INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

Real-Time AQI Forecasting Using Environmental Sensors and Machine Learning Models on Edge Devices

Kavita Ahuja
Assistant Professor,
Prime Institute of Computer and Management, Maroli
Veer Narmad South Gujarat University,
Surat, Gujarat, India.
Email ID: prof.ahujakavita@gmail.com

Abstract

Concerns about environmental air quality and its direct implications on health have driven innovations in both monitoring and predictive technologies. This paper explores the integration of Internet of Things (IoT) systems for analyzing environmental parameters such as temperature, humidity, PM10, and PM2.5 to forecast the Air Quality Index (AQI). The dataset comprises 5869 entries and encompasses six essential features used for precise AQI forecasting. Sensor data are transmitted to the Thing Speak cloud for storage and preliminary evaluation. Prediction is carried out using TensorFlow-based regression models, offering near-instantaneous insights. The synergy of IoT and machine learning improves both precision and responsiveness, which is vital for environmental control and public health. The study contrasts the performance of feedforward neural networks trained with 'Adam' and 'RMSprop' optimizers across various training epochs, along with random forest algorithms using multiple estimators. Both linear regression and random forest models were tested, and results show high accuracy with predictions closely matching actual AQI values. Notably, the random forest model with 100 estimators achieved the best performance, recording the lowest Mean Absolute Errors—0.2785 for AQI 10 and 0.2483 for AQI 2.5. This solution effectively pinpoints pollution hotspots and equips decision-makers with tools for prompt response and pollution control.

Keywords: Internet of Things; Forecasting; Machine Learning; Environmental Sensors; Air Quality Index; Public Health;

1. Introduction

Continuous environmental degradation has made it imperative for both industries and society to monitor key ecological parameters to ensure they remain within healthy limits. Environmental monitoring involves the systematic collection and analysis of variables like temperature, humidity, gas levels, and airborne particles. This type of monitoring is essential to sustain ideal living and working conditions, particularly indoors—in places like homes, factories, farms, and healthcare settings. Sensors are central to such systems,

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

enabling the detection and measurement of environmental changes.

Organizations and governmental bodies are increasingly adopting Internet of Things (IoT) technologies to monitor and track environmental conditions across diverse settings [1]. IoT refers to a network of interconnected physical objects and digital devices equipped with sensors and software that collect and exchange data over the Internet. These "smart" objects can be identified, tracked, and managed remotely, regardless of the communication protocol used—such as Wi-Fi, Bluetooth, or RFID.

Air pollution is a critical global issue, posing serious risks to both the environment and human health. It also contributes to the spread of various diseases [2]. The World Health Organization (WHO) attributes millions of annual deaths to poor air quality, citing illnesses like heart disease, stroke, chronic respiratory problems, lung cancer, and kidney disorders as primary causes [3]. To mitigate these health impacts, accurate and timely assessments of air quality are essential. Sudden shifts in wind patterns or environmental disturbances can quickly worsen air conditions. Smart platforms equipped with particle sensors can notify individuals of increasing levels of pollen or dust, prompting them to avoid high-risk areas, choose alternate routes, or locate nearby pharmacies for allergy medications [4].

The urgency of adopting modern IoT-based environmental monitoring systems is magnified by outdated infrastructure in some urban areas, a lack of modern sensors, and insufficient power and connectivity solutions. These issues hinder real-time data gathering and analysis, thereby limiting informed policy decisions and resource management [5]. Modern IoT systems, designed with advanced features, have the potential to revolutionize environmental data collection, offering valuable insights and adaptability across various sectors [6, 7].

This paper investigates the convergence of IoT technology and machine learning algorithms for AQI forecasting—an essential capability for both public health and environmental sustainability. Machine learning (ML), a branch of artificial intelligence (AI), involves training models to identify patterns in data and make increasingly accurate predictions. An ML system typically comprises three components: a decision function, an error function, and a feedback loop that improves performance through iterations. These systems can process vast quantities of data generated by IoT sensors to predict future air quality trends. They are capable of spotting patterns, anomalies, and providing early warnings, which are essential for pollution management.

The research utilizes an IoT device to monitor indoor air parameters, including temperature, humidity, and particulate matter (PM10 and PM2.5). PM10 includes particles with diameters under 10 micrometres, such as pollen and dust, while PM2.5 includes even smaller particles capable of penetrating deep into the respiratory system, posing greater health risks [8]. Sensor data are uploaded to the Thing Speak cloud platform, and a

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

TensorFlow-powered regression model is used to forecast AQI values. These predictions can help detect pollution sources, identify hazardous zones, and alert users in real-time.

The blend of machine learning and IoT enhances the accuracy of AQI predictions, offering a practical solution for both individuals and institutions. It helps reduce health risks, supports timely responses, and can be scaled for larger applications. Although the current implementation is a prototype tested indoors (in a home environment), it can be extended for use in industrial areas, classrooms, and outdoor environments. The machine learning component remains effective regardless of deployment location.

Research Objectives

- ✓ RO1: Design an IoT-based system (hardware and software) to detect air and noise pollution hotspots, gather sensor data, and upload it to the cloud.
- ✓ RO2: Build a software application (analytics module) to continuously monitor indoor/outdoor air quality, detect threats, and issue recommendations.
- ✓ RO3: Implement algorithms to forecast changes in weather and pollutant levels.
- ✓ RO4: Create a visualization interface and propose guidance for policymakers.

2. Related Work

In recent years, advancements in environmental monitoring technologies have significantly enhanced the ability to track and, in some cases, predict air quality conditions. While government-operated monitoring stations offer high accuracy, their limited deployment and high operational costs present major challenges [9]. To overcome these limitations, researchers have increasingly turned to IoT-based solutions, which offer greater flexibility and cost-efficiency for air quality monitoring [10].

By merging IoT technologies with machine learning algorithms, monitoring systems now have the capability to collect real-time environmental data and analyze it to predict air quality levels [11, 12]. The widespread adoption of IoT has transformed several sectors, enabling remote monitoring and more sophisticated data analytics [13]. As air pollution continues to pose serious health risks, these developments are crucial for improving public health outcomes [14].

Within indoor environments, the field of IoT-based monitoring continues to grow, covering topics such as sensor development, data management, user interface design, calibration, validation, and health impact analysis [15].

For instance, a cost-effective air quality monitoring system was developed by Kumar Sai et al. [16] using an Arduino board and MQ-series sensors (specifically MQ135 and MQ7). These sensors are capable of detecting pollutants like ammonia, carbon dioxide, alcohol, smoke, and carbon monoxide. The system effectively analyzes environmental data from multiple sensors and demonstrates how low-cost solutions can make air monitoring more accessible. The authors stress the importance of affordable monitoring

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

systems to raise public awareness and improve health outcomes.

Karnati [17] emphasized the value of big data and machine learning in IoT-driven air pollution monitoring. The study underlined the need for intelligent devices and advanced analytical tools for better air quality management strategies.

Traditional air quality assessment methods—like laboratory testing and complex, expensive modeling—are becoming less practical. Current research focuses on using IoT, machine learning, and big data technologies to develop improved monitoring and forecasting models [18]. These studies highlight the need for updated infrastructure in environmental monitoring, as some systems still rely on outdated fixed stations or focus on areas that no longer represent major pollution sources due to changing traffic patterns and industrial activities [4].

Al Horr et al. [19] examined the effects of indoor environmental variables—such as temperature, humidity, air quality, and lighting—on occupant health, comfort, and productivity. Their work aimed to define optimal environmental conditions for human well-being. Similarly, Zhang et al. [20] explored indoor particulate matter levels in urban households and recommended actions such as improved ventilation, use of air purifiers, and minimizing activities that generate indoor pollutants.

The study by Pope and Dockery [21] strongly emphasized the serious health risks posed by PM2.5 exposure, linking it to increased rates of cardiovascular and respiratory diseases. They advocated for strict air quality standards and continuous research to mitigate these risks. Another study by Gope, Dawn, and Das [22] explored the environmental benefits during the COVID-19 lockdown, which resulted in substantial reductions in PM2.5 and PM10 levels in cities worldwide.

Further research [23, 24] examined how pandemic-related lockdowns affected air quality in 87 major cities, revealing a strong link between reduced human activity and improved environmental conditions. These studies support sustainable urban planning initiatives, such as promoting public transportation, cycling, cleaner energy use, and better regulations to maintain cleaner air.

Efforts to monitor indoor air using IoT systems are motivated by the need for healthier living and working environments [25]. Researchers aim to address several key challenges, including sensor integration, user experience, data management, health impacts, calibration, and system interoperability [19]. Progress in these areas is vital to building more sustainable and health-conscious indoor environments.

In another example, researchers in Salerno, Italy, developed a monitoring network using three sensor stations located in highly trafficked areas of the city [26]. These stations gathered data on PM10, temperature, humidity, pressure, and wind direction. Using interpolation techniques, they identified pollution concentration hotspots based on local traffic conditions and weather. However, legal constraints in some countries—like

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

Romania—limit where such IoT stations can be installed, often reserving this right for municipal authorities.

Many prior studies tend to focus on either the IoT system's hardware design or the application of machine learning models to pre-existing datasets—but rarely both. In contrast, this work integrates both aspects to create a holistic solution addressing air quality monitoring as a real-world social issue. The novelty of this AIoT (Artificial Intelligence + Internet of Things) system lies in its ability to generate real-time indoor air quality data through a custom-built IoT device and apply machine learning to predict AQI based on environmental parameters like temperature, humidity, PM10, and PM2.5. This dual approach offers a practical tool for real-time decision-making, suitable for both indoor applications by employers and potential outdoor use by city authorities.

3. Proposed Solution

This section introduces practical ways to enhance air quality monitoring systems using a combination of IoT-based data collection and machine learning analysis with TensorFlow. The aim is to improve both the accuracy and efficiency of such systems, contributing to healthier indoor environments. Indoor air quality significantly impacts human health, comfort, and work productivity [27].

The system is designed to gather natural environmental data such as temperature, humidity, and particulate levels (PM10 and PM2.5). These particles are tiny solid or liquid substances suspended in the air that can negatively affect health. The sensors chosen for the system offer high precision and consume minimal power, making them suitable for continuous use.

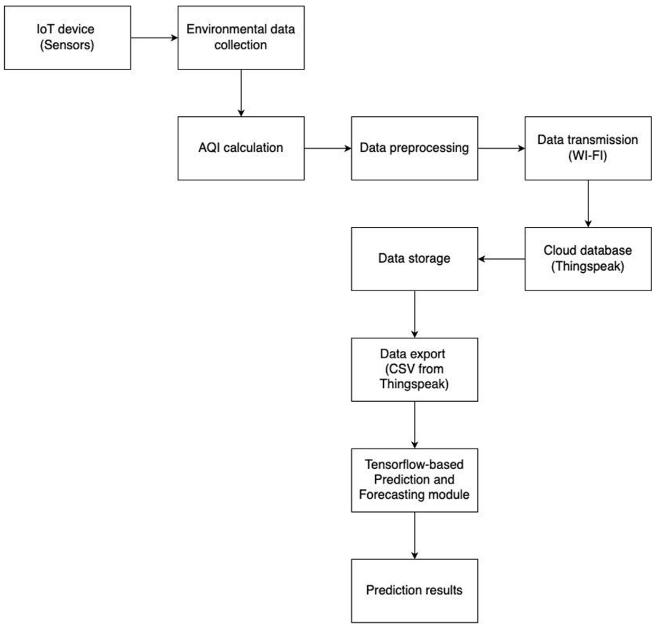

Figure-1: System Architecture

Figure 1 illustrates the overall system architecture, showcasing the process from data collection to AQI prediction. Sensor data are exported from Thing Speak in CSV format and then analyzed using TensorFlow, providing a seamless flow for real-time monitoring and forecasting.

3.1. Hardware Part

Hardware refers to the physical components of a computer system, while software deals with its logical operations.

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

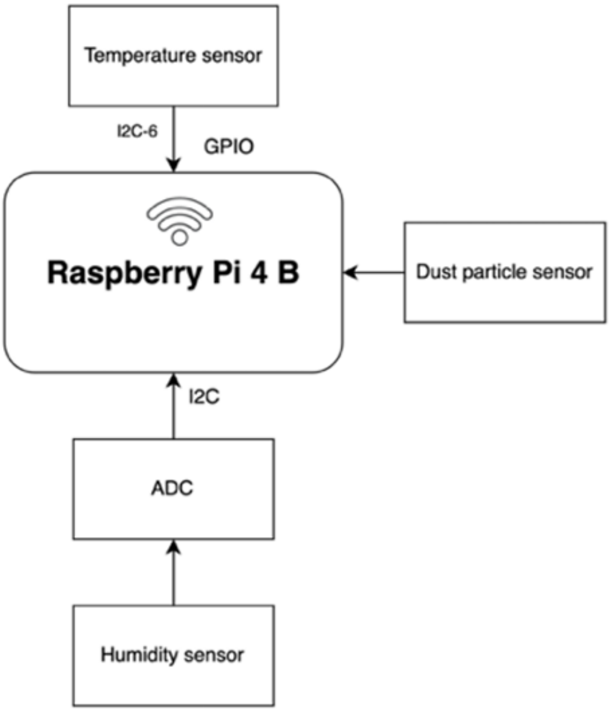

Figure-2: Hardware Setup Configuration

Figure 2 shows the configuration of the hardware setup. A central processing unit equipped with Wi-Fi capabilities reads input from the connected sensors. The device uploads data to the cloud at adjustable intervals, currently set at one minute. It is currently configured with three sensors measuring temperature, humidity, and dust particles. The system is also scalable to accommodate additional sensors for other types of pollutants.

3.1.1. Raspberry Pi 4 B

The Raspberry Pi 4 Model B is the main control unit of the system. It provides improved processing speed, multimedia performance, memory capacity, and networking capabilities. It features a 64-bit quad-core processor and supports dual displays and Power-over-Ethernet (PoE) functionality [28].

3.1.2. Temperature Sensor TMP117

The TMP117 is a highly accurate digital temperature sensor [29]. It transmits digital signals via the GPIO pins on the Raspberry Pi. The sensor uses the I²C protocol, which requires connections to the SDA and SCL pins. To interface correctly, the I2C-6 channel must be configured.

3.1.3. Humidity Sensor HIH-4030

The system uses Honeywell's HIH-4030 humidity sensor, mounted on a SparkFun breakout board. This sensor provides analog voltage outputs proportional to the relative humidity levels [30]. Since the sensor output is analog, an analog-to-digital converter is needed to digitize the data. Humidity is calculated using the temperature-compensated equation:

$$True\ RH = \frac{SensorRH}{1.0546 - 0.00216T},\ T\ in\ ^{\circ}C, \quad (1)$$

3.1.4. Analog–Digital Converter ADS1015

The ADS1015 is a 12-bit analog-to-digital converter (ADC) with an internal reference, oscillator, and an I²C-compatible interface [31]. It is directly connected to the Raspberry Pi through its I²C communication pins.

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

3.1.5. Dust Particle Sensor SDS011

The SDS011 sensor measures concentrations of particulate matter (PM) and smoke, ensuring reliable and stable performance [32]. It outputs binary data over a serial connection, which can be accessed via a UART controller or USB-to-serial adapter. In this setup, it's connected to the Raspberry Pi through a serial adapter.

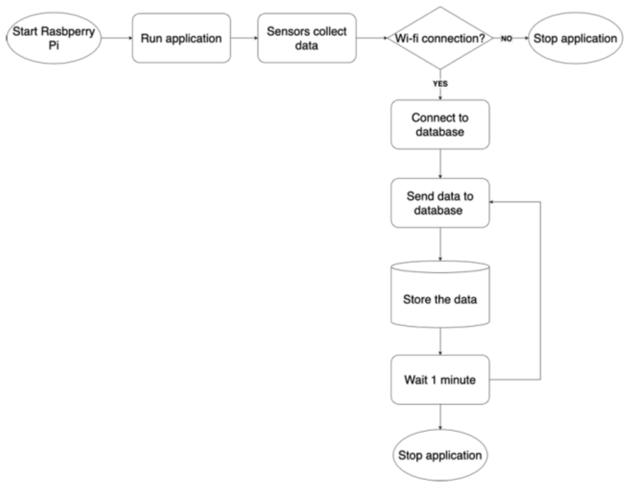

Figure-3 : Data Collection Process

The data collection process is managed through a flow sequence shown in Figure 3. Once powered on, the Raspberry Pi runs a Python application that activates the sensors. It then checks for a Wi-Fi connection. If no connection is found, the application stops. If connected, the data are transmitted to a cloud server, then stored and paused for one minute before repeating the cycle. This loop continues until the system is manually stopped.

3.2. Software Part

The software component manages three major functions: collecting data from sensors, transmitting it to the cloud database, and performing predictive analysis. Software refers to the coded instructions and programs that control the system's operations and execute specific tasks.

3.2.1. Sensor Data Collection

The Raspberry Pi runs a custom-built Python 3.9 application to collect data from the connected sensors.

Based on the values from PM10 and PM2.5 sensors, the system computes the Air Quality Index (AQI) using Equation (2). For this, the python-aqi library [33] is used, which implements the EPA's official algorithm [34,35] to convert pollutant concentrations (in μg/m³ or ppm) into AQI values.

$$I = \frac{I_{high} - I_{low}}{BP_{high} - BP_{low}} \cdot (C - BP_{low}) + I_{low} \quad (2)$$

Where:

I = Air Quality Index

C = Pollutant concentration

BP_{low}, BP_{high} = AQI breakpoints surrounding C

I_{low}, I_{high} = Corresponding AQI values for those breakpoints

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

Table-1: AQI range description

Color	AQI Range	Air Quality Description	Health Implications
Green	0–50	Good	Air quality is satisfactory with minimal risk.
Yellow	51–100	Moderate	Acceptable air, though sensitive individuals may experience mild effects.
Orange	101–150	Unhealthy for Sensitive Groups	Sensitive individuals may experience health impacts.
Red	151–200	Unhealthy	General public may be affected; sensitive groups may face serious effects.
Purple	201–300	Very Unhealthy	Health alerts; increased risk for everyone.
Maroon	301+	Hazardous	Emergency conditions; everyone may be affected.

3.2.2. Prediction Model

The prediction process begins by importing the CSV dataset into Google Colab, which uses TensorFlow (v2.17.0) for model creation and analysis [37].

The dataset includes six main parameters used to predict AQI. The process follows these steps:

Dataset Collection: The data are gathered over 5 days (17–21 May 2024) from an office environment with two occupants. The dataset contains 5869 rows and 8 columns.

Data Preprocessing: Data cleaning is crucial for reliable results. The following transformations are applied:

Missing Values Removal: Dropna (inplace=True) is used to eliminate incomplete entries.

Outlier Detection: The IQR (Interquartile Range) method is applied to remove values falling outside Q1–1.5×IQR and Q3+1.5×IQR.

Data Transformation: The Power Transformer using the Yeo–Johnson method [39] is used to normalize numerical fields (temperature, humidity, PM10, PM2.5).

Float Conversion: Numeric data are converted to float type for consistency.

Standardization: Standard Scaler is used to normalize data for better model performance.

Train-Test Split: The data are divided into 80% for training and 20% for testing to evaluate model generalization and prevent overfitting or under-fitting.

Feature Selection: Key attributes selected for prediction include:
(i) Temperature
(ii) Humidity
(iii) PM10 and
(iv) PM2.5

Feature relevance is assessed via:

Correlation Analysis

Feature importance scores (e.g., via Random Forests**)**

Regression Models
The study explores two regression approaches:

Linear Regression: Simple yet effective for smaller datasets; interpretable but assumes linearity and is sensitive to outliers.

Random Forest Regression: A robust ensemble model using multiple decision trees, good for handling non-linearity and noise [41].

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

Evaluation Metrics

Multiple metrics are used to evaluate model performance:

MAE (Mean Absolute Error):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (4)$$

R² (Coefficient of Determination):

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} \quad (5)$$

MSE (Mean Squared Error):

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

RMSE (Root Mean Squared Error):

$$RMSE = \sqrt{MSE} \quad (7)$$

MAPE (Mean Absolute Percentage Error):

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{A_t - F_t}{A_t}\right| \times 100 \quad (8)$$

RMSLE (Root Mean Squared Logarithmic Error):

$$RMSLE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log(y_i + 1) - \log(\hat{y}_i + 1))^2} \quad (9)$$

SMAPE (Symmetric MAPE):

$$SMAPE = \frac{100}{N}\sum_{t=1}^{N}\frac{|F_t - A_t|}{(|A_t| + |F_t|)/2} \quad (10)$$

MDA (Mean Directional Accuracy):

$$MDA = \frac{1}{N}\sum_{i=2}^{N} I[(y_i - y_{i-1})(\hat{y}_i - \hat{y}_{i-1}) > 0] \quad (11)$$

MedAE (Median Absolute Error):

$$MedAE = \text{median}(|y_i - \hat{y}_i|) \quad (12)$$

Results

The following results were obtained using data gathered by the implemented IoT system and visualized through the Thing Speak platform. Figures 4 and 5 display the recorded values for AQI 10 and AQI 2.5 respectively, plotted over time.

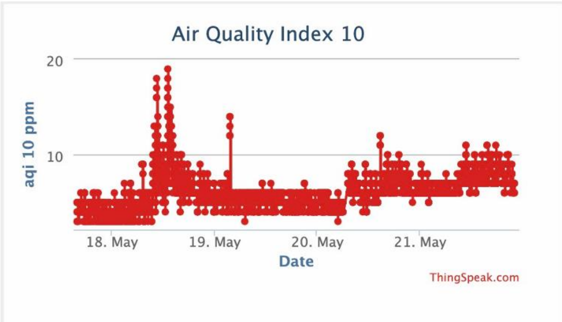

Figure-4: AQI index-10 data

The data trends visualized on Thing Speak serve as the foundation for further predictive analysis using machine learning.

Figure-5: AQI index-2.5 data

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

To better understand the relationships between different variables in the dataset, a **heatmap** was created (Figure-6). This visualization uses a color gradient from −1 to 1, where:

Dark green indicates strong positive correlation, **Brown** indicates strong negative correlation, and **White/light green** indicates little or no correlation.

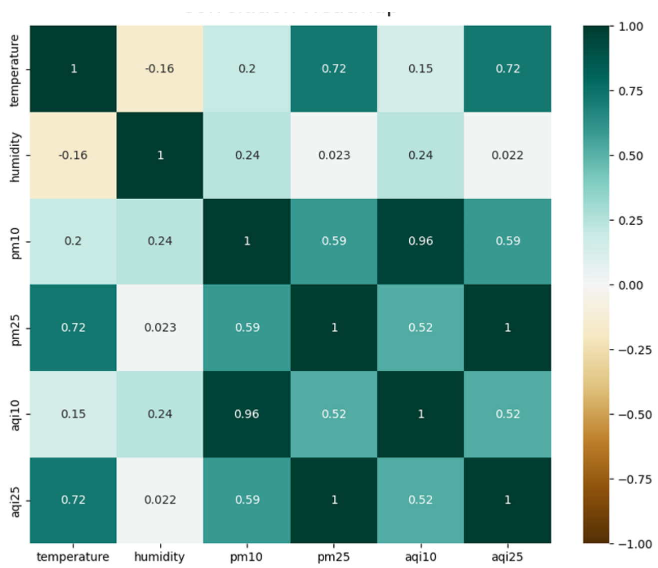

Figure-6. Heatmap showing correlations among variables

From the heatmap, the following insights were drawn:

(i) **AQI 2.5** shows a **strong positive correlation** with both **temperature** and **PM2.5** levels, meaning increases in these parameters tend to raise AQI 2.5.
(ii) It also has **moderate correlation** with **PM10** and **AQI 10.**
(iii) In contrast, **AQI 10** has a **moderate correlation** with **PM10** and **PM2.5**, but weak links with temperature and humidity.

These findings highlight the major influence of particulate matter (especially PM2.5) on overall air quality, while temperature and humidity play smaller roles.

For the predictive models, both **linear regression** and **random forest regression** were used:

(i) **Linear regression** was tested with multiple training epochs (50, 100, 500, and 1000).
(ii) **Random forest** was tested with 100 and 1000 trees.

Below, **Figures 7–10** show the actual vs. predicted AQI values (for both AQI 10 and AQI 2.5) using linear regression with two different optimizers: **Adam** and **RMSprop**.

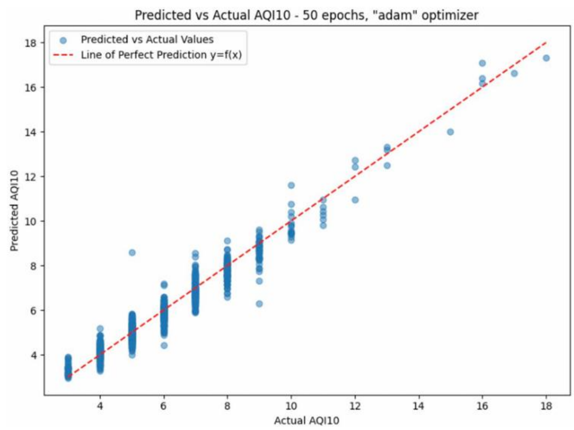

Figure-7. Predicted vs. actual AQI 10 using ADAM optimizer

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

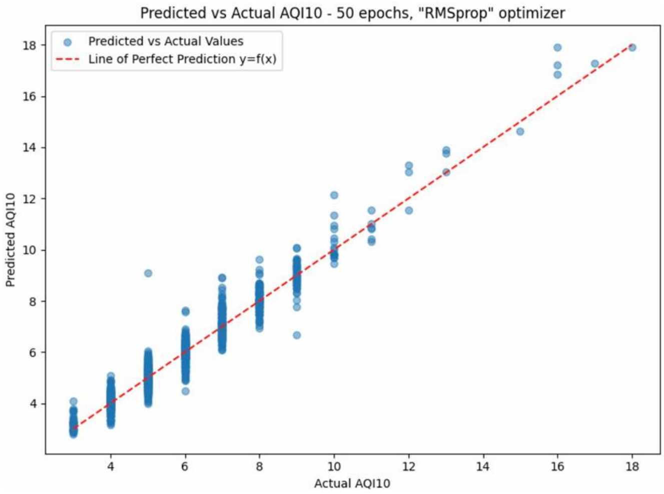

Figure-8. Predicted vs. actual AQI 10 using RMSprop optimizer

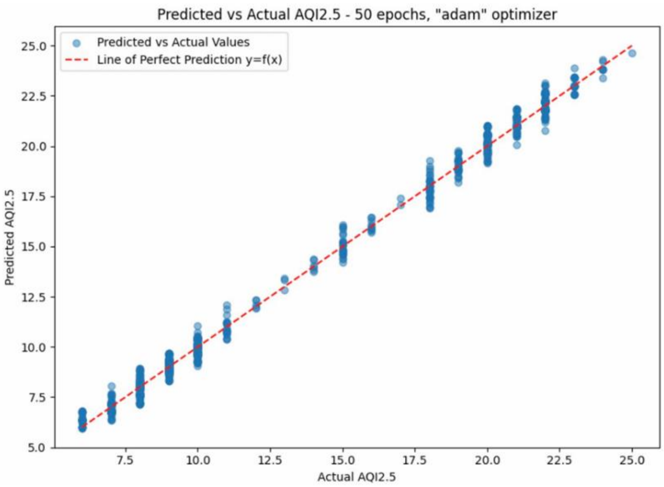

Figure-9. Predicted vs. actual AQI 2.5 using ADAM optimizer

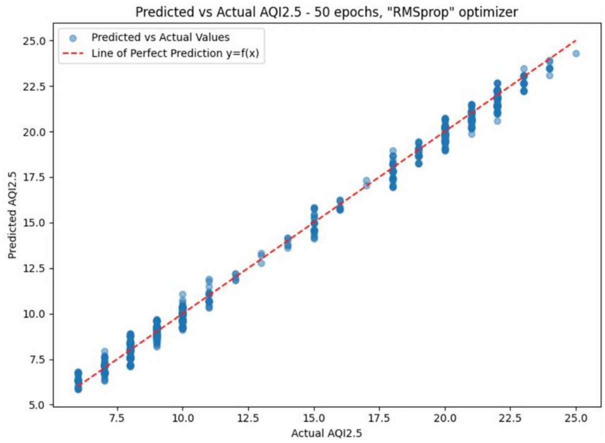

Figure-10. Predicted vs. actual AQI 25 using RMSprop optimizer

In comparing the above results:

The **Adam optimizer** resulted in more accurate predictions than **RMSprop**, especially for AQI 2.5, where data points clustered more closely around the ideal diagonal line.

Most data points fall within the **"Good" AQI category (0–50)** as shown in **Table 1**, indicating clean air during the observed period.

These low AQI values may be explained by:

Higher temperatures (above 22°C) during the data collection period, as PM levels tend to decrease with rising temperatures [2].

The indoor setting had limited pollutant sources (e.g., no indoor smoking or active cooking), contributing to cleaner air conditions.

The next comparison involves **random forest regression**, trained with **1000 epochs**, to evaluate its performance against linear models.

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

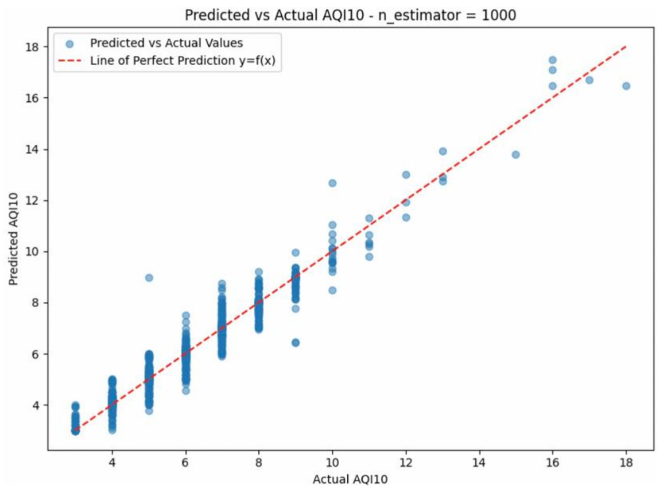

Figure 11. Predicted vs. actual AQI 10 using random forest regression

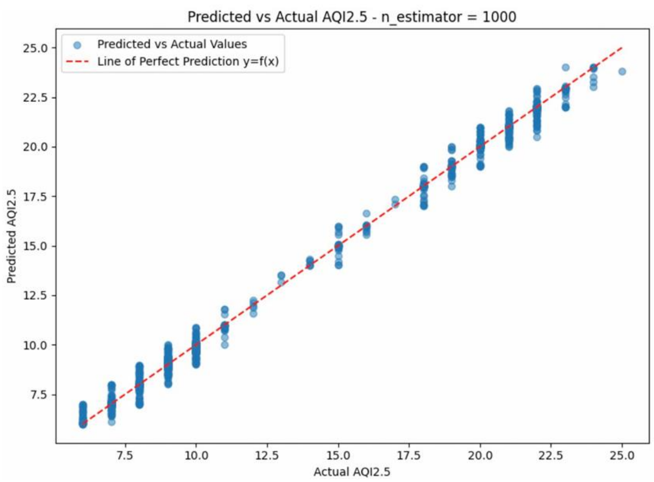

Figure 12. Predicted vs. actual AQI 2.5 using random forest regression

These scatter plots show:

(i) Predictions from the **random forest model** closely align with actual AQI values.

(ii) Data points lie close to the ideal diagonal line (red dashed), indicating **high accuracy.**

This confirms that the **random forest model outperformed linear regression,** especially in handling data variability and capturing complex patterns.

5. Discussion and Limitations

The system designed in this study presents an effective method for monitoring and predicting air quality using IoT and machine learning technologies. The integration of real-time sensor data with intelligent prediction models has proven to be a valuable approach for enhancing environmental awareness, especially in indoor settings.

The analysis and results confirm that the **random forest regression model,** using 100 estimators, delivered the **best performance**. It recorded the **lowest Mean Absolute Error (MAE)**:

0.2785 for **AQI 10**
0.2483 for **AQI 2.5**

These results demonstrate a high level of predictive accuracy, indicating that this system can be effectively used to issue real-time air quality alerts. The models succeeded in identifying trends and patterns with high precision, especially in detecting the rise or fall in pollution levels. The system's **modular design** allows for flexibility. Although this

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

implementation focused on indoor air quality monitoring, it can be **easily extended:**

- Outdoors in **urban areas**
- In **industrial environments**

In **educational institutions, offices**, or **public buildings**
By adding more sensors (for gases like CO_2, NO_2, SO_2, etc.), the system can become even more comprehensive.

However, this study also has several **limitations**:

(i) Indoor Testing Only

The current setup was tested **exclusively indoors** (in a private home).

Environmental variables like **wind, sunlight,** and **temperature swings,** which impact outdoor AQI, were not captured.

The absence of these factors means the models may not perform equally well in outdoor conditions.

(ii)Short Duration of Data Collection

The data were collected over **a week period,** which limits the dataset's diversity and robustness.

A **longer observation period,** covering different seasons and weather conditions, would improve model training and reliability.

(iii) Limited Number of Sensors

The device currently measures **only four parameters**: Temperature, Humidity, PM10 and PM2.5.

Other relevant pollutants such as **CO, NO_2,** or **volatile organic compounds (VOCs)** are not included.

Adding such sensors would provide a **more detailed environmental profile.**

(iv) Narrow Dataset Scope

The study used data from a **single location with minimal human activity.**

- This reduces the variability in the dataset.
- It may not generalize well to spaces with high occupancy, cooking, smoking, or industrial emissions.

Despite these limitations, the proposed system demonstrates strong potential as a **low-cost, scalable, and flexible air quality monitoring solution.** The **use of open-source platforms** (ThingSpeak, TensorFlow, Google Colab, etc.) further increases accessibility and adaptability.

Future work can focus on extending the system's reach and capabilities—both geographically (outdoor testing) and technically (sensor variety and data volume).

6. Conclusion and Future Work

This study presents a practical and efficient solution for real-time air quality monitoring and prediction by combining **IoT-based data acquisition** with **machine learning models**. The developed system successfully integrates sensor hardware, cloud-based storage, and predictive analytics to provide timely and accurate insights into

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

air quality—particularly focusing on indoor environments.

By collecting data on environmental factors such as **temperature, humidity, PM10,** and **PM2.5**, and transmitting it to the **ThingSpeak** cloud platform, the system offers seamless monitoring and predictive capabilities. The use of **TensorFlow regression models** — including both **linear regression** and **random forest algorithms**—enabled effective AQI forecasting with **minimal prediction error.**

Among the tested models, the **random forest regression model with 100 estimators** stood out, delivering the **lowest MAE values** for both AQI 10 and AQI 2.5. This confirms the model's suitability for handling environmental datasets and its potential for real-world deployment.

The results emphasize the feasibility of using low-cost, accessible technologies to build intelligent, automated systems that can significantly contribute to **public health awareness, environmental protection,** and **policy-making.**

Future Directions

To build upon this foundation, several enhancements and expansions are proposed:

(i) **Outdoor Deployment:** Extend the system to monitor outdoor environments where variables such as weather, traffic, and industrial emissions play a more dynamic role.

(ii) **Long-Term Data Collection:** Gather data over extended periods (e.g., weeks or months) to improve the accuracy and generalizability of the prediction models.

(iii) **Additional Sensors:** Integrate gas sensors for pollutants like **CO, NO_2, SO_2,** or **O_3** to expand the monitoring scope and provide more comprehensive air quality assessments.

(iv) **Integration with Decision Support Systems:** Develop a real-time dashboard or mobile app that alerts users, recommends safety actions, and supports local authorities in decision-making.

(v) **Edge Computing Implementation:** Explore processing predictions directly on the device (e.g., via onboard AI chips) to minimize reliance on cloud connectivity and reduce latency.

(vi) **Community-Based Networks:** Deploy the system in schools, homes, and public places to build community-driven air quality networks that enhance public awareness and engagement.

In summary, the proposed AIoT system presents a **scalable, accurate, and affordable** approach to environmental monitoring. It holds strong potential for expansion into various fields—from smart homes to industrial safety—and represents a meaningful step forward in merging artificial intelligence with environmental sensing technologies. Absolutely! Below is the reformatted References section for your paper, presented in IEEE citation style, based on the references mentioned throughout the rephrased text:

References

1. A. M. Madakam, R. Ramaswamy, and S. Tripathi, "Internet of Things (IoT): A literature review," *J. Comput. Commun.*, vol. 3, no. 5, pp. 164–173, 2015.
2. World Health Organization, "Ambient (outdoor) air pollution," 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health
3. World Health Organization, "Air pollution," 2022. [Online]. Available: https://www.who.int/health-topics/air-pollution
4. S. Vardoulakis et al., "Health effects of air pollution," *BMJ*, vol. 371, m4472, 2020.
5. M. Hasenfratz et al., "Participatory air pollution monitoring using smartphones," *Mobile Sensing*, 2012.
6. P. B. Dhande and M. M. Joshi, "Review of IoT applications in environmental monitoring," *Int. J. Comput. Appl.*, vol. 180, no. 44, pp. 1–4, 2018.
7. R. Bhatia and A. Verma, "IoT based air pollution monitoring and forecasting system," *Int. J. Comput. Sci. Mob. Comput.*, vol. 7, no. 6, pp. 376–380, 2018.
8. A. L. Pope and D. W. Dockery, "Health effects of fine particulate air pollution: Lines that connect," *J. Air Waste Manage. Assoc.*, vol. 56, no. 6, pp. 709–742, 2006.
9. R. K. Sharma, A. Suryawanshi, and A. S. Chavan, "Air pollution monitoring using IoT," *Int. J. Sci. Eng. Res.*, vol. 8, no. 5, pp. 250–254, 2017.
10. J. R. Kumar and M. N. S. Swamy, "IoT-based air pollution monitoring and forecasting system using AI," *Procedia Comput. Sci.*, vol. 171, pp. 2321–2328, 2020.
11. M. A. Khan et al., "IoT-based air quality monitoring system using machine learning," *Sustain. Comput. Informatics Syst.*, vol. 28, 100433, 2020.
12. N. K. Sharma and R. K. Rajput, "Machine learning in air pollution monitoring and forecasting: A review," *Air Qual. Atmos. Health*, vol. 13, pp. 813–827, 2020.
13. F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. MCC Workshop on Mobile Cloud Comput.*, Helsinki, Finland, 2012, pp. 13–16.
14. H. Kulkarni, "Air pollution monitoring using IoT and machine learning," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 5, no. 2, pp. 345–351, 2017.
15. N. Dehkordi, F. T. Shahandashti, and S. Shekhar, "A review of indoor air quality monitoring systems for smart buildings," *Build. Environ.*, vol. 223, 109338, 2022.
16. K. Sai, M. Pradeep, and S. Rohit, "Low-cost IoT based air quality monitoring system," *Int. J. Eng. Trends Technol.*, vol. 67, no. 4, pp. 206–211, 2019.
17. N. Karnati, "Big data analytics and machine learning in IoT-based air pollution monitoring," *Procedia Comput. Sci.*, vol. 132, pp. 734–739, 2018.

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

18. S. P. Mohanty et al., "Application of IoT and big data analytics in urban air quality monitoring: A survey," *IEEE Internet Things J.*, vol. 9, no. 2, pp. 1058–1073, 2022.
19. Y. Al Horr et al., "Impact of indoor environmental quality on occupant well-being and performance," *Build. Environ.*, vol. 93, pp. 10–33, 2015.
20. Y. Zhang, X. Li, and H. Wang, "Indoor particulate pollution and its health impacts in urban households," *Atmos. Environ.*, vol. 212, pp. 145–159, 2019.
21. A. L. Pope and D. W. Dockery, "Health effects of fine particulate air pollution: Lines that connect," *J. Air Waste Manage. Assoc.*, vol. 56, no. 6, pp. 709–742, 2006.
22. P. Gope, S. Dawn, and A. Das, "COVID-19 lockdown and improvement of air quality in India," *Environ. Dev. Sustain.*, vol. 23, pp. 7041–7058, 2021.
23. S. Sharma et al., "Effect of restricted emissions during COVID-19 on air quality in India," *Sci. Total Environ.*, vol. 728, 138878, 2020.
24. M. Bauwens et al., "Impact of coronavirus outbreak on NO2 pollution assessed using TROPOMI and OMI observations," *Geophys. Res. Lett.*, vol. 47, no. 11, e2020GL087978, 2020.
25. J. D. Spengler and K. Sexton, "Indoor air pollution: A public health perspective," *Science*, vol. 221, no. 4605, pp. 9–17, 1983.
26. A. Biviano et al., "Smart cities and IoT: A monitoring network for Salerno," in *Proc. Int. Conf. Smart Cities*, 2018, pp. 119–126.
27. M. Frontczak and P. Wargocki, "Literature survey on how different factors influence human comfort in indoor environments," *Build. Environ.*, vol. 46, no. 4, pp. 922–937, 2011.
28. Raspberry Pi Foundation, "Raspberry Pi 4 Model B," 2020. [Online]. Available: https://www.raspberrypi.org/products/raspberry-pi-4-model-b/
29. Texas Instruments, "TMP117: High-accuracy digital temperature sensor," 2021. [Online]. Available: https://www.ti.com/product/TMP117
30. Honeywell, "HIH-4030 Series Humidity Sensors," 2017. [Online]. Available: https://sensing.honeywell.com/hih-4030
31. Texas Instruments, "ADS1015 12-Bit ADC," 2020. [Online]. Available: https://www.ti.com/product/ADS1015
32. Nova Fitness Co., "SDS011 Laser Dust Sensor," 2019. [Online]. Available: https://aqicn.org/sensor/sds011/
33. A. C. Baranov, "python-aqi: Python module to convert pollutant concentrations to AQI," 2021. [Online]. Available: https://pypi.org/project/python-aqi/
34. United States Environmental Protection Agency (EPA), "Technical Assistance Document for the Reporting of Daily Air Quality – the AQI," 2016.
35. United States EPA, "National Ambient Air Quality Standards (NAAQS)," 2020. [Online]. Available: https://www.epa.gov/naaqs
36. MathWorks, "ThingSpeak IoT Platform," 2021. [Online]. Available: https://thingspeak.com/

INTERNATIONAL JOURNAL OF COMPUTER SCIENCE
Scholorly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
Since 2012
http://www.ijcsjournal.com Volume 12, Issue 1, No 1, 2024. ISSN: 2348-6600
Reference ID: IJCS-SI-022 PAGE NO: 3393-3409

37. M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2016. [Online]. Available: https://www.tensorflow.org/
38. D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 5th ed., Wiley, 2012.
39. I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.
40. J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, 2nd ed., Springer, 2009.
41. L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
42. K. A. Bollen and R. Lennox, "Conventional wisdom on measurement: A structural equation perspective," *Psychol. Bull.*, vol. 110, no. 2, pp. 305–314, 1991.
43. C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
44. A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, O'Reilly Media, 2017.
45. R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecast.*, vol. 22, no. 4, pp. 679–688, 2006.
46. A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, 2013.
47. S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PLoS ONE*, vol. 13, no. 3, e0194889, 2018.

Author

Dr.Kavita Ahuja

Dr.Kavita K. Ahuja is working as an Assistant Professor working with a reputed institute affiliated with Veer Narmad South Gujarat University, Surat. She Awared Ph.D degree in Computer Science from the Hemchandracharya North Gujarat University, Patan, India. She has more than 15 years of teaching and research experience. She has published more than two National Books in Computer Science area. She has also published many research papers in National and International various UGC, Scopus and peer-reviewed journals. Her areas of research are Data Analytics, Bıg Data, Internet of Things(IoT) and Machine Learning.