AI-Powered Energy Project Cost Estimator

by

Sakshi A. Tapkir

General Engineering, San Jose State University

ISE 200: Financial Methods for Engineers

Professor Gaojian Huang

May 2, 2025

Table of Contents

Executive Summary

This project presents an AI-driven solution for automated cost estimation in energy projects. Leveraging a synthetic dataset of 200 projects, I developed a machine learning model that predicts total project costs based on key parameters such as equipment, material, labor, regulatory costs, and inflation. The model is deployed via a Gradio web interface, enabling real-time, user-friendly predictions. Feature importance and SHAP analysis confirm that equipment and material costs are the primary cost drivers. The tool demonstrates high predictive accuracy ($R^2 \approx 0.99$) and offers actionable insights for stakeholders in the energy sector.

Background

Accurate and timely cost estimation is critical for the planning and execution of energy projects. Traditional methods are time-consuming, error-prone, and often fail to capture the complex relationships between project parameters. With the increasing complexity and scale of modern energy projects, there is a need for automated, data-driven tools that can provide fast, reliable, and interpretable cost estimates to support decision-making.

Alternative Problems Considered

Several alternative problems were considered, including:

- **Manual Excel-based Cost Models:** While accessible, these lack scalability and adaptability to new data.
- **Rule-Based Estimation Systems:** These are rigid and do not learn from historical data, limiting their predictive power.
- **Other Engineering Domains:** I considered cost estimation for manufacturing and construction but selected energy due to its data richness and industry impact.

The chosen problem-automated cost estimation for energy projects-was selected for its relevance, the availability of structured features, and its alignment with current industry needs.

Methods Used to Solve This Problem

1. Project Overview

The Energy Project Cost Estimator is an AI-powered tool designed to predict the total cost of energy projects based on key inputs. It leverages a machine learning model trained on data from 200 synthetic energy projects, covering various energy types (Solar, Wind, Natural Gas, Hydro, Nuclear) and cost parameters.

- Purpose: Automate cost estimation for energy projects to assist decision-makers in budget planning and scenario analysis.
- Deployment Platform: Gradio Interface for user-friendly interaction.
- Key Features:
  - Input fields for project details (e.g., size, type, location, costs).
  - Real-time prediction of total project cost.

- Example scenarios for testing.

2. Data Generation

A synthetic dataset of 200 energy projects was generated, including features such as:

- Project Size (MW)
- Energy Type (Solar, Wind, Natural Gas, Hydro, Nuclear)
- Location (Coastal, Rural, Urban)
- Material, Labor, Equipment, Regulatory Costs
- Inflation Rate, Project Duration, Year, Environmental Impact

3. Data Preprocessing

- **Encoding:** Categorical variables were one-hot encoded.
- **Scaling:** Numerical features were standardized using StandardScaler.
- **Feature/Target Split:** Features (X) and target (y: Total Cost) were separated for modeling.

4. Model Training

- **Algorithms:** Random Forest Regressor and XG Boost Regressor were trained and evaluated.
- **Metrics:** Mean Absolute Error (MAE) and R^2 score were used for performance assessment.
- **Best Model:** Random Forest was selected (MAE ≈ $367,636, $R^2 \approx 0.99$).

5. Economic Analysis

- **Net Present Value (NPV):** Calculated for each project to account for inflation and project duration.

6. Interpretability

- **Feature Importance:** Identified key cost drivers (Equipment and Material Costs).
- **SHAP Analysis:** Visualized how individual features impact predictions.

7. Deployment

- **Gradio Interface:** Developed for real-time, user-friendly cost estimation based on model predictions.

8. Prediction Output

- The model predicts the Total Cost ($) of the energy project based on the input parameters.
- Example Prediction:
  - Input Parameters:
    - Project Size (MW): 262.84
    - Energy Type: Solar
    - Location: Coastal
    - Material Cost ($): 1,577,025.97
    - Labor Cost ($): 202,806.27
    - Equipment Cost ($): 842,797.53
    - Regulatory Cost ($): 78,851.30
    - Inflation Rate (%): 4.9
  - Predicted Total Cost: $1,118,598.10

9.Example Scenarios

The interface includes sample scenarios to demonstrate functionality:

1. Example 1:
   - Project Size: 262.84 MW, Energy Type: Solar, Location: Coastal.
   - Predicted Cost: $1,118,598.10.

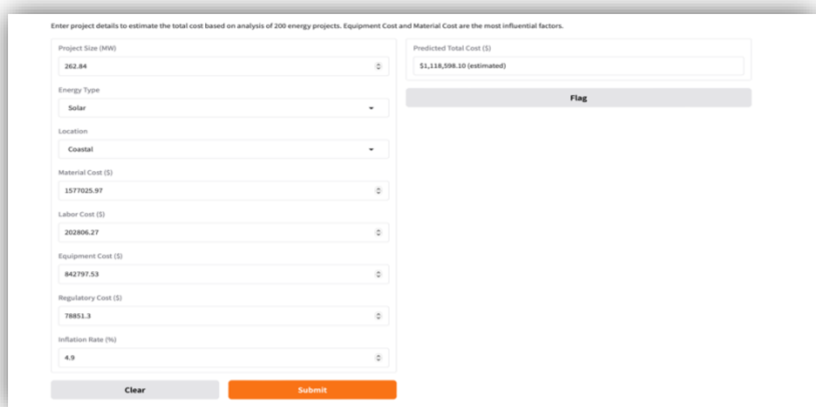

Image 1

2. Example 2:
    - Project Size: 115.54 MW, Energy Type: Wind, Location: Coastal.
    - Predicted Cost based on inputs.

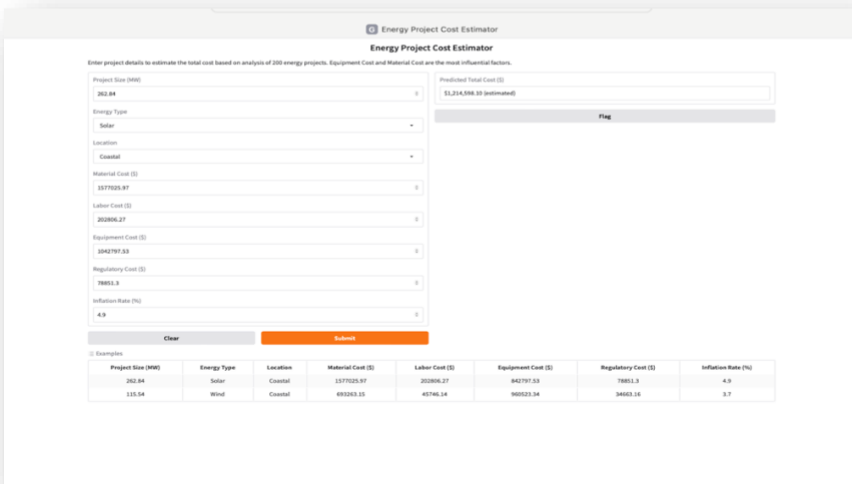

Image 2

These examples allow users to test the tool with pre-filled values.

Results

User Interface

- **Gradio App:** Allows users to input project details and receive instant cost estimates.
- **Example Output:** For a Solar project (262.84 MW, Coastal), the predicted total cost is **$1,118,598.10** (see Image 1).

Feature Importance

- **Chart (Image 3):** The image shows the **feature importance** from the Random Forest model.

1. **Key Cost Drivers**:
   - **Equipment Cost** (≈48%) and **Material Cost** (≈45%) are by far the most influential factors in determining total project costs
   - **Regulatory Cost** has a modest but noticeable impact (≈3%)
   - All other features combined account for less than 4% of cost prediction influence
2. **Low-Impact Features**:
   - Project size surprisingly has minimal direct influence
   - Energy type categories (Solar, Wind, etc.) show negligible importance
   - Location types and environmental impact are statistically insignificant

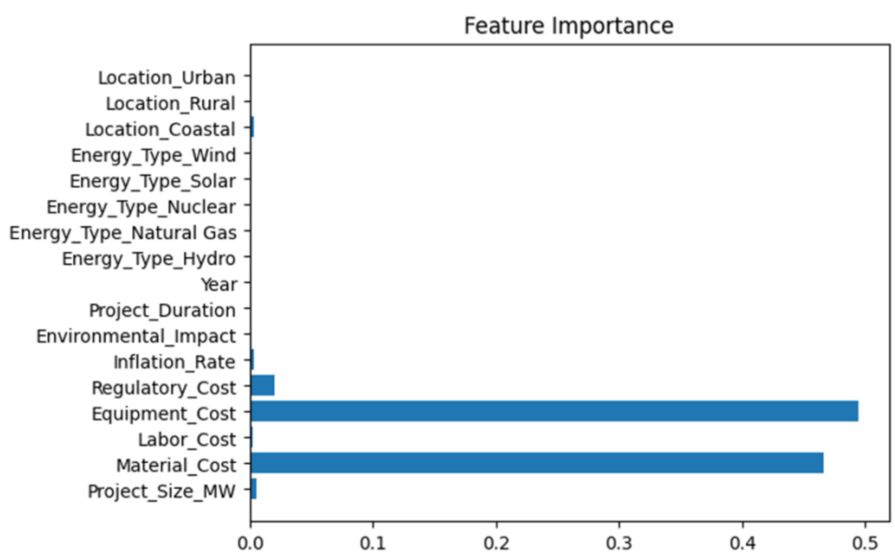

Image 3

SHAP Analysis

- **Summary Plot (Image 4):** The image shows the **SHAP (SHapley Additive exPlanations) values**:

1. **Feature Impact Distribution**:
   - Features 3 and 1 (likely corresponding to Equipment Cost and Material Cost) show the widest distribution of impacts
   - These features can both significantly increase costs (red dots toward right) or decrease costs (blue dots toward left)
   - Most other features cluster near zero impact
2. **Model Insights**:
   - The model's predictions are dominated by just two or three features
   - This aligns with the feature importance chart and confirms that our dataset has a clear cost structure

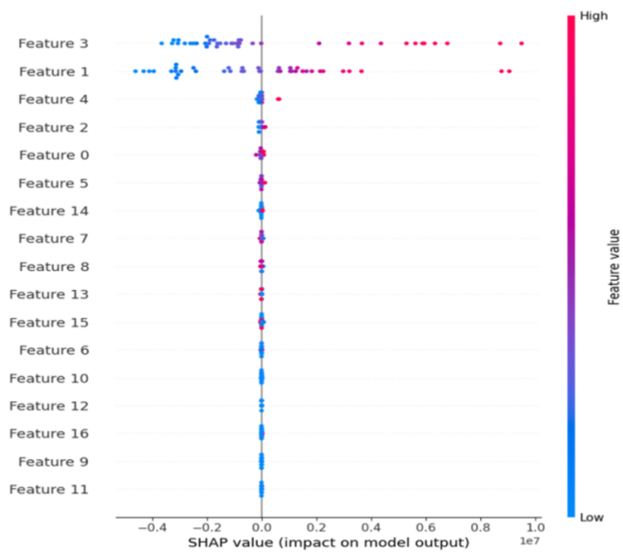

Image 4

Model Performance

- **Random Forest:** MAE (Mean Absolute Error) ≈ \$367,636; $R^2 \approx 0.99$
- **XG Boost:** MAE (Mean Absolute Error) ≈ \$524,358; $R^2 \approx 0.98$

Discussion

Key Insights

- **Dominant Cost Drivers:** Equipment and Material Costs are the most significant predictors of total project cost.
- **User Experience:** The Gradio interface provides a seamless, interactive experience for both technical and non-technical users.
- **Interpretability:** SHAP and feature importance analyses enhance trust and transparency in the model's predictions.

Limitations

- External Factors: Supply chain disruptions, policy changes, and other externalities are not modeled.
- Scalability: While Gradio is ideal for prototyping, enterprise deployment would require further development.

Future Work

- Expand features to include risk and sustainability metrics.
- Implement Monte Carlo simulations for uncertainty quantification.
- Deploy on scalable cloud platforms for enterprise use.

References

1. *Open Energy Information (Open EI). https://openei.org/*
2. *International Renewable Energy Agency (IRENA). https://www.irena.org/*
3. *Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30.*
4. *Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.*

Appendix

Appendix A: Python Code and explanations

Python Code:

- To demonstrate the depth and transparency of the cost estimation process, all Python code used in this project with explanations is included as a separate Word document attached to this report.

Appendix B: Excel Work

Excel File:

- The attached Excel file contains the underlying data and explanations for each project parameter used in the model.