International Journal of Technology and Engineering Studies
volume 4 issue 2 pp. 64-76 doi: https://dx.doi.org/10.20469/ijtes.4.10004-2

Improve Quality and Efficiency of Textile Process using Data-driven Machine Learning in Industry 4.0

Chia-Yun Lee
Engineering National Taiwan University,
Taipei Taiwan

Jia-Ying Lin*
Engineering National Taiwan University,
Taipei Taiwan

Ray-I Chang
Engineering National Taiwan University,
Taipei Taiwan

Abstract: This paper focuses on the relationship between key operation parameters and machine learning defects to design an Operation Parameters Recommender System (OPRS) in the textile industry. This paper integrates historic manufacturing process data from the perspective of data science, such as machine operation parameters from warping, sizing, beaming, weaving process, and management experience data, such as textile inspection results from the quality control section. Then, the regression models are applied to predict the textile operation parameters. This research also uses the classification models to predict the quality of the textile. Based on the ten-fold cross-validation testing, experimental results show that our model can achieve 90.8% accuracy on the quality level prediction. The best regression model for predicting weaving operation parameters can reduce the mean square error (MSE) 0.01%. By combining the above two models, the proposed OPRS can provide a completed analysis data of operation parameters. It provides good performance when comparing with previous stochastic methods. As the proposed OPRS can support technicians setting operation parameters more precisely, even for a new type of yarn, it can help to fix the tech skills gap in the textile manufacturing process.

Keywords: *Machine learning, Industry 4.0, cyber-physical system, process improvement*

Received: 20 November 2017; ***Accepted:*** 23 February 2018; ***Published:*** 13 April 2018

I. INTRODUCTION

Today, electronic devices with high acquisition rate sensors are widely used in the textile industry to collect the real-time and continuous data. Therefore, a large number of data are available to build a data-driven statistical model for predicting operation parameters in manufacturing process. However, the textile industry has one of the most complicated industrial chains [1] in the manufacturing industry as shown in Figure 1. Characterizing its operation parameters is complicated because of the wide variety of substrates, processes, machinery, and components used. Different types of yarns, methods of production, and finishing processes (preparation, printing, dyeing, and coating), all interrelate in producing a finished fabric and relate to each other. In addition, the modern textile industry also faces a great challenge in dealing with small-volume and large-variety customized order for this fast fashion generation. In this paper, we integrate historic manufacturing process data and management experience data to predict the operation parameters by the regression models. We also use the classification models to predict the quality of textile. Experimental results show that our model can achieve 90.8% accuracy on the quality level prediction and the best regression model for predicting operation parameters can reduce the MSE to 0.01%.

*Correspondence concerning this article should be addressed to Jia-Ying Lin, Engineering National Taiwan University, Taipei Taiwan. E-mail: blue945u@gmail.com

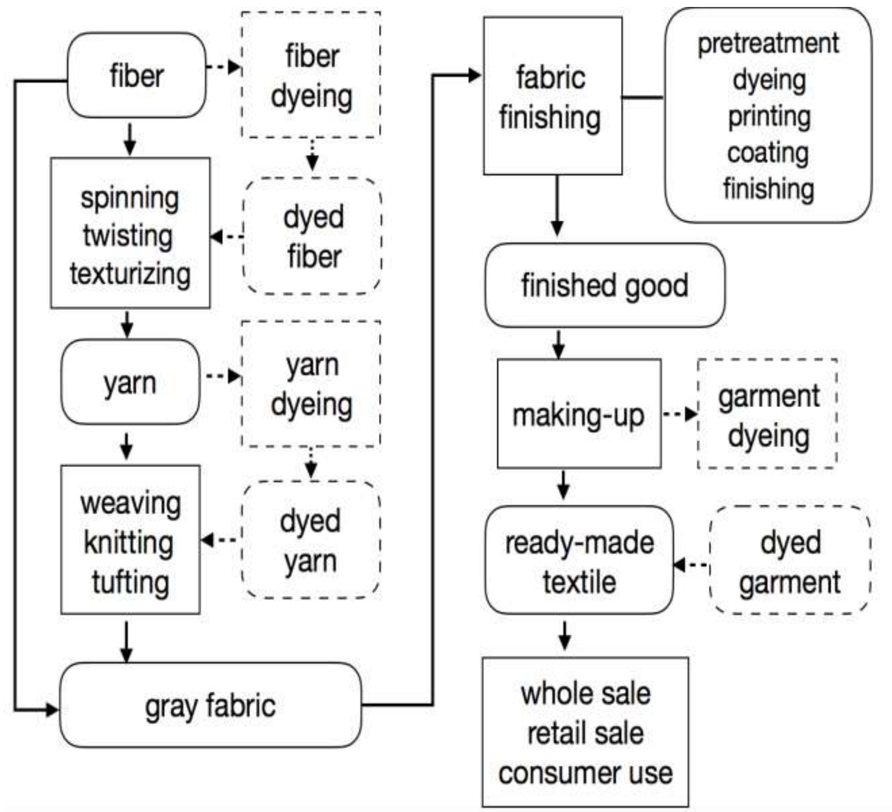

Fig. 1. The textile chain

In Industry 4.0 generation, lots of traditional manufacturing factories focus on importing the key technologies, including cyber-physical systems [2], Internet of Things [3], cloud computing [4], and cognitive computing [5]. Combining big data analysis and domain knowledge from field experts, these technologies bring in a new concept of the whole value chain which is customized and data-driven. In textile industry, the best strategy for the whole supply chain is to import the intelligent integration systems for analyzing global consumer trends, optimizing of manufacturing process by consumer demand-led, and achieving the goal of shortened delivery time and higher textile quality. The common way is collecting data into the Enterprise Resource Planning (ERP) [6] system and then uploading the data into the cloud platform for statistical analysis. However, current approaches only provide statistical data of previous operation parameters from known yarn type. Dealing with new type of yarn, technician could only set machines by trial and error. It is not only time-consuming and resource-wasting, but also impossible to deliver the tech skills to a newbie technician. To avoid the tech skills gap and increase the efficiency of the textile manufacturing process, this paper design an OPRS. It could help technician to set operation parameters more precisely when they get a new type of yarn.

The structure of this paper is organized as follows: According to the related works in Section II, this paper proposed a data-driven machine learning approach in Section III. The experimental results were shown in Section IV. Section V shows our OPRS design. Eventually, Section VI gives the discussions and the conclusion.

II. RELATED WORKS

The textile production manufacturing process involves four stages which are warping, sizing, beaming, and weaving. The success of the full textile operation is considerably influenced by the quality of yarn and the care taken during the preparatory weaving processes, such as warping and sizing.

A. *Warp*

Warp is the process of preparing a double flanged beam of warp yarns arranged parallel to each other. The objective of warping process is to convert the yarn packages into beam having desired width and containing requisite number of ends. Figure 2 shows the schematic side view of indirect warping process [7].

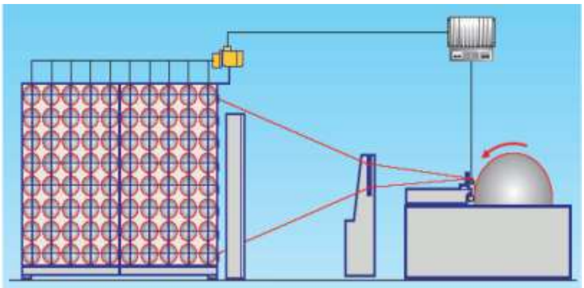

Fig. 2. The schematic side view of indirect warping process [7]

B. Sizing

The objective of sizing is to improve weavability of warp yarn. In order to do that, the sizing machine (shown in Figure 3) coats yarn surface with a suitable film forming polymeric material and penetrates the binding agent into the core of the yarn.

Fig. 3. Sizing machine

C. Weaving

Weaving is the most popular way of fabric manufacturing which is done by interlacing two orthogonal sets (warp and weft) of yarns in a regular and recurring pattern [8].

However, the important parameters depend on different processes in different textiles. For example, warping speed is important to warp process [9] but for weaving process, tension becomes a major parameter [10]. Breakage of yarns occurs if the tension is too high. In contrast, warp yarns attend to jam and break if the tension is too low. Therefore, the course of the tension during weaving is essential in order to perfectly set up a weaving machine.

According to the different raw material composition, the operation parameters are not the same. For example, there is an important effect of warp tension on fabric resistance (mechanical properties) like friction resistance, vertical tensile resistance, and horizontal tensile resistance [11]. In some cases, warp tension must be in high value, but in other cases, low warp tension is more appropriate. It depends on fabric variables like weft and warp density, weft and warp type, weft and warp count, and finally weave structure. Therefore, according to different yarn specifications, finding the appropriate value of the operation parameters of the machine is very important.

Another key factor of competitiveness in the modern textile industry is the quality of production. The reason is that while the unqualified textile production was detected, this production must be discarded. What's worse is that the whole manufacturing process has to redo in order to achieve a satisfactory quality of production. It will be time-consuming and resource-wasting.

Therefore, avoiding producing poor-quality textile production in advance is very important. To achieve this goal, a lot of research is related to the textile quality. Such as, [12] proposed a system to predict the quality of garment appearance from fabric mechanical properties. [13] implemented an application of intelligent control systems which predict a spun yarn quality by fiber properties, namely fiber strength. [14] studied on the correlation between fabric mechanical properties and the quality of seam appearance.

In addition to the fiber properties, setting the operation parameters of textile production manufacturing process would also significantly affect the quality of production. However, lacks of the studies on predicting the quality of textile from machine operation parameters of the manufacturing process.

This paper reports a study on predicting the quality level by classification machine learning methods from the operation parameters of the manufacturing process. Previous works indicate that the textile industry shows great interest in operation parameter prediction and textile quality prediction. With increasing computation power, machine learning approaches have extended the group of statistical interpolation techniques for industrial datasets during the last couple of years. In the following paragraphs, this paper reviews two types of machine learning approaches, which is aimed to solve two problems mentioned above respectively.

Regression machine learning algorithm has been widely used for predicting real-valued output problems. For instance, linear regression which has been used to measure the relationship between the variables. The evaluation of the correlation coefficients can be used to predict the value of the dependent variable from the independent variables by applying some form of historical data [15, 16, 17]. [18] conducted experiment to predict crop yield using regression analysis on agriculture in India by computing the linear regression model. In order to enhance the prediction accuracy and interpretability, LASSO presented by [19], which is an innovative variable selection method by minimizing the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. It is a well-known regression method which regularizes the parameter under sparse assumption and performs both variable selection and regularization [20] presented the LASSO-based fuel consumption prediction model to improve the calculation of the fuel consumption. Other regression method which was widely used such as Ridge and Elastic Net will be told in the next section.

From the previous paragraph, this paper reports the related works of regression algorithm. In the following, this paper reports the related works of classification algorithm.

Scholars conducted a lot of in-depth research on classification machine learning methods. In [20], these classification learning algorithms were divided into four categories. Logic-based algorithms, such as decision trees. Perceptron-based techniques, such as neural networks. Statistical learning algorithms, Bayesian networks. And instance-based learning, such as support vector machines. Performance of each algorithm largely depends on the properties of the input data. In this paper, the input data come from the machine operation parameters, which are mostly discrete categorical features. From [20] perspective of view, logic-based algorithms tend to perform better than others in this situation.

Compared with the basic classification methods which learn a single model, ensemble methods try to learn a compound classifier combined by combining a set of classifiers. Combining a set of models makes the ensemble methods more flexible [21, 22, 23]. According to the different ways of combining classifiers, ensemble methods were usually divided into two types, bagging ensemble methods and boosting ensemble methods.

Leo Breiman proposed a method “bagging” [24], which came from “bootstrap aggregating”. It is an averaging ensemble approach, which combined based classifier with equal weight. Random forest classifier is one of the classic bagging ensemble methods which combine a set of decision tree classifiers. Figure 4 shows the general architect of random forest. This method involved three steps. The first step is to randomly divide original data into k data sets. Then, k-decision trees would be built with the k data-sets. Finally, the final classification result would be decided from combining the results of each decision tree.

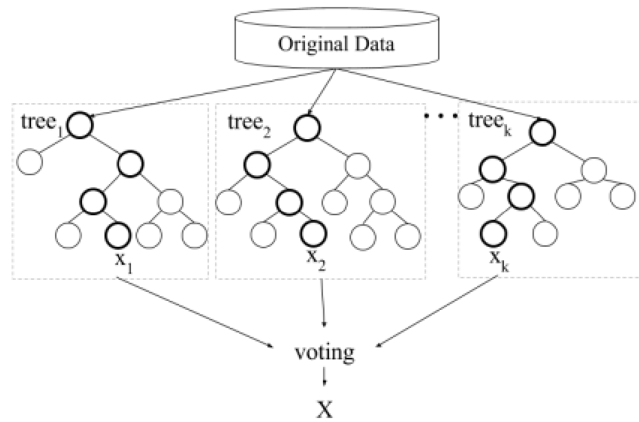

Fig. 4. General architect of random forest

Another type of ensemble methods, boosting methods, put emphasis on improving the classification accuracy of mis-classified data from previous classifier in each learning iteration. Generally, boosting methods performance better than bagging methods by training new classifiers with higher weighted mis-classified data [25, 26, 27]. There are two classic boosting methods, adaptive boosting and gradient boosting method, which weight mis-classified data differently. Adaptive boosting is usually abbreviated to AdaBoost [28]. Figure 5 shows the brief concept of AdaBoost algorithm. In the first step, AdaBoost tries to train a classifier with equally weighted data. From the next iteration, the data mis-classified by the previous classifier would be given a higher weight. AdaBoost would try to train a classifier which put more emphasis on the data with higher weight. In each iteration, AdaBoost added new classifier into the previous classifier to create a stronger classifier.

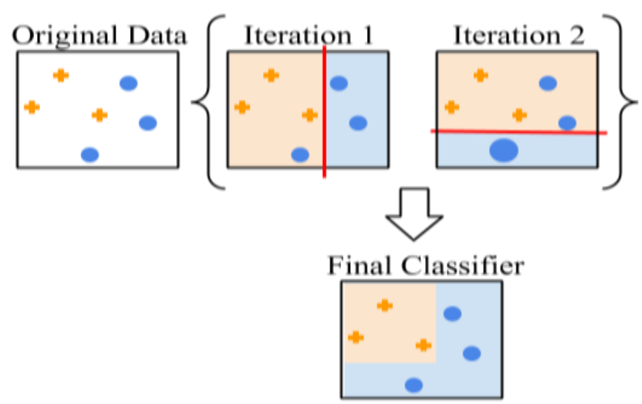

Fig. 5. Brief concept of AdaBoost

Comparing with AdaBoost which is the linear combination of each weak classifier, gradient boosting is additive training. Each iteration, gradient boosting algorithm optimized the new classifier by adding a based classifier which minimizes the loss function [29]. Recently, eXtreme Gradient Boosting, short for XGBoost [30], won many machine learning competitions. It is a new algorithm implemented under the Gradient Boosting framework [31]. The main advantages of this algorithm are high efficiency and good scalability. It avoids overfitting and improving accuracy by supporting many types of objective functions.

Considering the machine learning methods we mentioned above and the data we collected from Li Peng factory [32]. In order to predict operation parameters based on different yarn specifications and historical operating records, the alternative modelling of the four process in textile industry will be presented by four different regression methods. Moreover, five classification methods were used to predict the quality of textile based on the operation parameters predicted by regression models.

The next section shows the methodology proposed by this paper to train the regression and classification models by the algorithms mentioned above.

III. METHOD AND MATERIALS

The object of this paper is to improve the operation parameters setting process. In order to solve the real problem in operation parameters setting, this paper designs an OPRS. To build OPRS, this paper need to get the data from ERP [6] system of Li Peng factory and train two prediction models. One of the models is regression model, which aimed to predict operation parameters by yarn properties. Another is classification model, which is going to predict the quality of production by the operation parameters. Combining these two models, OPRS intellectualizes the way of setting operation parameters, instead of using traditional trial and error method. The architect of OPRS is shown in Figure 6.

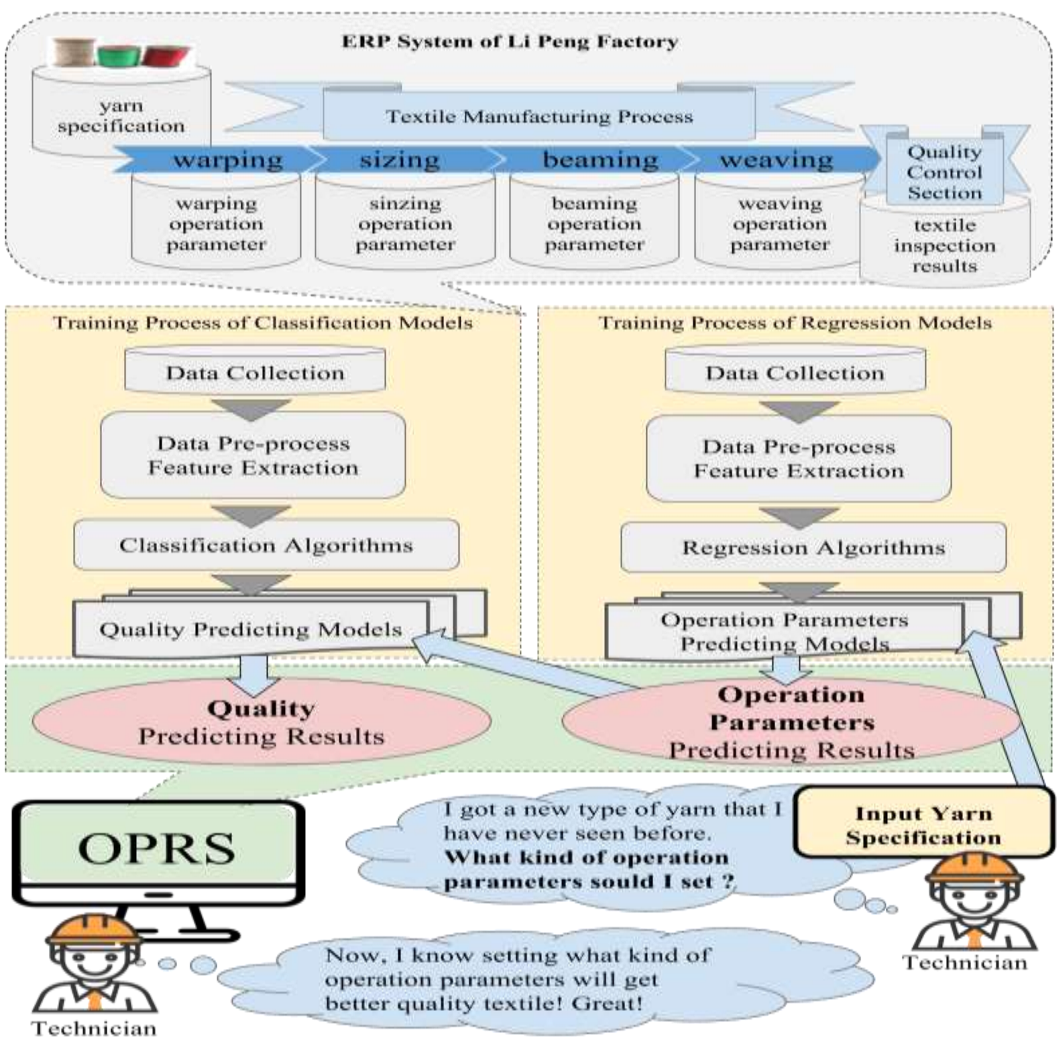

Fig. 6. Architect of OPRS

A. *Training Process and Validation of Regression Models*

The process of predicting machine operation parameter based on different yarn properties by regression models is listed as follows:

1. Step 1 Data Collection
2. Step 2 Data Preprocess
3. Step 3 Feature Extraction
4. Step 4 Model Training
5. Step 5 Cross Validation & Evaluation

1) Data Collection: All the data were collected from ERP [6] system of Li Peng factory, from April 2016 to August 2017. The overview of four process raw data is listed in Table 1.

TABLE 1
RAW DATA FROM THE ERP SYSTEM OF LI PENG FACTORY

File name (.csv)	Size(MB)	Number of column	Number of row	Description
Warpop	7.1	37	26103	Operation parameters in warping process
Sizeop	5	46	15950	Operation parameters in sizing process
Beamop	8.6	31	37089	Operation parameters in beaming process
Weaveop	50.9	25	161821	Operation parameters in weaving process

2) Data Preprocess: The problem is complicated by the fact that the databases in Li Peng factory are highly susceptible to noise, missing, and inconsistent data because of the multiple sources which are different sensors in Li Peng factory. So basically, the question to be addressed first is how we can provide accurate and clean operation parameters information from massive raw data which we retrieve from the sensor in Li Peng factory. There are several data preprocessing techniques [33] listed on Figure 7.

Data cleaning can be applied to remove noise and correct inconsistencies in data, shown in Figure 7-a. Data integration merges data from multiple sources into a coherent data store, such as a data warehouse, shown in Figure 7-b. Data reduction, shown in Figure 7-c, can reduce data size by, for example, aggregating, eliminating redundant features, or clustering.

However, variables tend to have ranges that vary greatly from each other. Therefore, data miners often normalize their numeric variables, in order to standardize the scale of effect each variable has on the results [34]. This process is called data transformation, shown in Figure Figure 7-d.

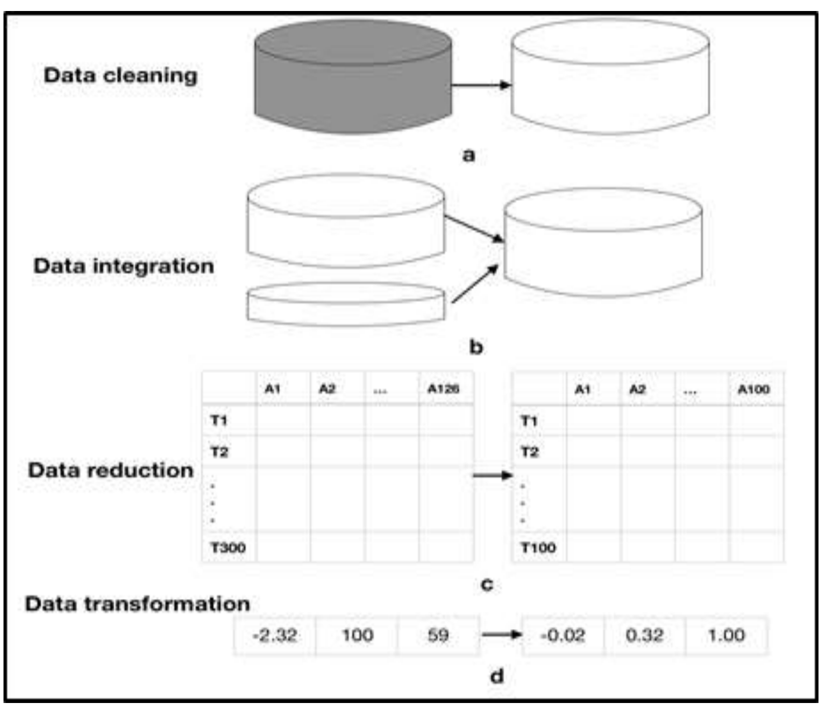

Fig. 7. Forms of data preprocessing

Applying data processing techniques before mining can substantially improve the overall quality of the patterns mined and/or the time required for the actual mining [33]. For our training data shown in Table 1, there is a special feature we need to transform into two different features; more specifically, YARNSPEC, which represents yarn specification, contained two valuable information such as denim and fiber. Therefore, regular expression was applied in this paper with defined pattern to extract those two important features.

For the four processes involved, features and the operation parameters we want to predict as label are listed in Table 2 and Table 3.

TABLE 2
FEATURES IN FOUR TEXTILE PROCESSES

Warping Process	Sizing Process	Beaming Process	Weaving Process
WEAVELISTNO	WEAVELISTNO	WEAVELISTNO	WEAVELISTNO
WARPTOTAL	WARPTOTAL	WARPTOTAL	YARNSPECDENIM
TOTALLENGTH	YARNSPECDENIM	YARNSPECDENIM	YARNSPECFIBERBASE
THEORYLENGTH	YARNSPECFIBERBASE	YARNSPECFIBERBASE	DENIM
YARNSPECDENIM	DENIM	DENIM	FIBERBASE
YARNSPECFIBERBASE	FIBERBASE	FIBERBASE	WEFTDENSITY
DENIM	WARPSTRIP		ROLLSERIALNO
FIBERBASE	WARPLENGTH		
UNITWEIGHT	SIZINGLENGTH		
GRANULARITY	SIZINGCOD		
WARPLENGTH			

TABLE 3
LABELS IN FOUR TEXTILE PROCESSES

Warping Process	Sizing Process	Beaming Process	Weaving Process
WARPSPEED	SIZINGSPEED	BEAMSPEED	WEAVEBTENSION
WARPPRES	SIZINGBPRES	BEAMATENSION	
SSTENSION	SIZINGATENSION	BEAMBTENSION	
WARPTENSION	SIZINGBTENSION	BEAMTENSION	
HYDRATENSION	CONSISTENCY		
	DENSITY		

Column Name	First Value	Value Type	Mean	Standard Deviation	Chart
YARNSPECDENIM	150	int64	187.52665418735248	92.34680836323687	
YARNSPECFIBERBASE	96	int64	131.79181248473998	135.89857090436692	
DENIM	155	int64	193.4753804834378	94.21880038490534	
FIBERBASE	10680	int64	3606.9746073085375	4838.099121443515	
WEFTDENSITY	86	int64	58.67209245544071	16.32947603082464	
ROLLSERIALNO	1	int64	2.4757060307642225	4.669340726167304	
MEASUREWHEEL	119	int64	109.35696264344429	5.948093082444352	
BEAMLENGTH	200	int64	3008.8665255961587	1472.7558832035866	
WEAVEBTENSION	397	int64	331.5845202246277	165.0205587002083	

Fig. 8. Data preview on InAnalysis

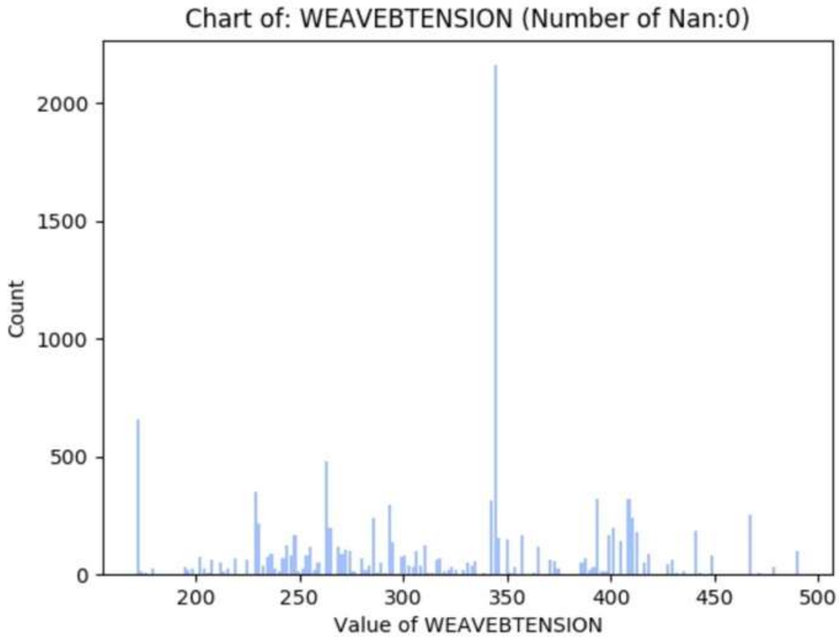

Fig. 9. Show the chart of WEAVEBTENSION on InAnalysis

Take weaving process as an example. Figures 8 and Figure 9 show the preview of weaveop.csv and the bar chart of label WEAVEBTENSION on our InAnalysis system [35]. InAnalysis contains many pre-designed tools for data mining, such as data pre-processing, feature selection, model training, model prediction, and evaluation. It is well-suited for developing and combining schemes for different machine learning applications.

3) Feature Selection: To solve the real problem, it is important to gain a deeper understanding of the actual field. Therefore, we have an interview with senior managers in Li Peng factory. From their previous experience, we know that to set the operation parameters in different processes need to consider different features. For example, there are two key factors, denim and fiber base, in the sizing process. Denim affects the thickness and the toughness of the fiber. Fiber base means the type of fiber, which represents different characteristics. According to the experience of senior managers, this paper chooses these key features for each process to train the operation parameters regression models.

4) Regression Model Training: Regression analysis has become one of the most widely used statistical tools for analyzing multi-factor data. It is appealing because it provides a conceptually simple method for investigating functional relationships among variables. In order to find optimized setting parameters for unknown articles, it is generally necessary to conduct experiments within the historical operation data. The regression algorithms we use in this paper are listed in Table 4.

TABLE 4
FORMULA OF REGRESSION ALGORITHM

Algorithm	Formula
Linear	$\hat{y} = \widehat{\beta}_j + \sum_{j=1}^{p} x_j \hat{\beta}_j$
Lasso	$(\hat{\alpha}_j) + (\hat{\beta}_j) = argmin\ \{\sum_{j=1}^{n}(y_i - \alpha - \sum \beta_j x_j)^2\}$ $subject\ to\ \sum_j \|\beta_j\| \leq t$
Ridge	Ridge regression solves the multicollinearity problem through shrinkage parameter λ by: $argmin + \|\|y - X_\beta\|\|_2^2 + \lambda \|\|\beta\|\|_2^2$
Elastic Net	Elastic Net is hybrid of Lasso and Ridge Regression techniques. It is trained with L1 and L2 prior as regularized. $\widehat{\beta} = argmin\ (\|\|y - X_\beta\|\|^2 + \lambda_2 \|\|\beta\|\|^2 + \lambda_1 \|\|\beta\|\|_1)$

5) Cross-Validation and Evaluation Model: In order to evaluate the regression models we trained, k-fold cross validation was applied in the training process to find the performance of regression. At every turn, data are divided into k subsets. Training set is formed by k-1 subsets of the k subsets and rest of the one subset is used in testing set. Compute MSE, takes the distances from the points to the regression line (these distances are the “errors”) and squares them. Therefore, the smaller MSE means the closer a regression line is to a set of points, which represents a better performance of regression model. After calculated across k turns, we will get the performance of the regression models as the average of scores. The results are shown in the next section.

B. Training and Validating Quality Classifiers

The goal of the textile quality classifiers is to predict the quality of production by the operation parameters setting. The whole process of training textile quality classifiers is listed as follows:

1. Step 1 Data Collection
2. Step 2 Data Preprocess
3. Step 3 Feature Extraction
4. Step 4 Model Training
5. Step 5 Cross Validation & Evaluation

The first step is to collect the training data and target data from the ERP [30] system of Li Peng factory. For the training data, this paper uses operation parameters of warping, sizing, beaming, and weaving process. We got 13157 records of warping process, 8266 records of sizing process, 15891 records of beaming process, and 15891 records of weaving process. For the testing data, we got 21164 records of textile inspection results from the quality control section. The quality of training data greatly affects the performance of the classifier. To avoid “Garbage in, Garbage out”, we will do the following data pre-process process to clean the raw data we got from ERP system. First of all, we removed the records with missing value to prevent training error. The next step is to combine these records from different resources. These records are all index by “WEAVE_LIST_NO”, which represents the ID of each textile product. Therefore, we used this number to combine the data from four manufacturing processes with the inspection results.

Because the variety of fabrics are available today, mixture of two or more types of fabrics is a good way to produce textiles with different characteristics. According to the set of different yarn specifications, more than one operation parameters sets have the same “WEAVE_LIST_NO”, which means more than one types of fabrics were used in one textile product. To keep these characteristics of the data, we created four additional features by calculating maximum, minimum, average and standard deviation of each parameter set which has the same “WEAVE_LIST_NO”.

In total, the training data contain 64 features, which are 16 features used in regression model with each feature having 4 values, maximum, minimum, average, and standard deviation

This paper uses the following supervised classification algorithms to train textile quality classifiers: Basic classification method: Decision Tree Bagging ensemble method: Random Forest Boosting ensemble method: AdaBoost, Gradient Boosting, XGBoost

In this paper, k-fold cross-validation is used in the training process to find the performance of classifier. At every turn, data are divided into k subsets. Training set is formed by k-1 subsets of the k subsets and rest of the one subset is used in the testing set. Compute the accuracy, which means the percentage of the data classified correctly over the total data. After calculated across k turns, we will get the performance of the classifier as the average of k accuracy.

IV. EXPERIMENTAL RESULTS

A. Results of Regression Models Training

Each regression model was evaluated by MSE which is computed by each parameter in process and K-cross validation. The results are shown in Tables 5, Tables 6, Tables 7, Tables 8.

TABLE 5
PERFORMANCE OF REGRESSION MODELS PREDICTING OPERATION PARAMETERS IN WARPING PROCESS

Algorithm	MSE of WARPSPEED	MSE of WARPPRES	MSE of SSTENSION	MSE of WARPTENSION	MSE of HYDRATENSION	10-fold cross-validation
Linear regression	7559.832	0.498	1.616	5.572	3.765	10127.988
Lasso regression	7559.843	0.509	1.626	5.582	3.776	9095.0066
Ridge regression	7559.832	0.498	1.616	5.571	3.765	10127.986
Elastic Net regression	7560.915	0.5112	1.630	5.595	3.791	10095.11

TABLE 6
PERFORMANCE OF REGRESSION MODELS PREDICTING OPERATION PARAMETERS IN SIZING PROCESS

Algorithm	MSE of SIZINGSPEED	MSE of SIZINGBPRES	MSE of SIZINGATENSION	MSE of SIZINGBTENSION	MSE of SIZINGBTENSION	MSE of DENSITY	10-fold cross-validation
Linear regression	4659.575	0.706	6.537	7.489	7.429	13.150	7.719
Lasso regression	4659.679	0.760	6.612	7.565	7.504	13.227	7.779
Ridge regression	4659.575	0.706	6.537	7.489	7.429	13.151	7.718
Elastic Net regression	4667.077	0.761	6.727	7.685	7.659	13.512	7.789

TABLE 7
PERFORMANCE OF REGRESSION MODELS PREDICTING OPERATION PARAMETERS INBEAMING PROCESS

Algorithm	MSE of BEAMSPEED	MSE of BEAMATENSION	MSE of BEAMBTENSION	MSE of BEAMTENSION	10-fold cross-validation
Linear regression	346.302	30660.307	41672.999	8195.679	31538.663
Lasso regression	346.302	30660.307	41672.999	8195.679	31066.511
Ridge regression	346.302	30660.307	41672.999	8195.679	31538.666
Elastic Net regression	346.306	30660.307	41673	8195.681	31538.228

TABLE 8
PERFORMANCE OF REGRESSION MODELS PREDICTING OPERATION PARAMETERS IN WEAVING PROCESS

Algorithm	MSE of WEAVEBTENSION	10-fold cross- validation
Linear regression	3.499	2.872
Lasso regression	0.000396	0.000353
Ridge regression	6.435	3.145
Elastic Net regression	3.697	0.00010375

The result showed that the operation parameter in weave process is more suitable to using regression model than other process, which MSE achieves 0.01% by ten-fold cross-validation. Figure 10 shows selecting models and setting k value to compare accuracy with k-fold cross-validation on InAnalysis. And Figure 11 shows that the system presents the corresponding results.

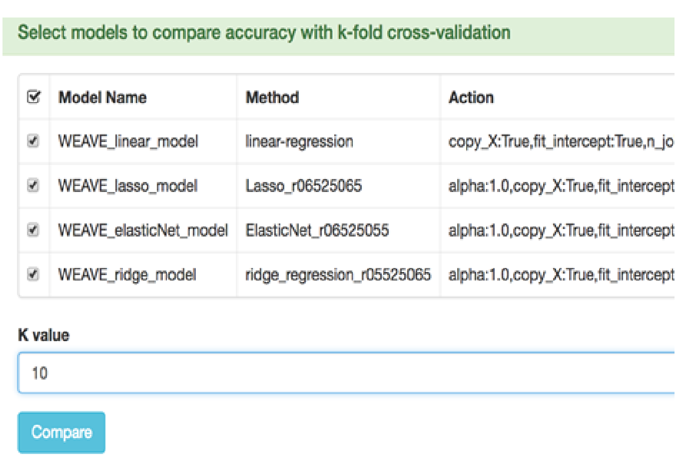

Fig. 10. Selecting models and set k value on InAnalysis before k-fold cross validation

1) Feature Selection: However, the performance from the same algorithm can vary slightly from the different parameters. For example, the weave process performs best in Lasso Regression model with alpha = 1.0. It can reduce MSE to 0.01%, shown in Table 9. The historical data of weaving operation parameters were shown in Figure 11.

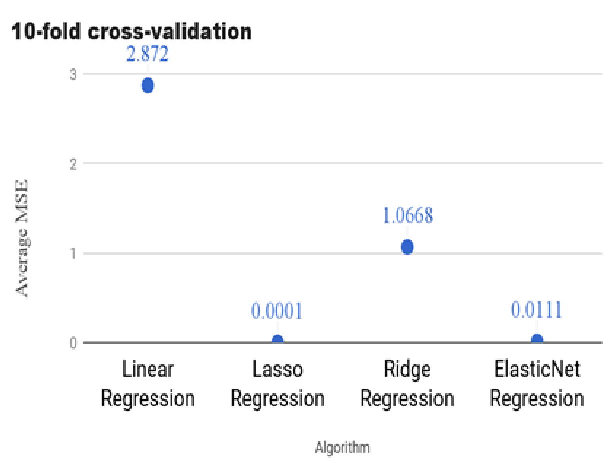

Fig. 11. Average MSE of four regression algorithms on InAnalysis

TABLE 9
PERFORMANCE VARIES FROM THE DIFFERENT ALPHA

Alpha	*MSE (10-fold cross-validation)*
0.1	0.000578
0.5	2.59508
1.0	0.0001113

B. *Results of Classification Models training*

The goal of this part is to find the best classification model by comparing the average prediction accuracy of eight different algorithms with k-fold cross-validation testing design. Figure 13 shows the models we want to evaluate by k-fold cross-validation, deploying the value of k by InAnalysis. The following Figure 13 shows the average prediction accuracy with 10-fold cross-validation provided by each model. It is obvious that the highest values of average prediction accuracy were provided by XGBoost, followed by AdaBoost, Gradient Boost, Random Forest, and Decision Tree, respectively.

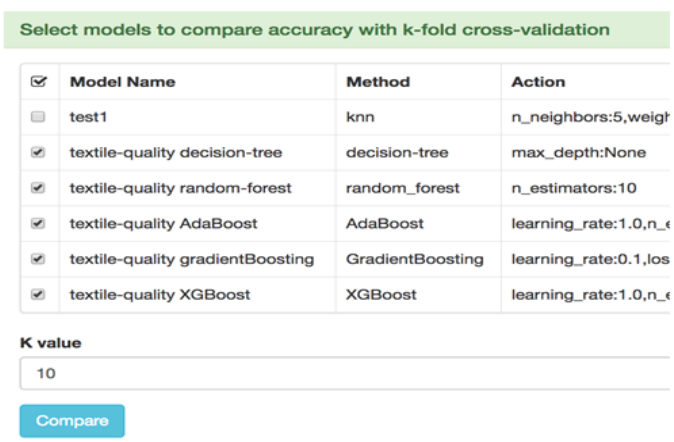

Fig. 12. Selecting models and setting k value on InAnalysis before k-fold cross-validation

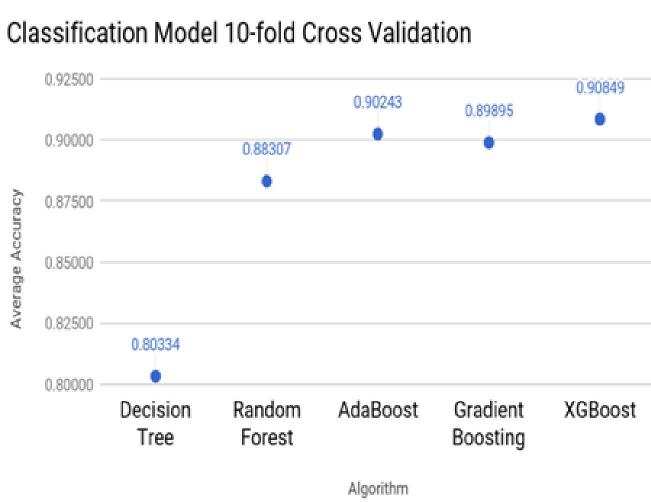

Fig. 13. Average accuracy of five classification algorithms on InAnalysis

C. *Summarization*

Based on the aforementioned experimental results and the previous discussion, the optimum set of predicting models can be summarized in the following: Using Lasso regression models for predicting each parameter performed better than other models and using XGBoost algorithm to train a classification model for predicting textile quality from parameters set.

V. TEXTILE OPERATION PARAMETERS RECOMMENDATION SYSTEM

By combining the two models above, we designed an operation OPRS. When the technician entered the fiber properties, such as denim and fiber base, our OPRS can provide a set of operation parameters in textile manufacturing process and the quality level prediction of this set of parameters. It provides good performance when comparing with traditional stochastic methods.

In today's competitive business environment, companies and factories are facing challenges in dealing with big data issues for improved productivity for the lack of smart analytic tools. With this issue in mind, this paper proposed the analytic system, OPRS, which supports technician setting operation parameters more precisely even for a new type of yarn. Eventually, it will help to fix the tech skills gap in the textile manufacturing process and reduce the cost by optimized machine operation parameters and maintenance scheduling.

VI. CONCLUSION AND RECOMMENDATIONS

It can be concluded that this paper intellectualizes the way of setting operation parameters on four stages, warping, sizing, beaming, and weaving. Instead of using traditional trial and error method, we imported the intelligent data-driven approach of setting operation parameters by our OPRS which integrated the historical data from ERP system and built models with machine learning algorithms. To achieve the goal of optimizing whole textile manufacturing process, this paper could be expanding into other textile manufacturing process, such as preparation of yarn, spinning, and finishing, based on the method built in this paper.

To summarize, the cyber-physical system and decision support system is a trend of the smart manufacturing and industrial big data environment. Industry 4.0 proposes the predictive manufacturing in the textile industry. Our OPRS could help textile factories in increasing their capabilities of self-awareness, self-prediction, and self-maintenance in Industry 4.0.

Declaration of Conflicting Interests

The authors make the declaration that no competing interests.

Acknowledgments

The work was partially supported by MOST, Taiwan, under grants 105-2410-H-002-099-MY3 and 107-3113-E-002 -011 -CC2.

REFERENCES

[1] A. Hasanbeigi, “Energy-efficiency improvement opportunities for the textile industry,” Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA, Tech. Rep., 2010.

[2] S. Wang, J. Wan, D. Li, and C. Zhang, “Implementing smart factory of industrie 4.0: An outlook,” *International Journal of Distributed Sensor Networks*, vol. 12, no. 1, pp. 158–168, 2016.

[3] A. V. Prasad, & Krishna, *Exploring the Convergence of Big Data and the Internet of Things*. Hershey, PA: IGI Global, 2017.

[4] R. Buyya, J. Broberg, and A. M. Goscinski, *Cloud computing: Principles and paradigms*. Hoboken, NJ: John Wiley & Sons, 2010.

[5] K. Hwang and M. Chen, *Big-Data Analytics for Cloud, IoT and Cognitive Computing*. Hoboken, NJ: John Wiley & Sons, 2017.

[6] T. M. S. Arik Ragowsky, “Enterprise resource planning,” *Journal of Management Information Systems*, vol. 19, no. 1, pp. 11–15, 2002. doi: https://doi.org/10.1080/07421222.2002.11045718

[7] M. Dorgham, “Warping parameters influence on warp yarns properties: Part 2: Warp yarn material and cone position on warping creel,” Ph.D. dissertation, Weaving and Knitting Department Faculty of Applied Arts, Helwan University, Cairo, Egypt, 2014.

[8] NPTEL, “Introduction to fabric manufacturing,” 2017. [Online]. Available: https://bit.ly/2BekVLK

[9] M. Dorgham, “Warping parameters influence on warp yarns properties: Part 2: Warp yarn material and cone position on warping creel,” *Journal of Textile Science & Engineering*, vol. 4, no. 5, pp. 164–170, 2014. doi: https://doi.org/10.4172/2165-8064.1000132

[10] Y. Gloy, W. Renkens, M. Herty, and T. Gries, “Simulation and optimisation of warp tension in the weaving process,” *Journal of Textile Science & Engineering*, vol. 5, no. 1, pp. 179–186, 2015. doi: https://doi.org/10.4172/2165-8064.1000179

[11] A. Karnoub, N. Kadi, Z. Azari, and E. Bakeer, “Find the suitable warp tension to get the best resistance for jacquard fabric,” *Journal of Textile Science & Engineering*, vol. 5, no. 6, pp. 222–232, 2015. doi: https://doi.org/10.4172/2165-8064.1000222

[12] J. Geršak, “Development of the system for qualitative prediction of garments appearance quality,” *International Journal of Clothing Science and Technology*, vol. 14, no. 3/4, pp. 169–180, 2002. doi: https://doi.org/10.1108/09556220210437149

[13] S. Yanık, C. Kahraman, and H. Yılmaz, “Intelligent process control using control chartsii: Control charts for attributes,” in *Intelligent Decision Making in Quality Management*. Springer, 2016, pp. 71–100.

[14] D. Z. Pavlinić, J. Geršak, J. Demšar, and I. Bratko, “Predicting seam appearance quality,” *Textile Research Journal*, vol. 76, no. 3, pp. 235–242, 2006. doi: https://doi.org/10.1177/0040517506061533

[15] B. Heshmaty and A. Kandel, “Fuzzy linear regression and its applications to forecasting in uncertain environment,” *Fuzzy Sets and Systems*, vol. 15, no. 2, pp. 159–191, 1985. doi: https://doi.org/10.1016/0165-0114(85)90044-2

[16] B. Savkovic, P. Kovac, I. Mankova, M. Gostimirovic, K. Rokosz, and D. Rodic, “Surface roughness modeling of semi solid aluminum milling by fuzzy logic,” *Journal of Advances in Technology and Engineering Studies*, vol. 3, no. 2, pp. 51–63, 2017. doi: https://doi.org/10.20474/jater-3.2.2

[17] F. Gongor, O. Tutsoy, and S. Colak, “Development and implementation of a sit-to-stand motion algorithm for humanoid robots,” *Journal of Advances in Technology and Engineering Research*, vol. 3, no. 6, pp. 245–256, 2017. doi: https://doi.org/10.20474/jater-3.6.4

[18] V. Sellam and E. Poovammal, “Prediction of crop yield using regression analysis,” *Indian Journal of Science and Technology*, vol. 9, no. 38, pp. 1–5, 2016. doi: https://doi.org/10.17485/ijst/2016/v9i38/91714

[19] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[20] S. Wang, B. Ji, J. Zhao, W. Liu, and T. Xu, “Predicting ship fuel consumption based on LASSO regression,” *Transportation Research Part D: Transport and Environment (In Press)*, 2017. doi: https://doi.org/10.1016/j.trd.2017.09.014

[21] J. Macek, “Incremental learning of ensemble classifiers on ECG data,” in *18th IEEE Symposium on Computer-Based Medical Systems (CBMS’05)*.

Dublin, Ireland: IEEE, 2005. doi: https://doi.org/10.1109/cbms.2005.69 pp. 315–320.

[22] P. Fergus, M. Selvaraj, and C. Chalmers, "Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using cardiotocography traces," *Computers in Biology and Medicine*, vol. 93, pp. 7–16, 2018. doi: https://doi.org/10.1016/j.compbiomed.2017.12.002

[23] F. Herrera, F. Charte, A. J. Rivera, and M. J. del Jesus, "Ensemble-based classifiers," in *Multilabel Classification*, F. Herrera, Ed. Cham, Switzerland: Springer, 2016, pp. 101–113.

[24] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996. doi: https://doi.org/10.1007/bf00058655

[25] I. I. Baskin, G. Marcou, D. Horvath, and A. Varnek, "Bagging and boosting of classification models," in *Tutorials in Chemoinformatics*, A. Varnek, Ed. Hoboken, NJ: Wiley, 2017, pp. 241–247.

[26] G. Ditzler, J. LaBarck, J. Ritchie, G. Rosen, and R. Polikar, "Extensions to online feature selection using bagging and boosting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 9, pp. 4504 – 4509, 2017. doi: https://doi.org/10.1109/tnnls.2017.2746107

[27] B. Sumana and T. Santhanam, "Optimizing the prediction of bagging and boosting," *Indian Journal of Science and Technology*, vol. 8, no. 35, pp. 1–13, 2015. doi: https://doi.org/10.17485/ijst/2015/v8i35/78449

[28] C. Ying, M. Qi-Guang, L. Jia-Chen, and G. Lin, "Advance and prospects of adaboost algorithm," *Acta Automatica Sinica*, vol. 39, no. 6, pp. 745–758, 2013. doi: https://doi.org/10.3724/sp.j.1004.2013.00745

[29] J. Son, I. Jung, K. Park, and B. Han, "Tracking-by-segmentation with online gradient boosting decision tree," in *Proceedings of the IEEE International Conference on Computer Vision*, Washington, DC, WA, 2015. doi: https://doi.org/10.1109/iccv.2015.350 pp. 3056–3064.

[30] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system." in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* San Francisco, CA: ACM, 2016. doi: https://doi.org/10.1145/2939672.2939785 pp. 785–794.

[31] A. Gómez-Ríos, J. Luengo, and F. Herrera, "A study on the noise label influence in boosting algorithms: AdaBoost, GBM and XGBoost." *International Conference on Hybrid Artificial Intelligence Systems,* La Rioja, Spain: Springer, 2017. doi: https://doi.org/10.1007/978-3-319-59650-1_23 pp. 268–280.

[32] Li Peng Enterprise Co., Ltd., "Libaolong relationship enterprise," 2017. [Online]. Available: https://bit.ly/2MfzFip

[33] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. New York, NY: Elsevier, 2011.

[34] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken, NJ: John Wiley & Sons, 2014.

[35] InAnalysis, "Inanalysis: Data science getting started best tools," 2017. [Online]. Available: https://inanalysis.github.io/