Mittuniversitetet
MID SWEDEN UNIVERSITY

Optimising Metal and Mineral Recovery from Waste Streams: A Data-Driven Approach

Jim Anderssén

Bachelor's thesis – Independent Project MX045G
Main field of study: Environmental Engineering
Credits: 15 Credits
Semester/year: Spring 2025
Supervisor: Robert Norgren
Examiner: Andreas Andersson

At Mid Sweden University, it is possible to publish the thesis in full text in DiVA (see appendix for publishing conditions). The publication is open access, which means that the work will be freely available to read and download online. This increases the dissemination and visibility of the degree project.

Open access is becoming the norm for disseminating scientific information online. Mid Sweden University recommends both researchers and students to publish their work open access.

I/we allow publishing in full text (free available online, open access):
☒ Yes, I/we agree to the terms of publication.

☐ No, I/we do not accept that my independent work is published in the public interface in DiVA (only archiving in DiVA).

Vasa, Finland, 30.5.2025 .....
Location and date

Environmental Engineering .....
Programme/Course

Jim Anderssén .....
Name (all authors names)

1996 .....
Year of birth (all authors year of birth)

Abstract

Global primary material use is expected to double by 2060, contributing to increased landfilled waste, environmental pollution and reduced access to metals and minerals. In order to address this challenge, a transition to a circular economy is essential. However, global circularity is decreasing annually, and circularity within the EU remains largely stagnant, indicating that this shift is not occurring at scale. Data-driven methods could be used as decision-making support for enabling development of waste recovery technologies to increase the input rate of secondary raw materials into the global economy.

This thesis presents a data-driven methodology consisting of: correlation matrix, clustering algorithms, domain confirmation and regression modeling to predict the oxide composition of ladle furnace slags more precisely compared to basic statistical methods. Evaluation using randomised datasets shows that the model improves prediction accuracy of oxide concentration by 37–45% for 4 out of 5 oxides with a variance reduction of 23–54%, compared to using statistical means. The result is lower error margins when estimating available quantity of oxides in ladle furnace slags, showing that the methodology can be used to better identify high potential mineral waste streams, improve the recovery of metals and minerals and thereby support the transition to a circular economy.

Table of Contents

1 Introduction

1.1 Observation

In recent decades the world has seen strong socio-economic developments. In the last half-century the global population has doubled to 8 billion, as shown by Ritchie et al. (2023) and the growth is projected to 10 billion by 2060 (OECD, 2019). The global economy is six times larger than half a century ago, due to an average annual growth rate of 4% (Kose & Ozturk, 2014). Much of this growth has raised living standards, enchanced life expectancy, increased availability of goods, services and energy, increased employment and improved education and healthcare (Circle Economy, 2024). While the population growth rate is expected to slow down in the coming decades, the living standards and income of the global population is still converging towards that of the most advanced countries, leading to projections of tripling the global GDP per capita by 2060 (OECD, 2019).

This growth and income convergence is expected to drive a strong increase in demand for material resources. An estimation by OECD (2019) suggests that global annual primary material use will nearly double, from 89 Gt in 2017 to 167 Gt, in 2060. While waste generation is commonly defined differently across the world and global data is inconsistent, Maalouf & Mavropoulos (2023) estimated that annual global solid waste in 2017 was 20 Gt, with the largest single waste sectors being the industrial non-metallic minerals (53%), tailings from ore processing (25%), biomass (15%), ashes (4%), metals (2%) and fossil materials (1%). The global circularity is currently at 6,9% and annually decreasing, as reported by Circle Economy (2025), suggesting that the rate of integrating secondary raw materials in the global economy is not exceeding the current economic growth, hence the linear use of primary materials; “take-make-waste”, is still the paradigm for enabling the socio-economic growth and high living standards.

1.2 Problem description

Industrial solid wastes (ISW) are commonly of inorganic composition, making permanent occupation of large areas of land inevitable (Guo et al., 2024). Incorrect handling of this accumulated waste leads to the surrounding environments receiving air, water and soil pollution by hazardous substances such as heavy metals, acidic or alkaline runoff and dispersed dust and gases (Liu et al., 2016).

While landfill waste contributes to environmental pollution, material consumption continues and requires further extraction from virgin land. Raw material extraction, originating mostly from mining processes, is a transformative activity which has a negative environmental impact as landscapes are permanently altered (Badakhshan et al., 2023). This increasing rate of extraction leads to further air, water and soil pollution, as a result of the inevitable extraction processes but also as a result of globally decreasing ore grades. While Rötzer & Schmidt (2018) claim that decreasing ore grades isn't necessarily a cause for concern in terms of global resource depletion, as technological improvements have made lower ore grades financially viable resources, they state it is a cause for concern that these technical improvements come with increased energy use, emissions and land requirements which lead to further ecological expenditure. It has been quantified that e.g. decreasing global copper ore grades correlates with increases in energy consumption and total material production (Calvo et al., 2016).

It is well documented that material consumption has an adverse impact on the environment, with material handling and use accounting for as much as 70% of global greenhouse gas (GHG) emissions (Circle Economy, 2021), and material extraction and use driving more than 90% of biodiversity loss and water stress (United Nations, 2020). Additionally, highly regionalised extraction and processing of metals and minerals create supply problems as many of these are irreplaceable for the green and digital transitions (European Commision, 2023). Due to this, the EU introduced a regulation on critical raw materials (Regulation 2024/1252, 2024), stating that raw materials should be assessed and given strategic

priority by their corresponding supply risk and economical importance. Examples of the vulnerability of these raw materials can be seen in the price fluctuations and supply chain disruptions observed during recent crises, such as the COVID-19 pandemic, Russia-Ukraine war and US-China trade tensions (Li et al., 2023).

To address these issues, an increased effort in developing strategies for efficient material usage is vital—as the relative input rate of primary raw materials in the global economy needs to decrease.

1.3 Literature review

Rethinking the linear approach to material usage has been described as a transition to a circular economy (CE). There are many aspects of this transition, with no end of opinions as to which approaches are the most effective.

While the CE concept has been considered a new and fashionable approach to the global economy and business models, the concept is not new at all, as Reike et al. (2018) argues that pre-industrial societies largely functioned on circular principles. Although developing industrial societies disregarded such principles, a systems perspective to material usage was successively developed from the 1970s, with the publishing of the first edition of Limits to Growth, with different terms such as: circular economy, industrial ecology, industrial symbiosis, closed-loop, cradle-to-cradle, eco-industrial park, cleaner production etc. (Reike et al., 2018). Unfortunately, a lack of clarity in defining this CE has been accompanied by some confusion across different actors in science and practice, as Kirchherr et al. (2017) highlight with a meta-analysis of CE definitions and consider whether this has contributed to the lack of implementation. Nevertheless, a circular economy can be described concisely by a set of tenets, such as "reduce, reuse, recycle, recover", or by a more thorough definition, as defined by Kirchherr et al. (2017):

"A circular economy describes an economic system that is based on business models which replace the 'end-of-life' concept with reducing, alternatively reusing, recycling and recovering materials in production/distribution and consumption processes, thus operating at the micro level (products, companies, consumers), meso level in (eco-industrial parks) and macro level (city, region, nation and beyond), with the aim to accomplish sustainable development, which implies creating environmental quality, economic prosperity and social equity, to the benefit of current and future generations."

In this thesis, the concept of using any method to re-introduce waste into the market as secondary raw materials, will be referred to as a CE principle.

These principles have influenced environmental policy, as noted in the EU Waste Framework Directive, which defines the Unions waste hierarchy as "prevention, preparing for re-use, recycling, recovery and disposal" (Directive 98/2008, 2018). This directive laid the regulatory foundation for waste management in the EU and additionally, since 2015, numerous revisions of a circular economy action plan have been created, which further emphasises CE principles (European Commision, 2020). The EU has reported a total circular material use rate (CMUR) of 11.8%, with a total increase of 1,4% during the last 15 years (Eurostat, 2024), and as the Union has defined a goal of doubling the CMUR from 2020 to 2030, it would entail an annual CMUR growth rate in the coming years which is 20 timer higher as compared to the last decade (ReBuilt, 2024).

To fulfil the definitions of the CE, the conversion levels of waste into secondary raw materials (SRM), to be used in industrial sectors must be raised as argued by Chioatto & Sospiro (2022). These are major challenges, as new modes of doing business are required, but Chioatto & Sospiro (2022) explain that integrating SRMs into production chains comes with multiple benefits: reduced environmental impacts, reduced costs of production and waste disposal, lowered supply risks of raw materials and decreased vulnerability to global price oscillations, hence should be considered an effective approach for the defined problem.

Utilizing waste as secondary raw material stocks has already taken place, as recovery technologies have been developed to extract valuable metals from specific waste streams, while also finding uses for the residual matrices, such as aggregates in civil engineering applications (Reuter et al., 2004; Spooren et al., 2020). This evolution has been driven by the stated supply risks and costs of specific metals, making it financially viable to valorise these waste streams (Spooren et al., 2020). This type of recovery does not only apply to newly-generated waste, rather it enables urban and industrial landfills to be considered as material resources (Quaghebeur et al., 2013). As reported by Jones et al. (2013), Europe has between 150,000–500,000 historic and active landfills. As the easy-access metal deposits have mostly been exhausted, this kind of prospecting and recovery of ISW represent an entirely new business model. This model not only fulfils the principles of CE, but is also becoming economically competitive to primary resources (Spooren et al., 2020). While these businesses are at the frontier of the CE sector, it requires a highly flexible strategy to process input materials with variable compositions and properties, rendering it a challenging effort to implement in practice (Spooren et al., 2020).

Data-driven methods could be used to optimise resource efficiency of industrial systems**.** While the fields of machine learning and data science have grown exponentially across sectors in recent years, gaps remain in data-driven methods for the CE, as noted by Tseng et al. (2018), who discuss the potential for optimisation models to provide decision-making support for CE businesses. The outcomes of analysis, prediction, planning and automation is highly increased by using data-driven methods, detailed by Liu et al. (2016), and numerous examples can be found in the CE and waste management literature**.** A sensor-based classification system was used to detect when recycling containers needed emptying (Rutqvist et al., 2020). City waste management performance was modeled by socio-economic development indices using regression models (Velis et al., 2023). An optimisation system was created for the re-manufacturing of recycled fibers from composite waste (Lopez-Garcia et al., 2022). End-of-life textile collection from households was modeled to increase the recycling potential of textile fibers (Jäämaa & Kaipia, 2022). The effect of institutional pressures to drive the development

of CE in firms was analysed with a combination of regression and neural networks (Arranz et al., 2022). Product life cycles can be designed using data classification for increasing the amount of returned products (Walter et al., 2023). Predicting the material properties of bituminous mixtures containing construction and demolition wastes and reclaimed asphalt pavement was undertaken by Rondinella et al. (2024), resulting in less need for experimental campaigns when using recyclates as feedstock.

As mentioned in the above, there is an abundance of different applications of data-driven methods connected to CE practices. Nevertheless, an important step in making smarter waste management decisions is undoubtedly to gain more knowledge of the chemical composition of the waste (Seyoum & Adeloju, 2008). Waste characterization has indeed been accepted as a requirement when creating industrial waste management plans (Misra & Pandey, 2005). In light of what has been discussed, waste characterization can be regarded as important for reducing the environmental impact of landfilled waste and natural resource extraction, identifying valuable metals for recovery and for finding use-cases for the resulting waste residuals.

1.4 Aim

The aim of this thesis is to develop a methodology that enables better decision-making regarding waste recovery development. By using data-driven methods, it aims to be agnostic to waste types, enabling rapid learning of waste properties even when previous knowledge is limited. The aim is to ease the ability to predict, while reducing the variance of those properties as compared to using basic statistical methods.

1.5 Research question

How can data-driven methods help us better identify high potential mineral waste streams and improve the recovery of metals and minerals?

2 Method

2.1 Delimitation

This paper is based on an exploratory project that started with few restrictions. The initial scope, in terms of which waste types to consider, was solid wastes from large volume-generating industrial sectors, such as metallurgical processing, paper and pulp, mining and energy production in the European region, with a special interest in the Nordics. Waste types from these sectors include slags, sludges, ashes and tailings. Initially, waste properties would include oxide composition, mineralogical composition, physical properties and various classifications; hazardous or non-hazardous and disposal or recovery. The scope of the methodology was to model these properties on a sector-level basis to be able to make generalisations of a waste originating from numerous facilities.

When applying a wide scope of waste types, other requirements were needed as a guiding light, and this was data quantity and quality. It was decided that enough quantity needs to be gathered, to be able to split the collected data into 2 parts, the first part being used to derive a method for analysis, and the other part for validation of the method. Furthermore, the data needed to be of sufficient quality that no extrapolation of missing values as input data for any modeling would be needed.

It was apparent that many industrial solid wastes have been thoroughly analysed, but many of the literature reviews included too little data for the stated requirements, as even concatenation of tables from multiple articles or literature reviews did not generate enough quantity. Considering the time-frame of the thesis, a decision was made to do a case study of ladle furnace (LF) slags, the refining process of steelmaking (secondary metallurgy). This was partly due to the discovery of a literature review with a relatively large amount of available data from Ouffa et al. (2024). This was chosen as it would enable the data-driven aspect of the methodology and partly as LF

slag had been described in literature as difficult to generalise due to it's varying chemical composition and as a result often disposed of in landfills (Jacob, 2024). As mentioned in the aim, an important condition for developing the methodology was that it must be applicable to waste even when previous knowledge is low. That condition was met as well for LF slag, as both the LF process and it's corresponding slag was new to the author. LF slag as a study object had a suitable balance of meeting the stated requirements but also serving a purpose for improving the knowledge of a waste that currently has low recovery rates.

The main region of interest was Europe, but as the requirements made the project dependent on collecting a certain quantity of data, samples originating from other regions were included as well.

Following the data collection, the selected literature review contained measurements in necessary quantity and quality for the 5 major oxides; CaO, SiO2, Al2O3, Fe2O3 and MgO, so the final scope was to develop the analysis method for the main oxides composition of LF slag.

2.2 Data collection

Scientific articles from scientific databases were analysed, such as Google Scholar and SCOPUS, with search queries including variations of boolean operators, e.g: ("material composition" OR "material characterisation") OR ("chemical characterisation" OR "chemical composition") AND "industrial solid waste" AND ("Ladle furnace" OR "LF") AND "slag" AND "review".

Tabulated data from selected articles was collected with the open-source pdf-extraction library 'gmft' for Python (Wei, 2024). This was done by extracting tables out of pdf files and directly transforming them to Pandas dataframes (tables) for which an extensive collection of data operations is available.

The collected data used for analysis had a total of 113 measured samples, where only the 5 major oxides, CaO, SiO2, Al2O3, Fe2O3

and MgO were measured across all of them. While it can be considered a small dataset, it was assumed big enough to be able to develop and test the methodology.

Exact data on generated quantity of LF slag was not available in the collected dataset, so general steel-producing data was gathered from reports by steel organisations, which specify the annual production of steel in Europe and the total number of producers. Estimations on generation of LF slag per tonne of produced steel was taken from scientific articles.

2.3 Data analysis

All analysis was performed using Jupyter notebooks with the programming language Python using its data science libraries: Pandas, Numpy, Scikit-learn, Seaborn and Matplotlib, for which detailed documentation is available.

To gain initial understanding of the data, a correlation matrix based on the Pearson-correlation coefficient was used for identifying linear relationships between oxides.

$$\boldsymbol{\rho_{X,Y}} = \frac{\boldsymbol{cov(X,Y)}}{\boldsymbol{\sigma_X \sigma_Y}} \tag{1}$$

Where,
$\boldsymbol{\rho_{X,Y}}$ = Correlation of variables X, Y
$\boldsymbol{cov(X,Y)}$ = Covariance of variables X, Y
$\boldsymbol{\sigma_X}$ = Standard deviation of variable X
$\boldsymbol{\sigma_Y}$ = Standard deviation of variable Y

Covariance was calculated as:

$$\boldsymbol{cov(X,Y)} = \frac{\boldsymbol{1}}{\boldsymbol{n-1}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\boldsymbol{x_i - \bar{x}})(\boldsymbol{y_i - \bar{y}}) \tag{2}$$

Where,
$\boldsymbol{x_i}$ = Value of variable $\boldsymbol{X}$ for sample $\boldsymbol{i}$
$\boldsymbol{y_i}$ = Value of variable $\boldsymbol{Y}$ for sample $\boldsymbol{i}$
$\boldsymbol{\bar{x}}$ = Mean of variable $\boldsymbol{X}$
$\boldsymbol{\bar{y}}$ = Mean of variable $\boldsymbol{Y}$
$\boldsymbol{n}$ = Number of samples

Standard deviation was calculated as:

$$\sigma_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (3)$$

Where,
$\boldsymbol{x_i}$ = Value of variable $\boldsymbol{X}$ for sample $\boldsymbol{i}$
$\boldsymbol{\bar{x}}$ = Mean of variable $\boldsymbol{X}$
$\boldsymbol{n}$ = Number of samples

Later, during the literature review of the study object, this matrix was referred as being able to confirm any assumptions about the relationships of the oxides in the LF slag.

All input vectors for modelling were normalised with standard scaling, so the vectors have a mean of 0 and std. dev of 1.

$$x' = \frac{x - \mu}{\sigma} \quad (4)$$

Where,
x' = Normalised vector
x = Vector
μ = Mean of vector
σ = Standard deviation of vector

2.3.1 Unsupervised learning

The parameter combinations from the dataset showing the strongest (negative or positive) correlations were used to classify the data using unsupervised learning. For this, both K-means and DBscan

clustering algorithms from the scikit library were used depending on the specific parameter combination. The euclidean distance was used as the distance metric, and the most optimal configuration was chosen by selecting the K-value for the K-means by the elbow method, i.e finding the smallest variance within clusters, and for DBscan by looping over the hyperparameters of the epsilon (distance constraint) and number of samples. The performance of the clustering algorithms was iteratively optimised by confirming that the clusters were visually distinct, and that the mean silhouette score of the clusters, ranging from -1 to 1, was at least moderate.

$$\boldsymbol{S_i} = \frac{\boldsymbol{b_i} - \boldsymbol{a_i}}{\boldsymbol{max}(\boldsymbol{a_i}, \boldsymbol{b_i})} \quad (5)$$

Where,
a_i = Mean distance between sample $\boldsymbol{i}$ and all other points in the same cluster.
b_i = Mean distance between sample $\boldsymbol{i}$ and all other points in the nearest cluster the sample is not a part of.

2.3.2 Connect with domain

A brief review of the study object (LF slag) was done to gain a basic understanding of it's origins and purpose in the industrial process that creates it.

Distinctive clusters were analysed by looking for explanations within the domain literature. If such explanations could be found, a generalisation that any sample of LF slag belonging in either one cluster or the other is highly probable, and can be labeled as such.

When the samples had been assigned labels, the domain literature was studied to find other process parameters which could be used as input for the model. The main criteria for any such process parameters was that it should be easily obtained from a facility of interest , either by public data, or through correspondence with the facility. While this was a defined criterion, it was not possible to confirm whether the chosen parameters actually can be easily

obtained from a facility, rather a general assumption had to be relied upon. Furthermore, any derived process parameter also needed to have some correlation to the oxide parameters, and was confirmed by computing the correlation between this and the oxide parameters. If the correlation between a process parameter and an oxide was moderate-strong (> ~0.5), it could be assumed that including this process parameter as input to the model would increase the predictive power for the oxide in question.

2.3.3 Supervised learning

The cluster labels and other easily obtained process parameters were used as input parameters to a K Nearest Neighbour (KNN) regression model, using the scikit library's functions, so that any oxide concentration could be predicted based on some input. The dataset was split into a training set (80%) and a test set (20%) to make sure that the model was able predict well on unseen data. Cross-validation of the training data was performed with 5 splits to tune the hyperparameters (amount of K neighbours and distance metric) of the model for optimal performance. The best input parameters for each oxide was found by iteratively testing different configurations. The best configuration of hyperparameters from the cross-validation results was used to train the model once more and final validation was done using the test set (remaining 20%). The model's performance was measured as the variance reduction of the predictions versus the mean of the full dataset's actual values. Performance scoring was selected as root mean square error (RMSE) and R2 score.

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{y_i} - \mathbf{y}_i)^2}{\boldsymbol{n}}} \tag{6}$$

Where:
$\widehat{y_i}$ = predicted value of variable y for sample i
y_i = true value of variable y for sample i
n = Number of samples

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (7)$$

Where:
SS_{res} = Sum of squared residuals
SS_{tot} = Total sum of squares

$$SS_{res} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \quad (8)$$

Where,
$\hat{y}_i$ = Predicted value of variable y for sample i
y_i = True value of variable y for sample i
n = Number of samples

$$SS_{tot} = \sum_{i=1}^{n} (\bar{y} - y_i)^2 \quad (9)$$

Where,
$\bar{y}$ = Mean value of variable y
y_i = Actual value of variable y for sample i
n = Number of samples

To assess the model's predictions compared to that of the full dataset's variance, the RMSE was normalised ($nRMSE$) by the standard deviation of the full dataset.

$$nRMSE = \frac{RMSE}{\sigma} \quad (10)$$

Where:
$RMSE$ = Root mean square error of the model's predictions
σ = Standard deviation of the full dataset

2.3.4 Evaluating model by scenarios

Hypothetical scenarios were used to assess strengths and weaknesses, while also highlighting a use case for the model.
A chosen amount of annually generated LF slag was assigned to each sample in the collected dataset. This was done as a rounded conservative annual estimate of 10 000 tons of generated LF slag per steel producing facility in Europe. In these scenarios, each sample represented the LF slag from a facility.

$$\boldsymbol{Slag_Q = Slag_P * \frac{Steel_P}{n}} \quad (11)$$

Where:
$\boldsymbol{Slag_Q}$ = Quantity of produced slag per facility (tonnes)
$\boldsymbol{Slag_P}$ = Amount of produced slag per tonne produced steel (kg/tonne). Estimated as 25 kg/tonne steel, aggregated from Ouffa et al. (2024)
$\boldsymbol{Steel_P}$ = Total annual steel production in EU, 126 million tonnes (EUROFER, 2024).
$\boldsymbol{n}$ = Number of steel producing facilities in EU (500 facilities) (European Commission, 2014).

Assessing strengths and weaknesses of the model based on a single scenario of the test dataset, containing 23 samples, was reasoned to give highly uncertain results. To eliminate this problem as much as possible, the method of bootstrapping was used to generate new datasets based on the test data.

2 hypothetical scenarios were constructed by bootstrapping samples from the test dataset, with 10 and 100 sample size datasets respectively and the scenarios were evaluated with a loop of 10 000 iterations. For each iteration, the samples were summed to get an simulate the available oxide quantity from every dataset, i.e from all included facilities.

The evaluation assessed the errors of the predicted values to the true values compared to a base case approach (using statistical mean from the training data). This was done by calculating the mean absolute error (MAE) per sample's oxide concentration and per total quantity of oxides.

$$\boldsymbol{MAE = \left| \frac{\sum_{i=1}^{n} \hat{y}_i - y_i}{n} \right|} \quad (12)$$

Where:
$\boldsymbol{y_i}$ = True value
$\boldsymbol{\hat{y}_i}$ = Either predicted value (model) or mean value (base case)
$\boldsymbol{n}$ = Number of observations

The total quantity was calculated for each iteration and case: predicted, base case and true, and the mean and σ of the quantity error rate of all iterations was calculated.

$$\boldsymbol{Quantity\ error\ rate} = \left(\mathbf{1} - \left(\frac{\boldsymbol{Current\ quantity}}{\boldsymbol{True\ quantity}}\right)\right) * \mathbf{100} \quad (13)$$

Where:
$\boldsymbol{Current\ quantity}$ = Either predicted quantity or the base case quantity (tonnes).
$\boldsymbol{True\ quantity}$ = Quantity from the true vector for each iteration (tonnes).

Finally, the accuracy of the model was specified with a counter for the amount of iterations the predicted quantity was closer to the true quantity than the base case quantity, and was calculated as % of iterations.

$$\boldsymbol{Counter} = \left(\mathbf{1} - \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i| < |\mu_{quantity} - y_i|}{\boldsymbol{n}}\right) * \mathbf{100} \quad (14)$$

Where,
$\hat{y}_i$ = Predicted quantity (tonnes)
y_i = True quantity (tonnes)
$\mu_{quantity}$ = Base case (mean) quantity (tonnes)
n = number of iterations

The error range for the calculated quantity was set to $2 * \sigma$, as a simple way to capture ~95% of the 10 000 iterations and the error range for the per sample error was set to σ, capturing ~67% of all predictions, with an assumption that the data follows a fairly normal distribution.

3 Results

The correlation matrix in Fig 1. shows a strong negative correlation of -0.77 between Al2O3 and SiO2. Additionally, there is a moderate negative correlation between CaO and Fe2O3 while the rest of the relationships have weak to no correlation.

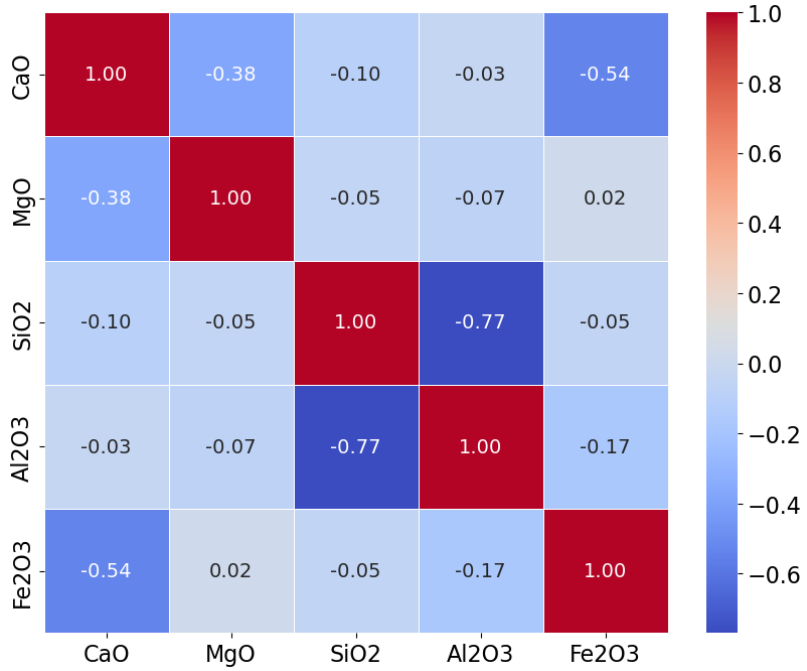

Figure 1. Correlation matrix of the 5 major oxides

3.1 Unsupervised learning

The boundary between the 2 clusters in the SiO2–Al2O3 system in Fig. 2 is arbitrary, where sample points could be assigned to either cluster, but otherwise it is a visually distinct result. This is confirmed with a moderate silhouette score of 0.59.

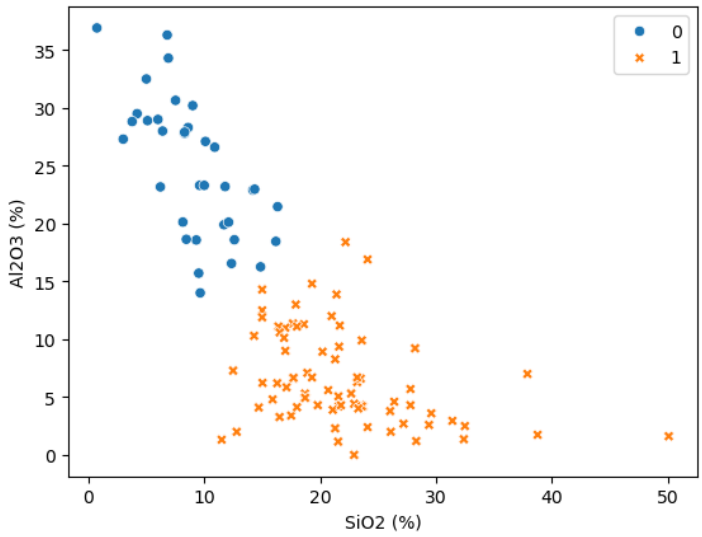

Figure 2. Clustering result of SiO2-Al2O3 parameters. Silhouette score 0.59.

The clustering of CaO-Fe2O3 oxides in Fig. 3 was less distinct than the SiO2-Al2O3 system, with a silhouette score of 0.65. Visually there is one major cluster, labelled 0, with a number of scattered points in the graph labelled -1.

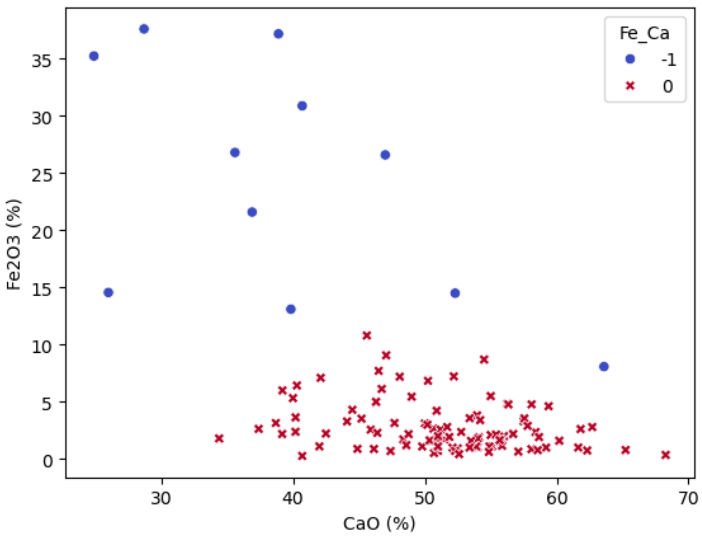

Figure 3. Clustering result of CaO-Fe2O3 parameters. Silhouette score 0.65.

The standard deviations of each parameter in Fig. 4 were reduced by 20-70% for all clusters apart from the High Fe, which can be explained by it consisting of noise points. Even though the result is weak for that particular cluster, it facilitates a reduction in the standard deviation of Fe2O3 by ~70% for the other cluster. There was no combination of parameters that could produce clusters of MgO, so it was not included in the result.

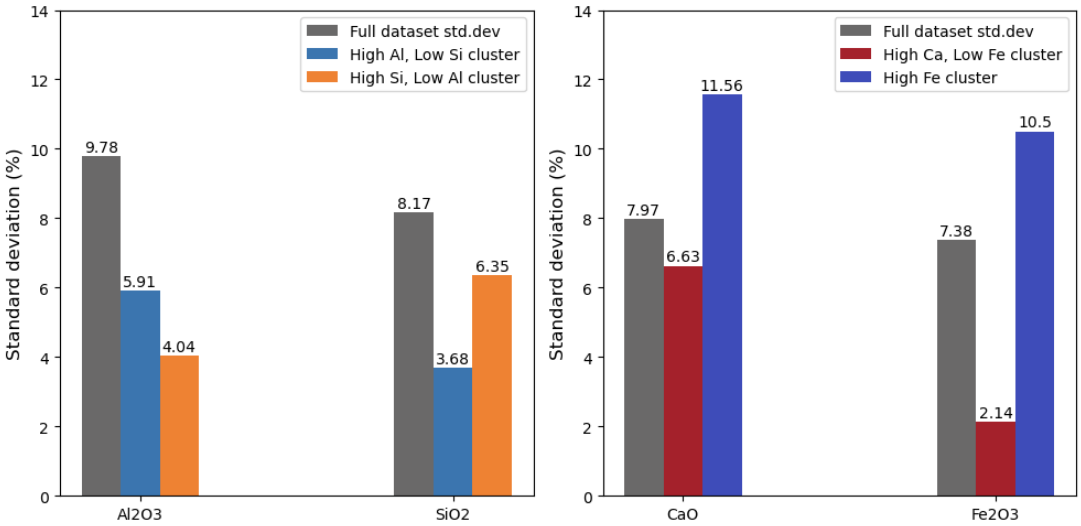

Figure 4. Standard deviations of oxide concentration (%) for selected parameters after clustering.

3.2 Connect with domain

LF slags are by-products of the refining stage of steelmaking, referred to as secondary metallurgy. Prior to this, the steel is created from pig iron in the primary metallurgy stage, which involves either an electric arc furnace (EAF) or a basic oxygen furnace (BOF). After the primary stage, the steel is transported to a ladle furnace and undergoes processes such as composition homogenisation, desulphurisation, deoxidation, degassing of nitrogen and hydrogen, decarburisation and inclusion removal (Ouffa et al., 2024). These processes remove the impurities in the steel by maintaining a floating layer of slag above the molten steel, consisting of a tightly controlled input of fluxes. Alloying elements are then added to the ladle to obtain the specified properties of the final product. After achieving the desired steel quality, the slag is skimmed off the ladle and becomes a waste-product (Jacob, 2024).

As previously mentioned, literature reviews on LF slags state that the chemical composition of these are significantly more varied than other steelmaking slags (Jacob, 2024; Setién et al., 2009). However, the distinct clustering result of the SiO2-Al2O3 parameters can be explained by domain literature as processing plants using either Si- or Al-containing fluxes as reducing agents for the deoxidation of steel in the ladle furnace (Santamaria et al., 2019). Fluxes used for this can be ferrosilicon, silicomanganese or aluminium (Turkdogan, 2010). This process is known as killing the steel, and the product becomes either Al-killed or Si-killed steel (Pretorius, n.d.).

There are also indications in literature that can explain the result between the CaO and Fe2O3 parameters. Large amounts of calcium-containing fluxes is added to the slag to enable the desulphurisation of the steel, making CaO the primary oxide in LF slag (Ouffa et al., 2024). A low Fe content in the slag is preferable, as high concentration of reducible oxides will impact the degree of deulphurisation (Pretorius, n.d.). This can be visualised in Fig. 3 where most of the collected samples belong to the cluster with low Fe2O3 content. However, during latter slag processing stages, Vlcek et al. (2013) state that uncast steel residues may get mixed with the

LF slag, as they are discarded in the same storage heap. When a sample is taken, it is a mixture of both LF slag and uncast steel residues, making the composition of Fe2O3 higher.

In the search for other process parameters to use for the model, the basicity ratio was found. The basicity of LF slag is a key processing parameter, as correctly managed it will reduce the steel's capacity to hold oxygen, sulphur, phosphorus and other minor impurities, while also protecting the refractory walls of the ladle furnace from excessive wear due to MgO dissolving into the slag (Pretorius, n.d.; Setién et al., 2009; Turkdogan, 2010). Two variations of basicity ratios used in the industry are the b5, for reduced slags, or b3, for oxidized slags (Pretorius, n.d.).

$$\boldsymbol{b5} = \frac{\%\boldsymbol{CaO} + \%\boldsymbol{MgO}}{\%\boldsymbol{SiO2} + \%\boldsymbol{Al2O3}} \quad (15)$$

$$\boldsymbol{b3} = \frac{\%\boldsymbol{CaO}}{\%\boldsymbol{SiO2} + \%\boldsymbol{Al2O3}} \quad (16)$$

Where:
$\%\boldsymbol{x}$ = Concentration (%) of oxide x in a sample.

The calculated b5 from the collected dataset has a moderate correlation with CaO and Al2O3, while being weakly correlated with the other oxides.

Table 1. b5 basicity ratio's correlation to main oxides.

	CaO	SiO2	Al2O3	Fe2O3	MgO
b5	0.43	-0.12	-0.46	0.03	0.22

The calculated b3 has a strong correlation to CaO, and no increased correlations to the other oxides compared to the b5 ratio.

Table 2. b3 basicity ratio's correlation to main oxides.

	CaO	SiO2	Al2O3	Fe2O3	MgO
b3	0.61	-0.11	-0.43	-0.03	-0.06

As additional parameters to the cluster labels, the b5 was used for predictions of Al2O3, SiO2, Fe2O3 and MgO, while b3 was used to predict CaO.

3.3 Supervised learning

The KNN regression model's range of R2 scores on the test data was 0.56–0.75 in Table 3, suggesting that it is able to capture over half of the variance in the dataset, and a maximum of 20% difference between all parameters, excluding MgO. The nRMSE values were 0.51–0.63 of the standard deviation, suggesting the error variance for predicted values was 37–49% lower than if predicted by the mean of the full dataset. No improved prediction was gained for MgO, so it was not included in the results. This suggests that the model should have moderate–strong predictive power for 3 out of 5 parameters (Al2O3, SiO2 and Fe2O3), moderate for CaO, and no predictive power for MgO.

Table 3. KNN regression model results

	Test R2	nRMSE	Test RMSE	Full Dataset std.dev	Full Dataset mean
Al2O3	0.71	0.51	4.96	9.78	12.38
SiO2	0.65	0.51	4.16	8.17	17.7
CaO	0.56	0.63	5.02	7.97	50.01
Fe2O3	0.75	0.63	4.63	7.38	4.89

3.4 Evaluating model by scenarios

3.4.1 Sample error

On a sample basis over 10 000 iterations, the mean absolute error of the oxide concentration is smaller for the model across all included parameters, with mean predictions being 37–45% closer to the true values than the base case in Fig. 5. The variance reduction is 23–54% for all parameters, with the smallest reduction for CaO, and largest for Al2O3 and Fe2O3.

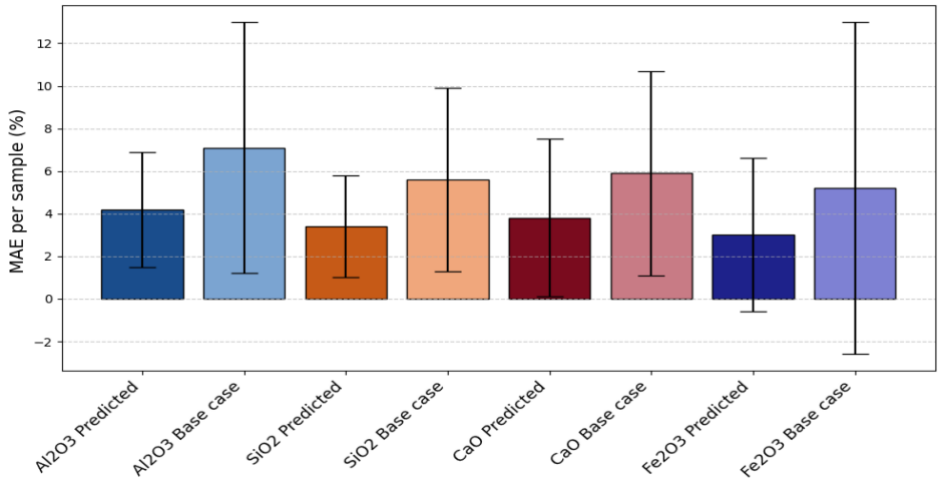

Figure 5. Mean absolute error per sample (%). Error bars are 1*std.dev, covering ~67% of the samples.

3.4.2 10-sample scenario

The variance of the quantity error is reduced by 32–56% for all 4 parameters when using the model in Fig. 6, with the largest variance reduction for Al2O3 and Fe2O3, and smallest for CaO. In a real scenario where the true quantity is unknown, a realistic estimate of oxide quantity for potential recovery is the error margin.

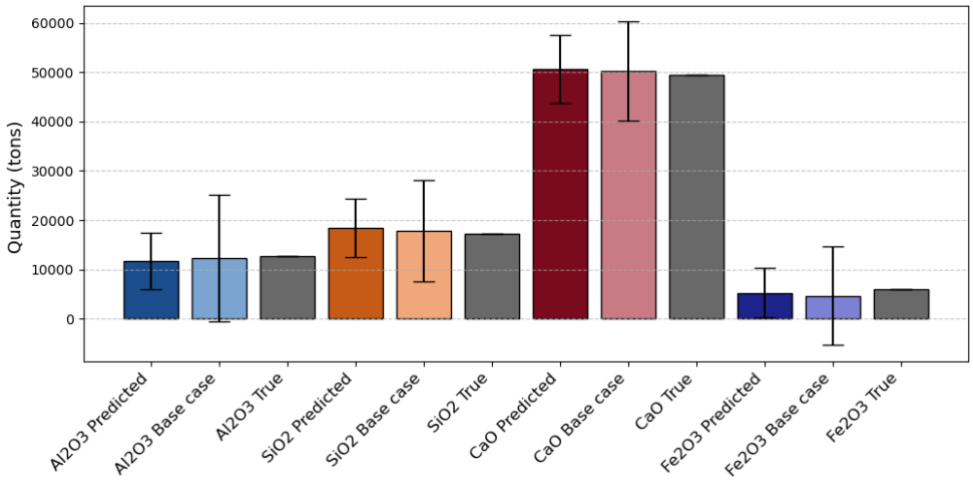

Figure 6. Estimation of oxide quantity from a 10 sample scenario, performed with 10 000 iterations. Error bars are 2*std.dev, covering 95% of the iterations.

3.4.3 100-sample scenario

The variance reduction of the quantity error is reduced by 32–50% for all 4 parameters using the model in Fig. 7, with the largest reduction for Al2O3 and Fe2O3, and smallest for CaO. Again, in a real scenario where the true quantity is unknown, a realistic estimate of oxide quantity for potential recovery is the error margin.

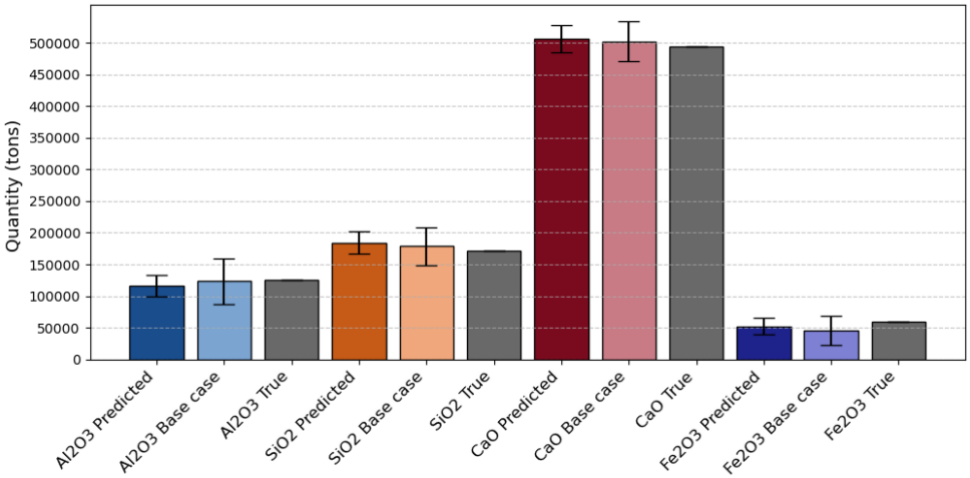

Figure 7. Estimation of oxide quantity from a 100 sample scenario, performed with 10 000 iterations. Error bars are 2*std.dev, covering 95% of the iterations.

3.4.4 Model accuracy

For the 10-sample scenarios in Fig. 8, the number of iterations the model predicts more accurately than the base case varies between 55–75% of the iterations for all 4 parameters. For the 100-sample scenarios, the base case is frequently more accurate than the model, except for the Fe2O3 parameter, where the model predicts more accurately for both sample sizes.

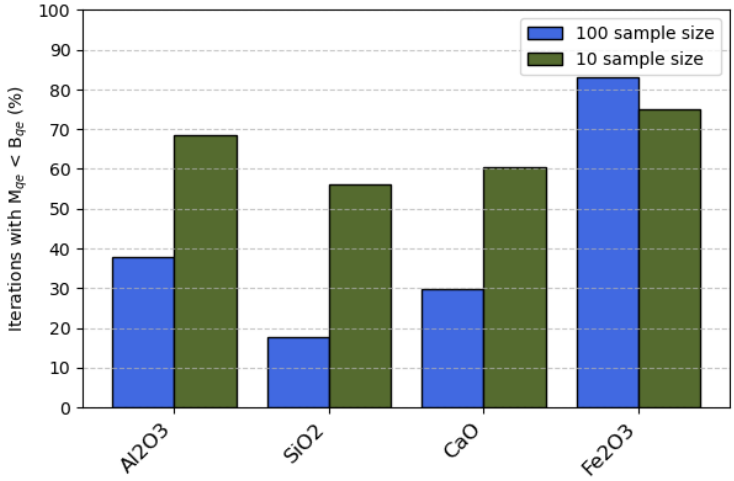

Figure 8. Number of iterations (%) where the model quantity error (M_{qe}) is smaller than the base case quantity error (B_{qe}).

4 Discussion

4.1 Result Discussion

As the strongest correlation between SiO2 and Al2O3 also corresponds to the strongest clustering result, the correlation matrix can be used as an initial rough proxy for how well the cluster algorithms separate the data into groups, thereby enabling predictions. This is likely to be applicable for other waste types as well. Although it is limited in this method to only explore linear correlations, other interactions could be explored, e.g. by computing the distance correlation between parameters (Edelmann et al., 2021).

The unsupervised learning method, which put samples into clusters, found structure in the dataset for 4 out of the 5 parameters, and reduced the variance for those, as can be seen in Fig. 4. Confirming the clusters in the domain literature, i.e different reducing agents (Al2O3 or SiO2) for steel refining show that the method is effective in increasing the learning rate of a waste properties. Instead of thoroughly studying the domain literature in the hopes of finding something which might explain the collected data, one can first apply the clustering, observe the results, and then look for confirmation of that result in the literature. Although it was effective for explaining the relationships between the parameters Al2O3, SiO2 and CaO, Fe2O3, there was no combination of parameters together with MgO that produced any reliable result, which indicates a limitation of the method.

The results of tuning the regression model, as shown in Table 3, suggest it is able to capture structure in the dataset resulting in reduced variance for predictions compared to using statistical means. The lack of any predictive power for MgO further confirms that there is a weakness for parameters of which no strong correlations can be found.

The MgO concentration is likely based on the varying use of dolomite, a commonly used flux in the steelmaking process (Turkdogan, 2010). As the dolomite mineral requires unusual

conditions to form, as discussed by (Warren, 2000), the usage of dolomite for steelmaking is likely dependent on it being regionally mined in the required quantities. A geography parameter is not included in this version of the model, therefore the MgO variance is not being accounted for, but could possibly be so in future work.

The results from the per sample errors and the 10-sample scenario can be interpreted as the model being both more precise, in Fig. 6, and more accurate in Fig. 8. The result from the 100-sample scenario was rather more complex, with the model reducing quantity error variance in Fig 7. while also being more frequently wrong in Fig 8. This suggests that the model was more precise but less accurate than the base case, as the base case errors were likely of greater magnitude when they occurred. Though the model was less accurate in this scenario, its performance can still be interpreted as good, since consistency is an indicator of a well-performing model (Levman et al., 2023). This signals that continued optimisation is possible to improve the accuracy.

4.2 Method Discussion

A wide scope at the start of this thesis was suitable as the methodology aims to be applicable across waste types, also for cases when previous knowledge of the waste is low. Restrictions at the start could have created arbitrary boundaries, and as a result the LF slag literature review, or other potential study objects might not have been found. While it introduced some uncertainty in the beginning of the thesis work, it was recovered in later stages when the selection of a case study, with the appropriate data quantity and quality did not need to be questioned for validity.

The oxide composition is not the only important material characteristic to specify for increasing the recovery of waste. Mineralogical composition has a large impact on the material properties of wastes and is important information when finding valid use cases for them (Yildirim & Prezzi, 2011). Understanding the mineralogical composition enables the selection of the optimal

technology to extract the valuable elements from the waste's mineral structures (Wu et al., 2022). While it is a shortcoming to only model oxide composition, mineralogical composition is usually harder to quantify, as it is possible to generalise that the amount of mineral combinations is greater than the amount of different oxides found within the waste. An example of this is that over 50 minerals have been identifed in LF slags, detailed by Ouffa et al. (2024), the reason being that the mineralogical composition is also affected by different cooling and weathering conditions after the slag has been discarded. Nevertheless, an analysis method could still be developed to find levels of occurrences of minerals within a specific waste, but the timeframe of this thesis dictated that such an analysis could not be included in the scope.

The data collection method helped during the early stage when the delimitation was broad, as it allowed to quickly gather data from many sources to get an overview of which waste types would meet the data quantity and quality requirements. Although it worked well, adjustments were needed to get the tables true to their sources, suggesting that future work should put emphasis on even better data collection methods. Compared to other methods such as importing pdf to Excel, creating tables with string formatting or GPTs, the selected method was considered reasonably time-efficient and functional.

The clustering method only uses 2 dimensions, after iterating and not finding any increases in performance when including more dimensions. In future analysis, principal component analysis (PCA) could be used to incorporate more parameters and consequently reduce dimensions (Salih Hasan et al., 2021). There is also a great variety of other unsupervised learning methods, apart from K-means and DBscan, which aims to find structure from unlabelled data (Naeem et al., 2023). The only motivation for why the selected algorithms were used instead of others is based on previous knowledge and experience.

Careful consideration is necessary when using this methodology for other waste types. The selection of parameters for finding clusters should always be done with a domain investigation whether such cluster labels could actually be inferred without having access to analytically measured data. A strong clustering result that lacks a realistic probability to correctly assign such a label to an unmeasured sample can be considered largely useless, as the sample would have to be analytically measured before any label is assigned and thereby eliminating the need for using the methodology. This is an inherently crucial step, and to some degree limits the rapid utilisation of the method across waste types.

The basicity ratios used are not definitive, as Turkdogan (2010) explains that different basicity ratios are used for different slags in the industry. A more detailed analysis is required to identify the most general ratio which is applicable to the majority of LF slags. However, the selected basicity ratios could be calculated from the collected dataset. An assumption was made that such ratios could be calculated e.g from import/export data, or that a facility, to a satisfactory degree, (1 decimal point), could specify what basicity ratio their slag has.

The choice of using a KNN model over other regression models is not definitive either and could be changed in future work. A KNN model was chosen as it is relatively simple yet efficient, being a non-parametric model that, when properly tuned, will ignore large outliers, as observed in the data (see section 3.3), and will only consider the K closest samples for each prediction. Although this means that the model must calculate the distance for every sample combination and with larger amounts of data will be computationally slower than other models (Taunk et al., 2019).

Evaluating the model by simulating scenarios should be considered as hypothetical, and serves a purpose most importantly by assessing the observed error margins of the model. The motivation behind this evaluation is that in a real scenario, the true values that the model is trying to predict are unknown, and any valid result needs to be

reported as a range of probable values. Understanding the size and frequency of the model's errors over a large amount of iterations makes it possible to specify this range of probable values. Although the simulation creates new data for every iteration, each generated dataset originates from the 23 samples of the test set, so while it reduces uncertainty by evaluating over many iterations, it can not be completely representative of a bigger sample size. Showing the available quantity of oxides in LF slag from the generated datasets was carried out to highlight a potential use case of the methodology, and does not serve a purpose beyond that as the numbers are hypothetical.

A weakness in the analysis method is that the model has only been tested by either labelled or calculated data from the collected samples' analytically measured oxide composition. As the real value lies in estimating sources of unknown composition by extrapolation of data, e.g using available public data or correspondence with facilities, which has not been done, the true performance of this method in an real scenario is largely unknown. In this thesis, it is assumed that this extrapolation of data is possible, but such an assumption might not be valid.

This method strictly evaluates performance of the model with equal quantity of waste from each source, 10 000 tonnes/year, but the true estimated quantity of LF slag is a widely discussed topic in literature, as noted by Ouffa et al. (2024). In an application context, the variance reduction in oxide composition gained from this methodology should be coupled with a method for variance reduction in available quantity. If this is not done, the available quantity becomes the limiting factor, and the benefit from reducing variance using this methodology is lessened.

4.3 Conclusion

The results show that the methodology reduces the variance in predicted oxide concentrations of LF slag, resulting in more precise estimates of total quantity of oxides for a given waste type. This shows that the methodology can be used to better identify high potential mineral waste streams and increase the possibilities for improved metal and mineral recovery.

4.4 Perspective

The methodology could be utilised with different perspectives for stakeholders in the circular economy. One use-case could be to evaluate feedstocks for a specific recovery technology. Another use-case could be prospecting, to predict valuable or strategic stocks and flows of secondary raw materials, either for supporting the development of recovery technology or for applying off-the-shelf technology. It could also be used to assess suitable locations for establishing recovery facilities. Continued research with this methodology could therefore be to use a real scenario, e.g. define a suitable feedstock, assess its stocks and flows within a geographical boundary and evaluate whether and where there is a business case for a recovery facility within that boundary. Specific emphasis should be placed on the extrapolation of data to infer the cluster labels without using analytically measured samples. Similarly, using this methodology to optimise predictions of waste quantity, rather than oxide concentration could also be attempted. Expanding the scope of the methodology to include more properties (e.g more elements, mineralogical composition or physical properties) could be another future topic.

A big challenge during this thesis has been to collect data of sufficient quantity and quality. If it was a challenge for a relatively small project like this, it is highly likely to be so for the CE industry as a whole. As discussed by Tseng et al. (2018), a properly functioning CE system includes levels of interdependence between stakeholders that can only be optimised by having access to the right data, and the lack of it will amplify a magnitude of disruptions in

such highly integrated systems. Considering the lack of increase in the CMUR in the EU over the last 15 years, even after the Union's defined goals in the circular economy action plan by the European Commision (2020), it is highly credible that the necessary levels of data sharing between stakeholders in and between value chains is not taking place at scale.

In the Nordics this issue of data sharing has been noticed, as Nordic Innovation (2021) defines 4 groups of barriers as: value-based, technical, legislative, and knowledge-lacking, while also stating that many companies are apprehensive towards data sharing due to the current regulations, data security, risk of manipulation or misrepresentation of data and vulnerability towards competitors. As stated in that report, every actor in a network not participating in data sharing reduces the potential for all other actors in that network to maximise value from circular processes. With this in mind, future research could examine how to overcome these barriers and what frameworks to be used to increase data sharing within and across value chains and industrial sectors. A prototype for a Nordic data sharing platform for primary and secondary mineral resources has been developed by Feltrin et al. (2024) and a similar structure could be a suitable starting point for such research. Another could be to assess the potential of federated learning in the CE, as mentioned by Lin & Wei (2023).

While this methodology's scope was to model waste properties on a sector level, facility-level modeling has also been done by other researchers to predict composition and quantity of produced slag by using more specific process parameters as input (Harada et al., 2013; Petrucciani et al., 2022). This is closely related to the concept of digital twins (Chen & Huang, 2021). It is uncertain whether such modeling could be scaled up and be better at predicting sector-level composition compared to the methodology of this thesis. This could be a topic for future research; building facility-level models/digital twins and expanding it to include as many facilities from a sector as possible. Again, acquiring high quality data should be at the

forefront in such a study, as pointed out by Petrucciani et al. (2022) and based on the experience gained during this thesis.

While working with data may seem too difficult and unfruitful at times, its potential remains undeniable. As the United Nations (2014) aptly noted: “Data are the lifeblood of decision-making.” Thus, providing the right information on the right things at the right time may lead us to the world we need — one that supports the prosperity of both people and planet.

5 Acknowledgements

I would like to express my sincere gratitude to my supervisors, Hussain Azeez Mohamed, Robert Norgren and Graham Aid at Ragn-Sells, for their guidance, support and feedback throughout the project. This exploratory project has been difficult in many regards, and would not have been possible without the valuable insights from industry professionals.

6 References

Badakhshan, N., Shahriar, K., Afraei, S., & Bakhtavar, E. (2023). Determining the environmental costs of mining projects: A comprehensive quantitative assessment. *Resources Policy*, *82*, 103561. https://doi.org/10.1016/j.resourpol.2023.103561

Calvo, G., Mudd, G., Valero, A., & Valero, A. (2016). Decreasing Ore Grades in Global Metallic Mining: A Theoretical Issue or a Global Reality? *Resources*, *5*(4), Article 4. https://doi.org/10.3390/resources5040036

Chen, Z., & Huang, L. (2021). Digital twins for information-sharing in remanufacturing supply chain: A review. *Energy*, *220*, 119712. https://doi.org/10.1016/j.energy.2020.119712

Chioatto, E., & Sospiro, P. (2022). Transition from waste management to circular economy: The European Union roadmap. *Environment, Development and Sustainability*, *25*, 1–28. https://doi.org/10.1007/s10668-021-02050-3

Circle Economy. (2021). *Circularity Gap Report 2021*. https://www.circularity-gap.world/2021

Circle Economy. (2024). *Circularity Gap Report 2024*. https://www.circularity-gap.world/2024

Circle Economy. (2025, May 19). *Circularity Gap Report 2025*. https://www.circularity-gap.world/2025

Directive 98/2008 (2018). http://data.europa.eu/eli/dir/2008/98/2018-07-05/eng

Edelmann, D., Móri, T. F., & Székely, G. J. (2021). On relationships between the Pearson and the distance correlation coefficients. *Statistics &*

Probability Letters, *169*, 108960. https://doi.org/10.1016/j.spl.2020.108960

EUROFER. (2024). *Annual Report 2024*. https://www.eurofer.eu/publications/reports-or-studies/annual-report-2024

European Commision. (2020). *Circular economy action plan*. https://environment.ec.europa.eu/strategy/circular-economy-action-plan_en

European Commision. (2023). *Critical Raw Materials Act*. https://single-market-economy.ec.europa.eu/sectors/raw-materials/areas-specific-interest/critical-raw-materials/critical-raw-materials-act_en

European Commission. (2014). *EUR-Lex—52014SC0020—EN* [Text/html; charset=UTF-8]. OPOCE. https://publications.europa.eu/resource/cellar/ba385885-8433-11e3-9b7d-01aa75ed71a1.0001.04/DOC_1?utm_source=chatgpt.com

Eurostat. (2024). *Circular material use rate*. https://ec.europa.eu/eurostat/databrowser/view/cei_srm030/default/table?lang=en

F.A. Arranz, C., Sena, V., & Kwong, C. (2022). Institutional pressures as drivers of circular economy in firms: A machine learning approach. *Journal of Cleaner Production*, *355*, 131738. https://doi.org/10.1016/j.jclepro.2022.131738

Feltrin, L., Andersson, M., & Eskil Larsen, B. (2024). *Information Management and Classification: Secondary Resources and their Critical Raw Material Potential in the Nordic Countries*.

Guo, X., Zeng, M., Yu, H., Lin, F., Li, J., Wang, W., & Chen, G. (2024). Critical review for the potential analysis of material utilization from inorganic industrial solid waste. *Journal of Cleaner Production*, *459*, 142457. https://doi.org/10.1016/j.jclepro.2024.142457

Harada, A., Maruoka, N., Shibata, H., & Kitamura, S. (2013). A Kinetic Model to Predict the Compositions of Metal, Slag and Inclusions during Ladle Refining: Part 1. Basic Concept and Application. *ISIJ International*, *53*(12), 2110–2117. https://doi.org/10.2355/isijinternational.53.2110

Jäämaa, L., & Kaipia, R. (2022). The first mile problem in the circular economy supply chains – Collecting recyclable textiles from consumers. *Waste Management*, *141*, 173–182. https://doi.org/10.1016/j.wasman.2022.01.012

Jacob, L. (2024). Ladle Furnace Slag: Synthesis, Properties, and Applications. *ChemBioEng Reviews*, *11*(1), 60–78. https://doi.org/10.1002/cben.202300024

Jones, P. T., Geysen, D., Tielemans, Y., Van Passel, S., Pontikes, Y., Blanpain, B., Quaghebeur, M., & Hoekstra, N. (2013). Enhanced Landfill Mining in view of multiple resource recovery: A critical review. *Journal of Cleaner Production*, *55*, 45–55. https://doi.org/10.1016/j.jclepro.2012.05.021

Kirchherr, J., Reike, D., & Hekkert, M. (2017). Conceptualizing the circular economy: An analysis of 114 definitions. *Resources, Conservation and Recycling*, *127*, 221–232. https://doi.org/10.1016/j.resconrec.2017.09.005

Kose, M. A., & Ozturk, E. O. (n.d.). *Taking stock of the past half century*.

Levman, J., Ewenson, B., Apaloo, J., Berger, D., & Tyrrell, P. N. (2023). Error Consistency for Machine Learning Evaluation and Validation with Application to Biomedical Diagnostics. *Diagnostics*, *13*(7), Article 7. https://doi.org/10.3390/diagnostics13071315

Li, Z.-Z., Meng, Q., Zhang, L., Lobont, O.-R., & Shen, Y. (2023). How do rare earth prices respond to economic and geopolitical factors? *Resources Policy*, *85*, 103853. https://doi.org/10.1016/j.resourpol.2023.103853

Lin, K., & Wei, S. (2023). Advancing the Industrial Circular Economy: The Integrative Role of Machine Learning in Resource Optimization. *Journal of Green Economy and Low-Carbon Development*, *2*(3), 122–136. https://doi.org/10.56578/jgelcd020302

Liu, Y., Guo, D., Dong, L., Xu, Y., & Liu, J. (2016). Pollution Status and Environmental Sound Management (ESM) Trends on Typical General Industrial Solid Waste. *Procedia Environmental Sciences*, *31*, 615–620. https://doi.org/10.1016/j.proenv.2016.02.111

Lopez-Garcia, P., Barrenetxea, X., García-Arrieta, S., Sedano, I., Palenzuela, L., & Usatorre, L. (2022). Compounding process optimization for recycled materials using machine learning algorithms. *Procedia CIRP*, *105*, 237–242. https://doi.org/10.1016/j.procir.2022.02.039

Maalouf, A., & Mavropoulos, A. (2023). Re-assessing global municipal solid waste generation. *Waste Management & Research: The Journal for a Sustainable Circular Economy*, *41*(4), 936–947. https://doi.org/10.1177/0734242X221074116

Misra, V., & Pandey, S. D. (2005). Hazardous waste, impact on health and environment for development of better waste management strategies

in future in India. *Environment International*, *31*(3), 417–431. https://doi.org/10.1016/j.envint.2004.08.005

Naeem, S., Ali, A., Anam, S., & Ahmed, M. (2023). An Unsupervised Machine Learning Algorithms: Comprehensive Review. *IJCDS Journal*, *13*, 911–921. https://doi.org/10.12785/ijcds/130172

Nordic Innovation. (2021). *Data Sharing for a Circular Economy in the Nordics*. Nordic Innovation. https://www.nordicinnovation.org/CEdatasharing

OECD. (2019). *Global Material Resources Outlook to 2060: Economic Drivers and Environmental Consequences*. OECD. https://doi.org/10.1787/9789264307452-en

Ouffa, N., Benzaazoua, M., Trauchessec, R., Belem, T., Taha, Y., & Diliberto, C. (2024). Potential Reuse of Ladle Furnace Slag as Cementitious Material: A Literature Review of Generation, Characterization, and Processing Methods. *Minerals*, *14*(12), Article 12. https://doi.org/10.3390/min14121204

Petrucciani, Alice, Zaccara, Antonella, Matino, Ismael, Colla, Valentina, & Ferrer, Maria. (2022). Flowsheet Model and Simulation of Produced Slag in Electric Steelmaking to Improve Resource Management and Circular Production. *Chemical Engineering Transactions*, *96*, 121–126. https://doi.org/10.3303/CET2296021

Pretorius, E. (n.d.). *FUNDAMENTALS OF EAF AND LADLE SLAGS AND LADLE REFINING PRINCIPLES*. Retrieved 26 May 2025, from https://www.academia.edu/35667652/FUNDAMENTALS_OF_EAF_AND_LADLE_SLAGS_AND_LADLE_REFINING_PRINCIPLES

Quaghebeur, M., Laenen, B., Geysen, D., Nielsen, P., Pontikes, Y., Van Gerven, T., & Spooren, J. (2013). Characterization of landfilled

materials: Screening of the enhanced landfill mining potential. *Journal of Cleaner Production*, *55*, 72–83. https://doi.org/10.1016/j.jclepro.2012.06.012

ReBuilt. (2024). CIRCULAR MATERIAL USE RATE IN EUROPE. *Interreg Central Europe*. https://www.interreg-central.eu/news/circular-material-use-rate-in-europe/

Regulation 2024/1252 (2024). https://eur-lex.europa.eu/eli/reg/2024/1252/oj

Reike, D., Vermeulen, W. J. V., & Witjes, S. (2018). The circular economy: New or Refurbished as CE 3.0? — Exploring Controversies in the Conceptualization of the Circular Economy through a Focus on History and Resource Value Retention Options. *Resources, Conservation and Recycling*, *135*, 246–264. https://doi.org/10.1016/j.resconrec.2017.08.027

Reuter, M., Xiao, Y., & Boin, U. (2004). Recycling and environmental issues of metallurgical slags and salt fluxes. *VII International Conference on Molten Slags Fluxes and Salts*, 349–356.

Ritchie, H., Rodés-Guirao, L., Mathieu, E., Gerber, M., Ortiz-Ospina, E., Hasell, J., & Roser, M. (2023). Population Growth. *Our World in Data*. https://ourworldindata.org/population-growth

Rondinella, F., Oreto, C., Abbondati, F., & Baldo, N. (2024). *A Deep Neural Network Approach towards Performance Prediction of Bituminous Mixtures Produced Using Secondary Raw Materials*. https://doi.org/10.3390/coatings14080922

Rötzer, N., & Schmidt, M. (2018). Decreasing Metal Ore Grades—Is the Fear of Resource Depletion Justified? *Resources*, *7*(4), Article 4. https://doi.org/10.3390/resources7040088

Rutqvist, D., Kleyko, D., & Blomstedt, F. (2020). An Automated Machine Learning Approach for Smart Waste Management Systems. *IEEE Transactions on Industrial Informatics*, *16*(1), 384–392. https://doi.org/10.1109/TII.2019.2915572

Salih Hasan, B. M., Abdulazeez, A. M., & Duhok Polytechnic University. (2021). A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Journal of Soft Computing and Data Mining*, *02*(01). https://doi.org/10.30880/jscdm.2021.02.01.003

Santamaria, A., Ortega-Lopez, V., Skaf, M., García, V., Gaitero, J., San-Jose, J., & González, J. (2019). *Ladle furnace slag as cement replacement in mortar mixes*. 430–440. https://doi.org/10.18552/2019/IDSCMT5048

Setién, J., Hernández, D., & González, J. J. (2009). Characterization of ladle furnace basic slag for use as a construction material. *Construction and Building Materials*, *23*(5), 1788–1794. https://doi.org/10.1016/j.conbuildmat.2008.10.003

Seyoum, T., & Adeloju, S. (2008). Analysis of Elemental Composition of Solid Industrial Wastes and its Use for the Assessment of Reuse/Recycling Options. *Journal of Solid Waste Technology and Management*, *34*, 56–69.

Spooren, J., Binnemans, K., Björkmalm, J., Breemersch, K., Dams, Y., Folens, K., González-Moya, M., Horckmans, L., Komnitsas, K., Kurylak, W., Lopez, M., Mäkinen, J., Onisei, S., Oorts, K., Peys, A., Pietek, G., Pontikes, Y., Snellings, R., Tripiana, M., … Kinnunen, P. (2020). Near-zero-waste processing of low-grade, complex primary ores and secondary raw materials in Europe: Technology development trends.

Resources, Conservation and Recycling, *160*, 104919. https://doi.org/10.1016/j.resconrec.2020.104919

Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). *A Brief Review of Nearest Neighbor Algorithm for Learning and Classification* (p. 1260). https://doi.org/10.1109/ICCS45141.2019.9065747

Tseng, M.-L., Tan, R. R., Chiu, A. S. F., Chien, C.-F., & Kuo, T. C. (2018). Circular economy meets industry 4.0: Can big data drive industrial symbiosis? *Resources, Conservation and Recycling*, *131*, 146–147. https://doi.org/10.1016/j.resconrec.2017.12.028

Turkdogan, E. T. (2010). *Fundamentals of steelmaking*. Institute of Materials.

United Nations. (2014, November 21). *A World That Counts: Mobilising a Data Revolution for Sustainable Development - World | ReliefWeb*. https://reliefweb.int/report/world/world-counts-mobilising-data-revolution-sustainable-development

United Nations. (2020). *Global Resources Outlook 2019: Natural Resources for the Future We Want*. UN. https://doi.org/10.18356/689a1a17-en

Velis, C. A., Wilson, D. C., Gavish, Y., Grimes, S. M., & Whiteman, A. (2023). Socio-economic development drives solid waste management performance in cities: A global analysis using machine learning. *Science of The Total Environment*, *872*, 161913. https://doi.org/10.1016/j.scitotenv.2023.161913

Vlcek, J., Tomkova, V., Ovcacikova, H., Ovcacik, F., Topinkova, M., & Matejka, V. (2013). *SLAGS FROM STEEL PRODUCTION: PROPERTIES AND THEIR UTILIZATION*.

Walter, S., Braun, A., & Louw, L. (2023). Development of a product return process in the context of the Circular Economy with the help of

Machine Learning. *Procedia CIRP*, *120*, 1498–1503. https://doi.org/10.1016/j.procir.2023.09.201

Warren, J. (2000). Dolomite: Occurrence, evolution and economically important associations. *Earth-Science Reviews*, *52*(1), 1–81. https://doi.org/10.1016/S0012-8252(00)00022-2

Wei, G. (2024). *Gmft* [Python]. https://github.com/conjuncts/gmft (Original work published 2024)

Wu, F., Liu, X., Qu, G., & Ning, P. (2022). A critical review on extraction of valuable metals from solid waste. *Separation and Purification Technology*, *301*, 122043. https://doi.org/10.1016/j.seppur.2022.122043

Yildirim, I. Z., & Prezzi, M. (2011). Chemical, Mineralogical, and Morphological Properties of Steel Slag. *Advances in Civil Engineering*, *2011*(1), 463638. https://doi.org/10.1155/2011/463638

7 Appendix

All code and relevant resources have been made publicly available via GitHub at: https://doi.org/10.5281/zenodo.15785322