Mpar-Cluster: Applied Algorithm of Geo-Selection for Optimization of the Credit Recovery of Electricity Supply

Augusto César da Silva Machado Copque and Mateus Prates de Andrade Rodrigues

Universidade Católica do SalvadorandCompanhia de Eletricidade do Estado da Bahia, Brazil

Abstract: This paper presents a study of optimization of operational recovery credit default with geoprocessing use through geoprocessing tools, developed in the Receivables Management Companhia de Eletricidade do Estado da Bahia-COELBA sector. The work was initially based on the application of Data Mining Tools for Software KNIME 2.9 and later use of the tool of GIS-ArcGIS 10.X/ESRI .Were evaluated and applied analytical processing algorithms, to improve the process of spatial selection and define the best sets logistical credit recovery. The focus study, based on geoprocessing use in cutting action is due to the fact that this process has the largest collection efficiency. It is understood that the efficiency of the cutting action, should the great importance that electricity has on modern life. The research was guided its evolution from analysis of algorithms agglutination and georeferenced database, whose focus was and is acting in the cutting action due to the fact that this process has the largest collection efficiency. For the implementation of the study, through some geoprocessing techniques, the Mpar-cluster, this optimization model was developed from a custom algorithm for spatial selection, that had with subsidy: information stored in geographic databases, database information alphanumeric, images (aerial photographs and satellite images), text files, and digital tables.

Key words: Geoprocessing, recovery credit, mpar-cluster.

1. Introduction

This article presents a study of operational and financial optimization of the supply of electricity cutting action from analysis of agglutination algorithms and basic georeferenced data. For the application of optimization modeling proposals were used data mining platforms and Geography Information System (GIS) or Geographic Information Systems (GIS).

The focus of this study, based on geotechnology[1]

Corresponding author: Augusto César da S. M. Copque, Geographer, master of urban environmental engineering, management expert planning and Professor at the Catholic University of Salvador-UCSAL, research fields: geoprocessing, geointeligence and geoscience.

[1] Geprocessing/Geotechnology is a GIS operation used to manipulate spatial data. A typical geoprocessing operation takes an inputdataset, performs an operation on that dataset, and returns the result of the operation as an output dataset. Common geoprocessing operations include geographic feature overlay, feature selection and analysis, topologyprocessing, rasterprocessing, and data conversion. Geoprocessing allows for definition, management, and analysis of information used to form decisions.

[5, 6] use in the cutting action is due to the fact that this process has the highest collection efficiency. It is understood that the cutting action of efficiency, it should be the great importance that the electricity has in modern life, with the electricity directly related to physiological items of the base of Maslow's needs pyramid [2].

However, the traditional model of disconnection of the power supply is carried out by a specialized workforce, the more expansive significantly the action.

Geoprocessing is a framework and set of tools for processing geographic and related data. The large suite of geoprocessing tools can be used to perform spatial analysis or manage GIS data in an automated way. A typical geoprocessing tool performs an operation on a dataset (such as a feature class, raster, or table) and creates a resulting output dataset. For example, the Buffer tool takes a map layer as input, creates areas around the layer's features to a specified distance, and writes those areas to a new output layer.

In addition to the suite of tools, geoprocessing also has a powerful framework that supports control of the processing environment and allows you to build custom tools that can further automate your work. You can use the geoprocessing tools included in ArcGIS as building blocks to create an infinite number of custom tools that automate repetitive tasks or solve complex problems.

Thus, It has been the cutting action a great setting for an optimization study, where an efficient and high operating costs action, provides scalable gains.

For the development the study optimization algorithms were applied in Data Mining as: K-means and *Equa k-means*[3]. Posteriorly, a second optimization model from a customized geographic selection algorithm was developed. This algorithm has been "baptized" as map-cluster or maximum pick amount per radius-cluster (a*glomerado de valor máximo de seleçãoporraio*)[2]. For the application of map-cluster were used some geotechnology techniques in a GIS environment (Fig. 1), with subsidy information stored in geographical databases, database of alphanumeric information, images (aerial photographs and satellite images) texts, tables and digital files.

Thus, this study demonstrated the possibility of better earnings by reviewing the collection of operational processes in the organization studied through the use of geographical knowledge, based on the geo-referenced location of the clients in the Bahia of state. The work was initially based on the application of tools for Data Mining Software KNIME 2.9 and later

Fig. 1   Estrutura de um SIG. Fonte Medeiros 2012.

the use of GIS tools — ArcGIS 10.x/ESRI. Were evaluated and applied analytical processing algorithms to improve the spatial selection process and define the best logistics sets of credit recovery. The study analysis focuses on demonstrate the results of operations more efficient strategic and financial charges on portfolios [4].

2. Materials and Methods

2.1 Structure

In Companhia de Eletricidade do Estado da Bahia (COELBA)[3], the operational process of default loan recovery is planned by the Management Unit of the Receivables — CRGR that uses different non-payment combat activities, the expected date of electric energy, planning actions such as warning notices, Short Message Service (SMS), Response Unit audible (URA), negativity, power cut, cut tracking, low administrative, billing advisory and judicial collection. The management of flow collection of default is managed through the Customer Success Story tool — Company CCS SAP, with "rules" specific collection. These rules are parameterized for different clients profiles and following the regulatory guidelines of Resolution No. 414/2010 of ANEEL [1].

The focus of this study was directed to supply cutting action of electric energy. The regulatory control of susceptibility to activity is managed by SAP CSS, the unit being provided a daily list of customers susceptible to power supply suspension action (cut). Currently the list of susceptible is handled in a proprietary system called Information Control and Collection (ICC). This system treats the list provided by SAP CCS, complementing financial information

---

[2] Algorithm developed by the authors to receivables management. The sea-cluster is used for geospatial selection of collection cutoff scores. The algorithm can also be applied to identification customers susceptible to new debts, negated and problems in power distribution

[3] The Companhia de Eletricidade do Estado da Bahia is the third largest electricity distributor in the country in number of customers and the sixth in power supplied volume. These same terms, ranks first among the North dealers — Northeast. Controlled by Neoenergia Group, one of the largest investors in the Brazilian electricity sector, Coelba is present in 415 of 417 municipalities in Bahia, serving more than 14 million inhabitants in a concession area of $563,000$ km^2. Today the company has over 5.6 million customers, 88% of these residential customers.

and automatically applies the logistical limitations of availability of Providers Outsourced Business Services — EPS.

For the implementation of the concept of limitation contractual logistics of EPS, the concession area COELBA is divided into 43 centers of work. ICC system (Fig. 2), these 43 work centers are divided into target zones, these zones represent sets of

neighborhoods in the municipal seat or group of municipalities — in the least populated areas. The zones are further subdivided into cluster, which represents sets of neighborhoods neighbors. These definitions of zones and clusters are parameterized empirically, based on the field knowledge of operational coordinators of the respective work centers.

Fig. 2   ICC. Source: COELBA, 2014.

The algorithm used by the ICC for selecting serial succinctly defined by the following steps:

(1) Select a particular zone Work Center — CT to set the selection;

(2) It makes up the summarization of the number of clients per group;

(3) Sort up zones of decreasing order by the charged total;

(4) Select first zones between the amount of service, depending on the logistical limitation;

(5) Repeats 1 through 4 the process to be performed to check banknotes for all areas of all TCs limited to daily target set for each zone;

The selected amount is increased by a percentage of 15% compared to the target area in order to facilitate the composition of sets of notes from classes by EPS. In addition, it adds up over 10% from the percentage of notes identified as paid during field service.

As an example, stands out in Table 1 the selection of the groups 22, 23, 1 and 2 of the sample under study referring to the day 05/20/2014. In Table 1, were summarized customers likely to cut according to the groups being classified the clusters in descending order of sum of debt collect. In this example are highlighted 1007 client where the program selected 1,000 customers, representing: 800 customers regarding the

cutting target this area of the work center, plus the 10% paid account and 15% of routing.

This selection of the ICC is directed SAP CCS, that generates the service and provides the following day the cutoff score for its EPS. In EPS, a staff with field experience brings together the notes by class and dispatches to their field technicians. The notes not dispatched by EPS are handled by SAP CCS and subsequently made available to a new selection of ICC.

In Fig. 3, there is an example of the spatial distribution of the notes being displayed on the Google Earth software, relative the generation and dispatch the classes on 05/20/2014. The icons (balloons) in shades of blue show the orders per group, coming from the susceptible portfolio selection process. In addition, perceives in the sameFig. 3 some stains yellow to red, representing, respectively, the smaller the higher concentrations of debts.

In a closer view of the EPS of routing, there is an overlap of activities of the groups, as well as an error of dispatch note orders that impact on an increase displacement in group 2, see Fig. 4.

Analyzing the current model selection, occur improvement points from a selection: more in line with the debt stains with smaller radius of action, with greater concentration and better order per group. From this universe of study, were evaluated two proposals for optimization and debt assemblage, one for analysis of statistical tools for Data mining using clustering algorithms, and another by the mpar-cluster own algorithm, implemented in ArcGIS/ESRI 10.

Table 1   The first lines of the zone groupings 1 do CT of Praça da Sé.

Grouping	Debt	Average	QTD
22	R$ 47'314.43	R$ 142.51	332
23	R$ 42'042.14	R$ 178.90	235
1	R$ 39'326.92	R$ 185.50	212
2	R$ 32'577.55	R$ 142.88	228
16	R$ 29'218.35	R$ 210.20	139
13	R$ 28'277.90	R$ 235.65	120
12	R$ 27'669.40	R$ 172.93	160
7	R$ 25'705.06	R$ 163.73	157

Fig. 3   Satellite image containing the selection of notes and debt stain. Prepared by the authors, 2014.
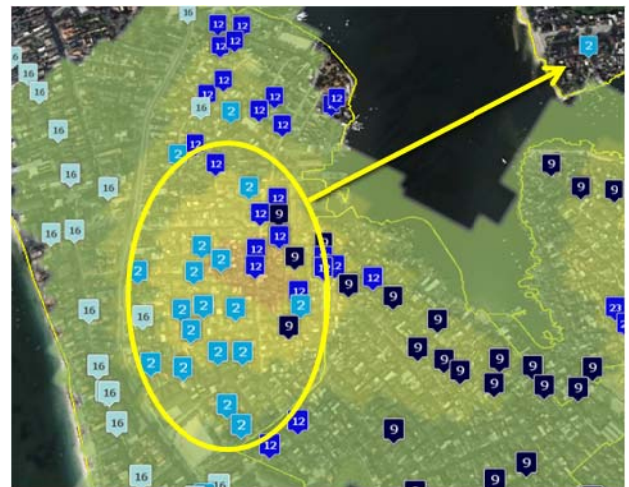

Fig. 4   Overlapping classes and error in the distribution of cutoff scores for group 2. Prepared by the authors, 2014.

2.2 Analysis of Selection of Optimization Proposals

2.2.1 Optimization in Data Mining

The term data mining (mining and mining data) is a part of Knowledge Discovery in Databases (KDD) (knowledge discovery in databases) that deals with knowledge discovery of methods in large amounts of data (Fayyad, Piatetsky-Shapiro, Smyth & Uthurusamy, 1996).

For the study in question was used KNIME software in the free version No 2.9 of the open source type (open source). The choice of the software in question is due to ease of use, with an extensive pre-installed library blocks drag and drop type. Among the available algorithms, there is the K-means. An agglutination or

clustering algorithm that aims to break "n" observations "k" sets where each observation belongs to the closest cluster average, see Fig. 5. This results in a division of the data space on a Voronoi diagram, a type of spatial decomposition where the sample segments from distance of relational analysis. In Fig. 5 it can be noted "n" points being divided into three colors, where for defining the color of each point was performed one distance calculation between the other points are assigned a color according to the proximity between individuals.

The geographic coordinates used in KNIME were previously treated in regional geodetic system for South America — South American Datum — SAD 1969 in decimal degrees. Was made necessary normalize latitude dimensions longitude and debt due to the large variation in the data of the latter compared the draft decimal variation of latitude and longitude coordinates, when converted to decimal degrees. This action was taken to mitigate agglutination trend of distance calculation for the dimension debt.

Even with the normalization of the values of debt shaft, it can be seen in the studies a grouping tendency by the distance of debt, getting more clusters grouped by this value due.

In a second stage of the study, it was removed from the size debt and applied the K-means the algorithm for dimensions latitude and longitude. During the study it

Fig. 5   Graphical representation of the result of a clustering. Source: dartmouth.edu, 2012.

was ratified a limitation of the k-means algorithm linked to the development of the same functionality; it is not possible to create clustering with predefined limits on the number of subjects per group.

From that point, we tried to search clustering algorithms that allow to define the amount of individuals by himself. Thus, it was identified and implemented a new mining algorithm developed by Pierre-David, titled EKmeans. This tool developed oriented programming language object — Java, which uses the concepts implemented in heuristics expression K-means, calculated according to the Euclidean distance, but limiting the number of individuals algorithm for clustering. This Java application was limited to two dimensions, which was already studied in line with the idea of using only the latitude and longitude dimensions for agglutination. By applying the algorithm EKmeans found interesting results which will be shown in the comparative topic between models

2.2.2 Geotechnology Optimization

In addition to the challenging modeling developed in Data Mining, we sought to develop and implement a solution focusing on geotechnology tool. The proposed algorithm developed and has had a focus: identify the client more limited debt amount to a maximum radius of a group action and enhance the clustering with higher note values to the limit of notes per group.

The use of GIS through ArcGIS/ESRI software was necessary due the search of geographical positioning of each customer or the proximity of this with the adjacent post. The location of the customers and the relationship of susceptible to charging by zone or neighborhoods of CT's were fundamental to define the best route to follow and efficiency in the collection, especially with regard to higher values.

The implementation of the designed algorithm was made with the partnership between an Electrical Engineer, postgraduate in Business Management (RODRIGUES, M.) and Geographer, specialist in GIS (COPQUE, A.). Thus, the use of GIS tools enabled the application of mpar-cluster algorithm in the
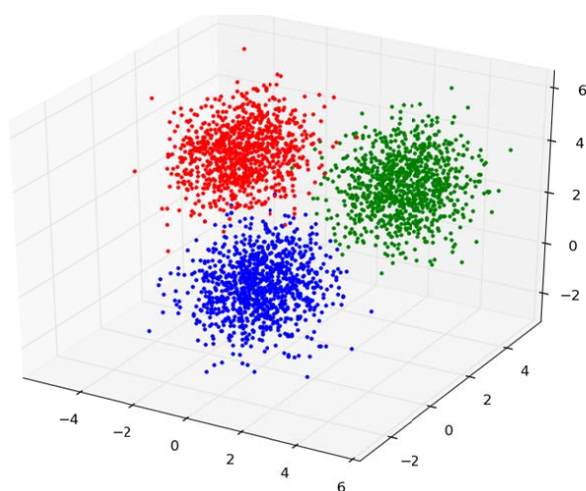
Company's business model, this application can be exemplified, succinctly, according to the sequence below:

(1) For each Work Center — CT (of the 43 existing) and zone list is georeferenced form all customers susceptible to cutting action;

(2) Among the list of susceptible, you select the highest debt value;

(3) Considering the previous study of a radius by zone if list to all customers around this greater debt to its maximum radius of action;

(4) Selects the highest debts in the action radius to form the amount of notes, the group will act;

(5) Repeat step 2-4 to form all zone groups concerned;

(6) Repeat step 2-4 to all classes of all the work centers by area;

In Fig. 6, there is the application of the algorithm mpar-cluster with the selection of two (2) clusters being displayed on the Google Earth software, referring to customers susceptible to cutting action on 05.20.2014 for CT Praça da Sé. The icons (points) in dark blue represent all customers susceptible to cutting action. The icons (points) in red represent the largest debts to be collected. The red circle represents the radius of action of group, defined from the best results of the field action history. Os ícones (pontos) azul claro representam os clientes suscetíveis e dentro do raio de proximidade da maior dívida. The orange icons (items) representing the likely customers, with larger values
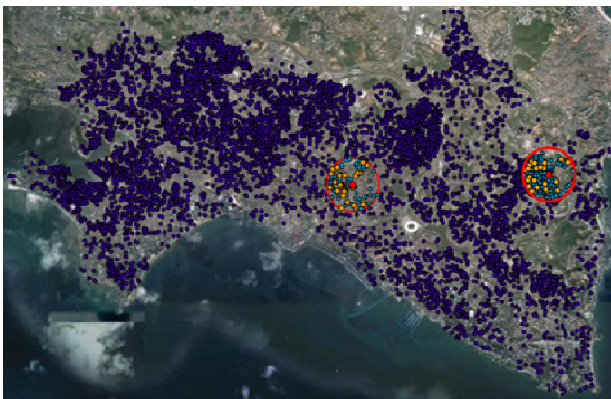

Fig. 6   Application of the algorithm mpar-cluster. Prepared by the authors, 2014.

and debt within the proximity distance higher debt to complete the total customer triggered by a group.

With the application of mpar-cluster were found algorithm further significant results and that will be as follows concerning the comparison between models.

3. Results and Discussions

3.1 Comparison between Models: Current, Mining, GIS

For the comparative evaluation of results of models were developed geographic representations of selection using the gadget ARCGIS 10.x/ESRI and displaying the software Google Earth. Furthermore, it was made a comparative Chart summarizing the information for a global day of action group.

Selecting the view of the ICC cutting portfolio it was presented in "2.1 The structure" through Fig. 3. In that image was observed optimization opportunities, ratified by divergent selection in relation to debt stain concentrations.

On the optimization in Data Mining, it was found in a study that the EKmeans algorithm had a better result than the study of application of K-means. Thus, it was made to Fig. 7, which depicts the application of EKmeans algorithm, referring to customers susceptible to cutting action on day 05.20.2014 for CT of Praça da Sé. The icons (balloons) in shades of blue show the orders per class, resulting from the application of EKmeans algorithm in the database of customers susceptible to cutting action. Furthermore, it appears the same spot due detailed in Fig. 3.

In a visual comparison between Fig. 7 and Fig. 3 it can be seen graphically EKmeans better performance compared to model the ICC. The clustering algorithm generated by EKmeans have a greater alignment between classes and the debt stains. However, it was still possible to see the application of the model EKmeans earning opportunities tied to debt stains not triggered and the dispersion of the class action radius. In this case, it was found for the cluster number 15 one

distance greater than Km 3, and a light scattering cluster number 13.

Fig. 8 was made for a graphic display of the application of Geotechnology optimization model. This image depicts the application of mpar-cluster algorithm in customer base susceptible to cutting action, on day

05.20.2014 for CT of Praça da Sé. The icons (points) in reds show orders per group, resulting from the application of mpar-cluster algorithm in the database of customers susceptible to cutting action. Moreover, a similar stain detailed debt is shown in Fig. 3.

Fig. 7 Application of the algorithm *EKmeans*. Prepared by the authors, 2014.
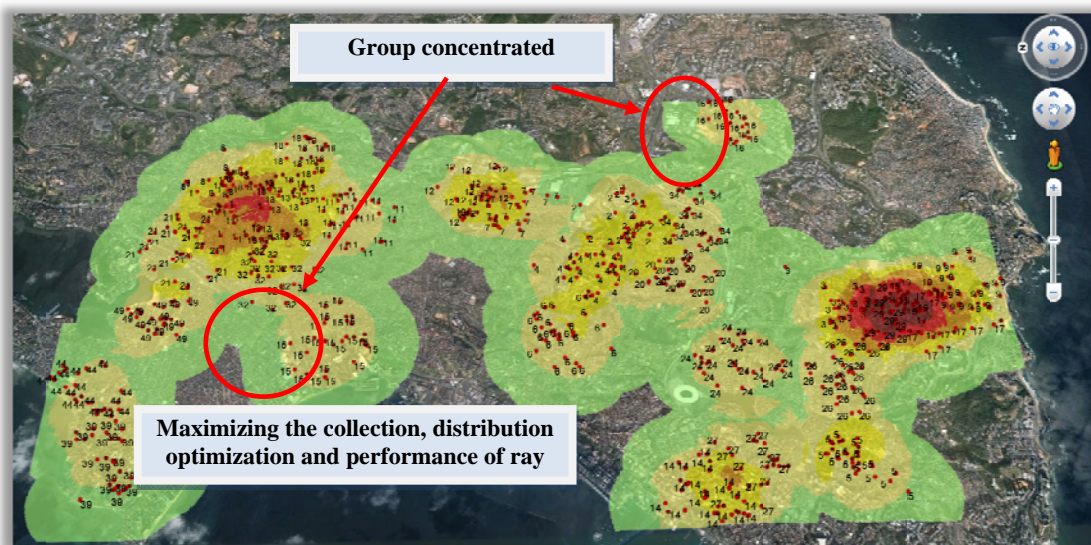

Fig. 8 Application of the algorithm mpar-cluster. Prepared by the authors, 2014.

By visually comparing Figs. 8, 7, 3 and Table 2, it validates graphically the best performance of the mpar-cluster algorithm (Fig. 8) compared to EKmeans (Fig. 7) and the ICC (Fig. 3). The clustering generated by mpar-cluster algorithm have a strong alignment between group and the debt stains. Besides having

more control over the radius of action of classes, thus avoiding cluster dispersion.

In Table 2 it is possible to identify the measured methods, the total charged by the mpar-cluster algorithm was 61% higher than the current model selection. This gain was recorded even though the

amount of notes 1.47% lower than that generated by the ICC selection. Regarding the performance of radius, it can be seen graphically that the mpar-cluster algorithm provides a reduction of the radius of action of group and may even increase the amount of notes executed a day ahead to earnings by reducing shifts.

3.2 Ratification of Geotechnologica Optimization Gains

Aiming to verify the gains from the application of mpar-cluster algorithm was made a new model of the application of simulation in Ilhéus work center in Canavieiras area. He was noticed in Fig. 9 and Table 3, a gain of over 60% compared to the average decrease of the group action radius. Moreover, with the same effective seven (7) classes it was possible to trigger the collection of R$ 187,900.56, representing a gain of 440% over the proposal generated by the ICC algorithm.

The tabulation of the geostatistical analysis used for CT Ilhéus, demonstrates the validity of the study in question and that this tends to bring organizational and financial gains for the sector and thus for the Company.
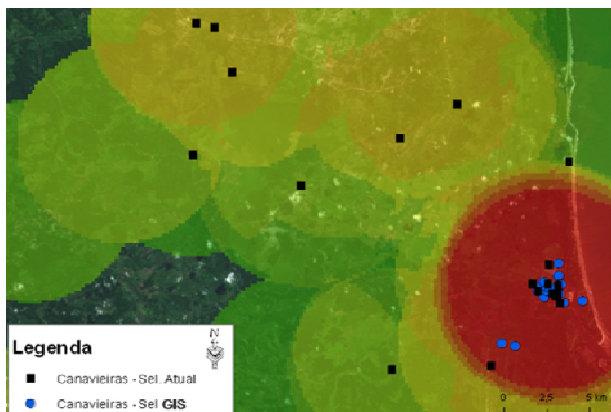

Fig. 10   Comparison between the current selection and GIS for CT Ilheus zone Canavieiras. Prepared by the authors, 2014.

Table 2   Comparative summary of the court of selection models.

Modelagem	QTD Turmas	QTD Notas	Total Cobrado	% Cobrado
Atual	23	747	R$135.095,03	100%
Ekmeans	23	732	R$164.590,76	122%
Algoritmo GIS	23	736	R$217.704,03	161%

Table 3   Comparing the Court of model selection in CT Ilheus.

Metodo	Turmas	Notas	Raio	Valor	%
SelecaoAtual	07	122	15 km	R$42.665,84	100%
Selecao GIS	07	126	5 km	R$187.900,56	440%

The implementation of the geographical concept (see Table 3 — Selection GIS) research was identified in substantial gains, as the amount of notes, action radius and percentage susceptible to collection.

Regarding the reduction of the radius of action, for example, there is a significant logistical gain when compared with the present situation and this selection leads geospatial/entail smaller displacements. In parallel with this it is also possible to have the exact location of the client maps (Digital Printing) from a structured base map.

4. Conclusions

The developed research sought to guide the potential of this study and the main geographical nature of returns, organizational and financial, to the detriment of the use of GIS. To this end, we have been tested and simulated clumping algorithms, focusing on optimizing logistics selecting customers susceptible to energy supply suspension action. All this study was also informed by concepts of GIS anld use of an organized base map.

During development of the research we were tested the applying various clustering algorithms, especially when the results achieved in implementing the mpar-cluster algorithm. The possible gains presented triggered increased value for the charge, which enables a proportional increase of the total amount recovered. In addition, the focus on defining the action radius, provides an increase in the amount of notes made by the forward group displacement economy circumvented by reducing the dispersion of notes.

After this study, it was hired by this company a Geographer/GIS Analyst, which collaborated with the implementation of a pilot project to use this optimization logistics. During deployment, already practical gains of mpar-cluster algorithm has been

validated, and other gains related to: basic treatment georeferenced customers; Studies of best access routes; Redefinition of areas of workplaces, based on customer sensitivity, secondary mesh network of energy and support roads;

The optimization pilot project selection via GIS was fully deploying in September 2014.

References

[1] Agência Nacional de Energia Elétrica – ANEEL, Resolução nº 414, Disponível em: http://www.aneel.gov.br/biblioteca/downloads/livros/REN _414_2010_atual_REN_499_2012.pdf, Accessed March 01, 2014.

[2] Maslow, A. 1962. Introdução à psicologia do ser. Rio de Janeiro: Eldorado.

[3] Jain, A. K., Murthy, M. N., and Flynn, P. J. 1999. "Data Clustering: A Review." ACM Computing Survey 31 (3): 264-323.

[4] Oliveira, D. De P. R. 2005. Planejamento estratégico – conceitos, metodologia, práticas (22nd ed.). São Paulo: Atlas.

[5] Davis, C., and Câmara, G. 2014. "Arquitetura de Sistemas de Informações Geográficas." Accessed: http://www.dpi.inpe.br/gilberto/livro/introd/cap3-arquitetu ra.pdf, July 01.

[6] Câmara, G., and Monteiro, A. M. V. 2014. "Conceitos Básicos em Ciência da Geoinformação." http://www.dpi.inpe.br/gilberto/livro/introd/cap2-conceito s.pdf, August 15.