

ULUSLARARASI 3B YAZICI TEKNOLOJİLERİ  
VE DİJİTAL ENDÜSTRİ DERGİSİ

INTERNATIONAL JOURNAL OF 3D PRINTING  
TECHNOLOGIES AND DIGITAL INDUSTRY

ISSN:2602-3350 (Online)

URL: <https://dergipark.org.tr/ij3dptdi>

## FABRIC AND PRODUCTION DEFECT DETECTION IN THE APPAREL INDUSTRY USING DATA MINING ALGORITHMS

**Yazarlar (Authors):** Taner Ersöz <sup>ID\*</sup>, Hamza Zahoor <sup>ID</sup>, Filiz Ersöz <sup>ID</sup>

**Bu makaleye şu şekilde atıfta bulunabilirsiniz (To cite to this article):** Ersöz T., Zahoor H., Ersöz F. "Fabric And Production Defect Detection In The Apparel Industry Using Data Mining Algorithms", *Int. J. of 3D Printing Tech. Dig. Ind.*, 5(3): 742-757, (2021).

DOI: 10.46519/ij3dptdi.1030676

Araştırma Makale/ Research Article

Erişim Linki: (To link to this article): <https://dergipark.org.tr/en/pub/ij3dptdi/archive>

# FABRIC AND PRODUCTION DEFECT DETECTION IN THE APPAREL INDUSTRY USING DATA MINING ALGORITHMS

Taner Ersöz<sup>a</sup> , Hamza Zahoor<sup>b</sup> , Filiz Ersöz<sup>b</sup> 

<sup>a</sup> Actuarial Science Department, Faculty of Business, Karabük University, TURKEY

<sup>b</sup> Industrial Engineering Department, Faculty of Engineering, Karabük University, TURKEY

\*Corresponding Author: [tanerersoz@karabuk.edu.tr](mailto:tanerersoz@karabuk.edu.tr)

(Received: 30.11.2021; Revised: 17.12.2021; Accepted: 29.12.2021)

## ABSTRACT

Nowadays, technology plays a crucial role in fabric production in the textile industry. The demand for high-quality products and rapidly changing economic conditions increase the significance of ready-made clothing manufacturers to produce the right quality product. In addition, in order to minimize production errors, to improve and maintain process performance, it is important to identify the sources of variability during manufacturing. The defective fabric is the main reason which is causing harm to the textile business. Therefore, the proper identification of manufacturing defects leads to a successful business. When it comes to extracting meaningful insights from data and knowledge discovery, data mining has proven its significance in various fields such as business, health, finance, and education. As in all other sectors, data mining is widely used in the textile sector too. In this study, it was aimed to determine the main causes of the error with defective production data of a company that produces clothing by using data mining methods. Decision Tree, Naive Bayes, Random Forest, and Gradient Boosted Trees Algorithms were used in the research. Accuracy rate and Cohen's kappa statistics were taken for comparison algorithms in the study. While determining the main reasons for defective products, the factors of which type of products the company produces for its customers, the sizes of the defective products, types of defects, and explanations were taken into consideration. The most common mistakes in sewing production and the main source of the error were evaluated. According to the results, suggestions were made for the company to take various measures.

**Keywords:** Textile, defect detection, machine learning, data mining, classification algorithms, decision tree.

## 1. INTRODUCTION

The rapidly advancing development and change of innovation and advancement in technology increase customer needs. In order for companies to survive for a long time in the sector where there are multiple alternatives, they must respond to customer needs at the most appropriate value and deadline. In a competitive environment, for companies to sustain their existence and profitability, reducing costs, responding to deadlines faster, and producing without sacrificing quality plays an important role. Firms are turning to various areas to use the resources they already have most efficiently, increase efficiency, capacity, and quality, and reduce costs.

Under global and national competitive conditions, quality and quality improvement activities play a major role in achieving permanent success in the apparel industry. Quality improvement: It is a process that consists of studies conducted to identify and eliminate factors that negatively affect product or service quality in order to increase the level of customer satisfaction. Businesses that apply a regular plan during their quality improvement activities go through the stages of periodic analysis of errors in production, elimination of errors, evaluation of risks, and taking precautions. Thanks to improvement activities, businesses reduce the cost of product error and aim to contribute to the enterprise, worker, and product reliability [1].

Textile is one of the important sectors in our country. Upon close examination, it is clear that the Marmara region, Aegean region, and southeastern Anatolia region are the most productive in Turkey

[2]. The textile products sector offers a wide range of uses, from ready-to-wear to home textile accessories and flags, which are the symbols of independence of the countries. All the product groups that offer this wide range of use come under the name of apparel products [3].

It is possible and even more optimum for textile companies to produce fabrics using technology that lowers the number of defective fabrics. This new method is of course more effective in less time-consuming than the traditional method that relied so mainly on human visuals. Thus, lately, there has been an increment in the usage of technology to improve the production process and also detect any defects in the fabrics. The sewing, knitting, and dyeing unit mostly cover both manual and automatic processes in the sector. It is a very difficult task to detect the defects in the fabric by hand and by visual human inspection. Its accuracy depends on the skill of the human factor.

Ready-to-wear consists of woven and knitted fabrics, clothes, and textile accessories produced for use in all areas of life for everyone. It is a form of clothing that is both comfortable in shape and affordable in price. With the rapid changes in the clothing industry and social-economic conditions, the growth in the technological field has shifted tailor-made work to the ready-to-wear sector [4]. It plays an important and major role in the economic development process of less developed countries due to the fact that ready-to-wear products are in the main and need consumption group, the capital requirement of this sector is lower than other sectors, and it provides the employment and export opportunities [5].

The ready-made clothing group, which is an important gear of the textile and apparel industry has been designed in various sizes and models, which are part of the daily routine, and has become an integral part of modern people. From traditionally used clothes to modern people, the clothes which have undergone many changes in shape, have created a huge sector from special designs to ready-made clothing as they present the needs of every age to human beings [6]. Considering the production stages of the ready-made clothing industry in general, there are stages such as designing the model, removing the mold, cutting, sewing, washing, ironing, and packaging. The clothes obtained through all these processes are offered to the market under the names of brands in different product groups. Even if some parts of the production stages of the ready-to-wear industry have received assistance from computer-aided systems, it still preserves its work based on human power [4].

In the apparel manufacturing industry, the main raw material is fabric, and it can also be defined as accessories and different types of trimmings. Major defects that occur in the apparel manufacturing process are sewing defects, seaming defects, placement defects, fabric defects, embroidery defects, pinhole rework, etc. Sewing defects are usually caused by the malfunctioning of sewing machines. Seaming defects are usually caused by the interaction of the operator and the machine in the processing of the garment. Placement defects are usually caused by marking and cutting and sewing operations in the sewing room. Fabric defects are usually caused by fabric processing processes such as knitting and dyeing. Embroidery defects are usually caused by the processing of garments [7].

Detecting defects and creating patterns have become easier with the help of the artificial eye of the computer which results in a decrement of manufacturing time and labor cost. It has a shorter cycle in using resources and time [8] with the help of minimum labor cost [9] and efficient automated systems. It is known in the fabric industry that these imperfect fabric defects account for 85% of the produced defects [10]. Producer enterprises achieve about 45-65% of their earnings from their second quality or lower quality products [11]. Thus, it also lowers the percentage of mistakes made during the knitting and weaving phase by 80%. In addition, defective production represents up to 5% of textile waste [12].

Asian countries such as China India and Bangladesh are leading the world when it comes to textile production that is an industry based on labor. As in many countries of the world, textile is the first and industrial branch established in Turkey [13]. The prominent products in Turkish ready-to-wear and apparel exports [14] are given in detail in Table 1 prepared according to the export sector report for February 2021.

**Table 1.** Prominent products in Turkish ready-to-wear and apparel production

Scope	2020 January- February	2021 January- February	Change (percentage) (%)	2021 Share (percentage) (%)
<b>Woven suits, jackets etc. for women / girls.</b>	544 141	487 601	-10.4	16.1
<b>T-shirts, athletes and the similar</b>	490 364	417 090	-14.9	13.7
<b>Sweaters, cardigans, vests and the similar</b>	258 747	286 989	10.9	9.5
<b>Woven suits, jackets etc. for men and boys.</b>	272 432	225 282	-17.3	7.4
<b>Knitted suits, jackets etc. for women / girls.</b>	198 061	204 704	3.4	6.7
<b>Bed linen, tablecloths, toilet and kitchen cloths</b>	180 485	200 617	11.2	6.6
<b>Socks and tights</b>	170 314	171 962	1	5.7
<b>Other ready-made items (including the disposable medical masks product group)</b>	52 861	99 634	88.5	3.3
<b>Ready-made garments made of plastic, rubber coated, impregnated fiber</b>	7 715	95 362	1136.1	3.1
<b>Woven blouses, shirts etc. for women / girls.</b>	117 065	87 367	-25,4	2.9
<b>Total ready to wear and apparel export record</b>	<b>3 007 231</b>	<b>3 035 477</b>	<b>0.9</b>	<b>100.0</b>

Within the scope of Turkish ready-to-wear and apparel products listed in Table 1, woven suits, jackets, etc. for women/girls are seen to have occupied a large area in exports with a share of 16.1% in February 2021 and provided 487 601 returns with a decrease of 10.4% compared to the same month of the previous year. T-shirts, undershirts and similar products are seen as the second most important export product with a share of 13.7%. The category product of sweaters, cardigans, vests, and the similar are the third most important product group with a share of 9.5% and a return of 286 989 by providing an increase of 10.9% as compared to the same times of the previous year.

As in many countries of the world, textile is the first established and developing industry branch in Turkey. The total trade volume of the Turkish ready-to-wear industry as of September 2020 is 12.12 billion dollars [15]. As given in Table 2, the data of ready-to-wear and apparel exports according to the general export performance are specified. The export values and rate of change between January and September in 2019 and 2020 are included in Table 2 [16].

**Table 2.** Distribution of ready-to-wear and apparel exports 2017-2020 (1000 \$)

	2017	2018	2019	2020
<b>Ready-to-Wear and Apparel Export</b>	12 629 097	13 236 851	13 277 200	12 123 142
<b>Industry Export</b>	88 271 637	100 172 335	102 212 731	88 921 495
<b>General Exports of Turkey (Including Warehouse and Free Zones Difference)</b>	115 046 252	129 199 568	132 792 571	118 355 460
<b>Special Export of Turkey</b>	-	122 941 962	121 754 717	109 018 223
<b>Share of Ready-to-Wear Exports in General Exports (%)</b>	11.0	10.2	10.0	10.2

As seen in the table, ready-to-wear exports shares of overall exports have declined from 2017 to 2020. Despite the closures and restrictions in the Covid-19 disease, which is declared as a global epidemic in 2020, ready-to-wear exports have increased in the overall export rate. When we look at the ready-to-wear and apparel exports, it is seen that 2020 is lower than 2019 and 2018. The conclusion here is that the decrease in ready-to-wear exports in 2020 has increased proportionally due to the fact that it is less than the decrease in other sectors.

Data analysis is important in the textile and apparel industry, both in the past and today. Many classical mathematical and statistical models have been used in many textile studies to process textile data. The latest technological innovation in artificial intelligence and data mining has proved to be amazingly

useful when it comes to increasing efficiency and smoothing the real-time production process. The use of AI from pre-production and in the process of identification n classification helps increase the overall effectiveness of the business.

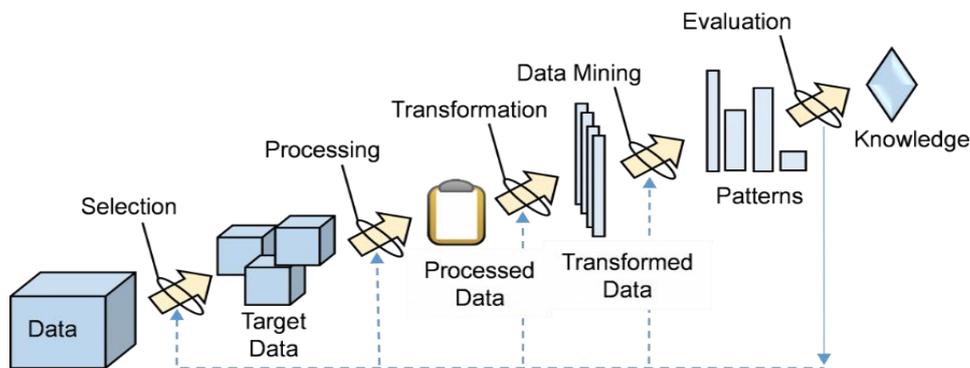
Data Mining is important for the textile industry in terms of analyzing occupational accidents in the sector, detecting errors, determining the number of products, their quality, and what affects customer's satisfaction. The introduction and application of data mining to textile production will increase the productivity of the processes in production and the quality of work. Therefore, in the textile sector, data mining techniques act as a catalyst in analyzing such big data and are adding to its popularity with each passing day.

Data Mining allows us to have an accurate interpretation of the data that the organizations acquire which eventually informs the business to make decisions to improve quality n increase customer satisfaction. Data mining is the process of extracting information from very big data sets. In other words, it is a powerful technology that has the potential to enable organizations to focus on important information in data warehouses [17].

Data mining is basically a technique that arises when numerous disciplines such as statistics, computer science, machine learning are used together. Data mining can be defined as a process of analyzing data stacks from different and diverse sources and converting this valuable information into business intelligence.

With data mining applications, the emergence of meaningful relationships between meaningless data, to predict the impact of the business's sales, profit, and customer satisfaction, and detailed and knowledge-based information such as customer profile analysis are produced. Data Mining services can help organizations interpret data accurately, which in turn helps the organizations make informed business decisions, improve processes, improve quality standards, and increase customer satisfaction [18].

The data mining cycle is completed by extracting and processing information from databases and interpreting the results of the analysis by decision-making techniques. Different steps of Knowledge Discovery and Data Mining [19] are given in Figure 1.



**Figure 1.** Different steps of data mining

In data mining, the data is passed through several stages before modeling, with an aim to achieve specific results from big and meaningless data stacks. As soon as the problem and the system are introduced, the first step is to clear the data before modeling. It's then followed by detection of outliers and extreme values. In this way, clean and high-quality data is obtained. Then, data are combined to achieve the meaningful patterns. Here, the selection of relevant and important variables and dimension reduction are carried out. As a result of the analysis, converting the data into a format suitable for re-use, evaluating the importance of data and relations and presenting the results to the decision-makers are the processes that complete the data mining [18].

The main objective of this study was to determine the main reasons of the error with the help of data mining methods. In the study, defective production data of a company producing clothing was used. In

order to determine the main causes of the defects during the production, the factors such as which type of product does the firm produce for its customer, the measurements of the defective products, different types of defects, and explanations related to them were taken into account. Customers, product kinds and sizes with the most erroneous production will be determined. However, it is aimed to identify the most frequent errors and locate their principal sources.

## 2. LITERATURE REVIEW

In the textile sector, many studies have been conducted for improving product quality and for the detection of fabric defects. For instance, machine learning studies have played a main role in understanding the stimulation of the manufacturing process and the identification of defects.

Data mining and machine learning applications are used to manage production quality. Moreover, the examples such as business intelligence platforms for monitoring productivity and real-time quality control, usage of artificial intelligence for detecting textile defects, and easy-to-install cameras and sensors for automatic inspection can be taken into account. These elements play a vital role in increasing performance and productivity. There are many academic studies, especially simulation [20], quality control [21-22], image perception [22], data mining [23, 24], decision making in apparel manufacturing [25-28], advancement of textile manufacturing processes [29], production and demand forecasting in textile industry [30-31]. In the literature review, the use of data mining in the textile industry and the studies on production defects were examined.

Murino et al. (2004), discussed the classification problem of defects in textile manufacturing. A new classification scheme was designed using different features extracted from the gray level histogram, shape, and concurrency matrices. The designed system has been tested using two textile databases showing promising results [32].

Mozafary and Payvandy (2014), examined the raw material testing and yarn quality parameters using the data from the worsted yarn mill. They estimated the yarn quality with the K-means algorithm, one of the artificial neural networks and clustering methods, and compared the results with the results obtained by the traditional method. The performance of data mining methods has been better [16].

Habib et al. (2016) have determined and classified various defects in textile fabrics in their studies. They extracted features using statistical techniques and used images of textile fabrics as an example. Using a suitable Bayes classifier to classify images into different defect property classes, they have achieved acceptable accuracy in the field of textile defect detection and classification [33].

Mohanty and Bag (2017) focused on image mining in their work. In the textile industry, they argued that flaw inspection in the fabric is a labor-prone and costly process that is performed in a semi-automatic way with human visual vision. In the research, texture analysis, association rule, and threshold segmentation were applied to detect textile defects. The feature determination is based on the Gray Level Co-Occurrence Matrix [34].

Ersöz et al. (2017) made a classification with decision tree algorithms using data for sewing products in a textile company. They found variables that affect production per capita and as a result, as the daily working time increases, per capita production decreases, and the relationship between total daily work and per capita production is negative [35].

In his studies in 2017, Tagluk emphasized the importance of determining the defects of fabrics woven on looms in textile, on a full-time basis and argued that fabric defects vary depending on the sources. For defect detection, they recommended a special filtering system considering the color, texture and morphological properties of the fabric, and defective products were separated at a rate of 99.10% [36].

Cevik et al. (2017), in his studies, in the textile sector, leather, fabric, etc. They mention that there are errors arising from manual production in the cutting of the products and discuss that these processes create a waste of time and risk. An image processing-based industrial quality control system that automatically determines cutting errors in textile products and distinguishes defective and error-free

products is explained. The described system has been tested on leather pieces purchased from a company in the leather industry and has been tested 150 times for 50 pieces of leather in 5 different templates in total, and these pieces are separated 149 times (99.33% success rate) for defective ones and it is decomposed [37].

In his studies in 2019, Blanes emphasized how crucial image processing is for the detection and classification of defects in the textile industry. In his study, he prioritized examining new methods and approaches for consistently improving each and every aspect of the production process [38].

Tosun (2020) made determinations to increase production efficiency in his study on a ready-to-wear producing firm. In this study, the simulation method is used to balance the assembly line and improve production processes. In the research, a balanced production line has been established with simulation to reduce the production cost to determine the necessary manpower and machine resources and use them effectively and efficiently to meet customer demands in a complete and timely manner. As a result of the research, it was observed that the capacity utilization rate of the enterprise increased significantly and the working capacity utilization rate increased with the more efficient use of the workers in the production line [39].

In their study, Amor et. al. (2021) emphasized the significant increment can be observed in the numbers of textile product applications for quality standards. In this study, textile products were classified by using polymer-based and fiber-reinforced polymer composites. Furthermore, many problems in artificial neural textile processes, fiber-reinforced polymer composites networks, genetic algorithms, and fuzzy logic methods were analyzed and their probable results were discussed [40].

### 3. METHODOLOGY

In this study, it was aimed to determine the main causes of the defect by using data mining methods. For this purpose, defective production data of a company that produces women's clothing was used. While determining the main reasons for defective products, the factors such as which type of products (pants, skirts, coats) the company produces for its customers, the sizes of the defective products, types of defects and explanations were taken into consideration.

The methodological approximation followed in the study is given in Figure 2. The research consists of three steps. The first step explains the business and data understanding. Second is the data pre-processing step that includes data cleaning, data integration, feature selection, and data reduction. The third step is providing details about the features and targets. In the fourth step, the classifiers were trained individually for each dataset. Lastly, the model-building step presents the development of the systems and the evaluation methods are used.

Generally, data classification consists of two-stage. The first indicates algorithm learning or training. This is the stage in which the model is developed by creating classes corresponding to the predetermined classes. Here it is assumed that each data represents a predefined class. The second stage is the testing phase. This phase uses a different set of data samples to estimate the accuracy of classification.

Training and testing phases were applied in the classification of the data. In the test phase, the model is tested to estimate the value of the class, and the actual and predicted values are compared. Also, no transformation is required for the data. The decision tree is evaluated by accuracy metric. %70 is used as a training dataset in decision tree learner operator and the remaining 20% data are used as test data by all algorithm predictors. The "Gini" measurements as quality measures were compared for performance. The basic principle of the decision tree is to divide the tree into two branches starting from each decision node. The first attribute to be divided from and the division value is calculated by assaying at the index value of Gini. The index value of Gini can be described as the ratio of attributes in the data set. If the Gini values of two assets show similarity, the distributions of the result are the same.

One of the most used algorithms for classification and prediction, decision trees was used in the study. It is the first approach used in the study in terms of convenient interpretation and for achieving a better understanding. A decision tree is an often-used data mining approach for classification and prediction. Decision Tree and Naive Bayes, Random Forest, and Gradient Boosted Trees algorithms were used in

the research. Among classifier techniques, Decision trees are one of the most widely used supervised learning algorithms. A classification approach is a structured approach having multi-classification. The accuracy of the algorithm results is quite high and the tree is formulated into a hierarchical structure during the data training phase [41,42]. Naive Bayes is a probabilistic machine learning classifier technique that makes classifications based on the Bayes' Theorem. In general, it is a fast, and easy-to-use technique [43]. Random forest is a supervised learning algorithm that can be used for both regression and classification. In the random forest algorithm, the bootstrap aggregating technique is applied to a large number of decision tree learners [44]. The bootstrap aggregating technique is the process of creating sub-training datasets using existing data with replacement [45]. Gradient Boosting algorithm is a regression, classification and machine learning technique that generates a predictive model by generalizing the produced model in stages. Gradient boosting makes it an additive training. At each iteration, the gradient boosting algorithm optimizes a new classifier by adding a new classifier [46].

The classification of the individual subsystems is discussed with a comparison between the performances of the four algorithms— Decision Tree, Naive Bayes, Random Forest, and Gradient Boosted Trees for each garment dataset.

In the model performance comparison, Cohen's Kappa measurement and model accuracy rates were used. It was determined that it was the decision tree that gave the highest values and the model results were interpreted.

Cohen's Kappa performance measure is a measure of how well the classifier is performing. The greater the difference between accuracy and zero error rate in this performance indicator, the higher the Kappa score. Cohen's Kappa is calculated with the following equation.

$$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e},$$

In the calculation of the model accuracy rate, the confusion matrix is used. Confusion Matrix is a performance measurement for machine learning classification. It is a table with four different combinations of predicted and actual values. The accuracy rate is calculated with the following equation.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN})$$

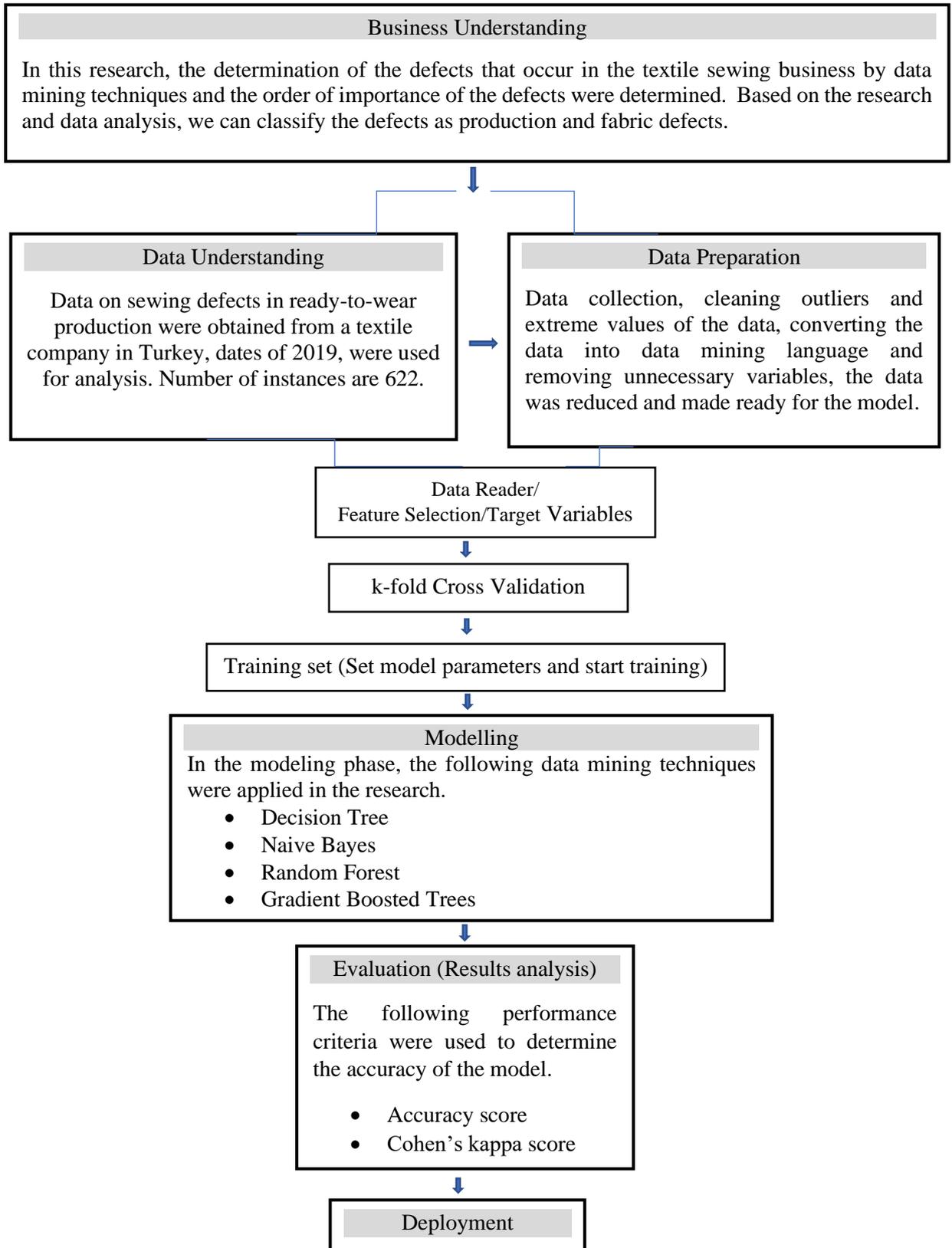
The basic concepts used in calculating the accuracy rate are explained below.

TP (True Positive): The value in the test data and the class valued by the model are the same. It has made the correct classification.

FN (False Negative): The value in the test data and the class produced by the model are different. The positive principle is classified as negative. It has made an incorrect classification.

FP (False Positive): While the true value is negative, it is classified as positive. It has made an incorrect classification.

TN (True Negative): While the true value is negative, it is classified as negative. It has made the correct classification.



**Figure 2.** Data mining methodology of the research

#### 4. DATA MINING APPLICATION

In this research, the determination of the defects that occur in the textile sewing business by data mining techniques and the order of importance of the defects were determined. Based on the research and data analysis, we can classify the defects as production and fabric defects. Since it is a classification problem, we predict certain types and output classes/labels based on sizes. The application study has been given under the data mining process.

- **Data Understanding:** This stage is understanding the data for modeling. Data on sewing defects in ready-to-wear production were obtained from a textile company. The datasets were downloaded for the purpose of classification model performance analysis of KNIME. The data set is in .xls format. The data used in the data mining application was taken from a textile manufacturing company and belongs to the year 2019. The number of samples is 622. The data table for datasets is presented in Table 3.

**Table 3.** Research dataset

Row ID	S Product...	I Model ...	I Product...	S Defect	S Defect ...	B Selecte...
Row0	Coat	1	6	Production	Sewing defect	false
Row1	Coat	1	8	Production	Sewing defect	false
Row2	Coat	1	8	Production	Scissors cut ...	false
Row3	Coat	1	10	Production	Iron usage ...	false
Row4	Coat	1	12	Production	Scissors cut ...	false
Row5	Coat	2	8	Production	Production-i...	false
Row6	Trousers	3	2	Production	Measureme...	false
Row7	Trousers	4	10	Production	pocket stitc...	false
Row8	Trousers	5	6	Production	Fabric_zedel...	false
Row9	Trousers	5	10	Production	Fabric_zedel...	false
Row10	Trousers	5	12	Production	Sewing defect	false
Row11	Coat	6	8	Production	Scissors cut ...	false
Row12	Coat	6	8	Production	Iron usage ...	false
Row13	Coat	6	8	Production	Sewing defect	false
Row14	Coat	6	10	Production	Stitch symm...	false
Row15	Coat	6	10	Fabric	Fabric textu...	false
Row16	Coat	6	12	Production	Sewing defect	false
Row17	Trousers	7	1	Production	Sewing defect	false
Row18	Trousers	7	1	Production	Stitch symm...	false
Row19	Trousers	7	2	Production	Scissors cut ...	false
Row20	Trousers	7	2	Production	Sewing defect	false
Row21	Trousers	7	2	Production	Stitch symm...	false
Row22	Trousers	7	2	Fabric	Thread pulli...	false
Row23	Trousers	7	3	Production	Scissors cut ...	false
Row24	Trousers	7	3	Production	Buttonhole ...	false
Row25	Trousers	7	3	Production	Thread pull ...	false
Row26	Trousers	7	4	Production	Sewing defect	false
Row27	Trousers	7	4	Fabric	Thread knot...	false
Row28	Trousers	7	5	Production	Sewing defect	false
Row29	Trousers	8	0	Production	Measureme...	false
Row30	Trousers	8	2	Production	Sewing defect	false
Row31	Trousers	8	2	Production	Production-i...	false
Row32	Trousers	8	3	Production	Sewing defect	false
Row33	Trousers	8	4	Production	Scissors cut ...	false
Row34	Blazer Jacket	9	4	Production	Stitch size d...	false
Row35	Blazer Jacket	9	6	Production	Scissors cut ...	false
Row36	Blazer Jacket	9	6	Production	Stitch symm...	false
Row37	Blazer Jacket	9	6	Production	Sewing defect	false

- **Data Preparation:** At the stage of model preparation; data collection, data integration, clearing outlier and extreme values of data, converting data into data mining language, and reducing the data by removing unnecessary variables entered the model.
- **Data Collection and Integration:** The data of the study was obtained from a ready-to-wear clothing firm. There is no integration in the data. Classified into defect type.

- Data cleaning and Outliers:** It means to improve data quality and identifying, cleaning or identifying missing, noisy, and inconsistent data in the data. To determine whether there is a deviation or abnormal value in the study data, the data was examined. There are no outlier and extreme values.
- Data Transformation and Reduction:** To prepare the data, it is necessary to remove (reduce) the variables that will not be used in the analysis. Descriptive statistical results of the data are given in Table 4.

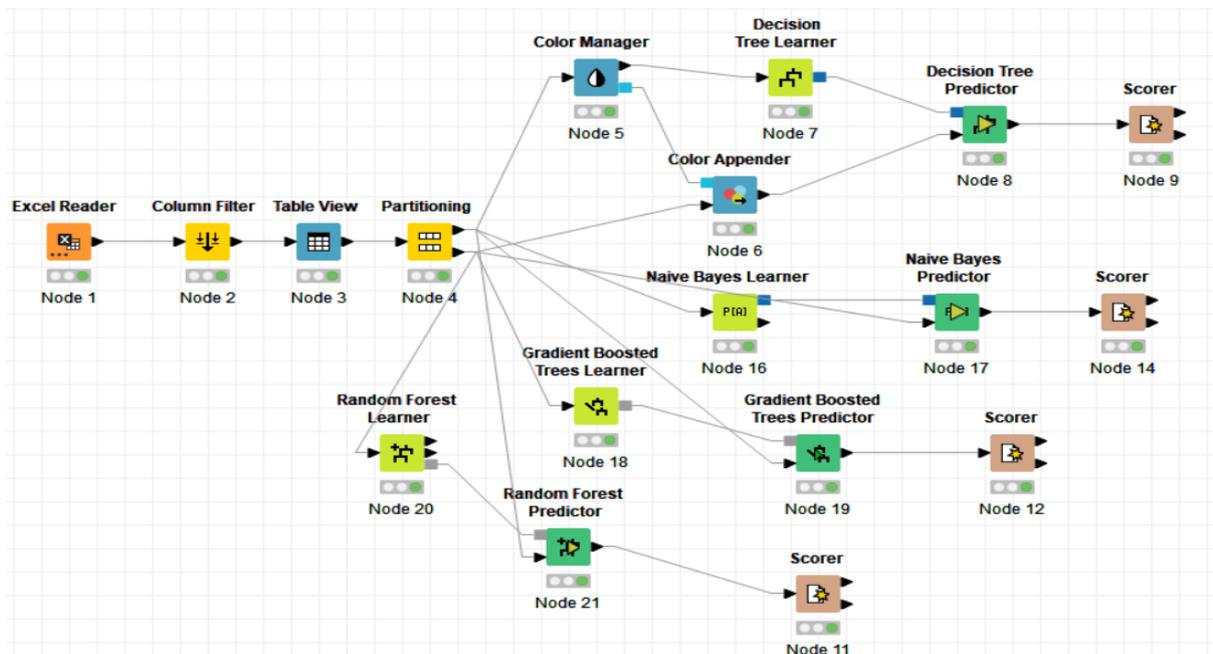
**Table 4.** Descriptive statistics

Field	Graph	Type	Min	Max	Mean	Std. Dev	...	Median	Mode	Unique	Valid
Product_T...		Set	--	--	--	--	--	--	Blazer Jacket	5	622
Model Nu...		Range	1.000	41.000	23.159	9.686	...	19.000	18.000	--	622
Product Size		Range	0.000	44.000	5.262	6.728	...	3.000	2.000	--	622
Defect Code		Set	101.000	205.000	--	--	--	--	202.000	20	622
Defect		Flag	--	--	--	--	--	--	Production	2	622
Defect Type		Set	--	--	--	--	--	--	Thread pulling defect in fabric	19	622

When Table 4 is examined, the most occurring defect related to stitching was seen in the "Blazer jacket". The maximum sewing production size is "2" coded product, the defected product code is "202"; the production defect is greater than fabric defect. The defect type with the greatest number of observed was found to be "Thread pulling defect in fabric".

- Data Modeling and Evaluation:** Data mining Decision Tree and Naive Bayes, Random Forest, and Gradient Boosted Trees algorithms were used in the research. At the end of the research, the accuracy rates of the model were compared. The decision tree algorithm giving the highest accuracy rate is interpreted.

The defects whether related to production or fabric itself are dependent on product type, model number, and product size. The aim of this research is to assay the type of defect and its possible causes. The screenshot of data mining in the Knime program is given in Figure 3.



**Figure 3.** The screenshot of data mining in the Knime program

**Performance of Classification Models:** Comparing the actual class of each observation with the predicted class, the accuracy, and other performance measures (Overall Accuracy, TP, FN, True Negative, Recall, Precision, Sensitivity, Specificity) of the predictive model are estimated. Accuracy level and Cohen’s kappa value were taken for comparing algorithms in the study. The decision tree is evaluated by accuracy metric. 80% of data are used as training datasets in the decision tree learner operator and the remaining 20% are used as test data by all algorithm predictors. In addition, the “Gini” and “Gain” measurements as quality measures were compared for performance and it was determined that the “*Gini*” quality measurement showed higher accuracy.

When the algorithm results were compared, it was seen that “Decision tree” algorithms had higher accuracy rates than other used classifier algorithms. Accuracy and Cohen’s Kappa statistics are reported in Table 5.

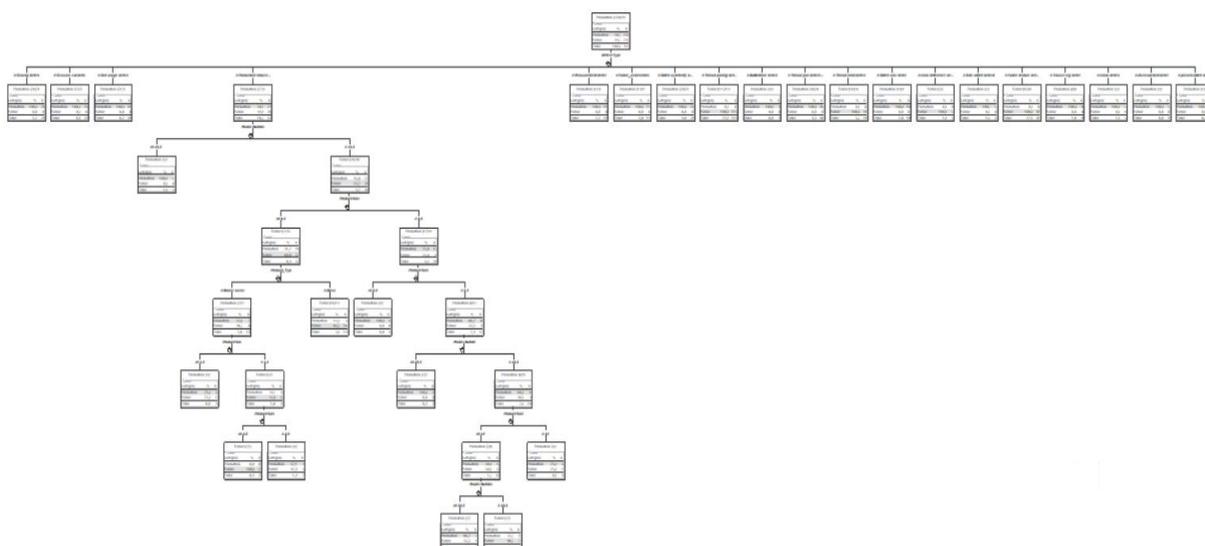
**Table 5.** Data mining algorithm results

Data Mining Algorithms	Accuracy Score (%)	Cohen’s Kappa Coefficient ( $\kappa$ )
<b>Decision Tree</b>	<b>0.960</b>	<b>0.918</b>
Naïve Bayes	0.952	0.904
Gradient Boosted Trees	0.956	0.911
Random Forest	0.952	0.903

Confusion Matrix is generated with the scorer operator in KNIME. When the algorithm results were compared, it was seen that “*Decision tree*” algorithms had higher accuracy rates than other used classifier algorithms. In the confusion matrix of the decision tree algorithm studied in the research, we clearly observed that the accuracy level is 0.96 % and Cohen’s kappa statistics is 0.918.

Whereas naïve Bayes’ accuracy level was observed to be 0.952, standing right under the decision tree. gradient boosted trees and random forest’s accuracy levels are 0.956 and 0.952, respectively. The aim is to classify the defect type and get to the most possible causes of the defects. With the help of the decision tree algorithm, we have reached some conclusions and the results for the decision tree algorithm screenshot and comment are given below.

According to the results of the research, the most important factor determining the characteristics of the defect is the "defect type" as an estimate. If the defect type is the measurement, pocket, fabric damage, sewing, stitching, or fabric texture defect, it means that 100 % of the defects are production-related. The decision tree algorithm results are given in Figure 4.



**Figure 4.** The screenshot of data mining in the Knime program

When the decision tree estimation model is examined, it has been observed that the most important error is the "defect type" variable. This variable is followed by model number, production size, and product type variables. All error types ("scissor cut defect", "sewing defect", "iron usage defect" etc.) except production-induced stain defect are 100 % production defect. The production-induced stain defect result is given in Figure 5.

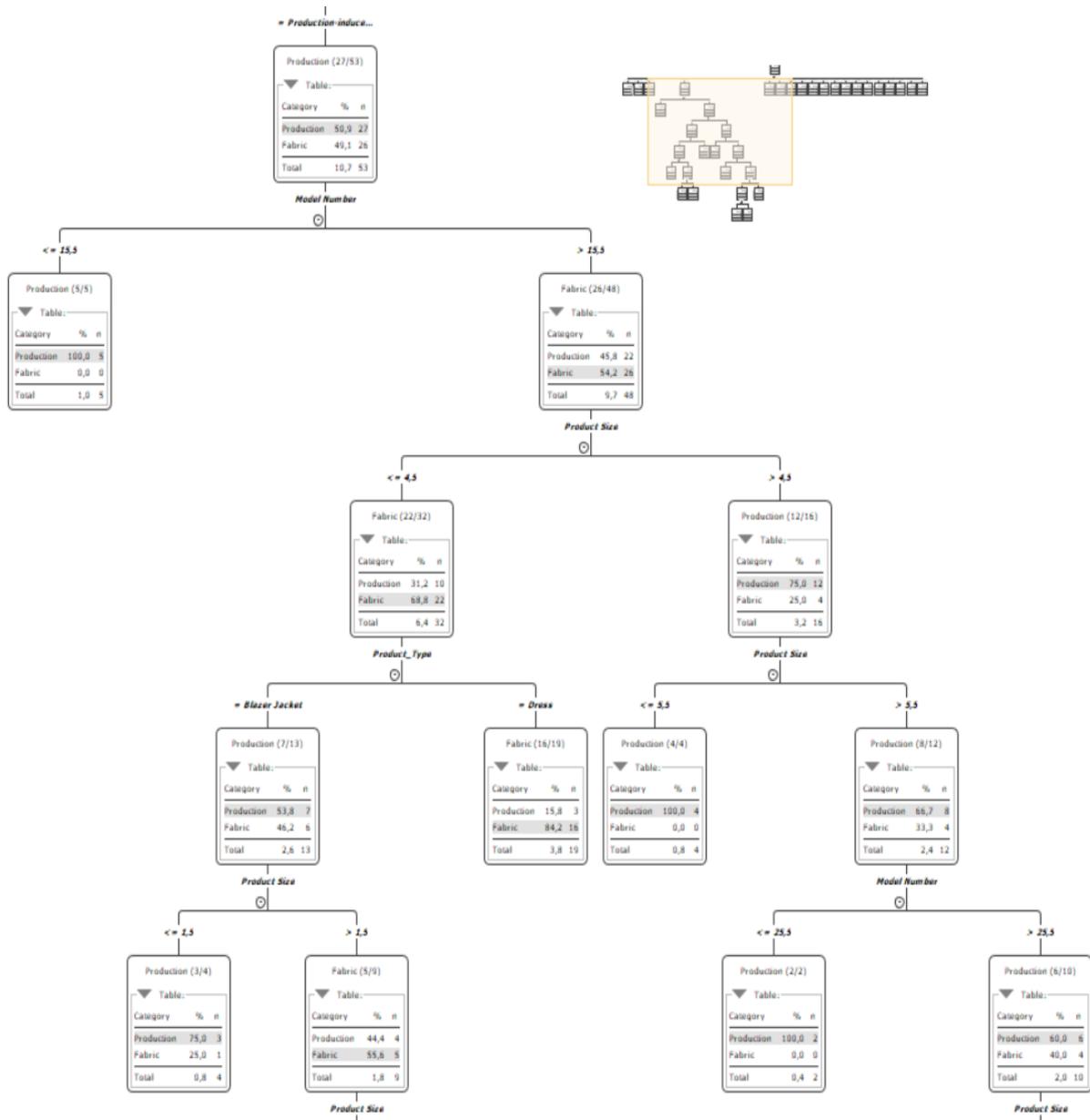


Figure 5. The decision tree algorithm results

If “production-induced stain defect” and the model number is “15” and below, it is 100% “production defect”. If the model number is “15” or higher, the production size is over 5, and the model number is less than 25, it is 100% “production defect. Model number 15 and higher, production size less than 5, and production type "dress" are 84.2% factory defect and 15.8% production defect. If the model number is “15 and above”, the production size is less than “5”, the production type is "blazer jacket" and the production size is less than “3”, it is 100% production defect. If the defect code is below 114, it is 100% production error, and above code 201 it is 100% fabric error. The defect code is between 114 and 201 and fabric texture defect is 100% production defect, and trouser leg defect is 100% fabric defect.

## 5. CONCLUSION

The increment in data in the textile industry day by day and the use of techniques such as data mining, machine learning, and artificial intelligence have become quite common. It is necessary to be involved in the process of digitalization that emerges with developing technology in order to ensure sustainability. Therefore, by revealing the business characteristics required for digitization and testing awareness, it can be revealed that how ready businesses are for the digitization process. The integration of computer software and hardware is now crucial in the apparel industry. The Internet of Things (IoT), where Industry 4.0 is involved in almost every business, production machines such as 3D printers in production, and data mining applications in productivity research and many other technology-oriented studies are now seen [47-51].

With the help of existing data and analytical approaches such as machine learning and data mining, developed large textile enterprises can increase productivity and employee performance, reduce costs and increase operating profit. To put it in a nutshell, the aim of the study was to determine the main causes of the defect faced in the textile industry by using data mining methods.

Data mining classification algorithms were applied in order to determine the core causes of the defective products. At the end of the research, the accuracy rates of the model were compared. It was seen that "Decision tree" algorithms had higher accuracy rates than other used classifier algorithms. It was observed that the main reasons of the defects were production-induced defects and the fabric defects. The occurrence of defects was determined in the defect type, model number, production size and product type, respectively. All seen defects, except for production-related stain defects, are production defects. According to the results of the research, the most important factor determining the characteristics of the defect is the "defect type" as an estimate. If the defect type is the measurement, pocket, fabric damage, sewing, stitching, or fabric texture defect, it means that 100 % of the defects are production-related.

It has been determined that the important variable affecting the production defect is the model number, model size, and then the product type. If the model code of blazer jacket production is less than 3, it has been determined that there is a production defect and if it is greater than 3, there is a fabric defect. In addition, it was observed that most of the garment production was fabric defects.

Sewing production enterprises should consider defects for sewing production found with data mining. When these errors are eliminated, the business will obtain efficient, high-performance, and quality production.

With the development in technology, many applications are brought into use. Thanks to the improvement in machine learning, artificial intelligence, and data mining technologies, more manufacturing and industrial applications will be used in the future of the textile manufacturing sector. Increasing awareness of industry 4.0 now benefits manufacturers in the apparel industry and will guide far beyond future trends.

## REFERENCES

1. Cheng, S.Y, Yuen C.W.M., Kan C.W., Cheuk K.K.L., "Development of cosmetic textiles using microencapsulation technology", Research Journal of Textile and Apparel, Vol. 12 Issue 4, Pages 41-51, 2008.
2. Ersoz, T., Tenbeli, R., Ersoz, F., "Visual analysis of textile sector of Turkey using gephi complex network", 11th International Symposium on Intelligent Manufacturing and Service Systems, May 2021.
3. Devlet Planlama Teşkilatı, "Tekstil ve hazır giyim sanayi özel ihtisas komisyonu raporu", DPT, Ankara, 1982.
4. Aydın, E., "İnternet tüketicilerinin hazır giyim satın alma davranışları üzerine bir araştırma", Yüksek Lisans Tezi Kastamonu Üniversitesi Sosyal Bilimler Enstitüsü, Kastamonu, 2019.
5. Atan, M. A., "Hazır giyim üretim planlamasında karşılaşılan sorunlar ve bir model önerisi", Yüksek Lisans Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara, 2011.

6. Akyol, A., "Tekstil ve hazır giyim sektörüne pazar oryantasyonu açısından genel bir bakış". Pazarlama Dünya Dergisi, Vol. 45, 2001.
7. Fibre2fashion website, Quality and productivity improvement in apparel industry, Pub. 2013, Retrived from <https://www.fibre2fashion.com/industry-article/7220/quality-and-productivity-improvement-in-apparel-industry>
8. Smith, M.L., Stamp, R.J., "Inspection of textured ceramic tiles using automation", Computers in Industry, Vol. 43 Issue 1, Pages 73-82, 2000.
9. Chin, R. T., Harlow, C. A., "Automated visual inspection, IEEE Trans, Pattern Anal. Machine Intell., Vol. 4 Issue 6, Pages 557-573, 1982.
10. Sengottuvelan, P., Wahi A., Shanmugam A., "Use of imaging systems for automatic defect analysis", Research Journal of Applied Sciences, Vol. 3 Issue 1, Pages 26-31, 2008.
11. Srinivasan, K., Dastoor, P.H., Radhakrishnaiah & Sundaresan Jayaraman., "FDAS: A knowledge-based frame detection work for analysis of defects in woven textile structures", Textile Institute's journal, Vol. 83, Issue 3, Pages 431-448, 1992.
12. Smartex website, 04 September 2021, "Retracing production cost", Retrieved from <https://www.smartex.ai/>
13. Uyanık, S., Çelikel, C.D., "Türk tekstil endüstrisi genel durumu", Teknik Bilimleri Dergisi, Vol. 9, Pages 32-41, 2019.
14. Hazır giyim ve konfeksiyon sektörü 2021 şubat aylık ihracat bilgi notu. <https://www.itkib.org.tr> Accessed April 2021.
15. Küheylan, Z., "İhracatın parlayan yıldızı hazır giyim sektörü", İzmir Ticaret Odası, İzmir, 2020.
16. İstanbul Tekstil Konfeksiyon İşletmeler Birliği, "Hazır giyim ve konfeksiyon sektörü 2020 raporu", İstanbul. 2020.
17. Mozafary, V., Payvandy, P., "Application of data mining technique in predicting worsted spun yarn quality", The Journal of The Textile Institute, Vol. 105 Issue 1, 2014.
18. Ersöz, F., "Veri Madenciliği Teknikleri ve Uygulamaları", Seçkin Yayınevi, Ankara, 2019.
19. Piatetsky-Shapiro, G., Fayyad, U., Smyth, P., "Knowledge discovery and data mining: towards a unifying framework", AAAI Press / The MIT Press, Menlo Park, CA. Pages 82-88, 1996.
20. Seçkin, M., Seçkin, A.Ç., Coşkun, A., "Simulation of production fault and forecasting from time series data with machine learning in glove textile industry", Journal of Engineered Fibers and Fabrics, Vol. 14, 2019.
21. Tyagi, S.K., Sharma B.K., "Data mining tools and techniques to manage the textile quality control data for strategic decision making, International Journal of Computer Applications (0975 – 8887), Vol. 13 Issue 4, January 2011.
22. Vajihe, M., Pedram, P., "Application of data mining technique in estimating worsted yarn quality", The Journal of The Textile Institute, Vol. 105 Issue 1, Pages 100-108, 2014.
23. Javed, A., Mirza, A.U., "Comparative analysis of different fabric defects detection techniques", International journal of image, Graphics and Signal Processing. Vol. 5, Issue 10, 2013.
24. Yıldırım, P., Birant, D., Alpyıldız, T., "Data mining and machine learning in the textile industry", WIREs data mining knowledge Discovery, DOI: 10.1002/widm.1228, 2018.
25. Colson, E., Coffey, B., Rached, T., Cruz, L. (n.d.), "Algorithm's tour: How data science is woven into the fabric of Stitch Fix", Retrieved from <https://algorithms-tour.stitchfix.com/> 03 September 2021.

26. Guo, Z. X., Wong, W. K., Leung S. Y. S., Li M., “Applications of artificial intelligence in the apparel industry: A review”, *Textile Research Journal*, Vol. 81 Issue 18, Pages 1871- 1892, 2011.
27. Jelil, R. A., “Review of artificial intelligence applications in garment manufacturing”, In *artificial Intelligence for Fashion Industry in the Big Data Era*. Springer, Singapore, Pages 97-123, 2018.
28. Sirovich, R., Craparotta G., Marocco, E., “An intelligent fashion replenishment system based on data analytics and expert judgment”, In *artificial Intelligence for Fashion Industry in the Big Data Era*, Springer, Singapore, Pages 173-195, 2018.
29. Tağman, A.B., “Sistem Simülasyonu ile süreç iyileştirme: bir tekstil işletme uygulaması”, Yüksek Lisans Tezi, Karabük Üniversitesi Lisansüstü Eğitim Enstitüsü, Karabük, 2021.
30. Tozak, E., “Veri madenciliği programları kullanılarak bir tekstil firmasının satış verilerinin değerlendirilmesi”, Yüksek Lisans Tezi, Karabük Üniversitesi Lisansüstü Eğitim Enstitüsü, Karabük, 2021.
31. Odabaş, S., “Tekstil sektöründe ihracat yapan bir firmada talep tahmini uygulaması”, Yüksek Lisans Tezi, Karabük Üniversitesi Lisansüstü Eğitim Enstitüsü, Karabük, 2021.
32. Murino, V., Bicego, M., Rossi, I.A., “Statistical classification of raw textile defects”, *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)* 1051-4651/04, 2004.
33. Habib, T. Md., Shuvo, S.B., Uddin, M.S., Ahmet, F., “Automated textile defect classification by bayesian classifier based on statistical features”, *Conference: International Workshop on Computational Intelligence (IWCI)*, 2016.
34. Mohanty, A.K., Bag, A., “Detection and classification of fabric defects in textile using image mining and association rule miner”, *International Journal of Electrical, Electronics and Computers (EEC Journal)*, Vol-2, Issue 3, 2017.
35. Ersoz, F., Ersoz, T., Guler, E., “Knowledge discovery and data mining techniques in textile industry”, *International Journal of Computer and Information Engineering*, Vol. 11, Issue 7, 2017.
36. Tagluk, M., “Dokuma tezgahlarında hatalı kumaş dokusunun tespiti için başarılı bir yöntem”, *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, Vol. 8, Pages 575-586, 2017.
37. Cevik, K.K., Koçer, E., “Computer aided control of cutting error in textile products”, *Textile and Apparel*, Vol. 27 Issue 3, Pages 300-308, 2017.
38. Silvestre-Blanes J., Albero-Albero, T., Miralles I., Pérez-Llorens, R., Moreno, J., “A public fabric database for defect detection methods and results”, *Autex Research Journal* Vol. 19, No.4, Pages 363-374, 2019.
39. Tosun, D., “Hazır giyim sektöründe benzetim tekniği kullanılarak üretim hattının dengelenmesi”, Yüksek Lisans Tezi, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, Denizli, 2020.
40. Amor, N., Noman M. T., Petru, M., “Classification of textile polymer composites: recent trends and challenges”, *Polymers*, Vol. 13 Issue 16, Pages 2592, 2021.
41. Liu, L., Su, J., Zhao, B., Wang, Q., Chen, J., Luo, Y., “Towards an efficient privacy-preserving decision tree evaluation service on the internet of things”, *Symmetry*, Vol. 12, Issue 103, 2020.
42. Song, Y.Y., Lu, Y., “Decision tree methods: Applications for classification and prediction”, *Shanghai Arch. Psychiatry*, Pages 130–135, 2015.
43. Mccallum, A., Nigam, K. A., “Comparison of event models for naive bayes text classification”, In *AAAI-98 Workshop on Learning for Text Categorization*, Madison, Wisconsin, Pages 41–48., 1998.
44. Siddiqui, M.F., Mujtaba, G., Reza, A.W., Shuib, L., “Multi-class disease classification in brain MRIs using computer-aided diagnostic system”, *Symmetry*, Vol. 9, Issue 37, 2017.

45. Aggarwal, C., "Data Classification: Algorithms and Applications", CRC Press: Boca Raton, FL, USA, 2014.
46. Son, J., Jung, I., Park, K., Han, B., "Tracking-by segmentation with online gradient boosting decision tree", In Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, WA, 2015.
47. Merdin, D., Ersöz, F., "Evaluation of the applicability of industry 4.0 processes in businesses and supply chain applications", 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Pages 1-10, 2019.
48. Aksoy, B., Uğuz, S., Oral, O., "Comparison of the data matching performances of string similarity algorithms in big data", Mühendislik Bilimleri ve Tasarım Dergisi, Vol. 7, Sep. No. 3, 608–618, 2019.
49. Ersoz, F., Merdin, D., Ersoz, T., "Research of industry 4.0 awareness: A case study of Turkey", Economics and Business, ISSN: 1407-7337, Vol: 32, Issue: 1, 247-263, 2018.
50. Yalçınkaya, S., Kılıçarslan, Y., Ersöz, F. ve diğerleri, "Sanayi 4.0 Teknolojik Alanları ve Uygulamaları", Pegem Yayınevi, 2019.
51. Ersoz, F., Ersoz, T., Artuc, B., "A general overview on industry 4.0 and society 5.0: A case study of awareness of industry 4.0", Journal of Turkish Operations Management, JTOM, Special issue, Pages 120-130, 2018.