

# کاهش متغیرهای ورودی در فرآیند مدل‌سازی تصادفات آزادراهها با استفاده از روش تجزیه و تحلیل مؤلفه‌های اصلی

عبدالرضا شیخ‌الاسلامی، استادیار، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران  
فاطمه باقری خلیلی (نویسنده مسئول)، دانش‌آموخته کارشناسی ارشد، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران  
عباس محمودآبادی، دانشجوی دکتری، دانشکده فنی و مهندسی، دانشگاه پیام نور، تهران، ایران

E-mail: f-bagheri@rmto.ir

دریافت: ۹۱/۰۶/۲۱ پذیرش: ۹۱/۱۱/۰۹

## چکیده

هدف اصلی در این مقاله استفاده از روش تجزیه و تحلیل مؤلفه‌های اصلی برای شناسایی متغیرهای تأثیرگذار در فرآیند مدل‌سازی تصادفات در آزادراههای برون‌شهری است. با توجه به توانایی مدل شبکه‌های عصبی در پیش‌بینی تصادفات رانندگی، مدل تعداد تصادفات آزادراههای برون‌شهری کشور با استفاده از مدل‌های شبکه‌های عصبی توسعه داده شده و متغیرهای مربوط به جریان ترافیک، سهمیه‌بندی بنزین و متغیرهای محیطی نیز به عنوان متغیرهای مستقل در نظر گرفته شده‌اند. در فرآیند تعیین متغیرهای مستقل جهت ورود به مدل روش تجزیه و تحلیل مؤلفه‌های اصلی و تجزیه و تحلیل عامل اصلی با بکارگیری معیارهای KMO و بارتلت، مورد استفاده قرار گرفته است.

ضمن بررسی کارایی مدل شبکه عصبی مصنوعی در برابر مدل رگرسیون لگاریتم طبیعی در مدل‌سازی تصادفات آزادراههای برون‌شهری، دقت مدل‌های ساخته شده در تصادفات آزادراهها شامل مدل‌های شبکه عصبی مصنوعی و رگرسیون لگاریتم طبیعی نیز قبل و بعد از حذف متغیرهای با اهمیت کمتر مورد بررسی قرار گرفته و نتایج نشان دهنده آن است که حذف متغیرهای کم اهمیت از طریق آنالیز مؤلفه‌های اصلی در فرآیند مدل‌سازی دقت مدل‌های ساخته شده را با تغییرات اساسی رو برو نکرده و ثابت شده است که میانگین حجم تردد روزانه و متوسط سرعت وسایل نقلیه بیشترین نقش را در تصادفات آزادراهها ایفا می‌کنند.

واژه‌های کلیدی: مدل پیش‌بینی تصادفات، تجزیه و تحلیل مؤلفه‌های اصلی، شبکه عصبی مصنوعی، رگرسیون لگاریتم طبیعی، آزادراه

## ۱. مقدمه

بکارگیری راهکارهای کاهش تصادفات و تلفات جاده‌ای مستلزم تجزیه و تحلیل تصادفات با استفاده از مدل‌های پیش‌بینی و تعیین میزان تأثیر پارامترهای گوناگون در وقوع آنها و تلاش در جهت بهبود وضعیت ایمنی ترافیک است. در ایران از مجموع ۷۸ هزار کیلومتر جاده تحت نظارت و مسئولیت وزارت راه و شهرسازی (بجز راه‌های روستایی)، ۱۹۵۷ کیلومتر را آزادراه‌ها تشکیل می‌دهند و به دلیل تمایل مردم به استفاده از راه‌های با سطح خدمت بالاتر، ایمنی آزادراه‌ها در اولویت‌های اولیه وزارت راه و شهرسازی قرار دارد [Statistical Yearbook of Transportation, 2010]. با توجه به اینکه آزادراه‌ها از شرایط هندسی بهتری نسبت به انواع دیگر راه‌ها (راه‌های اصلی، فرعی و بزرگراه‌ها) برخوردارند، مطالعات نشان می‌دهند بخش عمده تصادفات در آزادراه‌ها در نتیجه عوامل محیطی و شاخص‌های ترافیکی، عوامل انسانی نظیر خستگی و خواب‌آلودگی، سرعت غیرمجاز و اشتباهات رانندگی رخ می‌دهد [Baek et. Al., 2005] و پژوهشگران سعی دارند با استفاده از مدل‌های آماری، وضعیت حوادث رانندگی در آزادراه‌ها را مورد بررسی قرار دهند. روش‌های آماری متعددی برای توسعه مدل‌های پیش‌بینی تصادفات بکار گرفته می‌شوند. به عنوان مثال شبکه عصبی مصنوعی روش نوینی در پیش‌بینی تعداد تصادفات است که در سال ۱۹۹۵ برای توسعه مدل رفتار رانندگان، نگهداری روسازی راه و تعیین موقعیت مکانی وسایل نقلیه استفاده شده است [Dougherty, 1995]. عبدال‌تی و همکاران از یک شبکه عصبی احتمالی<sup>۱</sup> برای پیش‌بینی تصادفات در کریدور بین‌شهری اورلاندو استفاده و نشان دادند که حداقل ۷۰ درصد از تصادفات می‌توانند با مدل شبکه عصبی احتمالی به طور صحیح پیش‌بینی شوند [Abdel-Aty and Pande, 2005]. چانگ [Chang, 2005] نیز در مدل‌سازی تصادفات آزادراه‌ها در کشور تایوان از دو مدل شبکه‌های عصبی مصنوعی و دو جمله‌ای منفی استفاده کرد و با مقایسه کارایی دو مدل، به این نتیجه رسید که مدل شبکه‌های عصبی مصنوعی یک روش جایگزین مناسب برای تجزیه و تحلیل تصادفات آزادراه‌ها

است. انکووار و دوگان نیز برای پیش‌بینی تعداد تصادفات و تلفات جاده‌ای در راه‌های ترکیه از مدل‌های شبکه عصبی و مدل رگرسیونی غیرخطی استفاده کرده و به این نتیجه رسیده‌اند که مدل شبکه عصبی نتایج دقیق تری نسبت به مدل رگرسیون غیرخطی داشته است [Akgünger and Dogan, 2008]. باپاتا و همکاران نیز مدل‌سازی تصادفات ماهانه را در راه‌های این کشور انجام داده و دریافتند که با وجود پیچیدگی ویژگی تصادفات، مدل شبکه عصبی توانایی قابل قبولی در پیش‌بینی تصادفات دارد [Bayata, Hattatoglu and Karsli, 2011].

استفاده از شبکه‌های عصبی در مطالعات انجام شده در ایران نیز در ادبیات موضوع دیده می‌شود. عبدالمنافی [Abdolm - nafi and Afandizadeh, 2007] برای پیش‌بینی تعداد تصادفات در تقاطعات درون‌شهری تهران و نیز محمودآبادی [Mahmoudabadi, 2010] در برآورد تعداد تصادفات جاده‌ای در آزادراه کرج - قزوین و بررسی عوامل تأثیرگذار و همچنین برآورد روزانه تعداد تصادفات در شبکه راه‌های ایران [Mahmoudabadi and Safi, 2008] از شبکه‌های عصبی استفاده کرده و خروجی‌های مدل شبکه عصبی را با مدل‌های آماری دیگری مورد مقایسه قرار داده‌اند که در هر سه تحقیق مدل شبکه عصبی جواب‌های دقیق تری نسبت به مدل آماری داشته‌اند.

شناخت عوامل مؤثر بر تعداد تصادفات و شدت آنها در پژوهش‌های سال‌های گذشته با استفاده از مدل‌سازی تصادفات به‌ویژه در آزادراه‌ها و بزرگراه‌ها نیز به چشم می‌خورد. نیومن تأثیر عرض میانه راه‌های چهار خطه را روی نرخ تصادف با استفاده از توزیع دو جمله‌ای منفی بررسی کرد. علاوه بر تأثیر افزایش عرض میانه بر کاهش تصادفات میانه‌های عریض‌تر، نرخ تصادفات جلو به جلو و خروج از مسیر را بین خودروهای دو جهت به شدت کاهش می‌دهند و در نتیجه افزایش عرض تأثیر زیادی روی تصادفات شدید نسبت به تصادفات صرفاً با خسارات مالی داشت [Knuiman, Council and Reinfurt, 1993]. اوه و چانگ نیز رابطه بین نرخ تصادفات و حجم به ظرفیت (V/C) را برای تاسیسات و تسهیلات مختلف موجود در آزادراه‌ها به

به نام تجزیه و تحلیل مولفه‌های اصلی<sup>۳</sup> برای حذف متغیرهای کم اهمیت در وقوع تصادفات و انتخاب متغیرهای مهم تر برای ورود به مدل‌سازی تصادفات، استفاده کرد [Caliendo and Parisi, 2005] که تحقیق حاضر نیز بر مبنای این رویکرد استوار است. در این پژوهش سعی می‌شود تأثیر کاهش عوامل موثر بر تصادفات آزادراههای برون‌شهری در دقت مدل‌های پیش‌بینی حوادث رانندگی، با استفاده از روشهای آماری شناسایی شود. با انتخاب آزادراه تهران- قم به عنوان مطالعه موردی، متغیرهای مورد استفاده در این پژوهش شامل میانگین تردد روزانه وسایل نقلیه، میانگین سرعت وسایل نقلیه، درصد وسایل نقلیه سنگین (غیرسواری)، سهمیه‌بندی بنزین، ماه وقوع تصادف و وضعیت آب و هوایی در زمان وقوع تصادف (به صورت فصول گرم و سرد) به عنوان متغیرهای مستقل و تعداد تصادفات در ماه به عنوان متغیر وابسته و خروجی مدل تعریف شده‌اند. در این پژوهش تمرکز اصلی بر معرفی روشی جهت کاهش متغیرهای ورودی به مدل، با استفاده از تجزیه و تحلیل مولفه‌های اصلی است که از مزایای آن می‌توان به کاهش زمان لازم برای جمع‌آوری اطلاعات و همچنین مدل‌سازی در شبکه عصبی مصنوعی اشاره کرد [Lee, 2006]. منظور از وسایل نقلیه سنگین در این پژوهش، وسایل نقلیه شامل اتوبوس و مینی‌بوس (غیرسواری) است. با در نظر گرفتن حجم تردد وسایل نقلیه سنگین، نقش حجم ترافیک در رخداد تصادفات دقیق تر بررسی می‌شود. برای گردآوری آمار تصادفات، آزادراه تهران- قم به بخشهایی با مشخصات فنی یکسان تقسیم شده و در هر بخش تعداد تصادفات به عنوان متغیر وابسته و سایر متغیرهای مستقل گردآوری شده‌اند. سپس با استفاده از تجزیه و تحلیل مولفه‌های اصلی، متغیرهای با اهمیت برای ورود به مدل‌سازی مشخص شده‌اند. پیش از انجام روش تجزیه و تحلیل مولفه‌های اصلی، امکان استفاده از این روش با محاسبه مقادیر فاکتور<sup>۴</sup> KMO و آزمون بارتلت، بررسی شده است. در تعیین متغیرهای کم اهمیت نیز دو مدل شبکه عصبی مصنوعی و مدل رگرسیون لگاریتم طبیعی بر داده‌های تصادفات آزادراهها برازش

دست آوردند. تحقیقات آنها نشان داد که رابطه بین V/C و نرخ تصادفات در تمام مقاطع از یک تابع u شکل پیروی می‌کند [Oh et.al, 2001]. کانگ و همکاران [Kang and Lee, 2002] نیز به بررسی وضعیت ترافیکی آزادراهها، قبل از وقوع تصادف پرداخته و مانند مطالعات قبلی به این نتیجه رسیده‌اند که تغییرات سرعت و چگالی ترافیک عبوری در وقوع تصادفات مؤثرند. نتایج همچنین نشان‌دهنده آن است که برای بررسی دقیق تر اثر وضعیت ترافیکی لحظه‌ای بر وقوع تصادفات باید عوامل مربوط به طرح هندسی، وضعیت آب و هوایی و زمان وقوع تصادف نیز در مدل نظر گرفته شود. تحقیقات لردا و همکاران [Lorda, Manarab and Viziolib, 2005] نیز نشان می‌دهد مطالعاتی که در آن تنها متغیر حجم ترافیک به عنوان متغیر توصیفی وارد مدل می‌شود، نمی‌توانند به اندازه کافی روند وقوع حوادث ترافیک را در آزادراههایی که در محیطهای شهری و روستایی واقع شده‌اند توصیف کنند. عبدل آتی و همکاران در تحقیقاتی به این نتیجه رسیدند که متوسط سطح اشغال، انحراف معیار حجم، تغییرات در سرعت و شاخص میزان بارندگی در وقوع تصادفات آزادراهها موثر است [Abdel-Aty and Pemmanaboina, 2006]. در مقایسه با تحقیقاتی که اثر مشخصات هندسی راه و وضعیت ترافیکی را بر تصادفات نشان می‌دهند، تحقیقات مربوط به بررسی عوامل انسانی بر تصادفات آزادراهها اندک است. بانک و همکاران در سال ۲۰۰۵ مدلی را برای ارتباط بین طول زمان سفر و تصادفات واقع شده در آزادراهها ارائه و پتانسیل نرخ تصادفات<sup>۲</sup> را، با در نظر گرفتن توزیع فراوانی نرخ سفر در زمینه تجزیه و تحلیل تصادفات بررسی کردند و نتیجه موید این مطلب بود که بین پتانسیل نرخ تصادفات، حجم ترافیکی و تصادفات به وقوع پیوسته ارتباط تنگاتنگی وجود داشته است [Baek et.al., 2005]. در سالهای اخیر مدل‌های پیش‌بینی تصادفات جاده‌ای با استفاده از تکنیکهای مختلف تجزیه و تحلیل توسعه داده شده‌اند، ولی بکارگیری این مدلها زمانی که تعداد زیادی متغیر مد نظر است، به آسانی انجام نمی‌شود. به منظور رفع این مشکل می‌توان از ابزاری

نظر گرفته و دومین مؤلفه حداکثر واریانس در نظر گرفته نشده توسط اولین مؤلفه را پیش بینی کرده و این روند ادامه می‌یابد تا آخرین مؤلفه تمامی واریانس مورد نظر را در بر گیرد.

$$w_{i1}^2 + w_{i2}^2 + \dots + w_{ip}^2 = 1 \quad i = 1, \dots, \quad (2)$$

$$w_{i1}w_{j1} + w_{i2}w_{j2} + \dots + w_{ip}w_{jp} = 0 \quad \text{for all } i \neq j \quad (3)$$

برای انجام روش تجزیه و تحلیل عوامل اصلی مراحل زیر انجام می‌گیرد.

الف: استاندارد کردن متغیرهای ورودی: در این مرحله داده‌های ورودی به نحوی استاندارد می‌شوند که دارای میانگین صفر و انحراف معیار یک باشند. ماتریس Z که ماتریسی شامل مقادیر استاندارد شده پارامترهای است از رابطه (4) به دست می‌آید [Caliendo and Parisi, 2005].

$$Z_{ij} = \frac{(x_{ij} - \bar{x}_j)}{s_j} \quad (4)$$

For = 1, 2, ..., n and j = 1, 2, ..., q

که در این رابطه  $\bar{x}_j$  میانگین داده‌ها و از جنس  $x_j$  و  $s_j$  مقدار انحراف معیار مربوط هستند. داده‌ها از جنس  $x_j$  هستند.

ب: محاسبه فاکتور KMO: شاخص KMO مورد استفاده در دامنه صفر تا یک قرار دارد. این شاخص از رابطه (5) به دست می‌آید که در این رابطه  $r_{ij}$  ضریب همبستگی بین متغیرهای i و j، و  $a_{ij}$  ضریب همبستگی جزئی بین آنهاست [Hutcheson and Nick, 1999].

$$KMO = \frac{\sum \sum r_{ij}^2}{\sum \sum r_{ij}^2 + \sum \sum a_{ij}^2} \quad (5)$$

ج: محاسبه ماتریس همبستگی (کواریانس) برای متغیرهای اولیه: این ماتریس میزان همبستگی بین هر کدام از متغیرهای اولیه مورد استفاده را نشان می‌دهد. مقدار هر کدام از درایه این ماتریس  $a_{ij}$  بیانگر همبستگی بین متغیرهای i و j است که از رابطه شماره (6) حاصل می‌شود [Caliendo and Parisi, 2005].

$$R = \frac{1}{n} Z'Z \quad (6)$$

د: محاسبه مقادیر ویژه  $\lambda$  و بردارهای ویژه مربوطه از ماتریس همبستگی [Caliendo and Parisi, 2005]: با حل رابطه (7) و (8) مقادیر ویژه و بردارهای ویژه معادل هر مقدار ویژه محاسبه

داده شده است و سپس مؤلفه‌های برازش مدلها شامل آماره‌های توان دوم ضریب همبستگی  $R^2$  و میزان میانگین مربعات خطا (MSE) قبل و بعد از حذف متغیرهای کم اهمیت مورد بررسی قرار گرفته است. پس از حذف متغیرهای کم اهمیت کارآیی مدل شبکه عصبی مصنوعی در برابر مدل رگرسیون لگاریتم طبیعی در مدل‌سازی تصادفات آزادراهها مورد بررسی قرار گرفته و در نهایت به اعتبارسنجی مدل بهتر برای پیش‌بینی تصادفات آزادراههای برون‌شهری پرداخته شده است تا کارآیی آن در مدل‌سازی تصادفات آزادراههای برون‌شهری ثابت شود.

## ۲. تجزیه و تحلیل مؤلفه‌های اصلی

تجزیه و تحلیل مؤلفه‌های اصلی (PCA) نوعی از تجزیه و تحلیل آماری است که تعداد کمتری از عوامل را به نام مؤلفه‌های اصلی از میان عوامل اولیه گزینش می‌کند، به طوری که تعدادی از اطلاعات کم اهمیت حذف می‌شود. در صورتی که فاکتور تست KMO مربوط به این روش کمتر از 0/5 باشد، داده‌ها برای تجزیه و تحلیل عوامل اصلی مناسب نخواهند بود و اگر مقدار آن بین 0/5 تا 0/69 باشد باید با احتیاط بیشتر به تجزیه و تحلیل عوامل پرداخت. اما در صورتی که مقدار آن بزرگ تر از 0/7 باشد همبستگیهای موجود در بین داده‌ها برای تجزیه و تحلیل مناسب خواهد بود [Hutcheson and Nick, 1999]. با استفاده از این روش، ترکیباتی از p متغیر اولیه  $X_1, X_2, \dots, X_p$  برای ایجاد حداکثر p مولفه مستقل به صورت  $PC_1, PC_2, \dots, PC_p$  ایجاد می‌شود [Johnson and Wichern, 1982]. هر مولفه اصلی می‌تواند با دنباله ارائه شده در رابطه 1 مشخص شود:

$$PC_1 = w_{11}x_1 + w_{12}x_2 + \dots + w_{1p}x_p$$

$$PC_2 = w_{21}x_1 + w_{22}x_2 + \dots + w_{2p}x_p \quad (1)$$

...  
 $PC_p = w_{p1}x_1 + w_{p2}x_2 + \dots + w_{pp}x_p$   
 که در آن  $PC_1$  معرف مولفه مورد نظر،  $w_{ij}$  ضریب مربوط به متغیرهای اولیه و  $x_i$  نیز متغیر اولیه است. ضرایب  $w_{ij}$  طوری تخمین زده می‌شوند که اولین مؤلفه حداکثر واریانس داده‌ها را در

$$Y_i = \exp(\beta_0) \exp(\beta_1 x_{i1}) \exp(\beta_2 x_{i2}) \dots \exp(\beta_q x_{iq}) \quad (10)$$

در این مدل فرض می‌شود که لگاریتم تعداد تصادفات از توزیع نرمال با میانگین  $\mu_i$  و واریانس  $\sigma^2$  پیروی می‌کند. ضرایب  $\beta_0, \beta_1, \beta_2, \dots, \beta_q$  ضرایب رگرسیون خطی هستند که به وسیله روش کمترین مربعات خطا محاسبه می‌شوند. این مدل یک مدل کلاسیک از رابطه خطی چندگانه بین لگاریتم متغیر وابسته و  $q$  متغیر پیش‌بینی کننده مستقل است [Abdolmanafi and Afandizadeh, 2007].

### ۳-۲ مدل شبکه عصبی

شبکه‌های عصبی قادرند رابطه بین ورودیها و خروجیهای یک دستگاه فیزیکی را با شبکه‌ای از گره‌ها که همگی با هم متصلند، تعیین کنند [McCulloch and Pitts, 1943]. شکل ۱ شبکه‌های عصبی را با لایه‌های ورودی  $a_i$ ، وزنه‌های  $w_i$ ، لایه پنهان و یک خروجی نمایش داده است. یادگیری این مدلها که در واقع تعیین پارامترهای داخلی آنهاست، بر پایه قانون تصحیح خطا بنا شده است که تعمیم روش مشهور حداقل میانگین مربعات است و با استفاده از فرمول شماره (۱۱) محاسبه می‌شود که در آن:  $N$  تعداد عصبهای لایه خروجی که متناسب با تعداد مشاهدات تابع هدف،  $t_i$ : مقدار مشاهده شده برای  $i$  امین رکورد،  $td_i$ : مقدار خروجی شبکه برای  $i$  امین رکورد است [Yadav and Sharma, 2010].

$$MSE = \frac{1}{N} \sum_{n=1}^N (t_i - td_i)^2 \quad (11)$$

تقسیم داده‌های ورودی مدل به قسمتهای مختلف نیز به عنوان داده‌های آموزش و اعتبارسنجی در حین ساخت شبکه امری متداول است [Cansiz and Easa, 2011] و [Burden, B -] [ereton and Walsh, 1997]. با بکار بردن این روش که روش توقف آموزش<sup>۹</sup>، امکان استفاده از معماریهای پیچیده‌تر در طراحی شبکه برای کاربر فراهم می‌شود، بدون اینکه مشکل فوق‌براشی<sup>۱۰</sup> روی دهد و با قرار دادن معیارها آموزش متوقف می‌شود. به این ترتیب معماری مورد اشاره نقش مهمی در روش آموزش ایفا می‌کند [Coulibaly, Anctil and Bobee, 2000]. در این

می‌شوند. بردارهای ویژه به دست آمده، به ازای هر مقدار ویژه نیز به عنوان ضرایب متغیرهای اولیه در تشکیل مولفه مربوطه هستند. از حل رابطه (۷) که در آن  $I$  ماتریس واحد است، مقادیر ویژه  $\lambda_{ii}$  محاسبه می‌شوند.

$$\det(R - \lambda I) = 0 \quad (7)$$

مقدار واریانس هر کدام از مولفه‌های اصلی با استفاده از رابطه (۸) محاسبه شود.

$$\det(R - \lambda I) = V_h \quad (8)$$

ه: تعیین معیار استخراج تعداد عاملها

معیار مقدار ویژه<sup>۱</sup> و معیار درصد واریانس [Kalantari, 2009] و معیار تست بریدگی [Cattel, 1996] مهم ترین پارامترهای استخراج تعداد عوامل هستند که در این فرآیند بکار گرفته می‌شوند.

و: اجرای چرخش مناسب بر روی ماتریس ضرایب مؤلفه‌ها: این مرحله به تجزیه و تحلیل فاکتورهای اصلی<sup>۷</sup> نیز مشهور است. در این مرحله متغیرهایی که ضرایب بالایی در مولفه‌های اصلی استخراج شده دارند به عنوان متغیرهای مهم جهت ورود به مدل‌سازی انتخاب می‌شوند [Singha et al., 2004].

### ۳. مدل‌های پیش‌بینی تصادفات و معیارهای ارزیابی آنها

#### ۳-۱. مدل رگرسیون لگاریتم طبیعی<sup>۸</sup>

مدل رگرسیون لگاریتم طبیعی بر این فرض استوار است که لگاریتم طبیعی  $Y_i$  از یک توزیع نرمال با میانگین  $\mu_i$  و واریانس  $\sigma^2$  پیروی می‌کند. مدل رگرسیون لگاریتم طبیعی زمانی مورد استفاده قرار می‌گیرد که داده‌ها غیرمنفی و توزیع داده‌ها دارای انحراف مثبت و میانگین داده‌ها نسبتاً بزرگ باشد [Abdolmanafi and Afandizadeh, 2007]. در این مدل، رابطه بین تعداد تصادفات مورد انتظار در قطعه  $i$  و  $q$  متغیر پیش‌بینی کننده، به صورت رابطه (۹) و شکل نمایی نوشته به صورت رابطه (۱۰) می‌شود:

$$\ln(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_q x_{iq} \quad (9)$$

تحقیق از معیار میانگین مربعات خطا نیز برای مقایسه کارایی مدل‌های شبکه عصبی و رگرسیون لگاریتم طبیعی در پیش‌بینی تعداد تصادفات استفاده شده است.

#### ۴. محدوده مطالعه و گردآوری اطلاعات

به منظور بررسی آمار تصادفات و پارامترهای تأثیر گذار بر آن در آزاد راه‌های برون‌شهری، آزادراه تهران-قم به طول ۱۳۶ کیلومتر حد فاصل استان تهران و استان قم انتخاب و از بانک اطلاعاتی موجود در سازمان راهداری و حمل‌ونقل جاده‌ای مربوط به دوره زمانی مربوط به تاریخ ۸۴/۱۰/۰۱ تا ۸۷/۰۷/۳۰ استفاده شده است. در بخش جمع‌آوری اطلاعات، داده‌های مربوط به میانگین حجم تردد روزانه (ADT)، میانگین سرعت وسایل نقلیه و درصد تردد وسایل نقلیه غیرسواری در واحدهای زمانی ماهانه، از طریق دستگاه‌های تردد شمار موجود در آزادراه برای مسیر رفت و برگشت آزادراه به طور جداگانه جمع‌آوری شده است. با توجه به کامل بودن آمار و اطلاعات مربوط به سال ۸۵ و ۸۶ مقادیر مربوط به تصادفات آزادراه‌ها در مسیر تهران-قم در این سالها به ترتیب تعداد ۱۳۷۸ و ۱۱۱۰ است و در مسیر برگشت و (آزادراه قم-تهران) نیز به ترتیب ۱۴۱۵ و ۱۰۱۰ است. تعداد کل تصادفات این آزادراه در ماه‌های مختلف سال به عنوان متغیر وابسته و متغیرهای مستقل مربوط به هر ماه، شامل میانگین حجم

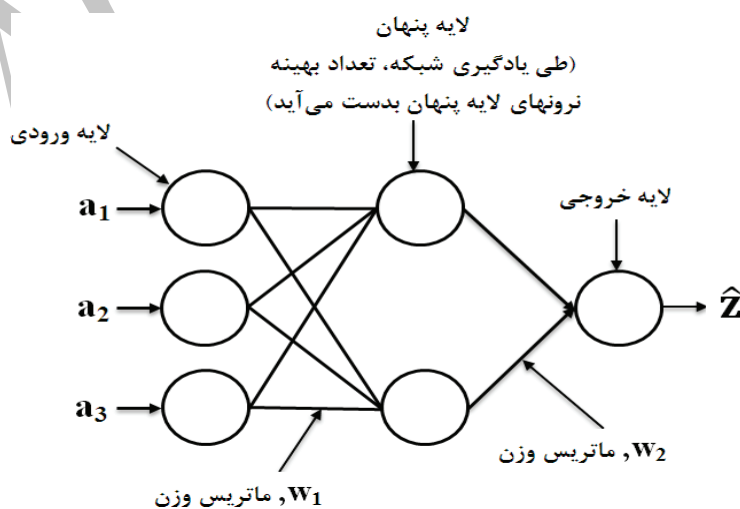
مقاله از الگوریتم آموزشی trainlm و روش توقف آموزش برای آموزش شبکه عصبی در پیش‌بینی تعداد تصادفات، استفاده شده است که در ادبیات موضوع توصیه شده است [Yadav and Sharma, 2010].

#### ۳-۳ معیارهای ارزیابی

یکی از معیارهای مورد بررسی اعتبار نتایج به دست آمده از مدل رگرسیونی و شبکه عصبی، معیار برازندگی ضریب همبستگی (R) و یا توان دوم آن (R<sup>2</sup>) است که مقدار آن به ترتیب برای R بین -۱ و +۱ و برای R<sup>2</sup> بین ۰ و +۱ تغییر می‌کند و از فرمول شماره (۱۲) مقدار R<sup>2</sup> محاسبه می‌شود [Yadav and Sharma, 2010].

$$R^2 = \frac{\sum_{m=1}^M (y_m - \bar{y})^2 - \sum_{m=1}^M (y_m - \hat{y}_m)^2}{\sum_{m=1}^M (y_m - \bar{y})^2} \quad (12)$$

که در آن  $y_m$ : متغیر وابسته مشاهده شده،  $\hat{y}_m$ : متغیر وابسته متناسب به متغیر مستقل  $x_m$ ،  $\bar{y}$ : متوسط داده‌ها که با فرمول  $y_m = \sum_{m=1}^M \frac{y_m}{M}$  محاسبه می‌شود که در آن نیز  $x_m$ : متغیر مستقل است. مقادیر مطلق نزدیک به ۱ تطابق بهتر داده‌های مشاهده‌ای و برآورد شده را نشان می‌دهند. ولی با توجه به این مساله که مقدار R<sup>2</sup> تحت تاثیر داده‌های خارج از محدوده است، باید از آن به اتفاق پارامترهای دیگر استفاده کرد. به همین دلیل در این



شکل ۱. نمونه از شبکه عصبی پیشرو [Yadav and Sharma, 2010]

### 5. تعیین عوامل تاثیرگذار

متغیرهای اولیه در آزادراه تهران - قم شامل تعداد تصادفات ماهیانه، میانگین تردد روزانه، درصد وسایل نقلیه سنگین، متوسط سرعت وسایل نقلیه، فصل وقوع تصادف و وضعیت سهمیه بندی بنزین به مقادیر استاندارد تبدیل شده‌اند. مقدار KMO برابر کل داده‌ها ۰/۴۷۸ (کمتر از ۰/۵) به دست آمده است. بنابراین نمی‌توان تجزیه و تحلیل مؤلفه‌های اصلی را روی کل داده‌ها اجرا کرد و نتایج حاصل از تحلیل مؤلفه‌های اصلی معتبر نیست. با حذف متغیر ماه وقوع تصادف از داده‌های ورودی، مقدار KMO عدد ۰/۵۲ به دست می‌آید که این مقدار وجود همبستگی لازم بین متغیرهای ورودی برای انجام آنالیز مؤلفه‌های اصلی را تایید می‌کند. در ادامه با استفاده از فرمول شماره (۶) ماتریس کواریانس متغیرها محاسبه شده و مقادیر و بردارهای ویژه مربوطه با استفاده از روابط (۷) و (۸) محاسبه می‌شوند. مقادیر ویژه  $(\frac{\lambda_h}{\sum \lambda_h})$  داده‌های مورد استفاده، درصد اطلاعات متغیرهای اولیه موجود در هر مولفه اصلی و درصد تجمعی اطلاعات مربوط به متغیرهای اولیه  $\sum (\frac{\lambda_h}{\sum \lambda_h} * 100)$  در جدول شماره (۳) ارائه شده است.

$$PC_1 = 0.61ADT + 0.512LV + 0.685AS + 0.798P + 0.396Y \quad (14)$$

جدول (۴) بردارهای ویژه را برای تشکیل هر مولفه نشان می‌دهد. برای تشکیل مولفه اول باید مقادیر پارامترهای اندازه‌گیری شده میانگین حجم تردد را در عدد ۰/۶۱، پارامترهای اندازه‌گیری شده درصد تردد غیرسواری را در عدد ۰/۵۱۲، پارامترهای

تردد روزانه، میانگین سرعت وسایل نقلیه و درصد تردد وسایل نقلیه غیر سواری، پارامتر فصل (به صورت فصول گرم شامل بهار و تابستان کد صفر، و فصول سرد شامل پاییز و زمستان کد یک) و همچنین تاثیر سهمیه‌بندی بنزین (به صورت کد صفر به ماههای قبل از اجرای طرح و کد ۱ به ماههای بعد از اجرای طرح) در نظر گرفته شده‌اند. متغیرها و نشانه‌های اختصاص داده شده به آنها برای مدل‌سازی در جدول شماره ۱ ارائه شده است.

از آن جا که در بعضی موارد تمامی پارامترهای مورد نظر در مدل‌های تصمیم‌گیری از ابعاد یکسانی برخوردار نیستند، برای یکسان‌سازی آنها از روشهای نرمال‌سازی یا استانداردسازی داده‌ها استفاده می‌شود [Mahmoudabadi and Seyedhosseini, 2012]. روشهای متعددی برای نرمال‌سازی داده‌های ورودی و خروجی مدل‌های تصمیم‌گیری وجود دارد که متداول‌ترین آنها داده‌های جمع‌آوری شده را به اعداد بین ۰ تا ۱ یا بین -۱ تا ۱ تبدیل می‌کنند [Sharma, 1996]. به منظور استانداردسازی داده‌ها با فرض نرمال بودن تابع توزیع آنها نیز از رابطه شماره (۱۳) استفاده می‌شود.

$$x_{new} = \frac{x_{old} - \bar{x}}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}} \quad (13)$$

که در آن  $X_{old}$  مقدار اولیه پارامتر،  $X_{new}$  مقدار نرمال شده یا استاندارد شده پارامتر و پارامترهای ورودی و خروجی با استفاده از فرمول (۱۳) در دامنه نرمال قرار گرفته است. آمارهای مربوط به مقادیر واقعی و نرمال شده داده‌ها در جدول (۲) آورده شده است.

جدول ۱. معرفی متغیرهای ورودی و نشانه‌های اختصاص داده شده به آنها

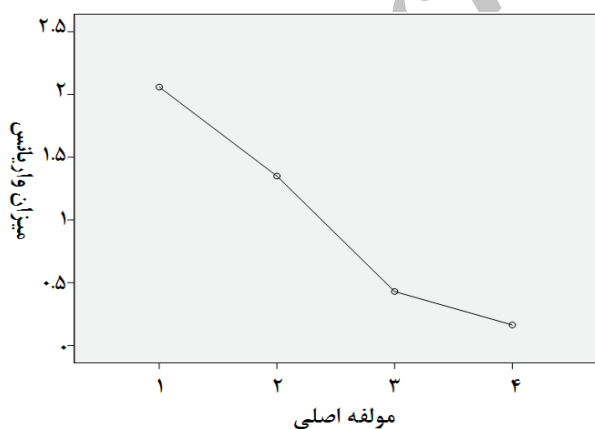
ردیف	نام متغیر	نشانه
۱	تعداد تصادفات در ماه	Y
۲	فصل وقوع تصادف	SEA
۳	متوسط حجم ترافیک روزانه در ماه	ADT
۴	درصد تردد وسایل نقلیه سنگین در ماه	L.V
۵	متوسط سرعت وسایل نقلیه عبوری در ماه	A.S
۶	وضعیت سهمیه بندی بنزین	P

جدول ۲. آمارهای مربوط به مقادیر واقعی و نرمال پارامترهای ورودی و خروجی

ردیف	نام پارامتر	مقادیر واقعی			مقادیر استاندارد شده			
		حداکثر	حداقل	میانگین	انحراف معیار	حداکثر	حداقل	میانگین
۱	میانگین حجم تردد روزانه در ماه	۵۰۴۷۵	۲۱۷۰۹	۳۵۱۰۲/۳۲	۷۰۰۵/۶۳	۱	۰/۴۷	۰/۲۴
۲	میانگین سرعت وسایل نقلیه	۱۱۶/۳	۷۳/۶	۹۴/۰۶	۱۳/۲۴	۱	۰/۵	۰/۲۹۶
۳	نسبت تردد غیرسواری	۰/۰۷۳	۰/۰۱۹	۰/۰۳۵	۰/۰۱۰۶	۱	۰/۳	۰/۱۹۶
۴	فصل از سال	۱	۰	۰/۳۸	۰/۴۹	۱	۰/۳۸	۰/۴۹
۵	وضعیت سهمیه بندی بنزین	۱	۰	۰/۶۸	۰/۴۷	۱	۰/۲۸	۰/۱۹
۶	تعداد تصادفات در ماه	۲۲۹	۵۰	۱۰۲/۰۱	۳۵/۶۶	۱	۰/۶۸	۰/۴۷

جدول ۳. مقادیر ویژه مربوط به ماتریس Z

ردیف	مقادیر ویژه $(\lambda h)$	مولفه اصلی $100 \left( \frac{\lambda h}{\sum \lambda h} \right)$	اولیه $100 \left( \frac{\lambda h}{\sum \lambda h} \right)$
۱	$\lambda_1$	۳۷/۹۴۲	۳۷/۹۴۲
۲	$\lambda_2$	۲۵/۵۳۲	۶۳/۴۳۷
۳	$\lambda_3$	۲۰/۰۳۴	۸۳/۵۰۸
۴	$\lambda_4$	۸/۹۶۶	۹۲/۴۷۴
۵	$\lambda_5$	۷/۵۲۶	۱۰۰



شکل ۲. نمودار آزمون بریدگی در روش تجزیه و تحلیل مولفه‌های اصلی برای تعیین متغیرهای اصلی در وقوع تصادفات آزادراهها از روش تجزیه و تحلیل عاملهای اصلی استفاده شده است. در این روش متغیرهای اصلی متغیرهایی هستند که حداقل یکی از ضرایب آنها

اندازه‌گیری شده متوسط سرعت وسایل نقلیه در عدد ۰/۶۸۵، پارامتر سهمیه‌بندی بنزین را در عدد ۰/۷۹۸ و تعداد تصادفات در ۰/۳۹۶ ضرب شده که بر این اساس اولین مولفه اصلی با فرمول شماره (۱۴) مشخص می‌شود. همان گونه که از رابطه (۱۴) مشخص می‌شود در تشکیل مولفه اول پارامترهای میانگین حجم تردد روزانه و متوسط سرعت وسایل نقلیه بالاترین ضرایب را دارند که نشان‌دهنده تاثیر بیشتر این پارامترها در تشکیل مولفه اول است. با توجه به اینکه چهار مولفه اول بیش از ۹۲ درصد از واریانس داده‌ها را تبیین می‌کند، در صورت استفاده از معیار درصد واریانس، چهار مولفه اول انتخاب می‌شوند. همچنین در شکل شماره (۲) که مربوط به نمودار تست بریدگی است، چهار مولفه اول به عنوان مولفه‌های اصلی انتخاب می‌شوند.



کاهش متغیرهای ورودی در فرآیند مدل‌سازی تصادفات آزادراهها با استفاده از ...

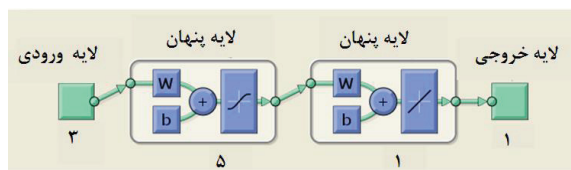
جدول ۴. مقادیر مربوط به بردارهای ویژه برای تشکیل هر مولفه

مولفه اول (PC <sub>1</sub> )	مولفه دوم (PC <sub>2</sub> )	مولفه سوم (PC <sub>3</sub> )	مولفه چهارم (PC <sub>4</sub> )	مولفه پنجم (PC <sub>5</sub> )	
۰/۶۱	۰/۳۲۲	-۰/۶۰۵	۰/۳۸۵	۰/۰۹۹	ADT
۰/۵۱۲	۰/۷۳۴	-۰/۰۵۴	-۰/۴۴۲	-۰/۰۰۷	L.V
۰/۶۸۵	-۰/۳۸۵	۰/۰۶	-۰/۱۶۵	۰/۳۹۸	A.S
۰/۷۹۸	-۰/۳۹۷	۰/۰۵۳	-۰/۰۰۷	-۰/۴۵	P
۰/۳۹۶	۰/۳۶۶	۰/۷۹۱	۰/۲۷۸	۰/۰۷۵	Y

جدول ۵. نتایج حاصل از ماتریس دوران یافته (مقادیر بردارهای ویژه در روش تجزیه و تحلیل عاملهای اصلی)

عامل اول (LF <sub>1</sub> )	عامل دوم (LF <sub>2</sub> )	عامل سوم (LF <sub>3</sub> )	عامل سوم (LF <sub>4</sub> )	
۰/۰۷۲	-۰/۰۲۷	۰/۹۶۶	۰/۲۱۲	ADT
-۰/۰۲۳	۰/۱۶۴	-۰/۲۱۴	۰/۹۶۲	L.V
۰/۹۴۹	۰/۰۳۷	۰/۰۷۲	-۰/۰۲۵	A.S
۰/۳۲۵	۰/۰۹	۰/۱۳۷	۰/۰۴۷	P
۰/۰۳۶	۰/۹۸۴	-۰/۰۱۴	۰/۱۵۳	Y

حداقل کردن میانگین مربعات خطا محاسبه شود. در مرحله اول در توسعه مدل شبکه عصبی، داده‌ها به سه دسته آموزش (۷۰٪)، اعتبارسنجی (۱۵٪) و آزمایش (۱۵٪) تقسیم شده‌اند. با توجه به این که با تغییر تعداد عصبهای لایه پنهان کارایی شبکه تغییر می‌کند حالت‌های مختلف شبکه با داشتن تعداد عصب متفاوت در لایه پنهان شبکه عصبی مورد اجرا قرار گرفته است و در نهایت شبکه با معماری به صورت شکل شماره (۳) به عنوان شبکه نهایی انتخاب شده است. سیستم در هر تکرار بر اساس یک سری از داده‌ها آموزش دیده و اعتبارسنجی شده و با باقیمانده داده‌ها با توجه به مقایسه مقادیر پارامترهای شبکه شامل MSE و R<sup>2</sup> مورد آزمون قرار می‌گیرد و مقادیر MSE و R<sup>2</sup> برای کل داده‌ها از مدل برداشت شده تا با مدل رگرسیونی مورد مقایسه قرار گیرد.



شکل ۳. وضعیت نهایی شبکه عصبی

که برای تشکیل عامل مربوطه استفاده می‌شود، دارای مقدار نسبتاً بالایی باشد. نتایج محاسبات در جدول شماره (۵) نشان داده شده و با توجه به پیچیدگی تصادفات این معیار معادل با ۰/۸ انتخاب شده است [Singha et.al, 2004]. با توجه به معیار در نظر گرفته شده و جدول شماره (۵) مشخص می‌شود که تمام متغیرها به جز متغیر سهمیه‌بندی بنزین ضریب بالای ۰/۸ را در چهار عامل استخراج شده دارند. بنابراین متغیر وضعیت سهمیه‌بندی بنزین با استفاده از آنالیز عاملهای اصلی، کم اهمیت ترین متغیر تأثیر گذار در تصادفات آزادراهها است. علت اصلی کم اهمیت تشخیص داده شدن این متغیر از طریق آنالیز فاکتورهای اصلی، همبستگی زیاد این متغیر با میانگین حجم تردد روزانه است.

## ۶. توسعه مدل‌های پیش‌بینی تصادفات

### ۶-۱ مدل شبکه عصبی مصنوعی

با توجه به این که سه متغیر ورودی و یک متغیر خروجی مد نظر است، مدل مورد نظر دارای سه عصب در لایه ورودی و یک عصب در لایه خروجی خواهد بود که تعداد عصب در لایه پنهان باید با

## ۶-۲ مدل رگرسیون لگاریتم طبیعی

برای توسعه مدل رگرسیون، لگاریتم آزمون آماری کولموگراف-اسمیرنو (KES) برای آزمایش تبعیت داده‌ها مورد استفاده قرار می‌گیرد. بر اساس تست کولموگراف-اسمیرنو، داده‌های مربوط به تصادفات با اطمینان بالای ۹۵ درصد از توزیع لگاریتم طبیعی پیروی می‌کنند. با توجه به داده‌های موجود، مقادیر  $\beta$  در مدل رگرسیون لگاریتم طبیعی برازش داده‌ها، برای هر یک از متغیرها به شرح جدول شماره (۶) است. مقدار معناداری ضرایب هر یک از متغیرها توسط آماره  $t$  مقایسه و نتیجه نهایی محاسبات به صورت رابطه (۱۵) و مقادیر جدول (۶) نشان داده شده است. آنچه که از رفتار مدلها در جدول شماره (۶) قابل استنتاج است، سرعت و حجم تردد از مهم ترین عوامل در بروز تصادفات آزادراهها به شمار می‌روند. این نتایج در تحقیقات گذشته نیز تصدیق شده است [Vogt and Bared, 1998]. همچنین این موضوع در مطالعات استفان هس و جان پلک در سال ۲۰۰۳ [Hess and Polak, 2003] و اولمستد [Olmstead, 2001] در سال ۲۰۰۱ نیز اشاره شده است که کنترل سرعت وسایل نقلیه باعث کاهش ۱۸ درصدی در میانگین تصادفات ماهیانه شده و ارتقای نظارت و کنترل حدود ۱۳/۲۶ درصد تصادفات جرحی را کاهش داده است و وجود تجهیزات مدیریت ترافیک آزادراهها در ایالت آریزونا باعث کاهش ۲۱ تا ۳۰ درصدی تصادفات شده است. در گزارش اداره حمل‌ونقل مینه سوتا نیز تصادفات آزادراههای درون‌شهری با گسترش دادن مدیریت آزادراهها حدود ۳۰ درصد، یعنی ۹۰۰ تصادف در سال، کاهش یافته [Lave, 1985]. مطالعه‌ای که در

مورد تجهیزات مدیریت ترافیک آزادراهها صورت گرفته است نشان می‌دهد که استفاده از تکنولوژی پیشرفته اعلام محدودیت‌های سرعت متغیر، سیگنال‌های کنترل خطوط و ردیابی اتوماتیک صف می‌تواند در کاهش کلی نرخ تصادفات بسیار مؤثر باشد [Inn - vative Traffic Control, 1999].

$$LN(Y) = 0/123 + 0/398 ADT + 0/484A.S - 0/28 L.V \quad (15)$$

$$Y = e^{0/123} \times e^{0/398ADT} \times e^{0/484A.S} \times e^{-0/28L.V}$$

## ۷. توصیف و تحلیل نتایج

### ۷-۱ ارزیابی و مقایسه خروجی مدلها

مدلهای ساخته شده شبکه عصبی و رگرسیون لگاریتم طبیعی بر اساس مقادیر میانگین مربعات خطا و توان دوم ضریب همبستگی مدل با استفاده از متغیرهای کاهش یافته مورد مقایسه قرار گرفته و نتایج در جدول (۷) نشان داده شده است. همان گونه که در این جدول مشاهده می‌شود میانگین مربعات خطا در مدل شبکه عصبی نسبت به مدل رگرسیونی کمتر بوده و مقدار  $R^2$  در آن از مدل رگرسیونی بیشتر است و این مطلب نشان‌دهنده آن است که مدل شبکه عصبی توانایی بیشتری در پیش‌بینی تعداد تصادفات آزادراهها نسبت به مدل رگرسیون طبیعی دارد.

### ۷-۲ بررسی تاثیر کاهش متغیرها

با توجه به توضیحات اشاره شده در قسمتهای قبل مربوط به حذف دو عامل متغیر فصل وقوع تصادف و سهمیه‌بندی بنزین، سنجش میزان کارایی روش تجزیه و تحلیل مولفه‌های اصلی در حذف

جدول ۶. نتایج حاصل از مدل رگرسیون لگاریتم طبیعی

مدل	ضرایب استاندارد نشده		ضرایب استاندارد شده		مقدار آماره $t$	مقدار Sig
	تخمین	خطای استاندارد	تخمین	تخمین		
مقدار ثابت	۰/۱۲۳	۰/۰۶۶	۰	۰	۱/۸۸۴	۰/۶۸
ADT	۰/۳۹۸	۰/۱۵۲	۰/۴۱۵	۰/۴۱۵	۲/۶۱۱۷	۰/۱۳
A.S	۰/۴۸۴	۰/۰۹	۰/۶۱۳	۰/۶۱۳	۵/۳۶۳	۰/۰۰۰
L.V	-۰/۲۸	۰/۱۵۴	-۰/۲۸۲	-۰/۲۸۲	-۱/۸۱۹	۰/۰۷۷

کاهش متغیرهای ورودی در فرآیند مدل‌سازی تصادفات آزادراهها با استفاده از ...

جدول ۷. مقایسه مدل رگرسیون لگاریتم طبیعی و مدل شبکه عصبی (برای داده‌های نورمال)

ردیف	نوع مدل	میانگین مربعات خطا (MSE)	ضریب همبستگی ( $R^2$ )
۱	شبکه عصبی	۰/۰۴۹	۰/۸۱
۲	رگرسیون لگاریتم طبیعی	۰/۲۹۴	۰/۵۶۵

یک ساله مربوط به سال ۸۶ در آزادراه قم- کاشان که در ادامه آزادراه تهران- قم قرار دارد، استفاده شده است. برای انجام اعتبارسنجی مدل با داده‌های تصادف سال ۸۶ آزادراه قم- کاشان، متغیرهای مستقل مدل در دوره زمانی یاد شده در ماههای مختلف گردآوری شده است. برای پیش‌بینی تصادفات از آماره میانگین مربعات خطا استفاده می‌شود که در این حالت مقدار آن عدد ۰/۰۵۶۷ به دست می‌آید که نشان دهنده کم بودن اختلاف میان میزان واقعی تصادفات دیده شده و میزان محاسبه شدن آن از طریق مدل است که نشان دهنده معتبر بودن مدل شبکه‌های عصبی توسعه داده شده در پیش‌بینی تصادفات آزادراهها است.

#### ۸. خلاصه و نتیجه‌گیری

در این مقاله اطلاعات تصادفات آزادراه تهران- قم در دوره زمانی ۳۴ ماهه از دی ماه ۸۴ تا مهر ماه ۸۷ برای مدل‌سازی استفاده شده است و اطلاعات مربوط به وضعیت ترافیکی مسیر شامل میانگین تردد روزانه، درصد تردد وسایل نقلیه سنگین (غیرسواری) و میانگین سرعت وسایل نقلیه از نهادهای مربوطه جمع‌آوری شده و پس از آن داده‌های مورد نظر بین ۰ و ۱ به صورت نورمال محاسبه شده است. با استفاده از روش تجزیه و تحلیل مولفه‌های

متغیرهای کم اهمیت در فرآیند مدل‌سازی، دو مدل شبکه‌های عصبی و مدل رگرسیون لگاریتم طبیعی بر داده‌های ۳۴ ماهه موجود از آزادراه تهران- قم در دو حالت متفاوت قبل از کاهش متغیرها و بعد از کاهش متغیرها صورت گرفته و میزان دقت مدل‌های ساخته شده در این دو حالت با استفاده از آماره‌های توان دوم ضریب همبستگی  $R^2$  و میزان میانگین مربعات خطا (MSE) مقایسه شده است. نتایج حاصل از حذف و عدم حذف متغیرهای یاد شده در مدل‌سازی تصادفات آزادراهها در جدول (۸) ارائه شده است. نتایج حاصل از جدول نشان دهنده این مطلب است که تغییری که طی فرآیند اجرای روش تجزیه و تحلیل مولفه‌های اصلی حذف شده است تأثیری قابل توجه در تعداد تصادفات آزادراهها نداشته و حذف متغیر یاد شده، تغییری در دقت مدل‌های ساخته شده ایجاد نمی‌کند. همچنین کاهش متغیرهای ورودی در مدل شبکه‌های عصبی مصنوعی باعث کاهش تعداد تکرارهای آموزش شبکه برای رسیدن به شبکه بهینه می‌شود و زمان رسیدن به شبکه عصبی بهینه کاهش می‌یابد.

#### ۷-۳ اعتبارسنجی مدل با داده‌های جدید

برای اعتبارسنجی مدل شبکه‌های عصبی به عنوان مدل برتر انتخاب شده در مدل‌سازی تصادفات آزادراهها، از اطلاعات

جدول ۸. نتایج حاصل از مدل‌سازی بعد و قبل از کاهش متغیرها

ردیف	نوع مدل	میانگین مربعات خطا (MSE)		ضریب همبستگی ( $R^2$ )	
		قبل از کاهش متغیرها	بعد از کاهش متغیرها	قبل از کاهش متغیرها	بعد از کاهش متغیرها
۱	شبکه عصبی	۰/۰۴۹	۰/۰۵۸	۰/۸۲۵	۰/۸۱
۲	رگرسیون لگاریتم طبیعی	۰/۲۶۹	۰/۲۹۴	۰/۵۷۳	۰/۵۶۵

## ۹. سپاسگزاری

از ریاست محترم و همکاران سازمان راهداری و حمل و نقل جاده‌ای که در انجام این تحقیق نهایت همکاری را با نویسندگان این مقاله داشته‌اند سپاسگزاری می‌شود

## ۱۰. پی‌نوشتها

1. PNN (Probabilistic Neural Network)
2. Potantial accident rate
3. Principal Component Analysis
4. Kaiser-Meyera-Olkin
5. Mean Square Error
6. Eigenvalue criterion
7. Principal Factor Analysis
8. Lognormal regression models
9. Stop training algorithm
10. Overfitting

## ۱۱. مراجع

- سالنامه آماری حمل و نقل جاده‌ای، دفتر فن‌آوری اطلاعات، سازمان راهداری و حمل و نقل جاده‌ای، ۱۳۸۹.

- عبدالمنافی، سیدابراهیم، احمدی‌نژاد، محمود و افندی‌زاده، شهریار (۱۳۸۶) "طراحی مدل پیش‌بینی تعداد تصادفات در تقاطعات درون‌شهری بر اساس مدل‌های آماری و شبکه عصبی"، پایان‌نامه کارشناسی ارشد، تهران: دانشگاه علم و صنعت ایران.

کلاتری، خلیل "پردازش و تحلیل داده‌ها در تحقیقات اجتماعی-اقتصادی"، (۱۳۸۸)، نشر فرهنگ صبا، چاپ سوم، ص ۳۰۳-۲۹۵

- محمودآبادی، عباس و صفی صمغ‌آبادی، اعظم‌دخت (۱۳۸۷) "برآورد روزانه تصادفات جاده‌ای با استفاده از شبکه‌های عصبی با تکیه بر وضعیت تردد"، دومین کنفرانس سیستم‌های فازی و هوشمند، تهران: دانشگاه صنعتی مالک اشتر، ۷ تا ۹ آبان ۸۷

-Abdel-Aty, M. and Pande, A. (2005) "Identification crash propensity using traffic speed condition", Journal of Safety Research, Vol 36, No. 1, pp. 97-108.

اصولی متغیرهای مهم تر برای ورود به مدل‌سازی از میان متغیرهای اولیه انتخاب شده‌اند. نتایج تحلیل مولفه‌های اصلی نشان‌دهنده آن است که متغیر فصل وقوع تصادف و وضعیت سهمیه‌بندی بنزین از متغیرهای کم اهمیت برای ورود به مدل‌سازی هستند. برای بررسی کارایی روش تجزیه و تحلیل مولفه‌های اصلی در کاهش متغیرهای کم اهمیت جهت ورود به مدل‌سازی، مدل‌های شبکه عصبی مصنوعی و رگرسیون لگاریتم طبیعی در دو حالت مختلف (قبل از کاهش متغیرهای کم اهمیت و بعد از کاهش متغیرها) بر داده‌های موجود برازش داده شده و نتایج حاصل از مقایسه آماره‌های نیکویی برازش مدل‌ها نشانگر این مطلب است که وجود متغیرهای کم اهمیت در مدل‌سازی تأثیر چندانی بر افزایش دقت مدل‌سازی نسبت به حالتی که این متغیرها در مدل‌سازی موجود هستند، ندارد.

برای بررسی میزان کارایی مدل شبکه عصبی خروجی‌های مدل توسعه داده شده با مدل رگرسیون لگاریتم طبیعی مورد ارزیابی قرار گرفته است. مقایسه بین دو مدل نشان داد که مدل شبکه عصبی کارایی بهتری نسبت به مدل رگرسیون لگاریتم طبیعی دارد. سپس مدل شبکه عصبی با داده‌های جدید یک ساله مربوط به آزادراه قم-کاشان مورد ارزیابی قرار گرفته است که نتایج حاصله نشان از معتبر بودن مدل ساخته شده در تصادفات آزادراهها دارد. همچنین مدل رگرسیون لگاریتم طبیعی برازش داده شده نشان‌دهنده این مطلب است که میانگین تردد روزانه و میانگین سرعت وسایل نقلیه تأثیر مستقیم در وقوع تعداد تصادفات در آزادراهها داشته‌اند ولی درصد تردد غیر سواری تأثیری معکوس در تصادفات آزادراهها داشته است.

به محققانی که علاقه‌مند به مطالعه در این زمینه هستند توصیه می‌شود که موضوع را در دیگر راههای کشور که دارای ویژگیهای مهندسی متفاوتی هستند مورد مطالعه قرار دهند و همچنین با بررسی دقیق تر موضوع در تصادفات آزادراهها عواملی نظیر عوامل انسانی موثر در تصادفات را با دقت بیشتری شناسایی کنند تا در جهت رفع آنها گامهای اساسی برداشته شود.

8, pp. 541-557.

-Coulibaly, P., Anctil, F., and Bobee, B. (2000) "Daily reservoir inflow forecasting using artificial neural networks with stopped training approach", *J. Hydrol.*, Vol. 230, PP. 244-257.

-Dougherty, M. (1995) "A review of neural networks applied to transport", *Transportation Research Part C* 3 (4), PP.247-260.

-Hess, S. and Polak, J. (2003) "An analysis of the effects of speed limit enforcement cameras on accident rates", submitted for presentation at the 2003 TRB Annual Meeting.

-Hutcheson, G. and Nick, S. (1999) "The multivariate social scientist: Introductory statistics using generalized linear models", Thousand Oaks, CA: Sage Publications.

-Innovative Traffic Control (1999) "Technology and practice in Europe", Federal Highway Administration, U.S. Department of Transportation, August.

-Johnson, R. A. and Wichern, D. W. (1982) "Applied multivariate statistical analysis, 3rd Ed", Prentice-Hall Inc., Englewood Cliffs, SA, 590.

-Kang, J. G. and Lee, S. K. (2002) "Traffic accident prediction model by freeway geometric types", *Journal of Korean Society of Transportation*, Vol. 20, No. 4, pp.163-175.

-Knuiman, M.W., Council, F. M and Reinfurt, D. W. (1993) "Association of median width and highway accident rates". *Transp. Res. Rec.*, 1401.

-Lave, C. (1985) "Speed, coordination, and 55 mph limit", *The American economic review*. December, pp.1159-1164.

-Lee, Y. (2006) "A computerized feature reduction using principal component analysis for accident duration forecasting on freeway", A research project sponsored by the National Science Council, Taiwan under the contract of NSC99-2628-E-451-001, Website: <http://cerc.wvu.edu/download/.../DMI2065.pdf>

-Abdel-Aty, M. A and Pemmanaboina, R. (2006) "Calibrating a real-time traffic crash-prediction model using archived weather and ITS traffic data", *IEEE Transportations on Intelligent Transportation Systems*, Vol. 7, No. 2, pp. 167-174.

-Akgüngör, A. P and Dogan, E. (2008) "Estimating road accidents of Turkey based on regression analysis and artificial neural network approach", *Advances in Transportation Studies, an International Journal*, Vol. 4, No.9, pp. 906-913.

-Baek, S., Chang, H., Kang, J.G. and Yoon, B. (2005) "An effect of trip length on freeway accident rate", *Journal of the Eastern Asia Society Transportation Studies*, Vol. 6, pp. 3467-3481.

-Bayata, H. F., Hattatoglu, F and Karsli, N. (2011) "Modeling of monthly traffic accidents with the artificial neural network method", *International Journal of the Physical Sciences* Vol. 6, No.2, January, pp. 244-254. Website: <http://www.academicjournals.org/IJPS>.

-Burden, F. R., Brereton, R. G. and Walsh, P. T. (1997) "Cross-validatory selection of test and validation sets in multivariate calibration and neural networks as applied to spectroscopy", *Analyst*, Vol.122, No. 10, pp.1015-1022.

-Caliendo, C. and Parisi, A. (2005) "Principal component analysis applied to crash data on multilane roads", Third international SIIV Congress, Bari, Italy, 20-22 September, ANCONA SIIV Vol. 1, Page 1-7, Website: <http://www.sed.siiv.scelta.com>.

Cansiz, O. F. and Easa, S. M. (2011) "Using artificial neural network to predict collisions on horizontal tangents of 3rd two-lane highways", *International Journal of Engineering and Applied Sciences*, Vol. 7, No.1, pp. 47-56.

-Cattel, R. B. (1996) "The scree test for the number of the factor", *Multivariate Behavioral Research*, Vol. 1, April, pp 245-276.

-Chang, Li-Y. (2005) "Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network", *Safety Science*, Vol 43, No.

- Olmstead, T. (2001) "Freeway management systems and motor vehicle crashes: A case study of Phoenix, Arizona", *Accident Analysis and Prevention*, Vol. 33, pp. 433-447.
- Sharma, S. (1996) "Applied multivariate techniques", Published by university of South Carolina.
- Singha, K. P., Malika, A., Mohana, D and Sinhab, S. (2004) "Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) a case study", *Water Research*, Vol. 38, pp. 3980-3992.
- Vogt, A, and Bared, J. (1998) "Accident models for two-lane rural segments and intersections", In *Transportation Research Record*, 1635, TRB, National Research Council, Washington, D.C., pp. 18-2.
- Yadav, D. and Veena Sharma, N. (2010) "Artificial neural network based hydro electric generation modelling", *International Journal of Applied Engineering Research*, Dindigul, Vol. 1, No. 3, pp.343-359.
- Mahmoudabadi, A. "Comparison of weighted and simple linear regression and artificial neural network models in freeway accidents prediction (Case study: Qom & Qazvin Freeways in Iran)", *Second International Conference on Computer and Network Technology*, Thailand, Bangkok, 23-25 April 2010, Part 7: *Traffic and Logistic Management*, pp. 392-396
- Mahmoudabadi, A. and Seyedhosseini, S. M. (2012) "Time-risk tradeoff of hazmat routing problem in emergency situation", *Proceedings of the Third International Conference on Industrial Engineering and Operations Management*, Istanbul, Turkey, July 3 – 6, 2012, pp. 344-351.
- McCulloch, W. S., and Pitts, W. (1943) "A logical calculus of the ideas imminent in nervous activity", *B. Math. Biophys.*, Vol. 8, pp. 115-133.
- Oh, C., Oh, J. S., Ritchie, S. G. and Chang, M. S. (2001) "Real-time estimation of freeway accident likelihood", *Annual Meeting CD-Rom 80th Annual Meeting of The Transportation Research Board*.

Archive of SID