Climate Change Detection System

Saurabh Gopal[1], Aditya Landage[2], Mayur Gawade[3], Gyaneshwar Donakonda[4] , Prof. Deepa Athawale[5]

[1,2,3,4]B.E student Department of Computer Engineering Bharat College of Engineering, Badlapur
[5] Professor, Department of Computer Engineering, Bharat College of Engineering, Badlapur, Thane, Maharashtra - 421503

Abstract—
Climate change is one of the most pressing global challenges, with wide-ranging impacts on natural ecosystems, weather patterns, and human livelihoods. Timely and accurate detection of climate change signals is essential for informed decision-making and policy development. Traditional statistical approaches, while effective, often struggle with the complexity and scale of modern environmental datasets.This paper explores the application of **machine learning (ML)** techniques to climate change detection, with a specific focus on the **Gradient Descent algorithm** as a foundational optimization method. Machine learning enables the analysis of large, high-dimensional climate datasets, such as temperature records, CO_2 levels, and satellite imagery, uncovering hidden trends and anomalies that may not be evident through conventional methods. Gradient Descent plays a critical role in training predictive models by iteratively minimizing error, thereby improving the accuracy and reliability of climate forecasts and anomaly detection.Through case studies and experimental results, we demonstrate how ML models optimized via Gradient Descent can effectively identify climate change indicators and support early warning systems. The paper also discusses the challenges of applying machine learning to climate science, including data quality, model interpretability, and computational constraints. Overall, this research highlights the transformative potential of machine learning in advancing climate change detection and enhancing environmental decision-making.

Climate change detection is critical for understanding the evolving patterns of environmental transformations and their impacts on ecosystems, economies, and human health. This paper presents a comprehensive climate change detection system that leverages advanced data analytics, machine learning algorithms, and environmental datasets to identify and monitor climate-related changes effectively. The system integrates data from diverse sources, including satellite imagery, weather stations, ocean buoys, and historical climate models, to enhance the accuracy of trend analysis. The model's performance is evaluated through statistical metrics like score and mean squared error to ensure reliability.

Keywords: Machine Learning, linear regression, exponential regression, polynomial regression

1. INTRODUCTION

Earth's climate has changed throughout history. Most of these Climate change poses one of the most significant threats to ecosystems, economies, and human societies in the 21st century. Detecting and understanding its patterns, causes, and future trajectory is crucial for effective mitigation and adaptation strategies. Traditional climate modeling techniques, while powerful, often struggle with the sheer complexity and volume of environmental data.In recent years, **machine learning (ML)** has emerged as a transformative tool in climate science. Leveraging large datasets from satellite imagery, weather stations, ocean buoys, and climate models, ML algorithms can uncover hidden patterns, enhance predictive accuracy, and automate anomaly detection in ways that traditional statistical methods may not.Machine learning techniques such as **neural networks, support vector machines, decision trees**, and **unsupervised clustering algorithms** have been applied to a wide range of climate-related tasks. These include identifying trends in temperature and precipitation, detecting extreme weather events, predicting sea-level rise, and analyzing carbon emissions. Furthermore, ML can assist in downscaling global climate models to regional scales, enabling more actionable insights for local policymakers. Despite its promise, the

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

integration of ML in climate change detection presents challenges—such as data quality, model interpretability, and the need for interdisciplinary collaboration. However, as computational power and data availability continue to grow, machine learning is set to play an increasingly critical role in advancing our understanding and response to climate change.

Paper Structure

Section 1: Provides an overview of the training process, detailing the dataset used, pre-processing techniques applied, and the algorithms implemented.

Section 2: Discusses the evaluation of different algorithms.

Section 3: Presents the inference results and predictive values.

Section 4: Concludes the paper with key findings and suggests potential future research directions.

2. TRAINING

Training a machine learning model for climate change detection involves feeding it historical and real-time climate data to learn patterns, trends, and anomalies associated with environmental changes. This process is essential for enabling the model to make accurate predictions and detect early signs of climate-related phenomena such as global warming, extreme weather Agriculture is highly sensitive to climate variability, and cereal crops such as wheat, rice, and maize are particularly vulnerable to changes in temperature, precipitation, and extreme weather events. Detecting climate change impacts on cereal yield is crucial for ensuring global food security. Machine learning (ML) offers powerful methods to model and predict the relationship between climatic factors and crop productivity.

Parameter number	Parameter Name
1	Total Forest Area (in square kilometers)
2	Percentage of Agricultural Land Under Irrigation (of total agricultural land)
3	Cereal Production Rate (measured in kilograms per hectare)
4	Proportion of Population with Electricity Access (percentage)
5	Carbon Dioxide Emission Intensity (in kilograms per kilogram of oil equivalent energy)
6	Carbon Dioxide Emissions from Gaseous Fuels (in kilotons)
7	Total Carbon Dioxide Emissions (in kilotons)
8	Carbon Dioxide Emissions from Liquid Fuels (in kilotons)
9	Per Capita Carbon Dioxide Emissions (in metric tons)
10	Carbon Dioxide Emissions from Solid Fuels (in kilotons)
11	Aggregate Greenhouse Gas Emissions (in kilotons of CO_2 equivalent)
12	Hydrofluorocarbon (HFC) Emissions (in thousand metric tons of CO_2 equivalent)
13	Methane Emissions (in kilotons of CO_2 equivalent)
14	Nitrous Oxide Emissions (in thousand metric tons of CO_2 equivalent)
15	Annual Freshwater Withdrawal (in billion cubic meters)
16	Total Population
17	Urban Population

TABLE I: Parameter name and their corresponding number

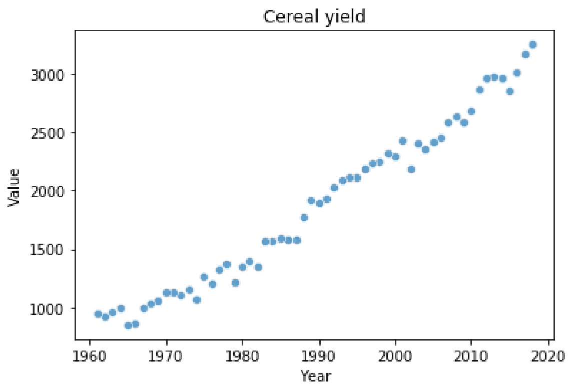


Fig. 1: Cereal yield (kilogram per hectare) per year

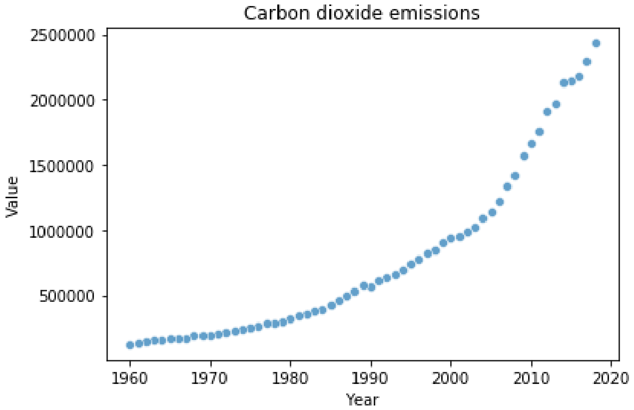
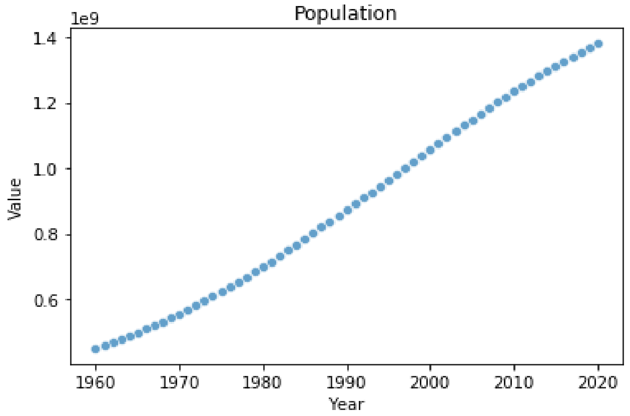


Fig. 2: Carbon dioxide emissions (kiloton) per year



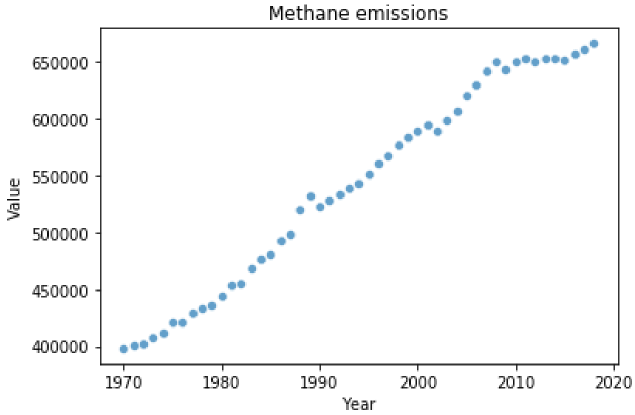
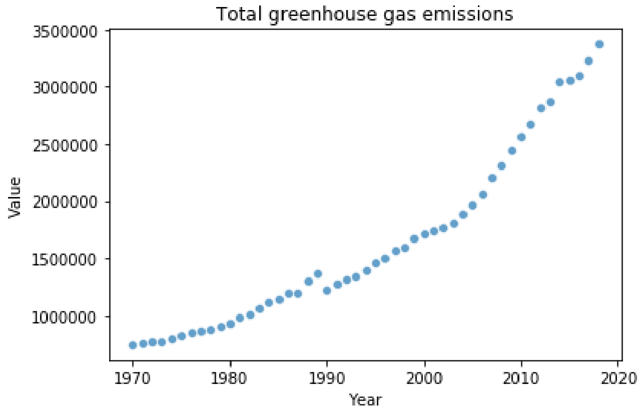


Fig. 3: Methane emissions (kiloton of carbon dioxide equivalent) per year



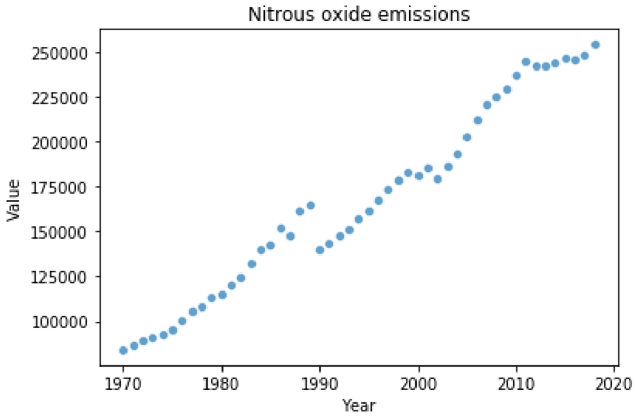


Fig. 4: Nitrous oxide emissions (thousand metric tons of CO2 equivalent) per year

A. Preprocessing

We have used Pandas library for pre-processing [5]. After dropping unwanted feature and selecting 17 important climate between 0 and 1 it is also known as Min – Max scaling . Here's the formula for normalization:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

In this context, XmaxX_{\text{max}}Xmax and XminX_{\text{min}}Xmin represent the highest and lowest values of the feature, respectively. When XXX corresponds to the minimum value in the dataset, the numerator becomes zero, resulting in a normalized value X′X'X′ of 0. Conversely, if XXX equals the maximum value, the numerator matches the denominator, yielding a normalized value of 1. For any value of XXX that lies between the minimum and maximum, the normalized output X′X'X′ will fall within the range of 0 to 1 Following data preprocessing, the dataset was analyzed using multiple regression techniques, including linear, polynomial, and exponential regression. The performance and outcomes of these models were then assessed to determine their effectiveness.

C. Algorithms

To detect climate change patterns and trends, we implemented three regression algorithms using Python's **Scikit-learn (sklearn)** library. These models help establish relationships between various climate parameters and track how they change over time.

1. Linear Regression

Linear regression is used to model the relationship between a single input feature (such as CO_2 emissions or temperature) and a target variable (like cereal yield or total greenhouse gas emissions). It fits a straight line to the data using the equation:

$$h(w_0, w_1) = w_0 + w_1 * X$$

2. polynomial Regression

Below is the hypothesis function for quadratic equation This model is more flexible than linear regression and can better fit non-linear patterns often found in climate data

$$h(w_0, w_1, w_2) = w_0 + w_1 * X + w_2 * X^2$$

3. Exponential Regression

Below is the hypothesis function for exponential function

$$h(A, B) = A * B^X$$

where

X: The input value

All adjustable parameters are optimized using the gradient descent algorithm. The hypothesis function is employed to generate predicted values based on the given inputs. These predicted values, compared with the actual target values, are used to compute the loss function. The gradient descent algorithm is then applied to minimize this loss function.

Algorithm: Gradient Descent

Input: Number of iterations N; n number of weights to be trained $w_0, w_1, \ldots w$; Loss function $J(w_0, w_1, \ldots w_n)$; Learning rate α

Output: Optimal weights which minimize the value of loss function $(w_0, w_1, \ldots w_n)$

1. **for** i = 1 to N **do**

Climate change detection involves analyzing vast and complex datasets, such as temperature records, atmospheric CO_2 levels, sea surface temperatures, and satellite imagery. Machine learning (ML) techniques provide powerful tools to extract patterns, make predictions, and detect anomalies in these environmental datasets. Among the core techniques used in ML, **Gradient Descent** plays a fundamental role in optimizing model performance.**Gradient Descent** is an optimization algorithm widely used in supervised learning models, such as linear regression, logistic regression, and neural networks. It helps these .

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

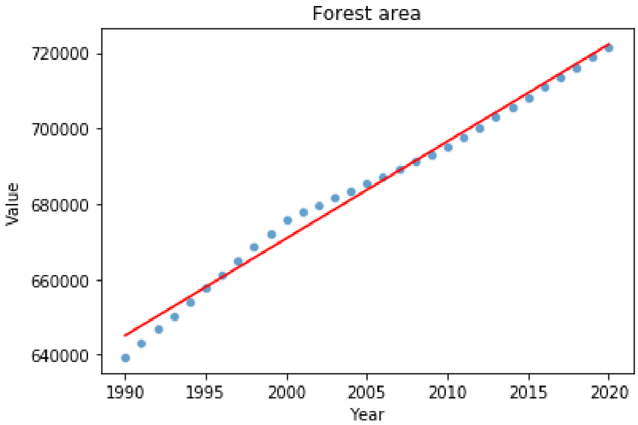


Fig. 7: Linear model for forest area

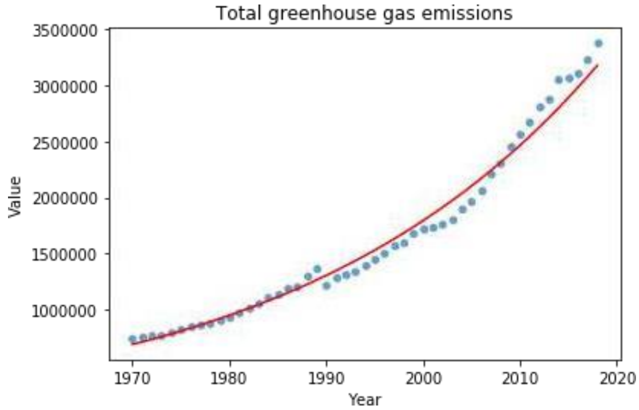


Fig. 10: Exponential model for total greenhouse emissions

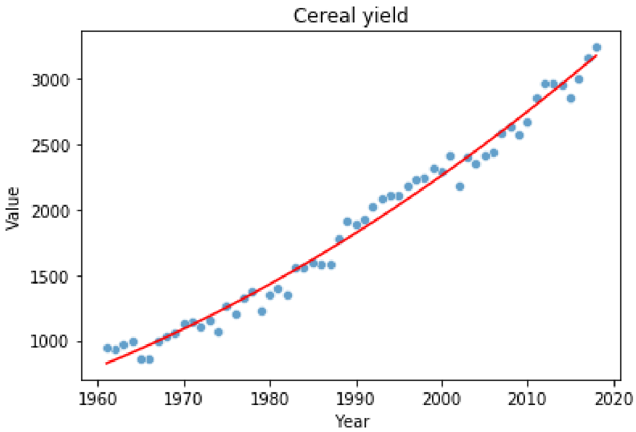


Fig. 8: Polynomial model for cereal yield

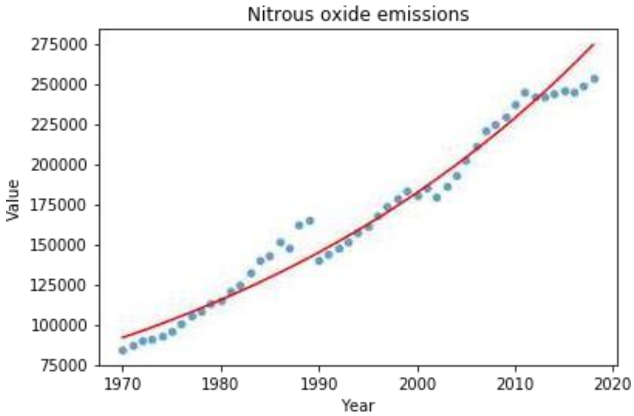


Fig. 11: Exponential model for nitrous oxide emissions

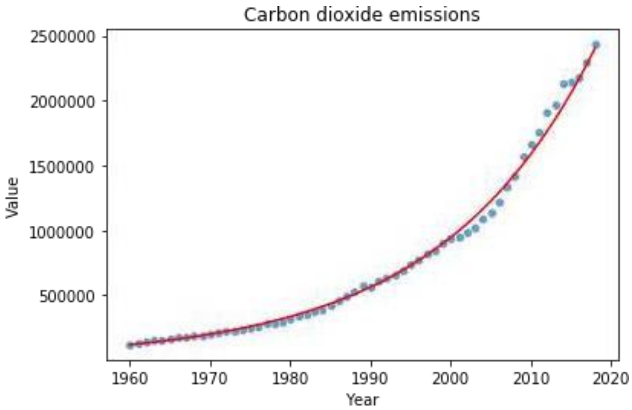


Fig. 9: Exponential model for carbon dioxide emissions

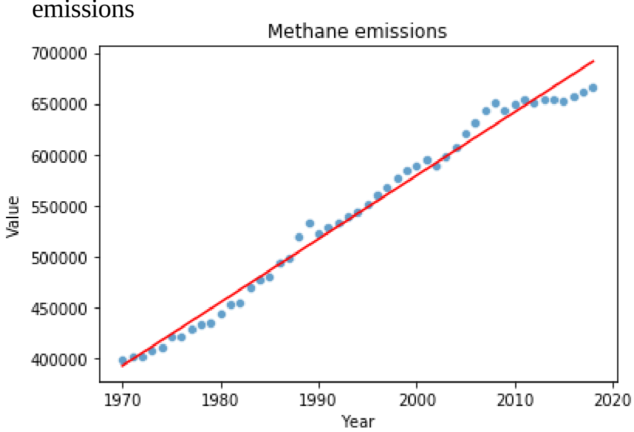


Fig. 12: Linear model for methane emissions

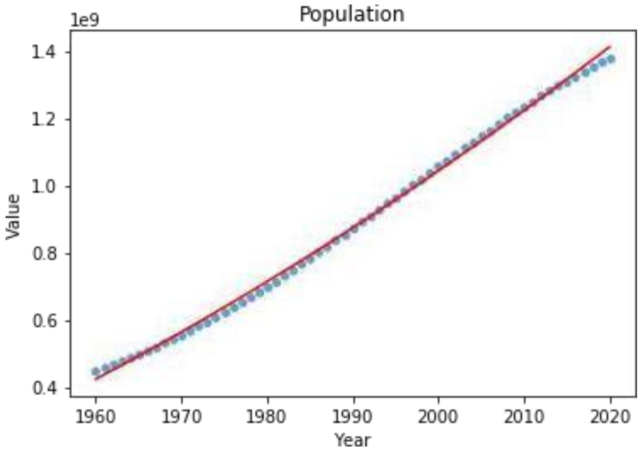


Fig. 13: Polynomial model for population

3. Evaluation

To assess the performance of the machine learning models used for climate change detection, we evaluated them using standard regression metrics. These metrics help determine In simpler terms, R2R^2R2 reflects how effectively the regression model fits the data, indicating the model's goodness of fit. The formula to calculate the R2R^2R2 score is as follows:

1. Root Mean Squared Error (RMSE)

The root-mean-square error (RMSE) is a frequently used measure of the differences between values predicted by a model and the values observed. The formula of root-mean- square error is as follows: how well the models fit the data and how accurately they can predict future values.

N: Total number of samples

We select the model among linear, polynomial and exponential which gives the best r-squared score and least root mean square error.

The R-squared score and root-mean-square error of the models are shown in table 2.

Parameter number	R-squared score	Root-mean-square error	Model Type
1	0.988	2532.218	Linear
2	0.912	0.657	Linear
3	0.984	87.071	Polynomial
4	0.973	2.298	Polynomial
5	0.986	0.0398	Linear
6	0.963	6159.992	Polynomial
7	0.997	48049.006	Exponential
8	0.986	28763.495	Exponential
9	0.986	0.065	Exponential
10	0.990	45896.375	Exponential
11	0.988	97892.936	Exponential
12	0.978	641.739	Linear
13	0.974	15899.680	Linear
14	0.975	8187.562	Exponential
15	0.861	42.361	Exponential
16	0.998	12983761.946	Polynomial
17	0.999	1578542.742	Polynomial

TABLE II: R-squared score and root-mean-square error of the models

4. Inference

From our analysis, it is evident that **machine learning algorithms can effectively detect patterns and trends in climate-related data**. By applying linear, polynomial, and exponential regression models to various climate parameters, such as greenhouse gas emissions, population growth, and cereal crop yields, we were able to observe significant correlations and long-term changes over time.

Among the models used:

Linear regression worked well for datasets showing steady, straight-line trends.

Polynomial regression captured more complex, non-linear relationships in the data, such as accelerating emissions or curving yield patterns.

Exponential regression was especially effective in modeling rapid growth trends, such as population increase or atmospheric CO_2 rise.

All models were trained using **Gradient Descent**, allowing them to learn from historical data and make accurate predictions. This demonstrates that machine learning is not only useful for modeling current climate behavior but also for forecasting future impacts and identifying critical thresholds.Overall, our findings support the idea that **machine learning can be a powerful tool in climate change detection and impact analysis**, providing valuable insights for researchers, policymakers, and environmental planners.

Parameter number	2025	2030	2035
1	735,124.03	747,984.96	760,845.89
2	42.80	45.25	53.09
3	3,580.98	3,882.18	4,195.81
4	115.02	132.03	150.93
5	2.81	2.95	3.09
6	147,058.58	174,585.04	204,456.59
7	3,454,882.16	4,477,007.86	5,801,529.11
8	1,099,238.78	1,467,701.88	1,959,673.24
9	2.05	2.41	2.83
10	2,282,348.27	2,932,672.65	3,768,298.20
11	3,965,426.38	4,647,366.03	5,446,579.74
12	22,812.13	25,822.29	28,832.44
13	769,936.74	816,603.17	866,098.08
14	279,436.02	297,558.13	315,680.25
15	800.52	853.40	909.78
16	1.509×10^9	1.610×10^9	1.713×10^9
17	5.357×10^8	5.936×10^8	6.548×10^8

TABLE III: Predicted values

5. CONCLUSION

Weather forecasting using the linear regression algorithm and the Naïve Bayes algorithm is critical for improving people's future results. The linear regression algorithm and the Naïve Bayes algorithm were used to forecast the weather using weather datasets. Using some selected input variables obtained from kaggle, GitHub we created a model to predict the weather. The issue with the current weather situation is that we are unable to organize ourselves and complete essential tasks.

As a result, this model was developed in order to know the weather scenario with high precision while taking into account all of the factors that influence the weather scenario.

REFERENCE

□*Weather: Forecasting from the Beginning"*. *InfoPlease. Retrieved January 14,*

2020

□University of California Museum of Paleontology. "Aristotle (384322 B.C.E.)

Archived November 20, 2016, at the Wayback Machine". Retrieved January 12,

2008.

□**David Pingree (2017, December 14).** *The Indian and Pseudo-Indian*

References in Greek and Latin Astronomical and Astrological Works (PDF).

□*pp. 141–195 [143–4]. Retrieved March 1, 2010.*[*permanent dead link*]

□ *Bible Gateway passage: Matthew 16:2-3 - English Standard Version"*. *Bible*

Gateway. Retrieved December 1, 2016.