REPORT

Project Title: Energy Consumption of Steel Industry

Abstract: Energy used by the Steel Industry requires constant supervision and management. We introduce an Energy control dataset in order to anticipate the energy consumption by these industries. We utilize 5 machine learning regression models to analyze the data and give insightful information including predictions. In addition, we examine which training models, out of those utilized, are more reliable for this set of data.

Keywords: Regression, Energy Consumption, Steel Industry, Current, Machine learning

Introduction:

In this project we are going to analyze energy consumption patterns in the steel industry using the Steel Industry Energy Consumption Dataset. The dataset provides comprehensive information on energy usage and all other factors in the steel production process. By employing machine learning techniques, we aim to gain insights, make predictions, and contribute to the understanding of energy consumption in the steel industry.

Proposed Methodology:

Dataset: [Steel Industry Energy Consumption Dataset](https://archive.ics.uci.edu/ml/datasets/Steel+Industry+Energy+Consumption+Dataset)

Link: https://archive.ics.uci.edu/ml/datasets/Steel+Industry+Energy+Consumption+Dataset

Data Set Characteristics:	Multivariate	**Number of Instances:**	35040	**Area:**	Computer
Attribute Characteristics:	N/A	**Number of Attributes:**	11	**Date Donated:**	2021-03-30
Associated Tasks:	Regression	**Missing Values?**	N/A	**Number of Web Hits:**	30788

Source:
Sathishkumar V E,
Department of Information and Communication Engineering,
Sunchon National University, Suncheon.
Republic of Korea.
Email: srisathishkumarve@gmail.com

Data Set Information:
The information gathered is from the DAEWOO Steel Co. Ltd in Gwangyang, South Korea. It produces several types of coils, steel plates, and iron plates. The information on electricity consumption is held in a cloud-based system. The information on energy consumption of the industry is stored on the website of the Korea Electric Power Corporation (pccs.kepco.go.kr), and the perspectives on daily, monthly, and annual data are calculated and shown.

Attribute Information:

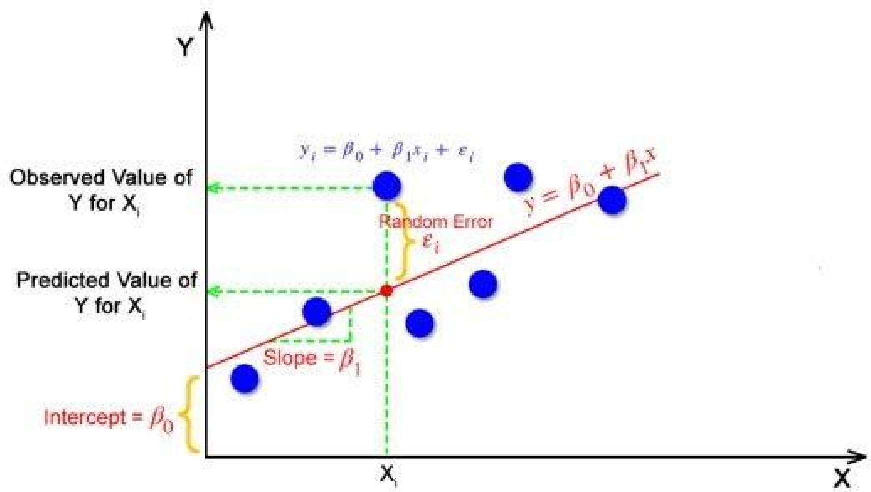
Data Variables	**Type**	**Measurement**
Date and Time	**Categorical**	DD/MM/YY
Industry Energy Consumption	**Continuous**	kWh
Lagging Current reactive power	**Continuous**	kVArh
Leading Current reactive power	**Continuous**	kVArh
tCO2(CO2)	**Continuous**	ppm
Lagging Current power factor	**Continuous**	%
Leading Current Power factor	**Continuous**	%
Number of Seconds from midnight	**Continuous**	S
Week status	**Categorical**	(Weekend (0) or a Weekday(1))
Day of week	**Categorical**	Monday to Sunday
Load Type	**Categorical**	Light Load, Medium Load, Maximum Load

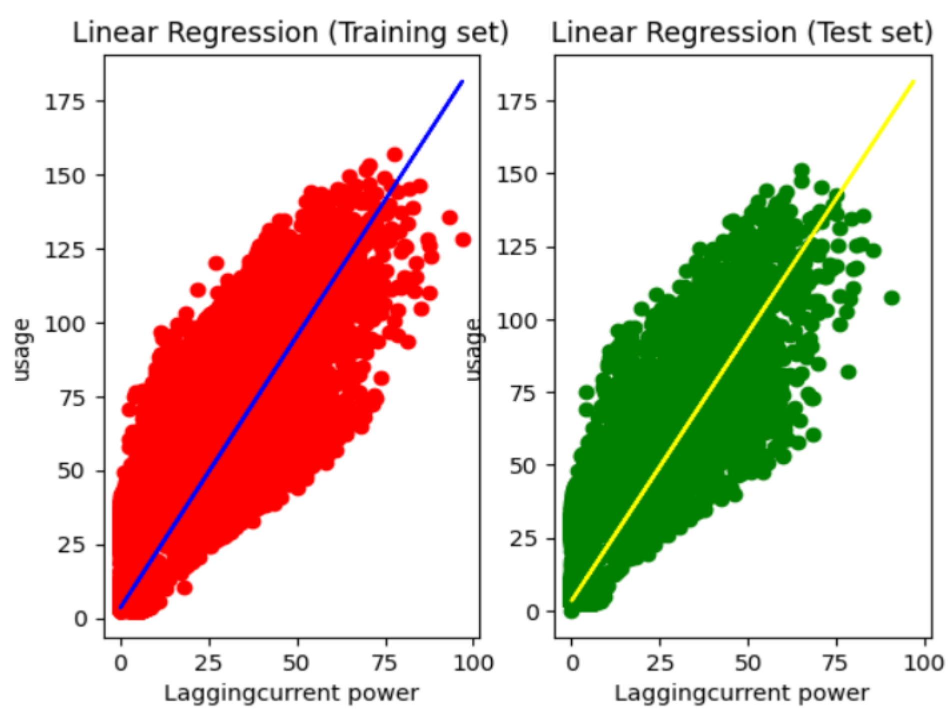
Here we are going to apply five types of regression on our dataset:

1. Simple Linear Regression
2. Multiple Linear Regression

3. Polynomial Regression
4. Lasso Regression
5. Ridge Regression

1. <u>Simple Linear Regression</u>
   Independent Variable: Lagging Current reactive power
   Dependent Variable: Industry Energy Consumption
   -Of the form y=mx+c
   -Univariate





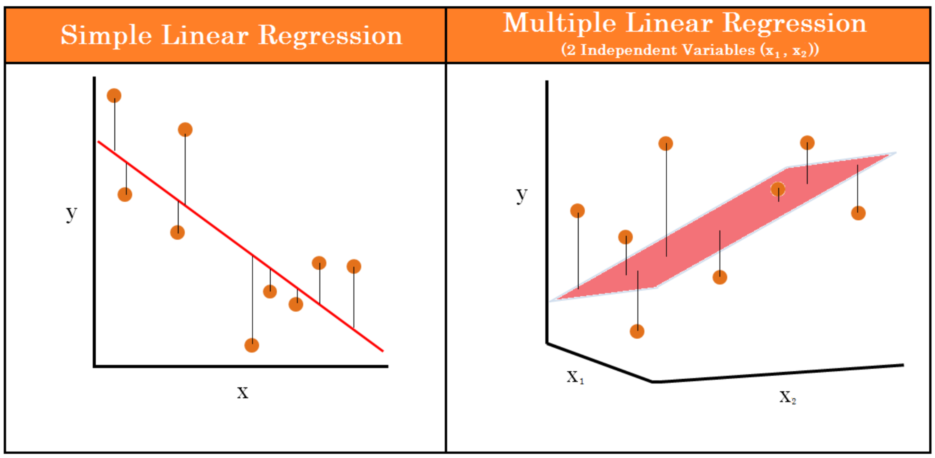
2. <u>Multiple Linear Regression</u>

   Independent Variable: Lagging Current reactive power, Leading Current reactive power, tCO2(CO2) , Lagging Current power factor, Leading Current Power factor

   Dependent Variable: Industry Energy Consumption

   -Of the form $y=m_1x_1+m_2x_2+......+m_nx_n+c$

   -Multivariate



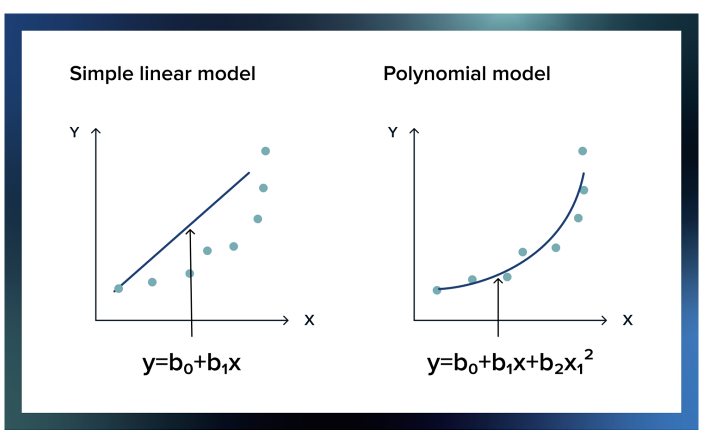
3. <u>Polynomial Regression</u>

   Independent Variable: Lagging Current reactive power, Leading Current reactive power, Lagging Current power factor, Leading Current Power factor

   Dependent Variable: Industry Energy Consumption

   -Of the form $y= b_0+b_1x_1+ b_2x_1^2+ b_2x_1^3+...... b_nx_1^n$

   -Multivariate



4. <u>Lasso Regression</u>

Independent Variable: Lagging Current reactive power, Leading Current reactive power, Lagging Current power factor, Leading Current Power factor

Dependent Variable: Industry Energy Consumption

Lasso regression is also called the Penalized regression method. This method is usually used in machine learning for the selection of the subset of variables. It provides greater prediction accuracy as compared to other regression models. Lasso Regularization helps to increase model interpretation.

The less important features of a dataset are penalized by the lasso regression. The coefficients of this dataset are made zero leading to their elimination. The dataset with high dimensions and correlation is well suited for lasso regression.

- Lasso Regression Formula:

D= Residual Sum of Squares or Least Squares Lambda * Aggregate of absolute values of coefficients

Lambda denotes the amount of shrinkage in the lasso regression equation.

The best model is selected in a way to minimize the least-squares.

Penalizing factor is added to form a lasso regression to the least-squares. The selection of the model depends upon its ability to reduce the above loss function to its minimal value.

All the estimated parameters are present in the lasso regression penalty, and the value of lambda lies between zero and infinity which decides the performance of aggressive regularization. Lambda is selected using cross-validation.

The coefficients tend to decrease and gradually become zero when the value of lambda is increased.

5. <u>Ridge Regression</u>

Independent Variable: Lagging Current reactive power, Leading Current reactive power, tCO2(CO2) , Lagging Current power factor, Leading Current Power factor

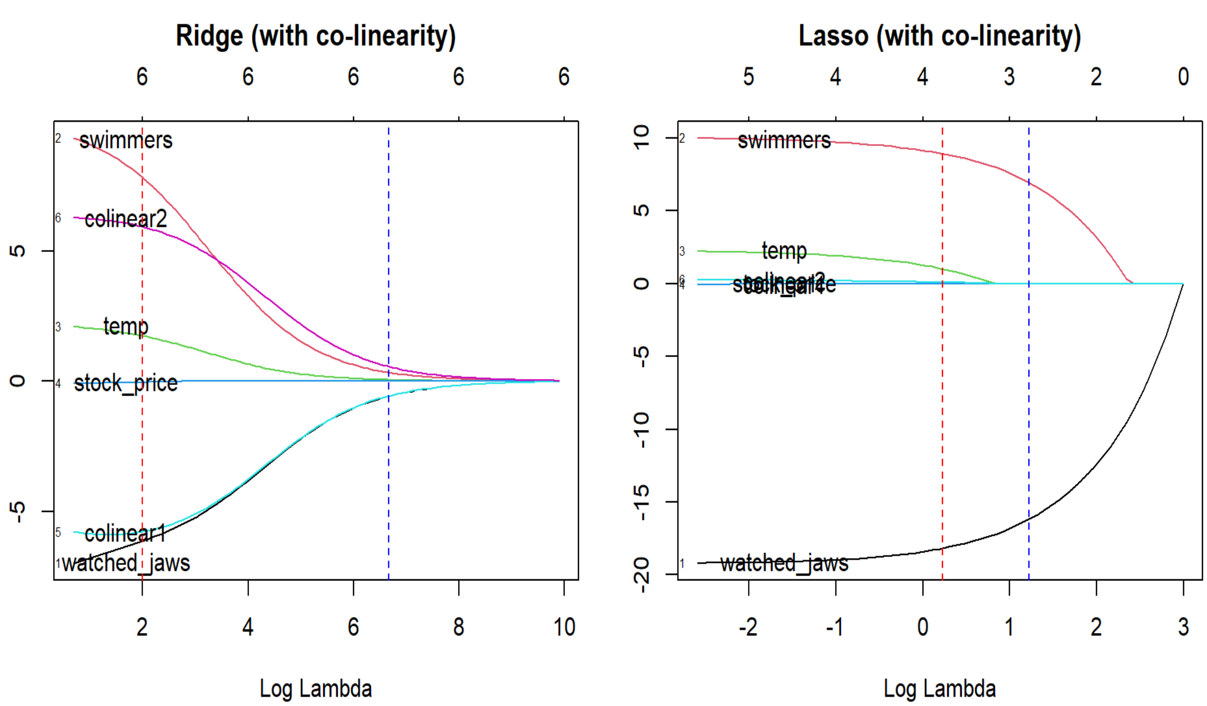
Dependent Variable: Industry Energy Consumption

In ridge regression, the first step is to standardize the variables (both dependent and independent) by subtracting their means and dividing by their standard deviations. For any type of regression machine learning model, the usual regression equation forms the base which is written as:

Y = XB + e

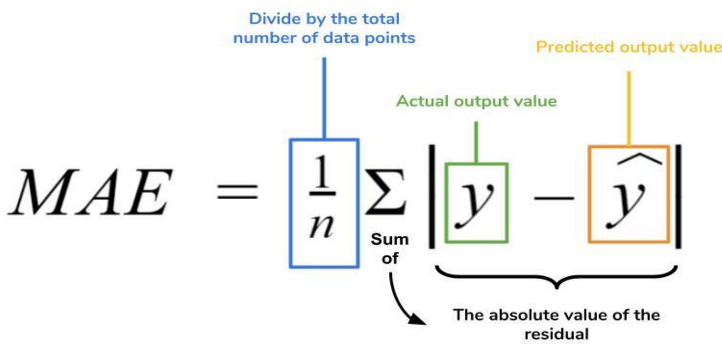
Where Y is the dependent variable, X represents the independent variables, B is the regression coefficients to be estimated, and e represents the errors are residuals.

Once we add the lambda function to this equation, the variance that is not evaluated by the general model is considered.



Comparison of all models used:

Using Mean Absolute Error:



Sr. no.	Regression Model	Mean Absolute Error
1.	Simple Linear Regression	10.630
2.	Multiple Linear Regression	2.569
3.	Polynomial Regression	1.113
4.	Lasso Regression	6.909
5.	Ridge Regression	6.885

Model with least MAE = Polynomial Regression

Model with most MAE = Simple Linear Regression

Result and Discussion:

For the purpose of predicting total energy usage in a steel power plant from the taken dataset, the most accurate prediction model (amongst the ones used in this project) is the Polynomial regression model with Mean absolute error of 1.113.

Conclusion and Future Work:

This study can further be studied and used for the noble purpose of analyzing energy consumption patterns in the steel industry, to gain insights, make predictions, and contribute to the understanding of energy consumption in the steel industry.

References:

1. Introduction to Categorical Data Analysis -ANUSHA ILLUKKUMBURA
2. CURVE and SURFACE FITTING with MATLAB. LINEAR and NONLINEAR REGRESSION -A. RAMIREZ
3. Holistic Theoretical Model For Optimal Multiple Linear And Multiple Nonlinear Regression Analysis -Mr Ramesh Chandra Bagadi
4. Regression Analysis -Jim Frost
5. Geeksforgeeks.com
6. Javatpoint.com
7. u-next.com
8. mygreatlearning.com
9. towardsdatascience.com
10. datacamp.com
11. archive.ics.uci.edu
12. scikit-learn.org
13. Hands-On Data Preprocessing in Python -Roy Jafari
14. The visual imperative -Lindy ryan