Data-driven applications to identify sustainable investments pathways in energy management and efficiency

Themistoklis Koutsellis,
Decision Support Systems Laboratory,
National Technical University of Athens,
Zografou, Greece,
tkoutsellis@epu.ntua.gr

Vangelis Marinakis,
Decision Support Systems Laboratory,
National Technical University of Athens,
Zografou, Greece,
vmarinakis@epu.ntua.gr

Zoi Mylona,
HOLISTIC IKE,
Ag. Paraskevi, Greece,
zmylona@holisticsa.gr

Aija Zučika,
Latvian Environmental Investment Fund,
Rīga, Latvia,
aija.zucika@lvif.gov.lv

Haris Doukas,
Decision Support Systems Laboratory,
National Technical University of Athens,
Zografou, Greece,
h_doukas@epu.ntua.gr

***Abstract*—Energy Efficiency projects are often fragmented, of high transaction costs, and fall below the minimum value that many private financial institutions are willing to consider. The availability of comparable, anonymised historical data pooled from major market segments, structured along major project characteristics, can encourage greater investment flow in energy management and efficiency. The aim of this paper is to identify investment financing patterns in a pool of provided projects in Latvia and discover possible Grand Financing Plans (GFP) for future use. These GFPs could improve the procedure of decision making in energy sector in terms of the percentage of grand financing per project. The improvement of the process of grand financing can attract and mobilise private funding on such projects, providing investors/financiers (e.g., commercial/green investment banks, institutional/insurance funds, etc.) and project developers (public/local authorities, energy providers, ESCOs, construction companies, etc.) with data and tools to identify sustainable investment pathways and decrease the investment risk.**

Keywords—Data Mining, Energy Efficiency, Energy Management, Energy Planning, Decision Making, Investment Efficiency, Grand Financing

I. Introduction

Energy Efficiency (EE) projects are often fragmented, of high transaction costs, and fall below the minimum value that many private financial institutions are willing to consider. It is common ground that financial institutions and investment funds, national and regional authorities, as well as energy solution providers need advanced decision-making tools [1] to assess projects in energy management and efficiency [2-4].

The finance community is lacking a tested, evidence-based platform, providing decision makers with support regarding the impacts of various investment criteria, risk-aware assessment, and performance applied on a pool of EE investments. The capability offered by emerging near big data analytics [4] to integrate cross-domain financial and energy consumption is key to building the necessary market confidence in EE projects and making them an attractive investment asset class [5].

Equivalently, it is of paramount importance to label an investment, based on given information before the investment (e.g. region, average temperature, age of building, innovation project activities etc.), in order to apply a pertinent percentage of grand financing [6]. In this respect, we need to: (i) Provide a labelling scale; (ii) Apply a mapping between labels and percentages of grand financing, called Grand Financing Plan (GFP); (iii) Train a predictive model, which classifies future projects in energy management and efficiency, based on the provided labels; and (iv) Finally, decide which percentage of grand financing should be applied, based on the predicted label and the GFP.

In this paper, we demonstrate the first two steps in an application in Latvia, against a pool of data provided by a fund. This database contains information for both public and private investment building innovations.

In Section II, we present the used sample data and its characteristics, through various charts. In Section III, we investigate whether there is an efficient pattern in the applied GFP (for both private and public buildings). It should be noted that the existing GFP is based on directives from the European Commission (EC). Therefore, any improvement in the GFP is an added value for both the fund and the EC. In Section IV, we propose an efficient GFP, in terms of investment costs.

This new GFP is based on a set of five class labels we assign to the EE projects {i.e. poor, decent, good, very good, excellent}. The procedure of labelling is based on the distribution of the provided sample. The latter is considered large enough and sufficient for statistical inferences. Finally, we provide a table of cost savings after applying the proposed GFP. It turns out that there is cost reduction of up to 55%. In the last section, we discuss results and future steps.

II. Data analysis

A. Sample data

A pool of 550 projects in energy management and efficiency were analysed. To further use this sample, we proceeded to data cleaning and analysis. By doing so, we guaranteed that the sample is sufficient for statistical inferences, class labelling, and machine learning model training.

The final sample consists of 337 projects. All of them have a baseline year, i.e. a year before the start date of the renovation project with recorded energy consumptions per project. The baseline year is necessary to compare the energy consumption before and after the innovation projects. This deviation, equivalently the energy reduction or the deviation percentage, provides a brute force metric of the investment efficiency. After the end of each project, multiple reporting years were provided. For all necessary metrics, we derived the average values.

In our analysis, we exclusively focused on energy projects with only one building. Projects with multiple buildings spread across multiple Latvian counties are not sufficient for statistical inferences, since they feature different building utilisations, energy consumption patterns, weather conditions, etc. This kind of deviation per building does not provide a homogenised set per project, which makes the labelling per project almost impossible.

In Table I, we provide all input data per project we used as input for our analysis.

TABLE I. SAMPLE DATA PER PROJECT

Input data per project
Project No.
County
No. of Buildings
No. of floors per project
Building function
Project activities
Meteorology station
Average Temperature before the project
Total project costs
Grant financing by LEIF
GrandFinancing/TotalCost (%)
Building year
End date of the innovation project
Energy Consumption before project (MWh)
Average Energy Consumption after the project (MWh)
Energy reduction (MWh)
% of Energy Reduction
Investment Efficiency (kWh/€)

In the following charts (Figures 1-3) there are visualisations of the sample characteristics. Most of the sample projects are Educational Institutions, concentrated in the county of Riga, which is the capital of Latvia, and consist mostly of 2 to 3 floors.

B. *Financing details of sample data*

Figures 4 and 5 visualise the financial details of the sample projects. The support of the project cost probability density function (i.e., where ~90% of the samples lies on) is approximately between zero and 1,210,000 € and the majority of the projects' costs 220,000 € to 550,000 € (Figure 4). However, the percentage of grand financing seems to have an arbitrary pattern; the corresponding probability density function (pdf) is close to a multimodal pdf (Figure 5). This indicates that there is not a clear pattern in the percentage of grand financing per project (see Section III). As such, there is significant room for improvement, in terms of cost savings.

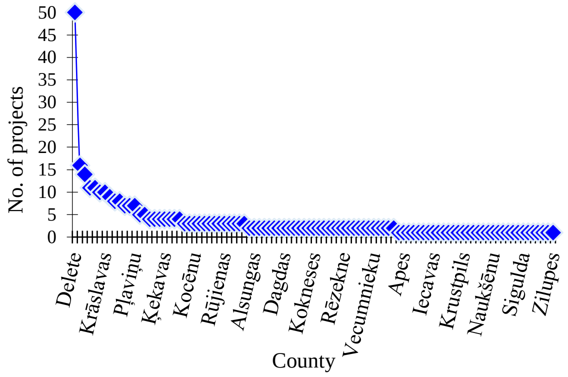

Fig. 1. Number of projects per county

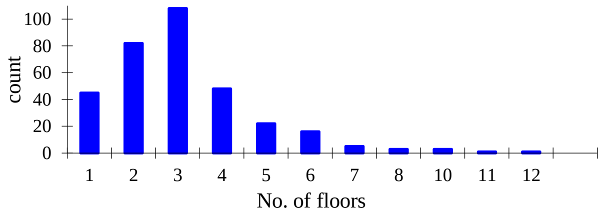

Fig. 2. Histogram of floors per project

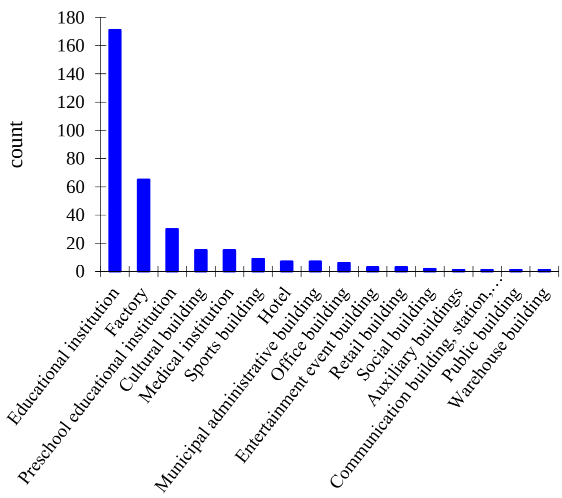

Fig. 3. Histogram of building function per project

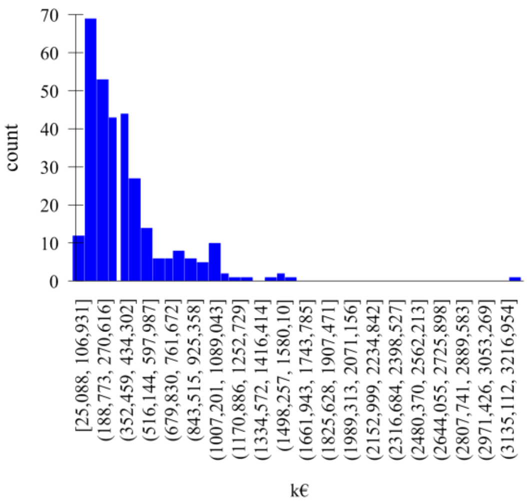

Fig. 4. Histogram of investemsnt cost

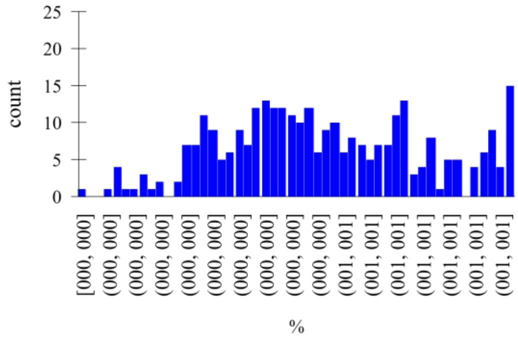

Fig. 5. Histogram of percentage of grand financing

III. Sample's Grand Financing Pattern

A key metric for our financing analysis is the kWh/€, which is based on the current practices of the fund. To identify whether there is a clear GFP, we plot the investment efficiency metric of kWh/€ against the percentage of grand financing (Figure 6 and 7). Figure 7 is a zoom-in of Figure 6, in the scale of (0, 2] kWh/€, i.e. where most of the projects lie on.

Figure 7 shows that there are multiple projects of low investment efficiency, which are grand financed with 85%, whereas other, more efficient projects, with only 35~40%.

We can also identify this discrepancy through the average percentage of grand financing per percentiles of investment efficiency (kWh/€). Based on discussions with the fund, the percentiles defined as 20%. If we split the kWh/€ axis into five equivalent intervals (with each interval containing the 20% of samples) we get the division depicted by the grey vertical lines in Figure 8. Bounds of these five intervals are shown in Table II (first column). If we assume that *the sample is large enough and sufficient for statistical inferences*, we can introduce a global labelling scale {poor, decent, good, very good, excellent} based on the division of the kWh/€ axis into 5 intervals of 20% percentiles (See Table II, second column).

The average percentage of grand financing is depicted in the third column of Table II as well as in Figure 8 (black solid line). We can see that the average line is still a better GFP than the blue dots (actual % of grand financing). However, the steps of the solid black line, from one percentile to another, lack a constant ascending slope: e.g., from 4^{th} to 5^{th} percentile — "very good" to "excellent" label or from (1.00, 1.40] kWh/€ to (1.40, 12.47] kWh/€ interval—the deviation of the percentage of grand financing is almost zero (~0.31%).

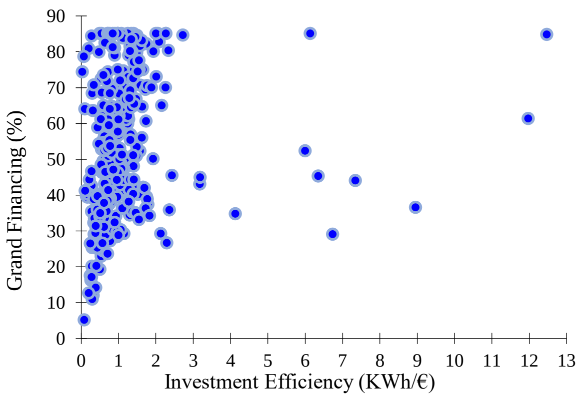

Fig. 6. Percentage of Grand Financing vs Energy Efficiency

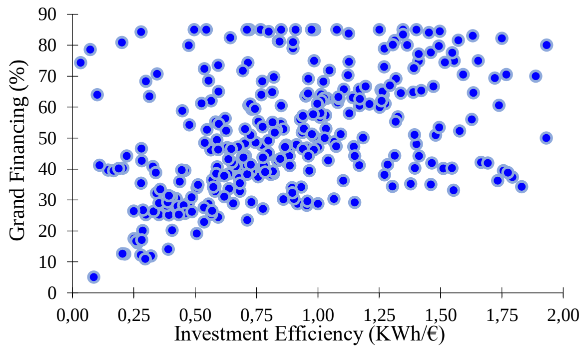

Fig. 7. Percentage of grand financing per project's investment efficiency

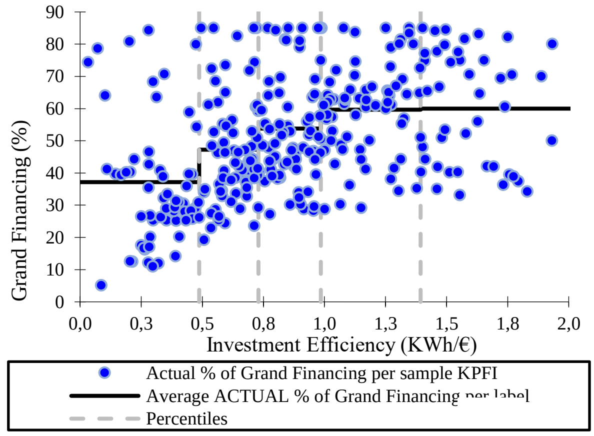

Fig. 8. Division of investment efficiency axis into 5 equivalent intervals

TABLE II. Actual Average Grand Financing per Label

20% percentiles: Intervals of kWh/€	Labels	Average % of actual grand financing
(0.00, 0.54]	Poor	37.17
(0.54, 0.76]	Decent	47.24
(0.76, 1.00]	Good	53.74
(1.00, 1.40]	Very good	59.67
(1.40, 12.47]	Excellent	59.98

In Section IV, we introduce a more efficient GFP than the average black solid line, which can further eliminate the financial cost for the same investment efficiency.

IV. Proposed Grand Financing Plan

Aim of the proposed GFP is to: (i) utilise the class labels (i.e., {poor, decent, good, very good, excellent}) and (ii) apply a rule with a constant step gradient per label. The chosen step gradient is 20%, e.g., from "poor" to "decent" investment efficiency, the increase of proposed % of grand financing is 20%. Similarly, from "decent" to "good", the increase is again 20%, and so forth. The new GFP is shown in Table III.

TABLE III. Proposed GFP

20% percentiles: Intervals of kWh/€	Labels	Average % of actual grand financing	Proposed % of grand financing
(0.00, 0.54]	Poor	37.17	**10**
(0.54, 0.76]	Decent	47.24	**30**
(0.76, 1.00]	Good	53.74	**50**
(1.00, 1.40]	Very good	59.67	**70**
(1.40, 12.47]	Excellent	59.98	**90**

In Figure 9, there is an illustration of the GFP concept. The black solid line represents the average % of actual grand financing; on the other hand, the red solid line represents the proposed GFP; the blue dots represent the actual status of the sample projects; the green diamonds are the corresponding position of blue dots, if the investment efficiency remains constant per project, but the grand financing follows the pattern of the proposed GFP (red solid line). The scattered blue dots are now transformed into a set of ordered green diamonds. The almost linear pattern of the proposed GFP guarantees that the less investment-efficient projects would get less % of grand financing (never as high as 85% as previously). Therefore, the proposed GFP is expected to yield an increase to the total savings.

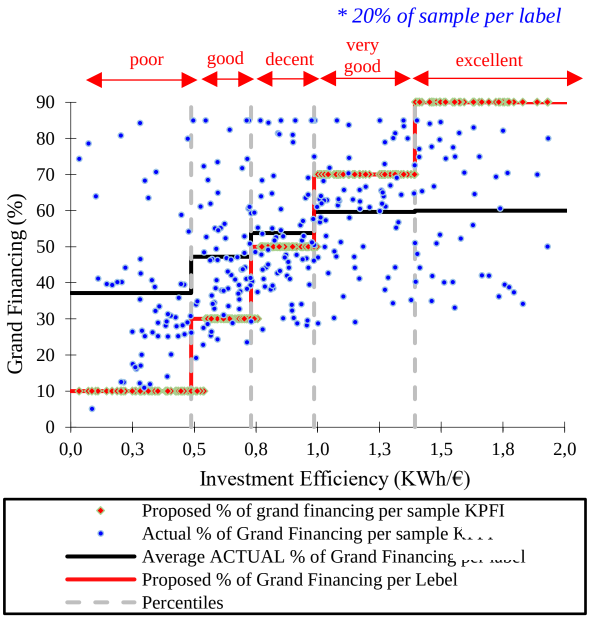

Fig. 9. Number of projects per county

Table IV depicts this increase. It provides the amount of € spent for the projects in the sample (first row). It also provides an estimate of total costs if the proposed GFP of Table III is applied in the same sample (second row of Table IV). The next two rows show the total savings. 55% of total savings is a significant amount of cost reduction and an indicator that the proposed GFP is efficient.

TABLE IV. COST SAVINGS AFTER APPLYING THE PROPOSED GFP

Total grand financing cost:	129 M€
Total proposed grand financing cost:	57 M€
Total savings:	71 M€
	55%

A precondition to apply the GFP, though, is to a priori know the class label of each project. To do so, a machine learning classifier should be trained from the given sample. For every new project, the classifier could predict its label and then the proposed GFP can be applied, yielding a reliable decision, in terms of the percentage of grand financing.

V. CONCLUSION & FUTURE WORKS

In this paper, we presented a new Grand Financing Plan (GFP) and a labelling classification, which can be useful for any financial institution in the energy sector. The proposed procedure yields significant investment savings (of up to 55%). In conclusion, this achievement could attract and mobilise private funding on such projects and help Member States to support fostering of green energy markets. In future work, we will explore various machine learning algorithms [7] to support the projects' labelling classification. We will also contact similar analysis for solely private sector buildings.

ACKNOWLEDGMENT

The work presented is based on research conducted within the framework of the project "BD4NRG: Big Data for Next Generation Energy", of the European Union's Horizon 2020 research and innovation programme under grant agreement no. 872613 (https://www.bd4nrg.eu/). The content of the paper is the sole responsibility of its authors and does not necessary reflect the views of the EC.

REFERENCES

[1] Nikas, A., Ntanos, E., & Doukas, H. (2019). A semi-quantitative modelling application for assessing energy efficiency strategies. Applied Soft Computing, 76, 140-155.

[2] Doukas, H. (2018). On the appraisal of "Triple-A" energy efficiency investments. Energy Sources, Part B: Economics, Planning, and Policy, 13(7), 320-327.

[3] Doukas, H., & Marinakis, V. (2020). Energy poverty alleviation: effective policies, best practices and innovative schemes. Energy Sources, Part B: Economics, Planning, and Policy, 15(2), 45-48.

[4] Forouli, A., Gkonis, N., Nikas, A., Siskos, E., Doukas, H., & Tourkolias, C. (2019). Energy efficiency promotion in Greece in light of risk: Evaluating policies as portfolio assets. *Energy*, *170*, 818-831.

[5] Marinakis V., Doukas H., Tsapelas J., Mouzakitis S., Sicilia Á., Madrazo L., Sgouridis S. (2020). From big data to smart energy services: An application for intelligent energy management. *Future Generation Computer Systems*, 110, 572-586.

[6] Marinakis, V. (2020). Big Data for Energy Management and Energy-Efficient Buildings. *Energies*, 13, 1555.

[7] Doukas, H., Xidonas, P., & Mastromichalakis, N. (2021). How Successful are Energy Efficiency Investments? A Comparative Analysis for Classification & Performance Prediction. *Computational Economics*, 1-20.

[8] Zekić-Sušac, M., Mitrović, S., & Has, A. (2021). Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities. International journal of information management, 58, 102074.