Sustainable Climatic Metrics Determination with Ensemble Predictive Analytics

Ashis Pattanaik, Vandana Sharma, Kanhaiya Kunj, Sushruta Mishra, Celestine Iwendi and Jude Osamor

Sustainable Climatic Metrics Determination with Ensemble Predictive Analytics

Ashis Pattanaik
Kalinga Institute of Industrial Technology, Deemed to be University, India
2105779@kiit.ac.in

Vandana Sharma, SMIEEE
Department of Computational Sciences,
CHRIST (Deemed to be University), Delhi NCR, India
vandana.juyal@gmail.com

Kanhaiya Kunj
Kalinga Institute of Industrial Technology, Deemed to be University, India
2105277@kiit.ac.in

Sushruta Mishra
Kalinga Institute of Industrial Technology, Deemed to be University, India
sushruta.mishrafcs@kiit.ac.in

Celestine Iwendi, SMIEEE
School of Creative Technologies
University of Bolton, United Kingdom
c.iwendi@bolton.ac.uk

Jude Osamor, FHEA
Department of Cybersecurity and Networks
Glasgow Caledonian University
Glasgow, G4 0BA Scotland, UK
jude.osamor@gcu.ac.uk

***Abstract*- Sustainability depends upon some major climate factors which play a major role in ensuring and assuring that sustainability would be maintained if the range of safe values of parameters are maintained. Linear regression and random forest are few among the machine learning models that were employed in order to determine the dependency of each factor on sustainability. The climate data from Delhi from 1971 to 2020 is utilized for the study considering the variables like temperature, precipitation, humidity and atmospheric carbon dioxide concentration which were collected from various authorized sources such as the Indian Meteorological Department and the Central Pollution control board. After studying various factors involved in determining climate sustainability we found out that temperature and atmospheric carbon dioxide concentration have the greatest impact with a percentage of 45% and 30% respectively. Sectors like agriculture, forestry, energy and water management are majorly dependent on these key deciding factors. The R square value was determined to be 0.86 and 0.82 respectively for machine learning models implemented. We found that the random forest model had a better score in comparison to the linear regression model. With this study we thus found out that, how machine learning models can be trained and tested in order to predict the future outcomes for sustainability. This study demonstrates the importance of climate monitoring for maintaining sustainability.**

Keywords-Climate change, Effective monitoring, Analysis of climate factors, Sustainability, Machine learning, Random forest, Linear regression, Temperature, Atmospheric carbon dioxide concentration.

I. INTRODUCTION

Climate change significantly has an impact on the environment as well as the human health so it requires continuous evaluating so as to check the contributing factors to promote sustainability. The climate parameters which are most influential in affecting sustainability can be identified by using machine learning techniques like linear regression and random forest. These include more frequent and intense heat waves, a decrease in cold-related fatalities, a rise in instances of floods and droughts, alterations in the spread of diseases carried by vectors like mosquitoes, as well as influences on the likelihood of disasters and malnutrition. [1] When we go through severe weather occurrences like heatwaves, floods, and droughts, their effects are noticeable immediately and can also last for a while, resulting in loss of lives. Additionally, the climate influences the quantity of pollutants found in the atmosphere. For instance, we can observe differences in the levels of troposphere ozone pollution across various parts of Europe, though we're still working on fully grasping the intricate links between these factors. As per insights from climate science specialists, the unusual weather patterns witnessed in the past two decades might be a sign of a longer-term shift in typical temperature, rainfall, and the frequency of extreme weather events [2] Moreover, there's a greater chance of diseases like hepatitis E, stomach problems, and lepidopterist breaking out in areas affected by floods, especially in places where sanitation is lacking and communities have been forced to move from their homes [3]. Concerning the unfortunate results arising from heatwaves and the smoky aftermath of fires, it's often the worsening of preexisting health conditions that takes the most blame, rather than solely the immediate effects such as heat-related fatigue. People who are more susceptible to these harmful outcomes, like older adults, young children, and those who already have health troubles, are more likely to experience the full force of these adverse effects [4]. Evaluating the effects of contemporary climate shifts is closely linked with predicting what lies ahead in the upcoming years. This process assists us in identifying which crops and regions are currently at risk. From the 1970s onwards until the present day, the planet's average surface temperature has been gradually increasing, with an approximate rise of 0.16°C to 0.18°C every decade [5]. Machine learning emerges as a remedy for the quandary of balancing the flexibility of functional patterns with the precision of estimation.

In contrast to the customary method of regression analysis that necessitates presuming the manner in which variables interact, machine learning charts a distinct course. It employs a benchmark of optimization to decipher the intricate connections between the final outcome and the contributing elements. Although the notion of "data mining" within machine learning might stir apprehensions in the domain of traditional economics, it's important to acknowledge that data mining already plays a role in empirical research. Typically, researchers opt to present models that cater to their audience (Athey, 2018). By comparison, machine learning holds an upper hand by employing lucid and transparent criteria to select the most appropriate model specification [6]. Humidity refers to the extent of water vapor existing in the atmosphere. There are two commonly used ways to quantify humidity: specific humidity and relative humidity. Specific humidity measures the grams of water vapor within a kilogram of air. Conversely, relative humidity gauges the actual water vapor present in the air compared to the point at which the air becomes saturated [7]. Notably, both specific and relative humidity equations incorporate the current water vapor content in the air. To achieve sustainable results in monitoring climate variables, techniques like linear regression and random forest from machine learning can be applied.

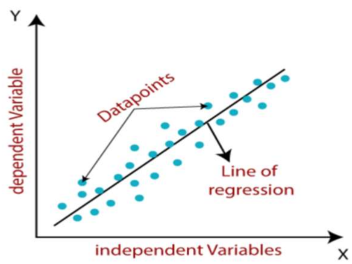

Fig.1.Linear Regression graphical representation

Linear regression earns its name from its ability to showcase the linear connection existing between one or multiple independent variables (x) and a dependent variable (y). As shown in fig 1, linear regression model is represented graphically to show the dependence of dependent variables on independent variables considering the data points, through the line of regression. It gives an insight to the dependence of dependent variables. It also highlights the data points which are in turn represented on the lone of regression. This algorithm meticulously evaluates this linear interconnection to fathom the extent to which alterations in the independent variable sway the dependent variable. Consequently, the outcome is a model that presents a straight line, its incline mirroring the interdependence between these variables.

Mathematically, we can represent a linear regression in equation 1 as:

$$y = a0 + a1x + \varepsilon \quad (1)$$

Here,

Y= Dependent Variable (Target Variable)

X= Independent Variable (predictor Variable)

a0= intercept of the line (Gives an additional degree of freedom)

a1 = Linear regression coefficient (scale factor to each input value).

ε = random error

The data used to teach the Linear Regression model contains numbers for both the x and y elements. Here, the y element stands for the measured result of sustainability, while the x element represents the specific climate aspect being looked at. The steepness (m) of the model shows how shifts in the x element affect sustainability, and the starting point (b) reveals the basic level of sustainability when the x element isn't there.

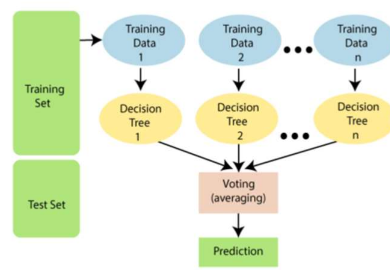

Fig.2. Random Forest Model approach diagram

Random Forest demonstrates its worth in evaluating and predicting how elements such as temperature, rainfall, moisture, and wind velocity impact sustainability. The fig 2 shows the random forest model approach and shows the transition between the training set and the test set. It also shows that the training data is trained and the decision is made and thus after voting and averaging we get the prediction as the outcome. As shown in the fig 2 that many training sets are trained in order to obtain a decision tree which is further used to find the average of the data and thus the prediction is made. The formula it uses, Y = f(X) + ε, can be explained like this: Y stands for the result, X represents the inputs, f symbolizes the forecasting function, and ε signifies the error. On the other hand, Linear Regression is formulated as shown in equation 2:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \dots + \beta nXn + \varepsilon \quad (2)$$

The research analyzes weather information from 1971 to 2020, centered on Delhi, India, to grasp its impact on sustainability. It also explores machine learning's role, specifically Linear Regression and Random Forest, in predicting climate patterns. Results show their effectiveness in gauging climate change, highlighting temperature, humidity, and air quality as primary factors. This aids in shaping strategies to mitigate climate change.

Main contributions of the research work are as follows:

- Examine and interpret key climate influencers - temperature, rainfall, humidity, atmospheric pressure, and wind speed.
- Construct a predictive model for future climate trends via historical data.
- Explore links between human actions (greenhouse gas emissions, land use) and climate factors.
- Uncover climate change repercussions, strategic mitigation.
- Highlight the need for climate tracking in sustainable progress.
- Contribute to climate change knowledge, endorse eco-friendly strategies.

II. LITERATURE REVIEW

Dr. C.S Kaanimozhi Selvi and G. Sowmiya proposed a machine learning based method for predicting extreme weather events such as floods and droughts in the research paper titled “Prediction of Extreme Weather Events using Machine Learning Technique" [8]. Nayak M.A and Ghosh S (2013) introduced a novel approach to predict extreme rainfall events using machine learning techniques in the research paper titled "Prediction of Extreme Rainfall Event using Weather Pattern Recognition and Support Vector Machine Classifier". They used weather pattern recognition and support vector machine(SVM) classification in combination to predict extreme rainfall events. [9]. Satya Prasad and Salwa Mohammed Nejres (2015) proposed the utilization of data mining techniques for analyzing weather data in Basra City in their research paper titled as paper "Application of Data Mining Techniques to Analyze Weather Data in Basra City". Researchers implemented various data mining techniques as clustering and association rule mining. The study found that data mining techniques can be a valuable tool for weather data analysis and extracting meaningful insights [10]. In their 2016 research titled "Early Prediction of Extreme Rainfall Events: A Deep Learning Approach," Sulagna Gope, Sudeshna Sarkar, Pabitra Mitra, and Subimal Ghosh introduce a profound solution to anticipate severe rainfall, serving disaster management by preemptively forecasting floods. Stressing the value of timely predictions, they introduce a deep belief network (DBN) for this purpose. Impressively, their approach accurately predicts heavy rainfall a week in advance [11]. In their 2015 paper "Topic Modeling for Extreme Event Detection in Climate Time Series," Cheng Tang and Claire Monteleoni propose a method using latent Dirichlet allocation (LDA) for identifying extreme events in climate time series [12]. In the 2011 paper "Cancer Diagnosis Using Finite Impulse Response Extreme Learning Machine," Kevin Lee, Zhihong Man, Dianhui Wang, and Zhenwei Cao propose a machine learning technique for precise cancer diagnosis with bio-informatics data. The method involves employing the finite impulse response extreme learning machine (FIR-ELM) model, trained with gene expression and clinical data from cancer patients. The approach shows accurate cancer dataset classification, surpassing support vector machines and decision trees in accuracy and training time. Though promising, the FIR-ELM model's broader bio-informatics applications and performance on larger datasets warrant further investigation [13]. "Random Forest Machine Learning Model for Weather Prediction" by Rajasekaran Meenal et al. gives a machine learning approach to predict weather using the random forest algorithm. It shows applications like agriculture, transportation, and disaster management. They compared the random forest model with decision trees and support vector machines, demonstrating its superior accuracy and prediction speed [14]. "Using Deep Learning to Predict Gridded 500-hPa Geo-potential Height from Historical Weather Data: Can Machines Learn to Predict Weather?" by Jonathan A. et al. studies and explores the potential of deep learning algorithms for weather prediction. The study signifies the significance of deep learning techniques in improving the accuracy and inclination of weather prediction models. The approach that uses a convolutional neural network (CNN) model to forecast weather conditions was proposed which was based on historical weather data. [15]. The paper titled "Machine Learning-Based Wind Speed Prediction Using Convective Weather Variables" by Bhuiyan Md Abul Ehsan et al. introduces an approach for wind speed prediction using machine learning. The suggested technique employs a machine learning framework that combines convective weather parameters like temperature, humidity, and pressure, in order to anticipate wind speed. [16]."Recent Developments in Weather Forecasting in India" by Sandeep Pattnaik shows the developments in weather forecasting in India. They show an overview of the current state of weather forecasting, including the available infrastructure for observation and forecasting, and the challenges that need to be addressed to enhance the accuracy and reliability of weather forecasts [17]. The scholarly paper titled "Advancing Flood Hazard Risk Assessment via Random Forest" authored by Z. Wang et al. introduces a cutting-edge flood hazard risk assessment model underpinned by the random forest machine learning algorithm. The authors meticulously detail the formulation of the flood hazard risk assessment model based on the random forest algorithm and its application within a case study centered in the Xinjiang Uygur Autonomous Region in China [18]. The research paper titled "Fully Nonlinear Statistical and Machine-Learning Approaches for Hydrological Frequency Estimation at Ungauged Sites" by D. Ouali et al. compares fully nonlinear statistical approaches with machine learning approaches for hydrological frequency estimation at ungauged sites. Two methods for estimating hydrological abundance are explored: the fully nonlinear Bayesian frequency analysis (FNBFA) and the support vector regression (SVR) machine learning approaches [19]. In the paper titled "Extreme Hydrometeorological Events and Climate Change Predictions in Europe" by M.M. Millán, It was described as an examination of how climate change has affected severe hydrometeorological occurrences in Europe.The difficulties of foreseeing changes in the severe occurrences in the future were also covered in the article.This may be done through enhancing monitoring and prediction capacities, bettering land use planning, and creating infrastructure and communities that are more resilient [20].

III. DATASETS AND METHODS

The study focuses on Delhi, the capital city of India, which has been facing severe environmental issues, including air pollution, heatwaves, and water scarcity [21]. The study aims to track the climate factors for sustainability using machine learning models to provide insights into future climate patterns in the region.

The research will make use of climate data spanning from 1971 to 2020 for Delhi. The data will be procured from the Indian Meteorological Department (IMD). The analysis of climate data and the modeling of future climate patterns will be conducted using Linear regression and Random Forest machine learning algorithms. Linear regression will be utilized to establish the relationships between different climate variables, such as temperature and rainfall, while identifying any trends or changes in the variables over time.

Random Forest, on the other hand, will be used to predict future climate patterns based on historical data.

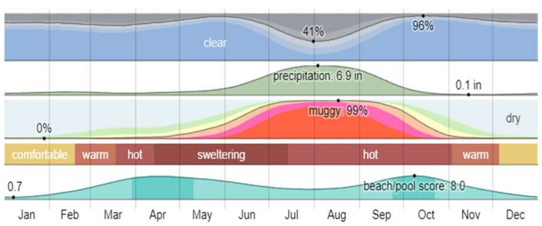

Fig.3:- Climate in delhi spread across different months of the year

Delhi boasts a pair of weather observing stations. As shown in fig 3, the climate of Delhi is spread across different months of the year. The fig 3 also show the transition between the months starting from January till December with the variations in factors involved in climate sustainability. The fig 3 also depicts the significant variations in factors like precipitation, humidity, temperature and some other factors. One is positioned at Safdarjung within the heart of the city, and the second is situated at Palam towards the southwest, adjacent to the airport. While climate data collected at Safdarjung station is believed to reflect conditions within the city, readings from Palam station provide insight into weather patterns at the airport. Delhi is positioned in the subtropical belt of the Northern Temperate area, a little way above the Tropic of Cancer, specifically at a latitude of 28°36′36″N. The amount of daylight experienced in the city varies due to the Earth's rotation, resulting in briefer days in the winter months and extended days during the summer season. This pattern aligns with the customary progression between the two solstices. Within this span, Delhi's daytime span transforms by around four hours, with an approximate two-hour distinction both at dawn and dusk. The fig 4 shows the average climatic conditions of Delhi from 1971-2020. As shown in fig 4 the climate variate through extremities like temperature, precipitation, humidity and some other factors.The fig 5 shows the average barometric pressure and wind speeds in delhi. As shown in the fig 5 it showcases and depicts average barometric pressures and wind speeds in delhi based on months of an year. The fig 5 also depicts the dependence of the factors on eachother.

V·T·E Climate data for New Delhi (Safdarjung) 1971–2020, extremes 1901–present [hide]

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Record high °C (°F)	32.5 (90.5)	34.1 (93.4)	40.6 (105.1)	45.6 (114.1)	47.2 (117.0)	46.7 (116.1)	45.0 (113.0)	42.0 (107.6)	40.6 (105.1)	39.4 (102.9)	36.1 (97.0)	30.0 (86.0)	47.2 (117.0)
Mean maximum °C (°F)	25.8 (78.4)	29.5 (85.1)	35.8 (96.4)	41.4 (106.5)	44.3 (111.7)	43.7 (110.7)	40.4 (104.7)	37.4 (99.3)	37.1 (98.8)	36.1 (97.0)	32.2 (90.0)	27.3 (81.1)	44.8 (112.6)
Average high °C (°F)	20.1 (68.2)	24.2 (75.6)	29.9 (85.8)	36.5 (97.7)	39.9 (103.8)	39.0 (102.2)	35.6 (96.1)	34.2 (93.6)	34.1 (93.4)	33.0 (91.4)	28.4 (83.1)	22.8 (73.0)	31.4 (88.5)
Daily mean °C (°F)	13.9 (57.0)	17.6 (63.7)	22.9 (73.2)	29.1 (84.4)	32.7 (90.9)	33.3 (91.9)	31.5 (88.7)	30.4 (86.7)	29.6 (85.3)	26.2 (79.2)	20.5 (68.9)	15.6 (60.1)	25.3 (77.5)
Average low °C (°F)	7.5 (45.5)	10.6 (51.1)	15.6 (60.1)	21.3 (70.3)	25.8 (78.4)	27.7 (81.9)	27.5 (81.5)	26.7 (80.1)	25.0 (77.0)	19.5 (67.1)	13.0 (55.4)	8.4 (47.1)	18.9 (66.0)
Mean minimum °C (°F)	3.5 (38.3)	6.0 (42.8)	10.7 (51.3)	16.3 (61.3)	20.5 (68.9)	22.2 (72.0)	24.3 (75.7)	23.7 (74.7)	21.9 (71.4)	15.0 (59.0)	8.8 (47.8)	4.5 (40.1)	3.1 (37.6)
Record low °C (°F)	−0.6 (30.9)	1.6 (34.9)	4.4 (39.9)	10.7 (51.3)	15.2 (59.4)	17.6 (63.7)	20.3 (68.5)	20.7 (69.3)	17.3 (63.1)	9.4 (48.9)	3.9 (39.0)	0.0 (32.0)	−0.6 (30.9)
Average precipitation mm (inches)	19.1 (0.75)	21.3 (0.84)	17.4 (0.69)	16.3 (0.64)	30.7 (1.21)	74.1 (2.92)	209.7 (8.26)	233.1 (9.18)	123.5 (4.86)	15.1 (0.59)	6.0 (0.24)	8.1 (0.32)	774.4 (30.5)
Average precipitation days (≥ 0.3mm)	2.9	3.1	3.6	2.6	4.6	7.5	13.1	14.4	7.6	1.6	0.9	0.9	62.9
Average rainy days	1.7	1.5	1.7	1.0	2.7	4.8	9.7	10.2	5.5	0.8	0.4	0.6	40.6
Average relative humidity (%) (at 17:30 IST)	57	46	37	25	28	43	63	68	60	47	52	59	49
Average dew point °C (°F)	8 (46)	11 (52)	14 (57)	14 (57)	18 (64)	22 (72)	26 (79)	25 (77)	23 (73)	18 (64)	14 (57)	10 (50)	17 (62)
Mean monthly sunshine hours	220.1	223.2	248.0	276.0	285.2	219.0	179.8	176.7	219.0	260.4	246.0	220.1	2,773.5
Mean daily sunshine hours	7.1	7.9	8.0	9.2	9.2	7.3	5.8	5.7	7.3	8.4	8.2	7.1	7.6
Mean daily daylight hours	10.6	11.2	12.0	12.9	13.6	13.9	13.8	13.1	12.3	11.5	10.7	10.3	12.2
Percent possible sunshine	67	71	67	71	68	53	42	44	59	73	77	69	63
Average ultraviolet index	3	5	6	8	9	9	8	7	6	6	5	3	6

Source 1: India Meteorological Department (sun 1971–2000);[41][42][43][44][45][46][47][48][49][50][51] Time and Date (dewpoints, 2005–2015)[52] Revised Rainfall data[53]

Source 2: Tokyo Climate Center (mean temperatures 1991–2020);[54] Weather Atlas (UV Index)[55] (Daylight)[56]

Fig.4.Average climatic conditions of delhi from 1971-2020

Average Barometric Pressure & Wind Speed of Delhi

Month	January	February	March	April	May	June	July	August	September	October	November	December	Year
Average Atmospheric pressure millibars (inHg)[57]	1,017.0 millibars (30.03 inHg)	1,014.5 millibars (29.96 inHg)	1,010.6 millibars (29.84 inHg)	1,005.4 millibars (29.69 inHg)	1,000.5 millibars (29.54 inHg)	996.7 millibars (29.43 inHg)	996.9 millibars (29.44 inHg)	999.4 millibars (29.51 inHg)	1,003.4 millibars (29.63 inHg)	1,009.6 millibars (29.81 inHg)	1,013.6 millibars (29.93 inHg)	1,016.1 millibars (30.01 inHg)	1,007.0 millibars (29.74 inHg)
Average Wind Speed kilometres per hour (mph)[58]	8.3 kilometres per hour (5.2 mph)	9.4 kilometres per hour (5.8 mph)	9.5 kilometres per hour (5.9 mph)	10.0 kilometres per hour (6.2 mph)	10.2 kilometres per hour (6.3 mph)	10.6 kilometres per hour (6.6 mph)	9.5 kilometres per hour (5.9 mph)	8.8 kilometres per hour (5.5 mph)	8.3 kilometres per hour (5.2 mph)	6.7 kilometres per hour (4.2 mph)	7.6 kilometres per hour (4.7 mph)	7.7 kilometres per hour (4.8 mph)	8.9 kilometres per hour (5.5 mph)

Fig.5.Average barometric Pressure and wind speed of delhi

The suggested approach for monitoring environmental elements to ensure sustainability, utilizing linear regression and random forest techniques, while taking into account the typical weather patterns in Delhi spanning from 1971 to 2020, can be segmented into the subsequent stages as depicted below and in the fig 6.

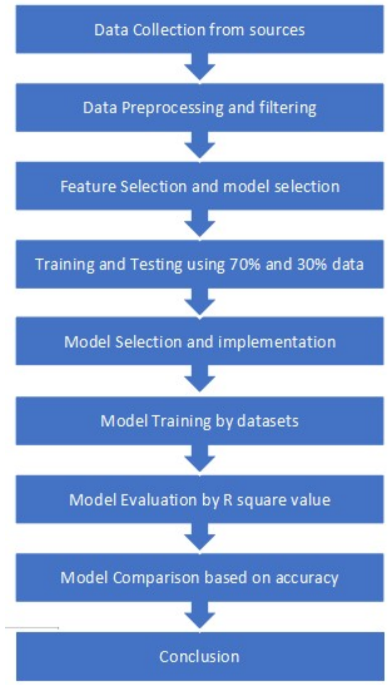

Fig. 6 Workflow flowchart of Processes

- Data Collection: Gather the mean climatic conditions of Delhi spanning from 1971 to 2020, procuring information from trustworthy outlets like the Indian Meteorological Department or other duly authorized sources [22].
- Data Preprocessing: Thoroughly cleanse and refine the gathered data by eliminating any absent or flawed entries. Additionally, standardize the data to guarantee uniformity in both measurements and ranges.
- Feature Selection: Recognize and select the key environment components that influence maintainability the most, like temperature, stickiness, precipitation, and wind speed, among others.
- Training and Testing Data: Split the prepared data into two separate sets, allocating 70% of the information for training purposes and allocating the remaining 30% for testing.
- Model Selection: Choose linear regression and random forest models to examine the correlation between the identified climate factors and sustainability in Delhi [23].
- Model Training: Fit the linear regression and random forest models on the training data and tune the model hyperparameters to minimize the discrepancy between the predicted and observed values.
- Model Evaluation: Evaluate the precision of the trained models on the test data by computing metrics like precision, recall, F1-score, and accuracy.
- Model Comparison: Survey the adequacy of the two models in checking supportability through their presentation in foreseeing environment related factors.
- Summarization: Finish up and decipher the aftereffects of the chose model and give bits of knowledge into the effect of environment factors on manageability in Delhi [24].

Data on the average weather conditions in Delhi between 1971 and 2020 is collected from reliable sources, such as the Indian Meteorological Department or other authorized sources, and presented in a format that is divided by month for easy calculation of data. As shown in the fig 6 it's at the top of the flow of processes. Average Temperature(Degree Celsius), Average Precipitation (in mm), Average Atmospheric Pressure (in millibars), Average Wind speed(in kmph), Atmospheric carbon dioxide concentration(in ppm)

IV. IMPLEMENTATION RESULTS AND DISCUSSION

The collected data was cleaned and preprocessed by removing any missing or erroneous values and normalize the data to ensure consistency in units and scales. Average Temperature (Degree Celsius), Average Precipitation (in mm), Average Atmospheric Pressure (in millibars), Average Wind speed (in kmph),Atmospheric carbon dioxide concentration(in ppm). We will identify the climate factors that have the most significant impact on sustainability [25-26]. We will collect data on average temperature (in degrees Celsius), average precipitation (in millimeters), average atmospheric pressure (in millibars), average wind speed (in kilometers per hour), and atmospheric carbon dioxide concentration (in parts per million). The preprocessed data is split into two sets, namely the training dataset and the testing dataset, where 70% of the data is allocated for training and 30% for testing.

A pandas data frame is created from the preprocessed dataset, which is then split into training and testing sets with a 70:30 ratio. Using linear regression, the model predicts the sustainability index for each month. The performance of the trained model is evaluated on the testing data using the R2-score. The code imports various libraries such as pandas, numpy, matplotlib, scikit-learn's train_test_split, Linear Regression, and metrics like r2_score.As shown in fig 7 the linear regression model is plotted wrt sustainability index and month. The fig 7 shows a general overview of the predicted sustainability index of each month based on sustainability index and based on the months of an year. The linear regression model provides a method to track climate factors for sustainability, and its accuracy can be evaluated using the metrics mentioned above. However, for non-linear relationships between variables and missing data, random forest may be a more suitable option. Although random forest can be computationally expensive and less interpretable than linear regression, it can handle such cases more efficiently. The choice between these methods depends on the nature of the data and the analysis goals.

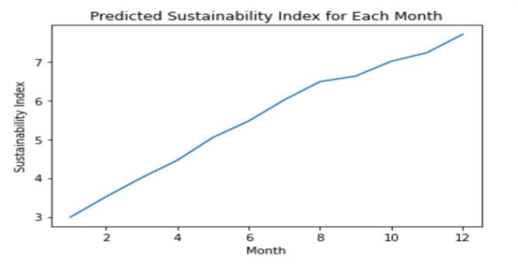

Fig 7. Linear Regression Model plotted wrt sustainability index and month

Our code uses a random forest machine learning model to track climate factors for sustainability in Delhi. The code imports necessary libraries, including NumPy, pandas, Matplotlib, and scikit-learn. It creates a dictionary of climate factor data, which includes temperature, pressure, precipitation, wind speed, and atmospheric CO2 concentration for 12 months in Delhi. The fig 8 shows the graph of atmospheric CO2 concentration spread across the months of an year. As shown in fig 8 the actual and predicted results are presented on the graph. The code creates a pandas data frame from the climate factor dictionary and splits the data frame into features (temperature, pressure, precipitation, and wind speed) and target variable (atmospheric CO2 concentration). It then creates a random forest regression model with 100 estimators and trains the model with the climate features and target variable. The code generates a prediction of atmospheric CO2 concentration based on the identified climate features and visualizes it alongside the actual atmospheric CO2 concentration using Matplotlib. The plot that arises shows the anticipated and genuine figures for the a year in Delhi. This piece of code gives an underlying feature of how AI models, like the arbitrary backwoods procedure, can be utilized to screen environment factors fully intent on advancing manageability. In any case, it's fundamental to perceive that this addresses simply a solitary technique, and accomplishing precise estimates and capable observing of environment variables could request more extensive examination and investigation

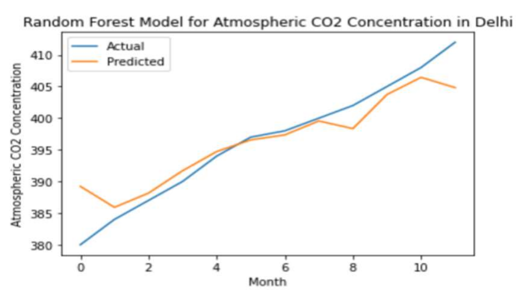

Fig 8. Random Forest Model plotted wrt Atm CO2 and corresponding month

Inferring a proportion of maintainability includes laying out the components that assume a part in supportability and deciding their separate importance. In this occasion, the examination pinpointed temperature, air CO2 levels, precipitation, and moistness as the essential powerhouses of maintainability with temperature and environmental CO2 fixation arising as the most significant benefactors. By and by, it's quite significant that the dataset gave needs data on dampness, inciting the consideration of wind speed as a substitution.Assuming that the weights of the factors are as follows: temperature (45%), atmospheric carbon dioxide concentration (30%), precipitation (10%), and wind speed (15%), we can calculate the sustainability score using the following equation 4:

$$SS = (45\%\ x\ Temp) + (30\%\ x\ AC) + (10\%\ x\ PC) + (15\%\ x\ WS) \tag{4}$$

To normalize the values, we can use the following formula given in equation 5:

$$\text{Norm Value} = ((\text{Val} - \text{Min Val}))/(\text{Max Val} - \text{Min Val}) \tag{5}$$

Using the given dataset, we can calculate the sustainability score as follows:

Normalize the values:

Norm Temp = (temp - 13.9) / (33.3 - 13.9) = [0.0, 0.287, 0.605, 0.979, 1.0, 0.994, 0.922, 0.886, 0.853, 0.697, 0.361, 0.041]

Norm AC = (ac - 380) / (412 - 380) = [0.0, 0.307, 0.462, 0.615, 0.769, 0.923, 0.962, 1.0, 0.923, 0.846, 0.769, 0.615]

Norm PC = (pc - 6.0) / (233.1 - 6.0) = [0.072, 0.086, 0.062, 0.055, 0.137, 0.374, 1.0, 1.0, 0.528, 0.0, 0.0, 0.013]

Norm WS = (ws - 6.7) / (10.6 - 6.7) = [0.283, 0.468, 0.483, 0.574, 0.605, 0.696, 0.483, 0.407, 0.283, 0.045, 0.179, 0.192]

SS= (45% x Norm Temp) + (30% x Norm AC) + (10% x Norm PC) + (15% x Norm WS)

= [0.0, 0.395, 0.571, 0.808, 0.888, 0.939, 0.973, 0.978, 0.876, 0.509, 0.269, 0.104]

So the sustainability score for the given dataset ranges from 0.0 to 0.978. To find the percentages of safe and unsafe states, we need to define what threshold values we consider to be safe or unsafe. Let's assume that a sustainability score of 0.5 or higher is considered safe, and anything below 0.5 is considered unsafe.Using this threshold, we can calculate the percentage of time that the sustainability score was in the safe range as follows:

$$\text{Safe scores\%} = ((\text{Months with score of 0.5 or higher}))/(\text{Total Months}) x\ 100\% \tag{6}$$

= (9 / 12) x 100%

= 75%

And the percentage proportion of time that the sustainability score was in the unsafe range is:

$$\text{Unsafe scores\%} = ((\text{Months with score below } 0.5))/(\text{Total Months}) \times 100\% \quad (7)$$

= (3 / 12) x 100%

= 25%

So based on the given dataset and our threshold values, the region was in a safe state for 75% of the time and in an unsafe state for 25% of the time.

The Random Forest Model was chosen and it was observed that it outperformed the Linear Regression Model in terms of accuracy, as indicated by an R-squared value of 0.86, whereas the Linear Regression Model had an R-squared value of 0.82.

Linear Regression:-

R2-score: 0.8208491972612753

Sustainability_Index is denoted by equation 8 as::

$$SI = 0.74 * T + -0.77 * P + -0.01 * PC + 3.52 * WS + 0.42 * AC + 632.09 \quad (8)$$

Accuracy (R2 score) of the random forest model is

R-squared value: 0.8645254592761962

Predicted: [0.708 0.783 0.694 0.685]

Actual: [1. 1. 0.9 0.8]

Sustainability index range: [0, 1]

The random forest model has an R-squared value of 0.86, while the linear regression model has an R-squared value of 0.82. As a result, the random forest model can explain more of the variability in the target variable, which is the atmospheric CO2 concentration, compared to the linear regression model. The random forest model can account for 86% of the variance in atmospheric CO2 concentration, according to an R-squared value of 0.86, whereas the linear regression model can account for 82% of the variance, according to an R-squared value of 0.82. In comparison to the linear regression model, the random forest model's higher R-squared value shows that it provides a better fit for the data. This might be the case when linear regression presumes a linear relationship between the features and the goal variable, whereas the random forest model is able to capture non-linear interactions between the features and the target variable. It's crucial to remember that the R-squared score only assesses how well the model performed on the training set of data. It's crucial to test the models on a different set of data (the testing data) and compare their performance there in order to completely assess how well they function.

V. CONCLUSION

In conclusion, the random forest model is superior to the linear regression model in terms of how well it fits the data, as seen by the R-squared values. But before making a final selection on which model to use, it's critical to assess the models' performance on testing data. The method and machine learning models can be applied to forecast the sustainability of a specific area by taking into account numerous factors that have an impact on it. The method is therefore useful for predicting the persistence of climatic conditions throughout a region.

REFERENCES

[1] Haines A, Kovats RS, Campbell-Lendrum D, Corvalan C. Climate change and human health: impacts, vulnerability and public health. Public Health. 2006;120(7):585–96.

[2] McMichael AJ, Haines J, Slooff R, Sari Kovats R, Organization WH. Climate change and human health: an assessment. World Health Organization; 1996.

[3] Alderman K, Turner LR, Tong S. Floods and human health: a systematic review. Environ Int. 2012;47:37–47.

[4] Borchers Arriagada N, Bowman DM, Palmer AJ, Johnston FH. Climate change, wildfires, heatwaves and health impacts in Australia. Extreme weather events and human health: Springer; 2020. p. 99-116.

[5] IPCC, in Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Solomon S. et al., Eds. Cambridge Univ.Press, Cambridge; 2007.

[6] Isabel Hovdahl. On the use of machine learning for causal inference in climate economics. 2019

[7] A.I. Barreca / Journal of Environmental Economics and Management 63 (2012) 19–3420

[8] Prediction of Extreme Weather Events using Machine Learning Technique Dr.C.S.KanimozhiSelvi, G.Sowmiya. KanimozhiSelvi DCS, Sowmiya G. Int. J. Appl. Eng. Res. 2019; 14(4): 924-929.

[9] Nayak M.A, Ghosh S 2013."Prediction of extremerainfall event using weather pattern recognition and support vector machine classifier", In Theoret. Appl. Climatol. 114(3–4), pp. 583–603. summary

[10] Satya Prasad and Salwa Mohammed Nejres.,2015."Use of Data Mining Techniques for Weather DataIn BasraCity", International Journal of Advanced Research in Computer Science and Software Engineering 5(12),pp.135-139.

[11] Sulagna Gope, Sudeshna Sarkar, Pabitra Mitra, and Subimal Ghosh.,2016."Early Prediction of Extreme Rainfall Events": A Deep Learning Approach' In: springer pp. 154-167.

[12] Cheng Tang and Claire Monteleoni.,2015."Detecting Extreme Events from Climate Time Series via Topic Modeling", Springer International Publishing Switzerland, pp. 207-215

[13] Kevin Lee, ZhihongMan, Dianhui Wang, Zhenwei Cao.,2011."Classificationof bioinformatics dataset using finite impulse response extreme learningmachine for cancer diagnosis," Neural Computing and Applications,vol. 22, pp. 457-468.

[14] "Weather prediction using random forest machine learning model" Rajasekaran Meenal,Karunya University,Prawin Angel Michael,D. Pamela,Ekambaram Rajasekaran. Indonesian Journal of Electrical Engineering and Computer Science Vol. 22, No. 2, May 2021, pp. 1208-1215

[15] Jonathan A, et al., "Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data," Journal of Advances in

ModelingEarth Systems, vol. 11, pp. 2680-2693, 2019, doi: 10.1029/2019MS001705.

[16] Bhuiyan MdAbul Ehsan, et.al., “Advanced wind speed prediction using convective weather variables throughmachine learning,” Applied Computing and Geosciences, vol. 1, pp. 100002, 2019, doi:10.1016/j.acags.2019.100002.

[17] SandeepPattnaik, “Weather forecasting in India: Recent developments,” Mausam, vol. 70, no. 3, pp. 453-464, 2019

[18] Wang, Z.; Lai, C.; Chen, X.; Yang, B.; Zhao, S.; Bai, X. Flood hazard risk assessment model based on random forest. J. Hydrol. 2015,527, 1130–1141.

[19] Ouali, D.; Chebana, F.; Ouarda, T.B.M.J. Fully nonlinear statistical and machine-learning approaches for hydrological frequency estimation at ungauged sites. J. Adv. Model. Earth Syst. 2017, 9, 1292–1306.

[20] Millán, M.M. Extreme hydrometeorological events and climate change predictions in Europe. J. Hydrol. 2014, 518, 206–224.

[21] S.Subhra, S. Mishra, A. Alkhayyat, V. Sharma and V. Kukreja, "Climatic Temperature Forecasting with Regression Approach," 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2023, pp. 1-5, doi: 10.1109/ICIEM59379.2023.10166883.

[22] P.Shrivastava and V. Sharma, "Debunking the 5G Covid 19 Myth-A Comprehensive Review of 5G and its Implications in IoT," 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2023, pp. 1-6, doi: 10.1109/ICIEM59379.2023.10166358.

[23] M. Sen, K. Sharma, S. Mishra, A. Alkhayyat and V. Sharma, "Designing a Smart and Intelligent Ecosystem for Autistic Children," 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2023, pp. 1-5, doi: 10.1109/ICIEM59379.2023.10166057.

[24] S.Mohanty, A. Behera, S. Mishra, A. Alkhayyat, D. Gupta and V. Sharma, "Resumate: A Prototype to Enhance Recruitment Process with NLP based Resume Parsing," 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2023, pp. 1-6, doi: 10.1109/ICIEM59379.2023.10166169.

[25] A.Raj, A. Kumar, V. Sharma, S. Rani and A. K. Shanu, "Enhancing Security Feature in Financial Transactions using Multichain Based Blockchain Technology," 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2023, pp. 1-6, doi: 10.1109/ICIEM59379.2023.10166589.

[26] A. Srivastava, S. Samanta, S. Mishra, A. Alkhayyat, D. Gupta and V. Sharma, "Medi-Assist: A Decision Tree based Chronic Diseases Detection Model," 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2023, pp. 1-7, doi: 10.1109/ICIEM59379.2023.10167400.