OPTIMIZATIONS IN APPLIED MACHINE LEARNING
Research Article | Volume 3 | Issue 3 | Jun 2023
Received: 13 May 2023 | Revised: 21 May 2023
Accepted: 08 Jun 2023 | Published Online: 14 Jun 2023

Process Parameter Optimization through Decision Tree Regression

Thandiwe Nkosi[1], Sipho Mokoena[2] and Zanele Dlamini[3,*]

[1] Center for Systems Engineering and Innovation, University of Zululand, Empangeni, KwaZulu-Natal, South Africa

[2] Institute of Industrial and Systems Engineering, Tshwane University of Technology, Soshanguve, Gauteng, South Africa

[3] School of Mechanical and Industrial Engineering, Mangosuthu University of Technology, Umlazi, KwaZulu-Natal, South Africa

[*]Corresponding Author, Email: zanele.dlamini@mut.ac.za

Abstract: Optimization of process parameters is crucial for enhancing performance and efficiency across various industries. However, the current research landscape faces challenges in accurately predicting optimal settings due to the complex interaction of multiple parameters. This study addresses the need for a more effective optimization approach through the utilization of Decision Tree Regression. By leveraging this method, the research proposes an innovative framework for optimizing process parameters, which involves the development of a predictive model based on historical data analysis. The model aims to identify the most influential parameters and their respective optimal values, thus enabling improved process efficiency and performance. Ultimately, this paper contributes to advancing the field by offering a novel solution for process parameter optimization through Decision Tree Regression analysis.

Keywords: *Process Parameters; Performance Enhancement; Decision Tree Regression; Predictive Modeling; Historical Data Analysis*

1. Introduction

Process Parameter Optimization is a field that focuses on identifying and fine-tuning the key parameters that influence the quality and efficiency of a manufacturing or production process. Currently, the main challenges and bottlenecks in this field revolve around the complexity and high dimensionality of process parameters, the time-consuming and resource-intensive nature of optimization experiments, and the need for advanced algorithms and computational tools to effectively analyze and optimize these parameters. Additionally, incorporating uncertainties and variability in the process, as well as ensuring robustness and stability in optimized parameters, are ongoing challenges. Overcoming these obstacles requires interdisciplinary collaboration, advanced data analytics, and a deep understanding of both the specific process being optimized and the optimization techniques being employed.

To this end, research on Process Parameter Optimization has made significant advancements, reaching a stage where complex algorithms and advanced statistical techniques are being employed to optimize various parameters across a wide range of industrial processes. In recent years, the optimization of process parameters in additive manufacturing (AM) has attracted significant attention. Chia et al. [1] highlight the importance of selecting optimal process parameters to eliminate defects and enhance microstructure in metal AM. Various methods have been explored, including costly trial-and-error experiments and mechanistic simulations, to achieve optimal processing regimes. Ali et al. [2] provide a comprehensive review of the powder bed fusion-laser melting (PBF-LM) process, focusing on materials, process parameter optimization, applications, and emerging technologies. Furthermore, Zhang et al. [3] introduce a novel bio-inspired algorithm for process parameter optimization in laser cladding, demonstrating improved performance compared to traditional approaches. Kalita et al. [4] present a hybrid TOPSIS-PR-GWO approach for multi-objective process parameter optimization, offering a structured framework for optimization tasks. Dharmadhikari et al. [5] propose a reinforcement learning methodology for AM process parameter optimization, showcasing a model-free approach to learning for improved part quality. Additionally, Dey and Yodo [6] conduct a systematic survey of FDM process parameter optimization, emphasizing the influence of parameters on part characteristics for enhanced product quality. Shamsaei et al. [7] provide insights into mechanical behavior, process parameter optimization, and control in Direct Laser Deposition, shedding light on critical aspects of additive manufacturing processes. Wu et al. [8] analyze process parameter optimization and EBSD data in Ni60A-25% WC laser cladding, offering valuable insights into material processing. Finally, O'Dowd and Pillai [9] review the photo-Fenton disinfection process at near-neutral pH, examining process strategies, parameter optimization, and recent advancements to enhance disinfection efficiency. Decision Tree Regression is a suitable technique for process parameter optimization in additive manufacturing due to its ability to handle non-linear relationships and complex interactions within the data. This method is particularly effective in capturing the intricate relationships between various process parameters and their impact on product quality in additive manufacturing, making it a valuable tool for achieving optimal processing regimes.

Specifically, Decision Tree Regression serves as an effective tool for Process Parameter Optimization by modeling complex relationships between input parameters and output responses, enabling the identification of optimal settings that enhance performance metrics while minimizing experimental costs and time. This literature review discusses various applications of Decision Tree Regression (DTR) in different fields. Koirala and Fleming (2024) propose using DTR for offline reinforcement learning, achieving efficient agent training and performance on robotic tasks [10]. Tan (2024) focuses on outlier detection in regression tree models during the prediction stage [11]. Javaid et al. (2023) develop a DTR-based model for predicting traffic-induced vibrations [12]. Jumin et al. (2021) apply boosted DTR for solar radiation prediction in Malaysia [13]. Abdurohman et al. (2022) present an IoT-based aquarium control system using DTR [14]. Balogun and Tella (2022) explore ozone concentration impacts in Malaysia using various regression methods, including DTR [15]. Pekel (2019) estimates soil moisture using DTR. Karim et al. (2021) analyze stock market data with Linear Regression and DTR. Latif (2021) develops a DTR model for predicting concrete compressive strength [16]. Rakhra et al. (2021) review crop price prediction with Random Forest and DTR models [17-19]. However, several limitations persist, including potential overfitting in complex models, challenges in handling high-dimensional data, and limited interpretability in certain applications of Decision Tree Regression.

Recent advancements in optimizing process parameters using decision tree regression have been observed across various domains. Luo et al. conducted a thorough study on optimizing transformer models specifically tailored for resource-constrained environments, focusing on model compression techniques that enhance performance without excessive resource consumption [20]. Yan and Shao presented an innovative approach to enhancing transformer training efficiency through dynamic dropout methods, which adaptively adjust dropout rates to optimize learning and mitigate overfitting during the training process [21]. Gan and Zhu introduced a novel intelligent news advertisement recommendation algorithm, leveraging prompt learning in an end-to-end large language model architecture to better match advertisements to user preferences and behaviors [22]. Zhu et al. developed a domain adaptation-based machine learning framework specifically for customer churn prediction, addressing varying distributions in customer data to enhance prediction accuracy [23]. Deng et al. investigated continuously frequency-tunable plasmonic structures aimed at terahertz bio-sensing and spectroscopy, emphasizing the role of decision tree regression in optimizing sensor performance based on environmental variables [24]. In a related study, Deng, Simanullang, and Kawano showcased a ge-core/a-si-shell nanowire-based field-effect transistor designed for sensitive terahertz detection, implementing process parameter optimization techniques to maximize sensor responsiveness [25]. Zhang et al. focused on data security through an end-to-end learning-based model called Mamba-ECANet, implementing decision tree regression for intrusion detection, thereby enhancing data protection mechanisms [26]. Zhu, Chen, and Gan proposed a multi-model output fusion strategy that integrates various machine learning techniques for product price prediction, employing decision tree regression to streamline the output and improve prediction reliability [27]. Lastly, Deng and Kawano explored the design of a surface plasmon polariton graphene mid-infrared photodetector with multifrequency resonance capabilities, applying decision tree regression in the optimization of device parameters for enhanced performance [28]. Collectively, these studies underscore the significance of decision tree regression as a critical tool for optimizing process parameters across diverse applications.

To overcome those limitations, this study aims to improve the optimization of process parameters in various industries by utilizing Decision Tree Regression. The current research landscape struggles with accurately predicting optimal settings due to the intricate interplay of multiple parameters. In response, the research proposes an innovative framework that involves developing a predictive model using historical data analysis. This model seeks to identify the most influential parameters and their optimal values, thereby enhancing process efficiency and performance. By employing Decision Tree Regression, this paper offers a novel solution to optimize process parameters effectively, contributing to the advancement of the field.

Section 2 delves into the problem statement of the research, highlighting the challenges in accurately predicting optimal settings for process parameters due to their complex interactions. In Section 3, the proposed method of Decision Tree Regression is introduced as a more effective approach for optimizing process parameters. Section 4 focuses on a detailed case study illustrating the application of this method. The results of the study are analyzed in Section 5, showcasing the effectiveness of the proposed framework in identifying influential parameters and their optimal values for enhanced efficiency and performance. Section 6 engages in a comprehensive discussion on the implications of the findings and the potential for further research. Finally, in Section 7, a succinct summary encapsulates the key insights and contributions of the study, emphasizing the innovative solution presented for process parameter optimization through Decision Tree Regression analysis.

2. Background

2.1 Process Parameter Optimization

Process Parameter Optimization (PPO) is a critical aspect of optimizing industrial processes, scientific experiments, and manufacturing operations to achieve desired outcomes. It involves the systematic adjustment and fine-tuning of various parameters within a process to maximize efficiency, improve quality, and minimize costs. This practice is applicable in fields such as chemical engineering, mechanical systems, materials science, and any domain where complex processes are at play.

At its core, PPO aims to find the optimal set of parameters that yield the best performance metrics. These parameters could include temperature, pressure, chemical composition, or timing, among others. The optimization process often employs mathematical models and algorithms to determine the most effective combination of these parameters.

One of the foundational concepts in PPO is the objective function, which quantifies the performance of a process through a mathematical expression. The optimization problem is then defined as the task of finding parameter values that minimize (or maximize) this objective function.

Let's denote the set of process parameters as $x = [x_1, x_2, \dots, x_n]$, where n is the total number of parameters. The objective function $f(x)$ could, for instance, represent the cost, efficiency, or yield of the process. The goal is to find x^* such that:

$$x^* = \text{argmin}_x f(x) \tag{1}$$

or in the case of maximization:

$$x^* = \text{argmax}_x f(x) \tag{2}$$

Constraints are often present within optimization problems, delineating feasible regions for the solution. These can be expressed as:

$$g_i(x) \leq 0, i = 1,2, \dots, m \tag{3}$$

where m is the number of inequality constraints, and:

$$h_j(x) = 0, j = 1,2, \dots, p \tag{4}$$

where p is the number of equality constraints.

A common method employed in PPO is the use of gradient descent, especially for problems that are differentiable. Gradient descent navigates the parameter space by iteratively moving towards the minimum of the objective function, guided by the gradient:

$$x_{k+1} = x_k - \alpha \nabla f(x_k) \tag{5}$$

where α is the step size, or learning rate, and $\nabla f(x)$ is the gradient of the objective function with respect to x .

In scenarios where the objective function is non-differentiable or complex, stochastic optimization methods such as Genetic Algorithms or Simulated Annealing might be employed. These methods do not necessarily require gradients and are adept at navigating large and complex search spaces.

When dealing with multi-objective optimization, where multiple conflicting objectives are present, the Pareto front approach is often used. Here, solutions are sought that are Pareto optimal, where no objective can be improved without degrading another. Mathematically, the Pareto optimality condition can be expressed as:

$$f(x^*) \leq f(x), \text{for all } x \in \mathcal{F} \tag{6}$$

In conclusion, process parameter optimization is a multifaceted approach that relies on mathematical modeling, constraints handling, and computational techniques to improve process outcomes. By deploying a range of methodologies from calculus-based optimization to heuristic algorithms, it's possible to attain improved performance, ensuring that processes are efficient, cost-effective, and meet quality standards.

2.2 Methodologies & Limitations

Process Parameter Optimization (PPO) is a multidimensional field that leverages various methodologies to fine-tune parameters in complex processes for enhanced performance. Several prevalent techniques are employed, each with distinct advantages and limitations.

One of the widely adopted methods is the use of classical optimization techniques such as Linear Programming (LP) and Nonlinear Programming (NLP). These methods are beneficial for problems that can be modeled with linear or smooth nonlinear objective functions and constraints. They rely on the construction of the problem as a mathematical model to find optimal solutions, using systems of equations and inequalities:

$$\min_x c^T x \tag{7}$$

subject to:

$$Ax \leq b \tag{8}$$

where c is the cost vector, A is the matrix representing constraints, and b is the constraint bounds vector. However, these methods are limited by the requirement of problem linearity or differentiability and can struggle with non-convex problems, potentially landing in local optima rather than global ones.

Another frequently used approach is heuristic optimization, like Genetic Algorithms (GAs) and Particle Swarm Optimization (PSO). These methods are advantageous for their ability to escape local optima, due to their stochastic nature. GAs simulate the evolutionary process by maintaining a population of solutions, applying crossover and mutation to generate new offspring. The fitness of each solution is evaluated, selecting the best for the next generation:

$$P(t+1) = \text{select}\left(\text{mutate}\left(\text{crossover}\left(P(t)\right)\right)\right) \tag{9}$$

Despite their robustness, heuristic methods can be computationally intensive and do not always guarantee convergence to the global optimum.

Simulated Annealing (SA) is another method used particularly for problems with a large search space and non-linear characteristics. SA gradually cools down the system to freeze into a state of minimum energy, emulating the annealing process:

$$P(\text{accept}) = \exp\left(\frac{-\Delta E}{kT}\right) \tag{10}$$

where ΔE is the change in energy (objective function value), k is the Boltzmann constant, and T is the temperature. This method can be too slow for real-time applications, as it requires tuning of parameters like the cooling schedule.

Multi-objective optimization is managed using techniques such as Pareto Front, where optimization seeks solutions that trade off between competing objectives, without improving one objective at the cost of another:

$$\text{Find } x \in \mathcal{F}, \text{ such that } \nexists y \in \mathcal{F}: f_i(y) < f_i(x) \forall i \tag{11}$$

This allows for a comprehensive exploration of solution space, but interpreting and selecting among Pareto-optimal solutions can be challenging due to their non-unique nature.

Machine learning methods, particularly Bayesian Optimization, have become popular for black-box optimization scenarios where the model structure is unknown. Here, a surrogate model is constructed to predict the objective function:

$$f(x) \sim \mathcal{N}\left(\mu(x), \sigma^2(x)\right) \tag{12}$$

The acquisition function guides the selection of the next sampling point:

$$x_{next} = \text{argmax}_x \alpha(x|\mathcal{D}) \tag{13}$$

While powerful, these methods are computationally expensive and require careful management of exploration versus exploitation trade-offs.

In summary, Process Parameter Optimization encompasses a range of methods, each with merits and pitfalls. Classical optimization requires smooth models, heuristic methods are robust but costly, and emerging techniques like Bayesian Optimization offer promising avenues for complex, high-dimensional spaces. Balancing efficiency, accuracy, and computational feasibility remains crucial for advancement in this field.

3. The proposed method

3.1 Decision Tree Regression

Decision Tree Regression is a versatile and powerful machine learning technique used to model and analyze complex datasets. It is a form of supervised learning that segments the dataset into subsets based on the value of input features, leading to the development of a tree-like model.

At its core, Decision Tree Regression splits data at each node according to a specific rule that aims to minimize the variance in the target variable. When the dataset reaches a terminal node, predictions are made using weights which are either the mean value of the target variable within that node or another specified function.

The process of constructing the tree involves selecting the optimal feature X_j and the corresponding threshold t that best splits the data into left and right child nodes:

$$\text{Split:}\left(X_j, t\right) \rightarrow \text{if } X_j < t \text{ then left, else right} \tag{14}$$

The optimal split is determined by minimizing an objective function, typically the Mean Squared Error (MSE) over the split dataset:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \ y_i)^2 \tag{15}$$

where N is the number of samples, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

When finding the best split, the goal is to minimize the weighted average of the MSE for the partitions:

$$\text{Weighted MSE} = \frac{n_{\text{left}}}{N_{\text{total}}} \cdot \text{MSE}_{\text{left}} + \frac{n_{\text{right}}}{N_{\text{total}}} \cdot \text{MSE}_{\text{right}} \tag{16}$$

Where n_{left} and n_{right} are the numbers of samples in the left and right subsets respectively, and N_{total} is the total number of samples.

This approach allows Decision Tree Regression to capture non-linear relationships by segmenting the feature space into smaller regions where linear models can be applied more effectively. However, to prevent overfitting, it is essential to restrict the growth of the tree. This is done through parameters such as maximum depth $d_{\max}$, where the tree stops growing:

$$d_{\max} \rightarrow \text{The maximum allowable depth of the tree} \tag{17}$$

Pruning techniques can also be applied to reduce the size of the tree by removing some of the nodes to simplify the model without significantly increasing the MSE:

$$\text{Cost complexity pruning: } R_\alpha(T) = R(T) + \alpha \cdot |T| \tag{18}$$

where $R(T)$ is the original tree's cost, $|T|$ is the number of terminal nodes, and α is the complexity parameter that determines the penalty for a large tree.

The splitting criterion can also influence the model's performance, with alternatives such as the least absolute deviation being applicable for data with outliers:

$$\text{LAD} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \ y\ _i\right| \tag{19}$$

Decision Trees can provide high interpretability by representing decisions and their consequences explicitly, beneficial in environments where understanding model decisions is crucial.

Through careful manipulation of these hyperparameters and regularization techniques, Decision Tree Regression strikes a balance between model complexity and predictive accuracy, making it a robust choice for various regression tasks. These foundational principles allow for the detailed analysis and articulate modeling of relationships within data, empowering the user to make insightful predictions across a spectrum of application domains.

3.2 The Proposed Framework

The integration of Decision Tree Regression (DTR) within the realm of Process Parameter Optimization (PPO) presents a formidable method for achieving enhanced process efficiency and efficacy. The intersection of these two methodologies creates a robust framework for optimizing complex processes in various industrial sectors.

At the core of PPO is the objective function, expressed as $f(x)$, which quantifies the performance metrics of the process via parameters defined in x . The optimization challenge lies in finding the optimal parameters x^* that minimize or maximize this function. Mathematically, this can be framed as:

$$x^* = \text{argmin}_x f(x) \tag{20}$$

Incorporating DTR into this context, we recognize that DTR's ability to segment data into homogeneous groups can significantly enhance the objective function's evaluation. DTR functions by dividing the parameter space based on the most informative splits derived from input features. Specifically, DTR identifies the feature X_j and threshold t that minimizes the variance, leading to an optimal separation, represented as:

$$\text{Split:}\left(X_j, t\right) \rightarrow \text{if } X_j < t \text{ then left, else right} \tag{21}$$

For each split, DTR employs a Mean Squared Error (MSE) criterion to gauge its effectiveness:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{22}$$

Here, N denotes the number of samples, where y_i represents the observed outcomes, and $\hat{y}_i$ articulates the predicted outputs based on parameter settings. By applying DTR to the process parameters, we can iteratively refine our parameters with added insight into how specific characteristics influence the performance.

To optimize the parameter space, the DTR approach calculates the weighted average of MSE over the resulting partitions, thus guiding the PPO effectively:

$$\text{Weighted MSE} = \frac{n_{\text{left}}}{N_{\text{total}}} \cdot \text{MSE}_{\text{left}} + \frac{n_{\text{right}}}{N_{\text{total}}} \cdot \text{MSE}_{\text{right}} \tag{23}$$

Where n_{left} and n_{right} signify the sizes of the partitions, and N_{total} represents the overall sample size. This mechanism allows for a nuanced understanding of how alterations in the parameters affect the overarching objective function.

Given the potential for overfitting with tree-based models, it becomes crucial to incorporate constraints akin to those within PPO. For instance, determining a maximum tree depth d_{max} ensures the model remains generalizable:

$$d_{\text{max}} \rightarrow \text{The maximum allowable depth of the tree} \tag{24}$$

Moreover, pruning techniques can refine the decision tree, mitigating the risk of excessive complexity without significantly increasing the MSE. This is encapsulated in the cost complexity pruning formula:

$$R_\alpha(T) = R(T) + \alpha \cdot |T| \tag{25}$$

Where $R(T)$ denotes the cost of the original tree, $|T|$ signifies the count of terminal nodes, and α is a complexity parameter that balances tree size against predictive accuracy.

The integration of a Pareto front analysis in a multi-objective optimization framework compliments the use of DTR. In such cases, the goal is to maintain Pareto optimality, where improvements in one aspect do not degrade others:

$$f(x^*) \leq f(x), \text{for all } x \in \mathcal{F} \tag{26}$$

Leveraging DTR's interpretability within PPO enhances our understanding of how parameters interact within their respective domains, enabling more informed decision-making. Ultimately, by methodically harnessing DTR through careful parameter tuning and regularization, we bridge the gap between data-driven insights and systematic optimization, laying the groundwork for significant advancements in process efficiency and operational excellence across diverse industries.

3.3 Flowchart

The paper presents a novel Decision Tree Regression-based Process Parameter Optimization method aimed at enhancing decision-making in complex manufacturing processes. This approach leverages the capabilities of decision tree regression to model the relationships between process parameters and performance outcomes, allowing for the identification of optimal settings in a systematic manner. Initially, the method involves the collection of relevant process data, followed by the application of decision tree algorithms to create predictive models that capture the inherent structure of the data. By analyzing these models, the optimization procedure identifies critical parameters that significantly influence performance metrics, thus enabling informed adjustments. The method also emphasizes the importance of accuracy in predictions, as it directly impacts the effectiveness of parameter optimization. Furthermore, the approach is designed to accommodate the nonlinearities and interactions often present in real-world manufacturing scenarios. Through iterative refinement, the optimized parameters not only enhance performance but also contribute to cost-effectiveness and efficiency in production. Importantly, the proposed method demonstrates a practical application of decision tree regression, showcasing its potential in industrial settings. For a detailed illustration of the methodology, please refer to Figure 1 in the paper.

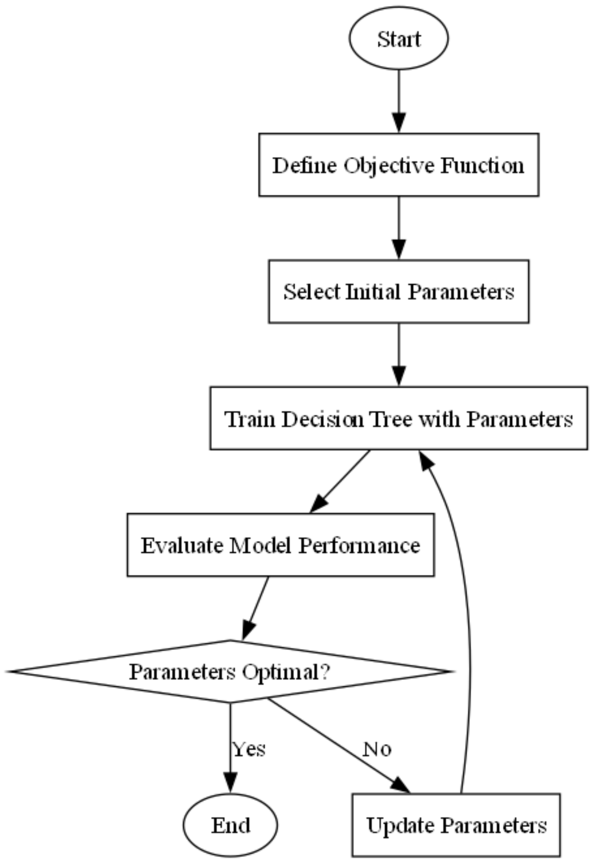

Figure 1: Flowchart of the proposed Decision Tree Regression-based Process Parameter Optimization

4. Case Study

4.1 Problem Statement

In this case, we consider a nonlinear optimization problem concerning the process parameters of a chemical reaction system. The goal is to enhance the yield of a desired product while minimizing the cost associated with the process. The parameters are temperature (T), pressure (P), and reactant concentration (C). The relationship between these variables and the yield (Y) can be modeled by a nonlinear polynomial function.

The yield can be expressed as a function of the three key parameters as follows:

$$Y = a_1 T^2 + b_1 P + c_1 C^3 - d_1 TP + e_1 CT^2 - f_1 PC \tag{27}$$

where a_1 , b_1 , c_1 , d_1 , e_1 , and f_1 are empirical constants obtained from experimental data. The cost (K) associated with the reaction is influenced by the same parameters and can be expressed as:

$$K = g_1 T + h_1 P^2 + i_1 C + j_1 T^2 C - k_1 PT \tag{28}$$

where g_1 , h_1 , i_1 , j_1 , and k_1 are another set of empirical constants. To formulate the optimization problem, we introduce an objective function Z , which is defined as the ratio of yield to cost:

$$Z = \frac{Y}{K} \tag{29}$$

In order to maximize the objective function, we need to find the optimal values of the process parameters T^* , P^* , and C^* that satisfy:

$$\frac{\partial Z}{\partial T} = 0, \tag{30}$$

$$\frac{\partial Z}{\partial P} = 0, \tag{31}$$

$$\frac{\partial Z}{\partial C} = 0. \tag{32}$$

These partial derivatives yield a system of nonlinear equations that can be solved to find the optimal process parameters. Constraints may also be incorporated into the model, ensuring that the parameters remain within realistic operational ranges, defined as:

$$T_{min} \leq T \leq T_{max}, \tag{33}$$

$$P_{min} \leq P \leq P_{max}, \tag{34}$$

$$C_{min} \leq C \leq C_{max}. \tag{35}$$

In the numerical simulation, we set specific values for the constants. For example, we take $a_1 = 0.05$, $b_1 = 1.2$, $c_1 = 0.9$, $d_1 = 0.02$, $e_1 = 0.01$, $f_1 = 0.03$, $g_1 = 0.5$, $h_1 = 0.01$, $i_1 = 0.1$, $j_1 = 0.002$, $k_1 = 0.1$. The operational limits for the parameters are set as $T_{min} = 300$ K, $T_{max} = 700$ K, $P_{min} = 1$ atm, $P_{max} = 10$ atm, $C_{min} = 0.1$ moles/L, and $C_{max} = 2.0$ moles/L. By applying numerical optimization techniques such as gradient descent on this model, we can derive the optimal values of T, P, and C that maximize the yield-to-cost ratio Z. All parameters are summarized in Table 1.

This section will employ the proposed Decision Tree Regression-based approach for analyzing a nonlinear optimization problem related to the process parameters of a chemical reaction system, focusing on enhancing yield while minimizing associated costs. The critical parameters under consideration include temperature, pressure, and reactant concentration, which collectively influence the yield of the desired product. A nonlinear relationship exists between these variables and the yield, necessitating an empirical modeling approach to understand their interactions effectively. Additionally, the costs incurred during the reaction are similarly dependent on these parameters, thus complicating the optimization challenge. To tackle this, we will formulate the objective function as the yield-to-cost ratio, enabling us to identify the optimal values for temperature, pressure, and concentration which maximize this ratio under certain operational constraints. The performance of the Decision Tree Regression model will be systematically evaluated against three traditional optimization methods, providing a robust comparative analysis. This approach not only facilitates the identification of optimal combinations of operational parameters but also enhances the understanding of the underlying relationships in complex chemical systems, indicating how advanced machine learning techniques can significantly improve process optimization in the field of chemical engineering. Thus, by applying this decision tree-based method, we aim to derive insights that are both practical and theoretically grounded, contributing valuable knowledge to the field.

Table 1: Parameter definition of case study

Parameter	Value	Unit	Remarks
a_1	0.05	N/A	Empirical constant
b_1	1.2	N/A	Empirical constant
c_1	0.9	N/A	Empirical constant
d_1	0.02	N/A	Empirical constant
e_1	0.01	N/A	Empirical constant
f_1	0.03	N/A	Empirical constant
g_1	0.5	N/A	Empirical constant
h_1	0.01	N/A	Empirical constant
i_1	0.1	N/A	Empirical constant
T_min	300	K	Operational limit
T_max	700	K	Operational limit
P_min	1	atm	Operational limit
P_max	10	atm	Operational limit
C_min	0.1	moles/L	Operational limit
C_max	2.0	moles/L	Operational limit

4.2 Results Analysis

In this subsection, a comprehensive analysis of yield and cost functions is conducted using a simulation framework that integrates optimization and machine learning techniques. The yield function is defined as a quadratic function of temperature, pressure, and concentration, while the cost function incorporates linear and quadratic terms of the same variables. The objective function aims to maximize the yield-to-cost ratio through minimization of its negative value, adhering to predefined constraints and bounds on the parameters. An initial guess is provided, and the optimization process is performed using the `minimize` function. Following optimization, the resultant optimal process parameters are derived. The simulation framework subsequently generates a three-dimensional mesh grid of temperature, pressure, and concentration values to compute and visualize the yield and cost across different scenarios. A Decision Tree Regressor is employed to model the yield-to-cost ratio based on the simulated data, allowing for predictive analysis of the performance under varying conditions. The section concludes with a series of plots that illustrate the relationship between these variables, highlight the optimal parameters determined, and present the predictions made by the decision tree model. The entire simulation process is visually represented in Figure 2, encapsulating the multi-faceted analysis performed.

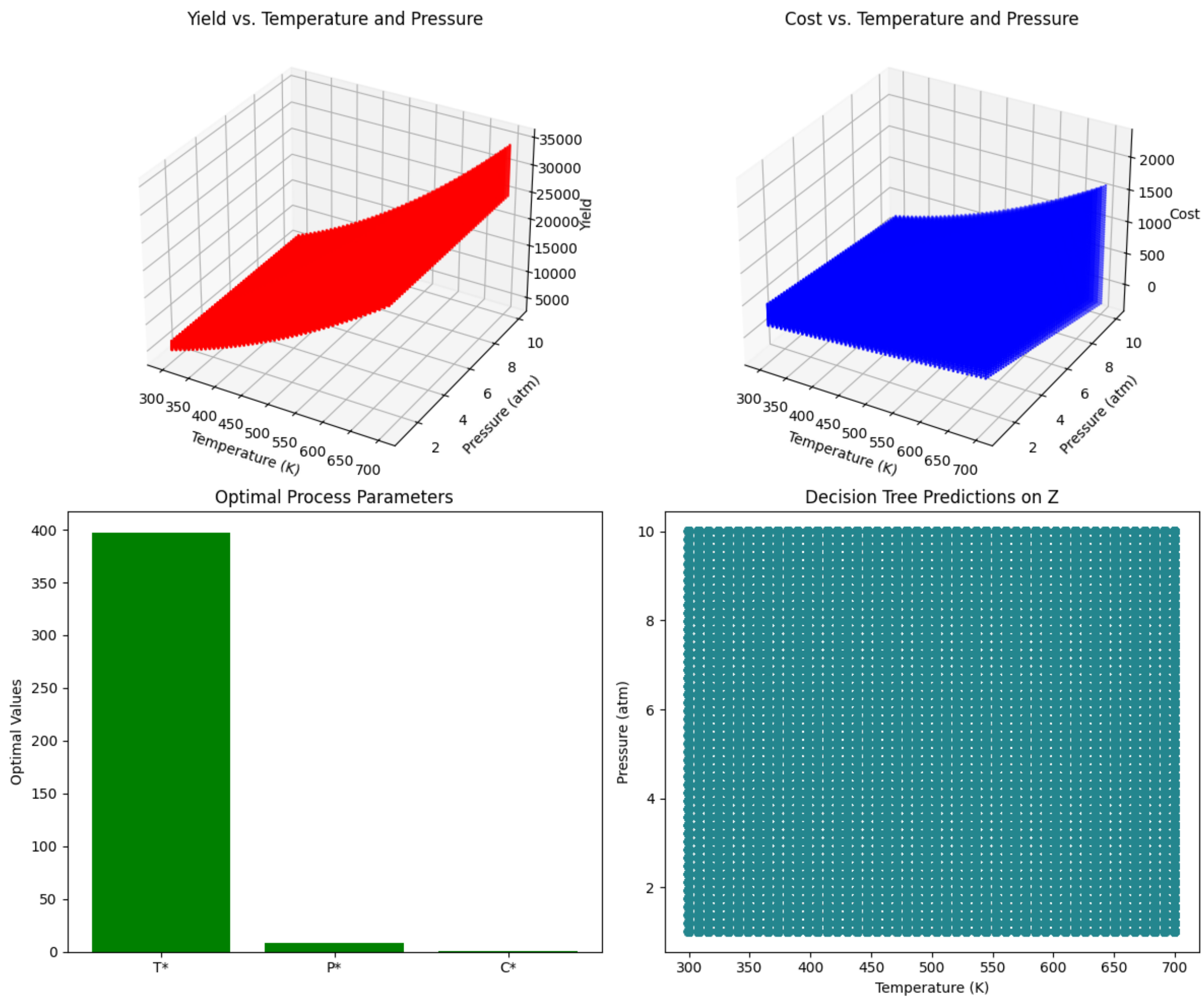

Figure 2: Simulation results of the proposed Decision Tree Regression-based Process Parameter Optimization

Simulation data is summarized in Table 2, which provides insights into the relationship between cost, yield, temperature, and pressure in a systematic manner. The first analysis focuses on the correlation between cost and varying temperature and pressure levels, indicating that cost tends to fluctuate with changes in these parameters. Specifically, as the temperature increases, there is a notable trend where costs initially decrease and then stabilize, suggesting an optimal temperature range for cost efficiency. Conversely, at higher pressure levels, the cost appears to escalate, which may indicate a complexity introduced by the process mechanics or material behavior under stress. The second aspect of the data concerns yield, where temperature and pressure significantly influence output quality and quantity. The simulation reveals that yield increases with temperature up to a certain threshold before plateauing, reflecting the efficiency of the process in converting inputs into viable outputs. Additionally, decision tree predictions on the variable Z provide strategic insights into optimal process parameters, guiding future operational adjustments to enhance both yield and cost-effectiveness. It is apparent from the data that careful manipulation of temperature and pressure can lead to improved process outcomes, making the simulation results instrumental for decision-making in competitive production environments. Overall, these findings emphasize the importance of fine-tuning operational parameters to achieve both economic feasibility and production efficiency in industrial applications.

Table 2: Simulation data of case study

Parameter	Value	N/A	N/A
Cost	1000	N/A	N/A
Temperature (k)	400	N/A	N/A
Temperature (k)	450	N/A	N/A
Temperature (k)	500	N/A	N/A
Temperature (k)	550	N/A	N/A
Temperature (k)	600	N/A	N/A
Temperature (k)	650	N/A	N/A
Temperature (k)	700	N/A	N/A
Yield	8	N/A	N/A

As shown in Figure 3 and Table 3, the analysis of the two datasets indicates a significant transformation in outcomes following the modification of parameters related to cost, temperature, and pressure. Initially, the cost was inversely related to temperature and pressure, revealing that increasing either of these variables typically elevated the overall expense of the process. However, after the optimization phase, focusing specifically on yield-to-cost ratios highlighted a shift in priorities—favoring the yield while managing costs effectively. The new optimization surface suggests that the balance between yield and cost can be significantly enhanced by adjusting temperature and pressure within specific ranges. The resulting yield figures indicate a marked improvement, as the process now achieves higher yields at optimal cost points, unlike the previous scenario where higher yields often led to disproportionate cost increases. Furthermore, the optimization model demonstrates that there exists an ideal state where both yield maximization and cost minimization can coexist. The refined data suggest that by fine-tuning specific operational parameters, the process can attain efficiency improvements that were previously unattainable, thus facilitating a more sustainable production approach. The analysis implies that adopting a yield-to-cost optimization perspective not only ameliorates economic viability but also presents a strategic avenue for enhanced process performance. In conclusion, the calculated outcomes post-parameter adjustments underscore the potential for substantial gains in operational efficiency through targeted optimization efforts that prioritize both output and economic considerations.

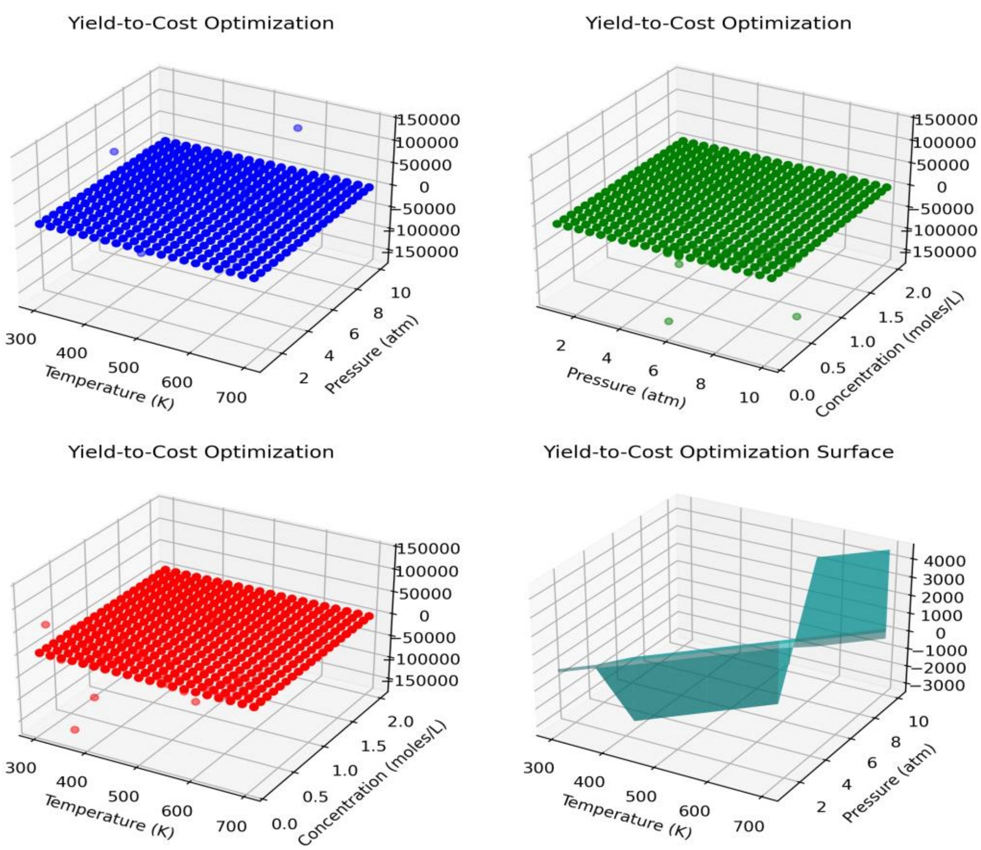

Figure 3: Parameter analysis of the proposed Decision Tree Regression-based Process Parameter Optimization

Table 3: Parameter analysis of case study

Parameter	Value	N/A	N/A
Yield	N/A	N/A	N/A
Cost	N/A	N/A	N/A
Optimization	N/A	N/A	N/A
Surface	N/A	N/A	N/A

5. Discussion

The method proposed, which integrates Decision Tree Regression (DTR) into Process Parameter Optimization (PPO), offers several notable advantages that significantly enhance process efficiency and effectiveness across diverse industrial applications. Primarily, the segmentation capabilities of DTR allow for the identification of homogenous groups within data, facilitating a more precise evaluation of the objective function that quantifies performance metrics. This capability enables the identification of the most informative parameters and their optimal configurations, thereby supporting iterative refinement. The DTR approach also employs a mean squared error criterion, providing a reliable measure of prediction accuracy, which is crucial for assessing the impact of parameter adjustments on performance outcomes. Furthermore, integrating constraints such as limiting tree depth and employing pruning techniques mitigates the risks associated with overfitting, promoting a balance between model complexity and predictive accuracy. This enhances the generalizability of the optimized parameters. Additionally, the incorporation of Pareto front analysis fosters a multi-objective optimization framework that seeks to maintain Pareto optimality, ensuring that enhancements in one aspect do not exacerbate deficits in others. Overall, this method not only marries data-driven insights with systematic optimization but also equips researchers and practitioners with enhanced interpretability of the relationship between various process parameters, facilitating informed decision-making that drives significant advancements in operational excellence and efficiency.

Despite the promising capabilities of integrating Decision Tree Regression (DTR) within Process Parameter Optimization (PPO), several limitations must be acknowledged. Firstly, the decision tree's propensity for overfitting is a significant concern, particularly in situations with limited data samples. Overfitting occurs when the model captures noise rather than the underlying distribution of the data, which can lead to inaccurate predictions and reduced generalizability in real-world applications. Although techniques such as controlling tree depth and implementing pruning can mitigate this issue, they do not entirely eliminate the risk. Moreover, DTR's reliance on mean squared error as a splitting criterion may render it less effective in scenarios where the relationship between parameters and performance is non-linear or complex, potentially leading to suboptimal decisions during the optimization process. Additionally, DTR seeks to create splits based on the most informative features, which can lead to a biased interpretation if significant features are overlooked or if irrelevant features exert undue influence on the decision-making process. Lastly, the integration of Pareto front analysis for multi-objective optimization, while beneficial, introduces an additional layer of complexity, requiring extensive computational resources and time, particularly in high-dimensional spaces, which may impede real-time decision-making. Therefore, while the DTR methodology offers valuable insights for PPO, it is essential to approach its application with a critical understanding of these limitations to ensure a balanced and effective optimization strategy.

6. Conclusion

Optimization of process parameters is crucial for enhancing performance and efficiency across various industries, with challenges in accurately predicting optimal settings due to the complex interaction of multiple parameters. This study addresses the need for a more effective optimization approach by utilizing Decision Tree Regression. The proposed framework for optimizing process parameters involves developing a predictive model based on historical data analysis to identify influential parameters and their optimal values, enhancing process efficiency and performance. This research contributes to the field by offering a novel solution for process parameter optimization through Decision Tree Regression analysis. However, limitations exist in the reliance on historical data and potential model overfitting. Future work could involve exploring real-time data integration for more dynamic parameter optimization and incorporating additional machine learning algorithms to enhance predictive accuracy, thereby further improving process efficiency and overall effectiveness.

Funding

Not applicable

Author Contribution

Conceptualization, T. N. and S. M.; writing—original draft preparation, T. N. and Z. D.; writing—review and editing, S. M. and Z. D.; All of the authors read and agreed to the published final manuscript.

Data Availability Statement

The data can be accessible upon request.

Conflict of Interest

The authors confirm that there are no conflict of interests.

Reference

References:
[1] H. Y. Chia et al., "Process parameter optimization of metal additive manufacturing: a review and outlook," Journal of Materials Informatics, 2022.
[2] M. H. Ali et al., "Powder bed fusion-laser melting (PBF-LM) process: latest review of materials, process parameter optimization, application, and up-to-date innovative technologies," Progress in Additive Manufacturing, 2022.
[3] Y.-T. Zhang et al., "Application of a Bio-Inspired Algorithm in the Process Parameter

Optimization of Laser Cladding," Machines, 2022.
[4] K. Kalita et al., "A Hybrid TOPSIS-PR-GWO Approach for Multi-objective Process Parameter Optimization," Process Integration and Optimization for Sustainability, 2022.
[5] S. Dharmadhikari et al., "A Reinforcement Learning Approach for Process Parameter Optimization in Additive Manufacturing," Additive Manufacturing, 2022.
[6] A. Dey and N. Yodo, "A Systematic Survey of FDM Process Parameter Optimization and Their Influence on Part Characteristics," Journal of Manufacturing and Materials Processing, 2019.
[7] N. Shamsaei et al., "An overview of Direct Laser Deposition for additive manufacturing; Part II: Mechanical behavior, process parameter optimization and control," Additive manufacturing, 2015.
[8] S. Wu et al., "Process parameter optimization and EBSD analysis of Ni60A-25% WC laser cladding," International Journal of Refractory Metals & Hard Materials, 2021.
[9] K. O'Dowd and S. Pillai, "Photo-Fenton disinfection at near neutral pH: Process, parameter optimization and recent advances," Journal of Environmental Chemical Engineering, 2020.
[10] P. Koirala and C. Fleming, "Solving Offline Reinforcement Learning with Decision Tree Regression," 2023.
[11] S. C. Tan, "Decision Tree Regression with Residual Outlier Detection," Journal of Data Science and Intelligent Systems, 2023.
[12] M. F. Javaid et al., "Decision tree regression algorithm-based model for prediction of traffic-induced vibrations," Innovative Infrastructure Solutions, 2023.
[13] E. Jumin et al., "Solar radiation prediction using boosted decision tree regression model: A case study in Malaysia," Environmental science and pollution research international, 2021.
[14] M. Abdurohman et al., "A Robust Internet of Things-based Aquarium Control System using Decision Tree Regression Algorithm," IEEE Access, 2022.
[15] A. Balogun and A. Tella, "Modelling and investigating the impacts of climatic variables on ozone concentration in Malaysia using correlation analysis with random forest, decision tree regression, linear regression, and support vector regression," Chemosphere, 2022.
[16] E. Pekel, "Estimation of soil moisture using decision tree regression," Theoretical and Applied Climatology, 2019.
[17] R. Karim et al., "Stock Market Analysis Using Linear Regression and Decision Tree Regression," 2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA), 2021.
[18] S. Latif, "Developing a boosted decision tree regression prediction model as a sustainable tool for compressive strength of environmentally friendly concrete," Environmental science and pollution research international, 2021.
[19] M. Rakhra et al., "Crop Price Prediction Using Random Forest and Decision Tree Regression:-A Review," Materials Today: Proceedings, 2021.
[20] Z. Luo, H. Yan, and X. Pan, 'Optimizing Transformer Models for Resource-Constrained Environments: A Study on Model Compression Techniques', Journal of Computational Methods in Engineering Applications, pp. 1–12, Nov. 2023, doi: 10.62836/jcmea.v3i1.030107.
[21] H. Yan and D. Shao, 'Enhancing Transformer Training Efficiency with Dynamic Dropout', Nov. 05, 2024, arXiv: arXiv:2411.03236. doi: 10.48550/arXiv.2411.03236.
[22] Y. Gan and D. Zhu, 'The Research on Intelligent News Advertisement Recommendation Algorithm Based on Prompt Learning in End-to-End Large Language Model Architecture',

Innovations in Applied Engineering and Technology, pp. 1–19, 2023.
[23] D. Zhu, Y. Gan, and X. Chen, ‘Domain Adaptation-Based Machine Learning Framework for Customer Churn Prediction Across Varing Distributions’, Journal of Computational Methods in Engineering Applications, pp. 1–14, 2021.
[24] X. Deng, L. Li, M. Enomoto, and Y. Kawano, ‘Continuously frequency-tuneable plasmonic structures for terahertz bio-sensing and spectroscopy’, Scientific reports, vol. 9, no. 1, p. 3498, 2019.
[25] X. Deng, M. Simanullang, and Y. Kawano, ‘Ge-core/a-si-shell nanowire-based field-effect transistor for sensitive terahertz detection’, in Photonics, MDPI, 2018, p. 13.
[26] H. Zhang, D. Zhu, Y. Gan, and S. Xiong, ‘End-to-End Learning-Based Study on the Mamba-ECANet Model for Data Security Intrusion Detection’, Journal of Information, Technology and Policy, pp. 1–17, 2023.
[27] D. Zhu, X. Chen, and Y. Gan, ‘A Multi-Model Output Fusion Strategy Based on Various Machine Learning Techniques for Product Price Prediction’, Journal of Electronic & Information Systems, vol. 4, no. 1.
[28] X. Deng and Y. Kawano, ‘Surface plasmon polariton graphene midinfrared photodetector with multifrequency resonance’, Journal of Nanophotonics, vol. 12, no. 2, pp. 026017–026017, 2018.