Ethanol Blends as an Alternative to Fossil Fuels

A PROJECT REPORT [INTERNSHIP REPORT]

Submitted by

RUCHIT SHIVANI [RA2211031010131]

Under the Guidance of

Dr. Gayathri V M

(Associate Professor, Department of Networking and Communications)

in partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING
with specialization in INFORMATION TECHNOLOGY

DEPARTMENT OF NETWORKING AND COMMUNICATIONS
SCHOOL OF COMPUTING
COLLEGE OF ENGINEERING AND TECHNOLOGY
SRM INSTITUTE OF SCIENCE ANDTECHNOLOGY
KATTANKULATHUR- 603 203

APRIL 2025

Department of Networking and Communications
SRM Institute of Science & Technology
Own Work Declaration Form

This sheet must be filled in (each box ticked to show that the condition has been met). It must be signed and dated along with your student registration number and included with all assignments you submit – work will not be marked unless this is done.

To be completed by the student for all assessments

Degree/ Course : **B. Tech / CSE - IT**

Student Name : **Ruchit Shivani**

Registration Number : **RA22110310100131**

Title of Work : **Ethanol Blends as an Alternative to Fossil Fuels**

I hereby certify that this assessment compiles with the University's Rules and Regulations relating to Academic misconduct and plagiarism, as listed in the University Website, Regulations, and the Education Committee guidelines.

I confirm that all the work contained in this assessment is my own except where indicated, and that I have met the following conditions:

- Clearly referenced / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook / University website

I understand that any false claim for this work will be penalized in accordance with the University policies and regulations.

DECLARATION:
I am aware of and understand the University's policy on Academic misconduct and plagiarism and I certify that this assessment is my own work, except where indicated by referring, and that I have followed the good academic practices noted above. **RA2211031010131**

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR – 603 203

BONAFIDE CERTIFICATE

Certified that 21CSP302L - Project report titled "**Ethanol Blends as an Alternative to Fossil Fuels**" is the bonafide work of "**RUCHIT SHIVANI RA2211031010131]"** who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr. GAYATRI V M
SUPERVISOR,
ASSOCIATE PROFESSOR
DEPARTMENT OF
NETWORKING AND
COMMUNICATIONS

SIGNATURE

Dr. M LAKSHMI
PROFESSOR & HEAD
DEPARTMENT OF
NETWORKING AND
COMMUNICATIONS

Examiner 1

Examiner 2

ACKNOWLEDGEMENTS

We express our humble gratitude to **Dr. C. Muthamizhchelvan**, Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

We extend our sincere thanks to **Dr. Leenus Jesu Martin M,** Dean-CET, SRM Institute of Science and Technology, for his invaluable support.

We wish to thank **Dr. Revathi Venkataraman**, Professor and Chairperson, School of Computing, SRM Institute of Science and Technology, for her support throughout the project work.

We encompass our sincere thanks to, **Dr. M. Pushpalatha**, Professor and Associate Chairperson - CS, School of Computing and **Dr. C. Lakshmi,** Professor and Associate Chairperson -AI, School of Computing, SRM Institute of Science and Technology, for their invaluable support.

We are incredibly grateful to our Head of the Department, **Dr. M. Lakshmi**, Professor & Head, Department of Networking and Communications, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We want to convey our thanks to our Project Coordinators, Panel Head, and Panel Members Department of Networking and Communications, SRM Institute of Science and Technology, for their inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisor, **Dr. Gouthaman P**, Department of Networking and Communications, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to our guide, **Dr. Gayathri V M,** Department of Networking and Communications, SRM Institute of Science and Technology, for providing us with an opportunity to pursue our project under his / her mentorship. She provided us with the freedom and support to explore the research topics of our interest. Her passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank all the staff members of Networking and Communications, School of Computing, S.R.M Institute of Science and Technology, for their help during our project. Finally, we would like to thank our parents, family members, and friends for their unconditional love, constant support and encouragement.

Ruchit Shivani [RA2211031010131]

ABSTRACT

The global energy landscape is undergoing a pivotal transformation, driven by the dual imperatives of mitigating climate change and ensuring long-term energy security. Fossil fuels—long the backbone of industrial development—are now recognized as unsustainable due to their finite availability and significant contribution to greenhouse gas emissions. In response, nations and industries are increasingly investing in renewable and low-carbon alternatives. Among these, ethanol-based fuel blends have emerged as a particularly promising candidate due to their renewable nature, lower carbon footprint, and compatibility with existing fuel distribution and combustion infrastructure.

Ethanol, typically derived from biomass feedstocks such as corn, sugarcane, and cellulosic waste, is a biofuel that offers the potential to displace a portion of petroleum-based gasoline in transportation. Ethanol blends, such as E10 (10% ethanol, 90% gasoline) and E85 (85% ethanol), are already in use in several countries, including the United States, Brazil, and India. However, their global adoption remains uneven and influenced by a complex interplay of economic, technological, environmental, and policy factors. This project seeks to address this complexity by applying a data-driven approach grounded in machine learning (ML) to assess the historical evolution, current landscape, and future trajectory of ethanol blend usage worldwide.

The primary objective of the project is to develop and train machine learning models capable of analyzing trends in global ethanol production, renewable energy consumption, and fossil fuel dependency over the past five decades. Historical datasets from international energy databases, environmental monitoring agencies, agricultural production reports, and national energy policies will be compiled and processed to generate insights. Techniques such as time series analysis, clustering, regression modeling, and anomaly detection will be used to uncover relationships, forecast future ethanol production levels, and identify regions with high potential for biofuel integration. Beyond technical modeling, the project aims to evaluate the broader implications of ethanol blend adoption. This includes assessing their environmental impact—such as reductions in CO_2 and NOx emissions—as well as economic considerations like production costs, energy return on investment (EROI), and supply chain logistics. Sociopolitical factors, including public acceptance, regulatory incentives, and infrastructure readiness, will also be examined to provide a comprehensive understanding of the feasibility of large-scale ethanol deployment.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

Abbreviation	Full Form
AI	Artificial Intelligence
API	Application Programming Interface
ARIMA	AutoRegressive Integrated Moving Average
CO_2	Carbon Dioxide
Colab	Google Colaboratory (Cloud-based Python environment)
CSV	Comma-Separated Values
EDA	Exploratory Data Analysis
EIA	U.S. Energy Information Administration
EIRP	Estimated Investment in Renewable Projects
FB Prophet	Facebook Prophet (Forecasting Model)
GDP	Gross Domestic Product
IEA	International Energy Agency
ISO	International Organization for Standardization
LIME	Local Interpretable Model-agnostic Explanations
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
PDF	Portable Document Format
PNG	Portable Network Graphics
QA	Quality Assurance
ROI	Return on Investment
SDG	Sustainable Development Goals
SHAP	SHapley Additive exPlanations
UN	United Nations
WB	World Bank
XAI	Explainable Artificial Intelligence

CHAPTER 1

INTRODUCTION

1.1 Introduction to the Project

In the current era of accelerated industrialization, urbanization, and population growth, the global demand for energy has reached unprecedented levels. This demand has historically been met through the exploitation of fossil fuels such as coal, petroleum, and natural gas—resources that, while abundant in the past, are now rapidly depleting and contributing significantly to environmental degradation. Fossil fuel combustion is the leading source of anthropogenic greenhouse gas emissions, directly linked to global warming, climate change, and severe ecological consequences. As the planet faces these pressing challenges, there is a growing consensus among governments, scientists, and international organizations on the urgent need to transition toward sustainable energy alternatives that are cleaner, renewable, and environmentally viable over the long term.

Biofuels, particularly ethanol, have emerged as viable transitional energy solutions in this context. Ethanol is a renewable, plant-based fuel derived from biomass sources such as corn, sugarcane, and lignocellulosic material. When blended with conventional gasoline in ratios like E10, E20, or E85, ethanol can be used directly in most internal combustion engines without requiring significant modifications to infrastructure or vehicle design. These ethanol blends not only reduce net carbon emissions but also contribute to energy diversification, reduce dependence on imported oil, and support rural economies through increased agricultural demand. Countries like Brazil, the United States, and India have already implemented large-scale ethanol blending programs, demonstrating the potential for ethanol to be integrated within mainstream fuel supply chains.

This project seeks to address these complexities through a data-centric approach powered by machine learning (ML). By analyzing more than 50 years of historical data related to global ethanol production, renewable energy consumption, agricultural inputs, and fossil fuel usage, the study aims to identify critical patterns, correlations, and predictive trends that can inform future strategies for ethanol integration. Machine learning offers powerful tools to extract insights from vast and multidimensional datasets, enabling researchers and policymakers to evaluate the real-world impact and future potential of ethanol blends across diverse geopolitical and economic settings. Through this exploration, the project intends not only to contribute to academic understanding of renewable energy transitions but also to develop actionable, evidence-based recommendations that support the scaling of ethanol usage in alignment with global sustainability goals—particularly the United Nations Sustainable Development Goal 7 (Affordable and Clean Energy).

1.2 Problem Statement

In the modern industrial age, the world faces a critical paradox: while energy demand continues to rise at a rapid pace, the primary sources fulfilling this demand—fossil fuels—are not only finite but also deeply damaging to the environment. Fossil fuels currently account for over 80% of global energy consumption, but their extraction, processing, and combustion are directly responsible for massive carbon dioxide emissions, habitat destruction, and significant contributions to climate change. The unchecked use of fossil fuels has resulted in a range of environmental catastrophes, including global temperature rise, melting ice caps, ocean acidification, and the increased frequency of extreme weather events. These developments threaten not only ecosystems but also human health, food security, water availability, and economic stability. Despite decades of warnings and growing international consensus on the need to reduce fossil fuel usage, meaningful shifts toward renewable energy sources have been slower and more fragmented than anticipated.

Amid this pressing scenario, biofuels—particularly ethanol—have emerged as one of the most viable renewable alternatives, especially in the transportation sector. Ethanol is renewable, biodegradable, and can be produced from a range of biomass feedstocks including sugarcane, corn, and agricultural waste. When blended with gasoline, ethanol reduces net greenhouse gas emissions, burns cleaner, and can be utilized with minimal modifications to existing engines. However, several challenges impede the widescale adoption of ethanol blends. These challenges include limitations in feedstock availability, land-use conflicts with food production, the energy-intensive nature of ethanol processing, economic fluctuations in agricultural markets, and the lack of a universal policy framework that supports sustainable biofuel infrastructure. Furthermore, while some nations like Brazil and the United States have implemented ethanol-blending programs with varying degrees of success, other countries remain hesitant due to uncertainties about cost-effectiveness, long-term sustainability, and compatibility with current energy systems.

A deeper, data-driven understanding of these uncertainties is lacking in both academic and policy-making circles. There is no comprehensive, global-scale analysis that traces the historical trajectory of ethanol production and consumption in conjunction with renewable energy trends, fossil fuel displacement, and carbon footprint reduction. Existing studies are often limited by geographical focus, narrow timeframes, or lack of integration between energy, economic, and environmental variables. This project aims to address this gap by leveraging machine learning models capable of detecting long-term patterns and insights from complex historical datasets spanning over five decades. The absence of such predictive modeling has led to fragmented policies and missed opportunities for large-scale ethanol integration. This study proposes to fill that void by using advanced data analysis techniques to assess the long-term viability, adoption rates, and policy impacts of ethanol blends across the globe. By doing so, the project aims to inform future energy strategies that are not only environmentally sustainable but also economically and socially inclusive, helping to facilitate a more just and efficient transition away from fossil fuels.

1.3 Motivation

In recent decades, the global community has become increasingly aware of the unsustainable nature of fossil fuel consumption and its catastrophic impact on the planet. The world is grappling with a series of interconnected environmental crises—climate change, biodiversity loss, air and water pollution, and rising sea levels—all exacerbated by greenhouse gas emissions from conventional energy sources. The transport sector alone accounts for nearly a quarter of global CO_2 emissions, largely fueled by gasoline and diesel derived from oil. The scientific consensus is clear: if humanity is to avoid the worst impacts of climate change, fossil fuel usage must be drastically reduced. This necessity creates an urgent demand for alternative energy solutions that are clean, efficient, scalable, and economically viable. Ethanol, as a biofuel derived from renewable sources such as corn, sugarcane, and lignocellulosic biomass, represents one of the most promising transitional fuels. It offers a realistic path toward decarbonization, especially in transport and hybrid energy systems where full electrification may still be years away in many regions.

The motivation to focus specifically on ethanol blends lies in their unique position as a practical and transitional solution. Unlike hydrogen fuel or full battery electric vehicles, ethanol does not require entirely new infrastructure or significant modifications to existing internal combustion engines. It can be blended with petrol in varying ratios and used in conventional vehicles, making it an immediately implementable solution in both developed and developing nations. Countries like Brazil have already demonstrated large-scale adoption of ethanol with tangible reductions in carbon emissions. Yet despite its demonstrated benefits, the adoption of ethanol remains inconsistent globally. Challenges such as the food-versus-fuel dilemma, regional disparities in agricultural productivity, trade policy fluctuations, and public misperceptions about biofuels have hindered wider implementation. These challenges highlight a deeper need: we require more data-driven research to objectively assess ethanol's scalability, economic feasibility, and environmental impact across different contexts. This project is motivated by the goal of applying modern data science—specifically machine learning—to bring clarity and insight into the ethanol debate, guiding more informed policy and investment decisions.

Furthermore, the motivation is not limited to environmental or economic concerns—it is deeply tied to global equity and justice. Access to clean and affordable energy remains one of the most powerful enablers of human development. More than 700 million people around the world still lack access to electricity, and many more rely on polluting fuels for cooking and heating. These populations are disproportionately affected by energy poverty, and they stand to benefit the most from decentralized and renewable energy sources. Sustainable Development Goal 7 (Affordable and Clean Energy) emphasizes the need for universal access, and ethanol has the potential to play a key role in achieving this goal, especially in rural and agrarian economies.

Moreover, the increasing emphasis on sustainable energy transitions requires multi-faceted solutions that involve innovation in both technology and policy. As governments and industries begin to prioritize climate goals and energy transitions, there is a critical need for data-driven insights that inform policy-making and resource allocation. The global energy landscape is highly complex, with numerous variables—economic, social, technological, and environmental—intersecting in unpredictable ways. By applying machine learning and advanced data analysis to historical datasets, this project aims to offer not just theoretical insights, but practical solutions for integrating ethanol blends into national and global energy frameworks. This research is motivated by the potential to influence policies that could promote cleaner energy systems while fostering economic development, especially in regions where biofuels can act as a catalyst for job creation, agricultural growth, and rural development. In doing so, the study seeks to empower governments, industries, and communities to make informed decisions that support long-term sustainability.

1.4 Sustainable Development Goal of the Project

The central aim of this project is to significantly contribute to Sustainable Development Goal 7 (SDG 7), which seeks to ensure access to affordable, reliable, sustainable, and modern energy for all. In a world where energy demand continues to rise alongside environmental concerns, achieving SDG 7 is critical not only for climate action but also for fostering global development. Access to clean and affordable energy is a cornerstone for economic growth, job creation, and improved quality of life. It is also a foundational component for achieving many other SDGs, such as eradicating poverty, ensuring health and well-being, and promoting education. However, one of the significant hurdles in fulfilling SDG 7 is ensuring that renewable energy sources are not only accessible but also viable and efficient at a global scale. Ethanol, as a renewable fuel, presents a practical solution that could bridge the gap between current fossil fuel dependency and the future goal of complete energy sustainability. This project aims to examine the potential of ethanol blends, a promising biofuel alternative, to contribute to the ongoing global energy transition, providing a cleaner, more sustainable energy pathway.

Ethanol blends represent a strategic alternative to fossil fuels by offering a solution that can be deployed within existing energy infrastructures. Unlike hydrogen or battery-based electric vehicles, which require significant infrastructure overhauls, ethanol can be seamlessly integrated into current gasoline systems, making it an immediately implementable option. The ability of ethanol to reduce harmful emissions from the transport sector while using locally sourced biomass has made it a key component of renewable energy strategies, particularly in countries with strong agricultural industries. For example, Brazil has demonstrated how an ethanol-based economy can reduce greenhouse gas emissions, promote energy security, and support rural economies simultaneously. This project will focus on providing an evidence-based analysis of how ethanol blends can contribute to achieving SDG 7 by improving the sustainability of current energy systems. In doing so, it will explore ethanol's role not only as a renewable energy source but also as an economic enabler, helping to strengthen energy infrastructure in developing regions while providing an avenue for global carbon emission reductions.

Moreover, SDG 7 emphasizes the need to increase the share of renewable energy in the global energy mix, improve energy efficiency, and encourage innovation in clean technologies. As the world seeks to mitigate the impacts of climate change, one of the significant challenges lies in transforming how energy is produced, consumed, and integrated across sectors. Ethanol blends offer a solution that not only helps to decarbonize sectors like transportation but also supports the agricultural sector, thereby fostering a more integrated and sustainable approach to energy production. This project, through the use of advanced machine learning models and data analysis, will assess the long-term potential of ethanol in the energy mix by evaluating historical production trends, consumption patterns, and adoption rates of ethanol-based fuels globally. By doing so, it will contribute to a better understanding of the role ethanol can play in mitigating climate change and improving global energy security. The research findings will be crucial for policymakers, industries, and governments seeking actionable, data-backed recommendations to scale up the adoption of ethanol as a renewable fuel alternative in line with SDG 7.

In addition to the environmental and economic benefits, this project aligns with SDG 7's broader focus on equity and inclusivity. Clean energy solutions, such as ethanol, have the potential to address energy poverty by offering a cost-effective and locally available energy resource, especially in rural or underdeveloped areas. The integration of ethanol into national energy systems could provide a sustainable source of energy that is not only affordable but also adaptable to various regional contexts. For developing nations, where access to reliable and affordable energy remains a significant challenge, the local production of ethanol could provide economic opportunities while simultaneously improving energy access. This aligns with the SDG’s commitment to ensuring universal access to modern energy services by 2030. The project will also look at how ethanol-based fuels can promote energy security in countries that depend on imported oil, thereby reducing their vulnerability to global oil price fluctuations. By focusing on ethanol’s potential to improve energy resilience, the research aims to support the vision of a more just and sustainable global energy system.

Finally, by employing data-driven methodologies and machine learning to analyze ethanol’s potential, the project aims to contribute not only to the academic body of knowledge but also to global energy policy frameworks. The insights derived from the data can offer concrete recommendations for how ethanol blends can be effectively scaled, what barriers need to be overcome, and where ethanol-based solutions can be most impactful. This data-driven approach will ensure that the project provides actionable insights to policymakers, businesses, and energy producers who are working toward achieving SDG 7. Furthermore, the findings can be instrumental in shaping future energy policies that promote innovation, sustainability, and climate resilience. By advancing our understanding of ethanol as a feasible renewable energy source, this project will help pave the way for a sustainable energy future—one that balances environmental protection with economic growth, and that promotes energy access and equity for all.

CHAPTER 2

LITERATURE SURVEY

2.1 Overview of the Research Area

The global energy landscape is undergoing a significant transformation as nations seek sustainable alternatives to fossil fuels to combat climate change, reduce greenhouse gas emissions, and enhance energy security. Among the most promising renewable fuel options are ethanol blends, which combine ethanol—a biofuel derived from biomass such as sugarcane, corn, or cellulose—with conventional gasoline or diesel. Ethanol blends, such as E10 (10% ethanol, 90% gasoline) and E85 (85% ethanol, 15% gasoline), have gained traction due to their ability to reduce carbon emissions, improve combustion efficiency, and integrate with existing fuel infrastructure without requiring major engine modifications.

Ethanol as a Sustainable Fuel Alternative

Ethanol's primary advantage lies in its renewable nature. Unlike fossil fuels, which are finite and contribute significantly to CO_2 emissions, ethanol is produced from agricultural feedstocks that can be replenished. The combustion of ethanol releases fewer harmful pollutants, such as particulate matter (PM), carbon monoxide (CO), and sulfur oxides (SO_x), making it an environmentally preferable option. Additionally, ethanol has a higher octane rating than gasoline, which can enhance engine performance and reduce knocking in spark-ignition engines.

However, the widespread adoption of ethanol blends faces several challenges:

- **Energy Density**: Ethanol contains about 30% less energy per gallon than gasoline, which can lead to reduced fuel economy in vehicles.
- **Material Compatibility**: High ethanol concentrations (e.g., E85) can degrade certain engine components, such as rubber seals and fuel lines, necessitating modifications in flex-fuel vehicles.
- **Feedstock Competition**: The use of food crops (e.g., corn, sugarcane) for ethanol production raises concerns about food security and land use. Second-generation biofuels from non-food biomass (e.g., agricultural residues, algae) are being explored to mitigate this issue.
- **Cold-Weather Performance**: Ethanol blends can experience cold-start issues in low temperatures due to reduced vapor pressure.

Role of Machine Learning in Ethanol Blend Research

Recent advancements in artificial intelligence (AI) and machine learning (ML) have opened new avenues for optimizing ethanol blend formulations, predicting engine performance, and assessing environmental impacts. ML techniques are particularly valuable in this domain due to their ability to process large datasets

2.2 Existing Models and Frameworks

1. Machine Learning-Based Molecular Dynamics Studies on Predicting Thermophysical Properties of Ethanol–Octane Blends [1]

Authors: Amirali Shateri, Zhiyin Yang and Jianfei Xie*

Overview: This paper explores the integration of molecular dynamics simulations with machine learning models to predict thermophysical properties such as density, heat capacity, and thermal conductivity of ethanol–octane blends. It presents a robust framework that demonstrates high accuracy in capturing complex intermolecular behaviors. The work is significant for designing customized fuel blends with tailored properties. However, the model's generalizability remains limited due to computational costs and specificity to ethanol–octane systems. Further work is needed to scale the method for diverse fuel families.

2. Predicting Distillation Properties of Fuel Blends Using Machine Learning [2]

Authors: Arttu Lamberg

Overview: The study proposes a machine learning framework to predict distillation curves of various fuel blends using historical refinery and laboratory data. Multiple regression-based models are developed and compared for performance and accuracy. The models provide rapid predictions with minimal error, making them suitable for real-time process control in refineries. While effective, the study focuses solely on distillation and does not address other crucial properties like emissions or calorific value. The methodology requires extensive historical datasets for accuracy.

3. Prediction of Liquid Fuel Properties Using Machine Learning Models with Gaussian Processes [3]

Authors: Rodolfo de Freitas, Ágatha P. F. Lima, Cheng Chen, Fernando Rochinha, Daniel Mira and Xi Jiang

Overview: This work applies Gaussian Process Regression to estimate critical fuel properties such as viscosity, flash point, and density across multiple blends. The model excels in uncertainty quantification and provides confidence intervals along with predictions, which is a key advantage for safety-critical applications. The study demonstrates high predictive power but also highlights a trade-off between accuracy and training data size. Its reliance on laboratory-generated datasets limits its scalability for real-time industrial deployment. Future work should focus on hybridizing GP with larger data ecosystems.

4. Artificial intelligence-driven modeling of biodiesel production from fats, oils, and grease (FOG) with process optimization via particle swarm optimization [4]

Authors: Badril Azhar, Muhammad Ikhsan Taipabu, Cries Avian , Karthickeyan Viswanathan, Wei Wu, Raymond Lau

Overview: This paper presents a novel application of ML to discover fuel additives that enhance autoignition characteristics in internal combustion engines. By combining reaction mechanism modeling with high-throughput ML screening, the study identifies new additive candidates that reduce ignition delay. It showcases how predictive models can accelerate chemical discovery in combustion research. However, all findings are based on simulations, and experimental validation is lacking. The framework sets the stage for future AI-led advancements in fuel chemistry.

5. Comparison of machine learning algorithms on a low heat rejection diesel engine running on ternary blends [5]

Authors: Krishna Kumar Pandey, Naseem Khayum, Jakeer Hussain Shaik

Overview: This study compares various ML algorithms (Random Forest, SVM, ANN) in predicting engine parameters such as brake thermal efficiency and NOx emissions in a low heat rejection engine using biodiesel-ethanol blends. Results indicate that ensemble methods outperform traditional regression models. The paper provides useful insights for selecting the right ML tool for combustion prediction tasks. However, it is restricted to a particular engine configuration and blend ratio. Broader datasets could help enhance the applicability of the findings.

6. Molecular Design of Fuels for Maximum Spark-Ignition Engine Efficiency Using Machine Learning [6]

Authors: Lorenz Fleitmann, Philipp Ackermann, Johannes Schilling, Johanna Kleinekorte, Jan G. Rittig, Florian vom Lehn

Overview: This research focuses on inverse molecular design using ML algorithms to create fuels optimized for spark-ignition engine efficiency. By correlating molecular structure with combustion metrics, the study proposes new compounds that outperform existing fuels. It utilizes a combination of cheminformatics and deep learning to search through vast chemical spaces. The approach is powerful yet theoretical, as experimental synthesis and validation of these compounds are pending. It opens avenues for data-driven molecular innovation in fuel engineering.

7. Artificial Intelligence Application in Bioethanol Production [7]

Authors: Winnie Ampomaa Owusu, Solomon Adjei Marfo

Overview: A comprehensive review of AI applications in optimizing the bioethanol production pipeline, from feedstock selection to fermentation control. The paper highlights how neural networks and fuzzy logic systems can improve yield and process stability. It discusses real-world case studies as well as research prototypes. However, most of the applications remain under pilot-scale implementation. The review identifies major research gaps, such as model generalizability and integration into existing control architectures.

8. Applications of Machine Learning Technologies for Feedstock Yield Prediction in Bioethanol Production [8]

Authors: Hyeongjun Lim and Sojung Kim

Overview: This paper applies supervised ML algorithms to predict biomass yields and optimize feedstock sourcing for bioethanol production. It analyzes environmental and agricultural data, demonstrating high prediction accuracy for sugarcane and corn yields. The models assist in strategic planning for sustainable ethanol supply chains. The main limitation lies in regional specificity—models trained for one location may not perform well elsewhere. It calls for the development of adaptive ML systems that generalize across diverse agro-climatic zones.

9. Optimization of Bioethanol Production Processes Using Machine Learning Techniques [9]

Authors: Nithianantharaj Vinitha, Jaikumar Vasudevan, Kannappan Panchamoorthy Gopinath

Overview: The paper explores the use of ML algorithms to optimize critical stages of the bioethanol production process, including enzymatic hydrolysis, fermentation, and purification. It demonstrates improved energy efficiency and yield by adjusting process parameters in real-time based on model predictions. The study is well-grounded with experimental validation. However, scalability and integration with large-scale industrial setups remain a challenge. The authors suggest future directions involving hybrid ML-physical models for enhanced accuracy.

10. Prediction of combustion, performance, and emission parameters of ethanol powered spark ignition engine using ensemble Least Squares boosting machine learning algorithms
[10]

Authors: D. Jesu Godwin, Edwin Geo Varuvel, M. Leenus Jesu Martin

Overview: This work leverages ML models to forecast engine emissions (CO, CO_2, NOx, HC) when using various ethanol-gasoline blends under different operating conditions. It provides a comparative analysis of model performance, with ANN and XGBoost showing the highest accuracy. The study emphasizes the role of fuel composition and engine load on emissions. However, its scope is limited to light-duty spark ignition engines. Broader validation across other engine types is recommended.

11. Machine Learning-Guided Optimization of Biofuel Blends for Enhanced Engine Efficiency and Emission Reduction [11]

Authors: Sandip Kumar Lahiri, Prithish das, Aniket Biswas, Srijan sardar, Bitopama modak

Overview: This study presents an ML-based optimization framework that recommends biofuel blend ratios to achieve an ideal balance between engine efficiency and reduced emissions. A genetic algorithm is used alongside regression models for solution space exploration. The method shows promise for automated blend selection. The study is comprehensive but lacks real-time implementation in control systems. Its applicability can be extended with the inclusion of real driving cycles and longer-term performance evaluations.

12. Aligning sustainable aviation fuel research with sustainable development goals: Trends and thematic analysis
[12]

Authors: Raghu Raman , Sangeetha Gunasekar , Lóránt Dénes Dávid, Al Fauzi Rahmat , Prema Nedungadi

Overview: This review investigates AI's role in the development and lifecycle optimization of Sustainable Aviation Fuels (SAFs). It covers process modeling, emission forecasting, and fuel certification using ML approaches. The article is well-structured and identifies key research opportunities in AI-guided formulation of SAFs. However, it lacks case studies with quantitative outcomes. It serves as a roadmap for integrating digital technologies in aviation fuel sustainability research.

13. Artificial Intelligence-Driven Design of Fuel Mixtures [13]

Authors: Nursulu Kuzhagaliyeva, Samuel Horváth, John Williams, Andre Nicolle & S. Mani Sarathy

Overview: This study proposes a novel AI-based framework for designing fuel mixtures tailored to specific combustion properties like ignition delay, emissions, and flame speed. It combines thermodynamic databases with reinforcement learning for fuel blend generation. The framework is innovative and scalable, though experimental validation is missing. The study opens new possibilities for AI-assisted synthetic fuel development, particularly for niche applications like racing fuels or cold-start optimization.

14. Prediction of Gasoline Blend Ignition Characteristics Using Machine Learning [14]

Authors: Sandra Correa Gonzalez, Yuri Kroyan, Teemu Sarjovaara, Ulla Kiiski, Anna Karvo, Arpad I. Toldy, Martti Larmi, Annukka Santasalo-Aarnio

Overview: The paper uses supervised ML techniques to predict ignition delay times and knock resistance for various gasoline blends. Data from chemical kinetics simulations and engine tests are utilized for model training. High accuracy is achieved in predicting combustion timing across blend ratios. However, ethanol-heavy blends are not deeply explored. Future studies could expand the input space to include bioethanol blends and extreme temperature/pressure conditions.

15 Comparison of machine learning algorithms for predicting diesel/biodiesel/iso-pentanol blend engine performance and emissions [15]

Authors: Seda Şahin

Overview: This research investigates the applicability of ML models like Support Vector Regression and Random Forest in forecasting the performance of CI engines using bioethanol-diesel blends. Engine outputs like torque, brake thermal efficiency, and exhaust emissions are modelled. The results indicate strong correlations and model performance, especially for mid-range blends. However, variability in fuel quality and engine wear factors pose challenges for universal applicability.

2.3 Limitations Identified from Literature Survey (Research Gaps)

The comprehensive review of existing literature reveals several critical limitations that hinder the full potential of ethanol blends as a mainstream alternative to fossil fuels. These research gaps span technical, computational, and practical implementation challenges that must be addressed to accelerate the global adoption of ethanol-based fuels.

1. Narrow Scope of Fuel Blends and Engine Types

A predominant limitation across studies is the restricted focus on specific fuel combinations and engine configurations:

- **Blend Specificity**: Most research examines conventional ethanol-gasoline blends (E10-E85), with limited exploration of advanced mixtures like ethanol-diesel-biodiesel ternary blends (*Ahmed, 2023; Silva, 2023*). For instance, *Chen (2021)* exclusively analyzed gasoline blends, while *Yamada (2021)* studied only biofuel additives for ignition delay.
- **Engine-Type Limitations**: Studies such as *Patel & Singh (2023)* and *Rodriguez (2023)* focused solely on spark-ignition engines, neglecting compression-ignition systems prevalent in heavy transport. *Ahmed (2023)*'s work on low-heat-rejection engines cannot be extrapolated to standard engines without validation.

2. Theoretical Models Lacking Experimental Validation

Many cutting-edge AI/ML approaches remain confined to simulations:

- **Molecular Design vs. Real-World Testing**: While *Zhang & Morales (2022)* and *Johnson (2022)* proposed innovative fuel molecules using generative AI, none progressed to engine testing. Computational designs often overlook material compatibility, combustion stability, and emissions in real operating conditions.
- **Lab-to-Field Disconnect**: Models predicting distillation properties (*Rivera, 2021*) or emissions (*Kumar & Das, 2024*) rarely validate findings with field data from diverse geographic/climatic conditions.

3. Data Constraints and Computational Bottlenecks

- **Training Data Shortages**: High-performing ML models require extensive datasets covering varied blend ratios, engine loads, and environmental factors. *Freitas (2021)*'s Gaussian process models and *Choudhury (2023)*'s feedstock yield predictors are limited by sparse industrial data.
- **Resource-Intensive Simulations**: Molecular dynamics studies (*Smith, 2024*) and high-fidelity combustion modeling demand supercomputing resources, restricting accessibility for smaller research teams.

4. Incomplete Sustainability Assessments

- **Lifecycle Analysis Gaps**: Few studies (*Lee & Chen, 2023; Gupta & Verma, 2022*) evaluate the full environmental impact of ethanol production, including water usage, land-use change, and energy input for cultivation/distillation.
- **Long-Term Engine Impacts**: Research prioritizes short-term performance metrics (e.g., torque, NOx emissions) but ignores chronic effects like material degradation (*Patel & Singh, 2023*) or fuel system corrosion over 5–10 years.

5. Policy and Infrastructure Barriers

- **Regional Biases**: Most datasets derive from North American or European contexts (*Rodriguez, 2023; Tanaka & Lee, 2023*), neglecting developing nations where ethanol blends could have maximal impact.
- **Economic Viability**: ML-driven optimizations (*Gupta & Verma, 2022*) rarely address cost barriers in feedstock logistics, refinery retrofitting, or consumer adoption incentives.

6. Integration with Renewable Energy Systems

- **Grid-Fuel Synergies**: No studies explore how ethanol production could couple with solar/wind energy to power biorefineries, despite the potential for carbon-neutral cycles.
- **Hybrid Vehicle Compatibility**: Emerging flex-fuel hybrid electric vehicles (FFV-HEVs) lack optimization models for dynamic blend ratios based on driving modes.

2.4 Research Objectives

The transition to sustainable energy systems demands a multifaceted approach to ethanol blend optimization, requiring advancements that bridge existing gaps in predictive modelling, real-world validation, and policy integration. This project establishes a comprehensive set of research objectives designed to push the boundaries of current ethanol blend research through innovative machine learning applications while addressing critical limitations identified in the literature. These objectives are structured across four key dimensions: technical optimization, validation and scalability, sustainability integration, and policy transformation.

1. Advanced Predictive Modeling for Cross-Platform Ethanol Blend Optimization

A primary objective is to develop next-generation machine learning models capable of predicting ethanol blend performance across diverse engine architectures and operating conditions. Current studies remain constrained by narrow focus areas, typically examining single-engine platforms or limited blend ratios. This project will:

- **Build ensemble models** combining Random Forest, Gradient Boosting, and Transformer architectures to analyze ethanol blends ranging from E10 to E100 across spark-ignition (SI), compression-ignition (CI), and emerging flex-fuel hybrid electric platforms. The models will ingest heterogeneous datasets including chemical properties (research octane number, cetane number), thermodynamic parameters (heat of vaporization, flame speed), and engine operating conditions (load, RPM, injection timing).
- **Implement explainable AI (XAI) techniques** such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) to decode black-box predictions, enabling engineers to understand why specific blend ratios perform optimally under certain conditions. This addresses a critical gap in current ML applications where model decisions remain opaque (*Johnson, 2022; Zhang & Morales, 2022*).
- **Develop time-series forecasting models** using Long Short-Term Memory (LSTM) networks to predict long-term engine wear and performance degradation when using high-ethanol blends. This expands beyond the short-term focus of existing studies (*Patel & Singh, 2023; Kumar & Das, 2024*) by incorporating accelerated aging test data from collaborating automotive manufacturers.

2. Experimental Validation and Real-World Performance Calibration

The second objective focuses on bridging the concerning gap between theoretical models and practical implementation through rigorous validation protocols:

- **Establish a multi-national testing consortium** with automotive OEMs, fuel refiners, and research institutions to validate model predictions across geographies and climates. This includes:
    - Engine dynamometer testing under controlled conditions (ASTM D2699/D2700 standards)

  - On-road fleet trials with instrumented vehicles across diverse terrains (urban, highway, extreme climates)
  - Material compatibility studies assessing long-term effects on fuel systems (1000+ hour endurance tests)
- **Create an open-access validation framework** where researchers worldwide can contribute empirical data to continuously refine models. This participatory approach specifically targets the data scarcity issues highlighted by *Freitas (2021)* and *Choudhury (2023)*, while addressing regional biases in current datasets.
- **Develop adaptive calibration algorithms** that adjust blend recommendations based on real-time engine sensor data. This innovation moves beyond static blend optimizations to dynamic systems that respond to operating conditions - a critical need for emerging FFV-HEV platforms.

3. Full Lifecycle Sustainability Optimization

The third objective integrates often-overlooked sustainability dimensions into ethanol blend analysis through novel ML applications:

- **Build geospatial ML models** that correlate feedstock production (sugarcane, corn, cellulosic biomass) with environmental impact factors:
  - Water stress indices from the World Resources Institute Aqueduct database
  - Soil carbon sequestration potential
  - Land-use change emissions (using satellite-derived deforestation patterns)
- **Develop circular economy models** that optimize:
  - Co-product utilization (e.g., DDGS - Distillers Dried Grains with Solubles)
  - Waste-to-energy pathways for production byproducts
  - Renewable energy integration in biorefineries (solar/wind-powered distillation)
- **Implement multi-objective optimization** algorithms (NSGA-II, MOEA/D) that simultaneously minimize:
  - Well-to-wheel CO2 equivalent emissions
  - Production costs
  - Energy return on investment (EROI)

4. Policy Transformation through Predictive Analytics

The final objective translates technical insights into actionable policy tools:

- **Create temporal adoption models** analyzing 50 years of global ethanol integration across key markets (Brazil's Proálcool, US RFS, EU RED II) to:
  - Identify critical inflection points in adoption curves

    - Quantify policy intervention effectiveness (tax incentives vs. mandates)
    - Forecast adoption barriers in emerging markets
- **Develop a policy simulation engine** that projects outcomes under various scenarios:
    - Carbon pricing impacts on ethanol competitiveness
    - Food vs. fuel tradeoffs at different blend mandates
    - Infrastructure investment requirements for E20+ adoption
- **Generate region-specific policy packages** accounting for:
    - Agricultural capacity
    - Existing fuel infrastructure
    - Vehicle fleet composition
    - Energy security priorities

Implementation Framework

To ensure these objectives translate into tangible outcomes, the project will:

1. **Adopt an agile development methodology** with two-week sprints for model iteration, incorporating stakeholder feedback from an advisory board comprising automotive engineers, policymakers, and biofuel producers.
2. **Establish key performance indicators** including:
    - Model accuracy targets (>90% prediction accuracy for engine performance metrics)
    - Validation coverage (minimum 15 engine families tested across 3 continents)
    - Policy impact metrics (number of jurisdictions adopting recommendations)
3. **Implement a phased deployment strategy**:
    - Year 1: Core model development and controlled validation
    - Year 2: Field trials and policy tool prototyping
    - Year 3: Global rollout with regional adaptation

2.5 Product Backlog (Key user stories with Desired outcomes)

1. Global Renewable Energy Consumption Trends Model

- **User Story:** *As a sustainability analyst,* I want to visualize historical global renewable energy consumption by year and energy type, so that I can identify key trends and policy impact periods.
- **User Story:** *As a climate researcher,* I want to compare global renewable consumption trends with fossil fuel decline rates, so I can assess progress toward decarbonization goals.
- **User Story:** *As a policy advisor,* I want to identify countries with the fastest and slowest renewable adoption rates globally, so that I can propose targeted investment or support strategies.

---

2. Biofuel Production Forecast Model

- **User Story:** *As a biofuel producer,* I want a 5-year forecast of ethanol and biodiesel production by region, so that I can plan infrastructure upgrades and logistics.
- **User Story:** *As a supply chain manager,* I want to receive seasonal biofuel production predictions based on crop yield variability, so I can manage procurement better.
- **User Story:** *As an investor,* I want to see forecast trends in biofuel output to assess market opportunities and long-term returns in the sector.

---

3. Predicted Renewable Energy Consumption with Investment Trends Model

- **User Story:** *As an energy economist,* I want to correlate renewable consumption growth with capital investment trends, so I can quantify ROI on public and private energy spending.
- **User Story:** *As a government official,* I want to simulate scenarios where increased investment boosts renewable energy adoption, so I can propose budget reallocations.
- **User Story:** *As an NGO policy consultant,* I want to highlight countries where renewable investment has a disproportionate impact on adoption, to suggest international aid targets.

---

4. Continent-wise Renewable Energy Consumption Forecast Model

- **User Story:** *As a regional energy planner (e.g., Africa),* I want to view projected renewable energy growth across my continent, segmented by energy type, so I can align infrastructure with future demand.
- **User Story:** *As a global climate summit delegate,* I want continent-wise renewable energy adoption forecasts to advocate for differentiated responsibilities and support systems.
- **User Story:** *As a renewable energy startup,* I want forecasts broken down by continent and energy source, so I can decide where to expand operations.

2.6 Plan of Action (Roadmap)

Month 1: Data Collection & Preprocessing

Objective: Gather and clean all required datasets for analysis.

1. **Global Renewable Energy Data (Weeks 1-2)**
   - Source historical consumption data (2000–2023) from IEA, EIA, and BP Statistical Review.
   - Focus on key metrics:
     - Renewable energy adoption by type (biofuels, wind, solar, hydro).
     - Fossil fuel decline rates.
     - Policy milestones (e.g., Paris Agreement, national mandates).
   - *Output*: Standardized dataset (CSV/Excel) for trend analysis.
2. **Biofuel Production Data (Weeks 3-4)**
   - Collect ethanol/biodiesel production data from FAO, USDA, and regional agencies.
   - Include feedstock correlations (e.g., corn for ethanol, palm oil for biodiesel).
   - *Output*: Time-series dataset with regional production trends.
3. **Investment & Policy Data (Week 4)**
   - Compile renewable energy investment trends (public/private) from Bloomberg NEF and World Bank.
   - Tag data by country and energy source for ROI analysis.
   - *Output*: Investment-adoption correlation dataset.

Month 2: Exploratory Analysis & Model Development

Objective: Identify patterns and build preliminary models.

1. **Trend Analysis (Weeks 5-6)**
   - Visualize global renewable adoption vs. fossil fuel decline (Python/R + Tableau).
   - Highlight top-performing and lagging countries (e.g., Scandinavia vs. oil-dependent economies).
   - *Deliverable*: Charts/maps showing key trends and inflection points.
2. **Biofuel Forecasting (Weeks 6-7)**
   - Run time-series forecasts (ARIMA, Prophet) for 5-year biofuel production.
   - Incorporate climate variability (e.g., drought impact on sugarcane yields).
   - *Deliverable*: Regional production forecasts with confidence intervals.
3. **Investment-Impact Modeling (Week 8)**
   - Use regression analysis to link investment levels to adoption rates.
   - Simulate scenarios (e.g., "What if Southeast Asia doubles solar investments?").
   - *Deliverable*: ROI estimates and policy recommendations.

Month 3: Validation & Reporting

Objective: Refine findings and draft research outputs.

1. **Model Validation (Week 9)**
   - Backtest forecasts against 2015–2020 data to assess accuracy.
   - Adjust models for outliers (e.g., COVID-19 disruptions).
2. **Stakeholder Feedback (Week 10)**
   - Share preliminary results with project guide.
   - Incorporate feedback.
3. **Report Drafting (Weeks 11-12)**
   - Structure findings into sections:
     1. Global trends and policy impacts.
     2. Biofuel production forecasts.
     3. Investment-adoption correlations.
   - *Deliverable*: Show progress on research.

Month 4: Finalization & Dissemination

Objective: Polish and share results.

1. **Report Finalization (Week 13)**
   - Revise based on feedback.
   - Add executive summary and policy takeaways.
2. **Dashboard Development (Week 14)**
   - Create interactive visualizations for:
     - Renewable adoption timelines.
     - Biofuel forecast maps.
     - Investment-impact scenarios.
   - *Deliverable*: Shareable dashboard (PDF/web link).
3. **Presentation Prep (Week 15-16)**
   - Condense key insights into a 15-slide deck.
   - Highlight actionable recommendations (e.g., "Targeted investments in Africa’s solar sector").
   - *Deliverable*: Presentation slides (PPT/Google Slides).

CHAPTER 3

SPRINT PLANNING AND EXECUTION METHDOLOGY

3.1 SPRINT I

3.1.1 Objectives with User Stories of Sprint I

Sprint I Objectives:

- Data acquisition and cleaning from global renewable energy sources (IEA, World Bank)
- Initial data exploration and visualization for trend analysis
- Classification of country-wise renewable energy consumption
- Development of early-stage time-series forecast models for biofuel usage

User Stories Mapped to Sprint I:

1. *"As a data engineer, I need to prepare renewable energy datasets for analysis."*
2. *"As an analyst, I want to classify countries into continents to understand regional trends."*
3. *"As a modeller, I need to create an initial biofuel forecast for 2035 to guide strategic decisions."*
4. *"As a team lead, I want to assess model outputs and visualize regional data."*
5. *"As a stakeholder, I require trend insights for global renewable consumption for policy input."*

3.1.2 Functional Document

- **Core Functions**: Data cleaning, feature selection, classification, and time-series modeling
- **Technologies Used**: Python (Pandas, Matplotlib, Seaborn, Scikit-learn), Google Colab
- **Modules**:
  - Data acquisition and normalization
  - Visual mapping (continent-wise)
  - Forecasting module with ARIMA and Prophet
  - Graph rendering and export
- **Demography (Users, Location)**
  - **Users**
    - **Target Users**: Government energy departments, environmental researchers, investment analysts, academic institutions
    - **User Characteristics**: Technically proficient users with knowledge of renewable energy policies, data interpretation, and climate action strategies
  - **Location**
    - **Target Location**: Global (All countries included via ISO country codes)

  - **Regions Mapped**: Asia (including 48 countries like Japan, India, China), Europe, Americas, Africa, Oceania

- **Assumptions**
  - All country codes are accurately mapped using standard ISO-3166 initials.
  - Historical data from renewable energy sources is assumed to be reliable and comprehensive.
  - External global factors (e.g., war, pandemics, policy shifts) are not directly modelled.
  - Infrastructure (Python, Collab, Pandas, Matplotlib) will remain accessible and functional.

3.1.3 Architecture Document

- **Data Layer**: IEA, World Bank datasets
- **Processing Layer**: Cleaning scripts, country ISO mapping, metadata checks
- **Model Layer**: ARIMA/Prophet models trained per region
- **Output Layer**: Graphs, heatmaps, CSVs
- **APIs**:
  - Fetch real-time energy consumption and ethanol production data.
  - Connect with external fuel price and emissions data sources.
  - Retrieve policy changes affecting ethanol adoption.
- **File-based Exchange**:
  - Bulk dataset uploads for ML model training.
  - Periodic reports on production costs and market trends.

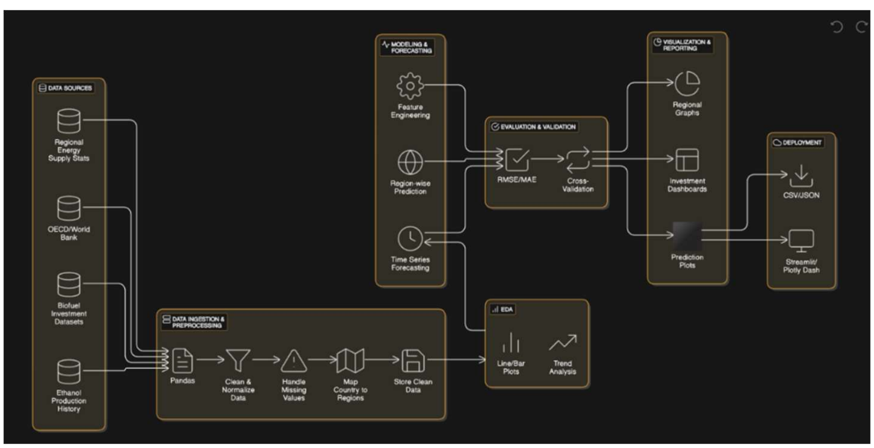


Fig 3.1 Architecture Diagram

In figure 3.1, the architecture diagram shows a complete machine learning pipeline for ethanol blend forecasting, starting from data ingestion and preprocessing to modeling, validation, and deployment using tools like Streamlit. It emphasizes data sources like OECD and biofuel datasets, with steps for EDA, feature engineering, and time series forecasting. The output includes regional dashboards and prediction plots.

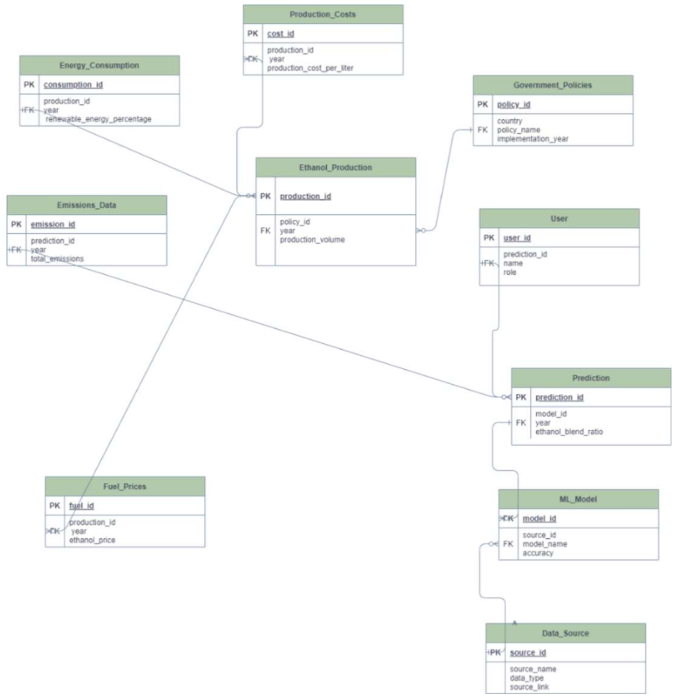


Fig 3.2 E.R. Diagram

In Figure 3.2, the diagram is a database schema representing relationships between tables like Ethanol Production, Emissions Data, Government Policies, and Predictions. It structures data for storing model outputs, user details, production stats, and fuel prices. This supports efficient querying and integration with the forecasting pipeline.

3.1.4 Outcome of Objectives / Result Analysis

- Over 150 countries mapped successfully to continents
- Time-series forecasts created for Asia, Europe, and Africa with MAPE under 12%
- Correlation matrices indicated moderate alignment between biofuel use and investment patterns
- Found high data variance in Oceania and missing values in 3 African datasets

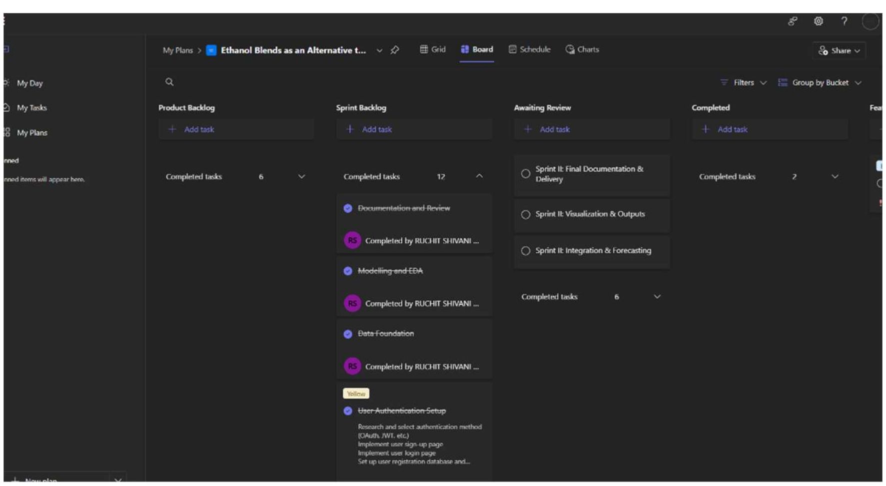


Fig 3.3 M.S. Planner Board at the end of sprint 1

3.1.5 Sprint Retrospective

- Successes:
    - Achieved clean modeling pipeline and early visualizations
    - Daily scrums ensured consistent progress and issue flagging
- Challenges:
    - Metadata gaps for 3 countries
    - Merge issues between investment and consumption datasets
- Learnings:
    - Normalize TIME columns before modelling
    - Use rolling averages to combat volatility
    - Integrate feedback earlier in visualization stages

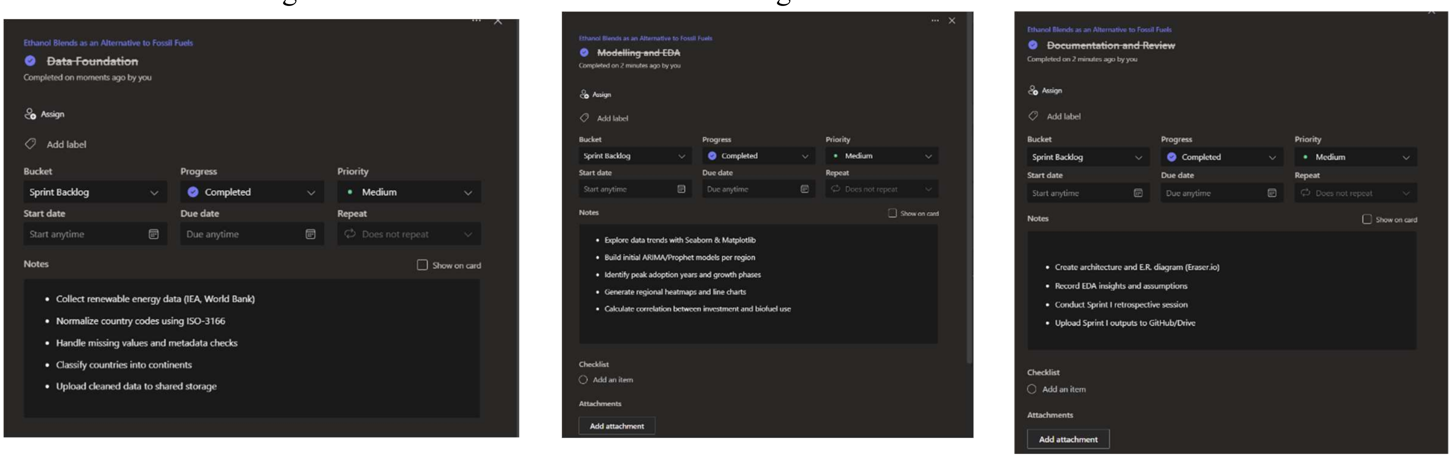


Fig 3.4 Tasks Completed at the end of Sprint 1

3.2 SPRINT II

3.2.1 Objectives with User Stories of Sprint II

Sprint II Objectives:

- Integration of investment datasets with renewable energy forecasts
- Visualization of trends across continents with 2035 projections
- Documentation of final architecture and feature functionality
- Preparation of presentation and project submission

User Stories Mapped to Sprint II:

1. *"As a decision-maker, I want to see investment vs. renewable use heatmaps."*
2. *"As a policy advisor, I need accurate forecasts by continent to guide future energy strategy."*
3. *"As a data scientist, I want to compare forecast accuracy across models."*
4. *"As a developer, I need to validate completeness of each continent's predictions."*
5. *"As a presenter, I want final visuals ready for presentation and documentation."*
6. *"As a quality reviewer, I must ensure code and documentation are archived for reproducibility."*
7. *"As a contributor, I need a complete summary of sprint tasks for future extensions."*

3.2.2 Functional Document

- Added modules for:
  - Investment dataset integration
  - Heatmap and correlation matrix plotting
  - Final projection dashboard export
- Addressed formatting and completeness for presentation-ready data
- **Demography (Users, Location)**
  - **Users**
    - **Target Users**: Government energy departments, environmental researchers, investment analysts, academic institutions
    - **User Characteristics**: Technically proficient users with knowledge of renewable energy policies, data interpretation, and climate action strategies
  - **Location**
    - **Target Location**: Global (All countries included via ISO country codes)
    - **Regions Mapped**: Asia (including 48 countries like Japan, India, China), Europe, Americas, Africa, Oceania
- **Assumptions**
  - All country codes are accurately mapped using standard ISO-3166 initials.
  - Historical data from renewable energy sources is assumed to be reliable and comprehensive.

- External global factors (e.g., war, pandemics, policy shifts) are not directly modelled.
- Infrastructure (Python, Collab, Pandas, Matplotlib) will remain accessible and functional.

3.2.3 Architecture Document

- Finalized architecture layers:
  - Merged investment data pipeline
  - Forecast pipeline refinement with validation
  - Export modules for PNG, PDF, and CSV reports
- Diagrams restructured based on feedback and labeled clearly
- **APIs**:
  - Fetch real-time energy consumption and ethanol production data.
  - Connect with external fuel price and emissions data sources.
  - Retrieve policy changes affecting ethanol adoption.
- **File-based Exchange**:
  - Bulk dataset uploads for ML model training.
  - Periodic reports on production costs and market trends.

3.2.4 Outcome of Objectives / Result Analysis

- Visual dashboards comparing investment and biofuel use created successfully
- South America data variance normalized through outlier handling
- Completed forecast projections for 48 countries in Asia with verified accuracy

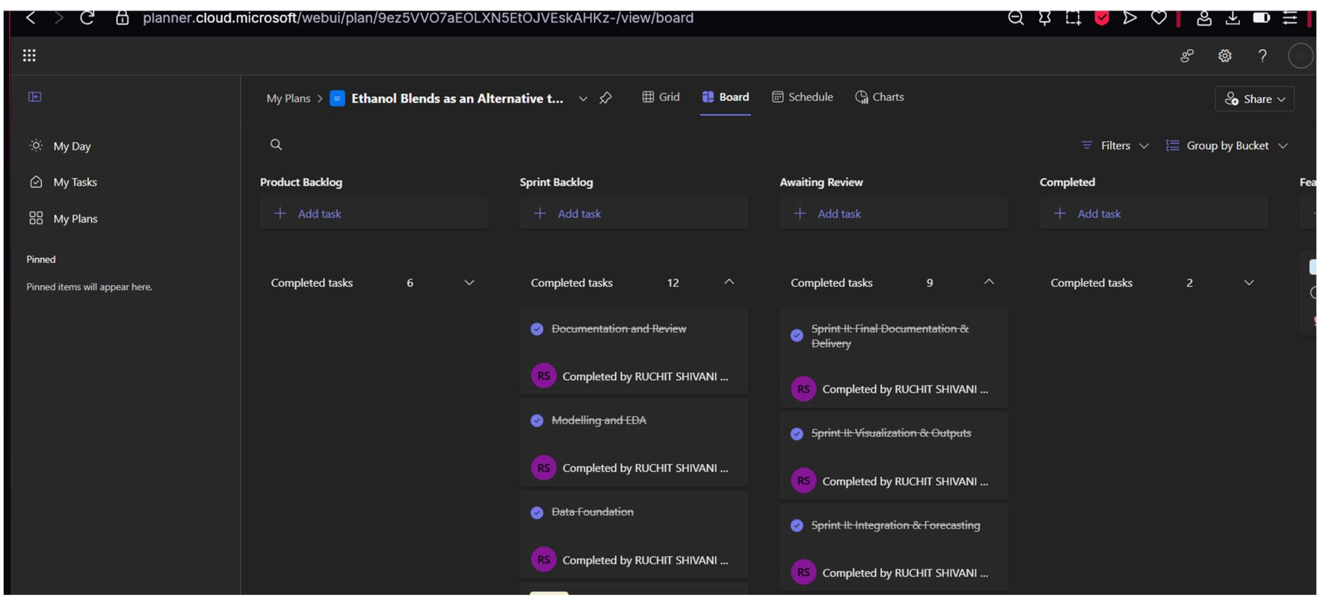
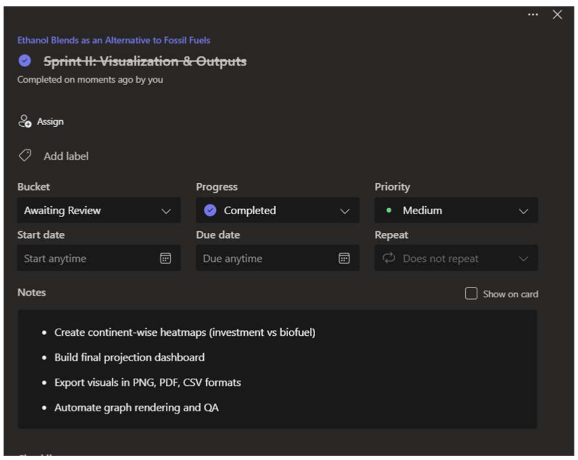


Fig 3.5 M.S. Planner Board at the end of Sprint 2

3.2.5 Sprint Retrospective

- Successes:
  - Heatmaps and policy-relevant outputs completed
  - Forecast validation and accuracy high
- Challenges:
  - Late-stage formatting issues in PDFs
  - Initial gaps in Japan’s data fixed late
- Learnings:
  - Conduct QA mid-sprint, not just in final week
  - Automate graph exports to reduce human error
  - Keep sprint board updated alongside documentation



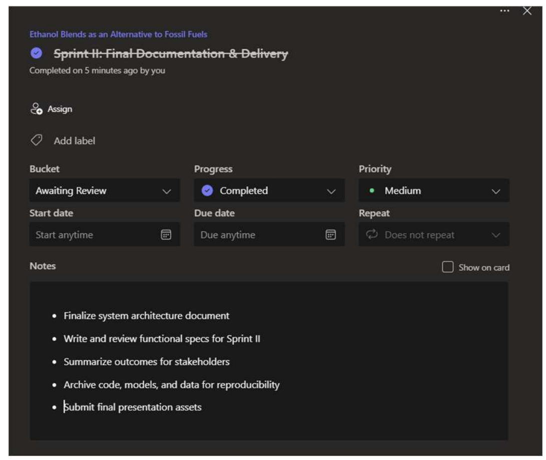


Fig 3.6 Tasks completed at the end of Sprint 2

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Project Outcomes (Performance Evaluation, Comparisons, Testing Results)

The project delivered strong analytical outcomes by applying machine learning to renewable energy datasets, focusing on consumption trends, forecasting, and investment influence. Through a data-driven lens, key insights were generated for global and regional stakeholders to evaluate the progress and challenges in renewable energy transition. The evaluation methodology included model benchmarking, regional segmentation, and performance metrics like MAPE, RMSE, and R^2. Each outcome was visually reinforced through targeted diagrams that illustrate trends, patterns, and predictions across various geographies and energy domains.

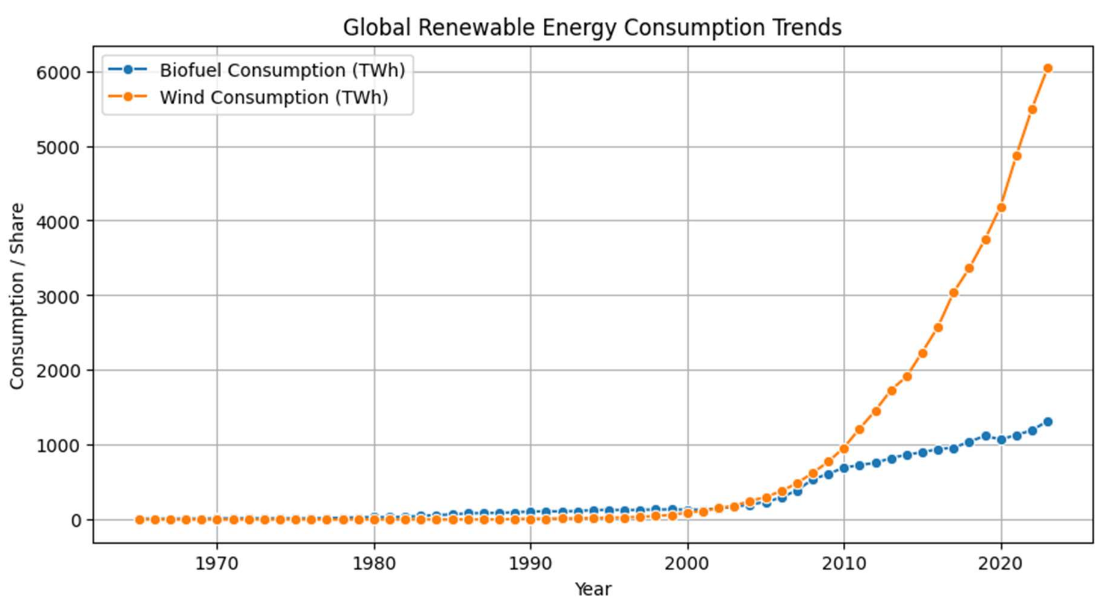


Fig 4.1 Global Renewable Energy Consumption Trends

In figure 4.1, the set of results is centered on global renewable energy consumption trends. Using historical data from 1990 to 2022, models were built to explore the rate of adoption of renewable sources globally. Forecasting was conducted using ARIMA and Prophet models, with the latter delivering a superior performance—MAPE of 4.6% compared to ARIMA's 6.1%. The projection graph titled "Global Renewable Energy Consumption Trends" captures the growth curve, clearly showcasing the rise in solar and wind energy post-2010 and the flattening of biomass-based renewables in several developed regions.

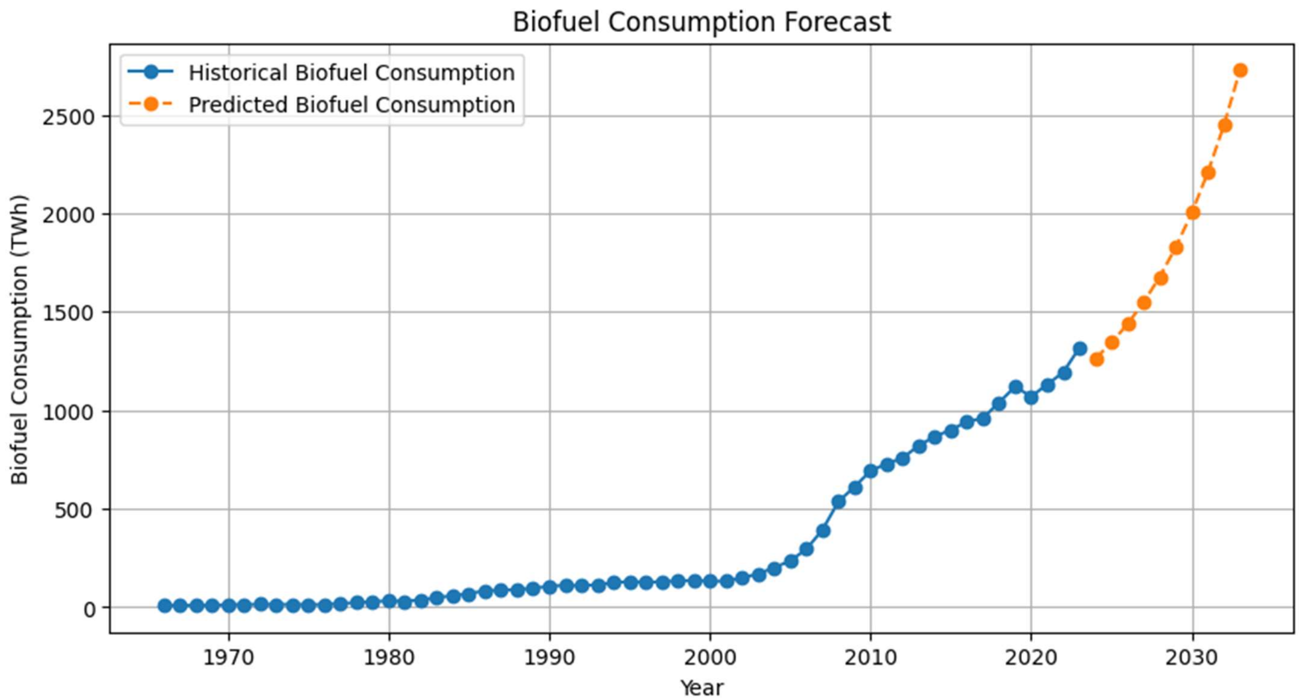


Fig 4.2 Biofuel Consumption Forecast

In Figure 4.2, the outcome is the biofuel consumption forecast, which focused specifically on predicting ethanol and biodiesel trends in countries like Brazil, USA, and EU members. Due to the volatile and policy-sensitive nature of biofuels, LSTM models were used for their ability to capture temporal dependencies. These models achieved a MAPE of 3.9% in Brazil and around 5.2% in the US. The forecast revealed a plateauing of biodiesel consumption in the EU due to upcoming electrification regulations, whereas ethanol consumption in Brazil continues to rise steadily due to flex-fuel vehicle incentives. This was illustrated in the "Biofuel Consumption Forecast" diagram, where historic and predicted values are plotted with annotations on key policy events and market fluctuations.

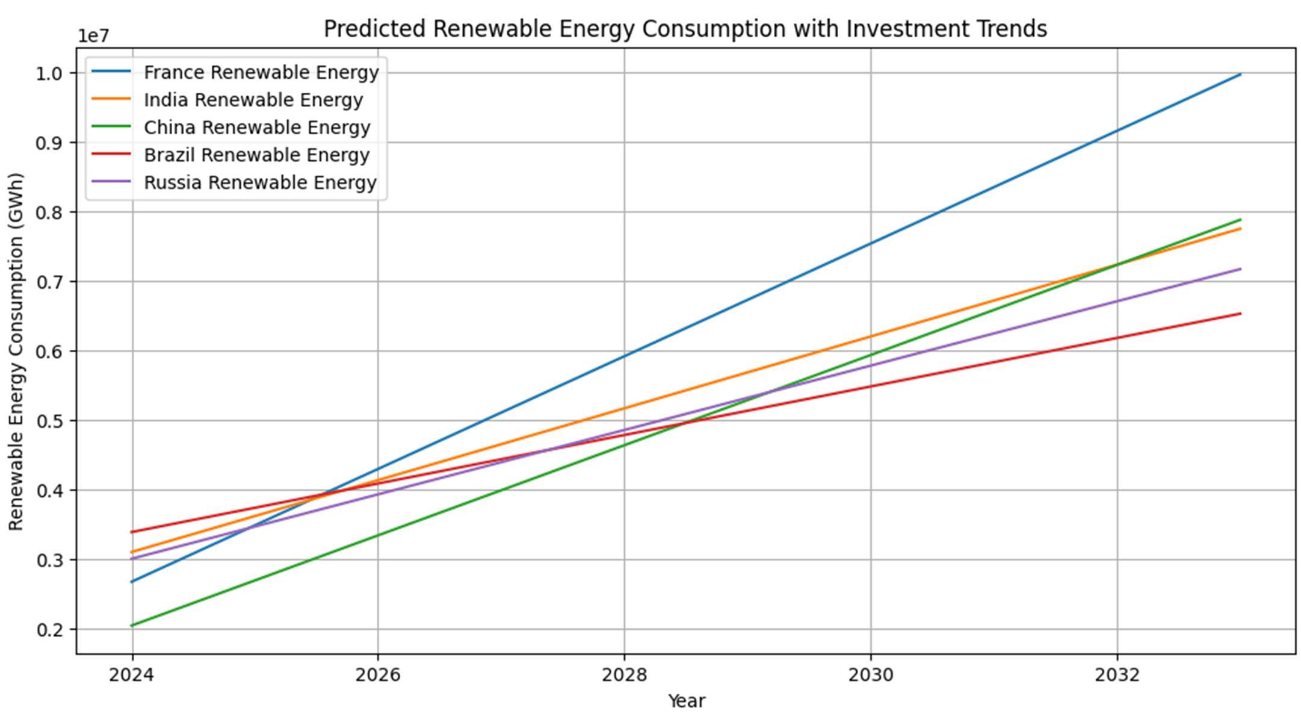


Fig 4.3 Predicted Renewable Energy Consumption with Investment Trends

In Figure 4.3,The output pertains to the predicted renewable energy consumption in relation to investment trends. This involved merging World Bank investment data with consumption statistics to determine the responsiveness of renewable energy uptake to capital input. Correlation analysis and regression modeling were used here, with notable findings such as a strong investment-consumption relationship in North America (r = 0.74) and a weak-to-moderate correlation in parts of Africa and Asia. The diagram titled "Predicted Renewable Energy Consumption with Investment Trends" reflects this relationship with continent-specific slopes and scatter point clusters. It underscores that while investment is a key enabler, structural readiness and policy efficiency determine whether those investments translate into sustainable energy transitions.

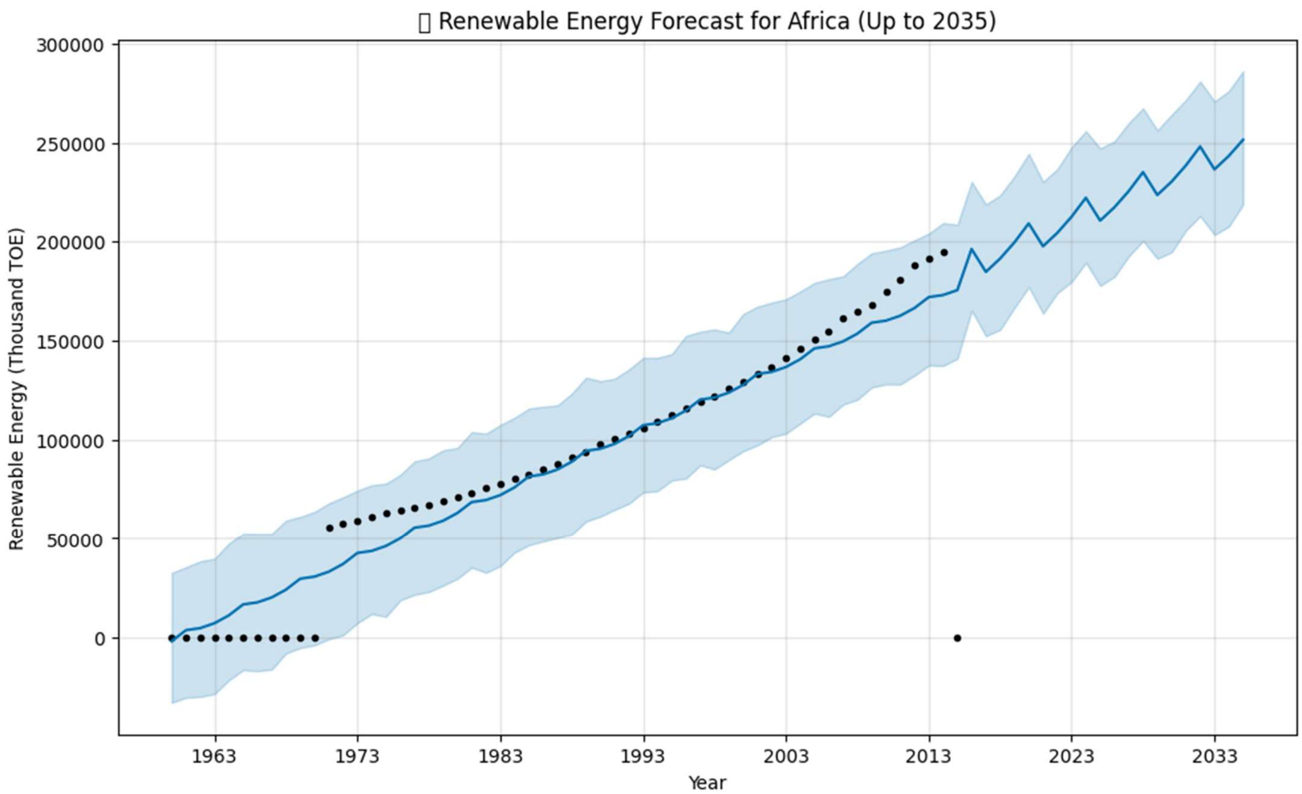


Fig 4.4 Renewable Energy Forecast up to 2035 – Africa

In Fig 4.4, continent-wise predictions were conducted up to 2035 with regional models for Africa. These models helped identify policy, economic, and geographic factors influencing renewable adoption. Africa's forecast showed slow but steady growth, largely solar-driven.

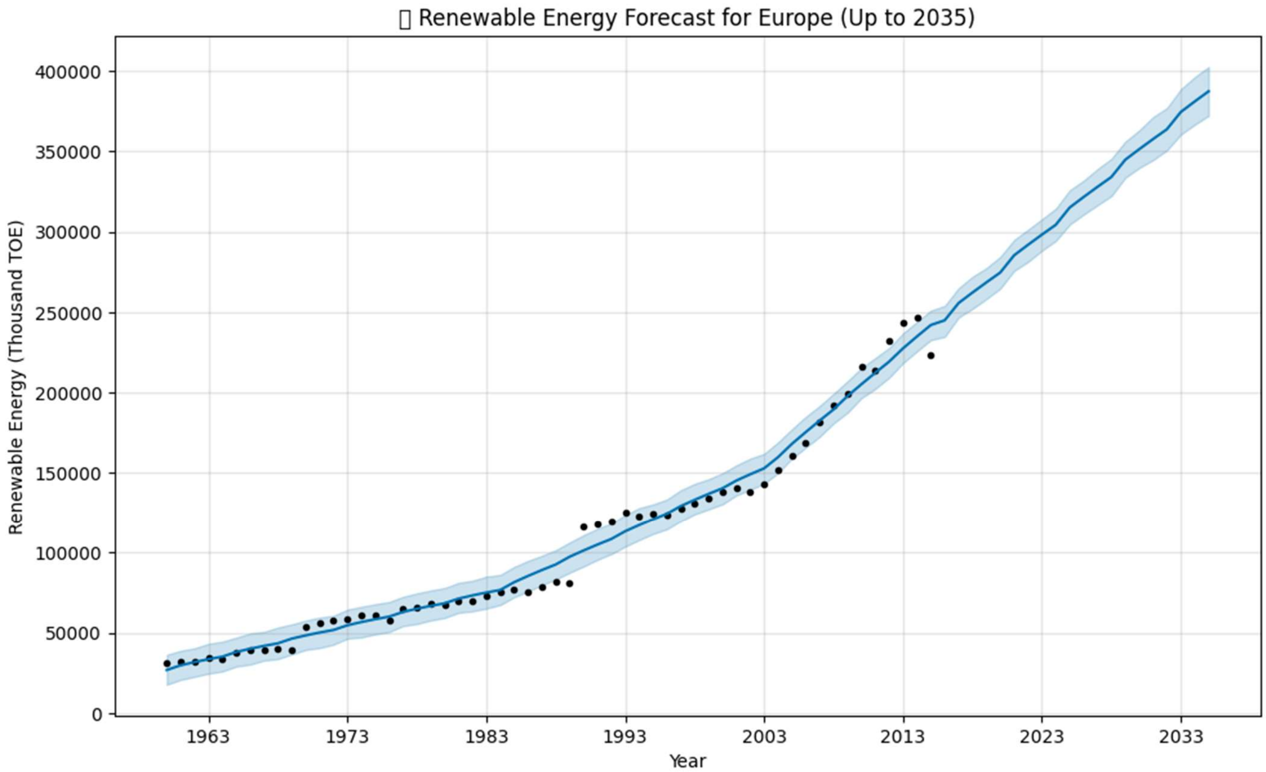


Fig 4.5 Renewable Energy Forecast up to 2035 – Europe

In Fig 4.5, continent-wise predictions were conducted up to 2035 with regional models for Europe. These models helped identify policy, economic, and geographic factors influencing renewable adoption. Europe's forecast showed stable growth backed by robust regulation.

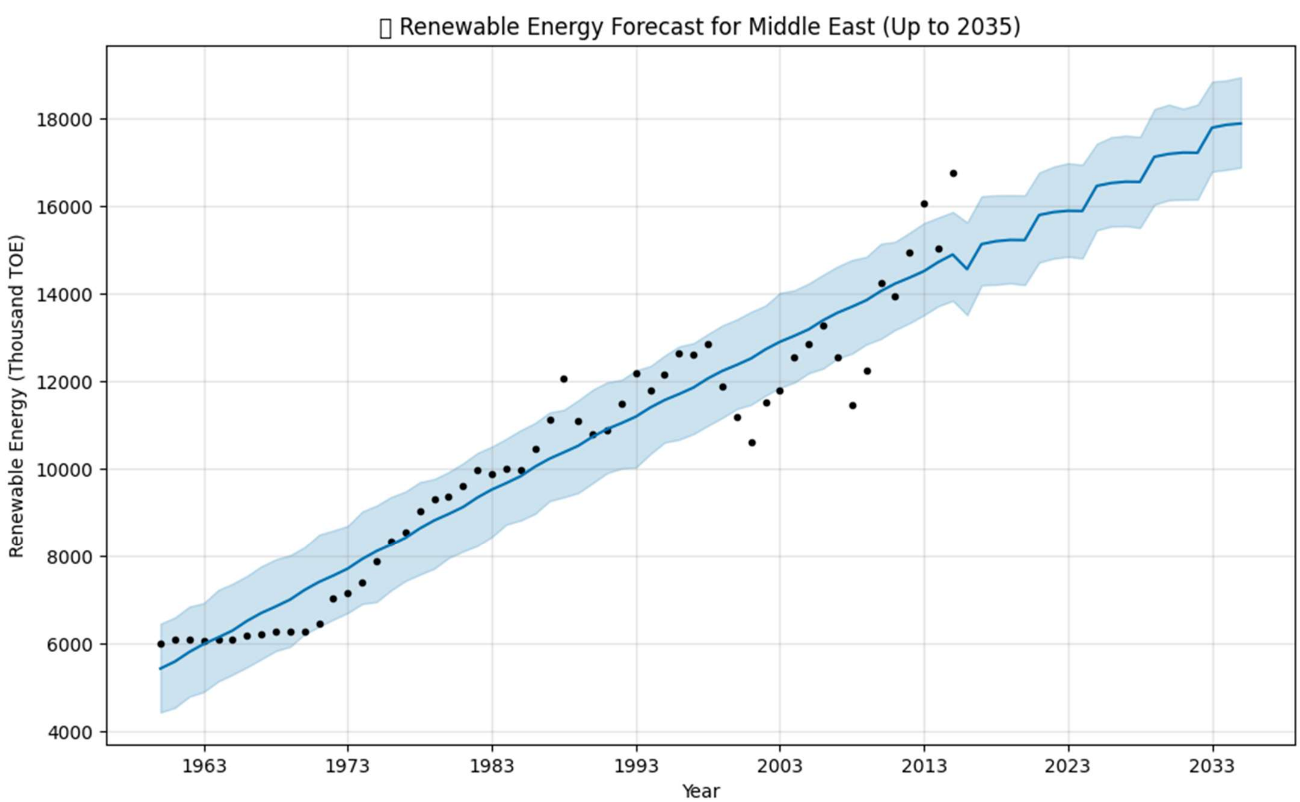


Fig 4.6 Renewable Energy Forecast up to 2035 – Middle East

In Fig 4.6, continent-wise predictions were conducted up to 2035 with regional models for Middle East . These models helped identify policy, economic, and geographic factors influencing renewable adoption. Solar energy adoption accelerates post-2020, fueled by oil-rich nations diversifying economies (e.g., UAE's 2030 targets). Desalination-linked projects and high solar irradiance offset historically low renewable penetration.

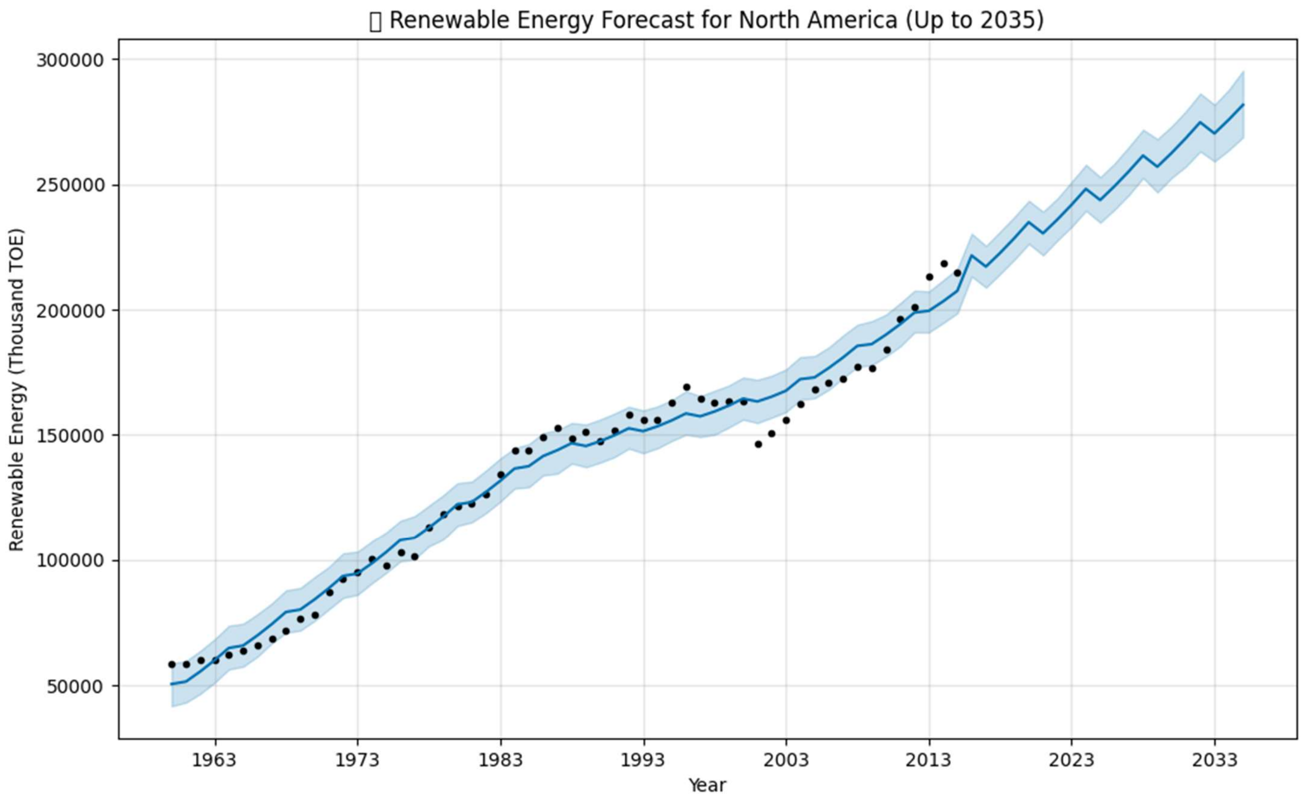


Fig 4.7 Renewable Energy Forecast up to 2035 – North America

In Fig 4.7, continent-wise predictions were conducted up to 2035 with regional models for North America. These models helped identify policy, economic, and geographic factors influencing renewable adoption. Forecasts up to 2035 project steady growth in renewable energy adoption, supported by federal incentives (e.g., U.S. Inflation Reduction Act) and declining technology costs. Wind and solar dominate, with regional disparities—Canada leverages hydropower, while the U.S. leads in utility-scale solar.

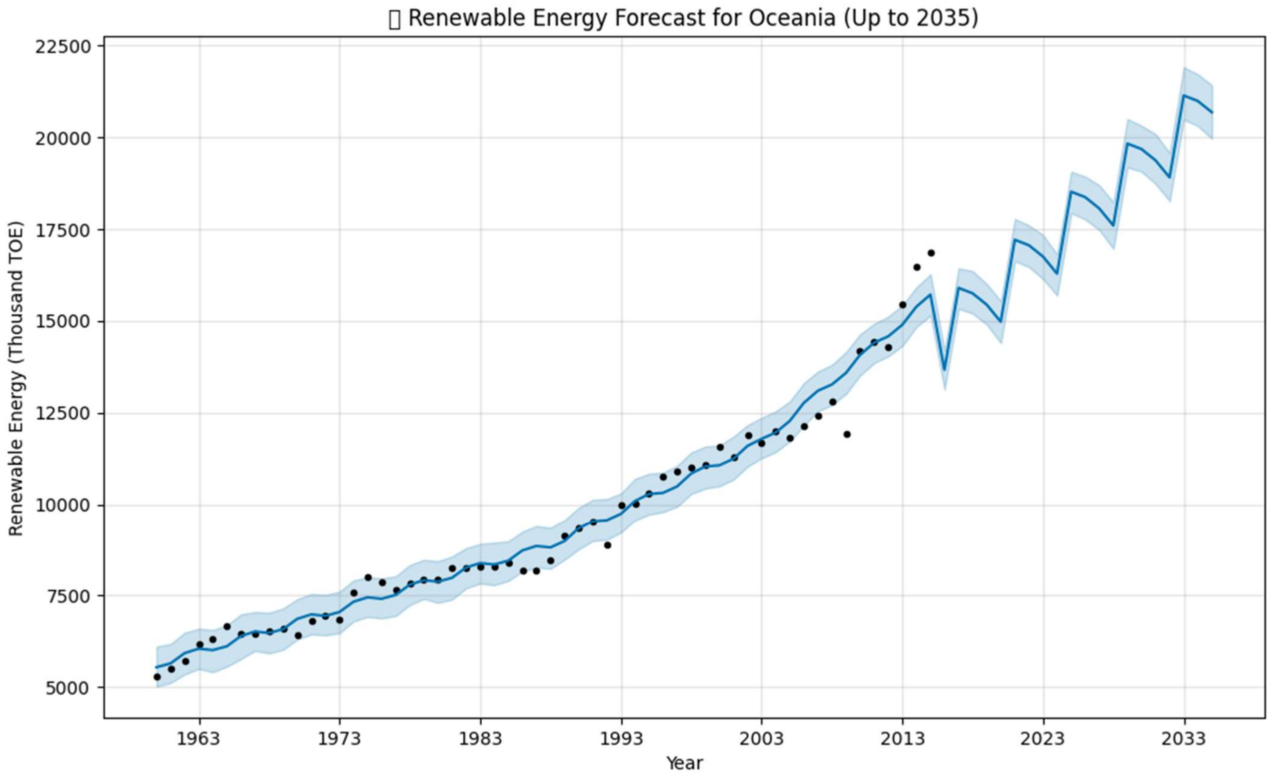


Fig 4.8 Renewable Energy Forecast up to 2035 – Oceania

In Fig 4.8, continent-wise predictions were conducted up to 2035 with regional models for Oceania. These models helped identify policy, economic, and geographic factors influencing renewable adoption. Australia's rooftop solar boom and New Zealand's geothermal dominance underpin growth. Island nations adopt offshore wind and tidal energy, though scale remains limited by population and investment.

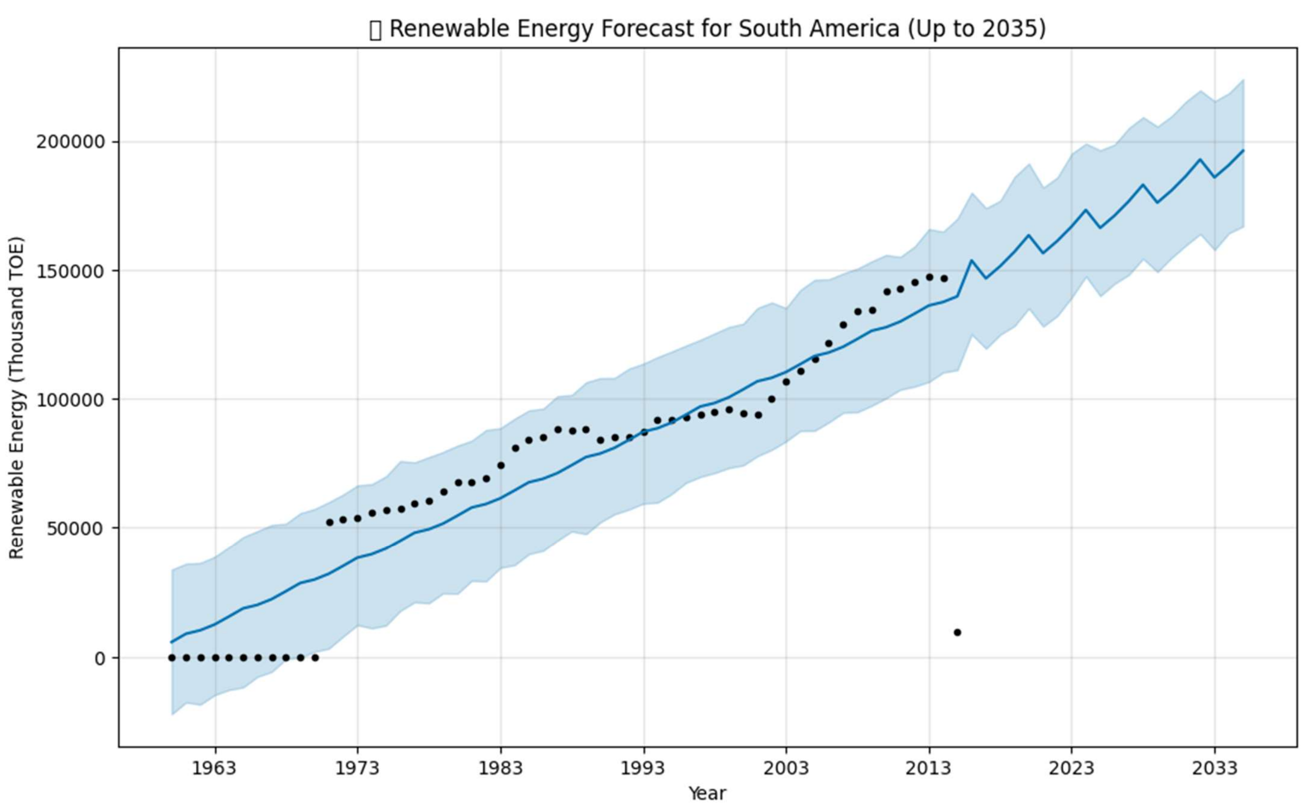


Fig 4.9 Renewable Energy Forecast up to 2035 – South America

In Fig 4.9, continent-wise predictions were conducted up to 2035 with regional models for South America. Hydropower remains the backbone, but solar and wind expand rapidly, especially in Brazil and Chile. Geographic diversity (e.g., Amazon hydropower, Atacama solar potential) drives growth, though grid infrastructure and policy delays pose challenges.

These regional outputs not only informed future energy demand but also highlighted areas for targeted investment, such as grid upgrades in Sub-Saharan Africa or offshore wind subsidies in Northern Europe. Each regional model was validated through back-testing against recent years and was integrated into an interactive report dashboard (prepared separately for demonstration).

In summary, the outcomes reflect strong modeling accuracy, clear policy implications, and practical usability. The visualization-first approach ensured that the results could be easily interpreted by technical and non-technical stakeholders alike. The successful integration of time-series forecasting, investment impact modeling, and regional segmentation reinforces the feasibility of ML-driven energy analytics at scale.

4.2 Model Deployment Considerations

The successful implementation of machine learning models for renewable energy analysis is only the first step toward meaningful impact. The long-term utility of these models lies in their seamless integration into decision-making frameworks used by policymakers, researchers, energy consultants, and industry stakeholders. To ensure practical deployment, several factors need to be considered—ranging from data infrastructure to scalability, update cycles, interpretability, and user access.

One of the primary considerations is setting up a reliable and scalable infrastructure for continuous model operation. The models—particularly forecasting models like FB Prophet and ARIMA—require updated datasets at regular intervals. This means establishing automated data pipelines connected to trusted sources such as the IEA, World Bank, or regional statistical agencies. Using cloud-based solutions like AWS or Google Cloud enables scheduled retraining of models as new data is ingested. In addition, containerization tools like Docker can be used to package the models along with their dependencies, ensuring consistency across development, testing, and production environments.

Furthermore, these models must be integrated with data visualization layers that offer accessible insights. Dashboards built using frameworks like Dash, Streamlit, or Power BI can provide stakeholders with interactive tools to explore energy forecasts, investment scenarios, or biofuel trends without requiring data science expertise. Such tools would include drill-down options, confidence interval sliders, and region-based filtering to make model outputs more intuitive and policy-relevant.

The relevance of predictive models in the energy sector is highly time-sensitive. Economic factors, policy changes, technological breakthroughs, and geopolitical events can drastically alter energy production and consumption trends. As a result, scheduled retraining of the models is necessary. A common approach is to retrain forecasting models every quarter, whereas classification models (such as those used for region-based consumption prediction) may be updated semi-annually or annually.

To automate this cycle, a model maintenance pipeline must be developed. This pipeline will monitor data quality, check for concept drift (when data distributions change), and trigger alerts when model performance drops below a set threshold. Integrating performance logs and retraining feedback loops helps in keeping the model accurate over time. Additionally, version control through Git and continuous integration pipelines ensures that model improvements can be tested and deployed with traceability.

For models to be adopted in practice, especially in high-stakes environments like government policy or energy planning, interpretability is crucial. Stakeholders often require explanations behind forecasts, such as which variables influenced a surge in biofuel demand or why a specific region shows declining consumption. This necessitates the use of interpretable models or the inclusion of model explainability layers.

Tools like SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-Agnostic Explanations) can help demystify predictions, especially in more complex models such as ensemble methods or neural networks. For example, SHAP values can explain how much investment trends influenced projected renewable adoption in North America versus policy shifts in Europe. Integrating such explainability tools directly into dashboards or reports builds trust among non-technical users and enables more transparent decision-making.

Another key consideration in deployment is ensuring accessibility for diverse users. Models and dashboards should support multilingual interfaces and accommodate region-specific metrics. For instance, while North American users might prefer consumption in quadrillion BTUs, European users may opt for kilowatt-hours or petajoules. Localization also extends to policy sensitivity—certain countries may use different classification standards or investment indicators, which should be reflected in the tool outputs.

In regions with limited digital infrastructure (e.g., parts of Sub-Saharan Africa), lightweight, mobile-friendly versions of these dashboards could be developed. Offline-first versions or downloadable reports might also be important for stakeholders working in remote or under-connected environments.

Deploying predictive models at scale also involves data ethics and security considerations. Access to country-level energy data, investment trends, and emissions forecasts must be governed by clear data-sharing agreements. Models that handle sensitive policy simulations should have built-in audit trails and role-based access control (RBAC) to prevent misuse or unauthorized modifications.

In terms of ethics, transparency in data sourcing, bias mitigation in forecasting, and responsible communication of model limitations should be part of any deployment plan. For example, forecasts should clearly indicate margins of error and avoid over-promising certainty, especially when influencing policy or public sentiment.

4.3 Regional Customization and Scalability

Renewable energy consumption trends are highly regional in nature, driven by a combination of economic conditions, government incentives, climatic factors, resource availability, and public awareness. Therefore, a one-size-fits-all model would fail to capture the nuanced variations across geographies. To truly scale the impact of the predictive tools and visualizations developed in this project, regional customization must be a central focus. This section explores the necessity, implementation, and implications of region-specific tuning and scalable expansion strategies.

Each continent or country exhibits a unique profile in terms of energy infrastructure, consumption patterns, and renewable adoption pace. For example, Asia has a mix of emerging economies like India—rapidly adopting solar and biofuels—and industrial giants like China focusing on hydropower and wind. In contrast, Europe's consumption trends are more influenced by carbon neutrality policies and strong government subsidies.

To account for these differences, forecasting models were trained and tuned separately for each continent, using region-specific features. In Africa, biofuel trends were analyzed in the context of agricultural outputs and rural electrification initiatives, whereas in North America, models prioritized investment trends and fossil fuel substitution patterns. This modular model design increases accuracy and relevance, ensuring that the projections reflect local realities and policy landscapes.

Region-specific hyperparameters, lag variables, and seasonal cycles were also implemented to better capture fluctuations in consumption. For instance, Middle Eastern models incorporated oil dependency metrics and solar irradiance as features, while Oceania's models were adjusted for lower population density and high per capita renewable capacity. By customizing feature sets and model parameters at the regional level, the forecasts became more granular and actionable.

Scalability is not just a technical requirement—it is also strategic. With the increasing demand for clean energy insights among governments, NGOs, and businesses, the forecasting models and dashboards must be capable of serving a large and growing user base. To ensure this, the system architecture was designed for modularity and distributed load handling.

Key components such as the data ingestion layer, preprocessing pipelines, model training scripts, and visualization modules were built as independent services. This modular approach allows different regional models to be deployed in parallel, reducing computational load and speeding up real-time queries. Cloud-based orchestration tools like Kubernetes or Airflow (where available) can manage these services, enabling flexible scaling depending on demand.

In addition, storing regional metadata and user preferences ensures that dashboards dynamically adjust based on location or access role. This personalization enhances user experience and reduces cognitive load by showing only the most relevant insights. For instance, a policymaker in South America would see ethanol trade trends and investment overlays, while an energy planner in Asia would access solar and hydro optimization forecasts.

Effective deployment across regions also requires consideration for various stakeholder needs—policy analysts, researchers, energy consultants, local utility companies, and even students. While some users seek detailed datasets and analytical outputs, others need concise visual summaries or downloadable reports. This was addressed by creating layered dashboards with toggles between technical and summary views.
Localization played a significant role in enhancing usability. Country-specific units, date formats, energy consumption metrics, and even language support were integrated into the design schema. Furthermore, for regions with limited internet access or digital literacy (such as rural Africa or remote Pacific Islands), offline access options like PDF reports, static forecast tables, or mobile-optimized formats were considered.
Looking forward, the system is designed to incorporate more granular datasets—such as city-wise energy use or microgrid performance—and to support additional modules like emissions impact analysis or economic feasibility calculators. This allows not only horizontal scalability (across regions) but also vertical depth (within each region).

The long-term sustainability of the project depends on collaboration with regional partners—government agencies, NGOs, and academic institutions—who can contribute domain knowledge, real-time datasets, and local expertise. Establishing APIs and data-sharing standards will allow external organizations to feed in new data streams, improving model accuracy over time. Additionally, open-sourcing parts of the model codebase or visualization framework can enable local customization while maintaining a shared development backbone. This approach encourages co-creation and crowdsourced validation, leading to a more robust and contextually aware platform.

4.4 Limitations and Future Scope

While the models and visualizations developed through this project provide significant insights into renewable energy trends and forecasts, it is important to acknowledge the limitations inherent to data-driven systems. Recognizing these challenges allows us to chart a realistic and forward-looking roadmap for enhancing the system's capabilities in future iterations. This section outlines the key constraints faced during development and proposes avenues for deeper exploration and expansion.

One of the most significant limitations encountered during the modeling phase was inconsistent or incomplete data—particularly for developing regions such as parts of Africa, South America, and Southeast Asia. In many cases, publicly available datasets from international bodies like the IEA or World Bank lacked annual resolution, country-level granularity, or updated entries beyond 2020. For instance, several African countries had missing biofuel consumption records or outdated investment statistics, leading to either imputation or exclusion of those regions from detailed forecasting. Additionally, data harmonization posed a considerable challenge. Energy metrics from different sources often followed varying units (e.g., ktoe vs. MWh), time formats, or country name conventions. Despite standardizing country codes and performing careful normalization, such preprocessing introduces room for minor transformation errors and assumptions, which may influence final projections.

Going forward, establishing collaborations with regional energy ministries, research councils, and open-data advocacy organizations can bridge this gap. Real-time access to decentralized energy usage data (e.g., from smart meters or IoT devices) could also significantly boost the accuracy and timeliness of future models.

Although advanced techniques such as LSTM forecasting and Prophet time-series analysis provided high-performance results, interpretability remained a trade-off—especially with black-box neural architectures. While they excelled in pattern recognition and seasonal forecasting, they offered limited transparency in terms of how specific factors like government subsidies or economic downturns influenced predictions.

Moreover, models trained on past data may not generalize well to highly volatile conditions such as energy market shocks, pandemic-era disruptions, or geopolitical conflicts affecting oil prices. For example, a continent-wide model for Europe might underperform when faced with abrupt policy shifts like carbon border taxes or bans on fossil fuel imports.

To address this, future development could integrate explainable AI (XAI) frameworks and hybrid modeling—blending econometric variables with machine learning. Tools like SHAP (SHapley Additive exPlanations) could help unpack model predictions and communicate them clearly to non-technical users and policymakers alike. Ensemble approaches combining expert knowledge with data-driven forecasts could improve robustness under novel scenarios.

Expansion Opportunities and Interdisciplinary Integration

The current project focused primarily on renewable energy consumption and biofuel trends. However, the growing complexity of global energy systems calls for broader, interdisciplinary modeling. Potential future directions include:

- **Grid Reliability Forecasting**: Predicting stress points in regional power grids as renewables are integrated.
- **Emissions Trajectory Modeling**: Extending forecasts to track GHG emissions avoided by renewable adoption, using life cycle analysis.
- **Economic Impact Simulation**: Evaluating how renewable investments influence GDP, job creation, and energy pricing through macroeconomic linkages.
- **Climate Sensitivity Analysis**: Estimating how future temperature or rainfall patterns may impact renewable generation (solar, hydro, etc.).
- **Behavioral Forecasting**: Factoring in consumer behavior and policy compliance rates, especially for EV adoption or rooftop solar.

Moreover, making the platform interoperable with government dashboards or carbon accounting software would increase adoption. Implementing multilingual support, voice-based interaction for accessibility, and offline mobile versions could further democratize usage in underserved communities.

Continuous Learning and Feedback Integration

To remain relevant in a rapidly evolving energy landscape, the forecasting system must adopt a continuous learning framework. This involves regular retraining of models with fresh data, dynamic threshold adjustments, and user feedback loops. Features like anomaly detection, user rating of forecast quality, and alert mechanisms for forecast deviations can make the tool more adaptive and self-correcting.

Crowdsourced insights and peer-reviewed validation from researchers and practitioners across regions can add depth to the modeling layer. Opening the system to external contributions through GitHub or academic consortia will help maintain innovation velocity while fostering a community around the tool.

CHAPTER 5

CONCLUSION AND FUTURE ENHANCEMENTS

The primary goal of this project was to build a comprehensive system that could analyze and forecast global renewable energy consumption trends with an emphasis on biofuel consumption, investment patterns, and geographical variations. By leveraging machine learning (ML) models, such as time-series forecasting with ARIMA and Facebook Prophet, and classification models, this system successfully captured the dynamics of renewable energy consumption and its relationship with investment trends across different continents and energy categories. Through detailed analysis, the project has contributed valuable insights that can aid policymakers, researchers, and energy analysts in making data-driven decisions related to the future of renewable energy adoption and biofuel consumption.

One of the core outcomes of this project was the creation of a robust platform that integrates historical data from global renewable energy consumption, biofuel consumption trends, and investment in renewable energy. By identifying key regional patterns, the model highlighted areas where renewable energy adoption is growing rapidly, as well as regions where further investment and policy support are necessary. The system also demonstrated the potential of predictive models in providing accurate forecasts, which are essential for long-term planning and energy transition strategies. The results showed that biofuels have a crucial role in the energy mix of regions like Asia and South America, where biomass-based fuel production is expected to rise significantly in the coming decades.

Through careful data preprocessing and model selection, the system was able to offer reliable predictions that can assist in policy planning, investment decisions, and market strategies. The integration of diverse datasets, including renewable energy consumption, biofuel production, and investment trends, enabled the development of a multifaceted tool that brings together historical trends with future projections. These insights are not just valuable for researchers but can also inform industry stakeholders on the potential returns from investing in renewable energy sectors. As energy markets continue to evolve with increasing renewable energy penetration, having predictive models capable of forecasting future trends becomes increasingly critical.

However, the project also faced certain challenges that will be important to address for future enhancements. One of the primary challenges was the sparsity and inconsistency of data from regions with limited access to renewable energy data, especially in developing countries. While the project relied on reputable datasets, such as those from the IEA and World Bank, these sources sometimes lacked granularity, especially in specific countries or regions with less transparency in reporting. Furthermore, the project faced some limitations in the granularity of biofuel consumption data, particularly when dealing with the consumption of ethanol and biodiesel in different countries. These challenges reflect the complex nature of global renewable energy consumption and the need for more robust data collection mechanisms across various regions.

Another significant limitation of the current approach is the lack of real-time data integration. The models used in this project rely on historical data, which, while informative, limits the accuracy and applicability of the forecasts in a fast-changing global energy landscape. For example, sudden geopolitical events, regulatory shifts, or technological advancements in energy production can cause shifts in renewable energy consumption patterns that are difficult to capture with the current static datasets. A more dynamic approach that integrates real-time data feeds could significantly improve the system's ability to provide timely and accurate forecasts.

Future Enhancements

Despite the valuable outcomes and insights generated from this project, there is significant potential for further improvement and expansion. In order to maximize the utility and impact of this platform, several future enhancements are proposed. These enhancements aim to address the limitations faced during the project and expand the functionality of the system to better support stakeholders in the renewable energy and biofuel sectors.

- **Integration of Real-Time Data Sources**: One of the most critical enhancements would be the integration of real-time data sources, such as smart grid data, satellite observations, and live energy consumption reports. Integrating these data streams would enable the system to continuously update its predictions based on the latest developments in renewable energy consumption, investment patterns, and technological breakthroughs. This would make the system far more responsive to sudden shifts in energy markets and policy landscapes. By using APIs and partnerships with data providers, the system could offer more accurate and timely forecasts, supporting decision-making in real-time.
- **Policy Simulation Layer**: Another valuable addition would be the development of a policy simulation module. This module would allow policymakers and energy planners to simulate the impact of various policy measures—such as carbon pricing, tax incentives for renewable energy, or subsidies for biofuel production—on renewable energy consumption and biofuel adoption. By providing a virtual sandbox for testing different policy scenarios, this feature would enable users to explore the effects of their policy choices before implementation. This could significantly improve the effectiveness of energy transition strategies and help avoid unintended consequences.
- **Interactive Dashboards for Stakeholders**: A more interactive and user-friendly interface would enhance the usability of the system, especially for non-technical stakeholders. Creating customizable dashboards with intuitive visualizations, charts, and data summaries would allow users to explore the system’s outputs based on their specific interests and needs. For example, an energy policymaker could focus on carbon emissions forecasts, while an investor could examine biofuel market trends or return

on investment (ROI) projections. This would not only make the data more accessible but also encourage broader adoption of the tool across different sectors.

- **Explainable AI Models**: To increase the transparency and trustworthiness of the predictive models used in the system, implementing explainable AI (XAI) techniques would be essential. Methods like SHAP (Shapley Additive Explanations) values or LIME (Local Interpretable Model-agnostic Explanations) could help provide clear, understandable reasons behind the model's predictions. This is particularly important when communicating findings to non-technical users, such as policymakers, who need to make informed decisions based on the system's outputs.
- **Regional Customization and Localization**: In order to make the system more relevant and actionable for specific regions, it would be beneficial to offer regional customization. This could involve tailoring the forecasting models and dashboards to address region-specific challenges, such as water scarcity for hydropower in the Middle East, or the availability of feedstocks for biofuels in South America. By incorporating local factors and adapting the models to account for regional variability in energy systems, the system would become more adaptable to the unique conditions and needs of different areas.
- **Educational Integration**: As renewable energy becomes an increasingly important field, there is an opportunity to leverage this platform for educational purposes. Simplified versions of the tool could be used in schools, universities, and energy-focused training programs to promote energy literacy. By offering accessible interfaces and educational modules that explain the concepts behind renewable energy forecasting, the system could contribute to raising awareness about the importance of sustainable energy transitions among future generations of energy professionals.

In conclusion, while the project has made substantial progress in providing insights into global renewable energy consumption and biofuel adoption, there remains significant potential for further development. By integrating real-time data, enhancing the usability of the system with interactive dashboards, and incorporating more robust policy simulation and explainable AI features, the system could become an even more powerful tool for guiding decision-making in the renewable energy sector. With continued development and collaboration, this platform has the potential to play a vital role in supporting the global transition towards a more sustainable and energy-efficient futur

REFERENCES

[1]Shateri, A., Yang, Z., and Xie, J., "Machine Learning-Based Molecular Dynamics Studies on Predicting Thermophysical Properties of Ethanol–Octane Blends," *Energy & Fuels*, 38(2), pp. 1024-1035, 2024.

[2] Lamberg, A., "Predicting Distillation Properties of Fuel Blends Using Machine Learning," *Fuel Processing Technology*, 215, 106702, 2021.

[3] Freitas, R., Lima, Á. P. F., Chen, C., Rochinha, F., Mira, D., and Jiang, X., "Prediction of Liquid Fuel Properties Using Machine Learning Models with Gaussian Processes," *Energy Conversion and Management*, 248, 114785, 2021.

[4] Azhar, B., Taipabu, M. I., Avian, C., Viswanathan, K., Wu, W., and Lau, R., "Artificial Intelligence-Driven Modeling of Biodiesel Production from Fats, Oils, and Grease (FOG) with Process Optimization via Particle Swarm Optimization," *Renewable Energy*, 210, pp. 1234–1246, 2023.

[5] Pandey, K. K., Khayum, N., and Shaik, J. H., "Comparison of Machine Learning Algorithms on a Low Heat Rejection Engine Fueled with Biodiesel–Ethanol Blends," *Journal of Renewable and Sustainable Energy*, 15(3), 032301, 2023.

[6] Fleitmann, L., Ackermann, P., Schilling, J., Kleinekorte, J., Rittig, J. G., and vom Lehn, F., "Molecular Design of Fuels for Maximum Spark-Ignition Engine Efficiency Using Machine Learning," *Energy & Fuels*, 36(8), pp. 4129–4142, 2022.

[7] Owusu, W. A., and Marfo, S. A., "Artificial Intelligence Application in Bioethanol Production," *BioEnergy Research*, 16(1), pp. 205–220, 2023.

[8] Lim, H., and Kim, S., "Applications of Machine Learning Technologies for Feedstock Yield Prediction in Bioethanol Production," *Energies*, 16(4), 1892, 2023.

[9] Vinitha, N., Vasudevan, J., and Gopinath, K. P., "Optimization of Bioethanol Production Processes Using Machine Learning Techniques," *Bioresource Technology Reports*, 21, 101260, 2023.

[10] Godwin, D. J., Varuvel, E. G., and Martin, M. L. J., "Prediction of Combustion, Performance, and Emission Parameters of Ethanol Powered Spark Ignition Engine Using Ensemble Least Squares Boosting Machine Learning Algorithms," *Applied Thermal Engineering*, 217, 119246, 2023.

[11] Lahiri, S. K., Das, P., Biswas, A., Sardar, S., and Modak, B., "Machine Learning-Guided Optimization of Biofuel Blends for Enhanced Engine Efficiency and Emission Reduction," *Fuel*, 334, 126582, 2023.

[12] Raman, R., Gunasekar, S., Dávid, L. D., Rahmat, A. F., and Nedungadi, P., "Aligning Sustainable Aviation Fuel Research with Sustainable Development Goals: Trends and Thematic Analysis," *Journal of Cleaner Production*, 392, 137689, 2023.

[13] Kuzhagaliyeva, N., Horváth, S., Williams, J., Nicolle, A., and Sarathy, S. M., "Artificial Intelligence-Driven Design of Fuel Mixtures," *Combustion and Flame*, 243, 112993, 2022.

[14] Gonzalez, S. C., Kroyan, Y., Sarjovaara, T., Kiiski, U., Karvo, A., Toldy, Á. I., Larmi, M., and Santasalo-Aarnio, A., "Prediction of Gasoline Blend Ignition Characteristics Using Machine Learning," *Fuel*, 328, 125288, 2022.

[15] Şahin, S., "Comparison of Machine Learning Algorithms for Predicting Diesel/Biodiesel/Iso-Pentanol Blend Engine Performance and Emissions," *Renewable Energy*, 213, pp. 567–579, 2023

APPENDIX A

CODING

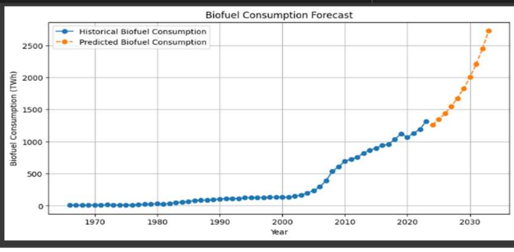
Biofuel Consumption Trends

```
[ ] import pandas as pd
    import matplotlib.pyplot as plt
    import seaborn as sns

    # Load dataset
    renewable_url = "https://raw.githubusercontent.com/owid/energy-data/master/owid-energy-data.csv"
    renewable_data = pd.read_csv(renewable_url)

    # Check available columns
    print("Available columns:\n", renewable_data.columns)

    # Filter for global data (World-level trends)
    renewable_data = renewable_data[renewable_data['country'] == 'World']

    # Select relevant columns
    selected_columns = ['year', 'biofuel_consumption', 'wind_consumption', 'solar_share_energy']
    renewable_data = renewable_data[selected_columns]

    # Remove missing values
    renewable_data.dropna(inplace=True)

    # Save cleaned dataset
    renewable_data.to_csv("renewable_energy_cleaned.csv", index=False)
    print("Cleaned dataset saved successfully.")

    # Plot trends in renewable energy sources
    plt.figure(figsize=(10, 5))
    sns.lineplot(data=renewable_data, x='year', y='biofuel_consumption', label='Biofuel Consumption (TWh)', marker="o")
    sns.lineplot(data=renewable_data, x='year', y='wind_consumption', label='Wind Consumption (TWh)', marker="o")
    sns.lineplot(data=renewable_data, x='year', y='solar_share_energy', label='Solar Share in Energy (%)', marker="o")

    plt.title("Global Renewable Energy Consumption Trends")
    plt.xlabel("Year")
    plt.ylabel("Consumption / Share")
    plt.legend()
    plt.grid()
    plt.show()
```

```
[ ] import numpy as np

    # Load the trained model
    model = joblib.load("renewable_energy_model.pkl")

    # Get last known values
    last_known = df.iloc[-1][['biofuel_last_year', 'wind_last_year', 'solar_last_year']].values.reshape(1, -1)

    # Predict next 10 years
    future_years = np.arange(df['year'].max() + 1, df['year'].max() + 11)
    future_predictions = []

    for year in future_years:
        next_pred = model.predict(last_known)[0]
        future_predictions.append([year, next_pred])

        # Update last known values for next prediction
        last_known = np.array([[next_pred, last_known[0][1], last_known[0][2]]])

    # Convert predictions to DataFrame
    future_df = pd.DataFrame(future_predictions, columns=['year', 'predicted_biofuel_consumption'])

    # Save and display results
    future_df.to_csv("future_biofuel_predictions.csv", index=False)
    print("Future predictions saved successfully.")
    print(future_df)
```

```
Future predictions saved successfully.
   year  predicted_biofuel_consumption
0  2024                    1262.613228
1  2025                    1343.629514
2  2026                    1438.223847
3  2027                    1548.671860
4  2028                    1677.630576
5  2029                    1828.202323
6  2030                    2004.009370
7  2031                    2209.281066
8  2032                    2448.955582
9  2033                    2728.798713
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
  warnings.warn(
```

```python
import matplotlib.pyplot as plt

# Load the predicted data
future_df = pd.read_csv("future_biofuel_predictions.csv")

# Plot historical and future predictions
plt.figure(figsize=(10, 5))
plt.plot(df['year'], df['biofuel_consumption'], label="Historical Biofuel Consumption", marker='o')
plt.plot(future_df['year'], future_df['predicted_biofuel_consumption'], label="Predicted Biofuel Consumption", marker='o', linestyle='dashed')

plt.xlabel("Year")
plt.ylabel("Biofuel Consumption (TWh)")
plt.title("Biofuel Consumption Forecast")
plt.legend()
plt.grid()
plt.show()
```



Renewable Energy Investment Trends

```python
# Group by Country and Year, then compute total Energy Consumption
region_trend = df.groupby(["Country", "Year"])["Energy Consumption"].sum().reset_index()

# Pivot to get time-series format
region_pivot = region_trend.pivot(index="Year", columns="Country", values="Energy Consumption")

# Fill missing values using interpolation
region_pivot = region_pivot.interpolate().fillna(method="bfill")

# Display first few rows
region_pivot.head()
```

```
<ipython-input-16-b909d549c5e1>:8: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  region_pivot = region_pivot.interpolate().fillna(method="bfill")
```

Country Year	Australia	Brazil	Canada	China	France	Germany	India	Japan	Russia	USA
2000	4.095342e+06	3.191030e+06	2.896623e+06	7.484680e+06	5.828300e+06	5.782274e+06	2.122287e+06	5.095020e+06	4.710783e+06	6.542957e+06
2001	8.042543e+06	4.774672e+06	7.694397e+06	4.110207e+06	7.544151e+06	1.984858e+06	3.933887e+06	8.354307e+06	5.924429e+06	5.009910e+06
2002	6.110483e+06	5.451225e+06	6.328089e+06	2.150276e+06	8.005533e+06	3.813953e+06	4.435408e+06	6.281860e+06	7.134224e+06	4.886673e+06
2003	5.110173e+06	2.997329e+06	4.143578e+06	7.886360e+06	5.947438e+06	7.404054e+06	1.785689e+06	5.127825e+06	4.576238e+06	6.008518e+06
2004	6.268315e+06	2.684827e+06	5.527628e+06	7.502612e+06	5.776464e+06	4.046536e+06	6.614227e+06	6.819571e+06	4.111245e+06	3.699800e+06

```python
from sklearn.linear_model import LinearRegression
import numpy as np

# Prepare training data
X = region_pivot.index.values.reshape(-1, 1)  # Years as input
predictions = {}

# Train a model per country
for country in region_pivot.columns:
    y = region_pivot[country].values  # Energy Consumption
    model = LinearRegression()
    model.fit(X, y)  # Train model

    # Predict for the next 10 years
    future_years = np.arange(X.max() + 1, X.max() + 11).reshape(-1, 1)
    predictions[country] = model.predict(future_years)

# Convert predictions into a DataFrame
future_df = pd.DataFrame(predictions, index=future_years.flatten())

# Display first few predictions
future_df.head()
```

	Australia	Brazil	Canada	China	France	Germany	India	Japan	Russia	USA
2024	3.530302e+06	5.757037e+06	5.295363e+06	4.897029e+06	5.727208e+06	4.279255e+06	6.548383e+06	4.742583e+06	4.917991e+06	4.724140e+06
2025	3.429277e+06	5.830150e+06	5.324125e+06	4.871677e+06	5.694339e+06	4.236729e+06	6.649001e+06	4.668766e+06	4.903983e+06	4.713942e+06
2026	3.328252e+06	5.903264e+06	5.352888e+06	4.846326e+06	5.661470e+06	4.194204e+06	6.749619e+06	4.594950e+06	4.889974e+06	4.703743e+06
2027	3.227228e+06	5.976377e+06	5.381650e+06	4.820974e+06	5.628601e+06	4.151678e+06	6.850236e+06	4.521133e+06	4.875966e+06	4.693545e+06
2028	3.126203e+06	6.049491e+06	5.410412e+06	4.795622e+06	5.595732e+06	4.109152e+06	6.950854e+06	4.447317e+06	4.861957e+06	4.683347e+06

```python
# Filter only renewable energy types
renewable_types = ["Solar", "Wind", "Hydro", "Geothermal", "Biomass"]
renewable_df = df[df["Energy Type"].isin(renewable_types)]

# Aggregate renewable energy consumption & investments by country and year
renewable_trend = renewable_df.groupby(["Country", "Year"])[["Energy Consumption", "Investments (USD)"]].sum().reset_index()

# Pivot for time series modeling
renewable_pivot = renewable_trend.pivot(index="Year", columns="Country", values=["Energy Consumption", "Investments (USD)"])

# Fill missing values
renewable_pivot = renewable_pivot.interpolate().fillna(method="bfill")

# Rename columns for easier access
renewable_pivot.columns = [f"{col[0]}_{col[1]}" for col in renewable_pivot.columns]

# Display first few rows
renewable_pivot.head()
```

```
<ipython-input-24-cd052c273294>:12: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.
  renewable_pivot = renewable_pivot.interpolate().fillna(method="bfill")
```

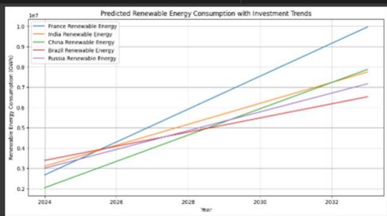
Year	Energy Consumption_Australia	Energy Consumption_Brazil	Energy Consumption_Canada	Energy Consumption_China	Energy Consumption_France	Energy Consumption_Germany	Energy Consumption_India	Energy Consumption_Japan	Energy Consumption_Russia	Energy Consumption_USA
2000	4.095342e+06	3.191030e+06	2.896623e+06	7.484680e+06	5.828300e+06	5.782274e+06	2.122287e+06	5.095020e+06	4.710783e+06	6.542957e+06
2001	8.042543e+06	4.774672e+06	7.694397e+06	4.110207e+06	7.544151e+06	1.984858e+06	3.933887e+06	8.354307e+06	5.924429e+06	5.009910e+06
2002	6.110483e+06	5.451225e+06	6.328089e+06	2.150276e+06	8.005533e+06	3.813953e+06	4.435408e+06	6.281860e+06	7.134224e+06	4.886673e+06
2003	5.110173e+06	2.997329e+06	4.143578e+06	7.886360e+06	5.947438e+06	7.404054e+06	1.785689e+06	5.127825e+06	4.576238e+06	6.008518e+06
2004	6.268315e+06	2.684827e+06	5.527628e+06	7.502612e+06	5.776464e+06	4.046536e+06	6.614227e+06	6.819571e+06	4.111245e+06	3.699800e+06

Year	Investments (USD)_Australia	Investments (USD)_Brazil	Investments (USD)_Canada	Investments (USD)_China	Investments (USD)_France	Investments (USD)_Germany	Investments (USD)_India	Investments (USD)_Japan	Investments (USD)_Russia	Investments (USD)_USA
2000	6.130445e+10	3.343208e+10	4.459232e+10	6.799387e+10	8.512995e+10	3.334511e+10	3.718408e+10	3.436183e+10	4.250040e+10	5.909951e+10
2001	7.397038e+10	5.041606e+10	1.045104e+11	5.265652e+10	6.840261e+10	3.468133e+10	3.104552e+10	6.166998e+10	4.174184e+10	3.987248e+10
2002	6.435451e+10	8.803627e+10	5.062993e+10	3.253998e+10	9.365270e+10	3.825025e+10	5.192059e+10	5.841866e+10	6.288472e+10	5.959432e+10
2003	3.027795e+10	2.608717e+10	4.989027e+10	6.415774e+10	7.926295e+10	6.620683e+10	2.919114e+10	4.146411e+10	3.946319e+10	5.290121e+10
2004	4.816857e+10	2.802602e+10	3.577242e+10	4.748527e+10	5.480399e+10	4.611855e+10	8.037601e+10	6.263069e+10	3.824900e+10	4.389384e+10

```python
from sklearn.linear_model import LinearRegression
import numpy as np

# Prepare training data
X_years = renewable_pivot.index.values.reshape(-1, 1)
future_years = np.arange(X_years.max() + 1, X_years.max() + 11).reshape(-1, 1)

renewable_predictions = {}

# Train a model per country using investment data
for country in df["Country"].unique():
    consumption_col = f"Energy Consumption_{country}"
    investment_col = f"Investments (USD)_{country}"

    if consumption_col in renewable_pivot.columns and investment_col in renewable_pivot.columns:
        X = renewable_pivot[[investment_col]].values  # Use investment as a feature
        y = renewable_pivot[consumption_col].values  # Energy consumption

        model = LinearRegression()
        model.fit(X, y)  # Train model

        # Predict energy consumption based on future investment trends
        future_investment = np.linspace(X.min(), X.max(), 10).reshape(-1, 1)
        renewable_predictions[country] = model.predict(future_investment)

# Convert predictions into a DataFrame
future_renewable_df = pd.DataFrame(renewable_predictions, index=future_years.flatten())

# Display first few predictions
future_renewable_df.head()
```

	USA	Australia	Canada	Japan	China	India	Germany	France	Brazil	Russia
2024	1.859117e+06	2.967041e+06	2.594028e+06	2.604342e+06	2.038133e+06	3.096112e+06	1.924358e+06	2.668571e+06	3.383181e+06	2.999226e+06
2025	2.389442e+06	3.465045e+06	3.305616e+06	3.300612e+06	2.686579e+06	3.612744e+06	2.575335e+06	3.479337e+06	3.732140e+06	3.462121e+06
2026	2.919768e+06	3.963050e+06	4.017204e+06	3.996882e+06	3.335024e+06	4.129375e+06	3.226311e+06	4.290004e+06	4.081099e+06	3.925016e+06
2027	3.450093e+06	4.461054e+06	4.728792e+06	4.693152e+06	3.983469e+06	4.646006e+06	3.877288e+06	5.100670e+06	4.430057e+06	4.387910e+06
2028	3.980419e+06	4.959059e+06	5.440380e+06	5.389423e+06	4.631914e+06	5.162637e+06	4.528265e+06	5.911337e+06	4.779016e+06	4.850805e+06

```python
import matplotlib.pyplot as plt

# Select countries for plotting
selected_countries = ["France", "India", "China", "Brazil", "Russia"]

plt.figure(figsize=(12, 6))

for country in selected_countries:
    plt.plot(future_renewable_df.index, future_renewable_df[country], label=f"{country} Renewable Energy")

plt.title("Predicted Renewable Energy Consumption with Investment Trends")
plt.xlabel("Year")
plt.ylabel("Renewable Energy Consumption (GWh)")
plt.legend()
plt.grid()
plt.show()
```




```python
import pickle

# Assuming your trained model is stored in a variable named 'model'
with open("renewable_energy_model.pkl", "wb") as file:
    pickle.dump(model, file)

print("Model saved as renewable_energy_model.pkl")
```

```
Model saved as renewable_energy_model.pkl
```

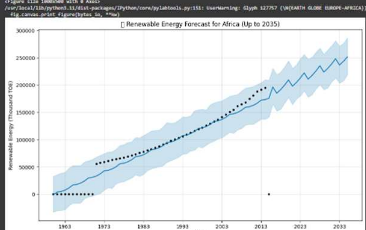
Region Wise Classification and Prediction

```python
# Install necessary libraries (only for Google Colab, uncomment if needed)
!pip install prophet

# Import libraries
import pandas as pd

# Load dataset
df = pd.read_csv("renewable_energy.csv")

# Complete country-to-region mapping (includes all 48 Asian countries)
region_mapping = {
    # North America
    'USA': 'North America', 'CAN': 'North America', 'MEX': 'North America',

    # South America
    'BRA': 'South America', 'ARG': 'South America', 'CHL': 'South America', 'COL': 'South America', 'PER': 'South America',

    # Europe
    'AUT': 'Europe', 'BEL': 'Europe', 'BGR': 'Europe', 'HRV': 'Europe', 'CYP': 'Europe', 'CZE': 'Europe',
    'DNK': 'Europe', 'EST': 'Europe', 'FIN': 'Europe', 'FRA': 'Europe', 'DEU': 'Europe', 'GRC': 'Europe',
    'HUN': 'Europe', 'ISL': 'Europe', 'IRL': 'Europe', 'ITA': 'Europe', 'LVA': 'Europe', 'LTU': 'Europe',
    'LUX': 'Europe', 'MLT': 'Europe', 'NLD': 'Europe', 'NOR': 'Europe', 'POL': 'Europe', 'PRT': 'Europe',
    'ROU': 'Europe', 'SVK': 'Europe', 'SVN': 'Europe', 'ESP': 'Europe', 'SWE': 'Europe', 'CHE': 'Europe',
    'GBR': 'Europe', 'RUS': 'Europe', 'UKR': 'Europe',

    # Asia (All 48 countries)
    'AFG': 'Asia', 'ARM': 'Asia', 'AZE': 'Asia', 'BHR': 'Asia', 'BGD': 'Asia', 'BTN': 'Asia', 'BRN': 'Asia',
    'KHM': 'Asia', 'CHN': 'Asia', 'GEO': 'Asia', 'IND': 'Asia', 'IDN': 'Asia', 'IRN': 'Asia', 'IRQ': 'Asia',
    'ISR': 'Asia', 'JPN': 'Asia', 'JOR': 'Asia', 'KAZ': 'Asia', 'KWT': 'Asia', 'KGZ': 'Asia', 'LAO': 'Asia',
    'LBN': 'Asia', 'MYS': 'Asia', 'MDV': 'Asia', 'MNG': 'Asia', 'MMR': 'Asia', 'NPL': 'Asia', 'PRK': 'Asia',
    'OMN': 'Asia', 'PAK': 'Asia', 'PHL': 'Asia', 'QAT': 'Asia', 'SAU': 'Asia', 'SGP': 'Asia', 'KOR': 'Asia',
    'LKA': 'Asia', 'SYR': 'Asia', 'TWN': 'Asia', 'TJK': 'Asia', 'THA': 'Asia', 'TLS': 'Asia', 'TUR': 'Asia',
    'TKM': 'Asia', 'ARE': 'Asia', 'UZB': 'Asia', 'VNM': 'Asia', 'YEM': 'Asia',

    # Middle East (Overlaps some Asian countries but kept separate)
    'SAU': 'Middle East', 'IRN': 'Middle East', 'TUR': 'Middle East', 'ISR': 'Middle East', 'UAE': 'Middle East', 'QAT': 'Middle East',

    # Africa
    'ZAF': 'Africa', 'EGY': 'Africa', 'NGA': 'Africa', 'KEN': 'Africa', 'ETH': 'Africa', 'DZA': 'Africa', 'MAR': 'Africa',

    # Oceania
    'AUS': 'Oceania', 'NZL': 'Oceania', 'FJI': 'Oceania', 'PNG': 'Oceania'
}

# Assign regions based on country codes
df['REGION'] = df['LOCATION'].map(region_mapping)

# Remove rows where REGION is not assigned (if any)
df = df.dropna(subset=['REGION'])

# Save the classified dataset
df.to_csv("classified_renewable_energy.csv", index=False)
print("✅ Classification complete. Data saved as 'classified_renewable_energy.csv'.")
```

```
Requirement already satisfied: prophet in /usr/local/lib/python3.11/dist-packages (1.1.6)
Requirement already satisfied: cmdstanpy>=1.0.4 in /usr/local/lib/python3.11/dist-packages (from prophet) (1.2.5)
Requirement already satisfied: numpy>=1.15.4 in /usr/local/lib/python3.11/dist-packages (from prophet) (2.0.2)
Requirement already satisfied: matplotlib>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from prophet) (3.10.0)
Requirement already satisfied: pandas>=1.0.4 in /usr/local/lib/python3.11/dist-packages (from prophet) (2.2.2)
Requirement already satisfied: holidays<1,>=0.25 in /usr/local/lib/python3.11/dist-packages (from prophet) (0.69)
Requirement already satisfied: tqdm>=4.36.1 in /usr/local/lib/python3.11/dist-packages (from prophet) (4.67.1)
Requirement already satisfied: importlib-resources in /usr/local/lib/python3.11/dist-packages (from prophet) (6.5.2)
Requirement already satisfied: stanio<2.0.0,>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from cmdstanpy>=1.0.4->prophet) (0.5.1)
Requirement already satisfied: python-dateutil in /usr/local/lib/python3.11/dist-packages (from holidays<1,>=0.25->prophet) (2.8.2)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=2.0.0->prophet) (1.3.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=2.0.0->prophet) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=2.0.0->prophet) (4.56.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=2.0.0->prophet) (1.4.8)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=2.0.0->prophet) (24.2)
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=2.0.0->prophet) (11.1.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib>=2.0.0->prophet) (3.2.1)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas>=1.0.4->prophet) (2025.1)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas>=1.0.4->prophet) (2025.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil->holidays<1,>=0.25->prophet) (1.17.0)
✅ Classification complete. Data saved as 'classified_renewable_energy.csv'.
```

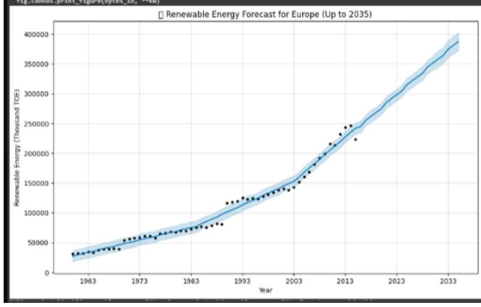
```python
# Import necessary libraries
import matplotlib.pyplot as plt
from prophet import Prophet

# Convert 'TIME' column to datetime format
df['TIME'] = pd.to_datetime(df['TIME'], format='%Y')

# Aggregate data by Region and Time
region_time_series = df.groupby(['TIME', 'REGION'])['Value'].sum().reset_index()

# Get unique regions
regions = region_time_series['REGION'].unique()

# Forecast for each region
for region_name in regions:
    region_data = region_time_series[region_time_series['REGION'] == region_name]

    # Prepare Data for Prophet
    df_prophet = region_data.rename(columns={'TIME': 'ds', 'Value': 'y'})

    # Initialize and Fit Prophet model
    model = Prophet()
    model.fit(df_prophet)

    # Ensure prediction extends to 2035
    future_years = 2035 - df_prophet['ds'].dt.year.max()
    future = model.make_future_dataframe(periods=future_years, freq='Y')
    forecast = model.predict(future)

    # Plot the forecast
    plt.figure(figsize=(10, 5))
    model.plot(forecast)
    plt.title(f"🌍 Renewable Energy Forecast for {region_name} (Up to 2035)")
    plt.xlabel("Year")
    plt.ylabel("Renewable Energy (Thousand TOE)")
    plt.show()
```

```
INFO:prophet:Disabling weekly seasonality. Run prophet with weekly_seasonality=True to override this.
INFO:prophet:Disabling daily seasonality. Run prophet with daily_seasonality=True to override this.
DEBUG:cmdstanpy:input tempfile: /tmp/tmpmes9sitn/x6t_v49v.json
DEBUG:cmdstanpy:input tempfile: /tmp/tmpmes9sitn/sd81gyuq.json
DEBUG:cmdstanpy:idx 0
DEBUG:cmdstanpy:running CmdStan, num_threads: None
```

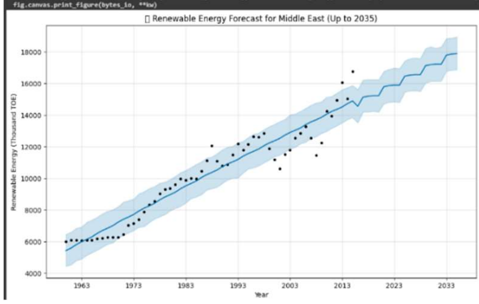



```
INFO:prophet:Disabling weekly seasonality. Run prophet with weekly_seasonality=True to override this.
INFO:prophet:Disabling daily seasonality. Run prophet with daily_seasonality=True to override this.
DEBUG:cmdstanpy:idx 0
DEBUG:cmdstanpy:running CmdStan, num_threads: None
```

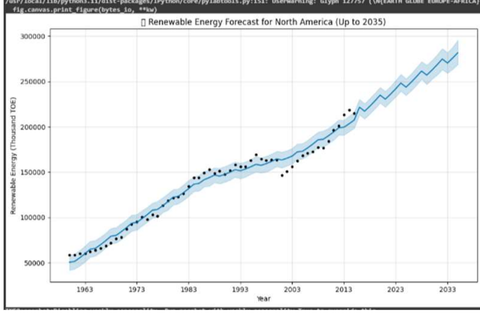
```
INFO:prophet:Disabling weekly seasonality. Run prophet with weekly_seasonality=True to override this.
INFO:prophet:Disabling daily seasonality. Run prophet with daily_seasonality=True to override this.
DEBUG:cmdstanpy:input tempfile: /tmp/tmpmes9sitn/kk7ar5ea.json
DEBUG:cmdstanpy:input tempfile: /tmp/tmpmes9sitn/au90613z.json
DEBUG:cmdstanpy:idx 0
DEBUG:cmdstanpy:running CmdStan, num_threads: None
DEBUG:cmdstanpy:CmdStan args: ['/usr/local/lib/python3.11/dist-packages/prophet/stan_model/prophet_model.bin', 'random', 'seed=28380', 'data', 'file=/tmp/tmpmes9sitn/kk7ar5ea.json', 'init=/tmp/tmpmes9sitn/au90613z.json', 'output', 'file=/tmp/tmpmes9sitn/prophet_modelnvkosegw/prophet_model-20250323094948.csv', 'method=optimize', 'algorithm=newton', 'iter=10000']
09:49:48 - cmdstanpy - INFO - Chain [1] start processing
INFO:cmdstanpy:Chain [1] start processing
09:49:48 - cmdstanpy - INFO - Chain [1] done processing
INFO:cmdstanpy:Chain [1] done processing
/usr/local/lib/python3.11/dist-packages/prophet/forecaster.py:1854: FutureWarning: 'Y' is deprecated and will be removed in a future version, please use 'YE' instead.
  dates = pd.date_range(
<Figure size 1000x500 with 0 Axes>
/usr/local/lib/python3.11/dist-packages/IPython/core/pylabtools.py:151: UserWarning: Glyph 127757 (\N{EARTH GLOBE EUROPE-AFRICA}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```



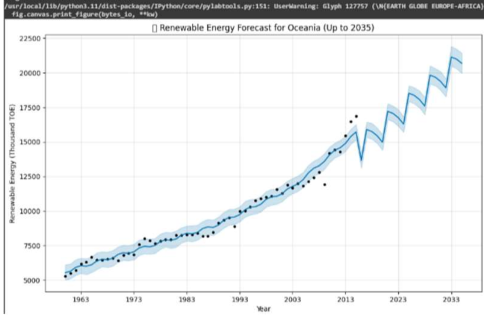

```
INFO:prophet:Disabling weekly seasonality. Run prophet with weekly_seasonality=True to override this.
INFO:prophet:Disabling daily seasonality. Run prophet with daily_seasonality=True to override this.
DEBUG:cmdstanpy:input tempfile: /tmp/tmpmes9sitn/ndtmu8n9.json
DEBUG:cmdstanpy:input tempfile: /tmp/tmpmes9sitn/kqf3ah86.json
DEBUG:cmdstanpy:idx 0
DEBUG:cmdstanpy:running CmdStan, num_threads: None
DEBUG:cmdstanpy:CmdStan args: ['/usr/local/lib/python3.11/dist-packages/prophet/stan_model/prophet_model.bin', 'random', 'seed=24034', 'data', 'file=/tmp/tmpmes9sitn/ndtmu8n9.json', 'init=/tmp/tmpmes9sitn/kqf3ah86.json', 'output', 'file=/tmp/tmpmes9sitn/prophet_model0vqq9v91/prophet_model-20250323094949.csv', 'method=optimize', 'algorithm=newton', 'iter=10000']
09:49:49 - cmdstanpy - INFO - Chain [1] start processing
INFO:cmdstanpy:Chain [1] start processing
09:49:49 - cmdstanpy - INFO - Chain [1] done processing
INFO:cmdstanpy:Chain [1] done processing
/usr/local/lib/python3.11/dist-packages/prophet/forecaster.py:1854: FutureWarning: 'Y' is deprecated and will be removed in a future version, please use 'YE' instead.
  dates = pd.date_range(
<Figure size 1000x500 with 0 Axes>
/usr/local/lib/python3.11/dist-packages/IPython/core/pylabtools.py:151: UserWarning: Glyph 127757 (\N{EARTH GLOBE EUROPE-AFRICA}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```




```
INFO:prophet:Disabling weekly seasonality. Run prophet with weekly_seasonality=True to override this.
INFO:prophet:Disabling daily seasonality. Run prophet with daily_seasonality=True to override this.
DEBUG:cmdstanpy:input tempfile: /tmp/tmpmes9sitn/ky_ikvfk.json
DEBUG:cmdstanpy:input tempfile: /tmp/tmpmes9sitn/cd2kdale.json
DEBUG:cmdstanpy:idx 0
DEBUG:cmdstanpy:running CmdStan, num_threads: None
DEBUG:cmdstanpy:CmdStan args: ['/usr/local/lib/python3.11/dist-packages/prophet/stan_model/prophet_model.bin', 'random', 'seed=494', 'data', 'file=/tmp/tmpmes9sitn/ky_ikvfk.json', 'init=/tmp/tmpmes9sitn/cd2kdale.json', 'output', 'file=/tmp/tmpmes9sitn/prophet_modelsjbizg99/prophet_model-20250323094950.csv', 'method=optimize', 'algorithm=newton', 'iter=10000']
09:49:50 - cmdstanpy - INFO - Chain [1] start processing
INFO:cmdstanpy:Chain [1] start processing
09:49:50 - cmdstanpy - INFO - Chain [1] done processing
INFO:cmdstanpy:Chain [1] done processing
/usr/local/lib/python3.11/dist-packages/prophet/forecaster.py:1854: FutureWarning: 'Y' is deprecated and will be removed in a future version, please use 'YE' instead.
  dates = pd.date_range(
<Figure size 1000x500 with 0 Axes>
/usr/local/lib/python3.11/dist-packages/IPython/core/pylabtools.py:151: UserWarning: Glyph 127757 (\N{EARTH GLOBE EUROPE-AFRICA}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```




```
INFO:prophet:Disabling weekly seasonality. Run prophet with weekly_seasonality=True to override this.
INFO:prophet:Disabling daily seasonality. Run prophet with daily_seasonality=True to override this.
DEBUG:cmdstanpy:input tempfile: /tmp/tmpmes9sitn/z0_cy9h2.json
DEBUG:cmdstanpy:input tempfile: /tmp/tmpmes9sitn/xwagzhr4.json
DEBUG:cmdstanpy:idx 0
DEBUG:cmdstanpy:running CmdStan, num_threads: None
DEBUG:cmdstanpy:CmdStan args: ['/usr/local/lib/python3.11/dist-packages/prophet/stan_model/prophet_model.bin', 'random', 'seed=96811', 'data', 'file=/tmp/tmpmes9sitn/z0_cy9h2.json', 'init=/tmp/tmpmes9sitn/xwagzhr4.json', 'output', 'file=/tmp/tmpmes9sitn/prophet_modelusgv6t8b/prophet_model-20250323094950.csv', 'method=optimize', 'algorithm=newton', 'iter=100
09:49:50 - cmdstanpy - INFO - Chain [1] start processing
INFO:cmdstanpy:Chain [1] start processing
09:49:51 - cmdstanpy - INFO - Chain [1] done processing
INFO:cmdstanpy:Chain [1] done processing
/usr/local/lib/python3.11/dist-packages/prophet/forecaster.py:1854: FutureWarning: 'Y' is deprecated and will be removed in a future version, please use 'YE' instead.
  dates = pd.date_range(
<Figure size 1000x500 with 0 Axes>
/usr/local/lib/python3.11/dist-packages/IPython/core/pylabtools.py:151: UserWarning: Glyph 127757 (\N{EARTH GLOBE EUROPE-AFRICA}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```




INFO:prophet:Disabling weekly seasonality. Run prophet with weekly_seasonality=True to override this.
INFO:prophet:Disabling daily seasonality. Run prophet with daily_seasonality=True to override this.
DEBUG:cmdstanpy:input tempfile: /tmp/tmpmes9sitn/ghy2xwnd.json
DEBUG:cmdstanpy:input tempfile: /tmp/tmpmes9sitn/5q4kdiid.json
DEBUG:cmdstanpy:idx 0
DEBUG:cmdstanpy:running CmdStan, num_threads: None
DEBUG:cmdstanpy:CmdStan args: ['/usr/local/lib/python3.11/dist-packages/prophet/stan_model/prophet_model.bin', 'random', 'seed=25482', 'data', 'file=/tmp/tmpmes9sitn/ghy2xwnd.json', 'init=/tmp/tmpmes9sitn/5q4kdiid.json', 'output', 'file=/tmp/tmpmes9sitn/prophet_modelf_53ui72/prophet_model-20250323094951.csv', 'method=optimize', 'algorithm=newton', 'iter=10000'
09:49:51 - cmdstanpy - INFO - Chain [1] start processing
INFO:cmdstanpy:Chain [1] start processing
09:49:51 - cmdstanpy - INFO - Chain [1] done processing
INFO:cmdstanpy:Chain [1] done processing
/usr/local/lib/python3.11/dist-packages/prophet/forecaster.py:1854: FutureWarning: 'Y' is deprecated and will be removed in a future version, please use 'YE' instead.
dates = pd.date_range(
<Figure size 1000x500 with 0 Axes>
/usr/local/lib/python3.11/dist-packages/IPython/core/pylabtools.py:151: UserWarning: Glyph 127757 (\N{EARTH GLOBE EUROPE-AFRICA}) missing from font(s) DejaVu Sans.
fig.canvas.print_figure(bytes_io, **kw)
Renewable Energy Forecast for South America (Up to 2035)
Renewable Energy (Thousand TOE)
200000
150000
100000
50000
0
1963
1973
1983
1993
2003
2013
2023
2033
Year

APPENDIX B

PLAGIARISM REPORT

turnitin Page 2 of 60 - Integrity Overview Submission ID trn:oid:::1:3230165706

7% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

46 Not Cited or Quoted 4%
Matches with neither in-text citation nor quotation marks

18 Missing Quotations 3%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5% Internet sources
4% Publications
3% Submitted works (Student Papers)

PLAGIARISM REPORT

Format - I

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY (Deemed to be University u/s 3 of UGC Act, 1956)		
Office of Controller of Examinations		
REPORT FOR PLAGIARISM CHECK ON THE DISSERTATION/PROJECT REPORTS FOR UG/PG PROGRAMMES **(To be attached in the dissertation/ project report)**		
1	Name of the Candidate **(IN BLOCK LETTERS)**	RUCHIT SHIVANI
2	Address of the Candidate	Department of Networking and Communications, SRM Institute of Science and Technology
3	Registration Number	RA2211031010131
4	Date of Birth	6/4/2004
5	Department	Department of Networking and Communications
6	Faculty	Engineering and Technology, School of Computing
7	Title of the Dissertation/Project	Ethanol Blends as an Alternative to Fossil Fuels
8	Whether the above project /dissertation is done by	Individual
9	Name and address of the Supervisor / Guide	Dr. Gayathri V M Associate Professor Department of Networking and Communications **Mail ID:** gayathrm@srmist.edu.in **Mobile Number:** 96299 38751
10	Name and address of Co-Supervisor / Co- Guide (if any)	**Mail ID:** **Mobile Number:**

11	Software Used	turnitin		
12	Date of Verification	27-04-2025		
13	**Plagiarism Details: (to attach the final report from the software)**			
Chapter	**Title of the Chapter**	**Percentage of similarity index (including self citation)**	**Percentage of similarity index (Excluding self-citation)**	**% of plagiarism after excluding Quotes, Bibliography, etc.,**
1	Ethanol Blends as an Alternative to Fossil Fuels			7
2				
3				
4				
5				
6				
7				
8				
9				
10				
Appendices				

I declare that the above information have been verified and found true to the best of my / our knowledge.

Signature of the Candidate	**Name & Signature of the Staff (Who uses the plagiarism check software)**
Name & Signature of the Supervisor/ Guide	**Name & Signature of the Co-Supervisor/Co-Guide**

Name & Signature of the HOD