Social Science and Humanities Journal
Everant Journal
https://sshjournal.com/
Impact Factor: 2024: 6.576
2023: 5.731
ISSN: 2456-2653
Volume 09 Issue 01 January 2025
DOI: https://doi.org/10.18535/sshj.v9i01.1459

Energy Consumption Prediction in the Steel Industry Using Principal Component Analysis and Regression Tree Methods

Mery Septiani[1, a *] | Genrawan Hoendarto[2,b] | Jimmy Tjen[3,c]

[1,2,3]Widya Dharma Pontianak University, Pontianak
[a]21421467_mery_s@widyadharma.ac.id, [b]genrawan@widyadharma.ac.id, [c]jimmytjen@widyadharma.ac.id

***Received* 09-12-2024**
***Revised* 10-12-2024**
***Accepted* 31-12-2024**
***Published* 03-01-2025**

Abstract:

The steel industry is a major contributor to global energy consumption and greenhouse gas emissions, driving the need for enhanced efficiency. This study explore the use of Principal Component Analysis (PCA) combined with Regression tree (RT) techniques to predict energy consumption in the steel industry. PCA reduces data dimensionality, while RT addresses complex, non-linear relationships. Tested on a Kaggle dataset, the PCA and RT model achieved high accuracy, with a Root Mean Square Error (RMSE) of 0.67 and an accuracy rate of 90.82%, outperforming other methods in a comparative analysis. The model's moderate training time of 1.18 seconds highlights its efficiency. Visual and comparative analysis confirmed the model's strong alignment with observed energy consumption values and its balance between accuracy and computational efficiency. The PCA and RT model is an effective tool for predicting energy consumption in the steel industry, offering a practical approach to improving energy efficiency and sustainability. Future research could explore advanced techniques to further enhance predictive accuracy and model robustness.

Keywords: Principal Component Analysis, Regression Tree, Prediction, Energy Consumption, Steel Industry.

Introduction:

The steel industry is one of the most energy-intensive sectors, contributing significantly to global energy consumption and greenhouse gas emissions. As the demand for steel continues to rise, particularly in emerging economies, there is a pressing need to enhance energy efficiency within this sector [1, 2]. The potential for energy conservation and emission reduction in the steel industry is substantial, yet realizing this potential requires the adoption of advanced analytical methods and technologies [3].

One promising approach to improving energy efficiency in the steel industry is the integration of PCA and RT. PCA is widely recognized for its ability to reduce dimensionality in datasets, allowing for more efficient data processing and interpretation [4]. When combined with RT, which is effective in handling complex, non-linear relationships in data, this hybrid approach can significantly enhance the accuracy of energy consumption predictions [5].

The focus on energy efficiency in the steel industry it not new. Historical analyses have emphasized the role of advanced technologies and process optimizations in reducing energy use [6]. For instance, the development of energy-efficient technologies in the U.S. iron and steel industry has been a significant area of research, highlighting the need for continuous innovation [7]. Similarly, studies on the energy and environmental performance of the steel industry in Sweden and China have underscored the importance of learning and adaptation in supporting emergent technologies [8, 9].

In recent years, life cycle assessments (LCA) and environmental analyses have become essential tools for evaluating the sustainability of steel production. These assessments provide critical insights into the energy and carbon footprints of different steelmaking scenarios, offering guidance for future improvements [10, 11]. Moreover, the integration of energy efficiency strategies with life cycle thinking is crucial for achieving long-term sustainability in the steel industry [12].

The significance of energy efficiency in steelmaking is also reflected in the efforts of global organizations such as the World Steel Association, which has highlighted the need for continuous improvement in energy use across the industry [13]. Recent advancements in energy efficiency, as detailed in white papers and technical reports, further demonstrate the industry's commitment to reducing its environmental impact [14].

Method

This section outlines the algorithms and research workflow employed to develop the energy consumption prediction model in the steel industry using the Principal Component Analysis (PCA) and Regression Tree (RT) methods. Please refer to [15] for the fundamentals of PCA and RT.

A. Energy Consumption Prediction Algorithm Using PCA and RT

Let $\boldsymbol{H} = [\boldsymbol{y}, \boldsymbol{X}]; \boldsymbol{H} \in \mathbb{R}^{m \times (n_1+1)}$ represent the dataset related to energy consumption in the steel industry, where $\boldsymbol{y} \in \mathbb{R}^{m}$ denotes energy consumption, and $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{n_1}]; \boldsymbol{X} \in \mathbb{R}^{m \times n_1}$ represents a set of numerical parameters affecting energy consumption (such as temperature, pressure, and other factors). Here, m is the number of samples in the dataset, and n_1 is the number of numerical parameters.

Step 1 : PCA transformation

First, the raw data $\boldsymbol{X}$ is transformed using PCA to reduce the dataset's dimensionality and capture the maximum variability in the data. The result of this transformation is a new set of orthogonal features, know as principal components. The PCA transformation can be formulated as Eq. (1):

$$\boldsymbol{X}' = \boldsymbol{X} . \boldsymbol{W} \tag{1}$$

Where $\boldsymbol{W}$ is the eigenvector matrix associated with the largest eigenvalues of the covariance matrix $\boldsymbol{X}$. The covariance matrix is computed as Eq. (2):

$$\boldsymbol{C} = \frac{1}{m-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X} \tag{2}$$

And $\boldsymbol{W}$ is obtained from the eigenvalue decomposition of $\boldsymbol{C}$ [15].

Step 2 : Building the Model with Regression Tree

After transformation, the next step is to build the prediction model using the Regression Tree (TR). In this step, the dimensionally reduced data, $\boldsymbol{X}'$ is used to train the RT model, which is then used to predict the energy consumption values $\boldsymbol{y}$.

The RT model is designed to find non-linear relationships between the transformed parameters and the predicted energy consumption. The RT prediction function can be expressed as Eq. (3):

$$\hat{y} = F_{RT}(\boldsymbol{X}') \tag{3}$$

Where F_{RT} represents the Regression Tree model function [15].

Step 3 : Model Evaluation

To evaluate the model's performance, the dataset is divided into two subsets: training data and testing data. The model is trained using the training data, and its accuracy is tested using the testing data. The evaluation methods employed include calculating the Root Mean Square Error (RMSE), Normalized RMSE (NRMSE), and accuracy to measure how well the model predicts energy consumption. The formulas for RMSE Eq. (4), NRMSE Eq. (5), and accuracy Eq. (6) are as follows:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \qquad (4)$$

$$NRMSE = \frac{RMSE}{range(y)} \qquad (5)$$

$$Accuracy = (1 - NRMSE) * 100\% \qquad (6)$$

Where y_i is the actual value, $\hat{y}_i$ is the predicted value, and m is the number of sample [15].

Step 4 : Determining the Best Model

The model with the smallest NRMSE and highest accuracy is considered the best model is then used to predict on datasets that model has not seen during training.

B. Data Source

The dataset used in this research is sourced from the publicly available Kaggle dataset titled "Steel Industry Energy Consumption" [16]. This dataset includes various parameters relevant to energy consumption in steel industry, such as temperature, pressure, and other parameters used in PCA and RT analysis.

Result and Discussion:

In this section, we present and discuss the results of the energy consumption prediction using the Principal Component Analysis (PCA) combined with Regression Tree (RT) to model energy consumption in the steel industry. The PCA and RT model was trained and tested on the preprocessed dataset, and its predictions were closely aligned with the observed energy consumption values.

A. Model Performance

The Regression Tree model was trained on 80% of the dataset, while the remaining 20% was used for testing. MATLAB was utilized as the software platform to implement the PCA and Regression Tree models for energy consumption prediction. The model's performance was evaluated using Root Mean Square Error (RMSE), Normalized RMSE (NRMSE), and accuracy metrics. The training time was recorded at 1.18 seconds, indicating the model's efficiency in handling the dataset. The results showed an RMSE of 0.67, which signifies a low prediction error. The NRMSE was calculated at 0.09, corresponding to an accuracy of 90.82%.

These result demonstrate that the PCA and RT model is highly accurate in predicting energy consumption within the steel industry, as indicated by the high accuracy percentage and low RMSE values. The use of PCA in the preprocessing stage effectively reduced dimensionality and enhanced the predictive capability of the Regression Tree.

A. Visual Analysis

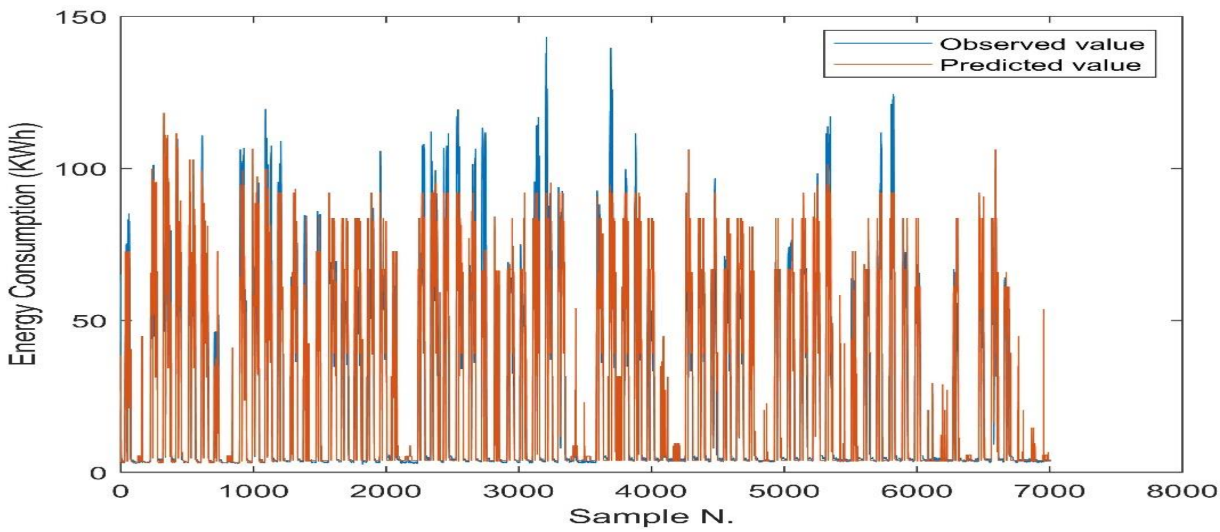

Figure 1 Comparison of Observed and Predicted Energy Consumption in the Steel Industry

The predicted values were plotted against the observed values to visually assess the model's performance. The plot, as shown in Figure 1, reveals a close alignment between the predicted and observed energy consumption values, further validating the model's accuracy. Minor deviations observed can be attributed to inherent noise in the data or limitations of the Regression Tree in capturing complex nonlinear relationships.

B. Comparison with Previous Studies

The model's accuracy of 90.82% is competitive when compared to other studies employing similar techniques in energy consumption prediction. The use of Regression Trees, combined with PCA, has been shown to improve prediction accuracy by addressing multicollinearity issues and enhancing the model's interpretability.

Table 1. Comparison of Different Methods for Energy Consumption Prediction

Method	Training Time	RMSE	NRMSE	Accuracy
PCA + RT	1.18	0.67	0.09	90.82
Regression Tree	0.24	2.55	0.1	89.88
RT + AR	0.23	4.71	0.19	81.31
Gradient Boosting	0.83	23.97	0.95	4.98
GB + AR	1.06	10.45	0.41	58.58
Random Forest	0.68	22.01	0.87	12.76
RF + AR	2.43	3.55	0.14	85.94

The comparison in Table 1 highlights that the PCA and RT method strikes an optimal balance between accuracy and training time compared to other methods. While methods like Gradient Boosting and Random Forest offer comparable accuracy, they come at the cost of significantly higher RMSE and training times. The PCA and RT approach, with its moderate training time and high accuracy, emerges as a reliable and efficient method for energy consumption prediction in the steel industry.

Conclusion:

This study investigated the use of Principal Component Analysis (PCA) combined with Regression Tree (RT) for predicting energy consumption in the steel industry. The PCA and RT model demonstrated high accuracy, with an RMSE of 0.67 and an accuracy rate of 90.82%, indicating its effectiveness in capturing the underlying patterns in energy usage data. The integration of PCA allowed for dimensionality reduction, which enhanced the model's predictive performance by mitigating multicollinearity issues.

The comparative analysis with other methods, including standalone Regression Trees, Random Forest, and Gradient Boosting, revealed that the PCA and RT model offer a favorable balance between computational efficiency and predictive accuracy. While some methods showed slightly higher accuracy, they required significantly longer training times, making PCA and RT a practical choice for real-time applications in industrial settings.

Overall, the PCA and RT model provides a robust and efficient tool for energy consumption prediction in the steel industry. Future research could explore the integration of more advanced ensemble techniques or alternative features selection methods to further enhance predictive accuracy and model robustness. The findings of this study contribute to the ongoing efforts to optimize energy management in industrial environments, offering a viable approach for improving energy efficiency and sustainability.

References:

[1] Y. Wang, X. He, and F. Jiang, "The energy conservation and emission reduction potentials in China's iron and steel industry: Considering the uncertainty factor," *J. Clean. Prod.*, vol. 413, p. 137519, Aug. 2023, doi: 10.1016/J.JCLEPRO.2023.137519.

[2] J. Kim *et al.*, "Decarbonizing the iron and steel industry: A systematic review of sociotechnical systems, technological innovations, and policy options," *Energy Res. Soc. Sci.*, vol. 89, Jul. 2022, doi: 10.1016/j.erss.2022.102565.

[3] Y. M. Wei, H. Liao, and Y. Fan, "An empirical analysis of energy efficiency in China's iron and steel sector," *Energy*, vol. 32, no. 12, pp. 2262–2270, Dec. 2007, doi: 10.1016/J.ENERGY.2007.07.007.

[4] I. T. Jollife and J. Cadima, "Principal component analysis: A review and recent developments," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0202.

[5] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification A nd R egression Trees Leo*. 2017.

[6] E. Worrell, "Advanced technologies and energy efficiency in the iron and steel industry in China," *Energy Sustain. Dev.*, vol. 2, no. 4, pp. 27–40, Nov. 1995, doi: 10.1016/S0973-0826(08)60148-9.

[7] N. Margolis and L. Sousa, "Energy and environmental profile of the U.S. iron and steel industry," *Proc. ACEEE Summer Study Energy Effic. Ind.*, no. August, pp. 103–113, 1997.

[8] B. Kiss and L. Neij, "The importance of learning when supporting emergent technologies for energy efficiency—A case study on policy intervention for learning for the development of energy efficient windows in Sweden," *Energy Policy*, vol. 39, no. 10, pp. 6514–6524, Oct. 2011, doi: 10.1016/J.ENPOL.2011.07.053.

[9] T. T. X. Hasanbeigi, Ali, William R Morrow III, Jayant A Sathaye, Eric R Masanet, "Assessment of Energy Efficiency Improvement and CO2 Emission Reduction Potentials in the Iron and Steel Industry in China," *Lawrence Berkeley Natl. Lab.*, 2012, [Online]. Available: https://energyanalysis.lbl.gov/publications/assessment-energy-efficiency-2

[10] J. Suer, M. Traverso, and N. Jäger, "Review of Life Cycle Assessments for Steel and Environmental Analysis of Future Steel Production Scenarios," *Sustain.*, vol. 14, no. 21, 2022, doi: 10.3390/su142114131.

[11] H. W. Kua, "Integrated policies to promote sustainable use of steel slag for construction—A consequential life cycle embodied energy and greenhouse gas emission perspective," *Energy Build.*, vol. 101, pp. 133–143, Aug. 2015, doi: 10.1016/J.ENBUILD.2015.04.036.

[12] S. Siitonen, M. Tuomaala, and P. Ahtila, "Variables affecting energy efficiency and CO2 emissions in the steel industry," *Energy Policy*, vol. 38, no. 5, pp. 2477–2485, May 2010, doi: 10.1016/J.ENPOL.2009.12.042.

[13] World Steel Association (WSA), "Fact Sheet: Energy use in the steel industry," *Worldsteel Assoc. Fact Sheet*, no. Lci, pp. 1–3, 2013.

[14] J. De Beer, E. Worrell, and K. Blok, "Future technologies for energy-efficient iron and steel making," *Annu. Rev. Energy Environ.*, vol. 23, no. 1, pp. 123–205, 1998, doi: 10.1146/annurev.energy.23.1.123.

[15] F. Smarra, J. Tjen, and A. D'Innocenzo, "Learning methods for structural damage detection via entropy-based sensors selection," *Int. J. Robust Nonlinear Control*, vol. 32, no. 10, pp. 6035–6067, 2022, doi: 10.1002/rnc.6124.

[16] Kaggle, "Steel Industry Energy Consumption Prediction." [Online]. Available: https://www.kaggle.com/datasets/joebeachcapital/steel-industry-energy-consumption