A STUDY ON VARIABLE SELECTIONS AND PREDICTION FOR CLIMATE CHANGE DATASET USING DATA MINING WITH MACHINE LEARNING APPROACHES

S. Ravishankar[1], Dr. P. Rajesh[2]

Article History: Received: 17.09.2022 **Revised:** 02.11.2022 **Accepted:** 18.12.2022

Abstract

Climate change is an intricate and pressing worldwide concern, demanding a synchronized endeavor across local, national, and international domains to diminish its repercussions and shift toward a more sustainable future. It presents substantial obstacles to ecosystems, economies, and human communities, thus ranking as one of the paramount issues of our era. Machine learning serves as a comprehensive concept encompassing the solution of problems that would be financially impractical to address through the creation of algorithms by human programmers. Instead, it empowers machines to uncover their algorithms autonomously, obviating the need for explicit guidance from human-developed algorithms. This paper considers climate change-related datasets like year, month, MEI, CO_2, CH_4, N_2O, CFC-11, CFC-12, TSI, aerosols, and temp. The machine learning approaches which is used to analyze and predict the dataset using linear regression, multilayer perceptron, SMOreg, M5P, random forest, random tree, and REP tree**.** Numerical illustrations are provided to prove the proposed results with test statistics or accuracy parameters.

Keywords: Machine learning, climate change, decision tree, correlation coefficient, and test statistics.

[1]Research Scholar, Department of Computer and Information Science, Annamalai University, Annamalainagar, Tamil Nadu, India
Email: thiru.ravishankar@gmail.com

[2]Assistant Professor, PG Department of Computer Science, Government Arts College, Chidambaram, (Deputed from Department of Computer and Information Science, Annamalai University, Annamalainagar) Tamil Nadu, India
Email: rajeshdatamining@gmail.com

DOI: 10.48047/ecb/2022.11.12.163

1. Introduction and Literature Review

Integrating machine learning and data mining into climate change research is pivotal for deepening our comprehension of the intricate mechanisms propelling climate change. It also plays a crucial role in formulating strategies to alleviate its consequences and adapt to a dynamic world. This approach empowers researchers to base their decisions on data and inform policies and mitigation endeavors. Integrating machine learning and data mining into climate change research is pivotal for deepening our comprehension of the intricate mechanisms propelling climate change. It also plays a crucial role in formulating strategies to alleviate its consequences and adapt to a dynamic world. This approach empowers researchers to base their decisions on data and inform policies and mitigation endeavors. Machine Learning entails the process of instructing a computer program to construct a statistical model from data. The primary aim of machine learning (ML) is to extract meaningful patterns and gain valuable insights from the given data.

An evaluation of downscaling is conducted to assess its intended purpose and value in providing information about potential climate change impacts and supporting regional risk assessments. The concepts of climate realism and the physical plausibility of change are introduced to qualify the general characteristics associated with various downscaling approaches. Climate realism assesses the effectiveness of different approaches in representing regional climate characteristics, while the physical plausibility of change evaluates their ability to simulate regional climate variations. Downscaling may capture aspects of change not accounted for by global climate models, particularly due to resolution or regional factors. If these aspects are of interest, downscaling can be valuable once its worth has been demonstrated. For situations where the primary concern is broad-scale changes in mean climate or where there's no demonstrated added value from downscaling, a wide range of regionalization methods are available for practitioners in the impact, adaptation, and vulnerability field [1].

The literature exploring the intersection of gender relations and climate change contributes significantly to our understanding of several key aspects. It sheds light on vulnerability and the impacts of climate change, adaptations in diverse contexts, the responsibility for greenhouse gas emissions, inequalities in climate governance, and the role of knowledge and social action in addressing climate change. Overall, this body of work establishes that gender relations are an integral component of the social transformations associated with climate change. Ignoring gender analysis in climate research overlooks crucial dimensions of social life in a changing climate. It is imperative to acknowledge the gendered nature of climate change and delve deeper into this aspect within the social sciences and humanities [2].

Consolidate research on the impact of climate change on sugarcane production and assess the implications for the sugarcane industry, along with potential response strategies. While existing research predominantly examines how climate changes affect primary production, limited attention has been given to the industry-wide consequences and economic outcomes across the sugarcane value chain. Among the 90 studies reviewed, 61 investigate observed and projected climate change impacts on sugarcane production, with varying conclusions based on methods, time frames, and geographic regions. Seventeen studies focus on adapting to climate change impacts, but evidence of successful adaptation remains limited. Additionally, some papers discuss mitigating energy use and greenhouse gas

emissions in sugarcane production. The review calls for further research, especially regarding the economic implications of climate change for the sugarcane industry [3].

Detecting and attributing climate change has proven unexpectedly challenging in the 21st century. Despite increasing greenhouse gas concentrations, the global surface temperature rose less rapidly from 2000 to 2015 compared to the latter half of the 20th century. This discrepancy is likely explained by the mitigating factors of La Niña cooling and declining solar irradiance. Physical climate models overestimated recent global warming because they failed to reproduce the observed La Niña cooling and possibly underestimated the impact of declining solar irradiance. Ongoing investigations are seeking alternative explanations for the divergence between simulated and observed climate change, known as the "global warming hiatus." This situation, exacerbated by media commentary, confounds policymakers, the public, and other stakeholders regarding the causes and reality of modern climate change. Understanding and communicating the causes of climate change in the coming 20 years may be equally challenging, as predictions about the modulation of anthropogenic warming by natural processes lack precision. Emerging feedbacks and tipping points, such as melting Arctic sea ice, could further impact Earth's global temperature, beyond the capabilities of current sophisticated climate models [4].

Studies aims to (a) provide a long-term perspective on the current variability and projected future climate changes in extratropical cyclone characteristics, (b) establish connections to natural external forces, and (c) deepen our understanding of cyclone intensification processes in past climate periods. The review summarizes the current knowledge regarding two historical periods: the last millennium and the Last Glacial Maximum (LGM), which occurred 21,000 years ago. These historical periods are compared to climate modeling results on global and regional scales, using limited data from paleo proxy archives. For example, significant changes in the climate mean state due to orbital forcing and associated feedbacks are shown to impact various cyclone characteristics, including a southward shift of storm tracks over the North Atlantic during the LGM. Other findings suggest that dynamic processes may play an equally important role as thermodynamic processes in influencing cyclone-induced precipitation variations, contrasting with future changes primarily driven by thermodynamic processes. This review highlights how a paleoclimatic perspective enhances our understanding of extratropical cyclones and their characteristics, offering insights into future changes in these systems [5].

Data mining is a powerful tool for uncovering hidden information within large existing databases, generating valuable insights. In this paper, we use a dataset where each row represents specific days, with attributes describing weather conditions on those days and a class indicating whether the conditions are suitable for playing golf. The attributes include Outlook, Temperature, Humidity, Windy, and a Boolean Play Golf class. All the data are used for training purposes, and seven classification algorithms, including J48, Random Tree (RT), Decision Stump (DS), Logistic Model Tree (LMT), Hoeffding Tree (HT), Reduce Error Pruning (REP), and Random Forest (RF), are employed to measure accuracy. Among these algorithms, Random Tree outperforms the others, achieving an accuracy of 85.714% [6].

Employs citation network analysis and a text mining approach to identify research trends and extract common research topics and emerging domains in urban climatology. The analysis reveals that aerosols and ozone, as well as the urban

heat island effect, are the most prominent topics. Additionally, four emerging fields are identified: secondary organic aerosols, urban precipitation, flood risk and adaptation, and greenhouse gas emissions. The study also examines the geographical information of city case studies to explore spatial-temporal patterns, particularly in the emerging topical areas. The growth of interdisciplinary research in urban climatology, as it interacts with urban hydrology, health, energy, and social sciences, is noted. The study highlights several knowledge gaps, including the need for long-term, high-temporal-resolution observational data for model validation, predictions of urban effects on precipitation and extreme flooding events under climate change, and the development of a framework for collaboration between physical sciences and social sciences in urban settings [7].

The stochastic agro-economic model GLOBIOM is employed to demonstrate the most effective design and evaluation of the CAP's financial and structural measures, both individually and in combination, considering inherent uncertainty and risk. This model simultaneously accounts for plausible shocks and derives robust measures against all potential shock scenarios. It can help prevent irreversible and sunk costs in unexpected scenarios. The distribution of CAP funds needs to consider exposure to risks, security targets, and the synergies between policy measures, including production, trade, storage, and irrigation technologies, to ensure adequate agricultural production [8]. Presents evidence of stochastic effects in laboratory experiments and provides an overview of stochastic climate theory from an applied mathematics perspective. The study also surveys the current use of stochastic methods in comprehensive weather and climate prediction models, suggesting that stochastic parameterizations have the potential to address existing biases in these models [9].

Data mining is a valuable technique for discovering concealed information using the stochastic sensing concept. This paper proposes an efficient assessment of groundwater levels, rainfall, population, food grains, and enterprises datasets by employing stochastic modeling and data mining approaches. The paper introduces a novel data assimilation analysis to effectively predict groundwater levels. Experimental results demonstrate the robustness of this approach in estimating expected groundwater levels [10] and [11]. The dataset for chronic disease data consists of specific locations as rows, with attributes such as topics, questions, data values, low confidence limits, and high confidence limits. All the data is utilized for both training and testing using five classification algorithms. The paper presents various analyses and evaluates the accuracy of five different decision tree algorithms, with the M5P decision tree approach emerging as the most effective algorithm for building models compared to other decision tree methods [12].

2. Backgrounds and Methodologies

A data mining decision tree is a widely used machine learning technique for classification and regression tasks. It visually depicts a sequence of decisions and their possible outcomes in a tree-like structure. Each internal node represents a decision based on a specific feature, and each branch corresponds to the potential result of that decision. The tree's leaf nodes represent the final decision or the predicted outcome. The "CART" (Classification and Regression Trees) algorithm is the most used algorithm for building decision trees [13].

2.1 Linear Regression

Linear regression is a statistical technique employed to comprehend and

forecast the connection between two variables by discovering the optimal straight line that most effectively aligns with the data points. It aids in ascertaining how alterations in one variable correspond to changes in another, proving valuable for predictions and trend recognition.

The core idea of linear regression is to find the best-fitting straight line (also called the "regression line") through a scatterplot of data points. This line represents a linear equation of the form:

$$y = m_x + b$$

Where:

- y is the dependent variable (the one you want to predict or explain).
- x is the independent variable (the one you're using to make predictions or explanations).
- m is the slope of the line, representing how much
- y changes for a unit change in x.

b is the y-intercept, indicating the value of y when x is 0.

2.2 Multilayer Perception

A Multilayer Perceptron (MLP) is an artificial neural network consisting of multiple layers of interconnected nodes or neurons. It's a fundamental architecture in deep learning and is used for various tasks, including classification, regression, and more complex tasks like image recognition and natural language processing. The architecture of an MLP typically includes three types of layers:

i. **Input Layer:** This layer consists of neurons receiving input data. Each neuron corresponds to a feature in the input data, and the values of these neurons pass through the network.

ii. **Hidden Layers:** These layers come after the input layer and precede the output layer. They are called "hidden" because their activations are not directly observed in the final output.

iii. **Output Layer:** This layer produces the network's final output. The number of neurons in the output layer depends on the problem type.

2.3 SMO

SMO stands for "Sequential Minimal Optimization," an algorithm used for training support vector machines (SVMs), machine learning models commonly used for classification and regression tasks. The SMO algorithm is particularly well-suited for solving the quadratic programming optimization problem that arises during the training of SVMs.

Step 1. **Initialization:** Start with all the data points as potential support vectors and initialize the weights and bias of the SVM.

Step 2. **Selection of Two Lagrange Multipliers:** In each iteration, the SMO algorithm selects two Lagrange multipliers (associated with the support vectors) to optimize.

Step 3. **Optimize the Pair of Lagrange Multipliers:** Fix all the Lagrange multipliers except the selected two, and then optimize the pair chosen to satisfy certain constraints while maximizing a specific objective function.

Step 4. **Update the Model:** After optimizing the selected pair of Lagrange multipliers, update the SVM model's weights and bias based on the new values of the Lagrange multipliers.

Step 5. **Convergence Checking:** Check for convergence criteria to determine whether the algorithm should terminate.

Step 6. **Repeat:** If convergence hasn't been reached, repeat steps 2 to 5 until it is.

2.4 M5P

M5P is a machine learning algorithm used for regression tasks. It is an extension of the decision tree-based model called M5, which Ross Quinlan developed. The M5 algorithm combines decision trees and linear regression to create more accurate and flexible regression models. M5P, specifically, stands for M5 Prime. It enhances the original M5 algorithm to improve its predictive performance. M5P uses a tree-based model to divide the data into subsets based on feature values recursively and then fits linear regression models to each of these subsets. The result is a piecewise linear regression model, where different linear regressions are used for other regions of the input feature space.

Steps involved in the M5P

Step 1. Building the initial decision tree (M5 model): Recursive Binary Splitting and Pruning (optional)
Step 2. Linear Regression Model: Leaf Regression Models and Model Parameters
Step 3. Piecewise Linear Regression: Piecewise Prediction
Step 4. Model Evaluation: Training and Testing.

2.5 Random Forest

Random Forest is a popular machine learning ensemble method for classification and regression tasks. It is an extension of decision trees and is known for its high accuracy, robustness, and ability to handle complex datasets. Random Forest is widely used in various domains, including data science, machine learning, and pattern recognition. The main idea behind Random Forest is to create an ensemble (a collection) of decision trees and combine their predictions to make more accurate and stable predictions. The following steps describe what Random Forest works like.

- Bootstrap Aggregating (Bagging)
- Decision Tree Construction
- Voting for Classification, Averaging for Regression

The key advantages of Random Forest are:

- Reduced overfitting
- Robustness
- Feature Importance

Steps involved in Random Forest

Random Forest is an ensemble learning method combining multiple decision trees to make more accurate and robust predictions for classification and regression tasks. The steps involved in building a Random Forest are as follows:

Step 1. Data Bootstrapping
Step 2. Random Feature Subset Selection
Step 3. Decision Tree Construction
Step 4. Ensemble of Decision Trees
Step 5. Out-of-Bag (OOB) Evaluation
Step 6. Hyperparameter Tuning (optional)

2.6 Random Tree

In machine learning, a Random Tree is a specific type of decision tree variant that introduces randomness during construction. Random Trees are similar to traditional decision trees but differ in how they select the splitting features and thresholds at each node. The primary goal of introducing randomness is to create a more diverse set of decision trees, which can help

reduce overfitting and improve the model's generalization performance. Random Trees are commonly used as building blocks in ensemble methods like Random Forests. The critical characteristics of Random Trees are as follows:

- ❖ Random Feature Subset
- ❖ Random Threshold Selection
- ❖ No Pruning
- ❖ Ensemble Methods

Steps involved in Random Tree

Step 1. Data Bootstrapping:
Step 2. Random Subset Selection for Features:
Step 3. Decision Tree Construction:
Step 4. Voting (Classification) or Averaging (Regression):

2.7 REP Tree

REP (Repeated Incremental Pruning to Produce Error Reduction) Tree is a machine learning algorithm for classification and regression tasks. A decision tree-based algorithm constructs a decision tree using a combination of incremental pruning and error-reduction techniques. The key steps involved in building a REP Tree are as follows:

- ❖ Recursive Binary Splitting
- ❖ Pruning
- ❖ Repeated Pruning and Error Reduction

Steps involved in REP Tree

REP Tree (Repeated Incremental Pruning to Produce an Error Reduction Tree) is a machine learning algorithm for classification and regression tasks. It is an extension of decision trees that incorporates pruning to reduce overfitting and improve the model's generalization performance. Below are the steps involved in building a REP Tree.

Step 1. Recursive Binary Splitting
Step 2. Pruning
Step 3. Repeated Pruning and Error Reduction
Step 4. Model Evaluation

2.8 Accuracy Metrics

The predictive model's error rate can be evaluated by applying several accuracy metrics in machine learning and statistics. The basic concept of accuracy evaluation in regression analysis is comparing the original target with the predicted one and using metrics like R-squared, MAE, MSE, and RMSE to explain the errors and predictive ability of the model [14]. The R-squared, MSE, MAE, and RMSE are metrics used to evaluate the prediction error rates and model performance in analysis and predictions [15] and [16].

R-squared (Coefficient of determination) represents the coefficient of how well the values fit compared to the original values. The values from 0 to 1 are interpreted as percentages. The higher the value is, the better the model is.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \quad \ldots (1)$$

MAE (Mean absolute error) represents the difference between the original and predicted values extracted by averaging the absolute difference over the data set.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}| \quad \ldots (2)$$

RMSE (Root Mean Squared Error) is the error rate by the square root of MSE.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \quad \ldots (3)$$

Relative Absolute Error (RAE) is a metric used in statistics and data analysis to measure the accuracy of a forecasting or predictive model's predictions. It is particularly useful when dealing with numerical data, such as in regression analysis or time series forecasting.

$$\mathrm{RAE} = \frac{\sum|y_i - \hat{y}_i|}{\sum|y_i - \bar{y}|} \quad \ldots (4)$$

Root Relative Squared Error (RRSE) is another metric used in statistics and data analysis to evaluate the accuracy of predictive models, especially in the context of regression analysis or time series forecasting.

$$\mathrm{RRSE} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}} \quad \ldots (5)$$

Equations 1 to 5 are used to find the model accuracy, which is used to find the model performance and error. Where Y_i represents the individual observed (actual) values, $\hat{Y}_i$ represents the corresponding individual predicted values, $\bar{Y}$ represents the mean (average) of the observed values and Σ represents the summation symbol, indicating that you should sum the absolute differences for all data points.

3. Numerical Illustrations

The corresponding dataset was collected from the open source Kaggle data repository. The climate change dataset include 11 parameters which have different categories of data like Year, Month, MEI, CO2, CH4, N2O, CFC-11, CFC-12, TSI, Aerosols, and Temp [17]. A detailed description of the parameters is mentioned in the following Table 1.

Table 1. Climate change sample dataset

Year	Month	MEI	CO2	CH4	N2O	CFC-11	CFC-12	TSI	Aerosols	Temp
1983	5	2.556	345.96	1638.59	303.677	191.324	350.113	1366.102	0.0863	0.109
1983	6	2.167	345.52	1633.71	303.746	192.057	351.848	1366.121	0.0794	0.118
1983	7	1.741	344.15	1633.22	303.795	192.818	353.725	1366.285	0.0731	0.137
1983	8	1.13	342.25	1631.35	303.839	193.602	355.633	1366.42	0.0673	0.176
1983	9	0.428	340.17	1648.4	303.901	194.392	357.465	1366.234	0.0619	0.149
...	...	...	...	...	...	...	...	...	...	...
2008	8	-0.266	384.15	1779.88	321.405	244.2	535.072	1365.657	0.0036	0.407
2008	9	-0.643	383.09	1795.08	321.529	244.083	535.048	1365.6647	0.0043	0.378
2008	10	-0.78	382.99	1814.18	321.796	244.08	534.927	1365.6759	0.0046	0.44
2008	11	-0.621	384.13	1812.37	322.013	244.225	534.906	1365.7065	0.0048	0.394

2008	12	-0.666	385.56	1812.88	322.182	244.204	535.005	1365.6926	0.0046	0.33

Table 2: Machine Learning Models with R2 Score

ML Approaches	**R2 Score**
Linear Regression	0.8553
Multilayer Perceptron	0.8102
SMOreg	0.8462
M5P	0.8637
Random Forest	0.9145
Random Tree	0.8551
REP Tree	0.8526

Table 3: Machine Learning Models with Mean Absolute Error and Root Mean Squared Error

ML Approaches	**MAE**	**RMSE**
Linear Regression	0.0721	0.0926
Multilayer Perceptron	0.0836	0.1065
SMOreg	0.0746	0.0954
M5P	0.0683	0.0901
Random Forest	0.0556	0.0724
Random Tree	0.0739	0.0979
REP Tree	0.0730	0.0944

Table 4: Machine Learning Models with Relative Absolute Error (%) and Root Relative Squared Error (%)

ML Approaches	**RAE (%)**	**RRSE (%)**
Linear Regression	48.2148	51.7449
Multilayer Perceptron	55.9342	59.4954
SMOreg	49.9501	53.2595
M5P	45.7068	50.3329
Random Forest	37.2348	40.4369
Random Tree	49.4283	54.6909
REP Tree	48.8539	52.7040

Table 5: Machine Learning Models with Time Taken to Build Model (Seconds)

ML Approaches	**Time taken (seconds)**
Linear Regression	0.0100
Multilayer Perceptron	0.2800
SMOreg	0.1300
M5P	0.0900
Random Forest	0.3300
Random Tree	0.0100
REP Tree	0.0300

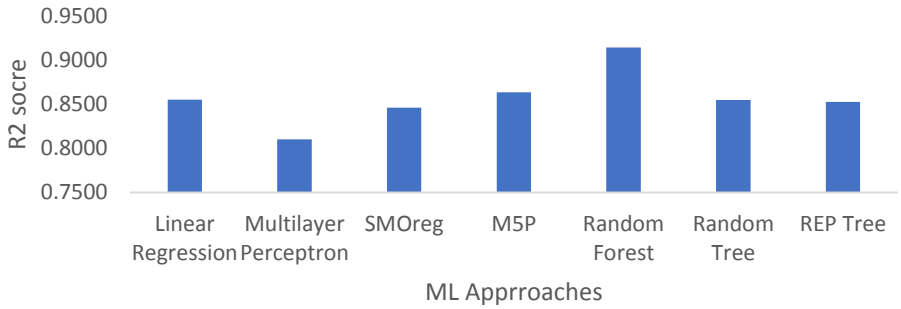

Fig. 1. R2 Score for Machine Learning Approaches

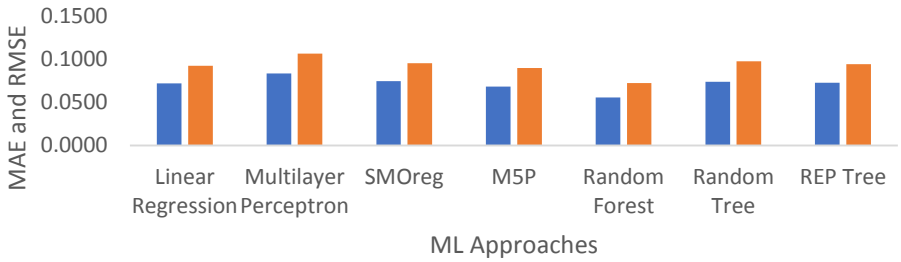

Fig. 2. Machine Learning Models with MAE and RMSE

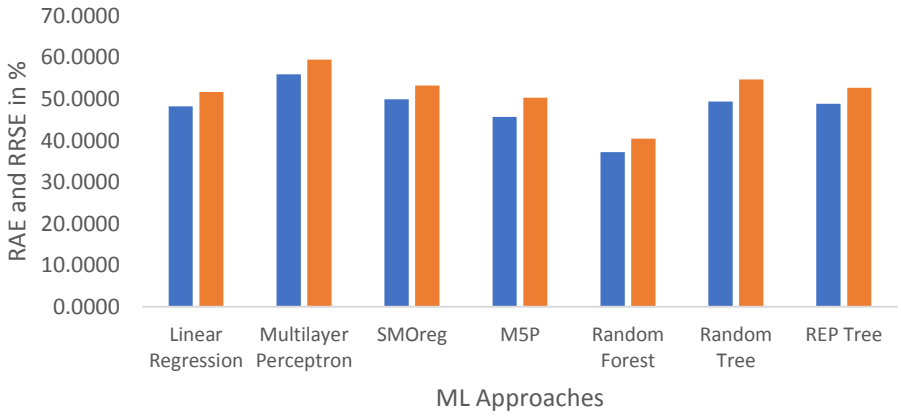

Fig. 3. Machine Learning Models with RAE (%) and RRSE (%)

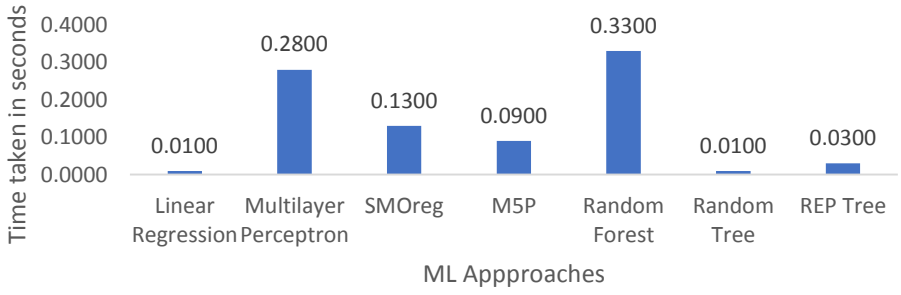

Fig. 4. Machine Learning Models and its Time Taken to Build the Model (Seconds)

4. Result and Discussion

Table 1 explains 11 parameters which include different categories of data like Year, Month, MEI, CO2, CH4, N2O, CFC-11, CFC-12, TSI, Aerosols, and Temp. Based on the dataset, it is evident that five different machine learning decision tree approaches are used to find the hidden patterns and which is the best or influencing parameter to decide future predictions. Related results and numerical illustrations are shown between Table 1 to Table 5 and Figure 1 to Figure 4.

They are based on Equation 1, Table 2, and Figure 1, which is used to find the R2 score or correlation coefficient by comparing 9 primary parameters. Numerical illustrations suggest that there may be a significant difference from one parameter to another. In this case, using seven different decision tree approaches among these results, Linear Regression, Multilayer Perceptron, SMOreg, M5P, Random Forest, Random Tree, and REP Tree, return a robust, strong positive correlation using various climate change-related parameters.

Further data analysis revealed a gradual improvement in test scores over time. The MAE is used to find model errors using Equations 2. Seven machine-learning algorithms will be used; all seven approaches return a minimum error for using MAE test statistics. The RMSE (root mean square error) measures the difference between predicted and actual values using Equation 3. In this case, also return a minimum error for using RMSE equation 3. The related numerical illustration is shown in Table 3 and Figure 2.

Relative Absolute Error (RAE) measures accuracy using equation 4 to compare the difference between predicted and actual values in percentage. In this research, considering seven ML classification algorithms, except Multilayer Perceptron, the remaining algorithms return a minimum error. Similar error approaches are reflected in RRSE. Similar numerical illustrations are shown in Table 4 and Figure 3.

Time taken is one of the significant tasks in machine-learning approaches. Based on Table 5 and Figure 4, except random forest and multilayer perceptron others take minimum time to build the model. Similar approaches are reflected in Table 5 and Figure 4. mentioned visualization.

5. Conclusion and further research

The findings presented in this study contribute to our understanding that all the parameters return robust positive correlations. In this research, the maximum of the machine learning approaches returns a minimum error with less processing time. Future studies can build upon these, finding the suitable variable for future prediction with increased accuracy using different machine learning and decision tree approaches.

6. Reference

1. Ekström, M., Grose, M.R. and Whetton, P.H., 2015. An appraisal of downscaling methods used in climate change research. Wiley Interdisciplinary Reviews: Climate Change, 6(3), pp.301-319.
2. Pearse, R., 2017. Gender and climate change. Wiley Interdisciplinary Reviews: Climate Change, 8(2), p.e451.
3. Linnenluecke, M.K., Nucifora, N. and Thompson, N., 2018. Implications of climate change for the sugarcane industry. Wiley Interdisciplinary Reviews: Climate Change, 9(1), p.e498.
4. Lean, J.L., 2018. Observation-based detection and attribution of 21st century climate change. Wiley Interdisciplinary Reviews: Climate Change, 9(2), p.e511.

5. Raible, C.C., Pinto, J.G., Ludwig, P. and Messmer, M., 2021. A review of past changes in extratropical cyclones in the northern hemisphere and what can be learned for the future. Wiley Interdisciplinary Reviews: Climate Change, 12(1), p.e680.
6. Rajesh, P. and Karthikeyan, M., 2017. A comparative study of data mining algorithms for decision tree approaches using the Weka tool. Advances in Natural and Applied Sciences, 11(9), pp.230-243.
7. Huo, F., Xu, L., Li, Y., Famiglietti, J.S., Li, Z., Kajikawa, Y. and Chen, F., 2021. Using big data analytics to synthesize research domains and identify emerging fields in urban climatology. Wiley Interdisciplinary Reviews: Climate Change, 12(1), p.e688.
8. Ermolieva, T., Boere, E., Biewald, A., Havlík, P., Mosnier, A., Leclere, D., Valin, H., Frank, S., Obersteiner, M. and Ermoliev, Y., 2019. Addressing climate change adaptation with a stochastic integrated assessment model: Analysis of common agricultural policy measures. Financial Statistics Journal, 1(2).
9. Franzke, C.L., O'Kane, T.J., Berner, J., Williams, P.D. and Lucarini, V., 2015. Stochastic climate theory and modeling. Wiley Interdisciplinary Reviews: Climate Change, 6(1), pp.63-78.
10. Rajesh, P., Karthikeyan, M. and Arulpavai, R., 2019, December. Data mining approaches to predict the factors that affect the groundwater level using a stochastic model. In AIP Conference Proceedings (Vol. 2177, No. 1). AIP Publishing.
11. Rajesh, P. and Karthikeyan, M., 2019. Data mining approaches to predict the factors that affect agriculture growth using stochastic models. International Journal of Computer Sciences and Engineering, 7(4), pp.18-23.
12. Rajesh, P., Karthikeyan, M., Santhosh Kumar, B. and Mohamed Parvees, M.Y., 2019. Comparative study of decision tree approaches in data mining using chronic disease indicators (CDI) data. Journal of Computational and Theoretical Nanoscience, 16(4), pp.1472-1477.
13. Kohavi, R., & Sahami, M. (1996). Error-based pruning of decision trees. In International Conference on Machine Learning (pp. 278-286).
14. Akusok, A. (2020). What is Mean Absolute Error (MAE)? Retrieved from https://machinelearningmastery.com/mean-absolute-error-mae-for-machine-learning/
15. S. M. Hosseini, S. M. Hosseini, and M. R. Mehrabian, “Root mean square error (RMSE): A comprehensive review,” International Journal of Applied Mathematics and Statistics, vol. 59, no. 1, pp. 42–49, 2019.
16. Chi, W. (2020). Relative Absolute Error (RAE) – Definition and Examples. Medium. https://medium.com/@wchi/relative-absolute-error-rae-definition-and-examples-e37a24c1b566
17. https://www.kaggle.com/datasets/econdata/climate-change