PROCEEDINGS

SEKE 2013

The 25th International Conference on Software Engineering & Knowledge Engineering

Sponsored by

Knowledge Systems Institute Graduate School, USA

Technical Program

June 27 - 29, 2013

Hyatt Harborside at Logan Int'l Airport, Boston, USA

Organized by

Knowledge Systems Institute Graduate School

ISBN-10: 1-891706-33-0 (paper)
ISBN-13: 978-1-891706-33-2

ISSN: 2325-9000 (print)

Additional Copies can be ordered from:
Knowledge Systems Institute Graduate School
3420 Main Street
Skokie, IL 60076, USA
Tel:+1-847-679-3135
Fax:+1-847-679-3166
Email:office@ksi.edu
http://www.ksi.edu

Proceedings preparation, editing and printing are sponsored by
Knowledge Systems Institute Graduate School

Printed by Knowledge Systems Institute Graduate School

Proposal and Validation of a Feasibility Model for Information Mining Projects

Pablo Pytel
PhD Program on Computer Science, UNLP - GISI UNLA - GEMIS UTN-FRBA, Buenos Aires, Argentina.
ppytel@gmail.com

Paola Britos
Information Mining Research Group. National University of Rio Negro at El Bolson, Argentina.
paobritos@gmail.com

Ramón García-Martínez
Information Systems Research Group, National University of Lanus, Remedios de Escalada, Argentina.
rgarcia@unla.edu.ar

***Abstract*— Information Mining projects are a special type of Software Engineering projects with the objective of extracting non-trivial knowledge from available data repositories. Information mining projects share similar problems with Software Engineering projects. Most of these problems should be handled at the initial activities of the project. But there is no model to analyze and evaluate the project feasibility which could be used to predict the main project risks. In this context, the objective of this paper is to propose and validate an ad-hoc model that can be used at the beginning of an Information Mining project in order to analyze its feasibility.**

Keywords-Feasibility Model; Information Mining; Software Engineering; Small and Medium-sized Enterprises.

I. INTRODUCTION

Information Mining projects are a special type of Software Engineering projects with the objective of extracting non-trivial knowledge which is located (implicitly) in the available data from different sources [1]. Commonly, instead of developing specific software, available software tools are used which include the necessary techniques and algorithms [2]. As a result, the features of Information Mining projects are different from Traditional Software Engineering projects and, also, from Knowledge Engineering projects, even though the algorithms are based on artificial intelligence methods [3]. The most used methodologies for Information Mining projects are CRISP-DM [4], SEMMA [5] and P^3TQ [6]. These methodologies are considered as proven by the community, but they exhibit problems when trying to define the phases related to project management [7]. Some elements of project management are mixed with project development process and tasks (such as project monitoring, verification and measurement) and others are not considered in the referenced methodologies.

In this context, it is detected the lack of a model to analyze and evaluate the project feasibility which could be used to predict the project risks. Therefore the objective of this paper is to propose and validate an ad-hoc model that can be used at the beginning of an Information Mining project in order to analyze its feasibility and then identify its strong and weak points. First, the problem is identified by analyzing the reasons of project failures (section II). Then the corresponding solution is proposed by specifying the feasibility model (Section III) which is validated using real project information (Section IV). Finally the main conclusions are presented (Section V).

II. ANALYSIS OF PROJECT FAILURE

Most Traditional Software Engineering projects can be considered (at least) partial failures because few projects meet all their cost, schedule, quality, or requirements objectives [8]. From the challenged or canceled projects, the average project was 189 percent over budget, 222 percent behind schedule, and contained only 61 percent of the originally specified features. In 2005, it has been considered that from 5 to 15 percent of projects were abandoned before or shortly after delivery as hopelessly inadequate [9]. In other words, few projects truly succeed. The most important reasons are, among others: inability to manage the project complexity, poorly defined system requirements, sloppy development practices, commercial pressures, unrealistic / unarticulated project goals and unmanaged risks.

Information Mining projects share similar problems. Conducted studies about Information Mining projects have detected that not all projects are successfully completed [10], ending most in failure [11]. In 2000, it has estimated that 85% of the projects have failed to achieve its goals [12]. In other words, this means that in average only 15 projects had been successfully completed from 100 developed ones. After five years working, the community has been able to decrease this project failure rate to approximately 60% [13]. Hence it can be said that the community is working on the right lane but there are still project elements that should be enhanced.

Most of these problems have to be detected in the initial activities of the project. Before starting any Traditional Software project, the organization normally decides whether it is appropriate performing it or not. Making such decisions is complex and depends on multiple factors as it is necessary to know both the impact of the software on the organization and the developing associated risks [14]. The project features are required to be analyzed by assessing the technical and economic feasibility of the project (commonly known as feasibility study). In Expert System development projects, something similar happens. As the initial specifications for these systems are often uncertain, incomplete, and inconsistent, it is necessary to develop several prototypes for coherently define the system functionality, performance, and interfaces

[15]. Since Knowledge Engineering (KE) projects use more resources than Traditional Software development projects [16], their feasibility study is highly more important in order to identify the risks that should be monitored and controlled during the project.

Information Mining project's initial tasks ought to be similar to a Traditional Software Development or KE projects. By early detection of the risks, its effects could be reduced during the project development. However, as the features of Information Mining projects are different from Traditional Software and KE projects, the models to study the feasibility cannot be reused for this type of projects and it is necessary to propose specific ones. In this domain, several studies have been conducted to identify success criteria [17-20], but there is no comprehensive model to analyze if the project is achievable or not. While in [21] a model that uses a fuzzy expert system has been proposed to measure the project success level from the quality used in each of its phases of CRISP-DM, this study may be performed once the project is completed because it requires the quality level applied in each phase to be known. On the other hand, in [22] a Bayesian analysis is used to determine whether the company is qualified to implement a data mining project (i.e. the enterprise characteristics are valued to decide whether data mining can be applied or not), but it does not consider important topics such as the business problem. Furthermore, this analysis does not make the classification of feasibility in different dimensions, considering it as a whole.

III. Proposed Feasibility Model

The proposal of the feasibility model for Information Mining projects requires the identification of the main conditions for considering a project as feasible (subsection A). Such a task is dependent on the project features which can be known in the initial stages. However, it is not usually easy to answer these conditions by answering 'yes' / 'no' questions (or by giving a numerical value). Thus the proposed feasibility model should be able to handle a range of linguistic values to answer each condition. From such values, and by applying a pre-defined process, it would be possible to determine the overall project feasibility (as detailed in subsection B).

This model has been based on the feasibility test defined for KE projects defined in [15, 16] which has been adapted and validated using actual Information Mining projects. These projects have been provided by researchers from the following research groups: GISI-DDPyT-UNLa, GEMIS-FRBA-UTN and GIEdI-UNRN. Be advised that all these projects had been performed by applying the CRISP-DM methodology [4] within Small or Medium-size Enterprises (SMEs). Therefore, the models can be considered reliable only for small and medium-sized Information Mining projects developed with this methodology.

A. Conditions

The main conditions are identified based on [17-24] and classified into three groups (or dimensions):

- *Conditions that determine the plausibility of the project* include the factors that make it possible to perform the information mining project. A project can be performed if the following conditions are met: the available data repositories have current and representative data of the business problem to be solve, the business problem is understood, and the team has a minimum knowledge about the information mining process.
- *Conditions that determine the adequacy of the project* include the factors that determine whether information mining is the appropriate solution for the identified business problem (i.e. it is the best solution for the problem). It is appropriate to apply information mining when the following conditions are met: the available data repositories have digital format (they are not only in paper), the business problem cannot be solved by using traditional statistical techniques, the business problem will not change during the project, and the data quality is good. The following metrics are used for assessing the data quality:
  - Quantity of attributes and records (measuring the availability of enough data to apply the data mining process).
  - Degree of credibility of the data (measures of how much you can trust on the data accuracy depending on the source and nature).
- *Conditions that determine the success of the project* include the factors ensuring the project accomplishment. An information mining project will be successful if the following conditions are met: data repositories are implemented with technologies allowing easy data access and manipulation (i.e. integration, cleaning, and formatting tasks), the project stakeholders (either high level managers, mid-level managers or end-users) support the project, it is possible to perform the project planning considering best practices with necessary required time, and the team has experience in similar projects.

B. Proposed Procedure

The following five steps are proposed to assess the project feasibility:

<u>Step 1:</u> *Determining the value of each project features.*
Looking for characterizing an Information Mining project and evaluating its feasibility, 13 features are used which are specified in Table I. Such features, based on the conditions identified in subsection A, should be answered from the interviews conducted with the organization stakeholders at the beginning of the 'Business Understanding' phase of CRISP-DM methodology. They should be valued by using one of the following words: 'nothing', 'little', 'regular', 'much', and 'all'.

For each feature the following attributes are defined:

- *Category*: used only to group the features according to what or who is concerned.
- *ID*: indicates a code to uniquely identify the property and the dimension to which it belongs (Plausibility, Adequacy, or Success).
- *Condition*: describes the question associated to the feature to be identified for characterizing the project.
- *Weight*: indicates the relative importance of each feature in the global model.

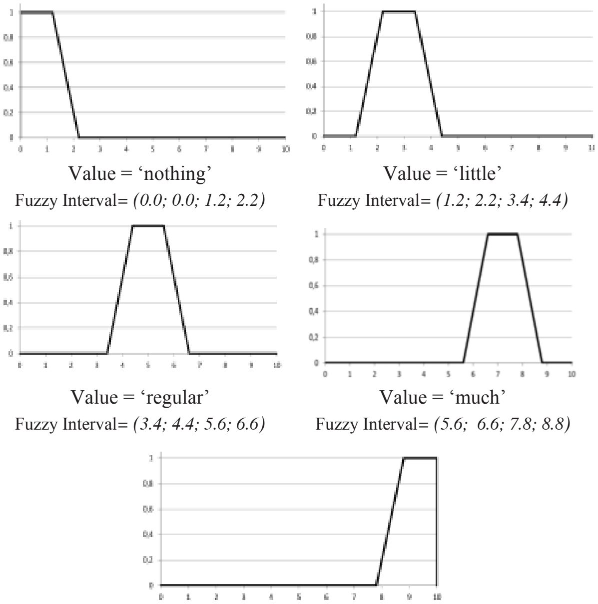
Note that features related to *plausibility* and *adequacy* dimensions must have a value equal to or bigger than 'little', otherwise the project can be considered as non-feasible. For *success* features, there is no minimum threshold (they can be valued with 'nothing').

TABLE I. PROJECT FEATURES EVALUATED BY THE MODEL.

Category	ID	Condition	Weight
Data	P1	How much actual is considered the data from the repositories?	*8*
	P2	How representative is considered the data in the repositories in order to solve the business problem?	*9*
	A1	How much the data repositories have digital format?	*4*
	A2	How many attributes and records are available in the data repositories?	*7*
	A3	How much credibility has the available data?	*8*
	S1	In which degree the repository technology supports the manipulation of the data?	*6*
Business Problem	P3	How much the business problem is understood?	*7*
	A4	In which degree the business problem cannot be solved by traditional statistical techniques?	*10*
	A5	How stable is considered the business problem during the project?	*9*
Project	S2	How much the stakeholders support the project?	*8*
	S3	In which degree the project plan considers the required time to perform best practices during the project?	*7*
Project Team	P4	How much knowledge has the team about information mining?	*6*
	S4	How much experience has the team in similar projects?	*6*

Step 2: *Converting feature values into fuzzy intervals.*
Once the linguistic values have been defined for each feature of Table I, they should be translated into numeric values to calculate the project feasibility. This transformation process is based on Fuzzy Expert systems [25]. For each word value, a fuzzy interval is specified that is expressed by four numbers (ranging from zero to ten) representing the breakpoints (or corner points) of the corresponding membership function. These intervals with the graphic representation of each membership function are shown in Figure 1.

Step 3: *Calculating the value of each dimension.*
To calculate the project dimension values, the fuzzy intervals (obtained in the previous step) are balanced considering their corresponding weight (already defined in Table I). The interval representing the value of each dimension (I_d) is calculated with the Formula #1 of Table II. This formula is formed by the combination of the harmonic mean and the arithmetic mean of the set of intervals (thus the influence of low values is reduced when calculating the dimension value). As a result of the formula, another fuzzy interval is achieved. To convert this interval into a single numeric value (V_d) the arithmetic average is used as specified in Formula #2 of the same table.



Value = 'nothing' Fuzzy Interval= *(0.0; 0.0; 1.2; 2.2)*

Value = 'little' Fuzzy Interval= *(1.2; 2.2; 3.4; 4.4)*

Value = 'regular' Fuzzy Interval= *(3.4; 4.4; 5.6; 6.6)*

Value = 'much' Fuzzy Interval= *(5.6; 6.6; 7.8; 8.8)*

Value = 'all' Fuzzy Interval = *(7.8; 8.8; 10.0; 10.0)*

Figure 1. Membership function graphical and fuzzy interval assigned.

TABLE II. FORMULAS USED BY THE MODEL.

#	Formula
1	$I_d = \left(\frac{1}{2} \cdot \frac{\sum_{i=1}^{n_d} W_{d_i}}{\sum_{i=1}^{n_d} \left(\frac{W_{d_i}}{F_{d_i}} \right)} \right) + \left(\frac{1}{2} \cdot \frac{\sum_{i=1}^{n_d} (W_{d_i} \cdot F_{d_i})}{\sum_{i=1}^{n_d} W_{d_i}} \right)$ where: I_d: represents the fuzzy interval calculated for the dimension *d* (using '$_P$' for plausibility, '$_A$' for adequacy, and '$_S$' for success). W_{di}: represents the weight of the feature *i* for the dimension *d*. F_{di}: represents the fuzzy interval that has been assigned to the feature *i* for the dimension *d*. n_d: represents quantity of features associated to the dimension *d*.
2	$V_d = \frac{\sum_{i=1}^{4} I_{d_i}}{4}$ where: V_d: represents the numeric value calculated for the dimension *d*. I_{di}: represents the value of the position *i* of the fuzzy interval calculated for the dimension *d* .
3	$OV = \frac{8 \cdot V_P + 8 \cdot V_A + 6 \cdot V_S}{22}$ where: *OV:* represents the overall project feasibility value. V_P: represents the value calculated for dimension plausibility. V_A: represents the value calculated for dimension adequacy. V_S: represents el value calculated for dimension success.

Step 4: *Calculating the overall project feasibility.*

The numerical values calculated in the previous step for each dimension (V_d) are combined by using a weighted arithmetic mean (Formula #3 of Table II) obtaining the overall project feasibility value (OV).

Step 5: *Interpreting the results.*

Finally, the numeric values for each dimension and the overall project feasibility value (already calculated in steps 3 and 4 respectively) are analyzed. As a way to interpret the results of the feasibility of each dimension, it is recommended to plot the corresponding membership function of the obtained fuzzy interval (I_d). The feasibility of a dimension can be considered as accepted if it exceeds the range of '*regular*' value (that has been shown in Figure 1). Examining the numeric value of the dimension is another way to do it. If the dimension value (V_d) is greater than 5, the dimension can be considered as accepted.

On the other hand, for analyzing the feasibility of the project, the following criteria can be used: whenever the three dimensions are accepted and the overall project feasibility (OV) is greater than 5, then the project is considered as feasible. Otherwise, it is not feasible. In both cases, the project weaknesses to be strengthened should be recognized by identifying the dimensions with lower values.

IV. Validation of the Proposed Model

In this section the validation of the model proposed in Section III is performed using the data of 25 collected information mining projects. The first twenty-two projects (i.e. *P1* to *P22*) have been satisfactory finished (with some minor problems) but the last three (i.e. *P23*, *P24* and *P25*) have been cancelled before completion.

To perform this validation the projects values calculated by the proposed model are compared to an appraisal provided by researchers who can be considered experts in the domain. On one hand, the projects have been characterized by the paper authors using the model's features and applying the corresponding steps to calculate the project dimensions and the project feasibility value. Because of the paper's strict length, the results of applying the model in each project cannot be reproduced here but it can be found in [26]. A summary of these results is included in Table III.

On the other hand, a survey has been issued to each researcher to assess one project. The researcher had examined the project information (including the plan, meeting notes, status reports among other things) and indicate a value between 1 and 10 (where 1 is the lowest value and 10 the biggest) to appraise each project dimension (i.e. plausibility, adequacy, success). Then the global feasibility has been calculated as the average of them. The obtained values are shown in Table IV.

As soon as the previous values has been collected, the chart graphs of Figure 2 have been prepared to show graphically the comparison between the values appraised by the researchers (shown with a light gray line) and the values calculated by the model (dark gray bar) per dimension. As it can be seen the values are very similar but not totally equal. This can also be noted in the boxplot graphs included in Figure 3. These graphs reflect the behavior of the values assigned by the researchers (locate in the left part of the graph) and the ones calculated by the model (right part) indicating the minimum and maximum values (thin line), standard deviation range (thick line) and average value (marker).

TABLE III. PROJECT VALUES CALCULATED BY THE PROPOSED MODEL.

#	Value of Plausibility (V_P)	Value of Adequacy (V_A)	Value of Success (V_S)	Overall Project Feasibility (OV)
P1	*7.20*	*6.11*	*5.25*	*6.27*
P2	*6.87*	*5.07*	*5.25*	*5.77*
P3	*5.90*	*5.67*	*5.31*	*5.65*
P4	*5.12*	*6.95*	*4.12*	*5.51*
P5	*5.12*	*7.82*	*6.81*	*6.56*
P6	*5.45*	*5.61*	*5.25*	*5.45*
P7	*5.45*	*5.56*	*5.42*	*5.48*
P8	*6.45*	*5.80*	*5.18*	*5.87*
P9	*7.20*	*5.61*	*5.57*	*6.18*
P10	*5.85*	*5.34*	*5.57*	*5.59*
P11	*6.22*	*6.56*	*5.42*	*6.14*
P12	*7.67*	*7.35*	*6.45*	*7.22*
P13	*5.93*	*5.09*	*7.05*	*5.93*
P14	*6.20*	*6.59*	*5.69*	*6.20*
P15	*8.72*	*6.89*	*7.66*	*7.77*
P16	*6.45*	*6.43*	*5.64*	*6.22*
P17	*6.14*	*5.83*	*5.42*	*5.83*
P18	*6.00*	*5.31*	*5.42*	*5.59*
P19	*7.01*	*6.89*	*5.58*	*6.58*
P20	*8.24*	*6.75*	*5.52*	*6.96*
P21	*8.05*	*6.45*	*5.25*	*6.70*
P22	*6.45*	*5.81*	*6.54*	*6.24*
P23	*4.66*	*5.34*	*3.25*	*4.52*
P24	*4.66*	*3.46*	*4.21*	*4.10*
P25	*4.63*	*2.81*	*3.01*	*3.52*

TABLE IV. PROJECT APPRAISAL PROVIDED BY RESEARCHERS.

#	Plausibility Value	Adequacy Value	Success Value	Global Feasibility Value
P1	*8.00*	*7.00*	*4.00*	*6.33*
P2	*7.00*	*6.00*	*5.00*	*6.00*
P3	*8.00*	*5.00*	*6.00*	*6.33*
P4	*6.00*	*6.00*	*4.00*	*5.33*
P5	*6.00*	*8.00*	*7.00*	*7.00*
P6	*6.00*	*5.00*	*5.00*	*5.33*
P7	*5.00*	*5.00*	*5.00*	*5.00*
P8	*6.00*	*5.00*	*6.00*	*5.67*
P9	*7.00*	*6.00*	*6.00*	*6.33*
P10	*6.00*	*5.00*	*6.00*	*5.67*
P11	*8.00*	*5.00*	*6.00*	*6.33*
P12	*7.00*	*8.00*	*7.00*	*7.33*
P13	*7.00*	*5.00*	*6.00*	*6.00*
P14	*7.00*	*7.00*	*6.00*	*6.67*
P15	*9.00*	*7.00*	*8.00*	*8.00*
P16	*7.00*	*6.00*	*5.00*	*6.00*
P17	*6.00*	*5.00*	*5.00*	*5.33*
P18	*5.00*	*5.00*	*6.00*	*5.33*
P19	*8.00*	*7.00*	*7.00*	*7.33*
P20	*9.00*	*7.00*	*5.00*	*7.00*
P21	*8.00*	*6.00*	*5.00*	*6.33*
P22	*7.00*	*6.00*	*6.00*	*6.33*
P23	*3.00*	*4.00*	*3.00*	*3.33*
P24	*5.00*	*3.00*	*2.00*	*3.33*
P25	*4.00*	*2.00*	*1.00*	*2.33*

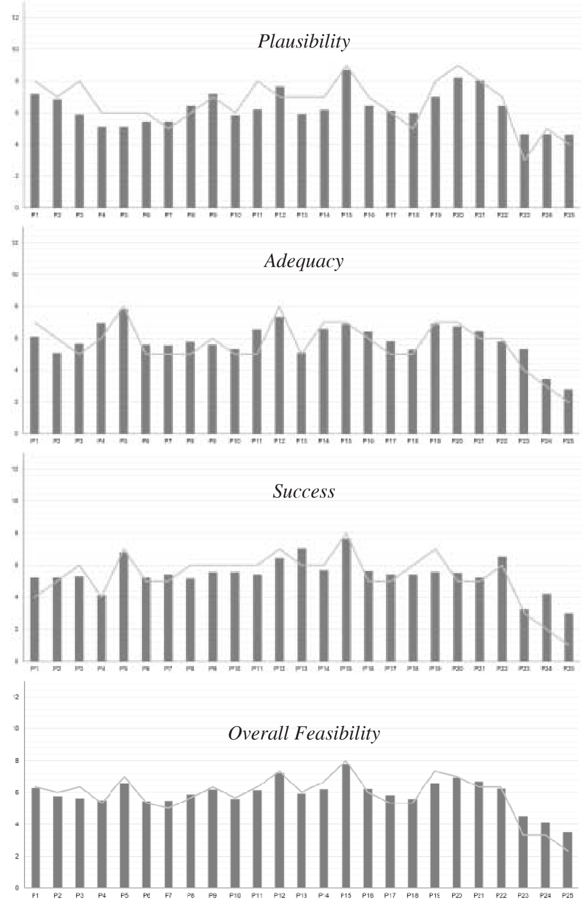


Figure 2. Comparison graph for each dimension

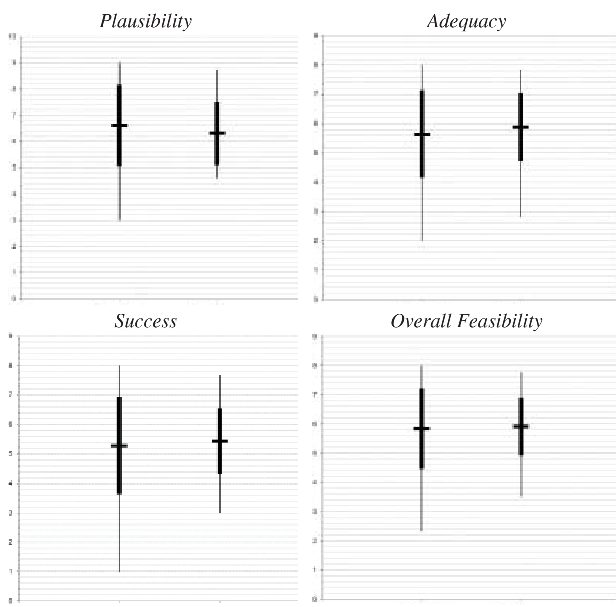


Figure 3. Boxplot graph for each dimension

As seen in the boxplot graphs of Figure 3, the model tends to be more conservative because the total range is shorter than the one assigned (particularly for the minimum values). But the standard deviation range and average values are almost the same (the bigger difference is lower than 0.30 for plausibility). Thus from this preliminary analysis it can be said that the model seems to be valid.

In order to assess the model, the Wilcoxon signed-rank test is applied [27]. This non-parametric statistical test allows to compare two related samples and define whether their population means differ (i.e. it is a paired difference test). It is an alternative to the paired Student's t-test when the population cannot be assumed to be normally distributed but there is a symmetric distribution of the differences around the median. In this test, each project dimension is handled independently. This means that for each dimension the values provided by the researchers are tested against the calculated by the model. The used null and alternative hypotheses are:

H_0: the valued assigned by the researchers and the values calculated by the model for each dimension have a median difference of zero (in other words, there are no meaningful differences between the researchers and the model values and they can be considered equivalent)

H_1: the median difference is not zero (i.e. the researchers and the model values are not equivalent)

The sums of signed-ranks generated by the application of the Wilcoxon test are shown in Table V for each dimension (where W^+ is the sum of all positive ranks and W^- is the sum of all negative ranks). All the auxiliary tables used in this test are available in [26].

TABLE V. RESULTS OF WILCOXON SIGNED-RANK TEST

Dimension	Sum Ranks$^+$ (W^+)	Sum Ranks$^-$ (W^+)	Quantity of non-zero pairs
Plausibility	*97*	*228*	*25*
Adequacy	*227*	*98*	*25*
Success	*175*	*150*	*25*
Overall Feasibility	*181*	*144*	*25*

The null hypotheses (H_0) is accepted or rejected based on comparison of the minimum sum of ranks (W) and a critical value extracted from the statistical reference table corresponding to the quantity of non-zero pairs and a level of significance. If W is lower than or equal to the critical value then the null hypotheses can be rejected (so in this case, it means that the model is not equivalent to the assessment of the researcher’s appraisal). Otherwise the null hypotheses can be considered as valid (and, in this case, the model can be considered as equivalent). For all dimensions, there are not any zero-value pair, so the total of pairs is 25 ($n=25$). As a 0.01 level of significance is selected, the critical value is equal to 68. This value is then compared to the minimum sum of ranks for each dimension:

- For *Plausibility*, the minimum sum of ranks (W) is equal to *97* because W^+ is lower than W^-. As 97 is bigger than 68, the null hypotheses is not rejected and then it can be said that there is no meaningful differences between the researchers and the model plausibility values and they can be considered equivalent.

- For *Adequacy*, the minimum value is of $W^{-} = 98$ which is also bigger than 68. This means that H_0 is not rejected and the model adequacy values are also valid.
- For *Success* happens something similar: $W = W^{-} = 150$ is also bigger than 68. This means that the success values are also significant.
- Finally, the *Project Overall Feasibility* values calculated by the model value can also be considered equivalent because $W = W^{-} = 144 > 68$.

Therefore it is confirmed that the proposed model has calculated values equivalent to the appraisal performed by the experts.

V. Conclusions

Information Mining projects are a special type of Software Engineering projects with the objective of extracting non-trivial knowledge from available data repositories. Conducted studies for these projects have detected that not all projects are successfully completed, ending most in failure.

This paper has the objective of proposing an ad-hoc model to be used at the beginning of Information Mining project in order to analyze its feasibility. Thirteen projects features (based on the project conditions) are defined and utilized in a procedure to calculate the project feasibility with three dimensions: *plausibility*, *adequacy* and *success*. As it is difficult to assign the features values at the beginning of the project, the proposed procedure considers using fuzzy intervals to calculate the project overall feasibility.

The proposed model has been validated using the information of 25 projects which have been appraised by expert researchers. A preliminary statistical comparison and the Wilcoxon signed-rank test have been applied. As a result it is found that the proposed model can estimate correctly the *plausibility*, *adequacy*, *success* and *overall feasibility* of the project in the initial steps.

Acknowledgements

The research reported in this paper has been partially granted by Research Projects 33A105, 33B102 and 33A167 within National University of Lanus, Research Project 40B133 within National University of Rio Negro, and Research Project EIUTIBA11211 within UTN at Buenos Aires.

References

[1] J. Schiefer, J. Jeng, S. Kapoor and P. Chowdhary, "Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence", Proceedings IEEE International Conference on E-Commerce Technology, pp. 162-169, 2004.

[2] R. García-Martínez, P. Britos, P. Pesado, R. Bertone, F. Pollo-Cattaneo, D. Rodríguez, P. Pytel and J. Vanrell, "Towards an Information Mining Engineering", Software Engineering, Methods, Modeling and Teaching, Sello Editorial Universidad de Medellín, pp. 83-99, 2011, ISBN 978-958-8692-32-6.

[3] R. Garcia-Martinez, P. Britos, F. Pollo-Cattaneo, D. Rodriguez and P. Pytel, "Information Mining Processes Based on Intelligent Systems", Proceedings of II International Congress on Computer Science and Informatics (INFONOR-CHILE 2011), pp. 87-94, 2011, ISBN 978-956-7701-03-2.

[4] P. Chapman, J. Clinton, R. Keber, T. Khabaza, T. Reinartz, C. Shearer and R. Wirth, "CRISP-DM 1.0 Step by step BI guide", Edited by SPSS, 2000. http://tinyurl.com/crispdm

[5] SAS. "SAS Enterprise Miner: SEMMA", 2008. http://tinyurl.com/semmaSAS

[6] D. Pyle, "Business Modeling and Business intelligence", Morgan Kaufmann, 2003.

[7] J. Vanrell, R. Bertone, R. García-Martínez, "Un Modelo de Proceso de Operación para Proyectos de Explotación de Información", Proceedings Latin American Congress on Requirements Engineering and Software Testing, pp. 46-52, 2012, ISBN 978-958-46-0577-1.

[8] L.J. May, "Major causes of software project failures", CrossTalk: The Journal of Defense Software Engineering, 11(6), pp. 9-12, 1998.

[9] R.N. Charette, "Why software fails", Spectrum, IEEE, 42(9), pp. 42-49, 2005.

[10] H.A.Edelstein and H.C. Edelstein, "Building, Using, and Managing the Data Warehouse", Data Warehousing Institute, Prentice-Hall PTR, EnglewoodCliffs (NJ), 1997.

[11] M. Strand, "The Business Value of Data Warehouses - Opportunities, Pitfalls and Future Directions". Ph.D. Thesis, Department of Computer Science, University of Skovde, 2000.

[12] U.M. Fayyad, "Tutorial report". Summer school of DM. Monash University (Australia), 2000.

[13] J.E. Gondar, "Metodología del Data Mining". Number 84-96272-21-4. Data Mining Institute S.L., 2005.

[14] R. Pressman, "Software Engineering: A Practitioner's Approach", Editorial Mc Graw Hill, 2004.

[15] R. García Martínez and P. Britos, "Ingeniería de Sistemas Expertos". Editorial Nueva Librería, 2004, ISBN 987-1104-15-4.

[16] A. Gómez, N. Juristo, C. Montes and J. Pazos, "Ingeniería del Conocimiento", Centro de Estudios Ramón Areces. S.A. (Madrid), 1997.

[17] J. Sim "Critical success factors in data mining projects". Ph.D. Thesis, University of North Texas, 2003.

[18] H.R. Nemati and C.D. Barko, "Key factors for achieving organizational data-mining success". Industrial Management & Data Systems, 103(4), pp. 282-292, 2003, doi:10.1108/02635570310470692.

[19] T.H. Davenport, "Make Better Decisions", Harvard Business Review, (November), pp. 117-123, 2009.

[20] U.Bolea, J. Jakličb, G. Papac and J. Žabkard, "Critical Success Factors of Data Mining in Organizations", Ljubljana, 2011.

[21] A. Nadali, E.N. Kakhky and H.E. Nosratabadi, "Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system", Electronics Computer Technology (ICECT), 3rd International Conference on Kanyakumari, Vol. 6, pp. 161-165, IEEE, 2011, doi:10.1109/ICECTECH.2011.5942073.

[22] G. Nie, L. Zhang, Y. Liu, X. Zheng and Y. Shi, "Decision analysis of data mining project based on Bayesian risk", Expert Systems with Applications, 36(3), pp. 4589-4594, 2009.

[23] L.L. Pipino, Y.W. Lee and R.Y. Wang, "Data quality assessment", Communications of the ACM, 45(4), pp. 211-218, 2002.

[24] N. Lavravc, H. Motoda, T. Fawcett, R. Holte, P. Langley and P. Adriaans, "Introduction: Lessons learned from data mining applications and collaborative problem solving", Machine learning, vol. 57, n.º 1, pp. 13-34, 2004.

[25] J.S.R. Jang, "Fuzzy inference systems", Upper Saddle River, NJ: Prentice-Hall, 1997.

[26] P. Pytel, "Datos Recopilados para Validación del Modelo de Viabilidad de Proyectos de Explotación de Información", Reporte Técnico GISI-TD-2011-01-RT-2012-02, Grupo de Investigación en Sistemas Información. Departamento de Desarrollo Productivo y Tecnológico, Universidad Nacional de Lanús (Argentina), 2012. http://tinyurl.com/valViab

[27] F. Wilcoxon, "Individual Comparisons by Ranking Methods", Biometrics 1, 80-83, 1945.