Final Model Documentation

Model Architecture:

The model used is a **Gradient Boosting** regressor. After hyperparameter tuning, the following best parameters were found:

- `learning_rate`: 0.1
- `max_depth`: 4
- `n_estimators`: 100
- `subsample`: 0.8

The hyperparameter tuning improved model performance by balancing the model's ability to generalize and reduce overfitting. By adjusting the depth and number of estimators, the model became more efficient at capturing the underlying patterns in the dataset without overfitting to noise. The tuned model also achieved a better cross-validation score, indicating its robustness.

Model Performance:

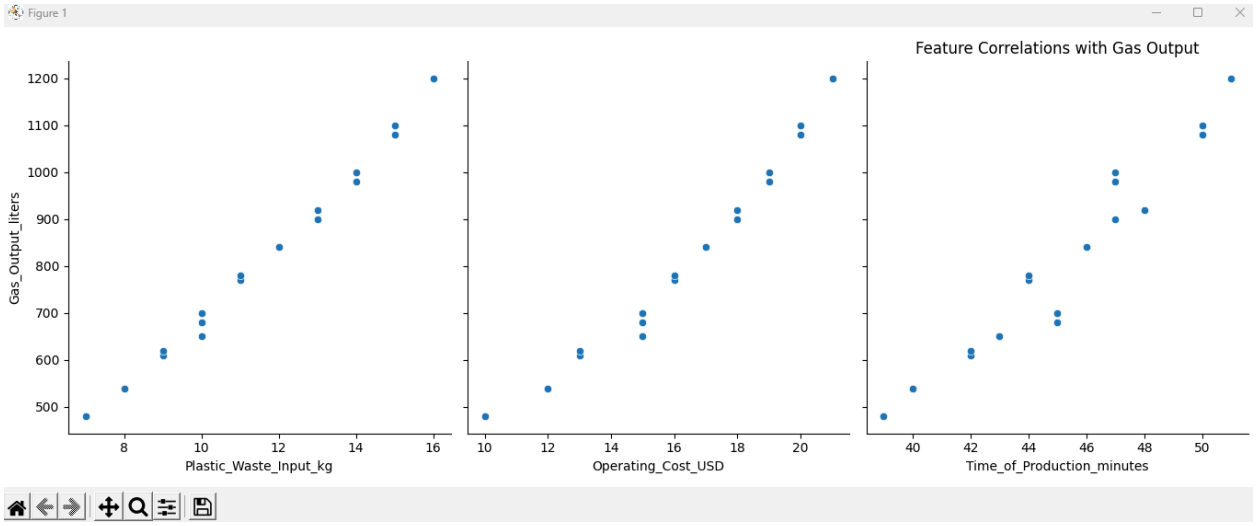
The model was evaluated using the following metrics:

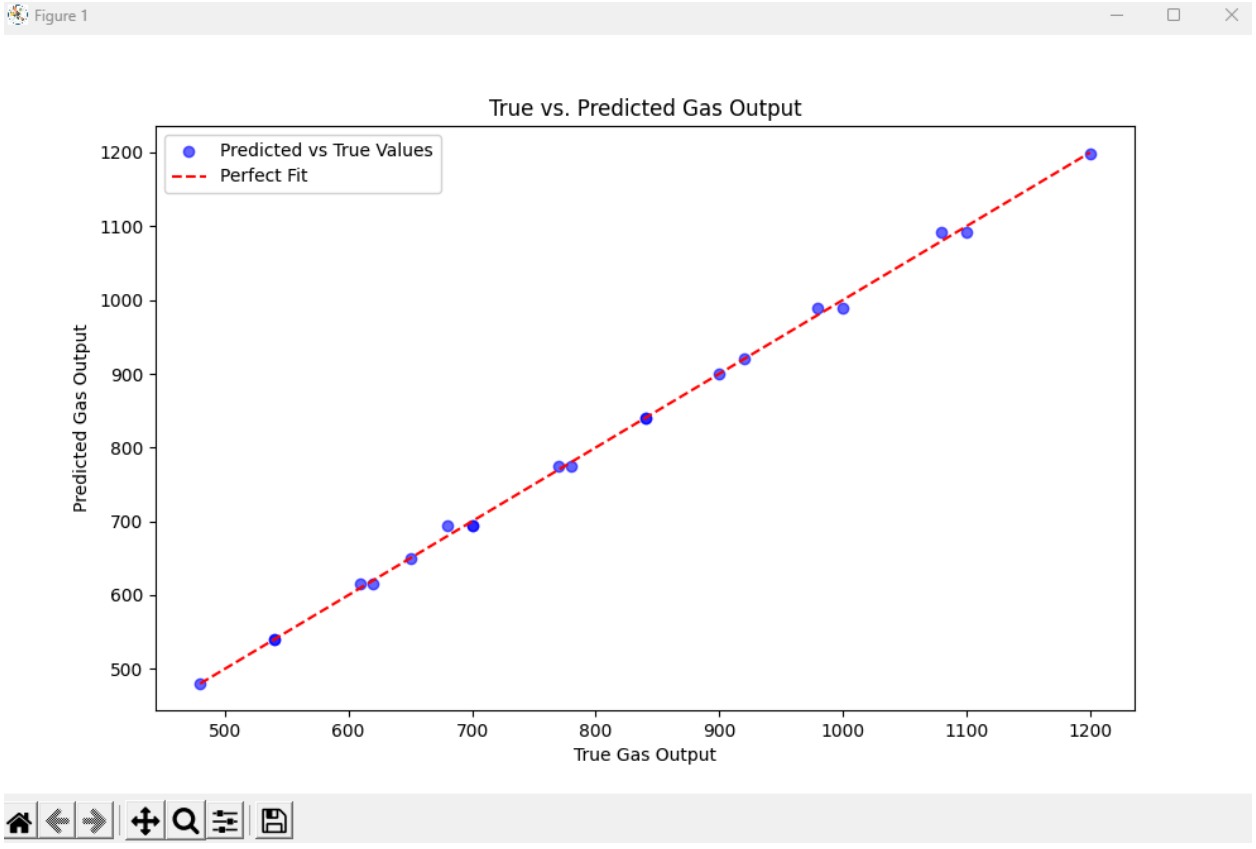
- **Mean Absolute Error (MAE):** 4.43
- **Mean Squared Error (MSE):** 39.11
- **R^2 Score:** 0.999

These metrics indicate that the model performs extremely well, with an almost perfect R^2 score, signifying that it explains nearly all the variance in the gas output data. The low MAE and MSE further indicate that the model's predictions are very close to the actual values, and the model can make accurate predictions on unseen data.

Data Visualization and Correlations:

Three key visualizations were generated to better understand the relationship between the input variables and the model's target variable (gas output in liters):





1. **Plastic Waste Input (kg) vs. Gas Output (liters):**
    - This plot indicates a positive correlation between the amount of plastic waste input and the gas output. As the plastic waste input increases, the gas output also increases, showing a clear linear relationship. This is significant because it demonstrates that the more plastic waste we process, the more gas we can produce efficiently, validating the scale-up potential of our process.
2. **Operating Cost (USD) vs. Gas Output (liters):**

- A positive correlation was also observed between the operating cost and gas output. As the operating costs rise, so does the gas output. This suggests that higher operating costs might be associated with better production efficiency or increased production capacity, implying that investments in better operational infrastructure could enhance gas production output.

3. **Time of Production (minutes) vs. Gas Output (liters):**
    - The plot shows a linear relationship between production time and gas output. The longer the production time, the more gas is produced. This indicates that extending production time allows for better utilization of inputs and resources, producing a higher yield of gas.

Conclusion:

The combination of the model's performance metrics and the insights from the visualizations demonstrate that the model is highly effective in predicting gas output from the input features. The positive correlations across all variables provide confidence in the scalability and efficiency of the Plas-tech process. Further optimization, such as reducing operating costs while maintaining production efficiency, can enhance overall output, making the process even more sustainable and economically viable.