Clearing up uncertainties in graduate programs candidate selection using a data science approach

Evaldo B. Costa[1], Giselle Raposo[1], José B. da S. Filho[1;2],
Luiz Paulo Carvalho[1], Wander dos S. Vasconcellos[1], Claudio Miceli de Farias[1]

[1]Programa de Pós-Graduação em Informática (PPGI) – UFRJ
Av. Athos da Silveira Ramos, 149, Rio de Janeiro, RJ – Brazil

[2]Centro de Instrução Almirante Graça Aranha (CIAGA) – Marinha do Brasil
Av. Brasil, 9020, Olaria, Rio de Janeiro, RJ – Brazil

`giselle.s.raposo@gmail.com, luiz.paulo.carvalho@ppgi.ufrj.br,`
`{evaldodacosta,josebarbosa,wander.vasconcellos,claudiomiceli}@ufrj.br`

Abstract. *Brazilian Graduate Programs are evaluated by a specific foundation, CAPES. This evaluation qualifies the respective program with a grade, and this directly influences the permissions and funds assigned to this program, such as having a doctorate level, grading 4 or higher. One of the indicators is the student body, as an external variable outside the control of the program. Based on the Design Science Research methodology, we present a research on the construction of an artifact that clears up the selection of candidates for the programs, categorizing them according to the profiles of previous students, based on descriptive statistics and data analytics.*

1. Introduction

The analysis of graduate program student profiles in informatics contextualizes this work, aiming to predict the profile of the student in a certain number of future years, for example, which is a possible approximation of the student's profile or identity entering the graduate program (GP) in the next year? Based on Data Science and Machine Learning [GRUS, 2018] [PROVOST AND FAWCETT 2013], we use the dataset with information already available from GPs as input, trying to minimize a cultural problem rooted in the mores, the subjectivity in the final decision of these candidate selections.

Part of the graduate experience is the application in selection processes in the academic trajectory, as we passed. In the undergraduate courses the National High School Examination (*Exame Nacional do Ensino Médio*-ENEM) result is enough to apply for access to the desired course. The Brazilian procedure to examine a graduate program requires an interview, regulated by a specific committee to examine the program, operated by one professor or more; curriculum analysis; foreign language test; and, in certain programs, proof of specific knowledge, traditional practical or written test.

The interview for the vacancy held by the candidate is scored based on subjective criteria inherent to each evaluator, whether justified or not. The evaluation is, as epistemologically expected by the human factor, subjective and skewed from the evaluator's indicators, as it will select its individuals' advisors or students who will compose its research group. Then the question arises, what is the weight of each

criterion or indicator in this note? Distance between where the candidate lives and the university? How old is he? Undergraduate training area, primary? Genre? Intent to get a research grant? Well-employed? Work with the evaluator's same research topic? How much do these responses impact the candidate's evaluation, if they do?

Decision making will not always be aligned with the GP strategies and respective goals so that it either remains ranked in the identical position or ascend, towards position 7 [CAPES 2017]. Decision Support Systems can assist, processing series or historical data, standardizing suggestions guiding behavioral decisions [ELAM et al. 1992].

From the Design Science Research (DSR) methodology [PEFFERS et al., 2007] [HEVNER, 2007] [WIERINGA, 2014], the proposal consists of a system to enrich the evaluation of candidates for graduate programs, complementing the interview or separately as its own stage, receiving data from graduating students and quality criteria by the regulatory parties and outputting a value, representative of the candidate. The dialogical ideation of the proposal involved deep debates in its process. It is untrivial to represent people as objective values through quantitative approaches, a discussion also included in the literature on Ethics and Data Science [SUMPTER, 2018].

The referral can be endorsed by the premise that Brazilian graduate programs are qualified through specific criteria, essentially quantitative, by the Coordination for the Improvement of Higher Education Personnel (*Coordenação de Aprimoramento de Pessoal do Ensino Superior-* CAPES)[1]. The quantitative approach follows the objective perception of reality [RECKER, 2013], allowing the construction of a model that represents it. The student is one of the variables composing this model, his productivity and performance influence the evaluation and consequently in the graduate programs perceived quality, according to CAPES criteria and indicators [CAPES, 2017]. In this way, the programs have power and influence over the admission of their candidates, not over the regulatory agency, CAPES, hierarchically superior. As a regulatory entity, it is CAPES 'responsibility to guide GPs by specific metrics and parameters that lead to better quality. It is necessary to reduce the uncertainty phenomena from structural mechanisms [PREMKUMAR et al. 2005], as norms, that direct the organizational behavior in a specific direction, by the dynamic complexity required by the Brazilian realities' plurality.

CAPES determine quantitative criteria that allow the representation of the Student Body and its components, as part of a singular quality indicator. There is no pragmatic approach to the debate on the student dehumanization as an objective value by the established indicators. This objectification occurs at a higher level, outside the influence of the GPs. In another light, a recommendation system can mature and assist the selection of candidates for graduate programs, with a non-exclusive approach. The act of a GP opposing or ignoring the criteria does not guarantee that they will be modified or neglected by the regulatory agency, which will run over them like a steamroller and officially discredit them for not meeting the criteria or not "presenting expected quality".

The proposal left the idea level for practice level when stakeholders from the social context. They found it relevant and useful for the presented problem, in the context of the graduate program, with more detail in Section 3. This paper aims to present a

1 Foundation associated to the Brazilian Ministry of Education (MEC), aims the expansion and consolidation of the stricto sensu graduate programs (master's and doctorate). Source: http://www.capes.gov.br/en/

research framework, present results of the proposal's effectiveness and promote debate about the matter, including the CAPES agenda impact.

The present work is structured as follows, Section 2 presents the theoretical foundation; Section 3, the methodology, Design Science Research; Section 4, the artifact fundamentals; Section 5, related works; Section 6 concludes this paper with discussion and conclusion.

2. Theoretical Foundation

In this Section the fundamental concepts used in the work are presented, the evaluation criteria of CAPES and Decision Support Systems.

2.1 CAPES evaluation criteria

The fundamental basis of the proposal is the 2017 CAPES Quadrennial Evaluation Regulation (2013-2016) [CAPES, 2017]. In this document, CAPES, the regulator of graduate and undergraduate courses, determines criteria scoring from 1 to 7 to Brazilian graduate programs. In all categories there is the criterion "Student Body", in a synthesized way it is evaluated if the graduates complete the courses, if they finish in the regular term and their intellectual production, except the respective thesis.

Of the CAPES evaluation criteria, only the Student Body represent an exogenous variable and the others are, directly or indirectly, the responsibility of the Faculty and Collegiate Body. In that manner, in the GP, the only completely independent aspect beyond the control of the GP Management, the Collegiate Body, remain the student.

The student represents a variable outside the direct control of the GP, despite this the criterion associated with the Student Body is decisive for the result of the evaluation, which reads: "ii. The lowest value among the grades obtained by the program in Questions 3 and 4 (student body and intellectual production) will define the limits of the final grade to be awarded." [CAPES, 2017]. The performance and performance of students influence the result of the GP assessment. For example, a GP may lose their doctorate program if drop from score 4 to 3.

When CAPES determine a quantitative assessment and the activities of a student are represented by a value, this comparative parameterization establishes an objective and rational reality, framed in the positivist paradigm [RECKER, 2013]. The evaluation does not score a student, but the collective formed by several. The student stops comprise a subjective individual to be an objective value, with a comparative purpose between constructs of reality. The assessment does not qualitatively check the student's well-being, his mental health, his satisfaction with the environment he attends, or what he believes or endures. The criteria are numerical and normative.

In proportion to the objective quantitative rigidity required by the regulator, the feasibility of designing systems that automate and support decision making is growing. By the Behavioral Decision Theory [ELAM et al., 1992], several factors intrinsic to individual subjectivity can influence the rational decision-making of the whole. The candidates' evaluation interviews conducted by teachers without the support of objective criteria can compromise the performance of the GP.

2.1. Decision Support Systems

Decision Support Systems (DSS) involves technological factors, Information and Communication Technology (ICT), human and procedural factors, categorized as an Information System (IS) [STAIR AND REYNOLDS, 2018]. In this work, we propose the behavior of a semi-structured DSS [TURBAN et al., 2011], that is, it combines standardized operation with personal evaluation; in this specific case, acquisition of human resources to compose the GP Student Body, among those who request access.

The ITC factor is based on automated systems, involving a data-based model, as a support for decision, stored and structured data imported as an input for solution, with the possibility of computerization. Human factor as candidates accessing a specific GP and the parties responsible for the selection, the ethical character of the solution and analysis by the humanistic bias of the CAPES evaluation method. Procedural factors, CAPES 'quality indicators and their details, the GP rules and procedures for selecting their candidates.

Students performance, collectively, influence the result of the respective PPG. At loftier levels the decision of which candidates to select becomes tactical and impacts the entire GP. It is not just a teacher selecting an oriented or member for their research group, this decision cannot be immature. When admitted students are not under the GP control, exercising their own freedom, adding risk or uncertainty to the problem in a dynamic and complicated scenario, complex to explain a priori [TURBAN et al. 2011].

Decision making should not be tacitly "the best", but structured reasonably and rationally [MARCH AND SIMON, 1993], based on raw data or feedback information [STAIR AND REYNOLDS, 2018], as shown in Figure 1. Mathematical models can help in this task, where complex relationships are built with inputs [JUNIOR AND CEGIELSKI, 2015]. For example, Clemen and Reilly (2014) use a decision tree in a similar scenario, exposing the cumulative risk in question.
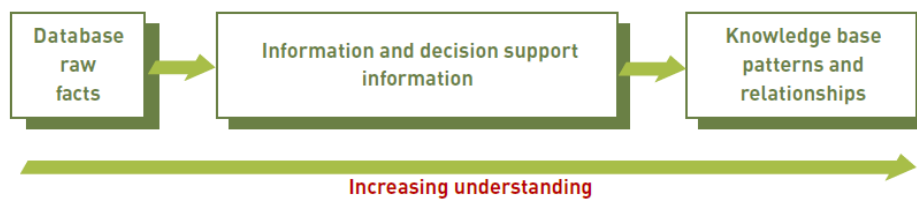
Figure 1: Relationships between data, information, and knowledge [STAIR and REYNOLDS, 2018]

By Organizational Information Processing Theory [PREMKUMAR et al. 2005], DSS built with relevant data and supported by an objective framework of reality, like the quantitative criteria of CAPES, improve organizations to deal with uncertainties and chaotic environments. In this context, in the conscious and justified selection of candidates to be approved in the selection process, based on data and to avoid persistence of errors with history recorded by the data. For example, if in the historical series all students who live far from the university performed very poorly, how to deal with the following candidates in this similar situation? Even if this specific issue is not crucial for final approval decision-making, other dimensions may be excellent. The initiative demonstrates awareness in producing the ultimate decision. After all, this candidate, if approved, will influence the GP collectivity and quantitative performance.

3. Design Science Research

DSR [WIERINGA, 2014] is the appropriate epistemology to conduct the research due to its nature, conferring epistemological rigor required by the scientific method [PIMENTEL et al., 2019]. Design Science is hitherto used in the Education area [PIMENTEL 2017] [DE VILLIERS AND HARPUR, 2013] with satisfactory scientific results. Artifact engineering aims to modify reality by making the usefulness of solving the problem relevant to the social context as a primary focus, followed necessarily by the development of scientific research around the artifact [WIERINGA, 2014].

Among present methods that guide research on DSR, we selected Peffers et al. (2007), well established in this scope [PIMENTEL et al., 2019]. Following the recommended process, we observe the interest in the social context. Then define the objective to solve the problem from the artifact, design and develop the logical-technological level and verify its applicability in a real case.

Figure 2 shows the partial methodological framework of this research, under development. Section 4 exposes the first evaluation already carried out, pointing the effectiveness of the artifact. At the beginning the focus is on the technical requirements of the computational mathematical artifact. Pragmatically, it is unfeasible to analyze the influence of the artifact without it having enabled and verified its functional requirements.
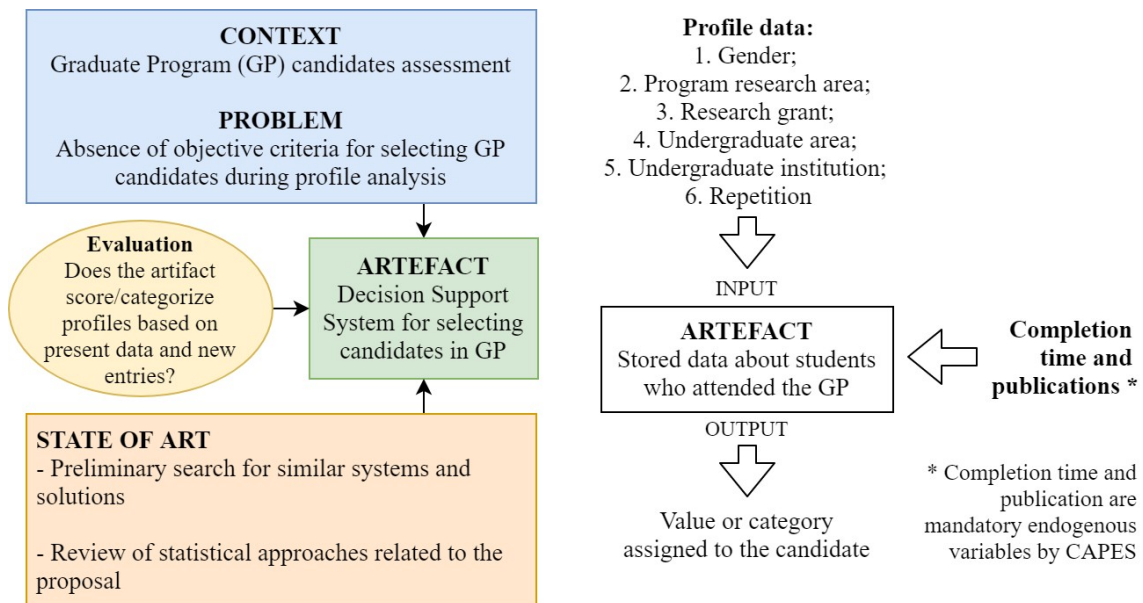

Figure 2: DSR methodological frame and artefact instance

4. A Decision Support System to assist GP candidates' assessments

Figure 2 illustrates the system utilizing the mathematical proposed model, instantiating its use for demonstration of effectiveness. The entry is the profile data of the candidate for the vacancy, which will be processed and compared with the profile data of former students in the GP, those who completed or evaded the program. Completion time and publication numbers are mandatory and considered key dimension

by CAPES [CAPES, 2017] regarding the GP evaluation criteria. These two dimensions must be present in the database for the artifact to be effective, additional data enrich the analysis [GRUS, 2018], the more relevant data and the longer the collection time, the better. The output is a numeric value or a category metaphorically representing ranges of values, simplified.

Other scenarios may have datasets with other dimensions, in greater or lowest relevant quantity selected for analysis. In the use case instance data were utilized between the years 2013 and 2018. Publications data were collected from the Sucupira platform, which has maintained them since 2013. If it were not for the publications data, we have had inputs since 2004. Personal data inputted: (i) **gender**, male or female, the program does not collect non-binary gender data; (ii) **research area**, representing the research area or research group; (iii) **research grant**, if the candidate indicates dependence on this item; (iii) **undergraduate area**, separated into considerable areas according to the candidate's completed degree course: or Informatics, or Science, Technology, Engineering and Mathematics (STEM) Non-Informatics (e.g., mathematics), or Not-STEM; (iv) **undergraduate institution**, candidate alma mater, "same" to indicate that he progressed in academic level in the same university; (v) **repetition**, analyzing whether the candidate represents a GP recidivist or is progressing at the education level.

The six dimensions are inputted for calculation, in other scenarios it is unlikely to be the same dimensions. These dimensions were selected because they capture economic and social factors that can impact the student's academic performance, current conjuncture of the society in which we live. For example, gender discrimination, academic capacity (graduate and research area), the need for financial research support (research grant), and their academic performance. Therefore, these dimensions represent diverse societal aspects, which may impact PPG performance.

We seek to reproduce the social context in which it is inserted, to try to recommend the individuals with more possibilities of success. It seeks to enable reproducibility and adaptability in this area. The determined CAPES success parameters were the ideal GP completion time and quality of student publications at renowned conferences and journals for computer science.

Using the parameters of completion time and publications quality, we divided the transition zones that all GP students go through or they can pass along the course in four student classification colors: green, yellow, red and black. The green area represents students who meet CAPES parameters; yellow, those who partially meet, did not produce the most influential publications or a little overdue; red, those who cannot handle the parameters, because they have far exceeded completion time; and black, those who do not meet any of the minimal parameters and have even been unable to complete their graduate studies.

The mathematical model developed during the research, based on descriptive statistics, was built in such a way to be evaluated as a managerial tool, K-Nearest-Neighbors (KNN) algorithm [Grus, 2018] has been used to graphically consolidate the database. with the function of identifying among graduate students, those whose profiles are close to the students classified in the green, yellow and red, looking for program excellence. Figure 3 illustrate the proposed model which can be written mathematically as a cross product with the following equation:

$$Grade = \vec{C} \cdot \vec{X} \Rightarrow Grade = \sum_{i=1}^{n} C_i X_i$$

Following the logical scenario, a student who has completed the GP with a high-quality publication, restricted quality (defined by CAPES as B1, A2, A1 in her scores) and within the regular completion time, without overdue, will be evaluated with the maximum value and his profile will be considered of the highest possible value. Students who do not have publications or publications without qualification in his area and overdue the completion time will be considered of the lowest possible value.

4.1. Coefficients definition

To define the C_i coefficients, several analyses were performed using descriptive statistics tools on the GP student's dataset. Table 1 shows the proportion between masters and doctors related to the duration of master's and doctorate courses. In both cases, the number of those who complete in the expected time is greater than the overdue number. With these data, the mean, standard deviation variance was calculated. Additionally, show the proportion of candidates classified by gender. The proportion of female candidates who complete the master's course in regular time is higher than male candidates. The proportion is inverse when we speak of the doctoral course, with more women than men overdue, which is a key indicator for CAPES [CAPES, 2017].

Table 1: Overdue and gender quantitative analysis

Overdue analysis			Gender analysis				
	Master's	**Doc.**		**Master's**		**Doc.**	
				Female	Male	Female	Male
Regular	173	20	Regular	48	125	6	14
Overdue	138	13	Overdue (number)	42	96	3	10
Total	311	33	Overdue (%)	47	43	33	42
Mean	1.90	3.75	Total	90	221	9	24
Variance	0.05	0.18					
Std. dev. *	0.22	0.42	*Standard deviation				

Table 2: Model dimensions and coefficients – Doctorate and master's

Student Dimension	Values	Doctorate			Master's		
		Qtt	(%)	Coef.	Qtt	(%)	Coef.
Gender	Male	24	72,73	12,12	221	71,06	11,84
	Female	9	27,27	4,55	90	28,94	4,82
Research area	SCA	6	18,18	3,03			
	ASC	13	39,39	6,57			
	GSC	14	42,42	7,07			
	IES				40	12,86	2,14
	MASI				35	11,25	1,88
	RCSD				29	9,32	1,55
	No info				68	21,86	3,64
	AMN				45	14,47	2,41
	SI				93	29,90	4,98
Recidivism	Recidivism	1	3,03	0,51	6	1,93	0,32
	Progression	16	48,48	8,08	305	98,07	16,35
	No repetition	16	48,48	8,08	93	29,90	4,98
Research Grant	Grant	21	63,64	10,61	47	15,11	2,52
	No grant	12	36,36	6,06	264	84,89	14,15
	Informatics	18	54,55	9,09	195	62,70	10,45
Undergraduate area	STEM (not-informatics)	14	42,42	7,07	57	18,33	3,05

	Other areas				25	8,04	1,34
	No info	1	3,03	0,51	34	10,93	1,82
Undergraduate Inst.	Same	14	42,42	7,07	121	38,91	6,48
	Public	8	24,24	4,04	57	18,33	3,05
	Private	10	30,30	5,05	98	31,51	5,25
	No info	1	0,32	0,51	35	11,25	1,88

Table 2 show the coefficients related to the dimensions evaluated, assigned a coefficient. In the present study, considering each of these six dimensions (n = 6) with the same relevance, we make the weights assume identical and unit values ($P_i = 1$) and using their probabilities of occurrence (Pr_i), the coefficient will be defined by:

$$C_i = \frac{Pr_i \times P_i}{\sum_{i=1}^{n} P_i}$$

Each coefficient represents a portion of a sum whose maximum value is 100%. In assessing the adherence indicator, we establish the relevant values for each dimension, thus simplifying by adding the coefficients that are defined in the individual under evaluation ($x_i = 1$):

$$Adherence = \sum_{i=1}^{n} C_i$$

This indicator shows the adherence degree to the current GP student profile, in a program with diversity it will never assume extreme values, 100% or 0% (100 or 0).

Figure 3 presents the graphs showing the students classification. According to CAPES criteria, in the master's degree, 9.4% of students perfectly fulfill the objective and deterministic criteria; in the doctorate 46.9% published in communications considered to be of higher quality and concluded the course in regular time.
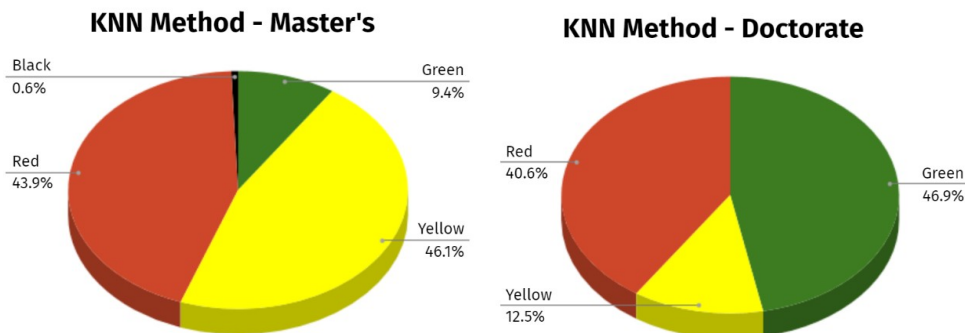

Figure 3: Dispersion of CAPES metrics for master's and doctoral degrees

In these experimental results it was not considered "counterexample" validation, i.e., an "unwanted" candidate, for ethical reasons and for limiting the available data set, based only on candidates accepted in the program. In this way, notes smaller than one is unseen. Due to the way we model the problem, there is no total adherence, i.e., scores higher than nine. Extreme scores would merely happen if the GP were homogeneous and exclusive, which in practice would be a program with little diversity, that is, qualitatively terrible and restricted.

5. Related Works

We performed a semi-structured bibliographic review in order to find out if the same artifact or solution exists or similar related works address the problem, even in different scenarios, allowing for transferability by similarity.

After a search in CAPES Journals Portal (Portal de Periódicos CAPES) and Google Scholar, using English and Portuguese terms associated with the proposal, we did not discover an equal solution. We believe that this phenomenon occurred predominantly because GP and Student Body assessment systems vary between countries, due to diverse and specific procedural and socio-technical aspects [STAIR AND REYNOLDS, 2018].

Ahmed et al. (2016) built a system to aid students select their universities around the world to study, based on their personal profiles and using Data Mining. In this work they select specific dimensions of the profile and identify, from the data, which are the best universities for the eager person to apply for, correlating information and analyzing their influence. They use previous profiles of successful students in the selections as input.

On the other hand, Benbassat and Baumal (2007) analyze, in a context different from that of Brazil, which criteria should be prioritized to approve candidates in undergraduate medical courses. The authors encourage the use of personal or sensitive data to assist or complement the selection decision.

In the work of Fong and Biuk-Aghai (2009) a Case Study is presented in Macau, China, using Data Mining for a DSS. They build a recommender that analyzes various data sources of high school students, to predict their chances of admission to specific universities. The authors also criticize the manual processes based on the intuition of the parts that select students, impairing accuracy.

In contrast, Rabelo et al. (2017) use Educational Data Mining applied to predict the performance of students in distance learning courses, starting from, also instrumentalizing Decision Tree. The system predicts whether a student will succeed or not during the subject course. This work is aimed at monitoring and follow-up, using personal actions of each one and not their data.

Finally, Pimentel (2017) exposes the way to develop scientific research in Informatics in Education using DSR for engineering of computational artifacts, observing the daily life of the GP, delving into the tensions related to the artifact and its conception and thinking-making conjectures.

6. Discussion and conclusion

This work presents the preliminary results in the construction of an artifact that objectively helps the selection of candidates for GPs. Made possible by the deterministic requirements stipulated by CAPES [CAPES, 2017].

Table 3 shows the simulation test performed on the artifact. Three doctoral candidates and three master's candidates were considered, with their respective attributes generated randomly, demonstrating the effectiveness of the proposal. The grade assigned by the artifact should be an indicator of similarity or synergy with the profiles of successful graduates according to CAPES metrics, making up a student

assessment.

Table 3: Artefact evaluation, candidates' assessment

Dimension	Doc. 1	Valor	Doc. 2	Valor	Doc. 3	Valor	Mast. 1	Valor	Mast. 2	Valor	Mast. 3	Valor
Gender	Male	12,12	Male	12,12	Male	4,55	Male	11,84	Male	11,84	Female	4,82
R. Area	SCA	3,03	ASC	6,57	GSC	7,07	IES	2,14	MASI	1,88	SI	4,98
Repetition	Recid.	0,51	Progr.	8,08	No Rep.	8,08	Recid.	0,32	Progr.	16,35	No Rep.	4,98
R. Grant	Grant	10,61	No grant	6,06	No grant	6,06	Grant	2,52	No grant	14,15	No grant	14,15
Und. Area	STEM	9,09	STEM (NI)	7,07	No info	0,51	STEM	10,45	STEM (NI)	3,05	Others	1,34
Und. Ins.	Same	7,07	Public	4,04	Private	5,05	Same	6,48	Public	3,05	Private	5,25
Score		42,43		43,94		31,32		33,75		50,32		35,52

Each candidate assumes their dimensions associated with their respective coefficients, shown in Table 2, scoring their own grade. Like so, the artifact would indicate candidates according to the decreasing order of grade, that is, candidates with the highest grades would have priority according to their research areas. Therefore, our proposal would display an expected tendency to indicate the best possible candidates.

Analyzing deeply the attributes of the candidates, we find the values of a dimension can vary substantially. However, we identified some similarities between the first placed in the overall classification. Both are men, are continuing their studies (same institution, from same undergraduate to same graduate area, i.e. progression), do not need a scholarship, are graduated in the STEM (Not Informatics, i.e., excluding Informatics) area and studied at public universities. Among the others, the attributes are more heterogeneous, and this occasioned them to lose some points. Therefore, it seems safe to say the profile of the first places would be predominant in the indications of our proposal. This tendency, according to Table 3, is more consolidated in the master's degree, where the second place's score is 32.93% lower than the first place's score, while in the Doctorate this difference falls to 3.44%, showing the trend can however vary. This is related to the number of examples used to describe the domains used. The Master's course already has a considerably larger student information than the doctorate, created years after.

The ethical and moral influences of data manipulation and analytics is a trending research topic, as exemplified by O'Neil (2016). From this scenario two requirements decisions were made: (i) issue exit notes for the possibility of "success" and not "failure", so that the system is unintended to lead candidates for elimination, the interest is sunny days and the potential of candidates to succeed; (ii) the output does not indicate an absolute binary "accept" or "reject" decision, but a value to assist those responsible for the selection in making a mature decision in accordance with their specific strategies.

This proposal is primarily aligned with the basic operationalizations of Data Science, using varied and undefined data in the construction of a future categorization [GRUS, 2018], additionally it can be used as a Business Intelligence approach, allowing a perception of how the student body is composed [LUHN, 1958]. Derived from this present work, Carvalho et al. (2019) is an example analyzing that in the studied PG there is a large numerical discrepancy between men and women approved in the assessment process, three men for one woman. Studying this reality, the PG can make decisions accordingly, in order to reduce the gender discrepancy in its numbers and committing to social responsibility initiatives.

Indirectly, the quantitative output of the model can expose specific phenomena from the data, through history, allowing for a theorization of reality based on that

[RECKER, 2013]. For example, if "gender" represents a dimension with a high impact on the grade, then there is an imbalance in the dimension, the GP may or may not strategize accordingly. In this way, a GP can adjust its selection to balance the dimensions. Unscrupulous use can lead to selection for a radical partial exclusionary behavior, which is already occurring today, however in a not-explicit way.

As future work, and the continuation of this research, we envision the accomplishment of the complementary study, to follow the candidates and graduates along the next cycles to refine the defined model; construction of a study with machine learning algorithms, implementing the proposed mathematical concept; application of the approach to the profile of other University and external programs, with other available dimensions, assessing their degree of applicability; critical evaluation of the GP objectification by CAPES evaluation criteria; deepening the ethical and human influences of this approach to numerical valuation of candidates.

7. Acknowledgments

References

Ahmed, S., Latiful Haque, A., Hasan, M., Tasmin, R., Abdullah, D., and Tabassum, A. (2016). Discovering knowledge regarding academic profile of students pursuing graduate studies in world's top universities. In 2016 International Workshop on Computational Intelligence (IWCI), pages 120–125. IWCI.

Benbassat, J. and Baumal, R. (2007). Uncertainties in the selection of applicants for medical school. Advances in Health Sciences Education, 12(4):509–521.

Carvalho, L. P., Raposo, G., Cappelli, C., Miceli, C. (2019) An analysis of female participation in informatics research at UFRJ's PPGI. In 2019 Escola Regional de Sistemas de Informação (ERSI), Duque de Caxias. SBC, Porto Alegre.

CAPES (2017). Ordinance nº 59, 21/03/2017, https://capes.gov.br/images/stories/download/avaliacao/27032017-Portaria-59-21-03-2017-Regulamento-da-Avaliacao-Quadrienal.pdf . BRAZIL, Education Ministry. Available in 17/03/2020.

Clemen, R. and Reilly, T. (2014). Making Hard Decisions with Decision Tools. Cengage Learning, 3rd edition.

De Villiers, M. R. R. and Harpur, P. A. (2013). Design-based research-the educational technology variant of design research: Illustrated by the design of an m-learning environment. In Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference, SAICSIT '13, pages 252–261, New York, NY, USA. ACM.

Elam, J. J., Jarvenpaa, S. L., and Schkad, D. A. (1992). Behavioral decision theory and DSS: New opportunities for collaborative research. In Information Systems and Decision Processes, pages 51–74, Los Alimotos, CA.

Fong, S. and Biuk-Aghai, R. (2009). An automated university admission recommender system for secondary school students. In The 6th International Conference on Information Technology and Applications (ICITA 2009). ICITA.

Grus, J. (2018). Data Science from Scratch: First Principles with Python. O'Reilly, USA.

Hevner, A. R. (2007). A three-cycle view of design science research. Scandinavian Journal of Information Systems, 19(4):87–92.

Junior, R. K. R. and Cegielski, C. (2015). Introdução a Sistemas de Informação. Apoiando e transformando negócios na era da mobilidade. CAMPUS, 5th edition.

Luhn, H. P. (1958). A Business Intelligence System. IBM Journal of Research and Development. 2(4):314–319. doi:10.1147/rd.24.0314.

March, J. G. and Simon, H. A. (1993). Organizations. Wiley-Blackwell, 2nd edition.

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group, New York, NY, USA.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3):45–77.

Pimentel, M. (2017). Design science research e pesquisas com os cotidianos escolares para fazer pensar as pesquisas em informática na educação. In Anais do SBIE 2017, pages 414–424. SBC.

Pimentel, M., Filippo, D., and Santoro, F. M. (2019). Design science research: fazendo pesquisas científicas rigorosas atreladas ao desenvolvimento de artefatos computacionais projetados para a educação. Metodologia de Pesquisa em Informática na Educação: Concepção da Pesquisa., 1(5):414–424.

Premkumar, G., Ramamurthy, K., and Saunders, C. S. (2005). Information processing view of organizations: An exploratory examination of fit in the context of interorganizational relationships. Journal of Management Information Systems, 22(1):257–294.

Provost, F. and Fawcett, T. (2013). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media.

Rabelo, H., Burlamaqui, A., Valentim, R., de Souza Rabelo, D. S., and Medeiros, S. (2017). Utilização de técnicas de mineração de dados educacionais para predição de desempenho de alunos de EAD em ambientes virtuais de aprendizagem. In Anais do SBIE 2017, pages 1527–1536. SBC.

Recker, J. (2013). Scientific Research in Information Systems A Beginner's Guide. Springer, Berlin.

Stair, R. and Reynolds, G. (2018). Principles of Information Systems. Cengage Learning, 13rd edition.

Sumpter, D. (2018). Outnumbered: From Facebook and Google to Fake News and Filter-bubbles – The Algorithms That Control Our Lives. Bloomsbury Sigma.

Turban, E., Delen, D., Sharda, R., and King, D. (2011). Business Intelligence: A Manage-rial Approach. Person Higher Ed, 4th edition.

Wieringa, R. (2014). Design Science Methodology for Information Systems and Software Engineering. Springer-Verlag Berlin Heidelberg