Predicting e-commerce orders returns and cancellations using machine learning

Master Thesis Business Analytics

HMN Yousaf

VU supervisor: Prof.dr. Sandjai Bhulai

VU second reader: Dr. Bram Gorissen

PVH supervisor: Nanne Sluis

Vrije Universiteit Amsterdam
Faculty of Sciences
Business Analytics
De Boelelaan 1081a
1081 HV Amsterdam

June 30, 2017

Preface

This thesis is composed as a part of the Business Analytics Master program at the Vrije Universiteit of Amsterdam. It contains the detailed analysis and findings of my internship task at PVH Corp, a company which operates a diversified portfolio of iconic lifestyle apparel brands led by Calvin Klein and Tommy Hilfiger.

The goal of this internship task was to apply the knowledge and latest methods learned from the theoretical work of Business Analytics program on a real business environment. It was an interesting task and I would like to thank Nanne Sluis, my manager at PVH, for introducing me to the company's work and for his guidance to complete this task successfully. Furthermore, I am much grateful to Alexey Chaplygin, a senior data scientist at PVH, for his directions and explaining me the background knowledge of the task. He provided a lot of insights and together with his useful aid on technical matters, I am able to get good results. I would also like to thank Prof.dr. Sandjai Bhulai, my supervisor from the Vrije Universiteit Amsterdam, for thinking along with my research and giving useful recommendations which helped me to select appropriate models for this research. Finally, I would like to thank Dr. Bram Gorissen for being my second reader.

Abstract

E-commerce business is booming everyday as internet has made the world a smaller place and facilitated long distance communication by making the process cheaper, faster, and easier. Fashion business houses have maintained websites to promote and sell their products. It is very common in online fashion business to expect the order returns as customers do not get chance to try them on before receiving those orders. The order returns can be irritating for online sellers as it costs a lot of time and money to process them. This research is completed to help PVH, a company that owns and operates the iconic lifestyle apparel brands led by Calvin Klein (CK) and Tommy Hilfiger (TH), in predicting e-commerce orders returns and cancellations. The machine learning models and techniques are applied to predict the TH and CK orders returns and cancellations with more than 95% accuracy. This study helped PVH to report rolling net sales accurately and plan marketing budget with high accuracy to push more campaign accordingly.

Contents

References

[1]. "Machine Learning" Tom Mitchell. McGraw-Hill, 1997

[2]. P. Langley and H. A. Simon. Applications of machine learning and rule induction. Communications of the ACM, 38(11):55–64, 1995

[3]. Schafer, J. Ben, Joseph A. Konstan, and John Riedl. "E-commerce recommendation applications." Applications of Data Mining to Electronic Commerce. Springer US, 2001. 115-153.

[4]. Chen, Le, Alan Mislove, and Christo Wilson. "An empirical analysis of algorithmic pricing on Amazon marketplace." Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2016.

[5]. Yu, Xiaobing, et al. "An extended support vector machine forecasting framework for customer churn in e-commerce." Expert Systems with Applications 38.3 (2011): 1425-1430.

[6]. Sun, Zhan-Li, et al. "Sales forecasting using extreme learning machine with applications in fashion retailing." Decision Support Systems 46.1 (2008): 411-419.

[7]. Efendigil, Tuğba, Semih Önüt, and Cengiz Kahraman. "A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis." Expert Systems with Applications 36.3 (2009): 6697-6707.

[8]. C.-H. Chiu, T.-M. Choi, and D. Li, "Price wall or war: the pricing strategies for retailers," IEEE Transactions on Systems, Man, and Cybernetics Part A, vol. 39, no. 2, pp. 331–343, 2009.

[9]. C.-H. Chiu and T.-M. Choi, "Optimal pricing and stocking decisions for newsvendor problem with value-at-risk consideration," IEEE Transactions on Systems, Man, and Cybernetics Part A, vol. 40, no. 5, pp. 1116–1119, 2010.

[10]. C. Frank, A. Garg, A. Raheja, and L. Sztandera, "Forecasting women's apparel sales using mathematical modeling," International Journal of Cl

[11]. Chang, P.-C., Liu, C.-H., & Lai, R. K. (2008). A fuzzy case-based reasoning model for sales forecasting in print circuit board industries. Expert Systems with Applications, 34(3), 2049–2058.

[12]. Wang, Wen-Chuan, et al. "A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series." Journal of hydrology 374.3 (2009): 294-306.

[13]. Liu, Na, et al. "Sales forecasting for fashion retailing service industry: a review." Mathematical Problems in Engineering 2013 (2013).

[14]. Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC press.

[15]. Breiman, Leo. "Random forests." Machine learning 45.1 (2001): 5-32.

[16]. Malley, James D., et al. "Probability machines: consistent probability estimation using nonparametric learning machines." Methods of Information in Medicine 51.1 (2012): 74.

[17]. Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. "Regularization paths for generalized linear models via coordinate descent." Journal of statistical software 33.1 (2010): 1.