تحلیل ساختار ترافیکی جاده‌های استان چهارمحال‌وبختیاری با استفاده از رویکردهای داده‌کاوی

وحیده احراری[1*]، رباب افشاری[2]

استادیار گروه علوم کامپیوتر، دانشکدهٔ علوم ریاضی، دانشگاه شهرکرد، شهرکرد، ایران[1*]

استادیار گروه آمار، دانشکدهٔ علوم، دانشگاه زنجان، زنجان، ایران[2]

چکیده

مدیریت ترافیک جاده‌ای یکی از چالش‌های اساسی در نظام حمل‌ونقل است که تأثیر مستقیمی بر ایمنی عمومی، بهره‌وری اقتصادی و پایداری محیط‌زیست دارد. استان چهارمحال‌وبختیاری به‌دلیل موقعیت جغرافیایی راهبردی خود، نقش مهمی در اتصال مناطق مختلف کشور ایفا می‌کند؛ ازاین‌رو، تحلیل دقیق ساختار ترافیکی جاده‌های این استان گامی ضروری برای ارتقای کیفیت برنامه‌ریزی‌ها و تصمیم‌گیری‌های مبتنی بر داده در حوزهٔ حمل‌ونقل محسوب می‌شود. در این پژوهش، داده‌های ترددشمار جاده‌ای استان در شهریورماه ۱۴۰۲ مورد استفاده قرار گرفته‌اند. این داده‌ها شامل متغیرهای کلیدی نظیر میانگین تردد وسایل نقلیه در پنج طبقهٔ مختلف، میزان تخلفات سرعت غیرمجاز، فاصلهٔ غیرمجاز، سبقت غیرمجاز و سرعت متوسط در محورهای جاده‌ای هستند. تحلیل داده‌ها بر پایهٔ سه الگوریتم یادگیری بدون نظارت شامل خوشه‌بندی k- میانگین، خوشه‌بندی سلسله‌مراتبی و رمزگذار خودکار انجام شده‌است. به‌منظور ارزیابی دقت و عملکرد این الگوریتم‌ها در خوشه‌بندی محورها، از سه شاخص معتبر سیلوئت، دیویس-بولدین و کالینسکی-هاراباسز بهره گرفته شده‌است. یافته‌ها نشان می‌دهند که مدل‌های رمزگذار خودکار و خوشه‌بندی سلسله‌مراتبی در مقایسه با روش سنتی k -میانگین، دسته‌بندی دقیق‌تری از جاده‌ها ارائه داده و ساختار ترافیکی پنهان را بهتر آشکار می‌سازند؛ بر اساس نتایج، جاده‌های استان به دو خوشهٔ متمایز تقسیم شده‌اند: خوشهٔ نخست شامل محورهایی با بالاترین نرخ تردد و بیشترین میزان تخلفات سرعت و فاصلهٔ غیرمجاز است که نشانه‌ای از رفتارهای پرخطر رانندگی و ریسک بالای تصادف محسوب می‌شود؛ درحالی که خوشهٔ دوم محورهایی با الگوهای ترافیکی ایمن‌تر را در برمی‌گیرد. نوآوری اصلی این پژوهش در ترکیب هم‌زمان چند الگوریتم پیشرفتهٔ خوشه‌بندی و استفاده از شاخص‌های متنوع ارزیابی عملکرد است که موجب افزایش دقت تحلیل و استحکام نتایج شده‌است؛ افزون‌بر این، تلفیق متغیرهای فنی و رفتاری ترافیکی در یک چهارچوب داده‌محور، امکان استخراج بینش‌های عمیق‌تری از الگوهای ترافیکی را فراهم ساخته است. چهارچوب ارائه‌شده، قابلیت تعمیم به سایر مناطق کشور را دارد و می‌تواند به‌عنوان الگویی نوین در هوشمندسازی مدیریت شبکه‌های جاده‌ای به‌کار گرفته شود. پژوهش حاضر با فراهم‌سازی ابزارهای تحلیلی دقیق و عملیاتی، می‌تواند به ارتقای آگاهی تصمیم‌گیرندگان، بهینه‌سازی تخصیص منابع، طراحی استراتژی‌های ایمنی و درنهایت کاهش نرخ تصادفات منجر شود. به‌منظور تکمیل و توسعهٔ این مسیر پژوهشی، پیشنهاد می‌شود مطالعات آتی به تحلیل‌های زمانی (فصلی)، ادغام متغیرهای انسانی، محیطی و توسعهٔ مدل‌های پیش‌بینی ترکیبی در حوزهٔ حمل‌ونقل بپردازند.

واژگان کلیدی: یادگیری بدون نظارت، خوشه‌بندی، تحلیل ساختار ترافیکی، رمزگذار خودکار، سرعت غیرمجاز، فاصلهٔ غیرمجاز، استان چهارمحال‌وبختیاری.

Analysis of the traffic structure of the roads of Chahar Mahal and Bakhtiari province using data-mining approaches

Vahideh Ahrari[1*], Robab Afshari [2]

Assistant Professor, Department of Computer Science, Faculty of Mathematical Sciences, Shahrekord University, Shahrekord, Iran[*1]

Assistant Professor, Department of Statistics, Faculty of Sciences, University of Zanjan, Zanjan, Iran[2]

[*] Corresponding author

[*] نویسندهٔ عهده‌دار مکاتبات

• تاریخ ارسال مقاله: ۱۴۰۳/۵/۹ • تاریخ پذیرش: ۱۴۰۴/۴/۳۰ • تاریخ انتشار: ۱۴۰۴/۹/۲۸ • نوع مطالعه: پژوهشی

Abstract

Road traffic management is a fundamental and multidimensional challenge within transportation systems, exerting a direct impact on public safety, economic efficiency, and environmental sustainability. The province of Chaharmahal and Bakhtiari, due to its strategic geographical location, plays a critical role in connecting various regions of Iran. Therefore, a precise analysis of the traffic structure of this province's roadways is essential for improving the quality of data-driven planning and decision-making in the transportation sector. In this study, road traffic-counter data collected during September were utilized. These data include key variables such as the average number of vehicles in five different classes, instances of traffic violations (namely speeding, tailgating, and illegal overtaking), as well as the average speed on various road segments. The data were analyzed using three unsupervised learning algorithms: k-means clustering, hierarchical clustering, and autoencoder-based clustering. To assess the accuracy and performance of these algorithms in segment clustering, three well-established evaluation metrics—Silhouette score, Davies-Bouldin index, and Calinski-Harabasz index—were employed. The results demonstrate that the autoencoder and hierarchical clustering models offer a more accurate classification of road segments compared to the conventional k-means method, revealing latent traffic structures more effectively. Based on the findings, the roads in the province were categorized into two distinct clusters: the first cluster includes segments with the highest traffic volume and the highest rates of speeding and tailgating violations—indicative of risky driving behaviors and elevated accident risk. The second cluster encompasses segments characterized by safer traffic patterns. The primary contribution of this research lies in the integrated application of multiple advanced clustering algorithms alongside diverse performance evaluation metrics, which significantly enhance the precision and robustness of the analysis. Furthermore, the combination of technical and behavioral traffic variables within a unified data-driven framework enables the extraction of deeper insights into traffic behavior patterns. The proposed framework is generalizable to other regions and can serve as a novel model for the intelligent management of roadway networks.By providing accurate and actionable analytical tools, this study has the potential to support decision-makers in raising awareness, optimizing resource allocation, designing safety strategies, and ultimately reducing accident rates. To extend and enrich this line of research, future studies are encouraged to incorporate temporal (e.g., seasonal) analyses, integrate human and environmental variables, and develop hybrid predictive models in the transportation domain.

Keywords: Unsupervised learning, clustering, traffic structure analysis, autoencoder, speeding violation, tailgating, Chaharmahal and Bakhtiari Province.

۱-مقدمه

مدیریت ترافیک یک جنبهٔ پیچیده و حیاتی در زیرساخت‌های شهری و روستایی است؛ بر اساس گزارش سازمان جهانی بهداشت[1] (WHO)، تصادفات جاده‌ای سالانه منجر به مرگ ۱/۱۹ میلیون نفر می‌شود که آن را به‌عنوان یکی از علل اصلی مرگ‌ومیر در سراسر جهان تبدیل می‌کند؛ به‌طوری که آسیب‌های ترافیکی جاده‌ای اکنون عامل اصلی قتل افراد پنج تا ۲۹ ساله است[۱]. جریان ترافیک در جاده‌ها شامل انواع وسایل نقلیه، چالش‌های مهمی برای ایمنی جاده‌ها و حمل‌ونقل کارآمد ایجاد می‌کند. سرعت غیرمجاز، رعایت‌نکردن فاصلهٔ مجاز و سبقت‌های نادرست این چالش‌ها را تشدید کرده و خطر تصادف و ازدحام را افزایش می‌دهد. پرداختن به این مسائل نیازمند رویکردی جامع است که باید عوامل مختلفی مانند طراحی راه، رفتار راننده و اجرای مقررات راهنمایی و رانندگی در آن گنجانده شوند؛ علاوه‌بر این، ادغام سامانه‌های مدیریت ترافیک پیشرفته و ارتقای آگاهی عمومی در مورد شیوه‌های رانندگی ایمن نیز از عناصر حیاتی در مدیریت مؤثر ترافیک هستند[۲].

استفاده از روش‌های یادگیری بدون نظارت برای تجزیه‌وتحلیل و درک الگوهای ترافیک، یک حوزهٔ پژوهشی فعال در سال‌های اخیر بوده‌است. در زمینهٔ روش‌های خوشه‌بندی بدون نظارت، الگوریتم‌های مختلفی در تحلیل داده‌ها به‌کار رفته‌اند که در بهبود دقت پیش‌بینی و پاک‌سازی داده‌ها نقش مؤثری داشته‌اند؛ برای مثال، کریمی و خدابخش [۳] از خوشه‌بندی بدون نظارت برای پیش‌بینی عملکرد پرس‌وجوها در موتورهای جست‌وجو استفاده کردند که بهبود دقت پیش‌بینی را نشان داد؛ همچنین، دانشپور و برزگری [۴] با ارائهٔ یک روش خوشه‌بندی سلسله‌مراتبی مبتنی بر توابع تشابه، دقت تشخیص رکوردهای تکراری را افزایش دادند.

دولت‌ها و مقامات حوزهٔ حمل‌ونقل به‌طور فزاینده‌ای از فناوری‌های پیشرفته مانند الگوریتم‌های خوشه‌بندی جاده‌ها برای بهبود مدیریت ترافیک استفاده می‌کنند. این الگوریتم‌ها داده‌های جریان ترافیک را برای شناسایی الگوهای بخش‌های جاده با ویژگی‌های ترافیکی مشابه تجزیه‌وتحلیل می‌کنند. این دانش می‌تواند فرایندهای تصمیم‌گیری در مورد مدیریت ترافیک، برنامه‌ریزی زیرساخت و تخصیص منابع را بهبود بخشد. با استفاده از این الگوریتم‌ها، دولت‌ها می‌توانند دربارهٔ زیرساخت‌های جاده‌ای، بهینه‌سازی سیگنال‌های ترافیکی و استراتژی‌های اجرایی برای افزایش ایمنی جاده‌ها و کاهش ازدحام، تصمیمات آگاهانه بگیرند[۵]. پژوهش‌های مختلفی برای بررسی ترافیک جاده‌ای بر اساس یادگیری ماشین انجام شده‌است که در ادامه به برخی از آن‌ها اشاره می‌شود.

وانگ و بیلجکی [۶] مروری سازمان‌یافته بر استفاده از روش‌های یادگیری ماشین بدون نظارت را در مطالعات

[1] World Health Organization

شهری ارائه کردند. آن‌ها ۹۳ مطالعهٔ منتشرشده بین سال‌های ۲۰۱۰ و ۲۰۲۰ را بررسی و داده‌های مربوطه را تجزیه‌وتحلیل کردند. در پژوهشی دیگر، براتساس و همکاران [۷] چگونگی عملکرد مدل‌های مختلف یادگیری ماشین مانند رگرسیون خطی، درخت‌های تصمیم‌گیری و شبکه‌های عصبی را در پیش‌بینی سرعت ترافیک شهری بررسی کردند و نشان دادند که مدل‌های پیشرفته، به‌ویژه مدل‌هایی که از یادگیری عمیق استفاده می‌کنند، از روش‌های سنتی در پیش‌بینی دقیق سرعت ترافیک تحت شرایط پیچیدهٔ ترافیک شهری بهتر عمل می‌کنند.

کوی و همکاران [۸] یک روش خوشه‌بندی مبتنی بر چگالی را با همسایگی فازی برای شناسایی الگوی مرکز اصلی در شبکه‌های جاده‌ای معرفی و اثربخشی این روش را بر داده‌های شبکهٔ جاده‌ای در دنیای واقعی بررسی کردند. آن‌ها نشان دادند این روش در مقایسه با الگوریتم‌های خوشه‌بندی سنتی کارایی بهتری دارد. یک سیستم شبکهٔ عصبی فازی ترکیبی برای کنترل ترافیک جاده‌ای کارآمد توسط آگاروال و همکاران [۹] ارائه شد. نتایج شبیه‌سازی آن‌ها نشان داد که رویکرد عصبی فازی یک‌پارچه از روش‌های کنترل ترافیک سنتی بهتر عمل می‌کند و منجر به بهبود جریان ترافیک و کاهش تراکم می‌شود؛ نویسندگان همچنین نشان دادند که روش پیشنهادی چهارچوبی قوی و سازگار برای مدیریت حمل‌ونقل هوشمند در شبکه‌های جاده‌ای پیچیده فراهم می‌کند. لیو و وو [۱۰] از یک مدل جنگل تصادفی برای پیش‌بینی تراکم ترافیک جاده‌ای استفاده کردند و نشان دادند که رویکرد جنگل تصادفی از روش‌های سنتی در پیش‌بینی تراکم ترافیک بهتر عمل می‌کند. اختر و مریدپور [۱۱] مروری جامع بر استفاده از روش‌های هوش مصنوعی برای پیش‌بینی تراکم ترافیک داشتند. نویسندگان، رویکردهای مختلف مبتنی بر هوش مصنوعی ازجمله مدل‌های یادگیری ماشین مانند شبکه‌های عصبی، ماشین‌های بردار پشتیبان، درخت تصمیم را بررسی کرده و بینش‌هایی را در مورد جهت‌گیری‌های پژوهشی آینده با استفاده از هوش مصنوعی پیشرفته برای برنامه‌های حمل‌ونقل هوشمند ارائه کردند. یانگ و همکاران [۱۲]، با استفاده از تحلیل خوشه‌بندی طیفی به بررسی الگوهای تغییرات زمانی ترافیک در یک شبکهٔ جاده‌ای شهری پرداخته و نشان دادند که خوشه‌های متمایز از بخش‌های جاده‌ای، الگوهای زمانی منسجمی از تغییرات وضعیت ترافیک را نشان می‌دهد. به‌تازگی، هانگ و همکاران [۱۳] برای تشخیص خطاها در داده‌های ترافیک رویکردی جدید مبتنی بر خوشه‌بندی خودکدگذار بر اساس فرایند بازسازی داده‌ها ارائه کردند؛ همچنین، شین و همکاران [۱۴] برای بررسی تأثیر رفتار رانندگی انسان در امر ترافیک از خوشه‌بندی مبتنی بر الگوریتم خودکدگذار و ارزیابی خطا استفاده کردند.

استان چهارمحال‌وبختیاری که در جنوب غربی ایران واقع شده‌است، از شمال و مشرق با استان اصفهان، از جنوب با استان کهگیلویه و بویراحمد، از مغرب با استان خوزستان و از شمال غرب با استان لرستان هم‌مرز است. این موقعیت جغرافیایی ویژه، جاده‌های استان را به یکی از شریان‌های اصلی ارتباطی کشور تبدیل کرده‌است؛ جاده‌هایی که نه‌تنها در توسعهٔ اقتصادی استان از طریق حمل‌ونقل کالا نقش مؤثری دارند، بلکه با ایجاد تعاملات فرهنگی، اجتماعی و گردشگری میان مردم منطقه و سایر نقاط کشور، به تقویت روابط منطقه‌ای ایران کمک می‌کنند. با توجه به ناهمواری‌های جغرافیایی استان و بار ترافیکی قابل‌توجه در محورهای ارتباطی، بررسی وضعیت این جاده‌ها از منظر نوع و چگونگی ترافیک، امری ضروری است. این تحلیل داده‌محور می‌تواند به مسئولین کمک کند تا در مدیریت بهینهٔ ترافیک جاده‌ای، تأمین زیرساخت‌ها، تخصیص منابع و بهبود حمل‌ونقل کارآمد در این منطقه گام‌هایی مهم، هدفمند و مثبت بردارند. این پژوهش با ترکیب سه الگوریتم خوشه‌بندی شامل K - میانگین، خوشه‌بندی سلسله‌مراتبی و رمزگذار خودکار و ارزیابی عملکرد آن‌ها با بهره‌گیری از شاخص‌های سیلوئت، دیویس-بولدین و کالینسکی-هاراباسز، در پی شناسایی الگوهای پنهان ترافیکی و دسته‌بندی محورهای جاده‌ای استان چهارمحال‌وبختیاری به دو گروه پرخطر و ایمن است. بر اساس بررسی پیشینهٔ موجود و مطالعات انجام‌شده توسط نویسندگان، تاکنون پژوهش جامع و داده‌محوری در این زمینه و در این منطقهٔ خاص صورت نگرفته است؛ ازاین‌رو، هدف این مطالعه تحلیل ساختار ترافیکی جاده‌های استان باتکیه‌بر سه رویکرد مذکور در خوشه‌بندی است؛ رویکردی که می‌تواند با ارائهٔ تصویری دقیق‌تر از رفتارهای ترافیکی در محورهای مختلف، بستر مناسبی برای ارتقای آگاهی تصمیم‌گیرندگان فراهم آورده و زمینه‌ساز تدوین راهکارهای کاربردی و هدفمند در مدیریت ایمن و هوشمند ترافیک منطقه شود. ساختار کلی مقاله به‌صورت زیر است: پس از معرفی داده‌ها و روش کار، الگوریتم‌های خوشه‌بندی و شاخص‌های ارزیابی عملکرد آن‌ها به‌ترتیب در بخش‌های دو، سه و چهار، به تحلیل توصیفی داده‌ها در بخش پنج پرداخته می‌شود.

بخش ششم به بررسی ساختار ترافیکی جاده‌های مربوط به استان چهارمحال‌وبختیاری بر اساس الگوریتم‌های خوشه-بندی و ارزیابی دقت آن‌ها اختصاص دارد. در پایان، نتایج حاصل در بخش هفت آورده می‌شود.

۲- داده‌ها و روش کار

با توجه به موضوع و هدف پژوهش، رویکرد مطالعه ترکیبی از رویکردهای توصیفی-تحلیلی است که به‌صورت مقطعی و

بر پایهٔ الگوریتم‌های یادگیری ماشین انجام شده‌است، پس از مطالعات نظری و آگاهی نسبی از پژوهش‌های انجام‌شده، با مراجعه به تارنمای مرکز مدیریت راه‌های کشور[1] ، اطلاعات ترددشمار و داده‌های مربوط به جاده‌های استان چهارمحال‌وبختیاری در شهریورماه سال ۱۴۰۲ دریافت شد.

این اطلاعات مربوط به داده‌های ثبت‌شده در شهریورماه سال ۱۴۰۲ در ۶۳ محور تردد شمار در جاده‌های استان چهارمحال‌وبختیاری است. شایان ذکر است که اطلاعات تردد شمارها از سال ۱۳۸۹ تاکنون به‌وسیلهٔ سازمان راهداری به‌صورت برخط از سامانهٔ ترددشماری برداشت می‌شود و در پایان هر ماه این اطلاعات به‌صورت برون‌خط به‌طور روزانه و ساعتی بر روی تارنمای مذکور قرار می‌گیرد. منظور از هر محور ترددشمار که با یک کد شش رقمی مشخص می‌شود یک جهت رفت یا برگشت مانند آزادراه تهران – قم (نعلبندان) است. متغیرهای ثبت‌شده در این مجموعه‌داده شامل نام محور، نوع یا دستهٔ وسیلهٔ نقلیهٔ عبوری، تعداد تخلف سرعت غیرمجاز، تعداد تخلف فاصلهٔ غیرمجاز، تعداد تخلف سرعت غیرمجاز و سرعت متوسط ثبت‌شده مربوط به ۶۳ محور ترددشمار است که اطلاعات ثبت‌شده مربوط به آن‌ها به‌طور خلاصه در پیوست آورده شده‌است.

در این مجموعه‌داده، متغیر نوع یا دستهٔ وسیله نقلیه شامل پنج دسته به‌شرح دستهٔ یک نشان‌دهندهٔ سواری و وانت، دستهٔ دو نشان‌دهندهٔ کامیونت و کامیون‌های کوچک و مینی‌بوس، دستهٔ سه نشان‌دهندهٔ کامیون‌های معمولی کمتر از ده متر و سه‌محوره‌ها، دستهٔ چهار مبین اتوبوس و درنهایت دستهٔ پنج نشان‌دهندهٔ تریلرها و باربرهای بالاتر از سه محور است.

متغیر متوسط سرعت، بر اساس میانگین سرعت همهٔ انواع وسایل نقلیهٔ عبوری در محور مربوطه ثبت شده‌است. منظور از تعداد تخلف فاصلهٔ غیرمجاز، تعداد خودروهایی است که فاصلهٔ زمانی دست‌کم دو ثانیه با خودروی پیشین را بر اساس سرعت وسیلهٔ نقلیه در محور مربوطه رعایت نکرده‌اند. منظور از تعداد تخلف سرعت غیرمجاز، تعداد وسایل نقلیه در یک بازه زمانی در محور مربوطه است که در آن محور سرعتی بالاتر از سرعت مجاز داشته‌اند؛ به‌دلیل اینکه این اطلاعات به‌صورت روزانه به‌وسیلهٔ سازمان راهداری ثبت می‌شوند، در پژوهش مورد نظر متوسط هر متغیر در ماه برای هر مسیر محاسبه و تحلیل‌ها بر اساس آن انجام شده‌اند. این متغیرها به‌طور خلاصه در جدول (۱) آورده شده‌است. در ادامه با بهره‌گیری از تحلیل توصیفی، ارزیابی اولیه از وضعیت تردد خودروها در منطقهٔ مورد نظر و در محورهای یادشده به‌عمل آمد؛ همچنین به‌منظور تشخیص و تعیین محورهای تردد با الگوهای مشابه، از مدل‌های یادگیری ماشین با تمرکز بر روش‌های غیرنظارتی استفاده شد؛ به‌منظور استخراج جاده‌های با رفتارهای مشابه و یا به‌عبارت‌دیگر تعیین خوشه‌ها، از سه رویکرد k-میانگین[2]، رمزگذار خودکار[3] و سلسه‌مراتبی[4] روی مجموعه‌داده استفاده شده‌است. در پایان، به‌منظور مقایسهٔ عملکرد این سه رویکرد در خوشه‌بندی داده‌ها و ارزیابی دقت آن‌ها نسبت به‌هم، از برخی شاخص‌های معروف شامل شاخص ضریب نیم‌رخ یا سیلوئت[5]، شاخص دیویس-بولدین[6] و شاخص کالینسکی-هاراباسز[7] بهره گرفته شده‌است. لازم به‌ذکر است در مطالعهٔ حاضر از نرم‌افزارهای پایتون[8] و تبلو[9] برای تحلیل داده‌ها استفاده شده‌است.

۳- الگوریتم‌های خوشه‌بندی

یکی از روش‌های یادگیری ماشین از نوع بدون نظارت، روش خوشه‌بندی است که در آن ساختار داده‌ها بدون داشتن برچسب‌هایی ازپیش تعیین‌شده بررسی می‌شود؛ یکی از کاربردهای این روش، کشف الگوها و گروه‌بندی طبیعی داده‌هاست. برای اجرای این نوع از روش یادگیری ماشین، الگوریتم‌های مختلفی توسط پژوهش‌گران ارائه شده‌است که ازجملهٔ آن‌ها می‌توان به سه روش سلسله‌مراتبی، k- میانگین و رمزگذاری خودکار اشاره کرد که در ادامه به معرفی آن‌ها پرداخته می‌شود.

۳-۱- خوشه‌بندی سلسله‌مراتبی

خوشه‌بندی سلسله‌مراتبی یک روش قدرتمند در پژوهش‌های کلان‌داده و داده‌کاوی است که برای ایجاد سلسله‌مراتبی از خوشه‌ها استفاده می‌شود؛ بر خلاف سایر روش‌های خوشه‌بندی که تعداد خوشه‌های ازپیش تعیین‌شده را ارائه می‌کنند، خوشه‌بندی سلسله‌مراتبی نمایش دقیقی از ساختار زیربنایی داده‌ها را از طریق یک نمودار درخت‌مانند به نام دندروگرام ارائه می‌کند. خوشه‌بندی سلسله‌مراتبی دو راهبرد اصلی دارد: تجمعی و تقسیمی. خوشه‌بندی تجمعی با هر مشاهده به‌عنوان یک خوشهٔ جداگانه شروع می‌شود و آن‌ها را در خوشه‌های بزرگ‌تر ادغام می‌کند تا زمانی که همهٔ نقاط داده در یک خوشهٔ واحد قرار گیرند؛ از سوی دیگر، خوشه‌بندی تقسیمی با تمام مشاهدات در یک خوشه شروع می‌شود و آن‌ها را به خوشه‌های کوچک‌تر تقسیم می‌کند[۱۵]. برای اطلاعات کامل‌تر در مورد خوشه‌بندی سلسله‌مراتبی، به کتاب ژانگ و همکاران [۱۶] مراجعه شود.

[1] https://141.ir

[2]. K-Means
[3]. Autoencoder
[4]. Hierarchical
[5] Silhouette coefficient
[6] Davis–Bouldin
[7] Calinski-Harabasz
[8] Python
[9] Tableau

۳-۲- خوشه‌بندی k- میانگین

خوشه‌بندی k - میانگین، یک الگوریتم یادگیری ماشینی بدون نظارت است که برای تقسیم یک مجموعه‌داده به k خوشهٔ مجزا استفاده می‌شود. این الگوریتم در ابتدا توسط استوارت لوید در سال ۱۹۵۷ پیشنهاد شد و در سال ۱۹۶۷ جیمز مک کوئین آن را [۱۷] اصلاح کرد. روش k - میانگین به‌دلیل سادگی و کارایی در زمینه‌های مختلفی مانند تشخیص الگو، تجزیه‌وتحلیل تصویر و داده‌کاوی استفاده می‌شود.

الگوریتم خوشه‌بندی k - میانگین با ایجاد خوشه‌های اولیه آغاز می‌شود و به‌دنبال آن نقاط داده به نزدیکترین مرکز خوشه تخصیص داده می‌شود، پس از آن، مراکز خوشه بر اساس نقاط اختصاص‌داده‌شدهٔ جدید، به‌روز می‌شوند. این چرخهٔ تخصیص نقاط و مراکز به‌روزرسانی تا زمانی ادامه می‌یابد که تخصیص‌های خوشه ثابت بماند یا تعداد تکرار ازپیش تعیین‌شده برآورده شود[۱۸].

برای کسب اطلاعات دقیق‌تر در مورد خوشه‌بندی k - میانگین، ازجمله زیربنای ریاضی و روش‌های مختلف اولیه‌سازی به استاینلی و بروسکو [۱۹] مراجعه شود.

(جدول- ۱): متغیرهای استفاده‌شده در تحلیل داده‌ها

(Table-1): Variables used in data analysis

نام متغیر	توضیحات
متوسط وسیلهٔ نقلیه دستهٔ یک (بر حسب ماه)	دستهٔ یک: سواری و وانت
متوسط وسیلهٔ نقلیهٔ دستهٔ دو (بر حسب ماه)	دستهٔ دو: کامیونت و کامیون‌های کوچک و مینی‌بوس
متوسط وسیلهٔ نقلیهٔ دستهٔ سه (بر حسب ماه)	دستهٔ سه: کامیون‌های معمولی کمتر از ده متر و سه‌محوره‌ها
متوسط وسیلهٔ نقلیهٔ دستهٔ چهار (بر حسب ماه)	دستهٔ چهار: اتوبوس
متوسط وسیلهٔ نقلیهٔ دستهٔ پنج (بر حسب ماه)	دستهٔ پنج: تریلرها و باربرهای بالاتر از سه محور
متوسط تخلف سرعت غیرمجاز (بر حسب ماه)	تعداد تخلف‌ها روزانه ثبت می‌شوند
متوسط تخلف فاصلهٔ غیرمجاز (بر حسب ماه)	تعداد تخلف‌ها روزانه ثبت می‌شوند
متوسط تخلف سبقت غیرمجاز (بر حسب ماه)	تعداد تخلف‌ها روزانه ثبت می‌شوند
سرعت متوسط (بر حسب ماه)	متوسط سرعت روزانه ثبت می‌شوند

۳-۳- شاخص دیویس-بولدین

رمزگذار خودکار نوعی معماری شبکهٔ عصبی است که برای کاهش ابعاد استفاده می‌شود. این رمزگذار به‌منظور فشرده‌سازی داده‌های ورودی برای نمایش در ابعاد پایین‌تر به‌نام فضای پنهان طراحی شده‌است. رمزگذار خودکار از دو بخش اصلی رمزگذار و رمزگشا تشکیل شده‌است؛ به‌طوری که رمزگذار، داده‌های ورودی را گرفته و آن‌ها را به یک نمایش با ابعاد پایین‌تر در فضای پنهان نگاشت می‌کند؛ سپس رمزگشا این نمایش را می‌گیرد و داده‌های اصلی را از طریق آن بازسازی می‌کند. هدف از رمزگذار خودکار، یادگیری نمایشی فشرده از داده‌های ورودی است که مهم‌ترین ویژگی‌ها را به تصویر می‌کشاند؛ به‌عبارت دیگر، هدف آن کاهش ابعاد داده‌ها در کنار حفظ اطلاعات ضروری است. منظور از کاهش ابعاد، کاهش تعداد ویژگی‌ها یا متغیرها در یک مجموعه‌داده است که بیشتر برای ساده‌کردن تفسیر و همچنین تسهیل محاسبات در تجزیه‌وتحلیل یا مدل‌سازی انجام می‌شود. چنانچه اشاره شد در رمزگذار خودکار، کاهش ابعاد با فشرده‌سازی داده‌های ورودی در یک نمایش با ابعاد پایین‌تر در فضای پنهان به‌دست می‌آید. فضای نهفته، فضای پنهان نمایشی با ابعاد پایین‌تر از داده‌های ورودی است که به‌وسیلهٔ رمزگذار خودکار آموخته می‌شود که می‌توان آن را به‌عنوان یک نسخهٔ فشرده از داده‌های اصلی در نظر گرفت. اندازهٔ فضای پنهان به‌طور معمول کوچکتر از ابعاد داده‌های ورودی است؛ در حقیقت رمزگذار خودکار یاد می‌گیرد تا داده‌های ورودی با ابعاد بزرگ‌تر را با حفظ بیشینهٔ ویژگی‌های مهم آن به فضای پنهان با ابعاد پایین‌تر نگاشت کند[۲۰]. در بیشتر موارد می‌توان از نمایش فشرده‌ای رمزگذار خودکار به‌عنوان ورودی یک الگوریتم خوشه‌بندی مانند k - میانگین استفاده کرد تا نقاط مشابه را بر اساس ساختار داده‌ها گروه‌بندی کند. رمزگذارهای خودکار با ارائهٔ نمایش معنادارتر و حفظ ساختار داده‌ها، دقت و قابلیت تفسیر نتایج خوشه‌بندی را افزایش می‌دهند[۲۱]. برای بررسی جامع‌تر رمزگذارهای خودکار در یادگیری عمیق، می‌توان به مقالهٔ برهمند و همکاران [۲۲] مراجعه کرد.

۴- شاخص‌های ارزیابی عملکرد الگوریتم‌های خوشه‌بندی

یکی از مهم‌ترین اقدامات پس از به‌کارگیری الگوریتم‌های خوشه‌بندی در بررسی ساختار داده‌ها، ارزیابی عملکرد و دقت الگوریتم‌های به‌کار گرفته‌شده است. برای این منظور شاخص‌های ارزیابی متفاوتی وجود دارد که در ادامه سه روش به‌کار گرفته‌شدهٔ معروف در این پژوهش معرفی می‌شوند.

۴-۱-شاخص ضریب نیم‌رخ

شاخص ضریب نیم‌رخ که به آن ضریب سیلوئت نیز گفته می‌شود، یک معیار پرکاربرد برای ارزیابی کیفیت نتایج

خوشه‌بندی است. فرض کنید مجموعه‌دادهٔ D با n مشاهده به k خوشه مانند $C_1, C_2, \dots, C_k$ افراز شده باشد. به‌ازای هر مشاهده $j \in D$ و $1 \le s \le k$ ضریب سیلوئت به‌صورت زیر تعریف می‌شود:

$$s(j) = \begin{cases} \frac{b(j)-a(j)}{\max\{b(j),a(j)\}}, & \text{if } |C_s| > 1 \\ 0 & \text{if } |C_s| = 1 \end{cases} \quad (۱)$$

که در آن $|C_s|$ تعداد مشاهدات در خوشهٔ sام، $a(j)$ متوسط فاصلهٔ بین مشاهدهٔ jام از تمام مشاهدات موجود در خوشه‌ای است که این مشاهده به آن تعلق دارد و $b(j)$ کمترین متوسط فاصلهٔ بین مشاهدهٔ jام و تمام مشاهدات موجود در هر خوشه‌ای است که این مشاهده به آن خوشه تعلق ندارد، یعنی:

$$a(j) = \frac{\sum_{j \in C_s, i \neq j} d(i,j)}{|C_s|-1} \quad (۲)$$

$$b(j) = \min_{C_l : 1 \le l \le k, l \neq s} \left\{ \frac{\sum_{i \in C_l} d(i,j)}{|C_l|} \right\} \quad (۳)$$

مقدار شاخص سیلوئت بین ۱- تا ۱+ تغییر می‌کند به‌طوری که مقدار نزدیک به یک بیان‌کنندهٔ این است که یک نقطه به‌خوبی با خوشهٔ خودش مطابقت داشته و با خوشه‌های مجاور هم‌خوانی ضعیفی دارد؛ به بیان دیگر، کوچک‌بودن مقدار ضریب سیلوئت، بیان‌کنندهٔ ضعیف‌بودن نتایج خوشه‌بندی است که ممکن است به علت انتخاب نامناسب تعداد خوشه‌ها (k) باشد[۲۳].

۴-۲-شاخص دیویس-بولدین

شاخص دیویس-بولدین معیار دیگری است که برای ارزیابی کیفیت الگوریتم‌های خوشه‌بندی استفاده می‌شود. این شاخص به‌صورت زیر تعریف می‌شود:

$$BI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i} \left(\frac{S_i + S_j}{D_{ij}} \right) \quad (۴)$$

که در آن S_i و S_j به‌ترتیب پراکندگی مشاهدات در درون خوشهٔ i و j بوده و D_{ij} پراکندگی بین این خوشه‌ها را نشان می‌دهد و داریم:

$$S_i = \sqrt{\frac{1}{|C_i|}\sum_{j \in C_i} d^2(j, v_i)} \quad (۵)$$

$$D_{ij} = \sqrt{\sum d^2(v_i, v_j)} \quad (۶)$$

که در آن v_i و v_j به‌ترتیب مراکز خوشهٔ i و j بوده و $d^2(.,.)$ بیان‌کنندهٔ فاصلهٔ اقلیدسی بین دو نقطه است.

مطابق تعریف، این شاخص متوسط بیشینهٔ نسبت پراکندگی درون خوشه‌ها را نسبت به پراکندگی بین خوشه‌ها بیان می‌کند؛ هرچه مقدار شاخص دیویس-بولدین کمتر باشد، نتیجهٔ خوشه‌بندی بهتر است؛ زیرا نشان می‌دهد که خوشه‌ها به‌خوبی از هم جدا شده و فشرده‌اند. این شاخص به‌ویژه برای تعیین تعداد بهینهٔ خوشه‌ها در یک مجموعه‌داده مفید است[۲۴].

۴-۳-شاخص کالینسکی-هاراباسز

شاخص کالینسکی-هاراباسز معیاری دیگر برای ارزیابی اعتبار خوشه‌بندی است که کیفیت آن را از طریق مقایسهٔ واریانس بین‌خوشه‌ای و واریانس درون‌خوشه‌ای ارزیابی می‌کند. این شاخص به‌صورت زیر تعریف می‌شود:

$$VRC = \frac{SSB/(k-1)}{SSw/(n-k)} \quad (۷)$$

که در آن k تعداد خوشه‌ها، n تعداد کل مشاهدات، SSB[1] و SSw[2] به‌ترتیب مجموع مربعات بین خوشه‌ای و مجموع مربعات درون خوشه‌ای است و داریم:

$$SSB = \sum_{i=1}^{k} |C_i| (m_i - m)^2 \quad (۸)$$

$$SSw = \sum_{i=1}^{k} \sum_{j \in C_i} (j - m_i)^2 \quad (۹)$$

که در آن $|C_i|$ تعداد مشاهدات واقع در خوشهٔ iام، m_i مرکز خوشهٔ iام و m میانگین کل مشاهدات است. مقدار بزرگ برای این شاخص نشان‌دهندهٔ کیفیت خوب خوشه‌بندی است؛ به‌عبارت دیگر، شاخص کالینسکی-هاراباسز با مقدار بزرگ نشان می‌دهد که خوشه‌ها به‌خوبی از یکدیگر متمایز شده‌اند. این شاخص به‌طور معمول در تحلیل خوشه‌ای برای تعیین تعداد بهینهٔ خوشه‌ها و ارزیابی اثربخشی الگوریتم‌های خوشه‌بندی استفاده می‌شود[۲۵].

۵- تحلیل توصیفی داده‌ها

چنانچه اشاره شد تعداد ۳۶ محور ترددشمار در استان چهارمحال‌وبختیاری وجود دارد. شکل‌های (۱) و (۲) به‌ترتیب نقشهٔ طرح کلی ترافیکی و طبقه‌بندی راه‌های استان چهارمحال‌وبختیاری را نشان می‌دهند.

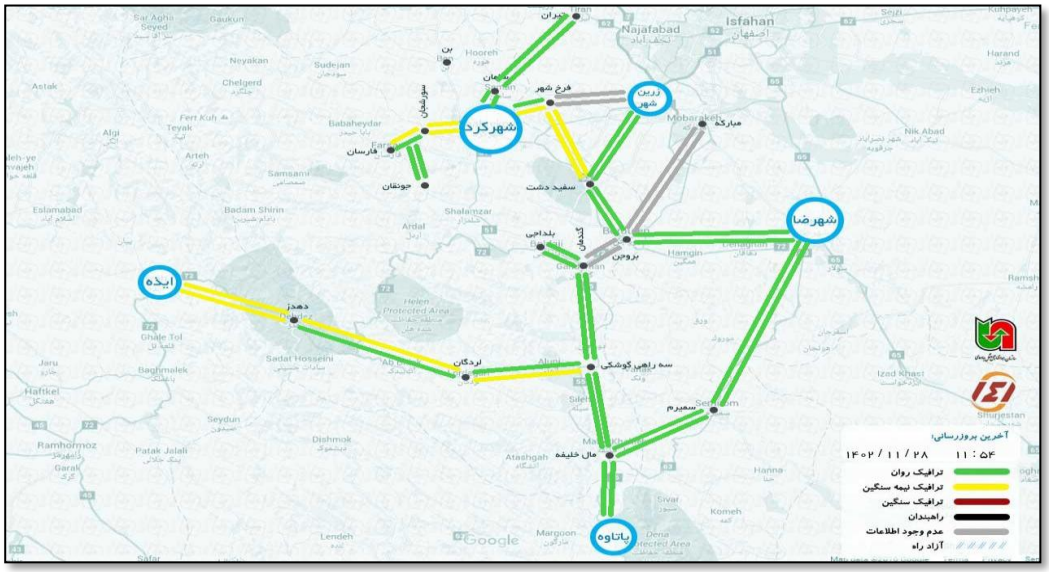
شکل (۳) سرعت متوسط و متوسط تخلف سرعت غیرمجاز را در شهریورماه سال ۱۴۰۲ برای محورهایی با بالاترین سرعت متوسط نشان می‌دهد؛ همان‌طور که مشاهده می‌شود، محورهای فارسان-سورشجان، شهرضا-بروجن (بروجن) و گوشکی-گندمان به‌ترتیب با سرعت‌های متوسط ۱۰۳/۶، ۹۸/۵ و ۹۶/۵ دارای بیشترین ثبت سرعت‌اند؛ همچنین، به‌طور متوسط بیشترین تعداد تخلف مربوط به سرعت غیرمجاز به‌ترتیب در محورهای فرخشهر-شهرکرد (۳۹۷۹ مورد) و فارسان- سورشجان (۲۸۱۸ مورد) بوده‌است.

شکل (۴) متوسط تعداد وسیلهٔ نقلیه مربوط به دستهٔ یک را که در شهریورماه سال ۱۴۰۲ در محورهای یادشده تردد داشته‌اند، نشان می‌دهد. بر اساس این شکل، به‌ترتیب از محورهای شهرکرد-فرخشهر، فرخشهر-شهرکرد و سورشجان-

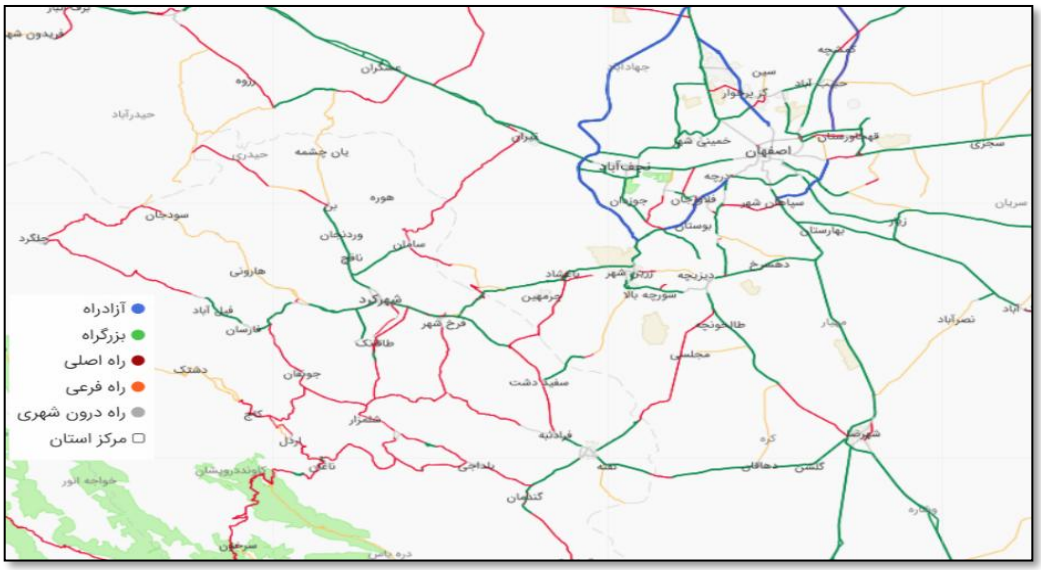
[1] Sum of Squares Between
[2] Sum of Squares Within

شهرکرد به‌طور متوسط بیشترین وسیلهٔ نقیله از نوع دستهٔ یک تردد داشته‌اند. بر اساس شکل (۵) (که متوسط تعداد وسیلهٔ نقلیه از نوع دستهٔ دو را که از این محورهای ۳۶ گانه عبور کرده‌اند را نشان می‌دهد)، به‌ترتیب محورهای گندمان-بروجن، شهرکرد-فرخ‌شهر و سفیدشت-بروجن به‌طور متوسط دارای بیشترین تعداد وسیلهٔ نقیله از نوع دستهٔ دو هستند. شکل (۶) متوسط تعداد وسیلهٔ نقلیه از نوع دستهٔ سه را در شهریورماه سال ۱۴۰۲ برای محورهایی که به‌طور متوسط دارای بیشترین وسیلهٔ نقلیهٔ دستهٔ سه در مقایسه با سایر محورها است، نشان می‌دهد. همان‌گونه که مشاهده می‌شود به‌ترتیب محورهای گندمان-بروجن، شهرکرد-فرخ‌شهر و فرخ‌شهر-شهرکرد به‌طور متوسط دارای بیشترین تعداد وسیلهٔ نقلیهٔ دستهٔ سه هستند. با توجه به شکل‌های (۷) و (۸) که به‌ترتیب متوسط تعداد وسیلهٔ نقلیهٔ گذری از نوع دستهٔ چهار و دستهٔ پنج را در شهریورماه سال ۱۴۰۲ نشان می‌دهند، محور گندمان-بروجن به‌طور متوسط دارای بیشترین تعداد وسیلهٔ نقلیهٔ دستهٔ چهار و دستهٔ پنج هست که از این محور در زمان مذکور عبور کرده‌اند.

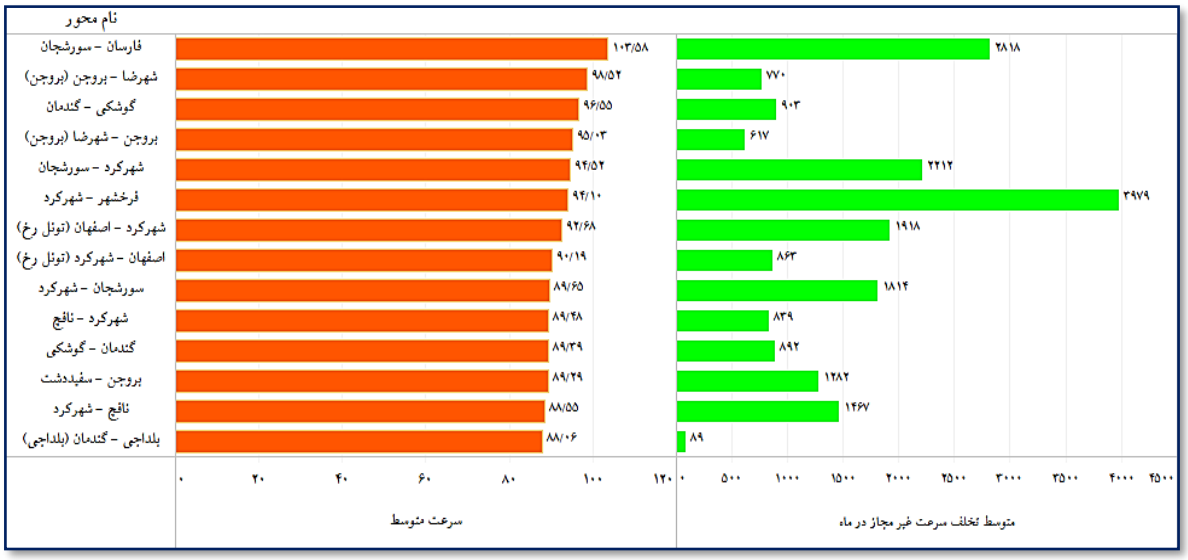
شکل (۹) متوسط تعداد تخلف سبقت و فاصلهٔ غیرمجاز را در شهریورماه سال ۱۴۰۲ برای محورهایی با بالاترین تخلف سبقت غیرمجاز نشان می‌دهد؛ همان‌گونه که مشاهده می‌شود، به‌طور متوسط محورهای لردگان-دهدز (تونل فتح)، فرح-پاتاوه و دهدز-لردگان بیشترین تخلف سبقت غیرمجاز را دارند که در بین این سه محور، محور دهدز-لردگان تخلف فاصلهٔ غیرمجاز بیشتری دارد.



(شکل-۱): نقشهٔ طرح کلی ترافیکی استان چهارمحال‌وبختیاری
(Figure-1): Schematic traffic map of Chaharmahal and Bakhtiari province

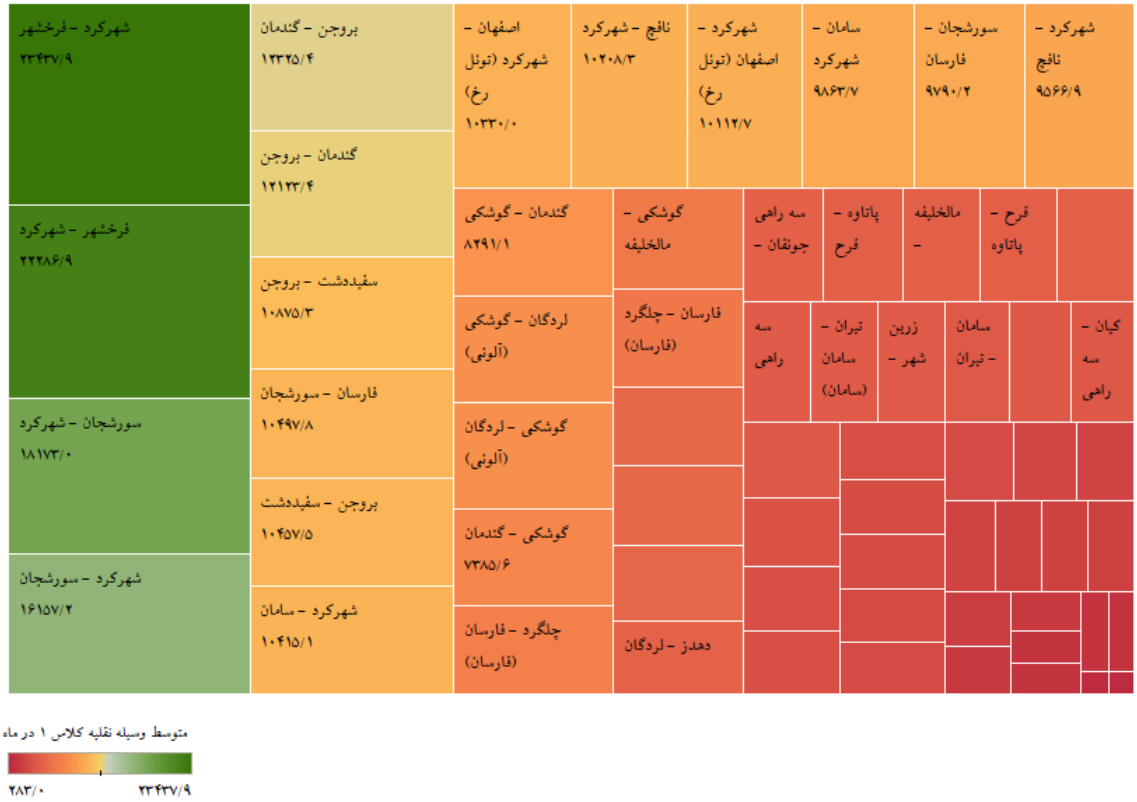


(شکل-۲): نقشه طبقه‌بندی راه‌های استان چهارمحال‌وبختیاری
(Figure-2): Classification map of the roads in Chaharmahal and Bakhtiari province



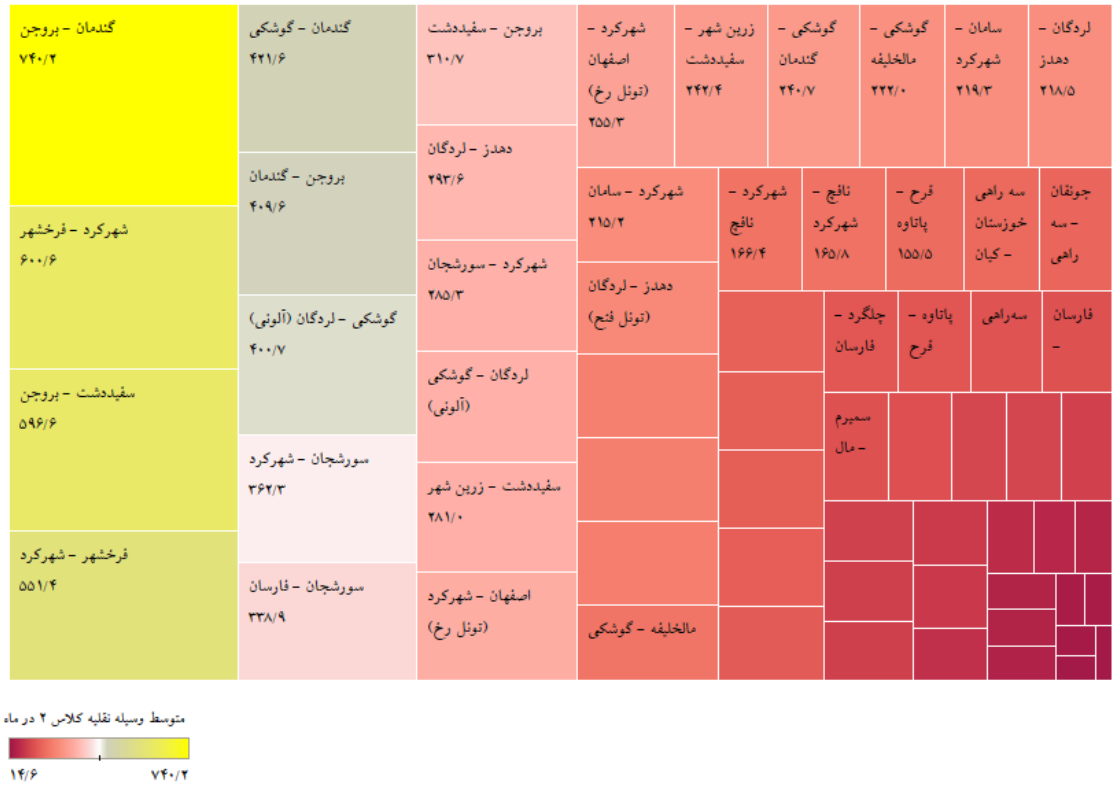
(شکل-۳): سرعت متوسط و متوسط تخلف سرعت غیرمجاز در شهریورماه ۱۴۰۲، برای محورهایی با بالاترین سرعت متوسط

(Figure -3): Average speed and average speed violations in September 2023 for routes with the highest average speed



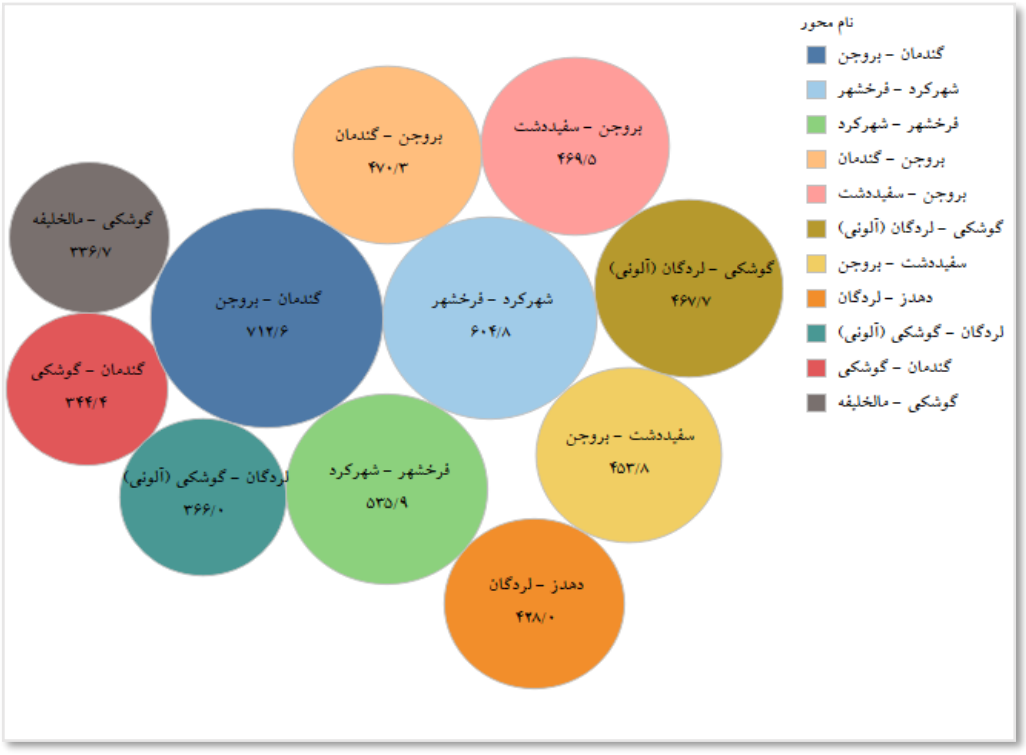
(شکل-۴): متوسط وسیلهٔ نقلیهٔ دستهٔ یک در شهریورماه سال ۱۴۰۲

(Figure-4): Average class 1 vehicle in September 2023



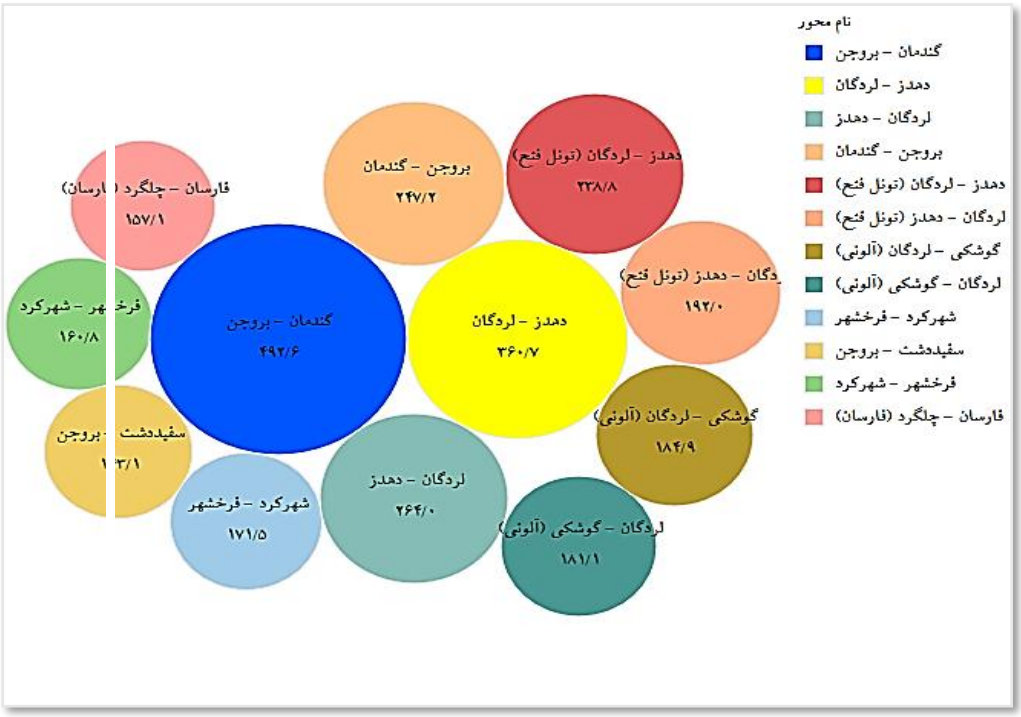
(شکل-۵): متوسط وسیلهٔ نقلیهٔ دستهٔ ۲ در شهریورماه سال ۱۴۰۲

(Figure-5): Average class 2 vehicle in September 2023



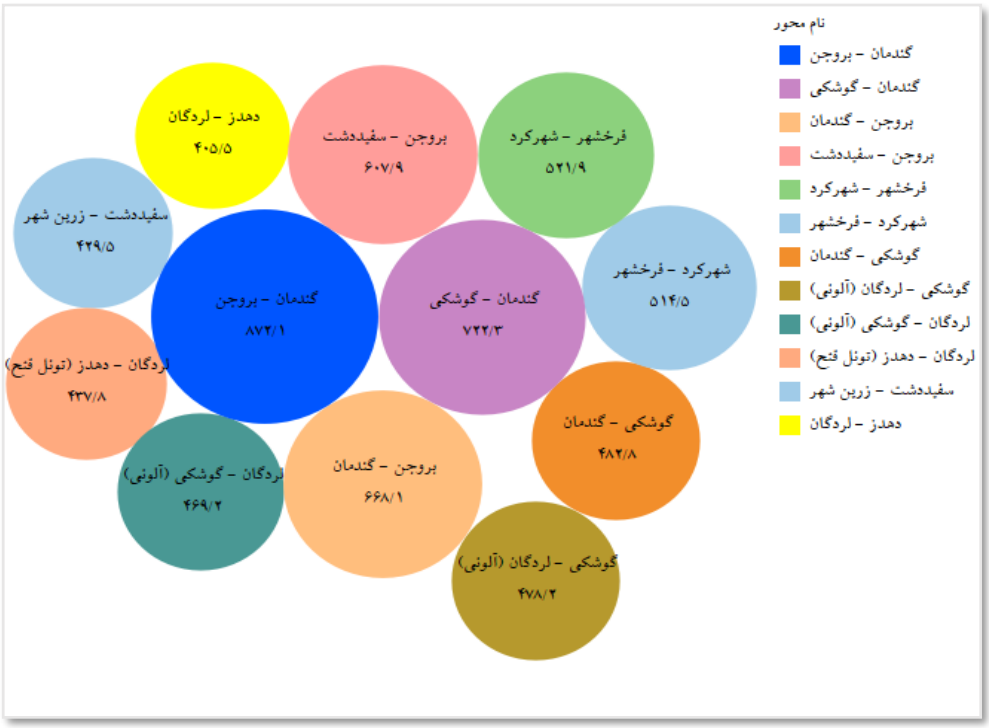
(شکل-۶): متوسط وسیلهٔ نقلیهٔ دستهٔ سه در شهریورماه سال ۱۴۰۲

(Figure-6): Average class 3 vehicle in September 2023



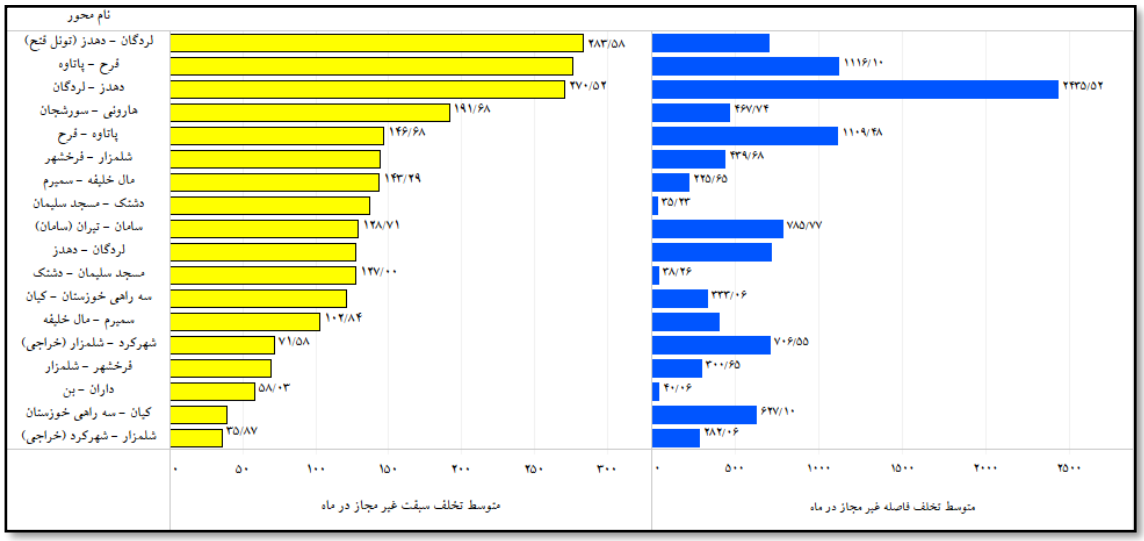
(شکل-۷): متوسط وسیلهٔ نقلیهٔ دستهٔ چهار در شهریورماه سال ۱۴۰۲

(Figure-7): Average class 4 vehicle in September 2023



(شکل-۸): متوسط وسیلهٔ نقلیهٔ دستهٔ پنج در شهریورماه سال ۱۴۰۲

(Figure-8): Average class 5 vehicle in September 2023



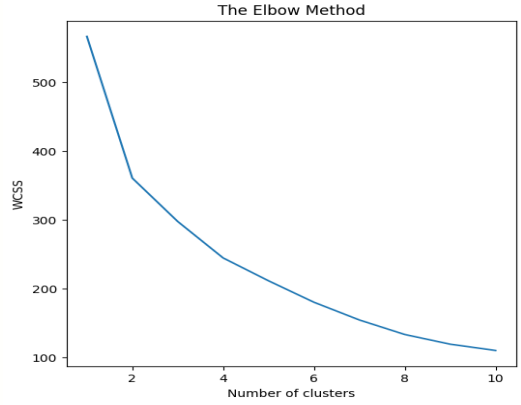
(شکل-۹): متوسط تخلف سبقت و فاصلهٔ غیرمجاز در شهریورماه ۱۴۰۲، برای محورهایی با بالاترین تخلف سبقت غیرمجاز

(Figure-9): Average overtaking violation and illegal distance in September 1402, for axes with the highest illegal overtaking violation

۶- تحلیل ساختار ترافیکی جاده‌ها بر اساس الگوریتم‌های خوشه‌بندی و مقایسهٔ عملکرد آن‌ها

از آنجا که در روش خوشه‌بندی مقدار فاصلهٔ بین مشاهدات (شباهت بین مشاهدات) در تشکیل و تعیین اعضای خوشه‌ها اهمیت دارد؛ بنابراین ضروری است داده‌های ورودی پیش از اجرای الگوریتم خوشه‌بندی هم‌مقیاس شوند؛ یکی از روش‌های موجود برای هم‌مقیاس‌سازی داده‌ها روش استانداردکردن داده‌ها است که با اجرای آن می‌توان اطمینان حاصل کرد که تأثیر مقیاس‌های مختلف درنتیجهٔ خوشه‌بندی کاهش یافته و عملکرد خوشه‌بندی بهبود می‌یابد. در مطالعهٔ حاضر پیش از اجرای الگوریتم‌های خوشه‌بندی، متغیرها به‌گونه‌ای استاندارد شدند که میانگین هر متغیر صفر و انحراف معیار آن یک باشد.

پس از استانداردکردن داده‌ها، گام مهم بعدی در اجرای الگوریتم k - میانگین تعیین تعداد بهینهٔ خوشه‌ها یعنی مقدار k است. برای این منظور روش‌های مختلفی توسط پژوهش‌گران ارائه شده‌است که در این پژوهش از روش آرنج[1] که مبتنی بر مینیم‌سازی (به کمترین مقدار ممکن رساندن) معیار مجموع مربعات فواصل درون‌خوشه‌ای[2] (WCSS) است، استفاده می‌شود. معیار WCSS میزان فشردگی داده‌ها در درون خوشه‌های تشکیل‌شده را نشان می‌دهد؛ به‌طوری که هرچه داده‌های واقع در خوشهٔ یکسان و فشرده‌تر، نتیجهٔ خوشه‌بندی بهتر است؛ ازاین‌رو هدف در روش آرنج، به کمترین حد رساندن مقدار این معیار برای یافتن تعداد خوشهٔ بهینه است؛ به‌طوری که اگر یک مقدار به تعداد خوشهٔ بهینه اضافه شود، نتیجه‌ای جز پیچیدگی مدل حاصل نشود. برای تشخیص مقدار k بهینه، به‌طور معمول نمودار WCSS به‌عنوان تابعی از k ترسیم شده و نقطه‌ای را که در آن شیب نمودار تند نیست تعیین می‌کنند (نقطه‌ای که در آن یک زاویه یا آرنج در نمودار ایجاد شده‌است). طول متناظر با این نقطه، معادل مقدار بهینه برای k است[۲۶].



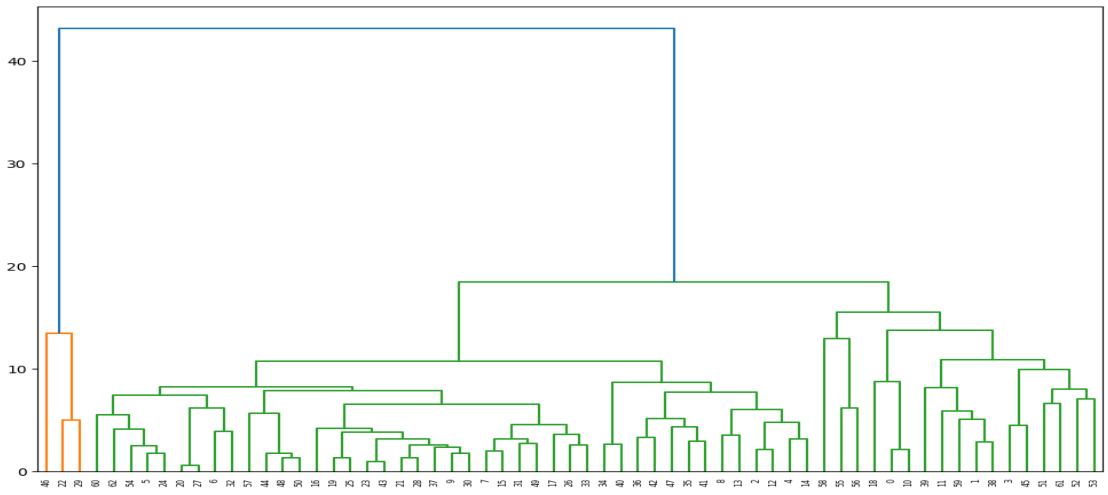
(شکل-۱۰): نمودار آرنج برای تعیین تعداد خوشه‌های بهینه

(Figure-10): Elbow diagram to determine the number of optimal clusters

شکل (۱۰) نمودار WCSS را در مقابل مقادیر مختلف k نشان می‌دهد. مطابق شکل (۱۰) و با توجه به آنچه که گفته شد مقدار بهینه برای تعداد خوشه‌ها برابر $k = 2$ به‌دست می‌آید. در الگوریتم خوشه‌بندی سلسله‌مراتبی برای تعیین خوشه‌ها از روش پیوند میانگین استفاده شده‌است. شکل (۱۱) نمودار درختی که دنبالهٔ ادغام‌ها و فواصل مربوط به این خوشه‌بندی است را نشان می‌دهد که در آن محور طول‌ها بیان‌کنندهٔ برچسب محورهای تردد و محور عرض‌ها فواصل

[1]. Elbow- Method
[2]. Within-Cluster Sum of Square

مربوط به خوشه‌های تشکیل‌شده در هر مرحله را نشان می‌دهد. با توجه به بیشینهٔ مقدار برای ضریب سیلوئت، تعداد خوشه‌ها از این طریق نیز برابر دو به‌دست می‌آید. با توجه به شکل (۱۱)، محورهای با برچسب ۴۶، ۲۲ و ۲۹ (که شامل محورهای شهرکرد-فرخشهر، فرخشهر-شهرکرد و گندمان-بروجن است) با هم‌دیگر تشکیل یک خوشه را داده و دیگر محورها با هم خوشهٔ دوم را تشکیل می‌دهند.



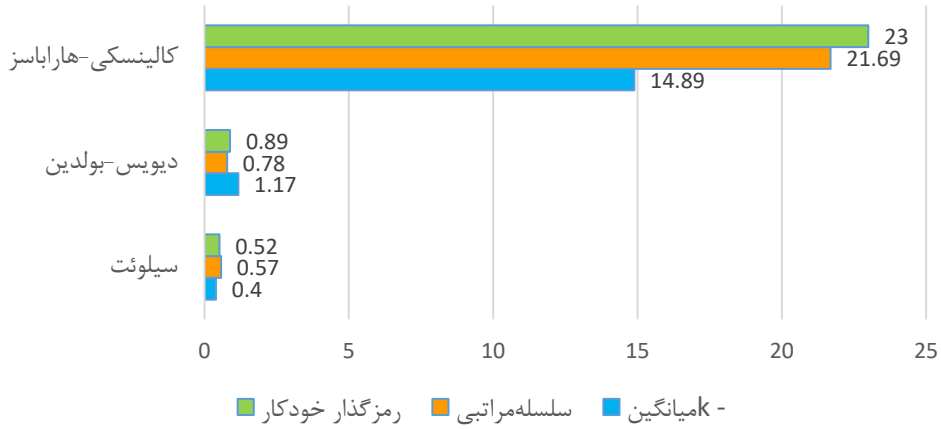
(شکل-۱۱): نمودار درختی در خوشه‌بندی سلسله‌مراتبی

(Figure-11): Dendrogram diagram in hierarchical clustering

(جدول- ۲): تعداد و اندازهٔ خوشه‌های حاصل از اجرای الگوریتم‌های خوشه‌بندی

(Table-2): Number and size of clusters resulting from clustering algorithms

روش خوشه‌بندی	تعداد خوشه‌ها	تعداد مشاهدات خوشهٔ دو	تعداد مشاهدات خوشهٔ یک
میانگین k-	۲	۱۶	۴۷
سلسله‌مراتبی	۲	۶۰	۳
رمزگذار خودکار	۲	۵۹	۴



(شکل -۱۲): نمودار مقایسهٔ مقادیر شاخص‌های ارزیابی ضریب سیلوئت، دیویس-بولدین و کالینسکی-هاراباسز

(Figure -12): Comparison chart of silhouette score, Davies-Bouldin index, and Calinski-Harabasz index

(جدول-۳): مقادیر شاخص‌های ارزیابی ضریب سیلوئت، دیویس-بولدین و کالینسکی-هاراباسز

(Table-3): Values of Silhouette, Davis-Bouldin and Calinski-Harabasz evaluation indices

روش خوشه‌بندی	ضریب سیلوئت	دیویس-بولدین	کالینسکی-هاراباسز
میانگین k -	۰/۴۰	۱/۱۷	۱۴/۸۹
سلسله‌مراتبی	۰/۵۷	۰/۷۸	۲۱/۶۹
رمزگذار خودکار	۰/۵۲	۰/۸۹	۲۳/۰۰

(جدول- ۴): ویژگی خوشه‌های به‌دست‌آمده از الگوریتم‌های خوشه‌بندی

(Table-4): Characteristic of the clusters obtained from the clustering algorithms

	متغیر	خوشهٔ یک			خوشهٔ دو		
		کمترین مقدار	بیشترین مقدار	میانگین	کمترین مقدار	بیشترین مقدار	میانگین
روش خوشه‌بندی سلسله‌مراتبی	متوسط وسیلهٔ نقلیه دستهٔ یک	۱۲۱۲۳.۳۵	۲۳۴۳۷.۸۷	۱۹۲۸۲.۷۲	۲۸۳.۰۰	۱۸۱۷۳.۰۰	۵۲۱۹.۰۷
	متوسط وسیلهٔ نقلیهٔ دستهٔ دو	۵۵۱.۳۵	۷۴۰.۲۳	۶۳۰.۷۳	۱۴.۶۱	۵۹۶.۵۸	۱۶۷.۸۳
	متوسط وسیلهٔ نقلیهٔ دستهٔ سه	۵۳۵.۸۷	۷۱۲.۶۵	۶۱۷.۷۸	۱۱.۱۶	۴۷۰.۲۶	۱۸۱.۸۱
	متوسط وسیلهٔ نقلیهٔ دستهٔ چهار	۱۶۰.۸۴	۴۹۲.۵۸	۲۷۴.۹۸	۰	۳۶۰.۶۸	۷۰.۰۳
	متوسط وسیلهٔ نقلیهٔ دستهٔ پنج	۵۱۴.۴۵	۸۷۲.۱۳	۶۳۶.۱۷	۰	۷۲۲.۳۲	۱۷۰.۱۷
	متوسط تخلف سرعت غیرمجاز	۱۹۷.۷۴	۳۹۷۸.۹۷	۲۱۲۱.۲۳	۳.۹۰	۲۸۱۸.۰۰	۵۰۶.۸۷
	متوسط تخلف فاصلهٔ غیرمجاز	۳۰۱۹.۳۵	۶۷۰۹.۲۶	۵۰۴۸.۷۴	۳۵.۲۳	۳۴۳۵.۰۳	۸۰۷.۷۸
	متوسط تخلف سبقت غیرمجاز	۰	۱.۸۷	۰/۶۲	۰	۲۸۳.۵۸	۴۲/۶۴
	سرعت متوسط	۶۵/۹۰	۹۴/۰۹	۸۲/۲۹	۵۳/۶۵	۱۰۳/۵۸	۷۷/۲۵
روش خوشه‌بندی رمزگذار خودکار	متوسط وسیلهٔ نقلیه دستهٔ یک	۱۲۱۲۳.۳۵	۲۳۴۳۷.۸۷	۱۷۵۴۳.۳۹	۲۸۳.۰۰	۱۸۱۷۳.۰۰	۵۰۹۸.۶۲
	متوسط وسیلهٔ نقلیهٔ دستهٔ دو	۴۰۹.۶۱	۷۴۰.۲۳	۵۷۵.۴۵	۱۴.۶۱	۵۹۶.۵۸	۱۶۳.۷۳
	متوسط وسیلهٔ نقلیهٔ دستهٔ سه	۴۷۰.۲۶	۷۱۲.۶۵	۵۸۰.۹۰	۱۱.۱۶	۴۶۹.۴۸	۱۷۶.۹۲
	متوسط وسیلهٔ نقلیهٔ دستهٔ چهار	۱۶۰.۸۴	۴۹۲.۵۸	۲۶۸.۰۴	۰	۳۶۰.۶۸	۶۷.۰۳
	متوسط وسیلهٔ نقلیهٔ دستهٔ پنج	۵۱۴.۴۵	۸۷۲.۱۳	۶۴۴.۱۵	۰	۷۲۲.۳۲	۱۶۱.۷۳
	متوسط تخلف سرعت غیرمجاز	۱۸۲.۲۹	۳۹۷۸.۹۷	۱۶۳۶.۴۹	۳.۹۰	۲۸۱۸.۰۰	۵۱۲.۳۷
	متوسط تخلف فاصلهٔ غیرمجاز	۲۲۲۷.۵۲	۶۷۰۹.۲۶	۴۳۴۳.۴۴	۳۵.۲۳	۳۴۳۵.۰۳	۷۸۳.۷۱
	متوسط تخلف سبقت غیرمجاز	۰	۲.۲۹	۱.۰۴	۰	۲۸۳.۵۸	۴۳.۳۲
	سرعت متوسط	۶۵.۹۰	۹۴.۱۰	۷۹.۷۴	۵۳.۶۵	۱۰۳.۵۸	۷۷.۳۴

رمزنگارهای خودکار به‌طور معمول شبکه‌های کم‌عمق هستند که رایج‌ترین آن‌ها دارای یک لایهٔ ورودی، یک لایهٔ میانی (مخفی) و یک لایهٔ خروجی است. آموزش در این الگوریتم بر مبنای تابع ضرر انجام می‌گیرد. در این الگوریتم، تابع ضرر میزان اطلاعات مربوط به ورودی را که طی فرایند رمزنگاری و رمزگشایی از دست رفته است اندازه‌گیری می‌کند؛ به‌طوری که هرچه مقدار ضرر کمتر باشد شبکه از قدرت بالایی برخوردار است. در مطالعهٔ حاضر پس از پیش‌پردازش و استانداردکردن داده‌ها از تابع ضرر بهینه‌ساز آدام[1] و آنتروپی متقاطع باینری[2] استفاده شد که کمترین مقدار برای آن به ازای تعداد لایهٔ میانی پنجاه‌تایی حاصل شد. در ادامه از خروجی این رمزگذار، به‌عنوان داده‌های ورودی خوشه‌بندی k- میانگین با $k = 2$ برای خوشه‌بندی استفاده شد. جدول (۲)، به‌طور خلاصه اندازه و تعداد خوشه‌های حاصل از اجرای سه الگویتم خوشه‌بندی سلسله‌مراتبی، k- میانگین و رمزگذاری خودکار را نشان می‌دهد؛ چنانچه در بخش‌های پیشین اشاره شد، در این مطالعه برای مقایسهٔ عملکرد و دقت الگوریتم‌های به‌کار گرفته‌شده در خوشه‌بندی داده‌ها، از سه شاخص ضریب نیم‌رخ، شاخص دیویس-بولدین و شاخص کالینسکی-هاراباسز استفاده می‌شود. در جدول (۳)، مقادیر به‌دست‌آمده برای این شاخص‌ها گزارش شده‌است. برای شهود و مقایسهٔ بهتر، شکل (۱۲) نمودار مقادیر گزارش‌شده در جدول (۳) را برای هر سه روش خوشه‌بندی به‌کار گرفته شده نمایش می‌دهد.

مطابق شکل (۱۲) و مقادیر گزارش‌شده در جدول (۳) برای معیارهای ارزیابی مختلف، دو روش خوشه‌بندی سلسله‌مراتبی و رمزگذار خودکار از عملکرد بهتری نسبت به روش k- میانگین در خوشه‌بندی داده‌ها برخوردارند. در روش خوشه‌بندی سلسله‌مراتبی، محورهای شهرکرد-فرخ‌شهر، فرخ‌شهر-شهرکرد و گندمان-بروجن رفتار مشابهی بر اساس متغیرهای در نظر گرفته شده، جدول (۱)، در مدل را داشته و اعضای خوشهٔ یک را تشکیل می‌دهند و دیگر محورها همگی در خوشهٔ دو قرار گرفته‌اند. در روش خوشه‌بندی رمزگذار خودکار، علاوه‌بر محورهای شهرکرد-فرخ‌شهر، فرخ‌شهر-شهرکرد، گندمان-بروجن، محور بروجن-گندمان نیز در شاخهٔ یک و سایر محورها در خوشهٔ دو قرار گرفته‌اند.

جدول (۴) ویژگی‌های مربوط به اعضای خوشه‌های یک و دو را به‌تفکیک برای دو مدل بهینهٔ سلسله‌مراتبی و رمزگذار خودکار، نشان می‌دهد. طبق جدول (۴)، به‌طور متوسط تعداد

[1]. Adam
[2]. Binary cross-entropy

تردد وسایل نقلیه از نوع دسته‌های یک تا پنج در خوشهٔ یک در مقایسه با خوشهٔ دو بیشتر بوده‌است؛ همچنین خوشهٔ یک، به‌طور متوسط تعداد تخلف سرعت و فاصلهٔ غیرمجاز بیشتری را در مقایسه با خوشهٔ دو به‌خود اختصاص داده‌است.

۷- نتیجه‌گیری

با توجه به رشد سامانه‌های هوشمند در شبکهٔ حمل‌ونقل جاده‌ای کشور و به‌منظور مدیریت بهینهٔ راه‌های ارتباطی، لازم است تحلیل مناسبی از وضعیت تردد فعلی جاده‌ها ارائه و مسیرهای با الگوهای حمل‌ونقل متفاوت شناسایی شود. استان چهارمحال‌وبختیاری به‌دلیل موقعیت جغرافیایی خاص خود، از اهمیت ویژه‌ای در شبکهٔ حمل‌ونقل جاده‌ای کشور برخوردار است. این استان در جنوب غربی ایران قرار گرفته است و پل ارتباطی بین شمال و جنوب، شرق و غرب کشور محسوب می‌شود. جاده‌های این استان نقش مهمی در اتصال استان‌های هم‌جوار و تردد بار و مسافر ایفا می‌کنند. بهبود وضعیت ایمنی و کارایی این جاده‌ها می‌تواند منجر به تسهیل حمل‌ونقل، افزایش ایمنی تردد، توسعهٔ اقتصادی و گسترش گردشگری در این استان شود؛ بنابراین مطالعه و طبقه‌بندی جاده‌های چهارمحال‌وبختیاری بر اساس ویژگی‌های ترافیکی، اهمیت راهبردی در مدیریت و برنامه‌ریزی حوزهٔ حمل‌ونقل این منطقه دارد. در این مطالعه با استفاده از سه الگوریتم خوشه‌بندی شامل k- میانگین، سلسله‌مراتبی و رمزگذار خودکار به طبقه‌بندی جاده‌های استان چهارمحال‌وبختیاری پرداخته شد.

خوشه‌بندی جاده‌ها بر اساس ویژگی‌هایی مانند متوسط تعداد وسیلهٔ نقلیهٔ دستهٔ یک تا پنج، متوسط تعداد تخلف سرعت غیرمجاز، متوسط تعداد تخلف فاصلهٔ غیرمجاز، متوسط تعداد تخلف سبقت غیرمجاز و سرعت متوسط انجام شدند. بر اساس شاخص‌های ارزیابی سیلوئت، دیویس-بولدین و کالینسکی-هاراباسز، دو روش خوشه‌بندی سلسله‌مراتبی و رمزگذار خودکار نسبت به روش خوشه‌بندی k- میانگین از دقت و عملکرد بهتری در طبقه‌بندی محورهای تردد برخوردارند. در طبقه‌بندی جاده‌ها، محورهای مواصلاتی به دو خوشهٔ مجزا تقسیم‌بندی شدند؛ نخستین خوشه شامل مسیرهایی با بیشترین تردد انواع وسایل نقلیهٔ دستهٔ یک تا پنج، بیشترین متوسط تعداد تخلف سرعت و فاصلهٔ غیرمجاز بوده که نشان‌دهندهٔ رفتار خطرناک در رانندگی و خطر تصادف بالقوه بالاتر است. یافته‌های این تحقیق می‌تواند در ارتقا و رشد آگاهی مسئولین امر در جهت تصمیم‌گیری هدفمند و بهینه برای افزایش کارایی و امنیت در جاده‌های استان چهار محال و بختیاری، مؤثر واقع شود. این روش می‌تواند در بررسی ساختار جاده‌های مواصلاتی دیگر استان‌ها به‌کار گرفته شده و چهارچوبی جامع برای درک و مدیریت شبکه‌های جاده‌ای فراهم کند. پژوهش حاضر از چند جهت واجد نوآوری و ارزش افزودهٔ علمی است؛ نخست آنکه برخلاف مطالعات پیشین که بیشتر به تحلیل ترافیک در سطح کلان یا با روش‌های کلاسیک آماری پرداخته‌اند، این پژوهش با بهره‌گیری از الگوریتم‌های یادگیری بدون نظارت و روش‌های نوین خوشه‌بندی، ازجمله رمزنگار خودکار و خوشه‌بندی سلسله‌مراتبی، به تحلیل و گروه‌بندی جاده‌های استان چهارمحال‌وبختیاری پرداخته است. این امر نه‌تنها موجب ارتقای دقت تحلیل در طبقه‌بندی مسیرها شده، بلکه امکان شناسایی الگوهای پنهان در رفتارهای ترافیکی را نیز فراهم آورده‌است؛ دوم آنکه استفادهٔ هم‌زمان از چندین شاخص ارزیابی معتبر ازجمله سیلوئت، دیویس-بولدین و کالینسکی-هاراباسز برای سنجش کیفیت خوشه‌بندی، موجب افزایش اعتبار و استحکام نتایج شده‌است. مورد سوم اینکه تمرکز پژوهش بر ترکیبی از متغیرهایی نظیر سرعت غیرمجاز، فاصلهٔ غیرمجاز و نوع وسیلهٔ نقلیه، رویکردی جامع و واقع‌بینانه نسبت به تحلیل ریسک‌پذیری جاده‌ها ارائه کرده‌است که می‌تواند در سیاست‌گذاری‌های مبتنی بر داده، به‌ویژه در حوزهٔ ایمنی جاده‌ای و توسعهٔ زیرساخت‌های حمل‌ونقل، مورد استفاده قرار گیرد. درنهایت، این مطالعه با ارائهٔ یک چهارچوب تحلیلی داده‌محور برای طبقه‌بندی جاده‌ها بر اساس الگوهای ترافیکی، گامی نوین در جهت بهره‌برداری هوشمند از داده‌های ترددشمار جاده‌ای در سطح استانی برداشته و می‌تواند الگوی مناسبی برای سایر مناطق کشور نیز باشد. در ادامه می‌توان موارد زیر را به‌عنوان مطالعات آتی در نظر گرفت:

۱. توسعهٔ مدل‌های ترکیبی خوشه‌بندی و یادگیری عمیق برای بهبود دقت در شناسایی الگوهای ترافیکی.
۲. تحلیل تأثیر سیاست‌های مدیریت ترافیک بر کاهش تخلفات مانند ارزیابی تأثیر نصب دوربین‌های کنترل سرعت، افزایش جریمه‌ها و آموزش رانندگان برای کاهش تخلفات در جاده‌های پرخطر.
۳. گسترش تحلیل به دوره‌های زمانی متفاوت و فصلی.
۴. مطالعهٔ تطبیقی بین استان‌های مختلف ایران از قبیل مقایسهٔ الگوهای ترافیکی و تخلفات در استان‌های کوهستانی (مانند چهارمحال‌وبختیاری) با استان‌های پرتردد (مانند تهران و اصفهان).
۵. بررسی تأثیر عوامل محیطی و اجتماعی بر الگوهای ترافیکی از قبیل ترکیب داده‌های ترافیکی با متغیرهایی مانند وضعیت آب‌وهوا، مناسبت‌های خاص یا داده‌های اجتماعی-اقتصادی مناطق که می‌تواند بینش عمیق‌تری نسبت به علل رفتارهای پرخطر رانندگان فراهم سازد.

۸-مراجع 8-References

[1] World Health Organization. (2018). Road traffic injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

[2] C. Musingura, G. Lee, Y. Ahn, and K. Kim, "Mitigating Road Traffic Crashes in Urban Environments: A Case Study and Literature Review-based Approach," Authorea Preprints, 2023.

[۳] کریمی، سیده فاطمه، خدابخش، مریم، «پیش‌بینی عملکرد نتایج پرس‌وجو با کمک روش‌های بدون نظارت»، *پردازش علائم و داده‌ها*، (۱) ۲۲، صفحات ۱۲-۳، ۱۴۰۴.

[3] S. Karimi, M. Khodabakhsh, "Unsupervised Methods for Predicting Query Performance",

Journal of Signal and Data Processing, 22 (1), pp. 3-12, 2025.

[۴] دانشپور نگین، برزگری علی، «روشی جدید در تشخیص تکراری رکوردها با استفاده از خوشه‌بندی سلسله‌مراتبی»، پردازش علائم و داده‌ها، (۴) ۱۸، صفحات ۳-۲۲، ۱۴۰۰.

[4] N. Daneshpour, A. Barzegari, "A New Method for Duplicate Detection Using Hierarchical Clustering of Records", *Journal of Signal and Data Processing*; 18 (4) , pp. 3-22, 2022.

[5] R. Rossi, M. Gastaldi, and G. Gecchele, "Comparison of clustering methods for road group identification in FHWA traffic monitoring approach: Effects on AADT estimates," *Journal of Transportation Engineering*, vol. 140, no. 7, pp. 04014025, Jul. 2014.

[6] J. Wang and F. Biljecki, "Unsupervised machine learning in urban studies: A systematic review of applications," *Cities*, vol. 129, p. 103925, 2022.

[7] C. Bratsas et al., "A comparison of machine learning methods for the prediction of traffic speed in urban places," *Sustainability*, vol. 12, no. 1, p. 142, 2019.

[8] X. Cui et al., "Extracting main center pattern from road networks using density-based clustering with fuzzy neighborhood," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 5, p. 238, 2019.

[9] A. Aggarwal, A. Purwar, and S. Gulati, "An efficient technique to control road traffic using fuzzy neural network system," in *Proc. 3rd Int. Conf. Reliability, Infocom Technol. Optim.*, Oct. 2014, pp. 1-6.

[10] Y. Liu and H. Wu, "Prediction of road traffic congestion based on random forest," in *2017 10th Int. Symp. Comput. Intell. Des.* (ISCID), Vol. 2, Dec. 2017, pp. 361-364.

[11] M. Akhtar and S. Moridpour, "A review of traffic congestion prediction using artificial intelligence," *J. Adv. Transp.*, vol. 2021, no. 1, p. 8878011, 2021.

[12] S. Yang et al., "Analysis of traffic state variation patterns for urban road network based on spectral clustering," *Adv. Mech. Eng.*, vol. 9, no. 9, p. 1687814017723790, 2017.

[13] Y. Haung, H. Zhen, and J.J. Yang, "Cluster-guided denoising graph auto-encoder for enhanced traffic data imputation and fault detection," *Expert Systems with Applications*, vol 261, p. 125531, Feb. 2025.

[14] D. Shin et al., "Clustering and investigation of human driving bahavior using autoencoder and risk assessment," IEEE Access, 2025 Jan 15.

[15] F. Nielsen and F. Nielsen, "Hierarchical clustering," in *Introduction to HPC with MPI for Data Science*, pp. 195-211, 2016.

[16] Z. Zhang et al., "Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R," *Annals of Translational Medicine*, vol. 5, no. 4, 2017.

[17] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, L. M. Le Cam and J. Neyman, Eds. California: University of California Press, pp. 281-297, 1967.

[18] Q. Wang, C. Wang, Z. Feng, and J. F. Ye, "Review of K-means clustering algorithm," *Electronic Design Engineering*, vol. 20, no. 7, pp. 21-24, 2012.

[19] D. Steinley and M. J. Brusco, "Initializing k-means batch clustering: A critical evaluation of several techniques," *Journal of Classification*, vol. 24, pp. 99-121, 2007.

[20] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," in *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, pp. 353-374, 2023.

[21] F. Ros and R. Riad, "Deep clustering techniques based on autoencoders," in *Feature and Dimensionality Reduction for Clustering with Deep Learning*, Cham: Springer Nature Switzerland, pp. 203-220, 2023.

[22] K. Berahmand et al., "Autoencoders and their applications in machine learning: a survey," *Artificial Intelligence Review*, vol. 57, no. 2, p. 28, 2024.

[23] I. Yatskiv and L. Gusarova, "The methods of cluster analysis results validation," *in Proc. Int. Conf. RelStat*, vol. 4, pp. 75-80, 2005.

[24] Y. A. Wijaya et al., "Davies bouldin index algorithm for optimizing clustering case studies mapping school facilities," *TEM J,* vol. 10, no. 3, pp. 1099-1103, 2021.

[25] X. Wang and Y. Xu, "An improved index for clustering validation based on silhouette index and calinski-harabasz index," in *IOP Conf. Ser., Mater. Sci. Eng*., vol. 569, Aug. 2019, Art. no. 052024.

[26] D. J. Ketchen Jr. and C. L. Shook, "The application of cluster analysis in Strategic Management Research: An analysis and critique," *Strategic Management Journal*, vol. 17, no. 6, pp. 441–458, 1996.



وحیده احراری استادیار دانشکدهٔ علوم ریاضی دانشگاه شهرکرد است. وی دانش‌آموختهٔ مقطع دکتری آمار از دانشگاه فردوسی مشهد است. زمینه‌های پژوهشی مورد علاقهٔ ایشان شامل علم داده، استنباط آماری، توزیع‌های طول عمر، اندازه‌های اطلاع و آزمون‌های نیکویی برازش است.

نشانی رایانامهٔ ایشان عبارت‌است از:

vahideh.ahrari@sku.ac.ir



رباب افشاری استادیار دانشکدهٔ علوم دانشگاه زنجان است. وی دانش‌آموختهٔ مقطع دکتری آمار از دانشگاه فردوسی مشهد است. زمینه‌های پژوهشی مورد علاقهٔ ایشان شامل کنترل کیفیت آماری، داده‌کاوی، آمار فازی، آمار صنعتی، آمار چندمتغیره و قابلیت اعتماد است.

نشانی رایانامهٔ ایشان عبارت‌است از:

afshari@znu.ac.ir