Prediction Model for Energy Consumption at Manufacture-Industry (The DAEWOO Steel Co. Case)

Yusril Sabilla Ashar
Introduction Machine Learning Project

Abstract:
Analysis of energy consumption (KWh) in manufacturing companies. Conducting basic analysis and modeling to find out the most appropriate model for consumption prediction for companies. The result is that the Random Forest Regression model is the best model that can be used in prediction.

I. Introduction

In the turbulent modern era, the development of the manufacturing industry has become one of the main pillars of global economic growth. Manufacturing companies continue to play an important role in meeting society's expanding demand for various consumer products, industrial equipment, and advanced technologies. Along with the rapid growth in production come serious challenges related to the heavy use of energy in the manufacturing process.

This paper will explore the latest developments in the manufacturing industry and how changes in society's demand behavior have affected manufacturing companies. In addition, it will discuss the impact of heavy energy use in manufacturing processes, as well as innovative efforts made by companies to reduce their carbon footprint and improve energy efficiency.

The manufacturing industry's central role in creating jobs and providing products that meet consumer needs makes manufacturing companies an important pillar of economic development. However, with this growth also comes great responsibility regarding the use of natural resources, including energy. Growing awareness of environmental issues and the demand for more sustainable consumers have pushed manufacturing companies to look for more efficient and environmentally-friendly ways to run their operations.

II. Relate Works

As the manufacturing industry rapidly grows in an increasingly smart urban environment, efforts to predict efficient energy consumption are becoming increasingly important. In order to achieve sustainability and energy efficiency goals, research has evolved to create reliable energy consumption prediction models. One important contribution in this domain is the work entitled "Efficient Energy Consumption Prediction Model for a Data Analytic-Enabled Industry Building in a Smart City" by Sathishkumar V E, Changsun Shin, and Yongyun Cho. This research presents an efficient energy consumption prediction model capable of providing valuable insights for manufacturing companies operating in a smart city environment.

This paper describes a concrete effort in integrating data analytics in industrial energy consumption prediction. By analyzing data collected from various sensors and information sources, this research enables companies to optimize their energy usage in a smarter and more sustainable way. The results of this research can serve as a foundation for further development of energy consumption prediction models in an increasingly technology-connected manufacturing industry.

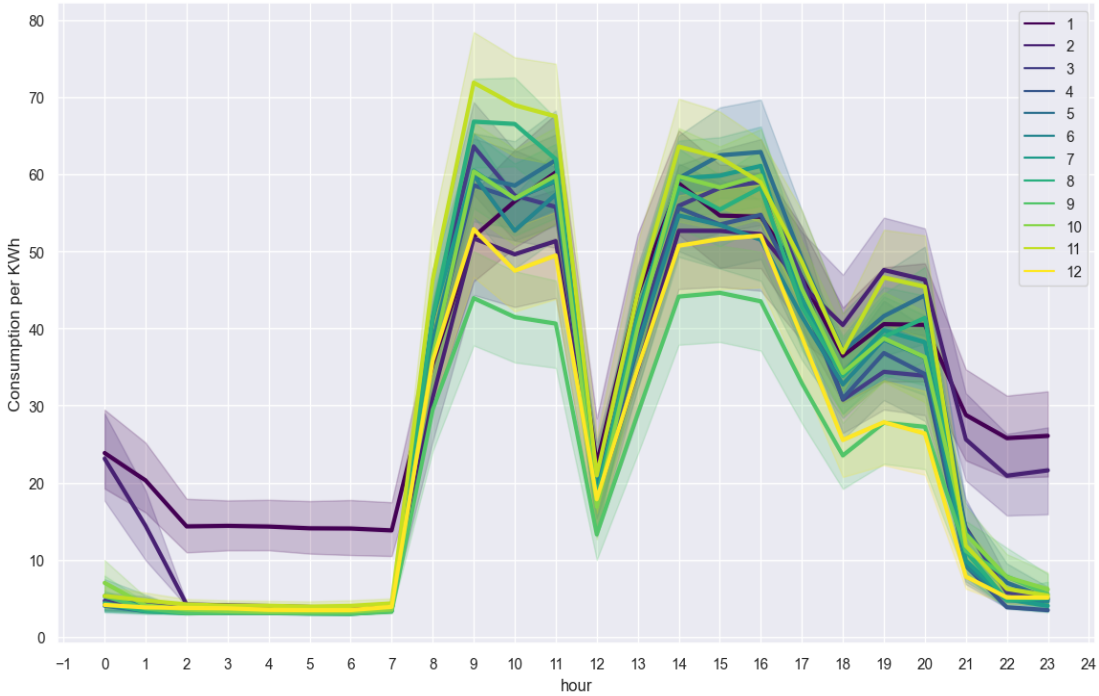
III. Methodology

I tried several different machine learning models to predict daily energy consumption. The models explored included:

- Linear Model: A linear regression model was used as the base model to understand the linear relationship between independent variables and energy consumption.
- Random Forest: Random Forest is an ensemble model that combines multiple decision trees to improve prediction accuracy.
- Gradient Boosting: Gradient Boosting model is used to overcome regression problems by combining predictions from weaker decision trees.
- AdaBoost: AdaBoost is an ensemble model that focuses on classifiers that were misclassified in previous iterations to improve prediction accuracy.
- Bagging: Bagging models are used to build multiple decision trees simultaneously and combine their results.
- Decision Tree: Decision Trees are used as the base model in some ensemble methods, as well as for comparison with other models.

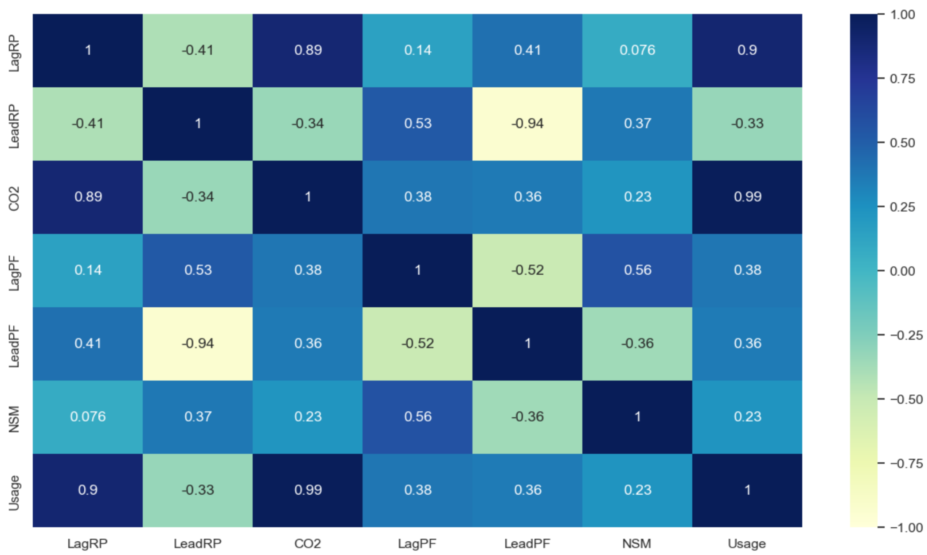
To compare the performance of different models, i use Mean Absolute Error (MAE) as an evaluation metric. MAE measures the average absolute value of the difference between the predicted and actual values.

IV. EDA



In the report, I observed KWh consumption trends throughout the day. Analysis of the data showed that between 7am and 11am, there was a significant increase in energy consumption. However, between 11am and 1pm, I observed a significant decrease in this energy usage, which can largely be attributed to the employees' break time. After 1pm, KWh consumption increased again and peaked at 4pm. This is an interesting phenomenon that reflects the daily activity patterns at the facility.

In analyzing the correlations between variables, I found some interesting relationships between various factors. Most striking, however, was the strong correlation between energy use (KWh) and CO2 emissions. The following correlation graph visualizes this relationship:



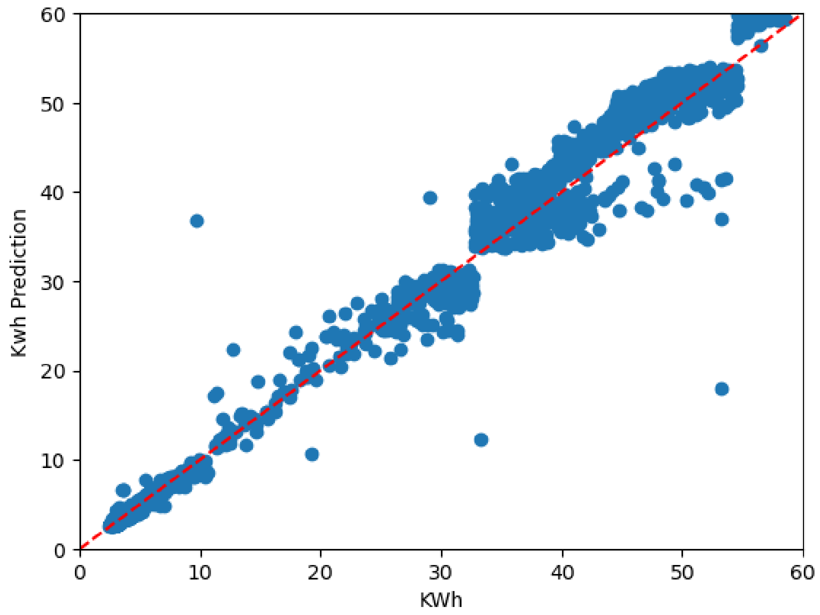
This graph illustrates with a jeta the significant correlation between energy use (KWh) and CO2 emissions, indicating that the higher the energy use, the greater the CO2 emissions. This highlights the importance of efforts to reduce energy consumption in order to reduce environmental impacts and carbon emissions.

V. Model

Model	**MAE Train**	**MAE Cross Validation**
Decision Tree	1.013998	1.196125
Bagging	0.149141	0.406396
Random Forest*	0.438629	1.124281
AdaBoost	2.603104	2.740038
Gradient Boost	6.399928	6.405598
Linear Regression	2.577146	2.580218

From these results, I can observe that the Bagging model gives the best performance with low MAE on both training and cross-validation data, demonstrating its ability to produce accurate predictions. The Random Forest model also showed good performance with low MAE on the training data, although the MAE on the cross-validation data was slightly higher. Meanwhile, the Decision Tree model showed promising results with a low MAE on the training data, although it was slightly higher on the cross-validation data.

However, there are some other models such as Gradient Boost, Linear Regression, and AdaBoost that seem to face difficulties in modeling the data well, indicated by higher MAE on both training and cross-validation data. In this context, the selection of a suitable model for daily energy consumption prediction is crucial, and further evaluation is needed to determine the best model for practical applications.



can be seen in the graph for good bagging performance at 10-30 KWh, therefore further research is needed, one of which checks for outliers in the data.