UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC
JOINVILLE
PROGRAMA DE PÓS-GRADUAÇÃO – COMPUTAÇÃO APLICADA

DANILO G. DE OLIVEIRA

PREDICTION OF ENERGY CONSUMPTION OF COLD ROLLING MILL USING MACHINE LEARNING

JOINVILLE
2022

DANILO G. DE OLIVEIRA

PREDICTION OF ENERGY CONSUMPTION OF COLD ROLLING MILL USING MACHINE LEARNING

Dissertação apresentada ao Programa de Pós–Graduação em Computação Aplicada do Centro de Ciências Tecnológicas da Universidade do Estado de Santa Catarina, como requisito parcial para a obtenção do grau de Mestre em Computação Aplicada.

Supervisor: Rafael S. Parpinelli

JOINVILLE
2022

Ficha catalográfica elaborada pelo programa de geração automática da Biblioteca Setorial do CCT/UDESC,

com os dados fornecidos pelo(a) autor(a)

Oliveira, Danilo Gonçalves de
Prediction of energy consumption of cold rolling mill using machine learning / Danilo Gonçalves de Oliveira. -- 2022.
66 p.

Orientador: Rafael Stubs Parpinelli
Dissertação (mestrado) -- Universidade do Estado de Santa Catarina, Centro de Ciências Tecnológicas, Programa de Pós-Graduação em Computação Aplicada, Joinville, 2022.

1. Energy prediction. 2. Machine learning. 3. Univariate regression. 4. Tandem Cold Mill. 5. Cold rolling. I. Parpinelli, Rafael Stubs. II. Universidade do Estado de Santa Catarina, Centro de Ciências Tecnológicas, Programa de Pós-Graduação em Computação Aplicada. III. Título.

DANILO G. DE OLIVEIRA

PREDICTION OF ENERGY CONSUMPTION OF COLD ROLLING MILL USING MACHINE LEARNING

Dissertação apresentada ao Programa de Pós–Graduação em Computação Aplicada do Centro de Ciências Tecnológicas da Universidade do Estado de Santa Catarina, como requisito parcial para a obtenção do grau de Mestre em Computação Aplicada.

Supervisor: Rafael S. Parpinelli

BANCA EXAMINADORA:

Rafael S. Parpinelli, Doutorado
Universidade do Estado de Santa Catarina (UDESC)

Membros:

Rafael S. Parpinelli, Doutorado
Universidade do Estado de Santa Catarina (UDESC)

Chidambaram Chidambaram, Doutorado
Universidade do Estado de Santa Catarina (UDESC)

Manassés Ribeiro, Doutorado
Instituto Federal Catarinense (IFC)

Joinville, 22 de Fevereiro de 2022

Aos meus filhos, pela inspiração para sempre seguir em frente.

ACKNOWLEDGEMENTS

Agradeço à ArcelorMittal pelo apoio para a condução desse trabalho. Agradeço em especial ao Sandro Sambaqui, que acreditou não só no potencial da parceria entre indústria e academia mas por sua inabalável confiança nas pessoas envolvidas. Minha gratidão também ao Marcello Torturela pela consideração, apoio e incentivo.

Ao José Francisco e Fabiano Miranda por moverem as montanhas que viabilizaram tudo isso. Mas, principalmente, por sua sabedoria, conselhos, direcionamentos e contribuições ao longo de todos esses meses. Sem os senhores, esse projeto, essa parceria, esse mestrado ficariam somente em um sonho distante.

Agradeço ao Professor Rafael Parpinelli por sua orientação enquanto eu trilhava esse caminho desafiador, pelas correções de curso em nossas inúmeras reuniões e pelas infindáveis tarefas, que foram fundamentais para que eu alcançasse os objetivos. Ao Eliton Machado e Pedro Serpa pela parceria e por dividirem a pressão; espero que tenham crescido tanto quanto eu nesses últimos meses.

Faço um agradecimento especial para minha esposa Janine Zytkuewisz. Seu apoio, companheirismo, compreensão, paciência e o seu amor foram os meus alicerces, meu porto-seguro. Foi você quem me ajudou a manter a sanidade e a serenidade nesse período tão conturbado de nossas vidas. Sem os momentos especiais em família para manter o equilíbrio eu, com certeza, sucumbiria.

E sou muito grato a Deus por Ele ter colocado em meu horizonte estas (e outras) pessoas tão especiais, cujos passos pude e sempre poderei seguir com a certeza de estar pisando em terreno firme e, principalmente, no caminho do bem.

ABSTRACT

Even though energy consumption has significant impact in the operational cost of tandem cold mills (TCM) of steel strips, not enough attention has been given to this important consumable throughout the years. Machine Learning techniques are becoming extremely common in the steel industry due to the high level of automation of the segment and large databases available. This document proposes a complete system capable of handling input data, training a machine learning algorithm, predicting the of energy consumption of a TCM and evaluating results. A comparison of the performance of Artificial Neural Networks (ANN), Random Forest (RF) and Extreme Gradient Boosting (XGB) algorithms with an existing statistical model concludes that the RF and XGB outperforms the other two on a product-to-product basis and on a monthly basis. The emulation of real-life model usage has also been carried out indicating that the proposed system is adequate to accurately predict energy consumption of TCM.

keywords: cold rolling mill, energy prediction, machine learning, univariate regression.

RESUMO

Apesar do impacto significativo do consumo de energia elétrica no custo operacional dos laminadores de tiras a frio (LTF) de chapas de aço, ainda não foi dada a devida atenção a esse importante consumível. Técnicas de aprendizagem de máquina estão ficando cada vez mais comuns na siderurgia devido ao elevado nível de automação do setor e vastos bancos de dados à disposição. Esse trabalho propõe o desenvolvimento de um sistema completo capaz de tratar dados de entrada, realizar o treinamento de algoritmo de aprendizagem de máquina, prever o consumo de energia de um LTF e avaliar os resultados. Um comparativo da performance de três algoritmos diferentes com um modelo estatístico existente permite concluir que as técnicas de aprendizagem de máquina apresentam resultados superiores tanto numa base bobina-a-bobina quanto numa base mensal. A simulação de aplicação do sistema em condições reais também demonstraram sua capacidade de previsão do consumo de energia do LTF.

Palavras-chave: laminação a frio, previsão do consumo de energia, aprendizagem de máquina, regressão univariável.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ABC	Artificial Bee Colony
ACO	Ant Colony Optimization
ACGDE	Adaptive Cross-Generation Differential Evolution
ANN	Artificial Neural network
ASE	Academic Search Engines
AWA GA	Adaptive Weight Approach Genetic Algorithm
AWPSO	Adaptive Weight Particle Swarm Optimization
BCO	Bee Colony Optimization
CART	Classification and Regression Trees
CGL	Continuous Galvanizing Line
DE	Differential Evolution
EX	Exclusion criteria
IBCO	Immune Bee Colony Optimization
IAGA	Improved Adaptive Genetic Algorithm
IBA	Commercial name given to specific data aquisition system
IC	Inclusion criteria
L2	Level 2
LDP	Load Distribution Problem
GA	Genetic Algorithm
GBM	Gradient Boosting Machines
MAE	Mean Absolute Error
ML	Machine Learning
MLP	Multilayer Perceptron Neural Network
MRE	Mean Relative Error
MSE	Mean Squared Error
NC	Natural Computing
OBL	Opposition-Based Learning
ONS	National System Operator
PIMS	Plant Integration and Management System
PLC	Programmable Logic Computer

PSO	Particle Swarm Optimization
REC	Regression Error Characteristic
RF	Random Forest
RMSE	Root Mean Squared Error
RQ	Research Questions
SD	Standard Deviation
SEC	Specific Energy Consumption
SI	Swarm Intelligence
SLR	Systematic Literature Review
SVM	Support Vector Machine
TCM	Tandem Cold Mill
TS	Tabu Search
TSP	Travel Salesman Problem
XGB	Extreme Gradient Boosting

CONTENTS

1 INTRODUCTION

Steel production is among the first industrial processes mankind has mastered. Because of its military and construction applications, steel has played a vital role in shaping the modern world (ROBERTS, 1978). However, steel production requires high amounts of energy and thus the adequate managing of this expansive resource has significant impacts on the business finance.

This work proposes the development of a system to predict the energy consumption of a tandem cold rolling mill by applying Machine Learning techniques. The next section details its motivations.

1.1 MOTIVATIONS OF THIS WORK

The cold rolling of steel strips is one of the highest consumers of electrical energy in the steel industry and this expense is among the highest in the yearly budget of the business, playing a significant role in the operational cost. The tandem cold rolling mill (TCM) studied in this work, composed of four mill stands and represented in Figure 1, has a total installed electrical power of 27.0MW. Consuming the equivalent to 45.000 family houses, this TCM is the highest consumer of electrical energy in the Santa Catarina State, Brazil.

The rolling process consists in reducing the steel strip thickness by plastic deformation, pressing it through a sequence of work rolls (typically ranging from four to five sequences) supported by backup roll, arranged in a mill stand, which can all be seen detailed in Figure 1. By reducing the strip thickness by up to 85%, the cold rolling process improves the surface quality and shape and provides internal energy for further steel processing at the galvanizing lines in which the final mechanical properties are reached with specific thermal cycles (ROBERTS, 1978).

The high energy consumption of the cold rolling process rests in the fact that steel presents a high resistance to plastic deformations. For this reason, steel is the material of choice of the automotive industry and is a crucial element in reinforcing the automobile structure to improve the safety of the driver and other occupants in the event of a car crash (BUCHMAYR; DEGNER; PALKOWSKI, 2018). Additionally, to achieve the required mechanical properties to comply with the safety regulations imposed to the automotive industry around the globe (BUCHMAYR; DEGNER; PALKOWSKI, 2018), novel steel grades are being developed every year with even higher resistance, increasing even further the energy requirements.

On a company level, the correct management of this expensive resource is key for the business financial results. However, it is even more critical to efficiently manage the electrical energy on a country wide perspective due to its implications on the daily lives of the entire population, such as blackouts, inflation, and others (ONS, 2021).

The adequate managing of electrical energy requires careful planning of the generation process since, in most cases, it is not possible to store the unused energy. In Brazil, Operador Nacional do Sistema (ONS, translated as System National Operator), a non-government organization, is responsible for managing the electrical energy generation and distribution (ONS, 2021). This agent requires that high energy consumers inform the expected consumption in advance to adequately plan the energy generation. Such procedure allows the organization to predict energy consumption peaks and prevent disruptions caused by insufficient generation (ONS, 2021). These high energy consumers, typically large industries, must purchase in advance energy they plan to consume at each specific moment and are subject to very high penalties if their actual consumption exceeds the purchased amount. On the other hand, there are no refunds if the consumption is lower than the contracted.

For all these reasons, predicting the energy consumption according to process and product conditions can have a significant impact on the production cost and, thus, on the profitability of the business.

Nowadays, basically every TCM relies on online classical rolling models for the prediction of many process set points, including electrical motor power. This variable can be integrated in the time domain and multiplied by the efficiency factors to yield the electrical energy consumption. These rolling models, however, as further detailed in Chapter 2, depend on many process variables, which are only available within minutes of product rolling. This short-term availability of process data imposes significant challenge for accurate application of classical rolling models for long-term energy prediction.

The current method for long-term energy prediction used by the TCM engineers is based on a statistical evaluation of the specific energy consumption which is ratio of energy divided and production weight. This method presents different constraints as it does not consider product or process information. When these conditions vary, as detailed in Chapter 5.1, this method presents significant prediction error.

Machine Learning (ML) is a field of research which allows the development of new computational tools for complex problem-solving (MOSAVI et al., 2019). The motivation, according to de Castro (CASTRO, 2007), is to provide alternative solution algorithms to problems that could not be satisfactorily resolved by traditional techniques. These algorithms are being applied in the steel industry since their early developments because this industry is highly automatized and sensored, building significant databases for decades (HU et al., 2019). An extensive systematic literature review following Petersen et al. (2008) methodology listed dozens of publications applying these techniques to TCM problems. However, none of these publications are related to the prediction of energy consumption, which supports even further this development.

The next section exposes the objectives of this work.

1.2 OBJECTIVES

As previously stated, the electrical energy has a significant impact on the operational costs of TCM and profitability. Thus, the industrial objective of this work is to support the reduction of energy costs by improving how it is managed and contracted. More specifically, better prediction of energy consumption could provide the means to reduce contract costs, minimize underutilization of contracted energy and overconsumption charges. In order to achieve this industrial objective, the main objective of this work is to predict the energy consumption of a TCM of steel strips for a month of production to reduce consumption charges and minimize the underutilization of contracted energy. As specific objectives, this work aims at defining the most adequate technique for modeling the energy consumption that would provide improved accuracy in comparison to the existing statistical method, adaptability to new rolling conditions such as new products, and that could be applicable to other steel processing lines and consumables, for example natural gas of galvanizing line furnaces.

A ML approach has been selected to solve the problem due to the complexity of the modeling and the long-term of the prediction. As this is a univariate regression problem, the author applies and compares the performance of three different algorithms: Artificial Neural Networks (ANN), Random Forest (RF) and Extreme Gradient Boosting (XGB). These algorithms are compared with the existing methodology for energy prediction with past (production) and future (sales) data to emulate the application of the system in real-industrial conditions.

1.3 STRUCTURE OF THIS DOCUMENT

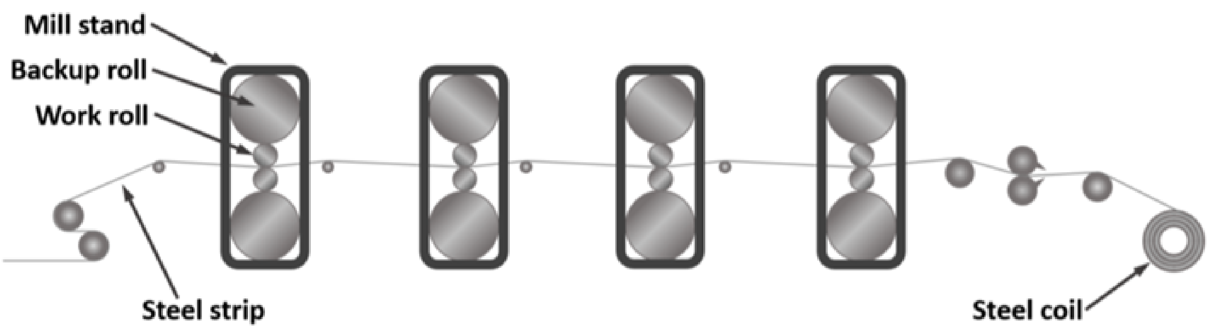
Chapter 2 presents a literature review, including a systematic review of the application of machine learning in the finishing lines of steel industries.

Chapter 3 describes the process of selecting the adequate database for this work while the model development is detailed in Chapter 4.

Model performance assessment, comparisons and emulation of real-life application of the system are exposed in Chapter 5.

Chapter 6 is reserved for final discussions and conclusions.

Figure 1 – Representation of a tandem cold mill.



Source: Elaborated by the author (2021).

2 LITERATURE REVIEW

The literature review of this work is divided in three main sections: i) the rolling theory state of the art and its application to the prediction of energy consumption (in Section 2.1), ii) an overview on machine learning algorithms applied to regression problems (Section 2.2) and iii) a systematic review of the application of these algorithms at the finishing lines of steel industries (Section 2.3).

2.1 ROLLING THEORY APPLIED TO ENERGY PREDICTION

Cold rolling is an efficient process for producing thin steel sheets for the subsequent stamping, where quality is a critical factor in terms of microstructure, surface texture and uniformity of mechanical properties and thickness (FRESHWATER, 1996).

There are several configurations of TCMs around the world. The main differences between them are in the number of rolls in each stand, which can range from two to twenty, and the number of stands itself, which can range from one to seven stands typically (ROBERTS, 1978). The Cold Strip Mill studied in this work has four stands, each with four rolls, being two work rolls (in direct contact with the steel strip) and two backup rolls.

Several theories for both cold rolling and hot rolling have been proposed in the last century after the pioneering works of Sibel and von Karman in 1924 and 1925 (ALEXANDER, 1972), in which they developed the first equations for predicting the rolling force and torque. The latter, in particular, is of great importance for forecasting the energy consumption considering that it allows to estimate the electric current when the rolling speed and the efficiency factor of the transformation of electrical energy into mechanical energy of the rolling mill drive train are also taken into account.

The most robust and complete rolling theory proposed in the 20th century is certainly the one proposed by Orowan in 1943 (OROWAN, 1943). In this famous article, cited by over 700 papers, Orowan (1943) describes the influence of complicating factors observed in the process: homogeneity of deformation, material flow variation during plastic deformation and the various friction regimes between the steel strip and the work rolls (ALEXANDER, 1972).

In Liu et al. (1985)'s perspective, despite being considered the "exact" theory of rolling, the approach Orowan adopted is based on a number of assumptions, including state plan of deformations and no elastic deformation of the steel sheet (LIU et al., 1985). The complexity of Orowan's approach has led several researchers to develop analytical solutions based on simplifications of key complicating factors raised by the original 1943 work (ALEXANDER, 1972).

The advent of modern electronic computers has improved the accuracy for industrial applications where the calculation time is critical to providing the mill with the necessary references for the adequate rolling of the next coil. However, Alexander (1972) concluded that "none of the existing rolling theories will be able to predict with precision the rolling torque"

(ALEXANDER, 1972) because the simplifications needed to solve the complex and non-linear equations involved in the rolling process lead to unacceptably incorrect results.

Freshwater (1996), on the contrary, was able to reach less than 1.0% precision between the predictions of the rolling torque and the values obtained experimentally in the rolling of copper under tension (FRESHWATER, 1996), a process very similar to the rolling of steel sheets, after adapting Alexander's solution (1972) method. However, his solution for torque calculation requires roll force, entry and delivery tensions, work roll radius and friction coefficient (FRESHWATER, 1996), variables that are only available within minutes of rolling each coil. This model has been extensively applied in different situations with excellent results (LIU et al., 2017) with minor adaptations over the years.

Liu et al. (2017) have modeled the rolling force based on an energy method, where the rolling force is approached as an energy minimization problem where the energy is the integration of the total power, decomposed as friction power, internal plastic deformation power, tangential velocity power, shear power and tension power. As in the Freshwater (1996) approach, these power compositions can only be calculated in very short time from the strip rolling.

Recent developments have focused on dynamic modeling of the rolling process which can be seen the in works of Hu e Ehmann (2000) and Kozhevnikov, Kozhevnikova e Bolobanova (2018). However, these are elaborated on the classical rolling theories thus requiring the same variables previously mentioned.

Finite element methods, which emerged in structural analysis, are rapidly expanding into several other areas where an exact solution cannot be found with traditional techniques, as stated by Liu et al. (LIU et al., 1985). When applied in cold rolling, this technique is able to simulate with high precision the plastic deformations suffered by the strip, the elastic deformation observed by the work rolls, and the rolling force and torque (SZűCS; KRáLLICS; LENARD, 2018). The accuracy of this method reported in several papers is incomparable to any other approach and finite elements has been used as a benchmark to assess the accuracy of other models (SZűCS; KRáLLICS; LENARD, 2018). However, the computational resources and/or the prolonged time required to reach solution convergence limit the usage of this finite element methods to case studies. The application of this method in online modeling for the prediction of process conditions is very unusual for the same reasons (SZűCS; KRáLLICS; LENARD, 2018).

As an alternative to the traditional analytical solution and the costly method of finite elements, modern approaches such as machine learning (ML) are becoming increasingly common in the steel industry (HU et al., 2019). In recent years, several works have focused on the application of artificial neural networks (ANN), random forest (RF) and other algorithms in the realm of ML, taking advantage of their versatility and great ability to generalize in complex and non-linear problems, reaching satisfactory results (SUN et al., 2018) in several problems in the steel industry, as exposed in the next section.

2.2 MACHINE LEARNING APPLIED TO STEEL INDUSTRY

The steel industry is one of the oldest industries in history of mankind, eventually determining the fate of entire civilizations by providing advantage to those mastering the steel technology and its production capacity in wars over the years (ROBERTS, 1978).

The finishing processes have been a part of the production chain of metals for many centuries. The first record of a rolling mill design is attributed to Leonardo da Vinci in 1480 (ROBERTS, 1978). Regardless of its longevity, the steel industry has a high level of automation driven by many reasons: pressures for improved health and safety of employees and higher productivity and profitability of the business are some examples.

As early as 1940 and 1950, the steel industry has made major investments in the instrumentation of the finishing processes and the application of early computers and data storages in the acquisition and processing of data and advanced modeling (ROBERTS, 1978). These investments allowed important advances in the mathematical formulation of process models and statistical analysis which led Orowan (1943) and Bland and Ford (1948) to publish their research, used to this day as important references to other researchers (LENARD, 2013).

Machine Learning (ML) can be described as an assignment of a specific task to a computer program and the machine learns if there is a measurable performance criteria which improves over time as the program acquires experience in performing the assigned task (RAY, 2019). So the machine learning process is based on data which positions the steel industry in a favorable condition for the application of such techniques. As an example, already in 1988, Miyabe, Biegl e Kawamura (1988) have applied the emerging artificial intelligence tool sets to detect faulty sensors in a hot strip finishing mill with significant improvements (MIYABE; BIEGL; KAWAMURA, 1988).

As a result of these early investments in automation, sensoring and database storage systems and application of cutting edge technology, steel finishing lines can be considered one of the most successful technological processes of modern industries (LENARD, 2013).

However, Hu et al. (2019) points out that the steel industry is composed of large-scale industrial complexes which make database integration extremely difficult. Even though very well instrumented and with a massive database at its disposal, steel enterprises have not achieved important milestones other sectors did years ago on process data integration and coordinated production optimization (HU et al., 2019) and this will be further detailed in Chapter 3.

With the objective of improving the intelligence level of the segment, Hu et al. (2019) affirms that some steel enterprises are building an overall management and control platform based on information technology (such as cloud computing, Internet of Things and big data) designed to collect, transform, process, monitor, manage and optimize data (HU et al., 2019). However, according to the author's point of view that these investments are aiming basically for improved cost control since this is an extremely competitive and cost driven industry.

The following section details the application of ML in general regression problems, which sheds lights on how to tackle the prediction of energy consumption.

2.2.1 Machine Learning applied to Regression Problems

Machine learning is considered as the study of computer algorithms enabling machines to learn and adapt to new data with little human intervention (AKANKSHA et al., 2021).

One of the different applications of machine learning is in regression problems in which it is used to model continuous variables (the model inputs) and make predictions of the behavior of other continuous variables (the model outputs or dependent variables). In regression problems there are labeled datasets and the output variable value is determined by input variable values (called supervised learning approach). The simplest regression algorithm is a linear regression where a straight line is fit to the dataset. However, real-life problems are usually more complex and more difficult to model (RAY, 2019). Typically, one dependent variable depends not only on one but on several factors. For example, the energy consumption of a house is dependent on the number of people living in that residence, on how many hours these people spend in the house, the type of equipment they have at their disposal, attached facilities and even at the average family income (GONZáLEZ-BRIONES et al., 2019). In summary, in a linear regression problem there is a one-to-one relationship between the input variable and the output variable. In a multiple linear regression, on the other hand, there is a many-to-one relationship, between a number of independent (input/predictor) variables and one dependent (output/response) variable (RAY, 2019).

Frequently, including more input variables in the regression algorithm does not mean the modeling will improve or provide better predictions since there could be a correlation between input variables (called multicollinearity). Multiple and simple linear regression have different use cases and one is not superior to the other. In some specific cases adding more input variables can make things worse as it results in over-fitting (RAY, 2019).

There are several data-driven algorithms that could be used to tackle the prediction of energy consumption of TCMs. In their review on the use of ML models for the modeling of electricity consumption, González-Briones et al. (2019) reported that the best performing algorithms were the Random Forests, Support Vector Machines, Decision Tree and Linear Regression (GONZáLEZ-BRIONES et al., 2019). However, this review was not focusing on an industrial application such as a tandem cold mill. Due to the already mentioned non-linearity and complexity of the problem at hand, the right approach has to be chosen otherwise there is risk of reaching wrong conclusions on pilot trials. A technology widely applied to the steel industrial sector is the artificial neural networks, described in the following section of this document. Another alternative that is explored in this work was the use of random forest algorithm, detailed in section 2.2.3.

2.2.2 Artificial Neural Networks

An Artificial Neural Network (ANN) is generally comprised of a collection of artificial neurons that are interconnected to perform some computation on input data and create output patterns (BROWNLEE, 2011). Even though these artificial neural networks are usually arranged in a complex architecture of several neurons arranged in different layers and with many interconnections, the calculations involved are relatively simple.

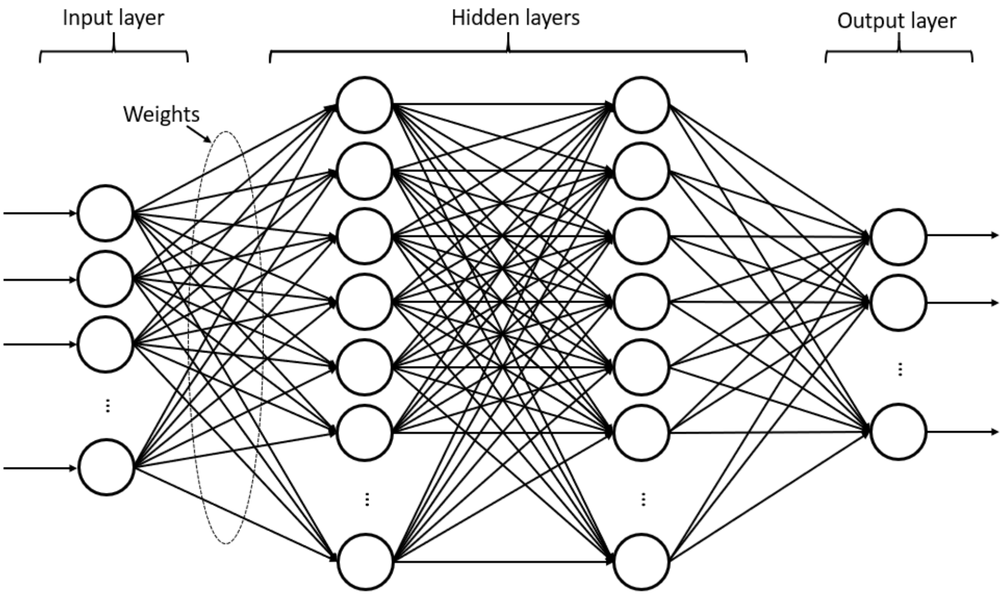
ANN is a type of model well established in machine learning and has also become competitive to regression and statistical models regarding usefulness (ABIODUN et al., 2018). Nowadays, they are widely used for uni- and multi-variable regression problems because of their excellent self-learning properties, fault tolerance, non-linearity, and advancement in input to output mapping (ABIODUN et al., 2018). Different artificial neural network architectures have been proposed since the 1980s, and one of the most influential is the multi-layer perceptrons (MLP) (ZHANG; PATUWO; HU, 1998). This architecture is the most used in many forecasting applications because of its inherent pattern-recognition capability, relatively simple configuration, and high speed (ZHANG; PATUWO; HU, 1998).

Typically, a MLP is composed of several layers of neurons, as indicated in Figure 2. The first layer is responsible for receiving the external information and feeding it to the ANN, while the last layer is an output layer where the problem solution is obtained (ZHANG; PATUWO; HU, 1998). There are one or more intermediate layers between these two layers, also called hidden layers, which are the core of the artificial neural network. The interconnections between the neurons located in the hidden layers are responsible for storing the input and output variables' relationship. Thus, ANNs are well suited for problems whose solutions require knowledge that is difficult to specify but for which there are enough data or observations available for the training of the network (ZHANG; PATUWO; HU, 1998).

Training is the iterative process that determines the weights (or the significance) of each input of each neuron in the network. These weights are initially defined in a stochastic fashion and are adjusted according to the error between the expected value (referred to as the target) and the actual output of the network calculation (ABIODUN et al., 2018). This scenario is known as supervised learning. As these weights are adjusted, and the error reduces in each iteration, one could say that the ANN is learning and that the weights express the knowledge the network is acquiring from the dataset. The problem with such data-driven modeling approaches, as ANNs, is that the underlying rules governing the behavior of the process are not always evident, and observations are often masked by noise in industrial conditions. Nevertheless, in some situations, this approach provides the only feasible way to solve real-world problems (ZHANG; PATUWO; HU, 1998).

Another advantage of the ANNs is their generalization capacity. After learning from the data presented to them during the training procedure, neural networks can often correctly infer the unseen part of a population even if the sample data contain noisy information (ZHANG;

Figure 2 – Representation of a multi-layer perceptron neural network.



Source: Elaborated by the author (2021).

PATUWO; HU, 1998). With the right architecture and configuration, they can also be used to learn and generalize from very complex, non-linear data-sets without any prior knowledge or assumptions about how each input variable influences the outputs.

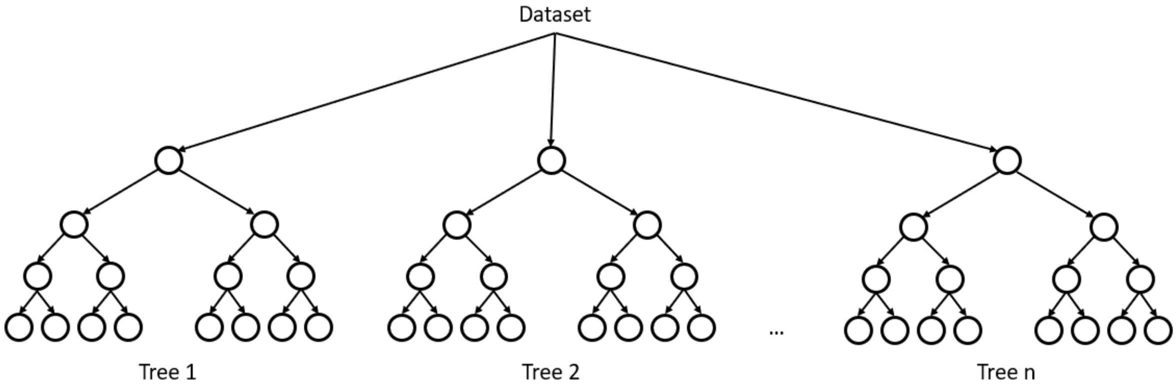
Rolling mills throughout the world are using ANNs in several applications (RATH et al., 2019). In the review conducted by Hu et al. (2019), they have listed 6 recent applications of ANN in the steel industry (HU et al., 2019). Rath et al. (2019) in their case-study have cited 31 researches where ANN have been applied to steel industry problems and Kim et al. (2021) have referenced 8 applications of the method (KIM et al., 2021).

2.2.3 Random Forest

Random Forest (RF) is a supervised machine learning algorithm methodology capable of performing both regression and classification tasks (MORE; RANA, 2017), introduced by Breiman in 2001, as an improvement of the Classification and Regression Trees (CART) method to improve stability (ANTONIADIS; LAMBERT-LACROIX; POGGI, 2021). Over the last two decades the use of the RF has received increasing attention due to the excellent results obtained and the speed of processing (BELGIU; DRAGUT, 2016).

In the RF algorithm, represented in Figure 3, the trees are created by drawing a subset of training samples through replacement (a bagging approach). This means that the same sample can be selected several times, while others may not be selected at all (BELGIU; DRAGUT, 2016). Typically, two thirds of the samples (in-bag samples) are used to train the trees while the remaining one third (out-of-the bag samples) are used in an internal cross-validation for estimating the model performance, according to Belgiu e Dragut (2016). Each decision tree is

Figure 3 – Representation of random forest.



Source: Elaborated by the author (2021).

independently produced without any pruning and each node is split using a user-defined number of features, selected at random. By growing the forest up also to a user-defined number of trees, the algorithm creates trees that have high variance and low bias and the final classification is taken by averaging the class assignment probabilities calculated by all trees (BELGIU; DRAGUT, 2016).

Differently from ANN which yields hundreds of search results related to the steel industry, there are very few applications of random forest algorithm in finishing lines of steel manufactures with less than a dozen being found when the term is searched in the title, abstract or keywords.

Kim et al. (2021) listed one single research paper applying this method in their review of artificial intelligence algorithms for industrial applications (KIM et al., 2021) while Hu et al. (2019) and Rath et al. (2019) have not even mentioned the algorithm. The reason for this unexpectedly low application of this powerful regressor in the steel industry are unknown to this author.

2.2.4 XGBoost

Similarly to RF, the Extreme Gradient Boosting (XGBoost or XGB) is a supervised machine learning algorithm based on CART. It is an improvement of the Gradient Boosting Machines (GBM) (CHEN; GUESTRIN, 2016) where the training of each tree (or estimator) depends on the previously trained trees and the learning procedure targets at constructing trees maximally correlated with the negative gradient of the loss function (SAGI; ROKACH, 2018). In GMB, the models usually have many shallow trees, while RF, on the contrary, has fewer but deeper trees (SAGI; ROKACH, 2018).

XGBoost optimizes the GMB algorithm basically in two aspects: i) use the first and second-order derivatives to obtain more accurate predictions (GMB uses only first-order derivative) and ii) prevent over-fitting by including a term to control model complexity (GUO; XU; YAO, 2021). Moreover, XGBoost includes several code optimizations that enhance parallelism, tree building and memory utilization (GUO; XU; YAO, 2021).

Even though XGBoost is a recently developed algorithm, dating back to only a few years (2016), researchers are already applying it in the cold rolling of thin strips. Chen et al. (2020) have applied XGBoost in the prediction of rolling force of dual-phase automotive steels (DP590) with high accuracy (CHEN et al., 2020).

2.3 SYSTEMATIC LITERATURE REVIEW

Mosavi et al. (2019) have systematically reviewed the literature focusing on the application of machine learning (ML) methods in energy systems. They have listed more than 100 references, but none are related to the steel industry (MOSAVI et al., 2019).

Hu et al. (2019) conducted a very comprehensive literature review on the use of flexible computing (or soft computing) applied to the steel industry, listing more than 60 references in the area. The authors noted that several papers were published using different techniques within soft computing to predict rolling forces and torques, flatness and its actuators, production schedule, thickness, strip and roll temperatures, mechanical properties of the finished material, internal stresses, among others process and quality variables, as can be seen by Figure 4. Hu et al. (2019) also observed a growing trend in recent years of using these techniques not only in modeling the process of production, but also in control strategies of its various variables, reducing the losses and improving the quality of the final product. Even though the work of Hu et al. (2019) provides valuable insight into recent publications in this field, the authors have not performed the review systematically.

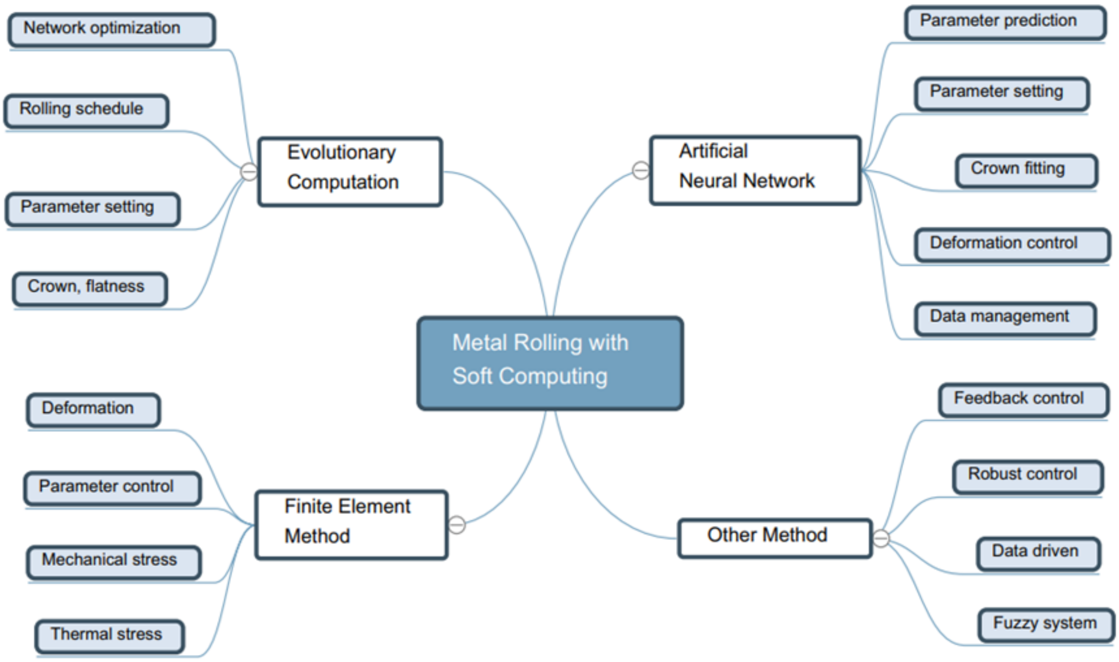
The manufacturing of steel coils mainly consists of three stages: the primary steelmaking, the coil making and finishing lines (VERDEJO; ALARCó; SORLí, 2009). The primary steelmaking stage consists in the production of semi-finished steel slabs from raw materials such as iron ore and coal. The coil making stage starts by reheating of slabs and rolling them into coils in accordance with customer requirements. The finishing lines apply selectively several processes to achieve final order specification to obtain the finished steel product. Operations such as cold-rolling, annealing, tempering, galvanizing, or coatings are executed in these lines (VERDEJO; ALARCó; SORLí, 2009). Due to the similarities in how the steel is handled and the type of equipment involved in these operations, this review is focusing only on the finishing lines of the steel productions chain.

Petersen et al. (2008) presents the methodology for performing a systematic literature review (SLR). According to them, the paper's research questions (RQs) reflect the main goals of the study which often can be to provide an overview of a research area or identify the type and quantity of results within that field, for example. In this review, finalized in March 2022 and comprising papers published until then, the author focuses in elucidating the following questions:

RQ1 What is the state-of-the-art of energy consumption prediction for tandem cold mills?

RQ2 What are the Machine Learning techniques being applied to predict or to optimize

Figure 4 – Soft computing approaches of metal rolling and their applications.



Source: HU et al. (2019).

energy consumption in the finishing processes (pickling line, tandem cold mill, batch or continuous annealing, galvanizing lines and temper rolling) of the steel industry?

RQ3 Any of these works are focused on the prediction of electrical energy consumption or power requirement in the tandem cold mill?

The objective of the RQ1 is to assess the state-of-the-art of classical models applied to tandem cold mills focusing on the energy consumption predictions. This search is relevant because, as previously mentioned in Section 2.1, energy was, for decades, a result of rolling torque, which is also calculated from rolling force and, hence, prone to the accumulation of modeling errors. Considering this point as a focus of research, the expectation is to find innovative model approaches that could provide a simpler, direct calculation of rolling energy.

RQs RQ2 and RQ3, on the contrary, intent to evaluate not classical modeling but the application of ML techniques to solve problems related to the finishing lines of steel industry and if any of these works have focused on the prediction of energy consumption or power requirements of a TCM.

After defining the RQs, the search for the relevant papers must be done. The selected Academic Search Engines (ASE) to perform such task were IEEE-Xplore and Elsevier's Scopus (Web of Science have also been searched but returned the same references already found on the other two ASE). The reason for this selection is based on the fact that these two ASE are recognized in the academia for their vast database of peer-reviewed documents, which significantly reduces the chance of including gray literature in the results. The next step is to define the search string which is used to automatize the initial filtering of the ASE database. Even though this search string will return a limited number of results, the scope of each paper must be aligned with the RQs this review intends to elucidate. To attend this objective, the following inclusion (IC) and exclusion criteria (EX) were defined. Documents fitting the IC are eligible for reviewing but only if not infringing any EX criteria.

In addition to satisfying all the criteria previously defined, a qualitative review of title, abstract and keywords of the publications are going to be assessed in order to verify they fit in the requirements of this review.

The inclusion criteria are:

IC1 Include paper if, and only if, it comprises the following criteria: energy consumption prediction for tandem cold mills OR Machine Learning techniques applied to energy related work in finishing line of the steel industry.

  IC1.1 Works related to redistribution of rolling loads and production schedules are being considered in this review as energy optimization techniques. Further details are provided in the next sections.

The exclusion criteria are:

EX1 Remove works that are focusing on other process variables as, for example, but not limited to, temperature, roll force, strip flatness, thickness, crown, etc.

EX2 Remove papers concentrated on upstream processes such as hot strip mill, plate mill, continuous casting, etc.

EX3 Remove works focusing on one specific steel grade, chemistry or metallurgy improvement, characteristics, or developments.

EX4 Remove papers developing algorithms and techniques to detect strip defects, improve or manage quality.

EX5 Remove papers to which the author of this review do not have full access.

EX6 Remove duplicates.

In order to answer the RQ1 the search strings shown in Table 1 were defined as inputs for each ASE. The number of publications returned by each ASE are also indicated in Table 1. The 146 documents resulted from the search were evaluated and only 19 passed the qualitative filter.

Bidabadi et al. (2019) have performed a thorough evaluation of the sensitivity of several rolling parameters on the energy consumption in a roll forming process. Legrand et al. (2010) estimated the impact of friction in energy consumption and proposed new cooling techniques to reduce it. All the other publications have focused on redistributing rolling forces and tensions to improve power balance, a classical rolling problem, but with an energy approach. However, none of these publications directly predicted energy consumption or proposed different approaches for rolling modeling, indicating that the classical rolling theory is still valid, as presented in Section 2.1.

The RQ2 and RQ3 are answered by the search strings explicit in Table 2. The "Search String 1" resulted in only 5 publications clearly indicating that, even though energy is an important factor in the operational cost of TMC, not enough attention has been given to it when

Table 1 – Search string defined to answer RQ1.

Connector	Where	What	Scopus	IEEE
	Title / Abstract / Keywords	cold OR tandem	621,085	22,882
AND	Entire Text	steel AND (mill OR line OR roll*)	7,141	512
AND	Title / Abstract / Keywords	electric* OR energy OR power OR load	1,087	309
AND	Entire Text	consumption	103	43

Source: Elaborated by the author (2021).

Table 2 – Search strings defined to answer questions RQ2 and RQ3.

Search String 1				
Connector	**Where**	**What**	**Scopus**	**IEEE**
	Title / Abstract / Keywords	cold OR tandem OR galvanizing OR hot dip OR annealing OR temper	1,184,484	64,324
AND	Entire Text	steel AND (mill OR line OR roll*)	8,860	867
AND	Title	Energy	74	10
AND	Entire Text	artificial intelligence OR swarm OR neural network OR genetic algorithm OR machine learning OR deep learning OR evolutionary algorithm	5	0
Search String 2				
Connector	**Where**	**What**	**Scopus**	**IEEE**
	Title / Abstract / Keywords	cold OR tandem OR galvanizing OR hot dip OR annealing OR temper	1,184,484	64,324
AND	Entire Text	steel AND (mill OR line OR roll*)	8,860	867
AND	Title / Abstract / Keywords	gas OR electric OR comb* OR furnace OR energy OR therm* OR power OR distr* OR load OR schedul*	4,345	839
AND	Entire Text	artificial intelligence OR swarm OR neural network OR genetic algorithm OR machine learning OR deep learning OR evolutionary algorithm	387	93
AND NOT	Title / Abstract / Keywords	hot strip OR hot rolling OR plate OR casting OR mechanical OR chemistry OR chemical OR defect OR surface OR tube	171	68

Source: Elaborated by the author (2021).

ML techniques are concerned.

The "Search String 2" was defined to broaden the search scope to not only electrical energy but to other sources of energy such as thermal energy, which are relevant for the finishing galvanizing lines. These documents have been deeply evaluated.

Many publications aimed at applying the ML techniques to the previously mentioned redistribution of TCM power balance. Others have modeled the galvanizing line furnace and its natural gas consumption. Other documents, following the innovative work of Verdejo, Alarcó e Sorlí (2009), focused on optimizing the line scheduling, a problem that consists of sequencing the different coils waiting to be produced in a cost-effective way. This SLR revealed that the algorithms applied to solve these problems tackled by the researchers (mainly the line scheduling and rolling load distribution) are approached by metaheuristic algorithms such as genetic algorithms, swarm particles, etc. answering to RQ2, as can be seen in the summary exposed in Table 3. This makes perfect sense when the nature of these problems is evaluated: multi-objective optimization.

It was also observed that major steelmakers are investing in research in this fields due to its importance on operational costs. This conclusion can be made from the level of confidentiality most papers treat the objective functions and how they describe the algorithms, often omitting important aspects of the implementation and functionality to protect intellectual property.

These are valuable documents as they expose the benefits of this kind of implementation and encourage other researchers in exploring ML algorithms and techniques to solve such complex problems. However, aside from a previous paper from this author (OLIVEIRA et al., 2020), which predicts power requirements of a TCM, all these publications provide only insights for future developments and minor contributions to this work.

Table 3 – Classification of publications according to algorithm and problem.

Algorithm	TCM power balance	Furnace of galvanizing line	Line scheduling optimization
Swarm Intelligence	Che et al. (2009), Li, Liu e Wang (2009), Wang et al. (2020)		Wang et al. (2011), Zhang e Zhu (2012), Fernandez et al. (2014), Arumugam, Chandramohan e Murthy (2011)
Genetic Algorithm	Wang et al. (2000), Wang, Tieu e D'Alessio (2005), Che et al. (2010), Wang et al. (2010), Poursina et al. (2012), Zhao et al. (2014)	Qu et al. (2018)	Kapanoglu e Koc (2006), Cohen, Foxx e Alul (2019), Martínez-de-Pisón et al. (2011)
Tabu Search	Bu et al. (2016), Bu, Yan e Zhang (2018)		Gao, Tang e Wang (2008), Yang e Tang (2008), Tang e Gao (2009)
Evolutionary Algorithm			Nastasi, Colla e Seppia (2014), Zhang, Zhao e Shao (2016)
Differential Evolution	Yong, Lei e Yu (2016)		

Source: Elaborated by the author (2021).

3 DATA GATHERING AND PROCESSING

Since the early 1980, ANNs and other ML techniques are being widely used in problems where solutions require knowledge that is difficult to specify but for which there are enough data or observations available for training of the network (ZHANG; PATUWO; HU, 1998). One of the greatest advantages of ML is that, after learning the underlying patterns of the data during training procedures, they can correctly infer the behavior of the unseen population of the data even if the data sample contain noisy information (ZHANG; PATUWO; HU, 1998). However, this low sensitivity of ML to noisy data does not mean unprocessed data should be used in the training of the model or in the prediction but the contrary: significant effort must be put to ensure the best quality in the dataset or the algorithm might learn patterns from noise, which will significantly reduce the model performance (SMITH, 1993).

For the work presented in this paper, three separate databases were initially available: Level 2 (L2), IBA (which is a commercial name defined by the developer of the system) and Plant Integration and Management System (PIMS). These datasets were collected directly from the TCM databases and are protected by non-disclosure agreements.

The first one, Level 2 database, is the official production database responsible for storing basic product information such as customer requirements, measured dimensions, date of production and quality information. It is organized in a coil-to-coil basis where each line of the dataset corresponds to one rolled coil (product) and it is used as the reference in this work to verify the other datasets. In addition to product information, the L2 also stores other type of data and the monthly energy consumption is also available in this database. However, this information is not granularly stored on a product level but only the total month consumption which makes it unsuitable for any model training.

IBA is a dataset resulting from the interaction of an electronic board connected directly to the programmable logic computer (PLC) of the production line. This PLC is responsible for the entire cold mill process automation including sensor reading, actuators and user interfaces. IBA, through this electronic board, is capable of reading the PLC memory in real time and communicating this information through optical fiber to a dedicated server equipped with solid state drives which store this information at a sampling rate as low as 40 micro-seconds. This architecture is very reliable and is currently used by the process engineers to assess product quality and process performance. However, due to high-installation costs, this equipment is constrained to the most essential PLCs. Many important energy consumers are not storing any information in IBA. Another downside of IBA for this work is that energy is not directly measured since the process PLCs have no interest in this information. The PLC measures, controls and stores other process variables such as electric current and voltage. Motor power and energy are the result of the interaction of these process variables from the equipment standpoint.

PIMS, on the other hand, acquire data not with optical fiber but regular Ethernet cables and at 1.0 second only (for some variables, most are at even slower sampling rate) in order to increase storage capacity. One important positive aspect of PIMS database is that it is connected directly to the system responsible for reporting the TCM energy consumption (called CCK), which provides a great advantage as detailed in the next subsections.

Figure 5 presents a summary of the main energy consumers and their presence in IBA and PIMS databases. As Figure 5 shows, IBA comprises only the main motors of the TCM and does not acquire any information related to approximately 30% of the energy consumed in the grid. On the other hand, PIMS comprises equipment such as overhead cranes, welding machines and air conditioning, which are not directly linked to process information and can significantly reduce the accuracy of the regression algorithm.

With these three databases available, dozens of process variables which are knowingly impacting the rolling torque, power and energy (FRESHWATER, 1996) were also at the disposal of the author to be selected and used in the model training and validation. However, most of these variables are highly dependent on specific process conditions such as work roll diameters, roll temperature and friction coefficient (FRESHWATER, 1996) and these conditions can only be accurately determined within hours of the time of production.

With the objective to reduce the model dependency on these short-term process variables the author decided to consider as inputs only the most basic product data and the process information that could be estimated from scheduling information. A preliminary set of variables has been investigated and the evaluation of the data indicated that the most relevant variables for model accuracy were: entry coil thickness, exit coil thickness, total coil reduction, coil width, coil weight, coil hardness, processing time and average speed (detailed in Table 4).

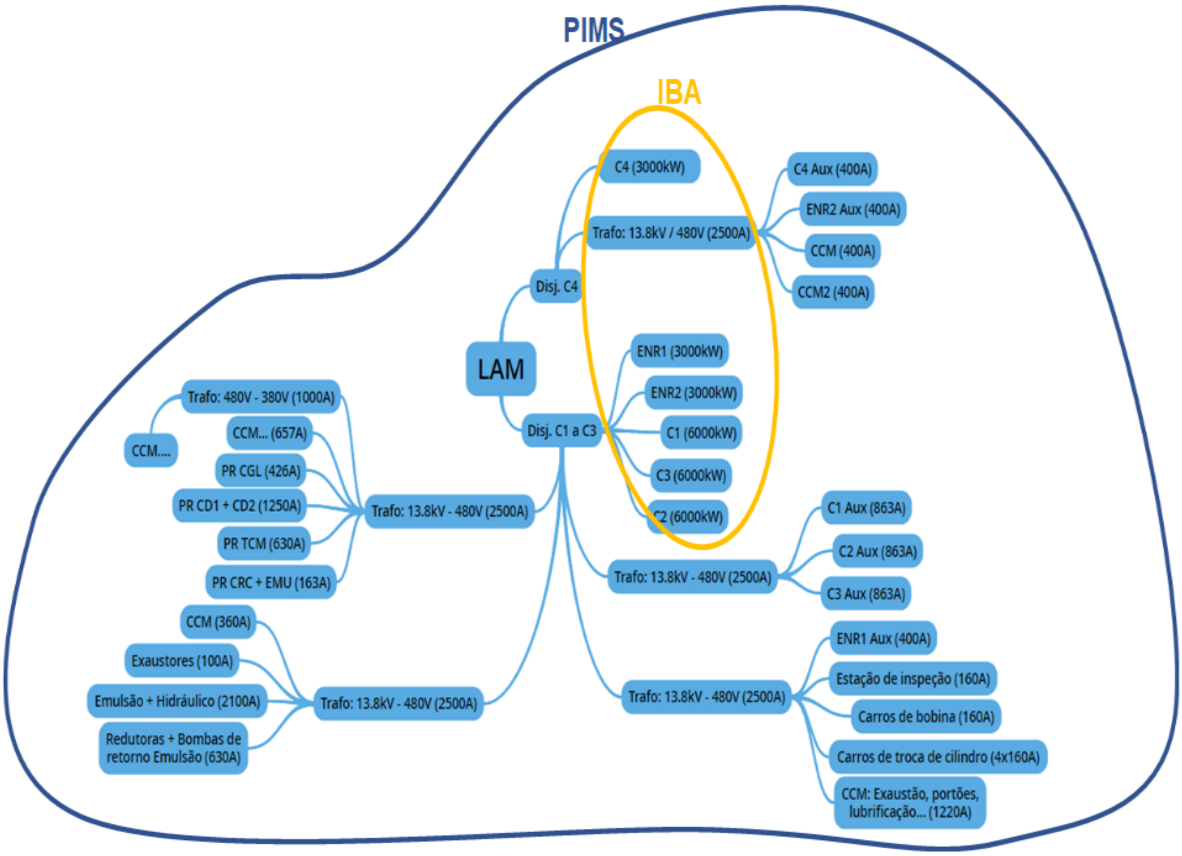
The following sections detail the gathering and processing of data from each of the previously mentioned systems.

Table 4 – Selected input and output variables for energy prediction.

Item	**Description**	**Domain**	**Range**	**Unit**	**Usage**
EntrTh	Entry coil thickness	Real	$[1.8 \sim 4.8]$	mm	Input
ExitTh	Exit coil thickness	Real	$[0.35 \sim 2.70]$	mm	Input
Red	Total coil reduction	Real	$[35.0 \sim 85.0]$	%	Input
Wd	Coil width	Real	$[750 \sim 1878]$	mm	Input
Wg	Coil weight	Real	$[0.6 \sim 37.5]$	t	Input
SH	Strip hardness	Real	$[64.0 \sim 150.0]$	kgf/mm^2	Input
CRT	Coil running time	Real	$[0.002 \sim 4.600]$	h	Input
AvgS	Average exit line speed	Real	$[85.0 \sim 910.0]$	mpm	Input
Energy	**Consumed Electric Energy**	**Real**	$[\mathbf{0 \sim 4000}]$	**kWh**	**Output**

Source: Elaborated by the author (2021).

Figure 5 – Representation of the TCM energy grid indicating the different electric circuits with current level and database coverage.



Source: Elaborated by the author (2021).

3.1 IBA DATABASE

The IBA database has been selected as the first approach for this work mainly because of its reliability and the excellent data quality, with low noise and few missing data. These two features of the dataset would allow a quick assessment of the performance of the system in the modeling of consumed energy because a poor performance with this high-quality data could indicate that ML is not a good approach to solve the proposed problem.

Another important aspect of the IBA system is the capability of extracting the data automatically on a coil-to-coil. Even though the IBA database is intrinsically stored in a time basis format, the system is capable of summarizing the measurements in statistical figures for each product such as average, standard deviation, maximum and minimum on a coil-to-coil basis simplifying the correlation with the Level 2 database and speeding up the regressor training and validation procedures.

However, as a first step, energy would have to be calculated since IBA does not store this information. In a previous work the electric motor power has been predicted by ANN with accurate results (OLIVEIRA et al., 2020) and its integration over time yields consumed energy. This method has been defined as the appropriate approach and an example for one product is exposed in Figure 6.

With calculated energy included in the database, the preprocessing of the data is simplified by the IBA interface, which provides a tool to automatize it. As the information is acquired directly from the PLC memory, communication issues are virtually nonexistent. IBA also performs an automatic filtering of the data, removing uninteresting conditions such as line stops or strip break events which reduces the preprocessing to simple condition checking for strip dimensions within production range to remove occasional incorrect sensor readings or malfunction.

3.2 PIMS DATABASE

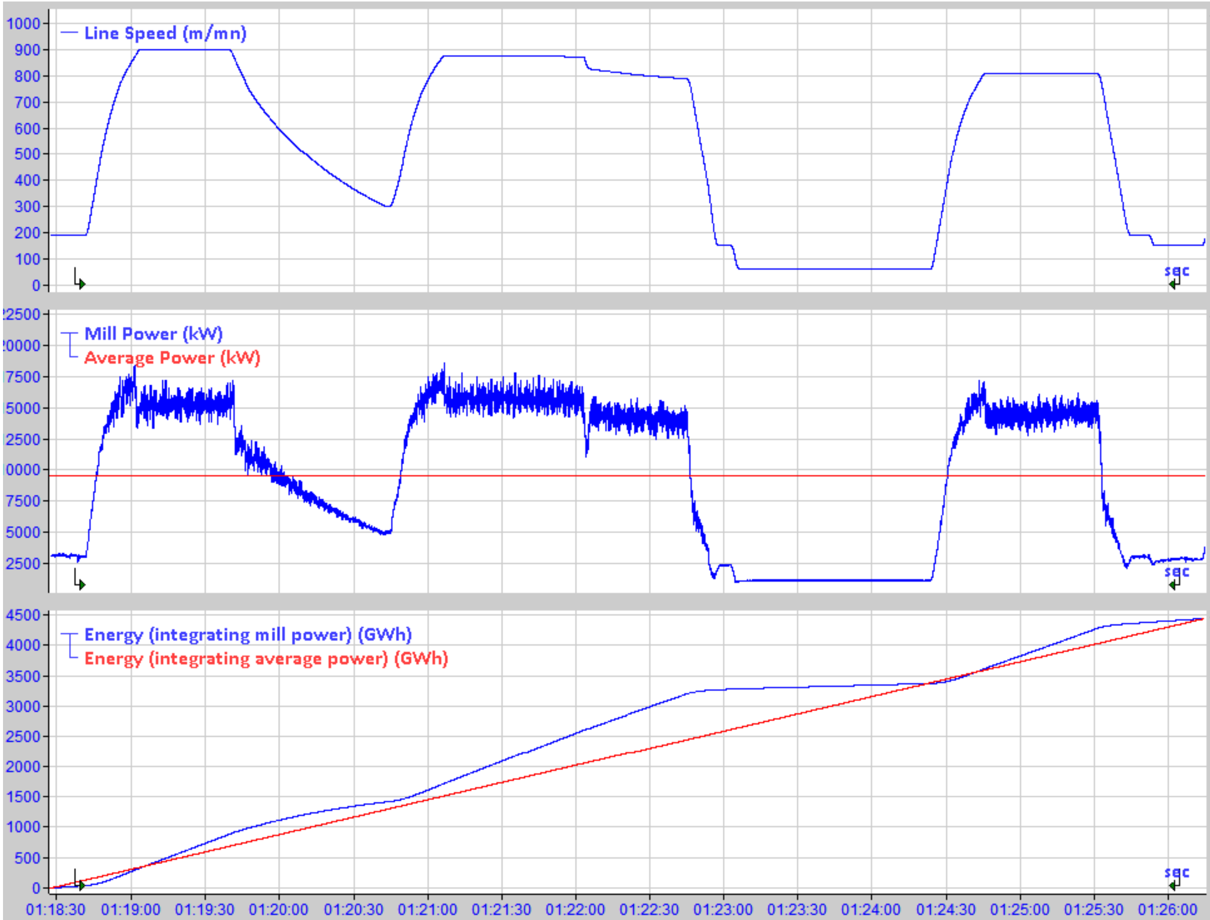
PIMS was designed to integrate the different industrial equipment, data acquisition systems and data storages in a single system to simplify data gathering. This advantage allowed to clearly correlate the energy measurements from CCK to the product being processed by the TCM, an impossible task without PIMS. However, due to its versatile purpose and many interfaces and communication protocols with different equipment, the system is not as reliable as IBA and frequently there are inconsistencies in the data stored by the PIMS such as empty rows, frozen values for several hours and days, unrealistic readings and so on.

In addition to these inconsistencies, unlike IBA, the dataset provided by PIMS is not summarized by product but on a time basis approach, as shown in Figure 7. This time-based dataset is unsuitable for handling this specific problem because the entire production management in the steel plant is made on a coil-to-coil basis. Additionally, the L2 and IBA databases are also only available on a coil-to-coil basis. Thus, the PIMS database required the development of an algorithm to group the data on a coil-to-coil basis for model training. This data grouping algorithm had to be extensively verified and tested to minimize the inconsistencies of the PIMS raw dataset.

Figure 8 shows the histogram of energy consumption in a coil-to-coil basis indicating a significant concentration of the data on low energy values. Figure 9 subdivides the previous chart in two only for visual purposes using a 4,000kWh threshold. Coils with lower or equal energy consumption than this threshold represents 99.6% of the database. An evaluation of the remaining 0.6% of the coils revealed that those were related to unusual events such as maintenance stops, machine setup, strip breaks etc. Even though these are not regular rolling conditions, the energy consumed in these coils could reach approximately 100,000kWh, significantly impacting monthly evaluations.

The algorithm developed to group the data on a coil-to-coil basis was capable to significantly improve the data quality but resulted in the removal of several lines from the original database: from 122,000 observations to 93,000 observations. Even though the data is enough data for model training and validation, the reduction of approximately 25% in the dataset has a significant impact when a full month of production is to be evaluated. Figure 10 compares the

Figure 6 – Example of power integration to calculate energy consumption.



Source: Elaborated by the author (2021).

total production of the TCM per month between Level 2 and PIMS dataset after the preprocessing. In some months, the difference has reached almost 48% which significantly impairs any direct comparison of the energy consumption: PIMS always indicates a much lower consumption than it was actually consumed by the TCM due to the removal of many lines of the database in the preprocessing phase.

After the selection of the sources and the extraction of the information from the databases, a preliminary set of experiments were conducted in order to determine the best data gathering approach for further model development, detailed in the next chapter. Before model training and validation, grid search and cross validation were used to select the hyperparameters of the algorithms followed by extensive model testing, also described in the next chapter.

Figure 7 – PIMS database sample.

Energia_RED_01.txt

	Data	Hora	Valor
2	01/04/2019	00:00:00	94051000.00000
3	01/04/2019	00:00:10	94051040.00000
4	01/04/2019	00:00:20	94051064.00000
5	01/04/2019	00:00:30	94051120.00000
6	01/04/2019	00:00:40	94051152.00000

Espessura_RED_01.txt

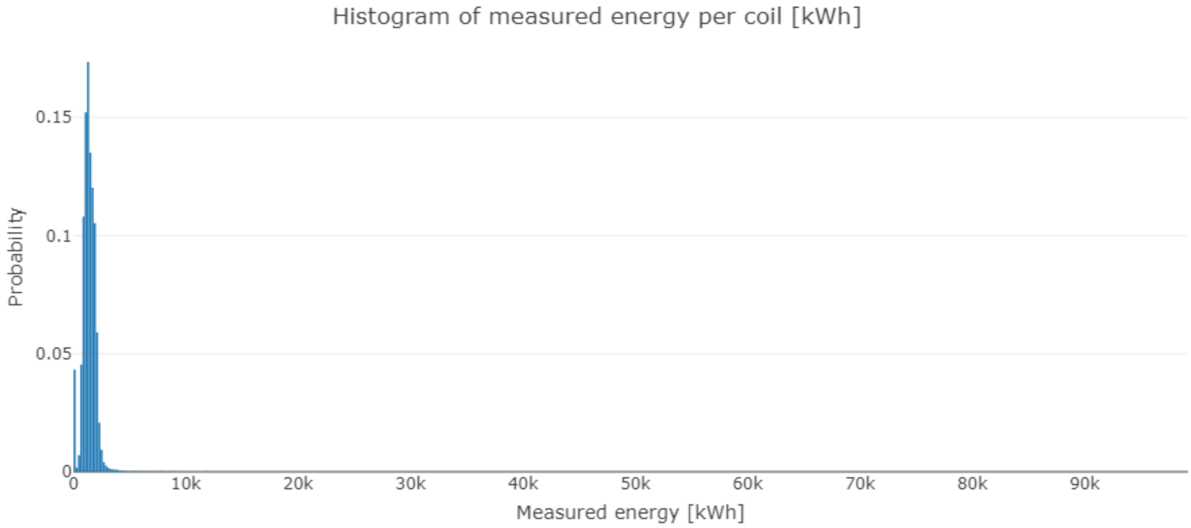
	Data	Hora	Valor
2	01/04/2019	00:00:00	0.38000000
3	01/04/2019	00:00:10	0.38000000
4	01/04/2019	00:00:20	0.38000000
5	01/04/2019	00:00:30	0.38000000
6	01/04/2019	00:00:40	0.38000000

Bobina_RED_01.txt

	Data	Hora	Valor
2	01/04/2019	00:00:00	H212083
3	01/04/2019	00:00:10	H212083
4	01/04/2019	00:00:20	H212083
5	01/04/2019	00:00:30	H212083
6	01/04/2019	00:00:40	H212083

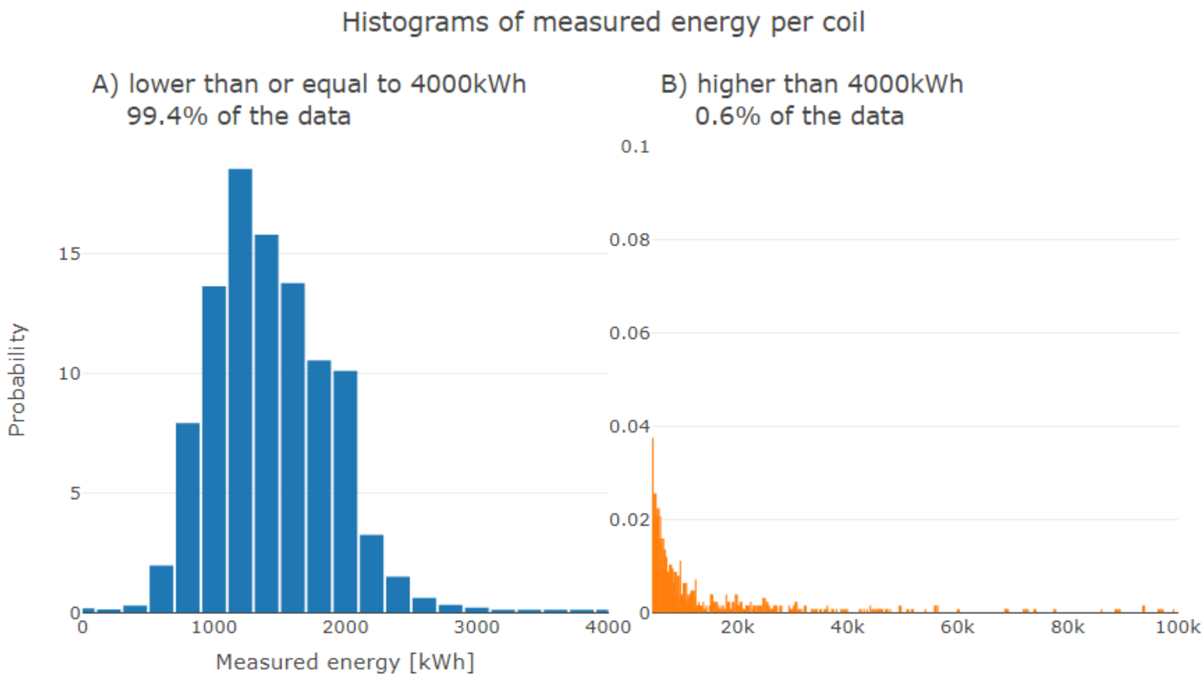
Source: Elaborated by the author (2021).

Figure 8 – Histogram of consumed energy measured store by PIMS on a coil-to-coil basis.



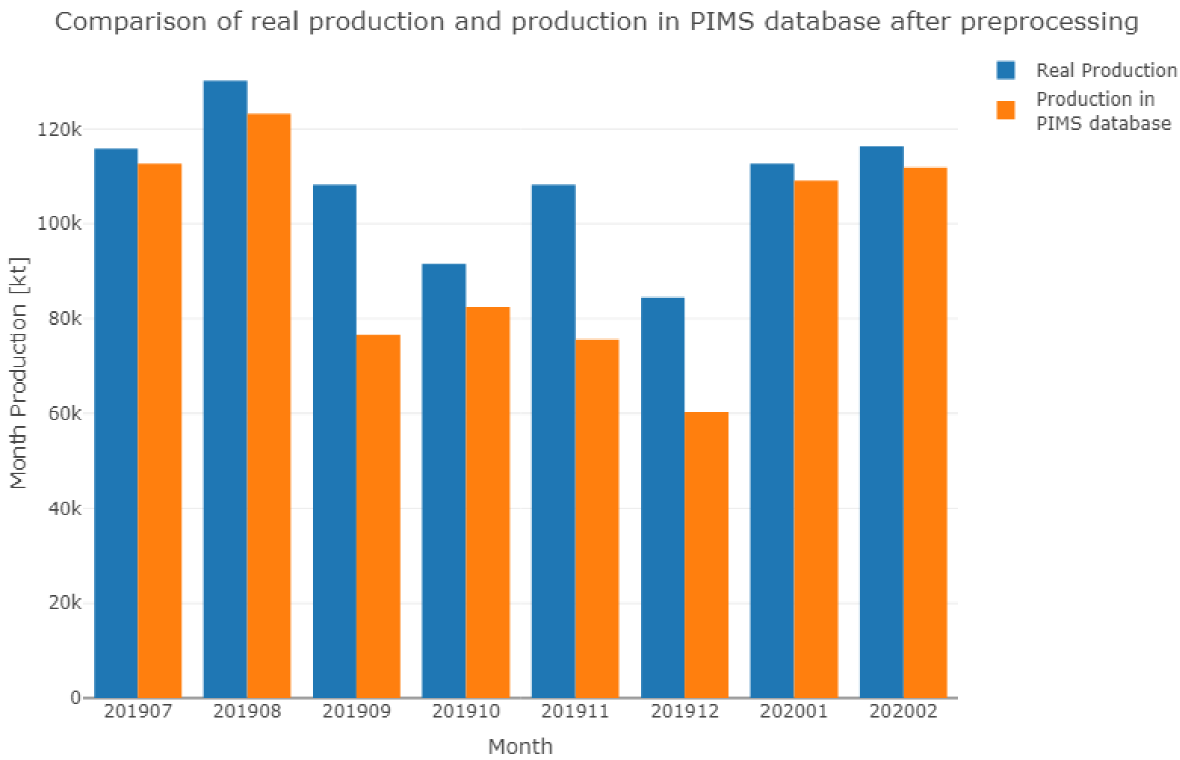
Source: Elaborated by the author (2021).

Figure 9 – Histogram of consumed energy on a coil-to-coil basis for A) lower or equat to 4000kWh and B) higher than 4000kWh.



Source: Elaborated by the author (2021).

Figure 10 – Comparison of real TCM production and PIMS database.



Source: Elaborated by the author (2021).

4 ENERGY MODEL DEVELOPMENT

The cold rolling of steel strips is a very complex, non-linear process (ROBERTS, 1978) and many scholars have dedicated a significant amount of time in the development of different rolling theories and models throughout the twentieth century (LENARD, 2013). The modeling complexity emerges from the elastic deformation of the work rolls in a contact interaction with elasto-plastic deformation of the steel strip under non-homogeneous friction regimes (ALEXANDER, 1972). The mathematical modeling of energy consumption is, to this day, still a challenge due to the dependency on calculated variables and the accumulation of modeling errors, a conclusion that can be made from the work of Freshwater (1996).

Computational methods have been widely used in solving such complex problems in engineering and science for decades (HU et al., 2019). However, they rely on extensive databases which allow the algorithm to learn the patterns hidden in the data. The flow chart of the system proposed in this work designed to predict the energy consumption of the TCM is indicated in Figure 11. This figure summarizes the storage and acquisition procedure previously discussed in Chapter 3, where process data coming from L2 and IBA are merged with energy data coming from PIMS to form a new database. This database is then processed for model training procedure followed by tuning of model hyperparameters, training and validation until desired accuracy is met. The model resulting from this procedure is saved for energy predictions. Two distinct datasets can be applied to the trained regressor: production data (composed of a list of previously processed coils, or past data) and schedule data (which is a list of coils expected to be produced in a given moment based on sales forecast and assumptions, or future data). These conditions are further detailed in Chapter 5.

This system is capable of reading and pre-processing the measurements database, train, store and validate the regression model or make predictions based on past (production) and future (schedule) data.

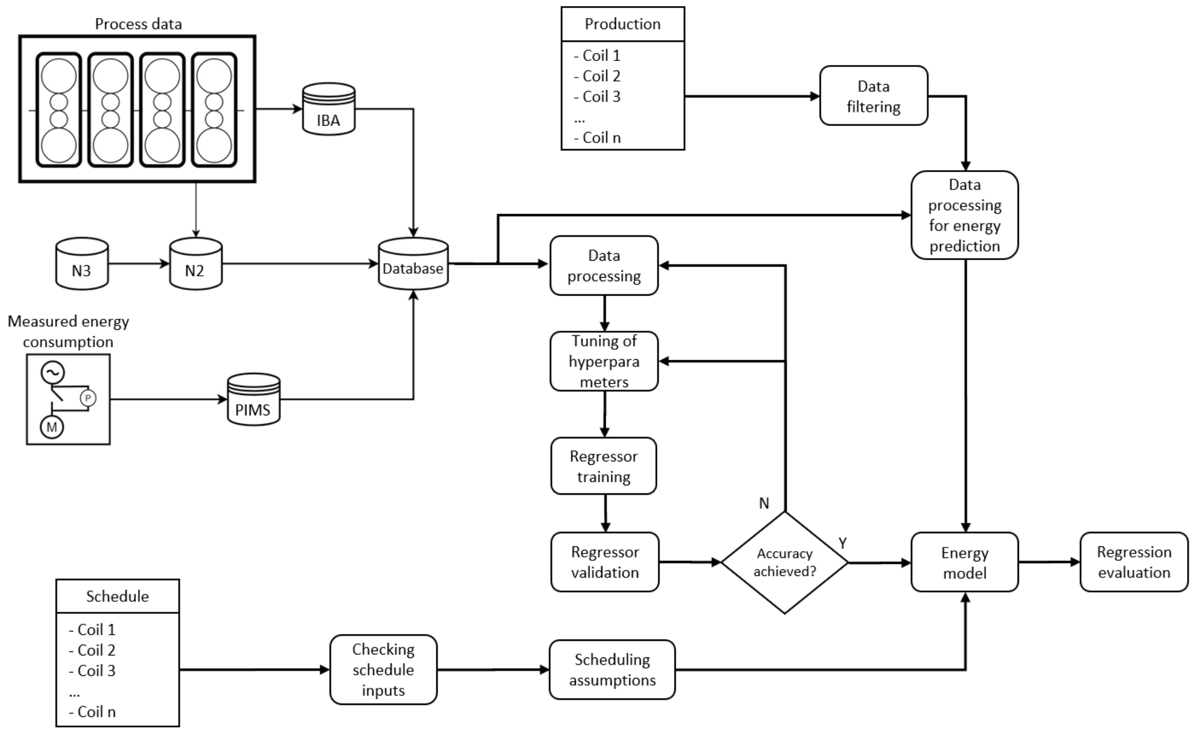
The accuracy of the learning process is highly dependent on the quality of the data (SMITH, 1993). Thus, several experiments were conducted to define the adequate database among those available for this work. This procedure is detailed in Section 4.1. The definition of model architecture is described in Section 4.2. Finally, a comparison between different models is made in Chapter 5 to ensure the best performing regressor is selected as energy predictor.

4.1 DATABASE SELECTION

Adequate databases are very important for most ML algorithms and the selection of the dataset requires special attention (SMITH, 1993). In industrial applications it is usual to encounter noisy signals, sensors indicating unexpected readings and missing data as a result to maintenance problems, harsh environment, etc.

Because of this, a preliminary set of experiments were conducted to support decision-

Figure 11 – Flow chart of energy prediction system.



Source: Elaborated by the author (2021).

making on the most adequate data storage system for model training.

The objective of these experiments was to quickly assess which of the two available datasets should be further explored for energy modeling considering their previously mentioned advantages and disadvantages: i) IBA with its robustness and high data quality but indirect energy data (calculated from motor power integrated over time) or ii) PIMS which reads real energy measurement from the entire mill grid but suffering with data integrity issues.

In order to compare the behavior of these two datasets, an ANN was selected due to its recognized capacity to solve complex regression problems (BROWNLEE, 2011) and applicability to energy predictions González-Briones et al. (2019). The same ANN architecture was employed in both cases: an MLP with eight input nodes representing the input variables, three hidden layers with 30 nodes each, and one output node representing the output variable. The hyperbolic tangent is used as activation function. The model architecture was empirically defined. The Adam optimizer was employed to train the model with a learning rate of 0.001.

The available data is split into training and test sets according to the date of production in order to emulate a real application of the model where the initial six months of the data has been applied in the training procedures while the remaining twenty months were reserved for testing of the trained ANN.

The data splitting have been defined based on the fact that the initial six months yielded 52000 products, which is sufficient for adequate model training. However, the model performance assessment and comparisons are not on a coil-to-coil basis (as shown in the next sections) but on a monthly fashion. For this reason, additional months were considered for the test dataset to allow longer-term evaluations.

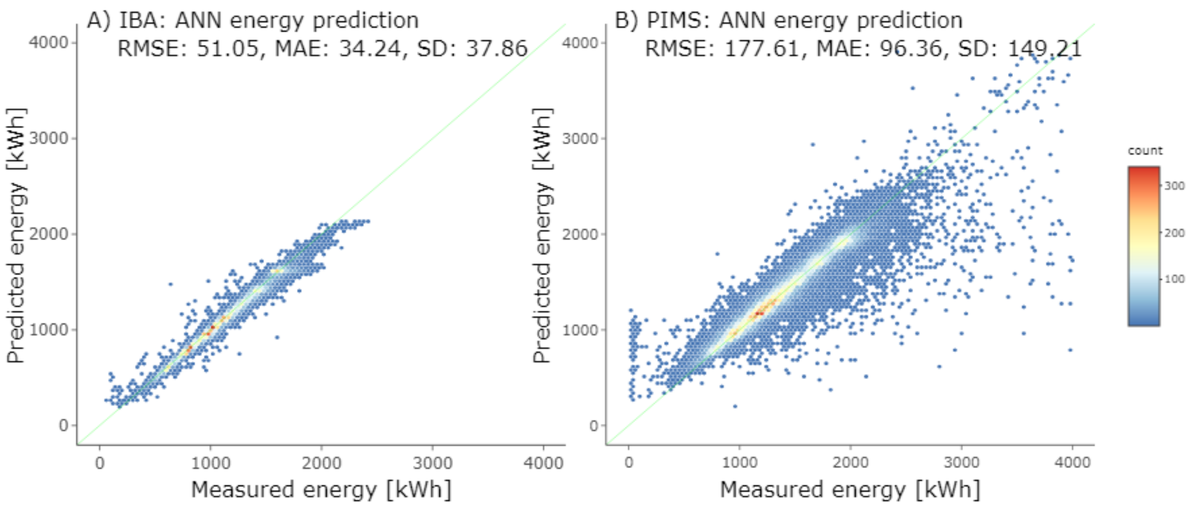
Regarding the training data, 20% is reserved and used for validation of the model. During the training procedures, the Mean Absolute Error (MAE) and the Mean Squared Error (MSE) are being constantly calculated for both the training and validation datasets (ZHANG; PATUWO; HU, 1998). The MSE of the validation was defined as the stop criteria to avoid overfitting.

The test data is then presented to the ANN for prediction of the energy consumption. The scatter plot of measured and predicted energy consumption of the trained ANN with IBA and PIMS datasets when presented with unseen (test) data can be seen in Figure 12. This clearly shows that IBA data significantly outperforms the PIMS extraction. The MAE and the root mean squared error (RMSE) are at least three times lower in IBA data. This performance difference was expected and can be attributed not only to the lower quality of PIMS database but also to the fact that it comprises energy consumers about which the inputs provide no information such as overhead cranes and air conditioning systems.

The regression error characteristic (REC) curve presented in Figure 13, a plot of the cumulative distribution function of the prediction relative error (HERNáNDEZ-ORALLO, 2013) which allows the comparison of different models accuracies over the entire dataset, also indicates a much better performance of the IBA ANN model. If the average consumption of 1407kWh is considered and that the PIMS RMSE reached 177.61kWh (12.62%) and an MAE of 96.36kWh (6.84%kWh) then it can be argued that both models have reached very good preliminary results.

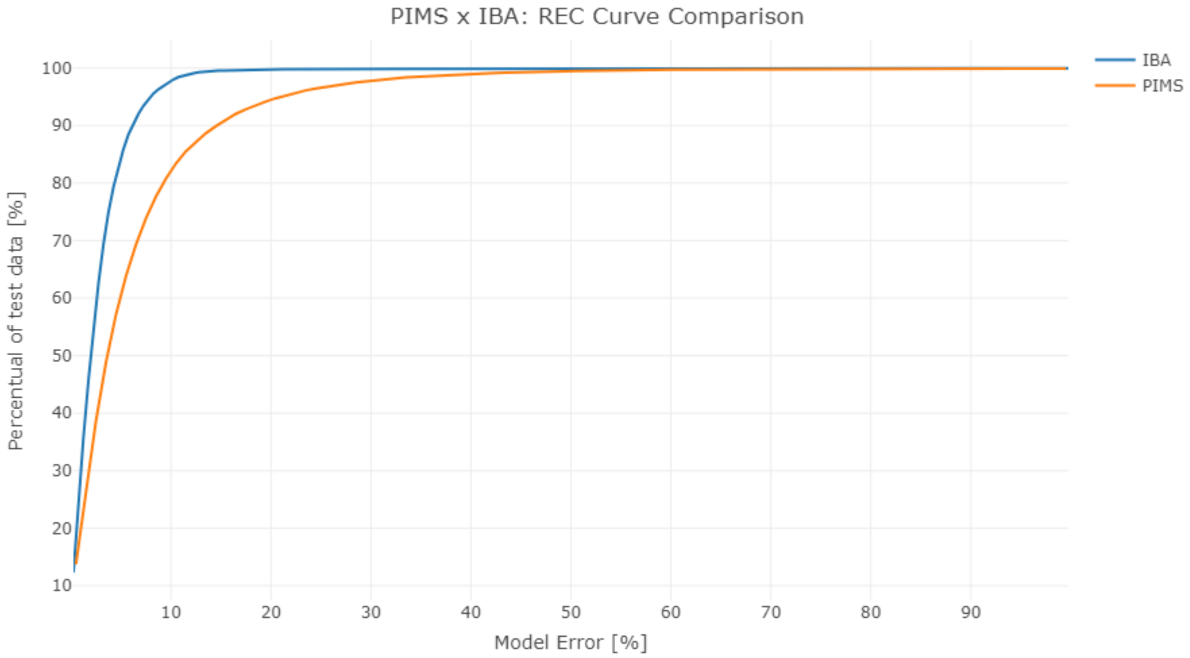
However, the objective of this work, as already stated, is not to assess the energy on a coil-to-coil basis, but to predict an entire month of energy consumption of the TCM grid. When the outputs of the ANN for each coil are summed to compute the month results and compared with the actual information provided by the electric energy bill, unacceptable discrepancies were

Figure 12 – Scatter plot of measured and ANN prediction of energy consumption for IBA (A) and PIMS (B) datasets.



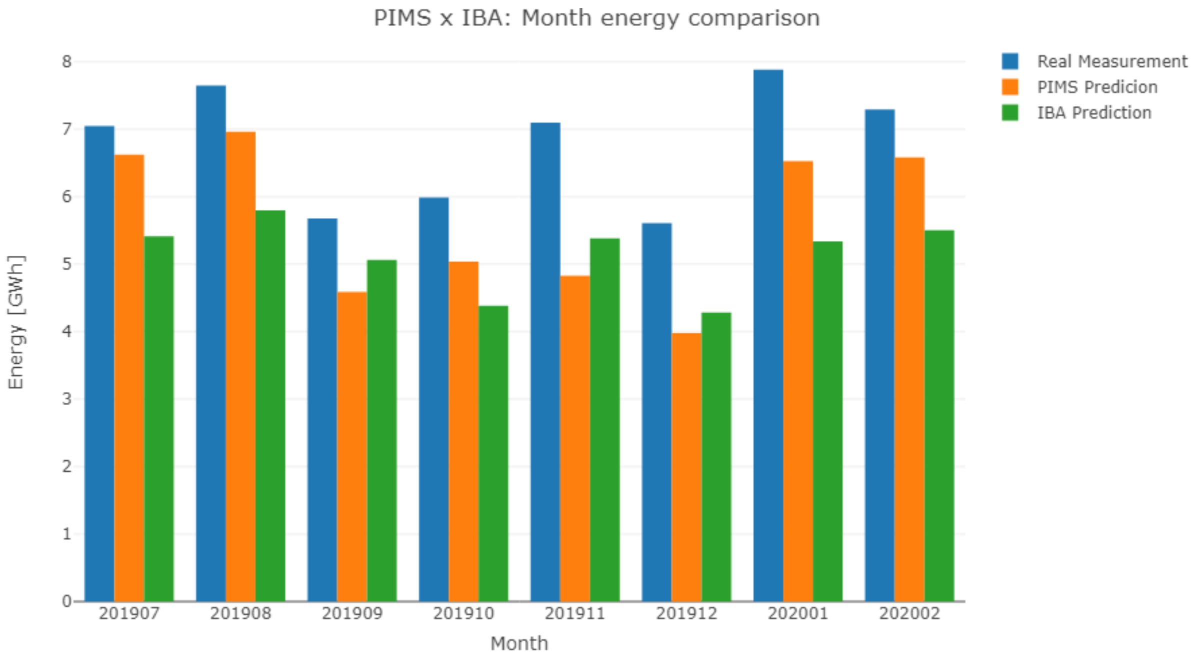
Source: Elaborated by the author (2021).

Figure 13 – REC curve for ANN model with IBA (blue) and PIMS (orange) datasets.



Source: Elaborated by the author (2021).

Figure 14 – Monthly energy consumption.

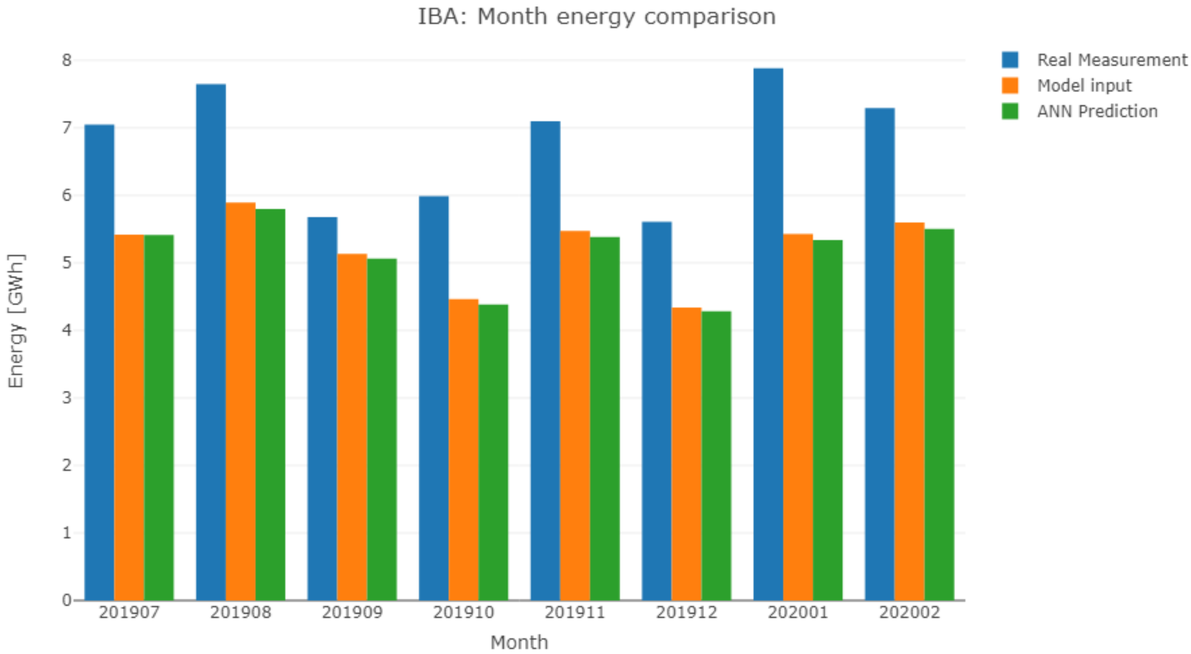


Source: Elaborated by the author (2021).

observed as shown in Figure 14.

These discrepancies arise not from modeling error, which is relatively low. For each database the root cause is different. In the case of IBA, this error is the result of energy consumption from equipment not comprised by the acquisition architecture mentioned in the previous section and exemplified in Figure 5. It is also made evident by the chart shown in Figure 15

Figure 15 – IBA: Monthly energy consumption.



Source: Elaborated by the author (2021).

where the model input and ANN prediction are almost identical, confirming the low modeling error and high difference from model input and actual energy measurement. This difference has proven that the energy calculation approach defined with the IBA database could not be exploited any further.

Similar conclusion could be drawn from the evaluation of Figure 16 since the behavior of PIMS database on a monthly basis is very similar to that of IBA but for different reason. In this case, the error comes from reliability issues on the data collecting process rather than measuring limitation.

Based on the experimental results and discussions exposed, a different approach was proposed to combine the IBA and PIMS databases into one large set to emphasize their strengths and reduce their weaknesses. In this new database, PIMS would provide only the energy data while the process variables would be extracted from IBA. With this approach, the ANN could be trained only on those products where PIMS data is consistent. For the testing, IBA process data is used as input to the trained ANN, which outputs the energy prediction. Since the model performance is made on a monthly basis, the lacking of energy data from PIMS becomes irrelevant.

In order to accomplish this, a robust algorithm had to be developed and exhaustively tested for this database merging procedure to ensure data quality since the two databases are inherently different as exposed in the previous chapter: IBA data is collected in a coil-to-coil fashion while PIMS is provided in a time basis format. For the merging of the two datasets, specific markers in variables common to both datasets were used to assess the time difference between them. This procedure allowed to identify the starting and ending point of each coil in the PIMS database corresponding to the same product in the IBA dataset.

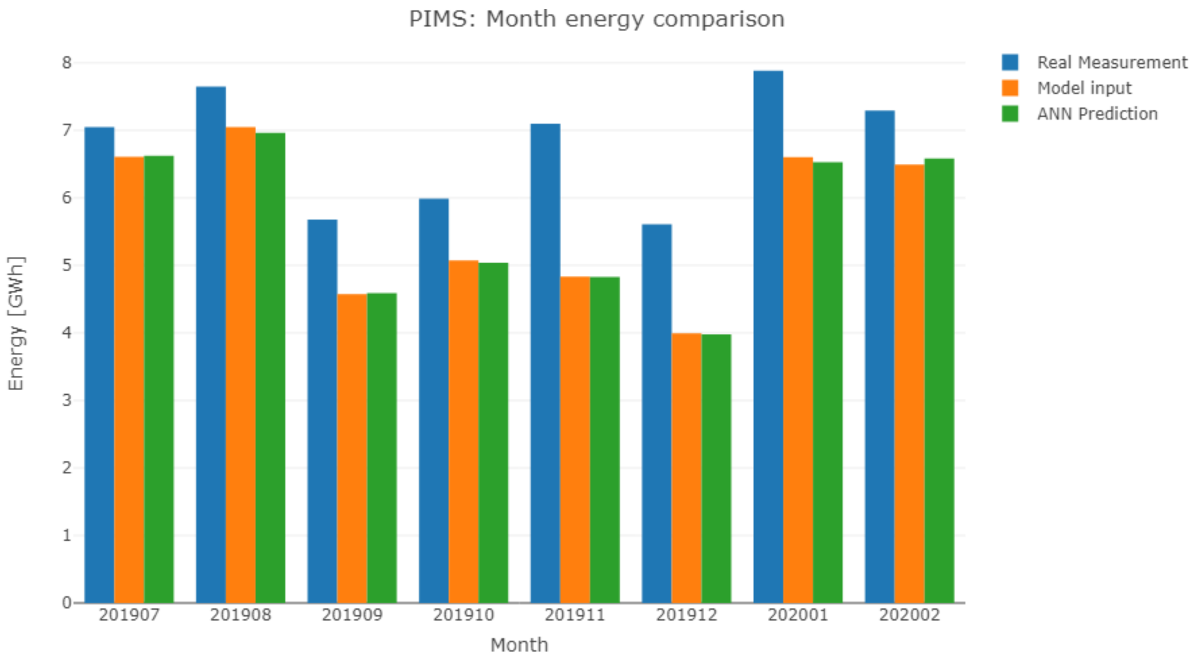
As energy readings from PIMS is being used for the ANN training procedures, the ANN performance on this merged approached is compared with data extracted only from PIMS in the scatter plot of Figure 17 which shows a slightly higher RMSE but lower MAE for the combined dataset in comparison with PIMS. A visual inspection of the plot indicates that the ANN trained on pure PIMS data has better performed, however the REC curve shown in Figure 18 shows otherwise: the ANN trained on the merged data has yielded a larger amount of data with lower error. This difference is evident at the 10% error point (horizontal axis). There is a notable advantage to the IBA+PIMS dataset since 88.5% of the predictions presented relative error lower or equal to 10% while only 82.1% of the predictions reached the same error level when the ANN was trained on pure PIMS database.

The comparison of the total energy consumption for a month of data is shown in the top chart of Figure 19, which also indicates the difference in modeling error on a monthly basis. The merged approach presented a significant improvement in the capacity of the system to predict energy consumption in this condition as well. The bottom chart of Figure 19 presents the model error relative to the real measurement and the mean relative error (MRE) reduced from 17.2% on the PIMS approach to 7.2% with the merged database.

As it can be seen from these preliminary results, the proven reliability of the IBA acquisition architecture combined with the actual energy measurement from the PIMS database have added significant value to the final dataset. For all these reasons, the merged dataset approach have been defined as the best method for continuing the model development.

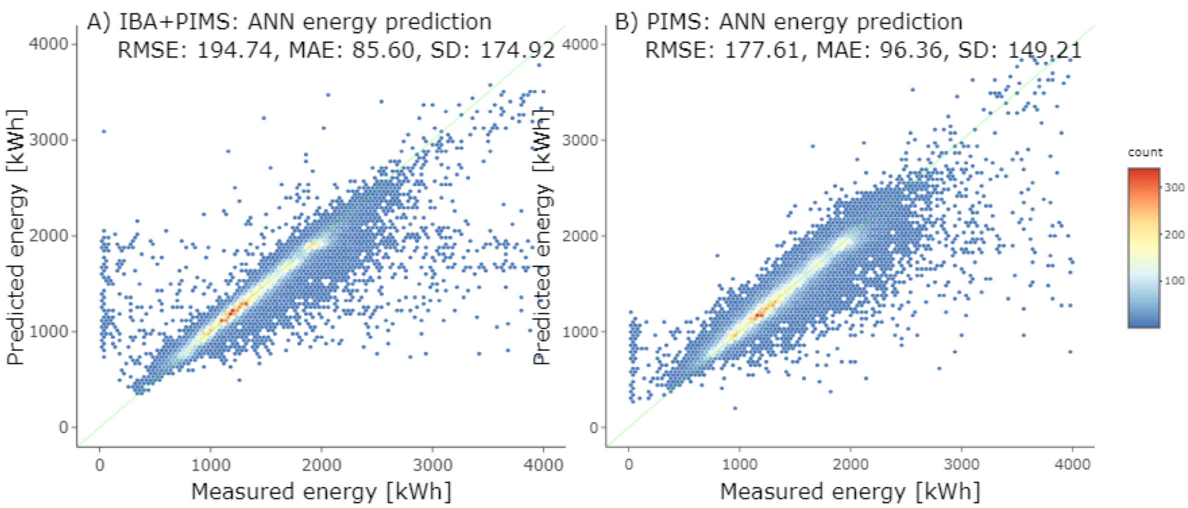
In order to determine model hyperparameters, a grid search have been carried out and the details are exposed in the following section.

Figure 16 – PIMS: Monthly energy consumption.



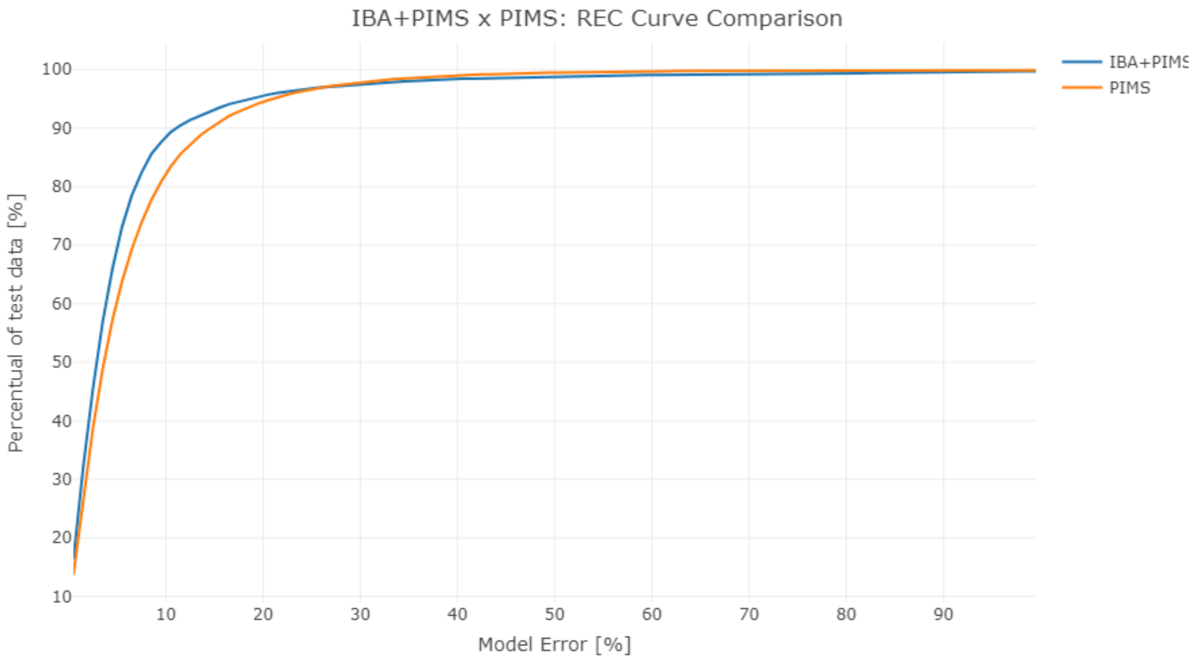
Source: Elaborated by the author (2021).

Figure 17 – Scatter plot of measured and ANN prediction of energy consumption for IBA+PIMS (A) and PIMS (B) datasets.



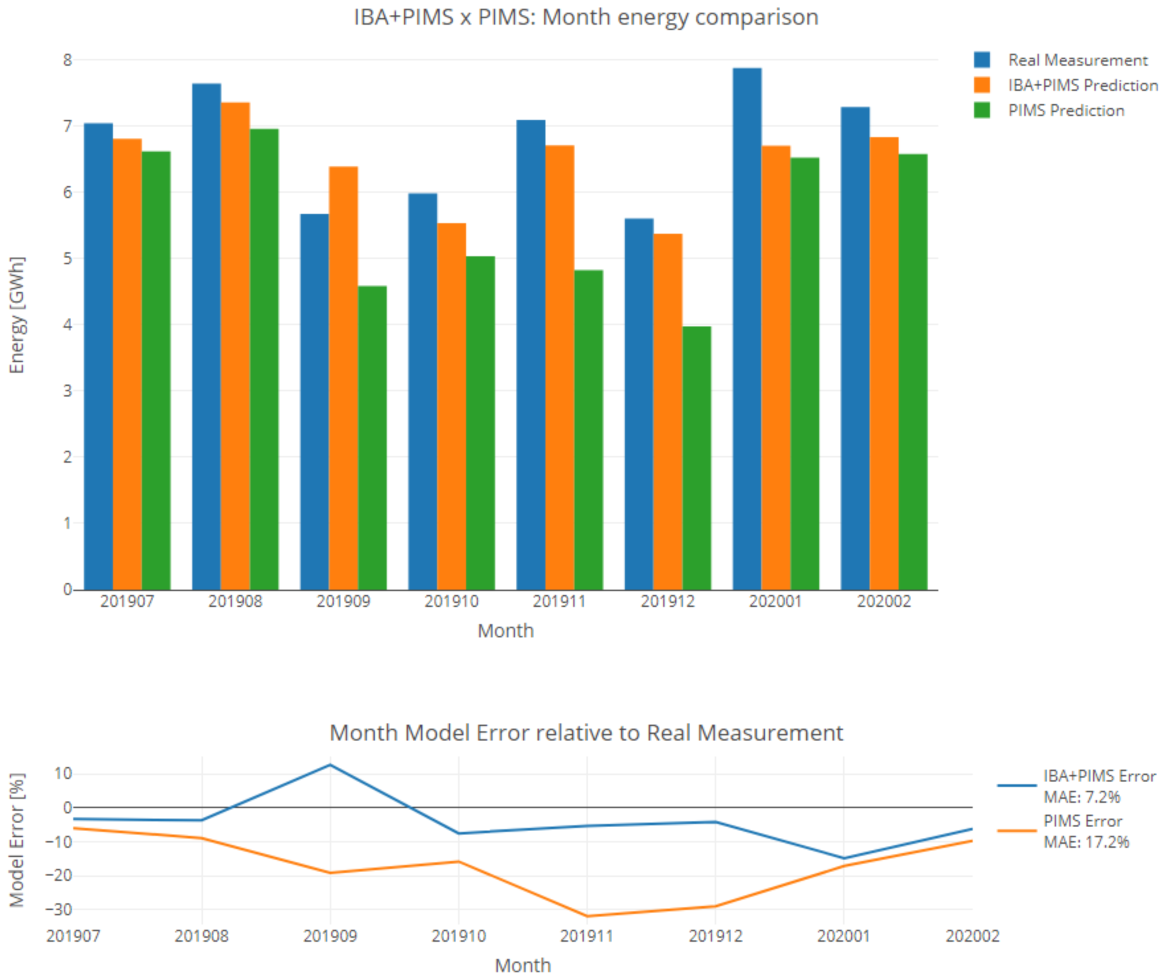
Source: Elaborated by the author (2021).

Figure 18 – REC curve for IBA+PIMS (blue) and PIMS (orange) datasets.



Source: Elaborated by the author (2021).

Figure 19 – Monthly energy consumption and model relative error comparison.



Source: Elaborated by the author (2021).

4.2 GRID SEARCH

The results presented in the previous section were reached after empirical definition of the ANN hyperparameters and architecture. Even though many experiments were conducted to reach the final configuration, no specific method was established in the process.

In order to clearly define this methodology and to fully explore the potential of the regressors selected for this work, a factorial combination grid search approach have been defined to identify the hyperparameters and improve the preliminary results.

The parameters tested for the ANN resulted in a total of 162 experiments and were:

a) Architecture (nodes in hidden layers): [30, 30, 30]; [30, 30, 30, 30]; [30]; [30, 30]; [8]; [64]; [512]; [128, 64]; [1024].
b) Activation Function: tanh; sigmoid; relu.
c) Optimizer: adam; sgd; rmsprop.
d) Loss Function: MSE; MAE.

For comparison reasons, a Random Forest (RF) and an Extreme Gradient Boosting (XGB) algorithms have also been trained and evaluated on the prediction of energy consumption for TCM, since these are very powerful regressor on such applications (BELGIU; DRAGUT, 2016). In order to find the best parameter settings for the regression on this particular problem, allowing adequate comparison between well-adjusted models, the grid search procedure was considered for the RF and XGB.

For the RF, 72 experiments were carried out:

a) Number of trees: 5; 10; 50; 100; 150; 200.
b) Minimum examples per node: 5; 10; 20.
c) Maximum tree depth: 16; 8; 4; 32.

The following parameters were tested for the XGB resulting in a total of 486 experiments:

a) Number of estimators (or trees): 5; 10; 50; 100; 150; 200; 400; 500.
b) Maximum depth: 3; 5; 6; 10; 15.
c) Learning rate: 0.01; 0.05; 0.10; 0.20.
d) Columns sample per tree: 0.1; 0.3; 0.5.
e) Objective function: linear error; squared error.

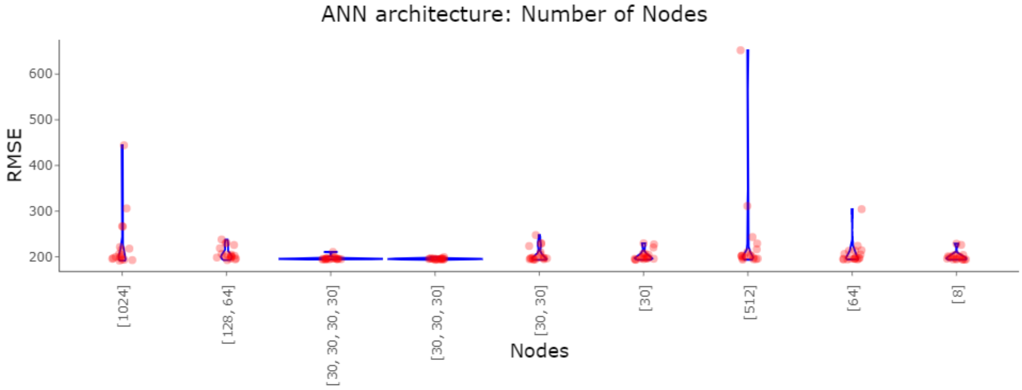
The grid search results for the ANN, RF and XGB are detailed the following sections.

4.2.1 ANN grid search

The evaluation of the RMSE of all the ANN experiments split by architecture shown in Figure 20 clearly indicates that those trials with 3 and 4 hidden layers with 30 neurons in each layer ([30, 30, 30] and [30, 30, 30, 30] respectively) presented the lowest scattering of the RMSE, regardless of other training parameters, while the experiment with one layer of 512 nodes ([512]) presented the highest scattering of RMSE.

Including the loss function and the optimizer algorithm in the splitting criteria of the

Figure 20 – Root Mean Square Error of the different ANN experiments split by architecture.



Source: Elaborated by the author (2021).

chart in Figure 21 does not provide any evidence that the loss function have a definitive impact on the variation of the RMSE of the experiments. However, it is noticeable that the optimizer sgd presents RMSE significantly more scattered than the others.

The evaluation of Figure 22 showing the influence of both optimizer algorithm and activation function clearly indicates that the interaction of the sgd algorithm with the sigmoid activation function resulted in the highest variations of the RMSE.

Removing the experiments with these parameters from the initial node evaluation indicates much less dependency of the RMSE with ANN architecture variation seen in Figure 23 as the vertical axis scale have been dramatically reduced. Regardless, it is still evident that the architectures with 30 nodes and three or more layers are less sensitive to training parameters. Because of these experiments and to reduce model complexity, it was decided to run additional trials with the ANN with three hidden layers and 30 neurons in each, with adam optimizer, tanh as activation function and MSE of loss function.

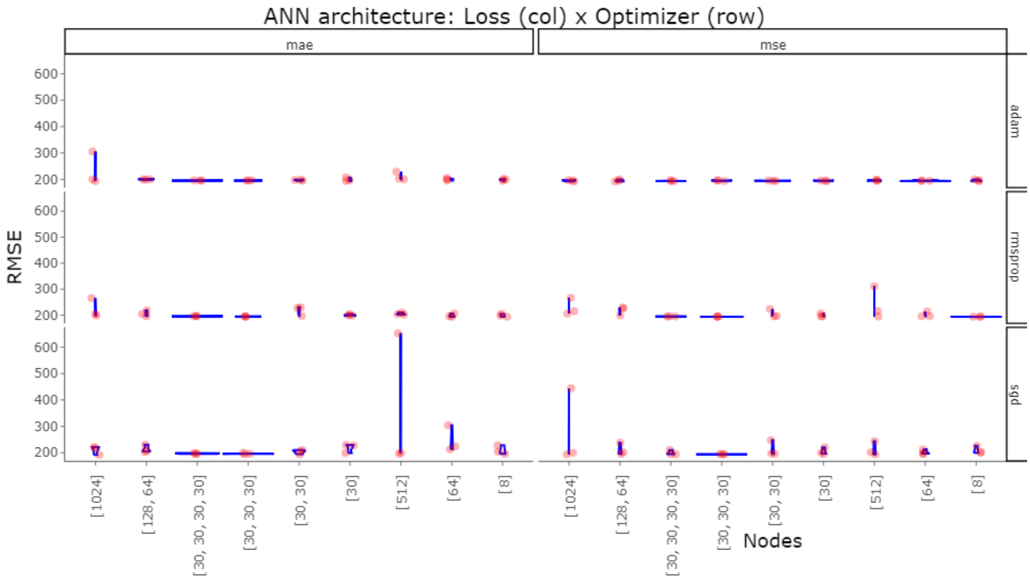
4.2.2 RF grid search

Splitting the RMSE results of the RF experiments by the number of decision trees in Figure 24 indicates that it can influence the error when less than 50 trees are selected for this specific problem. Another important observation is the significant scattering observed in all groups of data in the chart.

Adding maximum tree depth to the chart in Figure 25 shows that the experiments with depth of 4 presented a much worse result than the others, as expected. It can also be noted that the depth of 8 presented RMSE slightly higher in comparison to the remaining two other set of experiments.

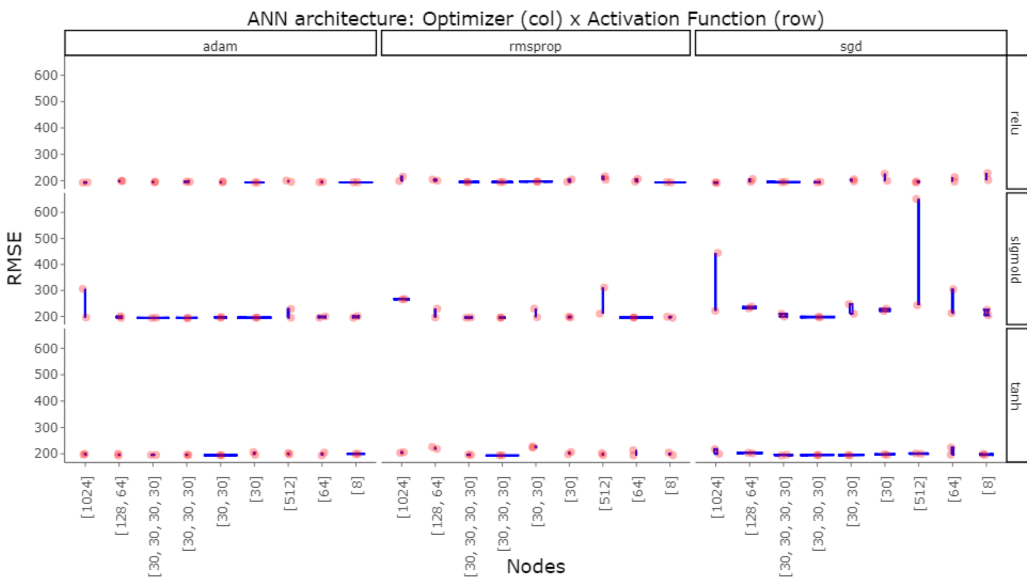
Excluding the experiments with depth of 4 from the evaluation and splitting the number of tree chart also by minimum examples in Figure 26 provides additional information for defining the final configuration of the RF. The evaluation of the plot indicates that increasing the minimum examples required to open a new node in the tree increases the RMSE. This behavior is expected

Figure 21 – Root Mean Square Error of the different ANN experiments split by architecture, loss function (columns) and optimizer algorithm (rows).



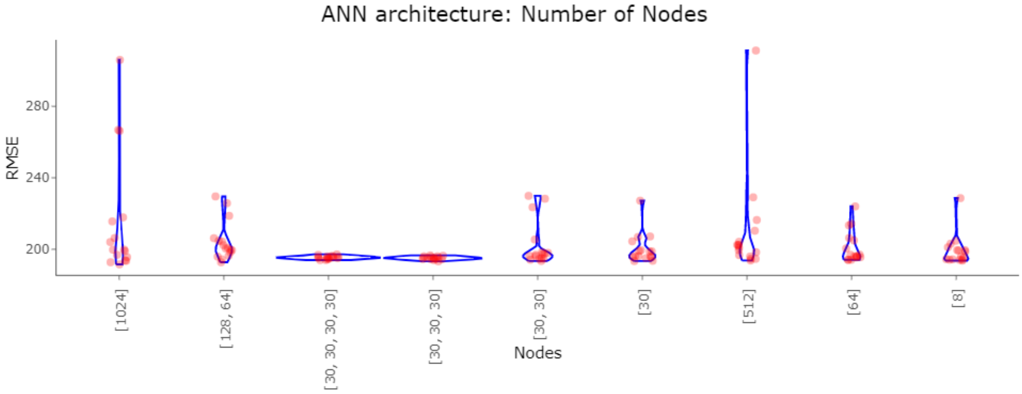
Source: Elaborated by the author (2021).

Figure 22 – Root Mean Square Error of the different ANN experiments split by architecture, optimizer algorithm (columns) and activation function (rows).



Source: Elaborated by the author (2021).

Figure 23 – Root Mean Square Error of the different ANN experiments split by architecture after removing sgd optimizer and sigmoid activation function.



Source: Elaborated by the author (2021).

since it can be argued that a tree with fewer nodes would have more difficulties reproducing complex data patterns. It can also be noted from Figure 26 that the 100 trees experiments presented slightly better RMSE than 50 trees in every condition.

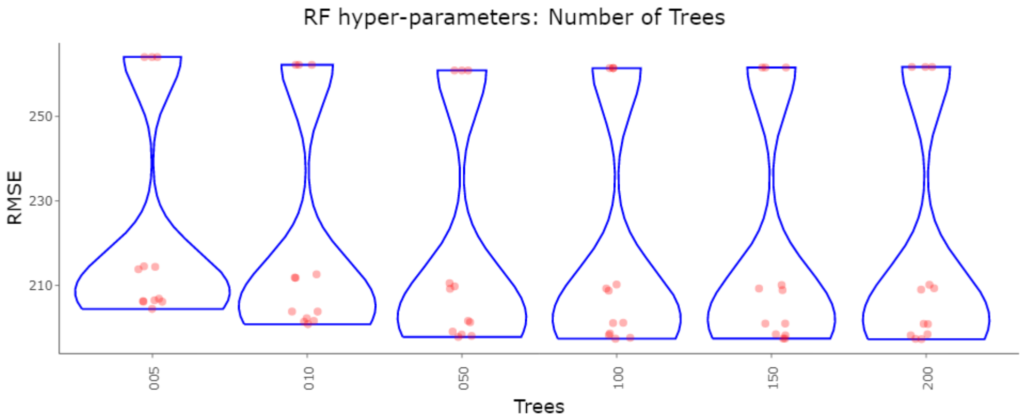
Based on these experiments results, the RF configuration selected was 100 trees, 5 minimum examples and 32 for maximum depth.

4.2.3 XGB grid search

The RMSE of the different experiments with the XGB split by number of estimators (or trees) is shown in Figure 27. As this figure indicates, the RMSE increases significantly when less than 50 estimators are used for the prediction of mill energy consumption. The chart also demonstrates that the scattering of the results significantly decreases with the increase of estimators, indicating lower influence of the other hyperparameters in the results.

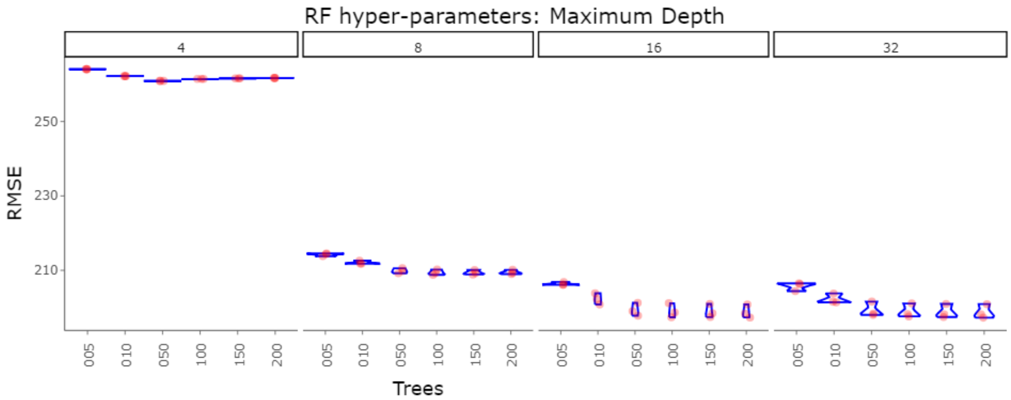
When the prediction errors are split by number of estimators and learning rate in Figure 28, similar conclusion can be drawn: the RMSE increases for lower learning rates and the

Figure 24 – Root Mean Square Error of the different RF experiments split by number of trees.



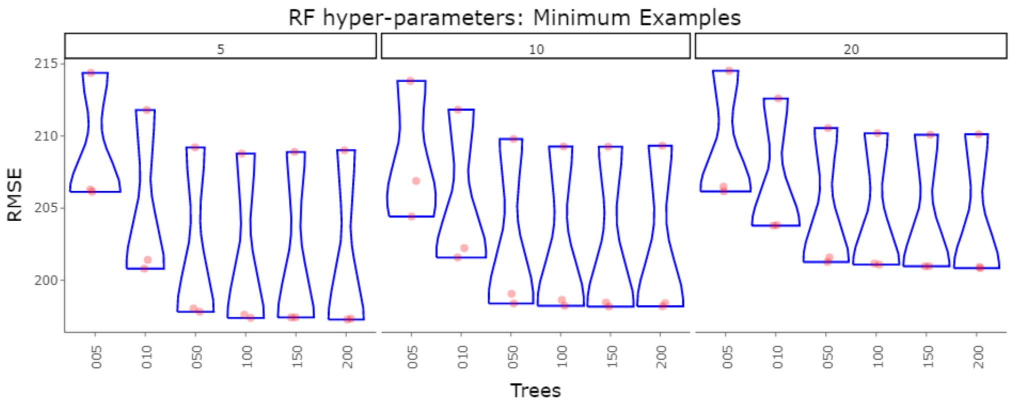
Source: Elaborated by the author (2021).

Figure 25 – Root Mean Square Error of the different RF experiments split by number of trees and maximum depth.



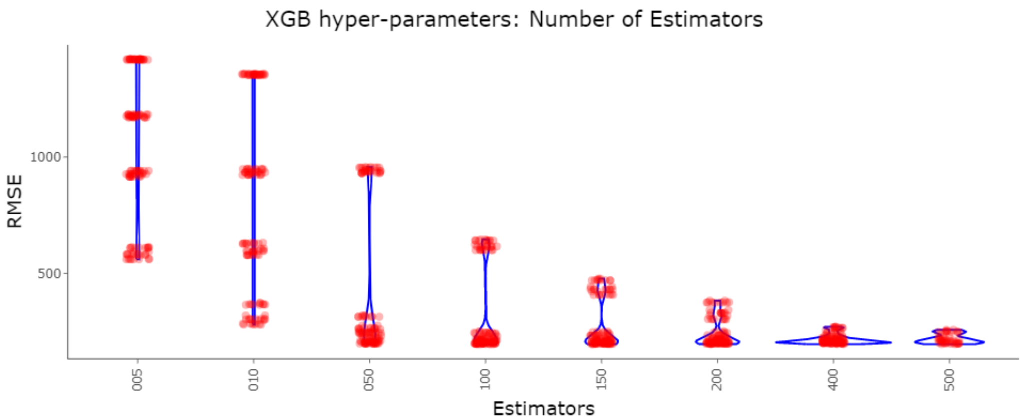
Source: Elaborated by the author (2021).

Figure 26 – Root Mean Square Error of the different RF experiments split by number of trees and minimum examples.



Source: Elaborated by the author (2021).

Figure 27 – Root Mean Square Error of the different XGB experiments split by number of estimators (trees).



Source: Elaborated by the author (2021).

scattering reduces with its increase.

Since the learning rate and number of estimators have a significant impact on model accuracy, the experiments with learning rate lower than 0.1 and number of estimators lower than 50 have been excluded from the evaluations of the other variables.

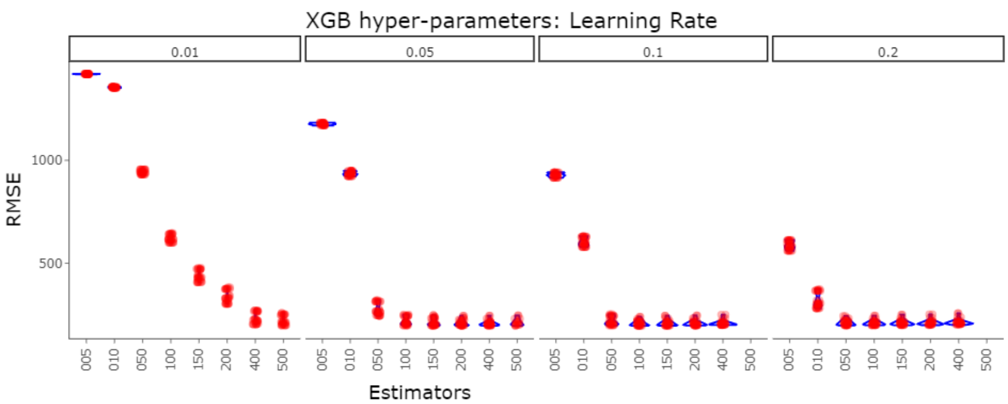
Surprisingly, the learning objective function, in this case, had indistinguishable impact in the model accuracy as Figure 29 indicates. The maximum depth and column samples per tree presented opposite behavior to one another: as the former increased the same behavior was observed in the RMSE (Figure 30) while opposite happened with the latter (Figure 31). It can also be seen in Figure 31 a significant reduction in the scattering of the RMSE with the increase in the column samples per tree, indicating improved model robustness in these experiments.

Evaluating only those experiments with column samples per tree equal to 0.5, Figure 32 shows that 0.1 learning rate presented lower RMSE, while the best performing results were reached with a maximum depth of 5 as can be seen in Figure 33. In both charts, it's clear that 100 estimators have presented the lowest RMSE in these experiments.

Based on these results and this discussion, the final configuration of the XGB algorithm selected was 100 estimators with a maximum depth of 5 and a learning rate of 0.1. The column samples per tree was defined as 0.5 and the learning objective function was squared error.

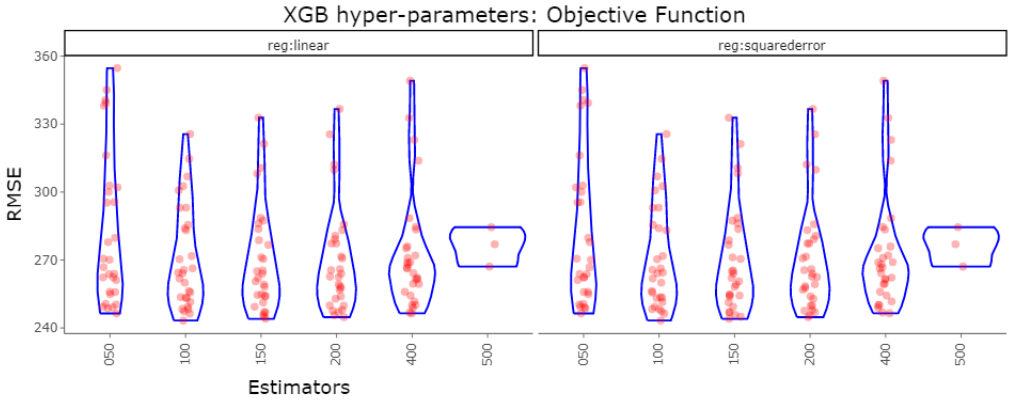
The next chapter details the training of the three regressors considering the configurations previously mentioned followed by detailed performance comparisons.

Figure 28 – Root Mean Square Error of the different XGB experiments split by learning rate.



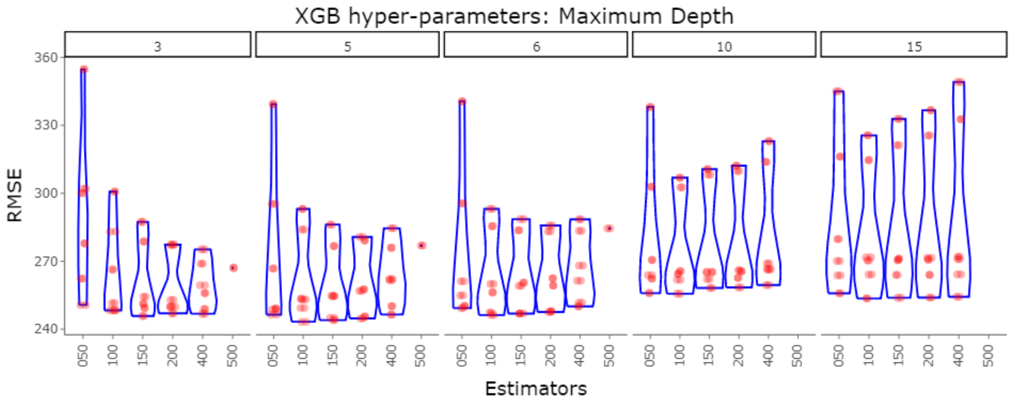
Source: Elaborated by the author (2021).

Figure 29 – Root Mean Square Error of the different XGB experiments split by learning objective function.



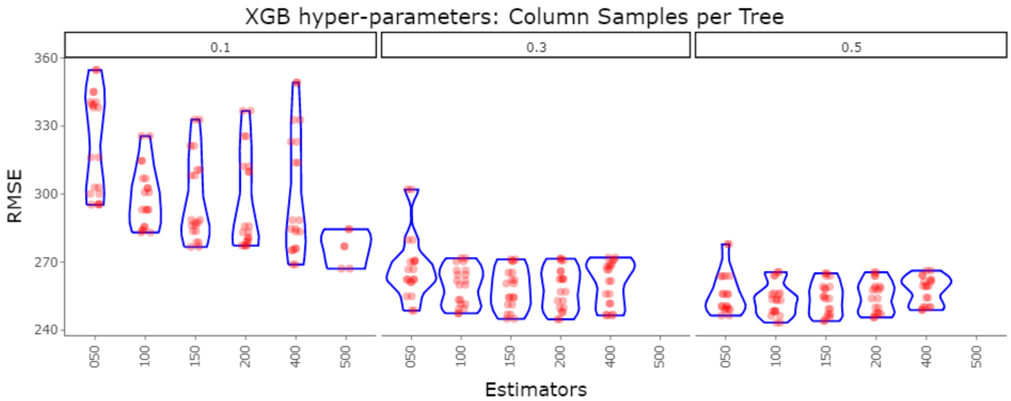
Source: Elaborated by the author (2021).

Figure 30 – Root Mean Square Error of the different XGB experiments split by maximum depth.



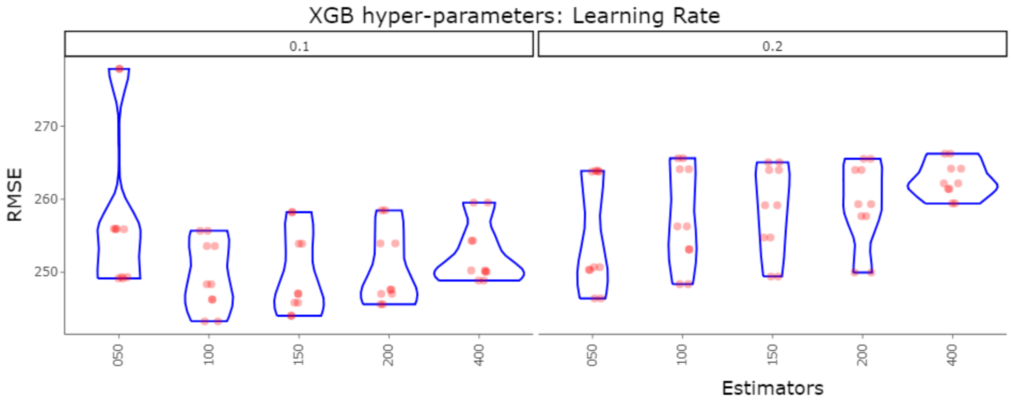
Source: Elaborated by the author (2021).

Figure 31 – Root Mean Square Error of the different XGB experiments split by column samples per tree.



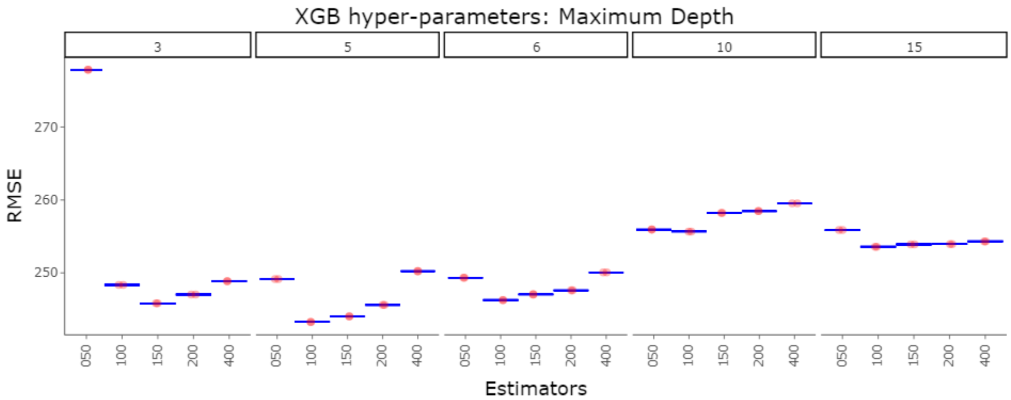
Source: Elaborated by the author (2021).

Figure 32 – Root Mean Square Error of the different XGB experiments split by learning rate for experiments with column samples per tree equal to 0.5.



Source: Elaborated by the author (2021).

Figure 33 – Root Mean Square Error of the different XGB experiments split by maximum depth for experiments with column samples per tree equal to 0.5.



Source: Elaborated by the author (2021).

5 MODEL PERFORMANCE COMPARISON

The database selection procedure exposed in the previous chapter with ANN showed that it presented preliminary promissing results. However, the No Free Lunch Theorem states that it's imperative to try different regressors in order to find the most suitable for each specific problem (WOLPERT; MACREADY, 1996). In this work, the RF and XGB algorithms have been selected as comparison parameters for the already mentioned ANN architecture. The configuration parameters of the RF and XGB regressors have been defined in a grid search procedure also detailed in previous chapter and the results are discussed below.

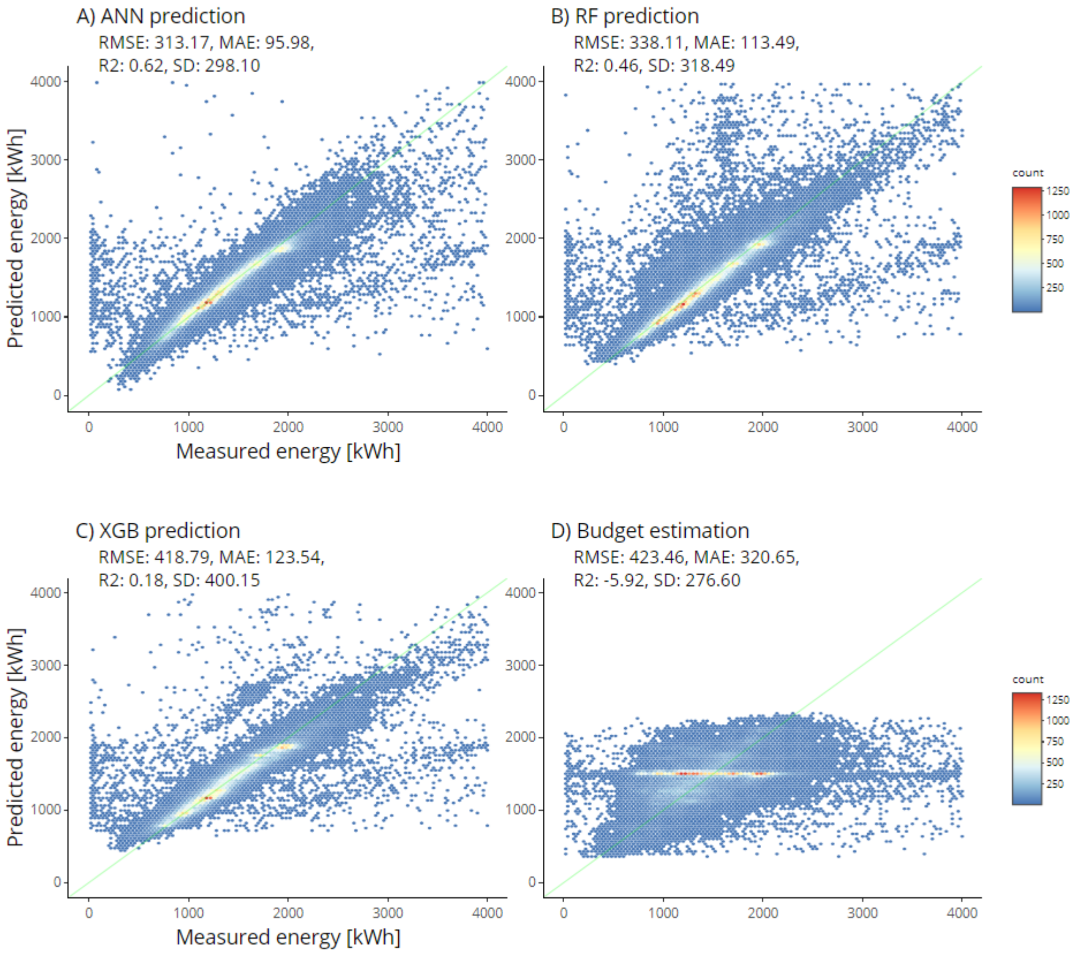
Another comparison basis used in this work is the current practice of budget preparation elaborated by the engineers of the operational team of the TCM. It is based on several years of historical evaluation and it consists in multiplying a constant to production weight to estimate expected consumption of energy by the process. This figure is then multiplied by the forecast of energy price for the next fiscal year to reach the financial budget for this important consumable in the company's financial report. This procedure has been used for many years and the averaging procedure to calculate the previously mentioned constant is frequently improved making its performance a very good benchmark for this model.

The scatter plot in Figure 34 shows that both the ANN and the RF regressors presented similar results on the energy prediction with a numerical advantage to the ANN, specially on the R-squared, which indicates the correlation between the variables in the chart. The XGB algorithm performed significantly worse than the other two and the budget estimation procedure is considerably less accurate than the proposed methods in coil-to-coil comparison.

Evaluation of the REC curve in Figure 35 confirms that ANN and RF regressors present similar performance for errors lower than 14% with the ANN notably better for errors higher than this. The XGB algorithm has a significantly worse performance than the RF for errors lower than 18% with significant improvement after. The REC chart also shows that the budget estimation method presents a much worse performance than the regressors on this coil-to-coil comparison, confirming the results observed on the scatter plots of Figure 34. On the other hand, the month comparison in Figure 36 (prediction error) shows how well-adjusted this procedure is for summarized data in monthly basis. In this condition, the budget estimation achieved a MAE of 0.31 and RMSE of 0.39, outperforming the ANN, which presented 0.34 and 0.40, respectively. The RF algorithm reached the best results in these metrics with 0.20 for MAE and 0.27 for RMSE while the XGB achieved a MAE of 0.24 and RMSE of 0.29. Similarly, Figure 37 shows that the RF algorithm presented the best results for mean relative error (MRE) with 2.90% followed by the XGB and the budget estimator with 3.41 and 4.49%, respectively, while the ANN presented an MRE of 4.81%, the worst performance in this metric too.

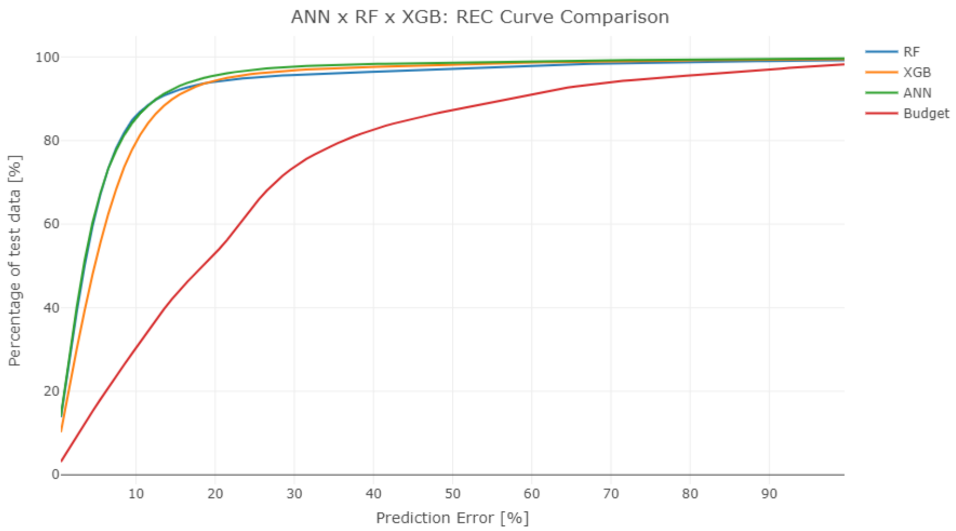
The history of training and validation errors for a typical ANN training can be seen in Figure 38. The procedure was set to finish at 1000 epochs, but it finished before epoch 700 because validation error ceased to reduce for 50 iterations. This figure indicates that overfitting

Figure 34 – Comparison of energy measurements against predictions for ANN model (A), RF model (B), XGB model (C) and budget estimation (D).



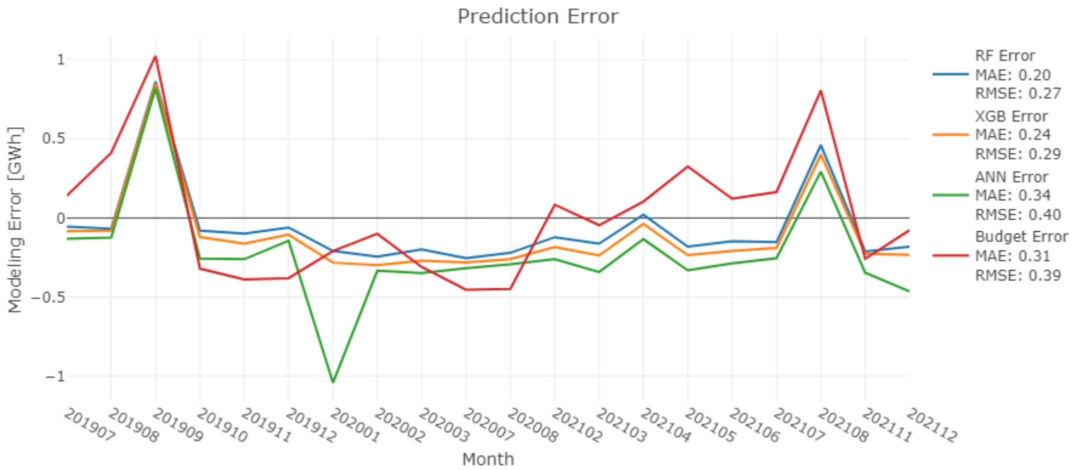
Source: Elaborated by the author (2021).

Figure 35 – REC curve of RF model (blue), XGB (orange), ANN (green) and budget reference (red).



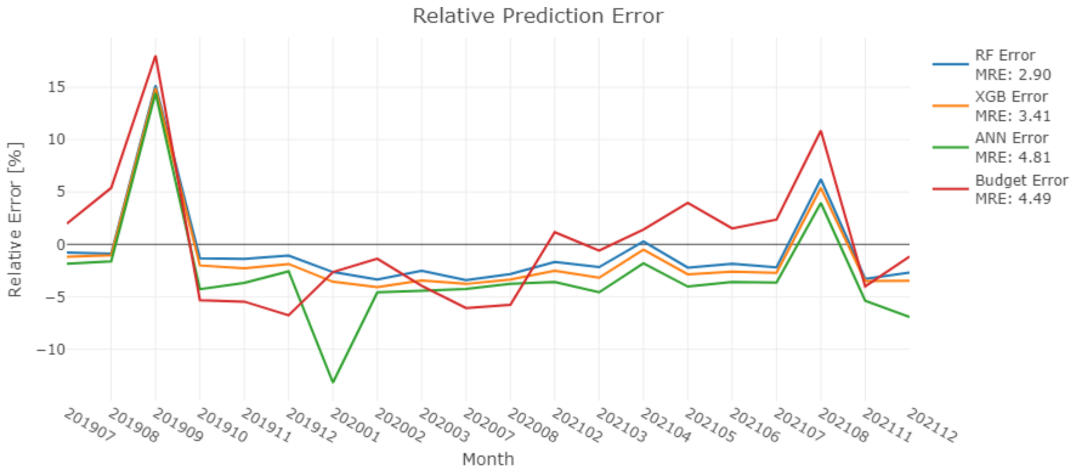
Source: Elaborated by the author (2021).

Figure 36 – Monthly prediction error of RF model (blue), XGB (orange), ANN (green) and budget reference (red).



Source: Elaborated by the author (2021).

Figure 37 – Monthly relative prediction error of RF model (blue), XGB (orange), ANN (green) and budget reference (red).

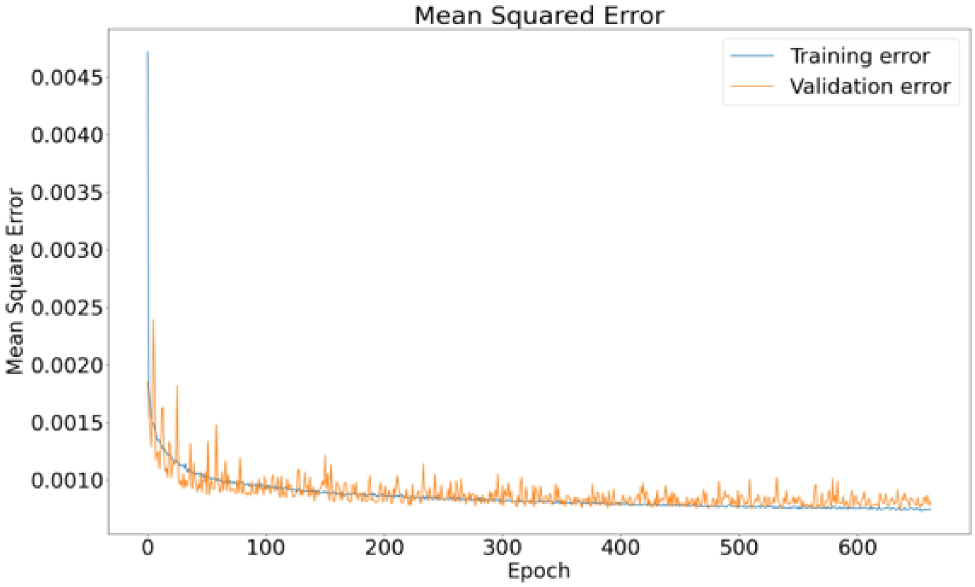


Source: Elaborated by the author (2021).

is not the cause of the unexpected lower performance of the ANN in this specific problem because significant separation between training and validation errors is not observed. The 162 experiments of the grid search procedure previously detailed in Chapter 3 shows that this architecture presented the best results. Recent developments on the ANN field have increased network complexity and improved performance in most cases. However, the literature review of Chapter 2 confirmed that the MLP is the most common in steel industry applications and similar problems, corroborating the selection of this architecture in this work.

Apart from outperforming the ANN on the month comparison and being just as good on a coil-to-coil reference, the RF and XGB presented another interesting advantage which was the training cost. Throughout the entire set of simulations executed to define model architecture, benchmark tests and experiments, it was noticeable that the required training time for the RF and

Figure 38 – History of training and validation errors in one typical training of the ANN.



Source: Elaborated by the author (2021).

XGB algorithms was ranging from 15 to 20 times faster than the ANN for the same database on a regular 16GB of RAM, 1.8GHz, 4 cores computer. The average training time with the current database was approximately 35 seconds for RF and XGB while the ANN needed over 10 minutes to complete the process. On the other hand, the ANN trained model required just a few kilobytes to be stored while the RF and XGB models required approximately 30 megabytes.

Based on this discussion, a decision was made to replace the ANN model with the RF model in the final version of the system in order to make it faster for the final users to retrain the model and perform their own experiments with new databases. The XGB algorithm was also included in the final version to provide additional information for decision-making.

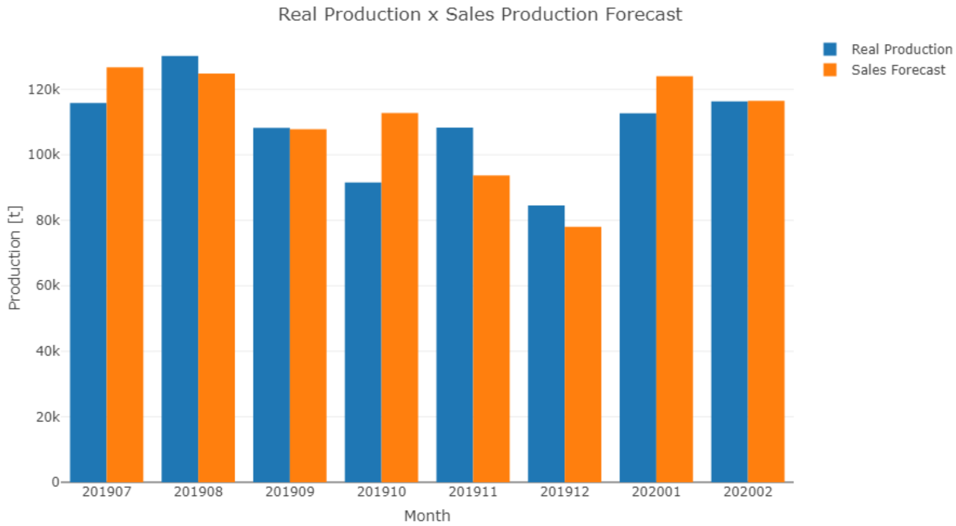
The next section of this work aims at simulating the practical application of this model with the objective of evaluating its functionality and capability of predicting the energy consumption.

5.1 MODEL APPLICATION

The objective of this model is to predict the energy consumption of a month of production which, in practice, must be done not with past data as in the investigations of Section 5. Instead, the energy prediction could consider for instance the forecast of the TCM production reviewed every month by the Scheduling Team based on orders placed by customer in the past weeks (since the delivery time of a coil to the final user is approximately 60 days) and the arrival of raw materials (coils from the hot strip mill) for the TCM.

This procedure of forecasting TCM production relies on assumptions that are frequently adjusted in order to have the most accurate production forecast possible, shown in Figure 39. One of these assumptions is, for example, the raw material delivery schedule. A shipment is not

Figure 39 – Monthly comparison of sales expected production and real production.



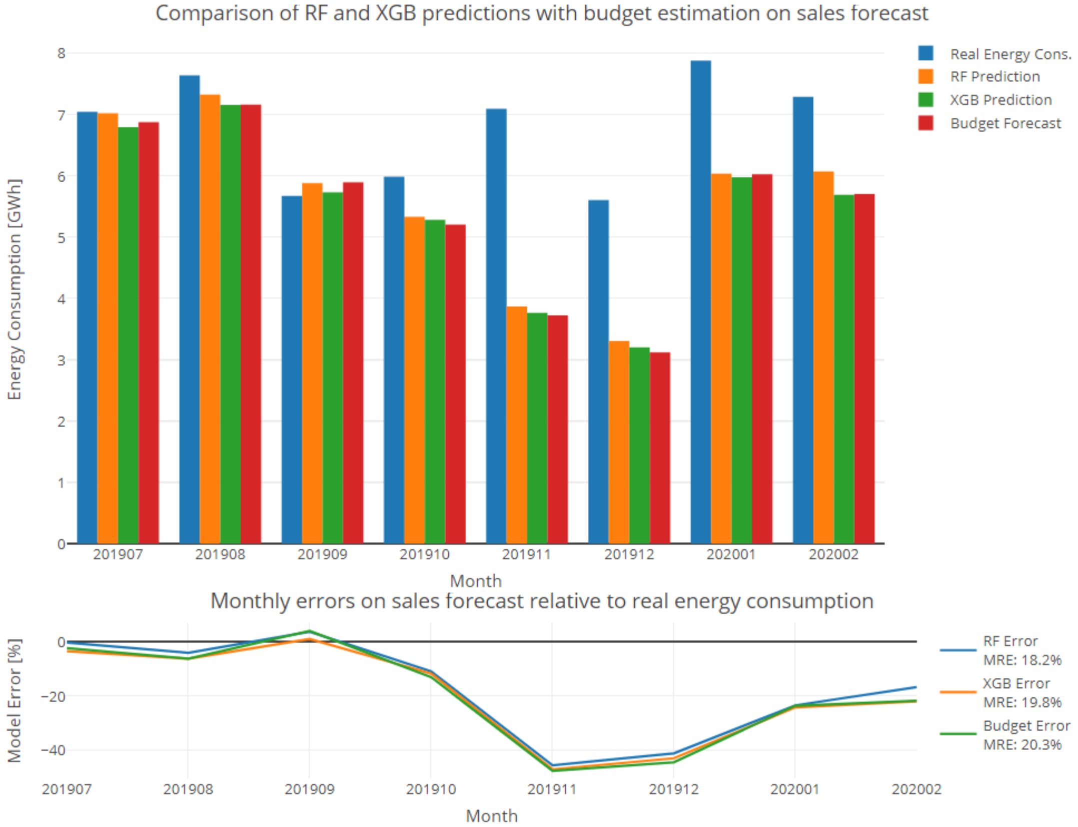
Source: Elaborated by the author (2021).

always feasible as planned in due time and one product that was supposed to be rolled in one specific month can be delayed to the next if needed (every coil is specially produced according to customer requirements). Another important assumption is that the TCM is going to perform following specific KPI's (such as unscheduled line stop ratio, line speed ratio, etc.) and the Operational and Maintenance teams are responsible for ensuring the production line will perform according to these figures. One of these KPIs is the productivity of the TCM used to estimate the time required to run the coils. This coil running time is a key figure to verify the TCM will be able to deliver the required total production of the month. Not by coincidence, the mill exit speed was among the input variables selected for the energy consumption model since this process variable can be easily calculated when the coil running time is available.

Based on the sales forecast and applying the budget preparation procedure already discussed in Section 5, the TCM engineering team is capable of forecasting the energy consumption of the TCM. The same assumptions for coil running time was used for calculating strip speed and inputted in the system to estimate the energy consumption for future TCM production. The model predictions are then compared with the actual energy measurement and the budget estimation and reproduced in Figure 40 (due to the confidentiality of this information, a smaller database was available). The MRE resulted from this evaluation are ranging from 18.2% to 20.3% in the short period investigated.

Part of this significant difference between the forecast and measured energy observed in Figure 40 is related to the production size deviations previously mentioned. For instance, a lower than expected production would consume less energy (month 2019/10 in Figure 39 is one example) or the contrary. However, production size does not explain the important variations observed in months 2019/11 and 12. Another possibility is a variation in the products expected to be rolled (called product mix) which influences the energy consumption since coil dimensions and hardness are important model inputs too. Separating these influences from others in the

Figure 40 – Monthly comparison of RF and XGB predictions and budget estimation using sales forecast data as model inputs.



Source: Elaborated by the author (2021).

results of the simulation observed in Figure 40 is not a trivial task.

A different approach has been proposed to evaluate the contribution of the inaccuracy of the prediction of production volume or mix by the sales forecast, or, in other words, to assess the accuracy of the assumptions of the sales forecasting methodology. It consists in applying these assumptions in actual production data by replacing the measured coil running time and speed by calculated values from the sales methodology. With this procedure, the production size and mix would not influence the simulation results while the process variables would still allow evaluating model sensitivity to sales assumptions.

The final results of this approach can be seen in Figure 41, where a significant improvement can be observed in comparison with Figure 40, as expected since part of the disturbance has been removed. The RF and XGB presented an MRE of 3.1% and 3.9%, respectively, while the budget estimation reached 5.9% in this approach. In Figure 41, to facilitate the comparison with real production data, this data from Section 5 is included in the chart for the period under investigation. It shows that the model predictions with sales assumptions have unexpectedly presented better performance than actual production data in this specific period for the RF algorithm. Evaluation of Figure 42, which compares the predicted energy in these two conditions for the RF and XGB algorithms clearly indicated that the sales assumptions had minimal influence in the

Figure 41 – Monthly comparison of RF and XGB predictions and budget estimation using production data but process variables recalculated with sales assumptions.

Source: Elaborated by the author (2021).

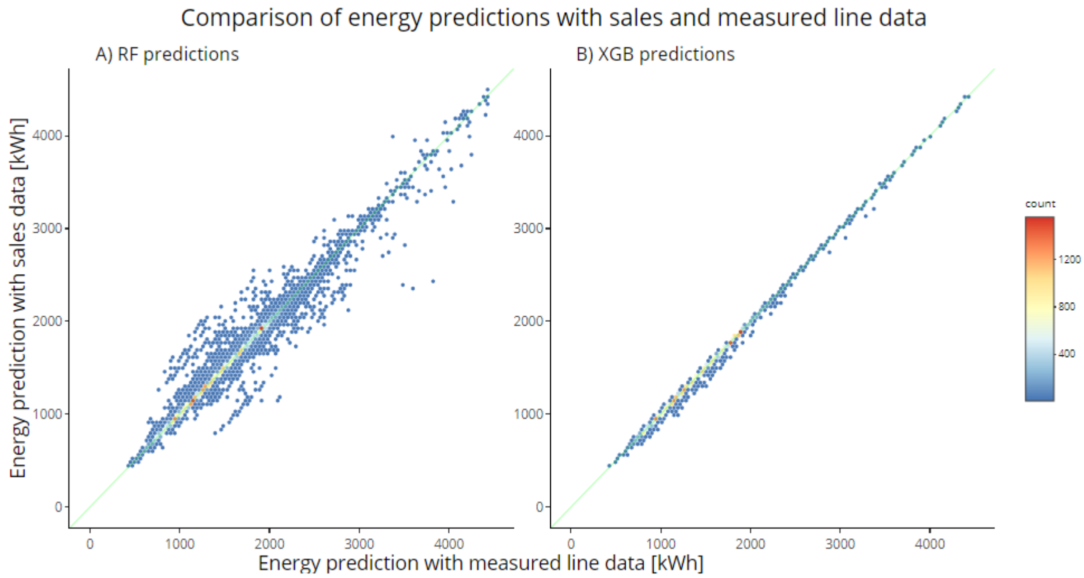
latter. In the former, the influence is noticeable but it was not possible to determine the reason for this unexpected behavior.

Other influencing factors in the accuracy of the energy budget prediction are energy cost, which can be easily assessed by accounting, and variations in the specific energy consumption (SEC), which is the constant value estimated by the engineers of the TCM operational team mentioned in Section 5. The developed energy prediction system is helpful in indicating variations in the SEC, as shown in Figure 43. In this chart the red line represents the SEC, currently estimated at 61.7kWh per ton of rolled steel. The other bars are obtained by dividing the monthly measured energy consumption and the model predictions by the total production of the TCM in each month. As can be seen both the real value and the model predictions present significant variation, and the predictions presented behavior very similar to the measurement.

This evaluation indicates that both the RF and XGB models detected variations in the input variables that caused oscillations in the SEC. This leads to the conclusion that even though the budget estimation procedure is robust for average evaluations it is not able to support deeper investigations when process or product conditions are varying.

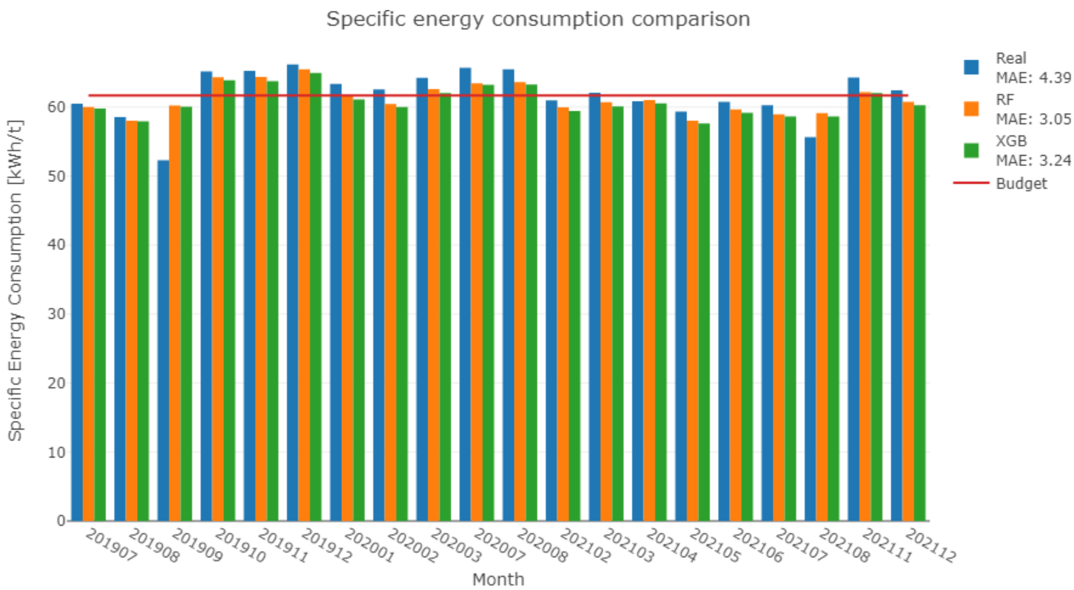
The next chapter presents closing remarks and conclusions on the work presented in this manuscript.

Figure 42 – Comparison of model energy prediction with actual line data and sales assumptions.



Source: Elaborated by the author (2021).

Figure 43 – Monthly comparison of real specific energy consumption and RF and XGB predictions in relation to budget reference.



Source: Elaborated by the author (2021).

6 CONCLUSIONS AND FUTURE WORK

The tandem cold rolling mill of steel strips is a high energy demanding process due to the high loads required to deform this strong material. Electrical energy has a significant impact on the financial balance of any steel enterprise and being able to accurately predict it provides strategic advantage in cost control, specially at this moment of prices on the rise all around the world.

The objective of this work was to develop a machine learning model capable of predicting the energy consumption of the tandem cold mill for an entire month of production and to make it as undependable as possible to very short-term process variables.

A systematic literature review has found dozens of references with the application of machine learning algorithms on the downstream lines of the steel industry however none of them is related to electrical energy prediction. This systematic review is a significant contribution of this work since it shows that this subject has not been given enough attention regardless of its recognized importance on the financial and environmental aspects of any steel manufacturer.

During the development of this work, it was also observed that the available energy database suffers from significant reliability issues. This demanded a set of experiments to identify the best methodology for data collecting and processing in order to improve model accuracy, reducing the mean absolute error in these preliminary experiments from 17.2% to 7.2%.

In addition to this, a thorough comparison between artificial neural networks, random forest and extreme gradient boosting regressors have been carried out to identify the best approach to tackle this univariate regression problem. A grid search on the hyperparameters of both algorithms allowed the selection of appropriate architectures for the modeling of energy consumption and the cross validation proved its robustness.

The random forest algorithm outperformed the artificial neural network in all the tested conditions on a monthly basis with an MRE of 2.90% while the neural network reached 4.81%. The XGB algorithm was not as good as the RF but also better than the ANN with 3.41% for MRE. The comparison parameter of this work, an energy estimation procedure based on historical data developed by the cold mill engineering team, presented an MRE of 4.49%, also better than the ANN. These comparisons are another important contribution of this document because they provide an example of a problem where the ANN is outperformed in every metric by a statistical model while RF and XGB present significantly better results.

This unexpected poor performance of the ANN in comparison to the other three references added to the considerably longer training time have justified removing the ANN algorithm from the final version of the system.

These results also showed that the existing energy estimation procedure presents relatively robust performance on a monthly basis, however, on a coil-to-coil basis its prediction accuracy was significantly lower than the proposed algorithms. This advantage of the system can be important when specific evaluations are required, for example, predicting energy consumption with unusual product mix.

The evaluation of the model performance with future data in a simulation of its practical use showed that the model is highly dependent on the accuracy of the production size estimation and product mix forecast. An evaluation of model sensitivity to sales assumptions for the calculation of process variables indicated that the model is robust to these procedures since both the RF and XGB presented a lower MRE in this situation than with production data for the investigation period.

This model has also proven to be relevant in detailed analysis since it was able to detect variations in the specific energy consumption, providing valuable information for the production engineers when unusual energy variations must be investigated.

However, the added value of the RF and XGB models was only reachable with the additional effort of extensive data manipulation and the merging of two independent databases (IBA and PIMS) to improve the quality of the acquired dataset. With the recent popularization of Industry 4.0 concepts, sensors and data acquisition systems are becoming less and less expensive making it highly advisable for the company to evaluate improving the energy data acquisition and storage to improve its quality and reliability.

One future opportunity to continue this development is to improve the procedure to define model architecture and consider self-adjustment algorithms, avoiding the grid search method. Extending this energy prediction system approach to the other steel processing lines such as the Pickling Line, Galvanizing Lines and Hot Strip Lines could also be considered. Another possibility is to explore the suitability of ML techniques to predict different energy sources such as natural gas, which is another significant cost to steel operations. Additionally, this model could be extended to focus not only on predicting energy consumption but also on supporting action plans to reduce it by adjusting process variables and rolling conditions.

These proposals will certainly require extensive research and dedication to be carried on (as this work did), but would definitely contribute to the academia and provide added value to the steel industry.

BIBLIOGRAPHY

ABIODUN, O. I. et al. State-of-the-art in artificial neural network applications: A survey. **Heliyon**, v. 4, n. 11, p. e00938, nov. 2018. ISSN 24058440. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2405844018332067>. Cited in page 21.

AKANKSHA, E. et al. Review on Reinforcement Learning, Research Evolution and Scope of Application. **2021 5th International Conference on Computing Methodologies and Communication (ICCMC)**, p. 1416–1423, abr. 2021. Cited in page 20.

ALEXANDER, J. M. On the Theory of Rolling. **Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences**, v. 326, n. 1567, p. 535–563, 1972. ISSN 0080-4630. Publisher: The Royal Society. Disponível em: <https://www.jstor.org/stable/77929>. Cited 3 times in page 17, 18, and 36.

ANTONIADIS, S.; LAMBERT-LACROIX, S.; POGGI, J. M. Random forests for global sensitivity analysis: A selective review. **Reliability Engineering & System Safety**, v. 206, p. 107312, 2021. Cited in page 22.

ARUMUGAM, M.; CHANDRAMOHAN, A.; MURTHY, G. On the optimal control of steel annealing processes via various versions of genetic and particle swarm optimization algorithms. **Optimization and Engineering**, v. 12, n. 3, p. 371–392, 2011. ISSN 13894420 (ISSN). Publisher: Kluwer Academic Publishers. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-79960264239&doi=10.1007%2fs11081-011-9143-5&partnerID=40&md5=0d858b1072ceffd64c10962a7589499f>. Cited in page 28.

BELGIU, M.; DRAGUT, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Phtogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016. Cited 3 times in page 22, 23, and 45.

BIDABADI, B. S. et al. Optimization of required torque and energy consumption in the roll forming process. **International Journal on Interactive Design and Manufacturing**, v. 13, n. 3, p. 1029–1048, 2019. ISSN 19552513. Publisher: Springer-Verlag France. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85064259116&doi=10.1007%2fs12008-019-00564-9&partnerID=40&md5=c4b0c20c55f950e9e40968d8e791f881>. Cited in page 26.

BROWNLEE, J. **Clever Algorithms: Nature-inspired Programming Recipes**. Melbourne, Australia: Jason Brownlee, 2011. Google-Books-ID: SESWXQphCUkC. ISBN 978-1-4467-8506-5. Cited 2 times in page 21 and 37.

BU, H. et al. Rolling-schedule multi-objective optimization based on inuence function for thin-gauge steel strip in tandem cold rolling. **Scientia Iranica**, v. 23, n. 6, p. 2663–2672, 2016. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85014025204&partnerID=40&md5=bdac0255b679b29aff114ccfd746a7cd>. Cited in page 28.

BU, H.-N.; YAN, Z.-W.; ZHANG, D.-H. Application of case-based reasoning-Tabu search hybrid algorithm for rolling schedule optimization in tandem cold rolling. **Engineering Computations (Swansea, Wales)**, v. 35, n. 1, p. 187–201, 2018. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85044322802&doi=10.1108%

2fEC-02-2017-0054&partnerID=40&md5=8f1e29f688d7a927bff3cc348a3b1d62>. Cited in page 28.

BUCHMAYR, B.; DEGNER, M.; PALKOWSKI, H. Future Challenges in the Steel Industry and Consequences for Rolling Plant Technologies. **BHM Berg- und Hüttenmännische Monatshefte**, v. 163, p. 1–8, jan. 2018. Cited in page 14.

CASTRO, L. N. de. Fundamentals of natural computing: an overview. **Physics of Life Reviews**, v. 4, n. 1, p. 1–36, mar. 2007. ISSN 1571-0645. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1571064506000315>. Cited in page 15.

CHE, H. et al. Optimization of schedule with multi-objective for tandem cold rolling mill based on IAGA. **2010 International Conference on Mechanic Automation and Control Engineering**, p. 3503–3506, jun. 2010. Cited in page 28.

CHE, H. et al. PSO algorithm-based schedule optimization for tandem cold mills. **2009 IEEE International Conference on Automation and Logistics**, p. 944–948, ago. 2009. ISSN: 2161-816X. Cited in page 28.

CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. San Francisco California USA: ACM, 2016. p. 785–794. ISBN 978-1-4503-4232-2. Disponível em: <https://dl.acm.org/doi/10.1145/2939672.2939785>. Cited in page 23.

CHEN, Y. et al. Research on Rolling Force Prediction Method of High Precision Cold Rolling Based on XGBoost Algorithm. **2020 7th International Forum on Electrical Engineering and Automation (IFEEA)**, p. 966–971, set. 2020. Cited in page 24.

COHEN, M. W.; FOXX, H.; ALUL, S. B. A decision support flexible scheduling system for continuous galvanization lines using genetic algorithm. **Production Engineering**, v. 13, n. 1, p. 43–52, 2019. ISSN 09446524 (ISSN). Publisher: Springer Verlag. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85055961798&doi=10.1007%2fs11740-018-0856-6&partnerID=40&md5=30acb29c1e26d8f449bbda18ff3e86d2>. Cited in page 28.

FERNANDEZ, S. et al. Scheduling a Galvanizing Line by Ant Colony Optimization. In: DORIGO, M. et al. (Ed.). **Swarm Intelligence**. Cham: Springer International Publishing, 2014. (Lecture Notes in Computer Science), p. 146–157. ISBN 978-3-319-09952-1. Cited in page 28.

FRESHWATER, I. J. Simplified theories of flat rolling—I. The calculation of roll pressure, roll force and roll torque. **International Journal of Mechanical Sciences**, v. 38, n. 6, p. 633–648, jun. 1996. ISSN 0020-7403. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0020740396800063>. Cited 4 times in page 17, 18, 30, and 36.

GAO, C.; TANG, L.; WANG, Y. Model and scheduling of a continuous galvanizing line. v. 2, p. 1829–1834, 2008. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-58049142329&doi=10.1109%2fSOLI.2008.4682827&partnerID=40&md5=038fa5d767a672a1fd5e85e78f1eff05>. Cited in page 28.

GONZáLEZ-BRIONES, A. et al. Machine Learning Models for Electricity Consumption Forecasting: A Review. **2019 2nd International Conference on Computer Applications Information Security (ICCAIS)**, p. 1–6, maio 2019. Cited 2 times in page 20 and 37.

GUO, C.; XU, Z.; YAO, Q. Information Fusion and XGBoost Algorithm Used for Bearing Remaining Useful Life Prediction. **2021 China Automation Congress (CAC)**, p. 1689–1693, out. 2021. ISSN: 2688-0938. Cited in page 23.

HERNáNDEZ-ORALLO, J. ROC curves for regression. **Pattern Recognition**, v. 46, n. 12, p. 3395–3411, dez. 2013. ISSN 0031-3203. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0031320313002665>. Cited in page 38.

HU, P.-H.; EHMANN, K. Dynamic model of the rolling process. part i: Homogeneous model. **International Journal of Machine Tools and Manufacture**, v. 40, n. 1, p. 1–19, 2000. Cited By 90. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0343618403&doi=10.1016%2fS0890-6955%2899%2900049-8&partnerID=40&md5=aca84b67964fa25e155fd2b4bc7d032f>. Cited in page 18.

HU, Z. et al. Optimization of Metal Rolling Control Using Soft Computing Approaches: A Review. **Archives of Computational Methods in Engineering**, nov. 2019. ISSN 1886-1784. Disponível em: <https://doi.org/10.1007/s11831-019-09380-6>. Cited 8 times in page 15, 18, 19, 22, 23, 24, 25, and 36.

KAPANOGLU, M.; KOC, I. A multi-population parallel genetic algorithm for highly constrained continuous galvanizing line scheduling. **HM**, v. 4030 LNCS, 2006. Pages: 41. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-33750079846&doi=10.1007%2f11890584_3&partnerID=40&md5=4fb6b761ebc2fd62af150949fb55a86c>. Cited in page 28.

KIM, S. W. et al. Recent Advances of Artificial Intelligence in Manufacturing Industrial Sectors: A Review. **International Journal of Precision Engineering and Manufacturing**, nov. 2021. ISSN 2005-4602. Disponível em: <https://doi.org/10.1007/s12541-021-00600-3>. Cited 2 times in page 22 and 23.

KOZHEVNIKOV, A.; KOZHEVNIKOVA, I.; BOLOBANOVA, N. Development of the model of cold rolling process in dynamic conditions. **Journal of Chemical Technology and Metallurgy**, v. 53, n. 2, p. 366–372, 2018. Cited By 5. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85042114171&partnerID=40&md5=952b54fe93f4ea7b1fa92c42a186f712>. Cited in page 18.

LEGRAND, N. et al. New cooling techniques to enhance roll bite lubrication, lower water usage and decrease energy consumption on cold strip mills. **Proceedings of the 10th International Conference on Steel Rolling**, 2010. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84889922172&partnerID=40&md5=de86514238418bcea108fbdce488ff6c>. Cited in page 26.

LENARD, J. G. **Primer on flat rolling**. 2nd edition. ed. Waltham, MA: Elsevier, 2013. ISBN 978-0-08-099418-5. Cited 2 times in page 19 and 36.

LI, Y.; LIU, J.; WANG, Y. An adaptive weight PSO for rolling schedules multi-objective optimization of tandem cold rolling. **Proceedings of the 2009 IEEE International Conference on Automation and Logistics, ICAL 2009**, p. 895–899, 2009. Cited in page 28.

LIU, C. et al. Elastic-plastic finite-element modelling of cold rolling of strip. **International Journal of Mechanical Sciences**, v. 27, n. 7, p. 531–541, jan. 1985. ISSN 0020-7403. Disponível em: <http://www.sciencedirect.com/science/article/pii/0020740385900438>. Cited 2 times in page 17 and 18.

LIU, Y. M. et al. Mathematical model for cold rolling based on energy method. **Meccanica**, v. 52, n. 9, p. 2069–2080, jul. 2017. ISSN 0025-6455, 1572-9648. Disponível em: <http://link.springer.com/10.1007/s11012-016-0569-x>. Cited in page 18.

MARTíNEZ-DE-PISóN, F. J. et al. Optimising annealing process on hot dip galvanising line based on robust predictive models adjusted with genetic algorithms. **Ironmaking & Steelmaking**, v. 38, n. 3, p. 218–228, abr. 2011. ISSN 0301-9233. Publisher: Taylor & Francis. Disponível em: <https://www-tandfonline.ez74.periodicos.capes.gov.br/doi/full/10.1179/1743281210Y.0000000001>. Cited in page 28.

MIYABE, Y.; BIEGL, C.; KAWAMURA, K. Methodologies for a real-time intelligent supervisory system for a hot strip mill finisher. In: **Proceedings of the 1st international conference on Industrial and engineering applications of artificial intelligence and expert systems - Volume 1**. New York, NY, USA: Association for Computing Machinery, 1988. (IEA/AIE '88), p. 483–491. ISBN 978-0-89791-271-6. Disponível em: <http://doi.org/10.1145/51909.51964>. Cited in page 19.

MORE, A. S.; RANA, D. P. Review of random forest classification techniques to resolve data imbalance. **1st International Conference on Intelligent Systems and Information Management (ICISIM)**, p. 72–78, 2017. Cited in page 22.

MOSAVI, A. et al. State of the Art of Machine Learning Models in Energy Systems, a Systematic Review. **Energies**, v. 12, n. 7, p. 1301, jan. 2019. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute. Disponível em: <https://www.mdpi.com/1996-1073/12/7/1301>. Cited 2 times in page 15 and 24.

NASTASI, G.; COLLA, V.; SEPPIA, M. D. A Route Planning Optimisation System for the Steelmaking Industry Based on Multi-objective Evolutionary Algorithms. **2014 European Modelling Symposium**, p. 326–331, out. 2014. Cited in page 28.

OLIVEIRA, D. G. de et al. Artificial Neural Network Model for Steel Strip Tandem Cold Mill Power Prediction. **International Conference on Applied Informatics**, p. 29–42, 2020. Cited 2 times in page 28 and 31.

ONS. **Operador Nacional do Sistema Elétrico**. 2021. Disponível em: <http://ons.org.br:80/paginas/sobre-o-ons/o-que-e-ons>. Cited 2 times in page 14 and 15.

OROWAN, E. The Calculation of Roll Pressure in Hot and Cold Flat Rolling. **Proceedings of the Institution of Mechanical Engineers**, v. 150, n. 1, p. 140–167, jun. 1943. ISSN 0020-3483. Publisher: IMECHE. Disponível em: <https://doi.org/10.1243/PIME_PROC_1943_150_025_02>. Cited in page 17.

PETERSEN, K. et al. Systematic Mapping Studies in Software Engineering. jun. 2008. Disponível em: <https://scienceopen.com/document?vid=6d552894-2cc3-4e2b-a483-41fa48a37ef8>. Cited 2 times in page 15 and 24.

POURSINA, M. et al. Application of genetic algorithms to optimization of rolling schedules based on damage mechanics. **Simulation Modelling Practice and Theory**, v. 22, p. 61–73, 2012. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84455171001&doi=10.1016%2fj.simpat.2011.11.005&partnerID=40&md5=5dd6f0560f89732b2cfffeaab66cdfe6>. Cited in page 28.

QU, Q. et al. Establishment and Optimization of Gas Flow Prediction Model for Annealing Furnace Based on GA-SVM. **2018 37TH CHINESE CONTROL CONFERENCE (CCC)**, p. 3486–3490, 2018. ISSN: 2161-2927. Cited in page 28.

RATH, S. et al. Application of Machine Learning in Rolling Mils: Case Studies. p. 17, 2019. Cited 2 times in page 22 and 23.

RAY, S. A Quick Review of Machine Learning Algorithms. **2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)**, p. 35–39, fev. 2019. Cited 2 times in page 19 and 20.

ROBERTS, W. L. **Cold rolling of steel**. New York: M. Dekker, 1978. (Manufacturing engineering and materials processing, 2). ISBN 978-0-8247-6780-8. Cited 4 times in page 14, 17, 19, and 36.

SAGI, O.; ROKACH, L. Ensemble learning: A survey. **WIREs Data Mining and Knowledge Discovery**, v. 8, n. 4, p. e1249, jul. 2018. ISSN 1942-4787. Publisher: John Wiley & Sons, Ltd. Disponível em: <https://wires-onlinelibrary-wiley.ez74.periodicos.capes.gov.br/doi/10.1002/widm.1249>. Cited in page 23.

SMITH, M. **Neural Networks for Statistical Modeling**. 1. ed. New York: Van Nostrand Reinhold, 1993. v. 1. Cited 2 times in page 29 and 36.

SUN, J. et al. Mathematical model of lever arm coefficient in cold rolling process. **The International Journal of Advanced Manufacturing Technology**, v. 97, n. 5, p. 1847–1859, jul. 2018. ISSN 1433-3015. Disponível em: <https://doi.org/10.1007/s00170-018-2078-7>. Cited in page 18.

SZűCS, M.; KRáLLICS, G.; LENARD, J. A Comparative Evaluation of Predictive Models of the Flat Rolling Process. **Periodica Polytechnica Mechanical Engineering**, v. 62, n. 2, p. 165–172, fev. 2018. ISSN 1587-379X. Number: 2. Disponível em: <https://pp.bme.hu/me/article/view/11847>. Cited in page 18.

TANG, L.; GAO, C. A modelling and tabu search heuristic for a continuous galvanizing line scheduling problem. **ISIJ International**, v. 49, n. 3, p. 375–384, 2009. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-67649961399&doi=10.2355%2fisijinternational.49.375&partnerID=40&md5=124f20002b8745a899e5f004a31cdf16>. Cited in page 28.

VERDEJO, V. V.; ALARCó, M. A. P.; SORLí, M. P. L. Scheduling in a continuous galvanizing line. **Computers & Operations Research**, v. 36, n. 1, p. 280–296, jan. 2009. ISSN 0305-0548. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0305054807001761>. Cited 2 times in page 24 and 28.

WANG, D. et al. Toward a heuristic optimum design of rolling schedules for tandem cold rolling mills. **Engineering Applications of Artificial Intelligence**, v. 13, n. 4, p. 397–406, ago. 2000. ISSN 09521976. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0952197600000166>. Cited in page 28.

WANG, D. D.; TIEU, A. K.; D'ALESSIO, G. Computational Intelligence-Based Process Optimization for Tandem Cold Rolling. **Materials and Manufacturing Processes**, v. 20, n. 3, p. 479–496, maio 2005. ISSN 1042-6914. Publisher: Taylor & Francis. Disponível em:

<https://www-tandfonline.ez74.periodicos.capes.gov.br/doi/full/10.1081/AMP-200053535>. Cited in page 28.

WANG, L. et al. Dynamic scheduling with production process reconfiguration for cold rolling line. **International Federation of Automatic Control**, v. 44, p. 12114–12119, 2011. Issue: 1 PART 1. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84866749756&doi=10.3182%2f20110828-6-IT-1002.01296&partnerID=40&md5=036a205c33599fd0b9455e3e584d715b>. Cited in page 28.

WANG, Y. et al. Multi-Objective Optimization of Rolling Schedule for Five-Stand Tandem Cold Mill. **IEEE Access**, v. 8, p. 80417–80426, 2020. ISSN 2169-3536. Conference Name: IEEE Access. Cited in page 28.

WANG, Z. et al. Design and development of intelligent optimization System for Cold Continuous Rolling Rules. **2010 8th World Congress on Intelligent Control and Automation**, p. 1371–1375, jul. 2010. Cited in page 28.

WOLPERT, D.; MACREADY, W. No Free Lunch Theorems for Search. mar. 1996. Cited in page 53.

YANG, Y.; TANG, L. Continuous annealing production scheduling in iron & steel industry. **Proceedings of the IEEE International Conference on Automation and Logistics, ICAL 2008**, p. 940–945, 2008. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-56449090422&doi=10.1109%2fICAL.2008.4636285&partnerID=40&md5=6447b0788c0b3764a2006f4cab15fb84>. Cited in page 28.

YONG, L.; LEI, F.; YU, W. Opposition learning adaptive cross-generation differential evolution algorithm based multi-objective optimization of rolling schedule for tandem cold rolling. **2016 IEEE International Conference on Information and Automation (ICIA)**, p. 319–323, ago. 2016. Cited in page 28.

ZHANG, F.; ZHAO, Y.; SHAO, J. Rolling force prediction in heavy plate rolling based on uniform differential neural network. **Journal of Control Science and Engineering**, v. 2016, 2016. ISSN 16875249 (ISSN). Publisher: Hindawi Publishing Corporation. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84978370286&doi=10.1155%2f2016%2f6473137&partnerID=40&md5=300f8d24b22d7604fc6015d74ba9a351>. Cited in page 28.

ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks. **International Journal of Forecasting**, v. 14, n. 1, p. 35–62, mar. 1998. ISSN 01692070. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0169207097000447>. Cited 4 times in page 21, 22, 29, and 38.

ZHANG, Y.; ZHU, J. Research on application of intelligent immune bee colony algorithm for production scheduling. **2012 International Conference on System Science and Engineering (ICSSE)**, p. 270–275, jun. 2012. ISSN: 2325-0925. Cited in page 28.

ZHAO, X. et al. Optimization of tandem cold rolling schedule based on competitive coevolution algorithm. **Proceedings of the 33rd Chinese Control Conference, CCC 2014**, p. 6500–6504, 2014. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84907921149&doi=10.1109%2fChiCC.2014.6896063&partnerID=40&md5=740f0a02f624b7491b2b5263a201625b>. Cited in page 28.