

# Demand Forecasting of Short Life Cycle Products using Data Mining Techniques

Ashraf A. Afifi<sup>1,2</sup>

<sup>1</sup> Department of Engineering, Design and Mathematics, Faculty of Environment and Technology, University of the West of England, Bristol, UK

<sup>2</sup> Industrial Engineering Department, Faculty of Engineering, Zagazig University, Zagazig, Egypt

Ashraf.Afifi@uwe.ac.uk

**Abstract.** Products with short life cycles are becoming increasingly common in many industries due to higher levels of competition, shorter product development time and increased product diversity. Accurate demand forecasting of such products is crucial as it plays an important role in driving efficient business operations and achieving a sustainable competitive advantage. Traditional forecasting methods are inappropriate for this type of products due to the highly uncertain and volatile demand and the lack of historical sales data. It is therefore critical to develop different forecasting methods to analyse the demand trend of these products. This paper proposes a new data mining approach based on the incremental *k*-means clustering algorithm and the RULES-6 rule induction classifier for forecasting the demand of short life cycle products. The performance of the proposed approach is evaluated using real data from one of the leading Egyptian companies in IT ecommerce and retail business, and results show that it has the capability to accurately forecast demand trends of new products with no historical sales data.

**Keywords:** Demand forecasting, Short life cycle products, Data mining, Clustering, Rule induction.

## 1 Introduction

Forecasting is an important and necessary tool that helps managers make decisions about future resourcing of their organisations. It plays a pivotal role in the effective planning of various operations in an organisation including production, inventory, budget, sales, personnel and facilities. Accurate forecasts can lead to significant cost savings, reduced inventory levels, improved customer satisfaction and increased competitiveness.

The fast pace of new product introduction continually drives product life cycles shorter. Products with life cycles of few weeks to few months are very common in fashion (e.g., toys and clothing) and high technology (e.g., mobile phones, computers and consumer electronics) retail industries. A typical demand pattern for such products is characterised by rapid growth, maturity, and decline phases [1]. The demand is also

highly uncertain and volatile, particularly in the introduction stage. An additional problem is the inadequacy of historical data due to the short period of sales. In case of new products, there is complete unavailability of any previous data related to the sales of such products. These characteristics make it difficult to forecast the demand of short life cycle products.

Traditionally, demand forecasting is accomplished by statistical methods such as moving averages, exponential smoothing, Bayesian analysis, regression models, Holt-Winters and Box-Jenkins methods, and autoregressive integrated moving average (ARIMA) models [2]. Statistical methods are popular because of their simplicity and fast speed, and they provide satisfactory results in many forecasting applications. However, these methods are not designed for application in the short life cycle environment, especially most of them require large historical data to ensure accurate estimation of their parameters and they are limited to linear relations [3, 4].

Artificial intelligence (AI) methods such as neural networks, evolutionary algorithms, support vector machines, and fuzzy inference systems are widely used in forecasting activities [5, 6]. AI methods cope well with complexity and uncertainty, and they have better forecasting accuracy compared to statistical methods. However, they usually require a long computational time which makes them less appealing to the fast changing market of fashion and high technology products [7, 8].

Recently, various hybrid methods such as neural fuzzy systems are proposed in the literature to enhance demand forecasting [9, 10]. Hybrid methods utilise the strengths of different models to form a new forecasting method. They can learn complex relations in an uncertain environment and many of them are considered to be more accurate and efficient than the pure statistical and AI models [11].

In an attempt to cope with the lack of historical data of short life cycle products, some references use the data of similar products for which sufficient history is available to forecast the demand of new products [12–16]. For example, a hybrid method based on the  $k$ -means clustering technique and a decision tree classifier was developed to estimate the demand of textile fashion products [12]. A similar approach that uses the self-organizing map and the neural networks techniques was introduced in [13]. This paper reports on a new forecasting method using alternative data mining techniques to improve demand forecasting in the context of a large retail company. In particular, the incremental  $k$ -means clustering technique [17] and the RULES-6 classification rule learning algorithm [18] are employed. These methods are simple, effective and computationally efficient, which make them powerful and practical tools for retail sales forecasting.

The outline of this paper is as follow. Section 2 reviews the basic concepts of data mining with particular focus on clustering and rule induction methods. Section 3 presents the proposed forecasting approach. Section 4 reports and analyses the experimental results obtained with a real retail data. Section 5 concludes the paper and provides suggestions for future work.

## 2 Overview of Data Mining

Data mining, or knowledge discovery in databases (KDD), aims at processing large data into knowledge bases for better decision making. It has been successfully used in many applications to uncover hidden patterns and predict future trends and behaviours. There are three main steps in data mining, namely, data preparation, data modelling, and post processing and model evaluation [19–21]. This section gives a brief description of the data modelling techniques employed in this research.

### 2.1 Clustering Techniques

Clustering techniques are concerned with partitioning of data sets into several homogeneous clusters. These techniques assign a large number of data objects to a relatively small number of groups so that data objects in a group share the same properties while, in different groups, they are dissimilar. Many clustering techniques have been proposed over the years from different research disciplines [22, 23]. *K*-means is one of the best known and commonly used clustering algorithms. The algorithm forms  $k$  clusters that are represented by the mean value of the data points belonging to each of them. This is an iterative process that searches for a division of data objects into  $k$  clusters to minimise the sum of Euclidean distances between each object and its closest cluster centre.

The  $k$ -means algorithm is relatively scalable and efficient in clustering large data sets because its computational complexity grows linearly with the number of data objects. However, it is sensitive to the initial selection of cluster centres and requires the number of clusters  $k$  to be specified before the clustering process starts. The algorithm has been improved to address many of its deficiencies [17, 24, 25]. In particular, a new version called incremental  $k$ -means was introduced to reduce the dependence of the  $k$ -means algorithm on the initialisation of cluster centres [17]. To validate the robustness of the new algorithm it has been tested on a number of artificial and real datasets. The results showed clearly that incremental  $k$ -means consistently outperforms the original algorithm [17]. Therefore, this algorithm is applied in this research to search for interesting and natural clusters in the retail data.

The incremental  $k$ -means algorithm is summarised in Figure 1. The algorithm starts initially with one cluster with the number of clusters  $k$  being incremented by 1 at each step thereafter. With each increase of  $k$ , a new cluster centre is inserted into the cluster with the highest distortion, and the objects are reassigned to clusters until the centres stop “moving”. The process is repeated until  $k$  reaches the specified number of clusters.

### 2.2 Classification Learning Techniques

Classification learning employs a set of pre-categorised data objects to develop a model that can be used to classify new data objects from the same population or to provide a better understanding of the data objects’ characteristics. Among the various classification learning techniques developed, inductive learning may be the most commonly used in real world applications [26]. The inductive learning techniques are relatively fast

Assign  $k = 1$ .  
 Phase 1: *Normal training*  
 Step 1: If  $k = 1$ , choose an arbitrary point for a cluster centre.  
         If  $k > 1$ , insert the centre of the new cluster in the cluster with the greatest distortion.  
 Step 2: Assign each object in the training set to the closest cluster and update its centre.  
 Step 3: If the cluster centre does not move, go to Phase 2.  
         Else, go to Phase 1, Step 2.  
 Phase 2: *Increasing the number of clusters*  
         If  $k$  is smaller than a specified value, increase  $k$  by 1 and go to Phase 1 – Step 1.  
         Else, stop.

**Fig. 1.** A pseudo-code description of the incremental  $k$ -means algorithm [17].

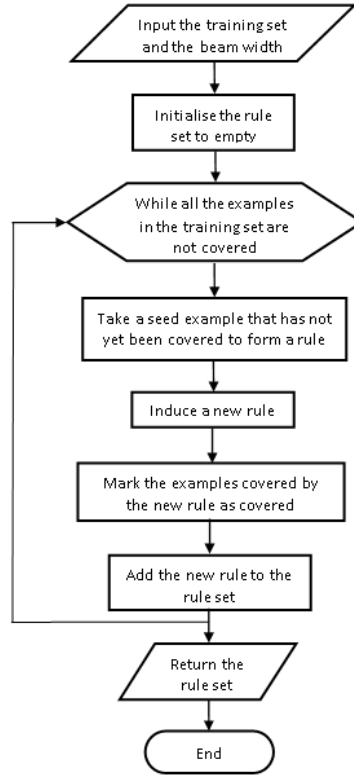
compared to other techniques. Another advantage is that they are simpler and the models that they generate are easier to understand.

In this study, a simple inductive learning algorithm called RULES-6 (RULe Extraction System – Version 6) [18] is used. RULES-6 extracts a set of classification rules from a collection of examples, each belonging to one of a number of given classes. The examples together with their associated classes constitute the set of training examples from which the algorithm generates the rules. Every example is described in terms of a fixed set of attributes, each with its own set of possible values.

In RULES-6, an attribute-value pair constitutes a condition. If the number of attributes is  $N_a$ , a rule may contain between one and  $N_a$  conditions, each of which must be a different attribute-value pair. Only conjunction of conditions is permitted in a rule and hence the attributes must all be different if the rule comprises more than one condition.

RULES-6 works in an iterative fashion. In each iteration, it takes a “seed” example not covered by previously created rules to form a new rule. Having found a rule, RULES-6 marks those examples that are covered by it and appends the new rule to its rule set. The algorithm stops when all examples in the training set are covered. In order to avoid the overfitting problem and to generate simple rules, the rules are pruned with a post pruning strategy [27]. A simplified description of RULES-6 is given in Figure 2.

To form a rule, RULES-6 performs a general-to-specific beam search for a rule that optimises a given quality criterion. It starts with the most general rule and specialises it in steps considering only conditions extractable from the selected seed example. The aim of specialisation is to construct a rule that covers as many examples from the target class and as few examples from the other classes as possible, while ensuring that the seed example remains covered. As a consequence, simpler rules that are not consistent, but are more accurate for unseen data, can be learned. RULES-6 uses effective search-space pruning rules to avoid useless specialisations and to terminate a non-productive search during rule construction. This substantially increases the efficiency of the learning process. It deals with continuous-value attributes using a pre-processing discretisation method. With this method, the range of each attribute is split into a number of smaller intervals that are then regarded as nominal values. A detailed description of the process by which RULES-6 induces a rule can be found in [18].



**Fig. 2.** A simplified description of RULES-6 [18].

### 3 Proposed Forecasting Approach

In this paper, it is proposed to forecast the demand of short life cycle products by applying data mining techniques. This forecasting approach includes the following steps (Fig. 3):

1. Preparing and transforming the available sales data into a format suitable for data mining techniques.
2. Using the incremental  $k$ -means algorithm to identify the sales profiles of historical products by grouping together the related time series of the products sales.
3. Applying the RULES-6 algorithm to build a classification model that links the descriptive attributes of the historical products and the sales profiles of the discovered groups in Step 2.
4. Using the classification model generated in Step 3 to forecast the sales profiles of the new products based on their descriptive attributes.

This section discusses the techniques applied in this research to realise the steps of the proposed approach.

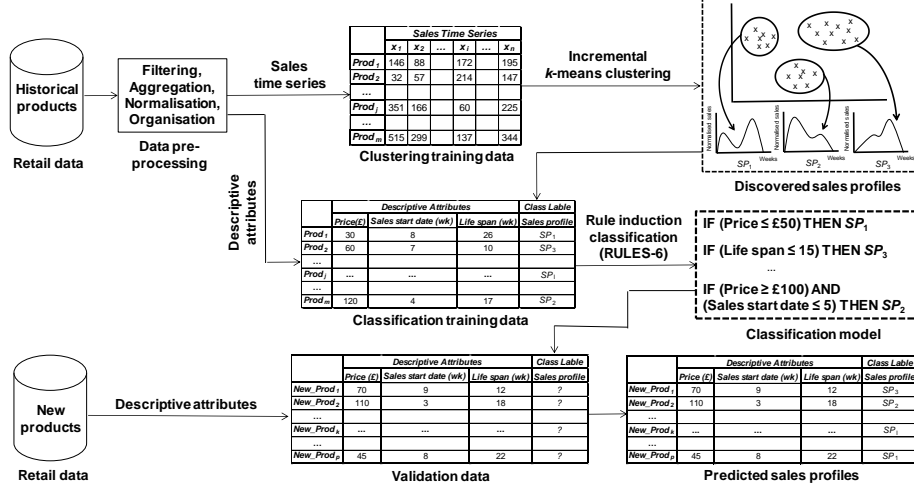


Fig. 3. A simplified description of the proposed forecasting approach.

### 3.1 Data Preparation

The data used in this research was obtained from a large retail Egyptian company called 2B that has 37 branches all over Egypt. The company sells a wide range of high technology products (e.g., laptops, mobiles, tablets, gaming, networking, cables, software, laptop and mobile accessories) that have a short life cycle ranging from a few weeks to a few months. Two sets of data were extracted. The first contains 507 products corresponding to year 2015 and it is used for the learning process. The second is composed of 253 products for year 2016 and it is used for testing. The data is available weekly and covers a period of 26 weeks. The selected descriptive attributes of the products are the price, the starting date of the sales and the life span of each product.

High quality data is a prerequisite for applying any data mining technique [28]. Prior to data modelling, the data needs to be prepared. The objective at this stage is two-fold: to convert the data into a format required by the data mining algorithms and also to expose as much information as possible to data modelling. To prepare the available data for further analysis the following pre-processing steps are implemented.

1. Aggregate the sales for each product into weekly time buckets in order to cope with high uncertainty and volatility in sales patterns.
2. Filter the data by removing the negative sales numbers for customer returns.
3. Normalise the sales volume of all products during the sales period to enable comparison between the time series of products having different sales volumes. The normalised sales at period  $i$ ,  $y_i$ , is computed by dividing the sale at period  $i$ ,  $x_i$ , by the sum of sales during the sales period:  $y_i = x_i / \sum_{k=1}^n x_k$  where  $n$  is the number of periods.
4. Normalise the life span of all products to 26 weeks to allow comparison between the time series of products having different life spans. Normalisation of the life span is computed using homothetic transformation [12].

5. Organise the time series sales and descriptive attributes of the products into a data object format that is suitable for data mining algorithms (examples of the structure of the training and validation data sets created in this way are shown in Fig. 3).

### 3.2 Data Modelling

This section discusses the data mining techniques that were used to analyse the retail data. The analysis is performed in two stages. First, the time series sales of historical products are grouped into several homogeneous clusters. Second, the sales profiles of the new products are predicted by mapping them to the clusters using the descriptive attributes.

**Sales Profiles Identification for Historical Products.** The incremental  $k$ -means clustering algorithm is applied on the pre-processed retail data to discover groups of products with similar time series sales. Each cluster centre defines a sales profile characterising the sales behaviour of the products belonging to the cluster. The distortion error was used to evaluate the clustering results [17], with a lower value of this measure indicative of better quality of clustering.

The incremental  $k$ -means algorithm requires users to specify a number of parameters, namely, the number of clusters and the termination conditions for stopping the clustering process. To find a satisfactory clustering result, a number of iterations were conducted where the algorithm was executed with different values of  $k$ , the number of clusters, and the  $k$  value producing the lowest distortion error was selected [17]. In this work,  $k$  values in the range of 2 to  $\sqrt{n}$ , where  $n$  is the size of the training data set [29], were considered and the optimal value was found to be 13. The clustering process could be stopped by specifying termination conditions such as a predefined number of iterations and the percentage reduction of the distortion errors in one iteration being smaller than a given value  $\varepsilon$ . In this work, these two termination criteria were used. In particular, the maximum number of iterations was set to 50 and  $\varepsilon$  to  $10^{-7}$  to stop the search process when one of these conditions is satisfied.

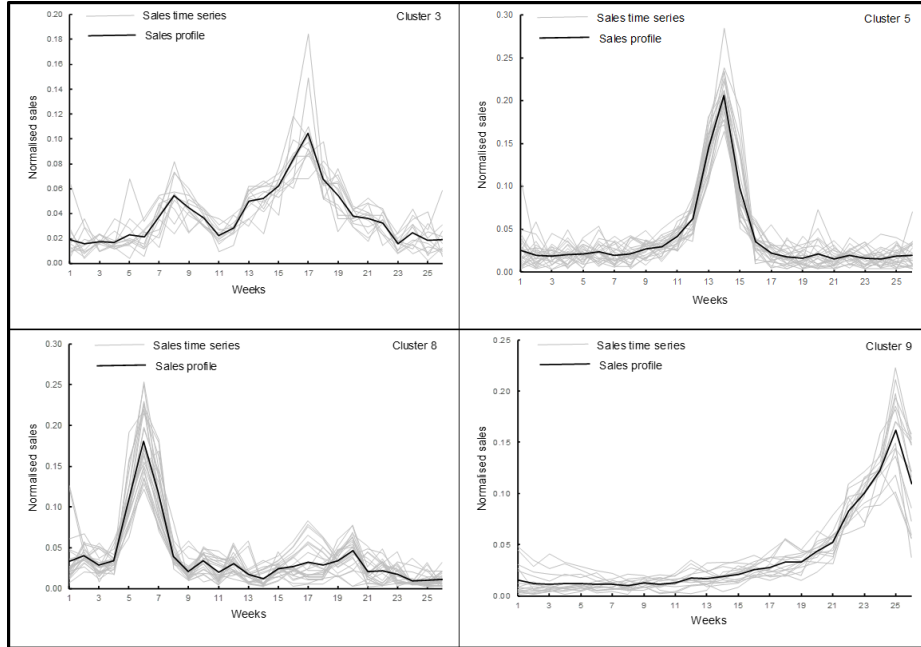
**Sales Profiles Prediction for New Products.** After clustering the time series data of the historical products, labels such as  $SP_1, SP_2, \dots, SP_1, \dots, SP_q$  representing the sales profiles of the formed groups, are assigned to the discovered clusters. A training set for the classification algorithm is then constructed using the descriptive attributes and the clusters labels. Each product is considered as a data point in the training set and it is described by the descriptive attributes as inputs and the label of the cluster to which it belongs is taken as its output.

The RULES-6 algorithm is used to extract if-then rules from the training data set. These rules provide a comprehensive insight into the data and are used to predict the sales profiles of the new products based on their descriptive attributes. RULES-6 has a number of parameters whose values determine the quality of the induced rule sets. In this research, the default settings were used [18].

## 4 Experimental Results

### 4.1 Clustering Results

The incremental  $k$ -means algorithm was applied to the training time series data and thirteen distinct clusters were created. Figure 4 shows four examples of the clustering results that have the maximum distortion errors. As can be seen in the figure, the incremental  $k$ -means clustering procedure is effective in producing accurate groups of products with similar time series sales. The sales behaviour of the products belonging to the different clusters are accurately described by the associated sales profiles (indicated in bold). It can then be concluded that a robust identification of the sales profiles has been achieved.



**Fig. 4.** Examples of the sales profiles created by the incremental  $k$ -means clustering algorithm.

### 4.2 Rule Induction Results

The RULES-6 algorithm was applied on the classification training data created from the descriptive attributes and the discovered sales profiles. Table 1 illustrates the produced set of rules to describe the training data. It is clear from the figure that the number of rules generated is significantly lower than the number of data points in the training set. Also, the number of features describing each rule is drastically reduced. The generated rule set is a compressed description of the training data that could be used to predict the sales profiles of new products.



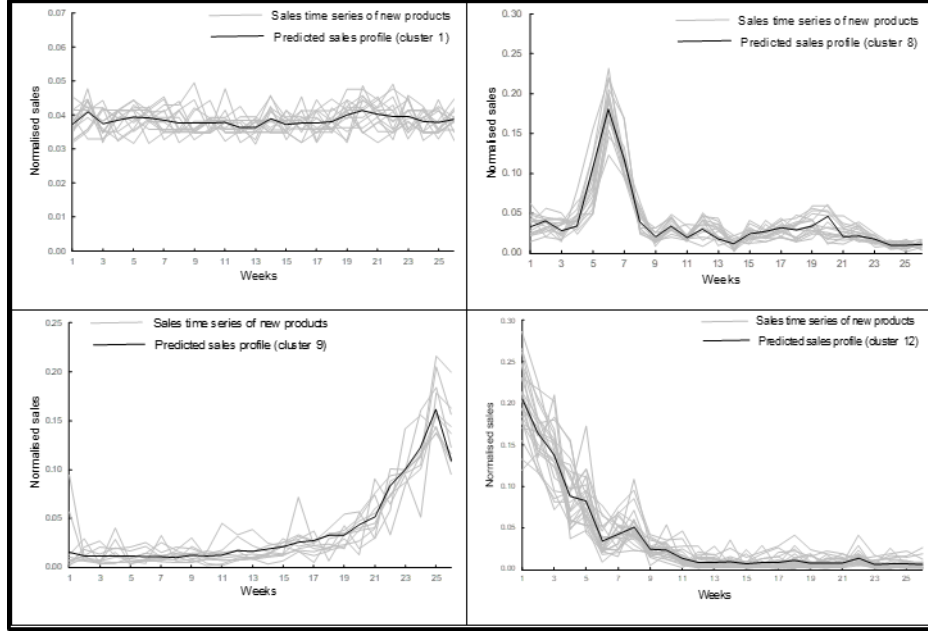
**Table 1.** Rules derived by RULES-6 from the retail data.

R01: IF (Price <= £44) THEN Sales profile = SP1
R02: IF (Sales start date = 4) THEN Sales profile = SP4
R03: IF (Sales start date >= 14) THEN Sales profile = SP2
R04: IF (£101 <= Price <= £147) AND (Life span >= 22) THEN Sales profile = SP3
R05: IF (Price <= £98) AND (Sales start date <= 2) THEN Sales profile = SP7
R06: IF (Price <= £99) AND (Sales start date <= 11) THEN Sales profile = SP11
R07: IF (Life span <= 11) THEN Sales profile = SP6
R08: IF (Life span = 25) THEN Sales profile = SP9
R09: IF (Price <= £92) THEN Sales profile = SP10
R10: IF (Price >= £300) THEN Sales profile = SP12
R11: IF (£200 <= Price <= £249) THEN Sales profile = SP5
R12: IF (Price >= £250) THEN Sales profile = SP13
R13: IF (Sales start date = 1) AND (11 < Life span <= 14) THEN Sales profile = SP8

To evaluate its performance, the RULES-6 algorithm was applied on the validation data to predict the sales profiles of the new products. Figure 5 shows a comparison between the actual sales of the new products and the sales profiles predicted by the RULES-6 algorithm. As can be seen in the figure, RULES-6 assigns the correct profiles to most products. The actual sales associated with profile SP9 of cluster 9 are quite different in some weeks. Also, the actual sales accompanied with profile SP8 of cluster 8 are relatively similar but are slightly different from the predicted sales profile. This could arise from the failure of the descriptive attributes used to build the rule set to explain the sales behaviour of these products.

To further test its performance, the prediction accuracy of the RULES-6 algorithm was compared with that of five different well-known classifiers. The classification methods are the OneR algorithm [30], the Naïve Bayes method [31], the  $k$ -nearest-neighbours classifier [32], the RIPPER rule induction classifier [33], and the C4.5 decision tree [34]. These methods are representative of the different classification techniques and widely used in other comparative studies. Each of the classification methods was first applied on the classification training data and the generated classification models were then tested on the validation data to predict the sales profiles of the new products. The default parameters of the tested methods were used. The forecasting errors of the predicted sales profiles were analysed using three of the most popular measures, namely, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) [35].

The forecasting error values of the predicted sales profiles for the different classification methods are given in Table 2. As shown in the table, RULES-6 is the most accurate algorithm. It achieved the smallest error values for the three measures and the values were lower by 39.2% for the MAE, 21.7% for the MAPE and 43.4% for the RMSE compared to the next best performing classifier. It could therefore be concluded that the proposed RULES-6 algorithm gives the best performance among the six methods tested.



**Fig. 5.** Comparison between the actual sales of new products and the sales profiles predicted by RULES-6.

**Table 2.** Forecasting errors of the predicted sales profiles for the tested classification methods.

	MAE	MAPE (%)	RMSE
OneR	70.7	5.2	84.3
Naïve Bayes	63.2	4.5	78.9
<i>k</i> -nearest-neighbours	51.3	3.7	66.8
RIPPER	16.4	2.9	21.1
C4.5	9.7	2.3	12.9
Proposed RULES-6	<b>5.9</b>	<b>1.8</b>	<b>7.3</b>

## 5 Conclusions and Future Work

The constraints of the retail market make the forecasting of products sales very challenging. This paper has proposed a forecasting approach based on the incremental *k*-means clustering algorithm and the RULES-6 rule induction classifier to estimate the sales profiles of new products. The performance of the proposed forecasting procedure has been tested on real sales data of a large retail Egyptian company specialised in selling and distribution of high tech products. The incremental *k*-means algorithm employed in this study has proved to be effective in discovering interesting groupings of historical products, from which general descriptions can be derived by applying the RULES-6 algorithm. From the clusters and their descriptions, sales profiles for new products could be predicted easily. However, some inaccurate profiles could result from

the inadequacy of the chosen descriptive attributes to discriminate the sales behaviours of all new products. The RULES-6 algorithm has also been tested against five other state-of-the-art classifiers and attained more accurate forecasts. These results suggest that the combined application of the incremental  $k$ -means and RULES-6 techniques is a valid approach for estimating the sales profiles of new products in retail businesses where historical sales data are not available.

Further research could be conducted to apply the proposed forecasting procedure to other short life cycle products and to compare the RULES-6 algorithm with other classifiers such as decision trees, neural networks, or genetic algorithms. Also, considering the various sources of uncertainty that arise in the retail market, it would be interesting to use fuzzy learning techniques which may help to produce more accurate forecasts.

**Acknowledgement.** The author wishes to thank the University of the West of England for providing a good environment, facilities and financial means to complete this paper.

## References

1. Kadam, S. and Apte, D.: A survey on short life cycle time series forecasting. *International Journal of Application or Innovation in Engineering & Management*, 4(5), 445–449 (2015).
2. Armstrong, J.S.: *Principles of Forecasting - A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers, Norwell, MA, USA (2001).
3. Xia, M. and Wong, W. K.: A seasonal discrete grey forecasting model for fashion retailing. *Knowledge-Based Systems*, 57, 119–126 (2014).
4. Beheshti-Kashia, S., Karimic, H. R., Thobenb, K.-D., Lütjenb, M. and Teuckeb, M.: A survey on retail sales forecasting and prediction in fashion markets. *Systems Science & Control Engineering*, 3, 154–161 (2015).
5. Giustiniano, L., Nenni, M.E. and Pirolo, L.: Demand forecasting in the fashion industry: a review. *International Journal of Engineering Business Management*, 5 (37) (2013).
6. Liu, N., Ren, S., Choi, T.-M., Hui, C.-L. and Ng, S.-F.: Sales forecasting for fashion retailing service industry: a review. *Mathematical Problems in Engineering*, ID 738675 (2013).
7. Choi, T. M., Hui, C.-L., Liu, N., Ng, S-F, & Yu, Y.: Fast fashion sales forecasting with limited data and time, *Decision Support Systems*, 59, 84–92 (2014).
8. Fajardo-Toro, C. H., Mula, I. and Poler, R.: Adaptive and hybrid forecasting models - a review. In: A. Ortiz et al. (eds.), *Engineering Digital Transformation*, Springer (2018).
9. Choi, T.-M., Yu, Y. and Au, K.-F.: A hybrid SARIMA wavelet transform method for sales forecasting. *Decision Support Systems*, 51, 130–140 (2011).
10. Thomassey, S.: Sales forecasts in clothing industry - the key success factor of the supply chain management. *International Journal of Production Economics*, 128, 470–483 (2012).
11. Choi, T.-M., Hui, C.-L. and Yu, Y.: *Intelligent Fashion Forecasting Systems: Models and Applications*. Springer-Verlag, Berlin Heidelberg (2014).
12. Thomassey, S. and Fiordaliso, A.: A hybrid sales forecasting system based on clustering and decision trees. *Decision Support Systems*, 42(1), 408–421 (2006).
13. Thomassey, S. and Happiette, M.: A neural clustering and classification system for sales forecasting of new apparel items. *Applied Soft Computing*, 7(4), 1177–1187 (2007).
14. Maaß, D., Spruit, M. and de Waal, P.: Improving short-term demand forecasting for short-lifecycle consumer products with data mining techniques. *Decision Analytics* (2014).

15. Gaku, R.: Demand forecasting procedure for short life-cycle products with an actual food processing enterprise. *International Journal of Computational Intelligence Systems*, 7, 85–92 (2014).
16. Basallo, M., Rodríguez-Sarasty, J. and Benitez-Restrepo, H.: Analogue-based demand forecasting of short life-cycle products: a regression approach and a comprehensive assessment. *International Journal of Production Research*, 55, 1–15 (2016).
17. Pham, D.T., Dimov, S.S. and Nguyen, C.D.: An incremental  $k$ -means algorithm. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 218(7), 783–795 (2004).
18. Pham, D.T. and Afify, A.A.: RULES-6: A simple rule induction algorithm for handling large data sets. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(10), 1119–1137 (2005).
19. Witten, I. H. and Frank, E.: *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers, USA (2000).
20. Han, J. and Kamber, M.: *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco, California, USA (2001).
21. Klösgen, W. and Żytkow, J. M.: *Handbook of Data Mining and Knowledge Discovery*. Oxford University Press, New York, USA (2002).
22. Kaufman, L. and Rousseeuw, P.J.: *Finding Groups in Data: An Introduction to Cluster Analysis*. 2nd edn. John Wiley & Sons, New Jersey, USA (2005).
23. Pham, D.T. and Afify, A.A.: Clustering techniques and their applications in engineering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 221(11), 1445–1459 (2007).
24. Pham, D.T., Dimov, S.S. and Nguyen, C.D.: A two-phase  $k$ -means algorithm for large datasets. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 218(10), 1269–1273 (2004).
25. Pham, D.T., Dimov, S.S. and Nguyen, C.D.: Selection of  $k$  in  $k$ -means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 215(1), 103–119 (2005).
26. Pham, D.T. and Afify, A.A.: Machine-learning techniques and their applications in manufacturing. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 219(5), 395–412 (2005).
27. Pham, D.T. and Afify, A.A.: A new minimum description length based pruning technique for rule induction algorithms. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 222(7), 1339–1352 (2008).
28. Pyle, D.: *Data Preparation for Data Mining*. Morgan Kaufmann Publishers, USA (1999).
29. Vesanto, J. and Alhoniemi, E.: Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3), 586–600 (2000).
30. Holte, R.C.: Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 1993, 11, 63–91.
31. Langley, P., Iba, W. and Thompson, K.: An analysis of Bayesian classifiers. In: *Proceedings of the 10<sup>th</sup> National Conference on Artificial Intelligence*, USA, 1992, pp. 223–228.
32. Aha, D.: Tolerating noisy, irrelevant, and novel attributes in instance-based learning algorithms. *International Journal of Man–Machine Studies*, 1992, 36 (2), 267–287.
33. Cohen, W.W.: Fast effective rule induction. In: *Proceedings of the 12<sup>th</sup> International Conference on Machine Learning*, Lake Tahoe City, California, USA, 1995, pp. 115–123.
34. Quinlan, J.R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann, USA, 1993.
35. Armstrong, J.S. and Collopy, F.: Error measures for generalizing about forecasting methods: empirical comparisons. *International Journal of Forecasting*, 1992, 8, 69–80.