A Recommender System for fund raising management

Luca Barzanti¹, Silvio Giove², and Alessandro Pezzi³

Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5, 40126, Bologna, Italy,

luca.barzanti@unibo.it

² Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, 30121, Venezia, Italy

sgiove@unive.it

School of Economics, Management and Statistics, University of Bologna, Piazzale della Vittoria 15, 47121, Forlì (BO), Italy a.pezzi@unibo.it

Abstract. In fund raising management the use of rigorous mathematical methods and decision support systems has been playing a more and more important role. These techniques develop the classical data base approach proposed by the operational literature. These improvements concern both the mathematical modeling associated with the information management techniques and the specific characteristics of the Associations, with the consequent specializations of the algorithms. In this approach, the role of the potential donors (i.e. contacts) and the process for turning the contacts into actual donors in the context of the so called "giving pyramid" was poorly investigated, despite its high importance in particular from a strategic point of view. In this contribution, we develop a recommender system that uses similarity measures to optimize the contacts' management. This is achieved by a proper use of the information contained in the Association data base, which concern the profiles of those donors that are suitable for the current campaign, and by matching these with the profiles of the contacts by similarity. The similarity among the (normalized) profiles will be realized by a suitable distance-based function. Numerical results show the effectiveness of the proposed approach.

Keywords. Fund raising management, similarity measures, recommender systems, Social Economics.

M.S.C. classification. 65K05, 68U35 J.E.L. classification. C61,C63

Mathematical Methods in Economics and Finance – m²ef Vol. 15/16, No. 1, 2020/2021

1 Introduction

The fund raising (FR) activity is particularly important for those Associations which operate in the context of Social Economics, see [1]. FR strategies are therefore crucial for the achievement of the mission and, specifically, for reaching the goal of the current campaign [18]. Quantitative methods employing data base technologies have been studied and developed in the literature for making these strategies more effective, see [14]. In fact, donors and contacts (i.e. potential donors) are normally managed by an organized Data Base (DB), in which much information about them is stored. The effective use of this information is crucial for optimizing the resources for the campaign by selecting the most promising contacts for the considered context. The ground of the so called "giving pyramid", [16], contains all those potential donors that are already known by the Association. One of the goals of a loyalty campaign, in addition to the achievement of its specific objective, is involving new people in the mission of the Association by their first donation. Operatively speaking, the goal is to make some contacts going up from the ground of the pyramid to the first level. This is particularly important from a strategic point of view. In fact, the effectiveness to achieve the step at the first gift request makes the new donor more prone to become soon a regular donor, establishing in this way a positive drift (the donor becomes "hot") towards a subsequent step to the second level of the pyramid. In this paper we focus our attention on this important aspect of the fund raising process, by determining a recommender system, a branch of Artificial Intelligence¹, that identifies the more suitable contacts to reach in the current campaign. In fact, reaching a person has obviously a cost (that depends on the type of request) and the budget constraint requires a choice. For this purpose, a suitable similarity measure, [10], can be used for matching the profile of the contacts with the profile of all the regular donors in the Association DB which usually give for similar campaigns. More specifically we suppose that the Association has stored a set of ("actual") donors which contains, for each element, a personal description in terms of a set of characteristics such as personal data, financial situation, risk aversion, etc., together with the record of requests and gifts (if any) for each previous campaign. The record specifies the type of campaign and the type of request (direct, mailing, e-mail, telemarketing, etc.). Moreover, the Association considers another set of contacts, formed by a list of elements where each one of them is described by the same set of actual donors. Since the elements of the second set have never given a gift in the past, the record of past requests and gifts is empty. Because of the request cost and the huge numbers of elements of both sets, given a pre-defined type of campaign, the aim of the Association consists in selecting the most promising contacts, so that only this reduced set of potential donors will be contacted. The type of request will also be considered and, for each contact, the system will provide a probability of the response to

¹ Refer to [19] for an exhaustive introduction to algorithms and aim of Artificial Intelligence.

the request in function of its typology. The output of this recommender system is a list of potential donors with gift probability higher than zero.

2 The fund raising problem in the context of non profit campaigns

For few years, the literature in the area of mathematical models and Decision Support Systems (DSS) has dealt with the fund raising problem in Social Economics, as it has been doing in for profit Economics. First attempts to develop a DSS for supporting FR management were presented in [2], [3]. At an operative level, the developed DSS are effective for large-sized Associations, that have a powerful organizational system. In [21] an analysis is developed in the same wake of this approach. On the other hand, the world of small-sized and medium-sized Organizations has also been studied (respectively in [4] and [5], by exploring their specific features and their consequences in modeling the FR process. Till now, only in [3] the problem of the contacts' management was considered, in the context of large-sized Organizations and when the budget is enough. By using the results obtained in Econometrics [12], [15], and more specifically in [9], [11], the variables of the personal profile that influence the gift probability are selected. Then by a fuzzy technique a fitness index is computed. This approach is quite generic, because it doesn't consider neither the particular characteristic of the current campaign nor the data of the Association DB regarding the specific profiles of donors that donated in the past for similar campaigns. Furthermore, from a strategic point of view it is particularly important that a contact becomes donor at the first gift request, and therefore it is important to solicit the gift in a campaign that is suitable for that contact. In fact, in this way not only the probability of a gift is higher, but also the probability of stabilizing the donor at the first level of the pyramid. Conversely, a refusal to the first solicitation will make the contact "cold" and less prone to take into consideration subsequent requests. It is therefore necessary to develop, with the appropriate mathematical techniques, a system that performs the task of promoting the contacts from the ground to the first level of the pyramid and that possibly stabilizes these new donors at this level. In this context, in the present paper we propose an approach that, for the first time, is optimized with respect to the mentioned objectives and uses all the available information.

3 The Fund Raising Recommender System; modeling and formalization

As presented above, two separate sets will be considered, the set of actual donors, say A, and the set of potential donor, P. The structure of the set E is as follows:

$$E = \{X, Y\} = \{(X_1, Y_1), ..., (X_n, Y_n)\}$$
(1)

being $X=(X_1,X_2,...,X_n), \ Y=(Y_1,Y_2,...,Y_n), \ \text{and} \ X_i=\{x_{i,1},...,x_{i,m}\}$ is the vector formed by the *personal* characteristics of the *i*-th donor, as age, level of education, financial situation, risk aversion and so on, while Y_i is the vector formed by a list of the N_i past requests for the i-th donor for which the donor has been contacted: $N_i \ 4-ples \ y_i, i=1,...,N_i$. Each $4-ple \ y_i$ is formed by $C_{i,k}, D_{(i,k)}, Q_{(i,k)}, R_{(i,k)}, (k=1,...,N_i)$, that are the type of campaign, the request date, the gift amount², and the type of the request (direct, mailing, telephone, e-mail, telemarketing...) respectively. Thus the structure of Y_i is:

$$Y_i = \{ y_i = (C_{i,k}, D_{(i,k)}, Q_{(i,k)}, R_{(i,k)}) \}_{(k=1,\dots,N_i)}$$
(2)

Moreover, the set of the M contacts is the set P (potential donors):

$$P = \{S_1, S_2, ..., S_M\} \tag{3}$$

where each vector S_k has the same structure of the first part of the vector D (personal characteristics): $S_j = \{x_{j,1}, ..., x_{j,m}\}$, and obviously it does not contain information on given gifts.

The proposed methodology consists of two phases. In the first phase - the donor's characterization - for each donor in the set E an empirically estimated gift probability is assigned, for each type of campaign and for each type of request³. This probability will be computed on the basis of the past gifts for each type of campaign and for each type of request, and the way of reasoning is case-based as in [8]. The following set E_1 is thus obtained:

$$E_1 = \{(X_1, \{z_1(c,r)\}), ..., (X_n, \{z_n(c,r)\})\}$$
(4)

being $z_i(c, r)$ a vector containing the estimated gift probabilities in function of the type of the c-th campaign and of the type of the r-th request (see below).

In the second phase, given the type of (current) campaign, all the elements in the set E will be compared with the donors in the set P. Thus the contact $S_i \in P$ is compared with the donor $D_j \in E$, using a suitably selected *similarity* measure Sim, which is computed between the two sets of the personal characteristics (of the considered contact and the donor): $s_{i,j} = Sim(X_i, X_j)$. The values of the similarities among all the donors are then used to compute an estimation of the empirical gift probability (see Section 5 below), using an adapted version of the K-NN algorithm (K-nearest neighbor), see [20], [13]⁴. The underlying idea is that the similarities are treated as suitable weights of a weighted averaging of the gift probabilities of all the donors estimated in the first phase. Given that the

 $^{^{2}}$ Q can be null if the donor hasn't given any gift for this request.

³ In this preliminary analysis, we consider only the probability of the elementary event "Gift/NoGift", while subsequently we intend to extend the model to a probability distribution, see [5].

⁴ We remark that the proposed approach is partially inspired to the general *lazy* learning methodology, see [17].

similarities are strictly correlated with the inverse of the distance, the proposed algorithm is a weighted K-NN where the weights are the similarities among the considered contact and all the donors in the Data Base, i.e. K=n.

4 The Actual donor's characterization

In this Section, we explain both the *personal* characterization - in the first Subsection - and the *performances* characterization of the Actual donors - in the other one.

4.1 The personal characterization

As mentioned in the Introduction, the econometric literature individuates the variables of the personal profile that influence the gift probability. These variables are commonly stored in an organized Data Base and can therefore be properly used.

The following personal data will be considered, i.e. the elements of the vector X_i :

- 1) Age (real),
- 2) Qualification (label), that can assume 4 possible ordered values, from PhD (P)
- the highest value to B (Bachelor), then H (High School) down to O (Other).
- 3) Presence of children (Boolean), Y or N.
- 4) Risk Aversion (*Integer*), *i.e.* the number of insurances (health, home, other) and the presence of a will.
- 5) Financial situation (Financial for brevity) (*Real*), measured in thousand of Euro, *i.e.* the global entry amount (income plus real estate and other possible financial income).

Each variable will be transformed into the common scale [0,1] using suitable value functions to allow a comparison between two different items. A 5-dimensional vector contains the donor's personal characteristics (the vector X_i for an Actual donor and the vector Y_j for a contact), and, after the normalization, the components of the vector are in [0,1], thus we can consider that each donor may be represented by a fuzzy set. The similarity between two fuzzy sets has widely been analyzed in the literature, among other see [10]; in this paper we adopt a distance-based similarity measure, see [6], thus the similarity functions between two fuzzy sets $A, B \in \mathbb{R}^5$ can be computed as:

$$Sim(A,B) = 1 - \sqrt{\frac{\sum_{i=1}^{5} (A(i) - B(i))^2}{5}}$$
 (5)

Three kind of variables are used in the description of the personal characteristics, so in order to normalize them, we consider that for the real variables (*Age* and *Financial*) two non decreasing *membership* functions will be defined for both

the variables, $f_A(Age)$, $f_E(Financial)$, where each of them represents the degree of truth of the propositions Age is High and Financial is High respectively. For the Boolean variable (Presence of children) the transformation leads to 0 if there is no children, and to 1 otherwise, but in the latter case a value $\omega < 1$ is assigned to this variable, as the relative importance of this variable with respect to the remaining ones. Finally for the integer variables (Qualification and Risk Aversion), a suitable set of weights is defined in function of the number of qualifications, $\eta_1, \eta_2, ..., \eta_{n_Q}$ and $\theta_1, \theta_2, ..., \theta_{n_P}$, being n_i the ith Qualification weight, $i = 1, ..., n_Q$ and θ_j the jth Risk Aversion weight, $j = 1, ..., n_P$.

4.2 The performances characterization

In order to identify the *performances* of the donors (set E), we consider the gift frequency, measured as the ratio between the given gift and the number of requests, together with an empirical evaluation of its robustness. For this purpose, we briefly review some basic notions about the methodological framework, more details can be found in [5]. Each Actual donor is represented by a string of data, (X_i, Y_i) collecting both personal information (X_i) and the time series of past requests and gifts for each kind of campaign (Y_i) , thus including the series of requests, the corresponding date, the type of campaign and the contingent gift⁵. A rough, simple but effective robustness measure is given by the number of the requests⁶. Then, as for the donor's characterization, we summarize all this information into a single indicator, that defines a more or less significant attitude to gift. This indicator is the gift frequency, the number of given gifts (positive responses) over the number of the total requests for a specific campaign type and following a specific type of request. The gift frequency reflects the attitude to be a frequent donor (independently of the gift amount)⁷. If n_i^+ , n_i^- are respectively the numbers of positive and negative responses for the i-th donor, the gift frequency $f_i(c,r)$ corresponding to the type of campaign c and the type of request r is:

$$f_i(c,r) = \frac{n_i^+(c,r)}{n_i^+(c,r) + n_i^-(c,r)}$$
(6)

In order to satisfy robustness, we require that the total number of past requests $n_i(c,r) = n_i^+(c,r) + n_i^-(c,r)$ is sufficiently high. This requirement can be obtained by a fuzzy membership - monotonic increasing function $sr_i(c,r) = g_R(n_i(c,r))$:

⁵ Possibly zero, if the donor gave no gift.

⁶ Few requests give poor information about the donor's attitude.

⁷ In this phase, the gift *value* - a measure of *generosity* - is not considered here. In fact, we are only interested in the selection of *plausible* new donors (yet the process of clustering in advance the set of possible donors is not an easy task), as specified in Section 2 with respect of the strategic point of view and the ground level of the pyramid.

 $R^+ \to [0,1]$ with the usual meaning: its value is the true value of the proposition The number of requests for the campaign type c and for the request type r is sufficiently high and can be used in the following aggregation phase. Thus, as for the performance characterization, the i-th donor is described by the couple $(f_i(c,r), sr_i(c,r))$.

5 Gift Probability and Contact ranking

The gift probability for a (new) contact, in function of the type of request and the type of campaign, are computed through the aggregation of the gift probabilities of donors which are weighted by the similarities among the contacts and the donors⁸. The aggregation of the similarities defined in (5), see [7], [10] can be used to compute a weighted sum of the gift probabilities of the donors, *i.e.* the empirical gift probability for the j-th contact, in function of the type of request and the type of campaign:

$$p_j(c,r) = \frac{\sum_{i=1}^n Sim(i,j) \times sr_i(c,r) \times f_i(c,r)}{\sum_{i=1}^n Sim(i,j) \times sr_i(c,r)}$$
(7)

In this formula, the term $Sim(i,j) \times sr_i(c,r)$ takes both the similarity and the robustness into account, and weights the gift frequencies of all the actual donors in the database in order to compute the gift probability for the j-th contact. The values $\{p_j(c,r)\}$ can be used to select the M contacts in function of the type of campaign and the type of request.

6 Computational results

SQL and Visual Basic are used to implement the Algorithm in a MS Access DB, which is limited to a small numbers of donors and contacts for clarity of exposition. For the same purpose, the type of campaign (i.e. children health) and request (i.e. leaflet) are fixed. The personal characteristics of the Actual donors, the set X, together with the whole user interface, are represented in the Figure 1. A graphical interface allows the fund raiser to set up the main parameters by which the similarities and the statistical robustness are calculated.

The performance data - set Y - for each donor is represented by the following list, containing the request date, denoted by an integer number, and the gift amount

$$d_1 = \{(12,0), (15,50), (18,60), (20,0), (25,0), (28,0), (33,0), (35,0), (45,0)\}$$

$$d_2 = \{(2,40), (4,0), (9,0), (14,0), (17,0), (19,100), (20,0), (23,0), (27,0)\}$$

⁸ Namely, also the robustness will be taken into account, see below.

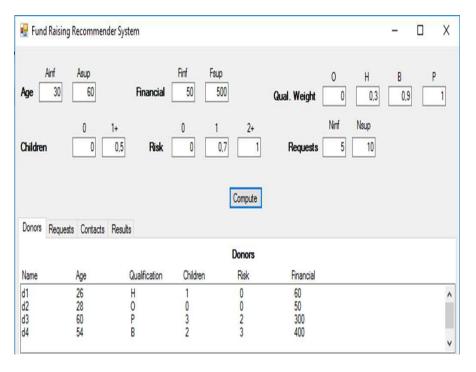


Fig. 1. The user interface with actual donors

$$(30,0), (32,0), (35,0), (38,0), (40,0), (44,80), (48,0), (50,0) \}$$

$$d_3 = \{(0,100), (5,100), (10,0), (12,100), (15,100), (21,100), (24,0), (28,100) \}$$

$$d_4 = \{(8,40), (13,45), (20,50), (23,0), (27,45), (32,40), (38,45), (43,50), (48,50), (54,0) \}$$

As membership functions $f_A(Age)$, $f_{Fin}(Financial)$, we consider here two continuous piecewise linear functions, which are defined through two suitable thresholds, a lower and an upper bound inf_{Age} , sup_{Age} , inf_{Fin} , sup_{Fin} . The values which are inferior to the lower bound $(inf_{Age}$ for Age, inf_{Fin} for Financial) mean completely unsatisfying values, on the other side the values which are greater than the upper bounds (sup_{Age}, sup_{Fin}) mean completely satisfying values. Finally, the degree of satisfaction moves linearly in the cases inside the bounds.

Again, the membership function $sr_i(c,r)$ is a piecewise linear function and is defined in the same way as $f_A(Age)$, $f_{Fin}(Financial)$ through a lower and an upper bound, inf_{sr} , sup_{sr} with the same interpretation as above. The param-

eters ω, η_i, θ_i can be subjectively assigned by the Decision Maker⁹; just as a simulation example, let us suppose that the parameters are chosen as follows:

$$inf_{Age} = 30, sup_{Age} = 60, inf_{Fin} = 50, sup_{Fin} = 500$$

 $inf_{sr} = 5, sup_{sr} = 10$
 $\omega = 0.5$
 $\eta_1 = 0, \eta_2 = 0.3, \eta_3 = 0.9, \eta_4 = 1$
 $\theta_1 = 0, \theta_2 = 0.7, \theta_3 = 1$

From these values, the data in Figure 1 will be modified as Figure 2 shows, that also includes the computation of donors' frequencies and robustness, denoted here by the variable gr.

Normalized Donors									
Name	Age	Qualification	Children	Risk	Financial	gr	Frequency		
d1	0	0,3	0.5	0	0.02	0.8	0,22	٨	
d2	0	0	0	0	0	1	0.18	10000	
d3	1	1	0.5	1	0.56	0.6	0.75		
d4	0.8	0,9	0,5	1	0,78	1	0,8	٧	

Fig. 2. Normalized Donors, robustness and gift frequencies

The three contacts are described by the Figure 3.

			Results	Contacts	Requests	Donors
	Contacts					
Financial	Risk	Children	Qualification	Age		Name
60	0	1	Н	26		c1
800	1	0	0	52		c2
800 400	3	2	P	60		c3

Fig. 3. Contacts

⁹ This is obviously a crucial phase of the procedure, but we do not enter into technical details (the interested reader can refer to the rich parameters assessment literature, see among others, [7]) because out of the scope of our proposal.

	Financial GIFT PROBABILITY
21 0 0,3 0,5 0	0,02
52 0,73 0 0 0,7	1
63 1 1 0,5 1	0,78

			Similanties
Contact	Donor	Value	
c1	d1	1	^
c1	d2	0,74	
c1	d3	0,26	
c1	d4	0,28	
c2	d1	0,32	
c2	d2	0,36	
c2	d3	0,43	
c2	d4	0,51	
c3	d1	0,22	
c3	d2	0,12	
c3	d3	0,9	
c3	d4	0,9	
			¥

Fig. 4. Normalized contacts, similarities and gift probabilities

Now we can proceed with the computation of the normalized contacts, of the similarity function among the contacts and the donors in the set E_1 and finally of the gift probability for each contact, as Figure 4 shows.

Notice that the probabilities are significantly different, in consequence of the different profiles of considered contacts, and give in this way a clear indication to the Decision Maker.

After the detailed description of the algorithm, that is clearly only possible by considering a very small number of Donors and Contacts, we now consider a complete numerical application, in a Data Base with 1000 Donors and 200 Contacts, achieving in this way a real context, developed in collaboration with Philanthropy Centro Studi¹⁰, with simulated data appropriately implemented by the criteria of the giving pyramid. In Figure 5, Contacts are classified in gift probability bounds. The lower part of the Figure specifies the Contacts in the highest two bounds, which can be selected by the Decision Maker in the implementation of the campaign.

Since this approach is novel, in particular with respect to the rigorous application of a quantitative method, a comparison with other numerical methods of the literature is not possible.

Research Centre of non profit, fund raising and social responsibility operative in the University of Bologna.

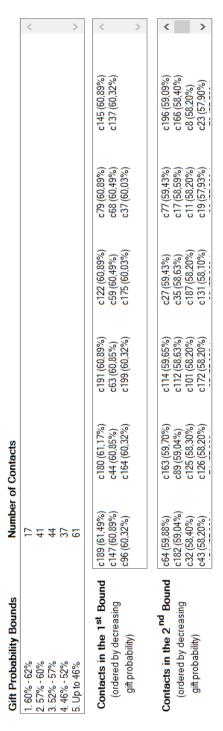


Fig. 5. Results of the complete numerical application

7 Conclusion

In this paper we develop a recommender system for fund raising management, that deals with the important issue of placing new donors in the first level of the giving pyramid. The similarity among profiles of actual donors and possible new donors is suitably used to focus the resources of the Association in reaching the most promising contacts. The profile variables which are taken into account are the most important according to the relevant literature. Numerical experiments give evidence of the effectiveness of the proposed approach.

Further developments include the consideration of similarity among different types of campaigns as well as different request modes, the use of combined strategies in order to improve their effectiveness and the fine tuning of the numerical parameters according to the actual requirements of the various Associations.

Acknowledgments We are grateful to the anonymous Referee who provided useful comments and suggestions.

References

- Andreoni, J.: Philantropy. In: L. A. Gerard-Varet, S. C. Kolm and J. Ythier (eds.), Handbook of giving, reciprocity and altruism. North Holland Amsterdam (2005)
- Barzanti, L., Dragoni, N., Degli Esposti, N., Gaspari, M.: Decision making in fund raising management: a knowledge based approach. In: Ellis, R., Allen, T., Petridis, M. (eds): Applications and Innovations in Intelligent Systems XV, Springer (2007) 189-201
- Barzanti, L., Gaspari, M., Saletti, D.: Modelling decision making in fund raising management by a fuzzy knowledge system. Expert Systems with Applications, 36 (2009) 9466-9478
- 4. Barzanti, L., Giove, S.: A decision support system for fund raising management based on the Choquet integral methodology. Expert Systems 29 (4) (2012) 359-373
- Barzanti, L., Giove, S.: A Decision Support System for fund raising management in medium-sized Organizations. M2EF - Mathematical Methods in Economics and Finance, 9-10 (2018) 3-12
- Beg, I., Ashraf S.: Similarity measures for fuzzy sets. Appl. and Comput. math, 2
 (4) (2009) 192-202
- 7. Beliakov G., Calvo T., James S.: Aggregation of Preferences in Recommender Systems. In: Ricci F., Rokach L., Shapira B. (eds.): Recommender Systems Handbook, Springer, New York (2011) 705-734
- 8. Bridge, D., Goker, M., McGinty, L., Smyth, B.: Case-based recommender systems. The Knowledge Engineering Review **20** (3) (2006) 315-320
- 9. Cappellari, L., Ghinetti, P., Turati, G.: On time and money donations. Journal of Socio-Economics 40 (6) (2011) 853-867
- 10. Couso, I., Garrido, L., Sánchez, L.: Similarity and dissimilarity measures between fuzzy sets: A formal relational study. Information Sciences **229** (2013) 122-141
- 11. Duffy, J., Ochs, J., Vesterlund, L.: Giving little by little: Dynamic voluntary contribution game. Journal of Public Economics 91 (9) (2007) 1708-1730
- 12. Duncan B.: Modeling charitable contributions of time and money. Journal of Public Economics **72** (1999) 213-242

- 13. Garcia E. K., Feldman S., Gupta M. R., Srivastava S.: Completely Lazy Learning. IEEE Transactions on Knowledge and Data Engineering DOI: 10.1109/TKDE.2009.159 **22** (9) (2010) 1274-1285
- 14. Kercheville, J., Kercheville, J.: The effective use of technology in nonprofits. In: Tempel, E. (eds): Hank Rosso's achieving excellence in fund raising, John Wiley & Sons (2003) 366-379
- Lee, L., Piliavin J. A., Call, V. R.: Giving time, blood and money: similarities and differences. Social Psychological Quarterly 62 (3) (1999) 276-290
- 16. Melandri, V.: Fundraising course materials. Bologna D.U. Press (in Italian) (2004)
- 17. Mitani Y, Hamamoto Y.: A local mean-based nonparametric classifier. Pattern Recognition Letters **27** (2006) 1151-1159
- 18. Nudd, S. P.: Thinking strategically about information. In Tempel, E. (eds.): Hank Rosso's achieving excellence in fund raising, John Wiley & Sons (2003) 349-365
- Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach (2nd ed.), Prentice Hall (2003)
- Sanchez J. S., Pla F., Ferri F. J.: On the use of neighbourhoodbased non-parametric classifiers. Pattern Recognition Letters 18 (11-13)(1997) 1179-1186
- 21. Verhaert, G. A., Van den Poel, D.: The role of seed money and threshold size in optimizing fundraising campaigns: Past behavior matters!. Expert Systems with Applications **39** (2012) 13075-13084

Mathematical Methods in Economics and Finance – m²ef Vol. 15/16, No. 1, 2020/2021

ISSN print edition: 1971-6419 — ISSN online edition: 1971-3878 Web page: http://www.unive.it/m2ef/ — E-mail: m2ef@unive.it