IoT to Efficiency: A Data-Driven approach to optimize Condensate Recovery Factor (CRF)

A Thesis

submitted to
the Department of Data Science,
Indian Institute of Science Education and Research, Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Shreshth Raj

Indian Institute of Science Education and Research
Pune - 411008, INDIA

April, 2025

Supervisor: Mr. Shripad Diwakar Kulkarni

© Shreshth Raj 2025

All rights reserved

Certificate

This is to certify that this dissertation entitled **IoT to Efficiency: A Data-Driven approach to optimize Condensate Recovery Factor (CRF)** towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by **Shreshth Raj** at Forbes Marshall India Pvt. Limited under the supervision of Mr. Shripad Diwakar Kulkarni, Principal Lead - Research and Developement, during the academic year 2024-2025.

S. D. Kulkarni

Mr. Shripad Diwakar Kulkarni

Committee:

Mr. Shripad Diwakar Kulkarni

Dr. Amit Apte

For the love of data and the pursuit of sustainable energy!

Declaration

I, **Shreshth Raj, Reg. No: 20201168**, hereby declare that the matter embodied in the report entitled **IoT to Efficiency: A Data-Driven approach to optimize Condensate Recovery Factor (CRF)** submitted to the Department of Data Science at the Indian Institute of Science Education and Research, Pune towards the partial requirement of **Master of Science** are the results of the work carried out by me at Forbes Marshall India Pvt. Limited, under the supervision of Mr. Shripad Diwakar Kulkarni and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.

Shreshth Raj

Shreshth Raj

Acknowledgement

No creation in this world is a solo effort and this work is no exception. From friends, family, and co-workers to the unsung hero at the coffee machine who made caffeine available every time inspiration faded, everyone has a role.

First and foremost, I'd like to thank Mr. Shripad Kulkarni and Mr. Prateek Muley (Forbes Marshall, India), for guiding me through this journey and giving me enough freedom to contribute my own ideas. The camaraderie and shared sense of purpose made this experience both enriching and enjoyable. I would also like to thank Dr. Amit Apte (IISER Pune), for his valuable suggestions and comments. Another thanks goes to all the domain experts of Forbes Marshall for providing essential domain knowledge and insightful feedback.

To my friends and family, thank you for being my unwavering support system, and for reminding me to stay grounded and focused. A special mention goes to my mom, Mrs. Neelam, whose endless love, patience, and unwavering faith in my abilities have been my greatest source of inspiration. Her encouragement and quiet strength have guided me through every challenge with grace and determination.

To everyone who played a role in shaping this work, whether through technical insights, moral support, or shared laughter, I am deeply grateful. This thesis is as much yours as it is mine.

Thank you all.

Abstract

In industrial process plants, optimization of the Condensate Recovery Factor (CRF) is critical with regard to energy efficiency and sustainability since it quantifies the reuse of condensate - a byproduct of a steam system, to reduce waste of resources and cost of operations. Despite the developments in IoT-enabled data gathering, there are challenges in transforming minute-scale steam production and the succeeding condensate recovery information into actionable conclusions. Traditional frameworks struggle with changing operational complexities such as varying rates of steam flow and system delay, whereas black-box machine learning methods lacks transparency, which hinders trust and practical implementation.

This research addresses these gaps by suggesting a hybrid framework that combines physics-informed machine learning with domain knowledge. A Fourier-inspired neural network structure is employed to represent temporal condensate recovery cycle patterns, assisted with genetic algorithms for hyperparameter adaptation. To bridge the gap between automation and human judgement, Explainable Artificial Intelligence (XAI) techniques like SHAP and LIME are leveraged to explain the model behavior, thus enabling engineers to compare predictions with operational limitations. The approach focuses on interpretability without compromising on predictive power, ensuring that the insights gained are consistent with industrial best practices.

By harmonizing IoT data streams with collaborative human-AI analysis, this work advances data-driven decision-making in steam system operations. It illustrates how Industry 4.0 technologies can transform underutilized datasets and convert it into strategic assets, driving sustainability through smarter resource utilization. The outcomes highlights the relevance of embedding domain knowledge into an AI system, and providing a scalable solution for industries to achieve operational excellence amidst the intricacies of digital innovations transformation.

Contents

1 Introduction

"In God we trust; all others bring data."
– W. Edwards Deming

1.1 Motivation

Efficient resource management stands as a pillar of industrial sustainability, most notably with respect to conserving energy and optimizing cost. With the advanced digitalization taking place, the manufacturing industry is collecting massive amount of data, most of which remains under-utilized. By leveraging these datasets, especially pertaining to steam generation and condensate recovery in the process plants, processes can be optimized to lower costs, improve quality, and enhance sustainability.
The **Condensate Recovery Factor (CRF)** stands as a pivotal metrics, reflecting the proportion of condensate effectively recovered and reused. In process plants, steam is converted to condensate after it has lost its latent heat of enthalpy in heat exchangers and thereafter recovered and returned to the boiler feedwater tank to be reused. Higher levels of CRF are preferred and strongly encouraged since it enhances plant efficiency. [1] Optimizing the CRF has far-reaching implications for energy efficiency, operational costs, and environmental impact, and, hence, it has become a priority for industries bent on improving its performance in an increasingly competitive and eco-conscious landscape.
Optimizing the CRF has far-reaching implications for energy efficiency, operational costs, and environmental impact, making it a priority for industries striving to enhance their performance in an increasingly competitive and eco-conscious landscape.

Despite its significance, accurate measurement and optimization of the CRF remains an area of challenge. Due to variations in steam flow rates, operational lags, and varying condensate recovery cycles, a multitude of challenges arise that prevent reliable estimation and improvement efforts. In this thesis, overcoming these challenges is discussed with the proposition of a robust framework for measurement, prediction and optimization of the CRF in process plants, focusing on data-driven approaches and advanced methodologies.

To extract meaningful insights from such large and complex datasets, a balance between automated analysis and human expertise must be fostered. While the algorithm may be able to identify trends with great efficiency, it is the domain knowledge of steam system engineers that creates the difference between identifying those patterns that are noteworthy or just noise.

While the CRF challenge is central to this research; however, its significance is further highlighted against the backdrop of Industry 4.0 and the ongoing digital transformation of the manufacturing milieu. Digitalization and Internet of Things (IoT) devices give prospects never seen before to capture, monitor, and analyze real-time data for smarter decision-making and operational efficiency enhancement.[2] These technologies are the ones that enable us to gather such intelligence, and in doing so, they present the capabilities to better deal with the complexities inherent in the CRF optimization problem.

1.2 Research Question and Sub-Questions

The main research question of this work is:

How can a data-driven approach optimize the Condensate Recovery Factor (CRF) in industrial plants while ensuring model interpretability and actionable insights for both AI experts and domain engineers?

To narrow down the research field of the thesis the following subquestions have been added:

1. *How can a machine learning model be designed for predicting CRF while ensuring its interpretability and transparency for non-expert stakeholders?*
2. *What key features and operational parameters significantly impact CRF, as identified through interpretable methods, ensuring alignment with industrial best practices and sustainability goals?*
3. *How can the insights from interpretable machine learning be leveraged to improve decision-making in condensate recovery and steam system efficiency?*

Based on the research questions, the goal of this work is twofold: first, to provide an overview of the current methodologies and advancements in the optimization of the Condensate Recovery Factor (CRF) for industrial plants, focusing on data-driven approaches such as machine learning models; and second, to design, develop, and evaluate a predictive model that optimizes CRF. This model will be trained on a dataset collected from multiple plants to accurately forecast CRF and identify key operational trends.

Additionally, to better understand the relationships between operational parameters and CRF, the neural network model will be analyzed using Explainable AI (XAI) methods. Such techniques will lead to model transparency, enhance interpretability, and so provide insight into the key factors affecting CRF. With the integration of XAI, this work ensures that the decision-making process is interpretable and therefore credible for the alternative optimization strategy in the industrial setting.

1.3 Outline

This research, conducted by Forbes Marshall India Pvt. Limited utilize real-world industrial data that have been obtained from multiple plants. This research acts as a bridge between theoretical advancements and their practical applications, demonstrating how data science and optimization tools can deliver real benefits to plant performance. This thesis thus addresses the CRF challenge and contributes to both the academic debate on resource efficiency and the practical world of industrial sustainability.
The following chapters build upon the basic concepts of Industry 4.0, digitalization of manufacturing processes, and, finally, implement innovative solutions toward optimizing the CRF. All these put together form a complete roadmap toward improving resource efficiency and operational excellence in process plants.

2 Preliminaries

This chapter discusses the basic theories of the concepts that were implemented in this thesis. These topics cover the most essential basics to be able to understand the main problem proposed in the thesis as well as the various approaches that were followed to address the problem.

Relevant concepts

2.1 Industry 4.0 and IoT

Industry 4.0 marks a shift in manufacturing by incorporating advanced technologies such as artificial intelligence, automation, and real-time data analytics into traditional systems. At the heart of this transformation is the Internet of Things (IoT). IoT refers to the network of interconnected devices and sensors that collect, exchange, and analyze data to facilitate its seamless exchange across the production environment. This interconnectedness enhances operational efficiency, reduces downtime, and promotes sustainability by optimizing resource use. [2]

In the context of Condensate Recovery Factor (CRF), Industry 4.0 and IoT play pivotal roles by:

- **Monitoring Steam and Condensate Flow:** Sensors provide real-time data on flow rates, pressures, and temperatures.
- **Analyzing Operational Data:** IoT systems enable predictive analytics, identifying inefficiencies or delays in condensate recovery.
- **Facilitating Automation:** Automated controls dynamically adjust system parameters to improve condensate recovery performance.

The integration of these technologies forms a foundation for our challenge of optimizing the Condensate Recovery Factor (CRF), ensuring a more efficient, intelligent, and environmentally responsible manufacturing process. [1]

2.2 Condensate Recovery Factor (CRF)

What is Condensate?

Steam is widely utilized in most process plants for heating and evaporation purposes. Once the steam is used, it releases its latent heat and transforms into condensate. This condensate, whether generated from a single application or multiple applications across various locations, is typically pumped at atmospheric pressure to a feed water tank in the boiler house, a condensate recovery header, or another suitable system. Recovering this condensate is crucial for optimizing overall efficiency and lowering the operational costs of the plant. [1]

Latent Heat vs. Sensible Heat

In industries that utilize steam, latent heat refers to the energy required to convert water into steam, also known as the enthalpy of vaporization. Water absorbs this energy to transform into steam, and when steam releases it, it returns to high-temperature water, referred to as condensate.
During the condensation process, at the exact point of phase change, the condensate retains the same temperature as the steam because only the latent heat has been removed, while all the sensible heat remains. This state is referred to as saturated water. Recovering and reusing this sensible heat, rather than letting it go to waste, is a key objective of condensate recovery efforts. [1]

When one kilogram of steam fully condenses, it produces an equivalent kilogram of condensate at the same pressure and temperature.

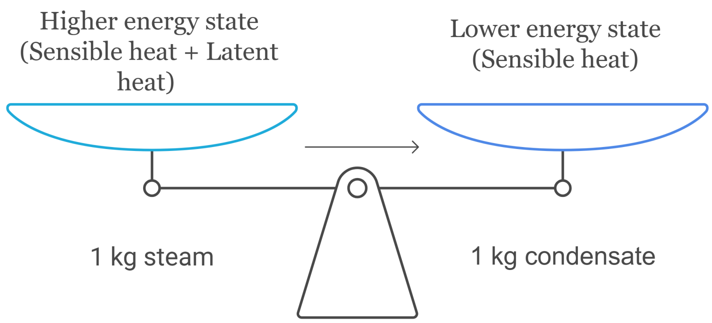

Figure 2.1: 1 kg of steam, when fully condensed, results in 1 kg of condensate.

What is Condensate Recovery?

When 1 ton per hour (t/h) of steam is supplied to equipment for heating, an equivalent amount of condensate i.e 1 ton per hour (t/h) is discharged during the process. Condensate recovery involves reusing the water and the sensible heat contained in this discharged condensate. Instead of discarding it, recovering condensate offers substantial benefits, including energy conservation, reduced chemical treatment costs and decreased demand for make-up water.

Mathematical Formula for CRF

The Condensate Recovery Factor (CRF) is defined as the ratio of the condensate recovered to the total steam supplied:

$$CRF = \frac{M_c}{M_s} \times 100\% \tag{1}$$

where:

- M_c = Mass of condensate recovered (kg/h),
- M_s = Mass of steam supplied (kg/h).

In an ideal system where all condensate is recovered, $CRF = 100\%$. However, in practical scenarios, due to system inefficiencies and losses, CRF is usually less than 100%.

Importance of Condensate Recovery

- **Economic Benefits:** Condensate is a valuable resource, and even small amounts are often worth recovering. The discharge from a single steam trap can lead to substantial savings when recovered. Without recovery, the condensate must be replaced by cold make-up water, which incurs additional costs for water treatment and fuel required to heat the water from a lower temperature.

- **Reduction in Water Charges:** Any unreturned condensate results in the need for make-up water, leading to increased water charges from local suppliers. Reducing water consumption also aligns with sustainable practices and lowers operational costs.

- **Maximizing Boiler Efficiency:** Colder feedwater in the boiler reduces its steaming rate, requiring more heat and fuel to raise the water temperature. This not only decreases boiler efficiency but also impacts the overall steam production.

- **Boiler Feedwater Quality:** Condensate is essentially distilled water, with very low total dissolved solids (TDS). Boilers need to be blown down periodically to manage TDS levels. Returning more condensate reduces the need for blowdown, minimizing energy loss.

2.3 Machine Learning (ML) and Explainable AI (XAI)

Supervised Learning Basics

Supervised learning involves training ML models to map input variables to output targets using labeled datasets. This paradigm is broadly categorized into **regression** and **classification**. In regression tasks, the objective is to predict continuous numerical values, such as the Condensate Recovery Factor (CRF), which quantifies condensate recovery efficiency as a percentage or mass flow rate (e.g., kg/hr). Classification, conversely, assigns discrete categorical labels to inputs (e.g., categorizing CRF as “High,” “Medium,” or “Low”).

For CRF prediction, regression is the appropriate framework, as the target variable is inherently continuous and requires precise numerical estimation for operational decision-making.
Central to regression is the use of **loss functions**, which quantify the discrepancy between predicted ($\hat{y}$) and observed (y) values. Two widely adopted loss functions are:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

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

where:

- MAE computes the average absolute deviation between predictions and ground truth. It is robust to outliers and aligns with industrial applications where sensor noise or transient anomalies may occur.
- RMSE penalizes larger errors quadratically, emphasizing accuracy in typical operating regimes but amplifying the impact of outliers.

For CRF prediction, MAE is favored due to its interpretability in absolute terms (e.g., ±5% error) and resilience to sporadic measurement faults common in industrial IoT systems. This choice ensures that models prioritize thermodynamic consistency (e.g., respecting CRF's physical bounds of 0–100%) while providing actionable insights for steam system optimization. The regression framework thus enables precise modeling of CRF dynamics, supporting granular operational adjustments such as blowdown cycle optimization or pump threshold tuning.

Deep Neural Networks

Neural networks containing multiple layers of differentiable functions are deep neural networks. These functions of multiple layers are the reason they are called deep. The neural networks that are relevant to the thesis are Feed Forward Neural Networks, which is explained below.

Feedforward neural networks are computational models designed to approximate complex relationships between input features and target outputs

through interconnected layers of neurons. The architecture begins with an **input layer** configured to accept a predefined number of features (e.g., sensor measurements), followed by **hidden layers** that transform inputs via weighted connections and non-linear activation functions. A distinguishing feature of this framework is the use of **parallel processing branches**, where distinct layers apply specialized transformations to the input data before merging their outputs.
This modular design enhances the network's ability to capture diverse patterns in industrial time-series data, such as cyclical trends or transient anomalies.

Activation Functions

Activation Functions form an important component of neural networks. They are the ones deciding whether the neuron should be activated or not which gets calculated by the weighted sum of inputs and adding bias to this value. To address domain-specific challenges, the network employs tailored activation functions:

- **Periodic Activations**: The sine activation function, defined as $f(x) = \sin(x)$, and the cosine activation function, $f(x) = \cos(x)$, are integrated into separate layers to model recurring patterns inherent to industrial processes. These functions decompose temporal signals into cyclical components, enabling the network to learn relationships such as:

$$y_t = A\sin(\omega t + \phi) + B\cos(\omega t + \phi)$$

  where A, B, ω, and ϕ are learned amplitudes, frequency, and phase shifts, respectively. This formulation captures periodic phenomena like daily operational cycles or machinery vibrations.

- **Non-Linear Activation**: A quadratic activation function, formulated as $f(x) = kx^2$ with $k = \textit{constant}$, introduces controlled non-linearity. The small constant k ensures minimal curvature, allowing the model to approximate relationships like:

$$y = kx^2 + bx + c$$

  without overwhelming linear or periodic terms. This is particularly effective for modeling weak non-linear dependencies (e.g., pressure-temperature hysteresis).

Genetic Algorithm for Hyperparameter Optimization

Genetic algorithms (GAs) are evolutionary optimization techniques used to automate the selection of hyperparameters—settings that govern model training and architecture. In this framework, GAs iteratively refine three critical hyperparameters:

- Batch size: Controls the number of samples processed before updating model weights.
- Number of periodic-activated units: Determines the capacity of layers dedicated to cyclical pattern detection.

Each candidate solution (individual) is encoded as a chromosome with three genes, corresponding to the hyperparameters above.

Evolutionary Operators

- **Crossover**: Combines genes from two parent chromosomes to generate offspring, preserving beneficial traits (e.g., high-performing batch sizes).
- **Mutation**: Introduces random adjustments to genes (e.g., altering the number of periodic units) to explore new regions of the hyperparameter space.
- **Selection**: Retains top-performing individuals based on validation loss, ensuring iterative population refinement.

Fitness Evaluation The fitness of each candidate is quantified using a composite score balancing three metrics:

$$\text{Fitness} = \alpha \times \text{Train Score} + \beta \times \text{Validation Score} + \gamma \times \text{Test Score}$$

where α, β, γ are weights prioritizing generalization performance. Validation loss (Validation Score) serves as the primary metric to mitigate overfitting, while the weighted aggregate ensures robustness across diverse operational scenarios. [3,4]

Explainable AI (XAI) for Model Transparency

Explainable AI (XAI) techniques bridge the gap between complex machine learning models and actionable human insights, ensuring transparency in critical industrial applications. [5]
Two pivotal methods that are used in this thesis are:

- **SHAP (SHapley Additive exPlanations)**: A game-theoretic method derived from cooperative game theory. SHAP assigns each feature x_i a Shapley value ϕ_i, quantifying its contribution to the difference between a model's prediction $f(x)$ and the baseline expectation $\mathbb{E}[f(x)]$:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{i\}) - f(S)]$$

  where F is the set of all features and S is a subset excluding i. SHAP values enable global interpretation of feature importance while maintaining consistency with physical principles. [6]

- **LIME (Local Interpretable Model-agnostic Explanations)**: Generates locally faithful explanations by approximating complex models with interpretable surrogates. For a prediction instance x', LIME solves:

$$g(x) = \arg\min_{g \in G} \mathcal{L}(f, g, \pi_{x'}) + \Omega(g)$$

  where $\mathcal{L}$ measures fidelity to the original model f, $\pi_{x'}$ is a proximity kernel, and $\Omega(g)$ penalizes surrogate model complexity. Linear surrogates ($g(x) = w^T x + b$) provide intuitive explanations for individual predictions. [7]

These methods transform opaque predictions into **interpretable, physics-aligned narratives**, enabling validation of model logic against domain knowledge (e.g., thermodynamic constraints) while maintaining operational relevance.

3 Broader Implications of Industry 4.0, IoT, and AI in Process Efficiency

3.1 The Dawn of Industry 4.0

Industry 4.0 marks the onset of the new industrial revolution, while introducing integration of the physical and digital dimensions into the production processes and economic capabilities. To understand its impact better, we can have a brief overlook at how the industrial revolutions have been evolving with time: the First Industrial Revolution, Industry 1.0 brought in steam power, which essentially changed the way people mechanised; the Second Industrial Revolution, Industry 2.0 revolutionised electric power through mass production; and the Third Industrial Revolution, Industry 3.0 brought in computers and automation that laid the foundation for digital systems.

Currently, Industry 4.0, the Fourth Industrial Revolution, builds on these bases by incorporating more complex technologies that include IoT, AI, and cyber-physical systems. This revolution enables communication and self-optimization of machines and systems, thus emerging with "smart factories," which is characterised with better efficiency and data processing in real-time, with an ability to manufacture adaption.

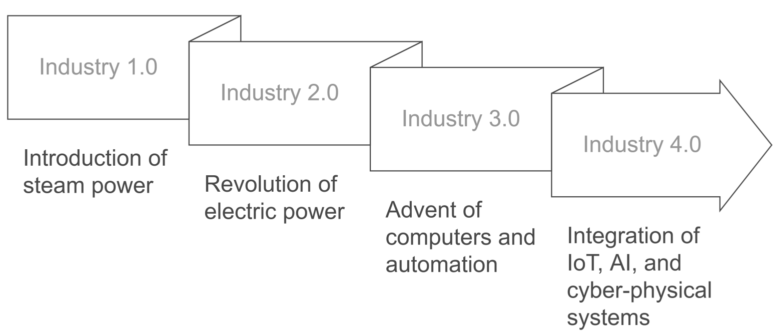

Figure 3.1: Evolution of Industrial Revolutions.

Industry 4.0, besides being an incremental advancement; it's also the revolution that combines machinery, human capital, and process operations into a highly responsive and adaptive system. As more companies begin to adopt this paradigm, the possibilities for increased efficiency, responsiveness, and cost-effectiveness sets a new standard, fundamentally reshaping the future of manufacturing and industry as a whole.

Building Blocks: Technologies that Power Industry 4.0

The transformative power of Industry 4.0 stems from a set of core technologies that are reshaping how businesses operate and compete. This set of fundamental technologies goes on to change the operation and competitive environment in which businesses operate - a fact that bestows a revolutionary potential upon Industry 4.0. Among them, the Internet of Things, artificial intelligence, cloud computing, and digital twins are the most representative. Each of these technologies offers particular functionalities, but together they create an integrated environment characterised by a data-centric industrial estate that maximises efficiency and flexibility.

Internet of Things connects all points across the production lines when it connects machines, sensors, and devices, thereby creating the real-time collection of a vast amount of data. These connectivities allow for predictive maintenance. IoT sensors can tell and alert a problem before its course turns into an expensive downtime.
Artificial Intelligence uses that information to monitor patterns, identify inefficiencies, and forecast needs. At the same time, the use of cloud computing—scalable solutions to store the data and computational capacity—enables AI to play around with these vast amounts of data across a geographic area to identify opportunities for optimization that might otherwise not be seen.

Together, these technologies form a self-sustaining loop of data gathering, analysis, and actionable intelligence that actually makes Industry 4.0 contexts not only more intelligent but also increasingly resilient and adaptive. The interaction between IoT, AI, cloud computing, and digital twins guides the promise of Industry 4.0 to drive an unprecedentedly efficient and responsive manufacturing system.

Challenges and Ethical Considerations

The advent of Industry 4.0 has its benefits. However, it also comes with notable drawbacks and contains critical ethical dilemmas. One of the most apparent challenges is the implementation costs. Transitioning to a

version that utilises high technologies such as IoT, Artificial intelligence, and automation necessitates adopting a costly infrastructure and acquiring more human resources, which can be very expensive especially for small to medium-sized enterprises. In addition to this, interconnectivity of systems also welcomes cybersecurity threats
Furthermore, ethical issues in this case arise because of the very nature of automation that can easily get rid of jobs that are basic and repetitive. Automating tasks that domain-skilled workers currently do will leave them without any employment, whereas those with technical capabilities will be in demand. However, it is not just the concern over automation; data privacy and responsible usage of artificial intelligence are also at the core as organisations have to draw a line between data collection and the right of people to privacy. All these calls for careful strategizing, appropriate policy formulation, and ethical use of technology.

The Road Ahead for Industry 4.0

The onset of Industry 4.0 is likely to revolutionise the present era of globalisation and industrialisation in the world by promoting efficiency, sustainable development and innovation. Industry 4.0 is about the incorporation of smart technologies in production processes by companies bearing in mind the need to eliminate waste, make better use of resources and cut emissions, paving the way for environmentally conscious operations.

It is evident this transition toward a smarter and greener industry is beneficial, however a course on responsible innovation must be in place. If managed correctly, Industry 4.0 has the potential to assist global industries in achieving more sustainable growth and development without purely focusing on the economy. The voyage of Industry 4.0 is on the genesis, with yet to go the full way of expectations for smart, flexible, and responsive industries all over the globe.

3.2 The Digital Shift in the Industry

With the aforementioned new industrial revolution , the face of industry has completely changed, for better; now, industry has evolved from analog systems to interconnected, data-driven ecosystems, powered by the influence of big data. In an earlier context, industry was not dominantly reliant on automation but saw it as part of manual processes with a good portion of human tamper, and independent, closed-loop production lines. Without real-time feedback, production lines were often isolated, operating independently and making optimization a challenging task. Today, the rise of digitalization has shifted this paradigm, where data is now at the heart of every operation, enabling smarter and more efficient workflows

A cornerstone of this transformation is the Internet of Things (IoT), enabling the real-time linkage and communications of devices and machines. IoT allows one to bring on board many data to analyse and optimise operations, predict maintenance needs, and improve decision-making processes. This transformation from traditional methods to digital solutions is greatly enhancing the operational efficiencies.

Internet of Things (IoT) Fundamentals: Building the Connected Ecosystem

IoT is one of the key enablers of the digital revolution, changing the very trajectory of industrial operations by integrating devices and providing a seamless flow of data. At its core, IoT is based on converting physical machinery, sensors, and equipment into interconnected nodes within the extensive digital network. In that regard, through real-time data gathering from different production lines, supply chains, and logistics - IoTs share and analyse data to optimise performance and decision-making.
Within the manufacturing settings, it monitors everything including but not limited to - machine performance to environmental conditions. Sensors placed on such equipment capture data and report this information to the central systems for analysis.

With a constant flow of data, businesses can detect abnormalities and immediately adjust to avoid disruptions. This brings with it invaluable predictability in the likelihood of equipment failure and subsequent downtime. Predictive analytics of IoT-based data can help to determine how much time was needed to initiate maintenance in advance of a breakdown, thus drastically

reducing unplanned downtime. Use of IoT in predictive maintenance has hence become one of the important strategies to gain operational efficiency and cost cutting.
To sum up, IoT transforms traditionally isolated systems into data-rich environments that enable organisations to work more efficiently, increase visibility, and enhance decision-making capabilities.

Data Power: Leveraging Analytics and AI

The true value of IoT data is unlocked through integration with artificial intelligence (AI) and advanced analytics. IoT devices are currently collecting real-time data from machines and production lines, but this raw data doesn't have all the insights for promoting operational excellence. AI and analytics combined can turn such kind of information into actionable intelligence-leading informed decision making, predictive maintenance, and operational optimization.

The role of AI becomes critical in analysing the huge amounts of collected data by the different IoT sensors. With machine learning algorithms, AI can analyse patterns and trends in the real time data, enabling systems to predict most issues even before they occur. For instance, in predictive maintenance, AI can note the slight and subtle changes such as temperature changes or unusual vibrations for an early detection of a malfunction in an equipment. Thus there would have been much less downtimes along with machinery utilities becoming more economical as downtime is reduced along with the losses caused by production interruptions. This is particularly significant in industries where downtime can be costly and detrimental to productivity.

In the long run, the combination of IoT, AI, and advanced analytics transform traditional approaches into proactive strategies that optimise performance, improves decision-making, and foster a more agile and efficient business environment.

3.3 Conclusion: The Future of IoT and Digitalization in Industry 4.0

As we look toward the future; innovation, efficiency, and sustainability across industries will still fundamentally be driven by IoT and digitization. These technologies are not just reshaping the process of business management in the present but are also laying the foundation for smarter and more adaptive industries in the future. By collecting and analyzing real-time data, IoT helps companies optimize production, improve customer experiences, and enhance supply chains. Digitalization, with its interconnected ecosystems, fosters collaboration and agility, which are crucial in today's fast-evolving markets. However, for IoT and digitalization to reach its full potential, challenges like data security, high implementation costs, and scalability needs to be addressed. Ensuring that IoT solutions are secure, affordable, and accessible to businesses of all sizes is critical for realizing their full impact.

Ultimately, IoT and digitalization will be indispensable for businesses striving to stay competitive, adaptable, and sustainable in the evolving landscape of Industry 4.0.

4 Dataset Description and Modeling Approach

To start off, a detailed dataset on steam generated and condensate recovered was needed for the time series forecasting of CRF. This was made available by Forbes Marshall's CRM (Condensate Recovery Meter), working along with its pressure-pumped package units across different locations. It assesses the total amount of steam pumped and condensate recovered over any given time period. The data obtained is stamped every minute and saved into a CSV file. This dataset served as the foundation for training the neural network model, allowing it to operate on realistic, current data obtained from different types of industries.
The list of data points collected for the analysis are mentioned below with their description:

Parameter	**Unit**	**Description**
Timestamp	minutely	Recorded in a minutely resolution, ensuring precise tracking of events and changes over time.
Steam Flow Rate	kg/h	Represents the amount of steam generated and utilized within the plant, a critical input for assessing condensate recovery.
Steam Total	kg	The cumulative steam flow across specific time periods, used to calculate efficiency metrics.
Total Condensate Recovered	kg	The volume of condensate collected from both direct and indirect steam consumers, crucial for calculating the CRF.
Condensate Recovery	kg	Reflects the contribution of condensate pumps in the system, allowing for detailed analysis of equipment-specific recovery.
Condensate Temperature	°C	Temperature measurements are necessary to assess thermal efficiency and potential heat losses in the condensate system.
Feed Water Temperature	°C	The temperature of water fed into the boiler, influencing steam generation and overall energy balance.
Blowdown TDS	kg	The concentration of dissolved solids in boiler blowdown, which impacts boiler performance and condensate quality.
Blowdown TDS Total	kg	The total mass of dissolved solids removed through blowdown.
Boiler ON/OFF Status	binary	This binary indicator helps differentiate between operational and non-operational states of the boiler, providing critical context for analyzing delays and dynamic behavior in the system.

Table 4.1: Key Data Points Collected for Analysis

4.1 Data Preprocessing and Feature Engineering

Condensate Recovery Factor or CRF is regulated by thermodynamic laws where steam flow dynamics, efficiency in recovering condensate, and boiler operational settings adheres to the conservation laws of energy and mass. In this systems, excessive preprocessing risks obscuring critical physical relationships, such as brief steam usage profiles or phase-change-driven condensate yield, that underpin CRF variability.[1] This philosophy aligns with physics-based modelling paradigms, where raw data fidelity retains mechanistic interpretability unadulterated by artificial distortions imposed by aggressive transformations.[8]

In Industrial IoT products such as Forbes Marshall's pressure-driven pump units and Condensate Recovery Meters (CRM); minutely telemetry of steam flow rates, condensate temperatures, and feedwater dynamics provides relevance information concerning system behavior. This is important in deciding anomalies including steam trap malfunction or boiler cycling inefficiencies.

Although the minutely recorded IoT data was largely clean, occasional sensor faults or communication failures resulted in missing values. These were addressed using appropriate imputation techniques such as forward-filling to maintain data continuity while ensuring the system's overall behaviour was not artificially modified. Since the raw readings were already structured and inherently meaningful, no additional feature engineering was required, except for Min-max scaling (it transforms data to a fixed range to preserves the original distribution while compressing extreme values) which was done purely to ease neural network training due to features like Steam Total (kg) and Total Condensate Recovered(kg) being in the range of 10^6. These large input values can lead to numerical instability in neural networks, causing issues like vanishing or exploding gradients. [9]
By leveraging the IoT data as close to its original form as possible, the predicting model remains aligned with the actual operational conditions of steam and condensate recovery, ensuring both accuracy and interpretability in optimizing CRF.

4.2 Neural Network Model Architecture

Predicting nonlinear trends in time-series data is a long-standing challenge with significant applications across industrial process optimization, finance, and control systems. Although traditional neural networks, according to their universal function approximation properties [10], can model any function, they frequently struggle at extrapolation and, rather, fit the data well yet mis-generalize beyond the observed patterns. Fourier-based approaches have been tried to circumvent this constraint, using sinusoidal activation functions to capture periodic trends in time-series forecasting [11,12,13,14]. However, ensuring training stability and preventing overfitting are key obstacles.

An essential element often overlooked in time-series modeling is the explicit treatment of time as a feature. Many of the traditional methods consider past observations alone to predict future behavior, making them susceptible to the influence of short-term fluctuation and noise [15,16]. Instead, incorporating time directly into the model and using it as a separate feature allows the better handling of periodicity and long-term trend forecasts, especially in such nonlinear dynamical systems as industrial steam management.

To overcome these challenges, we propose a Fourier-based Feedforward Neural Network (FNN) optimized by a Genetic Algorithm (GA) for the selection of hyperparameters.[11,4] The FNN is an effective function approximator, utilizing Fourier features to enhance extrapolation behavior. Meanwhile, GA effectively searches the space of hyperparameters, optimizing network structure and regularization parameters to improve forecast accuracy. This hybrid model approach balances precision and computation efficiency, which renders it especially suited for optimizing condensate recovery factors (CRF) in process plants.

By combining domain-specific insights with deep learning methods, this model is designed to enhance forecasting accuracy, better interpretability of system behavior, and improved robustness against process condition variations.

4.3 Model Implementation

A feedforward neural network was chosen for time series prediction due to its ability to learn complex periodic and nonlinear patterns from historical data. The architecture consisted of:

- **Input Layer:** Accepting multiple time series features, including time-based inputs and process variables.
- **Hidden Layers:** Three parallel fully connected layers, each with different activation functions to capture distinct data characteristics:
  - **Sine Activation:** Capturing periodic trends in the dataset.
  - **Cosine Activation:** Complementing sine activation for better Fourier representation.
  - **Quadratic Activation:** Modeling polynomial trends for non-periodic variations.
- **Concatenation Layer:** Merging outputs from the three hidden layers to integrate periodic and polynomial features.
- **Output Layer:** A final fully connected layer with a linear activation function to predict the next time step's Condensate Recovery Factor (CRF).

This hybrid approach enhances the model's ability to generalize across periodic and nonlinear trends, improving time-series forecasting accuracy.

4.4 Model Training and Evaluation

The dataset was divided into training, validation, and test sets based on temporal partitions to prevent data leakage. After the model was trained, its performance was evaluated using the following metrics:

- **Mean Absolute Error (MAE)**: Measures the average absolute difference between actual and predicted values:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

- **Root Mean Squared Error (RMSE)**: Captures the magnitude of prediction errors, penalizing larger deviations:

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

- **Mean Absolute Percentage Error (MAPE)**: Assesses relative prediction accuracy in percentage terms:

$$\mathrm{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

- **RMS Percentage Error (RMSPE)**: Provides insight into the model's consistency across different scales:

$$\mathrm{RMSPE} = \sqrt{\frac{100\%}{N}\sum_{i=1}^{N}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2}$$

4.5 Interpretability and Explainability

While the Fourier-inspired neural network (FNN) inherently provides interpretability through its additive structure, to enhance the model's transparency even further, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) were utilized. [6,7] SHAP, based on game theory, assigns each feature a contribution value, providing a global and local understanding of the model's predictions. LIME, on the other hand, builds locally interpretable models around individual predictions by perturbing inputs and observing their effects. These methods bridge the gap between the FNN's mathematical transparency and actionable operational guidance, ensuring stakeholders understand why and how the model makes predictions.

5 Neural Network Design and Evolutionary Optimization

The methods we implement include a neural network, inspired by Fourier analysis, and a genetic algorithm to search for optimal hyperparameters for the network. One key advantage of this approach is that it only requires the time variable to predict new values. Despite being a neural network, it is not a black box, as it contains only one hidden layer with simple activation functions. This simplicity allows for an easily extractable mathematical model from the network, as will be explained below.
Fourier Analysis decomposes a time series into a sum of sine and cosine functions, with different amplitudes and frequencies, revealing the components that make up the signal.

$$f(t_k) = c + \sum_{i=1}^{n-1} a_i \sin\left(\frac{2\pi i t_k}{T}\right) + b_i \cos\left(\frac{2\pi i t_k}{T}\right) \tag{1}$$

Here, n represents the number of points in f, and t_k denotes the discrete time variable corresponding to point k in the series. The total duration of the series is given by $T = t_{\max} - t_0$, where $t_{\max}$ and t_0 are the maximum and initial time points, respectively. The term c is a bias correction constant, which can also be interpreted as the $i = 0$ term in the summation.
The coefficients a_i and b_i are determined such that the summation of sine and cosine functions accurately reconstructs each point in the original time series. However, despite the perfect accuracy of this function for known points, it is not well-suited for forecasting future values. For times $t > t_{\max}$, the function simply repeats the values from t_0 to $t_{\max}$ periodically. This limitation arises because the summation consists solely of periodic functions, which restrict the representation to frequencies that do not exceed the fundamental period or domain of the original time series.
Nevertheless, many time series exhibit clear periodicity. To address this, we relaxed the rigidity of the original Fourier decomposition and allowed a neural network to learn the natural frequencies of the time series, including those related to arbitrarily long time periods, which is very useful for future predictions.
For instance, a time series used to train the network might cover a period of two months, but this doesn't prevent the network from learning patterns

and trends that could last for four months or even a year, in addition to the patterns already present in the series.
The architecture of the neural network used is shown in the following picture.

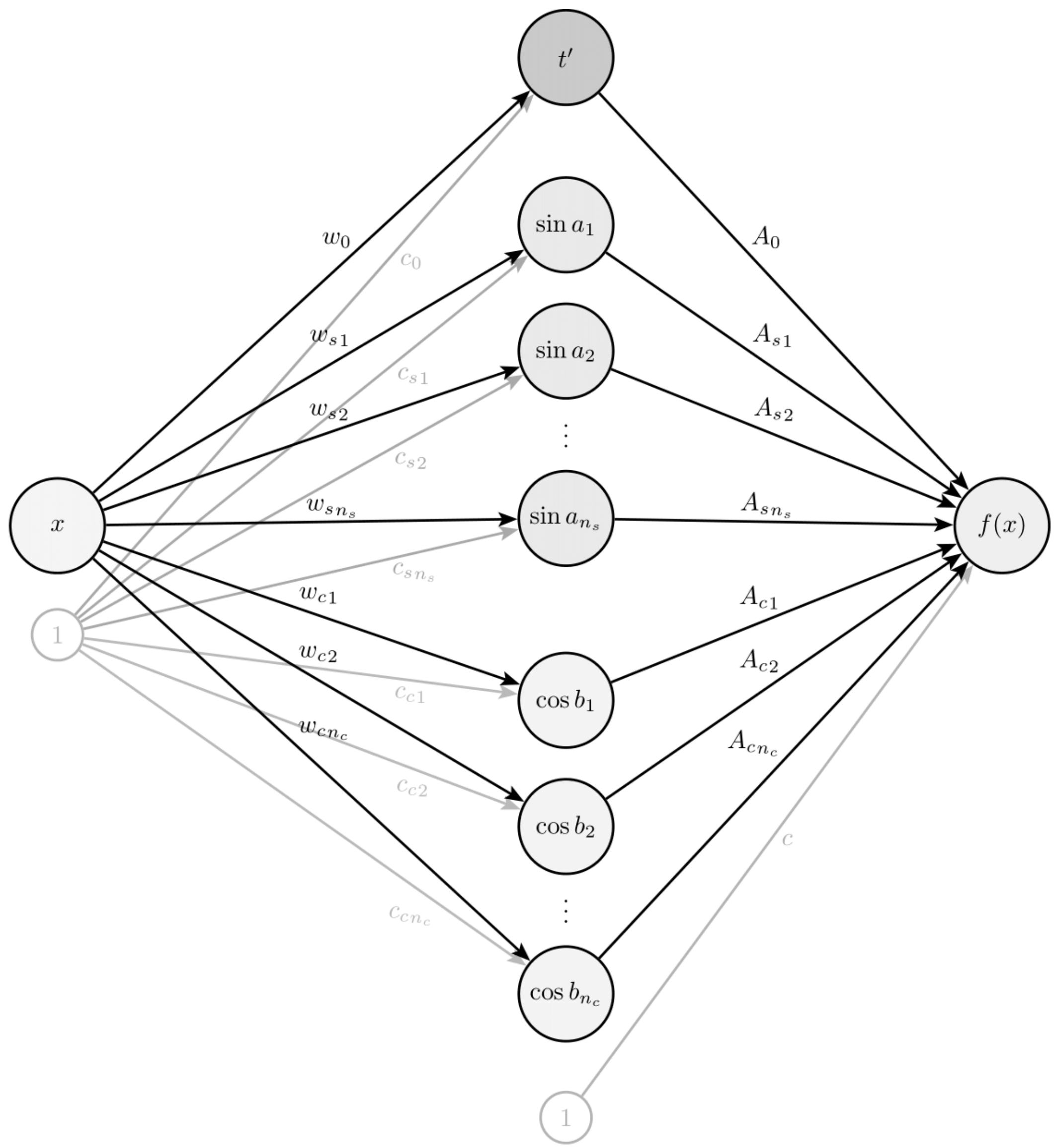

Figure 5.1: Fourier Neural Network (FNN) architecture with periodic and non-linear activation functions.

Here, t is the time-related value of a point in the series, so t values are only a linear succession of numbers (like [0,1,2,3 ...] or [0,2,4,6 ...], $f(t)$ is the net output, the value that is desired to match with the corresponding of the column to predict and:

$$a_i = w_{s_i} t + c_{s_i} \tag{2}$$
$$b_i = w_{c_i} t + c_{c_i} \tag{3}$$
$$t' = (w_0 t + c_0)^2 \tag{4}$$

After training the network to approximate $f(t)$ to the series to predict, we obtain the learned coefficients w's, c's, and A's. We can then make predictions for t greater than its maximum value given in the training. Mathematically, $f(t)$ is expressed as:

$$f(t) = c + A_0 t' + \sum_{i=1}^{n_s} A_{s_i} \sin a_i + \sum_{i=1}^{n_c} A_{c_i} \cos b_i \tag{5}$$

A quadratic term t' is included in the network because a linear or a very smooth quadratic long-term trend is observed in most of the series. It could be different from quadratic, but since the forecasts are not far from this trend, this approximation is sufficient.

5.1 Neural Network Architecture

5.1.1 Input Layer

- **Time-Only Model**: Scalar $t \in \mathbb{R}$
- **Time + Exogenous Model**: Augmented vector $\mathbf{z} = [t, \mathbf{x}] \in \mathbb{R}^{1+m}$, where $\mathbf{x} \in \mathbb{R}^m$ are exogenous features

5.1.2 Hidden Layers

1. **Sine and Cosine Terms**

   Each node computes :-

   $$\mathcal{S}_i(t, \mathbf{x}) = A_{s_i} \sin\left(\mathbf{w}_{s_i}^T \mathbf{z} + c_{s_i}\right), \quad \mathcal{C}_j(t, \mathbf{x}) = A_{c_j} \cos\left(\mathbf{w}_{c_j}^T \mathbf{z} + c_{c_j}\right)$$

   here,

- $\mathbf{w}_{s_i}, \mathbf{w}_{c_j} \in \mathbb{R}^{1+m}$: Frequency weights
- $c_{s_i}, c_{c_j} \in \mathbb{R}$: Phase biases
- $A_{s_i}, A_{c_j} \in \mathbb{R}$: Amplitudes

2. **Quadratic Trend Term**

$$\mathcal{Q}(t, \mathbf{x}) = A_0 \left(\mathbf{w}_0^T \mathbf{z} + c_0\right)^2$$

here, $\mathbf{w}_0 \in \mathbb{R}^{1+m}$ and $c_0 \in \mathbb{R}$.

5.1.3 Output Layer

$$y(t, \mathbf{x}) = c + \mathcal{Q}(t, \mathbf{x}) + \sum_{i=1}^{n_s} \mathcal{S}_i(t, \mathbf{x}) + \sum_{j=1}^{n_c} \mathcal{C}_j(t, \mathbf{x})$$

5.2 Time-Only Model (t)

For univariate input, the predicted CRF is:

$$CRF(t) = c + A_0 \cdot (w_0 t + c_0)^2 + \sum_{i=1}^{n_s} A_{s_i} \sin(a_i) + \sum_{j=1}^{n_c} A_{c_j} \cos(b_j), \tag{6}$$

$$\begin{aligned} a_i &= w_{s_i} t + c_{s_i}, \quad \forall i \in \{1, \ldots, n_s\}, \\ b_j &= w_{c_j} t + c_{c_j}, \quad \forall j \in \{1, \ldots, n_c\}, \\ t' &= (w_0 t + c_0)^2 . \end{aligned}$$

5.3 Time + Exogenous Variables $(t, \mathbf{X}_{\mathbf{exog}})$ Model

For multivariate inputs, let $\mathbf{X}_{\text{exog}} \in \mathbb{R}^m$ denote exogenous features. Define $\mathbf{z} = [t, \mathbf{X}_{\text{exog}}] \in \mathbb{R}^{1+m}$. Thus, the predicted CRF is:

$$CRF(t, \mathbf{X}_{\text{exog}}) = c + A_0 \cdot \left(\mathbf{w}_0^T \mathbf{z} + c_0\right)^2 + \sum_{i=1}^{n_s} A_{s_i} \sin\left(a_i\right) + \sum_{j=1}^{n_c} A_{c_j} \cos\left(b_j\right), \tag{7}$$

$$\begin{aligned} a_i &= \mathbf{w}_{s_i}^T \mathbf{z} + c_{s_i}, \quad \forall i \in \{1, \ldots, n_s\}, \\ b_j &= \mathbf{w}_{c_j}^T \mathbf{z} + c_{c_j}, \quad \forall j \in \{1, \ldots, n_c\}, \\ t' &= \left(\mathbf{w}_0^T \mathbf{z} + c_0\right)^2. \end{aligned}$$

5.4 Training Objective

The model minimizes the Mean Absolute Error (MAE):

$$\mathcal{L} = \frac{1}{N} \sum_{k=1}^{N} \left| f(t_k, \mathbf{X}_{\text{exog},k}) - \text{CRF}_{\text{true},k} \right|$$

where:

- $\mathcal{L}$: MAE loss

- $$f(\cdot) = \text{Dense}_{\text{linear}} \left(\text{Concatenate} \begin{bmatrix} \text{Dense}_{\text{sin}}(\mathbf{x}) \\ \text{Dense}_{\text{cos}}(\mathbf{x}) \\ \text{Dense}_{\text{quadratic}}(\mathbf{x}) \end{bmatrix} \right)$$

5.5 Variables Summary

Table 5.1: Symbol Definitions and summary of the variables

Symbol	Description	Notes
t	Time variable	- Scalar temporal feature - Dimension: $\mathbb{R}$
$\mathbf{X}_{\text{exog}}$	Exogenous feature vector	- Contains m external predictors - Dimension: $\mathbb{R}^m$
$\mathbf{z}$	Augmented input vector	- Combined $[t, \mathbf{X}_{\text{exog}}]$ - Dimension: $\mathbb{R}^{1+m}$
$\mathbf{w}_0$	Quadratic term weights	- Linear coefficients for constant term - Dimension: $\mathbb{R}^{1+m}$
$\mathbf{w}_{s,t}, \mathbf{w}_{c,j}$	Sine/cosine weights	- Frequency parameters per node - Dimension: $\mathbb{R}^{1+m}$
$c_0, c_{s,t}, c_{c,j}$	Phase biases	- Temporal offsets - Dimension: $\mathbb{R}$
$A_0, A_{s,t}, A_{c,j}$	Amplitudes	- Output scaling coefficients - Dimension: $\mathbb{R}$
n_s, n_c	Node counts	- Sine/cosine units in architecture - Dimension: $\mathbb{Z}^+$

5.6 Genetic Algorithm for Hyperparameter Optimization

Once the neural network architecture is finalized, the critical task is to identify optimal hyperparameters: the number of sine nodes (n_s), cosine nodes (n_c), and batch size (b).

The optimization of neural network hyperparameters represents a critical challenge in machine learning, particularly when the relationship between hyperparameters and model performance is nonlinear, discontinuous, or noisy. Traditional grid or random search methods, while straightforward, often prove computationally prohibitive and fail to exploit the latent structure of promising hyperparameter combinations. To address this, we employ a Genetic Algorithm (GA) that iteratively evolves hyperparameter configurations by balancing exploration of the search space and exploitation of high-performing candidates. The algorithm prioritizes minimizing the Mean Absolute Error (MAE) on validation data while explicitly guarding against overfitting through a weighted evaluation metric. Below, we formalize the steps of the GA.

5.7 Algorithmic Formulation

5.7.1 Hyperparameter Representation

Each candidate model is defined by a hyperparameter vector:

$$\theta = \begin{pmatrix} n_s \\ n_c \\ b \end{pmatrix} \in \Theta \subseteq \mathbb{N}^3,$$

where Θ is the bounded search space:

$$n_s \in [n_{s,\min}, n_{s,\max}], \quad n_c \in [n_{c,\min}, n_{c,\max}], \quad b \in [b_{\min}, b_{\max}].$$

5.7.2 Algorithm Workflow

The GA proceeds as follows:

1. **Initial Population**: Generate $N = 50$ models with uniformly sampled hyperparameters:

$$P_0 = \left\{\theta^{(i)}\right\}_{i=1}^{50}, \quad \theta^{(i)} \sim \mathcal{U}_{int}(\Theta).$$

2. **Parent Selection & Crossover**: Select $M = 25$ parent pairs. For each pair $(\theta^{(p_1)}, \theta^{(p_2)})$, generate a descendant $\theta^{(d)}$ and for each hyperparameter j, choose the crossover method via:

$$\beta_j \sim \text{Bernoulli}(0.5), \quad \text{where } \beta_j \in \{0, 1\}.$$

$$\theta^{(d)} = \begin{cases} \alpha_j \theta_j^{(p_1)} + (1 - \alpha_j)\theta_j^{(p_2)}, & \text{if } \beta_j = 1 \quad \text{(coin-toss inheritance)} \\ \left\lfloor \frac{\theta_j^{(p_1)} + \theta_j^{(p_2)}}{2} \right\rfloor, & \text{if } \beta_j = 0 \quad \text{(arithmetic mean inheritance)} \end{cases}$$

here, $\alpha_j \sim \text{Bernoulli}(0.5)$ governs the coin-toss inheritance for hyperparameter j.

3. **Mutation**: Perturb each descendant hyperparameter with probability $p_{\text{mut}} = 0.05$:

$$\theta_{\text{mut},j}^{(d)} \sim \begin{cases} \mathcal{U}_{int}(\Theta_j), & \text{with probaibility } 0.05, \\ \theta_j^{(d)}, & \text{otherwise,} \end{cases},$$

4. **Survival Selection**: Retain the top 50 models from the combined population $(P_t \cup D_t)$ based on validation MAE:

$$P_{t+1} = \underset{\theta \in P_t \cup D_t}{\arg \text{ top-50}} \left(-\text{MAE}_{\text{val}}(\theta) \right).$$

5. **Termination**: Repeat for $T = 10$ generations.

5.7.3 Final Model Selection

The final population P_T is evaluated using:

$$E(\theta) = 0.25 \cdot \text{MAE}_{\text{train}} + 0.25 \cdot \text{MAE}_{\text{val}} + 0.5 \cdot \text{MAE}_{\text{test}}. \tag{8}$$

To ensure robustness against overfitting, the optimal model θ^* is selected using a weighted evaluation metric E that emphasizes generalization to unseen data.

The optimal model is:

$$\theta^* = \underset{\theta \in P_T}{\arg \min} \, E(\theta). \tag{9}$$

Here, MAE_{test} is computed on a held-out test dataset representing the most recent segment of the time series. The weights reflect a design priority: test performance is twice as critical as training or validation performance.

5.7.4 Algorithm Advantages

1. **Adaptive Search**: The GA dynamically balances exploration (via mutation and stochastic crossover) and exploitation (via arithmetic averaging and elitist selection).

2. **Overfitting Mitigation**: The composite metric 'E' prioritizes test performance, discouraging over-optimization on validation data..

3. **Scalability**: Parallelizable training of candidate models reduces wall-clock runtime.

5.8 Variables Summary

Table 5.2: Summary of Genetic Algorithm Variables

Symbol	Description	Domain/Type	Notes
θ	Hyperparameter vector	$\mathbb{N}^3$	$\theta = \begin{pmatrix} n_s \\ n_c \\ b \end{pmatrix}$
n_s	Sine nodes	$[n_{s,\min}, n_{s,\max}]$	User-defined bounds
n_c	Cosine nodes	$[n_{c,\min}, n_{c,\max}]$	User-defined bounds
b	Batch size	$[b_{\min}, b_{\max}]$	User-defined bounds
P_t	Population at generation t	Set of θ	$\|P_t\| = 50$
T	Total generations	$\mathbb{N}$	$T = 10$
p_{mut}	Mutation probability	$[0, 1]$	$p_{\text{mut}} = 0.05$
$E(\theta)$	Weighted metric	$\mathbb{R}^+$	Prioritizes test MAE

6 Results and Discussion

In this chapter, the final stage of the thesis is outlined: testing and validation of the predictive model developed for condensate recovery factor (CRF). This chapter includes the major conclusions of the thesis and the rationale behind the methodologies used and the conclusions from the results attained. The chapter is segmented into sections according to the primary model that was used in CRF optimization.

6.1 Model Selection Process & Challenges

The process of selecting the most appropriate model to predict the Condensate Recovery Factor (CRF) involved several iterations, where different architectures were tested before settling on the chosen Fourier Neural Network (FNN). Each approach had its own strengths. However, they were ultimately found to be insufficient in properly capturing the unique temporal dynamics involved with the CRF.

Initial Approaches and Challenges

6.1.1 Statistical and Kernel-Based Methods (ARIMA, SARIMA & SVR)

ARIMA (AutoRegressive Integrated Moving Average) and its seasonally adjusted variant SARIMA (Box & Jenkins, 1976) were tested because of their strong performance in classical time series forecasting challenges. [17] These models seem to work very well with stationary data having clear trends and seasonality. However, CRF's continuously increasing nature caused the aforementioned models to perform poorly, as differencing and season adjustment couldn't accurately simulate its accumulation tendency. This obstacle is documented in Wang et al., 2019; *Why Are the ARIMA and SARIMA Not Sufficient*. [18]

Similarly, Support Vector Regression (SVR) can also capture nonlinearity in data through kernel mappings, but it also lacks the flexibility to capture extended long-term trends, as noted by Smola and Schölkopf (2004). [19]

6.1.2 Recurrent Neural Networks (RNNs) & Long Short-Term Memory (LSTM) Networks)

Although RNNs and LSTMs are intended to manage sequential dependencies, their performance on continuously increasing series like CRF is often affected by issues like vanishing gradients over extended sequences. This challenge is well-documented in (Hochreiter & Schmidhuber, 1997; Pascanu et al., 2013). [20,21] Our experiments confirmed that these models struggles to maintain long-term dependencies, which is critical for accurately forecasting CRF trends. To add, the high computational costs and sensitivity to hyperparameters rendered them impracticle for our industrial application.

6.2 Adoption of Fourier inspired Neural Network (FNN) as the Final Choice

Due to the aforementioned constraints, we eventually chose to proceed with the Fourier Neural Network (FNN) approach (Mingo et al, 2004). [11] The FNN complements time series forecasting by spectrally decomposing it into its frequency components to achieve periodicity and long-term trends. This approach not only provides better accuracy in forecasting but also increased interpretability for domain experts to appreciate the frequency-based drivers of CRF without the need for computational complexity as in the case of complex repeating patterns.

In short, the transition to FNNs provided us with the model that is accurate as well as computationally efficient, hence suitable for industrial real-time application.

Model Performance Evaluation

Model performance evaluation is essential in establishing the accuracy and effectiveness of predictive models. Known statistical metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Percentage Error (RMSPE), provides insights into the model's ability to learn patterns from data. Models that are evaluated well gains trust and, hence, are reliable in decision-making based on their predictions.

We trained the Time-Only model (t) for 500 epochs which took about 14 hours and the Time + Exogenous Variables (t, X_{exog}) model for 100 epochs which took about 12 hours in the 13th Gen Intel Core i5-1335U (12-core) processor with 16 GB RAM. The latter approach while fostering improved decision-making and trust, it came with an increased computational cost. This trade-off highlighted the need to assess whether the additional complexity is justified, especially in resource-constrained environments.
The findings from both these evaluation is shown in Figure 6.1 below.

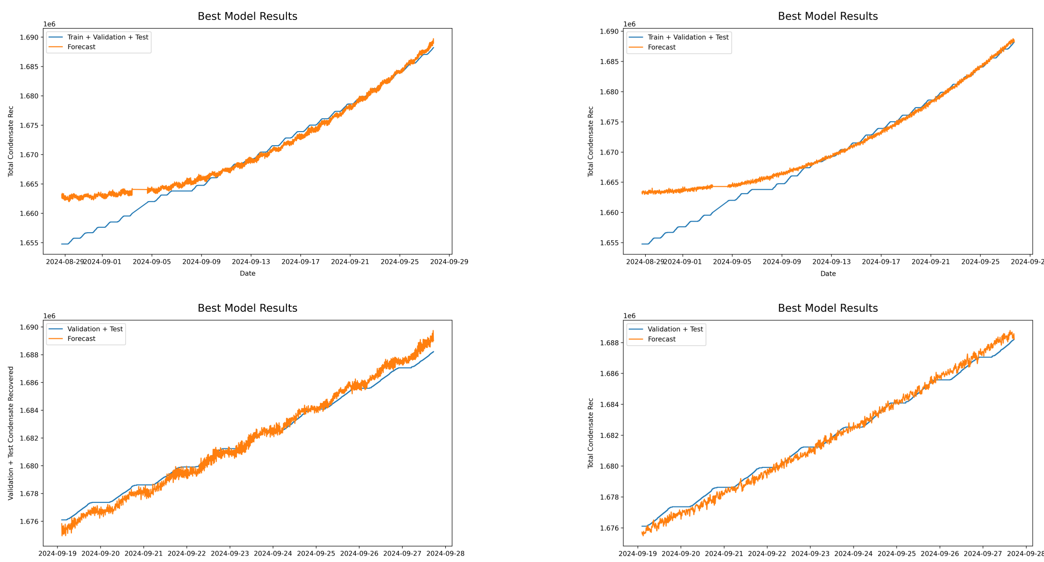

Figure 6.1: Forecasted vs. Actual Total Condensate Recovery (CRF) of Time-Only model (t) (left) and Time + Exogenous Variables (t, X_{exog}) model.

(The two graphs display the best model results for forecasting Total Condensate Recovery (CRF). In both, the forecast (orange) closely matches the actual data (blue). The top graph includes train, validation, and test data, while the second focuses only on the validation and test data. Both model's graph shows that the model accurately captures the CRF trend, demonstrating strong predictive performance across all datasets.)

Model	MAE	RMSE	MAPE (%)	RMSPE (%)
Time-Only (t) (500 epochs)	429.5146	510.7662	0.0257	0.0306
Time + Exogenous (t, X_{exog}) (100 epochs)	323.0992	390.4674	0.0190	0.0230

Table 6.1: Performance comparison of models with and without exogenous variables.

Performance Insights

The results indicate that the inclusion of exogenous variables significantly improves forecast accuracy. The Time + Exogenous (t, X_{exog}) model performs better than the Time-Only (t) model on all measures, lowering the MAE by 24.7% and RMSE by 23.5%, despite being trained for just 100 epochs versus the 500 epochs for the time-only model. This demonstrates the forecasting capability of external factors in optimizing CRF forecasting, and suggests that the additional tuning with domain-specific exogenous features could enhance model performance even further.

Accuracy vs. Computational Cost

The comparison between the two models highlights a vital trade-off in forecasting CRF: accuracy against computational cost. Although adding exogenous features leads to significantly enhanced predictive accuracy, it is done at the expense of additional resource. The inclusion of exogenous variables requires additional processing due to Genetic Algorithm (GA) tuning increases the computational complexities.

Despite training for only 100 epochs, the Time + Exogenous (t, X_{exog}) model achieves superior accuracy, indicating the strong impact of external factors on CRF dynamics. This trade-off indicates that while the inclusion of exogenous features is desirable for accuracy, deployability, however, is a function of computational resources available. Optimization of the GA process or the choice of most significant exogenous features would be a potential means of balancing accuracy and efficiency.

Model Advantages

- The model's design primarily depends on dynamics in time, eliminating the obligatory presence of exogenous factors. While incorporating exogenous factors can enhance performance in circumstances where previous observations are sparse, they are not strictly required for making accurate predictions.

- The chosen network structure facilitates the extraction of a mathematical representation of the series, effectively covering its built-in periodicities and its natural frequency elements.

- In contrast to complex architectures such as LSTMs, CNNs, or Transformers, this model is more interpretable and computationally efficient alternative while maintaining competitive predictive capabilities.
- The neural network architecture of the model makes it highly flexible. It can be easily integrated with other functions or reinforced with additional hidden layers, with a fine balance between complexity and explainability.
- Its applicability extends beyond specific datasets, making it a versatile tool for the analysis of a wide range of time series, regardless of domain.

Model Disadvantages

- The interpretability of the model's predictions is closely linked to their breakdown into natural frequency components, which may not always fit within specific explanatory frameworks relevant to the domain.
- The model struggles to anticipate infrequent, abrupt changes in the time series, such as sudden spikes or unpredictable disruptions.

6.3 Explainable AI Insights

This section discusses the results of the interpretability experiments conducted in relation to this work.

LIME (Local Interpretable Model-Agnostic Explanations)

LIME can be critical in explaining individual CRF predictions, particularly in instances where domain experts want to understand the fluctuations. LIME works by approximating the decision boundary of the trained Fourier Neural Network by fitting an interpretable, linear surrogate model to small perturbations around the instance in question.
This approach allows for a deeper analysis of the factors that affect CRF variations. The LIME results provides explanations tailored to each instance, emphasizing the variables that were linked with the observed fluctuations in CRF. These explanations can greatly aid in understanding the predictions, which is valuable for evaluating AI-driven CRF modeling.

To examine the model's behavior on a case-by-case basis, random data points were chosen, and LIME was utilized to create local explanations. The following plots in the Figure 6.2 demonstrate how different parameters contributed to the predicted CRF values, shedding light on the factors that influenced each prediction.

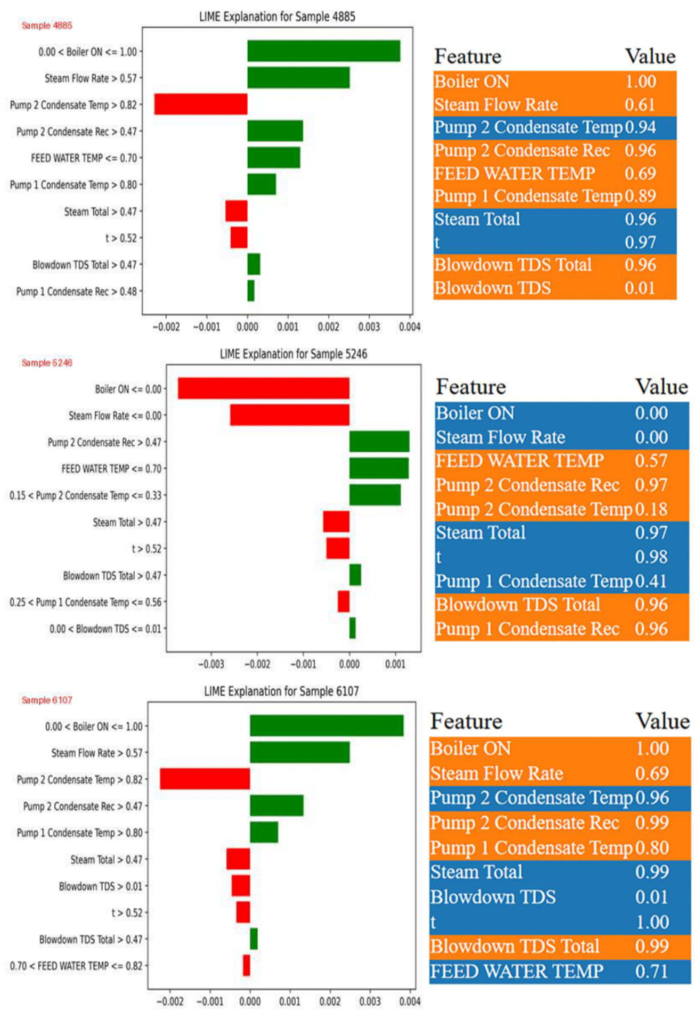

Figure 6.2: LIME Explanation for CRF Prediction

Each plots shows how various parameters either positively (in green/orange) or negatively (in red/blue) influence the predicted CRF value, shedding light on the model's decision-making process. The numeric values alongside the

features represent their respective contributions or presence in the analyzed instance.
The local explanation analysis gave valuable information on the behaviour of the Condensate Recovery Factor (CRF). When the boiler was ON it had notably positive effect, while OFF was negatively influencing CRF due to insufficient generation of steam. A high steam flow rate was desirable, but a low or zero flow rate negatively impacted CRF, reflecting limited steam availability for condensate recovery. The condensate temperature from Pump 2 showed a negative effect at higher levels, which might indicate inefficiencies in heat transfer or losses in the system.
On the other hand, the recovery of the condensate from Pump 2 has been proved to be beneficial to CRF, highlighting the need for effective condensate return. There was a mild positive effect from the feedwater temperature, likely because it helps to mitigate thermal shock in the boiler system. By contrast, the total dissolved solids (TDS) of blowdown and the overall blowdown losses usually stand for a negative impact, as excessive blowdown means steam loss and diminished recovery efficiency.

Furthermore, the positive contributions made by the total steam values reinforces the idea that more steam in the system leads to enhanced recovery.

These critical findings can help engineers and other stakeholders in analyzing predictions with greater nuance and understanding how various conditions affects the CRF estimates. This insight derived from these statements can prove beneficial in the assessment of model outputs vis-a-vis identifying critical operational factors affecting condensate recovery.

SHAP (SHapley Additive exPlanations)

While LIME assists the explanation of individual predictions, it is equally important to consider the general behaviors of the model for different types of test instances. This is where SHAP (SHapley Additive Explanations) comes in use, it provides a broader view on how much each feature is driving the model's predictions across the range of test samples.

To accomplish this global interpretability, we have investigated the SHAP summary plot, violin plot, and dependence plots. When taken together, these can provide insights about feature importance, distribution of effects, and interaction between variables.

The SHAP summary plot collects and ranks the contributions of features from all test samples according to their importance. Each dot indicates a SHAP value, and the value of the feature being considered is represented on a color gradient with red for high and blue for low. The SHAP layered violin plot complements this by showing the distribution of SHAP values for each feature, offering insights into the consistency and variability of their influence on predictions.

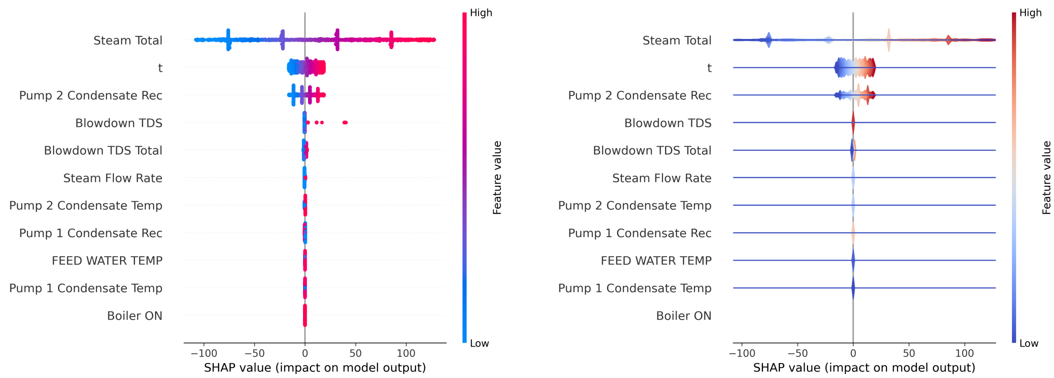

Figure 6.3: Global interpretability of feature importance using SHAP
(The SHAP summary plot (left) ranks features by their impact, with a color-coded distribution that indicates whether feature values are high (red) or low (blue). The layered violin plot (right) provides a further visualization of the spread and density of SHAP values, emphasizing the variability in feature influence across test samples.)

From these plots, we can infer that Steam Total, Time (t), and Feed Water Temperature are some of the most significant features. The violin plot also illustrates the variance in feature effects, indicating that while certain features have a consistent influence, others, such as Blowdown TDS and Pump Condensate Temperature display more spread-out effects. However, looking at global feature importance alone doesn't fully capture how specific feature values affect predictions. To gain insights into feature interactions and their nonlinear effects, we turn to SHAP dependence plots.

SHAP dependence plots can investigate how individual feature values affect model predictions while also uncovering interactions with other variables. Each scatter plot in the series depicts the relationship between a feature's actual value (x-axis) and its SHAP value (y-axis).

The dependence plots in Figure 6.4 below indicates that Steam Flow Rate, Steam Total, and Time have a strong linear effect on CRF, reinforcing

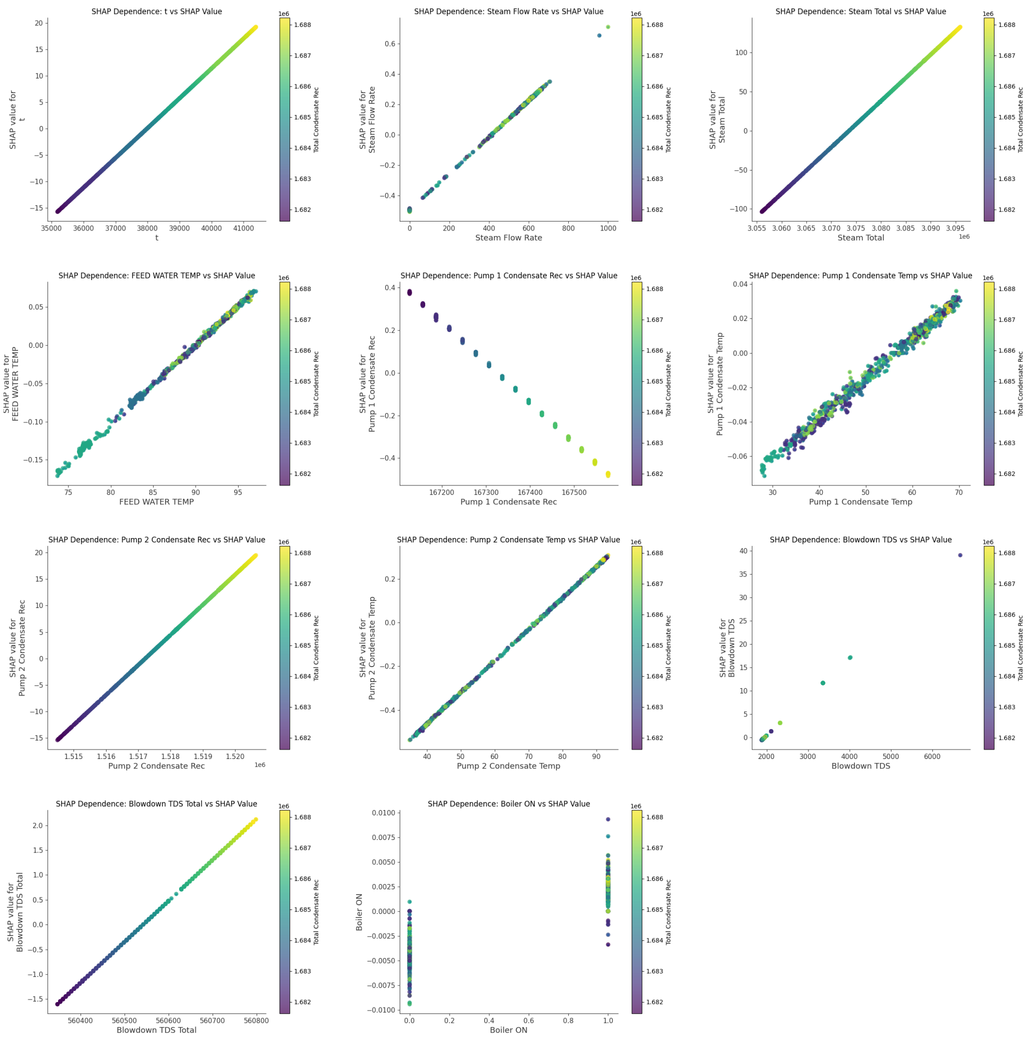

Figure 6.4: SHAP Dependence Plots of features impact on CRF
(Each scatter plot demonstrates how individual features relate to their SHAP values in relation to CRF, emphasizing the key contributors.)

their primary importance. Feed Water Temperature shows a nearly linear relationship, with slight variations suggesting interactions with other parameters.
The inverse relationship between Pump 1 and Pump 2 Condensate Recovery points to operational trade-offs. In contrast, Pump 1 and Pump 2 Condensate Temperatures positively influence CRF, emphasizing their role in heat retention. Blowdown TDS and its total have minor but structured effects, with higher values showing a slight impact on CRF.

The Boiler ON status introduces variability, indicating transient startup effects and delays in condensate recovery.

The next set of dependence plots in Figure 6.5 below explores how different features interact with each other and their combined effect on CRF.

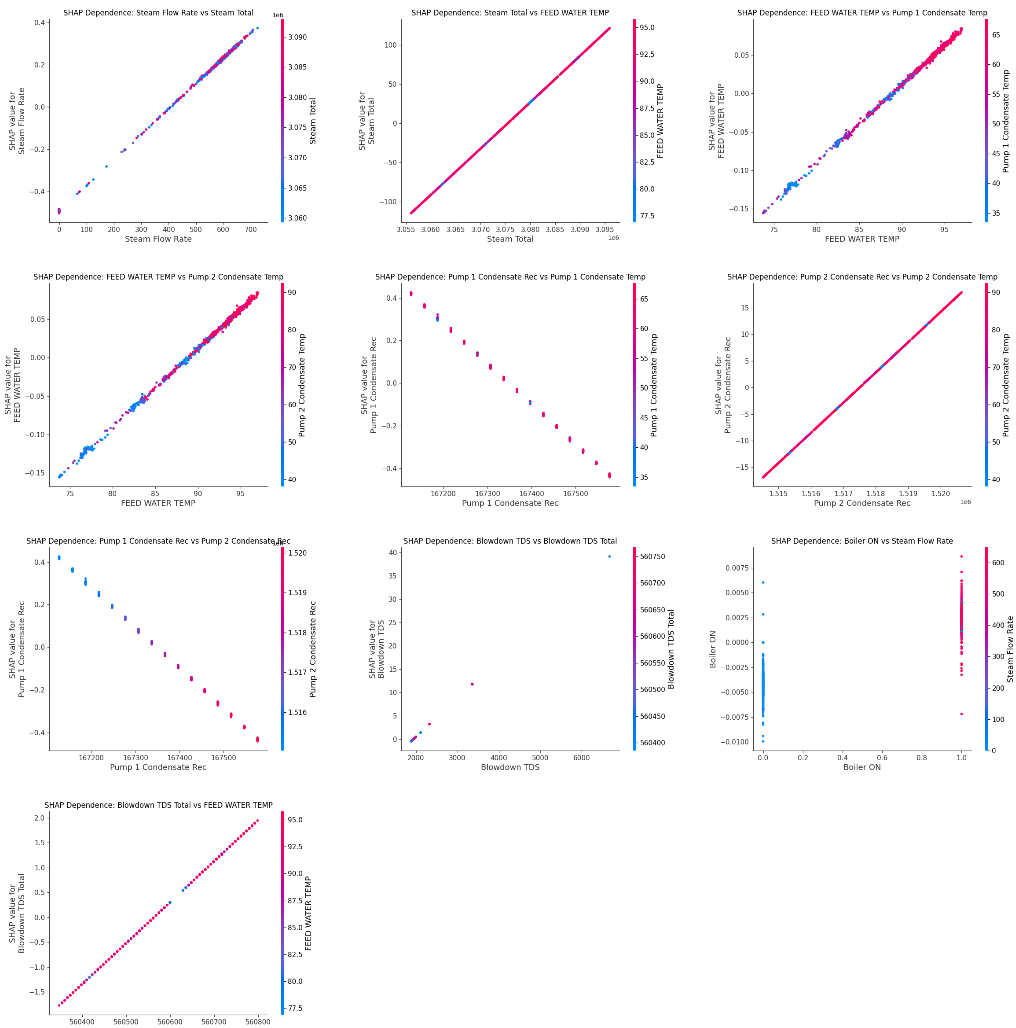

Figure 6.5: Feature interactions captured through SHAP dependence plots
(Each scatter plots demonstrate the connections between essential process parameters and their effects on the CRF, providing valuable insights into how each feature influences the model's predictions, crucial for understanding system dynamics.)

The interactions between features show a strong correlation between Steam

Flow Rate and Steam Total, as well as between Blowdown TDS and its total, confirming their interdependence. The Feedwater temperature impact on steam generation and condensate temperature tends to underline its significance in process efficiency. The reverse relationship of Pump 1 with Pump 2 Condensate Recovery highlights the importance of balancing their loading equilibrium. The farther-out pattern seen between Boiler ON status and Steam Flow Rate emphasizes the influence of transient steam behavior during startup on these interactions, which makes CRF optimization particularly intricate due to pump coordination, steam utilization, and transient states.

These dependence plots give a better indication of how key process parameters relates with the CRF, for future decision-making reference. The effect of change in parameters such as steam flow rate, feed water temperature, and condensate recovery has been illustrated for operational optimization, troubleshooting, and efficiency improvement. With the knowledge of these dependencies, better control strategies can further enhance the performance and sustainability of the plant.

6.4 Feature Sensitivity Analysis: Direct Model Response to Perturbations

Together with the model interpretability techniques i.e SHAP and LIME, an analysis of perturbation was also done to see how changes in key input variables influenced the output, Condensate Recovery Factor (CRF). The perturbation explored changes in the individual features by 20% either way while keeping all other features constant to observe model response. The inference is based on assessing the influence of this direct change of key input variables on the predicted CRF, in the below-shown Figure 6.6.

The results indicates:

- Steam Flow Rate and Feed Water Temperature show a strong positive correlation with CRF, implying that increasing these parameters enhances condensate recovery.
- Pump 1 and Pump 2 Condensate Temperatures also influence CRF but with different sensitivity levels.

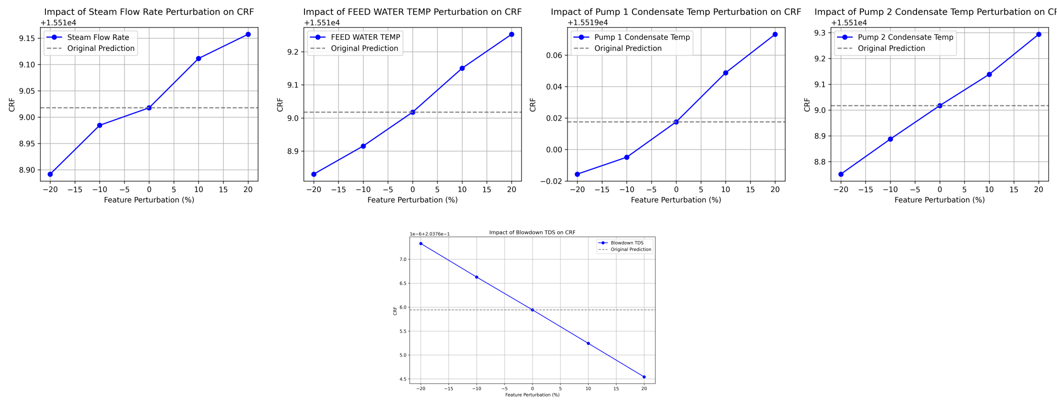

Figure 6.6: Impact of Feature Perturbations on CRF Predictions
(This analysis helps to understand how direct modifications to important input variables impact the predicted CRF. The observed trends suggest possible optimization strategies for enhancing condensate recovery.)

- Blowdown TDS, conversely, shows a negative correlation with CRF. An increase in this corresponds to a decrease in condensate recovery, underlining the importance of water quality for steam system efficiency.

The near-linear trends of all the features confirm that the model is behaving predictably, revealing the importance of these variables in plant improvement. Unlike SHAP, which inspects the model's global interpretability perspective, this perturbation analysis directly explores the model's predictive response, presenting insights to facilitate real-world adjustments towards the improvement of CRF.

6.5 Unified Framework for CRF Optimization

The integration of Fourier Neural Networks along with SHAP and LIME in CRF prediction creates an effective link between AI-driven modeling and practical applications in industry. The Fourier decomposition forms the basis interpretability by building trust in the periodic behaviours learned by the model. SHAP goes one step further by attributing more importance to individual parameters of the process, while LIME provides actionable real-time insights. Feature perturbation analysis also closely examines the model's sensitivity to variations in the key input parameters and provides review on the different ways process condition variations lead to the changes in

CRF predictions. This entire multi-layered approach helps pacify skepticism, by providing actionable insights for engineers and enhancing their trust in AI-based CRF optimization.

,

7 Conclusion and Future Works

7.1 Summary

This thesis aimed to enhance the prediction and optimization of the Condensate Recovery Factor (CRF) in industrial steam systems. Since the recovery factor is a measure of the overall energy efficiency of a system, and many existing methods either tend to take the route of overly simple heuristic approaches or use highly complex black box machine learning models, which can limit their effectiveness in dynamic industrial contexts. Thereby, we suggested a new methodology that merges Fourier-inspired neural networks (FNN) and genetic algorithms (GA) with minutely IoT data from steam and condensate recovery systems.

In the model design, we gave due consideration to domain knowledge and industry-specific practices pertaining to condensate recovery so as to logically interpret and physically embed our approach. The FNN architecture effectively decomposed temporal patterns into spectral components, such as sinusoidal terms that capture daily steam cycles, thus ensuring that predictions were consistent with real-world system behavior. Additionally, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) were also used to allow domain experts to assess feature importance and verify the model's predictions.

To ensure practical relevance, continuous feedback was integrated from steam system specialists, confirming that the model's outputs, such as spectral frequencies linked to boiler cycles, were consistent with empirical observations.

Ultimately this work demonstrates that physics-informed AI, when grounded in industrial telemetry and expert collaboration, can deliver both accurate predictions and actionable insights. By making spectral analysis and evolutionary optimization accessible to plant operators, this work reduces reliance on black-box tools, fostering data-driven, trust-based decision-making in energy-intensive industries.

7.2 Research Questions

The thesis was based on the following research questions and sub-questions, which we were able to answer after finishing the whole process, starting from researching existing approaches to developing and evaluating our own approach.

RQ: *How can a data-driven approach optimize the Condensate Recovery Factor (CRF) in industrial plants while ensuring model interpretability and actionable insights for both AI experts and domain engineers?*

To answer this question, this thesis establishes a novel framework that integrates physics-informed machine learning with industrial IoT data, enhancing both prediction accuracy and practical usability. By leveraging a Fourier-inspired neural network (FNN) to extract temporal patterns and a genetic algorithm (GA) for optimization, the model adapts dynamically to fluctuating plant conditions. SHAP and LIME further ensure transparency by highlighting influential parameters, allowing engineers to validate and refine operational strategies. The approach bridges the gap between data-driven insights and industry constraints, enabling more efficient, sustainable, and trust-driven decision-making in condensate recovery.

SQ1: *How can a neural network model be designed for predicting CRF while ensuring its interpretability and transparency for non-expert stakeholders?*

To enhance interpretability, the FNN architecture breaks down CRF fluctuations into distinct frequency components, mirroring real-world steam cycles. Unlike traditional black-box models, this design allows engineers to trace patterns to physical processes, making outputs more intuitive. Meanwhile, SHAP and LIME quantify the contribution of each input feature, ensuring that plant operators can confidently act on the model's recommendations. This ensures that the predictions remain aligned with industry expectations and operational realities.

SQ2: *What key features and operational parameters significantly impact CRF, as identified through interpretable methods, ensuring alignment with industrial best practices and sustainability goals?*

Through interpretable machine learning, the study pinpointed steam flow rate, boiler status, feed water temperature, and condensate recovery trends as dominant CRF drivers. SHAP and LIME confirmed their influence by exposing cause-and-effect relationships—steam fluctuations and boiler cycles directly impact recovery efficiency. These insights reinforce best practices by guiding efforts to reduce energy losses, optimize steam distribution, and maintain stable operations, aligning with broader sustainability and efficiency objectives.

SQ3: *How can the insights from interpretable machine learning be leveraged to improve decision-making in condensate recovery and steam system efficiency?*

The model explainability tools convert complex predictions into actionable insights, giving plant operators the data-driven support they need for decision making. By consistently analyzing CRF trends, stakeholders can predict inefficiencies, adjust boiler operation schedules, and carry out predictive maintenance while preventing losses before they arise. This strategy reduces the need for trial-and-error methods, ensuring that optimizations are based on real-time system behavior rather than static heuristics. As a result, steam system management becomes more proactive, flexible, and performance-oriented.

7.3 Future works

While this thesis has successfully illustrated a data-driven approach to optimizing the Condensate Recovery Factor (CRF), yet there are still several areas that could benefit from further investigation to enhance predictive accuracy, computational efficiency, and practical implementation in industrial settings.

One of the promising avenues is the application of digital twin technology, which could develop a virtual replica of the condensate recovery system. Synchronized in real time with data from the IoT sensors, the digital twin would enable operators to simulate situations, experiment with optimization strategies, and pre-emptively tune system parameters even before changes become concrete in the physical plant. This would significantly improve decision-making, fault detection, and predictive maintenance.

Another region of improvement is the embedding of real-time IoT data into adaptive predictive models. While this work used minute-scale IoT data, having an automated data pipeline with real-time streaming and anomaly detection would significantly enhance responsiveness to evolving plant conditions. With online learning methods, such as incremental training or reinforcement learning, the model would adapt online, refining predictions as new data arrive without the necessity for full retraining.

From a technical viewpoint, there is still room for improvements in both the neural network and genetic algorithm (GA). Upcoming developments can involve:

- **Enhanced Genetic Algorithm Optimization:** Instead of re-initializing model weights for every generation, GA can be modified to inherit learned weights from previous generations, allowing for more efficient convergence.
- **Multi-Objective Optimization:** Extending GA to simultaneously optimize multiple conflicting objectives, such as maximizing CRF while minimizing energy consumption and system wear.
- **Hybrid Architectures:** Exploring attention mechanisms or transformer-based time series models could help capture long-term dependencies in condensate recovery patterns.

Additionally, enhancing the feature set to incorporate external factors like weather conditions, production schedules, and variations in steam load could significantly boost the model's robustness. This would allow the framework to consider outside influences that impact the CRF, making predictions more aligned with the complexities of the real world.

From an industrial standpoint, scaling up deployment remains a key challenge. Future efforts should investigate cloud-based and edge computing solutions to implement models within industrial facilities while maintaining low-latency inference and security. Additionally, integrating automated explainability dashboards could help connect AI-driven insights with domain expertise, making predictive analytics more practical for plant operators.

By advancing these areas, future studies can extend AI-driven solutions beyond condensate recovery to include aspects like boiler efficiency, pressure control, predictive maintenance, and waste heat recovery, contributing to a fully optimized, self-regulating steam system that ensures sustainability, cost savings, and operational excellence in industrial steam plants.

List of figures

List of tables

Bibliography

[**1**] US Patent for Condensate and flash steam recovery system Patent (Forbes Marshall Pvt. Ltd. Patent # 9,976,809). Available at `https://patents.google.com/patent/US9976809B2/zh-CN`

[**2**] Wankhede, V.A., K.E.K, V., & Sahlot, P. (Eds.). (2025). Industry 4.0 for Manufacturing Systems: Concepts, Technologies, and Applications (1st ed.). CRC Press. Available at https://doi.org/10.1201/9781003470861.

[**3**] Holland, J.H. (1975). Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor. (2nd Edition, MIT Press, 1992.)

[**4**] Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning. Addison-Wesley Professional.

[**5**] Tripathy, B.K., & Seetha, H. (Eds.). (2024). Explainable, Interpretable, and Transparent AI Systems (1st ed.). CRC Press. Available at `https://doi.org/10.1201/9781003442509`.

[**6**] Lundberg, Scott & Lee, Su-In. (2017). A Unified Approach to Interpreting Model Predictions. Available at `https://arxiv.org/abs/1705.07874`.

[**7**] Ribeiro, Marco & Singh, Sameer & Guestrin, Carlos. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. 1135-1144. Available at `https://doi.org/10.1145/2939672.2939778`.

[**8**] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics, Volume 378, 2019, Pages 686-707, ISSN 0021-9991, Available at `https://doi.org/10.1016/j.jcp.2018.10.045`.

[**9**] Bengio, Y. & Glorot, X.. (2010). Understanding the difficulty of training deep feed forward neural networks. International Conference on Artificial Intelligence and Statistics. 249-256.

[10] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Syst. 2 (4) (1989) 303–314.

[11] L. Mingo, L. Aslanyan, J. Castellanos, M. Diaz, V. Riazanov, Fourier neural networks: an approach with sinusoidal activation functions, Int. J. Inf. Theory Appl. 11 (1) (2004).

[12] H. Tan, Fourier neural networks and generalized single hidden layer networks in aircraft engine fault diagnostics, J. Eng. Gas Turbin. Power 128 (4) (2006) 773–782.

[13] W. Zuo, Y. Zhu, L. Cai, Fourier-neural-network-based learning control for a class of nonlinear systems with flexible components, IEEE Trans. Neural Netw. 20 (1) (2009) 139–151.

[14] Michael S. Gashler, Stephen C. Ashmore, Modeling time series data with deep Fourier neural networks, Neurocomputing, Volume 188, ISSN 0925-2312. Available at `https://doi.org/10.1016/j.neucom.2015.01.108`.

[15] R.J. Frank, N. Davey, S.P. Hunt, Time series prediction and neural networks, J. Intell. Robot. Syst. 31 (1–3) (2001) 91–103.

[16] H.D. Abarbanel, R. Brown, J.J. Sidorowich, L.S. Tsimring, The analysis of observed chaotic data in physical systems, Rev. Mod. Phys. 65 (4) (1993) 1331.

[17] Box, George, and G. M. Jenkins. "Analysis: Forecasting and Control." San Francisco (1976).

[18] Wang, Shixiong & Li, Chongshou & Lim, Andrew. (2019). Why Are the ARIMA and SARIMA not Sufficient. Available at `https://arxiv.org/abs/1904.07632`.

[19] Smola, Alex and Bernhard Scholkopf. "A tutorial on support vector regression." Statistics and Computing 14 (2004): 199-222.

[20] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997, Available at `https://doi.org/10.1162/neco.1997.9.8.1735`.

[21] Pascanu, Razvan & Mikolov, Tomas & Bengio, Y.. (2012). On the difficulty of training Recurrent Neural Networks. 30th International Conference on Machine Learning, ICML 2013.