ISSN: 2586-7652 (Print) Vol. 03, No. 01, March 2020
ISSN: 2635-7607 (Online)
International Journal of Advanced Engineering
Source: http://ictaes.org
Manuscript received: December 10, 2019 ; Revised: December 30, 2019; Accepted: January 8, 2020

A study of the Demand Forecasting Model for Specific Books using Statistical Analysis and Machine Learning Method

Mee Hwa Park [1], Ill Chul Doo [2], Jong Sup Lee[3]*

[1,2] *Convergence Software Institute, Dongguk University-Seoul, Korea*
[3] *Development of smart community policing system(Googi) Center, Dongguk University-Seoul, Korea*

meehwap@dongguk.edu[1], mrdoo@dongguk.edu[2], jsleearmy@dongguk.edu[3]

Abstract

The purpose of this study is to identify the factors affecting the sale and return of specific books and to create a model to forecast sales demand. For this purpose, wd used the sales data of the books sold for 5 years (2012 ~ 2016) by A publishing company. In addition, we collected the data related to books in the Internet portal system and SNS site. As a result of this study, it was possible to identify the factors affecting the return and sales of a specific book, and it was possible to establish a sales order prediction model.
Keywords: *demand forecasting, Linear Regression Model., Gaussian Process Regression, Random Forest*

1. Introduction

Demand forecasting is an activity that uses historical data to establish a model that can identify trends and forecasts the future by using this model. Demand forecasting is widely used in a variety of business areas [1]. A company that publishes and sells books tries to generate more revenue by increasing book sales and reducing inventory losses. Using data analysis to predict accurate publishing demand and establishing countermeasures against factors that may cause returns can reduce the amount of losses incurred due to inventory control and returns.

2. FORECASTING METHOD

The purpose of this study is to identify the factors affecting the sale and return of specific books and to create a model to forecast sales demand. The study procedure is shown in Fig1. The variables used for prediction were used after standardization. The prediction models used in this study are Linear Regression, Gaussian Process Regression,

Corresponding Author : Jong Sup Lee
Author's affiliation : Dongguk University-Seoul
Email: jsleearmy@dongguk.edu
Copyright © ICT-AES

Support Vector Machine Regression, and Random Forest. The learning data used to generate the forecasting model is 256 books sold in 2014 and 2015, and the test data used to evaluate the accuracy of the model is 130 books in 2016. The variables used in the prediction are total 47, including the continuous variables and the categorical variables collected during the hypothesis verification process. The prediction model and accuracy are shown in Table 1. In the linear regression model, the explanatory power and prediction accuracy of the model are superior to those of the other models. The accuracy of the random forest model, which is a machine learning method, is the best.

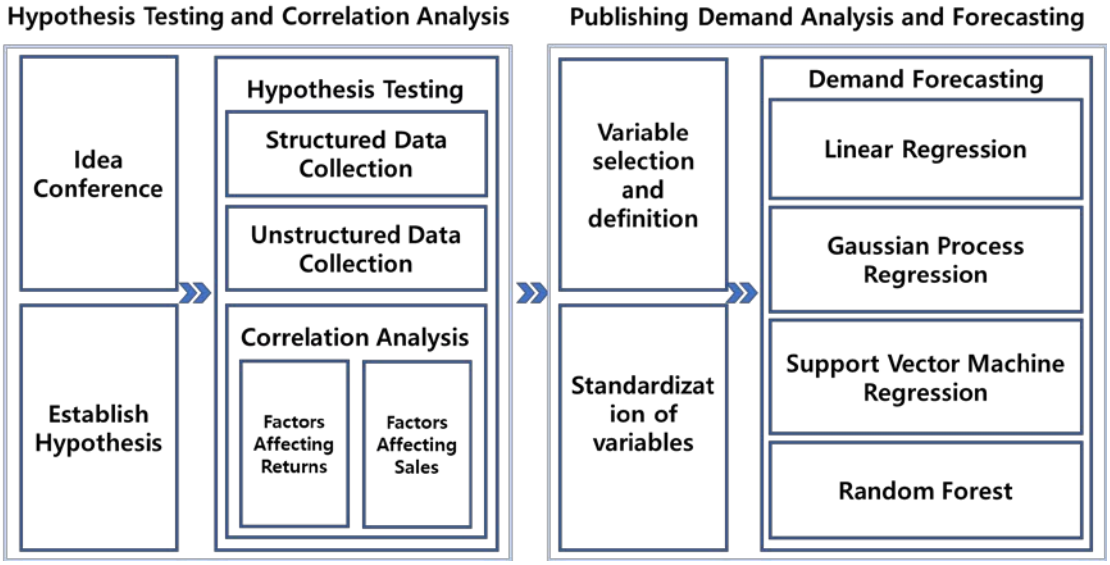


Figure 1. The Process of the Demand Analysis and Forecasting Research

Table 1. Accuracy of prediction models

Forecast model	Explanation(R^2)	Net sales Value	Predicted value	Accuracy
Linear Regression	0.96	11,603,352	11,953,805	97.0%
Gaussian Regression	0.93	11,603,352	11,214,397	96.6%
SVM Regression	0.87	11,603,352	11,017,538	94.9%
Random Forest	0.93	11,603,352	11,379,641	98.1%

3. CONCLUSION

The results of this study can be expected as follows. First, return loss cost decreases. Using company internal data will allow us to design a more accurate forecasting model and perform more accurate forecasting, which will reduce the cost of loss due to returns when used in publishing projects. Second, it can support the publishing and sales process. By analyzing various data, it is possible to understand the tendency of consumers and suppliers. Third, it can be used to forecast demand for various services. Based on the basic model proposed in this study, various services can be extended to the demand forecasting model. Using this information, we can find out the possibility of improving service by utilizing it in various business of company as well as publishing business.

ACKNOWLEDGEMENT

This research was supported by National Research Foundation of Korea (NRF) grant funded by the Korea

government (MSIT) (No. 2018R1A5A7023490).

References

[1] Schaer, O., et al., Demand forecasting with user-generated online information. International Journal of Forecasting (2018), https://doi.org/10.1016/j.ijforecast.2018.03.005.

[2] Getoor L, Taskar B, eds. Introduction to Statistical Relational Learning. Cambridge: MIT Press, 2007

[3] Magnani L. Abductive Cognition: the Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning. Berlin: Springer, 2009

[4] Mooney R J. Integrating abduction and induction in machine learning. In: Abduction and Induction. Dordrecht: Kluwer Academic Publishers, 2000. 181–191

[5] Muggleton S H, Bryant C H. Theory completion using inverse entailment. In: roceedings of the 10th International Conference on Inductive Logic Programming, London, 2000. 130–146

[6] Dai W-Z, Zhou Z-H. Combining logic abduction and statistical induction: discovering written primitives with human knowledge. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Francisco, 2017. 2977–2983

[7] Zhou Z-H. A brief introduction to weakly supervised learning. Natl Sci Rev, 2018, 5: 44–53