Prediction of yarn sales price using data mining techniques – a case of yarn manufacturing industry

DOI: 10.35530/IT.075.02.20234

MUHAMMAD ALI SAMAN HINA SHERAZ HUSSAIN SIDDIQUE RUKSHAN TAUFIQ LODHI

ABSTRACT – REZUMAT

Prediction of yarn sales price using data mining techniques - a case of yarn manufacturing industry

Data-driven knowledge is required for businesses to make better decisions that result in profit maximisation. In this study, it has been attempted to develop a model to predict yarn sales prices against cotton prices and other parameters. For this purpose, four different data mining techniques namely ARIMA (Autoregressive Integrated Moving Average), Multivariate regression, K-Nearest Neighbor (KNN) and Neural Networks (NN), were considered. The entire analysis was performed on thirty months of data that was collected from the ERP system of a yarn manufacturing industry. The unique aspect of this study is that before separately deploying data mining techniques, significant parameters that impact yarn sales prices were identified through Adjusted R-squared values. Seasonal and trend patterns were checked on yarn sales data, and seasonal adjustments were obtained through data mining algorithms. The performance of all four models was evaluated using Mean Absolute Error and Root Mean Square Error. The analysis shows that the KNN model, in the stated settings, is the most accurate as evident from MAE and RMSE values of 222.85 and 285.082, respectively. This study's unique combination of features and machine learning algorithms is envisaged to be valuable for decision-makers in the textile yarn manufacturing industry.

Keywords: yarn price, cotton price, prediction, K Nearest Neighbor, multivariate regression, Neural Network

Predicția prețului de vânzare a firelor folosind tehnici de extragere a datelor – un caz al industriei de fabricare a firelor

Cunoștințele bazate pe date sunt necesare pentru ca întreprinderile să ia decizii care au ca rezultat maximizarea profitului. În acest studiu, s-a încercat să se dezvolte un model care să preconizeze prețurile de vânzare a firelor în raport cu prețurile bumbacului și alți parametri. În acest scop, au fost luate în considerare patru tehnici diferite de extragere a datelor și anume ARIMA (Autoregressive Integrated Moving Average), regresia multivariată, algoritmul K-Nearest Neighbor (KNN) și rețelele neuronale (NN). Întreaga analiză a fost efectuată pe date colectate pe parcursul a treizeci de luni din sistemul ERP al unei industrii de fabricare a firelor. Aspectul unic al acestui studiu este că, înainte de implementarea separată a tehnicilor de extragere a datelor, parametrii semnificativi care influențează prețurile de vânzare a firelor au fost identificați prin valorile R-pătrat ajustate. Modelele sezoniere și tendințele au fost verificate pe datele vânzărilor de fire, iar ajustările sezoniere au fost obținute prin algoritmi de extragere a datelor. Performanța tuturor celor patru modele a fost evaluată folosind eroarea medie absolută și eroarea medie pătratică. Analiza arată că modelul KNN, în setările menționate, este cel mai precis, așa cum este evident din valorile MAE și RMSE de 222,85 și, respectiv, 285,082. Combinația unică de caracteristici și algoritmi de învățare automată a acestui studiu este considerată a fi valoroasă pentru factorii de decizie din industria producției de fire textile.

Cuvinte-cheie: prețul firelor, prețul bumbacului, predicție, algoritmul K Nearest Neighbor, regresie multivariată, rețea neuronală

INTRODUCTION

As supply chains are becoming more and more integrated, the criticality of this task of optimization at every stage of the realization of a product or service is increasing [1, 2]. This is because the decisions taken for one stage can have a direct and significant impact on all upstream and downstream operations, ultimately affecting the overall profit margins [3–5]. The present study is related to the textile yarn manufacturing sector in general and cotton yarn manufacturing in particular. In this sector, it is an established practice to keep a large inventory of raw material, i.e., cotton fibre, to cater for yarn production of up to several months [6]. It is imperative to mention here that several research studies in the past have dealt with the prediction of yarn properties, for instance, yarn strength, using artificial neural networks [7, 8].

However, very few studies regarding price prediction for cotton and cotton yarn are available. Thus, the focus of this research was to analyse the selling price of yarn concerning the variation that occurs in cotton price, labour cost, factory overheads, plant maintenance and electricity consumption in the manufacturing of yarn, etc. In addition, warehousing and delivery costs were also taken into account. Technical analysis of the aforementioned parameters was carried out using different methods to design a "predictive framework". Due to the limited literature available for cotton price prediction, the methods used for price prediction of other materials were also studied. For instance, price prediction for crude oil [9, 10] was reported by researchers who proposed a predictive model using 'Long Short Term Memory (LSTM)' in the deep learning domain to increase the accuracy of a large training set. Several forecasting-related studies report on the prediction of stock prices. In one such study, a 'Hybrid deep learning model' using 'LSTM' and 'Grated Recurrent Unit' for the prediction of stock prices was considered [11]. In this work, the investigators used traditional metrics including 'Mean Square Error - MSE', 'Mean Absolute Error - MAE' and 'Mean Absolute Percentage Error - MAPE' for evaluation of their proposed hybrid model. A known machine learning model, 'Support Vector Regression' was found effective in the prediction of stock prices if the model setting was updated repeatedly [12]. In another related study, a 'rough set' and 'Wavelet Neural Network' (WNN) were employed for the prediction of trends in stock prices [13]. In this study, the accuracy of the proposed integrated model was compared with other known methods such as 'Support Vector Machines-SVM' and 'Wavelet Neural Network - WNN when applied separately [13]'.

Several other forecasting models such as 'naïve methods' for three different settings including 'moving average', 'linear regression' and 'multiple regression', etc., were evaluated for an institutional food service facility. Among these, 'multiple regression' was found to be the best-suited model [14]. In another study, different forecasting methods were evaluated for hotel revenue management in which the most recommended methods were exponential smoothing, pick-up, and moving average models [15]. For the price prediction of gold, the application of 'The Autoregressive Integrated Moving Average Model' (ARIMA) has been reported in one study [16].

In one study about the price prediction of two different types of cotton yarn, researchers discussed the fluctuation in prices and the factors responsible for causing such fluctuations [17].

More specifically, 'Seasonality' and ARIMA Models were used to predict attribute prices and thus a composite model was formed by integrating Seasonality and ARIMA Models with the 'KNN Model' to forecast cotton yarn prices with improved accuracy. ARIMA model and time series analysis were utilized in one study about the prediction of Jute production in Bangladesh [18].

In another study, a 'composite' model for improved accuracy in the prediction of cotton prices has been reported [19]. In this study, monthly cotton price in two different periods namely the 'policy shock' period and 'normal market conditions' was predicted. Cotton price prediction in the policy shock period was done by 'vector auto-regressions' and in normal market conditions prediction was done using 'time series forecast'. In another study, researchers evaluated the 'ARIMA model', 'generalized autoregressive conditional heteroscedastic (GARCH) model' and 'exponential GARCH (EGARCH) model' for forecasting prices of edible oils (domestic and international) and cotton (international) [20]. EGARCH model outperformed the ARIMA and GARCH models in forecasting the international cotton price. This improved accuracy in forecasting by the EGARCH model was attributed primarily to its ability to capture asymmetric volatility patterns.

In light of the review of available literature presented above, the goal of this research was to analyse 'yarn sales price' while considering the variation in different components of the varn manufacturing costs. To the best of our knowledge, in the available literature, yarn sales price has not been incorporated in various forecasting/prediction methods as the primary target variable. The significant factors responsible for volatility in various components of manufacturing costs were taken into account [21, 22]. Furthermore, the fact that short-term forecasting is known to be more accurate for commodities, was also considered [23]. Therefore, it is envisaged that the present study will provide a useful insight into effective methods for the analysis of yarn manufacturing business in particular. However, the findings can potentially be applied to other manufacturing and services concerns for which the datasets are similar to the one used in this study.

METHODOLOGY

Based on the problem statement, certain steps were followed to build a prediction model for yarn sales price. In the first step, an evaluation of all the possible features that can create an impact on the manufacturing and selling of yarn was required. Secondly, features having the most pronounced effect on the performance of prediction models were identified. Lastly, effective algorithms for the dataset were investigated. The algorithms were decided based on seasonal and trend patterns in yarn sales prices, based on the influence of different features on the yarn prices and the basis of the relationship between yarn sales prices and the selected features.

Data collection and pre-processing

For the present study, data in the form of time series from July 2016 to December 2018 of a textile spinning mill located in Karachi, Pakistan, was collected from the ERP system of the mill. The data about two types of yarn ('carded' and 'combed') being manufactured by the said textile spinning mill was considered to account for the total cost of production and total sales for the local market.

The purpose of pre-processing operations is to transform and clean raw data to bring it into a meaningful and understandable format [24]. Incomplete data can potentially yield incorrect and misleading results. In this context, the following pre-processing operations were performed in the present study. From the obtained data, the month-wise and lot-wise data was extracted in MS Excel sheets. Some of the variables/parameters of interest were already available in the data while the remaining features that were

		Table 1
COST COMPONENTS AND) FEATURES CC	NSIDERED
Data component	Abbreviation	Unit
Yarn Sales Price	YSP	Rs/lb
USA Cotton Price	CP1	Rs
Pakistani Cotton Price	CP2	Rs
Yarn Production	YP	-
Cotton Consumption	CC	-
Raw Material Cost	RMC	Rs/lb
Electricity Cost	ELC	Rs
Salaries Cost	SC	Rs
Other Expenses Cost	EXC	Rs
Net Total Cost (per pound)	NTC	Rs/lb
Total Yarn Stock	TYS	lb

required for the present study were calculated using the equations as provided in the supplementary data to this study. After these calculations, the data about local sales was separated followed by transformation into numeric data so that it could be subsequently modelled and trained by regression techniques. All

the relevant cost components and features are tabulated in table 1 where Rs stands for Pakistani Rupee.

Analysis performed on yarn sale price

Graphical representation of data is often considered a preliminary step in data analysis. This approach is known to be helpful in the visualization of trends, seasonality, and cycles in the data as well as helps higher management to understand the inclination or declination of yarn sales prices in a better way. These features can then be incorporated into the forecasting methods described in the later section. Time series graphics are often regarded as the foremost method for this purpose [25]. In the present study, seasonal plot, time-series decomposition, and seasonal adjustment methods were employed to analyse the YSP.

- Seasonal plots: The available data was illustrated in the form of different formats of time series graphics. Firstly, a 'time plot' was constructed for the July 2016 to December 2018 period. The resulting plot is shown in figure 1. Secondly, a seasonal plot allows the underlying seasonal pattern to be seen more clearly and is especially useful in identifying seasons (for instance, 'years') in which the pattern changes [26]. In the present study, the seasonal plot of the data was constructed by setting the usual calendar year (Jan-Dec) as the 'season'. The resulting seasonal plot is shown in figure 2.
- Time series decomposition: It is a procedure where a single time series is broken down into its components, i.e., trend, seasonality, and error. In the present study, the available time series data was decomposed using two methods [27]; 'Seasonal and Trend Decomposition using Loess-STL' and 'Classical Decomposition' (both additive



Fig. 1. Time plot of YSP



industria textilă



Fig. 3. Time-series decomposition using: a – STL; b – Classical decomposition (type; additive); c – Classical decomposition (type; multiplicative)



and multiplicative). The resulting graphs are shown in figure 3.

• Seasonal adjustment: It helps model the data without "seasonal effects" [28]. Seasonally adjusted data is obtained by "omitting" the seasonal component from the original data. In the present study, forecasting of YSP via the ARIMA Model was carried out using the original data and the data from which the seasonal component was removed as follows. Seasonal adjustment was done using all the decomposition methods as described in the previous section. The time series re-constructed after seasonal adjustment with Seasonal and Trend Decomposition, Additive Decomposition and with Multiplicative Decomposition is shown in figure 4.

After analysing YSP, the selected features affecting YSP were identified. The selection was based on the 'Adjusted R2' values calculated for the individual parameters and their combinations thereof. Table 2

S

provides the most significant (combination of) parameters and the associated Adjusted R2 value for YSP. Firstly, original data for the months of Oct-Dec 2018 were removed from the 29 months of data (i.e., July 2016 - December 2018) as discussed in the first and second sections. YSP original values are tabulated in table 3. These values were used for testing purposes to find out better accuracy in the predictions and for constructing the best-fit model. To analyse the model's scalability, the dataset is split for 26 months into training and testing datasets listed in table 4.

	Table 3	
ORIGINAL VALUES OF YSP		
Month	YSP Original Value	
Oct-18	881.965	
Nov-18	1154.398	
Dec-18	954.81	

						Table 4
		Table 2				10,510
ELECTED PARAMETERS FOR YARN SALES PRICE			DIVISION OF DATASET (DATA FOR 29 MONTHS)			
OF LOCAL MARKET			Training set (%)	Testing set (%)	Division	
Target Parameters/Independent			60	40	A	
variable	variables	Adjusted R ²		70	30	В
YSP	CP1, CP2, YP, TYS	13.30%		75	25	С

			Table 5
YSP PREDICTIONS WITH ARIMA MODEL			
Month	Predicted value for STL	Predicted value for Additive Decomposition	Predicted value for Multiplicative Decomposition
Oct-18	1611.489	1564.422	1675.729
Nov-18	1594.881	1478.75	1675.729
Dec-18	1580.9	1478.75	1675.729

Implementation of algorithms for prediction of YSP

The first model that was considered to forecast YSP is the ARIMA Model. ARIMA model is a statistical analysis that is used for time series forecasting [29]. This model was built for both STL and classical decomposition (Additive/Multiplicative) methods. After the application of the ARIMA model, it was found that in comparison to the other methods, the YSP predicted using the additive decomposition was closer to the actual values for Oct, Nov, and Dec 2018. Values calculated for both cases (with and without seasonal component) were found to be the same for the dataset used in this research, as shown in table 5.

The second model that was applied is 'Multivarite Regression'. The predicted values of YSP obtained

			Table 6
YSP PREDICTIONS WITH MULTIVARIATE REGRESSION MODEL			
Month	Predicted value (Division A)	Predicted value (Division B)	Predicted value (Division C)
Oct-18	1536.8049	1333.0258	1509.7648
Nov-18	695.4393	490.4838	936.8201
Dec-18	683.3659	497.5379	893.1763

			Table 7
YSP PREDICTIONS WITH KNN MODEL			
Month	Predicted value (Division A)	Predicted value (Division B)	Predicted value (Division C)
Oct-18	934.502	1127.998	1029.166
Nov-18	908.404	890.2999	1029.166
Dec-18	908.404	890.2999	1029.166

YSP PREDICTIONS WITH NN MODEL			
Month	Predicted value (Division A)	Predicted value (Division B)	Predicted value (Division C)
Oct-18	865.0542	882.95	882.85
Nov-18	886.3313	881.22	881.29
Dec-18	889.4591	880.9	881

for different data set divisions are tabulated in table 6. The third model KNN was applied after data cleaning and partitioning as tabulated in table 4 and Euclidean distances were calculated for the predictors listed in table 2. The pair with the least distance is detected as the nearest neighbour and those pairs predicted the yarn sales price. After analysis, the results depicted that K=5 resulted in minimum RMSE and maximum model accuracy.

The predicted values of YSP obtained for different data sets are tabulated in table 7.

Lastly, the 'Neural Networks' model was employed. Different layers of the neural network models were proposed based on different configurations and partitions (table 4). After analysis, it was concluded that the least SSE between training and testing results was found for Model 3 (NN3) because it has two hidden layers with a total of 5 neurons which gives less error in NN3. YSP-relevant prediction results are provided in table 8.

RESULTS AND DISCUSSION

YSP predictions obtained from four models, as described in the preceding section, were analysed to identify the best-performing forecasting model for the dataset used in this research. This analysis was based on checking the error between the predicted and the actual values. For this purpose, the traditional metrics were considered, including 'Mean Absolute Error' (MAE) and 'Root Mean Square Error' (RMSE). The MAE and RMSE values for all models are tabulated in table 9 and table 10.

From these results, it is evident that for Division C of the dataset, the MAE and RMSE values are lower as compared to other Divisions and Methods.

Considering Division C, the MAE and RMSE values for the KNN Model were found to be lowest at 222.85 and 285.082, respectively. Thus, the optimal division of the dataset for all the models was KNN with Division C.

The previous studies used a combination of seasonality, ARIMA and KNN models to predict yarn prices [17]. However, this paper presents a new contribution to the literature. Firstly, the authors extended the features by including all possible data related to yarn sales prices. They evaluated each feature's impact on yarn prices using adjusted R2. Additionally, the authors analysed seasonal and trend patterns in yarn sales prices, which can assist higher management in making informed decisions in the textile industry.

Table 8

			Table 9
MEAN ABSOLUTE ERROR (MAE)			
Variable	Division A	Division B	Division C
Multivariate Regression	399.5926	397.3673	333.156
KNN	439.8063	327.1329	222.985
Neural Network	739.7477	572.9759	474.9306
Variable	STL	Additive Decomposition	Multiplicative Decomposition
ARIMA	598.699	510.249	678.671

Table 10

ROOT MEAN SQUARE ERROR (RMSE)			
Variable	Division A	Division B	Division C
Multivariate Regression	436.0006	438.623	398.502
KNN	571.7995	393.906	285.082
Neural Network	921.8554	740.0367	621.1331
Variable	STL	Additive Decomposition	Multiplicative Decomposition
ARIMA	610.524	530.868	688.37

Secondly, the authors utilized different machine learning algorithms to predict yarn sales prices. They used the ARIMA model to forecast prices based on historical trends and Multivariate Regression to get better results. To reduce errors, they also used the KNN model to predict sales prices. Additionally, they included neural networks in their study, as they are a widely studied tool for problem-solving in various fields and have previously been used to predict stock prices and yarn properties such as strength in the previous studies [8]. Finally, the authors evaluated the best-predicted model using traditional metrics that were not previously published in the literature.

CONCLUSION

The methods presented in this study can be applied to similar datasets to predict other variables of interest and subsequently facilitate the management of an industry in decision-making. This paper introduces a new approach to predicting yarn sales prices by using advanced data mining techniques and deep learning algorithms. Furthermore, trends and seasonal patterns in the data were analysed to predict yarn sales prices which is a novel approach compared to previous studies that used traditional techniques. This study's unique combination of features and machine learning algorithms is expected to provide valuable insights for decision-makers in the textile yarn manufacturing industry.

ACKNOWLEDGEMENTS

The authors are pleased to acknowledge NED University of Engineering & Technology, Karachi, for the moral support during the conduct of this study.

REFERENCES

- [1] Hwang, H.J., Seruga, J., An Intelligent Supply Chain Management System to Enhance Collaboration in Textile Industry, In: International Journal of u- and e- Service, Science and Technology, 2011, 4, 4, 47–62
- [2] Pranav, G.C., Jaju, S.B., *Designing innovative framework for supply chain performance measurement in textile industry*, In: International Journal of Logistics Systems and Management, 2014, 18, 2, 216–230
- [3] Pranav, G.C., Jaju, S.B., *Identification of Performance Measures for Textile Supply Chain: Case of Small & Medium Size Enterprise*, In: International Journal of Supply Chain Management, 2015, 4, 3, 50–58
- [4] Kogg, B., Greening a Cotton-textile Supply Chain: A Case Study of the Transition towards Organic Production without a Powerful Focal Company, In: Greener Management International, 2003, 43, 53–64
- [5] Diabat, A., Kannan, D., Mathiyazhagan, K., Analysis of enablers for implementation of sustainable supply chain management – A textile case, In: Journal of Cleaner Production, 2014, 83, 391–403
- [6] Ravindra Babu, M., Reddy, D.B.R., *Inventory Management Practices in Spinning Mills in South India*, In: International Journal of Management Studies, 2019, VI 1, 1, 109–113
- [7] Ramesh, M.C., Rajamanickam, R., Jayaraman, S., The Prediction of Yarn Tensile Properties by Using Artificial Neural Networks, In: The Journal of The Textile Institute, 1995, 86, 3, 459–469
- [8] Cheng, L., Adams, D.L., Yarn Strength Prediction Using Neural Networks: Part I: Fiber Properties and Yarn Strength Relationship, 1995, 65, 9, 495–500

- [9] Cox, J.E., Loomis, D.G., Improving forecasting through textbooks A 25 year review, In: International Journal of Forecasting, 2006, 22, 3, 617–624
- [10] Cen, Z., Wang, J., Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer, In: Energy, 2019, 169, 160–171
- [11] Hossain, M.A., et al., *Hybrid Deep Learning Model for Stock Price Prediction*, In: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), 2018
- [12] Henrique, B.M., Sobreiro, V.A., Kimura, H., Stock price prediction using support vector regression on daily and up to the minute prices, In: The Journal of Finance and Data Science, 2018, 4, 3, 183–201
- [13] Lei, L., Wavelet Neural Network Prediction Method of Stock Price Trend Based on Rough Set Attribute Reduction, In: Applied Soft Computing, 2018, 62, 923–932
- [14] Ryu, K., Sanchez, A., The Evaluation of Forecasting Methods at an Institutional Foodservice Dining Facility, In: The Journal of Hospitality Financial Management, 2003, 11, 1, 27–45
- [15] Weatherford, L.R., Kimes, S.E., *A comparison of forecasting methods for hotel revenue management*, In: International Journal of Forecasting, 2003, 19, 3, 401–415
- [16] Xiaohui, Y., The Prediction of Gold Price Using ARIMA Model, In: 2nd International Conference on Social Science, Public Health and Education (SSPHE 2018), Atlantis Press, 2019
- [17] Venkataraman, D., et al., Yarn price prediction using advanced analytics model, In: 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2016
- [18] Yamin, D.M.H., et al., *The Evaluation of Forecasting Methods at an Institutional Foodservice Dining Facility*, In: International Journal of Engineering and Management Research (IJEMR), 2018, 8, 4, 87–89
- [19] Chen, D.T., Bessler, D.A., Forecasting monthly cotton price: Structural and time series approaches, In: International Journal of Forecasting, 1990, 6, 1, 103–113
- [20] Achal, L., et al., *Modelling and Forecasting of Price Volatility: An Application of GARCH and EGARCH Models*, In: Agricultural Economics Research Review, 2015, 28, 1, 73–82
- [21] Crawford, A, Switha, V.G., *Price Volatility in the Cotton Yarn Industry: Lessons from India*, In: International Institute for Sustainable Development, 2007
- [22] Staff, H.T.T., Cotton Demand, Price Stabilizing, In: Home Textiles Today, 2007
- [23] Gargano, A., Timmermann, A., Forecasting commodity price indexes using macroeconomic and financial predictors, In: International Journal of Forecasting, 2014, 30, 3, 825–843
- [24] Han, J., Kamber, M., Pei, J., Data Mining Concepts and Techniques, 3rd ed., Elsevier, 2012, 86–90
- [25] Wang, X., Smith, K., Hyndman, R., Characteristic-Based Clustering for Time Series Data, In: Data Mining and Knowledge Discovery, 2006, 13, 3, 335–364
- [26] Hyndman, R.J., Athanasopoulos, G., Forecasting: principles and practice, in Chapter 2 Time series graphics, OTexts: Melbourne, Australia, 2018
- [27] Hyndman, R.J., Athanasopoulos, G., *Forecasting: principles and practice*, in Chapter 6 Time series decomposition, OTexts: Melbourne, Australia, 2018
- [28] Schlicht, E., A Seasonal Adjustment Principle and a Seasonal Adjustment Method Derived from This Principle, In: Journal of the American Statistical Association, 1981, 76, 374, 374–378
- [29] Hyndman, R.J., Athanasopoulos, G., Forecasting: principles and practice, in Chapter 8 ARIMA models, OTexts: Melbourne, Australia, 2018

Authors:

MUHAMMAD ALI¹, SAMAN HINA², SHERAZ HUSSAIN SIDDIQUE¹, RUKSHAN TAUFIQ LODHI³

¹NED University of Engineering & Technology, Department of Textile Engineering, University Road, 75270, Karachi, Pakistan

²NED University of Engineering & Technology, Department of Computer Science & IT, University Road, 75270, Karachi, Pakistan

³Contour Software, Karachi, Pakistan

Corresponding author:

MUHAMMAD ALI

e-mail: alimughal@neduet.edu.pk