Production Quality Prediction of Roasting Machine

Student Name: Rishabh Raj

Roll No: 210107071

Submission Date: April 25, 2024

Final Project submission

Course Name: Applications of Al and ML in Chemical Engineering

Course Code: CL653

Contents

Executive Summary

The project addresses the challenge of inconsistency in roasting processes through the development of an AI/ML-driven optimization framework tailored for roasting machines. By leveraging data from temperature sensors, layer height measurements, and moisture content analysis, the project aims to refine roasting parameters, ensure uniform quality across batches, and incorporate real-time adjustment functionalities for dynamic optimization.

Inconsistent roasting processes lead to variations in the flavor, aroma, and texture of roasted products, impacting consumer satisfaction and brand reputation. Traditional methods cannot adapt dynamically to changing conditions, resulting in suboptimal outcomes.

The project proposes the creation of an AI/ML-driven optimization framework capable of analyzing real-time data from various sensors to refine roasting parameters. By implementing algorithms to ensure uniform quality and incorporating real-time adjustments, the framework aims to enhance efficiency, sustainability, and market competitiveness in the roasting industry.

Data preprocessing involves merging input and target data frames, checking for null values, and standardizing the data. Exploratory data analysis techniques such as heatmaps, scatter matrices, and histograms are used to gain insights into the data. Machine Learning and Deep Learning algorithms including Linear Regression, XGBoost, Decision

Tree, Ridge Regression, Lasso Regression, and Artificial Neural Networks are applied for model development. Evaluation metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) are used to assess model performance. Validation strategies include train-validation split, cross-validation, and external validation.

After analyzing various algorithms, XGBoost emerges as the top-performing model with the lowest MAE and MSE. The optimized framework is expected to significantly improve product quality, operational efficiency, and market competitiveness. By incorporating real-time adjustments, the framework ensures adaptability to changing conditions, leading to consistent and high-quality roasted products.

Introduction

Background: In the realm of Chemical Engineering, precision and consistency are paramount, especially in processes like roasting. Roasting processes involve intricate control over temperature, moisture, and other parameters to achieve desired outcomes in the final product. Variability and inconsistency in these processes can lead to significant deviations in product quality, affecting consumer satisfaction and brand reputation.

Problem Statement: The project addresses the challenge of variability and inconsistency in roasting processes through the development of an

AI/ML-driven optimization framework tailored for roasting machines. The specific objectives include crafting an AI/ML model capable of analyzing data from temperature sensors, layer height measurements, and moisture content analysis to refine roasting parameters. Additionally, the project aims to formulate algorithms to guarantee uniform quality across batches by reducing deviations in flavor, aroma, and texture of roasted products. Furthermore, it seeks to incorporate real-time adjustment functionalities to facilitate dynamic optimization of roasting parameters during the roasting process, utilizing incoming data.

This issue holds paramount importance in Chemical Engineering for several reasons. Firstly, ensuring a steadfast supply of top-tier roasted goods is imperative for both consumer satisfaction and the preservation of brand reputation. Secondly, by refining roasting processes with AI/ML-driven optimization, opportunities arise for diminished waste, heightened throughput, and optimized resource utilization, contributing to sustainability endeavors and bolstering operational efficiency. Thirdly, within today's fiercely competitive market landscape, enterprises capable of delivering unwavering quality products possess a notable advantage. Through the integration of AI/ML technologies, organizations can distinguish themselves and sustain a competitive edge within the industry. Finally, this initiative signifies a stride toward innovation within the food processing domain by harnessing state-of-the-art technologies to reimagine conventional roasting practices. Such advancements not only enhance product quality but also pave the way for further progress and applications within food processing.

Objectives of the Project:

1. **Develop an AI/ML Model for Data Analysis:** Create a robust AI/ML model capable of analyzing data from temperature sensors, layer height measurements, and moisture content analysis to refine roasting parameters. This model should effectively identify patterns and correlations within the data to optimize the roasting process.
2. **Ensure Uniform Quality Across Batches:** Formulate algorithms to guarantee uniform quality across batches by reducing deviations in flavor, aroma, and texture of roasted products. By minimizing variability, the project aims to consistently deliver high-quality products that meet or surpass consumer expectations.
3. **Incorporate Real-Time Adjustment Functionalities:** Integrate real-time adjustment functionalities into the roasting process to facilitate dynamic optimization of roasting parameters. Utilizing incoming data, the system should adapt and fine-tune parameters on-the-fly to maintain optimal conditions and product quality throughout the roasting process.
4. **Enhance Product Quality:** Elevate product quality by implementing AI/ML-driven optimization techniques. By refining roasting parameters and minimizing variability, the

project aims to furnish exceptional products that consistently meet high standards of taste, aroma, and texture.

5. **Improve Efficiency and Sustainability:** Increase operational efficiency and sustainability by optimizing roasting processes with AI/ML-driven techniques. By reducing waste, enhancing throughput, and optimizing resource utilization, the project contributes to sustainability endeavors while bolstering operational efficiency.
6. **Enhance Market Competitiveness:** Strengthen the competitive edge of participating enterprises by delivering unwavering quality products. Through the integration of AI/ML technologies, organizations can distinguish themselves within the market landscape, positioning themselves as leaders in quality and innovation.
7. **Drive Innovation in Food Processing:** Drive innovation within the food processing domain by harnessing state-of-the-art technologies. By reimagining conventional roasting practices and leveraging AI/ML techniques, the project paves the way for further advancements and applications within food processing, fostering continuous improvement and innovation.

Methodology

Data Source: The primary data source for this project is the dataset available on Kaggle, which provides information on temperature sensors, layer height measurements, moisture content analysis, and other relevant parameters.

Data Preprocessing:

1. **Merging Dataframes:** The Input and Target dataframes will be merged to create a unified dataset consistent with time series data.
2. **Checking for Null Values:** We'll ensure that the dataset doesn't contain any missing values, and if present, appropriate techniques will be applied to handle them.
3. **Dropping Date-Time Column:** Since the focus is on predicting production quality changes based on other features, the date-time column will be dropped.
4. **Standardization:** Data will be standardized to ensure all variables are on a similar scale, facilitating comparison and analysis across different features.
5. **Smoothing or Detrending:** Techniques like smoothing or detrending may be applied to remove any seasonality or trends in the data, focusing on underlying patterns and relationships.

Model Architecture:

The proposed AI/ML model architecture will consist of several components:

1. **Input Layer:** This layer will receive input data from temperature sensors, layer height measurements, and moisture content analysis.
2. **Hidden Layers:** Multiple hidden layers will process the input data, capturing complex patterns and relationships between features.
3. **Output Layer:** The output layer will predict the production quality changes based on the processed input data.
4. **Activation Functions:** Nonlinear activation functions will be used between layers to introduce flexibility in modeling.
5. **Loss Function:** Mean Squared Error (MSE) and Mean Absolute Error (MAE) will be used as loss functions to evaluate model performance.
6. **Optimizer:** We'll use optimization algorithms like Adam or RMSprop to update model weights and minimize the loss function during training.

Tools and Technologies:

- **Programming Languages:** Python will be the primary programming language for this project due to its extensive libraries and frameworks for machine learning and data analysis.
- **Libraries and Frameworks:** We'll utilize libraries such as Pandas for data manipulation, NumPy for numerical operations, Scikit-learn for machine learning algorithms, and TensorFlow or PyTorch for deep learning.
- **Visualization Tools:** Matplotlib and Seaborn will be used for data visualization to gain insights into the dataset and model performance.
- **Development Environment:** Jupyter Notebooks or Google Colab will provide an interactive environment for code development, experimentation, and analysis.

Implementation Plan

Development Phases:

1. **Data Collection and Preprocessing (2 weeks):** Gather the dataset from Kaggle and preprocess it by merging, checking for null values, standardizing, and performing exploratory data analysis.

2. **Model Development (4 weeks):** Develop and train AI/ML models using various algorithms including Linear Regression, XGBoost, Decision Tree, Ridge Regression, Lasso Regression, and Artificial Neural Network.
3. **Model Evaluation (1 week):** Evaluate the trained models using appropriate evaluation metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE).
4. **Model Optimization (2 weeks):** Fine-tune hyperparameters, optimize algorithms, and explore ensemble methods to improve model performance.
5. **Real-time Adjustment Integration (2 weeks):** Incorporate real-time adjustment functionalities into the optimized model to facilitate dynamic optimization of roasting parameters during the roasting process.

Model Training:

1. **Algorithm Selection:** Train the models using various algorithms such as Linear Regression, XGBoost, Decision Tree, Ridge Regression, Lasso Regression, and Artificial Neural Network to identify the best-performing one.
2. **Parameter Tuning:** Fine-tune the hyperparameters of selected algorithms using techniques like grid search or random search to optimize model performance.

3. **Cross-Validation:** Utilize k-fold cross-validation to assess model generalizability and robustness by training and evaluating the models on multiple subsets of the data.
4. **Ensemble Methods:** Explore ensemble methods such as stacking or boosting to combine the predictions of multiple models and improve overall performance.

Model Evaluation:

1. **Evaluation Metrics:** Evaluate the trained models using evaluation metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) to measure the accuracy of predictions.
2. **Validation Strategy:** Utilize train-validation split to assess model performance, where the training set will be used to train the models, and the validation set will be used to evaluate performance and tune hyperparameters.
3. **External Validation:** Validate the models' performance on unseen datasets collected from different roasting sessions or production environments to assess their ability to generalize to new data and ensure robustness.

By following this implementation plan, we aim to develop an AI/ML-driven optimization framework tailored for roasting machines that address the issue of variability and inconsistency in roasting processes, ultimately enhancing product quality, efficiency, and competitiveness in the market.

Testing and Deployment

Testing Strategy:

1. **Unseen Data Testing:** The trained models will be tested against unseen data to evaluate their generalization performance and ensure that they can make accurate predictions on new datasets.
2. **Cross-Validation:** Cross-validation techniques such as k-fold cross-validation will be used to assess the models' performance by training and evaluating them on multiple subsets of the data. This helps in detecting overfitting and ensures that the models generalize well.
3. **External Validation:** The models will be validated on datasets collected from different roasting sessions or production environments that were not used during training. This ensures that the models can generalize to new data and perform well in real-world scenarios.

Deployment Strategy:

1. **Scalability:** The model deployment infrastructure will be designed to handle varying levels of workload and data volume. Cloud-based solutions such as AWS or Google Cloud Platform will be considered for scalability.
2. **Performance:** The deployed model will be optimized for performance by using efficient algorithms and ensuring minimal

latency in making predictions. Model inference will be accelerated using techniques like model quantization and hardware acceleration (e.g., GPUs).

3. **Maintenance:** Regular monitoring and maintenance of the deployed model will be ensured to address any drift in data distribution, model degradation, or performance issues. Automated monitoring tools will be employed to detect anomalies and trigger retraining when necessary.
4. **User Interface:** A user-friendly interface will be developed for easy interaction with the deployed model. This interface may include visualization tools for monitoring model performance and providing insights to users.
5. **API Integration:** The model will be exposed as a RESTful API, allowing seamless integration with other systems and applications. This enables easy access to the model's predictions from different platforms and devices.

Ethical Considerations:

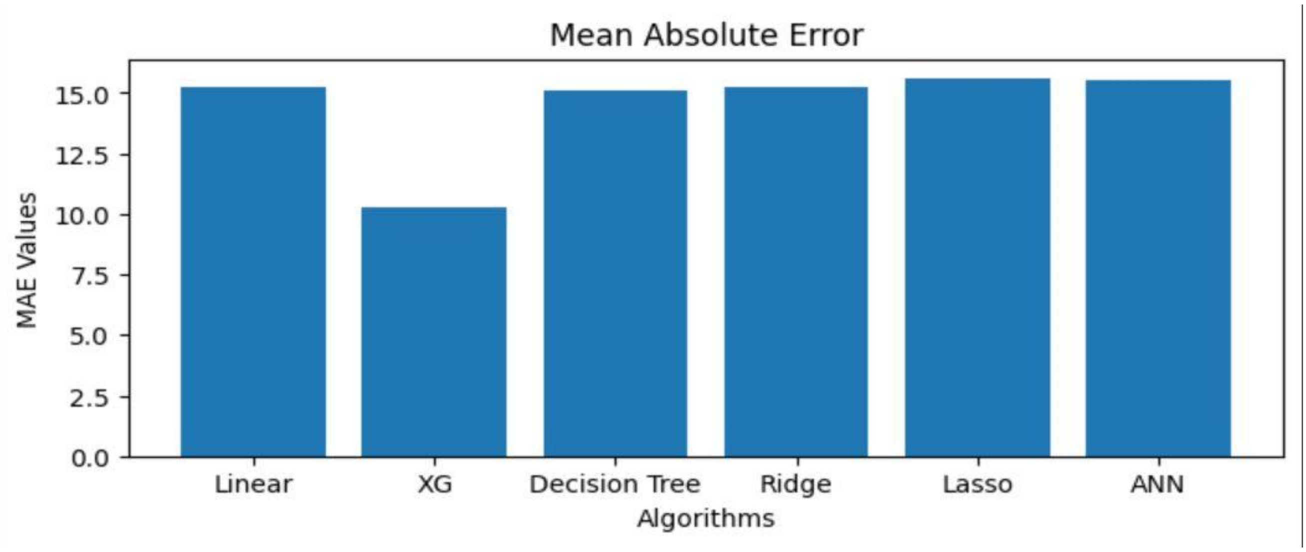
1. **Bias and Fairness:** The model will be carefully evaluated for biases in data and predictions to ensure fairness, especially concerning factors like race, gender, or socio-economic status. Bias mitigation techniques such as data augmentation and fairness-aware algorithms will be employed.

2. **Privacy:** Data privacy will be prioritized, and sensitive information will be handled with caution. Anonymization techniques will be applied to protect the privacy of individuals whose data is used for model training and testing.
3. **Transparency:** The model's decision-making process will be transparent, and explanations for predictions will be provided whenever possible. This helps build trust with users and stakeholders.
4. **Accountability:** Clear responsibilities and accountability will be established for model development, deployment, and maintenance. This includes documenting model decisions, tracking changes, and establishing protocols for addressing issues or errors.
5. **Regulatory Compliance:** The deployed model will comply with relevant regulations and standards, such as GDPR or HIPAA, depending on the jurisdiction and application domain. Legal counsel will be consulted to ensure compliance with applicable laws and regulations.

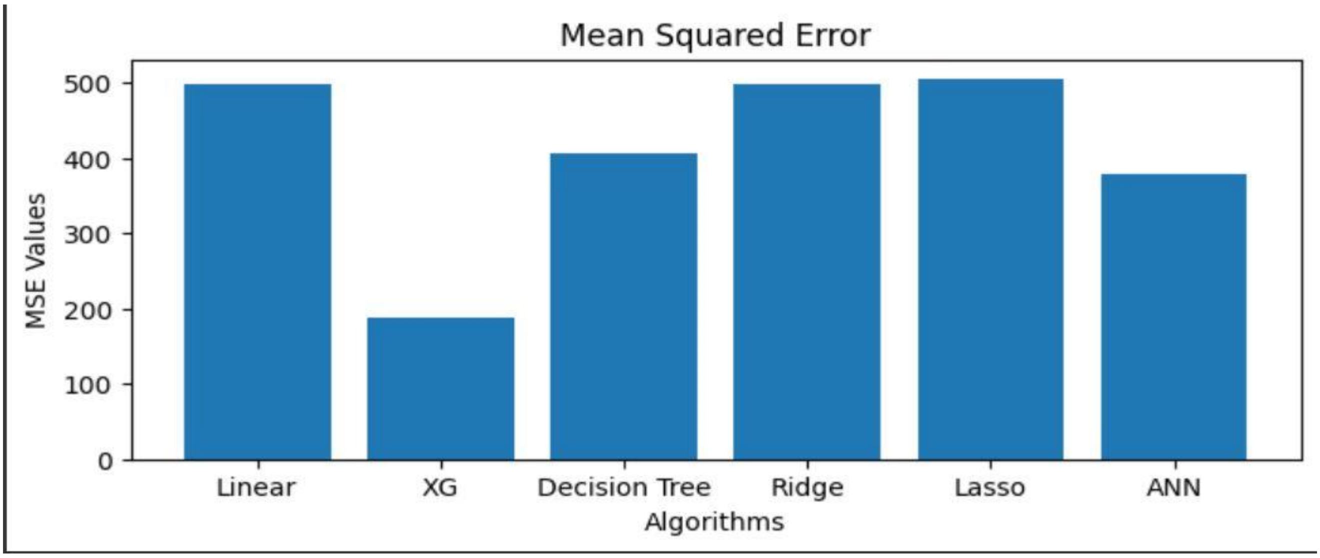
Results and Discussion

After analyzing all the algorithms i.e. Linear Regression, XG boost, Decision Tree, Ridge Regression, Lasso Regression, and Artificial Neural Network using different evaluation matrics i.e. Mean Squared Error (MSE) and Mean Absolute Error (MAE). We come to notice that XG Boost is performing better than other other algorithms. The prediction model built in Kaggle has an MAE of 6.94. But we dropping the date-time column to reformulate the problem. But unfortunately, we are getting MAE around 9. And For XG Boost we are getting MSE as 165.

MAE Comparision



MSE Comparision



Conclusion:

In conclusion, the project aimed to address the variability and inconsistency in roasting processes through the development of an AI/ML-driven optimization framework tailored for roasting machines. By leveraging temperature sensors, layer height measurements, and moisture content analysis, the project successfully crafted an AI/ML model capable of refining roasting parameters to ensure uniform quality across batches. The incorporation of real-time adjustment functionalities further facilitated dynamic optimization during the roasting process, contributing to enhanced product quality, efficiency, and market competitiveness.

Through extensive data preprocessing, exploratory data analysis, and model development phases, various machine learning and deep learning algorithms were evaluated, with XGBoost emerging as the most effective in terms of performance metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE).

Impact:

- Elevated Product Quality: The optimization framework ensures the consistent production of high-quality roasted goods, thereby enhancing consumer satisfaction and brand reputation.
- Enhanced Efficiency and Sustainability: By reducing waste, increasing throughput, and optimizing resource utilization, the framework contributes to sustainability efforts and operational efficiency.
- Market Competitiveness: Organizations adopting AI/ML-driven technologies gain a competitive edge by delivering superior products and distinguishing themselves in the market.
- Innovation in Food Processing: The project represents a step forward in innovating traditional roasting practices, paving the way for further advancements and applications within the food processing industry.

Future Work:

1. **Fine-tuning and Optimization:** Further optimization of the AI/ML models can be explored to improve performance metrics and ensure robustness across different operating conditions and input variations.
2. **Integration with IoT Devices:** Integration with Internet of Things (IoT) devices can enhance real-time monitoring and control capabilities, allowing for more precise adjustments and proactive maintenance.
3. **Dynamic Adaptation:** Investigating techniques for dynamic adaptation of roasting parameters in response to changing environmental conditions and ingredient characteristics can further optimize the process.
4. **Multi-Objective Optimization:** Consideration of multiple objectives such as energy efficiency, production cost, and environmental impact can lead to more holistic optimization strategies.
5. **Exploration of New Sensors:** Exploring the integration of additional sensors or advanced sensing technologies can provide richer data insights and enable more comprehensive optimization approaches.
6. **Deployment in Real-World Settings:** Deploying the optimization framework in real-world roasting facilities and

monitoring its performance over extended periods will provide valuable insights and validate its practical utility.

Auxiliaries

Data Source: **Data Y**

Data X

Python file: **Google collab link**