Pazmiño-Maji, R., García-Peñalvo, F. J., & Conde-González, M. Á. (2017). Statistical Implicative Analysis Approximation to KDD and Data Mining: A Systematic and Mapping Review in Knowledge Discovery Database Framework. In A. Schmidt, F. Laux, D. Hristovski, & S. Ohnishi (Eds.), Proceedings of the Ninth International Conference on Advances in Databases, Knowledge, and Data Applications, DBKDA 2017, (May 21 - 25, 2017 - Barcelona, Spain) (pp. 70-77). Wilmington, DE, USA: IARIA XPS Press.

Statistical Implicative Analysis Approximation to KDD and Data Mining:

A Systematic and Mapping Review in **K**nowledge **D**iscovery **D**atabase Framework

Rubén A. Pazmiño-Maji
Escuela Superior Politécnica de Chimborazo
Riobamba, Ecuador
Email: rpazmino@espoch.edu.ec

Francisco J. García-Peñalvo
Department of Computer Science
University of Salamanca
Salamanca, Spain
Email fgarcia@usal.es

Miguel A. Conde-González
Department of Computer Science
University of León
León, Spain
Email: miguel.conde@unileon.es

***Abstract*— According to Scopus, only in the year 2016, there were 15747 scientific papers about data mining and KDD. These have been and remain useful technologies. In this paper, we determine the approximation level of SIA to KDD and Data Mining. To this end, we have created an approximation framework based on definition and step process proposed by Fayyad. We use mapping review and systematic review from literature published in the last 5 years in bibliographic databases ACM, EBSCO, Google Scholar, IEEE, ProQuest, Scopus and WOS. We started with 200 papers and finally, 35 had all quality criteria. This paper also describes the SIA papers and identifies a series of future research in SIA, KDD and Data Mining.**

Keywords-Statistical Implicative Analysis; Knowledge Discovery Database; data mining, systematic review, mapping review.

I. INTRODUCTION

In recent years, our capacity to generate, transform, store, analyze and visualize data has increased, basically due to the high processing power and the low cost of the machines. However, within these huge and different types of data there is a lot of unknown information. The discovery of this information is possible thanks to Data Mining (DM), which among other sophisticated techniques applies artificial intelligence to find patterns and relationships within the data allowing the creation of models, that are abstract representations of reality. Common tasks in Knowledge Discovery in Databases (KDD) are rule induction, classification and clustering problems, recognition pattern, predictive modeling, dependency detection, and so on [1].

A. The KDD process for extracting knowledge

The first KDD workshop was in August 20, 1989, Detroit MI, USA, enabling researchers and practitioners to gather around KDD. We define the KDD process as [2]: "The non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data".

Data mining (DM) is a step in the general process to constituting the KDD process (see Figure 1). Data mining allows us to use specific algorithms for extracting patterns (including classification rules or trees, regression, clustering, sequence modeling, dependency, and line analysis) from data. Table I shows steps and subprocess in the KDD process [3]:

TABLE I: STEPS AND SUBPROCESS IN KDD

Step	Subprocess	Name
1		Learning the application domain
	A	Selection
2		Creating a target dataset
	B	Preprocessing
3		Data cleaning and preprocessing
	C	Transformation
4		Data reduction and projection
	D	Data Mining
5		Choosing the function of data mining
6		Choosing the data mining algorithm(s)
7		Data mining
	E	Interpretation/Evaluation
8		Interpretation
9		Using discovered knowledge

B. Statistical Implicative Analysis and the Knowledge discovery

Statistical Implicative Analysis (SIA) was created by Regis Gras [4], thirty eight years ago and has a set of data analysis tools that that allows us to approach knowledge on the basis of the information contained in the database (individuals and variables). The approach is performed starting from the generation of asymmetric rules [5] between variables and variables classes, represented by tables (clusters non-hierarchical) [6], graphs (association rules) [7] and dendrograms (hierarchical clusters, hierarchical oriented clusters) [8]. The statistical theory [9] and application of SIA are in continuous expansion and development. The SIA software tool is called CHIC [10] [11], the last Windows version is 7.0 and the CHIC free multiplatform version is called RCHIC [12]. SIA has an international group of active researchers since 2000 [13]. Usual CHIC functions are: Similarity Tree, Implicative Graph, Cohesion Tree and Reduction. Some of the complementary options implemented in CHIC are: the entropy is used when analyzing a large data sample; the supplementary variables are qualitative variables such as gender, education level or economic category; the contribution is used to know what are the subjects or classes of subjects more responsible for computed implications and the typicality indicates the typical subjects of the population for computed implications.

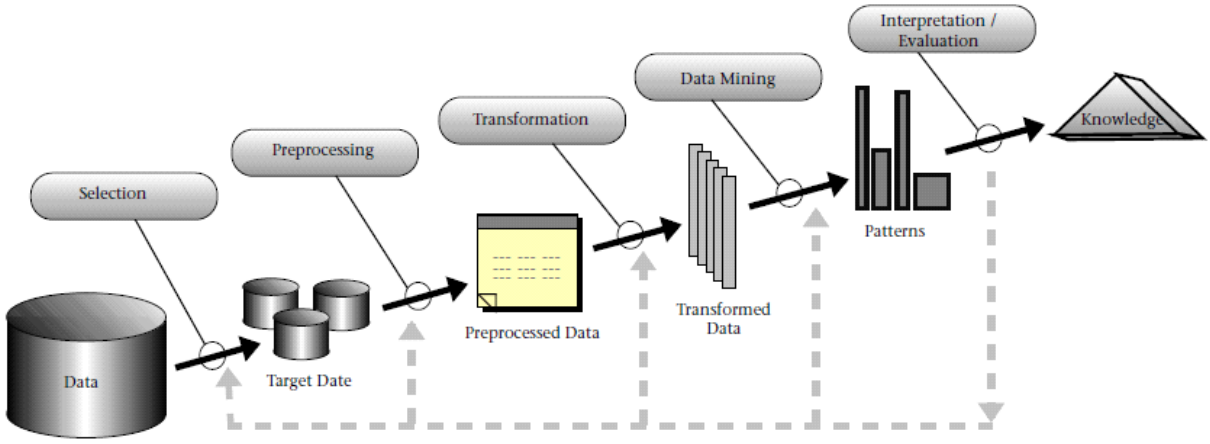


Figure 1. Overview of the steps constituting the KDD process [3].

The aim of this paper is to describe SIA papers in the last 5 years, to determine the position of SIA in KDD and DM, and identify a series of future research.

C. *Paper framework*

The approximation framework of Statistical Implicative Analysis to Knowledge Discovery and Data Mining are the KDD definition and the KDD steps proposed by Usama Fayyad [3]. This approximation allowed us to propose a new definition of SIA and to determine the steps that constitute the SIA process. We emphasize that this paper is a first approximation of SIA to Data Mining.

Section II describes the systematic and mapping review of literature and the steps in the research realized. Section III describes the results and their discussion. Finally, Section IV describes the conclusions.

II. METHOD

In this section we show in detail the steps of the methodology used.

A. *Systematic and mapping review of literature*

In the planning of systematic and mapping review the objectives were identified and the protocol was defined [14] [15]. The Protocol shows the method used in the systematic review and mapping to minimize the bias of researchers and that the methodology can be reproduced. Below we summarize the protocol used:

B. *Research questions*

The systematic mapping aims to answer these questions:
MQ1: What are the SIA papers by countries?
MQ2: Which are the SIA papers types?
MQ3: Which are the SIA papers Areas?
MQ4: What are the SIA papers tendency?

The systematic review aims to answer the question below:
RQ1: Which KDD step is the closest to SIA papers?
RQ2: Which KDD step is the furthest from SIA papers?
RQ3: How close are the SIA papers to KDD steps?
RQ4: How close are the SIA papers to KDD subprocess?
RQ5: How close are the SIA papers to Data Mining steps?
RQ6: Can you define the SIA based KDD definition?
RQ7: Can you define the process for SIA based KDD?

C. *PICOC method*

The paper of Petticrew and Roberts [16], proposed the PICOC method to define our scope:

- Population (**P**): Statistical implicative analysis papers in last five years (2012-2016).
- Intervention (**I**): SIA papers with explicit analysis process, in last five years (2012-2016).
- Comparison (**C**): No comparison intervention.
- Outcomes (**O**): SIA approximation percentages to KDD stages
- Context(**C**): SIA computational solutions.

D. *Time period*

The last 5 years (2012 to 2016)

E. *Sources*

The search was done in the following bibliographic databases [17]:

- EBSCO [18],
- Google Scholar [19],
- IEEExplore [20],
- ProQuest [21],
- Scopus [22],
- WOS [23],
- Web of Science [24],
- ACM [25].

To answer the research questions raised, the inclusion and exclusion criteria were defined. They also allowed us to select the source SIA papers.

F. *Inclusion and exclusion criteria*

The inclusion criteria (IC) [26] are presented below:

IC1: The papers used a SIA methods real application
IC2: The proposed solution is applied on specialized software (Chic, Rchic, etc.)
IC3: The SIA process application is possible to make explicit
IC4: The papers are written in English language
IC5: The papers are reported in peer reviewed Workshop or Conference or Journal or Technical Reports

The exclusion criteria are presented below:

EC1: The paper is essentially theoretical, historical or a literature review
EC2: The SIA process application is not possible to make explicit
EC3: The SIA papers analysis methods do not use computer programs
EC4: The papers are written in Spanish, Italian, Portuguese or French language.

G. *Search string*

The group of primary studies were defined [27]. The final search string was described as follows: ("statistical implicative analysis" OR SIA) AND (LIMIT-TO (PUBYEAR, 2016) OR (LIMIT-TO (PUBYEAR, 2015) OR (LIMIT-TO (PUBYEAR, 2014) OR (LIMIT-TO (PUBYEAR, 2013) OR (LIMIT-TO (PUBYEAR, 2012)) [28, 29] showed studies on control, if the search chain found relevant studies.

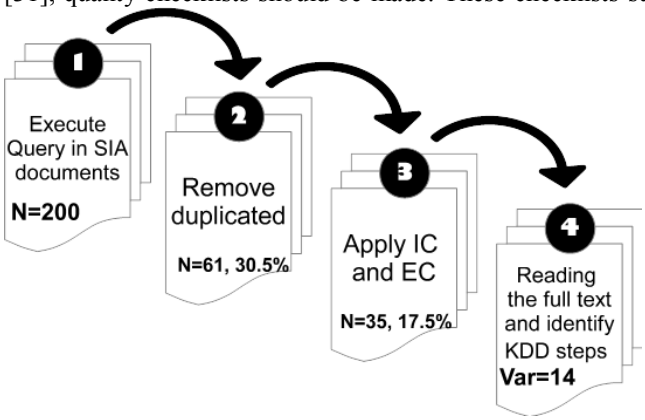
H. *Review and mapping steps*

The review and mapping were carried out and the details for this step are presented in the following subsection.

Figure 2 shows the steps of systematic and mapping review with SIA papers.

Figure 2. SIA papers mapping and review process

According to Kitchenham [30] and literature mapping [31], quality checklists should be made. These checklists su



pport the selection process. In this way, we produced the following checklist of quality.

I. *Quality assessment*

The quality assessment questions are presented below in Table II:

TABLE II: QUALITY ASSESSMENT QUESTIONS

Questions	Answers		
	Yes=1	*No=0*	*Half=0.5*
1. Are the SIA research goals clearly specified? 2. Are the research aims achieves? 3. Are the used data clearly described and their selection justified? 4. Are the pre-processed data in SIA papers clearly described? 5. Are the transformed data in SIA papers clearly described? 6. Are the SIA's papers algorithms clearly described and their selection justified? 7. Are the SIA's papers methods clearly described and their selection justified? 8. Is the data analysis process done by the computer? 9. How clear are the links between data, transformed data, analysis, interpretation and conclusions?			

III. RESULT AND DISCUSSION

In this section we show the results obtained in the mapping and systematic review process.

A. *Mapping literature review*

In this section we describe the results and their discussion about mapping and systematic literature review.

1) *What are the SIA papers by countries?*

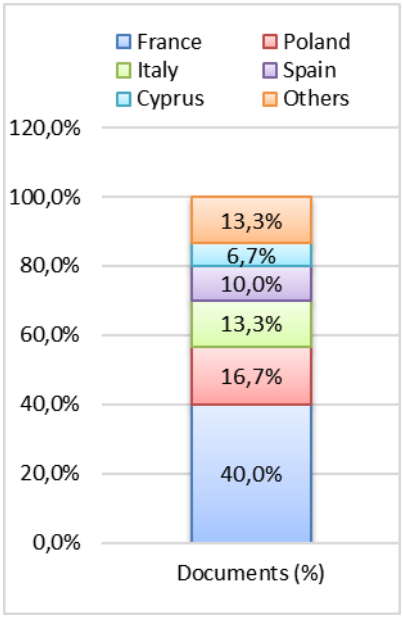


Figure 3. Percentage of SIA papers by countries [22]

Figure 3 shows that the most frequent countries in the selected literature are from France (40.0%) and Poland (16.7%). France is the most frequent country because SIA

theory was originated with Regis Gras born in France.

2) Which are the SIA papers types?

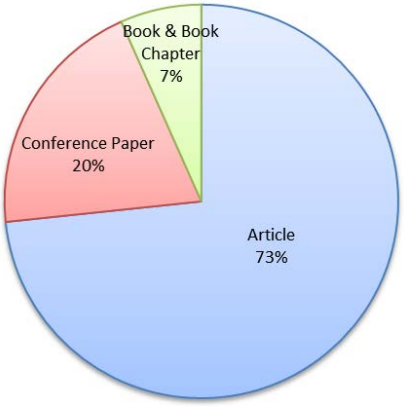


Figure 4. Percentage of SIA papers by type [22]

Figure 4 illustrates that studies from the chosen literature are 73% Articles, 20% Conference Papers, and 7% Book Chapter Books. This is because SIA international congress is producing new papers every two years.

3) Which are the SIA papers areas?

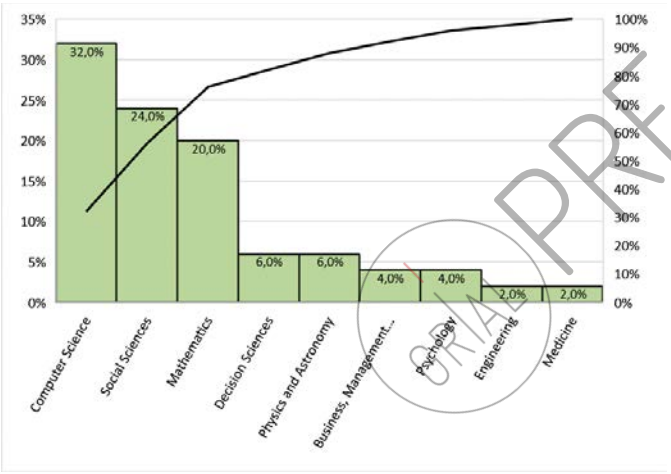


Figure 5. Percentage of SIA papers by Area [22]

Figure 5 illustrates that most of the studies have targeted

Computer Science (32.0%), Social Sciences (24.0%), Mathematics (23.6%) and Decision Sciences (6%). The four areas added are approximately 80%, this is because SIA theory was originated on didactic, mathematics and his methods are used in computer science.

4) What are the SIA papers tendency?

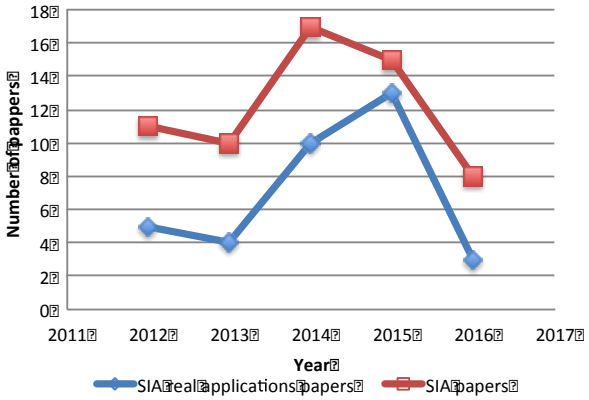


Figure 6. SIA papers by year

Figure 6 shows the tendency of SIA papers in general (in red) and SIA real applications papers (in blue), in the last five years. SIA real applications papers tend to increase because there are fewer the theoretical, historical or a literature review papers than real applications papers.

B. Systematic literarture review

In this section we describe the results and their discussion about systematic review.

1) Which KDD step is the closest to SIA papers.

Steps 1 (Learning the application domain) and steps 2 (Creating a target dataset) with 100 % both, are the steps closest to KDD steps. This is because, all SIA real applications papers need a goal and a target dataset.

2) Which KDD step is the furthest from SIA papers?

Step 7 (Data mining) with 31 %, is the step furthest to KDD steps. This is because, most papers showed not a Data Mining pattern clearly, had descriptive analysis, histograms, bar diagrams, box plots and SIA graphics like a similarity tree, cohesion tree, implicative graph or reduction.

3) How close are the SIA papers to KDD steps?

Table III shows the reference of 35 SIA papers, the compliance or not of KDD steps process 1 to 9 (see introduction, quality assessment and [18]). Also, Table III shows in the last two columns the numbers of positive compliance and the respective percentage. The final percentage is the 74,6 %, which is medium high close to KDD steps process. This means that KDD steps process and SIA steps process are the same in a 74,6 %.

TABLE III: SIA PAPERS & KDD STEPS RELATED TO TABLE I

SIA papers	KDD steps										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*N*	*%*
[32]	1	1	0	0	1	1	0	1	1	6	66,7
[33]	1	1	1	1	1	1	0	1	1	8	88,9
[34]	1	1	0	0	1	1	1	1	1	7	77,8
[35]	1	1	0	0	1	0	0	0	0	3	33,3
[36]	1	1	0	0	1	1	0	1	1	6	66,7
[37]	1	1	0	0	1	0	0	1	1	5	55,6
[38]	1	1	1	1	1	1	1	1	1	9	100,0
[39]	1	1	1	0	1	0	0	1	0	5	55,6
[40]	1	1	0	1	1	1	1	1	1	8	88,9
[41]	1	1	1	0	1	0	0	1	1	6	66,7
[42]	1	1	1	1	1	1	0	1	1	8	88,9
[43]	1	1	0	1	1	1	1	1	0	7	77,8
[44]	1	1	1	0	1	1	1	1	1	8	88,9
[45]	1	1	1	0	1	1	0	1	1	7	77,8
[46]	1	1	1	1	1	0	1	1	1	8	88,9
[47]	1	1	1	1	1	1	1	1	1	9	100,0
[48]	1	1	0	0	1	1	0	1	1	6	66,7
[49]	1	1	1	1	1	0	1	1	1	8	88,9
[50]	1	1	1	0	1	0	0	0	0	4	44,4
[51]	1	1	1	0	1	1	0	0	0	5	55,6
[52]	1	1	1	1	1	1	0	1	1	8	88,9
[53]	1	1	1	0	1	0	0	1	1	6	66,7
[54]	1	1	1	1	0	0	0	1	1	6	66,7
[55]	1	1	0	1	1	1	0	1	0	6	66,7
[56]	1	1	1	1	1	1	0	1	1	8	88,9
[57]	1	1	1	1	1	0	0	1	1	7	77,8
[58]	1	1	1	0	1	0	0	1	1	6	66,7
[59]	1	1	1	1	1	1	1	1	1	9	100,0
[60]	1	1	1	1	1	0	1	1	1	8	88,9
[61]	1	1	1	1	1	1	1	1	1	9	100,0
[62]	1	1	1	1	0	0	0	0	0	4	44,4
[63]	1	1	1	1	1	0	0	1	1	7	77,8
[64]	1	1	0	0	1	1	0	1	1	6	66,7
[65]	1	1	1	0	1	0	0	1	1	6	66,7
[66]	1	1	1	0	1	0	0	1	1	6	66,7

TABLE IV: SIA PAPERS & KDD SUBPROCESS RELATED TO TABLE I

SIA papers	KDD subprosses						
	A	*B*	*C*	*D*	*E*	*N*	*%*
[32]	1	1	1	1	1	5	100
[33]	1	1	1	1	1	5	100
[34]	0	1	1	1	1	4	80
[35]	1	1	1	0	1	4	80
[36]	1	1	1	1	1	5	100
[37]	1	1	1	1	1	5	100
[38]	1	1	1	1	1	5	100
[39]	1	1	1	1	1	5	100
[40]	1	1	1	1	1	5	100
[41]	1	1	1	1	1	5	100
[42]	1	1	1	1	1	5	100
[43]	1	1	1	0	1	4	80
[44]	1	1	1	1	1	5	100
[45]	1	1	1	1	1	5	100
[46]	1	1	1	1	1	5	100
[47]	1	1	1	1	1	5	100
[48]	1	1	1	1	1	5	100
[49]	1	1	1	1	1	5	100
[50]	1	1	1	0	1	4	80
[51]	1	1	1	0	1	4	80
[52]	1	1	1	1	1	5	100
[53]	1	1	1	1	1	5	100
[54]	1	1	0	1	1	4	80
[55]	1	1	1	1	1	5	100
[56]	1	1	1	1	1	5	100
[57]	1	1	1	1	1	5	100
[58]	1	1	1	1	1	5	100
[59]	1	1	1	1	1	5	100
[60]	1	1	1	1	1	5	100
[61]	1	0	1	1	1	4	80
[62]	1	1	0	0	1	3	60
[63]	1	1	1	1	1	5	100
[64]	1	0	1	1	1	4	80
[65]	1	1	1	1	1	5	100
[66]	1	1	1	1	1	5	100

4) How close are the SIA papers to KDD subprocess?

Table III shows the reference of 35 SIA papers, the compliance or not of KDD subprocess A to E (see introduction, quality assessment and [18]). Table IV shows in the last two columns the numbers of positive compliance and the respective percentage. The final percentage 94,2 %, means that KDD and SIA subprocess are very similar.

5) How close are the SIA papers to the Data Mining steps?

Observing step 5 (Choosing the function of data mining, 94%), step 6 (Choosing the data mining algorithm, 54%), step 7 (Data mining, 31%) and subprocess (Data mining, 86%) we have 66.4% of approach to Data Mining process. It is a medium high percentage. This is because the SIA methods can be similar to the Data Mining methods.

6) Can you define the SIA process based in KDD process?

The SIA definition based on the KDD process could be:

Statistical Implicative Analysis is the process of identifying valid, useful, and understandable, r-rules in data.

In the previous SIA definition, the words valid, useful, and understandable depend on the researcher interpretation of the SIA methods used and the results obtained. The novel word is not in a SIA definition because the use of SIA methods is novel in Data Mining.

7) Can you define the process for SIA based KDD?

Figure 7 shows the steps graphic constituting the SIA process (based in KDD).

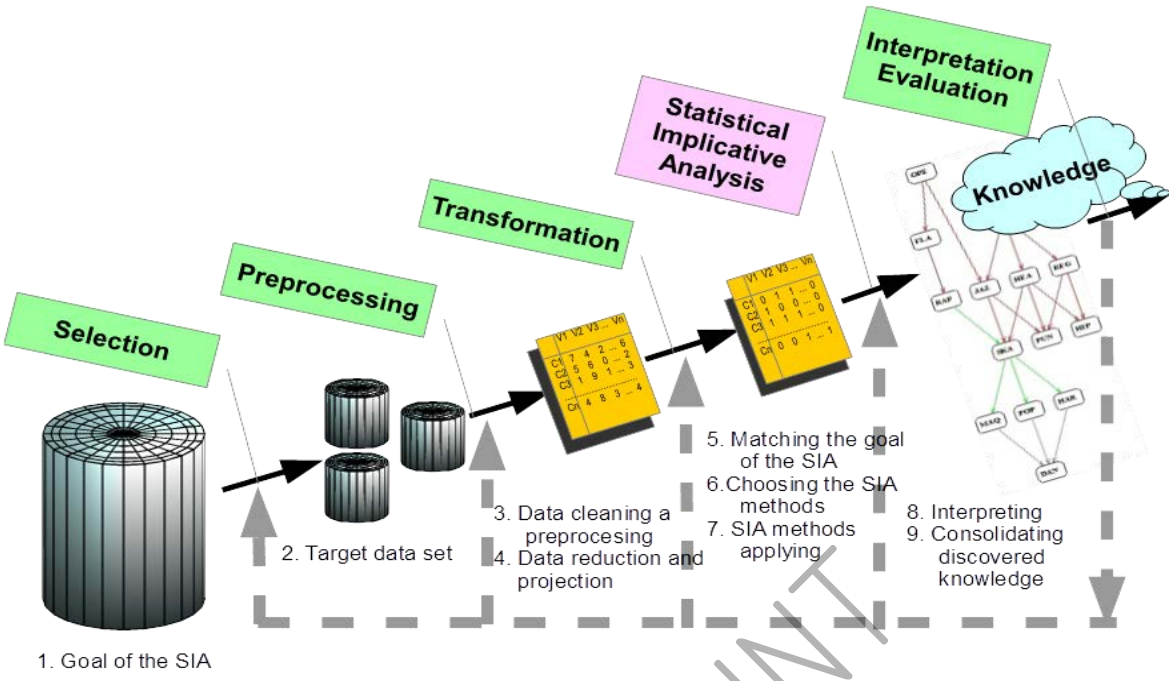


Figure 7. Proposal based in KDD, of the steps constituting the SIA process

IV. CONCLUSIONS

The aim of this paper is to describe SIA papers in last 5 years (2012-2016), to determine the approximation of SIA to KDD and DM, and identify a series of future research. To describe SIA papers, we use the mapping and systematic literature review methods. The most frequent country in the selected literature is France (40%); also 73% of the chosen literature are Articles. The studies were targeted in Computer Science (32.0%), Social Sciences (24.0%), Mathematics (23.6%), and Decision Sciences (6%). The four areas are approximately 80% of SIA papers total areas.

It is important to note that the results obtained depend closely on the approximation framework used, in this case the steps of the Faday process. It was determined that SIA and KDD are strongly related, with a contention relationship of SIA to KDD. In the first approximation, considering all steps without subprocesses, a contention of 74.6% corresponding to medium high approximation was obtained. In the second approach, considering all the subprocesses a contention of SIA to KDD of 94.2% was obtained that is high approximation to KDD process. This is summarized by indicating that the overall approximation rate considering the complete Faday process is 84.4% which is medium high approximation. The approximation achieved allowed to propose the steps that constitute the SIA process analysis, besides a new definition. The process D and steps 5 and 6 allowed us to give a first approximation of the SIA to data mining, obtaining 66.4% of contention between SIA and Data Mining corresponding to medium high percentage.

The future research in SIA, KDD and Data Mining can be to answer the following questions: Do SIA methods work with big data? Can we use similarity tree and cohesion tree like a data mining hierarchical cluster method? Can we use implicative graph like a data mining rule induction method? and Can we use reduction like a data mining cluster method?. For example we have the questions: Do SIA methods work with big data? Can we use similarity tree and cohesion tree like a data mining hierarchical cluster method? Can we use reduction SIA method like a data mining cluster method? and Can we use implicative graph like a data mining rule induction method?. The answer to the abode questions is future research in SIA, KDD and Data Mining.

ACKNOWLEDGMENT

We would like to thank the University of Salamanca PhD programme on Education in the Knowledge Society scope. Similarly, we want to thank Escuela Superior Politécnica de Chimborazo for funding to perform this research.

REFERENCES

[1] V. V. Asencios, "Data Mining y el descubrimiento del conocimiento," *Industrial Data*, vol. 7, pp. 083-086, 2014.

[2] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI magazine*, vol. 17, p. 37, 1996.

[3] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from

volumes of data," *Communications of the ACM,* vol. 39, pp. 27-34, 1996.

[4] T. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE translation journal on magnetics in Japan,* vol. 2, pp. 740-741, 1987.

[5] R. Gras, R. Couturier, F. Guillet, and F. Spagnolo, "Extraction de règles en incertain par la méthode statistique implicative," *Comptes rendus des 12èmes Rencontres de la Société Francophone de Classification,* pp. 148-151, 2005.

[6] R. Gras and S. A. Almouloud, "A implicação estatística usada como ferramenta em um exemplo de análise de dados multidimensionais," *Educ Mat Pesqui,* vol. 4, pp. 75-88, 2002.

[7] R. Gras, J.-C. Régnier, and F. Guillet, "Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités," *Toulouse (França): Cepadues,* 2009.

[8] G. Ritschard, "De l'usage de la statistique implicative dans les arbres de classification," *Troisieme Rencontre Internationale-Analyse Statistique Implicative,* pp. 305-316, 2005.

[9] M. Bailleul, "Des réseaux implicatifs pour mettre en évidence des représentations," *Mathématiques et sciences humaines. Mathematics and social sciences,* 2001.

[10] R. Couturier and S. A. Almouloud, "Historique et fonctionnalités de CHIC," ed, 2009.

[11] R. Couturier and R. Gras, "CHIC: traitement de données avec l'analyse implicative," in *EGC*, 2005, pp. 679-684.

[12] Fento-St. (2017, 2017/04/29/10:52:33). *Rchic - | Raphael Couturier*. Available: http://members.femto-st.fr/raphael-couturier/en/rchic

[13] J.-C. REGNIER. (2017, 2017-04-29 05:39:21). *A.S.I. 9 - 9th International Meeting Statistical Implicative Analysis*. Available: http://sites.univ-lyon2.fr/asi9/?page=1&lang=en

[14] F. W. Neiva, J. M. N. David, R. Braga, and F. Campos, "Towards pragmatic interoperability to support collaboration: A systematic review and mapping of the literature," *Information and Software Technology,* vol. 72, pp. 137-150, 2016.

[15] C. Okoli and K. Schabram, "A guide to conducting a systematic literature review of information systems research," *Sprouts Work. Pap. Inf. Syst,* vol. 10, p. 26, 2010.

[16] M. Petticrew and H. Roberts, *Systematic reviews in the social sciences: A practical guide*: John Wiley & Sons, 2008.

[17] C. Costa and L. Murta, "Version control in distributed software development: A systematic mapping study," in *Global Software Engineering (ICGSE), 2013 IEEE 8th International Conference on*, 2013, pp. 90-99.

[18] (2017/04/29/11:17:05). *EBSCOhost Login*. Available: http://search.ebscohost.com/

[19] (2017/04/29/11:22:57). *Google Académico*. Available: https://scholar.google.es/

[20] (2017/04/29/11:35:51). *IEEE Xplore Digital Library*. Available: http://ieeexplore.ieee.org/Xplore/home.jsp

[21] ProQuest. (2017, 2017/04/29/15:31:05). *ProQuest - Connect*. Available: http://search.proquest.com/

[22] (2017/04/29/11:41:48). *Scopus - Welcome to Scopus*. Available: https://www.scopus.com/home.uri

[23] (2017/04/29/11:45:14). *ScienceDirect.com | Science, health and medical journals, full text articles and books*. Available: http://www.sciencedirect.com/

[24] T. REUTERS. (2017, 2017/04/29/15:25:40). *Web of Science [v.5.24] - Colección principal de Web of Science*. Available: apps.webofknowledge.com

[25] A. f. C. Machinery. (2017, 2017/04/29/15:11:19). *ACM Digital Library*. Available: http://dl.acm.org/

[26] R. A. Pazmiño-Maji, F. J. García-Peñalvo, and M. A. Conde-González, "Approximation of statistical implicative analysis to learning analytics: a systematic review," in *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 2016, pp. 355-376.

[27] H. Zhang and M. Ali Babar, "On searching relevant studies in software engineering," 2010.

[28] A. Tolk, C. D. Turnitsa, and S. Y. Diallo, "Ontological implications of the levels of conceptual interoperability model," in *Proc. 10th World Multi-conf. on Systemics, Cybernetics and Informatics*, 2006, pp. 105-111.

[29] L. Kutvonen, "Tools and infrastructure facilities for controlling non-functional properties in inter-enterprise collaborations," in *Enterprise Distributed Object Computing Conference Workshops, 2008 12th*, 2008, pp. 423-432.

[30] B. Kitchenham, R. Pretorius, D. Budgen, O. P. Brereton, M. Turner, M. Niazi, and S. Linkman, "Systematic literature reviews in software engineering–a tertiary study," *Information and Software Technology,* vol. 52, pp. 792-805, 2010.

[31] C. A. Ellis, S. J. Gibbs, and G. Rein, "Groupware: some issues and experiences," *Communications of the ACM,* vol. 34, pp. 39-58, 1991.

[32] S. D. Anastasiadou, V. Batiou, and E. Valkanos, "Occupational Mobility Dimensions in Greece," *Procedia Economics and Finance,* vol. 19, pp. 325-331, 2015 2015.

[33] S. Anastasiadou and E. A. Panitsides, "AND NOW WHITHER..? EUROPEAN UNION LIFELONG LEARNING POLICY: A TWO LEVEL ANALYSIS," *The Economies of Balkan and Eastern Europe Countries in the changed World,* p. 41, 2014 2014.

[34] R. Belohlavek, D. Grissa, S. Guillaume, E. M. Nguifo, and J. Outrata, "Boolean factors as a means of clustering of interestingness measures of association rules," *Annals of Mathematics and Artificial Intelligence,* vol. 70, pp. 151-184, 2014 2014.

[35] N. Bonneton-Botte, H. Hili, F. De La Haye, and Y. Noel, "Drawings of the hand and numerical skills in children of preschool age," *Canadian Journal of Behavioural Science-Revue Canadienne Des Sciences Du Comportement,* vol. 47, pp. 207-215, Jul 2015.

[36] R. Coutrier, R. A. Pazmiño Maji, M. Á. Conde González, and F. J. García-Peñalvo, "Statistical implicative analysis for educational data sets: 2 analysis with RCHIC," 2015 2015.

[37] R. Couturier and R. Pazmiño, "Use of Statistical Implicative Analysis in Complement of Item Analysis," *International Journal of Information and Education Technology,* vol. 6, p. 39, 2016.

[38] T. Delacroix and A. Boubekki, "An application of multiple behavior SIA for analyzing data from student examsApplications multiples de l´ASI pour l´analyse des données des examens d´étudiants," *Educação Matemática Pesquisa,* vol. 16, 2014 2014.

[39] B. Di Paola, O. R. Battaglia, and C. Fazio, "Non-Hierarchical Clustering as a method to analyse an open-ended questionnaire on algebraic thinking," *South African Journal of Education,* vol. 36, pp. 1-13, 2016 2016.

[40] T.-N. Do, "Using Local Rules in Random Forests of Decision Trees," 2015, pp. 32-45.

[41] I. Elia, S. Özel, A. Gagatsis, A. Panaoura, and Z. E. Y. Özel, "Students' mathematical work on absolute value: focusing on conceptions, errors and obstacles," *ZDM - Mathematics Education,* vol. 48, pp. 895-907, 2016.

[42] C. Fazio, O. R. Battaglia, and B. Di Paola, "Investigating the quality of mental models deployed by undergraduate engineering students in creating explanations: The case of thermally activated phenomena," *Physical Review Special Topics-Physics Education Research,* vol. 9, p. 020101, 2013 2013.

[43] C. Fazio, O. R. Battaglia, and R. M. Sperandeo-Mineo, "Quantitative and qualitative analysis of the mental models deployed by undergraduate students in explaining thermally activated phenomena," *2013 ICPE-EPEC,* pp. 354-364, 2014 2014.

[44] C. Fazio, B. Di Paola, and I. Guastella, "Prospective elementary teachers' perceptions of the processes of modeling: A case study," *Physical review special topics-physics education research,* vol. 8, p. 010110, 2012 2012.

[45] C. Fernández and S. Llinares, "Implicative relations between strategies used in solving proportional and non-proportional problems," *Revista Latinoamericana de Investigacion en Matematica Educativa,* vol. 15, pp. 9-33, 2012.

[46] I. M. Gómez-Chacón, "Meta-emotion and mathematical modeling processes in computerized environments," in *From beliefs to dynamic affect systems in mathematics education*, ed: Springer, 2015, pp. 201-226.

[47] S. Guillaume, D. Grissa, and E. M. Nguifo, "Categorization of interestingness measures for knowledge extraction," *arXiv preprint arXiv:1206.6741,* 2012 2012.

[48] H. Khaled, S. Ghanem, and R. Couturier, "Analysis of Bejaia University Computer Science students' marks through the CHIC software and Statistical Implicative Analysis," in *2014 4th International Symposium ISKO-Maghreb: Concepts and Tools for knowledge Management (ISKO-Maghreb)*, 2014, pp. 1-8, 10.1109/ISKO-Maghreb.2014.7033473.

[49] I. Kohanova, "Analysis of University Entrance Test from mathematics," *Acta Didactica Universitatis Comenianae Mathematics,* vol. 12, pp. 31-46, 2012 2012.

[50] U. Kortenkamp and S. Ladel, "Flexible use and understanding of place value via traditional and digital tools," *RESEARCH REPORTS KNO-PI,* vol. 33, 2014 2014.

[51] I. C. Lerman and S. Guillaume, "Comparing two discriminant probabilistic interestingness measures for association rules," in *Studies in Computational Intelligence* vol. 471, F. Guillet, B. Pinaud, G. Venturini, and D. A. Zighed, Eds., ed, 2013, pp. 59-83.

[52] J. Melusova and K. Vidermanova, "Upper-secondary students' strategies for solving combinatorial problems," *Procedia-Social and Behavioral Sciences,* vol. 197, pp. 1703-1709, 2015 2015.

[53] K. Nikolantonakis and L. Vivier, "Positions numeration in any base for future elementary school teachers in France and Greece: one discussion via registers and praxis," *Menon, Florina,* vol. 2, pp. 99-114, 2013 2013.

[54] E. A. Panitsides and S. Anastasiadou, "Lifelong Learning Policy Agenda in the European Union: A bi-level analysis," *Open Review of Educational Research,* vol. 2, pp. 128-142, 2015 2015.

[55] D. Pasquier and R. Gras, "In the interest of the statistical analysis implicative (ASI) for exploratory research in psychology," *Psychologie Francaise,* vol. 57, pp. 161-173, 2012.

[56] D. Pasquier and L. Rioux, "Satisfaction et confort au travail. L'apport de la démarche implicative," *Psychologie du Travail et des Organisations,* vol. 20, pp. 275-293, 2014 2014.

[57] G. Pavlovicova and J. Zahorska, "The Attitudes of Students to the Geometry and Their Concepts about Square," *Procedia-Social and Behavioral Sciences,* vol. 197, pp. 1907-1912, 2015 2015.

[58] N. Q. Phan, H. X. Huynh, F. Guillet, and R. Gras, "Classifying objective interestingness measures based on the tendency of value variation," 2015, pp. 143-172.

[59] N. Pizzolato, C. Fazio, R. M. S. Mineo, and D. P. Adorno, "Open-inquiry driven overcoming of epistemological difficulties in engineering undergraduates: A case study in the context of thermal science," *Physical Review Special Topics-Physics Education Research,* vol. 10, p. 010107, 2014 2014.

[60] L. Rioux and D. Pasquier, "A longitudinal study of the impact of an environmental action," *Environmental Education Research,* vol. 19, pp. 694-707, 2013 2013.

[61] G. G. Stella and A. D. Sofia, "HUMAN RESOURCES DIMENSIONS: AN APPROACH OF GREEK MANAGERS," *The Economies of Balkan and Eastern Europe Countries in the changed World,* p. 59, 2014 2014.

[62] K. Ţilková, "Testing Pre-service Primary Education Teachers in Quadrilaterals," 2014.

[63] M. van den Heuvel-Panhuizen, I. Elia, and A. Robitzsch, "Kindergartners' performance in two types of imaginary perspective-taking," *ZDM Mathematics Education,* vol. 47, pp. 345-362, 2015.

[64] L. Zamora-Matamoros, D.-S. Jorge Rey, and L. Portuondo-Mallet, "Fundamental Concepts on Classification and Statistical Implicative Analysis for Modal Variables," *Revista Colombiana de Estadistica,* vol. 38, pp. 335-n/a, 2015 2015.

[65] K. Žilková, "Misconceptions in Pre-service Primary Education Teachers about Quadrilaterals," *Journal of Education, Psychology and Social Sciences,* vol. 1, 2015 2015.

[66] K. Zilková, J. Guncaga, and J. Kopácová, "(MIS) CONCEPTIONS ABOUT GEOMETRIC SHAPES IN PRE-SERVICE PRIMARY TEACHERS," *Acta Didactica Napocensia,* vol. 8, p. 27, 2015 2015.