Chapter VII
The Business Impact of Predictive Analytics

Tilmann Bruckhaus
Numetrics Management Systems, USA

ABSTRACT

This chapter examines the business impact of predictive analytics. It argues that in order to understand the potential business impact of a predictive model, an organization must first evalute the model with technical metrics, and then interpret these technical metrics in terms of their financial business impact. This chapter first reviews a set of technical metrics which can assist in analyzing model quality. The remaining portion of the chapter then shows how to combine these technical metrics with financial data to study the economic impact of the model. This know-how is used to illustrate how a business can choose the best predictive model from among two or more candidate models. The analysis techniques presented are illustrated by various sample models from the domains of insurance fraud prevention and predictive marketing.

INTRODUCTION

Metrics of financial business impact such as ROI and net profit are the key to success with predictive analytics in real world applications. Yet, current practices do not focus on such metrics and instead employ what may be termed "technical metrics." As we will see, technical metrics tell an organization little about whether a predictive model will benefit the organization. This chapter will review these topics, explain the difference between financial business impact metrics and technical metrics, and show how financial business impact metrics can be calculated with little effort. Throughout the chapter, we will illustrate the use of technical and financial business impact metrics with running examples in the areas of insurance fraud prevention and predictive marketing.

Copyright © 2007, IGI Global, distributing in print or electronic forms without written permission of IGI Global is prohibited.

The academic community has created an awesome arsenal of tremendously useful machine learning algorithms. However, when working with predictive modeling technology, the needs in academia and the commercial sector are different. The academic community strives to bring about and demonstrate algorithmic improvements, whereas the business community must generate financial profits. Algorithms are often designed by researchers to work across a wide variety of application areas. This makes it difficult to analyze the financial business impact because financial impacts vary greatly from one business context to the next. Because of this difficulty, and in order to allow for objective analysis of algorithms, a variety of benchmark datasets have been established. New algorithms are tested against these benchmarks, and researchers have developed metrics for analyzing model quality which can be applied in the absence of financial considerations. When data mining algorithms are then transferred into the business community the same technical metrics are transferred along with the algorithms. Practitioners in the business world are then able to evaluate predictive models developed with the available technical metrics. The dilemma is that, as we will see, these technical metrics do not answer the one question business practitioners need to answer: will this model have a beneficial impact on my business?

There are many types and versions of technical metrics of the quality of predictive models. Some of these are *accuracy*, *precision*, *top k precision*, *false alarms*, *recall*, *sensitivity*, *missed alarms*, *specificity*, *selectivity*, and we will review each of these in turn. As we introduce some of these technical metrics we will describe how each metric can be useful in a business context. Another group of technical metrics of model quality are based on graphical analysis of model performance, such as lift chart, gain chart, and even measurements of the size of the area under a geometric curves known as receiver operating characteristic (ROC). These chart-based metrics are out of the scope of this chapter and the reader is referred to other texts on data mining for more information on these. Some good texts on data mining technology are Berry and Linoff (1997), Han and Kamber (2005), Mitchell (1997), Quinlan (1993), Soukup and Davidson (2002), and Witten and Frank (2005).

Table 1. Overview of sample models discussed in this chapter; the calculation and meaning of the data presented will be discussed in the following sections

	Outcome Distribution and Information Entropy	Accuracy	Precision	Recall	Model Strategy
M_1^{Fraud}	Skewed H(X)=0.021	98.0%	40.0%	66.7%	Maximize business impact
M_2^{Fraud}	Skewed H(X)=0.021	98.5%	undefined	0.0%	Always predict negative
M_3^{Fraud}	Balanced H(X)=1.0	50.0%	undefined	0.0%	Always predict negative
M_4^{Fraud}	Skewed H(X)=0.021	50.0%	1.5%	50.0	Predict randomly
M_5^{Fraud}	Skewed H(X)=0.021	98.5%	100.0%	.7%	Maximize precision

23 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/business-impact-predictive-analytics/24904

Related Content

Reducing a Class of Machine Learning Algorithms to Logical Commonsense Reasoning Operations
Xenia Naidenova (2008). *Mathematical Methods for Knowledge Discovery and Data Mining (pp. 41-64).*
www.irma-international.org/chapter/reducing-class-machine-learning-algorithms/26132

Sentimental Analysis in Various Business Applications
Harshita Pateland B. Manjula Josephine (2019). *Sentiment Analysis and Knowledge Discovery in Contemporary Business (pp. 31-43).*
www.irma-international.org/chapter/sentimental-analysis-in-various-business-applications/210961

Aggregation and Maintenance of Multilingual Linked Data
Ernesto William De Luca (2012). *Semi-Automatic Ontology Development: Processes and Resources (pp. 201-225).*
www.irma-international.org/chapter/aggregation-maintenance-multilingual-linked-data/63903

A Study and Comparison of Sentiment Analysis Techniques Using Demonetization: Case Study
Krishna Kumar Mohbey, Brijesh Bakariyaand Vishakha Kalal (2019). *Sentiment Analysis and Knowledge Discovery in Contemporary Business (pp. 1-14).*
www.irma-international.org/chapter/a-study-and-comparison-of-sentiment-analysis-techniques-using-demonetization/210959

Dynamic Workload for Schema Evolution in Data Warehouses: A Performance Issue
Fadila Bentayeb, Cécile Favreand Omar Boussaid (2010). *Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development: Innovative Methods and Applications (pp. 28-46).*
www.irma-international.org/chapter/dynamic-workload-schema-evolution-data/39586