Data mining applied for national road maintenance decision support system

Andri Irfan Rifai[1,*], *Susanty* Handayani[2], and *Ronal* Al Rasyid[3]

[1]Directorate General of Highway, Ministry of Public Works & Housing, Indonesia
[2]Greater Jakarta Transportation Authority, Ministry of Transportation, Indonesia
[3]Jasa Marga Jalanlayang Cikampek Toll Road Company, Indonesia

Abstract. National roads are one of the main networks of a country's transportation system. To maintain the performance level of national roads requires a well-structured pavement management system (PMS). The decision support system (DSS) is inseparable in the modern PMS, which required the development of a new approach for the DSS in support of national road network maintenance. The proposed model integrates data mining (DM) and geographical information system (GIS) to construct a simple DSS. DM is used to developed road maintenance optimization models, and then integrated with DSS with the help of GIS as an interface application. Historical data on the national road network in West Java, Indonesia is used as a case study. Examples show that the proposed model can determine a decision support solution efficiently. In addition, a user-friendly computer interface is developed so that PMS stakeholders can plan pavement maintenance simply and effectively.

1 Introduction

The pavement management system (PMS) is carried out continuously, starting from design, planning, development, operation, maintenance, to the control stages. All stages of the pavement management system cycle have an equally important role and have a significant influence in maintaining road performance if implemented sustainably [1]. PMS is influenced by the nature and character of the pavement structures that can be patterned with various data approaches and other historical records. Higher pavement maintenance standards are currently required, in line with increasing road length, increased traffic volume, and other traffic space requirements [2].

The primary data required in the pavement management system is road performance data. Underperformance of the road can be reduced in proportion to the increasing age of pavement and the traffic load [3]. In general, the age of the pavement is determined based on the cumulative equivalent standard axle (CESA) estimated across the road pavement, calculated from the start of the construction and operation until the road pavement is categorized as damaged. Decreases in the overall road performance follow the function of volume increases and traffic load, changes in environmental conditions, and other conditions [4].

---

[*] Corresponding author: andrirfan@yahoo.com

Degradation of road performance does not take place in real time but gradually follows the time function. The speed and shape of performance changes have certain patterns and trends. Data collection in big data is absolutely necessary to produce a good and sustainable pattern [5]. New engineering approaches and the latest technology need to be used so that collected data can be utilized in a structured and scalable manner to support a better pavement management system through accurate interpretation and prediction of data. DM is one of the most recent approaches capable of performing accurate interpretations in predicting road service levels [6]. Regarding this matter, [7] has developed a road maintenance optimization model using information technology supported by a combination of a mathematical approach and genetic algorithm optimization.

Finally, a proven PMS is a system capable of providing a tool for users and decision makers to be able to understand and easily use. Coutinho-Rodrigues et al. [8] write that the easier a system is to use, the more the system can support the performance of the organization. In this regard, it is necessary to develop a DSS that is simple and easy to understand to optimise the existing pavement management system. The GIS approach is believed to be able to help simplify the DSS developed, as GIS contributes to the provision of a systematic method for the control of the maintenance and rehabilitation (M&R) process for paved networks [9].

Based on the description of some existing model concepts, the GIS approach is the main choice in explaining how optimization and prioritization are carried out. The development of this GIS model is expected to be an alternative complement to some of the existing concept models and other tools. The result of this GIS model development is expected to provide repair solutions for the improvement of the optimization model and the preparation of the DSS concept of road maintenance more comprehensive.

2 Literature review

DSS is needed to provide support to decision-makers in reaching the right policies. Likewise, in the issue of road maintenance, an approach is needed that can simplify the tools in making this decision. Maintenance optimization-based DSS is feasible for developing reasonable and consistent pavement maintenance [10]. In the process of pavement maintenance, a support system that can provide tools in decision-making is necessary. The decision support tools have neatly tiered and structured properties.

Some of the decisions that can be arranged into a system are strategic planning decisions, management control decisions, operational control decisions, and operational performance decisions. Various series of management activities, of course, require a system that can assemble all the node activities. Integration with existing systems such as computerized maintenance management systems and GIS is seen as the largest challenge for developing and using decision-support tools in the area of asset management [11].

DSS is often used in different contexts related to decision-making and refers to our support capabilities in making decisions; thus decision support is related to human decision-making. Generally, decision-making consists of three main components: intelligence, design, and choice [12].

The decision support model has been used in various fields, including pavement management. Moreover, plans and projects often impact multiple and contrasting interests in a complex institutional setting that result from decision-making processes involving several actors, both public and private [13]. Decision-making processes in infrastructure industries have increased due to the high level of inherent uncertainty. This is illustrated by the increasing complexity of the needs of decision support models, tools, and systems to assist the process. The decision-making model should also be applied to road infrastructure

investment. It is impossible to know exactly how accurate a particular investment decision is; therefore, DSS tools can assist in improving investment choices.

DSS is an interactive system-based computer system that helps decision-makers by utilizing data and models to solve unstructured and ambiguous problems. Network level infrastructure maintenance decision-making is a multi-factor and multi-criteria problem [14], which serves the needs of all levels of management decisions but is preferred for strategic decision support. The DSS function revolves around the scope of collecting and presenting information, and extrapolating, inferencing, and elaborating complex modelling. While the information system is based on the structure of analysis and decision support generated in the form of a unique answer, DSS emphasizes the importance of interactive activities and direct involvement of end users. Based on the feedback mechanisms inherent in the DSS, its use can improve the quality of decision-making and can optimize the limitations of dynamic moving resources. Furthermore, decision-making requires other supporting instruments besides technical approaches; economic, environmental, and social approaches require the same attention. Thus, in implementing effective management processes, the practical implication for making a decision is the dynamics of their collaborative networks. [15].

The role of the DSS is to help answer the question: "what is, what would, and what if". Without DSS it is quite difficult to reach the right decision. The problem of subjective presentation is sufficiently understood by the decision maker to be confronted with a statement that is inconsistent with the facts. Therefore, decision-making is a process based on knowledge and not just a "black box". To extract knowledge from hidden data, a tool is required that can interpret information with structure. DM extends the possibilities for decision support by discovering patterns and relationships hidden in the data and, therefore, enabling an inductive approach to data analysis [16]. The DSS model can be expected to have high accuracy if the approach model implanted in the system has high accuracy as well. All subsystems jointly or alone contribute to the various functions of the sub-system.

3 Method

The research method in this study consists of several components that can be synchronized with an input-process-output-process, which is the character of a system. The condition of road pavement is obtained from Directorate Genera Highway (DGH). Performance Index (PI), and road maintenance data are the main components of processes that use inputs to optimize road maintenance models. Pavement deterioration and maintenance models are the key components of the processes that use the inputs to optimize the model of pavement maintenance. Optimal road maintenance activities with minimum measures and costs must still meet minimum standards of job implementation. The iteration process of selecting maintenance actions starts from the easiest maintenance actions required each year to the end of the analysis period. The output of this model is a data mining-based DSS concept with PI prediction components and maintenance optimisation. The validity of a deterioration model is based on the accuracy and reliability of its data. This step entails taking several sources of data and combining them to create a comprehensive dataset.

The pavement condition data is mostly obtained from IIRMS. This data is a historical record of road conditions, road performance, and other relevant information, including roughness, cracks, ruts, potholes, Average Annual Daily Traffic (AADT), and Equivalent Single Axle Load (ESAL). Road condition data obtained from IIRMS dates from 2000 to 2017. Some of the data are incomplete, but the DM approach can be used to estimate lost or biased data in the database. Data used in this study, other than that obtained from IIRMS, is sourced from a report of Hawkeye vehicle surveys.

3.1 Data entry and acquisition

Data is one of the critical components in GIS; the methods available for adding or obtaining data are paramount. Methods to achieve this are importing digital information available in a compatible format, using a global positioning system (GPS) device, and digitalizing from analog data.

On the one hand, the development of information technology with the emergence of digital information and database has made GIS access easier, especially internet-based. On the other hand, compatibility between software is increasingly common, allowing one to convert data originating in one type of software to use in other formats, Formats such as CAD (e.g., DWG, DXF), vectorial, and raster data widely used in commercial GIS (ARC/Info, ARC/View, Intergraph MGE, etc.) and general image data (e.g., tiff, bmp, etc.) are some examples of data that can be added directly to most GIS software.

It is widely known that a method for GIS data acquisition is with the use of GPS. The method is performed by transmitting signals to multiple satellites and using triangulation to determine positions and altitudes with low error margins (e.g., under one meter). GPS can be connected to GIS for various purposes such as mapping, determining coordinate handling, and others. The data used to perform design simulation is pavement management data on Java Island, while for the simulation interface a map of the West Java region is used.

3.2 Data mining approach

In the development of this model, the researcher used a DM road-based performance prediction approach by using data from Java Island, Indonesia. Data is divided by province for calibration, learning, test, and validation purposes. Validation of road conditions and coordinate details were collected by direct data retrieval with Hawkeye in 2017. The road performance prediction model, in the form of PI with a DM approach without any assumption of limitation by considering input data is used to learning stages.

In this research, three DM techniques were trained using the previously described dataset to make predictions. The two DM algorithms are obtained by SVM and ANN models that have the same appearance. The performance of this model is confirmed by the values of R^2, MAD, and RMSE. Modelling results are presented with 95% confidence intervals in accordance with the t-distribution. Furthermore, SVM is adopted as a reference algorithm because it has a fairly high degree of accuracy with the number of iterations (20 iterations).

The DM technique, also known as association rule mining, has the purpose of finding associative rules between a combination of items. The importance of an associative rule can be identified by two parameters; the support, the percentage of the combination of attributes in the database, and the confidence, namely the strength of the relationship between attributes in the associative rules. The algorithm used in this research, paradigm generation and testing, makes candidate combinations of possible attributes based on a certain rule and is then tested. The combination of eligible attributes is called the frequent itemset, which is then used to create rules that meet the minimum confidence requirements.

3.3 Network implementation and analysis

Among the various GIS capabilities, network analysis is an important step in this research. The limited analysis is performed on vectorial data because its use may represent the availability of a road network, defined as a series of interconnected features, representing potential routes, vehicle characteristics, and road maintenance activities. At this stage, accurate connectivity and characteristics are more important than geographical views. The ability of the QGis Lisboa version is expected to perform the various minimum functions

required. The spatial approach in the pavement management system is able to simplify the extent of the area to be monitored by the system. Mapping, scribing, and area coverage presentations by conventional methods can lead to less accurate decision-making. This open source application has the capability and facilities to integrate with multiple systems. With expected results, DSS users no longer need additional applications when changing functionality.

The information system application software used in this research is implemented on a platform with Microsoft Windows 10 configuration, QGis Lisboa Version, MapServer, PostgreSQL, and PostGIS. While the modeling hardware implemented on the computer with computer specifications Core i7 with SDRAM memory / DDR 8 GB.

4 Discussion

The concept of DSS module development in this study aims to provide an overview of the implementation of integration between DM, optimization, and GIS. Using the GIS approach with open source software, we expect to make DSS development more interesting. The developed GIS module is fully integrated with the data mining presented in this module and contains many features in addition to the graphic displaying and report generating functions. The functions developed for the GIS module include the following capabilities: this concept is modelled as an interface application that is able to receive input in the form of numeric and coordinates and then provide output in text, numeric, tabular, and graphics forms.

4.1 Polygon feature

The Add-Relate method provided by the ESRI GIS Component is available for displaying polygon map features. This method creates a relationship between the graphics information in a map layer and the record sets generated from needs analysis results. Figure 1 shows the averaged IRI results in six provinces. The province can be associated with a polygon map and displayed using the Add-Relate method. The engineer can query needs analysis performance rating results (attribute) by clicking on the map. The GIS Base Map Preparation Function then automatically extracts the performance ratings of all the pavement projects associated with each province, computes the average performance ratings for each IRMS District, and creates a "Composite Rating in Year XXXX" field on the tab (menu) shown on the right side of the figure. The program also automatically creates a GDOT District map layer table shown on the left side of this figure.
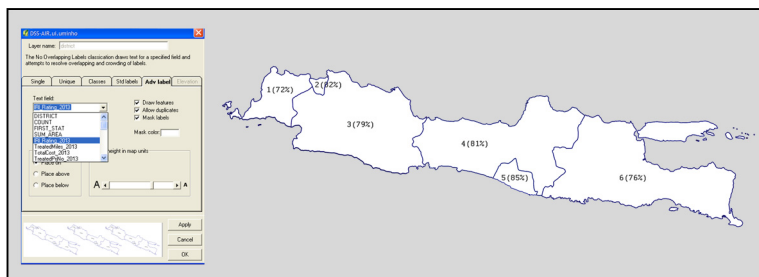

Fig. 1. Add-Relate method.

There is a common "District" field on both of the record-set tables, which uniquely identifies each record-set on the map layer and the results table. The Add-Relate method uses this common field to create a joint between the map layer and the results table, and creates a

new record-set, which contains all the records from the map layer together with all their attributes.

4.2 Map layer

The five basic map layers below are used in the GIS module. ANALYSIS RESULTS: Contains detailed project-level results, including project ratings in each future analysis year, treatment methods and costs, AADT, and spatial location information, such as SegmentNo, NetworkNo, ProvinceNo, Sta. From and Sta. To, District, Office, and others. It is created by the MDS method. STATEROUTE: This layer is provided by the Integrated Indonesian Integrated Road Management System (IIRMS). After data integration, the layer contains the complete information on state highway routes in Java Island. DISTRICT: Contains the detailed district information of Java Island. PROVINCE: Contains the IIRMS Province boundary information and the IIRMS District-level needs analysis results. NETWORK: Contains the Network boundary information.

The five basic map layers include most of the information generated from the maintenance model results that can be displayed on GIS maps. Results from the Project-level Analysis Module related to individual project information are displayed on the ANALYSIS RESULTS layer; see figure 2. Results from the Network-Level Analysis Module related to State Congressional Districts, provinces, and states are displayed respectively on the DISTRICT and PROVINCE map layers.
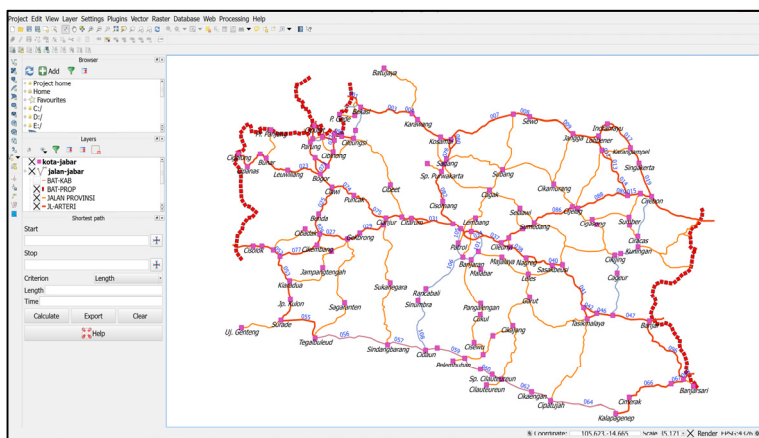

Fig. 2. GIS base map.

4.3 Visualization

GIS provides a powerful visualization and mapping capability, which is useful for pavement rehabilitation needs analysis. In order to facilitate the decision support on multi-year rehabilitation needs analysis, several advanced functions have been developed. The design concept of these functions is to facilitate pavement rehabilitation needs analysis using GIS visualization. The potential uses of the visualization and mapping functions are: Visualize spatial and temporal treatment strategies on one map; Identify projects with abnormal pavement conditions; Investigate detailed historical information and needs analysis results of the interested project dynamically and interactively; Make comparison among different jurisdictions; Monitor routes not surveyed.

Several potential GIS applications are presented in the following to demonstrate the potential uses of these functions to facilitate the decision-making for planning pavement rehabilitation needs activities.

4.4 Interactive analysis of maintenance

To facilitate decision support on pavement maintenance needs analysis, a maintenance scenario analysis is developed with GIS capabilities, including visualization, spatial identification, and analysis, with maintenance optimization model including barriers to segment level and road performance prediction, determination of improvement and priority maintenance work. The interactive pavement scenario maintenance function based on maps has been developed to make it easier for decision-makers to develop and evaluate different improvement scenarios intuitively and directly on GIS-based maps.

Some potential GIS applications are presented below to demonstrate the potential use of functions in facilitating decision makers in the planning of pavement rehabilitation activities. Information from the maintenance segment, not only the condition of the previous pavement but also the rehabilitation information, is important in helping the engineer make a good assessment of the pavement rehabilitation plan. Engineers can retrieve previous information with GIS maps dynamically and can be connected directly to GPS devices to import the latest data. Figure 3 shows an example.
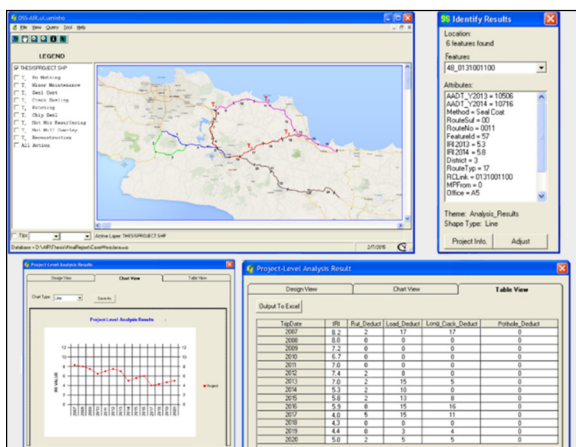

Fig. 3. Visualization of GIS module.

Using the above module, the user can select a special segment on the GIS-based map. In figure 3, there is sufficient information, including AADT, the year of improvement, rating, and repair methods on the selected segment shown in the table. By clicking on the "project info" button, various segment information, such as the level of the segment, the difference in the value of the segment, and AADT, can be retrieved from the database. Their relation to the road performance prediction information derived from the maintenance model results and stored in the database, can also be displayed in graphical or table formats.

This module is structured for users to understand that with technical knowledge, the management strategies generated by this DSS program can be improved. Good decisions can reduce the costs of mobilization of segments and congestion caused by the construction of two segments in a separate year, even reducing total construction costs. The module is considered to be an additional tool for stakeholders in implementing a pavement management system. The system was compactly designed for the road performance prediction model, optimization model, and DSS process and can be simply understood in an integrated manner.

5 Conclusion

The DSS concept was developed by integrating DM to develop PI prediction models and GIS to develop DSS interfaces, all capable of displaying simple interface applications and providing convenience to stakeholders to implement pavement management systems with simple steps. The DM approach that has been adapted to the needs of the road maintenance management system that has a wide area of coverage can simplify the constraint. The interface concept developed in this research is quite simple and flexible, and therefore can be developed in accordance with local needs.

References

1. T. Ding, L. Sun, & J. Chen, Optimal Strategy of Pavement Preventive Maintenance Considering Life-Cycle Cost Analysis. Social and Behavioral Sciences, **96**, 1679 – 1685, (2013)

2. R. R. Almuhanna, H. A. Ewadh, & S. J. Alasadi, Using PAVER 6.5. 7 and GIS program for pavement maintenance management for selected roads in Kerbala city. Case Studies in Construction Materials, **8**, 323-332, (2018)

3. P. Y. Wang, Y. Wen, K. Zhao, D. Chong, & A. S. Wong, Evolution and locational variation of asphalt binder aging in long-life hot-mix asphalt pavements. Construction and Building Materials, **68**, 172–182, (2014)

4. R. Hass, Good Technical Foundations Are Essential for Successful Pavement Management. Maintenance and Rehabilitation of Pavements and Technology Control. Guimaraes: Universidade do Minho-Escola de Engenharia. pp. 3-28, (2003)

5. M. Varela-González, M. Solla, J. Martínez-Sánchez, & P. Arias, A semi-automatic processing, and visualization tool for ground-penetrating radar pavement thickness data. Automation in Construction, **45**, 42-49, (2014)

6. 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, 5:751-761, (2015)

7. A. Rifai, S. P. Hadiwardoyo, A. G. Correia, & P. Pereira, Genetic Algorithm Applied for Optimization of Pavement Maintenance under Overload Traffic: Case Study Indonesia National Highway. Applied Mechanics and Materials ISSN: 1662-7482, **845**, Trans Tech Publications, Switzerland, pp 369-378, (2016)

8. J. Coutinho-Rodrigues, A. Simão, & C. H. Antunes, A GIS-based multicriteria spatial decision support system for planning urban infrastructures. Decision Support Systems, **51(3),** 720-726, (2011)

9. S. M. Bazlamit, H. S. Ahmad, & T. I. Al-Suleiman, Pavement Maintenance Applications using Geographic Information Systems. Procedia Engineering, **182**, 83-90, (2017)

10. F. Wang, Z. Zhang, & R. Machemeh, Decision-making problem for managing pavement maintenance and rehabilitation projects. Transportation Research Record: Journal of the Transportation Research Board, **1853**, 21-28, (2003)

11. M. Di Sivo, L. Daniela, Decision-support tools for municipal infrastructure maintenance management. Procedia Computer Science, **3**, 36-41, (2011)

12. Z. Turskis, G. Ambrasas, D. Kalibatas, & A. Barvidas, Multiple criteria decision support system model for construction works technological cards designing. In The 9th

International Conference "Modern Building Materials, Structures and Techniques, **1,** p. 3 (2007)

13. E. Cascetta, A. Carteni, F. Pagliara, & M. Montanino, A new look at planning and designing transportation systems: A decision-making model based on cognitive rationality, stakeholder engagement, and quantitative methods. Transport policy, **38**, 27-39, (2015)

14. H. Li, F. Ni, Q. Dong, & Y. Zhu, Application of analytic hierarchy process in network level pavement maintenance decision-making. International Journal of Pavement Research and Technology, (2017)

15. Q. Wen, M. Qiang, & P. Gloor, Speeding up decision-making in project environment: The effects of decision makers' collaboration network dynamics. International Journal of Project Management, (2018)

16. S. Khademolqorani, & A. Z. Hamadani, An adjusted decision support system through data mining and multiple criteria decision making. Procedia-Social and Behavioral Sciences, **73**, 388-395, (2013)