ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Cost-Efficient Sequential Predictions with a Hybrid Method of Topological Sorting and Boosting: A Case Study on LDPE-Property Prediction

Noparat Phongthakun[1], Akarapon Watcharapalakorn[2], None Wanitchollakit[3], Borworntat Dendumrongkul[4], Pana Wanitchollakit[5], Chayanin Kongsareekul[6], Sunisa Rimcharoen[7] and Nutthanon Leelathakul[8]

ABSTRACT

Determining low-density polyethylene (LDPE) properties typically requires extensive laboratory testing, which is time-consuming and costly. For instance, the conditioning phase alone for measuring Vicat softening temperature requires a minimum of 40 hours [1]. Predictive modeling can reduce these costs. Ensuring accuracy that meets manufacturing standards, however, is challenging. This paper introduces TopSABoost, a hybrid method that combines topological sorting and boosting techniques to perform sequential predictions of LDPE properties and minimize the overall laboratory-testing cost. This approach reduces laboratory testing costs by predicting one property first and using it to predict another. The complexity analysis demonstrates that the proposed algorithm is ideal for non-real-time determination of sequential predictions, as it computes the model offline once for repeated use without requiring recalculations, aligning with manufacturing needs. The experimental results demonstrate that TopSABoost achieves a maximum error of just 0.11%, satisfying strict manufacturing constraints. TopSABoost identifies that prioritizing the prediction of L-value, followed by Density, offers the most cost-efficient sequence and significantly reduces reliance on direct laboratory testing while maintaining adherence to error thresholds.

Article information:

Keywords: LDPE, Chemical Property, Sequential Prediction, Boosting Algorithm, Topological Sorting

Article history:
Received: December 24, 2024
Revised: January 23, 2025
Accepted: March 18, 2025
Published: April 5, 2025 (Online)

DOI: 10.37936/ecti-cit.2025192.260242

1. INTRODUCTION

Low-density polyethylene (LDPE) is a widely used thermoplastic polymer for producing various items, such as plastic bags, films, containers, and toys, due to its favorable properties, such as flexibility, chemical resistance, and ease of processing [2]. The performance of LDPE in these applications depends on a range of chemical and physical properties, such as melt flow index (MFI), density, tensile strength, and thermal stability. Accurately determining these properties is crucial for ensuring product quality, regulatory compliance, and customer satisfaction. However, the traditional approach to obtaining these properties through laboratory testing is time-consuming and costly, posing significant challenges for manufacturers seeking to optimize production efficiency [3].

The challenges of optimizing LDPE testing costs using machine learning techniques to predict LDPE properties arise from several interrelated factors. Firstly, the chemical properties of LDPE are highly interconnected, necessitating a comprehensive understanding of how different parameters influence one another. Secondly, high-quality, comprehensive real-world datasets encompassing a wide range of production conditions and resulting properties are often limited, hindering the development of reliable machine-learning models. Lastly, constraints on the accept-

[1,7,8]The authors are with the Faculty of Informatics, Burapha University, Thailand, Email: 65810010@go.buu.ac.th, rsunisa@buu.ac.th and nutthanon@buu.ac.th
[2]The author is with the Faculty of Optometry, Ramkhamhaeng University, Thailand, Email: akarapon.w@rumail.ru.ac.th
[3,4,5,6]The authors are with the Faculty of Engineering, Chulalongkorn University, Thailand, Email: 6332024921@alumni.chula.ac.th, 6632109921@student.chula.ac.th, 6532136721@student.chula.ac.th and 6532035021@student.chula.ac.th
[8]Corresponding author: nutthanon@buu.ac.th

able range of error that manufacturers can further complicate the implementation of predictive models for industries.

Previous attempts to estimate LDPE chemical properties have been constrained by the complexity of interconnected parameters and limitations in the data used for model construction. For instance, Zhong [4] employed regression models optimized with the Whale Optimization Algorithm to predict the mechanical properties of linear polyethylene; however, the reliance on a single input variable (oven residence time) may restrict the model's ability to capture more complex relationships influencing material behavior. Similarly, Shirazian *et al.* [5] used tree-based ensemble models optimized with the Firefly Algorithm to predict LLDPE mechanical properties. Still, their study was limited by a small dataset of only 25 samples. [6] depends on artificial neural networks that were used to predict diesel blends' cetane index and sulfur content. However, a small set of 35 samples were used to predict S50.

Existing hybrid prediction models have demonstrated effectiveness in various applications. For instance, Srikamdee *et al.* [22] proposed a collaborative learning framework for RNA secondary structure prediction, focusing on improving prediction accuracy. Similarly, Dey *et al.* [23] introduced a hybrid model for violence recognition in video streams, emphasizing real-time performance and accuracy. Jing *et al.* [24] presented a multi-granularity self-attention mechanism for few-shot learning, which enhances model adaptability. These works highlight the potential of hybrid approaches but underscore the gap in research focusing on cost-effective sequential prediction models, particularly in industrial contexts like LDPE property testing.

This paper introduces a hybrid approach that combines boosting algorithms [7] with a topological sorting algorithm [8] to address the challenges of cost-efficient sequential predictions in LDPE property testing. Our contributions are threefold:

1) We propose a hybrid method, TopSABoost, which integrates a topological sorting algorithm with boosting algorithms to serve as a cost-efficient sequential prediction model.
2) We construct and validate our model using real-world data from one of the largest chemical factories in Thailand, encompassing a diverse range of LDPE properties and characteristics.
3) The proposed technique effectively reduces the time and costs associated with laboratory testing of chemical properties while maintaining high prediction accuracy, ensuring compliance with manufacturing constraints.

By addressing these challenges, our work offers a viable alternative for the cost-effective determination of LDPE properties. This provides substantial economic and operational benefits for the chemical manufacturing industries and contributes to the broader adoption of machine learning techniques in chemical engineering and industrial processes.

2. BACKGROUND AND RELATED WORK

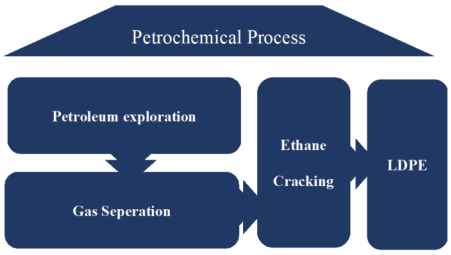

Fig. 1: *Overview of petrochemical process.*

The petrochemical industry encompasses a range of processes essential for transforming raw materials into valuable LDPE products, as shown in Fig. 1. Petroleum exploration is the initial phase, focusing on locating and extracting crude oil and natural gas reserves. Following extraction, gas separation techniques are employed to refine and isolate various hydrocarbons, including ethane, which serves as a key feedstock. Ethane cracking is a crucial step in this transformation, where ethane is converted into ethylene, a fundamental building block for numerous petrochemical products. One significant product derived from ethylene is low-density polyethylene (LDPE), which is widely used in packaging, plastic bags, and other applications due to its flexibility and durability.

2.1 Overview of Petrochemical Industries

The petrochemical industry plays a pivotal role in global commerce by facilitating the production of essential materials crucial to modern society. Beginning with rigorous exploration and extraction efforts using advanced technologies like seismic imaging and drilling, the industry identifies and extracts crude oil and natural gas, which undergoes meticulous gas separation processes such as distillation, absorption, and membrane techniques. These processes purify feedstocks and yield diverse hydrocarbons tailored for specific applications. Downstream, olefin cracking transforms heavier hydrocarbons into lighter olefins like ethylene, which is vital for synthesizing Low-Density Polyethylene (LDPE). LDPE's production involves controlled polymerization of ethylene molecules with catalysts, resulting in versatile polymers used in packaging, construction, healthcare, and agriculture. This journey exemplifies the synergy of science, technology, and industry expertise, essential for meeting global demand while navigating environmental sustainability challenges. Efforts toward sustainability include cleaner extraction technologies, process optimization, alternative feedstock

exploration, and advances in recycling, promoting a responsible and resilient petrochemical future.

2.2 Gradient Boosting Machine (GBM)

Gradient Boosting Machine (GBM) is a powerful ensemble learning technique that builds a model sequentially, each iteration aiming to correct the errors of the previous one. Introduced by Jerome Friedman in the late 1990s, GBM combines the predictive strength of multiple weak learners, typically decision trees, to create a robust model. The process begins with constructing an initial model, often a simple mean of the target variable. Subsequent models are then fitted on the residual errors of the initial model, effectively "boosting" its performance.

GBM's strength lies in its ability to optimize a differentiable loss function by iteratively adding models that minimize this loss. Each new model is trained to predict the residuals (errors) of the combined previous models, and the results are aggregated to form the final prediction. This method reduces bias and variance, making GBM particularly effective for complex datasets with nonlinear relationships.

In chemical property prediction [19], GBM has proven invaluable. It has been used to accurately predict properties such as solubility, boiling points, and reaction yields. Studies have shown that GBM can outperform traditional statistical methods due to its ability to capture intricate patterns and interactions within the data, providing a significant advantage in cheminformatics and materials science applications.

2.3 Extreme Gradient Boosting (XGBoost)

Adaptive Boosting (AdaBoost) is a prominent ensemble learning technique introduced by Freund and Schapire in 1995. Unlike other boosting methods, AdaBoost focuses on enhancing the performance of weak classifiers, typically decision stumps, by emphasizing the instances that are hardest to classify. The algorithm works iteratively, adjusting the weights of misclassified instances to ensure that subsequent classifiers focus more on these complex cases.

The process begins with all instances having equal weights. After each iteration, the weights of misclassified instances are increased, and a new weak classifier is trained on the reweighted data. This iterative process continues until a specified number of classifiers are created or the classification accuracy reaches a satisfactory level. The final model is a weighted majority vote of the individual classifiers, where each classifier's vote is weighted according to its accuracy.

AdaBoost has been effectively applied in various domains, including chemical property prediction. For example, it has been used to predict the flash points of organic compounds and the activity of molecular structures. Its ability to improve the accuracy of weak models makes it particularly useful in handling complex datasets where traditional methods might struggle.

In chemical property prediction [19], AdaBoost's strength lies in its capacity to adaptively enhance the focus on challenging prediction tasks, leading to more precise and reliable models. This adaptability and robustness make AdaBoost a valuable tool for advancing chemical property prediction and other related applications.

2.4 Adaptive Boosting (AdaBoost)

Adaptive Boosting, or AdaBoost, is a widely used ensemble learning technique that enhances the performance of weak classifiers to create a reliable model. Developed by Freund and Schapire in 1995, AdaBoost operates by iteratively adjusting the weights of training data based on the classification accuracy of previous models. Initially, all instances are assigned equal weights. After each iteration, the weights of misclassified instances are increased, prompting the next weak classifier to focus more on these challenging cases.

The core idea behind AdaBoost is to build a series of weak classifiers, such as decision stumps, each correcting the errors of its predecessors. This process continues for a predetermined number of rounds or until the model achieves the desired accuracy. The final output is a weighted majority vote of all the classifiers, where each classifier's vote is weighted by its performance.

AdaBoost's strength lies in its simplicity and effectiveness in improving classification accuracy. It is particularly effective for binary classification problems and has been successfully applied to various domains, including chemical property prediction. For instance, AdaBoost has been utilized to predict the flash points of organic compounds and to identify molecular activities, showcasing its ability to handle complex datasets.

In chemical predictions [3], AdaBoost's adaptive nature allows it to concentrate on difficult-to-classify chemical properties, thereby enhancing overall predictive performance. This makes it a valuable tool for researchers and professionals aiming to achieve precise and reliable predictions in chemical research and development.

2.5 Topological sorting

Topological sorting is a fundamental algorithmic technique to linearly order a directed acyclic graph (DAG) or a directed graph without cycles. The result of topological sorting is a linear ordering of the graph's vertices, such that for every directed edge (u, v) in the graph, vertex u comes before vertex v in the ordering.

The main application of topological sorting is scheduling tasks that have dependencies. For example, in project management, tasks must often be executed in a specific order based on their dependencies.

By performing topological sorting on a graph representing task dependencies, we can determine a valid order in which the tasks can be executed without violating any dependencies.

The algorithm for topological sorting typically involves systematically traversing the graph, such as depth-first search (DFS) or breadth-first search (BFS), while keeping track of visited vertices and their ordering. During the traversal, vertices are added to the ordering once all their incoming edges have been visited.

One crucial property of topological sorting is that it only applies to directed acyclic graphs (DAGs). If the graph contains cycles, there is no valid linear ordering of the vertices that satisfies all dependencies, and thus, topological sorting cannot be performed.

2.6 Sequential Machine Learning Models

Sequential prediction in machine learning involves predicting outcomes over a sequence of steps or periods, where each prediction relies on previous predictions or observed outcomes. This approach is fundamental in tasks where the order of predictions matters (such as forecasting stock prices and weather patterns) or in sequential decision-making processes. Models used for sequential prediction are designed to iteratively update their predictions based on new information, adjusting for dependencies and evolving patterns in the data. This iterative process allows the model to improve its accuracy over time by learning from past predictions and adapting to changing conditions. Applications of sequential prediction span various fields, including finance, natural language processing, and reinforcement learning, highlighting its significance in modeling complex, dynamic systems.

2.7 Related work

Machine learning has proven highly effective for predicting polymer characteristics by accommodating the nonlinear, multivariate relationships inherent in production data. Techniques such as artificial neural networks (ANNs), support vector regression (SVR), and ensemble methods like extreme gradient boosting (XGBoost) have demonstrated remarkable success. For example, [9] utilized ANNs to predict the mechanical and thermal properties of graphene-reinforced linear low-density polyethylene (LLDPE) composites, achieving high accuracy in tensile strength and thermal conductivity predictions. Similarly, [10] applied ensemble manifold learning to real-time melt index prediction in polyethylene production, demonstrating the adaptability of data-driven approaches to complex nonlinearities.

Boosting-based techniques have demonstrated notable effectiveness across diverse domains. Among these methods, prior research indicates that Adaboost and XGBoost are particularly promising for predictive tasks. For example, [11] showed the efficacy of the AdaBoost algorithm in reducing overfitting and enhancing the water temperature prediction. In [12], an ensemble of seven machine learning algorithms, one of which is XGBoost, was employed to forecast wheat yields. Moreover, XGBoost has been shown to excel in several applications, such as estimating surface water quality [13], forecasting retail sales [14], predicting wave heights [15], developing predictive maintenance models for wind turbines [16], interpreting factors affecting crude oil prices [17], and predicting software maintenance efforts [18]. In [19], GBM, XGBoost, and AdaBoost have been applied to predict only one LDPE characteristic, i.e., Vicat softening temperature.

Prior works on LDPE property prediction, such as [20] Wells and Ray (2005) utilizing a kinetic scheme to predict molecular weight distribution. More recently, [4] explored hyperparameter optimization using the Whale Optimization Algorithm (WOA) to enhance the predictive performance of regression models for LLDPE properties. [21] utilized regression, ensemble, distance-based, and regularized-based machine learning methods to predict polymer physical properties, including thermal, mechanical, and electrical characteristics.

Previous research works, just mentioned, have focused on predicting individual parameters at a time. While specific studies explored the prediction of multiple LDPE parameters, these approaches did not focus on sequential predictions nor minimizing the overall costs associated with predicting all parameters.

3. METHODOLOGY

The overview of our proposed hybrid method of topological sorting and boosting (TopSABoost) for cost-efficient sequential prediction is shown in Fig. 2. The proposed method consists of 4 main modules. The first module aims to determine the best predictive models among GBM, XGBoost, and AdaBoost for predicting each LDPE property, together with the models' top k important features. The second module depends on the models and the important features to construct a graph. The tail of a directed edge represents a feature utilized by one of the models to predict one of the target LDPE properties, which is represented by the associated arrowhead. The next module prunes the nodes that are not the target LDPE properties. The last module employs our proposed Minimum Cost Prediction Sequence algorithm (MCPS) to achieve cost-efficient sequential predictions: determining a sequence of predictions that maximizes the number of predicted properties while satisfying all factory requirements.

3.1 First-Level Prediction

This subsection explains the first module, called first-level prediction, as shown in Fig. 2 (a). In this

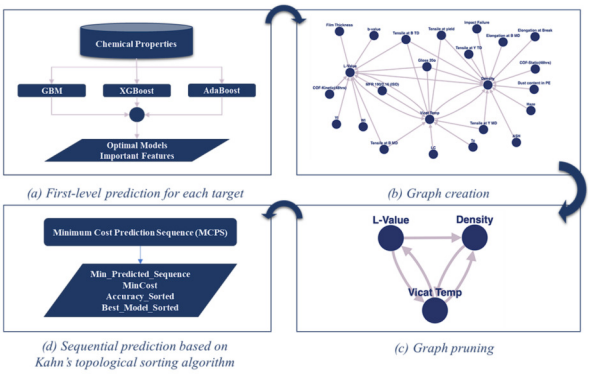

Fig.2: *Overview of TopSABoost.*

work, the LDPE properties a customer requires for laboratory testing are referred to as target properties $P = \{p_1, p_2, \ldots, p_t\}$, where t is the number of target properties. Let $attr_i, p_j$, and N denote the i^{th} attribute of data obtained from laboratory tests, a target LDPE property, and the number of attributes, respectively. In the first module, the $N-1$ attributes of data $\{attr_i | 1 \leq i \leq N$ and $attr_i \neq p_j\}$ serve as features for the M models (e.g., GBM, XGBoost, and AdaBoost) to predict each target property, p_j, where $1 \leq j \leq t$. The model with the best performance is selected for production, and its predicted property is incorporated as one of the features into the training data for the second-level prediction.

The relations between the k-most important features of the best-performing model (among GBM, XGBoost, and AdaBoost) and their associated predicted p_j are represented as a graph generated in the second module. Table 1 shows the example of the 10 most important features, where $P = \{$Lightness value (L-value), Density, Vicat Softening Temp$\}$, $t = 3$, and $k = 10$. p_j, one of the 10 most important features, is emphasized in bold and underlined.

Table 1: *The example of 10-most important features.*

Target LDPE Property	Best Model	10-most Important Features
L-value	AdaBoost	WI, Tensile at B, MD, YI, Gloss 20o, Tensile at B TD, COF-Kinetic(48hrs), MFR 190/2.16, Film Thickness, b-value, **Vicat Temp**
Density	AdaBoost	**Vicat Temp**, Gloss 20o, Tensile at Y TD, Tm, Tensile at Y MD, **L-value**, MFR 190/2.16, Impact Failure, Haze, Elongation at B MD
Vicat Temp	XGBoost	**Density**, Tensile at B MD, Gloss 20o, MFR 190/2.16, **L-value**, COF-Kinetic(48hrs), Tensile at yield, Impact Failure, COF-Static(48hrs), Tc

3.2 Graph Generation & Pruning

This subsection explains the process of generating a graph to represent the relationships among all the attributes and of pruning the graph structure to reduce running-time complexity.

The first module yields the k-most important features of the best model for predicting the value of each p_j. Such relations could be represented as a graph, as shown in Fig. 2(b). Our graph-based model effectively captures dependencies among target properties and their k-most important features, enabling the identification of the longest sequence of target property predictions and thereby minimizing laboratory testing costs. Specifically, predicted target properties can serve as attributes for prediction models to predict other target LDPE properties. For instance, in Fig. 2(b), Vicat Temp was among the 10 most important attributes for AdaBoost in predicting the values of L-value and Density. L-value was used by AdaBoost and XGBoost to predict density and Vicat values, respectively. XGBoost utilized density to estimate Vicat Temp values.

To focus solely on the target LDPE properties required by customers, we removed nodes that did not represent these target properties from the graph. The pruned graph, as exemplified in Fig. 2(c), reveals the direct dependencies and relationships among only the selected target properties, thereby reducing computational complexity and enhancing the efficiency of the analysis. It is important to note that the pruning process is solely intended to help determine the sequence of the target properties to be predicted. It does not exclude the removed properties from being used as input features for prediction.

Our graph-based algorithm, detailed in Subsection 3.6, effectively sorts the order of predicting the target properties to ensure they can be utilized in the most extended sequence possible while satisfying certain factor constraints.

3.3 Topological Sorting

Our proposed hybrid method leveraged Kahn's topological sorting algorithm [8] to determine the cost-effective prediction sequence. However, as shown in Fig. 2(c), the presence of cycles in the pruned graph made topological soring impossible. To address such an issue, we applied a brute-force approach and exhaustively evaluated all possible combinations of target-property prediction orderings. This technique allowed us to determine the longest cost-effective sequence possible while satisfying the factory's requirements.

3.4 Objective Functions & Constraints

We aim to minimize the costs associated with laboratory testing of chemical parameters while maintaining high prediction accuracy. To ensure that the pre-

diction outcomes of the models remain within practical thresholds, we consulted with expert chemists from the most prominent plastic manufacturing facility to obtain the error constraints. Specifically, the error in predicting Density, L-value, and Vicat softening temperature must not exceed 10%, 10%, and 2.9°C (approximately 2.9%), respectively. The key challenge of this research lies in determining the longest possible prediction sequence that meets these error constraints, as it would minimize the laboratory testing costs. The longer the prediction sequence with the acceptable errors, the greater the cost savings.

3.5 Sequential Prediction

This subsection introduces sequential predictions where a predicted LDPE property is used as a feature for the next model in the sequence to predict another LDPE property. Fig. 3 shows an n-level sequential prediction. Fig. 3 illustrates an n-level sequential prediction. All training data at the first level are obtained from laboratory tests and serve as features for the three models (i.e., GBM, XGBoost, and AdaBoost) to predict the first target LDPE property. The model with the best performance is selected for production, and its predicted property is incorporated as one of the features into the training data for the second-level prediction. This process is repeated to the n^{th} level, where n is the maximum level of prediction that could achieve the accuracy required by the factory.

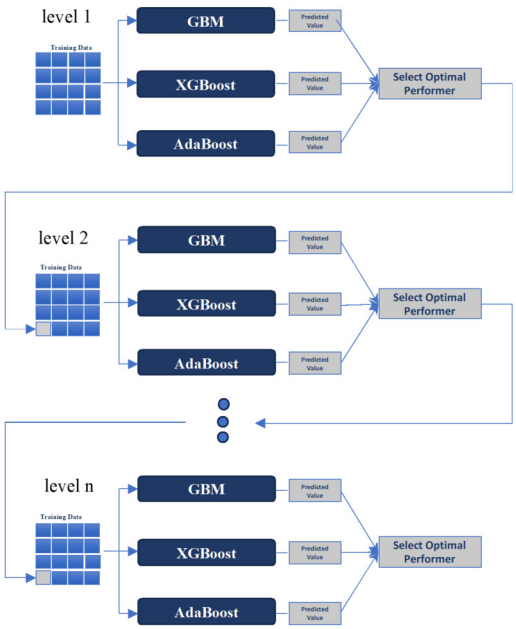

Fig.3: *Sequential Prediction.*

3.6 Algorithm & Running-Time Analysis

This subsection analyses the algorithm proposed in this research and evaluates its running-time complexity.

The proposed algorithm, TopSABoost, enhances the prediction of chemical parameters by combining the boosting method with a graph theory. The primary objective is to determine the most cost-effective forecasting sequence while ensuring compliance with the factory's accuracy requirements, which reduces laboratory testing costs.

The process begins by splitting the available dataset into training and testing datasets, following a standard 80-20 ratio. Once the data is split, the algorithm proceeds with the first level of prediction (Lines 3-12 of Algorithm 1). For each chemical parameter in the set of parameters to be predicted (Target_LDPE_Properties), the algorithm iterates over multiple methods, such as XGBoost, AdaBoost, and GBM. The method that provides the highest accuracy and satisfies the predefined accuracy constraints is selected as the best model for predicting that particular chemical parameter. In addition, the algorithm records the k-most important features used by the model, which are crucial for creating a graph subsequently.

After identifying the best models and the important features for each target chemical parameter, the algorithm constructs a directed graph G (Lines 15-18). In this graph, nodes represent features and target chemical parameters, while a directed edge indicates the dependency of target chemical parameters and the k-most important features. This graph serves as the foundation for determining the optimal sequence of predictions and minimizing the overall laboratory-testing cost.

To reduce the running-time complexity, we simplify the graph structure by pruning (Lines 21-23). Specifically, the algorithm removes any nodes that do not represent the target chemical parameters.

To find the optimal prediction sequence, the algorithm employs the Minimum Cost Prediction Sequence (MCPS) function (Lines 43-53). This function evaluates all possible combinations of nodes (representing target chemical parameters) within the graph, forming 2^t subgraphs, to guarantee the determination of the optimal prediction sequence. MCPS leverages the Kahn-based topological sorting method, CalculatePredSeq_Cost (Lines 60-93), to determine the dependency orderings and arrange the prediction sequence for each subgraph that meets the factory constraint requirements. CalculatePredSeq_Cost also calculates the cost of deleting vertices from the graph, corresponding to the cost of requiring lab testing for those associated parameters (Lines 89-91). Finally, after each subgraph is evaluated, the sequence that results in the lowest total cost and meets the accuracy constraints is selected as the optimal prediction sequence (Lines 48-52).

In certain circumstances where at least one cycle exists in the subgraph, the order of predicting the

target properties is unclear. CalculatePredSeq_Cost does not return the sequence that does not contain all nodes in the subgraph, which indicates that a cycle exists in the subgraph (Lines 88-93).

The running-time complexity of the algorithm is primarily dominated by the Minimum Cost Prediction Sequence (MCPS) function, which evaluates all possible combinations of target chemical parameters to determine the optimal prediction sequence, $O(2^t)$. The running-time complexity of the Kahn-based CalculatePredSeq_Cost is $O(|V| + |E|)$. This leads to a time complexity of $O(2^t(|V| + |E|))$, where $|V|$ the number of vertices and $|E|$ is the number of edges of the graph, G. The first level of prediction, which involves training models using various methods and testing their accuracy, has the time complexity of $O(t(B_1 + B_2 + \cdots + B_M))$, where t is the number of the target chemical parameters, and B_i is the running time of the i^{th} selected Boosting method. Graph construction and pruning operations add an additional complexity of $O(t|E|)$. While the overall complexity is exponential due to the combinatorial nature of MCPS, our algorithm remains efficient as it is designed for small-scale graphs and non-real-time applications. Optimization techniques that are more efficient for larger datasets still need to be investigated in the future.

4. EXPERIMENT RESULTS

4.1 Data Preparation

The dataset in this research is real-world data from one of the largest chemical factories in Thailand. The dataset comprises diverse properties and characteristics inherent to LDPE products. Specifically, LDPE products are to be sampled and tested in the laboratory to assess their properties prior to delivery, as these properties are essential to determine whether the products meet customers' requirements. To prepare the data, we conducted data transformation where the dataset's structure is modified to better suit analytical and modeling requirements as follows.

1. We queried raw data comprising 750,000 rows and 58 columns from the Lab Information Management System, which is sourced from Thailand's three major polymer manufacturing factories. The dataset covers 251 products, 71 testing methods, and 190 parameters.

2. Then, the columns deemed irrelevant to the analysis (e.g., test standard names, factory names, and test method versions) were removed. Of the 58 columns, 6 were selected for analysis: sample ID, sample status, sample type ID, dataset status, parameter ID, and value. Furthermore, duplicate records are systematically identified and eliminated to prevent redundancy, thereby enhancing the dataset's reliability and quality.

3. Each row of the previous step's data corresponded to a laboratory testing result of a single property per sample. We transformed the data further by ensuring that each row represents the laboratory testing results of all properties for each product sample. This transformation resulted in a dataset containing 4,500 rows and 71 columns.

4. In this step, we conducted the data cleansing by removing missing records.

Additionally, columns with a significant proportion of zero values, potentially including parameters Density-g/cm3, Density-Displacement, and Density-g/cm3-Displacement, are removed to reduce noise and optimize the dataset for accurate predictive modeling. The final dataset comprised 181 rows and 61 columns. Through label encoding, the categorical attributes, i.e., Deflect-detail and other-deflect, were encoded into 0, representing no defect, and 1, indicating the presence of defects.

4.2 Data Exploration Analysis

Table 2 describes the data. For instance, Density, ranging from 0.92 g/cm^3 to 0.925 g/cm^3 with an average of 0.9231 g/cm^3, highlights its mass per unit volume. Hardness, ranging from 43 to 50.7 with an average of 47.3449, indicates the material's resistance to indentation and scratching. Haze, ranging from

Algorithm 1: A hybrid method of topological sorting and boosting (TopSABoost)

Input:
Target LDPE Properties = $\{p_1, p_2, \ldots, p_t\}$, e.g., {'L-value', 'Density', 'Vicat'}
Methods = $\{m_1, m_2, \ldots, m_M\}$, e.g., {'XGBoost', 'AdaBoost', 'GBM'}
Accuracy Constraint = $\{C_1, C_2, \ldots, C_t\}$, e.g., {0.9, 0.9, .., 0.971}

Output:
Min_Prediction_Sequence, the ordered sequence $\{p'_1, p'_2, \ldots, p'_t\}$ of the target LDPE properties to be predicted
MinCost, the minimum cost of sequentially predicting all properties in Min_Prediction_Sequence
Accuracy_Sorted, the accuracy of sequentially predicting all properties in Min_Prediction_Sequence
Best_Model_Sorted, the sequence of the best models for predicting p'_i in Min_Prediction_Sequence

```
1  # First Level Prediction
2  {Chemical_Parameter_Training_Data, Chemical_Parameter_Testing_Data} = Split(Data, 80, 20)
3  foreach p in Target_LDPE_Properties:
4      max_accuracy[p] = 0
5      C = Accuracy constraint of p
6      foreach m in Method:
7          model = m.train(Chemical_Parameter_Training_Data)
8          accuracy = model.test(Chemical_Parameter_Testing_Data)
9          if accuracy > C and accuracy > max_accuracy[p]:
```

```
                max_accuracy[p] = accuracy
                bestModel[p] = model
                Import_Features[p] = model.getImportantFeature()

# Create a graph, G, based on maximum accuracies and important features
foreach p in Target_LDPE_Properties:
    if max_accuracy[p] != 0:
        foreach f in Import_features[p]:
            G.add_edge(f, p)  # Add the edge (f, p) to the graph

# Graph Pruning
foredge e in G:
    if src_node and dst_node of e are not in Target_LDPE_Properties:
        G.remove_edge(e)

# Update the accuracy of the node that could be sequentially predicted
Min_Prediction_Sequence = null
{Min_Prediction_Sequence, MinCost, Accuracy_Sorted, Best_Model_Sorted} = MCPS(G)
if Min_Prediction_Sequence != null:
    foreach p in Target_LDPE_Properties:
        if p in Min_Prediction_Sequence:
            accuracy[p] = Accuracy_Sorted[p]
            bestModel[p] = Best_Model_Sorted[p]
        else:
            accuracy[p] = max_accuracy[p]
else:
    # cannot determine any prediction sequences
    return -1

# Higher-Level Prediction
NodeSets = the combination of all nodes in G (i.e., the combination of all target LDPE properties)

# Function: Minimum Cost Prediction Sequence (MCPS) using binary representation
def MCPS(G):
    MinCost = infinite
    foreach NSet in NodeSets:
        SubG = a subgraph of G, which contains only nodes in NSet
        {Prediction_Sequence, Cost, tmp_accuracy, tmp_bestModel} = CalculatePredSeq_Cost(NSet, SubG)
        if Cost < MinCost:
            MinCost = Cost
            Min_Prediction_Sequence = Prediction_Sequence
            Accuracy_Sorted = tmp_accuracy
            Best_Model_Sorted = tmp_bestModel
    return {Min_Prediction_Sequence, MinCost, Accuracy_Sorted, Best_Model_Sorted}

# Function: Detemine the prediction sequence and the cost of deleting vertices
# (i.e., cost of parameter requiring lab testing)
# cost_del(i) = Cost of not including a vertex i in NSet (i.e., the laboratory-testing cost for p_i)
# which is the cost of parameter requiring lab testing
# deg(i) = In degree of vertex i (representing the i^th target LDPE property) in Nset
def CalculatePredSeq_Cost(NSet, G):
    Q = Queue
    PS = list()
    for v in NSet:
        if deg(v) == 0:
            Q.push(v)  # Push vertex v to the queue
    while not Q.empty():
        v = Q.front()  # Get the front element of the queue
        Q.pop()  # Remove the front element from the queue
        for t in NSet:
            if edge from v to t exists in G:
                deg(t) = deg(t)-1
                if deg(t) == 0 and t not in PS:
                    tmp_accuracy[t] = 0
                    C = Accuracy constraint of t
                    foreach m in Method:
                        model = Method.train(Chemical_Parameter_Training_Data)
                        accuracy = model.test(Chemical_Parameter_Testing_Data)
                        if accuracy > C and accuracy > tmp_accuracy[t]:
                            tmp_accuracy[t] = accuracy
                            tmp_bestModel[t] = model
                    if tmp_accuracy[t] == 0:
                        # The accuracy for predicting t does not meet the constraint
                        return {null, infinite, null, null}
```

```
85                          Q.push(t)  # Push the target property, t, to the queue
86                          PS.add(t)  # Add t to the list prediction sequence
87
88          if size(NSet) == len(PS):
89              foreach i in G but not in NSet
90                  total_lab_cost = cost_del(i)
91              return {PS, total_lab_cost, tmp_accuracy, tmp_bestModel}
92          else:
93              return {null, infinite, null, null}  # Return infinity as the graph is unsortable
```

5.69 to 9.133 with an average of 6.9611, indicates the degree of light scattering through the material, essential for achieving transparency and clarity in applications such as optical films and protective coatings. L-value, ranging from 81.71 to 84.66 with an average of 83.5399, reflects the lightness/darkness of the material, influencing its appearance in packaging, consumer goods, and architectural finishes. Melt flow rate (MFR 190/2.16), ranging from 0.197 to 4.505 g/10min with an average of 2.331 g/10min, assesses the flowability of products' melt. Vicat Temperature, ranging from 86.9°C to 96.333°C with an average of 93.5371°C, signifies the temperature at which products soften under load.

In summary, LDPE products exhibit diverse properties that collectively define their thermal stability, mechanical strength, optical clarity, processing characteristics, frictional behavior, chemical composition, and mechanical performance.

4.3 First-Level Prediction

The experiments are based on a 3.1GHz 6-Core Intel Core i5 processor, a Radeon Pro 575X 4GB graphics card, and 32GB of 2667 MHz DDR4 RAM. The hyperparameter settings are shown in Table 3.

The hyperparameter tuning process involved systematically testing different values for key parameters in the XGBoost, AdaBoost, and Gradient Boosting Machine (GBM) models to optimize their performance. The parameters included the number of boosting rounds (n_estimators), the maximum tree depth (max_depth), the minimum sum of instance weight needed in a child (min_child_weight), and the learning rate (learning_rate) for XGBoost, AdaBoost, and GBM as shown in Table 3.

Table 4 reports the optimal parameters. XGBoost, AdaBoost, and GBM were selected for predicting L-value, Density, and Vicat temperature using optimized hyperparameters. For L-value prediction, XGBoost achieved an MAE of 0.1576 and R^2 of 0.8588, AdaBoost showed an MAE of 0.1239 and R^2 of 0.9131, and GBM resulted in an MAE of 0.1589 and R^2 of 0.8594. For Density, XGBoost had an MAE of 0.00054 and R^2 of 0.7461, AdaBoost 0.00039 and R^2 of 0.7825, and GBM 0.00041 and R^2 of 0.8187. For Vicat temperature, XGBoost yielded an MAE of 0.3038 and R^2 of 0.9704, AdaBoost 0.3038 and R^2 of 0.9704, and GBM 0.3090 and R^2 of 0.9647.

The experimental results demonstrate that AdaBoost consistently outperforms XGBoost and GBM for most target properties. For predicting Vicat temperatures, XGBoost and AdaBoost perform equally well, yielding the same MAE and R^2, while GBM shows slightly inferior performance. AdaBoost's superior performance may stem from its iterative approach to minimizing errors by focusing on difficult-to-predict samples, allowing it to capture complex patterns in the data more effectively than other models.

According to the results of the first-level prediction, the 10 most important features for the selected models to predict the values of L-value, Density, and Vicat Temp are shown in Table 1.

4.4 Graph Generation & Graph Pruning

The dependencies of the predicted values and their important features (in Table 1) were used to generate a graph, as shown in Fig. 4. Then, we removed the nodes representing non-target LDPE properties, resulting in the pruned graph, as shown in Fig. 5.

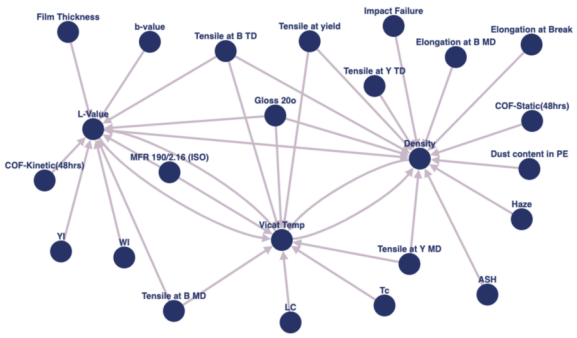

Fig.4: *Graph representing the relationship among target properties and important features.*

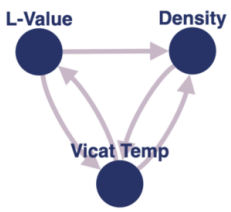

Fig.5: *Pruned Graph.*

Table 2: *LDPE property descriptions.*

PARAMETER NAME	MIN	MAX	AVERAGE	TYPE	PARAMETER NAME	MIN	MAX	AVERAGE	TYPE
%ASH	0.0450	0.1180	0.0759	Number	Haze	5.6900	9.1330	6.9611	Number
2 < PSD <4.0 mm	99.0200	100.0000	99.7492	Number	Impact Failure	96.0000	272.2500	131.7554	Number
a-value	-0.6300	-0.3100	-0.4559	Number	L-value	81.7100	84.6600	83.5399	Number
Appearance	88.0000	88.0000	88.0000	Number	LC	85.6900	88.0200	86.9904	Number
ASH	446.0960	1181.8000	759.7143	Number	MFR 190/2.16 (ISO)	0.1970	4.5050	2.3310	Number
b-value	-2.0800	-1.2500	-1.6389	Number	Other deflect	-	-	-	String
Blown film	-	-	-	String	PSD <2.0 mm	0.0000	0.1000	0.0061	Number
COF-Kinetic(48hrs)	0.0740	0.1420	0.1026	Number	PSD >4.0 mm	0.0000	0.9800	0.2447	Number
COF-Static(48hrs)	0.0820	0.1560	0.1115	Number	Secant Modulus	181.2000	269.2000	226.0138	Number
Deflect detail	-	-	-	String	Tc	91.3000	97.2500	94.6999	Number
Density	0.9200	0.9250	0.9231	Number	Tensile at B MD	16.8200	31.5000	22.2555	Number
Dust content in PE	0.0000	28.4860	5.6171	Number	Tensile at B TD	15.0000	32.4800	20.4504	Number
Elongation at B MD	347.0000	660.0000	508.3991	Number	Tensile at break	9.6000	21.0000	13.5531	Number
Elongation at B TD	565.4000	799.8000	708.6219	Number	Tensile at Y MD	9.5200	13.7000	11.4746	Number
Elongation at Break	452.0000	728.2000	556.8985	Number	Tensile at Y TD	9.4000	14.2800	11.3853	Number
Elongation at Y MD	11.0000	22.8000	16.3319	Number	Tensile at yield	9.0000	12.0000	10.7745	Number
Elongation at Y TD	8.8000	17.4000	13.5024	Number	Tm	108.8000	115.9100	112.3668	Number
Elongation at yield	13.0000	18.6000	15.5860	Number	Vicat Temp	86.9000	96.3330	93.5371	Number
Film Thickness	48.0000	70.3330	51.9485	Number	WI	68.5200	74.5600	71.4746	Number
Gloss 20o	36.9670	104.3330	81.8883	Number	YI	-5.1300	-2.9900	-3.9631	Number
Hardness	43.0000	50.7000	47.3449	Number					

Table 3: *Summary of hyperparameters.*

Hyperparameter	XGBoost	AdaBoost	GBM
n_estimators	50, 100, 150, 200, 250	50, 100, 150, 200, 250	50, 100, 150, 200, 250
max_depth	4, 5, 6, 7	N/A	4, 5, 6, 7
min_child_weight	1.1, 1.2, 1.3, 1.4, 1.5	N/A	N/A
learning_rate	0.51, 0.52, 0.53, 0.54, 0.55	0.51, 0.52, 0.53, 0.54, 0.55	0.51, 0.52, 0.53, 0.54, 0.55
loss	N/A	linear, square, exponential	N/A

4.5 Sequential Prediction

A. Minimum Cost Prediction Sequence (MCPS)

Minimum Cost Prediction Sequence (MCPS), which identifies that the most cost-effective sequence is {L-value → Density}, offers valuable insights into optimizing the prediction of Low-Density Polyethylene (LDPE) properties. By adopting this approach, businesses can deliver precise, high-quality results to customers in a more cost-effective manner, reducing both resource consumption and overall expenses without compromising the reliability and accuracy of the predictions.

The predictive sequence {L-value → Density} suggests that the L-value, which is associated with flow or viscosity properties, serves as a crucial precursor for accurately determining density. Predicting the L-value first may simplify or streamline the subsequent prediction of density, as these properties are inherently related. This approach demonstrates that the L-value provides valuable intermediate information, enhancing both the accuracy and efficiency of density predictions while minimizing the need for additional complex computations. In contrast, Vicat prediction was not included in the sequence. This might be because it caused errors more than the expected threshold.

B. Prediction Density and use Predicted L-value

Table 5 reports the optimal parameters for the 2-level model candidates. XGBoost, AdaBoost, GBM were configured with a learning rate of {0.53, 0.55, 0.55}, a maximum depth of {5, N/A, 7}, a min child weight of {1.1, N/A, N/A}, a loss function of {N/A, exponential, N/A}, and the number of estimators of {50, 200, 50}.

The above configuration resulted in a mean absolute error (MAE) of 5.44×10^{-4}, 3.61×10^{-4}, 3.42×10^{-4}, a mean squared error (MSE) of {$5.17\times^{-7}$, 4.17×10^{-7}, 2.9×10^{-7}}, a root mean squared error (RMSE) of {7.19×10^{-4}, 6.45×10^{-4}, 5.4×10^{-4}}. R^2 was {0.75, 0.8, 0.86}. The maximum error observed was {0.002, 0.002, 0.001} with a percentage maximum error of {0.22%, 0.22%, 0.11%}.

The AdaBoost model showed slightly better performance than XGBoost in terms of lower error metrics and a higher R^2 value, likely due to its iterative reweighting mechanism that prioritizes correcting high-error predictions, enhancing its ability to capture complex patterns. The GBM model outperformed both XGBoost and AdaBoost in terms of MAE, MSE, RMSE, and maximum error, as its ability to capture feature interactions allowed for superior predictive accuracy and reliability in density prediction.

5. DISCUSSION

A. Impact on accuracy if the predicted values obtained from the previous model are used

In Section 4, the L-value attribute in the training dataset for the sequential prediction model was obtained from laboratory testing. This section discussed the effect on accuracy if we instead used the predicted values obtained from the previous model in the training dataset.

Table 6 reports the optimal parameters for the 2-level model candidates. XGBoost, AdaBoost, GBM were configured with a learning rate of {0.53, 0.55, 0.53}, a maximum depth of {5, N/A, 6}, a min child weight of {1.1, N/A, N/A}, a loss function of {N/A, exponential, N/A}, and the number of estimators of {50, 200, 50}.

The above configuration resulted in a mean absolute error (MAE) of {5.44×10^{-4}, 3.89×10^{-4}, 4.06×10^{-4}}, a mean squared error (MSE) of {5.17×10^{-7}, 4.44×10^{-7}, 3.27×10^{-7}}, a root mean squared error (RMSE) of $7.19\text{x}10^{-4}$, 6.67×10^{-4}, 5.72×10^{-4}. R^2

was {0.75, 0.78, 0.84}. The maximum error observed was {0.002, 0.002, 0.001} with a percentage maximum error of {0.22%, 0.22%, 0.16%}.

Similar to the results in Section 4.5, the results show that GBM is superior in terms of predictive accuracy and reliability for density prediction. However, all measures confirm that using the laboratory-testing values in the training dataset yields better prediction performance than using the predicted ones. This might be because laboratory-testing values are direct, empirical measurements, which inherently contain less noise and uncertainty than predicted values. Predicted values may carry accumulated errors or approximations from earlier steps in the prediction process, which can reduce the overall accuracy when used as input in subsequent predictions. The precise nature of laboratory measurements thus provides a more reliable foundation for model-in-the-sequence training, leading to improved predictive performance.

B. Limitation

TopSABoost determines the topologically sorted sequence of all graph combinations to resolve cycles and identify the optimized prediction sequence. While this exhaustive approach ensures accuracy, its computational complexity increases exponentially with the number of LDPE properties and target parameters. It is well-suited for non-real-time prediction tasks, as the computation is performed only once, and the resulting sequence can be reused until new properties or parameters are required to be tested. Nevertheless, its scalability makes it impractical for large-scale or real-time applications. Future work could explore more efficient alternatives, such as heuristic-based optimization methods, dynamic programming, or machine learning-driven approaches for sequential prediction. Additionally, parallel computing or distributed processing techniques could be leveraged to reduce runtime and enable scalability. These advancements would enhance the algorithm's applicability to larger datasets and more complex industrial settings.

6. CONCLUSION

This paper introduces TopSABoost, a hybrid method that combines topological sorting with boosting techniques. The method comprises four modules: first-level prediction, graph generation, graph pruning, and sequential prediction. Its primary objective is to determine the cost-effective prediction sequence while adhering to the accuracy constraints of the manufacturing process, all while minimizing laboratory testing costs. Using an LDPE property (predicted by a former model) as an input for the next predictive model in the sequence effectively reduces reliance on costly laboratory tests. Experimental results demonstrated that the proposed method successfully identifies cost-effective predictive sequences and achieved a maximum error of 0.11%. Future work will focus on refining the algorithm by reducing the running time complexity, enabling its application to larger datasets and more complex scenarios, thereby broadening its utility in various manufacturing contexts.

ACKNOWLEDGEMENT

We express our gratitude to the team of expert chemists from the most prominent plastic manufacturing facility for providing invaluable data on constraints used in this research (i.e., the maximum prediction errors) in predicting various LDPE properties, including Density, L-value, and Vicat softening temperature, as well as for their insightful review of the analysis of the experimental results. We also would like to express our genuine gratitude to Mr. Sean Wanitchollakit, Mr. Natdanai Hirunsirikul, Mr. Nattanan Jeungtanasirikul, and Mr. Peraga Puangtong for reviewing the algorithm.

AUTHOR CONTRIBUTIONS

Conceptualization, N.P. and N.L.; methodology, N.P., S.R. and N.L.; software, N.P.; validation, N.P., A.W., N.W., B.D., P.W., C.K., S.R. and N.L.; formal analysis, N.P., A.W., N.W., B.D., P.W., C.K., S.R. and N.L.; investigation, N.P., S.R. and N.L.; data curation, N.P.; writing—original draft preparation, N.P.; writing—review and editing, N.P., S.R. and N.L.; visualization, N.P.; supervision, N.L. All authors have read and agreed to the published version of the manuscript.

References

[1] *Standard test method for Vicat softening temperature of plastics*, ASTM D1525-17E01, ASTM International, 2017.

[2] British Plastics Federation. "Polyethylene (Low Density) LDPE, LLDPE," bpf.co.uk. Accessed: Dec. 12, 2024. [Online.] Available: `https://www.bpf.co.uk/plastipedia/polymers/LDPE.aspx`

[3] D. Ledwani, I. Thakur and V. Bhatnagar, "Comparative Analysis of Prediction Models for Melt Flow Rate of C2 and C3 Polymers Synthesized using Nanocatalysts," *NanoWorld Journal*, vol. 08, no. S1, 2022.

[4] F. Zhong, "A theoretical approach based on machine learning for estimation of physical properties of LLDPE in moulding process," *Scientific Reports*, vol. 14, no. 1, pp. 1–13, Sep. 2024.

[5] S. Shirazian, T. Huynh, S. M. Sarkar and M. H. Zare, "Development and optimization of machine learning models for estimation of mechanical properties of linear low-density

polyethylene," *Polymer Testing*, vol. 137, pp. 108525–108525, Jul. 2024.

[6] F. M. de Oliveira *et al.*, "Predicting Cetane Index, Flash Point, and Content Sulfur of Diesel–Biodiesel Blend Using an Artificial Neural Network Model," *Energy & fuels*, vol. 31, no. 4, pp. 3913–3920, Mar. 2017.

[7] R. E. Schapire, "The strength of weak learnability," *Machine learning*, vol. 5, no. 2, pp. 197–227, 1990.

[8] A. B. Kahn, "Topological sorting of large networks," *Communications of the ACM*, vol. 5, no. 11, pp. 558–562, Nov. 1962.

[9] P. N. Khanam *et al.*, "Optimization and Prediction of Mechanical and Thermal Properties of Graphene/LLDPE Nanocomposites by Using Artificial Neural Networks," *International Journal of Polymer Science*, vol. 2016, pp. 1–15, 2016.

[10] Y. Liu, Y. Liang and Z. Gao, "Industrial polyethylene melt index prediction using ensemble manifold learning–based local model," *Journal of Applied Polymer Science*, vol. 134, no. 29, Mar. 2017.

[11] A. Taufiqurrahman, A. G. Putrada and F. Dawani, "Decision Tree Regression with AdaBoost Ensemble Learning for Water Temperature Forecasting in Aquaponic Ecosystem," *2020 6th International Conference on Interactive Digital Media (ICIDM)*, Bandung, Indonesia, pp. 1–5, 2020.

[12] J. Han *et al.*, "Prediction of winter wheat yield based on multi-source data and machine learning in China," *Remote Sensing*, vol. 12, no. 2, pp. 236, 2020.

[13] D. N. Khoi, N. T. Quan, D. Q. Linh, P. T. T. Nhi, and N. T. D. Thuy, "Using Machine Learning Models for Predicting the Water Quality Index in the La Buong River, Vietnam," *Water*, vol. 14, no. 10, pp. 1552, 2022.

[14] M. Baržić, N. -F. Munitić, F. Bronić, L. Jelić and V. Lešić, "Forecasting Sales in Retail with XGBoost and Iterated Multi-step Ahead Method," *2022 International Conference on Smart Systems and Technologies (SST)*, Osijek, Croatia, pp. 153–158, 2022.

[15] F. Anggraeni, D. Adytia and A. W. Ramadhan, "Forecasting of Wave Height Time Series Using AdaBoost and XGBoost, Case Study in Pangandaran, Indonesia," *2021 International Conference on Data Science and Its Applications (ICoDSA)*, Bandung, Indonesia, pp. 97–101, 2021.

[16] W. Udo and Y. Muhammad, "Data-Driven Predictive Maintenance of Wind Turbine Based on SCADA Data," in *IEEE Access*, vol. 9, pp. 162370-162388, 2021.

[17] M. Gumus and M. S. Kiran, "Crude oil price forecasting using XGBoost," *2017 International Conference on Computer Science and Engineering (UBMK)*, Antalya, Turkey, pp. 1100–1103, 2017.

[18] M. JayaBharath, N. L. Choudary, C. S. Pranay, M. D. Praveenya and B. R. Reddy, "An analysis of Software Maintainability Prediction Using Ensemble Learning Algorithms," *2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP)*, VIJAYAWADA, India, pp. 1–5, 2023.

[19] N. Phongthakun, S. Rimcharoen and N. Leelathakul, "Predictive Modeling of Vicat Softening Point for Low-Density Polyethylene Using GBM, XGBoost and AdaBoost: A Comparative Analysis," *Lecture notes in networks and systems*, pp. 23–32, May. 2024.

[20] G. J. Wells and W. H. Ray, "Prediction of Polymer Properties in LDPE Reactors," *Macromolecular Materials and Engineering*, vol. 290, no. 4, pp. 319–346, Apr. 2005.

[21] I. P. Malashin, V. S. Tynchenko, V. A. Nelyub, A. S. Borodulin and A. P. Gantimurov, "Estimation and Prediction of the Polymers' Physical Characteristics Using the Machine Learning Models," *Polymers*, vol. 16, no. 1, pp. 115–115, Dec. 2023.

[22] S. Srikamdee and P. Chongstitvatana, "Collaborative Learning of Estimation of Distribution Algorithm for RNA secondary structure prediction," *ECTI-CIT Transactions*, vol. 14, no. 1, pp. 92–102, Apr. 2020.

[23] A. Dey, S. Biswas, and L. Abualigah, "Efficient Violence Recognition in Video Streams using ResDLCNN-GRU Attention Network," *ECTI-CIT Transactions*, vol. 18, no. 3, pp. 329–341, Jul. 2024.

[24] W. Jing and C. S. Lim, "Multi-granularity self-attention mechanisms for few-shot Learning," *ECTI-CIT Transactions*, vol. 18, no. 4, pp. 522–530, Oct. 2024.

Table 4: *Summary of predicting evaluated value.*

XGBOOST										
Prediction Parameter	**learning_rate**	**max_depth**	**min_child_weight**	**n_estimators**	**MAE**	**MSE**	**RMSE**	**R2**	**MAX_ERROR**	**%_MAX_ERROR**
L-value	**0.51**	4	1.1	50	0.15755	0.04033	0.20082	0.85880	0.59406	0.71056
Density	**0.53**	5	1.1	50	0.00054	5.2E-07	0.00072	0.74607	0.00204	0.22101
Vicat	**0.51**	7	1.1	50	0.31893	0.16246	0.40306	0.96377	0.84879	0.90835
MFR	**0.515**	6	1.1	200	0.17775	0.09597	0.30979	0.94752	1.25402	52.3698
ADABOOST										
Prediction Parameter	**learning_rate**	**max_depth**	**loss**	**n_estimators**	**MAE**	**MSE**	**RMSE**	**R2**	**MAX_ERROR**	**%_MAX_ERROR**
L-value	**0.52**	N/A	square	250	0.1239	0.0248	0.1575	0.9131	0.3300	0.3947
Density	**0.55**	N/A	exponential	200	0.0004	4.4E-07	0.0007	0.7825	0.0020	0.2167
Vicat	**0.53**	N/A	exponential	200	0.3038	0.1326	0.3641	0.9704	0.7000	0.7491
MFR	**0.51**	N/A	linear	250	0.0523	0.0062	0.0786	0.9966	0.2310	9.6469
GBM										
Prediction Parameter	**learning_rate**	**max_depth**	**min_child_weight**	**n_estimators**	**MAE**	**MSE**	**RMSE**	**R2**	**MAX_ERROR**	**%_MAX_ERROR**
L-value	**0.51**	4	N/A	250	0.15886	0.04017	0.20043	0.85935	0.53399	0.63871
Density	**0.55**	7	N/A	200	0.00041	3.7E-07	0.00061	0.81868	0.00152	0.16446
Vicat	**0.53**	4	N/A	50	0.30904	0.15810	0.39761	0.96474	0.91464	0.97882
MFR	**0.54**	6	N/A	100	0.14716	0.12778	0.35747	0.93012	2.00247	83.6258

Table 5: *Density Prediction Errors (The train dataset contained the L-value obtained from Laboratory tests).*

Algorithm	Prediction Parameter	learning_rate	max_depth	min_child_weight	loss	n_estimators	MAE	MSE	RMSE	R2	MAX_ERROR	%_MAX_ERROR
XGBOOST	Density	0.53	5	1.1	N/A	50	5.44E-04	5.17E-07	7.19E-04	7.47E-01	2.04E-03	2.21E-01
ADABOOST	Density	0.55	N/A	N/A	exponential	200	3.61E-04	4.17E-07	6.45E-04	7.96E-01	2.00E-03	2.17E-01
GBM	Density	0.55	7	N/A	N/A	50	3.42E-04	2.90E-07	5.39E-04	8.58E-01	1.02E-03	1.11E-01

Table 6: *Density Prediction Errors (The train dataset contained the L-value predicted by the selected model).*

Algorithm	Prediction Parameter	learning_rate	max_depth	min_child_weight	loss	n_estimators	MAE	MSE	RMSE	R2	MAX_ERROR	%_MAX_ERROR
XGBOOST	Density	0.53	5	1.1	N/A	50	5.44E-04	5.17E-07	7.19E-04	7.47E-01	2.04E-03	2.21E-01
ADABOOST	Density	0.55	N/A	N/A	exponential	200	3.89E-04	4.44E-07	6.67E-04	7.82E-01	2.00E-03	2.17E-01
GBM	Density	0.53	6	N/A	N/A	50	4.06E-04	3.27E-07	5.72E-04	8.40E-01	1.47E-03	1.59E-01

Noparat Phongthakun received a B.Sc. degree in Computer Technology from King Mongkut's University of Technology North Bangkok, Thailand, in 2004, and a Master's degree in Business Administration from Dhonburi Rajabhat University, Thailand, in 2010.

Borworntat Dendumrongkul is currently pursuing a B.Eng. degree in computer engineering at Chulalongkorn University, Bangkok, Thailand. His academic interests include software development, artificial intelligence, and computer graphics. His current research focuses on machine learning, computer vision, and real-time computer graphics.

Akarapon Watcharapalakorn received a Doctor of Optometry (O.D.) degree with first-class honors and a gold medal in optometry from Ramkhamhaeng University, Thailand, in 2015. Akarapon has been a faculty member of the Faculty of Optometry, Ramkhamhaeng University, Bangkok, Thailand, since 2017. His current research interests include optometry, myopia, dry eye, blue light filtering lenses, artificial intelligence, machine learning and deep learning.

Pana Wanitchollakit is currently pursuing a Bachelor of Engineering (B.Eng.) degree in computer engineering at Chulalongkorn University. His academic interests include machine learning and deep learning. He has experience working with various machine learning frameworks and neural network architectures.

None Wanitchollakit received a B.Eng. degree with first-class honors in computer engineering from Chulalongkorn University, Bangkok, Thailand, in 2024. Since 2024, None has been a data analyst at TISCO Financial Group PCL. His current research interests include machine learning, association rule mining, and data analytics.

Chayanin Kongsareekul is currently pursuing a B.Eng. degree in computer engineering at Chulalongkorn University, Bangkok, Thailand. His academic interests include data structures and algorithms, with his current research focusing on these areas.

Sunisa Rimcharoen received a B.Sc. degree in computer science from Burapha University, Thailand, in 2003, an M.Sc. degree in computer science from Chulalongkorn University in 2005, and a Ph.D. degree in computer engineering from Chulalongkorn University in 2009. Dr. Sunisa has been a faculty member of the Faculty of Informatics, Burapha University, Chon Buri, Thailand, since 2009. Her current research interests include evolutionary computation, machine learning, prediction, and algorithmic trading.

Nutthanon Leelathakul received a B.Eng. degree with first-class honors in electrical engineering from Kasetsart University, Bangkok, Thailand, in 1999, an M.Eng. degree in electrical and computer engineering from Cornell University in 2004, and a Ph.D. degree in electrical and computer engineering from Carnegie Mellon University in 2010. Dr. Nutthanon has been a faculty member of the Faculty of Informatics, Burapha University, Chon Buri, Thailand, since 2010. His current research interests include computer networking, security in IoT, deep learning, and data analytics.