The prediction of product return rates with ensemble machine learning algorithms

DOI : 10.36909/jer.13725

Ayse Nur Adiguzel Tuylu[*] and Ergun Eroglu[**]

*Industrial Engineering Department, Engineering Faculty, Istanbul University-Cerrahpasa, 34320, Istanbul, Turkey

**Department of Quantitative Methods, School of Business, Istanbul University, 34320, Istanbul, Turkey

*Corresponding Author: aysenur.adiguzel@istanbul.edu.tr, ayseadiguzeltuylu@iuc.edu.tr;

ABSTRACT

There may not always be actual data available for planning. Predicted data are used especially for future planning. Due to errors in such planning based on prediction, many products enter the reverse logistics network without completing the shelf life. Especially in textile sector, because of fashion, it is the most important point of planning to be able to make accurate estimates in order to avoid unnecessary resource utilization and to provide minimum cost. It is difficult to establish a mathematical model because the prediction problems in real life have multivariate structure and unknown parameters. Most of the studies in literature have been based on time series prediction. But due to changing fashion and demands of consumers, there are significant differences between demand forecasts and real data. So, in the problems with unknown parameters and multivariate structure, Ensemble Machine Learning (EML) methods are preferred recently because they give more accurate results than other prediction methods.

Unlike other studies, the product return rate in textile sector has been predicted with the Stacking and Vote algorithms from EML methods in this paper. In this direction, it is aimed to concentrate on the returns of the products sold with the preferences of the customers and to predict the returns more accurately. In this way, the consumer information obtained as a result of the analyzes can provide more accurate planning in avoiding unnecessary production,

transportation and storage activities, reducing costs, resource utilization and environmental pollution. In addition, it is one of the main aims of the study to contribute to the literature by determining the parameters that can be used in predicting the return rates.
Highest performance results were obtained with Stacking algorithm. The obtained results were given comparatively and the correlation coefficient of 86.07% was reached.

Key words: Ensemble machine learning; Prediction modelling; Product return rate; Stacking algorithm; Vote algorithm.

INTRODUCTION

Some of the products that enter the reverse logistics network in the retail sector are renewed with added value or product parts are reused, some are kept for resale next season and about half are sold in the secondary retail market. Clearly, reverse logistics is a great asset, and suppliers, customers, partners, service providers, the environment, corporate earnings and shareholders are also somewhat affected by the reverse logistics process, as any company that produces, sends or sells goods to a customer is greatly affected by reverse logistics from all directions (Greve and Davis, 2012). One of the most fundamental requirements of sustainability in the market is to ensure customer satisfaction. It is very difficult to achieve this satisfaction with the ever-changing fashion perception and the developing technology every day. Because of the fashion of a different product, color, model, customer's interest in the product can change in a moment and shift to another product. As a result of this change in customer satisfaction and the mistakes made in planning activities such as sales forecasting, inventory management and logistics, products that have not yet completed their product life cycle are entering the reverse logistics network in order to lose their place in the market and regain value and take their place in the market. The right thing to do is to plan activities properly. More accurate planning can be made with the information obtained from product returns from customers or stores to the center. For example, the activities based on planning as products whose production cannot be predicted accurately, the customer's preferences, the change of these preferences depending on location, the accuracy of sales and marketing planning, the number of products sent to each store and the accuracy of product properties, the results of marketing strategies, etc. can be done more accurately with the analysis of information about the returned products.

The most important point that complicates the problem structure in product returns is uncertainty in demand. The demand forecast is actually the prediction of product returns. Accurate demand forecasting for returned products provides the company with strategic benefits in many key areas such as production, distribution and stock. Recent studies have shown that artificial intelligence and machine learning methods are more accurate than classical prediction methods in large complex data sets.

In the literature, there is a lack of studies to predict the return rates for the retail sector. It is aimed to contribute to the literature in this respect. The purpose of this study is to determine the effect of the sales point and product characteristics on product return by using EML algorithms, and to predict the product return rates from stores to the center more accurately. For this purpose, in order to study on the store and customer return data of a textile company operating around the world during the doctoral thesis study process, information about the parameters of the product and store that affect the product return was collected in consultation with business analysts and store personnel first. Data analysis, trial-error, and parameter selection algorithms have determined input parameters that can be used for the product return model (Adiguzel Tuylu, 2017). For more detailed analysis of product returns, a specific product group has been selected from a large quantity of product pool, with a wide range of product returns, consistent range of returns, not having missing and excessive end-to-end data. For this product group, the number of products sent back to the center of the store and the number of products sent from the center to the store were rated and the product return rates were calculated for all the products on the basis of the stores.

When we separated our data set as 90% training 5% validation and 5% hold-out test, the product return rates were predicted with EML algorithms. High prediction performances were obtained by using Stacking and Vote algorithms. When the correlation coefficient is examined, Stacking (SMOreg(M5P+M5Rules+SMOreg+DecisionTable+Linear Regression)) algorithm has the best prediction performance of 86.07% R.

Toktay (2003) conducted a study to emphasize the importance of estimating the time periods of product returns and the amount of returned product in reverse logistics. Krapp et al. (2013) aims to provide a general estimation approach to predict product returns. Agrawal et al. (2014) aimed to develop a model for the estimation of product returns in terms of quantity and time. Kumar et al. (2014) aimed at estimating the return products in the uncertainties. Temur et al. (2014) has developed a fuzzy expert system to designate the correct predict of the return amount. Zhu et al. (2018) proposed a local algorithm based on random walk to predict product return trend for each customer. Dzyabura (2018) showed that a using gradient boosted regression tree model can accurately predict the return rate using image processing techniques and incorporating image data improved the models' predictive accuracy. Cui et al. (2020) developed data-driven model for predicting return volume at the retailer.

MATERIAL AND METHOD

In this study, the data of the product return for the summer season from the stores to the warehouses of a company operating in the textile sector within the scope of reverse logistics were analyzed. The data which is the input parameters of the product return estimation model were determined by gathering information about the characteristics of the products and stores, affecting the product return by interviewing business analysts and store employees primarily.

Ensemble Machine Learning

Machine learning explores how computers can improve their learning status or performance based on data (Han et al., 2011). Supervised learning is a machine learning approach that is trained using labeled examples, such as an input where the desired output is known. In the supervised learning, the output variable labeled by the supervisor (Ongsulee, 2017). The system tries to find a relationship between the input and output variables to approach the desired output. While a continuous output variable presents a regression problem, a categorical output variable causes a classification problem (Izenman, 2008).

The concept of classification is to distribute the data to the classes in the data set according to their characteristics. The properties and number of these classes are predetermined. The values that specify these classes in the data set are called labels. The classification algorithms analyze the relationships between the class labels in the given training set and the other input features. As a result, it is decided by the help of this model classification model is to maximize the number of correctly assigned samples (Unlu, 2019). Significant improvements in predictive performance can be achieved by averaging the results of individual models or by voting if more than one model is combined into a single classifier (Simmons et al., 2008).

Stacking algorithm is based on the logic of training model that will combine the estimators together. Firstly, different predictors make predictions on the data, whereas meta-classifier/regressor makes the actual prediction with these predictions. The training set is divided into two sets to train the main predictor. The predictors in the first set are trained and then make each prediction for each data in the second set. The main predictor is trained in the training set created by the obtained predictions and the original data. After the main estimator is trained, take the classification predictions and make the final prediction.

The Stacking process is summarized that $\mathcal{L} = \{(y_n, x_n), n = 1, \dots, N\}$: data set; y_n : classification value; x_n: a vector that represents attribute values belonging to the *n*. instance, divided data into subgroups equal to $\mathcal{L}_1, \dots, \mathcal{L}_J$, $\mathcal{L}_J$ and $\mathcal{L}^{(-j)} = \mathcal{L} - \mathcal{L}_j$: testing and training sets for *J*-fold cross-validation; z_{Kn}; Estimate of $M_k^{(-j)}$on $x_n$for each x_n instance in $\mathcal{L}_j$. The K learning algorithms (level 0 models) call the k. algorithm of the rater on the data in the $\mathcal{L}^{(-j)}$ training set when creating an $M_k^{(-j)}$ model for $k = 1, \dots \dots, K$. At the end of the entire cross-validation process, the level1 data set collected from the outputs of the *K* models is calculated as, $\mathcal{L}_{CV} = \{(y_n, z_{1n}, \dots, z_{Kn}), n = 1, \dots, N\}$. From this data set, level 1 learning algorithms are used to derive the model $\widetilde{M}$, which is a function of $(Z_1, \dots \dots, Z_K)$. To complete the training process, M_k models are derived using all the data in $\mathcal{L}$. M_k models produce a vector $(Z_1, \dots \dots, Z_K)$ which is a level 1 $\widetilde{M}$ model that is the result of classification for this example when a new example is given (Erdogan, 2017).

Vote algorithm is the one of the simplest and most common EML methods that combines multiple classifier/regressors with a better classifier/regressor. The classification results $(L^1 \dots L^N)$ obtained by training multiple (N) different classifier/regressors $C^1 \dots C^N$, on the same data set $\mathcal{L}$ are evaluated. The most frequently repeated, that is, the most voting class is chosen as the class that makes up the prediction result, or if the results are in the form of probability values, the average of these results is obtained by the prediction result.

When the ratings of the classifier/regressors are evaluated, each of them can be given equal weight, or different weight values can be given in parallel with the accuracy values obtained through education using all data sets.

Table 1. Performance metrics

Correlation Coefficient (R): A common measure of how well the R curve fits the actual data. A value of 1 indicates a perfect fit between the actual values and the predicted values, meaning that the values have the same tendency.	$R = \frac{n\sum yy' - (\sum y)(\sum y')}{\sqrt{n(\sum y^2) - (\sum y)^2}\sqrt{n(\sum y'^2) - (\sum y')^2}}$
Mean Absolute Error (MAE): The MAE value is an amount used to measure how close predicts are to the final results. It calculates the average size of the errors between the predicted and actual values by ignoring the direction of the errors.	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y - y'\rvert$
Root Mean Square Error (RMSE): RMSE is calculated to find the square error of the predict compared with the actual values and to find the square root of the total value. Thus, it is the average distance of a data point from the fixed line measured along a vertical line. This tool effectively identifies undesirable large differences.	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\sum(y' - y)^2}{n}}$
Relative Absolute Error (RAE): The RAE value is the ratio of the absolute value of the difference between the predicted and actual values to the actual values.	$RAE = \frac{\lvert y'_1 - y_1\rvert + \cdots + \lvert y'_n - y_n\rvert}{\lvert y_1 - \bar{y}\rvert + \cdots + \lvert y_n - \bar{y}\rvert}$
Root Relative Squared Error (RRSE): The RRSE value is the square root of the sum of the squares of the differences between the predicted value and the actual value to the sum of the squares of the differences between the actual values and the mean value.	$RRSE = \sqrt{\frac{(y'_1 - y)^2 + \cdots + (y'_n - y_n)^2}{(y_1 - \bar{y})^2 + \cdots + (y_n - \bar{y})^2}}$

In this study are R, RMSE, MAE, RAE and RRSE were used as *performance metrics* to evaluate the prediction accuracy of the proposed models. Performance metrics are described in Table 1 where y' is the predicted value; y is the real value; and n is the number of data samples (Chou et al., 2015).

Data and Data Preprocess

These parameters were also analyzed by data analysts, trial and error methods and parameter selection algorithms and thus, the parameters that will be used in the model are presented.

According to these parameters, product return data was taken from the system and missing and extreme data were arranged in the obtained data. The aim of this study to determine the parameters affecting the return of the product is to provide contribution to the literature and to introduce a model that can be used in textile sector in general.

However, in order to make a more detailed analysis, "women's trousers" product group data was chosen to be used in our study due to the fact that it appeals to the same segment of the society, reasons for return, number of products, consistency in return rate, minimum of incomplete and extreme data among the data set consisting of 389,478 products. 9106 data were obtained through 3000 return activities from 313 stores to the center for 52 products belonging to the women's trousers product group. The product return rate was calculated by dividing the number of products sent from a store to the center with the number of products arriving at that store.

Input features may cause return activity in textile products are grouped under two headings:

- Store-related features; store name, location, channel, operational region, city, climate group name, store segment.
- Product features: style, color, color family, color type, size, life style, collection group, buyer group, line, fit, waist, price group, initial price group, product return rate.

Table 2. Correlation ranking values

Correlation Ranking	
Values	**Ranked attributes**
0,3028	10 Color Family
0,29061	9 Color Code
0,24051	20 Initial Price Group
0,1849	19 Price Group
0,18152	17 Fit
0,17704	18 Waist
0,17058	13 Life Style
0,16453	16 Line
0,12865	8 Style Code
0,11367	15 Buyer Group
0,11367	14 Collection Group
0,1095	11 Color Type
0,03256	7 Store Segment
0,03046	2 Location Id
0,02905	12 Size
0,01828	5 City Id
0,01662	6 Climate Group Name
0,01391	4 Operational Region Id
0,01126	1 Store Id
0,00934	3 Channel Id

As seen on Table 2, Correlation Attribute Eval method is used as feature ranking algorithm. This method evaluates the worth of a feature by measuring the correlation (Pearson's) between it and the class (Hall et al., 2009).

Modelling

Product return rates were predicted and obtained prediction performances by using Stacking and Vote algorithms from EML techniques. First, the WEKA program is set to use 90% of the data set for training algorithms for learning, 5% for validation and 5% for hold-out test for prediction. We used M5P, M5Rules, Decision Table, Linear Regression and SMOreg methods had been used in our previous study that is Adiguzel Tuylu and Eroglu (2019) and various combinations of these methods as a classifier/regressor for EML methods.

In the case that the data set is set as 90% training - 5% validation and 5% hold-out test, the prediction performances obtained from applied EML techniques are shown as Table 4. The flow chart of the application is shown in Figure 1.

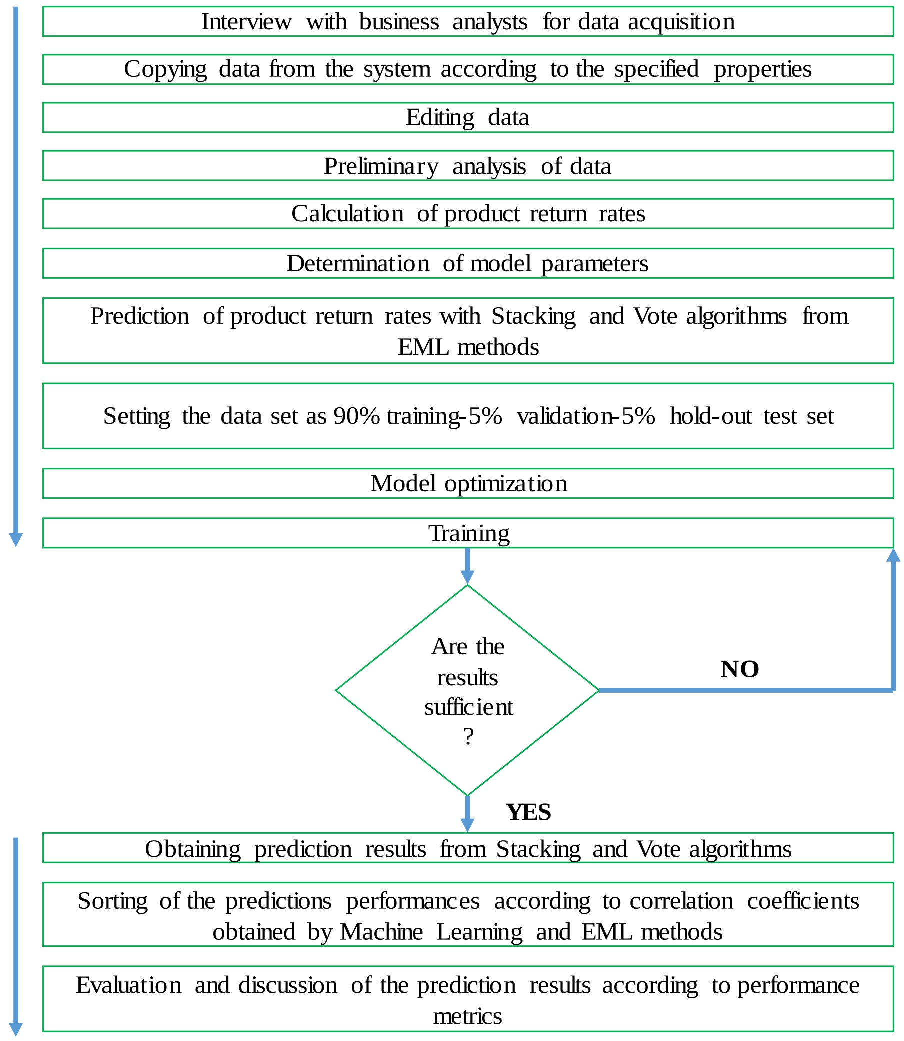

Figure 1. Flow chart

Grid search method used for hyperparameter tuning process. All ML model parameters and their values are shown in Table 3.

Table 3. Hyperparameters of ML algorithms

Model	**Parameters**
M5Rules	Minimum number of instance M: 4.0
M5P	Minimum number of instance M: 4.0
SMOreg	The complexity parameter C: 1, Regression optimizer: RegSMO Improved (Tolerance: 0.001, Seed:1, Epsilon: 1.0E-12, Epsilon parameter: 0.001, Use variant: 1), Poly Kernel (Exponent: 1.0, Cache size: 250007)
Decision Table	Cross validation: 1, Search: Best First (Look up cache size D: 1, Search termination N: 5)
Linear Regression	Ridge: 1.0E-8
Vote	Seed: 1, Combination rule: Average of probabilities
Stacking	Number of folds: 10, Seed: 1, Number of execution slots: 1

RESULTS

We obtained the performances metrics of the machine learning methods in Adiguzel Tuylu and Eroglu (2019), which is our previous study, and compared them according to the correlation coefficient. Our obtained best five algorithms; M5P, M5Rules, Linear Regression, Decision Table, SMOreg. These best performing machine learning algorithms are selected as sub- predictors in EML algorithms.

Table 4. The 10 best algorithms of validation and hold-out test set with respect to the correlation coefficient

THE 10 BEST ALGORITHMS ($R \geq 0.83$)	VALIDATION					HOLD-OUT TEST				
	R	MAE	RMSE	RAE	RRSE	R	MAE	RMSE	RAE	RRSE
Stacking SMOreg (M5P+M5Rules+DT+LR+SMOreg)	0.8521	0.0092	0.0129	43.54%	52.82%	0.8607	0.0094	0.0132	41.71%	52.74%
Stacking M5P (M5P+M5Rules+DT+LR+SMOreg)	0.8446	0.0101	0.013	48.26%	53.47%	0.8522	0.0104	0.0135	46.12%	52.29%
Stacking M5Rules (M5P+M5Rules+DT+LR+SMOreg)	0.8446	0.0101	0.013	48.31%	53.53%	0.8521	0.0104	0.0135	46.22%	52.31%
Stacking LR (M5P+M5Rules+DT+LR+SMOreg)	0.8444	0.0101	0.013	48.22%	53.51%	0.852	0.0104	0.0135	46.04%	52.35%
Stacking M5P (M5P+M5Rules+DT)	0.8471	0.01	0.0129	47.48%	53.10%	0.8516	0.0104	0.0135	46.14%	52.38%
Stacking M5Rules (M5P+M5Rules+DT)	0.847	0.01	0.0129	47.65%	53.11%	0.8514	0.0104	0.0135	45.99%	52.42%
Stacking M5Rules (M5P+M5Rules+SMOreg)	0,8474	0,01	0,0129	47.69%	53.04%	0.8493	0.0105	0.0136	46.22%	52.76%
Stacking M5P (M5P+M5Rules+SMOreg)	0,8473	0,01	0,0129	47.68%	53.06%	0.8493	0.0105	0.0136	42.38%	52.76%
Vote(M5P+M5Rules+LR)	0.8379	0.0106	0.0133	50.38%	54.62%	0.8435	0.0109	0.0139	48.11%	53.93%
Vote(M5P+M5Rules)	0.8479	0.0102	0.0129	48.29%	52.98%	0.8411	0.0107	0.014	47.47%	54.10%

EML algorithms using combinations of these machine learning algorithms as sub-predictors were trained and predicted with 90% training-5% validation- 5% hold-out test sets.

Table 4 shows the highest correlation coefficient and lowest error values of the validation and hold-out test set from the results of the EML Learning algorithms obtained from the study. Stacking SMOreg (M5P+M5Rules+DecisionTable+LinearRegression+SMOreg) algorithm has the best correlation coefficient and error values.

DISCUSSION AND CONCLUSION

Due to the mistakes made in production planning, sales forecasting, transportation, sales policy, inventory planning, packaging and distribution activities, many textile products may not be sold at the end of the season period and enter the reverse logistics network. While these products may not be sold as they cause the use of resources, energy and capital during the advanced logistics phase, they will continue to use both resources and capital consumption as they will cause many activities such as transportation, storage and value gaining when they enter the reverse logistics flow.

In this study EML techniques were used, rather than setting up a mathematical model or estimation of demand based on time series in the literature in general, due to the multi-parameter and multivariate structure of the prediction of the return rate on textile products or the changing demands of the fashion and consumers.

This study focuses on the returns of the products on sale and provides information on the consumer behavior information obtained from analyses to predict the returns in a correct way, and provides information on the correct body, color and model of the products to go to the stores and avoid unnecessary production, transportation and storage activities (stores are not available for sale due to inadequate stock formation, lack of stock, products to be returned to the center, warehouse or outlet stores, costs incurred in the transportation process such as fuel, worker and driver costs, new products to be returned to new products instead of storage in the warehouse, the activities to be carried out for these operations in the warehouse and the costs to be generated by these activities, renewal activities to bring value to the center the necessary costs, strategies and campaign activities for products that can't be sold).

In other words, with the more accurate estimation of product returns, all reverse logistics activities (unnecessary stock formation in stores; products cannot be sold due to excess stock; transport of return product to the center, warehouse or outlet stores, transportation activities as handling, packing and transportation costs as fuel, labor and driver cost; unnecessary using the areas to take place in the warehouse for storage of returned products instead of new products, activities to be carried out for these operations in the warehouse and the costs to be generated by these activities; renewal activities to add value to the products returned to the center and the costs required for this process; strategies and campaign activities for products can't be sold and returned;…) are minimized and thus all these costs and resource consumption are minimized. Thus, with a more accurate product return predict obtained as a result of our work, the company has many advantages in terms of manufacturing strategy determination, vehicle and warehouse capacity studies, vehicle routing, production planning, stock management, supplier selection, end-of-season product strategies, and correctly addressing the customer.

This is the first study to use Stacking and Vote algorithms from EML algorithms to estimate product return rates. It is expected to contribute to the literature and the textile sector with the preliminary study of selecting the parameters that will be input to the product return model by using expert opinions, data analysis, trial and error and parameter selection algorithms. The results show that the EML algorithms used have the ability to predict product return rates in the textile industry.

Prediction was made by using Stacking and Vote algorithms from EML algorithms, with Stacking (SMOreg (M5P+M5Rules+DecisionTable+LinearRegression+SMOreg)) algorithms, the most accurate prediction value was reached with a correlation coefficient of 86.07%. The results of the study show that better results are obtained with EML methods than Machine Learning methods in a way to support the recent studies on this subject in the literature.

When these results were evaluated by the company, the prediction performance was reached with the 70% correlation coefficient with the classical statistical prediction methods previously used by the company, but in this study the prediction performance has been reached 86.07% correlation coefficient with Stacking algorithm. With a predictive performance of 16.07%, the correct prediction of the 389,478 products in one season by Stacking algorithm means that the avoidance of the cost due to unnecessary transportation, storage, loss of value based on non-sale and material cost. The prices of the products vary, but they start at a minimum of 6 $. Even if we accept product prices at the lowest level, 16.07% increase in the prediction performance means the prevention of average cost of 375,534.69 $ in per season. Of course, except for this cost, when cost, labor and time loss caused by activities such as these products to the wrong stores and warehouses shipping, stocking, handling, shelf arrangement, etc. are added, it is clear making a more accurate prediction prevents the loss of millions of $ annually for the firm.

Since satisfactory results are obtained from prediction product return rate in the textile sector with EML methods, prediction product return can be made for other sectors in future studies. In addition to these, Feature engineering is key subject for Machine Learning based reverse logistics so it will be focused on feature engineering concept for future studies.

REFERENCES

Adiguzel Tuylu, A. N. 2017. An Optimization Study in Reverse Logistics Process, (Doctoral Dissertation), Istanbul, 2017.

Adiguzel Tuylu, A. N. & Eroglu, E. 2019. Using Machine Learning Algorithms For Forecasting Rate of Return Product In Reverse Logistics Process. Alphanumeric Journal, 7(1), 143-156.

Agrawal, S., Singh, R.K. & Murtaza, Q. 2014. Forecasting product returns for recycling in Indian electronics industry. J Adv Manag Res., 11(1), 102–14.

Chou, J. S., Ngo, N. T., & Pham A. D. 2015. Shear strength prediction in reinforced concrete deep beams using nature-inspired metaheuristic support vector regression. Journal of Computing in Civil Engineering, 30(1), 04015002.

Cui, H., Rajagopalan, S. & Ward, A. R. 2020. Predicting product return volume using machine learning methods. European Journal of Operational Research, 281(3), 612-627.

Dzyabura, D., El Kihal, S. & Ibragimov, M. 2018. Leveraging the power of images in predicting product return rates. SSRN Electronic Journal, 1-33.

Erdogan, Z. 2017. A Living Environment Prediction Model Using Ensemble Machine Learning Techniques Based on Life Quality Index, Istanbul University, Institute of Graduate Studies in Science and Engineering (M.Sc.Thesis), Istanbul, 2017.

Greve, C. & Davis, J. 2012. An Executive's Guide to Reverse Logistics: How to Find Hidden Profits by Managing Returns, editor no identificado.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H. 2009. The

WEKA Data Mining Software: An Update. SIGKDD Explorations, Volume 11, Issue 1.

Han, J., Pei, J. & Kamber, M. 2011. Data mining: concepts and techniques, Elsevier.

Izenman, A. J. 2008. Modern multivariate statistical techniques, Springer, New York, ABD.

Krapp, M., Nebel, J. & Sahamie, H. 2013. Forecasting product returns in closed-loop supplychains. Int J Phys Distrib Logist Manag, 43(8), 614–37.

Kumar, D.T., Soleimani, H. & Kannan, G. 2014. Forecasting return products in an integrated forward/reverse supply chain utilizing an ANFIS. Int J App Math Comput Sci.,24(3),69-82.

Ongsulee, P. 2017. Artificial intelligence, machine learning and deep learning. In 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE) (pp. 1-6). IEEE.

Simmons, K., Kinney, J., Owens, A., Kleier, D. A., Bloch, K., Argentar, D. & Vaidyanathan, G. 2008. Practical outcomes of applying ensemble machine learning classifiers to High-Throughput Screening (HTS) data analysis and screening. Journal of chemical information and modeling, 48(11), 2196-2206.

Temur, G.T., Balcilar, M. & Bolat, B. 2014. A fuzzy expert system design for forecastingreturn quantity in reverse logistics network. J Enterp Inf Manag. 27(3),316–28.

Toktay, B., 2003. Forecasting Product Returns. in Business Aspects of Closed-Loop Supply Chains, Ed: D. Guide, Jr. L.N. Van Wassenhove, Carnegie Bosch Institute, International Management Series:2.

Unlu, R. & Xanthopoulos, P. 2019. Estimating the number of clusters in a dataset via consensus clustering. Expert Systems with Applications, 125, 33-39.

Zhu, Y., Li, J., He, J., Quanz, B. L. & Deshpande, A. A. 2018. A Local Algorithm for Product Return Prediction in E-Commerce. In IJCAI (pp. 3718-3724).