Degree project

UNIVERSITY OF SKÖVDE

Multivariate Time Series prediction for endpoint prediction of temperature, phosphorus, and carbon in the Basic Oxygen Furnace

Master Degree Project in Informatics with a specialization in Data Science/Privacy, Information and Cyber Security

Second Cycle 30 credits

Spring term 2024

Student: Muhammad Usman

Supervisor: Gunnar Mathiason

Examiner: Juhee Bae

Abstract

To forecast the endpoint of the Basic Oxygen Furnace (BOF) process in steelmaking, we have employed deep learning techniques. However, our project faces limitations due to insufficient data, leaving key influencing factors undisclosed at the process's conclusion. The BOF process is intricate and multi-targeted, primarily managed manually by operators. It involves converting a blend of pig iron and recycled scrap into low-carbon steel. Our strategy involves deploying a joint neural network and comparing it against a static model to evaluate whether incorporating sequential data enhances predictive precision. Trained deep learning models exhibit proficiency in accurately predicting temperature, carbon, and phosphorus within predefined limits. We did SHAP analysis for finding the influential factors for target variables. Leveraging a comprehensive dataset, we conducted predictions on these target variables. One model relies solely on static data, while the other is a joint model integrating static and sparse sequential data. Surprisingly, the accuracy of the static model surpasses the joint model, with R2 scores of 0.92 for phosphorous, 0.79 for temperature, and 0.71 for carbon compared to lower R2 scores for the joint model, indicating that richer data can indeed enhance predictions in the BOF process.

Table of Contents

1 Introduction

The Basic Oxygen Steelmaking (BOS) process, executed within LD-converter furnaces, represents a sophisticated industrial effort aimed at the production of low-carbon steel from pig iron and recycled scrap. This intricate process involves a myriad of chemical and physical interactions, posing challenges in establishing clear correlations between process parameters and the final steel quality (Bae, Mathiason, Li, Kojola, & Ståhl, 2021). However, the rigorous operating conditions often hinder the availability of in-depth sensor data, resulting in a deficiency of comprehensive information. In the BOS steelmaking process, solid steel waste and liquid hot metal sourced from blast furnaces are introduced into the BOF converter vessel, initiating a batch process geared towards generating liquid steel. Subsequently, oxygen gas is introduced through a vertical lance mounted above the liquid metal surface at supersonic velocities. This oxygen injection leads to the formation of an oxide slag mixture, incorporating lime and calcined dolomite, as silicon, iron, and other minor metal components react with oxygen, including manganese and titanium. This slag formation serves multiple purposes, including phosphorus absorption and stabilization of the vessel's refractory lining.

During the oxygen injection phase, the vigorous gas flow induces violent splashing of liquid metal droplets into the slag. As silicon oxidation progresses, carbon in the metal reacts with oxygen, predominantly yielding carbon monoxide gas, which is expelled through the exhaust system. Concurrently, solid scrap undergoes melting due to the heat generated by oxidation reactions. Initially, the carbon content of the heated metal exceeds 4%, gradually decreasing to an average of 0.1% by the process's completion (Bae, Li, Ståhl, Mathiason, & Kojola, 2020). Presently, the furnace operates without solid scrap, with human operators injecting cooling agents such as ore or heating agents like ferrosilicon to maintain thermal balance during operation. The system mass encompasses the collective weight of materials introduced into the vessel, including ore, hot metal, lime, calcined dolomite, scrap, and additional hot metal. Given the high operational temperatures of up to 1750°C, the furnace experiences rapid cooling during downtime between batches. Hence, minimizing the interval between heats becomes imperative to uphold process predictability and efficiency. The overall process can be seen in figure 1 attached below. After the LD furnace process it goes to the RH-Process and CAS-OB. The continuous casting is being done and then steel slabs are made.

The central aim of this study revolves around the development of a resilient predictive model, leveraging the wealth of available data while integrating insights into the physical process of Luleå plant. Luleå is a coastal city in Swedish Lapland. The company had a branch in Luleå. The key objective is to create a model that maximizes the utilization of available data. By carefully capturing the specific execution patterns witnessed in each furnace, the envisaged model endeavours to provide precise predictions of process outcomes under a spectrum of operational scenarios. A critical aspect under scrutiny is the efficiency of utilizing time-series data within our framework. This exploration seeks to ascertain whether leveraging time-series data proves beneficial in our context or whether alternative approaches might yield superior results. Moreover, the study explores into the degrees of data collection setups, recognizing that variations therein can significantly impact the model's performance and its ability to faithfully represent the typical furnace processes. Through this comprehensive investigation and model development, we endeavour to enhance our understanding of the intricate relationship between data utilization, process dynamics, and predictive modelling, ultimately advancing the efficiency of steelmaking operations.

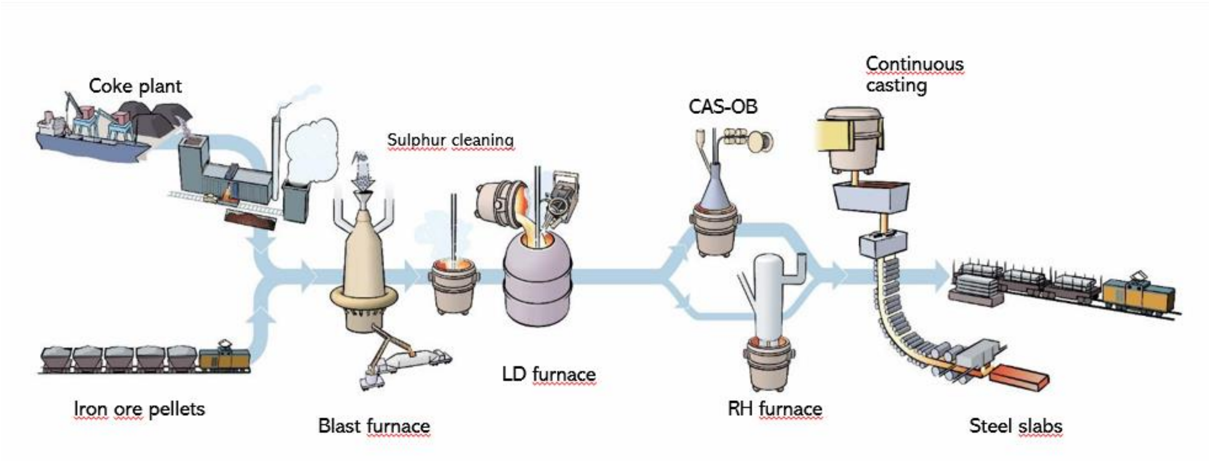

Figure 1: Overall Process

The novelty of our work lies in the integration of both sequential and static data to enhance predictive modeling within the steelmaking process. Previous research predominantly focused on either static or sequential data independently. By combining these two data types, our approach attempts to capture a more holistic view of the steelmaking process, potentially uncovering interactions and dependencies that single-data-type models might miss. Additionally, the use of Long Short-Term Memory (LSTM) networks in this context is innovative, as they are specifically designed to capture temporal dependencies and patterns in sequential data, which are crucial for processes like steelmaking that evolve over time. Our contributions to the data science community include:

- We present a methodology for integrating sequential and static data within a single predictive model, which can be applied to other industrial processes beyond steelmaking.
- The development and validation of a joint LSTM model tailored for complex industrial processes contribute to the body of knowledge on deep learning architectures.

The innovative approach presented in this study integrates statistical data (heat data) with time series data to create an advanced predictive model tailored for the Basic Oxygen Steelmaking (BOS) process. Initially, a baseline is established using a model solely reliant on static data, providing a foundational understanding of the process. However, recognizing the dynamic nature of the BOS process, our focus shifts towards harnessing the combined power of static and time series data. This integrated model possesses the capability to encapsulate both the static attributes linked to process parameters and the evolving patterns discernible over time. Within the BOS process, the behaviour of slag is a dynamic phenomenon, with its levels fluctuating and occasionally reaching critical thresholds. Additionally, chemical transformations occur continuously throughout the process, further highlighting its dynamic nature. To ensure the robustness of our model, we have engaged in extensive discussions with domain experts to set significant limitations and refine our approach accordingly. Our aim is to assess the efficiency of our model in producing predictions within desired ranges while scrutinizing whether the joint model enhances predictive accuracy and the potential value of its features for future applications.

The motivation to compare the static and joint models stems from the need to understand the benefits and limitations of integrating sequential and static data in predictive modeling. By comparing these models, we aimed to:

- **Evaluate Performance**: Assess the effectiveness of combined data types in improving prediction accuracy.
- **Identify Strengths and Weaknesses**: Determine scenarios where one model might be preferable over the other.
- **Guide Future Research**: Provide insights and directions for future work in integrating diverse data sources for industrial process modeling.

The effectiveness of our developed model is evaluated using real-world data collected from a single site, encompassing two furnaces with distinct data collection setups. Given the equivalence in static features across both furnaces, as well as in time series dataset, merging the data facilitates a unified analysis. Details regarding the data specifics will be explained in the subsequent Data section of this report. In our research, neural networks serve as the cornerstone for both static and joint models. These networks are meticulously constructed, with the initial layer incorporating inputs from both furnaces, subsequently propagated through successive layers. Rigorous testing and validation of the model across diverse scenarios have been undertaken to optimize performance and ensure reliability. Evaluation metrics, including the R2 score, are employed to gauge the accuracy of our models, with initial findings indicating that the proposed static model proficiently captures the intricacies of the BOS process. Throughout our research endeavour, we remain cognizant of the social and ethical dimensions inherent in our project. The following sections of this report will delve deeper into the complexities encountered and overcome while propounding the joint model, explaining on our journey towards enhancing predictive capabilities in steelmaking operations.

1.1 Problem definition

Steel production represents a complex and energy-intensive endeavour, characterized by a series of complex mechanical and chemical transformations. The ultimate quality of the steel product centers upon a myriad of factors, including temperature, pressure, chemical composition, and the duration of each stage within the production process. However, the hostile working environment poses challenges in the precise monitoring and control of these critical parameters. Consequently, this can result in suboptimal steel quality and inflated production costs, underscoring the pressing need for innovative solutions to enhance process efficiency and product quality. We have three target variables and needed to predict them at end of process. The target variables are temperature, carbon and phosphorous. We will do analysis on both the types of data. We will make a static model and joint model.

Research Question

➢ Can the utilization of time series data provide significant improvement in the quality of steel for BOF process based on sequential data analysis and static data analysis?

The objectives of the project are:

➢ Read the literature study and effective ways to design a model.

➢ Collect and pre-process sequential data and static data correspondingly from the steelmaking process to check the co-relation of the features.

➢ Establish a static model for the heat data provided.

➢ Develop and train a joint deep learning model to analyze the sequential data and predict the target variables.

➢ Evaluate the influential features for target variables using SHAP analysis.

➢ Evaluate the performance of both the static model and the joint deep learning model.

➢ Make sure that societal and ethical aspects are met along with the future considerations.

Chapter 2 of our report delves into an exploration of the Basic Oxygen Steelmaking (BOS) process, offering a thorough background to contextualize our research. Within this chapter, we embark on a journey through the complexities of BOS, solving its fundamental principles and operational degrees. Furthermore, we meticulously examine relevant literature to glean insights from previous studies, synthesizing a comprehensive understanding of the current state-of-the-art in BOS process optimization and predictive modelling. Moving forward, Chapter 3 includes the methodological framework supporting our research accomplishments. Here, we unveil a systematic and meticulously crafted approach employed in the collection and pre-processing of joint static and time series data sourced from multiple LD-converter furnaces. Each step of our methodology is detailed. By explaining our methodological underpinnings, we strive to foster transparency and reproducibility, ensuring the robustness and reliability of our research findings. Implementation of our models are discussed in chapter 4 along with the results and discussion in chapter 5 and chapter 6 simultaneously. In the end, conclusions and future considerations are also mentioned in report.

2 Background

In this section, we embark on an illuminating exploration of two pivotal concepts critical to laying the foundation for our subsequent research background: neural network for predictive modeling of target variables and LSTM (Long Short-Term Memory) networks for effectively modeling sequential data. By delving into these concepts, we aim to equip readers with the requisite understanding to grasp the complexities of our research actions.

Neural networks represent a pivotal tool in predictive modeling, enabling the extraction of complex relationships between independent features and a target variable. By leveraging neural network architectures, we aim to uncover complex patterns and forecast outcomes within our research sphere. Simultaneously, we explore the domain of Long Short-Term Memory (LSTM) networks, an advanced type of recurrent neural network known for its proficiency in identifying long-term dependencies in sequence data. By leveraging the capabilities of LSTM networks, we aim to extract valuable insights from the sequential data streams within our research area, thereby improving our predictive abilities.

2.1 Preliminaries

In this heading, we will give overview about the techniques being used in the thesis project.

2.1.1 Neural Network

Neural network stands as a statistical bedrock utilized to decipher and predict the complex interplay between a dependent variable, often referred to as the target variable, and one or more independent variables, commonly known as predictor variables. In the realm of our research focused on Basic Oxygen Steelmaking (BOS) processes, neural network emerges as a formidable way wielded to forecast pivotal target variables such as temperature, carbon content, and phosphorus content. Embracing the tenets of neural network, we embark on a journey of mathematical fitting, wherein a nuanced relationship is constructed between the available data points, enabling us to extrapolate predictions with precision. By harnessing the power of neural network, we try to unravel the enigmatic dynamics of the BOS process, peering into its complexities with clarity and foresight. This admired technique not only furnishes us with predictive ability but also gives interpretability, a cherished factor essential for interpreting the underlying mechanisms governing process dynamics. Indeed, the coefficients or weights assigned to each predictor variable serve as beacons of insight, revealing the relative influence and direction of their impact on the target variable. Amidst the vast expanse of industrial processes, finance, and social sciences, neural networks stand as stalwart sentinels, poised to unveil hidden patterns and empower decision-makers with actionable intelligence (He & Zhang, 2018). Their adaptability, capable of capturing both linear and non-linear relationships, renders them indispensable in navigating the complicated terrain of process optimization and performance enhancement.

2.1.2 LSTM (Long Short-Term Memory)

Sequential data embodies a narrative of continuity, where each data point is connected with its antecedents, forming a point of interdependent insights. Within the realm of Basic Oxygen Steelmaking (BOS) processes, the temporal rhythm of operations births a trove of time series data, each data points a testament to the evolving saga of steel production. To navigate this, we enlist the challenging Long Short-Term Memory (LSTM) network within the annals of our joint model.

One of the greatest recurrent neural networks (RNN) designs, LSTM, is ready to unlock the unsolved patterns in our time series data. In contrast to its traditional equivalents, LSTM networks excel in capturing long-term dependencies in sequential data due to their unique architecture, which includes cell states and gating mechanisms. This capability is crucial for the Basic Oxygen Steelmaking (BOS) process, where understanding the long-term interactions and effects of various stages is essential for accurate modeling and prediction. Traditional RNNs struggle with long-term dependencies because they are prone to the vanishing gradient problem, where gradients become too small during backpropagation through time, leading to ineffective learning of long-range patterns. Time delays and temporal dissonance are no longer limitations for LSTM, which skilfully traverses the ups and downs of temporal dynamics through its command of memory (Van Houdt, Mosquera, & Nápoles, 2020). Furthermore, LSTM can work in a temporal context as well, handling the distinction of short-term lapses and long-term narratives with equal skill. This double ability creates an assortment of insights that allow accurate forecasts based on the extensive collection of furnace data to be made. However, LSTM networks are specifically designed to mitigate the vanishing gradient problem through their gating mechanisms, ensuring that gradients are maintained during backpropagation. This makes them particularly suitable for deep architectures required to model complex time series data like that of the BOS process. While CNN-LSTMs can also mitigate this issue to some extent by combining convolutional layers with LSTM layers, they may introduce additional complexity and computational overhead. Although LSTMs are computationally intensive, they often strike a good balance between model complexity and performance for time series tasks. They provide robust performance without the additional layers and parameters introduced in CNN-LSTM architectures. RNNs can be more computationally demanding due to the added convolutional layers, which may not always yield proportional improvements in performance for sequential data without spatial dimensions. By using LSTM, we are able to dig deep into our data, finding valuable insights and unlocking its secrets.

2.1.3 SHAP (SHapley Additive exPlanations)

SHAP (SHapley Additive exPlanations) is a powerful and widely used method for interpreting complex machine learning models. It stands out among other feature-importance approaches for several reasons. Shap values provides a solid theoretical foundation ensuring fair and consistent feature importance values. Each feature's contribution to the prediction is calculated considering all possible combinations of features, which helps in capturing interactions between features. Features that do not change the prediction have a SHAP value of zero. Unlike methods that may provide biased or inconsistent attributions, SHAP ensures that if a model changes such that a feature contributes more to the output, its SHAP value will not decrease. This consistency is critical for reliable model interpretation. SHAP summary plots provide a comprehensive overview of feature importance and interactions across the entire dataset. SHAP can separate the effect of each feature value, showing not just the importance of a feature but how specific values of that feature influence predictions.

Positive SHAP Value: A positive SHAP value for a feature indicates that this feature is pushing the model's prediction higher compared to the average prediction. In other words, it contributes positively to the outcome.

Negative SHAP Value: A negative SHAP value for a feature means it is pushing the prediction lower compared to the average prediction, contributing negatively to the outcome.

Colours in SHAP Plots:

Red (Positive Influence): Features coloured in red are typically those with high values and are contributing positively to the model’s prediction. For example, if a feature is in red and has a high SHAP value, it means feature is pushing the prediction higher.

Blue (Negative Influence): Features coloured in blue generally represent low values and contribute negatively to the model's prediction. If feature is in blue and has a high (negative) SHAP value, it means feature is pushing the prediction lower.

2.2 Research Background

In predicting the final outcomes of the Basic Oxygen Steelmaking (BOS) process, researchers often rely on various regression models. These models come in different types, such as linear or nonlinear, and they are adept at forecasting crucial factors like melt temperature and carbon concentration in the melt. The basic regression model serves as a solid foundation, acting as a starting point for predictions. For example, some studies utilize the gray model alongside regression techniques. Additionally, there is a method called continual adaptation, where the regression parameters are continuously adjusted using the least-squares method, as proposed in certain studies. Furthermore, other researchers have explored different avenues. For instance, Huang et al. (Xie & Chai, 1999) conducted regression analysis on BOS data and put forward a model specifically targeting oxygen and temperature variables for prediction purposes. These approaches showcase the diverse strategies employed by researchers to tackle prediction challenges within the BOS domain.

Much research has been dedicated to predicting the endpoint of the Basic Oxygen Steelmaking (BOS) process using neural network-based models. These neural networks, or NNs for short, are pretty clever. They can learn on their own, handle complex relationships, and process data really fast. The examples of NNs being used to predict BOS process variables in various studies, like those mentioned in references (Gao, Shen, Liu, Wang, & Chu, 2019). Typically, these applications involve combining NNs with different optimization methods to make them work even better. There are also plenty of NN applications out there aiming to predict specific factors like stopping temperature and end blow oxygen in the LD converter. Some of these applications have seen improvements over time, as seen in reference (Yue, Yao, Zhao, & Wang, 2013). One notable advancement is the adaptive neural network fuzzy inference system (ANFIS), which has shown promise in enhancing predictions related to the BOS process.

The related studies conducted thus far exhibit variations across several dimensions, each of which influences their conclusions. These variances encompass factors such as the size and accuracy of the dataset, the reliability and sophistication of the machine learning (ML) algorithms employed, the breadth and diversity of information utilized, the target error range, the inclusion of both direct and estimated features, and the validation methodology employed (Bae, Mathiason, Li, Kojola, & Ståhl, 2021). For instance, the size and accuracy of the dataset play a pivotal role in the efficiency of the study's conclusions. Complicated ML algorithms may succumb to overfitting when applied to small datasets, thus compromising their reliability. Conversely, employing a simple ML method on a small dataset with limited variability may yield accurate predictions due to the reduced risk of overfitting. However, the utilization of limited techniques may not be suitable for real-world manufacturing models, as small datasets may fail to accurately capture the data distribution over an extended production period. Moreover, the scope and variations of the information utilized in these studies also influence the robustness of their conclusions. Studies incorporating a comprehensive range of relevant information are more likely to yield insights that generalize well to real-world scenarios. In summary, the choices made regarding dataset size, algorithm complexity, information scope, and validation methodology significantly impact the conclusions drawn from related studies. Careful consideration of these factors is essential to ensure the reliability and applicability of research findings in practical settings.

Previous literature studies serve as a bedrock of knowledge, providing invaluable insights into the research topic at hand. They give a broad view of the current landscape, summarizing key theories, concepts, methodologies, and findings pertinent to the field. For a newcomer without domain expertise, navigating the complex intricacies of the steelmaking process can prove daunting. Within this complex milieu, myriad reactions unfold, each holding the potential to impact the process in unforeseen ways. Consider a scenario where modifying a few features may yield improvements in one target variable while inadvertently compromising others. Such delicate balances underscore the necessity of a comprehensive approach to data analysis. Hence, we opt to conduct my comparisons using the entirety of the provided datasets. This holistic approach ensures that all records of the process are duly considered, safeguarding against potential blind spots or unintended consequences. Given my lack of practical experience in steelmaking, delving deep into the degrees of the Basic Oxygen Steelmaking (BOS) process was imperative. While past literature reviews have often relied on alternative datasets, my research benefits from access to the dataset spanning from 2023.

3 Method

In this section, we will outline the dataset utilized for creating the models and explain the rationale behind the chosen methods, explaining how they contribute to answering our research question.

3.1 Dataset

The SSAB group industrial LD-converter is where the data were taken from a real manufacturing of the year 2023. Data for this research was gathered from a BOF production line and distributed in two separate ways: as static data and time series data. As a result, there are two different data sets, each of which has some features in common with the others and some attributes that are specific to it. Some are gathered in the same manner while others are represented differently. The static datasets have 146 features. The dataset has 15,655 samples. For the time series dataset, 36 features are there. The time series dataset has 523956 samples. The only common thing in both static and time series datasets is the heat number represented by heatno. Time series datasets range for heatno varies a lot. It means that one heat number can have different samples and readings.

The features comprise of the parameters before the oxygen blow, during and after the oxygen blow from lance. The position of lance for blowing is also not constant sometimes it becomes closer to the scrap and sometimes taken far away which also is one of the reasons to consider for taking all parameters for the prediction models. Temperature, carbon, and phosphorus are our three targets. The goal of the model is to accurately predict the estimate of the end temperature (T) as well as the proportions of carbon (C) and phosphorus (P) in the finished product, which is similar to the target for the physical process. By having a discussion with domain experts, some limits are being set to see model predictions within certain range i.e. Temperature at 15 degree, Carbon at 0.01 and phosphorous at 0.003. With the physical process, the goal is for the steel temperature to be within a defined temperature range during each process execution (a heat), and for the carbon and phosphorus levels to be below predetermined criteria. There is position of oxygen lance, station numbers, blowing times, percentages of addition of different gases which includes argon, nitrogen, hydrogen etc. Time for tapping and stirring is also included along with distance of lance from the steel surface during that time. At the end of blow, some features are added to cool down the process i.e. Lime, Ore, Dolomite etc. In time series, bottom stirring from six different purge plugs are also included. Overall, it is a very complex process to understand and perform any approach. We tried with most of the features to see if time series can improve the accuracy of the target variables or provide significant results.

3.2 Approach

In this comparative analysis, we approach the problem by first examining a static dataset to establish a baseline, and then integrating it with a joint model. To achieve this, we make a unified dataframe. This consolidation ensures a comprehensive view of the data, facilitating a robust comparison. We refine the dataframe by dropping irrelevant columns and filtering rows based on specific criteria, such as station values and the presence of non-null values. These filtering steps ensure that only pertinent data points are retained for further analysis. Next, we preprocess the data by standardizing and normalizing it, a crucial step to ensure consistency and comparability across different features. Standardization involves adjusting each feature values to have a mean of zero and a standard deviation of one, while normalization scales the values within a certain range, typically between zero and one. Once the data is pre-processed, we extract the target variables and remove any redundant features identified through collaboration with domain experts. This precise curation of features ensures that our model focuses on the most relevant information, enhancing its predictive accuracy. Finally, we calculate descriptive statistics such as mean and standard deviation for each column in the dataset, providing insights into the data distribution and variability. These statistics serve as valuable indicators of the dataset characteristics and inform subsequent modeling decisions. Moving forward, we will delve into the design and implementation of a LSTM in joint sequential model.

In our data preprocessing pipeline for handling time series data within a static dataframe, we prioritize proper alignment with target heat numbers derived from the time series dataset. To achieve this, we meticulously sort the static dataframe based on the 'heatno' column. This sorting ensures that the sequence of heat numbers in the static dataframe matches that of the time series data, facilitating synchronization between the two datasets. To further enhance alignment, we create a new dataframe to store unique values of the 'heatno' column from the merged dataset. This serves as a reference point for identifying matching heat numbers between the static and time series datasets. To validate the alignment of heat numbers, we conduct a comparison to confirm consistency post-merge operation. By converting the dataframes into arrays and employing Boolean logic for equality checks, we verify that heat numbers are aligned correctly. Additionally, we introduce an engineered feature called the carbon escape time series into the dataframe to explore its potential impact on the dataset in 2023

Despite our efforts, however, we find that this feature does not significantly influence the predictive capabilities of the model. We also added CO/CO2 engineered feature that does not significantly improve our results.

Engineering the Carbon Escape Feature: We calculate a weighted sum of 'CO2 off gas' and 'CO off gas' using specific coefficients, guided by insights from domain experts. This sum is then divided by 'flow dry off gas' divided by 60 to derive the engineered feature.

Feature Selection: Some features deemed non-influential, such as lance position with respect to time and Oxygen flow minute, are dropped from the static data after discussing with the domain experts. This refinement leaves us with less features from the static data and the time series data.

Normalization of Time Series Data: All columns in the time series dataframe are normalized using the z-score normalization technique. This involves dividing each column's mean by its standard deviation after grouping them by heat numbers.

Padding of Time Series Sequences: We create a list containing feature time from new heat and check various statistics such as maximum length, minimum length, and mean length. Afterward, we employ the pad_sequences function from TensorFlow's Keras module to pad each feature individually to ensure uniform length.

Dimensionality Adjustment: As we are close to modeling, certain non-influential features from the static data are removed. After discussing with the domain experts, we removed some features containing time e.g. blowing time, time_sample, time_sublance etc. This leaves us with around 100 features on which we will build our joint neural network.

This thorough data preparation process ensures that our data is properly formatted and standardized for input into our joint neural network model. By incorporating relevant features and optimizing their structure, we aim to enhance the model predictive capabilities and gather actionable insights for the Basic Oxygen Steelmaking (BOS) process.

4 Implementation

In this section, the implementation of the models are discussed below.

4.1 Implementation of Static Neural Network

Under the implementation section, the groundwork laid in data preparation transitions seamlessly into the construction and training of the static model. Leveraging established libraries, we initialize crucial variables and data structures integral to subsequent model processing, encompassing training and evaluation phases. Central to our model architecture is the definition of the latent space dimensionality, set at 1024. This serves as the foundational framework for the succeeding deep learning model. Subsequently, we delineate the model architecture, comprising four hidden layers of sizes 512, 256, 128, and 64, respectively, encapsulated within a list structure.

Guided by the principles of reproducibility, we establish a seed value for the random number generator, ensuring consistent results across iterations. Through controlled random integer generation, the training and testing datasets are 90 percent and 10 percent simultaneously, laying the groundwork for cross-validation procedures. Within the model construction phase, a modified architecture is constructed, integrating dense layers and LeakyReLU activations to fill the model with robustness against overfitting. Dropout regularization, set at a rate of 0.4, is judiciously employed to mitigate the risk of model overfitting, enhancing generalization performance. Embracing best practices in model optimization, a model checkpoint mechanism is instituted to preserve weights corresponding to the lowest validation loss, thus safeguarding against convergence to local optima. The end of this process reveals in the compilation, training, and evaluation of the deep learning model, guided by the mean squared error loss function and the Adam optimizer.

Lastly, predictions are computed for the test data, with prediction errors meticulously logged for subsequent analysis. The four hidden layers are chosen to effectively capture the complexity of the input data. The first hidden layer with a large number of units (512) captures a broad range of features, while subsequent layers (256, 128, 64) progressively reduce the feature space, allowing the model to learn more abstract representations. This setup helps in handling high-dimensional data and learning hierarchical features, which is essential for accurate predictions in the Heat model. Through this rigorous implementation framework, we ascertain the model efficiency in capturing underlying patterns within the dataset, thereby advancing our understanding of the underlying data dynamics.

4.2 Implementation of Joint Neural Network

The model we are building relies on a neural network, and we start by importing all the necessary libraries for model construction, including layers, optimizers, callbacks, and regularization techniques. We differentiate between two types of input: sequential data, which has a time dimension with multiple features at each time step, and static data, which consists of fixed features for each sample.

Our sequential input data consists of 7,582 samples, each with 32 time steps and 32 features, while the static data comprises 7,582 samples with 85 features. We removed four features here which are related to timeid, timefromblowstart, maxtimefromnewheat, and lance position. Everything is closely discussed with the domain experts before removing. We create arrays to store the predicted values and validation results for target variables like Temperature, Carbon, and Phosphorous. Next, we define test fold indices for cross-validation, where the data is divided into folds for training and evaluation. Using 10 folds strikes a balance between accurate performance estimation and computational efficiency. Each fold is used as a test set exactly once, with the remaining folds used for training and validation. Then, we partition the data into training, and testing sets for both sequential and static data.

In model architecture, we apply a Masking layer to handle any padded values in the sequence input. We then use an LSTM layer with 128 units and apply L2 regularization to prevent overfitting. The LSTM layer output is concatenated with the static input and passed through dense layers with ReLU activation and dropout regularization to further prevent overfitting. We define three output layers for Temperature, Carbon, and Phosphorous prediction, each with a dense layer and linear activation function. The model is compiled using the Adam optimizer with a learning rate of 0.001 and mean absolute error as the loss function for all three outputs. The model is trained with a batch size of 128 and for 500 epochs.

In the subsequent phase, we execute computations to assess the performance of our model, comparing its predictions against actual values. Through systematic organization into lists, we carefully capture and quantify disparities between predicted and actual outcomes. Any absence is diligently handled, initializing lists to zero before populating them with computed differences. To evaluate model effectiveness, we contrast its performance against a rudimentary baseline model, which simply predicts mean values of training data for each target variable. Visual representation is harnessed through scatter plots, offering insights into the relationship between predicted and true values. Setting predefined limits for acceptable prediction ranges – 15 for temperature, 0.01 for carbon, and 0.003 for phosphorous – we steer the analysis towards discerning how accurately the model predicts within specified bounds. The LSTM model includes one LSTM layer followed by three Dense layers. The choice of four hidden layers (1 LSTM + 3 Dense) is to balance the complexity and the model's ability to capture intricate patterns in the time series data. The LSTM layer is responsible for capturing the temporal dependencies, while the Dense layers further process these learned representations to make predictions.

In general, this methodological framework serves as a conduit for addressing the research question, shedding light on the differential impacts and effectiveness of leveraging sparse sequential data within predictive modeling models.

5 Results

The results of our models are as follows:

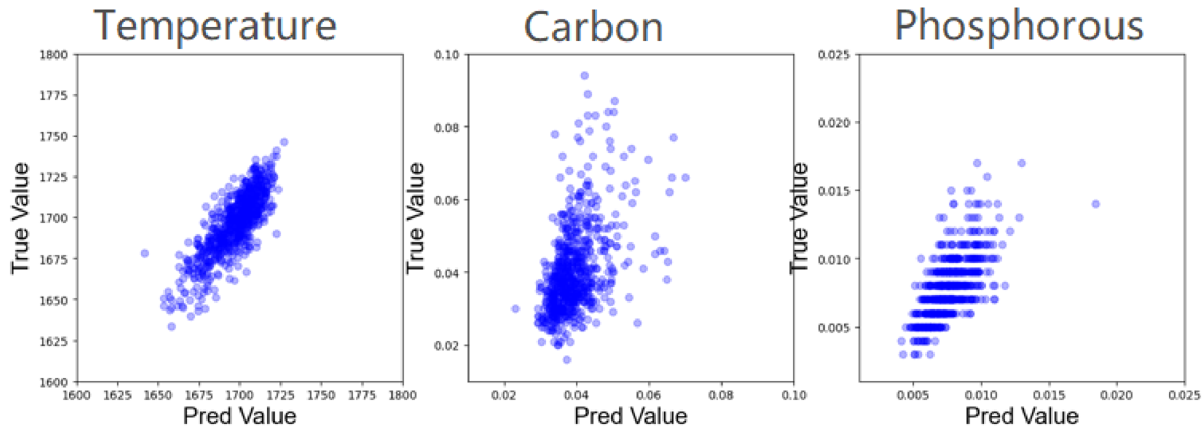

Figure 2: Scatter Plots for Temperature, Carbon, and Phosphorous

In Figure 2, we can observe a clustering of predicted values for temperature between 1640 and 1750, while for carbon, values range from 0.002 to 0.02. These bounds serve as benchmarks for assessing the model performance, particularly in capturing distinctions within the range of 0.02 to 0.015 for carbon and 0.002 to 0.02 for phosphorous. These scatter plots are made on the joint model. The dense clustering of points around the center of the diagonal line indicates that most of the predictions are close to the true values, signifying a well-performing model. The points are clustered along a diagonal line from the bottom left to the top right, showing a strong positive correlation. This means that as the predicted temperature increases, the actual temperature also increases, which suggests that the predictions are accurate. The data points are spread out more widely along the horizontal axis, which indicates a greater variance in the predicted values for carbon. Most points are clustered around the lower end of the predicted value range, suggesting that the model predicts lower carbon values more frequently. The data points are more tightly clustered along a diagonal line, indicating a strong positive correlation between the predicted and true values for phosphorous. This suggests that the model's predictions for phosphorous are quite accurate and consistent.

	R2 score for models			
	Static model	Joint model	Joint model after adding carbon escape	Joint model after adding CO/CO2
Temperature	0.79	0.66	0.68	0.66
Carbon	0.71	0.16	0.22	0.21
Phosphorous	0.92	0.41	0.43	0.43

Table 1 : R2 score for models

In Table 1, we can see the R2 score for all the models. R^2, also known as the coefficient of determination, provides a clear and interpretable value that indicates how well the model's predictions match the actual data. It ranges from 0 to 1, where 0 indicates that the model does not explain any of the variance in the target variable, and 1 indicates that the model explains all the variance. It provides an indication of the strength and direction of the linear relationship between the predicted and actual values, which is crucial for problems. R^2 is scale-invariant, meaning it does not depend on the scale of the target variable, making it versatile for different datasets and units of measurement. MAE treats all errors equally, which might not always be desirable. MSE is not as interpretable as R^2, and the squared units can make it less intuitive. Like MSE, RMSE does not provide a normalized measure of fit and can be influenced by outliers.

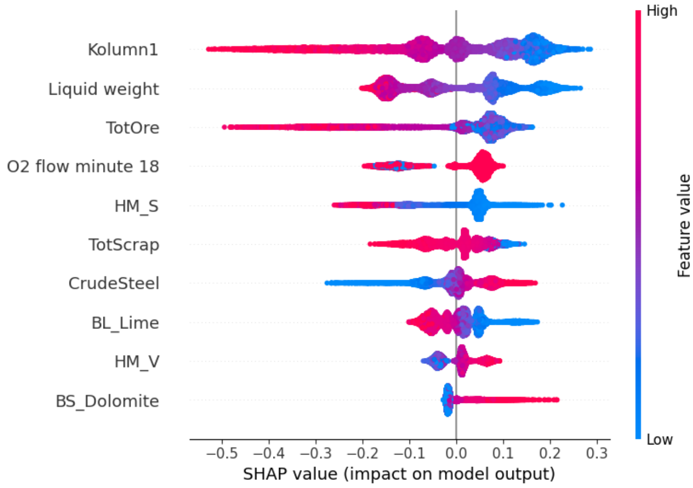

Figure 3: Influential factors for Phosphorous target

According to global model interpretation of phosphorous feature with SHAP using a summary plot, crude steel, dolomite is positively influenced with the target variable. When the crude steel, dolomite is high, target variable is driven positively and liquid weight, total ore, hot metal sulphur are negatively influenced with the target variable which means when these features are less then taget variable is influenced negatively.

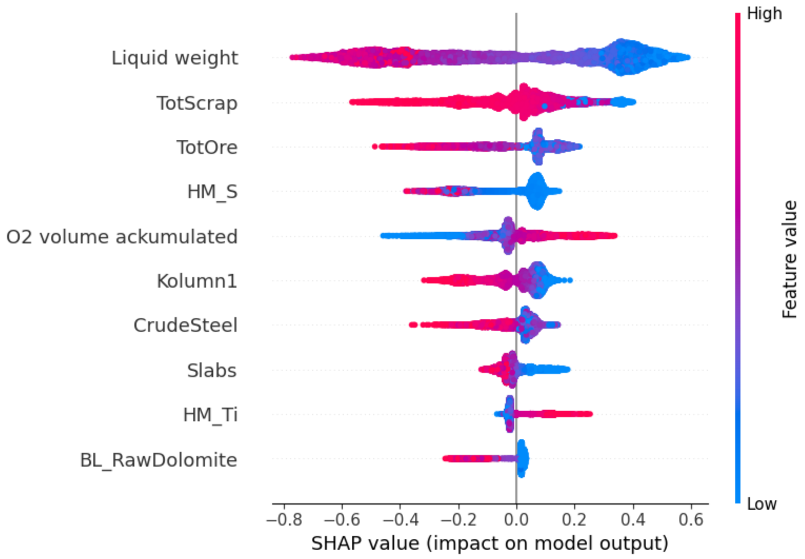

Figure 4: Influential factor for Temperature target

According to global model interpretation of temperature feature with SHAP using a summary plot, hot metal titanium, some amount of total scrap is positively influenced with the target variable and liquid weight, crude steel, ore, slabs are negatively influenced with the target variable. Kolumn 1 here is also a scrap and negatively influenced.

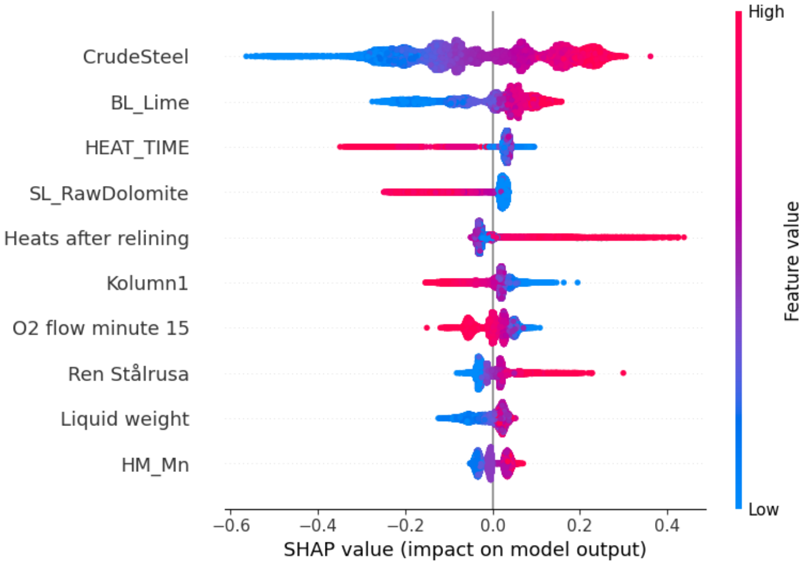

Figure 5: Influential factor for Carbon target

According to global model interpretation of carbon feature with SHAP using a summary plot, crude steel, Lime during oxygen blow, liquid weight is positively influenced with the target variable and Raw Dolomite, scrap are negatively influenced with the target variable. The selection of SHAP analysis will be explained in the discussion section of the report below.

6 Discussion

The finding that the static model's accuracy surpasses that of the joint model is unexpected. Typically, we might anticipate that leveraging multiple datasets through a joint model would yield better performance due to the enriched and diverse data inputs. However, the joint model did not outperform the static model, particularly for the carbon predictions, which was experienced. The hypothesis was that combining multiple datasets from different furnaces within the same company would provide a more comprehensive and robust dataset, thus reducing uncertainties and improving model accuracy. The basis was that the joint model could learn implicit representations and leverage complementary information from the diverse datasets, resulting in better performance. However, this was not observed in practice, indicating potential issues such as data heterogeneity, differing data distributions and ensuring the model could effectively learn from different data representations without overfitting or underfitting. The work aimed to develop such a resilient model by integrating multiple datasets and enhancing a neural network-based learning algorithm. While the static model demonstrated promising results, the joint model did not meet expectations, highlighting challenges in integrating and jointly training on diverse datasets. The carbon escape time series did not significantly improve the model accuracy, likely due to the late-stage calculation of carbon content in the steelmaking process. This suggests that the temporal patterns in carbon escape were not as predictive as expected. Similarly, CO/CO_2 ratios may not have provided the expected insights due to their complex and non-linear relationship with the target variables. These gases are influenced by various factors, and their direct relationship with the target outcomes may be too intricate to capture with the current model setup.

Future work could focus on more sophisticated methods of integrating and harmonizing multiple datasets, possibly using advanced data preprocessing and feature engineering techniques to better capture the underlying patterns. Further refinement in feature selection and engineering could improve model performance. This includes exploring additional relevant features, improving the representation of existing features, and possibly reducing dimensionality to enhance learning efficiency. For the carbon escape time series, more advanced time series analysis techniques or recurrent neural networks (RNNs) might capture the temporal dependencies more effectively. This aligns with the future consideration of using RNNs for potentially significant improvements. Iterative testing and validation with different configurations and settings can help identify optimal approaches. Continuous experimentation with various loss functions, evaluation metrics, and activation functions can lead to a more robust model. The lessons learned from this practice are as follows

- The integration of multiple datasets is challenging and requires careful handling to ensure the model benefits from the additional information.
- The choice of features and their representation plays a critical role in model performance.
- Advanced time series analysis and model selection are crucial for capturing complex temporal dynamics in industrial processes.
- Robust models require a balance between complexity and interpretability, ensuring robust performance across diverse conditions.

The lower R2 scores for the joint model suggest that this approach was less effective in capturing the variance in the target variables compared to the static model. Several potential reasons can be that the joint model trained on datasets from different furnaces, which may have had varying operational conditions, measurement techniques, and noise levels. These differences can introduce complexities and inconsistencies that hinder the model's ability to learn generalizable patterns. The interaction between features from different datasets might not be straightforward. The model might face difficulties in effectively combining these features to produce accurate predictions, especially if the relationships between features and target variables differ across datasets. The results suggest that different modeling approaches may be needed for different datasets. A one-size-fits-all model might not be the best strategy when dealing with diverse data sources. Tailored architectures that account for dataset-specific characteristics could improve performance. The lessons learned are to consider designing custom model architectures for different datasets or using ensemble methods that can separately learn from each dataset before combining their predictions. Explore advanced regularization techniques and hyperparameter tuning specifically tailored for joint models. Techniques such as multi-task learning, transfer learning, and domain adaptation could be beneficial.

6.1 Theoretical framework

In the past, researchers explored a wealth of data gathered during the production process to see how well traditional machine learning methods could handle the complexity of a Basic Oxygen Furnace (BOF) and predict a crucial production target. Surprisingly, even when carefully selecting the data, the results matched or even improve what was found in earlier studies. This suggests that while our dataset is promising, there is still room to improve how we train models for future predictions. The uncertainty in our data, caused by things like random process variations, uncertainties in sensor setups, and limitations of the sensors themselves, is a big challenge we face.

Our study is grounded in the theoretical framework of predictive modeling and deep learning, particularly focusing on Long Short-Term Memory (LSTM) networks for handling time series data. LSTM networks are well-regarded for their ability to capture temporal dependencies and manage long-term relationships within sequential data.

Use of LSTM Networks: Similar to prior studies, our research utilized LSTM networks to handle time series data, aligning with the established theoretical understanding of their capabilities. This approach is consistent with the literature, where LSTMs have been successfully applied to various predictive modeling tasks.

Focus on Predictive Accuracy: Our study prioritized enhancing predictive accuracy through advanced modeling techniques. The use of deep learning models for industrial process optimization is a common theme in the literature.

Static vs. Joint Model Performance: One notable difference is our finding that the static model outperformed the joint model, which contrasts with some studies that have reported improvements with joint models integrating multiple data types. This discrepancy suggests that the interaction between static and sequential data in our specific context might not have been effectively captured by the joint model.

Foundational Concepts: Reviewing foundational concepts in machine learning, such as supervised learning and regression analysis, provides a theoretical basis for understanding the methodologies employed in this research. Theories related to neural network architectures and optimization algorithms inform the design and implementation of deep learning models utilized in the prediction of steel quality.

Long Short-Term Memory (LSTM) Networks: Exploring the theoretical underpinnings of LSTM networks elucidates their ability to handle sequential data by capturing temporal dependencies and long-term relationships within time series data. This theoretical perspective is crucial for understanding how LSTMs can be leveraged to model the sequential aspects of the steelmaking process, where past events influence current outcomes.

R2 Score: The R2 score, or coefficient of determination, is a key metric used to evaluate the performance of predictive models. The theoretical foundation of the R2 score helps in understanding how well the predicted values align with the actual values, providing a measure of model accuracy and goodness-of-fit.

Temporal Dependencies: Theories in time series analysis inform the approach to modeling sequential data. Concepts such as autocorrelation, stationarity, and seasonality provide insights into the temporal patterns present in the steelmaking process data, guiding the development of models that can effectively capture and leverage these patterns.

Feature Importance and SHAP Analysis: SHAP provides a theoretical basis for interpreting the contributions of individual features to the model's predictions. This approach draws from cooperative game theory to attribute the prediction outcomes to specific features, helping in understanding the relative importance of different variables in the steelmaking process.

By synthesizing these theoretical perspectives, the research establishes a comprehensive framework that guides the formulation of hypotheses, the design of experiments, and the interpretation of findings. Furthermore, it provides a basis for advancing theoretical knowledge in the interdisciplinary intersection of machine learning, time series analysis, and industrial process control.

6.2 Results

This study delves into the idea that leveraging multiple datasets from within the same company's furnaces can help mitigate uncertainties arising from data scarcity and gathering constraints. However, the findings revealed that the joint model's performance was below chance and lower than that of the static model. This outcome indicates that the joint training approach did not enhance process forecasts as anticipated. Instead, the static model showcased better accuracy, particularly in predicting carbon, which had the least accuracy in the joint model due to its late-stage calculation in the process.

The enhanced findings underscore the advantage of enabling a learning system to discern implicit representations within each complementary dataset and concurrently train on multiple representations using a unified static training model. However, the joint model accuracies reveal that it did not outperform the static model, with carbon exhibiting the least accuracy due to its late-stage calculation in the process. Moreover, our engineered feature did not provide significant impact on model accuracy.

Conversely, the static model showcased promising results, underscoring the complex interaction between target variables and process reactions. The iterative process involved careful selection of the best model, loss function, evaluation metric, and activation functions, with extensive experimentation to discern their impact on outcomes.

We have done the SHAP analysis overall to find out the influential factors in our process for three different targets separately. The results are beneficial to the process experts. As seen in the figure 3 above, the expected features are related to the amount of ore, silicon, lime/dolomite, hot metal heat, and oxygen. For temperature shown in figure 4, the features common to both methods, like hot metal heat, oxygen levels, silicon and ore amounts, and heat losses, were anticipated by experts. Their consensus strengthens the notion that these factors indeed play a significant role in shaping the outcome. Additionally, from a theoretical perspective, nearly all these features were expected to be influential. For the carbon target feature shown in figure 5, the features are related to oxygen amount and carbon transfer.

We have the data from 2023 and still we came to the outcome that more data can be used to train the model and see the results. We will discuss the future consideration of RNN in the project in which we might have significant improvement. Evaluation was done using the R2 score method, with detailed results presented in the figures within the results section.

Furthermore, to optimize memory usage, we configured the TensorFlow session to allocate GPU memory dynamically, limiting usage to 50 percent of available memory. This adaptive allocation ensures efficient utilization without pre-emptively reserving the entire memory upfront.

6.3 Ethical and societal aspects

Throughout the duration of this project, utmost attention was devoted to ethical considerations and privacy safeguards. Given the focus on scrutinizing industrial data from the steelmaking realm, rigorous measures were implemented to anonymize and aggregate all data sets. This approach was pivotal in safeguarding the privacy and confidentiality of both individuals and entities involved.

While this project focuses on the steelmaking industry, the methodologies and insights gained can be applied to other industrial processes. This demonstrates the versatility and potential broad impact of the research. Ensuring that the models and methods developed are scalable and adaptable to different contexts is crucial for maximizing their societal benefits. This includes refining the models to handle diverse data types and sources effectively. By optimizing the steelmaking process and reducing errors, the amount of industrial waste generated can be minimized. This has positive implications for waste management and environmental conservation. The dataset is from last year 2023 so that the work should be done on latest advancements and be beneficial for the process. Some of the features are removed and discussed in the report above as well.

7 Conclusion

In this research, our primary objective was to create a robust deep learning model capable of analyzing both sequential and static data within the steelmaking process to predict steel quality. We synthesized insights from existing literature and methodologies for modeling intricate industrial processes. Using a rich dataset, we developed predictive models of high accuracy, focusing on deep learning models tailored to handle both statistical and sequential data. Leveraging LSTM networks, we aimed to capture temporal dependencies and deliver precise predictions regarding steel quality. Rigorous validation was central to our methodology, ensuring the reliability and robustness of our models. Despite achieving some success, particularly with the carbon target, the static model outperformed the joint model in predicting steel quality. This outcome suggests that integrating sequential and static data presents significant challenges. Nevertheless, our findings provide valuable insights for advancing predictive analytics within the steelmaking industry, fostering enhanced process control and optimization. Our SHAP analysis identified influential factors affecting the steelmaking process across three key targets: temperature, carbon, and phosphorus. Key features such as ore quantity, silicon levels, lime/dolomite usage, hot metal heat, and oxygen levels played a prominent role, aligning with expert expectations. In summary, this project represents a comprehensive effort underscored by valuable contributions from domain experts. Insights from past research on alternative datasets illuminate the inherent challenges of the Basic Oxygen Furnace (BOS) process, providing valuable context for our findings.

7.1 Future work

The advantages of accurate cutoff predictions in terms of energy efficiency, material usage, and environmental impact are undeniable. Previous efforts to enhance performance have faced challenges due to a lack of specificity, transparency, and limited access to extensive datasets and features. With various steel fabrication techniques in play, there is no standardized error range for targets, making it crucial to delve into contributing factors and the manufacturing process as a whole.

While domain specialists and researchers are keenly interested in understanding these elements, to date, there has been a notable absence of a dynamic sequential model, particularly one employing deep learning-based time-series modeling, capable of predicting targets at any given point during the process. Moreover, automating settings in real-time and identifying influencing factors pose additional complexities.

Furthermore, there is a pressing need to extend the model's applicability and address uncertainties stemming from factors like measurement inaccuracies, flawed modeling approaches, human errors, and the intricacies of both internal and external systems. By tackling these challenges head-on, we can unlock new avenues for enhancing process efficiency and advancing the steelmaking industry's sustainability objectives.

We tried to try creating some RNN model and checking the test MSE which is shown in figure below. We found out that RNN might perform well in future in this project. MSE measures the average squared difference between the predicted and actual values, providing a clear indication of the model's prediction error in absolute terms. This helps to understand the typical magnitude of prediction errors. The model is not complete and R2 score can be added in the future as well to the metrics. Testing with MSE can also serve as a robustness check, ensuring that the model not only fits well overall (high R2) but also performs consistently across all data points (low MSE).

```
Test MSE for t_eob: 0.43278224664708337
Test MSE for c_eob: 0.9479999457053039
Test MSE for p_eob: 0.6729336260873868
```

Figure 6: Test MSE for RNN model

```
Test MSE for t_eob: 0.4630727767944336
Test MSE for c_eob: 1.7416234016418457
Test MSE for p_eob: 1.0660016536712646
```

Figure 7: Test MSE for LSTM model

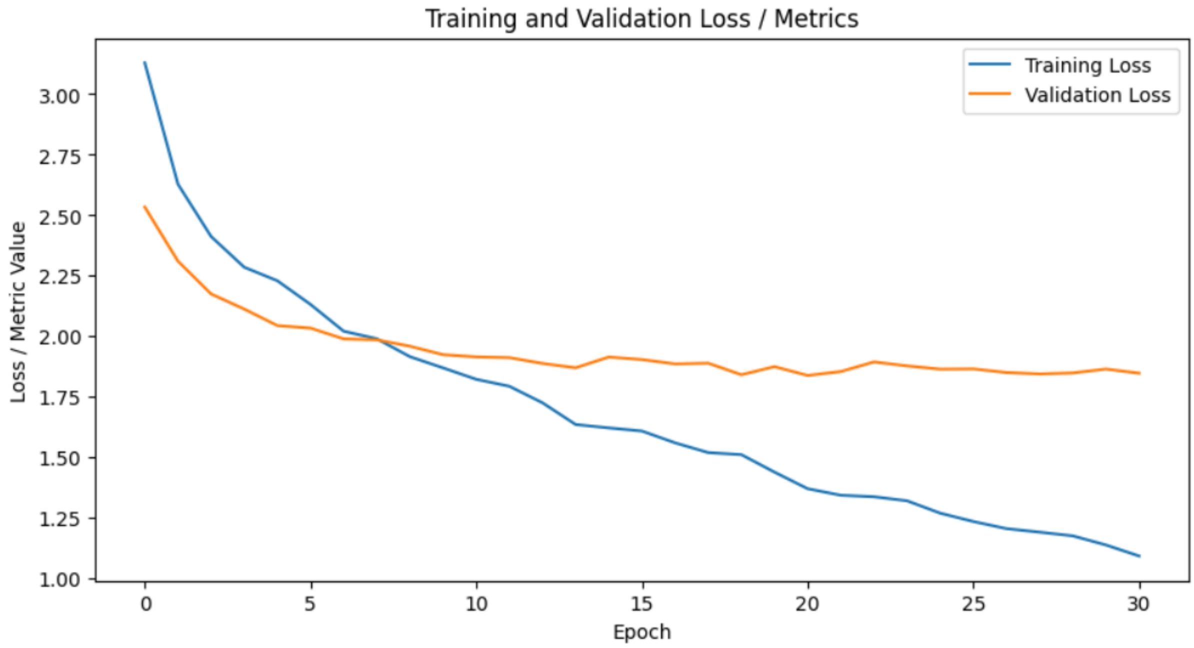

Figure 8: RNN model testing

In the future, this can be taken into account to create and check the accuracies of the model.

7.2 Acknowledgements

We would like to thank Mr. Patrik Wikstrom, and Mr. Lennart Gustavsson at SSAB, and Mr. Gunnar Mathiason from Högskolan i Skövde for providing the domain expert knowledge for this project.

7.3 Appendix

Table 2: Hyperparameters for Joint model

Hyperparameters	Selected Values	Tested Range
Learning Rate	0.001	$0.001-0.01$
Number of layers	$1\ LSTM + 3\ Dense$	$1-2\ LSTM + 2-4\ Dense$
Units per LSTM layer	128	$64,128,256$
Units per dense layer	$128,512,64$	$64,128,256,512$
Batch size	128	$16,32,64,128$
Number of epochs	500	$50,100,200,500$
Dropout rate	0.2	$0.1-0.5$
Optimization Algorithm	$Adam$	$SGD, Adam, RMSprop$
Regularization	0.001	$0.0001-0.01$
Validation monitor	val_loss	$val\_loss, val\_accuracy$
Cross validation folds	10	$5,10$

Table 3: Hyperparameters for static model

Hyperparameters	Selected Values	Tested Range
Learning Rate	0.001	$0.001-0.01$
Number of layers	$4\ Dense$	$2-4\ Dense$
Units per dense layer	$512,256,128,64$	$64,128,256,512$
Batch size	128	$64,128,256,512$
Number of epochs	500	$50,100,200,500$
Dropout rate	0.2	$0.1-0.5$
Optimization Algorithm	$Adam$	$SGD, Adam, RMSprop$
Activation Function	$Leaky\ ReLU$	ReLU, LeakyReLU, Tanh
Validation monitor	val_loss	$val\_loss, val\_accuracy$
Cross validation folds	10	$5,10$

8 References

Akundi, A., Euresti, D., Luna, S., Ankobiah, W., Lopes, A., & Edinbarough, I. (2022, February). State of Industry 5.0—Analysis and Identification of Current Research Trends. *Applied System Innovation, 5*, 27.

Bae, J., Li, Y., Ståhl, N., Mathiason, G., & Kojola, N. (2020, May). Using Machine Learning for Robust Target Prediction in a Basic Oxygen Furnace System. *Metallurgical and Materials Transactions B, 51*. doi:10.1007/s11663-020-01853-5

Bae, J., Mathiason, G., Li, Y., Kojola, N., & Ståhl, N. (2021, January). Understanding Robust Target Prediction in Basic Oxygen Furnace. *2021 The 2nd International Conference on Industrial Engineering and Industrial Management*.

Feng, K., Wang, H. B., Xu, A. J., & He, D. F. (2013, December). Endpoint temperature prediction of molten steel in RH using improved case-based reasoning. *International Journal of Minerals, Metallurgy, and Materials, 20*, 1148–1154.

Gao, C., Shen, M., Liu, X., Wang, L., & Chu, M. (2019, January). End-Point Static Control of Basic Oxygen Furnace (BOF) Steelmaking Based on Wavelet Transform Weighted Twin Support Vector Regression. *Complexity, 2019*, 1–16.

He, F., & Zhang, L. (2018, June). Prediction model of end-point phosphorus content in BOF steelmaking process based on PCA and BP neural network. *Journal of Process Control, 66*, 51–58.

Liu, Y., Hao, X., Zhang, B., & Zhang, Y. (2020, August). Simplified long short-term memory model for robust and fast prediction. *Pattern Recognition Letters, 136*, 81–86.

Liu, Z., Cheng, S., & Liu, P. (2022). Prediction model of BOF end-point P and O contents based on PCA–GA–BP neural network. *High Temperature Materials and Processes, 41*, 505–513. doi:doi:10.1515/htmp-2022-0050

Turóczy, Z., & Marian, L. (2012). Multiple Regression Analysis of Performance Indicators in the Ceramic Industry. *Procedia Economics and Finance, 3*, 509–514.

Van Houdt, G., Mosquera, C., & Nápoles, G. (2020, May). A review on the long short-term memory model. *Artificial Intelligence Review, 53*, 5929–5955.

Xie, S., & Chai, T. (1999, July). Prediction of BOF Endpoint Temperature and Carbon Content. *IFAC Proceedings Volumes, 32*, 7039–7043.

Yue, Y. J., Yao, Y. D., Zhao, H., & Wang, H. J. (2013, December). BOF Endpoint Prediction Based on Multi-Neural Network Model. *Applied Mechanics and Materials, 441*, 666–669.