Journal of Applied Engineering and Technological Science
Vol 7(1) 2025:274-295

PERFORMANCE IMPROVEMENT OF QUALITY MONITORING SYSTEMS IN IMBALANCED DATA CONDITIONS FOR FAT-FILLED POWDER QUALITY IN THE DAIRY INDUSTRY

Muhammad Asrol[1], Oki Pratama[2]
Industrial Engineering Department-BINUS Graduate Program, Bina Nusantara University
Jakarta, 11480, Indonesia[12]
muhammad.asrol@binus.edu[1], oki.pratama@binus.ac.id[2]

Received: 14 January 2025, Revised: 28 May 2025, Accepted: 07 October 2025
**Corresponding Author*

ABSTRACT
Fat-filled powder has the potential to substitute milk in meeting the nutritional needs of the community, but its product quality remains unstable during continuous production processes. A key challenge in fat-filled powder (FFP) production is the difficulty in quality monitoring, which is influenced by various uncertainty factors that affect product quality. Machine learning can be implemented for quality monitoring system, but the imbalanced data conditions require the development of algorithms with optimal performance. This study aims to design a quality monitoring system for FFP using a machine learning model under imbalanced dataset conditions and the influence of other uncertainty factors. A Random Forest (RF) machine learning model was developed for monitoring FFP quality. In the context of imbalanced datasets, the model was optimized through various scenarios, including data splitting for training and testing, as well as the Synthetic Minority Oversampling Technique (SMOTE) and Distribution Optimally Balanced – Stratified Cross Validation (DOB-SCV) schemes. The results showed that the SMOTE model achieved the best performance in terms of accuracy, precision, and recall with scores of 99.67%, 99.79%, and 99.24%, respectively, on the testing data. Statistically, the RF model with the SMOTE data manipulation scenario also showed significant differences compared to the DOB-SCV model and the traditional data splitting approach. The quality monitoring model for FFP developed in this study can be implemented in the dairy industry, offering more stable, accurate quality monitoring predictions that align with real conditions, helping to avoid quality uncertainties during the production process. The implementation of this model in the industry has the potential to facilitate a broader, more transparent, and optimized product quality evaluation process, which can also be conducted in real time under continuous production conditions.
Keywords: *Quality Monitoring, Dairy Industry, Imbalanced Data, Machine Learning, Synthetic Data Manipulation.*

1. Introduction

Fat-filled powder is a dairy industry formulated product with the potential to substitute milk. In developing countries with low income, fat-filled powder plays a crucial role in meeting daily protein needs, as these populations still face challenges in accessing nutritious food (Kene Ejeahalaka & On, 2020). The high potential of FFP as a substitute raw material for consumers and industries offers opportunities for developing a stable and efficient production process. Currently, the production process of FFP faces various challenges, particularly as it remains empirical in nature and has yet to achieve process stability to consistently deliver the highest quality FFP (Vignolles et al., 2010).

The production and formulation of fat-filled powder involve mixing skim milk as the base ingredient, substituted with a specific amount of vegetable fat. This mixing process faces challenges in ensuring proper homogenization and emulsification, which must be carried out effectively to facilitate further processing through a spray dryer (Finnegan et al., 2021). The mixing of various ingredients in FFP production carries the potential risk of product quality degradation, particularly due to the interaction between protein and fat (Purwantiningsih et al., 2022). This challenge becomes even more complex when large-scale production is required while maintaining cost-effectiveness. On the other hand, ensuring continuous production processes, robust quality monitoring, and stringent food safety measures is crucial to delivering products that meet consumer expectations at a competitive price.

The development of digital technology, supported by Industry 4.0, presents significant opportunities and impacts for controlling production processes in continuous and large-scale schemes like FFP. Key aspects of Industry 4.0, such as artificial intelligence, data analytics, and machine learning, hold great potential and play a crucial role in monitoring product quality during production. The challenge lies in monitoring FFP quality across various parameters and ensuring optimal data availability through a quality control system to guarantee a safe and controlled production process. Artificial intelligence and machine learning approaches have been widely utilized in product quality control, including: using convolutional neural networks (CNNs) to monitor fruit freshness without damaging the product (Han et al., 2022); employing sensors and k-nearest neighbour algorithms to monitor meat quality (Pounds et al., 2022); minimizing contamination in pasteurized dairy products through multi-method machine learning, such as random forest and logistic regression (Murphy et al., 2021); and improving agricultural product quality post-harvest using CNN-based U-Net (Singh et al., 2022).

Quality inspection of dairy products is a critical aspect, as these products are highly perishable. Rapid, efficient, and intelligent quality checks need to be implemented promptly. The use of machine learning and artificial intelligence technologies has been widely applied in the dairy industry, particularly for product quality inspection during or after production processes. Research by Ref. (Y. Zhang et al., 2022) utilized various machine learning preprocessing models and methodologies such as PCA, regression analysis, and classification to design a milk quality monitoring system. Ref. (Cockburn, 2020) applied multiple machine learning models not only for milk quality inspection but also for efficient dairy industry management. Ref. (R. Zhang et al., 2018) leveraged extreme learning machines and neural networks for food safety prediction, specifically for dairy industry products. Ref. (Feng et al., 2024) employed Raman spectroscopy and various classification machine learning models to analyse the nutritional components of dairy products. Additionally, Ref. (Mu et al., 2020) utilized an electronic nose alongside various regression and classification models to detect the origin of milk.

Various previous studies have shown that machine learning has been utilized in various business activities and downstream products within the dairy industry. However, FFP, as one of the dairy industry products resulting from the formulation and blending of various components, still faces limitations in further exploration, particularly in terms of quality and maintaining process stability. Yet, as mentioned earlier, this product holds significant economic potential as a substitute material in the industry or as a protein alternative for consumers. Thus far, prior research has mainly focused on the development of FFP formulations through blending engineering and the addition of supplementary ingredients via empirical experiments (Kene Ejeahalaka & On, 2020), (Vignolles et al., 2010).

The application of machine learning for monitoring the quality of FFP is an important consideration, given its high potential. Various machine learning models and approaches that have been applied to many other dairy industry products can be adopted and further developed for FFP quality monitoring. The main challenge in assessing FFP quality is no longer about selecting the appropriate algorithm but rather ensuring effective dataset handling to produce accurate machine learning models. This aligns with previous research findings, which highlight that in most product quality assessments, a common obstacle is the presence of imbalanced data, particularly in terms of quality classes for manufacturing products (Liu & Dai, 2022; Qin et al., 2022; Zhou et al., 2021) and perishable products (Kumar & Agrawal, 2024; Swain & Jenamani, 2023). Addressing imbalanced data to produce accurate quality prediction models must be a primary focus in model development. In this regard, training data must be well-organized to enable the model to learn all data conditions in line with real-world scenarios. Although various methods have been widely applied in managing and predicting product quality, there remains a critical problem in accurately predicting the quality of fat-filled powder (FFP) when faced with imbalanced data conditions. To address this gap, this study specifically addresses this problem by applying machine learning techniques combined with accurate data preprocessing to improve the reliability of FFP quality prediction models. A statistical analysis is also added to enrich the validation statement of FFP prediction model accuracy.

The objective of the research is to develop a quality prediction model for FFP products with various quality parameters during the production process under imbalanced data conditions, leveraging machine learning approaches and models. To complete the modeling, a data imbalance handling is developed under various scenarios to maximize model performance, ensuring that FFP quality monitoring performs effectively across diverse production process and product quality conditions. An inferential statistical analysis also completed the model to validate the model accuracy. This research seeks to contribute to the development of accurate monitoring and prediction of FFP quality under imbalanced dataset conditions, using machine learning approaches enriched with inferential statistical analysis. Accurate prediction of FFP quality is essential for ensuring high-quality products that meet consumer expectations.

2. Related Work

Food product quality is a crucial and primary aspect that consumers focus on, prompting extensive research aimed at improving quality control of products. Food items, such as dairy products, are highly prone to quality degradation, which can have significant impacts on human health. To prevent such issues, approaches for predicting and classifying product quality have been developed in various studies to effectively monitor and ensure product quality during production and delivery to consumer. In the following subsections, we provide a thematic related works to produce the research gap and contributions.

2.1. Challenges in predicting FFP quality

In this study, a machine learning approach is proposed to predict and monitor the quality of dairy products, specifically Fat Filled Powder (FFP). The quality classification of FFP is divided into two categories: good quality and bad quality. The primary challenge in such a dataset is the potential for imbalance, where one quality class dominates the other. This imbalance can cause the model to learn only the majority class while failing to adequately understand and classify the minority class, leading to misclassification and improper handling during the production process, ultimately resulting in customer dissatisfaction.

In an imbalanced dataset, the impact on the prediction model performance can be significant. Machine learning models tend to perform well on the dominant class but fail to classify the non-dominant class effectively. To improve the accuracy of prediction models, various strategies can be applied. However, the most common approach used by previous researchers is comparing different machine learning prediction methods to identify the one with the best accuracy. According to (Kaitlin et al., 2020; Mendez et al., 2019), the main challenge in imbalanced data is not the choice of algorithm but focusing on two key aspects: conditioning the data to improve accuracy and analysing metrics to evaluate the model's performance accurately.

On the other hand, two approaches can be implemented to address imbalanced data conditions: data reconstruction and the use of appropriate classification algorithms (Haixiang et al., 2017). Determining the right classification algorithm through a trial-and-error model scheme is not a prudent choice, as it still carries the risk of low and inaccurate performance. Previous researchers recommend developing models through data reconstruction, either by undersampling or oversampling, as these methods have a direct and significant impact on model performance (Zhou et al., 2021). In relation to data engineering and conditioning, adjustments are needed for imbalanced data to enable algorithms to learn in a balanced manner and fully comprehend the data conditions. Various methods for conditioning imbalanced data have been proposed, including data splitting, k-fold validation, Synthetic Minority Oversampling Technique (SMOTE), and Distribution Optimally Balanced – Stratified Cross Validation (DOB-SCV).

2.2. Model performance in imbalanced data conditions and the role of random forest model

In the development of prediction model, various approaches can be utilized for predicting product quality under imbalanced conditions. Ref. (Mahmudah et al., 2021) employed algorithms such as Support Vector Machine (SVM), decision tree, and random forest (RF) to classify public data across various sectors under imbalanced conditions. Ref. (Ganganwar, 2012) used oversampling or undersampling approaches combined with SVM modifications to handle classification models under imbalanced conditions. Ref. (Kokkinos et al., 2017) studied various algorithms for addressing imbalanced data conditions and found that data handling methods using SMOTE combined with SVM modeling demonstrated satisfactory performance for classification models. Previous research found that classification model based machine learning enable design a prediction model in imbalanced data conditions. They also highlighted the importance step in designing prediction model is not only on the modelling but also in data transformation steps. Moreover, in previous research were applied in many sectors and the model evaluation only focus on error adjustment.

The main point to consider in classification and prediction models for imbalanced data with binary target classes is that, according to Ref. (Mendez et al., 2019), the choice of algorithm is not the most critical factor in improving model performance. Instead, the focus should be on the dataset conditions and selecting the most appropriate metrics for evaluating model performance. A binary targeted dataset has high possibility to achieved imbalanced dataset condition since in the real-world industries were not possible to control and majority on one of class. Therefore, the data transformation and preprocessing steps are important to prepare the modelling stages. This has also been confirmed by Ref. (Thabtah et al., 2020), which emphasizes that to enhancing classification model performance lies in ensuring the training process is conducted effectively allowing the algorithm to learn data patterns accurately.

Additionally, Ref. (Kaitlin et al., 2020) states that comparing algorithms performance in prediction is not significant instead focus on how data management can be performed under imbalanced data conditions. Therefore, data preprocessing and transformation are provide in this study which applied in an optimum prediction model. Ref. (Kaitlin et al., 2020) also found that random forest performs well for binary datasets. In another study, Ref. (Bahel et al., 2020) also found that random forest outperformed logistic regression (LR), Naïve Bayes classifier, K-Nearest Neighbor, and decision tree classifier in binary classification tasks for breast cancer and Titanic datasets. Furthermore, Ref. (Naik & Purohit, 2017) stated that random forest performs well on massive datasets with binary classes compared to gradient boosting and decision trees. Therefore, since random forest found an optimum performance result, this research propose the random forest quality prediction in imbalanced data condition. A comprehensive data transformation to enhance the prediction model and statistical analysis to verify the model performance is also designed to prediction model development.

In another study, Ref. (Speiser et al., 2019) analyzed 311 datasets with various conditions, including binary datasets, high-dimensional features, and imbalanced datasets, and the results showed that random forest demonstrated excellent performance. For complex data such as binary X-ray datasets, random forest also outperformed other classification models (Kaitlin et al., 2020). Ref. (Mahmudah et al., 2021) proposed the SMOTE model for dataset handling in imbalanced conditions, combined with the random forest algorithm, which showed the best performance. Lastly, Ref. (Thabtah et al., 2020) applied 10-fold validation to address imbalanced data, emphasizing that imbalanced class classification negatively impacts model performance, making the pre-training phase the most critical step in managing excellent model performance. In this case, the dataset transformation to enhance the model performance with SMOTE and 10-fold validation are proposed.

2.3. Research gap and positioning of the proposed study

The issue of imbalanced data has been extensively studied in previous research. Imbalanced datasets are also possible to be identified in quality inspections of perishable food products. Misclassification of quality because of poor learning process by imbalanced dataset can lead to management losses and customer dissatisfaction. Previous studies have extensively predicted the quality of dairy products and their derivatives using various techniques. However, the quality of fat-filled powder (FFP), a derivative dairy product, has not been widely explored and presents a significant opportunity for further investigation. Moreover, the issue of imbalanced data in FFP quality monitoring is critical and remains underexplored, especially in real-world industrial settings. Therefore, to address this gap, the present study aims to predict FFP quality using machine learning under imbalanced data conditions. To solve these issues, this study contributes to the development of a classification model for perishable food products under imbalanced conditions using the random forest algorithm. Random forest was chosen for its simplicity, ability to identify data patterns, and its proven excellent performance on imbalanced datasets in various prior studies.

Furthermore, as highlighted by earlier research, the selection of an algorithm is not the primary challenge in addressing imbalanced data conditions. Therefore, this study focuses on handling imbalanced datasets through various approaches, including oversampling, undersampling, cross-validation, and splitting training and testing under normal conditions. A statistical approach is proposed to demonstrate that each method for handling imbalanced data has a significant impact on improving model performance. This ensures that the model developed under the constraints of imbalanced data can accurately classify product quality in real-world conditions.

3. Research Methods

3.1 Research Stages

The research stage is depicted in Figure 1. This research consists of several stages, namely data preprocessing and descriptions, random forest parameter tuning, model development, and model improvement. The data consists of seven features and one target variable that need to be described to examine the data condition and determine the appropriate handling during preprocessing. In the next stage, a quality classification model is determined along with its parameters. The parameters are set using a parameter tuning model with a grid partition method. In the next step, four models are developed to assess the performance of each model on an imbalanced dataset condition. Finally, the model with the best performance is selected through statistical tests to classify the quality of FFP. Each stage of the research is explained in detail in the following subsections.

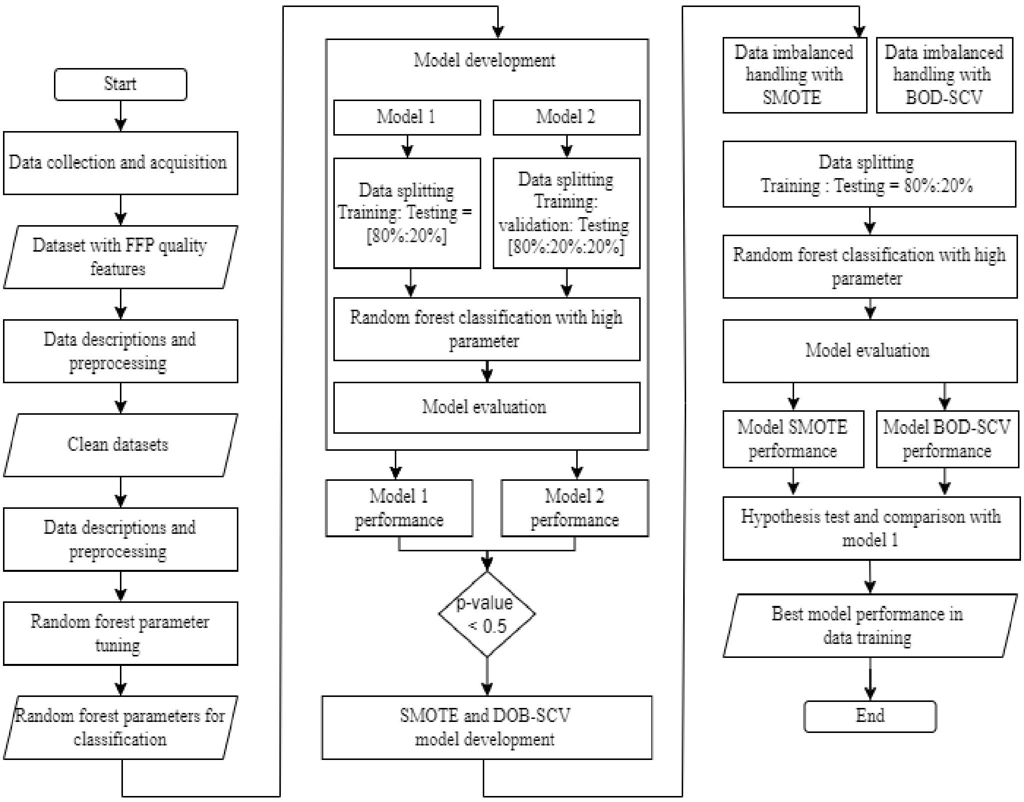


Fig. 1. Research Stage

3.2 Data description and preprocessing

Data description is needed to understand the statistical conditions of the data. Data description also ensures that the data is clean and ready to proceed to the next process. The data description identified from each data attribute includes central tendency and data distribution, which cover parameters such as mean, median, mode, skewness, and kurtosis. The data collected from the industry requires initial handling to proceed to the modelling stage. The product quality data collected from the industry has various forms and conditions. The preprocessing stages applied to the raw quality data are data cleaning and data integration. Data cleaning ensures that the data attributes entered into the next stages are relevant to the modeling goals. Data integration is necessary to combine data from various sources into an integrated and competed dataset, allowing the modeling process to continue.

3.3 FFP quality classification model

The challenge in developing machine learning models for product quality with binary classes is the low accuracy of the model. Previous research has conducted experiments on various algorithms that perform well, such as random forest (Mendez et al., 2019), ANN (Mendez et al., 2019), SVM (Deniz et al., 2017; Mendez et al., 2019; W. Zeng et al., 2011), logistic regression (Deniz et al., 2017). In this study, a quality prediction model will be developed using random forest due to its simplicity and model performance.

The random forest classification model is developed from the decision tree machine learning model. Random forest is an optimization model of decision trees, also known as ensemble learning, which utilizes multiple decision trees to generate a single prediction. The random forest model, first proposed by (Breiman, 2001), is adopted in this study. This study adopts the Random Forest algorithm due to its ability to handle complex data and correlations and provide accurate predictions. Given the nature of the target variable in this research—quality labels—previous studies have demonstrated that Random Forest outperforms other machine learning models in similar contexts (Bahel et al., 2020; Kaitlin et al., 2020; Naik & Purohit, 2017; Speiser et al., 2019). As an ensemble model, Random Forest is optimized for high predictive accuracy, capable of handling uncertainty, and effective in avoiding overfitting. Moreover, Random Forest has shown strong performance in dealing with imbalanced data, especially when combined with appropriate data transformation techniques (Mahmudah et al., 2021).

Suppose $X = \{x_1, x_2, \ldots, x_n\}$ represents the original dataset for training, and , X_i is a subset of X bootstrapped for training the i-th tree. Then, if the original dataset X consists of p attributes, m attributes will be randomly selected for evaluation at each node in the random forest tree, where $m < p$.

Each tree in the random forest will generate its classification prediction value, $h_i(x)$, based on the i-th tree and the input attributes x. The predictions from all trees in the random forest are then combined through majority voting among the trees. The voting result is represented by I, which equals 1 if all conditions are met and 0 otherwise. If there are T trees in the random forest, the prediction value for the input data x to produce class k can be described using Equation 1.

$$H_{(x)} = \arg max_k \sum_{i=1}^{T} I(h_i(x) = k) \tag{1}$$

3.4 Random forest parameters tuning for classification model

The performance of a classification model, such as Random Forest, is influenced by the model's parameter settings refer to the data's characteristics. Proper determination of model parameters enhance the model's accuracy in learning knowledge from the data and predicting classes accurately. Although Random Forest generally performs well with its default parameters, parameter tuning can further improve the model's performance in class predictions. Accurate class assignment is particularly critical in monitoring product quality in the dairy industry, as it directly impacts customer satisfaction.

The parameters that need to be set in a Random Forest model include the number of decision trees, the number of features at each node, the depth of each tree, the minimum number of samples required to split a node, the minimum samples at each leaf, or the bootstrap sampling settings. There are several methods for parameter tuning to determine the optimal parameters for Random Forest; in this study, the grid search method is adopted. Previous research has demonstrated that grid search is an effective method for optimally determining parameters with the best performance within an efficient search time (Anggoro & Mukti, 2021; Shekar & Dagnew, 2019).

The parameter search method using grid search is categorized as an exhaustive method, as it explores all possible combinations of parameters to find the best combination that recommends the optimal model performance. If Random Forest has *c* parameters in its classification modeling, the first step is to define the range of values for each parameter, both inbound and outbound. Grid search experiments with every combination of parameters to identify the combination with the best model performance. The mechanism for parameter search using grid search, as referenced in (Zhu et al., 2022), is illustrated in Figure 2. Meanwhile, the parameters and their respective value ranges defined for grid search in Random Forest can be seen in Table 1.

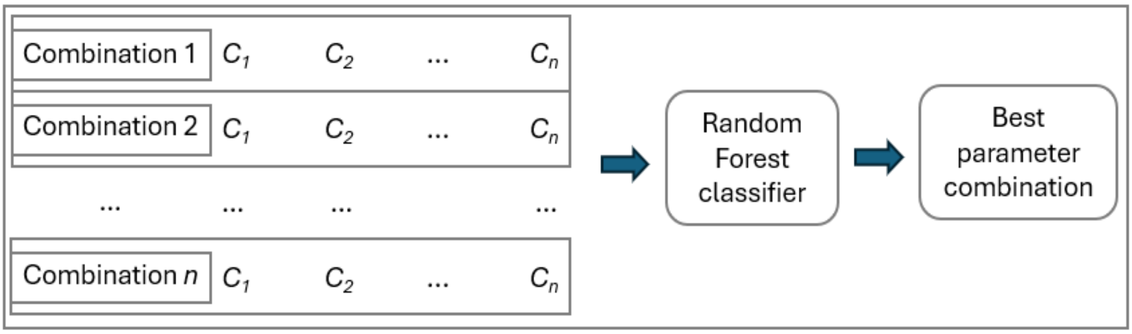


Fig. 2. Grid search mechanism

Table 1- RF parameters tuning for grid search

No	Parameters	Range
1	Number of tree	[50, 100, 200]
2	Maximum tree depth	[10, 20, 30]
3	Minimum sample split	[2, 5, 10]
4	Minimum sample leaf	[1, 2, 4]

3.5 Handling Imbalanced Dataset Scenario

Determining product quality in Fat Filled Powder (FFP) presents complex challenges, as it involves numerous features and two quality classes: *good* and *bad*. A common issue often encountered in binary attribute classification is the potential for an imbalanced dataset. Referring to (Mendez et al., 2019) (Kaitlin et al., 2020), the most critical aspect of addressing an imbalanced dataset lies not in selecting the right algorithm but in properly managing the dataset. Implementing the Random Forest algorithm with parameter tuning is sufficient to enhance model performance. Therefore, at this stage, several scenarios for configuring the training and testing datasets will be established to optimize the model in handling imbalanced datasets.

In this study, four scenarios are established for dataset configuration to enhance the performance of the Random Forest model with parameter tuning: data splitting, k-fold validation, Synthetic Minority Oversampling Technique (SMOTE), and Distribution Optimally Balanced – Stratified Cross Validation (DOB-SCV). Each scenario will be compared in terms of performance when applied to the training data to improve the model's performance on testing data. The details of each scenario are presented in Table 2.

Table 2 - Models for managing imbalanced dataset

No	Scenario	Detail
1	Model 1	Data splitting, training dan testing, 80:20
2	Model 2	Data splittting, training, validation dan testing, 60:20:20
3	SMOTE model	Oversampling in minority class for balance dataset (Chawla et al., 2002)
4	DOB SCV	Split data into subsamples while ensuring that there is no dominance of majority or minority classes within the subsamples.

3.6 K-fold validation

K-fold validation is one of the most reliable methods for improving model performance in classification tasks. It provides sufficient training data, allowing the model to learn knowledge from the dataset comprehensively, thereby maximizing model performance. Referring to [37], the k-fold validation scheme is illustrated in Figure 3.

In this study, a 10-fold validation is applied, with the data evenly divided across each fold. The performance of the model for each fold will be statistically compared with other models to assess the significance of performance improvements. Additionally, the data splitting approaches implemented in models 1 and 2 are combined with the k-fold validation method to achieve more stable and optimal results.

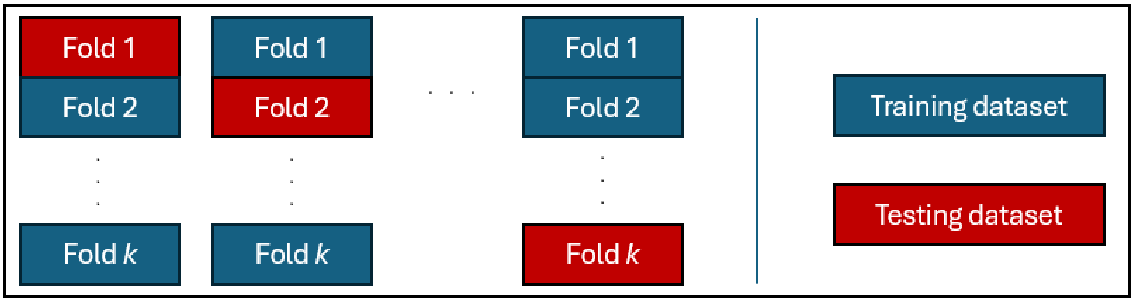


Fig. 3. K-Fold Validation Scenario

3.7 Synthetic Minority Oversampling Technique (SMOTE)

The SMOTE (Synthetic Minority Oversampling Technique) method was first proposed by Ref. (Chawla et al., 2002) to address issues in classification models with imbalanced class distributions. The main idea involves undersampling the majority class and oversampling the minority class to improve model performance and ROC value. According to (Chawla et al., 2002; Elreedy et al., 2024) (Burnaev et al., 2015), the SMOTE technique for generating a sample (s) can be described by Equation 2. The samples generated by SMOTE are linear combinations of two samples from the original data (x_o and x), where x_o is a randomly selected sample from five nearest neighbours of the original data, and $0 \leq u \leq 1$.

$$s = x + u \cdot (x^{r} - x) \tag{2}$$

The use of SMOTE for addressing imbalanced data conditions has been widely applied and proven to perform well in previous studies. Although SMOTE generates minority data synthetically, it achieves optimal classification performance when implemented with Random Forest (Bhagat & Patil, 2015). The synthetic data generated is also validly developed using the k-nearest neighbour (k-NN) approach, ensuring that it avoids outliers and maintains a distribution consistent with the original data.

3.8 Distribution Optimally Balanced – Stratified Cross Validation (DOB-SCV)

DOB-SCV (Distribution Optimally Balanced - Stratified Cross Validation) is an enhancement of cross-validation designed to improve classification model accuracy, as proposed by Ref. (X. Zeng & Martinez, 2000). This method is particularly suitable for imbalanced data since it divides the data into subsamples while ensuring that neither majority nor minority classes dominate within any subsample. Initially, Stratified Cross Validation was introduced as an improvement over k-fold validation to ensure that training data in each step was not dominated by the majority class. Subsequently, the SCV approach was optimized with Distribution Optimally Balanced (DOB) to ensure that the distribution of subsamples closely resembles the original data (Moreno-Torres et al., 2012).

To maintain the distribution of subsample data for training identical to the original data distribution, the DOB-SCV method ensures that neighboring samples within the same class are placed in different folds for training (X. Zeng & Martinez, 2000). According to (Moreno-Torres et al., 2012), the algorithm for DOB-SCV is detailed in Table 3.

Table 3- DOB SCV Algorithms adopted from (Szeghalmy & Fazekas, 2023a)

```
Require: k                  // number of folds
Require: C = {C_1, C_2, ..., C_n}   // classes
Ensure: F_1, F_2, ..., F_k          // generated folds

F_1 ← ∅, F_2 ← ∅, ..., F_k ← ∅
for i := 1 to n do
   while count(C_i) > 0 do
      x_1 ← randomly select sample from C_i
      F_1 ← F_1 ∪ {x_1}
      C_i ← C_i \ {x_1}
      for j := 2 to k do
         x_2 ← select the nearest neighbour of x_1 from C_i
         F_j ← F_j ∪ {x_2}
         C_i ← C_i \ {x_2}
         if count(C_i) = 0 then
            j ← k              // end for j
         end if
      end for
   end while
end for
```

3.9 Model evaluation

The model developed in this study is a classification model designed for imbalanced datasets. For classification models, evaluation can be conducted using several key parameters based on a bi-class confusion matrix. These parameters include accuracy, precision, specificity, and Receiver Operating Characteristic (ROC). The confusion matrix shows a comparison between the output of the classification model and the actual data. It calculates the number of outputs compared to the actual data based on true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The structure of the bi-class confusion matrix used in this study is presented in Table 4.

Table 4 - Confussion Matrix

		Predicted	
	Class label	Yes	No
Actual	Yes	TP	FP
	No	FN	TN

Accuracy represents the proportion of correctly classified results compared to the total classification data. Accurate classifications that match the actual class are indicated by True Positives (TP) and True Negatives (TN), compared against the classification results for all data. The calculation for accuracy is shown in Equation 3. Precision measures the model's performance with a focus on classifying positive classes. A high precision value indicates that the model can effectively classify data in the positive class. Precision is calculated as the total true positives divided by the sum of true positives and false positives, as shown in Equation 4.

Unlike precision, specificity focuses on evaluating the model's performance specifically for the negative class. Higher specificity indicates the model can effectively classify the negative class. Specificity is calculated using Equation 5. The Receiver Operating Characteristic (ROC) curve evaluates the performance of a classification model on binary target data, specifically showing the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR). The ROC provides a comprehensive view of the model's performance, even under imbalanced dataset conditions. The ROC's effectiveness is often summarized by the Area Under Curve (AUC), where higher values indicate better binary class classification performance. AUC is calculated using Equation 6. All evaluation parameters are measured for both positive and negative classes to assess the model's performance across both and ensure it functions well under imbalanced dataset conditions.

3.10 Performance improvement of the model across the four treatments is also tested statistically.

This testing ensures that there is a significant impact on the model's performance compared to previous models. The statistical test applied in this evaluation is the Mann-Whitney U test, which does not assume normal data distribution and compares the significance of performance differences between two models. The Mann-Whitney U test in this research is applied to analyse the performance of two models in predicting the FFP quality which were not normally distributed. To test the model with the mentioned statistical test, 10 folds model testing was proposed. This test is implemented due to the limited sample results available for performance testing and also to gain the validated result in providing a model recommendation for FFP quality evaluation for imbalanced dataset condition A confidence interval test is also applied to compare the results and ensure the significance effect of the model improvement.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

$$AUC = \int_0^1 TPR(FPR)d(FPR) \tag{6}$$

4. Results and Discussions

4.1 Data description

The data was obtained from a dairy milk industry that processes fat-filled powder (FFP). The statistical description of the data can be seen in Table 5, Figure 3, and Figure 4. The raw data collected from the field is clean and complete. The dataset contains seven features that determine the quality of the FFP product.

The main issue with the data for classification modeling lies in the condition of the target data and quality labels. It is evident that there is an imbalance in the sample sizes for the two classes: *good* (labeled as 1) and *bad* (labeled as 0). As shown in Figure 3, only 17.16% of the data has a value of 0 (poor quality), while the remaining 82.84% has a value of 1 (good quality). This kind of condition will pose new challenges in developing a prediction model that will only be accurate for the majority class, while errors will occur for the minority class.

Table 5 - Data descriptions

Features	count	mean	std	Min	25%	50%	75%	max
Fat	1525	32.67829	0.575109	27.03	32.19	32.82	33.13	34.64
Protein	1525	19.14013	1.126821	0.128576	18.9655	19.21042	19.42594	20.71904
Lactose	1525	39.14352	0.338847	37.7	38.94	39.08	39.33	42.68
Ash	1525	5.206193	0.304357	4.1	4.9	5.31	5.45	6.27
Moisture	1525	2.281738	0.136777	1.4	2.2	2.28	2.36	2.72
Acidity	1525	10.45084	0.466086	8.8227	10.1	10.32	10.82	12.09
Bulk Density	1525	0.547451	0.506464	0.454545	0.510204	0.520833	0.526316	9.85222
Label	1525	0.82623	0.379036	0	1	1	1	1

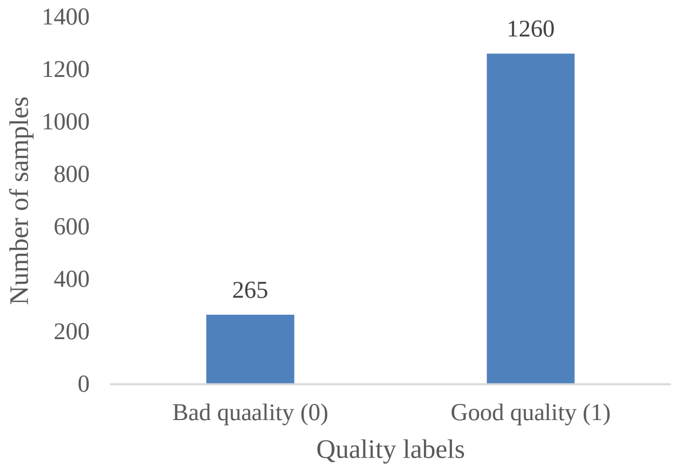


Fig. 4. Imbalanced data set description with FFP good quality (label 1) versus bad quality (label 0)

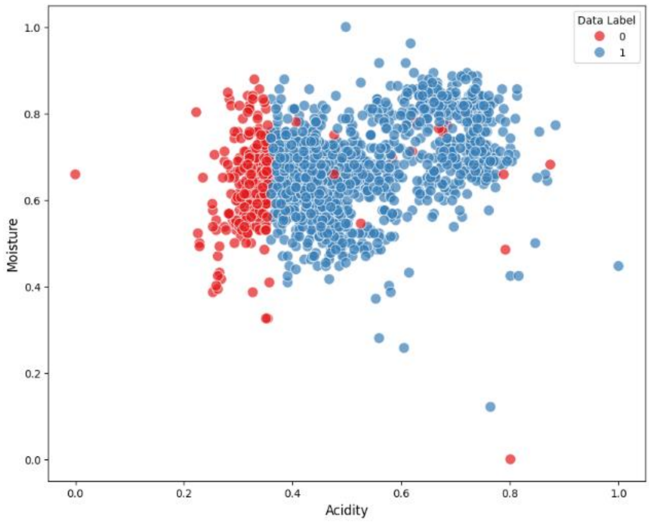


Fig. 5. Data distribution with normalized features

Figure 5 interprets that the majority of the data represents good quality compared to bad quality. Both quality types are clearly distinguished and clustered into separate groups. The minority class (bad quality) appears denser compared to the majority class (good quality). Additionally, there is an observed overlap where some minority class data points fall within the majority class cluster. This poses new challenges and indicates the potential presence of outliers or noise. Furthermore, a more detailed analysis is needed using other features to avoid greater model errors. Therefore, data transformation and preprocessing must be effectively applied to support the accuracy of the classification-based prediction model.

4.2 FFP quality prediction with random forest

Machine learning has strong capabilities in identifying data patterns, provided the data is complete, free of noise and accurate. However, in practice, various challenges are often encountered that affect the accuracy of quality predictions. The issue of imbalanced data samples, as found in this case, can impact the accuracy of quality predictions, which in turn affects customer satisfaction in achieving the product quality as standard. If the imbalance in the data is left unaddressed, the machine learning classification model will learn partially and perform well only for one of the label conditions, leading to suboptimal performance in quality classification tasks.

Improving the accuracy of the model for predicting quality across various product quality conditions needs to be addressed effectively. This study proposes the development of a baseline Random Forest model to evaluate the capability of machine learning in accurately predicting product quality. The Random Forest model is proposed to learn data patterns for predicting the quality of FFP under conditions of imbalanced class labels. Random forest is also proposed since it achieve an excellence performance to classify imbalanced dataset condition as mentioned in previous research.

The parameter tuning method using the grid partition technique is proposed to identify the best parameters for developing a Random Forest model to predict FFP (Fat-Filled Powder) quality under imbalanced conditions. After performing parameter tuning with the grid partition method, the parameters set for the quality classification model are presented in Table 6.

Table 6 - Parameter tuning for random forest using grid partition

No	Parameter	Value
1	Algorithm	Random forest classifier
2	Max depth	None
3	Minimum sample split	10
4	Minimum sample leaf	4
5	Number of estimators	100
6	Random state	42

The Random Forest model with parameter tuning was tested under two data splitting conditions: 80% training data (Model 1) and 60% training data (Model 2), as described in Table 2. This testing is necessary to evaluate the model's sensitivity and accuracy with respect to the amount of training data and its impact on testing data, ensuring practical implementation in the field. The results of the model testing are presented in Table 7.

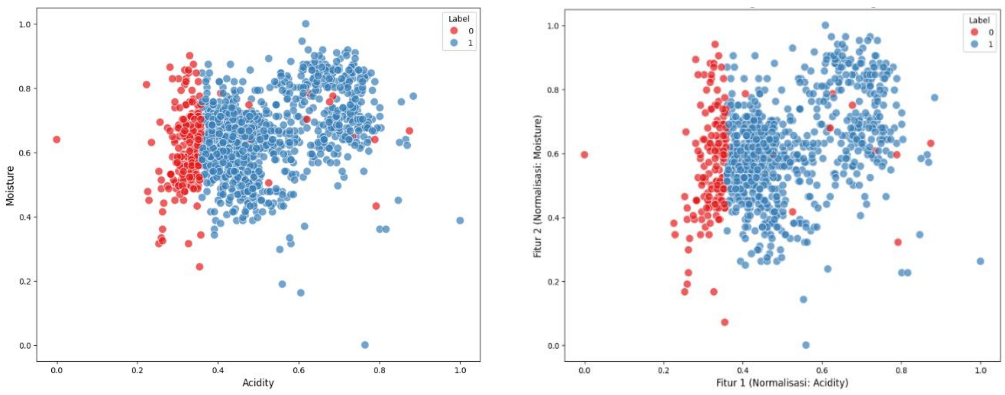
Table 7- Model performance in two data splitting scenarios

Parameter	Training		Testing	
	Model 1	Model 2	Model 1	Model 2
Jumlah data	80%	60%	20%	20%
Akurasi	0.9959	0.9967	0.9902	0.9902
ROC	1.0000	0.9999	0.9996	0.9992
Precision - class 0	1.0000	1.0000	1.0000	1.0000
Precision - class 1	0.9951	0.9961	0.9876	0.9876
Recall - class 0	0.9749	0.9800	0.9545	0.9545
Recall - class 1	1.0000	1.0000	1.0000	1.0000
F1-Score – class 0	0.9873	0.9899	0.9767	0.9767
F1-Score – class 1	0.9976	0.9980	0.9938	0.9938
Support – class 0	199	150	66	66
Support – class 1	1021	765	239	239

In general, the model's accuracy under both data conditions shows excellent score. However, this result can be biased since it affected by imbalanced data scenarios that accuracy is not the only evaluation metric that should be considered. For imbalanced datasets, it is essential to also consider precision, recall, and support to gain a comprehensive view of the model's performance (Thabtah et al., 2020).

Table 7 indicates that the model achieves good accuracy for both 80% and 60% training data splits. However, upon closer examination, the classification performance for recall in class 0—both in training and testing— shows no improvement. Special attention is needed for the classification of class 0 as minority class label.

The model has confirmed that under conditions of limited and imbalanced data, machine learning struggles to fully capture data patterns effectively. On the other hand, the test results also indicate that increasing the training data by 20% does not lead to any improvement in the model's recall performance during training or testing. The data distribution for Model 1 and Model 2 is illustrated more clearly in Figure 6.



(a) (b)

Fig. 6. Data training distribution for (a) Model 1 and (b) model 2

To analyze the model's performance and the differences between Model 1 and Model 2, an appropriate statistical analysis is required to recommend an accurate conclusion. To support this statistical analysis, the fold validation approach is utilized as a mechanism to gather model performance data across each fold. The models are developed using 10-fold validation to test the impact of increasing training data on model accuracy. The analysis is focused on recall performance since it able to capture model performance in imbalanced dataset condition. In the Random Forest model developed with 10-fold validation, the recall performance for Model 1 (80% training data) and Model 2 (60% training data) can be observed in Table 8.

The statistical test results for recall performance on class 1 (good quality) for Model 1 and Model 2 indicate that the performance is stable. The models effectively classify class 1 due to the sufficient amount of data available. Random Forest is able to capture data patterns effectively, resulting in satisfactory recall performance for the majority class. However, the recall performance for class 0 (poor quality) as minority class shows unsatisfactory and unstable results, with variance still remaining high. Both Model 1 and Model 2 exhibit unstable performance and have statistically insignificant recall performance differences. These results indicate that adjusting the amount of training data does not significantly impact the model's performance, particularly for the minority class. This highlights the need for specialized handling of the minority class to improve its classification performance.

Tabel 8 - Recall performance of training dataset for model 1 and model 2

Parameter	Class 0		Class 1	
	Model 1	Model 2	Model 1	Model 2
Recall fold 1	0.9791	0.9650	1.0000	1.0000
Recall fold 2	0.9738	0.9650	1.0000	1.0000
Recall fold 3	0.9843	0.9790	1.0000	1.0000
Recall fold 4	0.9791	0.9720	1.0000	1.0000
Recall fold 5	0.9634	0.9650	1.0000	1.0000
Recall fold 6	0.9686	0.9792	1.0000	1.0000
Recall fold 7	0.9634	0.9510	1.0000	1.0000
Recall fold 8	0.9634	0.9580	1.0000	1.0000
Recall fold 9	0.9737	0.9441	1.0000	1.0000
Recall fold 10	0.9842	0.9580	1.0000	1.0000
t-statistic	73.0		Insignificant	
p-value	0.0877		Insignificant	
Result	Insignificant		Insignificant	

4.3 Improving recall model performance in imbalanced data condition for minority class

Imbalanced data conditions are often encountered in datasets with binary target data or only two classes in classification. If imbalanced data is not handled properly, it can lead to inaccurate classifications, which may harm consumers when related to quality classification. In such cases, the algorithm tends to learn patterns primarily from the majority class, while the minority class is prone to misclassification (López et al., 2014).

Various models have been proposed to address issues with imbalanced datasets, including k-fold validation (López et al., 2014; Szeghalmy & Fazekas, 2023b), stratified cross validation (X. Zeng & Martinez, 2000) and SMOTE (Thabtah et al., 2020). The k-fold validation which is evaluated in this study showed no significant impact from increasing the amount of training data. Therefore, other approaches need to be formulated to optimize model performance in classifying imbalanced datasets. In this case, stratified cross-validation with distribution optimally balanced (SCV-DOB) and oversampling techniques through SMOTE are utilized.

The SCV-DOB method, which has been widely applied in previous studies, demonstrates strong potential for improving models through optimal data partitioning. The SCV-DOB technique effectively manipulates data samples, enabling the model to fully learn from all class labels, including both majority and minority classes. Given that the model's performance with traditional cross-validation still shows low recall metrics, the SCV-DOB model was tested using 10-fold validation to evaluate its impact on improving model performance. The results of the model evaluation are presented in Table 9.

It can be observed that the model using the SCV-DOB method still does not meet performance expectations and does not significantly improve evaluation performance, especially for the minority class. Pre-training stages, including data splitting through k-fold validation and stratified validation, have been attempted but have not yielded significant impacts. The data splitting approach with SCV-DOB also indicates no notable improvement in model performance compared to Model 1 (with 80% training data) evaluated using 10-fold validation.

The result also found that the SCV-DOB method does not yet demonstrate a significant improvement in model performance. The SCV-DOB model, in the preprocessing stage, does not fully perform data manipulation or transformation to support enhanced model performance. SCV-DOB is merely a data sampling approach conducted optimally to ensure that both minority and majority classes have equal opportunities to serve as the basis for model development in each testing fold. Under such conditions, since the "bad quality" class has a minority share of data, not all real-world data conditions can be adequately captured. As a result, the model's recall performance does not significantly differ from that of the basic model.

Table 9 - DOB-SCV model performance

Parameter	Class 0		Class 1	
	Model 1	DOB-SCV	Model 1	DOB-SCV
Recall fold 1	1.0000	0.9791	1.0000	1.0000
Recall fold 2	1.0000	0.9738	1.0000	1.0000
Recall fold 3	1.0000	0.9843	1.0000	1.0000
Recall fold 4	0.8519	0.9791	1.0000	1.0000
Recall fold 5	0.9259	0.9634	1.0000	1.0000
Recall fold 6	0.9615	0.9686	1.0000	1.0000
Recall fold 7	1.0000	0.9634	1.0000	1.0000
Recall fold 8	0.9615	0.9634	1.0000	1.0000
Recall fold 9	0.9231	0.9737	1.0000	1.0000
Recall fold 10	1.0000	0.9842	1.0000	1.0000
t-statistik	50.0		Insignificant	
p-value	1.0		Insignificant	
Result	Insignificant		Insignificant	

The data splitting approach using SCV-DOB has shown insignificant results. Oversampling approaches need to be considered as an option to improve model performance. Undersampling methods are not considered in this study because they risk removing some information from the data (Mujahid et al., 2024). This condition poses a risk of the model failing to accurately predict quality in alignment with real-world conditions.

In the subsequent modeling, the SMOTE method is considered as an oversampling approach to improve model performance. SMOTE performs oversampling on minority class samples by generating synthetic data close to the original data. This ensures that the number of samples in the minority class becomes comparable to the majority class. Under this condition, the model is expected to learn data patterns effectively, thereby optimizing its performance across both classes. The model's performance using the SMOTE method can be observed in Table 10.

Table 10 - SMOTE model performance

Parameter	Class 0		Class 1	
	Model 1	SMOTE	Model 1	SMOTE
Recall fold 1	0.9791	0.9946	1.0000	1.0000
Recall fold 2	0.9738	0.9946	1.0000	1.0000
Recall fold 3	0.9843	0.9967	1.0000	1.0000
Recall fold 4	0.9791	0.9946	1.0000	1.0000
Recall fold 5	0.9634	0.9946	1.0000	1.0000
Recall fold 6	0.9686	0.9946	1.0000	1.0000
Recall fold 7	0.9634	0.9946	1.0000	1.0000
Recall fold 8	0.9634	0.9946	1.0000	1.0000
Recall fold 9	0.9737	0.9946	1.0000	1.0000
Recall fold 10	0.9842	0.9946	1.0000	1.0000
t-statistic	0.0		Insignificant	
p-value	8.45		Insignificant	
Result	Significant		Insignificant	

To ensure that the test results are valid, a statistical confidence interval test was also conducted to demonstrate that the SMOTE model produces consistent and reliable results. A 95% confidence interval (CI) test was performed to verify that the recall performance of the model differs significantly from the baseline model. The CI test results for FFP quality classes are presented in Table 11.

Table 11 - Model confidence interval test

Parameter	Class 0		Class 1	
	Model 1	SMOTE	Model 1	SMOTE
Mean recall	0.9732	0.9945	1.00	1.00
CI lower	0.9673	0.9945	1	1
CI upper	0.9792	0.9946	1	1

The CI test results show that the recall score range for Class 0 (bad quality) in Model 1 still varies between 0.9673 and 0.9792, whereas the SMOTE model demonstrates highly consistent recall scores with minimal variation across tests. This finding also confirms that the confidence intervals of both models do not overlap, indicating that the SMOTE model has statistically significantly different recall performance compared to the baseline model (Model 1).

For Class 1, since the baseline model already performs well, the application of SMOTE does not yield a significant improvement but still maintains optimal performance. Therefore, the statistical analysis using both the Mann-Whitney U test and the confidence interval confirms that the SMOTE model, combined with the Random Forest algorithm, is effective in enhancing prediction performance for the minority class.

The test results for the model using SMOTE show that the model performs well for the majority class. Additionally, the model has demonstrated stable recall performance across all folds, particularly for class 0 (the minority class). To confirm the improvement in model performance, statistical tests have validated that there is a significant difference in recall performance between the baseline model and the model using the SMOTE oversampling technique. A comparison of recall performance across various models is illustrated in Figure 6, which highlights that the SMOTE model achieves the best performance.

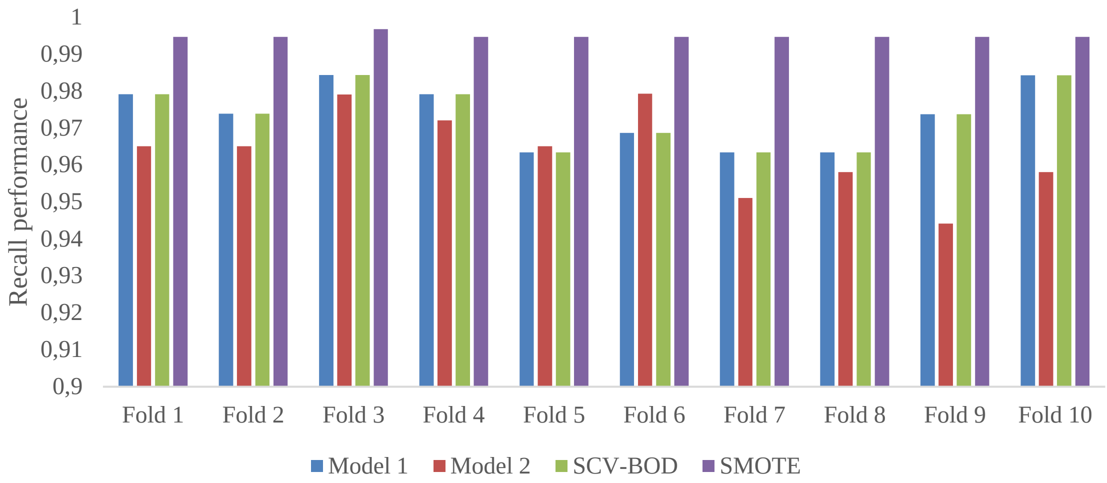


Fig. 7. Recall model performance on minority class by various model

The characteristics and performance of the SMOTE-based quality prediction model can be observed in Table 11. The SMOTE model performs oversampling on minority class data, resulting in an increased number of samples during model development. It is evident that the model achieves a balanced dataset for both classes, with 1,021 samples each. This balance is achieved through SMOTE oversampling the minority class, enabling the model to learn from a larger amount of data.

The model demonstrates strong performance in precision, recall, and F1-score for both majority and minority classes. Additionally, it exhibits excellent and balanced accuracy and ROC metrics between training and testing phases, effectively avoiding issues of overfitting or underfitting. This indicates that the SMOTE-enhanced model is well-suited for handling imbalanced datasets while maintaining robust and reliable predictions.

Table 11 - SMOTE model performance for FFP quality - imbalanced dataset

Parameter	Training	Testing
Accuracy	0.9976	0.9967
ROC	1.0000	0.9990
Precision - class 0	1.0000	1.0000
Precision - class 1	0.9951	0.9958
Recall - class 0	0.9951	0.9848
Recall - class 1	1.0000	1.0000
F1-Score – class 0	0.9975	0.9924
F1-Score – class 1	0.9976	0.9979
Support – class 0	1021	66
Support – class 1	1021	239

As a confirmation and comparison of model performance across all evaluation metrics, refer to Figures 8 and 9. It can be observed that the quality prediction model using Random Forest with SMOTE outperforms other models in both training and testing phases. This further strengthens confidence that this model can be effectively implemented for predicting the quality of FFP under imbalanced dataset conditions, a scenario commonly encountered in the industry.

The SMOTE model demonstrates the best performance among other preprocessing and data transformation models due to the distinct treatment applied to the data. SMOTE manipulates the minority class data using the k-NN technique so that the minority data points have similar characteristics. This helps the random forest prediction model to better learn the minority class, as the number of samples becomes more comparable to that of the majority class. In contrast, other models—such as those using 80% or 60% test data splits or even SCV-DOB—do not manipulate the data but merely select which data will be included in the prediction model.

Statistically, this approach has been proven not to improve the performance of the prediction model.

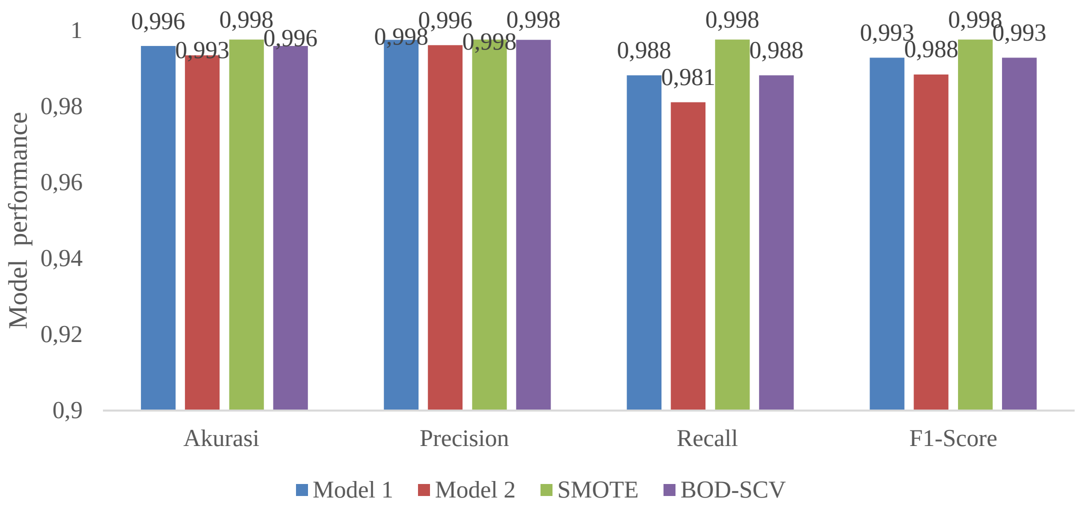


Fig. 8. Model performance on training dataset

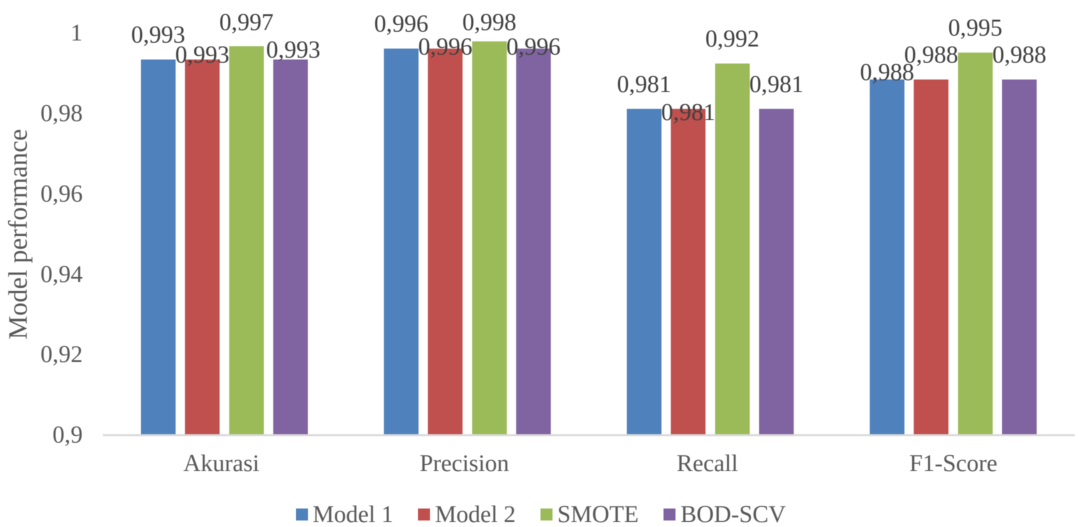


Fig. 9. Model performance on Testing dataset

4.4 Research limitations and real-world feasibility

This study has successfully developed a predictive model for FFP product quality under imbalanced data conditions. The research focused on imbalanced data transformation and the development of a prediction model. Four data transformation schemes were developed for model development: splitting the data into training and testing sets using 80% and 60% for training, the Synthetic Minority Oversampling Technique (SMOTE), and the Distribution Optimally Balanced – Stratified Cross Validation (DOB-SCV) scheme. The study shows that the SMOTE model can improve the performance of the predictive model using random forest.

Although this study has successfully enhanced the performance of the predictive model, attention must be paid to the model's generalizability in predicting product quality in real-world settings. As SMOTE generates synthetic minority data using the k-NN technique, there is a possibility that the synthetic data may not truly resemble the actual data. This could introduce bias, but it may still be helpful in developing predictive models for product quality under uncertainty. Ideally, minority class data should continue to be collected in real-world operations to gradually replace the synthetic data generated by SMOTE and it need to apply re-train model.

On the other hand, the predictive model developed in this study is based on the random forest algorithm. Random forest was chosen due to its advantages in predicting binary-class product quality under imbalanced data conditions. Therefore, all model performance results in this study refer to the predictions produced by the random forest model. Although the model's performance has been successfully improved, this does not guarantee that similar improvements will apply to other classification models. Further development and testing of alternative models are necessary to explore other potential predictive approaches.

As a practical step toward implementing this model, several aspects need to be considered. The model has been successfully developed to predict the quality of FFP products produced in the dairy industry and its derivatives. This model can be installed in an FFP quality monitoring system within the industry to monitor product quality in real time. To support this, sensor-based IoT systems should be considered for real-time data collection, enabling accurate and real-time product quality predictions.

5. Conclusion and recommendations

The quality of perishable products like fat-filled powder (FFP) is a crucial and essential aspect. Quality is vital for both producers and consumers to ensure that the product meets specifications and consumer expectations. As a manufactured product influenced by a series of processes, the quality of FFP is challenging to control. The continuous nature of production processes requires accurate, rapid, and high-precision quality classification. In reality, most products in the industry currently have good quality compared to those with poor quality. This is due to the industry's tendency to achieve production stability, resulting in fewer low-quality products. However, this does not mean that product quality should be neglected, as there is still a significant risk of producing defective products during the manufacturing process.

The primary challenge lies in the digital transformation within the industry. Process control requires a digitalization and optimization approach. Data analytics and machine learning techniques are essential for classifying product quality based on production patterns and historical quality data of FFP. However, the challenge arises from the limitations of historical data used as samples, where the amount of poor-quality data is significantly smaller than that of good-quality products. This imbalance makes it difficult for the model to learn data patterns effectively and increases the risk of failure in classifying product quality accurately based on real-world conditions.

This study proposed a machine learning model using Random Forest to classify the quality of FFP products. The issue of data imbalance was addressed using various approaches, including hyperparameter tuning, data splitting for training and testing, k-fold validation, oversampling, and stratified sampling. The use of a grid partition model for hyperparameter tuning successfully defined the optimal parameters for product quality classification. However, the model initially exhibited low performance in classifying the minority quality class, necessitating improvements in its performance. Various models were proposed to enhance performance, including increasing the training data proportion compared to testing, oversampling (SMOTE), and stratified sampling (DOB-SCV). The modeling results revealed that the oversampling model using the SMOTE technique achieved the best performance for both majority and minority classes compared to other models. The Random Forest model with hyperparameter tuning combined with the SMOTE oversampling method achieved a testing accuracy of 0.997 and a testing ROC of 0.999, indicating optimal performance. This model also demonstrated stable precision and recall for both majority and minority classes.

The utilization of SMOTE in model development aims to enable the model to comprehensively learn data patterns for both majority and minority classes. This predictive model has successfully classified product quality accurately, making it suitable for implementation in the FFP production industry. In practice, this model allows the industry to quickly determine product quality during continuous production processes, enabling rapid response in case of production errors. Quality failures can also be mitigated promptly, minimizing the number of poor-quality products and ensuring that consumer expectations are met.

This study successfully developed a classification model for the quality of perishable FFP products. However, the data collected was still manually retrieved from a database. Future research should focus on implementing a more practical model for industrial applications, enabling the model to quickly detect quality parameter values and accurately determine product quality in real-time. The utilization of Internet of Things (IoT) technology is necessary for detecting quality parameter values during production. Data collected via IoT-based sensors can be stored in the cloud for further processing. Implementing a model supported by IoT-based sensors will significantly accelerate the process of detecting product quality effectively. In addition, the same research approach can be extended to other dairy-derived products or other perishable goods. In the industry, imbalanced data conditions are commonly encountered, requiring special treatment in developing product prediction models to support quality and customer satisfaction. The framework developed in this study, which utilizes data transformation with SMOTE and predictive modeling with Random Forest, can be applied to develop predictive models for other dairy derivatives or perishable products in the industry.

Data availability statements

The data that support the findings of this study are openly available at: 10.6084/m9.figshare.29115317

Author contributions statements

Conceptualization: M.A., Methodology: M.A., M.A., Software: O.P., Formal Analysis: M.A., and O.P., Investigation: M.A., O.P., Data Curation: O.P., Writing Origital Draft: M.A., Writing – Review & Editing: M.A., O.P.

References

Anggoro, D. A., & Mukti, S. S. (2021). Performance Comparison of Grid Search and Random Search Methods for Hyperparameter Tuning in Extreme Gradient Boosting Algorithm to Predict Chronic Kidney Failure. *International Journal of Intelligent Engineering and Systems*, *14*(6), 198–207. https://doi.org/10.22266/ijies2021.1231.19

Bahel, V., Pillai, S., & Malhotra, M. (2020). A Comparative Study on Various Binary Classification Algorithms and their Improved Variant for Optimal Performance. *2020 IEEE Region 10 Symposium, TENSYMP 2020*, *January 2020*, 495–498. https://doi.org/10.1109/TENSYMP50017.2020.9230877

Bhagat, R. C., & Patil, S. S. (2015). Enhanced SMOTE algorithm for classification of imbalanced big-data using Random Forest. *Souvenir of the 2015 IEEE International Advance Computing Conference*, *IACC 2015*, 403–408. https://doi.org/10.1109/IADCC.2015.7154739

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1109/ICCECE51280.2021.9342376

Burnaev, E., Erofeev, P., & Papanov, A. (2015). Influence of resampling on accuracy of imbalanced classification. *Eighth International Conference on Machine Vision (ICMV 2015)*, *9875*(Icmv), 987521. https://doi.org/10.1117/12.2228523

Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*(2002), 321–357. https://doi.org/10.46880/jmika.vol4no1.pp67-72

Cockburn, M. (2020). Review: Application and prospective discussion of machine learning for the management of dairy farms. *Animals*, *10*(9), 1–22. https://doi.org/10.3390/ani10091690

Deniz, A., Kiziloz, H. E., Dokeroglu, T., & Cosar, A. (2017). Robust multiobjective evolutionary feature subset selection algorithm for binary classification using machine learning techniques. *Neurocomputing*, *241*, 128–146. https://doi.org/10.1016/j.neucom.2017.02.033

Elreedy, D., Atiya, A. F., & Kamalov, F. (2024). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning*, *113*(7), 4903–4923. https://doi.org/10.1007/s10994-022-06296-4

Feng, Z., Liu, D., Gu, J., & Zheng, L. (2024). Raman spectroscopy and fusion machine learning algorithm: A novel approach to identify dairy fraud. *Journal of Food Composition and Analysis*, *129*, 106090. https://doi.org/10.1016/J.JFCA.2024.106090

Finnegan, E. W., Mahomud, M. S., Murphy, E. G., & O'Mahony, J. A. (2021). The influence of pre-heat treatment of skim milk on key quality attributes of fat filled milk powder made therefrom. *International Journal of Dairy Technology*, *74*(2), 404–413. https://doi.org/10.1111/1471-0307.12758

Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. *International Journal of Emerging Technology and Advanced Engineering*, *2*(4), 42–47. http://www.ijetae.com/files/Volume2Issue4/IJETAE_0412_07.pdf

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, *73*, 220–239. https://doi.org/10.1016/J.ESWA.2016.12.035

Han, J., Li, T., He, Y., & Gao, Q. (2022). Using Machine Learning Approaches for Food Quality Detection. *Mathematical Problems in Engineering*, *2022*(1), 6852022. https://doi.org/10.1155/2022/6852022

Kaitlin, Trace, Smith, & Sadler, B. (2020). Identifying new X-ray binary candidates in M31 using random forest classification. *Monthly Notices of the Royal Astronomical Society*, *492*(4), 5075–5088. https://doi.org/10.1093/MNRAS/STAA207

Kene Ejeahalaka, K., & On, S. L. W. (2020). Effective detection and quantification of chemical adulterants in model fat-filled milk powders using NIRS and hierarchical modelling strategies. *Food Chemistry*, *309*, 125785. https://doi.org/10.1016/J.FOODCHEM.2019.125785

Kokkinos, K., Papageorgiou, E., Dafopoulos, V., & Ioannis, A. (2017). Efficiency in Energy Decision Support Systems Using Soft Computing Techniques. In A. Kumar, S. Ajith, A. Patrick, & S. Michael (Eds.), *Intelligent Decision Support Systems for Sustainable Computing* (pp. 33–52). Springer Nature.

Kumar, A., & Agrawal, S. (2024). Enhancing quality-based classification of perishable products: a convolutional neural network approach with statistical hyperparameter optimization. *Multimedia Tools and Applications*, 1–24. https://doi.org/10.1007/S11042-024-19700-Z/METRICS

Liu, F., & Dai, Y. (2022). Product Processing Quality Classification Model for Small-Sample and Imbalanced Data Environment. *Computational Intelligence and Neuroscience*, *2022*, 1–16. https://doi.org/10.1155/2022/9024165

López, V., Fernández, A., & Herrera, F. (2014). On the importance of the validation technique for classification with imbalanced datasets: Addressing covariate shift when data is skewed. *Information Sciences*, *257*, 1–13. https://doi.org/10.1016/j.ins.2013.09.038

Mahmudah, K. R., Indriani, F., Takemori-sakai, Y., Iwata, Y., Wada, T., & Satou, K. (2021). Classification of imbalanced data represented as binary features. *Applied Sciences (Switzerland)*, *11*(17), 7825. https://doi.org/10.3390/APP11177825/S1

Mendez, K. M., Reinke, S. N., & Broadhurst, D. I. (2019). A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification. *Metabolomics*, *15*(12), 1–15. https://doi.org/10.1007/s11306-019-1612-4

Moreno-Torres, J. G., Saez, J. A., & Herrera, F. (2012). Study on the impact of partition-induced dataset shift on k-fold cross-validation. *IEEE Transactions on Neural Networks and Learning Systems*, *23*(8), 1304–1312. https://doi.org/10.1109/TNNLS.2012.2199516

Mu, F., Gu, Y., Zhang, J., & Zhang, L. (2020). Milk source identification and milk quality estimation using an electronic nose and machine learning techniques. *Sensors (Switzerland)*, *20*(15), 1–14. https://doi.org/10.3390/s20154238

Mujahid, M., Kına, E., Rustam, F., Gracia Villar, M., Alvarado, E. S., De, I., Diez, L. T., & Ashraf, I. (2024). Open Access Journal of Big Data Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering. *Journal of Big Data*, *11*, 87. https://doi.org/10.1186/s40537-024-00943-4

Murphy, S. I., Reichler, S. J., Martin, N. H., Boor, K. J., & Wiedmann, M. (2021). Machine Learning and Advanced Statistical Modeling Can Identify Key Quality Management Practices That Affect Postpasteurization Contamination of Fluid Milk. *Journal of Food Protection*, *84*(9), 1496–1511. https://doi.org/10.4315/JFP-20-431

Naik, N., & Purohit, S. (2017). Comparative Study of Binary Classification Methods to Analyze a Massive Dataset on Virtual Machine. *Procedia Computer Science*, *112*(2017), 1863–1870. https://doi.org/10.1016/j.procs.2017.08.232

Pounds, K., Bao, H., Luo, Y., De, J., Schneider, K., Correll, M., & Tong, Z. (2022). Real-Time and Rapid Food Quality Monitoring Using Smart Sensory Films with Image Analysis and Machine Learning. *ACS Food Science and Technology*, *2*(7), 1123–1134. https://doi.org/10.1021/acsfoodscitech.2c00124

Purwantiningsih, T. I., Bria, M. A. B., & Kia, K. W. (2022). Levels Protein and Fat of Yoghurt Made of Different Types and Number of Cultures. *Journal of Tropical Animal Science and Technology*, *4*(1), 66–73. https://doi.org/10.32938/jtast.v4i1.967

Qin, W., Zhuang, Z., Guo, L., & Sun, Y. (2022). A hybrid multi-class imbalanced learning method for predicting the quality level of diesel engines. *Journal of Manufacturing Systems*, *62*(April), 846–856. https://doi.org/10.1016/j.jmsy.2021.03.014

Shekar, B. H., & Dagnew, G. (2019). Grid search-based hyperparameter tuning and classification of microarray cancer data. *2019 2nd International Conference on Advanced Computational and Communication Paradigms, ICACCP 2019, November*, 1–8. https://doi.org/10.1109/ICACCP.2019.8882943

Singh, A., Vaidya, G., Jagota, V., Darko, D. A., Agarwal, R. K., Debnath, S., & Potrich, E. (2022). Recent Advancement in Postharvest Loss Mitigation and Quality Management of Fruits and Vegetables Using Machine Learning Frameworks. *Journal of Food Quality*, *2022*(1), 6447282. https://doi.org/10.1155/2022/6447282

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, *134*, 93–101. https://doi.org/10.1016/j.eswa.2019.05.028

Swain, S., & Jenamani, M. (2023). Federated Learning for Temperature Break Identification in a Reefer Container from IoT Data. *2023 4th International Conference on Data Analytics for Business and Industry, ICDABI 2023*, 671–676. https://doi.org/10.1109/ICDABI60145.2023.10629315

Szeghalmy, S., & Fazekas, A. (2023a). A Comparative Study of the Use of Stratified Cross-Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced Learning. *Sensors*, *23*(4). https://doi.org/10.3390/s23042333

Szeghalmy, S., & Fazekas, A. (2023b). A Comparative Study of the Use of Stratified Cross-Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced Learning. *Sensors*, *23*(4), 1–27. https://doi.org/10.3390/s23042333

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences*, *513*, 429–441. https://doi.org/10.1016/j.ins.2019.11.004

Vignolles, M. L., Lopez, C., Le Floch-Fouéré, C., Ehrhardt, J. J., Méjean, S., Jeantet, R., & Schuck, P. (2010). Fat supramolecular structure in fat-filled dairy powders: A tool to adjust spray-drying temperatures. *Dairy Science and Technology*, *90*(2–3), 287–300. https://doi.org/10.1051/dst/2009057

Zeng, W., Jia, J., Zheng, Z., Xie, C., & Guo, L. (2011). A comparison study: Support vector machines for binary classification in machine learning. *Proceedings - 2011 4th International Conference on Biomedical Engineering and Informatics, BMEI 2011*, *3*(Vc), 1621–1625. https://doi.org/10.1109/BMEI.2011.6098517

Zeng, X., & Martinez, T. R. (2000). Distribution-Balanced Strati ed Cross-Validation for Accuracy Estimation 1 Introduction. *Journal of Experimental & Theoretical Artificial Intelligence*, *12*(1), 1–12.

Zhang, R., Zhou, L., Zuo, M., Zhang, Q., Bi, M., Jin, Q., & Xu, Z. (2018). Prediction of Dairy Product Quality Risk Based on Extreme Learning Machine. *Proceedings - 2nd International Conference on Data Science and Business Analytics, ICDSBA 2018*, 448–456. https://doi.org/10.1109/ICDSBA.2018.00090

Zhang, Y., Zhang, L., Ma, Y., Guan, J., Liu, Z., & Liu, J. (2022). Research on dairy products detection based on machine learning algorithm. *MATEC Web of Conferences*, *355*, 03008. https://doi.org/10.1051/matecconf/202235503008

Zhou, H., Yu, K. M., Chen, Y. C., & Hsu, H. P. (2021). A Hybrid Feature Selection Method RFSTL for Manufacturing Quality Prediction Based on a High Dimensional Imbalanced Dataset. *IEEE Access*, *9*, 29719–29735. https://doi.org/10.1109/ACCESS.2021.3059298

Zhu, N., Zhu, C., Zhou, L., Zhu, Y., & Zhang, X. (2022). Optimization of the Random Forest Hyperparameters for Power Industrial Control Systems Intrusion Detection Using an Improved Grid Search Algorithm. *Applied Sciences (Switzerland)*, *12*(20). https://doi.org/10.3390/app122010456