CS 5593 – Online Sections 995-999 - Fall 2025 - Dr. Le Gruenwald
PROJECT AND PROPOSAL

Project Proposal: Energy Consumption Forecasting & Optimization

1) Project Title

Energy Consumption Forecasting and Optimization in Steel Manufacturing using Data Mining Approaches

2) Names and Email Addresses of Authors

- Anvitha Reddy Thummalapally – anvitha.reddy.thummalapally-1@ou.edu
- Lakshmi Sahasra Jangoan – lakshmi.sahasra.jangoan-1@ou.edu
- Soumya Sri Kesani – soumya.sri.kesani-1@ou.edu

3) Category and Objectives of the Project

Category: (a) – 3 algorithms for 3 different data mining tasks.

Overall Objective:

Develop an integrated data mining application that forecasts energy consumption, identifies operating modes and detects anomalous energy usage in steel manufacturing. Building a decision support tool that transforms raw energy data into forecasts, operating profiles, and anomaly alerts, so industries can run more efficiently and sustainably.

Individual Objectives:

- Implement a Regression model from scratch to forecast energy consumption.
- Implement a clustering algorithm from scratch to identify operating modes and typical load patterns.
- Implement an outlier detection algorithm from scratch to detect anomalous energy usage.
- Integrate the implemented algorithms into an application that supports decision making for energy optimization.

4) The Significance of the Project

4.1) Application:

The proposed application is an energy consumption forecasting and optimization for steel manufacturing plants. It integrates multiple data mining tasks into a single decision support too. Using operational data, the system will forecast energy consumption using regression models, identify typical operating modes and load patterns through clustering, detect anomalous energy consumption behaviors through outlier detection.

Significance:

Energy is one of the largest cost drivers in steel production. Small improvements in forecasting accuracy and early detection of abnormal energy draws can reduce cost and industrial carbon dioxide emissions. This application combines forecasting, clustering, anomaly detection to give operators a single tool to forecast energy demand which helps ensure that the plant can schedule operations effectively, avoid sudden spikes in energy usage and identify operating modes to have insights into how the plant behaves under different production conditions. The application supports cost savings, operational efficiency, and sustainably in the steel industry.

Data Mining Questions to be answered:

1. How can we predict future energy consumption based on current and historical operational variables?
2. What are the typical operating modes or load profiles of the plant, and how do they differ in terms of energy usage?
3. Are there instances of abnormal energy usage that deviate from expected patterns, and what operational factors contribute to them?

4.2) Why these questions require Data Mining:

Question 1: How can we predict future energy consumption based on current and historical operational variables?

Energy consumption in steel plants is influenced by many interacting factors like load type, time of the day, weekday/weekend status, reactive power and power factors. These relationships are non-linear and too complex to capture. In this situation, data mining can help in learning patterns from thousands of historical records.

Required Data Mining task: We can use regression algorithms to learn these patterns from historical data and predict energy future usage.

Question 2: What are the typical operating modes or load profiles of the plant, and how do they differ in terms of energy usage?

Operating modes are not explicitly labeled in the dataset except for the categorical load type. However, the plant may actually have sub-modes or hidden patterns within these categories. Like weekday medium load may behave differently than weekend medium load. These cannot be discovered manually without a systematic analysis.

Required Data Mining task: Clustering algorithms can group records into natural clusters based on attributes like reactive power, CO2, power factors. These clusters reveal typical operating modes that help managers understand energy profiles and optimize operations.

Question 3: Are there instances of abnormal energy usage that deviate from expected patterns, and what operational factors contribute to them?

Anomalies in energy usage like sudden spikes or drops may signal equipment faults, inefficiencies, or abnormal operations. These events are rare, hidden in large volumes of time-stampered data. Manual inspection of thousands of records is infeasible.

Required Data Mining task: Outlier detection methods can automatically flag unusual points and correlate them with other variables. This helps pinpoint when and why abnormal consumption occurs.

4.3) Dataset and Sources

Number of datasets: 1

Name of the dataset: Steel Industry Energy Consumption

Source of the Dataset: UC Irvine Repository

Dataset Description: 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.

Number of Records: 35040

Number of Attributes: 11

File size: ~2.73 MB

Attributes (Name, Type, Meaning, Size):

There are 11 attributes.

- date- date - Timestamp of the observation (DD/MM/YYYY)
- Usage_kWh - Continuous - Actual energy consumed during the interval
- Lagging_Current_Reactive.Power_kVarh - Continuous - Reactive power consumed (lagging)
- Leading_Current_Reactive_Power_kVarh - Continuous - Reactive power consumed (leading)
- CO2(tCO2_ - Continuous - Estimated CO2 emotions associated with energy usage (in tons)
- Lagging_Current_Power_Factor - Continuous - Ratio indicating efficiency of power usage (lagging)
- Leading_Current_Power_Factor - Continuous - Ratio indicating efficiency of power usage (leading)
- NSM - integer - Number of seconds from midnight
- WeekStatus - Categorical - Indicates if the record is from weekday or weekend
- Day_of_week - Categorical - Name of the day (Monday - Sunday)
- load_type - Categorical - Indicates whether a plant is in Light load, Medium load, or Heavy load mode.

Average Record Size: ~78 bytes per row (2,731,389 bytes / 35,040 records)

Class/Dependent Attribute:

- For the regression task, the dependent variable is Usage_kWh.
- For clustering, multiple numerical features like Usage_kWh, reactive power, power factor, CO2 will be used as input attributes.
- For outlier detection, the target is to identify abnormal points in Usage_kWh.

Sample Records:

date	Usage_kWh	Lagging_Current_Reactive.Power_kVarh	Leading_Current_Reactive_Power_kVarh	CO2(tCO2)	Lagging_Current_Power_Factor	Leading_Current_Power_Factor	NSM	WeekStatus	Day_of_week	Load_Type
01/01/2018 00:15	3.17	2.95	0	0	73.21	100	900	Weekday	Monday	Light_Load
01/01/2018 00:30	4	4.46	0	0	66.77	100	1800	Weekday	Monday	Light_Load
01/01/2018 00:45	3.24	3.28	0	0	70.28	100	2700	Weekday	Monday	Light_Load

5) Implementation/Research Methodology and Time Table

General Plan of work:

This project will be implemented in phases to ensure structured progress. Every phase corresponds to a major deliverable colon, dataset preparation, algorithm development, integration and evaluation.

All algorithms will be implemented from scratch in python using only Fundamental libraries. Visualization and integration will be done by using Matplotlib or Plotly.

Task No.	Description	Deliverables	Start Date	End Date	Responsible person
1	Dataset Collection & Preprocessing	Cleaned Dataset, Data dictionary, Exploratory plots	Sept 22, 2025	Oct 1, 2025	Anvitha
2	Literature Review & Algorithm Selection	Literature review summary, justification for algorithm choice	Sept 22, 2025	Oct 3, 2025	Soumya
3	Core Algorithm Development	Working code prototypes of regression, clustering, outlier detection	Oct 2, 2025	Oct 23, 2025	Anvitha (Regression), Sahasra (Clustering), Soumya (Outlier Detection)
4	Regression Model (Forecasting)	Linear regression implementation, RMSE/MAE results	Oct 6, 2025	Oct 15, 2025	Anvitha
5	Clustering Model (Operating Modes)	K-Means clustering, cluster plots & silhouette score	Oct 10, 2025	Oct 18, 2025	Sahasra
6	Outlier Detection	Z-score/IQR anomaly detector,	Oct 12, 2025	Oct 22, 2025	Soumya

		flagged anomalies			
7	Integration & Dashboard Development	Streamlit dashboard with forecasting, clustering, anomaly modules	Oct 24, 2025	Nov 10, 2025	Sahasra (lead), Anvitha & Soumya (support)
8	Experimental Evaluation & Optimization	Evaluation report (RMSE, silhouette, anomaly precision/recall), optimization recommendations	Nov 11, 2025	Nov 25, 2025	Soumya (lead), All contribute
9	Final Report & Presentation	Final report, slides, demo video	Nov 26, 2025	Dec 12, 2025	All members (Anvitha – writing, Sahasra – slides, Soumya – demo)

6) References

1. [Fan, 2014] Fan, C., Xiao, F., & Wang, S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Applied Energy, Elsevier, Vol. 127, April 2014, pp. 1–10. Available at: https://www.sciencedirect.com/science/article/pii/S0306261914003596
2. [Jakubowski, 2024] Jakubowski, J., Wojak-Strzelecka, N., Ribeiro, R. P., Pashami, S., Bobek, S., Gama, J., & Nalepa, G. J. Artificial Intelligence Approaches for Predictive Maintenance in the Steel Industry: A Survey. Preprint on arXiv, May 2024, pp 4-11. Available at: https://arxiv.org/abs/2405.12785
3. [Tang, 2015] Tang, E., Shao, Y., Fan, X., Ye, L., & Wang, J. Application of Energy Efficiency Optimization Technology in Steel Industry. International Journal of Control and Automation, Science & Engineering Research Support Society (SERSC), Vol. 8, No. 3, March 2015, pp. 9 pages. Available at: https://www.researchgate.net/publication/274020102_Application_of_Energy_Efficiency_Optimization_Technology_in_Steel_Industry
4. [Zhang, 2021] Mingyang Zhang, Jihong Yan A data-driven method for optimizing the energy consumption of blast furnaces. Journal of Cleaner Production, Vol. 280, 2021, 13 pages ,Article 124317. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0959652620349064
5. [Zheng, 2023] Zheng, M., Geng, L., Zuo, B., & Nakata, T. A Dynamic Thresholds Based Anomaly Detection Algorithm in Energy Consumption Process of Industrial Equipment. In Proceedings of the 2023 7th International Conference on Big Data and Internet of Things (BDIOT 2023), Beijing, China, August 11–13, 2023. ACM, New York, NY, USA, 9 pages. Available at: https://doi.org/10.1145/3617695.3617706