

PROJECT TITLE: COMPARISON OF CLUSTERING ALGORITHMS ON ENERGY CONSUMPTION IN STEEL INDUSTRY USING WEKA

INTRODUCTION

Energy usage prediction in steel power plants involves analyzing historical data on energy consumption and other relevant variables using machine learning algorithms to create predictive models. These models can be used to forecast future energy usage and optimize energy management, resulting in increased energy efficiency and reduced costs.

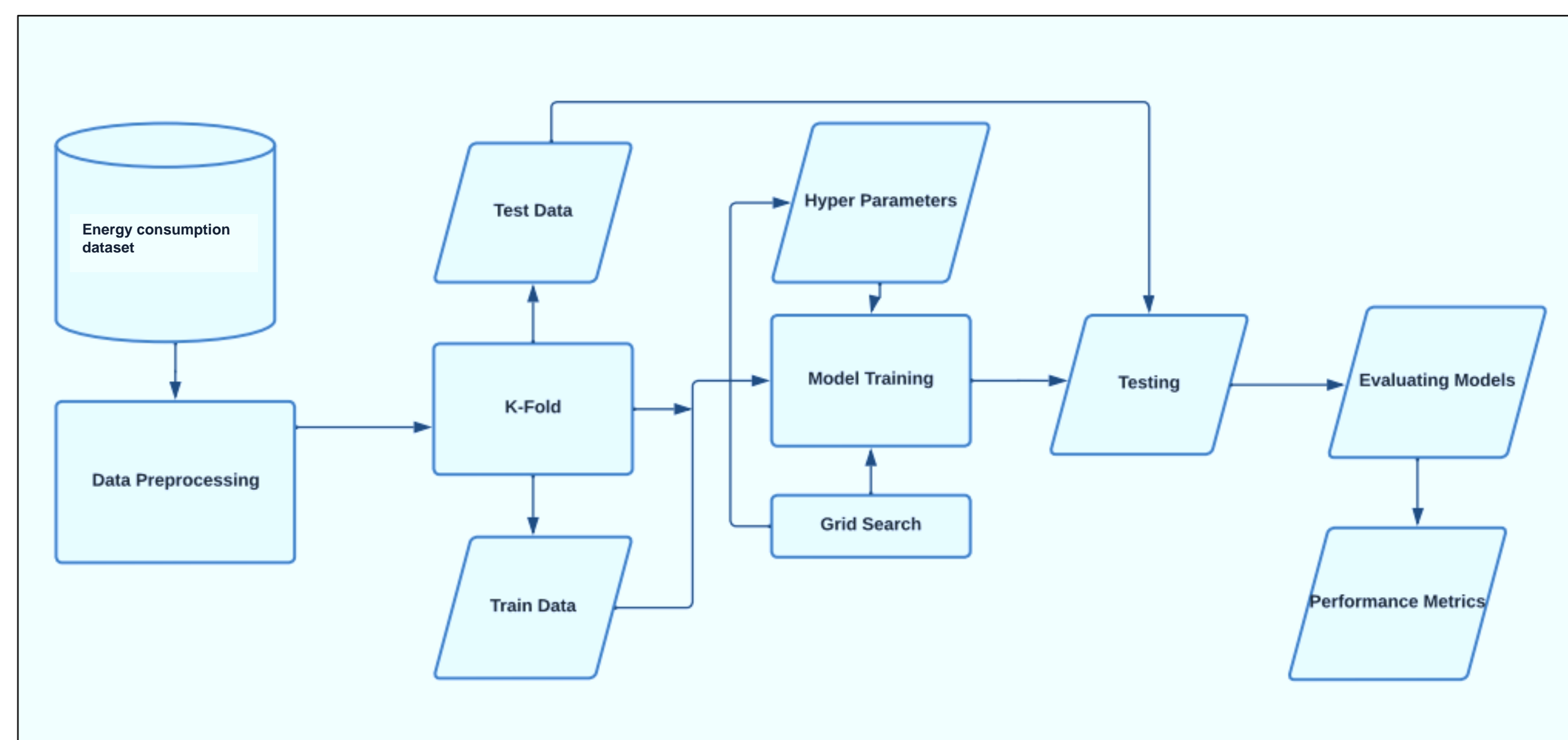
Weka is a popular machine learning tool that can be used to perform various data mining tasks, including clustering. In the case of energy usage clustering can help identify patterns and similarities about the various factors effecting the usage.

OBJECTIVES & AIMS:

- To explore and analyse clustering algorithms such as KMeans, DBSCAN and Agglomerative using WEKA tool.
- Overall, a comparison of clustering algorithms on energy consumption in a steel power plant using Weka can help identify the most effective algorithm for clustering energy consumption data and provide insights into energy usage patterns in the plant.

METHODOLOGY

- Data Preparation:** Energy consumption dataset contains various features such as usage, carbon dioxide emission, and several other technical features. The dataset can be pre-processed by removing any missing values, selecting relevant features, and normalizing the data.

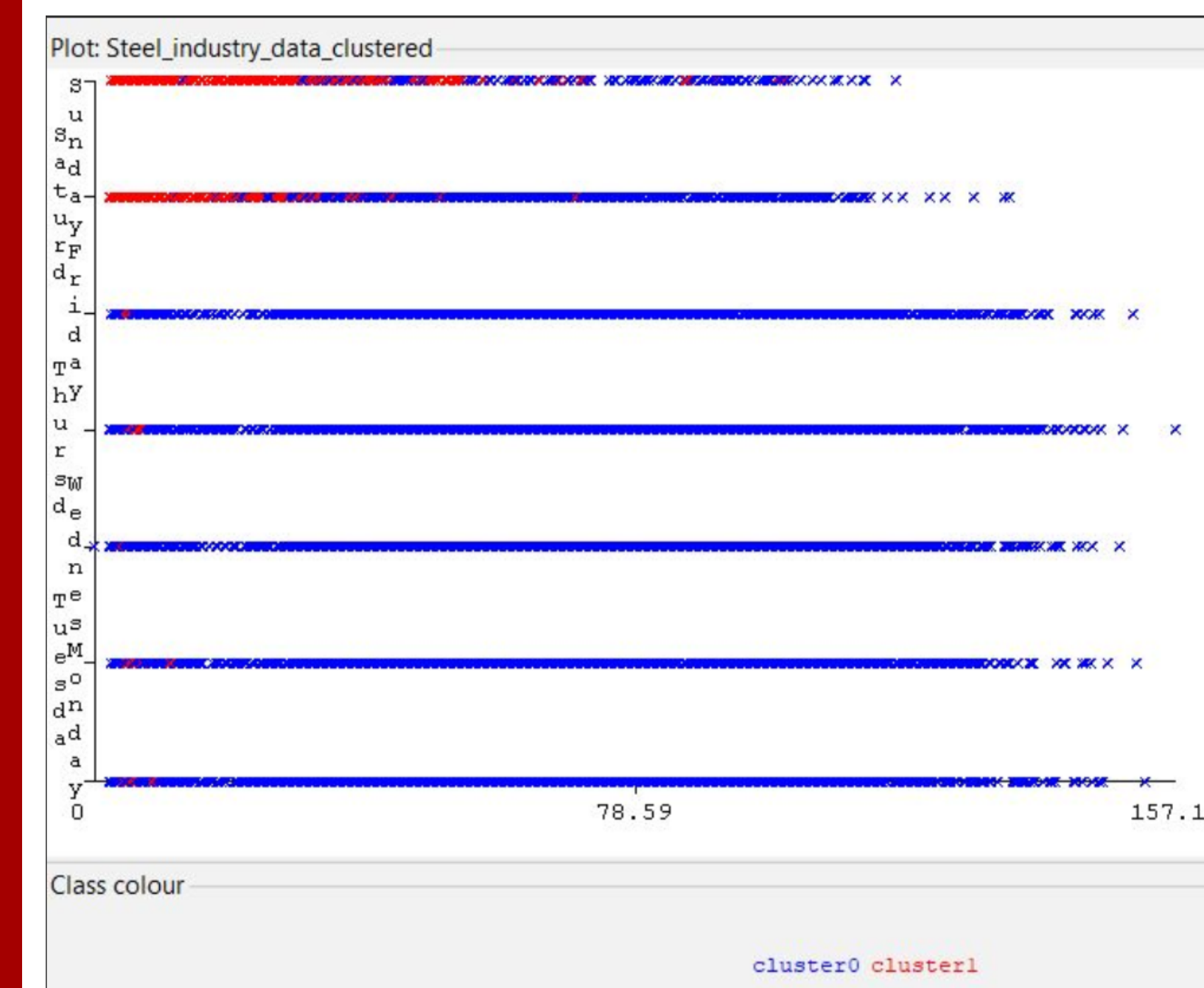


- Cluster Building:** After specifying the number of clusters, the algorithm can be trained on the pre-processed dataset.
- Model Testing:** The model built using the algorithms can be tested on the test dataset to evaluate its performance. Weka provides several evaluation metrics such as silhouette coefficient and Dunn index to evaluate the quality of the clusters formed and then finally visualised.

DATASET

The information gathered is from the DAEWOO Steel Co. Ltd in Gwangvang. South Korea. It produces several types of coils, steel plates, and iron plates. The information on electricity consumption is needing 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. It contains 11 columns and 35040 rows.

RESULTS:



Farthest First

| | Clustered | Instances |
|---|-----------|-----------|
| 0 | 25273 | 72% |
| 1 | 9767 | 28% |

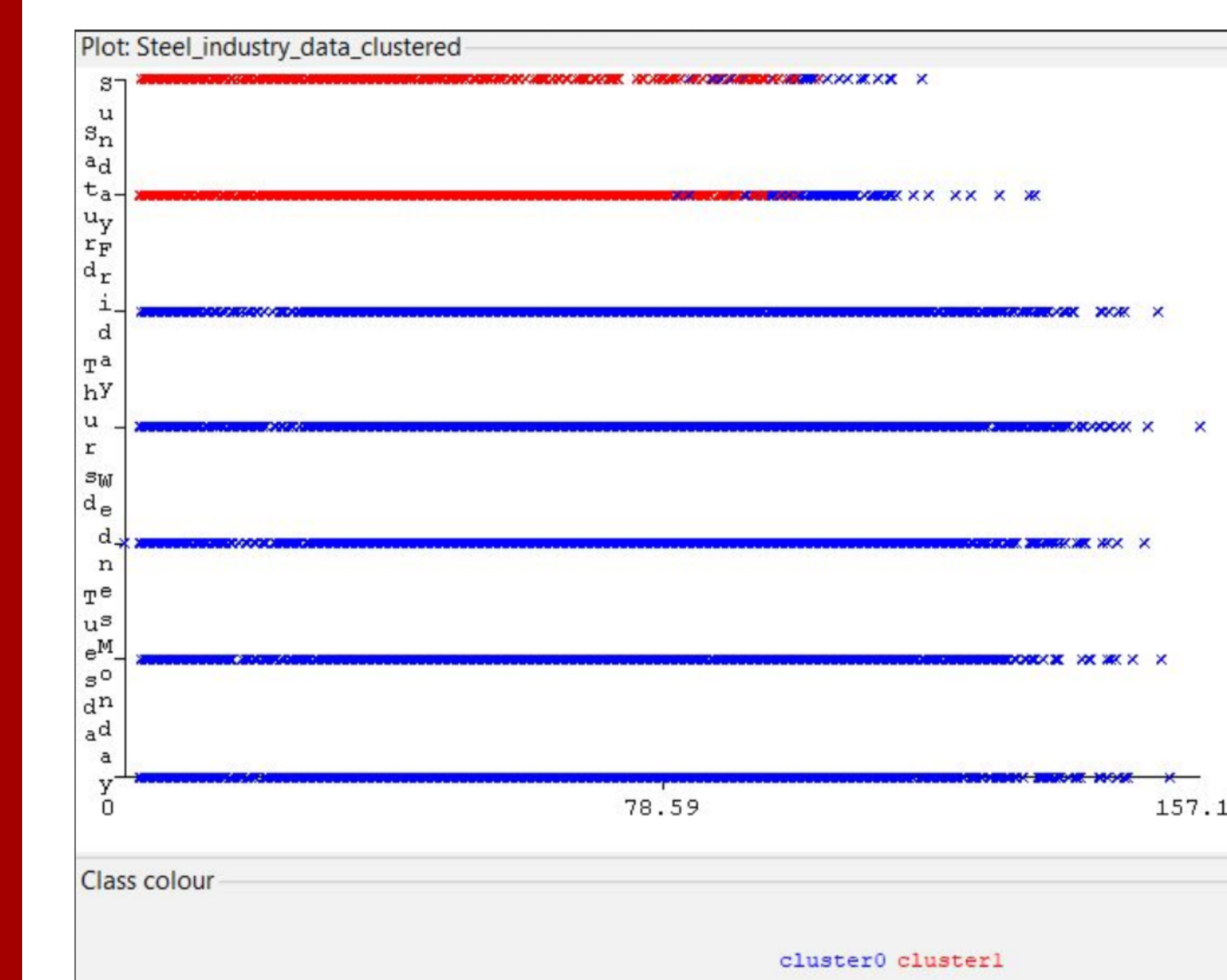
Time taken on training data: 0.27 s



KMeans Clustering

| | Clustered | Instances |
|---|-----------|-----------|
| 0 | 25056 | 72% |
| 1 | 9984 | 28% |

Time taken on training data: 0.27 s



DBSCAN Clustering

| | Clustered | Instances |
|---|-----------|-----------|
| 0 | 25178 | 72% |
| 1 | 9862 | 28% |

Time taken on training data: 0.35 s

CONCLUSION AND FUTURE SCOPE: Kmeans has the best clusters.

The use of machine learning tool like Weka enhances our understanding of Energy consumption in steel power plants and can potentially lead to energy consumption outcomes for a brighter future.