HARNESSING MACHINE LEARNING FOR ACCURATE AIR QUALITY PREDICTION: A TIME SERIES ANALYSIS

A MAJOR PROJECT-II REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
INFORMATION SYSTEMS

Submitted by:
THAMMALI GANGADHAR
2K22/ISY/22

Under the supervision of

Mrs. ANAMIKA CHAUHAN

DEPARTMENT OF INFORMATION AND TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

May, 2024

MASTER OF TECHNOLOGY (INFORMATION SYSTEMS) THAMMALI GANGADHAR 2024

DELHI TECHNOLOGICAL UNIVERSITY
(FORMERLY DELHI COLLEGE OF ENGINEERING)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Thammali Gangadhar, 2K22/ISY/22 studying MTech in Information Systems, hereby declare that the Major Project-II dissertation titled "**HARNESSING MACHINE LEARNING FOR ACCURATE AIR QUALITY PREDICTION: A TIME SERIES ANALYSIS**" which is submitted by me to the Department of Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any degree, Diploma Associateship, Fellowship, or other similar title or recognition.

Place: Delhi **THAMMALI GANGADHAR**
Date: 30/05/2024 **2K22/ISY/22**

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(FORMERLY DELHI COLLEGE OF ENGINEERING)
Bawana Road, Delhi-10042

CERTIFICATE

I hereby certify that the Major Project-II dissertation titled "**HARNESSING MACHINE LEARNING FOR ACCURATE AIR QUALITY PREDICTION: A TIME SERIES ANALYSIS**" which is submitted by Thammali Gangadhar, 2K22/ISY/22, Department of Information Technology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge, this work has not been submitted inpart or full for any degree or diploma to this University or elsewhere.

Place: Delhi
Date: 30/05/2024

Mrs. Anamika Chauhan
SUPERVISOR
Department of Information Technology
DELHI TECHNOLOGICAL UNIVERSITY

ACKNOWLEDGEMENT

I am grateful to **Prof. Dinesh K. Vishwakarma**, Head of Department (Department of Information Technology), Delhi Technological University (Formerly Delhi College of Engineering), New Delhi and all other faculty members of our department for their astute guidance, and constant encouragement and sincere support for this project work.

I would like to take this opportunity to express my profound gratitude and deep regard to my project supervisor **Mrs. Anamika Chauhan**, for her exemplary guidance, valuable feedback, and constant encouragement throughout the duration of the project. Her valuable suggestions were of immense help throughout our project work. Her perspective criticism kept us working to make this project in a much better way. Working under her was an extremely knowledgeable experience for us

I would also like to appreciate the support provided to me by my seniors and my peer group who aided me with all the knowledge they had regarding various topics. I hope that this project will serve its purpose to the fullest extent possible.

THAMMALI GANGADHAR
2K22/ISY/22

ABSTRACT

The existence of dangerous or excessive amounts of pollutants in the atmosphere that endanger human health, the environment, and general quality of life is referred to as air pollution. Short-term exposure to air pollution results in acute health issues such headaches, exhaustion, and irritation of the throat, nose, and eyes. Prolonged exposure raises the risk of heart disease, lung cancer, chronic respiratory disorders, brain, nerve, liver, and kidney damage. Children who are exposed for an extended period of time may also have problems. As a result, assessing and forecasting air quality is a crucial step in reducing environmental risk. To measure pollution levels, we use AQI. In recent years, there has been a growing interest in the use of machine learning (ML) and deep learning (DL) techniques for air pollution forecasting. These techniques have the potential to provide more accurate and timely predictions of air pollution levels, which can be used to inform public health interventions and environmental policy decisions. This thesis reviews the existing literature on the use of ML and DL techniques for air pollution forecasting. The thesis provides an overview of the different types of ML algorithms that have been used for this purpose. For the previous two years, Delhi, the capital of India, has been the most polluted city in the world. So, this research paper collected data on air pollution from CPCB (Central Pollution Control Board) specifically focusing on fine particles such as PM2.5, PM10, NO2, NH3, SO2, CO and OZONE from five different areas or stations in Delhi. Different machine learning algorithms such as Logistic Regression, K Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, Decision Tree and XGBoost are used to analyses the collected data. Evaluation metrics such as accuracy, Precision and are used. The comparison between the models is also discussed in this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Air pollution is a wide spread problem and one of the major environmental issues that not only effecting humans but also other living being of the planet. WHO research states that 7 million deaths globally occur each year as a result of diseases linked to air pollution, including cancer, bronchitis, asthma, disorders of the throat and eyes, and heart problems. The major pollutants present in the air and causing these diseases includes Particulate Matter (PM10 and PM2.5), CO, NO2, SO2, Suspended Particulate Matter (SPM), Ground level Ozone (O3).

Figure 1.1: pollutants

Due to the rapid development of the economy and the large number of people, air pollution has become a big problem in the city. According to the City Population website [1], 55,714% of people moved to the city in 2019. Population growth will cause many problems such as transportation, air pollution and health. Good weather forecasting and identifying when the weather is hazardous is important. It is necessary to predict weather conditions and identify situations when weather conditions exceed hazardous conditions. To do this, the government uses the air quality indicator AQI. Air Quality Index (AQI) is used to report air quality on a daily basis. AQI is divided into six categories, each with different health issues and a specific color. According to a survey of 7,000 cities worldwide by the Ministry of Health and the World Health Organization in August 2022 [3], India's urban areas, Delhi, have the worst air pollution among all major cities in the world. During March and September, Delhi's AQI generally falls into three categories: good (0-50), 1 satisfactory (51-100) and fair (101-1200). However, from October to February, due to the influence of many factors such as road dust and traffic, the weight dropped to "Weak" (201-300), "Severe" (301-400) or "Dangerous" (401 - 500). +)

levels. However, as machine learning progressed, everyday activities began to use very accurate machine learning models that were trained using probability functions and produced precise predictions. Machine learning has proven to be remark ably resilient in its ability to anticipate natural phenomena.

1.1 INDOOR AND OUTDOOR AIR POLLUTION

Air pollution is categorized into two types Indoor and Outdoor pollution. The pollutants mentioned above comes under outdoor pollution. The Indoor pollution refers to the contaminants that come from furniture, paints and adhesives inside the building. Tobacco smoke, household goods and insecticides, wood-burning stoves, and gases including CO, lead (pb), and NO2 are among the causes of indoor pollution. A study reveals that indoor pollution levels are 2-5 times high than outdoor pollution levels

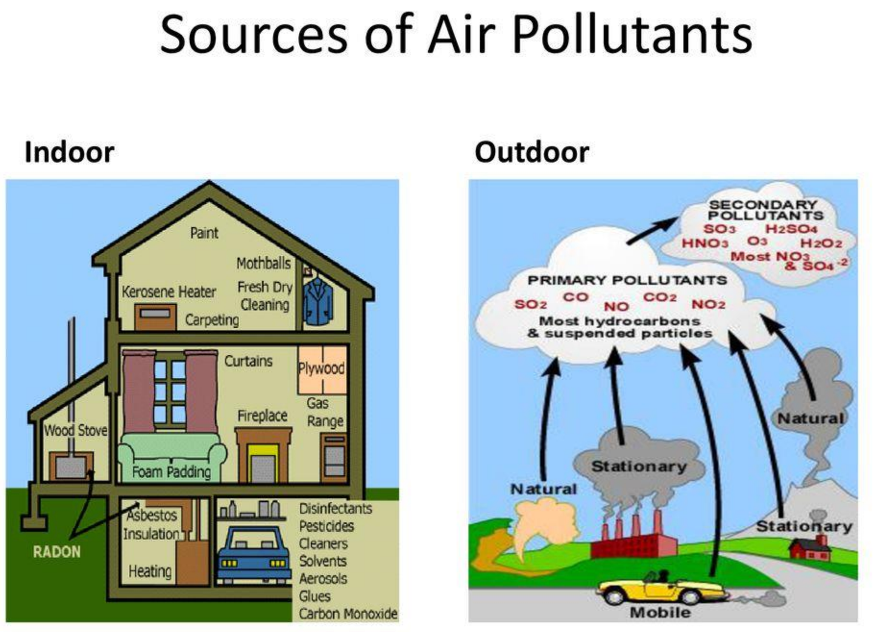


Figure 1.2: Sources of air pollutants

So, in order to control the emission of gases into the atmosphere, government is using air quality monitoring systems. The pollution prediction, forecasting and updation has gained tremendous importance in the recent years.

1.2 Air Quality Index (AQI)

The daily air quality is reported using the AQI. It alerts you to the level of pollution or purity in the air as well as any possible health risks. It focuses on potential health effects that breathing in polluted air may cause hours or days later. The AQI is calculated using twelve factors, according to the CPCB (Central Pollution Control Board). These parameters are NO2 (nitrogen dioxide), NH3 (ammonia), SO2 (sulphur dioxide), As (arsenic), CO (carbon monoxide), O3 (ozone), PM10, PM2.5, Ni (nickel), and Pb (lead). The AQI is commonly computed using the criteria pollutants (i.e., O3, CO, SO2, NO2, PM2.5, and PM10), even though it is recommended to calculate the AQI using several pollutants from the list of 12 pollutants. However, the AQI objectives, the averaging time, the data availability, the frequency of monitoring, and the measuring methods all influence the choice of pollutants. The health issues increase as the AQI rises. The purpose of the AQI is to raise public awareness. different contaminants have a unique number, name, and colour allocated to them. The AQI is categorised as Excellent, Acceptable, Mild, and Unsatisfactory. Each of these categories is determined by the levels of air pollution concentration in the atmosphere and the potential health effects they may have. In Figure 1.3, the AQI index is displayed.

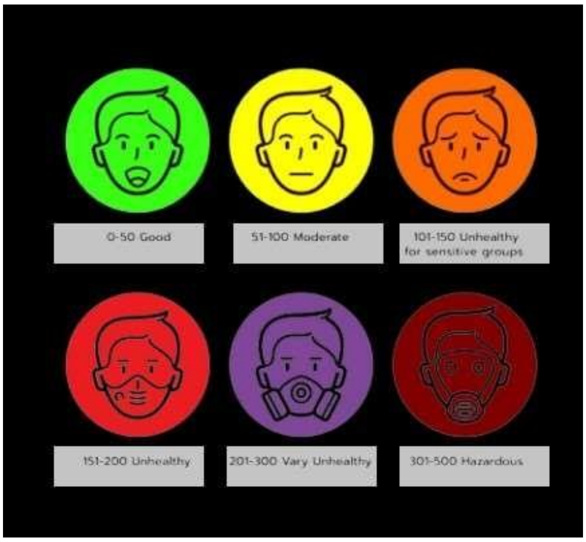


Figure 1.3: AQI Index

The AQI is divided into six categories, each of which has a distinct color and corresponds to a different health problem.

AQI is for five major pollutants. They are

1. Ground level ozone
2. Particle pollutants (Particulate Matter including $PM_{2.5}$ and PM10)

3. Carbon monoxide
4. Sulphur dioxide
5. Nitrogen dioxide

Table 1.1: AQI values

Daily AQI color	Levels of concern	Value of Index
Green	Good	0 – 50
Yellow	Moderate	51 – 100
Orange	Unhealthy for sensitive groups	101 -150
Red	Unhealthy	151 – 200
Purple	Very unhealthy	201 – 300
Maroon	Hazardous	301 and higher

1.3 Research problem

To forecast the air pollution in Delhi, this study uses machine learning models to predict and forecast the air pollution on a time series dataset which consists of AQI values of various pollutants such as PM2.5, PM10, NO2, SO2, NH3, CO and OZONE. The data of these pollutants is collected from CPCB (Central Pollution Control Board).

To achieve our problem statement, we have studied various research papers that uses machine learning models on a time series dataset and compared the accuracy of various models by running the models on our own dataset.

1.4 Purpose of Dissertation

The purpose of this dissertation is to forecast the air pollution in Delhi by collecting the data of various pollutants of five different stations or areas in Delhi from CPCB. This study provides a comprehensive overview and analysis of implementation of various machine learning models on the dataset. This research aims to address the need for prediction and forecasting of air pollution through extensive literature survey of existing work on forecasting air pollution. In addition to implementation of various machine learning models, this thesis also aims to provide a comprehensive review of various existing machine learning models used by researchers on various datasets and also provides the research gaps in those papers.

The outcome of this research can help the researchers in future for analyzing a time series data on machine learning models and help them in predicting better air quality.

CHAPTER 2
RELATED WORK

The data of various pollutants that we are getting from sensors, air monitoring systems requires rigorous analysis for getting accurate and reliable predictions, which only ML and DL algorithms can handle efficiently. This paper [5] is a survey of ongoing examinations which utilizes three strategies, Bibliometrics technique to survey the exploration status of anticipating air contamination techniques, Developmental trees for spatial transient information investigation and Markov chain technique to anticipate future exploration patterns for significant air pollutants. The two algorithms utilized in this research article [6] are Random Forest and Naive Bayes. 32516 record datasets are used to assess these models. A definite depiction of each algorithm utilizing pseudocodes was remembered for the distribution what's more, IBM SPSS form 21 was used to execute this paper what's more, Random Forest beats with an exactness of 99.154% contrasted with Naive Bayes. The biggest city in Australia, Sydney [7], has around 50 observing stations heavily influenced by the New South Grains Office of Climate and Legacy (OEH) that are utilized to constantly screen the city's air quality. Six air pollutants, including CO, NO2, O3, SO2, and two particle fractions (PM10 and PM2.5), are estimated at these checking areas by the OEH, which is responsible for keeping up with ecological wellbeing. The OEH has conveyed AQM 65, which estimates PM10, PM2.5, O2, NO2, CO, and SO2, because of the continuous ascent in air contamination levels. It just necessities a USB power source and a Wi-Fi access highlight be set up, making it exceptionally simple to utilize. Once associated, their guides quickly and progressively show the measures of air contamination [8]. The research paper [9] focusses on the following predictive models such as Random Forest, Gradient boosting regressor, KNN and Decision Tree regressor-CART for forecasting the air pollutant concentration of PM2.5 and are tested on the dataset "Taiwan Air Quality Monitoring Network dataset (TAQMN)".

The authors of [10] inspected examinations that pre-owned AI furthermore, enormous information examination to evaluate air quality and noted potential troubles and financing impediments for additional examination. The paper likewise pointed out the issue of equipment limitations affecting the accuracy of air quality screening and assessment. The authors recommended that more review be finished on robotized continuous approval and information quality displaying to defeat this issue.

The article gives an outline of the writing on a couple algorithms that have been applied to various information mining approaches for climate expectation by various scholastics. The work that numerous researchers in this field did has been examined furthermore, differentiated in a plain way. Decision trees and k mean clustering show unrivaled weather conditions conjecture execution also, exactness when contrasted with different information mining strategies [11].

An air quality observing framework with evaluation and estimating modules in light of information features such contamination levels, climate, and compound parts gotten from the WRF-Chem model was presented by the creators of [12]. The authors performed trials on 74 Chinese cities using several feature groups and classification techniques such random forest, SVM, and linear regression. Their outcomes showed that the consolidated model delivered expectations that were more precise than those of the individual models. The study [13] discusses air pollution and emphasises the serious health and financial risks it poses. The study centres on precise evaluation of air quality, specifically in Delhi, India, by utilising three machine learning models: SARIMAX, Prophet, and LSTM. SARIMAX is determined to be the most dependable of these. The Central Pollution Control Board provided the data, which was then pre-processed and managed by replacing missing values with mean values. The RSME values for the dataset mentioned above are as follows. 39.11- SARIMAX, 37.84 Prophecy and 120.70- LSTM. The highest performing model for univariate time series forecasting, according to the paper's conclusion, was SARIMAX. The performance of LSTM could be enhanced by additional training. Parameter adjustment for SARIMAX and extending the study to further places should be priorities for future development.

This research [14] examines how measurement errors in time series data might cause ambiguity and uncertainty in the data. Owing to the uncertainty in the data, different forecasting models may yield different results. In order to deal with data uncertainty, this work proposes fuzzy symmetry, which in turn uses triangular fuzzy numbers (TFN). After the data has been fuzzified, the midpoint of the fuzzy set is calculated to complete the defuzzification process. A basic average method is employed to determine the midway, however it is ineffective in cases where the data is unclear. Thus, this work uses a linear programming technique known as "defuzzification" to determine the middle point of fuzzy data. The suggested linear programming method outperforms in terms of accuracy, the article says. In order to test the dataset using machine learning models such as linear regression, lasso regression, random forest regression, and K-nearest neighbour regression, this article [15] collected data on PM2.5, SO2, NO2, and CO. The accuracy, mean absolute error (MAE), mean-square error (MSE), and root-mean-square error (RMSE) metrics are used to evaluate the findings. The research additionally examines the comparability of different models. Random forest regression is the most accurate and has the lowest RMSE of all the models.

This research paper [16] utilizes two algorithms Naive Bayes and Random Forest. These models are tried on 32516 record datasets. The paper gave a brief about every one of the algorithms with pseudocodes. This paper involved an IBM SPSS rendition 21 for execution. The research concludes by stating that, based on 2-tail analysis, Random Forest (99.154%) outperforms Naive Bayes (97.32%) in correlation with accuracy of 0.056 ($P>0.05$), recall 0.01 ($p<0.05$) and precision 0.02 ($p<0.05$). The research paper [17] proposes a Recurrent Neural Network model to help decision-makers and local authorities manage air pollution in Skopje, North Macedonia. The model uses various models, including IoT-related sensory arrays, cloud computing, statistical learning, Deep learning, ARIMA, and different DL architectures. The daily Air Pollution Index (API) was predicted using a Fuzzy Time Series Markov chain (FTSMC) model. For modelling jobs involving numerous time series, VARIMA models offer an alternative to auto-regressive approaches. The paper's contribution is threefold: it combines various data sensor sources in Skopje's area to increase prediction accuracy, leverages meteorological data from meteorological stations and historical data, and compares different architectures with the ARIMA model. This approach can help manage air pollution consequences and take proactive measures.

The research [18] introduces a deep learning-driven algorithm called Integrated Multiple Directed Attention (IMDA) and Variational Autoencoder (VAE) to improve air pollution forecast accuracy. The method is used to estimate concentrations of CO, NO, NO2, SO2, and O3. The accuracy of this methodology is assessed using six statistical metrics. The IMDA-VAE method outperforms various DL models like LSTM, bidirectional LSTM, GRU, and bidirectional GRU. The paper contributes to the development of IMDA-VAE based on attention mechanisms and conventional VAE, validates the suggested techniques using three distinct forecasting experiments, and evaluates the performance and capabilities of the proposed forecasting models and strong recurrent neural network models. This paper [19] explores the use of deep learning techniques in predicting air pollution in smart cities. It compares machine learning (ML) and deep learning (DL) methods, discusses various Deep

Learning model types, and provides a brief description of LSTM, SDL, CNN, and SAE. The paper highlights the rapid development of these technologies and their widespread application in daily life.

This study [20] investigates the effects of pollution levels in India before and after the lockdown, using various metrics to reveal geographical air quality patterns. The research focuses on air pollution causes, sources, long-term and short-term effects, and the major 8 air pollutants measured by AQI (PM2.5, PM10, NO2, CO, SO2, NH3, SO3, and lead). The "Air Quality data in India (2015-2020)" dataset, which includes 15 parameters like PM2.5, PM10, NO2, CO, SO2, NH3, SO3, etc., is used. The study found a significant decrease in air quality index (AQI) during lockdown, with the AQI of Indo-Gangetic cities improving. Concentration levels of PM2.5 and PM10 decreased to nearly half during the lockdown period. The study also compared pollution levels before and after lockdown, finding a decrease in PM10 and PM2.5 levels, a noticeable downward trend in July and August, and a significant drop in BTX levels in April. The study [21] proposes a CNN-LSTM-SVR forecasting model that integrates Deep Learning and Machine Learning techniques to accurately predict ambient air quality over 14 days. The model focuses on two major pollutants, PM10 and SPM, which pose serious threats to adults and children. The model uses ensemble modelling to reduce generalization error and minimize model training complexity. The model has three layers: feature extraction, time series analysis, and fine-tuning layer. The results show a 91-96% better performance compared to baseline models like bidirectional LSTM, LSTM, bidirectional GRU, GRU, and CNN for forecasting PM10 and SPM concentrations in Odisha.

Table 2.1: Literature Review

Reference	Year	Description	Model used	Dataset used	Performance evaluation	Result	Limitations
[1]	2021	Uses bibliometric, evolutionary and Markov chain methods to explain the research did so far	CMAQ, ADMS, ANN, multiple linear regression	Top 10 disciplines and journals that were most active in three periods (1990-1999, 2000-2009, 2010-2018)	Bibliometrics and geo-graphic evolutionary tree analysis	Hotspots and trends of air pollution prediction research	Focussed more on only three major pollutants. No further discussions are made on other pollutants.
[2]	2021	Proposes a new method- ology to	IMDA-based DL	Air pollutant data of four	For timesteps-3: RMS	Outperforms the recent	The dataset is limited to only

		forecast air pollution using a DL model called IMDA	model, variational autoencoder	US states	E-13.44 MAE-5.620 R2 -0.87369	LSTM, bidirectional LSTM models	four states
[3]	2020	Proposes a new method- ology to forecast air pollution using a DL model called CNN-LSTM-SVR to calculate PM10 and SPM concentrations for the next 14 days	CNN-LSTM-SVR DL model	Time series air pollution dataset of Odisha state	For PM10: RMSE -6.29 MSE-1317.32 MAE-35.52 For SPM: RMSE-2.52 MSE-6.36 MAE -2.11	Time series air pollution dataset of Odisha state	The dataset is limited to only one state.
[4]	2021	Investigate the studies related to DL under frame-work of smart cities	DL models	Research papers related to DL	RMSE, MAE	Hotspots and trends of air pollution in DL	Focused only on one pollutant
[5]	2021	Analyzes and compares three different models on the dataset to predict air quality.	Linear Regression, CNN and Random Forest	Kaggle's "Air quality data in India".	For city-day data: MSE-936 For city-hour data: MSE-1834	Random Forest yielded the best outcome for the data from City Day. CNN-produced the best result for city hour data.	Focused only on one pollutant and limited dataset
[6]	2020	Analyzed different ML techniques for forecasting PM2.5 on "Taiwan Air Quality Monitoring Net-work dataset	Random Forest, Gradient Boosting Regressor, K-neighbor Regressor,	Taiwan Air Quality Monitoring Network dataset (TAQMN)	R2 -0.9336 RMSE-0.1302	Using the TAQMN data, the gradient boosting regressor model performs better in	Focused only on one pollutant PM2.5. The difference between the results of the predictive models and other model are close to each

		(TAQMN)" dataset	decision tree regressor (CART).			predicting air pollution.	other. Not
[7]	2021	The paper studies the behavior of pollutants be- fore and during lockdown across the top 10 listed polluted cities.	Comparative evaluation of dataset	Air Quality data in India (2015-2020)	AQI	The pollution levels especially PM2.5 and PM10 have decreased during lockdown and the air quality is improved in major cities	No model is used
[8]	2021	Compared three different architectures LSTM, GRU and Simple RNN with ARIMA model on different sliding window lengths 6h, 12h and 24h and proposed the best model for PM concentration.	ARIMA, LSTM, GRU, Simple RNN	Measurements of the air quality made by sensors placed in various Skopje areas	RMSE for 24h-3.33 12h-24.77 6h-62.28	Predicted the best model for each time horizon	restricted dataset and the ARIMA model's performance was not assessed throughout a 24-hour period
[9]	2022	This paper uses two algorithms Naive Bayes and Random Forest on a 32516 record dataset and compare's the accuracy of the model	Naive Bayes and Random Forest	32516 record dataset	Accuracy Random Forest- 99.154% Naive Bayes- 97.32%	Random Forest outperforms Naive Bayes	only two models have been used.
[10]	2023	This paper suggests fuzzy symmetry to handle uncertainty in data which then involve TFN followed by defuzzification	ARIMA and Linear Programming	SMKPasir Gudang, SMK Bukit Rambai, Komplek Sukan Langkawi,SK Cenderawasih	RMSE of different datasets are compared	Utilizing LP to identify middle point is more re- liable	No other evaluation metrics other than RMSE is calculated

[11]	2023	This paper collected data from CPCB, Delhi and tested data on three models SARIMAX, Prophet and LSTM	SARIMAX, Prophet and LSTM	Data collected from CPCB contains 29531 rows of data	RMSE SARIMAX-39.10, Prophet-37.83 and LSTM-120.700	SARIMAX outperforms compared to other two models	For better prediction they have to expand the scope of study beyond Delhi
[12]	2023	This paper uses different machine learning models to analyze the data and compared the results of all models	Li near Regression, Lasso Regression, K Nearest Neighbors and Random Forest	Dataset accessible to the public from Sri Lanka. Data on PM2.5, NO2, C0, and SO2 concentrations are included in the collection, which spans the years January 2018 through December 2021.	Accuracy: Li near regression-89.23%, Lasso Regression-88.95%, KNN-92.65% and Random Forest-99.87%	When compared to other models, the Random Forest model yields the best overall accuracy and the lowest RMSE value.	For better prediction they would have gone for deep learning models as the dataset is large

CHAPTER 3
METHODOLOGY

For forecasting air pollution, we have prepared our own dataset and tested the dataset on the existing machine learning models. Various authors have proposed various methods but used the existing machine learning and deep learning models. The results of any model depend on the dataset used. For forecasting we need the time series dataset which consists of AQI values. So, depending upon the dataset and the model results may vary.

3.1. Dataset Description

In this research, we have prepared our own dataset that consists of AQI values of seven major pollutants such as PM2.5, PM10, NH3, NO2, SO2, CO and OZONE of five different stations or areas in Delhi. The algorithm used to determine air quality was trained using information obtained from the Central Pollution Control Board (CPCB). Here, the air quality index is the predicted output. Regression algorithms were utilized in this instance to train the dataset because the AQI value is a numerical value rather than a categorical classification. The features of the dataset are as follows:

- Station Name
- Date (d/m/y)
- PM2.5
- PM10
- NO2
- NH3
- SO2
- CO
- OZONE concentrations

3.2 Data Analysis and Model building

- **Preprocessing the data:**

Before training the dataset, it is important to check for the missing values in the dataset. This paper handled the missing values using “simple Imputer” from “sklearn.impute”, where the missing values are replaced with the mean of each feature

- **Training the model:**

Once the data preprocessing is done, the dataset is divided into two phases Training phase and Testing phase.

Training phase: After examining the data in the dataset, the ML system builds a model that is represented as a line or curve using the chosen ML technique. In our case, we just used 80% of the dataset for training.

Testing Phase: The system receives the inputs and is then evaluated for functionality. The precision is assessed. We used 20% of the dataset across all datasets for testing.

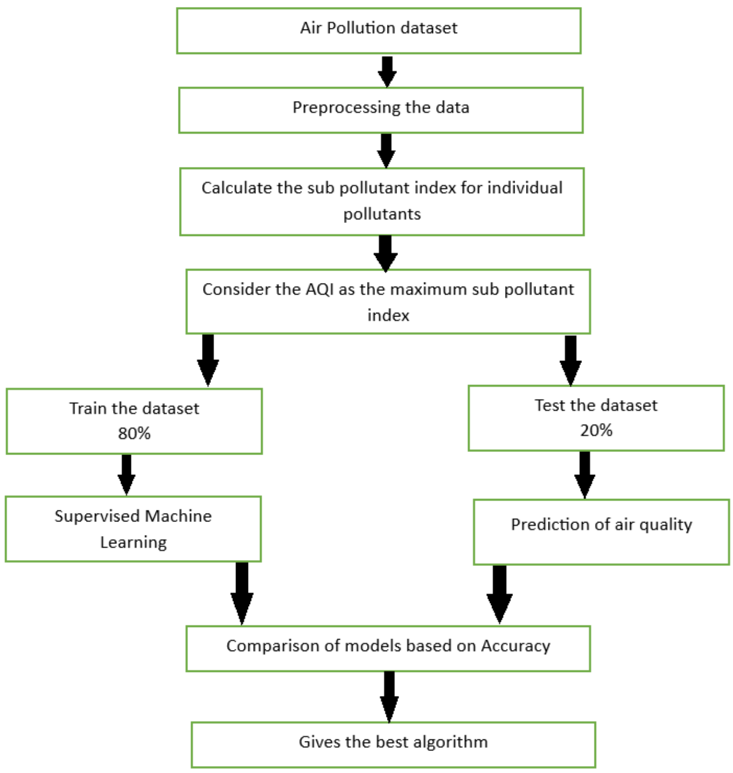


Figure 3.1: Architecture of the proposed model

The following are the algorithms that are used to train the dataset:

- Logistic Regression
- K-Nearest Neighbors
- Support Vector Machine
- Random Forest
- Decision Tree

- XGBoost

The data is trained according to the above method and used by python libraries (like pandas, sci-kit-learn etc.).

3.3 Model Evaluation

Using a pre-processed dataset, the model is used to predict the AQI value after the training phase is finished. Accuracy is used to determine which machine learning algorithm is the best. For regression problems, commonly used evaluation metrics are:

- Accuracy
- Precision
- Recall
- F1 Score

The formulae for the above evaluation metrics are shown below

Accuracy = (TP+TN)/ (TP+TN+FP+FN)

Precision = TP/(TP+FP)

Recall = TP/(TP+FN)

F1 Score = 2*((Precision * Recall)/ (Precision + Recall))

Table 3.1: Confusion matrix

	Predicted Negative (0)	**Predicted Positive (1)**
Actual Negative (0)	True Negative (TN)	False Positive (FP)
Actual Positive (1)	False Negative (FN)	True Positive (TP)

Where TP (True Positives): Correctly predicted positive instances

TN (True Negatives): Correctly predicted negative instances

FP (False Positives): Incorrectly predicted positive instances

FN (False Negatives): Incorrectly predicted negative instances

The two main pollutants, SO2 and NO2, are predicted by the author of [13] using a linear regression model. Five different machine learning techniques were employed by the author of [14], who compared the outcomes using the single parameter of RMSE values. However, six machine learning methods and seven contaminants were employed for prediction in this study.

The results were predicted based on comparison between the models using accuracy, recall, precision and F1 Score.

3.4 Machine Learning Models

1. **Logistic Regression:** Logistic regression is a statistical model used for applications involving binary categorization. It calculates the likelihood that a specific input is a member of a particular class. assumes that the log-odds of the result and the input features have a linear relationship. models the likelihood using a logistic function and outputs probabilities that are thresholded to produce binary forecasts. This model has shown an accuracy of 78%.

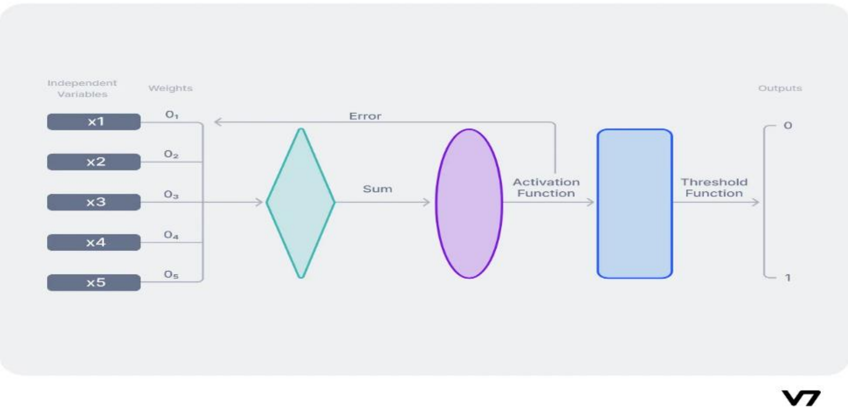


Figure 3.2: Logistic Regression architecture

2. **K Nearest Neighbor:** A non-parametric, instance-based learning approach called k-Nearest Neighbors (KNN) is employed for both regression and classification problems. Based on the similarity between the input sample and its k closest training samples, it generates predictions. The number of neighbors taken into account when making the forecast is indicated by the value of K. computationally demanding since every training instance must be stored and compared to. KNN has shown an accuracy of 91%.
3. **Support Vector Machine:** Support Vector Machine (SVM), a supervised learning technique, is used to solve regression and classification issues. It ascertains which feature space hyperplane most effectively splits the different classes. has the ability to use kernel functions for both linear and non-linear classification. increases the margin (support vectors) between the closest points of various classes. Specifically, though, categorization difficulties are resolved with it. an N-dimensional space that can separate

data points into classes (SVM). The term "16 hyperplane" refers to this ideal border. The extreme vectors that SVM selects to find the hyperplane are known as support vector. SVM has shown an accuracy of 84%.

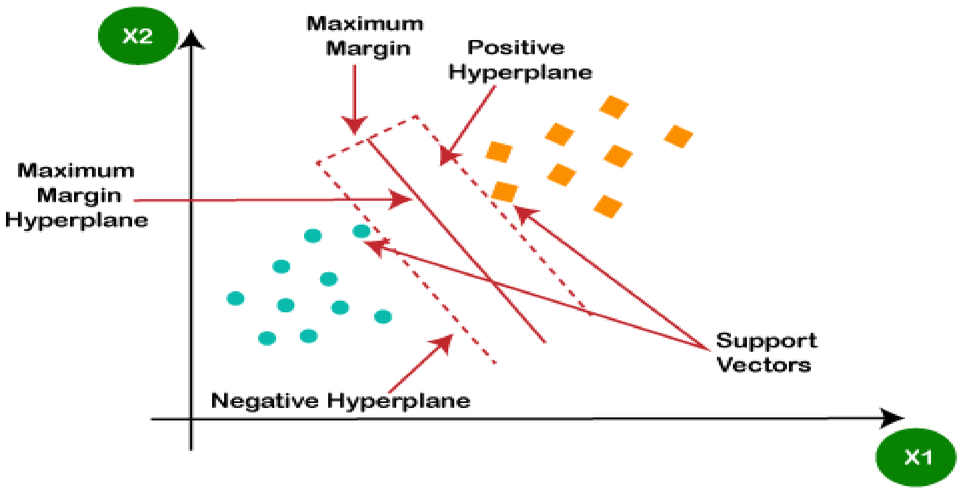


Figure. 3.3: Support Vector Machine

4. **Random Forest Regressor:** Random Forest is an ensemble learning strategy that creates many decision trees during training and combines their output to improve accuracy and decrease over-fitting. uses random feature selection and bootstrap aggregating (bagging) for robustness. lowers variance and enhances generalization through the use of multiple tree averages. provide measures of feature importance. This model has shown the accuracy of 99%, which is better compared to logistic regression, KNN and SVM.

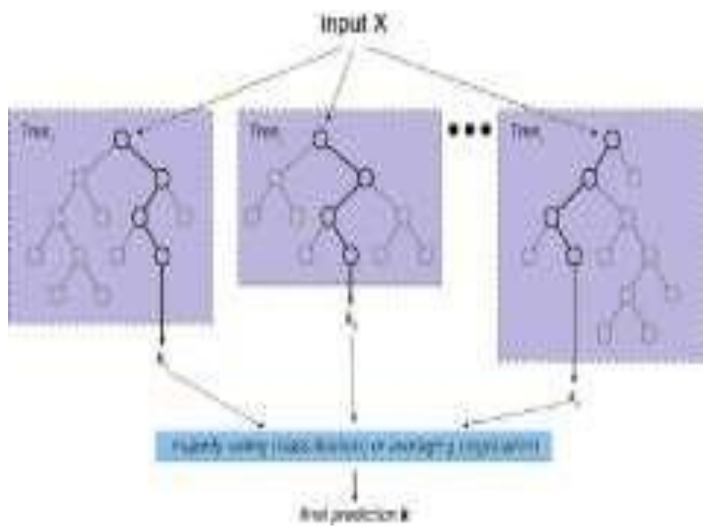


Figure 3.4: The general architecture of Random Forest

5. **Decision Tree Regressor:** The supervised learning method used to solve regression and classification problems is the decision tree. It creates a tree-like structure by dividing the data into subsets in order of importance. It is easy to understand and explain. Ability to handle confidential and digital information. It is so thin that it overfits, but pruning techniques help control this.

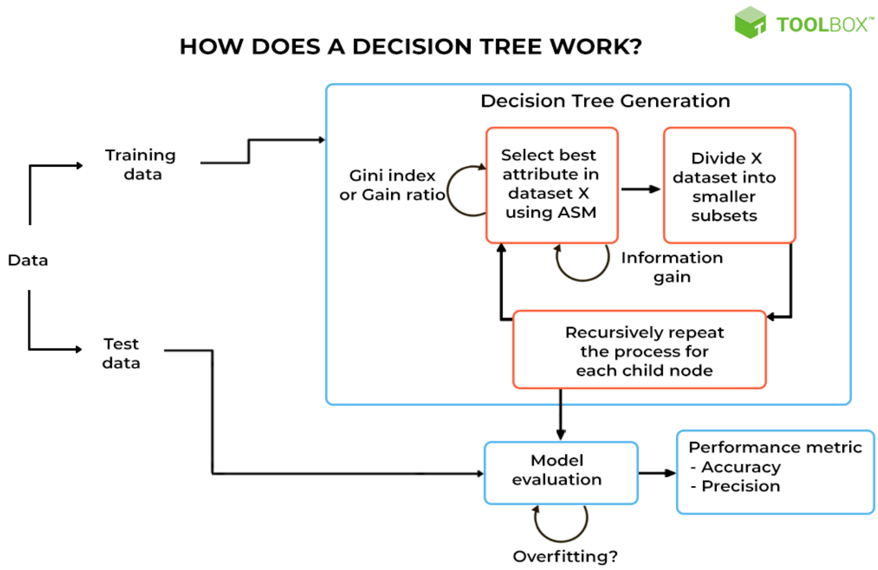


Figure 3.5: Decision Tree

6. **XGBOOST:** An effective and scalable gradient boosting framework implementation is called XGBoost (Extreme Gradient Boosting). It performs regression and classification in supervised learning problems. utilises a gradient boosting framework to merge multiple base learners' predictions and has regularisation built in to avoid overfitting. Extremely effective and frequently produces cutting-edge outcomes in machine learning contests.

CHAPTER 4
RESULTS AND DISCUSSION

Six machine learning models were used in the study to predict AQI based on different features: random forest regression, logistic regression, K-nearest neighbor, decision tree, XGBoost, and SVM. This data includes AQI values for seven major pollutants from five different monitoring stations in Delhi. Data from the Central Pollution Control Board (CPCB) was used to train the air quality algorithm. Since AQI values are numerical, the data set is trained using a regression algorithm.

Table 4.1 shows that we have 100% accuracy in the Decision Tree Regressor and XGBoost models for the given dataset. Since, in comparison to the other models, the Decision Tree Regressor and XGBoost models has the highest overall accuracy. It is possible to determine which model is most appropriate for this prediction process.

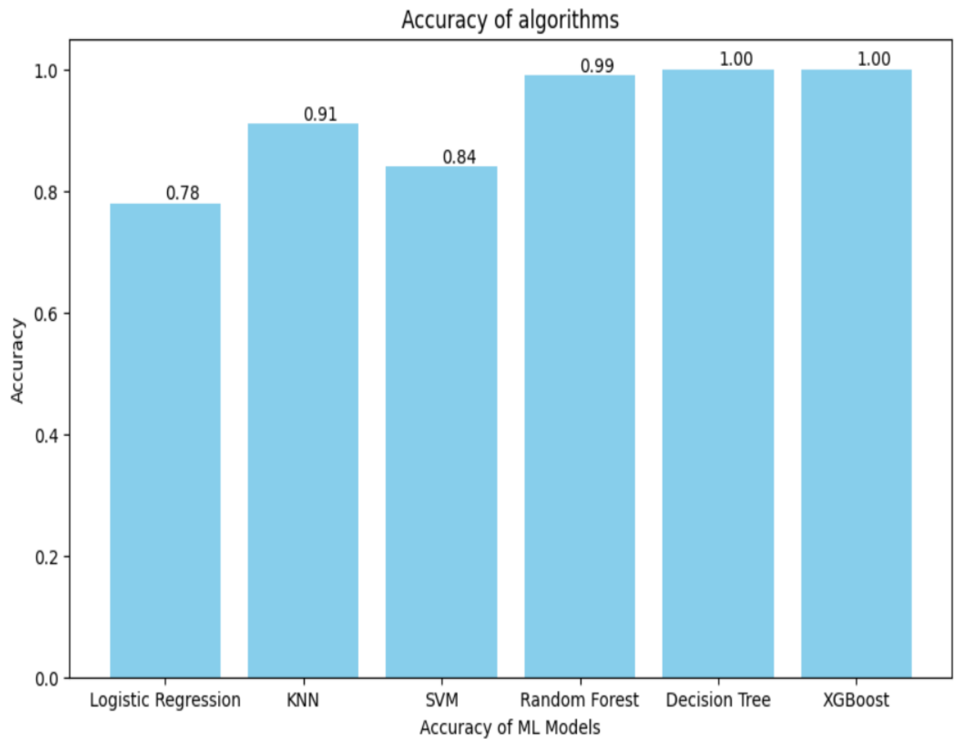


Figure 4.1: Accuracy of Machine Learning Models

The confusion matrix of machine learning models are shown below.

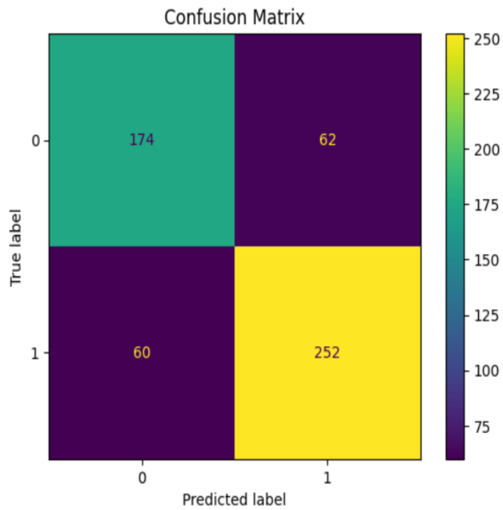


Figure 4.2: Logistic Regression

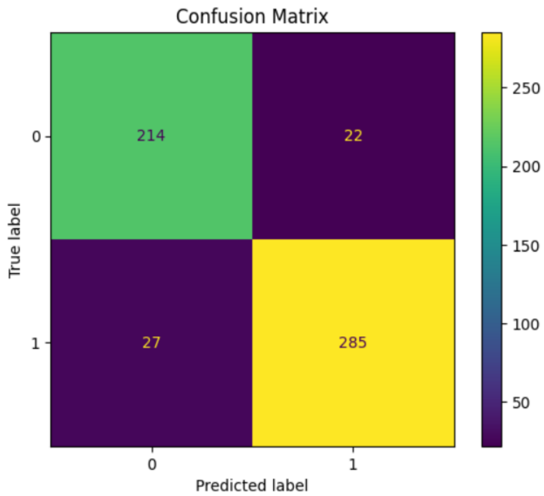


Figure 4.3: K Nearest Neighbor

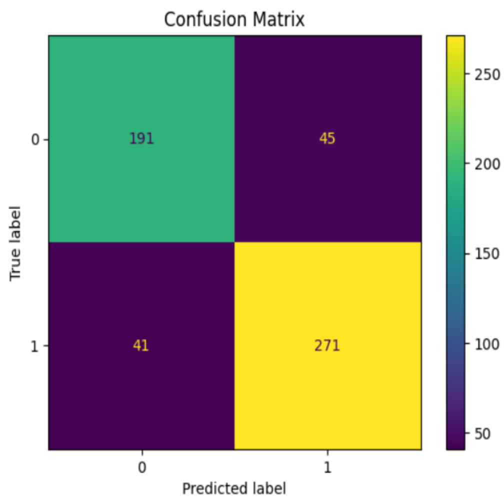


Figure 4.4: SVM

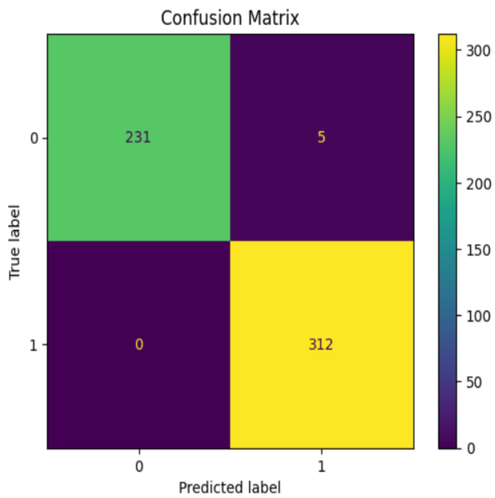


Figure 4.5: Random Forest

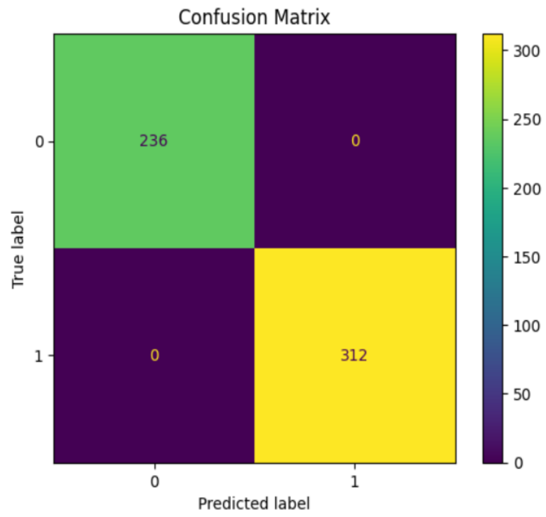


Figure 4.6: Decision Tree

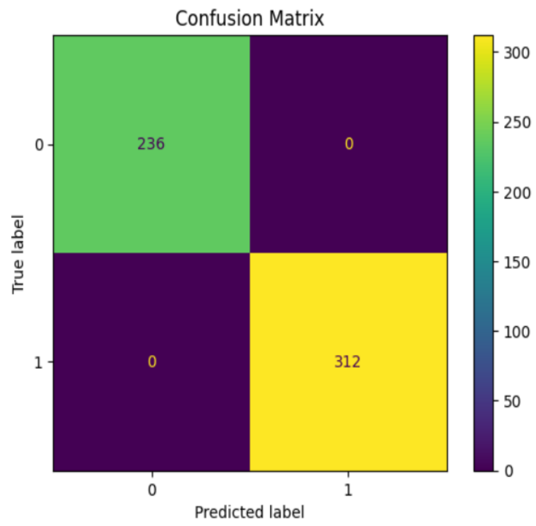


Figure 4.7: XGBoost

Table 4.1: Machine Learning Models comparison

Machine Learning Model	Precision	Recall	F1 score	Accuracy
Logistic Regression	0.80	0.81	0.81	0.78
KNN	0.93	0.91	0.92	0.91
SVM	0.86	0.87	0.86	0.84
Random Forest	0.98	1.00	0.99	0.99
Decision Tree	1.00	1.00	1.00	1.00
XGBoost	1.00	1.00	1.00	1.00

CHAPTER 5
CONCLUSION

In the twenty-first century, air pollution poses serious threats to public health as well as the environment. The accuracy of aviation forecasts and predictions is an effective way to shield individuals from pollution exposure by implementing early-warning systems. Precise projections of air pollution aid in increasing public consciousness, enabling marginalized communities to make arrangements, and informing legislators of health concerns. Different methodologies that were proposed in different papers were discussed. The approaches used by different researchers to improve the accuracy of forecasting the air pollution were seen and tabulated based on approach, dataset and model used. Through this thesis, how essential to measure the air quality to improve air pollution is understood. This thesis selected some of the areas in Delhi and AQI values for a number of pollutants from CPCB from January 1st, 2023 to December 31st, 2023, including PM2.5, PM10, SO2, CO, OZONE, NH3, and NO2 are gathered. The dataset is trained and tested on some of the ML models such as Logistic Regression, KNN, SVM, Random Forest, Decision Tree and XGBoost after handling the missing values.

The findings reveal that both Decision Tree and XGBoost provides the highest accuracy on the provided dataset compared to other models.

REFERENCES

[1] GI Drewil, RJ Al-Bahadili - Multicultural Education, 2021 - ijdri.com

[2] Urban Population (% of Total Population). Available online: https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?end=2019&name_desc=false&start 1960&view=chart

[3] Koshy, Jacob (17 August 2022). "Delhi's PM2.5 levels worst in the world". The Hindu. Retrieved 18 August 2022.

[4]" Out of here in November: In Delhi-NCR, smog break is becoming the new annual vacation". The times of India. 5 November 2023.

[5] Liao, Kuo, Xiaohui Huang, Haofei Dang, Yin Ren, Shudi Zuo, and Chensong Duan." Statistical approaches for forecasting primary air pollutants: a review." Atmosphere 12, no. 6 (2021): 686.

[6] P. D. Reddy and L. R. Parvathy, "Prediction Analysis using Random For est Algorithms to Forecast the Air Pollution Level in a Particular Loca tion," 2022 3rd International Conference on Smart Electronics and Com munication (ICOSEC), 10.1109/ICOSEC54921.2022.9952138. Trichy, India, keywords: 2022, pp. Machine 1585-1589, doi: learning algo rithms;Forestry;Programming;Air pollution;Prediction algorithms;Naive Bayes methods;Random forests;Machine learning;Innovative pollu tion;Prediction;Recall

[7] D. L. Robinson, "Air pollution in Australia: Review of costs, sources and potential solutions," Health Promotion Journal of Australia, vol. 16, no. 3, pp. 213-220, 2005.

[8] K. He, H. Huo, and Q. Zhang, "Urban air pollution in China: current status, characteristics, and progress," Annual review of energy and the environment, vol. 27, no. 1, pp. 397-431, 2002.

[9] Doreswamy, Harishkumar K S, Yogesh KM, Ibrahim Gad, Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models, Procedia Computer Science, Volume 171, 2020, Pages 2057-2066, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.04.221.

[10] 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, no. 2, p. 99, 2019.

[11] Chauhan, D., Thakur, J. (2014). Data mining techniques for weather prediction: A review. International Journal on Recent and Innovation Trends in Computing and Communication, 2(8), 2184-2189.

[12] X. Xi et al., "A comprehensive evaluation of air pollution prediction improvement by a machine learning method," in 2015 IEEE international conference on service operations and logistics, and informatics (SOLI), 2015: IEEE, pp. 176-181.

[13] S. Singh, V. Kumar, Z. Ahmed and K. Mittal, "Delhi Air Pollution Prediction: A Comparative Analysis using Time Series Forecasting," 2023 International Conference on

Disruptive Technologies (ICDT), Greater Noida, India, 2023, pp. 604-608, doi: 10.1109/ICDT57929.2023.10151445. keywords: {Machine learning algorithms;Biological system modeling;Time series analysis;Air pollution;Control systems;Calibration;Reliability;Air pollution;Machine Learning;SARIMAX;Prophet;LSTM}

[14] M. S. Che Lah, N. Arbaiy and P. -C. Lin, "ARIMA-LP: A Hybrid Model for Air Pollution Forecasting with Uncertainty Data," 2023 IEEE 8th International Conference On Software Engineering and Computer Systems (ICSECS), Penang, Malaysia, 2023, pp. 353-356, doi: 10.1109/ICSECS58457.2023.10256363. keywords: {Uncertainty;Time series analysis;Predictive models;Air pollution;Pollution measurement;Public healthcare;Forecasting;ARIMA forecast;fuzzy data;fuzzy data preparation;linear program;air pollution}

[15] M. Mihirani, L. Yasakethu and S. Balasooriya, "Machine Learning-based Air Pollution Prediction Model," 2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET), London, United Kingdom, 2023, pp. 1-6, doi: 10.1109/GlobConET56651.2023.10150203. keywords: {Radio frequency;Analytical models;Atmospheric modeling;Predictive models;Air pollution;Pollution measurement;Indexes;Air Pollution Prediction;Feature Engineering;Air Quality Index;Machine learning Models}

[16] P. D. Reddy and L. R. Parvathy, "Prediction Analysis using Random Forest Algorithms to Forecast the Air Pollution Level in a Particular Location," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1585-1589, doi: 10.1109/ICOSEC54921.2022.9952138. keywords: {Machine learning algorithms;Forestry;Programming;Air pollution;Prediction algorithms;Naive Bayes methods;Random forests;Machine learning;Innovative Random Forest algorithm;Naive Bayes algorithm;Air pollution;Prediction;Recall}

[17] Arsov, Mirche, Eftim Zdravevski, Petre Lameski, Roberto Corizzo, Nikola Koteli, Sasho Gramatikov, Kosta Mitreski, and Vladimir Trajkovik. 2021. "Multi-Horizon Air Pollution Forecasting with Deep Neural Networks" Sensors 21, no. 4: 1235. https://doi.org/10.3390/s21041235

[18] A. Dairi, F. Harrou, S. Khadraoui and Y. Sun, "Integrated Multiple Directed Attention-Based Deep Learning for Improved Air Pollution Forecasting," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-15, 2021, Art no. 3520815, doi: 10.1109/TIM.2021.3091511. keywords: {Atmospheric modeling;Forecasting;Predictive models;Air pollution;Deep learning;Data models;Time series analysis;Air pollution concentrations forecasting;deep learning;integrated multiple directed attention variational autoencoder (IMDA-VAE);multiple directed attention;time series}

[19] Drewll, Ghufran & Al-Bahadili, Riyadh. (2021). Forecast Air Pollution in Smart City Using Deep Learning Techniques: A Review. Multicultural Education. Volume 7. 10.5281/zenodo.4737746.

[20] H. Nandanwar and A. Chauhan, "Comparative Study Of Impact Of COVID-19 Lockdown On AQI Parameters Across Urban India," 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Pune, India, 2021, pp. 1-6, doi: 10.1109/SMARTGENCON51891.2021.9645784. keywords: {COVID-

19;Pollution;Greenhouse effect;Urban areas;Legislation;Air quality;Regulation;AQI;Covid-19;Air pollution;LocKdown;Human-health;Air quality;India;Pollutants}

[21] K. K. Rani Samal, K. Sathya Babu, A. Acharya and S. K. Das, "Long Term Forecasting of Ambient Air Quality Using Deep Learning Approach," 2020 IEEE 17th India Council International Conference (INDICON), New Delhi, India, 2020, pp. 1-6, doi: 10.1109/INDICON49873.2020.9342529. keywords: {Deep learning;Atmospheric modeling;Time series analysis;Predictive models;Air pollution;Reliability;Forecasting;Air quality;Forecasting;Deep Learning;LSTM;SVR}

[22] Gola, M., Settimo, G., & Capolongo, S. (2019). Indoor air quality in inpatient environments: a systematic review on factors that influence chemical pollution in inpatient wards. Journal of Healthcare Engineering, 2019.

[23] Sharma, S. B., Jain, S., Khirwadkar, P., & Kulkarni, S. (2013). The effects of air pollution on the environment and human health. Indian Journal of Research in Pharmacy and Biotechnology, 1(3), 391-396.

[24] Lah, Muhammad Shukri Che, and Nureize Arbaiy. 2020. “A Simulation Study of First-Order Autoregressive to Evaluate the Performance of Measurement Error Based Symmetry Triangular Fuzzy Number.” Indonesian Journal of Electrical Engineering and Computer Science 18(3): 1559–67.

[25] Battineni, G., Chintalapudi, N., & Amenta, F. (2020). Forecasting of COVID-19 epidemic size in four high hitting nations (USA, Brazil, India and Russia) by Fb-Prophet machine learning model. Applied Computing and Informatics, (ahead-of-print).

[26] Jiang, S. Y., Ma, A., & Ramachandran, S. (2018). Negative air ions and their effects on human health and air quality improvement. International journal of molecular sciences, 19(10), 2966.

[27] World Health Organization. Regional Office for Europe. (2004). Health aspects of air pollution: results from the WHO project "Systematic review of health aspects of air pollution in Europe". Copenhagen: WHO Regional Office for Europe.

[28] Bai, Lu, Jianzhou Wang, Xuejiao Ma, and Haiyan Lu. 2018. “Air Pollution Forecasts: An Overview.” International Journal of Environmental Research and Public Health 15(4): 1–44.

[29] Marć, M., Tobiszewski, M., Zabiegała, B., de la Guardia, M., & Namieśnik, J. (2015). Current air quality analytics and monitoring: A review. Analytica chimica acta, 853, 116-126.

[30] Kelly, F. J., Fuller, G. W., Walton, H. A., & Fussell, J. C. (2012). Monitoring air pollution: Use of early warning systems for public health. Respirology, 17(1), 7-19.

[31] Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018, December). A comparison of ARIMA and LSTM in forecasting time series. In 2018 17th IEEE international conference on machine learning and applications (ICMLA) (pp. 1394-1401). IEEE.

LIST OF PUBLICATIONS

[1] The paper entitled "A Review of Machine Learning and Deep Learning Techniques for Air Pollution Forecasting" has been accepted by the 2nd International Conference on Optimization Techniques in Engineering and Technology Engineering (ICOTET 2024). It will be published in the AIP Conference Proceedings (Scopus Index) and Springer Nature Conference Proceedings (Scopus Index). The conference is scheduled on June 14-15, 2024.

[2] The paper titled "Forecasting Air Pollution in Delhi: A Comparative Analysis of Machine Learning Models on Time Series Data" has been accepted for presentation and further publication with IEEE at the 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N- 24). The conference is scheduled on December 16-17, 2024.

Acceptance : ICOTET 2024

1 message

ICOTET2024 <icotetdgi2024@gmail.com> Tue, 21 May 2024 at 15:00
To: Gangadhar Sai <gangadharsai192@gmail.com>
Cc: icotet@gnindia.dronacharya.info

Greetings from ICOTET 2024!

Dear Author (s)

We are pleased to inform you that **Paper ID 2675** entitled **" A Review of Machine Learning and Deep Learning Techniques for Air Pollution Forecasting "** submitted by you has been accepted by the 2nd International Conference on Optimization Techniques in Engineering and Technology Engineering (ICOTET 2024).

You are advised to register for the conference by 22nd of May, 2024! Payment details for registration can be found at the bottom of this email.

You are requested to fill out the following Google form for the registration and payment information etc.:

https://forms.gle/mqyRFhx45hqkJ6cx7

All the registered and presented papers for the 2nd ICOTET 2024 will be published in the AIP Conference Proceedings (Scopus Index) and Springer Nature Conference Proceedings (Scopus Index). Please note that the plagiarism level of the paper should not exceed 15%.

For further details, please visit the official website: https://www.icotet.in/registration

Thanks & Regards

Organizing Committee

ICOTET 2024.

Bank Transfer Details:	
Bank Name	Canara Bank, Jagat Farm, Greater Noida
Account Name	Dronacharya Group of Institutions
Account Number	88951010000239
IFSC Code	CNRB0002807
MICR Code	110015485

SBI

Other Bank Transfer

INB Reference Number	IRY5903073	22-May-2024 [09:54 AM IST]
Debit Transaction Status	Processed	

Debit Account Details

SBI Account No	Account Type	SBI Branch	Amount	Commission Amount	Transaction Type	UTR Number
00000037451474350	Savings Account	GONEGANDLA	INR10,000.00	INR0.00	NEFT	SBIN324143832047

Credit Account Details

Account No.	Bank	Branch	Transfer Type	Amount	Purpose
88951010000239	CANARA BANK	GREATER NOIDA	NEFT	10,000.00	ICOTET conference fee

MoE's INNOVATION CELL (Government of India)

DRONACHARYA
Group of Institutions

Scopus

2nd International Conference on Optimization Techniques in Engineering and Technology

AIP Conference Proceedings

ICOTET-2024

Certificate of Participation

This certificate is awarded to **Prof./Dr./Mr./Ms.** ....Thammali Gangadhar........ for ~~participating~~/**presenting** a **paper titled** ...A Review of Machine Learning and Deep Learning Techniques for Air Pollution Forecasting...

In International Conference On Optimization Techniques In Engineering & Technology,

Organized by Dronacharya Group of Institutions, Gr. Noida, (U.P), India.

During 14th-15th June,2024

Prof. (Dr.) Arpita Gupta
Director

Prof. (Dr.) Shailesh Kr. Singh
Convener

Prof. (Dr.) Pawan Sharma
Organizing Chair

Gmail

Gangadhar Sai <gangadharsai192@gmail.com>

Registration Confirmation 1st IEEE ICAC2N-2024 : Paper ID 762 @ ITS Engineering College, Greater Noida

Microsoft CMT <email@msr-cmt.org> Thu, May 30, 2024 at 1:26 PM
Reply-To: "Dr. Vishnu Sharma" <vishnu.sharma@its.edu.in>
To: Thammali Gangadhar <gangadharsai192@gmail.com>

Dear Thammali Gangadhar,
DTU

Greetings from ICAC2N-2024 ...!!! Thanks for Completing your registration...!!

Paper ID- "762 “
Paper Title- “ Forecasting Air Pollution in Delhi: A Comparative Analysis of Machine Learning Models on Time Series Data “

This email is to confirm that you have successfully completed your registration for your accepted paper at ICAC2N-2024. We have received your registration and payment details. Further, your submitted documents will be checked minutely and if any action will be required at your end you will be informed separately via email.

For further updated regarding conference please keep visiting conference website www.icac2n.in or write us at icac2n.ieee@gmail.com.

Regards:
Organizing committee
ICAC2N – 2024

Note:
1. Transfer of E-copyright to IEEE and Presenting paper in conference is compulsory for publication of paper in IEEE. (For this you will be informed separately via email well before conference)
2. If plagiarism is found at any stage in your accepted paper, the registration will be cancelled and paper will be rejected and the authors will be responsible for any consequences. Plagiarism must be less then 15% (checked through Turnitin).
3. Change in paper title, name of authors or affiliation of authors is not allowed now.
4. Violation of any of the above point may lead to cancellation of registration.
5. Registration fee once paid is non-refundable.

Transaction Successful
11:55 am on 30 May 2024

Paid to

MS MANJEET SINGH ₹8,000
icac2n@ybl

Banking Name : Manjeet Singh

Transfer Details

Message

ICAC2N-2024 Paper ID-762 titled "Forecasting Air Pollution in Delhi: A comparative Analysis of Machine Learning Models on Time Series Data"

Transaction ID

T2405301155010758403576

Debited from

XXXXXX4350 ₹8,000

UTR: 451794979636

Powered by

UPI YES BANK

nmunication and Networking- ICAC2N • 1st International Conference on Advances in Computing, Communication and Networking- ICAC2N • 1st International Conference on Advances in Computing, Communication and Netw

ICAC2N IEEE UP SECTION (INDIA)

Submit Paper

HOME ABOUT GUIDELINES SPEAKERS COMMITTEE CALL FOR PAPERS REGISTER CONTACT

16th and 17th December 2024

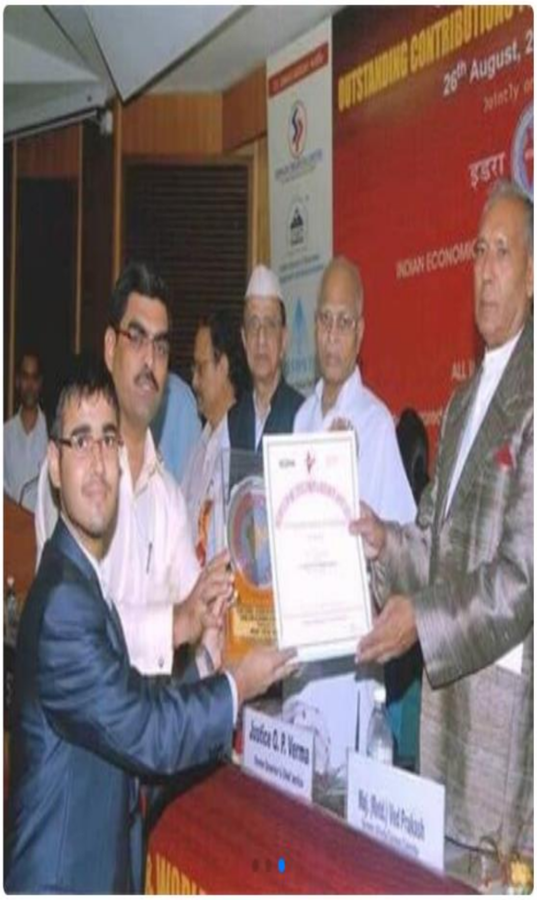
1st International Conference on Advances in Computing, Communication and Networking- ICAC2N

Conference Record Number #63387

IEEE XPLORE COMPLIANT ISBN No. 979-8-3503-5681-6

Computer Science and Engineering Department

ITS Enginerring College
Knowledge Park III, Greater Noida



About ICACCN

ICAC2N is a prestigious international conference that brings together top researchers, scientists, engineers, and scholars from around the world to share their latest research findings and experiences in computing, communication and networking. Featuring keynote speeches, technical sessions, and workshops, the conference covers a wide range of topics such as cloud computing, AI, wireless communication systems, IoT, and cybersecurity.

Important Dates

Paper Submission Starts New
15/03/2024

Paper Submission Deadline Important
31/08/2024

Notification of Acceptance
(within 4weeks from submission)
30/09/2024

Registration
31/10/2024

About College

I.T.S Engineering College, located in Greater Noida, is a distinguished institution for Engineering and Management, founded by leading professionals committed to delivering high-quality, industry-focused education at an affordable cost. With a rich history spanning around three decades, the ITS Education Group has emerged as a key player in higher education in North India, operating nine institutes with more than 20,000 students and a dedicated staff of over 1,250. The group excels in various disciplines, including Information Technology, Engineering, Management, Dental, Pharmacy, and Physiotherapy. Established in 2006 under the visionary leadership of Chairman Dr. R.P. Chadha, ITS Engineering College holds NBA accreditation, showcasing its dedication to providing top-notch education. Approved by AICTE and affiliated with AKTU-Lucknow, the college offers diverse programs, such as B. Tech in Civil, Computer Science & Engineering (CSE), CSE-Artificial Intelligence and Machine Learning, CSE-Data Science, Electronics and Communication Engineering, Electronics Engineering (VLSI Design and Technology), Electrical and Computer Engineering, Mechanical Engineering, as well as an MBA program in Marketing, Finance, Human Resource, Operations, International Business and Information Technology.

About Department of Computer Science

The Computer Science & Engineering Department at ITSEC offers a UG program in Computer Science & Engineering is designed to equip individuals with the necessary computer skills to succeed in the industry, and provide practical knowledge to prepare them for professional readiness. The program is structured to cover both the theoretical foundations and latest developments in computing solutions. Throughout the program, CSE students will develop expertise in algorithm and software design, development, and testing, and will learn to create effective computing solutions that address various social, economic, and organizational domains. It also offers B. Tech in Artificial Learning and Machine Learning at ITS Engineering College is an undergraduate programme with advanced learning solutions imparting knowledge of advanced innovations like Machine Learning, often called Deep Learning and Artificial Intelligence. Apart from CSE and AIML, ITSEC offers B. Tech in the specialization of Data Science which is a rapidly growing field that involves the analytical technique of extracting usable information from a pool of data for the benefit of businesses. Corporations require this valuable data for their strategic planning, forecasting, fraud detection, and decision-making amongst many other uses.

5,847

Visitors

Speakers for ICACCN 2024

National Speaker

Prof. (Dr.) Sri Niwas Singh

Director, Atal Bihari Vajpayee- Indian Institute of Information Technology and Management (ABV-IIITM), Gwalior, India / Chairman, India Council, IEEE India.

National Speaker

Prof. (Dr.) M.N.Doja

Director IIIT, Sonepat

National Speaker

Prof. (Dr.) S. K. Singh

Vice Chancellor, Rajasthan Technical University, Kota

National Speaker

Dr. Akhilesh Tiwari

Secretary IEEE and Associate Professor in IIIT Allahabad

National Speaker

Dr. Virender Ranga

Associate Professor, Department of Information Technology, Delhi Technological University

National Speaker

Prof. (Dr.) P. Nagabhushan

Vice-Chancellor, Vignan 's Foundation for Science, Technology &Research, Andhra Pradesh

National Speaker

Prof. (Dr.) Satish K. Singh

IIIT Allahabad / Section Chair, IEEE UP Section.

International Speaker

Prof. (Dr.) Shaymaa R. Tahhan

Laser and Optoelectronics Department, College of Engineering, Al-Nahrain University, Baghdad, Iraq

International Speaker

Prof. (Dr.) Anand Nayyar

School of Computer Science, Faculty of Information Technology, Duy Tan University, Da Nang, Viet Nam


ICOTET 2024
www.icotet.in
AIP Publishing
AIP Conference Proceedings
2nd International Conference on Optimization Techniques in Engineering and Technology
(June 14 - 15, 2024)
DRONACHARYA Group of Institutions
Scopus
Home
About Conference
About College
Success Story ICOTET-2022
Speakers
Important Dates
Schedule of Conferences
Registration
Authors Guidelines
Call For Paper
Committee
Gallery
Sponsors
Contact Us
Deposit the Registration Fee and Register for the conference on following link before : - 25 May, 2024
Registration Link: https://docs.google.com/forms/d/e/1FAIpQLSc3AJqhad9wycBQ_th-uqk-IMxRyhp_dDTN94mpEofYzbnF_w/viewform
ICOTET - 2024
Organized By :
Dronacharya Group of Institutions, Greater Noida
Let's Chat!
ICOTET 2024 CONFERENCE PROCEEDING WILL BE PUBLISHED IN AIP CONFERENCE PROCEEDING, SCOPUS AND WEB OF SCIENCE INDEXED (APPROVED)
ICOTET 2024
The 2nd International Conference on Optimization Techniques in Engineering and Technology Engineering (ICOTET 2024) is an inventive event organized with the motive of making available an open international forum for researchers, academicians, technocrats, scientists, engineers, industrialists and students around the globe to exchange their innovations, interact and share the research outcomes which may lead the young researchers, academicians and industrialist to contribute to the global society. ICOTET 2024 is being organized on 14th – 15th June 2024 at Dronacharya Group of Institutions, Greater Noida, UP, India.
About DGI
Dronacharya Group of Institutions, Greater Noida
To impart Quality Education, to give enviable growth to seekers of learning, and to groom them as World Class Engineers and Managers competent to match the expanding expectations of the Corporate World has been an ever-enlarging vision extending to new horizons of the Dronacharya Group of Institutions.
'Dronacharya Group of Institutions' is committed to serving society and improving the mode of life by imparting high-quality education in Engineering and Management, catering to the explicit and implicit needs of the students, society, humanity and industry. 'Shiksha evam Sahayata,' i.e. Education and Help, are the two words etched on our banner soaring higher year after year.
Vision
Instilling core human values and facilitating competence to address global challenges by providing Quality Te
Education.
Mission
Let's Chat!

Success Story : ICOTET-2022

The inaugural ICOTET-2022 took place at the Dronacharya Group of Institutions in Greater Noida, India. The event drew 150 research paper submissions from around the world, covering diverse subjects that reflect the conference's focus on emerging trends across various fields. With a commitment to academic rigour, a stringent peer-review process led to the acceptance of 110 research papers. These selected papers showcased innovative ideas and insightful findings, enhancing scholarly discussions. ICOTET serves as a testament to global knowledge pursuit, innovation, and collaboration. Each year, it aims to elevate academic discourse standards, encourage interdisciplinary engagement, and illuminate emerging trends across diverse areas of knowledge

Meet The Speakers

Prof. (Dr.) Pradeep Kumar

Professor, Department of Mechanical & Industrial Engineering, Indian Institute of Technology, Roorkee, Former Vice chancellor, Delhi Technological University, Delhi, India

Dr. Somnath Chattopadhyaya

Professor and Head, Department of Mechanical Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, India.

Let's Chat!

Dr. Ravinderjit Singh Walia

Professor, Punjab Engineering College (Deemed to be University), Chandigarh, India

Dr. Qasim Murtaza, PhD-Ireland, PDF-UK

Professor, Delhi Technological University, Delhi, India.

Dr. Alokesh Pramanik

Department of Mechanical Engineering, Curtin University, Bentley, Perth, WA, Australia

Dr Anish Kumar Sachdeva

Department of Industrial and Production Engineering
Dr B R Ambedkar National Institute of Technology Jalandhar, Punjab (India)

ANNEXURE-IV

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis HARNESSING MACHINE LEARNING FOR ACCURATE AIR QUALITY PREDICTION: A TIME SERIES ANALYSIS Total Pages 38 Name of the Scholar Thammali Gangadhar

Supervisor (s)

(1) Mrs. Anamika Chauhan

(2)____

(3)____

Department INFORMATION TECHNOLOGY, DELHI TECHNOLOGICAL UNIVERSITY

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: Turnitin Similarity Index: 13% , Total Word Count: 4402

Date: 30/05/2024

Candidate's Signature **Signature of Supervisor(s)**

PAPER NAME

HARNESSING MACHINE LEARNING FOR ACCURATE AIR QUALITY PREDICTION: A TIME SERIES ANALYSIS

AUTHOR

Thammali Gangadhar

WORD COUNT

4402 Words

CHARACTER COUNT

26220 Characters

PAGE COUNT

36 Pages

FILE SIZE

10.9MB

SUBMISSION DATE

May 30, 2024 8:35 PM GMT+5:30

REPORT DATE

May 30, 2024 8:36 PM GMT+5:30

13% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 8% Internet database
- 7% Publications database
- Crossref database
- Crossref Posted Content database
- 9% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 8 words)