Paper 1805-2014

Improving the Thermal Efficiency of Coal-Fired Power Plants: A Data Mining Approach

Thanrawee Phurithititanapong and Jongsawas Chongwatpol
NIDA Business School, National Institute of Development Administration, Bangkok, Thailand

ABSTRACT

Power producers are looking for ways not only to improve efficiency of power plant assets but also to grow concerns about the environmental impacts of power generation without compromising their market competitiveness. To meet this challenge, this study demonstrates the application of data mining techniques for process optimization in a coal-fired power plant in Thailand with 97,920 data records. The main purpose is to determine which factors have a great impact on both (1) heat rate (kJ/kWh) of electrical energy output and (2) opacity of the flue gas exhaust emissions. As opposed to the traditional excel-based regression analysis currently employed at the plant, more complex analytical models using SAS® Enterprise Miner™ help supporting managerial decision to improve the overall performance of the existing energy infrastructure while reduce emissions through a change in the energy supply structure.

1. INTRODUCTION

This coal-fired power plant (XYZ Company, name disguised) provides the electricity to the Electricity Generating Authority of Thailand (EGAT) to meet the electricity demands in Thailand with Independent Power Producers Agreement (IPP).

The company is looking for ways not only to improve efficiency of power plant assets but also to grow concerns about the environmental impacts of power generation without compromising their market competitiveness. To meet this challenge, this study demonstrates the application of data mining techniques for process optimization in a coal-fired power plant in Thailand. The main purpose is to determine which factors have a great impact on both (1) heat rate (kJ/kWh) of electrical energy output and (2) opacity of the flue gas exhaust emissions.

- The amount of fuel energy input needed to produce electrical energy output (heat rate, kJ/kWh) is the key factor to measure the overall efficiency of the plant.

- For the combustion process in a coal-fired power plant, the opacity of the flue gas exhaust emissions is one of the performance measures, which has to comply with the mandatory standards for environmental protection. This power station produces the electrical power by using the good quality of bituminous coal and is installed with Flue Gas Desulfurization (FGD), Electrostatic Precipitator (ESP), Low Nitrogen Oxide (NO_x) Burner, and environment management equipment.

Currently, the company employs only traditional excel-based regression analysis to monitor the power plant performance. Additionally, only several factors related to fuel properties are considered in the analysis. As a result, many potentially important variables related to operational properties are neglected, accordingly (see Figure 1).

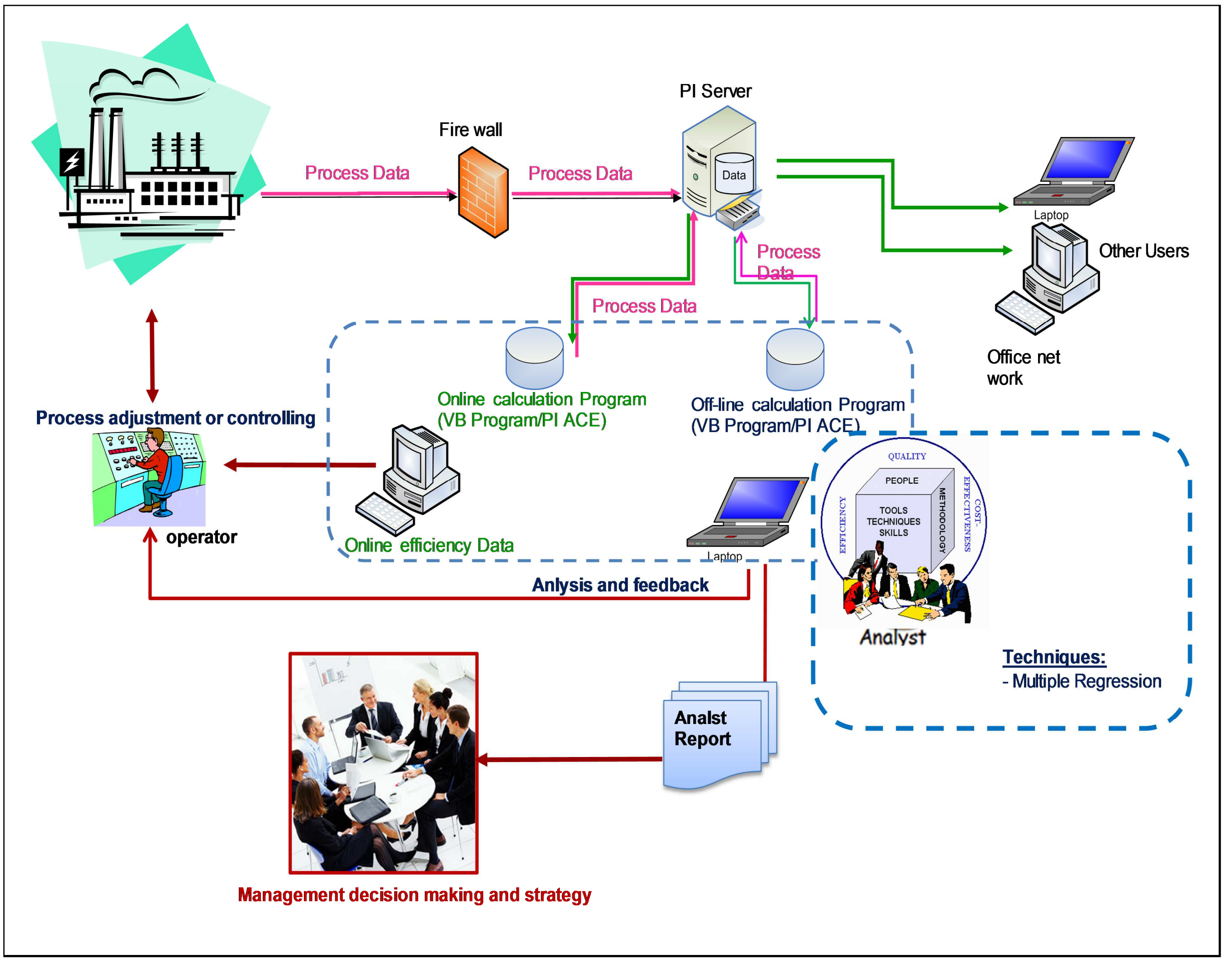


Figure 1: Existing Process

This study seeks to develop predictive models to support management decision making. Specially, two questions arise, "What are the most important factors that influence the overall efficiency and heat rate of power plant?" and "What factors have a great impact on stack emission when the coal qualities are changed in the combustion process?" To answer these research questions, we examine whether more complex analytical models using several data mining methodologies and algorithms can better predict and explain the variation of energy generation and gas exhaust emissions. We follow the CRISP-DM Model (Cross Industry Standard Process for Data Mining) as a guideline for this data mining project.

2. RESEARCH FRAMEWORK

In this study, we provide the in-depth analysis on how data mining approach can be a great help to improve overall plant performance. The data set contains a total of 105 variables with 17,376 records to determine which factors have a great impact on heat rate (kJ/kWh) of electrical energy output and a total of 60 variables with 29,952 records to determine which factors have a great impact on opacity of the flue gas exhaust emissions. The dataset is derived from approximately six months of power plant operations. Figure 2 presents the research framework of this study.

- Exploration: To study the historical operation data of the coal-fired power plant. The historical data must be verified and categorized based on the electricity output generation. The data obtained from DCS or data acquisition system, which includes the data necessary to calculate total boiler output, turbine output, fuel heat input, stack emission as well as the operational control of each major process component. Any missing values, errors, or extreme values are treated appropriately.

- Model building and validation: after partitioning data into training and testing dataset, three popular data mining techniques including multiple regression, decision trees, and neural networks are used to predict heat rate and opacity of flue gas exhaust emissions

- Compare the difference models: The complicity of the model is controlled by fit statistics calculated on the validation dataset. We use Average Squared Error (ASE) as criteria to select the best model.

- Deployment: using the model selected as best in the previous stage and applying it to new data in order to generate predictions or estimates of the expected heat rate and emission.

- Implications: To support managerial decision in improving the overall performance of the existing energy infrastructure while reduce emissions through a change in the energy supply structure.

3. DATA EXPLORATION

Table 1 presents an example of both dependent and independent variables of this study. Our first task is to get a sense of the dataset for any errors, inconsistencies, errors, or outlier (extreme values) in the data. Frequency distribution, descriptive statistics, and cross-tab analysis are used in this section. We then explore the historical data of the coal-fired power plant to understand the coal characteristic, which is the most impact on both opacity removable efficiency and the electricity output generation. The data is obtained from Data Acquisition System, which includes the data necessary to calculate the total boiler output, turbine output, fuel heat input, stack emission as well as the operations control of each major process component. The key findings are presented as follows:

Opacity

- There are two groups of parameters which are related to the inducted and reduced level of opacity in flue gas.
- Among these properties, the coal characteristics are the most influential factors on stack opacity as shown in the worth index and correlation statistic (Figure 3).

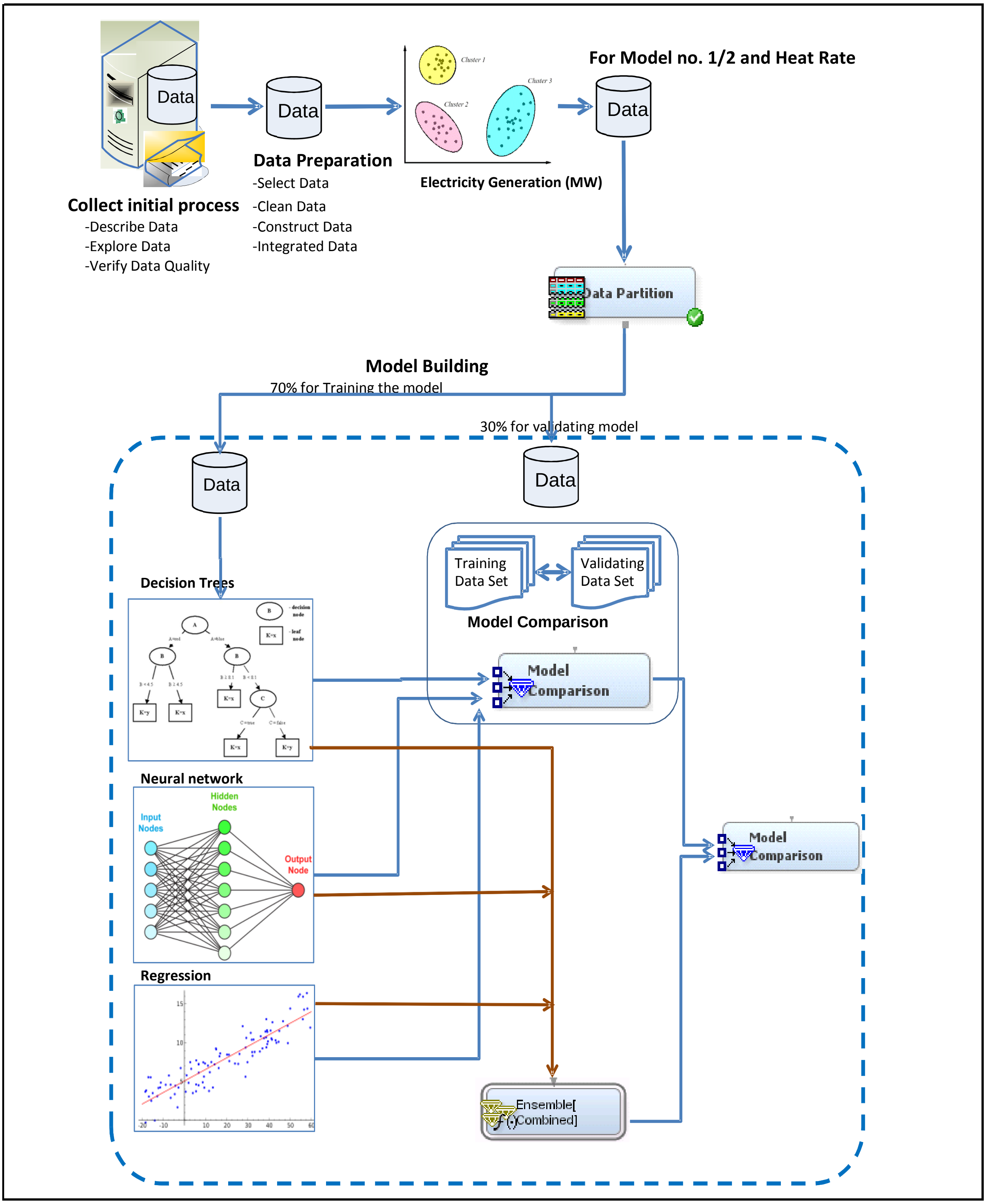
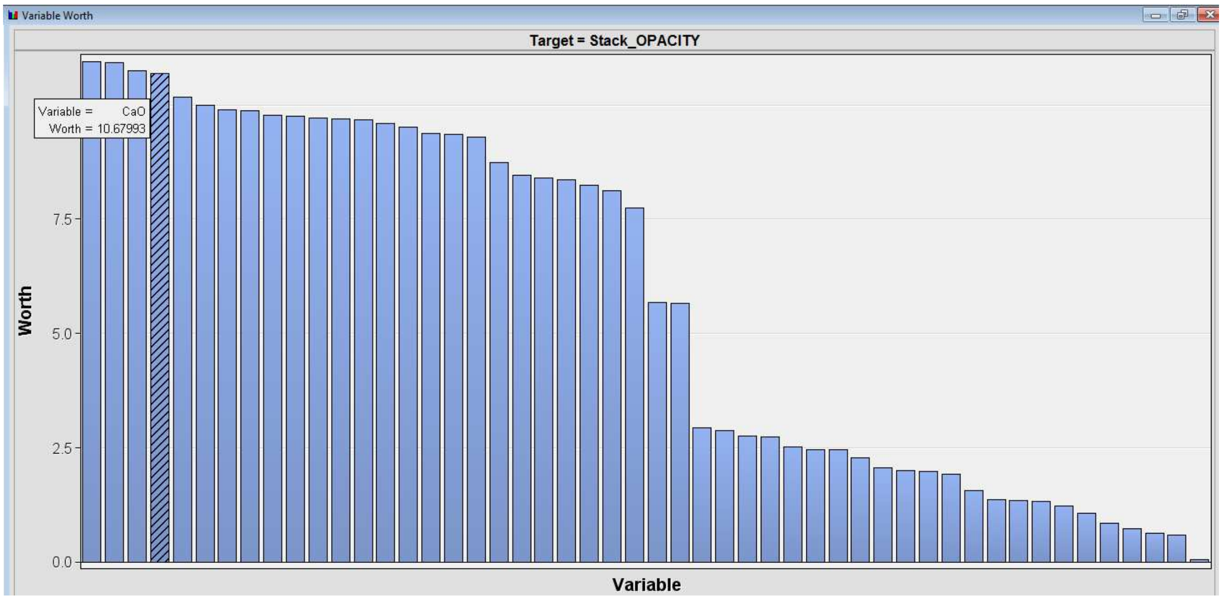


Figure 2: Research Frameworks

Table 1: An Example of Dependent and Independent Variables

Variable_Type	Variable	Model Role	Data_Type	Description
Dependent (Target)	Heat rate_kJ/kWh	Target model #1	numeric	The Energy input rate for producing one unit of electricity
	- Unit Boiler efficiency (%)	Target model #1.1	numeric	Unit Boiler Efficiency (%)
	- Unit Turbine efficiency (%)	Target model #1.2	numeric	Unit Turbine Efficiency (%)
	Emission_Opacity_%	Target model #2	numeric	Stack Emission Opacity (%)
Independent	Electricity Generation_MW	Input	categorical	Electricity Generation
	AH replacing element	Input	Binary	Air Heater replacing element
	HPHs Cleaning	Input	Binary	High Pressure Heaters Cleaning
Fuel_Properties	- coalPA_GCV_HtVl	Input	numeric	Coal analysis Higher Heating Value kJ/kg
	- coalPA_Moist_Mrat	Input	numeric	Coal analysis Total Moisture
	- coalPA_FC_Mrat	Input	numeric	Coal analysis Fixed Carbon
	- coalPA_VM_Mrat	Input	numeric	Coal analysis Volatile Matter (Dry Ash Free basis)
	- coalPA_Ash_MRat	Input	numeric	Coal analysis Ash
	- coalPA_TS_MRat	Input	numeric	Coal analysis Total Sulfur
	- coalPA_Cl_MRat	Input	numeric	Coal Ultimate Analysis percent Chloride
	- coalPA_HGI_MRat	Input	numeric	Coal analysis Hard Grove Index (HGI)
	- coalUA_C	Input	numeric	Coal Ultimate Analysis percent Carbon (As Received basis)
	- coalUA_H	Input	numeric	Coal Ultimate Analysis percent Hydrogen (As Received basis)
	- coalUA_N	Input	numeric	Coal Ultimate Analysis percent Nitrogen (As Received basis)
	- CInAsh_MRat	Input	numeric	Ash analysis percent Unburn Carbon
	- coalAsh_Resistivity	Input	numeric	Ash analysis Resistivity
	- coalAsh_Na2O	Input	numeric	Ash analysis Na2O
	- coalAsh_SO3	Input	numeric	Ash analysis Ash SO3
Operation conditions	- O2_Econ_Out	Input	numeric	Economizer outlet O2
	- Total coal flow	Input	numeric	Total coal flow
	- Stack inlet gas CO	Input	numeric	Stack inlet gas carbon mon- oxide
	- AH-A outlet flue gas temp	Input	numeric	AH-A outlet flue gas temp
	- AH-B outlet flue gas temp	Input	numeric	AH-B outlet flue gas temp
	- Temp_1ryinlet	Input	numeric	Primary Air inlet temperature
	- Temp_2ryinlet	Input	numeric	Secondary Air inlet temperature



Correlation Statistics
(maximum 500 observations printed)

Data Role=TRAIN Type=PEARSON Target=Stack_OPACITY

Input	Correlation
Al2O3	0.71583
TiO2	0.66822
FixCarbon	0.61431
ADJUSTABLE_CONTROL_DRIVE_AA_C	0.52552
ADJUSTABLE_CONTROL_DRIVE_AA_D	0.52215
HGI	0.51871
AUX_DAMPER_U_AA_D	0.34448
UPPER_AA_DAMPER_BIAS	0.30657
TotalMoisture	0.30157
AUX_DAMPER_U_AA_A	0.26841
AUX_DAMPER_U_AA_B	0.25996
AUX_DAMPER_U_AA_C	0.25829
Total_AirFlow	0.24990
AUX_DAMPER_L_AA_D	0.20623
C	0.20227
Total_CoalFlow	0.19009
CF_A	0.17823
Gross_MW	0.16509
LOWER_AA_DAMPER_BIAS	0.16337
AUX_DAMPER_L_AA_A	0.12460
N	0.12167
DPMist_Lower	0.11396
AUX_DAMPER_L_AA_B	0.11384
AUX_DAMPER_L_AA_C	0.11342
BUF_FLOW	0.07720
Amb_temp	0.03363
Furnace_Draft	0.00774

GCV	-0.07622
DPMist_Upper	-0.13601
O2_ECOout	-0.17306
Mn3O4	-0.17906
P2O5	-0.19387
AH_A_Fgouttemp	-0.20731
K2O	-0.23059
SeaWT_spray_Upper	-0.24176
SiO2	-0.26476
AH_B_Fgouttemp	-0.28253
Ash	-0.32558
Cl	-0.39538
VM	-0.43199
Fe2O3	-0.44895
H	-0.52816
Sulphur	-0.53197
SO3	-0.68995
MgO	-0.70216
FoulingIndex	-0.70677
ESP_K_Factor	-0.73033
CaO	-0.75081
SlaggingIndex	-0.75599
Na2O	-0.75660

Figure 3: Variable Worth and Correlation Statistic

- The coal properties are not normal distributed but there is no need to transform the data before building the predictive models since, commonly, the coal properties, which are fired in the combustion process, are varied within the acceptable range of the mutual-agreed purchasing contract.

Heat Rate

- There are two groups of parameters which are related to the inducted and reduced factors of the Heat Rate. The Figure 4 presents the Auxiliary load, which is the most influential factor to induce Heat Rate; meanwhile the Main Steam Pressure is the most influential factor to reduce Heat Rate.
- As presented in Figure 5, almost independent variables used to predict heat rate in the electricity generation are skewed and have a great impact on the operation control process.
- Additionally, more than 92% of the data of electricity generation is higher than 670 MW – 720 MW.

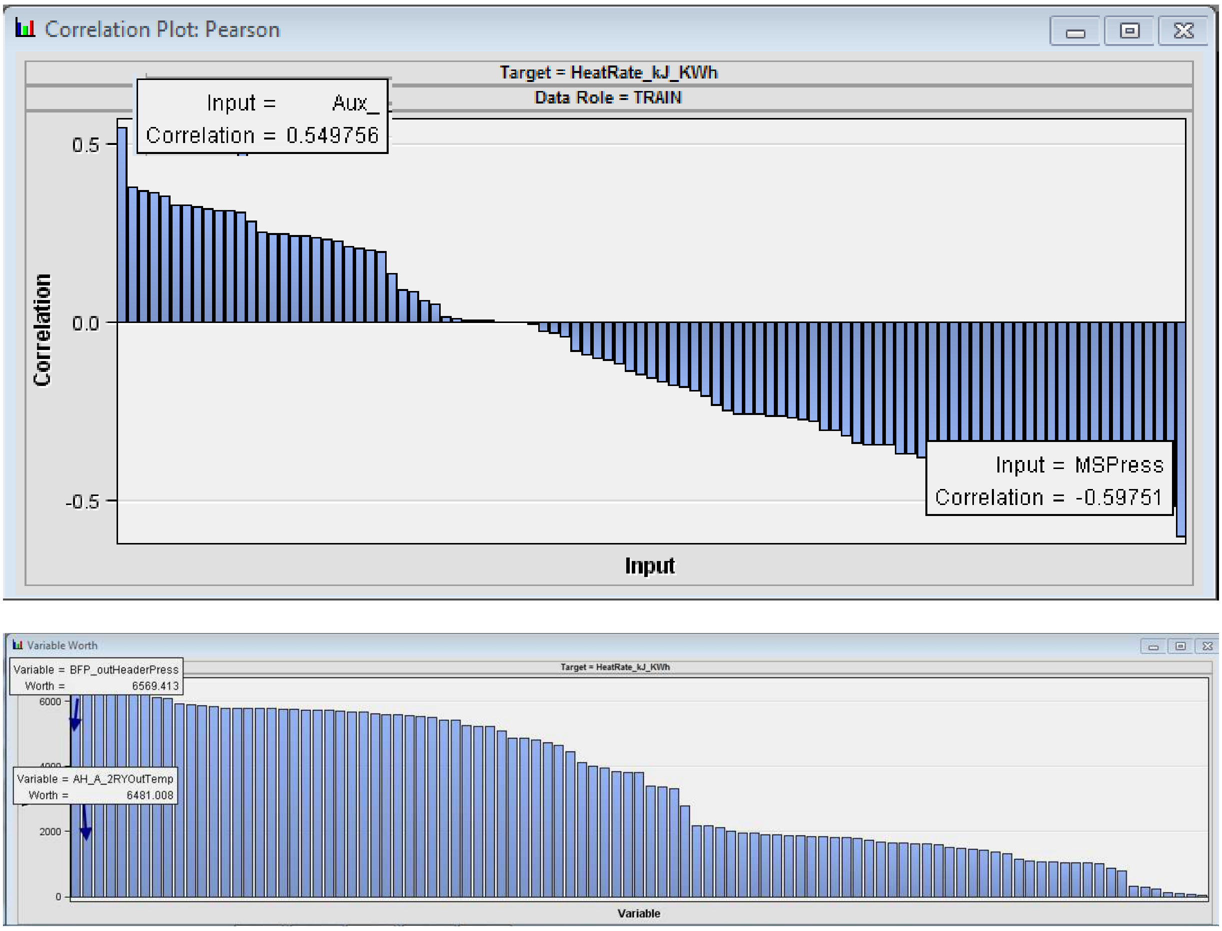


Figure 4: Variable Worth and Pearson Correlation Plot

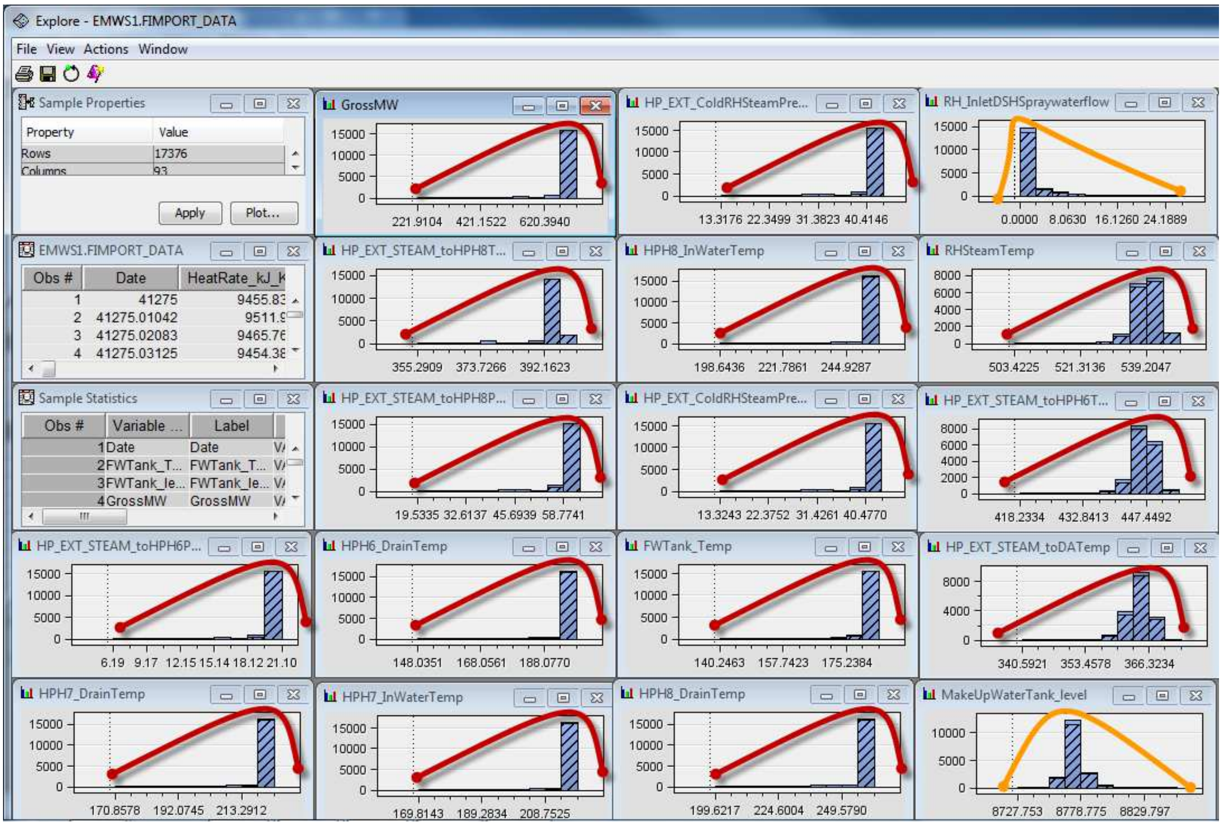


Figure 5: An Example of the Distribution of Independent Variables for Heat Rate Prediction Model

4. MODELING

After excluding variables with outliers and high missing values, we first develop predictive models on the original sample dataset with 105 variables for heat rate prediction model and 60 variables for opacity prediction model. For this preliminary analysis, we decided to do the pilot-model testing to get some senses of target variables, model development, and output analysis. We decide to partition our data into training (50%) and testing (50%) dataset. Figure5 presents an example of the process diagram of the heat rate prediction model. The total of 7 predictive models including (1) Decision Tree with maximum 2 branches, (2) Decision Tree with maximum 3 branches, (3) Neural Network with 1 hidden layer and 3 hidden units, (4) Neural Network with 1 hidden layer and 6 hidden units, (5) AutoNeural, (6) Linear Regression, and (7) Polynomial Regression models.

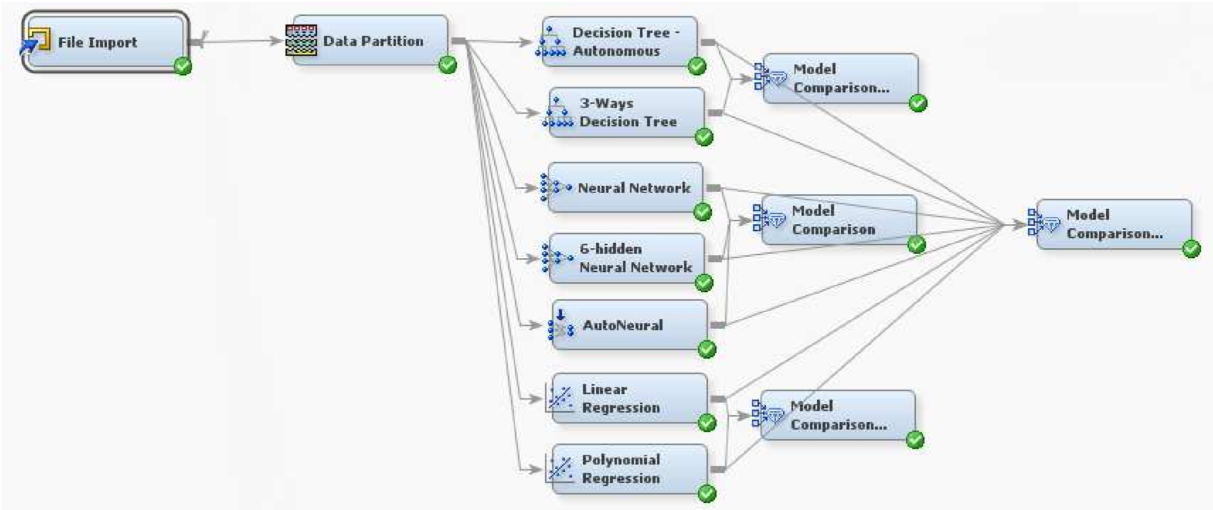


Figure 5: SAS® Enterprise Miner™ – Process Flow

5. RESULTS AND DISCUSSION

The Opacity of the Flue Gas Exhaust Emissions:

The Average Squared Error (ASE) is evaluated on the testing dataset. The lower the ASE, the better the model to be selected. In this study, the best model is Decision Tree with the lowest ASE of 1.9318, followed by Neural Network (ASE = 1.933) and Regression (ASE = 2.252). Figure 6 presents an example of the Decision Trees model. The following example illustrated a rule-based prediction:

For Combustion with Coal A:

IF “CaO” is less than 3.1025 ***AND*** “Sulphur” is greater than 0.4228 ***AND*** “N” is greater than 1.46 ***AND*** “HGI” is greater than 50 ***AND*** “total_Coal” is greater than 233.89 ***AND*** “DPMist_L” is less than 0.655, ***THEN*** the expected opacity emission is 15.2365%

For Combustion with Coal B:

IF “CaO” is greater than 3.1025 ***AND*** “N” is less than 1.322 ***AND*** “DPMist_L” is greater than 5.723 ***AND*** “N” is less than 1.2768 ***AND*** “C” is less than 62.087 ***AND*** “AH_B_Fgo” is less than 122.9119, ***THEN*** the expected opacity emission is 10.0673%

Figure 7 presents an example of applying the Decision Tree model to the new dataset in order to generate predictions or estimates of the expected opacity emission.

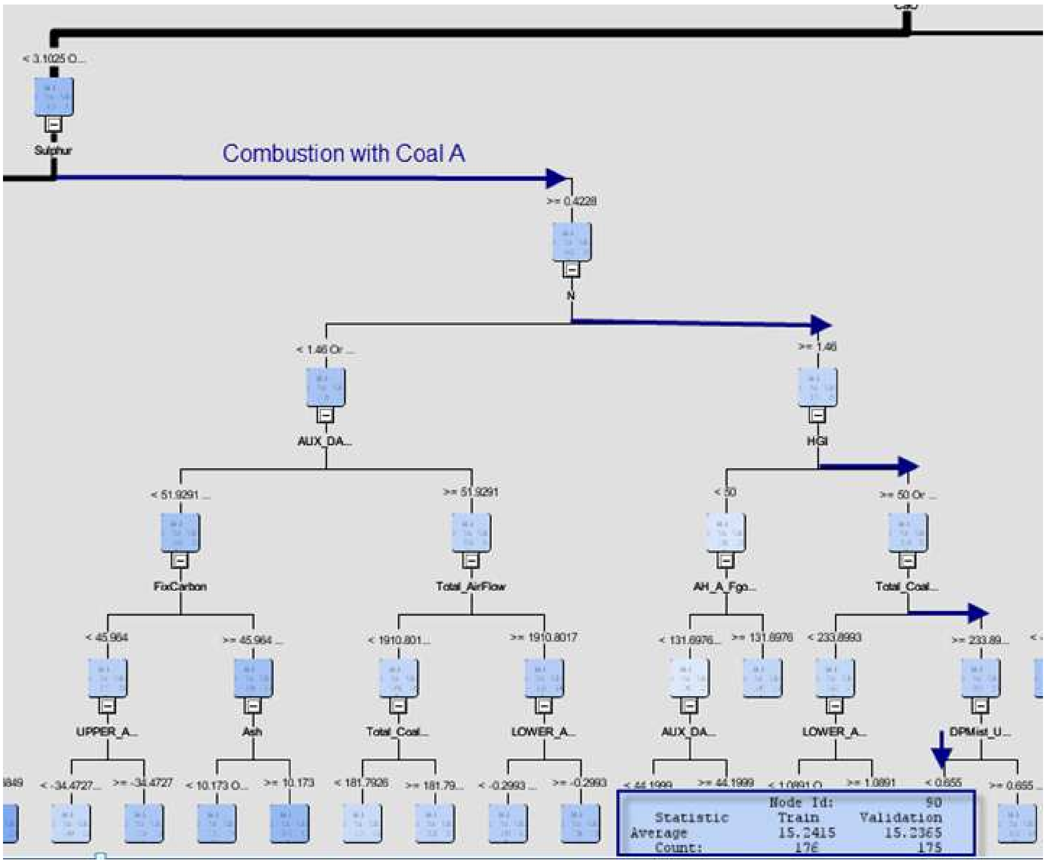


Figure 6: An Example of Decision Tree Models

Shipment	1	2	3
As Received	A	B	C
GCV (ar) (kJ/kg)	26531.75	21620.64	24857.03
Total Moisture	9.60	25.21	14.60
Fix Carbon	47.00	39.17	50.20
VM	30.30	31.43	24.83
Ash	13.10	4.20	10.40
Sulphur	0.49	0.34	0.27
Cl	0.02	0.01	0.02
HGI	50.00	51.00	70.00
C	64.30	53.01	61.64
H	4.08	3.65	3.34
N	1.49	0.59	1.47

Shipment No.	1	2	3
Ash Analysis	A	B	C
SiO_2	66.50	32.34	61.10
Al_2O_3	21.70	17.71	32.30
Fe_2O_3	5.00	4.69	1.74
CaO	1.04	14.49	0.40
MgO	0.88	3.36	0.33
Na_2O	0.47	6.45	0.11
K_2O	1.27	0.69	1.00
TiO_2	1.02	1.18	1.76
Mn_3O_4	0.05	0.04	0.02
SO_3	0.77	17.12	0.17
P_2O_5	0.20	0.25	0.28
Total	98.90	98.32	99.21
Fouling Index	0.050	3.740	0.004
Slagging Index	0.050	0.260	0.012
ESP K-Factor	28	26291	2

Figure 7: An Example of Predicted New Case

Heat Rate

The best model to predict heat rate is regression model with the ASE of 5374.13, followed by Decision Tree (ASE = 5801.77), and Neural Networks (ASE = 7118.92). Additionally, we also employ the Ensemble model, which combine all of models together to create a new model. The results show that the ASE of Ensemble decreases significantly at 2172.84. Figure 8 presents an example of stepwise regression when considering the gross generation MW greater than 50 MW. Out of the 105 independent variables, only 15 variables are left in the model. These important variables to predict heat rate in the electrical generation includes Aux, BFP_outHeaderTemp, CF_F, Cl, FixCArbon, HO_EXT_STEAM_toDATemp, IMP_AH_B_FGOut, IMP_condCWIn_TEMP, MSPress, MSTEmp, MgO, Na_2O , Pul_F_PA_flow. RHSteamTemp, and _RY_InSteamTemp_B

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	15473.9	408.7	37.86	<.0001
Aux_	1	95.5022	9.5909	9.96	<.0001
BFP_outHeaderTemp	1	16.7037	1.9714	8.47	<.0001
CF_F	1	-4.0114	0.2863	-14.01	<.0001
Cl	1	-589.0	188.0	-3.13	0.0017
FixCarbon	1	-12.0119	1.1871	-10.12	<.0001
HP_EXT_STEAM_toDATemp	1	-13.6242	1.3277	-10.26	<.0001
IMP_AH_B_FGOut	1	4.6191	0.3463	13.34	<.0001
IMP_condCWIn_Temp	1	25.8577	0.8490	30.46	<.0001
MSPress	1	-38.0055	1.8633	-20.40	<.0001
MSTemp	0	0	.	.	.
MgO	1	51.5915	6.2157	8.30	<.0001
Na2O	1	-32.4651	2.5998	-12.49	<.0001
Pul_F_PA_flow	1	7.3569	0.3238	22.72	<.0001
RHSteamTemp	1	8.4396	1.0076	8.38	<.0001
_RY_InSteamTemp_B	1	-11.8966	0.9438	-12.61	<.0001

Figure 8: Stepwise Regression Model

6. CONCLUSION AND IMPLICATIONS

This study demonstrates that data mining based approaches can be used to assess predictor variables influencing the stack opacity emission and heat rate in the energy generation process. As opposed to the traditional descriptive statistical analysis methods or the approaches adopting only expert-selected variables, the employment of regression, decision trees, or neural network models provide an interesting factors to understand the variation in both heat rate and opacity emission generated. For the opacity emission, Decision Trees explicitly shows better prediction results with the lowest average squared error (ASE); meanwhile, stepwise regression is the best model to predict heat rate.

Note that we present this study as a pilot study to determine if appropriate data is available, to understand the exploration of data mining approach in the coal-fired power plants, and to develop initial models to determine which factors influence the opacity of the flue gas exhaust emissions and heat rate of electrical energy output. Further analysis is required; especially when we can classify the sample data into subgroup based on the different range of electricity generation (heat rate model) or stack opacity emission (for opacity model) before building the predictive models and compare the results with the baseline model. Lastly, a larger sample sizes on both models will be tested to ensure the generalizability of our findings.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Name: Thanrawee Phurithititanapong
Enterprise: NIDA Business School, National Institute of Development Administration
Address: No. 9, I-8 Road, Map Ta Phut Industrial Estate, Map Ta Phut, Muang, Rayong, 21150, Thailand
Email: ya.janya@gmail.com

Name: Jongsawas Chongwatpol, Ph.D.
Enterprise: NIDA Business School, National Institute of Development Administration
Address: 118 Seri Thai Road, Bangkapi, Bangkok, 10240 Thailand
Email: jongsawas.c@ics.nida.ac.th, jong_tn@hotmail.com

Thanrawee Phurithititanapong is a performance engineer, technical services section in operations department at one of the top coal-fired power plants in Thailand. She is currently an MBA student in NIDA Business School at National Institute of Development Administration. She received her BE in mechanical engineering from Khonkaen University, Khonkaen, Thailand. Her research interest is in applying analytics to support managerial decision to improve the overall performance of the existing energy infrastructure while reduce emissions through a change in the energy supply structure.

Jongsawas Chongwatpol is a lecturer in NIDA Business School at National Institute of Development Administration. He received his BE in industrial engineering from Thammasat University, Bangkok, Thailand, and two MS degrees (in risk control management and management technology) from University of Wisconsin - Stout, and PhD in management science and information systems from Oklahoma State University. His research has recently been published in major journals such as Decision Support Systems, Decision Sciences, European Journal of Operational Research, and Journal of Business Ethics. His major research interests include decision support systems, RFID, manufacturing management, data mining, and supply chain management.