

Data Mining Solution for Assessing Brazilian Secondary School Quality Based on ENEM and Census Data

Paulo J. L. Adeodato

Centro de Informática, Universidade Federal de Pernambuco (UFPE),

Recife, Pernambuco, Brasil

`pjla@cin.ufpe.br`

Abstract: *This paper presents a data mining solution for assessing the quality of Brazilian private secondary schools based on the official school survey and students tests. Following the CRISP-DM method, after the problem interpretation and modeling, these two data sources yearly collected have been transformed to the school granularity level embedding data and expert's knowledge and have been integrated in a single data set with the national school code as primary key. Further transformations on the joint data set embedded additional knowledge and made the format compatible with the artificial intelligence techniques applied for knowledge extraction. Logistic regression was applied for producing a propensity score for good schools, decision tree applied for extracting the sequential decision making a human would follow and rules were induced for supporting the explanation of a decision based on the score. The AUC_ROC and Max_KS were used for assessing the propensity score performance and, coverage, confidence and lift were used for assessing the quality of the rules induced by the A Priori algorithm, together with the human knowledge available on the literature. The results showed that this domain-driven data mining approach was successful in modeling the problem and validating educational public policies.*

Keywords: *School quality assessment, Educational decision support system, CRISP-DM, Domain-Driven Data Mining, Educational Data Mining.*

Resumo: *Este artigo apresenta uma solução de mineração de dados para avaliação da qualidade das escolas secundárias privadas brasileiras com base nos dados do ENEM e do Censo Escolar. Seguindo o método CRISP-DM, após a interpretação problema e modelagem, essas duas fontes de dados anualmente levantados tiveram suas variáveis transformadas para a granularidade escola, incorporando conhecimento do especialista humano. Elas foram integradas em um único conjunto de dados tendo o código da escola como chave primária. Outras transformações no conjunto de dados integrado incorporou conhecimentos complementares e tornou o formato compatível com as técnicas de inteligência artificial aplicadas para extração de conhecimento. A regressão logística foi aplicada para produzir um escore de propensão para "boas" escolas, árvore de decisão foi aplicada para extrair o conhecimento que um ser humano usaria para tomar decisões de forma seqüencial e as regras foram induzidas por apoiar a explicação de uma decisão com base no escore de propensão. A AUC_ROC e o Max_KS foram utilizados para avaliar o desempenho do escore de propensão e, cobertura, confiança e lift foram utilizados para avaliar a qualidade das regras induzidas pelo algoritmo A Priori, juntamente com o conhecimento humano disponível na literatura. Os resultados mostraram que esta abordagem de mineração de dados focada no domínio de educação foi bem-sucedida na modelagem do problema e na validação de políticas públicas educacionais.*

Palavras-chave: *Avaliação da Qualidade das Escolas, Sistema de Suporte à Decisão para a Educação, CRISP-DM, Domain-Driven Data Mining, Educational Data Mining.*

1. Introduction

The quality of and accessibility to education are key aspects for the development of a fair society. In the United Nations (UN) and in all countries there are high level offices responsible for educational policies. These offices define several metrics for monitoring the quality of education at each of its stages. Several metrics and methods have been developed for quality assessment of education and, in recent decades, the ease of data gathering and their sheer volume has allowed for the application of new methods, particularly, those based on data mining approaches.

Despite these advances, the monitoring and the metrics for assessing the quality of education are often controversial but the government and regulatory bodies play an important role in defining standards to verify the effectiveness of public policies. An important issue considered is that the students performance is at the core of the institutional evaluation of learning. Furthermore, international efforts have advanced towards standardization with the Program in International Student Assessment (PISA) tests focusing on assessing the capabilities that the students have acquired with what they have learned [Darling-Hammond and Wentworth, 2010].

In Brazil, the National Institute for Educational Studies (Instituto Nacional de Estudos e Pesquisas Educacionais – INEP) produces the annual School Census which is a survey of the schools for secondary education in the country and the National Secondary School Exam (Exame Nacional do Ensino Médio – ENEM) that evaluates student performance at end of secondary education. In 2009, the ENEM has gained great importance when the government made it a mechanism for students' admission to higher education in public (free) universities. The increase in students enrollment and in their commitment to the ENEM exam have turned it into an excellent source of information on secondary education in the country. In addition to the technical knowledge of each student, ENEM captures information on socio-economic-cultural questionnaire on their profile [INEP, 2015]. The integration of information gathered by the School Census with the performance and profile of their students in ENEM enables the Federal Government to define and validate the public policies for national education [Travitzki, 2013].

The ENEM exam is also used to generate an annual ranking of schools from the average grades of their students on tests of the various areas of knowledge, considering performance only. Despite several technical limitations, this ranking method is an indicator of quality whose release promotes public awareness and influences society in general.

Even with the wealth of information available, little systematic knowledge has been extracted from these bases, among other reasons, for their large amount of data per year and because they are in different grains (schools, and students and teachers of the schools). This paper applies the work method CRISP-DM (**C**Ross **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining) [Shearer, 2000] to develop a data mining solution for extracting and expressing relevant knowledge about the secondary school education of Brazil.

The paper is organized in four more sections describing steps 2-5 of the applied CRISP-DM process followed by the conclusions. Section 2 presents the analysis of data from the problem interpretation considered in the introduction and the formal definition of the goal. It also details the transformation of variables to change the grain from students and teachers to the school grain, embedding human knowledge about the domain and statistical knowledge from the data. Section 3 presents the extraction of knowledge through three artificial intelligence techniques; logistic regression to estimate the chance of a school to be "good", rule induction to explain why a school was estimated "good" or "weak" and the decision tree to show the best form of sequential human decision. Section 4 presents the results and evaluates the performance of the solution measured by the AUC_ROC and Max_KS2 metrics. Section 5 highlights the main contributions and limitations of the paper and suggests future developments of the work.

2. Data Understanding, Goal Setting and Processing

This research has used two official public databases: Microdata from the National Secondary School Exam 2011 [ENEM, 2011] containing the students' socio-economic- cultural information and their grades on the test at the end of secondary education, and the School Census 2011 [Censo Escolar 2011] detailing the conditions of the schools, from physical infrastructure to faculty information. These databases refer to over 30,000 secondary schools across the country, but this paper will focus only on private schools.

2.1. Selection of the universe of schools (Scope)

By selecting only the focused private schools, the ENEM and the School Census databases were largely reduced. By integrating these databases through their unique school identification code (primary key), the intersection of private schools on the two databases dropped to 8,373 schools. This reduction was due both to new private schools which did not have any students sitting the ENEM exam and to private schools not surveyed by the school census that year.

Additional filters were applied to eliminate anomalies that could either divert from the goal or deteriorate the quality of the work. So, students who did not perform all the tests and those who took the exam without graduating secondary school that year were out of the scope of the research.

Back to the school grain, for having critical mass, only schools with 15 or more students were selected. After this last filter, the total definitely used in this research dropped to 4,844 secondary schools which are the dataset in this paper.

Additionally, all attributes that are valid exclusively for public schools (e.g. information on Education for Young Adults, EJA in Portuguese) were discarded.

2.2. Binary goal definition ("good" or "bad" school)

To effectively systematize this study, its objective about the quality of the schools was defined as a binary decision problem. In this process there are two controversial points: 1) which metrics should be used to evaluate the quality of the school and 2) which threshold should be adopted as a criterion to define what would be a "good" school [Travitzki, 2013; Stecher, 2010]. Keeping pragmatism

and fairness of the approach, this paper considered as a measure of school quality, the arithmetic mean of student grades, as already done in previous studies [Adeodato, 2015; Travitzki, 2013]. The grade for each student is the arithmetic average of their grades in each test while the school grade is the arithmetic mean of their students' grades considered above.

Once having defined the quality metrics, the most controversial point is to set the threshold to characterize what would be a "good" or "bad" school in the dichotomic objective. To circumvent the controversy and bring a higher level of abstraction that enables future comparison across years, regardless of the degree of difficulty of the exams, this study used statistics concepts for setting the threshold. Separatrices (quartiles, deciles, *etc.*) away from the tails of the distributions not only are robust against extreme values (outliers) [Johnson and Wichern, 2012] but also can be a straightforward data dependent dichotomizing criterion of interest for the application domain. The upper quartile has already been successfully used as threshold [Adeodato, 2015; Sousa *et al.* 2008] on a continuous goal variable for creating a binary target-variable thus converting the problem into a binary classification. If the research objective were to estimate the propensity for excellent schools, the top decile would be a more appropriate grade threshold. However, as the research objective is to estimate the propensity for schools being good, good schools will be those whose grades are in the upper quartile of the grades distribution of the schools included in the scope above.

2.3. Transformation of granularity of the attributes

The granularity of the attributes is a fundamental concept and its diversity brings great complexity to this work. How can one associate to each school its family income attribute from the distribution of family income of their students? How can one associate to each school its faculty education attribute from the distribution of faculty education of their teachers? These transformations represent a difficulty for teams without professionals specialized in developing data mining projects. This difficulty is due to both the sheer volume of data to be handled and the need to use artificial intelligence to embed knowledge of experts in education in the transformation of the attributes for granularity change in a process coined Domain-Driven Data Mining (D³M) [Cao, 2008].

To avoid combinatorial explosion, this study adopted the mean and mode to combine numeric and categorical attributes, respectively, in the School grain, despite existing richer transformations to embed knowledge [Adeodato, 2015]. Thus, in this paper the answers to the two above questions are 1) the average family income of the students considered and 2) the mode of faculty education of teachers in each school.

For binary variables, the ratio (proportion) of individuals from the target class on each school was the grain transformation performed.

2.4. Preprocessing

The missing data were filled by the median value in numeric attributes. In categorical variables, a "missing" category was created. In some cases dependent on the interpretation of the domain, when there was an order relation, the missing

data were filled by an extreme category (e.g. missing education level of parents was coded as "illiterate"). For numeric attributes, outliers were identified by dispersion beyond 3 standard deviations and replaced with this extreme value.

3. Knowledge Extraction and Results

In this paper the extraction of knowledge consisted of three basic functionalities aimed at building a Decision Support System (DSS): 1) an estimator propensity for a school to be "good", 2) a decision tree to explain how the best sequential decision process based on data would be from a human expert's point of view and 3) a rule base to explain the decisions and identify niches of high relevance to the field, even those representing small subpopulations ("nuggets").

3.1. Logistic regression

The technique chosen to generate the propensity score for a school being good was the logistic regression which has been successfully applied to binary classification problems in various areas such as bankruptcy [Kumar and Ravi, 2007], for instance. The technique identifies and quantifies the key attributes that significantly influence the binary target variable as a set of independent or explanatory variables (attributes). It is also well suited for small samples because it does not require a validation set for over-fitting prevention during training.

Logistic regression is adequate to study the behavior of a binary dependent variable based on a set of p independent variables x_p expressed by the logit function indicated in Equation (1) below. It presents explicitly the knowledge extracted from data in terms of the coefficients (β) validated by their statistical significance (p -value).

$$\log \left\{ \frac{\pi(x)}{1-\pi(x)} \right\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p, \quad (1)$$

where $\pi(x)$ is defined as

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}, \quad (2)$$

The output is a continuous score that expresses the propensity for a school to be good. This study used the forwardstepwise technique with the maximum likelihood criterion (LikelihoodRatio - LR) [Hair *et al.*, 2005], to select the attributes by their discriminating power in the model. Table 1 shows the 12 most influential attributes ranked by the absolute value of their parameter β that measures their influence among the 23 statistically significant attributes at a p -value < 0.05 for the model.

Table 1. Twelve most influential attributes in the logistic regression model with the forwardstepwise method.

Attribute	Beta	p-value	Var. Type	Grain
School has printer=Yes	2.54	0.009	Binary	School
School Located in the Southeast	1.11	0.000	Dummy	School
Quantity of Bathrooms in the Household (question q75)	0.83	0.000	Mode	Student

School has at least 1 PhD teacher	0.62	0.000	Transf.	School
School Located in the South	0.48	0.033	Dummy	School
School has Science Lab=Yes	0.36	0.003	Binary	School
Teachers Gender Predominance= G_1	-0.42	0.000	Mode	Teacher
Teachers Ethnic Predominance= E_2	-0.47	0.000	Mode	Teacher
School has Computer Lab=Yes	-0.59	0.000	Binary	School
School has TV set=Yes	-0.83	0.026	Binary	School
School Located in the North	-0.96	0.013	Dummy	School
Relevance of the ENEM Exam for Pursuing a Graduate Degree (question q25)	-1.41	0.033	Mode	Student

Some interesting knowledge has been explicated by these attributes' coefficients. Most attributes embed economic and financial conditions. From the most influential attributes, only the amount of bathrooms in the household appears among the main rules extracted by the artificial intelligence techniques below. One rule has pointed out that schools with science lab more often tend to be good while those with a computer lab tend to be weak, in a way similar to those having TV sets. That raises a question that should be further investigated by the education experts: "Are computer labs and TV sets being used for educational purposes or just used for the student's entertainment and social networking means, distracting the students from learning?"

3.2. Decision tree (C4.5/J48)

Decision Tree is a hierarchical structure of sequential decision represented as an inverted tree from the root to the leaves in which they partition the input space so as to maximize the information gain in relation to the goal. An important aspect of the tree is the generation of explicit knowledge embedded in the data in a human readable format of rules *if-then* [Safavian and Landgrebe, 1991]. Figure 1 illustrates only part of a branch of the tree, indicating the fraction of examples in that rule condition (support) expressed in percentage, followed by the concentration of good schools (confidence) and the ratio between the confidences of the rule condition and the population (lift) [Han *et al.*, 2011].

On top of the tree (left) are the most important attributes / conditions for the school to be in the first quartile. In the rules of three clauses, the concentration of good schools (confidence) ranges from 91.3% in the best extreme to 2.2% in the worst (shown in boldface). Once again, the influence of the economic and financial characteristics is very strong, either directly in the family income or indirectly by the ProUni option or the quantity of bathrooms in the household. When one sees that the 25% of the schools in a niche has a chance of 2,2% to be among those above the top quartile (less than 10% chance than by random selection), it becomes clear that the educational system in Brazil severely suffers from inequities and that the niche is directly related to economic condition of the population.

3.3. Classification rules

The classification rules were induced by the a priori algorithm, one of the most used algorithms for associative rule induction [Agrawal and Srikant, 1994]. The

rules were then filtered for preserving only those which had the target variable as consequent, thus characterized as classification rules.

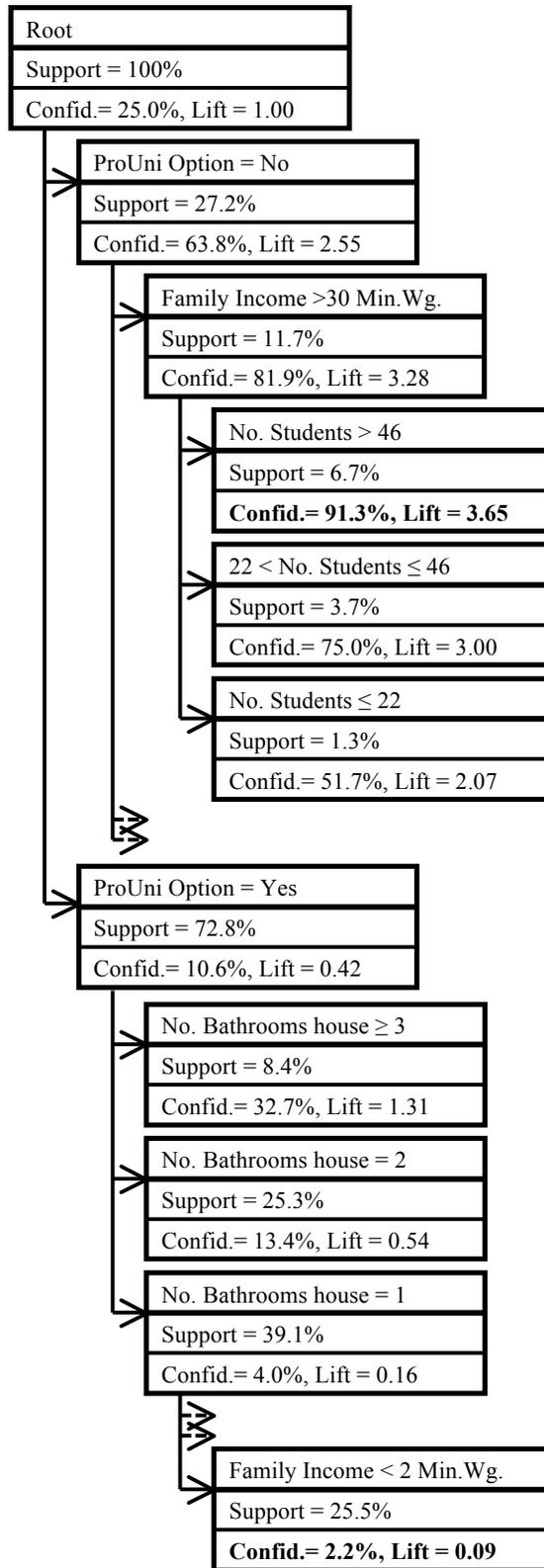


Figure 1. Branch of the Decision Tree Indicating Confidence and Support of each Rule Combined with Previous Ones until that Level.

Among the classification rules generated, the most relevant are those with the highest confidence diversion from the mean of the population either above or below that. These induced classification rules have their quality assessed independently of their support (coverage), thus allowing the discovery of small nuggets of knowledge which would pass undetected by the decision tree, for their low information gain which weighs the percentage of the population they represent (coverage).

The relevance of the rules was assessed by two metrics: confidence and departure of the lift from the average either above or below one (1). Confidence measures the quality of the rule by the relative frequency of examples of the target class within the hypercube selected. The lift is the ratio of the relative frequencies of the target class examples in the hypercube selected and in the population. The lift metrics indicates the increase of concentration of the target class examples in the selected niche compared to the sample average [Han *et al.*, 2011].

For verifying the statistical validity of the rules, a hypothesis test was applied to the relative frequency of the target class on each rule taken as a proportion [Jain, 1991]. The binomial distribution approximates the normal distribution in the calculation of the confidence interval with a given significance level ($1-\alpha=0.1$, in this work). That is when the rule's coverage (n) is taken into account, defining its confidence interval [Jain, 1991].

Further filtering eliminated the rules whose confidence intervals contained the relative frequency of the population, reducing greatly the amount of rules of interest. The last filter eliminated the rules with combined clauses that lost quality in relation to those formed by their component clauses taken individually. Clause combination already produces a loss of support (coverage) of the rules by their clauses' intersection. So it is not of interest to have additional loss of quality measured by the lift getting closer to one (sample average).

After sorting the remaining rules by decreasing order of their lift, Table 2 displays a few examples of manually selected rules.

Table 2. Classification rules induced with 1 and 2 combined clauses, assessed by the metrics support, confidence and lift.

Clause-1	Clause-2	Support	Confid.	Lift
Father Education \geq Graduate	School No. of Employees > 75	18.3%	62.1%	2.48
Father Education \geq Graduate		36.7%	51.6%	2.06
School No. of Employees > 75		32.5%	41.7%	1.66
School No. of Employees \leq 75		66.3%	11.6%	0.46
Mother Education \leq 5 th grade		4.2%	3.5%	0.14
Mother Education \leq 5 th grade	School No. of Employees \leq 75	3.8%	1.6%	0.07

The selected rules exemplify conditions which are not directly dependent on the economic power of the family; they are related to family education and school size. It is important to emphasize that the rules of combined clauses have improved lift

(compared to those of its component single clauses), in accordance to the filtering criterion above.

4. Performance Evaluation of the Propensity Score

Considering that the purpose of this paper is to present an application of a KDD solution, the main goals of the performance evaluation of the propensity score are checking the feasibility of the approach and estimating its performance. Different from algorithmic research which needs k-fold cross-validation process for performance comparisons, this work will only present a single estimate of the performance assessed on a held-out test sample.

The data sample within the scope of this research consists of 4,844 schools with the top quartile being "good" as already defined in Section II. This sample was partitioned in a modeling and a test sets containing 2/3 and 1/3 of the examples, respectively. The examples were randomly selected in a fashion stratified by the target variable, preserving the percentage of 25% of good schools in all sets.

Therefore, the propensity score performance evaluation was assessed on a statistically independent set with 1,616 examples having 404 from the target class of good schools.

Performance metrics

As there is not yet a specific use defined for this KDD solution in a decision support system, its operating point (decision threshold) has not been selected. In such scenarios, the performance assessment metrics should focus on a feature of the classifier itself, without assumptions on the operating point. Therefore, this paper will use the area under the ROC curve (AUC_ROC) [Provost and Fawcett, 2001], one the most widely accepted performance metrics for binary classifiers. This paper will also use the maximum Kolmogorov-Smirnov distance (Max_KS2) as a dissimilarity metrics [Conover, 1999] to assess continuous score classifier output, as commonly done in financial applications [Adeodato *et al.*, 2008].

The ROC Curve [Provost and Fawcett, 2001] is a non parametric performance assessment tool that represents the compromise between the true positive and the false positive example classifications based on a continuous output (the score) along all its possible decision threshold values (the operating point). In medical scenarios, the ROC curve equivalently expresses the compromise between sensitivity and specificity (actually, 1- specificity). Despite of the ROC curve being able to provide two performance metrics, this paper will use the AUC_ROC which is independent of the operating point. It is a performance indicator valid for assessing the performance throughout the whole continuous range of scores [Provost and Fawcett, 2001]. Considering binary decision on continuous range, the bigger the AUC_ROC, the closer the system is to the ideal classifier (AUC_ROC=1). If the ROC curve of a classifier appears above that of another classifier along the entire domain of variation, the former system is better than the latter.

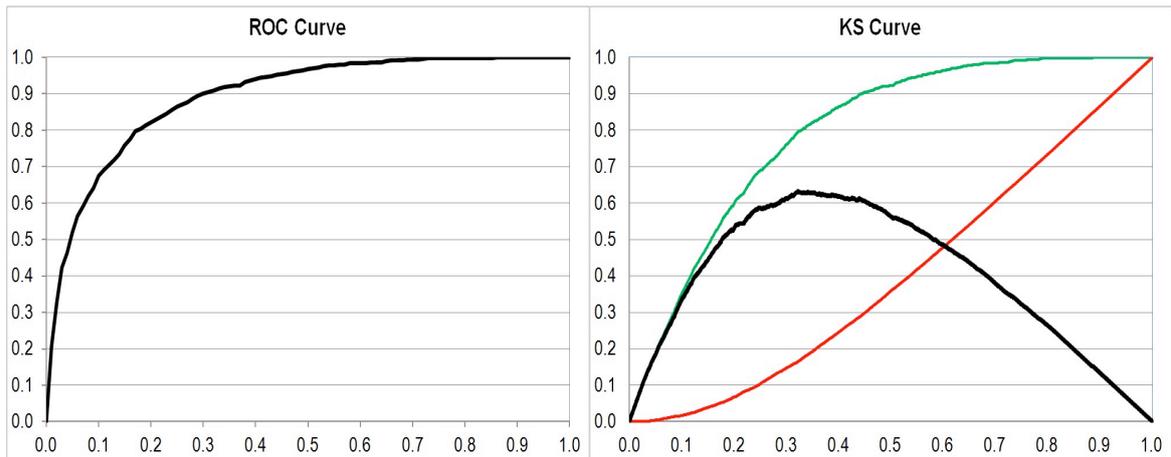


Figure 2. ROC (Receiver Operating Characteristic) and KS (Kolmogorov-Smirnov) Curves for assessing performance of the propensity score.

The Kolmogorov-Smirnov distribution has been originally conceived as an adherence test for distribution fitting to data [Conover, 1999]. In binary classification problems, it has been used as dissimilarity metrics for assessing the classifier's discriminant power measuring the distance that its score produces between the cumulative distribution functions (CDFs) of the two data classes [Adeodato *et al.*, 2008], known as KS2 for dissimilarity purposes. The usual metrics for both purposes is the maximum distance between the CDFs, which is invariant to score range and scale making it suitable for classifiers' comparisons. However, that is an assessment at a single operating point that, in general, does not suit the applications' needs.

Results and interpretation of the propensity score

The results presented here refer to the effect of the score assessed on a statistically independent dataset with 1,616 examples, being 404 from the target class, a random sample partitioned from the whole dataset, stratified according to the target variable. As the whole sample referred to a single year (2011), the generalization of these results should be used with caution.

Figure 2 shows the ROC and KS2 curves for the score variable. Their monotonically decreasing slopes confirm that the logistic regression model was able to learn the task consistently. The quality of the propensity score was measured by the performance indicators described in the previous subsection, namely the area AUC_ROC and the Max_KS2 which attained the values of 0.897 and 0.632, respectively.

5. Final Considerations

This paper has analyzed and extracted knowledge from the databases of ENEM and School Census 2011 for assessing the quality of the Brazilian private secondary schools. Even in such large bases and various grains, the CRISP-DM method along with the Domain-Driven Data Mining approach allowed to 1) characterize the "good" school as a goal and 2) extract knowledge focused on 3 functionalities of a decision support system towards the defined objective.

Logistic regression produced a classifier to generate a propensity score to estimate the success of a school, from its own characteristics and those of its faculty and students and families. The score attained high discriminating power measured by $AUC_{ROC} = 0.897$ and $Max_{KS2} = 0.632$. The coefficients showed that the economic factors are relevant appearing both directly, in family income and their household features and, indirectly, in the country's region where the school is located at. It has also shown that computer laboratory infrastructure is a bad school feature.

Both the economic influence and the computer lab should be interpreted with caution. For the former, frauds have been discovered in three out of the last five years of ENEM exams and they amplify the effect of the economic power on the results. For the latter, the knowledge about the labs is counter-intuitive and it needs to be checked if the labs have been used for educational purposes or just for entertainment (game playing and social networking), or even worse, used for entertainment during or instead of classes.

The decision tree extracted knowledge explaining how the human expert would decide sequentially using rules. Other techniques generated similar rules that corroborated the results and made clear that the main factors that influence the quality of schools are the economic and financial aspects, either directly (family income) or indirectly (No. of bathrooms in the household and ProUni option) or cultural aspects of the family (level of education of the mother or father).

The article has presented a data mining solution to evaluate and predict the quality of secondary schools, aimed at its future deployment in a Decision Support System for operation and navigation in real time. Similar study was developed for public schools and confirmed the effectiveness of the public policies for school free lunch and teacher pedagogic training.

Another study is underway to enhance the information gain in the transformations necessary to change granularity in the ETL (Extract, Transform and Load) process. Preliminary results on 2012 ENEM and Census data with the new approach [Adeodato, 2015] have shown that the impact of the students' parents and teachers education is much higher than the existing methods were able to capture.

References

Adeodato, P.J.L.: Variable Transformation for Granularity Change in Hierarchical Databases in Actual Data Mining Solutions. In Proc. International Conference on Intelligent Data Engineering and Automated Learning (IDEAL), Wroclaw, Poland, pp. 146–155, 2015.

Adeodato, P.J.L.; Vasconcelos, G.C.; Arnaud, A.L.; Cunha, R.C.L.V.; Monteiro, D.S.M.P.; Oliveira Neto, R.F.O.: The Power of Sampling and Stacking for the PAKDD-2007 Cross-Selling Problem, International Journal of Data Warehousing & Mining, 4(2), pp. 22–31, 2008.

Agrawal, R.; Srikant, R.: Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pp. 487–499, Santiago, Chile, 1994.

Cao, L.: Introduction to Domain Driven Data Mining. In: Cao, L., *et al.* (eds.) Data Mining for Business Applications, pp. 3–10, 2008.

Censo Escolar. Available (in Portuguese) at: <<http://portal.inep.gov.br/basica-censo>>. Last access:18 / 11 / 2014.

Conover, W.J.: Practical Nonparametric Statistics. Chap. 6, 3rd ed. New York: Wiley, 1999.

Darling-Hammond, L.; Wentworth, L.: Benchmarking learning systems: Student performance assessment in international context. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education, 2010.

ENEM. Available (in Portuguese) at: <<http://portal.inep.gov.br/web/enem/sobre-onem>>. Last access:18 / 11 / 2014.

Hair JR.; Anderson, R.E.; Tatham, R.L. *et al.*: Multivariate Data Analysis. 5th ed., Porto Alegre, Bookman, 2005.

Han, J.; Kamber, M.; Pei, J.: Data mining: Concepts and techniques. San Francisco: Morgan Kaufmann, 2011.

INEP. Available (in Portuguese) at: <<http://portal.inep.gov.br/>>. Last access:18 / 11 / 2014.

Jain, R.: “The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling”. Wiley-Interscience, 1991.

Johnson, R.A.; Wichern, D.W.: Applied Multivariate Statistical Analysis, 6th ed. Pearson Prentice Hall, Upper Saddle River, NJ, 2012.

Kumar, P.R.; Ravi, V.: Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. European Journal of Operational Research, 180, pp. 1–28, 2007.

Provost, F.; Fawcett, T.: “Robust Classification for Imprecise Environments.” Machine Learning Journal, vol.42, n.3. pp. 203–231, March 2001.

Safavian, S.R.; Landgrebe, D.: A survey of Decision Tree Classifier Methodology. IEEE Trans. Systems, Man and Cybernetics, vol. 21, pp. 660–674, May/June 1991.

Shearer, C.: The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing* , 5 (4), pp. 13–22, 2000.

Stecher, B.: *Performance Assessment in an Era of Standards-Based Educational Accountability*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education, 2010.

Sousa, M.U.R.S.; Silva, K.P.; Adeodato, P.J.L.: Data Mining Applied To The Processes Celerity Of Pernambuco's State Court Of Accounts. In: Proc. CONTECSI 2008 (in Portuguese).

Travitzki, R. "ENEM: limites e possibilidades do Exame Nacional do Ensino Médio enquanto indicador de qualidade escolar". Ph.D. Thesis, USP, São Paulo, 2013 (in Portuguese).