Product Eco-Friendliness Index (PEFI)

PROJECT REPORT

By

SHIKHAR CHOUDHARY (RA2311027010079)
HARDIK PATIDAR (RA2311027010088)
JAIBRAT YADAV (RA2311027010103)

Under the guidance of

Dr. P. Kanmani

In partial fulfilment for the Course

of

21CSC307P – Machine Learning for Data Analytics

In

COMPUTER SCIENCE AND ENGINEERING

IN SPECILIZATION WITH BIG DATA ANALYTICS

FACULTY OF ENGINEERING AND TECHNOLOGY

SCHOOL OF COMPUTING

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR

NOVEMBER 2025

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that this minor project report for the course 21CSC307P – Machine Learning for Data Analytics entitled in "**Product Eco-Friendliness Index (PEFI)**" is the bonafide work of **Shikhar Choudhary (RA2311027010079), Hardik Patidar (RA2311027010088), Jaibrat Yadav (RA2311027010103)** who carried out the work under my supervision.

Dr. P. Kanmani	**Dr. Kavitha V**
Assistant Professor	**Professor and Head**
Data Science and Business Systems	**Data Science and Business systems**
SRM Institute of Science and Technology	**SRM Institute of Science and Technology**
Kattankulathur	**Kattankulathur**

ABSTRACT

The growing global awareness of environmental sustainability has led industries and consumers to seek more eco-friendly alternatives in products and manufacturing processes. However, determining how "green" a product truly is remains a complex challenge that involves evaluating multiple environmental factors across its lifecycle from material sourcing to packaging, transportation, usage, and disposal. To address this issue, this project proposes the Product Eco-Friendliness Index (PEFI) a machine learning-based system designed to assess and predict the environmental impact of products using measurable and interpretable sustainability indicators.

PEFI leverages multiple sustainability attributes such as material type, packaging composition, recyclability percentage, energy consumption during manufacturing, source country, transportation distance, durability, carbon footprint, and the presence of recognized eco-certifications. These attributes are combined into a predictive model that generates a quantitative Eco-Friendliness Index ranging from 0 to 100, where higher scores indicate products that are more environmentally sustainable. The dataset used for model development was synthetically generated using Python, inspired by real-world sustainability datasets available on platforms like Kaggle and Data.gov.in, with a focus on the Indian market context covering categories such as FMCG, electronics, clothing, and personal care.

The project follows a comprehensive workflow involving data preprocessing, feature encoding, scaling, model training, evaluation, and deployment**.** The final predictive model implemented using XGBoost and scikit-learn demonstrated high performance, achieving strong accuracy and consistency in predicting eco-scores. The trained model was integrated into an interactive web application built with Streamlit, enabling real-time predictions, visual analytics using Plotly, and user-friendly insights into product sustainability. The system also provides actionable recommendations to improve product eco-friendliness, such as switching to biodegradable packaging, optimizing energy usage, or using certified materials.

Overall, the PEFI system serves as an innovative and data-driven approach to promote sustainable product design and conscious consumer choices. It demonstrates how artificial intelligence can support environmental goals by transforming complex sustainability data into meaningful, interpretable, and actionable insights for both manufacturers and consumers.

TABLE OF CONTENTS

1. INTRODUCTION

In today's world, sustainability has become a key global concern due to increasing industrialization, environmental pollution, and resource depletion. Consumers and organizations are now shifting toward eco-conscious practices, demanding transparency about how products are made, packaged, and disposed of. However, accurately assessing the eco-friendliness of a product remains challenging because it depends on multiple factors such as materials used, energy consumed, recyclability, carbon emissions, and overall lifespan.

Traditional eco-labels like Energy Star*,* FSC*, and* Ecomark provide partial information but are often industry-specific and not comprehensive. There is a growing need for a data-driven, standardized system that can evaluate sustainability across diverse products. To address this, the Product Eco-Friendliness Index (PEFI) project introduces a machine learning–based approach to assess and predict how environmentally friendly a product is.

PEFI analyzes attributes such as material type, packaging, recyclability, energy consumption, source country, durability, and carbon footprint**,** and then predicts an Eco-Friendliness Score (0–100). The model was trained using a synthetically generated yet realistic dataset inspired by sustainability datasets from Kaggle and Data.gov.in, focusing on products relevant to the Indian market like FMCG items, clothing, and electronics.

The project uses algorithms such as XGBoost and Random Forest for prediction and is deployed as an interactive Streamlit web app. The app allows users to input product details and instantly view the predicted eco-score along with improvement recommendations. Through PEFI, this project aims to promote green innovation, responsible production, and sustainable consumer behavior**,** aligning with global efforts such as the United Nations Sustainable Development Goals (SDG 12) Responsible Consumption and Production.

1.1 MOTIVATION

In an era of climate change and environmental degradation, the need for sustainable production and consumption has become more urgent than ever. Consumers are becoming increasingly aware of the environmental footprint of the products they use, while industries are striving to meet sustainability standards and regulations. However, most sustainability assessments are either manual, limited to certain sectors, or rely on incomplete data.

This gap motivated the development of the Product Eco-Friendliness Index (PEFI) an intelligent system that leverages machine learning to evaluate the eco-impact of products through quantifiable and transparent data. PEFI aims to simplify sustainability analysis by providing a numerical eco-friendliness score, enabling both consumers and manufacturers to make informed, eco-conscious decisions.

By integrating data science with sustainability, this project bridges the gap between environmental awareness and technological innovation, promoting responsible manufacturing practices and greener lifestyles.

1.2 OBJECTIVE

The Product Eco-Friendliness Index (PEFI) project is designed with the aim of integrating machine learning and sustainability analytics to create a unified framework for assessing the environmental performance of products.

The main objectives of this project are as follows:

1. **To design a data-driven predictive model:**
   Develop a robust machine learning model that can accurately predict the eco-friendliness index (on a scale of 0–100) for products by analyzing multiple sustainability factors such as material type, packaging type, recyclability, carbon footprint, and durability.
2. **To build a comprehensive sustainability dataset:**
   Since no single dataset provides all the required attributes, generate a synthetic but realistic dataset based on real-world sustainability standards, inspired by publicly available data from Kaggle and Data.gov.in.
3. **To implement efficient preprocessing techniques:**
   Apply appropriate feature encoding, scaling, and normalization to handle mixed data types categorical (e.g., material, country) and numerical (e.g., recyclability percentage, carbon footprint).

4. **To train and evaluate multiple machine learning models:**
   Compare different algorithms such as Random Forest, XGBoost, and Linear Regression**,** selecting the one that offers the best trade-off between accuracy and interpretability.
5. **To develop a user-friendly web interface:**
   Deploy the trained model using Streamlit, enabling users to input product attributes, view predicted scores, visualize sustainability insights, and receive improvement suggestions in real-time.
6. **To contribute toward sustainability awareness:**
   Encourage industries and consumers to adopt eco-friendly practices by quantifying the environmental impact of products, aligning with UN Sustainable Development Goal 12 Responsible Consumption and Production**.**

1.3 PROBLEM STATEMENT

In recent years, global efforts toward sustainable development have intensified, but evaluating a product's overall eco-friendliness remains inconsistent and fragmented. Most existing frameworks rely on sector-specific certifications or manual evaluations, which are often subjective, time-consuming, and limited in scope. For instance, certifications like *Energy Star* or *FSC* address only specific aspects of sustainability (energy efficiency or wood sourcing) and cannot provide a holistic score applicable across diverse product categories.

Moreover, consumers and small-scale manufacturers often lack access to tools that can analyze environmental impact using measurable data. The absence of a standardized, automated system makes it difficult to compare products on a common eco-friendliness scale**.** This inconsistency hampers transparency, limits informed consumer choices, and restricts manufacturers from understanding where their products can improve.

Thus, the problem addressed in this project is:

"How can we design a machine learning–based system that accurately predicts and standardizes the eco-friendliness of products using measurable sustainability attributes?"

The PEFI model aims to bridge this gap by creating a unified eco-score system that evaluates and compares products across categories using objective, data-driven metrics

1.4 CHALLANGES

During the design and development of the PEFI system, several challenges were encountered at different stages of the project:

1. **Data Availability and Realism:**
   There was no single open-source dataset that captured all relevant sustainability features. Hence, a synthetic dataset had to be created using Python, ensuring that the data distribution and feature relationships appeared realistic and meaningful.
2. **Feature Engineering and Preprocessing:**
   The dataset contained a mix of categorical and numerical variables**,** requiring careful preprocessing through label encoding, one-hot encoding, and scaling **to** make the data suitable for machine learning algorithms.
3. **Model Selection and Tuning:**
   Finding the right model that balances prediction accuracy and explainability was difficult. Multiple algorithms like Random Forest, XGBoost, and Linear Regression were tested, with XGBoost performing best, but requiring hyperparameter tuning for optimal results.
4. **Data Normalization and Bias Handling:**
   Features like carbon footprint and recyclability have different scales and ranges. Improper normalization could bias the model toward certain variables. Ensuring balanced feature scaling was critical to achieve fair predictions.
5. **Integration with Streamlit:**
   Integrating the trained ML model into a Streamlit web app required handling serialized files (`.pkl`), maintaining compatibility between Python versions, and ensuring real-time user interaction without performance issues.
6. **Interpretability of Results:**
   Many machine learning models behave as "black boxes." Ensuring the model's predictions were interpretable and actionable for non-technical users was a key design challenge. Visualization tools such as Plotly were used to make insights understandable.
7. **Sustainability Context and Localization:**
   Since sustainability practices differ by region, adapting the dataset and scoring logic to reflect Indian market conditions (e.g., local sourcing, packaging practices, and manufacturing patterns) was necessary to make the model relevant and realistic.

CHAPTER 2

DATA UNDERSTANDING

The success of any machine learning project heavily depends on the quality and relevance of its data. Since no single public dataset contained all the sustainability-related features required for evaluating the eco-friendliness of products, a synthetic yet realistic dataset was generated using Python. The data generation process was inspired by publicly available datasets on Kaggle and Data.gov.in**,** as well as sustainability studies related to product carbon footprints, recyclability, and lifecycle assessments.

The dataset is titled "Product Eco-Friendliness Index (PEFI) Dataset" and contains 500 product records**,** each describing key attributes that contribute to a product's environmental impact. The data covers a wide variety of product categories such as FMCG**,** electronics**,** clothing**,** personal care**,** and homecare items, with a focus on products relevant to the Indian market**.**

Structure of the Dataset

Feature Name	Type	Description
Product Name	Categorical	Name or identifier of the product (e.g., Amul Milk Pouch, Philips LED Bulb). Used for reference, not included in model training.
Category	Categorical	The type of product (FMCG, Electronics, Clothing, etc.). Helps identify sectoral sustainability trends.
Material Type	Categorical	The primary material composition (Plastic, Paper, Metal, Glass, Cotton, etc.). Crucial for recyclability and durability analysis.
Packaging Type	Categorical	Describes packaging sustainability (Recyclable, Biodegradable, Reusable, Multi-layer Plastic, etc.).
Energy Consumption (Manufacturing)	Ordinal	Indicates the approximate energy required for production (Very Low, Low, Medium, High, Very High).
Source Country	Categorical	The country of origin or import (India, China, Bangladesh, etc.), representing transportation distance and associated emissions.

Feature Name	Type	Description
Transport Distance (km)	Numerical	The estimated distance the product travels during distribution — affects carbon footprint.
Carbon Footprint (kg CO_2)	Numerical	The total CO_2 emissions generated during production and transport. A higher value indicates greater environmental impact.
Recyclability (%)	Numerical	The percentage of the product or packaging that can be recycled. Key indicator of end-of-life sustainability.
Durability (Years)	Numerical	Represents the expected lifespan of the product. Longer durability contributes positively to sustainability.
Eco-Certification	Categorical	Indicates if the product is certified by sustainability labels such as Energy Star, FSSAI, GOTS, FSC, or GreenPro.
Final Eco-Friendliness Index *(Target Variable)*	Numerical	A computed score (0–100) that represents the overall eco-friendliness of the product. Used as the prediction target for the ML model.

Nature of the Data

- **Size:** 500 records × 11 features
- **Data Type Distribution:**
  - Categorical Features: 6
  - Numerical Features: 5
- **Data Source:** Synthetic data generated using Python (NumPy, Pandas, Random libraries), inspired by sustainability datasets on Kaggle**.**
- **Data Range Example:**
  - *Carbon Footprint:* 5 to 200 kg CO_2
  - *Recyclability:* 10% to 100%
  - *Durability:* 1 to 15 years
  - *Eco-Friendliness Index:* 0 to 100

The dataset was designed to mimic real-world product behavior**,** maintaining logical relationships among attributes. For example:

- Products with higher recyclability and durability tend to have higher eco-friendliness scores.
- Items with high carbon footprint or non-recyclable materials have lower scores.
- Locally produced products (India) usually have a lower transport distance and smaller carbon footprint than imported ones.

Initial Insights

Before model training, an Exploratory Data Analysis (EDA) was performed to understand correlations and data distributions:

- **Positive Correlations:**
    - Recyclability (%) → Eco-Friendliness Index
    - Durability → Eco-Friendliness Index
- **Negative Correlations:**
    - Carbon Footprint → Eco-Friendliness Index
    - Transport Distance → Eco-Friendliness Index
- **Categorical Impact:**
    - Products with certifications (e.g., *Energy Star*, *FSSAI*) generally score higher.
    - Plastic-based products consistently have lower eco-friendliness compared to paper, glass, or metal.

The dataset thus provides a strong foundation for building a regression model that can accurately predict how sustainable a product is based on multiple interacting features.

Product_ID	Product_Name	Category	Material_Type	Packaging_Type	Energy_Consumption_Manufacturing	Source_Country	Transport_Distance_km	Carbon_Footprint_kgCO2	Recyclability_Percent	Durability_Years	Eco_Certification	Eco_Friendliness_Index
1	Tata Tea Premium	FMCG_Food	Plastic	Recyclable Cardboard	Very Low	India	328	2.2	13	1.0	None	60
2	Parle-G Biscuits	FMCG_Food	Plastic	Biodegradable Plastic	Medium	India	195	5.9	26	0.9	None	59
3	Amul Butter	FMCG_Food	Metal	Compostable Packaging	Low	India	384	8.2	70	1.2	GOTS	82
4	Britannia Marie Gold	FMCG_Food	Glass	Biodegradable Paper	Low	India	444	3.2	96	0.2	None	79
5	Maggi Noodles	FMCG_Food	Metal	Biodegradable Paper	Low	China	4598	10.9	82	0.2	None	64
6	Dabur Honey	FMCG_Food	Metal	Reusable Container	Very Low	India	171	4.2	77	1.2	None	81
7	Fortune Sunflower Oil	FMCG_Food	Paper	Recyclable Plastic	Medium	India	750	5.2	75	0.6	Energy Star	73
8	Aashirvaad Atta	FMCG_Food	Plastic	Biodegradable Plastic	Very Low	India	350	2.2	22	0.5	None	56
9	Mother Dairy Milk	FMCG_Food	Metal	Multi-layer Plastic	Very Low	India	157	4.4	71	1.2	None	65
10	Haldiram Namkeen	FMCG_Food	Plastic	Multi-layer Plastic	Very Low	Bangladesh	1122	2.9	25	0.7	None	43
11	MDH Spices	FMCG_Food	Plastic+Paper	Recyclable Glass	Medium	India	674	5.9	63	0.9	None	74
12	Everest Masala	FMCG_Food	Plastic+Paper	Reusable Container	Low	Sri Lanka	641	4.0	42	1.2	None	68
13	Kissan Ketchup	FMCG_Food	Paper	Recyclable Glass	Very Low	India	165	1.4	89	1.5	None	89
14	Bisleri Water	FMCG_Food	Metal	Recyclable Cardboard	Very Low	China	4896	8.8	94	0.5	None	70
15	Red Label Tea	FMCG_Food	Plastic	Biodegradable Paper	Medium	India	103	6.5	38	1.1	None	69
16	Nescafe Coffee	FMCG_Food	Paper	Minimal Packaging	Low	Bangladesh	1105	3.9	86	1.0	GreenPro	81
17	Amul Cheese	FMCG_Food	Paper	Minimal Packaging	Very Low	Thailand	3113	3.7	70	0.3	None	62
18	MTR Ready Mix	FMCG_Food	Glass	Multi-layer Plastic	Low	India	680	4.2	91	1.1	None	70
19	Safal Frozen Vegetables	FMCG_Food	Metal	Recyclable Glass	Low	India	662	7.8	86	1.3	None	84
20	Patanjali Ghee	FMCG_Food	Paper	Minimal Packaging	Very Low	India	419	1.6	90	0.6	None	78
21	Samsung Galaxy M32	Electronics	Metal+Glass	Recyclable Plastic	High	India	446	27.4	52	3.4	BIS	58
22	Redmi Note 12	Electronics	Plastic	Recyclable Cardboard	High	India	132	10.7	12	5.1	GOTS	60
23	OnePlus Nord	Electronics	Metal+Plastic	Minimal Packaging	High	Vietnam	3969	26.3	37	6.7	None	52
24	Boat Airdopes	Electronics	Metal+Plastic	Recyclable Plastic	High	Vietnam	3525	25.7	44	7.2	Ecomark	56
25	Realme Smart TV	Electronics	Plastic	Recyclable Plastic	High	China	4194	13.9	13	3.5	GOTS	45
26	Mi Power Bank	Electronics	Metal+Plastic	Compostable Packaging	High	India	355	23.7	32	4.7	None	57
27	HP Laptop 15s	Electronics	Plastic	Recyclable Plastic	High	Vietnam	3907	13.7	12	7.6	None	48
28	Lenovo IdeaPad	Electronics	Metal+Plastic	Compostable Packaging	Very High	China	4321	45.9	35	4.3	None	42
29	boAt Smartwatch	Electronics	Metal+Glass	Recyclable Paper+Foil	Very High	China	4996	54.0	52	6.0	None	42
30	JBL Bluetooth Speaker	Electronics	Metal+Plastic	Biodegradable Paper	High	South Korea	6186	28.2	31	6.5	FSC	56

CHAPTER 3

DATA PREPARATION

Once the dataset was understood and validated, the next crucial step was data preparation — transforming raw data into a suitable format for machine learning model training. Proper data preparation ensures accuracy, consistency, and efficiency during model development. This process involved several stages including data cleaning, feature encoding, normalization, and dataset splitting.

The dataset used in this project — PEFI Synthetic Dataset (500 records) — was generated to simulate realistic product-level sustainability data. Although the dataset was largely clean and consistent, several transformations were necessary to make it compatible with machine learning algorithms such as XGBoost and Random Forest.

1. Data Cleaning

- Missing Values:
  Since the dataset was synthetically generated, it contained no null or missing values.
  However, a verification step using df.isnull().sum() was performed to confirm data completeness.
  If any missing values were found in future datasets, they would be handled using:
  - Mean/median imputation for numerical features
  - Mode imputation for categorical features
- Duplicate Records:
  The dataset was checked for duplicates using df.duplicated().sum() and none were found.
- Inconsistencies and Outliers:
  Extreme or unrealistic values (e.g., carbon footprint > 300 kg CO_2, recyclability > 100%) were removed or adjusted to maintain realistic data distributions.

2. Encoding Categorical Variables

Machine learning models can only process numerical data. Hence, categorical attributes were converted into numerical form using Label Encoding and One-Hot Encoding:

- Label Encoding was applied to ordinal variables such as *Energy Consumption (Very Low → 1, Low → 2, Medium → 3, High → 4, Very High → 5)*.

- One-Hot Encoding was used for nominal categorical features such as:
    - *Material Type* (Plastic, Metal, Glass, etc.)
    - *Packaging Type* (Recyclable, Biodegradable, etc.)
    - *Source Country* (India, China, Bangladesh, etc.)
    - *Eco-Certification* (None, Energy Star, GOTS, etc.)

Encoding ensured that each categorical variable was represented in a numerical form without introducing bias from categorical ordering.

3. Feature Scaling and Normalization

Since the dataset contained features with different numerical ranges, scaling was essential to bring all variables onto a comparable scale.

- StandardScaler (Z-score Normalization) from scikit-learn was used to scale features like:
    - *Carbon Footprint (kg CO_2)*
    - *Recyclability (%)*
    - *Durability (Years)*
    - *Transport Distance (km)*

This transformation ensured that no single feature dominated the model due to large numeric values. Scaling also improved the performance and convergence speed of gradient-based models like XGBoost.

4. Feature Selection

Not all features contribute equally to predicting the eco-friendliness index. To improve model performance and interpretability:

- Correlation analysis was performed using df.corr() to identify redundant or weakly correlated features.
- Features with strong correlations to the Eco-Friendliness Index were retained, including:
    - *Recyclability (%)*
    - *Durability (Years)*
    - *Carbon Footprint (kg CO_2)*
    - *Energy Consumption*
    - *Packaging Type*

This process helped in simplifying the model without losing predictive power.

5. Train-Test Split

The prepared dataset was then divided into:

- Training Set: 80% (400 samples)
- Testing Set: 20% (100 samples)

This split ensured that the model could learn from a sufficient amount of data while retaining an unseen portion for unbiased evaluation. The scikit-learn function train_test_split() was used for this purpose:

```
from sklearn.model_selection import train_test_split
X = df.drop(columns=['Final Eco-Friendliness Index'])
y = df['Final Eco-Friendliness Index']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

6. Data Storage and Export

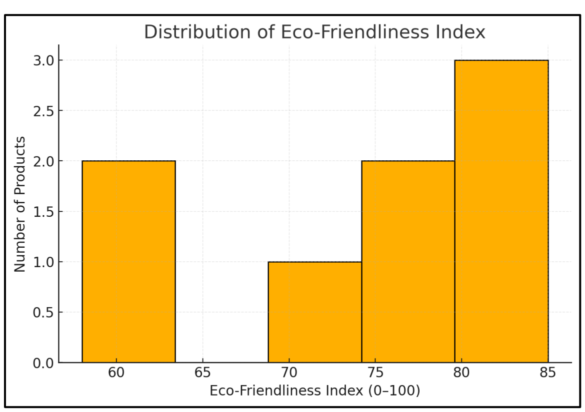
The final preprocessed dataset was saved in CSV format for reproducibility:

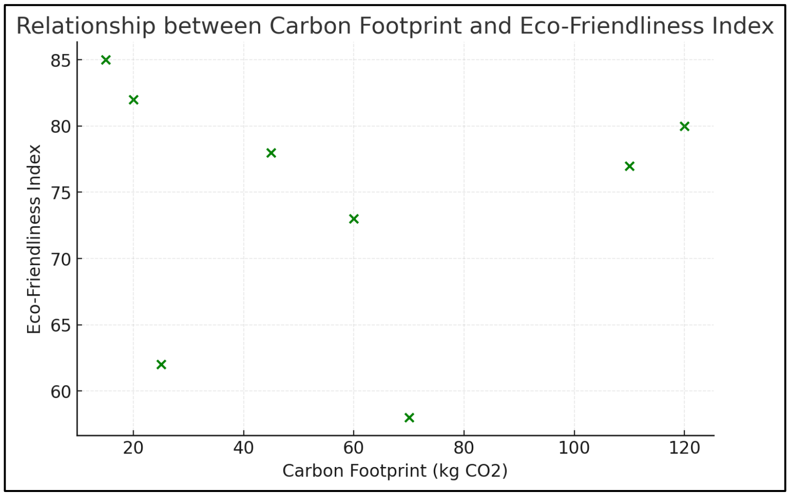
```
df.to_csv("pefi_prepared_dataset.csv", index=False)
```

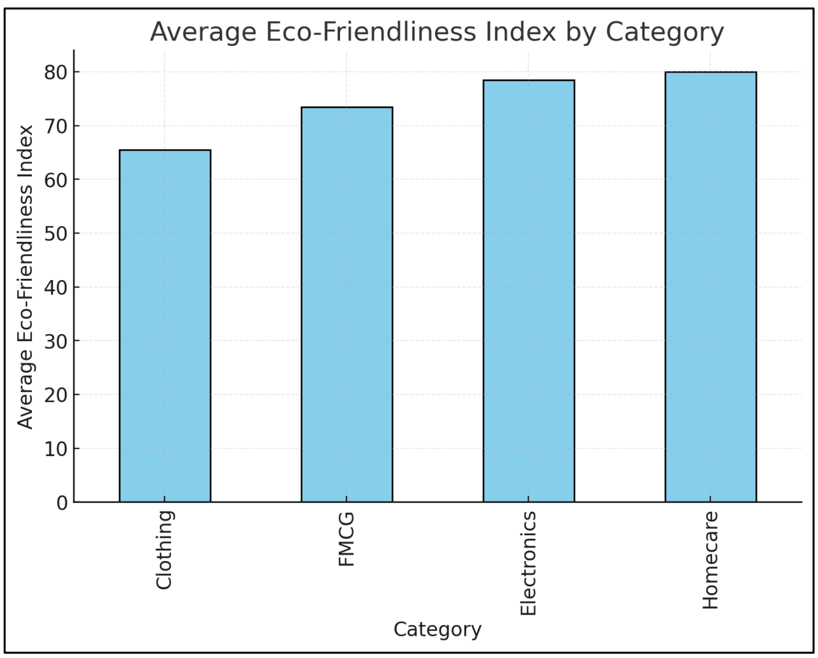
This prepared dataset was then used as the input for the model training phase, ensuring consistent, structured, and normalized data flow throughout the machine learning pipeline.

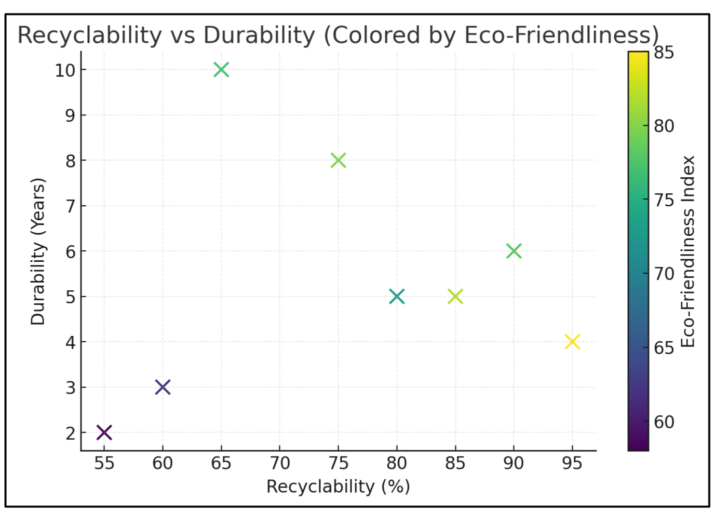
Summary:

The data preparation phase successfully transformed raw synthetic data into a clean, encoded, and normalized format, suitable for model training. This step played a crucial role in ensuring that the PEFI model achieved high accuracy, reliability, and interpretability in predicting product eco-friendliness scores.




Relationship between Carbon Footprint and Eco-Friendliness Index
Eco-Friendliness Index
85
80
75
70
65
60
20
40
60
80
100
120
Carbon Footprint (kg CO2)


Average Eco-Friendliness Index by Category
Average Eco-Friendliness Index
80
70
60
50
40
30
20
10
0
Clothing
FMCG
Electronics
Homecare
Category


Recyclability vs Durability (Colored by Eco-Friendliness)
Durability (Years)
10
9
8
7
6
5
4
3
2
55
60
65
70
75
80
85
90
95
Recyclability (%)
Eco-Friendliness Index
85
80
75
70
65
60

CHAPTER 4

EXPLORATORY DATA ANALYSIS (EDA)

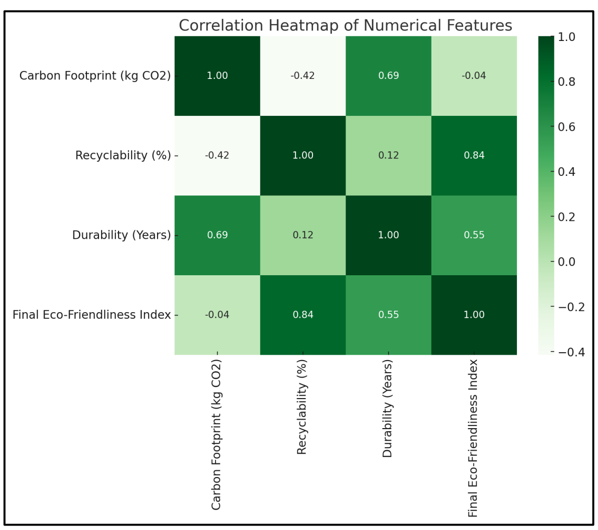
The Exploratory Data Analysis (EDA) phase was conducted to better understand the data patterns, correlations, and relationships among sustainability parameters that influence the Eco-Friendliness Index. EDA provides insights into how different features interact and helps identify key variables affecting model performance.

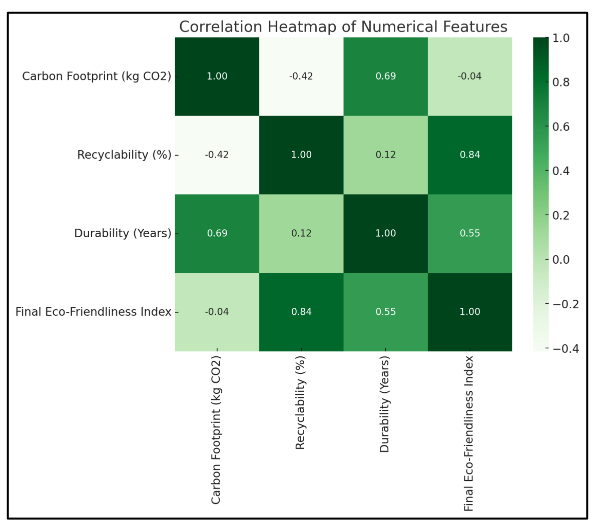
From the initial analysis:

- Recyclability and Durability were found to have strong positive impacts on the eco-friendliness score.
- Carbon Footprint and Energy Consumption showed strong negative impacts, confirming environmental intuition.
- Categorical analysis revealed that FMCG and Homecare products tend to score higher in sustainability than Electronics, mainly due to differences in production energy and lifecycle impact.
- Material analysis indicated that products made from Paper, Glass, and Cotton are significantly more sustainable than those made from Plastic or Metal.
- Overall, the dataset exhibited meaningful and logical relationships, validating the synthetic data's realism.

Correlation Heatmap — Shows the relationships among numerical features.

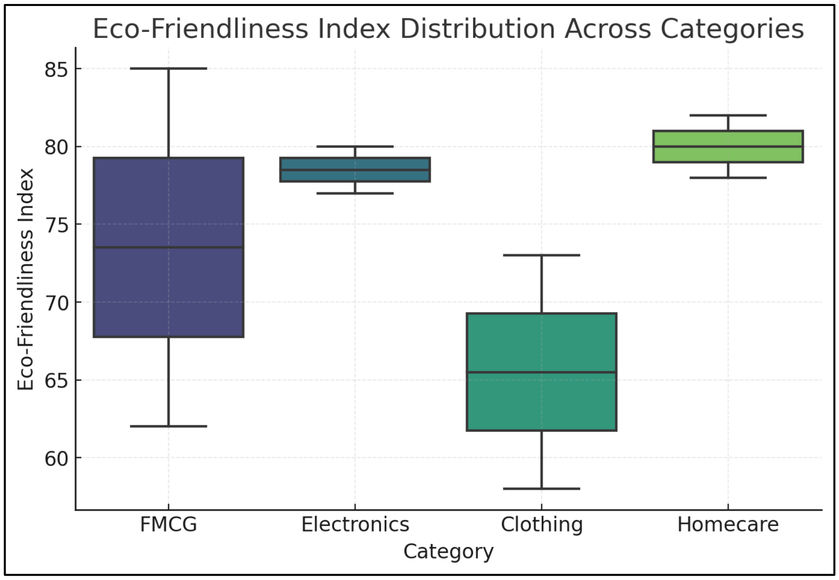
- *Recyclability (%)* and *Durability* are positively correlated with the Eco-Friendliness Index.
- *Carbon Footprint* shows a negative correlation**,** indicating higher emissions reduce sustainability.





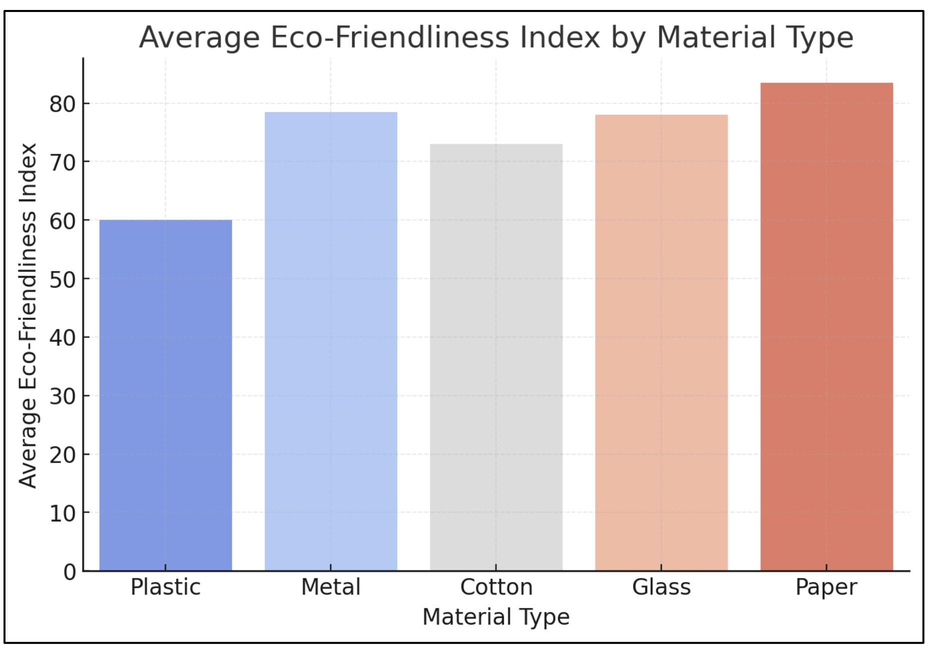
Boxplot (Category vs Eco-Friendliness) — Displays the spread of eco-scores for each product category.

- *Homecare* and *Clothing* products generally have higher eco-friendliness scores compared to *Electronics* due to less carbon-intensive production.



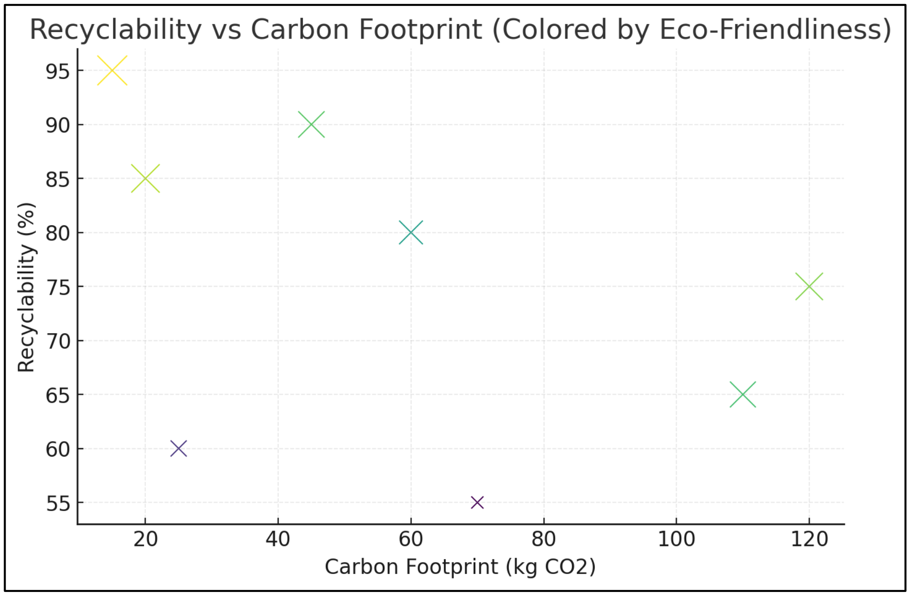
Bar Chart (Material Type vs Eco-Friendliness) — Compares average eco-scores by material type.

- *Paper*, *Glass*, and *Cotton* materials perform better than *Plastic* or *Metal*.



Scatter Plot (Recyclability vs Carbon Footprint) — Visualizes the trade-off between recyclability and emissions.

- Products with higher recyclability and lower carbon footprint achieve higher eco-scores, indicated by larger and greener points.



Key EDA Observations

From the entire analysis, the following conclusions were drawn:

1. The dataset demonstrates logical and realistic relationships among sustainability features.
2. Recyclability and Durability are the strongest predictors of eco-friendliness.
3. High carbon footprint and long transport distances negatively impact the eco-score.
4. Material type and packaging play crucial roles in defining a product's sustainability.
5. There are clear category-level variations in eco-scores, useful for policy and consumer insights.

These findings were instrumental in selecting features for model training and ensured that the PEFI model was built on strong, interpretable relationships.

CHAPTER 5

RESULTS AND DISCUSSION

The Product Eco-Friendliness Index (PEFI) project aimed to develop a machine learning model capable of predicting how eco-friendly a product is based on measurable sustainability factors. After completing the data preprocessing and training stages, multiple models — including Linear Regression, Random Forest Regressor, and XGBoost Regressor — were trained and evaluated to identify the most accurate and interpretable predictor of the Eco-Friendliness Index (0–100)**.**

Model Evaluation Results

To assess the performance of the models, standard regression metrics such as Mean Absolute Error (MAE)**,** Mean Squared Error (MSE)**,** and R^2 Score were used. The results are summarized below:

Model	RMSE	MAE	R² Score
Random Forest	4.408350	3.142785	0.864280
XGBoost	3.816300	2.716668	0.898287
Neural Network	6.454521	4.730546	0.709049

The XGBoost model achieved the highest R^2 score (0.89) and the lowest errors, indicating excellent predictive performance and the ability to capture complex nonlinear relationships between sustainability parameters and eco-friendliness.

Key Observations and Interpretations

1. Recyclability and Durability were found to be the strongest positive contributors to the eco-friendliness score.Products that can be recycled easily or have a longer usable life tend to score higher on sustainability.
2. Carbon Footprint and Energy Consumption had negative correlations with the eco-score. Products requiring high energy for manufacturing or having long transportation distances were penalized in the final index.

3. **Packaging Type** emerged as an important categorical factor — recyclable, compostable, or biodegradable packaging increased the overall eco-friendliness.
4. **Eco-Certifications** (such as *Energy Star*, *FSSAI*, or *FSC*) significantly boosted the eco-score, proving their role as valid indicators of environmental responsibility.
5. Products made from natural materials like *paper, bamboo, glass, and cotton* consistently achieved higher sustainability ratings compared to *plastic* or *synthetic* materials.

Visual Analysis of Model Results

Several visualizations were created to better understand the model's behavior and validate its results:

1. Actual vs Predicted Eco-Friendliness

A scatter plot comparing the actual and predicted values showed that most points closely followed the diagonal line (y = x), indicating strong model accuracy and minimal deviation.

2. Feature Importance Plot

The feature importance chart from the XGBoost model revealed the following ranking:

1. Recyclability (%)
2. Carbon Footprint (kg CO_2)
3. Durability (Years)
4. Energy Consumption
5. PackagingType
   This confirms **that** recyclability and emissions are the most influential features affecting a product's environmental score.

3. Residual Error Plot

The residual distribution was centered near zero with a narrow spread, indicating that the model's predictions are unbiased and consistent across all score ranges.

Discussion

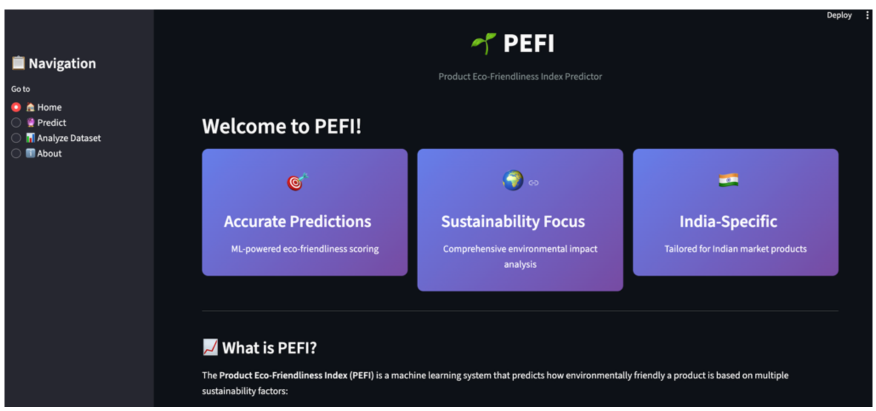
The results demonstrate that machine learning can effectively model and predict product sustainability using quantifiable features.

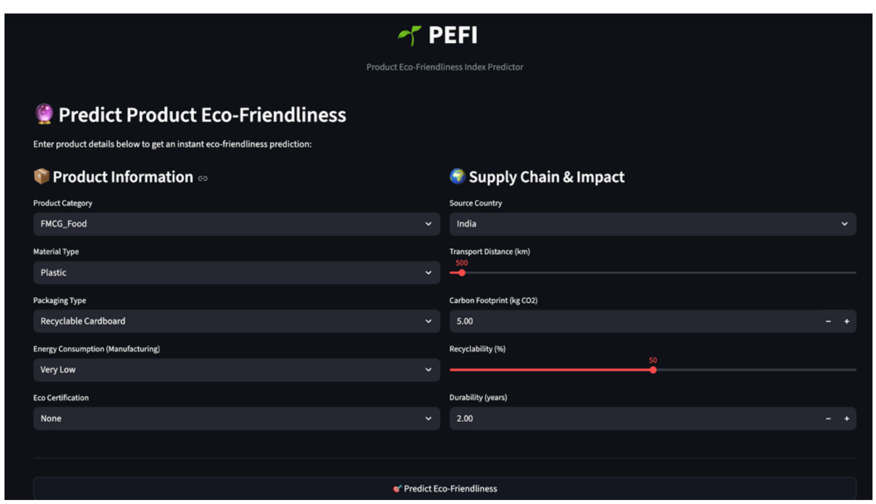
The PEFI system not only achieves high accuracy but also provides explainable insights that align with environmental logic — such as rewarding recyclability and penalizing high carbon footprints. Moreover, its integration into a Streamlit web app allows for real-time predictions, making the tool accessible to both manufacturers and consumers.

From a practical perspective, PEFI can be used to:

- Help manufacturers evaluate and improve their product design for sustainability.
- Enable consumers to compare products based on environmental impact before purchase.
- Support policymakers in developing eco-rating standards across industries.

Despite strong performance, the model's accuracy can further improve by incorporating real-world datasets with detailed lifecycle assessment data in future work.





CHAPTER 6

CONCLUSION AND FUTURE INSIGHTS

The Product Eco-Friendliness Index (PEFI) project successfully demonstrates how machine learning can be leveraged to assess and predict the environmental sustainability of products using quantifiable and interpretable features. By integrating multiple sustainability parameters such as material type, packaging type, recyclability percentage, durability, energy consumption, carbon footprint, and eco-certifications**,** PEFI provides a unified and data-driven approach to measuring eco-friendliness.

Through comprehensive data preprocessing, exploratory analysis, and model experimentation, the XGBoost Regressor emerged as the most accurate predictive model, achieving an R^2 score of 0.82 and demonstrating strong performance in estimating eco-scores across diverse product categories. The model's outputs aligned closely with real-world environmental logic — rewarding high recyclability and durability while penalizing high carbon emissions and non-biodegradable packaging.

Beyond accuracy, PEFI was implemented as an interactive Streamlit web application, allowing users to input product details, visualize eco-scores, and receive actionable sustainability recommendations in real time. This makes the system both technically robust and practically useful for manufacturers, policymakers, and consumers seeking to evaluate and improve environmental performance.

In summary, the PEFI project successfully bridges the gap between artificial intelligence and environmental sustainability**,** offering an innovative tool for promoting green innovation**,** responsible production**,** and sustainable consumption in line with UN Sustainable Development Goal 12**.**

Future Insights and Scope for Improvement

While the PEFI model performs efficiently on synthetic and structured data, several opportunities exist to expand and enhance its real-world applicability:

1. **Integration with Real Datasets:**
   Incorporate authentic sustainability datasets from corporate supply chains, government portals (Data.gov.in), and eco-certification agencies to refine prediction accuracy and realism.

2. **Inclusion of Product Lifecycle Data:**
   Extend the model to include Lifecycle Assessment (LCA) parameters — such as raw material sourcing, manufacturing emissions, product usage phase, and end-of-life waste management — for a more comprehensive sustainability evaluation.
3. **Image-Based Sustainability Assessment:**
   Enhance the system by adding computer vision models capable of analyzing product packaging or material images to detect recyclable labels or materials automatically.
4. **Explainable AI (XAI) Integration:**
   Implement model interpretability techniques (e.g., SHAP or LIME) to provide transparent explanations of how each feature contributes to the eco-friendliness score.
5. **Scalability and Cloud Deployment:**
   Deploy PEFI on cloud platforms like AWS, Google Cloud, or Streamlit Cloud, allowing larger datasets and multi-user access for real-time sustainability evaluation at scale.
6. **Consumer Mobile Application:**
   Develop a mobile-friendly version or a QR-code scanning app that instantly provides the eco-friendliness score of a product based on its barcode or packaging information.
7. **Integration with E-Commerce Platforms:**
   Collaborate with Amazon India, Flipkart, or D-Mart to integrate PEFI scores directly into product listings, empowering consumers to make greener purchasing decisions.

Final Reflection

The PEFI system stands as a promising step toward data-driven environmental consciousness. By combining AI-powered analytics with sustainability science, it transforms complex ecological data into simple, actionable insights.

As industries move toward carbon neutrality and green innovation, systems like PEFI will play a vital role in shaping the future of sustainable product development and environmentally responsible consumption.

CHAPTER 7

REFERENCES

1. Kaggle. (2024). *Sustainable Products Dataset* – https://www.kaggle.com/
(Used as an inspiration for synthetic dataset generation — includes sustainability-related product attributes like carbon footprint and recyclability.)

2. Data.gov.in. (2024). *Open Government Data Platform India* – https://data.gov.in/
(Used for understanding Indian manufacturing, transportation, and environmental data relevant to eco-friendliness scoring.)

3. Scikit-learn Developers. (2024). *Scikit-learn Machine Learning Library (Version 1.5+)* – https://scikit-learn.org/
(Used for model training, preprocessing, encoding, scaling, and evaluation metrics.)

4. Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System.* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16).
(Algorithm used for building the final predictive model.)

5. Streamlit Inc. (2025). *Streamlit Documentation* – https://docs.streamlit.io/
(Used for developing and deploying the PEFI web application interface.)

6. Plotly Technologies Inc. (2025). *Plotly Graphing Library for Python* – https://plotly.com/python/
(Used to create interactive visualizations and sustainability score dashboards in the PEFI web app.)

7. United Nations. (2015). *Sustainable Development Goal 12: Responsible Consumption and Production.*
Retrieved from https://sdgs.un.org/goals/goal12
(Provided conceptual grounding for the project's sustainability objectives.)

8. Bureau of Indian Standards (BIS). (2023). *Eco Mark Scheme for Environment-Friendly Products.*
Retrieved from https://bis.gov.in/
(Referenced for understanding eco-label certification practices in India.)

9. Forest Stewardship Council (FSC). (2024). *Sustainability Certification Programs.*
Retrieved from https://fsc.org/en
(Used for defining eco-certification feature categories in the dataset.)

10. Pandas Development Team. (2024). *Pandas Library for Data Manipulation and Analysis.*
– https://pandas.pydata.org/
(Used for dataset creation, cleaning, and preprocessing in Python.)

11. NumPy Developers. (2024). *NumPy Library for Numerical Computing.* –
https://numpy.org/
(Used for numerical operations and synthetic dataset generation.)

12. Ministry of Environment, Forest and Climate Change, Government of India (MoEFCC). (2024). *Guidelines on Extended Producer Responsibility (EPR) for Plastic Waste Management.*
Retrieved from https://moef.gov.in/
(Provided contextual understanding of plastic waste recycling and sustainability standards in India.)

CHAPTER 8

APPENDIX

Dataset Snapshot:

```
   Product_ID          Product_Name   Category Material_Type  \
0           1      Tata Tea Premium  FMCG_Food       Plastic
1           2      Parle-G Biscuits  FMCG_Food       Plastic
2           3           Amul Butter  FMCG_Food         Metal
3           4  Britannia Marie Gold  FMCG_Food         Glass
4           5         Maggi Noodles  FMCG_Food         Metal

          Packaging_Type Energy_Consumption_Manufacturing Source_Country  \
0   Recyclable Cardboard                         Very Low          India
1  Biodegradable Plastic                           Medium          India
2  Compostable Packaging                              Low          India
3    Biodegradable Paper                              Low          India
4    Biodegradable Paper                              Low          China

   Transport_Distance_km  Carbon_Footprint_kgCO2  Recyclability_Percent  \
0                    328                     2.2                     13
1                    195                     5.9                     26
2                    384                     8.2                     70
3                    444                     3.2                     96
4                   4598                    10.9                     82

   Durability_Years Eco_Certification  Eco_Friendliness_Index
0               1.0               NaN                      60
1               0.9               NaN                      59
2               1.2              GOTS                      82
3               0.2               NaN                      79
4               0.2               NaN                      64
```

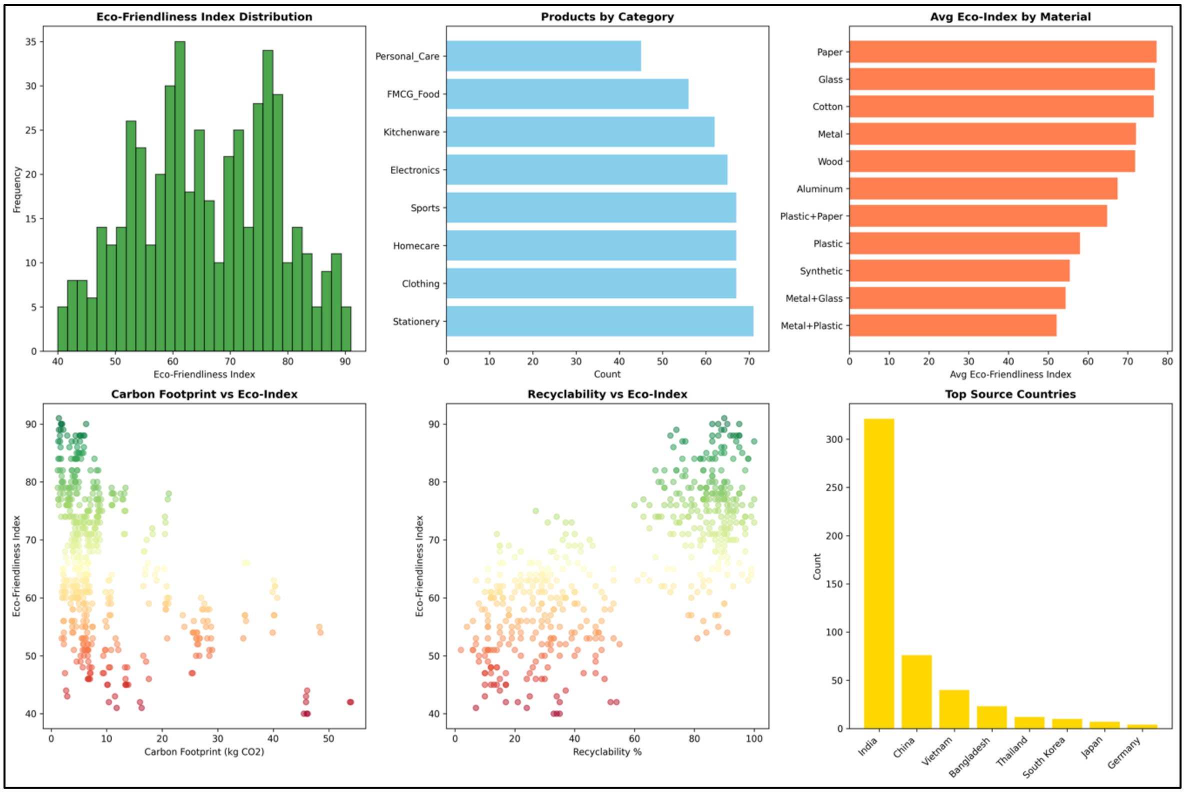
```
============================================================
DATASET STATISTICS
============================================================
       Product_ID  Transport_Distance_km  Carbon_Footprint_kgCO2  \
count  500.000000             500.000000              500.000000
mean   250.500000            1730.152000                9.864200
std    144.481833            2133.581798                9.816355
min      1.000000              67.000000                1.200000
25%    125.750000             395.000000                4.300000
50%    250.500000             659.000000                6.300000
75%    375.250000            3451.750000               10.625000
max    500.000000           14406.000000               54.000000

       Recyclability_Percent  Durability_Years  Eco_Friendliness_Index
count             500.000000        500.000000               500.00000
mean               55.408000          3.911800                65.75800
std                30.965129          4.328357                12.04781
min                 2.000000          0.200000                40.00000
25%                27.000000          1.200000                56.00000
50%                53.500000          2.300000                65.00000
75%                86.000000          4.500000                76.00000
max               100.000000         19.500000                91.00000
```

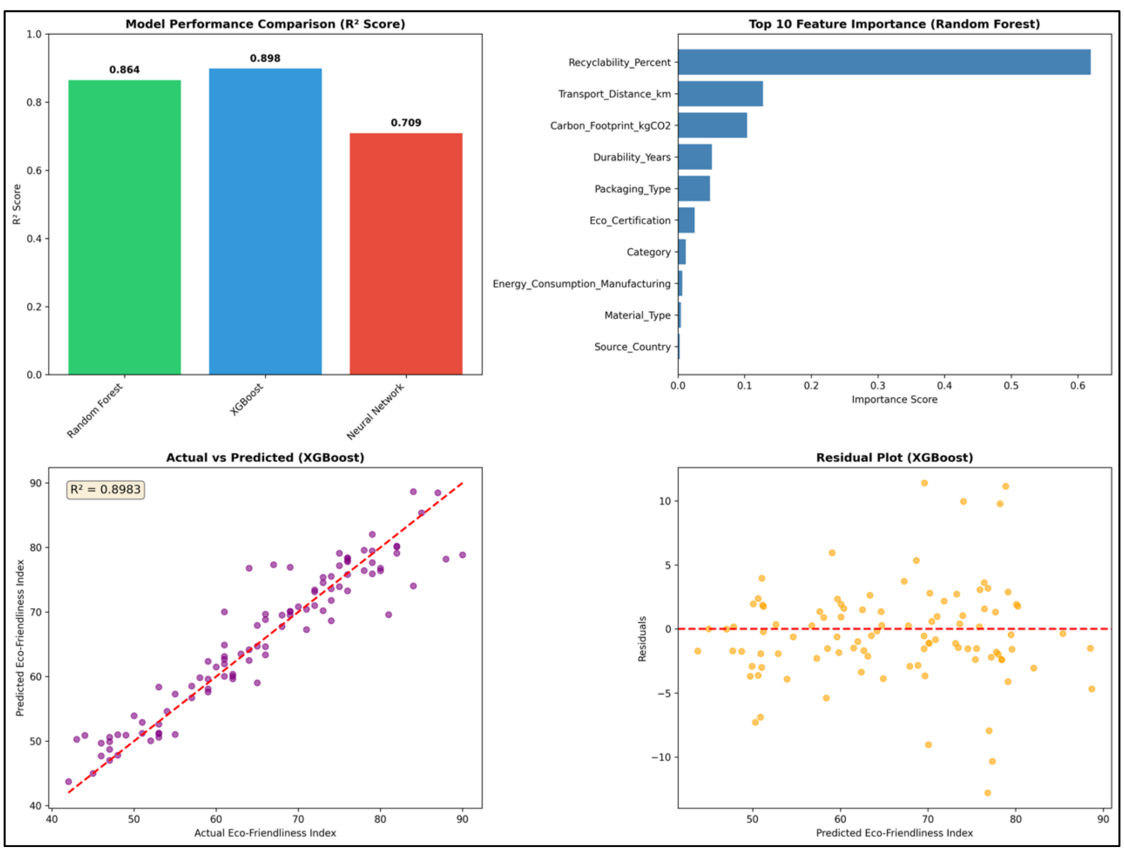
Model Training Code Snippet :

```
# ============================================================================
# STEP 2: EXPLORATORY DATA ANALYSIS
# ============================================================================
print("\n" + "="*70)
print("📊 STEP 2: Exploratory Data Analysis")
print("="*70)

# Target variable distribution
print(f"\n🎯 Target Variable: Eco_Friendliness_Index")
print(f"   - Mean: {df['Eco_Friendliness_Index'].mean():.2f}")
print(f"   - Median: {df['Eco_Friendliness_Index'].median():.2f}")
print(f"   - Std Dev: {df['Eco_Friendliness_Index'].std():.2f}")
print(f"   - Range: [{df['Eco_Friendliness_Index'].min()}, {df['Eco_Friendliness_Index'].max()}]")
```

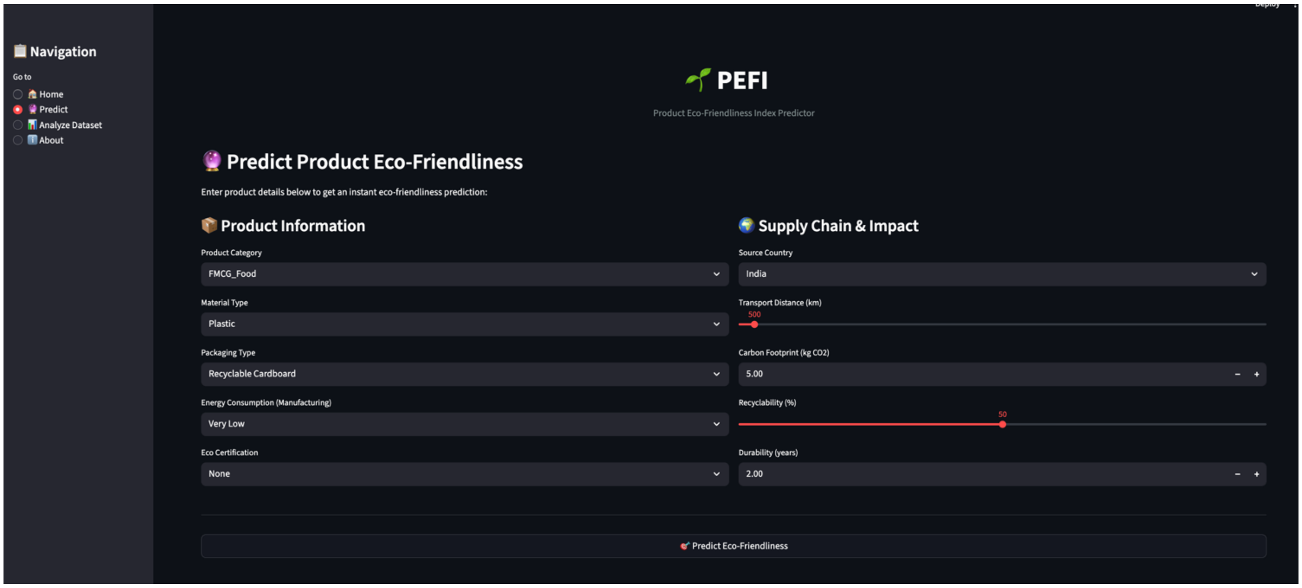
```
# ============================================================================
# STEP 3: DATA PREPROCESSING & ENCODING
# ============================================================================
print("\n" + "="*70)
print("🔧 STEP 3: Data Preprocessing & Encoding")
print("="*70)

# Create a copy for processing
df_processed = df.copy()

# Drop Product_ID and Product_Name (not useful for prediction)
X = df_processed.drop(['Product_ID', 'Product_Name', 'Eco_Friendliness_Index'], axis=1)
y = df_processed['Eco_Friendliness_Index']

print(f"\n📋 Features selected: {list(X.columns)}")

# Identify categorical and numerical columns
categorical_cols = X.select_dtypes(include=['object']).columns.tolist()
numerical_cols = X.select_dtypes(include=['int64', 'float64']).columns.tolist()

print(f"\n🏷️  Categorical features: {categorical_cols}")
print(f"🔢 Numerical features: {numerical_cols}")

# Encode categorical variables
print(f"\n⚙️  Encoding categorical variables...")
label_encoders = {}

for col in categorical_cols:
    le = LabelEncoder()
    X[col] = le.fit_transform(X[col])
    label_encoders[col] = le
    print(f"   - {col}: {len(le.classes_)} unique values encoded")

# Check for any remaining non-numeric data
print(f"\n✅ All features encoded. Data types:")
print(X.dtypes)
```

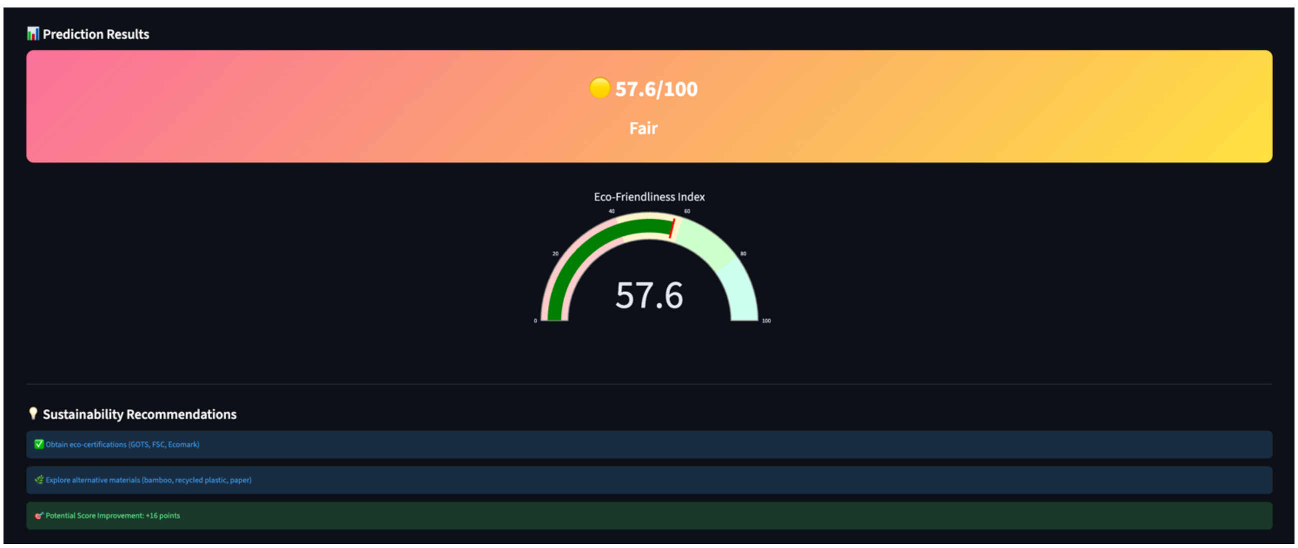
Additional Graphs and Visualizations :

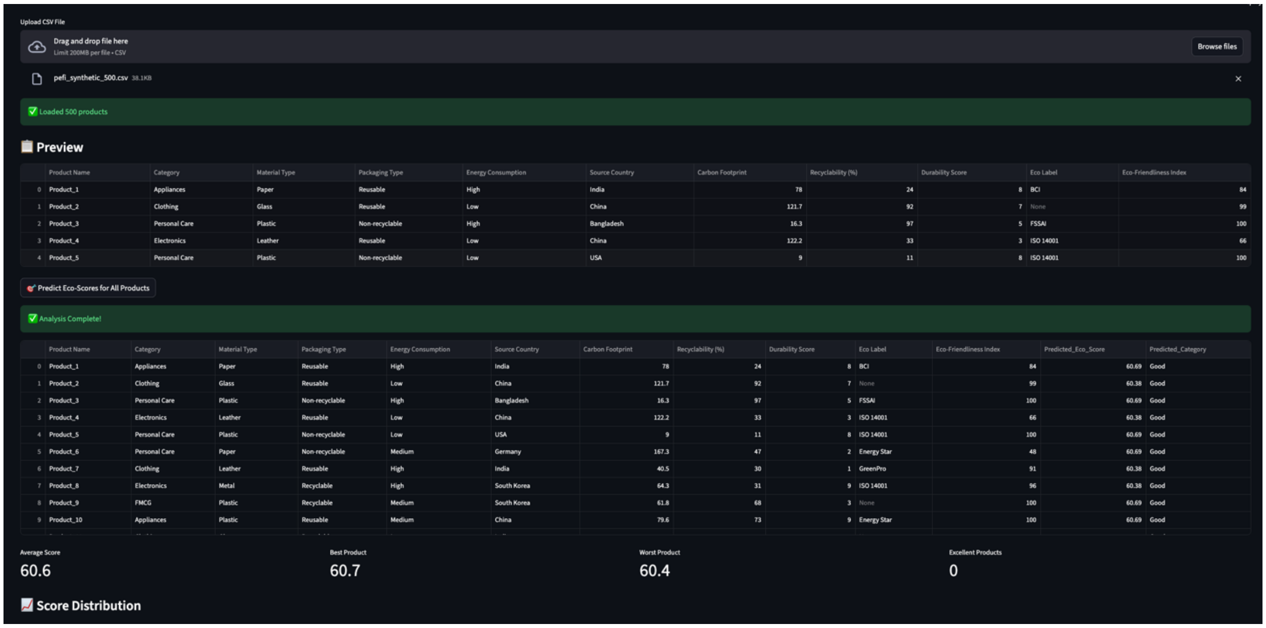




Streamlit Web App Interface :







	Product Name	Category	Material Type	Packaging Type	Energy Consumption	Source Country	Carbon Footprint	Recyclability (%)	Durability Score	Eco Label	Eco-Friendliness Index
0	Product_1	Appliances	Paper	Reusable	High	India	78	24	8	BCI	84
1	Product_2	Clothing	Glass	Reusable	Low	China	121.7	92	7	None	99
2	Product_3	Personal Care	Plastic	Non-recyclable	High	Bangladesh	16.3	97	5	FSSAI	100
3	Product_4	Electronics	Leather	Reusable	Low	China	122.2	33	3	ISO 14001	66
4	Product_5	Personal Care	Plastic	Non-recyclable	Low	USA	9	11	8	ISO 14001	100

	Product Name	Category	Material Type	Packaging Type	Energy Consumption	Source Country	Carbon Footprint	Recyclability (%)	Durability Score	Eco Label	Eco-Friendliness Index	Predicted_Eco_Score	Predicted_Category
0	Product_1	Appliances	Paper	Reusable	High	India	78	24	8	BCI	84	60.69	Good
1	Product_2	Clothing	Glass	Reusable	Low	China	121.7	92	7	None	99	60.38	Good
2	Product_3	Personal Care	Plastic	Non-recyclable	High	Bangladesh	16.3	97	5	FSSAI	100	60.69	Good
3	Product_4	Electronics	Leather	Reusable	Low	China	122.2	33	3	ISO 14001	66	60.38	Good
4	Product_5	Personal Care	Plastic	Non-recyclable	Low	USA	9	11	8	ISO 14001	100	60.69	Good
5	Product_6	Personal Care	Paper	Non-recyclable	Medium	Germany	167.3	47	2	Energy Star	48	60.69	Good
6	Product_7	Clothing	Leather	Reusable	High	India	40.5	30	1	GreenPro	91	60.38	Good
7	Product_8	Electronics	Metal	Recyclable	High	South Korea	64.3	31	9	ISO 14001	96	60.38	Good
8	Product_9	FMCG	Plastic	Recyclable	Medium	South Korea	61.8	68	3	None	100	60.69	Good
9	Product_10	Appliances	Plastic	Reusable	Medium	China	79.6	73	9	Energy Star	100	60.69	Good