Classification Approaches to Foster the Use of Distributed Generation with Improved Remuneration

Cátia Silva, Pedro Faria and Zita Vale
GECAD -Research Group on Intelligent Engineering and Computing for AdvancedInnovation and Development,
IPP - Polytechnic of Porto
Porto, Portugal
cvcds@isep.ipp.pt;pnf@isep.ipp.pt;zav@isep.ipp.pt

***Abstract*—There are currently efforts to implement the concept of smart grids throughout the electric sector. This will bring radical changes to the entire management of the sector. The energy market does not run away from the rule. In this way, virtual power players will be required to update their business models to introduce all the concepts that the context of smart grids imposes. Thus, in this article is proposed a method that aggregates distributed generation and consumers who belong to demand response programs. Optimized scheduling, resource aggregation and classification of possible new resources, rescheduling, and remuneration are the phases of the methodology proposed and presented in this article. The focus will be on classification phase and the main objective is to create rules, through a previously trained model, to be able to classify the new resources and help with the challenges that virtual power players may face. Thus, five classification methods were tested and compared: neural networks, Bayesian naïve classification, decision trees, k-nearest neighbor method, and lastly support vector machine method.**

Keywords— Distributed Generation, Naïve Bayesian, Neural Networks, k-nearest neighbor, Decision Trees, Support Vector Machine

I. Introduction

Currently, the electricity market is facing a major change. Companies in the sector should be concerned about the introduction of competition and so on the possibility of new competitors, market deregulation and new participants, for example, consumers with the possibility of producing their own energy. To succeed and obtain better results, it will be essential to understand their consumers, the main characteristics of their load profiles, the distributed resources, the market and all that can influence this sector, [1].

Business models must adapt to the constant innovation and the implementation of Smart Grids. In the current market, the advantages that application of this concept can imply are innumerable, being able to emphasize one of its main technologies, Demand Response (DR), [2]. The main benefits can be seen by reducing the load peaks and even changing the load diagram. With this, it would be possible to save investments necessary to carry out the application of this concept. Several researches and studies have been done in this area, considering residential and commercial consumers that can take part in DR programs but still a little unexplored. In this way, not only the part of consumers must be considered, the production is also part of this new concept. Smart Grids also encourage the use of distributed generation (DG).

Thus, virtual power players (VPP), aggregating these small resources, DG and DR, can enable the dissemination of their participation, and can make them more profitable through new approaches for the energy market. That said, there is a need to improve business models on their approach to the reward DG and DR for their participation. The methodology presented in this article can be considered a solution presenting all the necessary instruments for the management of resources, the minimization of operating costs and the remuneration due to each of the participants, [3].

In this article, the main focus will be in the classification phase, where several methods will be put to the test. In this way, the goal is to create ways to classify new resources through existing classification models belonging to the aggregator. The purpose of these models is to predict the class of the new scenarios through rules applied to input parameters and, therefore, avoiding another complex optimization, [4]. The models should be pre-trained with a set of data labeled as training. Subsequently, new objects will be presented to test the accuracy of the model.

Being considered a development of the article [5], in this paper were used results obtained from resource scheduling to test the different classification methods. A more detailed explanation will be presented in section III. Several scenarios compose the database studied. The variation of sample size for training and testing was also one of the tests faced by the methods.

Section I presents the theme related to the study carried out and presented throughout the article, in addition to the purpose of it. Section II refers to the type of approach the authors chose to ease the tasks performed by an aggregator, presenting a detailed description of the proposed method. Regarding section III, a brief description of each classification method studied is presented. The case study is only presented in section IV and section V, then the results will be displayed and later analyzed. Finally, section IV with the conclusions drawn from the study.

The present work was done and funded in the scope of the following projects: CONTEST Project (P2020 - 23575), and UID/EEA/00760/2013 funded by FEDER Funds through COMPETE program and by National Funds through FCT.

.

II. Approach

In this section will be presented the method proposed by the authors. This method is presented in Fig.1. and its divided into five main phases, which will be explained in detail throughout this segment.

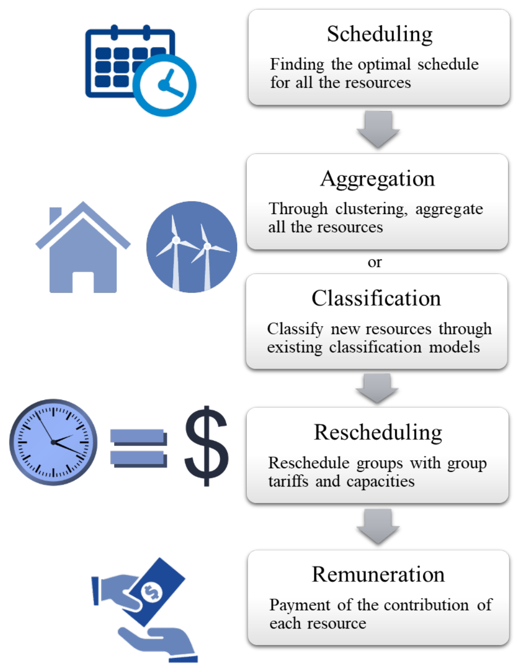


Fig. 1. Proposed Methodology

Fig.1. shows the method proposed by the authors to show how an aggregator introduce himself to the network infrastructure and how can handle the electric market change. In the first phase, scheduling, an optimization is performed to schedule all the energy resources – DR consumers, DG units and suppliers. The goal is to obtain the minimization of operating costs from the point of view of the VPP, considering price and operation constrains. In this way, a nonlinear problem was generated, and the solutions of this optimization were generated through the potential of MATLAB software, through the toolbox TOMLAB, [6].

The second phase of the method proposes the aggregation of all resources to create distinct groups so VPP can have a reasonable amount of energy to negotiate in the market. In this way, a clustering method is used to group resources. Classification is another way of generating groups. Through the results obtained from the optimization it will be possible to group new cases with rules from previous classification models created for the purpose. For both aggregation and classification, the generation of results came from software R.

Regarding third phase, called rescheduling, the goal is creating a tariff for each group of resources. In this way all elements in one group will be remunerated with the same value. Here maximum price was applied for the creation of the group tariff. The resource remuneration phase is used as motivation and as a way to encourage continued collaboration of all the resources associated with the aggregator in the operation of the network. The final remuneration is paid through the group tariffs obtained in the previous phase.

III. Methods

Classification can be explained as the task of assigning an object belonging to a database to a particular group, class, taking into account a set of attributes that are acquired and which describe this object, [7]. Supervised multiclass classification algorithms have the purpose of assigning a label to each new sample presented to the model, taking into account a training database where previously the algorithm found a learning pattern that classified each class, [8]. This problem can be solved by naturally extending the binary classification practices presented by certain algorithms. Some of these will be studied throughout this paper and are presented in Fig.1.

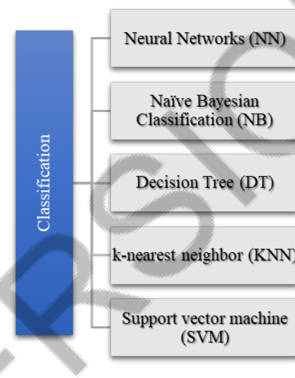


Fig. 2. Classification methods

The application and research of the topic related to neural networks (NN) is one of the most active in what concerns to theme of Classification. Indeed, NNs have been shown to be an extremely useful tool in classification and regression models, once challenged by Support vector Machine (SVM) and even Random Forests. However, with the introduction and the beginning of deep learning, the potential and superiority of NN was more noticeable for cases with more complex data, [5]. As the name implies, these networks simulate the neuronal structure of the brain. Neurons are nothing more than a set of input values and weights associated with them, a function that sums the weights and directs the results to an output. Composed by several layers: the input layer, the hidden layer, and the layer of output. The first consists of the registration values for the entry in the next layer. There may be several hidden layers, these being the intermediates an activation function produces the outputs consider the weighted inputs. In this study, the author considers two hidden layers with 4 and 3 neurons. Finally, the output layer that will take the object to the assigned class. One of the problems to which NN is the sensitivity to overfitting, being over-trained with the same samples for learning, may be adjusting to these and impairing their ability to predict, [3]. In this study all the methods studied were trained only once to compare in a similar base.

Bayesian methods are the root for traditional statistical classification methods. Here, an implicit probability model must be used to calculate the probabilities after the decision for classification. This may become one of the major disadvantages of this type of statistical method, since it only works effectively if several underlying assumptions or conditions are met. Thus, the analyst must know very well the data set that he presents and the capacity of the model. The set of naïve Bayesian classification (NB) - naïve models means that it is assumed that there is independence between terms, can be considered as simple models and can usually show quite good classification accuracy. NB assumes that each attribute is independent of the class it is in, in other words, each attribute node has as parent a class node, however there is no attribute node that can be related to another attribute node, [9].

Decision Trees (DT) are considered a technique with a lot of potential with respect to the Classification area. As the name implies, this method uses a tree structure to sort objects by dividing the dataset into smaller subsets. Each Leaf node then represents a decision that was made to divide the sets through the characteristics of the objects. DT may consider both categorical data and numerical data for decision. Two classifiers known for DT being these C4.5 and C5.0 may be mentioned. The C4.5 algorithm can handle both continuous and discrete attributes. This method handles continuous attributes by creating a boundary, then identifying the objects in the envelope and assigning them to these different lists. You can still deal with missing data by marking them with "?". C5.0 can be considered an improvement on C4.5, being faster, using memory more efficiently, and having lower error rates in cases of missing attributes. However, C5.0 creates more compact DTs relative to C4.5, [10]. The authors of this paper chose to apply C5.0. in this study.

The nearest neighbor method is considered very efficient and effective in several fields, namely pattern recognition and objects. Its simplicity is an advantage to be highlighted. The nearest neighbor rule recognizes the category of the new object through its "nearest neighbors" in which the class has already been previously recognized. T. M. Cover and P. E. Hart [11], first proposed k-nearest neighbor (kNN), in which the nearest neighbor would be calculated taking into account the value of k - defined as the total number of neighbors in each class. One of the disadvantages for the classification of a new sample, the distance for each of the previously used training samples will then be calculated. By identifying the k distances smaller and that better represent the class, they are considered as the exit label. A KNN method based on weights was also proposed, however, complexity and memory requirements remained the most critical points. One of the proposed solutions would be to reduce the size of the sample, removing data sets that would not add new information to the training database, [12].

Support Vector Machines (SVM) is said to be one of the most robust and one of the most successful within the Classification area. The SVM method is based on the idea of building an optimal hyper plan that contains the highest number of support vectors. In this way, it's possible to avoid local optimal solutions; a problem for neural network. This hyper plan must be selected from the set of hyper plans formed for the classification of the standards. It will be necessary to maximize the minimum distance between the chosen hyper plan and the nearest sample of each pattern. Regarding the type of patterns, there are distinct types, namely, linear, and non-linear. SVM can handle both types. Linear patterns will be easier to identify and hence easily sorted. Instead, nonlinear patterns are more complex to classify, so the need to be manipulated in order to achieve a simpler classification, [13]. One of the disadvantages SVM is, in case of large databases, that cannot distinguish the redundant information. Consequently, this will prejudice the learning process taking a lot of memory and reducing the speed of optimization, [11].

IV. CASE STUDY

The proposed methodology was applied to the case study presented in this section. It is a real distribution network located in Portugal and consists of 548 distributed producers and 20 310 consumers. This distribution network features 30 kV with only one 60/30 kV high voltage substation with a maximum capacity value of around 90 MVA. The distribution of the supra-referenced resources is done along 937 buses that make up the network.

The types of production in this case study are Wind, Biomass, Small hydro, co-generation (CHP), Photovoltaic, Fuel cell and Waste-to-energy (WtE). TABLE I. shows the characterization of each of the DGs, indicating the number of units in the network, the capacity of each in kWh and the operating price in m.u./kWh.

TABLE I. DISTRIBUTED GENERATION CHARACTERIZATION

Designation	Nº of units	Capacity (kW)	Price (m.u./kWh)
Wind	254	5 866.09	0.071
Co-generation	16	6 910.10	0.00106
Waste-to-energy	7	53.10	0.056
Photovoltaic	208	7 061.28	0.150
Biomass	25	2 826.58	0.086
Fuel cell	13	2 457.60	0.098
Small hydro	25	214.05	0.042
Total DG	548	25 388.79 kWh	

In this study, demand side management was able to rely on two programs: incentive-based (IDR) and price-based (RTP). In the first case, consumers are paid at a fixed price per kW of reduced load. In the second case, consumers change their consumption by responding to real-time electricity price changes.

The results obtained from the optimization of the present study, with 2592 scenarios, were later used to compare the different methods presented for the study. In this article, only the results obtained for the DG units will be used in the classification.

V. RESULTS

The classification methods studied will be tested and compared throughout this section. The goal is to group all resources considering the results from the first phase of the method, optimal schedule. Knowing the total amount of schedule power (Total P) for a selected scenario, being this the target in the classification, and all the other parameters presented later, it was possible to perform the proposed. In this way, for the classification, it was necessary to round to units the Total P values, in order to create more concise groups. The resulting data set was divided into seven groups, presented in TABLE II. This table also shows the number of elements in the database belonging to each group.

In a first phase, the percentage of the training data set size (T) and, consequently, the data set used for the forecast (P) was varying. In total, each method passed seven tests, and these are presented in TABLE III, where it is possible to verify the number of elements of each set. It should be noted that, in this article, only one training was performed for each model and for each method.

TABLE II. GROUPS FOR CLASSIFICATION

Group	Total P (kW)	Number of Elements
0	0	864
1	7 515	288
2	9 394	288
3	11 273	288
4	15 031	288
5	18 789	288
6	22 546	288

TABLE III. TESTS USED FOR CLASSIFICATION

Test	Training (#elements)	Test (#elements)
1: T70 / P30	1814	777
2: T50 / P50	1296	1296
3: T30 / P70	777	1814
4: T15 / P85	388	2203
5: T10 / P90	259	2332
6: T5 / P95	129	2462
7: T2 / P98	51	2540

In TABLE IV is presented the parameters variation used to represent the variables to be predicted. C_{reg} represents the cost of the regular supplier, C_{DG} represents the cost for the DGs, $P_{initial}$ represents the total load for each scenario, $P_{DR\_}^{MAX}$ represents the maximum reduction IDR, P_{Reduct} represents the maximum reduction RTP.

For α_{DG}, α_{DR} and α_{RTP} represent the maximum contribution of DG, IDR and RTP, respectively. With this, TABLE IV shows for each parameter the minimum, maximum and the step for the creation of different scenario. The first five parameters values are the percentage change.

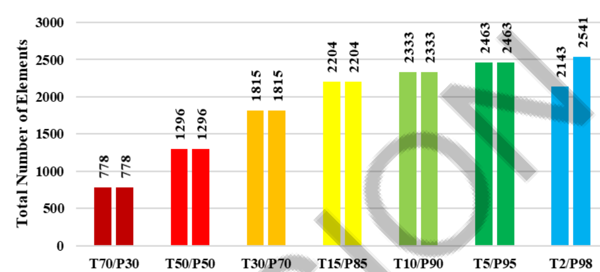
TABLE IV. PARAMETERS VARIATION USED FOR TESTS

Parameters	Min	Step	Max
C_{reg}	0.8	0.2	1.2
C_{DG}	1	0.2	1.2
$P_{initial}$	0.8	-0.2	1.2
P_{DR}^{MAX}	1	0.2	1.2
P_{Reduct}	1	0.2	1.2
α_{DG}	0	0.3	0.6
α_{DR}	0	0.15	0.3
α_{RTP}	0	0.05	0.15

Fig.3. represents the performance for NN method in predicting the variables under study. It is possible to verify, the method obtained very favorable results in most of the tests, sinning only in the last one, being understandable given the size of the training sample. The difference between the correct classification and the prediction was of about 398 elements, being these divided by the several groups. Group 6 would have been where this method the lower performance, hitting only 46% of the elements.

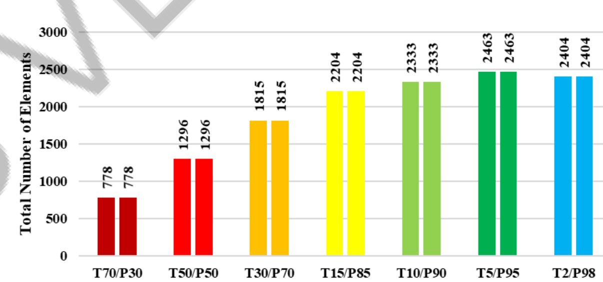
The same can be said regarding to the Naïve Bayesian classification, in Fig.4. The results obtained are consistent, reaching lower performance values only in the last test, such as NN. However, an improvement in results can be affirmed, since only 137 elements were not assigned correctly, as can be seen in Fig.4. The worst group in T2/T98 test was Group 2 accounting only 61% of the elements.

Through Fig.5. it is easily confirmed that the decision tree method can then overcome the two previously presented methods. This figure shows the performance of this method in correctly classifying elements, where it can be concluded that in the 6 of 7 tests performed, it obtained 100%. Still in this figure it is possible to verify that in the last test it did not assigned any of the elements to the group 5. Those elements were attributed to 4, the most similar in most of the parameters, where it allocated the 287 elements.



	T70/P30	T50/P50	T30/P70	T15/P85	T10/P90	T5/P95	T2/P98
Group 0	252	445	591	742	788	816	837
Group 1	96	150	210	244	256	276	205
Group 2	96	143	215	245	253	269	204
Group 3	70	135	192	246	253	270	283
Group 4	96	147	211	240	257	273	226
Group 5	96	139	193	254	269	278	104
Group 6	72	137	203	233	257	281	284

Fig. 3. Neural Networks comparison between test and prediction

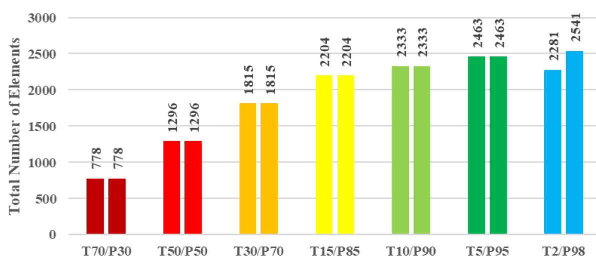


	T70/P30	T50/P50	T30/P70	T15/P85	T10/P90	T5/P95	T2/P98
Group 0	284	411	606	731	773	817	849
Group 1	77	160	202	251	258	279	271
Group 2	100	139	193	238	264	270	173
Group 3	77	157	196	242	265	278	278
Group 4	70	146	201	247	257	273	284
Group 5	86	149	200	248	260	277	270
Group 6	84	134	217	247	256	269	279

Fig. 4. Naïve Bayesian comparison between test and prediction

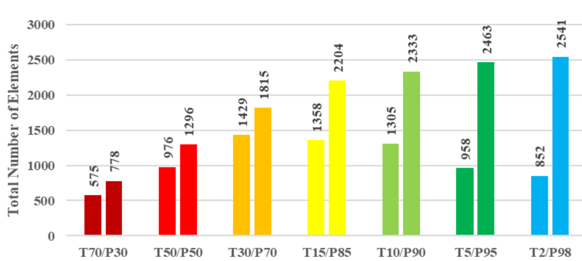
It is possible to affirm the possibility of not having elements belonging to group 5 in the training database, justifying the result obtained. Another possible conclusion of this test will be that the database doesn't have enough variation. Regarding KNN method, the number of k considered in this article was chosen taking into account the square root of the number of samples, resulting in a value of k = 51, [3].

In the same analysis of Fig. 5, the percentages of performance in this case are completely different, obtaining, in most cases, unsatisfactory results, as shown in Fig. 6, for this method. In addition to group 0, none of the other groups obtained the maximum value of performance. The worst test was the last one, T2 / P98, where it only correctly classified the elements of group 0. The remaining elements of the test database were also incorrectly attributed to this group.



	T70/P30	T50/P50	T30/P70	T15/P85	T10/P90	T5/P95	T2/P98
Group 0	260	425	599	736	777	819	857
Group 1	94	150	204	241	255	274	284
Group 2	92	150	209	240	257	275	285
Group 3	79	143	198	243	258	273	285
Group 4	86	142	202	246	262	275	287
Group 5	76	135	195	254	265	276	0
Group 6	91	151	208	244	259	271	283

Fig. 5. Decision Tree prediction performance tests



	T70/P30	T50/P50	T30/P70	T15/P85	T10/P90	T5/P95	T2/P98
Group 0	260	449	608	736	779	823	852
Group 1	37	75	134	93	23	15	0
Group 2	33	22	16	58	32	23	0
Group 3	44	69	128	72	16	4	0
Group 4	67	126	179	138	148	35	0
Group 5	58	103	172	67	156	39	0
Group 6	76	132	192	194	151	19	0

Fig. 6. K nearest neighbor prediction performance tests

Finally, the Support Vector Machine method. In the first four tests, was able to allocate all elements to the correct class even when the training database was less than 30% of the total. In Test 5, the difference between test and prediction, although not significant, generated the incorrect allocation of 11 elements. Regarding Test 6, the difference was more significant, about 387 elements, however, as you would expect, Test 7 was the one that performed worse because 1169 elements were not allocated correctly.

	T70/P30	T50/P50	T30/P70	T15/P85	T10/P90	T5/P95	T2/P98
Group 0	278	411	611	730	774	810	840
Group 1	97	153	205	246	258	268	255
Group 2	95	149	201	244	259	152	65
Group 3	67	137	205	247	254	185	95
Group 4	86	151	205	245	253	217	57
Group 5	76	150	194	256	261	209	19
Group 6	79	145	194	236	263	235	41

Fig. 7. Support Vector Machine comparison between test and prediction

Now, with all the tests for all the methods performed, you can then compare the performance between them in a detailed way. First, it should be noted that the size of the training sample is crucial, hence the T70 / P30 has obtained the best results. TABLE V and Fig.8. helped in this analysis. Fig.8. then presents the average percentage of performance of each method for each group to verify which one was the best for the database studied and which could be chosen one to ease the work of the aggregator. In this way, it is possible to see that DT has the best results. In all 7 tests performed, this method could easily predict all the elements' group. So, with this method will be possible to create rules for new and additional resources.

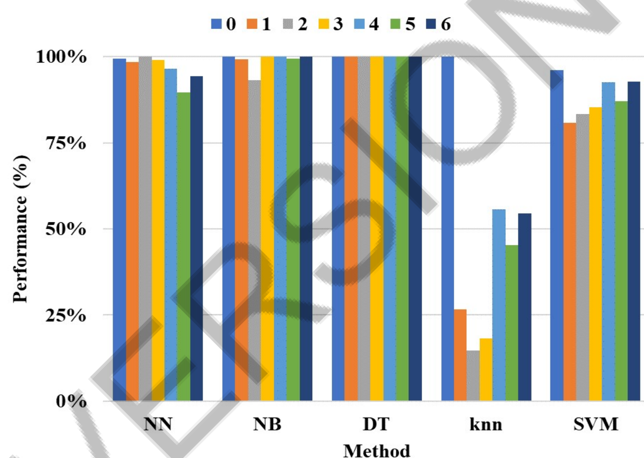


Fig. 8. Comparison between group performance for the studied methods

TABLE V. shows the average performance values of each method for each group. Group 0 did the best. In addition to the fact that there are a greater number of elements belonging to this group or simply the value of the parameters for this group is quite different, facilitating their classification.

By order of increasing performance, we can then start by talking about KNN. According to the literature, this method is extremely dependent on the value of K, and its results may vary with the value of K. It is also noted that when the samples are not evenly distributed, the determination of the K value in advance becomes complicated. Therefore, given that in this study, only training with k = 51 was performed, it may have influenced the results of this method, losing much of its performance, obtaining around 37.20% overall. Then SVM, with 88.09%, NN and NB obtained acceptable performance values, 96.74% and 98.80%, respectively. As shown in Fig.8. and TABLE IV, as previously mentioned, Decision Tree had a very satisfactory performance in all tests, accurately classifying all the elements. In this way, it will be a reliable method for the classification of new elements in the aggregation. The positive results of this study can be justified with the possibility of the database not having enough variation. Fig.9. presents the rules resulting from the study performed.

In this way, as already mentioned, each group defined a total amount of programmed power. Therefore, in case of knowledge of the input parameters, it will be possible to classify taking into account the rules presented in DT.

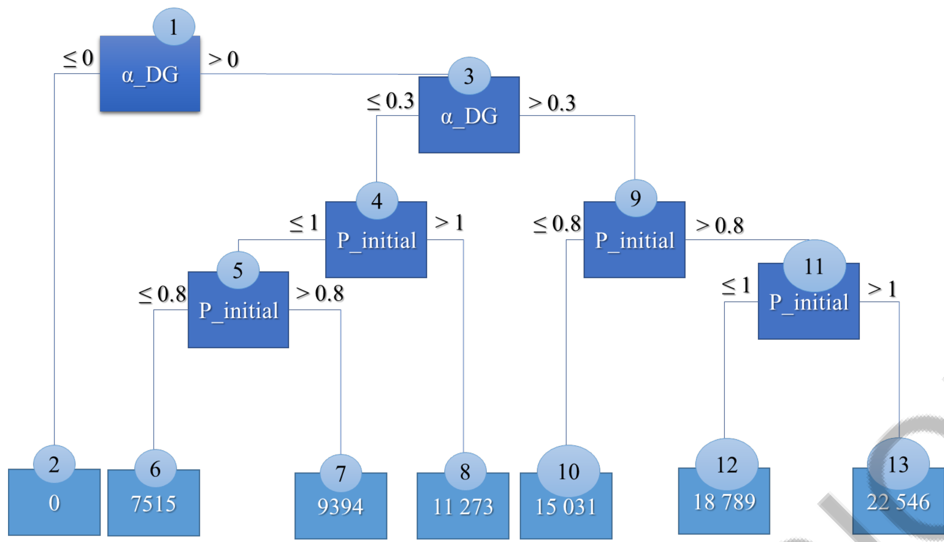


Fig. 9. Final Decision Tree

TABLE V. AVERAGE PERFORMANCE VALUES FOR EACH METHOD

Group	NN	NB	DT	KNN	SVM
0	99%	100%	100%	100%	96%
1	99%	99%	100%	27%	81%
2	100%	93%	100%	15%	83%
3	99%	100%	100%	18%	85%
4	97%	100%	100%	56%	93%
5	90%	99%	100%	45%	87%
6	94%	100%	100%	54%	93%

VI. CONCLUSIONS

The paper proposes a methodology to provide all the necessary instruments for a VPP. The first phase of this method was designed to comply with all the price and operational constraints, through an optimization schedule. After aggregate all small resources that collaborate with VPP, in phase two, there is a need to achieve a fair tariff for all groups. In the end, after a rescheduling, all members will be remunerated to continue collaboration. In this article, we focused on second phase, where classification can be introduced. Here were present five classification methods that were later tested and compared. The goal was to define a model that would aid the aggregator with the possibility of new members. With these rules it will be easier to aggregate a resource type from a result for the total amount of power. With this method it would be possible to avoid the need to perform a more complex optimization by following the rules presented.

REFERENCES

[1] S. Zhong and K. S. Tam, "Hierarchical Classification of Load Profiles Based on Their Characteristic Attributes in Frequency Domain," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2434–2441, 2015.

[2] P. Faria, Z. Vale, and J. Baptista, "Constrained consumption shifting management in the distributed energy resources scheduling considering demand response," *Energy Convers. Manag.*, vol. 93, pp. 309–320, Mar. 2015.

[3] A. B. Hassanat, M. A. Abbadi, G. A. Altarawneh, and A. A. Alhasanat, "Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach," *IJCSIS) Int. J. Comput. Sci. Inf. Secur.*, vol. 12, no. 8, 2014.

[4] Z. Vale, H. Morais, S. Ramos, J. Soares, and P. Faria, "Using data mining techniques to support DR programs definition in smart grids," in *IEEE Power and Energy Society General Meeting*, 2011, pp. 1–8.

[5] P. Faria, J. Spínola, and Z. Vale, "Aggregation and Remuneration of Electricity Consumers and Producers for the Definition of Demand-Response Programs," *IEEE Trans. Ind. Informatics*, vol. 12, no. 3, pp. 952–961, 2016.

[6] K. Holmstrom, A. Optimization, and M. G. Tom, "TOMLAB - An Environment for Solving Optimization Problems in MATLAB," *Matlab*, pp. 1–6, 2009.

[7] H. Langseth and T. D. Nielsen, "Classification using Hierarchical Naïve Bayes models," *Mach. Learn.*, vol. 63, no. 2, pp. 135–159, 2006.

[8] G. P. Zhang, "Neural networks for classification: a survey," *IEEE Trans. Syst. Man Cybern. Part C (Applications Rev.*, vol. 30, no. 4, pp. 451–462, 2000.

[9] N. Amor, S. Benferhat, and Z. Elouedi, "Naive bayes vs decision trees in intrusion detection systems," *ACM Symp. Appl. Comput.*, pp. 420–424, 2004.

[10] R. Pandya, J. Pandya, K. P. Dholakiya, and I. Amreli, "C5.0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning," *Int. J. Comput. Appl.*, vol. 117, no. 16, pp. 975–8887, 2015.

[11] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[12] N. Bhatia and C. Author, "Survey of Nearest Neighbor Techniques," *IJCSIS) Int. J. Comput. Sci. Inf. Secur.*, vol. 8, no. 2, pp. 302–305, 2010.

[13] A. Pradhan, "SUPPORT VECTOR MACHINE-A Survey," *Int. J. Emerg. Technol. Adv. Eng. Website www.ijetae.com*, vol. 2, no. 8, 2250.