Electric Arc Furnace Optimization

Data Mining Mini-project

Joaquin Velarde
Department of Data Science
University of Skövde
Skövde, Sweden
a24joave@student.his.se

***Abstract*—Electric Arc Furnaces play an important and promising role in modern steel production. For optimizing their operation, it is necessary to balance a lot of complexly related values, such as energy consumption, process temperature, gas and oxygen usage or material composition, among many others. This project explores the potential for prediction and optimization across a number of these variables.**

Firstly, we try to predict temperature based on energy usage. This will prove very challenging, forcing us to change our approach. After testing, we decided energy optimization was a more adequate problem. We performed regression on energy consumption using a number of independent variables, including gas usage, total material charged in the furnace and more. This is relevant because of the economic and environmental effects excess of energy usage can have.(*Abstract***)**

Keywords—furnace, energy, regression, optimization (*key words****)***

I. Introduction

Electric Arc Furnaces have been growing in importance over the last decades, and will likely continue to do so in the years to come, due mainly to their small environmental impact. Optimizing an Electric Arc Furnace from a Data Science perspective is a hard job, since there are a lot of possible questions that can be asked. Some of the possible problems one can aim to solve are: energy usage optimization, stabilization of final oxidation values, final temperature predicting, final chemical composition, etc.

To the inexperienced, it can be challenging to answer these questions, due to the broadness of EAF data. Some variables might be very relevant for predicting certain features, specially combined with other meaningful information, but others might be completely useless.

For a first approach, we will attempt to predict final temperature of a metal heat using as a predictor the energy used for processing the heat. This is meaningful since it would explicit the suspected strong relationship between energy spent and temperature, making it easy to automatize electricity consumption.

It turned out it was very difficult for us to find a strong relationship between energy and temperature, probably due to the influence of other variables, which we were not able to find. We explored a little bit to see if we could find such variables, to no success. Another probable cause could be a faulty engineering of the energy column, which we created manually.

This made us change approaches and optimize energy usage directly. Here, we had mild success. Our results can be used for performing further analysis on EAF-parameters optimization.

II. Methodology

First, we did some technical understanding. It was important to know what words like "heat" or "tapping" were. Also, it was important to understand what part of the process was reflected in each of our tables, and what variables could be relevant for predicting temperature or energy consumption. After deciding we wanted to predict temperature based mainly on energy data, we cleaned the temperature data and analyzed its distribution.

The energy variables were more tricky. There was no explicit variable describing energy consumption, so we had to create some variables that captured the information logged in the "eaf_transformer" file. It is unclear to what extent we did this correctly, since this can only be truly checked with an expert in the field. However, according to all the information we read, the variables we created made sense (which is not the same as saying they were optimal).

After cleaning and making sure our variables were logically robust, we started exploring the relationship between these variables and temperature. To our surprise, the relationships were really weak. This might have been caused by our engineering of the variables, but we are inclined to think energy is not as important as other factors, like carbon additions or gas usage.

We tried to explore how other variables influenced temperature, such as material composition and carbon additions, but none seemed to have a strong relationship with temperature. Actually, initial composition of our metals had almost no correlation with the final temperature. This is an important piece of information, since it tells us that material composition is almost perfectly handled in our process, at least with respect to temperature.

Since temperature forecasting was starting to seem very difficult and not as profitable, we decided to move on to another dependent variable. We think temperature forecasting is still worth trying, but, eager to get some modeling experience, we were not patient enough and changed our approach to energy consumption prediction.

This, in our minds, was a more practical and visual problem. If we could predict how much energy we were going to spend, it would be easy for us to cut energy costs.

We used one of the previously engineered columns as a dependent variable, and tried multiple independent variables until we found a configuration that gave good results.

Our final independent variables included: total amount of metal charged in the EAF through the main entrance, gas usage, oxygen usage and initial material composition.

In the preprocessing of these columns we found some interesting outliers. We dropped extreme values and applied logarithmic and square root transformations, but still felt like they could have been processed better. In a deeper analysis, maybe some clustering or other techniques can be applied to further understand the nature of this outliers and treat them in a more refined manner.

We started modeling then, trying to fit a regression model through our data.

Our main performance measures were R squared, MSE, RMSE and MAE, combined with sporadical graphs and other measurements of error.

Results were still not great, but after some preprocessing we were able to explain some of the variance of the dependent variable. It is worth noting that our dependent variable was quite widely distributed, which might contribute to the poor performance of some of our models.

We used Linear Regression, Ridge and Lasso Regression, XGBoost, Random Forest and Neural Networks to try to predict the energy values.

XGBoost, Random Forest and Neural Networks gave us the most performance. After some parameter tuning using Grid Search, Random Search, Hyperband, and manual tuning, we were able to build an ensemble method between the three that slightly bettered the previous performance of all three models.

III. Results

For concrete results, we have to describe our dependent variable. Our dependent variable was the megawatts per hour (MWh) spent on a batch. This variable, created by us, ranges from 0.5 to 18.

In our first steps predicting temperature, we found a linear correlation between temperature and our engineered energy consumption variable of about 0.18. Material composition features had less than 0.05 linear correlation with temperature.

Our baseline model, Linear Regression, got an R squared of less than 0.04 when testing this approach.

After changing our dependent variable, we had multiple correlations with energy consumption ranging from 0.10 to 0.20. We were able to increase R squared to 0.10 in our baseline model, Linear Regression.

XGBoost, Random Forest, and Neural Networks gave us 0.21, 0.25 and 0.20 R squared, which suggests some non-linearity.

After fine tuning and combining predictions, we were able to run a regression model that consistently gave a Mean Absolute Error of 1.77, RMSE of 2.29 and R squared of 0.27.

IV. Discussion

The first trials gave us insight on how the EAF works, and also highlighted the difficulty of capturing big amounts of information into few columns. We think it might be possible to do a better engineering of columns, leading to better results. This is the case in gas, oxygen, energy and amount of material columns, which were all engineered by us. For example, if instead of MWh we were to predict another measure or group of measures of energy, advised by an expert, a better performance might be reached.

Energy consumption and temperature had some correlation, but not nearly as much as we expected. This speaks to the enormous number of parameters that can influence temperature on an Electric Arc Furnace, and the complexity of the problem we had in our hands.

Material composition and temperature had almost no correlation, which tells us that no matter the material composition, the heat will be managed in a similar and succesful way, at least regarding temperature.

Trying more predictive variables for temperature forecasting could have been an option, but we did not pursue this path. This could be a suitable approach for further work into EAF parameter optimization.

We also found energy consumption can be roughly explained by our dependent variables. And that it followed non-linear and complex patterns, at least for our created columns. This could be of use to the EAF, if there is a wish to cut costs or reduce environmental impact.

V. Conclusion

Although metal composition and energy consumption do not seem to be enough, it seems worthwhile to conduct an analysis on temperature prediction through other variables in the EAF.

It is advisable, if work with energy or gas data is to be conducted, to contact an expert for feature engineering.

For further work, it would be interesting to explore outliers in most of the independent variables, specially oxygen usage and material composition.

It is possible to explain energy consumption in an EAF through a number of variables given before and while processing a metal batch.

Ethically and socially wise, the research gives a way for furnaces to be more respectful with the environment while cutting costs, which might make steel products cheaper and cleaner.