CSE3505 – Foundations of Data Analytics

J Component - Project Report

Review III

Energy Optimization for Sustainable Manufacturing

By

21MIA1068 – Kamesh E
21MIA1082 – Sonakshi Dinesh
21MIA1134 – Tharun Raj N

M.Tech CSE Integrated with Business Analytics

Submitted to

Dr.Asnath Victy Phamila Y,
Assistant Professor Senior,
SCOPE, VIT, Chennai

School of Computer Science and Engineering

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

November 2024

ABSTRACT

Sustainable manufacturing depends on effective energy management because it reduces expenses and environmental effects. The goal of this project is to employ prescriptive analytics and machine learning to forecast and optimize manufacturing energy use. In order to create predictive models that direct energy conservation, the study analyses time-series data on variables such as reactive power and CO_2 emissions to find patterns in energy utilization. Three machine learning models—Random Forest, Support Vector Machine (SVM), and Linear Regression—were used; Random Forest was the most accurate and thus best suited for this use case. In addition to forecasts, the study uses optimization methods, such as fuzzy logic, stochastic programming, and linear programming, to offer practical suggestions. When precise data is not available, fuzzy logic handles imperfect data inputs for flexible optimization, while stochastic programming deals with price and demand uncertainties to allow for flexible decision-making. When combined, these prescriptive models improve optimization by providing methods for lowering energy usage while taking imprecision and variability into consideration. By combining machine learning with cutting-edge optimization techniques to increase energy efficiency, this project supports the transition toward more responsible and sustainable industrial practices and advances sustainable manufacturing, which is in line with Sustainable Development Goal (SDG) 12 for responsible consumption and production.

INTRODUCTION

Energy optimization has emerged as a crucial element of sustainable production in today's industrial environment. Manufacturers are concentrating on effective energy management as a result of the increasing requirement to lower environmental effects and limit operating expenses. While prescriptive analytics offers practical suggestions for reducing usage, predictive analytics, which is driven by machine learning, allows businesses to forecast energy consumption based on historical data. This research integrates various strategies, employing optimization techniques like Linear Programming, Stochastic Programming, and Fuzzy Logic in conjunction with time-series data on energy usage to create predictive models like Linear Regression, SVM, and Random Forest. These methods establish a thorough framework for enhancing manufacturing energy efficiency by controlling inaccurate data inputs and addressing uncertainties. The project supports the shift to sustainable industrial practices and promotes responsible production and consumption through optimal energy use, which is in line with Sustainable Development Goal (SDG) 12.

Objectives:

1) **Descriptive Analytics**: Descriptive analytics involves preprocessing and cleaning the dataset to ensure accuracy and consistency, followed by exploratory data analysis (EDA) to summarise historical trends in energy consumption. Visualisations like histograms and box plots help understand patterns in usage, such as differences between weekday and weekend energy loads.

2) **Diagnostic Analytics**: Diagnostic analytics focuses on identifying the factors influencing energy consumption. Through correlation analysis, relationships between variables like reactive power and CO_2 emissions are explored to explain the key drivers behind consumption trends, providing insights into why these patterns exist.

3) **Predictive Analytics**: Predictive analytics use machine learning models such as Linear Regression, SVM, and Random Forest to forecast future energy consumption based on historical data. These models are evaluated using performance metrics like R^2, MAE, and RMSE to determine how accurately they predict energy usage.

4) **Prescriptive Analytics**: Prescriptive analytics uses optimisation techniques, such as linear programming, to provide actionable recommendations for reducing energy consumption. By leveraging model coefficients, it suggests strategies to minimise reactive power, ultimately helping to optimise overall energy usage based on predictive insights.

RELATED WORKS

Paper 1 :

Title : Resource saving by optimisation and machining environments for sustainable manufacturing

Methodology and Description : The study highlights sustainable machining through energy-efficient practices, reducing power consumption, and using environmentally friendly cooling methods like MQL, nanofluids, and cryogenic lubrication. Optimization techniques, including Taguchi and Grey Relational Analysis, enhance productivity while minimizing resource use. Emphasis is placed on reducing emissions, waste, and costs during machining processes. Challenges include balancing efficiency, cost, and environmental impact.

Paper 2 :

Title : A review on methods of energy performance improvement towards sustainable manufacturing from perspectives of energy monitoring, evaluation, optimisation and benchmarking

Methodology and Description : The study develops a framework for improving energy performance in manufacturing, focusing on four aspects: monitoring, evaluation, optimization, and benchmarking. It emphasizes using advanced technologies like IoT, big data, and digital tools to enhance energy efficiency, reduce CO_2 emissions, and achieve sustainability. A review of 166 research papers highlights methods and challenges, providing actionable strategies for energy-saving in production processes.

Paper 3 :

Title : Sustainability and Optimisation of Green and Lean Manufacturing Processes Using Machine Learning Techniques

Methodology and Description : The study explores the integration of sustainability, green manufacturing, and lean manufacturing with machine learning techniques to optimize energy use, reduce waste, and enhance efficiency. Case studies demonstrate applications like predictive maintenance, energy optimization, and circular economy practices. Challenges include data quality, implementation complexity, and ethical concerns, with trends like explainable AI and human-machine interaction offering solutions. The methodology combines case studies and literature reviews to provide actionable insights for researchers and practitioners.

Paper 4 :

Title : Predictive maintenance planning for industry 4.0 using machine learning for sustainable manufacturing

Methodology and Description : The study proposes an intelligent Predictive Maintenance (PdM) planning model for sustainable manufacturing, using machine learning techniques like SVM and RNN enhanced with the Jaya-based Sea Lion Optimization (J-SLnO) algorithm. Key steps include data cleaning, normalization, optimal feature selection, and prediction for forecasting equipment failures. The model demonstrated superior performance, achieving up to 95% better RMSE than traditional methods on datasets like aircraft engines and lithium-ion batteries. This methodology ensures improved resource efficiency, reduced failures, and enhanced sustainability in manufacturing

Paper 5 :

Title : Developing a hybrid evaluation approach for the low carbon performance on sustainable manufacturing environment

Methodology and Description : The study examines the positive impact of green sustainable practices in procurement, logistics, product design, and regulatory frameworks on low-carbon performance, driving sustainable manufacturing and societies. Using a hybrid methodology, it tests a theoretical model through PLS-SEM on 380 responses, validates it with machine learning, and further refines it using Item Response Theory. Results confirm the central role of carbon performance in linking green practices to societal sustainability.

METHODOLOGY

Data Collection and Preprocessing:
The data for this project comprises time-series information on energy consumption, including key variables such as reactive power, CO_2 emissions, and factors influencing energy usage patterns. To prepare this dataset for analysis, various preprocessing steps are undertaken:

- Cleaning: The dataset is checked for and addressed missing values, which may arise from sensor failures or intermittent data collection. Standard techniques such as imputation are applied to ensure the data is complete.

- Encoding Categorical Variables: If there are categorical variables in the dataset (e.g., day of the week or shift type), they are converted into numerical form using encoding techniques like one-hot encoding to ensure compatibility with machine learning models.

- Normalization/Standardization: Continuous variables are normalized or standardized to reduce the variance and improve model performance, especially for algorithms sensitive to feature scales.

Exploratory Data Analysis:
EDA is conducted to uncover initial insights and assess patterns in energy consumption. Various visualization techniques are applied to identify trends and patterns, such as:

- Histograms and Box Plots: These visualizations highlight distributions and potential outliers in energy consumption data, allowing for a deeper understanding of typical versus atypical energy use.

- Time-Series Plots: Plotting energy consumption over time reveals cyclical trends, such as weekday versus weekend usage patterns or seasonal variations.

- Correlation Analysis: This step examines relationships between variables, such as reactive power and CO_2 emissions, to identify the primary drivers of energy consumption. These insights guide the model selection process by highlighting significant predictors.

Model Training and Evaluation:

1. **Descriptive Analytics**:
   Utilizes data preprocessing and exploratory data analysis (EDA) to identify patterns and trends in historical energy consumption. Visual tools such as box plots and histograms offer clarity on usage behaviours.

2. **Diagnostic Analytics**:
   Focuses on understanding the causal factors influencing energy consumption. Correlation analysis helps identify relationships between variables such as reactive power and $CO2$ emissions, explaining key drivers behind energy trends.

3. **Predictive Models**:
   - Linear Regression: As one of the simplest models, Linear Regression is chosen for its interpretability and ability to model linear relationships between predictors (such as reactive power or CO_2 emissions) and the target variable (energy consumption). It provides a baseline for comparing more complex models.

   - Support Vector Machine (SVM): SVM is selected for its flexibility in handling complex relationships and is particularly useful when there may be non-linear dependencies in the data. With an appropriate kernel function, SVM can model these interactions, providing insights into energy consumption trends with moderate accuracy.

   - Random Forest: An ensemble learning method, Random Forest combines predictions from multiple decision trees, yielding high accuracy and robustness against overfitting. This model is particularly suited for capturing non-linear relationships and complex interactions, making it ideal for energy consumption data that may contain various types of predictors. Each model is evaluated using standard metrics: R^2 to measure fit, Mean Absolute Error (MAE) for average deviation from actual values, and Root Mean Squared Error (RMSE) for overall prediction accuracy.

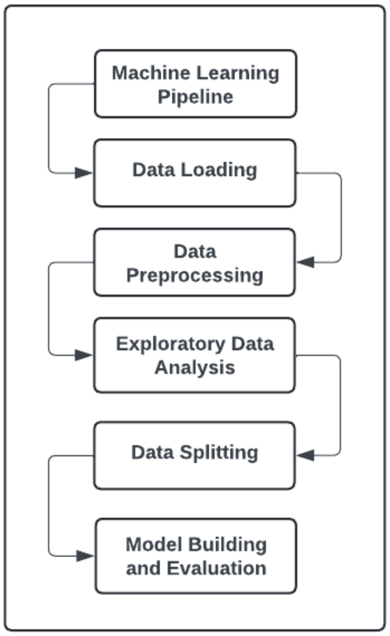
4. **Prescriptive Models**: To optimize energy consumption, the project applies three prescriptive models, each offering unique strengths for different scenarios:

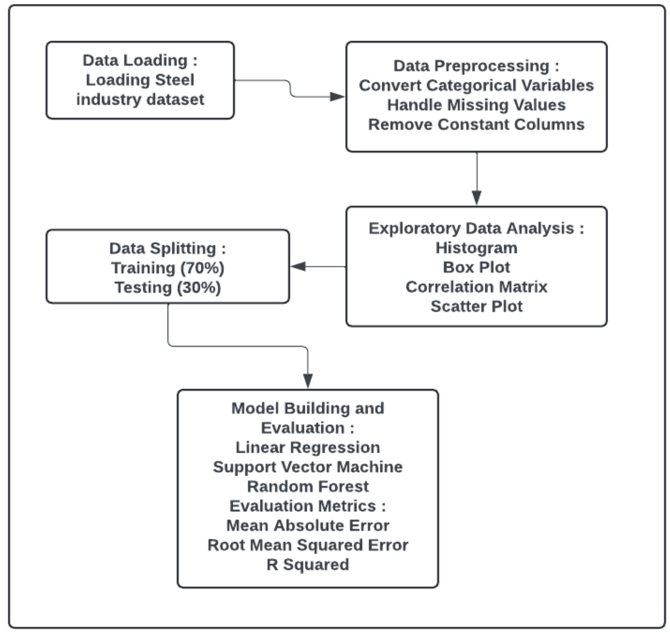
   - Linear Programming (LP): LP is used to minimize energy consumption within defined constraints, such as maintaining certain production levels or adhering to budget limits. This optimization model is deterministic, meaning it operates under the assumption that inputs are certain. LP helps establish a baseline for energy minimization and suggests specific adjustments to consumption patterns within fixed parameters.

   - Stochastic Programming: Recognizing that energy costs and production demands can fluctuate, Stochastic Programming introduces adaptability to the optimization process. It models these uncertainties by creating multiple "scenarios" with different probability distributions, allowing the system to adjust its energy usage based on potential future conditions. This approach makes it possible to optimize energy consumption dynamically, accommodating variations in real-world conditions.

   - Fuzzy Logic: Unlike traditional models that require precise data, Fuzzy Logic is designed to handle imprecise or ambiguous inputs. By creating if-then rules based on degrees of membership rather than exact values, fuzzy logic provides a flexible decision-making framework under variable conditions. For instance, it can optimize energy usage based on a range of CO_2 emission levels rather than a fixed amount, enabling it to adapt to fluctuating or uncertain data.

ARCHITECTURE DIAGRAM





The diagram illustrates a machine learning pipeline used in the steel industry for predictive maintenance. Data loading is the first step in the process, during which steel industry datasets are imported. The next step is data preprocessing, which includes actions like converting categorical variables, eliminating constant columns, and addressing missing values in order to guarantee data quality. The next step is exploratory data analysis (EDA), which looks for patterns and relationships in the data using tools including correlation matrices, box plots, scatter plots, and histograms. For model building, the data is subsequently divided into training (70%) and testing (30%) groups. Lastly, techniques like Random Forest, Linear Regression, and Support Vector Machine are used to develop and assess the model. Metrics like Mean Absolute Error, Root Mean Squared Error, and R-Squared are used to gauge the prediction performance of the models.

MODULE DESCRIPTION

This project integrates a diverse set of machine learning models and optimization techniques to comprehensively forecast and optimize energy consumption in sustainable manufacturing. By leveraging Linear Regression, Support Vector Machines (SVM), and Random Forest, it accurately predicts future energy usage, with each model addressing different complexities - Linear Regression offers simplicity and interpretability, SVM handles non-linear relationships, and Random Forest excels in managing variable interactions and diverse data types. Beyond prediction, prescriptive analytics is employed to optimize energy consumption. Linear Programming provides a deterministic framework for minimizing reactive power, while Stochastic Programming accounts for uncertainties such as fluctuating energy demand and supply, ensuring robust decision-making across various scenarios. Additionally, Fuzzy Logic enhances the system's adaptability by handling imprecise data and formulating flexible, rule-based optimization strategies. This integrated framework balances predictive accuracy with practical, actionable insights, enabling efficient energy management and supporting sustainable manufacturing practices.

1. Linear Regression: Linear regression serves as the foundational predictive model in this project. It is used to establish a baseline understanding of how key variables, such as reactive power and CO2 emissions, impact energy consumption. The model assumes a linear relationship between the dependent variable (energy consumption) and one or more independent variables. Linear regression provides interpretable coefficients, which quantify the direct effect of each predictor on energy usage. This model is particularly useful for stakeholders to understand basic consumption trends and the relative importance of different variables. However, its simplicity may limit its effectiveness in capturing more complex, non-linear patterns present in the data.

2. Support Vector Machines (SVM): Support Vector Machines are used for Support Vector Regression (SVR) in this project to capture more complex relationships between variables. Unlike linear regression, SVM can handle both linear and non-linear dependencies by utilizing kernel functions, which map the input data into higher-dimensional spaces where a linear relationship may exist. SVM focuses on a subset of data points (support vectors) that are most informative for defining the model. This makes it robust to outliers and well-suited for predicting energy consumption in scenarios where data points deviate significantly from typical patterns. SVM provides flexibility in modelling and improves prediction accuracy over simpler models in cases of non-linearity.

3. Random Forest: Random Forest is a powerful ensemble learning method employed to enhance the predictive accuracy of energy consumption forecasts. It constructs multiple decision trees during training and combines their outputs for final prediction. Each tree is built using a random subset of the data, reducing variance and preventing overfitting. Random Forest is particularly effective in handling complex, non-linear relationships and interactions between predictors. It also provides feature importance metrics, which help identify the most significant variables influencing energy consumption, such as reactive power and CO2 emissions. In this project, Random Forest demonstrated superior predictive performance, with high accuracy and low error metrics, making it an ideal choice for energy forecasting.

4. Linear Programming: Linear programming is used in the prescriptive analytics phase to optimize energy consumption. This mathematical optimization technique aims to minimize reactive power or energy costs while adhering to a set of linear constraints derived from predictive insights. For example, constraints could include operational limits on total energy usage or specific requirements for maintaining power factors. Linear programming provides an optimal solution under deterministic conditions, offering actionable strategies for energy efficiency. However, its effectiveness is contingent on the accuracy and reliability of the input data, which is assumed to be certain and unchanging.

5. Stochastic Programming: To address the uncertainties inherent in energy demand and supply, stochastic programming is integrated into the optimization process. Unlike linear programming, stochastic programming considers multiple possible scenarios, each with associated probabilities. These scenarios might include variations in energy demand during peak hours or fluctuations in renewable energy availability. By optimizing the expected value of the objective function across all scenarios, stochastic programming delivers robust solutions that perform well under different conditions. This ensures that the optimization strategies are resilient to real-world variability, making them more practical and reliable for dynamic manufacturing environments.

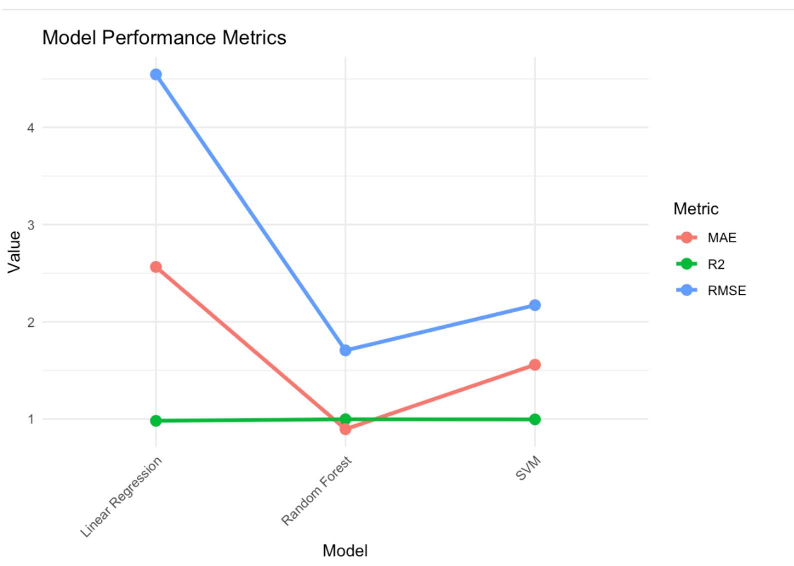
6. Fuzzy Logic: Fuzzy logic adds an adaptive layer to prescriptive analytics by dealing with vagueness and imprecision in the input data. Real-world variables like "high" reactive power or "moderate" CO_2 emissions do not always have crisp boundaries. Fuzzy logic uses membership functions to define these variables and allows for flexible rule-based decision-making. For example, a fuzzy rule might suggest reducing energy consumption when reactive power is "high" and CO_2 emissions are "moderate," with the exact level of reduction varying based on the degree of membership in these fuzzy sets. This adaptability makes fuzzy logic an excellent tool for formulating optimization strategies in environments where precise data is not always available.

RESULTS AND DISCUSSIONS

The implementation of various machine learning models and optimization techniques yielded significant insights into energy consumption patterns and optimization strategies. Among the predictive models, Random Forest emerged as the most accurate, achieving an R^2 of 0.997, Mean Absolute Error (MAE) of 0.895, and Root Mean Squared Error (RMSE) of 1.706. These metrics indicate that Random Forest can effectively capture complex, non-linear relationships and provide precise forecasts of energy consumption. In comparison, Support Vector Machines (SVM) also demonstrated good performance but with slightly higher MAE and RMSE, reflecting a marginally reduced predictive accuracy. Linear Regression, while providing a simpler and more interpretable model, exhibited lower accuracy with larger deviations from actual consumption values, underscoring its limitations in capturing non-linear dependencies.

In the prescriptive analytics phase, Linear Programming successfully optimized energy consumption by minimizing reactive power under deterministic conditions. However, the dynamic nature of manufacturing environments necessitated the use of Stochastic Programming, which incorporated uncertainty in energy demand and supply. This approach provided robust optimization strategies, ensuring efficiency across multiple probable scenarios. Additionally, the integration of Fuzzy Logic allowed for flexible decision-making in situations involving imprecise data. By applying fuzzy rules, the system could adapt to varying operational conditions, further enhancing the effectiveness of the optimization process.

In conclusion, the combination of advanced predictive and prescriptive methods not only improved forecasting accuracy but also enabled the formulation of robust and adaptive strategies for energy management. This integrated approach supports sustainable manufacturing by optimizing energy use, reducing costs, and minimizing environmental impact.



SUSTAINABLE DEVELOPMENT GOAL

Contribution of Results to SDG 12: Responsible Consumption and Production:

The results of this project significantly contribute to Sustainable Development Goal 12 (SDG 12) by optimizing energy consumption and promoting sustainable manufacturing practices. The integration of advanced predictive models (Linear Regression, SVM, and Random Forest) and prescriptive analytics (Linear Programming, Stochastic Programming, and Fuzzy Logic) has provided actionable insights into energy usage, enabling more efficient resource utilization. Here's how the results align with and support the targets of SDG 12:

1. Efficient Use of Resources: The project's optimization strategies directly address the efficient use of natural resources by minimizing reactive power and optimizing overall energy consumption. The Random Forest model's high accuracy ensures precise forecasting, which helps industries plan and allocate resources more effectively, reducing unnecessary energy use.
2. Waste Reduction: By identifying patterns in energy consumption and optimizing operations, the project minimizes energy waste. The prescriptive analytics framework ensures that energy is consumed efficiently, even under fluctuating demand and supply conditions, thereby reducing the carbon footprint and resource wastage.
3. Sustainable Production Practices: The project promotes sustainable production by integrating machine learning and optimization techniques that allow industries to meet their energy needs with minimal environmental impact. This aligns with SDG 12's focus on decoupling economic growth from resource use and environmental degradation.
4. Data-Driven Decision Making: The use of predictive analytics enables manufacturers to base their decisions on data-driven insights rather than estimates or assumptions. This precision supports the adoption of cleaner production techniques and fosters innovation in sustainable practices.
5. Sustainability in Supply Chains: The forecasting and optimization tools developed in the project can be extended to manage energy consumption across supply chains, ensuring sustainability at every stage of production. This aligns with SDG 12's goal of integrating sustainability into corporate practices.
6. Reduction of Greenhouse Gas Emissions**:** The project's focus on reducing reactive power indirectly contributes to lowering CO2 emissions. By optimizing energy use, industries can reduce their dependence on fossil fuels and improve their overall energy efficiency, thus supporting climate action under SDG 12 and SDG 13 (Climate Action).
7. Support for Long-Term Sustainability Goals: The insights gained from this project provide a foundation for continuous improvement in energy management, supporting long-term sustainability goals. The robust nature of stochastic programming ensures that industries can maintain optimal energy usage even in uncertain and dynamic environments.

CONCLUSION

The goal is completed by successfully implementing multiple machine learning models to predict energy consumption. Among the models, Random Forest demonstrated superior performance, establishing it as the best choice for this task. The project also highlights how advanced optimization techniques can reduce energy usage effectively based on predictive insights.

- Key factors influencing energy consumption include reactive power, CO_2 emissions, and the day of the week.
- Optimization through prescriptive analytics, now enhanced with stochastic programming and fuzzy logic, proves instrumental in minimizing reactive power and improving overall energy efficiency.
- Predictive models, particularly ensemble methods like Random Forest, are highly effective in forecasting complex time-series energy usage trends, making them invaluable for sustainable energy management.

REFERENCES

1. Pimenov, D. Y., Mia, M., Gupta, M. K., Machado, Á. R., Pintaude, G., Unune, D. R., ... & Kuntoğlu, M. (2022). Resource saving by optimisation and machining environments for sustainable manufacturing: A review and future prospects. Renewable and Sustainable Energy Reviews, 166, 112660.

2. Cai, W., Wang, L., Li, L., Xie, J., Jia, S., Zhang, X., ... & Lai, K. H. (2022). A review on methods of energy performance improvement towards sustainable manufacturing from perspectives of energy monitoring, evaluation, optimisation and benchmarking. Renewable and Sustainable Energy Reviews, 159, 112227.

3. Ingle, R. B., Swathi, S., Mahendran, G., Senthil, T. S., Muralidharan, N., & Boopathi, S. (2023). Sustainability and Optimisation of Green and Lean Manufacturing Processes Using Machine Learning Techniques. In Circular Economy Implementation for Sustainability in the Built Environment (pp. 261-285). IGI Global.

4. Abidi, M. H., Mohammed, M. K., & Alkhalefah, H. (2022). Predictive maintenance planning for industry 4.0 using machine learning for sustainable manufacturing. Sustainability, 14(6), 3387.

5. Ali, S. S., Kaur, R., Persis, D. J., Saha, R., Pattusamy, M., & Sreedharan, V. R. (2023). Developing a hybrid evaluation approach for the low carbon performance on sustainable manufacturing environment. Annals of Operations Research, 1-33.

DATA SOURCE

https://www.kaggle.com/datasets/csafrit2/steel-industry-energy-consumption