Application of Machine Learning Methods for Data Analytics in Social Sciences

DIJANA ORESKI
Faculty of Organization and Informatics,
University of Zagreb,
Varazdin, Pavlinska 2,
CROATIA

Abstract:-This article addresses the challenges in the application of artificial intelligence methods such as machine learning, computational intelligence and/or soft computing methods in social sciences. The literature review is performed in order to give a review of different approaches and methods that have been applied so far. The most used method in social sciences and management is the SWOT method, for the identification of strengths, weaknesses, opportunities, and threats when making strategic decisions. Two fundamental characteristics of previous approaches are the development of numerical models of utility functions and the possibility of upgrading these models by formalizing the intuition of strategic decision-makers. There are several shortcomings of the existing approaches. The application of computational intelligence and machine learning methods in social sciences is identified as one of the most challenging and promising areas, which could overcome identified shortcomings. The principles of one popular machine learning method, the decision tree, are explained and a demonstration is performed on the case study of churn prediction. Benchmarking data set from the publicly available repository is used to demonstrate the suggested approach Evaluation results measured through model accuracy and reliability gave promising results for further analysis. A developed predictive model could serve as a standalone tool or as support for decision-making in social sciences.

Key-words:-*Computational intelligence, data mining, data science, machine learning, social sciences, business.*

Received: April 14, 2022. Revised: January 9, 2023. Accepted: February 4, 2023. Published: March 7, 2023.

1 Introduction

This article seeks to contribute to solving one of the most challenging problems of artificial intelligence: application for planning in social sciences. Artificial Intelligence becomes the key technology of the XXI century, which will become the most significant economic branch in the next decade, with a great impact on all areas of human activity. Today, artificial intelligence demonstrates its superiority in solving well-structured lower intellectual-level issues such as machine learning (e.g. business analytics), visual recognition, speaking, translating, and converting text into speech. Solving complex problems such as planning in social sciences still needs to prove the ability of artificial intelligence. Modern business-based analytics, based on machine learning, is successful in supporting decision-makers and automating processes in the business of lower intellectual levels. Despite the supremacy of artificial intelligence concerning human beings in many areas, strategic planning is still unmanageable for people of their intuition and creativity, and the ability to see long-lasting changes that are not yet predictable to existing business analytics.Today's organizations, from the smallest (SME) to the global ones, base their actions on highly structured strategic planning methods. One of these techniques, SWOT analysis, is the most commonly applied management method. On the other hand, some approaches completely deny the applicability of strategic management, as they believe business systems are operating in a very dynamic environment where rigid planning can`t be long-term effects. Some newly-created technological giants base their business on employee creativity, and less on long-term planning. Most organizations still operate based on planning. Increased dynamism and speed of change in modern business (shortening product cycles) increasingly diminish the effectiveness of strategic planning, while at the same time increasing the need for it. Planning in social sciences (as a general term) is one of the affirmed areas of artificial intelligence. Decision makers are increasingly relying on analytics, but only partially in creating a strategy (e.g. 30%). Prediction based on historical data, achieved through machine learning techniques, is the basis for effective strategic planning. The existing

research was primarily concerned with the upgrading of qualitative techniques of strategic planning by quantitative methods (e.g. machine learning techniques, expert systems, and intelligent agents). A special challenge is to provide citizens with a rational decision-making that has significant long-lasting impacts on their lives. This research aims to prove the applicability of artificial intelligence and machine learning methods in social sciences. This paper is structured as follows. Section 2 provides a review of existing approaches. Section 3 explains used methodology and data. Section 4 gives an overview of the research results. Section 5 concludes the paper.

2 Related Literature

Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis is a widely used technique and one of the most common tools in management. SWOT is a brief list of statements or factors with descriptions of the present and future trends of both the internal and external environment. However, SWOT analysis has no means of determining the importance of each SWOT factor [1]. Thus, the utilization of SWOT alone in decision-making process is insufficient. Kurttila et. al. [2] recognized this limitation of SWOT analysis and its impreciseness of a quantitative examination. They created a hybrid SWOT-AHP method where SWOT analysis usability was improved. The limitation of the qualitative nature of SWOT analysis is then overcome with the quantitative SWOT-AHP method, but they still both stayed subjective, developed by the human decision-makers. SWOT-AHP has been used for strategic planning [3] in various domains, such as tourism [4] and manufacturing [5] In 1999. Houben at al. [6] described an interesting application of a knowledge-based system (KBS) to SWOT-analysis strategic planning in small and medium-sized enterprises. They are focused on the identification of internal strength and weakness factors recognized by this KBS from the financial situation of an organization. There are only a few papers so far that utilize the mainstream of a huge database growth and wide application of business intelligence and data mining to the definition of organizational strategies with the common and acceptable frame of SWOT analysis. Knowledge Discovery in Databases (KDD) and Data Mining (DM) techniques can model most complex systems accurately outperforming previously established linear methods. KDD and DM can develop models of complex systems represented by neural networks and decision trees. Furthermore, Milano et.al. [7] tried to cover “public policy issues in a wide variety of fields: economy, education, environment, health, social welfare, and national and foreign affairs. They are extremely complex, characterized by uncertainty, and involve conflicts among different interests.” Authors [7] also see the advantages of artificial intelligence as a solution for such complex problems. Athey [8] recognized big data potential in policy problems. Based on the aforementioned, this paper seeks to use the advantages of artificial intelligence and machine learning in order to solve strategic decision-making issues in social sciences.

3 Data and Methods

In the first two sections of the paper, we have described the recent developments and applications of strategic decision-making methods and artificial intelligence. The literature review demonstrated that present methods are insufficient for application in social systems that are nonlinear, complex, and based on complex dynamic laws, and variables in such systems are often not possible to measure exactly. This was the motivation for a new approach based on the application of artificial intelligence methods. Our design is based on the following models:

(i) Application of data mining and standard methods for conducting CRISP-DM.

(ii) Simulations driven by goals and data.

(iii) Evaluation and interpretation of predictive models.

The steps of the research based on the CRISP-DM methodology are explained in table 1.

Table 1. Research description through phases

Phase	**Steps**
Problem definition	Assessment of the environment Definition of business goals Assessment of the situation Determining performance criteria
Data understanding	Initial data collection Data description Basic statistical analysis Data quality assessment
Model structuring	Model structure development
Data preparation	Data set description Data selection Data cleaning Deriving attributes Data integration
Modeling	Choice of modeling technique Definition of model parameters Model description
Evaluation	Evaluation of data mining results in relation to business success criteria Model interpretation
Deployment	Application activity plan Implementation and performance control

The advantage of CRISP-DM is that combines the development of models by applying data mining techniques, and supplements the model with the knowledge and intuition of past data on the given topic and domain in social sciences.

4 Research Results

CRISP-DM standard is applied to data from one domain of social sciences, business. To demonstrate the application of machine learning in social science datasets about predicting whether a customer will change telecommunications provider, something is known as "churning", is used. The source of the dataset is the repository Kaggle [9]. Firstly, data description is performed through distribution representation for each attribute. Results are presented in Table 2.

Table 2. Data description

Variable	*Distribution*
CustomerID	
Gender	
Senior Citizen	
Partner	
Dependents	
Tenure	
PhoneService	
MultipleLines	
InternetService	
OnlineSecurity	
OnlineBackup	
DeviceProtection	
TechSupport	

This dataset is used in the modeling phase. The decision tree is applied to the dataset, as a machine learning algorithm for the development of predictive models. A decision tree is a well-known algorithm whose results are easy to understand. Different learning parameter settings were employed on the decision tree algorithm to get good models. In the end, active statistical pruning is used. The reliability of the active statistical pruning model is 73.46%. The attribute for which we make the prediction model is Churn, ie the departure of the client. We ask the question "Will the client leave or not?". The model we will choose for further analysis will be the active statistical pruning-based model since it is the most accurate and reliable of all models. The accuracy of such a model is very high because the error of the model is less than 4% (3.45%) Figure 1 depicts the model.

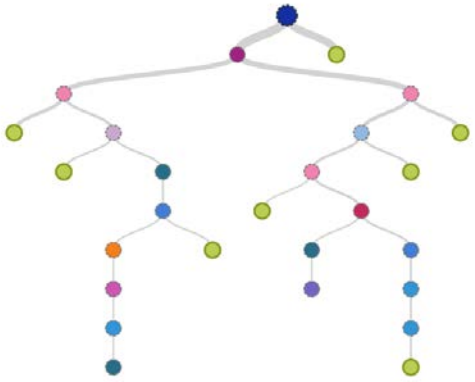

Fig. 1: Decision tree model

Sensitivity analysis is performed with the aim to detect the most important variables for churn prediction. The results are shown in figure 2.

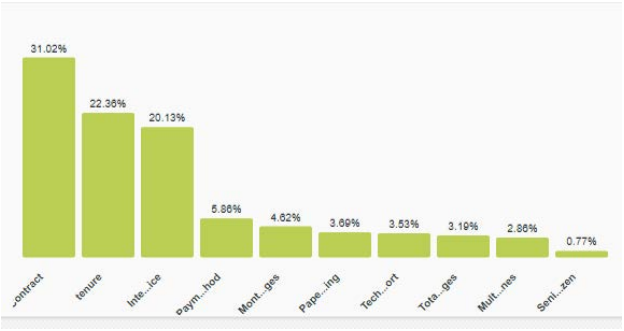

Fig. 2: Results of sensitivity analysis

The most important attribute is at the root, which is Contract, then Tenure, and then Internet Service. In the decision tree, we can see that the probability of the client leaving is highest when the attribute Contract = Month-to-month, Tenure <= 14, and InternetService = Fiber optic is 69.72%. Interpreting these data, we can conclude that most often customers who renew the contract period monthly, use the services of the company for less than 14 months and have fiber optics.

When developed, a predictive model can be used in prediction for new, unseen data. One example of prediction is given in figure 3.

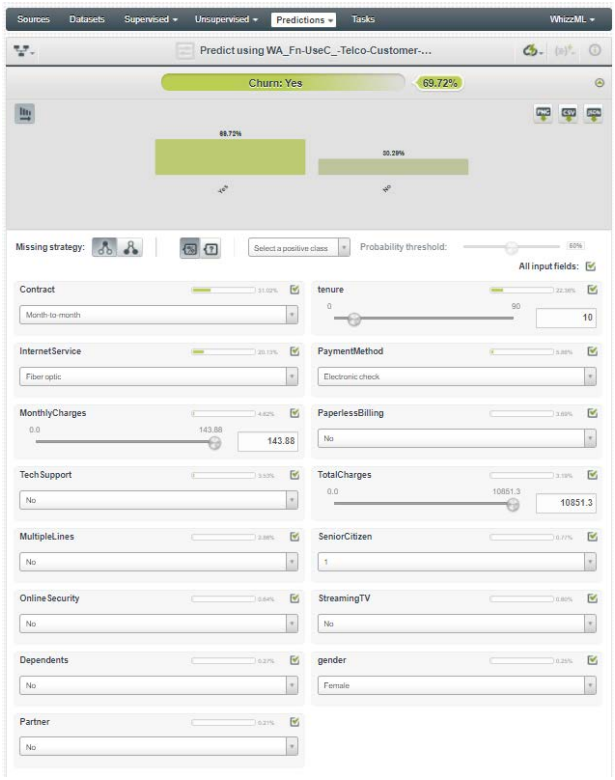

Fig. 3: Deployment phase: the model used in prediction for new clients

The decision tree model can be transferred into rules, which are easy to understand. Figure 4 gives the extraction of one such rule.

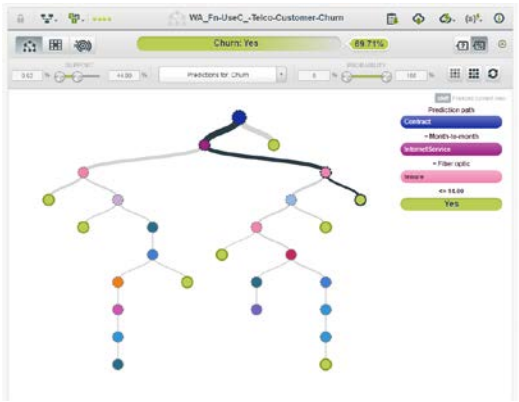

Fig. 4: Extracted rule

5 Conclusion

In this paper, we have demonstrated the application of machine learning in social sciences in the case of a decision tree algorithm for the prediction of churn. The results show the churn prediction could be accurate using machine learning. This analysis has shown that the proper machine learning application on churn data can be efficiently used for the vital extraction of valuable hidden knowledge from the vast amount of data generated on daily basis. Future studies should evaluate other machine learning algorithms, such as artificial neural networks, k-nearest neighbors, Bayesian approaches, support vector machines, and ensemble methods to assess whether performance is improved.

References:

[1] Shinno, H., Yoshioka, H., Marpaung, S., & Hachiga, S. (2006). Quantitative SWOT analysis on global competitiveness of machine tool industry. Journal of engineering design, 17(03), 251-258.

[2] Kurttila, M., Pesonen, M., Kangas, J., & Kajanus, M. (2000). Utilizing the analytic hierarchy process (AHP) in SWOT analysis—a hybrid method and its application to a forest-certification case. Forest policy and economics, 1(1), 41-52.

[3] Osuna, E. E., & Aranda, A. (2007). Combining SWOT and AHP techniques for strategic planning. Economic Journal. Instituto de Estudios Superiores de Administración (IESA) Avenida IESA, San Bernardino, Caracas–Venezuela.

[4] Jeon, Y., & Kim, J. (2011). An application of SWOT-AHP to develop a strategic planning for a tourist destination.

[5] Görener, A., Toker, K., & Ulucay, K. (2012). Application of combined SWOT and AHP: a case study for a manufacturing firm. Procedia-social and behavioral sciences, 58, 1525-1534.

[6] Houben, G., Lenie, K., & Vanhoof, K. (1999). A knowledge-based SWOT-analysis system as an instrument for strategic planning in small and medium sized enterprises. Decision support systems, 26(2), 125-135.

[7] Milano, M., O'Sullivan, B., & Gavanelli, M. (2014). Sustainable policy making: A strategic challenge for artificial intelligence. ai Magazine, 35(3), 22-35.

[8] Athey, S. (2017). Beyond prediction: Using big data for policy problems. Science, 355(6324), 483-485.

[9] Kaggle (2022), available at: https://www.kaggle.com/blastchar/telco-customer-churn

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This work has been supported in part by Croatian Science Foundation under the project UIP-2020-02-6312.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.