Available online at www.alphanumericjournal.com

alphanumeric journal

The Journal of Operations Research, Statistics, Econometrics and Management Information Systems

Volume 7, Issue 1, 2019

Received: March 18, 2019
Accepted: June 20, 2019
Published Online: June 30, 2019

AJ ID: 2018.07.01.OR.05
DOI: 10.17093/alphanumeric.541307
Research Article

Using Machine Learning Algorithms For Forecasting Rate of Return Product In Reverse Logistics Process

Ayşe Nur Adıgüzel Tüylü, Ph.D. *

Res. Assist., Department of Industrial Engineering, Faculty of Engineering, Istanbul University- Cerrahpasa, Istanbul, Turkey, aysenur.adiguzel@istanbul.edu.tr

Ergün Eroğlu, Ph.D.

Prof., Department of Quantitative Methods, School of Business, Istanbul University, Istanbul, Turkey, eroglu@istanbul.edu.tr

* İstanbul Üniversitesi-Cerrahpaşa Mühendislik Fakültesi, 34320 Avcılar, İstanbul, Türkiye

ABSTRACT

Many textile products are in reverse logistics network due to mistakes made in activities such as sales forecasting, inventory planning and distribution. In order to reduce resource usage and cost at first step, in addition to producing the correct quantity, these products must be sent to branches, in correct properties (amount, color, size, model…) and transportation planning and stock planning should be done correctly. Statistical methods, artificial intelligence and machine learning methods are used because of the difficulty of establishing mathematical models in multi-parameter and multi-variable problems. In general, all these activities are based on demand forecasts by time series, but there are important differences between these demand predictions and the actual demands because of fashion and consumers' requests change very quickly. Artificial intelligence and machine learning methods provide faster and more accurate results in complex data sets. The difference of this study from other studies is to estimate the product return rates in Reverse Logistics with Machine Learning. In this direction, it is aimed to predict the claims accurately by concentrating on the customers' preferences, their reasons and the replies of the products which are sold to the customers. Thus, the consumer information obtained as a result of these analyzes can provide us with more accurate planning in terms of avoiding unnecessary production, transportation and storage activities, and sending the products with the correct properties; amount, color, size and model, to the branches. Best results (the correlation coefficient value is 82.35% and lowest error metrics) of this study are obtained with M5P algorithms of machine learning techniques

Keywords: Reverse Logistics, Forecasting Rate of Return Product, Machine Learning, Textile

Tersine Lojistik Sürecinde İade Oranlarının Tahmini İçin Makine Öğrenme Algoritmalarının Kullanılması

ÖZ

Satış tahmini, stok planlama ve dağıtım gibi faaliyetlerde yapılan hatalar nedeni ile birçok tekstil ürünü tersine lojistik ağına girmektedir. Kaynak kullanımını ve maliyeti en başta azaltmak için doğru sayıda üretimin yanı sıra bu ürünlerin doğru şubelere doğru sayıda, renkte, bedende ve modelde gönderilmesi, nakliyesinin ve stok planlamasının doğru bir şekilde yapılması gerekmektedir. Çok parametreli ve çok değişkenli problemlerde matematiksel model kurmanın zorluğu nedeniyle istatistiksel yöntemler, yapay zeka yöntemleri ve makine öğrenme yöntemleri kullanılmaktadır. Genel olarak tüm bu faaliyetler zaman serisine dayalı talep tahminleri baz alınarak yapılır, fakat moda ve tüketicilerin çok çabuk değişen istekleri nedeniyle talep tahminleri ile gerçekleşen talepler arasında önemli farklılıklar doğmaktadır. Son dönemde yapılan çalışmalar gösteriyor ki bu şekilde karmaşık yapılı büyük veri setlerinde yapay zeka ve makine öğrenme yöntemleri diğer tahmin yöntemlerine göre doğruluğu daha yüksek sonuçlar vermektedir. Bu çalışmada diğer çalışmalardan faklı olarak Tersine Lojistikte ürün iade oranlarının ilk defa Makine Öğrenme yöntemleri ile tahmin edilmesi yapılmıştır. Bu kapsamda müşterilerin tercihleri ile birlikte satışa çıkan ürünlerin iadeleri ve nedenleri üzerinde yoğunlaşılıp iadelerin daha doğru bir şekilde tahmin edilmesi amaçlanmıştır. Elde edilen analizler sonucunda şubelere doğru beden, renk ve modelde ürünlerin gitmesi; gereksiz üretim, nakliye ve depolama faaliyetlerinden kaçınılması; maliyetin, kaynak kullanımının ve çevre kirliliğinin azaltılması; kaçınılamayan nakliye ve depolama maliyetlerinin tahmin edilmesi konularında daha doğru bir planlama yapılması sağlanmıştır. Makine Öğrenme tekniklerinden M5P algoritması ile en iyi tahmin performansına (% 82,35 korelasyon katsayısı ve en düşük hata ölçütleri) ulaşmıştır.

Anahtar Kelimeler: Tekstil, Tersine Lojistik, Ürün İade Oran Tahmini, Makine Öğrenme

1. Introduction

For the strong and sustainable development of today's textile market, it is necessary to succeed in the reverse logistics activities which will affect the most important parameters; decrease in costs and increase of production efficiency. Moreover, it is not only a cost advantage for firms to gain importance in reverse logistics but also the legal obligations, customer satisfaction, social responsibility and information confidentiality.

Rogers and Tibben-Lembke (2001) estimated that reverse logistics is an important part of US logistics costs and that logistics costs are about 9.9% of the US economy. For the companies examined in the study, reverse logistics activities accounted for 4% of the total logistics activities. In addition, reverse logistics costs were estimated to be 0.5% of the total US GDP for the period in which the survey was conducted (Rogers ve Tibben-Lembke, 2001).

As a result of the mistakes made in the planning of activities such as logistics, sales forecasting, inventory management and change in customer appreciation; products that have not yet completed their life span have entered the reverse logistics network in order to regain value when they lose their place in the market. It is necessary to benefit from the information obtained from reverse logistics activities. More accurate planning can be made with the information of the returns from customers or from stores to the center. For example; products in which production cannot be estimated correctly, customers' preferences, location based change of these preferences, accuracy of sales and marketing planning, the accuracy of the number of products and product properties sent to each store, planning-related activities such as the results of sales strategies can be performed more accurately by analyzing information about returned products.

The crucial point that complicates the problem structure in product returns is uncertainty. Due to the uncertainty of the quality of the products to be returned and the reason for the return, the planning is based on the estimates. The higher the accuracy of the estimations, the less the reverse logistics activities and the costs caused by these activities. Artificial intelligence and machine learning methods provide faster and more accurate results in complex data sets (Alpaydın, 2014: 3). In addition, machine learning is one of the most efficient research areas in both the application of new techniques and theoretical algorithms, as well as applying them to real life problems (Olivas et al., 2009).

While the first definitions related to logistics are made by Lambert and Stock (1981), the Logistics Management Council (CSCMP) has made its first known definition of logistics in the 1990s. Toktay (2003) carried out a case study with KODAK disposable cameras to emphasize the importance of estimating the time periods of product returns and the amount of returning products in reverse logistics. Efendigil et al. (2009) proposed a new predictive mechanism modeled by artificial intelligence approaches, including comparison of artificial neural networks and adaptive network-based fuzzy inference systems. Xiaofeng and Tijun (2009) proposed a new model based on wave function to estimate the amount of product returned by reverse logistics. Clottey et al. (2012) developed a general estimation approach to determine the distribution of return of products used. Krapp et al. (2013a) presented an approach based on Bayesian estimation techniques to predict product returns in closed loop supply chains. Krapp et al. (2013b) developed a general estimation framework for product returns and proposed a combination of adaptive Bayesian approach and Kalman filter concepts. Agrawal et al. (2014) applied Graphical Evaluation and Review Technique for estimation of recycling in terms of quantity and time. Kumar et al. (2014a) has developed an integrated two-phase methodology for estimating return products with its own open-loop supply chain; in the first phase, it introduced the Adaptive Network Based Fuzzy Inference System, and in the second stage, they optimized the proposed multi-layer, multi-product, multi-cycle, closed-loop supply chain network. Temur et al. (2014) has developed a fuzzy expert system for the accurate estimation of the amount of return in the reverse logistics network. Firstly, the most important factors affecting the return of the products have been defined, then the factors that are co-linear with the others are eliminated by using size redundancy analysis.

In the literature, there are studies for demand estimation with successful results by Machine Learning algorithms. Aha et al. (1991) describes a framework and methodology called sample-based learning, which produces classification estimates using only specific examples. Anyanwu and Shiva (2009) conducted an experimental analysis based on sample data records to review the serial applications of decision tree algorithms and evaluate the performance of these algorithms. Erpolat and Öz (2010) tested the success of machine learning methods in the classification of breast cancer data by using artificial neural networks and support vector machines. Deng and Yeh (2011) used the Least Squares method in this study to support the support vector machines (LS-SVM) method which solved the problem of estimating the production cost of body structural projects. Marques et al. (2012) aimed to determine classifiers according to each community approach in the context of credit score, for this purpose, the estimation performance of C4.5 decision tree, multi-layer sensor, logistic regression, the nearest neighbor and naive Bayes classifiers were evaluated. Lamrini et al. (2016) presented a dynamic model of the process based on artificial neural networks in order to estimate the temperature of the bread dough and the power required for kneading.

In our study, the estimate of product returns is actually a demand forecast. Products returned by consumers or retailers are considered to be a major problem by manufacturers and managers as they create inventory surplus. Reverse logistics and returns are an important link that is often overlooked in an organization's supply chain. Accurate demand forecasting for returned products provides the company with strategic benefits in many key areas such as production, distribution and stock. Demand estimation methods are divided into two parts as qualitative and quantitative methods. Quantitative methods are divided into two as Mixed Methods and Time Series Analysis; Mixed Methods are also divided into two as Regression Analysis and Data Mining / Heuristic Methods. In our study we use Machine Learning methods from data mining estimation methods. The aim of this study is to determine the effect of the point of sale and the properties of the product on the return of the product by using the Machine Learning methods. In the literature, there is a lack of studies aimed at estimating the return rates for the retail sector and we aimed to contribute to the literature in this respect. This study differs from other studies in the literature in terms of the fact that it is the first study on the estimation of product return rates in reverse logistics with machine learning methods.

In our study, 80% of the data set was trained - 20% of the data was tested and 90% of the data set was trained - 10% of the data was tested. Linear Regression and Support Vector Regression from the functional algorithms, M5P from decision tree algorithms and M5Rules and Decision Table algorithms from rule-based algorithms were the best results. The obtained results were given comparatively and the best estimation performance was obtained by taking into consideration the correlation coefficient as well as error measurements.

2. Methodology

Machine learning explores the ability of computers to learn based on data or improve their performance. The main area of research is that computer programs learn to recognize complex patterns automatically and make intelligent decisions based on data (Han et. al., 2011). Machine learning emerged from the subfields of computer science known as artificial intelligence. Because intelligence cannot be achieved without learning, machine learning plays a crucial role in artificial intelligence. The idea of learning from experience is the center of the problems related to various types of problems encountered in machine learning, especially classification. The general purpose of each of the problems is to find a systematic way of classifying a future sample (Izenman, 2008).

In the first step of our study, we met with business analysts of a textile company operating worldwide on the importance of estimating product returns in reverse logistics activities and analyzed product return data of the company with these business analysts. We selected a specific product group from a huge pool of data to review return rates in more detail. When choosing the range of returns to be estimated, we paid attention to the width of the product range, the consistency of the return rate range, the missing or extreme data is as low as possible. For the study, we analyzed and edited the data belonging to this product group by finding the appropriate female trousers product group. In the process of editing the data, together with the business analysts, we determined the properties of the products and stores, arranged the missing and the extreme data and we received information about the reasons for the return. We calculated the return rate of a product from a store to the center and the number of products that had been returned to that store.

We entered the edited data set in WEKA (Waikato Environment for Information Analysis) program and defined the data according to whether the data are categorical or numerical. WEKA is a program that allows application of standard machine learning techniques to real-world data sets.

Developed to provide an integrated environment that provides easy access to various machine learning techniques through an interactive interface to work with real-world datasets (Holmes et. al., 1994). WEKA includes regression, classification, clustering, relationship rule analysis and attribute selection methods for all standard data mining problems. All algorithms and methods take their inputs as a single relational table, which can be read from a file or produced by a database query. The system is written in Java programming language (Frank et. al., 2009).

We estimated the return rates of products using the classification algorithms from the Machine Learning Methods on the defined data and we achieved the estimation performance based on these estimates. We evaluated the performances of M5P, REPTree, Decision Stump, Random Tree, M5Rules, Decision Table, KStar, IBk, LWL, Linear Regression, SMOreg, Multilayer Perceptron methods appropriate to the data sets including both categorical and numerical values from the machine learning methods. First, we set the program will use the 80% of the data set to train the algorithms for learning, and 20% will estimate return rates. Next, we set the program to train the algorithms with 90% of the data set and estimate with 10%. In the event that the program sets the data set as both 80% training-20% test set and 90% training-10% test set, we showed the performance of the prediction obtained from applied machine learning algorithms as tables and we compared the methods with each other in terms of correlation coefficient (R) and error values (RRSE, RMSE, MAE, RAE).

The concept of classification is to distribute the data to the classes in the data set according to the qualifications. The properties and number of these classes are predetermined. The values that specify these classes in the data set are called labels. The classes of the items in the training set are defined and are used to create a model. The classification algorithms analyze the relationships between the class labels in the given training set and the other properties. The success of the model is measured by testing the items that are not in the model set. As a result, it is decided which class belongs to the newly arrived item and this model is tested with the help of this model.

2.1. Lazy Algorithms

The biggest difference between the other methods and lazy algorithms is to keep the learning set. The processes carried out during the learning phase in the other methods, are carried out in the estimation stage in this method.

- K * (K Star), is an example-based classifier, ie, the class of a test sample is based on a class of similar training examples, as determined by some similarity functions. Different from other sample-based learners using an entropy-based distance function (Cleary and Trigg, 1995).
- IBk (K-nearest neighbor), classifies the examples according to vote of the most of the most similar examples (Aha et. al., 1991). The distance of the neighbors is measured by Euclidean distance.
- LWL (Locally Weighted Learning), sets up a Naïve Bayes model using the cluster weight of learning samples in classifying a new sample, unlike other lazy methods.

2.2. Rule Based Algorithms

- Decision Tables, are a decision table that is formed and classified by the characteristics of the data in the training set. Its performance is good on some data sets with continuous features (Kohavi, 1995).
- M5Rules, is a rule-based learning technique and can estimate nominal and numerical values. M5 rule sets are formed from model trees. The rule algorithm works by repeating the model tree creation process and trying to select the best rule in each cycle (Ayaz et. al., 2015).

2.3. Decision Tree Algorithms

The decision tree algorithm is a data mining initialization technique that recursively splits the data set until all data elements belong to a particular class. A decision tree structure consists of root, inner and leaf nodes. The tree structure is used to classify unknown data records. Tree leaves consist of class labels where data items are grouped. The decision tree classification technique is carried out in two stages: tree growth and pruning. In tree growth, all data elements of the tree are separated until they arrive at the same class label. Pruning is used to improve the accuracy and estimation of the algorithm by minimizing detail in training data (Anyanwu and Shiva, 2009).

- Decision Stump is a one-step decision tree method. This algorithm classifies according to a single input property. In this method, the stem is directly attached to the leaves.
- In Random Tree algorithm, a tree structure is randomly selected from within the tree cluster.
- REPTree is used to sort numerical properties. When creating decision tree using information gain, pruning with reduced error pruning.
- M5P, Model tree called M5, has been introduced to cope with learning problems (Ayaz vd., 2015). M5P combines decision tree for data mining and multiple linear regression (Nikoo et. al., 2013).

2.4. Functional Algorithms Used in Classification

- Multilayer Perception (Artificial Neural Networks – ANN) is a computer system which is developed by inspiring the human brain, learning by imitating biological neural networks, connected to each other by means of weighted links and consisting of processing elements, each having its own memory, in parallel and distributed information processing structures ANN are developed with the ability to automatically acquire new information without any help through learning (Namlı, 2012).
- Support Vector Regression (SVR), is a statistical method that analyzes regression problems using this estimated linear or nonlinear function, based on the estimation of the most appropriate function to separate data from each other. SVR tries to find a function that minimizes the risk of regression (Namlı, 2012).
- Linear Regression, is the method that expresses the relationship between a variable and one or more variables that affect this variable with a linear model.

2.5. Performance Metrics (Chou vd., 2015)

Linear Correlation Coefficient (R): A common measure of how well the R curve fits the actual data. A value of 1 means that the values have the same tendency. y' is the estimated value; y real value; n is the number of data samples.	$R = \frac{n\sum yy' - (\sum y)(\sum y')}{\sqrt{n(\sum y^2) - (\sum y)^2}\sqrt{n(\sum y'^2) - (\sum y')^2}}$
The Mean Absolute Error (MAE) is an amount used to measure how close the estimates are to the final results.	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y - y'\rvert$
The Square Root of the Mean Square Error (RMSE) is calculated to find the square error of the estimation and the square root of the total value. That is, the average distance of a data	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\sum(y' - y)^2}{n}}$

point from a fixed line measured along a vertical line.	
The Relative Absolute Error (RAE) is the ratio of the absolute value of the difference between the estimated and actual values to the actual values.	$RAE = \frac{\lvert y'_1 - y_1 \rvert + \cdots + \lvert y'_n - y_n \rvert}{\lvert y_1 - \bar{y} \rvert + \cdots + \lvert y_n - \bar{y} \rvert}$
The Square Root of the Relative Square Error (RRSE) is the square root of the sum of the squares of the differences between the estimated value and the actual value to the sum of the squares of the differences between the actual values and the mean value.	$RRSE = \sqrt{\frac{(y'_1 - y)^2 + \cdots + (y'_n - y_n)^2}{(y_1 - \bar{y})^2 + \cdots + (y_n - \bar{y})^2}}$

3. Results

Firstly, the data set was divided into 80% - 20% for training and testing and the predictive performance of Machine Learning techniques was discussed. M5Rules algorithm as seen in Table 1., gave the best results in terms of performance metrics (R, RMSE, MAE, RAE and RRSE).

M5P		**REPTree**	
Correlation coefficient	**0,8018**	Correlation coefficient	0,6953
Mean absolute error	0,0114	Mean absolute error	0,0141
Root mean squared error	0,0151	Root mean squared error	0,0182
Relative absolute error	51,76%	Relative absolute error	64,03%
Root relative squared error	60,10%	Root relative squared error	72,39%
Model Building Duration	276,4	Model Building Duration	1,11
Decision Stump		**Random Tree**	
Correlation coefficient	0,4418	Correlation coefficient	0,6645
Mean absolute error	0,0192	Mean absolute error	0,0142
Root mean squared error	0,0225	Root mean squared error	0,0193
Relative absolute error	87,16%	Relative absolute error	64,52%
Root relative squared error	89,71%	Root relative squared error	77,06%
Model Building Duration	0,02	Model Building Duration	0,2
M5Rules		**Decision Table**	
Correlation coefficient	**0,8098**	Correlation coefficient	0,7412
Mean absolute error	0,0113	Mean absolute error	0,0131
Root mean squared error	0,0148	Root mean squared error	0,0169
Relative absolute error	51,13%	Relative absolute error	59,57%
Root relative squared error	58,82%	Root relative squared error	67,23%
Model Building Duration	408,96	Model Building Duration	1,38
KStar		**IBk**	
Correlation coefficient	0,6732	Correlation coefficient	0,6642
Mean absolute error	0,0141	Mean absolute error	0,0141
Root mean squared error	0,0188	Root mean squared error	0,0192
Relative absolute error	64%	Relative absolute error	64%
Root relative squared error	75%	Root relative squared error	76%
Model Building Duration	0	Model Building Duration	0,01
LWL		**Linear Regression**	
Correlation coefficient	0,5845	Correlation coefficient	0,7478
Mean absolute error	0,0176	Mean absolute error	0,0132
Root mean squared error	0,0206	Root mean squared error	0,0167
Relative absolute error	80%	Relative absolute error	59,73%
Root relative squared error	82%	Root relative squared error	66,44%
Model Building Duration	0,01	Model Building Duration	225,94

Table 1. Results from the Machine Learning algorithms (% 80 Training-% 20 Test Set)

The data set is divided into 90% -10% for training and testing, and the predictive performances of Machine Learning techniques are given in Table 2. M5P algorithm as shown in the table; It gave the best results in terms of performance metrics (R, RMSE, MAE, RAE and RRSE).

M5P		**REPTree**	
Correlation coefficient	**0,8235**	Correlation coefficient	0,7215
Mean absolute error	0,0106	Mean absolute error	0,0134
Root mean squared error	0,0141	Root mean squared error	0,0173
Relative absolute error	49,14%	Relative absolute error	62,05%
Root relative squared error	56,93%	Root relative squared error	69,53%
Model Building Duration	304,61	Model Building Duration	0,35
Decision Stump		**Random Tree**	
Correlation coefficient	0,4426	Correlation coefficient	0,6782
Mean absolute error	0,019	Mean absolute error	0,0138
Root mean squared error	0,0223	Root mean squared error	0,0187
Relative absolute error	87,87%	Relative absolute error	63,70%
Root relative squared error	89,72%	Root relative squared error	75,33%
Model Building Duration	0,06	Model Building Duration	0,15
M5Rules		**Decision Table**	
Correlation coefficient	**0,8222**	Correlation coefficient	0,7379
Mean absolute error	0,0107	Mean absolute error	0,013
Root mean squared error	0,0142	Root mean squared error	0,0168
Relative absolute error	49,44%	Relative absolute error	60,03%
Root relative squared error	57,24%	Root relative squared error	67,67%
Model Building Duration	450,42	Model Building Duration	1,74
KStar		**IBk**	
Correlation coefficient	0,6819	Correlation coefficient	0,6708
Mean absolute error	0,0137	Mean absolute error	0,0138
Root mean squared error	0,0183	Root mean squared error	0,0187
Relative absolute error	63%	Relative absolute error	64%
Root relative squared error	74%	Root relative squared error	76%
Model Building Duration	0	Model Building Duration	0
LWL		**Linear Regression**	
Correlation coefficient	0,582	Correlation coefficient	0,7423
Mean absolute error	0,0175	Mean absolute error	0,0131
Root mean squared error	0,0204	Root mean squared error	0,0167
Relative absolute error	81%	Relative absolute error	60,40%
Root relative squared error	82%	Root relative squared error	67,18%
Model Building Duration	0	Model Building Duration	253,31
SMOreg		**Multilayer Perceptron**	
Correlation coefficient	0,7252	Correlation coefficient	0,364
Mean absolute error	0,0129	Mean absolute error	0,0254
Root mean squared error	0,0175	Root mean squared error	0,0306
Relative absolute error	59,42%	Relative absolute error	117,24%
Root relative squared error	70,38%	Root relative squared error	123,31%
Model Building Duration	2331,86	Model Building Duration	8805,02

Table 2. Results from the Machine Learning algorithms (% 90 Training-% 10 Test Set)

Machine Learning Classifiers	**R**	**MAE**	**RMSE**	**RAE**	**RRSE**
M5P %80-20	0,8018	0,0114	0,0151	51,76%	60,10%
M5P %90-10	0,8235	0,0106	0,0141	49,14%	56,93%
REPTree %80-20	0,6953	0,0141	0,0182	64,03%	72,39%
REPTree %90-10	0,7215	0,0134	0,0173	62,05%	69,53%
Decision Stump %80-20	0,4418	0,0192	0,0225	87,16%	89,71%
Decision Stump %90-10	0,4426	0,019	0,0223	87,87%	89,72%
Random Tree %80-20	0,6645	0,0142	0,0193	64,52%	77,06%
Random Tree %90-10	0,6782	0,0138	0,0187	63,70%	75,33%
M5Rules %80-20	0,8098	0,0113	0,0148	51,13%	58,82%
M5Rules %90-10	0,8222	0,0107	0,0142	49,44%	57,24%
Decision Table %80-20	0,7412	0,0131	0,0169	59,57%	67,23%
Decision Table %90-10	0,7379	0,013	0,0168	60,03%	67,67%
KStar %80-20	0,6732	0,0141	0,0188	63,91%	74,81%

Machine Learning Classifiers	R	MAE	RMSE	RAE	RRSE
KStar %90-10	0,6819	0,0137	0,0183	63,47%	73,83%
IBk %80-20	0,6642	0,0141	0,0192	63,80%	76,32%
IBk %90-10	0,6708	0,0138	0,0187	63,76%	75,55%
LWL %80-20	0,5845	0,0176	0,0206	79,77%	82,23%
LWL %90-10	0,582	0,0175	0,0204	80,92%	82,33%
Linear Regression %80-20	0,7478	0,0132	0,0167	59,73%	66,44%
Linear Regression %90-10	0,7423	0,0131	0,0167	60,40%	67,18%
SMOreg %80-20	0,7352	0,0129	0,0173	58,43%	68,99%
SMOreg %90-10	0,7252	0,0129	0,0175	59,42%	70,38%
Multilayer Perceptron %80-20	0,4775	0,0314	0,0399	142,41%	159,10%
Multilayer Perceptron %90-10	0,364	0,0254	0,0306	117,24%	123,31%

Table 3. Performance Metrics of the Machine Learning methods

Table 3. presents the performance metrics obtained by the Machine Learning methods from the data set is divided into both the 80% - 20% data set and 90% to 10%.

When the performance of the machine learning algorithms is compared according to the correlation coefficient, the best value is obtained by M5P 90-10% algorithm and the worst result is artificial neural networks with% 90-10 algorithm.

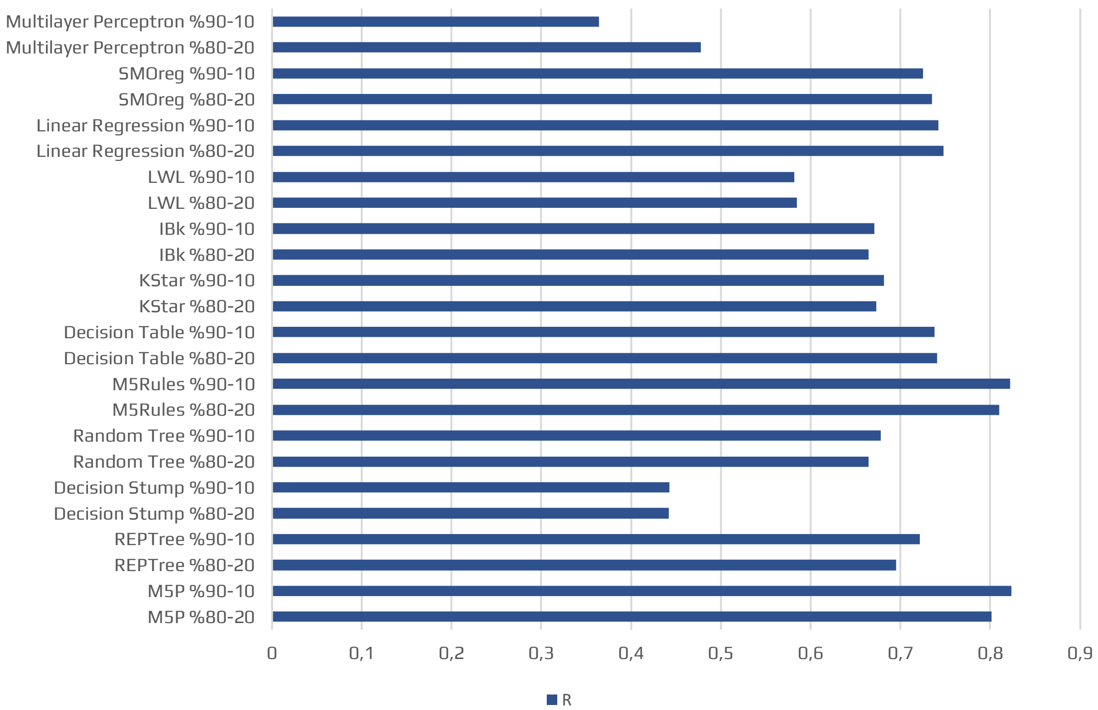

Figure 1. Correlation coefficients of results obtained by Machine Learning algorithms

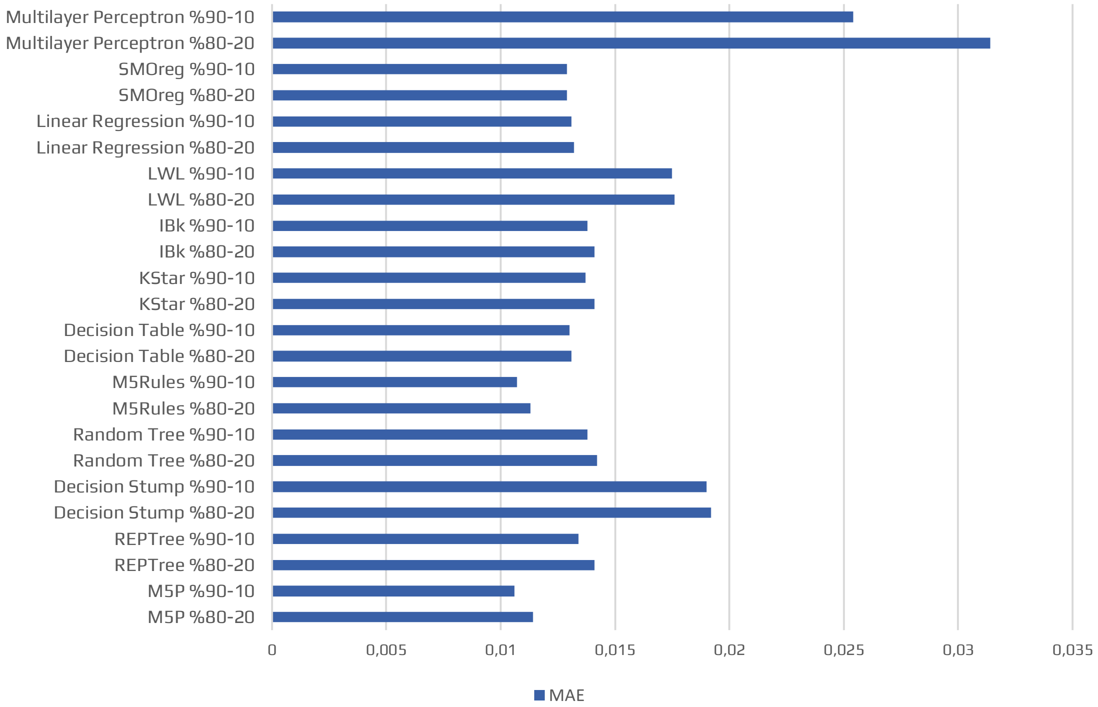

Figure 2. The Mean Absolute Errors of results obtained by Machine Learning algorithms

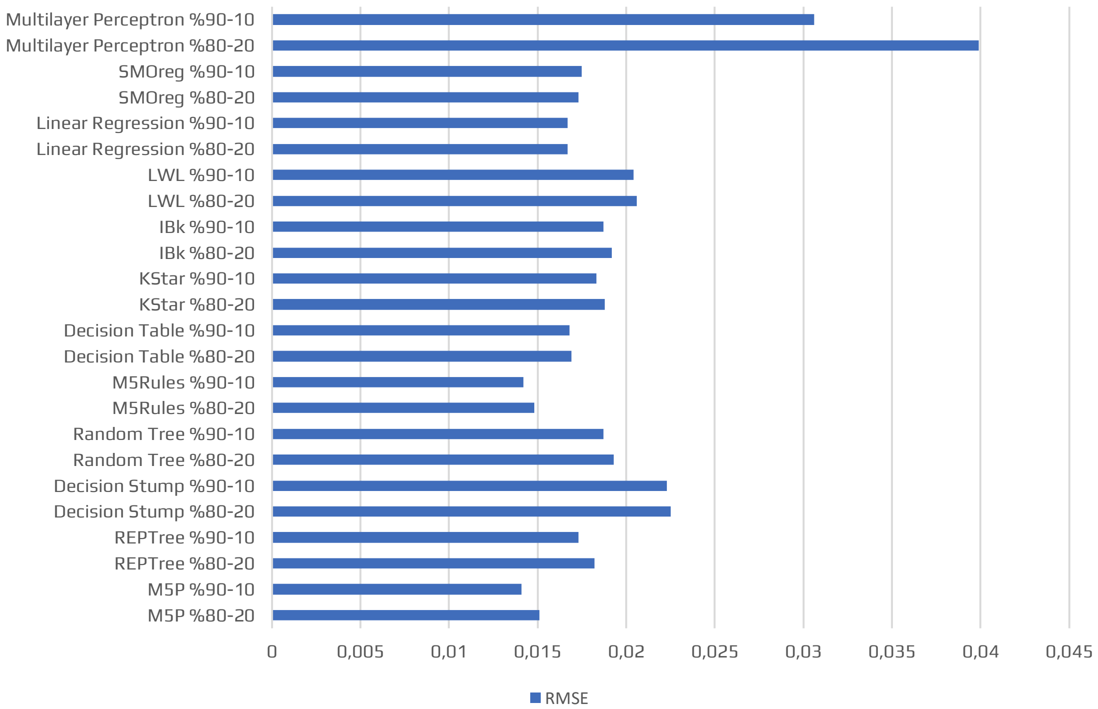

Figure 3. The Square Root of the Mean Square Error of results obtained by Machine Learning algorithms

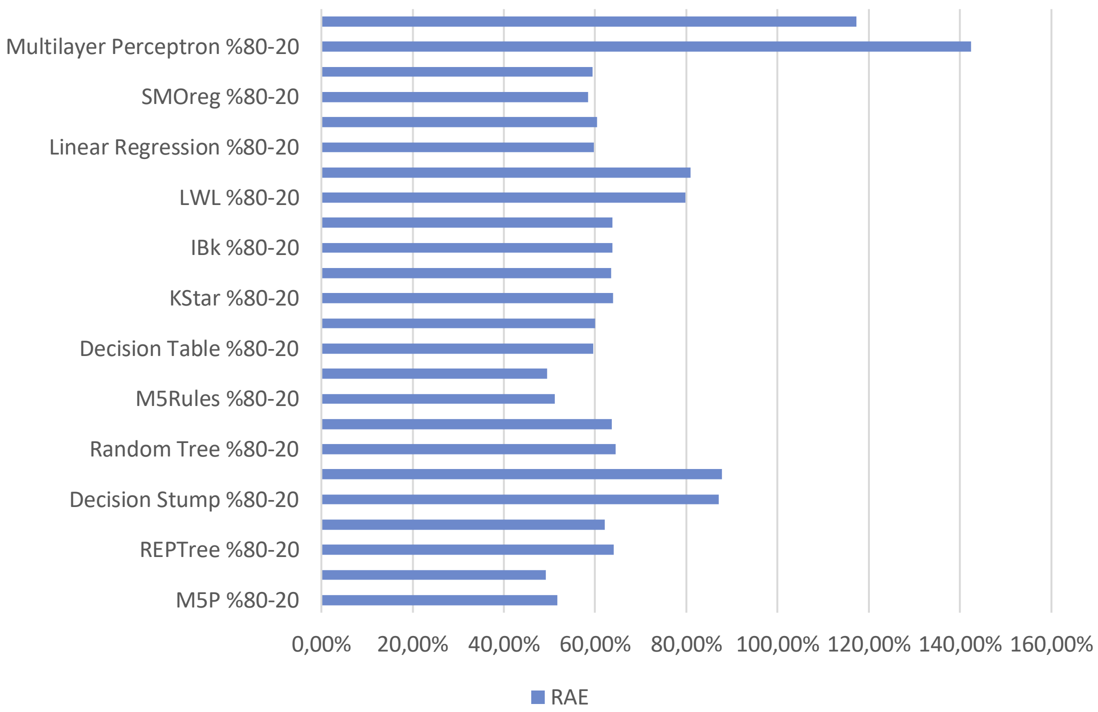

Figure 4. The Relative Absolute Error of results obtained by Machine Learning algorithms

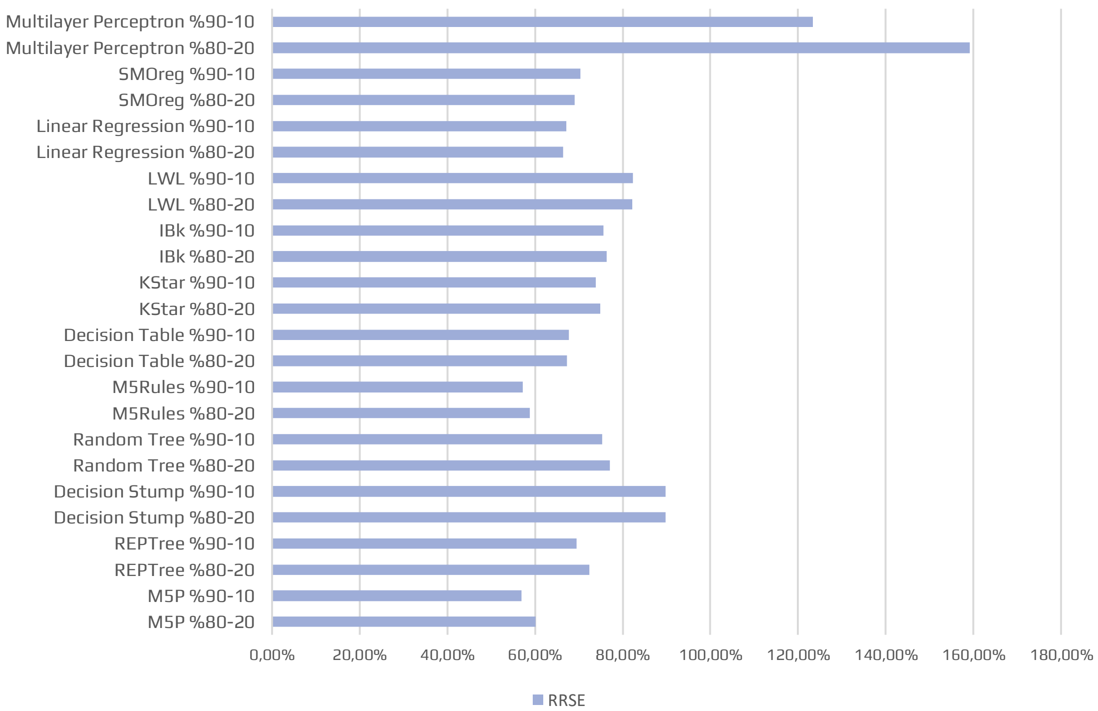

Figure 5. The Square Root of the Relative Square Error of results obtained by Machine Learning algorithms

When the error metrics of the results obtained from the machine learning algorithms are examined, the ANN 80-20% algorithm has the most error metrics, while M5P 90-10% algorithm has the least error metrics for all error metrics.

4. Discussion and Conclusion

Due to the mistakes made in production planning, sales forecasting, transportation, sales policy, stock planning, packaging and distribution activities, many textile products cannot be sold at the end of the sales period and entered into reverse logistics network. These products cause the use of resources, energy and capital in the logistics phase. When they enter the reverse logistics flow because of not being sold, 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.

Due to the multi-parameter and multivariate structure of the estimation of the rate of return on textile products, instead of building a mathematical model and because of the rapidly changing demands of the consumers and the fashion, as in the studies in the literature in general, instead of estimating the demand based on time series, Machine Learning methods were used which give faster and more accurate results in complex structured data sets. This is the first study to use Machine Learning Methods to estimate product return rates. The results show these the Machine Learning methods have the ability to estimate the return rates of the textile sector.

In this study, we focused on the return rates of the products with the preferences of the customers and the reasons of the returns, and the results of the analyzes made with the aim of correctly estimating the returns. In order to accurately estimate returns, consumer behavior information obtained from these analyzes may be ensured that the products are delivered to the stores in the right size, color and model, and unnecessary production, transportation and storage activities can be avoided. Thus, by means of a more accurate product return estimation obtained as a result of our work, the company can have many advantages in areas such as minimizing the costs and resource consumption, determination of production strategy, vehicle and storage capacity works, vehicle routing, production planning, supplier selection, by reducing all the reverse logistics activities (unnecessary stock formation in stores; products that cannot be sold due to lack of stock; transport of returned products to the center, warehouse or outlet stores; transportation, handling, packaging, transportation, fuel, labor and driver costs, such as transportation costs; redundant areas in the warehouse for storing returned products instead of new products; actions to be taken for these transactions in the warehouse and the costs of these activities; renewal activities to add value to the products returned to the center and the costs for this process; strategies for non-resale products and campaign activities) before they occur.

Machine learning classification techniques have been estimated by Linear Regresyon (LR), Support Vector Regression (SVR) and Artificial Neural Networks (ANN) from functional algorithms, M5P, REPTree, Random Tree, Decision Stump from decision tree algorithms, M5Rules and Decision Table from rule-based algorithms, KStar, IBk and LWL from lazy algorithims. Machine learning methods M5Rules and M5P showed the best performance in terms of both correlation coefficient and error metrics. The results obtained in the study show that high-performance results are obtained. By the machine learning methods and these results support the recent studies on this subject in the literature.

References

Agrawal, S., Singh, R.K., Murtaza, Q., (2014), "Forecasting product returns for recycling in Indian electronics industry", J Adv Manag Res., C.11, No:1, 102–14.

Aha, D. W., Kibler, D., Albert, M. K., (1991), "Instance-based learning algorithms", Machine learning, C.6, No:1, 37-66.

Akküçük, U., (2011), Veri Madenciliği: Kümeleme ve Sınıflama Algoritmaları, Yalın Yayıncılık, İstanbul.

Alpaydın, E., (2014), Introduction to machine learning, MIT press.

Anyanwu, M. N., Shiva, S. G., (2009), "Comparative analysis of serial decision tree classification algorithms", International Journal of Computer Science and Security, C.3, No:3, 230-240.

Ayaz, Y., Kocamaz, A. F., Karakoç, M. B., (2015), "Modeling of compressive strength and UPV of high-volume mineral-admixtured concrete using rule-based M5 rule and tree model M5P classifiers", Construction and Building Materials, C.94, 235-240.

Aydogmus, H. Y., Erdal, H. İ., Karakurt, O., Namli, E., Turkan, Y. S., Erdal, H., (2015), "A comparative assessment of bagging ensemble models for modeling concrete slump flow", Computers and Concrete, C.16, No:5, 741-757.

Chen, H., He, H., (2010), "Reverse logistics demand forecasting under demand uncertainty", IntConf Logist Eng Manag ASCE Conf Proc., 1:343.

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, C.30, No:1, 04015002.

Cleary, J. G., Trigg, L. E., (1995), "K*: An instance-based learner using an entropic distance measure", In Proceedings of the 12th International Conference on Machine learning, C.5, 108-114.

Clottey, T., Benton, W.C., Jr, Srivastava, R., (2012), "Forecasting product returns for remanufacturing operations", Decis Sci, C.43, No:4, 589–614.

Deng, S., Yeh, T. H., (2011), "Using least squares support vector machines for the airframe structures manufacturing cost estimation", International Journal of Production Economics, C.131, No:2, 701-708.

Efendigil, T., Onut, S., Kahraman, C., (2009), "A decision support system for demand forecastingwith artificial neural networks and neuro-fuzzy models: a comparative analysis", Expert Syst Appl, C.36, No:3, 6697–707.

Erpolat, S., Öz, E., (2010), "Kanser Verilerinin Sınıflandırılmasında Yapay Sinir ağları ile Destek Vektör Makineleri'nin Karşılaştırılması", İstanbul Aydın Üniversitesi Fen Bilimleri Dergisi, C.2, No:5, 71-83.

Fleischmann, M, Van Wassenhove, L.N., Van Nunen, J.A.E.E., van der Laan, E.A., Dekker, R., Bloemhof-Ruwaard, J.M., (1997), "Quantitative models for reverse logistics: a review", European Journal of Operation Research, C.103, No:1, 1–17.

Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I. H., Trigg, L., (2009), "Weka-a machine learning workbench for data mining", In Data mining and knowledge discovery handbook, Springer, ABD, 1269-1277.

Garner, S. R., (1995), "Weka: The waikato environment for knowledge analysis", In Proceedings of the New Zealand computer science research students conference, 57-64.

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

Holmes, G., Donkin, A., Witten, I. H., (1994), "Weka: A machine learning workbench. In Intelligent Information Systems", Proceedings of the 1994 Second Australian and New Zealand IEEE Conference on, 357-361.

Hsieh, S. L., Hsieh, S. H., Cheng, P. H., Chen, C. H., Hsu, K. P., Lee, I. S., Lai, F., (2012), "Design ensemble machine learning model for breast cancer diagnosis", Journal of medical systems, C.36, No:5, 2841-2847.

Izenman, A. J., (2008), Modern multivariate statistical techniques, 1. Baskı, Springer, New York, ABD.

Kohavi, R., (1995), Wrappers For Performance Enhancement And Oblıvıous Decısıon Graphs, Carnegie Mellon University, Department of Computer Science, Pittsburgh.

Krapp, M., Nebel, J., Sahamie, H., (2013), "Forecasting product returns in closed-loop supplychains", Int J Phys Distrib Logist Manag, C.43, No:8, 614–37.

Krapp, M., Nebel, J., Sahamie, H., (2013), "Using forecasts and managerial accountinginformation to enhance closed-loop supply chain management", OR Spectr, C.35, No:4, 975–1007.

Kumar, D.T., Soleimani, H., Kannan, G., (2014), "Forecasting return products in an integratedforward/reverse supply chain utilizing an ANFIS", Int J Appl Math Comput Sci., C.24, No:3, 669–82.

Lamrini, B., Della Valle, G., Trelea, I. C., Perrot, N., & Trystram, G., (2016), "A new method for dynamic modelling of bread dough kneading based on artificial neural network", Food control, C.26, No:2, 512-524.

Marqués, A. I., García, V., Sánchez, J. S., (2012), "Exploring the behaviour of base classifiers in credit scoring ensembles", Expert Systems with Applications, C.39, No:11, 10244-10250.

Marqués, A. I., García, V., Sánchez, J. S., (2012), "Two-level classifier ensembles for credit risk assessment", Expert Systems with Applications, C.39, No:12, 10916-10922.

Namlı, E., (2012), Proje yönetimi kapsamında risk tabanlı ve yapay zeka destekli bir maliyet tahmin modeli, İstanbul Üniversitesi Fen Bilimleri Enstitüsü (Doktora Tezi), İstanbul.

Nikoo, M. R., Karimi, A., Kerachian, R., Poorsepahy-Samian, H., Daneshmand, F., (2013), "Rules for optimal operation of reservoir-river-groundwater systems considering water quality targets: application of M5P model", Water resources management, C.27, No:8, 2771-2784.

Olivas, E. S., Guerrero, J. D. M., Sober, M. M., Benedito, J. R. M., Lopez, A. J. S., (2009), Handbook Of Research On Machine Learning Applications and Trends: Algorithms, Methods and Techniques, Information Science Reference, New York, ABD.

Rogers, D.S., Tibben-Lembke, R.S., (2001), "An examination of reverse logistics practices", J BusLogist, C.22, No:2, 129–48.

Rogers, D.S., Melamed, B., Lembke, R.S., (2012), "Modeling and analysis of reverse logistics", J BusLogist, C.33, No:2, 107–17.

Srivastava, S.K., Srivastava, R.K., (2006), "Managing product returns for reverse logistics", Int JPhys Distrib Logist Manag, C.36, No:7, 524–46.

Temur, G.T., Balcilar, M., Bolat, B., (2014), "A fuzzy expert system design for forecastingreturn quantity in reverse logistics network", J Enterp Inf Manag, C.27, No: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: Cilt:2.

Toktay, B., van der Laan, E.A. de Brito, M.P., (2004), "Managing Product Returns: Informational Issues and Forecasting Methods", in Reverse Logistics: Quantitative Methods for Closed-Loop Supply Chains, Ed: R. Dekker, Fleischmann, M., K. Inderfurth, L.N. Van Wassenhove, Springer Verlag.

Xiaofeng, X., Tijun, F., (2009), "Forecast for the amount of returned products based on wavefunction", Int Conf Inf Manag Innov Manag Ind Eng, C.2, 324–7.