

استفاده از الگوریتم‌های داده‌کاوی برای تحلیل سودمندی

دوربین‌های راهنمایی و رانندگی

بی‌بی هانیه دقیق نژاد، دانشجوی کارشناسی ارشد، مرکز آموزش‌های الکترونیکی، دانشگاه صنعتی شاهرود، شاهرود، ایران

هدی مشایخی (مسئول مکاتبات)، استادیار، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شاهرود، شاهرود، ایران

زهرا نوراله، دانش‌آموخته کارشناسی ارشد، دانشکده مهندسی صنایع، دانشگاه تربیت مدرس، تهران، ایران

E-mail: hmayekhi@shahroodut.ac.ir

پذیرش: ۱۳۹۶/۰۷/۲۲

دریافت: ۱۳۹۵/۰۵/۱۷

چکیده

مجمع عمومی سازمان ملل متحد، دهه ۲۰۱۱ تا ۲۰۲۰ میلادی را به عنوان دهه کاهش تلفات جاده‌ای معرفی کرد. این سازمان در طی قطعنامه‌های مختلفی از تمام کشورها خواسته است نسبت به کاهش تصادفات جاده‌ای توجه خاصی نمایند. عامل سرعت نقش کلیدی در کاهش ریسک تصادفات جاده‌ای دارد. یکی از عوامل بازدارندگی سرعت غیرمجاز خودروها، دوربین‌های ثبت تخلف پلیس راهور ناجا است. در سال‌های اخیر استفاده از این دوربین‌ها به منظور کاهش تخلفات و تصادفات جاده‌های برون‌شهری و بزرگراه‌ها گسترش زیادی داشته است. هدف از این پژوهش تحلیل سودمندی دوربین‌های ثبت تخلف راهنمایی و رانندگی بزرگراه تهران کرج است. این امر با بهره‌گیری از تکنیک‌های داده‌کاوی از قبیل مدلسازی با سری‌زمانی و رگرسیون صورت می‌پذیرد. در ابتدا با استفاده از تحلیل توصیفی، سیر تخلفات بزرگراه تهران کرج را به همراه حجم تردد و تعداد دوربین‌های فعال در آن مورد بررسی قرار می‌دهیم. در مرحله تحلیل استنباطی، با مدلسازی سری‌زمانی و مدل غیرخطی توانی رگرسیون، تخلفات بزرگراه تهران کرج را مورد تجزیه و تحلیل قرار می‌دهیم. در بررسی نتایج ارزیابی، مدل‌های برازش شده با سری‌زمانی از خطای پیش‌بینی کمتری نسبت به مدل‌های رگرسیونی برخوردار بودند. در پیش‌بینی صورت گرفته برای برخی ماه‌های سال ۹۳، میزان تخلفات بزرگراه تهران کرج به میزان ۳۲٪ نسبت به تخلفات ماه‌های نظیر خود در سال ۹۲ کاهش داشته است، این امر به همراه تحلیل‌های دیگر نشان‌دهنده سودمندی دوربین‌ها در بازه مورد بررسی است. از نتایج تحلیل و کاوش داده‌های دوربین‌های ثبت تخلف می‌توان برای بهبود مدیریت و برنامه‌ریزی شهری با هدف کاهش تخلفات و تصادفات استفاده نمود.

واژه‌های کلیدی: داده‌کاوی، سری‌زمانی، رگرسیون، دوربین ثبت تخلف

۱. مقدمه

نصب سامانه ترددشمار از نیمه دوم سال ۹۱ تاکنون و همچنین نصب دوربین‌های ثبت تخلف از سال ۸۸ و افزایش ۳۰ درصدی آنها در سالهای بعد، برآن شدیم تا سودمندی دوربین‌های ثبت تخلف سرعت غیرمجاز نصب شده در این بزرگراه را بررسی نماییم. بدین منظور در این مقاله از داده‌های سامانه‌های ترددشمار برای بررسی حجم وسایل نقلیه عبوری و دوربین‌های ثبت تخلف برای میزان تخلفات سرعت غیرمجاز وسایل نقلیه در این بزرگراه، استفاده گردیده است.

در ابتدا با استفاده از تحلیل توصیفی، سیر تخلفات بزرگراه تهران- کرج را همراه با تعداد دوربین‌های فعال در آن مورد بررسی قرار می‌دهیم. در مرحله تحلیل استنباطی، با مدلسازی سری‌زمانی و مدل غیرخطی توانی رگرسیون، تخلفات بزرگراه تهران کرج را مورد تجزیه و تحلیل قرار می‌دهیم. تشریح و مدلسازی تخلفات با سری‌زمانی و رگرسیون، تحلیل عوامل موثر بر آن مانند (حجم تردها و میانگین تعداد دوربین‌ها در ماه) و بررسی تأثیرگذاری عوامل بر یکدیگر از نوآوری‌های این پژوهش است.

در ادامه این مقاله، در بخش ۲ به بررسی پیشینه تحقیق پرداخته و در بخش ۳ مواد و روش‌های پژوهش را شرح می‌دهیم. در بخش ۴ به بررسی و تحلیل یافته‌ها می‌پردازیم و نتیجه‌گیری را در بخش ۵ ارائه می‌کنیم.

۲. پیشینه تحقیق

پژوهش‌ها و مطالعاتی تاکنون در مورد ارزیابی عملکرد دوربین‌های ثبت تخلف و نقش بازدارندگی آنها در کاهش تخلفات در ایران و کشورهای دیگر صورت گرفته است. در این پژوهش‌ها با استفاده از سری‌زمانی یا رگرسیون به مدلسازی داده‌های پیشین و پیش‌بینی میزان تخلفات یا تصادفات در ماه‌های آینده پرداختند. اما در این مقاله علاوه بر تشریح و مدلسازی متغیر تخلفات و عوامل موثر بر آن مانند (حجم تردها و میانگین تعداد دوربین‌ها در ماه) با سری‌زمانی، به بررسی تأثیرگذاری آنها بر یکدیگر نیز پرداخته شده است. با توجه به این که سری‌زمانی به صورت خطی مدلسازی می‌گردد، در این پژوهش متغیرهای مرتبط به صورت رگرسیون غیرخطی هم مدلسازی می‌شوند تا بهترین مدل جهت پیش‌بینی انتخاب شود. در ادامه به ترتیب به نمونه مطالعات انجام شده ملی و بین‌المللی اشاره می‌گردد.

بررسی کلی عوامل تأثیرگذار بر تصادفات، بخشی از پژوهش‌های داخلی را شامل می‌شود. با روشهای آماری همانند

جایگاه و نقش حمل و نقل در ابعاد مختلف اجتماعی، اقتصادی و سیاسی جوامع برکسی پوشیده نیست. با توسعه روزافزون صنعت خودرو و نیاز به حمل و نقل گسترده که شریان هر کشور محسوب می‌شود، باید سیستم حمل و نقل کارآمدی طراحی گردد که جوابگوی نیاز جامعه بوده و از حداکثر ظرفیت آن استفاده شود. این امر جز با توسعه سیستم حمل و نقل هوشمند (ITS) میسر نمی‌شود. استفاده از دوربین‌های کنترل سرعت که بخشی از تجهیزات ITS هستند به سالهای نه چندان دور برمی‌گردد. اولین بار اواخر سال ۱۳۸۵ دوربین‌های ثابت کنترل سرعت در شبکه جاده‌ای ایران در آزادراه‌های تهران قم و کرج قزوین استفاده گردید [Armanna, 2008].

برطبق اظهارات مقامات راه و شهرسازی و کنترل ترافیک کشور، نصب و راه‌اندازی دوربین‌های کنترل سرعت در اواخر سال ۹۳، دو برابر سال ۹۲ هزینه دربرداشته است و تعداد دوربین‌ها از ۴۰۰ عدد به ۱۸۸۰ تا پایان سال ۱۴۰۰ خواهد رسید. همچنین از خرداد ماه سال ۹۴ تا پایان آبان ماه همان سال بیش از ۶ میلیون تخلف سرعت غیر مجاز در بزرگراه‌های تهران ثبت شده است، که به طور متوسط در هر بزرگراه ۱۵۰ هزار تخلف سرعت ثبت شده و هر دوربین ۱۸ هزار دستگاه خودروی متخلف را ثبت کرده است. این آمار و ارقام نشان‌دهنده استفاده چشمگیر از دوربین‌های کنترل سرعت به منظور کاهش تصادفات و تلفات جاده‌های برون‌شهری و بزرگراه‌های شهری است که مسلماً نصب و نگهداری این تجهیزات هزینه‌های زیادی را دربرخواهد داشت.

با استفاده از داده‌کاوی بر روی داده‌های دوربین‌های کنترل سرعت، می‌توان دانشی تولید نمود که در جهت بهبود مدیریت و برنامه‌ریزی شهری مفید بوده و تا حد زیادی در کاهش ترافیک شهری و تصادفات نقش داشته باشد [Medghalchi and Rahimi, Ahadi and Abbasi, 2014]. جمع‌آوری این داده‌ها و تحلیل آنها، زمینه را برای استفاده از ابزارهای پیشرفته تجزیه و تحلیل فراهم کرده و می‌توان با تکنیک‌های کشف دانش و داده‌کاوی، سیاست‌های کنترل حوادث ترافیکی را تغییر داده و با تقویت برنامه‌های پیشگیرانه، نرخ مرگ و میر را در جامعه کاهش داد.

بزرگراه تهران-کرج به عنوان پرترددترین و پرتخلف‌ترین بزرگراه در میان مبادی ورودی و خروجی شهر تهران است. با توجه به

مشخص شد که در سال اول پس از نصب این سامانه‌ها، کاهش ۳۵ درصدی نرخ تصادفات را در پی داشته است [Poorhasan, Ghasemi and Samataei, 2014].

تحقیقات مشابهی در این زمینه در کشورهای دیگر صورت گرفته است که به برخی از آنها اشاره می‌شود. بخشی از تحقیقات به بررسی موثر بودن دوربین‌های مختلف در تقاطع‌های شهری می‌پردازند. ارزیابی عملکرد دوربین‌های نظارتی چراغ‌های قرمز در تقاطع‌های شهری در آمریکا با استفاده از اطلاعات جمع‌آوری شده مربوط به سال‌های ۲۰۰۰ تا ۲۰۰۳ با استفاده از الگوریتم‌های درخت تصمیم و خوشه‌بندی در مرجع [Jafari and Samadian, 2012] صورت پذیرفته است. در پژوهشی دیگر که در شهر وینپینگ واقع در کانادا صورت گرفت به بررسی سودمندی برنامه حفاظتی عکس برداری از سرعت در تقاطع‌ها که منجر به تصادف می‌گردد پرداخته شده است. در این مطالعه با بهره‌گیری از رگرسیون لجستیک و سری‌زمانی ماهانه، تحلیلی بر روی تصادفاتی که در هنگام چراغ قرمز با گردش به راست و برخورد با عقب وسیله نقلیه در تقاطع‌ها روی می‌دهد، و همچنین تصادفاتی که با سرعت بالا منجر به تلفات و یا خسارات مالی در سر تقاطع‌ها می‌شدند، انجام شده است. در این تحلیل نشان داده شد که پارامترهای مدل شناسایی شده در تصادفاتی که در زمان چراغ قرمز روی می‌دهند معنادار و در مدل تصادفات مرتبط با سرعت معنادار نیستند و نتیجه‌گیری شد که با توجه به کاهش چشمگیر تصادفات به میزان ۴۶٪ در گردش به راست بعد از اجرای طرح عکس برداری از تخلفات، این رویکرد موفق ارزیابی می‌گردد [Vanlaar, Robertson and Marcoux, 2014].

دسته‌ای از تحقیقات به بررسی تاثیرات قوانین ملی راهنمایی و رانندگی می‌پردازند. در پژوهشی اثر اجرای برنامه ملی چند بعدی ایمنی جاده (IMSEVI) در کشور مکزیک در سه شهر مختلف انجام شد. در این پژوهش با بررسی روند و مدل‌سازی نرخ مرگ و میر، صدمات مالی و تصادفات با سری‌زمانی ARIMA، در قبل و بعد و حین اجرای این برنامه ملی در هر سه شهر به بررسی کارایی برنامه ملی پرداخته است [Chandran et al. 2014].

اوه، ایون و کیم [2016] در پژوهشی تاثیر تغییرات قوانین اخذ گواهینامه را بر تصادفات جاده‌ای در کشور کره بررسی نمودند. در این پژوهش از سری‌زمانی منقطع استفاده شده و نشان داده

رگرسیون می‌توان نسبت به بررسی تاثیر متغیرهای مستقل و وابسته بر یکدیگر در داده‌ها یا پیش‌بینی متغیرها اقدام کرد. به طور مثال می‌توان به پژوهش رحیم اف و صادقی [2009] اشاره نمود که به تعیین عوامل موثر در شدت تصادفات آزادراهی و طراحی مدل آنها (مطالعه موردی: آزادراه تهران-کرج) در سال ۱۳۸۸ پرداختند. در این پژوهش با استفاده از معیارهای آماری و رگرسیون ترتیبی بر روی داده‌های تصادفات سال‌های ۱۳۸۴-۱۳۸۷ توسط نرم افزار spss، عوامل تاثیرگذار در شدت تصادفات بررسی گردید. همچنین در پژوهشی دیگر سری‌زمانی صدمات ترافیک جاده‌ای منجر به مرگ، در طی سال‌های ۲۰۰۴ تا ۲۰۱۱ در ایران مورد بررسی قرار گرفته است. مدل شناسایی شده Arima(0,1,1)(0,1,1)12 است که مدل فصلی ضربی است. فراوانی حوادث در طول زمان با استفاده از رگرسیون پواسون برآورد شده و عواملی مانند جمعیت، میزان وسایل نقلیه تولید شده به عنوان فاکتورهای مهم در میزان صدمات ترافیک جاده‌ای مورد بررسی قرار گرفته است [Bahadori Monfared, 2013].

دسته دیگری از پژوهش‌های داخلی به بررسی سودمندی دوربین‌های ثبت تخلف سرعت می‌پردازند. بررسی نقش بازدارندگی دوربین‌های ثبت تخلف سرعت نصب شده در آزاد راه زنجان-تبریز در مراجع [Medghalchi and Abbasi, 2014] و [Rahimi, Ahadi and Abbasi, 2014] انجام شده است. در این پژوهش‌ها با استفاده از تحلیل‌های آماری و روند تاثیر آن بر کاهش تصادفات در آینده با مدل‌سازی زمانی ARIMA، کاهش (۳±۲۰ درصد) در سرعت متوسط وسایل نقلیه و کاهش (±۳۰٪) در تخلف سرعت مجاز نتیجه‌گیری شده است. تجزیه و تحلیل آمار سامانه ترددشماری در محورهای استان مرکزی نشان‌دهنده کاهش سرعت متوسط و هفت درصدی تخلفات سرعت غیرمجاز نسبت به سال قبل بوده است. همچنین توقف رشد تصادفات و تلفات جاده‌ای نسبت به سال‌های قبل نشان‌دهنده سودمندی دوربین‌های ثبت تخلف است. [Abedi and Mohseni, 2014].

در پژوهشی دیگر اثرات نصب سامانه‌های کنترل سرعت بر کاهش تصادفات بزرگراه‌های شهر تهران (شامل محور یادگار امام، شمالی-جنوبی و محور شهید همت، شرقی-غربی)، و یکی از مبادی خروجی شهر تهران (محور امام رضا) بررسی گردید و

۳. مواد و روشها

در این بخش ابتدا به معرفی روش‌های مورد استفاده در این پژوهش و تعاریف مورد نیاز می‌پردازیم. سپس فرآیند تحلیل داده‌ها را شرح می‌دهیم.

به منظور تحلیل داده‌ها از استاندارد CRISP [Shearer, 2000] که متداول‌ترین استاندارد داده‌کاوی است، استفاده شده است. مطابق این استاندارد برای تحلیل داده بایستی گام‌های فهم تجاری، درک داده، آماده‌سازی داده، مدلسازی، ارزیابی و استقرار اعمال شوند. نحوه اجرای این استاندارد در بخش ۳-۳ شرح داده شده است. به طور خاص در مراحل مدلسازی و ارزیابی، از تحلیل‌های توصیفی و استنباطی استفاده شده است. در تحلیل توصیفی سیر تخلفات، تعداد دوربین‌های فعال و حجم تردد در کنار یکدیگر بررسی شده و نتیجه‌گیری شده است. در تحلیل استنباطی، به دلیل تغییر داده‌ها در طول زمان، از تحلیل سری زمانی و رگرسیون که تحلیل‌های مناسب و متداول این نوع داده‌ها است، استفاده شده است. در تحلیل سری زمانی از الگوی باکس-جنکینز استفاده شده است و در تحلیل رگرسیون بعد از بررسی پیش‌شرط‌های لازم، مدل‌های خطی و غیر خطی متعدد مورد بررسی قرار گرفته است در بخش‌های ۳-۱ و ۳-۲ مبانی نظری این دو روش شرح داده شده‌اند.

۳-۱ مبانی سری زمانی

سری‌زمانی مجموعه‌ای از مشاهدات پی‌درپی در طول زمان است. این مشاهدات می‌تواند به‌طور مداوم و پیوسته در طول زمان ایجاد شده باشد یا به صورت یک مجموعه گسسته از زمان باشد [Chatfield, 2000]. در تعریفی دیگر سری‌های زمانی زیرشاخه‌ای از آمار و احتمالات و مجموعه‌ای از مشاهدات هستند که برحسب زمان مرتب شده‌اند. زمان در سری‌زمانی با توجه به نوع متغیر مورد بررسی می‌تواند روز، هفته، ماه و یا سال باشد [Hassanzadeh, Najafi and Moradnazar, 2015]. اهداف تجزیه و تحلیل سری‌زمانی شامل توصیف و مدلسازی مشاهدات گذشته، و همچنین پیش‌بینی مقادیر آینده بر مبنای مشاهدات پیشین است. توصیف در تحلیل سری‌زمانی به توصیف روند، الگوی فصلی و دوره‌ای دنباله‌ای از مشاهدات که در طول زمان قرار گرفته‌اند می‌پردازد.

شده است که تسهیل قوانین مرتبط با دریافت گواهینامه منجر به افزایش مرگ و میر ناشی از تصادفات شده است.

گوین در پژوهش خود در سال ۲۰۰۶ با استفاده از سری‌زمانی تک متغیره به پیش‌بینی زمان مورد نیاز برای سفرهای درون‌شهری پرداخت. او جهت تست مدل شناسایی شده از داده‌های سیستم‌های تشخیص ویدئویی در شهر آتلانتا واقع در منطقه جورجیا جمهوری کرواسی استفاده کرد و با در نظر گرفتن بازه‌های شش ساعته در شبانه روز به صورت هفتگی، الگوی فصلی جهت شناسایی ساعات اوج ترافیک مشخص کرد [Guin, 2006]. کاظمی و همکاران (2016) با استفاده از سری زمانی به پیش‌بینی تعداد موارد فوتی در تصادفات پرداختند. در ادامه با تعیین هزینه هر فوت میزان هدررفت تولید ناخالص ملی را پیش‌بینی نمودند.

بررسی تخلفات ثبت شده توسط دوربین‌های کنترل سرعت و تاثیرات این دوربین‌ها در کاهش تخلفات و تصادفات موضوع بخش دیگری از پژوهش‌ها است. در پژوهشی با نصب دوربین‌های ثبت تخلف سرعت غیرمجاز به صورت مخفی در منطقه‌ای آزمایشی در نیوزلند در مقایسه با منطقه‌ای مشابه دیگر که دوربین‌ها به صورت آشکارا در معرض دید رانندگان قرار داشت به بررسی تأثیر دوربین‌ها در کاهش تصادفات و تلفات پرداختند [Novoa, Perez, and Santamarina-Rubio, 2009]. همچنین ارزیابی اثربخشی دوربین‌های سرعت‌سنج در کاهش تعداد تصادفات و افراد مجروح در جاده‌های بارسلونا با استفاده از روش تجزیه و تحلیل سری‌های زمانی صورت پذیرفت که در پایان کاهش ۲۶ درصد تصادفات ملاحظه گردید [Keall, Povey, and Frith, 2001]. قدوس [2016] با استفاده از کلاس جدیدی از مدل‌های سری زمانی (GLARMA) به بررسی تأثیر بزرگراه هوشمند بر تصادفات در کشور انگلستان می‌پردازد. وی کاهشی تا میزان ۳۴ درصد را در تصادفات مشاهده می‌کند. پژوهش‌هایی با اهداف مشابه با استفاده از سری‌های زمانی در کشورهای دیگر نظیر کامبوج [Commandeur et al. 2017]، جمهوری بوتسوانا [Sebego et al. 2014]، قبرس، یونان، مجارستان، نروژ، و سوییس [Antoniou, Papadimitriou, and Yannis, 2014]، و چین [Zhao et al. 2016] صورت گرفته است.

۳-۱-۱ تعاریف اولیه

معمولاً برای تحلیل توصیفی یک سری زمانی تغییرات چهار مولفه اصلی شامل روند، تغییرات فصلی، دوره‌ای، و نامنظم مورد بررسی قرار می‌گیرند. روند یا تمایل بلند مدت نشان دهنده کاهش یا افزایش بلند مدت یک سری زمانی است. تغییرات فصلی تغییراتی هستند که در دوره‌های تناوبی کوتاه پیش می‌آیند. این تغییرات مربوط به عواملی هستند که به طریقی منظم و چرخه‌ای روی یک دوره کمتر از یک سال عمل می‌کنند. حرکات نوسانی در یک سری زمانی با دوره نوسان بیشتر از یک سال را تغییرات دوره‌ای می‌نامند. یک دوره کامل معمولاً ۷ تا ۹ سال طول می‌کشد. در هر سری زمانی تغییرات نامنظم یا تصادفی می‌تواند توسط عوامل پیش‌بینی نشده رخ دهد.

یک سری زمانی را هنگامی ایستا یا مانا می‌نامند که میانگین و واریانس آن به زمان بستگی نداشته باشد و در طول زمان ثابت باشند [Khorrami, Bozorgnia, 2007]. چنانچه یک سری زمانی در میانگین نایستا باشد، مهمترین ابزار برای ایستا نمودن آن تفاضلی کردن سری است. در روش تفاضلی گیری هر مشاهده را از مشاهده قبلی کم می‌کنند [Reinert, 2010]. اگر به مرور زمان تغییر پذیری 7 یک سری افزایش یابد بدین معنی است که سری زمانی نسبت به واریانسش نایستا است. ممکن است یک سری هم در میانگین و هم در واریانس نایستا باشد، در این صورت ابتدا باید واریانس آن را ایستا نمود. جهت پایداری واریانس می‌بایست از تبدیلات لازم از قبیل باکس-کاکس و لگاریتمی استفاده کرد. فرمول باکس-کاکس از نوع تبدیل توانی که در سال ۱۹۶۴ توسط باکس و کاکس معرفی شده است در رابطه (۱) نشان داده شده است.

$$T(X_t) = \frac{X_t^\alpha - 1}{\alpha} \quad (1)$$

α پارامتر تبدیل و X_t مقدار مشاهده شده در زمان t و $T(X_t)$ مقدار جدید مشاهده بعد از پایداری در واریانس است [Khorrami, Bozorgnia, 2007]. در صورتی که پارامتر تبدیل یک باشد نیازی به استفاده از فرمول بالا نیست زیرا نمودار در واریانس پایدار است. [Brockwell, Davis, 2009].

خودهمبستگی در تأخیر k عبارت است از همبستگی بین مشاهداتی که k واحد زمانی با هم فاصله دارند. تابع

خودهمبستگی را با γ_{ACF} نمایش می‌دهند [Javidi, Sharifi, 2011]. همبستگی بین X_t و X_{t+k} بعد از حذف اثر متغیرهای $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$ را ضریب همبستگی جزئی می‌نامند. این تابع را که به اختصار با γ_{PACF} نشان می‌دهند در تشخیص الگوهای احتمالی مولد داده‌ها استفاده می‌شود [Khorrami, Bozorgnia, 2007].

۳-۱-۲ مدلسازی سری زمانی

یکی از معروف‌ترین مدل‌های مورد استفاده در پیش‌بینی سری‌های زمانی مدل ARIMA است. به صورت ساده‌ای این مدل فرض می‌کند که سری زمانی به صورت خطی بوده و توزیع مشخصی دارد. مدل‌های احتمالی در دو شاخه مدل‌های فصلی و غیرفصلی دسته‌بندی می‌شوند و زیرشاخه‌های آنها مدل خودرگرسیو مرتبه p AR(p)، میانگین متحرک مرتبه q MA(q)، خودرگرسیو میانگین متحرک q ARMA(p,q)، خودرگرسیو میانگین متحرک تفاضلی یافته ARIMA(p,I,q) و خودرگرسیو میانگین متحرک تفاضلی یافته فصلی SARIMA(p,I,q) در ادامه به اختصار معرفی شده‌اند [Kalantari, Khavarzadeh, Giasi, 2013].

مدل‌های سری زمانی غیرفصلی ایستا شامل مدل‌های خودرگرسیو، مدل میانگین متحرک و مدل مرکب اتورگرسیو میانگین متحرک می‌باشند [Chatfield, 2000]. مدل مرکب اتورگرسیو میانگین متحرک که نایستا باشد با تفاضلگیری تبدیل به مدل ایستا می‌شود که به این مدل، مدل مرکب اتورگرسیو میانگین متحرک تفاضلی شده می‌گویند.

• **مدل خودرگرسیو مرتبه p یا AR(p):** این مدل از مدل‌های متداول تصادفی است و همانگونه که از نام آن مشخص می‌باشد بر روی جملات خود رگرسیون را اعمال می‌کند. ساختار اصلی آن مطابق با رابطه (۲) می‌باشد [Javidi, Sharifi, 2011].

$$Z_t = \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \dots + \varphi_p Z_{t-p} + a_t \quad (2)$$

φ_i همان پارامترهای مدل AR، و a_t نویز سفید است. Z_t را مدل خودرگرسیو با مرتبه p می‌نامند. در واقع رابطه ۲ وضعیت فعلی Z_t را بر حسب p مقدار گذشته به شکل رگرسیون خطی به نمایش گذاشته است.

مشاهدات تکرار می شوند $s=3$ باشد. این مدل به شکل عمومی

رابطه (۶) تعریف می شود [Reinert, 2010].

$$\alpha_p(A)\beta_p(A^s)w_t = \gamma_q(A)\delta_q(A^s)a_t \quad (6)$$

مدل فوق را یک مدل فصلی ضرب پذیر از درجه p و q و α_p و β_p به ترتیب با مرتبه p و γ_q و δ_q از مرتبه q و Q تعریف شده اند. w_t سری مشاهدات است که با d و D مرتبه تفاضلی کردن سری اصلی بدست آمده است. a_t فرآیند تصادفی محض با میانگین صفر و واریانس ثابت است.

۳-۱-۳ ارزیابی مدل

یکی از بهترین روش های اندازه گیری خطا، شاخص اطلاعاتی آکائیک (AIC) است که با رابطه (۷) اندازه گیری می شود:

$$AIC = 2K - 2 \ln(L) \quad (7)$$

پارامتر K بیانگر تعداد بازه های زمانی مورد نظر برای پیش بینی و مقدار حداکثر درست نمایی است. مدل بهتر مدلی است که کمترین مقدار AIC را داشته باشد [Khorrami, 2007]. شاخص شوارتز یا بیزین (SBC) با رابطه (۸) اندازه گیری می شود [Hassanzadeh, Najafi, 2015].

$$SBC = K \ln(n) - 2 \ln(L) \quad (8)$$

در این روش پارامتر K بیانگر تعداد بازه های زمانی پیش بینی شده و n تعداد بازه های زمانی مشاهده شده و L مقدار حداکثر درست نمایی است.

۲-۳ مبانی رگرسیون

برای اندازه گیری رابطه بین دو یا چند متغیر، شاخص های متعددی وجود دارد که همه شاخص ها ضریب همبستگی بین دو متغیر را اندازه گیری می کنند در صورتی که رابطه بین متغیرها معنی دار باشد می توان رابطه ریاضی را نوشت که با استفاده از این رابطه می توان مقادیر متغیر وابسته را از روی متغیر (های) مستقل پیش بینی نمود. رابطه ریاضی در رگرسیون می تواند خطی یا غیرخطی باشد.

رگرسیون خطی ممکن است ساده یا چندگانه باشد. رگرسیون خطی ساده شامل یک متغیر وابسته و یک متغیر مستقل است ولی

• مدل میانگین متحرک مرتبه q یا $MA(q)$: فرم عمومی

این مدل مطابق رابطه (۳) است [Javidi, Sharfi, 2011]:

$$Z_t = \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_p a_{t-p} + a_t \quad (3)$$

θ_i همان پارامترهای مدل MA هستند و a_j ها نویز سفید با میانگین صفر و واریانس σ^2 می باشند. Z_t را مدل میانگین متحرک با مرتبه q می نامند.

• مدل خودرگرسیون و میانگین متحرک با مرتبه p و q

می توان این مدل را فرآیندهای مرکب نیز نامید که از ترکیب دو مدل قبلی بدست می آید. فرم عمومی مدل مطابق رابطه (۴) است [Khorrami, Bozorgnia, 2007]:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_p a_{t-p} \quad (4)$$

پارامترهای این مدل مانند مدل های AR و MA تعریف می شوند.

• مدل خودرگرسیون میانگین متحرک تفاضلی با مرتبه p, q

این مدل توانایی دسته بندی گسترده ای از سری های زمانی که ناپایستا هستند را دارد. با توجه به این که در عمل بیشتر سری های زمانی ناپایستا هستند در این مدل قرار می گیرند که با تفاضلی کردن از مرتبه I به سری ایستا تبدیل می شوند. فرم عمومی مدل مطابق رابطه (۵) است [Reinert, 2010]:

$$W_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_p a_{t-p} \quad (5)$$

سری W_t با d مرتبه تفاضلی کردن سری اصلی بدست آمده است. پارامترهای این مدل مانند مدل های AR و MA تعریف می شوند.

در عمل بیشتر سری های زمانی شامل یک مولفه فصلی دوره ای هستند که هر s مشاهده تکرار می شوند. s نشان دهنده طول دوره فصلی است. حال ممکن است این مولفه اصلی در مدل اتورگرسیون یا میانگین متحرک و یا هر دو نهفته باشد. در فرآیند خودرگرسیون میانگین متحرک تفاضلی فصلی با مرتبه p, q یا s SARIMA(p,d,q)(P,D,Q) مولفه دوره ای بودن هر کدام از پارامترهای مدل است. مثلاً برای داده های ماهانه ممکن است $s=12$ یا برای داده های فصلی که در پایان هر سه ماه

استفاده از الگوریتم‌های داده‌کاوی برای تحلیل سودمندی

این استاندارد، با بررسی اولیه به فهم مسئله و درک بهتری از داده‌ها دست یافتیم و محدودیت‌های موجود در داده‌ها شناسایی و بازه‌های زمانی مختلف جهت شناسایی بهتر الگوها مورد بررسی قرار گرفتند. در مرحله پیش پردازش داده‌ها، فیلدهای مورد نیاز پژوهش (شامل نوع تخلف، کدبزرگراه، دوربین ثبات، روز و ماه وقوع تخلف، بازه زمانی تخلف صورت گرفته در شبانه روز، نوع کاربری وسیله نقلیه، میزان جریمه و غیره) از بین فیلدهای پایگاه داده انتخاب و در بازه‌های زمانی ماهانه، هفتگی و روزانه مورد بررسی قرار گرفتند. حداقل تعداد مشاهدات در سری‌زمانی ۵۰ عدد است [Khorrami, Bozorgnia, 2007]. لذا انتخاب بازه ماهانه برای بزرگراه با توجه به بازه زمانی کلی داده‌ها (نیمه دوم سال ۹۰ تا پایان اردیبهشت ۹۳) مناسب نیست. با انتخاب بازه روزانه در سری‌زمانی روند نامنظم و تصادفی در سری مشاهده می‌گردد که استخراج الگو از آن کاری بس مشکل و تقریباً غیرممکن است. بهترین بازه زمانی بازه هفتگی است که به خوبی می‌توان سیر تغییرات را در شش ماهه اول و دوم سال مشاهده کرد.

قابل ذکر است در مرحله پیش‌پردازش داده‌ها علاوه بر تجمیع داده‌های مورد نیاز، داده‌های پرت از سری‌زمانی حذف و بازه‌های زمانی که فاقد داده هستند (مقادیر گمشده) شناسایی و درون‌یابی گردیدند. در این مرحله نیز کدگذاری‌های لازم برای داده‌های رشته‌ای جهت تحلیل نرم افزاری صورت پذیرفت. مرحله آماده سازی داده‌ها در داده‌کاوی، فرآیند زمان‌بر و با اهمیتی است که حدوداً ۶۰ تا ۷۰ درصد زمان انجام کل فرآیند داده‌کاوی را به خود اختصاص می‌دهد [Hajilou, 2005].

در این مقاله سری زمانی براساس الگوی باکس-جنکینز مدل‌سازی می‌گردد. باکس-جنکینز یک الگوی تکراری است تا بتواند مدل مناسب را شناسایی و تا انتخاب بهترین مدل ادامه پیدا می‌کند. فرآیند انتخاب مدل مناسب در شکل ۱ نشان داده شده است. به طور مختصر جهت مدل‌سازی سری زمانی به ترتیب به حذف مولفه‌های مانا، شناسایی مدل آریمای مناسب، برآوردکردن پارامترها، تحلیل باقیمانده‌ها و ارزیابی پیش‌بینی مدل با داده‌های تست و در نهایت به پیش‌بینی تخلفات ماه‌های خرداد تا شهریور سال ۹۳ می‌پردازیم.

در ادامه پژوهش، با برآورد مدل‌های خطی و غیرخطی رگرسیون برای متغیرهای مستقل و وابسته، مدلی که بیشترین

رگرسیون خطی چندگانه به ارزیابی رابطه یک متغیر وابسته و چند متغیر مستقل می‌پردازد. در رگرسیون غیرخطی، ارتباط بین متغیرهای مستقل با متغیر وابسته به صورت خطی نیست. ممکن است مدل از تابع لگاریتمی، نمایی، توانی، سهمی و غیره پیروی کند. انواع مدل‌های رگرسیونی غیرخطی مورد استفاده در جدول ۱ آورده شده‌اند. این مدل‌ها در کاربردهای دیگر نیز مورد استفاده قرار می‌گیرند [Hemmatfar et al. 2011]. در روابط جدول ۱، Y متغیر وابسته، X متغیرهای مستقل و β ضریب مدل و α یک ثابت است.

جدول ۱. مدل‌های رگرسیونی غیرخطی متداول

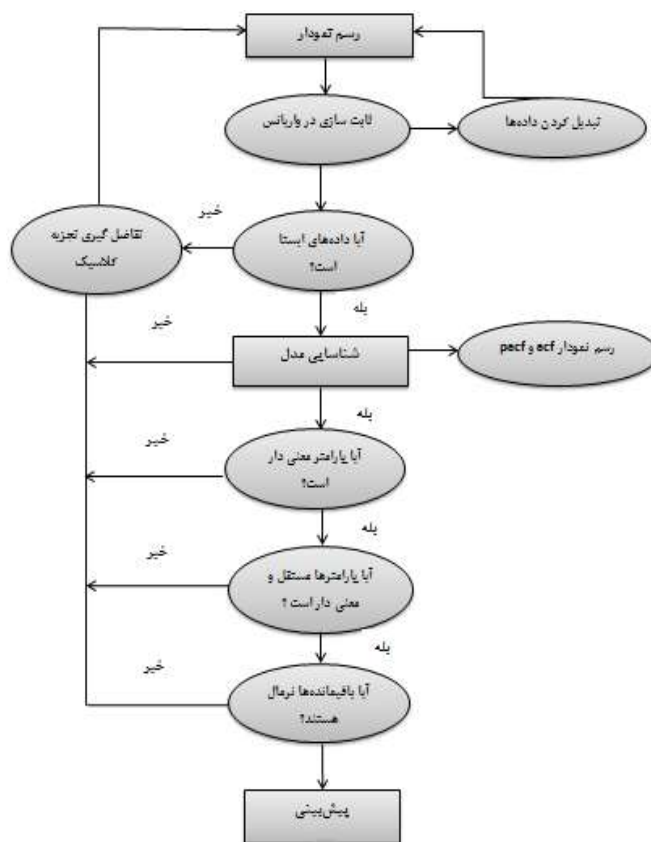
نام مدل	رابطه
Exponential regression	$Y = \alpha \exp(\beta X)$
S regression	$Y = \exp(\alpha + \beta/x)$
Quadratic regression	$Y = \alpha + \beta_1 X + \beta_2 X^2$
inverse regression	$Y = \alpha + \beta / X$
Cubic regression	$Y = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$
Compound regression	$Y = \alpha \beta^X$
Logarithmic Regression	$Y = \alpha + \beta \ln X$
Power regression	$Y = (\alpha X)^\beta$

۳-۳ فرآیند تحلیل

در این پژوهش داده‌های تخلفات ثبت شده توسط دوربین‌های کنترل سرعت بزرگراه تهران-کرج را در بازه (نیمه دوم سال ۹۰ تا پایان اردیبهشت ۹۳) مد نظر قرار داده‌ایم. ابتدا به تحلیل توصیفی نوسان تخلفات سرعت غیرمجاز در بزرگراه تهران-کرج، براساس داده‌های دوربین‌های فعال در ماه‌های نظیر در سال‌های متوالی می‌پردازیم. در مرحله بعد با استفاده از مدل‌سازی باکس جنکینز^{۱۱}، میزان تخلفات و ترددها را مدل نموده و اجزاء سری‌زمانی آنها را مورد بررسی قرار داده و در نهایت جهت پیش‌بینی میزان تخلفات مورد استفاده قرار می‌دهیم.

در این پژوهش از استاندارد CRISP [Shearer, 2000] که متداول‌ترین استاندارد داده‌کاوی است، استفاده شده است. مطابق

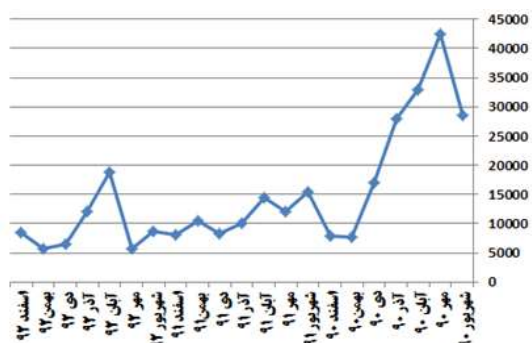
مقدار ضریب تعیین تعدیل شده و آماره F را دارد و مدل رگرسیونی در آن معنادار است، به عنوان مدل مناسب برآورد



شکل ۱. مراحل مدلسازی الگوی باکس-جنکینز [Hassanzadeh, Najafi, Moradnazar, 2015]

۴-۱ تحلیل توصیفی

در شکل های ۲ و ۳ سیر فراوانی تخلفات و دوربین های فعال این بزرگراه در بازه انتخابی آورده شده است.



شکل ۲. نمودار سیر تخلفات سرعت در بزرگراه تهران کرج

می شود. قابل ذکر است میزان تخلفات ثبت شده توسط دوربین ها به عنوان متغیر وابسته و حجم تردها و میانگین تعداد دوربین های فعال در ماه به عنوان متغیر مستقل در نظر گرفته شده است. همچنین بادر نظر گرفتن شرایط ثابت در تمامی ماه ها به بررسی سیر تخلفات بزرگراه تهران کرج با استفاده از متغیرهای وابسته و مستقل می پردازیم.

۴. یافته ها

بزرگراه تهران کرج در واقع قسمتی از آزادراهی است که از تهران به سوی شمال غرب تا شهر تبریز ادامه دارد و از شهرهای کرج، قزوین و زنجان می گذرد. این بزرگراه به طول ۵۱ کیلومتر از پرتددترین مبادی ورودی و خروجی شهر تهران است.

استفاده از الگوریتم‌های داده‌کاوی برای تحلیل سودمندی

همبستگی بزرگتر باشد ارتباط بین متغیرها قویتر است. با این احتساب در بزرگراه تهران کرج همبستگی بیشتری بین متغیرهای تخلفات و حجم تردها وجود دارد. مدل پیرسون در تعیین همبستگی مورد استفاده قرار گرفته است.

جدول ۲. بررسی همبستگی بین متغیرهای مستقل با وابسته

بزرگراه	متغیرهای مستقل	ضریب همبستگی	سطح معناداری
تهران	حجم تردد	۰/۲۹۵	۰/۰۰۸
کرج	میانگین تعداد دوربین‌های فعال	۰/۲۲۵	۰/۰۱۱

همبستگی بین متغیرهای مستقل هم در این آزمون محاسبه شد که هیچ رابطه معناداری بین متغیر تردها و میانگین دوربین‌های فعال وجود نداشت، و سطح معناداری بالای ۰,۰۵ به اندازه ۰,۴۶۳ با ضریب همبستگی پیرسون ۰,۱۱- محاسبه گردید. مقدار این ضریب بین ۱- تا ۱ تغییر می‌کند که ۱ به معنای همبستگی مثبت کامل، ۰ به معنی نبود همبستگی، و ۱- به معنی همبستگی منفی کامل است. حال با توجه به وجود همبستگی معنادار بین متغیرهای مستقل با وابسته که در این بخش بررسی گردید، می‌توان مدلسازی سری‌زمانی و رگرسیون را انجام داد.

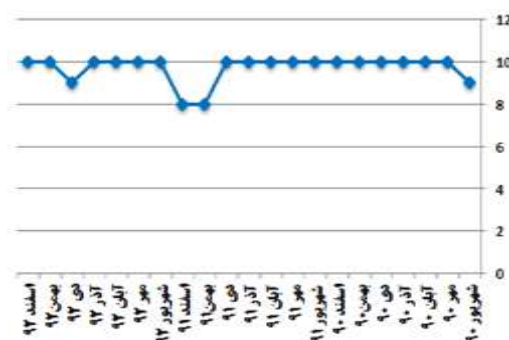
۳-۴ سری زمانی

در این بخش نتایج تحلیل سری زمانی را ارائه می‌دهیم.

۱-۳-۴ مدلسازی متغیر تخلفات

همان‌طور که در مدلسازی به روش باکس جنکینز بیان شد، مرحله اول در مدلسازی سری‌های زمانی رسم نمودار است. از روی نمودار می‌توان تا حدودی وجود الگوهای دوره‌ای، فصلی و روند آن و همچنین نیاز به ایستاسازی در واریانس یا میانگین را مشخص نمود.

در شکل ۴ نایستایی در میانگین و واریانس مشاهده می‌شود. بدین‌منظور با استفاده از آزمون باکس کاکس نایستایی آن در واریانس، و با استفاده از ترسیم نمودار acf و $pacf$ و تنزل کند



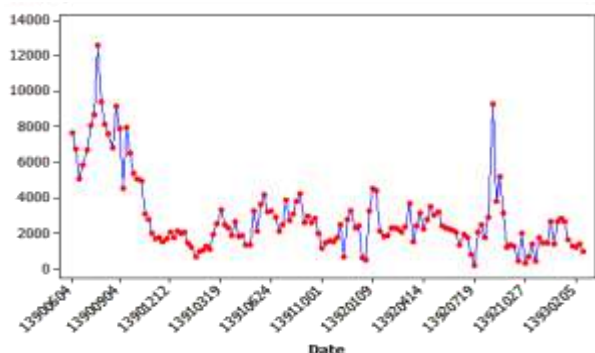
شکل ۳. سیر میانگین تعداد دوربین‌های فعال در بزرگراه تهران

کرج

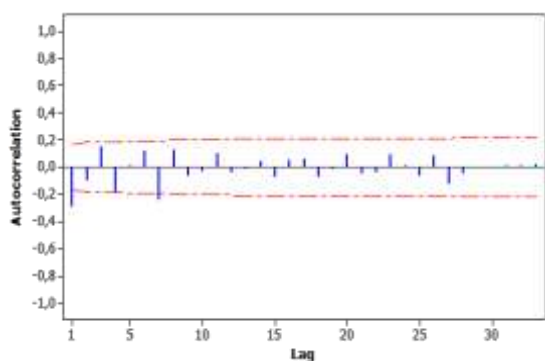
همان‌طور که در شکل ۲ نشان داده شده است، تخلفات در سال ۹۰ بیشتر از سال‌های ۹۱ و ۹۲ بوده است. در پایان سال ۹۰ در ماه‌های بهمن و اسفند شیب نمودار به طور قابل ملاحظه‌ای کاهش می‌یابد. در صورتی که در نمودار ۳ تعداد میانگین دوربین‌ها ثابت است. این خود می‌تواند دلیلی مانند افزایش تعرفه جریمه‌ها در دی ماه سال ۹۰ داشته باشد، که باعث کاهش قابل ملاحظه تخلفات شده است. در بهمن ماه و اسفند ماه سال ۹۱ کاهش دوربین‌ها را شاهد هستیم که در نمودار تخلفات هم کاهش تخلفات ملاحظه می‌شود. در آبان ماه سال ۹۲ افزایش تخلفات را داریم که با ثابت بودن میانگین دوربین‌ها در آن تاریخ همخوانی ندارد و ممکن است از شرایط دیگری مانند حجم تردها ناشی شده باشد. در کل می‌توان گفت اگر خطی را بر روی سیر تخلفات ترسیم کنیم شیب نزولی خواهد داشت.

۲-۴ تحلیل استنباطی

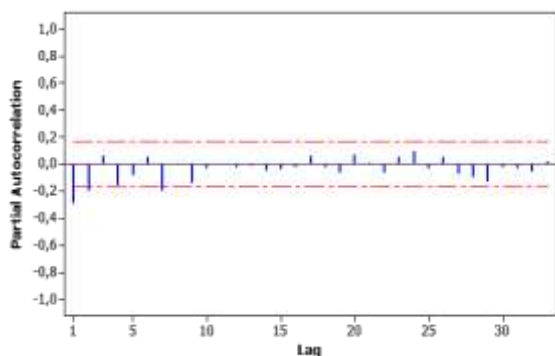
قبل از شروع به فرآیند مدلسازی می‌بایست رابطه معنادار بین متغیرهای مستقل و وابسته مورد بررسی قرار گیرد. قبل از محاسبه همبستگی بین متغیرها، مقادیر پرت شناسایی و با میانگین داده‌های قبل و بعد خود جایگزین شدند. داده‌های پرت مذکور در در سه ناحیه از متغیر تردد در بزرگراه تهران کرج مشاهده شدند، که جایگزینی مناسب صورت پذیرفت. با استفاده از نرم افزار SPSS همبستگی بین متغیرهای مستقل با وابسته پژوهش می‌شود. جدول ۲ نتایج مربوطه را برای بزرگراه تهران کرج نشان می‌دهد. متغیرهای مستقل رابطه معناداری با متغیر وابسته که همان تخلفات سرعت غیرمجاز است، دارند. هرچه ضریب



شکل ۴. نمودار سیر تخلفات سرعت غیرمجاز در بزرگراه تهران کرج (هفتگی)



شکل ۵. نمودار acf متغیر تخلفات در بزرگراه تهران کرج



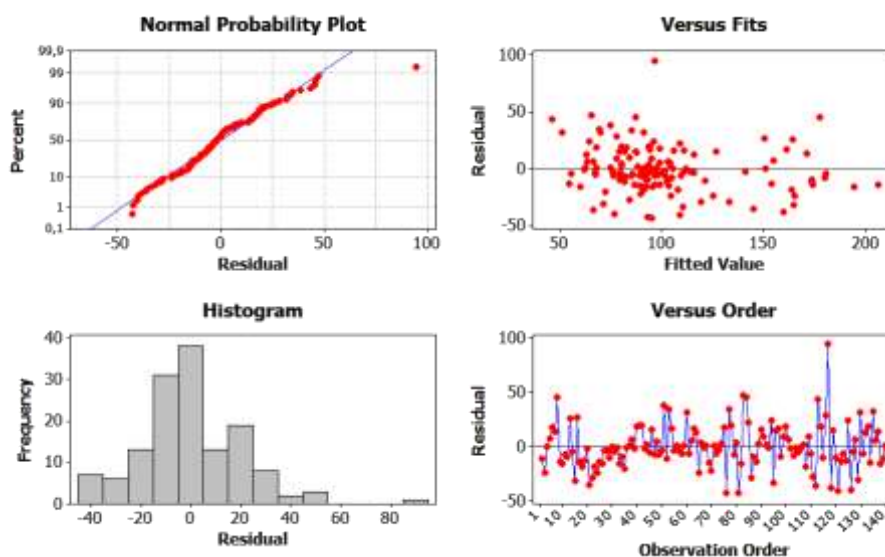
شکل ۶. نمودار pacf متغیر تخلفات در بزرگراه تهران کرج

نمودار acf نسبت به pacf نایستایی آن در میانگین تأیید می شود. در مرحله بعد به ترسیم نمودار acf و pacf از تخلفات که نسبت به واریانس و میانگین ایستا شده است، می پردازیم. همان طور که در شکل های ۵ و ۶ نشان داده شده است، در Lag 7 وقفه معنی دار وجود دارد. با توجه به این امر امکان الگوی فصلی در بازه ۷ هفته وجود دارد.

بعد از تخمین مدل، پارامترهای آن را از لحاظ معناداری بررسی می کنیم. مدل آزمایشی $ARIMA(1,1,1)(1,0,0)_7$ برآورد شد که پارامترهای آن بررسی شده و معنادار می باشند. در مرحله بعد به تجزیه و تحلیل باقیمانده های آن می پردازیم. در شکل ۷ نرمال بودن باقیمانده ها، از دو نمودار ستون چپ تصویر قابل مشاهده است. در نمودار احتمال نرمال نقاط در امتداد یک خط مستقیم قرار گرفتند و هیستوگرام باقیمانده ها کیفی شکل است که نرمال بودن باقیمانده ها را نشان می دهد. برای بررسی فرض ثابت بودن واریانس باقیمانده ها می بایست نمودار باقیمانده ها در مقابل مقادیر برازش شده بررسی شود. اگر ساختار خاصی را نشان ندهد و شکل کیفی مانند نداشته باشد، نشان دهنده ثابت بودن واریانس باقیمانده ها است. در شکل ۷، نمودار سمت راست بالا این امر نشان داده می شود. نمودار سمت راست پایینی هم، نمودار باقیمانده ها را در طول زمان نشان می دهد که دارای ساختار خاصی ناست و در اطراف سطح افقی صفر پراکنندگی مستطیلی بدون روندی را دارد.

با استفاده از آزمون پرت مانتو می توان خودهمبستگی باقیمانده ها را کنترل نمود. در جدول ۳ با بررسی همبستگی بین باقیمانده ها سطح معناداری آن نشان داده شده است. در صورتی که سطح معناداری (P-value) بیشتر از ۰/۰۵ باشد فرضیه همبستگی بین باقیمانده ها رد می شود.

استفاده از الگوریتم‌های داده‌کاوی برای تحلیل سودمندی



شکل ۷. نمودارهای تجزیه و تحلیل باقیمانده‌ها در مدل تخلفات بزرگراه تهران کرج

۸۷۷/۱۰۱	۸۷۱/۱۹	ARIMA(1,1,0)(0,0,1)	۳
۸۷۲/۵۷۶	۸۶۶/۶۶۶	ARIMA(0,1,1)(1,0,0)	۴
۸۷۵/۸۸۹	۸۶۷/۰۲۴	ARIMA(1,1,1)(0,0,1)	۵
۸۹۰/۳۳	۸۷۸/۵۱۰	ARIMA(1,1,1)(1,1,1)	۶
۸۷۱/۰۶۱	۸۶۵/۱۵	ARIMA(1,1,1)(1,0,0)	۷
۸۷۴/۵۴۵	۸۶۵/۶۸	ARIMA(2,1,0)(0,0,1)	۸

فرمول کامل مدل انتخاب شده به صورت فصلی ضربی در معادله (۹) نشان داده شده است.

$$(1 - 0.3485B)(1 + 0.2395B^7) w_t = a_t + 0.6923Ba_t \quad (9)$$

۴-۳-۲ مدل‌سازی متغیر تردها

بعد از انجام مراحل مدل‌سازی سری‌زمانی متغیر تردها، مدل آزمایشی 53 ARIMA(0,1,0)(0,0,1) پیشنهاد می‌گردد. و با بررسی مناسب بودن مدل‌های دیگر، انتخاب مدل فوق به عنوان مدل نهایی قطعی می‌گردد. قابل ذکر است در برازش جامع‌تر بسیاری از مدلها با دوره‌های فصلی احتمالی تست می‌شوند، اما

جدول ۳. نتایج آزمون همبستگی بین باقیمانده‌ها در مدل

تخلفات بزرگراه تهران کرج

IAG	۱۲	۲۴	۳۶	۴۸
Chi-Square	۵/۵	۱۰/۴	۲۳/۴	۲۹/۰
DF	۸	۲۰	۳۲	۴۴
p-value	۰/۷۰۴	۰/۹۰۶	۰/۸۶۴	۰/۹۶

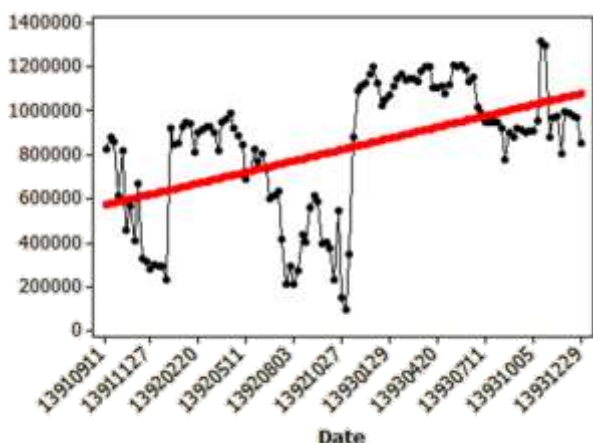
در نهایت با برازش جامع‌تر مدل‌ها، می‌توان مدل‌های بیشتری را مورد بررسی قرار داده و مراحل فوق را بر روی آنها انجام داد. در صورتی که چندین مدل مناسب شناخته شدند با استفاده از معیار آکائیک و شوارتز بیزین بهترین مدل از بین آنها مشخص می‌شود. جدول ۴ لیست مدل‌هایی که برای مدل تخلفات بزرگراه تهران کرج مناسب شناخته شده‌اند را به همراه معیارهای آن نشان می‌دهد. در نهایت مشاهده می‌شود که مدل آزمایشی دارای مقدار کمتری در معیار آکائیک (AIC) و شوارتز بیزین (SBC) است لذا مناسب‌تر شناخته شده است.

جدول ۴. بررسی مدل‌های متغیر تخلفات سرعت

ردیف	مدل	AIC	SBC
۱	ARIMA(0,1,1)	۸۷۰/۸۶	۸۷۷/۹۲۱
۲	ARIMA(1,1,1)	۸۷۲/۰۱	۸۷۳/۸۱۵

۲۶,۶۵۳۷ واحد زمانی است. با توجه به مدل $ARIMA(1,1,1)(1,0,0)$ 7 که در بخش قبل بدست آوردیم، مدل مربوطه دارای تغییرات فصلی در بازه ۴۹ روز است. یادآوری می‌گردد بازه زمانی تمامی مدل‌ها هفتگی است. بنابراین در کمتر از شش ماه الگوی تخلفات با خودرگرسیو فصلی در هر سال تکرار می‌شود. تغییرات نامنظم بعد از حذف روند و الگوهای فصلی برجای می‌ماند که نشان دهنده عدم وابستگی بین باقیمانده‌های مدل است. در شکل ۷ در تصویر چهارم تغییرات نامنظم باقیمانده‌ها نشان داده شده است که تصادفی بودن آنها مشاهده می‌شود.

۴-۳-۴ تحلیل اجزاء مدل سری زمانی متغیر تردها



شکل ۹. سری زمانی به همراه خط روند متغیر تردد

معادله خط روند نمودار سری زمانی شکل ۹ به صورت رابطه (۱۲) است:

$$y_t = 569169 + 4243t \quad (12)$$

ضریب تشخیص در رابطه (۱۲) برابر با ۰,۲۴ می‌باشد. با توجه به مدل $ARIMA(0,1,0)(0,0,1)$ 53 که در بخش قبل بدست آوردیم، مدل مربوطه دارای تغییرات دوره ای ۳۷۱ روزه است. که در مدت زمان تقریباً یک سال الگوی تردها با مدل میانگین متحرک فصلی تکرار می‌شود. در شکل ۹ مشهود است که حجم تردها در فروردین ماه هر سال به طور چشمگیری در این بزرگراه افزایش می‌یابد. از معادلات خط ۱۱ و ۱۲ در می‌یابیم، با افزایش روند تردها همچنان روند تخلفات در این بزرگراه رو به کاهش

در جدول ۵ فقط مدل‌هایی که پارامترهای معنادار و شروط لازم در تجزیه و تحلیل باقیمانده‌های مدل را دارند، ذکر گردیدند.

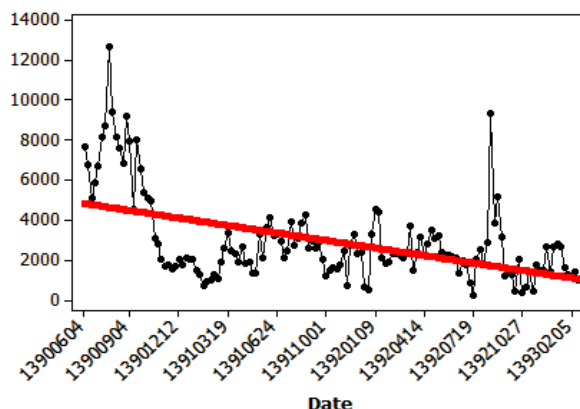
جدول ۵. بررسی مدل‌های متغیر تردها

ردیف	مدل	AIC	SBC
۱	$ARIMA(0,1,0)(0,0,1)$ 48	۲۸۵۱/۶۳۷	۲۸۵۴/۴۲۴
۲	$ARIMA(0,1,0)(1,0,1)$ 48	۲۸۳۵/۵۲۷	۲۸۴۱/۱۰۱۱
۳	$ARIMA(0,1,0)(0,0,1)$ 53	۲۷۸۴/۱۶۷	۲۷۸۶/۹۵۴
۴	$ARIMA(0,1,0)(1,0,1)$ 53	۲۸۰۹/۰۶۱	۲۸۱۴/۶۳۴
۵	$ARIMA(0,1,0)(0,0,1)$ 54	۲۷۹۹/۴۲۲	۲۸۰۲/۲۰۹
۶	$ARIMA(0,1,0)(1,0,1)$ 54	۲۸۰۶/۲۸۶	۲۸۱۱/۸۶۰

فرمول کامل مدل انتخاب شده در رابطه (۱۰) نشان داده شده است.

$$y = a_t + 0.7686B^{53} a_t \quad (10)$$

۴-۳-۴ تحلیل اجزاء مدل سری زمانی متغیر تخلفات



شکل ۸. سری زمانی به همراه خط روند متغیر تخلفات سرعت در بزرگراه تهران کرج

معادله خط روند نمودار سری زمانی شکل ۸ به صورت فرمول (۱۱) است:

$$y_t = 4878 - 26.6537t \quad (11)$$

ضریب تشخیص در رابطه (۱۱) برابر با ۰,۲۶ می‌باشد. همچنین فرمول (۱۱) نشان دهنده کاهش تخلفات در زمان به اندازه

$$V2 = MC^{B2} + \alpha = MC^{3.520} \quad (14)$$

با توجه نتایج جدول ۶ مدل توانی بدون عرض از مبدأ است لذا $(\alpha=0)$ است. بعد از بدست آوردن معادلات فوق، رگرسیون خطی که جهت محاسبه فرمول نهایی از مدل‌های فوق گرفتیم مطابق رابطه (۱۵) است:

$$V = 0.617V1 + 0.80V2 \quad (15)$$

قابل ذکر است قبل از مدلسازی پیش‌فرض‌های رگرسیون شامل (معنی دار بودن متغیرها، توزیع نرمال خطاها با میانگین صفر، مستقل بودن خطاها از یکدیگر، عدم وجود رابطه همخطی بین متغیرهای مستقل) بررسی گردیده است.

• معنی‌دار بودن متغیرها: که با بررسی همبستگی بین متغیرهای مستقل و وابسته تحقیق می‌شود. با توجه به اهمیت این پیش‌فرض در همبستگی ارتباط معنادار بین متغیرهای مستقل و وابسته کنترل گردید.

• توزیع نرمال خطاها با میانگین صفر: نمودار توزیع نرمال خطاها در بزرگراه‌های تهران کرج با بررسی هیستوگرام نرمال باقیمانده‌ها صورت پذیرفت. میانگین بزرگراه ۱۶-۳۷۱E است که تقریباً برابر با صفر است. انحراف معیار آن هم برابر با ۰,۹۸۶ که نزدیک به یک است.

• استقلال خطاها: عدم همبستگی بین خطاها با آزمون دوربین واتسون کنترل می‌شود. این آماره برابر با ۱,۵۲۴ هر آماره مقداری بین ۱,۵ تا ۲,۵ را دارد. بنابراین بین خطاها وابستگی وجود ندارد.

عدم همخطی بین متغیرهای مستقل: با بررسی آماره‌های ضریب سنجش تحمل^{۱۲} و عامل تورم تلورانس^{۱۳}(VIF) عدم همخطی را بین متغیرهای مستقل کنترل می‌کنیم. آماره تلورانس هر دو متغیر مستقل برابر ۰,۹۹۵ و عامل تورم آن برابر با ۱,۰۰۵ است. مقدار ضریب تلورانس بین ۰ تا ۱ و مقدار عمل تورم تلورانس هم بین ۱ تا ۱۰ است. بنابراین همخطی بین متغیرهای مستقل مشاهده نشد.

است. با توجه به این امر و بررسی که بر روی متغیر میانگین دوربین‌های فعال در این بزرگراه انجام دادیم، سودمندی دوربین‌های ثبت تخلف در کاهش تخلفات در این بزرگراه تضمین می‌شود. تغییرات نامنظم این سری هم مانند تخلفات از نمودار باقیمانده‌ها نسبت به زمان بدست می‌آید.

۴-۴ رگرسیون

در این بخش متغیرهای مستقل (حجم تردها و میانگین دوربین‌های فعال) نسبت به متغیر وابسته میزان تخلفات با مدل‌های مختلف رگرسیونی سنجیده می‌شود. همانگونه که در جدول ۶ مشاهده می‌شود، به ترتیب ضریب تعیین تعدیل شده، آماره مدل F، سطح معناداری مدل و نتیجه تایید و یا رد شدن مدل رگرسیونی برآورد شده و نشان داده می‌شود. سطح معناداری مدل کمتر از ۵ درصد نشان دهنده آن است که آماره F مدل رگرسیونی به دست آمده از مقدار بحرانی جدول بزرگتر است و در نتیجه آماره F مدل در ناحیه رد H_0 قرار می‌گیرد و معناداری مدل تأیید می‌شود. همچنین در جدول فوق نتایج برآورد مدل‌های خطی و غیرخطی رگرسیونی با در نظر گرفتن عدم یا وجود عرض از مبدأ مشاهده می‌شود. همان‌طور که ملاحظه می‌شود مقدار آماره F و ضریب تعیین تعدیل کننده مدل غیر خطی توانی در حالت بدون عرض از مبدأ، در هر دو متغیر مستقل از همه مدل‌ها بیشتر می‌باشند. لذا با توجه به بررسی‌هایی که بر روی همه مدل‌های رگرسیون صورت گرفته است مدل توانی بدون عرض از مبدأ به عنوان مدل رسمی رگرسیون بزرگراه تهران کرج انتخاب می‌شود. مدل غیر خطی توانی متغیر تخلفات V، با متغیرهای مستقل حجم تردها T و میانگین تعداد دوربین‌ها MC را به صورت جداگانه بدست آورده، و سپس رگرسیون خطی از مدل‌هایی که بدست آوردیم را آموزش می‌دهیم. فرمول مدل غیرخطی برای هر کدام از متغیرهای مستقل مطابق روابط (۱۳) و (۱۴) است.

$$V1 = T^{\beta1} + \alpha = T^{0.564} \quad (13)$$

بی بی هانیه دقیق نژاد، هدی مشایخی، زهرا نوراله

جدول ۶. نتایج برآورد مدل‌های خطی و غیرخطی در رگرسیون بزرگراه تهران کرج (A: بدون عرض از مبدأ، B: با عرض از مبدأ)

مدل‌های رگرسیونی										عنوان متغیرها
Power		Cubic		Quadratic		Logarithm		Linear		
B	A	B	A	B	A	B	A	B	A	
۰/۲۳۹	۰/۹۹۵	۰/۰۹۶	۰/۷۷۵	۰/۱۰۷	۰/۷۷۸	۰/۰۳۸	۰/۷۵۵	۰/۰۰۹	۰/۶۷۷	ضریب تعیین شده
۲۴/۵۲	۱۴۵۳۴/۸۲	۳/۶۳۳	۸۶/۹۹	۳/۶۳۳	۱۳۲/۲	۳/۹۵	۲۳۱/۷۵	۱/۶۷	۱۵۸/۳۰	تخلفات سرعت غیرمجاز
۰/۰۰۱	۰/۰۰۱	۰/۰۱۷	۰/۰۰۱	۰/۰۱۷	۰/۰۰۱	۰/۰۵	۰/۰۰۱	۰/۲	۰/۰۰۱	سطح معناداری نتیجه تأیید یا رد
تأیید	تأیید	تأیید	تأیید	تأیید	تأیید	تأیید	تأیید	رد	تأیید	ضریب تعیین شده
۰/۰۸۱	۰/۹۹۳	۰/۱۰۷	۰/۶۷۳	۰/۱۰۸	۰/۶۷	۰/۰۴۶	۰/۶۶۲	۰/۰۵۲	۰/۶۷	تخلفات سرعت غیرمجاز
۱۳/۳۷	۱۸۷۷۳/۳	۹/۴۱۸	۱۴۵/۹۵	۹/۴۵۹	۱۴۵/۹	۷/۶۹۶	۲۷۷/۷۲	۸/۶۴۳	۲۸۷/۶۲	میانگین تعداد دوربین‌های فعال
۰/۰۰۱	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱	۰/۰۰۲	۰/۰۰۱	۰/۰۰۶	۰/۰۰۲	۰/۰۰۱	۰/۰۰۱	سطح معناداری نتیجه تأیید یا رد
تأیید	تأیید	تأیید	تأیید	تأیید	تأیید	تأیید	تأیید	تأیید	تأیید	

۵. نتیجه گیری

در بخش‌های پیشین به تحلیل داده‌های تخلفات و تردد در بزرگراه تهران-کرج با استفاده از تحلیل سری زمانی و رگرسیون پرداختیم. مشاهده گردید که نمودار سری زمانی متغیر تخلفات شیب منفی و نمودار سری زمانی حجم تردد ها شیبی مثبت داشت. این امر نشان می‌دهد که در عین افزایش میزان حجم تردها در بزرگراه تهران کرج، کاهش تخلفات مشاهده می‌شود که خود دلیلی بر سودمندی دوربین‌های ثبت تخلف سرعت است.

در تحلیلی که در بخش (۴-۳-۳) صوت گرفت، نمودار تخلفات سرعت در بازه زمانی شهریور ماه سال ۹۰ تا اردیبهشت ماه سال ۹۳ به صورت هفتگی محاسبه گردید. با در نظر داشتن بازه زمانی پیوسته از سال ۹۰ الی ۹۳ به صورت هفتگی، معادله خط روند با شیب منفی ۲۶,۶۵۳۷ ملاحظه گردید. به منظور بررسی نمودار تخلفات در ماه‌های مشابه، هفت ماه آخر سال‌های ۹۰ الی ۹۲ را که داده‌های آن‌ها در دسترس می‌باشد، مورد بررسی قرار می‌دهیم. معادله خط روند به شرح ذیل بدست می‌آید:

$$y_t = 5226 - 119.897t \quad (16)$$

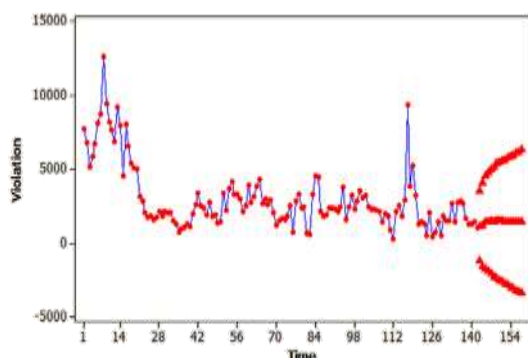
منفی بودن شیب خط نشان دهنده کاهش روند تخلفات در طول بازه زمانی انتخابی است. هرچه شیب کوچکتر (منفی) باشد نشان از کاهش بیشتر تخلفات است. نتیجه گیری فوق با در نظر داشتن شرایط ثابت حجم تردها و تعداد دوربین‌های فعال و همچنین عوامل دیگر در طول بازه زمانی، بررسی گردید. مقایسه بین شیب خط روند در هر دو معادله نشان می‌دهد که با فرض شرایط یکسان در ماه‌های مشابه سال‌های متوالی، سیر نزول تخلفات در سال ۹۳ بیشتر است که می‌تواند حاکی از موثر بودن دوربین‌های کنترل سرعت باشد.

همچنین در بررسی دیگر به پیش‌بینی متغیر تخلفات سرعت در بزرگراه تهران کرج پرداختیم. در جدول ۷ میزان خطای پیش‌بینی هر مدل آمده است. همان‌طور که در تحلیل استنباطی بیان گردید، بهترین مدل رگرسیون، مدل توانی غیرخطی بود. با توجه به این امر در جدول ۷، سری‌زمانی دارای خطای پیش‌بینی کمتری نسبت به مدل رگرسیونی است. بنابراین مدل سری زمانی مدل مناسبی جهت تحلیل داده‌های تخلفات است.

جدول ۷. خطاهای پیش‌بینی تخلفات در مدل سری‌زمانی و

رگرسیون		مدل
MAPE (%)	MAD (%)	
۰/۳۴	۵/۳۳	سری‌زمانی
۰/۶۸	۷/۸۸	رگرسیون غیرخطی

با توجه به پیش‌بینی که بر روی تخلفات بزرگراه تهران کرج در مدل سری زمانی از خردادماه سال ۹۳ تا پایان شهریور ماه همان سال صورت پذیرفت، میزان تخلفات در بازه تعیین شده نسبت به ماه‌های نظیرش در سال ۹۲، ۳۲ درصد کاهش داشته است. نتایج فوق همگی دلالت بر سودمندی دوربین‌های ثبت تخلف سرعت در بزرگراه تهران کرج دارند. در شکل ۱۰ نمودار پیش‌بینی تخلفات به همراه حدبالا و پایین پیش‌بینی نشان داده شده است.



شکل ۱۰. پیش‌بینی تخلفات در بزرگراه تهران کرج (زمان بر حسب شماره هفته)

در این مقاله به بررسی سودمندی ثبت تخلفات سرعت غیرمجاز در بزرگراه تهران کرج پرداختیم. با توجه به تحلیل‌های تفصیلی انجام‌شده و نتایج اخذ شده سودمندی دوربین‌ها محرز می‌گردد. پیشنهاد می‌گردد در کارهای آتی از شبکه‌های عصبی مصنوعی به همراه سری‌های زمانی و رگرسیون، برای پیش‌بینی میزان تخلفات بزرگراه‌های شهری به همراه میزان حجم تردد آنها و عوامل کیفی دیگر که می‌توانند به نحوی به صورت کیفی برآورد شوند استفاده گردد. همچنین سودمندی دوربین‌ها در دیگر

بزرگراه‌ها به صورت جداگانه و هم‌چنین به صورت مرتبط تحلیل گردد.

۶. پی‌نوشتها

1. Intelligent Transportation Systems
2. Autoregressive Integrated Moving Average
3. Autocorrelation function
4. Partial Autocorrelation function
5. Auto Regressive
6. Moving Average
7. Auto Regressive-Moving Average
8. Seasonal ARIMA
9. Akaike information criterion
10. Schwartz-Bayesian criteria
11. Box-jenkinz
12. Tolerance
13. Variance Inflation Factor

۷. مراجع

-آرمان نیا، ع. (۱۳۸۷) "نقش دوربین‌های مداربسته کنترل سرعت در کاهش تصادفات راه‌های برون شهری، نمونه موردی محور اندیمشک-اهواز-بندر امام"، پایان نامه کارشناسی ارشد، دانشکده مهندسی گروه عمران، دانشگاه شهید چمران اهواز.

-پورحسن، ف.، قاسمی، ب.، و سمثانی، ح.ر. (۱۳۹۳) "اثرات نصب سامانه‌های کنترل سرعت بر کاهش تصادفات بزرگراه‌های شهر تهران"، سومین کنفرانس ملی تصادفات جاده‌ای و سوانح ریلی و هوایی، دانشگاه آزاد اسلامی واحد زنجان، زنجان ص ۱۰-۱.

-جاویدی، ر.، و شریفی، م.ب. (۱۳۹۰) "استفاده از مدل‌های اتفاقی در شبیه‌سازی جریان رودخانه و پیش‌بینی دبی متوسط سالانه رودخانه توسط تحلیل سری‌های زمانی"، کنفرانس بین‌المللی منابع آب، شاهرود، ص. ۷-۵.

-جعفری، الف.، و صمدیان، م. (۱۳۹۱) "کاربرد داده کاوی در بررسی رفتار رانندگان متخلف در کلان شهرها"، مطالعات راهور-فصلنامه علمی-ترویجی، سال نهم، شماره ۱۷، ص. ۱۰۹-۱۳۲.

-حاجیلو، م. (۱۳۸۴) "نقش جرمه در کاهش تخلفات رانندگی درون شهری"، نخستین کنفرانس بین‌المللی حوادث رانندگی. تهران: دانشگاه تهران.

-حسن زاده، ج.، نجفی، ف.، و مرادنظر، م. (۱۳۹۴) "چگونه یک مدل مناسب برای داده‌های سری‌زمانی ایجاد کنیم؟" مجله تخصصی اپیدمیولوژی ایران، سال اول، شماره ۱۱، ص ۱.

-خرمی، م.، و بزرگ نیا، الف. (۱۳۸۶) "تجزیه و تحلیل سری‌های زمانی با نرم افزار MINITAB 14"، سال اول، شماره اول، انتشارات سخن گستر، مشهد، ص. ۴۵، ۱۵۰، ۴۴، ۱۵۲، ۱۴۵.

-رحیم اف، ک.، و صادقی کلینی، م. (۱۳۸۸) "تعیین عوامل موثر در شدت تصادفات آزادراهی و طراحی مدل آنها (مطالعه موردی: آزادراه تهران - کرج)"، پژوهشنامه حمل و نقل، سال هشتم، شماره ۲، ص. ۱۹۹-۲۰۹.

-رحیمی، م.، احدی، م. ر. و عباسی، و. (۱۳۹۳) "مطالعه تحلیلی میزان نقش بازدارندگی دوربین‌های ثبت تخلف سرعت در جاده‌ها در وقوع تصادفات (مطالعات موردی: آزاد راه زنجان- تبریز)"، نخستین همایش سیستم‌های حمل و نقل هوشمند جاده‌ای، سازمان راهداری و حمل و نقل جاده‌ای، تهران، ص ۱.

-عابدی، ح. و محسنی، م. (۱۳۹۳) "بررسی تاثیر دوربین‌های ثبت تخلفات رانندگی در مدیریت سرعت شبکه راه‌های استان مرکزی"، سومین کنفرانس ملی تصادفات جاده‌ای و سوانح ریلی و هوایی، دانشگاه آزاد اسلامی واحد زنجان، زنجان، ص. ۱.

-کاظمی، م.، صفارزاده، م.، موقری، ه. و فلاح زواره، م. (۱۳۹۵) "روش پیش‌بینی هزینه‌های متوفیات تصادفات ترافیکی در ایران"، فصلنامه مهندسی حمل و نقل، انتشار آنلاین از تاریخ ۲۰ تیر ۱۳۹۵.

-کلانتری، ن.، خاورزاده، ر.، و قیاسی، الف. (۱۳۹۲) "پیش‌بینی تعداد مسافران قطار سبک شهری مشهد به روش سری‌های زمانی فصلی"، سیزدهمین کنفرانس بین‌المللی مهندسی حمل و نقل و ترافیک، معاونت و سازمان حمل و نقل و ترافیک، ص ۱-۲، تهران

average time series model”, In Intelligent Transportation Systems Conference, ITSC'06-IEEE, pp. 493-498.

-Hughes, B. P., Newstead, S., Anund, A., Shu, C. C. and Falkmer, T. (2015) “A review of models relevant to road safety”, *Accident Analysis & Prevention*, Vol. 74, pp. 250-270.

-Keall, M. D., Povey, L. J. and Frith, W. J. (2002) “Further results from a trial comparing a hidden speed camera programme with visible camera operation”, *Accident Analysis & Prevention*, Vol. 34, No. 6, pp. 773-777.

-Novoa, A. M., Perez, K. and Santamarina-Rubio, E. (2009) “Effectiveness of speed enforcement through fixed speed cameras: a time series study”, *Injury prevention*, Vol. 16, No. 1, pp. 12-16.

-Oh, J., Eun, S. J., and Kim, H. S. (2016) “Effects of deregulation of the driving licence application process on road safety in Korea: an interrupted time-series analysis”, *Journal of Epidemiology and Community Health*, Published Online First: 26 May 201.

-Quddus, M. A. (2016) “Non-Gaussian interrupted time series regression analysis for evaluating the effect of smart motorways on road traffic accidents”, In *Transportation Research Board, 95th Annual Meeting*, No. 16-0157.

-Reinert, G. (2010) “Time series”, *Hillary Term*, p. 12.

Sebego, M., Naumann, R. B., Rudd, R. A., Voetsch, K., Dellinger, A. M. and Ndlovu, C. (2014) “The impact of alcohol and road traffic policies on crash rates in Botswana, 2004–2011: a time-series analysis”, *Accident Analysis & Prevention*, Vol. 70, pp. 33-39.

-Shearer, C. (2000) “The CRISP-DM model: the new blueprint for data mining”. *Journal of data warehousing*, Vol. 5, No. 4, pp. 13-22.

-Vanlaar, W., Robertson, R. and Marcoux, K. (2014) “An evaluation of Winnipeg's photo enforcement safety program: Results of time series analyses and an intersection camera experiment”, *Accident Analysis & Prevention*, Vol. 62, pp. 238-240

-مدقالچی، ع. و عباسی، و. (۱۳۹۳) “بررسی اثر پذیری دوربین‌های کنترل سرعت ثابت در تغییر رفتار رانندگان در ارتکاب به تخلفات سرعت با استفاده از روش آماری سری‌های زمانی بر روی داده‌های دستگاه‌های ترددشمار آنلاین (مطالعه موردی آزادراه زنجان-تیریز)”， نخستین همایش سیستم‌های حمل و نقل هوشمند جاده‌ای، سازمان راهداری و حمل و نقل جاده‌ای، تهران، ص. ۹-۱.

-همت فر، م.، حسینی، ع.ا.، شاه‌ویسی، ف. و نجفی، ی. (۱۳۹۰) “روابط خطی و غیرخطی بین متغیرهای حسابداری و بازده سهام شرکت‌های صنعت خودرو و ساخت قطعات”， پژوهشنامه حسابداری مال و حسابرسی، سال سوم، شماره دوازدهم، ص ۱۵۰-۱۴۵.

-Antoniou, C., Papadimitriou, E. and Yannis, G. (2014) “Road safety forecasts in five European countries using structural time series models”. *Traffic Injury Prevention*, Vol. 15, No. 6, pp. 598-605.

-Bahadorimonfared, A., Soori, H., Mehrabi, Y., Delpisheh, A., Esmaili, A., Salehi M. and Bakhtiyari M. (2013) “Trends of fatal road traffic injuries in Iran (2004–2011)”, *PLoS ONE*, Vol 8, No. 5: e65198, pp. 1-5.

-Brockwell P. J. and Davis R. A. (2009) “Time series: theory and methods”, *Springer*, pp. 44-79

-Chandran, A., Perez-Nunes, R., Bachani, A. M., Hajar, M., Salinas-Rodriguez, A. and Hyder, A. (2014) "Early impact of a national multi-faceted road safety intervention program in Mexico: results of a time-series analysis", *PloS ONE*, Vol. 9, No. 1: e8748, pp. 1-3.

-Chatfield, C. (2000) “Time-series forecasting”, *Chapman & Hall/CRC*, pp 20

-Commandeur, J. J., Wesemann, P., Bijleveld, F., Chhoun, V. and Sann, S. (2017) “Setting road safety targets in Cambodia: A methodological demonstration using the latent risk time series model”, *Journal of Advanced Transportation*, Article ID 5798174, 9 pages

-Guin, A. (2006) “Travel time prediction using a seasonal autoregressive integrated moving

traffic injuries in Guangzhou, China: a time-series study”, Journal of Epidemiology, Vol. 26, No. 8, pp. 433-439.

-Zhao, A., Chen, R., Qi, Y., Chen, A., Chen, X., Liang, Z. and Ye, J. (2016) “Evaluating the impact of criminalizing drunk driving on road-

استفاده از الگوریتم‌های داده‌کاوی برای تحلیل سودمندی

بی بی هانیه دقیق نژاد، درجه کارشناسی در رشته مهندسی کامپیوتر را در سال ۱۳۸۶ از دانشگاه پیام نور نیشابور و درجه کارشناسی ارشد در رشته مهندسی کامپیوتر را در سال ۱۳۹۴ از دانشگاه صنعتی شاهرود اخذ نمود.



هدی مشایخی، درجه کارشناسی در رشته مهندسی کامپیوتر را در سال ۱۳۸۵ از دانشگاه صنعتی شریف و درجه کارشناسی ارشد در رشته مهندسی کامپیوتر در سال ۱۳۸۷ را از دانشگاه صنعتی شریف اخذ نمود. در سال ۱۳۹۲ موفق به کسب درجه دکتری در رشته مهندسی کامپیوتر از دانشگاه صنعتی شریف گردید. زمینه های پژوهشی مورد علاقه ایشان داده کاوی، پردازش توزیع شده، پردازش جریان داده، یادگیری ماشین و سیستم های توصیه گر بوده و در حال حاضر عضو هیات علمی با مرتبه استادیار در دانشگاه صنعتی شاهرود است.



زهرا نوراله، درجه کارشناسی در رشته مهندسی کامپیوتر (نرم افزار) را در سال ۱۳۷۸ از دانشگاه آزاد اسلامی واحد تهران شمال و درجه کارشناسی ارشد در رشته مهندسی صنایع - مدیریت سیستم و بهره وری را در سال ۱۳۹۲ از دانشگاه تربیت مدرس اخذ نمود. زمینه های پژوهشی مورد علاقه ایشان داده کاوی و مدیریت سیستم های اطلاعاتی است.

