علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۲۰، شماره ۲، ۱۴۰۱
صفحات ۳۹-۵۱
مقاله پژوهشی

به‌کارگیری منطق فازی در پیش‌بینی شدت حوادث صنعت ساخت‌وساز

علی رحیم‌زادگان[۱]، منصور قلعه‌نوی[۲]*، علی‌اکبر سالاریان [۱]

*نویسنده مسئول، دریافت: ۰۰/۰۰/۰۰، بازنگری: ۰۰/۰۰/۰۰، پذیرش: ۰۰/۰۰/۰۰

[۱] دانشجوی کارشناسی ارشد، گروه مهندسی عمران، دانشکده مهندسی، دانشگاه فردوسی، مشهد، ایران
[۲] استاد، گروه مهندسی عمران، دانشکده مهندسی، دانشگاه فردوسی، مشهد، ایران

چکیده

ایمنی، همواره دغدغه‌ای جدی در کارگاه‌های ساخت‌وساز می‌باشد. در این راستا، مطالعه بیشتر شرایط ناایمن و حوادث پیشین می‌تواند به شناسایی خطرات آتی و اصلاح وضعیت موجود کمک کند. هدف این پژوهش، مدل‌سازی شدت حوادث صنعت ساخت‌وساز با منطق فازی می‌باشد. با کمک این مدل می‌توان شدت حوادث آینده را بر اساس اطلاعات موجود در حال حاضر پیش‌بینی کرد. ابتدا ورودی‌های مدل با توابع عضویت مختلف، به چهار حالت متفاوت تقسیم می‌گردند. سپس با استفاده از جدول جستجوی فازی، پایگاه قواعد برای هرکدام از حالات ایجاد می‌شود. در گام بعد، سیستم استنتاج فازی به‌منظور انجام پیش‌بینی، پیاده‌سازی می‌گردد. در پایان، عملکرد بهترین مدل ارائه شده با چند روش یادگیری ماشین رایج، مقایسه می‌گردد. یافته‌ها نشان می‌دهد که برخلاف سایر روش‌ها، متوازن‌سازی مجموعه‌داده تأثیر ناچیزی بر کاهش عملکرد روش فازی دارد و به برتری نتایج این روش منجر می‌شود. مقدار Accuracy و F1_score برای مدل فازی بهینه ۸۳/۴۳ و ۷۹/۶۹ درصد به‌دست می‌آید که حداقل ۱۹/۹۲ و ۱۰/۶۹ درصد نسبت به الگوریتم‌های دیگر، بیشتر است. این برتری برای مقادیر مذکور حداکثر به ۳۴/۹۶ و ۲۱/۵۳ درصد می‌رسد. توسعه ابزاری برای پیش‌بینی نتیجه حوادث، به متخصصان ایمنی کمک می‌کند تا اقدامات لازم برای پیشگیری از خسارات محتمل را برنامه‌ریزی کنند.

کلمات کلیدی: سیستم استنتاج فازی، جدول ارجاع فازی، یادگیری ماشین، حوادث ساختمانی، پیش‌بینی حوادث، صنعت ساختمان، مدیریت ایمنی

۱- مقدمه

هر روز افراد بسیاری در سراسر دنیا، حوادث شغلی را تجربه می‌کنند. بر اساس تخمین سازمان جهانی کار[۱]، سالانه حدود ۳۶۰ میلیون حادثه شغلی رخ می‌دهد که منجر به بیش از چهار روز دوری از کار می‌گردد و دست‌کم ۱/۹ میلیون نفر به علت این حوادث، جان خود را از دست می‌دهند [۱]. در ایران نیز در شش ماه نخست سال ۱۳۹۸، ۸۹۸ نفر قربانی حوادث ناشی از کار شده‌اند [۲] که این عدد در مقایسه با شش ماه نخست سال ۱۳۹۴، ۴۷ نفر افزایش‌یافته است [۳]. در بین صنایع مختلف، صنعت ساخت‌وساز به‌عنوان یکی از خطرناک‌ترین آن‌ها در بسیاری از نقاط دنیا شناخته می‌شود [۵،۴]. در کشور ما نیز ۴۰ تا ۵۰ درصد حوادث شغلی سال‌های اخیر در کارگاه‌های ساخت‌وساز رخ‌داده‌اند؛ درحالی‌که این مناطق، محل کار ۱۲ تا ۱۳ درصد شاغلان بوده‌اند [۲]. این مسئله کارگاه‌های ساخت‌وساز را در زمره خطرناک‌ترین مناطق از نظر وقوع آسیب‌های شغلی قرار می‌دهد. وقوع این‌گونه حوادث علاوه بر تحمیل بار مالی اضافه به پروژه و افزایش زمان اتمام پروژه، تأثیراتی همچون آسیب به خانواده فرد حادثه‌دیده، آسیب

به اعتبار مسئولان اجرایی پروژه و تحمیل هزینه به دولت را به همراه خواهند داشت. همچنین ماهیت حوادث به‌گونه‌ای است که عامل‌های متعددی در نحوه وقوع آن‌ها اثر دارند و اطلاعات موجود از حوادث دارای حجم زیادی هستند. این موارد، تحلیل حوادث و جلوگیری از آن‌ها را به یک مسئله پیچیده تبدیل می‌کند. با درنظرگرفتن سرمایه انسانی و خسارات فراوان ناشی از حوادث، نیاز به مطالعه بیشتر در زمینه ایمنی ساخت‌وساز احساس می‌گردد. درصورتی‌که این بررسی در زمینه پیش‌بینی حوادث آتی باشد، به مدیران ایمنی این امکان را می‌دهد تا در راستای جلوگیری از حوادث، حرکت کنند.

در رابطه با پیش‌بینی و مدل‌سازی حوادث، تلاش‌های زیادی صورت گرفته است. این مدل‌ها می‌توانند بر مبنای یادگیری آماری یا یادگیری ماشین توسعه یابند که با توجه به تعداد زیاد داده‌های در دسترس، مدل‌های مبتنی بر یادگیری ماشین عملکرد بهتری در انجام پیش‌بینی دارند [۶]. برای این منظور، الگوریتم‌های یادگیری ماشین مختلفی، به‌کاررفته است. برای مثال تیکسیر و همکاران [۷] از جنگل تصادفی [۲] و ارتقای گرادیان تصادفی [۳] بهره گرفتند تا نوع آسیب واردشده، قسمت آسیب‌دیده در بدن، منشأ حادثه از نظر نوع انرژی و شدت حادثه را پیش‌بینی کنند. نویسندگان، پارامترهای مدل‌ها را برای هرکدام از چهار خروجی، بهینه کردند. آنها دریافتند که به‌طور متوسط، ارتقای گرادیان تصادفی نسبت به جنگل تصادفی، عملکرد بهتری دارد. پو و همکاران [۸] مدلی برای پیش‌بینی وقوع حادثه و شدت آن توسعه دادند. آن‌ها از یک پیمانکار سنگاپوری، مجموعه‌داده مورداستفاده را جمع‌آوری کردند و سپس از الگوریتم‌های رگرسیون لجستیک [۴]، ماشین بردار پشتیبان [۵]، کی - نزدیک‌ترین همسایه [۶]، درخت تصمیم [۷] و جنگل تصادفی استفاده کردند. در مطالعه مذکور، جنگل تصادفی به‌عنوان بهترین مدل شناخته شد. ایهان و توکدمیر [۹] ایمنی در پروژه‌های ساخت‌وساز را بررسی کردند. نویسندگان برای کاهش ناهمگونی [۸] داده‌ها و افزایش دقت مدل از تجزیه‌وتحلیل خوشه‌بندی کلاس پنهان [۹]، استفاده کردند. سپس با شبکه عصبی مصنوعی [۱۰]، مدلی برای پیش‌بینی شدت حوادث ارائه دادند. مطالعه آن‌ها با ارائه اقدامات پیشگیرانه لازم، با توجه به نتایج پیش‌بینی‌شده به پایان رسید. چوی و همکاران [۱۰] با الگوریتم‌های تقویت تطبیقی [۱۱]، درخت تصمیم، جنگل تصادفی و رگرسیون لجستیک، خطر بالقوه تصادفات مرگبار در کارگاه‌های ساخت‌وساز را شناسایی کردند. آن‌ها سپس متغیرهای اصلی هر الگوریتم را برای انجام پیش‌بینی مشخص کردند. بیکر و همکاران [۱۱] نوع حادثه، نوع آسیب، شدت آسیب و قسمت آسیب‌دیده در بدن را با رویکرد یادگیری ماشین، مدل‌سازی کردند. آن‌ها با پردازش زبان طبیعی [۱۲]، اطلاعات حوادث را از گزارش‌ها استخراج کردند. سپس الگوریتم‌های ارتقای گرادیان شدید [۱۳]، ماشین بردار پشتیبان و درخت تصمیم را برای انجام پیش‌بینی، آموزش دادند. ژو و همکاران [۱۲] از هشت روش برای پیش‌بینی شدت حوادث ساخت‌وساز، استفاده کردند. این روش‌ها: رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، ماشین بردار پشتیبان، پرسپترون چندلایه [۱۴]، بیز ساده [۱۵]، کی - نزدیک‌ترین همسایه، یادگیری ماشین خودکار [۱۶] بود. آنها دریافتند که یادگیری ماشین خودکار قادر است بیش از ۸۴ درصد حوادث را درست طبقه‌بندی کند. جورج و همکاران [۱۳] اطلاعات حوادث صنعت ساخت‌وساز که سازمان ایمنی و بهداشت شغلی [۱۷] گزارش کرده است را بررسی کردند. نویسندگان ابتدا ویژگی‌های ورودی مهم را شناسایی و سپس توسط الگوریتم‌های درخت تصمیم و جنگل تصادفی، مرگبار بودن یا نبودن حوادث را پیش‌بینی کردند. آن‌ها موفقیت این دو مدل را با هشت معیار ارزیابی، مقایسه کردند.

مشکلی که هنگام استفاده از روش‌های یادگیری ماشین برای مسائل طبقه‌بندی رایج می‌باشد، وجود نامتوازنی در داده‌هاست. نامتوازنی در داده‌ها، بدین معنی است که تعداد نمونه‌ها در کلاس‌های متغیر هدف، تفاوت زیادی دارند و نمونه‌ها در کلاس اکثریت [۱۸] (کلاس دارای بیشترین تعداد نمونه) چندین برابر کلاس اقلیت [۱۹] (کلاس دارای کمترین تعداد نمونه) است. در یادگیری داده‌های نامتوازن، عموم الگوریتم‌ها نسبت به طبقه‌بندی کلاس اکثریت، جهت‌گیری دارند؛ درحالی‌که ممکن است طبقه‌بندی نمونه‌های کلاس اقلیت به‌اشتباه انجام شود [۱۶-۱۴]. به‌عبارت‌دیگر با توجه به اینکه این الگوریتم‌ها، طبقه‌بندی را طوری تنظیم می‌کنند که دقت افزایش یابد، کلاس‌های با تعداد نمونه کم نادیده گرفته می‌شوند و امکان دارد دقت به‌صورت کاذب، بالا گزارش شود. برای رفع این مشکل، محققان روش‌های بسیاری را به‌کار گرفته‌اند. برای مثال در زمینه ایمنی ساخت‌وساز پو و همکاران [۸] از بیش نمونه‌گیری [۲۰]، کنگ و ریو [۱۷] از کم نمونه‌گیری [۲۱]، چوی و همکاران [۱۰] از نمونه‌گیری تصادفی [۲۲] و بیکر و همکاران [۱۱] از وزن‌دهی به کلاس‌ها استفاده کردند. پس از رفع مشکل نامتوازنی، اهمیت همه کلاس‌ها برای الگوریتم یکسان می‌شود و جهت‌گیری در پیش‌بینی از بین می‌رود. در نتیجه، امکان سنجش عملکرد مدل فراهم می‌گردد.

در میان روش‌های یادگیری ماشین، به‌کارگیری الگوریتم‌هایی ارجحیت دارند که پس از فرایند آموزش، دانش به‌دست‌آمده را برای انسان ارائه می‌کنند و منجر به نتایج دقیق‌تر می‌شوند زیرا متخصصان ایمنی می‌توانند دانش فردی و دانش به‌دست‌آمده از ماشین را با یکدیگر ترکیب کنند و به‌طور مؤثرتری برای مقابله با وقوع حوادث آماده گردند. یکی از این روش‌ها در سیستم‌های هوشمند [۲۳]، منطق فازی [۲۴] می‌باشد که توسعه‌ای بر منطق بولی است. با توجه به جعبه‌سیاه نبودن منطق فازی، می‌توان این روش را در مسائلی که نیاز به تعامل بین انسان و ماشین است، به‌خوبی به‌کار گرفت. منطق فازی، سابقه‌ای طولانی در رشته‌های متعدد دارد اما کاربرد آن در زمینه مهندسی ساخت‌وساز نسبتاً جدید است [۱۸]. این تازگی در حوزه ایمنی ساخت‌وساز نیز وجود دارد. ازجمله مطالعاتی که در این حوزه از منطق فازی استفاده کرده‌اند، می‌توان به پژوهش گورکانلی و مونگن [۱۹] اشاره کرد. نویسندگان با استفاده از منطق فازی، به ارزیابی خطراتی پرداختند که کارگران در کارگاه‌های ساخت‌وساز در معرض آن‌ها هستند. ورودی‌های سیستم استنتاج فازی [۲۵] آن‌ها سطح ایمنی فعلی، شدت حادثه، احتمال حادثه و خروجی آن سطح خطر موجود بود. جانا و همکاران [۲۰] با مجموعه‌های فازی نوع دو بازه‌ای [۲۶]، مدل ارائه‌شده توسط گورکانلی و مونگن را توسعه دادند. آن‌ها نشان دادند که با این کار، دقت افزایش می‌یابد. ویناندا و همکاران [۲۱] خطر احتمالی سقوط کارگران از ارتفاع را بررسی کردند. آن‌ها با کمک سیستم استنتاج فازی، مدل‌سازی را انجام دادند و برای اعتبارسنجی مدل، از نظر کارشناسان بهره بردند. ایهان و

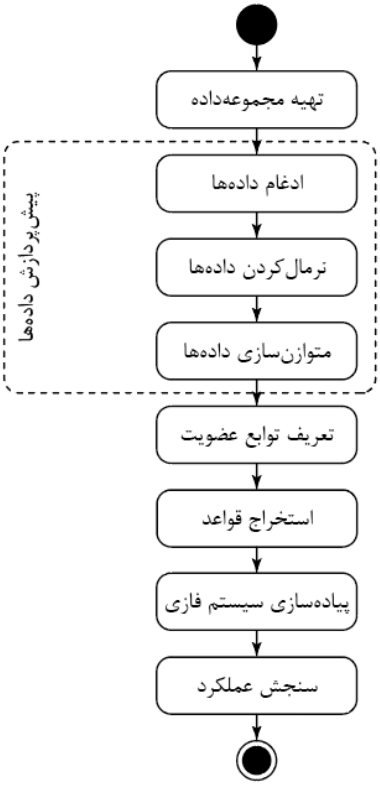
توکدمیر [۲۲] به‌منظور افزایش دقت پیش‌بینی از منطق فازی استفاده کردند. نویسندگان برای پیش‌بینی حوادث کارگاه‌های ساخت‌وساز، یک مدل مبتنی بر شبکه عصبی مصنوعی توسعه دادند. آن‌ها پس از انجام پیش‌بینی، با استفاده از ۳۵ قاعده فازی، نتایج به‌دست‌آمده را بهبود دادند.

در مطالعات ذکرشده در بالا، از منطق فازی به‌عنوان روشی برای بیان دانش انسانی و مدل‌سازی ریاضی آن استفاده شده است. این رویکرد درعین‌حال که مزایای زیادی دارد، به‌کارگیری منطق فازی را محدود می‌کند زیرا دانش انسانی برای تمامی مسائل وجود ندارد و در اختیارداشتن آن برای مسائلی که تعداد متغیرهای زیادی دارند یا دارای حجم زیادی از اطلاعات هستند، عملاً غیرممکن است. بنابراین هدف این مطالعه، به‌کارگیری منطق فازی به‌گونه‌ای است که دانش لازم برای پیش‌بینی شدت حوادث صنعت ساخت‌وساز توسط ماشین به‌دست بیاید. در این راستا، از روش جدول جستجوی فازی[۲۷] برای استخراج قواعد بر اساس اطلاعات موجود از حوادث گذشته و به‌صورت خودکار استفاده می‌گردد. بدین ترتیب نیاز به تنظیم دستی قواعد نمی‌باشد و فرایند آموزش مدل توسط ماشین انجام می‌گیرد. علاوه بر این مسئله، دانش به‌دست‌آمده توسط ماشین نیز در قالب قواعد فازی برای متخصصان ایمنی تفسیرپذیر و قابل‌ارائه است. شایان ذکر است که در حوزه پیش‌بینی حوادث، تا کنون مطالعه‌ای از روش جدول جستجوی فازی استفاده نکرده است. در این مطالعه، ابتدا داده‌ها متوازن می‌گردند. سپس فضای مسئله با استفاده از توابع عضویت به چهار صورت مختلف تقسیم‌بندی می‌شود. در گام بعد، جدول جستجوی فازی، پیاده‌سازی می‌گردد. پایگاه قواعد در حالت توازن و عدم توازن داده‌ها و برای تقسیم‌بندی‌های متفاوت فضای مسئله تنظیم می‌گردد. سپس سیستم استنتاج فازی با کمک پایگاه قواعد پیش‌بینی را انجام می‌دهد. برای سنجش عملکرد مدل فازی، الگوریتم‌های ریپر[۲۸]، درخت تصمیم، جنگل تصادفی، ماشین بردار پشتیبان و شبکه عصبی مصنوعی پیاده‌سازی می‌گردند. با مقایسه نتایج، پیش و پس از متوازن‌سازی داده‌ها میزان جهت‌گیری هر یک از مدل‌ها در حالت نامتوازنی داده‌ها مشخص می‌گردد. بر اساس اطلاعات نویسندگان، مقدار جهت‌گیری روش فازی در صورت استفاده از جدول جستجو در مطالعات سایر حوزه‌ها نیز بررسی نشده است و نخستین‌بار در این مطالعه موردتوجه قرار می‌گیرد. همچنین، با مقایسه حالات مختلف تقسیم‌بندی ورودی‌ها و خروجی، بهینه‌ترین مدل فازی تعیین می‌شود. در پایان، این مطالعه ابزاری را برای متخصصان ایمنی فراهم می‌آورد تا با کمک آن بتوانند شدت حوادث را پیش‌بینی کنند. با این رویکرد، به‌جای تمرکز بر روی واکنش دادن به حادثه هنگام وقوع، می‌توان اقدامات پیشگیرانه‌ی لازم را تهیه کرد و خسارات ناشی از حوادث را کاهش داد.

ساختار کلی مطالعه متشکل از چهار بخش می‌باشد که علاوه بر بخش مقدمه شامل مواد و روش‌ها، نتایج و بحث و نتیجه‌گیری است.

۲- مواد و روش‌ها

مراحل کلی طی‌شده در این بخش، در شکل ۱ نشان داده می‌شوند. ابتدا مجموعه‌داده به‌کاررفته، شناسایی می‌گردد و پیش‌پردازش‌های لازم بر روی آن انجام می‌شوند تا ورودی‌های مدل آماده گردند. در مرحله بعد، توابع عضویت فازی برای تقسیم‌بندی فضای مسئله تعریف می‌شود. سپس پایگاه قواعد با روش جدول جستجوی فازی تنظیم می‌گردد. در نهایت سیستم استنتاج فازی، فرایند پیش‌بینی را انجام می‌دهد و عملکرد آن ارزیابی می‌شود. در ادامه، توضیحات مربوط به هر بخش ارائه می‌گردد.



شکل ۱. مراحل روش تحقیق

۲-۱- معرفی مجموعه‌داده به‌کاررفته

در پژوهش حاضر، از مجموعه‌داده گردآوری‌شده در مطالعه ایهان و توکدمیر [۲۳] استفاده می‌شود. این مجموعه‌داده، به‌صورت برخط در دسترس عموم قرار گرفته است و دارای اطلاعات ۵۲۲۴ مورد از حوادث ساخت‌وساز می‌باشد که این حوادث در پروژه‌های عظیم[۲۹] منطقه اوراسیا رخداده‌اند. پروژه‌های عظیم، پروژه‌هایی هستند که معمولاً بیش از یک میلیارد دلار هزینه دارند و دارای ذی‌نفعان متعدد و زمان توسعه و ساخت طولانی هستند و میلیون‌ها نفر را تحت تأثیر خود قرار می‌دهند [۲۴]. برای توصیف شرایط وقوع حوادث در مجموعه‌داده، تعدادی ویژگی تعیین شده است که انتخاب این ویژگی‌ها، بر مبنای مطالعه دیگری از ایهان و توکدمیر [۲۲] و نظر کارشناسان بهداشت و ایمنی شغلی انجام گرفته است. در مطالعه حاضر برای مدل‌سازی شدت حوادث، از تجربه به همراه ۳۵ ویژگی دیگر، به‌عنوان اطلاعات حوادث استفاده می‌شود. این ۳۵ ویژگی در چهار دسته تقسیم‌بندی شده‌اند که این دسته‌ها عبارتند از: عوامل انسانی، عوامل محیط کار، رفتارهای پرخطر و موارد خطرناک. در ادامه تعدادی از ویژگی‌های مرتبط با هر دسته، بیان می‌شوند. دسته عوامل انسانی شامل مواردی مانند ناتوانی جسمی، ناتوانی روانی و ناتوانی در ارزیابی خطر موجود است. دسته عوامل محیط کار شامل مواردی مانند اقدامات حفاظتی نامناسب، سیستم‌های تحلیل حوادث ناکارآمد و ندیدن آموزش ایمنی و بهداشت شغلی می‌باشد. دسته رفتارهای پرخطر شامل مواردی مانند

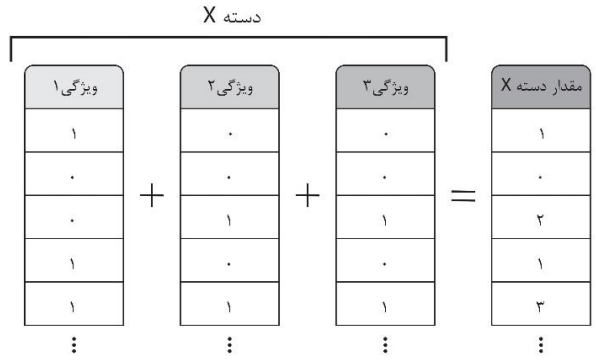
استفاده نادرست از تجهیزات و ابزارها، حرکت فیزیکی غلط و تخطی از خط‌مشی ایمنی است و دسته موارد خطرناک شامل مواردی مانند آتش‌سوزی و انفجار، خطرات شیمیایی و خطرات طبیعی می‌باشد. متغیر هدف نیز شدت حادثه می‌باشد که شش کلاس دارد. این کلاس‌ها و توزیع نمونه‌ها در آنان بدین صورت می‌باشد: ۲۴۴ مورد رفتارهای محتمل به ایجاد حادثه (کلاس یک)، ۱۱۶ مورد حوادث منجر به آسیب به مواد و مصالح (کلاس دو)، ۳۰۷۰ مورد حوادث نیازمند کمک‌های اولیه (کلاس سه)، ۱۳۹۳ مورد حوادث نیازمند به مداخله پزشکی (کلاس چهار)، ۳۹۷ مورد حوادث منجر به ازکارافتادگی (کلاس پنج) و ۴ مورد حوادث مرگبار (کلاس شش). خوانندگان برای مشاهده مشخصات کامل مجموعه‌داده می‌توانند به [۲۳] مراجعه کنند.

۲-۲- پیش‌پردازش داده‌ها

در این بخش، ابتدا موارد بدون مقدار در مجموعه‌داده شناسایی می‌گردند. مقدار تجربه در چهار حادثه نامشخص است که این حوادث از مجموعه‌داده کنار گذاشته می‌شوند. سایر مراحل این بخش عبارتند از: ادغام داده‌ها، نرمال‌کردن داده‌ها و متوازن‌سازی داده‌ها که در ادامه توضیح داده می‌شوند.

۲-۲-۱- ادغام داده‌ها

طول قواعد فازی متناسب با تعداد ویژگی‌هایی می‌باشد که به‌عنوان ورودی به‌کار می‌روند. تعداد ویژگی‌های به‌کاررفته در این مطالعه ۳۶ عدد می‌باشد که به نسبت زیاد است؛ بنابراین قواعد شامل بخش‌های زیادی خواهند بود و تفسیرپذیری قواعد برای انسان دشوار می‌گردد. باتوجه‌به اینکه به‌جز تجربه سایر ویژگی‌ها در چهار دسته عوامل انسانی، عوامل محیط کار، رفتارهای پرخطر و موارد خطرناک تقسیم گردیده‌اند، ویژگی‌های موجود در هر دسته با یکدیگر ادغام می‌شوند تا طول قواعد کاهش پیدا کند. برای این منظور، در هر حادثه، مقدار عددی ویژگی‌های موجود در یک دسته با یکدیگر جمع می‌گردند و نتیجه به‌عنوان مقدار عددی آن دسته لحاظ می‌شود. شکل ۲ فرایند ادغام ویژگی‌ها را به‌صورت شماتیک نشان می‌دهد. شایان ذکر است که برای تجربه نیز که یک متغیر طبقه‌ای[۳۰] می‌باشد، مرکز بازه به‌عنوان مقدار متغیر لحاظ می‌شود. با این عمل پنج متغیر خواهیم داشت که با اعداد به‌دست‌آمده، مقداردهی می‌شوند.



شکل ۲. شماتیک فرایند ادغام ویژگی‌ها

در مجموعه‌داده حاضر، توزیع نمونه‌ها در کلاس‌های مختلف بسیار متفاوت است. برای مثال ۰/۰۸درصد از حوادث ثبت‌شده، حوادث مرگبار و ۲/۲۲ درصد از آنها حوادث منجر به آسیب به مواد و مصالح است. در نقطه مقابل، حوادث نیازمند کمک‌های اولیه، ۵۸/۷۵ درصد از کل حوادث می‌باشند. این پراکندگی در توزیع نمونه‌ها باعث می‌گردد که یادگیری کلاس‌های با تعداد نمونه کم، برای ماشین با سختی روبه‌رو شود. در نتیجه برای کاهش پراکندگی نمونه‌ها، شدت حوادث بدین ترتیب با یکدیگر ادغام می‌گردد: حوادث کلاس یک و دو به‌عنوان حوادث خفیف، حوادث کلاس سه و چهار به‌عنوان حوادث متوسط و حوادث کلاس پنج و شش به‌عنوان حوادث شدید درنظر گرفته می‌شوند.

۲-۲-۲- نرمال کردن داده‌ها

نرمال‌کردن داده‌ها پیش از انجام ارزیابی، موضوعی رایج در یادگیری ماشین می‌باشد. در بسیاری از مواقع ویژگی‌های ورودی مختلف، محدوده تغییرات متفاوتی دارند اما این تفاوت لزوماً به معنی اهمیت کمتر یا بیشتر در فرایند پیش‌بینی نیست. بنابراین نرمال‌کردن داده‌ها، جهت‌گیری نسبت به ویژگی‌های با مقدار عددی بزرگ‌تر را به حداقل می‌رساند و مشارکت عددی ویژگی‌های مختلف در فرایند طبقه‌بندی را یکسان می‌کند [۲۵]. در این پژوهش مقادیر متغیرها، با استفاده از رابطه (۱) نرمال می‌گردند.

$$x = \frac{x_{raw} - min\ (x_{raw})}{max(x_{raw}) - min\ (x_{raw})} \quad (۱)$$

که x_{raw} بردار مقادیر متغیر قبل از اصلاح و x بردار مقادیر نرمال‌شده آن‌ها است.

۲-۲-۳- متوازن‌سازی داده‌ها

با توجه به ماهیت حوادث در صنعت ساخت‌وساز، داده‌های حوادث غالباً با مسئله نامتوازنی همراه هستند زیرا احتمال وقوع حوادث مرگبار نسبت به حوادث منجر به آسیب، کم می‌باشد. همچنین احتمال گزارش‌نشدن و ثبت حوادث جزئی، زیاد است. مجموعه‌داده حاضر نیز مشکل نامتوازنی دارد. در این مجموعه‌داده، تعداد نمونه‌ها در کلاس اکثریت (حوادث متوسط) ۱۲/۳۹ برابر تعداد نمونه‌ها در کلاس اقلیت (حوادث خفیف) می‌باشد. در پژوهش حاضر برای رفع مشکل نامتوازنی از روش بیش نمونه‌گیری اقلیت مصنوعی[۳۱] بهره برده می‌شود. این روش را چاولا و همکاران [۲۶] ارائه کرده‌اند و یکی از پرکاربردترین روش‌های متوازن‌سازی می‌باشد. هدف این روش، تولید نمونه‌های مصنوعی جدید با توجه به داده‌های موجود در یک کلاس می‌باشد تا بدین‌وسیله تعداد نمونه‌های آن کلاس افزایش یابد. در این پژوهش، ابتدا ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد برای آزمون درنظر گرفته می‌شود. سپس برای داده‌های آموزش نمونه‌های مصنوعی تولید می‌شود. فرایند متوازن‌سازی به سه صورت انجام می‌پذیرد. در یک حالت، نمونه‌های مصنوعی برای کلاس اقلیت تولید می‌گردند تا تعداد نمونه‌ها در کلاس اقلیت با کلاس اکثریت برابر شود. در این حالت کلاس میانی (کلاس دارای تعداد نمونه‌های بیشتر از اقلیت و کمتر از اکثریت) تغییری نمی‌یابد. در حالت دوم، بیش نمونه‌گیری برای کلاس میانی انجام می‌شود تا تعداد نمونه‌های آن به کلاس اکثریت برسد. در این حالت کلاس اقلیت بدون تغییر می‌ماند. در حالت سوم نیز با بیش نمونه‌گیری کلاس میانی و اقلیت، تعداد نمونه‌ها در هر سه کلاس یکسان می‌گردد. با این کار

داده‌های آموزش چهار وضعیت مختلف پیدا می‌کنند. این وضعیت‌ها عبارتند از: بدون نمونه مصنوعی (داده‌های خام)، دارای نمونه‌های مصنوعی برای کلاس میانی، دارای نمونه‌های مصنوعی برای کلاس اقلیت و دارای نمونه‌های مصنوعی برای کلاس‌های میانی و اقلیت.

۲-۳- تعریف توابع عضویت

پیش از پیاده‌سازی جدول جستجو و سیستم فازی، نیاز به تعیین توابع عضویت می‌باشد. این توابع به‌صورت‌های مختلفی قابل تعریف هستند. در اینجا توابع عضویت گوسی[۳۲] و زنگوله‌ای تعمیم‌یافته[۳۳]، مورداستفاده قرار می‌گیرند. فرم کلی این توابع در رابطه (۲) و (۳) بیان می‌گردد.

$$\mu_{gaussian}(x) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \quad (۲)$$

$$\mu_{gbell}(x) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \quad (۳)$$

که $\mu(x)$ بیانگر تابع عضویت و a ، b ، c و σ به‌ترتیب پارامتر کنترل‌کننده پهنا، پارامتر کنترل‌کننده شیب، مرکز تابع و انحراف از معیار آن است.

با کمک این توابع عضویت، فضای ورودی‌ها و خروجی مسئله تقسیم‌بندی می‌شوند. در اینجا توابع عضویت در چهار حالت مختلف تعریف می‌گردند: در حالت اول برای تقسیم هر متغیر ورودی از دو تابع عضویت گوسی استفاده می‌شود که بیانگر زیاد یا کم بودن متغیر است. خروجی نیز با سه تابع عضویت گوسی تقسیم می‌گردد که نشان‌دهنده حوادث خفیف، متوسط و شدید می‌باشند. در حالت دوم، تقسیم‌بندی ورودی‌ها با سه تابع عضویت گوسی انجام می‌گردد. این توابع نماینده کم، متوسط یا زیاد بودن ورودی هستند. تعداد توابع عضویت متغیر خروجی نیز برابر با سه می‌باشد که نشان‌دهنده‌ی حوادث خفیف، متوسط و شدید است. تعداد تقسیمات در حالت سوم و چهارم به ترتیب مشابه با حالت اول و دوم می‌باشد؛ با این تفاوت که در این جا از توابع عضویت زنگوله‌ای تعمیم‌یافته به‌جای گوسی استفاده می‌گردد.

۲-۴- استخراج قواعد

پایگاه قواعد اگر- آن‌گاه، نقشی کلیدی در یک سیستم فازی دارد زیرا با ترکیب قواعد می‌توان از ورودی‌ها به خروجی رسید و فرایند استنتاج را انجام داد. هر قاعده دارای دو بخش مقدم (اگر) و تالی (آن‌گاه) می‌باشد. بخش مقدم از چندین گزاره ساده فازی تشکیل می‌شود. هرکدام از این گزاره‌ها، وضعیت هر متغیر ورودی را نشان می‌دهند و با اتصال‌دهنده «و» با یکدیگر ترکیب می‌گردند. بخش تالی یک گزاره ساده فازی می‌باشد که وضعیت متغیر شدت حادثه را بیان می‌کند.

برای تشکیل پایگاه قواعد از روش جدول جستجو استفاده می‌شود. در این روش، قواعد بر اساس زوج‌های ورودی- خروجی تولید می‌گردند و دانش به‌دست‌آمده از اطلاعات در قالب قواعد بیان می‌شود. پیاده‌سازی این روش در سه گام انجام می‌پذیرد. در گام اول توابع عضویت به‌گونه‌ای تعریف می‌شوند که ورودی‌ها و خروجی را به‌طور کامل پوشش دهند. در گام دوم از هر داده که یک زوج ورودی- خروجی است یک قاعده تولید می‌گردد. بدین منظور، میزان تعلق هر متغیر به توابع عضویت تعریف‌شده، تعیین می‌گردد و هر تابع عضویت که دارای بیشترین میزان تعلق باشد، به آن متغیر نسبت داده می‌شود. در گام سوم قواعد پالایش می‌گردند. برای این کار، به هر قاعده تولیدشده، یک درجه نسبت داده می‌شود. این درجه برای حل ناسازگاری بین قواعد استفاده می‌گردد. از میان قواعد ناسازگار (قواعدی که بخش اگر یکسان و بخش آن‌گاه متفاوت دارند) فقط قاعده‌ای که بالاترین درجه را دارد، انتخاب می‌شود. در بین قواعد تکراری نیز فقط یکی از آن‌ها نگهداری می‌گردد. سه گام فوق برای توابع عضویت مختلفی که در بخش ۲-۳ توضیح داده شد، انجام می‌گیرد. توضیحات بیشتر در رابطه با این روش در [۲۷] ارائه شده است.

۲-۵- پیاده‌سازی سیستم فازی

سیستم فازی برای پیش‌بینی شدت حوادث نیاز به چهار قسمت دارد. این قسمت‌ها عبارتند از: فازی ساز[۳۴] ورودی، پایگاه قواعد فازی، موتور استنتاج فازی و غیرفازی ساز[۳۵] خروجی. در قسمت فازی ساز ورودی، فازی ساز منفرد[۳۶] به‌کار می‌رود. این فازی ساز، یک نقطه با مقدار حقیقی x^* را به یک مجموعه فازی می‌نگارد که مقدار تابع تعلق این مجموعه در نقطه x^* برابر یک و در نقاط دیگر برابر صفر است. پایگاه قواعد فازی از تعدادی قواعد اگر- آن‌گاه فازی تشکیل می‌شود. نحوه استخراج این قواعد توسط جدول جستجو در بخش ۲-۴ توضیح داده شد. موتور استنتاج فازی، قواعد اگر- آن‌گاه را ترکیب می‌کند و از مجموعه‌های فازی فضای ورودی به مجموعه‌های فازی فضای خروجی می‌رسد. در این پژوهش از موتور استنتاج مینیمم بهره جسته می‌شود. در این حالت، از استلزام[۳۷] مینیمم ممدانی و t-norm، مینیمم و s-norm ماکسیمم استفاده می‌شود. بدین صورت که ابتدا درجه عضویت قسمت‌های مقدم یک قاعده بر اساس ورودی، تعیین می‌شوند. سپس، مینیمم این مقادیر به‌عنوان درجه عضویت تالی آن قاعده درنظر گرفته می‌شود. این کار برای تمام قواعد انجام می‌پذیرد و نتیجه آن‌ها با استفاده از عملگر ماکسیمم تلفیق می‌گردند. با توجه به اینکه خروجی موتور استنتاج یک مجموعه فازی می‌باشد، غیرفازی ساز، این مجموعه را به یک مقدار عددی و قطعی تبدیل می‌کند. در این‌جا از غیرفازی ساز مرکز ثقل بهره گرفته می‌شود. در این حالت نتیجه خروجی از رابطه (۴) به‌دست می‌آید. در پایان، عدد به‌دست‌آمده از این رابطه، به نزدیک‌ترین کلاس شدت حوادث نسبت داده می‌شود.

$$y^* = \frac{\int y\mu(y)dy}{\int \mu(y)dy} \quad (۴)$$

که y^* مرکز ثقل، y متغیر شدت حادثه و $\mu(y)$ درجه تعلق متغیر y در مجموعه فازی به‌دست‌آمده از استنتاج می‌باشد.

۲-۶- سنجش عملکرد

برای مقایسه الگوریتم‌های طبقه‌بندی، از معیارهای مختلفی مانند Accuracy، Precision، Recall و F1-score می‌توان کمک گرفت. این معیارها، پس از انجام پیش‌بینی، از ماتریس درهم‌ریختگی[۳۸] قابل محاسبه می‌باشند. ماتریس درهم‌ریختگی، چگونگی توزیع داده‌ها در کلاس‌های واقعی و کلاس‌های پیش‌بینی‌شده برای آن‌ها را نشان می‌دهد. این ماتریس، یک ماتریس مربعی است که درایه‌های قطر اصلی آن، تعداد نمونه‌هایی را

نشان می‌دهند که کلاس واقعی و پیش‌بینی‌شده برای آن‌ها یکسان است. از آن‌جایی که متغیر شدت حادثه، سه کلاس دارد، ماتریس درهم‌ریختگی برای خروجی مسئله به‌صورت جدول ۱ است.

جدول ۱. ماتریس درهم‌ریختگی برای خروجی با سه کلاس

		مقادیر پیش‌بینی شده		
		C_1	C_2	C_3
مقادیر واقعی	C_1	E_{11}	E_{12}	E_{13}
	C_2	E_{21}	E_{22}	E_{23}
	C_3	E_{31}	E_{32}	E_{33}

بنابر نام‌گذاری صورت‌گرفته در جدول ۱، مقادیر معیارهای ارزیابی از روابط (۵) تا (۸) قابل محاسبه است.

$$Accuracy = \frac{\sum_{i=1}^{3} E_{ii}}{\sum_{i=1}^{3}\sum_{j=1}^{3} E_{ij}} \quad (۵)$$

$$Precision_1 = \frac{E_{11}}{\sum_{i=1}^{3} E_{i1}} \text{ (برای کلاس یک)} \quad (۶)$$

$$Recall_1 = \frac{E_{11}}{\sum_{j=1}^{3} E_{1j}} \text{ (برای کلاس یک)} \quad (۷)$$

$$F1_score_1 = \frac{2 \times Precision_1 \times Recall_1}{Precision_1 + Recall_1} \text{ (برای کلاس یک)} \quad (۸)$$

به‌طور مشابه با روابط (۶) تا (۸) مقادیر Precision، Recall و F1-score برای دو کلاس دیگر نیز محاسبه می‌شوند و میانگین وزن‌دار اعداد به‌دست‌آمده برای کلاس‌ها، به‌عنوان مقدار نهایی استفاده می‌گردد. پس از محاسبه مقادیر نهایی معیارها، مقایسه توسط مقادیر Accuracy و F1-score انجام می‌شود زیرا F1_score میانگین همساز[۳۹] (میانگین توافقی) Precision و Recall است و متأثر از مقدار این دو معیار می‌باشد.

در این پژوهش، عملکرد مدل فازی با دو دسته روش پرکاربرد در یادگیری ماشین مقایسه می‌گردد: دسته اول، مدل‌های جعبه سیاه می‌باشند که پس از فرایند آموزش، درک و توضیح رفتار این مدل‌ها دشوار یا غیرممکن است. دسته دوم، مدل‌های جعبه سفید هستند که تفسیر رفتار این مدل‌ها و فرایند انجام پیش‌بینی در آنان شفاف است. در دسته اول از الگوریتم جنگل تصادفی که توسط بریمن [۲۸]، ماشین بردار پشتیبان که توسط واپنیک [۲۹] و شبکه عصبی پرسپترون که توسط روزنبلات [۳۰] معرفی شده‌اند، استفاده می‌شود. در دسته دوم از الگوریتم ریپر که کوهن [۳۱] و درخت تصمیم که بریمن و همکاران [۳۲] معرفی کرده‌اند، استفاده می‌گردد. برای پیاده‌سازی الگوریتم ریپر از بسته RWeka در زبان برنامه‌نویسی R، منطق فازی از بسته Scikit-Fuzzy در زبان برنامه‌نویسی پایتون و سایر الگوریتم‌ها از بسته Scikit-Learn (با پارامترهای پیش‌فرض) در زبان برنامه‌نویسی پایتون بهره گرفته می‌شود.

۳- نتایج و بحث

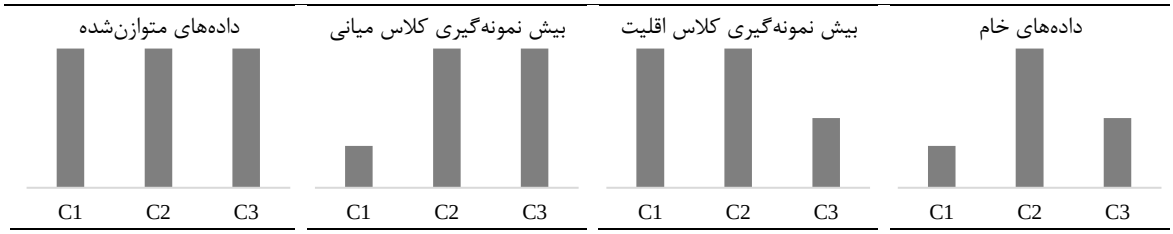
مجموعه‌داده، در چهار حالت شرح داده‌شده در بخش ۲-۲-۳ (داده‌های خام، بیش نمونه‌گیری کلاس اقلیت، بیش نمونه‌گیری کلاس میانی و داده‌های متوازن‌شده) آماده می‌شود تا در هرکدام از حالات به‌عنوان ورودی مدل‌ها استفاده گردد. برای پیاده‌سازی روش فازی، نخست جدول جستجوی فازی قواعد اگر- آن‌گاه را با توابع عضویت بیان‌شده در بخش ۲-۴ استخراج می‌کند. سپس با استفاده از قواعد به‌دست‌آمده، سیستم استنتاج فازی برای پیش‌بینی شدت حوادث، به‌کار گرفته می‌شود. نتایج روش فازی در جدول ۲ ارائه می‌گردد.

از جدول ۲ می‌توان دریافت که در همه حالات با افزایش تعداد تقسیم ورودی‌ها از دو به سه، تعداد قواعد افزایش قابل‌توجهی می‌یابد اما مقادیر Accuracy و F1-score کاهش پیدا می‌کند. بدین ترتیب انجام تقسیم‌بندی با دو تابع عضویت، منجر به دراختیارداشتن سیستم ساده‌تر و دقیق‌تر می‌گردد. همچنین در تمامی موارد، توابع عضویت گوسی نسبت به توابع عضویت زنگوله‌ای تعمیم‌یافته مقادیر Accuracy و F1-score بالاتری را در اختیار می‌گذارند. در نتیجه، بهینه‌ترین حالت برای مدل فازی استفاده از دو تابع عضویت گوسی برای تقسیم هر ورودی می‌باشد. بنابراین نتایج در این حالت، با سایر الگوریتم‌ها مقایسه می‌گردد.

موردی دیگر که از جدول ۲ قابل برداشت می‌باشد، تغییرات بسیار کم مقادیر Accuracy و F1-score با تغییر در توزیع داده‌ها است. به‌عبارت‌دیگر، حالت‌های مختلف نامتوازنی در مجموعه‌داده غالباً تأثیری در نحوه انجام پیش‌بینی ندارند؛ در موارد تأثیرگذار نیز تغییرات بسیار ناچیز می‌باشند. برای مثال در ردیف اول جدول ۲، مقادیر Accuracy، F1-score و تعداد قواعد در هر چهار حالت توزیع داده‌ها تغییری پیدا نمی‌کنند. در ردیف دوم نیز Accuracy و F1-score در سه حالت یکسان است و این اعداد تنها در حالت بیش نمونه‌گیری کلاس میانی کمی کاهش می‌یابند. در این ردیف، تعداد قواعد نیز از ۷۲ تا ۷۷ تغییر می‌کند.

جدول ۲. عملکرد منطق فازی در انجام پیش‌بینی

شماتیک توزیع داده‌های آموزش



تابع عضویت	تعداد تقسیم هر ورودی	داده‌های خام			بیش نمونه‌گیری کلاس اقلیت			بیش نمونه‌گیری کلاس میانی			داده‌های متوازن‌شده		
		Accuracy	F1-score	تعداد قواعد	Accuracy	F1-score	تعداد قواعد	Accuracy	F1-score	تعداد قواعد	Accuracy	F1-score	تعداد قواعد
گوسی	۲	۸۳/۴۳	۷۹/۶۹	۱۸	۸۳/۴۳	۷۹/۶۹	۱۸	۸۳/۴۳	۷۹/۶۹	۱۸	۸۳/۴۳	۷۹/۶۹	۱۸
گوسی	۳	۸۱/۷۱	۷۸/۹۴	۷۲	۸۱/۷۱	۷۸/۹۴	۷۵	۸۱/۶۱	۷۸/۸۱	۷۵	۸۱/۷۱	۷۸/۹۴	۷۷
زنگوله‌ای	۲	۸۰/۹۴	۷۸/۵۸	۱۸	۸۰/۹۴	۷۸/۵۸	۱۸	۸۰/۹۴	۷۸/۵۸	۱۸	۸۰/۹۴	۷۸/۵۸	۱۸
زنگوله‌ای	۳	۷۸/۸۳	۷۷/۶۷	۷۲	۷۸/۸۳	۷۷/۶۷	۷۵	۷۸/۷۴	۷۷/۶۱	۷۵	۷۸/۷۴	۷۷/۶۱	۷۷

توجیه این مسئله را می‌توان در نحوه یادگیری داده‌ها جستجو کرد. در مرحله استخراج قواعد، به‌ازای هر داده، یک قاعده تولید می‌شود. با توجه به اینکه در متوازن‌سازی، داده‌هایی مصنوعی، نزدیک به داده‌های موجود تولید می‌گردند که شباهت زیادی با آن‌ها دارند، قواعد منتج از آن‌ها نیز تکراری می‌شود. در بین قواعد تکراری فقط یکی از آن‌ها نگهداری می‌شود و سایر قواعد حذف می‌گردند. بنابراین با تولید نمونه‌های مصنوعی، تعداد داده‌های تقریباً مشابه افزایش می‌یابد. این افزایش تغییری در فرایند استخراج قواعد ایجاد نمی‌کند و در نهایت همه این داده‌ها منجر به تولید یک قاعده می‌گردند. پس از متوازن‌سازی، تنها ممکن است در بخش‌هایی از فضای ورودی که در ابتدا داده‌ای در آن‌ها وجود نداشته است، داده‌هایی مصنوعی قرار بگیرند. این داده‌ها منجر به تولید قواعد جدیدی می‌شوند که تعداد آن‌ها زیاد نیست. باتوجه‌به تغییرات اندک در پایگاه قواعد که مبنای فرایند استنتاج را شکل می‌دهند، نتیجه پیش‌بینی نیز تغییرات کمی را تجربه می‌کند. مزیت این که بتوانیم با استفاده از داده‌های خام نتایجی بسیار نزدیک به داده‌های متوازن‌شده در اختیار داشته باشیم، این است که مرحله تولید داده‌های مصنوعی قابل حذف است. به‌عبارت‌دیگر، می‌توان مدل‌سازی را در همان ابتدا و با داده‌های کم تعداد تر انجام داد که منجر به کاهش بار محاسباتی می‌گردد.

در ادامه، پایگاه قواعد در حالت تقسیم ورودی‌ها به دو قسمت با توابع گوسی که بهینه‌ترین حالت برای روش فازی است، در جدول ۳ نشان داده می‌شود.

جدول ۳. پایگاه قواعد در صورت تقسیم هر ورودی به دو قسمت با توابع گوسی

شماره قاعده	مقدم					تالی
	تجربه	عوامل انسانی	عوامل محیط کار	رفتارهای پرخطر	موارد خطرناک	شدت حادثه
۱	زیاد	کم	کم	کم	کم	خفیف
۲	زیاد	کم	زیاد	کم	کم	خفیف
۳	کم	کم	کم	کم	زیاد	خفیف
۴	کم	زیاد	زیاد	زیاد	کم	خفیف
۵	کم	کم	زیاد	زیاد	کم	خفیف
۶	زیاد	زیاد	کم	کم	کم	متوسط
۷	کم	زیاد	کم	کم	زیاد	متوسط
۸	کم	کم	زیاد	زیاد	زیاد	متوسط
۹	کم	زیاد	زیاد	کم	کم	متوسط
۱۰	کم	زیاد	کم	زیاد	کم	متوسط
۱۱	کم	کم	زیاد	کم	کم	شدید
۱۲	کم	زیاد	کم	کم	کم	شدید
۱۳	زیاد	کم	کم	زیاد	کم	شدید
۱۴	کم	کم	کم	کم	کم	شدید
۱۵	کم	کم	کم	زیاد	کم	شدید
۱۶	زیاد	زیاد	کم	زیاد	کم	شدید
۱۷	زیاد	کم	زیاد	زیاد	کم	شدید
۱۸	کم	زیاد	زیاد	کم	زیاد	شدید

اگر دانش انسانی برای کاهش شدت حوادث را زیادبودن تجربه و کم‌بودن هرکدام از عوامل انسانی، عوامل محیط کار، رفتارهای پرخطر و موارد خطرناک درنظر بگیریم، بررسی جدول ۳ بیانگر آن است که در بعضی موارد دانش انسانی با دانش به‌دست‌آمده از ماشین مطابقت دارد. برای مثال بخش مقدم قواعد شماره ۱۸ و ۹، دارای چهار گزاره یکسان است و با مقایسه آن‌ها مشخص می‌گردد که تغییر گزاره موارد خطرناک از زیاد به کم، باعث کاهش شدت حادثه از شدید به متوسط می‌شود. همین وضعیت

برای قواعد شماره ۱۱ و ۲ نیز اتفاق می‌افتد. در این‌جا با افزایش تجربه در قاعده ۲ نسبت به ۱۱، شدت حادثه از شدید به خفیف تغییر پیدا می‌کند. اما در بعضی موارد، دانش انسانی و دانش ماشین با یکدیگر همخوانی ندارند. برای مثال بخش مقدم قواعد شماره ۱۵ و ۱۰ فقط در گزاره عوامل انسانی با یکدیگر تفاوت دارند اما در اینجا با کاهش عوامل انسانی، شدت حادثه از متوسط به شدید تغییر می‌یابد. علت این امر می‌تواند وجود عوامل مؤثر دیگری در شدت حوادث باشد که اطلاعاتی از آن‌ها ثبت نگردیده است. برای مثال، شرایط آب‌وهوایی در هنگام حادثه یا وضعیت روحی فرد می‌توانند در نحوه وقوع حادثه اثر بگذارند؛ بااین‌حال، دررابطه‌با حوادث موردبحث اطلاعاتی از آن‌ها در دسترس نیست.

با توجه به توضیحات داده‌شده، می‌توان دریافت که شدت حوادث، پدیده‌ای تصادفی نیست بلکه الگوهایی وجود دارند که می‌توانند نتیجه حوادث را توصیف می‌کنند. این یافته برخلاف مطالعه تکسیر و همکاران [۷] است که ماهیت شدت حوادث را تصادفی می‌دانند. بااین‌حال تصادفی‌نبودن شدت حوادث عمدتاً در مطالعاتی که پس از مطالعه تکسیر و همکاران انجام‌گرفته است، تأیید شده است. برای مثال پو و همکاران [۸] نشان دادند که شدت حوادث به‌صورت تصادفی رخ نمی‌دهند. بیکر و همکاران [۱۱] توانستند شدت حوادث را با دقت بالایی پیش‌بینی کنند و نتیجه گرفتند که شدت حوادث، پدیده‌ای تصادفی نیست. ژو و همکاران [۱۲] دریافتند که آسیب‌های ناشی از حوادث به‌صورت تصادفی اتفاق نمی‌افتند، بلکه دارای الگوهای بالقوه‌ای هستند و این الگوها توسط یک مدل یادگیری ماشین قدرتمند قابل آشکارسازی هستند. با این هدف، نویسندگان توسط درخت تصمیم مجموعه‌ای از قواعد را برای ارزیابی شدت حوادث استخراج کردند.

به‌منظور سنجش عملکرد منطق فازی، نتایج به‌دست‌آمده از الگوریتم‌های ریپر، درخت تصمیم، جنگل تصادفی، ماشین بردار پشتیبان و شبکه عصبی مصنوعی در جدول ۴ نشان داده می‌شوند. شایان ذکر است که الگوریتم‌های جنگل تصادفی، ماشین بردار پشتیبان و شبکه عصبی مصنوعی فرایند مدل‌سازی را به‌صورت جعبه سیاه انجام می‌دهند و به‌طور مستقیم قاعده‌ای در اختیار نمی‌گذارند. این موضوع یک نقصان برای این روش‌ها نسبت به منطق فازی به‌حساب می‌آید. همچنین درخت تصمیم در دو حالت پیاده‌سازی می‌گردد. در حالت اول حداکثر عمق درخت به پنج، محدود می‌باشد زیرا بخش مقدم قواعد فازی، پنج قسمت دارد. در حالت دوم حداقل تعداد نمونه در گره برگ برابر یک انتخاب می‌گردد تا الگوریتم در انتخاب حداکثر عمق مناسب برای درخت، آزادی داشته باشد.

با توجه به جدول ۴ مشخص می‌گردد که برای داده‌های متوازن‌شده، مقادیر Accuracy و F1-score در روش فازی از همه مدل‌ها بالاتر می‌باشد. این در حالی است که نتایج به‌دست‌آمده برای داده‌های خام، در اغلب موارد نزدیک به روش فازی و بالاتر از آن است. این مسئله، وجود جهت‌گیری نسبت به کلاس اکثریت، در مدل‌سازی داده‌های نامتوازن توسط الگوریتم‌های جدول ۴ را تأیید می‌کند. با توجه به اینکه در ابتدا تعداد نمونه‌های کلاس‌های یک و سه کم می‌باشد، الگوریتم‌ها به‌منظور افزایش دقت، توجه کمتری به پیش‌بینی این کلاس‌ها دارند. هنگامی که تعداد نمونه‌های کلاس‌های یک و سه افزایش می‌یابد و برابر با تعداد نمونه‌های کلاس دو می‌شود، الگوریتم‌ها سعی در پیش‌بینی بهتر کلاس‌های مذکور می‌کنند. این اتفاق، باعث کاهش عملکرد مدل‌ها پس از متوازن‌سازی می‌شود. بنابراین مقایسه روش‌ها باید پس از متوازن‌سازی داده‌ها صورت پذیرد. در حالت استفاده از داده‌های متوازن‌شده علاوه بر اینکه مدل بهینه فازی عملکرد بهتری در انجام پیش‌بینی دارد، این کار را با تعداد قواعد کمتری نسبت به روش‌های ریپر و درخت تصمیم انجام می‌دهد. کم‌تر بودن تعداد قواعد، به معنی پیچیدگی کمتر مدل می‌باشد و باعث می‌شود تا مدل به‌دست‌آمده راحت‌تر برای انسان قابل‌درک باشد.

برای درک بهتر علت تفاوت عملکرد مدل‌ها در پیش‌بینی داده‌ها، پیش و پس از متوازن‌سازی آن‌ها، باید به تفاوت رفتار مدل‌ها در یادگیری داده‌ها توجه کرد. در الگوریتم‌های مختلف روش‌های متفاوتی برای یادگیری داده‌های آموزش و ایجاد نگاشتی بین فضای ورودی و خروجی مسئله وجود دارد. با استفاده از نگاشت مذکور، می‌توان برای ورودی‌های جدید، خروجی را پیش‌بینی کرد. برای مثال الگوریتم ریپر تعدادی قاعده از داده‌های آموزش استخراج می‌کند و سپس توسط این قواعد داده‌های آزمون را پیش‌بینی می‌کند. باتوجه‌به جدول ۴، هنگام یادگیری داده‌های خام توسط الگوریتم ریپر، تنها یک قاعده استخراج می‌گردد. در این قاعده، شدت حوادث متوسط (کلاس دو) بیان می‌شود. بنابراین این الگوریتم تمامی داده‌های آزمون را متوسط پیش‌بینی می‌کند. هنگام یادگیری داده‌های متوازن، ۲۵ قاعده توسط الگوریتم ریپر استخراج می‌شود. این قواعد، دارای شدت حوادث خفیف، متوسط و شدید هستند. بنابراین در این حالت، شدت حوادث داده‌های آزمون سه مقدار متفاوت پیش‌بینی می‌گردد. در نتیجه، هنگام متوازن بودن یا نبودن داده‌ها، فرایند یادگیری داده‌های آموزش تغییر می‌یابد و این تغییر باعث تفاوت عملکرد مدل در پیش‌بینی داده‌های آزمون می‌شود. رفتار سایر الگوریتم‌های جدول ۴ در یادگیری داده‌ها نیز کم‌وبیش شبیه به الگوریتم ریپر می‌باشد، به همین دلیل این الگوریتم‌ها پس از متوازن‌سازی، افت عملکرد را تجربه می‌کنند. در نقطه مقابل، روش فازی قرار دارد. قواعد مستخرج توسط جدول جستجو در حالت بهینه برای داده‌های نامتوازن ۱۸ عدد (جدول ۳) می‌باشند. در این حالت، قواعد مستخرج دارای هر سه کلاس شدت حوادث هستند. همان‌طور که توضیح داده شد، پس از متوازن‌سازی داده‌ها نیز قواعد مستخرج ۱۸ عدد و بدون تغییر نسبت به حالت قبل به‌دست می‌آیند. این موضوع که قواعد مستخرج در حالت توازن داده‌ها، از ابتدا در حالت عدم‌توازن در داده‌ها نیز به‌دست می‌آید، بیانگر آن است که روش فازی در یادگیری داده‌های نامتوازن جهت‌گیری ندارد. باتوجه به اینکه در این روش، مرحله یادگیری داده‌ها پیش و پس از متوازن‌سازی بدون تغییری باقی می‌ماند (تغییرنکردن قواعد)، پیش‌بینی انجام‌شده توسط مدل نیز تغییری نمی‌یابد.

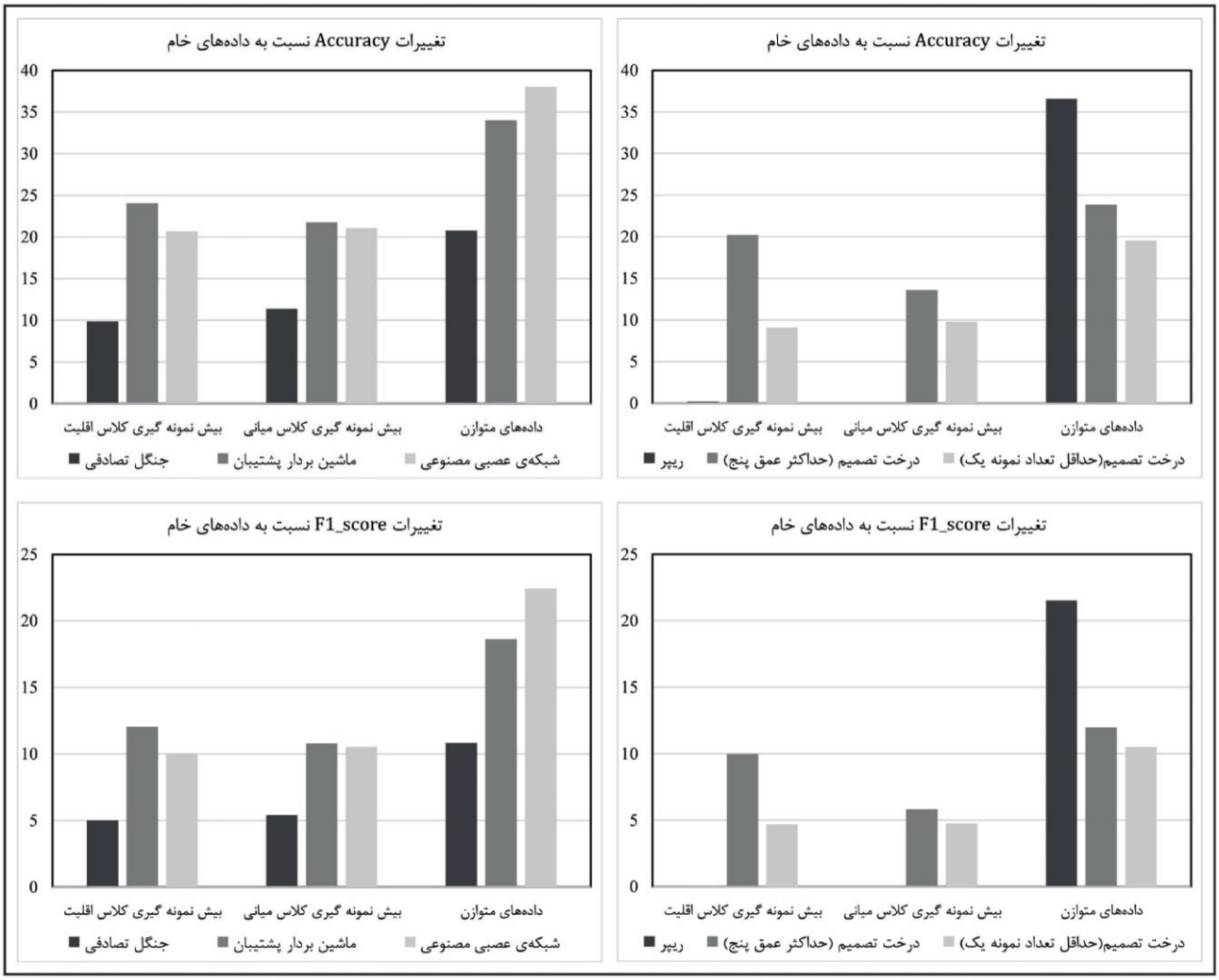
شکل ۳ میزان افت عملکرد مدل‌های جدول ۴ پس از بیش نمونه‌گیری در هر حالت را نشان می‌دهد. ستون‌های این شکل با کم‌کردن مقادیر Accuracy و F1-score در هر حالت متوازن‌سازی از مقادیر مربوطه در حالت داده‌های خام به‌دست می‌آیند.

باتوجه به شکل ۳، در تمامی موارد بیشترین افت عملکرد برای هر الگوریتم در حالت توازن در داده‌ها اتفاق می‌افتد زیرا در این حالت همه کلاس‌ها دارای تعداد نمونه یکسان و ارزش برابر در انجام پیش‌بینی

می‌باشند اما در دو حالت دیگر بیش نمونه‌گیری، یک کلاس وجود دارد که دارای تعداد نمونه‌های کمتری از بقیه است. همچنین بیشترین تغییرات در Accuracy و F1-score برای داده‌های متوازن مربوط به شبکه عصبی مصنوعی و کمترین تغییرات مربوط به الگوریتم درخت تصمیم با حداقل یک نمونه در گره برگ می‌باشد. میزان کاهش Accuracy و F1-score الگوریتم ریپر در حالت‌های بیش نمونه‌گیری از کلاس میانی و کلاس اقلیت ناچیز است اما در حالت داده‌های متوازن، پس از شبکه عصبی مصنوعی، بیشترین افت عملکرد مربوط به این الگوریتم می‌باشد. در مطالعات دیگری که در سایر حوزه‌ها بر روی پیش‌بینی داده‌های نامتوازن انجام‌گرفته نیز می‌توان افت عملکرد مدل‌های یادگیری ماشین پس از متوازن‌سازی را مشاهده کرد. برای مثال مجلی و همکاران [۳۳] حوادث ترافیکی را مدل‌سازی کردند و دریافتند که عملکرد تمامی مدل‌ها پس از متوازن‌سازی دچار کاهش می‌شوند. رفتاری مشابه در مطالعه کائو و همکاران [۳۴] نیز برای چندین مدل یادگیری ماشین مختلف قابل‌مشاهده است.

جدول ۴. عملکرد روش‌های یادگیری ماشین در انجام پیش‌بینی

الگوریتم	شماتیک توزیع داده‌های آموزش											
	داده‌های خام			بیش نمونه‌گیری کلاس اقلیت			بیش نمونه‌گیری کلاس میانی			داده‌های متوازن‌شده		
	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
	Accuracy	F1-score	تعداد قواعد	Accuracy	F1-score	تعداد قواعد	Accuracy	F1-score	تعداد قواعد	Accuracy	F1-score	تعداد قواعد
ریپر	۸۶/۴۹	۸۰/۲۳	۱	۸۶/۳۰	۸۰/۱۸	۱۱	۸۶/۴۹	۸۰/۲۳	۱۳	۴۹/۹۰	۵۸/۷۰	۲۵
درخت تصمیم (حداکثر عمق درخت برابر پنج)	۸۵/۹۲	۸۰/۱۶	۲۸	۶۵/۷۱	۷۰/۱۸	۲۱	۷۲/۳۲	۷۴/۳۳	۲۲	۶۲/۰۷	۶۸/۱۷	۳۰
درخت تصمیم (حداقل یک نمونه در گره برگ)	۸۲/۸۵	۷۹/۴۰	۴۲۹	۷۳/۷۶	۷۴/۷۰	۵۲۹	۷۳/۰۸	۷۴/۶۴	۵۳۵	۶۳/۳۱	۶۸/۸۷	۱۲۴۷
جنگل تصادفی	۸۴/۳۰	۷۹/۸۳	-	۷۴/۴۳	۷۴/۸۱	-	۷۲/۸۹	۷۴/۴۱	-	۶۳/۵۱	۶۹/۰۰	-
ماشین بردار پشتیبان	۸۶/۴۹	۸۰/۲۳	-	۶۲/۴۵	۶۸/۱۷	-	۶۴/۷۵	۶۹/۴۴	-	۵۲/۴۹	۶۱/۶۱	-
شبکه عصبی مصنوعی	۸۶/۴۹	۸۰/۵۹	-	۶۵/۸۱	۷۰/۵۶	-	۶۵/۴۲	۷۰/۰۷	-	۴۸/۴۷	۵۸/۱۶	-



شکل ۳. افت عملکرد مدل‌ها با بیش نمونه‌گیری داده‌ها

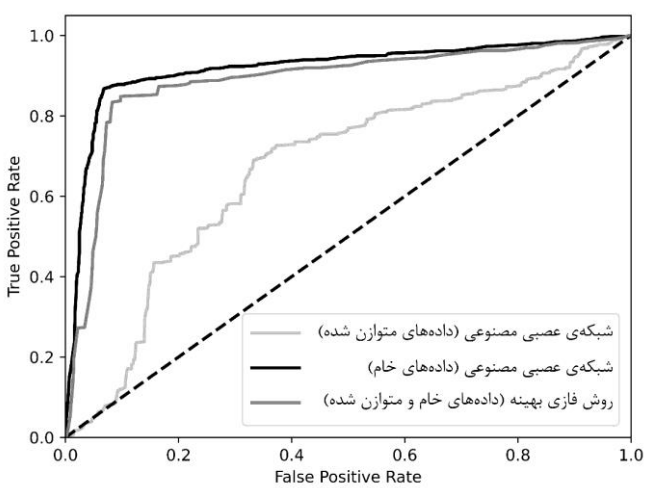
نمودارهای مشخصه عملکرد سیستم[۴۰] و Precision-Recall دو ابزار برای ارزیابی موفقیت مدل‌ها در انجام طبقه‌بندی داده‌ها می‌باشند. به‌منظور نمایش بهتر رفتار مدل‌ها پیش از متوازن‌سازی و پس از آن شکل‌های ۴ و ۵ ارائه می‌گردند. در شکل ۴ نمودار مشخصه عملکرد سیستم برای شبکه عصبی مصنوعی، به‌عنوان مدلی که پس از متوازن‌سازی بیشترین تغییرات را دارد و روش فازی بهینه، به‌عنوان مدلی که پس از متوازن‌سازی تغییری نمی‌یابد، ترسیم می‌شود. در این نمودار محور y، نرخ مثبت صحیح و محور x، نرخ مثبت کاذب می‌باشد. نرخ مثبت صحیح بر اساس نام‌گذاری انجام‌شده در جدول ۱، از رابطه (۷) و نرخ مثبت کاذب از رابطه (۹) به‌دست می‌آید.

$$False\ Positive\ Rate_1 = \frac{E_{21}+E_{31}}{\sum_{j=1}^{3}(E_{2j}+E_{3j})} \quad (\text{برای کلاس یک}) \qquad (۹)$$

به‌طور مشابه با روابط (۷) و (۹) مقادیر مذکور برای دو کلاس دیگر نیز محاسبه می‌شوند و میانگین وزن‌دار اعداد به‌دست‌آمده برای کلاس‌ها، به‌عنوان مقدار نهایی در نمودار استفاده می‌گردد.

شکل ۴. نمودار مشخصه عملکرد سیستم

در شکل ۵ نیز نمودار Precision-Recall برای شبکه عصبی مصنوعی و روش فازی بهینه ترسیم می‌گردد. در این نمودار محور y، Precision و محور x، Recall می‌باشد.



شکل ۵. نمودار Precision-Recall

باتوجه به شکل ۴، منحنی مرتبط با روش فازی بهینه پیش و پس از متوازن‌سازی تغییری نمی‌یابد. این منحنی پایین‌تر از منحنی شبکه عصبی مصنوعی برای داده‌های خام و بالاتر از منحنی شبکه عصبی مصنوعی برای داده‌های متوازن قرار دارد. این مسئله بدین معنی است که ابتدا عملکرد شبکه عصبی مصنوعی بهتر از روش فازی می‌باشد اما پس از متوازن‌سازی این عملکرد دچار افت زیادی می‌شود. برای ارزیابی عددی عملکرد مدل‌ها

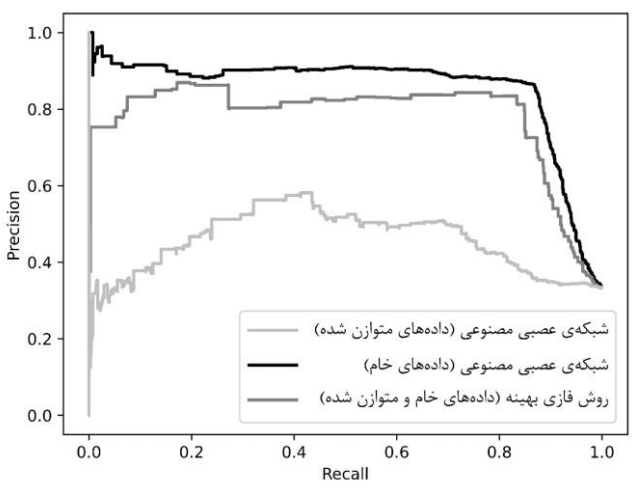
بر اساس نمودار مشخصه عملکرد سیستم، سطح زیر این نمودار محاسبه می‌گردد. بزرگ‌تر بودن این مقدار به معنی عملکرد بهتر مدل در انجام پیش‌بینی می‌باشد. سطح زیر منحنی برای روش فازی بهینه ۰/۸۹ به‌دست می‌آید. این عدد برای شبکه عصبی مصنوعی پیش از متوازن‌سازی ۰/۹۲ است و پس از متوازن‌سازی به ۰/۶۷ کاهش می‌یابد. این تغییرات در اعداد مذکور، به‌طور مشابه در بررسی مقادیر Accuracy و F1-score نیز مشاهده گردید. نتایج در شکل ۵ نیز کاملاً مشابه با نتایج قبلی می‌باشد. در این شکل منحنی مربوط به روش فازی بهینه، بین دو منحنی مربوط به شبکه عصبی مصنوعی قرار می‌گیرد و سطح زیر نمودار در روش فازی بهینه بیشتر از شبکه عصبی پس از متوازن‌سازی و کمتر از آن پیش از متوازن‌سازی می‌باشد. در این نمودار از میانگین وزن‌دار Precisionها برای خلاصه کردن شکل و ارزیابی عددی مدل‌ها استفاده می‌شود. وزن‌ها در هر گام، مقدار افزایش Recall در آن گام نسبت به گام قبل درنظر گرفته می‌شود. با توجه به شکل ۵، میانگین وزن‌دار Precision برای روش فازی بهینه پیش و پس از متوازن‌سازی، ۰/۷۹ می‌باشد. این مقدار برای شبکه عصبی مصنوعی پیش از متوازن‌سازی ۰/۸۶ است که پس از متوازن‌سازی به ۰/۴۵ کاهش می‌یابد.

در مطالعه دیگری که ایهان و توکدمیر [۲۳] بر روی مجموعه‌داده حاضر انجام داده‌اند، پس از خوشه‌بندی داده‌ها از شبکه عصبی مصنوعی و استدلال مبتنی بر مورد در چند حالت مختلف، استفاده شده است. نتایج عملکرد مدل‌های مطالعه مذکور بر اساس معیار میانگین درصد قدرمطلق خطا[41] گزارش شده است. نحوه محاسبه این پارامتر در رابطه (۱۰) بیان می‌گردد.

$$Mean\ Absolute\ Percentage\ Error = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{F_i - A_i}{A_i}\right| \quad (۱۰)$$

که A_i مقدار واقعی، F_i مقدار پیش‌بینی‌شده برای داده i ام و n تعداد داده‌ها می‌باشد.

براین‌اساس به‌منظور مقایسه عملکرد روش فازی پژوهش حاضر با مدل‌های ارائه‌شده توسط ایهان و توکدمیر جدول ۵ ارائه می‌شود. در این جدول میانگین درصد قدرمطلق خطا برای مدل فازی بهینه پژوهش حاضر و بهترین مدل شبکه عصبی مصنوعی و استدلال مبتنی بر مورد در پژوهش ایهان و توکدمیر نشان داده می‌شود. همچنین به‌عنوان نمونه، عملکرد شبکه عصبی مصنوعی پژوهش حاضر نیز در جدول ۵ ارائه می‌شود.



جدول ۵. مقایسه نتایج پژوهش حاضر و پژوهش ایهان و توکدمیر

پژوهش	بهترین عملکرد	
	مدل	میانگین درصد قدرمطلق خطا
مطالعه حاضر	فازی	٪۱۰/۱۲
	شبکه عصبی مصنوعی پیش از متوازن‌سازی داده‌ها	٪۸/۴۹
	شبکه عصبی مصنوعی پس از متوازن‌سازی داده‌ها	٪۲۹/۱۲
ایهان و توکدمیر [۲۳]	شبکه عصبی مصنوعی	٪۲۱/۰۴
	استدلال مبتنی بر مورد	٪۱۷/۴۱

بر اساس جدول ۵ میانگین درصد قدرمطلق خطای روش فازی پیش و پس از متوازن‌سازی داده‌ها تغییری نمی‌کند و کمتر از روش‌های شبکه عصبی مصنوعی و استدلال مبتنی بر مورد در مطالعه ایهان و توکدمیر است. این موضوع به معنی عملکرد بهتر روش فازی در انجام پیش‌بینی می‌باشد. همچنین برای مثال خطای شبکه عصبی در این مطالعه پیش از متوازن‌سازی داده‌ها ۸/۴۹ درصد می‌باشد اما پس از متوازن‌سازی به ۲۹/۱۲ درصد می‌رسد. با توجه به اینکه در مطالعه ایهان و توکدمیر از داده‌های نامتوازن برای مدل‌سازی استفاده شده است، می‌توان انتظار داشت که در صورت پیاده‌سازی روش تحقیق آن مطالعه و استفاده از داده‌های متوازن خطای حاصل بیشتر از مقدار گزارش شده باشد.

۴- نتیجه‌گیری

پژوهش حاضر، شدت حوادث صنعت ساخت‌وساز را با استفاده از منطق فازی، تجزیه‌وتحلیل می‌کند. با توجه به وجود مشکل نامتوازنی در مجموعه‌داده، روش بیش نمونه‌گیری اقلیت مصنوعی به‌کار گرفته می‌شود. پس از متوازن‌سازی داده‌های آموزش، فرایند توسعه مدل، آغاز می‌گردد. این فرایند شامل سه مرحله می‌باشد: تعریف توابع عضویت، استخراج قواعد فازی و پیاده‌سازی سیستم استنتاج فازی.

با مشاهده نتایج معلوم می‌گردد که برای مدل بهینه فازی، مقادیر Accuracy و F1-score به‌ترتیب ۸۳/۴۳ درصد و ۷۹/۶۹ درصد به‌دست می‌آید و این عملکرد با ۱۸ قاعده حاصل می‌شود. کم‌بودن تعداد قواعد، بیانگر پیچیدگی کمتر مدل و تفسیرپذیری راحت‌تر آن می‌باشد. این موضوع نشان می‌دهد که آسیب‌های ناشی از حوادث به‌صورت تصادفی اتفاق نمی‌افتند و الگوهای پنهان حاکم بر آن‌ها را می‌توان با کمک مدل‌های مبتنی بر ماشین آشکار کرد. مسئله مهم دیگر که از بررسی نتایج مشخص می‌گردد، تغییرات اندک عملکرد مدل‌های فازی، در صورت تغییر در متوازن بودن یا نبودن داده‌هاست. این رفتار مدل فازی، برخلاف سایر الگوریتم‌های موردمطالعه می‌باشد. در تمامی الگوریتم‌های جنگل تصادفی، ماشین بردار پشتیبان، شبکه عصبی مصنوعی، ریپر و درخت تصمیم پس از

متوازن‌سازی، عملکرد به‌طور قابل‌توجهی افت می‌کند. علت این افت، جهت‌گیری مدل‌ها نسبت به پیش‌بینی کلاس اکثریت در داده‌های خام می‌باشد که با متوازن‌سازی داده‌ها این جهت‌گیری از بین می‌رود اما با توجه به فرایند یادگیری در روش فازی، از ابتدا چنین مشکلی در مدل بروز نمی‌کند. افت قابل‌چشم‌پوشی عملکرد مدل فازی در کنار تغییرات زیاد روش‌های دیگر، باعث می‌شود تا در حالت توازن داده‌ها که نتایج واقعی هستند، عملکرد روش فازی نسبت به سایر الگوریتم‌ها بهتر باشد.

این مطالعه، سعی دارد تا به متخصصان ایمنی برای پیش‌بینی دقیق‌تر حوادث کمک کند. با توجه به ماهیت ذاتی حوادث که پراکندگی شدت آسیب در آن‌ها متفاوت است و حوادث شدید کمتر اتفاق می‌افتند، مسئله نامتوازنی در داده‌ها امری غیرقابل‌اجتناب است. در این زمینه، به‌کارگیری منطق فازی برای توسعه مدل، نتایج امیدوارکننده‌ای را علی‌رغم وجود نامتوازنی در داده‌ها نشان می‌دهد. مدیران ایمنی غالباً بر اساس تجربه شخصی خود از حوادث گذشته، در مورد خطرات موجود تصمیم‌گیری می‌کنند. باتوجه‌به تجربه محدود فردی، تصمیم‌گیری درباره وضعیت ایمنی موجود می‌تواند دارای خطای زیادی باشد و این امر باعث شود که حوادث احتمالی آتی به‌درستی شناسایی نگردند. باتوجه‌به جعبه‌سیاه نبودن منطق فازی، ارائه قواعد مدل می‌تواند به مدیران ایمنی در اخذ تصمیمات بهتر کمک کند. همچنین فعالان ایمنی می‌توانند قواعد مدل را متناسب با نظرات خود و شرایط حاکم بر هر پروژه تغییر دهند تا بهینه‌ترین مدل برای پیش‌بینی خطرات هر پروژه به‌طور خاص به‌دست بیاید.

درعین‌حال این مطالعه دارای محدودیت‌هایی نیز می‌باشد. ویژگی‌های بسیاری بر شدت حوادث اثر گذارند که در مجموعه‌داده گزارش نشده‌اند؛ مطالعات آینده می‌توانند بر جمع‌آوری جزئیات بیشتر از حوادث تمرکز کنند و میزان اثرگذاری آن‌ها بر شدت حوادث را بررسی کنند. همچنین حوادث مطالعه‌شده در اینجا، در حین فعالیت‌های مختلف مانند جوشکاری، بتن‌ریزی، حفاری و غیره رخ داده‌اند و این شرایط متفاوت می‌تواند بر نتیجه حوادث تأثیر بگذارد. بنابراین مدل‌سازی حوادث مربوط به هر فعالیت به‌صورت جداگانه می‌تواند در آینده موردتوجه قرار گیرد. استخراج قواعد با استفاده از روش‌های دیگر و پیاده‌سازی سیستم استنتاج فازی در حالت‌های متفاوت نیز موضوعاتی برای مطالعه بیشتر هستند. در انتها، روش پیشنهادی می‌تواند برای مجموعه‌داده‌های جدید، انواع حوادث و صنایع مختلف به‌کار رود تا راهنمایی‌های بهتری برای متخصصان ایمنی فراهم گردد.

۵- مراجع

[1] International Labour Organization, “Safety and health at work,” https://www.ilo.org/global/topics/safety-and-health-at-work/lang--en/index.htm.

[۲] ح. خانجانی، "آمار و هزینه‌های بالای حوادث ساختمانی،" ماهنامه پیام آبادگران، پیاپی ۳۸۹، ص ۳۳-۳۴، شهریور ۱۳۹۹.

[۳] ع. صنیع زاده، "آمارهای حوادث ساختمانی،" ماهنامه دانش نما ، پیاپی ۲۶۴-۲۶۲، ص ۱۱۰-۱۱۲، ۱۳۹۶.

[4] H. Y. Chong and T. S. Low, “Accidents in Malaysian construction industry: Statistical data and court cases,” *Int. J. Occup. Saf. Ergon.*, vol. 20, no. 3, pp. 503–513, 2014.

[5] Y. Kang, S. Siddiqui, S. J. Suk, S. Chi, and C. Kim, “Trends of Fall Accidents in the U.S. Construction Industry,” *J. Constr. Eng. Manag.*, vol. 143, no. 8, p. 04017043, Aug. 2017.

[6] S. Sarkar, S. Vinay, R. Raj, J. Maiti, and P. Mitra, “Application of optimized machine learning techniques for prediction of occupational accidents,” *Comput. Oper. Res.*, vol. 106, pp. 210–224, Jun. 2019.

[7] A. J. P. Tixier, M. R. Hallowell, B. Rajagopalan, and D. Bowman, “Application of machine learning to construction injury prediction,” *Autom. Constr.*, vol. 69, pp. 102–114, Sep. 2016.

[8] C. Q. X. Poh, C. U. Ubeynarayana, and Y. M. Goh, “Safety leading indicators for construction sites: A machine learning approach,” *Autom. Constr.*, vol. 93, pp. 375–386, Sep. 2018.

[9] B. U. Ayhan and O. B. Tokdemir, “Accident Analysis for Construction Safety Using Latent Class Clustering and Artificial Neural Networks,” *J. Constr. Eng. Manag.*, vol. 146, no. 3, p. 04019114, Mar. 2020.

[10] J. Choi, B. Gu, S. Chin, and J. S. Lee, “Machine learning predictive model based on national data for fatal accidents of construction workers,” *Autom. Constr.*, vol. 110, p. 102974, Feb. 2020.

[11] H. Baker, M. R. Hallowell, and A. J. P. Tixier, “AI-based prediction of independent construction safety outcomes from universal attributes,” *Autom. Constr.*, vol. 118, p. 103146, Oct. 2020.

[12] R. Zhu, X. Hu, J. Hou, and X. Li, “Application of machine learning techniques for predicting the consequences of construction accidents in China,” *Process Saf. Environ. Prot.*, vol. 145, pp. 293–302, Jan. 2021.

[13] M. Rijo George, M. R. Nalluri, and K. B. Anand, “Severity Prediction of Construction Site Accidents Using Simple and Ensemble Decision Trees,” *Lect. Notes Civ. Eng.*, vol. 171, pp. 599–608, May 2021.

[14] S. Das, S. Datta, and B. B. Chaudhuri, “Handling data irregularities in classification: Foundations, trends, and future challenges,” *Pattern Recognit.*, vol. 81, pp. 674–693, Sep. 2018.

[15] P. Vuttipittayamongkol, E. Elyan, and A. Petrovski, “On the class overlap problem in imbalanced data classification,” *Knowledge-Based Syst.*, vol. 212, p. 106631, Jan. 2021.

[16] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, “An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics,” *Inf. Sci.*, vol. 250, pp. 113–141, Nov. 2013.

[17] K. Kang and H. Ryu, “Predicting types of occupational accidents at construction sites in Korea using random forest model,” *Saf. Sci.*, vol. 120, pp. 226–236, Dec. 2019.

[18] A. R. Fayek, “Fuzzy Logic and Fuzzy Hybrid Techniques for Construction Engineering and Management,” *J. Constr. Eng. Manag.*, vol. 146, no. 7, p. 04020064-1, Jul. 2020.

[19] G. E. Gürcanli and U. Müngen, “An occupational safety risk analysis method at construction sites using fuzzy sets,” *Int. J. Ind. Ergon.*, vol. 39, no. 2, pp. 371–387, Mar. 2009.

[20] D. K. Jana, S. Pramanik, P. Sahoo, and A. Mukherjee, “Interval type-2 fuzzy logic and its application to occupational safety risk performance in industries,” *Soft Comput.*, vol. 23, no. 2, pp. 557–567, Oct. 2017.

[21] L. A. R. Winanda, T. W. Adi, N. Anwar, and F. S. Wahyuni, *Construction safety monitoring based on the project’s characteristic with fuzzy logic approach*, AIP Conf. Proc., vol. 1903, no. 1, p. 070009-1, Nov.2017.

[22] B. U. Ayhan and O. B. Tokdemir, “Predicting the outcome of construction incidents,” *Saf. Sci.*, vol. 113, pp. 91–104, Mar. 2019.

[23] B. U. Ayhan and O. B. Tokdemir, “Safety assessment in megaprojects using artificial intelligence,” *Saf. Sci.*, vol. 118, pp. 273–287, Oct. 2019.

[24] B. Flyvbjerg, “What you Should Know about Megaprojects

and Why: An Overview," *Proj. Manag. J.*, vol. 45, no. 2, pp. 6–19, Apr. 2014.
[25] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, Part B, p. 105524, Dec. 2020.
[26] N. v. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.
[27] L. X. Wang, *A Course in Fuzzy Systems and Control*, Prentice-Hall, 1997.
[28] L. Breiman, "Random forests," *Mach. Learn*, vol. 45, pp. 5–32, Oct. 2001.
[29] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1995.
[30] F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958.
[31] W. W. Cohen, "Fast Effective Rule Induction," *Proc. 12th Int. Conf. Mach. Learn.*, pp. 115–123, Jul. 1995.
[32] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, CRC Press, 1984.
[33] R. O. Mujalli, G. López, and L. Garach, "Bayes classifiers for imbalanced traffic accidents datasets," *Accid. Anal. Prev.*, vol. 88, pp. 37–51, Mar. 2016.
[34] Y. Cao, X. Fang, J. Ottosson, E. Näslund, and E. Stenberg, "A Comparative Study of Machine Learning Algorithms in Predicting Severe Complications after Bariatric Surgery," *J. Clin. Med.,* vol. 8, no. 5, p. 668, May 2019.

---

1. International Labour Organization (ILO)
2. Random Forest (RF)
3. Stochastic Gradient Boosting (SGB)
4. Logistic Regression (LR)
5. Support Vector Machine (SVM)
6. K-Nearest Neighbor (KNN)
7. Decision Tree (DT)
8. Heterogeneity
9. Latent Class Clustering Analysis (LCCA)
10. Artificial Neural Network (ANN)
11. AdaBoost
12. Natural Language Processing (NLP)
13. Extreme Gradient Boosting (XGBoost)
14. Multilayer Perceptron (MLP)
15. Naive Bayes (NB)
16. Automated Machine Learning (AutoML)
17. Occupational and Health Safety (OHS)
18. Majority Class
19. Minority Class
20. Oversampling
21. Undersampling
22. Randomsampling
23. Intelligent Systems
24. Fuzzy Logic
25. Fuzzy Inference System
26. Interval Type-2 Fuzzy Set
27. Fuzzy Lookup Table
28. Repeated Incremental Pruning to Produce Error Reduction(RIPPER)
29. Megaprojects
30. Categorical Variable
31. Synthetic Minority Oversampling Technique (SMOTE)
32. Gaussian
33. Generalized Bell
34. Fuzzifier
35. Defuzzifier
36. Singleton Fuzzifier
37. Implication
38. Confusion Matrix
39. Harmonic Mean
40. Receiver Operating Characteristic (ROC)
41. Mean Absolute Percentage Error (MAPE)