

کاربرد شاخص تراکم ترافیک و سرعت وزش باد در مدلسازی داده‌محور غلظت آلاینده های هوای شهری

رضا گلشن خواص^{۱*}، عبدالرضا شیخ‌الاسلامی^۲، مریم نیازی راد^۳

۱- عضو هیأت علمی دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران (rgolshan@iust.ac.ir)

۲- عضو هیأت علمی دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران

۳- فارغ التحصیل کارشناسی ارشد حمل‌ونقل، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران

چکیده:

آلودگی هوا یکی از چالش‌های مهم محیط‌زیستی و بهداشتی در شهرهاست و ترافیک شهری و شرایط هواشناسی می‌توانند بر غلظت آلاینده‌ها اثرگذار باشند. هدف این پژوهش، مدل‌سازی کوتاه‌مدت غلظت مونوکسید کربن (CO)، ذرات معلق با قطر کمتر از 10 میکرومتر (PM_{10}) و دی‌اکسید گوگرد (SO_2) با استفاده از شاخص تراکم ترافیک و سرعت باد است. برای این منظور، وضعیت ترافیک معابر در شعاع 100 متری ایستگاه سنجش آلودگی هوای میدان فتح تهران از لایه ترافیک گوگل میپس استخراج و با اختصاص ضرایب 1 تا 4 به سطوح مختلف تراکم، شاخص تراکم ترافیک محاسبه شد. داده‌ها برای 30 روز به صورت ساعتی جمع‌آوری شدند و 720 مشاهده در تحلیل به کار رفت. برای بررسی اثرات کوتاه‌مدت و تجمعی، پنج حالت شامل داده‌های ساعتی، میانگین و مجموع دوساعته و میانگین و مجموع سه‌ساعته بررسی شد. سه روش شبکه عصبی مصنوعی، واحد بازگشتی دروازه‌دار (GRU) و تقویت گرادیان حدی ($XGBoost$) برای پیش‌بینی آلاینده‌ها مقایسه شدند و عملکرد آنها با معیارهای ضریب تعیین، میانگین قدر مطلق خطا و ریشه میانگین مربعات خطا ارزیابی شد. نتایج نشان داد مدل $XGBoost$ در حالت داده‌های ساعتی بهترین عملکرد را دارد و مقادیر ضریب تعیین برای CO ، PM_{10} و SO_2 به ترتیب 0.65 ، 0.64 و 0.51 است. همچنین، تجمیع زمانی داده‌ها بهبود پایداری در دقت پیش‌بینی ایجاد نکرد. نتایج بیانگر ظرفیت داده‌های ترافیکی برخط و سرعت باد برای پیش‌بینی کوتاه‌مدت آلودگی هوا در مقیاس محلی است.

واژگان کلیدی:

یادگیری ماشین، تراکم ترافیک، $XGBoost$ ، سرعت باد، غلظت آلاینده های هوا

۱- مقدمه و تاریخچه تحقیقات

آلودگی هوا محصول جانبی توسعه شهرنشینی و فعالیت های انسانی است. منظور از فعالیت انسانی بخش زائد فعالیت هایی نظیر تولید کالاهای حمل و نقل، تولید انرژی، گرمایش، تولید نور و... می باشد. علاوه بر این، از آتش سوزی جنگل ها و فعالیت آتشفشان ها به عنوان منابع طبیعی آلودگی هوا یاد می شود. یک آلاینده هوا، گاز، قطرات مایع و ذرات جامد است که اگر با غلظت کافی در هوا پراکنده شوند، برای گیاهان، جانوران و انسان ها مضر خواهند بود.

آلودگی هوا یکی از چالش های جدی زیست محیطی است که تأثیرات منفی قابل توجهی بر سلامت انسان و اکوسیستم دارد. با افزایش جمعیت و توسعه شهری، ترافیک به عنوان یکی از عوامل اصلی آلودگی هوا شناخته می شود.

حمل و نقل همانند دیگر نیازهای بشر جزء لاینفک زندگی امروزی است که باعث مصرف انواع سوخت های فسیلی شده و دارای پیامدهایی چون تغییر در تعادل طبیعی محیط زیست، کاهش منابع سوخت، تولید انواع پسماندها و مواد زاید می شود [۱].

تراکم ترافیک منجر به افزایش انتشار آلاینده های هوا می شود، زیرا وسایل نقلیه در شرایط تراکم با سرعت کمتری حرکت کرده و توقف مکرر دارند. این وضعیت غلظت آلاینده ها را در هوا به ویژه در مناطق شهری افزایش می دهد، و خطرات سلامتی مانند بیماری های تنفسی و قلبی عروقی را تشدید می کند. علاوه بر این، هزینه های اقتصادی ناشی از اثرات بهداشت عمومی آلودگی هوا می تواند به میلیاردها دلار در سال برسد [۲]. در این راستا، بررسی و مدلسازی آلودگی هوا بر اساس شاخص های ترافیکی و شرایط جوی، به ویژه سرعت وزش باد، اهمیت ویژه ای دارد. تحقیقات نشان می دهد که ترافیک و شرایط جوی به طور مستقیم بر غلظت آلاینده ها تأثیر می گذارند. به عنوان مثال، افزایش تراکم ترافیک منجر به افزایش انتشار گازهای آلاینده مانند مونوکسید کربن، اکسیدهای نیتروژن و ذرات معلق می شود [۳]. از سوی دیگر، سرعت وزش باد می تواند به عنوان یک عامل مؤثر در پراکندگی و رقیق شدن این آلاینده ها عمل کند [۴].

در مواجهه با مشکلات جدی روزافزون آلودگی زیست محیطی، محققان حجم قابل توجهی از تحقیقات مرتبط را انجام داده اند و در این مطالعات، پیش بینی آلودگی هوا از اهمیت بالایی برخوردار بوده است. به عنوان یک اقدام احتیاطی، پیش بینی آلودگی هوا مبنایی برای انجام اقدامات مؤثر کنترل آلودگی است و پیش بینی دقیق آلودگی هوا به یک وظیفه مهم تبدیل شده است.

یکی از روش های علمی در بررسی این موضوع، مدلسازی است. مطالعات قبلی نشان داده اند که تحلیل داده محور کیفیت هوا با استفاده از تکنیک های یادگیری ماشین می تواند پیش بینی دقیق تری نسبت به مدل های سنتی ارائه دهد [۵]. همچنین، عوامل ترافیکی مانند تراکم و الگوهای حرکتی خودروها به طور معنی داری با غلظت آلاینده ها مرتبط هستند [۶]. مطالعات دیگری نیز به مدلسازی تعامل میان آلودگی هوا و ترافیک شهری پرداخته اند که نتایج آن ها با یافته های این تحقیق هم راستا بوده است [۷].

مدل های یادگیری ماشین به عنوان ابزارهای پیشرفته برای تحلیل داده ها و پیش بینی رفتار سیستم ها، به طور فزاینده ای در مدلسازی آلودگی هوا مورد استفاده قرار می گیرند. این مدل ها با توانایی پردازش حجم بالای داده ها و شناسایی الگوهای پیچیده، می توانند به بهبود دقت پیش بینی ها کمک کنند [۴].

در مطالعه ای شایکا و همکاران^۱ روش جدیدی را برای پیش بینی یک ساعته، ۸ ساعته و ۲۴ ساعته غلظت $PM_{2.5}$ برای ۳ سال (۲۰۰۸-۲۰۱۰) در شهر دهلی نو هند ارائه کردند. این مطالعه از ترکیب پارامترهای ورودی مختلف داده های هواشناسی، تعداد وسیله نقلیه و انتشار استفاده کرد. بعد از ارزیابی، مشاهده شد که روش $GRU-ED^2$ که یک مدل RNN^3 اصلاح شده است از روش های پیش بینی سنتی (نظیر رگرسیون، درخت تصمیم، رگرسیون لجستیک و ...) از نظر دقت بهتر عمل می کند [۸].

^۱ Shakya et al

^۲ Gated Recurrent Unit Based Encoder-Decoder

^۳ Recurrent Neural Network

فریمن و همکاران^۴ در مقاله‌ای به بررسی میانگین غلظت ۸ ساعته ازن سطحی پرداختند. در پژوهش ایشان، میانگین غلظت ۸ ساعته ازن سطحی (۳O) با استفاده از یادگیری عمیق متشکل از یک شبکه عصبی بازگشتی (RNN) با حافظه کوتاه مدت بلند مدت (LSTM^۵) پیش‌بینی شد. کیفیت هوا و داده‌های هواشناسی ساعتی برای آموزش و پیش‌بینی مقادیر تا ۷۲ ساعت با نرخ خطای کم استفاده شد. بعد از پیش‌پردازش داده‌ها، نتایج نشان داد یادگیری عمیق برای پیش‌بینی کیفیت هوا مناسب است [۹].

مطالعه‌ای تحت عنوان "پیش‌بینی هوشمند کیفیت هوا و پیش‌بینی آلودگی با استفاده از یادگیری ماشین" به بررسی مدل‌های پیش‌بینی آلودگی هوا پرداختند. این مدل‌ها شامل انواع مدل‌های رگرسیون، هوش مصنوعی و یادگیری ماشین می‌باشد. این مدل‌ها را برای پیش‌بینی غلظت آلاینده $PM_{2.5}$ با هم مقایسه کرده‌اند. از بین مدل‌های موجود، مدل‌های یادگیری ماشین (XGBoost^۶) نتایج قابل قبول‌تری را ارائه داده‌اند [۱۰]. در مطالعه دیگری به بررسی یک چارچوب تحلیلی و پیش‌بینی، برای پیش‌بینی سطوح غلظت $PM_{2.5}$ برای ۱۶ منطقه شهری در شانگهای ایجاد کردند که پیش‌بینی در این مطالعه هم توسط XGBoost انجام شد. نتایج نشان داد که این الگوریتم‌ها قادر به پیش‌بینی غلظت آلاینده است. [۱۱].

در سال ۲۰۲۳ در ماکائو چین، به بررسی مدل‌های پیش‌بینی آلودگی هوا پرداخته شد. در این مطالعه مدل‌های شبکه عصبی، تقویت گرادیان شدید (XGBoost) و مدل‌های رگرسیون مورد بررسی قرار گرفت. برای این کار از داده‌های کیفیت هوای ماکائو از سال ۲۰۱۶ تا ۲۰۲۱ استفاده شد. پیش‌بینی بصورت ساعتی برای آلاینده‌های PM_{10} ، $PM_{2.5}$ و CO انجام شد. نتایج نشان داد که این الگوریتم‌ها قادر به پیش‌بینی غلظت آلاینده‌ها در ماکائو هستند [۱۲].

مطالعاتی به بررسی کاربرد شبکه‌های عصبی مصنوعی (ANN) در پیش‌بینی کیفیت هوا پرداخته‌اند. در این مطالعات یک مدل ANN مبتنی بر رگرسیون توسعه داده شد که داده‌های هواشناسی را در پیش‌بینی آلاینده‌هایی چون CO ، NO ، NO_2 ، SO_2 ، PM_{10} و $PM_{2.5}$ لحاظ کردند. ارزیابی مدل نشان داد که ANN ها دقت بیشتری نسبت به مدل‌های خطی سنتی دارند و می‌توانند به بهبود مدیریت زیست‌محیطی و بهداشت عمومی کمک کنند [۱۳] و [۱۴].

نوآوری و سهم اصلی این پژوهش، استفاده از شاخص تراکم ترافیک استخراج شده از داده‌های برخط گوگل میپس در محدوده ۱۰۰ متری ایستگاه پایش و ترکیب آن با سرعت باد برای پیش‌بینی کوتاه‌مدت سه آلاینده CO ، PM_{10} و SO_2 است. همچنین، پنج نحوه بازنمایی زمانی داده‌های ترافیک و باد با سه مدل یادگیری ماشین مقایسه شده‌اند تا مشخص شود آیا میانگین‌گیری یا انباشت کوتاه‌مدت اطلاعات ورودی نسبت به داده‌های ساعتی، دقت پیش‌بینی را بهبود می‌دهد یا خیر.

۲- روش تحقیق

هدف این پژوهش مدلسازی آلودگی هوا با استفاده از مدل‌های یادگیری ماشین است که به بررسی تأثیر شاخص تراکم ترافیک و سرعت وزش باد بر غلظت آلاینده‌های هوا می‌پردازد. با تحلیل داده‌های ترافیکی و جوی، تلاش می‌شود تا الگوهای ارتباطی میان این عوامل شناسایی و پیش‌بینی‌های دقیقی از کیفیت هوا ارائه شود.

۲-۱- استخراج و نحوه محاسبه شاخص تراکم ترافیک

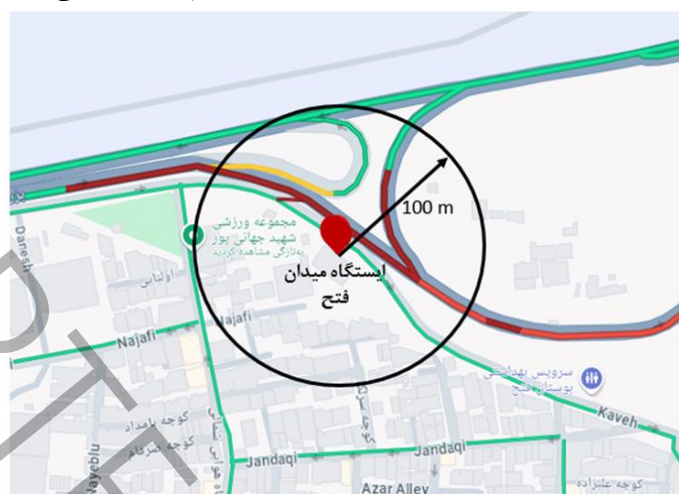
برای محاسبه شاخص تراکم ترافیک از داده‌های آنلاین، به‌ویژه Google Maps، استفاده می‌شود که اطلاعات لحظه‌ای از وضعیت ترافیک را ارائه می‌دهند. شعاع ۱۰۰ متری از ایستگاه سنجش آلاینده‌های هوا واقع در میدان فتح در جنوب غربی شهر تهران به عنوان منطقه مورد

^۴ Freeman et al

^۵ Long short-term memory

^۶ Extreme Gradient Boosting

بررسی انتخاب شده است تا تأثیرات ترافیکی در نزدیکی این ایستگاه به طور دقیق تری تحلیل شود. شکل ۱ محدوده اطراف ایستگاه میدان فتح را نشان می‌دهد. داده‌های رنگی ترافیک از Google Maps برای تحلیل وضعیت تراکم در این نواحی استفاده می‌شود.



شکل ۱: شعاع ۱۰۰ متری ایستگاه سنجش آلاینده‌های هوا در میدان فتح

Figure 1. A 100-m radius around the Fath Square air-pollutant monitoring station

برای محاسبه شاخص تراکم ترافیک رنگ‌های معابر در لایه ترافیک پلتفرم Google Maps به عنوان نمایانگر وضعیت ترافیک به کار رفتند:

سبز: ۱ (جریان روان)

نارنجی: ۲ (تراکم متوسط)

قرمز: ۳ (تراکم شدید)

قرمز تیره: ۴ (تراکم بسیار شدید)

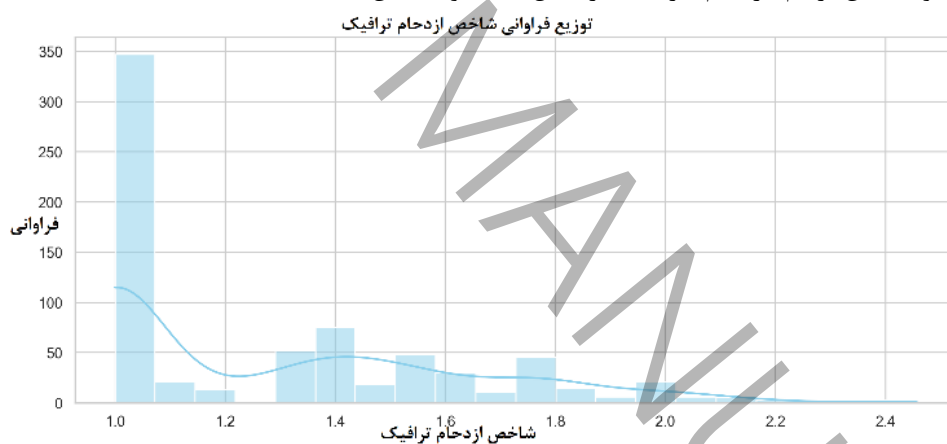
در پلتفرم Google Maps پس از کدگذاری مسیرها، طول هر رنگ با استفاده از ابزار Measure Distance محاسبه شد. سپس شاخص تراکم ترافیک از تقسیم مجموع حاصل ضرب کد رنگ هر قطعه در طول متناظر آن قطعه بر طول کل معابر محدوده به دست آمد. این روش، شاخص عددی برای تراکم هر معبر را در ناحیه مورد مطالعه نمایان می‌سازد. مطالعات پیشین نشان می‌دهد که استفاده از داده‌های Google Maps از دقت قابل قبولی برخوردار بوده و میزان دقت آن حدود ۹۰ درصد برآورد می‌شود [۱۵]. داده‌ها برای ماه آذر ۱۴۰۱ و به صورت ساعتی به مدت ۳۰ روز جمع‌آوری شدند. در این مطالعه از ۷۲۰ داده ساعتی به عنوان حجم نمونه استفاده شد. در سناریوهای مبتنی بر تجمیع زمانی، صرفاً متغیرهای هواشناسی در بازه‌های مورد نظر تجمیع گردیدند، در حالی که داده‌های آلاینده‌ها بدون تغییر و در همان تفکیک زمانی اولیه (ساعتی) حفظ شدند. بنابراین، تعداد مشاهدات در تمامی تحلیل‌ها ثابت باقی ماند. جدول ۱ شاخص تراکم محاسبه شده برای ایستگاه میدان فتح در تاریخ یکم آذر ۱۴۰۱ از ساعت ۹ صبح تا ۵ عصر را نمایش می‌دهد.

جدول ۱: بخشی از شاخص تراکم محاسبه شده

Table 1. A sample of the calculated traffic congestion index

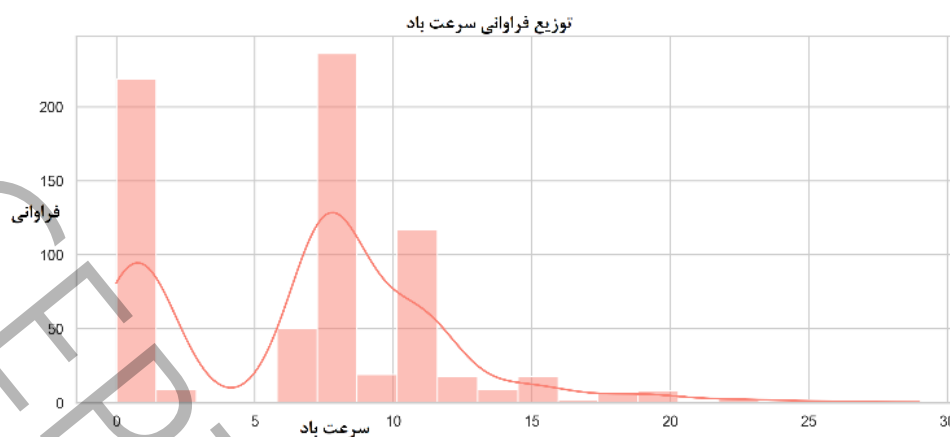
ایستگاه میدان فتح	زمان (ساعت)	تاریخ
۷۵۳۶۲۳/۱	۰۹:۰۰	۱۴۰۱/۰۹/۰۱
۵۹۴۲۰۳/۱	۱۰:۰۰	۱۴۰۱/۰۹/۰۱
۷۸۲۶۰۹/۱	۱۱:۰۰	۱۴۰۱/۰۹/۰۱
۳۹۸۸۴۱/۱	۱۲:۰۰	۱۴۰۱/۰۹/۰۱
۳۹۱۳۰۴/۱	۱۳:۰۰	۱۴۰۱/۰۹/۰۱
۱	۱۴:۰۰	۱۴۰۱/۰۹/۰۱
۲۱۷۳۹۱/۱	۱۵:۰۰	۱۴۰۱/۰۹/۰۱
۱	۱۶:۰۰	۱۴۰۱/۰۹/۰۱
۴۴۹۲۷۵/۱	۱۷:۰۰	۱۴۰۱/۰۹/۰۱

توزیع فراوانی سرعت باد و شاخص تراکم (ازدحام) ترافیک در شکل‌های ۲ و ۳ نشان داده شده است:



شکل ۲: توزیع فراوانی شاخص تراکم ترافیک

Figure 2. Frequency distribution of the traffic congestion index

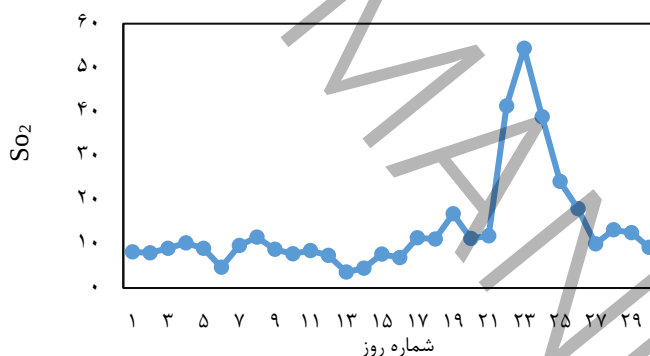


شکل ۳: توزیع فراوانی سرعت باد

Figure 3. Frequency distribution of wind speed

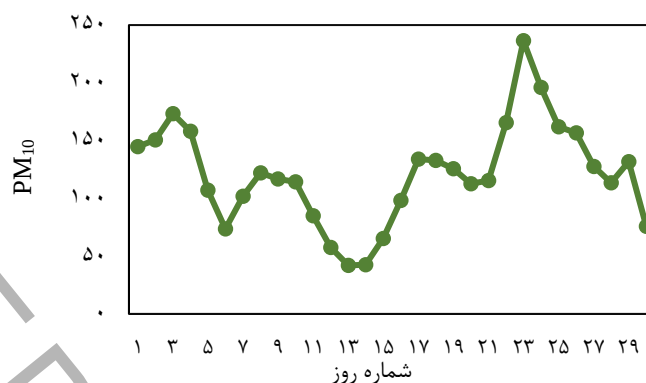
۲-۲- آماده سازی داده ها

برای مدلسازی آلودگی هوا سه آلاینده کلیدی شامل CO و SO_2 به عنوان نماینده های مهم کیفیت هوا در نظر گرفته شدند. به منظور تحلیل دقیق تر، داده های این سه آلاینده برای ایستگاه «میدان فتح» جمع آوری شد. این ایستگاه به دلیل تراکم ترافیک بالای خود در طی ۳۰ روز آذر ۱۴۰۱ به صورت ساعتی مورد بررسی قرار گرفت. توزیع میانگین روزانه این سه آلاینده در این ایستگاه در شکل های ۴ تا ۶ نشان داده شده است.



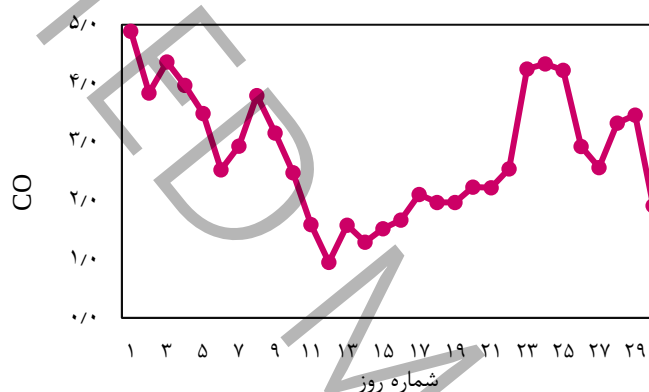
شکل ۴: توزیع میانگین روزانه آلاینده SO_2 در ایستگاه میدان فتح

Figure 4. Distribution of daily mean SO_2 concentration at Fath Square station



شکل ۵: توزیع میانگین روزانه آلاینده PM₁₀ در ایستگاه میدان فتح

Figure 5. Distribution of daily mean PM₁₀ concentration at Fath Square station



شکل ۶: توزیع میانگین روزانه آلاینده CO در ایستگاه میدان فتح

Figure 6. Distribution of daily mean CO concentration at Fath Square station

در مرحله آماده سازی داده ها برای مدلسازی آلودگی هوا با استفاده از شاخص تراکم ترافیک، داده ها شامل متغیرهای سرعت باد، آلاینده ها و شاخص تراکم به دقت سازماندهی شدند. برای بررسی تأثیر متغیرهای مختلف بر آلودگی هوا، پنج حالت مختلف در نظر گرفته شد:

- (۱) حالت کلی: در این حالت داده ها بصورت یک ساعته مورد بررسی قرار گرفت.
- (۲) میانگین دو ساعته: محاسبه میانگین مقادیر دو ساعته از شاخص تراکم و سرعت باد برای تحلیل نوسانات کوتاه مدت.
- (۳) مجموع دو ساعته: محاسبه مجموع مقادیر دو ساعته فقط برای شاخص تراکم به منظور بررسی تأثیر تجمعی ترافیک.
- (۴) میانگین سه ساعته: محاسبه میانگین مقادیر سه ساعته از سرعت باد و شاخص تراکم برای تحلیل تغییرات در یک بازه طولانی تر.
- (۵) مجموع سه ساعته: محاسبه مجموع سه ساعته برای شاخص تراکم به منظور درک تأثیرات تجمعی بر آلودگی.

۳-۲- اعتبارسنجی

برای ارزیابی عملکرد مدل های پیش بینی، معیارهای مختلفی وجود دارد. در این پژوهش، از سه معیار مهم $RMSE$ ، R^2 و MAE استفاده شد. هر یک از این معیارها به طور خاص به ارزیابی دقت پیش بینی های مدل کمک می کند.

RMSE: یکی از معیارهای محبوب برای ارزیابی دقت پیش‌بینی است. این معیار نشان‌دهنده میانگین خطای مربعی پیش‌بینی‌ها نسبت به مقادیر واقعی است و به صورت زیر محاسبه می‌شود:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

n : تعداد نمونه‌ها

y_i : مقدار واقعی

$\hat{y}_i$: مقدار پیش‌بینی شده

R^2 : معیاری است که نشان‌دهنده میزان واریانس توضیح داده‌شده توسط مدل در مقایسه با واریانس کل داده‌هاست. فرمول محاسبه R^2 به صورت زیر است:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

$\bar{y}$: میانگین داده‌های واقعی

y_i : مقدار واقعی

$\hat{y}_i$: مقدار پیش‌بینی شده

MAE: یک معیار دیگر برای ارزیابی دقت پیش‌بینی است که میانگین قدر مطلق خطاها را محاسبه می‌کند. فرمول MAE به صورت زیر است:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

n : تعداد نمونه‌ها

y_i : مقدار واقعی

$\hat{y}_i$: مقدار پیش‌بینی شده

۳ - نتایج و بحث

در ادامه، به تحلیل و تفسیر نتایج حاصل از مدلسازی آلودگی هوا با استفاده از شاخص تراکم ترافیک و سرعت وزش باد پرداخته خواهد شد.

۳-۱ - مدلسازی براساس شاخص تراکم و سرعت باد

نتایج هر کدام از پنج حالت مختلف مدلسازی در جداول زیر ارائه شده‌است.

جدول ۲: نتایج حالت (۱) برای ایستگاه میدان فتح

Table 2. Results of Scenario (1) for Fath Square station

CO	۱.PM	۱.SO	معیار / آلاینده
مدل NN			
۲۱/۰	۱۶/۰	۱۲/۰	'R
۴۸/۱	۷۲/۴۵	۷۸/۴	MAE
۹۹/۱	۵۸/۵۷	۲/۶	RMSE
مدل GRU			
۱۶/۰	۲۵/۰	۲۷/۰	'R
۲۹/۱	۳۰	۶/۵	MAE
۸۸/۱	۴۳/۳۸	۹۵/۷	RMSE
مدل XGBoost			
۶۵/۰	۶۴/۰	۵۱/۰	'R
۷۸/۰	۱۵/۱۵	۴۵/۳	MAE
۱۱/۱	۰۹/۲۱	۸۷/۴	RMSE

جدول ۳: نتایج حالت (۲) برای ایستگاه میدان فتح

Table 3. Results of Scenario (2) for Fath Square station

CO	۱.PM	۱.SO	معیار / آلاینده
مدل NN			
۰/۳۳	۰/۲۱	۰/۱	'R
۱/۴۳	۴۴/۱	۴/۹۷	MAE
۱/۹۱	۵۴/۲	۵/۸۸	RMSE
مدل GRU			
۰/۱۷	۰/۲۵	۰/۴۷	'R
۱/۲۹	۳۰	۵/۶	MAE
۱/۸۹	۳۸/۴۳	۷/۹۵	RMSE
مدل XGBoost			
۰/۶	۰/۵۸	۰/۵۲	'R
۰/۸۶	۱۷/۲۱	۳/۴۱	MAE
۱/۱۹	۲۲/۸۸	۴/۸۱	RMSE

جدول ۴: نتایج حالت (۳) برای ایستگاه میدان فتح

Table 4. Results of Scenario (3) for Fath Square station

CO	۱.PM	۱.SO	معیار / آلاینده
مدل NN			
۰/۲۸	۰/۲۱	۰/۱۱	'R
۱/۴۴	۴۳/۹۵	۴/۹	MAE
۱/۹۸	۵۴/۱۷	۵/۸۳	RMSE
مدل GRU			
۰/۲	۰/۱۶	۰/۳۶	'R
۱/۲۹	۳۱/۳۱	۵/۸۴	MAE
۱/۸۵	۴۰/۴۷	۸/۱۷	RMSE
مدل XGBoost			
۰/۶۲	۰/۵۸	۰/۵۲	'R
۰/۸۹	۱۷/۲۱	۳/۴۱	MAE
۱/۲۸	۲۲/۸۸	۴/۸۱	RMSE

جدول ۵: نتایج حالت (۴) برای ایستگاه میدان فتح

Table 5. Results of Scenario (4) for Fath Square station

CO	PM	SO	معیار آلاینده
مدل NN			
۰/۱۹	۰/۱۶	۰/۰۵	R^2
۱/۳۱	۳۶	۴/۹۲	MAE
۱/۸۷	۴۷	۶/۱۱	RMSE
مدل GRU			
۰/۱۹	۰/۲۳	۰/۳۳	R^2
۱/۳۱	۳۰/۴۹	۵/۶۷	MAE
۱/۸۷	۳۹/۲۸	۸/۱	RMSE
مدل XGBoost			
۰/۵۵	۰/۶۱	۰/۵۲	R^2
۰/۹	۱۶/۸۳	۳/۴۱	MAE
۱/۲۶	۲۲	۴/۸۱	RMSE

جدول ۶: نتایج حالت (۵) برای ایستگاه میدان فتح

Table 6. Results of Scenario (5) for Fath Square station

CO	PM	SO	معیار آلاینده
مدل NN			
۰/۲	۰/۱۶	۰/۰۵	R^2
۱/۲۶	۳۶/۱۸	۵/۰۴	MAE
۱/۸۶	۴۶/۸	۶/۲	RMSE
مدل GRU			
۰/۱۸	۰/۲۱	۰/۳۸	R^2
۱/۲۹	۳۰/۷۲	۵/۹۲	MAE
۱/۸۷	۳۹/۲۸	۸/۳۲	RMSE
مدل XGBoost			
۰/۵۸	۰/۶۱	۰/۴۷	R^2
۰/۹۵	۱۶/۸۳	۳/۶۷	MAE
۱/۳۳	۲۲	۵/۰۹	RMSE

۳-۲- ارزیابی معناداری متغیرها

در تحلیل‌های آماری، ارزیابی معناداری متغیرها اهمیت زیادی دارد و به تعیین تأثیر یک متغیر بر متغیر وابسته کمک می‌کند. یکی از روش‌های رایج برای این ارزیابی، استفاده از P-Value است؛ معیاری که احتمال تصادفی بودن نتایج را نشان می‌دهد.

$P < 0.05$: معناداری آماری، به این معنا که می‌توان فرضیه صفر (عدم وجود اثر) را رد کرد.

$P \geq 0.05$: عدم معناداری، که نشان می‌دهد شواهد کافی برای رد فرضیه صفر وجود ندارد.

جدول ۷: مقادیر P-Value در مدلسازی

Table 7. P-values of the variables used in the modeling

P-Value	متغیر
۰/۰۰۰۱۰۸	سرعت باد
۰/۰۰۰۰۰۹	شاخص تراکم

۳-۳- تفسیر نتایج

برای انتخاب مدل برتر در مدلسازی آلودگی هوا، نتایج پنج حالت مختلف با استفاده از معیارهای ارزیابی مانند R^2 ، RMSE و MAE بررسی شد. مدل با بهترین عملکرد، کمترین مقدار RMSE و MAE و بالاترین مقدار R^2 را دارد. نتایج نشان داد که مدل XGBoost بهترین عملکرد را نسبت به مدل های ANN و GRU در پیش بینی آلودگی هوا دارد. این مدل توانسته است بالاترین مقدار R^2 و پایین ترین مقادیر RMSE و MAE را به دست آورد، که دقت بالای آن در پیش بینی میزان آلاینده ها را تأیید می کند.

نتایج مدل XGBoost برای پیش بینی سه آلاینده SO_2 ، PM_{10} و CO در پنج حالت مختلف برای ایستگاه میدان فتح بررسی شد. مقدار R^2 برای آلاینده ها نشان دهنده توانایی مدل در توضیح واریانس داده ها است. برای آلاینده CO، بالاترین مقدار R^2 (۰/۶۵) در حالت اول مشاهده شد که دقت قابل قبولی را نشان می دهد. در مقابل SO_2 ، پایین ترین مقدار R^2 (۰/۴۷) را در حالت پنجم داشت که بیانگر چالش های بیشتر در پیش بینی این آلاینده است.

معیارهای MAE و RMSE برای سنجش دقت پیش بینی ها استفاده شدند. دامنه وسیع داده های PM_{10} ، که بین ۲۰ تا ۳۰۰ در نوسان است، می تواند مقدار زیاد MAE و RMSE را توجیه نماید. به ویژه برای PM_{10} ، MAE بالای ۱۵ و RMSE بالای ۲۰ نشان دهنده نوسانات بزرگ در پیش بینی این آلاینده هستند.

حالت (۱) بهترین عملکرد را در پیش بینی PM_{10} نشان داد، در حالی که SO_2 در حالت (۵) کاهش دقت و افزایش خطاها را تجربه کرد. به طور کلی، مدل XGBoost توانایی خوبی در پیش بینی CO دارد، اما در پیش بینی PM_{10} و SO_2 با چالش هایی مواجه است. نتایج نشان می دهند که تأثیر تجمعی تراکم ترافیک و سرعت ورزش باد بر غلظت آلاینده های هوا در محدوده ایستگاه مورد بررسی چندان معنی دار نبوده و استفاده از داده های ساعتی می تواند به پیش بینی های بهتری منجر شود.

۴- نتیجه گیری

در این تحقیق، به بررسی و مدلسازی آلودگی هوا با استفاده از شاخص تراکم ترافیک و سرعت باد با استفاده از ابزارهای یادگیری ماشین پرداخته شد. برای مدلسازی کوتاه مدت آلودگی هوا با استفاده از داده های مربوط به شاخص تراکم ترافیک و سرعت ورزش باد ابتدا شاخص تراکم محاسبه شد. به منظور محاسبه شاخص تراکم، از رنگ های موجود در Google Maps استفاده شد که به صورت ساعتی برای یک ماه جمع آوری گردید. این داده ها سپس به عنوان ورودی برای پیش بینی آلاینده های SO_2 ، PM_{10} و CO با استفاده از مدل های XGBoost، شبکه های عصبی (ANN) و مدل GRU به کار گرفته شدند. مدلسازی در پنج حالت مختلف انجام شد که هر یک از این حالات به منظور بررسی اثرات مختلف شاخص تراکم و سرعت ورزش باد طراحی گردید.

پس از مدلسازی، اعتبارسنجی مدل ها و تعیین معناداری متغیرها انجام شد. نتایج نشان داد که حالت (۱) برای ایستگاه میدان فتح و مدل XGBoost بهترین عملکرد را در پیش بینی غلظت آلاینده ها داشته اند. این امر به ویژه در پیش بینی آلاینده های CO و PM_{10} مشهود است و نشان دهنده قدرت بالای مدل XGBoost در تحلیل داده های پیچیده و غیرخطی است.

سهم علمی این پژوهش نشان دادن امکان استفاده از اطلاعات تراکم ترافیک برخط و سرعت باد برای پیش بینی کوتاه مدت آلودگی هوا در مقیاس محلی و نیز نشان دادن برتری داده های ساعتی نسبت به بازنمایی های تجمعی بررسی شده است. این چارچوب می تواند برای پشتیبانی از پایش و تصمیم گیری شهری توسعه یابد؛ با این حال، برای افزایش قابلیت تعمیم، بررسی دوره های زمانی طولانی تر، ایستگاه های بیشتر و متغیرهای هواشناسی و ترافیکی تکمیلی پیشنهاد می شود.

منابع

- [1] Kojima, M., Brandon, C., and Shah, J., 2000. Improving Urban Air Quality in South Asia by Reducing Emissions from Two-Stroke Engine Vehicles. World Bank, Washington, DC.

- [2] Levy, J.I., Buonocore, J.J., and Von Stackelberg, K., 2010. "Evaluation of the public health impacts of traffic congestion: a health risk assessment". *Environmental Health*, 9, pp. 1–12.
- [3] Darvishi, G., Yousefi Kebria, D., Ehteshami, M., Asadi-Ghalhari, M., and Golbabaee Kootenaei, F., 2020. "Prediction of traffic-induced air pollution in suburban roads using an ozone pollutant modeling with a regression method". *Journal of Health Research in Community*, 6(3), pp. 55–64. (in Persian)
- [4] Farhadi, R., Hadavifar, M., Moeen al-Dini, M., and Amin Toosi, M., 2019. "Prediction of air pollutant concentrations in Tehran based on meteorological variables using artificial neural networks and linear regression in hot and cold seasons". *Journal of Natural Environment*, 73(1), pp. 115–127. (in Persian)
- [5] Hamdi, M.A., Samiepour, M., Talebi, S.A., and Zare Shahne, M., 2025. "Data-driven analysis and predictive modeling of Air Quality Index (AQI) using machine learning techniques: a case study of Tehran (District 2), Iran". In 14th International Congress on Civil Engineering.
- [6] Jamali Donighi, S., Moradnezhad Hessari, M., and Ghodsipour, S.H., 2016. "Investigation of traffic factors affecting air pollution". In 3rd International Conference on Sustainable Development, Strategies and Challenges with an Emphasis on Agriculture, Natural Resources, Environment and Tourism. (in Persian)
- [7] Ahmadi Chalesara, A., Mehrzad, A., Ahmadi Chalesara, M., and Nourafkan, R., 2025. "Dynamic modeling of the interaction between air pollution and urban traffic (case study: Rasht City)". In 13th International Conference on Civil Engineering, Architecture, Urban Planning and Environment in the 21st Century. (in Persian)
- [8] Shakya, D., Deshpande, V., Goyal, M.K., and Agarwal, M., 2023. "PM_{2.5} air pollution prediction through deep learning using meteorological, vehicular, and emission data: A case study of New Delhi, India". *Journal of Cleaner Production*, 427, pp. 139278.
- [9] Freeman, B.S., Taylor, G., Gharabaghi, B., and Thé, J., 2018. "Forecasting air quality time series using deep learning". *Journal of the Air & Waste Management Association*, 68(8), pp. 866–886.
- [10] Kothandaraman, D., Praveena, N., Varadarajkumar, K., Madhav Rao, B., Dhabliya, D., Satla, S., and Abera, W., 2022. "Intelligent forecasting of air quality and pollution prediction using machine learning". *Adsorption Science & Technology*, 2022, pp. 5086622.
- [11] Wang, Z., Wu, X., and Wu, Y., 2023. "A spatiotemporal XGBoost model for PM_{2.5} concentration prediction and its application in Shanghai". *Heliyon*, 9(12), pp. e22569.
- [12] Lei, T.M., Ng, S.C., and Siu, S.W., 2023. "Application of ANN, XGBoost, and other ML methods to forecast air quality in Macau". *Sustainability*, 15(6), pp. 5341.
- [13] Hoffman, S., Filak, M., and Jasiński, R., 2022. "Air Quality Modeling with the Use of Regression Neural Networks". *International Journal of Environmental Research and Public Health*, 19(24), pp. 16494.
- [14] Maleki, H., Sorooshian, A., Goudarzi, G., Baboli, Z., Tahmasebi Birgani, Y., and Rahmati, M., 2019. "Air pollution prediction by using an artificial neural network model". *Clean Technologies and Environmental Policy*, 21, pp. 1341–1352.
- [15] Golshan Khavas, R., Shariat Mohaymany, A., and Rahimi, Z., 2022. "Evaluation of travel time data from Google Maps and Neshan services". In 13th National Congress on Civil Engineering, Isfahan, Iran. (in Persian)