WROCŁAW UNIVERSITY OF ECONOMICS AND BUSINESS

Maciej Wach

Model for the application of predictive analysis in the area of management of investment project portfolios in KGHM Polska Miedź S.A.

SUMMARY OF DOCTORAL DISSERTATION

Dissertation supervisor:

PhD. Dsc. Iwona Chomiak-Orsa,

Associated Professor at Wroclaw

University of Economics and Business

Wrocław, 2022

One of the basic problems in the processes of managing portfolios of investment projects is the assessment of the expected duration of projects and the determination of the related distribution of the budget necessary to implement it. According to the Project Management Institute (PMI, 2021), in the last dozen or so years, in global terms, the number of projects that have been completed on time varies in the range of 52-58% (depending on the flexibility of the organization), and slightly more (59-64%) are implemented within the planned budget. The mining industry is no exception in this case. Similar complications apply to investment projects implemented at KGHM Polska Miedź S.A., although a much more frequent and important problem is the extension of projects than the need to increase their financing (only 20% of projects in 2021), which in the period of low inflation, occurred mainly in cases of a significant increase in the projects' scope. Extending projects for subsequent years, even if they are implemented within the originally planned total budget, is a problem for many reasons. One of them is the benefits lost by not using the funds reserved in the Company's budget for delayed projects, which could be used for other investments, capable of generating revenue much earlier. The effects that are assumed from the implementation of the project in question are also shifted in time, which changes the structure of cash flow and reduces the planned NPV of the project. Another negative phenomenon is the growing financial liabilities. Too early reservation increases borrowing costs if it is one of the sources of financing for a project. Delays in the implementation of investments that are financed by external entities, such as investment banks, force in this case the need to annex previously signed agreements. As projects protract, so does the over-commitment of human resources. Projects block the assignment of specialists to subsequent investment tasks or stop their restoration to full operational work. This is particularly noticeable in matrix organizations, an example of which is KGHM Polska Miedź S.A., where a significant number of project managers combine their role with specialist work provided as part of basic employment. In the case of a large number of projects that make up the portfolio, the launch of new projects makes it necessary to train new, inexperienced project managers, what negatively affects the quality of their implementation. Also with the passing years, the prices of services, materials and equipment increase, exceeding the currently approved project budget and requires reserving additional funds in subsequent years, increasing the budget baselines for these periods. As a last resort, the organization may change the attitude towards the project. In the event of a change of sponsors or clients, e.g. due to the term of office of the Company's Management Board, there is a risk of losing support for their further implementation, and consequently abandoning projects and losing the value of already delivered products.

Creating portfolios of investment projects that are optimally aligned with the company's 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 investments in a project-based manner (Crispim, 2014). As part of KGHM's investment portfolios, which include not only mining plants, but also smelters and ore enrichment plants, more than 400 investment projects are under construction every year. Predicting the exact end date and budget needed for the implementation of these projects is in this case almost impossible from the position of high-level managers. This problem has been tried to be reduced in recent years by introducing the electronic Project Status Report (eRSP), which is a standardized report on the condition of the project, prepared once a month by project managers, integrating source data with current operational knowledge. However, this bottom-up method of prediction is subject to risks, which include, e.g.: the reserve of project managers towards reporting deviations before their actual occurrence, lack of knowledge about possible delays in external, independent processes (e.g. administrative proceedings, procurements, etc.), seeing an opportunity to catch up with the scope, reserving funds for unforeseen events, as well as errors in the use of the reporting tools themselves. Of these, external risks in particular are the cause of deviations that surprise managers, as they are not taken into account by them in the project planning process. In the vast majority of cases, project managers choose the most optimistic plan, without a time reserve, based only on the manager's expert knowledge about the area in which he implements the project. Due to this type of threats and the dependence of management information on the reliability and skills of project managers, it is justified to use the tool of predictive analytics, which in complex data sets is able to find dependencies leading to the selection of endangered projects based on unconventional inference. The basis for its implementation has been done in recent years through the development of access to large project data sets. In 2013, KGHM implemented a Project Management System (SZP), based on the Microsoft Project Server 2010 platform (in 2019 there was an update to version SZP 2.0 based on MS Project Server 2016) (Wach, Chomiak-Orsa, 2019), which with the entry into force of the KGHM Project Management Methodology (KGHMStep) developed on the basis of the TenStep® Project Management Process methodology - a project management standard fully compliant with the PMBOK® Guide methodology developed by the Project Management Institute (PMI, 2013), enabled project managers to run projects within a single system, able to combine functions related to scheduling, making changes to the project, accounting operations conducted in the SAP ERP system and the implementation of related contracts. In addition, there are three other tools in the organization that support the management of portfolios of investment projects, and which generate significant amounts of useful project data, these are: spreadsheets in which the plan of the Portfolio of Tangible Investments is drawn up as well as its monthly revisions; integrated with the SZP system, the previously mentioned eRSP application, in which monthly project reporting is carried out; portfolio reports from local (divisional) project management offices (LPMOs). Each of these sources provides additional analytical information and the generated data is assigned to the basic unit (project), hence aggregating them is possible and attempts to extend analytics are made, for example, in Business Intelligence class tools (in this case Microsoft Power BI). Such activities are defined as the so-called descriptive analytics, which is responsible for describing the current or past reality, which most often translates into data processing and visualization, facilitating their interpretation. Predictive analysis is increasingly becoming an important tool to support managers and other decision-makers in solving decision-making problems. Using advanced mathematical formulas and statistical algorithms to identify dependencies, relationships and formulas in datasets, it is able to generate future scenarios for a given set of input information. This makes it possible to predict, with some probability, future states and events, such as: device failures, material demand levels, market trends, customer financial decisions, medical diagnoses, etc. (IBM, 2017).

Due to the complexity, unpredictability, individualized approach and difficult parameterization of the process, project management, unlike other business areas such as: finance, marketing or healthcare, is, in global terms, an area with relatively low use of advanced analytical tools based on artificial intelligence. The analysis of literature sources regarding the use of predictive analytics in the area of project management showed that the existing solutions are significantly limited. Some of them, such as the Monte Carlo method (Plato, Constantinescu, 2014) or the Project Predictive Analysis approach (Fauser, Schmidthuysen, Scheffold, 2015), do not use machine learning algorithms belonging to the set of artificial intelligence methods, but are based primarily on statistical analysis. More advanced predictive models were used to estimate the cost, resources, and time needed for software development in IT projects (Tadayon, 2005; Kaushik, Soni, Sonia, 2013), or e.g. in engineering projects, where time, cost and quality are also analyzed, together with safety issues, in order to obtain an overall assessment of the project (Apanaviciene, Juodis, 2006) or key factors that influence the outcome of the construction are identified, such as: the competence of the project leader, the overall experience of the team, or the impact of decisions made during the project (Liu et al., 2009). The closest, in terms of the analytical goal of predicting projects' deviations, is the concept of using neural networks in the prediction of development projects in German industry (Schuh, Riesener, Dölle, 2017; Schuh, Riesener, 2017; Riesener et al., 2019). It assumes the decomposition of development projects into individual activities, for which groups of factors influencing the deviations of individual task parameters are identified. For each of these groups, predictive models are created that can determine the expected level of deviation for the cost, quality and duration of elementary tasks of new initiatives.

According to the author, there are no studies on the use of predictive analytics in project management, which would be dedicated to an environment similar to the one existing at KGHM. In the described solutions, there are no references to project portfolios, parameterization of projects and the use of project reports as data sources. At the same time, most of the existing solutions are based on the decomposition of projects into smaller tasks, which in the case of KGHM would mean the need to change the entire approach to generating project data, because at the moment there are no mechanisms to collect data on individual types of tasks in implemented projects. For this reason, it would also be impossible to conduct any research on the effectiveness of models based on such assumptions.

The author, wanting to fill this gap, on the basis of literature studies, internal regulations of the company and focus research on expert groups from the central project management office, developed an original model for the application of predictive analysis in the area of portfolio management of investment projects at KGHM Polska Miedź S.A., which modifies the existing environment of project portfolios in technical and procedural aspect, as well as conducted empirical research on the verification of this model and the assessment of predictive modeling capabilities based on currently available project data.

With regard to the discussed topic and the indicated research gap, the following **research problem** was formulated:

Is it possible to create a model for the use of predictive analytics, supporting decision-making processes in the area of portfolio and project management of tangible investments in KGHM Polska Miedź S.A.?

This question is the starting point for **the main hypothesis**:

There are predictive models which, when fed with project data, are capable of supporting decision-making processes in the area of portfolio and project management of tangible investments in KGHM Polska Miedź S.A.

And **specific hypotheses**:

- **H1:** The quality of the project data held is sufficient for the implementation of predictive analytics in the area of portfolio and project management of tangible investments in KGHM.
- **H2:** The main decision problems are modellable using predictive analysis.
- **H3:** Predictors related to the subject of the project are considered important in the predictive modeling process.
- **H4:** Predictive models increase prediction accuracy compared to project manager estimates.
- **H5:** Incorporating data from eRSP reports into the modeling process improves the quality of prediction.

In order to verify the assumed main hypotheses and detailed hypotheses, the following objectives were set to be achieved:

Main objective:

Development of a model for the use of predictive analytics supporting decision-making processes in the area of portfolio and project management of tangible investments in KGHM Polska Miedź S.A.

Specific objectives:

- **O1:** Identification of conditions for the implementation of predictive analytics at KGHM Polska Miedź S.A.
- **O2:** Assessment of existing project data sources in terms of their quality.
- **O3:** Development and modelling of decision-making scenarios for the KGHM investment area using predictive models.
- **O4:** Identification of key predictors.
- **O5:** Evaluation of the effectiveness of predictive models, based on actual project data.

Secondary objectives:

Epistemological objectives:

EO-1: Systematizing knowledge about predictive analytics and decision-making problems.

EO-2: Identify the applications of predictive analytics in project and portfolio management problems.

Utilitarian objective:

UO-1: Issuance of recommendations for the implementation of predictive analysis in the area of portfolio and project management of tangible investments in KGHM Polska Miedź S.A.

The following methods were used in the conducted research: predictive modeling, statistical data analysis, expert interview and the method of analysis and criticism of literature.

The proposed model is divided into two main perspectives: static and dynamic. As part **of the static perspective** defining the organization of the enterprise in area of investments, four main planes were defined, which correspond to the most important aspects affecting the efficient use of predictive models in the organizational conditions of KGHM: technical, regulatory, analytical and stakeholder.

As part of the technical plane, an approach to the integration of IT tools used in the area and its environment was proposed. It is a response to the insufficient quality of data currently generated by systems and applications, which, combined with the large volume of this data, limits the current analytical capabilities of organizational units and negatively affects the implementation capabilities of predictive models. Another of the planes, the analytical plane, would be fully responsible for the model (mathematical) side of the application of predictive analytics, including mainly the configuration of predictive models and the business contribution to its configuration. Due to the high formalization of all areas of the enterprise and ensuring the correct and timely flow of data within and between business areas, the regulatory plane was also included in the model. As part of it, formal documents of the company were defined, which will be affected by the occurrence of predictive analytics in management processes and which will require modification. Within the stakeholder plane, people and organizational units have been designated, which in the proposed model will be responsible for the efficiency and quality of functioning of predictive analytics in the area of project management and their target duties have been defined.

The dynamic perspective of the model, as opposed to the static perspective, describes the integration of the model into the management processes of the investment division, along with a proposal to modify these processes. For this purpose, management processes and specific decision-making problems were selected, for which the implementation of analytics is justified and will give a measurable effect in the form of decision support. Finally, the model proposed to modify five processes.

As part of the portfolio definition process, which is one of the most important processes in the area of investment, a change in three of its steps was recommended. At the stage of preparing the forecast of the planned level of expenditures and expenses of the portfolio, it was recommended to create an expense forecast based on the current assumptions of the activated projects and those planned to be launched. Time to include updates to predictive models at stage of approval of new Portfolio Rules by the Development Committee was also included in the process. The third modification is the inclusion of analytics in the planning of portfolio proposals by the Local Portfolio Owners by supporting the expected implementation of active projects and those already submitted to the plan. In the Process of Updating the Portfolio of Tangible Investments, it was recommended to support local PMOs through the Analytical Unit in the scope of generating information at the stage of verifying the feasibility of the proposed changes to the portfolio, including prediction regarding the implementation of the assumed project budget in the period declared by the applicant. In the Effects Plan Monitoring Process, indirect use of predictive analytics was introduced in the identification of projects at risk of lack of planned implementation of effects (cost or production), by generating predictions of project implementation in the assumed time, which translates into timely delivery of anticipated effects. The Consolidated Portfolio Reporting Process includes steps aimed at adapting models to planned changes in the parameterization of projects, as well as supplementing cyclically provided management information regarding current performance with forecasts created in the predictive modeling process. The changes proposed in the Project Management Process concern two stages of the process. The first is to introduce the forecast of the implementation of the new project to the applicant and to make the results of the consultation available to the Portfolio Owner before agreeing to include the project in the Portfolio. The next step is to consult and analyze the feasibility of implementing the proposed plan with the Project Manager and to deliver it to the Portfolio Owner before the decision to accept the change.

The dissertation consists of seven chapters, which complement the introduction and the ending summarizing it. The first three chapters are theoretical and systematize knowledge. Chapters four and five are a study of the conditions and the concept of implementing a model for the use of predictive analytics in the area of portfolio and project management of tangible investments in KGHM Polska Miedź S.A. Chapter six describes the course of research conducted with the use of predictive modelling and the assessment of the resulting models. The seventh chapter contains a summary in the form of implementation recommendations.

Chapter 1 characterizes a mining company such as KGHM Polska Miedź S.A., indicating the elements of the main production line and the most important products that are produced in it and then sold on international markets. The most important divisions of the Company and the companies belonging to the KGHM Group were presented. The definition of portfolios and investment projects, as well as project management methodologies, were reviewed. The content of the following methodologies functioning in KGHM is described in detail: KGHM Project Management Methodology and KGHM Portfolio Management Methodology, indicating their content, described processes and the most important definitions related to project roles and portfolios. The relationship between the company's strategy and portfolios of tangible investments was demonstrated. The structure of the investment division at the KGHM Head Office based on its elementary organizational units was described, indicating the main tasks carried out by them. The main formal documents regulating the functioning of the area, as well as the management processes defined in them and the related decision-making problems are indicated. Data on the performance of the area in 2019-2021 were presented, including, among others: the number of financed projects, the budget planned and implemented in individual years, the distribution of projects between portfolios and project statistics in the context of budget deviations in 2021. It also characterizes the IT tools that are used in the area of investments, both for project and portfolio management, as well as reporting and monitoring of their status. In addition, tools currently under development and used in supporting areas that collect data related to investment projects are described.

Chapter 2 focuses on the analysis of decision-making processes. The literature was reviewed in terms of: definition of decision-making processes, actors playing roles in these processes and approaches to the construction of their course and differentiation of the theory of decision models. Particular attention was paid to the issue of multi-criteria analysis. Its function as a tool supporting the solution of decision-making problems was characterized, along with known methods and types of applications. In particular, applications of multi-criteria analysis in project management and mining industry were recognized. The application of multi-criteria analysis in monitoring investment projects at KGHM was presented, which is based on a tool created by the author, based on BI class reports, powered by data from electronic project status reports (eRSP). Chapter 2 is complemented by a literature review related to decision support schemes. Their theoretical construction, classification according to the purpose of use and examples of application in the mining industry and project management were compiled.

Chapter 3, based on literature research, characterizes predictive analytics, its significance and relation to descriptive and prescriptive analytics. The stages of the predictive process are presented, which consists of: data collection and storage, modeling, analytics and reporting. Similarly, the process of predictive modeling is described, along with its stages: data division, selection of predictors, determination of performance measures, evaluation and selection of the model. Later in the chapter, machine learning is defined with a division into two basic categories: supervised and unsupervised learning, along with the techniques and basic algorithms that belong to them. The chapter is complemented by a discussion of the main applications of predictive modeling in mining and an analysis of attempts to apply it in project and portfolio management.

Chapter 4 characterizes the requirements for the implementation of predictive analytics in the area of KGHM's investments, that go beyond the knowledge of the modeling process. The main determinants were defined, which were grouped into three categories, such as: complexity of the problem, access to Big Data and business knowledge. As a part of the assessment of business knowledge and complexity of the problem, research problems resulting from direct research conducted on a group of experts were synthesized, the main formal documents on the basis of which the KGHM investment business area operates were described and then decision-making problems occurring in business processes were characterized in detail, identifying: business context, decision-makers, frequency of occurrence and the input data on the basis of which these decisions are made. Two main analytical issues are described that limit the use of predictive analytics with currently available tools, which are recording formal changes in projects and the ability to track the scope of projects. Next, referring to the level of access to data, the systems and applications that affect the collection and processing of data in the area are listed and then all available data sources, along with all the parameters contained in them, are described in detail. The last part of the chapter analyzes previously identified sources and decision problems, as a result of which data sources and decision problems that qualify for use in the predictive modeling process have been identified. At the same time, the reasons why some of the sources were rejected and which resulted from insufficient quality parameters of these data were described. The same has been done in the case of decision-making problems, for which it is not possible, under the current conditions, to create predictive models.

Chapter 5 defines the concept of a model for the use of predictive analytics in the management of portfolios of KGHM's investment projects. The model is structured into two key perspectives: a static perspective and a dynamic perspective. A static perspective describes how to reorganize the environment of the area of investments, which is required for the implementation of predictive models to work properly in it. In this perspective, 4 main areas were distinguished, which correspond to the most important aspects affecting the efficient use of predictive models in KGHM's organizational conditions: technical, regulatory, analytical and stakeholder. Within these areas, author defined: a scheme of integration of the IT tools used, which currently generate a large amount of project data, along with tools currently under development; responsibility for the model (mathematical) side of the application of predictive analytics, including mainly the configuration of predictive models and the business contribution to its configuration; formal documents of the enterprise, which will be affected by the existence of predictive analytics, crucial to ensuring the correct and timely flow of data within and between business areas; employees and organizational units that in the proposed model will be responsible for the efficiency and quality of functioning of predictive analytics in the area of project management, along with their designated duties.

As part of the dynamic perspective, the proposal for the implementation of predictive analytics was detailed, including a proposal to modify the management processes of the investments business area. The model proposes to modify five of the six processes that are currently defined in the Regulations for Portfolio and Investment Project Management at KGHM Polska Miedź S.A., these are the following processes: Defining of the Portfolio, Monitoring of the Effects Plan, Updating the Tangible Investment Portfolio, Reporting of the Consolidated Portfolio of Tangible Investments and Process of Management of Tangible Investment Projects. These processes define the additional role to be played by the Analytical Unit, as well as its tasks, along with the existing and newly added process steps to be the predecessors and successors for these tasks.

Chapter 6, at its beginning, presents the methodology of the research described later in the chapter and describes the conclusions of the model verification resulting from substantive research carried out on a group of experts. Subsequently, the project data were evaluated, in which the choice of data sources qualified for the research and the time range to which they relate were justified. For each of the three selected sources, a detailed quality assessment was carried out in terms of: availability, completeness, uniqueness, timeliness, accuracy, consistency, validity and problems occurring in their processing. The rest of the chapter describes the subsequent steps in the predictive modeling process. First, it shows how to import and preprocess batches of project data. Then, three predictive scenarios were defined, including one three-variant scenario, which correspond to the actual decision-making problems identified earlier in the area of KGHM's investments. For modeled problems, which were identified as classification problems, the choice of machine learning algorithms was justified and methods for assessing the quality of emerging predictive models were described. In the main part of the chapter, each decision scenario is characterized, the modeling process using IBM SPSS Modeler 18.1 software is described, related statistical data is presented, and the results obtained as a result of model testing are summarized. For each scenario, the effectiveness of the resulting models was evaluated and the fault matrix data containing the predicted and expected values for the test data set were analyzed. On the basis of these data, for each scenario, evaluation conclusions for the main model of application of predictive analytics in the area of KGHM's investments were presented.

In Chapter 7, the author summarized the recommendations resulting from the research and analysis carried out in the previous chapters. First of all, recommendations resulting from the model of using predictive analytics in the area of KGHM's investments were presented. These recommendations concern the requirements for the construction of component planes for a static perspective, as well as the proposed modifications of management processes described in the dynamic perspective of the model. The second part of the recommendation is a set of requirements for systems and applications functioning in the investment area, which can ensure the collection of project data of better quality. These requirements are divided into two groups – functional and integration requirements. The first group concerns identified deficiencies in applications that prevent the generation of data that is important in terms of predictive abilities or affect the fact that the generated data is of low quality. The second group of requirements concerns deficiencies in integration between applications, which affect the loss of consistency of existing data and significantly increase the workload on their processing. The last part of the chapter presents implementation recommendations for predictive models resulting from the synthesis of research results from chapter 6. The author presented the effectiveness of models and the impact of data set characteristics on individual decision scenarios. The results of the selection of predictors for individual cases were also presented, as well as conclusions from the assessment of individual predictive algorithms, based on measures of the quality of models built on the basis of these algorithms.

At the end of the work, research results and created recommendations were compiled in the context of the achieved goals, while justifying the verification of hypotheses related to these goals.

The specific objective O1 referred to the identification of conditions for the implementation of predictive analytics at KGHM Polska Miedź S.A. As part of its implementation, three main categories of conditions were defined: complexity of the problem, access to Big Data and business knowledge. As part of the assessment of business knowledge and complexity of the problem, the main formal documents on the basis of which the KGHM Investment Division operates were described, and then the decision-making problems occurring in business processes were characterized in detail, identifying: business context, decision-makers, frequency of occurrence and the input data on the basis of which these decisions are made. Two main analytical problems have been identified that limit the use of predictive analytics with currently available tools, such as identifying changes in projects and the ability to track the scope of projects. Next, referring to the level of access to data, the systems and applications that affect the collection and processing of data in the area are listed, and then all available data sources, along with all the parameters contained in them, are described in detail.

The specific objective O2 referred to the assessment of existing project data sources in terms of their quality. As a result of the analysis of previously identified data sources, those that are eligible for use in the predictive modeling process were selected. At the same time, the reasons why some of the sources were rejected and which resulted from insufficient quality parameters of these data were described. The elimination of sources took place due to: limited availability of project data and excessive aggregation (SAP Ariba and ePlan 4.0 procurment systems), timeliness of data (financial plan in the SZP system), or the level of implementation of the data generating tool (eWOPP application, risk register). Sources such as the Tangible Investment Portfolio Plan (PIR), data from electronic project status reports (eRSP) and aggregated data from portfolio reports were qualified for further research and detailed assessment. They are characterized by the fact that: they are updated at least once a month or with each change (timeliness and reliability), they apply to all investment projects of the consolidated portfolio (completeness), have the ability to combine them with data from other sources (consistency) and are collected in a form that allows their processing (availability). Due to the fact that not every source fully represents these qualitative features, for the purposes of the research experiment, selected sources were analyzed in detail in order to identify the main problems that hinder the data processing stage. These are: basing the data source on manual data entries (spreadsheets), incorrect project numbers in PIR plan, resulting from their complexity and lack of evaluation when manually entered, making it difficult to link data to the same project, separating projects in the PIR plan (and its revisions) into lines for phases, subprojects or components with different funding status, non-uniform field structure and their naming between corrections of the same plan and between all reports, changes in the number and permissible values of variables parameterizing the project between years, the need to convert units into budget variables, difficult identification of files containing PIR baselines and their revisions in network resources, text format of date variables, different from the one built into MS Excel. The results of achieving the O2 objective allowed to **negatively verify the H1 hypothesis**, denying that the quality of the project data currently available is sufficient for the implementation of predictive analytics in the area of portfolio and project management of tangible investments in KGHM. Many project data sources were rejected due to poor data quality, mainly in terms of timeliness, consistency and accessibility. In addition, data eligible for empirical research required a significant amount of manual work on their initial processing to eliminate emerging errors related to accuracy, uniqueness and consistency. This is a factor that disqualifies them from immediate implementation of predictive analytics, which must be based on on-line sources of project data, not requiring advanced processing.

The specific objective O3 referred to the development and modelling of decision-making scenarios for the KGHM investment area using predictive models. On the basis of the analysis and selection of previously identified decision-making problems as well as sources of project data, 3 predictive scenarios were created, including one three-variant, modelable and eligible for support of decision-makers in decision-making processes of the KGHM investment area. Scenario No. 1 concerned decisions related to the selection of projects for the portfolio up to the level of the set budget limit, taking into account active projects and related obligations, which are part of the Portfolio Update Process and the Tangible Investment Project Management Process. The target variable is the expected level of deviation of the budget at the end of the year from the budget baseline. Scenario No. 2 (in three variants) concerned decision-making problems related to reporting the status and updating of the portfolio of tangible investments, related to the processes of: Defining the Portfolio, Monitoring the Effects Plan and Reporting the Consolidated Portfolio of Tangible Investments, in which strategic decisions are made primarily related to the further direction of activities in portfolio management (e.g. intensification or reduction of expenses). The target variable was the expected level of deviation of the budget at the end of the year from the baseline budget and the expected level of deviation of the budget at the end of the year from the expected expenditures given in the eRSP reports (in scenario option 3). Scenario No. 3 concerned decision-making problems related to changes in the portfolio resulting from the introduction of changes in projects and proposals to include new projects in the portfolio during the calendar year, as part of the Tangible Investment Project Management Process. The target variable was the expected level of deviation of the budget at the end of the year from the baseline budget. The above scenarios were modeled using IBM SPSS Modeler 18.1 software, which in its functionality allows: import and processing of data; construction, testing and evaluation of predictive models; use of the created models with new data sets to generate predictions. The implementation of the O3 goal allowed to **positively verify the H2 hypothesis** and confirm that the main decision problems are eligible for predictive modeling.

The specific objective O4 concerned the identification of key predictors for modelled decision scenarios. The estimation of predictors was made on the basis of a built-in (IBM SPSS Modeler 18.1) predictor selection node, which reduced the set of expertly selected predictors by those that did not meet the significance criteria. The most frequently rejected predictors were: the amount of the total budget and the ratio of the annual budget to the total budget, which means that in the perspective of year-end deviations, the size of the project and plans in subsequent years don’t matter. For eRSP data, the models rejected various predictors related to the answers to the questions included in the report, among them those regarding the purpose of the prediction, such as the question of whether the annual budget would be executed as planned. This allows us to conclude that the quality of answers to the questions contained in the report is questionable and great emphasis should be placed on the understanding, correctness and reliability of the completed reports.

One of the most important predictors turned out to be the static parameters of projects, which are specified in the PIR plan and which refer to the diverse nature of projects, including: label, category or branch in which the project is implemented. The most important budget parameter is the distribution of the budget during the year, expressed by the weighted average of the monthly index against the planned monthly budgets. It is also important to know whether the project ends or starts in the current year. In the case of models concerning changes made in the project, the following are considered to be important predictors: the direction of change of the annual and total budget and the month of making the change. Among the distinguished predictors, two important sets can be distinguished, the first of them (static parameters) is responsible for the characteristics of the project and placing it in the company's structures, which means that the type of project product, its impact on production or the branch in which it is implemented is of considerable importance. The second group of predictors are predictors that detail the current situation of the project in the year for which the prediction is made, primarily the financial one. The results obtained as a result of achieving the O4 goal allowed to **positively verify the H3 hypothesis** and confirm that: predictors related to the subject of the project are considered important in the predictive modeling process.

The specific objective O5 referred to the evaluation of the effectiveness of predictive models, based on actual project data. As a result of modeling decision scenarios on the basis of selected algorithms, predictors and a set of learning data from 2018-2020, it was possible to evaluate the resulting models using a set of test data – project data from 2021. For the analysis of their widest possible range, as a result of a combination of algorithm parameters, in each analyzed case as many as 273 models based on 8 different algorithms were built. The most effective model based on a machine learning algorithm in most scenarios was definitely **the random trees model.** With the exception of Scenario 3, it had the best results in terms of efficiency the area under the ROC curve and lift, which translated into the fact that it was highly effective in indicating any deviations, as well as a better distribution of correct identifications between classes, which again translates into the ability to better select projects with potentially low deviation. Linear regression and LSVM models also had a high efficiency of the classifier, they showed a large number of correct matches, but succumbing to the tendency to unilaterally indicate all predicted deviations as high. They are therefore mainly effective in detecting excessive deviations, but their use in the selection of projects that will implement the plan is lower than in the case of random trees. In the results of the obtained models, they indicated the high ineffectiveness of neural networks, the SVM algorithm and random forest algorithms, which in all scenarios were not at the forefront of the ranked models. The Bayesian network model achieved satisfactory results in terms of all parameters only in Scenario No. 3.

Achieving the O5 objective allowed for **a positive verification of the H4 hypothesis**, confirming that predictive models increase the accuracy of prediction compared to the estimates of project managers. The overall accuracy of the estimation made by project managers (45.0-64.8%) is lower than the accuracy of model prediction for the same cases (64.5-79.6%). There is also a discrepancy in the effectiveness of indicating projects with low deviation, which in the case of predictive models is at least 45.6% (maximum 63.9%), and in the case of estimation of managers can be as low as 26.7% (maximum 63.4%). In the case of high deviations, managers' estimates are more narrow than predictions. 69.8-83.4% of deviations classified as high are confirmed, where in the case of prediction, this range is wider and amounts to 61.5-85.4%.

As part of the same objective, **the H5 hypothesis** that incorporating data from eRSP reports into the modeling process improves prediction quality **has been partially confirmed**. As the deviation threshold increased and the following months of the year increased, the effectiveness of indicating projects with low deviation by predictive models using reporting data showed an increase from 45% to 64%. This is a lower range than for the variant without reporting data, where the effectiveness increased from 60 % to 73 % depending on the model. Compared to models without reporting data, the tendency to incorrectly attribute low deviations as predictions of high deviation has decreased significantly, i.e. the so-called model sensitivity has increased. Using reporting data, the ratio of true positive responses (TP) to the sum of true positive and false negatives was between 32% and 79% depending on the month and the acceptable threshold, when without the use of reporting data it was only between 15% and 67%. This means that in the case of data from PIR plans, we have greater confidence that the project classified as a project with low deviation will actually be so, but in the case of reporting data, we have greater confidence that the deviation of the project with the ultimately low deviation will be classified as low, not high.

Based on the results of the presented research, it can be concluded that the implementation of the predictive analysis model in the area of portfolio management of investment projects of KGHM Polska Miedź S.A. is possible. However, it is necessary to meet a number of requirements related to the integration of the model into existing management processes at KGHM. These requirements were collected in the form of recommendations to technical requirements, formal regulations, stakeholder competences and the creation of analytical capabilities, as well as modifications of the currently defined processes of the portfolio management area of KGHM's investment projects. A derivative of the implementation requirements for the technical level are detailed recommendations for the development of systems and applications functioning in the investment area and its environment, which are of key importance in generating high-quality project data necessary to supply predictive models. An important part of the recommendation was also the creation of detailed conclusions for the construction of predictive models, which include: processing of current project data, effectiveness of the use of analytics in individual predictive scenarios against the background of PIR plans and eRSP reports, as well as the selection of predictors and machine learning algorithms for these scenarios. As a result of defining the above recommendations, **the utilitarian objective UO-1** was achieved, which was to issue recommendations for the implementation of predictive analysis in the area of portfolio and project management of tangible investments in KGHM Polska Miedź S.A.

REFERENCES

1. Apanaviciene, R. and Juodis, A. (2006) 'Predicting effectiveness of construction project management: Decision-support tool for competitive bidding', *Operational Research*, 6(3), pp. 347–360. doi:10.1007/BF02941262.
2. Crispim S. F., Siqueira L.D. (2014) "Alignment of IT projects with organization business models." Gestão & Produção, 21 (3) 621-634.
3. Fauser, J., Schmidthuysen, M., Scheffold, B. (2015). 'The Prediction of Success in Project Management. Predictive Project Analytics'. in *"Projektmanagement aktuell",* no. 5
4. IBM (2017) 'Descriptive, Predictive, Prescriptive : Transforming Asset and Facilities Management with Analytics' in Watson IoT: 1–8.
5. Kaushik, A., Soni, A. and Soni R. (2013), 'A Simple Neural Network Approach to Software Cost Estimation'
6. Liu, J. i *in.* (2009) 'Study on Engineering Project Management Based on Neural Network', in *2009 International Conference on Electronic Commerce and Business Intelligence*. IEEE, pp. 234–237. doi:10.1109/ECBI.2009.125.
7. Platon, V. and Constantinescu, A. (2014) 'Monte Carlo Method in Risk Analysis for Investment Projects', *Procedia Economics and Finance*, 15, pp. 393–400. doi:10.1016/S2212-5671(14)00463-8.
8. PMI (2013) The standard for portfolio management, 3rd ed. Project Management Institute, https://www.pmi.org/-/media/pmi/documents/public/pdf/certifications/standard-for-portfolio-management-third-edition.pdf?v=95a98c16-aebd-4f01-9e21-4769fbad4891 (access: 04.09.2022)
9. PMI (2021) 'Beyond Agility: Flex to the Future.', in *Pulse of the Profession, pp. 3-19. https://www.pmi.org/learning/library/beyond-agility-gymnastic-enterprises-12973 (access:04.09.2022)*
10. Riesener, M. et al. (2019) 'Implementing Neural Networks Within Portfolio Management to Support Decision-Making Processes', in *2019 Portland International Conference on Management of Engineering and Technology (PICMET)*. IEEE, pp. 1–7. doi:10.23919/PICMET.2019.8893760.
11. Schuh, G., Riesener, M. and Dölle, C. (2017) 'Concept for development project management by aid of predictive analytics', *PICMET 2016 - Portland International Conference on Management of Engineering and Technology: Technology Management For Social Innovation, Proceedings*, pp. 2040–2047. doi:10.1109/PICMET.2016.7806640.
12. Schuh, G., Riesener, M. and Dolle, C. (2017) 'Implementation and assessment of a predictive analytics model for development project management', in *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. IEEE, pp. 696–700. doi:10.1109/IEEM.2017.8289980.
13. Tadayon, N. (2005) 'Neural network approach for software cost estimation', in *International*

Conference on Information Technology: Coding and Computing (ITCC'05) - Volume II. IEEE, pp. 815-818 Vol. 2. doi:10.1109/ITCC.2005.210.

14. Wach, M. and Chomiak-Orsa, I. (2019) 'Improvement of investment processes in mining company by implementation of project management system', in *Mining Goes Digital*. London: CRC Press, pp. 47–56. doi:10.1201/9780429320774-6.