

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

چکیده

طبق گزارش سازمان بین‌المللی کار، حوادث شغلی حدود ۴ درصد از تولید ناخالص داخلی سالانه یا ۱.۲۵ تریلیون دلار برای اقتصاد هزینه دارد. این حوادث نه تنها باعث مشکلات جدی ایمنی و بهداشتی می‌شود، بلکه منجر به خسارات مالی هنگفتی نیز می‌شود. در نتیجه، تجزیه و تحلیل علل حوادث ساختمانی و آسیب‌های ناشی از آن برای بهبود عملکرد ایمنی بسیار مهم است.

مجموعه داده مورد بررسی، اطلاعات حوادث ساختمانی است که توسط اداره ایمنی بهداشت حرفه‌ای ایالات متحده بین سال‌های ۲۰۱۵ تا ۲۰۱۷ منتشر شده است. هدف اصلی این مطالعه، شناسایی عوامل مؤثر بر حوادث ساختمانی، تجزیه و تحلیل اکتشافی در داده‌ها و مدل‌سازی و پیش‌بینی حوادث ساختمانی با استفاده از الگوریتم‌های یادگیری ماشین می‌باشد.

تجزیه و تحلیل اکتشافی و استخراج دانش از مجموعه داده مورد بررسی در این پژوهش، نشان می‌دهد که ۸۵ درصد افرادی که از ناحیه سر دچار آسیب شده‌اند، فوت کرده‌اند. این موضوع، نشان از اهمیت استفاده از کلاه ایمنی مناسب در پروژه‌های ساختمانی دارد. همچنین اکثر افرادی که دچار سقوط جدی / ضربه شده‌اند (حدود ۹۰ درصد موارد)، فوت کرده‌اند. این موضوع نشان می‌دهد که ایمن‌سازی پرتگاه‌ها و قراردادن گاردهای ایمنی^۱ برای ایمن‌سازی در پروژه‌های ساختمانی می‌تواند به عنوان یک اولویت برای جلوگیری از حوادث منجر به مرگ در حوادث ساختمانی، در دستور کار قرار گیرد.

بررسی‌های آماری نشان می‌دهد که تجهیزات جابه‌جایی مواد و وضعیت چیدمان سطح کار کارگر، مجموعاً ۲۸ درصد مرگ و میرهای حوادث مورد بررسی را شامل می‌شوند که این موضوع نشان از اهمیت ایمن‌سازی حرکت تجهیزات و جابه‌جایی مواد و همچنین توجه به چیدمان سطح کار کارگر دارد.

بررسی‌ها نشان می‌دهد که آموزش نحوه قرارگیری مناسب به کارگران در محل مناسب، شبیه‌سازی موقعیت‌ها برای کارگر جهت تصمیم‌گیری بهتر کارگر و الزام کارفرمایان و کارگران به استفاده از تجهیزات حفاظت فردی می‌تواند برای کاهش خطای انسانی در حوادث ساختمانی بسیار مؤثر واقع شود.

از ۱۱ الگوریتم یادگیری ماشین در این پژوهش استفاده شد. با توجه به نتایج ارائه شده، از ۱۱ الگوریتم به کار گرفته شده در این مطالعه، ۸ الگوریتم دقتی بیش از ۸۰ درصد را دادند. می‌توان نتیجه گرفت که استفاده از الگوریتم‌های یادگیری ماشین برای پیش‌بینی درجه آسیب حوادث ساختمانی، تا حد خوبی می‌تواند مؤثر باشد و استفاده از آن می‌تواند سبب بهبود عملکرد در پروژه‌های ساختمانی شود. الگوریتم‌های جنگل تصادفی، دسته‌بندی تقویتی گرادیان و دسته‌بندی مبتنی بر رأی‌گیری به عنوان سه الگوریتم برتر در این پژوهش شناخته شدند.

^۱ Safety Guard

فهرست مطالب

۱- مقدمه	۱
۱-۱- مقدمه	۱
۱-۲- ضرورت انجام تحقیق	۱
۱-۳- اهداف تحقیق	۱
۱-۴- طرح مسئله	۲
۲- پیشینه تحقیق	۳
۲-۱- مرور ادبیات	۳
۲-۲- جدول مرور ادبیات	۱۳
۳- شکاف تحقیقاتی	۲۴
۳-۱- جدول شکاف تحقیقاتی	۲۴
۳-۲- تبیین شکاف تحقیقاتی	۲۶
۴- مبانی نظری	۲۷
۴-۱- حوادث بر اثر ساخت و ساز	۲۷
۴-۲- استرس	۲۷
۴-۲-۱- عوامل استرس‌زای شغلی	۲۷
۴-۳- عوامل مرتبط با رفتار کارگران در بروز حوادث	۲۸
۴-۴- داده کاوی	۲۸
۴-۵- شاخص‌های ایمنی پیش‌رو و پس‌رو	۲۸
۴-۶- یادگیری ماشین	۲۸
۴-۶-۱- انواع یادگیری ماشین	۲۹
۴-۶-۱-۱- الگوریتم‌های یادگیری ماشین بانظارت	۲۹
۴-۶-۱-۲- الگوریتم‌های یادگیری ماشین بدون نظارت	۲۹
۴-۶-۱-۳- یادگیری تقویتی	۳۰
۴-۶-۲- بیش برآزش، کم برآزش و برآزش مناسب	۳۰
۴-۶-۳- موازنه واریانس و بایاس	۳۲

- ۳۲-۷-۴ یادگیری ماشین در ساخت و ساز ۳۲
- ۳۳-۵ مطالعه موردی ۳۳
- ۳۴-۱-۵ روش تحقیق ۳۴
- ۳۴-۱-۱-۵ پیش پردازش داده ها: پاک سازی داده، کاهش ابعاد، انتخاب ویژگی ها ۳۴
- ۳۴-۱-۲-۵ تجزیه و تحلیل اکتشافی داده ها ۳۴
- ۳۵-۱-۳-۱-۵ استخراج قوانین انجمنی ۳۵
- ۳۶-۱-۳-۱-۵ Apriori الگوریتم ۳۶
- ۳۶-۱-۳-۲-۱-۵ معیارهای ارزیابی قوانین انجمن ۳۶
- ۳۷-۱-۴-۱-۵ پیاده سازی الگوریتم های یادگیری ماشین ۳۷
- ۳۷-۱-۴-۱-۵ آماده سازی مجموعه داده برای پیاده سازی الگوریتم های یادگیری ماشین ۳۷
- ۳۷-۱-۴-۲-۱-۵ تقسیم مجموعه داده به سه قسمت آموزشی، اعتبارسنجی و تست ۳۷
- ۳۸-۱-۴-۳-۱-۵ پیاده سازی الگوریتم های یادگیری ماشین ۳۸
- ۳۸-۱-۴-۳-۱-۵ رگرسیون لجستیک ۳۸
- ۳۹-۱-۴-۳-۲-۱-۵ K نزدیک ترین همسایه ۳۹
- ۴۰-۱-۴-۳-۳-۱-۵ درخت تصمیم گیری ۴۰
- ۴۱-۱-۴-۳-۴-۱-۵ جنگل تصادفی ۴۱
- ۴۲-۱-۴-۳-۵-۱-۵ ماشین بردار پشتیبان ۴۲
- ۴۳-۱-۴-۳-۶-۱-۵ بیز ساده گاوسی ۴۳
- ۴۴-۱-۴-۳-۷-۱-۵ بیز ساده برنولی ۴۴
- ۴۴-۱-۴-۳-۸-۱-۵ دسته بندی کیسه ای ۴۴
- ۴۴-۱-۴-۳-۹-۱-۵ دسته بندی تقویتی گرادبان ۴۴
- ۵۰-۱-۴-۳-۱۰-۱-۵ دسته بندی XGboost ۵۰
- ۵۱-۱-۴-۳-۱۱-۱-۵ دسته بندی بر اساس رأی گیری ۵۱
- ۵۱-۱-۵-۱-۵ اعتبارسنجی و بررسی صحت الگوریتم ها ۵۱
- ۵۲-۱-۵-۱-۱-۵ ماتریس اغتشاش ۵۲
- ۵۲-۱-۵-۲-۱-۵ دقت ۵۲

۵۲	۵-۱-۳- صحت
۵۳	۵-۱-۴- پوشش
۵۳	۵-۱-۵- امتیاز F_1
۵۳	۵-۱-۶- اعتبارسنجی متقابل
۵۴	۵-۱-۷- منحنی AUC – ROC
۵۵	۵-۱-۶- بررسی نتایج و مقایسه الگوریتم‌ها
۵۵	۵-۲- یافته‌های تحقیق
۵۵	۵-۲-۱- تجزیه و تحلیل اکتشافی
۶۴	۵-۲-۲- استخراج قوانین انجمنی
۶۵	۵-۲-۳- ارزیابی و مقایسه عملکرد الگوریتم‌های یادگیری ماشین
۶۶	۶- نتیجه‌گیری و پیشنهاد
۶۸	۷- منابع

فهرست اشکال

۸	شکل (۱-۲) دیاگرام پردازش اطلاعات کارگر
۸	شکل (۲-۲) عوامل مؤثر بر شدت حوادث ساختمانی
۹	شکل (۳-۲) مدل مفهومی ارائه شده برای شدت حوادث ساختمانی
۹	شکل (۴-۲) مدل مفهومی گروه‌بندی مضامین در ارتباط با حوادث کارگاه‌های ساختمانی
۱۰	شکل (۵-۲) قدرت شواهد عوامل مؤثر بر رفتارها و حوادث نایمن
۳۲	شکل (۴-۴) مصورسازی موازنه واریانس و بایاس برای برآوردگر
۳۴	شکل (۱-۵) روش تحقیق مطالعه
۳۹	شکل (۲-۵) مثالی از پیاده‌سازی الگوریتم K نزدیک همسایه برای طبقه‌بندی داده‌ها
۴۱	شکل (۳-۵) مثالی از پیاده‌سازی الگوریتم درخت تصمیم برای طبقه‌بندی داده‌ها
۴۲	شکل (۴-۵) مثالی از پیاده‌سازی الگوریتم جنگل تصادفی برای دسته‌بندی داده‌ها
۴۳	شکل (۵-۵) ابرصفحه‌ای با حداکثر حاشیه برای یک ماشین بردار پشتیبان
۴۴	شکل (۶-۵) نحوه عملکرد الگوریتم دسته‌بندی کیسه‌ای

شکل (۷-۵) مثالی از مسئله دسته‌بندی دو کلاسه	۴۵
شکل (۸-۵) نمایش صفحه پیش‌بینی به شکل سه‌بعدی	۴۶
شکل (۹-۵) شکل نمایش باقی‌مانده‌ها	۴۶
شکل (۱۰-۵) درخت ساخته شده برای متغیرهای x و باقی‌مانده r	۴۷
شکل (۱۱-۵) نمایش پیش‌بینی بروز شده	۴۸
شکل (۱۲-۵) نمایش باقی‌مانده‌های بروز شده	۴۹
شکل (۱۳-۵) درخت ساخته شده برای متغیرهای x و باقی‌مانده r بروز شده	۴۹
شکل (۱۴-۵) نمایش پیش‌بینی بروز شده	۴۹
شکل (۱۶-۵) فرآیند الگوریتم تقویتی گرادیان	۵۰
شکل (۱۷-۵) روند بهینه‌سازی در الگوریتم XGboost	۵۰
شکل (۱۸-۵) نحوه تجمیع مدل‌ها و پیش‌بینی الگوریتم دسته‌بندی بر اساس رأی‌گیری	۵۱
شکل (۱۹-۵) نحوه عملکرد اعتبارسنجی متقابل مونت کارلو	۵۳
شکل (۲۰-۵) همبستگی ویژگی‌ها در مجموعه داده	۵۶

فهرست جداول

جدول (۱-۲) مرور ادبیات	۱۳
جدول (۱-۳) شکاف تحقیقاتی	۲۴
جدول (۱-۵) تعریف ویژگی‌های مجموعه داده پاک‌سازی شده	۳۳
جدول (۲-۵) ماتریس اغتشاش	۵۲
جدول (۳-۵) ترتیب قوانین انجمنی استخراج شده بر اساس معیار پشتیبانی از زیاد به کم	۶۴
جدول (۴-۵) قوانین انجمنی استخراج شده از مجموعه داده، بر اساس معیار Lift	۶۴
جدول (۵-۵) مقایسه الگوریتم‌های یادگیری ماشین	۶۵

فهرست نمودارها

نمودار (۱-۲)	نمودار دایره‌ای شیوع آسیب‌های ناشی از کار در بین کارکنان ساختمانی	۳
نمودار (۱-۴)	منحنی بیش برآزش بر اساس چندجمله‌ای مرتبه ۱۵	۳۰
نمودار (۲-۴)	منحنی کم‌برآزش بر اساس چندجمله‌ای مرتبه ۱	۳۱
نمودار (۳-۴)	منحنی برآزش مناسب بر اساس چندجمله‌ای مرتبه ۴	۳۱
نمودار (۱-۵)	مثالی از منحنی رگرسیون لجستیک	۳۹
نمودار (۲-۵)	منحنی مشخصه عملکرد برای سه روش مختلف دسته‌بندی	۵۴
نمودار (۳-۵)	نواحی مطلوب و نامطلوب در منحنی ROC	۵۵
نمودار (۴-۵)	متغیر هدف مطالعه	۵۷
نمودار (۵-۵)	نمودار نواحی آسیب‌دیده در بدن	۵۷
نمودار (۶-۵)	نمودار نواحی آسیب‌دیده در بدن به تفکیک کشنده و غیرکشنده بودن حادثه	۵۸
نمودار (۷-۵)	نمودار میزان ماهیت آسیب در مجموعه داده	۵۸
نمودار (۸-۵)	نمودار میزان ماهیت آسیب به تفکیک کشنده یا غیرکشنده بودن حادثه	۵۹
نمودار (۹-۵)	نمودار میزان عوامل محیطی در مجموعه داده	۶۰
نمودار (۱۰-۵)	نمودار میزان عوامل محیطی به تفکیک کشنده یا غیرکشنده بودن حادثه	۶۰
نمودار (۱۱-۵)	نمودار عوامل انسانی مؤثر بر حوادث ساختمانی	۶۱
نمودار (۱۲-۵)	نمودار میزان عوامل انسانی مؤثر بر حوادث ساختمانی به تفکیک کشنده یا غیرکشنده بودن حوادث	۶۲
نمودار (۱۳-۵)	نمودار نوع وظیفه اختصاص یافته به کارگر به تفکیک کشنده یا غیرکشنده بودن حادثه	۶۲
نمودار (۱۴-۵)	نمودار تعداد حوادث ساختمانی در مجموعه داده طی روزهای هفته	۶۳
نمودار (۱۵-۵)	نمودار توزیع چگالی احتمال ساعات وقوع حوادث	۶۳
نمودار (۱۶-۵)	منحنی ROC برای الگوریتم‌های به کار گرفته شده	۶۵

۱- مقدمه

۱-۱- مقدمه

علیرغم معرفی اقدامات پیشگیرانه ایمنی متعدد در دهه های اخیر، ایمنی شغلی در صنعت ساخت و ساز همچنان نیازمند بهبود و پیشرفت است [۱]. صنعت ساخت و ساز دارای خطرناک ترین شرایط کاری است [۲]. در نتیجه، سطح ریسک آن در بسیاری از کشورها بالاترین سطح در نظر گرفته می شود [۳]. اشتغال در ساخت و ساز تقریباً ۷ درصد از نیروی کار جهانی را تشکیل می دهد، در حالی که نزدیک به ۳۵ درصد از کارگران جهان هر ساله در تصادفات مرگبار جان خود را از دست می دهند که منجر به کشته شدن حدود ۱۰۰.۰۰۰ کارگر در کارگاه های ساختمانی می شود [۴].

۱-۲- ضرورت انجام تحقیق

طبق گزارش سازمان بین المللی کار، حوادث شغلی حدود ۴ درصد از تولید ناخالص داخلی سالانه یا ۱.۲۵ تریلیون دلار برای اقتصاد هزینه دارد. این حوادث نه تنها باعث مشکلات جدی ایمنی و بهداشتی می شود، بلکه منجر به خسارات مالی هنگفتی نیز می شود [۵]. در نتیجه، تجزیه و تحلیل علل حوادث ساختمانی و آسیب های ناشی از آن برای بهبود عملکرد ایمنی بسیار مهم است.

نرخ بالای تصادف در صنعت ساخت و ساز را می توان به پیچیدگی این صنعت نسبت داد. فرآیند ساخت و ساز شامل فعالیت های متعددی است که ذینفعان مختلف را درگیر می کند. فعالیت های پروژه به شدت بر نیروی انسانی متکی است. شرایط جسمی و روحی کارگران مستعد حوادث شغلی است. علاوه بر این، تنوع ساخت و ساز با بزرگ تر و پیچیده تر شدن مقیاس ساخت و ساز بیشتر می شود. نوع و تعداد کارگران مورد نیاز نیز پشت سر هم رشد می کنند، بدین همین ترتیب موضوعات مدیریتی گسترده افزایش می یابد. در نتیجه، ریسک مرتبط با پروژه های ساختمانی افزایش می یابد [۶]. به همین دلیل است که عملیات ساخت و ساز خطرناک، پیچیده، نامطمئن و کار فشرده شناخته می شود [۷]. بنابراین توسعه رویکردهای پیشرفته برای درک وابستگی متقابل و ترکیب عوامل مرگبار یا عوامل ایمنی به دلیل پیچیدگی و انتزاع ذاتی آنها مورد نیاز است [۸].

تعیین علل صحیح حادثه ساختمانی کار دشواری است. حجم زیادی از اطلاعات به طور منظم در قالب قطعات داده جمع آوری می شود [۹]. تصور می شود که یادگیری از تصادفات روشی مؤثر برای پیشگیری از آسیب های آینده باشد [۱۰]. با بررسی ادبیات در این حوزه، در می یابیم که روش های مبتنی بر داده کاوی و مدل سازی عوامل تأثیرگذار بر حوادث ساختمانی و آسیب های ناشی از آن، با استفاده از یادگیری ماشین و پیش بینی متغیرهای مختلف یک حادثه، همچنان نیاز به توسعه دارند.

۱-۳- اهداف تحقیق

مجموعه داده مورد بررسی، اطلاعات حوادث ساختمانی است که توسط اداره ایمنی بهداشت حرفه ای^۲ ایالات متحده بین سال های ۲۰۱۵ تا ۲۰۱۷ منتشر شده است. هدف اصلی این مطالعه، شناسایی عوامل

^۲ OSHA: Occupational Safety and Health Administration

مؤثر بر حوادث ساختمانی، تجزیه و تحلیل اکتشافی در داده‌ها و مدل‌سازی و پیش‌بینی حوادث ساختمانی با استفاده از الگوریتم‌های یادگیری ماشین می‌باشد.

۱-۴- طرح مسئله

آنچه که به عنوان مسئله و سؤال برای این پژوهش در نظر گرفته شده است، به شرح زیر است:

۱. کدام یک از عوامل بیشترین تأثیر را بر یک حادثه ساختمانی دارند؟
۲. مدل‌های یادگیری ماشین تا چه میزان در مدل‌سازی و پیش‌بینی حوادث و آسیب‌های ناشی از آن در صنعت ساخت‌وساز کاربرد دارند؟

۲- پیشینه تحقیق

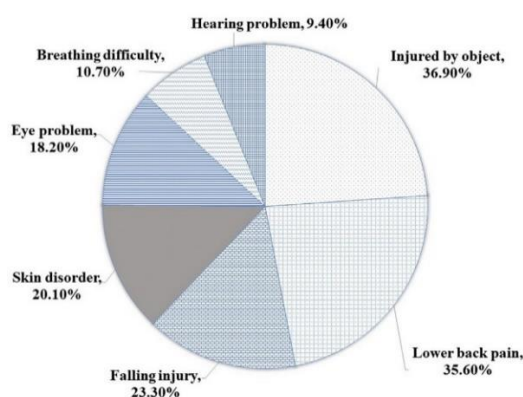
۲-۱- مرور ادبیات

۲-۱-۱-

این مقاله بر رفتارهای نایمن کارگران ساختمانی تمرکز دارد و یک مدل شناختی از رفتارهای نایمن کارگران ساختمانی را با اتخاذ یک فرم پنج مرحله‌ای توسعه می‌دهد. پنج مرحله این مدل شامل به‌دست‌آوردن اطلاعات، درک اطلاعات، درک پاسخ‌ها، انتخاب پاسخ و اقدام است. در مقایسه با سایر مدل‌های شناختی، بر اساس در نظر گرفتن کافی ویژگی‌های صنعت ساخت‌وساز، مدل ارائه شده در این پژوهش بر مرحله اول «کسب اطلاعات» و مرحله چهارم «انتخاب پاسخ» تأکید و تشریح می‌کند. کار ساخت‌وساز به طور ضعیفی تعریف و تنظیم شده است، بنابراین کارگران به‌شدت به مشاهدات و تجربیات خود برای شناسایی خطرات متکی هستند؛ بنابراین، مرحله اول شامل دو حالت مختلف (جستجوی عمدی و غیرعمدی) است تا به طور کامل فرایندهای جستجو و شکست‌های احتمالی در پشت هر فرایند را منعکس کند. علاوه بر مشاهدات مستقل، کارگران همچنین باید تصمیمات سریعی برای مقابله با خطرات توسط خود بگیرند [۱۱].

۲-۱-۲-

این مطالعه شیوع آسیب و عوامل مرتبط با آن را در بین کارگران ساختمانی در جنوب غربی اتیوپی ارزیابی کرده است. معاینه فیزیکی افراد مورد مطالعه در رابطه با آسیب شغلی انجام شده و شیوع کلی صدمات ناشی از کار در یک سال گذشته ۴۱.۴ درصد بوده است. پنج آسیب اول توسط جسم، کمردرد، آسیب ناشی از سقوط، اختلال پوستی و مشکل چشم پدید آمده‌اند. یافته‌های این مطالعه نشان داد که تهیه تجهیزات ایمنی و ارتقای استفاده از آن، اجتناب از اضافه‌بار کاری و کنترل استفاده از خات در محل کار می‌تواند به حداقل رساندن آسیب‌های ناشی از کار و بیماری‌های شغلی برای تضمین ایمنی کارگاه ساختمانی کمک کند [۱۲].



نمودار (۱-۲) نمودار دایره‌ای شیوع آسیب‌های ناشی از کار در بین کارکنان ساختمانی

۲-۱-۳-

تاکنون، مطالعات برای پیش‌بینی حوادث ساختمانی عمدتاً با روش‌های تحلیل آماری انجام شده است که مدل‌های خطی مانند تحلیل رگرسیون و سری زمانی را فرض می‌کنند که با ورود ویژگی‌های غیرخطی حوادث ایمنی ساختمانی به مدل، دشوار می‌شوند. تکنیک‌های یادگیری عمیق برای

تجزیه و تحلیل ویژگی‌های غیرخطی عوامل تأثیرگذار پیچیده استفاده می‌شود و هدف این مطالعه، توسعه یک مدل یادگیری عمیق برای پیش‌بینی حوادث ایمنی برای ساختمان‌سازی پایدار است. در این مطالعه، ۱۷۶۶ مورد حادثه واقعی توسط سازمان ایمنی شغلی کره طی دوره ۱۰ ساله از ۲۰۱۰ تا ۲۰۱۹ جمع‌آوری شد. هشت عامل مؤثر بر پیش‌بینی حادثه انتخاب شدند. سپس، قدرت پیش‌بینی بین مدل‌های یادگیری عمیق و مدل‌های رگرسیون چندگانه مرسوم مقایسه شد. نهایتاً، شبکه عصبی عمیق، قدرت پیش‌بینی را در میانگین خطای مطلق و در ریشه میانگین مربعات خطا در مقایسه با یک مدل رگرسیون چندگانه معمولی بهبود بخشید [۱۳].

۲-۱-۴-

هدف از این مطالعه بررسی امکان‌سنجی استفاده از یادگیری موجک-ماشین ترکیبی برای پیش‌بینی وقوع حوادث در صنعت ساختمان است. مشکل اصلی که توسط این مقاله حل می‌شود، کشف پیش‌بینی تصادف در پروژه‌های ساختمانی با استفاده از یادگیری موجک-ماشین ترکیبی است. این مطالعه بررسی کرد که چگونه الگوریتم‌های یادگیری ماشین ممکن است نقش مهمی در پیش‌بینی حوادث مربوط به ساخت‌وساز داشته باشند. نتایج نشان می‌دهد که استفاده از هیبریدهای نوآورانه یادگیری موجک-ماشین ممکن است به طور قابل توجهی نرخ تصادفات ساخت‌وساز را کاهش دهد. یادگیری ماشین ممکن است به دو روش مجزا برای بهبود ایمنی ساخت‌وساز استفاده شود. یک سیستم یادگیری ممکن است به عنوان ابزاری برای یادگیری و کسب اطلاعات در مورد حوادث استفاده شود. همچنین می‌توان از یک سیستم یادگیری به عنوان یک ابزار پشتیبانی تصمیم استفاده کرد تا با پیش‌بینی ویژگی‌های آن‌ها از حوادث جلوگیری شود [۱۴].

۲-۱-۵-

اهداف این مطالعه، بدین شرح است. اولین هدف، شناسایی علل و عوامل حوادث در صنعت ساخت‌وساز مالزی است که در سال‌های ۲۰۰۹-۲۰۰۰، ۴۲۷۷۵ حادثه را در صنعت ساخت‌وساز مالزی گزارش کرد. میانگین تعداد تصادفات در این ۱۰ سال برای مردان ۳۸۹۴ مورد و برای زنان ۳۸۴ مورد بود. با توجه به داده‌های به دست آمده رایج‌ترین نوع حادثه مورد اصابت قرار گرفتن با اشیاء یا ضربه خوردن با آن‌ها است که در هنگام حرکت تجهیزات ساخت‌وساز یا برخورد با کارگران اتفاق می‌افتد. بیش از ۵۰ درصد از تمام حوادث در صنعت مربوط به این دو دسته است. همچنین، در ادامه افتادن و کامیون‌ها یک عامل ارشد در قسمت حوادث صنعت ساخت‌وساز مالزی به حساب آمد. برای هدف دوم یک رویکرد تحقیق متفاوت برای شناسایی الگوهای قانونی اتخاذ شده، داده‌های مربوط به ۳۰ پرونده دادگاه از مجله حقوق مالزیایی و گزارش‌های مربوطه در لکسیس-نکسیس بازیابی شد [۱۵].

۲-۱-۶-

هدف این مطالعه کشف و آزمایش امکان‌سنجی با استفاده از داده‌های شتاب‌سنج و یادگیری ماشین برای طبقه‌بندی فعالیت کارگران ساختمانی در کنار گروه بزرگی از فعالیت‌های ساخت‌وساز پیچیده است. به این ترتیب، یک روش یادگیری ماشین برای آموزش و ارزیابی ۱۳ طبقه‌بندی با استفاده از ویژگی‌های مصنوعی استخراج شده از بخش‌های داده شتاب‌سنج خام توسعه داده شده که آن‌ها را تحت

۳ سناریو که این داده‌ها از شتاب‌سنج‌های پوشیدنی که هم بر روی مچ‌ها و پای کارگران قرار داشت به‌دست آمد. در ادامه با طبقه‌بندی و با تجزیه و تحلیل بهینه‌ترین چارچوب، انتخاب و اعمال شد. در مجموع ۶ موضوع مدار فعالیت اجرا که هر فرد با میانگین حدود ۸.۶ دقیقه در هر اجرا مدار را انجام داد که در مجموع ۹۳۰۸ نقطه به دست آمد. نتایج به‌دست آمده نشان می‌دهد که شتاب‌سنج‌ها می‌توانند برای ایجاد یک سیستم قوی برای تشخیص خودکار مجموعه بزرگی از فعالیت‌های پیچیده کارگران ساختمانی مورد استفاده قرار گیرند [۱۶].

۲-۱-۷-

هدف این مقاله، تجزیه و تحلیل حوادث شغلی رخ داده در سایت‌های ساختمانی اسپانیا با در نظر گرفتن ملیت کارگران است. وزارت کار، مهاجرت و تأمین اجتماعی داده‌های ناشناس تمام حوادث محل کار در اسپانیا را طی دوره ۲۰۰۳-۲۰۱۵ در اختیار قرار داد. پس از فیلترینگ مطالعه بر نمونه‌ای از ۱۵۲۵۸۶۵ تصادف از مجموع تصادف‌های گزارش شده در اسپانیا بین سال‌های ۲۰۰۳ و ۲۰۱۵ تمرکز شد. ویژگی‌هایی از جمله سن، شغل، مدت زمان فعالیت در شرکت، سبک شرکت، مقدار ساعت کاری، شرایط کاری و غیره در این مقاله در نظر گرفته شده‌اند. در نهایت نیز از قوانین تلازمی یا انجمنی برای دستیابی به نتیجه مقاله استفاده شده است. به عنوان یک نتیجه‌گیری کلی، جو ایمنی مثبت می‌تواند کارگران را برای رعایت مقررات ایمنی و استفاده از روش‌های کار ایمن ترغیب کند. برای اینکه این امر در یک محیط چندفرهنگی اتفاق بیفتد، موانع زبانی و فرهنگی باید از کل زنجیره تولید حذف شود تا اطمینان حاصل شود که اطلاعات بهداشتی و ایمنی به درستی منتقل می‌شود [۱۷].

۲-۱-۸-

موضوع مقاله، تجزیه و تحلیل تئوری سیستم‌ها برای پیشگیری از حوادث ساختمانی با مطالعه موردی گزارش‌های حوادث سازمان ایمنی و بهداشت شغلی ایالات متحده می‌باشد. این مطالعه به تئوری سیستم‌ها را در نظریه دومینوی هاینریش^۳ به منظور بررسی روابط متقابل ریسک‌ها و شکستن زنجیره علل حوادث می‌پردازد. این مقاله با روش تجزیه و تحلیل تجربی و آماری، ۹۳۵۸ حادثه که در صنعت ساخت و ساز ایالات متحده بین سال‌های ۲۰۰۲ تا ۲۰۱۱ رخ داده است را بررسی کرده و روابط بین حوادث و عوامل خطر کلیدی، ماهیت صدمات ساختمانی و عناصر آسیب را زیر ذره‌بین قرار داده است. در این مطالعه همبستگی بین عوامل خطر مختلف از جمله شرایط محیطی، رفتار کارگر و منبع آسیب از نظر آماری شناسایی شد و در نهایت ترکیب‌های عوامل خطر کلیدی را که در حوادث نقش دارند، تعیین شدند که از مهم‌ترین آن‌ها می‌توان به مواردی چون شرایط محیطی، رفتار کارگر و منبع آسیب اشاره نمود [۱۸].

۲-۱-۹-

در این مقاله به تجزیه و تحلیل ۱۱۱۷ گزارش شاهد خبره، ارائه شده به دادگاه‌های جنایی و کارگری به منظور اهمیت خطرات ساخت و ساز و سهم آن در حوادث در ترکیه پرداخته شده است. این گزارش‌ها از تمام مناطق کشور از سال ۲۰۰۸-۱۹۷۲ پوشش داده شده. حوادث بر اساس نتیجه حادثه، زمان و

^۳ Heinrich's domino

علت اصلی حادثه، نوع ساخت و ساز، شغل قربانی، فعالیت در زمان حادثه و شخص مسئول حادثه طبقه‌بندی شدند که افتادن، برخورد با شی پرتاب یا سقوط، فروپاشی ساختاری و برق‌گرفتگی چهار رتبه اول را به خود اختصاص دادند با توجه به نتایج، ۸۸.۴ درصد حوادث مربوط به ۴ حالت حادثه گفته شده، می‌باشد که نزدیک به دو سوم اقدامات اشتباه و مسامحه کار توسط کارفرمایان انجام شده، از دلایل اصلی این ناآگاهی اول نبوده است. همچنین در ادامه فرهنگ ایمنی و ثانیاً بازرسی‌های ناکافی دولتی گفته شده است [۱۹].

۲-۱-۱۰-

هدف از این مطالعه اعمال، تأیید و مقایسه عملکرد روش‌های مختلف یادگیری ماشین در طبقه‌بندی شدت حوادث ساختمانی در چین است. در تجزیه و تحلیل نتیجه، این مقاله عوامل بحرانی شدت حادثه را تعیین می‌کند، ماتریس سردرگمی را تجزیه و تحلیل می‌کند، دلایل خطای طبقه‌بندی الگوریتم را خلاصه می‌کند و از درخت تصمیم برای تولید مجموعه‌ای از قوانین ارزیابی پیامد حادثه استفاده می‌کند. در این مطالعه، هشت تکنیک یادگیری ماشین به نام‌های رگرسیون لجستیک، درخت تصمیم، ماشین بردار پشتیبان، Naive Bayes، نزدیک‌ترین همسایه K، جنگل تصادفی، پرسپترون چندلایه و AutoML برای پیش‌بینی عواقب تصادف مورد استفاده قرار گرفته و سپس ارزیابی می‌شوند. NB قوی‌ترین روش است و به دنبال آن رگرسیون لجستیک، درخت تصمیم، و جنگل تصادفی قرار دارند. بر اساس تجزیه و تحلیل مؤلفه‌های اصلی و سایر روش‌ها، «نوع حادثه» مهم‌ترین عامل است و پس از آن «گزارش حادثه» و «فرهنگ رسیدگی و ایمنی» از اهمیت بیشتری برخوردارند [۲۰].

۲-۱-۱۱-

این مطالعه یک روش خودکار برای تشخیص سطوح خطر درک شده کارگران ساختمانی با استفاده از سیگنال‌های فیزیولوژیکی به دست آمده از حسگرهای زیستی پوشیدنی نوع مچ‌بند در همراهی با یک الگوریتم یادگیری بانظارت ایجاد می‌کند. عملکرد مدل با سیگنال‌های فیزیولوژیکی به دست آمده از هشت کارگر ساختمانی که کار روزانه خود را انجام می‌دادند، مورد بررسی قرار گرفت. این مدل به دقت اعتبار ۸۱.۲ درصد برای تمایز بین سطوح پایین و بالای ریسک درک شده دست یافت. برای توسعه یک مدل یادگیری ماشینی برای شناسایی ادراک ریسک کارگران ساختمانی در محل، داده‌های فیزیولوژیکی (به عنوان مثال، EDA، PPG و ST) از هشت کارگر ساختمانی (دو نجار، یک کف‌ساز، چهار برق‌کار و یک سرکارگر) جمع‌آوری شد. مجموعه داده‌ها به ۱۰ قسمت تقسیم شد که در ابتدا با استفاده از ۹ قسمت آن مدل توسعه داده شد و سپس با بخش ۱۰ام صحت و دقت مدل مورد بررسی قرار گرفت [۲۱].

۲-۱-۱۲-

علت اصلی حوادث ساختمانی مرگبار، سقوط از ارتفاع است. در این مطالعه، سیستم تجزیه و تحلیل و طبقه‌بندی عوامل انسانی^۴ برای کمک به طبقه‌بندی علل اصلی حوادث مرگبار سقوط از ارتفاع در هنگ‌کنگ اتخاذ و اصلاح شد. ویژگی‌ها و علل اصلی حوادث با استفاده از تحلیل مرتبه تکرار مورد

^۴ HFACS: Human Factors Analysis and Classification System

بررسی قرار گرفت. ارتباط بین علل با استفاده از آزمون دقیق فیشر و تحلیل کلاس نهفته بررسی شد. باتوجه به نتایج تحلیل کلاس نهفته، عوامل ممکن است در چهار طبقه قرار گیرند که به نام‌های برنامه‌ریزی نادرست، تخلف، خطرات پنهان ایجاد شده توسط دیگران و کارکنان ناتوان نام‌گذاری می‌شوند. این مقاله با توسعه یک چارچوب HFACS اصلاح شده برای استخراج اطلاعات ساختاریافته از گزارش‌های حوادث ساختمانی، یافتن ویژگی‌های کلی و علل مدیریتی در پشت حوادث سقوط از ارتفاع و استفاده از تحلیل کلاس نهفته برای شناسایی ارتباطات بین علل مدیریتی، به مجموعه دانش در مدیریت ساخت‌وساز کمک می‌کند [۲۲].

۲-۱-۱۳-

هدف از این مطالعه، اضافه کردن به ادبیات نسبتاً پراکنده در مورد علت حوادث در صنعت ساخت‌وساز با شناسایی عواملی علی‌المرتبه و شناسایی ارتباط بین این عوامل می‌باشد. هفت عامل پرتکرار که در این مطالعه شناسایی شدند، عبارت‌اند از اقدامات کارگر، مدیریت ریسک، نظارت فوری، توانایی استفاده از مواد یا تجهیزات، خطرات محیطی، مهارت‌های کارگر و مدیریت پروژه. این عوامل به صورت دوه‌دو و چندگانه با یکدیگر همبستگی زیادی دارند و بر اهمیت پرداختن به ریسک در سطوح مختلف توسط بازیگران مختلف در پروژه‌های ساختمانی تأکید دارند. این مطالعه الگوهای قوی عوامل پرتکرار و ارتباط بین آن‌ها را نشان می‌دهد که بینش ارزشمندی را در مورد علت حوادث ساختمانی ارائه می‌دهد. این مطالعه نشان می‌دهد که اکثر حوادث ساختمانی چند علتی هستند و ترکیبات مختلفی از عوامل در حوادث وجود دارد. با این حال همه عوامل به یک اندازه اهمیت ندارند و اقدامات کارگران، نظارت فوری و مدیریت ریسک در پیشگیری از حوادث نقش بسزایی دارند [۲۳].

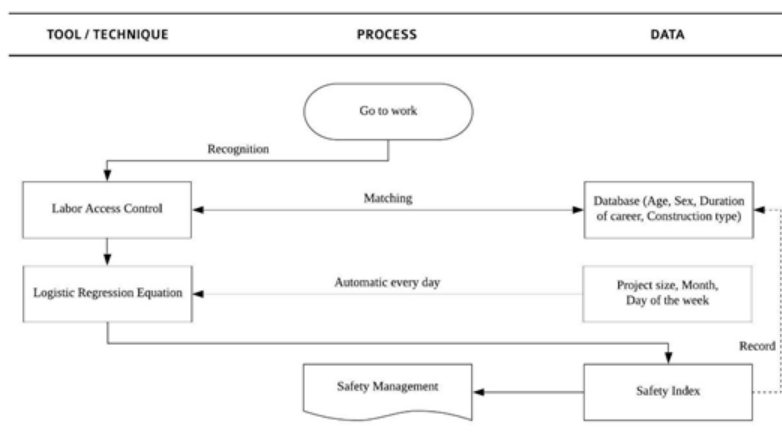
۲-۱-۱۴-

این مطالعه با استفاده از فناوری‌های یادگیری ماشین به بررسی وابستگی متقابل و رابطه ترکیبی بین علل حوادث کشنده در صنعت ساختمان می‌پردازد. اهداف این مطالعه عبارت‌اند از دسته‌بندی انواع مرگ‌ومیر ساختمانی بر اساس ویژگی‌های مرگبار آن‌ها، شناسایی روابط سلسله‌مراتبی بین علل مرگبار، و تعیین و کمی کردن ترکیبات علل بحرانی مرتبط با هر نوع حادثه مرگبار. برای رسیدن به این اهداف، یک مدل پیش‌بینی تصادفات مرگبار، بر اساس ۳۴ ویژگی کشنده برای استخراج رابطه ایجاد شد. این مقاله سپس به بهینه‌سازی مدل یادگیری ماشین برای تعیین روابط سلسله‌مراتبی بین هر حوادث مرگبار و علل مربوط به آن‌ها با استفاده از روش اهمیت جایگشت پرداخته است. در مجموع ۲۱ علت مؤثر بر مرگ‌ومیر شناسایی شدند. روش پیشنهادی این تحقیق به مدیران پروژه در شناسایی عوامل مرگبار بحرانی با اطلاعات قابل اعتماد کمک می‌کند [۲۴].

۲-۱-۱۵-

در جمهوری کره، نرخ تصادفات صنعتی و تعداد تلفات در صنعت ساختمان به مدت شش سال از سال ۲۰۱۱ به افزایش خود ادامه داده است. برای جلوگیری از حوادث، وزارت کار داده‌های تحلیلی مختلفی را از طریق گزارش سالانه حوادث صنعتی ارائه می‌کند. این مقاله داده‌های ۲۵۶۱۳ قربانی حوادث صنعتی صنعت ساختمان در کره (۲۵۱۱۴ تلفات تصادف، ۴۹۹ مرگ ناشی از تصادف) را

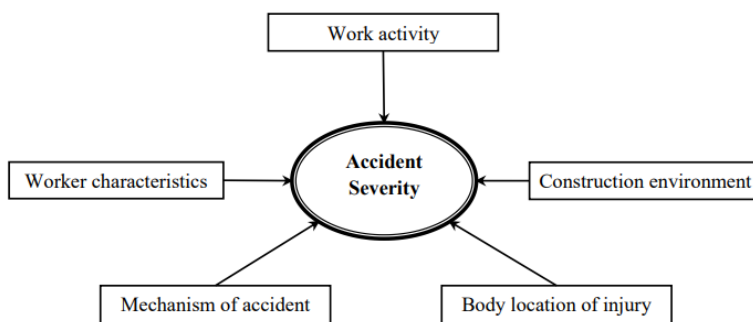
تجزیه و تحلیل کرد. این مطالعه داده‌ها را بر اساس مدل رگرسیون لجستیک مدل‌سازی کرد. در مورد تصادف منجر به فوت، احتمال وقوع در زمانی که سن بالاتر باشد، مدت زمان شغلی طولانی‌تر، اندازه پروژه بزرگ‌تر است، در روز یکشنبه و ماه آپریل بیش‌تر و احتمال وقوع در پروژه معماری کمتر است. اطلاعات بیومتریک به دستگاه تأییدکننده کار کارگر وارد می‌شود، شاخص ریسک محاسبه و گزارش می‌شود. بر اساس این شاخص، مدیر ایمنی می‌تواند گروه‌های خطر روزانه را به طور کارآمد طبقه‌بندی و مدیریت کند که انتظار می‌رود از حوادث مرگبار جلوگیری شود [۲۵].



شکل (۱-۲) دیاگرام پردازش اطلاعات کارگر

۱۶-۱-۲-

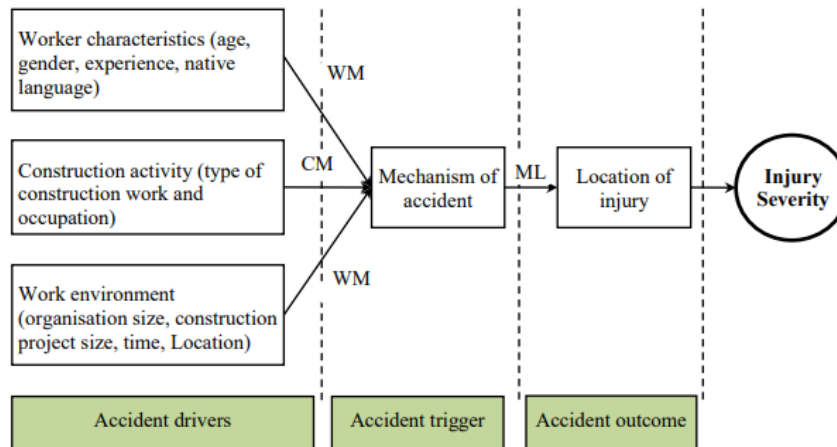
موضوع مقاله، عوامل مرتبط با شدت حوادث ساختمانی با مطالعه موردی در استرالیای جنوبی می‌باشد. این مطالعه به بررسی شدت حوادث ساختمانی و عوامل تأثیرگذار بر آن می‌پردازد. این مطالعه، ۲۴۷۶۴ حادثه ساختمانی بین سال‌های ۲۰۰۲ تا ۲۰۱۱ در استرالیای جنوبی تجزیه و تحلیل کرده است. عواملی چون سن، میزان تجربه، جنسیت، اندازه سازمان، مکان و اندازه پروژه، مکانیسم حادثه، محل آسیب بدن و زمان حادثه مورد بررسی قرار گرفته‌اند که نشان می‌دهند چرا برخی از حوادث ساختمانی تنها به شدت جزئی منجر می‌شوند در حالی که برخی دیگر کشنده هستند. در شکل زیر عوامل مؤثر بر شدت حوادث ساختمانی ارائه شده است.



شکل (۲-۲) عوامل مؤثر بر شدت حوادث ساختمانی

این مطالعه نشان می‌دهد که با افزایش سن کارگر، شدت حوادث نیز افزایش می‌یابد. همچنین زنان در میان حوادث حضور کم‌رنگی دارند و شدت حوادث نیز در آن‌ها کمتر است. این مطالعه بیان می‌کند

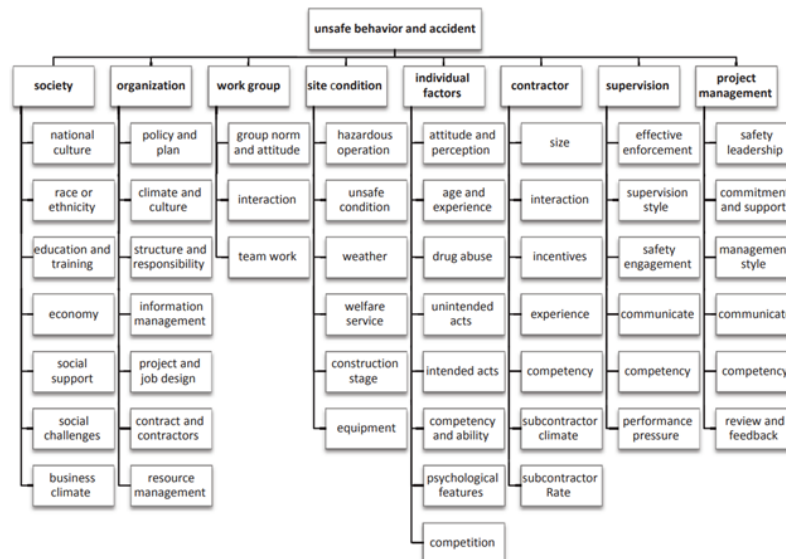
که اندازه سازمان، مکان و اندازه پروژه با شدت حادثه در ارتباط است و زمان حادثه ارتباط کمی با شدت حادثه دارد [۲۶].



شکل (۳-۲) مدل مفهومی ارائه شده برای شدت حوادث ساختمانی

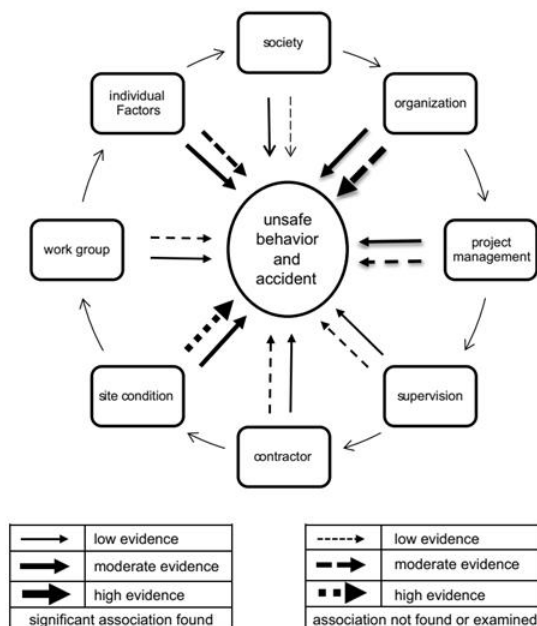
۲-۱-۱۷-

مطالعه حاضر تصویر کیفی دقیقی از ماهیت و طیف عوامل مؤثر بر رفتارهای نایمن و حوادث در سایت‌های ساختمانی ایجاد کرده است. این پژوهش داده‌های ۵۶ مطالعه را استخراج کرده و در مطالعات بررسی شده، ۹۴ متغیر برای تعیین ارتباط آن‌ها با رفتارهای نایمن و حوادث در کارگاه‌های ساختمانی مورد آزمایش قرار گرفتند. این متغیرها در ۵۰ موضوع و هشت عامل کمک‌کننده، یعنی جامعه، سازمان، مدیریت پروژه، نظارت، پیمانکار، شرایط سایت، گروه کاری و ویژگی‌های فردی گروه‌بندی شدند. شکل زیر، گروه‌بندی مضامین را در قالب یک مدل مفهومی خلاصه می‌کند.



شکل (۴-۲) مدل مفهومی مضامین در ارتباط با حوادث کارگاه‌های ساختمانی

این مطالعه از متاآنالیز برای بررسی روابط بین عوامل مؤثر و رفتارها و حوادث نایمن استفاده کرده است. نتایج نشان داده است که به ترتیب عوامل سازمان، مدیریت پروژه، وضعیت سایت، و ویژگی‌های شخصیت فردی شواهد بالایی از ارتباط با رفتارهای نایمن و تصادفات دارند. علاوه بر این، شکل زیر قدرت شواهد عوامل مؤثر بر رفتارها و حوادث نایمن را نشان می‌دهد [۲۷].



شکل (۲-۵) قدرت شواهد عوامل مؤثر بر رفتارها و حوادث ناایمن

۱۸-۱-۲-

این مطالعه با هدف بررسی علت ریشه‌ای نرخ بالای تصادفات در صنعت ساخت‌وساز از دیدگاه مدیریت استرس انجام شده است. مدل مفهومی شامل عوامل استرس‌زای شغلی، استرس (فیزیکی و روانی)، رفتار ایمنی و حوادث است. به‌منظور بررسی روابط بین این عوامل در بین کارگران ساختمانی، پرسش‌نامه‌ای طراحی شد که در مجموع ۵۰۰ نظرسنجی از طریق تماس شخصی از کارگران جمع‌آوری شد. سپس تحلیل‌های آماری بر روی داده‌ها اعمال شد. نتایج که با تحلیل همبستگی پیرسون و رگرسیون به‌دست‌آمده نشان می‌دهد استرس فیزیکی از طریق اطمینان شغلی، حمایت همکاران و تجهیزات ایمنی همراه می‌شود و همین‌طور استرس روانی به‌طور مثبت توسط حمایت سرپرست همراه می‌شود. همچنین از اطمینان شغلی و حمایت همکاران، و تجهیزات ایمنی می‌تواند استرس فیزیکی کارگران ساختمانی را به حداقل برسانند، نام برده شد [۲۸].

۱۹-۱-۲-

موضوع مطالعه، ادغام مهندسی ویژگی، الگوریتم ژنتیک و روش‌های یادگیری ماشین مبتنی بر درخت برای پیش‌بینی وضعیت ناتوانی کارگران ساختمانی پس از حادثه می‌باشد. هدف این مقاله، ایجاد یک چارچوب جامع برای پیش‌بینی وضعیت ناتوانی کارگران ساختمانی پس از حادثه است. مجموعه داده بررسی شده شامل ۴۷۹۳۸ حادثه ساختمانی ثبت شده در ترکیه است. پیش‌بینی‌ها از طریق چهار مدل یادگیری ماشین مبتنی بر درخت انجام شده است. یافته‌های این مطالعه نشان می‌دهد که کارگرانی که در معرض دماهای بالا و مواد شیمیایی قرار می‌گیرند، احتمال ناتوانی دائمی آن‌ها در حوادث بیشتر است. نتایج این مقاله به پیش‌بینی وضعیت ناتوانی کارگران ساختمانی پس از حادثه کمک می‌کند که منجر به محیط کار ایمن‌تر و برنامه‌ریزی بهتر در پروژه‌های ساختمانی می‌شود [۳].

۲-۱-۲۰-

هدف از این مطالعه، توسعه یک مدل پیش‌بینی است که با استفاده از یادگیری ماشین بر اساس داده‌های حوادث صنعتی جمع‌آوری شده توسط وزارت کار جمهوری کره از سال ۲۰۱۱ تا ۲۰۱۶، خطر بالقوه حوادث مرگ‌بار در کارگاه‌های ساختمانی را شناسایی می‌کند. این تحقیق اطلاعات ۱۳۷۳۲۳ مجروح و ۲۸۴۶ مورد مرگ را شامل می‌شود. پس از توصیف توزیع مجموعه داده، روش‌های یادگیری ماشین، مانند رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، و تحلیل‌های AdaBoost با استخراج متغیرهای اصلی مؤثر بر طبقه‌بندی در هر الگوریتم استفاده شدند. روش جنگل تصادفی از نظر طبقه‌بندی کارگرانی که ممکن است با خطر مرگ‌ومیر بالایی مواجه شوند، نرخ پیش‌بینی موفق‌تری نسبت به سایر روش‌ها داشته است. ماه (فصل) و طول اشتغال، تأثیرگذارترین عوامل هستند. به دنبال آن، سن، روز هفته و طول خدمت منجر به مرگ می‌شوند. نتایج به‌دست‌آمده می‌تواند به پیشگیری از حوادث ساختمانی کمک کند [۲۹].

۲-۱-۲۱-

عوامل زیادی در بروز آسیب‌های شغلی نقش دارند که در اکثر مطالعات تقسیم‌بندی و جدا شده‌اند. هدف این مطالعه، بررسی رابطه بین آسیب‌های ناشی از کار و عوامل متعدد شغلی و غیر شغلی در بین کارگران ساختمانی در ایالات متحده آمریکا است. داده‌های سال‌های ۱۹۸۸-۲۰۰۰ مرکز NLSY۷۹ بررسی شد. در معرض شغل بودن و رفتارهای بهداشتی مورد بررسی قرار گرفت و به‌عنوان متغیرهای مستقل در چهار مدل رگرسیون لجستیک چندمتغیره برای شناسایی ارتباط با آسیب‌های شغلی استفاده شد. پس از کنترل متغیرهای جمعیت‌شناختی، احتمال آسیب‌های شغلی ۱۸ درصد در ساختمانی بیشتر از غیرساختمانی بود. تلاش‌های فیزیکی شغلی، مشاغل متعدد و ساعات کاری طولانی باعث افزایش خطر در ساخت‌وساز می‌شوند. سیگار کشیدن، چاقی یا اضافه‌وزن و مصرف کوکائین به طور قابل‌توجهی خطر آسیب ناشی از کار را زمانی که فاکتورهای جمعیتی و شغلی ثابت نگه داشته شدند، افزایش داد [۳۰].

۲-۱-۲۲-

هدف این مقاله استخراج و رابطه بین عواملی که بر انواع حوادث شغلی در مناطق ساخت‌وساز کشور کره تأثیرگذارند، می‌باشد. در این مطالعه برای طبقه‌بندی و پیش‌بینی انواع حوادث شغلی از یک جنگل تصادفی استفاده شده که داده‌ها از آژانس سلامت و ایمنی شغلی کره جمع‌آوری شده که به دلیل وزن‌دهی نادرست و نبود دیتاهای مربوط به آب‌وهوا، به‌صورت جداگانه جمع‌آوری، و با هم ادغام شدند و مدل جنگل تصادفی با دقت پیش‌بینی ۷۱.۳ درصد به وجود آمد. مقادیر اهمیت با استفاده از مدل جنگل تصادفی به دست آمدند که به ترتیب: افتادن از ارتفاع به‌عنوان رایج‌ترین و به دنبال آن برخورد اشیاء، غلتک و اشیاء در حال سقوط بیشترین سهم از حوادث ساختمانی را به خود اختصاص داده‌اند؛ همچنین از دلایل بروز این حوادث تا حد زیادی ناشی از خطای انسانی و نظارت نامناسب و عدم وجود تأسیسات ایمنی نصب‌شده یا تجهیزات حفاظتی بود [۳۱].

۲-۱-۲۳-

اعمال نایمن کارگران (قضاوت یا عملکرد نادرست) زمانی که با شرایط کاری نایمن (شرایط سطح کار یا آبوهوا) در یک مکان ساختمان سازی ترکیب شود، به علت اصلی حوادث ساختمانی تبدیل می شود. هدف کلی این مقاله، کشف راه هایی برای جلوگیری از اعمال نایمن کارگران و کاهش احتمال وقوع حوادث ساختمانی است و هدف این مطالعه به طور خاص (۱) درک روابط بین عوامل خطر مرتبط با رفتار انسانی و عوامل خطر مرتبط با شرایط کاری، (۲) شناسایی عوامل مهم رفتار و شرایط و تأثیرات آن ها بر انواع حوادث (ضربه خوردگی، افتادن، شوک، استنشاق/بلع/جذب، یا نارسایی تنفسی) و شدت آسیب (مرگ و میر، بستری شدن در بیمارستان یا غیره) و (۳) تحلیل رابطه اساسی حادثه-آسیب، است. ۹۳۵۸ حادثه که در صنعت ساخت و ساز ایالات متحده بین سال های ۲۰۰۲ تا ۲۰۱۱ رخ داده بود، بررسی و مجموعاً ۱۷ ارتباط معنادار بین عوامل رفتار و شرایط شناسایی شد. نتایج تحقیق کمک می کند تا با حذف شرایط کاری نایمن، اعمال ناامن خاص کارگران کنترل شود [۱۸].

۲-۱-۲۴-

برای پرداختن به این موضوع، این مطالعه از رویکرد یادگیری ماشینی برای توسعه مدلی استفاده کرد که می تواند یک شاخص پیشرو ایمنی تولید کند که در پیش بینی خطرات ایمنی کارگاه های ساخت و ساز مفید است. با استفاده از مجموعه داده هفت ساله یک شرکت ساختمانی بزرگ و معتبر، چارچوب CRISP-DM شناخته شده در صنعت برای اعمال سیستماتیک فرآیندها و تکنیک های ML استفاده شد. به عنوان بخشی از مطالعه، لیستی از ۳۳ متغیر ورودی (همچنین به عنوان ویژگی ها یا متغیرهای مستقل شناخته می شود) شناسایی شد. پنج الگوریتم ML برای آموزش و اعتبارسنجی از طریق اعتبارسنجی متقاطع ترک یک استفاده شد. کاربرد ML در مدیریت ایمنی ساخت و ساز، به ویژه در حوزه شاخص پیشرو ایمنی، نسبتاً جدید است. این مدل به طور بالقوه می تواند برای راهنمایی مدیران و متخصصان ایمنی و بهداشت محیط کار برای شناسایی مکان های پرخطر استفاده می شود تا بتوان از حوادث پیشگیرانه پیشگیری کرد [۳۲].

۲-۱-۲۵-

موضوع مقاله، طبقه بندی حوادث سایت ساخت و ساز مبتنی بر متن کاوی با استفاده از یادگیری ماشینی تحت نظارت ترکیبی است. این مطالعه مدلی ترکیبی از مدل های تحت نظارت یادگیری ماشینی را ارائه می دهد که از پردازش زبان طبیعی^۵ برای پیش پردازش داده های متنی برای طبقه بندی حوادث استفاده شده است. این مقاله از ۱۰۰۰ گزارش واقعی حوادث ساخت و ساز به عنوان مطالعه موردی استفاده کرده است. در این مطالعه مدل SRGU^۶ به عنوان مدل ترکیبی از دو مدل GRU^۷ و SOS^۸ ارائه شده است. با استفاده از مدل ترکیبی پیشنهادی در این مقاله می توان از اشتباهات طبقه بندی و سوگیری ناشی از خطای انسانی جلوگیری کرد [۳۳].

^۵ NLP: Natural Language Processing

^۶ Symbiotic Gated Recurrent Unit

^۷ Gated Recurrent Unit

^۸ Symbiotic Organisms Search

۲-۲- جدول مرور ادبیات

جدول (۱-۲) مرور ادبیات

ردیف	عنوان مقاله	سال چاپ	نام مجله	نویسندگان	روش مورد استفاده	روش تحقیق مقاله	نتایج و دستاوردهای مقاله
۱	A Cognitive Model of Construction Workers' Unsafe Behaviours	۲۰۱۶	Journal of Construction Engineering and Management	Fang et al.	مدل شناختی (مقاله مروری)	این مقاله بر اساس نظریه‌های روان‌شناسی شناختی و اجتماعی، همراه با مدل‌های علت حوادث موجود، یک مدل پنج‌مرحله‌ای را به‌منظور توصیف فرآیندهای شناختی کارگران ساختمانی در هنگام مواجهه با خطرات بالقوه در سایت‌های ساختمانی توسعه داده است. این پنج مرحله شامل به‌دست‌آوردن اطلاعات، درک اطلاعات، درک پاسخ‌ها، انتخاب پاسخ و اقدام است. سپس نارسایی‌های شناختی مختلف هر مرحله توسط مدل توسعه‌یافته مورد مطالعه قرار گرفت.	بر اساس مدل توسعه‌یافته، شکست‌های شناختی که منجر به رفتارهای نایمن کارگران ساختمانی در مراحل مختلف شناختی می‌شود، به طور سیستماتیک تحلیل می‌شوند. از بین مراحل مدل، مرحله «کسب اطلاعات» و «انتخاب پاسخ» به‌عنوان دو مرحله کلیدی شناخته شده‌اند. مزیت اصلی توسعه چنین مدلی از دیدگاه شناختی در این است که می‌تواند به درک بیشتر مکانیسم‌های علت رفتارهای نایمن کمک کند و به توسعه پیشنهاد‌های مدیریت هدفمند برای کاهش مؤثر رفتارهای نایمن کمک کند.
۲	A survey of work-related injuries among building construction workers in southwestern Ethiopia	۲۰۱۸	International Journal of Industrial Ergonomics	Abate Lette et al.	نمونه‌گیری چندمرحله‌ای طبقه‌ای و نمونه‌گیری تصادفی ساده و تجزیه و تحلیل رگرسیون لجستیک دومتغیره	با استفاده از نمونه‌گیری چندمرحله‌ای طبقه‌ای و تصادفی ساده از بین ۱۰ سایت ساختمان‌سازی، ۵ تا انتخاب شده و سپس ۳۵۵ کارگر ساختمانی بر حسب نوع فعالیتشان انتخاب و مورد معاینات پزشکی قرار گرفتند. در انتها طبق معیارهای مشخصی مانند رضایت شغلی، استفاده یا عدم استفاده از وسایل حفاظت فردی، شیوع آسیب‌های ناشی از کار در بین کارکنان به‌دست آمده است.	شیوع کلی صدمات ناشی از کار در یک سال گذشته ۴۱.۴ درصد بوده است. پنج آسیب اول توسط جسم، کمردرد، آسیب ناشی از سقوط، اختلال پوستی و مشکل چشم پدید آمده‌اند. کار بدون تجهیزات حفاظت فردی، فقدان آموزش حرفه‌ای، جویدن ماده محرک خات و اضافه‌کاری به طور قابل توجهی احتمال آسیب‌های ناشی از کار را در بین کارگران ساختمانی افزایش داده‌اند. تهیه تجهیزات ایمنی و ارتقای استفاده از آن، اجتناب از اضافه‌بار کاری و کنترل استفاده از خات در محل کار می‌تواند به حداقل رساندن آسیب‌های ناشی از کار کمک کند.

۳	A Deep Learning Model Development to Predict Safety Accidents for Sustainable Construction A Case Study of Fall Accident in South Korea	۲۰۲۲	Sustainability	Ji-Myong Kim et al.	مدل یادگیری عمیق و شبکه عصبی عمیق	در این مطالعه، ۱۷۶۶ مورد حادثه واقعی توسط سازمان ایمنی شغلی کره طی دوره ۱۰ ساله از ۲۰۱۰ تا ۲۰۱۹ جمع‌آوری شد. هشت عامل مؤثر بر پیش‌بینی حادثه مانند روز پزشکی، میزان پیشرفت و مقیاس ساخت‌وساز انتخاب شدند. پس از آن، قدرت پیش‌بینی بین مدل‌های یادگیری عمیق و مدل‌های رگرسیون چندگانه مرسوم با استفاده از داده‌های حوادث واقعی در سایت‌های ساختمانی مقایسه شد.	در این مطالعه به‌عنوان نتیجه، یک شبکه عصبی عمیق قدرت پیش‌بینی را ۹.۳ درصد در میانگین خطای مطلق و ۱۰.۶ درصد در ریشه میانگین مربعات خطا در مقایسه با یک مدل رگرسیون چندگانه معمولی بهبود بخشید. نتایج این مطالعه دستورالعمل‌هایی را برای معرفی فناوری یادگیری عمیق به مدیریت ایمنی ساخت‌وساز ارائه می‌دهد. به‌کارگیری این مدل در سایت‌های ساختمانی می‌تواند در شناسایی ریسک‌ها و حداقل نمودن آسیب‌ها و حوادث ناشی از آن‌ها مؤثر واقع شود.
۴	Accident prediction in construction using hybrid wavelet machine	۲۰۲۱	International Journal of Creative Research Thoughts	Yeruva Ramana Reddy	استفاده از متدهای ترکیبی یادگیری ماشین و امکان‌سنجی به‌کارگیری آن‌ها	این سیستم برای تخمین احتمال وقوع حادثه با استفاده از مدل‌سازی احتمالی و تکنیک‌های تحلیل ریسک، بر تخمین موقعیت بی‌درنگ ردیابی UWB متکی است. بخش بعدی یک نمای کلی از سیستم کامل ارائه می‌دهد. با این حال بقیه این کار بر روی مدل پیش‌بینی متمرکز خواهد بود.	نتایج نشان می‌دهد که استفاده از هیبریدهای نوآورانه یادگیری موجک-ماشین ممکن است به طور قابل توجهی نرخ تصادفات ساخت‌وساز را کاهش دهد. یادگیری ماشین ممکن است به دو روش مجزا برای بهبود ایمنی ساخت‌وساز استفاده شود. به‌عنوان ابزاری برای یادگیری و همچنین به‌عنوان یک ابزار پشتیبانی تصمیم.
۵	Accidents in Malaysian Construction Industry Statistical Data and Court Cases	۲۰۱۵	International Journal of Occupational Safety and Ergonomics	Heap Yih Chong., Thuan Siang Low	مقایسه آماری ۲ پایگاه‌داده با نمودارهای آماری	این پژوهش از دو رویکرد تحلیلی کمی و کیفی استفاده شده داده‌های کیفی پرونده‌های دادگاه به داده‌های کمی تبدیل شدند تا تجزیه و تحلیل را در قالب نمودارهای میله‌ای راحت‌تر و قابل‌فهم‌تر کنند و در آخر داده‌های دو نتایج مقایسه شدند.	باتوجه به داده‌های آماری رایج‌ترین نوع حادثه مورد اصابت قرارگرفتن با اشیا و یا ضربه خوردن با آن‌ها بوده و باتوجه به پرونده‌های دادگاهی اکثر حوادث ناشی از استفاده نادرست و یا نگهداری ضعیف مصالح و تجهیزات، شایع‌ترین حوادث بودند (۱۴ مورد) و پس از آن، سقوط (۱۳ مورد)، برخورد (۱۰ مورد) و حوادث ناشی از فرو ریختن سازه‌های ساختمانی (۳ مورد) هستند.

۶	Activity classification using accelerometers and machine learning for complex construction worker activities	۲۰۲۰	Journal of Building Engineering	Luis Sanhudo et al.	یادگیری ماشین	<p>داده‌ها از افراد تست که یک مدار ۱۰ فعالیتی را به شکل یک مدار فعالیت با استفاده از سه شتاب‌سنج کم‌هزینه انجام داده، جمع‌آوری و برچسب‌زنی شد؛ سپس به ۳ گروه فعالیت: (۱: بنایی؛ رنگ‌آمیزی و ریخته‌گری. ۲: چکش‌کاری. ۳: آماده‌شدن.) دسته‌بندی و در یک چارچوب بهینه داده‌ها طبقه‌بندی شدند، در ادامه استخراج و انتخاب ویژگی‌ها و تنظیم هاپرپارامترها تعریف شد که با استفاده از روش صحت‌سنجی بهترین طبقه‌بندی انجام شده و در آخر بهترین طبقه‌بندی، رویکرد کاربر مستقل از همه فعالیت‌ها از چارچوب بهینه انتخاب و اعمال شد.</p>	<p>باتوجه به طبقه‌بندی داده‌ها در بهترین مدار، G۲ به بهترین عملکرد متوسط (۹۳.۵۵ درصد) دست‌یافت و پس از آن G۳ (۹۱.۱۲ درصد) و G۱ (۸۶.۳۱ درصد) قرار گرفتند. با این حال، هنگامی که با تمام فعالیت‌ها در یک‌زمان تست شد، عملکرد GrB به ۸۵.۵۴ درصد کاهش یافت، این نتایج نشان می‌دهد که شتاب‌سنج‌ها می‌توانند برای ایجاد یک سیستم قوی برای تشخیص خودکار مجموعه بزرگی از فعالیت‌های پیچیده مورد استفاده قرار گیرند.</p>
۷	An analysis of occupational accidents involving national and international construction workers in Spain using association rule technique	۲۰۲۲	International Journal of Occupational Safety and Ergonomics	María Martínez-Rojas et al.	به‌کارگیری قوانین تلازمی یا همان قوانین انجمنی	<p>جامعه مورد مطالعه اولیه شامل ۵,۴۹۵,۶۰۹ مورد از حوادث شغلی ثبت شده در دوره مذکور می‌باشد. از آنجایی که این مطالعه فقط حوادثی را که در حین فعالیت‌های ساخت‌وساز رخ می‌دهد در نظر می‌گیرد، پس از فیلترینگ، مطالعه بر نمونه‌ای از ۱۵۲۵۸۶۵ تصادف کاهش یافت. ویژگی‌هایی از جمله سن، شغل، مدت‌زمان فعالیت در شرکت، سایز شرکت، مقدار ساعت کاری، شرایط کاری و غیره در این مقاله در نظر گرفته شده‌اند. در نهایت نیز از قوانین تلازمی یا انجمنی برای دستیابی به نتیجه مقاله استفاده شده است.</p>	<p>جو ایمنی مثبت می‌تواند کارگران را برای رعایت مقررات ایمنی و استفاده از روش‌های کار ایمن ترغیب کند. برای اینکه این امر در یک محیط چندفرهنگی اتفاق بیفتد، موانع زبانی و فرهنگی باید از کل زنجیره تولید حذف شود تا اطمینان حاصل شود که اطلاعات بهداشتی و ایمنی به‌درستی منتقل می‌شود. علاوه بر این، کارگران بین‌المللی که معمولاً از نظر عاطفی آسیب‌پذیر هستند، باید احساس کنند که بخشی از برنامه‌های ایمنی و بهداشت شغلی هستند.</p>

۸	Analyses of systems theory for construction accident prevention with specific reference to OSHA accident reports	۲۰۱۳	International Journal of Project Management	Seokho Chi, Sangwon Han.	به کارگیری تئوری سیستم‌ها در مدل دومینوی هاینریش	این مقاله با روش تجزیه و تحلیل تجربی و آماری، ۹۳۵۸ حادثه که در صنعت ساخت و ساز ایالات متحده بین سال‌های ۲۰۰۲ تا ۲۰۱۱ رخ داده است را بررسی کرده و روابط بین حوادث و عوامل خطر کلیدی، ماهیت صدمات ساختمانی و عناصر آسیب را زیر ذره بین قرار داده است. در این مطالعه از آزمون مربع کای و آزمون دقیق فیشر برای تجزیه و تحلیل داده‌ها و درک ارتباط بین عوامل استفاده شده است. همچنین از روش تجزیه و تحلیل فراوانی برای شناسایی عوامل خطر کلیدی در حوادث ساختمانی استفاده شده است.	در این مطالعه همبستگی بین عوامل خطر مختلف از جمله شرایط محیطی، رفتار کارگر و منبع آسیب از نظر آماری شناسایی شد و در نهایت ترکیب‌های عوامل خطر کلیدی را که در حوادث نقش دارند، تعیین شدند. نتایج نشان می‌دهد که میزان مرگ و میر در حوادث ناشی از برق‌گرفتگی و نارسایی تنفسی به علت قرارگیری در معرض گازهای سمی از دیگر حوادث بیشتر است. بیشترین میزان حوادث مربوط به سقوط از ارتفاع، تماس با یک شی و گیرکردن بین دو یا چند شی گزارش شده است. همچنین یافته‌ها نشان می‌دهند که اکثر حوادث ساختمانی شدید هستند و خسارات جبران‌ناپذیری را برای کارگر و سازمان به دنبال دارند.
۹	Analysis of Construction Accidents in Turkey and Responsible Parties	۲۰۱۳	Department of Civil Engineering, Technical University of Istanbul	G.Emre Gurcanli, Ugur Mungen.	طبقه‌بندی بر اساس زمان و نوع حادثه و غیره همچنین طبقه‌بندی علل حوادث و تعداد پیشامد آن	با استفاده از تجزیه و تحلیل و دسته‌بندی ۱۱۱۷ گزارش خبره فعالیت‌هایی که بیشترین میزان حوادث را تشکیل می‌دهند شناسایی و حوادث بر اساس نتیجه حادثه، زمان و علل اصلی حادثه، نوع ساخت و ساز، شغل قربانی، فعالیت در زمان حادثه و شخص مسئول حادثه طبقه‌بندی شدند و اطلاعات حادثه مانند چگونگی، چرا، چه زمانی و چه کسی برای توسعه استراتژی‌های پیشگیری از تصادف مورد بررسی قرار گرفت.	باتوجه به داده‌ها نزدیک ۸۰٪ حادثه‌ها شامل افتادن، برخورد با شیء پرتاب / سقوط، فروپاشی ساختاری و برق‌گرفتگی می‌باشد که نزدیک به دو سوم اقدامات اشتباه و مسامحه کار توسط کارفرمایان انجام می‌شود که کشورها، به جز ترکیه، موفق به کاهش نرخ بروز شده‌اند و از دلایل اصلی این ناآگاهی اول نبود فرهنگ ایمنی و ثانیاً بازرسی‌های ناکافی دولتی نام برده شد.

۱۰	Application of machine learning techniques for predicting the consequences of construction accidents in China	۲۰۲۰	Process Safety and Environmental Protection	Zhu et al.	یادگیری ماشین، الگوریتم‌های رگرسیون لجستیک، درخت تصمیم، ماشین بردار پشتیبان، Naïve Bayes. نزدیک‌ترین همسایه K، جنگل تصادفی، پرسپترون چندلایه و AutoML	این تحقیق از تکنیک یادگیری ماشین برای تجزیه و تحلیل ۱۶ عامل حیاتی و ارزیابی تأثیر ترکیب‌های متنوع عوامل بر عملکرد پیش‌بینی شدن حوادث ساختمانی استفاده می‌کند. این پیش‌بینی با هشت الگوریتم انجام می‌شود: رگرسیون لجستیک، درخت تصمیم، ماشین بردار پشتیبان، Naïve Bayes، نزدیک‌ترین همسایه K، جنگل تصادفی، پرسپترون چندلایه و AutoML. این الگوریتم‌ها سپس از نظر عملکرد با یکدیگر مقایسه می‌شوند.	نتایج نشان می‌دهد که بر اساس ۱۶ عامل حادثه، رگرسیون لجستیک و روش Naive Bayes بهترین نتایج را در مجموعه داده‌های خام به دست می‌آورند. تجزیه و تحلیل ماتریس سردرگمی نشان می‌دهد که طبقه‌بندی ذهنی داده‌های اصلی و حوادث غیرمعمول خاص، منابع پیش‌بینی نادرست هستند. «نوع حادثه» و «گزارش و رسیدگی به حادثه» بحرانی‌ترین عوامل و «مدیریت اضطراری» و «آموزش ایمنی» زیرسیستم‌های مهمی هستند که هر دو به شدت بر شدت حادثه تأثیر می‌گذارند.
۱۱	Assessment of construction workers' perceived risk using physiological data from wearable sensors A machine learning approach	۲۰۲۱	Journal of Building Engineering	Gaang Lee et al.	جمع‌آوری داده با استفاده از ابزار حسگر، استفاده از روش طبقه‌بندی و الگوریتم یادگیری بانظارت	داده‌های فیزیولوژیکی (به عنوان مثال، EDA، PPG و ST) از هشت کارگر ساختمانی (دو نجار، یک کف‌ساز، چهار برق‌کار و یک سرکارگر) جمع‌آوری شد... روش جمع‌آوری داده‌ها توسط هیئت بررسی نهادهای دانشگاه میشیگان تأیید شد. سن افراد بین ۲۰ تا ۵۰ سال، و تجربه شغلی آنها بین ۳ تا ۲۵ سال بود. چهار الگوریتم ماشین بردار پشتیبان گاوسی نزدیک‌ترین همسایه درخت تصمیم و درخت کیسه‌بندی برای ایجاد الگوریتم انتخاب شدند.	اگرچه شناسایی سطوح ریسک درک شده توسط کارگران ممکن است به شناسایی تمام خطرات بالقوه در کارگاه‌های ساخت و ساز کمک نکند، اما برای شناسایی خطرات فوری و اساسی که کارگران در حین کار با آن‌ها مواجه می‌شوند مفید خواهد بود. مدیران ایمنی قادر خواهند بود با ادغام سطوح خطر درک شده کارگران و اطلاعات مکان، فعالیت‌های خطرناک را کنترل کنند. این به مدیران ایمنی کمک می‌کند تا اولویت نظارت ایمنی در محل را در سایت‌های ساختمانی بزرگ تعیین کنند.

۱۲	Association of Root Causes in Fatal Fall-from-Height Construction Accidents in Hong Kong	۲۰۱۶	Journal of Construction Engineering and Management	Louisa Wong et al.	طبقه‌بندی علل اصلی حوادث با HFACS و بررسی ارتباط بین علل با آزمون دقیق فیشر و تحلیل کلاس نهفته	۵۲ حادثه مرگبار ساختمانی سقوط از ارتفاع که بین سال‌های ۱۹۹۹ تا ۲۰۱۱ در هنگ‌کنگ رخ داده است مورد تجزیه و تحلیل قرار گرفت. برای طبقه‌بندی علل اصلی حوادث مرگبار از HFACS استفاده شده و ویژگی‌ها و علل اصلی حوادث با استفاده از تحلیل مرتبه تکرار مورد بررسی قرار گرفت. ارتباط بین علل با استفاده از آزمون دقیق فیشر و تحلیل کلاس نهفته بررسی گشت.	باتوجه به نتایج LCA، عوامل در تمام موارد ممکن است در چهار طبقه قرار گیرند که به نام‌های برنامه‌ریزی نادرست، تخلف، خطرات پنهان ایجاد شده توسط دیگران و کارکنان ناتوان نام‌گذاری می‌شوند. درصد بالایی از حوادث مربوط به افراد متولی است که در پروژه‌های کوچک کار می‌کردند؛ همچنین اکثر حوادث در مدت‌زمان کوتاهی پس از شروع کار رخ داده است و ۴۸٪ از حوادث مربوط به از دست‌دادن تعادل کارگران و عدم استفاده از دستگاه‌های محافظ سقوط بوده است.
۱۳	Causal factors and connections in construction accidents	۲۰۱۸	Safety Science	Stig Winge et al.	مقاله مروری	هدف از این مطالعه، بررسی ۱۷۶ حادثه شدید ساختمانی که توسط اداره نظارت بر کار در نروژ در سال ۲۰۱۵ بررسی شده بود و اضافه کردن به ادبیات نسبتاً پراکنده در مورد علت حوادث در صنعت ساخت‌وساز با شناسایی عواملی علی‌پرتکرار و شناسایی ارتباط بین این عوامل می‌باشد.	هفت عامل پرتکرار که در این مطالعه شناسایی شدند، عبارت‌اند از اقدامات کارگر، مدیریت ریسک، نظارت فوری، توانایی استفاده از مواد یا تجهیزات، خطرات محیطی، مهارت‌های کارگر و مدیریت پروژه. این عوامل به‌صورت دوه‌دو و چندگانه با یکدیگر همبستگی زیادی دارند و بر اهمیت پرداختن به ریسک در سطوح مختلف توسط بازیگران مختلف در پروژه‌های ساختمانی تأکید دارند. این مطالعه الگوهای قوی عوامل پرتکرار و ارتباط بین آن‌ها را نشان می‌دهد که بینش ارزشمندی را در مورد علت حوادث ساختمانی ارائه می‌دهد.
۱۴	Determining Critical Cause Combination of Fatality Accidents on Construction Sites with Machine Learning Techniques	۲۰۲۳	MDPI, Buildings	Shuang & Zhang	توسعه مدل یادگیری ماشین و الگوریتم آنالیز تکرار شونده	روش تحقیق این مطالعه شامل سه مرحله اصلی است: ۱. ایجاد چارچوب ویژگی علت مرگبار. ۲. توسعه فناوری‌های یادگیری ماشین برای پیش‌بینی انواع مختلف تصادفات ۳. استخراج رابطه سلسله‌مراتبی علل کشنده با استفاده از اهمیت ویژگی‌ها و ایجاد یک الگوریتم تجزیه و تحلیل تکرار شونده برای تعیین کمیت ترکیبات علل	نتایج تحقیق نشان داد که حتی اگر علل مرگبار حوادث سلسله‌مراتبی باشند، ترکیباتی در تحلیل آن‌ها وجود دارد. علاوه بر این، این مطالعه توصیه‌هایی برای بهبود مدیریت ایمنی و پیشگیری از حوادث شغلی ارائه می‌کند. یافته‌های این مطالعه، شرکت‌کنندگان در ساخت‌وساز را با ارائه علائم هشداردهنده اولیه عوامل کشنده و ناایمن راهنمایی می‌کند و در نهایت به پیشگیری از تلفات کمک می‌کند.

۱۵	Development of prediction model of construction workers accident occurrence through machine learning	۲۰۱۸	Creative Construction Conference	Jongko Choi et al.	استخراج داده‌های منتشر شده توسط وزارت کار، پیش‌پردازش و ارائه مدل رگرسیونی	ابتدا داده‌های کارگران ساختمانی را از MOEL جمع‌آوری کنید. دوم، داده‌های جمع‌آوری شده (سن، جنس، مدت‌زمان کار، اندازه پروژه، نوع ساخت‌وساز، تاریخ حادثه) از پیش‌پردازش شده است. سوم، یک معادله رگرسیونی استخراج کنید که تصادفات کشنده (مرگ) را با داده‌های تصادفات صنعتی یادگیری ماشینی از یک مدل رگرسیون لجستیک پیش‌بینی می‌کند.	در مورد تصادف منجر به فوت، احتمال وقوع در زمانی که سن بالاتر باشد، مدت‌زمان شغلی طولانی‌تر، اندازه پروژه بزرگ‌تر است، در روز یکشنبه و ماه آپریل بیش‌تر و احتمال وقوع در پروژه معماری کمتر است. اطلاعات بیومتریک به دستگاه تأییدکننده کار کارگر وارد می‌شود، شاخص ریسک محاسبه و گزارش می‌شود. بر اساس این شاخص، مدیر ایمنی می‌تواند گروه‌های خطر روزانه را به طور کارآمد طبقه‌بندی و مدیریت کند که انتظار می‌رود از حوادث مرگبار جلوگیری شود.
۱۶	Factors associated with the severity of construction accidents The case of South Australia	۲۰۱۳	The Australasian Journal of Construction Economics and Building	Jantane Dumrak et al.	مدل مفهومی توسعه‌یافته از طریق مرور ادبیات و بررسی تحلیلی پایگاه‌داده تصادفی	این مطالعه، ۲۴۷۶۴ حادثه ساختمانی بین سال‌های ۲۰۰۲ تا ۲۰۱۱ در استرالیای جنوبی (یکی از شش ایالت استرالیا) تجزیه و تحلیل کرده است. این داده‌ها توسط آژانس بهداشت حرفه‌ای، ایمنی و رفاه دولت ایالتی منتشر شده است. عواملی چون سن، میزان تجربه، جنسیت، اندازه سازمان، مکان و اندازه پروژه، مکانیسم حادثه، محل آسیب بدن و زمان حادثه مورد بررسی قرار گرفته‌اند که نشان می‌دهند چرا برخی از حوادث ساختمانی تنها به شدت جزئی منجر می‌شوند در حالی که برخی دیگر کشنده هستند.	این مطالعه نشان می‌دهد که با افزایش سن کارگر، شدت حوادث نیز افزایش می‌یابد. همچنین زنان در میان حوادث حضور کم‌رنگی دارند و شدت حوادث نیز در آن‌ها کمتر است. این مطالعه بیان می‌کند که اندازه سازمان، مکان و اندازه پروژه با شدت حادثه در ارتباط است و زمان حادثه ارتباط کمی با شدت حادثه دارد. در نهایت مدلی مفهومی برای شدت حوادث ساختمانی ارائه گردیده است.

۱۷	Factors Influencing Unsafe Behaviours and Accidents on Construction Sites A Review	۲۰۱۵	International Journal of Occupational Safety and Ergonomics	Khosravi et al.	آنالیز کیفی و متاآنالیز	در این پژوهش، نتایج و یافته‌های ۵۶ مطالعه پیشین مرتبط مورد بررسی قرار گرفته است. این تحقیقات بر اساس طرح، نوع، روش جمع‌آوری داده‌ها، روش‌های تحلیلی، متغیرها و یافته‌های کلیدی دسته‌بندی شدند. برای استخراج متغیرها، مضامین و عوامل از روش تحلیل محتوای کیفی استفاده شد. علاوه بر این، تمام مطالعات برای تعیین رتبه‌بندی کیفیت و ارزیابی قدرت شواهد ارائه شده مورد بررسی قرار گرفتند. در ادامه، این مطالعه از متاآنالیز برای بررسی روابط بین عوامل مؤثر و رفتارها و حوادث نایمن استفاده کرده است.	هدف از این بررسی روش‌شناختی و نظری، بررسی عوامل تجربی مؤثر بر رفتارهای نایمن و حوادث در سایت‌های ساختمانی است. در نتیجه، این بررسی تأیید کرد که علل رفتارهای نایمن و حوادث در کارگاه‌های ساختمانی چندعاملی به نظر می‌رسد و به‌طور کلی به جامعه، سازمان، مدیریت پروژه، نظارت، پیمانکار، وضعیت سایت، گروه کاری، و خصوصیات فردی مرتبط بودند. نتایج بازبینی از اهمیت عوامل جامعه، سازمان و مدیریت پروژه حمایت می‌کند که ممکن است در کاهش احتمال رفتارهای نایمن و حوادث فراتر از عوامل جزئی‌تر که شامل شرایط سایت و ویژگی‌های فردی می‌شود، نقش داشته باشند.
۱۸	Impact of Job Stressors and Stress on the Safety Behavior and Accidents of Construction Workers	۲۰۱۵	American Society of Civil Engineers	Mei-Yung Leung et al.	تحلیل همبستگی، تجزیه و تحلیل رگرسیون	به‌منظور بررسی روابط پرسش‌نامه‌ای طراحی که در مجموع ۵۰۰ نظرسنجی از طریق تماس شخصی از کارگرانی جمع‌آوری شد سپس تحلیل‌های آماری: عاملی، تحلیل قابلیت اطمینان، تحلیل همبستگی پیرسون و تحلیل رگرسیون خطی چندگانه بر روی داده‌ها اعمال شد که در ادامه برای کنترل کیفیت جمع‌آوری داده‌ها از نمونه‌گیری هدفمند و برای اجتناب از عوامل با عدم قطعیت از چرخش واریماکس استفاده شد.	نتایج نشان داده‌اند که اطمینان شغلی و حمایت همکاران، و همچنین تجهیزات ایمنی، می‌توانند استرس فیزیکی کارگران ساختمانی را به حداقل برسانند، همچنین استرس روانی به طور مثبت توسط حمایت سرپرست همراه می‌شود همچنین دیده شد رفتار ایمنی در میان کارگران ساختمانی با استرس فیزیکی مختل خواهد شد، و هنگامی که حمایت سرپرست مناسب فراهم شود، تقویت خواهد شد. همچنین خطر حوادث زمانی که کارگران رفتار ایمنی را انجام می‌دهند، کاهش می‌یابد

۱۹	Integrating feature engineering genetic algorithm and tree based machine learning methods to predict the post-accident disability status of construction workers	۲۰۲۱	Automation in construction	Kerim Koc et al.	یادگیری ماشین	مجموعه داده بررسی شده شامل ۴۷۹۳۸ حادثه ساختمانی ثبت شده در ترکیه است. سپس پاک سازی داده ها برای داده های تکراری و گمشده صورت گرفته است. در نهایت پیش بینی ها از طریق چهار مدل یادگیری ماشین مبتنی بر درخت شامل Random Forest, XGBoost, AdaBoost, and Extra Trees انجام شده است.	یافته های این مطالعه نشان می دهد که کارگرانی که در معرض دماهای بالا و مواد شیمیایی قرار می گیرند، احتمال ناتوانی دائمی آن ها در حوادث بیشتر است. نتایج این مقاله به پیش بینی وضعیت ناتوانی کارگران ساختمانی پس از حادثه کمک می کند که منجر به محیط کار ایمن تر و برنامه ریزی بهتر در پروژه های ساختمانی می شود. در پایان این مطالعه، راهنمایی برای اجرای این مدل برای شاغلین و همچنین پیشنهاداتی برای سیاست گذاران ارائه گردیده است.
۲۰	Machine learning predictive model based on national data for fatal accidents of construction workers	۲۰۱۹	Automation in Construction	Choi et al.	تجزیه و تحلیل رگرسیون لجستیک، تجزیه و تحلیل درخت تصمیم، تجزیه و تحلیل جنگل تصادفی، و تجزیه و تحلیل AdaBoost	روش به کار گرفته شده در این مطالعه شامل تجزیه و تحلیل داده های اکتشافی اولیه، قبل از توسعه مدل پیش بینی حوادث مرگبار بود. پس از آن، تجزیه و تحلیل رگرسیون لجستیک، تجزیه و تحلیل درخت تصمیم، تجزیه و تحلیل جنگل تصادفی، و تجزیه و تحلیل AdaBoost برای شناسایی کارگران با احتمال بالای مرگ و میر با یادگیری داده های جراحات و مرگ مرتبط با حوادث مرگبار قبلی، به منظور پیش بینی حوادث مرگبار بیشتر تشکیل شد.	این مطالعه داده های ملی حدود ۱۴۰۱۶۹ قربانی حوادث صنعتی در صنعت ساختمان را از سال ۲۰۱۱ تا ۲۰۱۶ (شامل داده های ۱۳۷۳۲۳ مصدوم و ۲۸۴۶ مرگ) تجزیه و تحلیل کرد تا یک مدل پیش بینی برای طبقه بندی گروه های خطر حوادث جدی به منظور افزایش کارایی در مدیریت ایمنی صنعت ساختمان ایجاد کند. عملکرد روش تجزیه و تحلیل جنگل تصادفی، گزارش داده است که ماه، طول مدت شغل، سن، روز و طول خدمت عوامل مهمی برای پیش بینی احتمال تصادف مرگبار هستند.

۲۱	Occupational and non-occupational factors associated with work-related injuries among construction workers in the USA	۲۰۱۵	International Journal of Occupational and Environmental Health	Xiuwen Sue Dong et al.	جمع‌آوری داده از مرکز NLSY۷۹ و مدل رگرسیون لجستیک چندمتغیره	داده‌های سال‌های ۲۰۰۰-۱۹۸۸ مرکز NLSY۷۹ بررسی شد. در معرض شغل بودن و رفتارهای بهداشتی مورد بررسی قرار گرفت و به‌عنوان متغیرهای مستقل در چهار مدل رگرسیون لجستیک چندمتغیره برای شناسایی ارتباط با آسیب‌های شغلی استفاده شد.	پس از کنترل متغیرهای جمعیت‌شناختی، احتمال آسیب‌های شغلی ۱۸ درصد در ساختمانی بیشتر از غیرساختمانی بود. چهار عامل ریسک شغلی یعنی مشاغل یقه آبی، تلاش‌های فیزیکی شغلی، مشاغل متعدد و ساعات کاری طولانی باعث افزایش خطر در ساخت‌وساز می‌شوند. سیگار کشیدن، چاقی/اضافه‌وزن و مصرف کوکائین به طور قابل توجهی خطر آسیب ناشی از کار را زمانی که فاکتورهای جمعیتی و شغلی ثابت نگه داشته شدند، افزایش داد. آسیب‌های محیط کار با بررسی هم‌زمان ویژگی‌های شغلی و غیر شغلی بهتر توضیح داده می‌شوند.
۲۲	Predicting types of occupational accidents at construction sites in Korea using random forest model	۲۰۱۹	Safety Science	Kyungsu Kanga., Hanguk Ryu.	مجموعه داده‌های بررسی حوادث جمع‌آوری شده از KOSHA و مدل منطقی و آزمون دقیق فیشر	در این مقاله به‌منظور اجرای مدل RF از مجموعه داده‌های بررسی حوادث جمع‌آوری شده که در مجموع ۹۷۹۶ حادثه از سال ۲۰۰۸ تا ۲۰۱۴ پوشش می‌دهد استفاده شد. در مجموع ۵۵ متغیر ورودی و انواع حوادث شغلی را به‌عنوان مقادیر خروجی در نظر گرفته شد برای حل نامتعادل بودن مسئله از تکنیک نمونه‌برداری تصادفی استفاده شد که کلاس کم‌تر از ۱۰٪ حذف و مجموعه داده‌ها به ۶۳۷۴ کاهش یافت و در آخر با روش جای‌گشت اهمیت ویژگی‌ها مورد تجزیه و تحلیل قرار گرفت.	باتوجه به نتایج مشخص شد که متغیرهای آب‌وهوایی تأثیر کمی بر انواع حوادث شغلی دارند. دلایل اصلی حوادث مربوط به سقوط از ارتفاع این بود که کارگران تجهیزات ایمنی نمی‌پوشیدند و همچنین انواع حوادث شغلی در سایت‌های ساخت‌وساز تا حد زیادی ناشی از خطای انسانی، نظارت نامناسب و عدم وجود تأسیسات ایمنی نصب شده یا تجهیزات حفاظتی بود و در آخر عوامل کلیدی مدیریتی به‌منظور کمک به کار متخصصان و محققان گفته شد.
۲۳	Relationship between Unsafe Working Conditions and Workers' Behavior and	۲۰۱۳	Journal of Construction Engineering and Management	Seokho Chi et al.	طبقه‌بندی حوادث طبق عامل رفتاری یا شغلی و بررسی همبستگی بین آن‌ها و تأثیرشان بر نوع حادثه و شدت آسیب	در این مطالعه ۹۳۵۸ حادثه را که در صنعت ساخت‌وساز ایالات متحده آمریکا بین سال‌های ۲۰۰۲ تا ۲۰۱۱ رخ داده بود، بررسی کرده است. در ادامه تحلیل آماری همبستگی میان شرایط کاری و عامل رفتاری با استفاده از آزمون مربع کای و آزمون	در بین داده‌های این تحقیق، ۳۳.۴٪ حوادث منجر به مرگ، ۵۵.۷٪ منجر به بستری و سایر حوادث غیربیمارستانی بوده‌اند. در مجموع ۱۷ ارتباط معنادار بین عوامل رفتار و شرایط شناسایی شد و عوامل خطر کلیدی مشخص شد که بر تعیین انواع حادثه و شدت آسیب تأثیر زیادی داشتند. نتایج تحقیق به

	Impact of Working Conditions on Injury Severity in U.S. Construction Industry					دقیق فیشر صورت گرفت و در نهایت تأثیر آن‌ها بر نوع حادثه و شدت آسیب بررسی شد.	مدیران ایمنی کمک می‌کند تا با حذف شرایط کاری ناایمن، اعمال ناامن خاص کارگران را کنترل کنند. آن‌ها همچنین می‌توانند عوامل خطر را اولویت‌بندی کنند و توجه بیشتری به کنترل آن‌ها برای دستیابی به محیط کاری ایمن‌تر داشته باشند.
۲۴	Safety leading indicators for construction sites A machine learning approach	۲۰۱۸	Automation in Construction	Yang Miang Goh et al.	جمع‌آوری دیتای گذشته و استفاده از یادگیری ماشین	<p>. با استفاده از مجموعه داده هفت‌ساله یک شرکت ساختمانی بزرگ و معتبر، چارچوب CRISP-DM شناخته شده در صنعت برای اعمال سیستماتیک فرآیندها و تکنیک‌های ML استفاده شد. به عنوان بخشی از مطالعه، لیستی از ۳۳ متغیر ورودی (همچنین به عنوان ویژگی‌ها یا متغیرهای مستقل شناخته می‌شود) شناسایی شد.</p> <p>پنج الگوریتم ML برای آموزش و اعتبارسنجی از طریق LOO استفاده شد. جنگل تصادفی به عنوان بهترین الگوریتم ML ظاهر شد.</p>	مجموعه ۱۳ متغیر ورودی انتخاب شده و شاخص پیش‌بینی‌شده توسط ML ("بدون تصادف"، "حادثه جزئی" و "حادثه بزرگ" که می‌تواند به ترتیب به عنوان خطر کم، متوسط و بالا نیز برچسب‌گذاری شود) به طور بالقوه می‌تواند برای راهنمایی مدیران و متخصصان ایمنی و بهداشت محیط کار برای شناسایی مکان‌های پرخطر استفاده می‌شود تا بتوان از حوادث پیشگیرانه پیشگیری کرد.
۲۵	Text mining-based construction site accident classification using hybrid supervised machine learning	۲۰۲۰	Automation in construction	Min-Yuan Cheng et al.	یادگیری ماشین، متن‌کاوی و پردازش زبان طبیعی	این مطالعه مدلی ترکیبی از مدل‌های تحت نظارت یادگیری ماشین را ارائه می‌دهد که از پردازش زبان طبیعی برای پیش‌پردازش داده‌های متنی برای طبقه‌بندی حوادث استفاده شده است. این مقاله از ۱۰۰۰ گزارش واقعی حوادث ساخت‌وساز به عنوان مطالعه موردی استفاده کرده است.	در این مطالعه مدل SRGU به عنوان مدل ترکیبی از دو مدل GRU و SOS ارائه شده است. با استفاده از مدل ترکیبی پیشنهادی در این مقاله می‌توان از اشتباهات طبقه‌بندی و سوگیری ناشی از خطای انسانی جلوگیری کرد.

۳- شکاف تحقیقاتی

۳-۱- جدول شکاف تحقیقاتی

جدول (۳-۱) شکاف تحقیقاتی

تکنیک‌های آماری برای شناسایی عوامل و ارزیابی داده‌ها				مدل				یادگیری ماشین	مطالعه موردی	مرور ادبیات	نویسندگان	ردیف
سایر	آزمون‌های آماری ^۹	نمونه‌گیری	تجزیه و تحلیل فراوانی	سایر	Naïve Bayes	DT, SVM, KNN, RF	رگرسیون لجستیک					
			*	مدل شناختی						*	Fang et al.	۱
		*	*				*		*		Abate Lette et al.	۲
MAE, RMSE				Neural Network					*	*	Ji-Myong Kim et al.	۳
				Hybrid Wavelet-ML				*		*	Yeruva Ramana Reddy	۴
نمودارهای آماری			*	مقایسه آماری					*		Heap Yih Chong., Thuan Siang Low	۵
ANOVA	*			MLP, GRB, LSVM, PSVM, ExT, Adb		*	*	*	*		Luis Sanhudo et al.	۶
قوانین انجمنی			*						*		María Martínez-Rojas et al.	۷
	*			مدل دومینوی هاینریش و تئوری سیستم‌ها					*		Seokho Chi, Sangwon Han.	۸
			*						*		G.Emre Gurcanli, Ugur Mungen.	۹
				پرسپترون چندلایه و AutoML	*	*	*		*	*	Zhu et al.	۱۰
						*			*	*	Gaang Lee et al.	۱۱

^۹ آزمون دقیق فیشر یا آزمون مربع کای یا نیکویی برازش

ردیف	نویسندگان	مرور ادبیات	مطالعه موردی	یادگیری ماشین	مدل				تکنیک‌های آماری برای شناسایی عوامل و ارزیابی داده‌ها			
					رگرسیون لجستیک	DT, SVM, KNN, RF	Naïve Bayes	سایر	تجزیه و تحلیل فراوانی	نمونه‌گیری	آزمون‌های آماری	سایر
۱۲	Louisa Wong et al.		*					HFACS	*		*	LCA آزمون برازندگی
۱۳	Stig Winge et al.	*	*						*			نمودارهای آماری
۱۴	Shuang & Zhang	*	*	*								الگوریتم آنالیز تکرارشونده
۱۵	Jongko Choi et al.	*	*				*					روش ۲۰-۸۰
۱۶	Jantane Dumrak et al.	*	*					بررسی تحلیلی	*	آزمون مونت کارلو		آزمون استقلال بین متغیرها
۱۷	Khosravi et al.	*						آنالیز کیفی				تحلیل محتوای کیفی
۱۸	Mei-Yung Leung et al.		*					نمونه‌گیری چندمرحله‌ای	*	نمونه‌گیری هدفمند		تحلیل همبستگی، رگرسیون
۱۹	Kerim Koc et al.		*			RF		XGBoost, AdaBoost, Extra Tree				Precision, Recall, Accuracy, F ¹ score, AUROC, Confusion matrix
۲۰	Choi et al.		*			DT, RF	*	AdaBoost				
۲۱	Xiuwen Sue Dong et al.		*				*			*	*	
۲۲	Kyungsu Kanga., Hanguk Ryu.		*			DT				*	*	
۲۳	Seokho Chi et al.	*	*					تحلیل همبستگی	*		*	
۲۴	Yang Miang Goh et al.	*	*	*			*	CRISP-DM				LOO
۲۵	Min-Yuan Cheng et al.		*			*	*	LSTM, GRU, SGRU	*			Precision, Recall, Accuracy, F ¹ score, Confusion matrix
	مطالعه حاضر											
		*	*	*	*	*	*	XGboost, Boosting, Voting, ...	*	*	*	Precision, Recall, Accuracy, F ¹ score, Confusion matrix, AUROC, Cross Validation

۳-۲- تبیین شکاف تحقیقاتی

همان‌طور که در جدول شکاف مشاهده می‌شود، باتوجه به اهمیت موضوع حوادث ساختمانی و ارتباط آن با جان انسان، اکثر مقالات از مطالعه موردی برای مطالعه خود استفاده کرده‌اند. زیرا مجموعه داده‌های بررسی شده مربوط به مناطق مختلفی هستند و همین امر سبب تفاوت در برخی از نتایج مقالات شده است. در این مطالعه، ما نیز از مجموعه داده حوادث ساختمانی که توسط اداره ایمنی بهداشت حرفه‌ای ایالات متحده بین سال‌های ۲۰۱۵ تا ۲۰۱۷ منتشر شده است، استفاده کردیم. تنها ۴ مورد از مقالات مورد بررسی از الگوریتم‌های یادگیری ماشین در مطالعه خود استفاده کرده‌اند که این موضوع نشان از وجود شکاف در این حوزه است. از روش‌های آماری متعددی برای شناسایی عوامل مؤثر بر حوادث ساختمانی و طبقه‌بندی این حوادث استفاده شده است. اما روش‌های مبتنی بر داده‌کاوی و مدل‌سازی عوامل با استفاده از یادگیری ماشین و پیش‌بینی متغیرهای مختلف یک حادثه ساختمانی، همچنان نیاز به توسعه دارند.

بنابراین، ما در این مطالعه تلاش کردیم ابتدا به بررسی و تجزیه و تحلیل اکتشافی داده‌های حوادث ساختمانی بپردازیم و در ادامه با استفاده از الگوریتم‌های یادگیری ماشین عوامل مؤثر بر حوادث ساختمانی را مدل‌سازی کرده تا بتوانیم حادثه و آسیب‌های ناشی از آن را پیش‌بینی و بررسی کنیم.

۴- مبانی نظری

صنعت ساخت‌وساز به‌عنوان یکی از خطرناک‌ترین صنایع در بسیاری از نقاط جهان شناخته شده است که با مرگ‌ومیر ناشی از کار، جبران خسارت کارگران، میزان جراحت و مرگ‌ومیر اندازه‌گیری شده است. برای جلوگیری از حوادث، باید علل حوادث در محیط کار مانند پروژه‌های ساخت‌وساز ذاتاً خطرناک، عوامل شخصی و پروژه و مکانیسم‌ها یا تجهیزاتی که منجر به حوادث می‌شوند را بدانیم. در اکثر کشورهای جهان، نرخ حوادث مرگبار در صنعت ساخت‌وساز بالاترین میزان است. از این‌رو، نیاز به شرکت‌های ساخت‌وساز برای کاهش مرگ‌ومیر ناشی از آن، موضوع فوریت است. همچنین اشاره شده است که حوادث گذشته می‌توانند به‌عنوان پایه‌ای برای جلوگیری از حوادث آینده عمل کنند [۱۵].

۴-۱- حوادث بر اثر ساخت‌وساز

ماهیت نامشخص پروژه‌های ساخت‌وساز مستلزم ریسک بالا با انواع پیامدها مانند تأخیر پروژه، هزینه‌های بیش از حد، حوادث شغلی منجر به معلولیت دائمی، و مرگ‌ومیرها در سطح جهانی است. حادثی که منجر به مرگ‌ومیر، ناتوانی دائمی، و اثرات سلامت غیرقابل برگشت می‌شوند، در بالاترین حد شدت جراحت قرار دارند که به آن‌ها "حوادث فاجعه‌بار" گفته می‌شود [۲۴]. همچنین هزینه‌های ناشی از حوادث شغلی از الگوی کوه یخ پیروی می‌کنند و شامل هزینه‌های مستقیم و غیرمستقیم است. هزینه‌های مستقیم بخش پیدای هزینه‌ها و قسمت کوچکی از هزینه‌های حوادث ناشی از کار بوده و مخارجی را شامل می‌شود که بابت آن پول پرداخت می‌شود، مانند هزینه پزشکی و درمان و غیره. اما هزینه‌های غیرمستقیم ناشی از حوادث شغلی که بخش پنهان، غیرقابل مشاهده و در بیش‌تر موارد قابل محاسبه نمی‌باشند، در یک تخمین عمومی ۵ تا ۱۱ برابر بیش‌تر از هزینه‌های مستقیم ناشی از حوادث شغلی هستند (وزارت کار، تعاون و رفاه اجتماعی، ۱۳۹۴).

۴-۲- استرس

استرس یک بخش اجتناب‌ناپذیر از تجربه انسان است و به هیچ حربه خاصی محدود نمی‌شود. کارگران ساختمانی معمولاً از انواع مختلفی از استرس، از جمله فیزیکی و روانی رنج می‌برند [۳۰].

۴-۲-۱- عوامل استرس‌زای شغلی

یک عامل استرس‌زا، برای تمایز محرک‌های استرس از پاسخ ایجاد شده به یک رویداد تهدیدکننده یا چالش‌برانگیز، اشاره دارد که می‌تواند منجر به استرس شود. عوامل استرس‌زا در محیط کار به‌عنوان عوامل استرس‌زای شغلی تعریف می‌شوند. به دلیل ویژگی‌های خاص شغل، مانند کار در ارتفاع، کار در مکان‌های شلوغ و کار در کارخانه و تجهیزات پیچیده، کارگران ساختمانی اغلب در موقعیت‌های خطرناک قرار می‌گیرند (به‌عنوان مثال سقوط از ارتفاع، برخورد با مصالح در حال سقوط، گرفتار شدن در ماشین‌آلات). در همین راستا تجهیزات ایمنی، مانند کلاه ایمنی و دستکش برای حفاظت سر و دست‌ها در برابر آسیب، نقش مهمی در حفاظت از سلامت و ایمنی کارگران ساختمانی در محل کار ایفا می‌کنند [۳۰].

۴-۳- عوامل مرتبط با رفتار کارگران در بروز حوادث

برای جلوگیری از صدمات، تلاش‌های پژوهشی قابل توجهی در بررسی عوامل انسانی مؤثر در حوادث ساختمانی صورت گرفته است. عوامل تعیین شده شامل ناآگاهی و فقدان دانش ایمنی، عدم رعایت رویه‌های ایمنی و نگرش نسبت به ایمنی که شامل عدم استفاده از تجهیزات حفاظت فردی، شرایط کار نایمن، عدم مهارت یا آموزش ایمنی و عدم شناسایی شرایط نایمن توسط کارگران در حین کار کردن است. سائورین و همکاران (۲۰۰۵) چنین خطاهای انسانی را به دو نوع اساسی طبقه‌بندی کرده‌اند:

۱. شکست‌های مربوط به عوامل شناختی؛ مانند ظرفیت محدود انسانی
۲. تخلف/انحراف از روش‌های کاری که به‌عنوان شرایط ایمن پذیرفته شده است.

۴-۴- داده‌کاوی

داده‌کاوی دانش را در یک مجموعه داده بزرگ بررسی می‌کند و آن را به یک ساختار قابل درک تبدیل می‌کند. رویکردهای مختلفی بر اساس هدفی که باید به آن دست‌یافت، وجود دارد، به‌عنوان مثال، کشف گروه‌هایی از داده‌ها (به‌عنوان مثال، تجزیه و تحلیل خوشه‌ای)، داده‌های غیرمعمول (مانند تشخیص ناهنجاری) و روابط بین متغیرها (به‌عنوان مثال، قوانین تلازمی) [۲۰].

۴-۵- شاخص‌های ایمنی پیش‌رو و پس‌رو

شاخص‌های پس‌رو شاخص‌هایی هستند که نتایج ایمنی و سلامت محیط کار مانند میزان آسیب و بیماری را اندازه‌گیری می‌کنند و شاخص‌های پیش‌رو شاخص‌هایی هستند که فعالیت‌ها، شرایط و رویدادهای محل کار را که به نتایج ایمنی و بهداشت محیط کار مرتبط هستند یا ممکن است تعیین کنند، اندازه‌گیری می‌کنند [۳۲].

۴-۶- یادگیری ماشین^{۱۰}

یادگیری ماشین شاخه‌ای از هوش مصنوعی و علوم کامپیوتر است که بر استفاده از داده‌ها و الگوریتم‌ها برای تقلید از روشی که انسان‌ها یاد می‌گیرند تمرکز دارد و به تدریج دقت آن را بهبود می‌بخشد. یادگیری ماشین جزء مهمی از حوزه روبه‌رشد علم داده است. از طریق استفاده از روش‌های آماری، الگوریتم‌ها، برای طبقه‌بندی یا پیش‌بینی و کشف بینش‌های کلیدی در پروژه‌های داده‌کاوی آموزش داده می‌شوند. این بینش‌ها متعاقباً تصمیم‌گیری را در برنامه‌ها و کسب‌وکارها هدایت می‌کنند و به طور ایده‌آل بر معیارهای رشد کلیدی تأثیر می‌گذارند. الگوریتم‌های یادگیری ماشین از داده‌های ساختاریافته و برچسب‌گذاری شده برای پیش‌بینی استفاده می‌کنند. به این معنی که ویژگی‌های خاصی از داده‌های ورودی برای مدل تعریف شده و در جداول سازمان‌دهی می‌شوند. این لزوماً به این معنی نیست که از داده‌های بدون ساختار استفاده نمی‌کنند. این فقط به این معنی است که معمولاً برای سازماندهی داده‌ها در قالبی ساختاریافته، داده‌ها فرآیند پیش‌پردازش را طی می‌کنند [۲۵].

^{۱۰} Machine Learning

۴-۶-۱- انواع یادگیری ماشین

- یادگیری نظارت شده^{۱۱}
- یادگیری نظارت نشده^{۱۲}
- یادگیری تقویتی^{۱۳}

۴-۶-۱-۱- الگوریتم‌های یادگیری ماشین بانظارت

این نوع از یادگیری یک نوع یادگیری ماشین است که در آن الگوریتم از داده‌های برچسب‌دار یاد می‌گیرد. داده برچسب‌گذاری شده به معنای مجموعه داده‌ای است که مقدار هدف مربوطه آن از قبل مشخص است. یادگیری تحت نظارت دو نوع دارد:

- **دسته‌بندی:** در این نوع از الگوریتم‌ها کلاس مجموعه داده بر اساس متغیر ورودی مستقل پیش‌بینی می‌شود. کلاس مقادیر مقوله‌ای گسسته است. مثلاً تصویر حیوان گربه یا سگ است.
- **رگرسیون:** در این نوع از الگوریتم‌ها متغیرهای خروجی پیوسته بر اساس متغیر ورودی مستقل پیش‌بینی می‌شود. مانند پیش‌بینی قیمت مسکن بر اساس پارامترهای مختلف مانند سن خانه، فاصله از جاده اصلی، موقعیت مکانی، مساحت و غیره.

۴-۶-۱-۲- الگوریتم‌های یادگیری ماشین بدون نظارت

در یادگیری نظارت بدون نظارت، الگوریتم باید خود به تنهایی به دنبال ساختارهای جالب موجود در داده‌ها باشد. به بیان ریاضی، یادگیری نظارت نشده مربوط به زمانی است که در مجموعه داده فقط متغیرهای ورودی وجود داشته باشند و هیچ متغیر داده خروجی موجود نباشد. به این نوع یادگیری، بدون نظارت گفته می‌شود؛ زیرا برخلاف یادگیری نظارت شده، هیچ پاسخ صحیح داده شده‌ای وجود ندارد و ماشین خود باید به دنبال پاسخ باشد. به بیان دیگر، هنگامی که الگوریتم برای کارکردن از مجموعه داده‌ای بهره گیرد که فاقد داده‌های برچسب‌دار (متغیرهای خروجی) است، از مکانیزم دیگری برای یادگیری و تصمیم‌گیری استفاده می‌کند. به چنین نوع یادگیری، بدون نظارت گفته می‌شود. یادگیری بدون نظارت قابل تقسیم به مسائل خوشه‌بندی و انجمنی است.

- **قوانین انجمنی:** یک مسئله یادگیری هنگامی قوانین انجمنی محسوب می‌شود که هدف کشف کردن قواعدی باشد که بخش بزرگی از داده‌ها را توصیف می‌کنند. مثلاً شخصی که کالای الف را خریداری کند، تمایل به خرید کالای ب نیز دارد.
- **خوشه‌بندی:** یک مسئله هنگامی خوشه‌بندی محسوب می‌شود که قصد کشف گروه‌های ذاتی (داده‌هایی که ذاتاً در یک گروه خاص می‌گنجند) در داده‌ها وجود داشته باشد. مثلاً، بخش‌بندی مشتریان بر اساس رفتار خرید آن‌ها.

^{۱۱} Supervised Learning

^{۱۲} Unsupervised Learning

^{۱۳} Reinforcement Learning

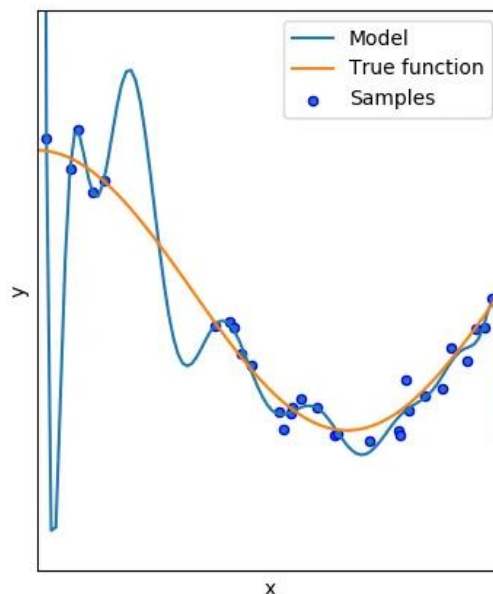
۴-۶-۱-۳- یادگیری تقویتی

یک برنامه رایانه‌ای که با محیط پویا در تعامل است باید به هدف خاصی دست یابد (مانند بازی کردن با یک رقیب یا راندن خودرو). این برنامه بازخوردهایی را با عنوان پاداش‌ها و تنبیه‌ها فراهم و فضای مسئله خود را بر همین اساس هدایت می‌کند. با استفاده از یادگیری تقویتی، ماشین می‌آموزد که تصمیمات مشخصی را در محیطی که دائم در معرض آزمون و خطا است اتخاذ کند.

۴-۶-۲- بیش برآزش^{۱۴}، کم برآزش^{۱۵} و برآزش مناسب

مدل بیش برآزش، مدلی بسیار پیچیده برای داده‌ها است. به این معنی که در تحلیل رگرسیونی، مدلی با بیشترین پارامترها ایجاد می‌شود. در چنین حالتی، مدل با تغییرات جهشی سعی در پوشش داده‌های حاصل از نمونه و حتی مقدارهای نویز می‌کند. در حالی که چنین مدلی باید منعکس‌کننده رفتار جامعه باشد. در این گونه موارد، اگر مدل رگرسیون به دست آمده، برای پیش‌بینی نمونه دیگری به کار رود، مقدارهای پیش‌بینی شده اصلاً مناسب به نظر نخواهند رسید.

در تصویر زیر، نمودار حاصل از بیش برآزش روی داده‌های حاصل از نمونه دیده می‌شود. خط آبی، نشان‌دهنده منحنی برآزش شده روی داده‌ها است و خط نارنجی تابعی است که مدل واقعی جامعه آماری را نشان می‌دهد. نقاط آبی‌رنگ نیز نمونه‌های تصادفی از جامعه آماری را نشان می‌دهند. در مدل بیش برآزش، نقطه‌های حاصل از نمونه بهترین برآزش را دارند و خط آبی تقریباً از همه آن‌ها عبور کرده است.



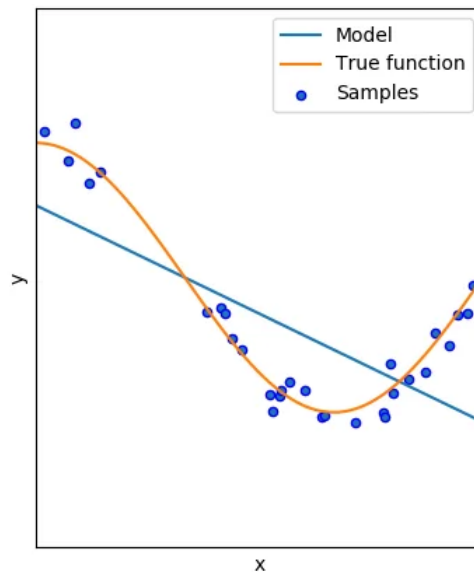
نمودار (۴-۱) منحنی بیش برآزش بر اساس چندجمله‌ای مرتبه ۱۵

همچنین در زمانی که پارامترهای مدل رگرسیونی به صورت کم‌برآزش برآورد می‌شوند، جانب احتیاط حفظ شده و مدل سعی می‌کند با کمترین پارامترها، عمل برآزش را انجام دهد. در نتیجه خطای حاصل از این مدل

^{۱۴} Overfitting

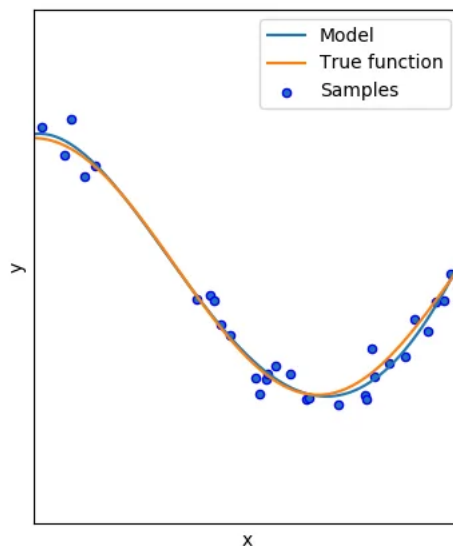
^{۱۵} Underfitting

حتی بر اساس نمونه‌های به‌کاررفته نیز بسیار زیاد است. در تصویر زیر، یک نمونه از مدل رگرسیونی کم‌برازش دیده می‌شود. درجه منحنی به‌کاررفته در این حالت ۱ است که معادله خط محسوب می‌شود.



نمودار (۲-۴) منحنی کم‌برازش بر اساس چندجمله‌ای مرتبه ۱

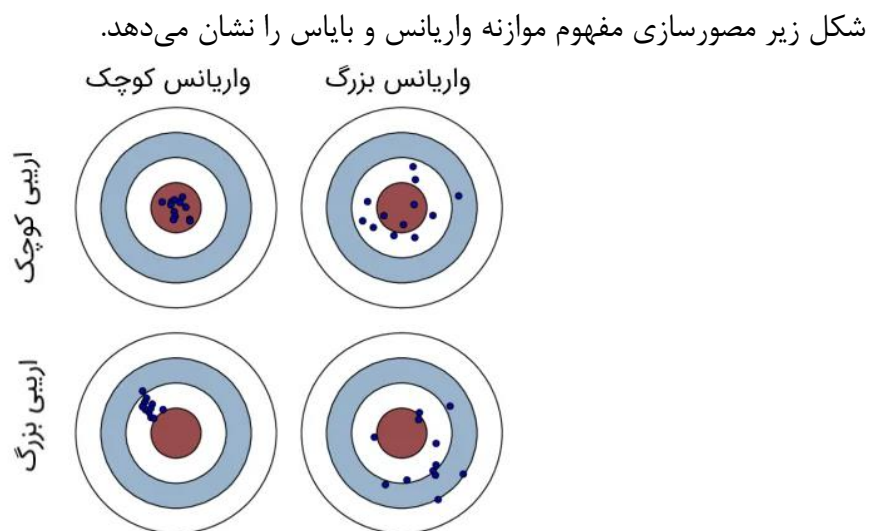
انتظار ما از یک تحلیل رگرسیون مناسب، ایجاد مدلی است که نه تنها بتواند برای داده‌های مربوط به نمونه برازش مناسب را انجام دهد، بلکه برای داده‌هایی جدید نیز امکان برآورد مناسب وجود داشته باشد. همان‌طور که در تصویر زیر دیده می‌شود مدل مناسب دارای خطای کوچکی است و قابلیت پیش‌بینی برای داده‌های جدید را دارد.



نمودار (۳-۴) منحنی برازش مناسب بر اساس چندجمله‌ای مرتبه ۴

۴-۶-۳- موازنه واریانس و بایاس^{۱۶}

- **خطای بایاس:** بایاس در واقع میزان اختلاف نقاط پیش‌بینی شده از متغیر هدف واقعی است. وجود فرضیه‌های مختلف روی مدل و الگوریتم یادگیری منجر به ایجاد خطای اریبی می‌شود. بزرگ بودن اریبی می‌تواند الگوریتم یا مدل آماری را از کشف روابط بین ویژگی‌ها و متغیر پاسخ باز دارد. اغلب بزرگ بودن خطای اریبی، منجر به کم‌برازش می‌شود.
- **خطای واریانس:** واریانس میزان پراکندگی نقاط را نشان می‌دهد. هر چه واریانس بیشتر باشد، پراکندگی داده‌ها بیشتر است. حساسیت زیاد مدل با تغییرات کوچک روی داده‌های آموزشی، نشانگر وجود واریانس زیاد است. این امر نشانگر آن است که اگر مدل آموزش داده‌شده را روی داده‌های آزمایشی به کار گیریم، نتایج حاصل با داده‌های واقعی فاصله زیادی خواهند داشت. متأسفانه افزایش واریانس در این حالت منجر به مدل‌بندی مقادیر نویز^{۱۷} شده و به جای پیش‌بینی صحیح، دچار پیچیدگی و مشکل بیش‌برازش می‌شود.



شکل (۴-۴) مصورسازی موازنه واریانس و بایاس برای برآوردگر

۴-۷- یادگیری ماشین در ساخت‌وساز

یادگیری ماشین می‌تواند فرآیند تصمیم‌گیری مدیریت را در طول برنامه‌ریزی بلندمدت و عملیات روزانه بهبود بخشد. باتوجه‌به ادبیات، یادگیری ماشین در زمینه‌های مدیریت پروژه ساخت‌وساز مانند زمان، هزینه، کیفیت، ایمنی و عملکرد عملیاتی کمک کرده است. برخی از مثال‌های این کاربرد شامل پیش‌بینی عملکرد هزینه پروژه‌های ساختمان تجاری، تمایل به اختلاف در پروژه‌های مشارکت عمومی و خصوصی و وقوع و شدت حادثه می‌شود. یادگیری ماشین اغلب در شرایطی که مشکل دامنه بر اساس ترکیبی از عوامل است، بهتر از روش‌های موجود عمل می‌کند. به‌عنوان مثال، حماد و همکاران گزارش کرده‌اند که تخمین مدت‌زمان ساخت فولاد با استفاده از یادگیری ماشین از روش تخمین موجود بهتر عمل کرده است.

^{۱۶} Bias-Variance Tradeoff

^{۱۷} Noise

۵- مطالعه موردی

در این مطالعه، برای تجزیه و تحلیل حوادث ساختمانی و پیاده سازی الگوریتم های یادگیری ماشین، ما از مجموعه داده اطلاعات حوادث ساختمانی که توسط اداره ایمنی بهداشت حرفه ای ایالات متحده بین سال های ۲۰۱۵ تا ۲۰۱۷ رخ داده است، منتشر شده است. این مجموعه داده در ابتدا شامل ۲۹ ستون (ویژگی^{۱۸}) و ۴۸۴۷ سطر (نمونه^{۱۹}) بوده است که پس از انتخاب ویژگی ها^{۲۰}، پاک سازی داده^{۲۱} و کاهش ابعاد^{۲۲} مجموعه داده، تعداد ستون ها به ۱۹ ستون و سطر ها به ۴۳۵۵ سطر تقلیل یافت.

در ادامه، در جدول (۵-۱) ویژگی های مجموعه داده پاک سازی شده، به صورت مختصر توضیح داده شده است.

جدول (۵-۱) تعریف ویژگی های مجموعه داده پاک سازی شده

نام ویژگی	تعریف
Event Date	تاریخ وقوع حادثه ساختمانی
Construction End Use	استفاده نهایی از سازه (ساختمان تجاری، کارخانه تولیدی، پل، خطوط انتقال انرژی و غیره)
Building Stories	تعداد طبقات سازه
Project Cost	هزینه پروژه (کمتر از ۵۰ هزار دلار، ۵۰ تا ۲۵۰ هزار دلار، ۲۵۰ تا ۵۰۰ هزار دلار، ۵۰۰ هزار تا ۱ میلیون دلار، ۱ تا ۵ میلیون دلار، ۵ تا ۱۰ میلیون دلار، ۱۰ تا ۲۰ میلیون دلار)
Project Type	نوع پروژه (سازه جدید، نوسازی، تعمیرات و مرمت، تخریب و غیره)
Degree of Injury	درجه آسیب دیدگی (کشنده و غیرکشنده)
Nature of Injury	ماهیت آسیب دیدگی (شکستگی، ضربه به سر، سقوط جدی از ارتفاع، خفگی، برق گرفتگی و غیره)
Part of Body	ناحیه آسیب دیده در بدن (سر، انگشتان، تمام بدن، دست، گردن، پا، قلب، دنده ها و غیره)
Event type	نوع حادثه (سقوط، ضربه خوردن، استنشاق، خراش، ساییدگی و غیره)
Environmental Factor	عوامل محیطی (آب و هوا، بلایای طبیعی، گرد و غبار، واکنش های شیمیایی و غیره)
Human Factor	عوامل انسانی (قضاوت نادرست، عدم استفاده از دستگاه های ایمنی، آموزش نادرست و غیره)
Task Assigned	نوع وظیفه تخصیص یافته (منظم و نامنظم)
Event DayOfWeek	روز هفته وقوع حادثه
Event Day	روز حادثه در ماه
Event Month	ماه حادثه در سال
Event Year	سال حادثه
Time	زمان حادثه
Hour	ساعت حادثه
Minutes	دقیقه حادثه

^{۱۸} Feature

^{۱۹} Instance (Record)

^{۲۰} Feature Selection

^{۲۱} Data Cleaning

^{۲۲} Dimension Reduction

۵-۱- روش تحقیق

روش تحقیق این مطالعه، در شکل زیر آمده است.



شکل (۵-۱) روش تحقیق مطالعه

۵-۱-۱- پیش پردازش داده‌ها: پاکسازی داده، کاهش ابعاد، انتخاب ویژگی‌ها

مجموعه داده توسط اداره ایمنی بهداشت حرفه‌ای ایالات متحده منتشر شده است. سه سطر از داده‌ها که تکرار بوده‌اند، حذف گردید. از متن خلاصه حادثه، زمان حادثه استخراج گردید و ستون‌های زمان، ساعت و دقیقه وقوع حادثه به مجموعه داده اضافه گردید. سطرهایی که داده‌های خالی داشتند، حذف گردید. در نهایت ستون‌های کدگذاری شده به دلیل اینکه دانشی را منتقل نمی‌کردند و امکان تحلیل بر روی آن‌ها وجود نداشت نیز حذف شدند.

۵-۱-۲- تجزیه و تحلیل اکتشافی داده‌ها

در این بخش، تجزیه تحلیل فراوانی روی ستون‌های (ویژگی‌های) موجود در مجموعه داده صورت گرفته است و با استخراج دانش، پیشنهاداتی برای بهبود ارائه گردیده است. همچنین همبستگی و ارتباط بین ستون‌ها نیز در این بخش مورد بررسی قرار گرفته است.

۵-۱-۳- استخراج قوانین انجمنی^{۲۳}

استخراج قوانین انجمنی، تکنیکی است که برای کشف روابط پنهان بین متغیرها در مجموعه داده‌های بزرگ استفاده می‌شود. این یک روش محبوب در داده کاوی و یادگیری ماشین است و کاربردهای گسترده‌ای در زمینه‌های مختلف مانند تجزیه و تحلیل سبد بازار، تقسیم‌بندی مشتریان و کشف تقلب دارد. هدف از استخراج قوانین انجمنی کشف قوانینی است که روابط بین متغیرهای مختلف در مجموعه داده را توصیف می‌کند.

برای مثال، مجموعه داده‌ای از حوادث در یک کارگاه ساختمانی را در نظر بگیرید. استخراج قوانین انجمنی می‌تواند برای شناسایی روابط بین عواملی که سبب بروز حادثه می‌شود، استفاده شود. برای مثال، قانون «اگر تجهیزات حمل مواد بالای سرکارگر حرکت کنند، ممکن است سبب آسیب به کارگر از ناحیه سر گردد.» یک قانون ارتباطی است که می‌تواند از این مجموعه داده استخراج شود. ما می‌توانیم از چنین قوانینی برای اطلاع از تصمیم‌گیری‌ها در مورد چیدمان تجهیزات، محل قرارگیری کارگر و مواردی از این دست استفاده کنیم.

الگوریتم‌های مختلفی برای استخراج قوانین انجمنی وجود دارند. در ادامه به پرکاربردترین آن‌ها اشاره می‌شود.

- **الگوریتم Apriori:** الگوریتم Apriori یکی از پرکاربردترین الگوریتم‌ها برای استخراج قوانین انجمنی است. این الگوریتم، ابتدا مجموعه موارد پرتکرار در مجموعه داده را شناسایی می‌کند (مجموعه‌هایی که در تعداد معینی از رکوردها ظاهر می‌شوند). سپس از این مجموعه موارد پرتکرار برای تولید قوانین انجمنی استفاده می‌کند. الگوریتم Apriori از یک رویکرد پایین به بالا استفاده می‌کند که از موارد جداگانه شروع می‌شود و به تدریج به مجموعه‌های موارد پیچیده‌تر می‌رسد.

- **الگوریتم FP-Growth^{۲۴}:** الگوریتم رشد الگوی پرتکرار، یکی دیگر از الگوریتم‌های محبوب برای استخراج قوانین انجمنی است که با ساختن یک ساختار درخت مانند به نام FP-tree کار می‌کند که مجموعه موارد پرتکرار در مجموعه داده را رمزگذاری می‌کند. سپس از FP-tree برای ایجاد قوانین انجمنی به روشی مشابه الگوریتم Apriori استفاده می‌شود. الگوریتم رشد الگوی پرتکرار به‌طور کلی سریع‌تر از الگوریتم Apriori است.

- **الگوریتم ECLAT^{۲۵}:** الگوریتم خوشه‌بندی کلاس هم ارز و پیمایش شبکه از پایین به بالا، نوعی از الگوریتم Apriori است که از رویکرد بالا به پایین به جای رویکرد از پایین به بالا استفاده می‌کند. با تقسیم موارد به کلاس‌های معادل بر اساس پشتیبانی آن‌ها (تعداد رکوردهایی که در آن‌ها ظاهر می‌شوند) کار می‌کند. سپس قوانین انجمنی با ترکیب این کلاس‌های هم ارزی در یک ساختار شبکه مانند ایجاد می‌شود. این یک نسخه کارآمدتر و مقیاس پذیرتر از الگوریتم Apriori است.

در این مطالعه ما از الگوریتم Apriori برای استخراج قوانین انجمنی استفاده کرده‌ایم. در ادامه نحوه عملکرد این الگوریتم را بیان می‌کنیم.

^{۲۳} Association Rule Mining

^{۲۴} Frequent Pattern Growth

^{۲۵} Equivalence Class Clustering and bottom-up Lattice Traversal

۵-۱-۳-۱- Apriori الگوریتم

این الگوریتم، با تنظیم حداقل آستانه پشتیبانی^{۲۶} شروع می‌شود. این عدد حداقل تعداد دفعاتی است که یک مورد باید در پایگاه داده رخ دهد تا بتوان آن را به عنوان مجموعه موارد پرتکرار در نظر گرفت. سپس الگوریتم هر مجموعه مواردی را که حداقل آستانه پشتیبانی را برآورده نمی‌کنند، فیلتر می‌کند.

سپس الگوریتم لیستی از تمام ترکیبات ممکن از مجموعه موارد پرتکرار ایجاد می‌کند و تعداد دفعاتی که هر ترکیب در پایگاه داده ظاهر می‌شود را می‌شمارد. سپس الگوریتم فهرستی از قوانین مرتبط را بر اساس ترکیبات پرتکرار مجموعه موارد تولید می‌کند.

قدرت^{۲۷} قانون انجمنی با استفاده از معیار اطمینان^{۲۸} اندازه‌گیری می‌شود که احتمال وجود آیت ب با توجه به وجود آیت الف است. سپس الگوریتم قوانین انجمنی را که حداقل آستانه اطمینان را برآورده نمی‌کند، فیلتر می‌کند. از این قوانین به عنوان قوانین انجمن قوی یاد می‌شود. در نهایت، الگوریتم لیستی از قوانین مرتبط قوی را به عنوان خروجی برمی‌گرداند. در ادامه به معیارهای ارزیابی و تحلیل قوانین انجمنی می‌پردازیم [۳۴].

۵-۱-۳-۲- معیارهای ارزیابی قوانین انجمن

در استخراج قوانین انجمنی، معمولاً از چندین معیار برای ارزیابی کیفیت و اهمیت قوانین کشف شده استفاده می‌شود. این معیارها را می‌توان برای ارزیابی کیفیت و اهمیت قوانین مرتبط و انتخاب مناسب ترین قوانین برای یک کاربرد خاص مورد استفاده قرار داد.

تفسیر نتایج معیارهای استخراج قواعد انجمنی مستلزم درک معنا و مفاهیم هر معیار و همچنین نحوه استفاده از آنها برای ارزیابی کیفیت و اهمیت قوانین مرتبط کشف شده است. در اینجا چند دستورالعمل برای تفسیر نتایج معیارهای استخراج قانون انجمنی اصلی آورده شده است.

- **Support:** پشتیبانی معیاری است که نشان می‌دهد یک مورد یا مجموعه موارد به دفعات در مجموعه داده ظاهر می‌شود و به شکل زیر محاسبه می‌گردد. پشتیبانی زیاد نشان می‌دهد که یک مورد یا مجموعه موارد در مجموعه داده مشترک است، در حالی که پشتیبانی کم نشان دهنده نادر بودن آن است.

$$Support(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Total\ number\ of\ transactions}$$

- **Confidence:** معیاری برای سنجش قدرت ارتباط بین دو مورد است. به عنوان تعداد رکوردهای حاوی هر دو مورد تقسیم بر تعداد رکوردهای حاوی اولین مورد محاسبه می‌شود. اطمینان بالا نشان می‌دهد که وجود مورد اول یک پیش‌بینی کننده قوی برای حضور مورد دوم است.

$$Confidence(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Transactions\ containing\ X}$$

- **Lift:** این معیار اندازه‌گیری قدرت ارتباط بین دو مورد با در نظر گرفتن فراوانی هر دو مورد در مجموعه داده است. به عنوان معیار اطمینان تقسیم بر معیار پشتیبانی مورد دوم محاسبه می‌شود. این معیار، برای مقایسه قدرت

^{۲۶} Minimum Support Threshold

^{۲۷} Strength

^{۲۸} Confidence

ارتباط بین دو مورد با قدرت مورد انتظار انجمن در صورتی که موارد مستقل باشند، استفاده می‌شود. مقدار بیشتر از ۱ نشان می‌دهد که ارتباط بین دو مورد قوی‌تر از حد انتظار بر اساس فراوانی اقلام است. این نشان می‌دهد که این ارتباط ممکن است معنی‌دار باشد و ارزش بررسی بیشتر را داشته باشد. مقدار کمتر از ۱ نشان می‌دهد که ارتباط ضعیف‌تر از حد انتظار است و احتمالاً کمتر قابل توجه باشد.

$$Lift(\{X\} \rightarrow \{Y\}) = \frac{(Transactions\ containing\ both\ X\ and\ Y) / (Transactions\ containing\ X)}{Fraction\ of\ transactions\ containing\ Y}$$

۵-۱-۴- پیاده‌سازی الگوریتم‌های یادگیری ماشین

۵-۱-۴-۱- آماده‌سازی مجموعه داده برای پیاده‌سازی الگوریتم‌های یادگیری ماشین

از آنجایی که ستون‌هایی که واریانس کمی دارند در مدل‌های پیش‌بینی تأثیر چندانی ندارند، برای کاهش زمان پردازش مدل، از مجموعه داده حذف شدند. این ستون‌ها شامل استفاده نهایی از سازه، تعداد طبقات سازه، هزینه پروژه و نوع پروژه می‌باشند؛ زیرا تعداد داده‌های وارد نشده آن‌ها بسیار زیاد می‌باشد. همچنین ستون‌های تاریخ وقوع حادثه و زمان وقوع حادثه به علت اینکه نوع داده‌های آن‌ها از نوع عددی^{۲۹} نمی‌باشند نیز از مجموعه داده حذف گردیدند. زیرا الگوریتم‌های یادگیری ماشین تنها روی داده‌های عددی قابل اجرا هستند.

برای تبدیل سایر ستون‌های غیر عددی طبقه‌بندی شده^{۳۰}، به ستون‌های عددی از یک روش کدگذاری طبقه‌ای^{۳۱} استفاده شده است. کدگذاری طبقه‌ای فرآیندی است که در آن داده‌های طبقه‌بندی به داده‌های عددی تبدیل می‌شوند. تکنیک‌های کدگذاری طبقه‌بندی زیادی وجود دارد، در این مطالعه، ما از روش کدگذاری هش^{۳۲} استفاده کرده‌ایم. رمزگذاری هش مبتنی بر تابع هش انجام می‌شود و داده‌های طبقه‌بندی را به عددی تبدیل می‌کند. مزیت اصلی استفاده از رمزگذاری هش این است که می‌توان تعداد ستون‌های عدد مطلوب را کنترل کرد. این کار باعث می‌شود تا هنگام پردازش مجموعه داده توسط مدل‌های یادگیری ماشین، میزان حافظه^{۳۳} کمتری اشغال گردد و مدل در زمان کمتری اجرا شود.

در این مطالعه، ما فرآیند رمزگذاری هش را روی ستون‌های ماهیت آسیب‌دیدگی، ناحیه آسیب‌دیده در بدن، نوع حادثه، عوامل محیطی، عوامل انسانی، نوع وظیفه اختصاص یافته و روز هفته وقوع حادثه پیاده‌سازی کردیم و تمام ستون‌های غیر عددی را به ستون‌های عددی تبدیل کردیم.

۵-۱-۴-۲- تقسیم مجموعه داده به سه قسمت آموزشی^{۳۴}، اعتبارسنجی^{۳۵} و تست^{۳۶}

پس از آماده‌سازی مجموعه داده برای پیاده‌سازی الگوریتم‌های یادگیری ماشین، مجموعه داده را به سه قسمت آموزشی، اعتبارسنجی و تست تبدیل کرده‌ایم. ۷۰ درصد مجموعه داده، به عنوان مجموعه داده آموزشی، ۱۰ درصد

^{۲۹} Numerical

^{۳۰} Categorical

^{۳۱} Categorical Encoding

^{۳۲} Hash Encoding

^{۳۳} Memory

^{۳۴} Train Dataset

^{۳۵} Validation Dataset

^{۳۶} Test Dataset

برای اعتبارسنجی اولیه هر مدل و ۲۰ درصد از مجموعه داده به عنوان مجموعه داده تست در نظر گرفته شده‌اند که این مجموعه داده برای اندازه‌گیری صحت و دقت مدل‌ها کنار گذاشته می‌شود و مدل در نهایت روی آن تست می‌گردد.

۵-۱-۴-۳- پیاده‌سازی الگوریتم‌های یادگیری ماشین

برای پیش‌بینی کشنده یا غیرکشنده بودن حوادث از الگوریتم‌های دسته‌بندی بانظارت زیر استفاده شده است. عدد ۱ به معنی غیرکشنده بودن حادثه و عدد ۰ به معنی کشنده بودن حادثه می‌باشند.

۵-۱-۴-۱- رگرسیون لجستیک^{۳۷}

در این مدل، احتمالاتی که نتایج احتمالی یک رویداد را توصیف می‌کنند، با استفاده از یک تابع لجستیک مدل‌سازی می‌شوند. رگرسیون لجستیک، یک مدل آماری رگرسیون برای متغیرهای وابسته دوسویی مانند بیماری یا سلامت و مرگ یا زندگی است. این مدل را می‌توان به عنوان مدل خطی تعمیم‌یافته‌ای که از تابع لجستیک^{۳۸} به عنوان تابع پیوند استفاده می‌کند و خطایش از توزیع چندجمله‌ای پیروی می‌کند، به حساب آورد. منظور از دو سویی بودن، رخ داد یک واقعه تصادفی در دو موقعیت ممکنه است. در این مطالعه متغیر هدف یعنی کشنده یا غیرکشنده بودن حادثه نیز یک متغیر وابسته دوسویی است. رابطه زیر، رابطه تابع لجستیک را نشان می‌دهد.

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i},$$

همچنین p در رابطه فوق برابر است با:

$$p = \Pr(y_i = 1).$$

$$p = \Pr(y_i = 1 | \vec{x}_i; \vec{\beta}) = \frac{e^{\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}}}{1 + e^{\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i})}}.$$

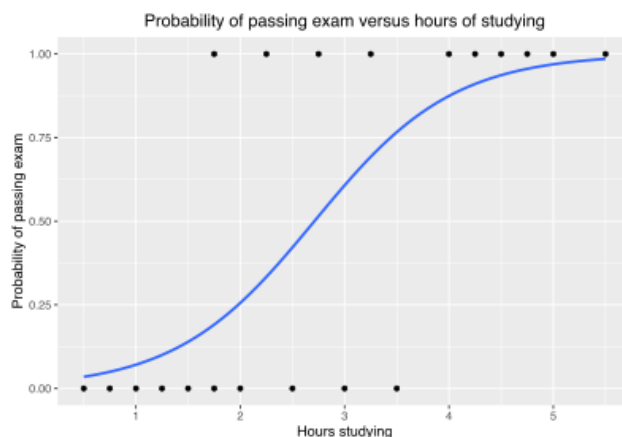
در نهایت رگرسیون لجستیک را می‌توان به شکل زیر بازنویسی کرد.

$$\Pr(y_i = 1 | \vec{x}_i; \vec{\beta}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i})}} = \sigma(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i})$$

نمودار زیر، یک منحنی رگرسیون لجستیک را نشان می‌دهد که احتمال قبولی در امتحان را در مقابل ساعات مطالعه مورد بررسی قرار داده است.

^{۳۷} Logistic Regression

^{۳۸} logit

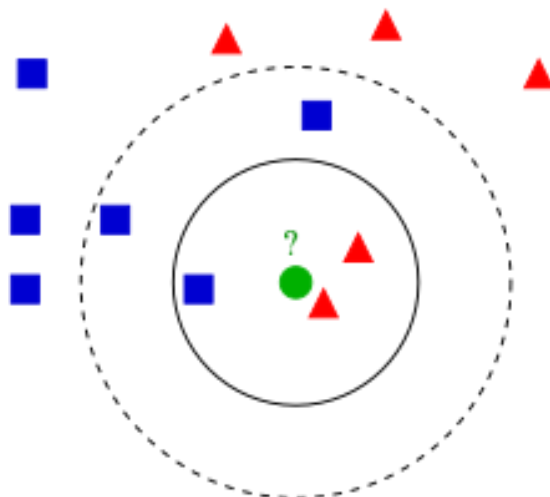


نمودار (۱-۵) مثالی از منحنی رگرسیون لجستیک

۵-۱-۲-۳-۴-۵ K نزدیک‌ترین همسایه^{۳۹}

روش K نزدیک‌ترین همسایه که روش آماری مبتنی بر فاصله است که برای طبقه‌بندی آماری و رگرسیون استفاده می‌شود. در هر دو حالت، K شامل نزدیک‌ترین مثال آموزشی در فضای داده‌ای می‌باشد و خروجی آن بسته به نوع مورد استفاده در طبقه‌بندی و رگرسیون متغیر است.

در حالت طبقه‌بندی با توجه به مقدار مشخص شده برای K، به محاسبه فاصله نقطه‌ای که می‌خواهیم برچسب آن را مشخص کنیم با نزدیک‌ترین نقاط می‌پردازد و با توجه به تعداد رأی حداکثری این نقاط همسایه، در رابطه با برچسب نقطه مورد نظر تصمیم‌گیری صورت می‌گیرد. برای محاسبه این فاصله می‌توان از روش‌های مختلفی استفاده کرد که یکی از مطرح‌ترین این روش‌ها، فاصله اقلیدسی است. در حالت رگرسیون نیز میانگین مقادیر به‌دست‌آمده از K خروجی آن می‌باشد. از آنجاکه محاسبات این الگوریتم بر اساس فاصله است نرمال‌سازی داده‌ها می‌تواند به بهبود عملکرد آن کمک کند. شکل زیر، مثالی از پیاده‌سازی این الگوریتم را نشان می‌دهد.



شکل (۲-۵) مثالی از پیاده‌سازی الگوریتم K نزدیک همسایه برای طبقه‌بندی داده‌ها

^{۳۹} K-Nearest Neighbors

در شکل فوق، نقطه سبز رنگ نمونه تست می‌باشد که باید به مربع‌های آبی یا قرمز طبقه‌بندی شود. اگر $K=3$ در نظر گرفته شود، ای داده به مثلث‌های قرمز نسبت داده می‌شود. اما اگر $K=5$ باشد، به مربع‌های آبی اختصاص داده می‌شود.

۵-۱-۴-۳- درخت تصمیم‌گیری^{۴۰}

درخت تصمیم یک مدل سلسله‌مراتبی پشتیبانی تصمیم است که از یک مدل درخت‌مانند از تصمیمات و پیامدهای احتمالی آن‌ها، از جمله نتایج رویدادهای شانس، هزینه‌های منابع و مطلوبیت استفاده می‌کند. درخت‌های تصمیم معمولاً در تحقیقات عملیاتی، به‌ویژه در تجزیه و تحلیل تصمیم‌گیری برای کمک به شناسایی استراتژی که به احتمال زیاد به یک هدف می‌رسد، استفاده می‌شوند.

درخت تصمیم دارای اجزای زیر است:

- **گره اصلی^{۴۱}:** ویژگی کلیدی در مجموعه داده
- **گره داخلی^{۴۲}:** گره‌هایی که یک یال ورودی و دو یا چند یال خروجی دارند.
- **گره برگ^{۴۳}:** گره پایانی بدون یال خروجی

درخت تصمیم از یک گره اصلی شروع می‌شود و با بررسی شرایط مختلف و اختصاص آن به سایر گره‌ها ادامه می‌یابد. درخت تصمیم زمانی کامل می‌شود که تمام شرایط به یک گره برگ منتهی شوند. گره برگ حاوی برچسب طبقه‌بندی می‌باشد.

برای تقسیم بهینه ویژگی‌ها دو روش وجود دارد:

- **روش شاخص جینی^{۴۴}:** ناخالصی جینی تعداد برچسب‌گذاری اشتباه هر عنصر مجموعه داده را هنگامی که به طور تصادفی برچسب‌گذاری می‌شود، اندازه‌گیری می‌کند. در شکل زیر فرمول شاخص جینی مشاهده می‌شود که در آن p_j احتمال کلاس j است. حداقل مقدار شاخص جینی ۰ است. این زمانی اتفاق می‌افتد که گره خالص باشد، به این معنی که تمام عناصر موجود در گره از یک کلاس منحصر به فرد هستند؛ بنابراین، این گره دوباره تقسیم نخواهد شد. تقسیم بهینه توسط ویژگی‌هایی با شاخص جینی کمتر انتخاب می‌شود. علاوه بر این، زمانی که احتمال دو کلاس یکسان باشد، حداکثر مقدار (۰.۵) را دریافت می‌کند.

$$GiniIndex = 1 - \sum_j p_j^2$$

- **روش آنتروپی^{۴۵}:** آنتروپی معیاری از اطلاعات است که نشان‌دهنده بی‌نظمی ویژگی‌ها با متغیر هدف است. مشابه شاخص جینی، تقسیم بهینه توسط ویژگی با آنتروپی کمتر انتخاب می‌شود. مقدار آنتروپی زمانی حداکثر مقدار خود (۱) را به دست می‌آورد که احتمال دو کلاس یکسان باشد و هنگامی که یک گره خالص باشد،

^{۴۰} Decision Tree

^{۴۱} Root Node

^{۴۲} Internal Node

^{۴۳} Leaf Node

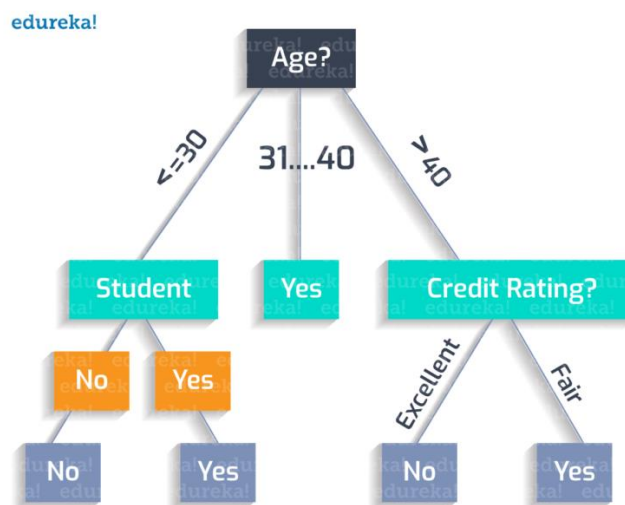
^{۴۴} Gini Index

^{۴۵} Entropy

مقدار آنتروپی حداقل مقدار خود یعنی ۰ است. فرمول محاسبه آنتروپی به شکل زیر است که در آن p_j احتمال کلاس j است.

$$Entropy = - \sum_j p_j \cdot \log_2 \cdot p_j$$

شکل زیر نمونه‌ای از یک درخت تصمیم با ویژگی‌ها با روش شاخص جینی را نشان می‌دهد که هدف آن پیش‌بینی خرید لپ‌تاپ توسط کاربر می‌باشد. همان‌طور که مشاهده می‌شود، ویژگی سن^{۴۶} به‌عنوان گره اصلی انتخاب شده است و سایر ویژگی‌ها در گره‌های داخلی قرار دارند و با بررسی شرایط مختلف، گره‌های برگ مشخص شده‌اند.



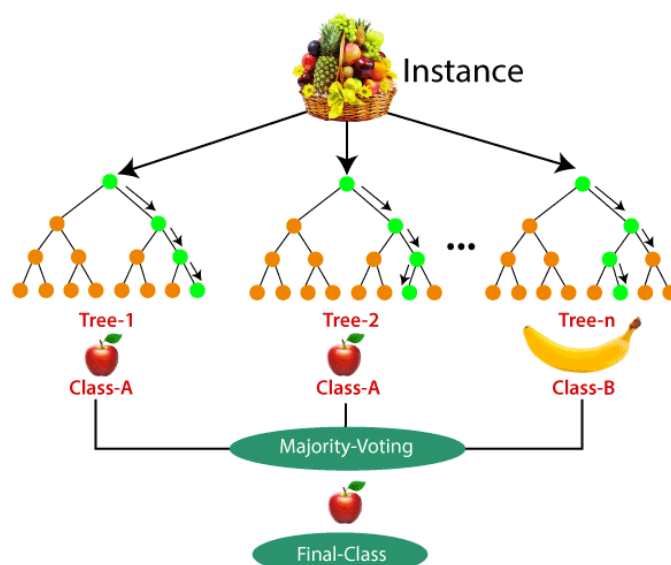
شکل (۵-۳) مثالی از پیاده‌سازی الگوریتم درخت تصمیم برای طبقه‌بندی داده‌ها

۵-۱-۴-۳-۴-۵- جنگل تصادفی^{۴۷}

جنگل تصادفی یک روش یادگیری ترکیبی برای دسته‌بندی و رگرسیون می‌باشد که بر اساس ساختاری متشکل از شمار بسیاری درخت تصمیم، در زمان آموزش عمل می‌کند. عملکرد الگوریتم جنگل تصادفی معمولاً بهتر از الگوریتم درخت تصمیم است، اما این بهبود عملکرد تا حدی به نوع داده هم‌بستگی دارد. برای کاربرد دسته‌بندی، خروجی جنگل تصادفی کلاسی است که توسط اکثر درختان انتخاب شده است. شکل زیر مثالی ساده از پیاده‌سازی الگوریتم جنگل تصادفی را بر روی نمونه‌ای از میوه‌ها نشان می‌دهد.

^{۴۶} Age

^{۴۷} Random Forest



شکل (۴-۵) مثالی از پیاده‌سازی الگوریتم جنگل تصادفی برای دسته‌بندی داده‌ها

۵-۱-۴-۳-۵- ماشین بردار پشتیبان^{۴۸}

ماشین بردار پشتیبان یکی از روش‌های یادگیری بانظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌شود. مبنای کاری دسته‌بندی این الگوریتم، دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها، سعی می‌کند ابرصفحه‌ای را انتخاب کند که حاشیه اطمینان بیشتری داشته باشد. حل معادله پیدا کردن خط بهینه برای داده‌ها به وسیله روش‌های برنامه‌ریزی غیرخطی که روش‌های شناخته شده‌ای در حل مسائل محدودیت‌دار هستند، صورت می‌گیرد. قبل از تقسیم خطی برای اینکه ماشین بتواند داده‌های با پیچیدگی بالا را دسته‌بندی کند، داده‌ها به وسیله تابع فی^{۴۹} به فضای با ابعاد خیلی بالاتر برده می‌شود. برای اینکه بتوان مسئله ابعاد خیلی بالا را با استفاده از این روش‌ها حل کنیم، از قضیه دوگانگی لاگرانژ^{۵۰} برای تبدیل مسئله مینیمم‌سازی موردنظر به فرم دوگانگی آن که در آن به جای تابع پیچیده فی که ما را به فضایی با ابعاد بالا می‌برد، تابع ساده‌تری به نام تابع هسته^{۵۱} که ضرب برداری تابع فی است ظاهر می‌شود، استفاده می‌کنیم. از توابع هسته مختلفی از جمله هسته‌های نمایی، چندجمله‌ای و سیگموئید^{۵۲} می‌توان بدین منظور استفاده نمود. شکل زیر، مثالی از عملکرد الگوریتم ماشین بردار پشتیبان را نشان می‌دهد. به ابرصفحه‌های حاشیه، بردارهای پشتیبان گفته می‌شود.

^{۴۸} Support Vector Machine

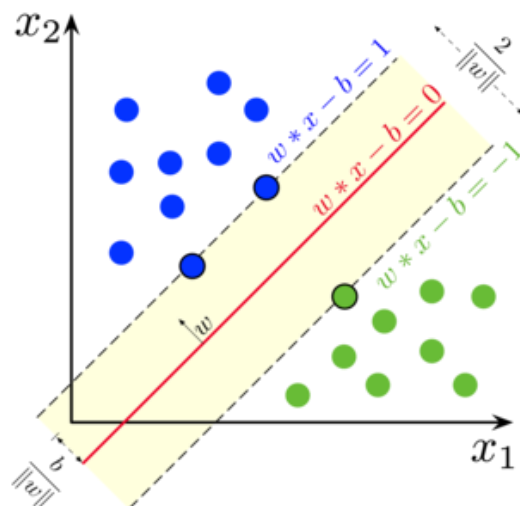
^{۴۹} Phi Function

^{۵۰} Lagrange Duality Theorems

^{۵۱} Kernel

^{۵۲} Sigmoid Function

تابع سیگموئید تابعی حقیقی، یکنوا، کران‌دار و مشتق‌پذیر است که به ازای کلیه مقادیر حقیقی قابل تعریف بوده دارای مشتق نامنفی است که دارای یک نقطه‌ی عطف است. این تابع به لحاظ گرافیکی شکلی شبیه حرف S انگلیسی و سیگما در یونانی دارد. دامنه توابع سیگموئید شامل تمامی اعداد حقیقی بوده و مقدار بازگشتی این تابع نیز به طور یکنواخت از ۰ تا ۱ یا باتوجه به نوع تابع از ۱ تا -۱ تغییر می‌کند.



شکل (۵-۵) ابرصفحه‌ای با حداکثر حاشیه برای یک ماشین بردار پشتیبان

۵-۱-۴-۳-۶- بیز ساده گاوسی^{۵۳}

در ابتدا توضیح کوتاهی درباره الگوریتم بیز ساده^{۵۴} ارائه می‌گردد. بیز ساده، یک الگوریتم یادگیری ماشین احتمالی است که می‌تواند در طبقه‌بندی چندگانه استفاده شود. کاربردهای معمولی بیز ساده طبقه‌بندی اسناد، فیلتر کردن هرزنامه‌ها، پیش‌بینی و غیره است. این الگوریتم بر اساس اکتشافات توماس بیز به این نام گذاشته شده است. مزیت اصلی این الگوریتم این است که یک الگوریتم ساده و درعین حال قدرتمند است.

قانون بیز فرمول احتمال متغیر Y را در شرایط X ارائه می‌دهد.^{۵۵} اما در دنیای واقعی، ممکن است چندین متغیر X وجود داشته باشد. وقتی ویژگی‌های مستقلی دارید، قانون بیز را می‌توان به قانون بیز ساده تعمیم داد که در آن X ها مستقل از یکدیگر هستند.

هنگامی که متغیر X از توزیع نرمال و یا گاوسی پیروی کند، از تابع چگالی توزیع نرمال در فرمول بیز ساده استفاده می‌گردد که به آن بیز ساده گاوسی گفته می‌شود. فرمول بیز ساده گاوسی به شکل زیر است.

$$P(X|Y=c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(x-\mu_c)^2}{2\sigma_c^2}}$$

در فرمول فوق، μ میانگین متغیر پیوسته X برای کلاس c و σ واریانس متغیر پیوسته X برای کلاس c می‌باشند. تابع چگالی احتمال گاوسی را می‌توان با جایگزین کردن پارامترها با مقدار ورودی جدید متغیر برای پیش‌بینی استفاده کرد. در نتیجه تابع گاوسی، تخمینی برای احتمال مقدار ورودی جدید ارائه می‌دهد.

^{۵۳} Gaussian Naive Bayes

^{۵۴} Naive Bayes

^{۵۵} Bayes Rule = $P(Y|X) = P(X|Y) * P(Y) / P(X)$

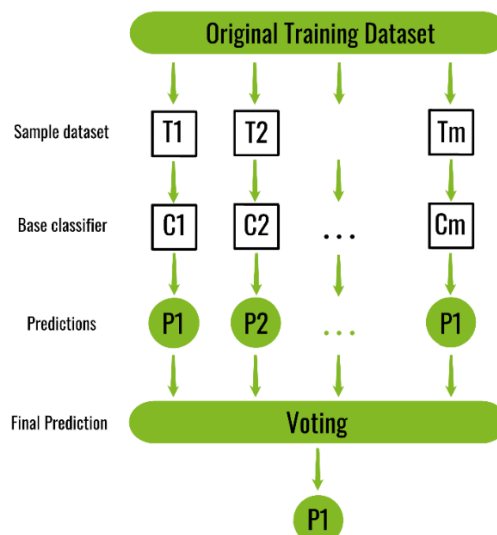
۵-۱-۴-۷- بیز ساده برنولی^{۵۶}

بیز ساده برنولی، یکی از الگوریتم‌های بیز ساده است که اساس توزیع برنولی می‌باشد و فقط مقادیر دودویی، یعنی ۰ یا ۱ را می‌پذیرد. اگر ویژگی‌های مجموعه داده دودویی^{۵۷} باشند، می‌توان از این الگوریتم استفاده کرد. فرمول بیز ساده برنولی به شکل زیر است.

$$P(x_i | y) = P(x_i = 1 | y)x_i + (1 - P(x_i = 1 | y))(1 - x_i)$$

۵-۱-۴-۸- دسته‌بندی کیسه‌ای^{۵۸}

الگوریتم دسته‌بندی کیسه‌ای یک فرابراوردگر^{۵۹} ترکیبی^{۶۰} است که هر کدام از طبقه‌بندی‌کننده‌های پایه را بر روی زیرمجموعه‌های تصادفی مجموعه داده اصلی قرار می‌دهد و سپس پیش‌بینی‌های فردی آن‌ها (چه با رأی‌گیری^{۶۱} یا با میانگین‌گیری^{۶۲}) را جمع‌آوری می‌کند تا یک پیش‌بینی نهایی را تشکیل دهد. چنین فرابراوردگر معمولاً می‌تواند به عنوان راهی برای کاهش واریانس تخمین گر جعبه سیاه^{۶۳} (به عنوان مثال، درخت تصمیم)، با ورود تصادفی به مراحل ایجاد آن و سپس ساختن مجموعه‌ای از آن استفاده شود.



شکل (۵-۶) نحوه عملکرد الگوریتم دسته‌بندی کیسه‌ای

۵-۱-۴-۹- دسته‌بندی تقویتی گرادیان^{۶۴}

طبقه‌بندی‌کننده تقویتی گرادیان یک الگوریتم یادگیری ماشین است که بسیاری از مدل‌های یادگیری ضعیف را با هم ترکیب می‌کند تا یک مدل پیش‌بینی قوی ایجاد کند. معمولاً هنگام انجام الگوریتم دسته‌بندی تقویتی

^{۵۶} Bernoulli Naive Bayes

^{۵۷} Binary

^{۵۸} Bagging Classifier

^{۵۹} Meta-Estimator

^{۶۰} Ensemble

^{۶۱} Voting

^{۶۲} Averaging

^{۶۳} Black-Box Estimator

^{۶۴} Gradient Boosting Classifier Algorithm

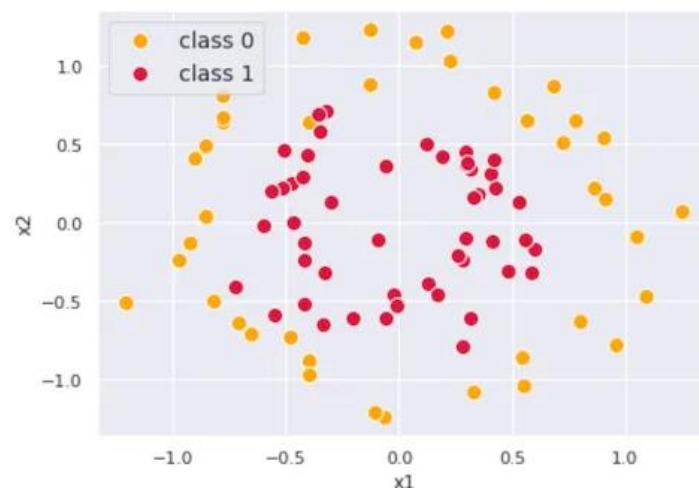
گرادیان از درختان تصمیم استفاده می‌شود. مدل تقویتی گرادیان به دلیل اثربخشی در طبقه‌بندی مجموعه داده‌های پیچیده، محبوب شده‌اند.

مدل تقویت گرادیان ترکیبی خطی از یک سری مدل‌های ضعیف است که به صورت تناوبی برای ایجاد یک مدل نهایی قوی ساخته شده است. این روش به خانواده الگوریتم‌های یادگیری گروهی تعلق دارد و عملکرد آن همواره از الگوریتم‌های اساسی یا ضعیف (مثلاً درخت تصمیم) یا روش‌های بر اساس کیسه‌گذاری (مانند جنگل تصادفی) بهتر است؛ اما این موضوع تا حدی از مشخصات داده‌های ورودی تأثیر می‌پذیرد.

روش این الگوریتم بدین ترتیب است که تابع هزینه^{۶۵} را به کمینه‌ترین مقدار خود برساند. در علم آمار، معمولاً تابع هزینه برای اینکه مشخص شود تخمین پارامترمان تا چه حد موفق بوده، استفاده می‌شود. تابعی است که برای سنجش میزان موفقیت تخمین‌گر از تخمین پارامتر نسبت به مقادیر واقعی از آن استفاده می‌شود. در مسائل طبقه‌بندی، تابع هزینه در اصل به نوعی تعداد طبقه‌بندی‌های اشتباه توسط تخمین‌گر را نمایان می‌کند.

الگوریتم یادگیری تقویتی یک الگوریتم تقویتی قدرتمند است که چندین یادگیرنده ضعیف را به یادگیرندگان قوی ترکیب می‌کند که در آن هر مدل جدید برای به حداقل رساندن تابع هزینه مانند میانگین مربعات خطا یا آنتروپی متقابل مدل قبلی، با استفاده از گرادیان نزول آموزش داده می‌شود. در هر تکرار، الگوریتم گرادیان تابع هزینه را با توجه به پیش‌بینی‌های مجموعه فعلی محاسبه می‌کند و سپس یک مدل ضعیف جدید را برای به حداقل رساندن این گرادیان آموزش می‌دهد. سپس پیش‌بینی‌های مدل جدید به مجموعه اضافه می‌شود و این فرآیند تا زمانی که یک معیار توقف برآورده شود، تکرار می‌شود. در این الگوریتم، وزن نمونه‌های آموزشی بهینه‌سازی نشده است، در عوض، هر پیش‌بینی‌کننده با استفاده از خطاهای باقی‌مانده قبلی به عنوان برچسب آموزش داده می‌شود.

در ادامه روند پیاده‌سازی مدل تقویتی گرادیان در قالب یک مثال آموزشی توضیح داده خواهد شد. شکل نمایش داده‌های دسته‌بندی در شکل زیر نشان داده شده است.



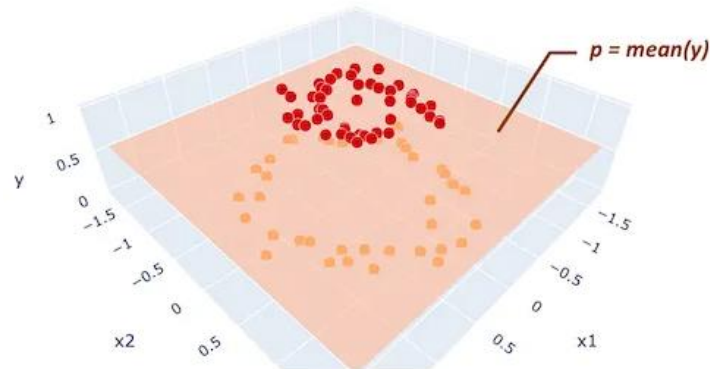
شکل (۵-۷) مثالی از مسئله دسته‌بندی دو کلاسه

^{۶۵} Loss Function

هدف ساخت یک مدل تقویتی گرادیان است که داده‌ها را به دودسته دسته‌بندی کند. اولین گام، ایجاد یک پیش‌بینی یکنواخت بر روی احتمال کلاس ۱ (ما آن را p می‌نامیم) برای تمام نقاط داده است که در واقع همان میانگین کلاس می‌باشد.

$$p = P(y = 1) = \bar{y}$$

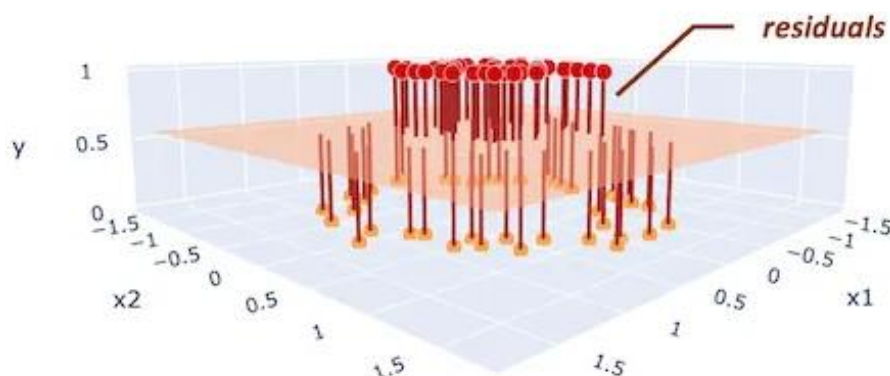
در اینجا یک نمایش سه‌بعدی از داده‌ها و پیش‌بینی اولیه آمده است. در این لحظه، پیش‌بینی فقط صفحه‌ای است که همیشه مقدار یکنواخت $p = \text{mean}(y)$ را در محور y دارد.



شکل (۵-۸) نمایش صفحه پیش‌بینی به شکل سه‌بعدی

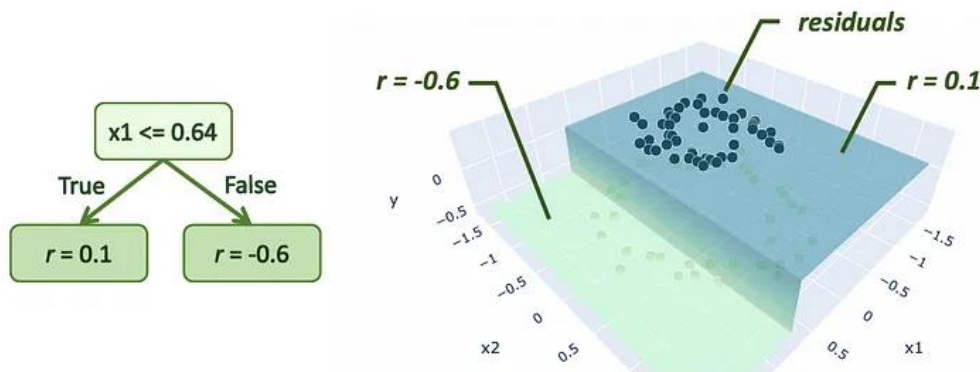
در این مثال، میانگین y ، ۰.۵۶ است. از آنجایی که بزرگ‌تر از ۰.۵ است، همه چیز با این پیش‌بینی اولیه در کلاس ۱ طبقه‌بندی می‌شود. ممکن است به نظر برسد که این پیش‌بینی ارزش یکسان، منطقی نیست، لازم به ذکر است که با اضافه کردن مدل‌های ضعیف بیشتر به آن، پیش‌بینی می‌یابد.

برای بهبود کیفیت پیش‌بینی، ممکن است روی باقی‌مانده‌ها (خطای پیش‌بینی) از پیش‌بینی اولیه تمرکز کنیم، زیرا این همان چیزی است که باید به حداقل برسد. باقی‌مانده‌ها به صورت $r_i = y_i - p$ تعریف می‌شوند (i نشان‌دهنده شاخص هر نقطه داده است). در شکل زیر باقی‌مانده‌ها به صورت خطوط قهوه‌ای نشان داده شده‌اند که خطوط عمود از هر نقطه داده به صفحه پیش‌بینی هستند.



شکل (۵-۹) نمایش باقی‌مانده‌ها

برای به حداقل رساندن این باقی مانده‌ها، یک مدل درخت رگرسیون با x_1 و x_2 به عنوان ویژگی‌های آن و باقی مانده r به عنوان هدف آن باید ساخته شود. اگر بتوان درختی ساخت که الگوهایی را بین x و r پیدا کند، می‌توان با استفاده از آن الگوهای یافت شده، باقی مانده‌های حاصل از پیش‌بینی اولیه p را کاهش داد. برای ساده کردن نمایش، درختان بسیار ساده‌ای که هر کدام فقط دارای یک تقسیم و دو گره برگ هستند، ساخته شده‌اند که به آن «استامپ»^{۶۶} می‌گویند. لازم به ذکر است که درخت‌های تقویت‌کننده گرادیان معمولاً درختان کمی عمیق‌تر مانند درخت‌هایی با ۸ تا ۳۲ گره برگ دارند. در اینجا ما اولین درخت ایجاد شده باقی مانده‌ها را با دو مقدار مختلف $r = (0.1, -0.6)$ پیش‌بینی می‌کند.



شکل (۵-۱۰) درخت ساخته شده برای متغیرهای x و باقی مانده r

در ادامه گاما طبق فرمول زیر محاسبه می‌شود. مقادیر گاما را به پیش‌بینی اولیه خود اضافه می‌کنیم تا باقی مانده‌ها را کاهش دهیم.

$$\gamma_j = \frac{\sum_{x_i \in R_j} (y_i - p)}{\sum_{x_i \in R_j} p(1 - p)}$$

γ is computed for each terminal node j

Aggregating for all the data points x_i that belongs to terminal node j

مقادیر گاما ۱ و گاما ۲ بدین ترتیب محاسبه می‌شوند.

$$\gamma_1 = \frac{\sum_{x_i \in R_1} (y_i - 0.56)}{\sum_{x_i \in R_1} 0.56 \cdot (1 - 0.56)} = 0.3$$

$$\gamma_2 = \frac{\sum_{x_i \in R_2} (y_i - 0.56)}{\sum_{x_i \in R_2} 0.56 \cdot (1 - 0.56)} = -2.2$$

برای اینکه گاما را به مقدار p اضافه شود، به شکل زیر عمل می‌کنیم. ابتدا مقدار $\log(\text{odds})$ را از p به دست می‌آوریم (به آن $F(x)$ گفته می‌شود). سپس گاما را به آن اضافه می‌کنیم.

^{۶۶} Stump

$$\log(\text{odds}) = \log\left(\frac{p}{1-p}\right)$$

برای اینکه مدل بیش از حد از آموزش^{۶۷} نبیند و خطای آن کاهش یابد می‌توان مقدار گاما را با یک وزنی (بین ۰ تا ۱) که به آن نرخ یادگیری^{۶۸} γ گفته می‌شود، ضرب کرد و سپس به مقدار $\log(\text{odds})$ یا همان $F(x)$ اضافه نمود تا پیش‌بینی بروز شود.

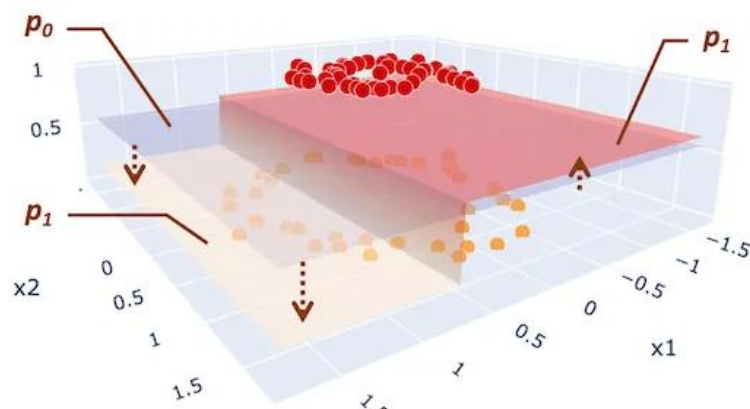
$$F_1(x) = F_0(x) + \gamma \cdot \gamma$$

Updated prediction
Initial prediction
Learning rate

در این مثال، ما از نرخ یادگیری نسبتاً بزرگ $\gamma = 0.9$ استفاده می‌کنیم تا فرآیند بهینه‌سازی را آسان‌تر درک کنیم، اما معمولاً قرار است مقادیر بسیار کوچک‌تری مانند ۰.۱ در نظر گرفته شود. با جایگزینی مقادیر واقعی برای متغیرهای سمت راست معادله بالا، پیش‌بینی به‌روز $F_1(x)$ را به دست می‌آید.

$$F_1(x) = \begin{cases} \log\left(\frac{0.56}{1-0.56}\right) + 0.9 \cdot 0.3 = 0.5 & \text{if } x_1 \leq 0.64 \\ \log\left(\frac{0.56}{1-0.56}\right) - 0.9 \cdot 2.2 = -1.7 & \text{otherwise} \end{cases}$$

اگر $\log(\text{odds})$ را دوباره به p تبدیل کنیم. شکی پله‌مانند از داده‌ها مانند شکل زیر به دست می‌آید.

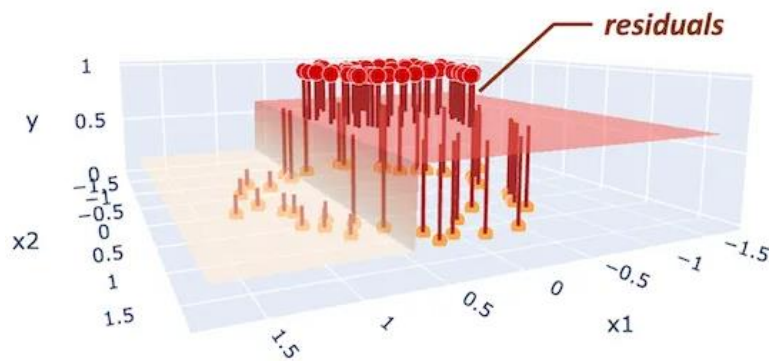


شکل (۵-۱۱) نمایش پیش‌بینی بروز شده

اکنون، باقی‌مانده‌های بروز شده r به شکل زیر است.

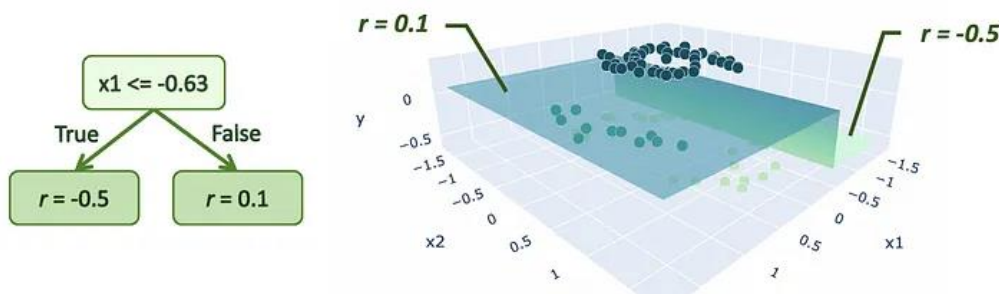
^{۶۷} Overfit

^{۶۸} Learning Rate



شکل (۵-۱۲) نمایش باقی‌مانده‌های بروز شده

مجدداً یک درخت رگرسیون با استفاده از همان x_1 و x_2 به عنوان ویژگی‌های ورودی برای باقی‌مانده‌های بروز شده ایجاد می‌کنیم.



شکل (۵-۱۳) درخت ساخته شده برای متغیرهای x و باقی‌مانده r بروز شده

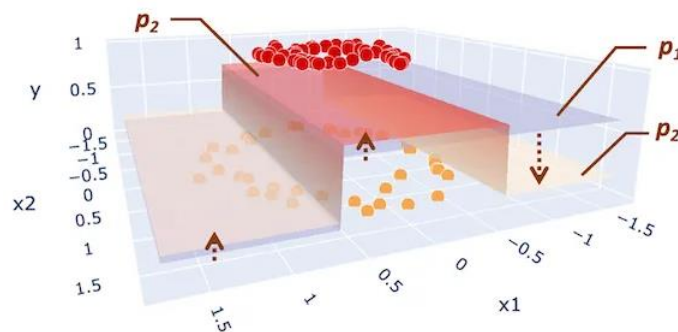
حال مجدداً گاما به همان روش قبل محاسبه کرده و $F_2(x)$ را به دست می‌آوریم.

$$F_2(x) = \begin{cases} F_1(x) - v \cdot 2.3 = 0.5 - 0.9 \cdot 2.3 = -1.6 & \text{if } x_1 \leq -0.63 \\ F_1(x) + v \cdot 0.4 = 0.5 + 0.9 \cdot 0.4 = 0.9 & \text{else if } -0.63 < x_1 \leq 0.64 \\ F_1(x) + v \cdot 0.4 = -1.7 + 0.9 \cdot 0.4 = -1.3 & \text{otherwise} \end{cases}$$

These are γ computed with this formula:

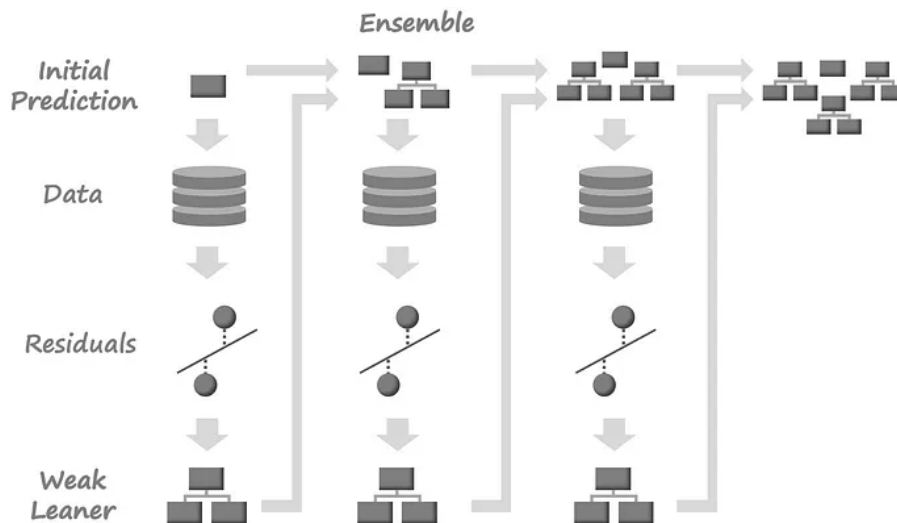
$$\gamma_j = \frac{\sum_{x_i \in R_j} (y_i - p)}{\sum_{x_i \in R_j} p(1 - p)}$$

سپس $F_2(x)$ را به $p_2(x)$ تبدیل می‌کنیم و به شکل زیر می‌رسیم.



شکل (۵-۱۴) نمایش پیش‌بینی بروز شده

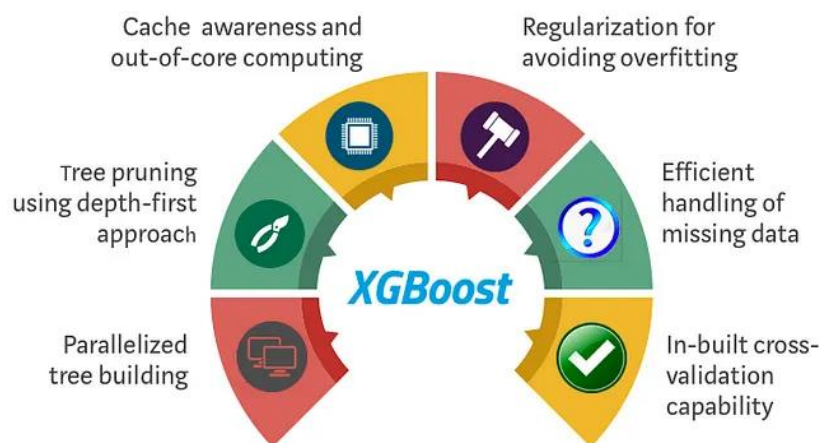
سپس، این مراحل را تکرار می‌کنیم تا زمانی که پیش‌بینی مدل متوقف شود. شکل‌های زیر روند بهینه‌سازی را از ۰ تا ۴ تکرار نشان می‌دهد. می‌توان دید که پیش‌بینی ترکیبی $p(x)$ به هدف ما نزدیک‌تر می‌شود؛ زیرا درخت‌های بیشتری را به مدل ترکیبی اضافه می‌کنیم. این روشی است که الگوریتم تقویتی گرادیان برای پیش‌بینی اهداف پیچیده با ترکیب چندین مدل ضعیف انجام می‌دهد. تصویر زیر به طور خلاصه کل فرآیند این الگوریتم را نشان می‌دهد.



شکل (۵-۱۶) فرآیند الگوریتم تقویتی گرادیان

۵-۱-۴-۳-۱۰- دسته‌بندی XGboost^{۶۹}

XGBoost یک الگوریتم یادگیری ماشین مبتنی بر درخت تصمیم است که از یک چارچوب تقویت گرادیان استفاده می‌کند. شکل زیر فرآیند بهینه‌سازی الگوریتم ماشین تقویت گرادیان^{۷۰} را توسط XGboost نشان می‌دهد.



شکل (۵-۱۷) روند بهینه‌سازی در الگوریتم XGboost

^{۶۹} XGboost Classifier

^{۷۰} Gradient Boosting Machines (GBMs)

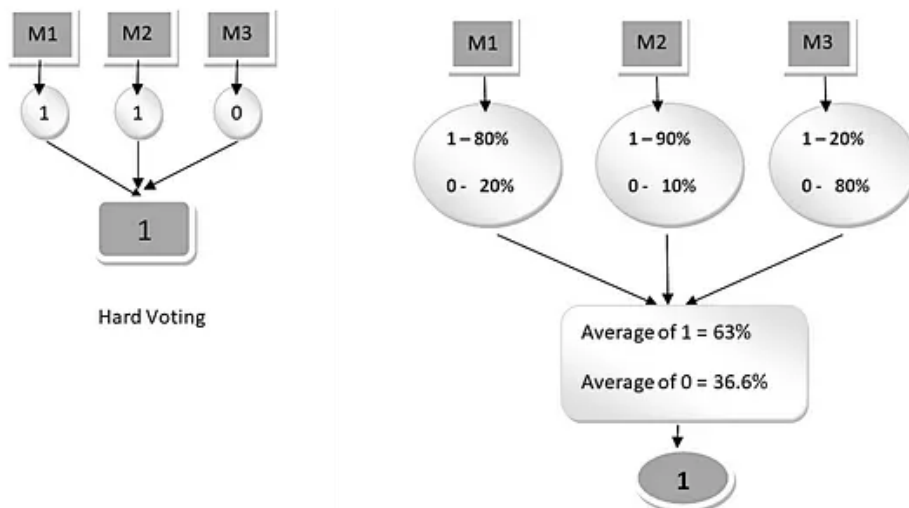
XGBoost مخفف واژه تقویت گرادیان شدید^{۷۱} است و به دلیل توانایی آن در مدیریت مجموعه داده‌های بزرگ و توانایی آن برای دستیابی به عملکرد پیشرفته در بسیاری از وظایف یادگیری ماشین، به یکی از محبوب‌ترین و پرکاربردترین الگوریتم‌های یادگیری ماشین تبدیل شده است.

در این الگوریتم درخت‌های تصمیم به صورت متوالی ایجاد می‌شوند. وزن‌ها نقش مهمی در XGBoost دارند. وزن‌ها به همه متغیرهای مستقل اختصاص داده می‌شوند که سپس به درخت تصمیم که نتایج را پیش‌بینی می‌کند، وارد می‌شوند. وزن متغیرهای پیش‌بینی شده اشتباه توسط درخت، افزایش می‌یابد و این متغیرها سپس به درخت تصمیم دوم تغذیه می‌شوند. سپس این طبقه‌بندی‌کننده‌ها یا پیش‌بینی‌کننده‌های منفرد برای ارائه یک مدل قوی و دقیق‌تر جمع می‌شوند.

۵-۱-۴-۱۱- دسته‌بندی بر اساس رأی‌گیری^{۷۲}

این الگوریتم، یک تخمین‌گر یادگیری ماشین است که مدل‌های پایه یا برآوردهای مختلفی را آموزش می‌دهد و بر اساس جمع‌آوری یافته‌های هر تخمین‌گر پایه، پیش‌بینی می‌کند. معیارهای تجمیع می‌تواند دو نوع باشد:

- سخت: دسته‌بندی بر اساس کلاس خروجی پیش‌بینی شده محاسبه می‌شود.
- نرم: دسته‌بندی بر اساس احتمال پیش‌بینی شده کلاس خروجی محاسبه می‌شود.



شکل (۵-۱۸) نحوه تجمیع مدل‌ها و پیش‌بینی الگوریتم دسته‌بندی بر اساس رأی‌گیری

۵-۱-۵- اعتبارسنجی و بررسی صحت الگوریتم‌ها

پس از پیاده‌سازی و آموزش مدل‌های روی داده‌های آموزشی، اعتبارسنجی الگوریتم‌ها با داده‌های اعتبارسنجی انجام شد. در ادامه به معیارهای اعتبارسنجی الگوریتم‌ها می‌پردازیم.

^{۷۱} Extreme Gradient Boosting

^{۷۲} Voting Classifier

۵-۱-۵-۱- ماتریس اغتشاش^{۷۳}

برای اینکه بتوانیم نتایج دسته‌بندی الگوریتم را با داده‌های واقعی مقایسه کنیم، از ماتریس اغتشاش استفاده می‌کنیم. جدول زیر، ماتریس اغتشاش را نشان می‌دهد.

جدول (۲-۵) ماتریس اغتشاش

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

سطرها مقادیر واقعی و ستون‌ها مقادیر پیش‌بینی شده را نشان می‌دهند. سلول‌های این ماتریس مفاهیم زیر را ارائه می‌دهند.

- **مثبت-صحیح^{۷۴}**: نشان می‌دهد که مدل یک نتیجه درست را پیش‌بینی کرده است و مشاهده واقعی درست بوده است.
- **مثبت-کاذب^{۷۵}**: نشان می‌دهد که مدل یک نتیجه درست را پیش‌بینی کرده است، اما مشاهده واقعی نادرست بوده است.
- **منفی-کاذب^{۷۶}**: نشان می‌دهد که مدل یک نتیجه نادرست را پیش‌بینی کرده است، درحالی‌که مشاهده واقعی درست بوده است.
- **منفی-صحیح^{۷۷}**: نشان می‌دهد که مدل یک نتیجه نادرست را پیش‌بینی کرده است، درحالی‌که نتیجه واقعی نیز نادرست بوده است.

۵-۱-۵-۲- دقت^{۷۸}

دقت معمولاً برای قضاوت در مورد عملکرد مدل استفاده می‌شود، فرمول محاسبه دقت به شکل زیر است.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

در واقع دقت، میزان پیش‌بینی درست مدل را بر کل محاسبه می‌کند.

۵-۱-۵-۳- صحت^{۷۹}

این معیار، اندازه‌گیری مثبت‌های واقعی نسبت به تعداد کل مثبت‌های پیش‌بینی شده توسط مدل را محاسبه می‌کند. در واقع این معیار، میزان مثبت‌بودن پیش‌بینی‌های مثبت مدل را اندازه‌گیری می‌کند.

^{۷۳} Confusion Matrices

^{۷۴} True Positive

^{۷۵} False Positive

^{۷۶} False Negative

^{۷۷} True Negative

^{۷۸} Accuracy

^{۷۹} Precision

$$Precision = \frac{TP}{TP + FP}$$

۵-۱-۴- پوشش^{۸۰}

معیار پوشش قادر به سنجش مثبت پیش‌بینی شده مدل نسبت به تعداد پیامدهای مثبت واقعی است. با استفاده از این معیار، می‌توان ارزیابی کرد که مدل چقدر قادر به شناسایی نتایج واقعی است.

$$Recall = \frac{TP}{TP + FN}$$

۵-۱-۵- امتیاز F۱^{۸۱}

این معیار، میانگین هارمونیک بین دقت و پوشش است. فرمول آن به شکل زیر است.

$$F1\ Score = \frac{2 * (Precision * Recall)}{Precision + Recall}$$

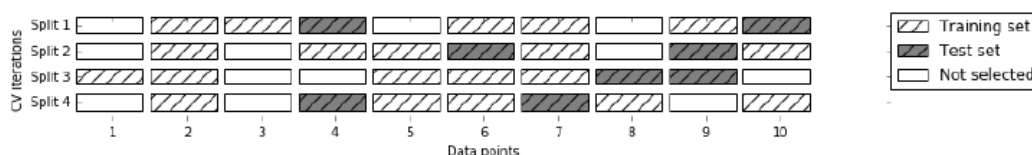
این معیار هنگامی که داده‌ها به صورت نامتوازن پخش شده‌اند، دید بهتری از عملکرد مدل ارائه می‌دهد.

۵-۱-۶- اعتبارسنجی متقابل^{۸۲}

اعتبارسنجی متقابل یک روش آماری است که برای تخمین عملکرد مدل‌های یادگیری ماشین استفاده می‌شود. این روش برای ارزیابی چگونگی تعمیم نتایج یک تحلیل آماری به یک مجموعه داده دیده نشده است. این روش، آموزش بیش از حد مدل را شناسایی می‌کند و با بررسی داده‌های دیده نشده نتیجه دقیق‌تری راجع به عملکرد مدل، ارائه می‌دهد.

در این پژوهش ما از روش اعتبارسنجی متقابل مونت کارلو^{۸۳} برای ارزیابی نهایی عملکرد الگوریتم‌ها روی داده‌های دیده نشده، استفاده کرده‌ایم. در ادامه به نحوه عملکرد این معیار اعتبارسنجی می‌پردازیم.

این معیار، یک استراتژی بسیار انعطاف‌پذیر برای اعتبارسنجی متقابل است. در این تکنیک، مجموعه داده‌ها به طور تصادفی به مجموعه‌های آموزشی و اعتبارسنجی تقسیم می‌شوند. درصدی از مجموعه داده‌ای را که قرار است به عنوان مجموعه آموزشی استفاده شود و درصدی که به عنوان مجموعه اعتبارسنجی استفاده می‌شود، را مشخص می‌کنیم. اگر مجموع درصدها به ۱۰۰ نرسد، از مجموعه داده باقی مانده استفاده نمی‌شود. سپس این تقسیم‌بندی به تعداد دفعاتی که مشخص می‌کنیم، تکرار می‌شود و دقت هر تکرار محاسبه می‌گردد. می‌توان میانگین نهایی دقت دفعات تکرار را به عنوان میزان دقت نهایی مدل روی داده‌های دیده نشده در نظر گرفت.



شکل (۵-۱۹) نحوه عملکرد اعتبارسنجی متقابل مونت کارلو

^{۸۰} Recall

^{۸۱} F۱ Score

^{۸۲} Cross Validation

^{۸۳} Monte Carlo Cross-Validation (Shuffle Split)

۵-۱-۵-۷- منحنی ROC – AUC^{۸۴}

این معیار ارزیابی روی داده‌های دیده‌نشده برای مقایسه نهایی مدل‌ها، پیاده‌سازی شده است. برای درک بهتر این معیار، مفاهیم زیر مطرح می‌گردد.

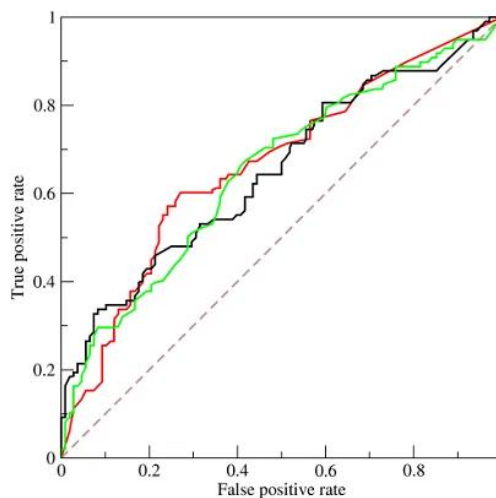
• **نرخ مثبت صحیح^{۸۵}:** یک معیار ارزیابی عملکرد می‌باشد. همان مفهوم معیار پوشش می‌باشد و مشخص می‌کند که به چه نسبتی پیش‌بینی صحیح صورت گرفته است. فرمول آن به شکل زیر است.

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN}$$

• **نرخ مثبت کاذب^{۸۶}:** یک معیار ارزیابی عملکرد می‌باشد و نشانگر تعداد شناسایی‌های مثبت از میان مشاهدات منفی است. فرمول آن به شکل زیر است.

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN}$$

• **منحنی مشخصه عملکرد^{۸۷}:** یک منحنی مشخصه عملکرد، یک نمودار برای نمایش توانایی ارزیابی یک سیستم دسته‌بندی دودویی محسوب می‌شود که آستانه تشخیص آن نیز متغیر است. این منحنی توسط ترسیم نرخ مثبت صحیح بر حسب نرخ مثبت کاذب، ایجاد می‌شود. نمودار زیر منحنی مشخصه عملکرد را برای سه مدل دسته‌بندی مختلف نشان می‌دهد.



نمودار (۵-۲) منحنی مشخصه عملکرد برای سه روش مختلف دسته‌بندی

باتوجه به نمودار زیر، بهترین عملکرد دسته‌بندی در این نمودار در نقطه‌ای با مختصات (۰,۱) رخ خواهد داد که در آن کمترین نرخ اشتباه و بیشترین نرخ بازیابی یا حساسیت را داریم. این نقطه بیانگر «بهترین دسته‌بندی^{۸۸}» است.

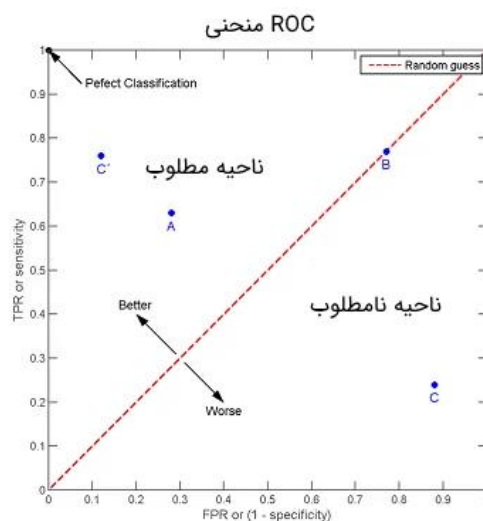
^{۸۴} AUC (Area Under the Curve) - ROC (Receiver Operating Characteristics) curve

^{۸۵} True Positive Rate (TPR)

^{۸۶} False Positive Rate (FPR)

^{۸۷} ROC: Receiver Operating Characteristics curve

^{۸۸} Perfect Classification



نمودار (۳-۵) نواحی مطلوب و نامطلوب در منحنی ROC

همچنین در نمودار فوق، خط منقطعی که از میان نمودار عبور کرده و نقطه $(0,0)$ را به $(1,1)$ پیوند می‌دهند، حدس تصادفی است که به صورت ناحیه $50\%-50\%$ نیز شناخته می‌شود. اگر نقطه‌ای روی این خط منقطع قرار گرفته باشد، تشخیص درستی نسبت به قرارگیری در هر گروه، برایش وجود ندارد. در حقیقت در نیمی از موارد می‌تواند در یک دسته و در نیمی از موارد نیز در دسته دیگری، طبقه‌بندی شود و نقشی در تعیین خطا نخواهد داشت. یکی از نمونه‌های معروف برای دسته‌بندی به صورت تصادفی، تصمیم تعلق نقطه به هر یک از دو گروه به وسیله پرتاب سکه است. هر چه تعداد نمونه‌ها در دسته‌بندی تصادفی بیشتر شود، این خط به قطر نواحی ROC نزدیک‌تر خواهد شد.

- AUC: مساحت زیر منحنی مشخصه عملکرد را AUC می‌گویند که نشان‌دهنده درجه یا معیار تفکیک‌پذیری است. این معیار نشان می‌دهد که مدل چقدر می‌تواند بین کلاس‌ها تمایز قائل شود. یک مدل عالی دارای $AUC = 1$ است

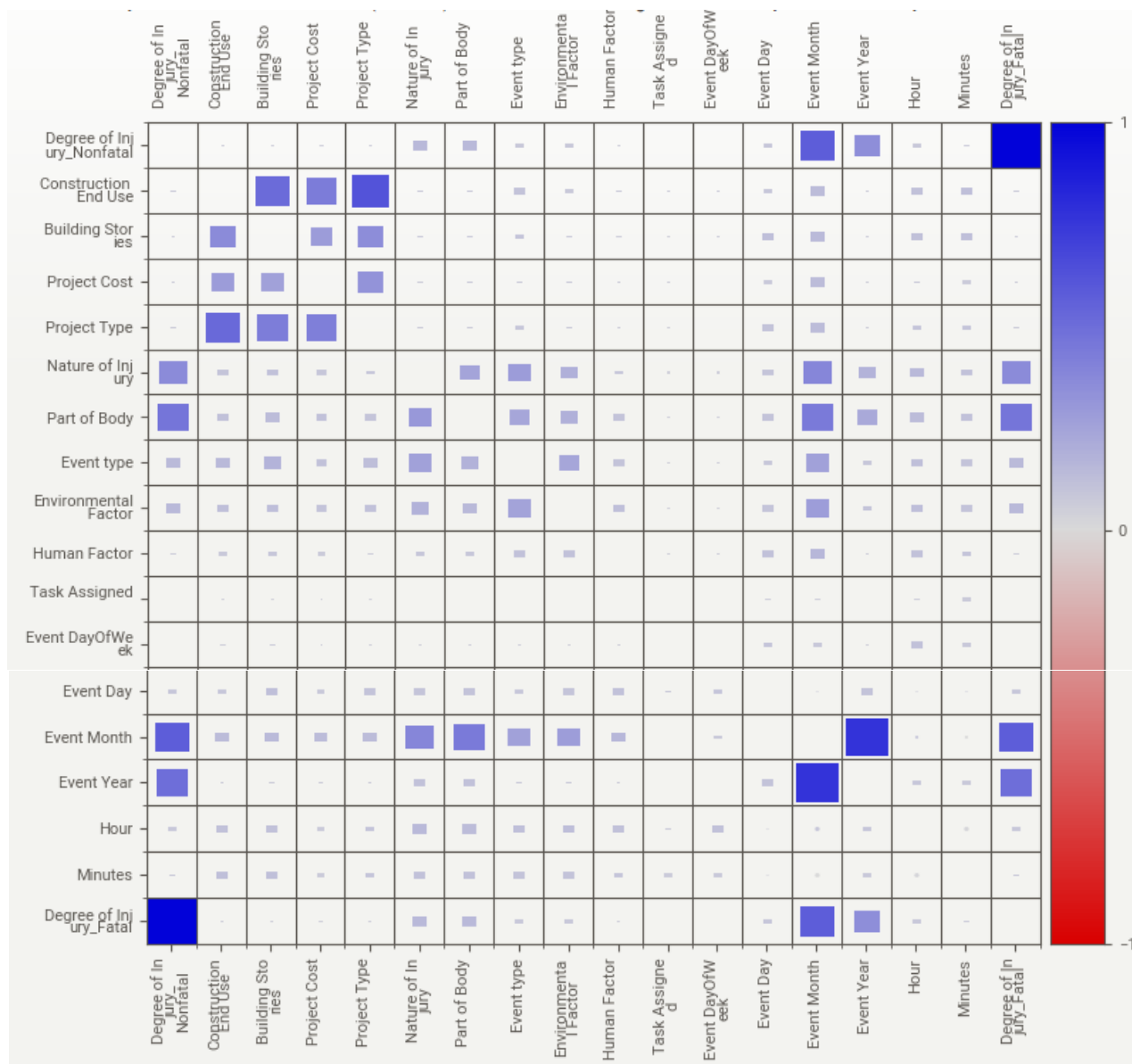
۵-۱-۶- بررسی نتایج و مقایسه الگوریتم‌ها

در این بخش، مقایسه الگوریتم‌ها بر اساس معیارهای اعتبارسنجی شده، صورت گرفت و نتایج در قالب جدول ارائه گردیده است.

۵-۲- یافته‌های تحقیق

۵-۲-۱- تجزیه و تحلیل اکتشافی

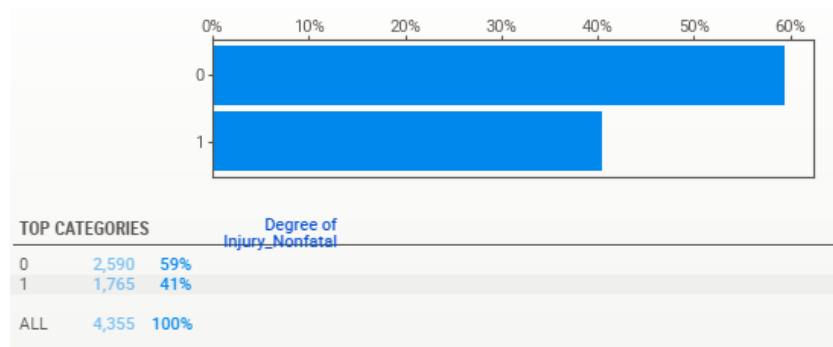
ابتدا همبستگی میان ویژگی‌ها بررسی شده است. شکل زیر، همبستگی بین ویژگی‌ها را نشان می‌دهد. همبستگی یک معیار آماری است که میزان ارتباط خطی دو متغیر را بیان می‌کند (به این معنی که آن‌ها با هم با یک نرخ ثابت تغییر می‌کنند). این یک ابزار رایج برای توصیف روابط ساده بدون اظهارنظر در مورد علت و معلول است. همبستگی عددی بین -1 و $+1$ است. همبستگی مثبت، نشان‌دهنده میزان افزایش یا کاهش آن متغیرها به صورت موازی است و همبستگی منفی نشان‌دهنده میزان افزایش یک متغیر با کاهش متغیر دیگر است.



شکل (۵-۲۰) همبستگی ویژگی‌ها در مجموعه داده

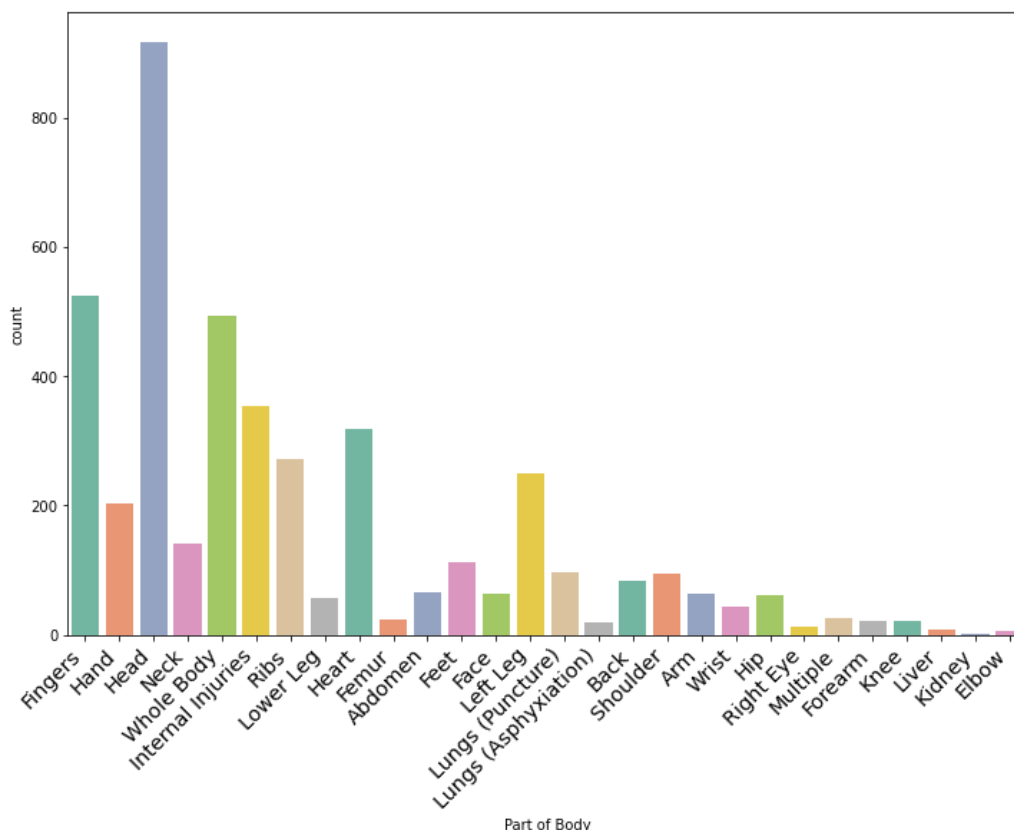
همان‌طور که از شکل فوق مشاهده می‌گردد، ویژگی‌های ماهیت آسیب‌دیدگی و ناحیه آسیب‌دیده در بدن همبستگی نسبتاً بالایی (حدود ۰.۳) با کشنده و غیرکشنده بودن حادثه دارند. همچنین ویژگی استفاده نهایی از سازه با ویژگی‌های هزینه پروژه، تعداد طبقات و نوع پروژه همبستگی نسبتاً بالایی دارد (بیشتر از ۰.۵). بین سایر ویژگی‌ها نیز همبستگی کمی مشاهده می‌گردد.

متغیر هدف در این مطالعه، کشنده یا غیرکشنده بودن یک حادثه می‌باشد. همان‌طور که در شکل زیر مشاهده می‌شود، در مجموعه داده مورد بررسی، ۴۱ درصد حوادث غیرکشنده و ۵۹ درصد حوادث کشنده بوده‌اند. این موضوع نشان می‌دهد که مجموعه داده مورد بررسی متعادل بوده و نیازی به متعادل‌سازی مجموعه داده نمی‌باشد.



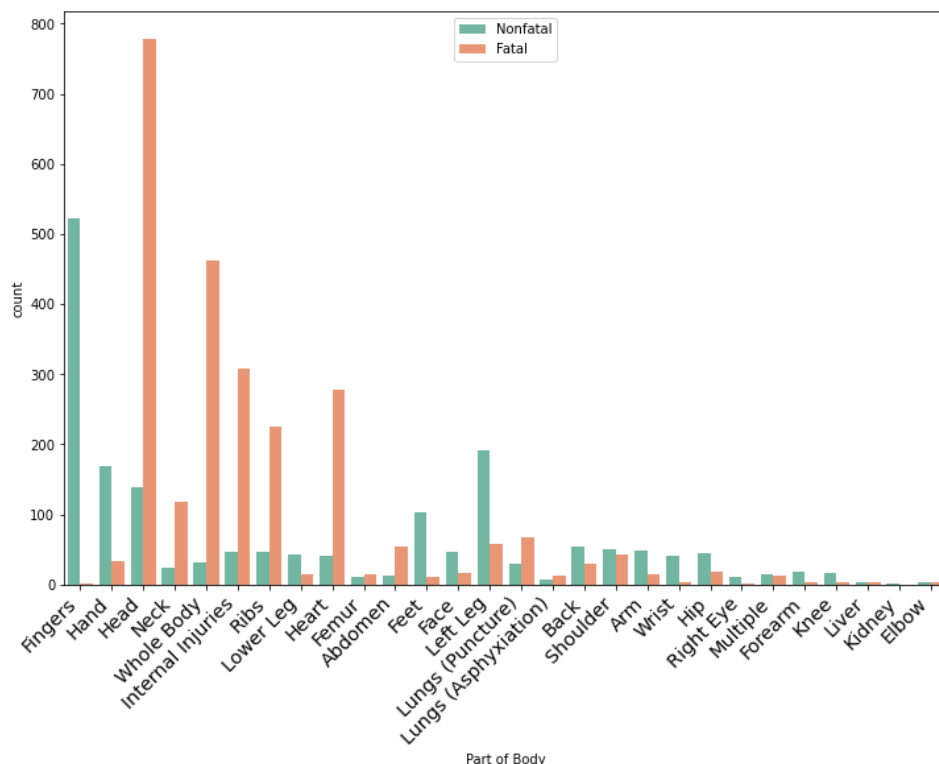
نمودار (۴-۵) متغیر هدف مطالعه

در ادامه به بررسی نواحی آسیب‌دیده در بدن در مجموعه داده پرداخته شده است.

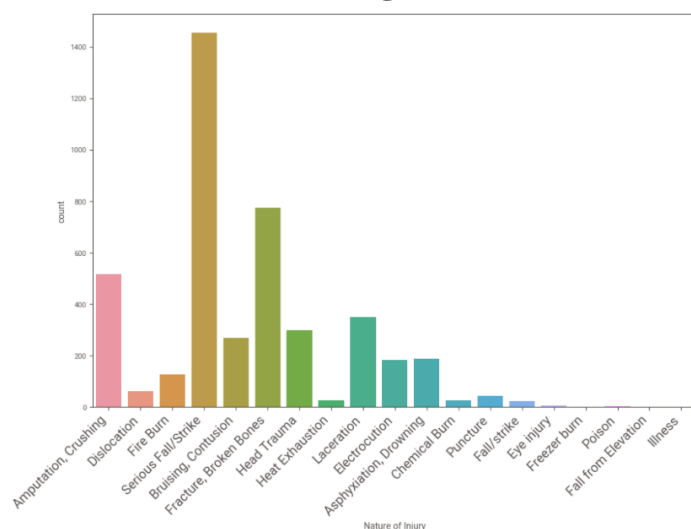


نمودار (۵-۵) نمودار نواحی آسیب‌دیده در بدن

همان‌طور که مشاهده می‌شود، بیشترین میزان آسیب‌دیدگی در این مجموعه داده، مربوط به آسیب به سر (۹۱۸ نفر معادل ۲۱ درصد افراد) می‌باشد. آسیب به انگشتان (۱۲ درصد معادل ۵۲۴ نفر)، آسیب به کل بدن (۱۱ درصد معادل ۴۹۳ نفر) و آسیب‌های داخلی (۸ درصد معادل ۳۵۴ نفر) رتبه‌های دوم تا پنجم را در میان نواحی آسیب‌دیده در بدن را در اختیار دارند. نکته قابل‌توجه در این زمینه این است که ۸۵ درصد افرادی که از ناحیه سر دچار آسیب شده‌اند، فوت کرده‌اند و این میزان نشان از اهمیت استفاده از کلاه ایمنی مناسب در پروژه‌های ساختمانی دارد. نمودار زیر نواحی آسیب‌دیده در بدن را به تفکیک کشنده یا غیرکشنده بودن حادثه نشان می‌دهد.

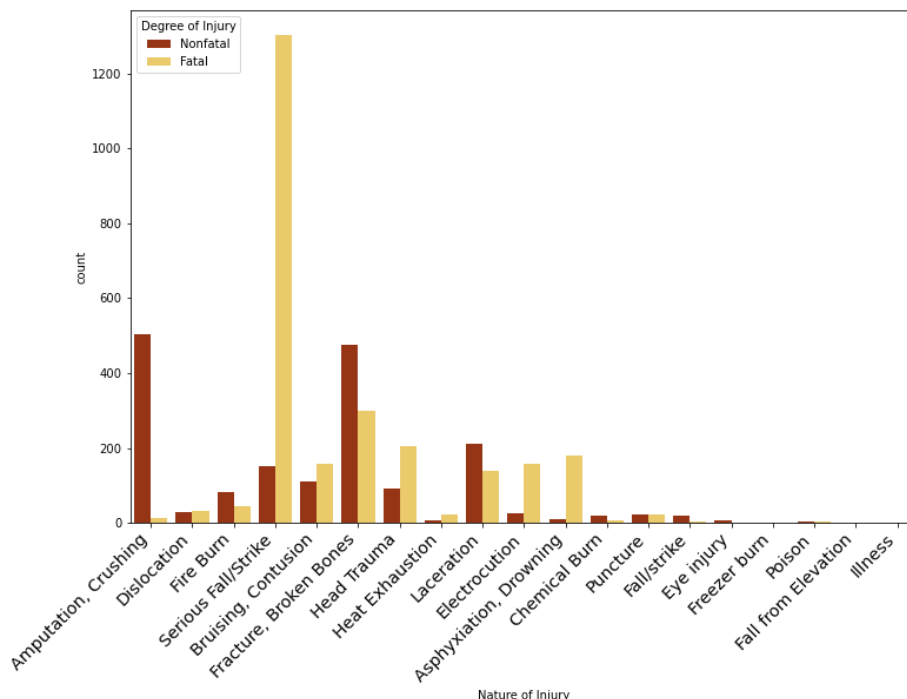


نمودار (۵-۶) نمودار نواحی آسیب دیده در بدن به تفکیک کشنده و غیر کشنده بودن حادثه در ادامه ماهیت آسیب در مجموعه داده مورد بررسی قرار گرفته است.



نمودار (۵-۷) نمودار میزان ماهیت آسیب در مجموعه داده

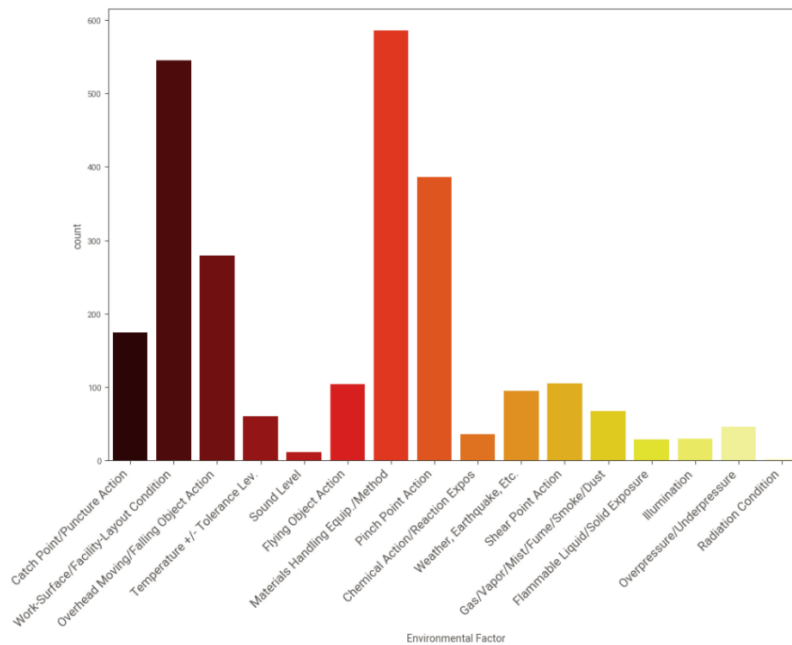
همان طور که در نمودار فوق مشاهده می گردد، سقوط جدی / ضربه بیشترین میزان حوادث (۳۳ درصد معادل ۱۴۵۴ نفر) را شامل می شوند. پس از آن شکستگی استخوان (۱۸ درصد معادل ۷۷۴ نفر)، قطع عضو (۱۲ درصد معادل ۵۱۶ نفر)، پارگی (۸ درصد معادل ۳۵۱ نفر) و ضربه به سر (۷ درصد معادل ۲۹۸ نفر) به ترتیب رتبه های دوم تا پنجم را در میان ماهیت آسیب در حوادث مورد بررسی در اختیار دارند.



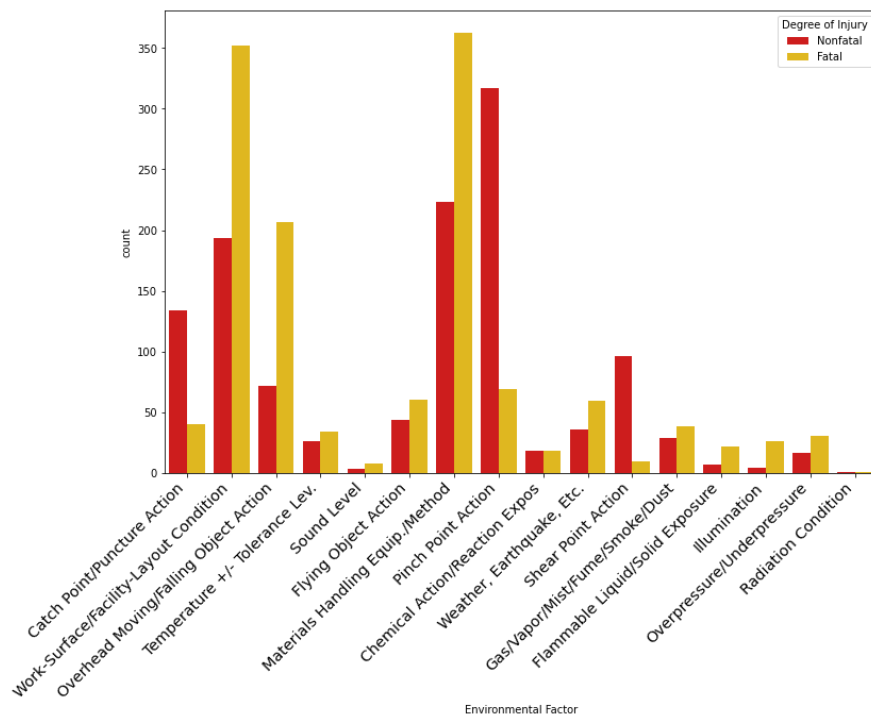
نمودار (۵-۸) نمودار میزان ماهیت آسیب به تفکیک کشنده یا غیرکشنده بودن حادثه

باتوجه به نمودار فوق، می توان متوجه شد که اکثر افرادی که دچار سقوط جدی / ضربه شده اند (حدود ۹۰ درصد موارد) فوت کرده اند. این موضوع نشان می دهد که ایمن سازی پرتگاه ها در پروژه های ساختمانی می تواند به عنوان یک اولویت برای جلوگیری از حوادث منجر به مرگ در حوادث ساختمانی در دستور کار قرار گیرد. سایر آسیب ها نیز به تفکیک کشنده یا غیرکشنده بودن حادثه در نمودار فوق قابل مشاهده است.

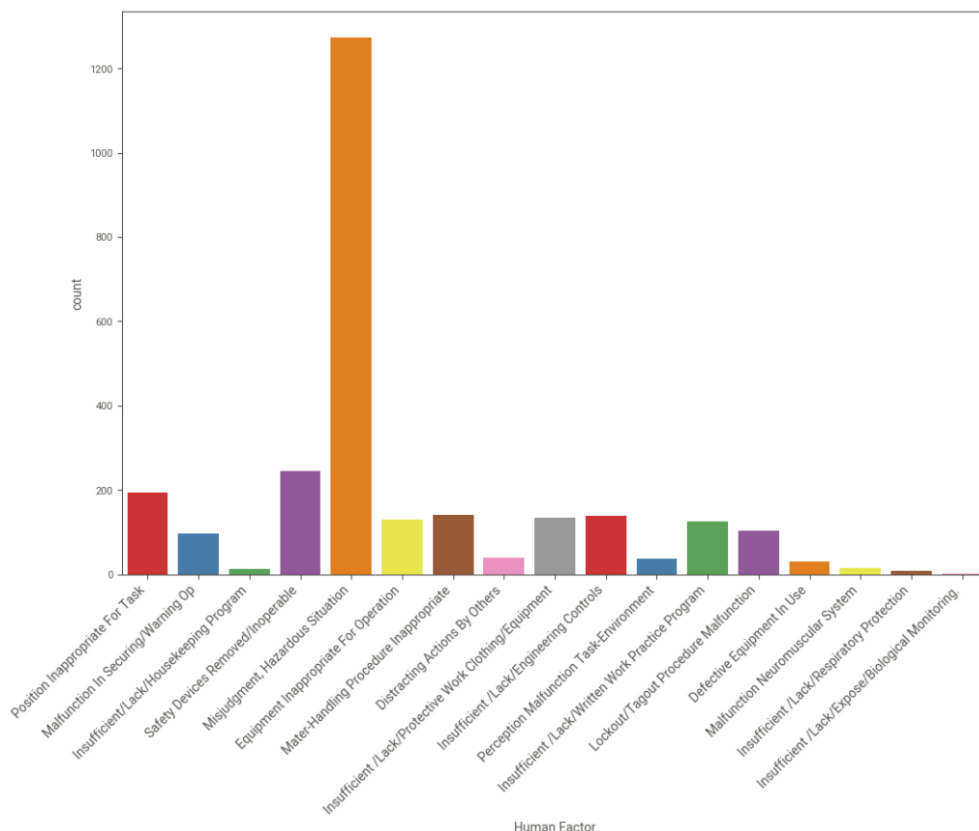
در میان عوامل محیطی ذکر شده در مجموعه داده، باتوجه به نمودار زیر، تجهیزات جابه جایی مواد (۱۳ درصد معادل ۵۱۳ نفر)، وضعیت سطح کار و چیدمان (۱۲ درصد معادل ۵۴۵ نفر)، رفتن دست در دستگاه (۴ درصد معادل ۱۷۴ نفر)، سقوط شی روی کارگر (۶ درصد معادل ۲۷۹ نفر) و سوراخ کردن (۴ درصد معادل ۱۷۴ نفر) بیشترین میزان را در میان عوامل محیطی داشته اند. با بررسی دقیق تر آمار می توان متوجه شد که تجهیزات جابه جایی مواد و وضعیت چیدمان سطح کار کارگر مجموعاً ۲۸ درصد مرگ و میرهای حوادث مورد بررسی را شامل می شوند که این موضوع نشان از اهمیت ایمن سازی حرکت تجهیزات و جابه جایی مواد و همچنین توجه به چیدمان سطح کار کارگر دارد. نمودارهای زیر، عوامل محیطی حوادث مورد بررسی و تفکیک آن ها را به کشنده یا غیرکشنده بودن حادثه را نشان می دهند.



نمودار (۵-۹) نمودار میزان عوامل محیطی در مجموعه داده

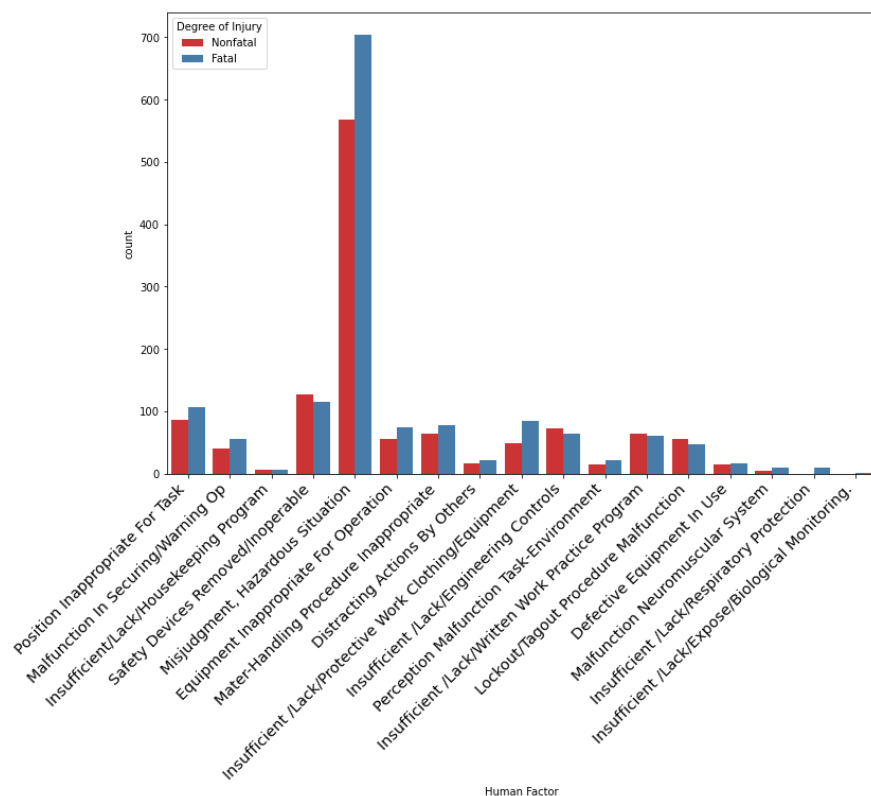


نمودار (۵-۱۰) نمودار میزان عوامل محیطی به تفکیک کشنده یا غیرکشنده بودن حادثه در گام بعدی، عوامل انسانی تأثیرگذار بر حوادث ساختمانی مورد بررسی قرار گرفته‌اند.

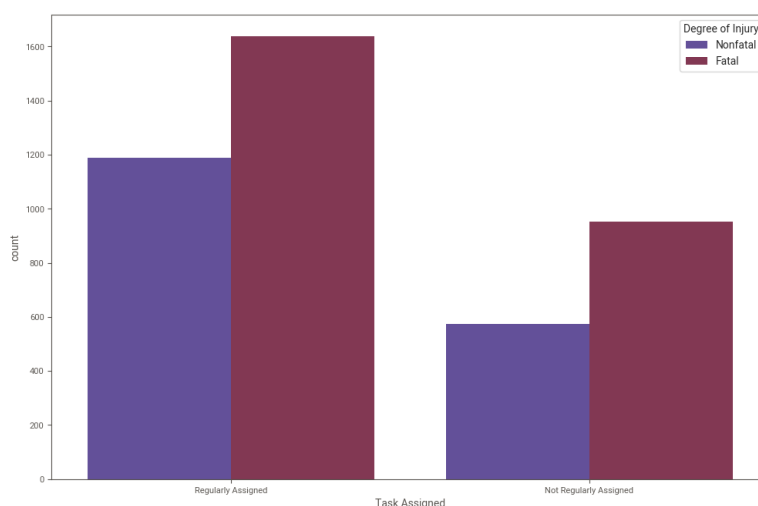


نمودار (۵-۱۱) نمودار عوامل انسانی مؤثر بر حوادث ساختمانی

همان طور که در نمودار فوق مشاهده می شود، قضاوت نادرست و قرارگیری در موقعیت خطرناک (۲۹ درصد معادل ۱۲۷۳ نفر) بیشترین میزان در عوامل انسانی را شامل می شود. آموزش قرارگیری درست به کارگر در موقعیت های مختلف و شبیه سازی موقعیت ها برای کارگر جهت تصمیم گیری بهتر کارگر می تواند بسیار مؤثر عمل کند. همچنین عدم استفاده از تجهیزات ایمنی (۶ درصد معادل ۲۴۴ نفر)، موقعیت نامناسب بر کار (۴ درصد معادل ۱۹۳ نفر)، استفاده از روش نامناسب جابه جایی مواد (۳ درصد معادل ۱۴۱ نفر)، نظارت های مهندسی ناکافی (۳ درصد معادل ۱۳۸ نفر) در رتبه های بعدی عوامل انسانی مؤثر بر حوادث ساختمانی قرار دارند.

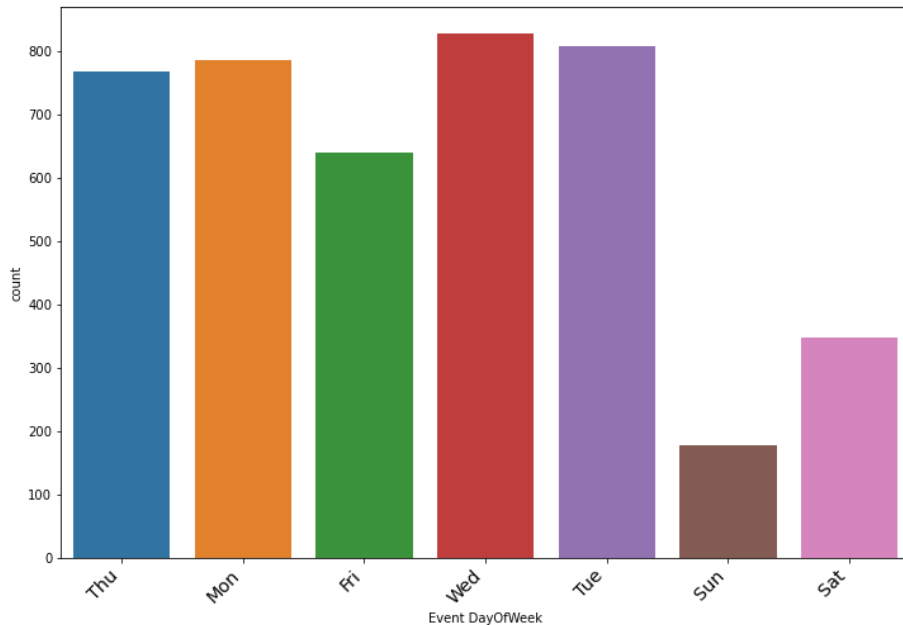


نمودار (۵-۱۲) نمودار میزان عوامل انسانی مؤثر بر حوادث ساختمانی به تفکیک کشنده یا غیرکشنده بودن حوادث در نمودار فوق، عوامل انسانی مؤثر بر حوادث ساختمانی به تفکیک کشنده یا غیرکشنده بودن حوادث نشان داده شده است. ۵۵ درصد از افرادی که قضاوت نادرست داشته‌اند و در موقعیت خطرناک قرار گرفته‌اند، فوت کرده‌اند که این موضوع همچنان بر آموزش درست این موارد به کارگران تأکید دارد. ویژگی بعدی که مورد بررسی قرار گرفته است، نوع وظیفه تخصیص یافته است. نمودار زیر نشان می‌دهد که اگر وظیفه به طور نامنظم به کارگر تخصیص داده شود احتمال فوت بر اثر حادثه برای کارگر بیشتر می‌شود.

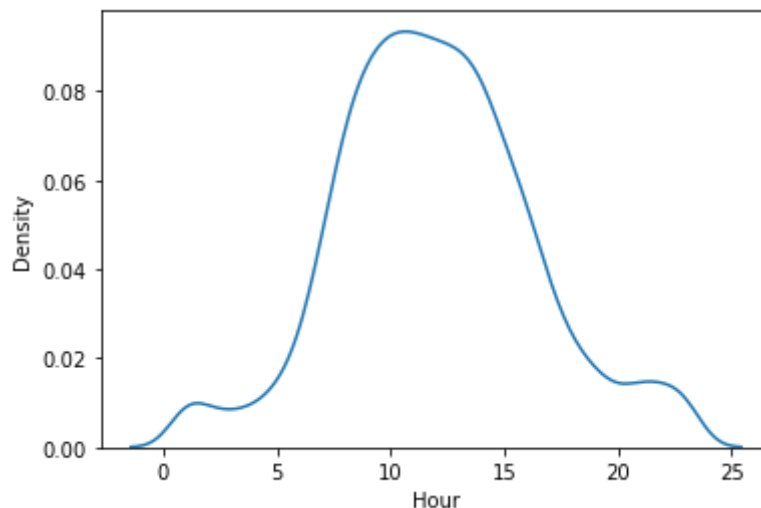


نمودار (۵-۱۳) نمودار نوع وظیفه اختصاص یافته به کارگر به تفکیک کشنده یا غیرکشنده بودن حادثه

در ۶۲ درصد حوادث منجر به مرگ، وظیفه اختصاص یافته به کارگر به شکل نامنظم بوده است. این آمار هنگامی که وظیفه اختصاص یافته به کارگر به طور منظم بوده است، حدود ۵۸ درصد است. نمودار زیر، تعداد حوادث ساختمانی را در روزهای هفته مورد بررسی قرار می دهد. همان طور که مشاهده می شود، در روزهای شنبه و یکشنبه میزان حوادث کمتر سایر روزهای هفته می باشد که این موضوع می تواند به علت تعطیلی این روزها در ایالات متحده باشد.



نمودار (۵-۱۴) نمودار تعداد حوادث ساختمانی در مجموعه داده طی روزهای هفته در پایان این بخش، شکل توزیع حوادث در ساعات مختلف روز مورد بررسی قرار گرفته است.



نمودار (۵-۱۵) نمودار توزیع چگالی احتمال ساعات وقوع حوادث

همان طور که در نمودار توزیع چگالی احتمال فوق مشاهده می شود، ساعات وقوع حوادث از توزیع نرمال پیروی می کند و بیشتر حوادث در بازه ۹ صبح الی ۳ بعد از ظهر رخ داده اند.

۵-۲-۲- استخراج قوانین انجمنی

الگوریتم Apriori برای استخراج قوانین انجمنی در این مطالعه استفاده شده است. حداقل حد آستانه پشتیبانی، ۰.۵ در نظر گرفته شده است. جدول زیر تعدادی از قوانین را نشان می‌دهد که بیشترین تکرار را در مجموعه داده داشته‌اند. در واقع بیشترین میزان پشتیبانی را دارند.

جدول (۵-۳) ترتیب قوانین انجمنی استخراج شده بر اساس معیار پشتیبانی از زیاد به کم

support	itemssets
2 0.873249	(Project Cost_)
1 0.851435	(Building Stories_)
9 0.814007	(Project Cost_ , Building Stories_)
0 0.780941	(Construction End Use_)
7 0.779564	(Project Cost_ , Construction End Use_)
6 0.773823	(Building Stories_ , Construction End Use_)
15 0.772675	(Project Cost_ , Building Stories_ , Construct...
3 0.769460	(Project Type_)
12 0.769001	(Project Cost_ , Project Type_)
8 0.769001	(Construction End Use_ , Project Type_)
17 0.768772	(Project Cost_ , Construction End Use_ , Proje...
10 0.766246	(Building Stories_ , Project Type_)
16 0.766016	(Building Stories_ , Construction End Use_ , P...
18 0.766016	(Project Cost_ , Building Stories_ , Project T...
20 0.765786	(Project Cost_ , Building Stories_ , Construct...
5 0.649139	(Task Assigned_Regularly Assigned)
4 0.594719	(Degree of Injury_Fatal)
14 0.559357	(Project Cost_ , Task Assigned_Regularly Assig...

همان‌طور که مشاهده می‌شود، هزینه پروژه و تعداد طبقات سازه، هزینه پروژه و استفاده نهایی از سازه، در میان موارد دوتایی بیشترین تعداد را داشته‌اند. جدول قوانین انجمنی استخراج شده از جدول فوق را نشان می‌دهد که معیارهای Confidence و Lift در آن محاسبه شده است. ترتیب این قوانین بر اساس بیشترین میزان معیار Lift می‌باشد.

جدول (۵-۴) قوانین انجمنی استخراج شده از مجموعه داده، ترتیب بر اساس معیار Lift از زیاد به کم

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
61	(Project Type_)	(Project Cost_ , Building Stories_ , Construct...	0.769460	0.772675	0.765786	0.995225	1.288026	0.171244	47.610362	0.969976
48	(Project Cost_ , Building Stories_ , Construct...	(Project Type_)	0.772675	0.769460	0.765786	0.991085	1.288026	0.171244	25.858860	0.983693
54	(Project Cost_ , Project Type_)	(Building Stories_ , Construction End Use_)	0.769001	0.773823	0.765786	0.995820	1.286883	0.170716	54.104724	0.965063
55	(Building Stories_ , Construction End Use_)	(Project Cost_ , Project Type_)	0.773823	0.769001	0.765786	0.989614	1.286883	0.170716	22.241889	0.985638
29	(Project Type_)	(Building Stories_ , Construction End Use_)	0.769460	0.773823	0.766016	0.995524	1.286500	0.170590	50.527899	0.965983
24	(Building Stories_ , Construction End Use_)	(Project Type_)	0.773823	0.769460	0.766016	0.989911	1.286500	0.170590	22.850544	0.984616
53	(Project Cost_ , Construction End Use_)	(Building Stories_ , Project Type_)	0.779564	0.766246	0.765786	0.982327	1.282000	0.168449	13.226598	0.997879
56	(Building Stories_ , Project Type_)	(Project Cost_ , Construction End Use_)	0.766246	0.779564	0.765786	0.999401	1.282000	0.168449	367.797933	0.941026
30	(Project Cost_ , Construction End Use_)	(Project Type_)	0.779564	0.769460	0.768772	0.986156	1.281620	0.168928	16.652808	0.996831
35	(Project Type_)	(Project Cost_ , Construction End Use_)	0.769460	0.779564	0.768772	0.999105	1.281620	0.168928	246.227325	0.953145
31	(Project Cost_ , Project Type_)	(Construction End Use_)	0.769001	0.780941	0.768772	0.999701	1.280123	0.168227	733.627095	0.947300

در سطر اول جدول مشاهده می‌گردد، هزینه پروژه با تعداد طبقات سازه و استفاده نهایی از سازه به هم وابسته هستند و در کنار یکدیگر می‌آیند. سایر موارد و قوانین استخراج شده نیز در فایل پیوست کدها قابل مشاهده است.

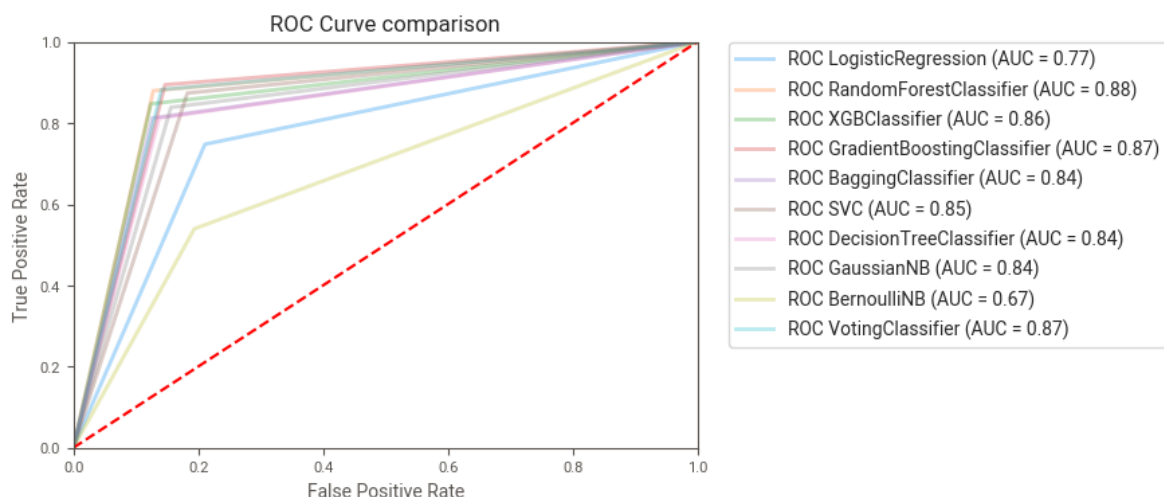
۵-۲-۳- ارزیابی و مقایسه عملکرد الگوریتم‌های یادگیری ماشین

در جدول زیر، مقایسه ارزیابی الگوریتم‌های اجرا شده روی داده‌های دیده نشده توسط مدل صورت گرفته است. همان‌طور که مشاهده می‌شود، الگوریتم جنگل تصادفی در میان تمام الگوریتم‌های پیاده‌سازی شده، بهترین عملکرد را داشته است.

جدول (۵-۵) مقایسه الگوریتم‌های یادگیری ماشین

	MLA used	Train Accuracy	Test Accuracy	Precision	Recall	AUC	F1-Score
1	RandomForestClassifier	1.0000	0.8750	0.815217	0.879765	0.875853	0.846262
3	GradientBoostingClassifier	0.9061	0.8693	0.796345	0.894428	0.873768	0.842541
9	VotingClassifier	0.9928	0.8681	0.800532	0.882698	0.870728	0.839609
2	XGBClassifier	0.9917	0.8647	0.814085	0.847507	0.861607	0.830460
4	BaggingClassifier	0.9900	0.8555	0.808023	0.826979	0.850401	0.817391
6	DecisionTreeClassifier	1.0000	0.8429	0.791429	0.812317	0.837420	0.801737
7	GaussianNB	0.8648	0.8417	0.775068	0.838710	0.841200	0.805634
5	SVC	0.8946	0.8394	0.754430	0.873900	0.845613	0.809783
0	LogisticRegression	0.7921	0.7729	0.694823	0.747801	0.768439	0.720339
8	BernoulliNB	0.7126	0.7018	0.641115	0.539589	0.672808	0.585987

همچنین نمودار زیر، منحنی ROC الگوریتم‌های به کار گرفته شده را نشان می‌دهد. مساحت زیر منحنی مربوط به مدل جنگل تصادفی بیشترین مقدار را در بین الگوریتم‌های پیاده‌سازی شده در اختیار دارد.



نمودار (۵-۱۶) منحنی ROC برای الگوریتم‌های به کار گرفته شده

۶- نتیجه‌گیری و پیشنهاد

تجزیه و تحلیل اکتشافی و استخراج دانش از مجموعه داده مورد بررسی در این پژوهش، نشان می‌دهد که ۸۵ درصد افرادی که از ناحیه سر دچار آسیب شده‌اند، فوت کرده‌اند. این موضوع، نشان از اهمیت استفاده از کلاه ایمنی مناسب در پروژه‌های ساختمانی دارد. همچنین اکثر افرادی که دچار سقوط جدی / ضربه شده‌اند (حدود ۹۰ درصد موارد)، فوت کرده‌اند. این موضوع نشان می‌دهد که ایمن‌سازی پرتگاه‌ها و قراردادن گاردهای ایمنی^{۸۹} برای ایمن‌سازی در پروژه‌های ساختمانی می‌تواند به‌عنوان یک اولویت برای جلوگیری از حوادث منجر به مرگ در حوادث ساختمانی، در دستور کار قرار گیرد.

بررسی‌های آماری نشان می‌دهد که تجهیزات جابه‌جایی مواد و وضعیت چیدمان سطح کار کارگر، مجموعاً ۲۸ درصد مرگ و میرهای حوادث مورد بررسی را شامل می‌شوند که این موضوع نشان از اهمیت ایمن‌سازی حرکت تجهیزات و جابه‌جایی مواد و همچنین توجه به چیدمان سطح کار کارگر دارد.

برای ایمن‌سازی حرکت مواد و تجهیزات می‌توان نکات زیر را مورد توجه قرار داد.

- انواع تجهیزات مورد استفاده برای جابه‌جایی مواد در سایت‌های ساخت و ساز عبارت‌اند از: بیل‌های هیدرولیک^{۹۰}، هندلرهای تلسکوپی^{۹۱}، جرثقیل‌ها^{۹۲}، لیفتراک‌ها^{۹۳}، دستگاه‌های بالابر^{۹۴} و غیره.

- از دریافت و بازرسی مواد گرفته تا ذخیره‌سازی، مونتاژ و استفاده، سیستم حمل و نقل مواد باید به‌خوبی هماهنگ و سازماندهی شده باشد تا همه افراد حاضر در محل از نحوه عملکرد آن آگاه باشند. هنگام جابه‌جایی مواد، ایمنی باید در اولویت قرار گیرد. بازرسی‌های قبل از شروع بسیار مهم است، از محدودیت‌های بار^{۹۵} نباید تجاوز کرد، دستورالعمل روش‌ها باید دنبال شود. حرکت مواد در اطراف پروژه مهم است که هنگام طراحی یک سیستم جابه‌جایی مواد برای یک پروژه ساختمانی از بهترین دستورالعمل‌ها پیروی کند.

- سیستم پیشنهادی جابه‌جایی مواد باید بر اساس نیازها، اهداف و مشخصات عملکردی تعریف شود.
- روش‌ها و فرآیندها باید استاندارد شوند تا از سردرگمی جلوگیری شود.
- دست‌زدن یا حرکت غیرضروری باید کاهش یا حذف شود.
- شرایط و روش‌های کار باید ایمنی کارگر را به‌عنوان هدف اصلی داشته باشد.
- بارهای واحد باید بهینه شود تا کار و ریسک کاهش یابد.
- مناطق ذخیره‌سازی باید مرتب و تمیز نگه داشته شوند، تاحدامکان تراکم را به حداکثر برسانند و آسیب به مواد را از بین ببرند.

- سایت‌ها باید ایمن، تمیز و به‌راحتی قابل جابه‌جایی باشند.

- تحویل مواد باید به‌سرعت دریافت و رسیدگی شود.

^{۸۹} Safety Guard

^{۹۰} Hydraulic Excavators

^{۹۱} Telescopic Handlers

^{۹۲} Cranes

^{۹۳} Forklift Trucks

^{۹۴} Lifting Devices

^{۹۵} Load Limits

- برنامه‌های مدیریت زباله سایت، باید ایجاد و حفظ شود.
 - در صورت امکان باید از فناوری‌های جابه‌جایی خودکار مواد^{۹۶} استفاده شود.
 - تجهیزات ایمنی مانند تجهیزات حفاظت فردی^{۹۷}، باید در دسترس باشند.
 - حمل دستی نادرست نیز یک عامل رایج در آسیب‌های ناشی از کار است. این موضوع به عواملی چون وزن مواد برای حمل، ماهیت تکراری حرکت، فاصله مواد در حال جابه‌جایی، محلی که مواد از آنجا یا به آنجا حمل می‌شود و وضعیت بدنی کارگر بستگی دارد.
- بررسی‌ها نشان می‌دهد که آموزش نحوه قرارگیری مناسب به کارگران در محل مناسب، شبیه‌سازی موقعیت‌ها برای کارگر جهت تصمیم‌گیری بهتر کارگر و الزام کارفرمایان و کارگران به استفاده از تجهیزات حفاظت فردی می‌تواند برای کاهش خطای انسانی در حوادث ساختمانی بسیار مؤثر واقع شود.
- باتوجه به نتایج ارائه شده، از ۱۱ الگوریتم به کار گرفته شده در این مطالعه، ۸ الگوریتم دقتی بیش از ۸۰ درصد را دادند. می‌توان نتیجه گرفت که استفاده از الگوریتم‌های یادگیری ماشین برای پیش‌بینی درجه آسیب حوادث ساختمانی، تا حد خوبی می‌تواند مؤثر باشد و استفاده از آن می‌تواند سبب بهبود عملکرد در پروژه‌های ساختمانی شود. الگوریتم‌های جنگل تصادفی، دسته‌بندی تقویتی گرادیان و دسته‌بندی مبتنی بر رأی‌گیری به عنوان سه الگوریتم برتر در این پژوهش شناخته شدند.
- در مطالعات آینده، پیشنهاد می‌شود از داده‌های حجیم‌تر برای پیاده‌سازی الگوریتم‌های یادگیری ماشین استفاده شود. همچنین به دلیل بهبود پیوسته و مداوم الگوریتم‌ها، می‌توان از الگوریتم‌های موجود دیگر مانند Adaboost برای پیش‌بینی حوادث استفاده کرد و نتایج را با نتایج این پژوهش مقایسه نمود.

^{۹۶} Automated Material Handling Technologies

^{۹۷} PPE: Personal Protective Equipment

٧- منابع

١. Ayhan, B.U. and O.B. Tokdemir, *Accident analysis for construction safety using latent class clustering and artificial neural networks*. Journal of Construction Engineering and Management, ٢٠٢٠. ١٤٦(٣): p. .٠٤٠١٩١١٤
٢. Kang, Y., et al., *Trends of fall accidents in the US construction industry*. Journal of Construction Engineering and Management, ٢٠١٧. ١٤٣(٨): p. .٠٤٠١٧٠٤٣
٣. Koc, K., Ö. Ekmekcioğlu, and A.P. Gurgun, *Integrating feature engineering, genetic algorithm and tree-based machine learning methods to predict the post-accident disability status of construction workers*. Automation in Construction, ٢٠٢١. ١٣١: p. .١٠٣٨٩٦
٤. Chiang, Y.-H., F.K.-W. Wong, and S. Liang, *Fatal construction accidents in Hong Kong*. Journal of Construction Engineering and Management, ٢٠١٨. ١٤٤(٣): p. .٠٤٠١٧١٢١
٥. Xu, N., et al., *Extracting domain knowledge elements of construction safety management: Rule-based approach using Chinese natural language processing*. Journal of Management in Engineering, ٢٠٢١. ٣٧(٢): p. .٠٤٠٢١٠٠١
٦. Lee, H.-S., et al., *Construction risk assessment using site influence factors*. Journal of computing in civil engineering, ٢٠١٢. ٢٦(٣): p. .٣٣٠-٣١٩
٧. Tixier, A.J.-P., et al., *Application of machine learning to construction injury prediction*. Automation in construction, ٢٠١٦. ٦٩: p. .١١٤-١٠٢
٨. Liu, M., et al., *Probabilistic-based cascading failure approach to assessing workplace hazards affecting human error*. Journal of Management in Engineering, ٢٠١٩. ٣٥(٣): p. .٠٤٠١٩٠٠٦
٩. Ubeynarayana, C. and Y. Goh, *An ensemble approach for classification of accident narratives*, in *Computing in Civil Engineering* ٢٠١٧. ٢٠١٧. p. .٤١٦-٤٠٩
١٠. Lukic, D., A. Littlejohn, and A. Margaryan, *A framework for learning from incidents in the workplace*. Safety Science, ٢٠١٢. ٥٠(٤): p. .٩٥٧-٩٥٠
١١. Fang, D., C. Zhao, and M. Zhang, *A cognitive model of construction workers' unsafe behaviors*. Journal of Construction Engineering and Management, ٢٠١٦. ١٤٢(٩): p. .٠٤٠١٦٠٣٩
١٢. Lette, A., et al., *A survey of work-related injuries among building construction workers in southwestern Ethiopia*. International journal of industrial ergonomics, ٢٠١٨. ٦٨: p. .٥٧٦٤
١٣. Kim, J.-M., et al., *A Deep Learning Model Development to Predict Safety Accidents for Sustainable Construction: A Case Study of Fall Accidents in South Korea*. Sustainability, ٢٠٢٢. ١٤(٣): p. .١٥٨٣
١٤. Reddy, Y.R., *Accident Prediction In Construction Using Hybrid Wavelet-Machine Learning*. International Journal of Creative Research Thoughts (IJCRT), ISSN, ٢٠٢١: p. .٢٨٨٢-٢٣٢٠
١٥. Chong, H.Y. and T.S. Low, *Accidents in Malaysian construction industry: statistical data and court cases*. International journal of occupational safety and ergonomics, ٢٠١٤. ٢٠(٣): p. .٥١٣-٥٠٣
١٦. Sanhudo, L., et al., *Activity classification using accelerometers and machine learning for complex construction worker activities*. Journal of Building Engineering, ٢٠٢١. ٣٥: p. .١٠٢٠٠١
١٧. Martínez-Rojas, M., et al., *An analysis of occupational accidents involving national and international construction workers in Spain using the association rule technique*. International journal of occupational safety and ergonomics, ٢٠٢٢. ٢٨(٣): p. .١٥٠١-١٤٩٠

۱۸. Chi, S. and S. Han, *Analyses of systems theory for construction accident prevention with specific reference to OSHA accident reports*. International journal of project management, ۲۰۱۳. ۳۱(۷): p. ۱۰۴۱-۱۰۴۷
۱۹. Gürcanli, G.E. and U. Müngen, *Analysis of construction accidents in Turkey and responsible parties*. Industrial health, ۲۰۱۳. ۵۱(۶): p. ۵۹۵-۵۸۱
۲۰. Zhu, R., et al., *Application of machine learning techniques for predicting the consequences of construction accidents in China*. Process Safety and Environmental Protection, ۲۰۲۱. ۱۴۵: p. ۳۰۲-۲۹۳
۲۱. Lee, G., et al., *Assessment of construction workers' perceived risk using physiological data from wearable sensors: A machine learning approach*. Journal of Building Engineering, ۲۰۲۱. ۴۲: p. ۱۰۲۸۲۴
۲۲. Wong, L., et al., *Association of root causes in fatal fall-from-height construction accidents in Hong Kong*. Journal of Construction Engineering and Management, ۲۰۱۶. ۱۴۲(۷): p. ۰۴۰۱۶۰۱۸
۲۳. Winge, S., E. Albrechtsen, and B.A. Mostue, *Causal factors and connections in construction accidents*. Safety science, ۲۰۱۹. ۱۱۲: p. ۱۴۱-۱۳۰
۲۴. Shuang, Q. and Z. Zhang, *Determining Critical Cause Combination of Fatality Accidents on Construction Sites with Machine Learning Techniques*. Buildings, ۲۰۲۳. ۱۳(۲): p. ۳۴۵
۲۵. Choi, J., B. Gu, and C. Sangyoon. *Development of prediction model of construction workers accident occurrence through machine learning*. in *Creative Construction Conference* ۲۰۱۸. ۲۰۱۸. Budapest University of Technology and Economics.
۲۶. Dumrak, J., et al., *Factors associated with the severity of construction accidents: The case of South Australia*. Australasian Journal of Construction Economics and Building, The, ۲۰۱۳. ۱۳(۴): p. ۴۹-۳۲
۲۷. Khosravi, Y., et al., *Factors influencing unsafe behaviors and accidents on construction sites: A review*. International journal of occupational safety and ergonomics, ۲۰۱۴. ۲۰(۱): p. ۱۲۵-۱۱۱
۲۸. Leung, M.-Y., Q. Liang, and P. Olomolaiye, *Impact of job stressors and stress on the safety behavior and accidents of construction workers*. Journal of Management in Engineering, ۲۰۱۶. ۳۲(۱): p. ۰۴۰۱۵۰۱۹
۲۹. Choi, J., et al., *Machine learning predictive model based on national data for fatal accidents of construction workers*. Automation in Construction, ۲۰۲۰. ۱۱۰: p. ۱۰۲۹۷۴
۳۰. Dong, X.S., X. Wang, and J.A. Largay, *Occupational and non-occupational factors associated with work-related injuries among construction workers in the USA*. International journal of occupational and environmental health, ۲۰۱۵. ۲۱(۲): p. ۱۵۰-۱۴۲
۳۱. Kang, K. and H. Ryu, *Predicting types of occupational accidents at construction sites in Korea using random forest model*. Safety Science, ۲۰۱۹. ۱۲۰: p. ۲۳۶-۲۲۶
۳۲. Poh, C.Q., C.U. Ubeynarayana, and Y.M. Goh, *Safety leading indicators for construction sites: A machine learning approach*. Automation in construction, ۲۰۱۸. ۹۳: p. ۳۸۶-۳۷۵
۳۳. Cheng, M.-Y., D. Kusoemo, and R.A. Gosno, *Text mining-based construction site accident classification using hybrid supervised machine learning*. Automation in Construction, ۲۰۲۰. ۱۱۸: p. ۱۰۳۲۶۵