خوشه‌بندی محورهای برون‌شهری بر اساس الگوهای تردد، سرعت، ترکیب ناوگان و تخلفات رانندگی با استفاده از داده‌های ترددشماری (مطالعه موردی: استان خراسان رضوی، سال ۱۴۰۴)

مقاله علمی-پژوهشی

سحر علیشاهی، دانش آموخته کارشناسی ارشد، گروه مهندسی عمران، دانشکده مهندسی، دانشگاه فردوسی، مشهد، ایران

*__سید علی ضیائی (نویسنده مسئول)__، دانشیار، گروه مهندسی عمران، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران

حسین محمدی، دانش آموخته کارشناسی ارشد، گروه مهندسی عمران، دانشکده مهندسی، دانشگاه فردوسی، مشهد، ایران

*__پست الکترونیکی نویسنده مسئول:sa-ziaee@um.ac.ir__

دریافت: ۱۴۰۵/۰۱/۲۲- پذیرش: ۱۴۰۵/۰۵/۰۵

صفحه ۳۵۴-۳۴۳

چکیده

مدیریت ایمنی و بهره برداری راه‌های برون شهری نیازمند شناخت تفاوت‌های رفتاری محورهای شبکه است. رتبه بندی ساده محور های پرخطر اگرچه برای اولویت بندی مفید است، اما قادر به آشکارسازی تیپ های رفتاری مشابه از نظر حجم تردد، سرعت، ترکیب ناوگان و الگوی تخلفات نیست. هدف این پژوهش، خوشه بندی محورهای برون شهری استان خراسان رضوی بر اساس داده های تردد شماری سامانه ۱۴۱ در سال ۱۴۰۴ است. در این مطالعه، ۱۳۶ محور یک جهته پس از کنترل کیفیت داده و تجمیع ویژگی های محوری وارد تحلیل شدند. ویژگی‌های مورد استفاده شامل حجم تردد سالانه، نرخ تخلف فاصله غیرمجاز، نرخ تخلف سرعت، نرخ تخلف سبقت، سرعت متوسط وزنی و سهم وسایل نقلیه تجاری/سنگین بود. برای کاهش اثر مقیاس، داده ها استاندارد سازی شدند و برای نمایش ساختار داده از تحلیل مؤلف ه‌های اصلی استفاده شد. سپس دو رویکرد K-Means و خوشه بندی سلسله مراتبی Ward برای تعداد خوشه های ۲ تا ۷ ارزیابی شدند. بر اساس شاخص Silhouette، مقدار k=٤ با ضریب ۰,۲۶۹ بهترین پیکربندی K-Means را ارائه کرد. نتایج چهار تیپ رفتاری شامل «پرریسک با تخلف فاصله بالا»، «پرتردد با ریسک رفتاری متوسط»، «تجاری/سنگین با سرعت کنترل‌شده» و «کم‌ریسک تر و پایدارتر» را نشان داد. خوشه نخست با ٤٧ محور، نرخ میانگین تخلف فاصله ۲۰۵,۵۵ تخلف به ازای هر ۱۰۰۰ وسیله نقلیه و نرخ وزنی ۲٤٤,۹۲ داشت و از دید مدیریت ایمنی در اولویت نخست پایش و مداخله قرار می‌گیرد. نتایج این پژوهش نشان می‌دهد خوشه بندی داده محور می‌تواند مکمل مناسبی برای شاخص‌های ریسک و مدل های پیش بینی باشد و به راهداری کمک کند محور های مشابه را با سیاست‌های هماهنگ‌تر مدیریت کند.

واژه‌های کلیدی: خوشه‌بندی، K-Means، ترددشماری، تخلف فاصله غیرمجاز، ترکیب ناوگان

۱-مقدمه

ایمنی راه و بهره‌برداری شبکه حمل‌ونقل، به‌ویژه در محورهای برون‌شهری، به شدت وابسته به شناخت تفاوت‌های رفتاری محورهای مختلف است. گزارش جهانی ایمنی راه نشان می‌دهد که تصادفات جاده‌ای همچنان یکی از چالش‌های جدی سلامت عمومی در جهان است و سالانه حدود ۱,۱۹ میلیون نفر در اثر تصادفات جان خود را از دست می‌دهند [Abduljabar et al,2025]. گزارش‌های بین‌المللی ایمنی راه نیز بر استفاده از داده‌های قابل اعتماد، پایش پیوسته عملکرد شبکه و تصمیم‌گیری مبتنی بر شواهد تأکید می‌کنند [Caliński et al,2024]. در چنین فضایی، داده‌های ترددشماری سامانه ۱۴۱ می‌توانند علاوه بر گزارش حجم تردد، برای شناسایی الگوهای پنهان رفتاری، طبقه‌بندی محورهای مشابه و طراحی

سیاست‌های مداخله‌ای هدفمند استفاده شوند.در مطالعات حمل‌ونقل، خوشه‌بندی یکی از ابزارهای اصلی یادگیری بدون ناظر برای کشف ساختارهای پنهان در داده‌هاست. خوشه‌بندی به پژوهشگر اجازه می‌دهد بدون تعریف قبلی برچسب ریسک، محورهایی را که از نظر حجم، سرعت، ترکیب ناوگان و تخلف رفتاری مشابه هستند در یک گروه قرار دهد. این نوع تحلیل برای شبکه‌هایی مانند خراسان رضوی اهمیت ویژه دارد، زیرا محورهای زیارتی، تجاری، حومه‌ای و بین‌شهری ممکن است از نظر میانگین تردد یا سرعت به تنهایی قابل تفکیک نباشند، اما در ترکیب چندمتغیره رفتارهای متفاوتی نشان دهند. نوآوری مقاله در تبدیل دیتابیس ساعتی/روزانه سامانه ۱۴۱ به یک پایگاه داده محور–محور برای خوشه‌بندی رفتاری است. سؤال اصلی پژوهش این است: آیا می‌توان محورهای برون‌شهری استان خراسان رضوی را بر اساس الگوهای چندبعدی تردد، سرعت، ترکیب ناوگان و تخلفات رانندگی به تیپ‌های رفتاری معنادار تقسیم کرد؟ سؤال فرعی دوم این است که هر خوشه چه دلالتی برای مدیریت ایمنی و اولویت‌بندی مداخلات دارد؟

داده‌ها و آماده‌سازی ویژگی‌ها

داده‌های این پژوهش از دیتابیس نهایی پژوهشی سامانه ۱۴۱ برای سال ۱۴۰۴ استخراج شد. واحد تحلیل، محور یک‌جهته است؛ بنابراین هر کد شش‌رقمی محور به عنوان یک جهت حرکت مستقل در نظر گرفته شد. این انتخاب با تعریف رسمی ترددشماری سازگار است و از ادغام نادرست رفت و برگشت جلوگیری می‌کند. پس از کنترل کیفیت داده‌های ساعتی، شاخص‌های محوری در سطح سالانه محاسبه شدند.تخلف فاصله غیرمجاز بر اساس تعریف سامانه، عدم رعایت حداقل فاصله زمانی دو ثانیه‌ای با وسیله نقلیه جلویی است. از آنجا که تعداد خام تخلف تحت تأثیر حجم تردد قرار دارد، در این پژوهش نرخ تخلف فاصله، سرعت و سبقت به ازای هر ۱۰۰۰ وسیله نقلیه محاسبه شد. همچنین برای کنترل اثر مقیاس محور، لگاریتم حجم تردد سالانه در کنار متغیرهای نرخ و سهم ناوگان استفاده شد.

۲- پیشینه تحقیق

۲-۱- خوشه‌بندی در تحلیل ترافیک

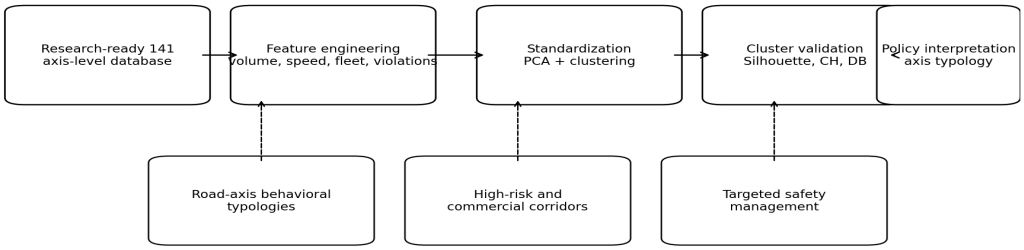
K-Means یکی از پرکاربردترین روش‌های خوشه‌بندی است که با کمینه‌سازی مجموع مربعات درون‌خوشه‌ای، داده‌ها را به k گروه تقسیم می‌کند [Davies et al, 1979 and Etemad, 2023]. خوشه‌بندی سلسله‌مراتبی Ward نیز با کمینه‌سازی افزایش واریانس درون‌خوشه‌ای، ساختاری درختی از گروه‌بندی داده‌ها ایجاد می‌کند [Federal Highway Administration Road Safety Fundamentals, 2024]. در کاربردهای ترافیکی، این روش‌ها برای شناسایی الگوهای مشابه جریان، سرعت، تقاضا و وضعیت شبکه استفاده شده‌اند. مطالعات جدید نیز نشان می‌دهند که ترکیب روش‌های خوشه‌بندی با داده‌های پرتعداد ترافیکی می‌تواند برای مدیریت هوشمند ترافیک، تشخیص الگوهای زمانی و برنامه‌ریزی عملیاتی مفید باشد [MacQueen,1967 and Rousseeuw 1987].

۲-۲- اعتبارسنجی خوشه‌ها

از آنجا که خوشه‌بندی برچسب مرجع ندارد، ارزیابی درونی خوشه‌ها اهمیت زیادی دارد. شاخص Silhouette انسجام درون‌خوشه‌ای و جدایی بین‌خوشه‌ای را با دامنه نظری −۱ تا +۱ می‌سنجد؛ مقادیر نزدیک به صفر معمولاً نشان‌دهنده همپوشانی خوشه‌هاست و در داده‌های واقعی و چندبعدی حمل‌ونقل انتظار مقادیر بسیار بالا همیشه واقع‌بینانه نیست [Federal Highway Administration ,2008 and Fredriksso et al, 2025]. شاخص Calinski–Harabasz نسبت جدایی بین‌خوشه‌ای به پراکندگی درون‌خوشه‌ای را نشان می‌دهد و مقدار بالاتر آن مطلوب‌تر است [Ikotun et al, 2023]. شاخص Davies–Bouldin نیز میانگین شباهت نسبی خوشه‌ها را می‌سنجد و مقدار کمتر آن مطلوب‌تر است [Road Safety Annual Report, 2024].

نقش در مقاله	متغیر/روش	گروه ویژگی
نمایش اندازه و شدت تقاضای محور	log_total_vehicles، total_vehicles	حجم تردد
تمایز محورهای آزادجریان و کنترل‌شده	avg_speed_weighted	رفتار سرعتی
بازنمایی رفتار پرریسک رانندگان	unsafe_distance_rate_per_۱۰۰۰، speeding_rate_per_۱۰۰۰، overtaking_rate_per_۱۰۰۰	تخلفات رفتاری
شناخت محورهای تجاری/سنگین	commercial_vehicle_share_class_۲_۵	ترکیب ناوگان
حذف اثر مقیاس و نمایش دوبعدی	StandardScaler، PCA	آماده‌سازی آماری
انتخاب تعداد خوشه و مقایسه الگوریتم‌ها	Silhouette، Calinski-Harabasz، Davies-Bouldin	اعتبارسنجی خوشه

جدول ۱. ویژگی‌های مورد استفاده در خوشه‌بندی محورهای برون‌شهری



شکل ۱. چارچوب روش‌شناسی مقاله خوشه‌بندی

۳-روش‌شناسی

فرایند روش‌شناسی شامل پنج مرحله بود: ۱) استخراج شاخص‌های محوری از دیتابیس research-ready، ۲) استانداردسازی متغیرها برای حذف اثر مقیاس، ۳) اجرای K-Means و Agglomerative-Ward برای k=۲ تا ۷، ٤) انتخاب تعداد خوشه با استفاده از شاخص‌های درونی، و ۵) تفسیر خوشه‌ها بر اساس پروفایل رفتاری آن‌ها. برای نمایش دوبعدی ساختار داده، تحلیل مؤلفه‌های اصلی نیز به کار رفت. دو مؤلفه نخست PCA در مجموع ٦٤,٦٧ درصد واریانس ویژگی‌ها را توضیح دادند که برای نمایش تصویری ساختار کلی داده مناسب است.در انتخاب تعداد خوشه، تمرکز اصلی بر Silhouette بود، اما Calinski-Harabasz و Davies-Bouldin نیز برای کنترل پایداری تفسیر گزارش شدند. با توجه به ماهیت واقعی داده‌های راه، همپوشانی رفتاری محورها طبیعی است؛ بنابراین انتخاب k صرفاً بر اساس بیشینه عددی شاخص‌ها انجام نشد، بلکه تفسیرپذیری مهندسی خوشه‌ها نیز معیار نهایی بود.

جدول ۲. ارزیابی الگوریتم‌ها و تعداد خوشه‌های مختلف

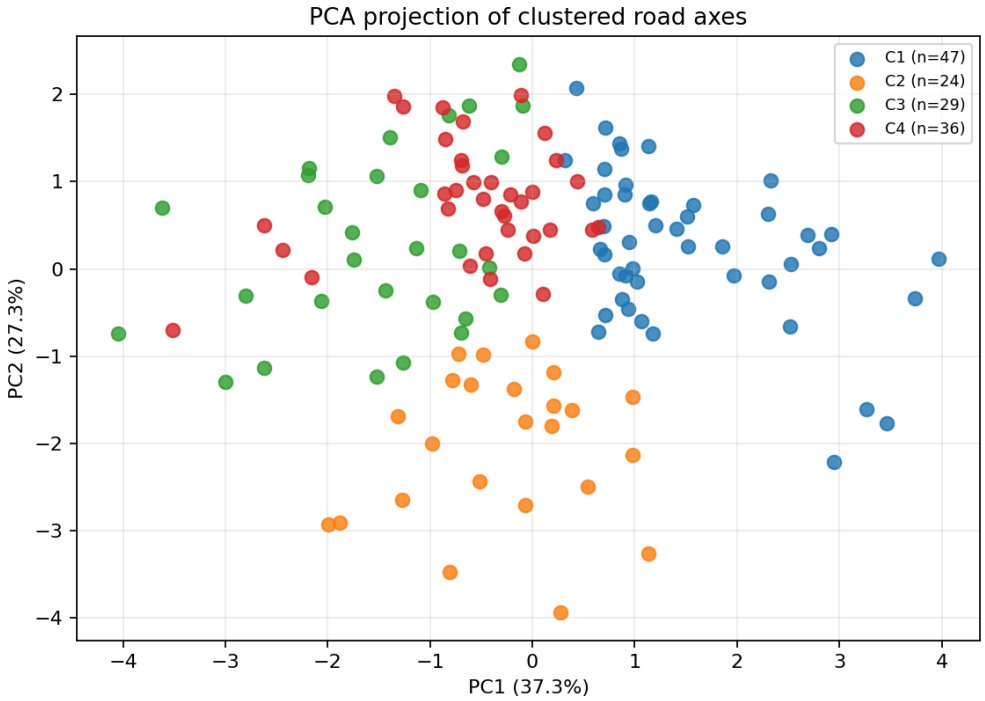
الگوریتم	تعداد خوشه	Silhouette	Calinski-Harabasz	Davies-Bouldin
KMeans	۲	۰,۲۶۶	۴۲,۸۱۵	۱,۶۱۰
KMeans	۳	۰,۲۴۲	۴۳,۰۳۰	۱,۳۸۰
KMeans	۴	۰,۲۶۹	۴۴,۳۹۷	۱,۲۱۵
KMeans	۵	۰,۲۵۱	۴۱,۹۹۰	۱,۲۵۱
KMeans	۶	۰,۲۵۵	۴۲,۴۸۳	۱,۱۷۶
KMeans	۷	۰,۲۵۵	۴۱,۰۷۶	۱,۰۸۹
Agglomerative-Ward	۲	۰,۲۶۴	۴۲,۴۸۷	۱,۶۱۷
Agglomerative-Ward	۳	۰,۲۰۲	۳۹,۰۹۲	۱,۴۳۰
Agglomerative-Ward	۴	۰,۲۲۴	۳۶,۴۰۳	۱,۵۵۰
Agglomerative-Ward	۵	۰,۲۳۵	۳۶,۱۲۲	۱,۲۸۳
Agglomerative-Ward	۶	۰,۲۲۷	۳۵,۷۹۴	۱,۲۰۷
Agglomerative-Ward	۷	۰,۲۱۴	۳۶,۰۳۳	۱,۲۲۳

شکل ۲. ارزیابی تعداد خوشه بر اساس شاخص‌های درونی

٤-تعداد خوشه و ساختار دوبعدی

نتایج نشان داد بهترین مقدار K-Means بر اساس Silhouette برابر با k=٤ و مقدار شاخص برابر با ۰,۲٦۹ است. شاخص Davies-Bouldin برای k=٤ برابر با ۱,۲۱۵ بود که نسبت به بسیاری از حالات دیگر جدایی مناسب‌تری را نشان می‌دهد. شکل ۳ نمایش PCA محورها را همراه با برچسب خوشه‌ها نشان می‌دهد. پراکندگی نقاط تأیید می‌کند که خوشه‌ها کاملاً جدا و مصنوعی نیستند، بلکه تیپ‌هایی از رفتارهای ترافیکی با مرزهای هم‌پوشان در شبکه واقعی هستند؛ این موضوع در تحلیل ترافیکی طبیعی و قابل انتظار است.

.



شکل ۳. نمایش دوبعدی خوشه‌ها با PCA

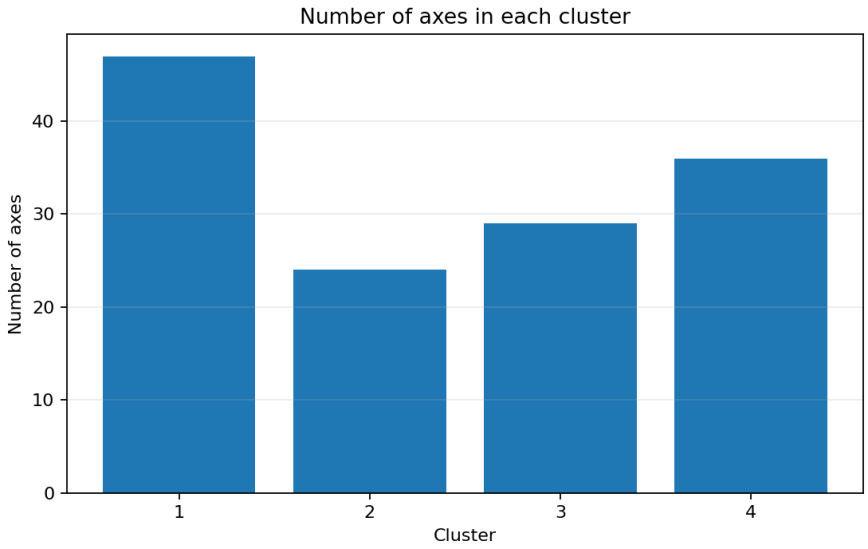
پروفایل و تفسیر خوشه‌ها

جدول ۳ و شکل‌های ٤ و ٥ پروفایل چهار خوشه نهایی را نشان می‌دهند. خوشه ۱ با ٤۷ محور و نرخ وزنی تخلف فاصله ۲٤٤,۹۲، پرریسک‌ترین خوشه از نظر فاصله‌گیری رانندگان است. خوشه ۲ شامل ۲٤ محور با ریسک رفتاری متوسط و حجم پایین‌تر است. خوشه ۳ با ۲۹ محور، نرخ سرعت بسیار بالا و نرخ فاصله کمتر دارد و می‌تواند نماینده محورهای آزادجریان‌تر یا دارای الگوی سرعتی خاص باشد. خوشه ٤ با ۳٦ محور، سهم تجاری بالاتر و نرخ فاصله پایین‌تر دارد و از منظر تخلف فاصله کم‌ریسک‌تر و پایدارتر تفسیر می‌شود.

جدول ۳. خلاصه و تفسیر چهار خوشه نهایی محورهای برون‌شهری

خوشه	برچسب تفسیری	تعداد محور	تردد کل	میانگین نرخ فاصله/۱۰۰۰	نرخ وزنی فاصله/۱۰۰۰	میانگین نرخ سرعت/۱۰۰۰	میانگین سرعت	سهم تجاری
۱	پرریسک با تخلف فاصله بالا	۴۷	۲۹۹،۶۸۳،۰۶۲	۲۰۵,۵۵۰	۲۴۴,۹۱۸	۱۰۸,۷۹۸	۸۶,۳۵۱	۰,۱۱۸
۲	پرتردد با ریسک رفتاری متوسط	۲۴	۲۹،۳۸۹،۴۴۸	۱۵۹,۷۰۹	۱۷۶,۶۲۱	۸۳,۷۰۹	۶۷,۲۲۸	۰,۱۱۵
۳	تجاری/سنگین با سرعت کنترل‌شده	۲۹	۳۲،۶۰۰،۸۰۶	۸۷,۱۰۰	۱۱۵,۳۸۷	۳۷۰,۹۶۶	۸۷,۲۱۴	۰,۱۱۴
۴	کم‌ریسک‌تر و پایدارتر	۳۶	۶۵،۸۶۷،۲۴۴	۶۲,۹۶۳	۷۳,۲۸۸	۱۳۰,۶۴۹	۸۸,۵۲۳	۰,۲۶۵

شکل ۴. پروفایل استانداردشده خوشه‌ها بر اساس ویژگی‌های اصلی



شکل ۵. تعداد محورهای عضو هر خوشه

محورهای شاخص در هر خوشه

برای کمک به تفسیر میدانی خوشه‌ها، سه محور دارای بیشترین نرخ تخلف فاصله در هر خوشه در جدول ٤ ارائه شده است. در خوشه پرریسک، محورهای مرتبط با مشهد–چناران، فریمان–مشهد و چناران–مشهد در صدر قرار دارند. این الگو می‌تواند ناشی از حجم تقاضا، جریان‌های حومه‌ای–زیارتی، نزدیکی به مراکز جذب سفر و فشردگی جریان باشد. البته برای نتیجه‌گیری قطعی، اتصال داده‌های هندسی، حد سرعت مجاز و داده تصادفات ضروری است. نتایج خوشه‌بندی نشان می‌دهد که محورهای شبکه خراسان رضوی را نمی‌توان صرفاً بر اساس حجم تردد یا سرعت متوسط طبقه‌بندی کرد.

برخی محورهای پرتردد دارای نرخ بالای تخلف فاصله هستند، در حالی که برخی محورهای دارای سهم بالاتر وسایل تجاری، نرخ فاصله پایین‌تری نشان می‌دهند. این یافته می‌تواند بیانگر تفاوت در ماهیت جریان، هدف سفر، ترکیب تقاضا و رفتار رانندگان باشد. از دیدگاه مهندسی ترافیک، خوشه ۱ باید در اولویت کنترل و پایش فاصله‌گیری قرار گیرد؛ خوشه ۳ نیازمند بررسی رفتار سرعتی و کنترل سرعت است؛ و خوشه ٤ می‌تواند به عنوان گروه مرجع برای مقایسه محورها استفاده شود.کاربرد مدیریتی این مقاله در طراحی بسته‌های سیاستی مبتنی بر تیپ محور است. به جای اعمال یک سیاست یکسان برای همه محورها، می‌توان برای هر خوشه بسته مداخله‌ای متفاوت طراحی کرد: در خوشه پرریسک، تابلوهای هشدار فاصله، کنترل هوشمند فاصله و اطلاع‌رسانی ساعات پرتخلف؛ در خوشه سرعتی، کنترل سرعت و بازنگری حد سرعت مجاز؛ در خوشه تجاری، پایش حمل‌ونقل سنگین و مدیریت سبقت؛ و در خوشه کم‌ریسک، پایش نگهدارنده و استفاده به عنوان الگوی مقایسه.

جدول ٤. سه محور شاخص در هر خوشه بر اساس نرخ تخلف فاصله غیرمجاز

تردد کل	نرخ سرعت/۱۰۰۰	نرخ فاصله/۱۰۰۰	نام محور	کد محور	برچسب خوشه	خوشه
۲۰،۱۸۰،۸۲۱	۱٤۲٫۵۵	٤۷۱٫۸۵	مشهد – چناران (مشهد – تقاطع آزادراه)	۳۱۳۱۰۳	پرریسک با تخلف فاصله بالا	۱
۳،۵۲۸،۳۸۱	۱۸٫۹۵	٤٤٦٫٤۹	فریمان – مشهد (عوارضی فریمان)	۳۱۳۳۵۲	پرریسک با تخلف فاصله بالا	۱
۱۷،۰۹٤،٤۲۲	۳٦٫۲۰	٤۰٤٫۹٤	چناران – مشهد (تقاطع آزادراه – مشهد)	۳۱۳۱۵۳	پرریسک با تخلف فاصله بالا	۱
۱،٦۹۱،۹۲۳	۱۱٦٫۷۵	۲۵٦٫۲٦	بردسکن – سبزوار (سبزوار)	۳۱٤۳۵۱	پرتردد با ریسک رفتاری متوسط	۲
۲،۱۷٦،٤۱٤	۱٤٫۰۳	۲٤۹٫۳۱	تایباد – تربت‌جام (تربت‌جام)	۳۱۳۷۵۱	پرتردد با ریسک رفتاری متوسط	۲
۱،٦۳۲،۵۰٦	۲۱٫۲۸	۲٤۸٫۸٦	اسفراین – سبزوار (سبزوار)	۳۱٤۳۵٤	پرتردد با ریسک رفتاری متوسط	۲
۱،٤۱٦،۸٦۰	۳۵۹٫۳۵	۱۹۳٫۹٦	گناباد – قائن (گناباد)	۳۱۵۵۰۲	تجاری/سنگین با سرعت کنترل‌شده	۳
۵،۷۳٤،۳۷۵	۵۸۵٫۹۰	۱۵٤٫۹۰	تربت حیدریه – مشهد (رباط سنگ – باغچه)	۳۱۳۸۵۲	تجاری/سنگین با سرعت کنترل‌شده	۳
۳،٦۸٤،۳۷۷	٤۳٦٫۲۸	۱٤۸٫۳٦	مشهد – تربت حیدریه (کمربندی تربت حیدریه)	۳۱۳۲۷٤	تجاری/سنگین با سرعت کنترل‌شده	۳
۱،٤٤٤،۰۹۰	۱۹٦٫٤۲	۲٤٤٫۰۷	قائن – گناباد (گناباد)	۳۱۵۵۵۲	کم‌ریسک‌تر و پایدارتر	٤
۳،٤۷٤،٤۵۰	۹۲٫۰۷	۱۱۵٫۸۲	نیشابور – سبزوار (زعفرانیه)	۳۱٤۳۷۰	کم‌ریسک‌تر و پایدارتر	٤
۳،۵۲۲،۱٤۱	۲۱٫۷۱	۱۱۳٫٤٤	سبزوار – نیشابور (نیشابور)	۳۱۵۷۵۱	کم‌ریسک‌تر و پایدارتر	٤

۵- نتیجه‌گیری

این پژوهش یک چارچوب داده‌محور برای خوشه‌بندی محورهای برون‌شهری استان خراسان رضوی با استفاده از داده‌های سامانه ۱۴۱ ارائه کرد. نتایج نشان داد چهار تیپ رفتاری قابل تفسیر در شبکه وجود دارد و خوشه‌بندی می‌تواند مکمل مناسبی برای مدل‌های پیش‌بینی تخلف و شاخص‌های ریسک باشد. مهم‌ترین نتیجه کاربردی پژوهش آن است که محورهای یک استان از نظر رفتار ترافیکی و تخلفات همگن نیستند و برنامه‌ریزی ایمنی باید بر اساس تیپ رفتاری محور انجام شود. بر اساس یافته‌ها، خوشه پرریسک با تخلف فاصله بالا، اولویت نخست برای بررسی میدانی و مداخلات مدیریتی است.

پیشنهادهای اجرایی

پیشنهادهای اجرایی عبارت‌اند از: ۱) به‌روزرسانی ماهانه خوشه‌بندی محورها برای پایش تغییر رفتار شبکه؛ ۲) طراحی داشبورد مدیریتی برای نمایش خوشه هر محور؛ ۳) تعریف برنامه کنترل فاصله برای خوشه پرریسک؛ ٤) اتصال نتایج خوشه‌بندی به داده‌های تصادفات برای اعتبارسنجی ایمنی؛ ۵) استفاده از خوشه‌ها برای تخصیص بهینه منابع راهداری و پلیس راه؛ و ٦) تعریف سناریوهای مداخله متفاوت برای هر خوشه.

٦-مراجع

-Abduljabbar, R., et al, (2025). Machine learning traffic flow prediction models for smart cities: Review and applications. *Infrastructures,* 7-10.

-Caliński, T., & Harabasz, J (2024). A dendrite method for cluster analysis. *Communications in Statistics*3(1), 1-27.

-Chai, A. B. Z., et al, (2024). Enhancing road safety with machine learning: Current state and future directions. *Engineering Applications of Artificial Intelligence*.

-Clustering Algorithms to Analyse Smart City Traffic Data, (2024). *International Journal of Advanced Computer Science and Applications,* (15)8.

-Davies, D. L., & Bouldin, D. W. A. (1979). cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* PAMI 2(1), 224-227.

-Etemad, S., et al, (2023).Clustering of urban traffic patterns by K-Means and Dynamic Time Warping: Case study. *arXiv 230909830.*

-Federal Highway Administration,Road Safety Fundamentals: *Measuring Safety* (2024).

-Federal Highway Administration (2008). Surrogate Safety Assessment Model and Validation Final Report.

-Fredriksson, H, (2025). Exploring spatio-temporal traffic performance variation in road networks. *Transportation Research\Preprint Report*.

-Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J,K (2023). means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Information Sciences 622, 178-220.

-International Transport Forum / OECD, Road Safety Annual Report (2024).

-Jain, A. K, :Data clustering 50 years beyond K-means. *Pattern Recognition Letters,* 31(18), 651-666.

-Jolliffe, I. T. .Principal Component Analysis (2002). *Springer,*

-Kwon, Y., Lee, M., Lee, M. J., & Son, S.-W, Cluster formations of free and congested flows in urban road networks. arXiv, 240808122.

-Lloyd, S,Least squares quantization in PCM. ,IEEE Transactions on Information Theory28(2), 129-137.

-MacQueen, J, (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium*,

-Morissette, L., & Chartier, S. (2013). The k-means clustering technique: General considerations and implementation in Mathematica. *Tutorials in Quantitative Methods ,for Psychology,* (9)1, 15-24.
-Pavlyshyn, V., et al, (2025). An adaptive machine learning approach to traffic pattern recognition. *Smart Cities,* (5)4, 151-152.
-Pedregosa, F., et al, (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, (12), 2825-2830.
-Rousseeuw, P. J,Silhouettes , (1987). A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics,* 20-53-65.
-Sohail, A. M. (2024). Unravelling traffic dynamics with K-means clustering and data preprocessing. *Information Systems and Smart Cities*.
-Ward, J. H., (2022). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*.
-World Health Organization,Global status report on road safety(2023).
- Xu, D., & Tian, Y(2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*(2), 165-193.
-Yumak, A. (2025). A machine learning approach to identify high-risk road segments. *Applied Sciences*(15)25, 1224-1225.

Clustering of Intercity Axes Based on Traffic Patterns, Speed, Fleet Composition, and Driving Violations Using Traffic Count Data (Case Study: Razavi Khorasan Province, 2025)

Sahar Alishahi, M.Sc., Grad., Department of Civil Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Iran.

Seyed Ali Ziaee, *Associate Professor, Department of Civil Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Iran.*

Hosein Mohammadi, M.Sc., Grad., Department of Civil Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Iran.

E-mail: sa-ziaee@um.ac.ir

Received: March 2026- Accepted: August 2026

ABSTRACT

Safety management and operation of intercity roads require understanding the behavioral differences among network corridors. Simple ranking of high-risk corridors, while useful for prioritization, fails to reveal similar behavioral patterns in terms of traffic volume, speed, fleet composition, and violation profiles. This study aims to cluster the intercity corridors of Razavi Khorasan Province based on traffic count data from the 141 system in the year 1404 (2025–2026). In this study, 136 unidirectional corridors were included in the analysis after data quality control and aggregation of corridor-level features. The features used included annual traffic volume, following-distance violation rate, speed violation rate, overtaking violation rate, weighted average speed, and the share of commercial/heavy vehicles. To reduce scaling effects, the data were standardized, and Principal Component Analysis was employed to visualize the data structure. Subsequently, two approaches—K-Means and Ward's hierarchical clustering—were evaluated for cluster counts ranging from 2 to 7. Based on the Silhouette index, k=4 with a coefficient of 0.269 provided the best K-Means configuration. The results revealed four behavioral types: "high-risk with high following-distance violations," "high-traffic with moderate behavioral risk," "commercial/heavy with controlled speed," and "lower-risk and more stable." The first cluster, comprising 47 corridors, had an average following-distance violation rate of 205.55 violations per 1,000 vehicles and a weighted rate of 244.92, placing it as the top priority for safety monitoring and intervention. The findings of this study indicate that data-driven clustering can serve as a suitable complement to risk indices and predictive models, assisting road authorities in managing similar corridors through more coordinated policies.

Keywords: Traffic Clustering, K-Means, Traffic Counting, Unsafe Following Distance Violation, Fleet Composition