Data mining applied for earthworks optimisation of a toll road construction project

Andri Irfan Rifai[1,*], *Yusuf* Latief[1], and *Leni* Sagita Rianti[1]

[1]Civil Engineering Department, Universitas Indonesia, Indonesia

Abstract. The length of the toll roads operating in Indonesia is still less than in other countries. Significant acceleration is needed to keep up with the country's traffic needs. Acceleration of development should be supported by the development capacities of road operators, one such capacity being earthworks. Data on earthworks can be utilised as a knowledge base and processed via a data mining approach, the results of which form the basis for interpretation and productivity predictions. The aim of this study is to develop a decision support system for the earthworks of a toll road construction project using the approach of data mining historical cases. The data mining approach used an artificial neural network and support vector machine analysis methods. The result is multi-objective optimisation with a genetic algorithm using real-world data from previous Indonesian toll road construction. This work aims to present a practical alternative for the optimisation of earthworks.

1 Introduction

The length of the toll roads operating in Indonesia is low compared to that of other countries. This is despite various parties' belief that the availability of infrastructure has an important role in the growth of the national economy [1]. The growth of transportation infrastructure is key to improving growth and development. From a number of infrastructure provisions, the transport infrastructure sector has a multidimensional impact on the economic growth of other sectors [2]. To realise overall growth, the government must ensure the availability of reliable and adequate transportation infrastructure, supporting its substantial and decisive role in the productivity of the country.

To reach the development level of other countries and achieve the toll road development target, various strategies can be utilised, one of which is simultaneous development in various locations. To make this possible, the toll road development organiser and all stakeholders must continue to develop themselves to maintain, expand and improve the performance of development activities. Simultaneous development in numerous locations can cause problems, one of which relates to resource constraints. For example, the resources required for earthworks require serious attention, because failure in this stage will have an impact on the overall development process [3].

The resources required for earthworks consist of methods of execution, mechanical and human resources, and costly equipment. Developments in technology and road construction methods have encouraged an increased use of machinery in every construction project. Indeed, mechanical tools are vital resources in construction projects [4]. However, the cost

---

* Corresponding author: andrirfan@yahoo.com

of earthwork equipment is high. Therefore, the role of construction management in the field of earthworks can have a strong influence on the overall efficiency and profitability of construction work.

The level of efficiency and effectiveness of a piece of earthwork equipment is based on the productivity of the tool. Productivity is used as a guide in determining the duration of each job and the amount of earthwork equipment required. An earthwork equipment management system is operated continuously for roadworks, comprising design, planning, development, operation, maintenance, and control. All stages of the construction management system cycle have equally important roles. The stages of an earthwork equipment management system have a significant influence on maintaining the performance of construction management if followed continuously over a long time. Construction engineering activities have a vital role in improving project performance in terms of meeting budgets; schedules; and safety, quality, and sustainability standards [5]. Earthwork equipment management systems can be developed using various data approaches and other historical records.

Historical data on the productivity and effectiveness of the earthwork equipment used in previous construction projects can be used to plan future work through accurate interpretation and prediction. Very large datasets are only information without meaning if not interpreted accurately and translated into accurate predictions. Thus, a model that can provide a good approach to the interpretation process is needed. Data mining (DM) is one of the most widely used approaches for data interpretation in various disciplines. Through the artificial intelligence (AI) approach, DM has enormous potential to assist in interpretation and prediction [6]. In construction management and earthworks, AI can serve as a better approach to analysis [7].

Various attributes of the construction management system for earthworks should receive balanced attention. Earthworks, a basic operation for any type of construction, depend heavily on equipment. The productivity and safety of earthwork equipment are determined by the effective management of the equipment [8]. All issues and objectives should be addressed and resolved thoroughly by the system on an ongoing basis. Where multiple objectives must be achieved simultaneously, a multi-objective optimisation (MOO) approach is required. In general, there is no single optimisation solution that can simultaneously generate minimum or maximum values for all objectives [9]. Finally, a good construction management system is a system capable of providing a tool for users and decision makers that allows them to easily understand and use the system.

Based on some of the above concepts, further development of earthwork optimisation is needed to take full advantage of DM for optimisation and priority determination as a strategy to increase the productivity of earthwork equipment. This approach is expected to be an alternative that complements some of the other existing model concepts. The optimisation result must be able to provide solutions for the improvement of earthwork optimisation models in a toll road construction project.

2 Literature review

Earthworks include all work related to digging, breaking, loosening, loading, hauling, transporting, dumping, filling, spreading, levelling, or compacting soil or rock using earthwork equipment [10]. These jobs are widely required in civil works such as construction of highways, dams, embankments, irrigation canals, canals, and airports. Although the common term is earthworks, it is not only limited to soil but sometimes also related to rocks. Indeed, earthwork equipment can be utilised for both soil and rock [11]. What is meant by soil here is the top layer of the earth's surface, being relatively soft, not very compact and

composed of loose granules, whereas rock is harder, more compact and composed of rock-forming minerals.

2.1 Earthwork productivity

Productivity is the ratio of generated output to the input resources used, based on some measure of value. In a construction project, the input for the productivity ratio is the value of the construction process, which can be separated into labour costs, materials, costs, methods and equipment. The success or failure of a construction project depends on the effectiveness of resource management [12]. In a system, generally "something" that works to run it is needed, namely organisation. Organisational effectiveness is the key characteristic that drives the success of the subsystems. The human factor becomes the determinant for achieving a defined level of productivity. To obtain the desired level of productivity and minimise any risk that may occur while prioritising safety and health, the project leaders must understand the capabilities and limitations caused by the condition of the project location [13].

According to another approach, productivity is the capacity of equipment per unit of time (m^3/h); by this measure, earthwork equipment is an important factor in projects, especially large-scale construction projects. The purpose of the use of earthwork equipment is to facilitate the work so that the expected results can be achieved more easily in a relatively short time compared to manual techniques. The productivity of the equipment depends on its capacity, cycle time, and equipment efficiency. The work cycle in material transfer is a recurring activity. The time required in the above activity cycle is called cycle time. The cycle time itself consists of several elements [14].

2.2 Optimisation of earthwork equipment productivity

An optimisation approach to earthworks is needed to optimise limited resources to meet the growing need for earthwork activities [15]. This is in line with the research undertaken by Parente et. al., who strengthened their research by deepening the detailed optimisation of earthwork as the scope of land work management for toll road construction [7]. The use of the latest technological approaches is growing, especially in the optimisation of earthwork management systems, for example, the use of case-based database reasoning [16]. This study examines the potential benefits of the record-keeping process and the historical data on road maintenance collected in the database, especially the decision-making process, which is then interpreted and used to model optimisation for subsequent decisions. This complements previous research by integrating case-based reasoning, eigenvector methods, and web technologies to use historical data and expert opinions in the field of road maintenance to create intelligent systems with a mathematical approach and utilise the capabilities of the cloud as a database.

2.3 Data mining

The understanding and deepening of the field of science have an important influence on the success of designing a DM algorithm. In recent times, DM has begun to be used in scientific engineering and civil engineering [17]. A database is only a set of data without meaning if it is not analysed using the right algorithm approach [18]. Furthermore, Fu also said that based on reviews conducted in recent years, DM's ability to grow in a particular domain is dependent on the number of researchers who continuously develop a particular algorithm. In simple cases, scholarship can help identify the right features to model the data. The preparation of a scientific database can also help design business goals that can be achieved using in-depth database analysis.

DM tasks are established based on the ability of DM to solve various problems through interpretation and other statistical operations on the data [19]. Depending on the type of pattern found, DM tasks are usually classified into two categories, namely predictive and descriptive. The predictive approach uses inference on the data to predict unknown values of the output variables, taking into account the known values of the input variables [20]. The descriptive approach characterises and summarises the various general properties of the data to improve the understanding and provision of extensive information. The utility of a DM task depends on the ability of the user to identify the initial problem and the purpose of completion.

3 Methods

Some tools, equations, algorithms and source code used to answer research questions, develop models, construct syntheses, and display modelling results will be described in detail in this section. To achieve accurate results from research, the work should be performed appropriately and systematically.

3.1 Data

Productivity data was mostly obtained from toll road business entities and the Toll Road Regulatory Agency. This data represents a historical record of earthwork projects, earthwork equipment productivity, and other important information. The earthwork equipment productivity data obtained from toll road business entities range from 2010 to 2017. Some of the data is not complete, but the DM approach can be used to estimate lost or biased data in the database. In addition to the information obtained from toll road business entities, data was also sourced from earthwork equipment standard specifications. The case studies in this work used data from the construction of the trans-Java toll road, as shown in Table 1.

Table 1. Data distribution for modelling.

No	ID	Segment	Model Development			
			Learning	Validation	Optimisation	GIS (next)
1.	CP	Cipali	√			
2.	PP	Pejagan Pemalang	√			
3.	SS	Semarang-Solo		√	√	√
4.	SK	Solo Kertosono	√			
5.	KM	Kertosono-Mojekerto	√			
6.	MS	Mojokerto Surabaya	√			
7.	SB	Serpong-Balaraja	√			

3.2 Allocation of earthwork equipment

The allocation of earthwork equipment determines the duration and cost of the construction required, so the optimal use and placement of the equipment are very important. The allocation of equipment should not only take into account the minimisation of time and cost of construction but also maximise the efficiency of the equipment. In turn, using equipment efficiently maximises a project's sustainability. The earthwork equipment to be allocated consists of trucks, excavators, spreaders, and compactors. Equipment allocation simulation was performed for one equipment plant that is considered to have the same organisation.

4 Discussion

This study developed a predictive model of earthwork equipment productivity using a DM approach, without any restrictions on the input data considered. Through consideration of the classification or regression approach, alternative evaluation steps may also be undertaken. For regression, the evaluation process is based on the difference between the observed value and the estimated value (error value). In general, the lower the error value, the better the prediction model of earthwork equipment productivity, where the error (value = 0) is the ideal value to be achieved.

In this model, three measurements were taken, namely MAD, RMSE and R^2. Models with low MAD and RMSE values and R^2 values close to unit values can be interpreted as models with high prediction accuracy. RMSE is more sensitive to extreme values compared to MAD because RMSE uses the squared value of the difference between the measurement results and the predicted model results. Compared to MAD, the RMSE for a model is likely to be larger. Comparing these error values for models will provide different perspectives on which to base model selection.

In the domain of scientific engineering, in addition to requiring a high degree of accuracy, the ability to interpret the modelling results is critical. The ability of a DM approach to interpret a dataset is strongly influenced by the power of the data-driven model for that purpose. When a black box DM approach is implemented with multiple regression (MR), artificial neural network (ANN) and support vector machine (SVM) algorithms involving complex mathematical expressions, then the data-driven procedure should be able to model the data. In this case, the interpretation of the model is done to obtain the input variable measurement for the productivity prediction model.

This model is evaluated with a confidence level of 95% according to the t-student distribution. All DM models with MR, ANN and SVM training algorithms use four input variable attributes. Table 2 presents the predictive capacities of all training outcomes, comparing their performance in terms of earthwork equipment productivity prediction scores based on MAD, RMSE, and R^2. This table shows that productivity values can be accurately predicted by each of the three DM models, especially the ANN and SVM models.

Table 2. Error metrics model DM.

Model	MAD	RMSE	R2
MR	0.72 ± 0.01	0.91 ± 0.02	0.54 ± 0.00
ANN	0.48 ± 0.02	0.62 ± 0.01	0.89 ± 0.02
SVM	0.57 ± 0.01	0.73 ± 0.01	0.77 ± 0.01

The table 2 shows the standard error and R^2 for each model developed. The DM model that uses the ANN algorithm has the smallest MAD and RMSE values as well as the highest R^2 value. The performances of the predictive models using ANN and SVM algorithms are acceptable and appropriate to be used in calculating road performance predictions because they have R^2 values greater than 0.70. In this research, the selected prediction model for the productivity of earthworks equipment was the DM model using the ANN algorithm.

The interpretation of the regression analysis used in DM (package rminer) provides a graphical interpretation tool consisting of a regression error characteristic (REC) curve, with the error tolerance illustrated on the x-axis and the percentage prediction value of road performance depicted on the y-axis. The resulting curve describes the error rate in the form of a cumulative distribution function (CDF). The error rate here is defined as the difference between the predicted f(x) and actual earthwork productivity at every point (x, y). The approach is also a squared residual $(y - f(x))^2$ or absolute deviation $|y - f(x)|$ based on error metric mapping.

The REC analysis results describe the effect of the main attributes that move dynamically. In the road performance prediction model with ANN, this attribute is the earthwork equipment plant consisting of trucks, excavators, spreaders, and compactors. Productivity increased following the allocation of prepared equipment. The overall changes in productivity values in the prediction model are illustrated in Figure 1.
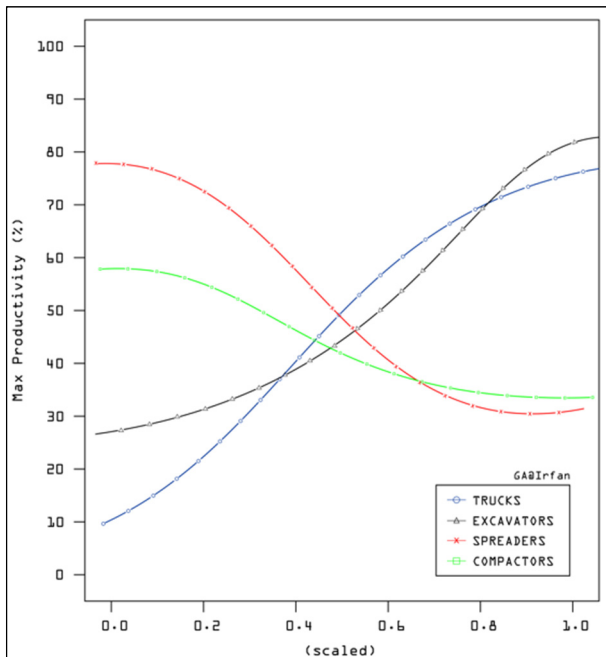

Fig. 1. REC curveꜱ for the prediction model.

The developed DM model can assess the contribution rate of each variable as well as the attribute that becomes the input data in the model. A parameter vector is selected in the DM model to explain that it is a uniform function and not parameters as in the parametric approach. The only condition for a uniform function is to generate a matrix of non-negative definite variance. There are several methods that can be used to predict hyperparameter values. The value of θ can be estimated in this DM using a cross-validation method. The hyperparameters used are H (2, 4, …, 10) and γ (2-15, 2-13, …, 23). This value produces the most precise model with optimal run time. Further models can be developed by trying other hyperparameters. The contribution of each attribute and dimension is its relative importance in modelling.

The Pareto solution approach is used to determine the DM-based optimisation model. The solution is structured to produce optimisation of the production value and the magnitude of the mechanical displacement cost. The optimisation results used as the basis for decision making will be illustrated in the application of the model. The toll road project data summarised in Table 1 was used as a simulation section. The optimisation was performed with various allocation equipment scenarios. The optimal earthwork equipment allocation program was chosen using the Pareto solution approach. The selected Pareto model used to construct the earthwork movement optimisation model can be seen in Figure 2.
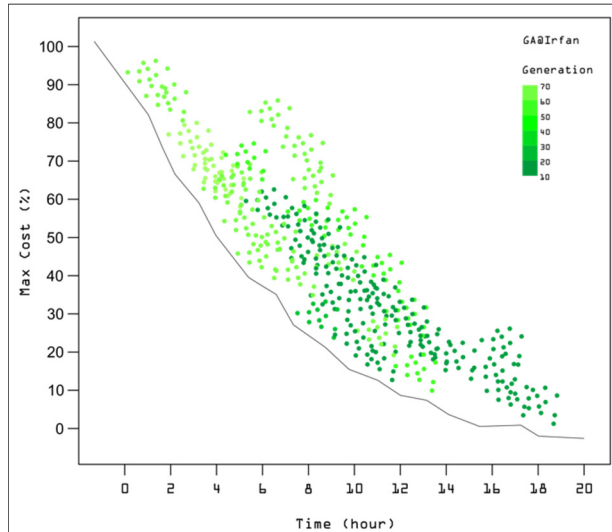
Fig. 2. Pareto front.

By using this choice model, this system is capable of achieving a high impact on both earthwork duration and project cost for a toll road project. In Figure 3, it can be seen that the level of work in each group of earthwork equipment in the form of the original distribution arrangement is not well structured, while the optimal distribution is quite well arranged. In the original distribution of equipment, the limited productivity of the excavator team (approx. 350 m^3/hr) resulted in the trucks, which have a much higher potential productivity of almost 2,000 m^3/h, being forced to wait for the material to be extracted before being able to transport it to a stockpiled and compacted area.
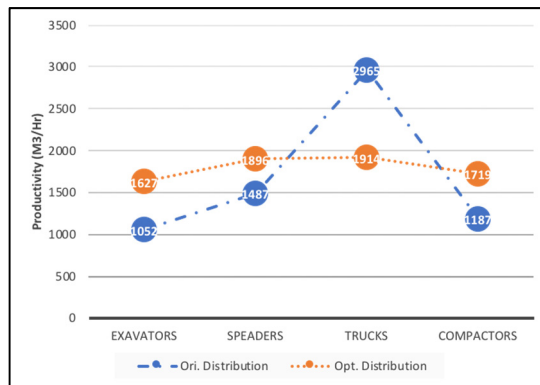

Fig. 3. Comparison between original and optimal equipment distributions.

Figure 4 is an illustration of the optimisation of the use of mechanical equipment via the Pareto approach. The optimisation is performed by utilising a genetic algorithm approach based on data mining [21]. In this figure, each point represents a viable distribution solution and optimal equipment for earthworks projects, evaluated in terms of the associated duration (in hours) and cost (maximum cost. The output system presents several solutions that correspond to optimal trade-offs between cost and duration, where the maximum sustainability is guaranteed in accordance with the methodology mentioned above. The completion of the earthwork depends on the conditions of the field, the weather and the presence of the equipment, so some flexibility in the model is needed.
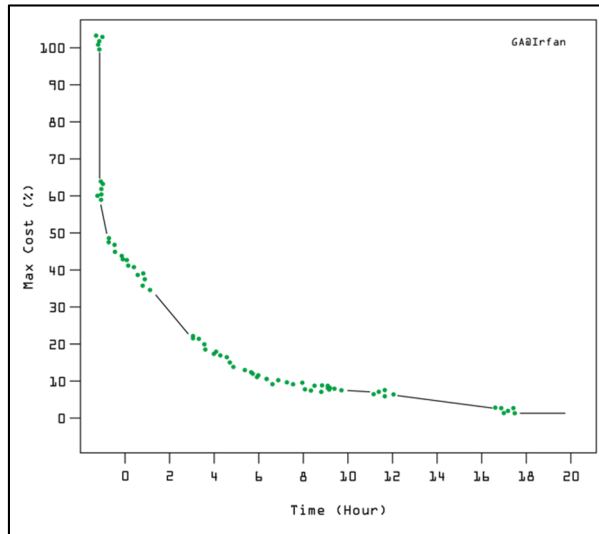
Fig. 4. The output of an optimisation run.

5 Conclusion

This paper presents a model for utilising big data on earthworks for a toll road construction project to obtain optimisation of equipment productivity. We began by arranging the allocation of equipment in each workgroup. The productivity of each arrangement of equipment is predicted by utilising DM techniques and, in particular, the ANN technique, a model with excellent predictive capacity for large data. Furthermore, we performed the optimisation using the Pareto approach with multiple generation options. With the Pareto approach, we obtained options for optimal allocation of equipment at minimal cost.

References

1. M. Berawi, T. Zagloel, P. Miraj, & H. Mulyanto, *Producing Alternative Concept for the Trans-Sumatera Toll Road Project Development using Location Quotient Method. Procedia Engineering*, **171**, 265-273, (2017)

2. A. Ansar, B. Flyvbjerg, A. Budzier, & D. Lunn, *Does infrastructure investment lead to economic growth or economic fragility? Evidence from China*. Oxford Review of Economic Policy, **32(3)**, 360-390, (2016)

3. N. Pradhananga, & J. Teizer, *Cell-based construction site simulation model for earthmoving operations using real-time equipment location data*. Visualization in Engineering, **3 (1)**, 12, (2015)

4. A. Rashidi, H. Nejad, & M. Maghiar, *Productivity estimation of bulldozers using generalized linear mixed models*. KSCE Journal of Civil Engineering, **18(6)**, 1580-1589, (2014).

5. C. B. Tatum, *Construction engineering research: Integration and innovation*. Journal of Construction Engineering and Management, **144(3)**, 04018005, (2018).

6. P. Cortez, *Data mining with neural networks and support vector machines using the r/rminer tool*. Advances in Data Mining: Applications and Theoretical Aspects, 10th Industrial Conference on Data Mining, **83**, Berlin, Germany: J In P. Pemer, editor, (2010)

7. M. Parente, A. G. Correia, & P. Cortez, *Artificial Neural Networks Applied to an Earthwork Construction Database*. In: Toll D, Zhu H, Osman A, et al (eds) Second Int.Conf. Inf. Technol. Geo-Engineering. IOS Press, Durham, UK, 200–205, (2014)

8. S. S. Lee, S. I. Park, & J. Seo, *Utilization analysis methodology for fleet telematics of heavy earthwork equipment. Automation in Construction*, **92,** 59-67, (2018)

9. P. Saha, & K. Ksaibati, *A risk-based optimisation methodology for pavement management system of county roads*. International Journal of Pavement Engineering, 1-11, (2015)

10. A. Alshibani, & O. Moselhi, *Productivity based method for forecasting cost & time of earthmoving operations using sampling GPS data*. Journal of Information Technology in Construction (ITcon), **21(3),** 39-56, (2016)

11. F. Vahdatikhaki, & A. Hammad, *Framework for near real-time simulation of earthmoving projects using location tracking technologies*. Automation in Construction, **42,** 50-67, (2014)

12. A. Sheikh, M. Lakshmipath, and A. Prakash, *Application of Queuing Theory for Effective Equipment Utilization and Maximization of Productivity in Construction Management*. International Journal of Applied Engineering Research, **11(8)**, 5664-5672, (2016)

13. A. A. Tsehayae, & A. R. Fayek, *Developing and Optimizing Context-Specific Fuzzy Inference System-Based Construction Labor Productivity Models*. Journal of Construction Engineering and Management, **142(7),** 04016017, (2016)

14. C. Koo, T. Hong, & S. Kim, *An integrated multi-objective optimization model for solving the construction time-cost trade-off problem*. Journal of Civil Engineering and Management, **21(3)**, 323-333, (2015)

15. Y. Pan, & L. Hou, *Lifting and parallel lifting optimization by using sensitivity and fuzzy set for an earthmoving mechanism*. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 0954407016660454, (2016)

16. J. –S. Chou, *Applying AHP-Based CBR to Estimate Pavement Maintenance Cost*. Tsinghua Science and Technology, 114-120, (2008)

17. S. Ahn, P. Dunston, A. Kandil, & J. Martinez, *Process Mining Technique for Automated Simulation Model Generation Using Activity Log Data*. In Computing in Civil Engineering, 636-643, (2015)

18. T. C. Fu, *A review on time series data mining. Engineering Applications of Artificial Intelligence*, **24(1)**, 164-181, (2011)

19. A. A. Freitas, *Data mining and knowledge discovery with evolutionary algorithms*. Springer Science & Business Media, (2013)

20. X. Wu, X. Zhu, G. Q. Wu, & W. Ding, *Data mining with big data*. Knowledge and Data Engineering., IEEE Transactions on, **26(1)**, 97-107, (2014)

21. A. I. Rifai, S. P. Hadiwardoyo, A. G. Correia, P. Pereira, & P. Cortez, *Data Mining Applied for The Prediction of Highway Roughness under Overloaded Traffic*. International Journal of Technology **5**:751-76, (2015)