

Optimizing Energy Efficiency in EAF Steel Production: A Data-Driven Approach

Yubin Zhao, Yize Qi*, Jose Diogo Franco Viveiros

School of Engineering, Cardiff University, The Parade, Cardiff, Wales, CF24 3AA, UK

* Corresponding author

Abstract: The article's main research mission is making predictions about energy usage within an EAF steel production operation and being able to reliably predict energy consumption. Besides, we analysed the relationship between energy consumption in electric arc furnace (EAF) production processes by the method of data mining, so we could obtain accurate forecasts to optimize production costs and maximize energy efficiency. Last but not least, by using decision tree, random tree algorithms and multilayer perceptron models, we can easily achieve the intended goals of efficient use of energy and provide a better plan for electric arc furnace (EAF).

Keywords: EAF, Weka, Data analysis, Data modelling.

1. Introduction

Nowadays, with the development of modern industry, we can't ignore that the steel industry is still playing a very important role in the industry of any country around whole world. Britain and other European countries are the earliest batch of countries to develop the steel during the First Industrial Revolution, so over the past century, scientist developed many ways in order to do the steel production, such as Smelting Reduction, Gas-based Direct Reduction and Direct Reduction Ironmaking, or DRI. Besides, one of the most widely used is the Electric Arc Furnace.

Although the Electric Arc Furnace is the most widely used way of steel production, there are still many problems occurred during the process, which can cause a series of problems. Like Furnace process control, Slag control and improvement of energy efficiency. In order to solve the problems, we must take the full advantages of date, which, in that case, a suitable type of data mining formulation has to be considered, and it not only can deal with the problems we meet during the steel production, but also have potential opportunity, which can lead the steel industry into a brighter future.

2. Data Understanding

2.1. Data Exploration and Description

We can analyse the data given in appendix(Fig 1, Fig2), this way making any change to the data, if needed, and get a brief introduction about dataset.

The dataset includes 21 variables and 3499 data points, including precise measures of energy usage, the manufacture of recycled steel, and the utilization of metal materials. When having a preliminary investigation of the values imported into WEKA, it is notable cases of unclear and missing data.

Instead of having a focus on production, the main objective of this data mining project for steel production in EAF's is to optimize the energy use. In terms of optimizing energy use, if "Billet Tons" represent the amount of steel produced per batch, then this parameter may won't have a significant impact on optimizing the energy use. This attribute ("Billet Tons") didn't have a significant contribution on predicting the energy

consumption and was considered for us as unimportant for the feature selection in the pre-processing stage, so the decision was to remove it. Furthermore, each sample was assigned a serial number called "Heat Number", which was optional and didn't have any significance importance in aiding the optimization.

2.2. Data processing

With the intention of further reducing the attribute and simplifying the model even more, the option "CorrelationAttributeEval" was selected in the attribute evaluator panel and the cut-off value was set to 0.02.



Figure 1. Select attributes tab on Weka.

The properties with values between -0.02 and 0.02 showed a low relevance to the experiment, so they were deleted together with the attribute "Argon (kg)".

We also discovered during the data filtering process that the Steel Grade attribute has two outliers of 00/01/1900 and four consecutive outliers of 0. We made the decision to include a

filter so the outliers can be managed so we can keep them from throwing off the algorithm.

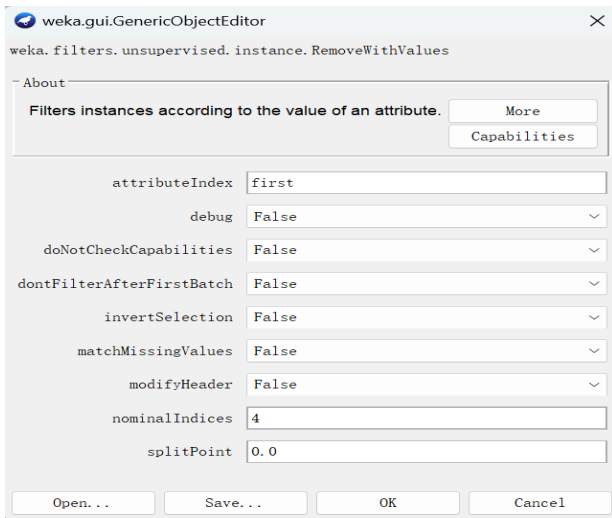


Figure 2. GenericObjectEditor Tab on Weka.

Thus, the missing data can now be checked and handled,

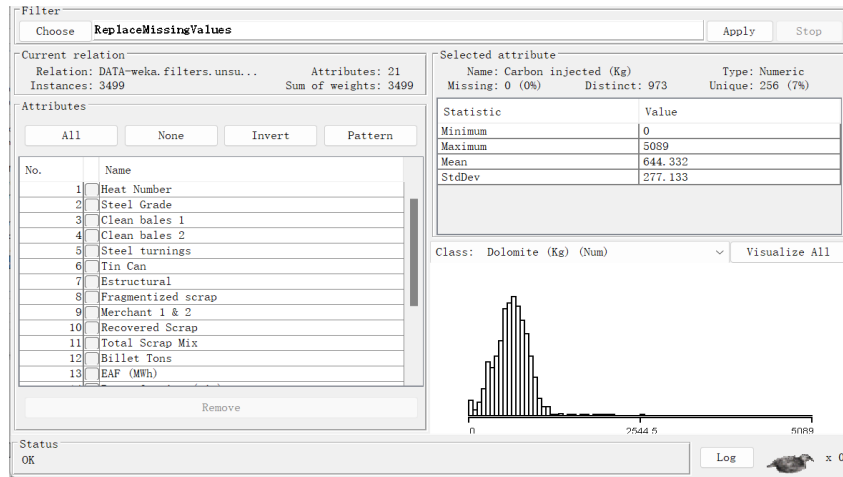


Figure 3. Process tab in Weka.

3. Methodology

The algorithms that we decided to choose to analyse the data were the J48 for Decision Trees and RandomTree. We choose these two models because they both have great advantages, such as high accuracy when dealing with small and medium sized datasets. Also, the model Multiplayer Perceptron was used. This model has its unique advantages when dealing with complex energy systems, where the model can adjust to the number of layers and neurons to better adapt to the more complex non-linear relationships. These three models were chosen because they are very distinct learning and pattern recognition approaches. Furthermore, this will allow us to have a more comprehensive understanding of the relationship between energy consumption and attributes.

- **J48 Decision Tree:** This is a decision tree algorithm, which it is often used the divide-and-conquer method in order to divide the instances, this way according to the similarity. The J48 algorithm is an application of the C4.5 algorithm, which is a very stable and reliable algorithm, as we can determine through long-term verification.

- **Multiplayer Perceptron:** This is a very common multi-layer feed-forward artificial neural network model. It's an

which can be judged by the values displayed by “Missing” in the selected attribute. The “ReplaceMissingValues” function in Weka can be responsible for handling the missing data.

In conclusion, those things which can get from the dataset are as following:

1) There is a moderate positive correlation between Heat Number and EAF (MWh) about 0.56, which may show that an increase in energy consumption as production continues.

2) The Carbon injected (Kg) and EAF (MWh) have moderate positive correlation about 0.39, which may suggested that the amount of carbon used may have an impact on energy consumption.

3) As for Total Scrap Mix and EAF (MWh), their moderate positive correlation is only 0.34, so it's goes without saying that total waste may be a key factor affecting energy consumption.

So in order to improve energy efficiency. The following elements need to be optimized, such as The time of the heat, amount of carbon injected and scrap mix. Besides, Focusing on carbon usage and waste mix optimization may help reduce energy consumption, thereby lowering costs and increasing production efficiency.

algorithm often used for classification and prediction of data, and it includes one or more hidden layers that allows it to learn non-linear patterns in the data, being a acceptable tool for deep learning.

- **RandomTree:** This algorithm can be used for both classification and regression problems. It is an important algorithm in trees. Each of its internal nodes represents a test of a characteristics, each branch represents the result of the test, and each leaf node represents a prediction.

When writing up the methodology, we always tried to maintain the importance of validating the results of this experiment to ensure the accuracy and stability of the results. In addition, we also paid attention to the iterative improvement of this experiment. The algorithm was replaced several times to find the most suitable algorithm for this experiment.

4. Model Building

4.1. MultilayerPerceptron

The MultiplayerPerceptron is a neutral model in Weka, commonly used in data mining and machine learning. This algorithm has its own benefits when working with small to

medium sized data, even though it is not as accurate as deep learning algorithms when working with larger datasets.

We determined that it was best to split the data into ten sections, nine for training and one for testing, based on our

experience. With this said, all the data shown below was acquired using a cross-validation of 10.

Note that the target attribute must be changed to (Num)EAF (MWh)

Table 1. Model training results after data segmentation.

Learning Rate	0.3	0.3	0.3	0.3	0.4	0.5	0.3	0.3	0.3
momentum	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.4	0.5
percent split	60	66	76	86	66	66	66	66	66
Correlation coefficient	0.8081	0.7924	0.8166	0.8142	0.7857	0.7644	0.7738	0.7921	0.7911
mean absolute error	2.0465	1.7199	1.9788	1.8565	1.7555	1.7918	1.7573	1.681	1.6878
Root mean squared error	2.6707	2.4518	2.54	2.4501	2.5537	2.5789	2.5564	2.4446	2.4931
Relative absolute error	70.321	58.078	66.7905	62.2219	59.2815	60.5079	59.34	56.766	56.9932
Root relative squared error	70.8842	63.3657	67.0566	64.2228	65.9995	66.6507	66.07	63.1805	64.4343

```

Classifier output
  Attrib Lime and Dolomite (Kg)  -0.13824517565169875
  Attrib Dolomite (Kg)         -2.255510068388434
Class
  Input
  Node 0

Time taken to build model: 8.72 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correlation coefficient          0.7921
Mean absolute error             1.681
Root mean squared error         2.4446
Relative absolute error         56.766 %
Root relative squared error     63.1805 %
Total Number of Instances      1189
  
```

Figure 4. Classify tab on Weka.

Looking at the data above in fig.6, we can see that when the values of the Learning Rate is 0.3, the Momentum is 0.4 and the Percentage split is 66%. The values of Mean absolute error, Root mean squared error, Relative absolute error and

Root relative squared error in the data are minimum, so it can easily get that the error between the actual value and the predicted value is minimum in this data.

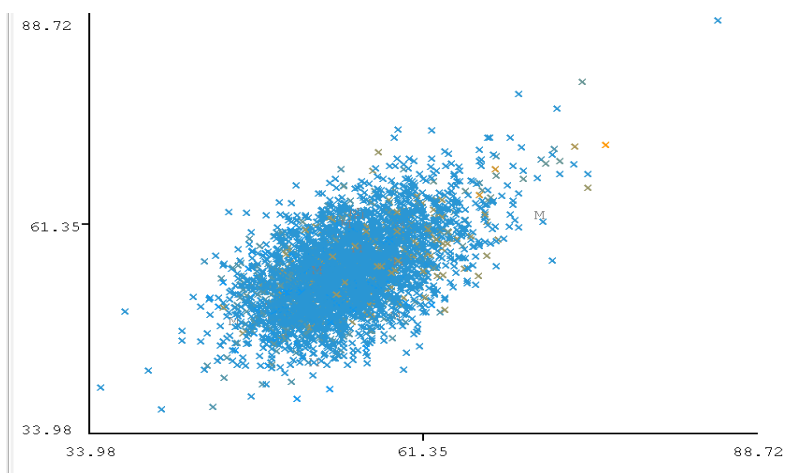


Figure 5. Program execution process and regression diagram.

4.2. J48

The J48 algorithm offers a high degree of accuracy when analysing small to medium sized datasets, a clear structure

and, results that are simple to read and comprehend. The algorithm

can also handle attributes with values missing and has a high fault tolerance. It usually struggles to manage

characteristics that often require constant performance, tough. When using the default parameter, we must first use the option Discretize in order to change the attribute from

numeric to nominal. Alternatively, we can also use the NumericalToNominal attribute to change the attribute from numeric to nominal.

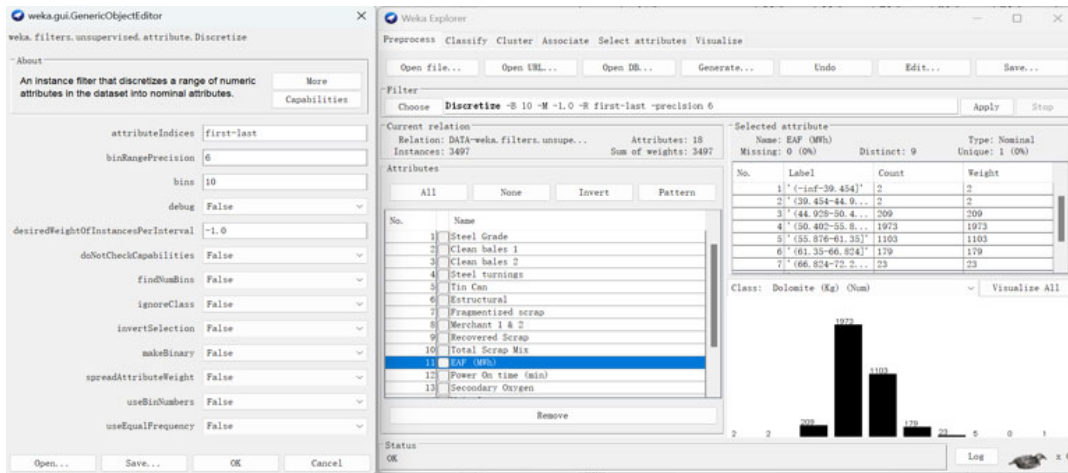


Figure 6. EAF data type conversion process.

We then turned Cross-validation to 10 and used the algorithm J48 to run different percentage split values.

Table 2. Model training results after data segmentation.

percent split	56	66	76	86	96
Correctly Classified Instances	68.2911	69.6883	68.2956	70.2041	69.2857
Incorrectly Classified Instances	31.7089	30.3117	31.7044	29.7959	30.7143
Mean absolute error	0.0808	0.0806	0.0808	0.0795	0.0816
Root mean squared error	0.2239	0.2157	0.2225	0.2169	0.2164
Relative absolute error	69.8711	69.8533	69.1381	67.8949	68.4355
Root relative squared error	92.7772	89.8878	91.0512	88.7971	87.1663

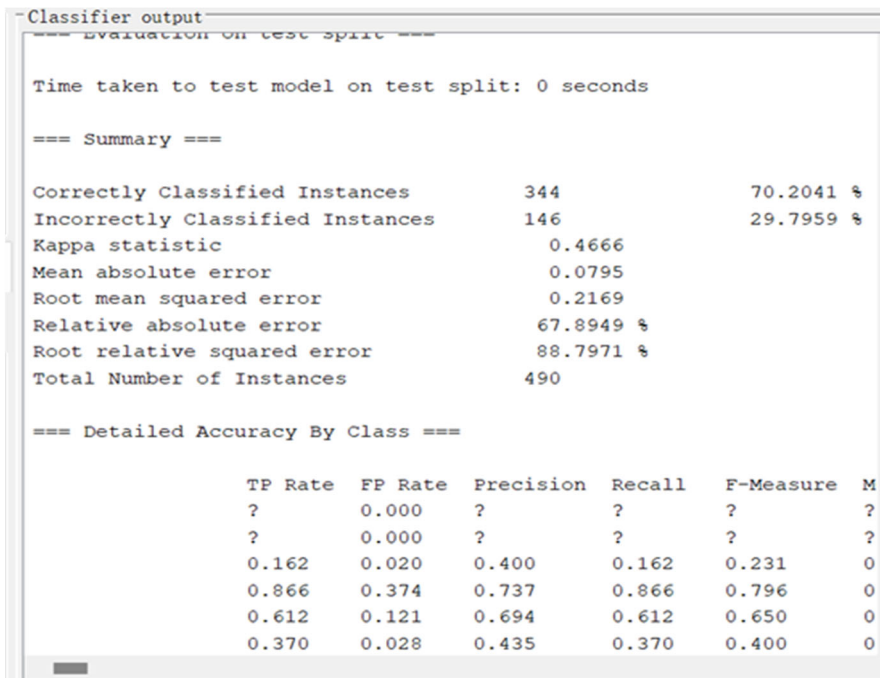


Figure 6. Classify tab on Weka.

We observed the following values through the window Classifier on output.

We can deduce from the size of the values in the above table that the accuracy is the highest when the percentage split

value is 56%. In Weka, selecting Visualise offers a more user-friendly way to present the data so that we can see how various attributes relate to one another. Moreover, users can also spot outliers by examining the image.

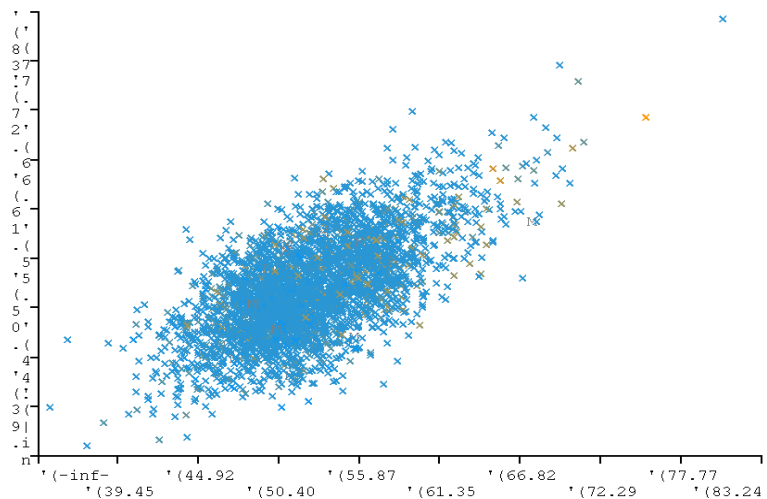


Figure 7. Program execution process and regression diagram

4.3. RandomTree

In Weka, the RandomTree algorithm is often used for machine learning and also data mining. It produces a decision tree model with an easy-to-understand structure. It can handle a variety of data types, these including nominal and numerical

data, which makes the operation simpler, although there is some randomness present and the accuracy is not the highest.

The cross-validation is set, Folds 10 and then the RandomTree algorithm is used to classify the data. The arithmetic resulting from varying the percentage split values produces the following data.

Table 3. Model learning training results after data segmentation

percent split	56%	66%	76%	86%	96%
Correctly Classified Instances	64.9773	63.3305	62.5745	66.3265	68.5714
Incorrectly Classified Instances	35.0227	36.6695	37.4255	33.6735	31.4286
Mean absolute error	0.0728	0.0755	0.076	0.068	0.0657
Root mean squared error	0.2559	0.2635	0.2633	0.2488	0.2376
Relative absolute error	62.9523	64.9629	65.0131	58.0774	55.0952
Root relative squared error	106.0335	108.4655	107.7565	101.8892	95.6949

From the data shown in the above table we can conclude

that when the value of the percentage split is 96%, the accuracy of this data is the highest and the error is the least.

```

Classifier output
| | | Clean bales 2 = '(89.874-inf)' : '(-inf-39.454]' (0/0)
| | | Power On time (min) = '(61.67-inf)' : '(-inf-39.454]' (0/0)

Size of the tree : 11306

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances          96          68.5714 %
Incorrectly Classified Instances       44          31.4286 %
Kappa statistic                        0.4725
Mean absolute error                    0.0657
Root mean squared error                0.2376
Relative absolute error                 55.0952 %
Root relative squared error            95.6949 %
Total Number of Instances              140

=== Detailed Accuracy By Class ===

```

Figure 8. Classify tab with the Percentage split set to 96 on Weka.

Thus, in order to be able to observe the data in a better visually way, the graphical decision tree can also be opened.

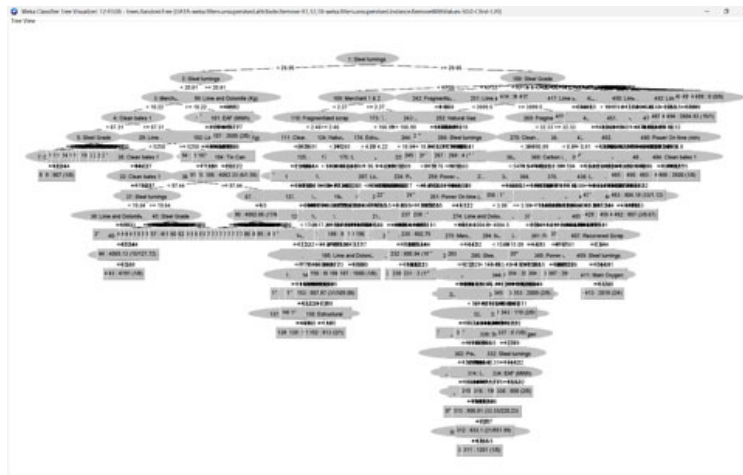


Figure 9. Visualize tree.

5. Results, Performance and Evaluation

In this part, the performance results and evaluation outcomes of the models will be discussed, the ones we used on our group's case study. In order to accomplish the target of energy consumption prediction in the EAF production process, we took into account the next ways:

Performance Metrics Rationale:

- **Accuracy and Error Metrics:** For the target of providing a comprehensive picture of the model's prediction accuracy and error size. We took the metrics such as Mean Absolute Error, Root Mean Squared Error, and Correctly Classified Instances.

- **Relative Errors:** For the Relative Errors, we choose Relative Absolute Error and Root Relative Squared Error, so we could know how our model would perform against baseline predictors, so we could get a better result.

Performance Results:

- **Multiplayer Perceptron:** I believe our model performs well in terms of accuracy, but we can still see from the error rate that there is some variability in the predictions. Besides, we should also notice that the correlation coefficient is significantly high, which means there is a relationship between the predicted and actual values.

- **J48 Decision Tree:** Based on the database we have, there is even a high tolerance for missing attribute values, we think it's still beneficial to the nature of our dataset.

- **RandomTree:** With the help of the RandomTree, it's easier to deal with various data types, including nominal and numerical.

After we conducted a comprehensive data analysis by Weka, we can get the performance metrics. In addition, the error rates, accuracy percentages, and the correlation coefficients were derived from tests and validations.

As for the pictures and graphs which are referenced in the Methodology section, they provide a representative and quantitative evidence of each model's performance.

Last but not least, from the presentation analysis of the three different models, it goes without saying that every one of them have their advantages and disadvantages. We believe that in order to do a good prediction of the energy consumption, we should take all of them into account, which would be the most effective way.

6. Conclusion and Reflection

This study has as its main aims to reveal the complex relationships of energy consumption in the EAF production process and provide more accurate predictions to optimize production costs and maximize energy efficiency. Besides, this article emphasized the importance of using different models in Weka. This not only upgrades the existing production model, but also helps the new insights for the development of smart manufacturing in the era of Industry 4.0.

Throughout the hole development of this case study in Weka, different models were used, each one being a different method of solving from each other, to make sure the prediction of the energy consumption would be the best. But there are still several things which need to be consideration.

1) In order to get the best prediction model, different software should be taken into use.

2) In the section of Data understanding, mathematical knowledge should be used as a tool to assist analysis

Finally, this case study showed that with the right data pre-processing, processing and with the right models, it is possible to get a high accuracy.

References

- [1] Huang, F. 2021. Network Activities Recognition and Analysis Based on Supervised Machine Learning Classification Methods Using J48 and Na^vive Bayes Algorithm. Available at: <https://arxiv.org/abs/2105.13698> [Accessed: 20 November 2023].
- [2] IOPscience. 2017. IOP Conference Series: Materials Science and Engineering - IOPscience. Available at: <https://iopscience.iop.org/journal/1757-899X>
- [3] Raheli, B., Aalami, M.T., El-Shafie, A., Ghorbani, M.A. and Deo, R.C. 2017. Uncertainty assessment of the multilayer perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: a case study of Langat River. Environmental Earth Sciences 76(14). doi: <https://doi.org/10.1007/s12665-017-6842-z>

Appendix

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	
	Heat Num	Steel Grade	Clean bal	Clean bal	Steel Eurs	Tin Con	Structural	Fragment	Mechanic	Recovered	Total Scrap	Blind	Losses/TAF	WAF	Power On	Secondary	Main Oxy	Natural	G/Argon	Kg	Carbon In	Loss and	Dutoms
1	1	KP08	0	13.4	73.79	50.73	2.82	0	14.94	0	164.672	133.31	70.68	57.65	1544	4284	227	1.09	2642	4257	808		
2	2	KP08	0	15.52	82.29	52.02	3.18	0	14.10	0	169.17	149.57	52.76	41.98	952	3500	202	1.01	324	4251	808		
3	3	KP08	0	10.58	80.05	55.86	3.58	0	15.42	0	165.483	141.14	80.82	40.88	932	3456	204	98	541	4254	808		
4	4	KP08	0	15.58	73.81	55.86	2.83	0	16.18	0	164.747	155.86	53.9	40.57	944	3433	203	84	530	4256	808		
5	5	KP08	0	9.85	81.76	54.33	3.25	0	15.41	0	164.602	150	51.01	40.47	816	3424	208	118	535	4253	808		
6	6	KP08	0	11.27	78.95	52.1	2.59	0	17.32	0	165.269	148.57	50.74	41.32	941	3533	225	80	551	4251	808		
7	7	KP08	0	11.91	78.14	55.88	3.38	0	15.87	0	164.988	148.57	51.57	41	958	3546	202	108	560	4453	807		
8	8	KP08	0	11.14	81.56	52.58	3.19	0	16.51	0	165.075	139.25	51.11	41.1	958	3689	209	83	558	4474	809		
9	9	KP08	0	10.76	81.13	55.65		0	15.67	0	165.147	144.45	52.06	40.85	960	3480	200	81	564	4655	808		
10	10	KP08	0	11.63	79.88	53.63	2.51	0	13.49	0	164.17	152.17	51.04	41.07	1010	3508	208	97	576	4654	806		
11	11	KP08	0	14.49	77.80	56	2.12	0	15.51	0	165.81	144.45	51.35	41.28	1118	3653	207	80	570	4653	807		
12	12	KP08	0	11.71	81.89	53.63	3.08	0	14.01	0	164.276	144.45	51.44	40.02	914	3427	194	104	575	4654	807		
13	13	KP08	0	13.19	80.25	55.19	2.71	0	14.29	0	165.959	158.2	52.25	40.98	922	3505	192	106	589	4655	807		
14	14	KP08	0	12.16	80.53	55	2.0	0	15.11	0	165.392	144.85	51.58	40.12	1142	3468	196	95	588	5210	914		
15	15	KP08	0	12.51	79.54	55.79	2.49	0	16.22	0	166.748	151.21	51.47	40.2	1064	3303	202	85	571	4655	807		
16	16	KP08	0	14.02	76.22	54.66	3.09	0	18.07	0	165.928	151.21	51.01	39.67	998	3399	201	108	467	4656	808		
17	17	KP08	0	11.39	76.07	56.96	2.42	0	16.45	0	165.328	151.21	50.77	39.57	942	3381	189	101	497	4657	807		
18	18	KP08	0	11.89	80.93	54.9	2.62	0	16.48	0	165.919	154.45	51.48	40.17	1016	3376	202	85	557	4456	807		
19	19	KP08	0	12.94	74.85	56.4	3.71	0	14.86	2.16	164.428	158.2	50.54	39.52	1022	3215	194	116	449	4487	808		
20	20	KP08	0	13.78	71.79	59.22	1.92	0	15.01	0	161.706	147.74	52.88	41.07	950	3381	186	107	591	4451	807		
21	21	KP08	0	16.28	80.53	53.24	2.26	0	14.11	0	165.409	147.24	51.26	39.85	895	3410	200	113	575	4457	809		
22	22	KP08	84.89	71.5	0	0	0	0	4.53	0	160.254	159.86	50.56	38.84	1034	3358	198	81	379	5258	2610		
23	23	KP08	80.7	79.35	0	0	0	0	2.84	0	162.887	162.58	61.84	40.3	958	3455	202	87	600	5058	2610		
24	24	KP08	79.34	78.39	0	0	0	0	3.95	0	161.718	163.80	62.11	40.37	1098	3397	202	81	513	5067	2609		
25	25	KP08	80.76	78.15	0	0	0	0	4.4	0	163.362	153.86	52.52	41.17	876	3463	196	91	677	5062	2612		
26	26	KP08	80.68	74.35	1.91	0	0	4.52	0	161.468	161.45	62.21	40.58	934	3434	188	110	688	5057	2610			
27	27	KP18	0	111.24	5.11	2.57	31.07	16.12	1.62	165.959	167.2	58.8	41.23	1032	3576	203	117	192	4295	9			
28	28	KP18	0	115.58	3.51	1.94	33.75	15.45	0	170.257	175.99	55.57	43.43	1194	3723	205	109	472	4203	5			
29	29	KP18	0	118.47	4.02	2	42.03	18.77	0	182.132	147.96	57.09	48.7	1138	4685	183	80	970	4205	6			
30	30	KP18	0	114.06	3.83	2.13	28.21	18.88	0	164.87	147.96	53.87	43.35	1192	3732	185	120	1306	4206	6			
31	31	KP18	0	113.37	4.02	3.03	38.17	19.29	0	175.451	143.33	57.62	46.82	1058	4011	189	113	142	4205	6			
32	32	KP18	0	118.3	5.19	2.08	24.51	15.47	1.62	165.141	147.96	51.93	40.06	968	3513	178	113	1072	4204	6			
33	33	KP18	0	110.7	4.05	2.21	31.39	15.14	1.98	165.475	147.96	52.44	40.06	1068	3819	185	113	848	4202	5			
34	34	KP18	0	114	4.03	2.1	29.17	15.29	2.39	166.852	147.96	52.04	40.93	1008	3634	189	105	232	4003	5			
35	35	KP18	0	108.81	4.33	3.26	29.53	16.38	3.3	164.593	152.58	51.77	39.12	938	3382	171	108	290	4000	6			
36	36	KP18	0	110.8	4.09	2.51	21.7	16.05	1.87	166.81	162.58	51.41	39.8	954	3454	178	118	255	4002	5			
37	37	KP18	0	106.15	4.12	2.91	31.59	17.79	2.14	164.704	147.96	51.2	38.54	914	3325	172	84	256	4505	110			
38	38	KP18	0	115.59	5.11	2.01	23.97	19.5	2.95	166.15	147.96	51.25	38.85	1010	3317	178	117	255	3499	7			

Fig. 1 Dataset in an Excel file.

Weka																
Location: DML																
No	1: Heat Number	2: Steel Grade	3: Clean bal	4: Clean bal	5: Steel Eurs	6: Tin Con	7: Structural	8: Fragmented scrap	9: Mechanic	10: Recovered Scrap	11: Total Scrap	12: Dillet loss	13: WAF	14: Power On	15: Free Ca loss	16: Secondary
1	1	KP08	0	13.4	73.79	50.73	2.82	0	0	164.672	133.31	70.68	57.65	1544	4284	227
2	2	KP08	0	15.52	82.29	52.02	3.18	0	0	169.17	149.57	52.76	41.98	952	3500	202
3	3	KP08	0	10.58	80.05	55.86	3.58	0	0	165.483	141.14	80.82	40.88	932	3456	204
4	4	KP08	0	15.58	73.81	55.86	2.83	0	0	164.747	155.86	53.9	40.57	944	3433	203
5	5	KP08	0	9.85	81.76	54.33	3.25	0	0	164.602	150	51.01	40.47	816	3424	208
6	6	KP08	0	11.27	78.95	52.1	2.59	0	0	165.269	148.57	50.74	41.32	941	3533	225
7	7	KP08	0	11.91	78.14	55.88	3.38	0	0	164.988	148.57	51.57	41	958	3546	202
8	8	KP08	0	11.14	81.56	52.58	3.19	0	0	165.075	139.25	51.11	41.1	958	3689	209
9	9	KP08	0	10.76	81.13	55.65		0	0	165.147	144.45	52.06	40.85	960	3480	200
10	10	KP08	0	11.63	79.88	53.63	2.51	0	0	164.17	152.17	51.04	41.07	1010	3508	208
11	11	KP08	0	14.49	77.80	56	2.12	0	0	165.81	144.45	51.35	41.28	1118	3653	207
12	12	KP08	0	11.71	81.89	53.63	3.08	0	0	164.276	144.45	51.44	40.02	914	3427	194
13	13	KP08	0	13.19	80.25	55.19	2.71	0	0	165.959	158.2	52.25	40.98	922	3505	192
14	14	KP08	0	12.16	80.53	55	2.0	0	0	165.392	144.85	51.58	40.12	1142	3468	196
15	15	KP08	0	12.51	79.54	55.79	2.49	0	0	166.748	151.21	51.47	40.2	1064	3303	202
16	16	KP08	0	14.02	76.22	54.66	3.09	0	0	165.928	151.21	51.01	39.67	998	3399	201
17	17	KP08	0	11.39	76.07	56.96	2.42	0	0	165.328	151.21	50.77	39.57	942	3381	189
18	18	KP08	0	11.89	80.93	54.9	2.62	0	0	165.919	154.45	51.48	40.17	1016	3376	202
19	19	KP08	0	12.94	74.85	56.4	3.71	0</								