

## پیش‌بینی ماهیت حریق مبتنی بر یادگیری ماشین: رگرسیون لجستیک یک الگوریتم تفسیرپذیر

فاطمه مسلمان زاده<sup>۱</sup>، حمیدرضا کوشا<sup>۲</sup>، کاظم صاعدی<sup>۳\*</sup>

<sup>۱</sup> کارشناس نوسازی و تحول سازمانی، سازمان آتش نشانی و خدمات ایمنی، شهرداری مشهد

<sup>۲</sup> استادیار مهندسی صنایع، دانشکده مهندسی، دانشگاه فردوسی مشهد

<sup>۳</sup> مسئول امور مطالعات و پژوهش، سازمان آتش نشانی و خدمات ایمنی، شهرداری مشهد

### مقاله پژوهشی

### چکیده

یکی از وظایف سازمان‌های آتش‌نشانی ارائه گزارش حریق و حوادث به مراجع قضائی، بیمه و سایر نهادهای درخواست‌کننده جهت تصمیم‌گیری و پرداخت خسارت است، لذا یافتن ماهیت حریق به نحوی که مؤلفه‌های غیرعملیاتی در تصمیم کارشناسان آتش‌نشانی کمترین تأثیر را داشته باشد، اهمیت این پژوهش را بیشتر خواهد نمود. با توجه به اینکه حدود ۱ درصد از گزارش‌ها حریق این سازمان ماهیت نامعلوم دارند، این موضوع باعث سردرگمی در ارائه خدمات مناسب به ارباب رجوع را داشته و تصمیم‌گیری با مشکل مواجه شده است. هدف از این پژوهش پیش‌بینی ماهیت حریق مبتنی بر الگوریتم‌های یادگیری ماشین در شهر مشهد است. در این پژوهش ابتدا مجموعه داده حریق ۷ ساله (۱۳۹۵-۱۴۰۱) مورد بررسی و واکاوی قرار گرفت و پس از آن با توجه به مسئله و ادبیات موضوع و با انجام پیش پردازش و مهندسی ویژگی مجموعه داده‌ای با تعداد ۴۶ ویژگی و ۲۸۹۳۰ نمونه تهیه شد. در مرحله بعد برای پیش‌بینی ماهیت حریق از سه الگوریتم یادگیری ماشین با ناظر استفاده شد و نتایج آن‌ها با هم مقایسه شد که الگوریتم رگرسیون لجستیک با ۷۹٫۶۶ درصد دقت با زمان اجرای ۱ ثانیه نتیجه بهتری را بین سه الگوریتم جهت پیش‌بینی ماهیت حریق ایجاد نموده است.

### تاریخ دریافت:

۱۴۰۳/۰۸/۰۸

### تاریخ پذیرش:

۱۴۰۳/۱۰/۱۵

### کلیدواژه‌ها:

ماهیت حریق، عمدی، غیرعمدی، یادگیری ماشین، رگرسیون لجستیک

### نویسنده مسئول:

saedi@mashhad.ir

doi : 10.22034/abmir.2025.22313.1068

E-ISSN: [2821-2037](https://doi.org/10.22034/abmir.2025.22313.1068) /© 2023. Published by Yazd University This is an open access article under the CC BY 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).



## ۱- مقدمه

یا غیر عمدی حریق مبتنی بر الگوریتم‌های یادگیری ماشین در شهر مشهد است.

برای این منظور ابتدا ضمن بررسی ادبیات موضوع و کارهای انجام شده و آشنایی با روش‌های علت‌یابی حریق و تکنیک‌های به کار گرفته سه الگوریتم یادگیری ماشین با ناظر جهت تحلیل داده‌ها استفاده و سپس نتایج آن‌ها را مقایسه و بهترین الگوریتم را جهت تشخیص ماهیت حریق پیشنهاد نمودیم.

## ۲- مرور ادبیات

آتش‌سوزی همواره از عواملی بوده است که خسارت‌های بسیاری به انسان وارد آورده است، از آتش‌سوزی‌های مهیب لندن در سال ۱۶۶۶، مسکو در سال ۱۷۵۶، آتش‌سوزی شهر لوس آنجلس در سال ۱۹۹۲، تا آخرین آتش‌سوزی گسترده، در سال ۲۰۰۳ که در آن، کتابخانه ملی عراق که گنجینه عظیم آثار مکتوب تمدن بین‌النهرین به شمار می‌آمد، در آتش سوخت و آسیب عمده‌ای دید و همچنین در سال ۱۳۸۴ بخش مرجع کتابخانه دانشکده حقوق دانشگاه تهران بر اثر اتصالی سیم‌های برق آتش گرفت و حدود ۷ الی ۸ هزار نسخه کتب نفیس آن سوخت [۱].

حریق عمدی یا آتش‌افروزی در سطح جهان همچنان یک معضل جدی است که منجر به هزینه‌های زیادی به لحاظ خسارت‌داری، مصدومیت و مرگ افراد می‌شود. بسیاری از محققان و متخصصین پزشکی قانونی آتش‌سوزی گمان می‌کنند که نرخ واقعی آتش‌سوزی در ایالات متحده باید بسیار بیشتر از آن باشد به طور معمول گزارش می‌شود. آن‌ها دلایل مختلفی بیان می‌کنند که چرا تعداد حریق‌های عمدی در کشورهای مختلف کمتر از حد گزارش می‌شوند [۲].

آتش‌افروزی عمدی به‌عنوان یک جرم پیچیده و منحصر به فرد، چالش‌های زیادی را در علت‌یابی حریق ایجاد می‌کند، به‌ویژه به دلیل از بین رفتن شواهد در صحنه آتش‌سوزی. شناسایی منشأ آتش‌سوزی یکی از مهم‌ترین وظایف بازرسان و کارشناسان حریق است و این امر مستلزم رویکردی سیستماتیک و جدی گرفتن هر صحنه آتش‌سوزی به‌عنوان جرم بالقوه است. انگیزه‌های متنوعی

در تمام عملیات‌های آتش‌سوزی فرماندهان عملیات یک هدف اصلی دارند و آن نجات جان و مال انسان‌هاست. اما هدف ثانویه آن‌ها این است که به بهبود کارایی و اثربخشی عملیات کمک کنند. یکی از این زمینه‌های بهبود، ثبت و گزارش کامل و دقیق داده‌های حریق است. به‌هرحال به نظر می‌رسد درحالی‌که حجم وسیعی داده توسط آتش‌نشانان و سرپرستان عملیات ثبت می‌شوند، اما هنوز مقدار قابل توجهی داده نادرست هستند یا گزارش نشده‌اند.

برای جلوگیری از ثبت داده‌های نادرست به‌غیر از دقت در ثبت اطلاعات عملیات و ثبت گزارش کارشناسی، ضرورت وجود یک سیستم در جهت بررسی و پیش‌بینی ماهیت حریق (عمدی یا غیرعمدی) وجود دارد که این موضوع تا حد زیادی زمان انتظار ارباب‌رجوع حریق را افزایش و دقت در علت‌یابی را کاهش می‌دهد.

در این پژوهش، مدل پایه، تحلیل پیشرفته داده است و به پژوهشی که در سال ۲۰۲۱ توسط ساراوانان و همکاران به بررسی، تحلیل و پیش‌بینی جرم با استفاده از داده‌کاوی و تکنیک‌های یادگیری ماشین پرداختند اقتباس شده است. ساراوانان و همکاران با استفاده از روش‌های هوش مصنوعی (بالاخص یادگیری ماشین) عملکرد مغز انسان را در هنگام رویارویی با صحنه جرم شناسایی کرده و دانش موجود در داده‌های جرائم را از طریق داده‌کاوی پیش‌بینی کردند و همچنین الگوها و طراحی مدل‌های محاسباتی مناسب را برای پیش‌بینی جرم بکار گرفتند. در این پژوهش نیز برای تشخیص اینکه حریق دارای ماهیت عمدی یا غیرعمدی است نیاز به دانشی است که علاوه بر ذهن خبرگان و کارشناسان آتش‌نشانی، در سوابق گزارش‌ها گذشته نیز موجود است.

به‌منظور پیاده‌سازی و تحلیل این دانش، نیاز به استفاده از روش‌های هوش مصنوعی است تا اولاً از داده‌های موجود به دانشی برسیم که آن تعیین ماهیت عمدی یا غیرعمدی بودن حریق است و ثانیاً از بین بردن کثرت مؤلفه‌هایی است تا بتوانیم ماهیت حریق را در آینده تشخیص دهیم. این کار تاکنون برای تشخیص ماهیت حریق صورت نپذیرفته و هدف از پژوهش حاضر پیش‌بینی ماهیت عمدی

فضایی و انواع مختلف حریق شناسایی کنند. عمر ساختمان، ارزش متوسط ملک، درصد واحدهای پر از سکنه، درصد واحدها با سکونت مالک، درصد املاک تجاری، درصد جمعیت سیاه‌پوست، درصد جمعیت بالای ۶۵ سال، نرخ فقر، ساختمان مسکونی، تعداد متوسط نفرات در هر اتاق، تراکم جمعیت یا ملک، ثبات خانواده و درصد جمعیت خارجی این متغیرها را تشکیل می‌دهند [۱۳].

در مطالعه‌ای دیگر بابوا و همکاران (۲۰۲۴) در پژوهشی به ارزیابی عملکرد مدل‌های رگرسیون لجستیک و یادگیری ماشین در پیش‌بینی استراتژی تخلیه ساختمان در زمان وقوع حریق جهت افزایش ایمنی ساکنان در مواقع بحران پرداخته است. هفت الگوریتم یادگیری ماشین شامل درختان طبقه‌بندی، نایو بیس، نزدیک‌ترین همسایه، ماشین‌های بردار پشتیبان، جنگل تصادفی، تقویت متقابل و شبکه‌های عصبی مصنوعی مورد آزمایش قرار گرفتند. داده‌ها از ۸۰۷ شرکت‌کننده جمع‌آوری شد و عملکرد مدل‌ها با معیارهایی مانند دقت و نمره F1 ارزیابی گردید.

نتایج نشان داد که رگرسیون لجستیک بالاترین ارزش مساحت زیر منحنی را دارد، درحالی‌که الگوریتم تقویت متقابل بهترین دقت و اختصاصی بودن را ارائه داد. همچنین، مدل همسایگی نزدیک بهترین فراخوانی و شبکه عصبی مصنوعی بالاترین نمره F1 را داشت. این مطالعه به شناسایی عوامل تأثیرگذار بر تصمیمات تخلیه، از جمله تأثیر حضور افراد نزدیک، پرداخته و بر لزوم تحقیقات بیشتر برای اعتبارسنجی یافته‌ها و بررسی عوامل فرهنگی تأکید دارد [۱۴].

لو و همکاران (۲۰۰۸) به بررسی واکنش‌های پیش از تخلیه افراد در شرایط آتش می‌پردازد و نشان می‌دهد که در یک مطالعه پس از حملات به مرکز تجارت جهانی، ساکنان بین ۱ تا ۳ ساعت برای خروج از ساختمان‌های ۱۱۰ طبقه زمان نیاز دارند. واکنش‌های پیش از حرکت می‌تواند بیش از دو سوم زمان کل تخلیه را شامل شود. به دلیل طبیعت تصادفی و ناواضح رفتار انسانی، روش‌های پیش‌بینی سنتی نمی‌توانند به طور مؤثر پاسخ‌های افراد را پیش‌بینی کنند. به همین دلیل، سیستم استنتاج فازی مبتنی بر شبکه‌های عصبی

برای وقوع حریق عمدی وجود دارد، از جمله کلاهبرداری در بیمه، کجکاوی و انتقام شخصی از افراد یا نهادها.

شانه‌های مشخصی مانند خسارت شدید، الگوهای سوختن غیرمعمول و آسیب به سیستم‌های ایمنی می‌تواند ظهور عمدی بودن آتش‌سوزی را نشان دهد. به‌علاوه، باتوجه‌به پیچیدگی‌های موجود، استفاده از روش‌های نوین داده‌کاوی و یادگیری ماشین می‌تواند در تحلیل و شناسایی الگوهای مرتبط با آتش‌افروزی‌های عمدی بسیار مؤثر باشد [۴،۳].

کرت و همکاران (۲۰۱۶) الگوهای مکانی ویژگی‌های کنترل حریق در جنوب فرانسه را با استفاده از عوامل محرک محیط زیستی و مردم‌شناختی مدل‌سازی کردند و نتیجه گرفتند که عوامل اقتصادی-اجتماعی تا حدی حریق را با اثرگذاری بر زمان بندی، توزیع مکانی و اندازه بالقوه آن‌ها کنترل می‌کنند [۵]، بسیاری از پژوهش‌های پیشین بر اهمیت عوامل اقتصادی-اجتماعی بر حریق ساختمان [۶]، آسیب‌پذیری خانواده‌های محروم و اثر عوامل خانوار مثل ترکیب خانواده بر ماهیت آتش‌سوزی تأکید کرده‌اند [۷].

چندین مطالعه تاکنون ارتباط بین شرایط اقتصادی-اجتماعی و وقوع حریق را نشان داده‌اند [۸،۹]. در سال‌های اخیر، هر دو روش فضایی و آماری برای تحلیل ارتباط بالقوهٔ حوادث حریق با داده‌های اقتصادی-اجتماعی و شناسایی جوامع آسیب‌پذیر استفاده شده است [۱۰،۱۱]. با استفاده از ترکیبی از تکنیک‌های آماری-فضایی و تحلیل رگرسیون، مطالعهٔ کوکوران و همکاران (۲۰۰۷) همبستگی بین حوادث حریق و محرومیت اقتصادی-اجتماعی و همچنین از تکنیک‌های نقشه برای شناسایی جوامع آسیب‌پذیر استفاده کردند [۱۲]. آن‌ها همچنین در یک مطالعه دیگر یک شاخص محرومیت مشترک برای تعیین تفاوت بین بافت‌های شهری معرفی کردند.

بسیاری از پژوهش‌های قبلی شامل تحلیل تفاوت ریسک حریق در بین مناطق جغرافیایی بودند. جنینگز (۱۹۹۹) در پژوهش خود مدل‌سازی حوادث حریق را با تأکید بر آتش‌سوزی‌های ساختمان شهری مرور کرد. جنینگز نتیجه گرفت که ۱۴ متغیر مختلف می‌توانند همبستگی‌های معنادار را بین خصوصیات اجتماعی -

می‌دهد. با این حال، کاربرد این روش محدود به هندسه‌های ساده است که با مدل‌های آتش زون قابل مدل‌سازی باشند. به‌طور کلی، این متد پیشرفت مهمی در مدل‌سازی پویای آتش محسوب می‌شود که تعادل مناسبی بین دقت و کارایی محاسباتی فراهم می‌کند [۱۷]. آتش‌سوزی‌های جنگلی، به‌ویژه آتش‌سوزی‌های انسان‌ساخت، تهدیدات جدی برای جهان به همراه دارند و خسارات مالی زیادی ایجاد می‌کنند. اولاد صیاد و همکاران (۲۰۱۹) در مقاله‌ای به بررسی استفاده از هوش مصنوعی، به خصوص داده‌های کلان، یادگیری ماشین و سنسج از دور، برای نظارت و پیش‌بینی آتش‌سوزی‌ها می‌پردازد. روش شناسی مورد استفاده شامل تحلیل داده‌های سنسج از دور از تصاویر ماهواره‌ای MODIS با تمرکز بر وضعیت محصولات (NDVI)، عوامل و شناسی (LST) و نشانه‌های آتش است.

آزمایش‌ها بر روی پلتفرم Databricks دقت پیش‌بینی ۹۸٫۳۲٪ را به همراه داشت. یافته‌ها نشان می‌دهند که آتش‌سوزی‌ها نه تنها جنگل‌ها را تخریب می‌کنند، بلکه اکوسیستم و تنوع زیستی را نیز تهدید می‌کنند. هدف این مطالعه بهبود پیش‌بینی آتش‌سوزی‌ها و کاهش خطرات مرتبط با آن بوده است [۱۸].

در مطالعه‌ای هونگ و همکاران (۲۰۲۲)، یک روش جدید ترکیبی از ابررایانه‌ها و یادگیری ماشین برای پیش‌بینی سریع رفتار دود آتش در تونل‌های معدن ارائه نمودند. با شبیه‌سازی ۱۰۰۰ مورد از آتش‌های سه بعدی تحت شرایط مختلف تهویه و حرارتی، چهار مدل آموزش دیدند. نتایج نشان داد که مدل‌های جنگل تصادفی (RFV) و ماشین بردار پشتیبان (SVM) بهترین عملکرد را در پیش‌بینی وقوع برگشت دود داشتند، در حالی که مدل شبکه‌های عصبی مصنوعی (ANN) در پیش‌بینی طول لایه گذاری دود مؤثرتر بود. همچنین، سرعت تهویه و زاویه شیب تونل به عنوان عوامل کلیدی در دینامیک آتش شناسایی شدند. این روش می‌تواند به بهبود ایمنی و مدیریت ریسک در عملیات معدنی کمک کند [۱۹].

(ANFIS) به عنوان یک رویکرد جایگزین پیشنهاد شده است. این مدل می‌تواند بر اساس داده‌های رفتاری انسان و قوانین فازی آموزش ببیند و در شبیه‌سازی رفتار انسانی در شرایط آتش کاربرد داشته باشد. مدل ANFIS به ویژه برای داده‌های نامنظم و پیچیده مناسب‌تر از مدل‌های آماری سنتی است. آن‌ها برای بهبود این مدل، جمع‌آوری داده‌های بیشتر و انتخاب مجموعه‌های فازی مناسب را پیشنهاد نموده‌اند [۱۵].

در پژوهشی دیگر ناگو و همکاران (۲۰۲۲) به ارائه یک روش چندپارامتری برای تشخیص آتش پرداخته‌اند که بر پایه استخراج عمیق ویژگی‌ها و مدل یادگیری انبوه (استکینگ) بنا شده است. این روش با استفاده از پارامترهایی چون دما، غلظت دود و CO، می‌تواند خسارات ناشی از آتش‌سوزی را بیش از ۸۰٪ کاهش دهد. با تحلیل اهمیت ویژگی‌ها از طریق الگوریتم‌های مختلف، سه نوع داده کلیدی شناسایی شده و ویژگی‌های عمیق آن‌ها با تکنیک‌های پیشرفته استخراج شده است. نتایج نشان می‌دهد که این روش در مقایسه با مدل‌های سنتی دقت و کارایی بالاتری دارد و به کاهش هشدارهای کاذب و عدم هشدار کمک می‌کند [۱۶].

در پژوهشی دیگر هودگز و همکاران (۲۰۱۹) به رویکردی مبتنی بر داده برای پیش‌بینی دما و سرعت‌های فضایی در یک محفظه آتش بر اساس مدل‌سازی آتش زون یک‌بعدی با استفاده از شبکه عصبی کانولوشن معکوس (TCNN) می‌پردازند. این رویکرد از مجموعه‌ای شامل ۱۳۳۳ شبیه‌سازی در شبیه‌ساز دینامیک آتش (FDS) استفاده کرده و به بررسی شرایط مختلفی از جمله مکان آتش، اندازه آتش، شرایط تهویه و هندسه محفظه‌ها پرداخته است. نتایج نشان می‌دهد که ۹۵٪ از پیش‌بینی‌های دما و سرعت TCNN با پیش‌بینی‌های FDS انطباق خوبی دارد، نه تنها در مورد محفظه‌های دوگانه بلکه در تست‌های پیچیده‌تر با چند محفظه نیز این موضوع صدق می‌نماید.

این رویکرد با ارائه پیش‌بینی‌های فضایی دما و سرعت بدون افزایش قابل توجه در الزامات محاسباتی، مزایای زیادی را ارائه

<sup>6</sup> Land Surface Temperature

<sup>7</sup> Random Forest

<sup>8</sup> Support Vector Machine

<sup>9</sup> Artificial Neural Network

<sup>1</sup> Adaptive Neuro-Fuzzy Inference System

<sup>2</sup> Transpose Convolutional Neural Network

<sup>3</sup> Fire Dynamics Simulator

<sup>4</sup> Moderate Resolution Imaging Spectroradiometer

<sup>5</sup> Normalized Difference Vegetation Index



فن و همکاران (۲۰۲۳) در پژوهشی به معرفی xFlashNet، یک مدل یادگیری ماشین قابل تفسیر برای پیش بینی فلش اوور در ساختارهای چند محفظه‌ای می‌پردازند. این مدل با استفاده از داده‌های دمای ثبت شده از بیش از ۱۷,۳۶۵ مورد آتش‌سوزی مصنوعی توسعه یافته و تأثیر داده‌های گمشده ناشی از حسگرهای حرارتی را در نظر می‌گیرد. این مدل در مقایسه با سه مدل پیشرفته دیگر آزمایش شده و دقت کلی آن حدود ۹۲,۹٪ است.

تحلیل با استفاده از dCAM نشان می‌دهد که کدام نواحی دما تأثیر زیادی بر تصمیمات مدل دارند. همچنین، مدل با داده‌های واقعی آتش‌سوزی آزمایش شده و هدف آن افزایش ایمنی آتش‌نشانان و کاهش آسیب‌ها و تلفات است. در نهایت، xFlashNet یک گام مهم به سوی بهبود پیش‌بینی فلش اوور در کاربردهای ایمنی آتش و فراهم آوردن تفسیرپذیری در فرایند تصمیم‌گیری است [۲۲].

در پژوهشی دیگر تیلور و همکاران (۲۰۲۴) از الگوریتم بیزی ساده جهت تحلیل اطلاعات آتش‌سوزی خانگی و آسیب‌های غیرکشنده ناشی از آن استفاده نمودند. آن‌ها با استفاده از داده‌های آتش‌نشانی شهر مرزساید کشور انگلیس از سال ۲۰۱۱ تا ۲۰۲۲، مدلی برای برآورد احتمال آسیب غیرکشنده در پاسخ به آتش (فرار، بازگشت به آتش، تلاش برای خاموش کردن آتش) تحت شرایط خاص (سن، جنس، وجود دکتور دود) توسعه دادند.

نتایج نشان می‌دهد که بیشتر آسیب‌ها در افرادی که تلاش به خاموش کردن آتش کردند، در ساختمان‌هایی با دکتور دود (۸۲٪) رخ داده است. این تحقیق به شناسایی تفاوت‌های سنی و جنسیتی و مدل‌سازی احتمال پاسخ‌های مختلف به آتش‌سوزی‌ها کمک کرده است [۲۳].

در تحقیق دیگر ارزیابی ظرفیت تیرهای فولادی با مقاطع نازک در دماهای بالا با استفاده از مدل‌های یادگیری ماشین توسط کوتو و همکاران (۲۰۲۲) انجام شد. نتایج نشان داده است که مدل‌های یادگیری ماشین مانند شبکه‌های عصبی مصنوعی (ANN) و جنگل‌های تصادفی (RF)، عملکرد بهتری نسبت به مدل‌های تحلیلی قدیمی دارند. به خصوص،  $R^2$  برای مدل ANN برابر با

در تحقیقی دیگر وانگ و همکاران (۲۰۲۱) به بررسی استفاده از رگرسیون برداری پشتیبان (SVR) برای پیش‌بینی وقوع Flash Over در آتش‌سوزی‌های دارای چند اتاق می‌پردازند. با تولید داده‌های مربوط به دما از حسگرهای حرارتی در ۱۰۰۰ شبیه‌سازی، مجموعاً ۸ میلیون داده برای توسعه مدل P-Flash گردآوری شده است. حسگرها در دمای ۱۵۰ درجه سانتی‌گراد از کار می‌افتند، که به‌واقع گرایمی مدل کمک می‌کند.

مدل P-Flash از داده‌های ثبت شده برای دما در اتاق‌های مجاور برای تخمین دمای اتاق اصلی استفاده می‌کند و از دو تکنیک جدید به نام‌های بخش‌بندی توالی و یادگیری از برازش بهره می‌برد تا مشکلات ناشی از دست‌دادن داده‌ها را حل کند. نتایج نشان می‌دهند که این مدل توانایی پیش‌بینی Flashover با دقت تقریباً ۸۳٪ و ۸۱٪ در شرایط فعلی و آینده را دارد. مدل P-Flash به آتش‌نشان‌ها اطلاعات قابل‌اعتمادی ارائه می‌دهد تا ایمنی و کارایی عملیات آتش‌نشانی را افزایش دهد [۲۰].

ارزیابی ریسک آتش‌سوزی در ساختمان‌های بلند اهمیت زیادی در مدیریت ریسک دارد و می‌تواند اطلاعات حیاتی برای آتش‌نشانی‌ها و شرکت‌های بیمه ارائه دهد. در پژوهشی دیگر گابی و همکاران (۲۰۰۹) یک سیستم امتیازدهی پیشنهاد می‌دهند که با استفاده از عوامل ریسک شناسایی شده برای ساختمان‌ها، امکان ارزیابی احتمال بروز آتش‌سوزی را فراهم می‌کند. با به‌کارگیری فرایند تحلیل سلسله‌مراتب (AHP)، وزن‌دهی به متغیرهای مؤثر در ریسک آتش‌سوزی انجام شده و سازگاری آن‌ها بررسی شده است. برای تأیید کارایی سیستم امتیازدهی، داده‌های واقعی مورد استفاده قرار گرفت و نتایج نشان‌دهنده همخوانی بین طبقه‌بندی سطح ریسک و نرخ حوادث آتش‌سوزی بودند. همچنین، مدل ماشین‌های بردار پشتیبان (SVM) به‌عنوان ابزاری برای پیش‌بینی میزان ریسک انتخاب شده که دقت سیستم را تأیید می‌کند. هدف نهایی این است که از کارت امتیاز برای برنامه‌ریزی شهری استفاده شود و مکان‌یابی ایستگاه‌های آتش‌نشانی و تعداد خودروهای آن‌ها بر مبنای سطح ریسک آتش‌سوزی به‌طور بهینه صورت گیرد [۲۱].

<sup>1</sup> Support Vector Regression

ناصر و همکاران (۲۰۲۰) به معرفی تکنیک‌های کلیدی یادگیری ماشینی (ML) و هوش مصنوعی (AI) مانند یادگیری عمیق، درخت‌های تصمیم، جنگل‌های تصادفی و ماشین‌های بردار پشتیبانی، اصول MI پرداخته و آن‌ها را تبیین می‌کنند. سپس، رویه‌های پیشنهادی برای آماده‌سازی پایگاه‌های داده و انجام تحلیل‌های مرتبط با MI در زمینه آتش، با استفاده از نمونه‌های عملی، به پژوهشگران و کارآموزان کمک می‌کند. در پایان، چالش‌ها و نگرانی‌های موجود مورد بررسی قرار می‌گیرد تا تحقیقات آینده در حوزه مهندسی و علوم آتش را تسریع کند و به حل معضلات چندبعدی و پیچیده این حوزه کمک کند [۲۴،۲۵].

۰،۹۹۹ و برای RF برابر با ۰،۹۹۸ بود، درحالی‌که مدل‌های تحلیلی نمرات کمتری کسب کردند. مدل‌ها همچنین در برابر شرایط بارگذاری و پشتیبانی مختلف عملکرد خوبی نشان دادند، هرچند نگرانی‌هایی نیز درباره احتمال بیش برآزش وجود دارد. این یافته‌ها پتانسیل بالای استفاده از ML<sup>۱</sup> در بهبود ایمنی طراحی سازه‌های فولادی در شرایط آتش‌سوزی را نمایان می‌سازد و نیاز به پژوهش‌های بیشتر در این حوزه را تأکید می‌کند، بررسی پتانسیل استفاده از هوش ماشینی (MI<sup>۲</sup>) در مطالعه رفتار آتش و ارزیابی عملکرد مواد و سازه‌ها در برابر آتش نیز موضوعی دیگر است که با وجود موفقیت MI در دیگر زمینه‌های پیچیده، کاربرد آن در مهندسی آتش محدود بوده است.

جدول (۱): پژوهش‌های مرتبط انجام شده

پارامترها و داده‌های مورد استفاده	هدف	روش تحلیل یا الگوریتم مدل‌سازی
برداشت پارامترهای موجود در صحنه عملیات	علت‌یابی حریق	مشاهدات عینی
عوامل محرک محیط زیستی و مردم‌شناختی	کنترل حریق	تحلیل مکانی
عوامل اقتصادی و اجتماعی و اقشار آسیب‌پذیر و محروم	بررسی شرایط اقتصادی و اجتماعی و تأثیر آن بر ماهیت و کنترل حریق	تحلیل مکانی و آماری
عوامل اقتصادی و اجتماعی و اقشار آسیب‌پذیر و محروم	ارتباط بروز حریق و شرایط اقتصادی و اجتماعی	تحلیل مکانی و آماری / رگرسیون
عمر ساختمان، ارزش متوسط ملک، درصد واحدهای پر از سکنه، درصد واحدها با سکونت مالک، درصد املاک تجاری و...	تحلیل تفاوت ریسک حریق در بین مناطق جغرافیایی	تحلیل ریسک و مدل‌سازی
عوامل فرهنگی ساکنین ساختمان	پیش‌بینی استراتژی تخلیه ساختمان در زمان وقوع حریق	یادگیری ماشینی / رگرسیون لجستیک
واکنش‌های انسان در زمان آتش	بررسی واکنش‌های پیش از تخلیه افراد در شرایط آتش	استنتاج فازی مبتنی بر شبکه‌های عصبی (ANFIS)
دما، غلظت دود و CO	ارائه یک روش چندپارامتری برای تشخیص وقوع حریق	استخراج عمیق ویژگی‌ها و مدل یادگیری انبوه (استکینگ)
مکان آتش، اندازه آتش، شرایط تهویه و هندسه محفظه آتش	پیش‌بینی دما و سرعت‌های فضایی در یک محفظه آتش	شبکه عصبی کانولوشن معکوس (TCNN)
تصاویر ماهواره‌ای هواشناسی، عوامل هواشناسی، نشانه‌های آتش	نظارت و پیش‌بینی آتش‌سوزی جنگل و کاهش خسارات	هوش مصنوعی / داده‌های کلان / یادگیری ماشینی

<sup>2</sup> Machine Intelligence

<sup>1</sup> Machine Learning



سرعت تهویه و زاویه شیب تونل، میزان دود	پیش‌بینی سریع رفتار دود آتش در تونل‌های معدن و افزایش ایمنی	الگوریتم‌های یادگیری ماشین RF/ANN, SVM
داده‌های دریافتی از حسگرهای حرارتی	پیش‌بینی وقوع Flash Over در آتش‌سوزی‌های دارای چند اتاق جهت بهبود ایمنی	رگرسیون برداری پشتیبان (SVR)
سطح ریسک و نرخ حوادث آتش‌سوزی	ارزیابی احتمال بروز آتش‌سوزی	یادگیری ماشین / تحلیل سلسله‌مراتبی (AHP)
داده‌های دریافتی از حسگرهای حرارتی	پیش‌بینی Flash Over در ساختارهای چند محفظه‌ای جهت افزایش ایمنی آتش‌نشانان و کاهش آسیب‌ها و تلفات	یادگیری ماشین قابل تفسیر
سن، جنس، وجود دکتور دود	تحلیل اطلاعات آتش‌سوزی خانگی و آسیب‌های غیرکشننده در پاسخ به آتش (فرار، بازگشت به آتش، تلاش برای خاموش کردن آتش)	الگوریتم بیزی ساده
عملکرد مواد و سازه‌ها در برابر آتش	بهبود ایمنی طراحی سازه‌های فولادی در شرایط آتش‌سوزی	الگوریتم‌های یادگیری ماشین RF/ANN
پایگاه‌های داده موجود در زمینه آتش	آماده‌سازی پایگاه‌های داده و انجام تحلیل‌های مرتبط با MI در زمینه آتش	یادگیری عمیق، درخت‌های تصمیم، جنگل‌های تصادفی و ماشین‌های بردار پشتیبانی

یادگیری ماشین با ناظر<sup>۱</sup> استفاده شده است. الگوریتم‌های یادگیری ماشین که جهت مدل‌سازی از آن‌ها استفاده شد به شرح جدول (۲) می‌باشند.

جدول (۲): الگوریتم‌های یادگیری ماشین

شماره	نام الگوریتم
۱	Linear Regression
۲	Logistic Regression
۳	K-NN

جدول (۳): ویژگی‌های مجموعه داده

ردیف	نام ویژگی	نوع
۱	ایستگاه	Numerical
۲	شیفت کاری	polynomial
۳	نوع عملیات	polynomial
۴	زمان اعلام به ایستگاه (ماه میلادی)	Numerical
۵	زمان اعلام به ایستگاه (ماه شمسی)	Numerical
۶	زمان اعلام به ایستگاه (فصل میلادی)	Numerical
۷	زمان اعلام به ایستگاه (فصل شمسی)	Numerical

### ۳- روش تحقیق

پژوهش فعلی به منظور پیش‌بینی و تشخیص عمدی یا غیرعمدی بودن حریق در سطح شهر مشهد و بر روی ۴۶ ویژگی و ۲۸۹۳۰ نمونه از داده‌های مربوط به گزارش‌ها حریق در بازه زمانی سال‌های ۱۳۹۵ الی ۱۴۰۱ که در سامانه ثبت گزارش‌ها حریق و حوادث سازمان آتش‌نشانی و خدمات ایمنی شهرداری مشهد [۲۶]. برداشت شده، انجام شد. در این پژوهش جهت انجام تمامی مراحل شامل جمع‌آوری داده‌ها، پیش‌پردازش داده‌ها، پاک‌سازی داده‌ها، تحلیل و مدل‌سازی داده‌ها، اعتبارسنجی و مدل نهایی از نرم‌افزار RapidMiner 10.1 و اکسل ۲۰۱۶ و سیستم مورد استفاده با مشخصات ذیل جهت مدل‌سازی و پردازش به کارگیری شد یک دستگاه PC با حافظه 8G پردازنده: Cori3 3.9 GHZ از آنجاکه برای پیش‌بینی عمدی یا غیرعمدی بودن حریق کلیه داده‌ها پس از مرحله پیش‌پردازش، دارای برجسب عمدی یا غیرعمدی هستند جهت مدل‌سازی در این پژوهش از الگوریتم‌های

<sup>1</sup> Supervised Learning



Polynomial	زیر تصرف	۳۸
Polynomial	سری عملیات	۳۹
Polynomial	دسته عملیات	۴۰
Polynomial	ماده سوختنی	۴۱
Polynomial	منشأ حرارت	۴۲
Polynomial	اولین ماده سوختنی درگیر در حریق	۴۳
Polynomial	کد علت عملیات	۴۴
Numerical	فاصله جغرافیایی Geo	۴۵
Bionominal	ماهیت (این ویژگی به عنوان برجسب جهت پیش بینی استفاده می‌شود)	۴۶

در الگوریتم‌های جدول (۲)، هدف ایجاد مدلی است که بتواند با بیشترین دقت و اطمینان ضمن بررسی وابستگی بین ویژگی‌ها ماهیت حریق را پیش بینی نماید. از آنجاکه الگوریتم‌های یادگیری ماشین با نظارت طیف وسیعی دارند و با توجه به مجموعه داده موجود استفاده از الگوریتم‌های مبتنی بر رگرسیون می‌تواند کمک زیادی به حل مسئله مربوط به این پژوهش نماید. در این پژوهش، هدف بررسی رابطه متغیرهای موجود است. در این جا نیز از تحلیل رگرسیون باهدف پیش بینی تغییرات یک یا چند متغیر وابسته (ویژگی‌های موجود در جدول ۳) با توجه به تغییرات متغیرهای مستقل (ماهیت حریق) و الگوریتم K-NN به دلیل آنکه یک الگوریتم چندمنظوره است و هم برای مسائل رگرسیون کاربرد دارد و هم مسائل طبقه‌بندی و دارای پیاده‌سازی آسان و نتایج ساده‌ای است، استفاده می‌نماییم.

باتوجه به موارد ذکر شده در خصوص انتخاب الگوریتم‌های پیش‌بینی، حال مجموعه داده انتخاب شده جهت این پژوهش که دارای ویژگی‌های به شرح جدول (۳) است را مورد بررسی و تحلیل قرار خواهیم داد.

رویکرد مهندسی ویژگی برای ایجاد ویژگی‌های بهتر در مجموعه داده و به جهت افزایش عملکرد مدل‌های خروجی است، به همین دلیل برای ویژگی‌های T103 در مجموعه داده که از نوع تاریخ و زمان هستند ویژگی‌هایی جدید شامل سال، فصل، ماه، روز در سال، روز در ماه، روز در هفته، هفته در سال، ساعت و همچنین اختلاف بین هرکدام از زمان‌های رسیدن به محل حریق با سایر زمان‌ها نیز به‌عنوان ویژگی‌های جدید به مجموعه داده اضافه شد. علت اضافه شدن ویژگی‌های فوق به این دلیل است که تفاوت معناداری بین

Numerical	زمان اعلام به ایستگاه (هفته میلادی)	۸
Numerical	زمان اعلام به ایستگاه (هفته شمسی)	۹
Numerical	زمان اعلام به ایستگاه (روز در ماه میلادی)	۱۰
Numerical	زمان اعلام به ایستگاه (روز در ماه شمسی)	۱۱
Numerical	زمان اعلام به ایستگاه (روز در هفته شمسی)	۱۲
Numerical	زمان اعلام به ایستگاه (ساعت)	۱۳
Numerical	زمان بازگشت به ایستگاه (فصل میلادی)	۱۴
Numerical	زمان بازگشت به ایستگاه (ساعت)	۱۵
Numerical	مدت‌زمان رسیدن به محل حریق (T103 - T104)	۱۶
Numerical	مدت‌زمان حرکت از ایستگاه تا شروع عملیات (T103 - T105)	۱۷
Numerical	مدت‌زمان بین حرکت از ایستگاه تا برگشت به ایستگاه (T103 - T107)	۱۸
Numerical	مدت‌زمان اطفاء حریق (T104 - T105)	۱۹
Numerical	مدت‌زمان بین حرکت از ایستگاه تا جمع‌آوری تجهیزات (T103 - T106)	۲۰
Numerical	مدت‌زمان بین رسیدن به محل حریق تا جمع‌آوری تجهیزات (T104 - T106)	۲۱
Numerical	مدت‌زمان کاوش (T105 - T106)	۲۲
Numerical	مدت‌زمان بین رسیدن به محل حریق تا رسیدن به ایستگاه (T104 - T107)	۲۳
Numerical	مدت‌زمان بین اطفاء حریق تا رسیدن به ایستگاه (T105 - T107)	۲۴
Numerical	مدت‌زمان رسیدن به ایستگاه (T106 - T107)	۲۵
Numerical	فاصله محل حریق از ایستگاه به متر	۲۶
Numerical	فاصله از محل حریق بر اساس موقعیت موجود و نقشه نرم‌افزار نشان	۲۷
Numerical	مدت‌زمان رسیدن به محل حریق بر اساس موقعیت موجود و نقشه نرم‌افزار نشان	۲۸
Numerical	دمای هوا	۲۹
Bionominal	وضعیت باد	۳۰
Numerical	سرعت وزش باد	۳۱
Polynomial	جهت وزش باد	۳۲
Numerical	میزان رطوبت هوا	۳۳
Polynomial	آب‌وهوا	۳۴
Numerical	فشار هوا	۳۵
Numerical	منطقه شهرداری	۳۶
Polynomial	تصرف	۳۷



برای الگوریتم‌های جدول (۲) از اعتبارسنجی متقاطع با تعداد فولد ۱۰ استفاده شد. در این پژوهش برای ارزیابی مدل ایجاد شده در هر الگوریتم، از اعتبارسنجی متقاطع استفاده شده است، علت این انتخاب ایجاد مدلی است که امکان تعمیم آن به گزارش‌ها حریق خارج از مجموعه داده را داشته باشد.

در ارزیابی‌های متداول الگوریتم‌های یادگیری ماشین، مجموعه داده با نسبت ۸۰ درصد آموزش و ۲۰ درصد آزمایش تقسیم می‌شوند. یادگیری پارامترهای یک تابع پیش بینی و آزمایش آن بر روی همان داده‌ها یک اشتباه روش‌شناختی است به این دلیل که مدل، فقط برچسب (ماهیت حریق) نمونه‌هایی را که به‌تازگی مشاهده کرده است را تکرار می‌کند. در این روش ممکن است که معیار صحت مناسبی نیز در ارزیابی ارائه شود، اما نمی‌تواند با اضافه شدن نمونه‌های جدید به مجموعه داده‌ها، پیش بینی درستی ارائه نماید که به این مشکل بیش برآزش داده می‌گویند.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FN} + \text{FP} + \text{TN}) \quad (1)$$

پارامتر صحت، متداول‌ترین، اساسی‌ترین و ساده‌ترین معیار اندازه‌گیری کیفیت یک دسته‌بند است و عبارت است از میزان تشخیص صحیح دسته‌بند در مجموع دودسته. این پارامتر در واقع نشانگر میزان الگوهای است که درست تشخیص داده شده‌اند و به‌صورت رابطه (۱) فرموله و تعریف می‌شود.

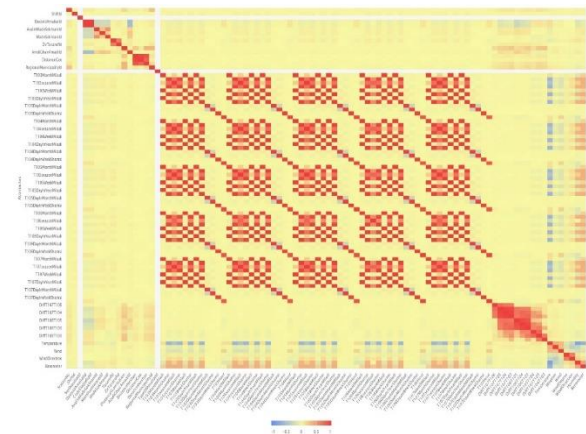
$$\text{Sensitivity (TPR)} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

البته، پارامتر صحت معمولاً به‌صورت درصد بیان می‌شود. اما پارامترهای دیگری نیز علاوه بر معیار صحت وجود دارند که می‌توان به‌سادگی از این ماتریس استخراج کرد. یکی از متداول‌ترین آن‌ها، معیار حساسیت است که آن را «نرخ پاسخ‌های مثبت درست» نیز می‌گویند. حساسیت به معنی نسبتی از موارد مثبت است که آن‌ها را به‌درستی به‌عنوان نمونه مثبت تشخیص داده است. این پارامتر به‌صورت رابطه (۲) محاسبه می‌شود.

در واقع، «حساسیت» همان معیار بحث شده در مورد مثال بالا است. معیاری که مشخص می‌کند دسته‌بند، به چه اندازه در تشخیص تمام حریق‌های عمدی موفق بوده است. همان‌گونه که از رابطه فوق مشخص است، تعداد حریق‌های غیرعمدی که توسط دسته‌بند به‌اشتباه به‌عنوان عمدی تشخیص داده می‌شوند، هیچ تأثیری در

روزهای هفته، فصول، ماه‌ها و ساعت رخداد حریق، تعطیل یا غیرتعطیل بودن روز رخداد حادثه، در خصوص عمدی یا غیرعمدی بودن آن می‌تواند ایجاد نماید.

در شکل (۱) که ماتریس همبستگی بین داده‌های موجود در این پژوهش است، همبستگی مثبت بین دما و رطوبت نشان‌دهنده این است که با افزایش دما، رطوبت نیز افزایش می‌یابد، همچنین در رابطه رطوبت و فشار بارومتریک، همبستگی منفی بین این دو به این معنی است که افزایش رطوبت معمولاً با کاهش فشار بارومتر همراه است، در رابطه‌ای دیگر همبستگی مثبت بین سرعت باد و دما نشان‌دهنده این است که در شرایط خاص، وزش باد می‌تواند دما را افزایش دهد. بین زمان اعلام عملیات (ماه شمسی) با فشار اتمسفر و رطوبت هوا همبستگی مثبت و بین همین ویژگی با دما همبستگی منفی وجود دارد؛ یعنی بالارفتن فشار اتمسفر در ماه‌های سرد سال اتفاق می‌افتد. در ویژگی دیگر، همبستگی مثبت بین زمان رسیدن (DiffT104T103) و فاصله محل حریق تا ایستگاه موردنظر وجود دارد.



شکل (۱): ماتریس همبستگی داده‌ها

پس از بررسی بر روی مجموعه داده، مشخص شد برخی از ویژگی‌ها مفهوم مناسب در مدل‌ها ایجاد نخواهند کرد به‌عنوان مثال زمان رخداد حریق که برای رفع این مشکل از تکنیک گسسته‌سازی جهت تبدیل زمان رخداد که به‌صورت پیوسته (ساعت، دقیقه، ثانیه) بود به بازه‌های گسسته نیمه‌شب: (از ساعت ۱۲ شب تا ۶ صبح)، صبح: (از ساعت صبح ۶ تا ۱۲ ظهر)، عصر: (از ساعت ۱۲ تا ۱۸ عصر) و شب: (از ساعت ۱۸ تا ۲۴) استفاده شد.

توضیح که پارامتر ارزش اخباری مثبت را اصطلاحاً دقت، و حساسیت را اصطلاحاً یادآوری (صحت) می‌نامند، «معیار F» به صورت رابطه (۶) قابل محاسبه است.

### ۱-۳ یافته‌ها

در این مرحله پس از تنظیم پارامترهای مؤثر هر یک از الگوریتم‌های جدول (۲)، نتایج اجرای الگوریتم‌ها بر روی ۶۶ ویژگی موجود در داده‌های جدول (۳) قابل مشاهده است.

نتیجه اجرای الگوریتم Linear Regression

جدول (۴): پارامترهای تنظیم شده برای الگوریتم Linear Regression

Regression	
اعتبارسنجی الگوریتم	
اعتبارسنجی	مقاطع
نوع نمونه برداری	خودکار
تعداد Fold	۱۰
پارامترهای الگوریتم	
Feature Selection	M5 Prime
حذف ویژگی‌های دارای هم خطی	فعال
Min Tolerance	فعال
Use Bias	فعال

این پارامتر، روش انتخاب ویژگی مورد استفاده در طول رگرسیون خطی را نشان می‌دهد. گزینه‌های زیر برای روش انتخاب شامل: M5 prime، حریص، T-Test، T-Test، تکراری، پارامتر Min Tolerance زمانی امکان فعال شدن را دارد که گزینه حذف ویژگی‌های دارای هم خطی فعال باشد. پارامتر Use Bias نشان می‌دهد که آیا مقدار عرض از مبدأ باید محاسبه شود یا خیر.

جدول (۵): ماتریس درهم‌ریختگی Linear Regression

		عمدی	غیرعمدی
ماهیت حریق	عمدی	TP = ۱۱۳۰۲	FN = ۳۹۲۶
	غیرعمدی	FP = ۲۰۵۸	TN = ۱۱۶۴۴

در جدول (۵) نتایج پیش‌بینی‌های الگوریتم Linear Regression نشان داده شده است. در این جدول از تعداد ۲۸۹۳۰ نمونه، تعداد ۱۱۶۴۴ نمونه عملیات‌های غیرعمدی بوده‌اند که الگوریتم فوق آن را به درستی غیرعمدی (منفی صحیح) تشخیص داده است. همچنین تعداد ۳۹۲۶ مورد از نمونه‌ها که ماهیت عمدی داشته‌اند را در کلاس غیرعمدی (منفی غلط) دسته‌بندی نموده است. در این الگوریتم

محاسبه این پارامتر ندارد و در واقع زمانی که از این پارامتر به عنوان پارامتر ارزیابی برای دسته‌بند استفاده می‌شود، هدفش دستیابی به نهایت صحت در تشخیص نمونه‌های کلاس مثبت است.

$$\text{Specificity (TNR)} = \text{TN} / (\text{TN} + \text{FP}) \quad (۳)$$

در نقطه مقابل این پارامتر، ممکن است در مواقعی صحت تشخیص کلاس منفی حائز اهمیت باشد. از متداول‌ترین پارامترها که معمولاً در کنار حساسیت بررسی می‌شود، پارامتر خاصیت، است که به آن «نرخ پاسخ‌های منفی درست» نیز می‌گویند. خاصیت به معنی نسبتی از موارد منفی است که آن‌ها را به درستی به عنوان نمونه منفی تشخیص داده است. این پارامتر به صورت رابطه (۳) محاسبه می‌شود.

این دو پارامتر (حساسیت و خاصیت) نیز مشابه معیار صحت، معمولاً به صورت درصد بیان می‌شوند. واضح است که پیش‌بینی عالی، پیش‌بینی است که مقادیر حساسیت و خاصیت مربوط به آن، هر دو صددرصد باشند؛ اما احتمال وقوع این اتفاق در واقعیت بسیار کم است و همیشه یک حداقل خطایی وجود دارد.

دو پارامتر مقدار پیش‌بینی شده مثبت و مقدار پیش‌بینی شده منفی هستند که برای بیان «نسبت پاسخ‌های درست در هر دسته» استفاده می‌شوند.

$$\text{PPV} = \text{TP} / (\text{TP} + \text{FP}) \quad (۴)$$

$$\text{NPV} = \text{TN} / (\text{TN} + \text{FN}) \quad (۵)$$

ارزش اخباری مثبت، بیان‌کننده این است که چند درصد از الگوهایی که دسته‌بند آن‌ها را مثبت تشخیص داده، در واقعیت هم مثبت هستند و به همین ترتیب، ارزش اخباری منفی نشان می‌دهد که چند درصد از نمونه‌هایی که عضو دسته منفی تشخیص داده شده‌اند، در واقعیت هم عضو همین دسته هستند. این دو پارامتر نیز به سادگی از روی ماتریس درهم‌ریختگی به صورت رابطه (۴) و (۵) قابل محاسبه می‌باشند.

$$\text{F-measure} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \quad (۶)$$

پارامتر مهم دیگری به نام «معیار F» وجود دارد که برای ارزیابی عملکرد دسته‌بندها بسیار مورد استفاده قرار می‌گیرد و از ترکیب دو پارامتر حساسیت و ارزش اخباری مثبت حاصل می‌شود. با این

در شناسایی حریق‌های عمدی است و این نمودار داده‌های موجود در ماتریس درهم‌ریختگی شکل شماره (۵) را تأیید می‌نماید.

نتیجه اجرای الگوریتم Logistic Regression

جدول (۶): پارامترهای تنظیم شده برای الگوریتم Logistic Regression

اعتبارسنجی الگوریتم	
اعتبارسنجی	مقاطع
نوع نمونه‌برداری	خودکار
تعداد Fold	۱۰
پارامترهای الگوریتم	
Add intercept	فعال
Computer p-values	فعال
مدیریت مقادیر از دست‌رفته	MeanImputation
استانداردسازی	فعال

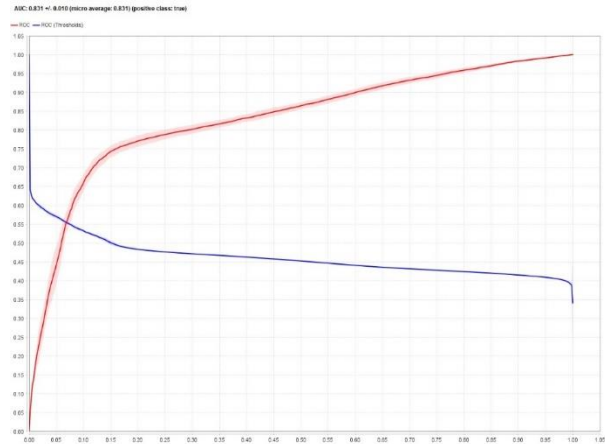
جدول (۷): ماتریس درهم‌ریختگی Logistic Regression

ماهیت حریق	عمدی		غیرعمدی	
	عمدی	غیرعمدی	TP = ۱۱۵۸۲	FN = ۳۶۴۶
	غیرعمدی	FP = ۲۲۳۸	TN = ۱۱۴۶۴	

در جدول (۷) نتایج پیش‌بینی‌های الگوریتم Logistic Regression نشان داده شده است. در این جدول از تعداد ۲۸۹۳۰ نمونه تعداد ۱۱۵۸۲ نمونه عملیات‌های غیرعمدی بوده‌اند که الگوریتم فوق آن را به درستی غیرعمدی (منفی صحیح) تشخیص داده است، همچنین تعداد ۳۶۴۶ مورد از نمونه‌ها که ماهیت عمدی داشته‌اند را در کلاس غیرعمدی (منفی غلط) دسته‌بندی نموده است. در این الگوریتم تعداد ۲۲۳۸ مورد حریق غیرعمدی در کلاس عمدی دسته‌بندی شده است (مثبت غلط) و در نهایت تعداد ۱۱۵۸۲ نمونه به درستی عمدی (مثبت صحیح) دسته‌بندی نموده است.

نتایج تولیدشده از خروجی‌های مدل منتج از الگوریتم لجستیک رگرسیون که در شکل (۳) نمایش داده شده است مشابهت زیادی با نتایج حاصل از الگوریتم رگرسیون خطی در شکل (۱) دارد نتایج نمودار نشان‌دهنده عملکرد یک مدل طبقه‌بندی با AUC برابر با

تعداد ۲۰۵۸ مورد حریق غیرعمدی در کلاس عمدی دسته‌بندی شده است (مثبت غلط) و در نهایت تعداد ۱۱۳۰۲ نمونه به درستی عمدی (مثبت صحیح) دسته‌بندی شده است.



شکل (۲): خروجی AUC الگوریتم رگرسیون خطی

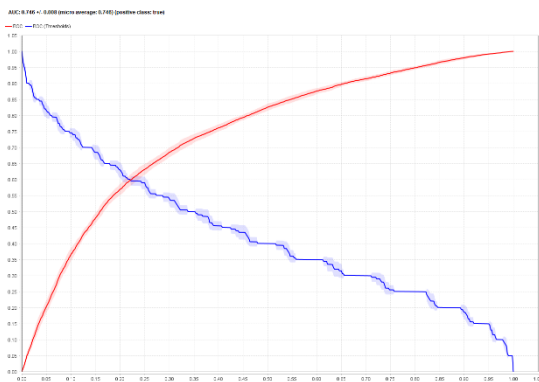
همان‌طور که در شکل (۲) مشاهده می‌شود مدل تولیدشده بر مبنای الگوریتم رگرسیون خطی و نتایج حاصل از آن دارای عملکرد خوبی است، مساحت زیر منحنی قرمز رنگ به عنوان یک معیار برای ارزیابی عملکرد دسته‌بند مورد استفاده قرار می‌گیرد، در حالت ایده‌آل، مساحت زیر منحنی برابر با بیشترین مقدار خود، یعنی یک است. بنابراین، هر چه مساحت زیر نمودار به عدد یک نزدیک‌تر باشد، به معنای بهتر بودن عملکرد دسته‌بند است. که با امتیاز AUC برابر با ۰.۸۳۱ نشان‌دهنده توانایی مناسب آن در شناسایی نمونه‌های مثبت و منفی (عمدی و غیرعمدی) است. انحراف معیار  $+/- 0.10$  نشان‌دهنده پایداری و عدم نوسان زیاد در عملکرد مدل است. همچنین، میانگین میکرو نیز ۰.۸۳۱ بوده که عملکرد یک‌نواخت مدل را در تمامی کلاس‌ها تأیید می‌کند. نمودار ROC<sup>۲</sup> نشان‌دهنده توانایی مدل در تفکیک حریق‌های عمدی و غیرعمدی است.

بالین‌حال، محور عمودی یا نرخ مثبت درست (TPR) و محور افقی یا نرخ مثبت غلط (FPR) نیز توازن در شناسایی نمونه‌ها را به‌خوبی نمایان می‌سازد. به‌طورکلی، نتایج حاکی از دقت بالای مدل

<sup>2</sup> Receiver Operating Characteristic

<sup>1</sup> Area Under Curve

در جدول (۹) نتایج پیش‌بینی‌های الگوریتم k-NN نشان داده شده است. در این جدول از تعداد ۲۸۹۳۰ نمونه تعداد ۸۷۴۷ نمونه عملیات‌های غیرعمدی بوده‌اند که الگوریتم فوق آن را به درستی غیرعمدی (منفی صحیح) تشخیص داده است، همچنین تعداد ۴۵۸۰ مورد از نمونه‌ها که ماهیت عمدی داشتند را در کلاس غیرعمدی (منفی غلط) دسته‌بندی نموده است. در این الگوریتم تعداد ۴۹۵۵ مورد حریق غیرعمدی در کلاس عمدی دسته‌بندی شده است (مثبت غلط) و در نهایت تعداد ۱۰۶۴۸ نمونه به درستی عمدی (مثبت صحیح) دسته‌بندی نموده است.

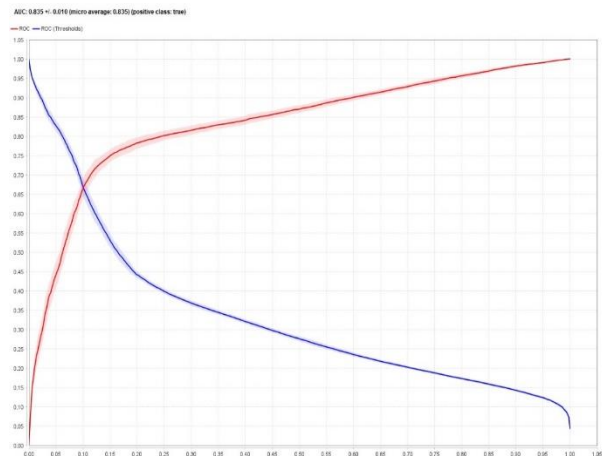


شکل (۴): خروجی AUC الگوریتم KNN

نتایج تولیدشده از خروجی‌های مدل منتج از الگوریتم K-NN که در شکل (۴) نمایش داده شده است با دو الگوریتم دیگر تفاوت آشکاری دارد، نتایج نمودار نشان‌دهنده عملکرد یک مدل طبقه‌بندی با AUC برابر با ۰,۷۴۶ است که نشان‌دهنده عملکرد ضعف تر الگوریتم در شناسایی نمونه‌های مثبت و منفی (عمدی و غیرعمدی) است. نکته دیگر نوسان و عدم پایداری در عملکرد مدل است.

در جدول (۱۰) که نتیجه پیش‌بینی و دسته‌بندی هر الگوریتم است، علیرغم اینکه بهترین عملکرد در تشخیص ماهیت درست حریق‌های غیرعمدی توسط الگوریتم رگرسیون خطی ارائه شده است اما در تشخیص ماهیت عمدی درست الگوریتم لجستیک رگرسیون عملکرد بهتری را ارائه داده است و الگوریتم K-NN در همه گزینه‌ها تشخیص ضعیف‌تری نسبت به دو الگوریتم دیگر ارائه کرده است.

۰,۸۳۵ است که عملکرد یکنواخت مدل را در تمامی کلاس‌ها حدود ۰,۰۴ نسبت به الگوریتم رگرسیون خطی بهبود داده است و انحراف معیار ۰,۰۱۰ است، نمودار ROC حساسیت و اختصاصی بودن مدل را در آستانه‌های مختلف بررسی کرده و کارایی مطلوب آن را در تفکیک بین کلاس‌ها ثابت می‌کند. و در نهایت، این نتایج نشان‌دهنده قدرت و پایداری مدل در شناسایی نمونه‌های مثبت هستند.



شکل (۳): خروجی AUC الگوریتم لجستیک رگرسیون

نتیجه اجرای الگوریتم k-NN

جدول (۸). پارامترهای تنظیم شده برای الگوریتم k-NN

اعتبارسنجی	
اعتبارسنجی	مقاطع
نوع نمونه‌برداری	خودکار
تعداد Fold	۱۰
الگوریتم k-NN	
K	۲۰
رأی وزن‌دار	فعال
نوع اندازه‌گیری	Mixed Measures
اندازه‌گیری	فاصله اقلیدسی

جدول (۹). ماتریس درهم‌ریختگی k-NN

		عمدی	غیرعمدی
ماهیت حریق	عمدی	TP = ۱۰۶۴۸	FN = ۴۵۸۰
	غیرعمدی	FP = ۴۹۵۵	TN = ۸۷۴۷

جدول (۱۰): نتیجه پیش بینی و دسته‌بندی الگوریتم‌ها

TN+TP	غیرعمدی (FN) نادرست	عمدی (FP) نادرست	غیرعمدی (TN) صحیح	عمدی (TP) صحیح	نام الگوریتم
۲۲۹۴۶	۳۹۲۶	۲۰۵۸	۱۱۶۴۴	۱۱۳۰۲	Linear Regression
۲۳۰۴۶	۳۶۴۶	۲۲۳۸	۱۱۴۶۴	۱۱۵۸۲	Logistic Regression
۱۹۳۹۵	۴۵۸۰	۴۹۵۵	۸۷۴۷	۱۰۶۴۸	K-NN

جدول (۱۱): دامنه تغییرات معیارهای ارزیابی الگوریتم‌ها

Specificity	F_measure	Recall	Precision	Accuracy	درصد دامنه تغییرات نام الگوریتم
+/- ۰,۹۳	+/- ۰,۶۶	+/- ۱,۰۲	+/- ۰,۷۸	+/- ۰,۶۰	Linear Regression
+/- ۰,۸۹	+/- ۱,۱	+/- ۱,۲۶	+/- ۰,۸۸	+/- ۰,۹۶	Logistic Regression
+/- ۰,۹۷	+/- ۰,۸۶	+/- ۱,۳۵	+/- ۰,۶۴	+/- ۰,۷۷	K-NN

در جدول (۱۱) مشاهده می‌شود که بیشترین میزان دامنه تغییرات در معیار دقت، صحت و معیار F را الگوریتم Logistic Regression جهت تشخیص عمدی یا غیرعمدی بودن دارد. از آنجا که از تعداد ۲۸۹۳۰ مورد نمونه موجود در مجموعه داده، ۴۸,۳۶ درصد یعنی تعداد ۱۳۷۰۲ مورد مربوط به کلاس عمدی و

۵۲,۶۴ درصد یعنی تعداد ۱۵۲۲۸ مورد مربوط به کلاس غیرعمدی است نتیجه خواهیم گرفت که توزیع داده‌ها در این دو کلاس تقریباً نرمال است و بهترین معیار جهت ارزیابی مدل، معیار Accuracy است.

جدول (۱۲): درصد کیفیت معیارهای ارزیابی الگوریتم‌ها

AUC	Specificity	F_measure	Recall	Precision	Accuracy	نام الگوریتم
۰,۸۳۱	۸۴,۹۸	۷۹,۰۷	۷۴,۲۲	۸۴,۶	۷۹,۳۲	Linear Regression
۰,۸۳۵	۸۳,۶۷	۷۹,۷۴	۷۶,۰۶	۸۳,۸۱	۷۹,۶۶	Logistic Regression
۰,۷۱۳	۶۳,۸۴	۶۹,۰۷	۶۹,۹۲	۶۸,۲۴	۶۷,۰۴	K-NN

برای انتخاب بهترین مدل از بین مدل‌های موجود، تنها به معیار صحت هر الگوریتم نمی‌توان بسنده کرد، معیارهای مهم دیگری نیز در انتخاب مدل می‌توانند تأثیرگذار باشند. به‌عنوان مثال زمان اجرای مدل یکی از معیارهای مهم برای انتخاب مدل نهایی است که خود تابعی از منابع سخت‌افزاری و نرم‌افزاری (حافظه، پردازنده و...) است. معیار انتخاب بعدی تفسیرپذیری مدل است، یعنی بتوان از

روی مدل‌های واقعی، مدل‌های تولیدشده توسط الگوریتم را مورد تست و بازبینی قرارداد. در جدول (13) برای مقایسه بهتر مدل‌ها به‌غیر از معیار Accuracy، معیارهای تفسیرپذیری و زمان اجرای مدل نیز به جهت اهمیت، در انتخاب بهترین مدل در نظر گرفته شد. الگوریتم Linear Regression و K-NN هر دو تفسیرپذیر

موردبررسی قرار گرفت که نتایج، بهترین مدل برای پاسخ به پرسش این پژوهش را مدل‌های تولیدشده از الگوریتم Logistic Regression نشان می‌دهد.

خطای پیش‌بینی‌های انجام شده برای ماهیت‌های عمدی حریق با مجموعه داده انتخابی در این الگوریتم برابر با ۱۶,۲ درصد و برای ماهیت حریق غیرعمدی برابر ۲۴,۲ درصد است. با توجه به طیف گسترده الگوریتم‌های یادگیری ماشین این امکان وجود دارد تا در پژوهش‌های آتی، داده‌های این پژوهش را توسط سایر الگوریتم‌ها مانند الگوریتم‌های مرتبط با شبکه‌های عصبی، درخت تصمیم، الگوریتم‌های تقویتی موردبررسی و تحلیل قرار داد و این نتایج را مطابق با معیارهای معرفی شده مقایسه و بهترین الگوریتم را برای ایجاد مدل انتخاب نمود.

## ۵- قدردانی

با تشکر از سازمان آتش‌نشانی و خدمات ایمنی شهرداری مشهد برای همکاری خوب و صمیمانه

## References

- [1] M. Niknam, "Tehran: Organization of Records and National Library of the Islamic Republic of Iran, Research and Education Management Publications; 2006," *Combating fire*, 2006.
- [2] <http://www.firedata.org>.
- [3] Gorbett, G.E., Meacham, B.J., Wood, C.B. et al. "Use of damage in fire investigation: a review of fire patterns analysis, research and future direction". *Fire Sci Rev* 4, 4 (2015). <https://doi.org/10.1186/s40038-015-0008-4>
- [4] NFPA (2014) NFPA 921-"Guide for Fire and Explosion Investigations". National Fire Protection Association, Quincy (USA)
- [5] C. R. Jennings, "Socioeconomic characteristics and their relationship to fire incidence: a review of the literature," *Fire Technol.*, vol. 35, no. 1, pp. 7-34, 1999, doi: 10.1023/A:1015330931387.
- [6] T. Curt, T. Fréjaville, and S. Lahaye, "Modelling the spatial patterns of ignition causes and fire regime features in southern France: Implications for fire prevention policy," *Int. J. Wildl. Fire*, vol. 25, no. 7, pp. 785-796, 2016, doi: 10.1071/WF15205.
- [7] Y. Zhang, "Federal Emergency Management Agency (FEMA)," *Encycl. Glob. Heal.*, 2012, doi: 10.4135/9781412963855.n451.

می‌باشند اما از لحاظ زمان اجرا این دو الگوریتم دارای اختلاف زمانی زیادی می‌باشند.

مدل‌های تولیدشده از روی الگوریتم‌های غیر تفسیرپذیر مانند یک جعبه سیاه هستند که محتوای داخل جعبه غیرقابل مشاهده است و صرفاً ورودی‌ها و خروجی‌ها دیده خواهند شد. از آنجاکه اهمیت مشاهده محتوای این جعبه‌ها جهت رصد و بررسی صحت خروجی‌های مدل در داده‌های سازمان آتش‌نشانی حائز اهمیت است، در نتیجه بهترین و کاراترین مدل که سه معیار سرعت بالای پیش‌بینی، دقت بالا در پیش‌بینی عمدی یا غیرعمدی بودن حریق و تفسیرپذیری مدل را با هم دارد، الگوریتم Logistic Regression است و می‌توان گفت که خروجی‌های این مدل بهترین خروجی را بین این سه الگوریتم ایجاد نموده است و ضعیف‌ترین نتایج را الگوریتم K-NN در خصوص تشخیص ماهیت حریق داشته است.

جدول (۱۳): انتخاب مدل براساس معیارهای سنجش

نام الگوریتم	تفسیرپذیری	زمان اجرا (به ثانیه)	AUC
Linear Regression	×	۰	۰,۸۳۱
Logistic Regression	✓	۱	۰,۸۳۵
K-NN	✓	۳۴	۰,۷۱۳

## ۴- نتیجه‌گیری

از آنجاکه توجه به ایمنی شهر مقدس مشهد به‌عنوان دومین کلان‌شهر کشور و پایتخت مذهبی ایران از اهمیت بسزایی برخوردار است و آمار آتش‌سوزی در این شهر هر ساله روبه افزایش است، لذا باتوجه به حجم انبوه داده، با استفاده از الگوریتم‌های یادگیری ماشین نسبت به پیش‌بینی ماهیت حریق از منظر عمدی یا غیرعمدی بودن آن پرداختیم و نتایج آن به شرح ذیل است:

خروجی‌های منتج از مدل‌های تولیدشده در این پژوهش که شامل الگوریتم‌های پرکاربرد یادگیری ماشین با ناظر یعنی رگرسیون خطی، لجستیک و K-NN بود توسط نرم‌افزار ریپید ماینر تحلیل شد که این خروجی‌ها دارای دقت و صحت مطلوبی جهت پیش‌بینی ماهیت حریق بودند. جهت شناسایی بهترین الگوریتم ابتدا معیارهای دقت و صحت، زمان اجرا و تفسیرپذیری مدل‌ها





- Fire Saf. J.*, vol. 108, p. 102854, 2019, doi: [10.1016/j.firesaf.2019.102854](https://doi.org/10.1016/j.firesaf.2019.102854).
- [18] Y. O. Sayad, H. Mousannif, and H. Al Moatassime, "Predictive modeling of wildfires: A new dataset and machine learning approach," *Fire Saf. J.*, vol. 104, pp. 130–146, 2019, doi: [10.1016/j.firesaf.2019.01.006](https://doi.org/10.1016/j.firesaf.2019.01.006).
- [19] Y. Hong, J. Kang, and C. Fu, "Rapid prediction of mine tunnel fire smoke movement with machine learning and supercomputing techniques," *Fire Saf. J.*, vol. 127, 2022, doi: [10.1016/j.firesaf.2021.103492](https://doi.org/10.1016/j.firesaf.2021.103492).
- [20] J. Wang *et al.*, "P-Flash – A machine learning-based model for flashover prediction using recovered temperature data," *Fire Saf. J.*, vol. 122, no. June 2020, p. 103341, 2021, doi: [10.1016/j.firesaf.2021.103341](https://doi.org/10.1016/j.firesaf.2021.103341).
- [21] J. Gubbi, S. Marusic, and M. Palaniswami, "Smoke detection in video using wavelets and support vector machines," *Fire Saf. J.*, vol. 44, no. 8, pp. 1110–1115, 2009, doi: [10.1016/j.firesaf.2009.08.003](https://doi.org/10.1016/j.firesaf.2009.08.003).
- [22] L. Fan, W. C. Tam, Q. Tong, E. Y. Fu, and T. Liang, "An explainable machine learning based flashover prediction model using dimension-wise class activation map," *Fire Saf. J.*, vol. 140, 2023, doi: [10.1016/j.firesaf.2023.103849](https://doi.org/10.1016/j.firesaf.2023.103849).
- [23] "A Bayesian analysis of domestic fire response and fire injury".
- [24] C. Couto, Q. Tong, and T. Gernay, "Predicting the capacity of thin-walled beams at elevated temperature with machine learning," *Fire Saf. J.*, vol. 130, no. February, p. 103596, 2022, doi: [10.1016/j.firesaf.2022.103596](https://doi.org/10.1016/j.firesaf.2022.103596).
- [25] M. Z. Naser, "Mechanistically Informed Machine Learning and Artificial Intelligence in Fire Engineering and Sciences," *Fire Technol.*, vol. 57, no. 6, pp. 2741–2784, 2021, doi: [10.1007/s10694-020-01069-8](https://doi.org/10.1007/s10694-020-01069-8).
- [26] Incidents statistics reports for the I. M. Fire & Safety service organization, "<http://www.mahar.mashhad.ir>".
- [8] P. Chhetri, J. Corcoran, R. J. Stimson, and R. Inbakaran, "Modelling potential socio-economic determinants of building fires in South East Queensland," *Geogr. Res.*, vol. 48, no. 1, pp. 75–85, 2010, doi: [10.1111/j.1745-5871.2009.00587.x](https://doi.org/10.1111/j.1745-5871.2009.00587.x).
- [9] S. E. Chandler, A. Chapman, and S. J. Hollington, "Fire Incidence, Housing and Social Conditions - the Urban Situation in Britain.," *Fire Prev.*, no. 172, pp. 15–20, 1984.
- [10] M. Duncanson, A. Woodward, and P. Reid, "Socioeconomic deprivation and fatal unintentional domestic fire incidents in New Zealand 1993-1998," *Fire Saf. J.*, vol. 37, no. 2, pp. 165–179, 2002, doi: [10.1016/S0379-7112\(01\)00033-9](https://doi.org/10.1016/S0379-7112(01)00033-9).
- [11] J. Corcoran, G. Higgs, C. Brunson, and A. Ware, "The use of comaps to explore the spatial and temporal dynamics of fire incidents: A case study in South Wales, United Kingdom," *Prof. Geogr.*, vol. 59, no. 4, pp. 521–536, 2007, doi: [10.1111/j.1467-9272.2007.00639.x](https://doi.org/10.1111/j.1467-9272.2007.00639.x).
- [12] J. Corcoran, G. Higgs, and A. Higginson, "Fire incidence in metropolitan areas: A comparative study of Brisbane (Australia) and Cardiff (United Kingdom)," *Appl. Geogr.*, vol. 31, no. 1, pp. 65–75, 2011, doi: [10.1016/j.apgeog.2010.02.003](https://doi.org/10.1016/j.apgeog.2010.02.003).
- [13] J. Corcoran, G. Higgs, C. Brunson, A. Ware, and P. Norman, "The use of spatial analytical techniques to explore patterns of fire incidence: A South Wales case study," *Comput. Environ. Urban Syst.*, vol. 31, no. 6, pp. 623–647, 2007, doi: [10.1016/j.compenvurbysys.2007.01.002](https://doi.org/10.1016/j.compenvurbysys.2007.01.002).
- [14] A. Balboa, A. Cuesta, J. González-Villa, G. Ortiz, and D. Alvear, "Logistic regression vs machine learning to predict evacuation decisions in fire alarm situations," *Saf. Sci.*, vol. 174, no. April 2023, 2024, doi: [10.1016/j.ssci.2024.106485](https://doi.org/10.1016/j.ssci.2024.106485).
- [15] S. M. Lo, M. Liu, P. H. Zhang, and R. K. K. Yuen, "An artificial neural-network based predictive model for pre-evacuation human response in domestic building fire," *Fire Technol.*, vol. 45, no. 4, pp. 431–449, 2009, doi: [10.1007/s10694-008-0064-6](https://doi.org/10.1007/s10694-008-0064-6).
- [16] N. Qu, Z. Li, X. Li, S. Zhang, and T. Zheng, "Multi-parameter fire detection method based on feature depth extraction and stacking ensemble learning model," *Fire Saf. J.*, vol. 128, 2022, doi: [10.1016/j.firesaf.2022.103541](https://doi.org/10.1016/j.firesaf.2022.103541).
- [17] J. L. Hodges, B. Y. Lattimer, and K. D. Luxbacher, "Compartment fire predictions using transpose convolutional neural networks,"



## Predicting the nature of fire based on machine learning: Logistic regression is an interpretable algorithm

Fatemeh Mosalmanzadeh<sup>1</sup>, Hamidreza Koosha<sup>2</sup>, Kazem Saedi<sup>3\*</sup>

<sup>1</sup>Organizational modernization and transformation expert, Fire&Safety Services Department, Mashhad municipality

<sup>2</sup>Assistant Professor of Industrial Engineering, Faculty of Engineering, Ferdowsi University of Mashhad

<sup>3</sup>Responsible for studies and research, fire department and safety services organization, Mashhad municipality

### Article Information

### Abstract

#### Original Research Paper

#### Received:

2024 October 29

#### Accepted:

2025 January 4

#### Keywords:

Nature of fire, intentional, unintentional, machine learning, logistic regression

#### Corresponding Author\*:

saedi@mashhad.ir

One of the duties of firefighting organizations is to submit fire and accident reports to judicial authorities, insurance, and other requesting institutions to make decisions and pay damages, therefore, finding the nature of the fire in such a way that non-operational components have the least impact on the decision of firefighting experts increases the importance of this research. Considering that about 1% of the fire reports of this organization are unknown, this issue has confused providing proper services to Arbab-Rojoo and decision-making has been difficult. This research aims to predict the nature of fire based on machine learning algorithms in the city of Mashhad. In this research, the 7-year fire data set (1395-1401) was first examined and analyzed, and then, according to the problem and the literature, a data set with 46 features and 28930 samples was prepared by pre-processing and feature engineering. In the next step, to predict the nature of the fire, two machine learning algorithms were used with the observer and their results were compared. The logistic regression algorithm, with 79.66% accuracy and an execution time of 1 second, created a better result between the two algorithms in predicting the nature of the fire.



: 10.22034/abmir.2025.22313.1068

E-ISSN: [2821-2037](https://doi.org/10.22034/abmir.2025.22313.1068) /© 2023. Published by Yazd University This is an open access article under the CC BY 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).

