

## استفاده از ابزار برنامه‌نویسی پایتون در توسعه مدل‌های داده‌کاوی

### شدت تصادفات برون‌شهری استان اصفهان

#### مقاله علمی - پژوهشی

علی توکلی کاشانی\*، دانشیار، دانشکده مهندسی عمران، تهران، دانشگاه علم و صنعت ایران، تهران، ایران  
مهدی هادی‌نیا، دانشجوی کارشناسی ارشد، دانشکده مهندسی عمران، تهران، دانشگاه علم و صنعت ایران، تهران، ایران  
محمدعلی صلواتی، دانشجوی دکتری، دانشکده مهندسی عمران، دانشکده فنی دانشگاه تهران، تهران، ایران

\*پست الکترونیکی نویسنده مسئول: Alitavakoli@iust.ac.ir

دریافت: ۱۴۰۳/۰۳/۰۵ - پذیرش: ۱۴۰۳/۰۹/۲۰

صفحه ۱۱۴-۱۰۷

#### چکیده

مطابق آمارهای پزشکی قانونی ایران، آمار مرگ‌ومیر و مصدومین بر اثر حوادث ترافیکی از سال ۱۳۹۶ تا ۱۴۰۱ افزایشی بوده است. این روند نه تنها در کل کشور بلکه در استان‌ها نیز قابل مشاهده است. روند افزایشی مرگ‌ومیر و مصدومین حوادث ترافیکی در طی سالیان اخیر می‌تواند زنگ خطری برای مسئولین برای تحلیل عوامل انسانی، وسیله نقلیه‌ای، راه و محیطی بر تصادفات و یافتن راهکارهایی در جهت کاهش آمار مرگ‌ومیر و مصدومین حوادث ترافیکی استان‌ها باشد. از این جهت در این پژوهش به بررسی عوامل موثر بر شدت تصادفات برون‌شهری استان اصفهان پرداخته شده است. در این مطالعه برای یافتن عوامل موثر بر شدت تصادفات، از مدل رگرسیون لجستیک استفاده شده که با ۲۱ هزار داده مدل نهایی ساخته شده و بر روی نتایج آن تحلیل‌های آماری صورت گرفته است. نتایج این تحلیل‌ها نشان می‌دهد متغیرهایی نظیر ماه وقوع تصادف، ساعت تصادف، نوع راه، نوع برخورد، جنسیت و علت نامه تصادف معنادار شناسایی شدند. نتایج نشان می‌دهند احتمال وقوع یک تصادف شدید در ساعات اولیه روز از باقی ساعات روز بیشتر است؛ همچنین احتمال وقوع یک تصادف شدید در فروردین نسبت به بقیه ماه‌های سال از بقیه بیشتر است؛ و یا نتایج نشان دادند که نقض راه می‌تواند احتمال وقوع تصادف شدید را ۵٫۲۷ برابر بیشتر از باقی علت نامه تصادفات افزایش دهد (نسبت بخت=۵٫۲۷) و در نهایت احتمال وقوع یک تصادف شدید برای مردان، ۱٫۴۱ درصد بیشتر از زنان است. در انتها سعی شد در رابطه با تعدادی از این متغیرهای اشاره شده، راهکارهایی برای انجام مطالعات بیشتر پیشنهاد شود.

واژه‌های کلیدی: شدت تصادف، رگرسیون لجستیک، داده‌کاوی، تصادفات جاده‌ای

#### ۱- مقدمه

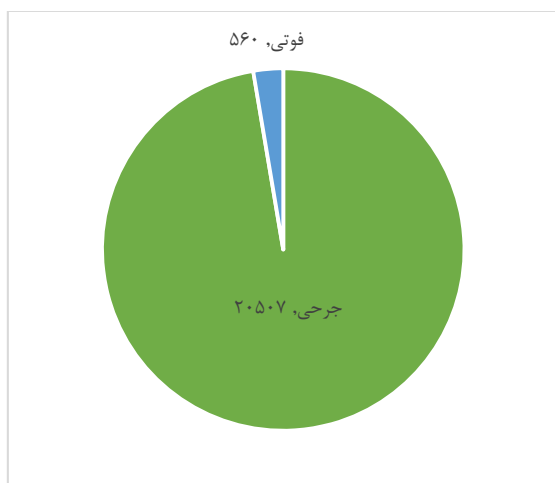
در ایران و استان اصفهان افزایشی بوده و باید چاره‌ای برای مهار آن اندیشیده شود (سازمان پزشکی قانونی، آمار متوفیات و مصدومین حوادث رانندگی). آئین‌نامه ایمنی راه‌های آمریکا (Highway Safety Manual)، سه عامل اصلی در تصادفات را عامل انسانی، وسیله نقلیه و راه و محیط اعلام می‌کند. عوامل انسانی شامل سن، جنسیت، قدرت تصمیم‌گیری، مهارت‌های رانندگی و خستگی؛ عوامل راه شامل سازنده، نوع وسیله نقلیه

طبق آمارهای سازمان بهداشت جهانی، برآورد می‌شود هر ساله نزدیک به ۱٫۳ میلیون نفر جان خود را در تصادفات رانندگی از دست می‌دهند؛ و همچنین بین ۲۰ تا ۵۰ میلیون نفر در جهان از آسیب‌های ناشی از تصادفات رنج می‌برند (World Health Organization, Road traffic injuries, 2022). همچنین مطابق آمار سازمان پزشکی قانونی ایران بین سال‌های ۱۳۹۶ تا ۱۴۰۱، آمار مرگ‌ومیر و مصدومین تصادفات جاده‌ای

شده است (Zafri et al., 2020). همچنین مطالعات نشان داده‌است که علت تصادف نیز یکی از پارامترهای تاثیر گذار در شدت تصادفات در راه‌های برون‌شهری ایران است (Tavakoli & Kashani & Besharati, 2016; شریعت‌مهیمنی و توکلی‌کاشانی، ۱۳۸۹). در این مطالعه با استفاده از مدل رگرسیون لجستیک به بررسی هر یک از عوامل تاثیر گذار بر شدت تصادفات، به بررسی رابطه هر یک از متغیرها به صورت تکی پرداخته و عوامل موثر بر شدت تصادف را شناسایی کرده و با مشخص کردن علت هر یک از عوامل تاثیر گذار بر شدت تصادف، سعی در پیشنهاد راهکاری برای کاهش شدت تصادفات می‌پردازیم.

## ۲- داده‌ها

در این مطالعه از حدود ۲۵ هزار داده‌ی مربوط به تصادفات برون‌شهری استان اصفهان طی سالهای ۱۳۹۶ تا ۱۴۰۰ که توسط پلیس‌راه استان اصفهان جمع‌آوری شده است، استفاده شده‌است. پس از پاکسازی داده‌ها و حذف رکوردهایی با مقادیر نامشخص و سایر، نزدیک به ۲۱ هزار داده باقی ماند که فراوانی آنها در شکل ۱ مشخص است. همچنین بر اساس جدول ۱، این پایگاه داده شامل متغیرهایی نظیر روز وقوع تصادف، ماه وقوع تصادف، ساعت وقوع تصادف، نوع راج، جنسیت مقصر، سن مقصر، نوع وسیله نقلیه مقصر، نوع برخورد و علت تامه تصادف می‌باشد.



شکل ۱. فراوانی تصادفات استان اصفهان در پایگاه داده،

از سال ۱۳۹۶ تا سال ۱۴۰۰

(از منظر سواری، سنگین و ...)، نگهداری و طراحی آن؛ عوامل راه و محیط شامل وضعیت آب‌وهوایی، نوع راه، وسایل کنترل ترافیک، سطح راه و علامت‌گذاری می‌شوند ( Highway Safety Manual, 2010). در مطالعات گذشته، مدل‌سازی شدت تصادف بر اساس مدل‌های آماری و یادگیری ماشین انجام شده است؛ که در برخی از این مطالعات، مدل‌های یادگیری ماشین دقت بیشتری نسبت به مدل‌های آماری ارائه داده‌اند (توکلی‌کاشانی و همکاران، ۱۴۰۱)؛ همچنین اکثر مطالعات نیز به بررسی سه عامل انسان، وسیله و راه و محیط بر شدت تصادفات پرداخته‌اند. مطالعات نشان داده‌اند رگرسیون لجستیک یک روش بهینه برای مدل‌سازی یک مدل با نتایج دودویی (باینری) است (Zafri et al., 2020). بدارد و همکاران نشان دادند که با افزایش سن، شدت تصادف رخ داده افزایش می‌یابد، آنها گزارش دادند که رانندگان بالای ۸۰ سال ۴,۹۸ برابر بیشتر از رانندگان ۴۰ تا ۴۹ سال آسیب‌مرگبار داشتند (Bédard et al., 2002). این نتایج با مطالعات وانگ و همکاران نیز تطابق دارند، آنها نشان دادند که رانندگان و دوچرخه‌سواران بالای ۵۵ سال احتمال آسیب را افزایش می‌دهد (Wang et al., 2015). همچنین بسیاری از مطالعات نشان داده‌اند که جنسیت عاملی اثرگذار بر روی شدت تصادفات است، بطوریکه مردان تصادفات شدید بیشتری را نسبت به زنان تجربه می‌کنند (Albalate & Fernández-Villadangos, 2010; Bédard et al., 2002; Benlagha & Charfeddine, 2020). همچنین برخی مطالعات نشان داده‌اند که نوع وسیله نقلیه نیز در شدت تصادف موثر است؛ آلبالاته و فرناندز نشان دادند که موتور سیکلت آسیب‌پذیری بیشتری در تصادفات در مقایسه با دیگر وسایل نقلیه دارد (Albalate & Fernández-Villadangos, 2010); و یا ظفری و همکاران طی مطالعاتشان به این نتیجه رسیدند که خودروهای سنگین تصادفات شدیدتری را تجربه می‌کنند (Zafri et al., 2020). زمان تصادف نیز عاملی موثر در شدت تصادف است؛ به عنوان مثال در مطالعه ظفری و همکاران، زمان تصادف یک متغیر معنادار شناسایی شده و احتمال رخداد یک تصادف شدید در شب، ۵,۳۳ برابر رخداد تصادف در هنگام روز است همچنین احتمال رخداد یک تصادف شدید در هنگام غروب و سپیده‌دم نیز ۱,۷۳۹ برابر احتمال رخداد تصادف در هنگام روز برآورد

جدول ۱. توصیف متغیرها و دسته‌بندی آنها

دسته‌بندی	متغیر
	۱: جراحی ۲: فوتی
	۱: شنبه ۲: یکشنبه ۳: دوشنبه ۴: سه‌شنبه ۵: چهارشنبه ۶: پنجشنبه ۷: جمعه
	۱: فروردین ۲: اردیبهشت ۳: خرداد ۴: تیر ۵: مرداد ۶: شهریور ۷: مهر ۸: آبان ۹: آذر ۱۰: دی ۱۱: بهمن ۱۲: اسفند
	۱: ۱۲ صبح تا ۶ صبح ۲: ۶ صبح تا ۱۲ ظهر ۳: ۱۲ ظهر تا ۶ عصر ۴: ۶ عصر تا ۱۲ شب
	۱: آزادراه ۲: بزرگراه ۳: اصلی ۴: فرعی ۵: روستایی
	۱: زن ۲: مرد
	۱: کمتر از ۱۸ سال ۲: ۱۸ تا ۳۵ سال ۳: ۳۵ تا ۶۰ سال ۴: ۶۰ سال به بالا
	۱: عابر ۲: موتورسیکلت و دوچرخه ۳: سواری ۴: سنگین
	۱: تصادفات گروه عابر ۲: تصادفات گروه وسایل نقلیه دوچرخ ۳: تصادفات سایر وسایل نقلیه ۴: واژگونی
	۱: انحراف به چپ ۲: انحراف به راست ۳: تجاوز از سرعت مجاز ۴: تجاوز به چپ ۵: تغییر مسیر ناگهانی ۶: حرکت با دنده عقب ۷: حرکت در خلاف جهت ۸: دور زدن در محل ممنوع ۹: عبور از محل ممنوع ۱۰: عدم توانایی در کنترل وسیله ۱۱: عدم توجه به جلو ۱۲: عدم رعایت حق تقدم ۱۳: عدم رعایت فاصله طولی و عرضی ۱۴: نقص راه ۱۵: نقص فنی وسیله نقلیه ۱۶: نقض ماده (۲۱۱ و ۲۱۲) رعایت مقررات ایمنی تعمیر و نگهداری راه ۱۷: توقف وسیله نقلیه در خطوط عبور ۱۸: نقض مقررات حمل بار ۱۹: گردش به طرز غلط ۲۰: سرعت نامطمئن و زیاد ۲۱: تقارن دو خطا

### ۳- روش تحقیق

شدت تصادفات تصادفات است. در مدل‌های شدت تصادف دو پارامتر حائز اهمیت هستند، یکی سطح معناداری و دیگری نسبت بخت<sup>۱</sup> که در این مطالعه به بررسی هر یک از این دو پارامتر پرداخته و بر اساس آن به سهم عواملی نظیر انسان، وسیله نقلیه و راه و محیط را مشخص خواهد شد. بنابراین، یک مدل رگرسیون لجستیک دودویی (باینری) برای شناسایی عوامل مؤثر بر شدت تصادف، برای توصیف رابطه بین شدت تصادف و عوامل شناسایی شده با استفاده از معادله زیر توسعه داده شد.

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) = \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i \quad (1)$$

که در رابطه (۱)،  $P$  احتمال وقوع تصادف فوتی،  $\alpha_0$  ضریب ثابت،  $\beta_i$  ضریب مدل و  $x_i$  نیز متغیر غیروابسته است (Zafri et al., 2020). همچنین برای ساخت مدل رگرسیون لجستیک در پایتون از کتابخانه statsmodels.formula.api استفاده شده است.

برای دستیابی به هدف تحقیق، با استفاده از زبان برنامه‌نویسی پایتون ابتدا به بررسی استقلال یا عدم استقلال شدت تصادف با متغیرهای مشخص شده در جدول ۱ پرداخته و سپس با ساخت مدل رگرسیون لجستیک، برای متغیرهایی که مستقل از شدت تصادف نباشند، رابطه هر یک از متغیرها با شدت تصادف را بررسی می‌کنیم. برای بررسی استقلال دو متغیر از یکدیگر، از آزمون استقلال دو متغیر طبقه‌بندی استفاده می‌کنیم که با استفاده از آزمون کای‌دو می‌توان همبستگی میان آنها را مورد بررسی قرار داد (توکلی کاشانی و بشارتی، ۱۳۹۹). در این آزمون فرض صفر استقلال دو متغیر از یکدیگر تعریف شده و در صورتی که مقدار p-value از ۰,۰۵ کمتر شود، یعنی این دو متغیر از یکدیگر مستقل نبوده و یا به عبارت دیگر با سطح اطمینان ۹۵ درصد دارای رابطه معنادار با یکدیگر هستند. برای انجام این آزمون در پایتون از کتابخانه scipy.stats استفاده شده است. همانطور که در مطالعات پیشین اشاره شد، مدل رگرسیون لجستیک یکی از مدل‌های مطمئن برای ساخت مدل

#### ۴- نتایج

همانطور که در روش تحقیق گفته شد، ابتدا با آزمون استقلال دو متغیر طبقه‌بندی شده شده، مستقل یا معنادار بودن رابطه دو متغیر وابسته و غیروابسته با یکدیگر را کنترل کرده و در صورت احراز شدن معنادار بودن رابطه آن دو، با استفاده از مدل رگرسیون لجستیک به بررسی هر یک از متغیرها می‌پردازیم. نتایج این آزمون در جدول ۲ قابل مشاهده است.

جدول ۲. نتایج آزمون استقلال دو متغیر طبقه‌بندی شده (متغیر پاسخ: شدت تصادف)

متغیر	p-value
روز وقوع تصادف	۰,۲۴۱
ماه وقوع تصادف	<۰,۰۰۱
ساعت تصادف	۰,۰۰۱۵
نوع راه	<۰,۰۰۱
جنسیت مقصر	۰,۰۵۲۷
سن	۰,۲۶۳۹
نوع وسیله نقلیه مقصر	<۰,۰۰۱
نوع برخورد	<۰,۰۰۱
علت تامه تصادف	<۰,۰۰۱

بیشتر از باقی ساعات شبانه‌روز است ( $OR=1$ )؛ که علت آن می‌تواند خستگی و خواب آلودگی رانندگان در این ساعت باشد و همچنین نبود روشنایی نیز می‌تواند مزیدی بر این علت باشد. نوع راه نیز تنها راه روستایی معنادار شناسایی شده که احتمال وقوع تصادف شدید در آن به میزان ۶۰,۶ درصد از آزادراه کمتر است. در ماه‌هایی که با شدت تصادف در آنها معنادار شده‌اند، هیچکدام به اندازه فروردین ماه احتمال بالایی ندارند ( $OR=1$ ) که این موضوع می‌تواند به علت تعطیلات نوروز باشد که باعث افزایش سفرهای بین‌شهری شده و به همین علت تلفات رانندگی در این ماه نیز بیشتر از سایر ماه‌های سال است (سازمان پزشکی قانونی، آمار متوفیات و مصدومین حوادث رانندگی). در متغیر نوع برخورد، بیشترین احتمال وقوع تصادفات شدید (فوتی و یا جرحی) مربوط به تصادفاتی است که عابر در آن نقش دارد ( $OR=1$ ) و این می‌تواند بخاطر آسیب‌پذیری بالای انسان در برخورد با وسایل نقلیه متحرک باشد؛ در گام دوم نیز تصادفاتی که به علت واژگونی اتفاق افتاده‌اند، بیشترین احتمال آسیب‌پذیری را دارند ( $OR=0,587$ ). به این ترتیب نتایج اشاره شده در نوع برخورد هیچکدام با نتایج اشاره شده در مطالعات پیشین همخوانی ندارد. همانطور که در مطالعات پیشین اشاره شد، احتمال وقوع تصادفات شدید در مردان از زنان بیشتر است

همانطور که از جدول ۲ مشخص است، با استفاده از نتایج آزمون استقلال دو متغیر طبقه‌بندی شده و با سطح اطمینان ۹۵ درصد، متغیر روز وقوع تصادف و سن از شدت تصادف را مستقل از یکدیگر تشخیص می‌دهیم، اما متغیر جنسیت با قدرت خیلی کمی فرض صفر آن تأیید شده و نمی‌توان با اطمینان از مستقل بودن جنسیت و شدت تصادفات سخن گفت؛ و در نهایت می‌گوییم به احتمال ۹۵ درصد متغیرهایی نظیر ماه وقوع تصادف، ساعت تصادف، نوع راه، نوع وسیله، نوع برخورد و علت تامه تصادف از شدت تصادف مستقل نیستند؛ یا به عبارت دیگر سهم دسته‌های هر متغیر با وابستگی معنادار با شدت تصادف در تصادفات فوتی و جرحی متفاوت است. نتایج مدل رگرسیون لجستیک نیز در جدول ۳ مشخص است (در این جدول فقط پارامترهایی که در سطح ۹۵ درصد معنادار شده‌اند، آورده شده است).

با استفاده از جدول ۳ می‌توان گفت تنها روز جمعه در روز تصادف معنادار شناسایی شده و احتمال وقوع تصادف شدید در آن نیز ۳۷,۶ درصد از روز شنبه بیشتر است ( $OR=1,376$ ). ساعت وقوع تصادف در تمام دسته‌بندی‌ها معنادار شناسایی شده و به مانند مطالعه ظفری و همکاران که در مطالعات پیشین به آن اشاره شد، احتمال وقوع تصادف در شب (۱۲ صبح تا ۶ صبح)،

معنادار شناسایی شدند، همانطور که در مطالعات پیشین علت تامه تصادفات یکی از متغیرهای مهم در شدت تصادفات راه‌های برون‌شهری شناسایی شده‌اند. اما مهم‌ترین دسته در علت تامه تصادفات، نقش نقض راه در شدت تصادفات است که با اختلاف معناداری نسبت به سایر دسته‌بندی‌ها در این متغیر معنادار شده‌است (OR=۵,۲۶۸).

که این نتیجه با نتایج بدست آمده در این مطالعه همخوانی دارد و احتمال آن که مردان تصادفات شدیدتری را نسبت به زنان تجربه کنند، ۴۱,۲ درصد بیشتر است (OR=۱,۴۱۲). همچنین برخلاف نتایجی که در مورد متغیرهای سن و وسیله نقلیه مقصر که در مطالعات پیشین به آن اشاره شد، هیچکدام از این دو متغیر در مدل رگرسیون لجستیک ساخته شده در این مطالعه معنادار نشدند. رابطه علت تامه تصادفات و شدت تصادفات با یکدیگر

جدول ۳. نتایج مدل رگرسیون لجستیک برای شدت تصادف جاده‌های برون‌شهری استان اصفهان

متغیر		Coef	Std error	p-value	OR
روز تصادف	شنبه	-	-	-	۱
	جمعه	۰,۳۱۹۵	۰,۱۵۸	۰,۰۴۴	۱,۳۷۶
ساعت تصادف	۱۲ صبح تا ۶ صبح	-	-	-	۱
	۶ صبح تا ۱۲ ظهر	-۰,۳۱۳۵	۰,۱۳۶	۰,۰۲۱	۰,۷۳۱
	۱۲ ظهر تا ۶ عصر	-۰,۵۳۷۶	۰,۱۳۱	۰,۰۰۰	۰,۵۸۴
	۶ عصر تا ۱۲ شب	-۰,۴۴۰۵	۰,۱۳۳	۰,۰۰۱	۰,۶۴۴
نوع راه	آزادراه	-	-	-	۱
	روستایی	-۰,۹۳۰۳	۰,۲۵۴	۰	۰,۳۹۴
ماه وقوع تصادف	فروردین	-	-	-	۱
	خرداد	-۰,۶۱۳۱	۰,۱۸۱	۰,۰۰۱	۰,۵۴۲
	تیر	-۰,۷۱۳۲	۰,۱۸۴	۰,۰۰۰	۰,۴۹۰
	مرداد	-۰,۶۰۴۴	۰,۱۷۸	۰,۰۰۱	۰,۵۴۶
	شهریور	-۰,۷۹۸۲	۰,۲۴۰	۰,۰۰۰	۰,۴۵۰
	مهر	-۱,۰۶۹۸	۰,۲۱۶	۰,۰۰۰	۰,۳۴۳
	آبان	-۱,۲۳۰۱	۰,۲۳۳	۰,۰۰۰	۰,۲۹۲
	آذر	-۰,۸۶۳۹	۰,۲۲۷	۰,۰۰۰	۰,۴۲۲
	دی	-۰,۵۰۲۴	۰,۱۹۶	۰,۰۱۰	۰,۶۰۵
	بهمن	-۰,۷۴۷۸	۰,۲۱۰	۰,۰۰۰	۰,۴۷۳
	نوع برخورد	تصادفات گروه عابر	-	-	-
موتورسیکلت		-۰,۹۲۰۹	۰,۱۸۹	۰,۰۰۰	۰,۳۹۸
سایر وسایل نقلیه		-۰,۸۹۹۲	۰,۲۲۳	۰,۰۰۰	۰,۴۰۷
واژگونی		-۰,۵۳۳۴	۰,۱۵۹	۰,۰۰۱	۰,۵۸۷
علت تامه تصادف	انحراف به چپ	-	-	-	۱
	تغییر مسیر ناگهانی	-۱,۲۹۰۴	۰,۲۹۲	۰,۰۰۰	۰,۲۷۵
	حرکت با دنده عقب	-۱,۹۹۸۰	۰,۷۲۸	۰,۰۰۶	۰,۱۳۶
	حرکت در خلاف جهت	-۱,۲۶۷۹	۰,۴۷۹	۰,۰۰۸	۰,۲۸۱
	عدم توانایی در کنترل وسیله نقلیه	-۰,۵۱۳۹	۰,۱۸۶	۰,۰۰۶	۰,۵۹۸
	عدم توجه به جلو	-۰,۶۵۴۳	۰,۱۷۲	۰,۰۰۰	۰,۵۲۰
	عدم رعایت حق تقدم	-۱,۴۹۷۴	۰,۲۳۶	۰,۰۰۰	۰,۲۲۴

	عدم رعایت فاصله طولی و عرضی	-۱,۴۷۳۸	۰,۴۴۱	۰,۰۰۱	۰,۲۲۹
	نقض راه	۱,۶۶۱۷	۰,۴۹۲	۰,۰۰۱	۵,۲۶۸
	نقض فنی وسیله نقلیه	-۱,۲۶۴۷	۰,۴۷۹	۰,۰۰۸	۰,۲۸۲
	گردش به طرز غلط	-۱,۱۱۱۷	۰,۳۹۱	۰,۰۰۴	۰,۳۲۹
	سرعت نامطمئن و زیاد	-۱,۶۴۸۸	۰,۳۷۲	۰,۰۰۰	۰,۱۹۲
جنسیت	زن	-	-	-	۱
	مرد	۰,۳۴۴۷	۰,۱۷۲	۰,۰۴۵	۱,۴۱۲

## ۵- نتیجه گیری

داده‌های فوتی ثبت شده توسط پلیس راه دارای مقادیر نامشخص بوده که این داده‌ها از پایگاه داده کنار گذاشته شده است. همچنین برای مطالعات آینده می‌توان پیشنهادهای نظیر ساخت مدلی متفاوت از رگرسیون لجستیک استفاده شده در این مطالعه، و یا ساخت مدلی با داده‌های جدیدتر و مقایسه نتایج این مطالعه با مدل جدید و در نهایت ساخت مدل رگرسیون لجستیک با ترکیب متغیرهای غیروابسته ارائه کرد.

در این پژوهش به شناسایی و بررسی عوامل موثر بر شدت تصادفات راه‌های برون‌شهری استان اصفهان و مقایسه آن با مطالعات پیشین پرداخته شد و برای این کار از مدل رگرسیون لجستیک استفاده شد. در این مدل متغیرهایی نظیر ماه وقوع تصادف، ساعت تصادف، نوع راه، نوع برخورد، جنسیت و علت تامة تصادف در رابطه با شدت تصادفات معنادار شناسایی شدند. همچنین در گروه علت تامة تصادفات، نقض راه با اختلاف بسیار زیاد بیشترین نسبت بخت را به خود اختصاص داده بود که نمایانگر بحرانی بودن نقض راه در احتمال وقوع یک تصادف شدید (جرحی و فوتی) است که می‌تواند سازمان‌های ذی‌ربط نظیر سازمان راهداری و حمل‌نقل جاده‌ای استان اصفهان را معطوف به تعمیر و نگهداری راه‌های استان اصفهان در وهله اول و در وهله دوم به بررسی عوامل موثر در کاهش شدت تصادف نظیر گاردریل و سیستم‌ها ضربه‌گیر معطوف کند. همچنین تصادفات در ساعات اولیه روز (۱۲ صبح تا ۶ صبح) نیز بیشترین سهم را در احتمال وقوع تصادفات شدید دارند که می‌تواند موضوع مطالعه‌ای برای تشخیص عوامل موثر بر وقوع تصادفات و علت‌یابی آن در این ساعت از شبانه‌روز در استان اصفهان باشد. همچنین احتمال بیشترین رخداد تصادفات شدید در ماه فرودین که تقریباً نیمی از آن به علت تعطیلات نوروز تعطیل است، که می‌توان راهکارهایی را در سازمان و نهادهای ذی‌ربط برای کاهش تصادفات، نظیر افزایش گشت‌های پلیس و یا استقرار دوربین‌های نظارت جاده‌ای فعال، جست‌وجو کرد.

## ۶- پی‌نوشت‌ها

1. Significance level
2. Odd Ratio
3. Binary Logistic Regression

## ۷- مراجع

-توکلی کاشانی، علی، امیری فر، سعیده، مدقالچی، علی، محمدی، اعظم، و جزوقتی، محمد (۱۴۰۱). پیش‌بینی شدت تصادفات جاده‌ای توسط روش‌های یادگیری ماشین، مطالعه موردی استان زنجان. *نوزدهمین کنفرانس بین‌المللی مهندسی حمل‌ونقل و ترافیک*، تهران.

-توکلی کاشانی، علی، و بشارتی، محمدمهدی (۱۳۹۹). آمار و احتمالات در مهندسی حمل‌ونقل. *دانشگاه علم و صنعت*.

-سازمان پزشکی قانونی، آمار متوفیات و مصدومین حوادث رانندگی. <https://www.lmo.ir>

-شریعت‌مهمینی، افشین، و توکلی کاشانی، علی (۱۳۸۹). تحلیل شدت مصدومیت ناشی از تصادف‌ها در راه‌های دوخطه

از نکات قابل توجه در این مقاله می‌توان به داده‌های اندک تصادفات فوتی (۳ درصد از کل داده‌ها) در مقایسه با داده‌های تصادف جرحی اشاره نمود که می‌تواند باعث ایجاد بایاس در مدل ساخته شده شود. این مشکل از آنجا ناشی می‌شود که

- Highway Safety Manual. (2010). (first ed.). American Association of State Highway Transportation Officials (AASHTO) .
- Tavakoli Kashani, A., & Besharati, M. M. (2016). An Analysis of Vehicle Occupants' Injury Severity in Crashes Occurred On Rural Freeways and Multilane Highways in Iran. *International Journal of Transportation Engineering*, 4(2), 137-146.  
**doi.org/10.22119/ijte.2016.40538**
- Wang, C., Lu, L., & Lu, J. (2015). Statistical Analysis of Bicyclists' Injury Severity at Unsignalized Intersections. *Traffic Injury Prevention*, 16(5), 507-512.  
**doi.org/10.1080/15389588.2014.969802**
- World Health Organization, Road traffic injuries. (2022). World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>
- Zafri, N. M., Prithul, A. A., Baral, I., & Rahman, M. (2020). Exploring the factors influencing pedestrian-vehicle crash severity in Dhaka, Bangladesh. *International Journal of Injury Control and Safety Promotion*, 27(3), 300-307.  
**doi.org/10.1080/17457300.2020.1774618**
- برون شهری با استفاده از مدل‌های داده کاوی. پژوهشنامه حمل و نقل. ۱۶۵-۱۵۳، (۲)۷.
- <https://sid.ir/paper/83930/fa>
- Albalate, D., & Fernández-Villadangos, L. (2010). Motorcycle Injury Severity in Barcelona: The Role of Vehicle Type and Congestion. *Traffic Injury Prevention*, 11(6), 623-631.  
**doi.org/10.1080/15389588.2010.506932**
- Bédard, M., Guyatt, G. H., Stones, M. J., & Hirdes, J. P. (2002). The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. *Accident Analysis & Prevention*, 34(6), 717-727.  
**doi.org/https://doi.org/10.1016/S0001-4575(01)00072-0**
- Benlagha, N., & Charfeddine, L. (2020). Risk factors of road accident severity and the development of a new system for prevention: New insights from China. *Accident Analysis & Prevention*. 136, 105411.  
**doi.org/10.1016/j.aap.2019.105411**

# Using the Python Programming in Developing Data Mining Models for the Accident Severity of Rural Crashes in Isfahan Province

*Ali Tavakoli Kashani, Associate Professor, School of Civil Engineering,  
Iran University of Science and Technology, Tehran, Iran.*

*Mahdi Hadinia, M.Sc., Student, School of Civil Engineering, Iran University of Science and  
Technology, Tehran, Iran.*

*Mohammad Ali Salavati, Ph.D., Candidate, School of Civil Engineering,  
College of Engineering, University of Tehran, Tehran, Iran.*

*E-mail: Alitavakoli@iust.ac.ir*

Received: August 2024- Accepted: December 2024

## **ABSTRACT**

According to Iran's legal medical statistics, the death and injury rate due to traffic crashes has increased from 1396 to 1401. This trend is observable in the whole country and the provinces. It can be a warning sign for the authorities to analyze the factors on crashes and find solutions to reduce the death and injury rate. This research examines the factors affecting the severity of rural crashes in Isfahan province. Logistic regression model is used with 21 thousand data, and statistical analysis is performed on the results. The results show that variables such as month, time, Type of road, type of collision, gender and cause of crash are significant. The probability of a severe crash is higher in the early hours of the day and in Farvardin than the rest of the months of the year; road violation can increase the probability of a severe crash by 5.27 times more than other causes (OR=5.27); and men have 41 percent higher probability of a severe crash than women. Some suggestions are made for further studies on some of these variables.

**Keywords:** Crash Severity, Logistic Regression, Data Mining, Rural Crashes