ISSN 1507-3858
e-ISSN 2450-0003

Maciej Wach
KGHM Polska Miedź SA
e-mail: maciej.wach@ue.wroc.pl
ORCID: 0000-0002-5219-6266

THE APPLICATION OF PREDICTIVE ANALYSIS IN THE MANAGEMENT OF INVESTMENT PROJECT PORTFOLIOS

ZASTOSOWANIE ANALIZY PREDYKCYJNEJ W ZARZĄDZANIU PORTFELAMI PROJEKTÓW INWESTYCYJNYCH

DOI: 10.15611/ie.2021.4.05
JEL Classification: O22, D81.

Quote as: Wach, M. (2021). The application of predictive analysis in the management of investment project portfolios. *Business Informatics*, (4).

Abstract: The purpose of this paper is to indicate the possibilities of applying predictive analytics in the area of investment project portfolio management. In the article, deductive reasoning, critical analysis of the literature and the analysis of selected decision cases were used as the research method. The author presents the process of the investment project portfolio management. The decision-making problems are highlighted, along with the consequences that poor portfolio management may have on the enterprise. Predictive analytics is characterised as a tool for aiding decision-makers together with basic requirements for its application in any organization. As a result, a framework is presented, which uses new approach, where project is considered as a parametrised object that follows patterns created by past cases. Predictive algorithms are suggested for specific decision-making problems met by portfolio managers. The author also discusses the limitations of the proposed solutions.

Keywords: predictive analysis, investment project portfolio management.

Streszczenie: Celem niniejszego artykułu jest wskazanie możliwości zastosowania analityki predykcyjnej w obszarze zarządzania portfelami projektów inwestycyjnych. W pracy zastosowano wnioskowanie dedukcyjne, a jako metodę badawczą wykorzystano krytyczną analizę literatury oraz analizę wybranych przypadków decyzyjnych. Autor przedstawia proces zarządzania portfelami projektów inwestycyjnych z podkreśleniem problemów decyzyjnych oraz wpływu złego zarządzania portfelami na przedsiębiorstwa. Analiza predykcyjna scharakteryzowana została jako narzędzie wspomagające decydentów wraz z wymaganiami co do jej zastosowania w dowolnej organizacji. W efekcie stworzono modelowe podejście zastosowania tej analizy, w którym projekt traktowany jest jako sparametryzowany obiekt, który podąża za wzorcami stworzonymi przez wcześniej realizowane projekty. Określonym problemom decyzyjnym przypisano sugerowane alorytmy predykcyjne. Dodatkowo omówiono najważniejsze ograniczenia proponowanych rozwiązań.

Słowa kluczowe: analiza predykcyjna, zarządzanie portfelami projektów inwestycyjnych.

1. Introduction

Creating investment project portfolios that are optimally aligned with corporate strategy and contribute to achieving the best possible results expected by the organization, has been identified as one of the biggest challenges for companies managing their investments in project manner (Crispim and Siqueira, 2014). One of the basic problems in the decision-making processes of managing investment projects portfolios is the assessment of the expected duration of each project and budget distribution needed for its implementation. According to the Project Management Institute (2018), in the last decade, the number of global projects completed on time ranged from 50% to 55%, and slightly more (55%-60%), were implemented within the planned budget. Extensive project delays or miscalculation of the budgets is a multidimensional problem for the project portfolio owner (manager) and the whole company. Opportunities are lost due to the unused financing, the project's effects are postponed what decreases its NPV and ROI, external liabilities grow, specialists are tied up to ineffective tasks, resources change value over time and finally the attitude towards the project may change in the organization, resulting in the abandonment of the project and the loss of the value of the already delivered project products.

In recent years, predictive analysis has become a significant asset in decision-making processes, giving managers representing various business areas, an insight into so far hidden dependencies, relations and patterns in data sets. With the use of predictive analysis, it is possible to support the strategic decisions made by portfolio managers by detecting projects endangered with extensive delays and with inappropriate budget execution. In contrast to the existing methods that lean towards the risk analysis of single tasks, or thorough projects assessment, the framework for portfolio predictive analysis presented in this article is based only on the projects' parametrisation and reporting defined by the project management methodology applied in the organization.

2. Project portfolio management

Early definitions of project portfolio simply characterized it as a group of projects managed under common sponsorship. Currently, its relation to the company's strategy is strongly emphasised when describing project portfolios. Assuming this approach, project portfolio is best defined by the Project Management Institute (PMI) standard as "a component collection of programmes, projects or operations managed as a group to achieve strategic objectives" (Project Management Institute, 2019). What distinguishes project portfolio from a programme, which is also a set of projects launched to pursue a common goal, is the number and variability of its objectives, as well as the level and complexity of management. The programmes once established in terms of objective and project composition are less likely to be redefined, and mainly focus on the optimal execution of component projects. Project portfolio management (PPM) is responsible for translating the strategic goals of the company into programmes or single projects. Thus project portfolios may serve multiple strategic purposes, and have to react to changes made in company strategy, and align to it by making constant changes in their structure. Cooper defined six portfolio performance measures (Cooper, Edgett, and Kleinschmidt, 1992) that reflect the main goals of portfolio management as: maintaining the number of projects adequate to the resources available; gridlock avoidance that allows projects to be completed on time; inclusion of high-value projects (value maximisation); balancing the portfolio in terms of the risk and time perspective; keeping the alignment with business strategy; planning a spending's breakdown that mirrors strategic priorities.

2.1. Project portfolio management processes and decision-making

Although PPM occurs as a dynamic, continuous process, it is commonly defined as a partially-cyclical sequence of stages, each of them closely related with specific decision-making problems. The Project Management Institute divides this process into five stages (Oltmann, 2008):

- **Goal clarification** – establishing a strategic goal (or goals) that the portfolio is meant to follow as well as the project valuation method. These are required to properly execute the next steps resulting in the project selection.
- **Capture and research** – projects that will form future portfolios have to be gathered, as well as the data required to perform their evaluation. Project ideas may be sourced from the company's management, customer requests or project managers and other employees. Ongoing projects are also included in this phase and usually revised.
- **Project selection** – based on the established evaluation criteria, projects are selected to form a portfolio that maximises total project value and are balanced in terms of risk and length.
- **Validation and initiation** – before projects that make up the portfolio are initiated, the interdependencies, budget constraints, resources availability and organizational capabilities to perform these projects, have to be checked. If any of those are unrealistic, the portfolio has to be reprioritised.
- **Management and monitoring** – the most complex phase of the portfolio management process that includes daily basis activities performed by portfolio

managers supported by Project Management Offices (PMOs): portfolio performance monitoring, resource assignment, scope and budget changes, managerial reporting, and portfolio composition changes.

In considering the described phases, the first phase is incidental and its occurrence relies on changes in corporate strategy. Phases 2 to 4 often appear cyclically, as companies review their portfolios to achieve periodical performance indicators. The 5th phase is an ongoing process, that in a broad perspective, requires the highest level of human work and is characterised by a high frequency and variety of decision-making situations.

Baptestone and Rabechini (2019) evaluated the relation between project portfolio management and decision-making. The latter is positively influenced by the former if the portfolio is strongly aligned with the organization's strategy, generates financial returns and has a positive impact on the business. The described phases of the management process are then closely followed by the organizations. Baptestone and Rabechini (2019) also indicated that this theoretically perfect workflow is disturbed by the influence of intuitive and political decisions. These are described in detail by Elbanna (2006), as those made relying on judgment, experience and 'gut-feeling' (intuitive decisions) or from a strongly empowered position (political decision).

3. Predictive analysis

The role of predictive analysis has acquired significant importance in supporting decision-makers over the past two decades. Predictive analytics utilises advanced mathematical formulas, statistical algorithms, as well as IT tools and services to identify dependencies, relationships and patterns in data sets or reduce their complexity (Dinov, 2018). This outcome is achieved by the creation of predictive models that are able to determine, on the basis of historical data, the probability of future events in alternative situations represented by the given input data sets (Williams, 2011).

3.1. Machine learning

Machine learning, a subset of a wider field – the artificial intelligence (AI) – is the most commonly used set of techniques in the process of creating predictive models. It is defined as a computer science responsible for developing algorithms and techniques for automated problem-solving difficult to program using conventional programming methods (Rebala, Ravi, and Churiwala, 2019). Other definitions of machine learning describe it as a set of tools that use algorithms to identify hidden patterns in data, giving them a previously unidentified meaning (El Morr and Ali-Hassan, 2019). The idea of machine learning assumes that natural, human learning methods should not be taken as the only possible cognitive pathways and its main goal is to explore alternative learning mechanisms (Klix, 1985). The essence of machine learning techniques is to supply a selected model with a set of training data, on the basis of which its parameters are being adjusted by the selected algorithm. The process of adjusting model parameters is called ‘training’. The created model is expected to respond to future input data based on the patterns detected in the training dataset. The division of machine-learning techniques into two main categories of supervised learning and unsupervised learning, is determined by the differences in the structure and utilisation of the training data sets.

The unsupervised learning category shows the lack of expected model responses in the training data set, which results in delegating the responsibility for its creation to the automatic learning process. The main goal of unsupervised learning is to detect natural patterns in the data set (Etaati, 2019) without any external suggestions. The algorithm identifies clusters or groups of similar objects to which it fits new data sets in the future (the allocation is the response of a model). There are two main techniques used in unsupervised learning, namely clustering and dimensionality reduction. The main task of clustering is to identify hidden patterns and divide the data sets into clusters, which is especially employed in image recognition and market analysis (El Morr and Ali-Hassan, 2019), while dimensionality reduction is a technique that simplifies the dataset by reduction of the number of input variables in order to better fit a predictive model.

Table 1. Classification of machine learning algorithms

Learning type	Technique	Algorithm
1	2	3
Supervised learning	Classification	Artificial Neural Networks
		Bayesian Networks
		Classification Trees
		Fuzzy classification
		K-nearest neighbour
		Linear Discriminant Analysis (LDA)
		Logistic Regression
		Random Forests
		Support Vector Machine (SVM)
	Regression	Artificial Neural Networks
		Bayesian Networks
		Decision Trees
		Fuzzy classification
		Generalized Linear Model
		K-nearest neighbour
		Linear Regression
		Multiple linear regression
		Support Vector Machine (SVM)

Table 1, cont.

1	2	3
Unsupervised learning	Clustering	Artificial Neural Networks
		Genetic algorithm
		Hidden Markov Model
		K-Means Clustering
		Self-Organizing Map
	Dimensionality reduction	Principal Component Analysis
		Linear Discriminant Analysis
		Multidimensional Statistics
		Random Projection

Source: (El-Morr and Ali-Hassan, 2019).

In the case of supervised learning, the training data set is complemented with correct (actually observed) responses. The model parameters are adjusted to achieve the best fit between the estimated and observed responses based on the so-called testing dataset, detached from the initial training dataset and not used in the model parametrisation. For each new set of input data, the learned model tries to generate a response known from the pool of training data. Supervised learning methods are divided on the basis of the nature of the model response, which can be continuous (regression techniques) or categorical (classification techniques). The classification of machine-learning techniques and of the corresponding algorithms is presented in Table 1. Some of the algorithms can be applied in a different manner, so that they can serve different techniques or even learning types.

3.2. Requirements for implementation

The implementation of predictive analysis in an organization, apart from knowledge about statistical algorithms and machine learning, requires the fulfillment of three additional conditions (McCarthy, McCarthy, Ceccucci, and Halawi, 2019):

- **Advanced business knowledge –** knowledge and experience that allows an organization to successfully embed analytics in a given business area by correctly defining the purpose of the analytics, understanding the relations between the variables used in it, as well as possessing the ability to capitalise its results.
- **Complex business problem** – the problem solved by the analyst must be complex and significant enough for the organization to make the implementation of the analytics profitable and not replaceable by simpler tools.
- **Access to Big Data** – possession of extensive databases and warehouses, regularly supplied with project data, is required in order to achieve a sufficient level of reliability and accuracy of predictive analytics.

Bandara (Bandara, Behnaz, Rabhi, and Demirors, 2019) introduced similar requirement classification, defining them as knowledge, strategic and data requirements respectively. This division is based on the five categories introduced by Taylor (2015), which further decomposes strategic requirements into decision requirements, performance and business context. The taxonomy proposed by Bandara (Bandara et al., 2019) and Taylor (2015) is more detailed and takes into account initial decisions that are obligatory to properly design the analytics process. Knowledge requirements should not only concern business, but also analytical knowledge in order to properly select models and identify independent variables. Similarly, strategic requirements should not only properly define a business problem, but also define expected prediction accuracy in reference to performance measures, the maximising of which is ultimately the main purpose of introducing predictive analytics.

4. Application of predictive analytics in project portfolio management

Although already widely introduced in areas like banking, healthcare, insurance and industry, there are limited applications of predictive analytics in project portfolio management. The approach that has recently drawn increasing attention is the Project Predictive Analysis developed by Deloitte in partnership with the Helmsman Institute in Australia (Fauser, Schmidthuysen, and Scheffold, 2015). It is based on a profound project assessment that takes 4-6 weeks, and offers the possibility to predict potential risks at any stage of the project and identify areas where projects are most vulnerable to errors. During the five-step assessment process, which includes interviews with stakeholders, document review, risk and complexity assessment, the project data are compared against the database of over 2,000 similarly evaluated projects, which allows to predict potential risk throughout the whole project lifecycle.

Another well-known approach is the Monte Carlo Method. This simulation method created the first half of the 20th century, can be applied to analysing investment projects through estimating borderline values for each scheduled task in terms of duration and required budget. Probability distribution is then applied to each task in order to generate random project outcomes in hundreds or even thousands of simulations. The results are then analysed using histograms, confidence intervals and other statistical indicators resulting from the simulation (Platon and Constantinescu, 2014).

Both the described approaches focus on risk assessment and require a lot of input from project stakeholders. Moreover, in the case of the Monte Carlo Method, portfolio data are not used and all external dependencies must be taken into account by those who provide the estimates to elementary tasks.

4.1. Framework for predictive project portfolio analysis

The proposed framework for predictive project portfolio analysis is based on the approach commonly seen in healthcare or finance. Projects that make up the portfolio are treated like patients or customers in these areas. Each project is assigned with basic parameters gathered when the project proposal is submitted to the portfolio owner or project management office (a situation similar to e.g. a patient admission to hospital or a loan application evaluation). These parameters should be defined individually in every organization based on its business context. Often they are standardised by project methodologies but companies tend to gather additional project information in order to supply it to other business areas that these projects affect (such as environmental, legal, production and real estate data). Therefore the projects can be parametrised by, among others, project size, length, budget, estimated quality, product type, team experience, NPV, ROI, etc. Based on these parameters, project evaluation measures must be established to create predictive models built with the use of historical data in the process of predictive modelling that consists of five steps: data selection, choice of predictors, determination of performance measures, model evaluation and model selection (Kuhn and Johnson, 2013).

In the selection stage of portfolio management process (Figure 1), the aforementioned parameters gathered from new projects are used as an input data for predictive models that support portfolio owners (managers) in making the decision of whether the project should be included in the portfolio or not, and what priority the project should be assigned with. These two are typical classification problems, which can be solved using classification algorithms such as: classification trees, random forests, association rules and logistic regression. The classification rules generated in the case of these algorithms are more understandable than those created by, for example, artificial neural networks, so that the decision-maker is more aware of the reasons standing behind the model response. After the project selection, validation and initiation phases, predictive analytics may support further decisions during management and monitoring phase. Most decision-making problems related to project portfolio management occur in this phase, including scope of changes, critical problems exceeding budget and time constraints, and also unexpected project proposals. Similarly to project selection, these situations generate classification problems that require decisions as to whether to extend the project, assign additional resources, close the project or initiate a new one. Again, algorithms like classification trees can be useful at this stage, but sometimes the complexity of the situation may require more advanced but less understandable solutions, such as artificial neural networks.

The described situations can be defined as human-initiated, and predictive analysis is thus only a tool supporting a reaction to their occurrence. During the monitoring phase, predictive analytics serves its basic purpose, namely the detection of hidden patterns in the data. It may be applied to detect future project failures such

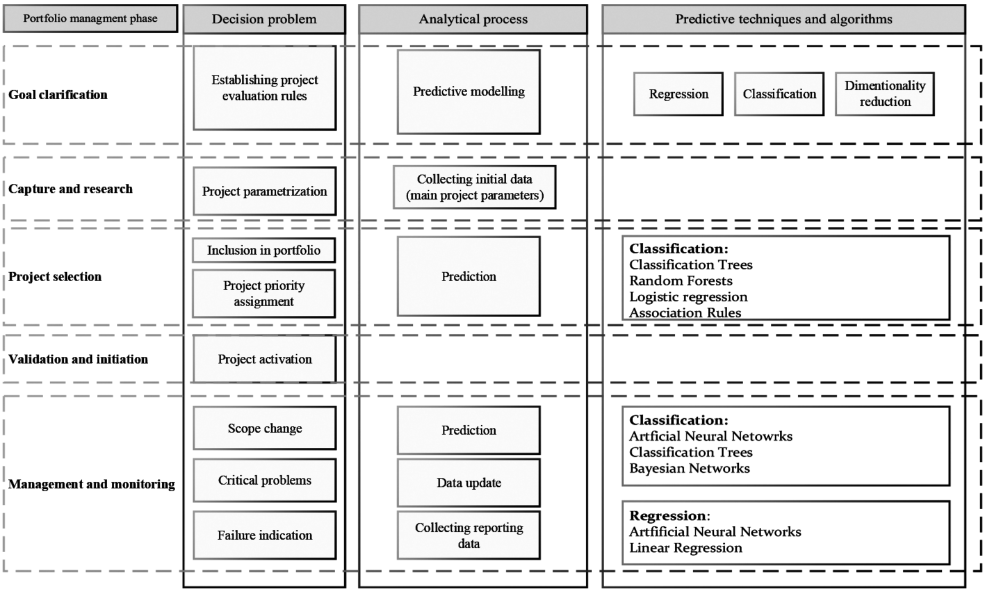

Fig. 1. Framework for predictive project portfolio analysis

Source: own work.

as overspend or delay, before this event is predicted by project manager and reported to portfolio managers or is intentionally undisclosed. The input data for predictive models at this stage require not only static project parameters but also dynamic ones, i.e. current budget deviation, scope completion, procurements status, etc., gathered in the process of project reporting. In this case, it is sometimes expected that predictive analytics is capable of providing decision-makers with estimates regarding final budget deviation, length of delay, etc. that are more detailed than categorical values. To achieve this level of information, regression algorithms such as linear regression or artificial neural networks have to be applied, however this does not rule out using classification algorithms if such level of prediction is sufficient for the portfolio management.

4.2. The limitations of the solution

When applying the proposed framework, one should be aware of its main limitations. As was mentioned in Section 3.2, access to Big Data is essential, as well as expert knowledge both in the area of project management and predictive analysis. The data required for the proper implementation of predictive analytics should contain information from tens or hundreds (depending on project uniqueness) of past project gathered and stored in an organized manner, so that effective and accurate models can be created in the modelling process. The smaller (or more unique) the dataset, the less reliable and more overfitted the models. Hence companies willing to introduce predictive analytics should be highly developed and experienced in the area of project management and IT. This includes the capability to introduce and maintain other project management tools that allow gathering project information (schedules, budgets, reports) which then, after initial preprocessing constitute the input data for effective predictive models. The level of the company's maturity, in terms of project management, must be high enough to enforce standardised and up-to-date reporting to ensure that the data used in the analysis do not contribute to false or inaccurate outcomes.

5. Conclusion

The proposed framework is an attempt to introduce predictive analysis in the area of project portfolio management, that takes a different approach than those used in solutions based on the risk analysis of each scheduled task or thorough project assessments that take weeks to conduct. The solution suggested by the author, follows an analytical approach familiar from other business areas such as healthcare and banking, and considers a project as a parametrised object that follows patterns created by past cases, i.e. projects already completed by the organization. The proposed framework focuses on decision-making problems met by portfolio managers at different stages of project portfolio management process. The most important ones are assigned with algorithms that are most appropriate for the task and should be considered when creating predictive models for these stages. Finally, the author discusses the limitations of the proposed solutions that mainly concern high requirements for project management and IT maturity of the organization willing to apply it, as well as the size of the historical data needed in the process of predictive modelling that precedes any practical use of predictive analytics.

References

Bandara, M., Behnaz, A., Rabhi, F. A., and Demirors, O. (2019). From requirements to data analytics process: An ontology-based approach. *Business Process Management Workshops,* 543-552. Cham: Springer International Publishing.

Baptestone, R., and Rabechini, R. (2019). The portfolio management and influence in projects of decisions. In *New global perspectives on industrial engineering and management* (pp. 293-302). Springer International Publishing.

Cooper, R , Edgett, S., and Kleinschmidt, E. (1992). New product portfolio management: Practices and performance. *Journal of Product Innovation Management,* (16), 333-351.

Crispim, S. F., and Siqueira, L. D. (2014). Alignment of IT projects with organization business models. *Gestão & Produção*, *21*(3), 621-634.

Dinov, I. D. (2018). *Data science and predictive analytics: Biomedical and health applications using R data science and predictive analytics: Biomedical and health applications using R.* Springer International Publishing.

Elbanna, S. (2006). Strategic decision-making: Process perspectives. *International Journal of Management Reviews*, (8), 1-20.

El Morr, C., and Ali-Hassan, H. (2019). *Analytics in healthcare: a practical introduction*. Springer International Publishing.

Etaati, L. (2019). *Machine learning with microsoft technologies*. *Selecting the right architecture and tools for your project*. Berkeley, CA: Apress.

Fauser, J., Schmidthuysen, M., Scheffold, B. (2015). The prediction of success in project management. Predictive project analytics. *Projektmanagement Aktuell*, *26*, 66-74.

Klix, F. (1985). Machine learning. An artificial intelligence approach. *ZAMM – Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, *65*(11), 568-568.

Kuhn, M., and Johnson, K. (2013). *Applied predictive modeling*. New York, NY: Springer New York.

McCarthy, R. V., McCarthy, M. M., Ceccucci, W., and Halawi, L. (2019). *Applying predictive analytics: Finding value in data.* Cham: Springer International Publishing.

Oltmann, J. (2008). *Project portfolio management: how to do the right projects at the right time.* Newtown Square, PA: Project Management Institute.

Platon, V., and Constantinescu, A. (2014). Monte Carlo method in risk analysis for investment projects. *Procedia Economics and Finance*, (15), 393-400.

Project Management Institute. (2017). *The standard for portfolio management – fourth edition*.

Project Management Institute. (2018). Success in disruptive times: Expanding the value delivery landscape to address the high cost of low performance. *Pulse of the Profession*, 35.

Rebala, G., Ravi, A., and Churiwala, S. (2019). *An introduction to machine learning, deep learning and neural networks.* Springer International Publishing.

Taylor, J. (2015). *Framing requirements for predictive analytic projects with decision modeling.* International Institute for Analytics.

Williams, G. (2011). *Data mining with rattle and R*. *The art of excavating data for knowledge discovery*. New York: Springer.