UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Programa de Pós-Graduação em Ciência da Computação

Mychell Laurindo de Souza Sá

Data-Driven Soft Sensor for Prediction of 430 Steel Coil Temperature in an Annealing Line Furnace

Belo Horizonte
2024

Mychell Laurindo de Souza Sá

Data-Driven Soft Sensor for Prediction of 430 Steel Coil Temperature in an Annealing Line Furnace

Final Version

Thesis presented to the Graduate Program in Computer Science of the Federal University of Minas Gerais in partial fulfillment of the requirements for the degree of Master in Computer Science.

Advisor: Adriano Alonso Veloso

Belo Horizonte
2024

Sá, Mychell Laurindo de Souza.

S111d Data-driven soft sensor for prediction of 430 steel coil temperature in an annealing line furnace [recurso eletrônico] / Mychell Laurindo de Souza Sá - 2024.

1 recurso online (71 f. il., color.) : pdf.

Orientador: Adriano Alonso Veloso.

Dissertação (Mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação.

Referências: f. 68-71

1. Computação – Teses. 2. Aprendizado do computador – Teses. 3. Sistemas orientados a dados – Teses. 4. Simulação por computador - Processos de fabricação – Teses. 5. Gêmeos digitais (Simulação por computador) – Teses. I. Veloso, Adriano Alonso. II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Computação. III. Título.

CDU 519.6*86(043)

Ficha catalográfica elaborada pela bibliotecária Irénquer Vismeg Lucas Cruz
CRB 6/819 - Universidade Federal de Minas Gerais - ICEx

UNIVERSIDADE FEDERAL DE MINAS GERAIS

DATA-DRIVEN SOFT SENSOR FOR PREDICTION OF 430 STEEL COIL TEMPERATURE IN AN ANNEALING LINE FURNACE

MYCHELL LAURINDO DE SOUZA SA

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

Prof. Adriano Alonso Veloso - Orientador
Departamento de Ciência da Computação - UFMG

Prof. Frederico Gadelha Guimarães
Departamento de Ciência da Computação - UFMG

Doutor Tarcisio Reis de Oliveira
R&D - Aperam South America

Belo Horizonte, 23 de maio de 2024.

Documento assinado eletronicamente por **Adriano Alonso Veloso**, **Professor do Magistério Superior**, em 11/06/2024, às 09:53, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Frederico Gadelha Guimaraes**, **Professor do**

Magistério Superior, em 24/06/2024, às 22:07, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Tarcisio Reis de Oliveira**, **Usuário Externo**, em 20/08/2024, às 10:55, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3260634** e o código CRC **D02A6981**.

Referência: Processo nº 23072.227940/2024-81 SEI nº 3260634

This work is dedicated to my loving parents, their support was essential to my journey.

Acknowledgments

My special thanks goes to several individuals that made this work possible:

- Aperam South America for the cooperation in this work, providing the data, and being supportive of my commitment to the course.
- To my advisor Professor Adriano Veloso, who always put me on the right track.
- To my colleague Sandey Novy, from Aperam Quality Team, for his multiple explanations about the annealing line explored in this work.
- To my colleague Caio Moura, from Aperam R&D, for providing the physical model later implemented in this study
- To my parents for providing all the education I needed to be where I am right now.
- To my animals, who kept me company every night while I was working on my projects.

“*Prefer knowledge to wealth,*
for the one is transitory,
the other perpetual.”
(Socrates)

Resumo

Este estudo adentra no campo do sensoriamento baseado em dados dentro dos processos industriais de fabricação de aço, focando especificamente na Linha 4 de Recozimento e Decapagem da Aperam South America. O objetivo principal é desenvolver um gêmeo digital para o Pirômetro 4, um componente crítico no processo de recozimento que mede a temperatura da tira, utilizando algoritmos de aprendizado de máquina e um modelo físico para empregar uma estratégia de modelagem de caixa cinzenta.

A metodologia de pesquisa abrange uma exploração abrangente, incluindo revisão de literatura, entendimento de dados, transformação, limpeza, treinamento e avaliação de modelos. Vários algoritmos de aprendizado de máquina, como Regressão Linear, Support Vector Machines, Random Forest, e XGBoost, são avaliados por sua capacidade de criar um regressor capaz de prever leituras de temperatura e replicar o comportamento do pirômetro. Além disso, o conhecimento específico do domínio é integrado para construir modelos híbridos com o objetivo de melhorar a precisão da predição.

Por meio de uma avaliação e comparação minuciosas desses modelos, conhecimentos valiosos sobre suas forças, limitações e aplicações potenciais são obtidos. O estudo enfatiza a importância de empregar uma abordagem de modelagem híbrida, que combina modelos impulsionados pela física com técnicas de aprendizado de máquina, para desenvolver gêmeos digitais robustos e precisos. Por fim, a pesquisa visa contribuir para o avanço de soluções baseadas em dados em ambientes industriais, facilitando a tomada de decisões e a otimização de processos.

Palavras-chave: sensores virtuais; orientado a dados; aprendizado de máquina; gêmeo digital.

Abstract

This study delves into the realm of data-driven soft sensing within industrial steel manufacturing processes, specifically focusing on Aperam South America's Annealing and Pickling Line 4. The primary objective is to develop a digital twin for Pyrometer 4, a critical component in the annealing process that measures strip temperature, by utilizing machine learning algorithms and a physical model to employ a grey-box modeling strategy.

The research methodology encompasses a comprehensive exploration, including a literature review, data understanding, transformation, cleaning, model training, and evaluation. Various machine learning algorithms, such as Linear Regression, Support Vector Machines, Random Forest, and XGBoost, are assessed for their ability to create a regressor capable of predicting temperature readings and replicating the pyrometer behavior. Moreover, domain-specific knowledge is integrated to construct hybrid models aimed at enhancing predictive accuracy.

Through thorough evaluation and comparison of these models, valuable insights into their strengths, limitations, and potential applications are gained. The study emphasizes the importance of employing a hybrid modeling approach, which combines physics-driven models with machine learning techniques, to develop robust and accurate digital twins. Ultimately, the research aims to contribute to the advancement of data-driven solutions in industrial settings, facilitating decision-making and process optimization.

Keywords: soft sensor; data driven; machine learning; digital twin.

List of Figures

List of Tables

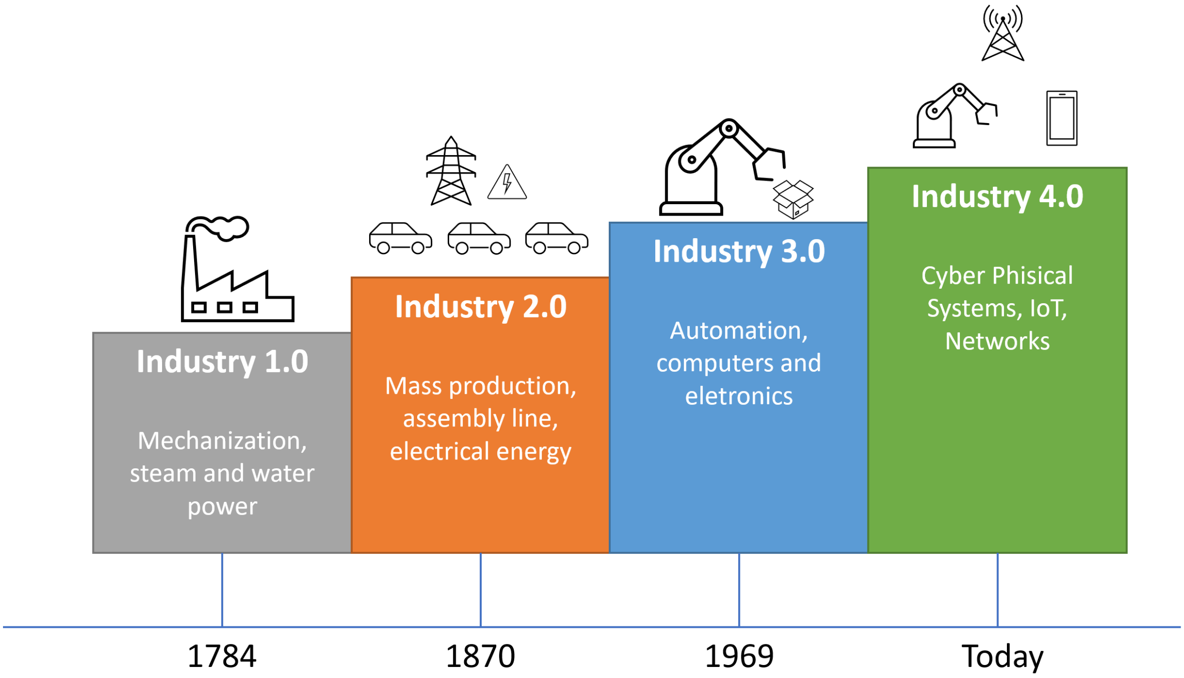
Contents

Chapter 1

Introduction

The concept of Industry 4.0 originated in Germany and has spread extensively throughout the world, primarily due to its ability to enhance production efficiency through smart services in smart factories [Sisinni et al., 2018]. It is regarded as the fourth industrial revolution, characterized by the capacity of industrial plant components to communicate with the entire facility through the Internet, thereby exchanging substantial amounts of information. This is in stark contrast to Industry 3.0, where such components were mostly isolated and lacked awareness of their surroundings [Wagner et al., 2017]. The evolution of the industry and the major breakthroughs it has brought are illustrated in Figure 1.1.

Figure 1.1: Infographic of the Four industrial revolutions and its advances.

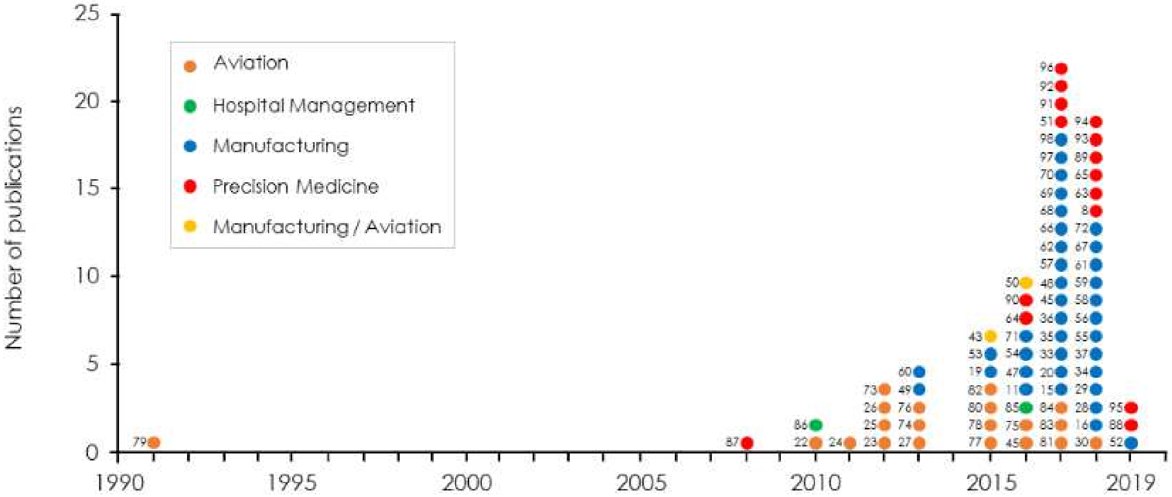

Industry 1.0
Mechanization, steam and water power
Industry 2.0
Mass production, assembly line, electrical energy
Industry 3.0
Automation, computers and eletronics
Industry 4.0
Cyber Phisical Systems, IoT, Networks
1784
1870
1969
Today

Industry 4.0, the fourth industrial revolution, has three fundamental components: the Internet of Things (IoT), Cyber-Physical Systems (CPS), and Smart Factories [Hermann et al., 2016]. The IoT is characterized by ubiquitous connectivity, facilitating information exchange among connected devices, including data on themselves and their surroundings. This disruptive technology has the potential to address a range of current issues, including smart cities, manufacturing, pollution, and health [Sisinni et al., 2018].

To interconnect these devices, CPSs play a crucial role. CPSs are ”integrations of computation and physical processes” and involve embedded computers and networks that monitor and control physical processes through feedback loops, where physical processes influence computations and vice versa [Hermann et al., 2016]. CPSs provide digital descriptions of real-world physical objects, which are then stored and modeled as ”digital twins” with their own identities in the virtual space [Sisinni et al., 2018].

Digital twins, along with IoT, data mining, and machine learning technologies, have opened up new possibilities for revolutionizing today’s manufacturing, making it more intelligent [Min et al., 2019]. A review of papers on the development of digital twins across multiple domains has been conducted by Barricelli et al. [2019], and the findings are presented in Figure 1.2.

Figure 1.2: Timeline of the papers analyzed for this study. The labels are the reference numbers. Each point in the chart has a color that refers to the application domain in which the study is framed.



Author: Barricelli et al. [2019]

1.1 Motivation

The extensive interest and increasing adoption of Industry 4.0 concepts by companies have resulted in a plethora of opportunities to be explored. There exists a multitude of interconnected devices, sensors, cameras, processes, and systems that generate vast volumes of data. [Gartner, 2017] analyst Thomas Oestreich points out that the challenge that companies face in this 4.0 scenario is how to handle this massive volume of data, and to unleash its full potential, new algorithms need to be developed. The aim of this study is to leverage these new technologies to create a data-driven soft sensor for a physical pyrometer of an existing Annealing and Pickling Line Furnace.

Operational and process control decisions in the company are based on the values of strip temperature readings given by pyrometers. If these readings are outside the specified limits, the product could become unusable or require modification for another application. Several factors can affect the quality of the product, including the possibility of a fault in the pyrometer or one of the furnace zones not providing the desired output, requiring an operator to manually change it. The operator has to perform trial and error tests to check if the strip temperature is meeting the required specifications.

Although mathematical models could be applied, they tend to be more general in nature and do not account for all the specificities of the production line. Hence, having a digital twin could aid in simulating the output based on inputs without a trial and error approach, while it still could take advantage of existing physical models. It might also be used as a reference to verify the data quality from the physical sensor, and automatically adjust the parameters.

1.2 Objectives

The primary aim of this project is to develop a data-driven soft sensor capable of measuring the temperature of a steel strip in an annealing line by exploring various machine-learning models for regression, exploring a grey-box approach. The project encompasses specific objectives:

- Conduct analysis of regression techniques and models suitable for building a virtual sensor in an industrial setting.
- Verify the capability of a grey-box modeling strategy.
- Compare the performance between models using the proposed evaluation metrics, Root Mean Squared Error (RMSE).
- Build a data-driven soft sensor that is a digital twin of the pyrometer of the studied annealing line furnace.

Chapter 2

Literature Review

The literature review chapter of this research provides an overview of essential domains central to the experiment. It begins with an examination of the RB4 Annealing and Pickling Line, offering insights into its operation and highlighting the significance of the Pyrometer within this context.

Machine Learning is the subsequent focus, where its principles, techniques, and applications are explored. It serves as a foundational element to create an accurate digital twin that mirrors the Pyrometer's behavior.

The exploration extends to Digital Twins, their nature, and their potential to revolutionize the comprehension and control of intricate industrial processes, and how Machine Learning plays an important role in leveraging practical on-site data collection to build AI-enabled twins.

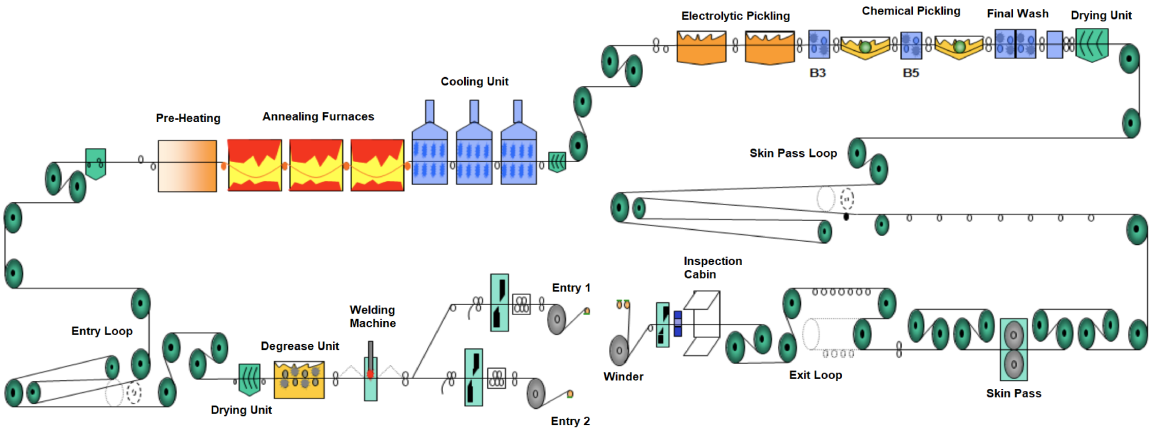
Lastly, it delves into the Grey-Box Modeling Approach. This approach mixes physics-driven models with Machine Learning methods to construct data-driven digital twins. These soft sensors play a pivotal role in bridging the divide between theoretical models and real-world observations.

2.1 Annealing and Pickling Line

Annealing is a heat treatment process in steel production, wherein the material undergoes high-temperature exposure to achieve its desired properties. As explained by Askeland and Pradeep [2009], cold rolling results in an increase in hardness due to the growth of dislocation density, and annealing can be employed to enhance ductility and counter these effects. Moreover, by controlling the thermo-mechanical processing, it is possible to obtain materials with improved mechanical properties and in usable shape. At Aperam, the RB4 production line is dedicated to the Annealing process, which comprises three furnaces. A schematic of the complete production line is illustrated in Figure 2.1.

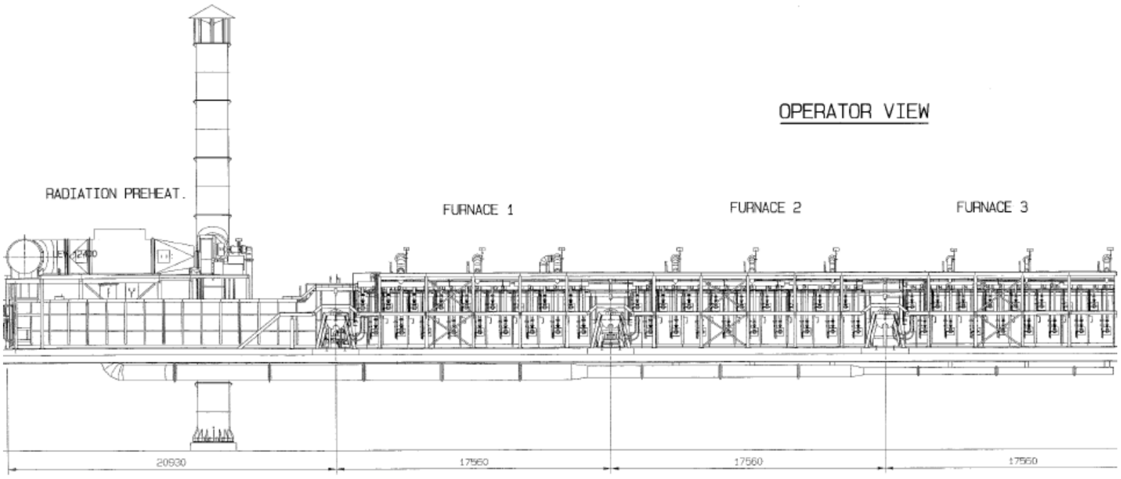
In RB4, the annealing process is carried out continuously, during which the strip

Figure 2.1: Flow chart of the Annealing and Pickling Line 4 (RB4).



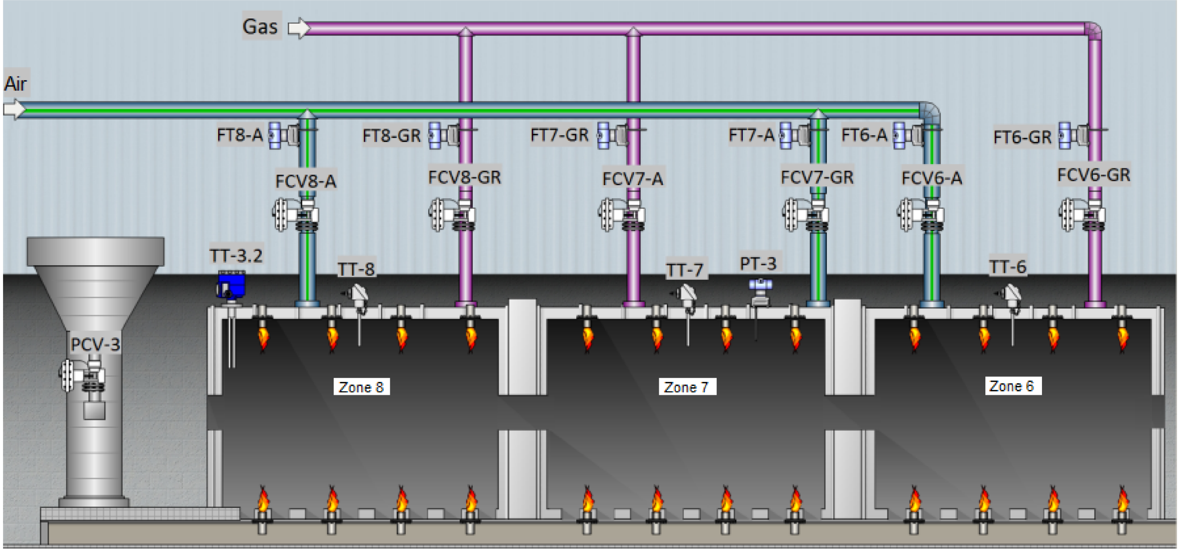
is subjected to temperature treatment through three consecutive furnaces. Prior to the first furnace, there is a Pre-heating zone, followed by the first furnace, which comprises zones 1, 2, and 3. The first furnace aims to recover grain through preheating. The second furnace consists of zones 4, 5, and 6, which are responsible for the recrystallization of the grains. Finally, zones 7, 8, and 9 in the third furnace contribute to grain growth through heating and soaking. An overview of the furnaces can be seen in Figure 2.2.

Figure 2.2: Side view of the annealing furnace of RB4.



Ensuring that the strip temperature profile meets the required specifications is of paramount importance to attain the desired mechanical properties and prevent the occurrence of defects. In his thesis on "Optimal Control for Continuous Annealing Furnace," Teixeira da Silveira [2009] provides a detailed account of the employment of a pyrometer in the exit of a sister line of RB4 in Aperam, called RB1. The "TT-3.2" pyrometer, as illustrated in Figure 2.3, is used to monitor this important parameter of the process and provide real-time feedback for the necessary process interventions.

Figure 2.3: Detailed components of RB1's Furnace 3.

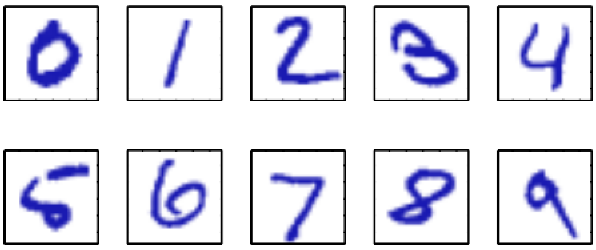


Source: Teixeira da Silveira [2009]

2.2 Machine Learning

According to Bengio [2009], allowing computers to model our world to a degree that we perceive as intelligence has been the focus of research for half a century. Achieving this entails digitally storing data in any format. However, manually formalizing all information for machines to generalize to novel contexts would be arduous. As a result, researchers have turned to learning algorithms. This field of study, known as Machine Learning, has been defined by Murphy [2013] as a set of methods that can automatically detect patterns in data and use them to predict future events or make other types of decisions. One noteworthy example is the human ability to read various handwritten symbols through generalization. According to Bishop [2006], the essential aspect of a machine learning algorithm is its capacity for generalization. To demonstrate this, he presents an algorithm capable of accurately categorizing a digit based on an image, as shown in Figure 2.4.

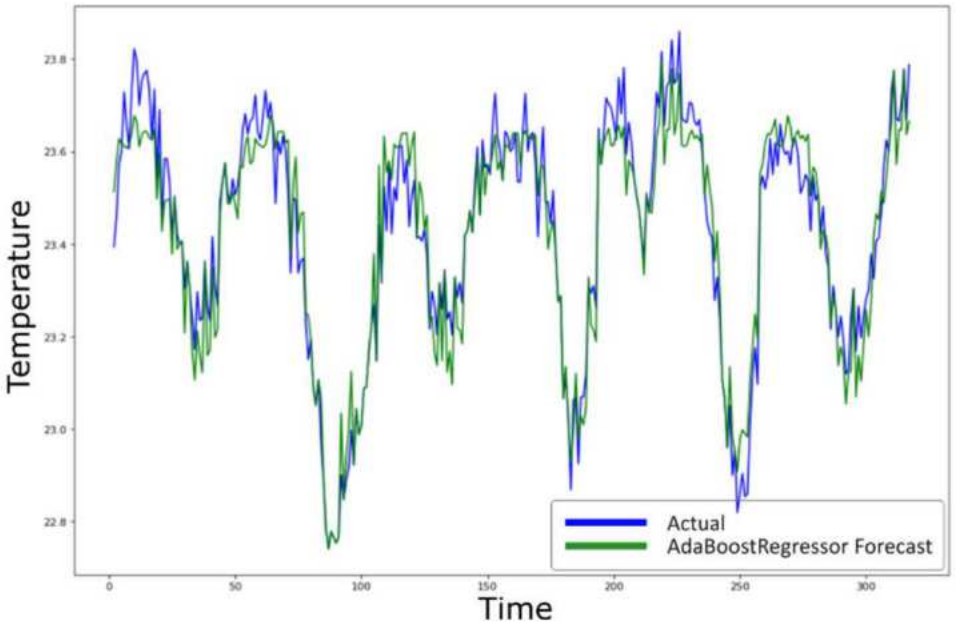
Figure 2.4: Examples of hand-written digits taken from US zip codes.



Source: Bishop [2006]

Müller and Guido [2016] define Machine Learning algorithms that learn from input/output as supervised learning algorithms. These algorithms are supervised by a "teacher" who provides target output for each example they learn from. If a dataset can be created with the target outcomes and the problem can be formulated as a supervised learning one, it can be solved using this strategy. Bishop [2006] categorizes supervised problems into two categories. The first category is classification, which aims to assign each input vector to one of a finite number of discrete categories, as seen in the digit recognition problem. The second category is regression, which involves predicting an output made up of continuous variables. In a study conducted by Taufiqurrahman et al. [2020], a regression algorithm known as AdaBoost was used for water temperature forecasting in an Aquaponic Ecosystem. The resulting model is illustrated in Figure 2.5.

Figure 2.5: AdaBoost regression predictive data compared with actual data.



Source: Taufiqurrahman et al. [2020]

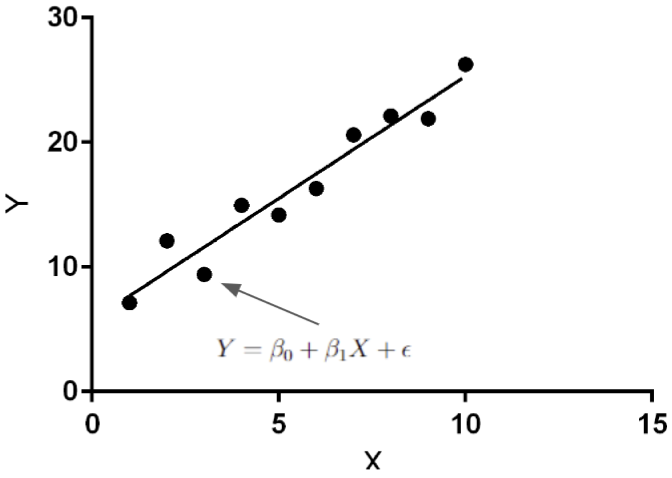
2.2.1 Linear Regression

Linear regression is a classic and widely used statistical technique for modeling the relationship between a dependent variable and one or more independent variables. The fundamental assumption is that this relationship can be approximated by a linear equation. In a simple linear regression, the model is represented as:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Montgomery et al. [2012] explains it in detail, stating that Y is the dependent variable, X is the independent variable, β_0 is the intercept, β_1 is the slope, and ϵ represents the error term. The goal of linear regression is to estimate the coefficients β_0 and β_1 that minimize the sum of squared differences between the observed and predicted values. These elements are illustrated in Figure 2.6.

Figure 2.6: Linear Regression elements.



Linear Regression can be expanded to predicting a dependent variable that is based on multiple predictors. Montgomery et al. [2012] presents the general form of a multiple linear regression model with n predictors, which is given by:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon$$

The goal of multiple linear regression is to estimate the coefficients $(\beta_0, \beta_1, \ldots, \beta_n)$ that minimize the sum of squared differences between the observed and predicted values.

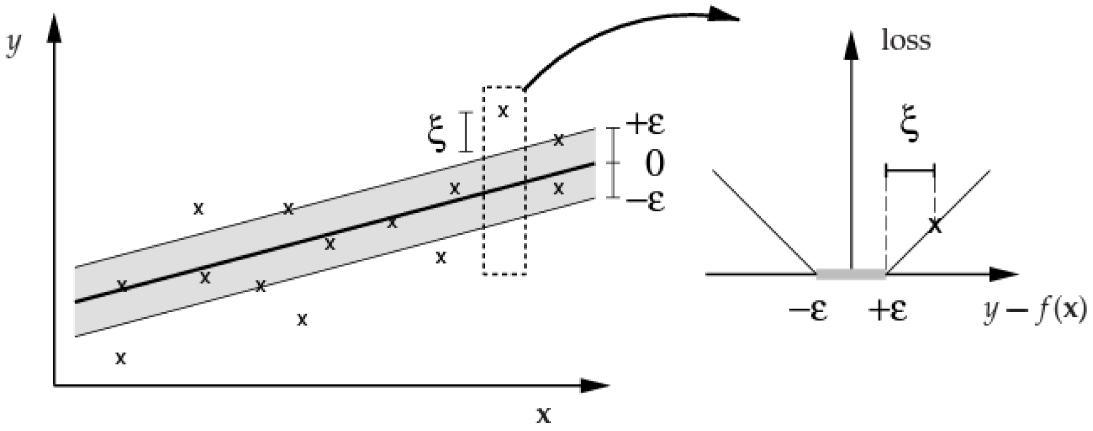
Linear regression is characterized by its simplicity and interpretability. It provides insights into the strength and direction of relationships between variables. However, its effectiveness may be limited when dealing with complex, non-linear relationships.

2.2.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful machine learning algorithm used for both classification and regression tasks. It operates by finding the hyperplane that best separates the data points of different classes.

Support Vector Regression (SVR) is a regression technique that extends the principles SVMs to predict continuous output. SVR involves finding a hyperplane in a high-dimensional space that captures the relationship between input features (**X**) and the continuous target variable (y). Scholkopf and Smola [2001] details the fundamental idea of the algorithm, which is to maximize the margin around the predicted values while allowing for a certain degree of error within an epsilon-insensitive zone, illustrated in Figure 2.7.

Figure 2.7: In SV regression, a tube with radius ϵ is fitted to the data. The trade-off between model complexity and points lying outside of the tube.



Source: Scholkopf and Smola [2001]

The SVR model is represented as:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$$

subject to the constraints:

$$|y_i - f(\mathbf{x}_i)| \leq \varepsilon$$

.

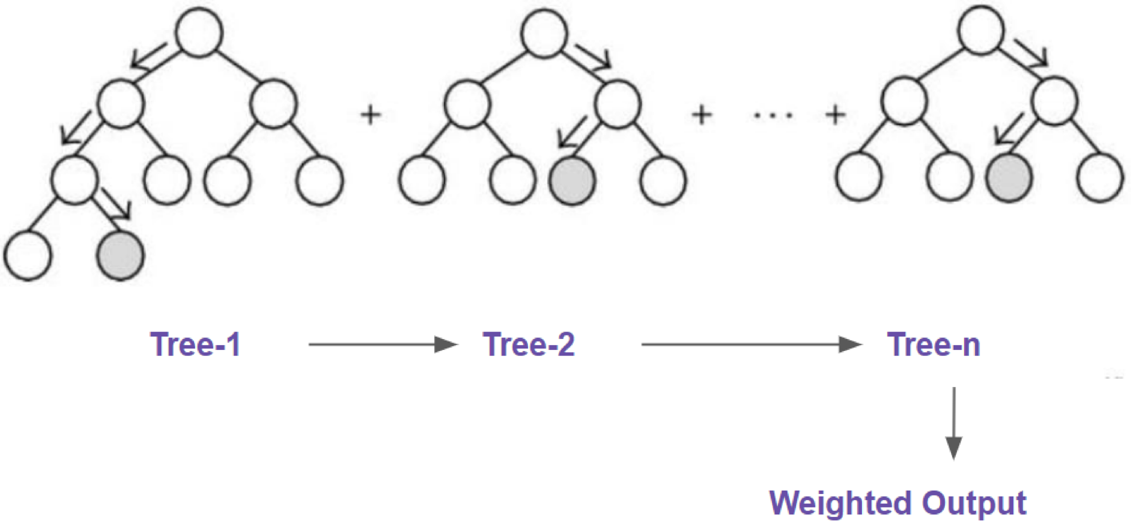
SVMs for regression are effective in capturing complex relationships and are particularly useful when dealing with datasets with high dimensionality.

2.2.3 XGBoost

Boosting-based algorithms, often recognized as a potent learning paradigm over the past two decades Hastie et al. [2009], belong to the ensemble model family. These algorithms strategically amalgamate multiple weak models to create a robust and accurate predictive model, capitalizing on the principle that ”unity is strength.”.

XGBoost, short for eXtreme Gradient Boosting, proposed and explained by Chen and Guestrin [2016], expands on the boosting strategy by incorporating gradient boosting principles. It involves iteratively adding weak learners to the ensemble, with each new learner correcting errors made by the existing ensemble as depicted in Figure 2.8, and employing gradient descent optimization to minimize the loss function. XGBoost takes this concept further by introducing a regularized objective function and employing a more sophisticated mechanism for tree construction. This algorithm has demonstrated exceptional scalability and performance, making it a preferred choice in various machine-learning applications.

Figure 2.8: XGBoost deployment strategy.

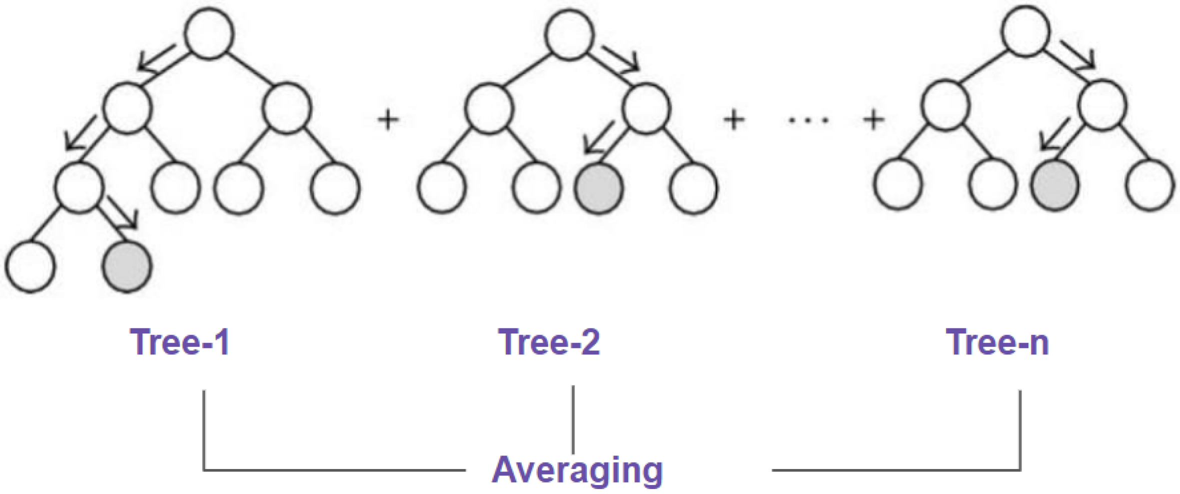


2.2.4 Random Forest

Another ensemble model that will be evaluated in this study is the Random Forest. It is an ensemble learning method that belongs to the family of bagging algorithms, a term derived from ”Bootstrap Aggregating.” The primary idea behind bagging is to train multiple models independently on different subsets of the training data and then combine their predictions to produce a more robust and accurate model. It was introduced by Breiman [2001], and extends the bagging strategy by using decision trees as the base learners.

In a Random Forest, the bagging strategy involves constructing multiple decision trees, each trained on a randomly sampled subset of the training data. The random sampling is performed with replacement, meaning that the same data point can appear multiple times or not at all in a given subset. Additionally, at each node of the tree, a random subset of features is considered for splitting, providing diversity among the trees. This randomness helps mitigate overfitting and decorrelates the individual trees, as stated by Hastie et al. [2009]. This all results in an ensemble model as represented in Figure 2.9 that is more resilient and generalizes well to unseen data.

Figure 2.9: Random Forest deployment strategy.

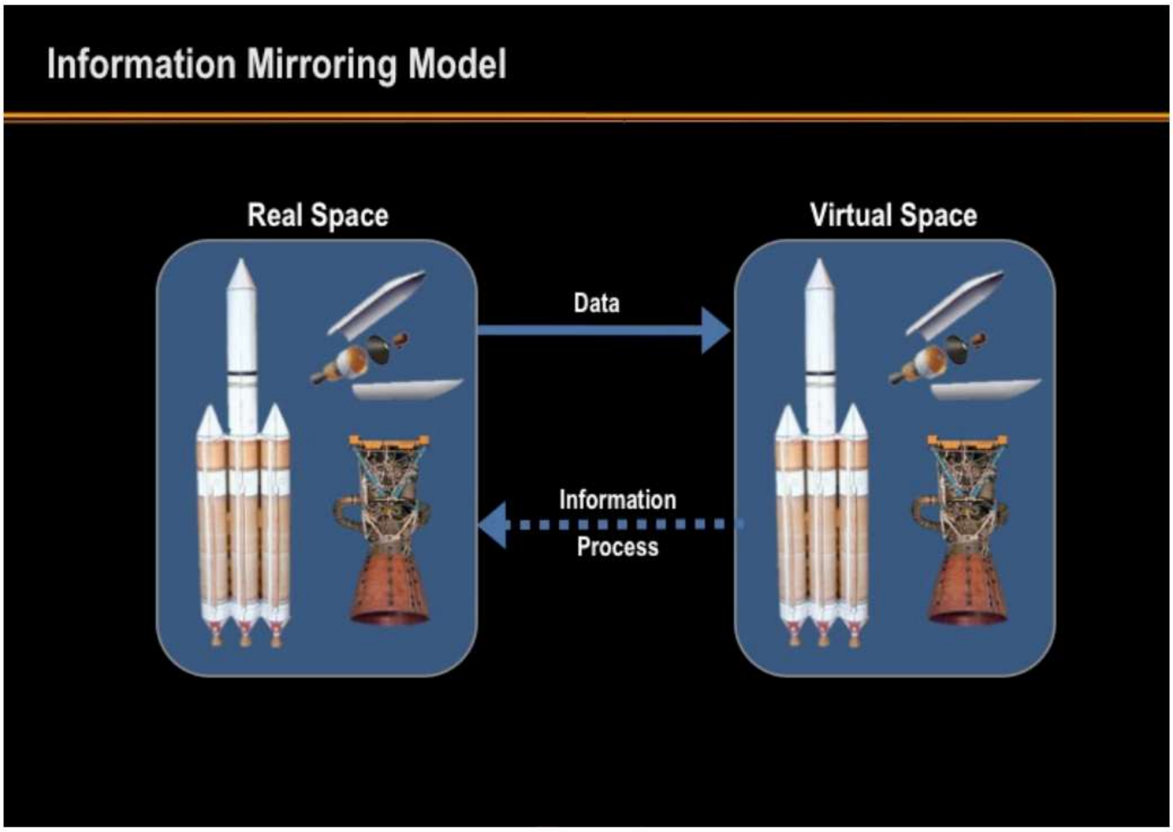


2.3 Digital Twin

In 2003, the terminology of digital twin was first introduced in the presentation about Product Lifecycle Management (PLM) by Grieves [2014] at the University of Michigan. He documented some of his ideas in a white paper where he presented a concept model of a digital twin comprising three components: the physical product, the virtual product, and data flux that binds them together. The concept is graphically depicted in Figure 2.10.

Numerous explanations have been offered for the term "digital twin," but the most noteworthy and commonly used among scholars is similar to the one put forward by Glaessgen and Stargel [2012] in a NASA publication. In this paper, a digital twin is characterized as "an integrated multiphysics, multiscale, probabilistic simulation of an as-built vehicle or system that uses the best available physical models, sensor updates,

Figure 2.10: Information Mirroring Model.



Information Mirroring Model

Real Space

Virtual Space

Data

Information Process

Source: Grieves [2014]

fleet history, etc., to mirror the life of its corresponding flying twin". This definition was strongly linked to a virtual representation of a space asset. A more recent interpretation by Liu et al. [2018] proposed a similar definition, describing a digital twin as "a living model of the physical asset or system that continually adapts to operational changes based on collected online data and information and can forecast the future of the corresponding physical counterpart". What is consensus, and acknowledged by Wright and Davidson [2020], is that a key aspect that sets a digital twin apart from a mere model is its association with an existing physical object.

2.4 AI-enabled Digital Twins

Machine Learning (ML) has emerged as a fundamental technology for Digital Twins (DTs), offering a variety of algorithms that form the basis for constructing models. According to Huang et al. [2021], ML models can be trained as surrogate models to enhance the efficiency of complex numerical simulations at both the process and material scales, as well as to expedite production ramp-up and create soft sensors for inline-quality monitoring. In the same paper, it was presented a list of applications where Machine Learning has been applied to process and material applications of Digital Twins. Table 2.1 presents various supervised learning methods and algorithms employed by multiple authors.

Key Methods	Application Case	Ref
PIO, SVM	Prediction of surface roughness	Zhao et al. [2022]
Ensemble methods, ANN	Modeling of the rheological behavior of drilling fluids	Samnejad et al. [2020]
ANN	Prediction of stress and fatigue damage (FE surrogate) of flexible risers	Repalle et al. [2020]
DNA-based computing, Markov chain	Prediction of surface roughness	Ghosh et al. [2020]
SVM	Prediction of the occurrence of defects in metal AM (LPBF, LMD)	Gaikwad et al. [2020b]
CNN, LSTM, RNN	Quality assurance in metal AM (LPBF)	Gaikwad et al. [2020a]
CART	Prediction of additive manufacturability	Ko et al. [2019]
HMM	Model adaptivity and quality assessment of laser material removal processes	Stavropoulos et al. [2020]
CNN, transfer learning	Detection of dry points in the production of carbon fiber reinforced plastics	Stieber et al. [2020]
AdaBoost, XGBoost, RF	Prediction of temperature distribution of thermoplastic composites	Hürkamp et al. [2020]
DNN	FE surrogate for a composite textile draping process	Pfrommer et al. [2018]
PML	Prediction of material properties of a composite material system	Ghanem et al. [2020]

Table 2.1: Summary of AI-enabled DTs in smart manufacturing: process and material level. Source: Huang et al. [2021]

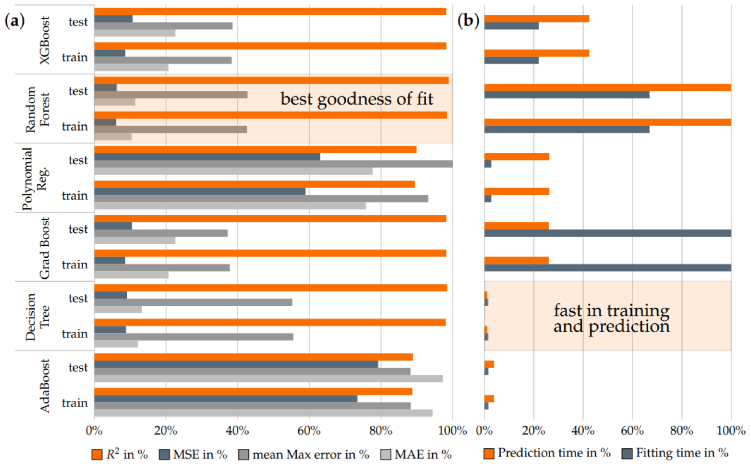
2.5 Grey-Box Modelling Approach

Hürkamp et al. [2020] worked on the development of a physics-based digital twin, they noted that while complex FEM simulations can provide virtual insights into structural characteristics and interface conditions, including factors such as temperature distribution, they often incurred substantial computational costs, rendering them impractical for deployment as a digital twin.

To address this challenge, their research turned to the development of surrogate models, which could rapidly and accurately navigate an extensive parameter space. These surrogate models exhibited the remarkable ability to generate suitable representations within milliseconds, thus overcoming the computational limitations of the traditional method. The construction of these surrogate models was achieved through a synergistic blend of FEM simulations and machine-learning.

To assess the efficacy of their proposed approach, a comparative analysis involving six distinct data-driven methods was undertaken, presented in Figure 2.11. This rigorous evaluation affirmed the overall feasibility of their approach, with notable promise demonstrated by the Random Forest and Decision Tree methods.

Figure 2.11: Comparison of model evaluation metrics without outliers for all investigated data-driven approaches for the demonstrator structure: (a) Error (R2, mean squared error (MSE), mean MAX error, mean absolute error (MAE)); (b) times for training and prediction.



Source: Hürkamp et al. [2020]

Looking into this potential for the proposed soft sensor of this work, Aperam Research Department was contacted, and it has been developing a physical model for strip temperature. Access was provided to the second version of this model.

Aperam model aims to simulate the heat transfer process occurring within the Annealing Furnace of RB4. This modeling approach draws inspiration from the strategies outlined by [Bitschnau and Kozek, 2009] in his research paper ”Modeling and Control of an Industrial Continuous Furnace”, in which it emphasized that solving the three-dimensional heat transfer problem numerically requires significant computational resources. To mitigate this computational intensity, he proposed simplifying the problem using a one-dimensional heat equation and making specific assumptions, thus reducing the computational burden. His approach also involved discretizing the strip into N spatial and temporal elements.

Recognizing the computational challenges associated with employing numerical solutions, the Aperam team adopted a similar modeling strategy based on Bitschnau's work. This strategy streamlines the modeling process while addressing the computational complexities inherent in simulating heat transfer within the annealing furnace.

Chapter 3

Methodology

This chapter provides a detailed description of the methodologies employed in this study, focusing on the data aspects, modeling strategy, and experimental setup. The first section delves into the data, covering its acquisition, understanding, transformation, and cleaning processes. Additionally, it explores the pyrometer data and independent variables to establish a solid foundation for subsequent analyses. The next section outlines the modeling approach, detailing the overall strategy and the processes involved in model training and validation. Finally, section three describes the setup of the experiments, ensuring that the procedures are thoroughly documented and reproducible.

3.1 Data

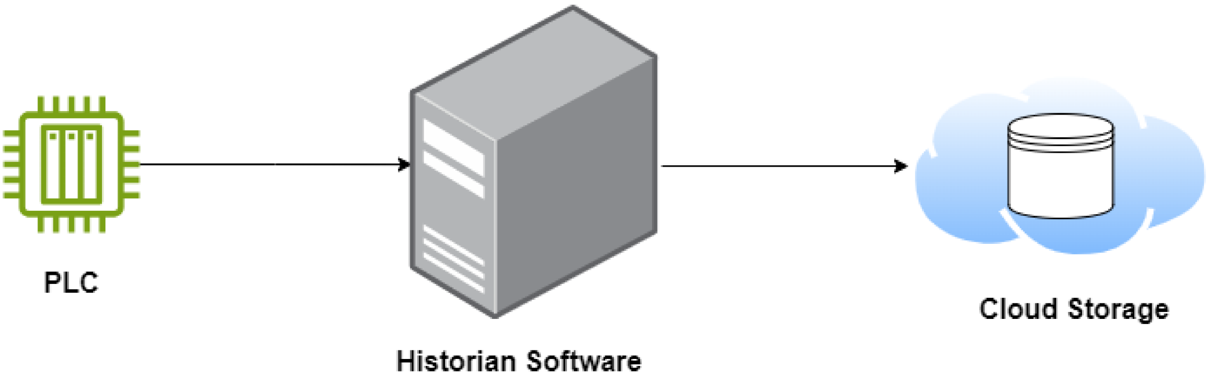
Data is the fundamental component of a Digital Twin, and its preparation was a significant workload in this research. The present section is dedicated to explaining its meaning and exploring each step involved in its processing, transformation, and cleaning. Comprehending these characteristics is essential for conducting meaningful analyses and providing valuable insights in the subsequent chapters of the thesis.

3.1.1 Data Acquisition

The dataset used in this research comprises time series and descriptive variables, which were collected from the Furnace of a Continuous Annealing line (RB4) of Aperam South America plant, spanning from January 1st, 2022, to December 31st, 2022.

The equipment features multiple sensors that are distributed throughout its structure, and all of them are connected to a Programmable Logic Controller (PLC). The PLC processes the data and sends it to the Historian Software, which compresses the data before storing it in a database. The frequency at which the data is stored is defined by the user. Subsequently, the data is sent to the cloud via a Representational State Transfer (REST) Application Programming Interface (API) and stored in big data storage as illustrated in Figure 3.1.

Figure 3.1: Data acquisition flow.



3.1.2 Data Understanding

The raw dataset includes 38 process variables, which are essential for analyzing and understanding the industrial processes under investigation. These variables predominantly include temperature, air, and gas flow parameters of various zones within the production line. Additionally, the dataset incorporates measurements of furnace pressure and line speed, which play crucial roles in the overall process dynamics. Also, there are variables related to the product, namely thickness and width, were also included in the dataset as they are significant quality attributes. Finally, the main focus of the thesis study is the Pyrometer, which serves as the primary focus of the digital twin replication. It measures temperature at the exit of the furnace and provides valuable insights into the thermal behavior of the strip inside the furnace.

To ensure a high level of temporal resolution, the data was collected at a frequency of 1 second. This granular sampling rate allows for detailed analysis and examination of the industrial processes and their associated variables. All the data entries within the dataset are exclusively related to a particular steel grade within the Stainless Steel family. This intentional selection ensures that the analysis and findings of the thesis study are specifically modeled to the characteristics and requirements of this particular product.

A compilation of the variables contained within the dataset is presented in Table 3.1, accompanied by their corresponding engineering unit, data type, and tag.

Variable	Unit	Type	Tag
Product ID	-	String	UM
Steel Strip Thickness	mm	Float	Esp
Steel Strip Width	mm	Float	Larg
Temperature from Pyrometer 4	°C	Float	Pir_4
Line Speed	m/min	Float	Vel
Temperature of Pre-Heating 1	°C	Float	Pre_1
Temperature of Pre-Heating 2	°C	Float	Pre_2
Temperature of Zone 1 to 3 of Furnace 1	°C	Float	Z1, Z2, Z3
Temperature of Zone 4 to 6 of Furnace 2	°C	Float	Z4, Z5, Z6
Temperature of Zone 7 to 9 of Furnace 3	°C	Float	Z7, Z8, Z9
Air flow of Zone 1 to 3 of Furnace 1	Nm^3/h	Float	VA_Z1, VA_Z2, VA_Z3
Air flow of Zone 4 to 6 of Furnace 2	Nm^3/h	Float	VA_Z4, VA_Z5, VA_Z6
Air flow of Zone 7 to 9 of Furnace 3	Nm^3/h	Float	VA_Z7, VA_Z8, VA_Z9
Glas Flow of Zone 1 to 3 of Furnace 1	Nm^3/h	Float	VG_Z1, VG_Z2, VG_Z3
Glas Flow of Zone 4 to 6 of Furnace 2	Nm^3/h	Float	VG_Z4, VG_Z5, VG_Z6
Glas Flow of Zone 7 to 9 of Furnace 3	Nm^3/h	Float	VG_Z7, VG_Z8, VG_Z9
Pressure of Pre-Heating	mmCA	Float	P_Pre
Pressure of Furnace 1	mmCA	Float	P1
Pressure of Furnace 2	mmCA	Float	P2
Pressure of Furnace 3	mmCA	Float	P3

Table 3.1: Description of Dataset Variables.

3.1.3 Data Transformation

The production process of the line involves the movement of the strips through various sections, starting from the pre-heating section and progressing towards Furnace 3, with a total line length of 73.13 meters. The pre-heating section spans 20.59 meters, while Furnaces 1 and 2 have lengths of 17.56 meters each, and Furnace 3 spans 17.42 meters.

The objective of the data transformation step was to produce a dataset that represented each meter of processed material. This was achieved by calculating the average values of the variables measured by the line sensors, from the beginning of the strip's journey inside the furnace (starting from the pre-heating section) to the end of the line. Additionally, the average temperature measured by the pyrometer for each meter was computed, as it represents the target variable that we aim to replicate through modeling.

To determine the position of each meter within the furnace, a combination of line tracking and speed information was utilized. As each coil enters the furnace, a tracking sensor assigns a unique Product ID to the material being processed. By dividing the line speed (given in meters per minute) by 60, we obtained the rate of material progression per second. This enabled us to determine the position of each meter inside the furnace at any given time. Utilizing this positional information, measurements for each meter were collected while it passed through zones where the sensors were located.

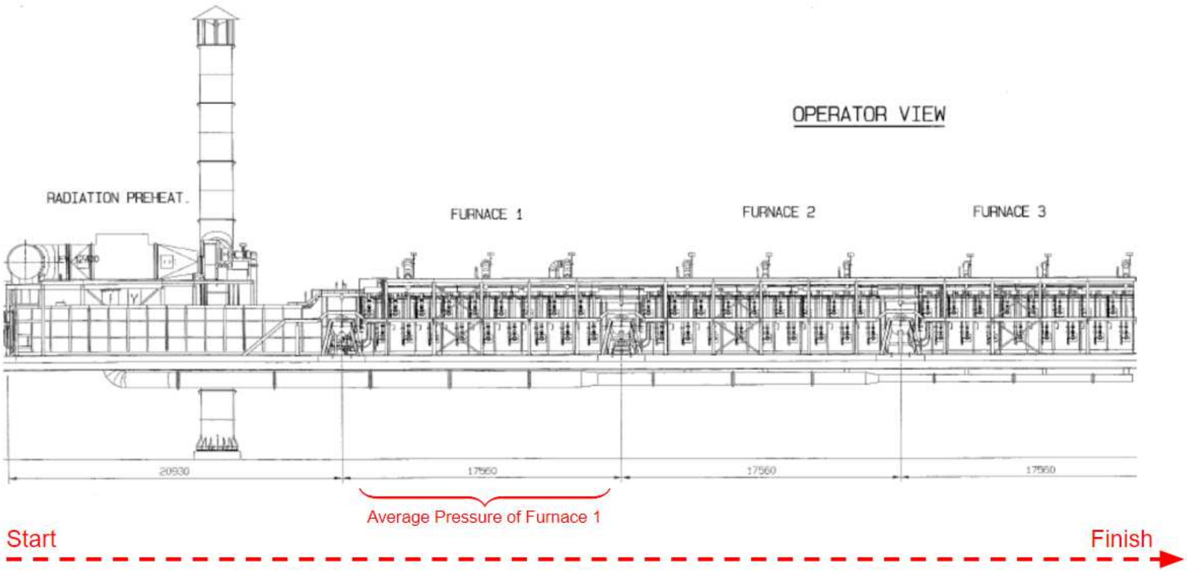
The data processing procedure generates a representation of the material's characteristics as it progresses through the production line. This representation facilitates subsequent modeling and the replication of the target variable.

To illustrate this process, let's consider the calculation of the average pressure in Furnace 1. As the strip starts its journey from the pre-heating section and traverses the furnace, meter by meter, the average furnace pressure is computed while each of them are inside Furnace 1, as exemplified in Figure 3.2. These calculated averages are then appended to the transformed database. It's important to note that this same strategy is employed for calculating every variable associated with each meter, with the only variation being the range reference. Detailed information on all variables and their respective calculation range can be found in Table 3.2.

Variable	Aggregation Range (m)
Temperature of Pre-Heating 1	[0 , 3.5]
Temperature of Pre-Heating 2	[3.5 , 20.59]
Temperature, Air Flow and Gas Flow of Zone 1	[20.59 , 26.5]
Temperature, Air Flow and Gas Flow of Zone 2	[26.5 , 31.5]
Temperature, Air Flow and Gas Flow of Zone 3	[31.5 , 37.75]
Temperature, Air Flow and Gas Flow of Zone 4	[37.75 , 44]
Temperature, Air Flow and Gas Flow of Zone 5	[44 , 49]
Temperature, Air Flow and Gas Flow of Zone 6	[49 , 55.25]
Temperature, Air Flow and Gas Flow of Zone 7	[55.25 , 61.5]
Temperature, Air Flow and Gas Flow of Zone 8	[61.5 , 66.5]
Temperature, Air Flow and Gas Flow of Zone 9	[66.5 , 71.6]
Pressure of Pre-Heating	[0 , 20.59]
Pressure of Furnace 1	[20.59 , 37.75]
Pressure of Furnace 2	[37.75 , 55.25]
Pressure of Furnace 3	[55.25 , 71.6]
Temperature from Pyrometer 4	[70.6 , 72.6]

Table 3.2: Aggregation range of process variables.

Figure 3.2: Strip movement across the Furnace.



3.1.4 Data Cleaning

Data cleaning plays a pivotal role in ensuring the accuracy and reliability of the subsequent modeling processes. It serves as a critical step in the data preprocessing pipeline, aimed at enhancing the overall model quality.

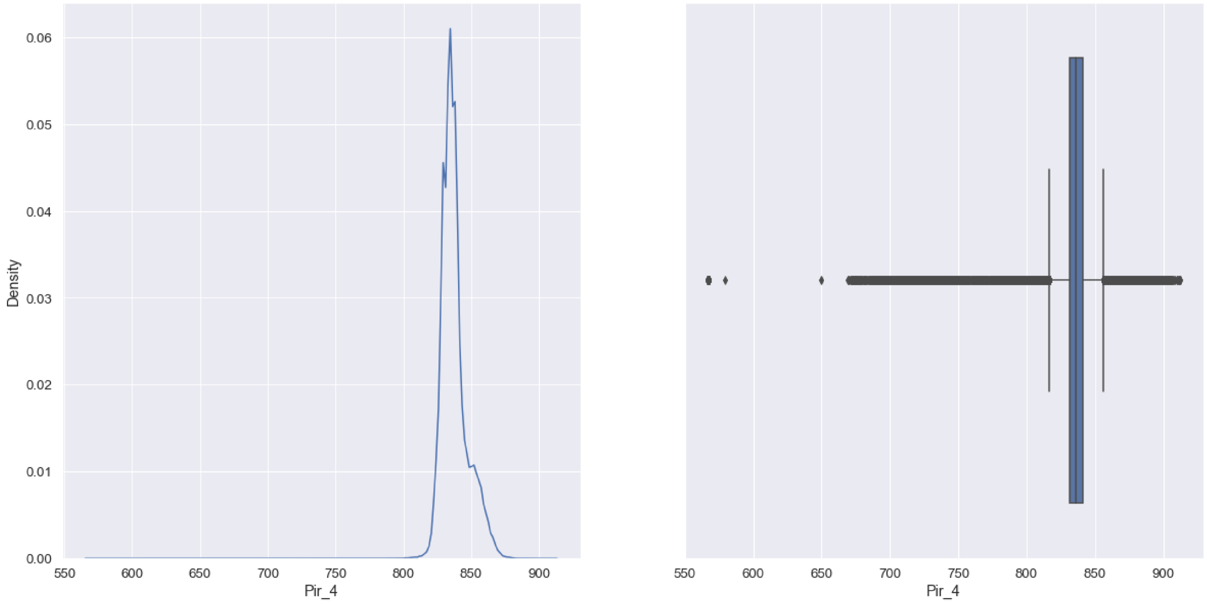
In the context of the 430st production line, it is imperative to note that the operational speed is typically expected to surpass a minimum threshold of 50 meters per minute (m/min). This benchmark provides a crucial reference point for gauging the line's efficiency and productivity. However, the line speed may not always maintain a consistent pace. It can exhibit sudden and pronounced drops, often serving as a telltale sign of underlying issues within the production line or potential defects in the product being processed. These deviations in speed carry the potential to significantly impact the performance of any predictive model, introducing a level of noise and variability that can challenge the model's robustness.

To mitigate the potential impact of speed variations on the analysis and modeling, a deliberate decision has been made to consider only those products that consistently maintained a line speed exceeding 50 m/min. This selective approach filters out products that fall below this speed threshold, ensuring that the subsequent analysis is conducted on a more homogenous and reliable dataset.

Another important aspect is that the dataset employed in this analysis originates from physical sensors embedded within a production line. Such sensors are susceptible to generating erroneous data. During the data transformation phase, erroneous data within the time series were meticulously eliminated. The most common sources of this aberrant data were related to sensor failures or temporary shutdowns.

However, it should be noted that the issue of outliers within the dataset remained unaddressed. These outliers may manifest when the manufacturing process deviates from its predefined setpoints. Furthermore, they could be attributed to sensor misconfigurations and other unforeseen anomalies. An illustrative example of this outlier presence can be observed in Figure 3.3, which depicts Pyrometer readings. The data predominantly clusters within the range of 810 to 870, as evident in the concentration of data points within the interquartile range. Nonetheless, the presence of numerous data points beyond the whiskers in the boxplot signifies the existence of outliers.

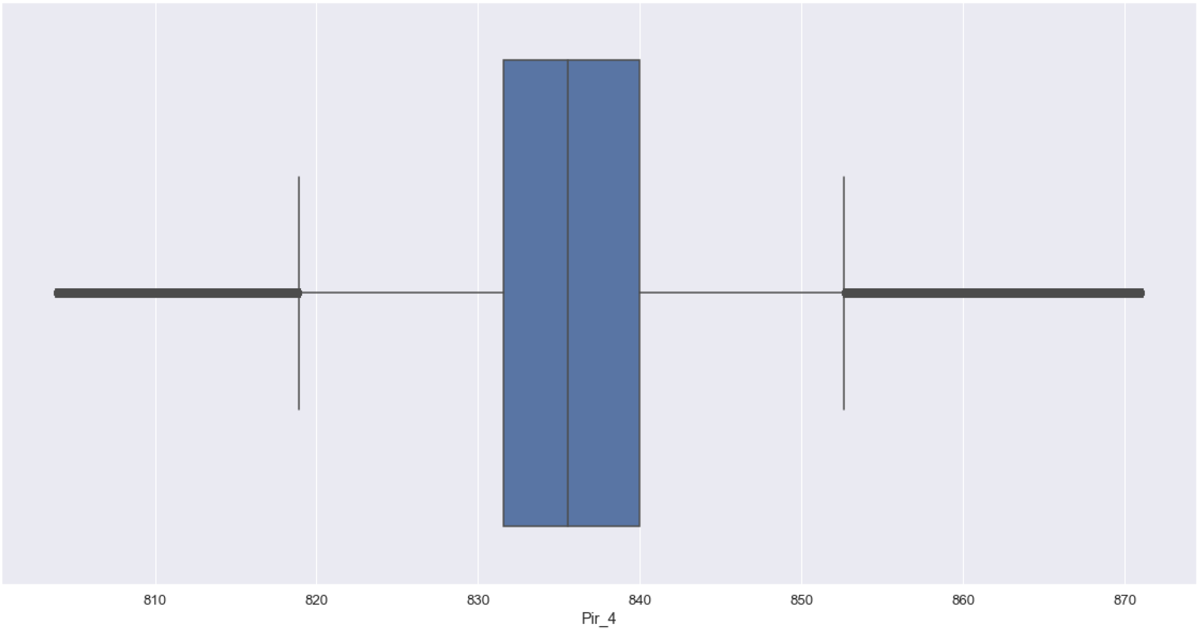
Figure 3.3: Presence of outliers in the Pyrometer readings.



It should be acknowledged that these outliers possess the potential to adversely impact the performance of any subsequent modeling endeavors due to their classification as uncommon patterns. Therefore, a pivotal step within the data processing pipeline involves their meticulous removal. To execute this critical process, the Z-Score was computed, and data points found to deviate by three standard deviations from the sample mean were systematically eliminated.

The results of the data cleansing process are portrayed in Figure 3.4, where a distinct transformation in the dataset's distribution is observed. This same rigorous procedure was replicated for Zone Temperatures, signifying a comprehensive approach to ensure the quality and reliability of the dataset employed for subsequent analyses.

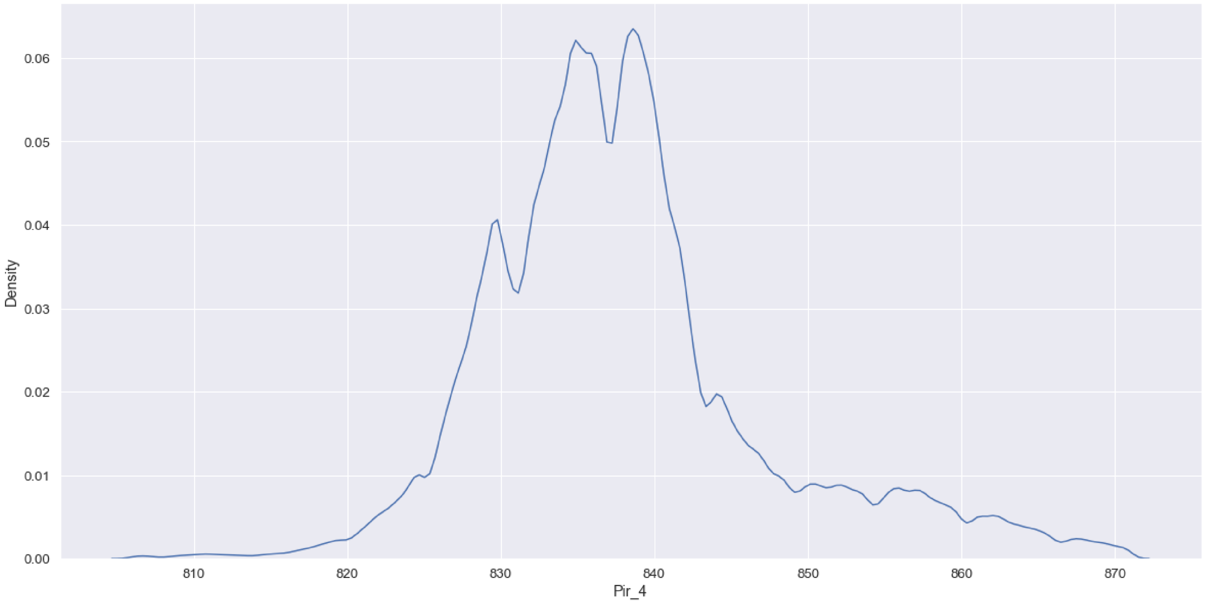
Figure 3.4: Pyrometer 4 bloxplot after removed outliers.



3.1.5 Pyrometer Data Exploration

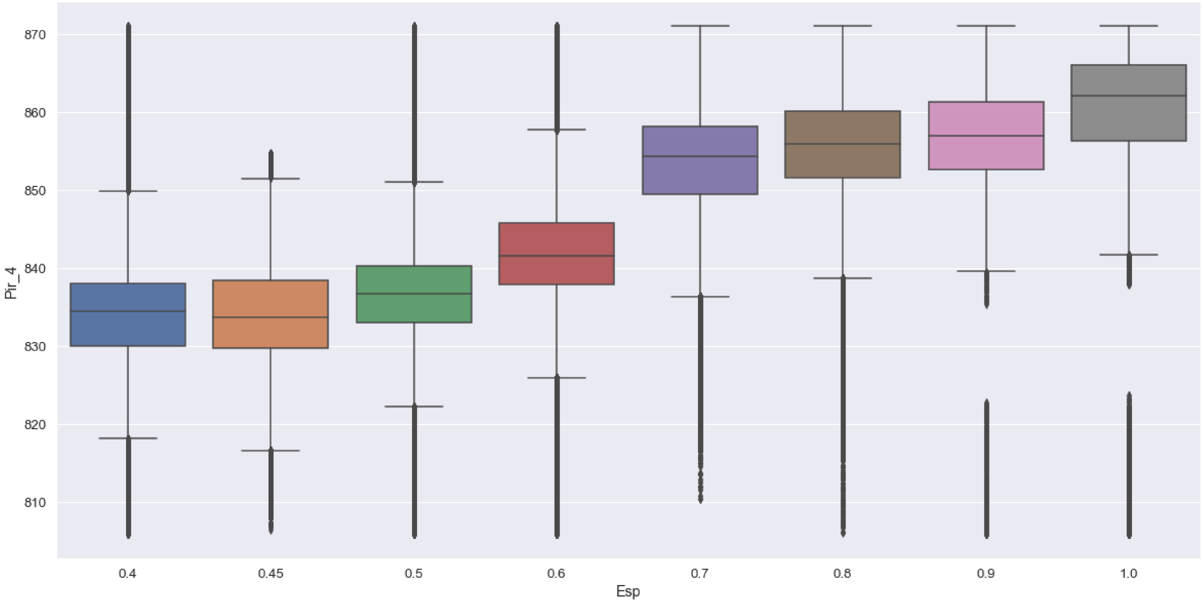
Firstly, an examination of pyrometer measurements led to the revelation of one important characteristic of its data. As illustrated in 3.5, it became evident that the distribution of these measurements deviated from a Gaussian.

Figure 3.5: Pyrometer 4 data distribution.



This departure from normality can be attributed to the existence of various furnace setups contingent upon the material's thickness. Subsequently, 3.6, portraying box plots for distinct thickness categories, starkly underscores the influence of these disparate furnace configurations on the distribution of pyrometer measurements.

Figure 3.6: Pyrometer 4 boxplot by Thickness.



Recognizing the distinct distribution patterns within each thickness group, our forthcoming sections on model performance evaluation will include a detailed analysis of results according to individual thickness categories. It is essential to acknowledge that models are expected to yield consistent predictions across all thickness ranges. However, there is a possibility that they might exhibit superior performance within specific thickness intervals. This granularity in the assessment will provide a deeper insight into the models' effectiveness in handling variations across different thickness groups.

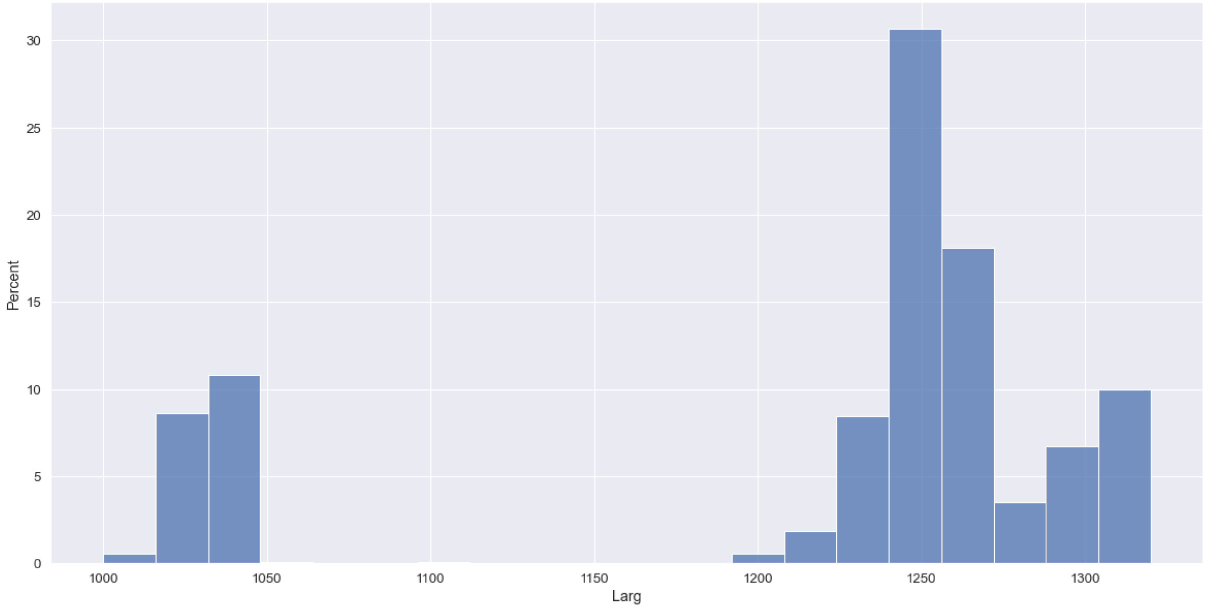
After all the transformation and cleaning processes, the final size of the training dataset is an extensive 3,214,145 records, with each entry representing a distinct meter of material processed in the Furnace of RB4. The abundance of data in the training dataset is advantageous for machine learning models as it allows for a more efficient learning of patterns and relationships within the data, enhancing the model's predictive capabilities. Similarly, the test dataset, following the same meticulous cleaning procedures, comprises a substantial 1,415,696 records. This ample amount of data in the test set is invaluable for evaluating model performance, ensuring robustness, and demonstrating resilience to overfitting, ultimately validating the model's ability to generalize well to new, unseen data.

3.1.6 Independent Variables Data Exploration

After data cleaning and exploration of the target variable, this section delves into an in-depth examination of the independent variables. To safeguard Aperam Intellectual Property, data related to furnace process variables will be presented in a normalized form.

As elucidated earlier, this study encompasses a total of eight distinct thickness groups, each with its intrinsic significance. Additionally, another critical dimensional variable is the width of the material, which spans from 1000mm to 1040mm and from 1200mm to 1320mm, as visually depicted in Figure 3.7. The interplay of width and thickness provides valuable insights into the mass of the processed material, rendering it a pivotal consideration in the annealing process, both of them are also present in the physical model.

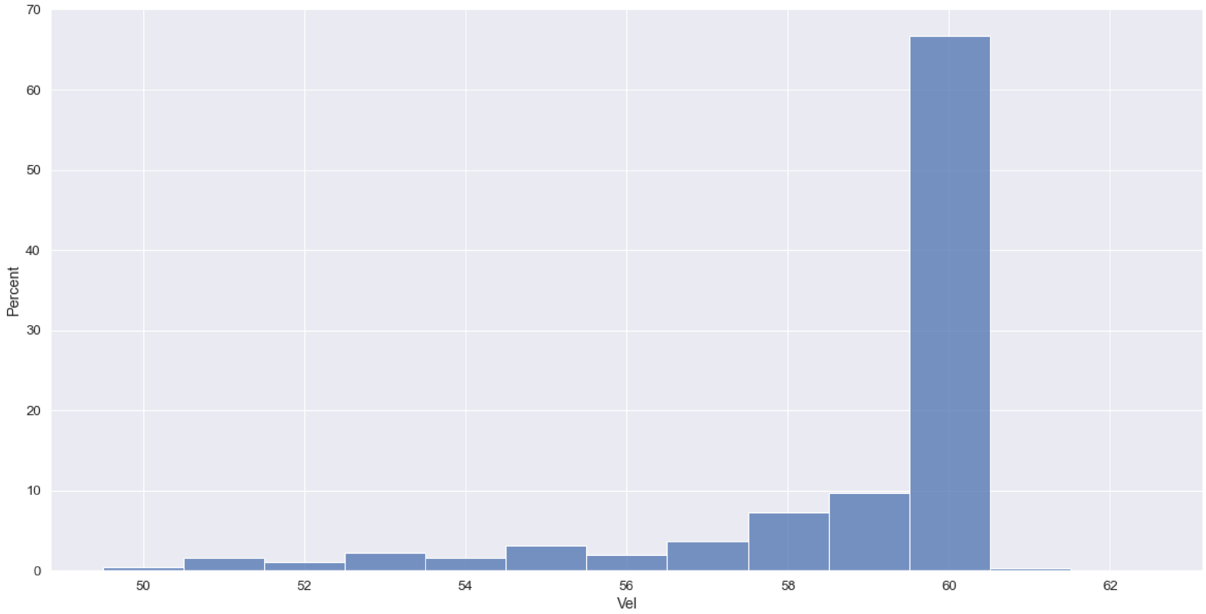
Figure 3.7: Histogram of Width in the training dataset.



The line speed, another pivotal variable in the process, significantly impacts the pyrometer readings. The proposed setpoint for speed in the studied product is approximately 60 m/min, ensuring that, with the correct temperature input in each zone, the steel strip's heating curve aligns with expectations. Instances of quality or mechanical issues may necessitate a speed drop or even a complete line stop, rendering the product unsuitable for its intended application. Given the focus on replicating a stable line in this study, only speeds above 50 m/min were considered, effectively eliminating these outliers. The distribution of line speed data is visually represented in Figure 3.8, demonstrating that the majority of readings cluster around 60 m/min.

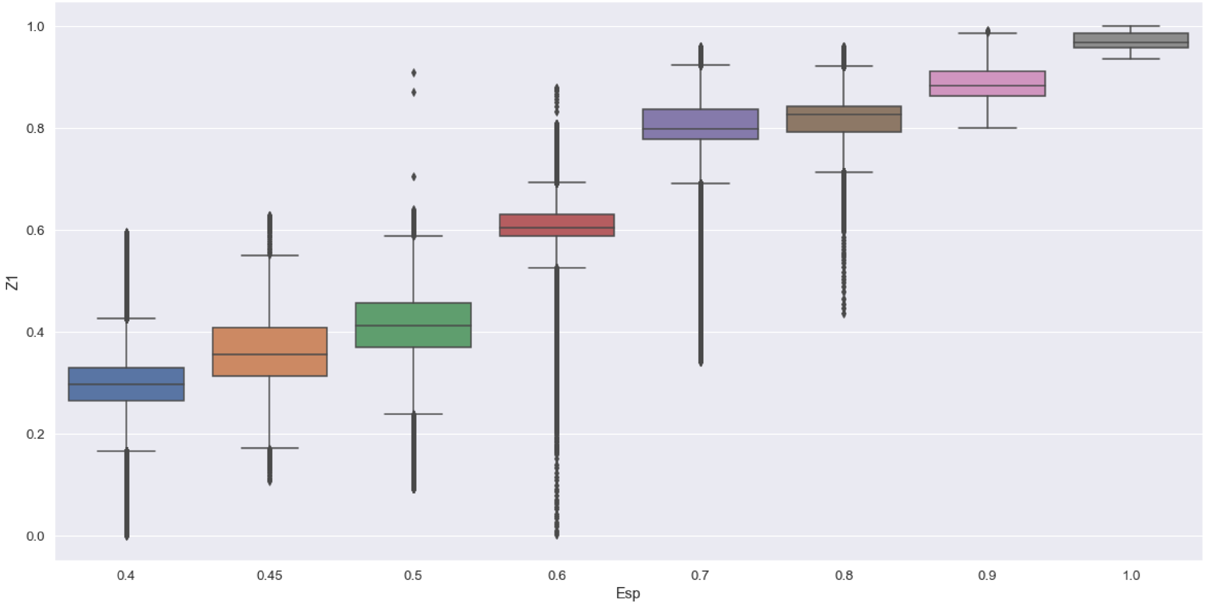
Zone temperatures assume a pivotal role in the experiments, serving as crucial in-

Figure 3.8: Histogram of line speed in the training dataset.



puts for both the physical models and the proposed grey-box models. Their distributions, at majority, do not adhere to a normal distribution, owing to distinct configurations for different thickness groups, mirroring the pattern observed in pyrometer temperatures (as shown in Figure 3.9). Box-plots of Zone 1 temperature for each thickness group illustrate this variation, showcasing an increasing trend. This variance is necessary to accommodate the heating requirements of different mass loads.

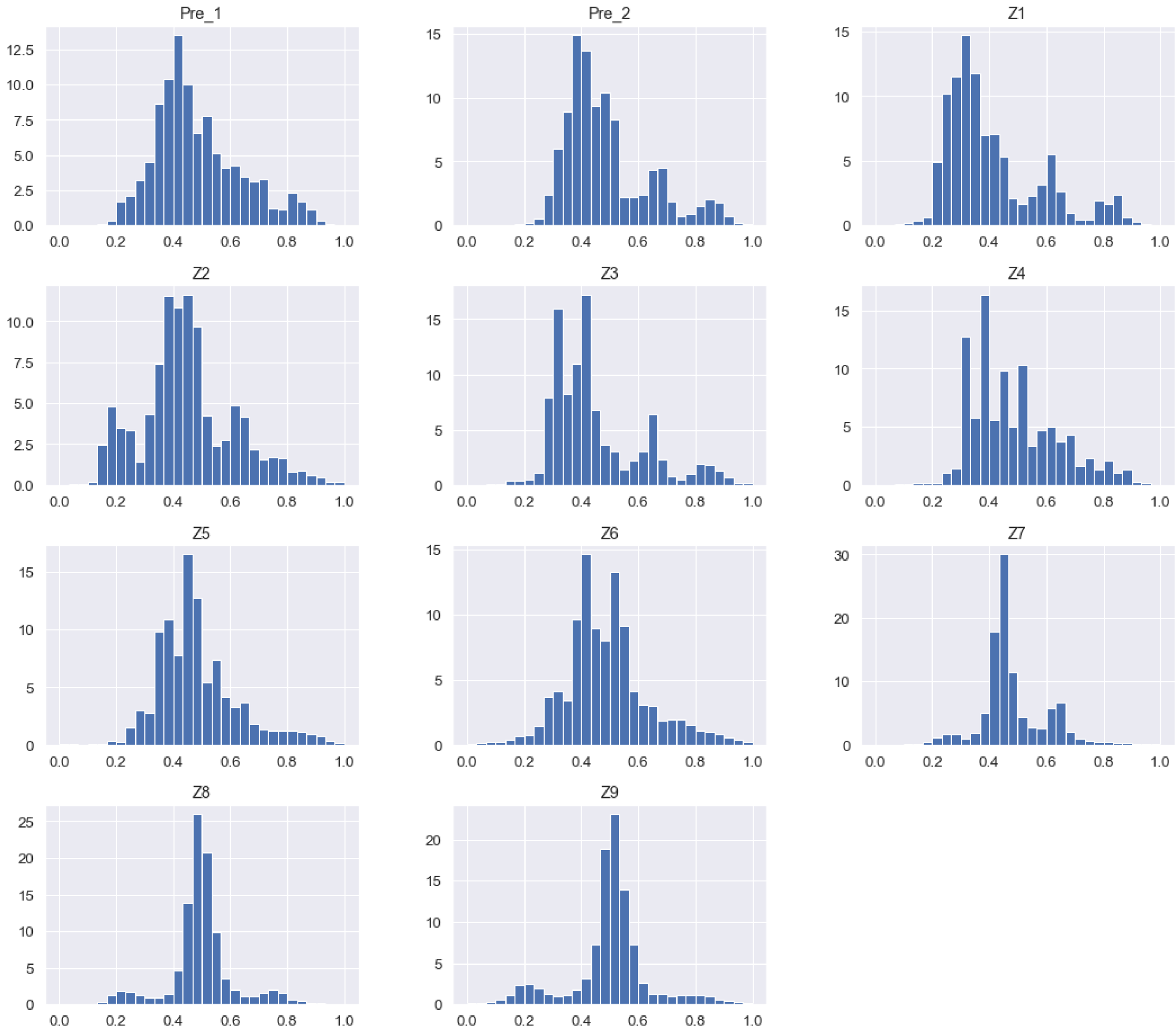
Figure 3.9: Boxplots of Zone 1 temperature by Thickness.



There exists one temperature measurement for each zone, with two additional representations for a pre-heating section. As explained their distribution, depicted in

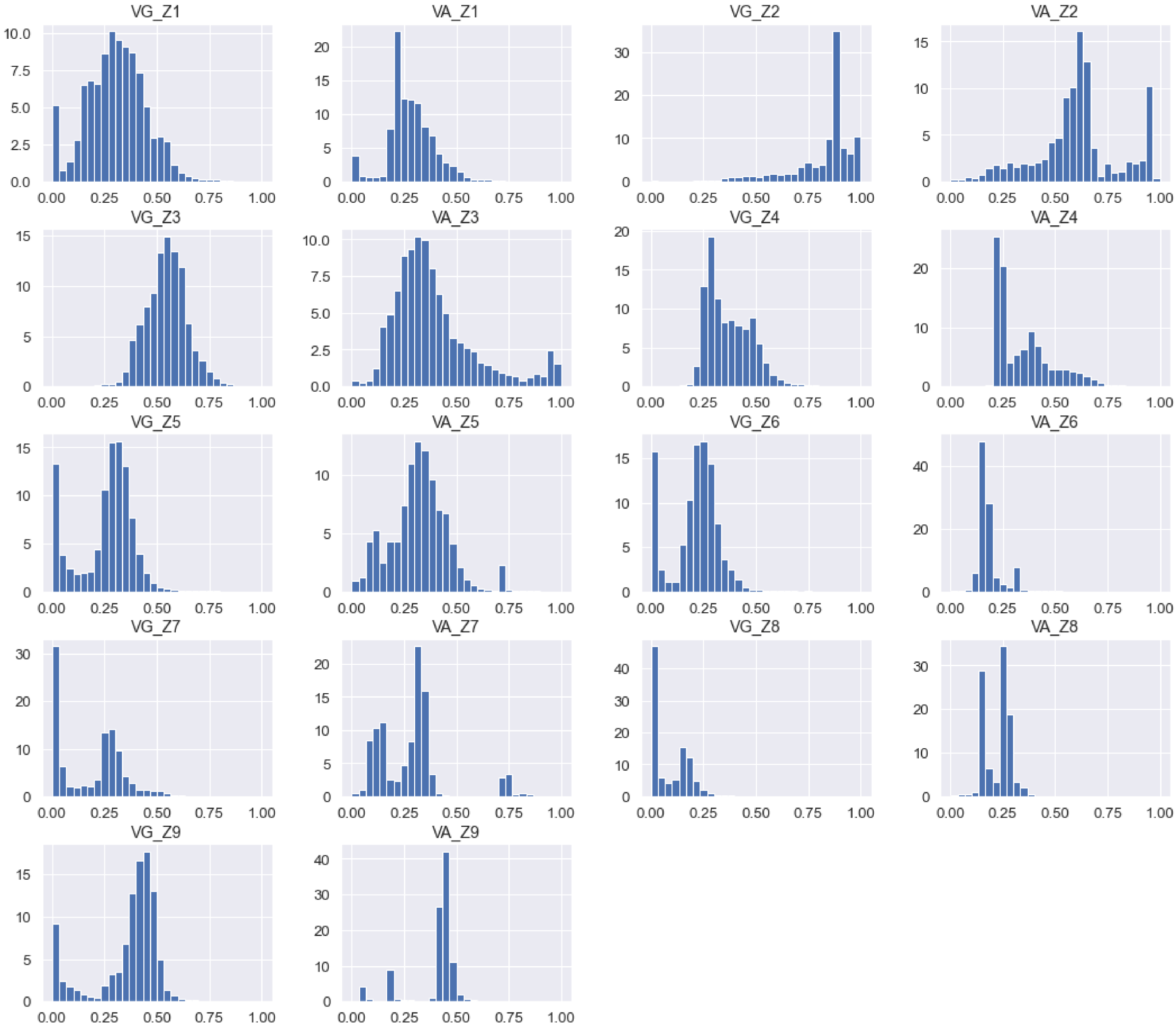
Figure 3.10 exhibits diverse formats. Given that the material traverses these temperatures from the furnace's inception to its end, the accurate configuration of zones also holds paramount significance. A failure or malfunction in the setup of any zone can disrupt the strip temperature curve, leading to deviations from the desired outcome.

Figure 3.10: Distribution of Zones temperatures.



This study extends its exploration to variables outside the physical model, delving into process-related factors. Among these, Air and Gas Flow measurements assume significance and are employed to achieve and regulate zone temperatures. The distributions of these measurements, depicted in Figure 3.11, reveal two discernible patterns. Firstly, final zones exhibit a more restricted operational range, characterized by a narrower distribution around their means. This indicates a lesser degree of operational interference during the annealing process. Secondly, certain zones, like Zone 1 and Zone 5, exhibit numerous zero values in Gas Flow, which are also actual zero values in the original distribution. This is not a data error but reflects instances of zero gas input during the process, highlighting operational characteristics.

Figure 3.11: Distribution of Air and Gas Flow variables.

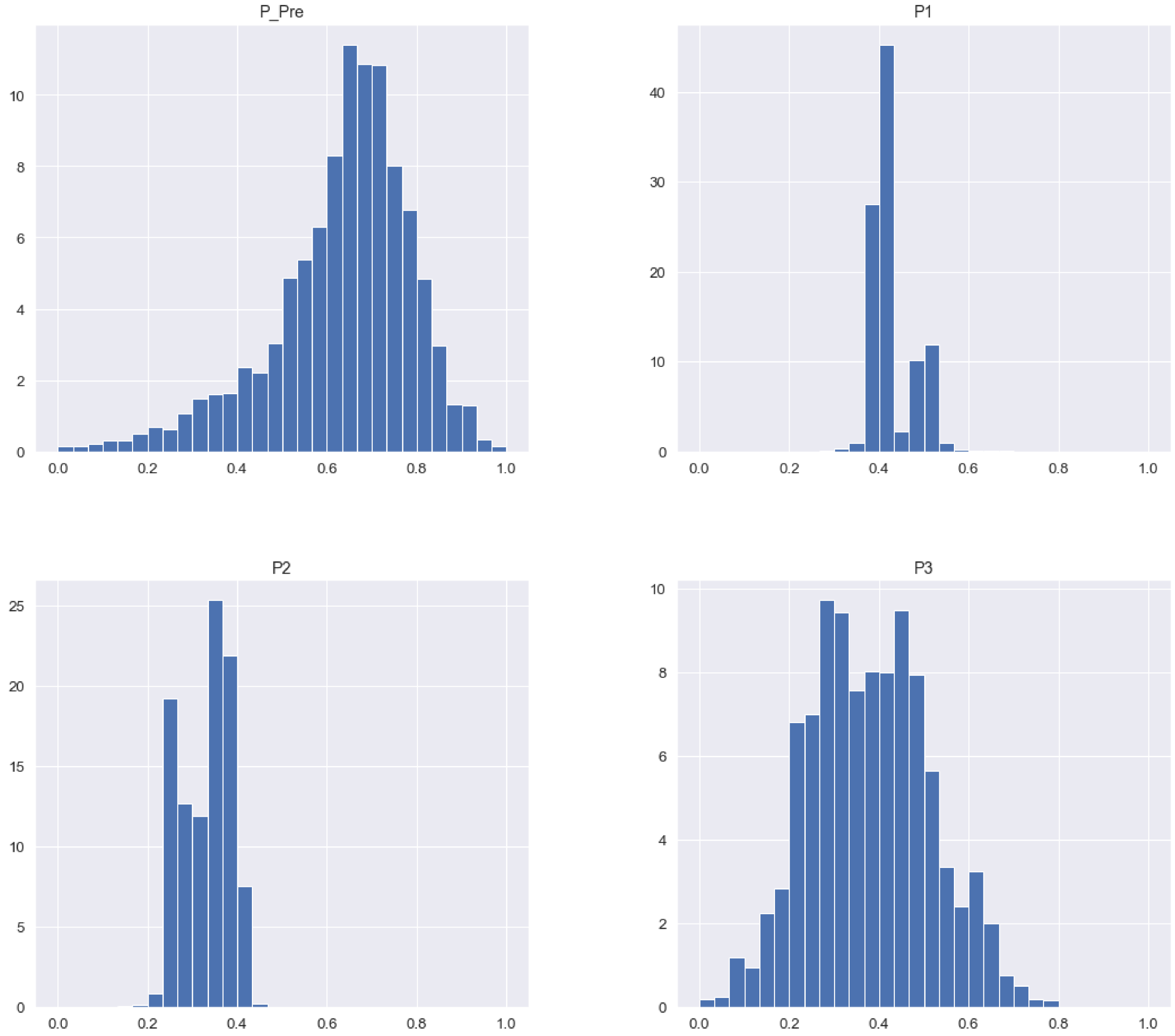


While attempting to physically model all gas movements presents a formidable challenge, these variables might influence the furnace atmosphere. Thus, their inclusion in hybrid modeling, alongside zone temperatures, can empower machine learning algorithms to capture the complexities of the annealing process more comprehensively than relying solely on temperatures.

The final set of variables under consideration includes Furnace Pressure, although not explicitly incorporated into the physical model, they play a pivotal role in the RB4 process. As demonstrated in Figure 3.12, the distribution of Furnace Pressure reveals distinctive patterns, particularly in the pre-heating zone. An exhauster in this zone facilitates the expulsion of combustion by-products like CO2, creating a counter flux flow in the line, moving from the end to the beginning of the line.

These pressure variables bear significance in ensuring the proper movement of gases within the furnace. Anomalies in their behavior could lead to deviations in heating curves and potentially result in quality-related issues in the processed material. Thus,

Figure 3.12: Distribution of Furnace Pressure variables.



understanding and monitoring Furnace Pressure is crucial for maintaining the integrity of the annealing process and the desired product quality.

In concluding the exploration of independent variables, a spectrum of crucial factors influencing the annealing process in RB4 has been traversed. From material dimensions such as thickness and width to operational parameters like line speed, zone temperatures, and environmental conditions such as air and gas flow, each variable bears significance in achieving the desired temperature curves for the steel strip. The distribution analyses have unearthed distinctive patterns, providing insights into the operational nuances and potential challenges within the furnace. Beyond the physical model's constraints, the inclusion of variables like air and gas flow, as well as furnace pressure, showcases the richness of the process dynamics that can enhance the capabilities of machine learning algorithms in this context. This thorough exploration sets the stage for the subsequent integration of these variables in the modeling endeavors.

3.2 Modeling

3.2.1 Modeling Strategy

Four distinct machine learning algorithms were explored in this study, specifically XGBoost, Random Forest, Linear Regression, and Support Vector Machine (SVM). The training of these models was carried out using the widely-used open-source Python library, scikit-learn by Pedregosa et al. [2011]. These models were selected mainly due to their explainability, since having an overview of the underlying features used by the models, alongside their importance to each model, would be useful to understand their different results and approaches to the grey-box strategy.

In order to enhance the model's performance, an exploration of parameters optimization was undertaken. The exploration range for each of the investigated algorithms is presented in Table 3.3.

Model	Hyperparameters	Explored range
XGBoost	gamma subsample n_estimators max_depth learning_rate	[0, 0.5, 2, 4] [0.6, 0.8, 1] [50, 100, 200, 400] [8, 10, 12] [0.1, 0.3]
Random Forest	max_features max_depth n_estimators	['auto', 'sqrt'] [6, 8, 10] [50, 100, 200, 400]
Linear Regression	-	-
SVM	C	[0.1, 0.5, 1]

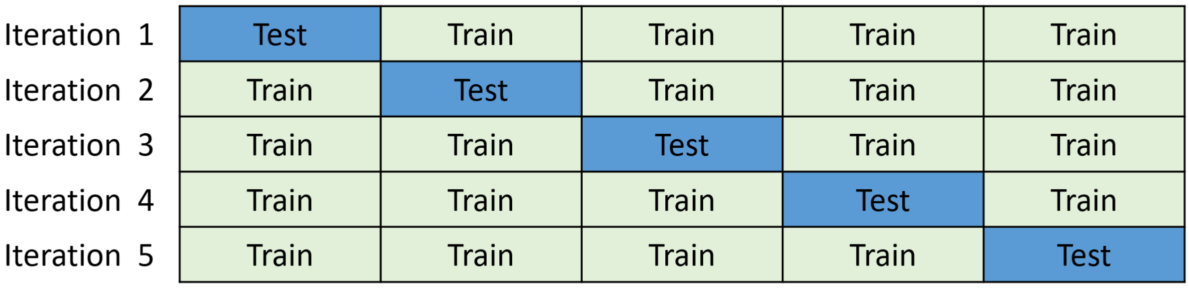
Table 3.3: Overview of modeling strategy.

3.2.2 Model Training and Validation

In order to obtain reliable estimates of model performance and prevent overfitting, two strategies will be implemented to evaluate the model. The first strategy involves splitting the available data into training and testing datasets. Approximately 70% of the available data will be used to develop the model, which comprises dates between January 1st, 2022, and September 30th, 2022. The remaining 30%, which is data collected from October 1st, 2022 to December 31st, 2022 will be used to assess the performance of the model's predictions on unseen data. This process can provide insights into the model's variance and its ability to generalize to new data.

The second strategy involves using cross-validation during model training, as described by Hastie et al. [2009]. This method involves dividing the training data into k equally sized sub-samples, using one sub-sample for validation and the remaining sub-samples to build the model. This process is repeated k times, with each sub-sample used once for validation. The results of cross-validation can be used to detect overfitting and make adjustments to the model. This approach allows for the use of as much data as possible during training, which is particularly important when data is scarce. Since multiple ranges of hyperparameters will be explored, the number of folds in this experiment will be fixed at 3. An example of k-folding is exemplified in Figure 3.13.

Figure 3.13: Cross Validation with 5 k-folds.



Iteration 1	Test	Train	Train	Train	Train
Iteration 2	Train	Test	Train	Train	Train
Iteration 3	Train	Train	Test	Train	Train
Iteration 4	Train	Train	Train	Test	Train
Iteration 5	Train	Train	Train	Train	Test

Through the experiments, the metric that will be used to evaluate how well the models fit the data is the root mean squared error (RMSE). It is a widely used measure that gives important feedback about the regressor's performance.

The RMSE gives an insight into the distance between the predicted values from the real ones in a dataset, being the lower the root mean squared error, the better the model. The formula to calculate it is given as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}$$

3.3 Experiments Setup

The experimental process will be executed through a series of well-defined steps presented.

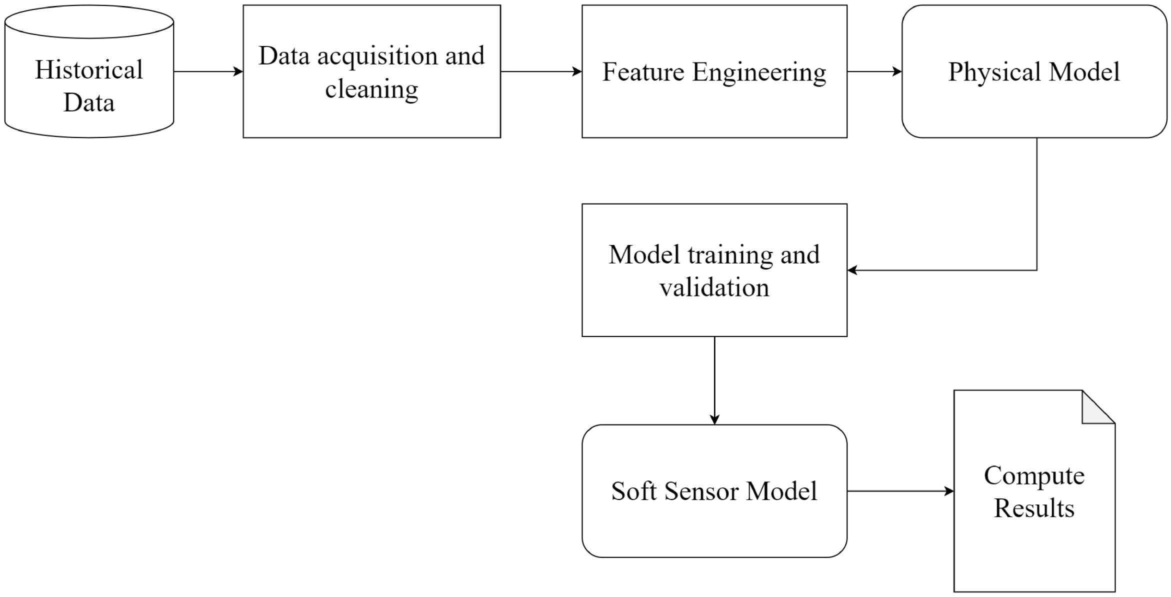
First and foremost, data will be collected from the PIMS systems, subsequently cleaned, and then removed from any faulty or missing data.

After data collection and initial cleaning, focusing mainly on outlier removal, a feature engineering process will be applied to the gathered data. This process will culminate in the creation of a curated dataset, ready for data modeling.

Following the feature engineering stage, the physical model will be trained and validated. Then, the probabilistic model will undergo a similar process, exploring a grey modeling approach, followed by the validation of the final soft sensor model.

Upon the completion of these preceding steps, each individual model will be subject to evaluation on test data, thereby facilitating an assessment of their performance and reliability. A complete overview of the entire process is represented in Figure 3.14

Figure 3.14: Experiment steps flow.



Chapter 4

Results and Discussion

This chapter is dedicated to the presentation of the experimental results. It encompasses the outcomes derived from the deployment of the physical model, provided by Aperam R&D for RB4. Additionally, the results of the Grey model approach, employing machine learning models in conjunction with the physical model to predict pyrometer temperature, will be expounded upon. Lastly, the section encompasses the presentation of findings obtained through the application of a pure probabilistic approach and a comparison between the models.

4.1 Physical model deployment

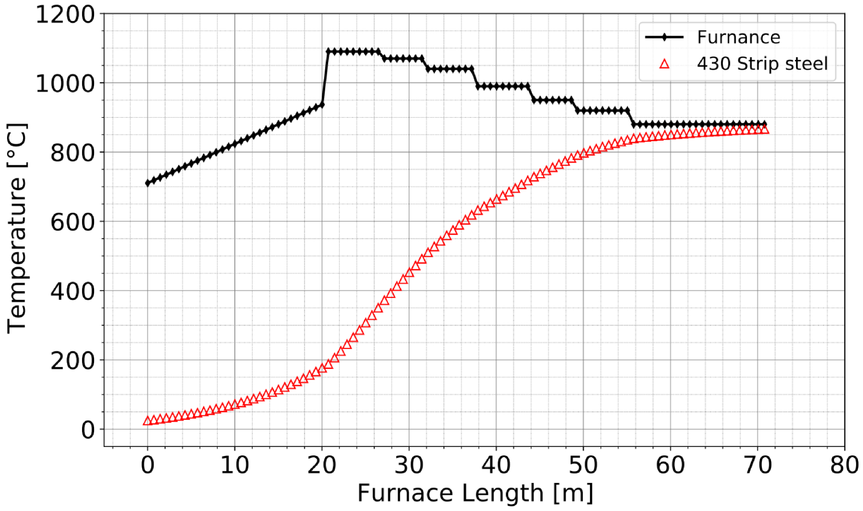
The initial phase involved the integration of the physical model, provided by Aperam R&D, into the dataset. This model operates with specific input parameters, which include material thickness, material width, line speed, and zone temperatures. Other parameters linked to the overall phenomenon, rather than the material itself, are either pre-calculated or held constant within the model.

The process of deploying the physical model employs the Finite Method Difference, to solve the differential heat conduction exchange equation explicitly, yielding the strip temperature profiles represented in Figure 4.1. In this experiment, a total of 100 FDM elements were taken into account to discretize the temperature distribution across the length of the furnace.

To align the model's results with the Pyrometer readings, which are situated at the end of the production line, the data corresponding to the final element was collected.

The post-processing of the dataset involved assigning input parameters to every meter of material processed within the specified time frame. Consequently, individual temperature curves were meticulously computed for each of these meters.

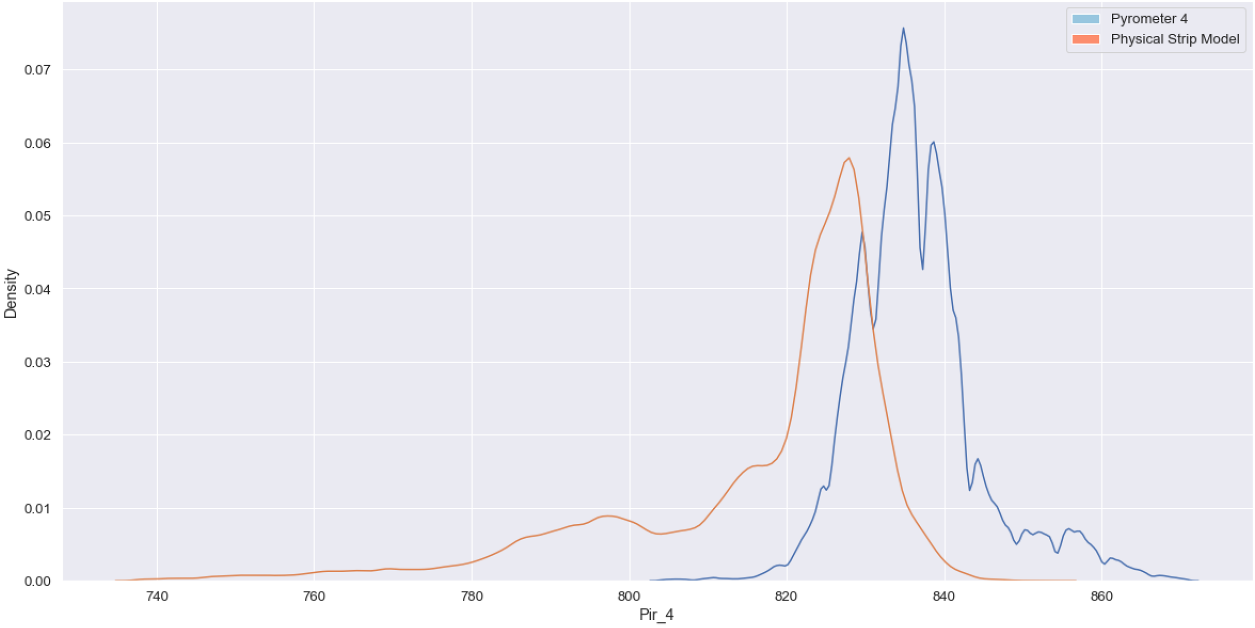
Figure 4.1: Strip temperature curve calculated by physical model.



4.1.1 Evaluation of the physical model

When comparing the distribution of strip temperatures derived from the provided physical model with actual pyrometer measurements, discrepancies between these distributions are apparent, as depicted in Figure 4.2.

Figure 4.2: Pyrometer 4 vs Physical Model.



Several factors may contribute to these disparities. Firstly, it's plausible that the physical model may not encapsulate all relevant phenomenological aspects of the annealing process, since as stated it is a simplified approach. Moreover, variations in pyrometer configurations may also contribute to these deviations.

When examining the physical model's temperature estimations. Contrary to the behavior of Pyrometer 4, Figure 4.3 illustrates that as material thickness increases, the physical model predicts a decrease in temperature.

Figure 4.3: Physical model temperature estimations boxplot by product thickness.

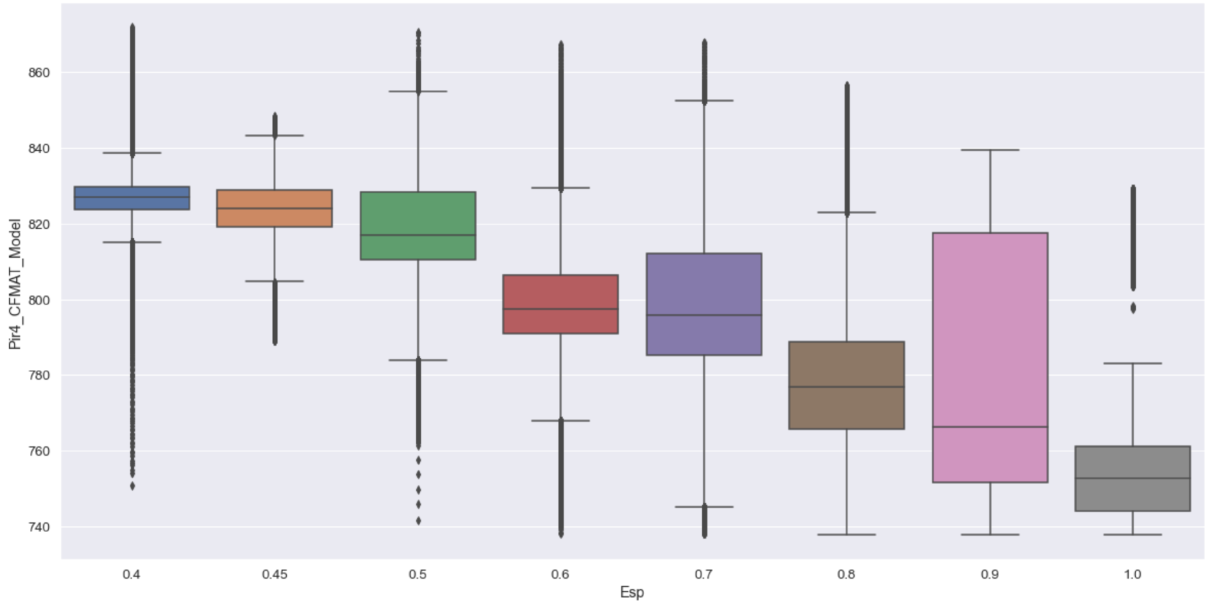
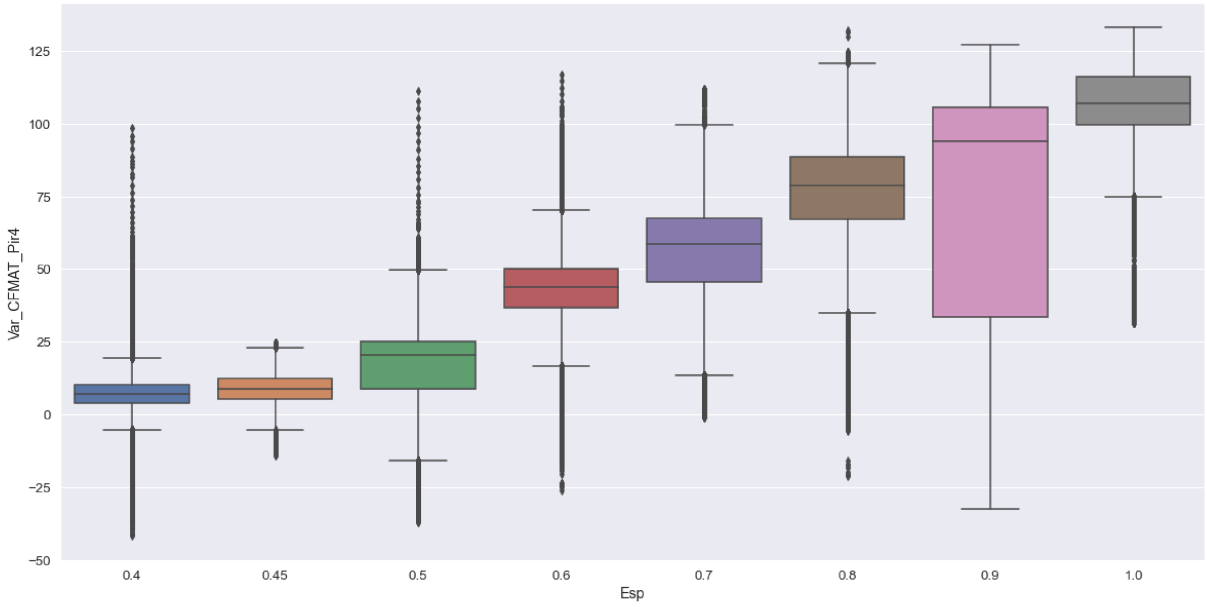


Figure 4.4 further illustrates this by revealing that errors increase in a linear fashion along with thickness. This observed pattern underscores the non-random nature of these errors and paves the way for further exploration, particularly in the realm of Grey Box modeling, which may leverage linear models to predict pyrometer temperature.

Figure 4.4: Error by Thickness.



The overall performance of the model in train and test datasets is presented in Table 4.1

Model	Train data RMSE	Test data RMSE
Physical Model	29.66	33.36

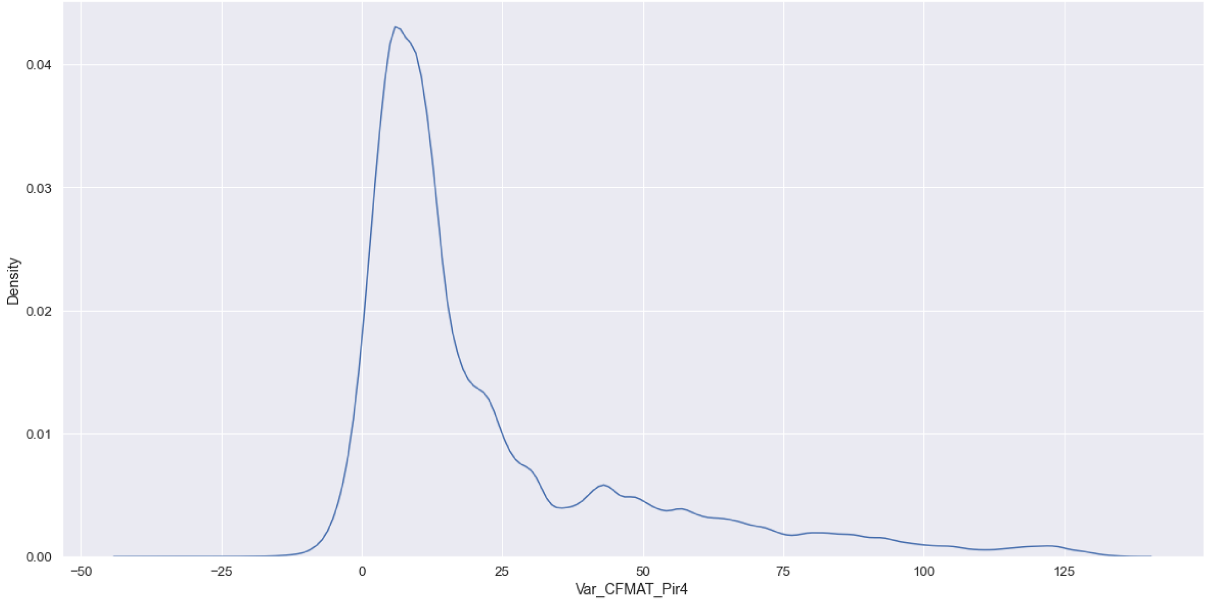
Table 4.1: Physical Model Performance.

4.2 Grey box modeling

This section aims to provide a presentation of the strategy and results derived from each of the models employed in the study. This will be done by showcasing their individual performance using the chosen metric through different optics, to elucidate their strengths, and weaknesses across multiple thickness ranges, and their underlying strategy regarding feature selection.

The objective of the proposed grey box modeling approach is to use process variables and the physical model to forecast Pyrometer 4 temperature. Consequently, the target variable chosen for this purpose is the variation between the Pyrometer reading and the physical model's prediction, which can be observed in Figure 4.5.

Figure 4.5: Variation between Pyrometer 4 and Physical Model.



This selection is motivated by the behavior of certain machine learning models, such as XGBoost, which may not utilize all available features when constructing a decision tree. This strategy, employed by the algorithm, is primarily aimed at mitigating overfitting. However, for the present experiment, it is essential to ensure that the Physical Model contributes to the final prediction consistently, thereby enabling a meaningful comparison with the pure probabilistic approach. Ultimately, the Pyrometer Temperature can be reconstructed by adding the physical model's value with the forecasted error component.

4.2.1 Models Perfomance

Table 4.2 presents the evaluation of the models proposed using the grey-box modeling strategy.

Interestingly, the evaluation revealed some unexpected outcomes. It was observed that relatively simple linear models outperformed ensemble algorithms. Each model will be individually explored, aiming to gain a deeper understanding of their behavior and the specific results observed during evaluation.

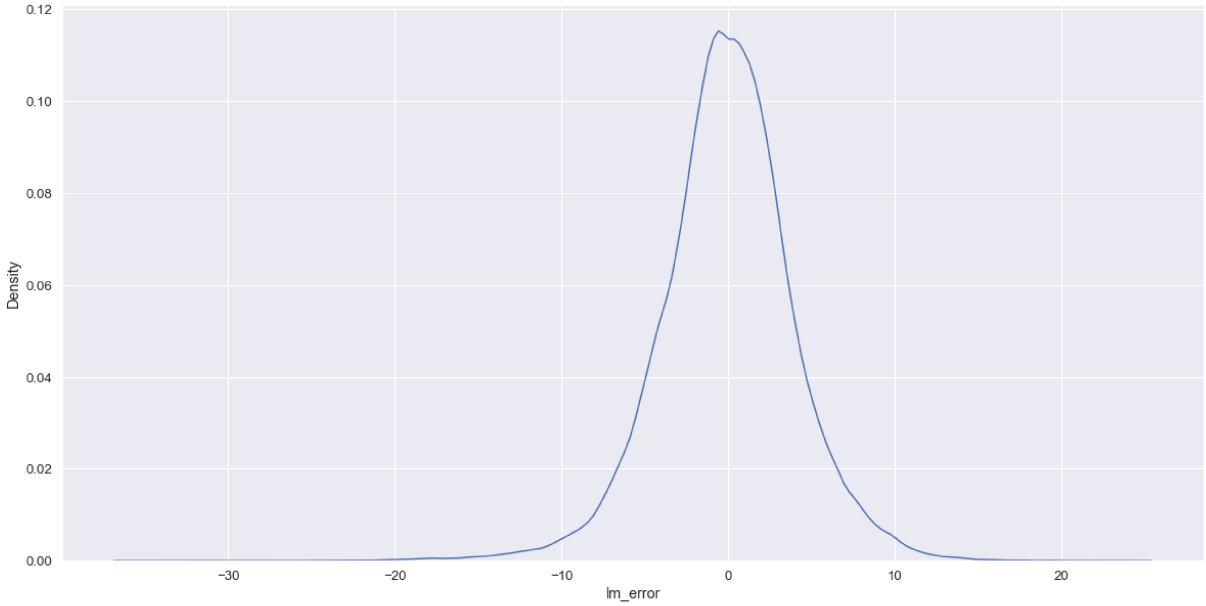
Model	Best hyperparameters	Train RMSE	Test RMSE
XGBoost	gamma: 0.5 subsample: 1 n_estimators: 200 max_depth: 6 learning_rate: 0.1	4.75	4.57
Linear Regression	-	4.86	4.01
SVM	C: 0.1	4.97	3.94
Random Forest	max_features: 'sqrt' max_depth: 12 n_estimators: 100	5.63	4.92

Table 4.2: Overview of model results.

4.2.2 Evaluating Linear Regression

In Figure 4.6 the errors associated with the linear model exhibit a distribution that appears to be reasonably centered around zero, with mean and standard deviation being $\overline{x} = -0.07$ and $s = 4.01$. Nevertheless, the presence of a few outliers on the left side of the distribution is notable, suggesting certain deviations from the norm.

Figure 4.6: Error distribution for Linear Regression grey-box model.



The residual plot in Figure 4.7 unveils an intriguing trend. It becomes apparent that errors for lower temperatures primarily manifest as negative values, and as temperature increases, the errors transition towards increasingly positive values. This observation highlights an unexpected correlation between temperature and error magnitude.

Figure 4.7: Residual plot for Linear Regression grey-box model.

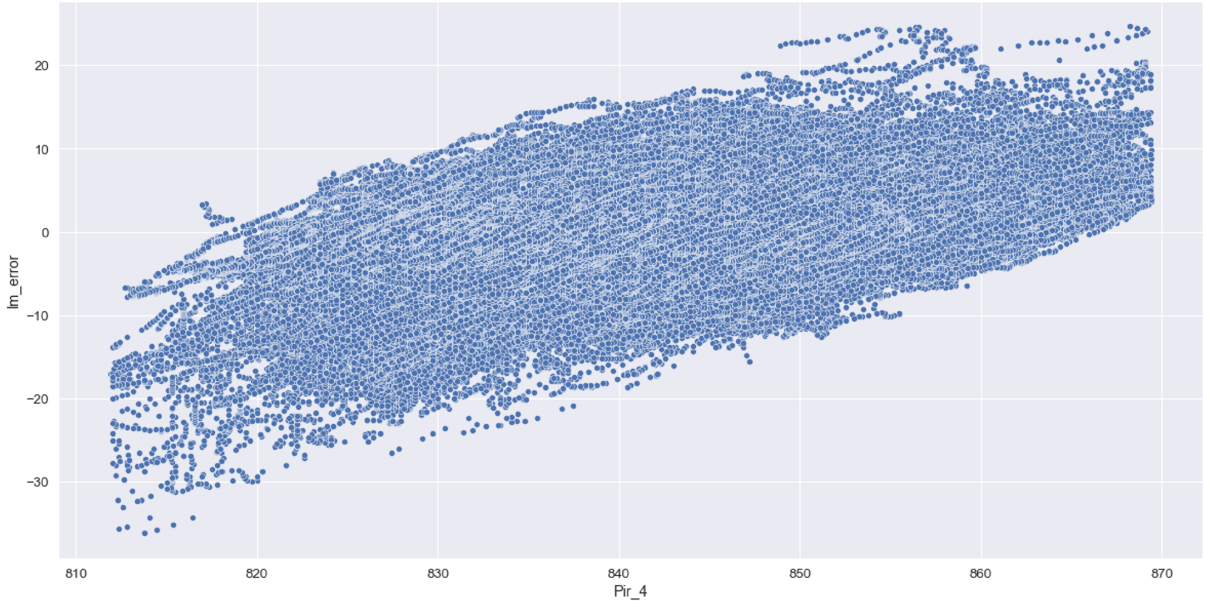
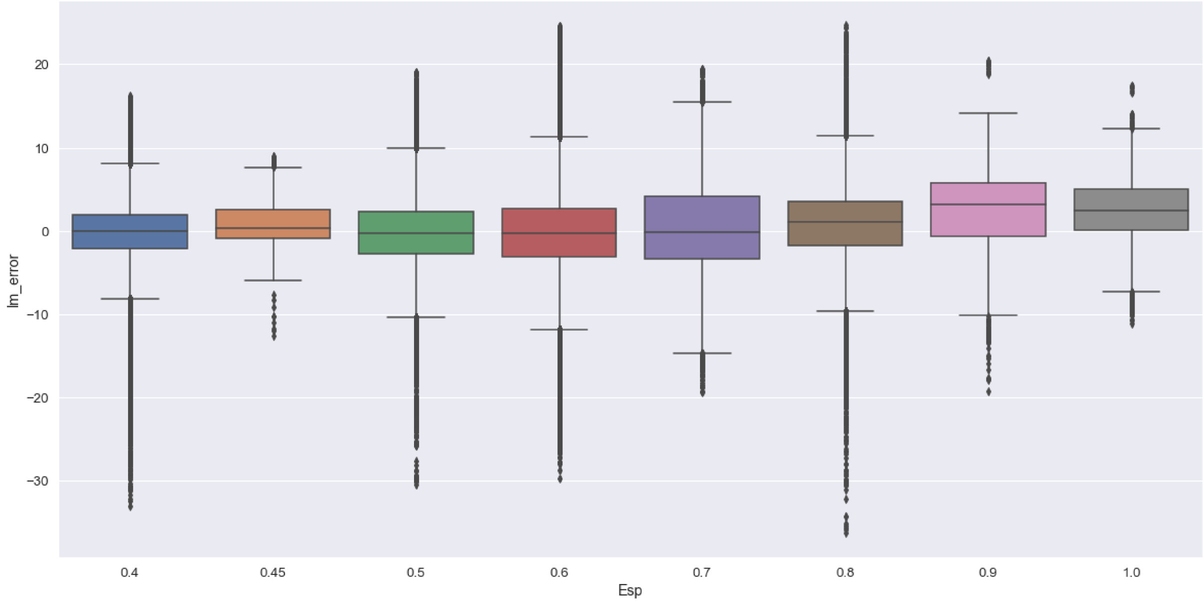


Figure 4.8 allows us to explore the influence of thickness on the errors. Here, we observe a maximum error range of approximately -10 to +10 across different thicknesses. However, an interesting pattern emerges: most of the outliers are skewed towards the negative side. Additionally, when the material thickness surpasses 0.8mm, the median line of errors shifts to a position above zero. These combined findings provide valuable

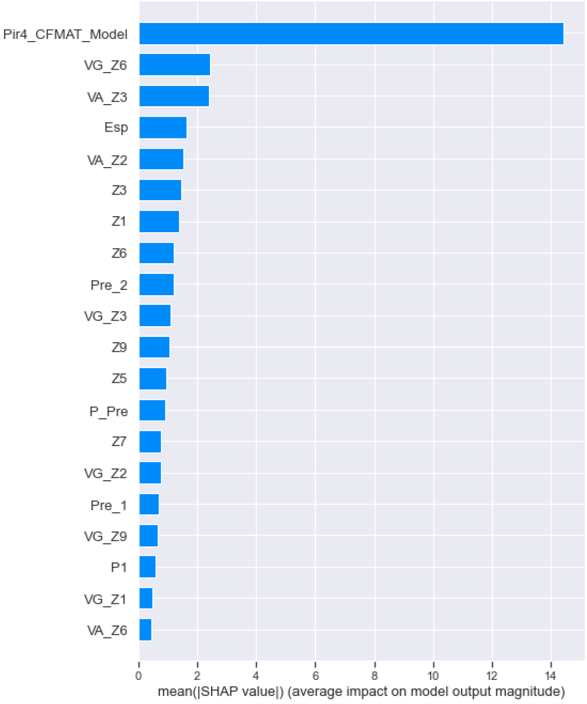
insights into the factors affecting the residuals as seen in the residual plots, since it was already stated that higher temperature readings are in higher thickness.

Figure 4.8: Boxplot of errors by thickness for Linear Regression grey-box model.



Looking at the feature importance of Linear Regression in Figure 4.9, the actual physical model temperature emerges as a prominent feature, as anticipated, given its direct correlation with the target variable. The model employs this temperature as the primary driver for its predictions, indicating its pivotal role in shaping the output.

Figure 4.9: Feature importance for Linear Regression grey-box model.

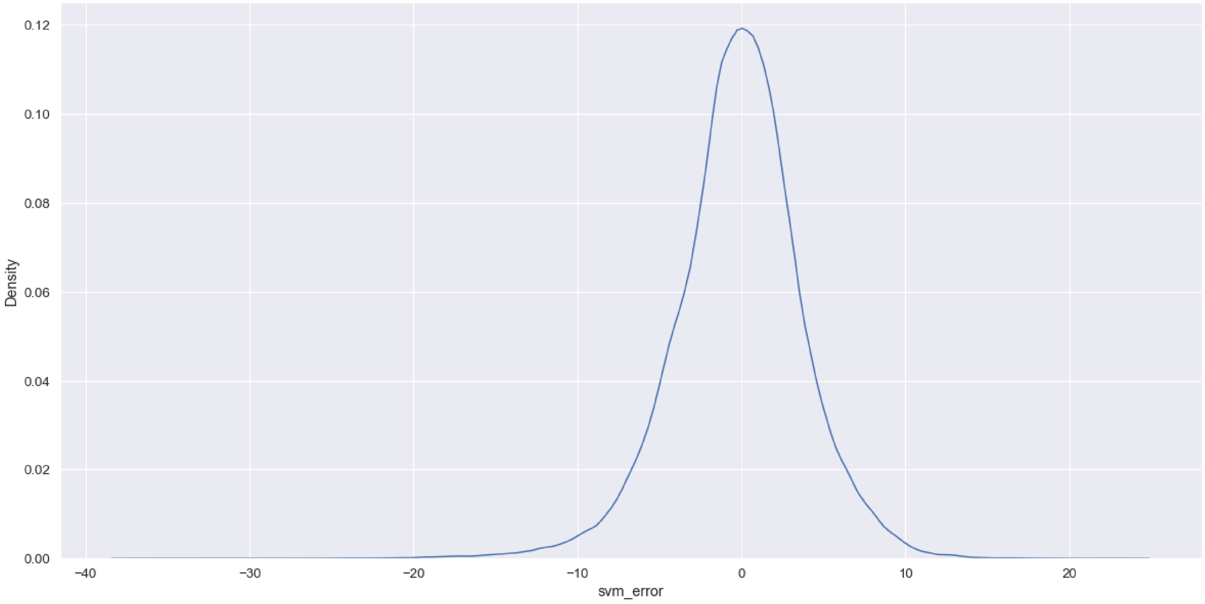


However, the model doesn't solely rely on the physical model temperature. It leverages additional features to enhance the precision of its predictions. Among these, zone temperatures and material thickness significantly contribute to the model's output. Additionally, some measurements related to Air and Gas Flow also exert a substantial impact, as indicated by their presence in the top 5 features in the model.

4.2.3 Evaluating Linear SVM

SVM exhibits a slightly better performance compared to Linear Regression. When examining the error distribution of the SVM model, it remains concentrated within the -10 to +10 range, with calculated statistics of $\overline{x} = -0.18$ and $s = 3.94$, as depicted in Figure 4.10. What's noteworthy is that the distribution's tip is smoother when compared to that of Linear Regression.

Figure 4.10: Error distribution for SVM grey-box model.



When observing the feature importance in the SVM model, some distinctions from the Linear Regression model come to light, as evident from Figure 4.11. While the physical model temperature retains its position as the most influential feature, the SVM model introduces variations in the importance of other features. In this case, Air and Gas flows, along with Zone temperatures exhibit substantially higher importance than thickness.

It's worth emphasizing that Air and Gas flows are also featured in the linear model. Nevertheless, their pronounced role in the SVM model is rationalized by the recognition that these parameters play a vital role in influencing the overall energy input and the atmospheric conditions within the furnace. Consequently, this could explain their contributions to the predictive capacity of the model.

Figure 4.11: Feature importance for SVM grey-box model.

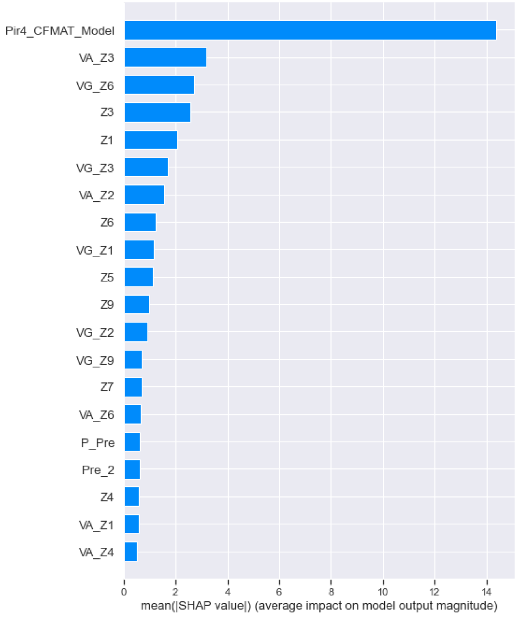
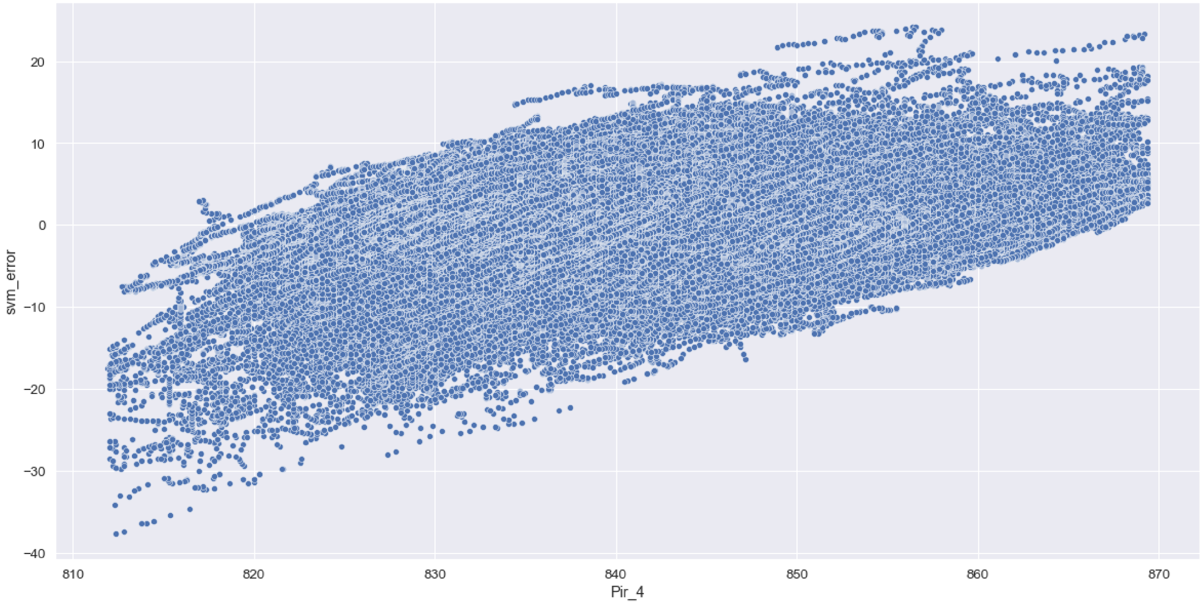


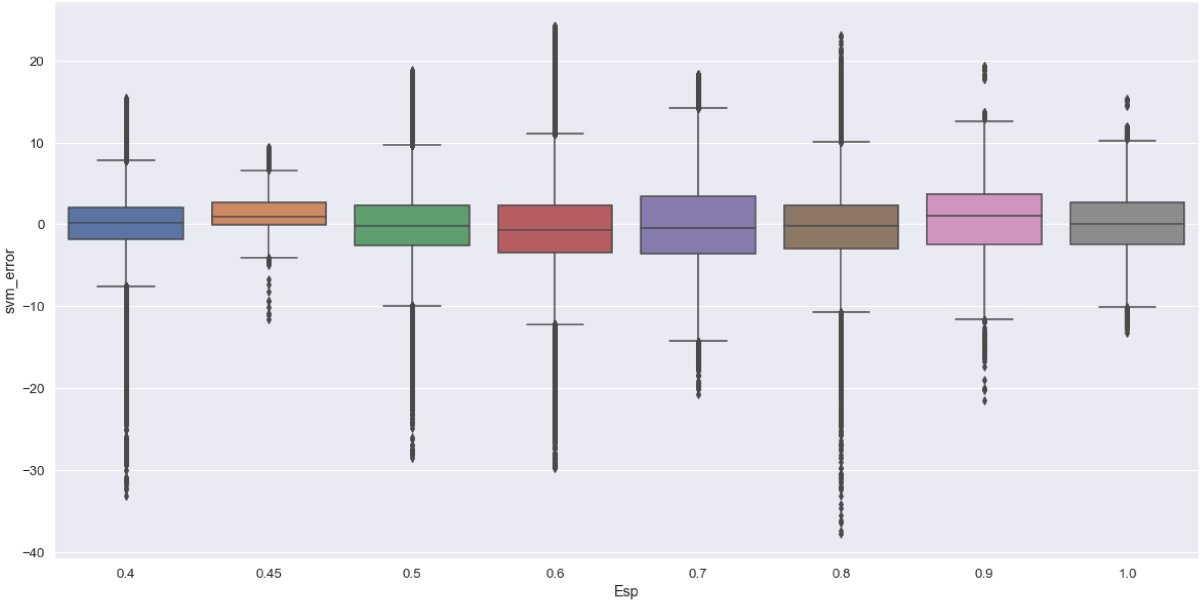
Figure 4.12 provides insights into the residual plot of the SVM regressor, revealing a pattern that closely resembles the one observed in the Linear Regression model. The similarity in the residual plot patterns suggests some commonality in error behavior.

Figure 4.12: Residual plot for SVM grey-box model.



SVM's superior performance compared to Linear Regression can be attributed to a notable characteristic within the error distribution—specifically, the position of the median. A careful examination of Figure 4.13 reveals a significant shift in the median's location for thicknesses exceeding 0.8mm. This movement towards closer proximity to zero signifies a noteworthy enhancement that plays a pivotal role in driving SVM's improved overall performance.

Figure 4.13: Boxplot of errors by thickness for SVM grey-box model.



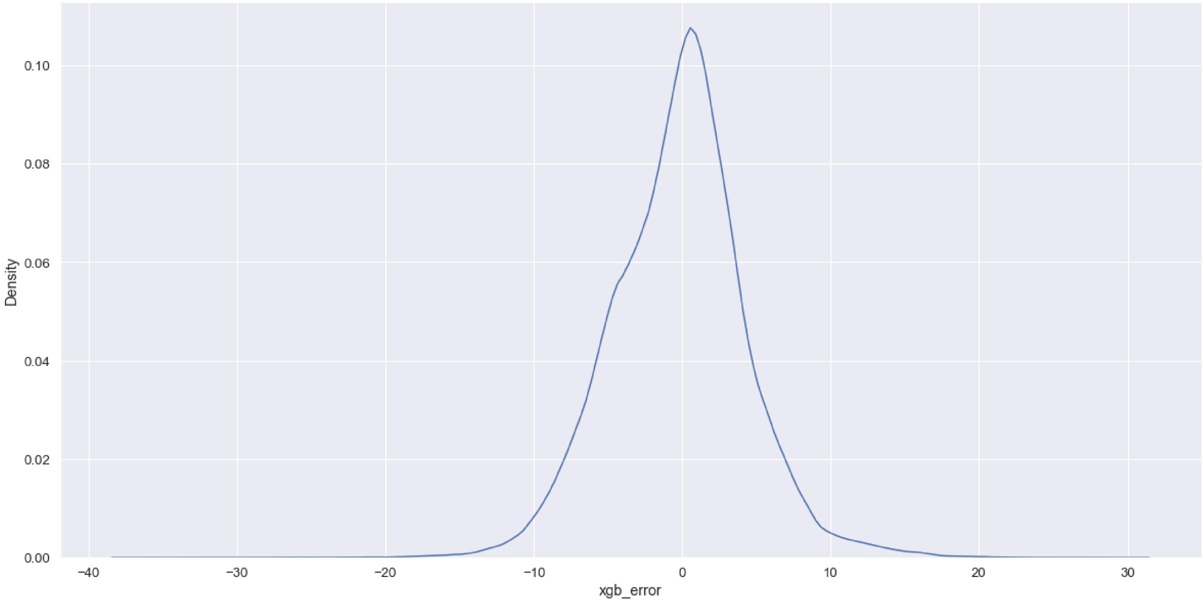
LinearSVR emerges as a strong contender for deployment as a soft sensor, demonstrating notable performance characteristics during our evaluation. While some inefficiencies were detected through the residual plot, it exhibits considerable strengths in its predictive capabilities.

The error distribution of the model is a key indicator of its reliability. Across various material thicknesses, SVR's errors are predominantly centered around zero. This consistent pattern assures us of the model's ability to provide accurate temperature predictions. Additionally, the standard deviations of the errors provide an estimate that the deviations from the actual values are likely to stay within a reasonable range, typically within +10 and -10 degrees of temperature. This level of precision is particularly promising, given the inherent complexities of pyrometry as a temperature measurement technique.

4.2.4 Evaluating XGBoost

XGBoost exhibits a marginally inferior performance compared to linear models, as indicated by the distribution depicted in Figure 4.14. This is affirmed by the statistics of this distribution, with a mean of -0.22 and a standard deviation of 4.56.

Figure 4.14: Error distribution for XGBoost grey-box model.



The feature selection process employed by XGBoost varies considerably from the previous models. Figure 4.15 illustrates the prominence of the physical model as the most crucial feature, followed by dimensional attributes such as thickness and width. Other features receive relatively less importance in this model's predictions.

Figure 4.15: Feature importance for XGBoost grey-box model.

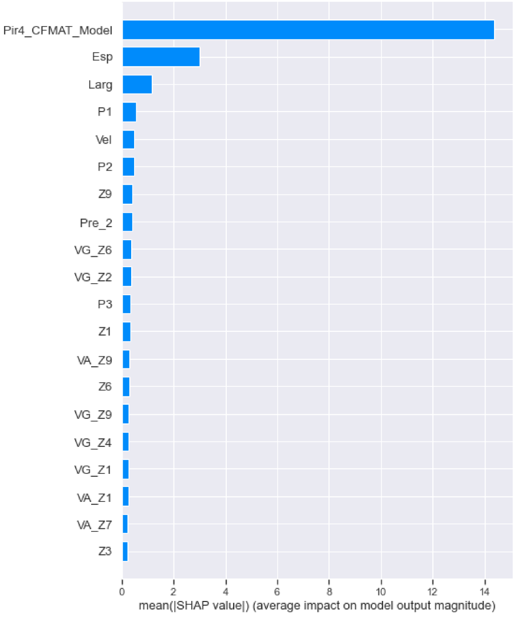
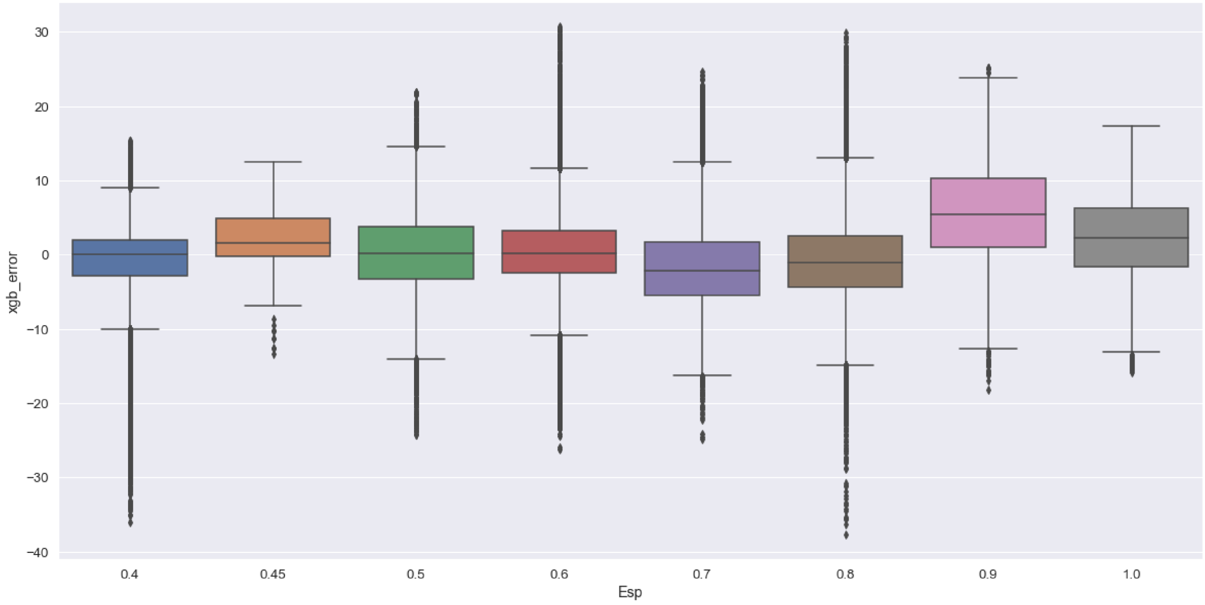


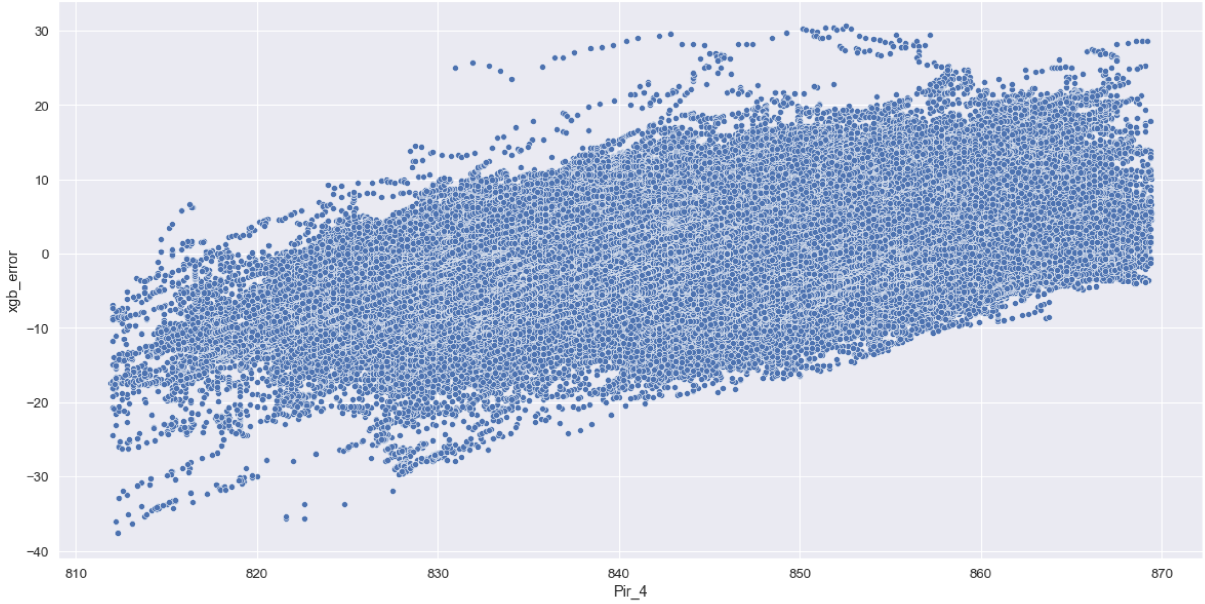
Figure 4.16, displaying the error by thickness, demonstrates that XGBoost exhibits more significant deviations from zero, particularly in thicker products. This divergence in error distribution is a plausible explanation for the model's comparatively lower performance, particularly in the context of thicker materials.

Figure 4.16: Boxplot of errors by thickness for XGBoost grey-box model.



In Figure 4.17, the residual plot maintains a similar pattern observed earlier. Errors continue to exhibit an increasing trend with temperature. And it should be remarked that XGBoost is a non-linear model, it operates differently from the previous linear models.

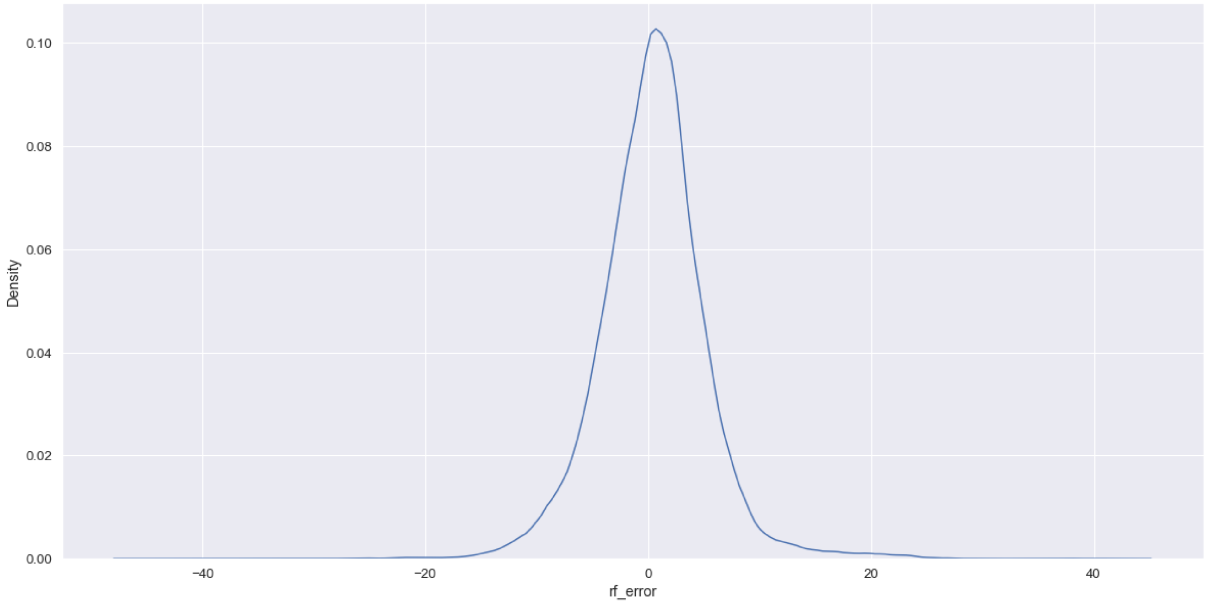
Figure 4.17: Residual plot for XGBoost grey-box model.



4.2.5 Evaluating Random Forest

Random Forest demonstrated the lowest performance on testing, with an error distribution as depicted in Figure 4.18, exhibiting mean and standard deviation of $\overline{x} = 0.43$ and $s = 4.9$.

Figure 4.18: Error distribution for Random Forest grey-box model.

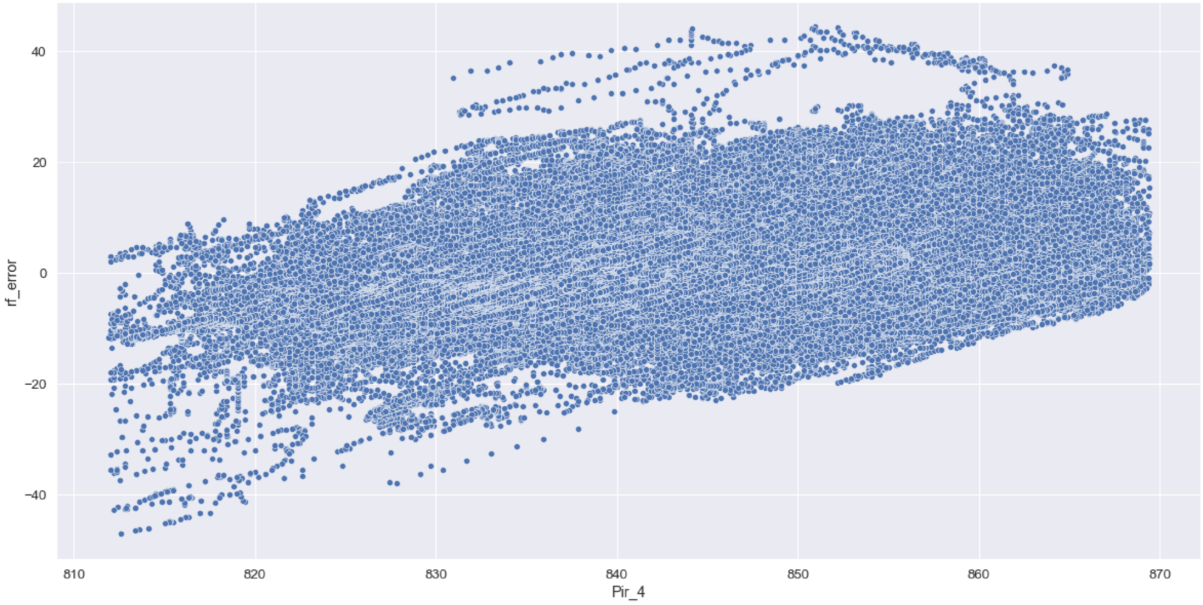


Random Forest, like XGBoost, is another non-linear model. While it underperformed when pitched with the other algorithms for the given task, it is interesting to explore its results to understand how it converges or differs from XGBoost. Both models are ensemble methods, but they employ different principles. Random Forest operates based on bagging, whereas XGBoost relies on boosting, resulting in distinct internal workings.

The residual error pattern, as seen in Figure 4.19, aligns with that of the other models, displaying the same increasing trend. The consistency of this pattern across four different models suggests that this trend is more likely associated with the underlying data characteristics rather than the choice of the modeling algorithm. There is the possibility of missing predictor variables within the model, which were not included in the collected data. This might contribute to the discrepancies observed in the model's performance.

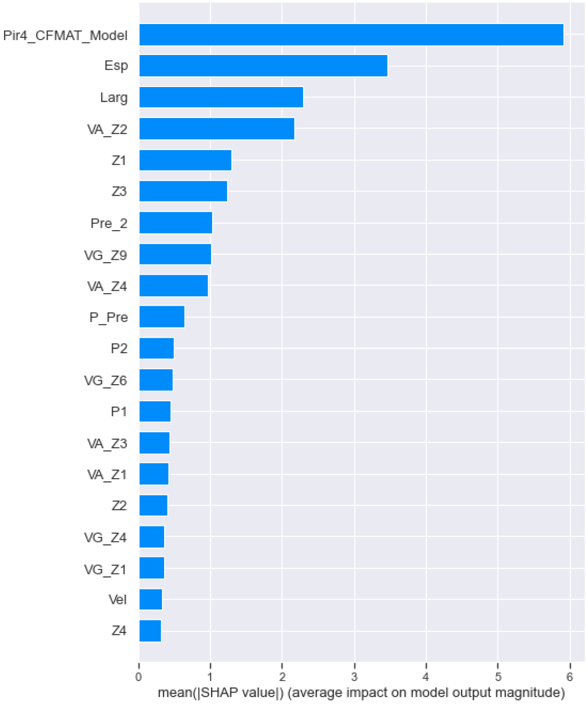
An interesting difference between Random Forest and XGBoost is their feature selection. XGBoost primarily focuses on the physical model and the dimensional attributes of the steel strip as its main features, while Random Forest incorporates other variables alongside these. Although Random Forest had the flexibility during hyperparameter optimization to explore using either root square or all features for each split, it achieved better results using the former approach. This led to an exploration of more features,

Figure 4.19: Residual plot for Random Forest grey-box model.



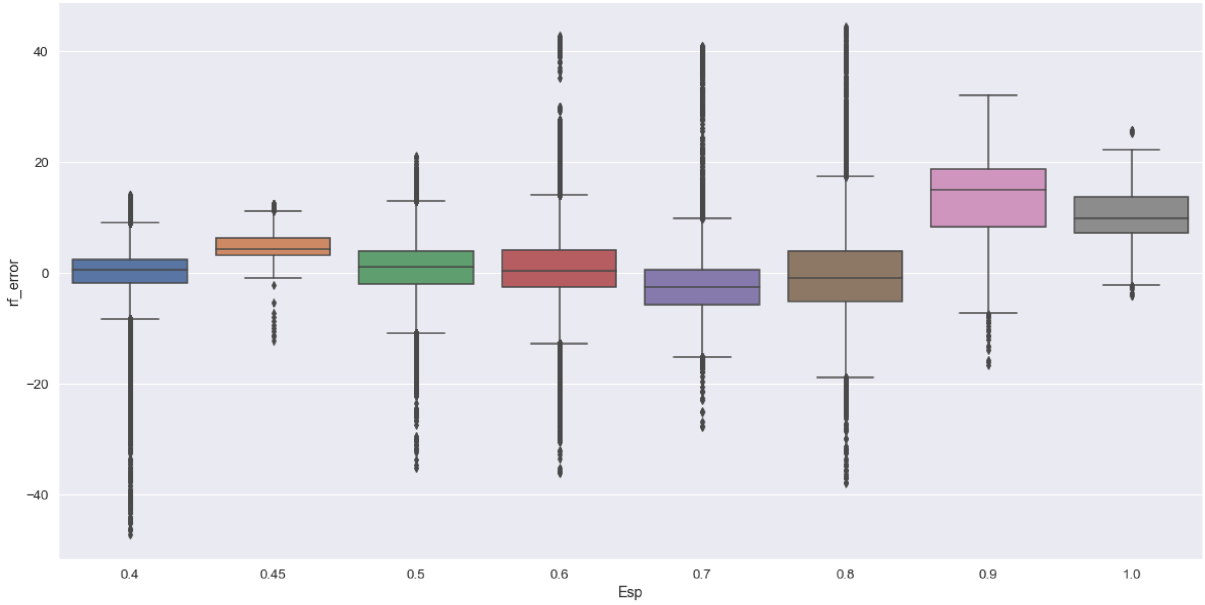
introducing additional variables into the process, as indicated in Figure 4.20.

Figure 4.20: Feature importance for Random Forest grey-box model.



Despite these differences in feature selection, the results for different thicknesses were strikingly similar to those of XGBoost, with performance degrading as thickness increased, as visualized in Figure 4.21.

Figure 4.21: Boxplot of errors by thickness for Random Forest grey-box model.



4.3 Final Observations

So far, the evaluation has primarily focused on the RMSE metric to gauge the performance of our models. However, it is essential to extend the analysis beyond numerical accuracy and look into the temporal characteristics of the reconstructed time series generated by the models. This allows to not only assess their quantitative performance but also examine the resemblance of the reconstructed curves to the actual temperature profiles.

In this investigation, the initial fifteen thousand records within the test dataset. As previously highlighted, the physical model follows the inverse general trend of the target temperature. Figure 4.22 also shows that Physical Model's output follows a similar trajectory, exhibiting corresponding fluctuations and variations.

SVM efficiently utilizes the physical model alongside other pertinent features to try to align itself with the Pyrometer temperature readings. The graph presented in Figure 4.23 underscores this behavior, showcasing that SVM succeeds in narrowing the gap between its predictions and the Pyrometer temperature. This visual inspection enhances our understanding of the models' capabilities and their ability to reproduce the underlying temporal patterns.

To provide a more in-depth perspective on the effectiveness of the grey-box modeling approach relative to a purely probabilistic one, an experiment was conducted. In this experiment, a linear model was constructed using only the Thickness variable as the predictor. The purpose of this exercise is to elucidate the impact of incorporating

Figure 4.22: Pyrometer 4 and Physical model temperature curves for the first fifteen thousand records of the test dataset.

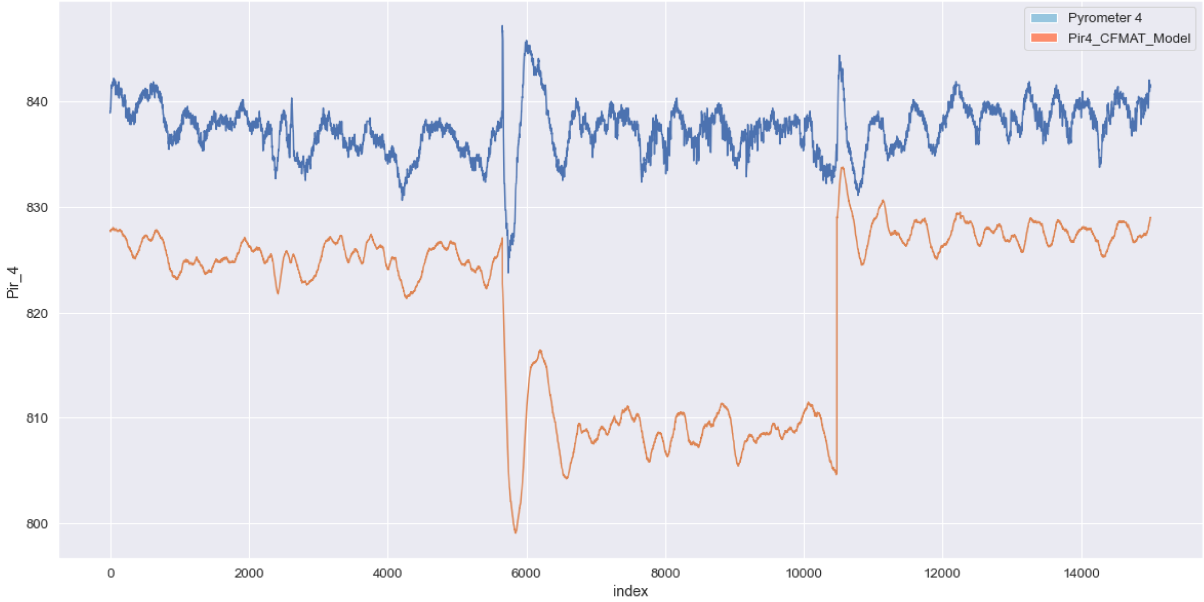
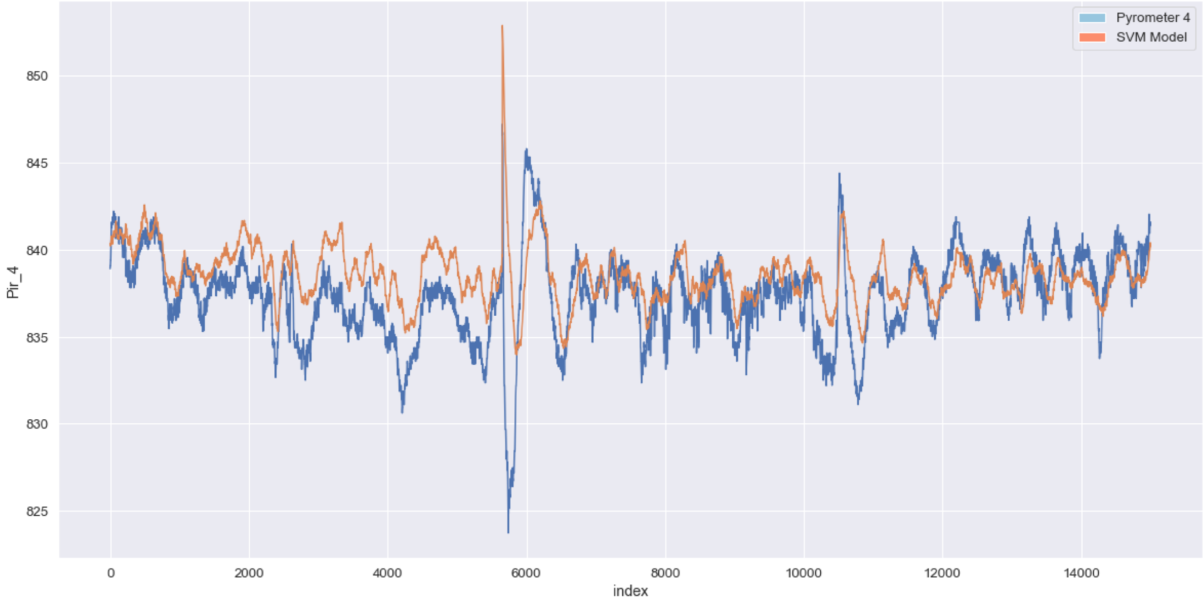


Figure 4.23: Pyrometer 4 and SVM temperature curves for the first fifteen thousand records of the test dataset.

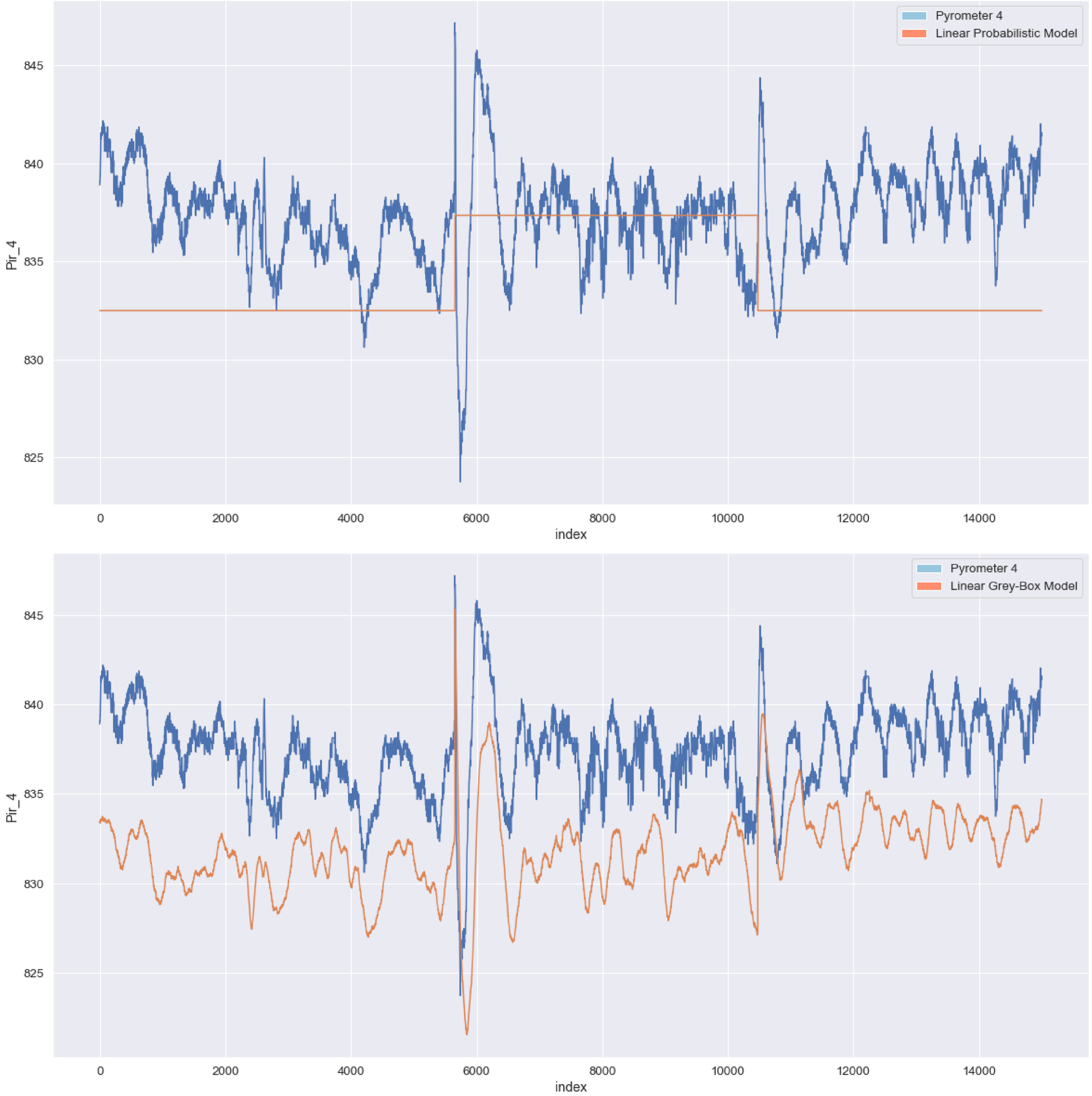


physical model information on the model's performance and ability to capture the target temperature profile using restricted data dimensionality.

Upon inspecting Figure 4.24, it's possible to observe that the linear model, when developed without the integration of physical model insights, exhibits a capacity to estimate the y-axis position of the temperature, as expected due to previous observations of the linear influence of thickness in the pyrometer temperature, but it lacks the accuracy and fidelity required for a loyal representation.

In contrast, the grey-box model, which uses the predicted error between the physical and real temperatures to construct the time series, and relies solely on the Thickness variable, displays a distinct behavior. It progressively converges towards the correct y-axis position of the temperature profile while simultaneously preserving the fundamental shape of the Pyrometer time series. This behavior underlines the capability of the grey-box model to enhance its representation of the time-series.

Figure 4.24: Linear Regression and Grey-Box Linear Regression model fitted with just thickness feature.



When comparing the tested models—Support Vector Machine (SVM), Linear Model, Random Forest, and XGBoost—distinctive patterns and performance characteristics emerge, providing an understanding of their efficacy in predicting Pyrometer 4 temperatures.

Support Vector Machine (SVM) reveals commendable performance despite inefficiencies noted in its residual plot. The distribution of errors across various material thicknesses is noteworthy, with errors predominantly centered around zero. Standard deviations indicate deviations within a manageable range of +10 and -10 degrees, highlighting SVM's robustness in providing accurate temperature predictions.

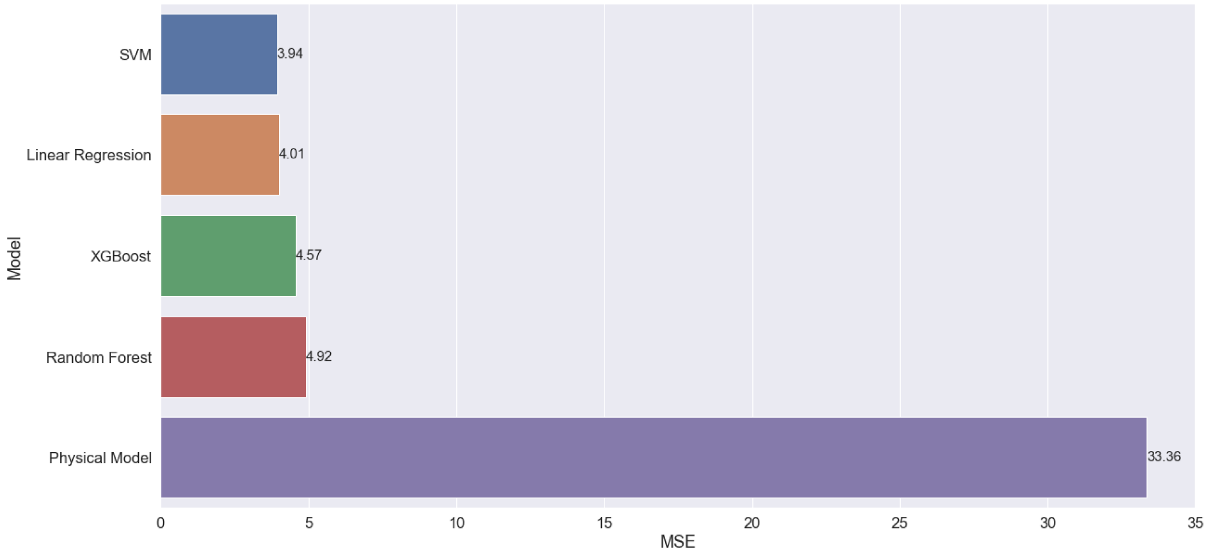
Linear Regression exhibits great generalizability, particularly across diverse product thicknesses, showcasing lower errors that contribute to a significant level of consistency. This emphasizes the model's reliability in predicting Pyrometer 4 temperatures.

XGBoost, while exhibiting slightly inferior performance compared to linear models, introduces a unique dimension by primarily focusing on the physical model and dimensional features. But has a more pronounced deviation from zero, especially for thicker products.

Random Forest emerges as the least performing model during testing, evident in the distribution's mean of 0.43 and a standard deviation of 4.9. The residual error closely mirrors patterns observed in other models, emphasizing that the observed trend is more inherent to the dataset than the specific algorithm employed. A distinctive feature of Random Forest is its emphasis on additional variables alongside the physical model and dimensional attributes.

The grey-box models exhibit deployment-ready results with consistently low errors compared to the pure physical model, as evident in Figure 4.25. Linear models, especially, showcase strong generalization across diverse product thicknesses. A hypothesis emerges that the inherent errors in the physical pyrometer readings might impact the performance of complex tree-based models, potentially explaining their superior training results. Further exploration is required to dissect the influence of these errors on overall model performance.

Figure 4.25: Mean Squared Error (MSE) comparison across models



Chapter 5

Conclusion

In this work, the ambition to create a Digital Twin for the Pyrometer 4 at the heart of the Annealing Furnace in RB4, led to the exploration of an array of models and strategies. The pursuit went through the deployment of the Physical Model, the exploration of the Grey-Box Modeling approach, and an in-depth evaluation of various machine learning models.

The objective of the deployment of the Physical Model provided by Aperam R&D was to have a model that incorporated phenomenological aspects of the process, aiming to capture the essence of the Annealing Furnace. Yet, it was clear that while a remarkable starting point, it could not single-handedly yield the desired accuracy.

The introduction of Grey-Box Modeling marked a turning point. It brought together the best of both worlds, combining the real-world physics encapsulated in the Physical Model with the adaptability and learning capabilities of machine learning models. This approach allowed to address the discrepancies between the model and the real-world Pyrometer readings more effectively. It acted as a bridge, minimizing the delta between the Physical Model and our target, the Pyrometer temperature.

Across various models, each demonstrated results that were satisfactory for it to be chosen as the final model. Linear Regression exhibited an unexpected resilience, outperforming more complex models in certain scenarios. SVM showcased promise, exhibiting consistent errors distributed within a relatively narrow range.

XGBoost and Random Forest are two non-linear models that were evaluated for the digital twin creation of the Pyrometer in Furnace 3. While they both exhibited less satisfactory performance compared to the linear models, they provided valuable insights into the complexity of the problem. The underperformance of these two models might be attributed to the inherent challenge of replicating the Pyrometer data due to the natural errors in its readings. These errors can introduce noise and incorrect patterns, which more complex tree-based models might try to wrongly replicate. Moreover, the behavior of the models regarding product thickness highlights a degradation in performance for thicker products, which could be explored in future works.

An examination of the key features highlighted the significance of the Physical Model in the prediction of Pyrometer temperature. It acted as a guiding light, illuminating the path for the machine learning models, and combining it with additional process variables for enhanced accuracy.

All the compiled results showed the transformative potential of Hybrid Modeling. By integrating the knowledge encoded within the Physical Model with the data-driven capabilities of machine learning, a powerful synergy was unlocked. This integration offers a bridge between the deterministic domain of physics and the probabilistic nature of data. It allows the model to harness the strengths of both realms while mitigating their inherent weaknesses.

5.1 Future Works

Expanding the dataset by collecting more parameters or introducing additional sensors could enhance the richness of the data. Moreover, exploring models with increased complexity, perhaps delving into deep learning architectures, may uncover even finer details of the Pyrometer's temperature prediction.

During the elaboration of this work, feedback was provided to Aperam R&D team regarding the results of the physical model, which could lead to new improved versions to be used in the construction of the grey-box model, in order to produce better performance.

Also, incorporating real-time data and online learning capabilities will enable the Digital Twin to adapt to changing conditions within the Annealing Furnace. Such adaptive models hold immense potential for robust predictive power.

The ultimate vision of a Digital Twin involves seamless integration with industrial control systems. Maturing the proposed architecture and coupling it with the furnace's control infrastructure, would make the Digital Model a decision-making companion for the industrial process.

References

R. D. Askeland and P. F. Pradeep. *Essentials of Materials Science and Engineering).* CL Engineering, Toronto, Canada, 2009. ISBN 0495244465.

B. R. Barricelli, E. Casiraghi, and D. Fogli. A survey on digital twin: Definitions, characteristics, applications, and design implications. *IEEE Access*, 7:167653–167671, 2019. doi: 10.1109/ACCESS.2019.2953499.

Y. Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009. ISSN 1935-8237. doi: 10.1561/2200000006.

C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics).* Springer, Berlin, Heidelberg, 2006. ISBN 0387310738.

L. Bitschnau and M. Kozek. Modeling and control of an industrial continuous furnace. In *2009 International Conference on Computational Intelligence, Modelling and Simulation*, pages 231–236, 2009. doi: 10.1109/CSSim.2009.26.

L. Breiman. Random forests. *Machine Learning*, 45:5–32, 10 2001. doi: 10.1023/A:1010950718922.

T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322. doi: 10.1145/2939672.2939785. URL https://doi.org/10.1145/2939672.2939785.

A. Gaikwad, B. Giera, G. M. Guss, J.-B. Forien, M. J. Matthews, and P. Rao. Heterogeneous sensing and scientific machine learning for quality assurance in laser powder bed fusion – a single-track study. *Additive Manufacturing*, 36:101659, 2020a. ISSN 2214-8604. doi: https://doi.org/10.1016/j.addma.2020.101659. URL https://www.sciencedirect.com/science/article/pii/S2214860420310319.

A. Gaikwad, R. Yavari, M. Montazeri, K. Cole, L. Bian, and P. Rao. Toward the digital twin of additive manufacturing: Integrating thermal simulations, sensing, and analytics to detect process faults. *IISE Transactions*, 52(11):1204–1217, 2020b. doi: 10.1080/24725854.2019.1701753. URL https://doi.org/10.1080/24725854.2019.1701753.

Gartner. Gartner says by 2020, at least 30 percent of industrie 4.0 projects will source their algorithms from leading algorithm marketplaces. URL: https://www.gartner.com/en/newsroom/press-releases/2017-03-21-gartner-says-by-2020-at-least-30-percent-of-industrie-4-projects-will-source-their-algorithms-from-leading-algorithm-marketplaces, 2017. Accessed: 2022-03-12.

R. Ghanem, C. Soize, L. Mehrez, and V. Aitharaju. Probabilistic learning and updating of a digital twin for composite material systems. *International Journal for Numerical Methods in Engineering*, n/a(n/a), 2020. doi: https://doi.org/10.1002/nme.6430. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.6430.

A. K. Ghosh, A. S. Ullah, A. Kubo, T. Akamatsu, and D. M. D'Addona. Machining phenomenon twin construction for industry 4.0: A case of surface roughness. *Journal of Manufacturing and Materials Processing*, 4(1), 2020. ISSN 2504-4494. doi: 10.3390/jmmp4010011. URL https://www.mdpi.com/2504-4494/4/1/11.

E. Glaessgen and D. Stargel. The digital twin paradigm for future nasa and u.s. air force vehicles. 04 2012. ISBN 978-1-60086-937-2. doi: 10.2514/6.2012-1818.

M. Grieves. Digital twin: manufacturing excellence through virtual factory replication. *White paper*, 1:1–7, 2014.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, 2009. ISBN 9780387848846. URL https://books.google.com.br/books?id=eBSgoAEACAAJ.

M. Hermann, T. Pentek, and B. Otto. Design principles for industrie 4.0 scenarios. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 3928–3937, 2016. doi: 10.1109/HICSS.2016.488.

Z. Huang, Y. Shen, J. Li, M. Fey, and C. Brecher. A survey on ai-driven digital twins in industry 4.0: Smart manufacturing and advanced robotics. *Sensors*, 21(19), 2021. ISSN 1424-8220. doi: 10.3390/s21196340. URL https://www.mdpi.com/1424-8220/21/19/6340.

A. Hürkamp, S. Gellrich, T. Ossowski, J. Beuscher, S. Thiede, C. Herrmann, and K. Dröder. Combining simulation and machine learning as digital twin for the manufacturing of overmolded thermoplastic composites. *Journal of Manufacturing and Materials Processing*, 4(3), 2020. ISSN 2504-4494. doi: 10.3390/jmmp4030092. URL https://www.mdpi.com/2504-4494/4/3/92.

H. Ko, P. Witherell, N. Y. Ndiaye, and Y. Lu. Machine learning based continuous knowledge engineering for additive manufacturing. In *2019 IEEE 15th International Con-*

ference on Automation Science and Engineering (CASE), pages 648–654, 2019. doi: 10.1109/COASE.2019.8843316.

Z. Liu, N. Meyendorf, and N. Mrad. The role of data fusion in predictive maintenance using digital twin. volume 1949, page 020023, 04 2018. doi: 10.1063/1.5031520.

Q. Min, Y. Lu, Z. Liu, C. Su, and B. Wang. Machine learning based digital twin framework for production optimization in petrochemical industry. *International Journal of Information Management*, 49:502–519, 2019. ISSN 0268-4012. doi: https://doi.org/10.1016/j.ijinfomgt.2019.05.020. URL https://www.sciencedirect.com/science/article/pii/S0268401218311484.

D. Montgomery, E. Peck, and G. Vining. *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, 2012. ISBN 9780470542811. URL https://books.google.com.br/books?id=0yR4KUL4VDkC.

A. Müller and S. Guido. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, 2016. ISBN 9781449369897. URL https://books.google.com.br/books?id=vbQlDQAAQBAJ.

K. P. Murphy. *Machine learning : a probabilistic perspective*. MIT Press, Cambridge, Mass. [u.a.], 2013. ISBN 9780262018029 0262018020. URL https://www.amazon.com/Machine-Learning-Probabilistic-Perspective-Computation/dp/0262018020/ref=sr_1_2?ie=UTF8&qid=1336857747&sr=8-2.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

J. Pfrommer, C. Zimmerling, J. Liu, L. Kärger, F. Henning, and J. Beyerer. Optimisation of manufacturing process parameters using deep neural networks as surrogate models. *Procedia CIRP*, 72:426–431, 2018. ISSN 2212-8271. doi: https://doi.org/10.1016/j.procir.2018.03.046. URL https://www.sciencedirect.com/science/article/pii/S221282711830146X. 51st CIRP Conference on Manufacturing Systems.

N. Repalle, R. Thethi, P. Viana, and E. Tellier. Application of Machine Learning for Fatigue Prediction of Flexible Risers - Digital Twin Approach. volume Day 1 Tue, November 17, 2020 of *SPE Asia Pacific Oil and Gas Conference and Exhibition*, 11 2020. doi: 10.2118/202461-MS. URL https://doi.org/10.2118/202461-MS. D013S104R005.

M. Samnejad, M. Gharib Shirangi, and R. Ettehadi. A digital twin of drilling fluids rheology for real-time rig operations. volume Day 1 Mon, May 04, 2020 of *OTC Offshore*

Technology Conference, 05 2020. doi: 10.4043/30738-MS. URL https://doi.org/10.4043/30738-MS. D011S010R006.

B. Scholkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001. ISBN 0262194759.

E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund. Industrial internet of things: Challenges, opportunities, and directions. *IEEE Transactions on Industrial Informatics*, 14(11):4724–4734, 2018. doi: 10.1109/TII.2018.2852491.

P. Stavropoulos, A. Papacharalampopoulos, and L. Athanasopoulou. A molecular dynamics based digital twin for ultrafast laser material removal processes. *The International Journal of Advanced Manufacturing Technology*, 108(1):413–426, May 2020. ISSN 1433-3015. doi: 10.1007/s00170-020-05387-7. URL https://doi.org/10.1007/s00170-020-05387-7.

S. Stieber, A. Hoffmann, A. Schiendorfer, W. Reif, M. Beyrle, J. Faber, M. Richter, and M. Sause. Towards real-time process monitoring and machine learning for manufacturing composite structures. In *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, volume 1, pages 1455–1458, 2020. doi: 10.1109/ETFA46521.2020.9212097.

A. Taufiqurrahman, A. G. Putrada, and F. Dawani. Decision tree regression with adaboost ensemble learning for water temperature forecasting in aquaponic ecosystem. In *2020 6th International Conference on Interactive Digital Media (ICIDM)*, pages 1–5, 2020. doi: 10.1109/ICIDM51048.2020.9339669.

F. A. Teixeira da Silveira. Controle Ótimo para forno de recozimento contÍnuo. Master's thesis, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 2009.

C. Wagner, J. Grothoff, U. Epple, R. Drath, S. Malakuti, S. Grüner, M. Hoffmeister, and P. Zimermann. The role of the industry 4.0 asset administration shell and the digital twin during the life cycle of a plant. In *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pages 1–8, 2017. doi: 10.1109/ETFA.2017.8247583.

L. Wright and S. Davidson. How to tell the difference between a model and a digital twin. *Advanced Modeling and Simulation in Engineering Sciences*, 7(1):13, Mar 2020. ISSN 2213-7467. doi: 10.1186/s40323-020-00147-4. URL https://doi.org/10.1186/s40323-020-00147-4.

Z. Zhao, S. Wang, Z. Wang, S. Wang, C. Ma, and B. Yang. Surface roughness stabilization method based on digital twin-driven machining parameters self-adaption adjustment: a case study in five-axis machining. *Journal of Intelligent Manufacturing*, 33(4):943–952, Apr 2022. ISSN 1572-8145. doi: 10.1007/s10845-020-01698-4. URL https://doi.org/10.1007/s10845-020-01698-4.