VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering

Project Report on

Predictive Analysis of Carbon Footprint for Forest Cover in India

In partial fulfillment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer

Engineering at the University of Mumbai

Academic Year 2023-24

Submitted by
Namrata Avhad (D17-A , Roll no - 03)
Sonal Belani (D17-A , Roll no - 05)
Utsav Gavli (D17-A , Roll no - 17)
Nikkita Gurnani (D17-A , Roll no - 20)

Project Mentor

Mrs.Lifna CS

(2023-24)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering

Certificate

This is to certify that **Namrata Avhad (D17A/03) , Sonal Belani (D17A/05), Utsav Gavli (D17A/17),Nikkita Gurnani (D17A/20)** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on **"Predictive Analysis of Carbon footprint for forest cover in India"**as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor **Prof.Mrs.Lifna CS** in the year 2023-24 .

This project report entitled **Predictive Analysis of Carbon footprint for forest cover in India** by **Namrata Avhad, Sonal Belani , Utsav Gavli , Nikkita Gurnani** is approved for the degree of **B.E Computer Engineering**

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:

Project Guide:

Project Report Approval
For
B. E (Computer Engineering)

This project report entitled **Predictive Analysis of Carbon footprint for forest cover in India** by **Namrata Avhad (D17A/03) , Sonal Belani (D17A/05), Utsav Gavli (D17A/17),Nikkita Gurnani (D17A/20)** is approved for the degree of **B.E Computer Engineering**

Internal Examiner

External Examiner

Head of the Department

Principal

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

---------------------------------------   ---------------------------------------

Namrata Avhad (D17A/03)   Sonal Belani (D17A/05)

---------------------------------------   ---------------------------------------

Utsav Gavli (D17A/17)   Nikkita Gurnani (D17A/20)

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Prof. Lifna C.S** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solutions for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop a professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Index

7

List of Figures:

List of Tables:

Abstract

Forests cover about 30% of the Earth's land surface. As forests grow, their trees take in carbon from air and store it in wood, plants and under the soil. If not for forests, much of the carbon would remain in the atmosphere in the form of carbon dioxide (CO_2). In most of the developing countries, the increasing rate of GHG emissions is considered as a major cause of concern. India is leading in terms of GHG emissions as compared to other countries. The vegetation cover comprises only 24.39% of the geographic area of India. The primary objective of this proposal is to identify the relationship between the increase in GHG emissions and forest cover in metropolitan cities. An additional objective is to predict the amount of afforestation required for each area to cope up with the GHG emissions over the next 25-years.It can be achieved using time series forecasting models like ARIMA ,SARIMAand other machine learning models. The proposal provides suggestions on optimal techniques like use of renewable resources, such as Solar power for sustainable afforestation and to cope up with loss of GHG emission.

Keywords: GHG Emission , Vegetation Cover , ARIMAX , SARIMA , Solar Power , Machine learning , Renewable Energy

Chapter 1: Introduction

1.1 Introduction

A carbon footprint is the total amount of greenhouse gases that are generated by human activities. Some of the major greenhouse gases are CO2 i.e carbon dioxide, CH4 i.e methane, N20 i.e nitrous oxide.

Annually, India generates 2.88 billion tonnes of carbon dioxide through various activities, Nitrous Oxide generated is 272 million tonnes , Methane generated is 897 million tonnes. Out of India's total population, 141 crores (about 35% of total) live in urban regions. The average urban Indian generates 1.32 tons of carbon dioxide over the course of a year. The 493 million people who live in metropolitan areas in India are responsible for production of 650 million tons(about 22% of the total) of CO2.

According to the survey of the Global Climate Risk Index (CRI) of 2017, India is 14th on the list of most vulnerable countries. India ranked the second highest for the rate of deforestation after losing 668,400 hectares of forest cover in the last 30 years. India also topped the chart for biggest increase in deforestation between 1990 and 2020 with a difference of 284,400 hectares in forestry loss. According to Global Forest Watch, from 2001 to 2022, India lost 2.19 Mha of tree cover, equivalent to a 5.6% decrease in tree cover since 2000, and 1.11 Gt of CO_2e emissions.This research aims to calculate the content of carbon across various states of India and calculate its relationship with forest cover. This helps to determine the requirements of the future and shift from fossil fuels which are a major source of carbon emissions to renewable energy such as solar power. As of the latest India State of the Forest Report (ISFR) 2021, the total aboveground carbon stock in India is estimated to be around 7,204 million tonnes. This includes carbon stored in both natural forests and Trees Outside Forests (TOF).

Here's a breakdown of the carbon stock:

Forest Carbon Stock: 6,674.53 million tonnes, primarily stored in:

Trees: 6,385.06 million tonnes

Bamboo: 205.64 million tonnes

Mangroves: 83.83 million tonnes

Trees Outside Forests (TOF) Carbon Stock: 529.47 million tonnes, stored in trees outside designated forest areas, including urban trees, plantations, and agroforestry systems.

This research, coupled with the integration of solar power dynamics, is essential for several reasons. Firstly, it enables the assessment of the impact of forest cover changes and solar power expansion on carbon sequestration and emissions. Secondly, it provides insights into the complex interactions between land-use changes, renewable energy deployment, and their combined impact on carbon dynamics. Thirdly, it offers a basis for formulating holistic policies and strategies that balance forest conservation, renewable energy development, and climate change mitigation efforts.

1.2 Motivation

The research on predictive analysis of carbon footprint and its impact on forest cover in India represents a critical academic pursuit with profound implications. India, characterized by its diverse ecosystems and burgeoning population, confronts substantial challenges in harmonizing economic development with environmental preservation. This research endeavors to unravel the intricate interplay between carbon emissions and changes in forest cover, offering invaluable insights for policymakers, environmental advocates, and industrial entities. The outcomes of this study hold the potential to inform targeted interventions and promote sustainable practices, thereby guiding India towards a more environmentally conscious and sustainable trajectory.

Furthermore, this research serves as a clarion call for heightened environmental consciousness and concerted action among all stakeholders. The data-centric approach of predictive analysis provides a methodical and objective framework for assessing the ramifications of human activities on the environment. By quantifying the carbon footprint and its repercussions on forest cover, this study underscores the imperative of adopting sustainable practices and curtailing carbon emissions. It transcends academic realms to underscore the urgent need for collective responsibility and proactive measures to safeguard our environment for posterity.

1.3 Problem Definition

Over the years, the annual rate of global CO2 emissions has gradually increased. Global CO2 emissions have been reaching or exceeding 40 billion metric tons per year in recent decades. Carbon emissions have played a crucial role in the rise of climate change as a global crisis in recent years. In order to lessen the negative effects of climate change, forest cover must be managed and preserved. Ongoing deforestation, alterations in land use, and natural disasters, however, may cause the release of carbon dioxide that has been sequestered in the atmosphere to return to the atmosphere, aggravating global warming. The aim of our project is to carry out an extensive analysis of carbon emissions related to changes in forest cover. In order to understand how deforestation, carbon release, and potential climatic impacts relate to one another, the analysis will require gathering and analyzing data on forest cover, land-use changes, and carbon emissions. Our main goal is to develop a system that will forecast and predict carbon footprint levels using Machine Learning algorithms. The system would include a comparative Analysis of algorithms such as XGBoost and Random Forest on the vegetation cover in various states of India and would recommend simple and easy ways to reduce CO2 emissions. Our aim is to incorporate data visualizations such as pie charts, heat diagrams, line graphs, scatter plots, etc.

1.4 Existing systems

Several existing systems and approaches are dedicated to the predictive analysis of carbon footprint and its impact on forest cover in India, leveraging a range of technologies and methodologies. Remote sensing and Geographic Information Systems (GIS) play a crucial role in monitoring forest cover changes and estimating carbon stocks. These systems analyze satellite imagery and spatial data to detect deforestation, forest degradation, and carbon emissions. Additionally, carbon sequestration models like the Integrated Carbon Observation System (ICOS) and the Forest Carbon Index (FCI) are used to estimate the carbon sequestration potential of forests. These models take into account factors such as forest type, biomass, and land use changes to predict carbon sequestration rates, providing valuable insights into the role of forests in mitigating climate change.

Moreover, machine learning algorithms and artificial intelligence are increasingly being applied to predict carbon footprint and forest cover changes in India. These models can analyze complex datasets and identify patterns to predict future trends in carbon emissions and forest cover. Furthermore, various carbon footprint calculators are available online that allow individuals and organizations to estimate their carbon emissions. By considering factors such as energy consumption, transportation, and waste generation, these calculators provide insights into individual and collective carbon footprints, encouraging sustainable practices and informed decision-making.

1.5 Lacuna of the existing systems

- Existing systems or projects may rely on outdated data sources, which can lead to   inaccurate predictions and hinder the ability to respond to current environmental challenges.
- Some existing systems or projects may focus on specific regions or states in India, leaving out other critical areas where monitoring and intervention are needed.
- Data may be stored in isolated silos within different agencies or organizations, making it difficult to integrate and analyze comprehensively.
- :Scalability challenges may hinder the expansion of existing systems to cover larger areas or accommodate increasing volumes of data.

1.6 Relevance of the Project

The project on predictive analysis of carbon footprint and its impact on forest cover in India holds immense relevance in the context of sustainable development and environmental conservation. By understanding the intricate relationship between carbon emissions and forest cover changes, this research can inform policy decisions and conservation efforts aimed at mitigating climate change. Additionally, as India grapples with the dual challenges of economic growth and environmental preservation, this project can provide valuable insights into balancing these objectives. Furthermore, the findings of this research can guide stakeholders in adopting sustainable practices and promoting biodiversity conservation, thus ensuring a greener and more sustainable future for India and the planet.

Chapter 2: Literature Survey

A. Brief Overview of Literature Survey

The literature survey chapter presents insights from prior research on predictive analysis of carbon footprint for forest cover in India. It serves to inform the project's development by identifying methodologies, technologies, tools, and challenges from existing work, thereby guiding the implementation of effective strategies and ensuring a robust foundation for the proposed solution.

B. Related Works

2.1 Research Papers Referred

2.1.1 Jemyung Lee , Oliver Taherzadeh, , Keiichiro Kanemoto , "The scale and drivers of carbon footprints in households, cities, and regions across India", Research Institute for Humanity and Nature, 2021.[1]

a) Abstract : The paper "The scale and drivers of carbon footprints in households, cities, and regions across India" provides a comprehensive analysis of the carbon footprints in households and districts in India. The study aims to understand the total contribution of different entities to climate change in India, focusing on the carbon footprint analysis at the household level across 623 districts in India, based on micro consumption data from 203,313 households. The study reveals significant differences in carbon footprints between different socio-economic and demographic groups within the country, shedding light on the disparities in consumption patterns and their associated impacts on carbon emissions.

b) Inference : The findings suggest that while poverty alleviation in India may not significantly increase the country's overall carbon footprint, economic expansion and the shift from medium to high expenditure households could lead to a substantial increase in carbon emissions. The study underscores the importance of understanding the interplay between economic development, poverty alleviation, and climate change mitigation in India to achieve sustainable and equitable growth. The researchers emphasize the need for targeted interventions and policy measures that consider the interplay between economic development and environmental sustainability to ensure a balanced and inclusive approach to sustainable growth in the country.

2.1.2 Qi Huanga, Heran Zheng , Jiashuo Li , Jing Mengc, Yunhui Liud, Zhenyu Wange,Ning Zhanga, Yuan Li ,, Dabo Guanc, "Heterogeneity of consumption-based carbon emissions and driving forces in Indian states, Advances in Applied Energy, ,Volume 4, November 2021 [2]

a) Abstract : As the second most populous country in the world, India is on the way to rapid industrialization and urbanization, possibly becoming the next carbon giant. With its vast territory and high regional heterogeneity in terms of development stages and population, state-level consumption-based emissions patterns and driving forces are critical but unfortunately, remain far from completed. This paper explores the regional heterogeneity of consumption-based carbon emissions and driving forces in Indian states. The study uses a multi-regional input-output model to ascertain the heterogeneity in consumption-based emissions and track carbon flows in the inter-state supply chain, based on a newly constructed Indian multi-state input-output table for 2015, using the Flegg location quotient method.

b) Inference : The findings reveal that household consumption dominates consumption-based emissions at state levels, accounting for a significant proportion of total consumption-based emissions, while investment-led emissions are relatively higher in developed regions than in developing regions. Moreover, the study highlights the significant spillover effect, indicating that more than 30% of consumption-based emissions in developed states were imported from less developed states with higher carbon intensity, suggesting a substantial inter-state cooperation is recommended for mitigating emissions.

2.1.3 Ishma Amin, Meenakshi Agarwal, Stevert Lobo, Rahul Gurnani, and Prof. (Mrs.) Priya R.L, "Concept for Mapping Carbon footprint with Change in Vegetation Cover and Population in India" ,ITM Web Conf. , Volume 32, 2020 International Conference on Automation, Computing and Communication 2020 (ICACC-2020) [3]

a) Abstract : In most of the developing countries, the increasing rate of Carbon emissions is considered as a major cause of concern. India is leading in terms of CO2 emissions as compared to other countries. The vegetation cover comprises only 24.39% of the geographic area of India. Metropolitan cities in India are witnessing rapid urbanization. The primary objective of this proposal is to identify the relationship between the increase in carbon emissions and deforestation in metropolitan cities. An additional objective is to predict the amount of afforestation required for each area to cope up with the carbon emissions over the next 25-years. It can be achieved by using statistical models like ARIMA, LSTM and machine learning techniques such as Random Forest. The

proposal provides suggestions on optimal places and techniques for sustainable afforestation to the concerned authorities using artificial intelligence.

b) Inference : This study consists of three distinct phases. The first phase is to identify how carbon footprint growth is associated with deforestation. The second phase estimates the additional vegetation cover required over the next 25 years to cope up with the increase in carbon footprint and deforestation during that period. The third phase of this research suggests optimal techniques for sustainable afforestation according to region specific needs.This study addresses the optimization and management of tree ecosystems in major metropolitan cities. Considering the growing population of metropolitan cities, the aim is for sustenance along with development.

2.1.4 Surbhi Kumari, ,Sunil Kumar Singh , " Machine learning-based time series models for effective CO2 emission prediction in India" , Environmental Science and Pollution Research , June 25, 2022 [5]

a) Abstract : China, India, and the USA are the countries with the highest energy consumption and CO2 emissions globally. As per the report of datacommons.org, CO2 emission in India is 1.80 metric tons per capita, which is harmful to living beings. Increasing CO2 emissions can affect human health in two ways; directly and indirectly. It affects directly when inhaled in high dosage and can be the cause of serious diseases such as breathlessness, blindness, dizziness, and even delirium .Global problems such as climate change, acid rain, and global warming can also be seen in the indirect form of high CO2 emissions .All these forms of emissions are highly hazardous for human beings and the environment.So this paper presents India's detrimental CO2 emission effect with the prediction of CO2 emission for the next 10 years based on univariate time-series data from 1980 to 2019.

b) Inference : In this paper they have used three statistical models; autoregressive-integrated moving average (ARIMA) model, seasonal autoregressive-integrated moving average with exogenous factors (SARIMAX) model, and the Holt-Winters model, two machine learning models, i.e., linear regression and random forest model and a deep learningbased long short-term memory (LSTM) model. This paper brings together a variety of models and allows us to work on data prediction. The performance analysis shows that LSTM, SARIMAX, and Holt-Winters are the three most accurate models among the six models based on nine performance metrics. Results conclude that LSTM is the best model for CO2 emission prediction with the 3.101% MAPE value, 60.635 RMSE value, 28.898 MedAE value, and along with other performance metrics. The deep learning-based LSTM model is suggested as one of the most appropriate models for CO2 emission prediction.

2.1.5 B. Mohan Kumar , Sreejith Aravindakshan , "Carbon footprints of the Indian AFOLU (Agriculture, Forestry, and Other Land Use) sector: a review " , Arunachal University of Studies, May 24, 2022 [6]

a) Abstract : Stabilizing greenhouse gas emissions from croplands is crucial for mitigating climate change, with the Agriculture, Forestry, and Other Land Use sector playing a significant role in both emitting and absorbing carbon, making it essential for India to focus on reducing emissions from this sector to achieve its goal of net-zero emissions by 2070. Depending on management, the Agriculture, Forestry, and Other Land Use (AFOLU) sector can be both a source as well as a net sink for carbon. Currently, it contributes 25% of the global anthropogenic carbon emissions. Although India's emissions from this sector are around 8% of the total national GHG emissions, it can contribute significantly to the country's aspirations of reaching net-zero emissions by 2070. In this review, they explained the carbon footprints of the AFOLU sector in India, focusing on enteric fermentation, fertilizer and manure management, rice paddies, burning of crop residues, forest fires, shifting cultivation, and food wastage. Furthermore, using the standard autoregressive integrated moving average method, they projected India's AFOLU sector emission routes for 2070 under four scenarios: business as usual (BAU) and three emission reduction levels, viz., 10%, 20%, and 40% below BAU.

b) Inference : The AFOLU sector contributed to about 8% of India's total GHG emissions in 2015 , which is much lower than the global share of AFOLU . This sector encompasses GHG emissions and removals arising from carbon stock changes in biomass, CO_2 and non-CO_2 GHG emissions from detritus and mineral soils, and CO_2 and non-CO_2 GHG emissions from fire . Rice paddies and livestock production systems (enteric fermentation) emit the most CH_4, managed soils emit N_2O, and manure management systems emit both CH_4 and N_2O . The photosynthetic process in plants, especially in forest trees, represents a significant CO_2 removal mechanism. While permanent removal of trees will lead to increasing emissions, increasing forest cover and other tree-based production systems, such as agroforestry can contribute substantially to CO_2 removals .

2.1.6 T.V. Ramachandra , Shwetmala , "Decentralized carbon footprint analysis for opting climate change mitigation strategies in India" , Renewable and Sustainable Energy Reviews , Volume 16, Issue 8, October 2012 [4]

a) Abstract : This Paper deals with the study on emission of carbon dioxide from various emission inventories in an educational institution. Carbon dioxide is the chief greenhouse gas that results from human activities and causes global warming and climatic change. The burning of the organic materials in fossil fuels produces energy and releases carbon dioxide and other compounds into the

earth's atmosphere.Greenhouse gases can be emitted through transport, land clearance, production and consumption of food, fuels, manufactured goods, materials, wood, roads, buildings and services etc. A carbon footprint is the measure of the amount of greenhouse gases, measured in units of carbon dioxide, produced by human activities. A carbon footprint can be measured of an individual or an organization, and is typically given in tons of carbon dioxide equivalent per year. The study has been undertaken in the college campus in order to evaluate the amount of carbon dioxide produced and to suggest remedial measures for the reduction of emissions as a part of social commitment.

b) Inference : The paper discusses the carbon flow in Indian forests and variations in forest biomass and productivity. It highlights the importance of forest cover in carbon storage and the potential for sequestering carbon in forest soil. It mentions the need for a decentralized inventory of carbon emissions and sequestration potential at disaggregated levels. This would help in implementing carbon capture strategies and developing region-specific mitigation measures. Improvements in agricultural practices can increase the quantity of organic carbon in soil, further contributing to carbon sequestration .A decentralized inventory of carbon emissions and sequestration potential at disaggregated levels is necessary for implementing effective carbon capture strategies. This requires sector-wise analysis of sources and sinks at regional levels and the development of region-specific mitigation measures. The paper also references other studies on air pollution in urban India, emissions from India's transport sector, and the environmental impacts of trade liberalization on India's automobile sector.5.The paper concludes with carbon status, results and discussion, and overall conclusions.

2.1.7 Jingwei Han , Zhixiong Tan , Maozhi Chen , Liang Zhao , Ling Yang and Siying Chen , "Carbon Footprint Research Based on Input–Output Model—A Global Scientometric Visualization Analysis " International Journal of Environmental Research and Public Health [7]

a) Abstract : Reducing the effect of mankind's activities on the climate and improving adaptablity to global warming have become urgent matters. With the development of industrialization, global warming has caused irreversible damage to the environment that human beings depend on in terms of sea level rise, food crises, water shortages and so on. Scientific research shows that greenhouse gas (GHG) emissions are the main cause of global warming. Reducing GHG emissions has become the worldwide target since the first Intergovernmental Panel on Climate Change (IPCC) assessment report was released in 1990 .The carbon footprint (CF), derived from the concept of ecological footprint, has been used to assess the threat of climate change in recent years. As a "top to bottom" method, input–output analysis (IOA) has become a universally applicable CF assessment tool for

tracing the carbon footprint embodied in economic activities. A wide range of CF studies from the perspective of the IOA model have been presented and have made great progress. The purpose of this paper is to explore the knowledge structure and frontier trends in respect of the IOA model applied to CF research using scientometric visualization analysis.

b) Inference : The document focuses on the analysis of the knowledge flow and evolution in the field of IOA-related CF (carbon footprint) assessment. It aims to provide guidance for future research in this area. They conducted a comprehensive review of 491 studies retrieved from the WOS (Web of Science) database from 2008 to 2021 using bibliometric analysis methods. Through information visualization, knowledge-network analysis, and knowledge-evolution analysis, the research field of IOA-related CF became clearer. The paper highlights the importance of bibliometric analysis in exploring the knowledge structure and trends of a discipline. It identifies three main knowledge-flow paths in the field of CF assessment with IOA, with knowledge carriers distributed among journal clusters on the left and knowledge sources distributed among journal clusters on the right. At last the paper mentions the application of scientometric visualization analysis in exploring the intellectual structure and evolution history of related fields, such as CF research, energy and environment research, and carbon neutralization goals.

2.1.8 Andréea Lorena Radua , Marian Albert Scrieciua , Dimitriu Maria Caracotaa , "Carbon Footprint Analysis: Towards a Projects Evaluation Model for Promoting Sustainable Development " , International Economic Conference of Sibiu 2013 Post Crisis Economy: Challenges and Opportunities, IECS 2013 [9]

a) Abstract : Environmental protection has now become a major concern, especially following the significant negative consequences involved by the economic development promoted since the industrial revolution. People become progressively aware of their activities' implications on the environment, and are increasingly interested in reducing and correcting the adverse effects. A growing number of studies, research and collected data, reveal the existence of a direct relationship between climate change and carbon dioxide emissions (CO_2) (IEA, 2012). According to the Fourth Assessment Report prepared by Intergovernmental Panel on Climate Change (IPCC), activities of all nations generate increasingly more GHG emissions, having significant negative impacts on climate change due to alterations taking place in the compositional level of the atmosphere, and also on rising the average global temperature since the mid of the 20th Century (IPCC, 2007).The main elements that generate large amounts of carbon dioxide are fossil fuels (especially oil and coal), through burning them for obtaining energy. Of all the greenhouse gases, CO_2 has the largest share. Thus, emissions of other greenhouse gases (CH_4, N_2O, HFC, PFC, SF_6) are converted in units of

CO2 equivalent (CO2e), using the warming potential related to each gas. Climate change and global warming are internationally recognized as current issues, driving negative effects on humanity, and being mainly caused by GHG emissions generated both from industrial activities, and from other anthropogenic activities. Restoring the ecological balance requires urgent action to reduce GHG emissions. In this respect, the European Union has set the target to reduce the GHG emissions by 20% until 2020, compared to 1990 level. This paper presents a methodology to develop a model for carbon footprint calculation, for assessing and reducing GHG emissions generated by European funds financed projects.

b) Inference : The Paper recognizes Climate change and global warming as current issues with negative effects on humanity and the environment and highlights the urgency to address and mitigate the impacts of greenhouse gas (GHG) emissions. The paper proposed a methodology for developing a model to calculate carbon footprints. This model aims to assess and reduce GHG emissions generated by projects financed by European funds. Afforestation, the process of planting trees, is highlighted as a means to reduce the effects of GHG emissions. Trees convert carbon dioxide into oxygen and other organic compounds, helping to restore the ecological balance. The paper mentions the importance of technology and materials that generate fewer gases in reducing emissions. It also discusses the concept of compensating for emissions through the creation of absorption capacity for carbon emissions..The paper acknowledges the asymmetrical trajectory of GHG emissions in relation to economic growth. Emissions tend to increase at a higher rate than economic growth but decrease more slowly compared to economic decline.Two methods to combat the effects of GHG emissions are mentioned: reducing the level of emissions and utilizing flexible trading mechanisms in the carbon certificates market.

2.2 Patent search

2.2.1 System And Method For Managing And Forecasting Power From Renewable Energy Sources (US20150186904A1)
Inventor: Supratik GuhaHendrik F. HamannLevente I. KleinSergio A. Bermudez Rodriguez

The patent provides a comprehensive framework for managing and forecasting power from renewable energy sources, primarily focusing on solar and wind power. It begins by highlighting the increasing demand for renewable energy due to rising costs and environmental concerns associated with traditional energy sources like coal, oil, and natural gas. Despite the advantages of renewable energy, such as being environmentally friendly, its intermittent nature poses challenges for reliable power supply. To address these challenges, the patent introduces novel methods and systems for managing power from renewable sources.

One aspect of the invention involves a computer-implemented method that creates a list of tasks to be performed within a specific timeframe. Each task is associated with a power load, and task performance is prioritized based on the power load and the availability of power from the renewable energy source during the timeframe. Additionally, the patent describes a system designed for managing power use in buildings equipped with appliances powered partially or entirely by renewable energy sources. This system includes sensors associated with each appliance and a controller that receives data from these sensors. The controller utilizes the data to create a task list for the appliances within a given timeframe, associates a power load with each task, and prioritizes task performance based on the power load and the availability of power from renewable sources. Furthermore, the patent discusses the potential applications of these techniques in various contexts, such as residential, commercial, and industrial settings.

2.2.2 Method and system for predicting solar energy production (US20050039787A1)
Inventor: James Bing

The patent is a system, method and computer program product to assist in managing the physical plant mechanisms and market finances for a deregulated electricity grid or regulated utility grid, populated with solar electric generation capacity. This system provides tools to assist grid operators in the scheduling and dispatch of generation resources in an electrical grid populated with solar electric generation capacity, a week in advance, on an hourly basis. It also provides tools to assist companies engaged in generation, distribution, and energy marketing, in the electrical power industry, to manage their contractual supply obligations in the day-ahead hourly wholesale market and the spot market, in an electrical grid populated with solar electric generation capacity. This process can also be used to predict solar loading of building structures, using forecast irradiance data as inputs to common building energy modeling programs, a week in advance, on an hourly basis.

2.3 Comparison with the existing system

1. Existing systems primarily focus on historical data, lacking the ability to forecast future trends or propose solutions. In contrast, the proposed system not only estimates carbon footprint levels but also utilizes predictive modeling to forecast future trends and recommends afforestation strategies for each region to mitigate projected emissions.

2. The proposed system incorporates diverse fields like data science, machine learning, and environmental science to deliver a more comprehensive analysis which may be lacking in existing systems .

3. Scalability challenges may hinder the expansion of existing systems to cover larger areas or accommodate increasing volumes of data.

4. Issues related to data quality, such as inaccuracies, missing values, or outdated information, can compromise the reliability of existing systems.

5. Unlike existing systems that may focus on only solar energy, the proposed system enables dynamic comparisons between renewable energy sources, allowing for more informed and balanced decision-making.

6. By covering major Indian cities and offering insights into solar power generation potential, the proposed system provides a more comprehensive and holistic approach to energy planning compared to existing systems that may have limited geographical coverage .

Chapter 3: Requirement Gathering for the Proposed System

In this chapter we are going to discuss the resources we have used and how we analyzed what the user actually needs and what we can provide. We will also discuss the functional and non-functional requirements and finally the software and hardware used.

3.1 Introduction to requirement gathering

The Requirement Gathering is a process of requirements discovery or generating list of requirements or collecting as many requirements as possible by end users. It is also called as requirements elicitation or requirement capture.

3.2 Functional Requirements

1) Data Collection and Management:

● Climate and Weather Data: Historical climate and weather data to assess how these factors influence carbon dynamics.

● Land Use and Land Cover Data: Information on land use changes, deforestation, and land cover transitions over time.

2) Data Preprocessing and Cleaning:

● Tools and software for data cleaning, standardization, and integration to ensure data accuracy and consistency.

3) Predictive Analytics and Machine Learning:

● Machine learning algorithms and frameworks for developing predictive models based on historical data.

● Tools for feature selection, model training, validation, and evaluation.

4) Visualization and Reporting:

● Data visualization tools to create informative maps, graphs, and charts.

● Reporting tools to generate comprehensive reports and summaries of findings.

5) Scalability and Performance:

● Ensuring that the tools and software used can handle large datasets and perform calculations efficiently.

3.3 Non-Functional Requirements

1) Performance:

● Response Time: The system should provide timely responses for data processing, analysis, and reporting, even when handling large datasets.

● Scalability: The project should be scalable to accommodate growing datasets and evolving research needs.

● Resource Efficiency: Minimize resource consumption (CPU, memory, storage) to ensure efficient performance.

2) Reliability:

● Availability: The system should be available for use consistently, with minimal downtime for maintenance or updates.

● Data Integrity: Ensure that data remains accurate and intact throughout the project lifecycle.

3) Usability and User Experience:

● User-Friendly Interface: Design an intuitive and user-friendly interface for ease of use.

● Accessibility: Ensure that the project is accessible to users with disabilities, following accessibility standards.

● Training and Documentation: Provide user training and comprehensive documentation to assist users in utilizing the system effectively.

 4) Scalability and Flexibility:

● Adaptability: The system should be adaptable to changing research requirements and evolving technologies.

● Interoperability: Ensure that the project can integrate with external systems or data sources as needed.

3.4 Hardware & Software Requirements

3.4.1 Hardware Requirements:-

a. Minimum 8 GB RAM

b. Core I5 7th Gen processor

c. NVIDIA GPU

d. Disk space of 4GB

3.4.2 Software Requirements:-

   a. Python

   b. Google colab

   c. Python Basic Stack:

   Numpy

   Matplotlib

  Tensorflow

   Pandas

  Transformer

3.4.3 Techniques -

- **Python:-** Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation.

3.4.4 Tools Required:-

- **Vscode:-** Visual Studio Code is a streamlined code editor with support for development operations like debugging, task running, and version control. It aims to provide just the tools a developer needs for a quick code-build-debug cycle and leaves more complex workflows to fuller featured IDEs, such as Visual Studio IDE.

- **Google Colab:-** Colaboratory, or "Colab" for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs.

3.5 Constraints

- Data Availability and Quality: Limited access to up-to-date and high-quality data on forest cover, carbon emissions, and related factors can hinder the accuracy of your analysis.Incomplete or inconsistent data from various sources may introduce biases into your results

- Data Resolution and Spatial Variability: Data at fine spatial resolutions may not be readily available or may require significant computational resources to process. Differences in data resolution can lead to challenges in integrating and comparing datasets.

- **Modeling Complexity**: Developing accurate predictive models for carbon footprint and forest cover involves dealing with complex environmental systems, which may require advanced modeling techniques and computational power.

- **Resource Constraints**: Limited access to computational resources, such as high-performance computing clusters or cloud computing resources, may restrict the scale and speed of your analysis.

- **Model Validation and Uncertainty**: Validating predictive models for carbon footprint and forest cover can be challenging, and uncertainties in predictions should be carefully addressed and communicated.

Chapter 4: Proposed Design

4.1 Block diagram of the system

The block diagram as seen in Fig. 4.1 offers a detailed breakdown of the system's design, elucidating the working modules. This concise presentation aims to offer a clear understanding of the project's structure.
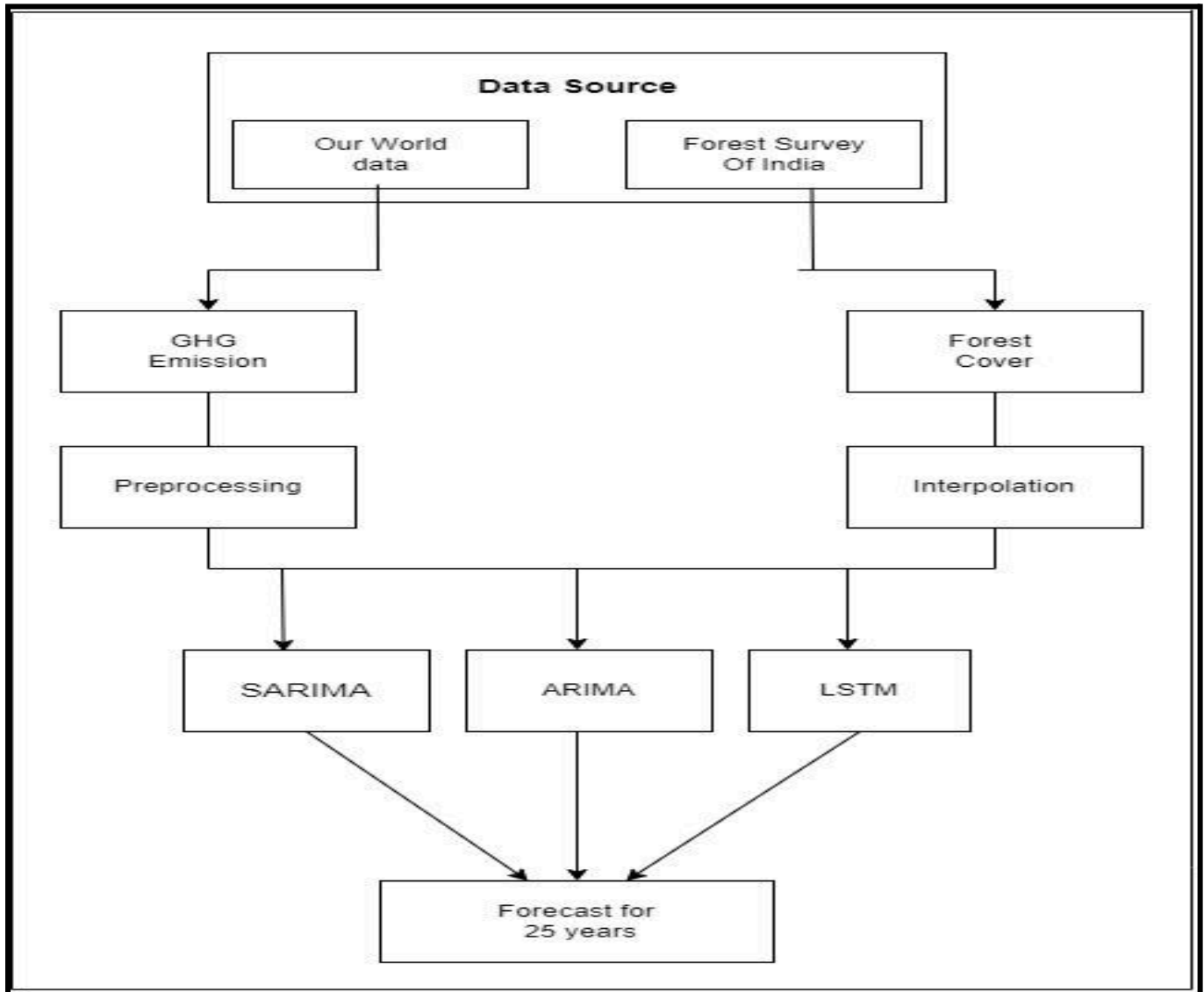
Fig 4.1: Block Diagram

A block diagram , comprising two primary data sources: Our World Data and the Forest Survey of India. The data from Our World Data is directed towards greenhouse gas (GHG) emissions analysis, undergoing preprocessing steps, while the data sourced from the Forest Survey of India is directed towards forest cover analysis, undergoing interpolation procedures. Subsequently, three distinct forecasting models—SARIMA, ARIMA, and LSTM—are employed to predict trends for a 25-year period. This structured approach enables the synthesis of insights regarding the relationship between forest cover and carbon footprint in India, facilitating informed decision-making and policy formulation in environmental management.

Data Sources: Two main sources of data are utilized in the project: Our World Data and the Forest Survey of India. Our World Data provides information related to greenhouse gas emissions, while the Forest Survey of India offers data concerning forest cover in India.

Data Analysis Flow: The data from Our World Data undergoes preprocessing steps before being utilized for greenhouse gas emissions analysis. Similarly, the data sourced from the Forest Survey of India undergoes interpolation procedures to analyze forest cover.

Forecasting Models: Three distinct forecasting models—Seasonal Autoregressive Integrated Moving Average (SARIMA), Autoregressive Integrated Moving Average (ARIMA), and Long Short-Term Memory (LSTM)—are employed to predict trends in carbon footprint and forest cover for a period of 25 years.

Objective: The overarching goal of the project is to conduct predictive analysis regarding the carbon footprint associated with forest cover in India. By employing various forecasting models, the project aims to provide insights into the future trends of carbon footprint and forest cover, aiding in environmental planning and decision-making processes.
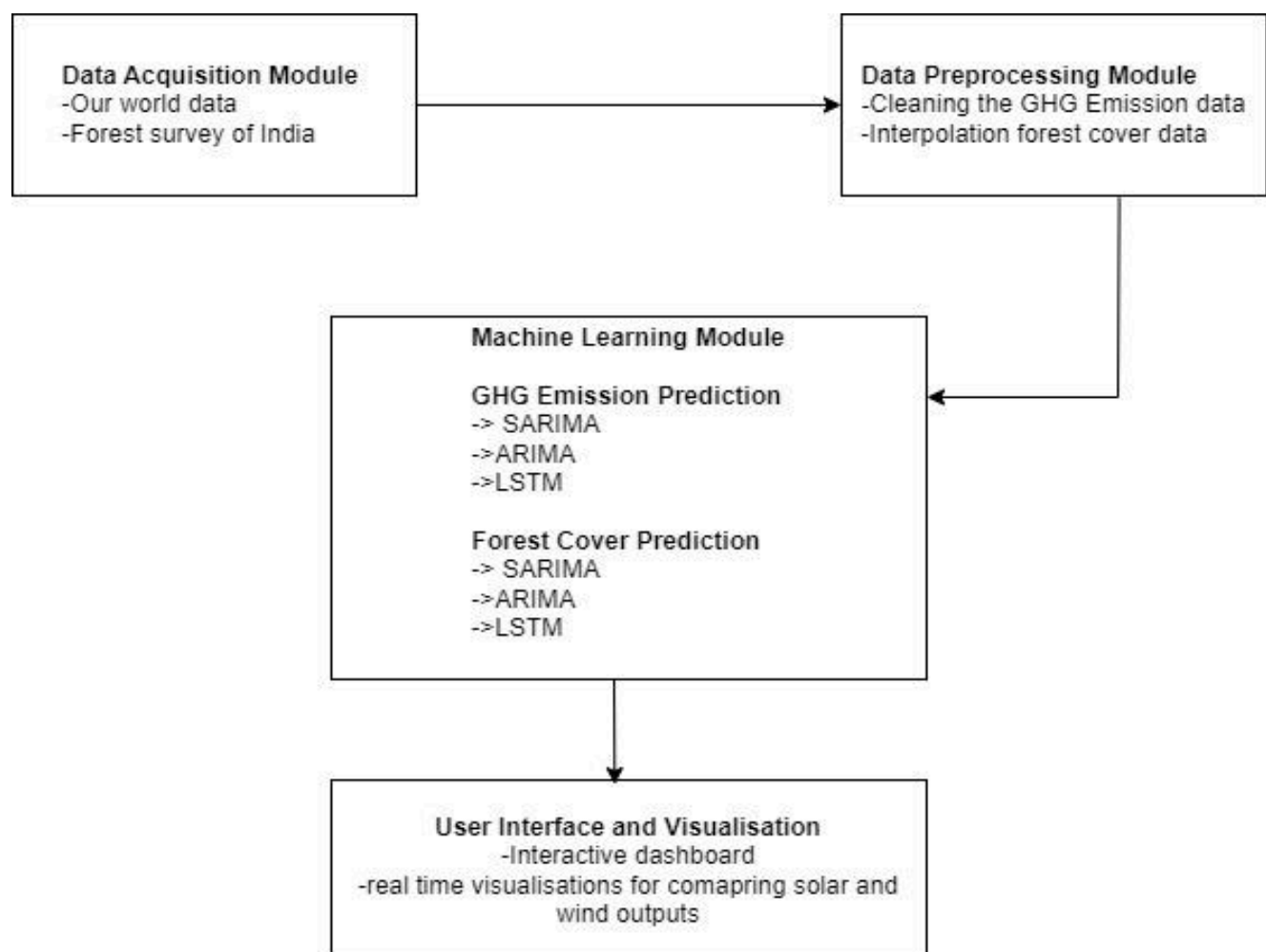
4.2 Modular design of the system

The modular diagram as seen in Fig 4.2 consists of various modules which are explained below.

Data Acquisition: The system collects real-time environmental data from sources like the Our world data. The data acquisition module utilized by the Forest Survey of India is designed to systematically gather, process, and analyze spatial and attribute data for comprehensive forest assessment and management.

Preprocessing: Data preprocessing involves cleaning, handling missing values, and interpolation forest cover data to prepare the data for analysis.

Machine Learning: The system employs machine learning models for GHG Emission prediction and Forest cover Prediction.For GHG Emission Prediction and forest cover Prediction, SARIMA , ARIMA , LSTM models utilized.

User Interface: A web application built using Python Dash Library.The Python Dash library is a framework for building interactive web applications. It works by allowing users to create web applications entirely in Python, leveraging libraries like Plotly for data visualization .

4.3 Project Scheduling & Tracking using Timeline / Gantt Chart

The gantt chart as seen in Fig. 4.3 visualizes the timeline of the different phases of the project, starting from defining the project scope to the deployment of the project.
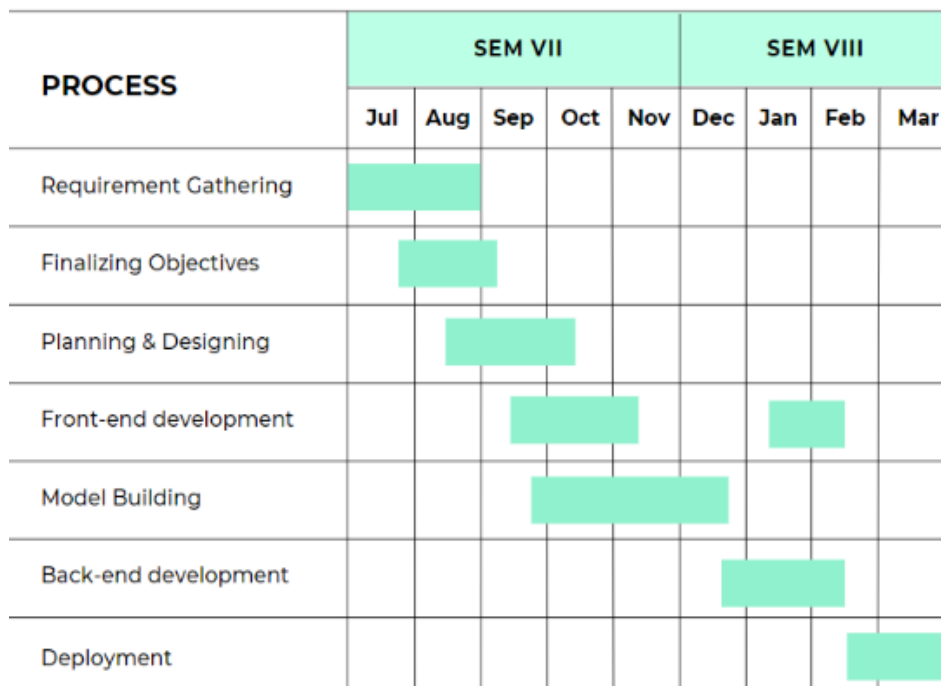

Fig 4.3 : Gantt Chart

Chapter 5: Implementation of the Proposed System

5.1. Methodology employed for development

5.1.1.Dataset Collection:

1. Greenhouse Gas (GHG) Emission in India:

   Data on greenhouse gas emissions in India from 1850 to 2021 was obtained from Our World in Data, an online publication that provides research and data on various global issues. The dataset includes comprehensive information on the emission trends of greenhouse gases such as carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O) over the specified period.

2. Forest Cover Data of India:

   The forest cover data for India spanning from 1975 to 2021 was sourced from the Forest Survey of India (FSI). The FSI periodically conducts assessments to monitor the extent and changes in forest cover across the country. The dataset comprises information on forest cover, including categories such as dense forest, open forest, and collected on alternate years.

3. Carbon Sequestration Data:

   Carbon sequestration data from 2011 to 2021 was acquired from the Forest Survey of India (FSI). This dataset provides insights into the amount of carbon dioxide absorbed by forests and other vegetation in India during the specified time frame, contributing to the mitigation of greenhouse gas emissions.

4. State-wise GHG Emission in India:

   The greenhouse gas emission data for each state in India from 2005 to 2015 was obtained from the Center for Study of Science, Technology and Policy (CSTEP), an independent think tank based in India. This dataset offers state-specific information on the emission of greenhouse gasses, aiding in the analysis of regional variations and contributing factors.

5. State-wise Solar Power Capacity:

   This dataset about the solar power capacity of each state was sourced from the Press Information Bureau, it depicts the solar power capacity of each state in MegaWatts(MW) over the past 5 years.

This dataset represents the funds allocated(in crores) to each state for solar power generation for the past 5 years. This data was fetched from the Press Information Bureau.

For the purpose of this project, a comprehensive dataset was gathered from various reputable sources

5.1.2 Data Cleaning:

In the process of extracting data from various resources, it was observed that while comprehensive information was available for most countries across different parameters, certain datasets, specifically those related to greenhouse gas (GHG) emissions and renewable energy, were not universally accessible. Consequently, a preprocessing step was necessary to ensure the inclusion of India's data spanning from 1850 to 2021, as per the scope of the study. This entailed meticulous filtering and extraction techniques to isolate and retain India-specific information, allowing for a focused analysis within the designated timeframe. Similarly, the renewable energy dataset underwent a similar treatment, ensuring that only data pertinent to India over the specified period was retained for further analysis and interpretation. This preprocessing phase was crucial in maintaining the integrity and relevance of the data, enabling a targeted examination of India's trajectory in GHG emissions and renewable energy adoption within the broader context of global trends. By consolidating and refining the datasets in this manner, the subsequent analysis can provide valuable insights into India's historical and contemporary contributions to climate change mitigation efforts and its transition towards sustainable energy practices.

5.1.3 Data Preprocessing:

To ensure continuity in the forest cover dataset spanning from 1975 to 2021, interpolation techniques were employed to estimate values for the intervening years where alternate-year data was available, thereby providing a comprehensive and continuous representation of forest cover trends over the specified period. Interpolation is a mathematical method used to estimate unknown values within a range of known data points. In this study, linear interpolation was utilized, which calculates the values between two known data points based on a straight line connecting them. The formula for linear interpolation is:

$$Y = Y_1 + (X - X_1)(Y_2 - Y_1)/(X_2 - X_1)$$

Where:
- Y represents the estimated forest cover value for the year of interest.

- Y_1 and Y_2 are the forest cover values for the years X_1 and X_2, respectively, where X_1 and X_2 are the years closest to the year of interest for which data is available.
- X represents the year of interest."

The data before interpolation:

Table 1. Data Before Interpolation

Year	GA	VDF	MDF	OF	TOTAL
1975	3287469	71584	233458	250138	555180
1981	3287469	52381	210345	200744	463470
1987	3287469	84847	280686	276508	642041
1989	3287469	84855	297870	257409	640134
1991	3287469	85017	304235	249930	639182
1993	3287469	85112	304720	250275	640107
1995	3287469	84533	305756	249311	639600
1997	3287469	74827	297260	261310	633397
1999	3287469	74811	293861	250306	637293
2001	3287469	67,044	3,49,765	2,58,729	6,75,538
2003	3287469	51,285	3,39,279	2,87,769	6,78,333
2005	3287469	83472	319948	286751	690171
2007	3287469	83510	319012	288377	690899
2009	3287469	83428	320238	288738	692394
2011	3287469	83,471	3,20,736	2,87,820	6,92,027
2013	3287469	83,502	3,18,745	2,95,651	6,97,898
2015	3287469	85904	315374	300395	701673
2017	3287469	98158	308138	301797	708093
2019	3287469	99278	308472	304499	712249
2021	3287469	99779	306890	307120	713789

The data after interpolation:

Table 2. Data after Interpolation

YEAR	GA	VDF	MDF	OF	TOTAL
1975	3287469	71584	233458	250138	555180
1976	3287469	70456	229501	239938	539895
1977	3287469	69234	220501	234875	524610
1978	3287469	68381	216200	224744	509325
1979	3287469	67381	210417	216242	494040
1980	3287469	66786	203716	208253	478755
1981	3287469	62381	200345	200744	463470
1982	3287469	68935	211865	212432	493232
1983	3287469	73934	224493	224567	522994
1984	3287469	75584	233434	243738	552756
1985	3287469	78746	246116	257655	582517
1986	3287469	82236	261078	268965	612279
1987	3287469	84847	280686	276508	642041
1988	3287469	84851	289278	266959	641088
1989	3287469	84855	297870	257409	640134
1990	3287469	84936	301053	253669	639658
1991	3287469	85017	304235	249930	639182

2011	3287469	83471	320736	287820	692027
2012	3287469	83487	319741	291735	694963
2013	3287469	83502	318745	295651	697898
2014	3287469	84703	317060	298023	699786
2015	3287469	85904	315374	300395	701673
2016	3287469	92031	311756	301096	704883
2017	3287469	98158	308138	301797	708093
2018	3287469	98718	308305	303148	710171
2019	3287469	99278	308472	304499	712249
2020	3287469	99529	307681	305810	713019
2021	3287469	99779	306890	307120	713789

5.1.4 Model selection

The impact of carbon footprint on forest cover in India, the selection of ARIMA, SARIMA, and LSTM models was based on their specific strengths in analyzing time series data.

ARIMA :

The ARIMA model was chosen for its ability to capture the temporal dependencies and trends in the data. It is well-suited for understanding the dynamics of carbon footprint and forest cover changes over time. The simplicity of the ARIMA model makes it easy to interpret and implement, while still being able to capture both short-term and long-term patterns in the data. However, ARIMA does assume that the underlying data is stationary, which may not always hold true for complex time series data.

SARIMA :

To address the seasonality present in environmental datasets like carbon footprint and forest cover, the SARIMA model was also utilized. SARIMA extends ARIMA to handle seasonal patterns in the data, making it more suitable for modeling the complex seasonal variations in carbon footprint and forest cover data. By capturing both the seasonal and non-seasonal components of the data, SARIMA provides a more comprehensive analysis of the patterns and trends in the data.

LSTM :

Lastly, the LSTM model was employed for its ability to capture long-term dependencies in the data. As a type of recurrent neural network (RNN), LSTM is well-suited for learning patterns in sequential data, such as time series data. This makes LSTM particularly useful for understanding the relationship between carbon footprint and forest cover over time, as it can automatically learn relevant features from the data. However, LSTM models can be computationally expensive to train and require careful tuning of hyperparameters.

5.1.5.Model Building:

We have used 3 methods for model evaluation to predict and forecast the values which are:
Direct method,Addition method and Linear method for yearly data.

Direct method :

The Direct method involved a straightforward approach of directly inputting the target value, Total_ghg_emission, into the model for prediction and forecasting. This method streamlined the process by using the target variable without any intermediary steps, providing a clear and direct path to forecasting.

Addition Method:

In contrast, the Addition method took a more intricate approach by applying the model individually to the independent variables CO2, CH4, and NO. The predicted and forecasted values of these variables were then summed to determine the total greenhouse gas emissions. This method provided a detailed breakdown of the contributions of each independent variable to the overall greenhouse gas emissions, offering a more nuanced understanding of the data.

Linear Method :

Similarly, the Linear method employed linear regression to predict and forecast the values of CO2, CH4, and NO. These predicted values were then aggregated to obtain the total greenhouse gas emissions forecast. Interestingly, both the Linear and Addition methods yielded the same output, suggesting that both approaches were equally effective in forecasting total greenhouse gas emissions.

SARIMA - Direct / Addition

Seasonal Autoregressive Integrated Moving Average (SARIMA) model stands out as a robust methodology, particularly adept at handling datasets with distinct seasonal patterns. Key Components of SARIMA include

Seasonal Component (S): The seasonal aspect of SARIMA encompasses three fundamental elements:

Seasonal Autoregressive (SAR) Terms: These terms account for the relationship between the current observation and past observations from the same season.

Seasonal Differencing (I): This component addresses seasonality by differencing the series at the seasonal lag.

Seasonal Moving Average (SMA) Terms: SMA terms capture the relationship between the error term and past error terms at the seasonal lag.

Autoregressive Component (AR): SARIMA's autoregressive component explores the dependency between the current observation and previous observations at non-seasonal lags.

Moving Average Component (MA): The moving average component models the error term as a linear combination of past error terms at non-seasonal lags.

The SARIMA model is denoted as SARIMA(p,d,q)(P,D,Q)[S], where:

p, d, q are the non-seasonal ARIMA parameters.

P, D, Q are the seasonal ARIMA parameters.

S is the seasonal period.

ARIMA - Direct / ARIMAX- Addition

The Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model emerges as a powerful extension of the traditional ARIMA model, offering the ability to incorporate external variables, or "exogenous" variables, into the forecasting process. This unique feature makes ARIMAX particularly well-suited for predicting carbon emissions.

LSTM

An LSTM (Long Short-Term Memory) model was utilized for analyzing the impact of carbon footprint on forest cover in India. The LSTM model is a type of recurrent neural network (RNN) that is well-suited for learning patterns in sequential data, making it suitable for analyzing time series data such as carbon footprint and forest cover data.

The LSTM model was configured with 50 epochs and 50 units.

Epochs: An epoch refers to one complete pass through the entire training dataset. In this case, the LSTM model was trained over 50 epochs, meaning that the model went through the entire dataset 50 times during the training process.

Units: The units parameter in an LSTM model refers to the number of memory cells or neurons in the layer. Each unit is responsible for learning and remembering patterns in the input data. Having more units can allow the model to capture more complex patterns in the data, but it also increases the computational complexity of the model.

5.1.6.Model Evaluation:

 Multiple metrics were used to evaluate the performance of the models, including RMSE (Root Mean Square Error), MAE (Mean Absolute Error), R-squared (R2), and MSE (Mean Squared Error). These metrics provide different insights into how well the models are performing in terms of accuracy and precision.

RMSE: RMSE measures the average magnitude of the errors between predicted and actual values. It provides a sense of how spread out the errors are. A lower RMSE indicates better model performance.

MAE: MAE is similar to RMSE but calculates the average of the absolute errors. It gives a more direct interpretation of the average error magnitude. Like RMSE, a lower MAE indicates better model performance.

R-squared (R2): R-squared is a measure of how well the model fits the actual data. It represents the proportion of variance in the dependent variable (forest cover) that is predictable from the independent variable (carbon footprint). A higher R2 value indicates a better fit.

MSE: MSE calculates the average of the squared errors between predicted and actual values. It penalizes larger errors more than smaller ones. Like RMSE, a lower MSE indicates better model performance.

The model with the highest R-squared value was selected for further forecasting. This decision was likely made because R-squared provides a measure of how well the model explains the variance in the data, making it a good indicator of overall model performance. By selecting the model with the highest R-squared value, we chose the model that best captures the relationship between carbon footprint and forest cover in India, providing more accurate forecasts for future scenarios.

GHG Emission data -

Metrics Dictionary:
{**'Total_ghg_emission_direct'**: {**'MSE'**: 2.467685522629966e+16, **'MAE'**: 138668750.51080364, **'R2'**: 0.9391930609594895, **'RMSE'**: 157088685.86343086},

'Total_ghg_emission_addition': {**'MSE'**: 3.552970362405361e+16, **'MAE'**: 159426029.13032925, **'R2'**: 0.9124502493294727, 'RMSE': 188493245.56613058},

'lstm': {**'MSE'**: 2.3726318019291315e+17, **'MAE'**: 459414175.5659034, **'R2'**: 0.4153530665782684, 'RMSE': 487096684.64578277},

'Total_ghg_emission_arimax_exog': {'MSE': 2.908702916621613e-07, 'MAE': 0.000537895020984468, **'R2'**: 1.0, **'RMSE'**: 0.000539323920906686},

'Total_ghg_emission_arimax': {**'MSE'**: 2.628490096836622e+16, **'MAE'**: 105112464.01152399, **'R2'**: 0.9352306298265313, **'RMSE'**: 162126188.41003516}}

Here we choose the SARIMA Direct Method, we ignored the maximum R2 which is 1 in ARIMAX because it overfits the data.

Forest cover data:-

Metrics Dictionary:

{**'forest_cover_prediction_sarima_direct_SARIMA'**: {**'MSE'**: 170707096.93882734, **'MAE'**: 11571.53548611463, **'R2'**: -3.1075180807023486, **'RMSE'**: 13065.492602226192},

'forest_cover_prediction_sarima_SARIMA': {**'MSE'**: 745045436.4426779, **'MAE'**: 24738.470539017453, **'R2'**: -16.927125796238702, **'RMSE'**: 27295.520446451974},

'forest_cover_prediction_arimax_exog': {**'MSE'**: 11788706.400398275, **'MAE'**: 3189.18558041387, **'R2'**: 0.7163426386128545, **'RMSE'**: 3433.4685669739683},

'forest_cover_prediction_arimax': {**'MSE'**: 6212194.794118999, **'MAE'**: 2415.7131483803505, **'R2'**: 0.8505234820621859, **'RMSE'**: 2492.4274902429956}}

Here we choose the model with highest R2, which is ARIMA for predicting.

The carbon emitted in the year 2021 is 3.9 billion tonnes and if we don't take necessary steps to reduce it it will reach 5.18 billion tonnes. On the contrary the forest cover of India is 713789 sq km in the year 2021 and it will only reach 752254.17 sq km. The carbon sequestrated in the Indian forest is 7.204 billion tonnes in the year 2021. The annual gain over the last 10 years is 541 million tonnes. The carbon sequestrated in the year 2021 is 39.4 million tonnes. The ideal forest cover required to offset 3.9 billion tonnes of carbon_emission is 7012329.97 sq km.which is larger than the geographical area of India As the rate of increase in carbon emission is significantly more than the rate at which the carbon sequestration and forest cover is growing.

Table 3. Predicted Budget of Solar Power for Next 10 years

```
Predicted Budgets for the Next 10 Years:
      Predicted Budget
2023         986.590000
2024         858.530000
2025         730.470000
2026         602.410000
2027         474.350000
2028         346.290000
2029         218.230000
2030         106.408084
2031        -374.689984
2032        -498.878051
```

5.2 Algorithms and flowcharts for the respective modules

5.2.1 Long Short-Term Memory Model

The Long Short-Term Memory (LSTM) model is a type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem of traditional RNNs and capture long-range dependencies in sequential data. LSTM models are specifically designed for sequential data, such as time series or natural language text. They process input data one element at a time, maintaining an internal state that captures information from previous elements in the sequence. The key component of an LSTM model is the memory cell, which stores information over time and controls the flow of information through the network. Each memory cell contains three main components: an input gate, a forget gate, and an output gate. The input gate regulates the flow of new information into the memory cell, the forget gate controls the retention or deletion of information from the cell's internal state, and the output gate determines the information that is passed on to the next time step.

5.2.2 ARIMA: Uncovering Trends in Time Series Data

ARIMA, or Autoregressive Integrated Moving Average, stands as a cornerstone in time series analysis. Its strength lies in uncovering the underlying patterns within data collected over time. Imagine you have stock market data for the past few years. ARIMA can analyze this data, considering the influence of past stock prices (autoregressive component) and accounting for random fluctuations (moving average component). By incorporating these factors, ARIMA can generate forecasts about future stock prices. The magic lies in the "integrated" part. This refers to a process that transforms the data to ensure it exhibits constant mean and variance over time, a crucial requirement for ARIMA to function accurately.

Beyond simply generating forecasts, ARIMA offers valuable insights into the data's inherent structure. By analyzing the autoregressive and moving average components, you can understand how strongly past values influence future values and how much randomness is present in the data. This knowledge can be crucial for various applications, such as predicting sales trends, analyzing economic indicators, or even understanding weather patterns.

5.2.3 SARIMA: Addressing Seasonality in Forecasts

While ARIMA excels at uncovering trends, it has a blind spot: seasonality. Many time series exhibit recurring patterns based on time of year, month, or even holidays. SARIMA, or Seasonal ARIMA, addresses this limitation by incorporating seasonality explicitly into the model. Imagine monthly retail sales data with a significant spike in December due to holiday shopping. A standard ARIMA model might miss

this crucial pattern. SARIMA, however, can account for this seasonal trend, leading to more accurate forecasts compared to ARIMA, especially for data with strong seasonal components.

The beauty of SARIMA lies in its ability to capture the cyclical nature of the data. It not only considers the influence of past values like ARIMA but also factors in the influence of past seasonal values. This allows SARIMA to identify and account for recurring seasonal patterns, leading to more robust and informative forecasts. SARIMA proves particularly valuable for analyzing data in fields like tourism, agriculture, or any domain where seasonality plays a significant role.

5.3 Datasets source and utilization

To facilitate the analysis of key factors influencing carbon footprint and environmental sustainability in India. The following datasets were meticulously collected and curated:

1. Forest Cover Data (1975 - 2021):

The Forest Survey of India (FSI) served as the primary source for acquiring data on forest cover spanning from 1975 to 2021. This dataset provides detailed information on the extent and changes in forest cover across different regions of India over the specified period. Categories such as dense forest, open forest, and others were included to capture the diverse landscape of forest cover in the country.

2. Greenhouse Gas (GHG) Emission Data (1850 - 2021):

Our World in Data, a reputable online publication, was the source of greenhouse gas emission data covering the period from 1850 to 2021. This dataset offers comprehensive insights into the historical trends and patterns of major greenhouse gasses such as carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O) in India. The data enables a thorough examination of the factors contributing to GHG emissions over time.

3. Renewable Energy Distribution Data:

Our World in Data also provided data on renewable energy distribution in India. This dataset encompasses information on the adoption and distribution of renewable energy sources such as solar, wind, and hydroelectric power across different regions of India. Analyzing this data facilitates the assessment of India's progress towards sustainable energy practices and its potential impact on reducing carbon emissions.

By integrating these diverse datasets, our analysis aims to uncover meaningful insights into the interplay between forest cover dynamics, greenhouse gas emissions, and the adoption of renewable energy solutions in India. This comprehensive approach allows for a holistic understanding of the challenges and opportunities in mitigating carbon footprint and promoting environmental conservation efforts in the country**.**

Chapter 6: Testing of the Proposed System

This chapter focuses on testing the functionality, reliability, and usability of the developed system. It aims to validate the system's performance and identify any issues for improvement, ensuring its effectiveness in addressing stakeholders' needs and establishing its credibility and reliability.

6.1 Introduction to testing :

Testing is an essential phase in the software development lifecycle (SDLC) that ensures the reliability, functionality, and quality of the developed software. It involves systematically executing the software components or system to identify any defects or errors that may affect its performance or user experience.

The primary objective of testing is to validate whether the software meets the specified requirements and behaves as expected under different conditions. In the context of our project, testing plays a crucial role in verifying the correctness and robustness of the software solution we have developed. By subjecting the software to various test scenarios, we aim to uncover any discrepancies between the actual and expected behavior, enabling us to rectify issues and enhance the overall quality of the product.

The testing process encompasses several phases, starting from the early stages of development and continuing throughout the project lifecycle. Furthermore, testing is not limited to functional aspects alone but also encompasses non-functional aspects such as performance, security, usability, and compatibility. These aspects are equally important in delivering a successful software solution that meets the needs and expectations of end-users.

In our project, we have adopted a comprehensive testing approach that combines both manual and automated testing techniques. Manual testing allows for human judgment and exploration of the software's behavior, while automated testing aids in repetitive and regression testing tasks, ensuring efficiency and consistency.

6.2 Types of tests considered :

In our project, we have adopted a comprehensive testing strategy that encompasses various types of tests to ensure thorough validation of the software solution. Two fundamental approaches that we have employed are black box testing and white box testing.

<u>Black Box Testing:</u>

Black box testing, also known as functional testing, focuses on evaluating the software's functionality without considering its internal structure or implementation details. Test cases are designed based on the system's specifications and requirements, treating the software as a black box. During black box testing, testers interact with the software through its interfaces, inputs, and outputs to validate whether it behaves as expected under different conditions. This approach helps identify issues related to incorrect or missing functionality, user interface flaws, and integration problems, ensuring that the software meets the end-users' requirements and expectations.

<u>White Box Testing:</u>

White box testing, in contrast to black box testing, examines the internal structure and logic of the software system. Also referred to as structural testing or glass box testing, this approach involves analyzing the source code, design documents, and architecture to devise test cases that exercise specific paths, conditions, and branches within the code. White box testing aims to uncover defects related to logic errors, coding mistakes, and performance bottlenecks that may not be apparent through black box testing alone. By gaining insight into the internal workings of the software, white box testing enables us to verify the correctness of individual components, ensure code coverage, and optimize the software's efficiency and maintainability.

By incorporating both black box and white box testing techniques into our testing strategy, we aim to achieve comprehensive test coverage and maximize the detection of defects across different layers of the software. This hybrid approach allows us to address functional and structural aspects effectively, resulting in a more robust and reliable software solution.

6.3 Various test case scenarios considered :

<u>Black box testing</u>

1. **User Interface Testing:**

   Ensure all user interface elements are functional, including buttons, input fields, dropdown menus, and navigation links.

   Verify that the user interface is responsive and displays correctly on different devices and screen sizes.

   Validate the consistency of design elements, typography, colors, and layout across all pages of the application.

2. **Real-Time Data Accuracy Testing:**

Collect real-time data from external sources or APIs and compare it with the data displayed in the system.

Verify that the real-time data updates correctly and reflects the latest information available from the source.

Test the system's ability to handle fluctuations in real-time data and ensure that it updates consistently and accurately.

Validate the timestamps associated with real-time data to ensure they match the current time and date.

3.Forecast Testing:

Input historical data into the system and generate forecasts for future time periods.

Compare the forecasted values with actual values observed after the forecast period to assess accuracy.

Test the system's forecasting algorithms under various scenarios and conditions to evaluate performance.

Verify that the system provides reliable and timely forecasts that align with expected trends and patterns.

4. Visualization Testing:

Verify that data visualizations, such as charts, graphs, and maps, are generated correctly and accurately represent the underlying data.

Test the interactivity of visualizations, such as zooming, panning, and filtering, to ensure they respond smoothly to user interactions.

Validate the accessibility of visualizations for users with disabilities, such as providing alternative text for images and supporting screen reader navigation.

Ensure that visualizations are displayed consistently across different browsers and devices and that they maintain clarity and legibility at various screen resolutions.

White Box Testing:

1. **Unit Testing:**

Writing unit tests for individual functions or modules responsible for handling solar and wind energy prediction. Testing different scenarios such as valid input, invalid input, and edge cases to ensure robustness.

2. **Integration Testing:**

Testing the integration between frontend and backend components. Verifying that data is passed correctly between different layers of the application.

3. **Performance Testing:**

   Measuring the performance of the application when handling real-time data uploads and processing for solar and wind energy forecasting.

6.4 Inference drawn from the test cases :

User Interface Testing:

The user interface is intuitive, easy to navigate, and responsive across different devices and screen sizes. The proposed system provides a user-friendly interface that enhances user experience and usability.

Real-time Data Accuracy Testing:

Real-time data displayed on the system matches the actual data obtained from external sources with minimal latency. The system effectively retrieves, processes, and presents real-time data, ensuring accuracy and timeliness for decision-making.

Forecast Testing:

Forecasted energy generation trends closely align with historical data and exhibit reasonable predictions for future periods. The forecasting algorithms employed in the system demonstrate reliability and effectiveness in predicting energy generation patterns.

Visualization Testing:

Data visualizations, such as charts, graphs, and maps, are visually appealing, informative, and convey insights effectively. The visualization components enhance data understanding and facilitate data-driven decision-making for energy planning and management.

Unit Testing:

Individual components and functions of the system behave as expected and produce correct outputs for various input conditions. The core functionalities of the system are implemented correctly and exhibit the desired behavior in isolation.

Integration Testing:

Different modules and components of the system integrate seamlessly and exchange data accurately without compatibility issues. The system architecture is well-designed, and the integration points function correctly, ensuring proper communication between system elements.

Performance Testing:

The system maintains acceptable performance levels under different load conditions, with response times and throughput meeting performance criteria. The system is capable of handling expected user loads and provides satisfactory performance in terms of speed and scalability.
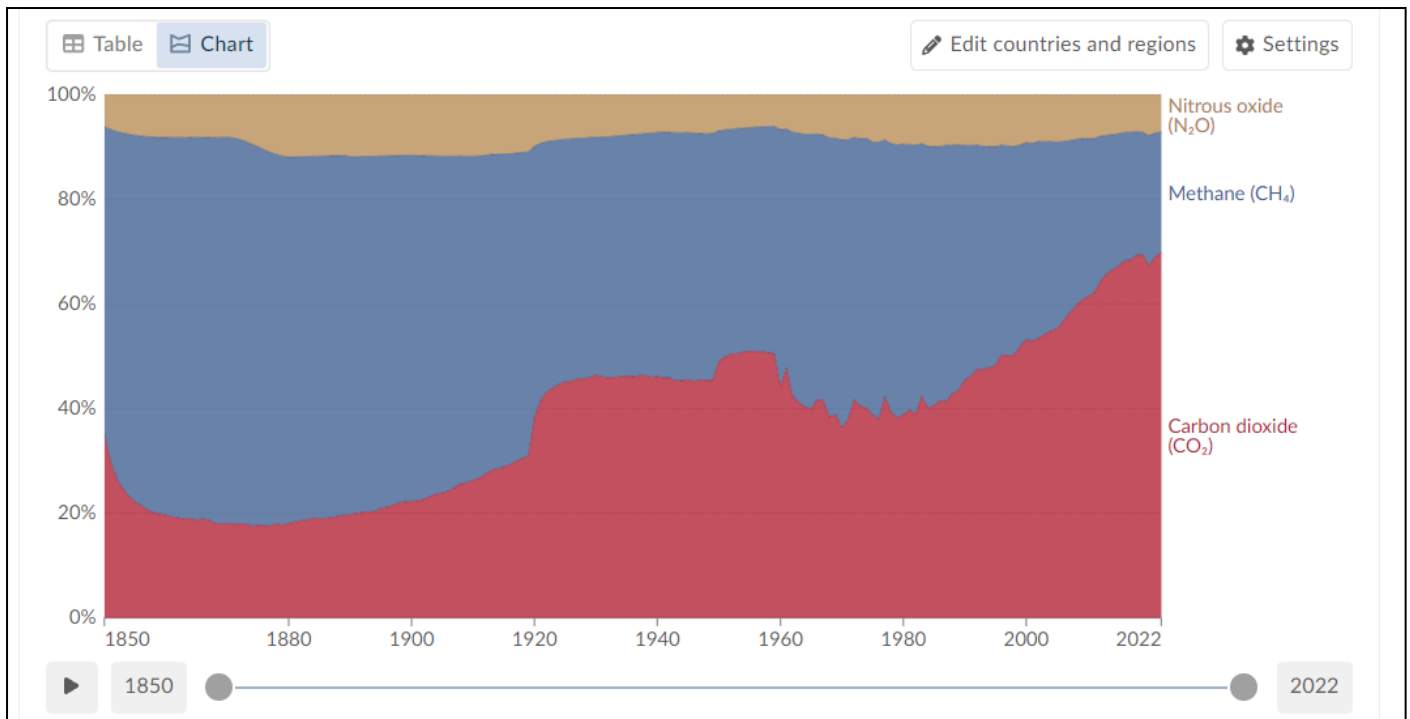
Chapter 7: Results and Discussion

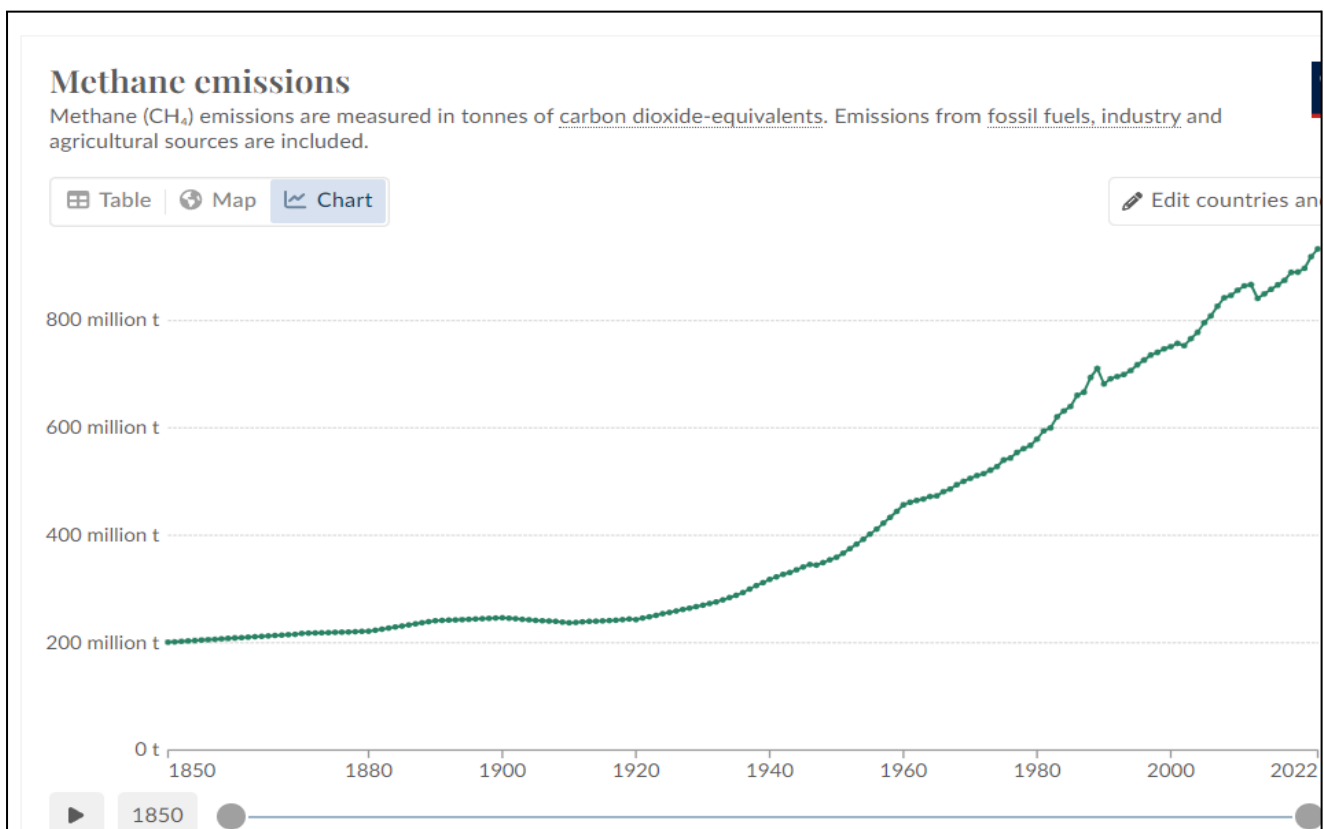## 7.1 Screenshots of User Interface (UI):
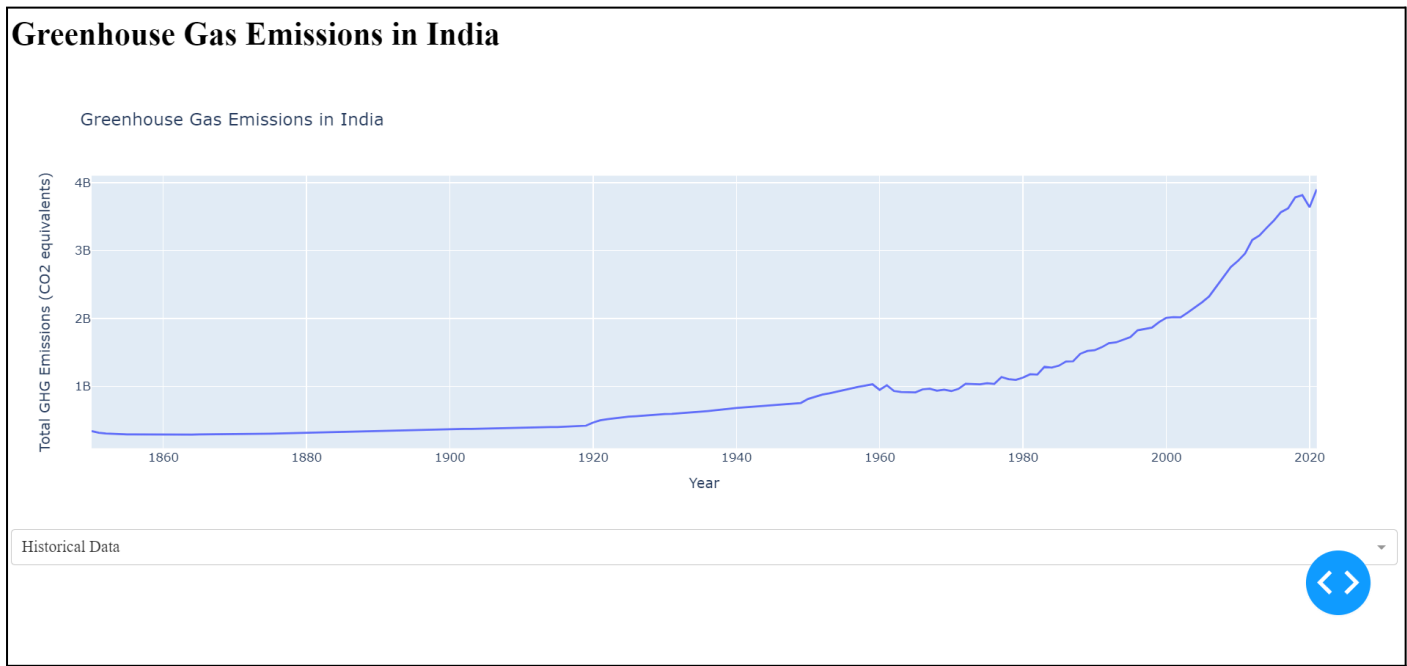
7.1.1.Greenhouse Gas Emissions in India
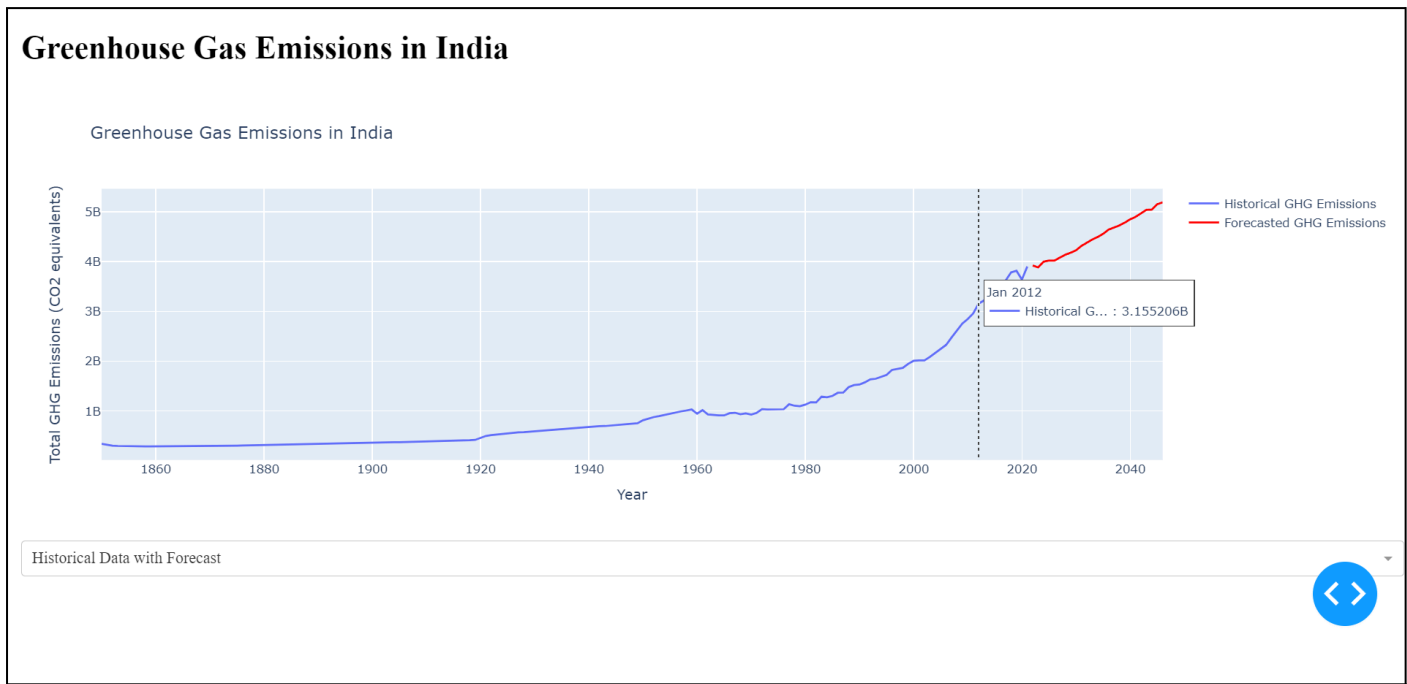
7.1.2 Greenhouse Gas Emissions with forecasting data

7.1.3.Distribution of GHG in india

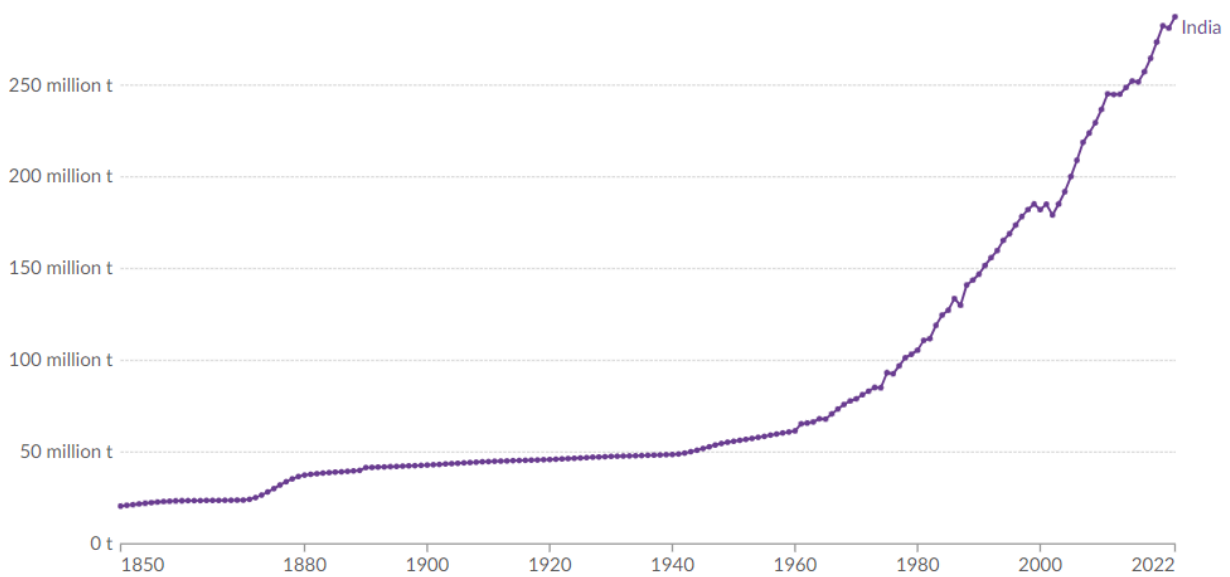7.1.4. Annual methane emission in india from 1850 to 2021

7.1.5. Annual Nitrous Oxide emissions in india from 1850 to 2021

7.1.6. Annual Carbon dioxide emission in india from 1850 to 2021

7.2. Performance Evaluation measures

1. **Mean Squared Error (MSE)**

   MSE measures the average squared difference between the actual and predicted values.

   It penalizes large errors more than smaller ones.

   Lower MSE indicates better model performance, with zero representing a perfect fit.

   $MSE = \Sigma(y_i - p_i)^2 / n$

2. **Mean Absolute Error (MAE):**

   MAE measures the average absolute difference between the actual and predicted values.

   It provides a more intuitive understanding of error magnitude compared to MSE.

   Lower MAE indicates better model performance.

   $$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right|$$

3. **R-squared (R2):**

   R2 represents the proportion of variance in the dependent variable (target) that is explained by the independent variables (features) in the model.

   It ranges from 0 to 1, where 1 indicates a perfect fit.

   Higher R2 values indicate better model performance in explaining the variance.

   $$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

4. **Root Mean Squared Error:**

   The root mean square error (RMSE) formula is a measure of the differences between values predicted by a model or an estimator and the values actually observed. It is commonly used in various fields such as statistics, machine learning, and signal processing to evaluate the accuracy of a model's predictions.

   The formula for RMSE is as follows**:**

7.3. Input Parameters / Features considered

GHG emission dataset:

- Annual nitrous oxide emissions in CO_2 equivalents:nNO emission from 1850 to 2021.
- Annual methane emissions in CO_2 equivalents: CH4 emission from 1850 to 2021
- Annual CO_2 emissions: CO2 emission from 1850 to 2021

Forest cover dataset:

- Very Dense Forest: Tree cover exceeding 70% density.
- Moderately Dense Forest: Tree cover between 40% and 70% density.
- Open forest : Tree cover between 10% and 40% density

State-wise Solar Power Budget dataset:

- Funds allocated by Government of India in Crores
- Data is divided state-wise in fiscal years from 2018 to 2022

7.4. Graphical and statistical output:

7.4.1. GHG emission:

Table 4. Comparison table for GHG emission in india for (2001 to 2021)

| Year	original_emission	Total_ghg_emission_direct	Total_ghg_emission_linear	Total_ghg_emission_addition	Total_ghg_emission_arimax_exog	Total_ghg_emission_arimax
2001	2019021660	2.090044e+09	2.102618e+09	2.102618e+09	2.019022e+09	2.082245e+09
2002	2017562400	2.144205e+09	2.135139e+09	2.135139e+09	2.017562e+09	2.149533e+09
2003	2088233880	2.188557e+09	2.166190e+09	2.166190e+09	2.088234e+09	2.215228e+09
2004	2162795810	2.279903e+09	2.240605e+09	2.240605e+09	2.162796e+09	2.298520e+09
2005	2239350740	2.332360e+09	2.313255e+09	2.313255e+09	2.239351e+09	2.374339e+09
2006	2327488030	2.397872e+09	2.366712e+09	2.366712e+09	2.327488e+09	2.454412e+09
2007	2475741210	2.489791e+09	2.476096e+09	2.476096e+09	2.475741e+09	2.547122e+09
2008	2623861870	2.581686e+09	2.548739e+09	2.548739e+09	2.623862e+09	2.645661e+09
2009	2756778930	2.655058e+09	2.627465e+09	2.627465e+09	2.756779e+09	2.751006e+09
2010	2847361420	2.722331e+09	2.708652e+09	2.708652e+09	2.847361e+09	2.851172e+09
2011	2957348500	2.814150e+09	2.788263e+09	2.788263e+09	2.957348e+09	2.954544e+09
2012	3155205640	2.918987e+09	2.890869e+09	2.890869e+09	3.155206e+09	3.062352e+09
2013	3220519650	3.028120e+09	3.007014e+09	3.007014e+09	3.220520e+09	3.178170e+09
2014	3333282800	3.117678e+09	3.066173e+09	3.066173e+09	3.333283e+09	3.304699e+09
2015	3438725700	3.204190e+09	3.149638e+09	3.149638e+09	3.438726e+09	3.434132e+09
2016	3564861130	3.321146e+09	3.252332e+09	3.252332e+09	3.564861e+09	3.570347e+09
2017	3621199130	3.418503e+09	3.347016e+09	3.347016e+09	3.621199e+09	3.705819e+09
2018	3783796140	3.519806e+09	3.438700e+09	3.438700e+09	3.783796e+09	3.842072e+09
2019	3815779680	3.649375e+09	3.569540e+09	3.569540e+09	3.815780e+09	3.986561e+09
2020	3638071190	3.775579e+09	3.689244e+09	3.689244e+09	3.638071e+09	4.135568e+09
2021	3900927010	3.886618e+09	3.798895e+09	3.798895e+09	3.900927e+09	4.297858e+09

Metrics Dictionary:

{**'Total_ghg_emission_direct'**: {**'MSE'**: 2.467685522629966e+16, **'MAE'**: 138668750.51080364, **'R2'**: 0.9391930609594895, **'RMSE'**: 157088685.86343086},

'Total_ghg_emission_addition': {**'MSE'**: 3.552970362405361e+16, **'MAE'**: 159426029.13032925, **'R2'**: 0.9124502493294727, 'RMSE': 188493245.56613058},

'lstm': {**'MSE'**: 2.3726318019291315e+17, **'MAE'**: 459414175.5659034, **'R2'**: 0.4153530665782684, 'RMSE': 487096684.64578277},

'Total_ghg_emission_arimax_exog': {'MSE': 2.908702916621613e-07, 'MAE': 0.000537895020984468, **'R2'**: 1.0, **'RMSE'**: 0.000539323920906686},

'Total_ghg_emission_arimax': {**'MSE'**: 2.628490096836622e+16, **'MAE'**: 105112464.01152399, **'R2'**: 0.9352306298265313, **'RMSE'**: 162126188.41003516}}

Here's a breakdown of this dictionary:

- Different models are evaluated: There are entries for 'lstm', 'arimax', 'arimax_exog', and some variations related to emission types ('direct' and 'addition').
- Performance metrics are provided: Each model entry includes MSE (Mean Squared Error), MAE (Mean Absolute Error), R-squared (coefficient of determination), and RMSE (Root Mean Squared Error). These metrics offer insights into how well each model performs in predicting emissions.
- 'arimax_exog' stands out: This model achieves a perfect score (R-squared of 1.0 and very low MSE and MAE) which suggests it might be overfitting the data (meaning it performs well on the training data but may not generalize well to unseen data).
- Other models show trade-offs: 'arimax' and 'direct/addition' variations seem to have a good balance between R-squared (indicating fit) and error metrics (MSE, MAE, RMSE). 'lstm' has a lower R-squared but potentially lower errors, suggesting a different strength - it might capture more complex patterns but may not perfectly match the overall trend.
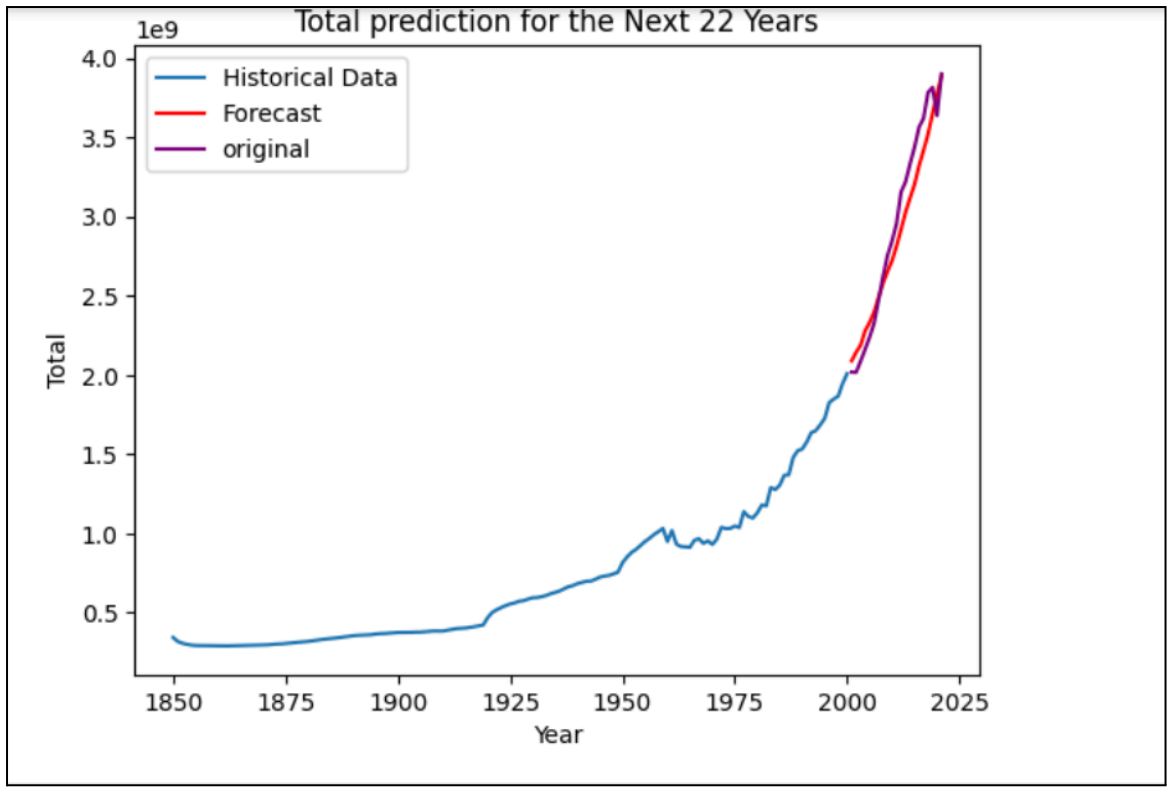
7.4.1.1 GHG emission line graph for prediction from 2001 to 2021

Based on the r2 score we have selected the Sarima model(direct approach)
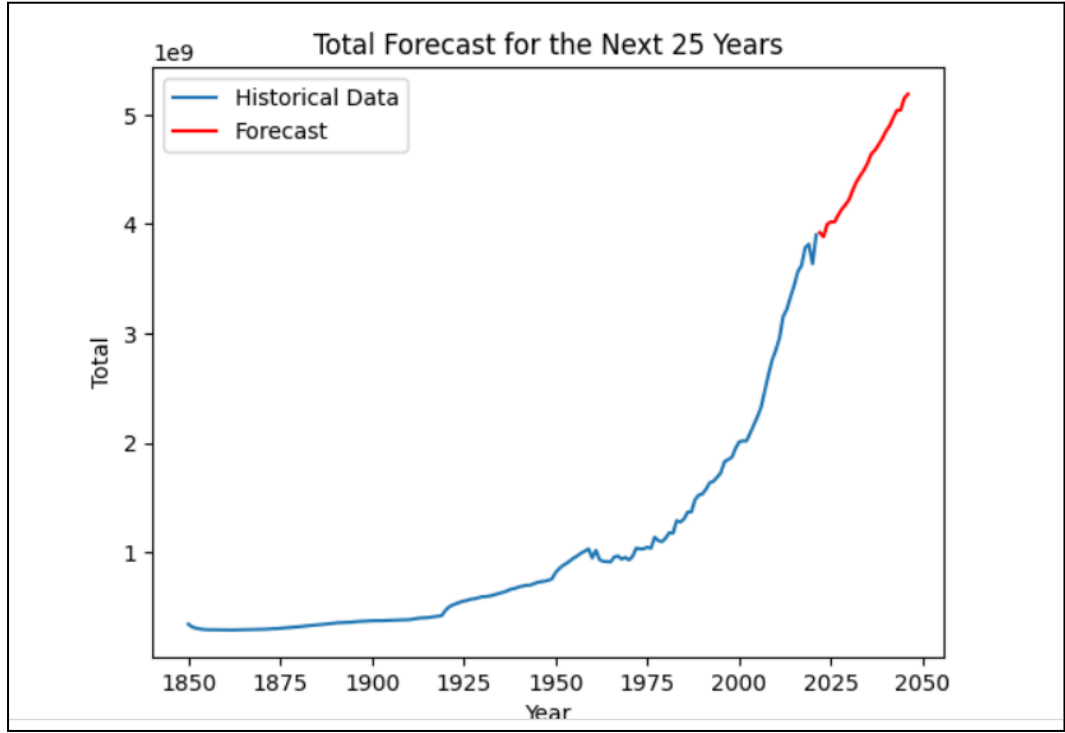
7.4.1.2GHG emission line graph for forecast of next 25 years

Based on the r2 score we have selected the Sarima model(direct approach)

7.4.2. Forest cover:

Table 5. Comparison table for forest cover in india(2012 to 2021)

| Year	forest_cover_prediction_sarima_direct	forest_cover_prediction_sarima_addition	forest_cover_prediction_arima	forest_cover_prediction_arima_exog	Total_original_forest
2012	706453.646236	718854.451285	692640.567451	693969.539355	694963
2013	720881.285471	745398.505691	693926.863326	695903.342608	697898
2014	716825.853442	723291.994013	695755.864303	697329.311004	699786
2015	712769.845237	739285.653923	698011.277156	698751.769196	701673
2016	713377.906936	737265.176953	700601.728484	702413.755466	704883
2017	713986.032734	735359.480904	703455.424776	706074.859155	708093
2018	709280.096190	731338.563120	706515.956748	707818.193863	710171
2019	704574.234601	727495.545727	709739.002747	709561.224664	712249
2020	700410.842642	724239.742651	713089.738554	710036.001238	713019
2021	696246.041762	721379.591124	716540.802240	710508.871965	713789

Metrics Dictionary:

{**'forest_cover_prediction_sarima_direct_SARIMA'**: {**'MSE'**: 170707096.93882734, **'MAE'**: 11571.53548611463, **'R2'**: -3.1075180807023486, **'RMSE'**: 13065.492602226192},

'forest_cover_prediction_sarima_SARIMA': {**'MSE'**: 745045436.4426779, **'MAE'**: 24738.470539017453, **'R2'**: -16.927125796238702, **'RMSE'**: 27295.520446451974},

'forest_cover_prediction_arimax_exog': {**'MSE'**: 11788706.400398275, **'MAE'**: 3189.18558041387, **'R2'**: 0.7163426386128545, **'RMSE'**: 3433.4685669739683},

'forest_cover_prediction_arimax': {**'MSE'**: 6212194.794118999, **'MAE'**: 2415.7131483803505, **'R2'**: 0.8505234820621859, **'RMSE'**: 2492.4274902429956}}

Here's a breakdown of the above dictionary:

- There are four models being compared: 'sarima_direct_SARIMA', 'sarima_SARIMA', 'arimax_exog', and 'arimax'. Each likely refers to a specific statistical method used for forecasting.

- The models are named based on the prediction target ("forest_cover_prediction") and the method used. 'direct' might indicate a variation focused on a specific aspect of forest cover.

- Performance metrics are included: Each model entry contains MSE (Mean Squared Error), MAE (Mean Absolute Error), R-squared (coefficient of determination), and RMSE (Root Mean Squared Error). Lower error values (MSE, MAE, RMSE) and a higher R-squared value generally indicate better model performance.

- 'arimax' and 'arimax_exog' outperform 'sarima' models: Based on R-squared, 'arimax' and 'arimax_exog' seem to fit the data considerably better. Their error metrics (MSE, MAE, RMSE) are also lower.

- 'arimax_exog' might be the best option: While 'arimax' shows good results, 'arimax_exog' has the lowest overall error and the highest R-squared, suggesting it might be the most accurate predictor of forest cover in this case.

- Negative R-squared values for 'sarima' models: It's interesting to note that the 'sarima' models have negative R-squared values. This typically indicates the model performs worse than just predicting the average forest cover.
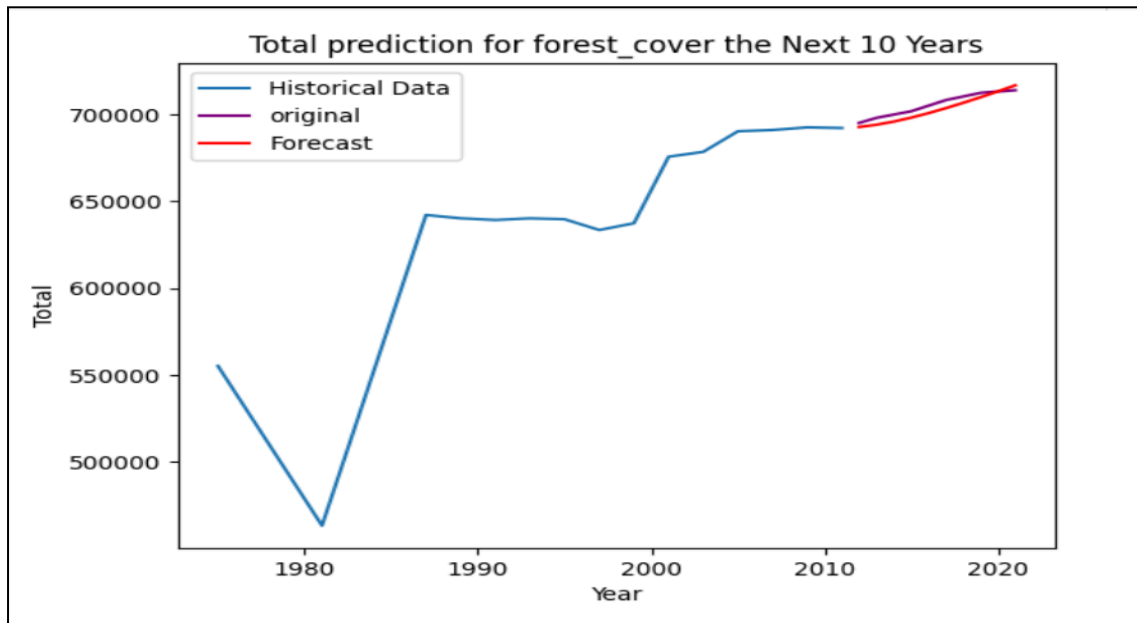
7.4.2.1Forest cover line graph for prediction from 2012 to 2021
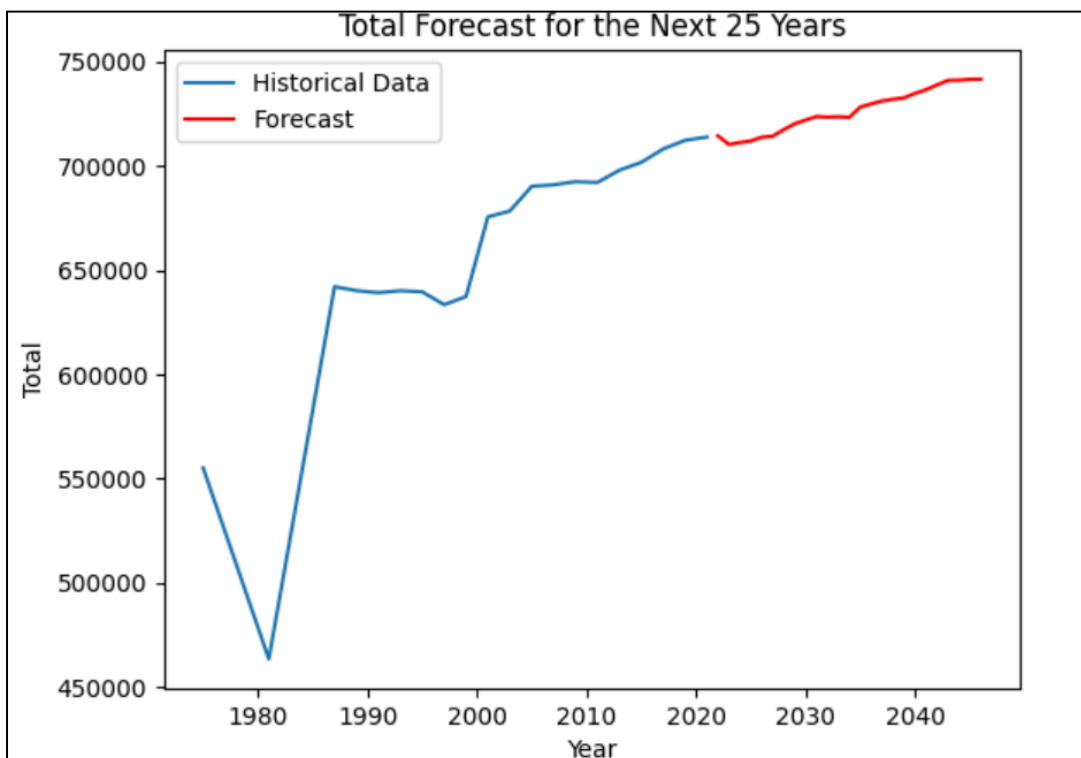
7.4.2.2 Forest cover forecast for next 25 years

Based on the r2 score we have selected the arima model.

7.5. Comparison of results with existing systems

Table 6. Comparison of Existing and proposed system

Existing System	Proposed System
Analyzes forest cover (may not consider impact factors)	Focuses on impact of carbon footprint on forest cover in India
Uses various methods (may not be time series specific)	Employs time series analysis (ARIMA, SARIMA)
Does not consider carbon footprint explicitly	Includes carbon footprint to assess impact on forest cover
May not account for seasonal variations	Explicitly considers seasonal variations in forest cover
May or may not generate forecasts	Generates forecasts considering carbon footprint and seasonality
Provides basic understanding of forest cover	Offers in-depth analysis of carbon footprint's impact on forest cover

7.6. Inference drawn

The issue of carbon emissions poses a significant challenge globally, with projections indicating a concerning trajectory if urgent action is not taken. In 2021 alone, carbon emissions reached a staggering 3.9 billion tonnes, with forecasts suggesting a potential increase to 5.18 billion tonnes if corrective measures are not implemented swiftly. This rapid rise in emissions far outpaces the rate at which nature's carbon sequestration mechanisms, particularly forests, can mitigate the impacts. India, for instance, possesses a substantial forest cover, amounting to 713,789 square kilometers in 2021, with projections indicating a modest increase to 752,254.17 square kilometers. Despite the significant carbon sequestration capacity of these forests, estimated at 7.204 billion tonnes in 2021, the rate of growth is not keeping pace with escalating emissions. Over the past decade, the annual gain in carbon sequestration has averaged 541 million tonnes, with 2021 seeing a sequestration of 39.4 million tonnes. However, to offset the 3.9 billion tonnes of carbon emissions, an ideal forest cover of 7,012,329.97 square kilometers would be required—a geographical area exceeding that of India itself. This stark disparity underscores the urgent need for comprehensive measures to curb carbon emissions and enhance carbon sequestration efforts to mitigate the looming threat of climate change.

In light of the challenges associated with rapidly expanding forest cover, alternative approaches to mitigate carbon emissions and address climate change have become imperative. One such strategy involves the widespread adoption of renewable energy sources, particularly solar energy. Solar energy presents a promising solution due to its abundance, sustainability, and scalability. Unlike fossil fuels, which contribute significantly to carbon emissions, solar energy harnesses the power of sunlight, converting it into electricity

through photovoltaic cells or concentrating solar power systems.The adoption of solar energy offers numerous advantages. First and foremost, it provides a clean and renewable source of electricity, emitting no greenhouse gases or harmful pollutants during operation. This significantly reduces the carbon footprint associated with energy production, thereby helping to mitigate climate change. Additionally, solar energy systems can be deployed across diverse geographical locations, making it accessible to both urban and rural communities. This decentralized approach to energy generation enhances energy security and resilience, reducing dependence on centralized power grids.Furthermore, solar energy systems require minimal maintenance once installed, leading to lower operational costs compared to traditional fossil fuel-based power plants. This economic viability, coupled with declining costs of solar technology, makes solar energy increasingly competitive and attractive for widespread deployment. Moreover, the growth of the solar industry stimulates job creation and economic growth, fostering innovation and investment in clean energy technologies.As countries transition towards a low-carbon future, solar energy plays a pivotal role in diversifying the energy mix and reducing reliance on fossil fuels. Through strategic investments in solar infrastructure, coupled with supportive policies and incentives, nations can accelerate the transition to a sustainable energy system while simultaneously mitigating the impacts of climate change. By harnessing the power of the sun, societies can pave the way towards a cleaner, greener, and more resilient future for generations to come.

Chapter 8: Conclusion

8.1 Limitations

1. Dependency on Data Quality: The accuracy and reliability of our system heavily rely on the quality of the input data, including real-time sensor readings and historical datasets. Poor data quality or inconsistencies could affect the performance of the forecasting models.

2. Sensitivity to Environmental Factors: Our system's predictive accuracy may be impacted by sudden environmental changes or extreme weather events that are challenging to forecast accurately. Variations in solar irradiance, wind speed, or ambient temperature could lead to deviations from predicted energy generation.

3. Regional Limitations: Our system's applicability may be limited to regions where reliable real-time data is available and where solar and wind energy generation is viable. It may not be suitable for areas with limited renewable energy potential or inadequate infrastructure for data collection.

4. Regulatory and Policy Constraints: Regulatory policies and market dynamics in the renewable energy sector could impact the adoption and implementation of our system. Compliance with regulations and navigating policy frameworks may pose challenges for deployment in some regions.

8.2 Conclusion

In this project, we conducted a predictive analysis of carbon footprint for forest covers in India, aiming to forecast greenhouse gas (GHG) emissions for the next 25 years. We employed three different models, SARIMA, ARIMA, and LSTM, to forecast GHG emissions and forest cover trends. Our analysis revealed crucial insights into the relationship between forest cover and GHG emissions, highlighting the limitations of relying solely on forest cover to mitigate GHG emissions. Our findings indicate that forest cover alone may not effectively address the challenge of GHG emissions in India. Despite efforts to preserve and expand forest areas, the GHG emission levels continue to rise. This underscores the need for alternative strategies to mitigate carbon emissions and transition towards renewable sources of energy.

To address this issue, we proposed a cost estimation solution focused on deploying solar panels to harness renewable energy and reduce carbon emissions. By leveraging data on renewable energy distribution in India, we demonstrated the potential of solar energy as a viable alternative to fossil fuels. The cost estimation model provided insights into the financial implications of implementing solar panel projects and highlighted the economic feasibility of transitioning towards renewable energy sources.

Overall, our project emphasizes the importance of proactive measures to combat climate change and reduce carbon emissions. While forest conservation remains crucial for biodiversity and ecosystem preservation, complementary efforts such as promoting renewable energy adoption are essential for achieving sustainable development and mitigating the adverse effects of climate change in India.

8.3 Future Scope

1. Our next work is to focus on reducing carbon emission by using renewable energy like hydro,solar wind biogas and geothermal energy sources.

2. We will calculate how much energy is produced by coal(which is the largest contributor to carbon emission (around 1.8 billion tonnes in 2021) and other non-renewable energy sources.

3. We will calculate the annual cost spent on these non renewable energy sources.

4. Then we will propose our solution which will replace fossil fuels with renewable energy sources at around the same cost.

5. We are planning on targeting 2 to 3 locations in india which has major emission problem and try to solve it(by planting windmills,solar panels)

References

[1] Jemyung Lee , Oliver Taherzadeh, , Keiichiro Kanemoto , "The scale and drivers of carbon footprints in households, cities, and regions across India", Research Institute for Humanity and Nature, 2021.

[2] Qi Huanga, Heran Zheng , Jiashuo Li , Jing Mengc, Yunhui Liud, Zhenyu Wange,Ning Zhanga, Yuan Li ,, Dabo Guanc, "Heterogeneity of consumption-based carbon emissions and driving forces in Indian states, Advances in Applied Energy, ,Volume 4, November 2021

[3] Ishma Amin, Meenakshi Agarwal, Stevert Lobo, Rahul Gurnani, and Prof. (Mrs.) Priya R.L, "Concept for Mapping Carbon footprint with Change in Vegetation Cover and Population in India" ,ITM Web Conf. , Volume 32, 2020 International Conference on Automation, Computing and Communication 2020 (ICACC-2020

[4] T.V. Ramachandra , Shwetmala , "Decentralized carbon footprint analysis for opting climate change mitigation strategies in India" , Renewable and Sustainable Energy Reviews , Volume 16, Issue 8, October 2012

[5] Surbhi Kumari, ,Sunil Kumar Singh , " Machine learning-based time series models for effective CO2 emission prediction in India" , Environmental Science and Pollution Research , June 25, 2022

[6] B. Mohan Kumar , Sreejith Aravindakshan , "Carbon footprints of the Indian AFOLU (Agriculture, Forestry, and Other Land Use) sector: a review " , Arunachal University of Studies, May 24, 2022

[7] Jingwei Han , Zhixiong Tan , Maozhi Chen , Liang Zhao , Ling Yang and Siying Chen , "Carbon Footprint Research Based on Input–Output Model—A Global Scientometric Visualization Analysis " International Journal of Environmental Research and Public Health

[8]Tao Gao, Qing Liu and Jianping Wang , "A comparative study of carbon footprint and assessment standards" ,Institute of Project Management, School of Mechanics and Civil Engineering, China

[9] Andréea Lorena Radua , Marian Albert Scrieciua , Dimitriu Maria Caracotaa , "Carbon Footprint Analysis: Towards a Projects Evaluation Model for Promoting Sustainable Development " , International Economic Conference of Sibiu 2013 Post Crisis Economy: Challenges and Opportunities, IECS 2013

Plagiarism Report

Project Review Sheet 1

Inhouse/Industry_Innovation/Research:

Sustainable Goal: Affordable & clean Energy (7) / Responsible consumption & production (12)

Project Evaluation Sheet 2023 - 24

Class: D17 A/B7C

Group No.: 30

Title of Project: Predictive analysis of carbon footprint for forest cover in India

Group Members: Utsav Gavli (7), Nikkita Gurnani (20), Namrata Avhad (032), Sonal Belani (05) D17A

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
4	4	4	3	4	2	2	2	2	2	2	2	3	2	2	40

Comments: _____

Name & Signature   Reviewer1
Pragna Solanke P...

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
4	4	4	3	4	2	2	2	2	2	2	2	2	2	3	40

Comments: - Good work done
           - Paper Publication required.

Date: 10th february, 2024

Name & Signature   Reviewer 2
Sunita Suralkar

Project Review Sheet 2

Inhouse/ Industry _Innovation/Research:

Sustainable Goal: Affortable & clean energy (7)) **Project Evaluation Sheet 2023 - 24**

Class: D17 A/B/C

Group No.: 30

Title of Project: Predictive Analysis of carbon footprint for forest cover in India.

Group Members: Nikhita Gurnani (20), Utsav Gavli (17), Namrata Athad (03), Sonal Belani (05)

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life-long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
4	4	4	2	4	2	2	2	1	2	3	3	2	2	3	40

Comments: ① How to correlate ym and ① GHG emmins c frret cover.
② Complete Paper c Comparative Study. SARIMAX & ARIMA / LSTM

Name & Signature   Reviewer1

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life-long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
4	4	4	2	4	2	2	2	2	2	3	3	2	2	3	41

Comments: _____

Date: 9th March, 2024

Sunita Suralkar

Name & Signature   Reviewer 2