A Decision Tree Application in Tourism-based Regional Economic Development

Curtis, Panayiotis G. and Kokotos, Dimitris X.

Technological Institute of Chalkida, University of Piraeus

18 May 2008

A DECISION TREE APPLICATION IN TOURISM BASED REGIONAL ECONOMIC DEVELOPMENT

Panayiotis G.Curtis
Technological Institute of Chalkida

Dimitris X. Kokotos
University of Piraeus

The reasons for the deficient performance of Ionian Islands tourism in terms of in coming revenues expressed in USD and return on equity capital invested on hotels is explored. The answers to the questioner of management of the hotels surveyed were analyzed with the use of the Decision Tree tool. The issue of competitiveness of the tourism product was assessed. The development of alternative forms of tourism is proposed as a means of improving competitiveness and restoring sustainability in the sector.

Keywords: *Decision trees, tourism, strategy, regional economic development*

INTRODUCTION

Ionian islands is a well known tourism destination of Greece. It is a region that its GDP is approximately 30 % lower than the corresponding average one for the European Union. The plans for regional economic development of Ionian are primarily based on tourism which is the most important sector of the local economy. These plans aim to close the gap with the other regions of the Union.

The consistently inferior return to Equity capital achieved by the hotel sector (the main pillar of tourism) on Ionian islands, compared to the one of the entire country of Greece for the period 1993-2004 is examined, in an attempt to explain sector's economic performance. Inferior revenue development and lower (negative in the case of Ionian islands) return on equity indicate that the strategy followed is not sustainable, since the cost of equity is not covered by the return on it (Curtis, 2004). Superior return on capital can't be achieved and sustained

without competitive operation (Porter, 1996 ; Grant, 2002). Therefore the present strategy of tourism in the region reduces its share in the global tourism and destroys the wealth, since the opportunity cost of capital is greater than its return. The tourism industry in Ionian demonstrate rates of revenue development and profitability, that according to Marakon matrix explain the problem reflected in financial statements (Hax and Majuf, 1996).

The issue of underperformance is explored through the appropriate questioner among the management staff of hotels companies that represent the main pillar of the tourist sector of the region's economy. Among other tools used towards that aim, data mining technique was also applied (Curtis, 2004).

Data mining is a powerful new technology, which adds value across a broad spectrum of industries. It is defined as the process of extracting valid, previously unknown and ultimately comprehensive information, from large databases and using it to make critical decisions (Hand et al, 2001). In the current and emerging environment of massive data availability and collection, powerful multiprocessing computers and data mining algorithms, data mining is considered as one of the emerging technologies. The importance of data mining has been recognized by information intensive industries such as production and logistics, banking, health, insurance, marketing, retail telecommunication and tourism.

A decision tree is a model/ tool employed in data mining, other models being neural networks, rule induction, genetic algorithms, statistical inference, data visualization etc. Amongst them, decision trees and neural networks are the most widely utilized models in data mining applications. The major advantage of decision trees over neural networks is that they are faster in processing and presenting more comprehensible output/results.

The structure of this is paper is as follows: Section 2 presents the predictive variables of the model for the data set. Section 3 presents a short overview of classification tree theory, describes the data preparation and presents the final classification tree structure. Section 4 discusses the results and strategy for the competitiveness of the Ionian Islands tourism. Section 5 gives a summary of the paper and discusses issues concerning potential applications based on the proposed classification tree

THE PREDICTIVE VARIABLES AND THE DATA SET

Original statistical methods are the basic data analysis tools used so far for the development of the applications. In this paper we present a

classification tree application for answering to the questions concerning the appropriateness of the present strategy given the conditions of internal and external environment of tourism. Based on the decision tree tool was made possible to examine the influence of a number of factors (Hand et al, 2001) towards the direction of improving the competitiveness of the sector of tourism in the region. This particular approach of using classification trees has not been investigated in depth in previous research efforts and applications.

The issue of strategy towards competitiveness, its appropriateness and its improvement were examined, as well as major sources of competitive advantage that are either the lower cost (economic effectiveness) or the product differentiation (Porter, 1985).

In order to build an explanatory model a preparation phase of this study identified a number of factors. In a preliminary stage of the analysis, a number of factors used as predictive variables, have been identified. For these factors, a data set has been generated from an existing large dataset (Curtis, 2004).

The chosen variables include the dependent and the independent ones. "Competitiveness of tourism in Ionian islands compared to the one of other countries" is the depended variable for the analysis and it is defined as a dichotomous variable accepting values "yes" and "no". Competitiveness examined with respect to low cost or differentiation (quality) strategy. Low cost strategy was not considered as viable in the future since tastes change, create threats for Ionian mass tourism for summer vacations. It needs differentiation through the creation of alternative forms of tourism which posses a significant market share, exhibit higher rate of annual growth and the region holds strengths (resources and competencies) in their development (conventions, sea and agrotourism).

The predictive variables used are:

- "The quality of services offered compared to Italy". The values for this variable are: lower, higher and the same quality.
- "The development of alternative forms for tourism". The main values for this variable are: congressional, sea tourism and agro tourism.
- The values for variable "Protection of the Environment" are: existence of biological cleaning, hotels connected to sewerage system or nothing from the two

We should note that the question with respect to quality of services offered in Ionian islands, explored in comparison to other comparable tourist counties. The values for this variable were Italy, Spain and

Turkey. The quality of services offered in the region compared to these countries was characterized as lower, higher and the same. The comparison to Spain and Turkey was considered as less important compared to the Italy one according to the answers given and is not presented on the tree.

The chosen method of classification tree was compatible to form of the data and depends on the structure of this data files.

DECISION TREE METHODOLOGY

The Classification Tree is data mining technique for predicting membership of cases in classes that are defined by a dependent variable usually of categorical type. Each case is measured along a number of predictor variables. The implementation of a classification tree is achieved through a training process (induction) in which a specific algorithm is applied to a sample data set composed by the predictor variables. A typical induction algorithm works in two phases: the splitting phase and the pruning phase. The splitting phase is an iterative top-down process that expands the tree by defining nodes connected with branches. The nodes at the end of the braches are called leafs. The first node at the top of the tree is the root node. At every node, the splitting algorithm creates new nodes by selecting a predictor variable so that the resulting nodes to be as far as possible from each other.

There are many different algorithms in the literature for expanding classification trees and values for initial parameters have been tested (Kokotos & Smirlis, 2005). The more popular induction algorithm is C&RT. A short description of Classification and Regression tree algorithm is that generates only binary trees. It constructs the tree by examining at each node, all possible splits for each predictor variable and uses the goodness of fit measurement criterion to find the best split. It assumes scale, ordinal or nominal types in the predictor and dependent variables.

The distance measurement used for the splitting is much depended on the specific splitting algorithm. In this paper, because of categorical dependent variable, we used "gini " from specification of algorithm like impurity measure.

One important feature of the splitting algorithm refers to the ability of the algorithm to look forward in the tree in order to examine if another combination of splitting could produce better overall classification result. The nodes at the lowest part a branch, that cannot be split further in other

nodes due to the fact that they contain cases with only one outcome, are called pure leafs.

The splitting phase terminates when a stopping rule, initially selected by the user, is satisfied. Stopping rules may include the maximum number of nodes, the number of variables in a node considered for splitting, a minimum number of cases per node etc.
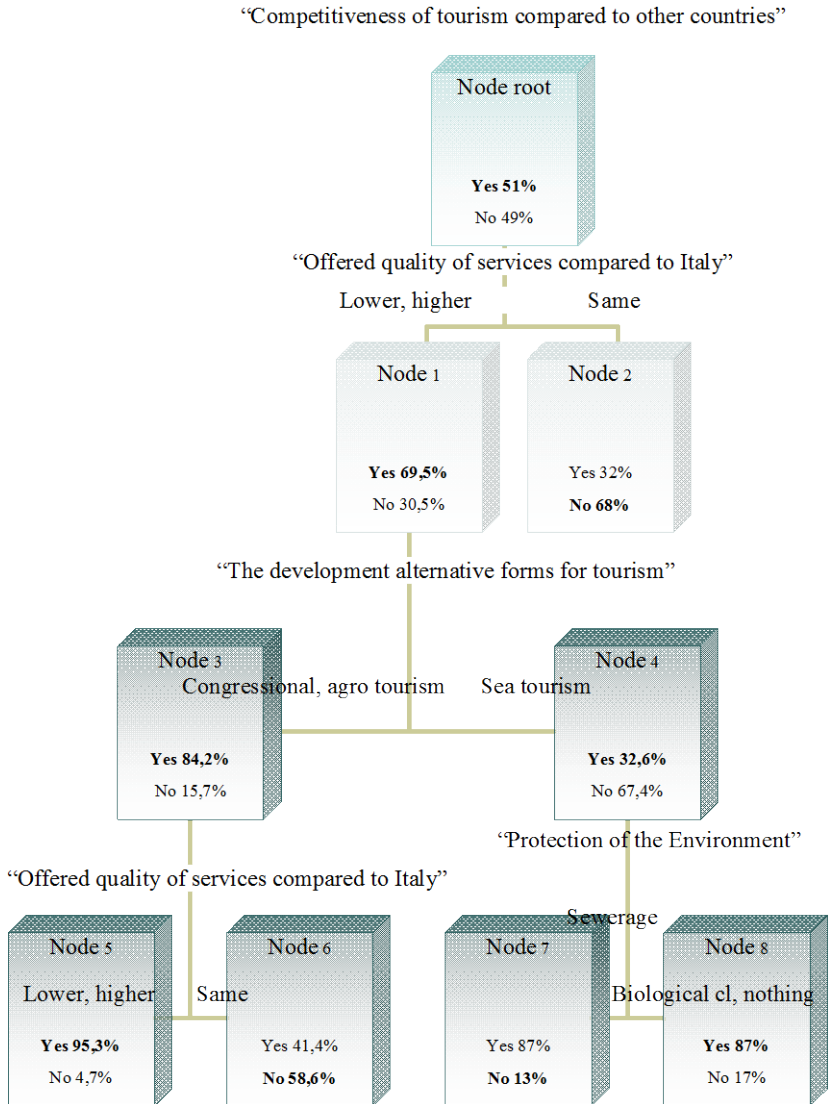
We have chosen as maximum tree depth 3 levels below root. Also, as minimum number 100 cases for parent node and 50 cases for child node. After the development of the structure of the tree, pruning is sometimes required, to make the tree more applicable to other similar data sets or to exclude nodes that seem to be inappropriate for the specific data set or the application. Pruning was not required in these data sets.

To ensure that the tree will perform as good as in the training sample, a validation procedure can be applied. The most preferred type of validation is the testing with a sample data set, taken from the original data set, especially when this data set is large enough. The size of the sample can be approximately one third or a half of the learning data set . When no sample data set can be available, the validation could be on subsets of the original learning sample.

In all the tests carried out in this analysis, the classification tree algorithms were applied in a sample training set of 66 % of the total number of cases. The rest 34 % were considered as the test data set used for validation.

In the case when the misclassification costs are not close enough to costs of the learning sample, the size and the splitting of tree have to be reconsidered. In this paper there was not problem with costs, thus the misclassification costs was chosen equal for all categories.

The following figure presents the final tree structure produced by the Classification and Regression tree algorithm after the specification adjustments during the training phase. Each node is identified by the node number and the percentage of cases included in the node for the classes.

"Competitiveness of tourism compared to other countries"

Node root

Yes 51%

No 49%

"Offered quality of services compared to Italy"

Lower, higher   Same

Node 1

Yes 69,5%

No 30,5%

Node 2

Yes 32%

No 68%

"The development alternative forms for tourism"

Node 3

Congressional, agro tourism   Sea tourism

Yes 84,2%

No 15,7%

Node 4

Yes 32,6%

No 67,4%

"Protection of the Environment"

"Offered quality of services compared to Italy"

Sewerage

Node 5

Lower, higher   Same

Yes 95,3%

No 4,7%

Node 6

Yes 41,4%

No 58,6%

Node 7

Yes 87%

No 13%

Node 8

Biological cl, nothing

Yes 87%

No 17%

Classification trees are directly compared to other traditional statistical methods, such as Regression, as they can classify cases depended on classes defined by a depended variable, but it has less stringent requirements as it assumes no linearity of relationship between

the independent variables and the dependent and no normally distributed variables.

STRATEGY ASSESSMENT FOR THE COMPETITIVENESS OF THE IONIAN ISLANDS TOURISM

Based on the tree constructed, the corresponding rules were formed which contribute to combine many seemingly unrelated at first glance data in order to produce conclusions related to issues under scrutiny that are of interest in determining the competitiveness of Ionian island tourism product.

The decision tree methodology revealed that the operational effectiveness (Porter, 1996) and the development of thematic forms of tourism (product differentiation) are the decisive factors to resolve the problems that Ionian tourism is confronted. It is reflected in low spending per incoming tourist from abroad (expressed in US $) and negative average return on capital for the period 1993-2004 (Curtis, 2004). The only viable growth in the long run is the profitable one that is characterized by return to equity capital greater than the cost of capital employed (Hax & Majluf, 1996). Product differentiation primarily combined with operational effectiveness are deem as necessary to restore competitiveness, profitability and viability in the long run.

All the above policy prescriptions are in accord with the view of the respondents that the tourism product of the region is reaching the maturity stage of its cycle and slow growth is expected for the future. So, it is imperative that the product be diversified and upgraded. At the same time the operational effectiveness has to be bolstered so that the net profit margin is adequate to support an acceptable return to equity with respect to capital employed in the sector (Curtis, 2004).

The stage of the product cycle that mass tourism for summer vacation is going through in Ionian islands, renders prices a critical success factor (CSF). To the extend that the tourism product of the region still remains undifferentiated, high price elasticity of demand and the competitive pressures from new destinations, as well as the existing ones, are increasing. Scale utilization, idle capacity utilization, cost control, efficient use of labour, reduction in transportation cost, reduction in seasonality are critical factors of success. None of these factors, is favorable according to opinion of the management staff of the hotels operating in the region.

Even more, development of the existing product is not proposed without reservations due to a) the change in the structure of demand and

tastes, b) the increasing competition, c) the falling prices, d) the increase in costs, and e) the unbearable (in same cases) dependence on tourist operators.The later take advantage of the situation asking for lower prices, while at the same time exploit the last minute booking tactics to create additional uncertainty and anxiety in the industry, which can not store its excess production capacity that is lost due to the insufficient demand, and so exert more pressures on prices.

Market penetration and development is proposed initially, but since cost advantage does not exist, the carrying capacity of the islands involved is exhausted and pollution threats on the coasts are probable, product development and diversification in to new segments (and markets), with greater prospects that match resources and capabilities of the region, must be sought. The external environment, industry conditions, strategy and positioning in tourism define the critical factors necessary for the success of the hotel sector in the future which do not secure successful low cost competitive advantage type of strategy since the region does not posses cost advantage any more compare to existing and new destinations in central and east Europe and Mediterranean sea. The absence of a) economies of scale, b) low personnel cost, c) efficiency in operations and d) adequate capacity utilization, which are the critical success factors necessary for the success of the existing strategy, do not permit the continuation of it. Instead product differentiation and enrichment is proposed as the appropriate strategy.

The existing "stuck in the middle" cloudy and vague strategy in Ionian islands is responsible to a great extend for the unsatisfactory performance in terms of profitability, since it reflects the lack of specific strategy and it is dangerous as a procedure because it hampers the focusing on the market and allows relaxing as opposed to intense *(Valuebasedmanagement.net).*

CONCLUSIONS

Ionian's hotel sector return on equity is below the average one for the entire country of Greece reflecting "stuck in the middle" strategy. Ionian hotels do not earn attractive profits, since the structure of the mass tourism industry is not favorable due to the harmful operation of the "five forces", which engender fierce competition in the sector. Based on the data from the answers to the questioner, the existing strategy towards competitiveness, its appropriateness and its improvement were examined using the Classification and Regression tree algorithm. The decision tree technique explains convincingly and effectively, that underperformance is

attributed to vagueness of the present course of tourism strategy applied in the region. According to the model product differentiation (through the development of thematic forms of tourism), coupled with operational effectiveness is the appropriate course of action to restore competitiveness in Ionian tourism, since the product cycle of mass tourism in the region has reached the stage of maturity and the decline in some areas.

Higher quality of services and product development based on the unique characteristics (historical, cultural, physical, etc) of the region, will allow Ionian tourism to differentiate the product, prolong the tourism period each year, abate seasonality, diffuse tourism development in the region and attract higher income visitors who will spend more per night. At the same time it will decrease the price elasticity of demand and ameliorate the corresponding income one. It will also contribute to an increase in revenues, asset turnover, profit margin and restore a well grounded and lasting viability. This strategy is deeming to be suitable, feasible and acceptable.

The specific type of strategy has been identified as more appropriate to restore financial soundness in the sector and in the case of the analysis of strategy formation and evaluation with the use of business matrices (Curtis 2004, Ch.4). Product differentiation and higher quality policy are the elements in accordance with the strategy proposed by EU and it more suitable to regions composed of islands, where the carrying capacity is more limited and ecosystems more fragile. European Union fosters quality as a means of satisfying tourists and increasing competitiveness and proposes a system of integrated quality management-IQM (European Commission 2000, 13).

REFERENCES

Biggs, D., deVille, B. & Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, Vol. 18, No.1, pp.49-62.

Brieman, L., Friedman, J., Olshen, R.A. & Stone, C. (1993). *Classification and Regression trees.* CA, Belmont.

Curtis, P. (2004). Strategy and Competitiveness in tourism towards regional development–The case of Ionian Islands. Unpublished PhD Thesis, Harokopion University, Athens.

Drucker, H. In Sharkey (Eds), *Boosting using neural networks* (pp.51-78).

European Commission. (2001). *Towards quality coastal tourism. Integrated quality Management of coastal tourist destinations*. Brussels, Enterprise Directorate-General, Tourism Unit.

Grant, R.M. (2002). *Contemporary Strategy Analysis. Concepts, techniques and applications* (4th ed.). Blackwell Business.

Hand, D., Mannila, H. & Smyth, P. (2001). *Principles of Data Mining*. Massachusetts, M.I.T Press.

Hax, A. & Majluf, N. (1996). *The Strategy concept and process a pragmatic Approach*. Prentice Hall.

Kokotos D.A. (2003). Data mining – Decision tree analysis upon vessel accidents. Phd Thesis, Department of Maritime Studies, University of Piraeus.

Kokotos X.D. & Smirlis G.Y. (2005). A classification tree application to predict U.S.A total ship loss. *Journal of Transportations and Statistics*, A Peer-Review Innovative Technology Administration, Bureau of Transportation. *Journal Department of Transportation, Research and Statistics*, Vol. 8., No.2, pp.31-42.

Porter, M.E. (1996). *What is strategy*? Harvard Business Review Nov-Dec., pp. 61-78.

Ripley, B.D. (1994). *Neural networks and related methods for classification* (with Discussion) J.R. Statist. Soc., pp. 409-456.

Witten, H. & Frank, I. (2000). Data Mining - Practical Machine Learning Tools. San Francisco and Techniques with Java Implementations. Morgan Kauffmann Publishers.

Valuebasedmanagement.net

Panayiotis Curtis (pcurtis@teihal.gr) is an associate professor at the Technological Institute of Chalkida, Department of Accounting, Psahna, 34400, Eubea, Greece.

Dimitris Kokotos (dkokotos@unipi.gr) is a Lecturer at the University of Piraeus, Department of Maritime Studies, 40 Karaoli & Dimitriou St, 18532, Piraeus, Greece.