نشریه علمی مهندسی فناوری اطلاعات مکانی

سال دهم، شماره چهارم، زمستان ۱۴۰۱

Vol.10, No.4, Winter 2023

۶۷ -۸۵

مقاله پژوهشی

DOR: 20.1001.1.20089635.1401.10.4.4.4

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]

[DOR: 20.1001.1.20089635.1401.10.4.4.4]

[DOI: 10.61186/jgit.10.4.67]

تهیه نقشه طبقه‌بندی و پیش‌بینی آلاینده $PM_{2.5}$ با استفاده از روش‌های یادگیری ماشین و استخراج قوانین انجمنی

محمدرضا حیدری [۱]، پرهام پهلوانی [۲*]، بهناز بیگدلی [۳]

۱- دانش آموخته کارشناسی ارشد سیستم‌های اطلاعات مکانی، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی دانشگاه تهران

۲- دانشیار دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، دانشکدگان فنی دانشگاه تهران

۳- دانشیار دانشکده مهندسی عمران، دانشگاه صنعتی شاهرود

تاریخ دریافت مقاله: ۱۴۰۱/۰۳/۰۱ تاریخ پذیرش مقاله: ۱۴۰۲/۰۲/۲۴

چکیده

آلودگی هوا ناشی از وجود آلاینده‌های گوناگون در هوا می‌باشد که بیش‌تر آن مربوط به وجود ذرات معلق هوا به خصوص آلاینده ذرات معلق کم‌تر از ۲٫۵ میکرون $(PM_{2.5})$ است. پیش‌بینی و شناسایی مکان‌هایی که تمرکز آلودگی در آنجا بیش‌تر است به مدیریت و برنامه‌ریزی صحیح کمک خواهد کرد. از این رو برای شناسایی این مکان‌ها نیاز به تهیه نقشه طبقه‌بندی و پیش‌بینی کلاس آلودگی ذرات معلق هوا می باشد. در این مقاله از روش‌های ماشین بردار پشتیبان، شبکه عصبی و درخت تصمیم به عنوان روش‌های یادگیری ماشین نظارت‌شده برای تهیه نقشه طبقه‌بندی و پیش‌بینی غلظت آلاینده $PM_{2.5}$ شهر تهران استفاده گردید. در ادامه نیز برای تحلیل تاثیر پارامترهای مکانی از روش استخراج قوانین انجمنی استفاده می‌گردد. روش ماشین بردار پشتیبان با دقت کلی ۸۷٫۳ درصد و میزان کاپا ۸۱٫۵ درصد به عنوان روش برتر انتخاب گردید. از این روش برای پیش‌بینی غلظت آلاینده تا ۷۲ ساعت آینده استفاده شد که این روش توانست با دقت کلی ۸۰٫۷ درصد و میزان کاپا ۷۱٫۱ درصد به پیش‌بینی کلاس آلاینده در روز سوم بپردازد. یافته‌ها حاکی از آن است که روش ماشین بردار پشتیبان مدل‌سازی و پیش‌بینی را با دقت بالاتری نسبت به بقیه روش‌ها انجام می‌دهد. هم‌چنین با توجه به تاثیر پارامترهای مکانی در قوانین انجمنی قوی‌تر، میزان آلاینده نزدیک‌ترین دو همسایگی، وضعیت توپوگرافی، دما، فشارهوا، میزان بارش، شدت وارونگی دما، رطوبت نسبی، سرعت‌باد، جهت باد، ماه‌سال، روزهفته، ساعت‌روز به ترتیب بیش‌ترین تاثیر را در تعیین کلاس آلاینده دارد.

کلید واژه‌ها : آلودگی هوا، آلاینده $PM_{2.5}$ ، پارامترهای مکانی، یادگیری ماشین نظارت شده، قوانین انجمنی.

* نویسنده مکاتبه کننده: دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی دانشگاه تهران.

تلفن: ۶۱۱۱۴۵۲۴

Email: pahlavani@ut.ac.ir

۱- مقدمه

آلودگی هوا در چند دهه اخیر تاثیرات فراوانی بر سلامت انسان‌ها داشته‌است که دلیل آن پیش‌رفت صنعت و شهرنشینی می‌باشد[۱]. آلودگی هوا مساله‌ای است که هم اکنون گریبان‌گیر کلان‌شهرها می‌باشد. طبق آمار سازمان بهداشت جهانی سالانه ۷ میلیون نفر در جهان جان خود را بر اثر آلودگی هوا از دست می‌دهند[۱] که از این آمار ۳۳ هزار نفر نیز در ایران جان خود را از دست می‌دهند. به صورت کلی آلودگی هوا ناشی از وجود آلاینده‌های دی اکسید نیتروژن، ازن، مونواکسید کربن، دی اکسید گوگرد، ذرات معلق کم‌تر از ۱۰ و ۲٫۵ میکرون می‌باشد[۲]. ذرات معلق هوا بیش از هر نوع آلاینده‌ها هوا مردم را تحت تأثیر قرار می‌دهند. اجزای اصلی تشکیل‌دهنده ذرات معلق هوا شامل سولفات‌ها، نیترات‌ها، آمونیوم، سدیم کلرید، کربن سیاه، ذرات معدنی و آب می‌شود. به عبارت دیگر ذرات معلق هوا یک مخلوط پیچیده از ذرات جامد و مایع متشکل از مواد آلی و معدنی معلق در هوا هستند. مهم‌ترین منابع ایجاد آلودگی این ذرات به ترتیب وسایل نقلیه به خصوص وسایل نقلیه گازوئیلی و انتشار توسط صنایع می‌باشند. چنانچه انسان به صورت مزمن با ذرات معلق هوا مواجه باشد به بیماری‌های قلبی ـ عروقی، ریوی و هم‌چنین سرطان ریه دچار می‌شود[۳]. ذرات معلق کم‌تر از ۲٫۵ میکرون با توجه به زمان در حال تغییر می‌باشند که در این پژوهش وابستگی به زمان(ساعات روز، روزهای هفته و ماه‌های سال) لحاظ شده‌است[۴]. البته عوامل مکانی چون پارامترهای هواشناسی، میزان شدت وارونگی هوا و میزان غلظت آلاینده برای نزدیکترین دو ایستگاه همسایه برای هر نقطه جغرافیایی، در طبقه‌بندی و پیش‌بینی کلاس آلاینده ذرات معلق کم‌تر از ۲٫۵ میکرون تاثیرگذار می‌باشد. هم‌چنین در این پژوهش با توجه به اینکه حس‌گرهای موجود برای استخراج داده‌های مورد استفاده در فضای باز هستند و اندازه‌گیری‌ها نیز در فضای باز انجام می‌شود، در نتیجه این تحقیق به صورت خارج از ساختمان[1] می‌باشد.

تاکنون تحقیقات بسیاری در زمینه طبقه‌بندی و پیش‌بینی آلودگی ذرات معلق هوا و استخراج قوانین انجمنی صورت گرفته است.

از روش شبکه عصبی بدلیل انجام پردازش‌های پیچیده به طور وسیعی در پیش‌بینی سری‌های زمانی استفاده شده است[۵ و ۶]. البته شبکه عصبی با محدودیت‌هایی همچون بیش‌برازش[2]، بهینه محلی و پردازش‌های زمانبر در مواجهه با حجم بالای داده‌های ورودی روبه‌رو می‌باشد[۷, ۸]. فنگ و همکاران از روش شبکه عصبی برای پیش‌بینی آلودگی ذرات معلق کم‌تر از ۲٫۵ میکرون بر مبنای مدل جغرافیایی و تبدیل موجک استفاده کردند. آن‌ها در این تحقیق برای بهبود پیش‌بینی آلودگی ذرات معلق کم‌تر از ۲٫۵ میکرون عوامل موثری همچون بیشینه و کمینه دما ساعتی، رطوبت، جهت باد در دو سوی محورها، روز از سال، روز از هفته و ۱۰ حالت مختلف برای آب و هوا نیز در نظر گرفتند[۹]. در تحقیق دیگری الانگاسینگ و همکاران با در نظر گرفتن پارامترهای هواشناسی شامل سرعت باد، جهت باد، شدت تابش، دما، رطوبت نسبی، ساعت در روز، روز در هفته و ماه در سال با روش شبکه عصبی به مدل‌سازی غلظت آلاینده دی اکسید نیتروژن پرداختند[۴]. در پژوهشی دیگر، رستمی فصیح و همکاران به پیش‌بینی شاخص کیفیت هوا برمبنای متغیرهای هواشناسی و مولفه‌های خودهمبسته با استفاده از شبکه عصبی مصنوعی پرداختند و مشخص‌شد که در بین متغیرهای هواشناسی، میزان دید افقی و میزان بارندگی تأثیر بیشتری بر مقدار شاخص داشتند[۱۰]. هم‌چنین چاکر و همکاران به ارزیابی عملکرد شبکه عصبی در پیش‌بینی غلظت آلاینده‌های هوا در نیکوزیا پرداختند. آن‌ها در تحقیق خود از پارامترهای غلظت آلاینده، فشار هوا، سرعت باد،

[1] *Outdoor*
[2] *Overfitting*

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

رطوبت نسبی و دما بین سالهای ۲۰۱۲ تا ۲۰۱۵به عنوان ورودی مدل شبکه عصبی استفاده کردند. نتایج تحقیق آن‌ها نشان داد که شبکه عصبی عملکرد بسیار خوبی داشته‌است. هم‌چنین برای آن‌ها مشخص‌شد که می‌توانند از مدل‌های انتشار رو به عقب شبکه عصبی ساخته‌شده برای پیش بینی آلاینده‌ها استفاده کنند[۱۱]. در تحقیقی چاعه و همکاران برای پیش‌بینی آنی ذرات معلق کم‌تر از ۱۰ و ۲٫۵ میکرون از نوعی مدل شبکه عصبی استفاده کردند و آن را بر داده‌های کیفیت هوا و هواشناسی درون‌یابی شده اعمال کردند و در نهایت عملکرد پیش‌بینی با شاخص ارزیابی R^2 بالاتر از ۰٫۹۷ و خطای کم‌ترین مربعات ریشه تقریبا ۱۶٪ انحراف معیار بود[۱۲]. ماشین بردار پشتیبان از دیگر روش‌های مطرح در طبقه‌بندی بوده که به عنوان روشی که قابلیت‌های محاسباتی بالا دارد و در حل بسیاری از مسائل مورد استفاده قرار می‌گیرد شناخته می‌شود[۱۳]. قائمی و همکاران برای پیش‌بینی مکانی-زمانی آلودگی هوای شهر تهران از روش ماشین بردار پشتیبان استفاده نمودند آن‌ها در این تحقیق بر اساس داده‌های سری زمانی غلظت آلاینده، شرایط آب و هوایی، پارامترهای جغرافیایی همچون ترافیک، انحنای سطح و ارتفاع محلی پیش‌بینی صورت گرفت که ارزیابی خروجی‌ها نشان داد مدل دارای دقت کلی ۰٫۷۱، خطای ریشه کم‌ترین مربعات ۰٫۵۴، و شاخص R^2 ۰٫۸۱ می‌باشد[۱۴]. لیونگ و همکاران به پیش‌بینی شاخص کیفیت هوا با استفاده از روش ماشین بردار پشتیبان پرداختند[۱۳]. در پژوهش دیگری لو و همکاران از روش ماشین بردار پشتیبان و شبکه تابع پایه شعاعی کلاسیک[1] برای پیش‌بینی پارامترهای آلاینده هوا استفاده کردند[۱۵]. هم‌چنین دلاور و همکاران برای پیش‌بینی آلودگی هوا از روشهای یادگیری ماشین همچون ماشین بردار پشتیبان و شبکه عصبی مصنوعی استفاده کردند و درنهایت به این نتیجه رسیدند که روش ماشین بردار پشتیبان نسبت به روش شبکه عصبی مصنوعی دقت بالاتری دارد اما زمان پردازش بالاتری را طلب می کند[۱۶].

درخت تصمیم نیز به دلیل سادگی، نمایش بهتر و فهم ساده‌تر مورد استفاده قرار می‌گیرد. در این راستا در پژوهشی جمال و همکاران برای پیش‌بینی شاخص کیفیت هوا بر مبنای پارامترهای هواشناسی از روش های شبکه عصبی و درخت تصمیم استفاده کردند که در انتها مشخص‌شد که روش شبکه عصبی دقت بالایی در پیش‌بینی شاخص کیفیت هوا دارد[۱۷]. در پژوهش دیگری هیلال و همکاران برای پیش بینی شاخص کیفیت هوا از روش های یادگیری ماشین نظیر درخت تصمیم و درخت گرادیان تقویت شده استفاده کردند و به این نتیجه رسیدند که روش درخت تصمیم با دقت ۹۳٫۷۲ درصد به پیش بینی شاخص کیفیت هوا می پردازد[۱۸]. پس از مدل‌سازی و پیش‌بینی کلاس آلاینده، هدف استخراج قوانین می‌باشد. قوانین انجمنی به کشف روابط ارتباطی، الگوهای مکرر یا همبستگی بین مجموعه داده‌ها یا عناصر در پایگاه‌های داده می پردازد[۱۹]. از این رو پایوس و همکاران در تحقیقی به کاربرد استخراج قوانین انجمنی برای بیماران تنفسی با استفاده از پایگاه داده آلودگی هوا پرداخت و موارد مناسبی را شناسایی کرد[۱۹]. در تحقیق دیگری شفیع‌زاده و همکاران با استفاده از قوانین تصمیم استخراج شده آلودگی هوا به پیش‌بینی آلودگی هوا پرداختند[۲۰].

در این پژوهش برای پیش‌بینی و طبقه‌بندی آلودگی ذرات معلق کم‌تر از ۲٫۵ میکرون از پارامترهای هواشناسی، وضعیت توپوگرافی، میزان شدت وارونگی و میزان غلظت آلاینده در نزدیکترین دو همسایگی به عنوان پارامترهای ورودی استفاده گردیده است و پارامتر خروجی کلاس آلودگی آلاینده در نظر گرفته شده است. پیش‌بینی و طبقه‌بندی با استفاده از متداول ترین روش‌های یادگیری ماشین نظارت شده شامل درخت تصمیم، شبکه عصبی و ماشین بردار پشتیبان

[1] *RBF*

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

انجام گرفته است. سپس از روشی که دارای دقت بالاتری بود، قوی‌ترین قوانین انجمنی استخراج گردید و با توجه به قوی‌تر بودن هر قانون به تحلیل نتایج و تعیین روابط بین پارامترهای مکانی با نوع کلاس آلودگی ذرات معلق هوا پرداخته شد.

اهداف انجام شده در این پژوهش که آن را از پژوهش‌های پیشین متمایز می‌نماید به شرح ذیل می‌باشد:

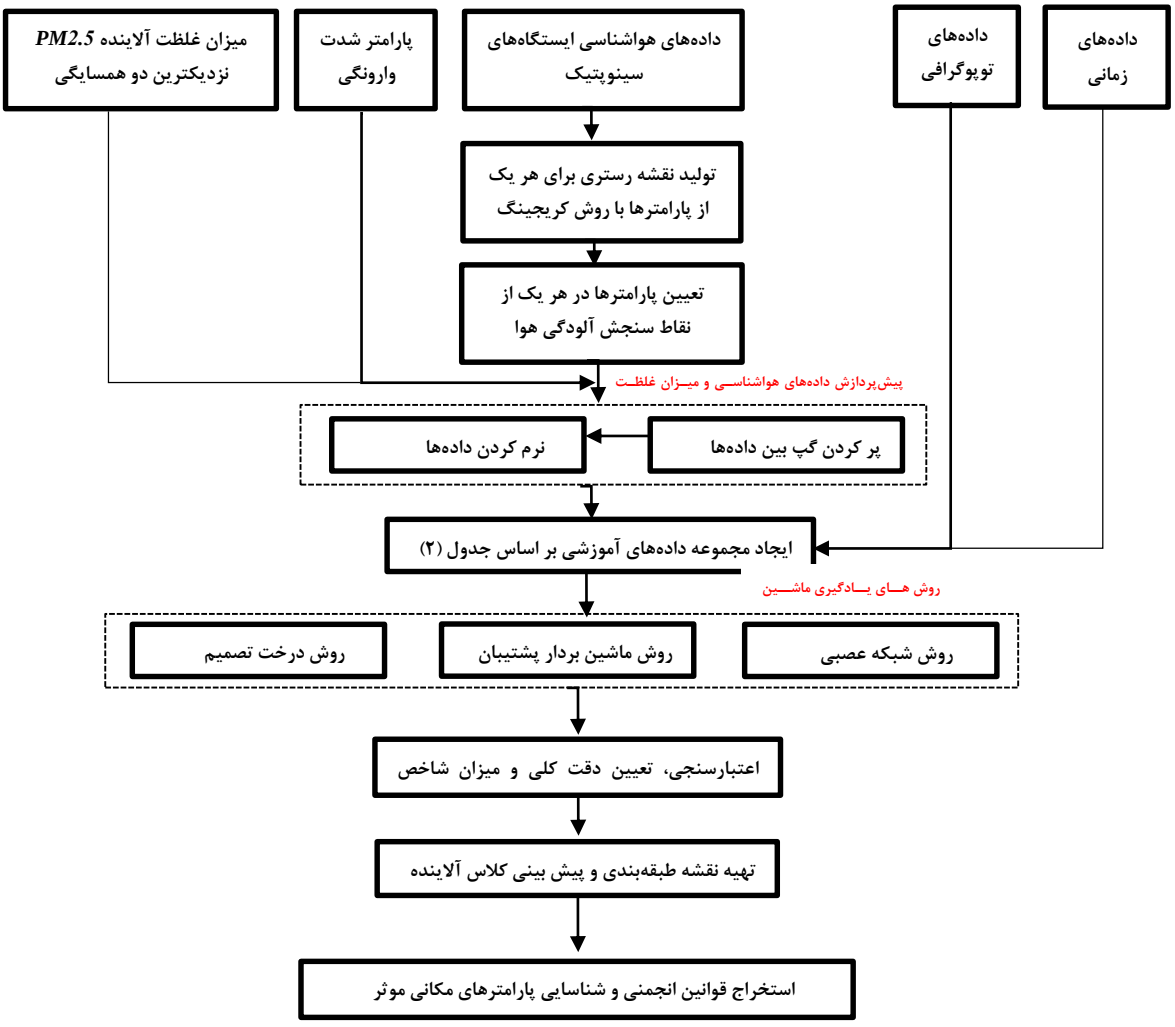
۱- بررسی الگوریتم‌های کلاسه‌بندی مبتنی بر یادگیری ماشین بر روی داده‌های آلودگی ذرات معلق کم‌تر از ۲٫۵ میکرون برای شناسایی پارامترهای مکانی موثر

۲- استخراج دانش از طریق تولید قوانین انجمنی موثر میان پارامترهای ورودی و شاخص کیفیت هوا به همراه تحلیل قوانین

در ادامه، در بخش ۲ به بررسی روش تحقیق پیشنهادی، الگوریتم و آماده سازی داده‌ها پرداخته شده است. بخش ۳ به بررسی منطقه مورد مطالعه، پیاده سازی و ارزیابی نتایج پرداخته شده است و در نهایت بخش ۴ به نتیجه‌گیری می‌پردازد.

۲- روش پیشنهادی

روند کلی پیشنهادی پژوهش حاضر در شکل(۱) ارائه شده است.



شکل۱: فلوچارت مراحل انجام پژوهش

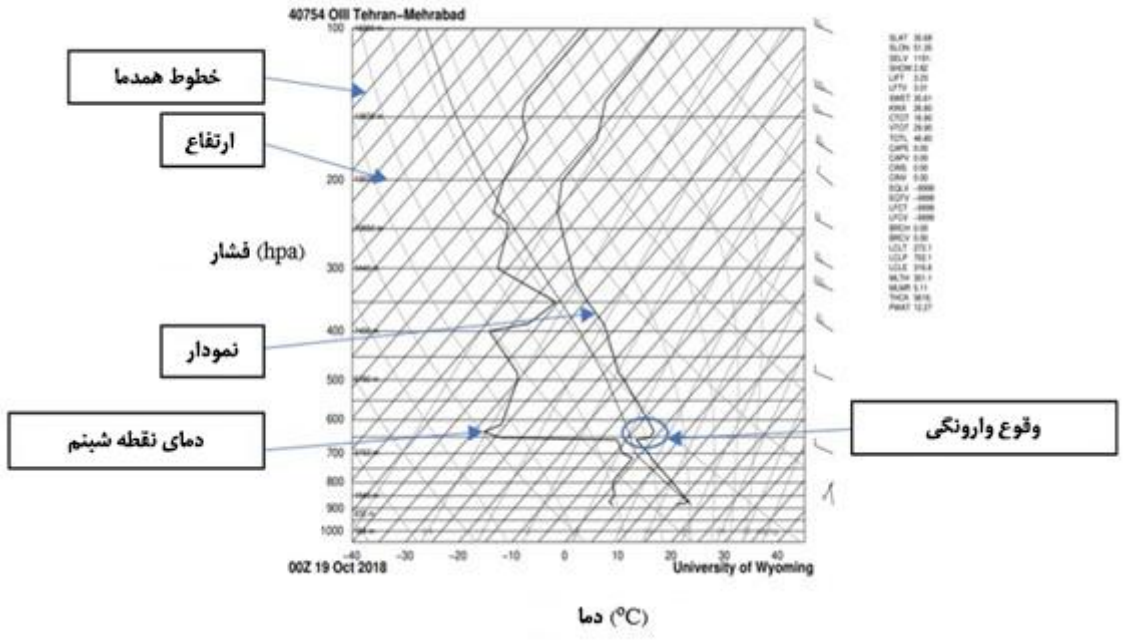
[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

در این تحقیق همانطورکه در شکل (۱) نمایش داده شده است ابتدا نیاز به انجام پیش‌پردازش و آماده‌سازی داده‌های ورودی شامل داده‌های زمانی(ماه، روز و ساعت)، داده‌های توپوگرافی(طول جغرافیایی، عرض جغرافیایی و ارتفاع)، پارامترهای هواشناسی (سرعت باد،جهت باد، رطوبت نسبی، فشار هوا، دما، میزان بارش)، پارامتر شدت وارونگی دما و میزان غلظت آلاینده نزدیکترین دو همسایگی می‌باشد، به این‌صورت که داده‌هایی که تعداد نقاط ایستگاه سنجش آلودگی هوا ۲۸ می‌باشد که برخی از ایستگاه‌ها به خاطر گپ-های موجود ناشی از ایرادات فنی در دستگاه های ثبت یا حس‌گرها غلظت $PM_{2,5}$ را در ساعتی ثبت نمی-کنند و در اختیار نمی‌باشد که آن‌ها با استفاده از روش کریجینگ، تعیین‌شود. در این پژوهش عوامل موثر علاوه بر فاکتور دسترس پذیر بودن با استفاده از مطالعه و جمع‌بندی پژوهش‌های پیشین انجام شده در این راستا انتخاب شده‌اند و به نوعی این پژوهش تکمیل شده‌ی پژوهش‌های پیشین می‌باشد.

یکی از پارامترهای ورودی مدل در این پژوهش شدت وارونگی دما است. وارونگی دما (*Temperature inversion*) در روزهای آلوده به یک میزان و شدت نیست بلکه در برخی روزها از لحاظ شدت و عملکرد متفاوت می‌باشند. به همین منظور از پارامتری به نام شدت وارونگی برای بیان میزان وارونگی دما استفاده می‌شود که از رابطه(۱) قابل محاسبه می‌باشد[۲۱]. پارامترهای موجود در رابطه(۱) با استفاده از نمودار *skew-T* موجود در سایت دانشگاه وایومینگ[۲۲] استخراج گردید که در شکل (۲) نمونه‌ای از آن آورده شده است.

$$dI = \frac{\Delta T}{\Delta Z} \qquad \text{رابطه(۱)}$$

در رابطه(۱)، ΔTاختلاف دمای وارونگی و ΔZاختلاف ارتفاع لایه وارونگی می‌باشد.



شکل۲: نمایش نمونه‌ای از یک نمودار *skew-T* [۲۱]

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

در ادامه با استفاده از جدول(۱) که آستانه آلاینده ذرات معلق کم‌تر از ۲٫۵ میکرون آورده شده، مقادیر غلظت آلاینده با توجه به محدوده آن به کلاس‌های طبقه‌بندی آلودگی استاندارد (پاک، سالم، ناسالم برای گروه‌های حساس، ناسالم، بسیار ناسالم، خطرناک) تبدیل می‌شود. اطلاعات مربوط به جدول(۱) از سایت آژانس حفاظت از محیط زیست ایالات متحده آمریکا اخذ شده است[۲۳]. برای تهیه و تعیین دقت مدل و هم‌چنین تهیه مدل پیش‌بینی کلاس آلاینده، ساختار داده‌ها در این تحقیق به صورت جدول (۲) تنظیم و وارد مدل گردید.

جدول۱: آستانه آلاینده ذرات معلق کم‌تر از ۲٫۵ میکرون[۲۳]

دسته‌بندی شاخص کیفیت هوا	بازه شاخص کیفیت هوا		نقاط شکست آلاینده $PM_{2.5}$	
	کم‌ترین	بیش‌ترین	کم‌ترین	بیش‌ترین
پاک	۰	۵۰	۰	۱۲
سالم	۵۱	۱۰۰	۱۲٫۱	۳۵٫۴
ناسالم برای گروه حساس	۱۰۱	۱۵۰	۳۵٫۵	۵۵٫۴
ناسالم	۱۵۱	۲۰۰	۵۵٫۵	۱۵۰٫۴
بسیار ناسالم	۲۰۱	۳۰۰	۱۵۰٫۵	۲۵۰٫۴
خطرناک	۳۰۱	۴۰۰	۲۵۰٫۵	۹۹۹۹٫۹

جدول۲: ساختار داده‌های ورودی و خروجی مدل در روش‌ها برای پیش بینی

داده‌های ورودی												خروجی
moy^{1}_{t-h}	dow^{2}_{t-h}	hod^{3}_{t-h}	xyz	wd^{4}_{t-h}	ws^{5}_{t-h}	p^{6}_{t-h}	t^{7}_{t-h}	u^{8}_{t-h}	r^{9}_{t-h}	dI^{10}_{t-h}	$2\text{-}NN_{t-h}$	$class_t$

در جدول (۲) ساختار پارامترهای ورودی و خروجی برای مدل‌سازی و پیش بینی به اختصار در پانویس آورده‌شده‌است. پارامترهای ورودی به ترتیب از چپ به راست، شامل ماه، هفته، روز، وضعیت توپوگرافی، جهت باد، سرعت باد، فشار هوا، دما، رطوبت، میزان بارش، میزان شدت وارونگی، میزان غلظت نزدیکترین دو همسایگی در زمان $t\text{-}h$ می باشند و پارامتر خروجی در جدول (۲) نیز کلاس آلودگی در زمان t می‌باشد. در جدول (۲) منظور از h تاخیر زمانی(ساعت) می‌باشد که برای تعیین دقت مدل از $h=0$ استفاده می‌شود. همچنین در برخی زمان‌ها داده‌های مربوط به نزدیکترین دوهمسایگی، داده‌های هواشناسی و شدت

[1] *Month of year*
[2] *Day of week*
[3] *Hour of day*
[4] *Wind direction*
[5] *Wind speed*
[6] *Air pressure*
[7] *Temperature*
[8] *Humidity*
[9] *Rain*
[10] *Inversion intensity*

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

وارونگی هوا موجود نمی‌باشد (توسط سنجنده برداشت نشده است)؛ به همین منظور برای ورود به مدل، باید گپ میان داده‌ها پر گردد. برای پر کردن این گپ روش‌های مختلفی همچون سری فوریه و اسپیلاین وجود دارد. تحقیقات بسیاری نشان دادند که روش اسپیلاین برای پر کردن گپ میان داده‌ها مناسب است[۲۴]. در این پژوهش نیز از روش تابع اسپیلاین استفاده می‌شود. در مرحله‌ی بعد به دلیل نوسان بالای پارامترهای هواشناسی،شدت وارونگی و میزان غلظت آلاینده ذرات معلق کم‌تر از ۲٫۵ میکرون ایستگاه‌ها نیاز به نرم کردن آن‌ها وجود دارد. برای این منظور در این پژوهش از روش ساویتزکی گولای (*Savitzky-golay*) استفاده گردید[۲۵]. در جدول(۳) نمونه ای از داده‌های ورودی و خروجی مدل آورده شده است.

جدول۳: نمونه داده‌های ورودی و خروجی مدل‌سازی

moy	*dow*	*hod*	*x*	*y*	*h*	*wd*	*ws*	*p*	*t*	*u*	*r*	*dI*	*pm2nn*	*class*
۱	۶	۶	۵۳۴۹۰۵	۳۹۵۲۶۸۶	۱۲۸۷	۲٫۱۶۸	۹۷٫۱	۱٫۸۷۱	۶٫۱۶	۳٫۴۱	۰	۰	۴۰	سالم
۱	۱	۰	۵۴۳۷۴۷	۳۹۶۱۴۱۶	۱۵۱۹	۱۳۵	۳۳٫۲	۸٫۸۴۸	۱٫۱۷	۳٫۶۲	۴٫۰	۰	۱۹	پاک
۳	۳	۱۲	۵۲۹۹۸۳	۳۹۵۷۶۳۵	۱۴۷۰	۲٫۱۸۹	۷۶٫۱	۸٫۸۶۷	۷٫۱۳	۵٫۴۴	۰	۰۳٫۰	۶۸	ناسالم گروه حساس
۲	۱	۰	۵۳۸۷۳۹	۳۹۵۹۴۳۶	۱۵۰۰	۶۲٫۸۷	۴٫۰	۲٫۸۵۵	۰۶٫۷	۶٫۷۵	۰	۰۴٫۰	۱۹	ناسالم

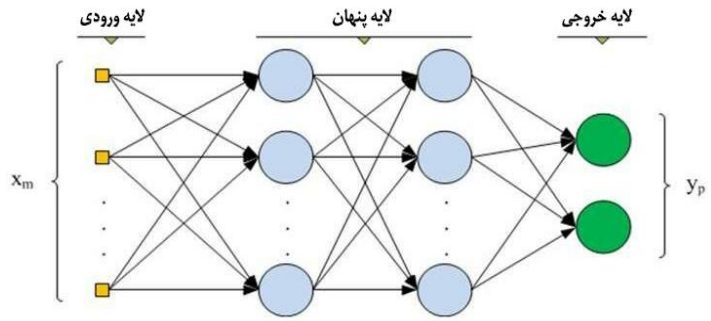
در بحث آلودگی هوای ذرات معلق کم‌تر از ۲٫۵ میکرون استفاده از روش‌های یادگیری ماشین نظارت شده در پژوهش‌های فراوانی [۱۶] توصیه شده است که از این بین، روش‌های درخت مبنا، شبکه های عصبی و ماشین بردار پشتیبان بیش‌ترین کاربرد و تکرار را در این پژوهش ها داشته‌است. سپس به مقایسه دقت روش‌ها پرداخته می‌شود. هم‌چنین در این پژوهش برای بیان دقت، از دقت کلی و شاخص کاپا استفاده شده است. در ادامه نیز نقشه پیش‌بینی برای هر روش تولید می‌گردد. در انتها نیز قوانین انجمنی استخراج و از این قوانین، پارامترهای موثر در تعیین کلاس آلودگی آلاینده شناسایی می‌شود.

۲-۱- روش شبکه عصبی پرسپترون چندلایه[1]

پرسپترون چندلایه دسته‌ای از شبکه‌های عصبی پیشخور است. این شبکه با پردازش روی داده‌ها، دانش یا قانون موجود نهفته در ورای داده‌ها را به ساختار شبکه منتقل می‌کنند که به این عمل یادگیری می‌گویند. در شکل(۳) معماری یک شبکه عصبی پرسپترون چندلایه نمایش داده می‌شود که در سمت چپ لایه ورودی، در وسط لایه یا لایه های پنهان قراردارد و در سمت چپ لایه خروجی مشاهده می‌گردد. در لایه ورودی، پارامترهای ورودی برنامه وارد می‌شود و با انجام محاسبات، یک خروجی تولید می‌شود. با توجه به تفاوت بین خروجی تولید شده و خروجی واقعی، خطا در شبکه پخش می‌شود و این روند تا رسیدن به سطح مطلوب یادگیری شبکه ادامه می یابد. بعد از مشخص‌شدن اینکه شبکه به اندازه کافی آموخته است، با دادن ورودی‌های تست به شبکه، می‌توان خروجی‌های پیش‌بینی را تولید کرد [۲۶].

[1] *Multi-Layer Perceptron Neural network*

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]



شکل۳: معماری یک شبکه عصبی پرسپترون چندلایه

فرآیند شبکه عصبی که برای مدل داده استفاده می‌شود شامل دو مرحله مختلف است:

- انتشار رو به جلو[1]

در این حالت ورودی به صورت رو به جلو از طریق لایه پنهان به لایه خروجی منتقل می‌شود. این روش مقادیر خروجی پرسپترون چند لایه را با استفاده از یک تابع فعالسازی[2] ترسیم می‌کند. تابع فعالسازی، هم‌چنین به عنوان تابع انتقال نیز شناخته می‌شود، که غیرخطی بودن را به شبکه معرفی می‌کند زیرا بدون تعریف غیرخطی بودن تابع، شبکه عصبی در همگرایی ناکام خواهد بود. تابع غیرخطی به جز گره‌های ورودی به هر لایه دیگر معرفی می‌شود. در این پژوهش از تابع تانژانت هایپربولیک استفاده می‌شود که خروجی این تابع فعالسازی عددی بین $[-1,1]$ می‌باشد. یکی از توابع فعالسازی، تابع تانژانت هایپربولیک است که در رابطه (۲) آمده است[۲۷].

$$\tanh(x)=\frac{e^{2x}-1}{e^{2x}+1} \qquad \text{رابطه(۲)}$$

اگر ورودی را با x_i نمایش داده و وزن ورودی‌ها با w مشخص‌شده باشد و مقدار بایاس با b نمایش داده شود ورودی خالص n با رابطه (۳) بیان می‌شود[۲۸]:

$$n=\sum_{i=1}^{R}\left(w.x_i+b\right) \qquad \text{رابطه(۳)}$$

- به‌روز رسانی وزن و بایاس

انتشار رو به عقب[3] فرایند محاسبه تابع خطا و به روزرسانی وزنهای سیناپسی[4] گره‌های ورودی برای کاهش خطای تابع می‌باشد. در رابطه(۴) و (۵) با تغییر نرخ یادگیری در هر مرحله وزن و بایاس اصلاح می‌شود تا جایی که بهینه‌ترین وزن و بایاس به دست بیاید[۲۹].

$$w_i \leftarrow w_i-\eta\frac{\Delta E}{\Delta w_i} \qquad \text{رابطه(۴)}$$

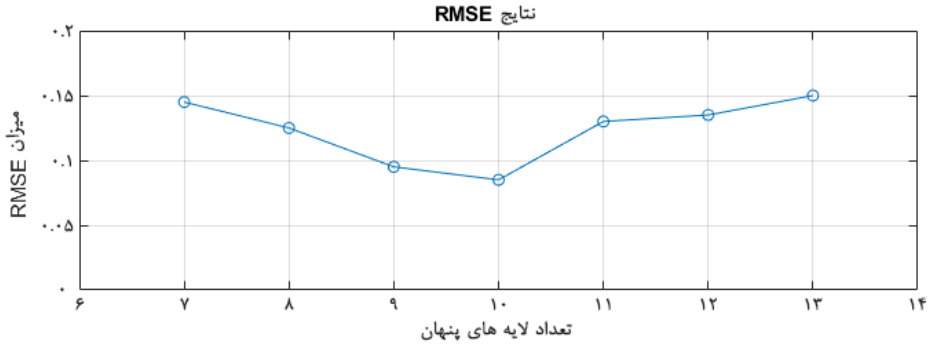
$$b_i \leftarrow b_i-\eta\Delta E \qquad \text{رابطه(۵)}$$

در روابط (۴) و (۵)، w_i وزن هر مرحله و η نرخ یادگیری و E خطای تابع می‌باشد.

در این تحقیق تابع فعالسازی تانژانت هایپربولیک، تعداد ورودی‌ها ۱۴ پارامتر، تعداد لایه‌ها برابر با ۳ و تعداد نورون‌ها در لایه مخفی با توجه به شکل (۴) در مقدار ۱۰ برابر با کم‌ترین میزان خطای میانگین مربعات ریشه[5] می‌باشد که در نتیجه تعداد نورون ها در لایه مخفی برابر با ۱۰ در نظر گرفته شد. همچنین از روش لونبرگ-مارکوارت[6] برای آموزش شبکه و به‌روز رسانی وزنها و بایاس‌ها استفاده گردید[۲۹].

[1] *Feed-forward propagation*
[2] *Activation function*
[3] *Backward propagation*
[4] *Synapse*
[5] *RMSE*
[6] *Levenberg-Marquardt*

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]



شکل۴: رابطه مقادیر خطای میانگین مربعات ریشه با تعداد نورون‌ها در لایه مخفی

۲-۲- روش ماشین بردار پشتیبان[1]

ماشین بردار پشتیبان اولین بار توسط وپنیک به عنوان یک طبقه‌بندی کننده باینری مطرح گردید[۳۰][۳۱]. در حالتی، ماشین بردار پشتیبان خطی فرض می‌شود که مجموعه‌ای از نمونه‌های آموزشی تفکیک‌پذیر وجود دارند که می‌توان آن‌ها را با y_i برچسب زد. در این حالت نمونه‌ها به شکل زوج مرتب‌های (x_i, y_i) بیان می‌شوند که در آن x_i متعلق به R^n و y_i دارای مقادیر [۱-و۱] می‌باشد[۳۲]. در حالتی که دو کلاس به صورت خطی قابل جداسازی باشند، ماشین بردار پشتیبان ابرصفحه‌ای را ایجاد می‌کند که کلاس‌ها را به گونه‌ای از هم جدا کند تا فاصله میان نزدیکترین نمونه های دو کلاس، در راستای عمود بر مرز تصمیم‌گیری، بیشینه شود. در حالت خطی، طبقه‌بندی با استفاده از رابطه (۶) حاصل می‌شود[۳۳]:

$$f(x) = sign\{w.x_i + b\}$$ رابطه(۶)

به نحوی که رابطه (۷) برقرار باشد:

$$\sum_{i=1}^{l} \alpha_i y_i = 0, \alpha_i \geq 0$$ رابطه(۷)

در روابط(۶) و (۷)، α_iضرایب لاگرانژ، wبردار نرمال عمود بر صفحه و bفاصله از مبدا مختصات می‌باشد. بردار نرمال عمود بر ابرصفحه و فاصله از مبدا مختصات از روابط (۸) و (۹) به‌دست می‌آیند[۳۴]:

$$w = \sum_{i=1}^{l} \alpha_i x_i y_i$$ رابطه(۸)

$$b = y_i - \sum_{i=1}^{l} y_i \alpha_i x_i x_e$$ رابطه(۹)

تنها نمونه‌هایی که ضرایب لاگرانژ آن‌ها مخالف صفر است در تشکیل ابرصفحه مشارکت دارند. این داده‌ها که نزدیکترین نمونه‌ها به ابرصفحه هستند، بردارهای پشتیبان نامیده می‌شوند. سایر نمونه‌ها در شکل‌گیری ابرصفحه تاثیری ندارند[۳۳]. روابط (۸و۹) برای طبقه-بندی در حالت خطی کاربرد دارند، در حالیکه بسیاری از پدیده‌های طبیعی رفتار خطی ندارند در این حالت نیاز است داده‌ها به فضایی با ابعاد بالاتر منتقل شود در چنین شرایطی باید از تابع کرنل[2] استفاده شود. در فضای جدید داده‌ها به صورت خطی قابل جداسازی هستند. برای حالت خطی تابع به صورت رابطه (۱۰) تعریف شده و ضرایب لاگرانژ و فاصله از مبدا مانند حالت خطی محاسبه می‌شود[۳۳].

$$f(x) = sign\left\{\sum_{i=1}^{l} \alpha_i y_i k(x_i, x_e) + b\right\}$$ رابطه(۱۰)

که در رابطه (۱۰)، $k(x_i, x_e)$ از رابطه(۱۱) قابل محاسبه می‌باشد[۳۲]:

$$k(x_i, x_e) = e^{(-\gamma ||x_i - x_e||)^2}, \gamma > 0$$ رابطه(۱۱)

به نحوی که رابطه (۱۲) برقرار باشد.

$$\sum_{i=1}^{l} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C$$ رابطه(۱۲)

[1] *Support vector machine*

[2] *Kernel*

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

که در روابط(۱۰)، (۱۱) و (۱۲)، $k(x_i, x_e)$ تابع کرنل و C عدد ثابتی است که مشخص می‌کند چه میزان خطا در طبقه‌بندی قابل صرف نظر کردن است. در این پژوهش نوع تابع کرنل RBF ، مقدار C برابر با ۱۰۰۰ و مقدار γ برابر با ۰٫۰۱درنظرگرفته می‌شود که این مقادیر با استفاده از روش جستجوی گریدی[1]بدست آمد[۳۴].

۲-۳- روش درخت تصمیم[2]

درخت تصمیم یکی از متداول ترین روش‌های یادگیری ماشین نظارت شده می‌باشد. علاوه بر این، می‌توان برای کارهای طبقه‌بندی و پیش‌بینی استفاده کرد. ایده اصلی این روش استفاده از نمایش درخت برای حل مسئله پیشنهادی است که در آن هر گره برگ با یک برچسب کلاس خاص مطابقت دارد[۳۵]. در این پژوهش درخت با عمق ۴۵ در نظر گرفته می‌شود تا بهترین دقت حاصل شود.

۲-۴- روش های ارزیابی نتایج

شاخص‌های مختلفی برای ارزیابی نتایج وجود دارد که در این تحقیق از دقت کلی و شاخص کاپا استفاده می‌گردد.

شاخص‌های آماری نظیر دقت کلی و شاخص کاپا، بر اساس ماتریس مقایسه و عناصر موجود در آن بیان می‌شوند. در شاخص کاپا بر خلاف دقت کلی تمامی عناصر ماتریس مقایسه مورد استفاده قرار می‌گیرند که این مزیت مهمی برای این روش به حساب می‌آید.

دقت کلی و شاخص کاپا با توجه به روابط (۱۴) و (۱۵) قابل محاسبه هستند[۳۶]:

$$OA = \frac{\sum_{i=1}^{c} P_{ii}}{\sum_{i=1}^{c}\sum_{e=1}^{c} P_{ie}} \qquad \text{رابطه(۱۴)}$$

$$Kappa = \frac{\sum_{i=1}^{c} P_{ii} - \sum_{i=1}^{c}\sum_{t=1}^{c} P_{it}P_{ti}}{1-\sum_{i=1}^{c}\sum_{t=1}^{c} P_{it}P_{ti}} \qquad \text{رابطه(۱۵)}$$

در روابط(۱۴) و (۱۵)، i و e بیانگر حالت‌های موجود برای هر درایه در ماتریس مقایسه، P_{ii}بیان‌گر تعداد درایه‌هایی که در واقعیت در حالت iبوده و در کلاسه-بندی نیز در حالت iهستند، به همین ترتیب P_{ie}بیان-گر تعداد درایه‌هایی که در واقعیت در حالت i بوده و در کلاسه‌بندی نیز در حالت eهستند، P_{it}بیان‌گر مجموع درایه های حالت iدر واقعیت و P_{ti}بیان گر مجموع درایه‌های حالت iدر کلاسه‌بندی هستند. هر چه مقدار کاپا به عدد یک نزدیکتر باشد یعنی کلاسه‌بندی و مدل‌سازی به خوبی انجام شده است[۳۷].

۲-۵- قوانین انجمنی[3]

قوانین انجمنی[۳۸] روابط و وابستگی‌های متقابل بین مجموعه‌ای بزرگ از داده‌ها را نشان می‌دهند. اگر مجموعه داده‌ای به نام I وجود داشته باشد، $A \to B$ گزاره ای از یک قانون انجمنی می‌باشد به شرطی که اولا Aو Bزیر مجموعه‌ای از مجموعه داده I باشند و ثانیا با یکدیگر نیز اشتراکی نداشته باشند. قوانین انجمنی باید از لحاظ ارزش و معیار مقبولیت مورد بررسی قرار بگیرند که در این راستا برای بررسی ارزش و معیار مقبولیت قوانین انجمنی به ترتیب از دو پارامتر مهم پشتیبان و اطمینان استفاده می‌شود. پارامتر پشتیبان نشانگر آن است که یک مجموعه اقلام چند بار در پایگاه داده تکرار شده است. این مقدار به عنوان کسر رکوردهای شامل $X \cup Y$ بر کل تعداد رکوردها در پایگاه داده است. پارامتر اطمینان به نسبت تعداد تراکنش‌های حاوی $X \cup Y$ به کل تعداد رکوردهای شامل x گویند. این مقدار، مقیاسی از استحکام قواعد وابستگی است. پارامترهای پشتیبان و اطمینان با توجه به روابط (۱۶) و (۱۷) به دست می‌آیند [۳۹]:

$$P(A \cup B) = Support(A \to B) \qquad \text{رابطه(۱۶)}$$

$$P(A \mid B) = Confidence(A \to B) \qquad \text{رابطه(۱۷)}$$

[1] *Grid search*
[2] *Decision tree*
[3] *Associative Rules*

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

قوانینی که حد پایین پشتیبان[1] و حد پایین اطمینان[2] را داشته باشند، قوانین انجمنی قوی نامیده می‌شوند[۴۰].

در این پژوهش برای کشف قوانین انجمنی از الگوریتم ایپریوری[3] استفاده می‌شود.

الگوریتم ایپریوری یک الگوریتم توانمند برای کاوش مجموعه آیتم‌های پرتکرار برای کشف قوانین انجمنی بولی می‌باشد[۴۱]. هدف اصلی این الگوریتم یافتن مجموعه آیتم‌های پرتکرار است. یک زیرمجموعه از یک مجموعه آیتم پرتکرار نیز لازم است یک مجموعه آیتم پرتکرار باشد تا بتوان مجموعه آیتم‌های پرتکرار را با کاردینالیتی از ۱ تا k به صورت تکراری پیدا نمود و از این مجموعه آیتم‌های پرتکرار جهت ایجاد قوانین انجمنی استفاده نمود. ایپریوری از یک روش تکراری برای یافتن مجموعه عناصر پرتکرار استفاده می‌کند به این صورت که برای یافتن $(k+1)$ -مجموعه آیتم‌ها از k –مجموعه آیتم‌ها استفاده می‌کند. ابتدا ۱-مجموعه آیتم‌ها پیدا می‌شوند که با L_1 نمایش داده می‌شوند. L_1برای یافتن L_2 که ۲ –مجموعه آیتم‌ها می‌باشند استفاده می‌شود و همینطور این فرآیند ادامه دارد تا هیچ k –مجموعه آیتمی یافت نشود. یافتن L_k نیاز دارد تا کل پایگاه یکبار پیمایش شود [۴۱].

۳- پیاده‌سازی و ارزیابی

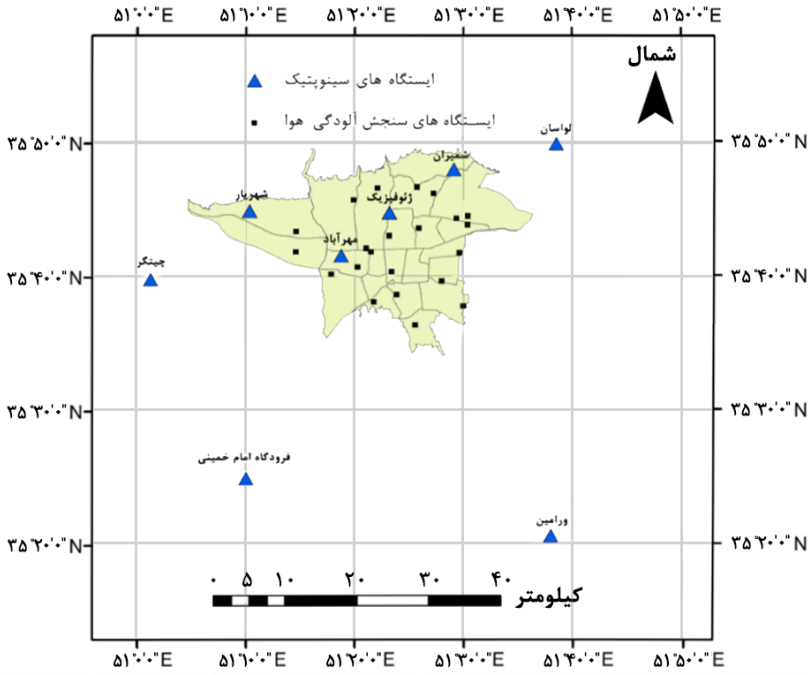
شهر تهران در کوهپایه‌های جنوبی رشته کوه البرز، حد فاصل طول جغرافیایی′۵۳ °۵۱-′۵ °۵۱ شرقی و عرض جغرافیایی ′۵۹ °۳۵-′۳۴ °۳۵ شمالی با حدود ۷۰۰کیلومتر مربع مساحت گسترده شده است. ارتفاع شهر تهران بین ۱۰۵۰-۲۰۰۰ متر بالاتر از سطح دریا می‌باشد.

در شکل(۵) موقعیت جغرافیایی شهر تهران و توزیع ایستگاه‌های سینوپتیک و سنجش آلودگی هوا نمایش داده می‌شود.

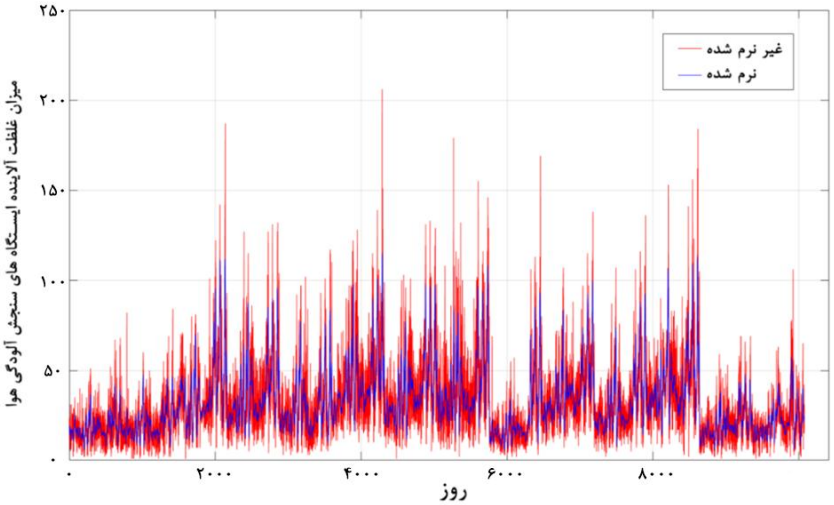
این پژوهش بر روی داده‌های مربوط به فصل پاییز سال‌های ۹۷ و ۹۸ انجام شده است. داده‌های استفاده شده شامل داده‌های هواشناسی (سرعت باد، جهت باد، رطوبت نسبی، میزان بارش و دما) مربوط به ایستگاه‌های سینوپتیک واقع در استان تهران که از سازمان هواشناسی کل کشور[۴۲] اخذ گردید. سپس پارامترهای ایستگاه‌های سینوپتیک را با روش کریجینگ[4][۴۳] با اندازه پیکسل ۱۰۰ متر به صورت پیوسته تولید کرده و مقدار آن‌ها برای تمام نقاط ایستگاه‌های سنجش آلودگی هوای شهر تهران تعیین می‌شود. با توجه به وجود همبستگی فاصله‌ای داده‌ها برای درون یابی از روش کریجینگ استفاده می‌شود. مقادیر غلظت آلاینده ذرات معلق کم‌تر از ۲٫۵ میکرون توسط ایستگاه‌های فعال سنجش آلودگی هوا در شهر تهران اندازه‌گیری می‌شود که این اطلاعات از سایت کیفیت هوای شهر تهران[۴۴] دریافت گردید. برای پر کردن گپ میان داده‌ها از روش تابع اسپیلاین استفاده گردید. در ادامه داده‌ها به دلیل داشتن نوسان بالا همانطور که در داده‌های میزان غلظت آلاینده ذرات معلق کم‌تر از ۲٫۵ میکرون در شکل(۶) دیده می‌شود نمودار آن برای دست‌یابی به دقت بالا با استفاده از روش ساویتزکی گولای[5]نرم گردید[۲۵].

[1] *Min-sup*
[2] *Min-conf*
[3] *Apriori*
[4] *Kriging*
[5] *Savitzky-golay*

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]



شکل۵: نمایش موقعیت ایستگاه های سینوپتیک و سنجش آلودگی هوا



شکل۶: نمایش نوسانات میزان غلظت آلاینده $PM_{2.5}$ ایستگاه‌ها غیر نرم‌شده و نرم‌شده

برای ارزیابی روش‌ها و مدل‌ها، ابتدا از شاخص KMO[1][۴۵] برای اطمینان از کافی بودن میزان نمونه داده‌ها استفاده می‌شود که اندازه شاخص برای این مجموعه داده ۰٫۸۳ بود که نشان دهنده‌ی کفایت نمونه‌گیری می‌باشد سپس در ادامه برای به دو گروه داده‌های آموزشی و داده‌های آزمایشی تقسیم شدند. داده‌های آموزشی به‌عنوان ابزار اصلی مدل‌سازی و آموزش روش، استفاده می‌شوند و قسمت عمدهٔ داده‌ها

[1] *Kaiser-Mayer-Olkin*

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

متعلق به این گروه است. دستهٔ دیگر، داده‌های آزمایشی است که در فرایند آموزش دخالتی ندارند و تنها برای آزمودن آموزش صورت‌گرفته، استفاده می‌شوند. از مجموع داده‌های پاییز ۹۷ و پاییز ۹۸،که به صورت ساعتی برداشت شده اند، به‌طور یکسان برای هر سه روش،۷۰ درصد داده‌ها در گروه داده‌های آموزشی ۳۰ درصد داده‌ها در گروه داده‌های آزمایشی قرار گرفتند که انتخاب این نقاط با استفاده از روش اعتبارسنجی متقابل (*Cross-validation*) صورت می‌گیرد[۴۶و ۴۷]. اعتبارسنجی متقابل یک روش ارزیابی است که مشخص می‌کند نتایج یک تحلیل آماری بر روی یک مجموعه داده تا چه اندازه قابل تعمیم و مستقل از داده‌های آموزشی است. به طور کلی اعتبارسنجی متقابل شامل افراز داده‌ها به دو زیرمجموعه مکمل، انجام تحلیل بر روی یکی از آن زیرمجموعه‌ها (داده‌های آموزشی) و اعتبارسنجی تحلیل با استفاده از داده‌های مجموعه دیگر (داده‌های اعتبارسنجی یا آزمایشی) است. برای کاهش پراکندگی، عمل اعتبارسنجی چندین بار با افرازهای مختلف انجام و از نتایج اعتبارسنجی‌ها میانگین گرفته می‌شود. در این پژوهش داده‌ها به ۵ زیرمجموعه (*5-Fold*) افراز شدند به این صورت که هربار چهار زیرمجموعه برای آموزش و زیرمجموعه باقیمانده برای اعتبارسنجی به کار می‌روند. در نهایت میانگین نتایج به عنوان نتیجه نهایی درنظر گرفته می‌شود. پس از انجام اعتبارسنجی متقابل، داده‌ها وارد مدل‌ها می‌شوند و با استفاده از روش‌های ارزیابی نتایج با یکدیگر مقایسه شدند که در جدول (۴) آمده است.

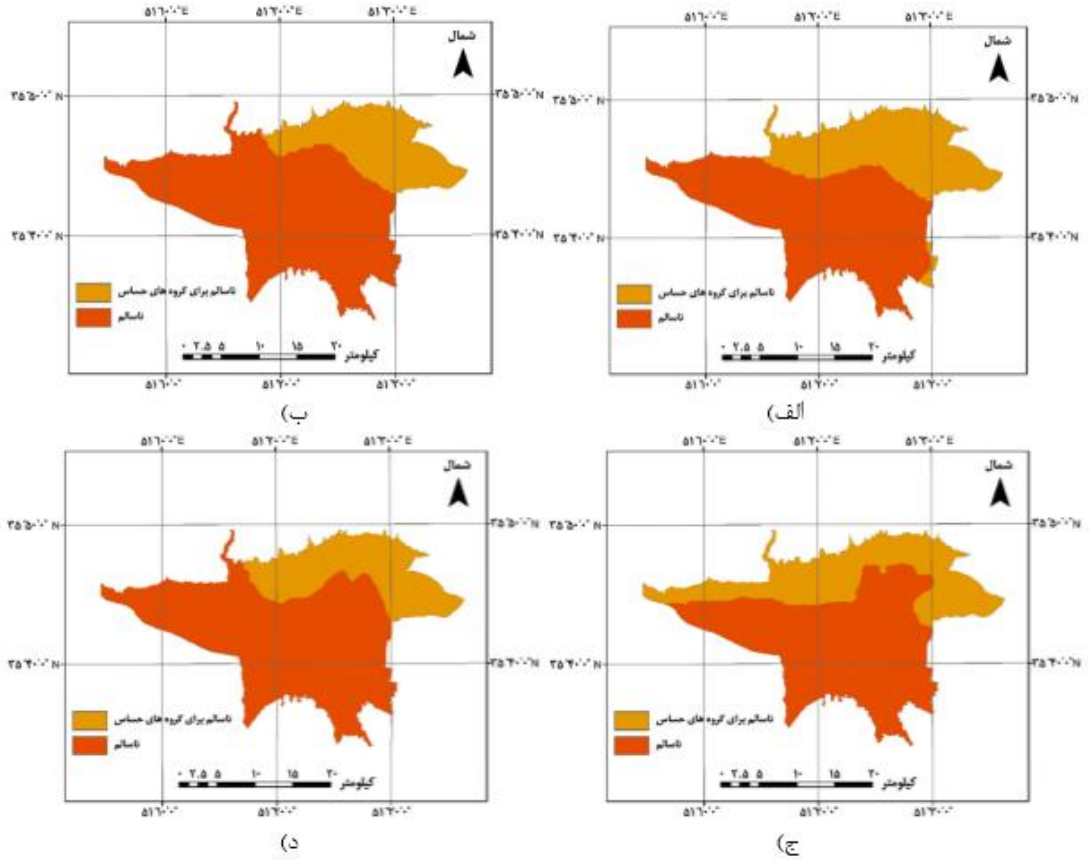
سپس با استفاده از هر روش، نقشه پیش‌بینی طبقه-بندی کلاس آلودگی آلاینده به طور نمونه و برای یک روز خاص تولید گردید. به عنوان نمونه شکل (۷) نقشه پیش‌بینی طبقه‌بندی تولید شده مربوط به ساعت ۱۲، روز ۲۰ آبان، سال ۱۳۹۸ با استفاده از داده‌های ۶ ساعت قبل (تاخیر زمانی ۶ ساعته) را نمایش می‌دهد.

با توجه به شکل (۷) که برای یک روز و زمان خاص تولید شده است می‌توان دریافت که مدل ماشین بردار پشتیبان، مدل واقعی را بهتر پیش‌بینی می‌کند. با توجه به تاخیر زمانی(ساعت) ذکر شده در ساختار جدول(۲)، داده‌های ورودی برای پیش‌بینی مورد ارزیابی قرار گرفتند. در جدول (۵) پیش بینی‌ها تا ۷۲ ساعت آینده به همراه دقت کلی و میزان شاخص کاپا بیان شده است که پس از ۷۲ ساعت دقت کلی به کم‌تر از ۸۰ درصد افول پیدا کرد.

جدول۴: دقت کلی هر یک از روش‌ها و شاخص کاپا

پارامتر / روش	دقت‌کلی	شاخص‌کاپا
شبکه عصبی	۸۲/۷	۷۳/۸
ماشین بردار پشتیبان	۸۷/۳	۸۱/۵
درخت تصمیم	۷۵/۶	۷۱/۴

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]



شکل۷: نقشه پیش‌بینی طبقه‌بندی میزان آلودگی $PM_{2.5}$، (الف)مدل واقعی (ب)ماشین بردار پشتیبان (ج) شبکه عصبی د)درخت تصمیم

جدول۵: دقت کلی و میزان شاخص کاپا برای پیش‌بینی کلاس آلاینده با استفاده از روش ماشین بردار پشتیبان

تاخیر زمانی	دقت کلی	شاخص کاپا
۶	۸۶٫۹	۸۰٫۲
۱۲	۸۶٫۳	۷۸٫۳
۱۸	۸۵٫۸	۷۷٫۹
۲۴	۸۵٫۵	۷۷٫۱
۳۰	۸۵٫۰	۷۶٫۲
۳۶	۸۴٫۸	۷۵٫۳
۴۲	۸۴٫۴	۷۴٫۱
۴۸	۸۴٫۱	۷۳٫۴
۵۴	۸۳٫۹	۷۲٫۹
۶۰	۸۳٫۲	۷۲٫۳
۶۶	۸۲٫۵	۷۱٫۸
۷۲	۸۰٫۷	۷۱٫۱
۷۸	۷۶٫۲	۶۷٫۵
۸۴	۷۱٫۴	۶۲٫۰

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

پس از تهیه مدل پیش‌بینی، یافتن دقت کلی و شاخص کاپا تا ۷۲ ساعت آینده، قوانین انجمنی قوی از روش برتر استخراج شدند. در این پژوهش تنها به استخراج قوانین انجمنی موثر(منظور از قوانین انجمنی موثر قوانینی هستند که تاثیر بیش‌تری بر روی تعداد داده-های نمونه داشتند) پرداخته شده است که مقدار حدپایین اطمینان و حدپایین پشتیبان به ترتیب برابر با ۰٫۵ و ۰٫۱ را دارا باشند. تعداد کل قوانین انجمنی استخراج شده برابر با ۴۶ قانون بود که از بین آن‌ها فقط تعداد ۱۹ قانون از قوانین انجمنی موثر بوده-اند(جدول(۶)).

بازه‌های موجود در جدول(۶) توسط الگوریتم ایپریوری و با توجه به بازه داده‌های موجود و میزان تکرارپذیری این داده‌ها در بازه مورد نظر تبدیل به قانون گردیده است که در انتها حد پایین اطمینان و پشتیبان برای هر قانون در جدول آمده است تا میزان دقت و اعتمادپذیری مورد بررسی قرار گیرد.

جدول۶: قوانین انجمنی موثر استخراج شده از روش برتر پژوهش

شماره	اگر	آنگاه	پشتیبان	اطمینان
۱	$ws \in [۱.۸, \infty] \& wd \in [۳۱۲.۱, ۳۶۰]$	پاک	۰٫۱۱۹	۰٫۶۸۹
۲	$ws \in [۱.۸, \infty] \& r \in [۰.۱, \infty]$	پاک	۰٫۱۳۲	۰٫۷۵۲
۳	$moy \in [۷,۸] \& dow \in [۵,۷] \& hod [۰,۶]$	پاک	۰٫۱۰۵	۰٫۵۷۶
۴	$moy \in [۷,۸] \& t \in [۱۵.۶, \infty] \& u [۴۲.۲, \infty]$	پاک	۰٫۱۰۲	۰٫۵۱۶
۵	$p \in [۸۳۰, ۸۶۰.۷] \& dI \in [-۰.۷, -۰.۳]$	پاک	۰٫۱۵۶	۰٫۷۱۱
۶	$ws \in [۱.۳, ۱.۸] \& wd \in [۲۶۰, ۳۱۲.۱]$	سالم	۰٫۱۰۷	۰٫۷۲۳
۷	$dI \in [-\infty, -۱.۶]$	سالم	۰٫۱۳۱	۰٫۷۱
۸	$y \in [۳۹۵۲۹۹۱.۴, ۳۹۵۸۴۶۱.۳] \& ۲ - NN \in [۰, ۱۹.۱]$	سالم	۰٫۱۲۱	۰٫۷۵۹
۹	$u \in [۴۲.۲, ۶۲.۲] \& r \in [۰, ۰.۱]$	سالم	۰٫۱۱۱	۰٫۶۹۴
۱۰	$wd \in [۱۴۸.۹, ۲۶۰] \& ws \in [۱.۳, \infty]$	ناسالم برای گروه های حساس	۰٫۱۰۸	۰٫۶۵۱
۱۱	$p \in [۸۶۰.۷, ۸۷۴.۲] \& ۲ - NN \in [۳۷.۲, ۵۵]$	ناسالم برای گروه های حساس	۰٫۱۱۶	۰٫۷۶۸
۱۲	$r \in [۰, ۰.۱] \& t \in [-\infty, ۹.۵] \& dI [-۰.۵, ۱.۳]$	ناسالم برای گروه های حساس	۰٫۱۱۸	۰٫۷۹۵
۱۳	$moy \in [۸,۹] \& dow \in [۱,۳] \& hod \in [۱۲,۱۸]$	ناسالم برای گروه های حساس	۰٫۱۰۱	۰٫۵۴۳
۱۴	$dow \in [۱,۳] \& ۲ - NN \in [۴۲.۵, \infty]$	ناسالم	۰٫۱۰۲	۰٫۶۱۹
۱۵	$p \in [۸۷۴.۲, \infty] \& ۲ - NN \in [۴۲.۵, \infty]$	ناسالم	۰٫۱۰۷	۰٫۹۳۱
۱۶	$y \in [-\infty, ۳۹۵۲۹۹۱.۴] \& ۲ - NN \in [۴۲.۵, \infty]$	ناسالم	۰٫۱۰۵	۰٫۸۴۶
۱۷	$r \in [۰] \& t \in [-\infty, ۹.۵] \& dI [۱.۳, \infty]$	ناسالم	۰٫۱۴۶	۰٫۹۰۱
۱۸	$x \in [-\infty, ۵۳۰۸۶۵.۵] \& ۲ - NN \in [۴۲.۵, \infty]$	ناسالم	۰٫۱۳۴	۰٫۹۲۹
۱۹	$u \in [۶۲.۲, \infty] \& t \in [-\infty, ۹.۵]$	ناسالم	۰٫۱۰۱	۰٫۶۹۸

۳-۱- ارزیابی نتایج

همانطور که در جدول(۴) مشاهده می‌شود، مدل بردار پشتیبان دارای دقت بالاتری نسبت به بقیه روش‌ها می-باشد. پس از آن با استفاده از این مدل و تهیه جدول(۲) به پیش‌بینی کلاس آلاینده ذرات معلق کم‌تر از ۲٫۵ میکرون طی زمان‌های آتی پرداخته شد. همانگونه که در جدول(۵) مشاهده می‌شود این مدل توانایی پیش‌بینی غلظت آلاینده را تا ۳ روز آینده با

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

دقت کلی بالای ۸۰ درصد را دارد و پس از ۷۲ ساعت، دقت کلی به کم‌تر از ۸۰ درصد افول می‌کند. این پیش‌بینی امکان مدیریت مخاطره را بدست می‌دهد. پس از آن نقشه پیش‌بینی طبقه‌بندی آلاینده با سه روش یادگیری ماشین نظارت شده به‌کار گرفته شده برای یک زمان خاص به طور مثال ساعت ۱۲، روز ۲۰ آبان، سال ۱۳۹۸ با داده‌های مربوط به ۶ ساعت قبل تولید گردید و مشخص‌شد که مدل ماشین بردار پشتیبان شباهت بیش‌تری را به مدل واقعی و مدل درخت تصمیم کم‌ترین شباهت را به مدل واقعی دارد. در ادامه به استخراج قوانین انجمنی از بهترین روش پرداخته شد که از بین این قوانین که تعداد آن‌ها برابر با ۴۶ قانون بود، قوانین موثر با مقدار حدپایین اطمینان و حدپایین پشتیبان به ترتیب برابر با ۰٫۵ و ۰٫۱ انتخاب شدند که در جدول (۶) قابل مشاهده می‌باشند. گام بعد تحلیل پارامترهای مکانی موثر است که با استفاده از قوانین انجمنی قوی‌تر و با توجه به میزان بیش‌ترین تکرار پارامترها در قوانین، پارامترهای میزان آلاینده نزدیکترین دو همسایگی، وضعیت توپوگرافی، دما، فشار هوا، میزان بارش، شدت وارونگی دما، رطوبت نسبی، سرعت باد، جهت باد، ماه، روز، ساعت به ترتیب بیش‌ترین میزان تاثیر را در پیش‌بینی غلظت آلاینده $PM_{۲٫۵}$ داشته اند.

۴- نتیجه گیری

هدف از انجام این پژوهش طبقه‌بندی کلاس‌های آلودگی آلاینده، استخراج روابط میان پارامترهای مکانی و کلاس آلودگی آلاینده با استفاده از قوانین انجمنی و پیش‌بینی کلاس آلودگی آلاینده برای ۷۲ ساعت آینده می‌باشد که امکان مدیریت و برنامه‌ریزی صحیح را به ما می‌دهد. از میان روش‌های یادگیری ماشین، دقت روش ماشین بردار پشتیبان از بقیه روش‌ها بالاتر بود واین مدل پیش‌بینی بهتری از واقعیت دارد. به همین دلیل قوانین انجمنی موثر از این روش استخراج گردید و نتایج زیر به ترتیب از قوانینی که حد پایین پشتیبان و اطمینان بالاتری را نسبت به قوانین دیگر دارند، به دست می‌آید:

- پارامتر میزان غلظت نزدیکترین دو ایستگاه همسایه بیش‌ترین تاثیر را در تعیین کلاس آلودگی آلاینده دارد. (قوانین شماره ۸، ۱۱، ۱۴، ۱۵، ۱۶، ۱۸)
- اگر باد از مناطق شمال‌غرب و غربی شهر بوزد، میزان آلودگی آلاینده کم خواهد بود و اگر باد از سمت جنوب و جنوب‌غربی بوزد، به دلیل وجود کارخانجات صنعتی در آن قسمت، آلاینده به سمت شهر در حال حرکت خواهد بود. (قوانین شماره ۱، ۶، ۱۰)
- سرعت باد از لحاظ تاثیر در میزان آلودگی آلاینده اکثرا وابسته به اینکه جهت باد به کدام سمت است، می‌باشد اما به طور کلی هرچه سرعت وزش باد بیش‌تر باشد، آلاینده جابجا شده و میزان آلودگی آلاینده کاهش خواهد یافت. (قوانین شماره ۱، ۲، ۶، ۱۰)
- بارش باران تاثیر بسیار به سزایی در کنترل آلودگی دارد به اینصورت که بارش‌های بالای ۰٫۱ میلی‌متر باعث پاک شدن هوا می‌شود و در صورتی که بارش نباشد یا میزان بارش کم‌تر از ۰٫۱ میلی‌متر باشد، آلودگی آلاینده بیش‌تر خواهد شد. (قوانین شماره ۲، ۹، ۱۷)
- می‌توان نتیجه گرفت که با کاهش دما در فصول سرد سال امکان وقوع پدیده وارونگی دما وجود دارد، این پدیده به شدت باعث افزایش میزان غلظت آلاینده می‌شود. وارونگی دما با تغییرات دما و پایداری و ناپایداری وضعیت هوا رابطه مستقیم دارد. (قوانین ۵، ۷، ۱۲)
- با افزایش رطوبت نسبی، غلظت آلاینده افزایش می‌یابد البته باید در نظر گرفته شود اگر رطوبت نسبی به دلیل رخ دادن بارش باشد غلظت آلاینده به طور چشم‌گیری کاهش می‌یابد. (قوانین شماره ۴، ۹، ۱۹)
- تغییرات فشار هوا باعث پایداری هوا شده و افزایش غلظت آلودگی آلاینده $PM_{۲٫۵}$ را به همراه دارد. (قوانین شماره ۵، ۱۱، ۱۵)
- وضعیت توپوگرافی نیز یکی از عوامل موثر در تعیین

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

کلاس غلظت آلاینده است و نواحی مرتفع‌تر کلاس های آلودگی پایین‌تر و مناطق جنوبی و مرکزی شهر نیز آلودگی بیش‌تری نسبت به مناطق دیگر تجربه می‌کنند. (قوانین شماره ۸، ۱۶، ۱۸)

- غلظت آلاینده نیز با توجه به زمان تغییر می‌کند و یکسان نیست. ۱-افزایش غلظت آلاینده در ماه‌های سرد سال بیش‌تر رخ می‌دهد. ۲-روز در هفته نیز عامل تاثیرگذار دیگری است که مطابق با انتظار اواسط هفته، میزان غلظت آلاینده معمولا نسبت به روزهای اول و آخر هفته کم‌تر خواهد بود. ۳-میزان غلظت آلاینده در بازه زمانی صبح و شب کم‌تر از بازه زمانی ظهر می‌باشد به گونه‌ای که معمولا ساعات ۶ الی۱۸ بالاترین میزان آلاینده را در روز تجربه می کنند. (قوانین شماره ۳، ۴، ۱۳، ۱۴)

در این پژوهش به علت حجم بالای داده‌های موجود و در دسترس نبودن برخی داده‌ها برای هر پارسل به منظور کلاسه بندی تمامی پارامترها در نظر گرفته نشده است که در پژوهش‌های آتی می‌توان پارامترهای متنوعی از جمله ترافیک، تراکم جمعیت یا شبکه راه و تنوع پوشش گیاهی را برای مدل‌سازی و پیش‌بینی مورد استفاده قرار داد.

مراجع

[1] *Y.-S. Chang, et al., "An LSTM-based aggregated model for air pollution forecasting", Atmospheric Pollution Research, Vol.11(8), pp. 1451-1463, 2020.*

[2] *K. Harishkumar, K. Yogesh, and I. Gad, "Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models", Procedia Computer Science, Vol.171, pp. 2057-2066, 2020.*

[3] *D.W. Dockery, "Health effects of particulate air pollution", Annals of epidemiology, Vol.19(4), pp. 257-263, 2009.*

[4] *M.A. Elangasinghe, et al., "Development of an ANN–based air pollution forecasting system with explicit knowledge through sensitivity analysis", Atmospheric pollution research, Vol.5(4), pp. 696-708, 2014.*

[5] *M. Dong, et al., "PM2.5 concentration prediction using hidden semi-Markov model-based times series data mining", Expert Systems with Applications, Vol.36(5), pp. 9046-9055, 2009.*

[6] *M. Elangasinghe, et al., "Complex time series analysis of PM10 and PM2.5 for a coastal site using artificial neural network modelling and k-means clustering", Atmospheric Environment, Vol.94, pp. 106-116, 2014.*

[7] *H. Niska, et al., "Evolving the neural network model for forecasting air pollution time series", Engineering Applications of Artificial Intelligence, Vol.17(2), pp. 159-167, 2004.*

[8] *K.P. Singh, S. Gupta, and P. Rai, "Identifying pollution sources and predicting urban air quality using ensemble learning methods", Atmospheric Environment, Vol.80, pp. 426-437, 2013.*

[9] *X. Feng, et al., "Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation", Atmospheric Environment, Vol.107, pp. 118-12, 2015.*

[10] *Z. Rostami Fasih, et al., "Forecasting the air quality index based on meteorological variables and autocorrelation terms using artificial neural network", Razi Journal of Medical Sciences, Vol.22(137), pp. 31-43, 2015.*

[11] *Cakir, S. and Sita, M., 2020. "Evaluating the performance of ANN in predicting the concentrations of ambient air pollutants in Nicosia". vol.11, no.12, p. 2327-2334.*

[12] *Chae, S., Shin, J., Kwon, S. et al. PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network. Sci Rep 11, 11952 (2021). https://doi.org/10.1038/s41598-021-91253-9.*

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

Pahlavani. "Prediction of pm2.5 concentrations using temperature inversion effects based on an artificial neural network." The ISPRS international conference of Geospatial information research, Vol. 15. 2014.

[22]http://weather.uwyo.edu/upperair/sounding .html.

[23]https://aqs.epa.gov/aqsweb/documents/code tables/aqi_breakpoints.html.

[24]M.R. Delavar, et al., "A novel method for improving air pollution prediction based on machine learning approaches: a case study applied to the capital city of Tehran". ISPRS International Journal of Geo-Information, Vol.8(2), pp. 99, 2019.

[25]D. Acharya, et al., "Application of adaptive Savitzky–Golay filter for EEG signal processing", Perspectives in science, Vol.8, pp. 677-679, 2016.

[26]A. K. Jain, Jianchang Mao and K. M. Mohiuddin, "Artificial neural networks: a tutorial," in Computer, vol.29(3), pp. 31-44, 1996.

[27]V.N. Subramanian, "Data analysis for predicting air pollutant concentration in Smart city Uppsala", Dissertation, 2016.

[28]M. Cai, Y. Yin, and M. Xie, "Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach". Transportation Research Part D: Transport and Environment, Vol.14(1), pp. 32-41, 2009.

[29]J.J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory", in Numerical analysis, Heidelberg: Springer ,1978, pp. 105-116.

[30]V.N. Vapnik, "An overview of statistical learning theory". IEEE transactions on neural networks, Vol.10(5), pp. 988-999, 1999.

[31]W. Wang, C. Men, and W. Lu, "Online prediction model based on support vector machine", Neurocomputing, Vol.71(4-6), pp. 550-558, 2008.

[13]W. Leong, R. Kelani, and Z .Ahmad, "Prediction of air pollution index (API) using support vector machine (SVM)", Journal of Environmental Chemical Engineering, Vol.8(3): pp. 103208, 2020.

[14]Z. Ghaemi, M. Farnaghi, and A. ALIMOHAMMADI, "An Online Approach for Spatio-Temporal Prediction of Air Pollution in Tehran using Support Vector Machine", Scientific Information database, Vol.3(4), pp.43-63, 2016.

[15]W. Lu et al., "Air pollutant parameter forecasting using support vector machines", Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290), , vol.1, pp.630-635, 2002.

[16]Delavar, M.R., et al., 2019. "A novel method for improving air pollution prediction based on machine learning approaches: a case study applied to the capital city of Tehran". vol.8, no.2, p. 99.

[17] A. Jamal, and R.N. Nodehi, "Predicting air quality index based on meteorological data: A comparison of regression analysis, artificial neural networks and decision tree". Journal of Air Pollution And Health, Vol.2(1),2017.

[18]A.M. Hilal, et al., Machine learning-based Decision Tree J48 with grey wolf optimizer for environmental pollution control, Environmental Technology, 2022. DOI: 10.1080/09593330.2021.2017491.

[19]C. Payus, et al., "Association rules of data mining application for respiratory illness by air pollution database", Int J Basic Appl Sci, Vol.13(3), pp. 11-16, 2013.

[20]E. Sahafizadeh and E. Ahmadi, "Prediction of Air Pollution of Boushehr City Using Data Mining," 2009 Second International Conference on Environmental and Computer Science, Dubai, 2009, pp. 33-36.

[21]R.A. Bahari, R. Ali Abbaspour, and P.

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

[32]C. Cortes, and V. Vapnik, "Support-vector networks". Machine learning, Vol.20(3): pp. 273-297, 1995.

[33]C.J. Burges, "A tutorial on support vector machines for pattern recognition". Data mining and knowledge discovery, Vol.2(2), pp. 121-167, 1998.

[34]B. Bigdeli, H. Amini Amirkolaee, and P. Pahlavani, "Deep feature learning versus shallow feature learning systems for joint use of airborne thermal hyperspectral and visible remote sensing data". International Journal of Remote Sensing, Vol.40(18), pp. 7048-7070, 2019.

[35]D. Zhu, et al., "A machine learning approach for air quality prediction: Model regularization and optimization" .Big data and cognitive computing, Vol.2(1), pp. 5, 2018.

[36]H. Askarian Omran, and P. Pahlavani, "Using of Markov Chain, MOLA, and Neighborhood filter for developing and increasing the efficiency of Logistic Regression to predict multiple land-use changes, a case study: Tehran". Engineering Journal of Geospatial Information Technology, Vol.3(2), pp. 89-109, 2015.

[37]Pijanowski, B., et al., "Urban expansion simulation using geospatial information system and artificial neural networks", International journal of environmental research(IJER), Vol.3(4), pp.493-502, 2009.

[38]S. Brin, R. Motwani ,and C. Silverstein. "Beyond market baskets: Generalizing association rules to correlations", in Proceedings of the 1997 ACM SIGMOD international conference on Management of data, pp.265-276, 1997.

[39]Zhang, Chengqi, and Shichao Zhang. "Association rule mining: models and algorithms", Association rules, Heidenberg: Springer, Vol. 2307, 2003, pp. 25-197.

[40]R. Ramezani, M. Saraee and M. A. Nematbakhsh, "Finding association rules in linked data, a centralization approach," 2013 21st Iranian Conference on Electrical Engineering (ICEE), Mashhad, 2013, pp. 1-6.

[41]R. Agarwal, and R. Srikant. "Fast algorithms for mining association rules". in Proc. of the 20th VLDB Conference, Vol.1215, 1994, pp. 487-499.

[42]http://www.irimo.ir/.

[43]J.-P. Chilès, and N. Desassis, Fifty years of kriging, in Handbook of mathematical geosciences, Springer, Cham. 2018,

[44]http://airnow.tehran.ir/.

[45]A. Ruano, Motter FR, Lopes LCDesign and validity of an instrument to assess healthcare professionals' perceptions, behaviour, self-efficacy and attitudes towards evidence-based health practice: I-SABEBMJ Open 2022;12:e052767. doi: 10.1136/bmjopen-2021-052767.

[46]D.M. Allen, "The relationship between variable selection and data agumentation and a method for prediction". technometrics, Vol.16(1), pp. 125-127, 1974.

[47]M. Stone, "Cross-validatory choice and assessment of statistical predictions", Journal of the Royal Statistical Society: Series B (Methodological), Vol.36(2), pp. 111-133, 1974.

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]

K. N. TOOSI UNIVERSITY OF TECHNOLOGY
FACULTY OF GEODESY AND GEOMATICS ENGINEERING

Journal of Geospatial Information Technology
Vol.10, No.4, Winter 2023

Research Paper

Providing The Classification And Prediction of $PM_{2.5}$ Pollutant Map Using Machine Learning Methods And Extracting Association Rules

Mohammadreza Heydari [1], Parham Pahlavani [2*], Behnaz Bigdeli [3]

1- GIS M.Sc. Student at School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran
2- Associate Professor at School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran
3- Associate Professor at School of Civil Engineering, Shahrood University of Technology, Shahrood, Iran

Abstract

Air pollution is caused by the presence of various pollutants in the air, which is mostly related to the presence of particulate matters, especially particulate pollutant concentrations which are smaller than 2.5 microns (PM2.5). Predicting and identifying the infected areas will help us in managing and planning. Therefore, in order to identify these places, it is necessary to provide the maps of classification and the prediction maps of the PM2.5 pollution. The Supervised Machine Learning Methods used in this study, were Support Vector Machine, Multilayer Neural Network, and Decision Tree for classifying and predicting the PM2.5 pollutant maps in Tehran city. Moreover, to identify the effect of the spatial parameters, the Association Rules Mining Method was used. The Support Vector Machine Method with 87.3 percent for overall accuracy and 81.5 percent for the Kappa index was selected as the best classifier. This method was used to predict the concentration of the pollutants on the third day, which was able to predict the third day with 80.7 percent for the overall accuracy and 71.1 percent for the Kappa index. The findings indicate that the Support Machine Vector Method performs modeling and predicting with higher accuracy than the other methods. Attention to the influence of the spatial parameters in stronger association rules, the amount of the pollution of the nearest two neighborhoods, topography, temperature, air pressure, rainfall, intensity of air inversion, relative humidity, wind speed, wind direction, month of the year, day of the week, hour of the day had the greatest impact on determining the pollutant class.

Key words: *Air Pollution; PM2.5 pollutant; Spatial parameters; Supervised Machine Learning; Associative Rules.*

Correspondence Address*: GIS Group, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran.*
Tel*: +98 21 61114524.*
Email: *pahlavani@ut.ac.ir*

[Downloaded from jgit.kntu.ac.ir on 2025-07-05]
[DOR: 20.1001.1.20089635.1401.10.4.4.4]
[DOI: 10.61186/jgit.10.4.67]