RESEARCH ARTICLE

SAMRIDDHI Volume 15, Issue 1, 2023 Print ISSN: 2229-7111 Online ISSN: 2454-5767

Big Data Analysis for Revenue and Sales Prediction using Support Vector Regression with Auto-regressive Integrated Moving Average

Gagan K. Sharma*, Sunil Patil

Department of Computer Science and Engineering RKDF University, Bhopal, Madhya Pradesh, India

Abstract

In e-commerce industry, customers' demands get fluctuated throughout the year depending on the purchasing behavior and season. It may be a repetition period in the year, where sales may generally be down, moderate, and whilst some periods are extremely high. Studies reveal that machine learning techniques boosted much e-commerce industry, from supply chain management to business planning. In this paper, a hybrid big data analytical model which integrates Support Vector Regression (SVR) with Auto-Regressive Integrated Mov- ing Average (ARIMA) is proposed to predict product sales and revenues. The simulation results show that the proposed model presents lower relative error rate and higher accuracy that can be utilized for business planning and strategies.

Keywords: Big Data, Predictive Analysis, Machine Learning.

SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology (2023); DOI: 10.18090/samriddhi.v15i01.01

Introduction

Prediction is the fundamental process for business planning, sourcing, production, product distribution, customer service, and whole supply chain management. Prediction has become more critical and complex due to increasing competition in markets, wide variety of production, and globalization of supply chain.[1] The advanced digital technologies such as cloud computing, Internet of things (IoT), and social media arecollecting vast volume of real-time data that is referred to as "big data".[2-4] The collected vast volume of data (big data) can be analyzed to predict customers' demands, to improve apprehension of the customers' behavior, and to implement effective supply chain management.[5,6]

Before the beginning of the sales, customers demands must be predicted for successful business and e-commerce. The under-stocking (when supply of specific products do not fulfill demand of consumers) issues must be avoided by determining the inventory level needs of the customers. The sales predic- tions have direct impact on customer management, corporate financial market planning, and e-commerce strategies. Sales prediction is indispensable step for modern e-commerce in- dustries and business organizations, but, it experiences many issues and challenges.[7–10] However, there are many factorsthat directly or indirectly affect the sales prediction of the business. Figure 1 shows the factors that affect the sales prediction.

The basic economic factors such as national income, interestrates, profit rates, and price level decide the business potentialsand affect sales prediction. Business factors such as national plans, public opinion, taxation, and loan policies strike the revenues. The internal strategies and plans adopted by business industries, sociological factors (education level, rural-urban habitation, family income, demographic conditions), psycho- logical factors (learning skills, attitudes, habits, perceptions), and competitive factors (quality

Corresponding Author: Gagan K. Sharma, Department of Computer Science and Engineering RKDF University, Bhopal, Madhya Pradesh, India, e-mail: gagansharma.cs@gmail.com

How to cite this article: Sharma, G.K., Patil, S. (2023). Big Data Analysis for Revenue and Sales Prediction using Support Vector Regression with Auto-regressive Integrated Moving Average. *SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology,* 15(1), 1-8.

Source of support: Nil

Conflict of interest: None

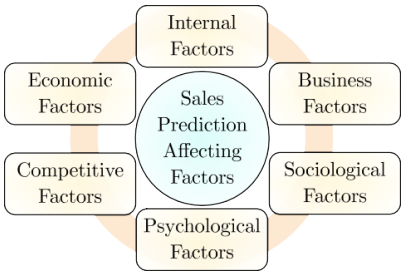

Figure 1: Sales prediction affecting factors

improvement, advertisement, technical competence) create impact on the revenue and sales prediction.[11]

Future customers' demands can be predicted based on historical information and data over certain time-period. Proper consumer demand prediction is an essential process for po- tential market growth plans, future sales, revenues, decision- making, and strategies. The effectiveness to predict business risks and bankruptcy is essential for financial institutions, as erroneous and inaccurate decisions may lead to financial con-sequences.[13] Risky business making decisions can impact on profitability and supply chain management without proper consumer's demand predictions. An appropriate prediction model can be applied based on the fresh products demands in e-commerce using big data analysis, and thus business can be optimized using future sales prediction.[14] The machine learning is applied for monitoring business processes and risk reduction, to predict the degree of risk based on the historical big data corresponding to the risk management.[15]

This article is structured as follows; Section II explains the essential impact of machine learning based big data analysis on financial risks and sales prediction in modern e-commerce business. Section III presents related works and literature surveys in the field of revenue and sales prediction using data analysis based on machine learning methods. Section IV presents the proposed integrated "support vector regression (SVR)" with "auto-regressive integrated moving average (ARIMA)" technique, formulation, and overall workflow to evaluate the results. Section V visualizes the python simulation results and comparisons with other methods. Finally, Section VI summarizes the conclusion of the paper with future recommendations.

Financial Risks and Sales Prediction in E-Commerce

In the e-commerce industry, financial risks may arise, andthey can affect business revenue, funds, manufacturing, supply chain management, distribution cycle, and also social redistri-bution. Financial risks are earnestly influencing e-commerce industries and society. Industries are facing difficulties in the form of financial risks, liquidations and bankruptcies are also raising. Therefore, the predictive analysis and the early cautionary techniques for financial risk management must be perpetually improved.[16] Consequently, Profitable business activities for the economy and society can be maintained using machine learning based big data analysis that can prevent the financial risks effectively.[17]

Predictive analysis on vital business risk management can avoid financial loss, and it can maximize the profit. Machine learning can be an auspicious source to business management as the process heavily relies on data-oriented decision-making predictions. The large adoption of machine learning basedpredictive techniques can be seen in recent year for many business risk management tasks.[18]

The risk management is directly associated with the predictive analysis because the effective prediction can significantly minimize the business risks. Based on the previous historical data, the revenues can be efficiently predicted using machine learning-based data analysis. The "Statista" (https://www.statista.com) company specialized in consumer and market data generate predictions about product revenues and sales of upcoming year using machine learning techniques. According to this company, its database contains over 1,000,000 statistical data on above 80,000 subjects from around 22,500 sources and 170 industries. As a sample, Figures 2 and 3 represent apparels and smartphones revenue report of India from the year 2013 to 2026. Based on the previous revenues and sales data from the year 2013 to 2021, the revenues of the year 2022 to 2026 were predicted using machine learning methods.

For financial risk protection in e-commerce business innovation, modern machine learning based big data analytical techniques are the significant landmark. In this paper, "support vector regression (SVR)" with "auto-regressive integrated moving average (ARIMA)" technique is applied to big data of Indian apparels and smartphone sales records for revenue prediction.

Related Work

With the progression of Internet technologies, the habit of customers have changed. The online marketing and sales are very common today that create the vast amount of data. Various researchers have proposed their methods to predict business strategies using big data analytics.

For newspaper/magazines' sales prediction issues, Yu *et al.*[19] used SVR method. They accomplished the lower structural risks in place of the lower empirical risks. Challa *et al.* [20] applied ARIMA to predict risks for companies listed in "Bombay stock exchange (BSE)".

Using traditional neural networks, Buyar and Abdel-Raouf[21] presented a model to predict sales of products. Their research helped business organizations for product stock evaluation, expansion and dividend-price ratio, using future sales prediction. Their model predicted with 3.54% of "mean absolute percentage error (MAPE)" using traditional neural networks and it perfored well in training data.

Sentiment analysis and time series analysis-based price prediction model was proposed by Tseng *et al.*[22] for e-commerce products. To enhance the prediction accuracy of their algorithm, they performed text analysis related to product information affected by electronic news. Their outcomes indicate that remarkable events and news influence the e-commerce products sale prices.

Based on the "Stable Seasonal Pattern (SSP)" and "Support Vector Regression (SVR)", Ye and Eskenazi[23] proposed a hybrid learning technique to enhance the prediction accuracy in stable trend and seasonality constraints. They applied SVR for sales value prediction and applied SSP technique for seasonality index calculation in monthly basis.

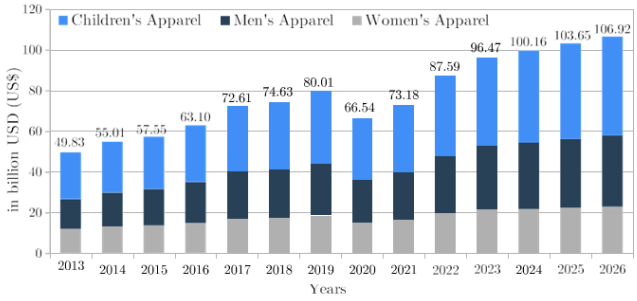

Figure 2: Apparel Report 2021 of India from "Statista Consumer Market Outlook"[12]

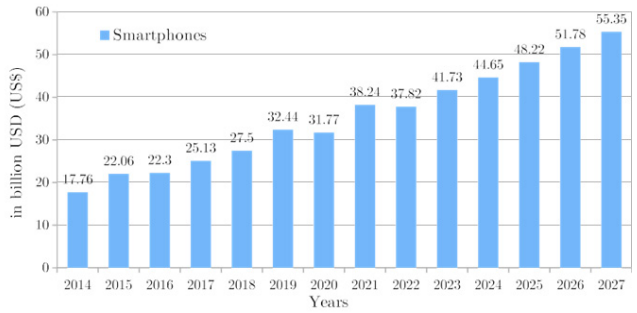

Figure 3: Smarphone report 2021 of India from "statista consumer market outlook"[12]

Duo *et al.*[24] applied differential ARIMA for prediction of short-term prices based on current e-commerce conditions. They established "back-propagation neural network" model forprediction of the long-term prices.

Dong *et al.*[25] established a model for sales prediction using historical e-commerce data. They analyzed the influenceof various characteristics that create the immense impact on the product sales. To channelize the business decision-making and to enhance the business profits, their established model is very helpful.

For analyzing the product demands and predictions in e- commerce, Chong *et al.*[26] used neural network modeling with big data technology. In their analysis, they used variables that are effective sales predictors of online products for better results. Their results reveal that the online customers reviews, promotions, and recommendations extracted from historical data analysis are applicable to develop a model for online e-commerce sales prediction. Their research shows that big data supported techniques to collect large scale data-sets can be applied efficiently, and the results help to predict demands of customers in e-commerce. This research shows that the demands of the products in the business can be predictedthrough the online marketing. Nowadays, the main platform for product selling is the e-commerce and business companies are focusing to improve product sales prediction.

Palanimalai and Paramasivam[27] discussed the several schemes to analyze Big data which can be exploited to channelize values of the business and modern perceptiveness using higher dimensional big data. For driving sales prediction

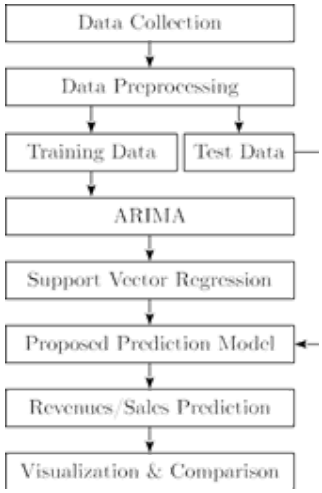

Figure 4: Over workflow of proposed method

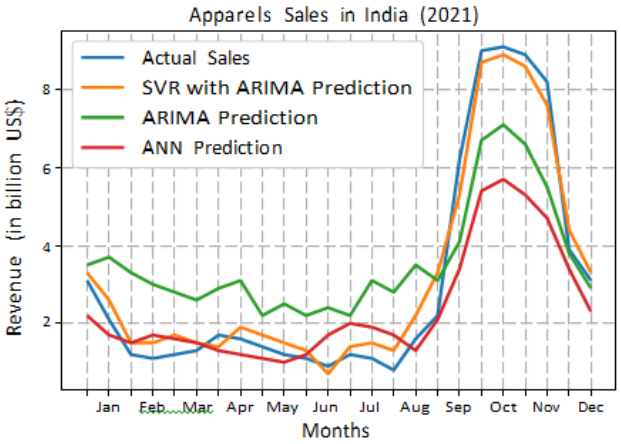

Figure 5: Apparel sales prediction.

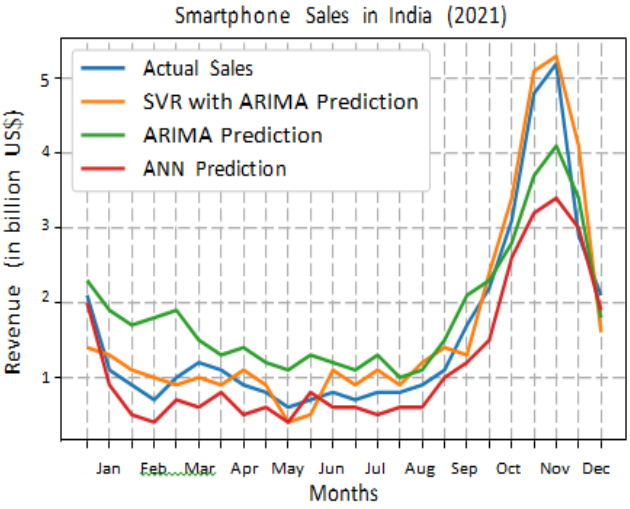

Figure 6: Smartphone sales prediction.

method, they conducted the experiment using "customer rela-tionship management (CRM)" database that used predicted,target and actual magnitude to aim business

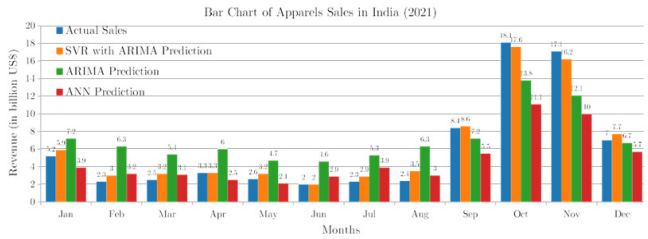

Figure 7: Bar plot of apparel sales prediction.

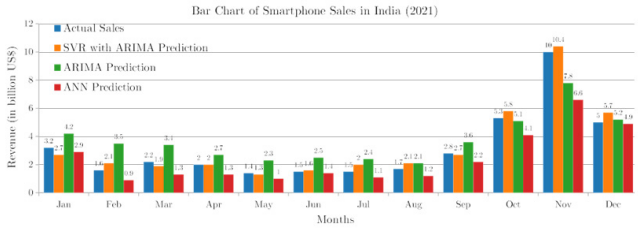

Figure 8: Bar Plot of Mobile Sales Prediction.

Table 1: Simulation software

Operating System	*Ubuntu 22.04 (Linux)*
Programming Language	Python 3.9 (Anaconda)
Dataset	India apparels, smartphone https://www.statista.com
Framework	PySpark
Data Preprocessing	Pandas, numpy
Machine Learning	scikit-learn, tensorflow, auto-arima, forecaster
Visualization	Matplotlib, seaborn

Table 2: Comparative Analysis For Indian Apparels Sales (2021)

	SVR with ARIMA Prediction	*ARIMA Prediction*	*ANN Prediction*
Revenue (billion US$)	77.10	85.60	56.90
Relative Error Rate	5.33%	16.94%	22.27%
Accuracy	94.67%	83.06%	77.73%

plannings andstrategies. The business analytical models help organizations torecognize and visualize patterns and trends, which can presentperceptive solutions on performance of the modern business. For prediction of product ratings, customers' sentiments, and recommendations, Woo and Mishra[28] investigated thedatabases of Amazon customer reviews and applied severalmachine learning algorithms using Spark and Azure. Forconventional machine learning with small datasets as well asmassive big data, both Spark and Azure are very effective tools.

To solve the complicated real-world business-to-business (B2B) sales issues, Bohanec *et al.*[29] demonstrated a novel black-box machine learning model-based application. Their

Table 3: Comparative Analysis For Indian Smartphone Sales (2021)

	SVR with ARIMA Prediction	*ARIMA Prediction*	*ANN Prediction*
Revenue (billion US$)	40.30	44.80	28.90
Relative Error Rate	5.49 %	17.28 %	24.34 %
Accuracy	94.51 %	82.72 %	75.66 %

results demonstrated the usability and effectiveness of the proposed model. A significant recommender system in theproduct sales strategies and measure action of sellers are the advantages of their proposed model.

Using supervised learning technique, Ma and Fildes[30] proposed a novel "convolution neural network (CNN)" based meta-learner that automatically acquire characteristics from unstructured time-series sales data and their essential factors. Their applied meta-learner significantly improves prediction of the retail sales.

Boldt *et al.*[31] tested the possibility of precise sales prediction for *Nike* products using *Facebook* data and observedthe effect of user's activities of Nike's Facebook posts onsales revenues. This research can help in business marketing strategies by assessing the information extracted from big social data.

Mishra[32] reviewed and discussed various big data analysis tools developed by computer scientists for significant business intelligent applications in prediction of financial development and management. The business intelligence toolsare effective for data-centric business organizations that can create practicable intelligent decision-making results in theanalyzable financial scenario.

Niu[33] proposed a novel technique for prediction of "Walmart" sales using "XGBoost" algorithm and feature engi-neering computation on datasets of Walmart sales. The results depict that this research can effectively extract significant features from data and performed better as compared to "Ridge Regression" and "Logistic regression". Chen *et al.*[34] proposed a neural network model for Walmart sales predictionusing their database. To train their proposed model, they applied "Tensor Flow" and "Keras", and the results represent that neural network based model outperforms as compared to support vector machine (SVM) and linear regression algo-rithm. Moreover, they utilized "SHapley Additive exPlanations(SHAP)"[35] framework to examine the significant features for their proposed model.

For fashion products sales prediction, Tehran and Ahrens used data analysis methods. In this research, the simula- tions were carried out using basic linear regression method. They also exploited the "unsupervised machine learning" techniques for target fitting of datasets.

Sohrabpour *et al.*[37] applied genetic programming

model to fluctuating export sales record for Middle Eastern business organizations and some considerable important features. It resulted in a genetic programming-based export sales model that assess a prediction for six weeks and sensitivity analysis of the variables. The error metrics mean absolute error, mean squared error, correlation coefficient, and R-squared goodness of fit were calculated to assess the model's prediction quality.

Qu *et al.*[38] proposed a multi-factor "grey wolf optimizer-support vector regression (GWO-SVR)" based model for sales prediction of automobile using external affecting factors and star rating of customers. The effective generalization, basic elementary structure and global optimization are the advantages of SVR model. Significantly, it is well suited for small datasets and multi-dimensional large datasets, that are appropriate in sales prediction considering influencing factors and features. Consequently, the SVR method is applied to train model. The GWO algorithm is further applied to minimize the year 2021. Indian apparels and smartphone sales mainly depends on the seasonal festivals, so the ARIMA based time series analysis with SVR is the best suited method for sales prediction.

ARIMA

For prediction in the time series analysis, "Auto-Regressive Integrated Moving Average (ARIMA)"[39,40] model was established. It basically predicts the future values based on the past values using time-series analysis over a specified time-period. A special statistical method is applied to present the series based on the time-series auto-correlation analytics. With the help of present and past data of time-series, the future values are predicted. Instability of noise data and investigation method's independency is considered for the ARIMA models prediction process.

The ARIMA (*p,d,q*) model is expressed in polynomial form using *I*(d) process having d^{th} integer differences that examines the fixed ARMA (*p,q*) process[41,42] as presented in Equation 1:

$$T_t = \sum_{i=1} \phi_i T_{t-1} + \sum_{i=1} \vartheta_i \varepsilon_{t-1} + \varepsilon_t \qquad (1)$$

Where T_t = The value of difference time series.
Φ, θ = The unknown variables.
ε = Error term with zero mean which is identically and independently distributed.

T_t is usually presented as the past and the current values and also known as the past error term. The three essential conditions of ARIMA model known as stationarity, invertibility and parsimony are applied to identify, estimate and verify the characteristics. Over the time period, mean, variance and co-variance remain constant. This can be completed using differing the first or second order integral. Using "convergent auto-regressive process", T_t is calculated in invertibility step.

Finally, the parsimony step produces considerable prediction with additive coefficient. The performance of this model will be better than the parameterized models.

The actual data of previous sales records is denoted by the matrix D [a, b], where a, b represents the dimensions of the data, the sales of the companies in i^{th} year and j^{th} is denoted by S (*i, j*), Where $i = 1, 2,..., n$ represents the years and $j = 1, 2,..., 12$ represents the months. So, the total sales in i^{th} year is calculated as

$$T_i = \sum_{j=1}^{12} S(i,j) \qquad (2)$$

The total sales from beginning to n^{th} year is calculated as

$$T_n = \sum_{i=1}^{n} \sum_{j=1}^{12} S(i,j) \qquad (3)$$

The seasonal index I_S *(i, j)* of the j^{th} month in the i^{th} year is calculated as:

$$I_{S(i,j)} = \frac{S(i,j)}{T_n} \qquad (4)$$

where, $i = 1, 2,..., n$; $j = 1, 2,..., 12$. The monthly average seasonal index represents more accurate seasonality if big data of sales record pattern is stable in periodic season. So, the monthly basis seasonal index is calculated as: The seasonal index $I_{S(i,j)}$ of the j^{th} month in the i^{th} year is calculated as:

$$I^{S(j)} = \frac{1}{n} \sum_{i=1}^{n} I^{S(i,j)} \qquad (5)$$

Support Vector Regression (SVR)

The support vector regression (SVR)[43,44] generally uses the kernal-based transformations. For minimization of regression error margin and maximization of "between class" margin, this kernal-based transformation is used from "original attribute space" to the "higher dimensional space" based on the "Vapnik's statistical learning theory".[45] The non-linear problem can be transformed into the higher dimensional space by applying a selected kernel function.

SVR goes through the "support vector machine" for regression. The basic form of linear regression to express the predicted value y is:

$$y(x) = w^T \phi(x) + c \qquad (6)$$

The regular error function *c* can be minimized as:

$$c = \frac{1}{2} \sum_{n=1}^{N} \{y_n - D_n\}^2 + \frac{\lambda}{2} \|w\|^2 \qquad (7)$$

The multiple independent variables from input vector x, with the static feature-spaced transformations Φ_x, and the actual data of previous sales records *D* are applied to calculate predicted value *y*.

Simulation Result and Analysis

For customers, the Apparel markets in India contain all types of clothing segments which are manufactured, and it is categorized into the main three products: Women's Apparels, Men's Apparels and Children's Apparels. The datasets contain sales record of clothes only, and it does not contain additional personal accessories such as handbags, jewelries,

wristwatches and similar items. Along-with the world's most popular brands, many small local, middle, and large-scale brands that work successfully in all regional market in India are included in the datasets. The simulations under Ubuntu (Linux) operating system using Python programming language in PySpark framework are performed, the detailed software and tools used to generate results is presented in Table 1.

According to the https://www.statista.com/[12], in 2022, the expected revenue from the apparel market in India amount to 87.59 billion US$ and the expected annual growth from 2022 to 2026 is 5.1%. The women's apparel contains the largest portion of the total apparels market volume.

The simulations are carried out using big data sales records (datasets) from Indian market of apparels and smartphone.

The results of proposed method is compared with alone ARIMA prediction and "artificial neural network (ANN)" prediction. The Figure 5 represents simulation results that shows the revenue prediction of Indian apparels market. The actual revenue is represented by blue curve, the predicted revenue is represented by the orange curve of proposed model, whereas, green and red curves represent the alone ARIMA and (ANN) based predictions. The Figure 6 represents simulation results to show Indian smartphone market's revenue prediction.

Similarly, blue, orange, green, and red curves represent actual, proposed, alone ARIMA, and ANN predictions respectively. The alternative bar chart presentation of the result from Figure 5 and 6 is visualized in Figure 7 and 8 respectively. By observing the simulation results from Figure 5-8, it is obvious that the proposed integrated method (SVR with ARIMA) represents more accurate prediction as compared to alone ARIMA and ANN.

Finally, the relative error rate E_R is calculated as:

$$E_R = \frac{|V_A - V_P|}{V_A} \times 100\% \qquad (8)$$

where, V_A and V_P denotes the actual and predicted value respectively. The accuracy is calculated as:

$$\text{Accuracy} = 100\% - E_R \qquad (9)$$

In the year 2021, the actual revenue generated in Indian apparels sales is 73.2 billion US $, and the actual smartphone sales in India is 38.2 billion US $ (https://www.statista.com). Table 2 and 3 shows the results (relative error and accu- racy) generated by ARIMA, ANN, and proposed (SVR with ARIMA) method. By taking Indian apparel sales dataset of the year 2021, the alone ARIMA model predicted 85.6 billion US $ revenue that is 83.06% accurate, the ANN predicted 56.9 billion US $ revenue that is 77.73% accurate, and the proposed model (SVR with ARIMA) predicted 77.1 billion US $ revenue that is 94.67% accurate (Table 2). Similarly, by taking Indian smartphone sales dataset of the year 2021, the alone ARIMA model predicted 44.8 billion US $ revenue that is 82.72% accurate, the ANN predicted 28.9 billion US $ revenue that is 75.66% accurate, and the proposed model (SVR with ARIMA) predicted 40.3 billion US $ revenue that is 94.51% accurate (Table 3). The ARIMA method is perfect for time-series analysis based prediction, and SVR supported ARIMA improves the prediction accuracy significantly. The ANN is very effective method in many fields, however it is not suitable choice for this prediction. As presented in Table II and III, the relative error rate is also minimum for proposed method as compared to alone ARIMA and ANN methods taking Indian apparels and smartphone sales data of the year 2021.

Conclusion

The machine learning techniques with big data analysis created significant impact on the progressive growth, strategy, plan- ning, and management in e-commerce business industries. The ARIMA is an effective model to predict time-series seasonal data. This research presented an integrated SVR with ARIMA method to predict sales and revenues of Indian apparel and smartphone for the year 2021. The simulation results represent that the revenues and sales prediction in proposed method are near to the actual record. As compared to alone ARIMA and ANN, the proposed method predicts revenues more accurately with minimum relative error rate.

References

[1] T. Boone, R. Ganeshan, A. Jain, and N. R. Sanders, "Forecasting sales in the supply chain: Consumer analytics in the big data era," International Journal of Forecasting, vol. 35, no. 1, pp. 170–180, 2019, special Section: Supply Chain Forecasting. doi: https://doi.org/10.1016/j.ijforecast.2018.09.003

[2] B. Schmarzo, Big Data: Understanding How Data Powers Big Business, 1st ed. Wiley Publishing, 2013.

[3] RKDF University Journal of Science and Engineering, vol. 01, no. 3, pp. 5–8, Dec 2018.

[4] S. Makkar and S. Jaiswal, "Predictive analytics on e- commerce annual sales," in Proceedings of Data Analytics and Management, D. Gupta, Z. Polkowski, A. Khanna, S. Bhattacharyya, and O. Castillo, Eds. Singapore: Springer Nature Singapore, 2022, pp. 557–567. doi: https://doi.org/10. 1007/978-981-16-6289-8 46

[5] R. N. Doug Hudgeon, Machine Learning for Business: Using Amazon SageMaker and Jupyter, 1st ed. Manning Publications, 2020.

[6] D. S. B. Kai R. Larsen, Automated Machine Learning for Business. Oxford University Press, 2021. doi: https://doi.org/10.1093/oso/9780190941659.001.0001

[7] G. Sharma and S. Patil, "Extreme gradient boosting model-based forecasting of big data online sales record," SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology, vol. 14, no. 1, pp. 112–119, Mar 2022. doi: https://doi.org/10.18090/samriddhi.v14i01.18

[8] A. Diamantopoulos and B. Mathews, "Factors affecting the nature and effectiveness of subjective revision in sales forecasting: An empirical study," Managerial and Decision Economics, vol. 10, no. 1, pp. 51–59, 1989. doi: https://doi.org/10.1002/mde.4090100106

[9] Statista Consumer Market Outlook, "Apparel report 2021," https://www.statista.com/outlook/cmo/apparel/india#revenue, 2021.

[10] W.-Y. Lin, Y.-H. Hu, and C.-F. Tsai, "Machine learning in financial

crisis prediction: A survey," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 4, pp. 421–436, 2012. doi: https://doi.org/10.1109/TSMCC.2011.2170420

[11] M. Zhang, Z. Zhou, Z. Liu, and J. Dai, "Case teaching of e-commerce fresh product demand forecasting in logistics system forecasting," in Advances in Computer Science for Engineering and Education, Z. Hu, I. Dychka, S. Petoukhov, and M. He, Eds. Cham: Springer International Publishing, 2022, pp. 437–442. doi: https://doi.org/10.1007/978-3-031-04812-8 37

[12] S. Jomthanachai, W.-P. Wong, and C.-P. Lim, "An application of data envelopment analysis and machine learning approach to risk management," IEEE Access, vol. 9, pp. 85 978–85 994, 2021. doi: https://doi.org/10.1109/ACCESS.2021.3087623

[13] Y. Cao, Y. Shao, and H. Zhang, "Study on early warning of e-commerce enterprise financial risk based on deep learning algorithm," Electronic Commerce Research, vol. 22, no. 1, pp. 21–36, Mar 2022. doi: https://doi.org/10.1007/ s10660-020-09454-9

[14] K. Peng and G. Yan, "A survey on deep learning for financial risk prediction," Quantitative Finance and Economics, vol. 5, no. 4, pp. 716–737, 2021. doi: https://doi.org/10.3934/QFE. 2021032

[15] A. Mashrur, W. Luo, N. A. Zaidi, and A. Robles-Kelly, "Machine learning for financial risk management: A survey," IEEE Access, vol. 8, pp. 203 203–203 223, 2020. doi: https://doi.org/10.1109/ACCESS.2020.3036322

[16] X. Yu, Z. Qi, and Y. Zhao, "Support vector regression for newspaper/magazine sales forecasting," Procedia Computer Science, vol. 17, pp. 1055–1062, 2013. doi: https://doi.org/10.1016/j.procs.2013.05.134

[17] M. L. Challa, V. Malepati, and S. N. R. Kolusu, "Forecasting risk using auto regressive integrated moving average approach: an evidence from s&p bse sensex," Financial Innovation, vol. 4, no. 1, p. 24, Oct 2018. doi: https://doi.org/10.1186/ s40854-018-0107-z

[18] V. Buyar and A. Abdel-Raouf, "A convolutional neural networks-based model for sales prediction," in Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control, ser. AIRC '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 61–67. doi: https://doi.org/10.1145/3388218.3388228

[19] K.-K. Tseng, R. F.-Y. Lin, H. Zhou, K. J. Kurniajaya, and Q. Li, "Price prediction of e-commerce products through internet sentiment analysis," Electronic Commerce Research, vol. 18, no. 1, pp. 65–88, Mar 2018. doi: https://doi.org/10.1007/s10660-017-9272-9

[20] F. Ye and J. Eskenazi, "Sales forecast using a hybrid learning method based on stable seasonal pattern and support vector regression," in Emerging Technologies for Information Systems, Computing, and Management, W. E. Wong and T. Ma, Eds. New York, NY: Springer New York, 2013, pp. 1251–1259. doi: https://doi.org/10.1007/978-1-4614-7010-6 139

[21] Z.-w. Dou, M.-x. Ji, M. Wang, and Y.-n. Shao, "Price prediction of pu'er tea based on arima and bp models," Neural Computing and Applications, vol. 34, no. 5, pp. 3495–3511, Mar 2022. doi: https://doi.org/10.1007/s00521-021-05827-9

[22] W. Dong, Q. Li, and H. V. Zhao, "Statistical and machine learning-based e-commerce sales forecasting," in Proceedings of the 4th International Conference on Crowd Science and Engineering, ser. ICCSE'19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 110–117. doi: https://doi.org/10.1145/3371238.3371256

[23] A. Y. L. Chong, E. Ch'ng, M. J. Liu, and B. Li, "Predicting consumer product demands via big data: the roles of online promotional marketing and online reviews," International Journal of Production Research, vol. 55, no. 17, pp. 5142–5156, 2017. doi: https://doi.org/10.1080/00207543.2015.1066519

[24] S. Palanimalai and I. Paramasivam, "Big data analytics bring new insights and higher business value - an experiment carried out to divulge sales forecasting solutions," Int. J. Adv. Intell. Paradigms, vol. 8, no. 2, pp. 207–218, apr 2016. doi: https://doi.org/10.1504/IJAIP.2016.075728

[25] J. Woo and M. Mishra, "Predicting the ratings of amazon products using big data," WIREs Data Mining and Knowledge Discovery, vol. 11, no. 3, p. e1400, 2021. doi: https://doi.org/10.1002/widm.1400

[26] M. Bohanec, M. Kljajic´ Borsˇtnar, and M. Robnik-Sˇikonja, "Explaining machine learning models in sales predictions," Expert Systems with Applications, vol. 71, pp. 416–428, 2017. doi: https://doi.org/10.1016/j.eswa.2016.11.010

[27] S. Ma and R. Fildes, "Retail sales forecasting with meta- learning," European Journal of Operational Research, vol. 288, no. 1, pp. 111–128, 2021. doi: https://doi.org/10.1016/j. ejor.2020.05.038

[28] L. C. Boldt, V. Vinayagamoorthy, F. Winder, M. Schnittger, M. Ekran, R. R. Mukkamala, N. B. Lassen, B. Flesch, A. Hussain, and R. Vatrapu, "Forecasting nike's sales using facebook data," in 2016 IEEE International Conference on Big Data (Big Data), 2016, pp. 2447–2456. doi: https://doi.org/10.1109/BigData.2016.7840881

[29] S. Mishra, "Financial management and forecasting using business intelligence and big data analytic tools," International Journal of Financial Engineering, vol. 05, no. 02, p. 1850011, 2018. doi: https://doi.org/10.1142/S2424786318500111

[30] Y. Niu, "Walmart sales forecasting using xgboost algorithm and feature engineering," in 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), 2020, pp. 458–461. doi: https://doi.org/10.1109/ICBASE51474.2020.00103

[31] J. Chen, W. Koju, S. Xu, and Z. Liu, "Sales forecasting using deep neural network and shap techniques," in 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), 2021, pp. 135– 138. doi: https://doi.org/10.1109/ICBAIE52039.2021.9389930

[32] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proceedings of the 31st International Conference on Neural Information Processing Systems, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 4768–4777.

[33] A. F. Tehrani and D. Ahrens, Improved Forecasting and Purchasing of Fashion Products based on the Use of Big Data Techniques. Wiesbaden: Springer Fachmedien Wiesbaden, 2016, pp. 293–312. doi: https://doi.org/10.1007/ 978-3-658-08809-5 13

[34] V. Sohrabpour, P. Oghazi, R. Toorajipour, and A. Nazarpour, "Export sales forecasting using artificial intelligence," Technological Forecasting and Social Change, vol. 163, p. 120480, 2021. doi: https://doi.org/10.1016/j.techfore.2020.120480

[35] F. Qu, Y.-T. Wang, W.-H. Hou, X.-Y. Zhou, X.-K. Wang, J.-B. Li, and J.-Q. Wang, "Forecasting of automobile sales based on support vector regression optimized by the grey wolf optimizer algorithm," Mathematics, vol. 10, no. 13, 2022. doi: https://doi.org/10.3390/math10132234

[36] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control, 4th ed., ser. Wiley Series in Probability and Statistics. Wiley, 2008. doi: https://doi.

org/10.1002/9781118619193
[37] J. D. Hamilton, Time Series Analysis. Princeton University Press, 1994.
[38] P. J. Brockwell and R. A. Davis, Eds., ARMA Models. New York, NY: Springer New York, 2002, pp. 83–110. doi: https://doi.org/10.1007/0-387-21657-X 3
[39] P. A. Moran, "Hypothesis testing in time series analysis," Journal of the Royal Statistical Society: Series A (General), vol. 114, no. 4, pp. 579–579, 1951. doi: https://doi.org/10. 2307/2981095
[40] M. Awad and R. Khanna, Support Vector Regression. Berkeley, CA: Apress, 2015, pp. 67–80. doi: https://doi.org/10.1007/ 978-1-4302-5990-9 4
[41] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2011.
[42] F. F. Soulie´ and D. Girard, "Applications of vapnik's theory for prediction," in ICANN 98, L. Niklasson, M. Bode´n, and T. Ziemke, Eds. London: Springer London, 1998, pp. 41–50. doi: https://doi.org/10.1007/978-1-4471-1599-1 4