش اجرای تحقیق

روش پیشنهادی تحقیق بر اساس تلفیق تحلیل‌های مکانی، خوشه‌بندی، قوانین انجمنی، درخت تصمیم CHAID و طبقه‌بندی کننده Naïve Bayes جهت تحلیل شدت تصادفات

و شناخت فاکتورهای مؤثر در آن و ارتباطات متغیرهای مختلف استوار است. شکل ۳ مدل مفهومی این پژوهش را نشان می‌دهد. با مشخص نمودن موقعیت تصادفات بر روی محور راه و الصاق ویژگی‌های مکانی و غیر مکانی به داده‌های جمع‌آوری شده عوامل تاثیرگذار و ارتباطات آن‌ها بر شدت تصادفات بررسی می‌گردد.

R	Variable	Type	Description
1	Crash severity	Nominal	Injured & Deceased
2	Weather Conditions	Nominal	Clear & Foggy & Cloudy & Rainy & Snowy
3	Lighting conditions	Nominal	Sunrise & Daylight & Sunset & Night
4	Topographic conditions	Nominal	Flat & Mountain
5	Marginal Land use	Nominal	Agriculture & Commercial & Educational & Industrial & Non – Residential & Recreational & Others
6	Road surface conditions	Nominal	Dry & Frost & Wet
7	Road geometry	Nominal	Straight, flat & Screw, flat & Screw, up / down &
8	Human Factor Accident	Nominal	Failure to comply with priority & Fatigue and drowsiness & Hurry up and accelerate & Ignoring the rules & Lack of familiarity with the road
9	Vehicle manufacturer	Nominal	Benz bus & Benz truck & Citroen & Daewoo & FAW & Ferguson & HOWO & Zamyad & Volvo truck & Volvo Bus & Toyota & Tiba & Suzuki & Sepand & Pride & Proton & Man & Mitsubishi & Motorcycle & Muso & MVM & Nissan & Samand & Rio & Renault & Peugeot &
10	Age	Continuous	Driver's age
11	Gender	Nominal	Female & Male
12	Solar month	Continuous	Farvardin & Ordibehesht & Khordad & Tir & Mordad & Shahrivar & Mehr & Aban & Azar & Dey & Bahman & Esfand
13	Hour	Continuous	00 - 2400
14	Distance _ Accident	Continuous	Distance of accident location from Qazvin - Zanjan road police station

جدول ۱: ویژگی داده‌های تصادفات



شکل ۳: مدل مفهومی

خوشه‌بندی

خوشه‌بندی تقسیم داده‌ها به گروه‌های اشیاء مشابه است. نمایش داده‌ها با خوشه‌های کمتر اگر چه بعضی از جزئیات را نادیده می‌گیرد ولی باعث تحلیل‌های ساده‌تر می‌گردد (برخین^۱، ۲۰۰۶: ۲۶، مریست و چارتیر^۲، ۲۰۱۳: ۱۸).

خوشه‌بندی یک الگوریتم داده کاوی غیر نظارت شده است. یکی از الگوریتم‌های محبوب خوشه‌بندی K-Means بوده که داده‌ها را به K خوشه تقسیم‌بندی می‌نماید. با توجه به ماهیت، نوع و تعداد داده‌ها روش‌های مختلفی جهت کلاسترینگ موجود است (تیواری و همکاران، ۲۰۱۷: ۴۱۰، کومار و توشنیوال، ۲۰۱۶: ۱۷). تعیین تعداد خوشه‌ها یکی از مهمترین مشکلاتی است که در بسیاری از الگوریتم‌های خوشه‌بندی نظیر K-Means وجود دارد (میزوتانی و کانای^۳، ۲۰۱۷: ۸، پلج و مور^۴، ۲۰۰۰: ۱۷، هاسن و یو^۵، ۲۰۰۱: ۱۳). کلاسترینگ مکانی نیز یک تکنیک مهم برای داده کاوی مکانی به شمار می‌رود و هدف آن

^۱ Berkhin

^۲ Morissette & Chartier

^۳ Mizutani & Kanai

^۴ Pelleg & Moore

^۵ Hasen & Yu

بخش‌بندی داده‌های مکانی به تعدادی خوشه به نام Spatial Cluster می‌باشد (جایابابو^۱ و همکاران، ۲۰۱۷: ۶).

رویکردهای مبتنی بر Bayes Information Criteria (BIC) و Minimum Description Length (MDL) Akiake Information Criterion (AIC) به منظور تعیین تعداد بهینه خوشه‌ها به کار می‌رود (جاین^۲، ۲۰۱۰: ۶۵۹).

الگوریتم خوشه‌بندی X-Means یکی از الگوریتم‌های خوشه‌بندی است که خودش تلاش می‌کند که بهترین تعداد خوشه‌ها را تشخیص دهد. این الگوریتم بر اساس BIC روند خوشه‌بندی را انجام می‌دهد و یکی از ضعف‌های روش K-Means را می‌تواند پوشش دهد (برخین، ۲۰۰۶: ۲۳).

$$\text{BIC}(M_j) = L_j(D) - \frac{P_j}{2} * \log R \quad \text{رابطه ۱}$$

Where $L_j(D)$ is the Log-Likelihood of the data according to the j -th model and taken at the maximum-Likelihood point.

P_j is number of parameters at M_j

الگوریتم X-Means می‌تواند برای مجموعه داده‌های کوچک استفاده شود (پلج و مور، ۲۰۰۰: ۱۷). در این پژوهش به منظور اجرای خوشه‌بندی از الگوریتم X-Means استفاده گردیده است.

کاوش قوانین انجمنی

الگوریتم قوانین انجمنی یک متدلوژی مناسب به منظور شناسایی ارتباطات مهم بین داده‌های ذخیره شده در یک پایگاه داده است (لی و همکاران، ۲۰۱۷: ۲۵). قوانین انجمنی یک مجموعه از قوانینی که بر اساس الگوهای خاصی که بین داده‌ها است را کشف می‌کند. قانون $A \rightarrow B$ نشان می‌دهد که اگر A اتفاق بیفتد آنگاه B رخ خواهد داد (کومار و توشنیبال، ۲۰۱۵: ۱۱، شتی و همکاران، ۲۰۱۷: ۱۴). در قوانین انجمنی یک قانون به عنوان

^۱ Jayababu

^۲ Jain

یک قانون قوی در نظر گرفته می‌شود اگر دارای Confidence و Support بالایی باشد (جایابابو و همکاران، ۲۰۱۷: ۵۶۸).

Support یک قانون نشان می‌دهد که به چه میزانی تراکنش‌های A و B با هم رخ می‌دهند (رابطه ۲).

$$\text{Confidence Support} = p(A \cap B) \quad \text{رابطه ۲}$$

یک قانون نیز نشان می‌دهد که در C% از تراکنش‌ها وقتی A رخ می‌دهد آنگاه B نیز اتفاق می‌افتد و از طریق رابطه ۳ محاسبه می‌شود.

رابطه ۳

$P(A|B) = \frac{P(A \cap B)}{P(B)}$ یکی دیگر از معیارهایی که برای یک قانون اندازه‌گیری می‌شود Lift می‌باشد که از طریق رابطه ۴ محاسبه می‌شود. اگر مقدار Lift در یک قانون $Lift > 1$ باشد این قانون می‌تواند با ارزش تلقی گردد.

$$Lift = \frac{P(A \cap B)}{P(A)P(B)} \quad \text{رابطه ۴}$$

در این مقاله جهت کشف قوانین انجمنی بین مجموعه داده‌ها از الگوریتم Apriori استفاده گردیده است.

دو مرحله مهم Association Rule Mining تولید آیت‌مست‌های تکراری و تولید قوانین انجمنی هستند. در مرحله اول هر کدام از ارزش‌های ویژگی‌ها که تعداد بیشتری در مجموعه داده‌ها دارند به‌عنوان آیت‌مست تلقی می‌گردند و در مرحله بعد قوانین انجمنی از آیت‌مست‌های تکراری تولید می‌شود که به آن Confidence نیز می‌گویند (جایابابو و همکاران، ۲۰۱۷: ۵۶۷).

طبقه‌بندی کننده بیزین

این طبقه‌بندی کننده یک الگوریتم یادگیری نظارت شده به شمار می‌رود که از روش‌های آماری و احتمالاتی به‌منظور طبقه‌بندی داده‌ها استفاده می‌کند (گپتا^۱ و همکاران، ۲۰۱۶: ۱۵، آتافو

^۱ Gppta

و کومار^۱، ۲۰۱۷: ۱۸). این روش طبقه‌بندی بر اساس تئوری بیز (رابطه ۵) و فرضیه حداکثر احتمال (Maximum Posteriori) عمل می‌نماید.

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \quad \text{رابطه ۵}$$

از مزایای این الگوریتم سادگی و سرعت مناسب آن در مجموعه داده‌های بزرگ (احمد^۲ و همکاران، ۲۰۱۷: ۱۱، لونگ^۳، ۲۰۰۷: ۱۷).

و نیز نیاز به داده‌های کم جهت آموزش و ارزیابی پارامترها است (تیواری و همکاران، ۲۰۱۷: ۱۷). الگوریتم Naïve Bayes متغیرهای گسسته و پیوسته و مسائل غیر خطی را پوشش داده (شینده و پال^۴، ۲۰۱۷: ۱۶) و در موقعیت‌های جهان واقعی به خوبی کار می‌کند (چاراسیا^۵ و همکاران، ۲۰۱۴: ۲۷).

درخت تصمیم

Decision Tree یکی از قوی‌ترین تکنیک‌های داده کاوی و دسته‌بندی داده‌ها محسوب می‌گردد و الگوهای مختلفی از قبیل Gini، Entropy، Logworth، CART و CHAID به‌منظور تولید درخت تصمیم موجود است (ساپری^۶ و همکاران، ۲۰۱۷: ۶۵). یک درخت تصمیم مجموعه داده‌ها را به زیرگروه‌هایی بر اساس ارتباطات بین متغیرهای ورودی تقسیم و جدا سازی می‌کند (پراتی^۷ و همکاران، ۲۰۱۷: ۱۴). هدف از الگوریتم درخت تصمیم تولید یک الگو برای پیش‌بینی ارزش‌های متغیر هدف بر اساس ارزش‌های ورودی است. مجموعه داده‌های آموزشی به‌منظور تولید درخت و مجموعه داده‌های تست نیز به‌منظور ارزیابی دقت درخت تصمیم استفاده می‌شود (ماهاجان و کار^۸، ۲۰۱۶: ۱۷).

¹ Atnafu & Kaur

² Ahmed

³ Leung

⁴ Shinde & Pal

⁵ Chaurasia

⁶ Sapri

⁷ Prati

⁸ Mahajan & Kaur

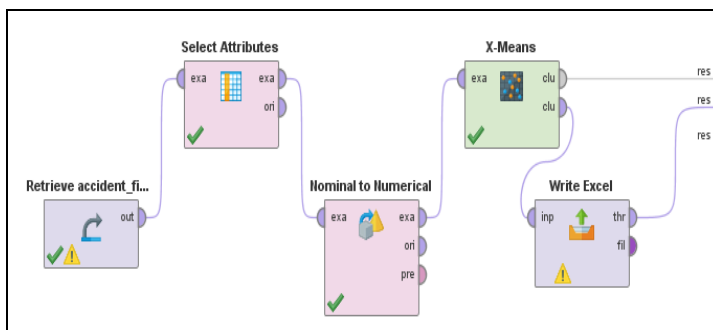
درخت تصمیم CHAID یکی از تکنیک‌های طبقه‌بندی قانون مبنا است که قوانین بین داده‌ها را به صورت ساختار درختی ارائه داده و در نهایت مقدار ارزش Accuracy بین متغیرها را نمایش می‌دهد (پراتی و همکاران، ۲۰۱۷: ۴۸). این تکنیک در آفریقای جنوبی توسعه یافت و در سال ۱۹۸۰ توسط Gordon V. Kass منتشر شد. در این پژوهش به منظور تشکیل درخت تصمیم از الگوریتم CHAID استفاده شده است.

نتایج و بحث

تحلیل خوشه‌بندی

نیازمندی اصلی برای آنالیز خوشه‌بندی تعیین تعداد خوشه‌ها است. در این پژوهش از الگوریتم X-Means برای خوشه‌بندی داده‌ها استفاده گردیده است. این الگوریتم بر اساس BIC (فرالی و رافتری^۱، ۱۹۹۸: ۵۸۰) بهترین تعداد خوشه‌ها بین داده‌های موجود را مشخص می‌کند.

بر اساس این الگوریتم کلیه داده‌های تصادفات به ۶ کلاستر خوشه‌بندی گردید. شکل ۴ روند اجرا را در نرم افزار Rapid Miner و جدول ۲ و ۳ نتایج حاصل از این خوشه‌بندی را نشان می‌دهد. جدول شماره ۲ مراکز خوشه‌ها، تعداد آبجکت‌های هر خوشه و جدول ۳ وضعیت هر تصادف و تعلق آن به خوشه مربوطه را مشخص نموده است.



شکل ۴: عملیات کلاسترینگ ((X-Means))

^۱ Fraley & Raftery

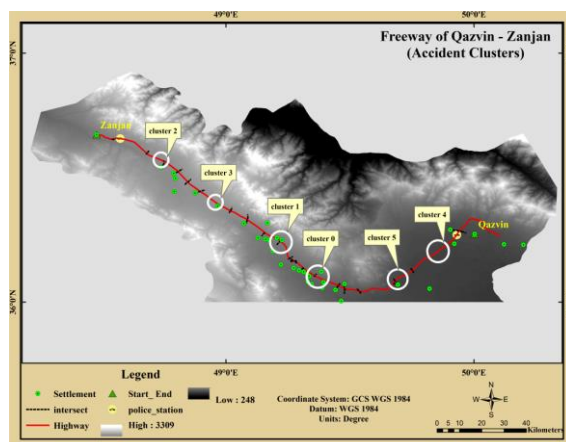
جدول ۲: کلاسترینگ به روش (X means)

Attribute	Cluster0	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Crash severity = Deceased	0.205	0.221	0.115	0.181	0.141	0.121
Crash severity = Injured	0.795	0.779	0.885	0.819	0.859	0.879
Weather Conditions = rainy	0.076	0.030	0.034	0	0.186	0.085
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
Solar month	5.99	6.61	6.52	6.74	6.48	5.93
Crash distance	61013	85804	143000	114578	7780	34035
Total	219	226	87	88	220	146

جدول ۳: خوشه‌ها

Row No	cluster	Crash severity = Deceased	Crash severity = Injured	Weather Conditions = rainy	000	000	Solar month	Crash distance
1	Cluster0	1	0	1	000	000	1	64000
2	Cluster1	1	0	0	000	000	5	96000
0	0	0	0	0	000	000	0	0
0	0	0	0	0	000	000	0	0
0	0	0	0	0	000	000	0	0
986	Cluster4	0	1	0	000	000	11	2000

با توجه به خوشه‌بندی انجام گرفته و نیز ویژگی فاصله محل وقوع حادثه از ایستگاه پلیس، هر خوشه روی آزادراه مطابق شکل نمایش داده شده است. جدول ۴ توصیفات کلی هر خوشه را بیان می‌نماید.



شکل ۵: خوشه تصادفات

جدول ۴: اندازه و شرح خوشه‌ها

Cluster number	Cluster description	Size (%)
0	در فاصله ۶۱۰۱۳ متری از پلیس راه قرار داشته و تراکم مراکز صنعتی و جمعیتی در حاشیه راه و تقاطع‌ها بالا است. تعداد ۲۱۹ تصادف با ۴۶ کشته و ۱۷۴ زخمی را در ثبت نموده است.	22.9
1	در فاصله ۸۵۸۰۴ متری از پلیس راه قرار داشته و تراکم مراکز جمعیتی در حاشیه راه و تقاطع‌ها نسبتاً بالا بوده و تعداد ۲۲۶ تصادف با ۵۰ کشته و ۱۷۶ زخمی را شامل می‌شود.	22.2
2	در فاصله ۱۴۳۰۰ متری از پلیس راه قرار داشته و تراکم مراکز صنعتی و جمعیتی در حاشیه راه و تقاطع‌ها نسبت به سایر خوشه‌ها کمتر بوده و تعداد ۸۷ تصادف با ۱۰ کشته و ۷۷ زخمی را در بردارد.	8.8
3	در فاصله ۱۱۴۵۷۸ متری از پلیس راه قرار داشته و تراکم مراکز صنعتی و جمعیتی در حاشیه راه و تقاطع‌ها کمتر بوده و تعداد ۸۸ تصادف با ۱۶ کشته و ۷۲ زخمی را شامل می‌شود.	8.9
4	در فاصله ۷۷۸۰ متری از پلیس راه قرار داشته و تراکم مراکز جمعیتی در حاشیه راه بالا بوده و نزدیک‌ترین خوشه به پلیس راه محسوب گردیده و هندسه جاده مطلوب نیست و تعداد ۲۲۰ تصادف با ۳۱ کشته و ۱۸۹ زخمی را شامل می‌شود.	22.3
5	در فاصله ۳۴۰۳۵ متری از پلیس راه قرار داشته و تراکم مراکز جمعیتی در حاشیه راه بالا بوده و وضعیت هندسه راه نیز قابل تأمل است. تعداد ۱۴۶ تصادف با ۱۸ کشته و ۱۲۸ زخمی را در بردارد.	14.8

کشف قوانین انجمنی

به‌منظور تولید قوانین انجمنی از الگوریتم Apriori با حداقل Support ۱۵٪ برای هر خوشه و نیز خوشه کل داده‌های تصادفات استفاده گردیده و با استفاده از ارزش Confidence، هر قانون ارزیابی می‌گردد. تعداد ۱۲ قانون برتر برای هر خوشه انتخاب و تحلیل می‌گردد. جدول ۵ این قوانین را برای هر خوشه نشان داده است.

جدول ۵: قوانین انجمنی در خوشه‌ها

R	Consequent	Antecedent	Support %	Confidence %
Cluster 0				
1	Crash severity = Injured	Lighting conditions = daylight and Marginal Landuse = Non-Residential and Human Factor Accident = Hurry up and accelerate and Weather Conditions = clear and Road surface conditions = Dry	18.7	100.0
2	Crash severity = Injured	Solar month = Shahriyar and Weather Conditions = clear and Road surface conditions = Dry and Road geometry = Straight, flat	15.06	96.9
3	Human Factor Accident = Hurry up and accelerate	Lighting conditions = night and Marginal Landuse = Non-Residential and Topographic conditions = flat and Road geometry = Straight, flat	18.8	92.6

4	Crash severity = Injured	Vehicle manufacturer = Peugeot and Weather Conditions = clear and Road surface conditions = Dry and Road geometry = Straight, flat	24.6	90.7
5	Crash severity = Injured	Vehicle manufacturer = Pride and Human Factor Accident = Hurry up and accelerate and Road geometry = Straight, flat	19.6	90.6
6	Crash severity = Injured	Vehicle manufacturer = Peugeot	29.6	86.1
7	Crash severity = Injured	Human Factor Accident = Fatigue and drowsiness and Road surface conditions = Dry and Road geometry = Straight, flat	18.7	82.9
8	Crash severity = Injured	Vehicle manufacturer = Pride	31.05	82.3
9	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate and Road geometry = Straight, flat	64.3	79.4
11	Crash severity = Injured	Weather Conditions = clear	81.7	79.3
12	Crash severity = Injured	Weather Conditions = clear and Road geometry = Straight, flat	81.2	79.2
13	Crash severity = Injured	Lighting conditions = Sunrise	21.4	74.6
14	Crash severity = Deceased	Lighting conditions = night and Marginal Landuse = Non-Residential and Human Factor Accident = Hurry up and accelerate and Road surface conditions = Dry	15.5	47.05
Cluster 1				
15	Crash severity = Injured	Vehicle manufacturer = Pride and Human Factor Accident = Hurry up and accelerate	16.8	94.7
16	Crash severity = Injured	Vehicle manufacturer = Pride and Human Factor Accident = Hurry up and accelerate and Topographic conditions = flat	15.9	94.4
17	Crash severity = Injured	Vehicle manufacturer = Pride and Human Factor Accident = Hurry up and accelerate and Road geometry = Straight, flat and Topographic conditions = flat	15.5	94.3
18	Crash severity = Injured	Marginal Landuse = Agriculture and Human Factor Accident = Hurry up and accelerate and Weather Conditions = clear and Road surface conditions = Dry	15.0	94.1
19	Crash severity = Injured	Vehicle manufacturer = Pride and Weather Conditions = clear and Road surface conditions = Dry and Topographic conditions = flat	21.2	89.6
20	Human Factor Accident = Hurry up and accelerate	Marginal Landuse = Non-Residential and Weather Conditions = clear and Road surface conditions = Dry and Topographic conditions = flat	37.6	88.2
21	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate	61.06	82.6
22	Crash severity = Injured	Vehicle manufacturer = Peugeot	27.4	82.2

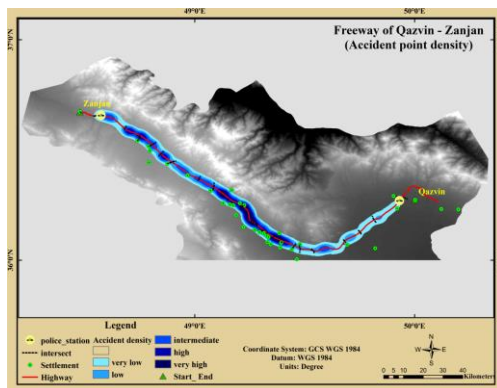
23	Crash severity = Injured	Solar month = Shahrivar and Weather Conditions = clear and Road surface conditions = Dry	16.3	78.3
24	Crash severity = Injured	Lighting conditions = Sunset	16.8	76.3
25	Crash severity = Injured	Road geometry = Straight, flat	90.3	75.5
26	Crash severity = Injured	Road geometry = Straight, flat	90.3	75.5
27	Crash severity = Injured	Lighting conditions = Sunrise	21.2	62.5
28	Crash severity = Injured	Human Factor Accident = Fatigue and drowsiness and Road surface conditions = Dry and Road geometry = Straight, flat	23.4	58.5
Cluster 2				
29	Crash severity = Injured	Vehicle manufacturer = Pride and Weather Conditions = clear and Topographic conditions = flat	22.9	100.0
30	Crash severity = Injured	Human Factor Accident = Fatigue and drowsiness and Weather Conditions = clear and Road surface conditions = Dry	25.2	95.4
31	Crash severity = Injured	Vehicle manufacturer = Peugeot and Weather Conditions = clear and Road geometry = Straight, flat	22.9	95.0
32	Crash severity = Injured	Lighting conditions = Sunrise and Marginal Landuse = Agriculture and Weather Conditions = clear and Road surface conditions = Dry and Road geometry = Straight, flat	20.6	94.4
33	Crash severity = Injured	Lighting conditions = Sunrise and Topographic conditions = flat	18.3	93.7
34	Crash severity = Injured	Vehicle manufacturer = Peugeot	27.5	91.6
35	Crash severity = Injured	Weather Conditions = clear and Road surface conditions = Dry	87.3	89.47
36	Crash severity = Injured	Human Factor Accident = Fatigue and drowsiness	32.18	89.2
37	Crash severity = Injured	Human Factor Accident = Fatigue and drowsiness and Topographic conditions = flat	32.18	89.28
38	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate	31.03	88.8
39	Crash severity = Injured	Weather Conditions = clear and Road surface conditions = Dry and Road geometry = Straight, flat and Topographic conditions = flat	79.3	88.4
40	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate and Weather Conditions = clear and Road surface conditions = Dry	28.7	88.0
41	Crash severity = Injured	Topographic conditions = flat	95.4	87.9
42	Crash severity = Injured	Lighting conditions = daylight	44.8	84.6
Cluster 3				

43	Crash severity = Injured	Lighting conditions = Sunset	15.9	100.0
44	Crash severity = Injured	Vehicle manufacturer = Pride	22.7	100.0
45	Crash severity = Injured	Vehicle manufacturer = Pride and Marginal Landuse = Agriculture and Topographic conditions = flat	19.3	100.0
46	Crash severity = Injured	Lighting conditions = daylight and Weather Conditions = clear and Road surface conditions = Dry	38.6	91.1
47	Crash severity = Injured	Human Factor Accident = Ignoring the rules	36.3	90.6
48	Crash severity = Injured	Topographic conditions = flat	98.8	81.6
49	Crash severity = Injured	Road surface conditions = Dry	96.5	81.1
50	Crash severity = Injured	Weather Conditions = clear and Road surface conditions = Dry and Road geometry = Straight, flat and Topographic conditions = flat	90.9	80.0
51	Crash severity = Injured	Marginal Landuse = Agriculture and Road geometry = Straight, flat and Topographic conditions = flat	85.2	78.6
52	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate	25.0	77.27
53	Crash severity = Injured	Human Factor Accident = Fatigue and drowsiness and Weather Conditions = clear and Road surface conditions = Dry and Road geometry = Straight, flat and Topographic conditions = flat	35.2	74.1
54	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate and Weather Conditions = clear and Road surface conditions = Dry and Road geometry = Straight, flat and Topographic conditions = flat	20.4	72.2
55	Crash severity = Deceased	Lighting conditions = Sunrise and Marginal Landuse = Agriculture and Topographic conditions = flat	21.5	52.6
56	Crash severity = Deceased	Lighting conditions = Sunrise	23.8	47.6
Cluster 4				
57	Crash severity = Injured	Lighting conditions = night and Road surface conditions = Dry and Human Factor Accident = Hurry up and accelerate and Road geometry = Straight, flat	16.	100.0
58	Crash severity = Injured	Vehicle manufacturer = Pride and Road surface conditions = Dry	25.0	98.1
59	Crash severity = Injured	Vehicle manufacturer = Pride and Road surface conditions = Dry and Topographic conditions = flat	22.7	98.0
60	Crash severity = Injured	Weather Conditions = clear and Road surface conditions = Dry and Road geometry = Straight, flat	63.6	97.8
61	Crash severity = Injured	Marginal Landuse = Non-Residential and Road surface conditions = Dry and Human Factor	41.4	97.8

Accident = Hurry up and accelerate				
62	Crash severity = Injured	Road surface conditions = Dry	80.9	97.7
63	Crash severity = Injured	Weather Conditions = clear and Human Factor Accident = Hurry up and accelerate and Road geometry = Straight, flat	52.7	97.4
64	Crash severity = Injured	Vehicle manufacturer = Peugeot	24.5	92.5
65	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate	84.5	84.4
66	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate and Road geometry = Straight, flat and Topographic conditions = flat	75.0	83.03
67	Crash severity = Deceased	Weather Conditions = rainy and Marginal Landuse = Non-Residential and Human Factor Accident = Hurry up and accelerate	15.0	75.7
68	Crash severity = Injured	Lighting conditions = daylight	38.6	75.3
69	Crash severity = Deceased	Road surface conditions = wet and Weather Conditions = rainy	17.2	71.05
70	Crash severity = Deceased	Weather Conditions = rainy and Topographic conditions = flat	18.1	67.5
Cluster 5				
71	Crash severity = Injured	Vehicle manufacturer = Peugeot and Human Factor Accident = Hurry up and accelerate and Road geometry = Straight, flat	19.1	100.0
72	Crash severity = Injured	Lighting conditions = night and Weather Conditions = clear and Road geometry = Straight, flat and Marginal Landuse = Non-Residential	19.8	93.1
73	Crash severity = Injured	Vehicle manufacturer = Peugeot	23.2	91.1
74	Crash severity = Injured	Vehicle manufacturer = Peugeot and Human Factor Accident = Hurry up and accelerate	22.6	90.9
75	Crash severity = Injured	Vehicle manufacturer = Pride and Weather Conditions = clear and Human Factor Accident = Hurry up and accelerate and Topographic conditions = flat	30.1	90.9
76	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate and Road geometry = Straight, flat and Marginal Landuse = Non-Residential and Topographic conditions = flat	80.8	90.6
77	Crash severity = Injured	Lighting conditions = night	29.4	88.3
78	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate	90.4	87.8
79	Crash severity = Injured	Weather Conditions = clear and Road surface conditions = Dry and Human Factor Accident = Hurry up and accelerate and Topographic conditions = flat	67.8	87.8

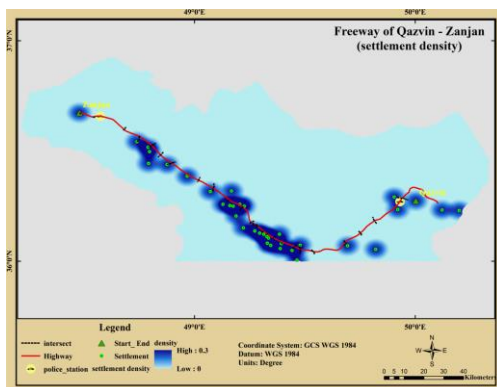
80	Crash severity = Injured	Vehicle manufacturer = Pride and Weather Conditions = clear and Road surface conditions = Dry and Marginal Landuse = Non-Residential and Topographic conditions = flat	32.1	87.2
81	Crash severity = Injured	Lighting conditions = night and Human Factor Accident = Hurry up and accelerate and Road geometry = Straight, flat and Marginal Landuse = Non-Residential and Topographic conditions = flat	26.02	86.8
82	Crash severity = Injured	Lighting conditions = daylight and Human Factor Accident = Hurry up and accelerate	41.7	85.2
83	Crash severity = Injured	Lighting conditions = daylight	47.2	84.05
84	Crash severity = Injured	Vehicle manufacturer = Pride and Road surface conditions = Dry and Human Factor Accident = Hurry up and accelerate Total Data	32.8	83.3
85	Crash severity = Injured	Vehicle manufacturer = Pride and Weather Conditions = clear and Road surface conditions = Dry and Topographic conditions = flat	23.7	90.5
86	Crash severity = Injured	Vehicle manufacturer = Pride and Road surface conditions = Dry and Topographic conditions = flat	25.3	90.0
87	Crash severity = Injured	Vehicle manufacturer = Pride and Road surface conditions = Dry	26.5	89.3
88	Crash severity = Injured	Lighting conditions = daylight and Weather Conditions = clear and Road surface conditions = Dry and Road geometry = Straight, flat	31.4	88.06
89	Crash severity = Injured	Vehicle manufacturer = Pride	29.6	88.01
90	Crash severity = Injured	Lighting conditions = daylight and Weather Conditions = clear and Road surface conditions = Dry and Road geometry = Straight, flat and Topographic conditions = flat	30.1	87.8
91	Crash severity = Injured	Vehicle manufacturer = Peugeot	26.4	86.9
92	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate and Road surface conditions = Dry	55.9	86.4
93	Crash severity = Injured	Weather Conditions = clear and Road surface conditions = Dry	80.8	84.3
94	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate	65.6	83.6
95	Crash severity = Injured	Weather Conditions = clear and Road geometry = Straight, flat and Topographic conditions = flat	73.2	83.5
96	Crash severity = Injured	Human Factor Accident = Hurry up and accelerate and Road geometry = Straight, flat and Topographic conditions = flat	58.8	82.7
97	Crash severity = Injured	Topographic conditions = flat	94.6	82.3
98	Crash severity = Injured	Marginal Landuse = Non-Residential and Human Factor Accident = Hurry up and accelerate and Topographic conditions = flat	46.9	80.5

ذکر ویژگی‌های مکانی راه در شدت تصادفات و خسارات بسیار حائز اهمیت است. عواملی از قبیل تقاطع‌ها، پیچ‌ها، تراکم مراکز صنعتی، خدماتی و جمعیتی در حاشیه راه از علل تشدید تصادفات در بسیاری از موارد است. در این پژوهش به‌منظور تحلیل قوانین استخراجی از داده‌های تصادفات به عوامل مکانی نیز توجه می‌گردد. شکل ۶ تراکم تصادفات در طول محور مورد مطالعه را در GIS نشان می‌دهد.

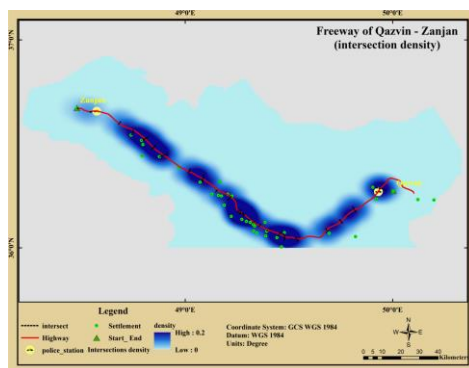


شکل ۶: تراکم تصادفات

کلیه ویژگی‌های مکانی و غیرمکانی به محل وقوع حادثه در GIS به هم الصاق گردیده و از تحلیل‌های مجاورت و فاصله مراکز خوشه‌ها با عوامل مکانی ذکرشده استفاده می‌گردد. شکل ۷ تراکم مراکز جمعیتی حاشیه آزاد راه قزوین - زنجان و شکل ۸ میزان تراکم تقاطع‌های موجود در طول آزادراه را نمایش می‌دهد.

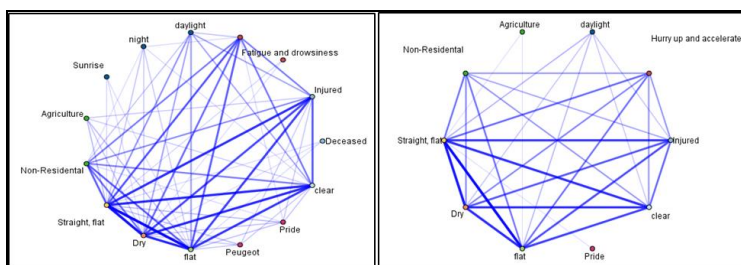


شکل ۷: تراکم سکونتگاهها



شکل ۸: تراکم تقاطع‌ها

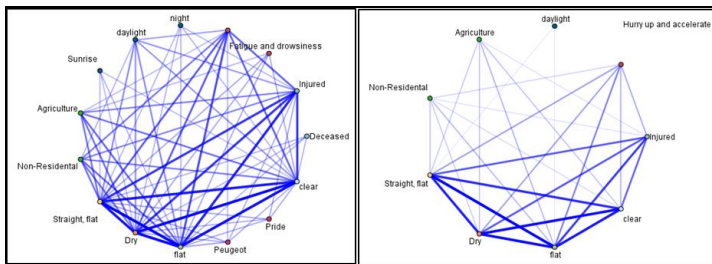
قوانین کلاستر صفر نشان می‌دهد که بیشترین تصادفات منجر به صدمات، در طول روز و مکان‌های با کاربری غیرمسکونی و به دلیل عجله و شتاب بیش از حد انجام گرفته است و دو نوع خودروی پراید و پژو در صدر بیشترین تلفات قرار دارند. این خوشه با تعداد ۲۱۹ تصادف و ثبت تعداد ۴۶ کشته از خطرناک‌ترین قسمت‌های آزاد راه محسوب می‌گردد. اگر به نقشه‌های عوامل مکانی توجه گردد، حداکثر تراکم مراکز جمعیتی و صنعتی در این خوشه قرار گرفته است و مجاورت و نزدیکی این عوامل به جاده و مرکز این خوشه می‌تواند از علل افزایش تصادفات و یا شدت آن باشد. در شکل ۹ شدت روابط بین فاکتورهای مؤثر در تصادفات با دو آستانه متفاوت را نشان می‌دهد.



شکل ۹: شدت فاکتورهای تصادفات در خوشه صفر

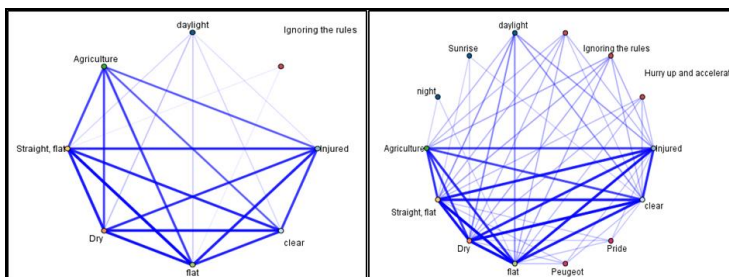
قوانین کلاستر یک نشان می‌دهد که تصادفات صورت گرفته در این مکان در شرایط هوایی نسبتاً مطلوب و به دلیل عجله و شتاب بی‌مورد و نیز خستگی و خواب‌آلودگی رانندگان بوده و درصد زمان وقوع حوادث در زمان طلوع و غروب خورشید افزایش معناداری را

نشان می‌دهد. با توجه به فاکتورهای مکانی، تراکم سکونتگاه‌ها و مراکز جمعیتی و همچنین تقاطع‌ها در این خوشه بالا بوده و هندسه جاده نیز به صورت مستقیم نبوده و انحنا و پیچ را در اینجا شاهد می‌باشیم که خود می‌تواند در شدت حوادث و یا وقوع آن‌ها حائز اهمیت باشد. شکل ۱۰ روابط بین عوامل مختلف تصادفات در این خوشه را نشان می‌دهد. این خوشه دارای تعداد ۲۲۶ تصادف با ۵۰ کشته است.



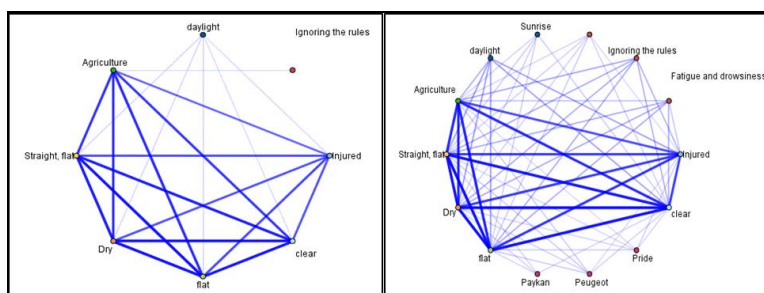
شکل ۱۰: شدت فاکتورهای تصادفات در خوشه ۱

قوانین کلاستر دو نشان می‌دهد که تصادفات در طول روز و موقع طلوع بالاترین نرخ را داشته و نادیده گرفتن قوانین و سرعت بالا از فاکتورهای انسانی وقوع تصادف بوده است. خودروهای پراید و پژو همچنان در این خوشه بالاترین نرخ تصادفات را به خود اختصاص داده‌اند. با توجه به عوامل مکانی هندسه راه که بیشتر به صورت صاف و بدون پیچ بوده و کم تراکم بودن تقاطع‌ها و مراکز جمعیتی خود می‌تواند از علل کم بودن تعداد تصادف در این خوشه یاد کرد این خوشه با تعداد ۸۷ تصادف و ثبت تعداد ۱۰ کشته کمترین نرخ تصادف را در برداشته است. شکل ۱۱ شدت روابط بین عوامل مختلف وقوع تصادفات را نشان می‌دهد.



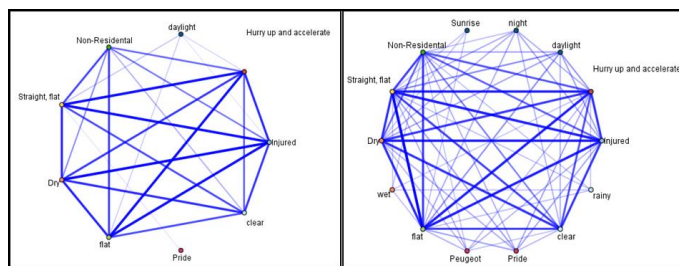
شکل ۱۱: شدت فاکتورهای تصادفات در خوشه ۲

قوانین کلاستر ۳ با تعداد ۸۸ مورد تصادف و ۱۶ کشته نشان می‌دهد که بیشترین تلفات در زمان طلوع خورشید و در شرایط توپوگرافی هموار و یکنواخت افزایش چشمگیری داشته است. همچنین خستگی و خواب آلودگی در شرایط یکنواخت جاده و نادیده گرفتن قوانین رانندگی می‌تواند از علل تشدید تصادفات در این خوشه یاد نمود. خودروی پراید حداکثر تصادفات منجر به جرح در این خوشه را به خود اختصاص داده است. تراکم سکونتگاه‌ها و مراکز خدماتی و صنعتی در این خوشه نسبتاً پایین است. شکل ۱۲ روابط بین عوامل مختلف تصادفات را در این خوشه نشان می‌دهد.



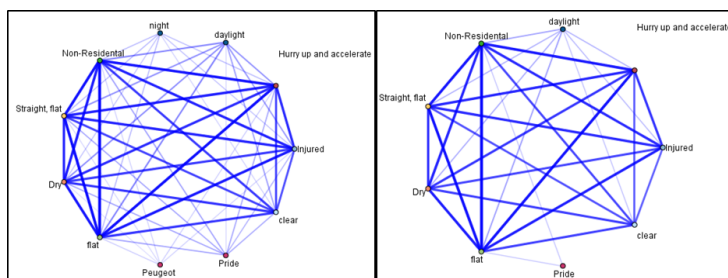
شکل ۱۲: شدت فاکتورهای تصادفات در خوشه ۳

قوانین کلاستر ۴ نشان می‌دهد، در این خوشه شرایط جوی ناپایدار و سطح خیس جاده آمار تلفات منجر به فوت را به شدت افزایش داده و بیشترین عوامل انسانی دخیل نیز عجله و شتاب بالا و بی‌توجهی به قوانین رانندگی بوده است. هندسه جاده در این خوشه دارای بیشترین پیچ‌های عمودی (سرازیری و سربالایی) است که خود می‌تواند در شدت تصادفات مؤثر باشد. تراکم تعداد تقاطع و مراکز جمعیتی نیز می‌تواند مد نظر قرار گیرد. شایان ذکر است این خوشه نزدیک‌ترین موقعیت به شروع یا پایان آزاد راه بوده که خستگی پایان سفر و حجم ترافیک در ورودی‌ها از مسائلی است که می‌تواند شدت رخداد تصادفات را افزایش دهد. این خوشه دارای تعداد ۲۲۰ تصادف با ۳۱ کشته بوده که خودروهای پراید و پژو بیشترین تصادفات منجر به جرح و فوت را ثبت نموده‌اند. شکل ۱۳ شدت روابط بین عوامل مختلف تصادفات را در این خوشه نشان می‌دهد.



شکل ۱۳: شدت فاکتورهای تصادفات در خوشه ۴

قوانین خوشه ۵ نشان می‌دهد تصادفات این خوشه بیشتر در شرایط مطلوب آب و هوایی رخ داده است. نرخ وقوع تصادفات در زمان شب افزایش را نشان می‌دهد. هندسه جاده در این خوشه دارای یک پیچ بوده که تمرکز تصادفات در آن محل نسبتاً بالا است. عجله و شتاب بی‌مورد از علل انسانی مهم در وقوع تصادفات بوده و نیز عوامل مکانی از قبیل تراکم و نزدیکی به مراکز جمعیتی و خدماتی از علل مهم وقوع حوادث و شدت آن در این محل است. با توجه به قرار گرفتن شهر نسبتاً پرجمعیت تاکستان در مجاورت این خوشه، خود می‌تواند افزایش سوانح رانندگی را در بر داشته باشد. این خوشه با تعداد ۱۴۶ تصادف تعداد ۱۸ کشته را در پی داشته است. شکل ۱۴ شدت روابط بین فاکتورهای مختلف تصادف را در این خوشه نشان می‌دهد.

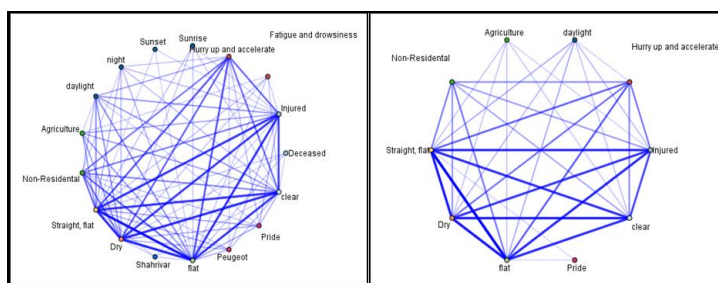


شکل ۱۴: شدت فاکتورهای تصادفات در خوشه ۵

قوانین حاصل از کاوش کلیه داده‌ها به‌عنوان یک خوشه نشان می‌دهد که میزان تصادفات در شرایط عادی آب و هوا و سطح جاده خشک و هموار نرخ بالاتری را به همراه داشته است و این می‌تواند به دلیل کم توجهی رانندگان، شتاب و عجله بیش از حد مجاز و نیز عدم رعایت قوانین باشد. بیشترین نرخ تصادفات مربوط به خودروهای پراید، پژو و پیکان بوده

که آمار تلفات فوتی و جرحی آن‌ها به مراتب بالاتر است. در بین خوردوهای حمل و نقل عمومی اتوبوس اسکانیا بیشترین آمار تصادفات و مجروحین را به خود اختصاص داده است. در بین ماه‌های سال، شهریور دارای فراوانی بالایی بوده و به همان نسبت نیز شدت تصادفات و سوانح منجر به فوت نیز بیشتر است که نرخ سفرها نیز در این ماه رشد بالاتری را نشان می‌دهد.

بعد از زمان روشنایی روز، تصادفات منجر به فوت در زمان طلوع بیشترین نرخ را دارد که ارتباط آن با عامل انسانی خواب آلودگی شدت بالاتری را نسبت به سایر علل داشته است. در مجموع از تعداد ۹۸۶ تصادف انجام گرفته در طول این آزاد راه تعداد ۱۷۱ کشته ثبت گردیده است که تعداد ۹۶ کشته (بیش از ۵۶٪) در فاصله ۶۰ تا ۸۵ کیلومتری از ایستگاه پلیس بوده که منطبق بر خوشه‌های صفر و یک است. در این دو خوشه بیشترین تمرکز مراکز جمعیتی، صنعتی، خدماتی، تقاطع و پیچ‌ها از نظر هندسی، واقع شده‌اند که خود باعث افزایش تعداد و تشدید تصادفات می‌گردد. شکل ۱۵ شدت روابط بین عوامل مختلف و مؤثر در تصادفات بین داده‌ها را نشان می‌دهد.



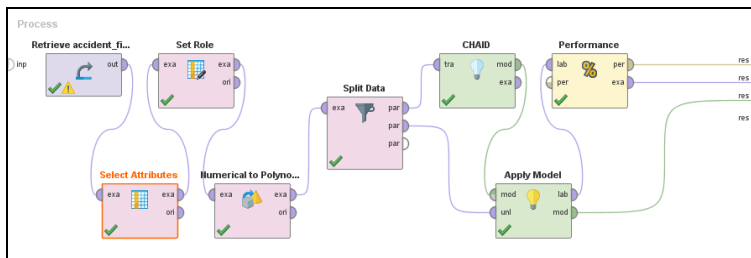
شکل ۱۵: شدت فاکتورهای تصادفات در کل داده‌ها

درخت تصمیم^۱ CHAID (Chi-square automatic interaction detection)

درخت تصمیم، تصویری از پیش‌بینی تصادفات را به صورت مناسب و قابل درک ارائه می‌کند (لاوانیا و دیویا^۱، ۲۰۱۷: ۱۹). در این پژوهش از الگوریتم درخت تصمیم CHAID به منظور پیش‌بینی تصادفات استفاده گردیده است.

^۱ Decision Tree

ماتریس خطا برای محاسبه دقت هر الگوریتم استفاده می‌شود. این ابزاری مفید برای تجزیه و تحلیل طبقه‌بندی کننده به منظور شناخت کارایی تاپل‌ها در کلاس‌های مختلف است (تیواری، ۲۰۱۷: ۱۶، لاونیا و دیویا، ۲۰۱۷: ۲۰). به منظور اجرای طبقه‌بندی، داده‌ها به دو دسته شامل مجموعه داده‌های آموزشی و داده‌های تست تقسیم می‌شوند. از تعداد ۹۸۶ تصادف ۷۰٪ داده‌ها (۶۹۰ تصادف) به عنوان داده آموزشی و ۳۰٪ داده‌ها (۲۹۶ تصادف) به عنوان تست به صورت تصادفی انتخاب گردیدند. شکل ۱۶ روند اجرای طبقه‌بندی را نشان می‌دهد.



شکل ۱۶: عملیات درخت تصمیم

جدول ۶ نتیجه پیش‌بینی حاصل از ۲۹۶ داده تست را بر اساس داده‌های آموزشی و جدول ۷ ماتریس خطا و دقت الگوریتم طبقه‌بندی (Accuracy) را به همراه مقدار کاپا (Kappa) نشان می‌دهد.

جدول ۶: پیش‌بینی داده‌های آزمایشی

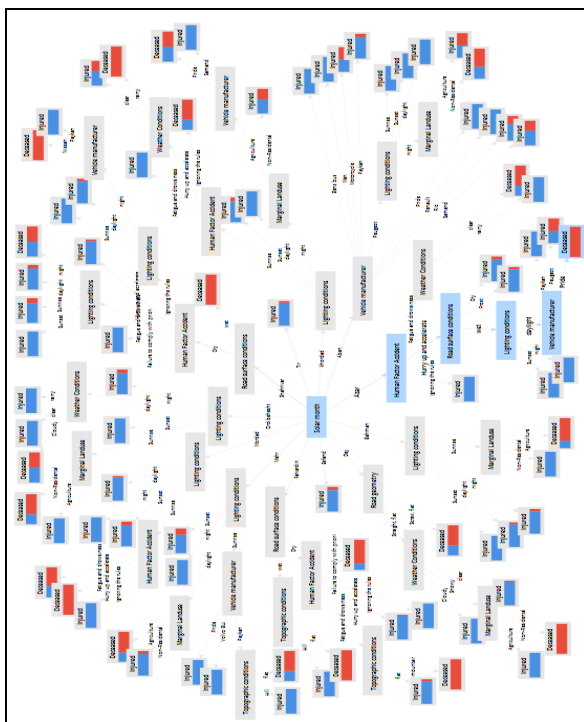
row	Crash severity	prediction	Confidence (injured)	Confidence (deceased)	Weather condition	Solar month
1	Injured	Injured	0.9	0.1	clear	Farvardin
2	Injured	Injured	0.9	0.1	clear	Farvardin
3	Deceased	Injured	0.8	0.14	clear	Ordibehesht
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
295	Injured	Injured	0.9	0.1	clear	Bahman
296	Injured	Injured	0.8	0.1	Cloudy	Esfand

¹ Lavanya & Divya

جدول ۷: ماتریس خطا

	true Injured	true Deceased	class precision
pred. Injured	237	37	86.50%
pred. Deceased	8	14	63.64%
class recall	96.73%	27.45%	
Accuracy: 84.80%		Kappa: 0.312	

شکل ۱۷ درخت تصمیم CHAID ساخته شده بر اساس کلیه داده‌های تصادف را نشان می‌دهد. رنگ قرمز نسبت کشته‌ها و رنگ آبی موارد جرحی را در تصادفات نشان می‌دهد.

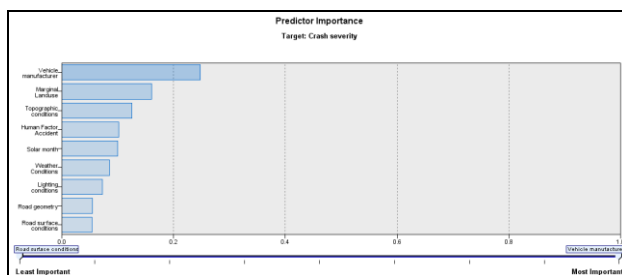


شکل ۱۷: درخت تصمیم CHAID

با نگاهی دقیق‌تر به این درخت می‌توان قواعدی را ملاحظه نمود. برای نمونه خودروی پراید در روشنایی روز و شرایط سطح جاده خیس و عجله و شتاب راننده به‌عنوان عامل انسانی و در آذر ماه باعث شدت تصادف بالاتری خواهد بود. یا اینکه در شرایط آب و هوای بارانی به همراه عامل انسانی عجله و سرعت بیش از حد و در زمان طلوع در خردادماه صدمات و تلفات جانی بیشتری را در بر داشته و شدت تصادفات بالاتر است.

همچنین در شرایط جوی برفی و در زمان طلوع، شدت تصادفات در دی‌ماه افزایش خواهد داشت. بنابراین با دنبال نمودن شاخه‌ها و برگ‌ها می‌توان احتمال تصادفات منجر به جرح یا منجر به فوت را به سادگی مشاهده نمود.

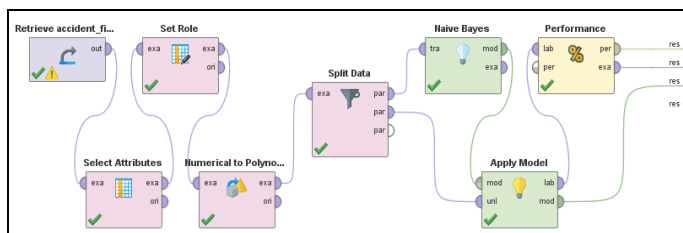
با استفاده از این الگوریتم (CHAID) میزان اهمیت هر کدام از عوامل به دست می‌آید. نوع خودرو (۰/۲۵)، کاربری‌های حاشیه‌ای جاده (۰/۱۶)، شرایط توپوگرافیکی محل (۰/۱۳)، عامل انسانی تصادف (۰/۱)، ماه وقوع حادثه (۰/۱)، شرایط آب و هوایی (۰/۰۸)، شرایط روشنایی و ساعت وقوع تصادف (۰/۰۷)، هندسه جاده (۰/۰۵) و شرایط سطح جاده (۰/۰۵)، به ترتیب مهم‌ترین پیش‌بینی کننده‌ها در تصادفات استخراج گردیدند (شکل ۱۸).



شکل ۱۸: مهم‌ترین پیش‌بینی کننده‌ها در تصادفات

طبقه‌بندی کننده بیزین

به‌منظور اجرای طبقه‌بندی به روش Naïve Bayes، داده‌ها به دو دسته شامل مجموعه داده‌های آموزشی و داده‌های تست تقسیم می‌شوند. از تعداد ۹۸۶ تصادف ۷۰٪ داده‌ها (۶۹۰ تصادف) به‌عنوان داده آموزشی و ۳۰٪ داده‌ها (۲۹۶ تصادف) به‌عنوان تست به‌صورت تصادفی انتخاب گردیدند. شکل ۱۹ روند اجرای طبقه‌بندی را نشان می‌دهد.



شکل ۱۹: اپراتور بیزین

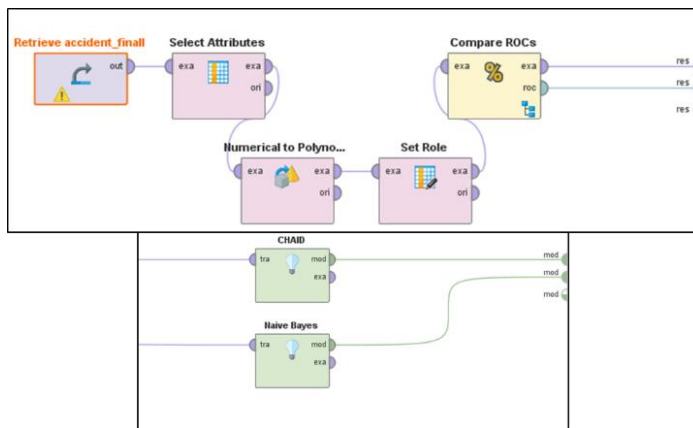
جدول ۸، نتیجه حاصل از ۲۹۶ داده تست را براساس داده‌های آموزشی و جدول ۷ ماتریس خطا و دقت الگوریتم طبقه‌بندی (Accuracy) را به همراه مقدار کاپا (Kappa) نشان می‌دهد.

جدول ۸: پیش‌بینی داده‌های آزمایشی

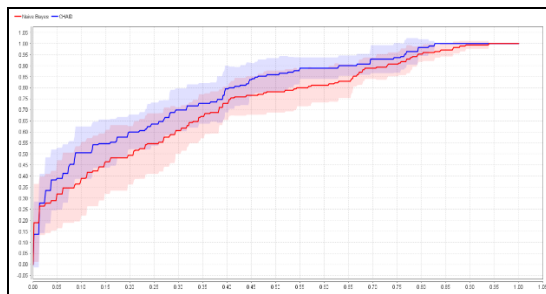
row	Crash severity	prediction	Confidence (injured)	Confidence (deceased)	Weather condition	Solar month
1	Injured	Injured	0.79	0.2	clear	Ordibehesht
2	Injured	Injured	0.83	0.1	clear	Ordibehesht
3	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
295	Injured	Injured	0.9	0.06	clear	Bahman
296	Injured	Injured	0.95	0.04	Cloudy	Esfand

جدول ۹: ماتریس خطا

	true Injured	true Deceased	class precision
pred. Injured	230	31	88.12%
pred. Deceased	15	20	57.14%
class recall	93.88%	39.22%	
Accuracy: accuracy: 84.46%		Kappa: kappa: 0.378	



شکل ۲۰: روند اجرا



شکل ۲۱: مقایسه منحنی ROC (Naïve Bayes & CHAID)

در ادامه صحت الگوهای ارائه شده CHAID DT و Naïve Bayes به وسیله منحنی ROC مقایسه می‌گردند. در منحنی ROC هر چقدر قوس منحنی به سمت گوشه بالا و سمت چپ نمودار نزدیک‌تر باشد، وضعیت مدل مناسب‌تر خواهد بود. شکل ۲۰ روند اجرا و شکل ۲۱ مقایسه این دو مدل را نشان می‌دهد. مدل درخت تصمیم CHAID وضعیت مناسب‌تری را نشان می‌دهد.

نتیجه‌گیری و پیشنهادها

با توجه به مطالعات انجام شده، داده‌کاوی ثابت کرده است که یک تکنیک بسیار مناسب به منظور آنالیز تصادفات جاده‌ای است. در این مقاله شدت تصادفات جاده‌ای با استفاده از الگوریتم‌های داده‌کاوی از قبیل خوشه‌بندی، کشف قوانین انجمنی، درخت تصمیم CHAID و طبقه‌کننده Naïve Bayes در آزادراه قزوین- زنجان با تلفیق عوامل مکانی و غیرمکانی تاثیرگذار بر تصادفات، ارزیابی، تحلیل و پیش‌بینی گردید.

با اجرای عملیات خوشه‌بندی، داده‌های تصادفات به صورت مجموعه‌های همگنی استخراج گردیده و سپس با استفاده از الگوریتم Apriori، برای هر خوشه و نیز تمام داده‌ها فاکتورهای مرتبط با هم در هر خوشه مشخص گردیدند. همچنین به منظور انجام پیش‌بینی شدت تصادفات از دو الگوریتم درخت تصمیم CHAID و طبقه‌کننده Naïve Bayes

استفاده و داده‌های تصادفات به دو مجموعه آموزشی و تست تقسیم و میزان صحت این دو الگو ارزیابی گردید. نتایج نشان می‌دهد که الگوی درخت تصمیم، شدت تصادفات را در مقایسه با روش بیزی با صحت بالاتری پیش‌بینی می‌نماید. با انجام تحلیل حساسیت به‌منظور تعیین اهمیت نسبی متغیرهای ورودی و مؤثر در شدت تصادفات، مشخص گردید که به ترتیب عواملی از قبیل: نوع خودرو، کاربری‌های حاشیه‌ای جاده، شرایط توپوگرافیکی محل، عامل انسانی تصادف، ماه وقوع حادثه، شرایط آب و هوایی، شرایط روشنایی و ساعت وقوع تصادف، هندسه جاده و شرایط سطح جاده بیشترین نقش را در شدت تصادفات جاده‌ای داشته‌اند.

با توجه به تحلیل‌های انجام گرفته، خودروهای پراید، پژو، پیکان و نیسان وانت به همراه اتوبوس اسکانیا در صحنه تصادفات بیشترین مرگ و میر و جراحات را به خود اختصاص داده‌اند. این در حالی است که شدت تصادفات در شرایط هندسه نامناسب جاده، مناطق مجاور با مراکز جمعیتی، خدماتی، صنعتی و تقاطع‌ها به‌طور محسوس افزایش را نشان می‌دهد. ماه‌های شهریور، مرداد، خرداد و فروردین میزان تصادف و مرگ و میر ناشی از آن، افزایش معناداری را نشان می‌دهد که افزایش حجم سفرها را نیز می‌بایست لحاظ نمود. توجه به تحلیل مکانی و فاکتورهای مرتبط به آن از قبیل: تراکم تقاطع‌ها، شیب جاده، تراکم مراکز جمعیتی، خدماتی، صنعتی و نیز کاربری‌های حاشیه راه در کنار سایر عوامل محیطی، انسانی و خودرویی از موارد مهمی است که در تحلیل الگوریتم‌ها و قوانین استخراج شده مورد استفاده قرار گرفته است.

این مطالعه به متخصصین ایمنی راه در اصلاح زیرساخت‌ها و عملیات بهسازی جاده، پلیس راه، سازمان‌های امداد رسانی و نیز سایر سازمان‌های مرتبط کمک می‌نماید تا با رفع نقایص موجود و یا تصحیح آیین نامه‌ها و بهینه‌سازی محل نصب علائم هشداردهنده و نظارتی از قبیل دوربین‌های کنترل سرعت و غیره از تعداد تصادفات و جراحات ناشی از آن بکاهند. آگاه نمودن رانندگان و استفاده کنندگان از جاده‌ها در مورد عوامل مخاطره‌آمیز جاده‌ای خاص می‌تواند شدت و تعداد تصادفات را کاهش دهد. همچنین کشف دانش حاصل از

داده کاوی این داده‌ها می‌تواند به متولیان صنعت خودرو جهت برطرف نمودن نقایص فنی، اصلاح ساختارهای تولید و طراحی مناسب‌تر و ایمن‌تر یاری رساند.

فهرست منابع

- Ahmed, W., TabrezNafis, M. and Biswas, S.S., 2017. Performance Analysis of Naïve Bayes Algorithm on Crime Data using Rapid Miner. *International Journal*, 8(5).
- Akomolafe, D.T. and Olutayo, A., 2012. Using Data Mining Technique to Predict Cause of Accident and Accident Prone Locations on Highways. *American Journal of Database Theory and Application*, 1(3), pp.26-38.
- Atnafu, B. and Kaur, G., 2017. Analysis and predict the nature of road traffic accident using data mining techniques in Maharashtra, India. *Analysis*.
- Berkhin, P., 2006. A survey of clustering data mining techniques. *Grouping multidimensional data*, 25, p.71.
- Chaurasia, V. and Pal, S., 2014. Data mining approach to detect heart diseases.
- Effati, M., Rajabi, M.A., Hakimpour, F. and Shabani, S., 2014. Prediction of crash severity on two-lane, two-way roads based on fuzzy classification and regression tree using geospatial analysis. *Journal of Computing in Civil Engineering*, 29(6), p.04014099.
- Fraley, C. and Raftery, A.E., 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. *The computer journal*, 41(8), pp.578-588.
- Gupta, A., Mohammad, A., Syed, A. and Halgamuge, M.N., 2016. A comparative study of classification algorithms using data mining: crime and accidents in Denver City the USA. *Education*, 7(7).
- Han, J., Pei, J. and Kamber, M., 2011. *Data mining: concepts and techniques*. Elsevier.
- Jain, A.K., 2010. Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8), pp.651-666.
- Jayababu, Y., Varma, G.P.S. and Govardhan, A., 2017. Mining Spatial Association Rules to Automatic Grouping of Spatial Data Objects Using Multiple Kernel-Based Probabilistic Clustering. *Journal of Intelligent Systems*, 26(3), pp.561-572.
- Kass, G.V., 1980. An exploratory technique for investigating large quantities of categorical data. *Applied statistics*, pp.119-127.
- Kashyap, J. and Singh, C.P., 2016. Mining road traffic accident data to improve safety on road-related factors for classification and prediction of accident severity. *International research journal of engineering and technology*, 3(10), pp.221-226.
- Kumar, S. and Toshniwal, D., 2016. A data mining approach to characterize road accident locations. *Journal of Modern Transportation*, 24(1), pp.62-72.
- Kumar, S. and Toshniwal, D., 2015. A data mining framework to analyze road accident data. *Journal of Big Data*, 2(1), p.26.
- Lavanya, B. and Divya, B., 2017. Predictive Analytics on Accident Data Using Rule Based and Discriminative Classifiers. *Advances in Computational Sciences and Technology*, 10(3), pp.461-469.
- Leung, K.M., 2007. Naive bayesian classifier. Polytechnic University Department of Computer Science/Finance and Risk Engineering.

- Li, D., Wang, S., Yuan, H. and Li, D., 2016. Software and applications of spatial data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 6(3), pp.84-114.
- Li, L., Shrestha, S. and Hu, G., 2017, June. Analysis of road traffic fatal accidents using data mining techniques. In *Software Engineering Research, Management and Applications (SERA), 2017 IEEE 15th International Conference on* (pp. 363-370). IEEE.
- Mahajan, N. and Kaur, B.P., 2016. Analysis of Factors of Road Traffic Accidents using Enhanced Decision Tree Algorithm. *Analysis*, 135(6).
- Martín, L., Baena, L., Garach, L., López, G. and de Oña, J., 2014. Using data mining techniques to road safety improvement in Spanish roads. *Procedia-Social and Behavioral Sciences*, 160, pp.607-614.
- Mizutani, H. and Kanai, R., 2017. A description length approach to determining the number of k-means clusters. *Ar Xiv preprint arXiv: 1703.00039*.
- Morissette, L. and Chartier, S., 2013. The k-means clustering technique: General considerations and implementation in Mathematica. *Tutorials in Quantitative Methods for Psychology*, 9(1), pp.15-24.
- Pelleg, D. and Moore, A. 2000. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In *Proceedings 17th ICML, Stanford University*.
- Prati, G., Pietrantonio, L. and Fraboni, F., 2017. Using data mining techniques to predict the severity of bicycle crashes. *Accident Analysis & Prevention*, 101, pp.44-54.
- Sanmiquel, L., Rossell, J.M. and Vintró, C., 2015. Study of Spanish mining accidents using data mining techniques. *Safety science*, 75, pp.49-55.
- Sapri, F.E., Nordin, N.S., Hasan, S.M., Yaacob, W.F.W. and Nasir, S.A.M., 2017. Decision Tree Model for Non-Fatal Road Accident Injury. *International Journal on Advanced Science, Engineering and Information Technology*, 7(1), pp.63-70.
- Shetty, P., Sachin, P.C., Kashyap, S.V. and Madi, V., 2017. Analysis of road accidents using data mining techniques.
- Shinde, T.A. and Prasad, J.R., 2017. IoT based Animal Health Monitoring with Naive Bayes Classification. *IJETT*, 1(2).
- Tiwari, P., Dao, H. and Nguyen, G.N., 2017. Performance Evaluation of Lazy, Decision Tree Classifier and Multilayer Perceptron on Traffic Accident Analysis. *Informatica*, 41(1), p.39.
- Tiwari, P., Kumar, S. and Kalitin, D., 2017, March. Road-User Specific Analysis of Traffic Accident Using Data Mining Techniques. In *International Conference on Computational Intelligence, Communications, and Business Analytics* (pp. 398-410). Springer, Singapore.
- World Health Organization, 2015. Global status report on road safety 2015. World Health Organization.
- World Road Association, 2003. Road safety manual. Piarc technical committee on road safety.