Chapter 14
Understanding Customers' Behaviour through Web Data Mining Models

Valerio Veglio
University of Southampton, UK

ABSTRACT

Companies have realized that the customer knowledge contained in web marketing database represent one of the main key to forecast business performance in today's competitive landscape. Appropriate web data mining models are one the best supporting approach to make different marketing decision. Analysing and understanding in advance customers' behaviour can represent the main corporation's strength in planning marketing forecasting. This research want to demonstrate as predictive web data mining models are accurate patterns in predicting marketing performance compared to traditional statistical methods in global business. In addition, particular attention is paid on the identification of the main marketing drivers performed by potential customers before purchasing a given service online. Finally, the criteria based on the loss functions confirm the high predictive power of the web data mining models in detecting the probability of customer conversion.

INTRODUCTION

The main challenge for companies is to identify the best models able to forecast marketing performance, especially in today's competitive landscape. In fact, a fundamental part of managing organizations is planning future market trends. Nowadays, the long-run success of corporations is closely related to how well management is able to anticipate future market tendencies and draw up competitive marketing and business strategies. According to Sato (2000) data mining analysis differs from the statistical data analysis. In fact, statisticians study the population parameters drawing observation sample by estimation, testing and predictions with the main risk to forecast wrong future market trends. For this reason, the development of predictive data mining models is becoming one of the main priorities for managers in every industry. Data mining analysis has become

an astonishing approach for marketers given that the meaningful knowledge is often hidden in huge customer database and most traditional statistical methods could fail to uncover such knowledge. Bueren et al. (2004) define customer knowledge as a strategic intangible asset useful to obtain a competitive advantage in the markets. An efficient utilization of the latter determines the competitive development of corporations. This is particularly true in marketing field due to the dramatic proliferation of e-customer data available online. Indeed, companies have acknowledged that their marketing strategies should focus on identifying those customers who are likely to churn in order to minimize the churn risk probability and maximize the probability of customer conversion in the short, medium and long run (Hadden, Tiwari, Roi & Ruta 2005). Churn management is defined as a set of techniques that enable corporations to keep their profitable customers in order to increase customer loyalty (Lejeune, 2001).

Prediction of behaviour, customer value, customer satisfaction and customer loyalty are example of some of the information that can be extracted from the data that should already be stored within a customer's database. In fact, one of the main tools able to help marketers in the mentioned approach is data mining due to the large amount of data available in this database (Khak & Glolamain, 2010). Finally, the key goal for businesses, regardless their size, is to discover strategic knowledge within large database in order to minimize the probability of customer's risk of churn.

The purpose of this research is to demonstrate the strategic importance of web data mining models in estimating the probability of customer's risk churn in competitive landscape. Special attention is paid on General Linear Models (GLMs) such as Logistic Regression Models based on 'Enter Methods'. The Criteria based on the Loss Functions such as the Confusion Matrix and the Receiver Operating Characteristic (ROC) curve have been used to evaluate the goodness of fit of the predictive models developed in this work.

This chapter is further structured as follows. Section 1 provides a theoretical framework of the main difference between statistical data analysis and data mining analysis due to the big lack in the literature. Section 2 describes the research methodology implemented in this research as well as the data collection process. Section 4 shows the main empirical findings with particular attention of the criteria based on the loss functions. Section 5 highlights conclusions. Finally, Section 6 shows the main future research directions.

THEORETICAL FRAMEWORK

Sato (2000) defines the main features that differentiate both statistical and data mining analysis. First, data mining analysis is governed by the need to uncover, in a timely manner, emerging trends, whereas statistical data analysis is related to historical fact and it is based on observed data. Second, statistical data analysis focused on finding and explaining the major source of variation in the data instead data mining analysis endeavours to discover, not the obvious source of variation, but rather the meaningful, although currently overlooked information. Third, statistical data analysis manages data related to a specific research questions while data mining analysis explores data collected for different purposes other than the aim of the research. In addition, Giudici (2010) observes that data mining is not just about analysing data; it is a much more complex process where data analysis is just one the aspect. Instead, Turban, Aronson, Liang and Sharda (2008, p. 305) define data mining as "the process that uses statistical, mathematical, artificial intelligence and machine-learning techniques to extract and identify useful information and subsequently gain knowledge in databases". "To apply a data mining methodology means following and integrated

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/understanding-customers-behaviour-through-web-data-mining-models/130571

Related Content

Analysing WhatsApp and Instagram as Blended Learning Tools
Alberto Andujar (2023). *Research Anthology on Applying Social Networking Strategies to Classrooms and Libraries (pp. 649-663).*
www.irma-international.org/chapter/analysing-whatsapp-and-instagram-as-blended-learning-tools/312946

A Survey of Research in Real-Money Trading (RMT) in Virtual World
Mohamed Nazirand Carrie Siu Man Lui (2017). *International Journal of Virtual Communities and Social Networking (pp. 34-53).*
www.irma-international.org/article/a-survey-of-research-in-real-money-trading-rmt-in-virtual-world/180674

Social Networks as a Tool for E-Collaborative Learning and its Effects on Knowledge Acquisition and Satisfaction Among Al-Aqsa University Students
Hasan R. Mahdiand Mohanad A. Al-Taban (2017). *International Journal of Virtual Communities and Social Networking (pp. 24-45).*
www.irma-international.org/article/social-networks-as-a-tool-for-e-collaborative-learning-and-its-effects-on-knowledge-acquisition-and-satisfaction-among-al-aqsa-university-students/182730

Social Networking Engagement and Crisis Communication Considerations
Mitchell Marovitz (2017). *Social Media Performance Evaluation and Success Measurements (pp. 130-155).*
www.irma-international.org/chapter/social-networking-engagement-and-crisis-communication-considerations/174023

Lessons from the Virtual Campus: The Life and Death of a Second Life University
Kathryn E. Stevensand Scott P. Stevens (2012). *International Journal of Virtual Communities and Social Networking (pp. 1-18).*
www.irma-international.org/article/lessons-virtual-campus/75776