

A study on the simulation and estimation of the urban air quality using artificial intelligence (case study: Gholhak Station, Tehran)

Ehsanzadeh Alireza

Graduate student, Department of Environmental Engineering, Yazd University of Iran.

Farhad Nejadkoorki

* Faculty member of the Department of Environmental Engineering, Yazd University of Iran (Corresponding Author) f.nejadkoorki@yazd.ac.ir

Ali Talebi

Faculty member of the Department of Watershed Management Engineering, Yazd University of Iran.

Received: 15 December 2015

Accepted: 2 February 2016

ABSTRACT

Background & objective: Air pollution in Tehran, because of high concentration of pollutants, has caused various diseases and many problems concerning the public health and welfare of citizens and also damages to the environment and living organisms.

Materials & Methods: Air Quality Index (AQI) is a key tool to monitor the air quality, to realize the effects of air pollution on health and to choose methods against air pollution. This study aimed at modeling and estimation AQI by CART algorithm and adaptive boosting algorithm (AdaBoost). Hourly data on concentration of air pollutants and meteorological parameters related to Gholhak stations in Tehran was used for modeling and estimation of AQI.

Results: The results showed that CART model had better performance than AdaBoost model. To evaluate these models, root mean square error (RMSE), mean absolute error (MAE), mean squared error (MSE) and correlation coefficient (R) of the CART model for the test, were respectively, 0.75, 0.101, 0.563, and 0.99 when compared to the AdaBoost model (RMSE=7.1, MAE=5.11, MSE=50.52 and R=0.95) which implies the absolute superiority of the CART model than the AdaBoost model.

Conclusion: The results of this study showed that regression decision tree model can be used as an efficient model for modelling and estimation of urban air quality index.

Keywords: Simulation, Air Quality Index, Air Pollution, Decision Tree, Adaptive Boosting Algorithm.

► **Citation:** Ehsanzadeh A, Nejadkoorki F, Talebi A. A study on the Simulation and Estimation of the urban air quality by using artificial intelligence (case study: Gholhak Station of Tehran. *Journal of Research in Environmental Health*. Winter 2015;1 (4) : 285-296.

شبیه‌سازی و برآورد کیفیت بهداشتی هوای شهری با استفاده از هوش مصنوعی (مطالعه موردی: ایستگاه قلهمک شهر تهران)

چکیده

زمینه و هدف: کیفیت نامطلوب هوا ناشی از وجود غلظت‌های زیاد آلاینده‌ها در کلان شهر تهران موجب ایجاد بیماری‌های مختلف و مشکلات فراوان برای سلامتی و رفاه عمومی این کلان‌شهر گردیده است و همچنین موجب آسیب به محیط زیست و موجودات زنده می‌شود.

مواد و روش‌ها: شاخص کیفیت هوا (AQI) یک ابزار کلیدی جهت آگاهی از کیفیت هوا، نحوه اثر آلودگی هوا بر سلامت و روش‌های محافظتی در برابر آلودگی هوا است. هدف اصلی این تحقیق مدل‌سازی و برآورد شاخص کیفیت هوا با درخت تصمیم و الگوریتم بوستینگ تطبیقی است. از داده‌های ساعتی غلظت آلاینده‌های هوا و پارامترهای هواشناسی ایستگاه قلهمک شهر تهران به منظور مدل‌سازی و برآورد AQI استفاده شد.

یافته‌ها: نتایج تحقیق نشان داد که مدل درخت تصمیم عملکرد بهتری نسبت به مدل بوستینگ تطبیقی دارد. برای ارزیابی نتایج مدل‌های مذکور، ریشه میانگین مربعات خطا (RMSE)، میانگین مطلق خطا (MAE)، میانگین مربعات خطا (MSE) و ضریب همبستگی (R) در مدل درخت تصمیم برای مرحله آزمون به ترتیب ۰/۷۵، ۰/۱۰۱، ۰/۵۶۳ و ۰/۹۹ حاصل شد که در مقایسه با مدل بوستینگ تطبیقی (۰/۷۱، ۰/۱۱، ۰/۵۲، ۰/۹۵) حاکی از برتری مطلق نتایج درخت تصمیم رگرسیون نسبت به مدل بوستینگ تطبیقی است.

نتیجه‌گیری: نتایج حاصل از این مطالعه نشان داد که می‌توان از مدل درخت تصمیم رگرسیون به عنوان یک مدل کارآمد جهت مدل‌سازی و برآورد شاخص کیفیت هوا شهری استفاده نمود.

کلیدواژه‌ها: شبیه‌سازی، شاخص کیفیت هوا، آلودگی هوا، درخت تصمیم، الگوریتم بوستینگ

علیرضا احسان زاده

دانشجوی کارشناسی ارشد مهندسی محیط زیست،
دانشکده منابع طبیعی، دانشگاه یزد

فرهاد نژادکورکی

* دانشیار گروه مهندسی محیط زیست، دانشکده منابع
طبیعی، دانشگاه یزد (نویسنده مسئول)
f.nejadkoorki@yazd.ac.ir

علی طالبی

دانشیار گروه مهندسی آبخیزداری، دانشکده منابع
طبیعی، دانشگاه یزد

◀ **استناد:** احسان زاده ع، نژادکورکی ف، طالبی ع. شبیه‌سازی و برآورد کیفیت بهداشتی هوای شهری با استفاده از هوش مصنوعی (مطالعه موردی: ایستگاه قلهمک شهر تهران). *فصلنامه پژوهش در بهداشت محیط*. زمستان ۱۳۹۴؛ ۱(۴): ۲۸۵-۲۹۶.

تاریخ دریافت: ۱۳۹۴/۰۹/۲۴

تاریخ پذیرش: ۱۳۹۴/۱۱/۱۳

کیفیت هوا مرتبط با وضعیت شیمیایی اتمسفر در هر زمان و مکان است. همانند آب و هوا، کیفیت هوا نیز بر همگان اثرگذار است. کیفیت نامطلوب هوا، دارای پتانسیل ایجاد بیماری‌های تنفسی و قلبی عروقی در انسان می‌باشد (۱). کیفیت نامطلوب هوا ناشی از وجود غلظت‌های زیاد آلاینده‌ها در کلان شهر تهران موجب ایجاد بیماری‌های مختلف و مشکلات فراوان برای سلامتی و رفاه عمومی این کلان‌شهر گردیده است و همچنین موجب آسیب به محیط زیست و موجودات زنده می‌شود. راهبردهای مختلفی جهت کنترل و مدیریت آلودگی هوا وجود دارد. شاخص‌های سنجش کیفیت هوا به‌طور گسترده در طرح‌های کنترل کیفیت هوا مورد استفاده قرار می‌گیرند. این شاخص‌ها کیفیت هوا را برحسب میزان آلودگی و آلاینده‌های مختلف، طبقه‌بندی می‌نماید. اولین شاخص سنجش کیفیت هوا، شاخص آلودگی هوا (PSI) نام دارد، که توسط آژانس حفاظت محیط زیست آمریکا (US-EPA) توسعه داده شد، که غلظت آلاینده‌های اصلی هوا نظیر ذرات معلق، منوکسیدکربن، ازن، دی اکسید گوگرد و دی اکسید نیتروژن را به شاخص استاندارد آلودگی هوا تبدیل می‌نماید. در سال ۱۹۹۷ شاخص PSI توسط US-EPA کاملتر گردید و تحت شاخصی به نام شاخص کیفیت هوا (AQI) جایگزین گردید (۲). مدیریت پایش و نظارت بر کیفیت هوا در شهرهای بزرگ داده‌های مربوط به کیفیت هوا را به AQI تبدیل می‌کند و اطلاعات مورد نیاز را در اختیار عموم مردم قرار می‌دهد. بنابراین شاخص کیفیت هوا یک ابزار کلیدی جهت آگاهی از کیفیت هوا، نحوه اثر آلودگی هوا بر سلامت و روش‌های محافظتی در برابر آلودگی هوا است (۳). یکی از اقدامات اولیه‌ای که جهت کنترل آلودگی هوا می‌بایست صورت پذیرد، اندازه گیری غلظت آلاینده‌های هوا از جمله ذرات معلق، منوکسیدکربن، ازن، دی اکسیدگوگرد و اکسیدهای نیتروژن می‌باشد، ولی صرف اندازه گیری غلظت این آلاینده‌ها جز داده‌های خام چیزی حاصل ندارد، استفاده از روشی که رابطه میان غلظت آلاینده‌ها و سطح سلامت جامعه و اقدامات کنترلی مربوط

به آلاینده‌های هوا را تعیین نماید، امروزه تحت شاخصی بنام AQI مورد استفاده قرار می‌گیرد (۴). از نمونه مطالعاتی که از روش‌های هوش مصنوعی و آماری جهت پیش‌بینی و مدل‌سازی پارامترهای کیفیت هوا به کار برده شده می‌توان به موارد زیر اشاره کرد: استفاده از روش تحلیل مؤلفه‌های اصلی و استفاده از شبکه عصبی مصنوعی جهت پیش‌بینی شاخص روزانه کیفیت هوا (۵)، استفاده از سیستم استنتاج فازی و مدل اتورگرسیون به منظور ارزیابی و پیش‌بینی کیفیت هوا (۶)، کاربرد رگرسیون خطی چندگانه مبتنی بر تحلیل مؤلفه‌های اصلی جهت پیش‌بینی کوتاه مدت شاخص کیفیت هوا (۷)، کاربرد الگوریتم یادگیری جمعی جهت شناسایی منابع آلاینده هوا و مقدار شاخص کیفیت هوا (۸)، کاربرد شبکه عصبی مصنوعی و روش انتخاب متغیرهای تصادفی به منظور پیش‌بینی کیفیت هوای شهری (۹)، پیش‌بینی غلظت روزانه مونوکسیدکربن با استفاده از روش رگرسیون خطی چندگانه براساس تحلیل مؤلفه‌های اصلی و همچنین استفاده از شبکه عصبی مصنوعی (۱۰) و پیش‌بینی و مدل‌سازی غلظت آلاینده مونوکسیدکربن با تلفیق شبکه عصبی - فازی تطبیقی و سیستم اطلاعات جغرافیایی (۱۱). نتایج اکثر این تحقیقات حاکی از برتری مدل‌های هوش محاسباتی نسبت به مدل‌های آماری است. با توجه به بررسی نویسندگان تاکنون در داخل و خارج کشور استفاده از درخت تصمیم و الگوریتم بوستینگ در برآورد شاخص کیفیت هوا استفاده نشده است. با این حال یکی از بهترین روش‌ها که در زمینه تعیین کیفیت هوا امروزه در سراسر دنیا مورد استفاده قرار می‌گیرد، تبدیل غلظت آلاینده‌ها به شاخص کیفیت هوامی‌باشد، در واقع AQI یک شاخص مفید جهت آگاهی از کیفیت هوا، میزان اثر آلاینده‌ها بر سلامت و روش‌های مختلف کنترلی در برابر آلودگی هوا را مشخص می‌نماید (۱۲). کیفیت نامطلوب هوا ناشی از وجود غلظت‌های زیاد آلاینده‌ها در کلان شهر تهران موجب ایجاد بیماری‌های مختلف و مشکلات فراوان برای سلامتی و رفاه عمومی این کلان‌شهر گردیده است. با این رویکرد برآورد و مدل‌سازی کیفیت هوای شهری که دارای ماهیت

غیرخطی است و همچنین تعیین عناصر موثر بر آن و اطلاع رسانی به عموم مردم از ضروری ترین برنامه های زیست محیطی در کلان شهر تهران به شمار می رود. هدف اصلی این تحقیق مقایسه کارایی درخت تصمیم و الگوریتم بوستینگ تطبیقی جهت شبیه سازی و برآورد شاخص کیفیت هوا در ایستگاه قلعهک شهر تهران می باشد.

روش کار

در این تحقیق که از نوع کاربردی می باشد از داده های ساعتی غلظت آلاینده های هوا و پارامترهای هواشناسی مربوط به ایستگاه قلعهک شهر تهران بدلیل اینکه یکی از آلوده ترین مناطق شهر تهران می باشد و همچنین پارامترهای هواشناسی و آلودگی هوا را به صورت همزمان ثبت می نماید، جهت برآورد و شبیه سازی AQI استفاده خواهد شد. هدف اولیه استفاده از دستورالعمل US-EPA و مرکز سلامت و محیط کار ایران جهت محاسبه شاخص کیفیت هوا براساس غلظت های ساعتی مربوط به تک تک آلاینده ها می باشد و در مرحله بعد با استفاده از سری های زمانی مربوط به داده های هواشناسی، آلودگی هوا و میزان AQI محاسبه شده، نسبت به ایجاد و توسعه مدل های برآوردگر و شبیه ساز کیفیت هوا با استفاده از روش های درخت تصمیم و الگوریتم بوستینگ تطبیقی در محیط نرم افزار MATLAB اقدام خواهد شد. در اولین مرحله غلظت تک تک آلاینده ها به عنوان ورودی الگوریتم محاسبه AQI بوده و خروجی این الگوریتم که شاخص کیفیت هوای مربوط به هر آلاینده و شاخص کلی کیفیت هوا است همراه با داده های هواشناسی برای توسعه مدل ها استفاده خواهد شد و هدف نهایی شبیه سازی و برآورد شاخص کیفیت هوا برای ایستگاه قلعهک شهر تهران است. در آخر مقایسه ای بین مدل های به کار رفته در تحقیق خواهد شد و مدلی که نتایج بهتری را به منظور برآورد و شبیه سازی ارائه نماید، معرفی می شود.

داده های تحقیق

داده های مورد استفاده در تحقیق شامل غلظت های ساعتی

مربوط به آلاینده های هوا شامل: مونوکسیدکربن (CO)، ذرات معلق کوچکتر از ۱۰ میکرون (PM₁₀)، دی اکسید نیتروژن (NO₂)، ازن (O₃)، دی اکسید گوگرد (SO₂)، مونو کسید نیتروژن (NO)، کل هیدروکربن ها (THC) و متان (CH₄) است. همچنین شامل برخی پارامترهای هواشناسی نظیر: سرعت باد (WS)، جهت باد (WD)، دمای هوا (T)، فشار هوا (P) و رطوبت هوا (H) است. داده های موجود مربوط به دوره زمانی ۱۳۸۵ تا ۱۳۹۰ است.

شاخص کیفیت هوا

مدیریت پایش و نظارت بر کیفیت هوا در شهرهای بزرگ داده های خام اندازه گیری شده توسط دستگاه های سنجش غلظت آلاینده های هوا را به AQI تبدیل نموده و اطلاعات مورد نیاز مربوط به سطوح سلامتی انسان و اثرات بهداشتی آلاینده های هوا را توسط شاخص AQI در اختیار مردم قرار می دهند. AQI یک ابزار کلیدی جهت آگاهی از کیفیت هوا، نحوه اثر آلودگی هوا بر سلامت و روش های محافظتی در برابر آلودگی هوا است. AQI برای پنج آلاینده اصلی هوا یعنی ذرات معلق، دی اکسید نیتروژن، ازن سطح زمین، منو کسید کربن و دی اکسید گوگرد محاسبه می شود (۱). بمنظور محاسبه این شاخص از معادله (۱) استفاده می شود، که غلظت آلاینده ها و نقاط شکست جهت محاسبه طبقه کیفیت هوا مطابق با دستورالعمل مرکز سلامت و محیط کار ایران است. US-EPA به منظور درک راحت تر مقدار شاخص کیفیت هوا و سطوح بهداشتی مختلف با آن و همچنین دستورالعمل های کنترلی مربوطه با مقادیر مختلف AQI را به شش دسته طبقه بندی نموده است که هر دسته را به سطوح مختلف سلامت انسان مربوط می سازد؛ این شش دسته به شرح جدول (۱) می باشد.

$$p = \frac{I_H - I_b}{B_H - B_b} (C_p - B_b) + I_b \quad \text{معادله (۱)}$$

که در این رابطه :

Ip = شاخص کیفیت هوا برای آلاینده Cp، P = غلظت

اندازه گیری شده آلاینده BPHi، P_i = نقطه شکستی که بزرگتر یا CP باشد، I_{Hi} = مقدار AQI منطبق با I_{Lo} ، BPHi، مقدار AQI مساوی Cp است، BPLo = نقطه شکستی که کوچکتر یا مساوی منطبق با BPLo.

جدول ۱: نقاط شکست برای محاسبه AQI

نقاط شکست $BP_{Lo} - BP_{Hi}$							AQI $I_{Lo} - I_{Hi}$
$O_3^{(i)}$ (ppm) ۸ ساعته	O_3 (ppm) یک ساعته	PM_{10} ($\mu g/m^3$) ۲۴ ساعته	$PM_{2.5}$ ($\mu g/m^3$) ۲۴ ساعته	CO (ppm) ۸ ساعته	SO ₂ (ppm) ۲۴ ساعته	NO ₂ (ppm) یک ساعته	
۰-۰/۰۵۹	-	۰/۰-۱۵/۴	۰-۵۴	۰/۰-۴/۴	۰/۰-۰/۰۳۴	۰-۰/۰۵۳	۰-۵۰
۰/۰۶۰-۰/۰۷۵	-	۱۵/۵-۳۵	۵۵-۱۵۴	۴/۵-۹/۴	۰/۰۳۵-۰/۱۴۴	۰/۰۵۴-۰/۱	۵۱-۱۰۰
۰/۰۷۶-۰/۰۹۵	۰/۱۲۵-۰/۱۶۴	۳۵/۱-۶۵/۴	۱۵۵-۲۵۴	۹/۵-۱۲/۴	۰/۱۴۵-۰/۲۲۴	۰/۱۰۱-۰/۲۶۰	۱۰۱-۱۵۰
۰/۰۹۶-۰/۱۱۵	۰/۱۶۵-۰/۲۰۴	۶۵/۵-۱۵۰/۴	۲۵۵-۳۵۴	۱۲/۵-۱۵/۴	۰/۲۲۵-۰/۳۰۴	۰/۳۶۱-۰/۶۴۰	۱۵۱-۲۰۰
۰/۱۱۶-۰/۳۷۴	۰/۲۰۵-۰/۴۰۴	۱۵۰/۵-۲۵۰/۴	۳۵۵-۴۲۴	۱۵/۵-۳۰/۴	۰/۳۰۵-۰/۶۰۴	۰/۶۵۱-۱/۲۴	۲۰۱-۳۰۰
(۲)	۰/۴۰۵-۰/۵۰۴ ۰/۵۰۵-۰/۶۰۴	۲۵۰/۵-۳۵۰/۴ ۳۵۰/۵-۵۰۰/۴	۴۲۵-۵۰۴ ۵۰۵-۶۰۴	۳۰/۵-۴۰/۴ ۴۰/۵-۵۰/۴	۰/۶۰۵-۰/۸۰۴ ۰/۸۰۵-۱/۰۰۴	۱/۲۵-۱/۶۴ ۱/۲۶-۲/۰۴	۳۰۱-۴۰۰ ۴۰۱-۵۰۰

(۱) در بیشتر مناطق AQI بر اساس مقادیر ازن ۸ ساعته گزارش می‌شود اما در برخی مناطق AQI بر اساس مقادیر ازن یک ساعته به احتیاط نزدیک‌تر است. در این شرایط AQI می‌بایست هم برای مقادیر ازن ۸ ساعته و هم برای ازن ۱ ساعته محاسبه شود هر کدام بیشتر بود گزارش شود.
(۲) وقتی غلظت ازن ۸ ساعته از 0.374 ppm فراتر رود مقادیر AQI بیش از ۳۰۰ باید با استفاده از غلظت ازن ۱ ساعته محاسبه شود.

تصمیم CART جهت مدل‌سازی و برآورد شاخص کیفیت هوا استفاده شده است. ساختار این درخت بر سه اصل استوار است: ۱. مجموعه‌ای از سوالات به شکل $x \leq d$? که در آن x یک متغیر مستقل و d یک مقدار ثابت است و جواب هر سوال بله/خیر است. ۲. بهترین معیار شاخه زدن جهت انتخاب بهترین متغیر مستقل برای ایجاد شاخه جدید. ۳. ایجاد آمار خلاصه برای گره انتهایی.

$$SS(t) = \sum_{i=1}^{N(t)} [(y_i(t) - \hat{y})(t)]^2$$

معیارهای متفاوتی جهت ایجاد شاخه و تولید درخت تصمیم وجود دارد (۱۳، ۱۴). اما از آنجا که پژوهش حاضر به استفاده از درخت تصمیم رگرسیونی پرداخته، معیار مورد استفاده در این مدل براساس مجموع مربع خطا (SSE) می‌باشد. نتیجه تقسیم باید به طور ایده آل منجر به مجموعه‌هایی شود که مقدار پاسخ به میانگین گروه نزدیک باشد. کمترین مقدار SSE برای گروه نشان دهنده نزدیک بودن مقادیر گروه به میانگین مجموعه است. برای هر شکاف بالقوه مقدار SSE برای هر گره حاصل محاسبه می‌شود. امتیاز هر شکاف به وسیله مجموع مقادیر SSE محاسبه شده برای

درخت تصمیم رگرسیون

درختان تصمیم قادر به تولید توصیفات قابل درک برای انسان، از روابط موجود در یک مجموعه داده‌ای هستند و می‌توانند برای وظایف دسته‌بندی و پیش‌بینی بکار روند. این ساختار تصمیم‌گیری می‌تواند به شکل تکنیک‌های ریاضی و محاسباتی که به توصیف، دسته‌بندی و عام‌سازی یک مجموعه از داده‌ها کمک می‌کنند نیز معرفی شوند. داده‌ها در رکوردهایی به شکل $(x, y) = (x_1, x_2, x_3, \dots, x_k, y)$ داده می‌شوند. با استفاده از متغیرهای $x_1, x_2, \dots, x_k$ سعی در درک، دسته‌بندی یا عام‌سازی متغیر وابسته Y داریم. انواع صفات در درخت تصمیم به دو نوع صفات دسته‌ای و صفات حقیقی بوده که صفات دسته‌ای، صفاتی هستند که دو یا چند مقدار گسسته می‌پذیرند (یا صفات سمبلیک) درحالی که صفات حقیقی مقادیر خود را از مجموعه اعداد حقیقی می‌گیرند (۱۳). درخت CART نامی است که به هر دو روال بالا اطلاق می‌شود. نام CART سرنام کلمات درختان رگرسیون و دسته‌بندی است. در این تحقیق از الگوریتم درخت

$$MAE = 1/N \sum_{i=1}^N |P_i - O_i| \quad \text{معادله (۷)}$$

$$MSE = 1/N \sum_{i=1}^N (P_i - O_i)^2 \quad \text{معادله (۸)}$$

$$R = \frac{\sum_{i=1}^N (P_i - \bar{O})^2 - \frac{(\sum_{i=1}^N (P_i - O_i))^2}{N}}{\sum_{i=1}^N (P_i - \bar{O})^2} \quad \text{معادله (۹)}$$

$$R^2 = (R)^2 \quad \text{معادله (۱۰)}$$

N: تعداد کل داده‌ها، P_i مقادیر پیش‌بینی شده، O_i مقادیر مشاهده یا محاسبه شده، $\bar{O}$ میانگین مقادیر مشاهده یا محاسبه. شاخص IA میزان صحت مدل را نشان می‌دهد که عدد ایده‌آل آن ۱ است، شاخص FB نشان دهنده این است که مدل بالاتر از مقدار محاسبه شده برآورد می‌نماید یا پایین‌تر از آن، عدد ایده‌آل آن صفر است، شاخص‌های MAE، RMSE و MSE میزان خطای مدل در مرحله آموزش و آزمون را نشان می‌دهد که مقدار عدد مطلوب آن ۰ است، شاخص R میزان همبستگی بین مقادیر برآورد شده و محاسبه شده را نشان می‌دهد که بین ۱- تا ۱+ متغیر است و هرچه به عدد ۱+ نزدیک‌تر باشد مطلوب‌تر است و شاخص R^2 ضریب تبیین مدل را نشان می‌دهد و عدد ۱ مقدار ایده‌آل آن است.

یافته‌ها

نتایج حاصل از بررسی آماری داده‌های ایستگاه قل‌هک در جدول (۲) و در ادامه نتایج حاصل از مدل‌سازی و برآورد شاخص کیفیت هوای شهری در ایستگاه پایش کیفیت هوای قل‌هک ارائه گردیده است. شبیه‌سازی و برآورد به وسیله روش‌های درخت تصمیم رگرسیون و الگوریتم بوستینگ تطبیقی انجام شد.

جدول ۲: مشخصات آماری هر یک از پارامترهای ایستگاه قل‌هک

نام پارامتر	WS	WD	T	P	H	THC	CH _۴	NO	O _۳	CO	SO _۲	NO _۲	PM _{۱۰}
واحد	m/s	Deg	DegC	mBar	RH%	ppm	ppm	ppm	ppm	ppm	ppm	ppm	ug/m ^۳
حداقل	۰/۰۰۳	۰/۰	-۷/۶۲۹	۴۰۳/۹	۰/۳۷	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱	۰/۰۱۸	۰/۰۰	۰/۰۰۱	۰/۰۰
حداکثر	۴/۵	۳۲۵/۷	۳۹/۵۱	۹۵۷/۶	۹۵/۳۴	۱۷/۳۳	۴/۷۹۰	۰/۳۸۶	۰/۰۵۰	۹/۰۹۲	۰/۰۸	۰/۴۲۸	۳۱۶/۸
میانگین	۰/۹۰	۱۱۶/۳	۱۸/۲۸	۸۷۳/۹	۴۴/۳۳	۳/۵۳	۲/۱۱۲	۰/۰۸۸	۰/۰۱۵	۳/۵۵	۰/۰۳۸	۰/۱۲۰	۸۹/۶۹
جولگی	۰/۹۷	۰/۰۶	-۰/۲۴	-۲/۲۴	۰/۳۲	۰/۹۷	۰/۰۴۴	۱/۰۲	۱/۲۱	۰/۵۶۲	۰/۰۵	۱/۲۴	۱/۳۹
کشیدگی	۱/۱۳	-۱/۳۱	-۰/۷۸	۲۱/۷	-۰/۶۳	۲/۹	-۰/۱۹	۰/۷۹	۰/۸۸	۰/۲۰	۰/۲۳	۱/۶۵	۲/۰۵

ورودی و خروجی مدل‌ها

در این پژوهش از داده‌های ساعتی آلاینده‌های هوا و پارامترهای هواشناسی ساعتی به عنوان ورودی مدل‌ها در مراحل آموزش و آزمون استفاده شد. همچنین خروجی تمامی مدل‌ها مقدار شاخص کیفیت هوا در نظر گرفته شد، این شاخص به صورت ساعتی در محیط نرم افزار MATLAB محاسبه و پس از آن به عنوان پارامتر هدف در مدل‌ها مورد استفاده قرار گرفت. در این پژوهش ۸۰ درصد کل داده‌ها جهت آموزش و ۲۰ درصد داده‌ها جهت آزمون مدل‌ها مورد استفاده قرار گرفت. تعداد کل نمونه‌ها شامل ۱۹۸۴۰ ساعت می‌باشد. تقسیم بندی داده‌ها بصورت تصادفی و بعد از حذف داده‌های گمشده و پرت انجام شد.

ارزیابی اعتبار مدل‌ها

جهت ارزیابی عملکرد مدل‌ها و مقایسه نتایج بدست آمده در مراحل آموزش و آزمون از شاخص‌های آماری نظیر شاخص توافق (IA)، بایاس (FB)، جذر میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE)، میانگین مربعات خطا (MSE)، ضریب همبستگی (R) و ضریب تبیین (R^2) که معادلات مربوطه به شرح زیر است.

$$\text{معادله (۴)}$$

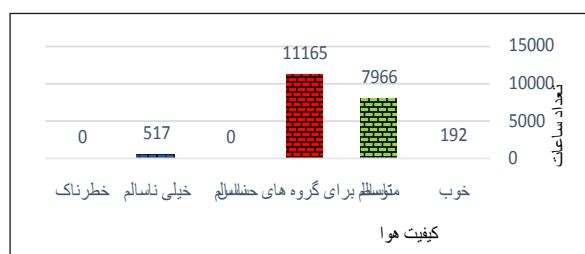
$$\tilde{H} = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N \sum_{i=1}^N (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad \text{معادله (۵)}$$

$$B = 1/N \sum_{i=1}^N (P_i - O_i) / (P_i + O_i) / 2 \quad \text{معادله (۶)}$$

$$RMSE = (1/N \sum_{i=1}^N (P_i - O_i)^2)^{0.5}$$

برآورد شاخص کیفیت هوا

نتایج تحلیل حاصل از محاسبه شاخص کیفیت هوا در ایستگاه قلپک در شکل (۲) نشان داده شده است. بیشترین طبقه کیفیت هوا مربوط به طبقه ناسالم برای گروه‌های حساس جامعه با تعداد ساعات ۱۱۱۶۵ می‌باشد و آلاینده مسئول ایجاد کیفیت نامطلوب در اکثر اوقات دی اکسید نیتروژن می‌باشد، که عامل اصلی آن حجم ترافیک بالای خودروها و جمعیت بیشتر و احتراق سوخت در موتور خودروها در محدوده این ایستگاه می‌باشد. با توجه به شکل (۲) شرایط خیلی ناسالم شامل ۵۱۷ ساعت می‌باشد، مواقعی که وارونگی هوا و مه‌دود فتوشیمیایی اتفاق می‌افتد، عامل ایجاد شرایط خیلی ناسالم کیفیت هوا می‌باشد.



شکل ۲: مقدار شاخص و طبقه کیفیت هوای ایستگاه قلپک

مدل‌سازی شاخص کیفیت هوا

نتایج مدل‌سازی نشان داد که کارایی روش‌های به کار رفته در تحقیق جهت برآورد شاخص کیفیت هوای شهری عملکرد متفاوتی دارند. برای مدل‌سازی و برآورد شاخص کیفیت هوا از تمام پارامترهای هواشناسی و آلودگی هوای ثبت شده در ایستگاه قلپک شهر تهران استفاده شد. شاخص کیفیت هوا با دستورالعمل سازمان حفاظت محیط‌زیست آمریکا و مرکز سلامت و محیط کار ایران محاسبه شد. جهت ساخت مدل‌ها داده‌ها به صورت

تصادفی به دو دسته آموزش و آزمون تقسیم گردید. تعداد نمونه‌ها در مرحله آموزش ۱۵۸۷۲ و در مرحله آزمون ۳۹۶۸ و شامل ۱۳ پارامتر ورودی و یک پارامتر خروجی است.

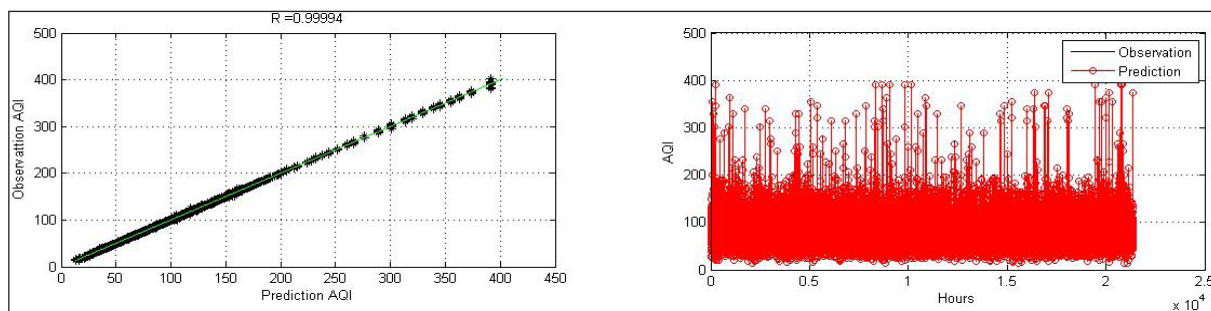
مدل درخت تصمیم

در این قسمت نتایج حاصل از کاربرد الگوریتم CART در مدل‌سازی و برآورد شاخص کیفیت هوای ایستگاه قلپک شهر تهران ارائه شده است. در جدول (۳) مقادیر شاخص‌های آماری حاصل از سه مرحله آموزش، آزمون قبل از هرس و آزمون بعد از هرس درخت تصمیم ارائه شده است. جهت هرس درخت از معیار پیچیدگی استفاده شد و مقدار آن پس از آزمون و خطای انجام شده بر روی درخت به مقدار ۰/۰۳ جهت تعیین بهترین ساختار درخت مشخص گردید. نتایج بدست آمده حاکی از این است که الگوریتم CART از عملکرد بسیار بالایی در شبیه‌سازی شاخص کیفیت هوا برخوردار است. بطوری که همبستگی بین مقادیر شبیه‌سازی شده و مشاهده شده بسیار نزدیک به عدد ۱ است. مقادیر خطای درخت در سه مرحله آزمون و آموزش بسیار ناچیز و مشابه است. مهمترین پارامترهای تاثیر گذار جهت ساخت درخت تصمیم رگرسیون در برآورد AQI در ایستگاه قلپک ذرات معلق کوچکتر از ۱۰ میکرون با ۱۸/۵ درصد، دی‌اکسید نیتروژن با ۳۲/۲ درصد و مونوکسید نیتروژن با ۶ درصد تاثیر می‌باشند. در شکل (۳)، (۴) و (۵) بترتیب نتایج شبیه‌سازی کیفیت هوا در مراحل آموزش، آزمون قبل از هرس و بعد از هرس درخت تصمیم رگرسیون ارائه شده است. در شکل (۶) مهمترین پارامترهای تاثیر گذار در برآورد AQI ارائه شده است.

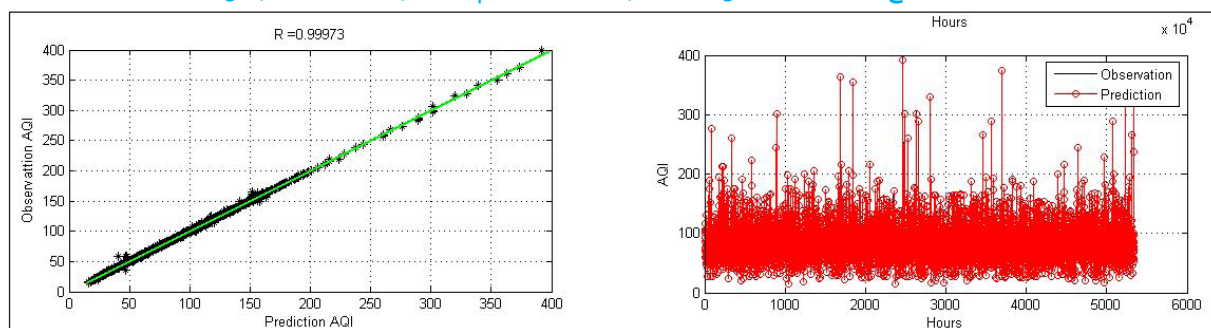
جدول ۳: نتایج حاصل از ارزیابی مدل درخت تصمیم

مدل درخت تصمیم	MSE (۰)	RMSE (۰)	MAE (۰)	R (۱)	R ² (۱)	FB (۰)	IA (۱)
مرحله آموزش	۰/۰۶۷	۰/۲۵۹۳	۰/۰۴۰۸	۰/۹۹۹۹	۰/۹۹۹۹	۰	۱
مرحله آزمون قبل از هرس درخت	۰/۵۶۳	۰/۷۵	۰/۱۰۱۲	۰/۹۹۹۶	۰/۹۹۹۱	۰	۱
مرحله آزمون بعد از هرس درخت	۱/۷۳	۱/۳۱	۰/۸۸۴	۰/۹۹۸۶	۰/۹۹۷۲	۰	۱

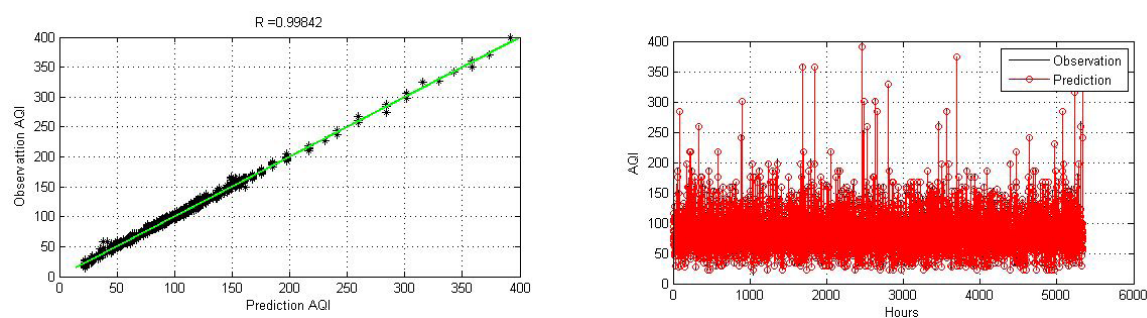
اعداد داخل پرانتز مقادیر ایده‌آل هر شاخص را نشان می‌دهد



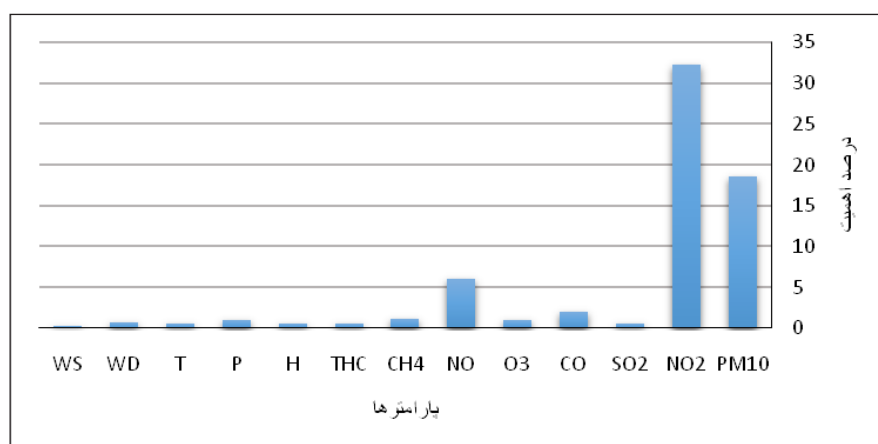
شکل ۳: نتایج شبیه‌سازی شاخص کیفیت هوا با درخت تصمیم رگرسیون در مرحله آموزش



شکل ۴: نتایج شبیه‌سازی شاخص کیفیت هوا با درخت تصمیم رگرسیون در مرحله آزمون قبل از هرس



شکل ۵: نتایج شبیه‌سازی شاخص کیفیت هوا با درخت تصمیم رگرسیون در مرحله آزمون بعد از هرس



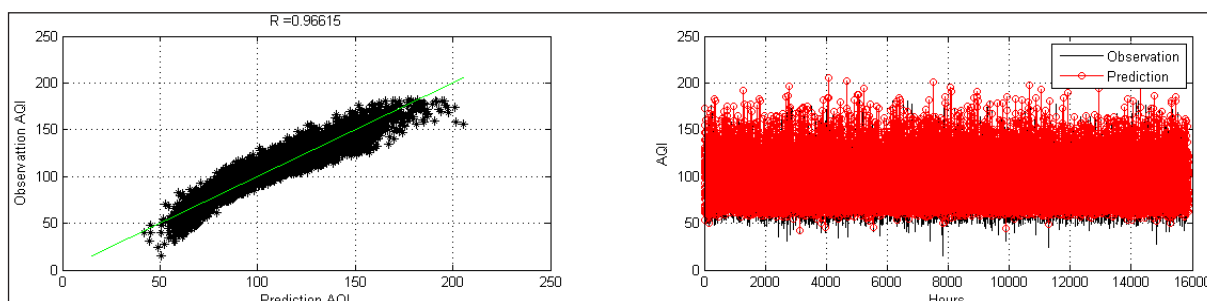
شکل ۶: مهمترین پارامترهای تاثیرگذار در برآورد شاخص کیفیت هوا

مدل بوستینگ تطبیقی

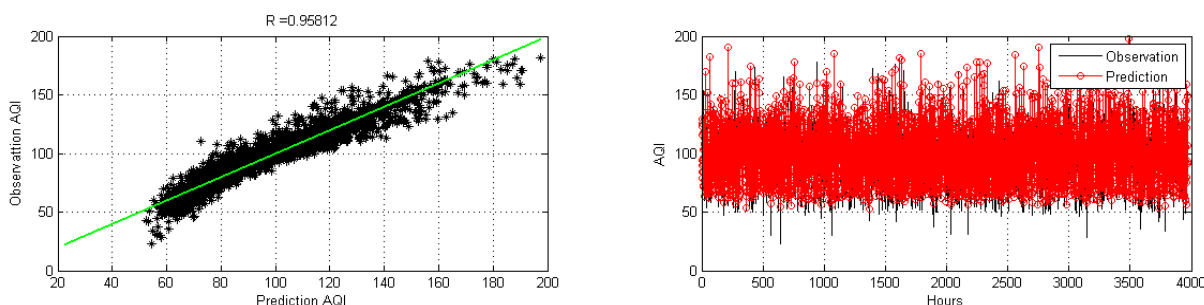
در این قسمت نتایج حاصل از کاربرد الگوریتم بوستینگ تطبیقی در مدل سازی و برآورد شاخص کیفیت هوای ایستگاه قلعهک شهر تهران ارائه شده است. در جدول (۴) مقادیر شاخص های آماری حاصل از مرحله آموزش و آزمون ارائه شده است. نتایج بدست آمده

جدول ۴: نتایج حاصل از ارزیابی مدل بوستینگ تطبیقی

مدل بوستینگ تطبیقی	(-) MSE	(-) RMSE	(-) MAE	(-) R	(-) R ²	(-) FB	(-) IA
مرحله آموزش	۴۲/۰۳	۶/۴۸	۴/۷۲	۰/۹۶۶۲	۰/۹۳۳۵	۰	۱
مرحله آزمون	۵۰/۵۲	۷/۱	۵/۱۱	۰/۹۵۸۱	۰/۹۱۸۰	۰/۰۰۱۹	۱



شکل ۷: نتایج شبیه سازی شاخص کیفیت هوا با بوستینگ تطبیقی در مرحله آموزش



شکل ۸: نتایج شبیه سازی شاخص کیفیت هوا با بوستینگ تطبیقی در مرحله آزمون

بحث

مقدار آماره های خطای نظیر RMSE و MAE مدل رگرسیون درخت تصمیم در بهینه ترین حالت در مرحله آموزش به ترتیب برابر ۰/۲۵۹۳ و ۰/۰۴۰۸ و در مرحله آزمون به ترتیب ۰/۷۵ و ۰/۱۰۱۲ می باشد و مقدار آماره های خطای نظیر RMSE و MAE مدل بوستینگ تطبیقی در بهینه ترین حالت در مرحله آموزش برابر ۶/۴۸ و ۴/۷۲ و در مرحله آزمون به ترتیب ۷/۱ و ۵/۱۱ می باشد. با مقایسه این آماره ها نشان می دهد که مدل درخت تصمیم

رگرسیون از عملکرد بهتری نسبت به مدل بوستینگ تطبیقی برخوردار است. آماره های ضریب همبستگی R و ضریب تبیین R² مدل درخت تصمیم رگرسیون در مرحله آموزش ۰/۹۹۹۹ و ۰/۹۹۹۹ و در مرحله آزمون ۰/۹۹۹۶ و ۰/۹۹۹۱ می باشد و در مدل بوستینگ تطبیقی در مرحله آموزش ۰/۹۶۶۲ و ۰/۹۳۳۵ و در مرحله آزمون ۰/۹۵۸۱ و ۰/۹۱۸۰ می باشد. با توجه به معیارهای همبستگی نشان از این دارد که بین مقادیر برآورد شده مدل و محاسبه شده کیفیت هوا در مدل درخت تصمیم رگرسیون

فازی با مقدار شاخص همبستگی 0.74 و ریشه میانگین مربعات خطا $3/19$ می‌باشد (۶).

نتیجه‌گیری:

در این تحقیق مطالعه در سطح یک ایستگاه پایش کیفیت هوا انجام شد، به دلیل اهمیت بسیار زیاد موضوع کیفیت هوا در شهرهای آلوده نظیر تهران پیشنهاد می‌شود در تحقیقات مشابه آتی مطالعه در سطح بسیار گسترده‌تر انجام شود و جهت درک بهتر موضوع کیفیت هوا در نقاط بسیار آلوده سطح شهر نقشه‌های پهنه‌بندی کیفیت هوا از طریق روش‌های زمین آمار و مدل‌های برآوردگر کیفیت هوا نظیر شبکه عصبی مصنوعی و درخت تصمیم در نرم افزار سامانه اطلاعات جغرافیایی (GIS) تولید گردد. با توجه به اینکه یکی از مشکلات همیشگی در تحقیقات وجود داده‌های گمشده در بین مجموعه داده‌های موجود است، لذا پیشنهاد می‌شود با استفاده از روش درخت تصمیم اقدام به تخمین داده‌های گمشده و بازسازی داده‌های ناقص اقدام گردد. جهت واسنجی دقیق‌تر درخت تصمیم به عنوان مناسب‌ترین مدل در این مطالعه توصیه می‌شود در چند ایستگاه با شرایط محیطی مختلف این مدل اجرا شده تا در صورت اثبات دقت این مدل در برآورد شاخص کیفیت هوا بتوان این روش را به عنوان یک روش پایه در مطالعات پایش و برآورد کیفیت هوا در سطح کشور معرفی نمود. دست‌آوردهای این تحقیق می‌تواند به عنوان یک ابزار مورد استفاده مدیران و برنامه‌ریزان شهری جهت بهبود و اقدامات کنترلی برای پیشگیری از آلودگی هوای شهری و اطلاع رسانی کیفیت و درجه سلامت هوای تنفسی به شهروندان مناطق آلوده شهری قرار گیرد.

تشکر و قدردانی

بدینوسیله از سازمان حفاظت محیط زیست به خاطر در اختیار گذاشتن داده‌های تحقیق تشکر و قدردانی می‌نمایم.

اختلاف بسیار ناچیز است. آماره‌های ضریب همبستگی R و ضریب تبیین R^2 در مدل درخت تصمیم رگرسیون نزدیک به عدد ۱ است که نشان از توانایی بالای مدل درخت تصمیم رگرسیون در برآورد شاخص کیفیت هوای شهری دارد. نتایج ارزیابی مدل‌های مورد استفاده جهت شبیه‌سازی و برآورد شاخص کیفیت هوای ایستگاه قلعه‌ک نشان می‌دهد که مدل درخت تصمیم رگرسیون عملکرد بهتری نسبت به مدل بوستینگ تطبیقی دارد. به طوری که مدل درخت تصمیم رگرسیون ۹۹ درصد مقادیر شاخص کیفیت هوا را در مرحله آزمون به صورت صحیح برآورد نموده است و مدل بوستینگ تطبیقی ۹۵ درصد مقادیر شاخص کیفیت هوا را در مرحله آزمون به صورت صحیح برآورد نموده است. نتایج این تحقیق همسو با مطالعه «سینگ» و همکاران که در سال ۲۰۱۳ با استفاده از روش‌های یادگیری جمعی نظیر درخت تصمیم رگرسیون و استفاده از پارامترهای آلودگی هوا و هواشناسی شهر لاکنو هند انجام شد می‌باشد (۸). همچنین، با بررسی شاخص‌های آماری مطالعه‌ای که توسط «کومار و گوپال» که در سال ۲۰۱۳ با روش شبکه عصبی مصنوعی و تحلیل مؤلفه‌های اصلی نسبت به پیش‌بینی شاخص کیفیت هوا در شهر دهلی هند انجام شده است مشخص شد که درخت تصمیم رگرسیون و الگوریتم بوستینگ عملکرد بهتری دارند و عوامل آلوده‌کننده هوا در پیش‌بینی شاخص کیفیت هوا تاثیر بیشتری نسبت به پارامترهای هواشناسی دارند (۵). همچنین، مطالعه‌ای که توسط «کومار و گوپال» در سال ۲۰۱۱ با استفاده از روش رگرسیون خطی و تحلیل مؤلفه‌های اصلی در جهت توسعه مدل پیش‌بینی کننده شاخص کیفیت هوا است، مطابقت دارد و شاخص‌های آماری R و $RMSE$ حاکی از عملکرد مطلقاً برتر مدل بوستینگ تطبیقی و درخت تصمیم استفاده شده در این مطالعه نسبت به مدل رگرسیون خطی است (۱۱). همچنین، با مقایسه آماره‌های خطا و همبستگی حاصل از این تحقیق با مطالعه «کارباجال» و همکاران که در سال ۲۰۱۱ در شهر مکزیکو سیتی انجام شد، نتایج بیانگر عملکرد بهتر مدل درخت تصمیم رگرسیون استفاده شده در این مطالعه نسبت به مدل منطق

References:

1. Zhang Y, Bocquet M, Mallet V, Seigneur C, Baklanov A. Real-time air quality forecasting, part I: History, techniques, and current status. *Atmospheric Environment*. 2012;60:632-55.
2. Sowlat MH, Gharibi H, Yunesian M, Tayefeh Mahmoudi M, Lotfi S. A novel, fuzzy-based air quality index (FAQI) for air quality assessment. *Atmospheric Environment*. 2011;45(12):2050-9.
3. Künzli N, Kaiser R, Medina S, Studnicka M, Chanel O, Filliger P, et al. Public-health impact of outdoor and traffic-related air pollution: a European assessment. *The Lancet*. 2000;356(9232):795-801.
4. Cheng W-L, Chen Y-S, Zhang J, Lyons T, Pai J-L, Chang S-H. Comparison of the revised air quality index with the PSI and AQI indices. *Science of the total environment*. 2007;382(2):191-8.
5. Kumar A, Goyal P. Forecasting of air quality index in Delhi using neural network based on principal component analysis. *Pure and Applied Geophysics*. 2013;170(4):711-22.
6. Carbajal-Hernández JJ, Sánchez-Fernández LP, Carrasco-Ochoa JA, Martínez-Trinidad JF. Assessment and prediction of air quality using fuzzy logic and autoregressive models. *Atmospheric Environment*. 2012;60:37-50.
7. Kumar A, Goyal P. Forecasting of air quality index in Delhi using principal component regression technique. *Atmospheric Pollution Research*. 2011;2:436-44.
8. Singh KP, Gupta S, Rai P. Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmospheric Environment*. 2013;80:426-37.
9. Russo A, Raischel F, Lind PG. Air quality prediction using optimal neural networks with stochastic variables. *Atmospheric Environment*. 2013;79:822-30.
10. Nouril, R.E., K. Ashrafi, and A. Azhdarpour, Comparison of ANN and PCA based multivariate linear regression applied to predict the daily average concentration of CO: A case study of Tehran. 2008. *Journal of the Earth and Space Physics*. 2008;34(2).
11. Tanaka, K., M. Sano, and H. Watanabe, Modeling and control of carbon monoxide concentration using a neuro-fuzzy technique. *Fuzzy Systems, IEEE Transactions on*, 1995. 3(3): p. 271-279.
12. Lee CC, Ballinger TJ, Domino NA. Utilizing map pattern classification and surface weather typing to relate climate to the Air Quality Index in Cleveland, Ohio. *Atmospheric Environment*. 2012;63:50-9.
13. Breiman L, Friedman JH, Olshen R, Stone C. Classification and regression trees. New York: Chapman & Hall; 1993.
14. Otterbein, L.E., et al., Carbon monoxide has anti-inflammatory effects involving the mitogen-activated protein kinase pathway. *Nature medicine*, 2000. 6(4): p. 422-428.
15. Naslmosavi S, Aghaei Chadegani A, Mehri M. Comparing the Ability of Bayesian Networks and Adaboost for Predicting Financial Distress of Firms Listed on Tehran Stock Exchange (TSE). *Australian Journal of Basic and Applied Sciences*. 2011;5(10):629-34..