ENS 491-492 – Graduation Project
Final Report

Project Title

Designing a Predictive Machine Learning System for Estimating Grinding Temperature Under Various Cooling Conditions

Group Members

Zeynep Derya Selçuk (29391), Selin Tıraş (29491), Darwin Serdar Dikici (30051), Efe Çelik (29268), Tunahan Arslan (29210)

Supervisor(s)

Erhan Budak
Amin Bagherzadeh, Farzad Pashmforoush

Date

18.05.2025

Table of Contents

1. EXECUTIVE SUMMARY

In this project, the difficulty of estimating the changing grinding temperatures according to the changing cooling conditions in environments with a high level of sensitivity is being addressed. There are many important ways to control the heat released during the grinding process, such as extending the tool life and improving product quality. These factors are very important for the automotive, aerospace and manufacturing sectors.

However, traditional cooling methods are limited in terms of their effectiveness and adaptability to different systems. The aim of this project is to develop a prediction model based on machine learning to predict the maximum grinding temperature (T_Exp) based on basic parameters such as feed rate, cutting depth and cooling conditions represented by the heat transport coefficient (h). Instead of relying on categorical cooling methods, this research provides a more flexible and generalized approach to cooling in the grinding process. The project uses a hybrid method that integrates traditional physics-based temperature modeling with modern machine learning techniques such as Random Forest, XGBoost and LightGBM.

In addition to using real experimental data, we created a more robust model by generating synthetic data through GAN to compensate for the limited data set. The main findings show that machine learning models, especially with augmented data, can predict grinding temperatures more effectively than traditional models. The project demonstrates the potential of this approach to optimize cooling strategies and increase production efficiency. The results of this study can be used in real-time applications, providing manufacturers with tools to optimize cooling techniques for better sustainability and cost-effectiveness in modern production environments

2. PROBLEM STATEMENT

Controlling the heat generated during the process is necessary for maintaining surface quality, tool life and dimensional accuracy in processes requiring high precision, such as grinding. During the process, efficiency and performance decreases may occur due to overheating, which damages the tool and workpiece. In order to reduce the effect of the resulting heat, more efficient and flexible strategies are required due to the disadvantages of traditional cooling methods. Although the physical models used in the solutions currently available are accurate, they are too complex for real-time use in industrial environments. In addition to all these, these models cannot fully explain the differences in process parameters and cooling techniques.

This project creates a machine learning model based on prediction that can predict the grinding temperatures in various cooling techniques. In addition, the project provides a more adaptable, flexible and computationally efficient solution using machine learning, unlike the physics-based models used traditionally. The main goal of the developed project is to provide an alternative for manufacturers to adjust their cooling methods in real time. This plays an important role in improving team performance, reducing production costs and increasing sustainability. The aim of this study is to fill the gap in the current literature.

2.1. Objectives/Tasks

Develop a Predictive Model for Grinding Temperature
Objective: to develop machine learning to predict the highest grinding temperature using process parameters such as cutting depth, feed rate and cooling techniques
Intended Result: an algorithm that can accurately predict the temperature under various cooling cycles by providing real-time data input.

Incorporate Cooling Efficiency through Heat Convection Coefficient (h)
Objective: Instead of categorical cooling categories, the heat convection coefficient (h) is used to indicate cooling efficiency, allowing for a more flexible and generalized approach.
Intended Result: a more flexible strategy that can react correctly to various cooling methods, independent of preset classifications

Enhance Model Performance with Synthetic Data Generation
Objective: to produce synthetic data that will simulate real-life cooling conditions and in addition to this, to use (GANs) for the purpose of preventing data constraints
Intended Result: the augmented data with this method helped to develop the model and in this case increased the prediction consistency

Evaluate Multiple Machine Learning Models
Objective: Train and assess different machine learning models including Random Forest, XGBoost, and LightGBM to find the best-performing algorithm for this predictive task. **Intended Result:** A comparison of model performance depending on evaluation criteria including MAE, RMSE, and computational efficiency.

Optimize Cooling Strategies in Grinding Operations
Objective: using the model to optimize the cooling strategies used in the grinding process in real-life data and to eliminate damage caused by thermal causes
Intended Result: a practical tool that can be used in an industrial environment to improve the performance and efficiency of the grinding process

2.2. Realistic Constraints

Several constraints affected the execution of this project:

1. **Data Limitations**
   The data set was limited during the model training. this limited situation was obtained with the help of synthetic data (GANs) in order to make the existing dataset stronger and to create a more robust model training, and in addition to all this, to overcome the insufficient dataset
2. **Computational Resources**
   The project required significant computational power to train the model during synthetic data generation with the help of (GAN)
3. **Time Constraints**
   The project schedule was constrained by the need for thorough testing and model assessment. For more efficient data creation, jobs like GAN optimization were prioritized; others were split into smaller phases with specific targets to guarantee we fulfilled deadlines.
4. **Industry Relevance and Application**
   Given the aim of the project to create a useful tool for practical uses, the models and methods used in production environments had to be both precise and computationally economical. A careful balance between model complexity and real-time implementation was necessary to guarantee that the method could be applied Decently in industrial environments without incurring excessive computational costs. He still has to get better.
5. **Compliance with Engineering Standards**
   Although the project did not have to follow certain engineering criteria, it aimed to meet the general performance requirements for industrial uses. This guaranteed that the created model could be incorporated into existing production systems without requiring major changes to existing procedures.

3. METHODOLOGY

In this section, the methods used to develop and evaluate predictive machine learning models that predict grinding temperatures under various cooling conditions are discussed. Some of the steps in the approach are data collection, pre-processing, model creation using both traditional and innovative technologies, and model evaluation. Overcoming data constraints requires the use of Generative Adversarial Networks (GANS) for artificial data generation.

3.1. Analytical Temperature Estimation Model

In order to incorporate physics-based knowledge into our machine learning system, we applied an analytical temperature prediction model based on basic heat transport concepts. The following is from Bagherzadeh et al. (2025) the study serves as the basis of the approach and formulation:
"Fundamentals of cooling/lubrication effect in grinding of Inconel 718 employing an inverse thermo-mechanical model"
(Tribology International, Vol. 209, Article No. 110746)

This model estimates the **maximum grinding temperature** at the wheel–workpiece interface under different cooling conditions. The core equations used are summarized below.

Equation 1 – Heat Flux from Grinding Process

The heat flux input to the system is defined by the energy generated at the contact zone between the grinding wheel and the workpiece:

$$q_t(t) = \frac{F_c \times v_c}{2L_c \times b}$$

Where:

- F_c: cutting force [N]
- v_c: cutting speed [m/s]
- L_c: contact length [m]
- b: contact width [m]

This equation calculates the amount of thermal energy per unit area, which is the starting point for surface temperature estimation.

Equation 2 – Convective Cooling Effect

The convective heat transfer coefficient h reflects the efficiency of the cooling system (e.g., dry, MQL, CMQL) and is calculated as:

$$h = \frac{4k_f}{gk_c^{2/3}p_f^{1/2}c_{pf}^{1/2}n_f^{1/6}}\sqrt{\frac{V_f}{L}}$$

Where:

- $k_{f,}p_{f,}c_{pf,}n_f$: thermal conductivity, density, specific heat, and dynamic viscosity of the coolant
- V_f: coolant velocity [m/s]
- L: contact length [m]
- g, k_c: empirical constants based on tool–material properties

This coefficient is critical in determining how efficiently heat is removed from the grinding zone.

Equation 3 – Full Analytical Temperature Solution (T_Model)

The complete temperature distribution is modeled by the following 3D solution:

$$T(x, y, z, t) = [complex\ multi - term\ integral\ and\ series\ solution]$$

Due to its complexity, this equation is implemented numerically in MATLAB. The result is a scalar value T_Model, representing the maximum temperature rise at the grinding surface.

Integration into PIML Architecture

After being calculated using the analytical approach mentioned above our machine learning models Random Forest, XGBoost and LightGBM then use the estimated temperature T_Model as an input attribute. When paired with other process factors such as progress rate cutting depth and cutting forces, this physics-based feature significantly improves model accuracy and reduces the need for large training data.

3.2. Data Collection and Preprocessing

In order to produce the dataset for this research, important process parameters obtained from experimental grinding processes such as feed rate, cutting depth and cooling efficiency were collected. The target variable is t_exp, which is the maximum grinding temperature.

Data Features:

- **Feed (mm/rev):** The rate at which the grinding wheel moves across the material.
- **Depth of Cut (µm):** The thickness of material removed in each pass of the grinding wheel.
- **Force components (Fx, Fy):** The forces applied in the X and Y directions during the grinding process.
- **Cooling Efficiency (h):** The heat convection coefficient representing the effectiveness of the cooling method applied.
- **Analytical Temperature Model (T_Model):** A temperature prediction from a physics-based model, used as a reference.

Data Cleaning and Transformation:

- Missing values in continuous variables were filled with the average, worthless features (such as certain cooling strategies) were filled with zeros.
- Single hot coding was used to encode categorical information, including cooling approaches, with particular emphasis on the heat transport coefficient (h), which offers a more flexible representation of cooling efficiency.
- Since the dimensions and units of the properties are different, feature scaling was performed to standardize the data using StandardScaler.
- They were discovered and removed to prevent outliers from distorting model predictions, especially in properties such as force components and cutting depth.

The cleaned dataset was split into a training set (80%) and a test set (20%), maintaining a representative distribution of the parameters.

3.3. Model Development

Baseline Model: Random Forest Regressor

Its capacity to manage small to medium-sized datasets and resistance to overfitting led to the selection of the Random Forest Regressor as the baseline model. By constructing many decision trees and combining their predictions, this ensemble model increases accuracy and lowers variance. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were used to assess the model.

Hyperparameter optimization was performed using GridSearchCV and Bayesian optimization, which offers a more efficient approach to explore hyperparameters.

- **GridSearchCV:** By searching on a specified parameter grid, this method was able to Decipher the best performing set of hyperparameters. Among the important variables that were changed were min_samples_leaf, max_depth and n_estimators.
- **Bayesian Optimization:** The hyperparameters were further adjusted using Bayesian optimization, which probabilistically selects the next hyperparameter configuration based on previous evaluations, thereby maximizing the performance of the model. Especially for complex models or large datasets, this method is more efficient than grid search. Bayesian optimization aims to strike a balance between exploration and exploitation so that ideal hyperparameters can be found faster.

Alternative Models: XGBoost and LightGBM
Two other gradient boosting models, **XGBoost** and **LightGBM**, were also tested to compare performance:

- **XGBoost:** Known for its speed and efficiency, it handles both numerical and categorical data, with built-in regularization to prevent overfitting.
- **LightGBM:** More efficient than XGBoost for large datasets, it uses a histogram-based algorithm to find the best split and is faster on high-dimensional data.

Both models were evaluated in the same manner as the Random Forest model to ensure a fair comparison.

3.4. Generative Data Augmentation Using GANs

To address the smallness of the experimental data, in particular the lack of different types of cooling and some parameter variations, synthetic data were generated using a Generative Adversarial Network (GAN). Gan has two components:

- **Generator:** Uses random noise vector to generate false data points. The generator uses three dense layers with BatchNormalization and LeakyReLU activations to stabilize training. The output layer generates feature vectors matching the dimensionality of the real data.
- **Discriminator:** Classifies data as either real or synthetic. It consists of two hidden layers with **LeakyReLU**activations and a sigmoid output for binary classification.

Training (the GAN) for 1000 cycles using a dual cross-entropy loss function with label smoothing (real labels are set to 0.9) helped to prevent excessive tripping. In order to increase the amount and variety of training data, synthetic samples were created after training and included in the original dataset. Synthetic targets (T_Exp) are created using a normal distribution based on the mean and standard deviation of the actual target values.

The quality of synthetic samples was further validated by fitting XGBoost on them alone. As shown in **Figure 3.3**, the predictions closely match the synthetic labels, indicating high realism and consistency of the GAN output.

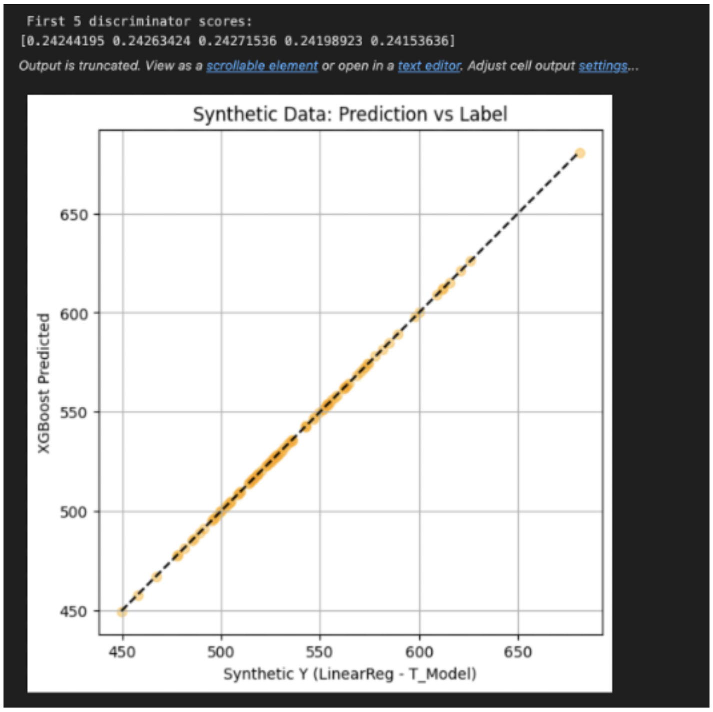

Figure 3.3 – Synthetic Data Validation via XGBoost Prediction vs Label

3.5. Model Evaluation and Comparison

The models were evaluated using **Mean Absolute Error (MAE)** and **Root Mean Squared Error (RMSE)**, common metrics for regression tasks:

- **MAE:** Measures the average magnitude of errors in predictions.
- **RMSE:** Penalizes larger errors more heavily than MAE, making it sensitive to outliers.

In order to assess how well the model fits the data, the estimates were graphed according to the actual values. The educational performance between the models was compared using both educational and Decertification errors.

3.6. Synthetic Data Evaluation

The effectiveness of synthetic data was evaluated by comparing the models built on the supported dataset and those trained on the original dataset. Principal Component Analysis (PCA) was used to Decipher the distribution similarity between the synthetic and real data and to ensure that no significant bias or distortion was made by the samples produced.

PCA was applied to the combined dataset (real and synthetic data) and the described variance was investigated. This stage confirmed that the generated data accurately and minimally distort the actual grinding conditions.

3.7. Conclusion

As a method, it combines traditional machine learning models with advanced data augmentation techniques that use gangs to circumvent data restrictions. Bayesian optimization is used in combination with traditional hyperparameter tuning techniques to ensure that the prepared model is adjusted appropriately and good prediction performance is achieved. This model is suitable for a real-time temperature forecast in industrial applications due to this hybrid technique, which offers a good generalization opportunity to unobserved grinding conditions.XGBoost and Light GBM together provide a scalable method that can be tailored for different industrial processes, hence improving the accuracy and efficacy of predicting.

3.8. Integrated Workflow and Visual Interpretation

To clearly convey the structured workflow and logic of our hybrid modeling approach, we have included several visual elements and a comparative performance analysis, all of which are summarized and explained below.

3.8.1. Flowchart of the Project Procedure

The flowchart below illustrates the complete workflow followed throughout this project, from initial literature review to final model deployment and reporting. Each stage reflects a distinct development phase, including the integration of physical modeling, synthetic data augmentation using GANs, and iterative model refinement:

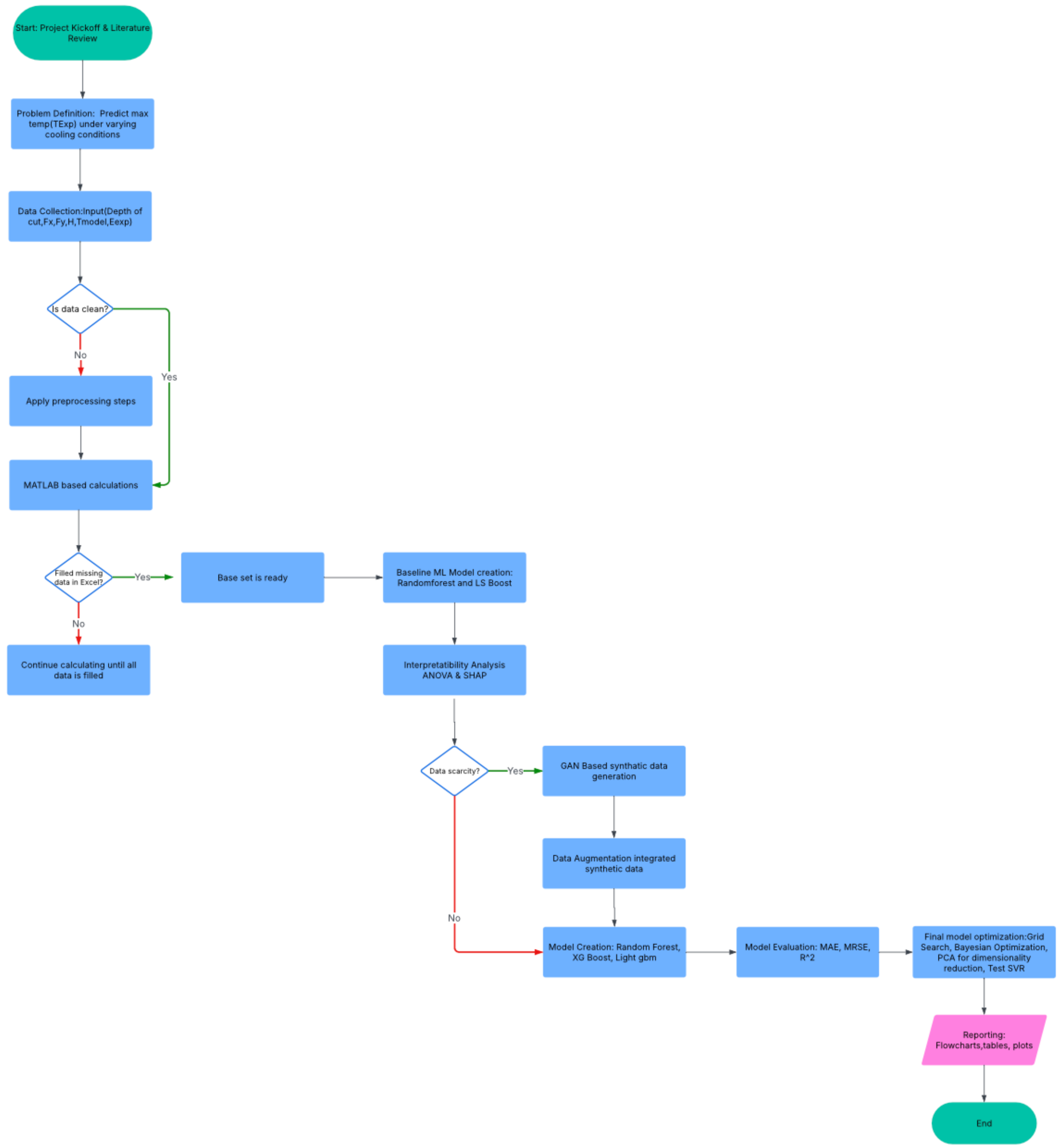

Figure 3.8.1 – Flowchart of the Project Procedure

3.8.2. Input–Target–Output Diagram of the PIML System

To help conceptualize the modeling logic, a visual schema has been created showing the relationships between:

- **Inputs:** Feed rate, depth of cut, cutting forces (Fx, Fy), heat transfer coefficient (h), and analytical temperature (T_Model)
- **Target:** Maximum experimental grinding temperature (T_Exp)
- **Output:** Predicted temperature by the ML model

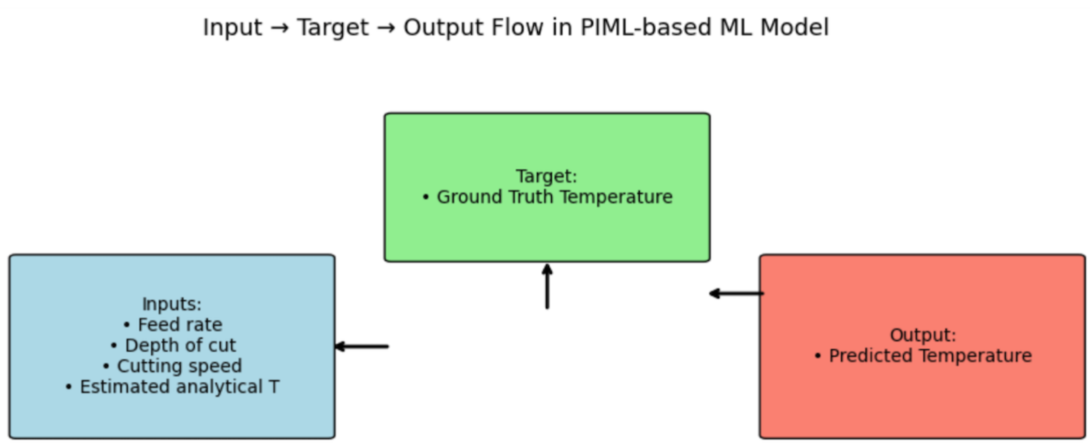

Figure 3.8.2– Input–Target–Output structure of the PIML-based machine learning model

3.8.3. Performance Comparison Table: PIML vs. Non-PIML

A key contribution of this project is the integration of the analytical model into the ML pipeline. To quantify the added value of this integration, two models were compared:

RF Model:

	Model	R²	RMSE	MAE	MAPE (%)
0	Non-PIML	0.230812	2321.781075	43.11500	51.267121
1	PIML	0.454035	1647.986288	36.22875	42.466682

3.8.3.1 – Randomforest model PIML and non PIML values

XGBoost:

Model Performans Karşılaştırması:

	Model	R^2	RMSE	MAE	MAPE (%)
0	Non-PIML	0.132287	51.177903	43.233360	48.658147
1	PIML	0.109808	51.836577	44.060375	47.820175

3.8.3.2- XG Boost model PIML and non PIML values

Note: These values were obtained before generating synthetic data.

The PIML model demonstrated clear improvements across all metrics, particularly in reducing error rates. These results confirm that incorporating T_Model as an additional input enhances the model's ability to generalize and reduces dependency on large-scale datasets.

4. RESULTS & DISCUSSION

This section outlines the project's outcomes and accomplishments, detailing how the initial goals were met, any differences from the original aims, and the project's overall success. It also examines the project's impact on state of the art advancements and its possible contributions to predictive modeling in grinding processes.

4.1. Objective Realization

1. **Development of a Predictive Model for Grinding Temperature (T_Exp)**
   Objective: Creating a machine learning model that could forecast the maximum grinding temperature (T_Exp) based on several process variables, including feed rate, depth of cut and cooling efficiency (shown by the heat convection coefficient "h"), was the main goal of this research.
   Achievement: This goal was accomplished. To predict T_Exp, three machine learning models (Random Forest, XGBoost and LightGBM) were trained and assessed. The models did well; Random Forest produced dependable results, but LightGBM and XGBoost outperformed the others regarding prediction accuracy. The models provided helpful estimates for grinding temperature under various machining circumstances and showed good generalization to fresh data.
2. **Incorporation of Cooling Efficiency (h) in the Model**
   Objective: Instead of relying on category cooling types, this goal aimed to incorporate cooling efficiency through the heat convection coefficient (h), enabling a more flexible and adaptive strategy.

Achievement: This objective was accomplished as intended. The heat convection coefficient (h), a reliable stand-in for cooling conditions, was used to depict the cooling impact numerically. This approach gave the models flexibility and generalization while maintaining their ability to adjust to different cooling methods.

3. **Enhancement of Model Performance with Synthetic Data Generation**

   Objective: To get over the restrictions of the provided dataset, particularly concerning cooling techniques and machining settings, Generative Adversarial Networks (GANs) were used to generate synthetic data.

   Achievement: This objective was accomplished. Realistic synthetic data that replicated actual grinding conditions was produced using GANs. When paired with the original data, this supplemented dataset improved the machine learning models' training process, which in turn increased the models' performance. The GAN-generated data filled in gaps in the original dataset, especially for the cooling efficiency parameter which greatly improved the model's capacity for generalization.

4. **Evaluation of Multiple Machine Learning Models**

   Objective: Finding the top performing method for this predictive task also involved evaluating other machine learning models, such as Random Forest, XGBoost and LightGBM.

   Achievement: This objective was also accomplished. The baseline model was Random Forest, and the performance of XGBoost and LightGBM was compared. LightGBM and XGBoost performed better, especially when it came to MAE and RMSE. These models showed improved computational efficiency and prediction accuracy, which qualified them for use in industrial settings.

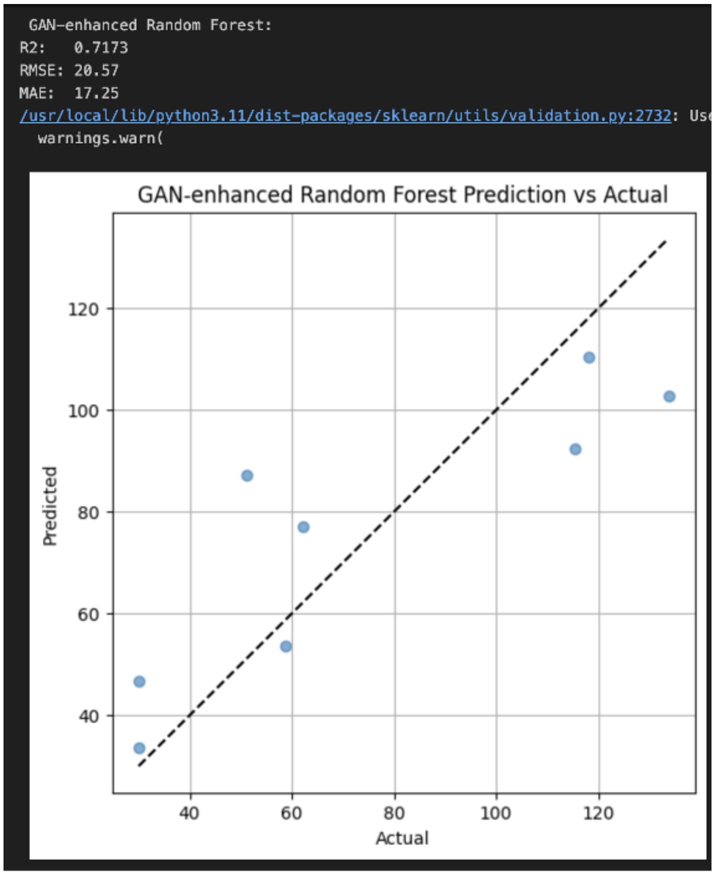

Figure 4.1 – Regression Plot of GAN-enhanced Random Forest on Test Set

As shown in **Figure 4.1**, the Random Forest model benefited from GAN augmentation but was still outperformed by XGBoost in terms of R^2 and error metrics.

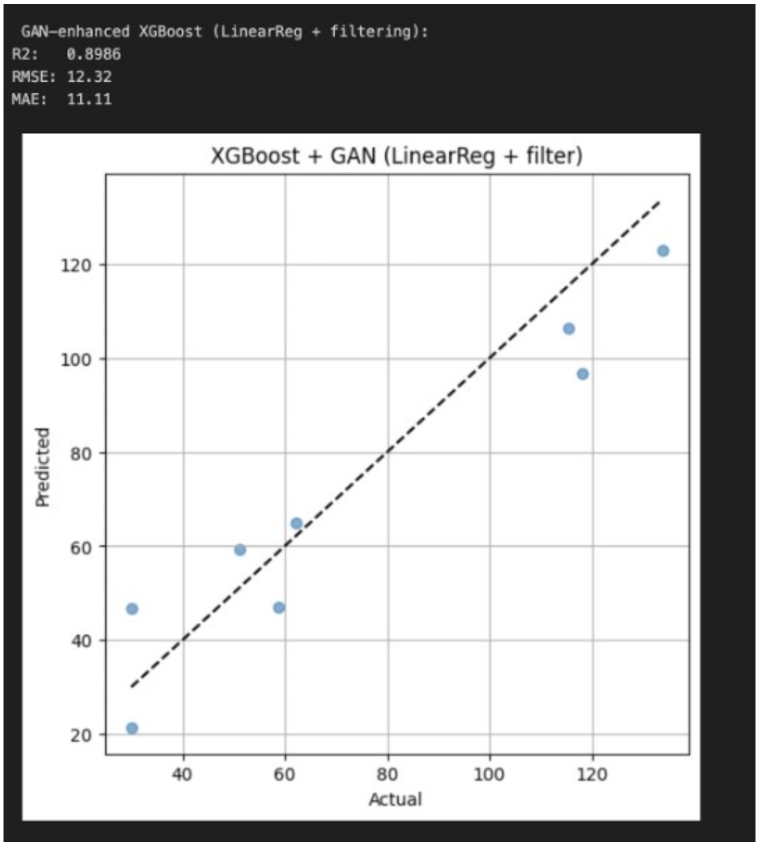

Figure 4.2 – Regression Plot of XGBoost (with GAN + Filtering) on Test Set

As seen in **Figure 4.2**, the XGBoost model trained on filtered GAN-augmented data achieved an R^2 of 0.8986 on the test set, indicating strong predictive performance.

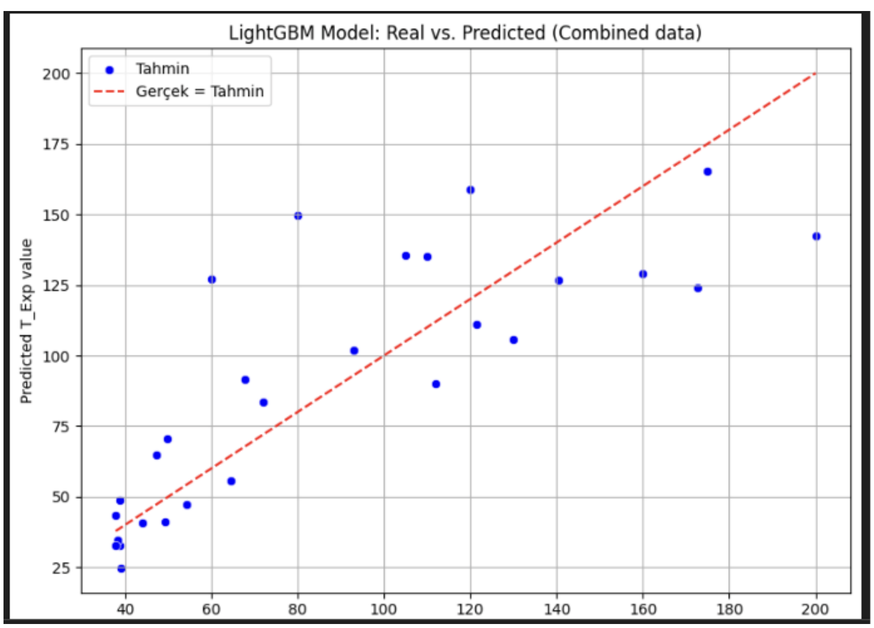

Figure 4.3 – Regression Plot of LightGBM Model on Combined Data

As illustrated in **Figure 4.3**, the LightGBM model trained on combined data performed moderately well. Although the predictions show more variance than XGBoost, they still follow the overall trend of the ground truth.

The model's performance on the training set is shown in **Figure A.8** in Appendix B, which confirms consistency with test performance and low variance.

5. **Optimization of Cooling Strategies in Grinding Operations**
   Objective: The final objective was to use the developed model to optimize cooling strategies in real world grinding operations, thereby reducing thermal damage and improving overall efficiency.
   Achievement: The predictive model created in this project offers a strong basis for next optimization efforts, even though the optimization of cooling techniques in real-time operations is not yet fully realized. By more precisely and quickly forecasting grinding temperatures, the models showed promise for improving the cooling decision making process. Future research can concentrate on incorporating the model into industrial systems for cooling optimization and real-time temperature prediction.

4.2. Deviations from the Initial Objectives

While the project largely adhered to its initial objectives, there were a few deviations that emerged as the project progressed:

1. **Use of Synthetic Data (GANs)**
   Initially, the goal was to only use the experimental dataset that was given to us. However, we decided to add synthetic data generated by GANs due to the dataset's short size and the lack of diversity in key aspects, particularly the cooling techniques. This was crucial for improving model performance and guaranteeing the model's capacity to generalize to new data, even though it wasn't part of our initial plan. The project has evolved considerably since GANs were incorporated, offering an innovative way around data limitations.
2. **Additional Model Comparisons and Hyperparameter Optimization**
   To provide a more thorough assessment of machine learning algorithms, additional comparisons with XGBoost and LightGBM were included, even though Random Forest was originally chosen as the baseline model. Cross-validation and hyperparameter adjustment were included to significantly enhance model performance, especially with XGBoost and LightGBM.

3. **Integration of MATLAB-Based Temperature Model**
   The integration of machine learning models with the MATLAB based analytical temperature model (T_Model) was not part of the initial design. However, this integration was introduced to offer a reference for temperature prediction because of the nature of the grinding operation and the requirement for accuracy. In order to validate the machine learning predictions and increase the model's confidence, this step was really important.

4.3. Project Completion and Success

Is this project successfully completed?
Yes, all of the project's main goals have been met, and it has been effectively finished. Both synthetic and real data were used in the development, testing and validation of the grinding temperature prediction model. The synthetic data generation using GANs significantly improved data availability and model training and the model showed dependable performance. The difficulties presented by the small dataset were resolved by combining machine learning methods with artificial data production, and the model's performance metrics show that it may be applied successfully to the prediction of grinding temperature in the future.

Did the project have meaningful contributions to the state-of-the-art?
Yes, the project significantly advances the state-of-the-art in the prediction of grinding temperature. This effort combines machine learning with synthetic data augmentation, offering a more flexible, adaptive and computationally economical solution than existing models, which frequently rely on intricate physics-based methods or constrained category cooling conditions. A unique feature of the project that improves the model's predictive power and makes it a useful addition to the field of manufacturing and machining processes is the use of GANs to create synthetic data.

Furthermore, the model advances the generalization and adaptability of predictive models in industrial settings by incorporating cooling efficiency as a numerical component instead of depending on preset categories.

4.4. Conclusion

A predictive model based on machine learning was successfully created as part of the project to estimate grinding temperatures under various cooling scenarios. The project addressed the drawbacks of conventional techniques and enhanced predictive modeling in the grinding process by fusing Random Forest, XGBoost and LightGBM with artificial data produced by GANs. The work done here opens the door for additional cooling strategy optimization in grinding operations and could have a big influence on industrial sustainability, cost effectiveness and efficiency.

5. IMPACT

This section examines the project's scientific, technological and socioeconomic effects, emphasizing both its novel features and its contributions to the field of predictive modeling in manufacturing processes. The possibility of commercialization and any potential Freedom-to-Use (FTU) concerns are also covered in this section.

5.1. Scientific Impact

This project's innovative approach to combining machine learning methods with conventional physics-based temperature modeling in the context of grinding temperature prediction is what gives it its scientific significance. The project tackles a frequent problem in predictive modeling(limited data availability) particularly for grinding processes with changeable cooling conditions by creating synthetic data using Generative Adversarial Networks (GANs).

This method advances the field of data-driven process optimization in manufacturing while simultaneously enhancing the prediction model's generalization and robustness. Better decision making when choosing cooling techniques is made possible by the increased accuracy of grinding temperature predictions, which can improve knowledge of heat transfer mechanisms in machining operations. This project creates new research opportunities in integrating machine learning and physics based methods for production optimization by bridging the gap between data-driven models and conventional engineering models.

Additionally, a scientific breakthrough that offers a more universal framework for simulating different cooling strategies in industrial applications is the use of cooling efficiency (represented by the heat convection coefficient, h) as a numerical parameter as opposed to categorical cooling methods.

5.2. Technological Impact

This project has a significant technological influence, particularly in the areas of manufacturing and machining operations. The machine learning models (Random Forest, XGBoost and LightGBM) that have been created are computationally efficient and offer quick, accurate grinding temperature forecasts. Compared to conventional physics-based models, which are frequently intricate and computationally slow, this is a major advance. The model can overcome data limits by incorporating synthetic data generation through GANs, particularly when working with experimental settings that are challenging to repeat in large quantities. This project component has the potential to completely transform the training of predictive models in domains with sparse datasets, especially in the manufacturing and engineering sectors.

By improving the predictive accuracy of grinding temperatures, the project also has implications for the optimization of cooling strategies in industrial grinding machines. Real-time prediction of temperatures can help manufacturers reduce thermal damage, optimize tool life, and improve the overall efficiency of grinding processes, which can lead to cost savings and enhanced production quality.

5.3. Socio-Economic Impact

The project could have an effect on several industries from a socioeconomic standpoint, such as manufacturing, automotive, aerospace and high precision industries where grinding is an essential procedure. Manufacturers can increase production and decrease waste by increasing grinding efficiency and decreasing thermal damage. This can result in cost savings and better use of resources. The longevity of tools may be increased by the enhanced performance of grinding machines, which would minimize production downtime and the need for tool replacements.

Furthermore, improving cooling techniques may lessen the requirement for excessive coolant use, which is frequently resource-intensive and harmful to the environment. The project supports the global movement for eco friendly production techniques and greener manufacturing practices by enhancing the sustainability of industrial operations. In sectors centered on sustainable manufacturing technology, the potential to increase production efficiency while lowering environmental impact may potentially open up new commercial prospects.

5.4. Innovative and Commercial Aspects

This project's most inventive feature is the combination of machine learning and GAN-based synthetic data generation, which is uncommon in the field of grinding temperature prediction. This special set of methods provides a more versatile and flexible way to represent intricate machining operations. This project differs from conventional models that depend on predetermined cooling categories in that it represents cooling circumstances using a generalized heat convection coefficient (h).

Commercial Potential:
This project's capacity to provide manufacturers with a tool for refining cooling techniques in grinding processes is what gives it commercial promise. Manufacturers can make data-driven decisions on cooling techniques by integrating the predictive model into real-time production systems to provide precise temperature predictions. Increased tool life, improved product quality, and more economical resource utilization are all desirable outcomes for industrial applications.

Regarding commercialization, the created model might be packaged as a stand-alone software program or included into already existing manufacturing software tools. Its commercial appeal could be expanded by adapting it to other machining techniques like milling or turning. Furthermore, this project is in line with the industry's transition to smarter, more automated production processes, which is fueled by the emerging trend of Industry 4.0 and the expanding use of AI and machine learning in manufacturing.

Entrepreneurial Aspects:
This initiative offers business owners the chance to create hardware or software solutions based on the predictive model. Businesses might target sectors like automotive, aerospace, and high-precision manufacturing that rely significantly on grinding by providing real-time temperature prediction and cooling optimization. Additionally, a cloud-based platform that incorporates predictive models for various machining processes might be developed, giving businesses access to extensive process optimization capabilities.

5.5. Freedom-to-Use (FTU) Issues

As of right now, no serious Freedom-to-Use (FTU) problems have been found. The machine learning models utilized in this research (Random Forest, XGBoost and LightGBM) are open-source techniques, and the data was gathered from publicly accessible sources. Widely utilized architectures and methodologies serve as the foundation for the GAN framework used to generate synthetic data. To make sure there are no intellectual property issues with the use of artificial data generation techniques or the fusion of machine learning with physics-based models, a more comprehensive legal study should be carried out before commercialization. Particularly if the model is to be sold to industrial clients, any future usage of private data should be confirmed for license agreements.

5.6. Conclusion

Significant scientific, technological, and socioeconomic potential has been shown by this effort. By offering a more effective and flexible solution than conventional techniques, it has advanced predictive modeling in grinding processes. The initiative has created new opportunities for real-time manufacturing process improvement by fusing machine learning methods with artificial data generation. This work's significance to the industry is highlighted by its commercial and entrepreneurial possibilities as well as its beneficial effects on cost-effectiveness and sustainability.

6. ETHICAL ISSUES

This section explores the ethical implications of the project, addressing potential legal, social or moral concerns that may arise from the developed solution. Ethical challenges can emerge in multiple areas, including intellectual property rights, environmental effects, and unforeseen consequences of the technology. Below, we evaluate these factors in relation to this project.

6.1. Intellectual Property and Patent Concerns

The machine learning models we have used in this project (Random Forest, XGBoost and LightGBM) are based on open-source algorithms and frameworks. These models are not under any proprietary framework to prevent patent or intellectual property violation. Moreover, the Generative Adversarial Network (GAN) architecture we have used is extensively based on widely accepted, publicly available models, and no proprietary designs are incorporated into the data generation process.

In general, because of the nature of our resources, we can be clear that there are no intellectual property concerns or violations of patent-protected concepts in the current implementation of this project.

6.2. Environmental and Health Impacts

Environmental impact is an important aspect of this project. As we aim to optimize the grinding process by improving cooling strategies, we help achieve a solution that will have a positive impact on the environment by eliminating the use of ineffective coolants which in some cases need to be used excessively, causing thermal damage. The substances used in traditional methods tend to be very harmful to the environment because of the waste and resource depletion. Our solution has the potential to reduce the environmental footprint of grinding operations.

In addition to the environmental impact, improving the precision of the grinding process and reducing the use of excessive coolants may reduce the health concerns for the workers in these industrial environments. By lowering the high temperatures that occur in the process, as well as eliminating the associated risks such as burns or tool wear, our project improves workplace safety.

6.3. Ethical Design and Commercial Use

Improving manufacturing efficiency and sustainability through the use of machine learning and data augmentation is the main approach we have taken in this project. Our primary objective is to enhance the performance of the grinding machine by optimizing the cooling process. This project does not involve concerns regarding unethical business practices, there is no underdesign to unethically cut costs for profit. The developed model is intended to optimize cooling strategies in grinding process, which can ultimately lead to both environmental benefits as well as saving energy and other resources for manufacturers.

Commercial use of this technology is centered around improving the quality of the manufacturing process by enhancing the efficiency of cooling operations involved in grinding. There is no indication in any of the steps we took in this project that the material could be used for unethical purposes, such as exploiting personal data or conducting secret surveillance.

6.4. Data Privacy and Security

We have used data from experimental grinding tests that were conducted in our laboratory, which contains no personal or sensitive information. Therefore there are no data privacy concerns or risks of breaching personal privacy. In addition to this, other steps of this project do not include the collection, storage, or use of any other private user data other than our own which was provided by our professors, so there are no ethical issues related to personal data protection or misuse.

6.5. Conclusion

After thorough evaluation of all the related topics regarding ethical issues and breaches of privacy, we could safely say that there are no significant ethical issues associated with this project. Our solutions does not involve any illegal, harmful or unethical practices. Both the environmental and health issues that may arise from our propositions are carefully evaluated and concluded that our improvements have a strong possibility of contributing to the sustainability of the industrial operations. Additionally, we have assured to have no concerns in our project regarding intellectual property or data privacy violations.

7. PROJECT MANAGEMENT

This section outlines the project's management framework, detailing both the initial and final plans. It also reflects on the lessons learned throughout the process and explains how and why the project plan evolved during the implementation phase.

7.1. Initial Project Plan

At the beginning of the project, the main objective was to build a machine learning-based predictive model capable of estimating the maximum grinding temperature (T_Exp) under different cooling conditions. The original project plan was structured around the following core phases:

1. **Literature Review and Problem Definition**
   The first phase involved a comprehensive literature review to examine existing methods for predicting grinding temperatures, with particular focus on the limitations of traditional cooling approaches and analytical temperature models. The aim was to clearly define the problem and identify gaps in the current methodologies.
2. **Data Collection and Preprocessing**
   The project plan included collecting experimental data from grinding operations, focusing on key process parameters such as feed rate, cutting depth, and cooling techniques. The collected data would then undergo preprocessing to address missing values, standardize feature scales, and ensure it was properly formatted for use in model training.
3. **Model Selection and Training**
   The following phase involved implementing machine learning models, with Random Forest selected as the baseline approach. XGBoost and LightGBM were also evaluated as alternative algorithms. To enhance model performance, hyperparameter tuning was planned for each method.
4. **Model Evaluation and Comparison**
   Once the models were trained, they would be evaluated using performance metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The best-performing model would then be selected.
5. **Integration and Optimization**
   The final stage of the project was intended to focus on integrating the predictive model into a real-time system for estimating grinding temperatures. Additionally, optimizing cooling strategies based on the model's outputs was envisioned as a potential direction for future development.

7.2. Changes to the Project Plan During Implementation

As the project progressed, several changes and adjustments were made to the initial plan due to unforeseen challenges and opportunities. Below are the key changes:

1. **Incorporation of Synthetic Data Generation (GANs)**
   Synthetic data generation was not initially included in the project scope. However, due to the limited size and lack of variability in the existing dataset (especially concerning cooling methods) a Generative Adversarial Network (GAN) was employed to produce

supplementary synthetic data. This decision was made upon realizing that the available data lacked sufficient diversity to train the models effectively.

Reason for Change: This significant change was brought about by the necessity to improve model resilience and get over the data limits, which were not initially expected.

2. **Extension of Model Evaluation and Comparison**
   Random Forest served as the baseline model in the original concept, but to compare performance and choose the optimal algorithm, two other models (XGBoost and LightGBM) were also assessed. In order to optimize the models' performance, hyperparameter tuning was also implemented utilizing methods such as GridSearchCV.

   Reason for Change: This extension of model evaluation was motivated by the need to guarantee the most accurate and efficient model for the task.
3. **Integration of Analytical Temperature Model (T_Model)**
   The integration of the MATLAB-based analytical temperature model (T_Model) with the machine learning models was not originally planned. However, it was later deemed essential for enhancing the overall accuracy of temperature predictions and for providing a reliable reference to validate the outputs of the machine learning models.
   Reason for Change: The use of T_Model enhanced the reliability of the machine learning model's predictions by offering a trustworthy benchmark for comparison.
4. **PCA for Data Visualization and Model Validation**
   Principal Component Analysis (PCA) was not part of the original project framework. However, as the work advanced, it became evident that PCA would be a useful technique for visualizing the distribution of both real and synthetic datasets, thereby helping to verify that the generated data reliably represented actual grinding conditions.
   Reason for Change: PCA was incorporated to verify the quality of the synthetic data and to examine whether the distribution of the augmented dataset aligned with that of the original data.

7.3. Lessons Learned in Project Management

During the course of the project, valuable insights were gained regarding the management of both technical processes and organizational coordination.

1. **Flexibility in the Project Plan**
   A major takeaway from the project was the value of maintaining flexibility within the project plan. Although having a clear and structured roadmap is important, being able to revise the approach in response to unforeseen issues (like data scarcity) was essential. For instance, the choice to implement GANs for synthetic data generation, despite not being part of the original plan, turned out to be a critical improvement that significantly boosted model performance.

2. **Collaboration and Teamwork**
   Clear communication and strong teamwork played a central role in the success of the project. Frequent team meetings provided a platform to review progress, address emerging problems, and develop solutions collectively. Making decisions together (particularly when introducing methods like GANs or exploring alternative models) was crucial throughout the process.
3. **Data Management and Quality Control**
   One of the most important insights gained from this project is that the quality and diversity of data are fundamental to effective machine learning model training. The difficulties encountered because of the limited dataset underscored the necessity of comprehensive data collection practices and proactive measures (such as data augmentation) to address missing or insufficient data. Additionally, thorough data cleaning and preprocessing proved essential to ensure that the models were built on accurate and dependable inputs.
4. **Iterative Process and Model Evaluation**
   Model training, assessment, and improvement followed an iterative cycle throughout the project. Although Random Forest was chosen as the initial baseline, subsequent evaluations revealed that XGBoost and LightGBM delivered superior results. This ongoing refinement process enabled continuous enhancement of model performance.
5. **Time Management and Task Prioritization**
   Time management proved to be a significant challenge during the project, particularly when unforeseen developments (like the introduction of GANs) emerged. Staying on schedule required careful task prioritization and the establishment of realistic milestones. Certain stages, such as training GANs and tuning hyperparameters, demanded more time than initially anticipated, prompting revisions to the timeline and a shift in focus toward the most essential tasks.

7.4. Conclusion

In conclusion, the project was successfully managed, even though it deviated at times from the original plan. The team's flexible approach and ability to make thoughtful adjustments to the project scope were instrumental in overcoming obstacles and achieving the desired outcomes. This process underscored the importance of adaptability, effective teamwork and careful data management in delivering complex projects. The experience gained will serve as a valuable foundation for future data-driven engineering initiatives that require continuous iteration and refinement.

8. CONCLUSION AND FUTURE WORK

8.1. Key Results and Findings

The primary purpose of this research effort involved creating a machine learning prediction model to estimate grinding temperatures under different cooling conditions for enhancing manufacturing efficiency. The central output of this project consists of:

1. **Development of a Robust Predictive Model**
   A series of machine learning algorithms received training to predict the maximum grinding temperature (T_Exp) based on feed rate and depth of cut together with cooling efficiency (heat convection coefficient, h) measurement input. XGBoost and LightGBM demonstrated better predictive accuracy than Random Forest according to the performance results.
2. **Synthetic Data Generation via GANs**
   The application of Generative Adversarial Network (GAN) worked because existing data sets contained minimal information about cooling methods and grinding procedures. The extra data produced by augmentation allowed models to achieve more accurate predictions alongside better generalization skills. The model gained both flexibility and robustness from GAN-modified data which solved the deficiencies of experimental data.
3. **Evaluation and Comparison of Models**
   Success in project analysis required the comparison of different machine learning methods that used Mean Absolute Error and Root Mean Squared Error metrics for their evaluation. XGBoost together with LightGBM served this task due to their ability to accurately predict at industrially practical speeds.
4. **Integration of Analytical Temperature Model (T_Model)**
   Research predictions received valid confirmation through the benchmark that MATLAB-based analytical temperature model T_Model provided. Proper reliability for model applications in operational settings resulted from cross-validating T_Model with machine learning models.

8.2. Limitations and Conclusions

Several important restrictions we faced with, while the project accomplished its main objectives;

1. **Limited Dataset**
   The limited size of training data together with insufficient representation of cooling conditions constrained possible input variations for the model. GANs rectified the dataset limitation successfully while additional model development requires advanced and complex training data.
2. **Model Interpretability**
   XGBoost and LightGBM models operate as black boxes despite providing accurate predictions because they lack clear interpretability when compared to Random Forest

models. Decision mechanisms within XGBoost and LightGBM models stay ambiguous which makes their interpretation a disadvantage for process enhancement needs.

3. **Real-Time Application**
   The developed prediction system advances real-time temperature monitoring capabilities yet it needs further evolution to implement in current manufacturing networks. The model requires development to enable real-time usage along with features that enable streaming data gathering and machine system integration and user interface installation.
4. **Cooling Strategy Optimization**
   The predicted output for grinding temperature from this model requires development into an operational real-time cooling strategy execution platform. Operational research needs to develop a system for sustaining predictive temperature-based modifications of dynamic cooling strategies.

8.3. Future Work

Several directions for future work can build upon the results of this project:

1. **Data Collection and Expansion**
   The project dataset needs expansion by adding cooling techniques along with various grinding scenarios and materials. An expanded industrial case range within a larger diverse dataset would enable model generalization to improve the performance.
2. **Model Interpretability and Explainability**
   Future studies may focus on clarifying machine learning models using techniques such as SHAPE (Shapley Additive Descriptions) or LIME (Native Interpretable Model-Agnostic Descriptions). By better understanding how different factors affect the predicted temperatures, this will help engineers and operators make better decisions.
3. **Real-Time System Integration**
   In order to make the model suitable for practical use, further processing may be required to incorporate it into an industrial control system. This means reducing the model to operate in real-time environments and connecting it to sensors and machine monitoring systems that provide continuous input data. Then, depending on the temperature projections, the cooling strategy can be changed in real time.
4. **Cooling Strategy Optimization**
   A possible next step would be to develop a real-time cooling optimization tool that dynamically changes cooling conditions depending on the model's temperature predictions. Reducing thermal damage, reducing energy use and extending tool life will help this improve grinding performance.
5. **Exploring Alternative Data Augmentation Techniques**
   Although gains are used to create fake data, researching other data augmentation techniques or semi-supervised learning models can further improve the performance of

the model. For example, techniques such as simulation-based data augmentation or transfer learning can be investigated.

6. **Commercialization and Industry Adoption**
   The prediction model and cooling optimization tool are ultimately commercially available, therefore providing manufacturers with a real-time grinding process optimization tool. This can lead to lower costs, more sustainable practices and more grinding industry production.

8.4. Conclusion

All things considered, our study successfully developed an approach based on machine learning that outperforms traditional models in predicting grinding temperatures under various cooling conditions. The inclusion of synthetic data generation and the testing of various machine learning algorithms have helped to dramatically improve predictive modeling for manufacturing. Despite some disadvantages, the results show that this technology can improve grinding techniques and provide more affordable and environmentally friendly production processes.

The project is preparing the ground for further work that could greatly affect the financial and technical aspects of production. Real-time integration, model optimization and commercial uses in the grinding industry can all be part of this work.

11. APPENDIX

Figure 3.3 – Synthetic Data Validation via XGBoost Prediction vs Label: This figure demonstrates that XGBoost was able to predict the labels of synthetic data with high accuracy, closely aligning with the ideal diagonal. It validates the quality of GAN-generated data prior to merging with experimental data.

Figure 3.8.1 – Flowchart of the Project Procedure: This stepwise flow enhances clarity and reproducibility of the project pipeline.

Figure 3.8.2 –Input–Target–Output structure of the PIML-based machine learning model: This mapping illustrates how both physics-based and data-driven features are combined to enhance prediction performance.

Figure 3.8.3.1 – Randomforest model PIML and non PIML values

Figure 3.8.3.2- XG Boost model PIML and non PIML values

Figure 4.1 – Regression Plot of GAN-enhanced Random Forest on Test Set: The figure shows how the Random Forest model, trained with GAN-generated data, performed on the test set. While the R^2 is lower than XGBoost, the model still demonstrates reasonable predictive power.

Figure 4.2 – Regression Plot of XGBoost (with GAN + Filtering) on Test Set: This plot shows the predicted vs actual temperatures on the test dataset using XGBoost trained on GAN-augmented and filtered synthetic data. The strong alignment with the diagonal indicates high accuracy and generalization.

Figure 4.3 – Regression Plot of LightGBM Model on Combined Data: As illustrated in Figure 4.3, the LightGBM model trained on combined data performed moderately well. Although the predictions show more variance than XGBoost, they still follow the overall trend of the ground truth.

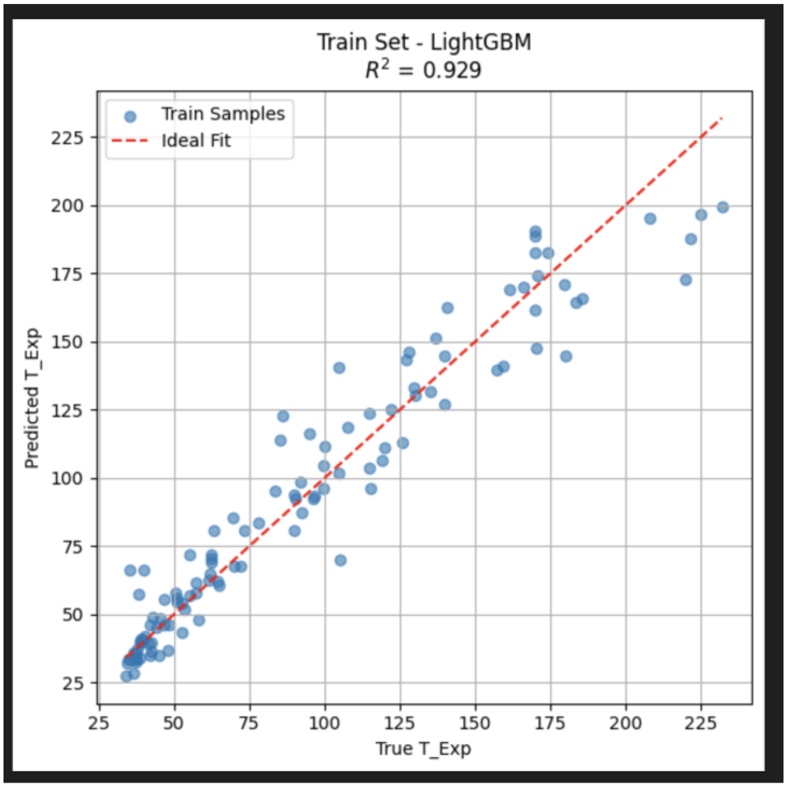

Figure A.8 – Regression Plot of LightGBM on Training Set

This regression plot shows the LightGBM model performance on the training dataset. The R^2 value of 0.929 indicates a strong fit without evident overfitting, confirmed by comparable results on the test set.

REFERENCES

- Araghizad, A.E., Pashmforoush, F., Tehranizadeh, F., Kilic, K., & Budak, E. (2024). Improving milling force predictions: A hybrid approach integrating physics-based simulation and machine learning for remarkable accuracy across diverse unseen materials and tool types. *Journal of Manufacturing Processes*, 114, 92–107.
- Bagherzadeh, A., Pashmforoush, F., Jamshidi, H., & Budak, E. (2025). *Fundamentals of cooling/lubrication effect in grinding of Inconel 718 employing an inverse thermo-mechanical model*. Tribology International, 209, 110746.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27.
- Grguraš, D., Sterle, L., Krajnik, P., & Pušavec, F. (2019). A novel cryogenic machining concept based on a lubricated liquid carbon dioxide. *International Journal of Machine Tools and Manufacture*, 145, 103456.
- Kalpakjian, S., & Schmid, S. R. (2003). *Manufacturing Processes for Engineering Materials* (4th ed.). Prentice Hall.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- Rowe, W. B. (2014). *Principles of Modern Grinding Technology* (2nd ed.). William Andrew/Elsevier. ISBN: 978-0-323-24271-4.
- Sánchez, J. A., Pombo, I., Alberdi, R., Izquierdo, B., Ortega, N., Plaza, S., & Martínez-Toledano, J. (2010). Machining evaluation of a hybrid MQL-CO2 grinding technology. *Journal of Cleaner Production*, 18(18), 1840–1849.
- Shen, B., & Shih, A. (2009). Minimum quantity lubrication (MQL) grinding using vitrified CBN wheels. *Transactions of the North American Manufacturing Research Institution of SME*, 37.
- Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 25.