EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Assessment of Global Forest Coverage through Machine Learning Algorithms

P. S. Metkewar[1,*], Ravi Chauhan[1], A. Prasanth[2] and Malathy Sathyamoorthy[3]

[1]Symbiosis Institute of Computer Studies and Research (SICSR), Symbiosis International (Deemed University) (SIU), Model Colony, Pune, Maharashtra, India
[2]Department of Electronics and Communication Engineering, Sri Venkateswara College of Engineering, Sriperumbudur, Tamilnadu, India
[3]Department of Information Technology, KPR Institute of Engineering and Technology, Coimbatore, India

Abstract

This exploration of paper presents an investigation of the Forest Region Inclusion Dataset that gives data on the backwoods inclusion of different nations overall from 1990 to 2020. The dataset contains country-wise information on population, population density, population development rate, total population rate, and forest region inclusion. We examined this dataset to decide the patterns in woodland region inclusion across various nations and mainlands, as well as the connection among populace and backwoods region inclusion. Our discoveries show that while certain nations have essentially expanded their forest region inclusion, others have encountered a decline. Besides, we found that population density and development rate are adversely related with forest area coverage. Authors have implemented four machine learning algorithms that are Linear Regression, Decision Tree, Random Forest and Support Vector Machine on the dataset.

Keywords: Forest Coverage, Deforestation, Remote Sensing, Ground Surveys, Environmental Issues, Climate Change, Machine Learning, Mean Squared error, R2 Score, Mean Absolute Error, Root Mean Square Error

Received on 30 November 2023, accepted on 04 February 2024, published on 15 February 2024

doi: 10.4108/eetsis.5122

*Corresponding author. Email: pravin.metkewar@gmail.com

1. Introduction

Timberlands are one of the world's most significant assets, giving an extensive variety of environmental, monetary, and social advantages. They are home to a critical extent of the world's biodiversity, give crucial environment administrations like carbon sequestration and water guideline, and backing the jobs of millions of individuals all over the planet.

Nonetheless, timberlands are under danger, with deforestation and woodland debasement happening at a disturbing rate in many regions of the planet. Deforestation is the long-lasting transformation of forested land to non-forested land, while woodland corruption alludes to the decrease in the quality and amount of backwoods assets, like lumber, fuelwood, and non-wood timberland items.

To resolve the issue of deforestation and woodland corruption, having exact and state-of-the-art data on the degree of timberland inclusion and changes in backwoods region after some time is fundamental. The Woodland Region Inclusion dataset gives this data, with information on backwoods inclusion in various nations from 1990 to 2020.

This examination paper expects to investigate the Woodland Region Inclusion dataset, analysing the progressions in timberland region throughout recent many years, the drivers of deforestation and backwoods debasement, and strategy and the board ways to deal with address the issue. The paper additionally tries to feature the significance of safeguarding and dealing with the world's timberlands for their natural, financial, and social worth.

The paper is organized as follows. The writing audit gives an outline of the drivers of deforestation and timberland debasement, as well as strategy and the executive's ways

to deal with address the issue. The strategy segment portrays the information sources and techniques used to dissect the Forest Region Inclusion dataset. The outcomes and conversation area presents the discoveries of the examination and talks about their suggestions. At last, the paper closes with a synopsis of the principal discoveries and their importance for timberland protection and the board.

2. Literature Review

Deforestation and backwoods corruption are intricate and complex issues that outcome from various drivers, including agribusiness extension, logging, mining, framework improvement, and urbanization. The review distributed in [1] investigate about issue of deforestation which has expanded as of late, bringing about the new expansion in tropical clammy backwoods aggravations (regular and anthropogenic debasement or deforestation). Without Agribusiness development is the main driver of deforestation, representing roughly 80% of worldwide woodland misfortune. This is frequently determined by interest for items, for example, palm oil, soybeans, and meat, which are utilized in food creation and different ventures.

Logging and mining likewise add to deforestation and backwoods debasement and carbon dioxide discharge displayed in the review distributed in [2], as they include the expulsion of trees and the annihilation of timberland natural surroundings. Framework improvement, like street development and hydroelectric dams, can likewise prompt deforestation, as it frequently includes getting huge regions free from forested land. Urbanization, in the interim, can prompt woodland debasement, as it frequently includes the transformation of forested land into metropolitan regions and the discontinuity of timberland environments.

To resolve the issue of deforestation and backwoods corruption, different arrangement and the board approaches have been proposed and carried out. One methodology is the foundation of safeguarded regions, for example, public parks and untamed life holds, which are intended to monitor woods and their biodiversity. In any case, the viability of safeguarded regions differs, contingent upon variables like administration, institutional limit, and financing.

One more methodology is the advancement of feasible backwoods the board rehearses, which look to adjust the financial, biological, and social upsides of timberlands. The examination displayed in [3] tells us economically oversaw timberlands give fundamental labour and products and hence have a crucial impact in supportable turn of events. Reasonable woods the executives include gathering lumber and non-wood backwoods items in a manner that keeps up with timberland wellbeing and efficiency, advances biodiversity preservation, and supports the jobs of neighbourhood networks.

Certificate plans, like the Forest Stewardship Board (FSC) and the PEFC, give a method for checking that woodland items have been delivered utilizing maintainable backwoods the executives rehearse. These plans have been fruitful in advancing reasonable woods the board and expanding interest for ensured items and concentrate in tells [4] the best way to decrease environmental fossil fuel by-products from tropical deforestation is at present viewed as a savvy choice for relieving environmental change.

Notwithstanding, the adequacy of these strategy and the executives' approaches differs among nations and areas, contingent upon variables like administration, institutional limit, and financial circumstances. In certain nations, powerless administration, defilement, and absence of authorization can prompt the disappointment of safeguarded regions and the proceeded with development of horticulture and different drivers of deforestation.

In rundown, deforestation and timberland corruption stay critical dangers to the world's woods, with farming extension, logging, mining, framework improvement, and urbanization being the essential drivers. Study was directed by which prepared and looked at different AI models including Choice trees, [10] Straight Relapse, [18] Irregular Woodland and SVM and recorded the most elevated precision with Arbitrary Timberland of 97.50%. Powerful strategy and the board draw near, like the foundation of safeguarded regions, practical timberland the executives, and confirmation plans, can assist with resolving the issue, however their viability relies upon a scope of variables. The Woodland Region Inclusion dataset gives significant data that can be utilized to illuminate this arrangement and the executives' choices and track progress towards accomplishing timberland preservation and reasonable administration objectives. Showing the carbon lack in Guyana [5] shows carbon capacity in the three safeguarded species at concentrate on destinations situated in Berbice (Locales 5 and 6), Demerara (Districts 3 and 4) and Essequibo (Areas 1 and 2), Guyana during the period 2014-16. One more [6] shows woodland cover change in the Gerecse Slopes, Hungary.

Authors observed that earlier no papers talk about the implementation of multiple machine learning algorithms due to this reason authors have taken up this study into consideration and have implemented four machine learning algorithms on the dataset. Also have noted the accuracy of all algorithms. This successful implementation of multiple machine learning algorithms helped in bridging the research gap.

3. Methodology

3.1 Data Set

We obtained the Forest Area Coverage Dataset, which contains country-wise data on forest area coverage, population, population density, population growth rate,

world population percentage, and other related variables from the World Bank's Open Data Portal. Authors have used Python and various data visualization tools such as [22] Matplotlib and Seaborn to analyse and visualize the dataset. Initially 214 rows and 40 columns were there and after pre-processing 214 rows and 12 columns were left altogether. Moreover, authors say that during pre-processing authors came across to null values for forest area coverage for the different years although, null values were less than 10, have deleted these null values and applied the algorithms on the cleaned data.

3.2 Experiment Design

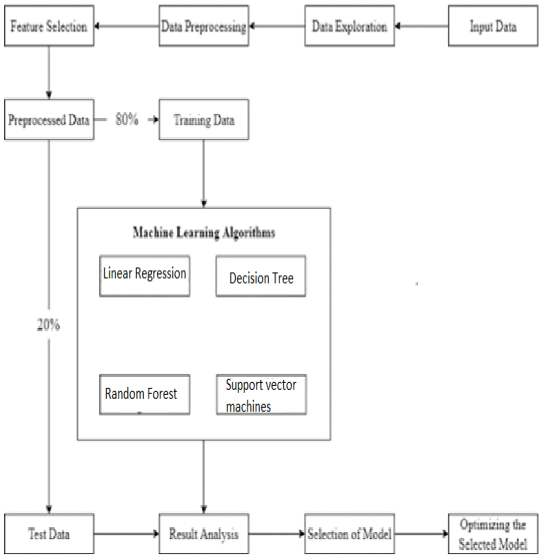

Figure 1. Experiment Design

The proposed system in Fig. 1 includes parting the dataset 80:20 into training and testing/approval information keeping a consistent irregular state as 0 in the wake of preprocessing, exploratory information examination, and component choice. The preparation subset of the information is utilized to prepare the different calculations and create a model. Upon effective preparation, the model is checked for exactness against the testing information. The other assessment boundaries utilized in the review are R2 score, Mean Absolute Error, Mean Squared Error and Root Mean Square Error. After testing the models and recording the expectations for each model, the anticipated worth and the genuine worth is looked at and the outcomes are examined utilizing R2 score as precision. Since an assortment of AI calculations and procedures can be applied to take care of the forest coverage inclusion issue, the expectation strategies considered for examination in this review [9], [11], [20] are Linear-Regression, [13], [15] Decision-Trees, [16], [17] Random-Forest- Regressor, and [20] Support-Vector-Machines. The main goal is to identify the optimum regressor with the most accurate results. Authors say that out of the four machine learning algorithms applied on the dataset, random forest is the best algorithm as it has the highest accuracy among all the algorithms so, random forest is the best algorithm to apply and predict the output for the dataset used in this research paper.

3.3 Data Pre-processing and Representation

As discussed, the dataset contains only 50 missing values and all were numeric so, we filled them all with the help of 'mean' strategy by using Simple Imputer, the categorical attributes containing a constant value are identified and dropped from the data. Since this problem falls under Prediction problem i.e.., Predicting the forest area coverage in upcoming years.

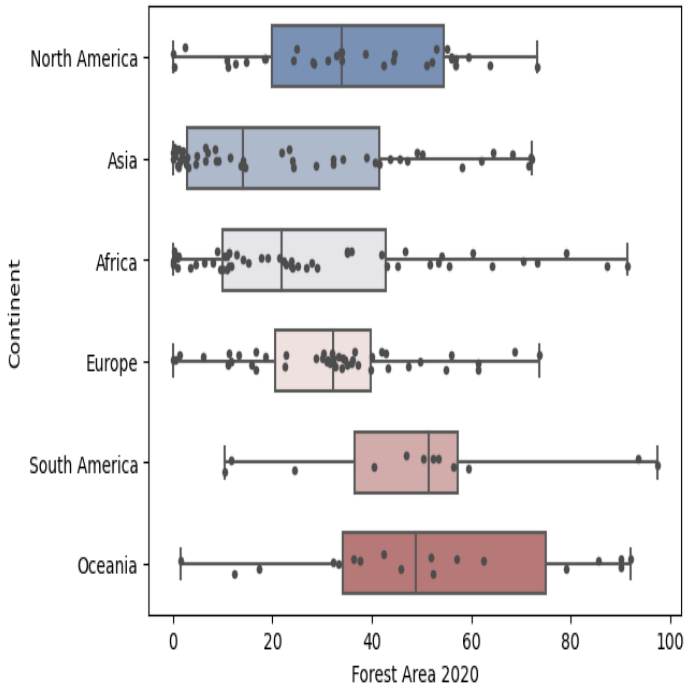

Figure 2. Continent wise Forest Area in the year 2020

The data is visualized in figure 2 highlights the continent wise forest area coverage in the year 2020.

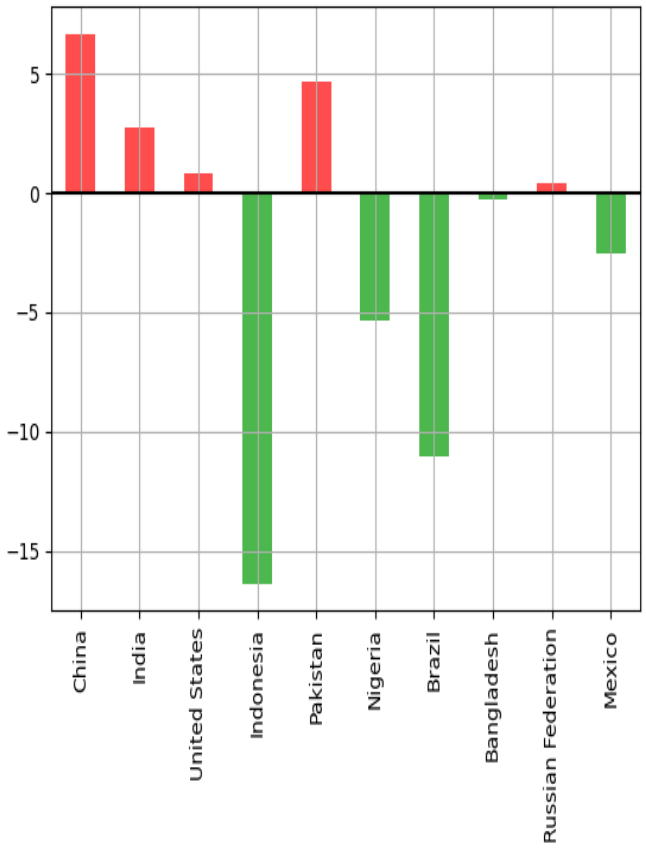

Figure 3. Top 10 countries which have the highest population rank.

As observed in figure 3. we can see population rank wise top 10 countries.

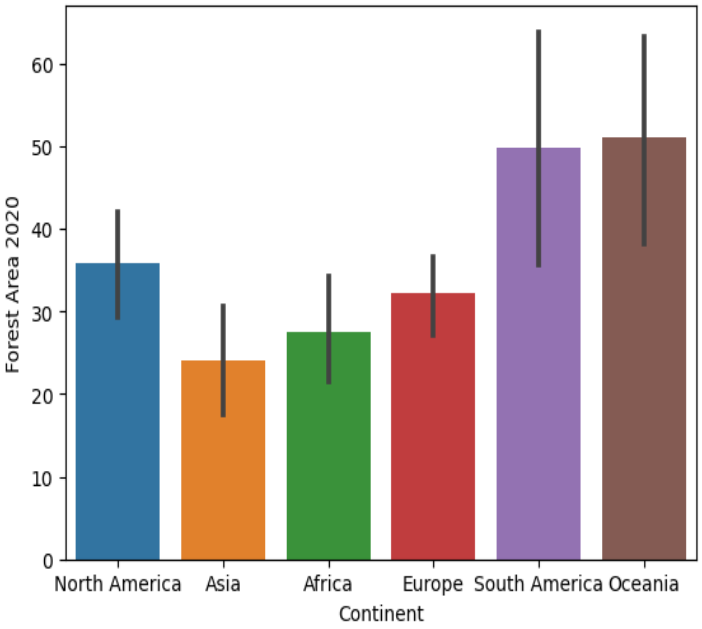

Figure 4. Continent wise forest area coverage in year 2020

As observed from figure 4, can see the continent wise forest area coverage in the year 2020 and can say that the Oceania is the highest with 50% approx. and Asia the least with approx. 25%.

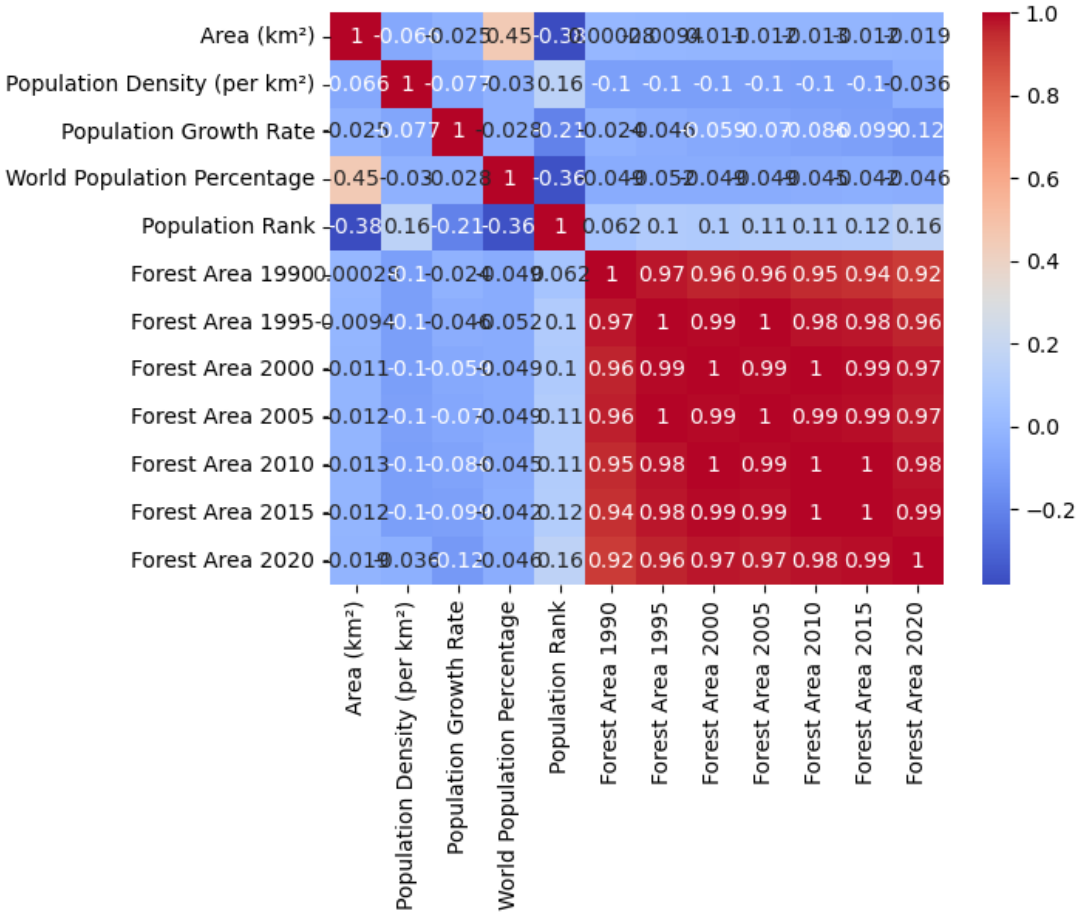

Figure 5. Correlation Heat Map

As observed from figure 5, the '*Forest Area 1990'* and '*Forest Area 1995'* have a strong positive correlation as the forest area in 1990 increases the forest area in 1995 as well.

3.4 Feature Selection

In our dataset, all the features are related with each other so, there was no need to select any feature. Moreover, authors have dropped the columns consisting of forest area for many years and only kept the forest area for the years which were divisible by 5 just to make the analysis job easier.

4. Results

After investigation and preprocessing of the information, the most effective highlights were chosen, the information was prepared and tried and the exactness was created of each model, Irregular Woodland had the most elevated precision among all the regressors contemplated, It was adjusted to show up at a precision of 97.50%, Different calculations additionally gave promising outcomes, Direct Relapse recorded an exactness of 92.95%. Decision Tree likewise gave exceptionally encouraging outcomes, Decision Tree recorded an exactness of 95.79 even though Decision Tree it is inclined to overfitting, the outcomes recorded are practical with a Mean Absolute Error 5.70 and SVM was the most exact as far as forecast and recorded the Root Mean Square error 20.12.

Table 1. Score analysis of all four algorithms

Index Number	Algorithms	R2 Score	Mean Absolute Error	Mean Square Error	Root Mean Square Error
0	Linear Regression	0.929097	9.860466	265.267929	16.287048
1	Decision Tree	0.953452	5.709302	174.151163	13.196635
2	Random Forest	0.970697	5.008140	109.631279	10.470496
3	Support Vector Machine	0.891751	12.627732	404.993029	20.124439

In view of the examination of the forest area coverage dataset in table.1 and Fig. 6, utilizing four distinct calculations - linear regression, [12], [14] decision tree, random forest, and [19] SVM - it has been resolved that arbitrary woodland is the best calculation for this specific dataset.

This end was arrived at after contrasting the presentation of every calculation utilizing fitting assessment measurements, for example, mean squared error, root mean squared error, and R-squared score. The outcomes showed that the arbitrary woods calculation had the most reduced mean squared error and root mean squared error, demonstrating that it had the option to even more likely foresee the timberland region inclusion than different calculations. Also, the R-squared score was higher for the irregular backwoods' calculation, demonstrating that it could make sense of a greater amount of the change in the information.

Subsequently, it very well may be reasoned that the irregular backwoods calculation is the most appropriate model for anticipating the timberland region inclusion in this dataset. This data could be significant for dynamic cycles connected with backwoods the board, protection, and ecological strategy.

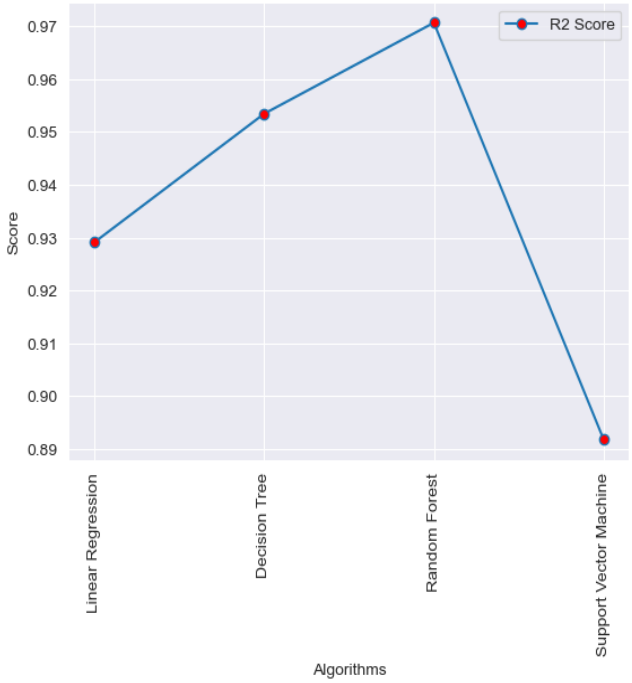

Figure 6. Graph shows R2 Score for all the four algorithms

Our examination of the Forest Area coverage Inclusion Dataset uncovers a few intriguing bits of knowledge into the patterns with regards to timberland region inclusion across various nations and mainland's. A portion of our key discoveries are as per the following:

1. Forest Region Inclusion Patterns: Our examination demonstrates that the woods region inclusion has been generally steady in certain nations, while others have encountered a huge increment or decline in backwoods region inclusion. For instance, the woodland region inclusion in Brazil, which had declined steeply until the mid-2000s, has been expanding beginning around 2015. Conversely, nations, for example, Indonesia and Malaysia have encountered a decrease in their timberland region inclusion.
2. Population and Woods Region Inclusion: We found a negative connection between population density and woodland region inclusion, demonstrating that nations with higher populace densities will generally have lower backwoods region inclusion. Likewise, we found a negative connection between population development rate and timberland region inclusion, demonstrating that nations with higher populace development rates will more often than not have lower woods region inclusion.
3. Regional Contrasts: Our examination additionally uncovers local contrasts in timberland region inclusion patterns. For example, while Africa has the most elevated woods region inclusion as a level of land region, it has likewise encountered the main decrease in woodland region inclusion. Conversely, North America

has the most minimal woods region inclusion yet has encountered an expansion in backwoods region inclusion as of late.

5. Discussion

The Forest Area Coverage Inclusion dataset gives significant data on the degree of backwoods inclusion in various nations throughout the course of recent many years. The dataset shows that the worldwide woods region has declined by 178 million hectares somewhere in the range of 1990 and 2020, addressing a total deficit of 5.2 million hectares each year.

The dataset additionally features huge contrasts in backwoods inclusion among nations, for certain nations having exceptionally high timberland inclusion, while others have extremely low woods inclusion. For instance, Suriname, Gabon, and Guyana have the most elevated woodland inclusion, with more than 90% of their property region covered by backwoods, while nations like Iceland, Qatar, and Saudi Arabia have no timberland inclusion by any stretch of the imagination.

The dataset additionally shows that the total populace has developed by over 1.8 billion individuals starting around 1990, coming to 7.8 billion of every 2020. The populace development rate changes generally among nations, for certain nations encountering high populace development rates, while others have negative development rates. The populace development rate can influence backwoods inclusion, as high populace development rates can prompt expanded interest for rural land, fuelwood, and other woodland assets.

The conversation of this exploration paper centres around the drivers of deforestation and woods corruption, strategy, and the board ways to deal with address the issue of deforestation and woodland debasement. The concentrate in [7] shows the assurance and feasible utilization of woodland biological variety is of extraordinary importance to backwoods security and supportable administration. The writing survey features that deforestation is basically determined by horticultural extension, logging, mining, foundation advancement, and urbanization.

To resolve the issue of deforestation, different strategy and the executives' approaches have been proposed and carried out, for example, laying out safeguarded regions, advancing reasonable woods the board practices, and accreditation plans. Be that as it may, the viability of these methodologies' changes among nations and areas, contingent upon variables like administration, institutional limit, and financial circumstances.

All in all, the Forest Region Inclusion dataset gives fundamental data on the degree of woodland inclusion in various nations throughout recent many years. The dataset features massive contrasts in timberland inclusion among nations, for certain nations having exceptionally high woods inclusion, while others have extremely low woodland inclusion. The drivers of deforestation and woods debasement are mind boggling and fluctuate among nations and districts, requiring fitted strategy and the executives' ways to deal with address the issue.

Besides, in the review [8] it is shown that the timberlands give a scope of biological system administrations fundamental for human prosperity.

Conclusion

Forest Area Coverage Inclusion dataset gives a complete outline of the condition of timberlands all over the planet. The dataset uncovers that worldwide woodland region has declined by 178 million hectares somewhere in the range of 1990 and 2020, addressing an overall deficit of 5.2 million hectares each year. The dataset likewise features massive contrasts in woods inclusion among nations, for certain nations having extremely high backwoods inclusion, while others have exceptionally low timberland inclusion.

The drivers of deforestation and woods corruption are intricate and fluctuate among nations and locales, with horticulture extension, logging, mining, framework advancement, and urbanization being the essential drivers. The fast development of the total populace likewise adds to the issue of deforestation, as it prompts expanded interest for timberland assets.

To resolve the issue of deforestation, different strategy and the executives' approaches have been proposed and carried out, for example, laying out safeguarded regions, advancing economical backwoods the board practices, and certificate plans. In any case, the adequacy of these methodologies shifts among nations and areas, contingent upon variables like administration, institutional limit, and financial circumstances.

Obviously, there is nobody size-fits-all answer for the issue of deforestation and timberland debasement. All things being equal, there is a requirement for custom fitted strategy and the board moves toward that consider the particular drivers of deforestation and the financial states of the impacted locales. The Timberland Region Inclusion dataset can be utilized to illuminate this arrangement and the board choices by giving cutting-edge data on woodland inclusion and changes in backwoods region after some time.

This examination paper features the significance of saving and dealing with the world's woods, for their biological worth as well as for their financial and social worth. It is indispensable that states, associations, and people cooperate to track down viable answers for the issue of deforestation and timberland corruption to guarantee the proceeded with presence of these fundamental biological systems for people in the future.

Acknowledgements.
We would like to acknowledge SICSR management for providing necessary infrastructure to work on this research project.

References

[1] Vancutsem C, Achard F, Pekel JF, Vieilledent G, Carboni S, Simonetti D, Gallego J, Aragao LE, Nasi R. Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. Science Advances. 2021 ;7(10):eabe1603.

[2] DeFries RS, Rudel T, Uriarte M, Hansen M. Deforestation driven by urban population growth and agricultural trade in the twenty-first century. Nature Geoscience. 2010 ;3(3):178-81.

[3] FAO F. Agriculture Organization: Global Forest Resources Assessment. FAO, Rome, Italy. 2010.

[4] Abdulqader AW. Predicting Carbon Dioxide Emissions with the Orange Application: An Empirical Analysis. Mesopotamian Journal of Computer Science. 2023 ;2023:56-66.

[5] JAIKISHUN S, Ansari AA, DASILVA P, HOSEN A. Carbon storage potential of mangrove forest in Guyana. International Journal of Bonorowo Wetlands. 2017;7(1):43-54.

[6] Kucsicsa G, Popovici EA, Bălteanu D, Dumitraşcu M, Grigorescu I, Mitrică B. Assessing the potential future forest-cover change in Romania, predicted using a scenario-based modelling. Environmental Modeling & Assessment. 2020 ;25:471-91.

[7] Li J, Zhou A, Liao Y, Zhao Z, Mao X, Zhang S. Forest ecological diversity change prediction discrete dynamic model. Discrete Dynamics in Nature and Society. 2022 ;2022:1-1.

[8] Pecchi M, Marchi M, Moriondo M, Forzieri G, Ammoniaci M, Bernetti I, Bindi M, Chirici G. Potential impact of climate change on the forest coverage and the spatial distribution of 19 key forest tree species in Italy under RCP4. 5 IPCC trajectory for 2050s. Forests. 2020 ;11(9):934.

[9] Reddy SR, Malathi P. Linear Regression and Artificial Neural Networks based Efficient Sales Forecasting Model with Increased Prediction Accuracy. In2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA) 2022 (pp. 1100-1105). IEEE.

[10] C. G. Raju, V. Amudha and S. G.: Comparison of Linear Regression and Logistic Regression Algorithms for Ground Water Level Detection with Improved Accuracy. Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2023, pp. 1-6.

[11] Cao Z, Liu L, Markowitch O. Comment on “highly efficient linear regression outsourcing to a cloud”. IEEE Transactions on Cloud Computing. 2017 6;7(3):893-.

[12] Zhang B. Tactical decision system of table tennis match based on C4. 5 decision tree. In2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) 2021 (pp. 632-635). IEEE.

[13] Mitrofanov S, Semenkin E. An approach to training decision trees with the relearning of nodes. In2021 International Conference on Information Technologies (InfoTech) 2021 (pp. 1-5). IEEE.

[14] Yang FJ. An extended idea about decision trees. In2019 International Conference on Computational Science and Computational Intelligence (CSCI) 2019 (pp. 349-354). IEEE.

[15] Euldji R, Boumahdi M, Bachene M. Decision-making based on decision tree for ball bearing monitoring. In2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH) 2021 (pp. 171-175). IEEE.

[16] Miah MO, Khan SS, Shatabda S, Farid DM. Improving detection accuracy for imbalanced network intrusion classification using cluster-based under-sampling with random forests. In2019 1st international conference on advances in science, engineering and robotics technology (ICASERT) 2019 (pp. 1-5). IEEE.

[17] Guo Y, Zhou Y, Hu X, Cheng W. Research on recommendation of insurance products based on random forest. In2019 international conference on machine learning, big data and business intelligence (MLBDBI) 2019 (pp. 308-311). IEEE.\

[18] Sapra V, Sapra L, Vishnoi A, Srivastava P. Identification of Brain Stroke using Boosted Random Forest. In2022 International Conference on Advances in Computing, Communication and Materials (ICACCM) 2022 (pp. 1-5). IEEE.

[19] Zhili C, Guiyan J. Application of multiple SVM classifier fusion technique in freeway automatic incident detection. In2008 27th Chinese Control Conference 2008 (pp. 581-585). IEEE.

[20] Yang Y, Wang J, Yang Y. Exploiting rotation invariance with SVM classifier for microcalcification detection. In 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI) 2012 (pp. 590-593). IEEE.

[21] Rao NS, Thangaraj SJ, Kumari VS. Flight Ticket Prediction Using Gradient Boosting Regressor Compared With Linear Regression. In 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) 2023 (pp. 1-6). IEEE.

[22] Rasoarahona R, Wattanadilokchatkun P, Panthum T, Thong T, Singchat W, Ahmad SF, Chaiyes A, Han K, Kraichak E, Muangmai N. Optimizing Microsatellite Marker Panels for Genetic Diversity and Population Genetic Studies: An Ant Colony Algorithm Approach with Polymorphic Information Content. Biology 2023; 12, 1280.

[23] Arvindhan M, Rajeshkumar D, Pal AL. A review of challenges and opportunities in machine learning for healthcare. Exploratory Data Analytics for Healthcare. 2021; pp 67-84.

[24] Chandraprabha M, Dhanaraj RK. Machine learning based Pedantic Analysis of Predictive Algorithms in Crop Yield Management. In2020 4th International conference on electronics, communication and aerospace technology (ICECA) 2020 (pp. 1340-1345). IEEE.