LOUIS DUCLOS-GOSSELIN

HOW MANAGERS CAN USE PREDICTIVE ANALYSIS AND MATHEMATICAL MODELS AS DECISION MAKING TOOLS

Mémoire présenté
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de maîtrise en Marketing analytique
pour l'obtention du grade de maître ès sciences (M. Sc.)

DÉPARTEMENT DE MARKETING
FACULTÉ DES SCIENCES DE L'ADMINISTRATION
UNIVERSITÉ LAVAL
QUÉBEC

2011

Assessing Selling Units' Market Performance over Time

Abstract

This paper proposes a simple measure of the market performance of various selling units (e.g., sales territories, sales regional offices, or the whole sales organization). The measure is easy to estimate and to be understood by management. Along with other indicators, it can be used as a diagnosis tool by comparing the market performances of various units, taking into account the conditions prevailing in the different markets (such as competition relative effectiveness, sales penetration, or local market fluctuations). Combining sales volume, market share, and profit variations data into an Index of Sales Unit Market Performance (ISUMP), can provide managerial guidance for selling units' evaluation or resource allocations among units. This index can also account for a firm's selected market strategy (market penetration, market skimming, etc.) An implementation of this method in a large North American insurance company is reported.

Key words: Selling units, market performance, sales performance, market shares, market strategies, selling units' control.

Résumé (French)

Cet article propose une mesure simple (un modèle mathématique) évaluant la performance de différentes entités de vente (comme les vendeurs, les territoires de vente, les bureaux de vente régionaux ou l'ensemble des ventes de l'organisation) est proposée. Cette mesure est facile à estimer et elle peut facilement être comprise par les gestionnaires. De fait, elle peut être utilisé pour comparer les performances des différentes entités de vente, en tenant compte des conditions prévalant dans les différents marchés (tels que l'efficacité de la concurrence, la pénétration des ventes, ou les fluctuations du marché local). Les résultats de la mise en œuvre de cette mesure dans une grande entreprise d'assurance dommage nord-américaine sont présentés.

Mots clefs: Évaluation des ventes, rendement des ventes, parts de marché, stratégies de marché, contrôle de la force de vente

How Health Sciences Practitioners Can Use Data Mining for Predicting Individuals' Risks of Contracting Nosocomial Pneumonia

Abstract

From the beginning of health management sciences, managers have tried to find better techniques to classify and qualify risks in order to effectively manage health operations, including reducing costs while increasing patient wellbeing. This paper explains how managers can use data mining techniques for solving problems related to individual risks of contracting nosocomial pneumonia (within 48-72 hours after hospital admission).

Keywords: Nosocomial Pneumonia, Nosocomial Infection, Health Management, Health Science, Health Informatics, Decision Support System, Predictive analysis, Data Mining.

Résumé (French)

Depuis le début de sciences de la gestion de la santé, les gestionnaires ont toujours tenté de trouver de meilleures techniques pour classer et qualifier les risques afin de gérer efficacement les activités de leurs établissements, et ce, afin de réduire les coûts, tout en augmentant le bien-être des patients. Cet article présente une solution analytique qui prédit les risques qu'un patient contracte une pneumonie nosocomiale. Cette solution a été testée sur une base de données d'un grand hôpital américain.

Mots clefs: Pneumonie nosocomiale, infections nosocomiales, gestion de la santé, sciences de la santé, l'informatique de santé, aide à la décision, analyse prédictive, Data Mining.

Avant-Propos / Introduction (English)

Starting from earlier management science studies, managers have always tried to find better techniques to effectively manage operations in order to reduce the costs while increasing customers' / employees' satisfaction, organization's performance (in this case) and patients' wellbeing (in this case).

This MSc research shows how managers can use mathematical models and predictive analysis to address such problems.

First, a simple measure (mathematical model) of the market performance of various selling units (e.g., sales people, sales territories, sales regional offices, or the whole sales organization) is proposed. An implementation of this method in a large North American insurance company is reported.

Second, a new predictive analysis technique for solving problems related to the risk of an individual contracting nosocomial pneumonia is presented. The technique has been tested on a large health database.

The two proposed articles have been co-signed with Benny Rigaux-Bricmont and René Yves Darmon. Benny Rigaux-Bricmont is professor of Marketing at Laval University and director of this thesis. René Yves Darmon is an affiliate professor of Marketing at Laval University, professor of Marketing at ESSEC Business School and co-director of this thesis.

Louis Duclos-Gosselin is the principal author of the second article.

In the first article (Assessing Selling Units' Market Performance over Time), Louis collaborated with Benny Rigaux-Bricmont and René Yves Darmon to construct the proposed measure. He also worked during one year to implement the reported measure in a large North American insurance company. A different version of this article has been submitted to the EUROPEAN JOURNAL OF OPERATIONAL RESEARCH.

In the second article (How Health Sciences Practitioners Can Use Data Mining for Predicting Individuals' Risks of Contracting Nosocomial Pneumonia), Louis developed the proposed predictive analysis solution and tested it on a large health database. In addition, Louis developed the C++ code behind the predictive analysis solution used. Benny Rigaux-Bricmont and René Yves Darmon wrote the article with Louis.

Both papers have been accepted and presented in full or partial form at various international management science conferences: INFORMS Annual Meeting, Washington, 2008; INFORMS International Meeting, Toronto, 2009; INFORMS Annual Meeting, San Diego, 2009.

Avant-Propos / Introduction (French)

Depuis toujours, les gestionnaires s'affairent à trouver de meilleures moyens pour gérer efficacement les opérations, afin de réduire les coûts, tout en augmentant la satisfaction des clients / employés satisfaction.

Cette recherche de maîtrise s'inscrit dans ce courant en illustrant comment les gestionnaires peuvent utiliser des modèles mathématiques et des solutions analytiques pour résoudre certains de leur problèmes de gestion.

Tout d'abord, une mesure simple (un modèle mathématique) évaluant la performance de différentes entités de vente (comme les vendeurs, les territoires de vente, les bureaux de vente régionaux ou l'ensemble des ventes de l'organisation) est proposée. Les résultats de la mise en œuvre de cette mesure dans une grande entreprise d'assurance dommage nord-américaine sont présentés.

Deuxièmement, une solution analytique prédisant les risques qu'un patient contracte une pneumonie nosocomiale est illustrée. Cette solution a été testée sur une base de données d'un grand hôpital américain.

Version history

Version	Date	Description	From
1,0	2009-11-03	Preliminary version	Louis Duclos-Gosselin
2,0	2011-03-04	Final version	Louis Duclos-Gosselin

The author wants to acknowledge Bernard Lamond and Alain Martel for theirs useful comments on the Preliminary version.

Warning

This version is under review.

Author(s)

Name	Organization	Fonction
Louis Duclos-Gosselin	Laval University	Student

Collaborators

The author wants to acknowledge these people for their contribution to this work.

Name	Organization	Fonction
Benny Rigaux-Bricmont	Laval University	Professor of Marketing
René Yves Darmon	Laval University	Professor of Marketing

Table of contents

Table of figures

List of tables

Assessing Selling Units' Market Performance over Time

Abstract

This paper proposes a simple measure of the market performance of various selling units (e.g., sales territories, sales regional offices, or the whole sales organization). The measure is easy to estimate and to be understood by management. Along with other indicators, it can be used as a diagnosis tool by comparing the market performances of various units, taking into account the conditions prevailing in the different markets (such as competition relative effectiveness, sales penetration, or local market fluctuations). Combining sales volume, market share, and profit variations data into an Index of Sales Unit Market Performance (ISUMP), can provide managerial guidance for selling units' evaluation or resource allocations among units. This index can also account for a firm's selected market strategy (market penetration, market skimming, etc.) An implementation of this method in a large North American insurance company is reported.

Key words: Selling units, market performance, sales performance, market shares, market strategies, selling units' control.

Résumé (French)

Cet article propose une mesure simple (un modèle mathématique) évaluant la performance de différentes entités de vente (comme les vendeurs, les territoires de vente, les bureaux de vente régionaux ou l'ensemble des ventes de l'organisation) est proposée. Cette mesure est facile à estimer et elle peut facilement être comprise par les gestionnaires. De fait, elle peut être utilisé pour comparer les performances des différentes entités de vente, en tenant compte des conditions prévalant dans les différents marchés (tels que l'efficacité de la concurrence, la pénétration des ventes, ou les fluctuations du marché local). Les résultats de la mise en œuvre de cette mesure dans une grande entreprise d'assurance dommage nord-américaine sont présentés.

Mots clefs: Évaluation des ventes, rendement des ventes, parts de marché, stratégies de marché, contrôle de la force de vente

Article

Introduction

Many managerial marketing decisions (e.g., budget and resource allocations) and sales force decisions aim at maintaining or improving the performance of various *selling units* over time (Brown and Peterson 1993; Churchill *et al.* 1985). A selling unit is defined as any entity responsible for some output performance (such as sales, profits, market share). It may be, for instance, a sales territory (either assigned to a sales team or to an individual salesperson), a regional or district sales office, or the whole sales organization. Every selling unit is characterized by its *market performance*, defined here as the market outcomes resulting from the accomplishment of the selling tasks, in its specific environment, over some long- and/or short-term periods of time.

In order to properly enhance selling units' market performance, managers must rely on valid and accurate measures of the concept. Inaccurate assessments may lead to poor managerial decisions, feelings of inequity, morale problems, lack of organizational commitment, and dysfunctional turnover (Barksdale *et al.* 2003; Jaramillo *et al.* 2006; McKay *et al.* 1991; Morris *et al.* 1991). For that purpose, managers generally collect and analyze large amounts of data, often frequently supplied by CRM applications. However, it is often difficult to devise adequate measures of market performance: this concept covers, among others, short- and long-term variations in sales, profits, customer satisfaction, loyalty building, and customer relationship development; it is multidimensional (Avila *et al.* 1988; Chonko *et al.* 2000) and consequently, no single measure can account for its various dimensions. This justifies the large number of indices used for that purpose (Brown and Peterson 1993; Churchill *et al.* 1985).

In spite of this large amount of data, managers still lack simple and meaningful ways of tracking selling units' market performance over time, and to assess up to what point every selling unit has taken advantage of market opportunities (or reacted to shrinking opportunities) during a considered period of time.

The aim of this paper is to supplement the sales managers' tool kit by proposing a simple measure of a selling unit's market performance over time. When considered jointly with extra-role performance measures (MacKenzie *et al.* 1998), this market performance measure can allow managers to make better situation diagnoses, and consequently, to provide better managerial guidance for selling units' evaluation, organization, compensation, and/or resource allocations (Silver *et al.* 2006).

After discussing the concept of selling units' market performance measurement, this paper outlines the proposed procedure. An application illustrates the concept. Finally, the advantages and limits of the proposed procedure are discussed.

Selling Units' Market Performance Measures

Assessing market performance is fraught with difficulty, and sales force researchers and sales managers have typically used different approaches to that purpose (see for instance Berry 1986; Boles, Donthu, and Lothia 1995; Cron and Levy 1987; Dalrymple and Strahle 1990; Dubinsky *et al.* 1989; Edwards *et al.* 1984; Gentry *et al.* 1991; Jobber *et al.* 1993; McFarland and Kidwell 2006).

Sales Force Researchers' Approaches

Over the last decades, a major thrust of sales force research has been to gain better understanding of salespeople's performance and to find appropriate ways to enhance it (Muczyk and Gable 1987). The seminal work by Walker *et al.* (1977) that proposed a model of the major determinants and consequences of sales force performance has been followed by many studies sharing similar concerns (see for instance DeCarlo *et al.* 1997; Krishnan *et al.* 2002; Plank and Greene 1996; Plank and Reid 1994; Weitz 1981; Weitz *et al* 1986). These studies have investigated the antecedent and/or consequences of salespeople's sales performance and/or effectiveness (Avlonitis and Panagopoulos 2006; Chonko *et al.* 1986; Flaherty *et al.* 1999).

In most cases, researchers have measured selling units' performance either using objective sales data (see for instance MacKenzie *et al.* 1998) or evaluative scales (see for instance Jaramillo *et al.* 2006). Measurement instruments have been completed either by managers or by the salespersons themselves (self-reports) (see for instance Behrman and Perreault 1982; 1984; Hunter and Perreault 2006; Silver *et al.* 2006) or by both (Giacobbe *et al.* 2006; Jaramillo *et al.* 2006; Patton and King 1985). It seems that, in practice, research suggests that subjective and objective measures lead to different results (Bommer *et al.* 1995; Heneman 1886; Marshall *et al.* 1992; Marshall and Mowen 1993; Mowen *et al.* 1980; Rich *et al.* 1999). Unfortunately, because they are highly subjective and may well yield biased judgments (Jaramillo *et al.* 2004; Sharma *et al.* 2004), their validity is often questionable (Pettijohn *et al.* 2000; Rich *et al.* 1999), and they can have only limited practical value. Consequently, such self-report evaluative scales are seldom used alone in the practice of sales management.

Sales Managers' Approaches

In order to assess selling unit market performance, managers have typically used quantitative and/or qualitative measures (Jackson *et al.* 1983). Here again, the former measures (such as sales volume) tend to be objective and easy to assess. Unfortunately, they seldom capture the qualitative aspects of the selling job, especially the extra-role performance aspects (MacKenzie *et al.* 1998). In addition, they are inadequate for assessing the unit's performance over time, taking into account the effects of environmental variables on market size. A selling unit's sales performance results from many other factors than the marketing resources deployed in this unit (Pilling *et al.* 1999), for instance, each unit's environmental conditions (competitive strength, economic conditions, etc.).

Alternatively, in order to explain management controls of the sales force, some authors have made a distinction between behavior- and outcome-based sales force controls (Challagalla and Shervani 1996; Cravens *et al.* 1993; Oliver and Anderson 1994). Although controlling salespeople's behaviors is an important and necessary aspect of sales management (Bartkus et al 1989; Cravens *et al.* 1993), it would be worthless if the proper behaviors did not eventually translate into sales, profits, market shares, strengthened customer relationships, or clients' satisfaction. In practical situations, managers do evaluate selling unit market performance from at least some objective quantitative measures of sales performance (Hyman and Sager 1999), and outcome performance is always part of a sales force control process (Jaworski 1998; Oliver and Anderson 1995; Ouchi and McGuire 1975). Although outcome performance measures are always used by management, few, if any, can provide a fair assessment of salespeople's market performance.

Territory sales volume, market shares, or *profits*, which are typically watched by managers, are inadequate market performance measures when considered individually. As stated above, although selling units are instrumental in building sales volume over time, sales are also influenced by a host of factors which are beyond the unit's control. In most cases, managers cannot easily disentangle what parts of such outcomes must be attributed to a selling unit's marketing program (attractiveness of the offers, advertising, company's reputation, etc.) or to environmental factors (competitive strength, economic trends, etc.). As a result, sales volume alone may give a distorted picture of this unit's market performance (Hunter and Perreault 2006; Mowen *et al.* 1986). Although territory market shares bring the additional dimensions of industry sales, competition's performance, and the firm's market penetration, they suffer from the same problems as sales volume. In addition, all outcome measures of performance tend to over-emphasize short- over long-term performances. Finally, selling unit profit measures are useful to watch, in as much as the units are given responsibility to negotiate prices and/or are left to allocate their efforts and resources among several product lines with different profit margins (Farley 1964). Being linked to sales, profit measures have the same drawback as sales as market performance measures.

Sales variations over some benchmark period (after removing seasonal effects) are frequently used and provide better measures because they indicate a unit's success (or failure) in sustaining a given level of sales over time. Taken alone, however, sales variations can be misleading because they may also be caused by many factors beyond a selling unit's control (for instance, a shift in territory environmental conditions, windfall sales in any of the two periods, or simply random factors affecting sales in one or both periods of time).

Customer satisfaction could be a more appealing measure, because it involves an important long-term firm's objective. Unfortunately, customer satisfaction variations are not always easy to measure. In addition, research suggests that salespersons' confusion can result from control systems that combine both customer satisfaction and short-term performance measures such as sales (Sharma and Sarel 1995). Not surprisingly, a survey has shown that only a relatively small proportion of firms (about 25 percent) use customer satisfaction as a basis for selling unit performance evaluation and rewards (Cohen 1997).

Market share variations in a selling unit's territory provide a clear indication of a firm's standing in its market. It measures a company's market penetration, and, to some extent, the selling unit is responsible for it. Market share brings the additional dimension of industry sales, and consequently, of competitive performance. Competitive sales performance constitutes a natural benchmark for evaluating selling units' market performances. In other words, it is essential to account for the evolution of the competition's sales levels for evaluating selling unit market performance.

When *territory sales, industry sales, market shares, and profit variations* are considered jointly, they provide a more complete assessment of a selling unit market performance. Various market conditions require different efforts and abilities (Giacobbe *et al.* 2006). Increasing or even maintaining sales volume and/or profits in a declining market are indicative of higher market performance than the same achievement in a fast expanding market. A few methods, such as regression analysis and data envelopment analysis (DEA) (Boles, Donthu, and Lohtia 1995; Charnes, Cooper and Rhodes 1978) have been proposed and sometimes used in practice for addressing the problem of evaluating selling units' market performance with multiple criteria (Parsons 1992). These techniques rely on linear aggregations of various input and output criteria and compare every selling unit to the "best" performing one. This technique, however, has a number of limitations, the most serious ones being its complexity and difficulty to make it understood by practitioners, its reliance on many subjectively selected criteria, and its sensitivity to measurement errors (Boles, Donthu, and Lothia 1995).

Another important dimension of selling unit market performance is the time frame over which it is measured. While building long-term relationships with customers has become a prevalent strategy, short-term sales have often become inadequate measures of market performance. The current emphasis on sales activity-based at the expense of outcome-based controls is a logical consequence of this evolution. Consequently, there is a need for devising selling unit's market performance measures that (1) are meaningful to managers, (2) are based on easily observable and quantifiable data, (3) are valid and reflect the ability of selling units to progress, taking into account its market evolution, and (4) can accommodate short- as well as long-run market performance assessments.

Like DEA, the selling unit market performance measure proposed in this paper is also a multi-criterion procedure. It presents, however, several advantages over DEA: (1) it measures "true" market performance, discarding environmental factors that influence the sales results of every selling unit; (2) unlike DEA which is an extreme point method, it is not sensitive to some unusually high or low unit performances, and consequently to outlying observations; (3) it requires much less computations (against one linear program per selling unit in DEA); and (4) it is more easily understood and applicable by managers.

Proposed Method

Principle

Managers assess a selling unit's market performance at time T_1 by comparing the performance this unit has achieved over some period of time t (typically a few weeks or months, ending at time T_1) to some selected *benchmark*. Depending on the situation and management's goals, the selected benchmark can be (1) the outcomes achieved during a comparable reference period of length t (such as the preceding period T_0 or some other selected period, or an average of several selected periods) or (2) some target outcome levels to be achieved during the considered period of time (for instance, some sales objectives or quotas). Note that this analysis can be carried with several types of appropriate benchmarks.

The choice of the benchmark is crucial and requires careful managerial consideration. Selecting a preceding similar period is appropriate whenever, during this period of time, external factors have not *differentially* and *significantly* affected the various selling units. For instance, it would not be wise to use as a reference period some quarter in which a storm has affected only one part of the selling units. In such cases, a more appropriate benchmark could be outcome objectives set by management and taking into account the specific market situations of the different selling units (in this example, sales/profit quotas objectives taking into account the effects of the storm on the concerned selling units).

At some time T_1, four main factors reflect different aspects of a selling unit's current market performance variation over the benchmark: (1) the total industry sales (IS) variation in this unit: $is_1 = (IS_1 - IS_0)/IS_0$; although a selling unit cannot be held responsible for industry sales variations, one should recognize that it is more difficult to improve a firm's market position in a declining than in an expanding market; (2) the unit's sales volume (S) variation $s_1 = (S_1 - S_0)/S_0$ over the benchmark situation; and (3) the corresponding selling unit's market share (MS) variation $ms_1 = (MS_1 - MS_0) / MS_0 = (S_1/IS_1 - S_0/IS_0) / S_0/IS_0$, because performance should account for a selling unit's ability to improve or maintain the firm's market position in its territory; and (4) the corresponding gross profit variations $\pi_1 = (\Pi_1 - \Pi_0)/ \Pi_0$. Being measures of variation (in decimal form) is_1, s_1, and ms_1 can be either positive or negative (or null) and can vary between -1 and $+\infty$. π_1 is also assumed to vary between -1 and $+\infty$.[1]

In order to provide a simpler explanation of the underlying concepts and rationale of the proposed method, a simpler case that does not consider profit variations will be discussed first. Then, in a following section, the method will be generalized to the case that includes profit variations.

Selling Units' Market Performance (Excluding Gross Profits)

Table 1 provides an eight-cell cross-tabulation of selling units' market performance situations according to the variation directions of the territory's industry sales ($is_1 > 0$ or < 0), sales volumes ($s_1 > 0$ or < 0), and market share ($ms_1 > 0$ or < 0). (From here on, for simplification purpose, the subscript 1 will be omitted, unless necessary.)

[Insert Table 1]

These three dimensions are somewhat interrelated. Increased sales in a decreasing industry sales territory imply an increased market share (Case 1). Decreased sales in an increasing industry sales market imply a decreased market share (Case 6). No inference about market share variations can be made, however, on the basis of sales and industry sales variations alone when both sales and market industry sales increase (decrease). In such cases, market share increases or decreases, depending on the relative sizes of the sales and industry sales increase (decrease) rates. Therefore, one should explicitly take market share variations into consideration. The six possible cases defined in Table 1 provide clear indications about how a selling unit market position has evolved in its territory over time, taking into account factors over which this unit has no control (such as market size variation, competitive strength, or other factors).

The six situations described above are represented graphically in Figure 1 (Zone 1 to 6). This diagram relates a selling unit's territory sales (s) and market share (ms) variations over the benchmark. For instance, a selling unit A may have experienced a sales increase (s_a) and a market share increase (ms_a) and consequently be positioned at point A in Figure 1. The *slope* of the vector NA reflects the corresponding industry sales variation (is_a) during that period of time (see Analytical Formulation section). Note that every selling unit's vector must start at point N (-1, -1) (a 100% sales decrease always implies a 100% market share decrease). Because industry sales variations are generally beyond their control, selling units cannot choose the slope of the vector along which they can move. However, how far they move on their respective vectors in the direction of the arrow does reflect the quantity and quality of the work accomplished by the firm in this territory. The *length* of the vector NA reflects Selling Unit A's market performance. In other words, the unit's market performance (the *length* of Vector NA) has been separated from uncontrollable market size variations (the *slope* of Vector NA).

[Insert Figure 1]

Although theoretically, s, is, and ms can possibly vary between -1 and infinity, in most usual cases, their values are likely to be relatively close to zero (NO = *status quo*, i.e., no change over the benchmark: s = is = ms = 0). The six zones shown in Figure 1 correspond to the six cases in Table 1. Because the length of the vector is characteristic of a given selling unit market performance, iso-market performance curves are quarter of circles centered on N (here, other choice are possible). A larger radius of the circle reflects a higher market performance level (a higher progression).

In order to compare the market performance of various selling units, one can compare the percentages by which each one has moved on its vector compared with the *status quo* N0. A formal measure called ISUMP (Index of Selling Unit Market Performance) can be used to that effect.

Analytical Formulation

LET SUCCESSIVELY BE SALES VARIATIONS (S), INDUSTRY SALES VARIATIONS (IS) AND MARKET SHARE VARIATIONS (MS) IN SELLING UNIT I'S TERRITORY (ALL IN DECIMAL FORM) (SUBSCRIPT I OMITTED UNLESS NECESSARY):

(1) $s = (S_1 - S_0) / S_0, \quad s \geq -1$

(2) $is = (IS_1 - IS_0) / IS_0, \quad is \geq -1$

(3) $ms = (MS_1 - MS_0) / MS_0, \quad ms \geq -1$, with $MS_0 = S_0 / IS_0$, and $MS_1 = S_1 / IS_1$

where:

S_1 = sales level achieved by Selling unit i in period 1, in dollars

S_0 = sales level for this unit in the benchmark situation, in dollars

IS_1 = industry sales in period 1, in dollars

IS_0 = industry sales in the benchmark situation, in dollars

MS_1 = market share achieved in this unit's territory in period 1, in decimal form

MS_0 = market share for this selling unit's territory in the benchmark situation, in decimal form.

Replacing S_1 and IS_1 by their values in (1) and (2) in Equation (3) leads to:

(4) $$ms = [1 / (is + 1)]\, s - [is / (is + 1)]$$

Given specific values of is ($-1 \leq is < +\infty$), one can determine the linear relationship linking market share ms to sales volume s variations, as shown in equation (4). Thus, for is = -1, ms => $+\infty$; for is = - ½, ms = 2s + 1; for is = 0, ms= s, and for is = 1, ms = s/2 – ½. In the same way, because the three concepts are interrelated, each one can be expressed as a function of the other two:

(5) $$s = ms\,(is + 1) + is \quad \text{and} \quad is = (s - ms) / (ms + 1)$$

As shown in Figure 1, every selling unit can move along a vector characteristic of its territory situation, and market performance is reflected by the position an entity has reached on its vector. By simple application of the Pythagorean Theorem, the *status quo* (benchmark situation) iso-performance curve has a radius equal to the square root of 2, i.e. $(2)^{1/2}$.

In order to compare the market performance achieved by various selling units, one must assess by which percentage each unit has moved on its vector from the *status quo*. For a given unit i which has achieved sales variations of s_{i0} and market share variation of ms_{i0} in period 1, the market performance increase (or decrease) in comparison with the benchmark situation 0 (or Index of Selling Unit Market Performance, ISUMP) is given by:

(6) $$ISUMP = 100\, \{[(1 + s_{i0})^2 + (1 + ms_{i0})^2] / 2\}^{1/2}$$

By replacing s_{i0} and ms_{i0} by their values in (1-3) and rearranging the terms leads to another expression of the ISUMP:

(7) $$ISUMP = (100/IS_1)\,(S_1/S_0)\,[(IS_1^2 + IS_0^2)/ 2]^{1/2}$$

ISUMP = 100 means no market performance improvement, ISUMP > 100, some improvement, and ISUMP < 100 some market performance decrease. In order to assess how a selling unit's market performance compares with the overall higher order entity (e.g., a given sales territory within its corresponding branch office) market performance, this index can be supplemented by an adjusted ISUMP defined as[2]:

(8) Adjusted ISUMP = 100 (entity's ISUMP) / (higher order entity's ISUMP)

The analysis can be supplemented by locating the position of every selling unit on the six-zone map (tips of market performance vectors in Figure 2). For instance, in the reported application (see the following section), the market performance vectors of twenty-eight sales territories (ST1 to ST28) constituted of a 713 salesperson sales force have been cast into one single common graph and can be compared directly. For greater visibility, however, only representative cases from each Zone are shown in Figure 2.

[Insert Figure 2]

Selling Units' Market Performance (Including Gross Profits)

The same principles apply when profit margin variations are added to the analysis. In this case, there are twelve situations (and twelve corresponding zones); each case identified in Table 1 being characterized either by a profit margin (pm) increase or decrease over the benchmark. Consequently, all previous zones are split according whether they are characterized by a profit margin increase (indexed a) or a profit margin decrease (indexed b). As shown in Figure 3, the market performance vectors are cast into a three dimensional space characterized by the s, ms, and pm variations.

[Insert Figure 3]

There are two ways through which a selling unit can increase profits: increasing sales and/or the profit margins on the goods or service sold (either through negotiation of higher prices and/or selling more profitable products). The profit situation is therefore characterized by:

(9) $$\pi_1 = (\Pi_1 - \Pi_0)/ \Pi_0, \text{ with } \Pi_0 = PM_0\ S_0 \text{ and } \Pi_1 = PM_1\ S_1$$

where PM_0 and PM_1 are the profit margin rates respectively in the benchmark situation and achieved in Period 1. Defining the profit margin variation as $pm = (PM_1 - PM_0)/ PM_0$ leads to:

(10) $$\pi = pm + pm\ s + s \quad \text{or} \quad pm = (\pi - s) / (s + 1)$$

A simple extension of the previous analysis leads to an estimate of the Index of Selling Unit Market Performance (ISUMP):

$$(11) \qquad ISUMP = 100 \left\{[(1 + ms_{i0})^2 + (1 + s_{i0})^2 + (1 + pm_{i0})^2] / 3\right\}^{½}$$

or using equation (10):

$$(12) \qquad ISUMP = 100 \left\{[(1 + ms_{i0})^2 + (1 + s_{i0})^2 + (1 + \pi_{i0})^2 / (1 + s_{i0})^2] / 3\right\}^{½}$$

Assessing Selling Units' Performance at Implementing a Firm's Market Strategy

In many cases, a firm may equally value sales, market share, and profit achievement. Alternatively, when introducing a new product line, a firm may select a strategy of fast market penetration. In this case, sales and market shares may be assigned a higher weight than immediate profits. In other instances, the firm may select a market skimming strategy. In this case, profits may be given more importance relative to market share. Other strategies may be pursued by assigning different weights reflecting the relative importance of the three outcomes. Let:

α = weight assigned to market share increases
β = weight assigned to sales volume increases
γ = weight assigned to profit margin increases

with

$$(13) \qquad \alpha + \beta + \gamma = 1$$

In this case, a straight application of these weights to the corresponding dimensions lead to a new expression of the Index of Selling Unit Market Performance (ISUMP):

$$(14) \qquad ISUMP = 100 \{[\alpha^2 (1 + ms_{i0})^2 + \beta^2 (1 + s_{i0})^2 + \gamma^2 (1 + pm_{i0})^2] / (\alpha^2 + \beta^2 + \gamma^2)\}^{1/2}$$

In addition, when management wishes to assess selling units' performance at implementing different market strategies for different product lines, the analysis may be carried out for the various product lines separately. Then, management may assign different weights to the product lines reflecting their relative strategic importance. A weighted market performance index could be computed for every unit, and compared to the corresponding higher level selling entity's ISUMP.

Implementation

This procedure has been implemented in a very large North-American insurance company which sells directly to customers, with no intermediary involved, automobile and habitation insurance products. Like many North American insurance companies, this firm lacked adequate means to properly assess the market performance of the various selling units (sales agents and sales managers) at developing their territories over time. Thus, the general sales manager was highly concerned about the current market performance evaluation procedure where sales managers were essentially relying on one single criterion, i.e. the sales volume achieved in each territory. Note that his lack of formal reliance on modern CRM applications is not unusual in the North American insurance industry. As a result, several salespersons grew dissatisfied with this criterion: they argued that even though their sales grew slowly (or even decreased) their territory market share was increasing. Market share data were collected from internal services and communicated to the sales personnel each month.

In this case, the application involved all the branch offices covering the North American market. The company used 713 insurance agents under the supervision of 28 sales managers. Because of space constraints, only the results of the 28 sales offices are reported here. In other words, the branch office has been selected as the selling unit. For that purpose, the required data are the summated results across territories belonging to a selling unit. Those were calculated by cumulating the results of the insurance agents they supervise.

Cross-Sectional Market Performance Analysis

Selecting an adequate time period length to assess performance variations is an important decision. Too short a period of time (e.g., monthly data) could lead to wide market performance variations and may hide the true longer term performance of some selling units. Alternatively, too long a period of time (e.g., more than one year) may not be flexible enough, especially if management bases some financial rewards on short-term performance. This is why it may be worthwhile to carry the analysis with various time period lengths, whenever possible (this will be show in the following sections). In this application, quarterly performance results were not judged by management to be stable enough. Consequently, a one-year time period length (t = one year) was deemed most appropriate.

In addition, because the sales territories of the 28 branches had been pretty similar in terms of competitive intensity levels and economic conditions, management believed that market performances achieved during the previous year were appropriate benchmarks for comparing current market performances and were used in this analysis. During that year, no new product had been launched, marketing investments had been normal and the market had remained pretty stable.

For every sales territory (ST1 to ST28), the first eight rows of Table 2 give the sales levels, industry sales (estimated by the internal services of the company), the territory market shares, and the gross profit margins during both years 2007 (Y_1) and 2006 (benchmark situation Y_0).

[Insert Table 2]

The percentage variations (in decimal form) in sales (s), industry sales (is), market shares (ms), gross profits (π) and gross profit margins (pm) in 2007 over 2006 have been used for computing the Index of Selling Unit Market Performance (ISUMP), using Equation (11) or (12). In this application, gross profit margins were estimated as:

Gross profit margins = revenues - (sales cost + general administration cost + costs of customer claims)

The interpretation is straightforward: the ISUMP index relates market performance in every sales territory, relative to the benchmark year, *taking into account the evolution of industry sales and consequently, the impacts of competition in the territory*.

Considering first the 2007 results excluding profits, as can be seen in Figure 2, among the four highest performing territories (ST25, ST26, ST27, and ST28), three are located in Zone 1. Overall, nine sales territories (ST28, ST26, ST25, ST22, ST20, ST16, ST14, ST12, and ST11) (or 32 percent) are high sales performers and are located in Zone 1: the firm's position is aggressively strengthened in a declining opportunity market.

In ten sales territories (ST27, ST24, ST23, ST21, ST19, ST18, ST17, ST15, ST10, and ST9), i.e., 36 percent, market shares and sales volumes could increase while demand had been slightly increasing. These territories are located in Zone 2 and their managers have been able to take advantage of the industry sales growth in their territories. They include the second best performing territories (ST27) in which sales significantly increased in an almost stagnant market (borderline of Zone 2).

Six sales territories (ST13, ST8, ST7, ST6, ST4, and ST2), i.e., 21 percent, fall into Zone 4. Their managers do not seem to have properly exploited the growth opportunities of their markets: their sales volumes increased in an expanding market, but not sufficiently to maintain the firm's market position at its original level. In addition, two sales territories (ST3 and ST1), i.e., 7 percent, performed pretty poorly: their management failed to exploit the market growth opportunities and weakened the firm's position in an expanding market (Zone 6). Only in one sales territory (ST5) (4 percent) management could strengthen the firm's market position in a decreasing market demand, but not enough to keep the same sales rate (Zone 3). No sales territory has been observed in Zone 5.

The graphic illustration of Figure 2 allows a better diagnostic of the relative market performances in the different sales territories. Using these results, the top management learned that territories operated in different zones and that it was desirable to adopt different strategies for the territories in the various zones. For instance, it was decided to better reward the efforts of the best performing agents in Zone 1, who aggressively strengthened the firm's position in a declining market.

Adding the profit dimension to this analysis (bottom of Table 2) provides a somewhat different picture. Immediately, one can see that two sales territories are outliers. ST11 changes a substantial loss in 2006 to a large profit in 2007. As mentioned above (in the Proposed Method SECTION), computing a profit percentage variation is meaningless in this case, and consequently, ST11 has been dropped from this part of the analysis. One advantage of this analysis is that it raised a warning flag for management that could then carefully review and diagnose ST11's situation. In this case, it was found that the wide profit variation was mainly the result of a too liberal risk selection by ST11 in 2006 in an attempt to secure sales growth too easily. As a consequence the general sales manager required a more drastic risk selection in ST11, and at the same time, increased prices in this territory. These two actions have led to an unusual profit level in 2007 (compared to the loss in 2006).

ST4 is another outlier. Although sales, industry sales and market shares have not varied substantially in these territories between 2006 and 2007, profit margins have been multiplied by a factor of 10! Here again, this situation could be explained by a very liberal risk selection in ST4 in 2006. As a consequence, a more rigorous risk selection imposed by higher level management was followed by extra selling efforts in ST4 and by a drastic decrease in claim costs in 2007. These two cases highlight the need to interpret the indices qualitatively, as mentioned above.

Of the 27 sales territories kept in this analysis, 22 (81 percent) fall into zone a (profit margin increase). Only five (ST20, ST13, ST9, ST8, and ST1) or 19 percent, fall in a zone b and experience a profit margin decrease. Consequently, because of these overall good results, the whole company experiences a high ISUMP index of 119.51. The highest performing sales territories (all factors accounted for) are successively ST18 (205.95), ST17 (164 78), ST5 (147.84), ST14 (147.80), and ST15 (142.18). The three lowest performing territories are ST20 (87.38), ST9 (98.26), and ST1 (99.02). This shows that including the profit dimension into the analysis sheds some new light on the various territories' market performances. For example, ST5 with a below average sales performance (adjusted ISUMP = 98.81), has a superior performance when profits are also considered (adjusted ISUMP = 123.71).

Although this analysis has highlighted the use of the ISUMP, one should not overlook that this index must be considered along series of other indicators in order to obtain a complete picture of every selling unit market performance. Like with any other quantitative method, the results from this analysis should also be assessed qualitatively. Sales territories that depart sharply from an ISUMP value of 100 should be considered for further diagnosis and possible corrective actions.

Comparison with the Previous Evaluation Method

One of the most frequently used bases for evaluating a unit's sales performance is sales increase/decrease over the last (similar) period. In the reported case study, variations of selling unit's sales as a measure of market performance does not fully account for the different territory situations (industry sales variations or market share evolution). As can be seen in Figure 1, a given level of sales decrease (for instance s = 0.90) can yield quite different market performance assessments, depending on whether the performance vector ends up in Zone 1, 2, or 4.

Table 3 provides a comparison of the sales territory rankings based on their ISUMP scores (including profits) and based on sales variations exclusively. The Spearman coefficient of correlation is a low 0.100. In addition, in the present case, several sales territories experienced a low market performance with the ISUMP index excluding profit and a very good market performance with the index including profit (ST4, ST5).

[Insert Table 3]

In the insurance industry, the profitability of a company is characterized by a rigorous risk selection. Consequently, one observes substantial variations of profitability from one year to another depending on the rigor of the risk selection decisions. Sometimes, insurance agents subscribe bad risks in order to increase their sales. For this reason, great variability may exist between the indices depending on whether profits are taken into account or not. In other industries, one may observe that the two indices are closely connected. In a former study involving a pharmaceutical company, the Spearman coefficient of correlation between the two indices was a higher 0.757 (sales variations are a major factor in both methods). This highlights the need for managers to specify the weights they want to give to sales growth versus profitability. This point is further discussed in a subsequent section.

Longitudinal Market Performance Analysis

The proposed procedure can be applied to track a selling unit's market performance over time. Such long-term dynamic analyses provide an evolutionary evaluation of sales territories' market performance. In this case study, the yearly ISUMP indices have been computed for every territory over a three year period (2005 - 2007) and the yearly evolution has been represented. For illustrative purpose, the ISUMP for only one sales territory (ST28) as well as for the overall sales office, are reported in Table 4 and Figure 4 over the three-year period (Y1= 2006 (versus 2005) and Y2=2007 (versus 2006)).

[Insert Table 4 and Figure 4]

A quick glance at the left part of the chart reveals that ST28 (which is the best market performer in 2006 and 2007) experiences a decrease in market performance over time, although remaining above the office average levels that remain about constant.

The right part of Figure 4 provides a visual comparison between the evolutions of the ISUMP indices for ST28 depending on whether the indices account for the profit variations (right part) or not (left part). The two sets of indices display different patterns: when profits are taken into account, ST28 and the overall office do perform at higher levels, but both decline over the two-year period. ST28 remains only slightly above average for each of the two consecutive years.

Strategic Market Performance Analysis

For illustrative purpose, the analysis has been extended to the cases where the firm pursues either market penetration or market skimming strategies (versus an undifferentiated strategy giving equal importance to all three objectives). In such cases, top management could assign different weights to the three objectives (sales volume, market share, and profits). In these occurrences, the firm's management should clearly communicate those weights to all the concerned managers, and inform them of the strategic market priorities. As an example, management could have assigned the following weights:

Weights for:	**Market skimming strategy**	**Market penetration strategy**
Market share increases	$\alpha = 0.1$	$\alpha = 0.6$
Sales volume increases	$\beta = 0.3$	$\beta = 0.3$
Profit increases	$\gamma = 0.6$	$\gamma = 0.1$
Total	**1.0**	**1.0**

The results obtained after application of Equation (14) are shown in Table 5 and Figure 5. This firm would achieve a better overall market performance if it had followed a market skimming strategy (ISUMP= 145.46 (2006) and 140.64 (2007)), and a lower market performance if it had followed a market penetration strategy (ISUMP =102.82 (2006) and 102.24 (2007)) or an undifferentiated strategy (ISUMP= 122.15 (2006) and 119.50 (2007)). In others words, this firm would have a better performance if it followed a market skimming strategy. ST28's market performance assessment would follow a very similar pattern. As a result, the adjusted ISUMP for ST28 remains very stable, irrespective of the strategy being followed (respectively 109.63, 107.93, and 108.26 for 2006; 108.15, 103.77, and 106.45 for 2007). Consequently, weighting the different elements included in the ISUMP index so as to reflect the selected market strategy makes it possible to assess how effective every selling unit has been at implementing this strategy.

[Insert Table 5 and Figure 5]

Following this implementation, the top managers in charge of business development in North America were quite impressed by the outcomes of this project. For the first time, they could make a sound analysis of the market performance in the different sales territories in terms of business development, which was not done before. For the 2007 annual sales territories' evaluation, top management used the results of this analysis for making a better allocation of its resources. They had at their disposal a sound basis for denying additional resources requested by some sales agents and assigning them to sales territories that could profitably take advantage of market opportunities. This method provided top management with a powerful tool that could help assess their own strategies and decisions and point to possible resource allocation improvements. In addition, managers could identify the zones with large untapped company potential for growth. Although the ISUMP indices are indicators of market performance, they could allow management to identify and question the sales managers in charge of every territory in order to find proper explanations for their market performance and plan actions to take advantage of every market opportunity.

Advantages and Limits

The proposed procedure has several advantages:

First, it provides short-run and/or long-run sales market performance assessments that, as can be observed frequently in practice, are dimensions that many sales managers like to watch very closely. It reveals also which selling units could and/or should take advantage of market opportunities in order to reinforce the firm's market position, a typically long-term objective, generally implying customer relationship and loyalty building.

Second, the proposed market performance measure is simple to compute and understand within a firm and its various selling units. It is based upon the three major ingredients of market performance, i.e., increases/decreases of (1) sales, (2) market share, and (3) profit, compared with some selected benchmark(s). These are elements over which selling units are generally recognized to exert direct influence, and that are easy to measure. As a result, this method can be implemented easily and at low cost as part of a CRM application or business intelligence programs.

Third, the proposed market performance measurement process is fair. It provides selling units with evaluations that are commensurate with their actual market performance, by making the generally reasonable assumption that environmental conditions equally affect a firm and its competitors.

Fourth, the procedure is dynamic, and can be applied over several consecutive periods of time, and/or with various time lengths. As a result, it can be a useful device for tracking selling unit market performance over time, for shorter or longer periods of time, depending on the objectives.

Fifth, the ISUMP index can be applied to assess the market performance of various sales entities (from territories assigned to a sales teams or an individual salesperson, regional or district sales offices, to the whole sales organization) and for various product lines, providing a common basis for making useful comparisons among and across those entities.

Finally, a firm could use the ISUMP indices for allocating financial rewards (such a quarterly bonus) for short-run market performance (McAdams 1987). As can be seen from equation (7), this amounts to providing a reward that is proportional to current sales (S_1), but at a rate that is specific to each selling unit and that accounts for basic territory characteristics. Note that when selling teams are involved, one part of the bonus may be allocated according to the unit (e.g., regional office) market performance index, and one part according to individual market performance measures.

The proposed procedure has also a few limitations:

First, this method is applicable in cases where several firms compete in the same market and when the considered firm does not hold too large a market position compared to its competitors.

Second, random sales variations due to environmental uncertainties or unusual circumstances are accounted for only implicitly. For instance, an unexpected windfall sale would increase S_1, s, and consequently, provide too high a market performance evaluation for that period. Like in the cases of ST11 and ST4 in the reported case study, the ISUMP index can serve as a flag to alert management that further investigation should be carried out. Thus, such unusual occurrences will be recognized by management and judgmentally included in the analysis. In other words, the ISUMP indices are only market performance diagnosis tools. If they point to more and/or less efficient selling units taking the territory conditions into account, they do not diagnose why such market performance levels have been reached. In other words, the ISUMP index can identify possible problems, but does not provide a diagnosis.

Third, as a corollary, such market performance analyses should not prevent management from watching other more qualitative aspects of selling units' market performances. For instance many human aspects (such as, for instance, the characteristics of the salespersons or the sales teams in charge of the selling units, e.g., their experience or career stage) should be kept in mind when carrying such an analysis. In fact, the ISUMP analyses should be considered as only one dimension (although a powerful one) of market performance assessments.

Finally, the proposed method requires access to sufficiently reliable market share data for every selling unit. Note, however, that in many industries (like the pharmaceutical industry) firms have access to such syndicated data on a regular basis.

Summary and Conclusion

This paper has described a rather simple procedure for assessing various selling units' market performance, not only in the short-run, but also accounting for a firm's market position improvement (or decay) a typical long-term firm's marketing concern and a more relevant selling unit's market performance measure. The assessment formula is simple to explain and easy to administer. It requires only three sets of data: selling units' sales volumes, market shares, and profits in the corresponding territories. In addition, the outcomes of a selected market strategy can be assessed easily. A firm can easily and at low cost integrate this procedure into its CRM or other sales intelligence system. This procedure has been illustrated by an actual case study. It has been shown to provide more adequate results (from a marketing point of view), and more equitable market performance assessments than more complex (or even simpler) comparable procedures.

There are several ways in which the proposed procedure could be extended or refined. For instance, it could be modified to account for various variables, such as random factors and uncertainties. It should be kept in mind, however, that these refinements would come at the cost of making the procedure more complex. Consequently, they should be included only if the new benefits are worth the additional complexity.

Tables and figures

Table 1 - Six Possible Strategic Performance Situations According to Territory Sales, Market Industry sales, and Market Share Variations

	Sales increase (s > 0)		Sales decrease (s < 0)	
	Market share increase (ms > 0)	**Market share decrease (ms < 0)**	**Market share increase (ms > 0)**	**Market share decrease (ms < 0)**
Industry sales increase (is > 0)	**Case 2** Sales increase at a higher rate than market potential **The salesperson strengthens the company position in an expanding market demand territory**	**Case 4** Sales increase, but at a lower rate than market industry sales **The salesperson weakens the company position in an expanding market demand territory**	**IMPOSSIBLE**	**Case 6** Sales and market share decrease in an expanding market **The salesperson cannot take advantage of the market opportunities in the territory**
Industry sales decrease (is < 0)	**Case 1** Sales and market share increase as the market shrinks **The salesperson strengthens the company position in a declining market demand territory**	**IMPOSSIBLE**	**Case 3** Sales decrease but at a lower rate than the market industry sales **The salesperson offers a strong resistance against the declining market demand in the territory**	**Case 5** Sales decrease at a higher rate than the market industry sales **The salesperson weakens the company position in a declining market demand in the territory**

Table 2 - Sales and Territory Market Share Variations over a Selected Benchmark Period for the Twenty Eight Sale

Sales Data	SALES TERRITORIES									
	ST28	**ST27**	**ST26**	**ST25**	**ST24**	**ST23**	**ST22**	**ST21**	**ST20**	**ST19**
S_0	38375	22251	27438	44621	40582	39002	48201	46524	21521	30578
IS_0	170860	119913	194451	205975	192147	200388	229116	203825	114844	149982
MS_0	0.2246	0.1856	0.1411	0.2166	0.2112	0.1946	0.2104	0.2283	0.1874	0.2039
Π_0	3640118	1307837	3058069	3653074	3255811	3634367	2999814	5207487	2443306	2498190
S_1	39730	23106	28092	45462	41571	39885	48938	47271	21852	31281
IS_1	168894	120262	194174	204352	193758	201415	228524	203913	114837	152288
MS_1	0.2352	0.1921	0.1447	0.2225	0.2146	0.198	0.2141	0.2318	0.1903	0.2054
Π_1	6144282	1612006	3341885	4133169	6063437	4011017	5351855	6407254	1186444	4001957
S	0.0353	0.0384	0.0238	0.0188	0.0244	0.0226	0.0153	0.0161	0.0154	0.023
Is	-0.0115	0.0029	-0.0014	-0.0079	0.0084	0.0051	-0.0026	0.0004	-0.0001	0.0154
Ms	0.0474	0.0354	0.0253	0.0269	0.0159	0.0174	0.0179	0.0156	0.0154	0.0075
Π	0.6879	0.2326	0.0928	0.1314	0.8623	0.1036	0.7841	0.2304	-0.5144	0.6019
Pm	0.6303	0.1870	0.0674	0.1105	0.8179	0.0792	0.7572	0.2109	-0.5218	0.5659
ISUMP Based on Sales and Market Share Variations Only										
ISUMP	104.14	103.69	102.46	102.29	102.01	102	101.66	101.58	101.54	101.53
Adjusted*	103.08	102.64	101.42	101.25	100.98	100.97	100.63	100.55	100.51	100.5
Zone	Zone 1	Zone 2	Zone 1	Zone 1	Zone 2	Zone 2	Zone 1	Zone 2	Zone 1	Zone 2
ISUMP Based on Sales, Profit, and Market Share Variations										
ISUMP	126.85	108.92	103.90	105.29	133.99	104.01	131.08	108.48	87.38	122.66
Adjusted **	106.14	91.15	86.94	88.10	112.12	87.03	109.69	90.77	73.12	102.64
Zone	Zone 1a	Zone 2a	Zone 1a	Zone 1a	Zone 2a	Zone 2a	Zone 1a	Zone 2a	Zone 1b	Zone 2a

Table 2 - (cont'd)

Sales Data	SALES TERRITORIES									
	ST14	ST13	ST12	ST11	ST10	ST9	ST8	ST7	ST6	ST5
S_0	41465	27782	31351	27064	27808	30022	32791	37302	38239	28685
IS_0	161817	123364	150232	124592	160238	159464	154253	199180	154391	156766
MP_0	0.2562	0.2252	0.2087	0.2172	0.1735	0.1883	0.2126	0.1873	0.2477	0.183
Π_0	2298704	2666088	1458881	-922997	3099928	3074894	1889793	4865202	4583632	1758809
S_1	41842	28333	31551	27156	27894	30143	32955	37492	38339	28584
IS_1	161557	125861	149625	124336	160255	159893	155486	200932	155715	156205
MP_1	0.259	0.2251	0.2109	0.2184	0.1741	0.1885	0.2119	0.1866	0.2462	0.183
Π_1	4927893	3135235	1688500	3735137	3241197	2906008	3476994	5861578	4746430	3744195
S	0.0091	0.0198	0.0064	0.0034	0.0031	0.004	0.005	0.0051	0.0026	-0.0035
Is	-0.0016	0.0202	-0.004	-0.0021	0.0001	0.0027	0.008	0.0088	0.0086	-0.0036
Ms	0.0107	-0.0004	0.0105	0.0055	0.003	0.0013	-0.003	-0.0037	-0.0059	0.0001
Π	1.1438	0.176	0.1574	NA°	0.0456	-0.0549	0.8399	0.2048	0.0355	1.1288
Pm	1.1245	0.1532	0.1500	NA°	0.0424	-0.0587	0.8307	0.1987	0.0328	1.1363
ISUMP Based on Sales and Market Share Variations Only										
ISUMP	100.99	100.98	100.84	100.44	100.3	100.27	100.1	100.07	99.84	99.83
Adjusted*	99.96	99.95	99.82	99.42	99.28	99.25	99.09	99.05	98.82	98.81
Zone	Zone 1	Zone 4	Zone 1	Zone 1	Zone 2	Zone 2	Zone 4	Zone 4	Zone 4	Zone 3
ISUMP Based on Sales, Profit, and Market Share Variations										
ISUMP	147.80	105.97	105.78	NA°	101.63	98.26	133.61	107.08	101.00	147.84
Adjusted **	123.67	88.68	88.51	NA°	85.04	82.23	111.80	89.60	84.51	123.71
Zone	Zone 1a	Zone 4b	Zone 1a	NA°	Zone 2a	Zone 2b	Zone 4b	Zone 4a	Zone 4a	Zone 3a

*Office ISUMP Index: 101.03 (Zone 2)
**Office ISUMP Index: 119.51 (Zone 2a)

°Because profit (and gross profit margin) variations are meaningless in this case, ST11 has been

Table 3 - Comparison of Managers' Market Performance under Sales Variations Versus Sales, Profits, and Market Share Variations

Sales territories	Sales, Profits, and Market Share Assessment Basis			Sales Assessment Basis	
	Index	Zone	Ranking*	Index	Ranking*
ST4	582.45	Zone 4a	1	0.0010	23
ST18	205.95	Zone 2a	2	0.0176	9
ST17	164.78	Zone 2a	3	0.0125	13
ST5	147.84	Zone 3a	4	-0.0035	26
ST14	147.80	Zone 1a	5	0.0091	16
ST15	142.18	Zone 2a	6	0.0123	14
ST16	139.81	Zone 1a	7	0.0111	15
ST24	133.99	Zone 2a	8	0.0244	3
ST8	133.61	Zone 4a	9	0.0050	19
ST22	131.08	Zone 1a	10	0.0153	12
ST28	126.85	Zone 1a	11	0.0353	2
ST19	122.66	Zone 2a	12	0.0230	5
ST2	114.31	Zone 4a	13	0.0005	24
ST27	108.92	Zone 2a	14	0.0384	1
ST21	108.48	Zone 2b	15	0.0161	10
ST7	107.08	Zone 4a	16	0.0051	18
ST13	105.97	Zone 4a	17	0.0198	7
ST12	105.78	Zone 1b	18	0.0064	17
ST25	105.29	Zone 1a	19	0.0188	8
ST23	104.01	Zone 2a	20	0.0226	6
ST26	103.90	Zone 1a	21	0.0238	4
ST10	101.63	Zone 2a	22	0.0031	21
ST6	101.00	Zone 4a	23	0.0026	22
ST3	100.19	Zone 6a	24	-0.0028	25
ST1	99.02	Zone 6b	25	-0.0140	27
ST9	98.26	Zone 2a	26	0.0040	20
ST20	87.38	Zone 1a	27	0.0154	11
Total Office	**119.51**	**Zone 2a**			

***Spearman's Rank Correlation = 0.100**

Table 4 - Sales Territory ST28 and Industry Sales Data over a Two-Year Period

Sales Data	**SALES TERRITORY ST28**	
	Year1 (2006)	**Year2 (2007)**
S_0	36972	38375
IS_0	182036	170860
MS_0	0.2031	0.2245
Π_0	3276106	3640118
S_1	38375	39730
IS_1	170860	168894
MP_1	0.2245	0.2352
Π_1	5837068	6144282
S	0.03794	0.0353
Is	-0.06139	-0.0115
Ms	0.10583	0.0474
Π	0.78170	0.6879
Pm	0.71657	0.6303
ISUMP Based on Sales and Market Share Variations Only		
ISUMP	107.24	104.14
Adjusted*	105.51	103.08
Zone	Zone 1	Zone 1
ISUMP Based on Sales, Profit, and Market Share Variations		
ISUMP	132.24	126.85
Adjusted **	108.26	106.14
Zone	Zone 1a	Zone 1a
TOTAL SALES OFFICE		
S_0	996641	933985
IS_0	4788307	4623362
MS_0	0.1939	0.2020
Π_0	64857067	70496812
S_1	1013779	944995
IS_1	4796080	4637304
MS_1	0.1964	0.2037
Π_1	102556347	106829528
S	0.0171	0.0117
Is	0.0016	0.0030
Ms	0.0126	0.0087
Π	0.5812	0.5153
Pm	0.5546	0.4977
ISUMP Based on Sales and Market Share Variations Only		
ISUMP	101.63	101.03
Zone	Zone 2	Zone 2
ISUMP Based on Sales, Profit, and Market Share Variations		
ISUMP	122.15	119.51
Zone	Zone 2a	Zone 2a

Table 5 - Impact of Various Market Strategies upon ISUMPs (Yearly Data)

Sales Data	Year 1 (2006)	Year 2 (2007)
SALES TERRITORY ST28		
S	0.0379	0.0353
Ms	0.1058	0.0473
Π	0.7817	0.6879
Pm	0.7166	0.6303
ISUMP Based on a Market Skimming Strategy (α=0.1, β=0.3, γ=0.6)		
ISUMP	159.48	152.11
Adjusted	**109.63**	**108.15**
ISUMP Based on a Market Penetration Strategy (α =0.6, β =0.3, γ =0.1)		
ISUMP	110.98	106.10
Adjusted	**107.93**	**103.77**
ISUMP Based on Equal Weight for Sales, Profits, and Market Share Variations (α =0.33, β =0.33, γ =0.33)		
ISUMP	132.24	126.84
Adjusted	**108.26**	**106.14**
TOTAL OFFICE		
S	0.0171	0.0117
Ms	0.0126	0.0087
Π	0.5812	0.5153
Pm	0.5546	0.4977
ISUMP Based on a Market Skimming Strategy (α =0.1, β =0.3, γ =0.6)		
ISUMP	145.46	140.64
ISUMP Based on a Market Penetration Strategy (α =0.6, β =0.3, γ =0.1)		
ISUMP	102.82	102.24
ISUMP Based on Equal Weight for Sales, Profits, and Market Share Variations (α =0.33, β =0.33, γ =0.33)		
ISUMP	122.15	119.50

Figure 1 - The Six Performance Zones

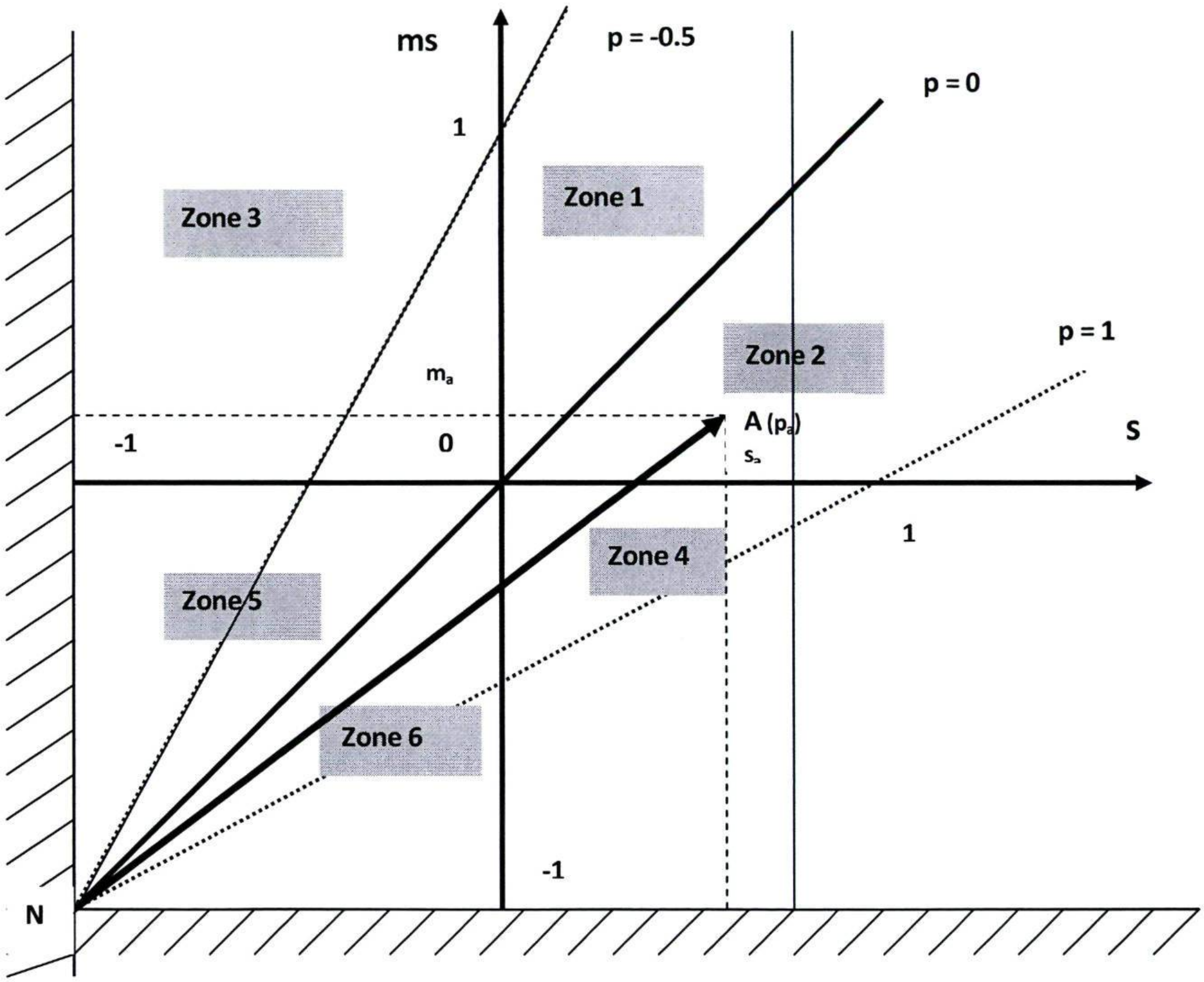


Figure 2 - Positions (Tips of Market performance Vectors) of Seventeen Selected Sales Manager on the Six-Zone Grid and Overall Office (Darker Arrow)

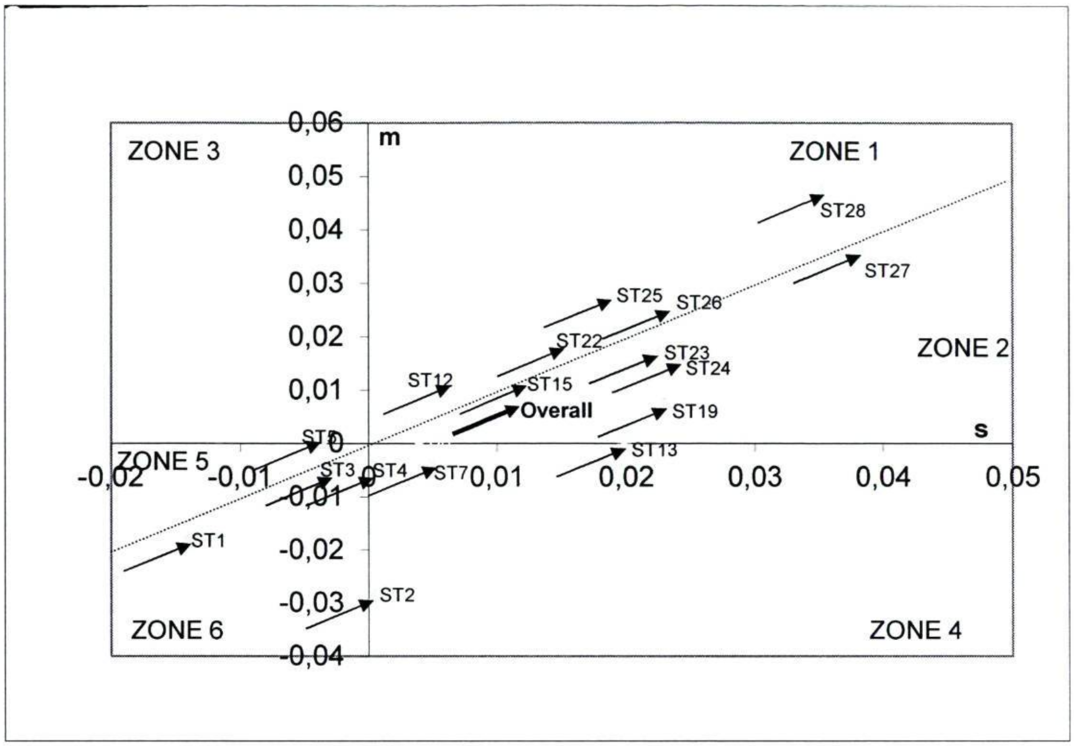


Figure 3 - Selling Unit's Market Performance Vector in a Three Dimensional Space

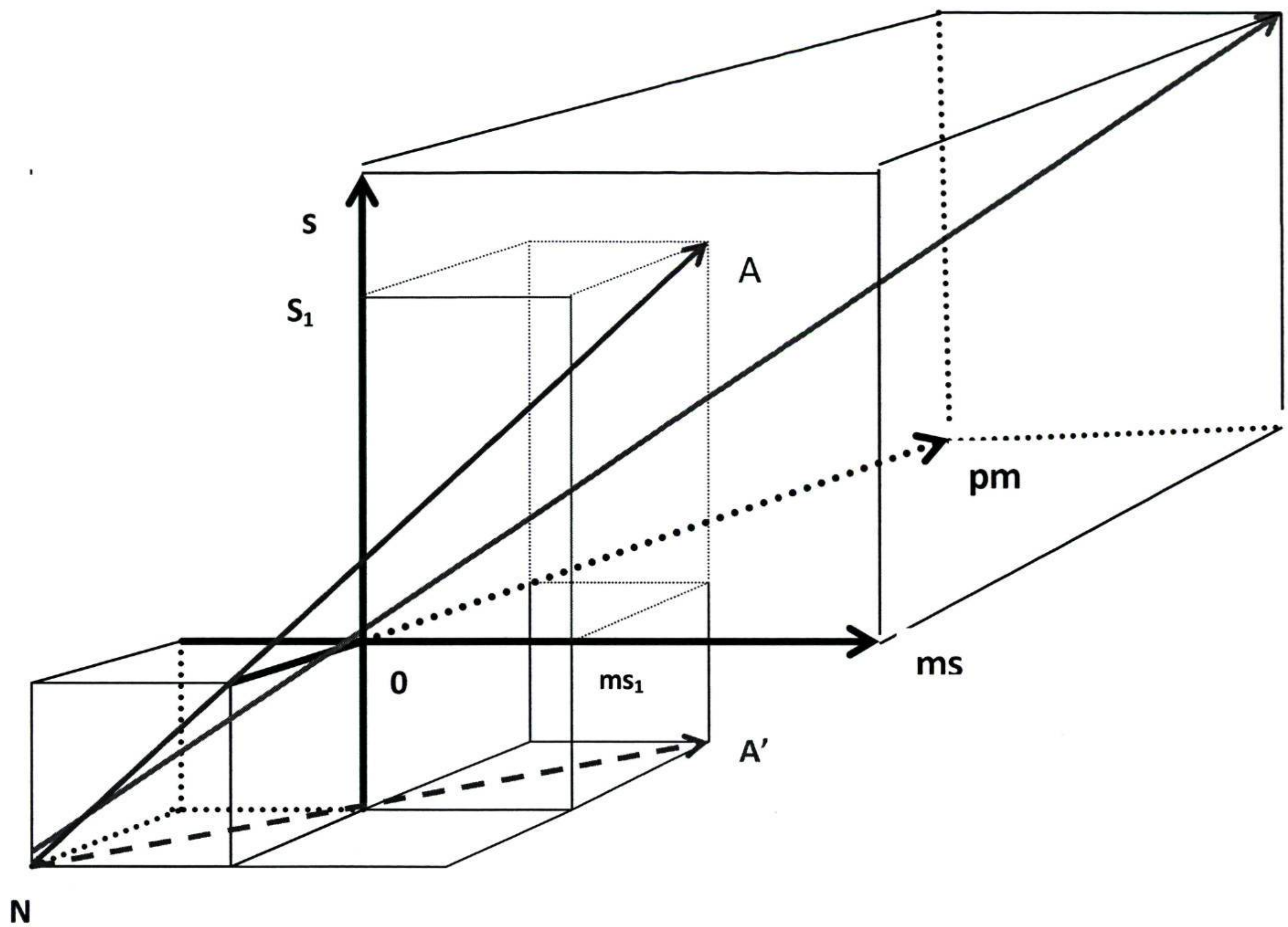
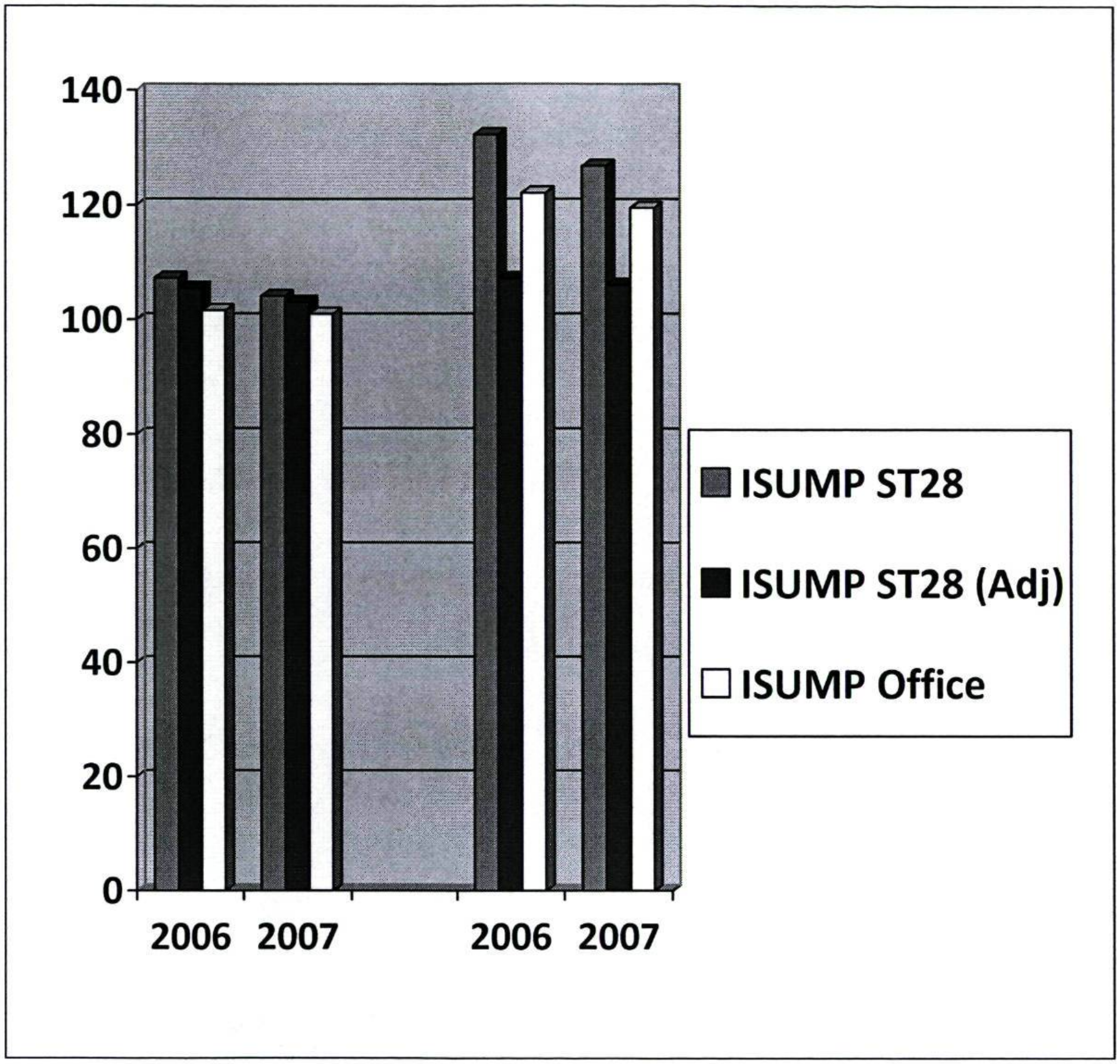


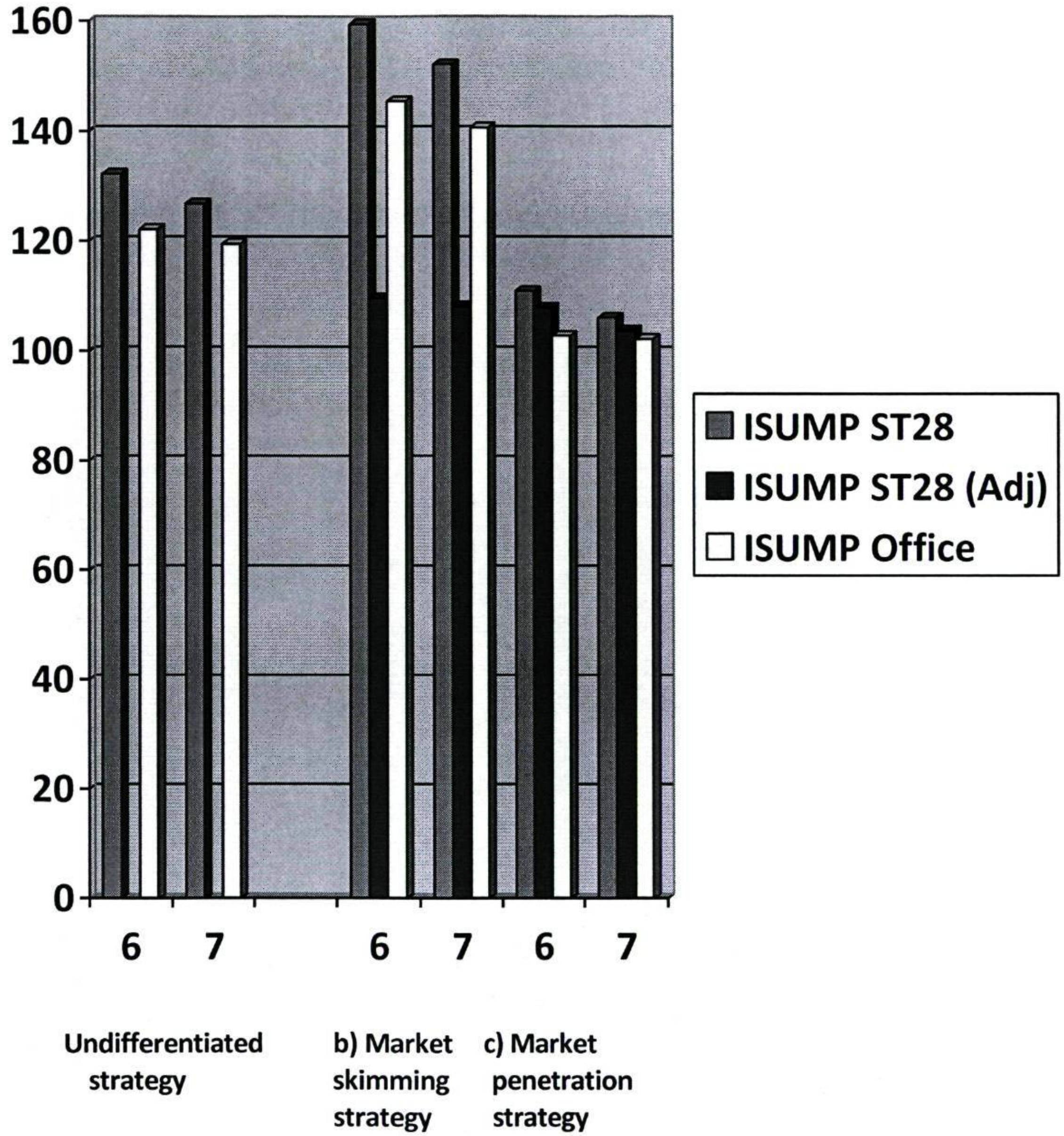
Figure 4 - Market Performance in Sales Territory ST28 and Sales Office over a Two-Year Period



a) **Sales and market share variations**

b) Sales, market share and profit variations

Figure 5 - Evolution of the ISUMP Index over a Two-Year Period under Three Different Market Strategies



How Health Sciences Practitioners Can Use Data Mining for Predicting Individuals' Risks of Contracting Nosocomial Pneumonia

Abstract

From the beginning of health management sciences, managers have tried to find better techniques to classify and qualify risks in order to effectively manage health operations, including reducing costs while increasing patient wellbeing. This paper explains how managers can use data mining techniques for solving problems related to individual risks of contracting nosocomial pneumonia (within 48-72 hours after hospital admission).

Keywords: Nosocomial Pneumonia, Nosocomial Infection, Health Management, Health Science, Health Informatics, Decision Support System, Predictive analysis, Data Mining.

Résumé (French)

Depuis le début de sciences de la gestion de la santé, les gestionnaires ont toujours tenté de trouver de meilleures techniques pour classer et qualifier les risques afin de gérer efficacement les activités de leurs établissements, et ce, afin de réduire les coûts, tout en augmentant le bien-être des patients. Cet article présente une solution analytique qui prédit les risques qu'un patient contracte une pneumonie nosocomiale. Cette solution a été testée sur une base de données d'un grand hôpital américain.

Mots clefs: Pneumonie nosocomiale, infections nosocomiales, gestion de la santé, sciences de la santé, l'informatique de santé, aide à la décision, analyse prédictive, Data Mining.

Article

Introduction

From the beginning of health management sciences, managers have tried to find better techniques for classifying and qualifying risks, in order to prevent infections, identify infected people (Vason, B. J. 2004), understand the factors responsible for diseases (Harper, P. 2005), give better care, increase the performance of resources and reduce costs (Fitzpatrick, M. A. 2006). Relative to this, we tested data mining techniques to predict individuals' risks of contracting nosocomial pneumonia (NP) infections. The results are based on data from a large American (U.S.) hospital (655 000 patient admissions in two years).

This paper aims to expose health managers and related health sciences practitioners to the usefulness of employing data mining techniques to predict individuals' risks of contracting NP.

Data mining is a promising approach for solving this problem because it can handle large numbers of observations, variables, outliers, unbalanced data sets and over-fitting, while being highly accurate.

This paper has been organized in the following way: first, previous work on data mining applied to health care problems is presented; second, NP is explained; third, the managerial benefits of using data mining to predict NP risk are exposed; fourth, the methodology used is detailed; fifth, we show how to transform data mining results into managerial decisions; sixth, the data mining performance results are illustrated; seventh, we present our conclusions.

Previous work

The use of data mining in the healthcare industry for medical decision support is relatively new. In fact, in recent years, the data mining community has principally limited its work to this area of application:

[Insert Figure 1]

Our contribution in this area is to develop a solution to allow healthcare professionals to better predict and detect, immediately upon admission, the probability of a patient contracting NP during the hospital stay.

Other data mining based solutions have been developed for medical decision support (to find important factors responsible for myocardial infarction (Burn, T. 1998), spread infection (Brosset,S. E. 1998), schizophrenia (Michael, G. M. 2010) and other related diseases (Mauricio, C.-R. 2010)), as well as for healthcare fraud problems (Rasim, M. 2010 and Milley, A. 2000).

Nosocomial Pneumonia infection

NP refers to any pneumonia contracted within 48-72 hours after being admitted to a hospital.

As a result, the patient generally (Gerald, L. et al 2004):

- Stays in the hospital one to two weeks longer than expected;
- Requires additional care and treatment;
- Loses work-days and quality of life.

NP affects over a million people each year in the United States and Europe (Boulder, C. 1998) and 88 000 of them die every year (4th Decennial International Conference on Nosocomial and Healthcare-Associated Infections 2000).

NP is one kind of nosocomial infection. Ten per cent (10%) of all people admitted to health care facilities suffer from a nosocomial infection. As a consequence, 1% of all people admitted die of this infection (Ministère de la santé et des services sociaux, 2006).

Studies have shown that it is possible to reduce by 33% the number of newly infected patients using structured prevention programs, which include data mining techniques (Ministère de la santé et des services sociaux, 2006).

National health departments generally develop and transmit action plans to combat nosocomial infection in healthcare facilities. These plans generally include guidance, principles and frameworks. However, the responsibility generally belongs to healthcare facilities for developing their own program of prevention and control of nosocomial infections (Ministère de la santé et des services sociaux, 2006).

Data mining techniques are tools which aim to help health care facilities to implement their program in a cost efficient manner.

Managerial benefits

Using data mining techniques, like those presented in this paper, allows healthcare facilities managers to assist stakeholders in their decision-making and to reduce costs.

In fact, data mining techniques can recommend stakeholders perform a particular action on a patient according to his or her probability of contracting a NP. Consequently, managers can act more rapidly to fight the infection (Vason, B. J. 2004).

More precisely, data mining techniques assist stakeholders to formulate recommendations for preventive measures to be administered to patients. These recommendations are developed from management information and from data mining techniques explained by the two following figures.

[Insert Figure 2] [Insert Figure 3]

For example, in a context where isolation rooms, a common infection prevention protocol, are limited, data mining techniques could advise stakeholders on which patients to send to these rooms, i.e. the one with the highest predicted risk (probability).

In addition, data mining techniques help to reduce costs by restricting high cost protocols only to patients with the highest probability of acquiring an NP. Thus, stakeholders save resources and reduce ongoing costs.

Let's say, for example, we have one heath care facility which has the following characteristics:

- 655 000 admissions;
- 65 500 admissions suffer from a NP (10% of all people admitted);
- 6 550 admissions die because of they contracted NP
- Each infected patient remains in hospital four days for treatment of his NP (examinations and treatments) (Ministère de la santé et des services sociaux, 2008);
- It costs $741 to treat a minor NP (Ministère de la santé et des services sociaux, 2008).

The estimated NP cost for this heath care facility is $48,535,500 US and 717 beds per year (262 000 nights).

Knowing that a structured prevention program, assisted by data mining techniques would reduce by 33% the number of newly infected patients, the heath care facility could save $16 ,016, 715 US and 236 beds per year (86 460 nights) by using data mining techniques.

Methodology

In this study, the managerial and information process used to predict NP infections is as follows:

[Insert Figure 4]

On the managerial process side, the patient comes to the health care facility, he completes his registration and immediately after, with the data mining technique results, the stakeholders are informed about the recommendation (protocol) to take to prevent nosocomial pneumonia.

On the information side, the information about the patient is subjected to the Data Mining techniques which integrate the results into the health care facility systems.

On the technical side, the process used to construct the Data Mining techniques is as follows:

[Insert Figure 6]

First, we developed the Data Mining techniques by:

- Making the different databases communicate with each other to collect all the information on the back office databases;
- Extracting the needed data and dealing with missing data and outliers;
- Preparing the data for the Data Mining algorithm;
- Writing the code relative to the Data Mining algorithm in the heath care facility systems;
- Predicting the data from the test database, to validate the performance of the data mining techniques;
- Writing the code relative to the implementation of all the processes on the heath care facility systems.

To verify that the Data Mining algorithm will work when we implement it, we predicted (scored) the test database observations and validated the performance. The Data Mining algorithm should have similar performance in the training database and in the test database to be considered as superior performance. The Data Mining algorithm with poor performance will not be implemented and the Data Mining algorithm with superior performance will be implemented. The test database represented 30% of our database.

When the implementation of the Data Mining techniques is complete, one code runs upon the registration of the patient. The code collects, from the database, the different information needed, applies the Data Mining algorithm, predicts the risk of the patient contracting NP and recommends the appropriate protocol to prevent the patient from contracting NP. Figure 4 illustrates this process. The risk of the patient contracting NP is the output score of the Data Mining algorithm. It could be interpreted as "probability" in the case of linear Data Mining algorithm.

Data Description

In this study, two years of data from more than 655 000 patient visits were available. The objective was to predict what will happen the following year. Four types of data sets involving 140 variables (information) on the patients were used (see the Appendix for the important variables available in the database).

Among the 140 variables available for each patient, the proposed model identified the eight variables that are the most discriminating (Figure 5), and two of them are extremely important in this model: the past medication taken by the patient and the patient's physical state (The International Classification of Diseases of the patient) before the current hospital visit. Both variables have an extremely important impact on the risk of contracting NP. In other words, some drugs and certain health conditions may involve higher risks of contracting NP.

[Insert Figure 5]

Other variables also appear moderately related to the risk of contracting NP, namely the predicted facility cost paid by Medicare, the predicted total reimbursement, the predicted total charges, the predicted primary procedure, the predicted payment facilities, and the reasons why a patient entered the hospital.

Previous research has found that NP is usually caused by a bacterial infection (Warrell, D. A. et al 2003). Urinary tract infections are the second biggest cause of nosocomial infections (Fauci, A. S. et al 2008). For this application, information about urinary tract infection was not available and consequently, this risk factor could not be confirmed. Moreover, the literature also suggests other possible causes of NP (Table 1).

[Insert Table 1]

Unlike other studies, the age of the patient was not found to be an important risk factor in this analysis. A possible interpretation could be that our database had more important risk factors than age. Consistent with the extant literature, the present analysis confirmed that the medications and the presence of other diseases increased the risk of contracting NP.

The discriminating power of the variables has been measured by (a) the evaluation of the reduction of predictive accuracy after a random permutation of the values assumed by the variable; and (b) the total heterogeneity reduction produced by the variable on the response variable (binary nosocomial pneumonia variable in this case) (Breiman 2002)

Data Preparation

In this application, the data set was characterized by a lot of missing and outlying values. Moreover, the most important variables had the largest number of missing values. Consequently, those values had to be replaced in order to avoid decreasing their discriminating power.

For missing values in numeric variables, the Expectation-Maximization (EM) algorithm was used. The EM algorithm (Rom, J. F. 1992) imputes the missing values by using an iterative process which calculates the maximum likelihood of a missing value (Dempster, A. et al 1977). In other words, this algorithm replaces the missing values by the most probable value, given the available data. For categorical data, the surrogate algorithm (He, Y. 2006, Liu, W. Z. 2006), was used, because it usually provides the best results with tree-based techniques. The imputation of a missing value is a prediction made by the gini decision trees technique. This prediction is based on the values of other variables in the database (He, Y. 2006, Liu, W. Z. 2006). Software like SAS, SPSS, R, etc. could execute those algorithms.

No special technique was used to control extreme values because the gradient boosting tree technique generally deals with this problem. In addition, because health data display a large variance, highly variable by nature, it is difficult to manage the extreme values without damaging the structure of the data.

Model

Health managers face very tight budgets. As a result, health managers must use the best techniques in order to be more efficient and to improve their services to citizens, within their budget constraints.

Consequently, the real issue is how to build a model, from a Data Mining algorithm, which maximizes the area under the curve (a popular data mining evaluation criterion). In other terms, we want a model which assigns higher probabilities to patients who will contract NP.

Let's explain how the area under the curve is calculated. Since predicting if patients will contract NP or not is a binary problem, the possible outcomes can be represented in a confusion matrix, where tp (true positive), fn (false negative), tn (true negative) and fp (false positive) represent all the possible outcomes:

[Insert Table 2]

After building a confusion matrix, the sensitivity (also called true positive rate or hit rate) and the specificity (true negative rate) is defined as:
Sensitivity = tp/pos
Specificity = tn/neg
where pos=tp+fn is the total number of positive examples and neg=tn+fp the total number of negative examples.

The area under the curve corresponds to the area under the curve - plotting sensitivity against specificity by varying a threshold on the prediction values to determine the classification result. The curve is given by {(1,0),(tn/(tn+fp),tp/(tp+fn)),(0,1)}. Health sciences practitioners could use software like SAS, R (an open source statistical software), etc. to calculate area under the curve. The higher the area under the curve is, the better the model. A perfect model will correctly predict the outcomes, in other words there will be only true positive and true negative cases (so, the area under the curve will be 1.0).

In this study, using our own code, two data mining techniques have been used jointly in order to maximize area under the curve (to build better models). Health sciences practitioners could use software like SAS, R, etc. to build a Data Mining model.

More precisely, the genetic algorithm has been used for providing an optimal choice of parameters for gini boosting type decision tree models. Those parameters are controlled genetically: number of trees, depth of trees, trimming factor, cross-validation (to avoid over-learning, i.e., the possibility for the model to learn too much from the training sample data, thus impairing its predictive validity), proportion of the population used, and the minimum size for splitting a node. The link between genetic algorithm and gini boosting type decision tree models is that we used a search algorithm (genetic algorithm) to find the best parameters of one Data Mining algorithm (gini boosting type decision tree models).

First, a model with the gini boosting type decision tree technique was built. The idea behind this technique is to build successive trees (Friedman, J. H. 1999a). Each new tree in the series is based on a sample of incorrectly classified observations (of the current series) and from random observations. In addition, the probability associated with one person is the sum (or the weighed sum) of the values this person obtained at each terminal node of each tree (Friedman, J. H. 1999b). This technique could not be done manually by practitioners.

A major issue is: how to choose the best parameters for this technique? There are many possibilities, resulting in diverging performances. Furthermore, the choice of these parameters is very important because it makes it possible to avoid the over-learning and to manage the performance. Table 3 shows the parameters that must be chosen when using gini boosting type decision tree technique (Friedman, J. H. et al 2001):

[Insert Table 3]

There are multiple parameters to select in order to obtain a model with acceptable performance and that does not over-learn. We used genetic algorithm with the selection of these parameters. Genetic algorithm is a technique which identifies the possibilities with the greatest chances of being the best choices. In addition, it tries to avoid overlooking certain zones of possibility (Haupt, R. L. et al 2004).

Starting with initial groups of choices (possibilities), these groups evolve in order to keep the most accurate groups of choices, i.e., those that are likely to include the best solutions. The worst groups of choices are eliminated from the pool. Consequently, at the end, when the algorithm converges, a group of choices (possibilities) which is closest to the best solution is obtained, because the worst groups of choices are eliminated after each iteration (Mitchell, M. 1999). In brief, multiple gini boosting type decision tree technique models (with multiple parameters) are evaluated in order to obtain the model with the best performance.

The parameters which have been simultaneously selected with the assistance of the genetic algorithm are given in Table 4.

[Insert Table 4]

From data mining results to managerial decisions

This section explains in detail how we build the recommendations.

To build the recommendations with stakeholders we build a table as follows:

- During development of the Data Mining techniques, we store some data about patients to estimate the performance of our Data Mining techniques during the implementation. We called this data the test database;
- Prior to implementation, we use the Data Mining algorithm found to predict the probability of patients (included in the test database) to contract a NP;
- Subsequently, we compare the predictions with what actually happened, by building a list of patient s in descending order, according to their likelihood of contracting the NP (so, those that we determine the most at risk are found at the top of the list);
- Then we divide this list in increments of 10% that we called bin;
- Then, in each slice, we calculate the % who actually had the infection;
- Finally, we construct a table that shows the cumulated results.

This is the table resulting from our work on the test database:

[Insert Figure 9]

Data mining performance results

From a more technical point of view, it is interesting to see that the joint use of the genetic algorithm and the gini boosting type decision tree clearly controls over-learning.
In addition, this model produces a very low misclassification rate.

[Insert Figure 7]

The model could correctly identify 86.17% of the patients who actually contracted NP (since the separation between NP and pneumonia was not indicated in the data, we assume that all cases were NP cases). Consequently, given that managers can develop special protocols for decreasing the risks of contracting NP, they can, for example, require that these patients be placed in an isolation room.

Figure 3 shows the graph illustrating the area under the curve (A.U.C.). It shows that the model is very powerful compared to a random model with an area under the curve of 74.50 %.

[Insert Figure 8]

This model made it possible to discover new risk factors to be taken into account in tariff setting and in medical research. In addition, managers can better understand the important variables that increase the risk of NP (Mosley, D. 2005).

Conclusion

In the context of NP prediction, Data Mining techniques can be used for many purposes: segmenting the risks; discovering new risk factors; understanding the important variables; developing a special process in order to decrease the risks to contract NP. In this application, 86% of the patients who will contract NP were located in 10% of the population. This provides a powerful tool to managers. This technique can help with: preventing infections; identifying the infected people; understanding and discovering the factors responsible of disease, so giving better care; increasing the performance of resources and reducing the costs.

In the case of a real implementation, some caution is required.

In fact, the implementation of such solutions should address some challenges on the management side.

In fact, our experience in health industry tells us that it's difficult to change the current process/behavior in this industry. In fact, changing the current nosocomial infection process will require a lot of "change management" skill. In this case, a possible mitigation strategy could be to:

- Involve, at the beginning of the project, a highly specialized "change management" consultant to elaborate, execute and monitoring a "change management" strategy;
- Involve, at the beginning of the project, all identified stakeholders;
- Obtain the support from the direction of the organization.

Moreover, a potential challenge for this kind of implementation is the not availability of intern employees. In fact, if intern employees are not available to collaborate with the project team in building the solution, at determined critical moment, the project could fail. In this case, a possible mitigation strategy could be to:

- Evaluate, at the beginning of the project, the availability needed from identified intern employees;
- Define, at the beginning of the project, with the organization, a strategy to make the identified intern employees available when needed;
- Align, at the beginning of the project, the project working plan to the identified intern employees' current duties;
- Define, at the beginning of the project, a strategy to replace the identified intern employees if needed.

Another potential challenge for this kind of implementation is the not availability of important variables needed in the database of the organization. In this case, a possible mitigation strategy could be to:

- Inform the organization, at the beginning of the project, about the important variables needed according to this paper;
- If important variables needed are missing, develop a strategy to acquire it (ex.: collect from the others database, collect it upon the registration of the patient or use others variables).

Tables and figures

Table 1 - Elements causing the nosocomial pneumonia

Risk factors	Present in the dataset	Is important in the model
Old age (Harris, J. R. 1996)	✓	✗
Medications (Monson, K. 2007, WebMD 2003)	✓	✓
Neurologic, or other disease (Crouse, D. T. 1992) states that result in respiratory tract obstruction	✓	✓
Mechanical ventilation (Schleder, B. J. 2003)	✗	N / A
Intrinsic respiratory (Warrell, D. A. et al 2003)	✗	N / A

- ✓ **In the first column = Present in the dataset**
- ✓ **In the second column = Is important in the model**
- ✗ **In the first column = Not p**resent in the dataset
- ✗ **In the second column = Is not important in the model**

Table 2 - Confusion matrix of area under the curve

		Prediction	
		Class +1 (contracted NP)	Class 0 (Did not contracted NP)
Truth	Class +1 (contracted NP)	Tp	Fn
	Class 0 (Did not contracted NP)	Fp	Tn

Table 3 - Parameters of gini boosting type decision tree technique

Parameters	**Explanations**
Number of trees	When we use the gini boosting type decision tree technique we must determine the maximum number of trees to use in the series. Too many trees will increase the over-learning, not enough will reduce the performance.
Depth of individual trees	In addition, we must manage the depth of individual trees. This parameter determines how many splits (to the maximum) the individual tree can have. Here again, too many splits will increase the over-learning, not enough will reduce the performance.
Minimum size node to split	Also, we must choose the minimum size node to split in each individual tree. Too small a node will increase the over-learning.
Shrink factor	The shrink factor is the weighting coefficient applies to each tree in the series to reduce the learning rate. In order to find the best model, the learning rate should not be too high or too low.
Prior probabilities	Moreover, we must deal with the right prior probabilities of the distribution for the target categories variable. The performance of modeling is sensitive to that.
Proportion of observations for each tree	The proportion of observations for each tree means the proportion of the rows that are chosen randomly from the full set of rows. Too high will increase the over-learning and too low will reduce the performance.
Influence trimming factor	This factor specifies the observations which not being problematic, will not be included in the following sample. This parameter must be well selected. If it is not, it has an unquestionable effect on the performance.
Random percent	This parameter manages the percentage of the sample to use to test the model and to obtain an idea of the performance of the model (so to control over-learning). The higher this percentage is, the less data we have to build the model (It is thus more difficult to build a good model and avoid over-learning).
Smooth minimum spikes	This element averages each error-rate/tree-size value with its neighbouring values (in validation process). Consequently, if one allots the good value to this parameter, this improves its chances of finishing the training of the algorithm in the right place.
Minimum trees in the boost	This prevents the algorithm from stopping too quickly. So this prevents the algorithm from falling into false local minima.

Table 4 - Parameters chosen by genetic algorithm.

Parameters	Parameters
Number of trees (80)	Proportion of observations for each tree (60%)
Depth of individual trees (3)	Influence trimming factor (0.009)
Minimum size node to split (10)	Random percent (19%)
Shrink factor (0.04)	Smooth minimum spikes (5)
Prior probabilities (50%)	Minimum trees in the boost (15)

The values between brackets represent the final values

Figure 1 – Healthcare Data Mining Community Applications

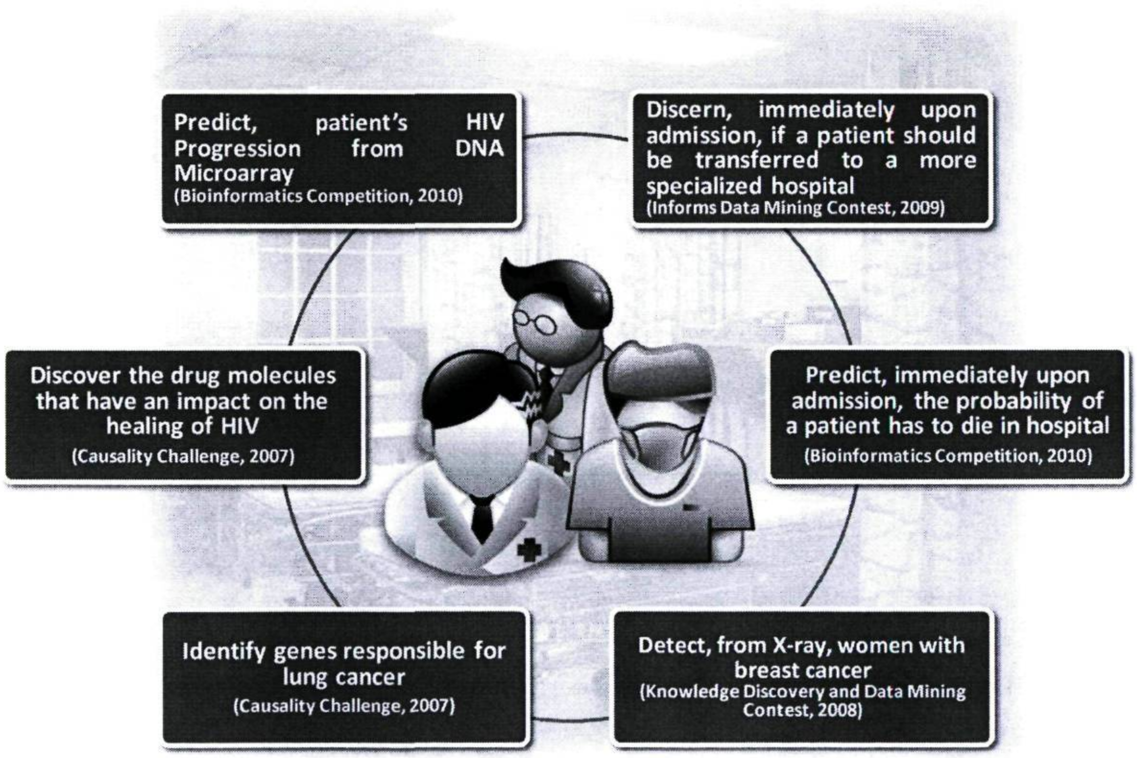


Figure 2 – Data Mining techniques recommendation

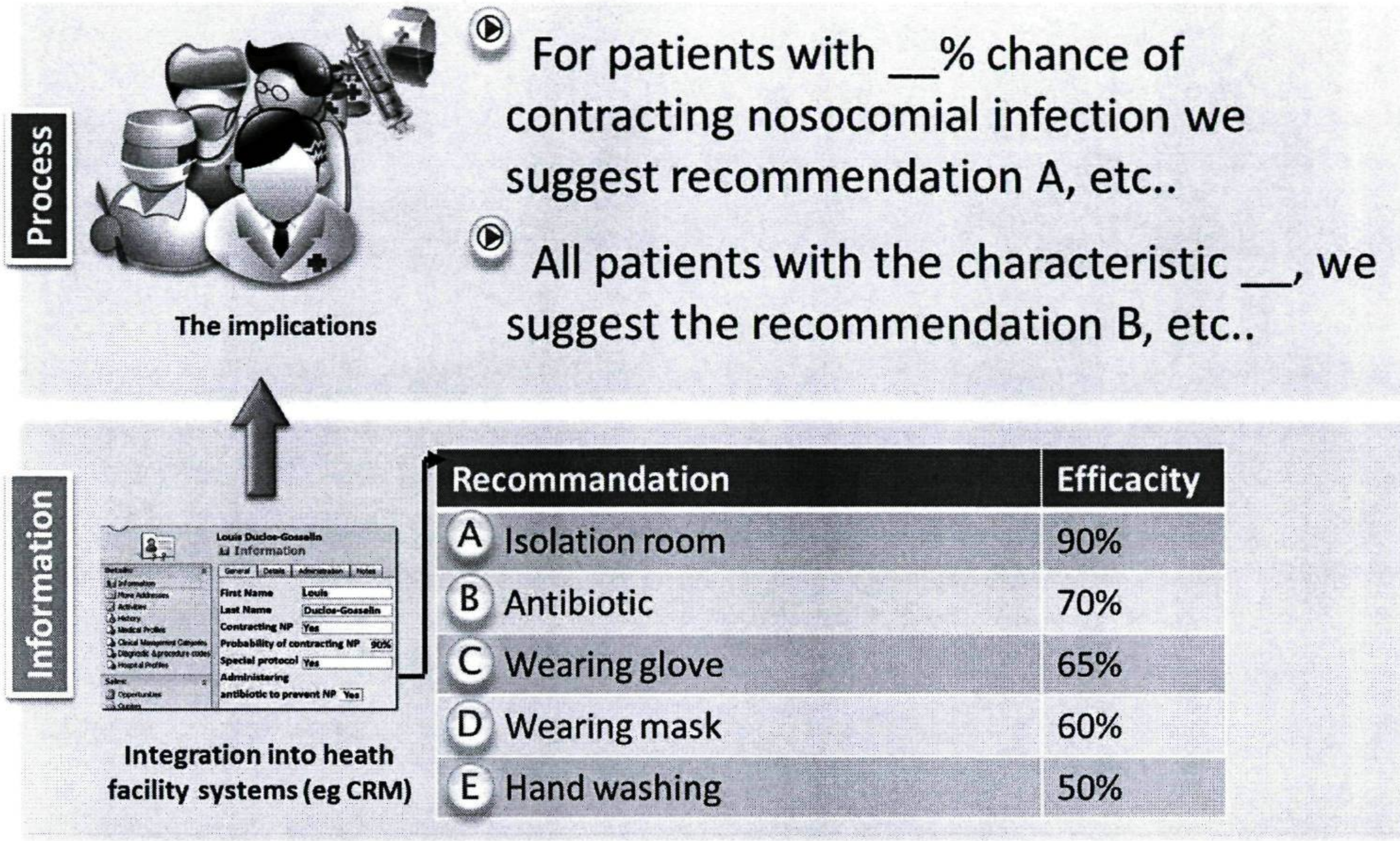


Figure 3 – Data Mining techniques outputs

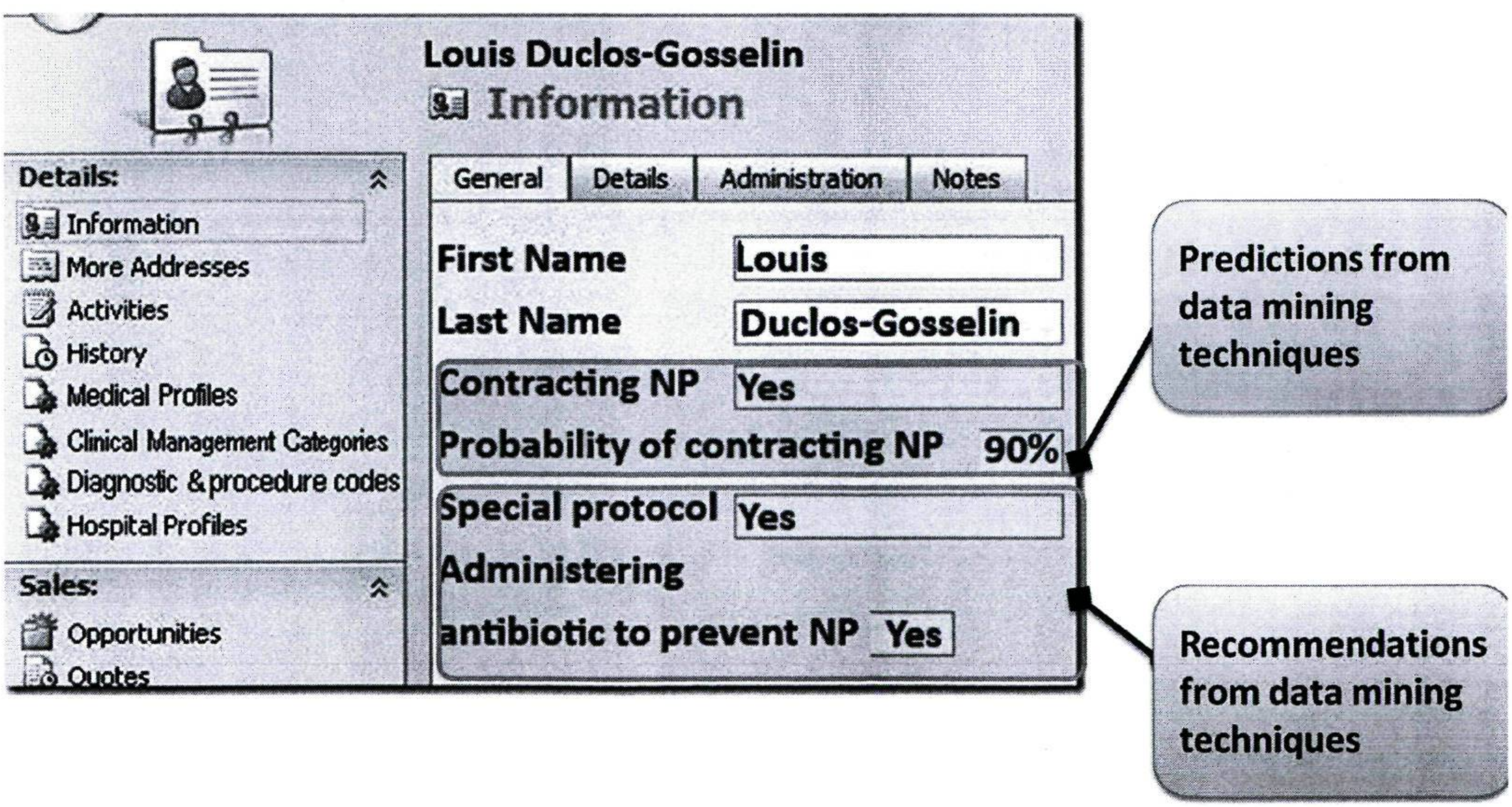


Figure 4 – Data Mining techniques process

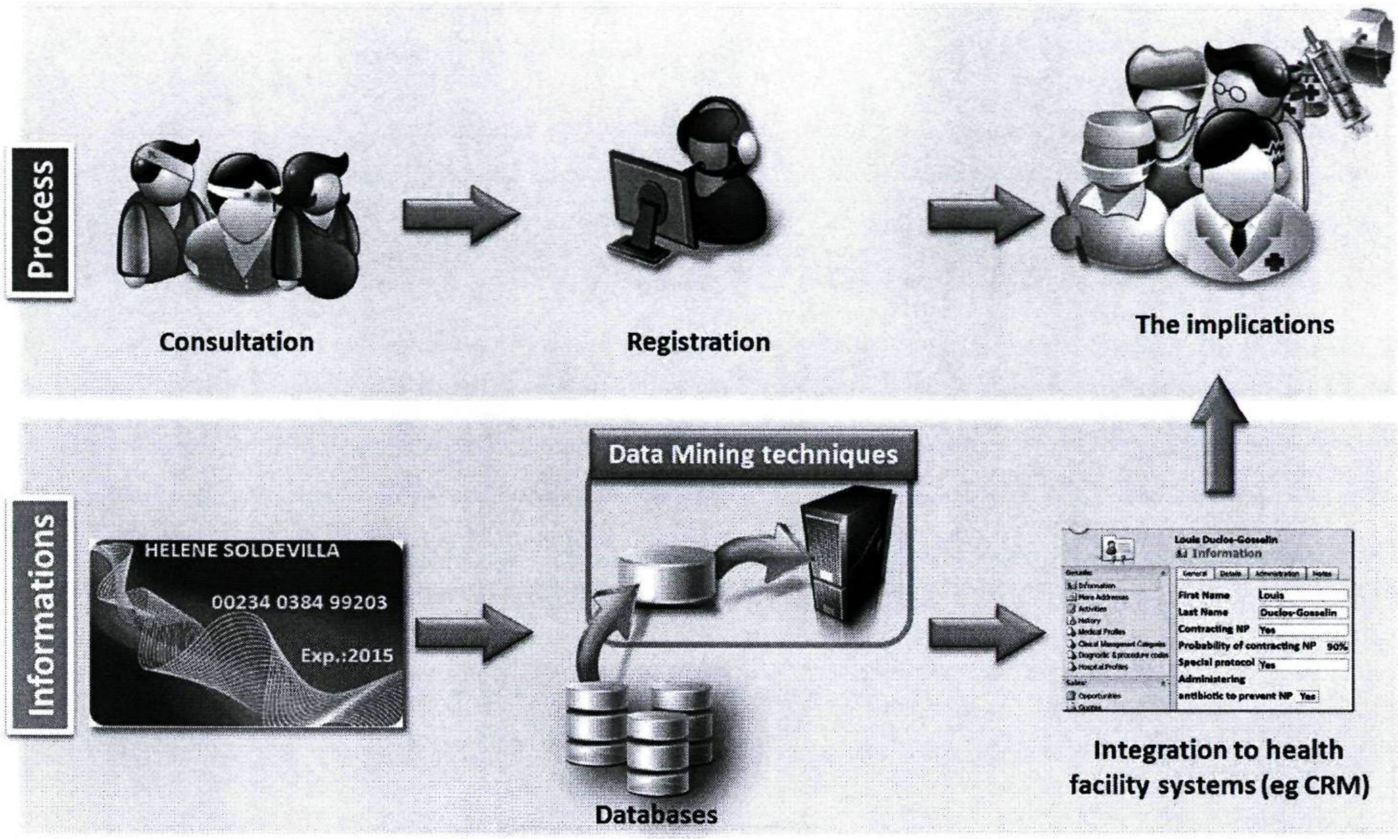


Figure 5 - Relative importance of variables

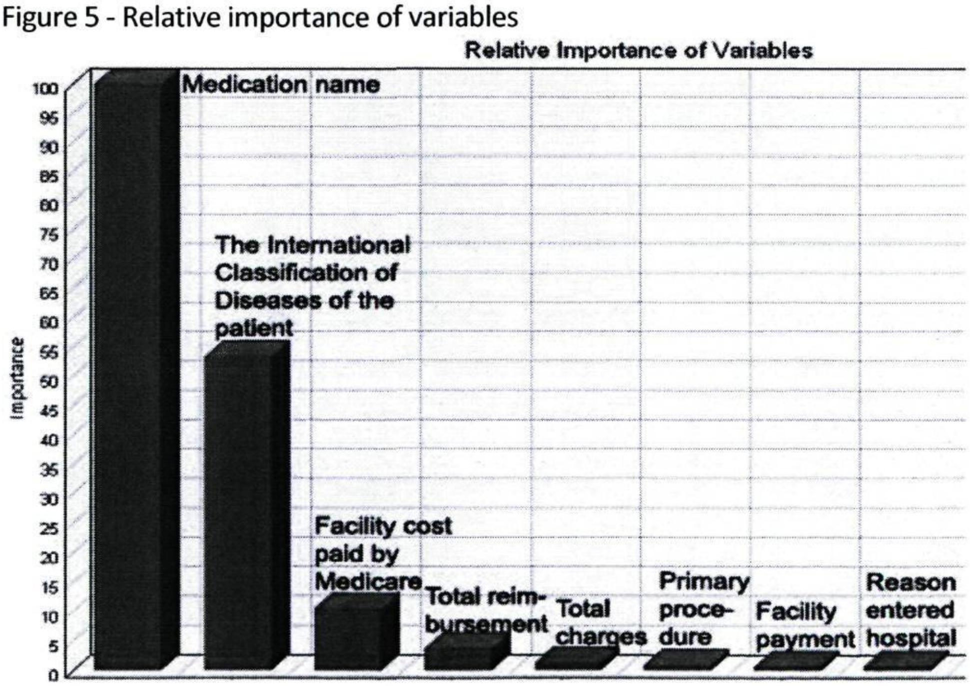


Figure 6 – Building recommendation

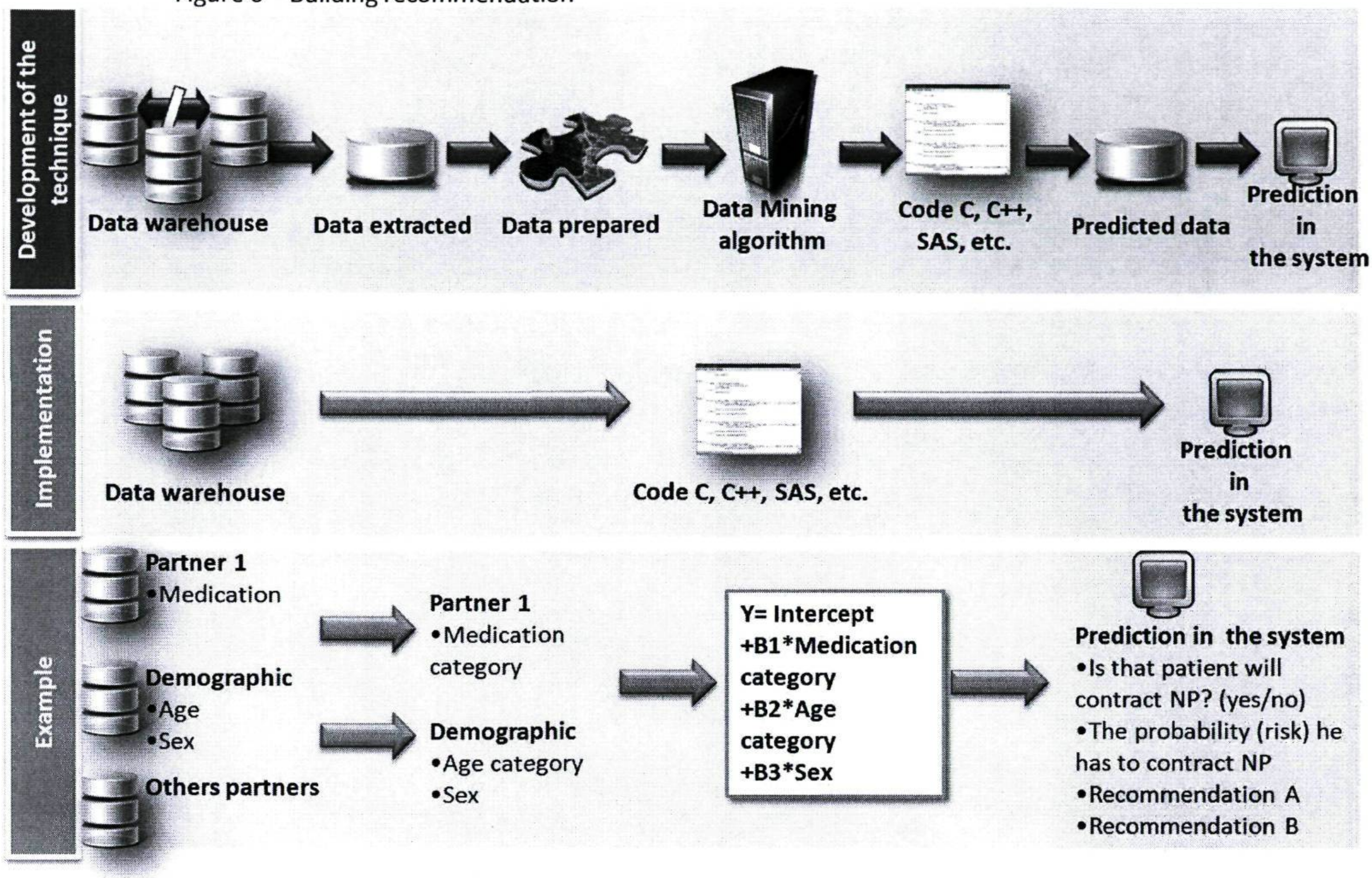


Figure 7 - Model Size and Error Rate

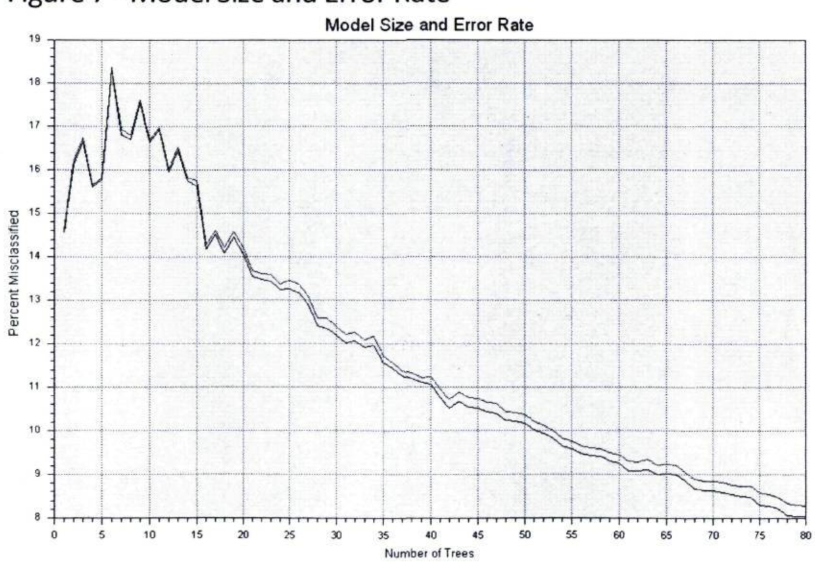


Figure 8 - Area under the curve (A.U.C.)

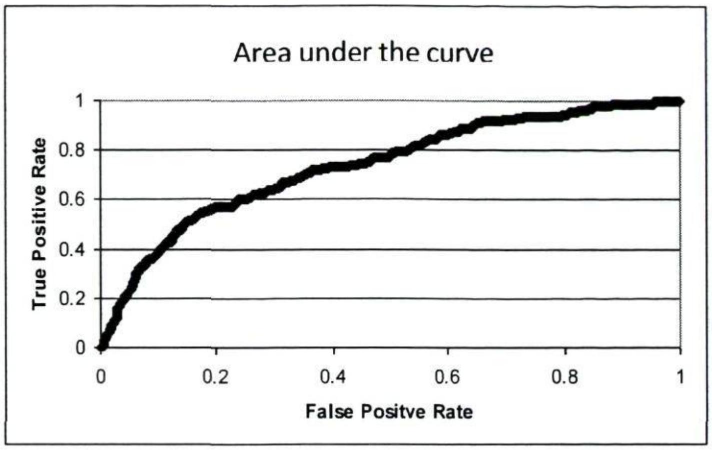


Figure 9 – Recommendation building

Cumulative % of population	Cumulative % of the population which contracted NP
10	**86.17**
20	90.69
30	93.98
40	95.95
50	97.08
60	97.84
70	98.78
80	99.25
90	99.81
100	100

1. The health care managers may decide to include in the patient file the following recommendation when the Data Mining techniques predict that the patient is in this bin: placing the patient in an isolation room. The interpretation of this decision is as follows: in treating 10% of patients admitted to the health care facility with this recommendation (those predicted in this bin), it reached 86.17% of all patients who will contract the infection.

2. Moreover, the health care facility would make the following recommendation in the patient file when the Data Mining techniques predict that the patient is in this bin: an antibiotic. Here, the interpretation of this decision is: treating 20% of patients admitted to the health care facility (those predicted in this bin and in the first bin), it reached 90.69% of all patients who will contract the infection.

3. Without the Data Mining techniques to achieve 90.69% of all patients who will contract the infection, the hospital would have to treat 90.69% of all patients admitted.

Appendix A : The Key Variables for Each Data Type

Patient conditions dataset	**Patient demographic dataset**
The International Classification of Diseases of the patient	Birthday
Patient hospital dataset	Sex
Did the patient stay begin with emergency room visit?	Race
Number of nights stayed at provider	Civil status
Number of nights in hospital	Level of education
Reason for entered at the hospital	Income
Hospital stay related to condition	Level of poverty
VA Facility flag	Are any operations or surgeries performed?
Predicted facility cost paid by Medicaid	Is medicines prescribed at discharge?
Predicted facility cost paid by medicare	**Patient medication dataset**
Predicted facility cost paid by other federal	Payment claim filing information
Predicted facility cost paid by other insurance	Has insulin or diabetic equipment/supply?
Predicted facility cost paid by other public	Use generic medication?
Predicted facility cost paid by other private	Type of pharmacy provenance
Predicted facility cost paid by private insurance	Pregnancy category
Predicted facility cost paid by individual	How much prescription purchased/obtained?
Predicted facility cost paid by state	When the person started taking his medicine
Predicted facility cost paid by Tricare	Prescription form
Predicted facility cost paid by Veteran's Administration	Unit of prescription form
Predicted facility cost paid by Workmen's compensation	Quantity of medication prescript
Predicted total facility cost	Strength of the dose prescript
Facility payment	Past medication name
Primary diagnosis	Doctor identifier
Secondary diagnosis	Therapeutic class of the medication
Predicted physician cost paid by Medicaid	Medication expenditure
Predicted physician cost paid by Medicare	Amount paid by Medicaid
Predicted physician cost paid by other federal	Amount paid by Medicare
Predicted physician cost paid by other insurance	Amount paid by other federal agency
Predicted physician cost paid by other public	Amount paid by other insurance
Predicted physician cost paid by other private	Amount paid by other public agency
Predicted physician cost paid by private	Amount paid by other private insurance

insurance	
Predicted physician cost paid by individual	Amount paid by private insurance
Predicted physician cost paid by state	Amount paid by the individual
Predicted physician cost paid by Tricare	Amount paid by the state
Predicted physician cost paid by Veteran's Administration	Amount paid total
Predicted physician cost paid by Workmen's Compensation	
Predicted total physician cost	
Predicted physician reimbursement	
Predicted primary procedure	
Secondary procedure	
Predicted total charges	
Predicted total reimbursement	
Predicted total charged	
Predicted tricare payment	
VA payment	
Predicted workman's Comp payment	

Conclusion

As we have seen, managers can use mathematical models and predictive analysis to effectively manage operations.

In this MSc research, we tested two different models which specifically address some insurance and health management problems. The mathematics behind this kind of models could also be used to address managerial problems of other activities.

For example, we implemented the predictive analysis model presented in the second paper on financial, fraud, marketing, library, physic, health, lottery, medical, tourism and time series data as demonstrated by the following results of various international competitions:

- 2010 : Tourism Forecasting Contest II (19th place on 257) Tourism Forecasting Problem
-
- 2010 : University of California at San Dieogo Data Mining Contest (25th place on 127 preliminary result) Appetency, Upselling Problem
- 2010 : Tourism Forecasting Contest I (47th place on 257) Tourism Forecasting Problem
- 2010 : IEEE ICDM Contest (13th place on 575) Road Traffic Prediction for Intelligent GPS Navigation Problem
- 2010 : Bioinformatics Competition (67th place on 295) HIV Progression Prediction Problem
- 2010 : Data Mining Cup Contest (56th place on 115) Revenue Maximisation by Intelligent Couponing Problem
- 2010 : Pacific-Asia Knowledge Discovery and Data Mining Competition (39th place on 45) Credit Scoring Problem
- 2010 : RSCTC Discovery Challenge (76th place on 96) DNA Microarray Problem
- 2009 : Informs Data Mining Contest (23th place on 250) Identify Transfers to Tertiary Hospitals Problem
- 2009 : Pacific-Asia Conference on Knowledge Discovery and Data Mining (88th place on 141) Financial Problem

- 2009 : University of California at San Diego Data Mining Contest (88th place on 143) Fraud Problem
- 2009 : Knowledge Discovery and Data Mining Contest (1099th place on 3304) Marketing Problem
- 2009 : Data Mining Cup Contest (20th place on 52) Library Problem
- 2008 : Informs Data Mining Contest (5th place on 8) Health Problem
- 2008 : University of California at San Diego Data Mining Contest (22th place on 41) Physic Problem
- 2008 : Knowledge Discovery and Data Mining Contest (124th place on 246) Medical Problem
- 2008 : Data Mining Cup Contest (114th place on 208) Lottery Problem
- 2007 : World Congress on Computational Intelligence pour le Causality Challenge (20th place on 31) Medical Problem
- 2007 : World Congress on Computational Intelligence pour le Causality Challenge (16th place on 29) Medical Problem
- 2007 : World Congress on Computational Intelligence pour le Causality Challenge (20th place on 31) Marketing Problem
- 2007 : International Joint Conference on Neural Networks (2th place on 793) Marketing Problem
- 2007 : Pacific-Asia Conference on Knowledge Discovery and Data Mining (30th place on 252) Marketing Problem
- 2007 : Data Mining Cup Contest (135th place on 688) Marketing Problem
- 2006 : Artificial Neural Network & Computational Intelligence Forecasting Competition (11th place on 44) Time series Problem

Bibliography

4th Decennial International Conference on Nosocomial and Healthcare-Associated Infections. March 5-9,. 2000, Atlanta, GA.

Avila, Ramon A., Edward F. Fern, and O. Karl Mann (1988), "Unravelling Criteria for Assessing the Performance of Salespeople: A Causal Analysis," *Journal of Personal Selling & Sales Management,* 8 (May), 45-54.

Avlonitis, George J. and Nikolaos G. Panagopoulos (2006), "Role Stress, Attitudes, and Job Outcomes in Business-to-Business Selling: Does the Type of Selling Situation Matter?" *Journal of Personal Selling & Sales Management*, 26 (winter), 67-77.

Barksdale, Hiram C., Jr., Danny N. Bellenger, James S. Boles, and Thomas G. Brashear (2003), "The Impact of Realistic Job Previews and Perceptions of Training on Sales Force Performance and Continuance Commitment: A Longitudinal Test," *Journal of Personal Selling & Sales Management,* 23 (spring), 125-138.

Bartkus, Kenneth R., Mark F. Peterson, and Danny N. Bellenger (1989), "Type A Behavior, Experience, and Salesperson Performance," *Journal of Personal Selling & Sales Management,* 9 (summer), 11-18.

Behrman, Douglas N. and William D. Perreault, Jr. (1982), "Measuring the Performance of Industrial Salespersons," *Journal of Business Research*, 10 (September), 355-369.

---- and ---- (1984), "A Role Stress Model of the Performance and Satisfaction of Industrial Salespersons," *Journal of Marketing*, 48 (fall), 9-21.

Berry, Dick (1986), "Sales Performance: Fact or Fiction?" *Journal of Personal Selling & Sales Management,* 6 (August), 71-79.

Bommer, William H., Johnathan L. Johnson, Gregory A. Rich, Philip M. Podsakoff, and Scott B. MacKennzie (1995), "On the Interchangeability of Objective and Subjective Measures of Employee Performance: A Meta-Analysis," *Personnel Psychology,* 48 (3), 587-605.

Boles, James S., Naveen Donthu, and Ritu Lohtia (1995), "Salesperson Evaluation Using Relative Performance Efficiency: The Application of Data Envelopment Analysis," *Journal of Personal Selling & Sales Management*, 15, 3 (summer), 31-49.

Boulder, C., NeXstar Begins Third Phase II Trial for MiKasome. PR Newswire,1, 1998.

Breiman, L., Manual on setting up, using, and understanding Random Forests v3.1. Technical Report, ftp://ftp.stat.berkeley.edu/pub/users/breiman/Using_random_forests_v3.1.pdf, 2002

Brown, Steven P. and Robert A. Peterson (1993), "Antecedents and Consequences of Salesperson Job Satisfaction: Meta-Analysis and Assessment of Causal Effects," *Journal of Marketing Research*, 30 (February), 63-77.

Brosset, S. E. and al, (1998), "Association Rules and Data Mining in Hospital Infection Control and Public Health Surveillance", *Journal of the American Medical Informatics Association*, (5:4), 373-381

Challagalla, Goutam and Tasadduq Shervani (1996), "Dimensions and Types of Supervisory Controls: Effects on Salesperson Performance and Satisfaction," *Journal of Marketing,* 60 (January), 89-105.

Charnes, Abraham, William W. Cooper, and Edwardo Rhodes (1978), "Measuring the Efficiency of Decision-Making Units", *European Journal of Operational Research* vol. 2, 429-444.

Chonko, Lawrence B., Roy D. Howell, and Danny N. Bellenger (1986), "Congruence in Sales Force Evaluations: Relation to Sales Force Perceptions of Conflict and Ambiguity," *Journal of Personal Selling & Sales Management,* 6 (May), 35-48.

----, Terry W. Low, James A. Roberts, and John F. Tanner (2000), "Sales Performance: Timing of Measurement and Type of Measurement Make a Difference," *Journal of Personal Selling & Sales Management,* 20 (winter), 23-36.

Churchill, Gilbert A., Jr., Neil M. Ford, S. W. Harley, and Orville C. Walker, Jr. (1985), "The Determinants of Salesperson Performance: A Meta-Analysis, *Journal of Marketing Research,* 22 (May), 103-118.

Cohen, Andy (1997), "General Electric," *Sales and Marketing Management,* 149 (October), 57.

Cravens, David W., Thomas N. Ingram, Raymond W. LaForge, and Clifford E. Young (1993), "Behavior-Based and Outcome-Based Salesforce Control Systems, *Journal of Marketing,* 57, 4, 47-59.

----, Raymond W. LaForge, Gregory M. Pickett, and Clifford E. Young (1993), "Incorporating a Quality Improvement Perspective into Measures of Salesperson Performance," *Journal of Personal Selling & Sales Management,* 13 (winter), 1-14.

Cron, William L. and Michael Levy (1987), "Sales Management Performance Evaluation: A Residual Income Perspective," *Journal of Personal Selling & Sales Management,* 7 (August), 57-66.

Crouse, D. T. ,Characterization of Ureaplasma urealyticum pneumonia in human neonates and a murine model, Ph.D. dissertation, The University of Alabama at Birmingham, United States, 1992.

Dalrymple, Douglas J. and William M. Strahle (1990), "Career Path Charting: Framework for Sales Force Evaluation," *Journal of Personal Selling & Sales Management,* 10 (summer), 59-68.

DeCarlo, Thomas E., R. Kenneth Teas, and James C. McElroy (1997), "Salesperson Performance Attribution Processes and the Formation of Expectancy Estimates," *Journal of Personal Selling & Sales Management,* 17 (summer), 1-17

Dempster, A., Laird, N. and Rubin, D.. 1977, Likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society, Series B 39(1),1–38, 1977.

Dubinsky, Alan J., Steven J. Skinner, and Tommy E. Whittler (1989), "Evaluating Sales Personnel: An Attribution Theory Perspective," *Journal of Personal Selling & Sales Management,* 9 (Spring), 9-21.

Edwards, Mark R., W. Theodore Cummings, and John L. Schlacter (1984), "The Paris-Peoria Solution: Innovations in Appraising Regional and International Sales Personnel," *Journal of Personal Selling & Sales Management,* 4 (November), 26-38.

Farley, John U. (1964), "An Optimal Plan for Salesmen's Compensation," *Journal of Marketing Research,* 1, 39-44.

Fauci, A. S., Braunwaldm, E. , Kasper, D. L. , Hauser, S. L., Longo D. L. and Jameson, J. L., Harrison's Principles of Internal Medicine (17th Edition, McGraw- Hill, 2008).

Fitzpatrick, M. A. Using data to drive performance improvement in hospitals, Health Management Technology 27, 12, 2006

Flaherty, Theresa B., Robert Dahlstrom, and Steven J. Skinner (1999), "Organizational Values and Role Stress as Determinants of Customer-Oriented Selling Performance," *Journal of Personal Selling & Sales Management,* 19 (Spring), 1-18.

Friedman, J. H., Greedy function approximation: A gradient boosting machine, Technical report, Dept. of Statistics, Stanford University, United States, 1999a.

Friedman, J. H., Stochastic gradient boosting, Technical report, Dept. of Statistics, Stanford University, United States, 1999b.

Friedman, J. H., Hastie T. and Tibshirani, R., The elements of statistical learning, Data mining Inference and Prediction (Springer, 2001).

Gentry, James W., John C. Mowen, and Lori Tasaki (1991), "Salesperson Evaluation: A Systematic Structure for Reducing Judgmental Biases," *Journal of Personal Selling & Sales Management,* 11 (Spring), 27-38.

Giacobbe, Ralph W., Donald W. Jackson, Jr., Lawrence A. Crosby, and Claudia M. Bridges (2006), "A Contingency Approach to Adaptive Selling Behavior and Sales Performance: Selling Situations and Salesperson Characteristics," *Journal of Personal Selling & Sales Management*, 26 (spring), 115-142.

Gerald, L. , Mandell, Bennett, J. E. and Dolin, R., Mandell's Principles and Practices of Infection Diseases (6th Edition, Churchill Livingstone, 2004).

Harper, P. Combining data mining tools with health care models for improved understanding, Clinical and Investigative Medicine 28, 6, 2005.

Harris, C. Uncovering data, Canadian Underwriter 74, 2, 2007.

Harris, J. R., Predictors of nosocomial pneumonia in critically ill trauma patients, Ph.D. dissertation, University of Maryland, Baltimore, United States, 1996.

Haupt, R. L. and Haupt, E. S., Practical genetic algorithms (Wiley, 2004).

He, Y., Missing data imputation for tree-based models, Ph.D. dissertation, University of California, Los Angeles, United States, 2006.

Heneman, Herbert G. (1986), "The Relationship between Supervisory Ratings and Result-Oriented Measures of Performance: A Met-Analysis," *Personnel Psychology,* 39 (winter), 11-826.

Hunter, Gary K. and William D. Perreault, Jr. (2006), "Sales Technology Orientation, Information Effectiveness, and Sales Performance," *Journal of Personal Selling & Sales Management*, 26 (spring), 95-113.

Hyman, Michael R. and Jeffrey S. Sager (1999), "Marginally Performing Salespeople: A Definition," *Journal of Personal Selling & Sales Management,* 19 (fall), 67-74.

Jackson, Donald W., Jr., Janet E. Keith, and John L. Schlacter (1983), "Evaluation of Selling Performance: A Study of Current Practices," *Journal of Personal Selling & Sales Management,* 3 (November), 42-51.

Jaramillo, Fernando, François A. Carrillat, and William B. Locander (2004), "Response to Comments: Starting to Solve the Method Puzzle in Salesperson Self-Report Evaluation," *Journal of Personal Selling & Sales Management,* 24 (spring), 141-155.

----, ----, and ---- (2005), "A Meta-Analytic Comparison of Managerial Ratings and Self-Evaluations," *Journal of Personal Selling & Sales Management,* 25 (fall), 315-328.

----, Jay Prakash Mulki, and Paul Solomon (2006), "The Role of Ethical Climate on Salesperson's Role Stress, Job Attitudes, Turnover intention, and Job Performance," *Journal of Personal Selling & Sales Management,* 26 (summer), 271-282.

Jaworski, Bernard J. (1988), "Toward a Theory of Marketing Control: Environmental Context, Control Types, and Consequences," *Journal of Marketing*, 52 (July), 23-39.

Jobber, David, Graham J. Hooley, and David Shipley (1993), "Organizational Size and Salesforce Evaluation Practices," *Journal of Personal Selling & Sales Management,* 13 (spring), 37-48.

Krishnan, Balaji, C., Richard G. Netemeyer, and James S. Boles (2002), "Self-Efficacy, Competitiveness, and Effort as Antecedents of Salesperson Performance," *Journal of Personal Selling & Sales Management,* 22 (fall), 285-295.

Liu, W. Z., Techniques for dealing with missing values in classification, Advances in Intelligent Data Analysis Reasoning about Data, 527-536, 2006.

MacKenzie, Scott B., Philipp M. Podsakoff, and Michael Aherne (1998), "Some Possible Antecedents and Consequences of In-Role and Extra-Role Salesperson Performance," *Journal of Marketing*, 62 (July), 87-98.

Maimon, O. and Rokach, L., The Data Mining and Knowledge Discovery Handbook (Springer, 2005).

Marshall, Greg W. and John C. Mowen (1993), "An Experimental Investigation of the Outcome Bias in Salesperson Performance Evaluations," *Journal of Personal Selling & Sales Management,* 13 (summer), 31-47.

----, ----, and Keith J. Fabes (1992), "The Impact of Territory Difficulty and Self Versus Other Ratings on Managerial Evaluations of Sales Personnel," *Journal of Personal Selling & Sales Management,* 12 (fall), 35-47.

Mauricio, C.-R. (2010), "Identification of Potential Cancer Biomarkers Using Multiple Criteria
Optimization in Microarray Experiments", Proceedings of the 5th INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2010) Austin TX, November 2010

McAdams, Jerry (1987), "Rewarding Sales and Marketing Performance," *Management Review,* (April), 33-38.

McFarland, Richard G. and_Blair Kidwell (2006), "An Examination of Instrumental and Expressive Traits on Performance: The Mediating Role of Learning, Prove, and Avoid Goal Orientation," *Journal of Personal Selling & Sales Management*, 26 (spring), 143-159.

McKay, Sandra, Joseph F. Hair, Jr., Mark W. Johnston, and Daniel L. Sherrell (1991), "An Exploratory Investigation of Reward and Corrective Responses to Salesperson Performance: An Attributional Approach," *Journal of Personal Selling & Sales Management,* 11 (spring), 39-48.

Michael, G. M. (2010), "A Robust Hierarchical Model for Schizophrenia Detection in High Dimensional fMRI", Proceedings of the 5th INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2010) Austin TX, November 2010

Milley, A. (2000), "Healthcare and data mining", Health Management Technology, 21 (8), 44-47

Ministère de la santé et des services sociaux (2006), "Plan d'action sur la prévention et le contrôle des infections nosocomiales 2006-2009"

Ministère de la santé et des services sociaux (2008), "Rapport dur comité d'examen sur la prévention et le contrôle des infections nosocomiales"

Mitchell, M., An introduction to genetic algorithms (M.I.T. Press, 1999).

Monson, K. and Schoenstadt, A., Seroquel Warnings And Precautions, http://bipolar-disorder.emedtv.com/seroquel/seroquel-warnings-and-precautions.html, 2007.

Morris, Michael H., Duane L. Davis, Jeffrey W. Allen, Ramon A. Avila, and Joseph Chapman (1991), "Assessing the Relationships Among Performance Measures, Managerial Practices, and Satisfaction When Evaluating the Salesforce: A Replication and Extension," *Journal of Personal Selling & Sales Management,* 11 (summer), 25-35.

Mosley, D. , 2005, Detecting a pattern, Best's Review 106, 1, 2005.

Mowen, John C., Stephen W. Brown, and Donald W. Jackson, Jr. (1980), "Cognitive Biases in Sales Management Evaluations," *Journal of Personal Selling & Sales Management,* 1 (fall/winter), 83-89.

----, Fabes, Keith J. and Raymond W. LaForge (1986), "Effects of Effort, Territory Situation, and Rater on Salesperson Evaluation," *Journal of Personal Selling & Sales Management,* 6 (May), 1-8.

Muczyk, Jan P. and Myron Gable (1987), "Managing Sales Performance through a Comprehensive Performance Appraisal System," *Journal of Personal Selling & Sales Management,* 7 (May), 41-52.

Oliver, Richard L. and Erin Anderson (1994), "An Empirical Test of the Consequences of Behavior- and Outcome-Based Sales Control Systems," *Journal of Marketing,* 58 4, 53-67.

---- and ---- (1995), "Behavior- and Outcome-Based Sales Control Systems: Evidence and Consequences of Pure-Form and Hybrid Governance," *Journal of Personal Selling & Sales Management,* 15 (fall), 1-16.

Ouchi, William G. and M. Mcguire (1975), "Organizational Control: Two Functions," *Administrative Science Quarterly,* 20 (December), 559-569.

Parsons, Leonard M. (1992), "Productivity versus Relative Efficiency in Marketing: Past and Future?" In Gary Lilien, Gilles Laurent, and Bernard Pras (Eds.), *Research Traditions in Marketing*. Kluwer, Amsterdam, 169-196.

Patel, N. R., Data Mining: Finding the Data and Models That Create Value: Comparison of Data Mining techniques, MIT Statistics course (15.062), 2003.

Patton, W. E., III and Ronald H. King (1985), "The Use of Human Judgment Models in Evaluating Sales Force Performance," *Journal of Personal Selling & Sales Management,* 5 (May), 1-14.

Perlich, C. and Provost, F. , ACORA: Distribution-based aggregation for relational learning from identifier attributes, Special Issue on Statistical Relational Learning and Multi-Relational Data Mining, Journal of Machine Learning 62, 65-105, 2006.

Pettijohn, Charles E., Linda S. Pettijohn, and Albert J. Taylor (2000), "An Exploratory Analysis of Salesperson Perceptions of the Criteria Used in Performance Appraisals, Job Satisfaction, and Organizational Commitment," *Journal of Personal Selling & Sales Management,* 20 (spring), 77-80.

Pilling, Bruce K., Naveen Donthu, and Steve Henson (1999), "Accounting for the Impact of Territory Characteristics on Sales Performance: Relative Efficiency as a Measure of

Salesperson Performance," *Journal of Personal Selling & Sales Management,* 19 (spring), 35-45.

Plank, Richard E. and David A. Reid (1994), "The Mediating Role of Sales Behavior: An Alternative Perspective of Sales Performance and Effectiveness," *Journal of Personal Selling & Sales Management*, 14 (3), 43-56.

---- and J. N. Greene (1996), "Personal Construct Psychology and Personal Selling Performance," *European Journal of Marketing* 30, 7, 25-48.

Rasim, M. (2010), "Data Mining Methods to Investigate Medicare Fraud Within Unsupervised Databases", Proceedings of the 5th INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2010) Austin TX, November 2010

Rich, Gregory A., William H. Bommer, Scott B. MacKenzie, Philip M. Podsakoff, and Johnathan L. Johnson (1999), "Methods in Sales Research: Apples and Apples or Apples and Oranges? A Meta-Analysis of Objective and Subjective Measures of Salesperson Performance," *Journal of Personal Selling & Sales Management*, 19 (fall), 41-52.

Rom, J. F., Using the EM algorithm to replace missing observations in multiple linear regression: A Monte Carlo study, Ph.D. dissertation, Cornell University, United States, 1992.

Roosevelt, M., Detecting a pattern, Best's Review, 106, 1, 2005.

Schleder, B. J., Taking charge of ventilator-associated pneumonia, Nursing Management 34(8), 27-32, 2003.

Sharma, Arun, Gregory A. Rich, and Michael Levy (2004), "Comment: Starting to Solve the Method Puzzle in Salesperson Self-Report Evaluations," *Journal of Personal Selling & Sales Management*, 24 (spring), 135-139.

---- and Dan Sarel (1995), "The Impact of Customer Satisfaction Based Incentive Systems on Salespeople's Customer Service Response: An Empirical Study," *Journal of Personal Selling & Sales Management*, (summer), 17-29.

Sherwood, M. T., Individual risk rating (Foundations of Casualty Actuarial Science, 4th Edition, Casualty Actuarial Society, Chapter 4, 2001).

Silver, Lawrence S., Sean Dwyer, and Bruce Alford (2006), "Learning and Performance Goal Orientation of Salespeople Revisited: The Role of Performance-Approach and Performance-Avoidance Orientations," *Journal of Personal Selling & Sales Management*, 26 (winter), 27-38.

Yeo, A. C., A mathematical programming approach to optimise insurance premium pricing within a data mining framework, Journal of the Operational Research Society 53, 1197–1203, 2002.

Vason, B. J. Mine data to discover infection control trends, Nursing Management 35, 6, 2004.

Walker, Orville C., Gilbert A. Churchill, and Neil M. Ford (1977), "Motivation and Performance in Industrial Selling: Present Knowledge and Needed Research," *Journal of Marketing Research*, 14 (May), 156-158.

Warrell, D. A., Cox , T. M. and John, D., The Oxford Textbook of Medicine (4th Edition, Oxford University Press, 2003).

WebMD, Rituximab, http://www.webmd.com/drugs/drug-5208-Rituximab+IV.aspx?drugid=5 208&drugname=Rituximab+IV, 2003.

Weitz, Barton A. (1981), "Effectiveness in Sales Interactions: A Contingency Framework," *Journal of Marketing*, 45, (January), 85-104.

----, Harish Sujan, and Mita Sujan (1986), "Knowledge, Motivation, and Adaptive Behavior: A Framework for Improving Selling Effectiveness, *Journal of Marketing*, 50 (September), 174-191.

Notes

[1] All the selling units are assumed to make at least some profit in T_0 and T_1. In some instances, however, it is possible to observe $\pi_1 < -1$ whenever a loss rather than a profit occurs in either T_0 or T_1. In such cases, a profit (or loss) percentage variation is meaningless. Whenever this situation arises, the problem can be handled in different ways: (1) excluding such selling units from this analysis and dealing with them separately; (2) If the loss is small and affects a relatively small number of selling units, the profit variation can be approximated by the -1 value (this would slightly overestimate these selling units' market performance); (3) the reference period could be changed (or replaced by objectives) in order to suppress the problem; and (4) the same profit amount could be added to the profit results in T_0 and T_1 for all the selling units, in such a way as to suppress the loss for the largest losing selling unit.

[2] Like in DEA, every selling unit market performance could be related to the "best practice".