Public Institutions' Investments with Data Mining Techniques

ADELA BÂRA
Academy of Economic Studies, Bucharest, ROMANIA
bara.adela@ie.ase.ro, ba_adic@yahoo.com
ION LUNGU
Academy of Economic Studies, Bucharest, ROMANIA
ion.lungu@ie.ase.ro
SIMONA VASILICA OPREA
University Polytechnic of Bucharest, ROMANIA
simona.oprea@transelectrica.ro

Abstract: - Developing Decision Support Systems in public institutions such as the National Power Grid Companies require applying very efficient methods in order to support the decisions. The decision support system in the National Power Grid Companies can integrate the energy prediction achieved by some data mining algorithms that help managers to fundament their investment decisions in order to justify the financial feasibility. This is done by estimating all the benefits and costs during the life cycle. In order to estimate the revenues, we need to know with certain accuracy the output of these power plants. Due to the fact that the wind speed significantly fluctuates even during a day at the same location, the wind power output is difficult to be forecasted by statistical methods. In this paper we apply the data mining techniques on the available measured weather data in order to predict the wind power output and determine the financial feasibility of investment.

Key-Words: - Data Mining (DM), Decision support systems (DSS), Wind Power Plant (WPP), wind power forecast, measured weather parameters

1 Introduction

Developing Decision Support Systems involve time, high-costs and human resources, efforts and the success of the system can be affected by many risks like: system design, data quality, and technology obsolescence. Especially in public institutions there can be another risk factor: the lack of funds to support these high costs required by a DSS system [5,6]. Under these circumstances the investment process in public institutions without any decision support system can be serious affected. These systems can assist the managers and executives to make decision regarding the benefit of investment, budgeting cash flows and financial planning, especially in the case of public funds, as we also presented in [7, 10].

As a result of the massive incentives given by the governmental authorities, many public institutions are concerned with their investment projects in wind power plants. To justify the feasibility of these projects, the cost – benefit analysis is an approach of making decisions of this kind. One the one part, the benefits include the revenues from wind power output plus the green certificates or feed-in tariff, according to each supporting scheme applied in the respective country. On the other part, the costs include the investment, maintenance, operational costs, etc.

As a result of the specific characteristics of wind generation, it is necessary in the system to be other generators for wind reserve generation, which should allow quick startup and loading up to maximum load and to have at any time sufficient primary resources (fuel or water).

This power reserve is required both in periods when wind generators are in operation (in this case is necessary an available of power that can be quickly started in the event that the wind speed comes out of the utility limits: less than 3.5 m / s to or greater than 30 m / s) and in decreasing, if they do not work (in this case is necessary an available power that can be quickly turned off if the wind speed falls in the utility limits mentioned above).

In order to determine the wind power output for the entire life cycle, we need to perform an accurate forecast of the electrical energy produced by the WPP.

In most cases there is an enormous amount of data, with correlation between attributes, but also with many information hiding within it that overwhelm traditional methods of data analysis. Data mining provides a way to get at the information hidden in data; it creates models to find hidden patterns in large, complex collections of data, patterns that sometimes are not taken into account by traditional statistical approaches because of the large number of attributes, the complexity of patterns, or the difficulty in performing the analysis [9]. Data mining techniques consist in a set of algorithms that are used to build the output model based on historical data and the system is trained to learn from these previous records and then to predict the future record values [11]. Data mining algorithms are applied on data sources stored in databases. Data from initial measured weather parameters is filtered, normalized, sampled, transformed and eventually used as input to data mining algorithms. Data is stored as tables in a database, so that data preparation can be performed using database facilities. Data mining models have to be built, tested, validated and deployed in the application domain environments. The data mining results may need to be post-processed as part of domain specific computations (for example, calculating estimated risks, expected utilities, and response probabilities) and then stored into permanent databases or data warehouses.

2 The necessity of wind forecasting

Hourly wind electricity prognosis continues to record large errors, which can reach several thousand of MW. In determining the predicted load degree of the installed WP in a electrical energy system, should consider the following aspects:

- Most wind generators have at terminals a voltage of 690 V. Each wind generator is equipped with a step-up transformer, and the links inside the WPP are performed, usually, at medium voltage (20 kV). From experience, it resulted that losses in a wind park are approximately of 2.5%.
- Backwater effect leads to the decreasing of power of a wind park with approx. 15%.
- Wind speed is a size that significantly differs even within the same emplacement.

From the above considerations, it has resulted that the power produced by a WPP is always lower than the installed power.

Setting the predicted load degree of the installed WPP in National Energy System (NES) can be determined through statistical or empirically methods. Statistics experience on the operation of WPP in Romania is insignificant. Therefore, estimating the load degree of the installed WP in NES is empirically done. The load degree of the installed WPP in NES will be determined by statistical methods when the installed power in wind parks in Romania will allow this.

From the maps showing the wind potential in Romania shows that the wind speed in Romania is of 7-8 m / s.

Based on the power characteristics of the generators in operation in other systems, it is known that wind generators operate when wind speed reaches a certain minimum value of 3 - 5 m / s and does not exceeds the maximum value of 25 - 30 m / s.

Figure 1 shows typical power characteristic considered for a wind generator:

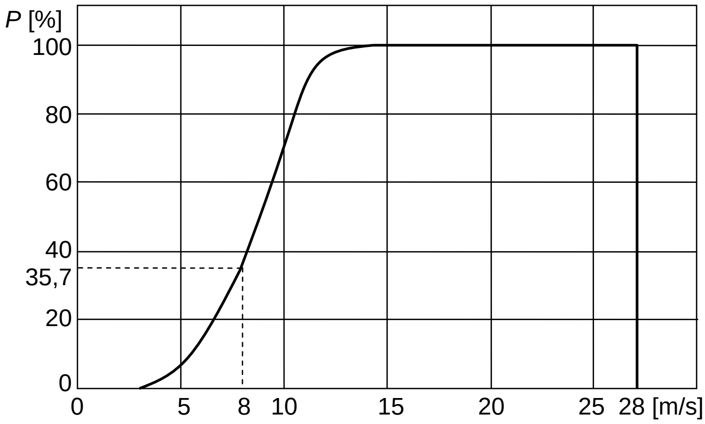

Fig. 1. Power characteristic of a wind generator

Maximum power that can be taken from the energy of air masses is [2]:

$$(1)\quad P_{WA,\max} = \frac{8}{27} \cdot A \cdot \rho \cdot v^3$$

where:

A – scanned area by the wind generator group paddles;

ρ– air density (ρ_{air} = 1,2 kg/m3);

v – wind speed.

It results that the available power of a WPP is proportional to the air density, wind speed to third power and with scanned area by the generator paddles.

If we consider P_n of a wind generator of 1,000 kW and wind speed at which P_n is reached is of 12 m / s, then the produced power by the wind generator, according to the formula above, when the wind speed is of 8 m / s is of 296 kW, which means 30% of P_n. It results that WPPs in Romania will work, on average, with 30% of installed power, as in (2).

$$(2)\quad k_{use} = 0,3$$

For rapid tertiary reserve (RTR) (that is the reserve generating units that can be mobilized in 15 minutes when the system is in need) determination at loading, are considered: hydro groups with barrier lakes and moulded reserve of the thermal groups.

To determine RTR at discharging are considered: hydro groups with barrier lakes and thermal groups that do not work with cogeneration, to minimum technical power. At hydro power plants, the minimum technical power is considered 0.

It is determined the scheduled RTR (RTRprog) as being equal to the RTR amount resulted as available in the scheduling process of the previous calendar year. According to the Commercial Code in force, the market participants have the obligation to transmit to the Transmission System Operator (TSO) for each common diagram system (h) active available power (Pdisp), minimum technical power (Pmin) and physical notification (PNF) for each dispatchable unit. Based on these values the RTR scheduled to increase (RTRprog+h) and RTR scheduled to decrease (RTRprog-h) are determined, by summation of these values, on each time interval, for each i dispatchable unit (n is the number of dispatchable units), as in (3) and (4):

$$(3)\quad RTR_{prog+h} = \sum_{i=1}^{n}\left(\mathrm{P}_{\mathrm{disp}_\mathrm{i}} - \mathrm{P}_{\mathrm{NF}_\mathrm{i}}\right)$$

$$(4)\quad RTR_{prog-h} = \sum_{i=1}^{n}\left(\mathrm{P}_{\mathrm{NF}_\mathrm{i}} - P_{\min_i}\right)$$

Scheduled RTR in NES is established for each h time interval (RTRprogh) as being equal to the minimum amount of RTR, at increasing or decreasing, observed in the respectively time interval in the previous calendar year, as in (5).

$$(5)\quad RTR_{progh} = \min\left(RTR_{prog+h}, RTR_{prog-h}\right)$$

The available RTR in NES is established as the maximum amount of RTR, which was available in the NES in the previous year 8322 hours (representing 95% of 8760 hours). It results that the generation park of NES ensures RTR necessary for the operation of WPP (which works on average with 30%) in 95% during the year, for 5% of the time being possible the production limitation in WPP on the not ensuring the reserve criterion [12].

The National Company - Romanian Power Grid Company Transelectrica S.A. annually establishes the maximum installed power in WPP, acceptable in terms of availability of rapid tertiary reserve (RTR) existent in NES, as in (6).

$$(6)\quad P_{i\max CEE} = \frac{RTR_{avail}}{k_{use}}$$

RTR is designed to ensure rapid recovery (maximum 15 minutes) of the secondary adjustment reserve, participating in adjusting the frequency and the balance to the direction value.

RTR must cover most power disconnectable almost simultaneously from NES, plus a maximum acceptable vagueness of the consumption prognosis (estimated value). It results that need of RTR in Romania, until the emergence of WP, is of 800 MW (to cover the trigger of a unit at Cernavoda NPP, at which it adds 100 MW for consumption prognosis impreciseness).

WPPs in Romania will be grouped on the criterion of common operating mode due to coincidence conditions of wind speed for components wind groups. Each such group will be considered as a equivalent generator group, which will work on average with 30% of installed power. In the event that equivalent generating groups with production higher than the largest group in the system will result, they will become the reference for establishing the necessary RTR in the system.

Since the volume of indented installed power in WP is much higher than available RTR in NES, the RTR value is the limitation criterion of the power in WP.

Maximum installed power value in WP for which technical permission can be given, is determined so that the maximum estimated power to be on / off in WP almost cvasi-simultaneously (PsimultWP) to be able to be compensated by loading / unloading of the available RTR (RTRavail) in NES, as in (7).

$$(7)\quad P_{simultCEE} = RTR_{avail}$$

RTR available in NES is periodically determined, because can suffer changes caused by the structure of the production park and market opportunities.

Since WPP is characterized by a low degree of utilization of installed power (kuse), the simultaneously produced power (estimated to be on / off) in all the WPPs is estimated as a percentage of installed power, which allows installation in NES of a power greater than the permissible one to be in operation at a time [12].

3 Data mining - a possible solution

Due to the problems described above, we need a method that allow us to forecast with a very good accuracy the wind speed in order to determine the wind power that will be produced. The characteristics of the wind speed make it difficult to determine the speed based only on statistical methods. So, we decided to study and test the data mining techniques and to see if we can achieve a better prediction.

Data mining algorithms can be divided into two major categories: *supervised* and *unsupervised*. Supervised algorithms are used to predict a value based on historical data. In order to lead the system to learn from previous data, these functions require the specification of a target which is either a binary attribute with true/false value or indicating yes/no decisions (wind turbine in/out of operation) or multi-class target indicating a preferred alternative (certain wind output levels). Unsupervised functions are used to find the intrinsic relations in data and do not use a target. For example, clustering algorithms can be used to find naturally occurring groups in data.

Data mining algorithms can also be classified as *predictive* or *descriptive*. Predictive data mining constructs one or more models that are used to predict outcomes for new data sets. Predictive data mining functions are classification and regression. Descriptive data mining describes a data set in a concise way and presents interesting characteristics of the data. Descriptive data mining functions are clustering, association models, and feature extraction.

Different algorithms serve different purposes; each algorithm has advantages and disadvantages. A given algorithm can be used to solve different kinds of problems as we presented also in [9]. For example, Naive Bayes is one algorithm used for predictive data mining that can be used to predict a yes/no target, but if you want to forecast a range of values the best algorithm is regression. In our analysis we will focused on supervised data mining techniques such as:

- ***Classification*** which consist in grouping items into discrete classes and predicting which class an item belongs to;
- ***Regression*** used to approximate and forecast continuous values;
- ***Attribute Importance*** that can help us to identify the attributes those are most important in predicting results.

To be effective in data mining, it is necessary to fulfill *four major steps*:

1. Problem definition - This is the step where the an abstract objective of the mining process is translating into a target set of variables and into a more tangible and useful data mining problem statement e.g. "Which meteorological conditions are most likely to produce the favorable wind speed?" To build a predictive model that predicts what factors are most likely to produce the target wind speed, we first must have data that describes the weather conditions that have produced the target wind speed in the past. Then we can begin to prepare the data for mining.

2. Data gathering and preparation - In this step, we take a closer look at our available data and determine what additional data we will need to address our problem. We will begin by working with a reasonable sample of the data, thousands or millions of cases recorded through a period of 2-3 years. We'll consider four major variables first: temperature, air pressure, nebulosity, wind speed.

3. Model building and evaluation - Once steps 1 and 2 have been properly completed; this step is where the data mining algorithms sift through the data to find patterns and to build predictive models. We will build several models and change mining parameters in an attempt to build the best or most useful models.

4. Knowledge deployment - Once we'll find a useful model that adequately models the data, we will distribute the new insights and predictions to the main wind energy application.

4 Algorithm implementation

Based on three-site measurement data recorded in five years in Romania, it has been proved that there is a good wind potential in those areas. We will use these data as input to forecast the wind power output. To forecast the wind power we need to determine the wind speed. The wind speed that has an impact on the wind power output can vary between 3.5 and 25 m/s. However, the wind velocity greatly fluctuates even within a day and month (see fig. 2 and fig. 3).

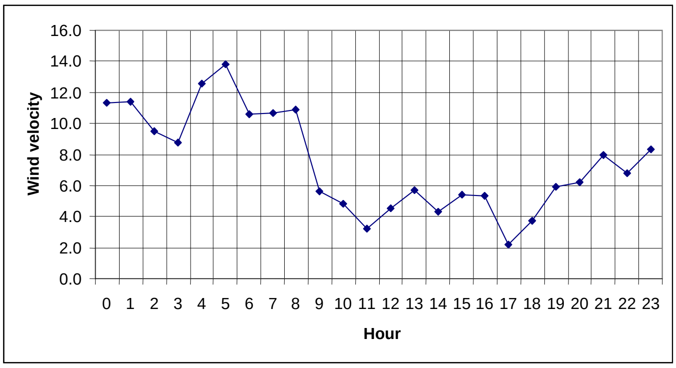

Fig. 2 Fluctuations of wind velocity within a day

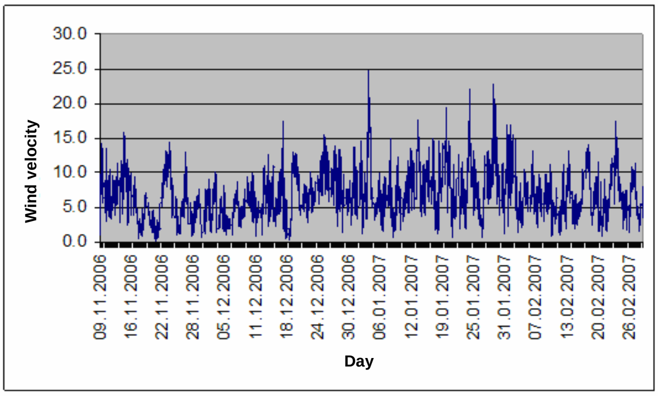

Fig. 3 Fluctuations of wind velocity over a few months

The wind power forecast will be performed by applying a couple of data mining techniques.

In our data mining process we will apply supervised learning techniques. The supervised-learning technique then sifts through data trying to find patterns and relationships between the independent attributes or predictors such as temperature and pressure and the dependent attribute – wind speed which will be the target attribute. Then the algorithm builds a model that best represents the functional relationships. The data is separated into three parts — one for building and training, table *wind_build* containing approximately 10000 records, one for testing, table *wind_test* with 4600 records and another one, split into three smaller parts: *wind_apply* table with 1100 records, *wind_apply1* table with 144 records and *wind_apply2* table with 144 records for applying and evaluating. The initial model is built using the first, larger sample of the data and then the model is tested on the second sample to assess the accuracy of the model's predictions. Because we already know the outcome — when the wind speed is favorable and when not— we can evaluate the model's accuracy and make decisions about the efficiency of the model. Finally, the model is applied and assessed on the third data set.

In order to obtain the smallest error we will apply several data mining techniques to test which is the best way to obtain the prediction model. *In every case the output of the data mining model is the wind power out.* In the following section we will describe the methods that is apply in every case.

Naïve Bayes algorithm finds patterns and relationships in the data by counting the number of times various conditions are observed. It then builds a data mining model to represent those patterns and relationships. NB is useful for building data mining models that can be used to classify and predict a variety of purposes, such as identifying whether weather conditions are likely to produce a favorable wind speed or not. Naïve Bayes algorithm makes predictions using Bayes' Theorem that assumes that each attribute is conditionally independent of the others. NB affords fast model building and scoring and can be used for both binary and multi-class classification problems. We applied the NB algorithm in order to predict the target attribute: 0 – no energy is produced, 1 – energy is produced. The accuracy of our predictions in the case of the three data sets is 93% with 9% errors for wind_apply table (104 errors), 1% errors for wind_apply1 table (1 error) and 9% errors for wind_apply2 table (13 errors) (fig 4).

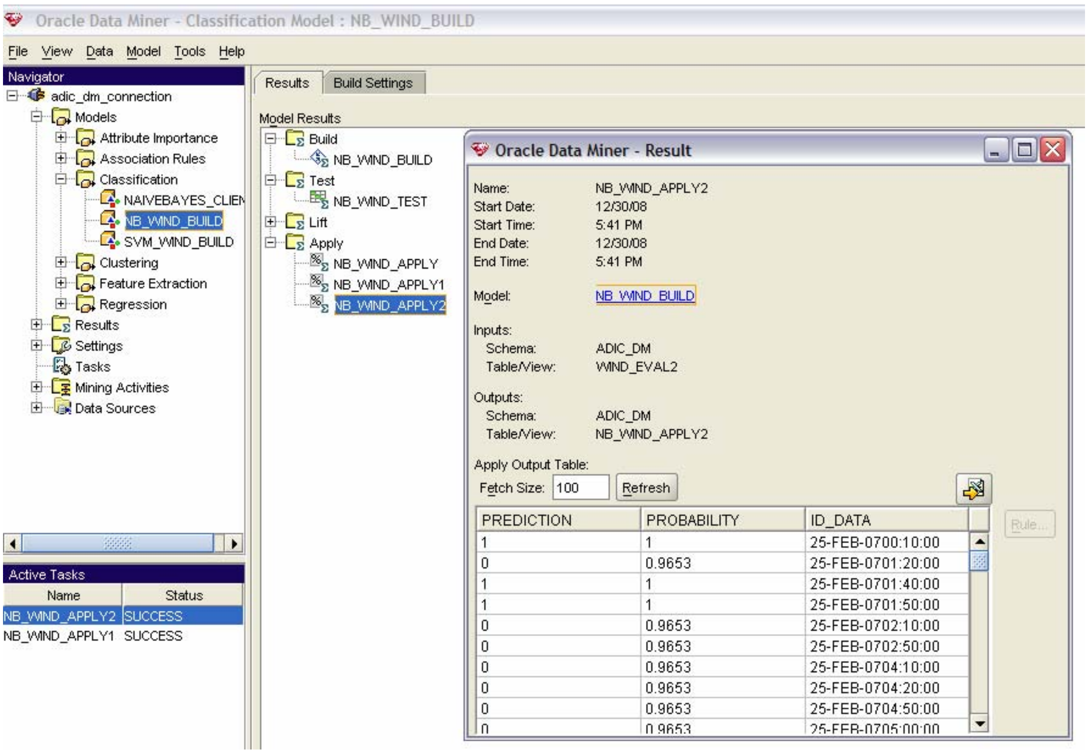

Fig. 4. The results of the Naive Bayes tests

In conclusion, the error rate is below 10%, which means that the model has a good accuracy and can be considered satisfactory.

Decision Tree method uses the ID3 algorithm to build the top-down decision nodes based on the training input. We applied the DT algorithm in addition to the NB to predict the 0/1 target value. The accuracy is 99%, better then the NB: only 0.6% errors for wind_apply table, no errors for wind_apply1 table and 0.7% errors for wind_apply2 table (fig. 5)

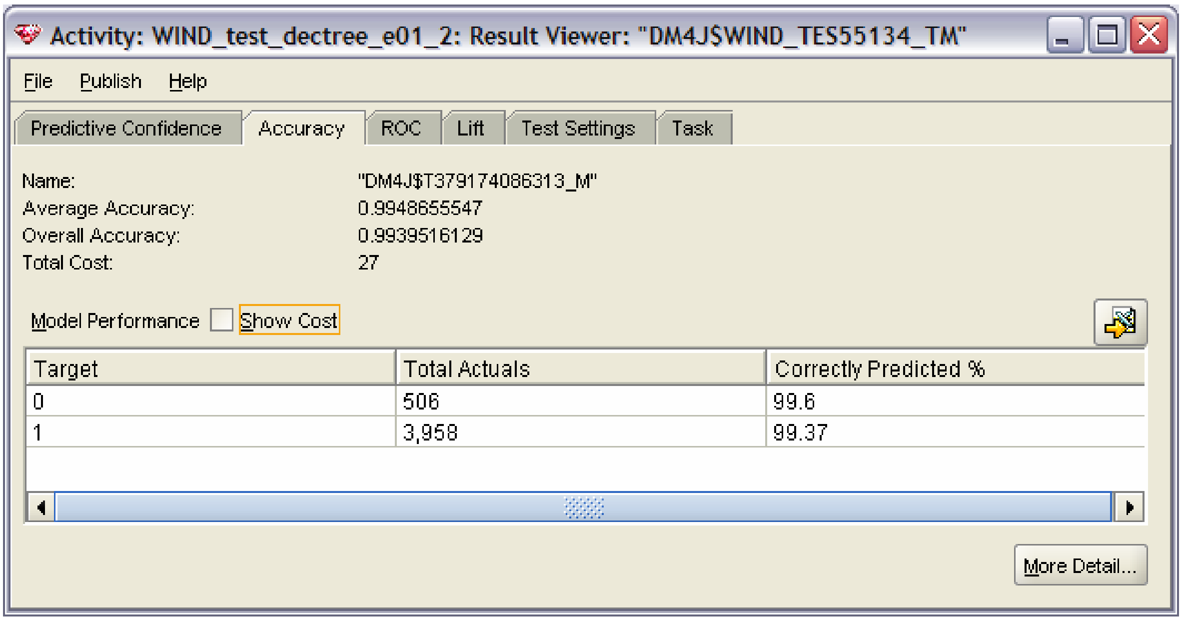

Fig. 5. The result of the Decision Tree tests

In conclusion, the error rate is below 1%, which means that the model has an excellent accuracy and can be considered very reliable.

But the output of these data mining techniques is the prediction of the 0/1 value of the target attribute: the wind speed will be favorable (the speed will be between 3.5 and 25 m/s) and the energy can be produced or not. In order to make a decision concerning the efficiency of the investment we need to predict the energy value. So we need to apply the DM algorithms that can predict discrete values. First, we applied the ***Attribute Importance*** algorithm to identify the attributes that have the greatest influence on a target attribute. In our case, knowing which attributes are most influential helps us to better understand and manage the wind forecast and simplify modeling activities. The output of the algorithm indicates that the wind speed, temperature, pressure and wind direction are the most influent factors.

Regression with Support Vector Machines algorithm supports prediction of the discrete values of the target attribute. This algorithm is useful in predicting the energy that will be produced based on the wind speed instead of predicting if the wind will blow or not. We use this algorithm to see if we ca predict the actual energy that can be produced based on the wind speed, the atmospheric pressure, temperature and wind direction. First, we build and test the regression to predict the energy as a discrete value.

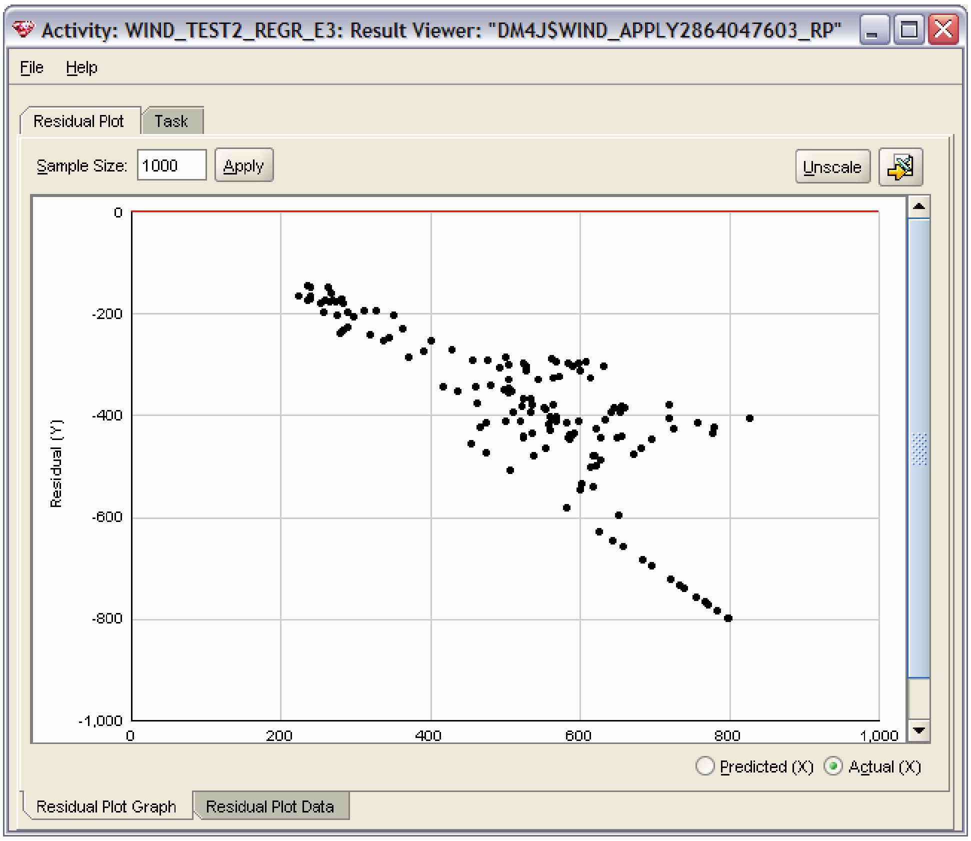

Fig 6: Residual plot data results of the regression with discrete values

The accuracy of the model is 57.68% and the residual plot shown in Fig 6 indicates that the errors are significant; the differences between the actual values and the predicted values can be of - 400 kW that means if the actual value is 1000 kW, the system predict around 600 kW.

So, we decide to build another regression model that can be applied to predict a value based on wind speed thresholds. We introduced another attribute called E_THRESHOLD that is calculated based on the variation of 0.5 m/s of the wind speed. For instance, we considered that there is no energy produced if the wind velocity is between 0 and 3.5 m/s, and there is a certain amount of energy accordingly to the power characteristic of the wind generator [1, 2].

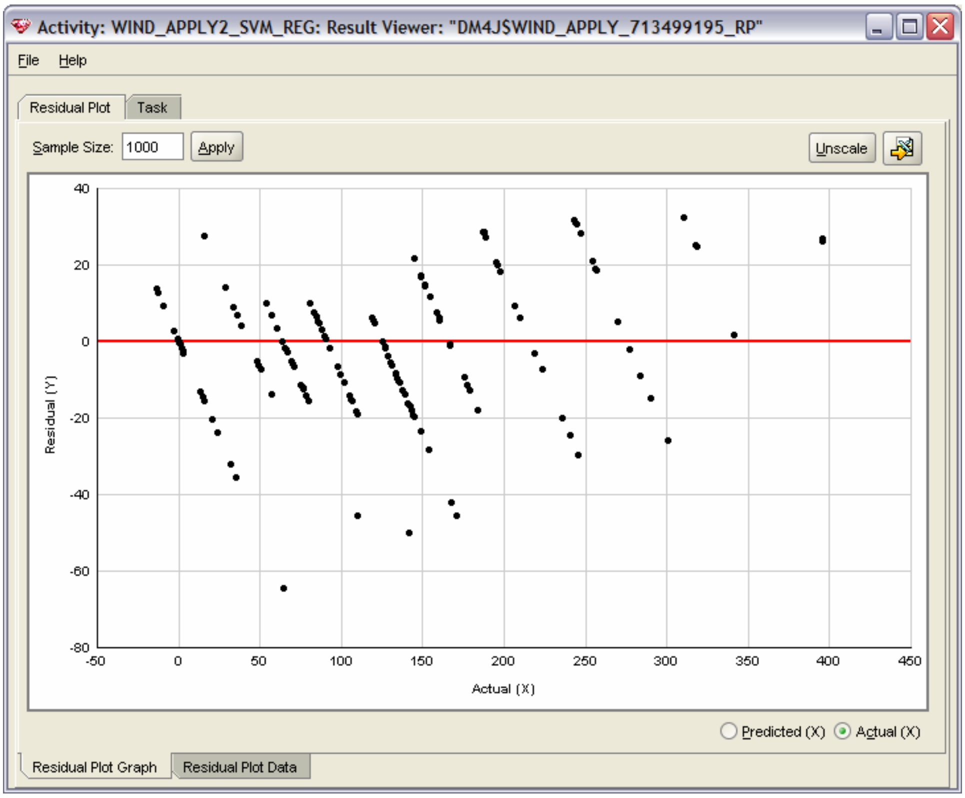

Fig 7: Residual plot data results of the regression with thresholds values

The accuracy of this model is 93.73% and the residual plot (Fig 7) shown that in the most cases the errors are between ± 20 kW, that means if the actual value is 1000 kW, the system provide a prediction between the range of 980 kW and 1020 kW.

In the following table we collected the results provided by the two regression models and, as you can observe, the second model is much better then the first one in terms of errors.

Table 1 Results of the regression models within 2 hours

No.	***Time***	***Energy (E)***	***Predicted E***	***E_ Threshold***	***Predicted E_ Threshold***
1	5:00	91.13	554.5	91	89.6
2	5:10	166.38	580.8	166	158.2
3	5:20	262.14	653.7	216	235.9
4	5:30	140.61	557.8	125	137.7
5	5:40	274.63	654.8	275	244.4
6	5:50	343	719.9	343	310.6
7	6:00	157.46	559	125	148.6
8	6:10	103.82	536.4	91	101.6
9	6:20	0	626.4	0	0
10	6:30	85.18	523.7	64	78.1
11	6:40	59.32	537.5	43	36
12	6:50	79.51	523.7	64	64
13	7:00	166.38	533.5	166	151.4

The results of the data mining process are shown in table 2:

Table 2 The accuracy of the models

No	***Algorithm***	***Accuracy***
1	Naive Bayes	93%
2	Decision Tree	99%
3	Regression with discrete values	57%
4	Regression with thresholds values	94%

The regression model with threshold values data is suitable for prediction so we'll use this method to calculate the amount of energy that can be produced and the economic efficiency of the power plant. We can assume that this model can be largely applied by public institutions for there cost benefit analysis.

5 Estimating the financial feasibility

There are many methods that can be applied in order to determine the financial feasibility of a project. The most important ones are: payback, net present value and internal rate of return.

The ***payback method*** gives the necessary time in order to get the initial investment back based on the predicted cash flow. It does not take into account the depreciation.

The ***net present value*** (NPV) is the updated cash flow, therefore taking into account the real value. For a profitable project NPV should be greater then zero.

The ***internal rate of return*** IRR is obtained when NPV = 0. These three methods are applied in details for the project that the public institution is concerned about.

Let's consider that the public institution would like to invest in a middle-scale wind power plant. Here is a short definition of its project:

- number of turbines: 20
- nominal power: 3 MW
- total installed capacity (Pi): 60 MW
- life expectancy: 10 years
- output expectancy: 50%*Pi
- required rate of return: 10%
- investment: 90 mil. Euro
- the revenue is given by the electricity price plus green certificate value
- yearly cash flow (as shown in table 3) is given by reducing the revenue by the operational costs.

The revenue has been partially estimated based on the prediction using the data mining algorithms. Considering the very good results obtained out of the wind power forecast (see table 1 and table 2), our estimation has a high level of accuracy.

Based on the payback method and the cash flow (see table 3), it is expected to entirely recover the initial investment after 7 years. However, this method can be easily applied, but it has several drawbacks. Therefore, it is a must to consider some other methods as well.

The main reason for applying NPV and IRR is that these two methods are quite relevant because they are calculated based on the updated cash flow.

By applying the following formula, each value has been updated in order to consider the depreciation:

$$Updated\_value_t = \frac{Initial\_value}{(1+Rate)^t} \quad (8)$$

where *i* is 0 for the investment year, 1 for the first year of operation and so on.

As a result, in this case for the required rate of return, NPV > 0, therefore the project is profitable.

Table 3 NPV and IRR calculation

Year	***Cash flow***	***0.08***	***0.09***	***0.10***	***0.11***	***0.12***	***0.13***
0	-90	-90	-90	-90	-90	-90	-90
1	30	27.8	27.5	27.3	27.0	26.8	26.5
2	20	17.1	16.8	16.5	16.2	15.9	15.7
3	20	15.9	15.4	15.0	14.6	14.2	13.9
4	-10	-7.4	-7.1	-6.8	-6.6	-6.4	-6.1
5	12	8.2	7.8	7.5	7.1	6.8	6.5
6	10	6.3	6.0	5.6	5.3	5.1	4.8
7	10	5.8	5.5	5.1	4.8	4.5	4.3
8	10	5.4	5.0	4.7	4.3	4.0	3.8
9	15	7.5	6.9	6.4	5.9	5.4	5.0
10	30	13.9	12.7	11.6	10.6	9.7	8.8
NPV		***10.6***	***6.5***	***2.8***	***-0.7***	***-3.9***	***-6.9***

Figure 8 indicates the way IRR is graphically determined. When NPV becomes negative the rate of return is greater than 10%. According to these results, we can conclude that the project is profitable and can be successfully implemented.

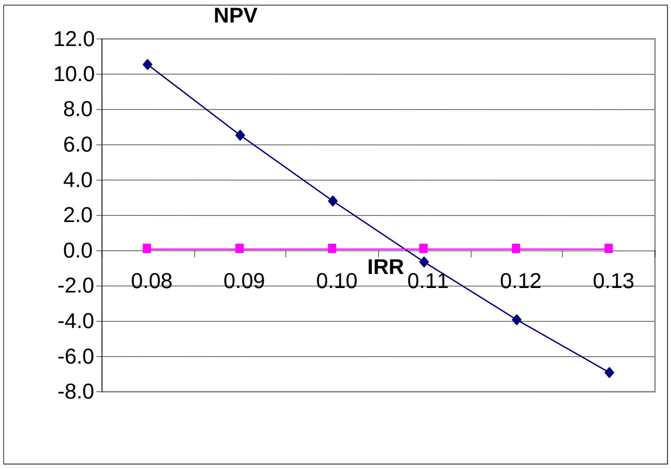

Figure 8 IRR representation

By applying the payback method, NPV and IRR, the most important three methods that show weather a project brings enough benefits or not, we have demonstrated that the project chosen by the public institution is feasible from the financial point of view.

6 Conclusion

The challenge of making decisions determines an orientation to a very good system design and conceptualization of an enterprise's true business needs. A solution for covering these risks is Business Intelligence technologies, like data warehouse, OLAP, data integration and data mining.

In our case, to determine the economic efficiency of investments is a very common problem. The solution has to be widely accepted, but business oriented. The data mining techniques can be a solution; the algorithms can be applied to complex business problems like predicting the wind power and, after that, to build a DSS system to integrate the data mining process and determine the economical efficiency. In this paper we propose such solution that can be used in public institutions.

As a result of the specific characteristics of wind generation, it is necessary in the system to be other generators for wind reserve generation, which should allow quick startup and loading up to maximum load and to have at any time sufficient primary resources (fuel or water).

This power reserve is required both in periods when wind generators are in operation (in this case is necessary an available of power that can be quickly started in the event that the wind speed comes out of the utility limits: less than 3.5 m / s to or greater than 30 m / s) and in decreasing, if they do not work (in this case is necessary an available power that can be quickly turned off if the wind speed falls in the utility limits mentioned above).

First, we applied the data mining process to predict the wind power and tested each of these algorithms. After we provided different types of output we decided that regression with thresholds values algorithm can be suitable for predicting the wind power. The data mining algorithm can be customize and modified depending on business needs.

After we have the wind power predictions, we can integrate these data mining algorithms in the DSS and determine the efficiency.

From the feasibility point of view, applying the payback, NPV and IRR methods, it is clear that the proposed project is profitable for any public institution interested in investing in WPP. The public institution can include these methods into its feasibility study and justify the project efficiency.

The efficiency of the project has been proved by the very good results out of the wind power forecast. Nevertheless, the financial efficiency

References:

[1] Ackerman T.- Wind Power, 2003

[2] Burton T., Sharpe D. - Wind Enegy Handbook, Wiley, 2001

[3] Bâra A, Lungu I., Velicanu M, Diaconița V, Botha I - Improving query performance in virtual data warehouses, WSEAS TRANSACTIONS ON INFORMATION SCIENCE AND APPLICATIONS, May 2008, ISSN: 1790-0832;

[4] Bâra A, Lungu I., Velicanu M, Diaconița V, Botha I – Extracting data from virtual data warehouses – a practical approach of how to improve query performance, Proceedings of the 7th WSEAS International Conference on ARTIFICIAL INTELLIGENCE, KNOWLEDGE ENGINEERING and DATA BASES (AIKED '08), Cambridge UK, 20-22 feb. 2008, ISBN: 978-960-6766-41-1, ISSN, 1790-5109;

[5] Bologa A.R, Bologa R., Bâra A. - Technology Vs Business Needs In Business Intelligence Projects, Proceedings of the International Conference on e-Business (ICE-B 2008), 26-29 July, Porto, Portugal, 2008, ISBN 978-989-8111-58-6;

[6] Diaconita V., Botha I., Bâra A., Lungu I., Velicanu M. - Two Integration Flavors in Public Institutions, WSEAS TRANSACTIONS ON INFORMATION SCIENCE AND APPLICATIONS, May 2008, ISSN: 1790-0832;

[7] Diaconita V., Botha I., Bâra A., Lungu I., Velicanu M. – Portal oriented integration in public institutions, Proceedings of the 7th WSEAS International Conference on ARTIFICIAL INTELLIGENCE, KNOWLEDGE ENGINEERING and DATA BASES (AIKED '08), Cambridge UK, 20-22 feb. 2008, ISBN: 978-960-6766-41-1, ISSN, 1790-5109;

[8] Lungu I., Velicanu M., Bâra A., Diaconita V., Botha I. – Practices for designing and improving data extraction in a virtual data warehouses project, Proceedings of ICCCC 2008, International conference on computers, communications and control, Baile Felix, Oradea, Romania, 15-17 May 2008, pag 369-375, published in International Journal of Computers, Communications and Control, vol 3, 2008, ISSN 1841-9836;

[9] Lungu I., Bâra A. - Executive Information Systems, Academy of Economic Studies Publisher, 2007, ISBN: 978-973-594-690-6;

[10] Lupu A. R., Bologa R., Lungu I., Bâra A. – The Impact of Organization Changes on Business Intelligence Projects, New Advances in Simulation, Modelling and Optimization - Proceedings of the 7th WSEAS International Conference on Simulation, Modelling and Optimization (SMO '07), Beijing, China, 15-17 Septembrie 2007, ISBN: 978-960-6766-05-3, ISSN: 1790-5117

[11] ORACLE Corporation – Oracle Database Concepts 10g, www.oracle.com;

[12] Operational procedure „CEE integration from the system flexibility point of view", COD: TEL - 07.38, 2008