**GHG Emission Prediction Project**

# **Day 1 -** Understanding the Project

# **Date -** 17 June 2025 (Week 1)

# **Internship Name** - Edunet-Shell Skills4Future AICTE Internship

# **Intern Name:** Abhinay Singh

# **📌 Objectives for Day 1**

* Understand the theoretical background of GHG emissions and the greenhouse effect.
* Understand the problem statement and project goals.
* Identify whether the project is classification or regression.
* Understand the dataset structure (inputs and output).
* Clarify personal doubts and build a foundational mindset for the internship.

📘 **What I Learned Today**

**1. Theoretical Understanding of the Greenhouse Effect**

* The greenhouse effect is a natural phenomenon that traps heat and keeps Earth livable (~14°C).
* Without it, Earth's surface would be too cold (~-18°C) to support life.
* But **human activities** are increasing GHGs like CO₂, CH₄, and N₂O, which **trap excessive heat**, leading to global warming and climate change.

**2. Real-World Importance of the Project**

* This project aims to predict GHG emission factors for different industries and commodities.
* By doing this, industries and governments can **identify high-emission areas** and **make cleaner choices**.
* It's a **data-driven contribution to sustainable development** and fighting climate change.

**3. Project Type and Tools**

* **✅** It is a Supervised Learning Regression Problem.
* **Output** (target) is numeric: GHG Emission Factor (with or without margin)
* **Tools/Libraries**: Python, Pandas, Scikit-learn, Matplotlib, Seaborn

**4. Input and Output Clarification**

**Input Features** (independent variables):

* Industry, Commodity, Substance, Unit
* Quality metrics like Reliability, Temporal/Geographical/Technological correlation, etc.

**Output Feature** (target):

* GHG Emission Factor with Margin → a numerical value (kg CO₂ equivalent).

## **❓ Doubts Asked Today and Clarified**

|  |  |
| --- | --- |
| **Question** | **Answer** |
| Why is the greenhouse effect both good and bad? | It's essential for life (natural), but humans are increasing it excessively, causing global warming. |
| How does this project help in real life? | It helps industries and governments make decisions based on GHG data — reducing emissions and promoting eco-friendly practices. |
| Is this project classification or regression? | Regression — because the output is a continuous numeric value. |
| What are the input and output variables? | Inputs = Industry, Commodity, Reliability, etc.; Output = GHG Emission Factor (numeric). |

**✅ Conclusion for Day 1**

* Today was a strong start to the internship. I understood the importance of GHG data, the project goals, and the structure of the machine learning task.
* I feel confident about my theoretical understanding and excited to begin working on the dataset and Week 1 code.

## **🔗 References**

* **What are greenhouse gases?**

[**https://www.nationalgrid.com/stories/energy-explained/what-are-greenhouse-gases**](https://www.nationalgrid.com/stories/energy-explained/what-are-greenhouse-gases)

* ChatGPT guidance for theory and concepts.

# 