

## بررسی اثر متغیرهای طرح اختلاط بر عملکرد تراکم مخلوط آسفالتی با کمک الگوریتم‌های داده کاوی

### مقاله پژوهشی

محمدحمیدی‌نیا، دانشجوی کارشناسی ارشد، دانشکده مهندسی عمران، دانشگاه یزد، یزد، ایران

محمد مهدی خیبری\*، دانشیار، دانشکده مهندسی عمران، دانشگاه یزد، یزد، ایران

مهدی مخبری، استادیار، دانشگاه آزاد اسلامی، واحد استهبان، استهبان، ایران

\*پست الکترونیکی نویسنده مسئول [mkhabiri@yazd.ac.ir](mailto:mkhabiri@yazd.ac.ir)

دریافت: ۹۹/۰۱/۲۹ - پذیرش: ۹۹/۰۷/۱۵

صفحه ۷۵-۸۸

### چکیده

عوامل یا پارامترهای زیادی بر تراکم میدانی مخلوط آسفالتی، تاثیر می‌گذارد از آنجا که تعداد متغیرهای موثر بر تراکم زیاد و برخی بر یکدیگر اثر می‌گذارند، تعیین یک رابطه‌ی درونیابی ثابت تقریباً غیر ممکن است. داده کاوی و تکنیک‌های آن راهکاری برای کشف دانش پنهان بین متغیرهای وابسته و مستقل است، همچنین روابط غیر مستقیم و غیرخطی را می‌تواند با تقسیم بندی داده‌ها در گروه‌ها یا برگ، در درخت تصمیم شناسایی کند. در این مطالعه، تکنیک‌های رایج داده کاوی در مهندسی عمران، شامل شبکه عصبی، رگرسیون خطی و درخت تصمیم، به عنوان ابزار تحلیلی مورد استفاده قرار می‌گیرد. با تاکید بر کاربرد روش درخت تصمیم، با هدف بررسی مدل‌های کشف دانش و ارائه پیش بینی، سایر ابزارهای داده کاوی نامبرده برای کمک به ساخت و ارزیابی مدل ساخته شده، بکار گرفته می‌شوند. متغیرهای توضیحی که در سه ابزار آماری این مطالعه مورد استفاده قرار می‌گیرند، شامل: درصد فضای خالی، مقاومت مخلوط آسفالتی، دانه بندی سنگدانه‌ها، درصد قیر، نرمی مخلوط آسفالتی می‌باشد. نتایج نشان می‌دهد متغیرهای درصد فضای خالی مصالح سنگی، درصد عبوری الک شماره‌ی ۲۰۰ و شماره‌ی ۴ و درصد قیر تاثیر بیشتری بر نتایج تراکم و چگالی مخلوط آسفالتی داشته‌اند. همچنین مدل رگرسیون خطی چندگانه با ضریب همبستگی نزدیک به یک بین چگالی مخلوط آسفالتی و متغیرهای تاثیرگذار بر آن ارائه شد.

واژه‌های کلیدی: طرح اختلاط، مخلوط آسفالتی، داده کاوی، کنترل کیفیت تراکم، خرابی روسازی، درخت تصمیم

### ۱- مقدمه

همچنین روابط غیر مستقیم و غیر خطی را می‌تواند با تقسیم بندی داده‌ها در گروه‌ها یا برگ‌های درخت تصمیم شناسایی کند. داده کاوی راهی است برای تحلیل خودکار داده‌ها و شناسایی الگوهای پنهان که انجام آن به صورت دستی امکان پذیر نیست (بریمان و همکاران، ۱۹۸۴). اهمیت کشف این ارتباط، از این لحاظ است که بسیاری از عوامل مهم در فرآیند طرح اختلاط شناسایی می‌شوند و در مرحله اجرا از اثرات نامطلوب آن با کنترل کیفیت دقیق جلوگیری می‌شود. گائو و همکاران (۲۰۱۴) در تحقیق خود بر روی خصوصیات

امروزه خرابی روسازی یکی از مشکلات صنعت حمل و نقل جاده ای است، که برطرف کردن عیوب آن مستلزم از بین بردن دلایل وقوع آن است، از جمله دلایل خرابی زودرس، عدم تراکم مناسب روسازی است. بر تراکم میدانی مخلوط آسفالتی، عوامل یا پارامترهای زیادی تاثیر می‌گذارد از آنجا که تعداد متغیرهای موثر بر تراکم زیاد و برخی بر یکدیگر اثر می‌گذارند، تعیین یک رابطه‌ی خطی معین تقریباً غیر ممکن است. داده کاوی و تکنیک‌های آن راهکاری برای کشف دانش پنهان بین متغیرهای وابسته و مستقل است،

تراکم به این نتیجه رسیدند که کیفیت تراکم مخلوط آسفالتی داغ بر عملکرد آن موثر است. تراکم با کیفیت بالا ظرفیت تحمل، ثبات دما و دوام روسازی را بهبود می‌بخشد و تراکم ضعیف باعث تخلخل، ناهمواری، نفوذ پذیری و شیار شدگی می‌شود. بسیاری از خرابی‌های پیش از موعد با تراکم بد و نامناسب در ارتباط است. یک درصد افزایش هوا بیشتر در اثر عدم تراکم کافی از میزان طراحی در مخلوط آسفالتی باعث کاهش ۳۵ درصدی عمر خستگی روسازی و نفوذ پذیری دو برابر روسازی می‌شود. کاسم و همکاران (۲۰۱۲) در تحقیق خود در رابطه با ارزیابی کامل تراکم روسازی‌های آسفالتی و یک روش مناسب برای پیش‌بینی تراکم به این نکته اشاره نمودند که تراکم مخلوط‌های آسفالتی تأثیر بسیار زیادی بر عملکردشان دارد و تراکم ناکافی، حتی با وجود این که تمامی پارامترهای طرح اختلاط مورد قبول باشد، موجب خرابی‌هایی مانند تغییر شکل دائمی پیش از موعد، پیر شدگی شدید و خرابی‌های ناشی از رطوبت می‌شود. در حال حاضر الگوی تراکم مورد نیاز برای رسیدن به دانسیته (چگالی) مورد نظر براساس تجربه و سعی و خطا بدست می‌آید که این روند زمان بر و پرهزینه می‌باشد. در این مطالعه چند نوع مخلوط آسفالتی با روش‌های مختلف تراکم مورد ارزیابی واقع شدند. از نتایج این آزمایش‌ها برای ارزیابی تأثیرات دمای تراکم، روش تراکم، طرح اختلاط و نوع اساس بر تراکم پذیری مخلوط‌های آسفالتی استفاده شد، نتایج نشان داد، تعداد متغیرهای تأثیرگذار بر تراکم زیاد با روابط پیچیده بین آنهاست. لیو و همکاران (۲۰۱۸) در تحقیقی درباره شبیه سازی مخلوط‌های آسفالتی متراکم شده توسط روش‌های مختلف به این نتیجه رسیدند که مهمترین هدف در تراکم مخلوط آسفالتی، افزایش چگالی و ایجاد فضای خالی مناسب می‌باشد که این امر، باعث بهبود مشخصات مکانیکی (مقاومت، روانی، استحکام کششی و مدول الاستیسیته و سایر پارامترهای مهم) شده و همچنین دوام مخلوط آسفالتی (به معنی کاهش نفوذ آب و هوا برای جلوگیری از پدیده عریان‌شدگی و کاهش پیرشدگی) را در هنگام بهره‌برداری روسازی افزایش می‌دهد. کوک و همکاران (۲۰۱۷) در تحقیقی با هدف مطالعه اثرات شرایط تراکم میدانی بر ویژگی‌های مکانیکی مخلوط آسفالتی و سپس مقایسه این ویژگی‌ها با ویژگی‌های نمونه‌های آماده شده با تعداد چکش‌های مارشال متفاوت به این نتیجه رسیدند که نمونه‌های

آزمایشگاهی و میدانی به جهت‌های مختلف تحت تأثیر سطح انرژی و میزان تراکم و دما قرار می‌گیرند. با توجه به اهمیت موضوع ایمنی و حوادث رانندگی، در سال‌های گذشته روش‌ها و مدل‌های گوناگونی جهت ارزیابی، پیش‌بینی وقوع سوانح رانندگی ارائه شده است؛ از جمله این روش‌ها می‌توان به روش شبکه عصبی مصنوعی و روش نوین درختان تصمیم‌گیری یا مدل‌های درختی، در پیش‌بینی حوادث رانندگی اشاره کرد. مدل درخت تصمیم بر خلاف مدل شبکه عصبی مصنوعی به تولید قانون می‌پردازد. پیش‌بینی به دست آمده از ساختار درخت تصمیم، در قالب یک سری قواعد و قوانین دانشی جدید توضیح داده خواهد شد. همچنین بر خلاف شبکه عصبی مصنوعی در درخت تصمیم، این امکان وجود دارد که داده‌ها به صورت غیر عددی ویا اسمی باشند. از ویژگی درختان تصمیم این است، که ابزاری هستند که قابلیت پاسخگویی به مسائل پیچیده و غیر خطی و با متغیرهای زیاد دارند و از این توانایی در علوم دیگر از جمله مهندسی مخابرات و الکترونیک که اساساً دارای مسائل غیر خطی هستند به خوبی توانسته‌اند مسائل مربوطه را حل نمایند، لذا این الگوی جدید و ابزار آماری نو به علوم مهندسی دیگر که دارای پدیده‌های غیر خطی و پیچیده و حتی تصادفی هستند نیز باز شده است. حجم زیادی از داده‌ها توسط سازمان‌های کنترل و پایش نظیر آزمایشگاه‌ها امروزه در مقیاس‌های مکانی و زمانی در پروژه‌های مختلف امروزه ثبت می‌گردد. نظر به گسترش سیستم پایگاه داده و ایجاد ابزارهای متعدد برای ذخیره حجم بالایی از داده‌ها، داده کاوی به عنوان شاخه علوم بین رشته‌ای توسعه روز افزونی یافته است. درخت تصمیم<sup>۱</sup> در روش‌های داده‌کاوی<sup>۲</sup> به جستجو و کشف مدل‌های گوناگون، خلاصه‌سازی و دریافت مقادیر از بانک داده‌ها و استخراج مدل به منظور تشریح خوشه یا رسته‌های مهم داده‌ها، فهم و پیش‌بینی رفتار آنها در آینده به کار می‌رود. مدل‌های رسته‌ای در تحلیل داده‌های گسسته و طبقه‌ای، و مدل‌های پیش‌بینی یا درونیابی بیشتر روی داده‌های غیرگسسته کار می‌کنند، از جمله تحقیقات بین‌المللی و داخلی صورت گرفته بر روی کارایی درخت تصمیم در زمینه علوم منابع طبیعی می‌توان به مطالعات دستورانی و همکاران، اشاره نمود (دستورانی و همکاران، ۲۰۰۵)، البته در زمینه مهندسی عمران و راه و شهرسازی نیز مطالعات مختلفی صورت گرفته است.

هدف از این مطالعه این است که آیا می‌توان از روش درخت تصمیم برای برای تجزیه و تحلیل رابطه بین عوامل خطر و تصادفات استفاده نمود یا خیر. این کار با مقایسه نتایج روش درخت تصمیم با مدل رگرسیون دوجمله‌ای منفی انجام شد (مسلمی و همکاران، ۲۰۰۵). در مطالعه‌ای خیبری و همکاران (۲۰۱۰)، به بررسی تاثیر خرابی روسازی راه‌های درون‌شهری و شرایط ترافیکی موثر بر تصادفات پرداخته‌اند. از آنجایی که ناهمواری روسازی، تاثیر بسزایی بر تصادفات جاده‌ای داشته و لازم است در برنامه‌ریزی ایمنی جاده‌ها و بهبود وضعیت روسازی‌ها در نظر گرفته شود، به صورت موردی از بین خرابی‌های روسازی، فرورفتگی‌ها و برآمدگی‌های سطح روسازی مورد مطالعه قرار گرفتند. اطلاعات ۱۸ نقطه از روسازی دارای خرابی و برای هر نقطه عکس‌العمل و سرعت ۳۰ راننده در حین مواجه با آن ثبت نمودند. جهت تعیین عوامل موثر بر میزان انحراف رانندگان از روش درخت تصمیم استفاده شده است. درخت حاصل نشان می‌دهد متغیرهای مساحت خرابی، طول و عرض آن در مقایسه با متغیرهای حجم ترافیک، سرعت وسیله‌نقلیه، فاصله مرکز خرابی تا گوشه سمت راست خط عبوری و نوع ناهمواری (فرورفتگی یا برآمدگی) تاثیر بیشتری بر انحراف رانندگان از مسیر را داشته‌اند. مطالعات دیگری نظیر هوآنگ و همکاران (۲۰۱۹) و موسی و همکاران (۲۰۱۹) با ابزار یادگیری ماشین و نیز سایر الگوریتمهای داده کاوی برای دستیابی به دانشهای جدید در روسازی نظیر عملکرد روکش رویه بتنی در سالیان اخیر در دست انجام است. در مجموع با بررسی منابع علمی بر روی موضوع کاربرد درخت تصمیم بر عملکرد تراکم لایه‌های روسازی و متغیرهای مرتبط با آن کمتر تمرکز شده است، بنابراین، هدف از انجام این مطالعه، استفاده از تکنیک داده کاوی درخت تصمیم CART برای تعیین اینکه کدام پارامترها تاثیر بیشتری بر کیفیت، حساسیت و میزان تراکم رویه آسفالتی در پروژه‌های اجرا شده می‌گذارند و بر اساس آن‌ها کنترل‌های دقیق‌تری در میدان بر روی پارامترها انجام شود. در نهایت این تحقیق از الگوریتم CART به عنوان یکی از انواع درختان تصمیم رگرسیونی جهت کشف دانش و رابطه بین متغیرهای موثر در تراکم میدانی، استفاده می‌شود.

به عنوان یک مورد برجسته، به تحقیقات ناث و همکارانش (۲۰۱۰)، با استفاده از تکنیک داده‌کاوی مکانی و الگوریتم خوشه‌بندی به روش K-MEANS به برآورد زمان سفر اشاره می‌شود. روش این الگوریتم این گونه است که داده‌های مشابه را در گروه‌های مجزا دسته بندی می‌نمایند به طوری که گروه‌های قرار گرفته شده در هر دسته هیچ مشابهتی با دسته‌های دیگر ندارند. داده‌های این تحقیق با استفاده از دستگاه GPS<sup>۳</sup> ای که درون خودروها می‌باشد، جمع‌آوری شده‌اند. در روش K-MEANS یک عدد یا یک عارضه به عنوان مرکز دسته شناخته می‌شود و اعداد دیگر یا عارضه‌های دیگر با استفاده از متد فاصله و یا مشابهت به آن مرکز دسته تعلق پیدا خواهند کرد. مزیت تکنیک خوشه‌بندی K-MEANS نسبت به تکنیک‌های دیگر خوشه بندی در این می‌باشد که در هر تکرار انجام شده برای به دست آوردن داده‌های مشابه با هم باشند. در نهایت با استفاده از فرمول‌های آماری زمان سفر مشخص می‌گردد. چودهاری و همکارانش (۲۰۰۹)، با استفاده از دو تکنیک میانگین متحرک<sup>۴</sup> و میانگین ترتیبی<sup>۵</sup> به برآورد زمان سفر پرداخته‌اند. این دو تکنیک بر پایه یک سری فرمول‌های ریاضی ساده استوار می‌باشند. در این تحقیق برای برآورد زمان سفر از داده‌های پیش بینی که توسط خودروها جمع آوری شده است، استفاده شده است و با استفاده از روش میانگین خطای نسبی مطلق دقت زمان سفر برآورد شده را ارزیابی می‌نمایند. روش میانگین متحرک نسبت به روش میانگین ترتیبی دارای دقت بهتری می‌باشد. از مزیت‌های این دو الگوریتم می‌توان اشاره به سادگی، دقت بالا، هزینه‌های پایین و آنکه پیچیدگی سیستم به علت اجرای ساده هر دو روش پایین آمده و نوسانات یا نویزهایی که در مجموعه داده ها وجود داشته حذف می‌گردند. در تحقیق انجام شده توسط چنگ و همکاران (۲۰۰۵) بر روی داده کاوی مدل‌های درختی برای تجزیه و تحلیل نرخ تصادفات آزادراه از روش درخت تصمیم<sup>۶</sup> (CART) که یکی از تکنیک‌های کاربردی و گسترده داده کاوی بوده و در صنعت و مهندسی نیز کاربرد دارد، استفاده شد. مدل درخت تصمیم به هیچ رابطه از پیش تعریف شده‌ای بین متغیر مستقل و متغیر وابسته نیاز نداشته و یک ابزار قدرتمند برای پیشبینی و طبقه بندی مشکلات می‌باشد. اما به دلیل استفاده کم از روش درخت تصمیم در تجزیه و تحلیل مشکلات ایمنی ترافیک،

## ۲- الگوی آماری درخت تصمیم

الگوهای اصلی داده کاوی به دو دسته اصلی قابل تقسیم می‌باشند: توصیفی و پیشبینانه. در روش توصیفی وظیفه این است که خواص عمومی داده‌ها مشخص می‌شود. هدف از توصیف، یافتن الگوهایی پنهان در داده‌هاست که برای انسان و درک مغز بشری به لحاظ پیچیدگی و تعداد روابط به راحتی قابل تفسیر نباشد. از الگوی پیشبینانه به منظور تخمین رفتارهای آینده متغیرها استفاده می‌شوند. منظور از پیش‌بینی به کارگیری چند داده و متغیر یا فیلد در پایگاه داده‌ها برای پیش‌بینی مقادیر آینده یا ناشناخته دیگر متغیرهای مورد نظر است (دستورانی و همکاران، ۲۰۰۵). الگوی درخت تصمیم یکی از روش‌های داده‌کاوی و از ابزارهای قدرتمند و رایج برای دسته‌بندی و پیش‌بینی می‌باشد که در نقطه مقابل شبکه‌های عصبی به تولید قانون می‌پردازد. بدین معنی که درخت تصمیم پیش‌بینی خود را در قالب یکسری قوانین ارائه می‌دهد. این در حالیکه در شبکه‌های عصبی مصنوعی فقط متغیر یا متغیرهایی پیش‌بینی بیان می‌شود و روابط پیش‌بینی آن در شبکه پنهان باقی می‌ماند. علاوه بر آن در ساخت درخت تصمیم بر خلاف شبکه‌های عصبی، می‌توان از داده‌های اسمی یا غیر عددی نیز استفاده نمود. ساخت این الگو بر سه اصل استوار است:

۱- مجموعه‌ای از سوالات به شکل یک تابع در رابطه (۱) که در آن  $X$  یک متغیر مستقل و  $d$  یک مقدار ثابت و جواب هر سوال مثبت یا منفی است.

$$x \leq d^+ \quad (1)$$

۲- بهترین معیار ایجاد شاخه جهت انتخاب بهترین متغیر مستقل برای شاخه‌زدن.

۳- ایجاد آمار خلاصه در هر گره انتهایی (بریمان و همکاران، ۱۹۸۴)

معیارهای مختلفی جهت ایجاد شاخه و تولید درخت تصمیم وجود دارد، این مطالعه به استفاده از درخت تصمیم رگرسیونی پرداخته، که معیار مورد استفاده‌ای آن روش انحراف حداقل مربعات<sup>۷</sup> (LSD) نام دارد. تعریف معیار مورد استفاده به صورت رابطه‌ی (۲) ارائه می‌شود (مظفری و امیدوار، ۲۰۱۵):

$$SS(t) = \sum_{i=1}^{Nt} (y_i(t) - \bar{y}(t))^2 \quad (2)$$

$Nt$ : تعداد داده‌ها در هر گره برگ،

$Yt$ : مقدار خروجی (متغیر هدف در گره برگ)،

$\underline{Y}t$ : میانگین مقادیر متغیر هدف برای همه گره‌ها.

در این مرحله متغیر ورودی  $S$  زمانی بهترین متغیر برای ایجاد شاخه جدید در گره  $t$  می‌باشد که مقدار  $Q(s,t)$  را بیشینه نماید.

$$Q(s,t) = SS(t) - SS(t_R) - SS(t_L) \quad (3)$$

متغیرهای بکاررفته در این رابطه عبارتند از:

$SS(t_i)$ : میزان در شاخه سمت راست و

$SS(t_r)$ : سمت چپ گره  $t$  می‌باشد.

پارامترهای پیچیدگی، خصوصیات شکل درخت را تعیین می‌کنند و به اندازه‌ی نمونه بستگی ندارند. پارامترهای پیچیدگی یک درخت به صورت تعداد برگهای درخت، تعداد گره‌های داخلی آن و بیشترین طول مسیر از ریشه تا یک برگ می‌باشند. همچنین می‌توان از طول یک مسیر خارجی استفاده کرد که به صورت تعداد شاخه‌هایی تعریف می‌شود که یک درخت کامل را تشکیل می‌دهند. بهترین درخت با توجه به این شرط باید دارای کمترین مقدار  $Q(S,T)$  باشد. از روشی که در آن بیشترین پیچیدگی مجاز برای درخت تعیین می‌شود به طور همزمان با جستجوی دقیقترین درخت هم می‌توان استفاده کرد. در باب کیفیت یک درخت تصمیم، توسط پاک‌گوهر و همکاران (۲۰۰۸) بیان می‌شود، اگر یک درخت تصمیم با نمونه‌ای از  $N$  شی وجود داشته باشد. امکان انتخاب دو نوع اصلی از پارامترهای توضیح دهنده کیفیت آن درخت وجود دارد، نوع اول پارامترهای دقت هستند و نوع دوم پارامترهای پیچیدگی درخت می‌باشد. پارامترهای دقت یک درخت را می‌توان با کمک دقت در خوشه‌بندی نمونه تعریف کرد و کیفیت تقسیم اشیاء در کلاس‌های مختلف (در مورد یک مسئله تشخیص)، یا اندازه بزرگی خطا (در مورد یک مسئله تحلیل رگرسیون) را تعیین کرد. عدد نسبی (فراوانی) خطاها به معنای کسری از اشیا است که توسط درخت به طور اشتباه به یک کلاس نسبت داده شده، از رابطه‌ی (۴) محاسبه می‌شود.

$$\hat{P}_{err} = \frac{N_{err}}{N} \quad (4)$$

$$N_{err} = \sum_{s=1}^M \sum_{i=1}^K N_i^s \quad (5)$$

$$i \neq \hat{Y}(S)$$

و  $K$  تعداد کلاسه‌ها است.

واریانس نسبی برای یک درخت تصمیم را می‌توان به صورت رابطه‌ی (۶) محاسبه کرد:

$$d_{an} = \frac{d_{oc}}{d_o} \quad (6)$$

که در این رابطه:

$$d_{oc} = \frac{1}{N} \sum_{S=1}^M \sum_{i \in Data^S} (\hat{Y}(S) - y^i)^2 \quad (7)$$

واریانس باقیمانده است، واریانس اولیه به صورت رابطه‌ی (۸) تعریف می‌شود:

$$d_o = \frac{1}{N} \sum_{i=1}^N (y^i - \bar{y})^2 \quad (8)$$

که در آن نیز رابطه‌ی (۹) برقرار است.

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y^i \quad (9)$$

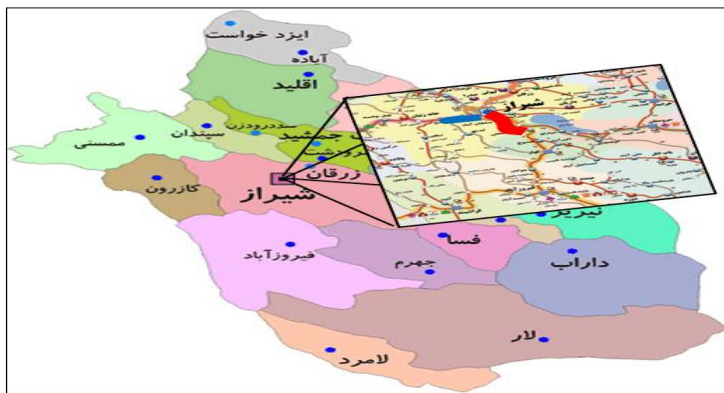
باتوجه به این توضیحات پارامترهای پیچیدگی، خصوصیات شکل درخت را تعیین می‌کنند (پاک‌گوهر و همکاران، ۲۰۰۸) و به ابعاد و حجم نمونه بستگی ندارند.

### ۳- داده‌ها و روش‌شناسی

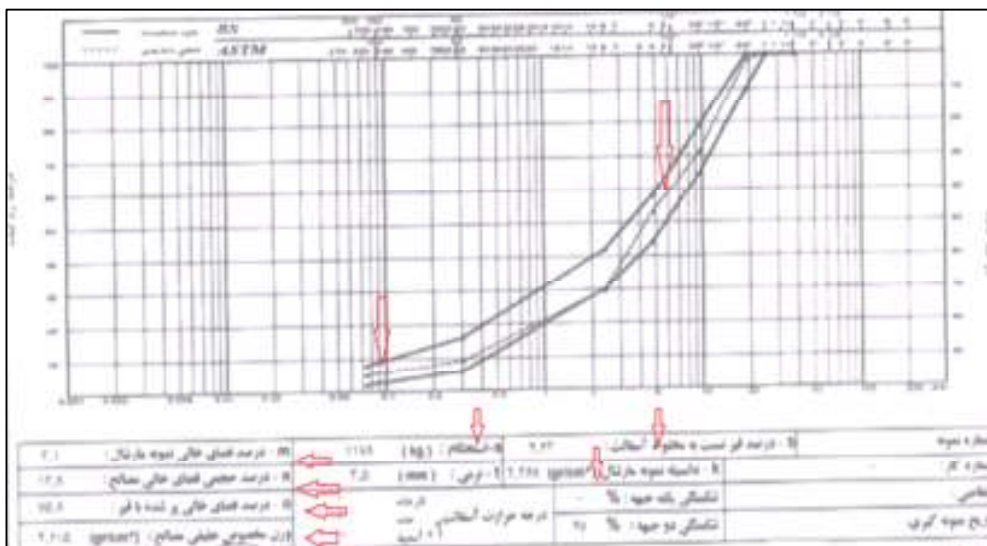
در این تحقیق به علت پیچیده بودن روابط کلی و تغییرات پارامترها برای شناسایی دلایل عدم تراکم کافی روسازی می‌توان از تکنیک‌های داده کاوی استفاده کرد. داده کاوی یک هنر و علم استخراج اطلاعات پنهان، از مجموعه‌ای از داده‌های فراوان است. این روشها به عنوان مجموعه‌ای از روشها در فرایند کشف دانش است که برای تشخیص الگوها و رابطه‌های نامعلوم در داده‌ها مورد استفاده قرار می‌گیرد. داده کاوی راهکاری است برای یافتن

الگوهای جدید و پنهان و غیرمنتظره که در داده‌ها استفاده می‌شود. ابتدا پارامترهای طرح اختلاط و اجرایی چندین پروژه جمع آوری شده و سپس اطلاعات برداشت شده با استفاده از مدلسازی رایانه‌ای و نرم افزارهای آماری مربوطه مورد تجزیه و تحلیل قرار می‌گیرند. جهت اجرای مدل درختان تصمیم از آمار و اطلاعات مربوط به برخی پارامترهای آزمایشگاه مخلوطهای آسفالتی واقع در استان فارس مطابق شکل (۱) ثبت شده، استفاده گردید. نمونه‌ای از این شیت‌های آزمایشگاهی و متغیرهای مورد استفاده در شکل (۲) نمایش داده شده‌است. داده‌های مورد استفاده در این تحقیق مربوط به اطلاعات یک آزمایشگاه مکانیک خاک و روسازی در شیراز

شیراز می‌باشد به تفکیک پروژه‌های جاری و پایان یافته در دسترس گرفته است. شیت داده‌های آزمایشگاهی از ۳۰ پروژه اجرایی به صورت رکوردهایی برای هر پروژه شامل متغیرهای عمومی درصدفضای خالی، درصد قیر، درصدعبوری از الک شماره‌ی ۴ و الک شماره‌ی ۲۰۰ و چگالی مخلوط و چگالی مصالح سنگی و فضای خالی بین مصالح سنگی از پایگاه داده مذکور استخراج شدند. جدول (۱) مشخصات متغیرهای تحقیق را نشان می‌دهد. در مجموع تعداد داده‌ها بالغ بر ۲۷۰ رکود اطلاعاتی در محاسبات استفاده گردید. انتخاب داده‌های آموزش و آزمون به صورت سیستماتیک (۵۰-۵۰) و یا می‌تواند توسط کاربر انجام شود و بدیهی است که داده‌های آزمون در مرحله آموزش مورد استفاده قرار نمی‌گیرند. در این فرایند داده‌ها به دو دسته داده‌های آموزشی و داده‌های آزمایشی تقسیم شدند. در این تحقیق ۷۵٪ از کل داده‌ها به آموزش مدل تعلق گرفت و ۲۵٪ باقیمانده به عنوان داده‌های آزمون به مدل معرفی گردید.



شکل ۱. موقعیت جغرافیایی تهیه نمونه‌های آزمایشگاهی (استان فارس)



شکل ۲. مشخصات متغیرهای استخراج شده از یک شیت نمونه‌ی آزمایشگاهی

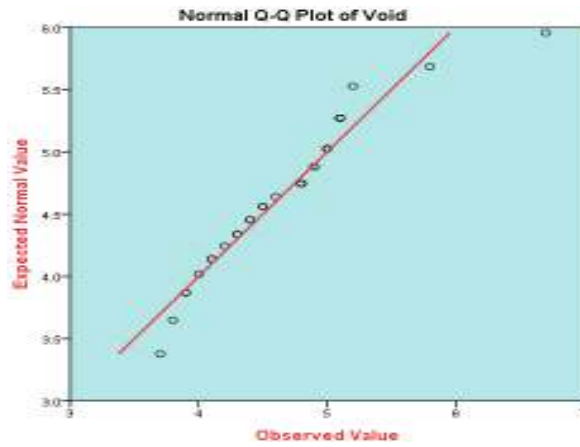
نمایش توزیع فراوانی میانگین درصد فضای خالی مشاهده شده، به عنوان متغیر نماینده تراکم، و به دست آمده از توزیع، احتمال و بیول را برای داده‌های دریافت شده، به ترتیب از دو شکل (۳) و (۴) استفاده شده است.

برای تعیین اولویت متغیرها در مطالعه‌ی حاضر در ابتداء از شبکه عصبی مصنوعی در محیط نرم افزار SPSS، که یکی از روش‌های محاسباتی داده کاوی است، استفاده شد. این روش کمک فرآیند یادگیری و با استفاده از پردازشگرهای ساده تلاش می‌کند با شناخت روابط ذاتی میان داده‌ها، نگاهی میان فضای ورودی (لایه ورودی) و فضای مطلوب (لایه خروجی) ارائه دهد. شکل (۵) اولویت بندی متغیرهای تاثیرگذار در این مطالعه را به این روش نمایش می‌دهد. ملاحظه می‌شود، اولویت بندی بیشترین تاثیر در متغیر وابسته فضای خالی، ناشی از فضای خالی مصالح سنگی، درصد قیر و دانه بندی مصالح سنگی و درجه‌ی نرمی مخلوط آسفالتی است.

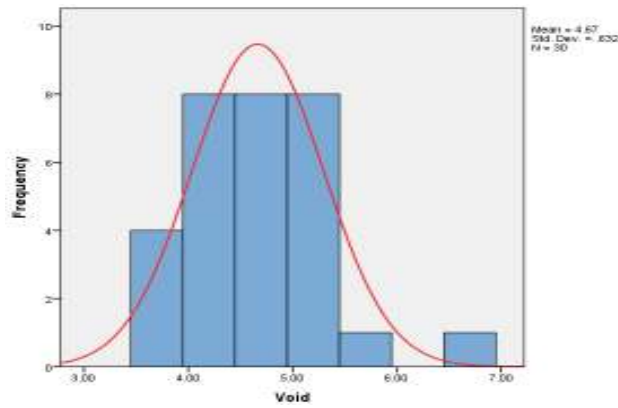
داده‌های بکار رفته در این مطالعه به صورت ادامه معرفی می‌شوند: درصد فضای خالی مخلوط: این متغیر نمایانگر هوای موجود در مخلوط آسفالتی است که به کیفیت تراکم مخلوط آسفالتی وابسته می‌باشد. چگالی مخلوط آسفالتی: این متغیر نمایانگر وزن مخصوص واقعی مخلوط می‌باشد. هرچه تراکم نمونه آسفالتی بیشتر باشد، چگالی مخلوط نیز افزایش می‌یابد. درجه نرمی مخلوط آسفالتی: این متغیر نمایانگر میزان انعطاف پذیری مخلوط است و حدود مورد پذیرش آئین نامه‌ها بین ۲-۳/۵ میلی‌متر است. مقاومت مخلوط آسفالتی: این متغیر بیانگر میزان مقاومت مخلوط است، که حدود مورد پذیرش آئین نامه‌ها بیش از ۸۰۰ کیلوگرم-نیرو است. درصد فضای خالی مصالح سنگی: این متغیر نمایانگر فضای آزاد بین مصالح سنگی است. دانه بندی مصالح سنگی: این متغیر نمایانگر دانه بندی مصالح سنگی است. که از درصد عبوری دو الک مرزی شماره ۲۰۰ و شماره ۴ برای بیان این متغیر استفاده شده است. جهت بررسی نرمال بودن و

جدول ۱. معرفی متغیرهای موثر از شیت‌های آزمایشگاهی و داده‌های برداشت شده

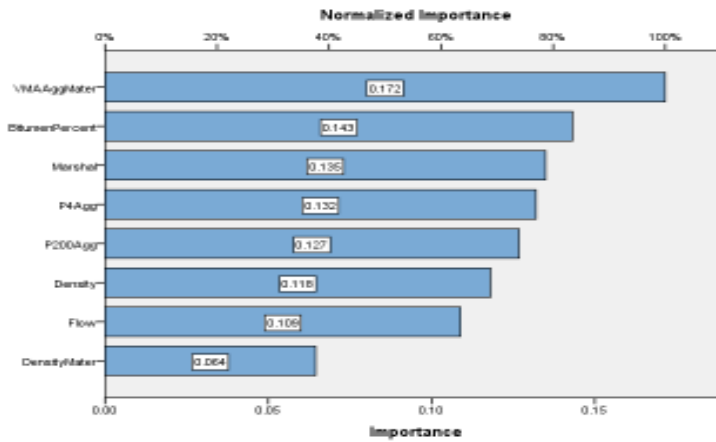
کد متغیر در مجموعه داده و واحد	تشریح متغیر	حدود تغییرات
Density(gr/cm <sup>3</sup> )	چگالی مخلوط آسفالتی	(۲/۲۹-۲/۳۵)
Marshal(kgf)	مقاومت مارشال	(۱۲۶۰-۱۸۷۵)
Flow(mm)	روانی	(۲/۱-۲/۸)
BitumenPercent(%)	درصد قیر در مخلوط	(۳/۴۶-۴/۵۴)
Void(%)	درصد فضای خالی آسفالت	(۳/۷۰-۶/۷۰)
VMAAggMater(%)	درصد فضای خالی مصالح سنگی	(۱۲/۱۰-۱۴/۹۰)
P200Agg(%)	درصد عبوری الک ۲۰۰	(۱-۱۰)
P4Agg(%)	درصد عبوری الک ۴	(۴۵-۵۳)



شکل ۳. فراوانی متغیر درصد فضای خالی مشاهده شده و محاسبه شده با استفاده از تابع ویبول (آزمون نرمال)



شکل ۴. بافت نگار فراوانی متغیر درصد فضای خالی در نتایج آزمایشگاهی



شکل ۵. بافت نگار متغیر درصد فضای خالی در نتایج آزمایشگاهی در مدل شبکه عصبی

#### ۴- کاربرد الگوریتم‌های داده‌کاوی در تعیین متغیرهای موثر

درخت تصمیم که تعداد شاخه‌ها ۵ و تعداد گره‌ها، ۹ تا است، از این جهت که مقایسه‌ای گرافیکی با متغیر درصد فضای خالی بوجود آید، ارائه گردید. با ملاحظه این شکل گسترش اندک و حذف تعدادی از متغیرهای وارد شده در ساخت مدل مشاهده می‌شود. جدول (۳) متغیرهای دارای اهمیت بالاتر بر میزان چگالی مخلوط آسفالتی را نمایش می‌دهد، در این جدول تعداد متغیرهای موثر کاهش یافته است. متغیرهای مشترک در این دو جدول (۲) و (۳) درصد فضای خالی مصالح سنگی، درصد عبوری الک شماره‌ی ۲۰۰ و شماره‌ی ۴ و درصد قیر است.

در مرحله ترسیم نمودار درختی، با تحلیل آماری و مصورسازی داده‌های تحقیق، یک دانش مقدماتی از داده‌ها به دست می‌آید، که این دانش مقدماتی در مراحل بعدی به تشریح و تفسیر نتایج مدل سازی کمک شایانی خواهد نمود. جدول ۲ برخی از پارامترهای آماری مهم موثر بر متغیر اصلی تحقیق را از روش درخت تصمیم نشان می‌دهد. شکل (۶) در محیط گرافیکی نرم‌افزار SPSS، مدل درخت تصمیم را نمایش می‌دهد، مدل ساخته شده دارای ۱۷ گره و ۱۸ شاخه و ۳۵ برگ است. مدل درخت تصمیم دیگری برای متغیر چگالی که به عنوان نماینده دیگری از تراکم مخلوط آسفالتی محسوب می‌شود، ارائه شده در شکل (۷) توسعه یافت، این

جدول ۲. تعیین متغیرهای مهم و میزان تاثیرگذاری هر یک در مدل درخت تصمیم در فضای خالی مخلوط آسفالتی

اولویت	نام متغیر	میزان اهمیت	میزان اهمیت نرمال شده
۱	Density	۰/۲۶۳	۱۰۰
۲	VMAAggMater	۰/۱۸۳	۶۹/۷
۳	Marshal	۰/۱۵۵	۵۹
۴	P200Agg	۰/۰۸۵	۳۲/۵
۵	P4Agg	۰/۰۷۶	۲۸/۸
۶	BitumenPercent	۰/۰۱۶	۶/۹
۷	Flow	۰/۰۱۴	۵/۵

جدول ۳. تعیین متغیرهای مهم و میزان تاثیرگذاری هر یک در مدل درخت تصمیم

اولویت	نام متغیر	میزان اهمیت نرمال شده
۱	VMAAggMater	۱۰۰
۲	Void	۷۷/۲
۳	Flow	۵۹/۸
۴	P200Agg	۲۷/۹
۵	P4Agg	۱۵/۲
۶	BitumenPercent	۱۱/۰۶

#### ۴-۱- قانون انجمنی از اجرای مدل درخت تصمیم در شرایط استفاده از اصل داده‌ها

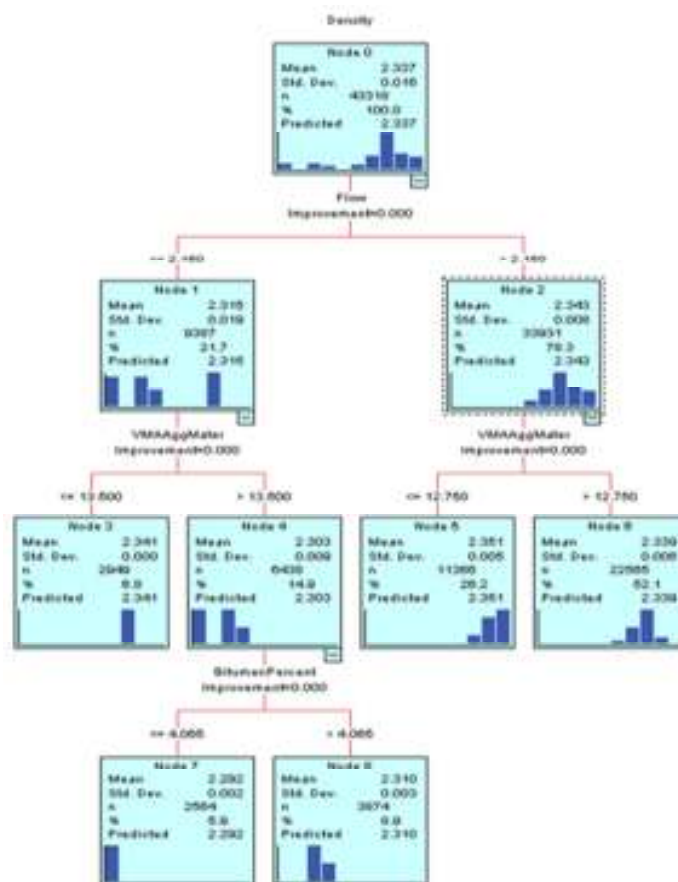
شکل‌های (۶) و (۷) البته به صورت گویاتری برای متغیرهای وابسته مختلف قابل مشاهده است. بررسی این قوانین و مقایسه آن با نتایج سایر محققین نشان می‌دهد که درصد فضای خالی آسفالت تابع درصد فضای خالی مصالح سنگی و دانه‌بندی و درصد قیر است که سایر منابع علمی این موضوع را از جمله نتیجه مطالعات ژئو و القادی، تأیید می‌کنند (ژئو و همکاران، ۲۰۱۹). افزایش درصد فضای خالی مخلوط آسفالتی و افزایش ریزدانه در مخلوط چگالی مخلوط آسفالتی طبق قوانین فوق بر چگالی یا ترکم مخلوط آسفالتی تأثیر مستقیم دارد، که همراستا با نتایج مطالعات اسماعیل و همکاران، است (اسماعیل و همکاران، ۲۰۱۹).

قوانین غیر آشکار و ارتباطات بین متغیرهای تحقیق توسط روش‌های درخت تصمیم قابل ارائه است، از مدل‌های درخت تصمیم ساخته شده در این مطالعه سه قانون پنهان و رابطه‌ی علی-معلولی بین متغیرهای مستقل (درصد فضای خالی و چگالی مخلوط آسفالتی) و متغیرهای مستقل (درصد فضای خالی مصالح سنگی، درصد‌های عبوری در دانه‌بندی سنگدانه‌ها، مقاومت مارشال و درصد قیر و سایر متغیرها) که در مراحل ساخت مدل‌های درخت تصمیم در محیط خروجی نرم‌افزار حاصل شده است، در جدول (۴) ارائه و بحث می‌شود. تعداد قوانین پیش‌بینی متعددی از درخت تصمیم می‌توان حاصل نمود که نمایش‌های گرافیکی





شکل ۶. مدل درخت تصمیم برای درصدفضای خالی در نتایج آزمایشگاهی



شکل ۷. مدل درخت تصمیم برای چگالی مخلوط آسفالتی حاصل از نتایج آزمایشگاهی

جدول ۴. قوانین انجمنی مستخرج از داده کاوی در مدل درخت تصمیم توسعه یافته

شرح	قانون انجمنی
اگر درصد عبوری الک شماره ۴ کمتر از ۴/۷۵٪ باشد و درصد فضای خالی مخلوط آسفالتی از ۵/۵٪ بیشتر باشد و درصد عبوری الک شماره ۲۰۰ از ۱/۵٪ کمتر باشد و درصد فضای خالی مصالح سنگی بیش از ۱۴/۸٪ باشد و مقاومت مارشال از ۱۲۵۵ کیلوگرم-نیرو کمتر باشد، به طور متوسط چگالی مخلوط آسفالتی ۲/۳۲۴ پیش‌بینی می‌شود.	الف پیش‌بینی چگالی مخلوط آسفالتی
اگر درصد عبوری الک شماره ۴ بیشتر از ۴/۷۵٪ باشد و درصد فضای خالی مخلوط آسفالتی از ۵/۵٪ کمتر باشد و درصد عبوری الک شماره ۲۰۰ از ۱/۵٪ بیشتر باشد و درصد فضای خالی مصالح سنگی کمتر از ۱۴/۸٪ باشد و مقاومت مارشال از ۱۲۵۵ کیلوگرم-نیرو بیشتر باشد، به طور متوسط چگالی مخلوط آسفالتی ۲/۳۳۹ پیش‌بینی می‌شود.	ب
اگر درصد عبوری الک شماره ۴ بیش از ۴/۷۵٪ باشد و چگالی مخلوط آسفالتی از ۲/۳۰۱ بیشتر و درصد فضای خالی مصالح سنگی از ۱۴/۶۵٪ بیشتر باشد و مقاومت مارشال از ۱۲۵۵ کیلوگرم-نیرو کمتر باشد و درصد عبوری الک شماره ۲۰۰ از ۱/۵٪ کمتر باشد و درصد قیر در مخلوط آسفالتی از ۲/۵۲۵ بیشتر باشد، نتیجه‌ی پیش‌بینی درصد فضای خالی مخلوط آسفالتی به طور متوسط ۴/۰۹٪ خواهد شد.	پیش‌بینی درصد فضای خالی مخلوط آسفالتی

۴-۲- توسعه مدل رگرسیون خطی چندگانه

مقدار منفی آن وجود رابطه خطی معکوس، مقدار مثبت آن وجود رابطه خطی مستقیم و مقدار صفر آن عدم وجود رابطه بین دو متغیر را نشان می‌دهد که نتایج آن در جدول (۵) شده‌است. همچنین رابطه‌ی (۱۰) مدل رگرسیون خطی چندگانه حاصل را نمایش می‌دهد.

با استفاده از تحلیل آماری با نرم افزار SPSS می‌توان مدل رگرسیون خطی و میزان همبستگی متغیرهای تحقیق را به دست آورد. ابزاری آماری برای تعیین نوع و درجه رابطه یک متغیر کمی با متغیر کمی دیگر، ضریب همبستگی است. ضریب همبستگی شدت رابطه و همچنین نوع رابطه (مستقیم یا معکوس) بین دو متغیر در بازه (۱-۱) را نشان می‌دهد.

$$Density = 2.601 + 7.794 \times 10^{-7} \text{ Marshal} - 0.001 \text{ Flow} + 0.02 \text{ BitumenPercent} - 0.001 \text{ Void} - 0.25 \text{ VMAAggMater} + 0.001 \text{ P200Agg} \quad (10)$$

$$R^2 = 0.999$$

جدول ۵. نتایج اعتبارسنجی رگرسیون خطی درونیابی به کاررفته

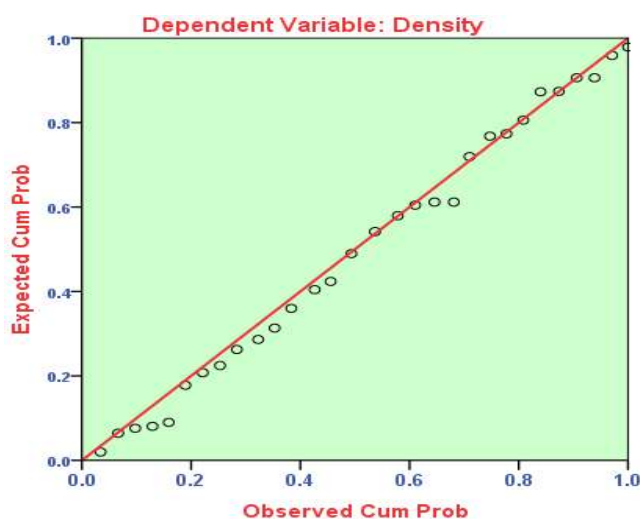
مدل	Sum of Squares	df	Mean Square	آماره‌ی F	سطح معناداری Sig.
Regression	11.353	7	1.622	4787494.559	.000
Residual	.015	43310	.000		
Total	11.367	43317			

خلاصه‌ای از آماره‌های مدل‌سازی Model Summary				
مدل	R	R square	Adjusted R Square	Std. Error of the Estimate
1	.999	.999	.999	.00058

ضرایب همبستگی بین متغیرها Correlations								
	P4Agg	P200Agg	VMAAggMater	Void	BitumenPercent	Flow	Marshal	Density
Pearson Correlation	.143	-.263	-.951	-.785	-.171	.465	.165	1
Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	
N	43318	43318	43318	43318	43318	43318	43318	43318



شکل ۸. مدل گرافیکی مقایسه داده‌های چگالی مخلوط آسفالتی حاصل از مدل درونیابی خطی چندگانه و نتایج آزمایشگاهی

از داده‌های پایش شده می‌باشد، که در صورت منظورکردن دامنه تغییرات داده‌های ورودی قادر است با ضریب اطمینان بالایی میزان تراکم مخلوط آسفالتی را برآورد نماید. هدف اصلی این مطالعه بررسی سودمندی استفاده از روش‌های آماری و تکنیک‌های داده‌کاوی در ثبت داده‌های آزمایشگاهی و میدانی است. هرچند بسیاری از متغیرهای موثر بر تراکم همانگونه که در نتایج تحقیقات قبلی نیز شناسایی شده، لیکن اولویت‌بندی و یافتن رابطه پیچیده‌ی بین آنها به عنوان مهمترین هدف این مطالعه انجام گرفت. یکی از مواردی که می‌تواند به این تحقیق و تحقیقات آتی در زمینه کاربردی داده‌های آزمایشگاهی و کنترل کیفیت اجرایی در داده‌کاوی کمک کند، ایجاد پرونده‌های الکترونیکی ساماندهی شده از این داده‌ها در آزمایشگاه‌های استاندارد و در ادارات با نقش کارفرمایی نظیر شهرداری و راه و شهرسازی و سازمان‌های مرتبط است و با دسترسی به این داده‌ها می‌توان با دقت بالاتری به منظور کنترل کیفیت و کشف دانش‌های جدید دست یافت. مدلسازی حاضر برای داده‌هایی از پروژه‌های انجام گرفته که شرایط ایده‌آل و مورد پذیرش از دیدگاه طرح اختلاط و عملیات اجرایی داشتند، ولی در موارد و پروژه‌های مشابه دارای اشکالات اجرایی و فنی و کیفی می‌توان گستره‌ای از داده‌ها را فراهم آورد تا نتایج مدلسازی دقیق‌تر و انطباق بیشتری با شرایط واقعی داشته باشند.

#### ۶- سپاسگزاری

بدین‌وسیله نویسندگان مراتب قدردانی خود را از مسئولین معاونت پژوهشی و مرکز تحقیقات حمل و نقل انسان محور و شرکت آسفالت‌سازان شیراز به دلیل حمایت از پژوهش و آزمایشگاه ژئوتکنیک و مقاومت مصالح وابسته به استان فارس به دلیل همکاری‌های اجرایی و گردآوری داده‌ها به عمل می‌آورند.

ضمن اینکه مدل خطی درونیابی برای کل داده‌های چگالی، در سطح دقت معادل ۹۵ درصد دارد. اهمیت متغیرها بکاررفته در رابطه‌ی (۱۰) در این روش در قالب متغیرهای مقاومت مارشال، درصد فضای خالی مصالح سنگی و دانه‌بندی آن (درصد عبوری الک شماره‌ی ۲۰۰) قابل ارائه است.

#### ۵- نتیجه‌گیری

تکنیک‌های روش‌های داده‌کاوی در سال‌های اخیر در حوزه‌های مختلفی از مهندسی از جمله مهندسی عمران، به منزله‌ی دانش جدید در حال رشد و استفاده گسترده است، به گونه‌ای که استخراج نتایج کاربردی و توسعه مدل‌هایی از انبوه داده‌ها خود زمینه‌ای روبه رشد در دانش مهندسی است. در زمینه تشخیص و پیشگیری اشکالات فنی و پیش‌بینی نکاتی که به دقت بالاتری در طرح اختلاط و اجراء مخلوط آسفالتی به طور گسترده‌ای مورد استفاده قرار گرفته است. در این تحقیق با استفاده از تکنیک‌های مختلف داده‌کاوی و SPSS درخت تصمیم در نرم افزار آماری و داده‌کاوی SPSS مدل‌هایی به منظور و پیش‌بینی و دسته‌بندی متغیرهای موثر در تراکم مخلوط آسفالتی ساخته شد. بهترین مدل ساخته شده، مربوط به پیش‌بینی فضای خالی مخلوط آسفالتی با در نظر گرفتن تعداد متغیرهای بیشتر و با تعداد گره و شاخه بیشتر می‌باشد. در تمام مدل‌سازی‌های صورت گرفته، پارامتر فضای خالی مصالح سنگی و درصد عبوری الکها (دانه‌بندی) دخیل بوده است. این موضوع نشان دهنده اهمیت این پارامترها در پیش‌بینی تراکم مخلوط آسفالتی است. با توجه به تحلیل‌های تفصیلی انجام‌شده و نتایج موجود در منابع علمی موجود محرز است درخت تصمیم رگرسیونی CART، مدلی نسبتاً کارآ در پیش‌بینی میزان تراکم و متغیرهای اصلی آن (چگالی- درصد فضای خالی) با استفاده

#### ۷- پی‌نوشت‌ها

1. Decision Tree
2. Data Mining
3. Global Position System (GPS)
4. Moving Average
5. Chain Average
6. Classification and Regression Trees (CART)
7. Least Significant Difference (LSD)

- Computers, 35(2), pp.487-498. <https://doi.org/10.1007/s00366-018-0611-9>.
- Kassem, E., Scullion, T., Masad, E., and Chowdhury, A., (2012), "Comprehensive evaluation of compaction of asphalt pavements and a practical approach for density predictions". *Transportation Research Record*, 2268(1), pp.98-107, <https://doi.org/10.3141/2268-12>.
- Khabiri, M.M. and Elahizadeh, M., (2015), "Evaluation of the effect of pavement failure on the probability of accidents using decision tree method (case study: pavement unevenness)", *Transportation Engineering Journal*, Vol. 7, Issue 4, summer, pp. 579-605.
- Liu, P., Wang, D., Otto, F., Hu, J., and Oeser, M. 2018. "Application of semi-analytical finite element method to evaluate asphalt pavement bearing capacity". *International Journal of Pavement Engineering*, 19(6), pp. 479-488.
- Moslemi Najkarlaei, F., Meysam Efati M., Farzin Naseri F., Rajabi A., (2015), "Prediction of travel time in outbound routes using spatial data mining techniques Case Study: Ghaemshahr Route to Babylon and Sari to Qaemshahr", *Scientific and Research Journal of Soft Computing and Information Technology JDL* 4, Issue 3, Vol. 43, pp.3-15.
- Mozafari, G. Omidavr, K. 2015, "Evaluation of the efficiency of regression decision tree model in drought prediction Case Study: Sanandaj synoptic station", *Environmental hazards journal*, year 4, issue 6, winter, pp.1-19.
- Mousa, M., Elseifi, M. A., and Abdel-Khalek, A. 2019. "Development of Tree-Based Algorithm for Prediction of Field Performance of Asphalt Concrete Overlays". *Journal of Transportation Engineering, Part B: Pavements*, 145(2), 04019011. <https://scelibrary.org/doi/abs/10.1061/JPEODX.000112>.
- Nath, D. R., Lee, H., Chowdhury, N., Chang, J., (2010), "Modified K-Means Clustering for Travel Time -Prediction Based on Historical Traffic Data", *Proceedings of the 14th international conference on Knowledge-based and intelligent information and engineering systems*, pp. 524-544.
- Pakgohara, R., and Sadeghikia, A., (2008), "Analysis of Statistical Data on Driving Accidents by Decision tree", *Journal of Traffic Management Studies*, Third Issue, Vol.3, No. 8, spring 2008, pp. 27-57.
- Zhao, S., & Al-Qadi, I. L., (2019), "Algorithm development for real-time thin asphalt concrete overlay compaction monitoring using ground-penetrating radar". *NDT & E International*, Volume 104, pp. 114-123. <https://doi.org/10.1016/j.ndteint.2019.04.008>.
- Breiman, L., Friedman J., Olshen R., and Stone. C., (1984), "Classification and Regression Trees", Chapman & Hall/CRC Press, Boca Raton, FL. Development of a decision tree modeling approach, *Geoderma* 139, pp.227-287.
- Chang, L. Y., & Chen, and W. C., (2005), "Data mining of tree-based models to analyze freeway accident frequency". *Journal of safety research*, 36(4), pp.365-375. <https://doi.org/10.1016/j.jsr.2005.06.013>
- Chowdhury, Ch., Deb Nath, R., Lee, H., and Chang, J., (2009), "Development of an Effective Travel Time -Prediction Method Using Modified Moving Average Approach", *Proceedings of the 13th International conference on Knowledge-Based and Intelligent Information and Engineering systems*, pp. 430-437.
- Cook, M. D., Ghaeezadah, A., and Ley, M. T., (2017), "Impacts of coarse-aggregate gradation on the workability of slip-formed concrete". *Journal of Materials in Civil Engineering*, 30(2), 04017265. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0002126](https://doi.org/10.1061/(ASCE)MT.1943-5533.0002126).
- Dasturani, M. T., Habibipour, A. Proprietary M. R. Talabi A.A., and Johari J., (2011), "Evaluation of the efficiency of decision tree model in prediction of precipitation (Case study of cytopathic station in Yazd)", *International Journal of Research*, Vol. 8, No. 3, pp. 14-23.
- Ismail, S., Hassan, N. A., Yaacob, H., Hainin, M. R., Ismail, C. R., Mohamed, A., and Satar, M. K. I. M., (2019), "Properties of dense-graded asphalt mixture compacted at different temperatures". In *IOP Conference Series: Earth and Environmental Science*, Vol. 220, No. 1, pp. 120-130. <https://doi.org/10.1088/1755-1315/220/1/012010/meta>.
- Gao, Y., Huang, X., and Yu, W. 2014. "The compaction characteristics of hot mixed asphalt mixtures". *Journal of Wuhan University of Technology-Mater. Sci. Ed.*, 29(5), pp.956-959. <https://doi.org/10.1007/s11595-014-1027-z>.
- Hoang, N. D., and Nguyen, Q. L., (2019), "A novel method for asphalt pavement crack classification based on image processing and machine learning". *Engineering with*

# Investigating the Effects of Mixing Design Variables on Performance of Asphalt Compaction Using Data Mining Algorithms

*Mohamad Hamidinia, M.Sc., Student, Engineering Faculty, Yazd University, Yazd, Iran.*

*Mohammad Mehdi Khabiri, Associate Professor, Engineering Faculty, Yazd University, Yazd, Iran.*

*Mehdi Mokhberi, Department of Civil Engineering, Estahban Branch, Islamic Azad University, Estahban, Iran.*

*E-mail: Email: mkhabiri@yazd.ac.ir*

Received: June 2020-Accepted: November 2020

## **ABSTRACT**

Many factors or parameters influence the field density of asphalt mixes. Since the number of variables affecting a large and somewhat different compacts affects each other, it is almost impossible to determine a constant interpolation relationship. Data mining and its techniques are a way of discovering hidden knowledge between dependent and independent variables, as well as indirect and nonlinear relationships can be identified by dividing the data into groups or leaves in the decision tree method. In this study, commonly used data mining techniques in civil engineering, including the neural network, logistic regression and decision tree. By emphasizing the application of decision tree method, with the aim of exploring knowledge models and providing predictions, other data mining tools are used to assist in constructing and evaluating the developed statistical model. The explanatory variables used in the three models of this study are the percentage of void, asphalt mix strength, aggregate size, bitumen percentage, asphalt mix flow. The results show that the percentage of void content of stone materials, the percentage of passage of sieve 200 and 4, and the bitumen percentage had a greater effect on the density and compaction of asphalt mixture. Also, a multiple linear regression model with a correlation coefficient of nearly one between the density of asphalt mixture and variables were presented.

**Keywords:** Mixing Design, Asphalt Mix, Data Mining, Concentration Quality Control, Pavement Failure, Decision Tree