Journal of Soft Computing Paradigm (JSCP) (2020)
Vol.02/No.02
Pages: 101-110
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2020.2.007

Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics

Dr. V. Suma,
Professor,
Department of Information Science & Engineering,
Dayananda Sagar College of Engineering,
Shavige Malleshwara Hills, Kumarswamy Layout,
Bangalore, India.
E-mail id: suma-ise@dayanandasagar.edu

Abstract: There has been an increasing demand in the e-commerce market for refurbished products across India during the last decade. Despite these demands, there has been very little research done in this domain. The real-world business environment, market factors and varying customer behavior of the online market are often ignored in the conventional statistical models evaluated by existing research work. In this paper, we do an extensive analysis of the Indian e-commerce market using data-mining approach for prediction of demand of refurbished electronics. The impact of the real-world factors on the demand and the variables are also analyzed. Real-world datasets from three random e-commerce websites are considered for analysis. Data accumulation, processing and validation is carried out by means of efficient algorithms. Based on the results of this analysis, it is evident that highly accurate prediction can be made with the proposed approach despite the impacts of varying customer behavior and market factors. The results of analysis are represented graphically and can be used for further analysis of the market and launch of new products.

Keywords: Regression tree; Machine learning; Data mining; Holdout Cross Validation; Refurbished Electronics; Electronic gadgets;

1. Introduction

The revenue generation through refurbished products has greatly impacted the manufacturing industry [1]. Sales of products, especially electronic goods has increased rapidly over the past decade all across India. The large amount of environmental and economic benefit gained from these products act as the driving force for this rapid evolution [2]. The benefits of refurbished products include low price, reduced energy consumption, reduced

ISSN: 2582-2640 (online)
Submitted: 20.03.2020
Accepted: 16.05.2020
Published: 22.05.2020

Journal of Soft Computing Paradigm (JSCP) (2020)
Vol.02/No.02
Pages: 101-110
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2020.2.007

manufacturing emission, lesser raw materials and lesser production cost [3]. Due to the demand and return uncertainty, refurbishing cannot be the solution for a sustainable and promising business prospect. The uncertainty of the product quality and improper or lack of information regarding various aspects of the product with respect to its performance and looks are some of the causes for return uncertainty [4].

Testing, inspection, disassembling and cleaning of refurbished products cannot be observed, hence the quality of the product as well as the market demand for the product cannot be evaluated [5]. For efficient operation of closed-loop supply chain (CLSC), the demand of refurbished product has to be predicted in an accurate manner. Progress in marketing strategy and demand prediction is lesser on implementation of CLSC for refurbishing operational issues and management of acquisition of returned products. Return of product, product price and demand are some of the uncertain parameters in the CLSC technique that are to be quantified accurately by implementing sophisticated prediction techniques [6]. From these studies, it is evident that refurbishing domain requires a sophisticated analytical model for understanding the product acceptance in the market and facilitating development of products with higher relevance to the industry.

2. Related Works

Refurbished products related research has gained increasing attention during the past decade with the increasing demand for such products [7]. Revenue management, pricing analysis, warranty strategies and sales channels are the major areas that are considered while dealing with refurbished products. The comparison between the behavior of customers towards new products and refurbished ones is an emerging topic among the various areas of interest. A comparison of the machine learning approaches used for analyzing the behavior of the consumers with respect to the refurbished products is performed [8].

Traditionally, logistic regression techniques were used for processing the information. The drawbacks of these techniques are overcome by the machine learning approach and data-driven decision-making is achieved in this era of big-data [9]. Profitability as well as productivity of companies are improved significantly in companies using this technique as suggested by empirical evidences. Parametric approach is used in the logistic regression method where there is a predetermined model structure and simplified assumptions are fir by the relationship between the input and outputs [10]. However, machine learning technique determines the relationship between

ISSN: 2582-2640 (online)
Submitted: 20.03.2020
Accepted: 16.05.2020
Published: 22.05.2020

Journal of Soft Computing Paradigm (JSCP) (2020)
Vol.02/No.02
Pages: 101-110
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2020.2.007

the input and output based on datasets using algorithms and does not begin from the structure of the model. In case of high dimensional and noisy datasets, the unknown and complex non-linearity can be approximated better than the statistical techniques by implementing machine learning approaches [11].

The demand prediction accuracy of refurbished products can be improved by implementing sophisticated machine learning techniques as suggested by several literature surveys [12]. These techniques help in establishing an efficient marketing strategy for refurbished products. The major objectives covered in this paper includes implementation of machine learning technique for improving the accuracy and robustness of demand prediction of refurbished products as well as obtaining a thorough understanding of consumer behavior of refurbished products by analysis. Partial dependency plots and variable importance ranking can be used for this purpose. This helps in developing a practical and optimal marketing strategy [13]. These objectives are realized using three datasets from various e-commerce websites that are well known in the Indian market and 100 products are compared in this regard.

3. Proposed Work

Data mining and Cross Industry Standard Process is used for data analytics in this paper. The fundamental approaches used in this technique includes converting the refurbished product demand prediction based business objectives into problems related to data mining. The problem related variables are obtained and data source is identified by data understanding. Before data analysis, the data is structured appropriately using transforming and several data extraction techniques. This procedure is termed as data preparation. Validation, hyperparameter tuning, model development and selection of variables is done under predictive modelling. Measurement of various predefined error values are used for comparing the predictive performance and in evaluation of model. The managerial decision making is assisted by the deployment of model. These techniques complete the proposed framework as represented in Figure 1.

ISSN: 2582-2640 (online)
Submitted: 20.03.2020
Accepted: 16.05.2020
Published: 22.05.2020

Journal of Soft Computing Paradigm (JSCP) (2020)
Vol.02/No.02
Pages: 101-110
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2020.2.007

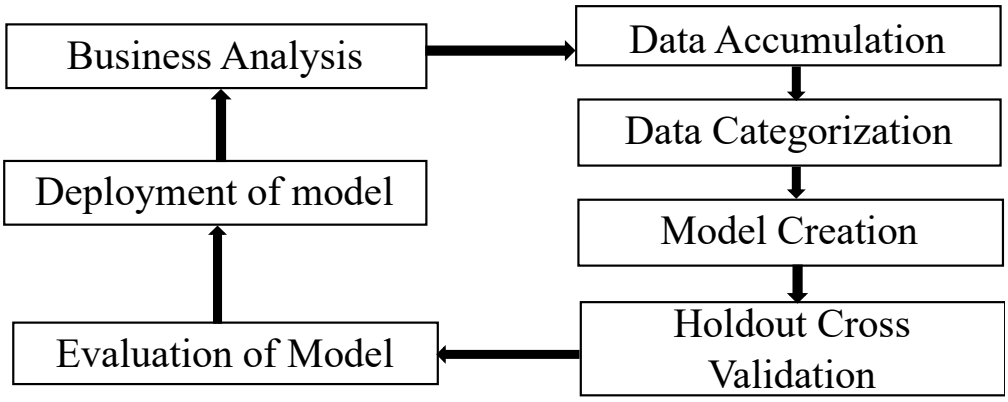

Figure 1: Proposed Framework

4. Data Collection and Processing

The data study sources are identified to be Amazon, Flipkart and Snapdeal. These are among the top e-commerce websites in India. Web crawlers are used for identifying 100 random refurbished electronic products in these websites for the purpose of analysis. The products are generally categorized as new, refurbished and used products. Original equipment manufacturers test and certify the refurbished products. They may also be qualified similar to new products and in appropriate working condition by third parties that refurbishes the product. These certifications have direct impact over the products similar to the closed-loop supply chain core products. In this paper, we focus on only the electronic products. The sales and transaction details are not revealed by these e-commerce websites. Hence, the customer demand is used as an indicator for predicting the sales performance. The customer demand and sales rank of the product are inversely proportional to each other.

Rather than the level, the natural logarithm transformation (ln) is used for estimation of sales rank due to the scale effects evident in the data. $Customer\ Demand = 1 \ln(Sales\ rank)$ is the represented as the expression of dependent variables. Here the product demand, termed as customer demand is a crucial component in the proposed model. The historical data related to the listed refurbished products are gathered using the python based

ISSN: 2582-2640 (online)
Submitted: 20.03.2020
Accepted: 16.05.2020
Published: 22.05.2020

Journal of Soft Computing Paradigm (JSCP) (2020)
Vol.02/No.02
Pages: 101-110
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2020.2.007

web crawler from all the three mentioned e-commerce websites related to the Indian market. The complete dataset that is publicly available on the product page is captured using this web crawler for the listing of each product. The data is monitored for a period of one month in order to reflect the variation in the recent sales and the updates in the sales ranks. Among the dependent and independent variables, the potential issue od simultaneity can be removed by prediction of sales rank. The time duration of one month is represented as t which is also the time of sales rank prediction. The prediction has been done over a period between Nov and Dec 2019. Electronic products that were compared includes GPS navigation, cameras, computers, laptops, cell phones and so on.

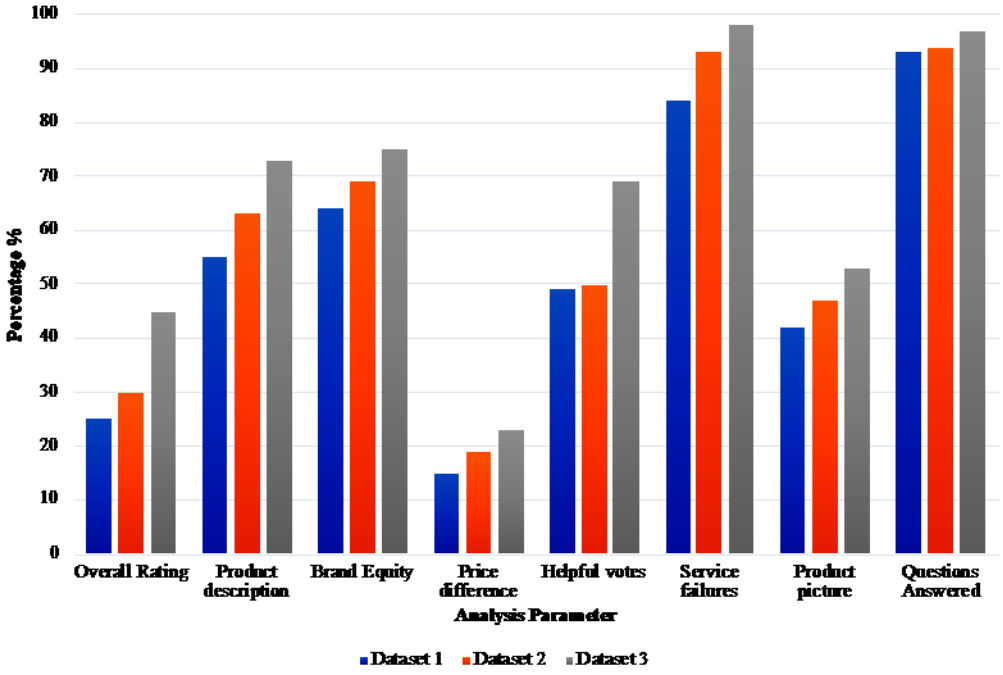

Figure 2: Ranking of importance of variables using sensitivity analysis scheme

ISSN: 2582-2640 (online)
Submitted: 20.03.2020
Accepted: 16.05.2020
Published: 22.05.2020

Journal of Soft Computing Paradigm (JSCP) (2020)
Vol.02/No.02
Pages: 101-110
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2020.2.007

5. Results and Discussion

The demand prediction model for refurbished products make use of certain variables that are selected using the aggregated and prepared datasets that are well structured. The machine learning algorithms tend to damage the performance of prediction, consume large amount of resources and sometimes slow down the algorithm due to several noisy and irrelevant variables, especially in huge datasets. In order to avoid these issues, it is crucial to select an appropriate variable in the machine learning application. The variable relevance concept is applied for selection of variables. The issues with respect to selection of variables are classified as all-relevant and minimal-optimal problem. The former focuses on identifying all the variables that are relevant both strongly and weakly; the later involves removal of redundant data that are available in the weakly relevant variable subsets while keeping the strongly relevant variable subset intact. Figure 2 represents the implementation of the sensitivity analysis scheme for ranking the importance of variables involved in the prediction of demand. The answered questions, product picture, service failure, helpful votes, price difference, brand equity, product description and overall rating for the three datasets are compared.

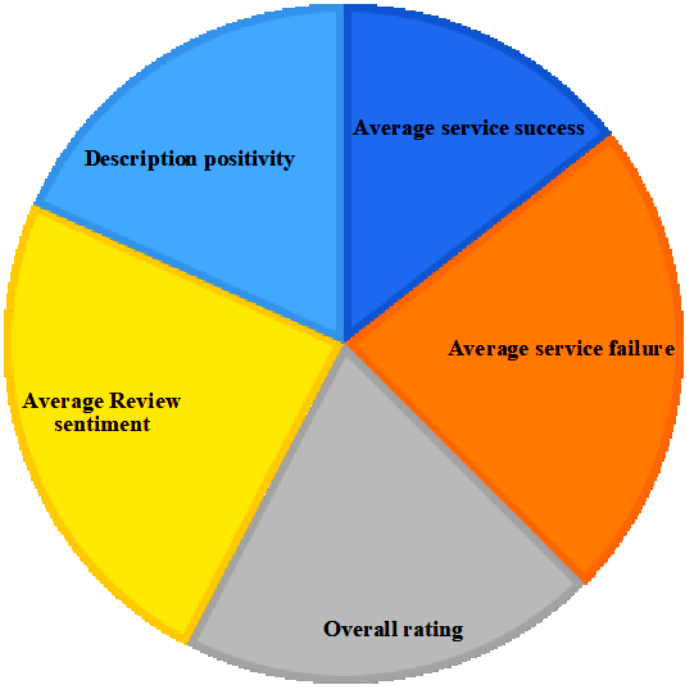

Figure 3: Consolidation of partial dependence plot analysis of various parameters

ISSN: 2582-2640 (online)
Submitted: 20.03.2020
Accepted: 16.05.2020
Published: 22.05.2020

Journal of Soft Computing Paradigm (JSCP) (2020)
Vol.02/No.02
Pages: 101-110
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2020.2.007

The all 3 datasets contain 100 random electronic products each from Amazon, Flipkart and Snapdeal. Out of the compared products in the first iteration, 10% of the products does not have new products that can be used for comparison of price. These products are removed from the dataset and are removed from the list and replaced with other products for which comparison is possible. It is also found that around 15% of the refurbished products does not contain any data regarding customer reviews. Since the reviews are crucial for predicting the demand of the products, these products are further removed and replaced with different products in the next iteration. With this, an appropriate dataset is created with products that contain all the required components for estimation and analysis. Figure 3 represents the comparison of parameters like overall rating, positivity of the product description, average success and failure of service for the product, and the average review sentiment of the product whose partial dependence plots are consolidated into a pie chart.

We have 3 databases of 300 products in total and 300 product pictures as well as 15,323 reviews of Indian customers. The dataset is further split into two subsets namely training and testing set of 60 and 40% of the overall data. Figure 4 represents the testing and training prediction results of the three datasets along with their weights. The sampling bias is reduced by repeating holdout cross validation scheme thrice during the entire process of modelling.

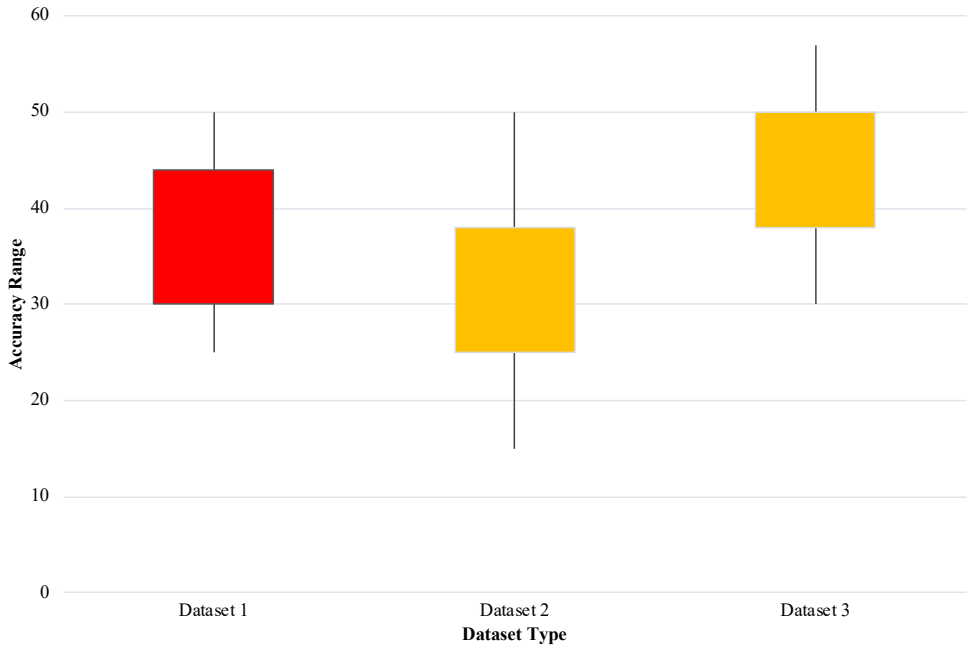

Figure 4: Testing and Training prediction results for the three datasets

ISSN: 2582-2640 (online)
Submitted: 20.03.2020
Accepted: 16.05.2020
Published: 22.05.2020

Journal of Soft Computing Paradigm (JSCP) (2020)
Vol.02/No.02
Pages: 101-110
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2020.2.007

6. Conclusion

For marketing strategy development and demand prediction of refurbished products in the Indian market, a lucid data mining scheme is proposed in this paper. For this purpose, real datasets from three prominent e-commerce websites are used. Robust and accurate results are obtained by the prediction using the regression tree model based on the results obtained. Using the prediction model, the market factors that greatly affect the product demand are determined. Partial dependency plots and variable importance ranking schemes are used for the purpose of prediction of these market factors. This work helps in identifying several insights into the market scenario based on the obtained data, which can be used for developing efficient marketing strategies and providing managerial guidelines for the refurbished products. On combining other research work with this. It is possible to perform higher end business analytics in the field of refurbished products. The results of the three datasets form the basis of this research. It is evident from the aggregation that the customer behavior changes based on the e-commerce platform as well as the product deal. The overview of the results make it look incomprehensible. It is essential to develop an efficient machine learning model that can help in understanding the pattern and predicting the market with greater accuracy. Closed-loop supply chain using reverse logistics and data mining schemes offers promising prospects in this domain. Future work is focused on improvising the machine learning algorithm to analyze the pattern in the big data for more accurate prediction of future market demands for refurbished products.

ISSN: 2582-2640 (online)
Submitted: 20.03.2020
Accepted: 16.05.2020
Published: 22.05.2020

Journal of Soft Computing Paradigm (JSCP) (2020)
Vol.02/No.02
Pages: 101-110
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2020.2.007

References

[1] Rosenbloom, R. S. (2000). Leadership, capabilities, and technological change: The transformation of NCR in the electronic era. Strategic Management Journal, 21(10-11), 1083-1103.

[2] Antonenkov, D. V., & Solovev, D. B. (2017, October). Mathematic simulation of mining company's power demand forecast (by example of "Neryungri" coal strip mine). In IOP Conference Series: Earth and Environmental Science (Vol. 87, No. 3, p. 032003). IOP Publishing.

[3] Bhuie, A. K., Ogunseitan, O. A., Saphores, J. D., & Shapiro, A. A. (2004, May). Environmental and economic trade-offs in consumer electronic products recycling: a case study of cell phones and computers. In IEEE International Symposium on Electronics and the Environment, 2004. Conference Record. 2004 (pp. 74-79). IEEE.

[4] Pathak, P., & Srivastava, R. R. (2017). Assessment of legislation and practices for the sustainable management of waste electrical and electronic equipment in India. Renewable and Sustainable Energy Reviews, 78, 220-232.

[5] Linoff, G. S., & Berry, M. J. (2011). Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons.

[6] Chen, Y. L., Chen, J. M., & Tung, C. W. (2006). A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales. Decision support systems, 42(3), 1503-1520.

[7] Brühl, B., Hülsmann, M., Borscheid, D., Friedrich, C. M., & Reith, D. (2009, July). A sales forecast model for the german automobile market based on time series analysis and data mining methods. In Industrial Conference on Data Mining (pp. 146-160). Springer, Berlin, Heidelberg.

[8] Changchien, S. W., Lee, C. F., & Hsu, Y. J. (2004). On-line personalized sales promotion in electronic commerce. Expert Systems with Applications, 27(1), 35-52.

[9] Banks, D. L., & Said, Y. H. (2006). Data mining in electronic commerce. Statistical Science, 234-246.

ISSN: 2582-2640 (online)
Submitted: 20.03.2020
Accepted: 16.05.2020
Published: 22.05.2020

Journal of Soft Computing Paradigm (JSCP) (2020)
Vol.02/No.02
Pages: 101-110
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2020.2.007

[10] Joseph, S. I. T., & Thanakumar, I. (2019). Survey of data mining algorithm's for intelligent computing system. Journal of trends in Computer Science and Smart technology (TCSST), 1(01), 14-24.

[11] Wang, H. (2019). SUSTAINABLE DEVELOPMENT AND MANAGEMENT IN CONSUMER ELECTRONICS USING SOFT COMPUTATION. Journal of Soft Computing Paradigm (JSCP), 1(01), 49-56.

[12] Bashar, A. (2019). INTELLIGENT DEVELOPMENT OF BIG DATA ANALYTICS FOR MANUFACTURING INDUSTRY IN CLOUD COMPUTING. Journal: Journal of Ubiquitous Computing and Communication Technologies September, 2019(01), 13-22.

[13] Pandian, A. P. (2019). Artificial intelligence application in smart warehousing environment for automated logistics. Journal of Artificial Intelligence, 1(02), 63-72.

ISSN: 2582-2640 (online)
Submitted: 20.03.2020
Accepted: 16.05.2020
Published: 22.05.2020