Machine Learning Algorithms with Intermittent Demand Forecasting: An Application in Retail Apparel with Plenty of Predictors

İlker GÜVEN[1] (iD) 0000-0002-2754-6893
Özer UYGUN[2] (iD) 0000-0002-8437-7678
Fuat ŞIMŞIR[3] (iD) 0000-0001-7001-5951

[1,3] Karabük University / Department of Industrial Engineering / Karabük
[2] Sakarya University / Department of Industrial Engineering / Sakarya

Corresponding Author: Fuat Şimşir, fuatsimsir@karabuk.edu.tr

ABSTRACT

Demand forecasting is a key factor for apparel retail stores to sustain their business, especially where there are variety of products and intermittent demand. In this study, two of the most popular machine learning methods, random forest (RF) and k-nearest neighbour (KNN), have been used to forecast retail apparel's intermittent demand. Numerous variables that may have an effect on the sales, have been taken into account one of which is defined as "special day" that might trigger intermittency in the demand. During the forecast application, four different datasets were used to provide reliability. 28 different variables were used to increase accuracy of the forecasting and experience of the behaviours of the algorithms. Root mean square error (RMSE) was used to evaluate performance of the methods and as a result of this study, RF showed better performance in all four datasets comparing to KNN.

1. INTRODUCTION

Forecasting customer demand is challenging considering the variables' effects on the demand, and constantly changing needs of customers. Also, it is both mandatory and challenging for retailers to make operational decisions [1]. In case of wrong prediction during the application, the results might be losing customer and sales as well as losing prestige of the brand. To avoid loss of sales, companies make plans in the short term and one of the most important inputs of this plan is demand forecasting. However, making prediction is not easy because many variables that affect the demand differ according to the sector and are shaped outside the supplier [2]. Predicting the amount of sales and product types for future periods correctly is a serious investment for companies. According to the forecasting figures, inventory, stock and labor cost can be optimized [3].

Missing data, fluctuating demand during special days, economic crisis, missing variables that affect the prediction and so on can be counted among the challenges of demand forecasting. Since it is a complex and important topic, many methods were suggested to forecast the demand. A classical forecast can be made based on the time series of past sales figures [4], while advanced methods take several features into account. Hence, classical methods such as moving averages, exponential smoothing or linear regression are still favorite in industrial practice [3].

Demand forecasting is important for all kinds of businesses, but more challenging for retailers in terms of uncertainty and complexity. All factors that can influence the forecasting performance are out of control of the business owner or cannot be identified [5]. These factors can depend on the product in terms of its color, price, size, number of stores, customers in terms of fashion, or external factors such as weather, economy and so on. Besides, past sales

figures can also be debatable [6]. One of the characteristics of the apparel retailers is having intermittent demand where many of the sales periods have zero demand, which increases the complexity in the demand [7]. The reason for the intermittent demand can be expressed as seasonality.

Intermittent demand is the most frequently seen demand type in the apparel retailer for the items. Due to the sales period or the seasons of the year, customers don't need to buy specific type of products, thus sales of these particular products occur as zero. Traditional forecasting methods have trouble catching these zero sales periods of the goods. Nikolopoulos has made a point regarding intermittent demand and pointed out few things; forecasting literature focuses on the fast time series, common sense neglects the intermittent demand's importance, special sales periods can be forecasted by using intermittent demand models [8].

In the literature, there are significant number of studies regarding demand forecasting which emphasize the importance of it. However, some of these studies follow traditional methods, while others use artificial intelligence methods or other methods. Recent developments and attention paid to the artificial intelligence led the researchers to use these methods such as artificial neural networks (ANN) for forecasting [9]. Güven and Şimşir have used ANN and support vector machines (SVM) in their study where they examined whether the color of the product has importance on the sales. According to results of the study, methods matter regarding the performance of the forecasting with color parameter [2]. Ma and Fildes have used meta-learning methods to forecast retail sales in 2020. They have proposed the first evaluation of meta-learning methods in the retailers and this method is based on the deep learning neural networks [10]. In another study, Loureiro et al., also have used deep neural networks for sales forecasting in apparel retailers. In the study, they have performed comparison of the deep neural networks with other artificial intelligence techniques such as decision trees, random forest (RF), SVM, ANN and linear regression has been given. Obtained results of the study have showed that deep learning outperformed other techniques in some of the evaluation metrics, however it does not significantly outperform some other techniques such as RF [11]. In 2017 Yu et al., compared SVM and RF for real time radar-derived rainfall forecasting and revealed that SVM based forecasting method outperformed RF based method [12]. Cui et al. in 2018 applied random forest along with other methods such as gradient boosting and SVM for daily sales forecast for online apparel retailers, and according to results of the study they outperformed linear models [13]. K-nearest neighbor (KNN) algorithm is another method used for forecasting in the literature. Martinez et al., have used KNN to deal with seasonality in the time series forecasting problem by assigning specialized KNN for every season and made forecasting [14]. Souza et al., have used KNN regression model to forecast trunk volume in a city and their opinion is using KNN unless classical methods are far away from providing good results [15]. In another study,

Ghiassi et al., have forecasted water demand of a metropolitan using KNN besides other machine learning methods. They have obtained significantly good results from all the methods; however dynamic artificial neural network was the best model in the study [16].

Traditional forecasting methods such as ARIMA, exponential smoothing, linear regression have been applied to forecast problems successfully, and according to studies in the literature there is no certain proof that they are less good than machine learning models [1]. One of the traditional methods, regression, has a wide application area in the literature as single application or combined/improved with other approaches regarding forecasting in different domains [17]–[20]. ARIMA is another one of the most common methods used for demand forecasting in the literature [21], [22]. ARIMA is often used with ANN [23]–[25]. Also, Bayesian methods [26], [27], moving average [28]–[30] and exponential smoothing [31]–[34] are commonly used for demand forecasting problems. Moving average and exponential smoothing are one of the most popular intermittent demand forecasting methods [7].

Based on the literature review, making accurate prediction especially in the retail apparel requires high consideration of the variables that may have an effect on the sales. Taking only past sales into account might neglect the reality of the sales. Changes between sales terms cannot be explained by the change itself; but there must also be other factors that can involve in the sales. It is a well-known fact that on special days or events such as holidays, Eid, religious days sales increase. Ignoring these factors' effect on the sales would decrease the chance of valid prediction. These variables should be chosen and determined carefully. Unfortunately, traditional methods are unsuitable to take these variables into account. However, according to the literature [1], [2] traditional methods are still competitive in terms of forecasting for some cases or in short terms.

In this study, demand forecasting has been made in color of the item/product level for the retail apparel industry with RF and KNN as machine learning methods. Thus, we are able to compare the results between two machine learning methods in the intermittent demand environment with large sets of the predictors that may influence the sales. One of the reasons for selecting RF is that we can obtain the importance of the variables that are used in the study. That could lead to the elimination of the variables from less effective to more effective ones, and for the future predictions effective variables could be taken into account. The data handled in the study has high standard deviation at the output variable which means real sales are highly variable. That is the biggest challenge in intermittent demand forecasting. Variables, in data, have non-parametric values which limit the problem applications to certain algorithms. Thus, RF and KNN, non-parametric algorithms, have been chosen for the application. Therefore, both algorithms are successful for numerous data according to literature review.

The contribution of the study is testing the success of machine learning methods applied for retail apparel industry when the sales are intermittent and the standard deviation is high, and indicating how to decide optimum parameters for the method used for prediction. Besides, since determining the variables that affect the sales in the retail apparel industry is quite difficult and vary according to the type of the product, identifying the most effective variables on prediction performance by the values obtained from RF method contributes to the literature. Review of the literature showed us that most of the studies use different variables, however, the effect of those variables on prediction performance are not examined. Therefore, in this study, the effect of variables to the prediction is identified, and they could be useful for the future studies.

In this context, forecasting models with defined methods are built and results are obtained from the datasets. "Special day" variable has been added between predictors to be able to reflect intermittence. Thus, methods have been compared based on the evaluation criteria which is the root mean square error (RMSE). Thereby, results of the study show the influence of the variables other than past sales figures on the forecasting performance.

In the second section of the study, RF and KNN have been explained theoretically. In the following sections, a forecasting study has been made. Providing the reliability of the study, forecasting models have been applied to four datasets, and results have been obtained accordingly. Then, conclusions of the study and suggestions for further studies are given in the last section.

2. MATERIALS AND METHODS

2.1 Random Forest

Random forest algorithm is one of the well-known and most popular methods used in the literature, among decision trees. It is mostly used for classification; however, it has good performance for predicting problems as well. Random forest forms several decision trees, thus, it eliminates the disadvantages of the decision trees [35]. RF was proposed in 2001 by Brieman [36]. The core idea of the RF was selecting random features to create a subset of features to optimize the length of the tree and overfitting [37]. RF tries to combine the information coming from small trees that are created during the training of the algorithm. If the problem type is classification, RF uses medians of the clusters, otherwise RF uses mean of the predictions for regression problems.

RF algorithm uses the Classification and Regression Trees (CART) in the creation of the decision tree. GINI index are used to evaluate performance of the decision trees that are created by CART and to create new branches [38]. Differently, RF doesn't use all features while creating new trees but choose the features randomly to create them. This random selection of the features is based on their performance of providing the most successful splitting. The

GINI index's equation is given in the Equation (1) where p is the data set, j is the frequency and $p(j/t)$ is the probability at the t node for j cluster [36].

$$Gini(t) = 1 - \sum_{j} \left[p\left(\frac{j}{t}\right) \right]^2 \tag{1}$$

RF method has some advantages such as limiting the overfitting, training speed, ability to deal with numerous features; however, optimizing the parameters directly affects the prediction performance. That is the challenging part of the RF applications [35], [37]. RF can be easily implemented to real-world problems and performs well compare to newer methods [39]. Also, it handles the outliers and is good at handling big data, and distribution of the data is irrelevant [40]. General steps of RF can be found as below[37]:

1. Choosing the n sample randomly from dataset that has same size as the dataset or training dataset where 2/3 of this sample to be used as training dataset and 1/3 of it to be used as test dataset.

2. Creating subsets of features by selecting K features from total M features. K can be calculated as K=M/3 for regression and K=M^{1/2} for classification.

3. Creating decision tree by using datasets from step 1 and features from step 2. GINI index is used to evaluate the performance of trees.

4. Repeating the first 3 steps until reaching the maximum tree number and calculating the forecasting error for each tree.

Figure 1 shows the RF algorithm where predictions for different sample datasets and features are obtained, then the average of all trees' prediction is calculated, and finally the calculation of the RF is obtained.

2.2 K-Nearest Neighbor

KNN is a commonly used non-parametric learning algorithm. It has a simple algorithm that searches the K most similar variable vectors within the historical database to make prediction. KNN is sensitive to distance function due to the spontaneous sensitivity to irrelevant features [41].

KNN was proposed by Fix and Hodges in 1951 for the first time as a classification method for nonparametric discriminant analysis; however, its popularity started after the study of Cover and Hart in 1967. In 1977, Stone extended this method to regression problems [3].

The reason for the using KNN can be explained as below [42]:

1. Interpretability of the model. KNN takes into account past data based on their similarity to the current situation according to distance metric. A result of the prediction is calculated as the average of the nearest neighbors.

2. Cyclic factors treatment. KNN algorithm can be tuned to work with cyclical data such as year, month which a sales data would have it.

3. Single learning. KNN algorithm doesn't need to be trained multiple times when new data is introduced to model. Expanding the search instances without recalculating the model would work fine.

Another advantage of the KNN is being fast because it saves all training patterns. Thus, KNN is useful for the problems which have numerous data [43].

KNN uses distance measure to find k nearest neighbors. Therefore, in the application we use Euclidean distance due to its popularity for forecasting problems and high performance compare to other distance measure. Also selecting k is another thing to be done by the user and should be done by considering some experimentation [14].

3. IMPLEMENTATION

RF and KNN are presented to forecast of apparel retailer's demands. Datasets consist of the products which have 12 months sales period and seasonality. Sales that are used in the datasets cover 212 weeks between 2014 and 2018. Both of the methods are applied to datasets separately, and the dataset's information is given in Table 1. Input variables such as color, week, gender, special day and so on, and output variables such as sales unit are given in Table 2 [2]. Color variable's level depends on the datasets and goes up to 10. Two different models are built which are RF and KNN.

The variable type column in Table 2 is taken from R as is. Thus, categorical variables are named as factor and numeric type is used for rational numbers.

RMSE, which is one of the most common performance measures, is calculated according to Equation (2). Here, y_i is the observed data, $\hat{y}_i$ is the predicted value and n is the number of observation [44], [45].

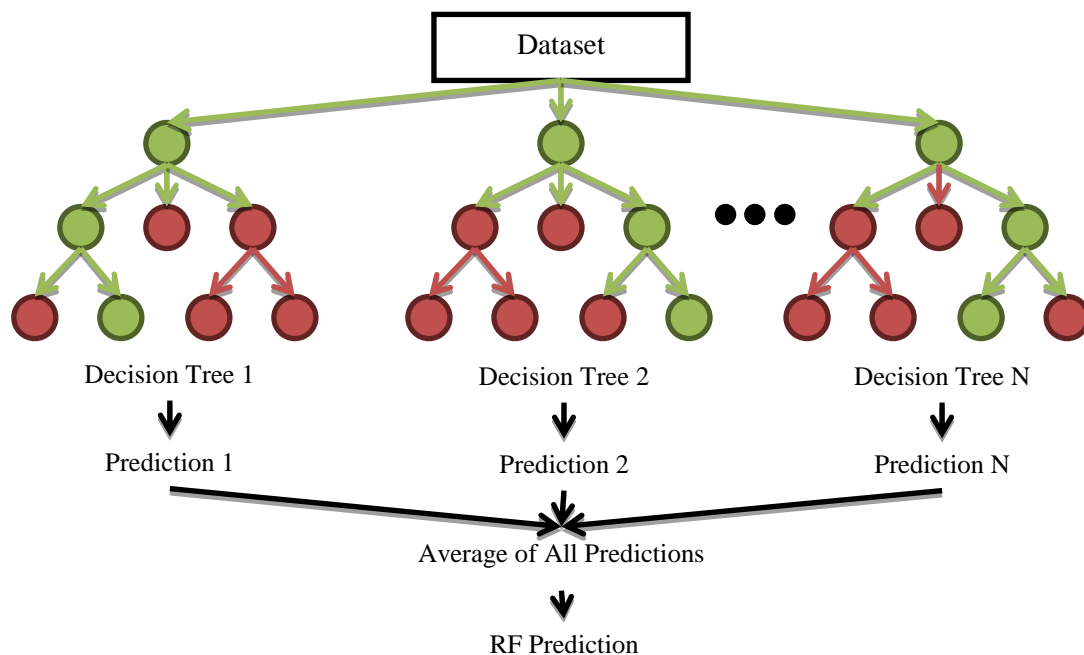$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

$$(14)$$

Figure 1. Random Forest algorithm

Table 1. Datasets and descriptive statistics

Dataset Name	Description	Number of Observation	Mean	Median	Std. Dev.
CNS	Women's basic cardigan	542	690	69	1428
MLN	Women's crew neck basic long sleeve shirt	1590	1099	217	1656
RTK	Men's polo neck basic short sleeve T shirt	1738	1564	150	2722
TNT	Women's basic jersey trousers	703	1376	555	1746

Table 2. Variables used in the study

Variable Name	Description	Variable Type	Level
Color	Color of the product.	Factor	10*
Week	Sales week	Integer	-
Gender	Gender group of the product	Factor	2
Special Day	Special days of the sales week	Factor	2
Weather	Weighted weather for the week in C	Numeric	-
Number of Stores	Number of stores that product has been sold	Integer	-
Minimum Wage	Minimum wage in the sales month (Monthly)	Numeric	-
Clothing Industry PI	Clothing production index (Monthly)	Numeric	-
Number of Tourists	Number of visitors to the country (Monthly)	Integer	-
Spending Per Tourist	Average expenditure per capita (USD-Monthly)	Integer	-
Gold Gram	Gold gram sale price (TL-Monthly)	Numeric	-
CPI Clothing	Consumer price index for clothing (Monthly)	Numeric	-
PPI Clothing	Domestic manufacturer price index for clothing (Monthly)	Numeric	-
CI	Real sector confidence index (Monthly)	Numeric	-
Clothing Capacity	Production level of clothing (Monthly)	Numeric	-
Unemployment	Unemployment rate (Monthly)	Numeric	-
Employment	Employment rate (Monthly)	Numeric	-
VIP	Product status	Factor	2
Clothing Expenses	Bank card and credit card spending amount for clothing and accessories (Weekly-Thousand TL)	Integer	-
USD Currency	US Dollar exchange rate (Weekly-TL)	Numeric	-
Interest	Weighted average interest rates applied to general purpose loans extended to banks (Weekly)	Numeric	-
Loan Given	Loan amount is given (Weekly-Thousand TL)	Integer	-
Ray Stock	Stock quantity at the end of the day in the relevant stores	Integer	-
Gross Profit	Gross profit (TL) obtained from products sold in relevant stores.	Numeric	-
Amount	Amount of product sold (TL)	Numeric	-
Price	Actual sales price of the relevant product	Numeric	-
RSP	Retail sale price of the relevant product	Numeric	-
Store Cover	Rate of stock quantity to sales quantity in the department in the relevant stores during the day	Numeric	-
Sales Quantity	Total number of sales of the product	Integer	-

K-fold cross validation technique is used for both methods to increase learning during the training process of the algorithms. It is although an optional process in general where there is enough data for training and test process, machine learning applications take cross validation serious which allows them to test the results of training data in k amount sub samples according to literature review[46], [47]. Therefore, k-fold cross validation divides the training data into k sub data. Then, one of these k sub data is chosen for validation data and the rest of them, k-1 sub data, are used as training data. Iterations continue k times, thus, each of the sub data is used as validation and training data. As a result of this process, a single estimation can be resulted as a combination of the k folds' results which minimize the forecasting error [47], [48]. According to literature review, generally, k value is set to 10 [48]–[50], thereof, in the study, data sets were randomly divided into 10 sub data, therefore k=10. Thus, 10 different data sets were generated from the training dataset and used during the training process, along with one final hold-out dataset for the final test process after training.

R language version 4.0.3 is used with R studio version 1.3.1093 during the implementation of the algorithms. Caret package, classification and regression training, is used to run the algorithms [51]. The package contains many machine learning methods and has a standard way to use these methods with data preparation tools as well as model tuning tools.

3.1 RF Application

The performance of the RF model depends on the tree number, number of predictors (number of variable - mtry) and node number. Thus, different combinations for each parameter are defined, and RF has been run separately for each parameter with the dataset RTK as trial. Obtained results for different number of predictors and RMSE values are given in Figure 2.

In Table 3, different number of predictors and their RMSE performance are shown. According to Figure 2 and Table 3, the number of randomly selected predictors is defined as 13 for each tree creation.

Table 3. RMSE values for different number of predictors in RF

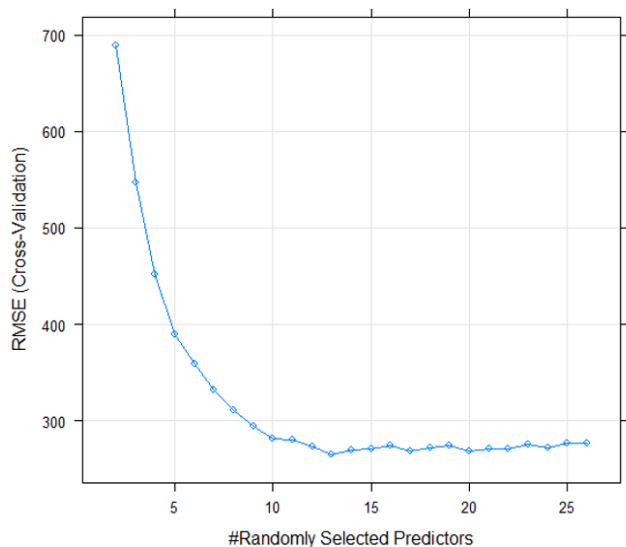| mtry	RMSE	mtry	RMSE	mtry	RMSE
2	688,543	14	269,705	20	268,863
5	389,655	15	271,426	22	271,517
9	294,998	16	274,224	24	272,164
13	**265,579**	18	271,760	26	272,164

Figure 2. RF trials for a different combination of predictors

RMSE performance of RF with different number of trees and number of nodes are given in the Table 4 respectively. As seen in Table 4, the best performance which has the lowest RMSE is obtained with 29 nodes and 550 trees in RF model. Each parameter has been applied to the model separately and the decided value is held steady while other parameters are trained. Thus, parameters for RF model have been determined and used for datasets.

Table 4. RMSE values for different parameters in RF

Number of Node	RMSE	Number of Tree	RMSE
10	535,216	250	332,556
13	469,801	300	331,869
17	394,660	350	331,755
21	387,054	400	331,736
25	357,645	450	331,330
27	338,669	**550**	**330,535**
29	**331,757**	800	331,046
30	333,920	1000	322,856

Selected RF model was run for the all datasets and results were obtained as in Figure 3. Horizontal axis represents the observed data while vertical axis shows the predicted values. Ideally, blue dots which represent the intersection of forecasted values and actual values, must be on the diagonal line where each forecast perfectly represents the actual value. However, due to forecasting error, normally there could be deviations as expected. Nevertheless, it is seen that predictions were quite accurate for all datasets despite some outliers.

The first 200 rows of the observed data and RF predictions for all datasets are shown in Figure 4. Thus, Figure 4 can give an idea about the accuracy of the prediction. Regarding each dataset, RF algorithm was able to make good prediction where the actual data is not extreme, as shown in Figure 4. Also, it is clear that RF performed successful prediction for values with intermittent period. In RTK dataset, where the sales amount is relatively higher than the rest of the data which is between 1000 and 1500, RF forecasting error was high. Thus, RTK has the highest RMSE value as shown in Table 5.

As results of RF model, RMSE values for all datasets are shown in Table 5. It is seen that RF has showed relatively poor prediction performance when observation number goes up among 4 datasets. Therefore, RF has shown better performance for CNS dataset with the lowest RMSE value.
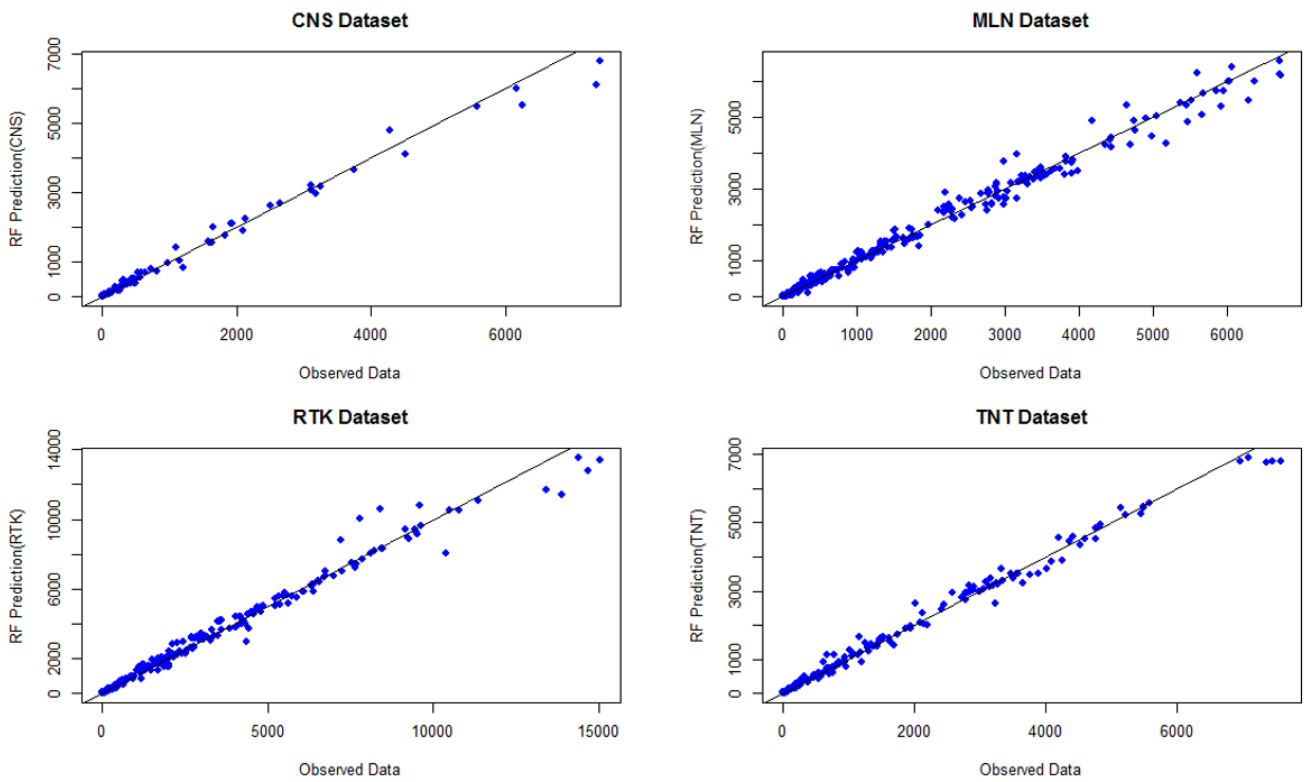
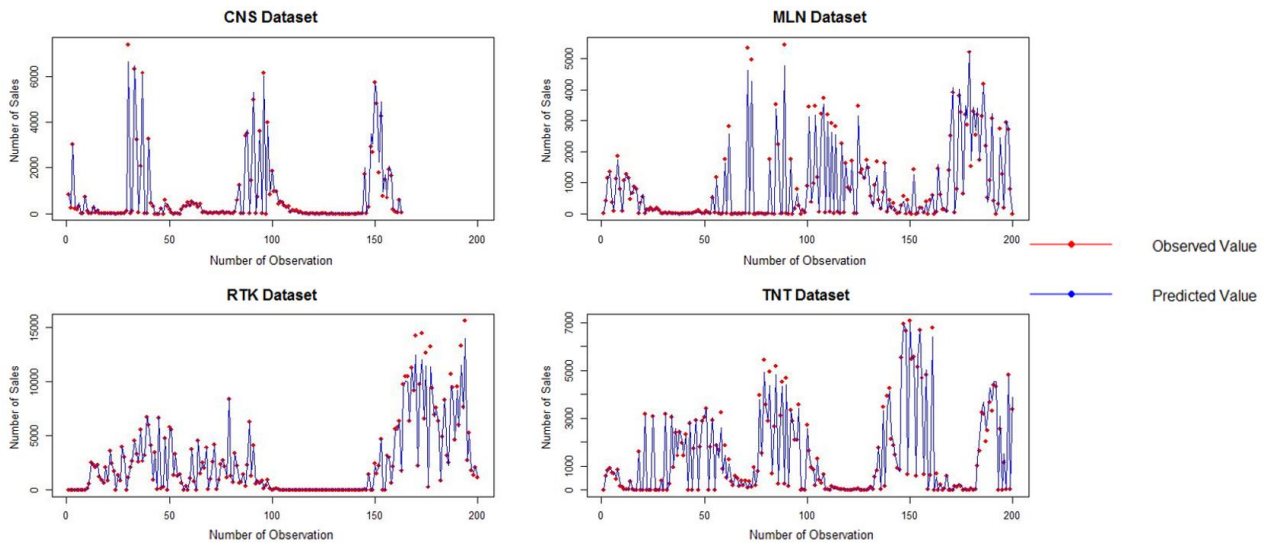**Figure 3.** RF prediction and observed data for all datasets

Figure 4. First 200 observation and RF prediction data for all datasets

Table 5. RF results for all datasets

Data Set	Number of Observations	RMSE
CNS	542	119,305
MLN	1590	227,979
RTK	1738	322,846
TNT	703	166,579

3.1.1 Importance of Variables

As mentioned in the Introduction section, one of the reasons for using RF is to obtain the importance of the variables that are used in the study. This is one of the main aspects of the study since we have used plenty of variables as predictors to be able to make acceptable forecasting. Thus, RF algorithm has provided a list of variables and their effect on the forecast results as in Table 6.

Table 6 is computed from permutation of test data. In the study, the number of mtry which is the number of random variables for each iteration in RF is determined as 13 as shown in Table 3. Thus, randomly 13 predictors have been taken into consideration in each iteration and their effects on the forecasting error is calculated. Accordingly, in Table 6, these effects are shown for each variable on each dataset. Higher value represents higher importance and higher impact on the forecasting error, therefore, "Amount" and "Gross Profit" variables are the most important variables for each dataset since they have the highest effect on the forecasting error. "Number of Stores" variable also seems relatively more important than other variables for all dataset except MLN, since "Week" and "PPI Clothing" has higher values for MLN dataset.

3.2 KNN Application

The most important parameter of KNN is k, and defining k directly affects the KNN performance. Because of this reason KNN has been run with different parameter of k with the dataset RTK as trial, and results are given in Figure 5. RMSE has been taken into account to evaluate the best performance of different k values for KNN.

Some of the RMSE values for different k values are given in Table 7. According to results, the lowest RMSE value is obtained when k is 4 for the KNN model. Randomly selected predictors are defined as 13 for each iteration.
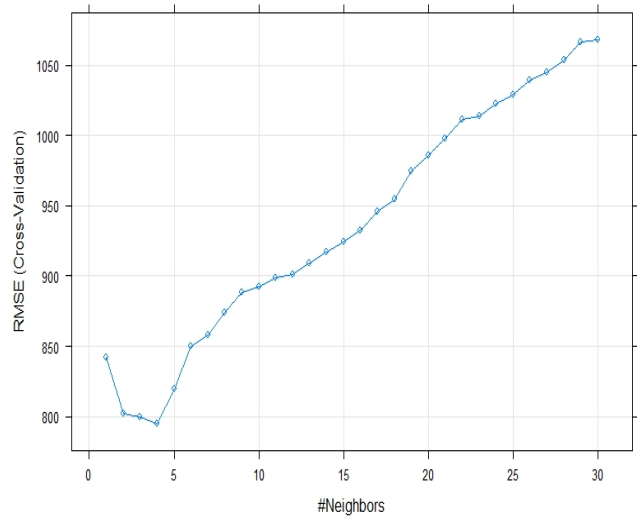

Figure 5. KNN trials for a different combination of k

KNN model was run for the all datasets with k value as 4 and results were obtained as in Figure 6 where the horizontal axis represents the observed data while vertical axis shows the predicted values. Similar to RF method, prediction and real values are presented so that the deviation between forecast and actual data can be seen. Intersection points of forecast values and observed values spread far from the diagonal line which means KNN forecasting performance was poor.

Table 6. Importance of variables obtained from RF for each dataset

Variable Name	CNS Importance	MLN Importance	RTK Importance	TNT Importance
Amount	25.39	26.31	35.34	37.18
Gross Profit	24.17	31.67	23.66	21.21
Number of Stores	11.12	10.73	11.66	11.18
Store Cover	3.87	5.81	3.91	5.32
Ray Stock	3.61	6	5.11	6.29
Color	3.41	9.43	2.87	7.44
Gold Gram	2.98	7.01	9	5.59
Week	2.32	10.86	10.34	6.79
Price	2.27	7.83	3.54	7.28
Interest	2.18	7.46	8.72	8.37
CPI Clothing	2.16	5.54	7.59	5.11
PPI Clothing	2.12	10.74	8.73	5.97
Weather	2.09	4.55	2.9	1.1
Employment	2.08	5.47	2.16	2.84
Unemployment	1.89	2.06	3.91	10.4
Minimum Wage	1.57	2.53	7.86	1.48
RSP	1.46	10.27	7.73	6.05
USD Currency	1.45	5.83	8.34	6.16
Number of Tourist	1.43	6.98	2.86	3.48
Spending Per Tourist	1.26	6.86	5.19	4.18
Clothing Capacity	1.13	6.12	4	5.6
CI	1.09	2.63	2.79	8.22
Clothing Industry PI	1.06	2.83	4.65	3.21
Special Day	0.98	0.9	-1.27	0.06
Loan Given	0.59	9.19	9.5	7.57
Clothing Expenses	-0.5	6.43	3.75	4.8

Table 7. RMSE values for different k values in KNN

k	RMSE	k	RMSE
1	842,345	12	901,343
2	801,919	14	916,754
4	**795,023**	17	945,931
6	849,920	20	985,916
10	892,466	30	1067,924

The first 200 rows of the observed data and KNN predictions for all datasets are shown in Figure 7. Thus, prediction error of the KNN could be seen. When the data given in Figure 7 is examined, KNN was not able to make good prediction overall, regarding each dataset. Distance between observed and predicted values are quite far, thus, leads to higher prediction error. However, when the number of sales is low or zero, KNN was able to catch the pattern of sales and predict it.
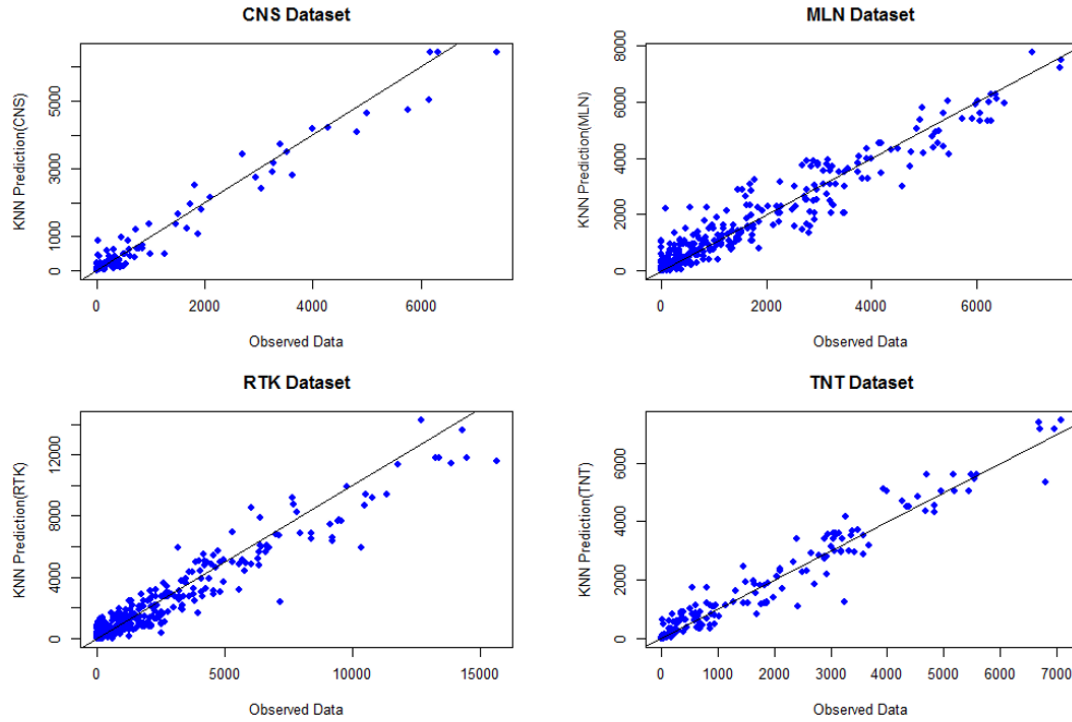

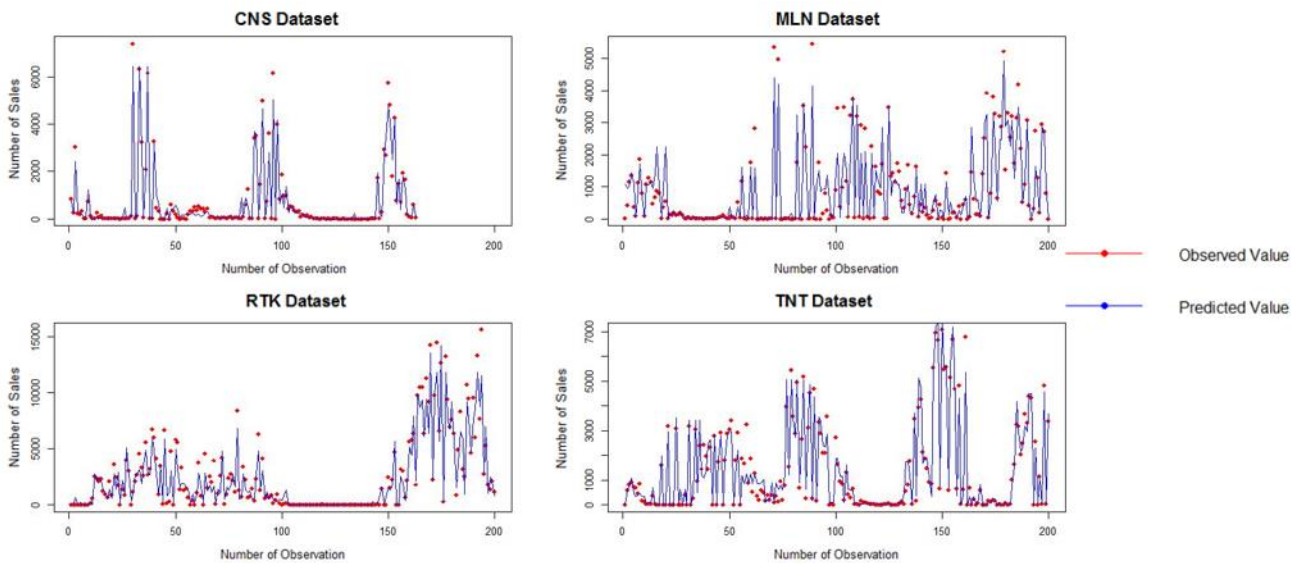**Figure 6.** KNN prediction and observed data for all datasets

Figure 7. First 200 observation and KNN prediction data for all datasets

Obtained results of the KNN models are shown in Table 8 as in RMSE values. Similar to RF method, RMSE values increase when the number of observations high. Therefore, KNN has shown better performance for CNS dataset with the lowest RMSE value.

Data Set	Number of Observations	RMSE
CNS	542	263,260
MLN	1590	442,169
RTK	1738	732,553
TNT	703	390,341

4. RESULTS AND DISCUSSION

The results of both RF and KNN models clearly indicate that RF has outperformed KNN in regard of the RMSE value. Comparison of the RMSE values for RF and KNN models for all datasets is shown in the Figure 9.

According to Figure 9, RF has performed better than KNN for all datasets and in terms of relatively big data. It is seen that, KNN has poorly performed in RTK dataset which has higher standard of deviation than other datasets, as it is 2722.

In Table 9, total sales amount of the test data as observation value and total prediction values obtained with each method for the same test data are given. In the percentage column, deviations of prediction values from observation values are shown. As it is seen from the table deviation between sales and prediction is reasonable despite the RMSE values were obtained high because total sales amounts are excessively high. Considering the demand type in apparel retail and high standard deviation between sales periods in the data as a result of demand type, algorithms have had difficulties to predict sharp transitions between sales where they have changed from hundreds to thousands and then to hundreds again. However, low deviation rate shows the success of these algorithms.

5. CONCLUSION

Forecasting customer demands is necessary to be able to make short-term plans regarding inventory, labor, transportation and other expenses. Additionally, considering production and sourcing cost, companies place so much importance on making correct prediction. If customers cannot find goods in stores or the goods in stores do not appeal to their taste, sales lost may occur, thus the company would lose sales and prestige. Thus, unsold goods would remain as stock and intervene next year purchase while bringing liability to the company. Forecasting future demands is a key factor of the companies for survival.

Forecasting is a challenging job for all kinds of sectors, but complexity and intermittent sales make it thorny in apparel retail. Numerous products' range doesn't favor this challenging task either. Traditional methods work partly well in demand forecast where the product range is limited and factors that have an effect on the demand are limited and can be controlled. In this way they are not applicable for most of the real world problems and big data where there are so many variables affecting the demand. Thus, in this study, 28 variables that may affect the sales have been taken into consideration to reflect the complexity of the retail apparel in using two commonly used machine learning models as RF and KNN. Established models have been run with four different datasets and analysis has been carried out on the product basis to the color level.

Results of the study showed that RF model outperformed KNN in four datasets out of four. That means RF has absolute success over the KNN. Average RMSE of RF was obtained as 209 while the value for KNN is 457, thus, RF's performance was better 118% than KNN. It has been revealed that RF is more successful to make accurate forecasting when the predictors are in high numbers which is up to 28. Also, it needs to be indicated that intermittent demand and sharp transitions between sales terms are highly challenging prediction to do so, and in order to increase success of the prediction, variables which present intermittency should be chosen well.
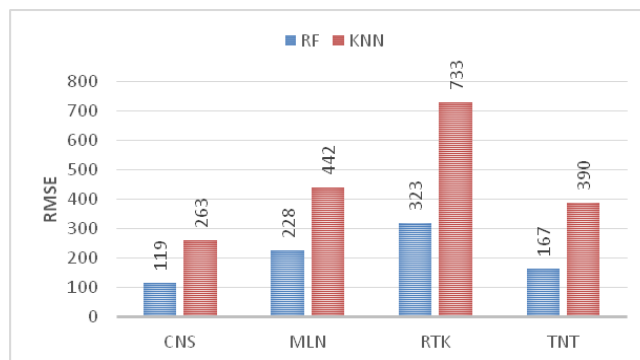

Figure 9. RMSE results of both RF and KNN algorithms

Table 9. Comparison between methods based on some of test data and predictions for all datasets

Dataset	Observation Value	RF Prediction			KNN Prediction		
		RMSE	Prediction	%	RMSE	Prediction	%
CNS	112400	119.31	116530.09	*3.67%*	263.26	108124.25	*-3.80%*
MLN	524406	227.98	523216.83	*-0.23%*	442.17	539464.40	*2.87%*
RTK	816417	322.85	804360.20	*-1.48%*	732.55	760202.75	*-6.89%*
TNT	290253	166.58	294534.88	*1.48%*	390.34	299547.25	*3.20%*

Another important output of this study can be stated as obtaining the importance of the variables. Thus, the variables have more effect on the prediction performance than others are revealed. Therefore, these variables can be used as predictors for other studies, and it is believed these variables are good inferences for future studies. Thus, researchers can inspect the relation between these variables and comment them in their studies. Also, the variables have negative impact on the forecast performance should be taken out of the model. It is important to note that each dataset has its own characteristic, therefore, demand for each of them can be varied. Considering the fact, one variable can be important to one particular dataset while not important to the other one. Therefore, researchers should take into account dataset's descriptive statistics while following this paper in terms of determining the predictors.

The main purpose of this study was to consider a vast amount of the predictors and their behavior during the forecasting process in two different non-parametric methods with the intermittent demand environment. Obtaining knowledge of the suitable predictors that contribute to accuracy is also an important outcome of the study and a significant contribution to the literature. Therefore, the most important point of the study is to see the prediction of the zero-sales period. In Figure 4 and Figure 7, it is seen where the sales are down to zero, that machine learning models were good enough to catch intermittence in all datasets and adapt the prediction accordingly. However, the study has been made in the apparel retailer where there might be other key predictors to be found out, that influence the sales, in this sense, future studies might investigate this. The reason of the intermittence of the demand can be analyzed and predictors can be added to the models. Considering the size of the data in retail, deep learning methods can be applied to build more accurate models. Also, efficiency of the predictors can be investigated by further studies to increase accuracy.

Acknowledgement

REFERENCES

1. J. Huber and H. Stuckenschmidt, 2020. Daily retail demand forecasting using machine learning with emphasis on calendric special days. *Int. J. Forecast.* vol. 36, no. 4. pp. 1420–1438. doi: 10.1016/j.ijforecast.2020.02.005.

2. İ. Güven and F. Şimşir, 2020. Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods. *Comput. Ind. Eng.* vol. 147. p. 106678. doi: 10.1016/j.cie.2020.106678.

3. M. Kück and M. Freitag, 2021. Forecasting of customer demands for production planning by local k - nearest neighbor models. *Int. J. Prod. Econ.* vol. 231. p. 107837. doi: 10.1016/j.ijpe.2020.107837.

4. C.-H. Wang and J.-Y. Chen, 2019. Demand forecasting and financial estimation considering the interactive dynamics of semiconductor supply-chain companies. *Comput. Ind. Eng.* vol. 138. p. 106104. doi: 10.1016/j.cie.2019.106104.

5. P. Vroman, M. Happiette, and B. Rabenasolo, 1998. Fuzzy Adaptation of the Holt–Winter Model for Textile Sales-forecasting. *J. Text. Inst.* vol. 89, no. 1. pp. 78–89. doi: 10.1080/00405009808658668.

6. S. Thomassey, M. Happiette, N. Dewaele, and J. M. Castelain, 2002. A Short and Mean Term Forecasting System Adapted to Textile Items' Sales. *J. Text. Inst.* vol. 93, no. 3. pp. 95–104. doi: 10.1080/00405000208658360.

7. C. Li and A. Lim, 2018. A greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing. *Eur. J. Oper. Res.* vol. 269, no. 3. pp. 860–869. doi: 10.1016/j.ejor.2018.02.029.

8. K. Nikolopoulos, 2020. We need to talk about intermittent demand forecasting. *Eur. J. Oper. Res.* doi: 10.1016/j.ejor.2019.12.046.

9. S. Beheshti-Kashi, H. R. Karimi, K.-D. Thoben, M. Lütjen, and M. Teucke, 2015. A survey on retail sales forecasting and prediction in fashion markets. *Syst. Sci. Control Eng.* vol. 3, no. 1. pp. 154–161. doi: 10.1080/21642583.2014.999389.

10. S. Ma and R. Fildes, 2021. Retail sales forecasting with meta-learning. *Eur. J. Oper. Res.* vol. 288, no. 1. pp. 111–128. doi: 10.1016/j.ejor.2020.05.038.

11. A. L. D. Loureiro, V. L. Miguéis, and L. F. M. da Silva, 2018. Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decis. Support Syst.* vol. 114. pp. 81–93. doi: 10.1016/j.dss.2018.08.010.

12. P.-S. Yu, T.-C. Yang, S.-Y. Chen, C.-M. Kuo, and H.-W. Tseng, 2017. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *J. Hydrol.* vol. 552. pp. 92–104. doi: 10.1016/j.jhydrol.2017.06.020.

13. R. Cui, S. Gallino, A. Moreno, and D. J. Zhang, 2018. The Operational Value of Social Media Information. *Prod. Oper. Manag.* vol. 27, no. 10. pp. 1749–1769. doi: 10.1111/poms.12707.

14. F. Martínez, M. P. Frías, M. D. Pérez-Godoy, and A. J. Rivera, 2018. Dealing with seasonality by narrowing the training set in time series forecasting with k NN. *Expert Syst. Appl.* vol. 103. pp. 38–48. doi: 10.1016/j.eswa.2018.03.005.

15. D. V Souza *et al.*, 2019. k -Nearest Neighbor Regression in the Estimation of Tectona G randis Trunk Volume in the State of Pará, Brazil. *J. Sustain. For.* vol. 38, no. 8. pp. 755–768. doi: 10.1080/10549811.2019.1607391.

16. M. Ghiassi, F. Fa'al, and A. Abrishamchi, 2017. Large metropolitan water demand forecasting using DAN2, FTDNN, and KNN models: A case study of the city of Tehran, Iran. *Urban Water J.* vol. 14, no. 6. pp. 655–659. doi: 10.1080/1573062X.2016.1223858.

17. N. J. Johannesen, M. Kolhe, and M. Goodwin, 2019. Relative evaluation of regression tools for urban area electrical energy demand forecasting. *J. Clean. Prod.* vol. 218. pp. 555–564. doi: 10.1016/j.jclepro.2019.01.108.

18. T. Fang and R. Lahdelma, 2016. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Appl. Energy.* vol. 179. pp. 544–552. doi: 10.1016/j.apenergy.2016.06.133.

19. M. Sebri, 2016. Forecasting urban water demand: A meta-regression analysis. *J. Environ. Manage.* vol. 183. pp. 777–785. doi: 10.1016/j.jenvman.2016.09.032.

20. F.-L. Chu, 2014. Using a logistic growth regression model to forecast the demand for tourism in Las Vegas. *Tour. Manag. Perspect.* vol. 12. pp. 62–67. doi: 10.1016/j.tmp.2014.08.003.

21. P. Ramos, N. Santos, and R. Rebelo, 2015. Performance of state space and ARIMA models for consumer retail sales forecasting. *Robot. Comput. Integr. Manuf.* vol. 34. pp. 151–163. doi: 10.1016/j.rcim.2014.12.015.

22. W. Anggraeni, R. A. Vinarti, and Y. D. Kurniawati, 2015. Performance Comparisons between Arima and Arimax Method in Moslem Kids Clothes Demand Forecasting: Case Study. *Procedia Comput. Sci.* vol. 72. pp. 630–637. doi: 10.1016/j.procs.2015.12.172.

23. I. Khandelwal, R. Adhikari, and G. Verma, 2015. Time Series Forecasting Using Hybrid ARIMA and ANN Models Based on DWT Decomposition. *Procedia Comput. Sci.* vol. 48. pp. 173–179. doi: 10.1016/j.procs.2015.04.167.

24. L. A. Díaz-Robles *et al.*, 2008. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmos. Environ.* vol. 42, no. 35. pp. 8331–8340. doi: 10.1016/j.atmosenv.2008.07.020.

25. M. Khashei and M. Bijari, 2011. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Appl. Soft Comput.* vol. 11, no. 2. pp. 2664–2675. doi: 10.1016/j.asoc.2010.10.015.

26. W. J. Wang and Q. Xu, 2014. A Bayesian Combination Forecasting Model for Retail Supply Chain Coordination. *J. Appl. Res. Technol.* vol. 12, no. 2. pp. 315–324. doi: 10.1016/S1665-6423(14)72347-8.

27. A. Kulshrestha, V. Krishnaswamy, and M. Sharma, 2020. Bayesian BILSTM approach for tourism demand forecasting. *Ann. Tour. Res.* vol. 83. p. 102925. doi: 10.1016/j.annals.2020.102925.

28. F.-L. Chu, 2008. A fractionally integrated autoregressive moving average approach to forecasting tourism demand. *Tour. Manag.* vol. 29, no. 1. pp. 79–88. doi: 10.1016/j.tourman.2007.04.003.

29. K. Nakade and Y. Aniyama, 2019. Bullwhip Effect of Weighted Moving Average Forecast under Stochastic Lead Time. *IFAC-PapersOnLine.* vol. 52, no. 13. pp. 1277–1282. doi: 10.1016/j.ifacol.2019.11.374.

30. Z. Michna, S. M. Disney, and P. Nielsen, 2020. The impact of stochastic lead times on the bullwhip effect under correlated demand and moving average forecasts. *Omega.* vol. 93. p. 102033. doi: 10.1016/j.omega.2019.02.002.

31. G. Sbrana and A. Silvestrini, 2019. Random switching exponential smoothing: A new estimation approach. *Int. J. Prod. Econ.* vol. 211. pp. 211–220. doi: 10.1016/j.ijpe.2019.01.038.

32. T. M. Dantas and F. L. Cyrino Oliveira, 2018. Improving time series forecasting: An approach combining bootstrap aggregation, clusters and exponential smoothing. *Int. J. Forecast.* vol. 34, no. 4. pp. 748–761. doi: 10.1016/j.ijforecast.2018.05.006.

33. 33. G. Sbrana and A. Silvestrini, 2013. Forecasting aggregate demand: Analytical comparison of top-down and bottom-up approaches in a multivariate exponential smoothing framework. *Int. J. Prod. Econ.* vol. 146, no. 1. pp. 185–198. doi: 10.1016/j.ijpe.2013.06.022.

34. R. Tsaur, 2009. Seasonal forecasting of a decomposed fuzzy exponential smoothing model using grey estimated values. *J. Chinese Inst. Eng.* vol. 32, no. 1. pp. 17–31. doi: 10.1080/02533839.2009.9671479.

35. Y. Zhu, W. XU, G. Luo, H. Wang, J. Yang, and W. Lu, 2020. Random Forest enhancement using improved Artificial Fish Swarm for the medial knee contact force prediction. *Artif. Intell. Med.* vol. 103. p. 101811. doi: 10.1016/j.artmed.2020.101811.

36. E. Izquierdo-Verdiguier and R. Zurita-Milla, 2020. An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* vol. 88. p. 102051. doi: 10.1016/j.jag.2020.102051.

37. X. Zhou, P. Lu, Z. Zheng, D. Tolliver, and A. Keramati, 2020. Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree. *Reliab. Eng. Syst. Saf.* vol. 200. p. 106931. doi: 10.1016/j.ress.2020.106931.

38. J. E. Pesantez, E. Z. Berglund, and N. Kaza, 2020. Smart meters data for modeling and forecasting water demand at the user-level. *Environ. Model. Softw.* vol. 125. p. 104633. doi: 10.1016/j.envsoft.2020.104633.

39. M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, 2014. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *J. Mach. Learn. Res.* vol. 15, no. 90. pp. 3133–3181. [Online]. Available: http://jmlr.org/papers/v15/delgado14a.html.

40. 40. M. Aghaabbasi, Z. A. Shekari, M. Z. Shah, O. Olakunle, D. J. Armaghani, and M. Moeinaddini, 2020. Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques. *Transp. Res. Part A Policy Pract.* vol. 136. pp. 262–281. doi: 10.1016/j.tra.2020.04.013.

41. X. Wang, K. An, L. Tang, and X. Chen, 2015. Short Term Prediction of Freeway Exiting Volume Based on SVM and KNN. *Int. J. Transp. Sci. Technol.* vol. 4, no. 3. pp. 337–352. doi: 10.1260/2046-0430.4.3.337.

42. E. Mangalova and E. Agafonov, 2014. Wind power forecasting using the k-nearest neighbors algorithm. *Int. J. Forecast.* vol. 30, no. 2. pp. 402–406. doi: 10.1016/j.ijforecast.2013.07.008.

43. L. A. Teixeira and A. L. I. de Oliveira, 2010. A method for automatic stock trading combining technical analysis and nearest neighbor classification. *Expert Syst. Appl.* vol. 37, no. 10. pp. 6885–6890. doi: 10.1016/j.eswa.2010.03.033.

44. Z. Pang, F. Niu, and Z. O'Neill, 2020. Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. *Renew. Energy.* vol. 156. pp. 279–289. doi: 10.1016/j.renene.2020.04.042.

45. H. Tang, P. Dong, and Y. Shi, 2019. A new approach of integrating piecewise linear representation and weighted support vector machine for forecasting stock turning points. *Appl. Soft Comput.* vol. 78. pp. 685–696. doi: 10.1016/j.asoc.2019.02.039.

46. A. L. Pomerantsev and O. Y. Rodionova, 2021. Procrustes Cross-Validation of short datasets in PCA context. *Talanta.* vol. 226. p. 122104. doi: https://doi.org/10.1016/j.talanta.2021.122104.

47. A. Rohani, M. Taki, and M. Abdollahpour, 2018. A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). *Renew. Energy.* vol. 115. pp. 411–422. doi: https://doi.org/10.1016/j.renene.2017.08.061.

48. Z. Xiong, Y. Cui, Z. Liu, Y. Zhao, M. Hu, and J. Hu, 2020. Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. *Comput. Mater. Sci.* vol. 171. p. 109203. doi: https://doi.org/10.1016/j.commatsci.2019.109203.

49. F. Maepa, R. S. Smith, and A. Tessema, 2021. Support vector machine and artificial neural network modelling of orogenic gold prospectivity mapping in the Swayze greenstone belt, Ontario, Canada. *Ore Geol. Rev.* vol. 130. p. 103968. doi: https://doi.org/10.1016/j.oregeorev.2020.103968.

50. A. Kolus, D. Imbeau, P.-A. Dubé, and D. Dubeau, 2015. Adaptive neuro-fuzzy inference systems with k-fold cross-validation for energy expenditure predictions based on heart rate. *Appl. Ergon.* vol. 50. pp. 68–78. doi: https://doi.org/10.1016/j.apergo.2015.03.001.

51. M. Kuhn, 2008. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* vol. 28, no. 5. doi: 10.18637/jss.v028.i05.