ORIGINAL ARTICLE

TEKSTİL VE KONFEKSİYON

VOL: 36, NO. 1
DOI: 10.32710/tekstilvekonfeksiyon.1489112



Multi-Output Regression in the Textile Industry: Time, Waste and Price Prediction

Ceren ATIK[1] 0000-0002-0843-7026
Recep Alp KUT[2] 0000-0002-5781-334X

[1] Dokuz Eylul University, Graduate School of Natural and Applied Sciences, Izmir, Türkiye
[2] Dokuz Eylul University, Department of Computer Engineering, Izmir, Türkiye

Corresponding Author: Ceren Atik, ceren.atik@ogr.deu.edu.tr

ABSTRACT

Machine learning and data mining techniques provide businesses with cutting-edge data-driven decision-making capabilities. Their popularity is growing because they enable more accurate and consistent evaluation and prediction of current and future situations based on previous data. This study used machine learning methodologies to address three of the textile industry's most pressing concerns: lead time, cloth waste, and price. A multi-output regressor model in which three subjects are predicted simultaneously is also investigated, in addition to training individual models for each subject. XGBoost is the model with the best lead time prediction results, with an R^2 of 0.86 and an MAE of 8.35. When all three subjects are predicted at the same time, XGBoost achieves R^2 of 0.88 and MAE of 3.79. These findings indicate that, in addition to single models, multi-output models are also promising.

ARTICLE HISTORY
Received: 23.05.2024
Accepted: 08.12.2025

KEYWORDS

Machine learning, ensemble learning, multi-output regression, textile, apparel

1. INTRODUCTION

Customer happiness and providing excellent products and services are the top priorities in the textile industry, as they are in any other industry. Planning and resource management are crucial to meeting these goals. With the advancement of technology and its incorporation into many aspects of our lives, it has contributed a new dimension to attaining these goals, generating products, and assuring sustainability. When research in the field of textiles is reviewed, it is discovered that issues such as quality management [1-3], performance improvement [4,5], and demand forecasting [6,7] are mostly covered. In addition to this, there is research [8-10] that may be considered as decision support systems.

This study utilizes machine learning technologies to enhance process management and enable more precise decision-making. These strategies offer company-specific solutions to challenges. In terms of process management, it enables them to assess their sub-steps in terms of production time, production quantity planning, and product unit price regarding previous orders. Each of the subjects covered in this study is a focus of the sector's continuous efforts to improve. At the planning stage, an accurate estimate of production time is critical for the most effective utilization of capacity and labor. Production amount is one of the elements that influences production time. When determining the production amount, situations apart from the order quantity must be considered. Losses that may occur throughout the production process must be considered to provide the order requested by the client in the required quality and quantity. Printing and embroidery, for example, increase the error rate. As a result, mistakes and losses that may occur in all processes must be considered. When determining production quantity, the textile industry often initiates cutting operations by adding a standard rate to the order amount. Since this ratio is unaffected by additional characteristics like color, fabric type, or sub-processes, it results in under- or overproduction. This study tackles this challenge by estimating an accurate cloth waste rate based on past orders. In addition to the time and quantity of production, the unit price of the generated product is one of the issues that must be resolved. This study also addresses the topic of predicting the unit price of a textile product based on its most fundamental attributes. Since the models used in the

To cite this article: Atik C, Kut R A. 2026. Multi-output regression in the textile ındustry: time, waste and price prediction. *Tekstil ve Konfeksiyon,* 36 (1), 60-70.

study will help determine the relevant values for the three subjects during the production planning stage, they may be evaluated under decision support systems. This study not only used a single model to predict each subject, but also tested multi-output regressor models.

We can describe the aim of the study as follows:

- Comparison of individual performance of ensemble learning models in predicting three subjects.
- Comparison of the results of multi-output models with binary combinations of three subjects and versions in which all three are predicted simultaneously.
- To select the model that will assist in making more precise decisions by considering a variety of order-dependent fabric and process factors.
- Predicting lead time, cloth waste, and price all together using real manufacturing data.
- Creating data-based solutions for planning production with faster and more accurate results.
- Preventing either overproduction or underproduction to guarantee effective resource usage.

We structure the rest of the article as follows: The related work section covers studies in the subject of textiles. The material and method section describes the data collection, models, and assessment measures that were employed. The experimental results section examined the findings obtained and their interpretation. The conclusion section offered an overall assessment of the study.

2. RELATED WORK

Various research on quality, profitability, and process management is being conducted in the textile industry. From the beginning of manufacturing until the end user, numerous concerns may be addressed, such as fabric quality, product production time, production faults, profitability, employee and customer profile analysis. Despite the fact that they are regarded as diverse areas, the goal of the study is the same: improvement.

To prevent stock overruns or shortages, it is essential to predict product demand for the upcoming time by considering market trends, client orders, sales history, and other relevant factors. The initial phase of production planning involves forecasting future demands through demand analysis. Demand forecasting is complex due to internal factors specific to the textile industry and external influences such as macroeconomic issues. This study examines the importance of both internal and external factors in demand modeling and forecasting [11]. Demand estimates were made using different machine learning and deep learning models, and the significance of various factors in predicting demand was analyzed using generalized least squares and single-layer perceptron models on time series data. In another study [12], Google Trends was utilized for demand forecasting with additional data to enhance time series data for training machine learning models. It was determined that machine learning models yielded superior outcomes compared to traditional statistical methods. The SVR and Ridge models demonstrated optimal performance in the problem-specific context, characterized by the minimal error rate.

Demand forecasting is crucial for production planning, inventory management, and supply chain optimization. In this context, demand forecasting and supply chain management are essential elements that collaborate to enhance efficiency and customer satisfaction in the textile industry. Production processes encompass interdependent factors. Efficient supply chain management is essential as it directly influences production duration. This study [13] evaluates lead time based on the supply chain and seeks to address the issue using linear programming instead of artificial intelligence technology. Their findings indicate that forecasting production time can reduce it by fifty percent. They contend that the delay will also affect production costs. Another study [14] evaluates production time utilizing artificial intelligence technologies with conventional methods. The Smart Supply Chain Big Data Analysis collection comprises structured and unstructured data from supply, production, sales, commercial distribution activities, and additional product categories. The expected production time for this dataset utilizes metaheuristic and deep learning methodologies. The Eagle Strategy Slime Mold Algorithm (ESSMA) clustering is compared to other deep learning models like General Regression Neural Network (GRNN), Convolutional Neural Networks (CNN), and Generative Adversarial Networks (GAN). Among the models shown, ESSMA-GRNN is the best fit for the problem, showing the least amount of error (MSE of 0.19) and the highest success rate (93% explained variance). Another study [15] addressing the same issue utilized ensemble learning algorithms to predict the production time of textile products. Utilizing the bagging approach, they achieved optimal performance of 85%. In an additional study [16], the manufacturer of dyed knitted fabric from yarn uses Artificial Neural Networks (ANN) to ascertain the standard delivery time. To achieve this objective, the manufacturing processes are delineated, focusing on those that influence production duration. The manufacturing of textile fabric and the allocation of orders to clients are recognized as critical processes among the four primary steps. The cumulative durations for each procedure are aggregated, indicating the total manufacturing time. This study [17] aimed to predict delays in textile manufacture, asserting that such delays directly affect production time. The decision tree model for regression was developed using data from multiple sources to predict delay time. The trained model attains 99% accuracy.

A study [18] aimed to predict when warp breaks, weft breaks, and yarn bursts happen, using real data from IoT sensors and a textile company's production system database, because manufacturing errors can affect production time. The researchers examined the effectiveness of four distinct AutoML algorithms for predicting one target and three targets simultaneously. The models exhibited superior performance in predicting specific targets utilizing the H2O

CC BY NC ND

tool. Examining the same subject from another angle, ascertaining the optimal production quantity is a factor that directly affects production duration, and another study [19] also explored this issue. Machine learning models have been developed to aid in calculating the production quantity by incorporating an additional rate to the order volume and factoring in anticipated losses for each order. The bagging model yielded the highest results, achieving an 86% success rate. This study [20] employs multilayer perceptron (MLP) and linear regression models to estimate production time, arguing that the performance of the machinery and workers participating in the processes must also be taken into account. The MLP model predicts the production time of textile goods with an accuracy exceeding 85%, including factors such as sewing machine type, worker skills, material type (e.g., fabric), and product design specifications.

Quality control is a significant focus of textile research. Error detection is essential in fabric production, the fundamental materials utilized in textile manufacturing and its subprocesses. The automation of quality control is essential for meeting client requirements and preventing time and financial losses associated with quality issues. This study [21] attempts to quantify yarn quality utilizing three metrics obtained from textile machines. The collected data were classified into normal and anomalous categories, and machine learning models were executed for each classification. They used the ensemble regressor chain model to lower prediction errors because they found that unusual cases were harder to identify. This study [22] observed that fabric classification research predominantly relies on fabric images and that incorporating fundamental fabric properties is essential to enhance the success rate. The CNN model was employed to analyze fabric characteristics and perform recognition. They assert that more successful outcomes will be achieved in many practical applications, such as garment identification, by employing models trained with a combination of essential traits and photographs. A recent study [23] introduces a new method called Optimal Artificial Neural Network-based Fabric Defect Detection and Classification (OANN-FDDC), which includes steps like preparing the data, extracting important features by hand, using artificial neural networks, and fine-tuning the classification with root mean square propagation. The methodology was assessed utilizing the Fabric Detection dataset and the ZJU-Leaper dataset, yielding accuracies of 96.97% and 94.78%, respectively.

The consideration of all these issues during production necessitates handling profitability as an additional factor. This study [24] indicates that sellers assess the tactile qualities of textile products during in-person sales, yet no metric exists to ascertain product softness in online transactions, which may influence sales outcomes based on customer preferences. They created a dataset that included the weaving characteristics of the fabric along with mechanical factors for this objective. The samples were categorized into three thickness classifications (light, medium, and thick), and the factors affecting the softness value of each classification were examined. The regression model they employed achieved success rates of 83.2% for light materials, 88.7% for medium materials, and 84.1% for thick materials.

Effective pricing is essential, as achieving profitability through competitive pricing and minimizing inventory are significant issues. Pricing seasonal or pre-owned products, such as determining the final price through an assessment of costs and profit margins, requires a distinct analysis. This study [25] aims to predict the pricing of seasonal retail Christmas merchandise. Data from an online retail platform is examined using statistical and machine learning methods. To make a profit, activities are implemented to enhance sales or retain current clients.

This study [26] utilizes description information from the Sefamerve.com website to evaluate the price of apparel and home textile products. They compiled these detailed explanations into a single document and used it to estimate the cost. Deep learning approaches based on recurrent neural network (RNN) and CNN were compared for performance. Their results show that the long short-term memory-based model outperforms the CNN model. Profiting from selling at the right price and liquidating stocks are important considerations [27]. In this context, the researchers investigated the appropriate price for buying and selling used clothing in a flea market environment. It is critical that users' expectations are met based on the various features of clothing and that sellers profit from their sales by melting their stocks. Each product's price is determined by the seller. Since similar products have varying price ranges, they developed a model that can accurately predict the sales price in order to meet customer demand. The random forest and multiple linear regression models were used, with the random forest model achieving the highest accuracy rate.

As can be seen, attempts have been undertaken over the years to provide answers to various problems utilizing the machine learning technique. In this study, models based on production procedures were developed to assist the firm in making more precise decisions based on past data. There has been no study of a decision support system that simultaneously assesses the lead time, cloth waste, and price of textile items, as in this study. Research was conducted during the production planning phase to predict the cloth waste rate and the lead time to create products with the proper amount and quality. At the same time, an attempt was made to estimate the price of one product to assist in deciding the price information, which is another key issue of the planning process. We used the firm's data to model these challenges and reviewed the outcomes.

3. MATERIAL AND METHOD

This study focuses on estimating time, cloth waste, and price subjects independently and together using machine learning models, as well as evaluating the performance of these models. Data was originally obtained for this purpose via the company's Canias ERP system. This section

CC BY NC ND

discusses the dataset's attributes, the models utilized, and the assessment measures.

3.1 Data Collection

The procedure for producing a textile product is as follows: Upon receipt of an order, the manufacture of textile products commences with design and planning. This phase encompasses capacity planning, delivery time estimation, process definition, production flow management, and scheduling. Upon creation of the plan, the supply process commences, encompassing inventory management and material procurement.

A work order initiates the production process. As illustrated in Figure 1, knitted dyed fabrics are subjected to quality control during the manufacturing process. Faultless fabrics are set out, and cutting procedures are carried out in accordance with the order's model design. Cut works are forwarded to sewing after undergoing sub-processes such as printing, embroidery, or washing, if required in the order. After completion of the sewing process, the finished products undergo further quality control measures before packaging and delivery to the consumer.

To guarantee successful production, each phase must be executed correctly. Inadequately managed procedures, erroneous judgments, and defective products extend the production timeline, misallocate resources and expenses, and increase the environmental impact of manufacturing processes.

Consultations with qualified employees of the company were conducted in order to determine the features of the dataset, as high-quality and well-arranged data strongly influence model success. The production data gathered from the relevant report screens in the Canias ERP system was combined using ETL processes. In particular, basic features of the products included in the orders (such as model, color, and customer), process information (printing, embroidery, etc.), and parameters that need to be evaluated specific to the targeted purpose were collected. These reports exported from the screens in the ERP system were combined specific to the relevant orders, and a new dataset was created.

Table 1. Sample Order from Dataset

Attributes	Values
Gender	Woman
Product Type	Dress
Customer ID	3262
Country	SE
Color ID	3914
Season	Spring-Summer
Ordered	2629
Sewing Center	In-house
Knitted Fabric Type	Single Jersey
Embroidery	0
Print	1
Washing Type	None
Fabric Blend Ratio (%)	95% Organic Cotton 5% Elastane
Fabric Finishing	None
Cutting Etude Time (Minutes)	1.5
Sewing Etude Time (Minutes)	15.5
Packing Etude Time (Minutes)	3.5
Broidery Etude Time (Minutes)	0
Print Etude Time (Minutes)	3.5
Difficulty Level of Model	2
Actual Lead Time (Days)	241
Actual Cloth Waste Rate	2.5%
Unit Price	12.25 EUR

From the collected features, we generated and added the following new features to the dataset:

- The actual production time was computed as the difference in days between the order placement date and the shipping date and then included in the dataset.
- The waste rate was determined by subtracting the quantity of cut pieces from the order quantity, dividing the result by the order quantity, and including it in the dataset.
- If the sewing location information was processed internally, it was designated as "in-house" for the relevant orders; conversely, if it was executed by a subcontractor, it was labeled as "external company" for the relevant orders and included in the dataset.

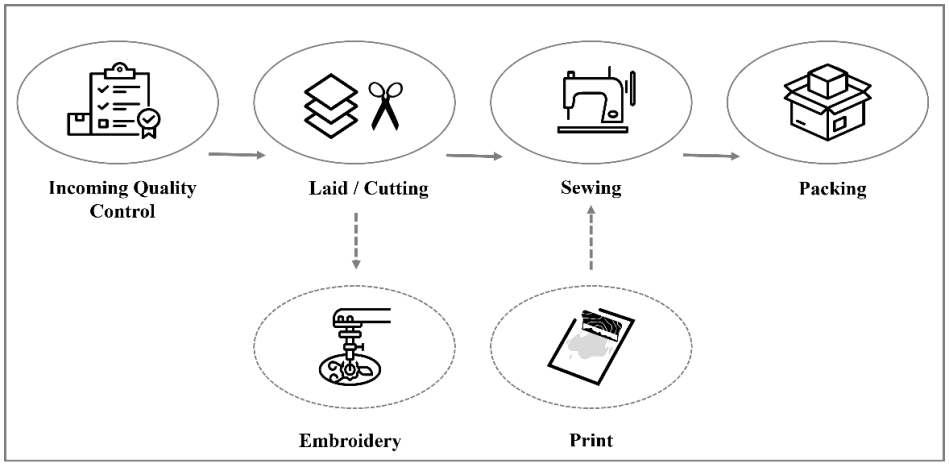


Figure 1. Production Process Flow

CC BY NC ND

Ultimately, values with incorrect and incomplete information were eliminated from the dataset and utilized to predict the three headings jointly that are the focus of the study. The final dataset of the study encompasses 749 orders purchased and delivered as of September 2022. Table 1 shows an example from the obtained dataset.

3.2 Models

This research discusses cases in which lead time, cloth waste, and price are calculated separately and together. Ensemble learning models were utilized in the situation where each title was modeled one by one. We utilized a multi-output regressor model for the simultaneous prediction of all titles.

Ensemble learning is a method in which many models are trained concurrently or sequentially with subsets of the dataset, and the results are merged [28]. It is still widely used because it produces better outcomes than individual models. Ensemble learning consists of three basic methods: bagging, boosting, and stacking. This study selected models based on bagging and boosting.

The bagging method involves training identical (homogeneous) basic learner models in parallel, independently of one another, then combining their outputs by averaging them [29]. Thus, the desired outcome is achieved. This study employs the bagging approach using Random Forest and Bagging models. When using the boosting approach, the initial weight of each observation is equal. The tree model is the base learner. The weightings are changed in such a way that the weight of misclassified observations rises, and the weight of uncommon, misclassified observations falls as the tree community evolves [30]. This study makes use of the XGBoost model, which applies the boosting strategy.

Multi-output regression is a label solution that provides many predictions at the same time. Multiple output regression employs two approaches: algorithm adaptation and problem transformation. Problem transformation techniques break a multi-output problem into smaller single-output problems, which are then solved using a single-output regression algorithm, while algorithm adaptation methods change a single-output approach to work with multi-output datasets [31].

Figure 2 displays the general flow diagram for the multi-output models used in this research. Following gathering, the dataset underwent preprocessing and was split into training and test sets. The training set was used to train multi-output regression models, and the outcomes of the models trained on the test set were assessed.

Tree-based models were selected for the study because they are easy to understand, can identify complex relationships between features and targets, do not depend on the scale of the features, can work with numerical and categorical data, and are strong against outliers [32]. Bagging, Random Forest, and XGBoost models all employ decision tree as their base learners; hence, these ensemble learning models were included in the study. Another rationale for using these tree-based models is that they are naturally suited to the multi-output regressor structure.

3.2.1 Decision Tree

Decision trees [33] are hierarchical structures made up of nodes and edges that, through the examination of their attribute values, recursively divide data (records) into different groups. Among the supervised learning methods, decision trees are suitable for tasks involving both regression and classification. A variation of the decision tree classifier, called decision tree regression, is used to solve the regression problem in this study because the target values that need to be predicted are continuous. The pseudocode of the decision tree regression algorithm is found in Algorithm 1.

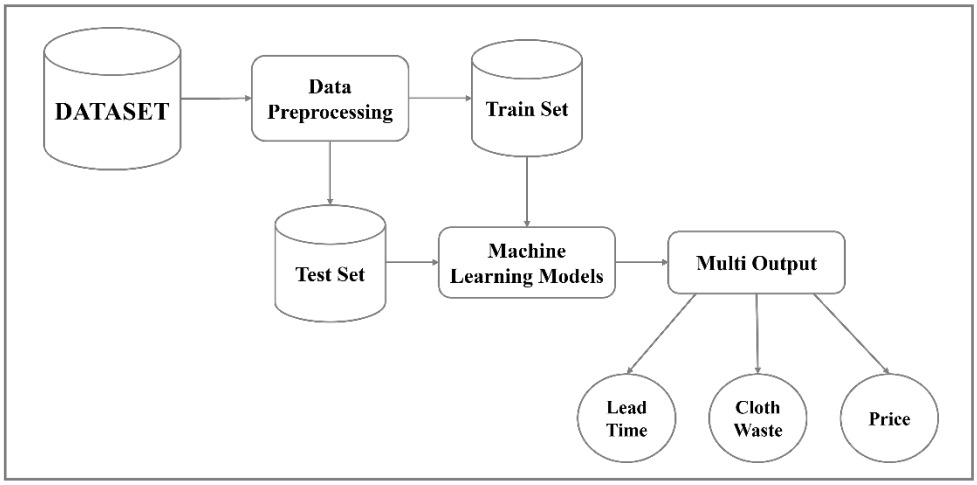


Figure 2. Schema for Multi-Output Regression

CC BY NC ND

Regression trees make use of *Mean Squared Error (MSE)*, which measures the average of the squares of the errors in Equation (1) and choose the feature and split value that leads to the minimum MSE at each node.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_{actual} - y_{predicted})^2 \quad (1)$$

Assume that dataset D has n instances and that $D=\{(x_1,y_1), .. ,(x_n,y_n)\}$ with $F = \{f_1,f_2,f_3,..,f_m\}$ the set of features. Regression tree construction is an iterative process that divides the data into divisions over all features F. The structure of the tree is initially determined using all of the training data. Next, the algorithm runs over every pair of successive feature values. At each iteration, the algorithm divides the data into two partitions and chooses the split that minimizes the total MSE in each partition. It also calculates average target y_{right} for right partitions and y_{left} for left partitions. The tree is obtained by repeatedly performing each of these steps.

3.2.2 Bagging

An ensemble meta-estimator Bagging [29] fits base learner B individually on arbitrary subsets D_j of the original dataset D, then averages between the individual predictions in ensemble E to provide a final prediction y. The bagging pseudo-code can be found in Algorithm 2. The decision tree is the default base learner used by this meta-estimator.

3.2.3 Random Forest

Random Forest, as its name implies, is a set of several decision tree algorithms that use random sampling. The bagging process often trains the collection of decision trees that form the forest [34]. The algorithm improves the model's randomness as the trees grow. When splitting a node, the best feature from a set of randomly picked features is chosen rather than the most important. Algorithm 3 provides the random forest pseudo-code.

Assuming the dataset $D = \{(x_1,y_1), .. ,(x_n,y_n)\}$ has $F = \{f_1,f_2,f_3,..,f_m\}$ feature set with m features, the method randomly selects k features with $k < m$. Next, by selecting a node with the lowest MSE among the k features, the algorithm will determine which node is the root. Subsequently, the algorithm divides the node into descendant nodes and goes through this procedure j times. In this way, a forest with j trees develops. Each tree is built from various sets of samples, and each node's split is based on a different set of attributes. Each tree makes its own predictions. A final prediction y from the model is an average of the predictions produced by the trees T in the forest.

Algorithm 1. Decision Tree

Input:
Dataset, $D = \{(x_1,y_1), .. ,(x_n,y_n)\}$ with features $F = \{f_1,f_2,f_3,..,f_m\}$
Output:
Decision Tree, DecisionTree
Begin
for f = 1 **to** F:
Define best parameters; feature, split value, weighted MSE, target y_{right}, target y_{left}
for n=1 **to** N:
Compute the average of two consecutive rows for split value
Split D into two partitions
Compute the weighted MSE of both partitions
Calculate average target y_{right} for right partitions and y_{left} left partitions
Update and store best parameters based on minimum weighted MSE
end for
DecisionTree = Grow tree recursively left and right subtrees with best parameters.
End

Algorithm 2. Bagging (Boostrap Aggregation)

Input:
Dataset, $D = \{(x_1,y_1), .. ,(x_n,y_n)\}$ with features $F = \{f_1,f_2,f_3,..,f_m\}$
Output:
Predicted Value, y
Begin
for j = 1 **to** J:
Draw a boostrap sample D_j of size n, randomly with replacement from D
Train a base learner B_j using D_j, and add it to the ensemble E.
end for
$y = \frac{1}{J}\sum_1^J E_j(x)$
End

Algorithm 3. Random Forest

CC BY NC ND

Input:
Dataset, $D = \{(x_1,y_1), .. ,(x_n,y_n)\}$ with features $F = \{f_1,f_2,f_3,..,f_m\}$
Output:
Predicted Value, y
Begin
for j = 1 **to** J:
Construct a boostrap sample D_j with the size *n* from *D*
Create a Random Forest Tree T_j with Dj until the minimum node size is reached by following the recursively;
Choose randomly *k* variables from the *F*
Select the best split among the *k*
Split the node into two descendant nodes
end for
$y = \frac{1}{J} \sum_1^J T_j(x)$
End

3.2.4 XGBoost

Gradient boosting is a unique approach in which each estimator corrects the inaccuracy of its previous. XGBoost is an implementation of gradient boosting methods. It generates decision trees consecutively and weights all non-dependent factors. The weight of the variables that the tree mistakenly predicted is increased, and they are subsequently added to the secondary decision tree [35].

As demonstrated in Algorithm 4 [36], the base learner *b* is defined initially, and the base model always takes the average by default. Subsequently, the tree is generated, starting with the root node. To grow the tree, one seeks for the node with the best division. The algorithm looks for the best split with the maximum gain specified in Equation (2) for all features *F*.

$$\mathbf{Gain} = \frac{\mathbf{1}}{\mathbf{2}} \left[\frac{\mathbf{G_{Left}^2}}{\mathbf{H_{Left}}} + \frac{\mathbf{G_{Right}^2}}{\mathbf{H_{Right}}} - \frac{\mathbf{G^2}}{\mathbf{H}}\right] \quad (2)$$

G_j and H_j for leaf *j* with sample *n* are computed as in Equation (3);

$$\mathbf{G_j} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{g_{ji}(x)}, \mathbf{H_j} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{h_{ji}(x)} \quad (3)$$

where g_{ji} first order partial derivative and h_{ji} is second order partial derivative of instance *i* in leaf *j*. Then, the optimal leaf weights for a fixed structure are then computed as $w^* = -\frac{G}{H}$. After identifying the gain and weight for the base learner *b(x)* the additive model is iteratively updated with boosting as $f_j(x) = f_{j-1}(x) + b(x)$. The final model is the sum of base learners, as shown in Equation (4):

$$\boldsymbol{f(x)} = \sum_{\boldsymbol{j=0}}^{\boldsymbol{J}} \boldsymbol{f_j(x)} \quad (4)$$

for j= 1, 2, ..., J where J is the number of the base learners. These base learners then combine to produce a strong model.

3.2 Evaluation Metrics

R^2, as calculated in Equation (5), is a popular statistic that evaluates the percentage of variance in the dependent variable that can be explained through the independent variable thus assesses the performance of regression models.

$$\boldsymbol{R^2} = \mathbf{1} - \frac{\sum(\boldsymbol{y_{actual}} - \boldsymbol{y_{predicted}})^2}{\sum(\boldsymbol{y_{actual}} - \boldsymbol{y_{mean}})^2} \quad (5)$$

Algorithm 4. XGBoost
Input:
Dataset, $D = \{(x_1,y_1), .. ,(x_n,y_n)\}$ with features $F = \{f_1,f_2,f_3,..,f_m\}$
Output:
Predicted Value, y
Begin
Initialize $f_0(x)$
for j = 1 **to** J:
Calculate G_j and H_j
Construct the structure by selecting splits that maximize gain.
Compute the weights w* of leafs
Initialize the b(x)
Add them $f_j(x) = f_{j-1}(x) + b(x)$
end for
$y = f_J(x)$
End

CC BY NC ND

In Equation (6), *Mean Absolute Error Rate (MAE)* is calculated by dividing the sum of all absolute errors between the measured value and the true value by the sample size.

$$MAE = \frac{1}{n}\sum |y_{actual} - y_{predicted}| \quad (6)$$

K-fold Cross Validation (CV) [37] is a resampling strategy that divides a given dataset into k groups and then evaluates machine learning models across these groups.

As shown in Figure 3, the dataset's randomly shuffled samples are split into *k* groups. Iteratively, *k-1* group is used as training data and the remaining group is used as testing data.

The model outputs are averaged to determine the final evaluation as in Equation (7).

$$CV = \frac{1}{k}\sum_{i=1}^{k} R_k \quad (7)$$

Ten has been selected as the study's value of *k*. In this study, the metrics indicated as R_k that are included in the evaluation at the end of every iteration are determined as R^2 and MAE.

4. EXPERIMENTAL RESULTS

On the basis of 2023 seasonal data acquired from the textile firm, research on lead time, cloth waste, and price prediction was conducted. To that end, situations in which each subject was modeled independently and those in which all subjects were predicted using a single model were compared.

To compare the success of the models, R^2 and MAE values acquired by 10-fold cross-validation were utilized. The dataset is split into 90% training and 10% test. The models were initially run with the default parameters. When the results in Table 2 are examined, it is clear that the XGBoost model has the highest R^2 value and the lowest MAE value for cloth waste and lead time prediction. In the price prediction problem, the Random Forest model achieves the highest R^2 value and the lowest MAE value.

Table 2. Performance of Models for Each Task

Models	Cloth Waste		Lead Time		Price	
	R^2	MAE	R^2	MAE	R^2	MAE
Decision Tree	0.77	2.21	0.59	9.40	0.80	0.94
Random Forest	0.77	2.80	0.85	8.77	**0.91**	**0.86**
Bagging	0.76	2.95	0.83	9.26	0.90	0.87
XGBoost	**0.87**	**2.21**	**0.86**	**8.35**	0.89	0.88

While all the models used are tree-based, their methods for responding to the problem differ. Despite both the Random Forest and Bagging models being based on the bagging approach, the Random Forest model's randomization in the feature set distinguishes it from the Bagging model, resulting in a greater success rate for each task. When XGBoost creates a series of models, each new model tries to fix the mistakes of the previous one, which makes the XGBoost model better than the other tree-based models used in this study for predicting cloth waste and lead time. Figure 4 also provides a visual representation of the R^2 values for the models used in each task.

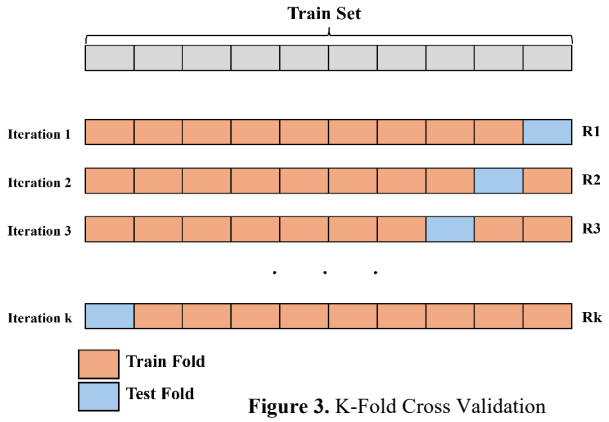


Figure 3. K-Fold Cross Validation

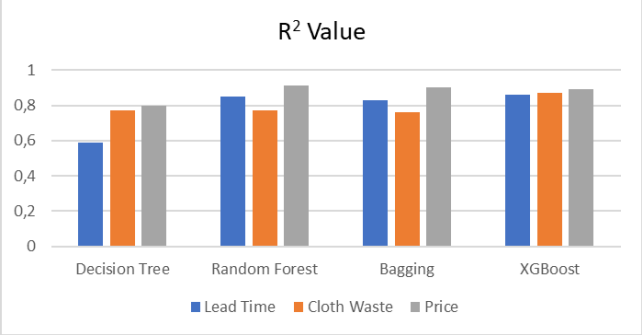


Figure 4. Comparison of R^2 Values of Models for Each Task

To provide a solution to the company's needs, the results of models predicting binary combinations of subjects are compared alongside predictions made using individual models for each subject. Finally, we investigated the effectiveness of models that simultaneously estimated three subjects. We trained the developed models using the default parameters. Table 3 summarizes the findings of multi-output regressor models for predicting cloth waste and pricing.

Table 3. Performance of Multi-Output Regression Models for Cloth Waste and Price Prediction

Models	R^2	MAE
Multi-output Regressor with Decision Tree	0.78	1.58
Multi-output Regressor with Random Forest	0.84	1.83
Multi-output Regressor with Bagging	0.83	1.91
Multi-output Regressor with XGBoost	**0.89**	**1.50**

We first analyzed models that assessed cloth waste and price values jointly, as these were the two issues with the highest success values in individual models. According to the results, the multi-output regressor with XGBoost performed the best, with an R^2 of 0.89 and an MAE of 1.50. Ensemble learning is a methodical strategy for combining the predictive capacity of several learners. The Random Forest model outperforms the Decision Tree model because it consists of multiple decision trees. Again, the randomness of Random Forest leads to stronger predictive ability than bagging.

CC BY NC ND

Additionally, since individual models make more mistakes when predicting lead time compared to other topics, the results of the combined model that merges lead time with cloth waste and price were also examined. Table 4 displays the performance of the models that predict both lead time and price simultaneously. It appears to have a lower error rate and a greater success rate than models that estimate lead time individually.

Table 4. Performance of Multi-Output Regression Models for Lead Time and Price Prediction

Models	R^2	MAE
Multi-output Regressor with Decision Tree	0.70	5.17
Multi-output Regressor with Random Forest	0.88	4.82
Multi-output Regressor with Bagging	0.87	5.07
Multi-output Regressor with XGBoost	**0.88**	**4.62**

When lead time and price prediction are attempted to be estimated at the same time, ensemble learning approaches outperform the single decision tree model in terms of success and error rates. Despite the bagging strategy shared by random forest and bagging models, random forest performs better in this situation as well due to the randomness impact. XGBoost performs better in this scenario because it has the lowest error rate (4.62 of MAE), even though its R^2 value (0.88) is the same as that of Random Forest.

Table 5 presents the results of models that estimate cloth waste and lead time together. The error rate in the findings shown here is also lower than the error rate in the models in which the lead time is predicted individually. In the case where cloth waste and lead time are predicted together, decision tree has 0.68 of R^2 and 5.81 of MAE values, compared to 0.59 of R^2 and 9.40 of MAE in lead time estimation. The XGBoost model, which has the best lead time estimation performance (0.86 of R^2 and 8.35 of MAE), has a lower MAE value of 5.23, although it has the same R^2 value in this case when lead time and cloth waste are predicted together.

Table 5. Performance of Multi-Output Regression Models for Cloth Waste and Lead Time Prediction

Models	R^2	MAE
Multi-output Regressor with Decision Tree	0.68	5.81
Multi-output Regressor with Random Forest	0.81	5.79
Multi-output Regressor with Bagging	0.80	6.10
Multi-output Regressor with XGBoost	**0.86**	**5.23**

On the basis of cloth waste and lead time prediction problems, ensemble learning models outperform the individual decision tree model, the random forest model outperforms bagging model due to its randomness, and the XGBoost model outperforms all others due to its boosting approach. The outcomes of the models trained on binary combinations of the three main subjects are displayed together in Figure 5. It is observed that when time estimation, which has the largest error rate of any individual model, is combined with cost and waste, the error rate decreases.

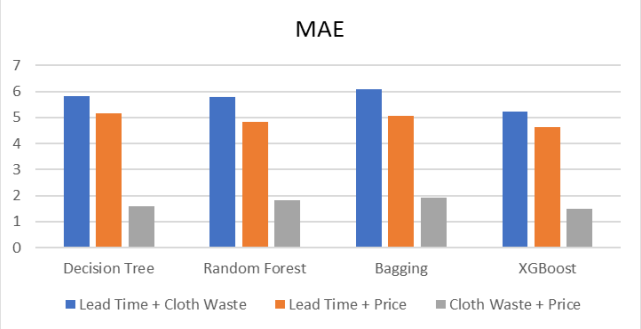


Figure 5. Comparison of MAE of Multi-Output Regression Models with Binary Combination of Tree Tasks

We created multi-output regressor models with default parameters, which predict three subjects simultaneously. Table 6 displays the results of estimating three subjects at the same time. In the simultaneous prediction of three subjects, the multi-output regressor with the XGBoost model has the highest R^2 of 0.88 and the lowest MAE of 3.79. Multi-output models with three subjects calculated together have the lowest error rate of all multi-output models that contain lead time.

Table 6. Performance Of Multi-Output Regression Models for Lead Time, Cloth Waste and Price Prediction

Models	R^2	MAE
Multi-output Regressor with Decision Tree	0.72	4.18
Multi-output Regressor with Random Forest	0.84	4.14
Multi-output Regressor with Bagging	0.83	4.36
Multi-output Regressor with XGBoost	**0.88**	**3.79**

When trying to predict combinations of three subjects, ensemble learning models work better than single decision tree models. Random Forest does better than the Bagging model in success and error rates. Thanks to the boosting technique, XGBoost achieves the highest success rate and the lowest error rate, outperforming both bagging models and individual decision tree models. Random forest outperforms the bagging model in terms of success and error rates. On this issue foundation, XGBoost has the highest success rate and lowest error rate due to the boosting technique, surpassing bagging and individual decision tree models. Figure 6 shows the separate models and models that predicted all three subjects together.

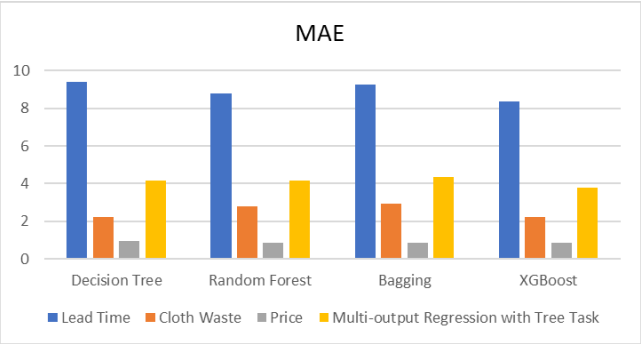


Figure 6. Comparison of MAE of Single Task and Multi-Output Regression Models with Tree Task

CC BY NC ND

Individual models for each problem can be employed, or multi-output regressor models might be selected because there are two major benefits of using multi-output regression trees versus creating a separate regression tree for each subject. A multiple-output regression tree is often substantially smaller than the overall size of separate single-task trees for all variables, and it recognizes relationships between multiple target variables more effectively.

For this reason, models that predict multiple subjects simultaneously can also be used, even though the error rates of the models for individual tasks are lower. Models with acceptable error rates that are regarded as successful for the firm can be employed, and separate models can continue to be used for the other subjects, especially when the outcomes of the models in which two subjects are predicted are analyzed.

Estimating lead time helps enhance planning and time management. Effective capacity planning can be achieved through cloth waste estimation during the determination of production quantities, minimizing cost and resource waste by averting stock excesses or shortages. Price estimation enables the creation of more precise and competitive pricing proposals. The predictive model developed from previous production data can estimate the lead time, cloth waste, and price associated with a new order. Production managers utilize a real-time decision support system to base their decisions on data, enabling immediate production improvement.

The production process can be overseen by promptly presenting predictions from machine learning models using dashboards. Furthermore, model outputs can be included in the ERP system. This study enables the organization to optimize essential indicators, including manufacturing time, production amount, and pricing estimation, thereby enhancing efficiency, lowering costs, and strengthening decision support systems.

5. CONCLUSION AND FUTURE WORK

Considering cost, quality, and time are the three most essential problems in the textile sector, the correctness of decisions made on these issues in terms of cost and welding process is critical. In the light of past data, more accurate and consistent outcomes may be produced when the basic aspects of a product and the procedures it will go through are taken into account. Machine learning models have been employed as a solution to this problem since they allow for the evaluation of additional factors at the decision stage and the generation of several outputs at the same time.

This study used machine learning algorithms to predict three major themes in the textile sector. The results show that not only do the ensemble learning models perform well on their own regarding price, time, and waste, but the multi-output regressor model also successfully predicts three topics. It has been demonstrated that more accurate order-specific results can be obtained by taking into account more attributes rather than adhering to fixed ratios, and machine learning models can be utilized to do this.

The utilization of a dataset derived from the company's daily operations enhanced the validity of the machine learning model for practical applications. However, although this situation brings with it some generalizability limitations, it offers a remarkable starting point due to the significant advantages and potential opportunities it contains. It has the potential to provide a solid basis for advanced analysis and comparative research involving different companies.

Successful modeling outcomes derived from a limited dataset demonstrate the scalability of machine learning methodologies and their potential for integration with real-world data; hence, they assist in the development of sector-wide roadmaps.

With an expanded dataset, future research can explore the following areas:

- Comparison of the results of a multitask learning study utilizing deep learning models with traditional machine learning models
- Classifying production processes can enhance production optimization by categorizing items according to their manufacturing techniques.
- Creation of dynamic decision models that assess the impacts of various scenarios through simulations of price and production quantities.
- Insights into product creation for new strategies, especially regarding the efficiency and pricing of various product types, are available for investigation.

Acknowledgement

This study has been supported by the project numbered 119C078 carried out within the scope of the TUBITAK 2244 Industry Doctorate Program. The authors would like to thank TUBITAK and TYH Textile for their assistance and support.

REFERENCES

1. Lee C-Y, Lin J. 2018. Improve quality and efficiency of textile process using data-driven machine learning in industry 4.0. *International Journal of Technology and Engineering Studies* 4(2), 64-76.
2. Hui Han P, rong Li X, Liu R, Zhang S, Yuan C. 2024. Prediction method of carding process production quality based on digital twin technology. *Textile Research Journal* 94(5-6), 713-724.
3. Su Z, Lu Y, Wu J, Zhang H, Li P. 2023. FocusDet: towards high-quality digital printing fabric defect detection. *Textile Research Journal* 93(23-24), 5237–5248.
4. Sikka MP, Sarkar A, Garg S. 2024. Artificial intelligence (AI) in textile industry operational modernization. *Research Journal of Textile and Apparel* 28(1), 67-83.
5. Wang Z, Wang J, Zeng X, Sharma S, Xing Y, Xu S, Liu L. 2021. Prediction of garment fit level in 3D virtual environment based on artificial neural networks. *Textile Research Journal* 91(15-16), 1713–1731.
6. Lorente-Leyva Leandro L, Alemany MME, Peluffo-Ordóñez DH, Araujo RA. 2021. Demand forecasting for textile products using statistical analysis and machine learning algorithms. In: Nguyen Ngoc Thanh and Chittayasothorn S and ND and TB (ed) *Intelligent Information and Database Systems*. Springer International Publishing, 181–194.

CC BY NC ND

7. Xu J, Zhou Y, Zhang L, Wang J, Lefloch D. 2021. Sportswear retailing forecast model based on the combination of multi-layer perceptron and convolutional neural network. *Textile Research Journal* 91(23-24), 2980–2994.
8. Lu Z, He Z, Tran KP, Thomassey S, Zeng X, Hong M. 2021. Decision support systems for textile manufacturing process with machine learning. *Machine Learning and Probabilistic Graphical Models for Decision Support Systems.* CRC Press, 17.
9. Hong Y, Zeng X, Bruniaux P, Chen Y, Zhang X. 2018. Development of a new knowledge-based fabric recommendation system by integrating the collaborative design process and multi-criteria decision support. *Textile Research Journal* 88(23), 2682–2698.
10. Fadara TG, Wong KY. 2023. A decision support system for sustainable textile product assessment. *Textile Research Journal* 93(9-10), 1971–1989.
11. Yasir M, Ansari Y, Latif K, Maqsood H, Habib A, Moon J, Rho, S. 2022. Machine learning–assisted efficient demand forecasting using endogenous and exogenous indicators for the textile industry. *International Journal of Logistics Research and Applications* 0(0), 1–20.
12. Medina H, Peña M, Siguenza-Guzman L, Guaman R. 2022. Demand forecasting for textile products using machine learning methods In: Botto-Tobar, M., Montes León, S., Torres-Carrión, P., Zambrano Vizuete, M., Durakovic, B. (eds) *Applied Technologies,* 301–315. Springer International Publishing.
13. Ghosh S, Hossain S, Rahman H, Zoha N, Islam M. 2020. Developing a linear programming model to maximize profit with minimized lead time of a composite textile mill. *ITEGAM-JETIA* 6(22), 18–21.
14. Fri, M. 2024. Lead time prediction using advanced deep learning approaches: a case study in the textile industry. *Logforum* 20(2), 149–159.
15. Atik C, Kut RA, Birol S. 2021. Estimating lead time using machine learning algorithms: a case study by a textile company. In: Innovations in Intelligent Systems and Applications Conference (1-5). Elazig, Turkey.
16. Susanto S, Tanaya PI, Soembagijo AS. 2012. Formulating standard product lead time at a textile factory using artificial neural networks. In: 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering (99–104). Jalarta, Indonesia.
17. Xia M, Wu Y, Wang Z. 2022. Time delay identification in dynamical systems based on interpretable machine learning. *Journal of Donghua University (English Edition)* 39(4), 332-339.
18. Azevedo J, Ribeiro R, Matos LM, Sousa R, Silva JP, Pilastri A, Cortez P. 2022. Predicting Yarn Breaks in Textile Fabrics: A Machine Learning Approach. *Procedia Computer Science* 207, 2301–2310.
19. Atik C, Kut A, Birant D, Birol S. 2022. Prediction of cloth waste using machine learning methods in the textile industry. In: 9th International Conference on Electrical and Electronics Engineering (165–169). Alanya, Turkey.
20. Onaran, E, Yanık S. 2020. Predicting cycle times in textile manufacturing using artificial neural network. In S. and C. O. S. and O. B. and T. A. C. and S. I. U. Kahraman Cengiz and Cebi (Ed.), *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making*, 305–312. Springer International Publishing.
21. Demirel KC, Şahin A, Albey E. 2019. Ensemble learning based on regressor chains: a case on quality prediction. In: 8th International Conference on Data Science, Technology and Applications (267-274). Prague, Czech Republic.
22. Ma X, Ren M, Wang X. 2023. Research on fabric recognition method based on its properties. In: Batista P, Bilas Pachori R (eds) International Conference on Image, Signal Processing, and Pattern Recognition (1270736). Changsha, China.
23. Priya L, Surendar A. 2024. Optimizing dead stock management in the textile industry through machine learning. 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2024 - Proceedings, 1996–2003.
24. Fang J, Du J, Liu Z. 2022. Digital prediction modeling of fabric softness based on thickness classification. In: Zhao F (ed) 5th International Conference on Computer Information Science and Application Technology. Chongqing, China.
25. Mohamed M. A, El-Henawy I. M, Salah A. 2022. Price prediction of seasonal items using machine learning and statistical methods. *Computers, Materials and Continua* 70(2), 3473–3489.
26. Saraçlar B, Kuyumcu B, Delil S. 2022. Text2Price: deep learning for price prediction. *Artificial Intelligence Theory and Applications* 2(2), 28-38.
27. Katai Y, Hasuike T. 2019. Sales price prediction of used clothes considering the market price of flea market application. In: 8th International Congress on Advanced Applied Informatics (750–755). Toyama, Japan.
28. Mienye ID, Sun Y. 2022. A Survey of ensemble learning: concepts, algorithms, applications, and prospects. *IEEE Access* 10, 99129–99149.
29. Breiman L. 1996. *Bagging Predictors*. Kluwer Academic Publishers.
30. Schapire RE. 1990. The Strength of Weak Learnability. *Machine Learning* 5, 197-227.
31. Borchani H, Varando G, Bielza C, Larrañaga P. 2015. A survey on multi-output regression. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 5(5), 216–233.
32. Rokach, L, Maimon, O. 2005. Decision Trees. In: Maimon, O., Rokach, L. (eds) *Data Mining and Knowledge Discovery Handbook*. Springer.
33. Breiman L, Friedman J, Olshen RA, Stone CJ. 1984. *Classification and regression trees (1st ed.)*. Chapman and Hall/CRC
34. Breiman L. 2001. Random forests. *Machine Learning* 45, 5-32.
35. Chen T, Guestrin C. 2016. XGBoost: A scalable tree boosting system. In: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (785–794). New York, USA.
36. Choi DK. 2019. Data-driven materials modeling with xgboost algorithm and statistical inference analysis for prediction of fatigue strength of steels. *International Journal of Precision Engineering and Manufacturing* 20(7), 129–138.
37. Kohavi R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: 14th International Joint Conference on Artificial Intelligence (1137-1143). San Francisco, USA.

CC BY NC ND