We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

5,500
Open access books available

136,000
International authors and editors

170M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Data Mining Applied for Community Satisfaction Prediction of Rehabilitation and Reconstruction Project (Learn from Palu Disasters)

Andri Irfan Rifai

Abstract

Natural disasters can occur anytime and anywhere, especially in areas with high disaster risk. The earthquake that followed the tsunami and liquefaction in Palu, Indonesia, at the end of 2018 had caused tremendous damage. In recent years, rehabilitation and reconstruction projects have been implemented to restore the situation and accelerate economic growth. A study is needed to determine whether the rehabilitation and reconstruction that has been carried out for three years have met community satisfaction. The results of further analysis are expected to predict the level of community satisfaction for the implementation of rehabilitation and other reconstruction. The method used in this paper is predictive modeling using a data mining (DM) approach. Data were collected from all rehabilitation and reconstruction activities in Palu, Sigi, and Donggala with the scope of the earthquake, tsunami, and liquefaction disasters. The analysis results show that the Artificial Neural Network (ANN) and the support vector machine (SVM) with a DM approach can develop a community satisfaction prediction model to implement rehabilitation and reconstruction after the earthquake-tsunami and liquefaction disasters.

Keywords: Community Satisfaction, Data Mining, Disasters, Reconstruction, Rehabilitation

1. Introduction

The Palu earthquake, Indonesia, on September 28, 2018, caused severe damage with a reasonably broad impact. At the time of this writing, the atmosphere of grief and trauma of the people affected directly and indirectly began to disappear. The earthquake has a complete phenomenon in the movement of faults, tsunamis, landslides, and liquefaction events. Simultaneous liquefaction in several locations is unique in the world. This liquefaction phenomenon has received attention from the people in the world because the mudflow event during liquefaction has devastated infrastructure and housing on a massive scale [1].

Palu City and its surroundings based on topographic, geological, and seismological conditions can suffer damage due to earthquakes, including secondary disasters

(tsunami, liquefaction, and cliff landslides). The earthquake in Palu on May 20, 1938, with a magnitude of 7.6 SR, was the previous incident with many fatalities. Studying, analyzing, and estimating all the supporting factors and the potential for disasters of such magnitude, the government needs to empower all components of society. The role of stakeholders in providing thoughts and recommendations is not accurate. Before and after an earthquake disaster occurs, they are better prepared psychologically and physically to reduce the impact of the disaster [2].

After a disaster with a significant impact, as mentioned above, various parties immediately carried out rehabilitation and reconstruction work, one of which was in transportation infrastructure. There are rehabilitation and reconstruction works on several roads, handling roads affected by liquefaction, including drainage systems, construction of retaining walls, construction of bridges, maintenance of bridges, and construction of access roads to permanent residences for disaster victims. According to its stages, the implementation of the rehabilitation and reconstruction was carried out, starting from recovery, trauma healing, permanent planning up to the overall reconstruction. The trauma healing stage is the starting point for the rehabilitation and reconstruction directly related to the community [3].

The implementation of rehabilitation and reconstruction due to natural disasters has not been completed yet. In early 2020 the Palu area could not avoid the non-natural disasters that plagued the world as a whole, namely the COVID-19 pandemic. This condition adds to the pressure to complete all stages of rehabilitation and reconstruction, especially work productivity which is directly impacted by restrictions on the labor movement. The decline in performance was mainly due to limited employee interactions with concerns and the potential risk of being exposed to the coronavirus. Covid-19 is transmitted by shedding droplets when an infected person coughs or exhales. Then, the released droplets will fall on nearby objects and surfaces, thereby polluting the surrounding environment [4].

Mitigation management and natural disaster recovery are an inseparable series of activities, starting from planning, mitigation, trauma healing, rehabilitation, and reconstruction, to socio-cultural recovery of the community. The speed and accuracy of planning play an essential role in achieving the success of post-disaster management. A thorough understanding and mapping are required in determining the plan that can be implemented appropriately in the field. Planning and implementation of work must consider the latest conditions taking into account the potential for recurring disasters. A thorough and well-targeted evaluation is required to ensure that the rehabilitation and reconstruction process runs according to the community's expectations. One of the evaluations that can be done is to measure community satisfaction at the job site. Because community satisfaction is one of the essential things in measuring the success of rehabilitation and reconstruction, the valuable experience from this disaster incident can be developed by a community satisfaction prediction model. The model that is built is expected to be an improvement step in the process of implementing rehabilitation and reconstruction in other activities.

2. Literature review

This section describes the literature review by conducting an integrated study of various information collected from library sources to provide a background for scientific development in rehabilitation and reconstruction. If necessary, comments and current knowledge trends will be included to show that the development of this knowledge can be included in the development of professionalism. In several sections, there is further information presented in different forms in implementing

post-disaster rehabilitation and reconstruction. All information obtained from this literature review is used as a background to understand community satisfaction.

This paper will discuss about community satisfaction using a data mining approach. It is hoped that data mining can interpret and predict the data collected pre-during-post rehabilitation and reconstruction after the earthquake, tsunami, and liquefaction disaster. The use of data mining is believed to be able to provide a new approach in determining a better satisfaction level for the implementation of similar disaster management.

2.1 Disaster vulnerability

Apart from being famous for its wealth and natural beauty, Indonesia is also a country that is prone to disasters. This condition is because Indonesia is in a dynamic volcanic area and continental plates. This position also causes the shape of Indonesia's relief to varying widely, from mountains with steep slopes to gently sloping areas along very long coastlines, all of which are susceptible to landslide, flood, abrasion, and tsunami hazards. Various hydrometeorological conditions sometimes threaten flooding and landslides, hurricanes or tornadoes, drought-related forest fires, etc. Another threat is disasters caused by various technological failures.

The condition of Indonesia with a reasonably high risk of natural disasters such as Sulawesi Island is a complex area. The location of the Sulawesi is a meeting place for three large plates. The plate is the Indo-Australian Plate moving north, the Pacific Plate moving west, the Eurasian Plate moving south-southeast, and the smaller plate, the Eurasian plate, which moves south-southeast, and the smaller plate, the Philippine Plate. Sulawesi, a young island in Indonesia, is located where subduction and collisions are still active. Based on existing rock blocks, the island of Sulawesi can be divided into three parts of the geological area. The first is West Sulawesi, where tertiary deposits and magma rocks are the dominant parts. Second, Central and Southeast Sulawesi mainly consisting of rocks from the early Cretaceous era. Thrid, East Sulawesi ophiolitic nappe covered Mesozoic and Paleozoic era sedimentary rocks [5].

Palu City is one of the capital cities in Sulawesi, which has a high risk of disaster. Palu was also passed by a significant fault that divides the city firmly on the surface. This fault is often referred to as the Palu-Koro fault, originally called the Fossa Sarassina fault. All geologists and geophysicists who are familiar with the Palu-Koro fault agree that this fault is active. An active fault will experience an earthquake at the exact location of the period. Several studies show repeated earthquakes for hundreds and thousands of years [6]. These faults are thought to have caused the history of earthquakes in the area to be quite long. The history of earthquakes in central Sulawesi has been recorded since the 19th century. Several major earthquakes with a sufficiently large record were in 1968 with 6.7 SR, 1993 at 5.8 SR, and 2005 at 6.2 SR. Meanwhile, the tsunami occurred in 1927 in Palu Bay with a wave height of 15 m, 1968 in Malaga as high as 10 m, and 1996 in Simuntu Pangalaseang as high as 3.4 m [7].

This condition causes Palu's vulnerability to earthquakes to be very high. The studies about earthquake vulnerability by conducting a microtremor test in Palu City based on the earthquake's epicenter from the United States Geological Survey (USGS), magnitude 6.3, which occurred on January 23, 2005 [5]. Microtremor survey to estimate the distribution of solid earthquake vibrations. From the survey, the peak acceleration, velocity, and earthquake susceptibility index were obtained. From these observations, it can be concluded that Palu City has soil conditions with shear wave velocity Vs. < 300 m/s. The peak acceleration can reach more than 400-gal, resulting in significant damage to the building. From microtremor research, it

is found that the vulnerability index in hilly areas is low and vice versa. The earthquake vulnerability index in the alluvium area is very high.

2.2 Rehabilitation and reconstruction

Rehabilitation is the repair and recovery of all public or community services to an adequate level in post-disaster areas. The main target of rehabilitation is to normalize or run fairly all aspects of government and community life in post-disaster areas. Rehabilitation is carried out by improving the environment in the disaster area, repairing public infrastructure and facilities, and providing assistance for community housing repairs. Rehabilitation activities also include socio-psychological recovery, health services, reconciliation and conflict resolution, socio-economic and cultural recovery, restoration of security and order. Furthermore, several other main activities that should not be neglected are restoring government functions and public services [8].

The implementation of rehabilitation includes physical repair activities and restoration of non-physical functions. Rehabilitation activities are carried out in areas affected by the disaster and other areas where it is possible to become target areas for rehabilitation activities. Rehabilitation activities must pay attention to building construction standards, social conditions, customs, culture, and economy. Repair of public infrastructure and facilities is an activity to repair public infrastructure and facilities to meet the transportation, smooth economic activities, and the socio-cultural life of the community [9].

Socio-economic and cultural recovery is part of the rehabilitation phase, aimed at helping communities affected by disasters to restore their social, economic, and cultural conditions to pre-disaster conditions. Social, economic, and cultural recovery activities are carried out by helping communities to revive and reactivate social, economic, and cultural activities through advocacy and counseling services, activity stimulant assistance, and training. This rehabilitation activity does not only concentrate on physical work but focuses more on social recovery. So the success of rehabilitation is not only measured by the recovery of physical conditions and infrastructure, but rather by the recovery of all community activities [10].

The next stage after or simultaneously with post-disaster management rehabilitation is reconstruction. In terms of handling reconstruction, a proper reconstruction process is needed, based on sound planning, so that it is right on target and orderly in the use of funds. It can increase community resilience to the threat of disasters in the future through disaster risk reduction efforts. A good post-disaster reconstruction process must recover community conditions, both physically, mentally, socially, and economically, and reduce vulnerability to disasters, not exacerbate existing vulnerability conditions that lead to disasters. For the reconstruction process to run well, it is necessary to involve non-governmental organizations and the general public [11]. The objective was to ensure the reconstruction process was planned on time, on quality, and budget, and following its objectives.

The reconstruction objective is to permanently rebuild part or all of the physical and non-physical facilities and infrastructure, along with the entire institutional and service system damaged by the disaster, so that conditions are restored. Their functions can run well, and the community can be better protected. From various catastrophic threats [12]. Resource mobilization, including human, equipment, material, and financial resources, is carried out by considering the available resources. Human resources who understand and have professional skills are indispensable in all post-disaster rehabilitation processes and activities. Resources in the form of equipment, materials, and funds are provided and ready to be allocated to support the rehabilitation and reconstruction process.

2.3 Community satisfaction

Monitoring of post-disaster rehabilitation and reconstruction is required to monitor disaster recovery processes and activities continuously. The steering committee and government elements carry out monitoring of rehabilitation and reconstruction activities. It may involve planning agencies at the national and regional levels as an overall ingredient in the implementation of rehabilitation [13]. Each rehabilitation program must meet specific achievement indicators, mainly so that each component of public infrastructure and facilities can function adequately again to support the resumption of the social and economic life of the people in the disaster area.

Disaster management activities are an inseparable series. One of the rehabilitation and rehabilitation phase implementations is an activity that must be linked to other stages. In this understanding, rehabilitation and reconstruction relate to the pre-disaster and emergency stages and trauma healing. The whole series of activities can be successful if each stage is carried out with strict monitoring and control. Therefore, disaster management should not be positioned as a goal but to achieve the efficiency and effectiveness of disaster management as a whole [14]. This condition is a necessity that obliges stakeholders to ensure that the planning, preparation, post-rehabilitation, and reconstruction stages are carried out under sound management principles.

In the rehabilitation and reconstruction phase, it is necessary to consider the available local resources to meet various implementation needs. Human resources who understand and have professional skills are indispensable in all post-disaster rehabilitation processes and activities. In addition, resources in equipment, materials, and funds are needed and are ready to be allocated to support the rehabilitation process [15]. Rehabilitation and reconstruction activities involving local communities can indirectly assist the community to revive social, economic, and cultural activities. It is hoped that the active involvement of the community in rehabilitation and reconstruction will make the community feel recognized as part of the community and ensure that community expectations are appropriately fulfilled.

The various steps taken during the rehabilitation and reconstruction phase must be ensured that they have met the community's needs or have not. In its stages, a community satisfaction survey is needed in connection with some of the above. This is a comprehensive measure of the level of community satisfaction with the quality of rehabilitation and reconstruction services provided by public service providers [16]. It is necessary to conduct a survey to determine the weaknesses of each indicator of public services. In addition, it can be used to determine the performance of the rehabilitation and reconstruction that has been carried out [17].

2.4 Data mining

Currently, soft computing methods are carried out by mimicking processes found in nature, such as the brain and natural selection [18]. Soft computing techniques make it possible to perform data processing to reduce uncertainty, imprecision, and ambiguity. In the mid-early 1960s, a new branch of computer science began to attract the attention of most scientists. This new branch, referred to as artificial intelligence (AI), can be defined as the study of how making computers drive the quality of people's work better. The AI approach encourages the development of soft computing in various fields, one of which is the development of data mining.

The development of the information technology industry is speedy, and knowledge in data collection is proliferating. Large databases are not a problem if they can take advantage of computer technology with various primary and supporting applications. All data collected and stored in a suitable database can be

precious knowledge (for example, trend models, behavior models) that can support decision-making and optimize action [19]. Classical statistics have limitations for performing large amounts of data analysis or complex relationships between data variables. The solution for this problem and its limitations is to develop computer-based data analysis tools with more excellent capabilities and are automatic [20]. With the development of semi-automatic approaches in various fields of science, in recent decades, there has been an increase and across disciplines, such as AI, statistics, and information systems. This field is formally defined as knowledge discovery from the database (KDD). That in its development, KDD is increasingly known as DM [21].

One step in developing a community satisfaction prediction model in rehabilitation and reconstruction is processing the satisfaction data for each stage in a KDD process to form a DM prediction model. DM is a logical combination of data knowledge and statistical analysis developed in knowledge or a business process that uses statistical techniques, mathematics, artificial intelligence, artificial intelligence, and machine learning to extract and identify valuable information for related knowledge from large databases. The DM approach continues to be developed in various scientific fields. In recent times the use of DM for predicting social problems is increasing [22]. At the KDD stage, the DM algorithm has equipped a dataset used during the learning-phase, to be developed into a data-driven model. The model can be described as the relationship between input and output, which can provide helpful information.

Understanding and deepening the scientific field has an essential influence on the success of designing the DM algorithm. The database is only a meaningless set of data if an appropriate algorithm is not approached [23]. Furthermore, Fu also said that reviews carried out in the last few years show that DM's ability is growing in specific domains and depends on continuously developing specific algorithms. In simple cases, science can help identify the right features to model the data that underlie the compilation of scientific databases. Knowledge can also help design business goals that can be achieved using in-depth database analysis.

In this study, the database collects data on various satisfaction variables in the pre, during, and post-rehabilitation and reconstruction. Stages summarized in a post-disaster management system can be defined, and algorithms can be compiled to become real information support in improving mitigation management. The development of a system like this has a significant impact on the scientific development of disaster management, and even if the prediction accuracy is only a little, it is still better than random guessing. The availability of a complete database can provide a better and more reliable satisfaction prediction model [24].

3. Research method

In developing community satisfaction prediction models, complete information is needed about the characteristics of the type of work carried out. In general, community satisfaction at each stage is relatively easy to obtain if data is collected regularly and routinely. Community satisfaction is generally easy to compile and has several measurement methods to evaluate overall community satisfaction objectively. Meanwhile, data satisfaction that is outside the existing standard stages is a little more challenging to obtain and requires a long time. For example, data on community satisfaction pre-handling rehabilitation and reconstruction, compared to other stages, is more difficult to obtain. Existing data is more subjective, so that the quality of the data obtained depends on the ability of stakeholders to see and see analyze the conditions of these stages.

This section will describe the methods used to predict community satisfaction. This analysis is not mathematical, but it is carried out to obtain illustrations to show the argument that the proposed method is a more effective model. The community satisfaction prediction model is considered very important in completing a natural disaster management system. In addition, information related to the characteristics of community satisfaction includes pre, during, and post-rehabilitation and reconstruction, which are variables that are considered to have a significant influence on overall community satisfaction.

The community satisfaction model can be used in each stage, analyze disaster management, and determine the rehabilitation and reconstruction methods needed. Disaster management can analyze the existing conditions of the disaster management stages required to complete each disaster management step. This is linked to decision-making in management regarding the best and alternative methods for implementing post-disaster rehabilitation and reconstruction. In developing this model, researchers will use a DM-based community satisfaction prediction approach using data collected from the rehabilitation and reconstruction work locations in Palu, Sigi, and Donggala. Data is divided according to the handling area for calibration, learning, test, and validation purposes.

3.1 Model approach

This study will develop a community satisfaction prediction model with the DM approach without any restrictive assumptions by considering the input data sourced from the questionnaire results. The preparation of a community satisfaction prediction model with DM follows the following stages and processes. It was first cleaning and researching data that can be used in the deterioration model. The data cleaning process includes deleting inappropriate and irrelevant data from the database. This process can include writing errors, ensuring that the writing format remains consistent, and deleting records with incomplete data.

Second, check the data. The first step is to make a histogram or bar chart to determine the frequency of each variable. After that, the relationship of each data must be found. Knowing the distribution and correlation between existing variables helps researchers choose the proper form of data and be more efficient in evaluating the mode to be formed. In data checking, discrepancies and inaccuracies can be found so that further data cleaning is required. The level of correlation refers to the relationship between two variables. A high level of correlation indicates that the two variables are closely related, where if one of these variables changes, the other variables will also change proportionally. If the variables are continuous, these variables will form a line if drawn together. A low level of correlation indicates that the two variables change randomly and are not related. Most of the data fall between two extreme values. The correlation level test is shown through the correlation matrix.

Third, choosing the type of model. After considering each type of model previously studied (deterministic, probabilistic, and artificial intelligence). In this research, the development of the selected AI-based model. Developing a community satisfaction model is carried out through iteration stages by changing aspects of the model to form the best model based on the available data. Model development is done by adjusting aspects to the type of model and the available software. Several factors influence the shape of the model, among others, the basic equation, the variables used in the model, and the grouping of these variables into groups.

Fourth, look for parameter values. Determination of values and parameters is required in model development. In general, this step is completed using an optimized algorithm equation. However, for simple models (for example, a linear regression model using the least square method), this value can be manually

optimized using a spreadsheet program. The *rminer* provides a complete menu option in determining the parameter value with the command:> contribution.

Finally, after the parameter values are obtained and the model has been formed, the model must be evaluated. The evaluation method will depend on the type of model selected. If, after evaluation, the model is not feasible, then the type of model must be reconsidered. If the type of model is still deemed inadequate, the form of the model must be changed and redeveloped. If the evaluation results conclude that the model type is unsuitable for the available data, then the model type must be reconsidered. There are several ways to evaluate statistical models. One of the initial actions that must be considered in evaluating a model is estimating parameter values. The parameter values must be reasonable and significant.

3.2 Model evaluation

By considering the classification or regression approach, other alternative evaluation steps can also be taken. The evaluation process is carried out for regression based on the difference between the observed value and the estimated value (error value). In general, the lower the error value, the better the community satisfaction prediction model, where the error value = 0 is the ideal value to be achieved.

In this study, three measurements were taken: the mean absolute deviation (MAD) root mean squared error (RMSE). Models with low MAD and RMSE values and R^2 values close to the unit value can be interpreted as models with a high level of prediction. RMSE is more sensitive to extreme values than MAD, and this is because RMSE uses the square value of the difference between the measurement results and the predicted model results. Compared to MAD, RMSE is more likely to produce a more significant error value in a model. Looking at the differences, measuring the error value through the two models will provide a different perspective on the proposed model to be used as a comparison.

Furthermore, different DM regression models can be easily compared by drawing a regression error characteristic (REC) graph, which depicts the tolerance for error values on the x-axis compared to the error tolerance percentage values estimated on the y-axis. The representation of the feasibility level of the model is also used in this study. All outputs are collected for evaluation. The integration of the R application with other reporting applications can be facilitated by compiling additional scripts.

3.3 R Tools

The satisfaction pattern through the community satisfaction prediction model is designed to be dynamic with various algorithm choices. The choice of the Multiple Regression (MR), ANN, and SVM algorithms is expected to provide various approaches to community satisfaction with the rehabilitation and reconstruction stages. The results of developing a community satisfaction model will be evaluated and adjusted throughout the disaster management stages until a model can translate the dynamics of existing data. The prediction model must be dynamic and respond to changing conditions [25].

Getting a fit model has carried out a whole iteration of all possible combinations between all variables. In this study, iterations were carried out with consideration of 25 variables and combination exploration. The model selection stage, especially during the feature selection stage, is only applied to the SVM algorithm. The advantage of this approach lies in the fact that the three SVM hyperparameters (c, γ, ϵ) can be set automatically and are urgently needed during the feature selection process.

During the learning phase (after selecting the input variables), the ANN algorithm in this study will use the overall multilayer perception relationship, with one

hidden layer using H processing units, relationship predictions, and logistic activation functions 1 / (1 + e (−x)). The best value of H can be found by range {2, 4, …, 10}, under the internal value (amount of training data used), around 5-fold cross-validation has been performed [26]. Based on tracing the built network, the value of H, which produces the smallest MAD value, has been selected, and ANN is retested using all training data. For the SVM algorithm, to reduce search space, this study uses the Gaussian kernel approach and the proposed heuristics approach to determine *complexity penalty parameter* = 3, and sizes for *incentive tube*, $\epsilon = \frac{\hat{\sigma}}{\sqrt{N}}$, where $\hat{\sigma} = \frac{1,5}{N} \cdot \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$, y_i ais the amount of data used [27]. The most critical parameters in SVM are *kernel parameter* γ, used in the search scope {2^{-15}, 2^{-13}, …, 2^3}, below the minimum 5-*fold cross-validation* [26].

Completing the modeling of the ANN and SVM algorithms, in this study, the MR model was tested as a comparison. The entire DM algorithm consisting of ANN, SVM, and MR is implemented with the R-Tool (R Development Core Team, 2009) and *rminer library* [28]. Furthermore, before fitting the ANN, SVM, and MR models, all data are tested with standard statistics, and then the output is tested for inverse transformation.

4. Experiment and discussion

As study material in this paper used data from the earthquake incident on September 28, 2018, in Palu, Sigi, and Donggala. This choice takes into account that the disaster has a reasonably broad impact on damage. In general, the damage can be divided into several phenomena. One of them is the damage caused by fault movements, fractures, and earthquake shocks. The fault movement is an offset where the left side moves north and the right side shifts to the south. The length of the most considerable shear on the right side is about 4 m, while the left side shifts to the north along 3 m. This shift is visible on the map visible on Google map. Of course, buildings that are traversed by faults will suffer significant damage and soil fractures, where fractures can be the impact of the movement of faults (or reactivated faults) with a smaller offset. Earthquake shocks are in the form of vibrations both horizontally and vertically. In general, in Palu City, the impact of damage due to shocks was not too much, except for buildings of low quality.

Therefore, is the phenomenon of damage due to the tsunami. The impact of a tsunami is the result of inundation (submerged buildings) and tsunami currents (speed or force acting to push or pull buildings). The impact of current velocity is mainly the scouring of the subgrade. If it is loose sand, the erosion rate is very high. Generally, buildings with shallow foundations fail because the scour reaches the base of the foundation. The buildings are relatively light, so they are easily carried away by the flow of water. Another damage is due to the tsunami and at the same time carrying debris to cars and ships, so collisions with these objects often result in heavy damage.

Lastly is the phenomenon of damage due to liquefaction. There are 4–5 locations that are pretty prominent and wide, namely in Balaroa, Petobo, Jono Oge, Lolu village (also in Jono Oge), and Sibalaya. Although some spots also occur liquefaction in the sand boil, it is not prominent and is not recorded. In addition, landslides in the sea can occur due to liquefaction. This kind of avalanche is induced by liquefaction. The landslides in Balaroa and Sibalaya were a phenomenon of liquefaction-induced landslides. It is possible that the submarine landslides that occurred in Palu Bay which caused the tsunami impact had the exact mechanism as in Sibalaya.

4.1 Community satisfaction prediction model

This section presents the modeling framework and procedures used to develop the ANN and SVM approach models. Similar to the traditional modeling process, where the goal is to estimate set coefficients in the form of a particular function. The main objective of the ANN model in this study is to obtain a set of matrices, which are abstract basic knowledge of the available data after going through the training loop. However, to use ANN in solving real-world problems, it is necessary to design a framework following the characteristics of a problem. The framework design aims to define the required ANN architecture and the relationships between the components in the framework. After completing the design framework, the next stage is to design the architecture of each ANN sub-model. The ANN architectural design process is a decision-making process, which includes determining the number of layers, the number of neurons in each layer, the variables entered into the input layer and the output layer. After completing the ANN architectural design, the design results need to be tested and validated.

In general, a neural network is made up of millions (even more) of the basic structures of interconnected and integrated neurons so that they can carry out activities regularly and continuously as needed. The imitation of a neuron in an artificial neural network structure is a processing element that can function as a neuron. The number of input signals is multiplied by the corresponding weight w. Then do the sum of all the results of the multiplication and the resulting output is passed into the activating function to get the degree of the output signal f (a, w). Although it is still far from perfect, the performance of this neuron clone is identical to that of the cell biology we know today. The collection of neurons is made into a network that functions as a computational tool. The number of neurons and the network structure for each problem solved is different.

Furthermore, this model was developed by activating the entire network in ANN. Activating an artificial neural network means activating every neuron used in that network. Many functions can be used as activators, such as goniometric and hyperbolic functions, step unit functions, impulses, sigmoid, etc. Of the several commonly used functions is the sigmoid function because it is considered closer to the human brain's performance. The algorithm activation process during iteration can be monitored, and its movement pattern can be seen.

In contrast to the neural network strategy, which seeks to find a hyperplane that separates classes, SVM tries to find the best hyperplane in the input space. The basic principle of SVM is a linear classifier. It is further developed to work on non-linear problems by incorporating the concept of a kernel trick in a high-dimensional workspace. This development encourages research in modeling to explore the potential capabilities of SVM theoretically and in terms of application. Currently, SVM has been successfully applied to real-world problems, and in general, provides a better solution than conventional methods.

4.2 Community satisfaction data

The model built is verified using data from questionnaire collection around the rehabilitation and reconstruction project. The questionnaire result dataset includes 625 results from 2 rehabilitation and reconstruction projects and 25 input parameters referred to as influencing parameters in an empirical study of community satisfaction. These parameters are given a sequence code based on the pre-during-post stage as input, as shown in **Table 1** below. All data obtained based on the level of importance and level of performance of each parameter asked the correspondent.

No	Code	Satisfaction Indicator
A. Before the rehabilitation and reconstruction		
1	A1	Information and socialization about reconstruction & rehabilitation
2	A2	The time the reconstruction program began
3	A3	Road & bridge damage identification process
4	A4	Participation in the reconstruction & rehabilitation process
5	A5	Collaboration between local communities in reconstruction & rehabilitation
6	A6	The wishes of the people are fulfilled by the reconstruction & rehabilitation
7	A7	Easy administration/disbursement process
8	A8	The role of government in the reconstruction process
B. During the rehabilitation and reconstruction		
9	B1	The role of the facilitator in the reconstruction & rehabilitation process
10	B2	Labor availability
11	B3	Work experience and skills
12	B4	Availability of material for reconstruction & rehabilitation
13	B5	Quality material available for reconstruction & rehabilitation
14	B6	Quality of road & bridge
15	B7	Community participation in the reconstruction & rehabilitation
C. After the rehabilitation and reconstruction		
16	C1	With the results of existing assistance
17	C2	The current state of the road & bridge is compared to the past
18	C3	The road & bridge become earthquake-resistant
19	C4	The comfort of road & bridge compared to before
20	C5	The quality of the road & bridge now compared to before
21	C6	The road & bridge was had been as a community wish
22	C7	Satisfaction with the current design
23	C8	The access road to residence compared to before the reconstruction & rehabilitation
24	C9	Current availability of street/environment lighting
Result		
25	CS	Community Satisfaction

Table 1.
Input code.

4.3 Stages of learning and modeling test

Forming a dataset is carried out to form three datasets that can be used immediately to learn, test, and validate. The database is divided into two datasets. The first set includes all the information. The dataset of both questionnaires was collected,

which will be used for validation purposes. The entire dataset used for learning and test purposes is further divided into two subsets to obtain learning datasets. One set contains 80% of the data used for learning and 20% of the data used for testing. It is statistically independent data from the dataset used during learning and testing based on separating the dataset for the validation process. Therefore, verification of the DM model by using a separated dataset can be considered a control to check the performance of the DM model. The learning process is carried out with the number of epochs (10,000 times). The iteration process produces an ANN model that has an optimal weight between neurons.

After the learning phase is complete, the model development step is continued to the test stage to check the effectiveness of the learning process. The dataset used in the test stage becomes the DM input. The algorithm used in this stage uses a learning algorithm that has been recorded in the DM application when the learning process is running. The test process can calculate the error rate that occurs. If the error level of the test stage is still within an acceptable level, then the DM model is considered reasonable. A comparison of the model's accuracy is made by comparing the average MSE values during the test phase. Finally, the DM model with the lowest MSE error rate and the highest R^2 is selected. Finally, after the learning and test process is complete. Furthermore, the verification and validation of the model are carried out using the data that has been prepared with the prediction model of the community satisfaction learning and test results. Different dataset details were selected for model validation.

4.4 Model interpretation

In engineering science, apart from requiring a high level of accuracy, it also requires interpreting the modeling results. The ability to interpret DM is greatly influenced by the power of the data-driven model for this purpose. When the DM black box is implemented with ANN, SVM, and MR algorithms that involve complex mathematical expressions, the data-driven application procedure provided must translate the model. In this case, the results of the model interpretation are carried out to obtain a measurement of the input variables of the community satisfaction prediction model.

The first stage of model interpretation is to believe in the ability and accuracy of the model. The prediction model of community satisfaction using community satisfaction as the leading prediction parameter is first checked for modeling accuracy. There are several methods for evaluating predictive models, one of which uses the sum of absolute errors. The sum of the absolute errors often referred to as the absolute deviation of the average or MAD, is measuring forecasting accuracy by averaging the forecast errors using their absolute values. MAD is beneficial for analyzing and measuring the prediction error in the same unit of measure as the original data. In addition, the resulting process modeling criteria are stated in the RMSE, provided that the smaller the resulting RMSE (close to the value 0) will result in a better output prediction model.

This model is structured with a confidence level of 95% according to the t-student distribution. All DM models with ANN, SVM and MR algorithms are trained using 12 input variable attributes. **Figure 1** shows the predictive capacity of all training outcome models, comparing their performance in predicting the value of community satisfaction based on MAD, RMSE, and R^2. This table shows that the value of community satisfaction can be predicted accurately by each of the three DM models, especially by the ANN and SVM models.

Figure 1 above shows the standard error, and R^2 for each model developed. The DM model with the SVM algorithm has the smallest MAD value and RMSE value,
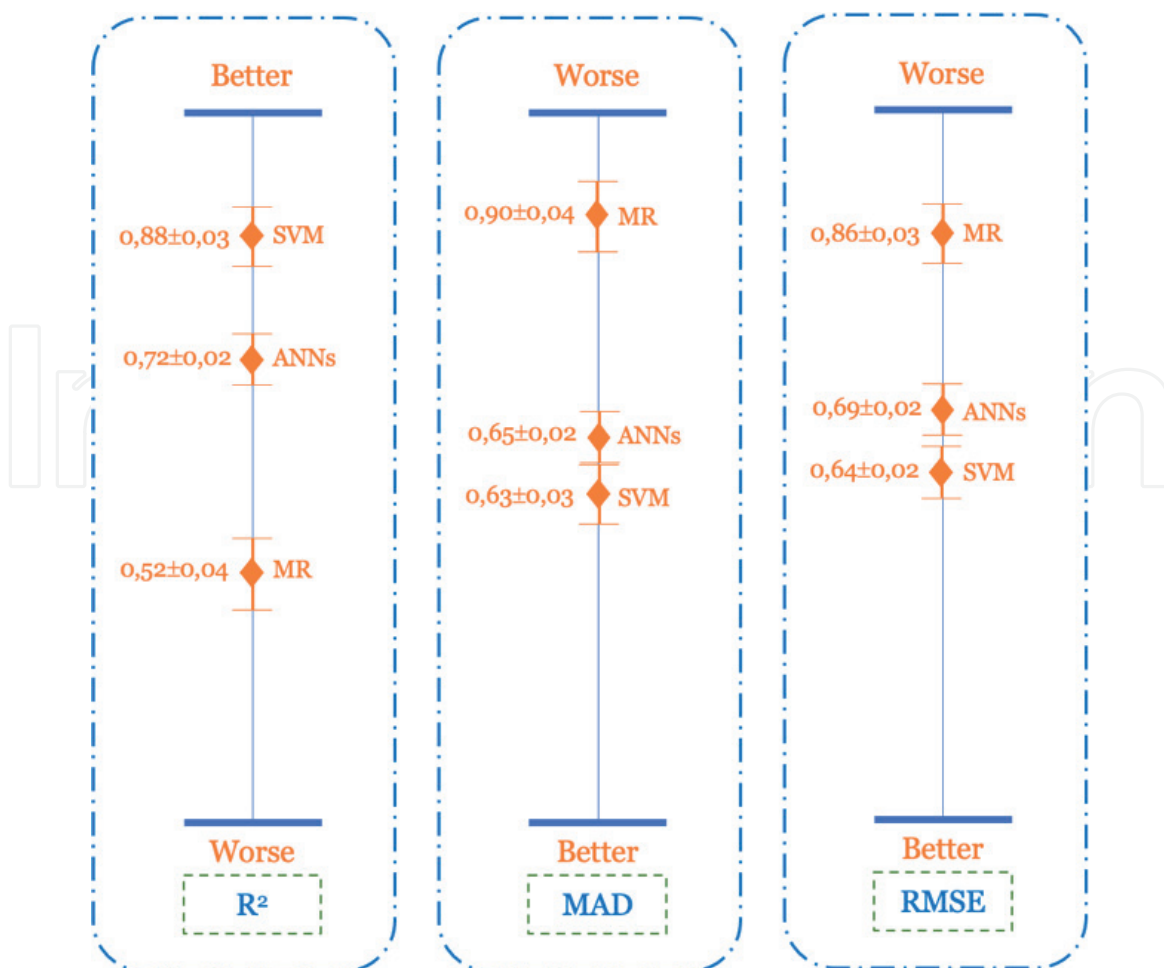
Figure 1.
Performance measured.

and the highest R^2 value. The prediction model with the ANN and SVM algorithms is acceptable and can be used in calculating community satisfaction predictions because it has R^2 close to 1. The following community satisfaction prediction model used in this study is the DM model with the SVM algorithm.

DM technique, also known as association rule mining, can find associative rules between a combination of items. Two parameters can determine the importance of an associative rule. The parameter is the percentage combination of these attributes in the database and confidence, namely the strength of the relationship between attributes in the associative rule. With the generate and test paradigm, the algorithm used in this study is making candidate combinations of attributes based on specific rules and then tested. Combining attributes that meet these requirements is called a frequent itemset, which is then used to create rules that meet the minimum confidence requirements.

By analyzing **Figure 2** (the scatterplot of the community satisfaction value prediction of the SVM algorithm with the questionnaire results), the variables that have been determined have a significant relationship with the change in the value of the questionnaire community satisfaction. **Figure 2a** shows the scatter-plots of learning results in the SVM model, and **Figure 2b** shows the results of the validation stages.

In the validation stage, the library feature *rminer* is used to describe and obtain the relative contribution value of each input value. The confirmed model has R^2, MAD, and RMSE values in the performance validation stage, such as **Figure 1**, with 20 runs performed, while the best hyperparameters to achieve a fit SVM model are used. $\epsilon = 0.07 \pm 0.01$ and $\gamma = 0.05 \pm 0.00$. Whereas the hyperparameters for ANN used $H = 3 \pm 1$.
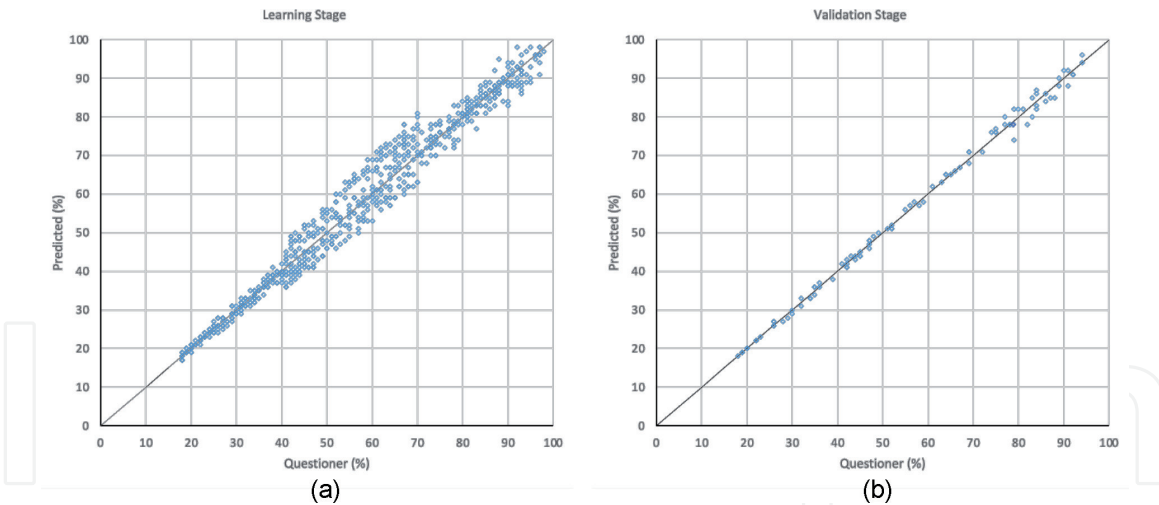
Figure 2.
Community satisfaction prediction outputs. a. Learning stage, b. validation stage.

Furthermore, the interpretation of the regression analysis used in DM is carried out. Package *rminer*, provides a graphical interpretation tool, namely: REC curve, error tolerance depicted on the x-axis, while the percentage value of road performance predictions is depicted on the y-axis. The resulting curve describes the level of error in the form cumulative distribution function (CDF). The error level defined as the difference between the predicted values of community satisfaction f(x) with community satisfaction actual on every coordinate (x, y). This approach is also a squared residual $\left(y - f\left(x\right)\right)^2$ or absolute deviation $\left|y - f\left(x\right)\right|$ based on error metric mapping. **Figure 3** shown REC curve community satisfaction model with MR, ANN, dan SVM algorithm.

In **Figure 3** it can be analyzed that the REC curve describes the error tolerance on the x-axis and the level of accuracy of the regression function on the y-axis. The level of accuracy is defined as the percentage of modeling results that fit the specified tolerance. If the tolerance value is zero, only that value is considered to meet the model requirements. However, if you choose the maximum tolerance, other values can be used as reference for accuracy values. In the REC curve it is clear that the level of accuracy has a trade-off with tolerance. The greater the tolerance value given, the higher the accuracy value. Conceptually, the model with the lowest tolerance value with the highest accuracy is the model that has the best REC value.

The illustration of the REC curve depicts three different models. The curve shows that the SVM model has the highest accuracy value with the smallest tolerance value that moves consistently. This REC curve depicts the entire iteration process with 20 runs on the SVM model with hyperparameters as mentioned in the previous section. The shape of the REC curve can change shape when using different hyperparameters and the number of iteration runs is different.

4.5 Variable contribution

The DM model developed can assess each variable's contribution and attribute that becomes input data in the model. In this study, the variables or attributes consist of A1-C9. All attributes are then grouped into three dimensions pre, during, and post. A parameter vector in this DM model is chosen to explain that it is a variable function and not parameters as in the parametric approach. The only condition for a variance function is to be able to generate a non-negative definite variance matrix. Several methods can be used to estimate hyperparameter values. The value of θ can be predicted in this DM by using the cross-validation method. Hyperparameter used (H and γ) are H (2, 4, ..., 10) and γ (2–15, 2–13, ..., 23). This value produces
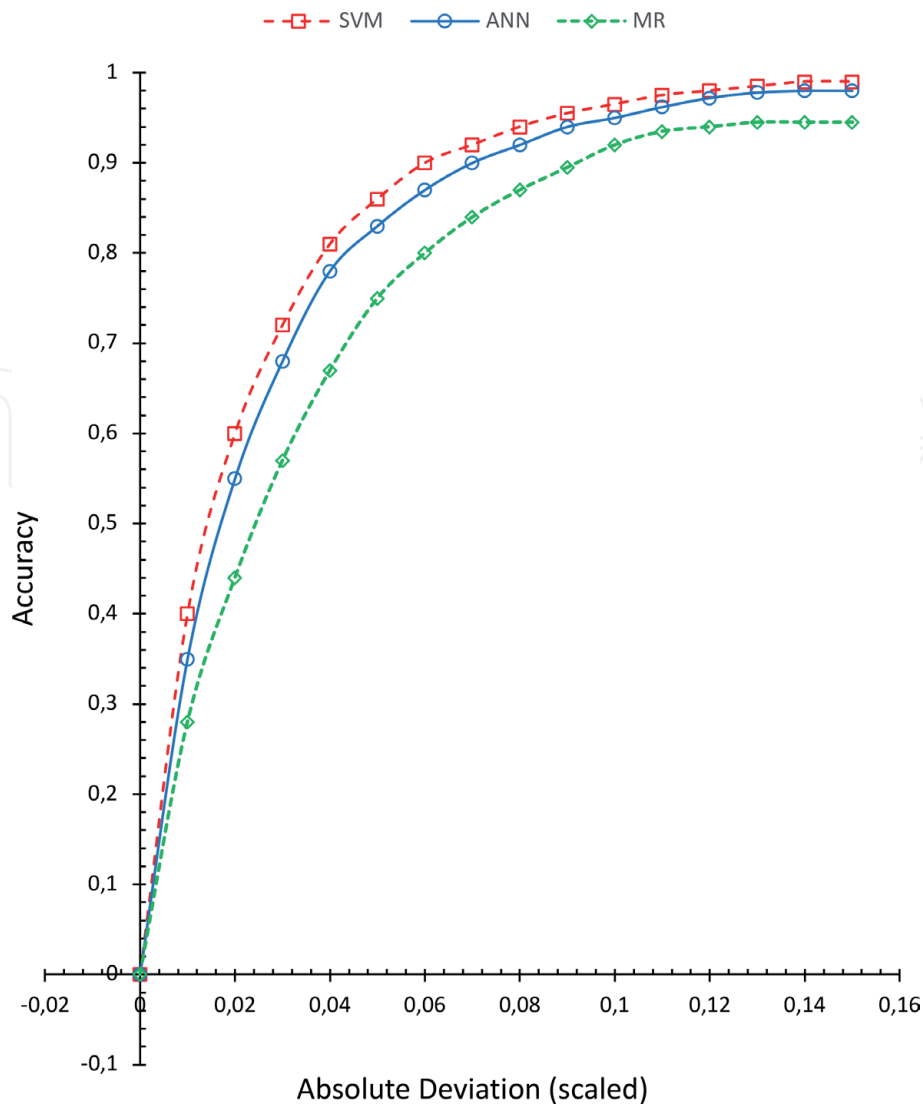
Figure 3.
The regression error characteristic curve.

the most precise model with optimal run time. For further model development, an approach can be used to try other hyperparameter values. The contribution of each attribute and dimension is of relative importance in composing the model.

The search results for the contribution value in DM can be simplified and displayed in **Figure 4**. This figure can display the relative importance on the x-axis for each attribute and dimension on the y-axis forming the community satisfaction prediction model with the DM model approach using the SVM, ANN, and MR algorithms.

Based on **Figure 4** above, each parameter has an almost even effect on community satisfaction in disaster management. When using a model that is considered the fittest, namely SVM, it can be seen that the most significant importance is the comfort of road and bridge compared to before (C4), and Collaboration between local communities in reconstruction and rehabilitation (A5). Therefore, the access road to residence compared to before the reconstruction and rehabilitation (C8), Participation in the reconstruction and rehabilitation process (A4), and Community Participation in the reconstruction and rehabilitation (B7). While pre-rehabilitation and reconstruction, the stage is the most critical dimension affecting community satisfaction.

The following model analysis is to compile an algorithm to select the main dimensions that affect the community satisfaction model and analyze the supporting variables that affect the community satisfaction prediction model that is not accommodated in this model. The results of VEC analysis illustrate the influence of the main attributes that move dynamically in the prediction model of community satisfaction
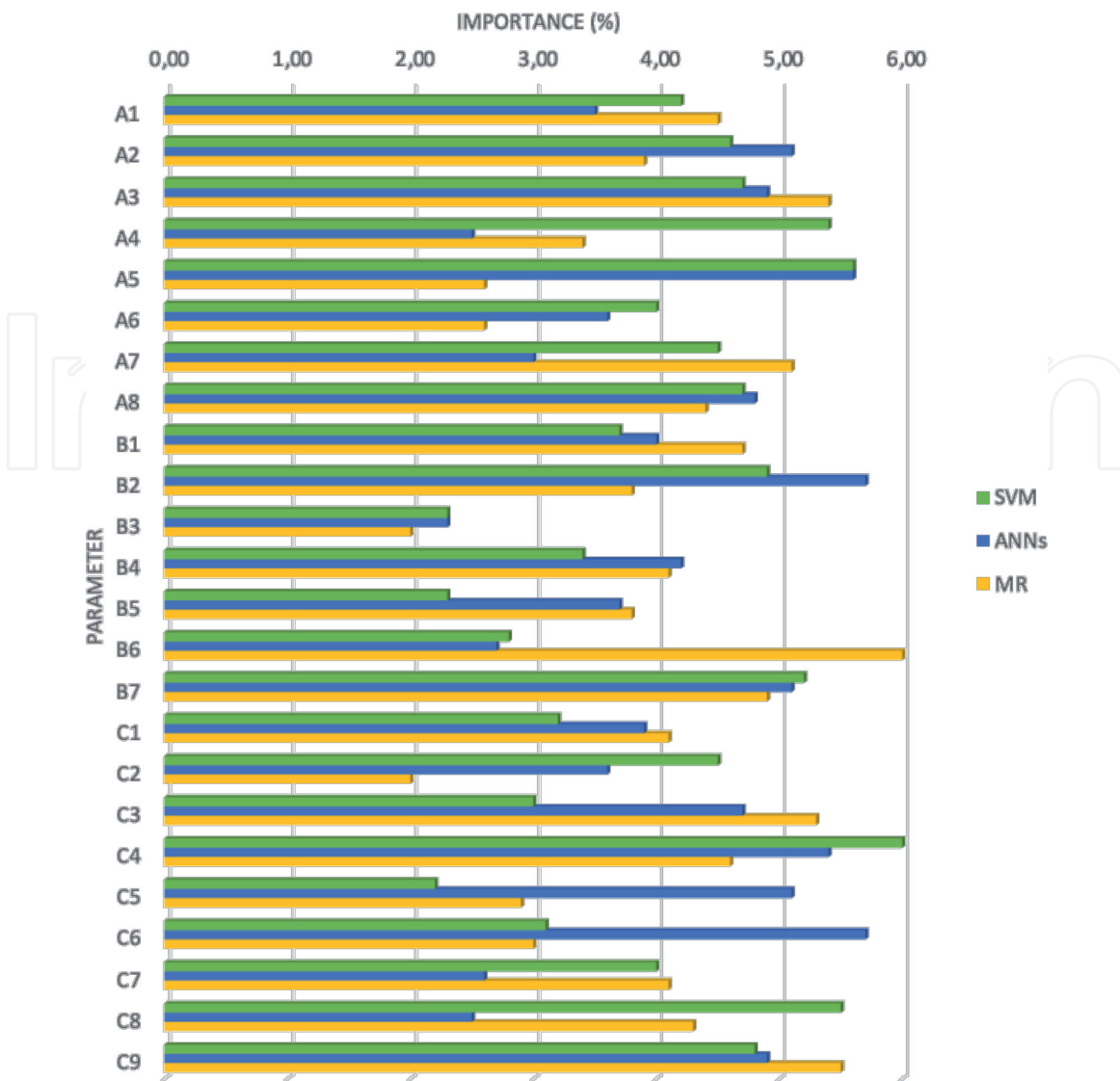
Figure 4.
Relative importance.

with this SVM model in the form of information and socialization about reconstruction and rehabilitation (A1), a pre-rehabilitation and reconstruction group. Decreased community satisfaction following the time of reconstruction program began (A2) and the role of the facilitator in the reconstruction and rehabilitation process (B1), and conversely, community satisfaction improved when performed the access road to residence compared to before the reconstruction and rehabilitation (C8).

5. Conclusion

The modeling process with the DM approach using the SVM, ANN, and MR algorithms produces a community satisfaction prediction model with a reasonably good model performance. The three model algorithms are compared with the questionnaire results. The REC curve shows the accuracy of each model used. Based on the resulting error matrix, it is believed that the SVM model is the best model to predict community satisfaction with a low iteration of 20 runs and has a good consistency. The most critical parameter in preparing the community satisfaction prediction model is the comfort of the road and bridge compared to before. Each attribute that affects the community satisfaction prediction model is successfully described with the algorithm of relative importance.

Acknowledgements

Author details

Andri Irfan Rifai
Universitas Internasional Batam, Batam, Indonesia

*Address all correspondence to: andri.irfan@uib.ac.id

IntechOpen

References

[1] Tim Peneliti Unpar (2018). *Menyelisik Untaian Bencana Palu-Dongala (in Bahasa).* Bandung: Universitas Katholik Parahyangan.

[2] Wu, X., Wang, Z., Gao, G., Guo, J., & Xue, P. (2020). Disaster probability, optimal government expenditure for disaster prevention and mitigation, and expected economic growth. *Science of the total environment, 709*, 135888.

[3] Meilianda, E., Munadi, K., Direzkia, Y., & Oktari, R. S. (2017). Assessment of post-tsunami disaster recovery of Banda Aceh city of Indonesia as window of opportunities for sustainable development. *IOP Conference Series: Earth and Environmental Science, Vol. 56, No. 1* (p. 012019). IOP Publishing.

[4] Jamaludin, S., Azmir, N. A., Ayob, A. F., & Zainal, N. (2020). COVID-19 exit strategy: Transitioning towards a new normal. *Annals of Medicine and Surgery*, 165-170.

[5] Thein, P. S., Pramumijoyo, S., Brotopuspito, K. S., Kiyono, J., Wilopo, W., Furukawa, A., & Putra, R. R. (2015). Estimation of S-wave velocity structure for sedimentary layered media using microtremor array measurements in Palu City, Indonesia. *Procedia Environmental Sciences, 28*, 595-605.

[6] Abdullah, A. I. (2020). A field survey for the rupture continuity of Palu-Koro fault after Donggala earthquake on September 28, 2018. *In Journal of Physics: Conference Series (Vol. 1434, No. 1)*, 012009.

[7] Widyaningrum, R. (2012). *Penyelidikan Geologi Teknik Potensi Liquefaksi Daerah Palu, Provinsi Sulawesi Tengah (in Bahasa)*. Bandung: Pusat Sumber Daya Air Tanag dan Geologi Lingkungan.

[8] Ahmad, M. I., & Ma, H. (2020). An investigation of the targeting and allocation of post-flood disaster aid for rehabilitation in Punjab, Pakistan. *International Journal of Disaster Risk Reduction, 44*, 101402.

[9] Nakamura, N., & Kanemasu, Y. (2020). Traditional knowledge, social capital, and community response to a disaster: resilience of remote communities in Fiji after a severe climatic event. *Regional Environmental Change, 20(1)*, 1-14.

[10] Islam, E., Abd Wahab, H., & Benson, O. G. (2020). Structural and operational factors as determinant of meaningful community participation in sustainable disaster recovery programs: The case of Bangladesh. *International Journal of Disaster Risk Reduction, 50*, 101710.

[11] Lu, Q., Zhong, D., & Zhang, Q. (2020). The evolving pattern of NGOs' participating in post-disaster community reconstruction in China: cases study on the 2008 Wenchuan earthquake and the 2013 Lushan earthquake. *Natural Hazards, 104(1)*, 167-184.

[12] Paudel, D., Rankin, K., & Le Billon, P. (2020). Lucrative Disaster: Financiali-zation, Accumulation and Post-earthquake Reconstruction in Nepal. *Economic Geography, 96(2)*, 137-160.

[13] Daly, P., Mahdi, S., McCaughey, J., Mundzir, I., Halim, A., & Srimulyani, E. (2020). Rethinking relief, reconstruction, and development: Evaluating the effectiveness and sustainability of post-disaster livelihood aid. *International Journal of Disaster Risk Reduction, 49*, 101650.

[14] Yong, Z., Zhuang, L., Liu, Y., Deng, X., & Xu, D. (2020). Differences in the disaster-preparedness behaviors of the general public and professionals: evidence from Sichuan Province, China. *International journal of environmental research and public health, 17(14)*, 5254.

[15] Santha, S. D. (2018). Social interfaces in disaster situations: Analyzing rehabilitation and recovery processes among the fisherfolk of Tamil Nadu after the Tsunami in India. In *The Asian Tsunami and post-disaster aid* (pp. 65-78). Singapore: Springer.

[16] Sofyan, M. (2019). Community Satisfaction of the Urban Flood Control System Improvement Project (UFCSI). *Ilomata International Journal of Social Science, 1(1)*, 29-34.

[17] Ophiyandri, T., Hidayat, B., & Ghiffari, C. (2020). Community satisfaction levels on the housing reconstruction project after Mentawai tsunami in 2010-a case study at Sipora Island. *IOP Conference Series: Materials Science and Engineering Vol. 933, No. 1,* (p. 012041). IOP Publishing.

[18] Tinoco, J., Correia, A. G., & Cortez, P. (2014). Support vector machines applied to uniaxial compressive strength prediction of jet grouting columns. *Computers and Geotechnics 55*, 132-140.

[19] Liu, H., Motoda, H., Setiono, R., & Zhao, Z. (2010). Feature Selection: An Ever Evolving Frontier in Data Mining. *Proceedings of the Fourth Workshop on Feature Selection in Data Mining*, (pp. 4-13). Hyderabad, India.

[20] Rahman, F. A., Desa, M. I., Wibowo, A., & Haris, N. A. (2014). Knowledge Discovery Database (KDD)-Data Mining Application in Transportation. *Proceeding of the Electrical Engineering Computer Science and Informatics, 1(1),* (pp. 116-119.).

[21] Wang, X. Z. (2012). *Data mining and knowledge discovery for process monitoring and control.* Springer Science & Business Media.

[22] Suma, V., & Hills, S. M. (2020). Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics. *Journal of Soft Computing Paradigm (JSCP), 2(02)*, 101-110.

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

[24] Zhao, Y., Xu, X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management, 76,* 111-121.

[25] Rifai, A. I., Hadiwardoyo, S. P., Correia, A. G., Pereira, P., & Cortez, P. (2015). Data Mining Applied for The Prediction of Highway Roughness under Overloaded Traffic. *International Journal of Technology*.

[26] Hastie, R. T., Tibshirani, & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer-Verlag New York, second edition.

[27] Cherkassky, V., & Ma, Y. (2004). Practical selection of svm parameters and noise estimation for svm regression. *Neural Networks, 17(1) ISSN 0893-6080.*, 113-126.

[28] Cortez, P. (2010). Data Mining with Neural Networks and Support Vector Machines Using the R/*rminer* Tool. *10th Industrial Conference on Data Mining (ICDM 2010)*, (p. Lecture Notes in Artificial Intelligence 6171). Advances in Data Mining.