



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *2019 International Conference on Engineering, Science, and Industrial Applications (ICESI), Tokyo, August 22-24, 2019.*

Citation for the original published paper:

Giri, C., Thomassey, S., Balkow, J., Zeng, X. (2019)

Forecasting New Apparel Sales Using Deep Learning and Nonlinear Neural Network Regression

In: *2019 International Conference on Engineering, Science, and Industrial Applications (ICESI)*

2019 International Conference on Engineering, Science, and Industrial Applications (ICESI)

<https://doi.org/10.1109/ICESI.2019.8863024>

N.B. When citing this work, cite the original published paper.

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:hb:diva-22822>

# Forecasting New Apparel Sales Using Deep Learning and Nonlinear Neural Network Regression \*

Chandadevi Giri, Sebastien Thomassey, Jenny Balkow and Xianyi Zeng

**Abstract**— Compared to other retail industries, fashion retail industry faces many challenges to foresee future demand of its products. This is due to ever-changing choices of their consumers, who get influenced by rapidly changing market trends and it leads to the short life cycle of a fashion product. Due to the advent of e-commerce business models, fashion retailers have to put a multitude of virtual product images along with their feature information on their websites in order for their customers to know the fashion products and improve their purchasing experience. It is imperative for fashion retailers to predict future consumer preferences in advance; however, they lack advanced tools to achieve this goal. To overcome this problem, this research work combines the historical information of products with their image features using deep learning and predicts future sales. Apparel images are converted into feature vectors and then are merged with historical sales data. We applied backward propagation neural network model to predict the sales of a new product. It is found that the model performs quite well despite the small size of the dataset. This approach could be promising for forecasting the new arrivals of apparels in the market, and fashion retailers could improve their efficiency and growth.

**Keywords:** Forecasting, Deep Learning, Neural Network, Apparel Industry.

## I. INTRODUCTION

In the era of internet and information services, where retailing is integrated online, consumers have multiple options to buy products of their choice. Consumers choices are influenced by multiple factors such as social media, fashion shows, fashion events, etc. Fashion retailers face various challenges to fulfill rapidly changing consumer demands in a very short period of time[1]. For this reason, it is important for fashion retailers to optimize their decision plans quite ahead of the upcoming replenishment of their stocks by accurately predicting future market trends. Accordingly, new products can be designed and produced in

\*Research supported and funded by European Commission under the framework of Erasmus Mundus Joint Doctorate Program, SMDTex (Sustainable Management and Design for Textiles).

Chandadevi Giri, a doctoral student, working collaboratively with University of Borås, Allegatan 1, 503 32 Borås, Sweden, Gemtex Ecole Nationale Supérieure des Arts et Industries Textiles, 59100 Roubaix, France, Université Lille Nord de France, F-59000 Lille, France and College of Textile and Clothing Engineering, Soochow University, 215168-Suzhou, China ([chanda.giri2@gmail.com](mailto:chanda.giri2@gmail.com)).

Sebastien Thomassey, an associate Professor, working with Ecole Nationale Supérieure des Arts et Industries Textiles, 59100 Roubaix, France 5 and Université Lille Nord de France, F-59000 Lille, ([sebastien.thomassey@ensait.fr](mailto:sebastien.thomassey@ensait.fr)).

Jenny Balkow, a Lecturer, working with University of Borås, Allégatan 1, 503 32 Borås ([jenny.balkow@hb.se](mailto:jenny.balkow@hb.se)).

Xianyi Zeng, a Professor, working with Ecole Nationale Supérieure des Arts et Industries Textiles, 59100 Roubaix, France and Université Lille Nord de France, F-59000 Lille, France ([xianyi.zeng@ensait.fr](mailto:xianyi.zeng@ensait.fr)).

a way that the future customer demands will be fulfilled and the efficiency of their business will be improved.

The outlook of apparel products varies in terms of style, pattern, design, etc. and it changes quickly with respect to time. Before products are delivered to the retailers, they pass through various supply chain layers starting from product design, manufacturing and to its distribution. Forecasting the optimal quantity of product demands is always interesting for the retailers, which is usually based on the current and historical sales pattern. Inability to forecast demand quantities of products correctly may cause a shortage or overstock of products, which may further lead to the excess inventory cost and loss of profit and customer loyalty.

Nowadays, the fashion retail industry generates an enormous amount of data, which can offer significant insights not only about their product features but also about their customers. Data generated in fashion retailers' databases exhibit complex format; broadly in image and relational sales transaction data format. Image data possess certain characteristics of the product in terms of color, style, design, etc., and sales data which defines the demand of the product in the market. Given the season-wise variations in the fashion products and consumer demands, historical product image data alone can't predict the features of future fashion products; neither can it forecast accurate information about their future sale. Considering the importance of a precise forecast of future product demands, we combine current sales data of the products with their images in order to predict the number of future sales.

Rest of the paper is organized as follows: Section II defines the aim of the study. A brief review of the related work is presented in section III. Experimental details and discussion about the model and results are presented in section IV and V respectively. Finally, section VI concludes the research study.

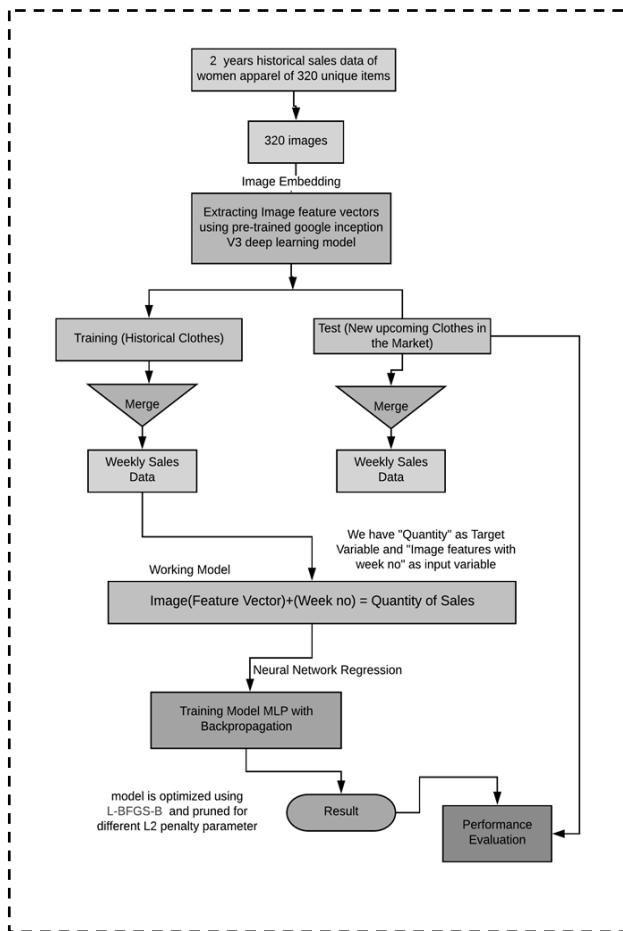
## II. AIM OF THE RESEARCH

The objective of this work is to predict the quantity of new product using historical sales data of female apparel clothing and their corresponding images. To achieve this aim, we used a deep neural network V3 inception model developed by Szegedy [2] for extracting the feature vectors of the images. Each image defines a unique item and 320 unique women apparel images are considered for this study. This will be followed by nonlinear Neural Network (MLP-Multilayer Perceptron) regression used in [3]. Schematic representation of the work can be seen in figure 1.

### III. LITERATURE WORK

A number of forecasting methods have been developed and employed in the fashion retail industry over recent years. Statistical approaches such as regression used by Papalexopoulos[4]; ARIMA and Exponential smoothing used in Healy [5]; Box & Jenkins methods [6] are popular for forecasting demand. However, these methods face the limitation of transforming data from qualitative to quantitative because sales pattern varies in the fashion retail industry [7]. Laney and Goes [8] suggest that with the introduction of big data, data mining has gained a lot of popularity in recent times. Data mining and artificial intelligence overcome the problems of the classical approaches in forecasting problems[9][10].

Figure 1. Schematic representation of Research Framework



Recently, the growing popularity of deep learning methods and their advantages in solving data mining problems over other methods has attracted the attention of the scientific community[11]. Deep learning has found a plethora of applications in many areas and significant research work has been done on it in the domain of medical field [12], transportation problems[13], electricity [14], etc. There are various studies such as [10], [15], in which the prediction was performed by considering various product categories. In

another interesting study by Thomassey [16], clustering and NN are combined to predict the fashion sales. CNN has gained popularity in image recognition since 2014 [2]. Taking into account all previous work in the domain of sales forecasting in the fashion retailing, this work proposes a novel approach to combine the deep learning and nonlinear regression using Neural network.

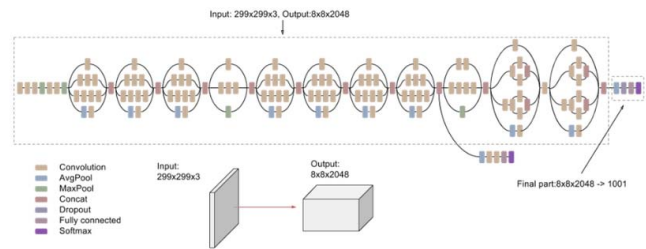
### IV. METHODOLOGY

This section elucidates the methodologies used to conduct this research study. Section IV.A explains the feature extraction of images using the inception v3 model and section IV.B describes the Multilayer perceptron used in this research work.

#### A. Feature Extraction:

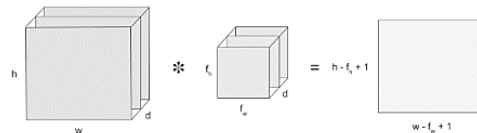
Image features are extracted using deep learning architecture of inception-v3 model [2]. Inception-v3 model has outperformed the previous image recognition and has accomplished accuracy for recognizing objects are trained on more than 1000 objects including clothes. The architecture of the inception-v3 model is shown in the below figure 2.

Figure 2. The architecture of Inception Model



1) *Convolution layer*- Image features are learned at the pixel level and preserved by convolution layers. It is the first level of image input to extract the features. It takes two inputs Image matrix  $(h \times w \times d)$  and filters  $(f_h \times f_w \times d)$  which maps the feature and give output as  $(h - f_h + 1) \times (w - f_w + 1) \times 1$  as depicted in Figure 3.

Figure 3. Feature mapping of Convolution Layer



2) *Pooling layers*: if the size of the input image is huge, then pooling layers helps to reduce the dimension by discarding the parameters. Mapping could be done using spatial pooling and it could be done in three ways namely “Max pooling”, “Average Pooling” and “Sum

Pooling”. “Average pooling” computes the average values of the feature map, “Max Pooling” computes the largest values from the feature map and “Sum Pooling” gives the aggregate value of all feature value.

3) *Concat layer*- Pooling layer are followed by concat layers, which concatenate the output from all layers and link it to the subsequent layer.

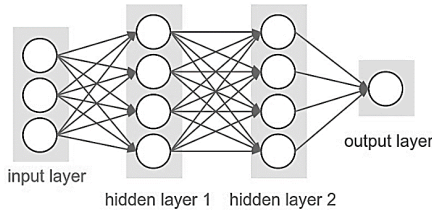
4) *Dropout layer* - The purpose of Dropout layer is to deal with the overfitting problems, so at this stage, nodes are dropped out of the model with a probability of “1-p” where “p” is dropping out neurons.

5) *Fully connected layer and softmax layer*- Lastly, features are passed through the fully connected layer and top-level feature interpretation is done here and features will be transformed into feature vectors, and these are used to create a model. For the classification task, activation function such as sigmoid or softmax is further used in the model.

### B. Multilayer perceptron (MLP)

This belongs to the group of “feedforward artificial neural network”. It consists of three layers input layer, weights, and activation function. The model uses weighted input using Non-linear activation function for giving outputs. MLP uses supervised learning for training using the back-propagation technique. Multiple layers and non-linear activation function differentiate MLP from a linear perceptron (LP). They have the ability to solve complicated regression and classification problems. Working of MLP is shown in the Figure 4.

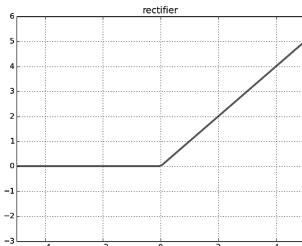
Figure 4. MLP Workflow



#### 1) Activation Function- ReLU (Rectified Linear Unit)

ReLU is popular non-linear activation for the deep learning task. If the input value is less than 0, the output will be 0, while if the value of the input is higher than 0, then the output will be equivalent to the input value. Graphical performance of ReLU is shown in Figure 5.

Figure 5. Graphical performance of ReLU



$$f(x) = \max(0, x) \quad (1)$$

#### 2) Weight Optimizer

L-BFGS-B [17], [18] is an optimizer that belongs to quasi-Newton algorithms and it is used for estimating parameters in Artificial intelligence methods. It is the extended version of L-BFGS and applied on the constraints of the system like  $m_i \leq x_i \leq n_i$ , where ‘ $m_i$ ’ and ‘ $n_i$ ’ are upper and lower bounds. The algorithm simply works by finding fixed and free variables at each step using the gradient descent method and then applying L-BFGS on free variables to reduce errors and this process keeps on repeating unless it achieves high accuracy.

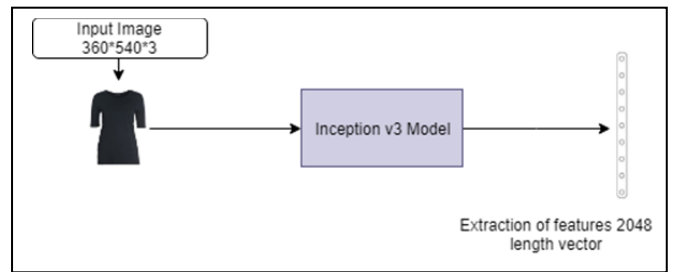
## V. EXPERIMENTAL DESIGN

This section discusses the sales prediction model formation using the presented methodologies discussed in the above section IV. Subsequent sections V.A and V.B outlines the data preparation and modelling.

### A. Data Preparation

This step was carried out on apparel images and its sales information data. 320 women apparel images are used for this study of a fashion apparel brand. Image features were extracted using deep learning CNN google inception model v3 developed in [2]. The dimension of each input images is 360 X 540 X 3. This feature of each images are retrieved using inception model v3 model shown in Figure 2, which gives 2048 feature vectors of the images represented in Matrix form as an output as shown in Figure 6. Each image is a product and has historical sales information. Quantity sold is aggregated weekly, so that weekly demand of the products could be modelled. The extracted image feature vectors are merged with its sales data. Before merging, data was split into training and test considering 90% of the items for training and 10 % of the items for testing, i.e. 289 images for training and 31 images for testing. After merging feature vectors of product images with sales data, the final number of observations was 7403 and 748 in training and test data respectively.

Figure 6. Image Embedding using Inception V3 Model (Feature Extraction)



## B. Apparel Sales Forecast Model

The final dataset has 2049 features as independent variables and ‘quantity of sales’ as a target variable. We used “Neural Network Multi-layer Perceptron” for this work as it has a proven ability to learn non-linear models [3]. It trains from function  $f(\cdot): I_n(\text{Input}) \rightarrow O_t(\text{Output})$  by training on input data, where “n” is the number of predictors and “t” is the number of target values for output. In this study, we considered features for Input layer  $\text{Input} = \{I, W\}$ , where, I is a set of Images  $I = \{i_1 + i_2 + \dots + i_n\}$  and each image is represented as a set of features extracted from inception model  $i = \{f_1 + f_2 + \dots + f_{2048}\}$ , and W is the set of weeks in which apparel products were sold and it could be presented as  $W = \{w_1 + w_2 + \dots + w_n\}$  and  $\text{Output} = \{Q\}$ , where Q is the quantity sold for a product with image i for a given week w.. The number of hidden layers is 100 followed by activation function “ReLU,” which is a rectified linear activation function and popular for deep learning [19], [20]. Weights are optimized using “L-BFGS” [18], [21], [22] and model is pruned for the different “ $\alpha$ ” values. Model parameters can be seen in Table 1. Model performance is evaluated based on MSE (Mean squared error), RMSE (Room mean Squared error), MAE (Mean Absolute Error) [23], and R2 (co-efficient of correlation) for different tuning parameters on test and trained data. Results are illustrated in the Table 2 and Table 3.

TABLE I. MODEL PARAMETERS

Model	Hidden Layers	Activation function	Tuning Parameter	Weight Optimization
Neural Network, MLP	100	ReLU	$\alpha = 0.00001, 0.01, 0.1$	L-BFGS-B

## VI. RESULTS AND DISCUSSION

We applied the “Neural network MLP model” on trained data for three different regularized “ $\alpha$ ” values to get the optimum performance of the model and also to illustrate the overfitting problem. We can observe that the values of the evaluation parameters for training and test dataset do not differ significantly. Hence, we can say that the model does not exhibit the overfitting problem. The value of “R2” gives information of the variance in the dependent attributes with independent attributes [24]. In our analysis, the model prediction on training data is positive (approx. 40 % of variation) whereas for the test data it shows the inverse relationship, which is negative. It is possible that R2 values can be negative [25], which is due to the nonlinear relationship between the regression variables, and therefore, we cannot consider R2 values for assessing the goodness of fit [24]. MSE and RMSE are two popular indicators in statistics to access the performance of regression models. In Table 3, the prediction performance on the test data for  $\alpha = 0.0001$  is 19.4, which is not too far from value of train data shown in Table 2. Therefore, RMSE value is acceptable data. Overall, the best prediction result is for the alpha value = 0.0001 for both trained and test dataset given the model performance on trained data.

This could be further improved by increasing the number of trained images. Mean Absolute Error (MAE) explains as to how close the predictions are to the actual data points [23]. MAE is commonly used for accessing the performance of forecasting models. It can be interpreted easily; closer to zero is the better. So, we can see that the MAE for test data is 13.197 and it could be improved by increasing the size of training data.

Figure 7. Actual vs Predicted Weekly Sales Quantity

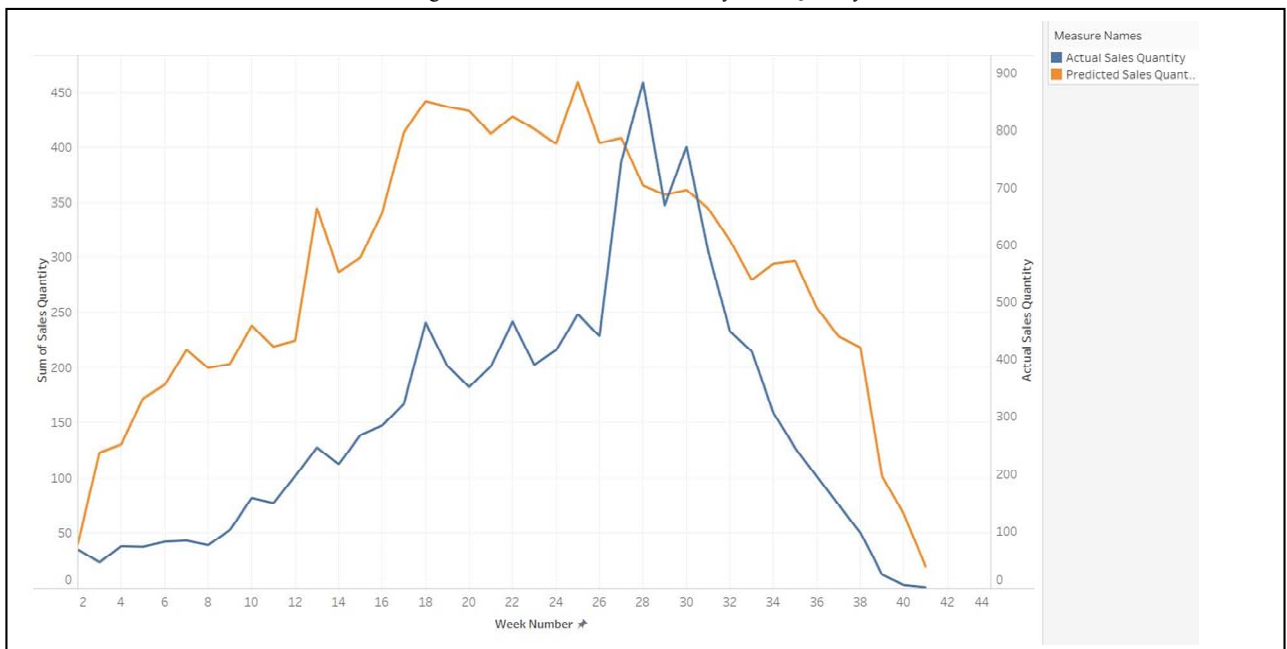




TABLE II. MODEL PREDICTION RESULTS ON THE TRAINING DATASET

Method	MSE	RMSE	MAE	R2	Tuning Parameter
Neural Network	238.857	15.455	9.556	0.441	$\alpha=0.0001$
Neural Network	245.532	15.669	9.742	0.425	$\alpha=0.1$
Neural Network	245.083	15.655	9.704	0.426	$\alpha=1$

TABLE III. MODEL PREDICTION RESULTS ON TEST DATASET

Method	MSE	RMSE	MAE	R2	Tuning Parameter
Neural Network	376.371	19.400	13.197	-0.246	$\alpha=0.0001$
Neural Network	394.798	19.870	13.605	-0.307	$\alpha=0.1$
Neural Network	414.740	20.365	14.462	-0.373	$\alpha=1$

Here, we can observe that few products sales are predicted quite accurately for some weeks and for some products, prediction results are less accurate since the predicted values differ from the actual values. This model could be further improved by enhancing the dataset with more number of images with variations in terms of product category, style, design, color, etc.

## VII. CONCLUSION

The novel approach to forecast the sales of fashion products based on deep learning and non-linear NN regression is presented in this paper. Results of this study seem promising for forecasting the future sales quantity of the products. The major limitation of this study lies in the fact that smaller dataset is used. In future work, we aim to overcome this limitation by selecting large image dataset in order to improve model performance. In the era of the big data and data mining, this approach could benefit the designer and fashion retailers. By carefully studying the sales of the current product and extracting the abstract features hidden in the particular product images, fashion designers and retailers can predict the nature and demand of the products in future and thereby can improve their business performance.

Figure 8. Predicted Vs Actual Quantity of sales for test items

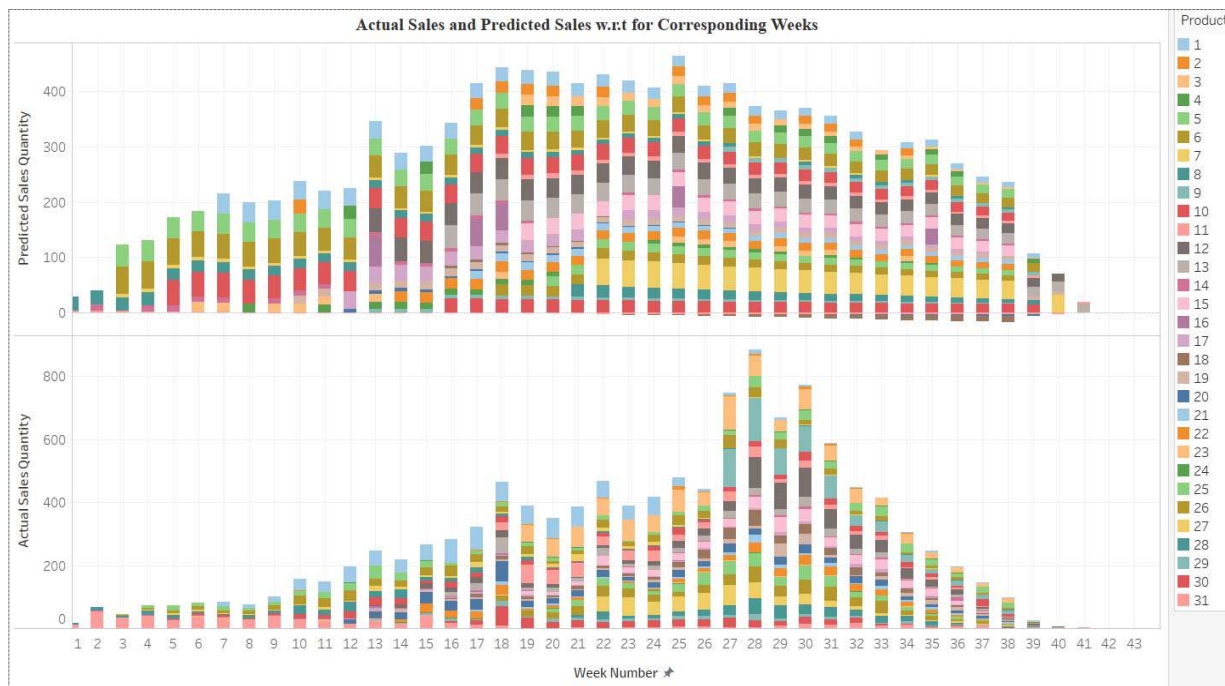


Figure 7 depicts the actual and predicted sales quantity on the test data. We can observe that the product sale in a given week is over predicted. Figure 4 illustrates the comparison for the actual and predicted the quantity of the sales for the 31 product images in the test data. In figure 8, each color in the bar represents test images which are not considered for the model training.

## ACKNOWLEDGMENT

This work is carried out under the framework of SMDTex-Sustainable Management and Design for textiles. We are thankful to the 'Evo Pricing' company and their team for helping us with the data.

## REFERENCES

- [1] C. Giri, S. Thomassey, and X. Zeng, "Customer Analytics in Fashion Retail Industry," in *Functional Textiles and Clothing*, Singapore: Springer Singapore, 2019, pp. 349–361.
- [2] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.
- [3] M. A. Nielsen, "Neural Networks and Deep Learning." Determination Press, 2015.
- [4] A. D. Papalexopoulos and T. C. Hesterberg, "A regression-based approach to short-term system load forecasting," in *Conference Papers Power Industry Computer Application Conference*, pp. 414–423.
- [5] M. J. R. Healy and R. G. Brown, "Smoothing, Forecasting and Prediction of Discrete Time Series.," *J. R. Stat. Soc. Ser. A*, vol. 127, no. 2, p. 292, 1964.
- [6] G. E. P. Box and G. M. Jenkins, *Time series analysis : forecasting and control*. Prentice Hall, 1976.
- [7] P. C. L. Hui and T.-M. Choi, "Using artificial neural networks to improve decision making in apparel supply chain systems," in *Information Systems for the Fashion and Apparel Industry*, Elsevier, 2016, pp. 97–107.
- [8] D. Laney, "3D Data Management: Controlling Data Volume, Velocity, and Variety."
- [9] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.
- [10] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*. Pearson Addison Wesley, 2005.
- [11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning.," *Nature*, vol. 521, no. 7553, pp. 436–44, May 2015.
- [12] S. Jiang, K.-S. Chin, L. Wang, G. Qu, and K. L. Tsui, "Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department," *Expert Syst. Appl.*, vol. 82, pp. 216–230, Oct. 2017.
- [13] J. ScienceDirect (Online service), H. Zheng, H. Yang, and X. (Michael) Chen, *Transportation research. Part C, Emerging technologies.*, vol. 85, no. 0. Pergamon Press, 2017.
- [14] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. J. Amaratunga, "Empirical Mode Decomposition based ensemble deep learning for load demand time series forecasting," *Appl. Soft Comput.*, vol. 54, pp. 246–255, May 2017.
- [15] S. Ren, T.-M. Choi, and N. Liu, "Fashion Sales Forecasting With a Panel Data-Based Particle-Filter Model," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 45, no. 3, pp. 411–421, Mar. 2015.
- [16] S. Thomassey and M. Happiette, "A neural clustering and classification system for sales forecasting of new apparel items," *Appl. Soft Comput.*, vol. 7, no. 4, pp. 1177–1187, Aug. 2007.
- [17] R. H. Byrd, L. Peihuang, and J. Nocedal, "A limited-memory algorithm for bound-constrained optimization," Argonne, IL, Mar. 1996.
- [18] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, Dec. 1997.
- [19] U. Michelucci, "Convolutional and Recurrent Neural Networks," in *Applied Deep Learning*, Berkeley, CA: Apress, 2018, pp. 323–364.
- [20] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for Activation Functions," Oct. 2017.
- [21] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A Limited Memory Algorithm for Bound Constrained Optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, Sep. 1995.
- [22] J. Morales, J. N.-A. T. M. Softw., and undefined 2011, "Remark on&quot; algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization&quot;," *researchgate.net*.
- [23] C. Willmott, K. M.-C. research, and undefined 2005, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Clim. Res.*
- [24] S. Glantz and B. Slinker, "Primer of applied regression and analysis of variance," 1990.
- [25] A. Cameron, F. W.-J. of econometrics, and undefined 1997, "An R-squared measure of goodness of fit for some common nonlinear regression models," *Elsevier*.