

Statistical Analysis and Forecast Modeling of PM_{2.5} Concentration Using Artificial Intelligence Based on Machine Learning in Mashhad (2016-2022)

Ahmad Makhdoomi

Master's Student in Environmental Health Engineering, Member of the Student Research Committee, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran.

Maryam Sarkhosh

* Associate Professor, Department of Environmental Health, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran.

(Corresponding Author)
SarkhoshM@mums.ac.ir

Somayyeh Ziaei

Postdoctoral Researcher, School of Health, Member of the Student Research Committee, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran.

Abstract

Background and Purpose: This study aims to forecast PM_{2.5} concentrations using four non-linear Machine Learning (ML) models.

Materials and Methods: The ML techniques employed include Light Gradient Boosting Machine (LGBM), Extreme Gradient Boosting Regressor (XGBR), Random Forest (RF), and Gradient Boosting Regressor (GBR). Meteorological and pollutant data were collected to predict the Air Quality Index (AQI) in Mashhad, Khorasan Razavi Province, Iran, for the period from 2016 to 2022.

Results: The ML models performed exceptionally well in predicting PM_{2.5} concentrations, with approximately 95% of their predictions falling within a factor of the observed values. Additionally, the predicted PM_{2.5} concentrations were compared with observed values to assess prediction accuracy. Among the four ML models, GBR demonstrated the best performance, achieving high accuracy metrics, including a coefficient of determination (R²) of 0.9802, a mean absolute error (MAE) of 0.54, a mean squared error (MSE) of 5.33, a root mean squared error (RMSE) of 2.31, and a mean absolute percentage error (MAPE) of 1.9%.

Conclusion: This study proposes a high-accuracy PM_{2.5} prediction method using ML, which can be beneficial for global air quality monitoring and improving acute exposure assessments in epidemiological research.

Keywords: Air quality index, PM_{2.5}, Machine learning, Non-linear Models

Open Access Policy: This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>

Received: 2024/09/27

Accepted: 2024/12/22

Doi:10.22038/jreh.2025.25727

► **Citation:** Makhdoomi A, Sarkhosh M, Ziaei S. Statistical Analysis and Forecast Modeling of PM_{2.5} Concentration Using Artificial Intelligence Based on Machine Learning in Mashhad (2016-2022). *Iranian Journal of Research in Environmental Health*. Winter 2025; 10 (4):22-35.

پیش‌بینی و آنالیز آماری غلظت PM_{2.5} با استفاده از هوش مصنوعی مبتنی بر یادگیری ماشین در شهر مشهد (۱۳۹۵-۱۴۰۱)

احمد مخدومی

دانشجو ارشد مهندسی بهداشت محیط، عضو کمیته تحقیقات دانشجویی دانشکده بهداشت دانشگاه علوم پزشکی مشهد، مشهد، ایران.

مریم سرخوش

* دانشیار گروه بهداشت محیط، دانشکده بهداشت، دانشگاه علوم پزشکی مشهد، مشهد، ایران. (نویسنده مسئول)

SarkhoshM@mums.ac.ir

سمیه ضیائی

دانشجو دوره تکمیلی پژوهشی دانشکده بهداشت، عضو کمیته تحقیقات دانشجویی دانشکده بهداشت دانشگاه علوم پزشکی مشهد، مشهد، ایران.

تاریخ دریافت: ۱۴۰۳/۰۷/۰۶

تاریخ پذیرش: ۱۴۰۳/۱۰/۰۲

نوع مقاله: پژوهشی

چکیده

زمینه و هدف: آلودگی هوای شهرهای صنعتی و بزرگ ناشی از وجود آلاینده‌های گوناگونی به‌خصوص ذرات معلق با قطر کمتر از ۲/۵ میکرون است که به هوای شهری وارد می‌شود. پیش‌بینی مکان‌هایی که آلاینده PM_{2.5} به میزان زیاد است به مدیریت و برنامه‌ریزی صحیح در راستای بهبود کیفیت هوا کمک شایانی خواهد کرد. هدف از این مطالعه پیش‌بینی غلظت PM_{2.5} با استفاده از چهار مدل غیرخطی هوش مصنوعی مبتنی بر روش یادگیری ماشین است.

مواد و روش‌ها: تکنیک‌های یادگیری ماشین مورد استفاده در این مطالعه شامل: ماشین تقویت گرادیان سبک، رگرسیون تقویت گرادیان پیشرفته، جنگل تصادفی و رگرسیون با تقویت گرادیان بود. داده‌های هواشناسی و غلظت ذرات معلق برای پیش‌بینی شاخص کیفیت هوا در بازه‌ی زمانی سال‌های ۱۳۹۵ تا ۱۴۰۱ در شهر مشهد جمع‌آوری گردید.

یافته‌ها: هر چهار مدل یادگیری ماشین در پیش‌بینی غلظت PM_{2.5} عملکرد بسیار خوبی را نشان دادند و حدود ۹۵ درصد از پیش‌بینی‌های آن‌ها در محدوده‌ی فاکتور غلظت مشاهده‌شده قرار داشت. نتایج این مطالعه نشان می‌دهد که مدل رگرسیون با تقویت گرادیان از میان چهار الگوریتم استفاده شده بر پایه یادگیری ماشین، عملکرد بهتری را نسبت به سایر مدل‌های غیرخطی با معیارهای دقت بالا از جمله ضریب رگرسیون ۰/۹۸۰۲، میانگین خطای مطلق ۰/۵۴، میانگین خطای مربعات ۵/۳۳، ریشه میانگین خطای مربع ۲/۳۱ و میانگین درصد خطای مطلق ۱/۹٪ را نشان می‌دهد.

نتیجه‌گیری: در نتیجه، این مطالعه روشی را برای به‌دست آوردن نتایج پیش‌بینی PM_{2.5} با دقت بالا با استفاده از هوش مصنوعی مبتنی بر یادگیری ماشین را پیشنهاد می‌کند که برای پایش کیفیت هوا در مقیاس جهانی و بهبود ارزیابی مواجهه حاد در تحقیقات اپیدمی مفید است.

کلیدواژه‌ها: آلودگی هوا، PM_{2.5}، یادگیری ماشین، مدل‌های غیرخطی

◀ **استناد:** مخدومی ا، سرخوش م، ضیائی س. پیش‌بینی و آنالیز آماری غلظت PM_{2.5} با استفاده از هوش مصنوعی مبتنی بر یادگیری ماشین در شهر مشهد (۱۳۹۵-۱۴۰۱). فصلنامه‌ی پژوهش در بهداشت محیط. زمستان ۱۴۰۳؛ ۱۰(۴): ۲۲-۳۵.

آلودگی هوای شهرهای بزرگ و صنعتی تبدیل به یکی از مهم‌ترین نگرانی‌های زیست محیطی شده است (۱). ذرات ریز با قطر آیرودینامیکی کمتر از 2.5 میکرومتر که یکی از انواع آلاینده‌های اصلی هوای شهری به شمار می‌رود و به‌شدت با بیماری‌های مختلفی، مانند سرطان ریه و مرگ‌های زودرس در انسان مرتبط است (۲). افزایش غلظت $PM_{2.5}$ علاوه بر این که تهدیدی برای سلامتی انسان‌ها به شمار می‌رود، باعث بروز هزینه‌های اقتصادی و تأثیر منفی بر حاصل‌خیزی خاک، چرخه‌ی غذایی، فتوسنتز گیاهان و منابع آبی می‌شود، بنابراین اندازه‌گیری و پایش میزان غلظت این ذرات امری ضروری است. با وجود مطالعه و بررسی این آلاینده در تحقیقات مختلف، متأسفانه ایستگاه‌های نظارت کم و محدودی جهت بررسی کیفیت هوا و دستیابی به توزیع مکانی دقیق غلظت $PM_{2.5}$ وجود دارد. پیش‌بینی غلظت $PM_{2.5}$ در آینده، امکان اطلاع‌رسانی به عموم مردم برای برنامه‌ریزی بهتر فعالیت‌های روزانه و همچنین تصمیم‌گیری به‌موقع و اقدامات پیشگیرانه برای جلوگیری از مشکلات این آلاینده را فراهم می‌سازد (۳، ۴).

روش‌های هوش مصنوعی مبتنی بر یادگیری ماشین^۱ (ML) در دهه اخیر به‌دلیل ظرفیت بالای خود در مدل‌سازی کیفیت هوا و پیش‌بینی غلظت‌های $PM_{2.5}$ به‌طور چشمگیری مورد توجه قرار گرفته‌اند (۵، ۶). تلاش برای یافتن مدل‌های بهتر نسبت به مدل‌های سنتی سنجش کیفیت هوا، منجر به انجام تحقیقات گسترده‌ای در زمینه استفاده از رویکردهای هوشمند مختلف یادگیری ماشین شده است که قادر به پیش‌بینی دقیق شاخص‌های کیفیت هوا هستند (۷). مدل‌های متعددی از جمله شبکه‌های عصبی مصنوعی، درختان رگرسیون تقویت‌شده، جنگل‌های تصادفی و ماشین‌های بردار پشتیبان به‌طور معمول برای ایجاد مدل‌های کیفیت هوا با دقت قابل قبول یا حتی بهتر به کار می‌روند (۸-۱۰). این الگوریتم‌ها می‌توانند با هزینه‌ی محاسباتی کمتری به‌دست آیند و نیازی به فرضیات خاصی

درباره فرآیندهای جوی ندارند (۱۱). الگوریتم‌های یادگیری ماشین قادر به شبیه‌سازی تعاملات پیچیده و غیرخطی بین پارامترهای کیفیت هوا هستند و مدل‌هایی تولید می‌کنند که در پیش‌بینی داده‌های ناشناخته عملکرد خوبی دارند (۱۲، ۱۳).

در دهه‌های اخیر، تکنیک‌های یادگیری ماشین تبدیل به یکی از ابزارهای قدرتمندی شده است که برای مدل‌سازی کیفیت هوا و پیش‌بینی غلظت ذرات به کار برده شده است. تحقیقات برای جستجوی مدل‌های با کیفیت جهت مدل‌سازی کیفیت هوا، منجر به ارائه‌ی مطالعات زیادی در مورد استفاده از روش‌های هوشمند مختلف یادگیری ماشین شده است که بتوانند به‌طور دقیق شاخص‌های کیفیت هوا را پیش‌بینی کنند و در مقالات مختلف به آن‌ها اشاره شده است (۱۴، ۱۵). هدف از این پژوهش بررسی کارایی هوش مصنوعی یادگیری ماشین با استفاده از چهار مدل رگرسیون غیرخطی جهت پیش‌بینی میزان غلظت $PM_{2.5}$ در فاصله‌ی سال‌های ۱۳۹۵-۱۴۰۱ در مشهد است.

روش کار

منطقه‌ی مورد مطالعه

منطقه‌ی مورد مطالعه، شهر مشهد، دومین کلان‌شهر بزرگ ایران و مرکز استان خراسان‌رضوی است. این شهر با مساحت تقریبی ۲۸۰ کیلومتر مربع و جمعیتی بالغ بر ۳،۱۳۱،۵۸۶ نفر (بر اساس آخرین سرشماری)، یکی از مهم‌ترین مراکز زیارتی و گردشگری ایران به‌شمار می‌رود. حضور حرم مطهر امام رضا (ع) موجب شده است که سالانه بیش از ۲۰ میلیون زائر از این شهر بازدید کنند. در سال‌های اخیر، توسعه سریع شهری و صنعتی مشهد همراه با افزایش تراکم جمعیت، منجر به تشدید انتشار آلاینده‌های هوا در این منطقه شده است. دوره‌ی زمانی این مطالعه، سال‌های ۱۳۹۵ تا ۱۴۰۱ را شامل می‌شود. شکل ۱ نقشه شماتیک منطقه مورد مطالعه را ارائه می‌دهد.

¹ Machine Learning

و حذف شدند تا تأثیر داده‌های غیرعادی بر عملکرد مدل‌ها به حداقل برسد. همچنین، برای جلوگیری از سوگیری و حفظ اعتبار نتایج، رکوردهای دارای مقادیر گم‌شده به‌طور کامل حذف شدند. به‌منظور رفع مشکلات ناشی از تفاوت مقیاس متغیرها و بهبود همگرایی مدل‌ها، داده‌ها با استفاده از روش نرمال‌سازی Min-Max به بازه [۰, ۱] انتقال یافتند.

تحلیل همبستگی

رابطه‌ی بین آلاینده‌های اندازه‌گیری‌شده با استفاده از همبستگی اسپیرمن محاسبه شد. در این همبستگی، روابط بین متغیرهای عددی اندازه‌گیری می‌شود. بنابراین، اساساً معیاری از رابطه‌ی یکنواخت بین آن دو متغیر را ارائه می‌دهد. دامنه همبستگی از -۱ تا ۱ است. اگر همبستگی نزدیک به ۱ باشد، متغیرها، همبستگی مثبت دارند، در حالی که -۱ به معنای همبستگی منفی است (۱۶، ۱۷). تحلیل همبستگی اسپیرمن برای بررسی رابطه یکنواخت بین آلاینده‌های اندازه‌گیری‌شده به کار گرفته شد. هدف از این تحلیل، شناسایی میزان ارتباط میان آلاینده‌ها و تعیین تأثیرات احتمالی متقابل آن‌ها بر یکدیگر است. این روش به‌ویژه برای داده‌هایی با توزیع غیرنرمال یا مقیاس‌های ترتیبی مناسب بوده و اطلاعات ارزشمندی درباره ارتباطات غیرخطی بین متغیرها فراهم می‌کند.

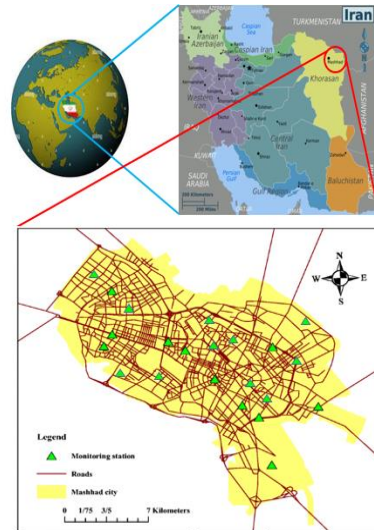
مدل‌های یادگیری ماشینی

ماشین تقویت گرادیان سبک^{۱۰} (LGBM)

الگوریتم LGBM یک ابزار قدرتمند و کارآمد برای یادگیری ماشینی است که می‌تواند هم برای کارهای رگرسیون و هم برای طبقه‌بندی استفاده شود، به‌ویژه برای مجموعه‌های داده در مقیاس بزرگ که کارایی مدل را بهبود می‌بخشد و استفاده از حافظه را کاهش می‌دهد. LGBM می‌تواند انواع مختلف داده مانند دسته‌بندی، عددی و متن را مدیریت کند. همچنین شامل ویژگی‌های داخلی برای پیش‌پردازش داده‌ها، اعتبارسنجی متقابل و تنظیم‌های پارامتر است که بهینه‌سازی مدل‌های خود را برای کاربران آسان‌تر می‌کند (۱۸). معادله LGBM به صورت زیر است:

$$\hat{y}_i = \operatorname{argmin}_f E_{y,x} L(y, f(x)) \quad (1)$$

¹⁰ Light Gradient Boosting Machine (LGBM)



شکل ۱. شماتیک منطقه‌ی مورد مطالعه

جمع‌آوری داده‌ها

در این مطالعه، از داده‌های هواشناسی و آلاینده‌های هوا شهر مشهد از سال ۱۳۹۵ تا ۱۴۰۱ برای تحلیل کیفیت هوا استفاده شد. به‌عنوان شاخص نهایی، غلظت $PM_{2.5}$ به‌دلیل ارتباط مستقیم آن با شرایط حاد کیفیت هوا انتخاب گردید. داده‌های مورد استفاده شامل اندازه‌گیری‌های ایستگاه‌های پایش کیفیت هوا^۱ بود که هر دو نوع ذرات معلق، PM_{10} و $PM_{2.5}$ ، را به‌دلیل اثرات قابل‌توجه آن‌ها بر کیفیت هوا شامل می‌شد. همچنین، متغیرهای هواشناسی از هر ایستگاه برای توسعه مدل‌های یادگیری ماشین استخراج شدند که عبارتند از: (حداقل دید افقی)^۲، (میانگین سرعت باد)^۳، (میزان بارش ۲۴ ساعته)^۴، (میانگین رطوبت نسبی)^۵، (جهت باد)^۶، و (تعداد گزارش گردوغبار)^۷. این داده‌ها از سازمان هواشناسی کشور گردآوری شدند و به‌دلیل نقش کلیدی در تعیین شاخص کیفیت هوا^۸ (AQI) در مناطق خاص، اهمیت بالایی در مدل‌سازی دارند.

در فرآیند پیش‌پردازش داده‌ها، ابتدا مقادیر پرت با استفاده از تحلیل فاصله بین چارک‌ها^۹ (IQR) شناسایی

¹ Air Quality Monitoring

² vvmn

³ ffm

⁴ rrr24

⁵ um

⁶ dd

⁷ nhz

⁸ Air Quality Index

⁹ Inter Quartile Range

حضور نویز یا تغییرات جزئی در داده‌های ورودی است. این ویژگی، RF را به ابزاری قدرتمند برای مسائل پیچیده رگرسیون و طبقه‌بندی تبدیل کرده است (۲۱). فرمول جنگل تصادفی را می‌توان به صورت زیر بیان کرد:

$$RF = \frac{1}{k} \sum_{k=1}^k h_k(x) \quad (3)$$

که در آن، K پیش‌بینی‌های تولید شده است و $h_k(x)$ به‌عنوان درخت‌های رگرسیون است.

رگرسیون با تقویت گرادیان^۳ (GBR)

رویکرد رگرسیون با تقویت گرادیان (GBR) به‌عنوان یک الگوریتم گرادیان تابعی عمل می‌کند که هدف آن به-حداقل رساندن تابع خطا از طریق ساخت یک مدل افزایشی است. ویژگی کلیدی GBR این است که در هر مرحله درخت‌های تصمیم قبلی اصلاح نمی‌شوند، بلکه مدل با افزودن درخت‌های جدید بر اساس باقیمانده‌ها، در نقاط ضعیف عملکرد خود به تدریج بهبود پیدا می‌کند (۲۲). فرمول رگرسیون تقویت گرادیان در رابطه ۴ نشان داده شده است

$$f_N(x_i) = \sum_n y_n h_n(x_i) \quad (4)$$

ارزیابی مدل‌ها

دقت پیش‌بینی PM_{2.5} مبتنی بر یادگیری ماشین در این مطالعه مورد ارزیابی قرار گرفت. به‌منظور سنجش کارایی مدل، پنج شاخص آماری برای انطباق، شامل میانگین مربعات خطا (MSE)، ضریب تعیین (R²)، ریشه میانگین مربعات خطا (RMSE)، میانگین مطلق خطا (MAE) و میانگین درصد خطای مطلق (MAPE) استفاده شد تا همبستگی بین غلظت‌های تخمینی و مشاهده شده PM_{2.5} را طبق معادلات (۵) تا (۹) ارزیابی کند (۲۳، ۲۴).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

هدف اصلی LGBM بهینه‌سازی مقادیر پیش‌بینی‌شده مدل ($f(x)$) با به‌حداقل رساندن تابع خطا (L) با استفاده از مجموعه آموزشی تحت نظارت (X) است.

رگرسیون تقویت گرادیان پیشرفته^۱ (XGBR)

الگوریتم XGBR یکی از روش‌های پیشرفته در حوزه یادگیری گروهی است که با ترکیب پیش‌بینی‌های مجموعه‌ای از مدل‌های پایه یا ضعیف، یک مدل نهایی قدرتمند و دقیق ایجاد می‌کند. این الگوریتم به دلیل توانایی بالای آن در مدیریت انواع داده‌ها، کشف روابط پیچیده، و تطبیق با توزیع‌های مختلف، به‌طور گسترده در مسائل یادگیری ماشین مانند طبقه‌بندی و رگرسیون مورد استفاده قرار می‌گیرد. علاوه بر این، XGBR به دلیل بهره‌گیری از طیف گسترده‌ای از فرآیندها، قابلیت تنظیم و بهینه‌سازی دقیقی دارد و برای تحلیل مجموعه داده‌های بزرگ و پیچیده بسیار کارآمد است (۱۹، ۲۰). فرمول زیر برای XGBR است:

$$\hat{Y}_i^t = \sum_{k=1}^t f_k(x_i) = \hat{Y}_i^{(t-1)} + f_t(x_i) \quad (2)$$

جایی که: \hat{Y}_i^t و $\hat{Y}_i^{(t-1)}$ به ترتیب پیش‌بینی‌های مرحله t و t₁ هستند، $f_t(x_i)$ یک یادگیرنده در مرحله t و x_i متغیر ورودی است.

جنگل تصادفی^۲ (RF)

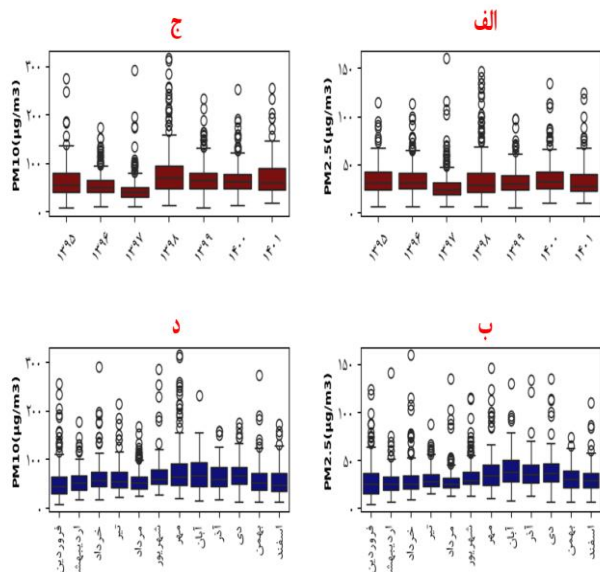
الگوریتم جنگل تصادفی یک روش یادگیری گروهی پیشرفته است که با استفاده از مجموعه‌ای بزرگ از درخت‌های تصمیم، خروجی نهایی را تعیین می‌کند. این الگوریتم با نمونه‌گیری تصادفی از زیرمجموعه‌های داده، شامل نمونه‌گیری از ردیف‌ها (داده‌ها) و ستون‌ها (ویژگی‌ها)، اقدام به ساخت درختان تصمیم می‌کند. چنین رویکردی امکان کاهش هم‌بستگی بین درخت‌ها و افزایش دقت پیش‌بینی را فراهم می‌آورد. یکی از مزایای کلیدی این روش، قدرت بالای آن در تعمیم‌دهی و تخمین داده‌های جدید، حتی در

¹ Extreme Gradient Boosting Regressor (XGBR)

² Random Forest (RF)

³ Gradient Boosting Regressor (GBR)

۱۶۰/۲۵ میکروگرم بر مترمکعب و ۳۱۵/۵۵ میکروگرم بر مترمکعب بود که نشان‌دهنده‌ی بروز دوره‌هایی از سطوح بالای آلودگی است. به‌طور کلی، این یافته‌ها الگوهای زمانی و تاثیر عوامل طبیعی و انسانی بر غلظت‌های $PM_{2.5}$ و PM_{10} را نشان می‌دهد (۲۵). رابطه‌ی بین شاخص کیفیت هوا (AQI) و غلظت‌های $PM_{2.5}$ اهمیت نظارت و توجه به ذرات‌معلق در ارزیابی و پیش‌بینی کیفیت هوا را برجسته می‌کند.



شکل ۲. مقادیر توزیع سالانه و ماهانه $PM_{2.5}$ (الف و ب) و PM_{10} (ج و د)

شاخص کیفیت هوا و پارامترهای هواشناسی

شاخص کیفیت هوا (AQI) معمولاً با مقدار ذرات‌معلق و آلاینده‌ها در هوا ارتباط نزدیکی دارد و این شرایط جوی می‌توانند منجر به افزایش سطوح AQI شوند. علاوه بر این، مقدار کل بارش نیز نقش مهمی در تعیین AQI در مشهد ایفا می‌کند. در مشهد دوره‌ی زمانی از آذر تا اردیبهشت به‌طور قابل توجهی بارش بیشتری نسبت به سایر ماه‌ها دارد و شاخص کیفیت هوا در این دوره‌ی زمانی بهبود می‌یابد (شکل ۳).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

$$\%MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

در اینجا، n به تعداد نقاط داده، y_i به مقدار واقعی و \hat{y}_i به مقدار پیش‌بینی شده اشاره دارد.

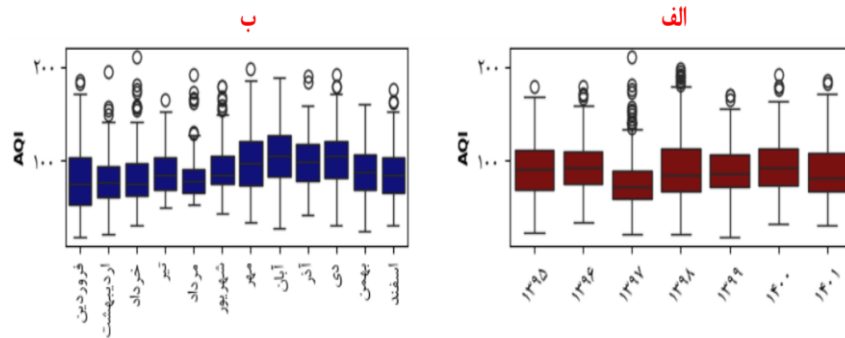
یافته‌ها

ذرات‌معلق (PM_{10} و $PM_{2.5}$)

الگوهای ماهانه و سالانه میانگین سری‌های زمانی داده‌های جمع‌آوری‌شده در این مطالعه (۱۳۹۵-۱۴۰۱) در شکل ۲ نشان داده شده‌اند که الگوی تغییرات غلظت $PM_{2.5}$ (الف) و PM_{10} (ب) را به‌ترتیب نمایش می‌دهند. نتایج نشان می‌دهد که غلظت‌های PM_{10} به‌طور مداوم در طی ماه‌های شهریور تا آذر و همچنین در ماه آبان از حد آستانه ۶۰ میکروگرم بر مترمکعب فراتر رفته است. این موضوع نشان‌دهنده‌ی دوره‌ای از سطوح بالای آلودگی در این ماه‌ها است. علاوه بر این، در سال‌های ۱۳۹۹ تا ۱۴۰۰، غلظت‌های پایین PM_{10} و $PM_{2.5}$ (به‌ترتیب کمتر از ۶۵ میکروگرم بر مترمکعب و ۳۰ میکروگرم بر مترمکعب) که ناشی از کاهش فعالیت‌های انسانی و انتشار آلاینده‌ها به‌دلیل تاثیرات ناشی از همه‌گیری COVID-19 بود.

بر اساس شکل ۲، میانگین سالانه‌ی غلظت‌های $PM_{2.5}$ و PM_{10} در مشهد بالاتر از دستورالعمل‌های کیفیت هوای سازمان جهانی بهداشت^۱ (WHO) بود. این موضوع نشان می‌دهد که پیش‌بینی $PM_{2.5}$ تأثیر قابل‌توجهی بر سطح شاخص کیفیت هوا (AQI) دارد. علاوه بر این، میانگین غلظت‌های PM_{10} و $PM_{2.5}$ در طی یک دوره‌ی هفت ساله از ۱۳۹۵ تا ۱۴۰۱ به‌ترتیب ۳۲/۸۴ میکروگرم بر مترمکعب و ۶۱/۸۸ میکروگرم بر مترمکعب محاسبه شد. این مقادیر میانگین، درک کلی از سطوح آلودگی در این بازه‌ی زمانی ارائه می‌دهند. همچنین، حداکثر غلظت‌های ثبت‌شده برای PM_{10} و $PM_{2.5}$ به‌ترتیب

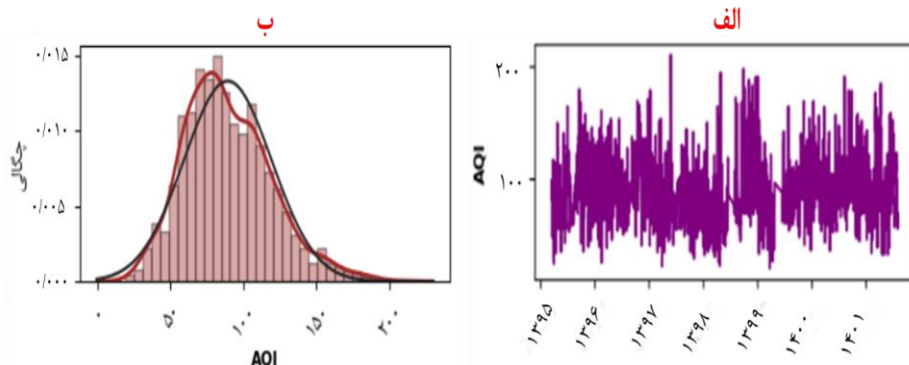
¹ World Health Organization



شکل ۳. توزیع مقادیر میانگین سالانه و ماهانه کیفیت هوا (AQI)

تجمع ذرات معلق در زمستان می‌شود. در شکل ۵ توزیع پارامترهای هواشناسی نشان داده شده است. همان‌طور که در این شکل مشاهده می‌شود، توزیع پارامتر هواشناسی به جز $ITR24$ تقریباً با حالت نرمال مشابه است.

شکل ۴ توزیع و تغییرات روزانه AQI را از سال ۱۳۹۵ تا ۱۴۰۱ نشان می‌دهند. داده‌ها کاهش مداوم AQI را در هر ماه و سال آشکار می‌کند. این الگو را می‌توان با عوامل مختلفی توضیح داد: از جمله وارونگی‌های دما که منجر به



شکل ۴. تغییرات روزانه شاخص کیفیت هوا (الف) و توزیع شاخص کیفیت هوا (ب) از سال ۱۳۹۵ تا ۱۴۰۱

مدل‌ها، از جستجوی شبکه‌هایی^۲ همراه با اعتبارسنجی متقابل^۳ استفاده شد که امکان یافتن بهترین ترکیب پارامترها را با ارزیابی چندگانه عملکرد فراهم کرد. در این روش، مجموعه‌ای از مقادیر مختلف برای هر هاپیر پارامتر مدل آزمایش شده و بهترین ترکیب پارامترها با کمترین خطای پیش‌بینی شناسایی می‌شود. اعتبارسنجی متقابل به‌منظور ارزیابی دقیق‌تر عملکرد مدل و جلوگیری از برازش بیش از حد^۴ به‌کار گرفته

مقایسه عملکرد مدل‌های یادگیری ماشین برای پیش‌بینی $PM_{2.5}$

در این مطالعه، چهار مدل یادگیری ماشین شامل GBR، RF، XGBR و LGBM برای برآورد غلظت $PM_{2.5}$ مورد استفاده قرار گرفتند. به‌منظور آموزش و ارزیابی، برای انتخاب ۸۰٪ از داده‌ها جهت آموزش مدل‌ها، از روش تقسیم تصادفی^۱ استفاده شد. در این روش، داده‌ها به‌صورت تصادفی به دو بخش آموزشی (۸۰٪) و آزمایشی (۲۰٪) تقسیم شدند تا مدل‌ها بتوانند در شرایطی شبیه به داده‌های واقعی آزمایش شوند. این تقسیم‌بندی تصادفی به جلوگیری از سوگیری انتخاب کمک کرده و تضمین می‌کند که هر نمونه، شانس مساوی برای قرارگرفتن در هر یک از بخش‌ها دارد. برای تنظیم بهینه‌ی پارامترهای

² Grid Search

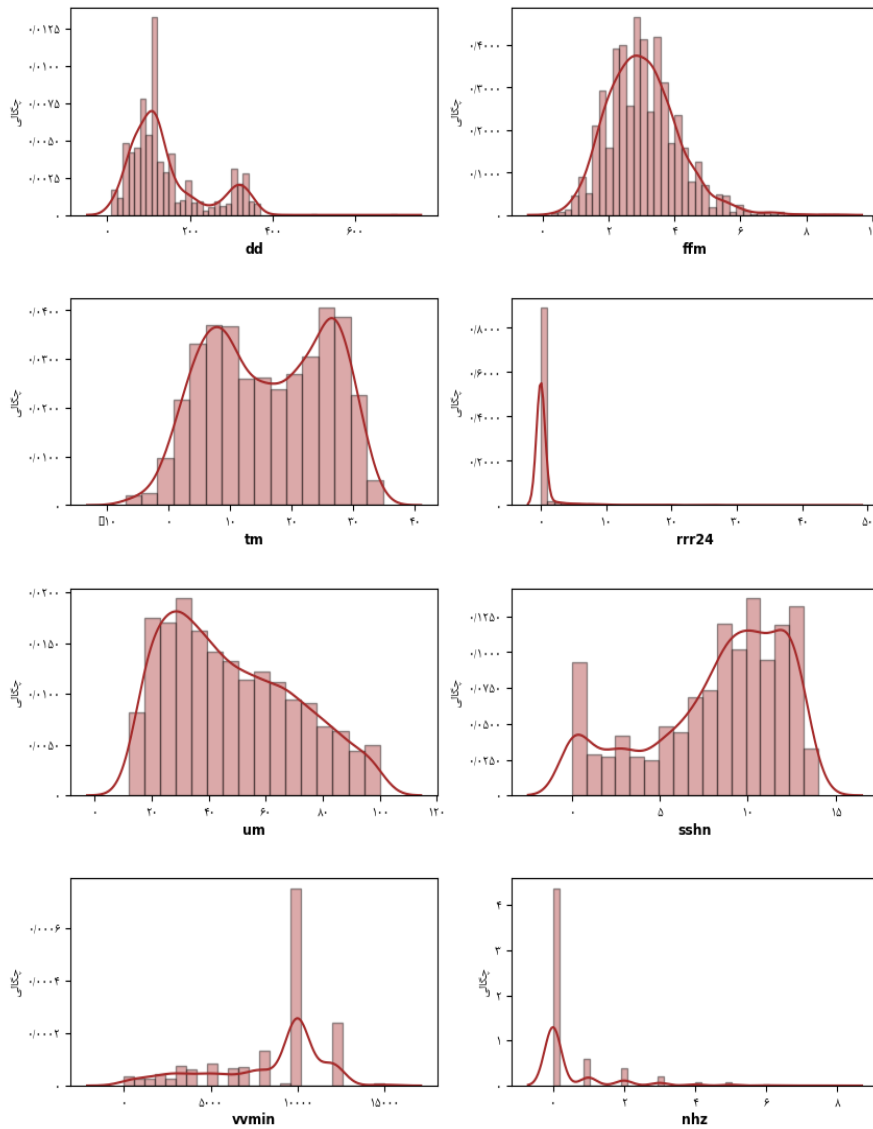
³ Cross-Validation

⁴ Overfitting

¹ Random Split

برازش، و توانایی ثبت تغییرات در غلظت $PM_{2.5}$ ارائه می‌دهند. این معیارها برای هر دو مجموعه داده‌های آموزشی و آزمایشی محاسبه شدند، اما از آنجا که مقادیر مجموعه آزمایشی معیار واقعی عملکرد مدل‌ها در داده‌های جدید محسوب می‌شوند، تحلیل اصلی بر این مجموعه متمرکز شد.

شد، به طوری که داده‌ها به k بخش تقسیم و مدل در هر دور روی داده‌های مختلف آموزش داده و ارزیابی می‌شود. جدول ۱ مقایسه معیارهای ارزیابی پیش‌بینی $PM_{2.5}$ برای هر الگوریتم یادگیری ماشین را نشان می‌دهد. این معیارها شامل R^2 و $MAPE$ ، $RMSE$ ، MSE ، MAE هستند که هر کدام اطلاعات مهمی درباره‌ی جنبه‌های مختلف عملکرد مدل‌ها، شامل دقت پیش‌بینی، کیفیت



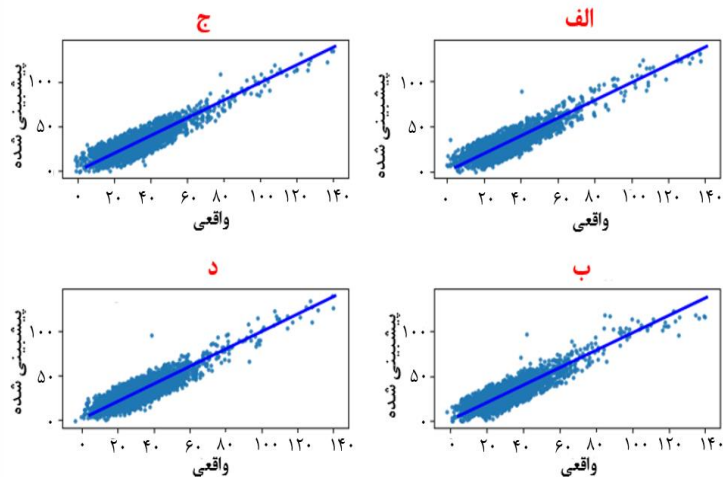
شکل ۵. مقایسه توزیع پارامترهای هواشناسی با حالت توزیع نرمال

جدول ۱. عملکرد مدل‌ها بر اساس شاخص‌های مختلف در پیش‌بینی $PM_{2.5}$

مجموعه تست (۲۰ درصد)					مجموعه آموزش (۸۰ درصد)					مدل
R^2	MAPE (%)	RMSE	MSE	MAE	R^2	MAPE (%)	RMSE	MSE	MAE	
۰/۹۶۱۱	۲/۶	۳/۱۹	۱۰/۲	۰/۹	۰/۹۷۸۴	۲/۳	۲/۴	۵/۷۴	۰/۷۵	LGBM
۰/۹۶۸۶	۱/۵	۲/۹۵	۸/۷۳	۰/۴۷	۰/۹۹۶۷	۰/۷	۰/۹۴	۰/۸۸	۰/۱۹	RF
۰/۹۸۰۲	۱/۹	۲/۳۱	۵/۳۳	۰/۵۴	۰/۹۹۶۷	۱/۷	۱/۱۸	۱/۴	۰/۴	GBR
۰/۹۷۸۱	۲/۷	۲/۴۳	۵/۹۲	۰/۸۱	۰/۹۹۸۱	۱/۵	۰/۷۳	۰/۵۳	۰/۴۱	XGBR

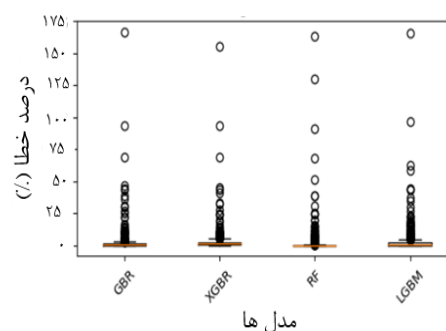
۹۸٪ از تغییرات $PM_{2.5}$ می‌تواند توسط متغیرهای مورد استفاده تعیین شود. به‌منظور مقایسه‌ی قدرت پیش‌بینی مدل‌ها، نمودارهای پراکندگی (شکل ۶) ترسیم شدند. همان‌طور که در این شکل مشاهده می‌شود، مدل LGBM دارای خطاهای بیشتری است، به‌ویژه در پیش‌بینی غلظت‌های بالاتر $PM_{2.5}$ ، نسبت به مدل‌های GBR، RF و XGBR دارد. پراکندگی نقاط داده نسبت به خط رگرسیون نیز در مدل‌های GBR و LGBM بیشتر از مدل‌های RF و XGB است. این موضوع نشان‌دهنده‌ی عملکرد بهتر مدل‌های RF و XGB در پیش‌بینی غلظت $PM_{2.5}$ نسبت به سایر مدل‌ها است.

بر اساس معیار MAE، مدل RF بهترین عملکرد را دارد، در حالی که مدل (LGBM ضعیف‌ترین عملکرد را نشان داده و کمترین MAE (۰/۹) را به‌دست آورده است. اگرچه بر اساس معیار MSE، مدل GBR عملکرد بسیار خوبی (۵/۳۳) داشته و LGBM بدترین عملکرد را با MSE برابر با (۱۰/۲) نشان می‌دهد. طبق معیار MAPE، بهترین عملکرد مربوط به مدل RF است، در حالی که مدل (XGBR ضعیف‌ترین عملکرد را دارد. علاوه بر این، مقدار R^2 در تمامی مدل‌ها بالای ۹۶٪ است و بر اساس این معیار، مدل GBR از قدرت بیشتری برخوردار بوده و LGBM بدترین عملکرد را دارد. مقدار R^2 برابر با ۰/۹۸۰۲ نشان می‌دهد که تقریباً



شکل ۶. نمودارهای پراکندگی غلظت $PM_{2.5}$ مشاهده‌شده و برآوردشده توسط مدل‌ها (الف) GBR (ب) LGBM (ج) RF (د) XGB

پراکندگی خطای مربوط به هر مدل در شکل ۷ نشان داده شده است. مدل GBR کمترین پراکندگی خطا را دارد و خطای آن عمدتاً در محدوده‌ی نزدیک به میانه متمرکز است، که بیانگر ثبات و دقت بالای این مدل در پیش‌بینی‌ها است. در مقابل، مدل RF بیشترین تعداد داده‌های پرت را نشان می‌دهد که حاکی از نوسانات بیشتر و کاهش ثبات در عملکرد آن است. وجود داده‌های پرت در تمامی مدل‌ها نشان‌دهنده‌ی چالش‌هایی در پیش‌بینی مقادیر خاص است، اما تمرکز خطاها در مدل GBR و LGBM در مقایسه با سایر مدل‌ها بیانگر توانایی بالاتر این دو مدل در ارائه‌ی پیش‌بینی‌های دقیق‌تر است.



شکل ۷. پراکندگی خطای پیش‌بینی مدل‌ها

بحث

غلظت بالای ذرات معلق در هوای مشهد را به عوامل مختلفی می‌توان نسبت داد، که مهم‌ترین آن‌ها ترافیک سنگین و ناهنجاری‌های حمل‌ونقل هستند (۲۶). استفاده‌ی بیش از حد از وسایل نقلیه و زیرساخت‌های ناکافی منجر به انتشار ذرات معلق در هوا می‌شود. وسایل نقلیه ذرات معلق تولید می‌کنند که می‌توانند از ذرات معمولی تا ذرات سمی متغیر باشند (۲۷، ۲۸). فعالیت‌های صنعتی نیز نقش مهمی در افزایش غلظت ذرات معلق در مشهد ایفا می‌کنند. انتشار گازها و ذرات از ماشین‌آلات، کارخانه‌ها و دیگر واحدهای صنعتی به آلودگی هوا از طریق رهاسازی ذرات معلق در جو کمک می‌کند (۲۹). استفاده از سوخت‌های فسیلی مانند زغال‌سنگ، نفت و گاز طبیعی یکی دیگر از عوامل افزایش غلظت ذرات معلق است. سوخت‌های فسیلی به‌طور گسترده به‌عنوان منابع انرژی و سوخت مصرف می‌شوند و در اثر احتراق، ذرات معلق را در هوا آزاد می‌کنند (۳۰). علاوه بر این، مشهد دارای اقلیمی خشک همراه با بادهای قوی است. برخی از

مطالعات نشان می‌دهند که قطرات باران می‌توانند گردوغبار موجود در هوا را جذب کنند. هنگامی که قطرات باران و گردوغبار تحت تأثیر نیروی جاذبه به زمین می‌ریزند، تعداد ذرات معلق در هوا کاهش می‌یابد. اثر کاهشی بارش باران بر غلظت‌های $PM_{2.5}$ در محدوده‌ای معین نیز در مطالعه ما مشاهده شد (۳۱). هم‌چنین باد نقش مهمی در انتقال، پراکنده‌سازی و انتشار $PM_{2.5}$ دارد. علاوه بر عوامل ذکرشده، نوسانات غلظت ذرات معلق در مشهد می‌تواند تحت تأثیر شرایط جوی و اقلیمی و هم‌چنین وارونگی دما در طول سال قرار گیرد. این عوامل به‌ویژه در فصل زمستان تأثیر قابل‌توجهی بر سطح آلودگی هوا دارند.

نتایج هم‌چنین نشان می‌دهد که در طول تابستان، دما نقش مهمی در سطوح $PM_{2.5}$ ایفا می‌کند. به‌طور کلی افزایش دما باعث افزایش ارتفاع لایه اختلاط می‌شود که به انتشار عمودی جو کمک می‌کند و در نتیجه فضای بیشتری برای پراکنده‌شدن آلاینده‌های سطحی فراهم می‌کند. هم‌چنین، اثر اختلاط آشفته ناشی از نیروهای حرارتی و دینامیکی سطح زیرین تأثیر مستقیمی بر مهاجرت و تغییر شکل آلاینده‌ها در لایه‌ی مختلط دارد (۳۲). هم‌چنین مطالعات نشان می‌دهند که پوشش گیاهی می‌تواند به‌طور مؤثری تعداد منابع $PM_{2.5}$ را با تثبیت خاک کاهش دهد و تأیید می‌کنند که افزایش سطح برگ، شاخه‌ها و سطوح ساقه‌ها، کارایی در مهار یا جذب $PM_{2.5}$ را در جنگل‌های مخروطی و پهن برگ افزایش می‌دهد و بدین ترتیب به‌طور مؤثری غلظت‌های $PM_{2.5}$ در هوا را کاهش می‌دهد. هم‌چنین، آلودگی هوای زمستانی در مشهد به یک مسئله‌ی جدی تبدیل شده است. این شرایط جوی، همراه با وارونگی دما، می‌تواند به تشکیل مه‌دود کمک کنند که با غلظت بالای ذرات معلق و آلاینده‌ها مشخص می‌شود. مه‌دود می‌تواند به‌طور قابل توجهی کیفیت هوا در مناطق اطراف را کاهش دهد و به افزایش سطح آلودگی هوا کمک کند.

بارش و AQI متغیرهای محیطی مرتبط نزدیکی هستند. اگرچه عوامل متعددی بر پیش‌بینی‌های AQI تأثیر می‌گذارند، بارش به‌عنوان یک عامل تعیین‌کننده‌ی مهم شناخته می‌شود. افزایش بارش منجر به کاهش AQI می‌شود و بالعکس، کاهش بارش نتیجه‌ی معکوس دارد (۳۳). تحلیل‌های آماری به‌وضوح نشان می‌دهند که AQI

حمل‌ونقل عمومی را افزایش دهیم، فضای سبز را گسترش دهیم، تعداد خودروهای قدیمی را کاهش دهیم و کیفیت سوخت را بهبود بخشیم. در مورد کلان‌شهرهایی مانند مشهد، تغییرات در مقادیر غلظت $PM_{2.5}$ به‌طور مستقیم با سلامت انسان مرتبط است؛ بنابراین، پیش‌بینی دقیق غلظت $PM_{2.5}$ بسیار حائز اهمیت است. علاوه بر این، داده‌های پیش‌بینی‌شده می‌توانند در سیاست‌های زیست‌محیطی برای نظارت و بهبود سلامت عمومی و کیفیت هوا مورد استفاده قرار گیرند.

نتیجه‌گیری

آلودگی هوا یکی از مهم‌ترین معضلاتی است که بشر امروز با آن مواجه شده است و پژوهشگران در سرتاسر جهان در تلاشند تا راه‌حلی برای آن بیابند. در این مطالعه، به‌دلیل نقش مهم $PM_{2.5}$ در آلودگی هوای مشهد، عملکرد مدل‌های غیرخطی یادگیری ماشین شامل RF، LGBM، XGBR و GBR برای پیش‌بینی دقیق غلظت‌های $PM_{2.5}$ در سال‌های ۱۳۹۵ تا ۱۴۰۱ مورد بررسی قرار گرفتند. پارامترهای ارزیابی دقت مدل‌های استفاده‌شده شامل R^2 ، MAPE، RMSE، MSE و MAE بودند که برای هر دو مجموعه داده‌های آموزشی و آزمایشی محاسبه شدند. نتایج تحلیل نهایی نشان داد که مدل GBR در پیش‌بینی غلظت‌های $PM_{2.5}$ بر اساس معیارهای R^2 ، MSE و RMSE عملکرد دقیق‌تری نسبت به سایر مدل‌های غیرخطی مورد مطالعه در این پژوهش داشت. هم‌چنین، بر اساس مدل غیرخطی RF با MAPE معادل ۱/۵٪ و MAE معادل ۰/۴۷، بهترین عملکرد را از خود نشان داد که نشان‌دهنده دقت بالای مدل در پیش‌بینی غلظت‌های $PM_{2.5}$ است. البته مقدار R^2 برای داده‌های تمامی مدل‌های غیرخطی داده بالای ۹۶٪ بود و نشان می‌دهد که مدل‌های پیش‌بینی $PM_{2.5}$ که در این مطالعه پیشنهاد شده‌اند، دقت بسیار بالایی دارند.

تشکر و قدردانی: این مقاله حاصل طرح پژوهشی با عنوان پیش‌بینی غلظت ذرات معلق PM_{10} و $PM_{2.5}$ در شهر مشهد با استفاده از مدل‌های یادگیری ماشین در مقطع

تمایل دارد در دوره‌های بارش شدید پایین‌تر باشد و بالعکس، در دوره‌های کاهش بارش بالاتر باشد. مشهد بیشتر بارش‌های خود را از اواسط پاییز تا اواسط بهار سال بعد تجربه می‌کند. این الگوی فصلی بارش ممکن است دلیل اصلی تغییرات مشاهده شده در AQI باشد. علاوه بر انتشار محلی آلاینده‌های هوا، غلظت و مدت زمان ورود PM نیز بسته به آرایش فشار جوی و شرایط آب‌وهوایی اطراف مشهد متفاوت است.

شایان ذکر است یافته‌های توزیع و تغییرات روزانه AQI نشان می‌دهد که در سال ۱۴۰۱، سطح آلودگی نسبت به سال‌های گذشته افزایش یافته است، با وجود این‌که در سال ۱۴۰۰ پایین‌تر از سال‌های قبل بود. این افزایش در سال ۱۴۰۱ ممکن است به فعالیت‌های صنعتی سنگین که پس از قرنطینه‌های ناشی از کووید-۱۹ دوباره آغاز شده‌اند، مرتبط باشد. از سرگیری فعالیت‌های صنعتی ممکن است به افزایش سطوح آلودگی هوا در سال ۱۴۰۱ کمک کرده باشد.

محدودیت‌ها و پیشنهادات پژوهش: از محدودیت‌های

این مطالعه می‌توان به عدم دسترسی به داده‌های بلندمدت و کامل‌تر از سایر آلاینده‌ها و برخی متغیرهای هواشناسی اشاره کرد که ممکن است بر دقت نهایی مدل‌ها تأثیرگذار باشند. علاوه بر این، تأثیر عوامل انسانی، صنعتی، و شرایط خاص محلی که به‌طور مستقیم بر غلظت $PM_{2.5}$ تأثیر می‌گذارند، در این مدل‌سازی به‌صورت مجزا لحاظ نشده است. هم‌چنین، مدل‌های یادگیری ماشین به داده‌های حجیم و متنوع برای پایداری و قابلیت اطمینان در پیش‌بینی‌ها نیاز دارند که ممکن است در تمامی مناطق به‌طور کامل در دسترس نباشند.

با توجه به مقادیر میانگین سالانه و ماهانه غلظت $PM_{2.5}$ اندازه‌گیری‌شده در مشهد و غلظت‌های بالاتر از استانداردهای EPA (۱۵) میکروگرم بر مترمکعب و WHO (۱۰) میکروگرم بر مترمکعب؛ توجه بیشتری به خطرات بهداشتی و اثرات منفی بر مناطق مسکونی مشابه منطقه مورد مطالعه ما ضروری است و نهادهای مربوطه باید اقدامات لازم را انجام دهند. در این راستا، می‌توانیم ناوگان

ملاحظات اخلاقی: این مقاله حاصل طرح پژوهشی با عنوان پیش‌بینی غلظت ذرات معلق PM₁₀ و PM_{2.5} در شهر مشهد با استفاده از مدل‌های یادگیری ماشین به کد اخلاق IR.MUMS.FHMPM.REC.1403.042 در مقطع کارشناسی‌ارشد است.

سهم نویسندگان: ۱. احمد مخدومی ۲. مریم سرخوش ۳. سمیه ضیائی در اجرای این پژوهش مشارکت داشته‌اند.

References

- Suleiman A, Tight MR, Quinn AD. Applying machine learning methods in managing urban concentrations of traffic-related particulate matter (PM₁₀ and PM_{2.5}). *Atmospheric Pollution Research*. 2019;10(1):134-44.
<https://doi.org/10.1016/j.apr.2018.07.001>
- He B, Xu H-M, Liu H-W, Zhang Y-F. Unique regulatory roles of ncRNAs changed by PM_{2.5} in human diseases. *Ecotoxicology and Environmental Safety*. 2023;255:114812.
<https://doi.org/10.1016/j.ecoenv.2023.114812> PMID:36963186
- McCarron A, Semple S, Braban CF, Swanson V, Gillespie C, Price HD. Public engagement with air quality data: using health behaviour change theory to support exposure-minimising behaviours. *Journal of Exposure Science & Environmental Epidemiology*. 2023;33(3):321-31.
<https://doi.org/10.1038/s41370-022-00449-2> PMID:35764891
PMCID:PMC10234807
- Yang H, Wang W, Li G. Prediction method of PM_{2.5} concentration based on decomposition and integration. *Measurement*. 2023;216:112954.
<https://doi.org/10.1016/j.measurement.2023.112954>
- Hardini M, Sunarjo RA, Asfi M, Riza Chakim MH, Ayu Sanjaya YP. Predicting Air Quality Index using Ensemble Machine Learning. *ADI Journal on Recent Innovation*. 2023;5(1Sp):78-86.
<https://doi.org/10.34306/ajri.v5i1Sp.981>

کارشناسی‌ارشد رشته مهندسی بهداشت محیط دانشگاه علوم پزشکی مشهد است که بدین‌وسیله از تمام افرادی که ما را در انجام این پژوهش یاری نمودند، تشکر و قدردانی میشود.

تعارض منافع: پژوهش حاضر دارای تعارض منافع نمی‌باشد.

حمایت مالی: از همکاری معاونت پژوهشی دانشگاه علوم پزشکی مشهد جهت انجام این پژوهش، تشکر و قدردانی می‌شود.

- Wu C-l, He H-d, Song R-f, Zhu X-h, Peng Z-r, Fu Q-y, et al. A hybrid deep learning model for regional O₃ and NO₂ concentrations prediction based on spatiotemporal dependencies in air quality monitoring network. *Environmental Pollution*. 2023;320:121075.
<https://doi.org/10.1016/j.envpol.2023.121075> PMID:36641063
- Doan QC, Chen C, He S, Zhang X. How urban air quality affects land values: Exploring non-linear and threshold mechanism using explainable artificial intelligence. *Journal of Cleaner Production*. 2024;434:140340.
<https://doi.org/10.1016/j.jclepro.2023.140340>
- Sun J, Gong J, Zhou J. Estimating hourly PM_{2.5} concentrations in Beijing with satellite aerosol optical depth and a random forest approach. *Science of The Total Environment*. 2021;762:144502.
<https://doi.org/10.1016/j.scitotenv.2020.144502> PMID:33360341
- Yang L, Xu H, Yu S. Estimating PM_{2.5} concentrations in Yangtze River Delta region of China using random forest model and the Top-of-Atmosphere reflectance. *Journal of Environmental Management*. 2020;272:111061.
<https://doi.org/10.1016/j.jenvman.2020.111061> PMID:32669259

10. Kim B-Y, Lim Y-K, Cha JW. Short-term prediction of particulate matter (PM₁₀ and PM_{2.5}) in Seoul, South Korea using tree-based machine learning algorithms. *Atmospheric Pollution Research*. 2022;13(10):101547. <https://doi.org/10.1016/j.apr.2022.101547>
11. Gardner MW, Dorling SR. Statistical surface ozone models: an improved methodology to account for non-linear behaviour. *Atmospheric Environment*. 2000;34(1):21-34. [https://doi.org/10.1016/S1352-2310\(99\)00359-3](https://doi.org/10.1016/S1352-2310(99)00359-3)
12. Berrocal VJ, Guan Y, Muyskens A, Wang H, Reich BJ, Mulholland JA, et al. A comparison of statistical and machine learning methods for creating national daily maps of ambient PM_{2.5} concentration. *Atmospheric Environment*. 2020;222:117130. <https://doi.org/10.1016/j.atmosenv.2019.117130> PMID:32863727 PMCid:PMC7451200
13. Ghahremanloo M, Choi Y, Sayeed A, Salman AK, Pan S, Amani M. Estimating daily high-resolution PM_{2.5} concentrations over Texas: Machine Learning approach. *Atmospheric Environment*. 2021;247:118209. <https://doi.org/10.1016/j.atmosenv.2021.118209>
14. Tang D, Liu D, Tang Y, Seyler BC, Deng X, Zhan Y. Comparison of GOCI and Himawari-8 aerosol optical depth for deriving full-coverage hourly PM_{2.5} across the Yangtze River Delta. *Atmospheric Environment*. 2019;217:116973. <https://doi.org/10.1016/j.atmosenv.2019.116973>
15. Wang W, Mao F, Du L, Pan Z, Gong W, Fang S. Deriving Hourly PM_{2.5} Concentrations from Himawari-8 AODs over Beijing-Tianjin-Hebei in China. *Remote Sensing [Internet]*. 2017; 9(8). <https://doi.org/10.3390/rs9080858>
16. Williams DR, Rast P. Back to the basics: Rethinking partial correlation network methodology. *British Journal of Mathematical and Statistical Psychology*. 2020;73(2):187-212. <https://doi.org/10.1111/bmsp.12173> PMID:31206621 PMCid:PMC8572131
17. Demir E, Bilgin MH, Karabulut G, Doker AC. The relationship between cryptocurrencies and COVID-19 pandemic. *Eurasian Economic Review*. 2020;10(3):349-60. <https://doi.org/10.1007/s40822-020-00154-1> PMCid:PMC7388435
18. Chen H, Li X, Feng Z, Wang L, Qin Y, Skibniewski MJ, et al. Shield attitude prediction based on Bayesian-LGBM machine learning. *Information Sciences*. 2023;632:105-29. <https://doi.org/10.1016/j.ins.2023.03.004>
19. Asselman A, Khaldi M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interactive Learning Environments*. 2023;31(6):3360-79. <https://doi.org/10.1080/10494820.2021.1928235>
20. Tran DA, Tsujimura M, Ha NT, Nguyen VT, Binh DV, Dang TD, et al. Evaluating the predictive power of different machine learning algorithms for groundwater salinity prediction of multi-layer coastal aquifers in the Mekong Delta, Vietnam. *Ecological Indicators*. 2021;127:107790. <https://doi.org/10.1016/j.ecolind.2021.107790>
21. T R P. A Comparative Study on Decision Tree and Random Forest Using R Tool. *IJARCCCE*. 2015:196-9. <https://doi.org/10.17148/IJARCCCE.2015.4142>
22. Otchere DA, Ganat TOA, Ojero JO, Tackie-Otoo BN, Taki MY. Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. *Journal of Petroleum Science and Engineering*. 2022;208:109244. <https://doi.org/10.1016/j.petrol.2021.109244>

23. Chicco D, Warrens M, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*. 2021;7:e623. <https://doi.org/10.7717/peerj-cs.623> PMID:34307865 PMCID:PMC8279135
24. Althoff D, Rodrigues LN. Goodness-of-fit criteria for hydrological models: Model calibration and performance assessment. *Journal of Hydrology*. 2021;600:126674. <https://doi.org/10.1016/j.jhydrol.2021.126674>
25. Liu X, Zou B, Feng H, Liu N, Zhang H. Anthropogenic factors of PM2.5 distributions in China's major urban agglomerations: A spatial-temporal analysis. *Journal of Cleaner Production*. 2020;264:121709. <https://doi.org/10.1016/j.jclepro.2020.121709>
26. Mohammadi M, Hatami M, Esmaeli R, Gohari S, Mohammadi M, khayami E. Relationships between Ambient Air Pollution, Meteorological Parameters and Respiratory Mortality in Mashhad, Iran: a Time Series Analysis. *Pollution*. 2022;8(4):1250-65.(Persian)
27. Harrison R. Airborne particulate matter. *Philosophical transactions Series A, Mathematical, physical, and engineering sciences*. 2020;378:20190319. <https://doi.org/10.1098/rsta.2019.0319> PMID:32981435 PMCID:PMC7536032
28. Aminiyan MM, Kalantzi O-I, Etesami H, Khamoshi SE, Hajiali Begloo R, Aminiyan FM. Occurrence and source apportionment of polycyclic aromatic hydrocarbons (PAHs) in dust of an emerging industrial city in Iran: implications for human health. *Environmental Science and Pollution Research*. 2021;28(44):63359-76. <https://doi.org/10.1007/s11356-021-14839-w> PMID:34231139
29. Maciejczyk P, Chen L-C, Thurston G. The Role of Fossil Fuel Combustion Metals in PM2.5 Air Pollution Health Associations. *Atmosphere [Internet]*. 2021; 12(9). <https://doi.org/10.3390/atmos12091086>
30. Bilal M, Hassan M, Tahir DBT, Iqbal MS, Shahid I. Understanding the role of atmospheric circulations and dispersion of air pollution associated with extreme smog events over South Asian megacity. *Environmental Monitoring and Assessment*. 2022;194(2):82. <https://doi.org/10.1007/s10661-021-09674-y> PMID:35013892
31. Pal S, Das P, Mandal I, Sarda R, Mahato S, Nguyen K-A, et al. Effects of lockdown due to COVID-19 outbreak on air quality and anthropogenic heat in an industrial belt of India. *Journal of Cleaner Production*. 2021;297:126674. <https://doi.org/10.1016/j.jclepro.2021.126674> PMID:34975233 PMCID:PMC8714179
32. Kanawade VP, Srivastava AK, Ram K, Asmi E, Vakkari V, Soni VK, et al. What caused severe air pollution episode of November 2016 in New Delhi? *Atmospheric Environment*. 2020;222:117125. <https://doi.org/10.1016/j.atmosenv.2019.117125>
33. Hamzeh NH, Karami S, Kaskaoutis DG, Tegen I, Moradi M, Opp C. Atmospheric Dynamics and Numerical Simulations of Six Frontal Dust Storms in the Middle East Region. *Atmosphere [Internet]*. 2021; 12(1). <https://doi.org/10.3390/atmos12010125>