INTERNATIONAL SOCIETY FOR
SOIL MECHANICS AND
GEOTECHNICAL ENGINEERING

*This paper was downloaded from the Online Library of the International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE). The library is available here:*

*https://www.issmge.org/publications/online-library*

*This is an open-access database that archives thousands of papers published under the Auspices of the ISSMGE and maintained by the Innovation and Development Committee of ISSMGE.*

*The paper was published in the proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR 2019) and was edited by Jianye Ching, Dian-Qing Li and Jie Zhang. The conference was held in Taipei, Taiwan 11-13 December 2019.*

# Predicting Slope Stability with Incomplete Data Using Naive Bayes Classifier

Xianda Feng[1], Shuchen Li[2], Chao Yuan[3], and Peng Zeng[4]

[1]School of Civil Engineering and Architecture, University of Jinan, No. 336, West Road of Nan Xinzhuang, Jinan 250022, Shandong Province, P.R. China. E-mail: feng.xianda@hotmail.com
[2]Geotechnical Structural Engineering Research Center, Shandong University, No. 17923, Jingshi Road, Jinan 250061, Shandong Province, P.R. China. E-mail: shuchenli@sdu.edu.cn
[3]Laboratoire 3SR, University Grenoble Alpes, France and CNRS, 3SR, F-38000 Grenoble, France.
E-mail: chao.yuan2012@gmail.com
[4]State Key Lab. of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, No. 1, Eastern Erxianqiao 3rd Road, Chengdu 610059, Sichuan Province, P.R. China.
E-mail: zengpeng1987@hotmail.com

**Abstract:** Landslides can cause serious loss of life and property, and being able to predict the stability of a slope are of primary concern in identifying potential landslide sections and mitigating damages caused by landslides. In this study, we employ a Naive Bayes classifier (NBC) to predict slope stability for a slope subjected to circular failures, based on six parameters: slope height ($H$), slope angle ($\alpha$), cohesion ($c$), friction angle ($\varphi$), unit weight ($\gamma$), and pore pressure ratio ($r_u$). The Naïve Bayes classifier is "learned", using the Expectation Maximization algorithm, with an incomplete data set of 69 slope cases. The model validation with 13 new cases shows that, when compared to the existing empirical approach, the proposed NBC yields higher accuracy and allows using incomplete data for the prediction. Moreover, it helps in estimating the probabilities of slope stability that are of interest to reliability-based design of slopes.

Keywords: Slope stability; naive Bayes classifier; incomplete data; circular failures.

## 1 Introduction

Landslides are one of the major geological disasters that can cause serious loss of life and property. Assessing and predicting the stability of a slope is a major concern for determining potential landslide profiles and mitigating damage caused by landslides. (Alimohammadlou et al. 2014; Rukhaiyar et al. 2017). Accurately predicting the stability of a slope is a challenging task because it depends on a variety of geotechnical and physical factors. Moreover, the interactions between these factors are complex and "often difficult to describe mathematically" (Ferentinou and Sakellariou 2007; Lu and Rosenbaum 2003; Xue 2017).

A number of methods have been proposed to predict slope stability, with limit equilibrium methods (LEM) and numerical methods being the most common methods (Liu et al. 2014; Xue 2017). Some other approaches include empirical equations (Bye and Bell 2001; Taheri and Tani 2010) and limit analysis approaches based on lower and upper bound theorems (Chen 1975). However, the above methods have certain shortcomings. For instance, limit equilibrium methods cannot reflect the actual stress conditions of the slip surfaces (Lenchman and Griffiths 2000) and their accuracy is affected by simplifying assumptions (Sakellariou and Ferentinou 2005). The numerical methods are usually time consuming and their accuracy is highly dependent on accurate estimates of geotechnical and physical parameters.

More recently, soft computing methods are increasingly used to predict slope stability. (Gordan et al. 2016; Li and Kong 2014; Rukhaiyar et al. 2017; Xue 2017). Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) are the most popular soft computing methods for predicting slope stability because (i) they do not require prior knowledge of specific model forms and have flexible nonlinear modeling capabilities (Alimohammadlou et al. 2014) and (ii) they perform better than the traditional analytical and regression methods in slope stability prediction (Erzin and Cetin 2013; Samui 2008).

Although predictive models developed based on ANN or SVM methods can sometimes produce more accurate predictions than traditional slope stability analysis methods, predicting slope stability is not feasible when data is incomplete, especially in the initial stages of slope design. Therefore, we propose a Naive Bayes classifier (NBC) (Ting et al. 2011) to predict the stability of slopes that are subject to circular failures. It has been proven that NBCs are particularly useful to deal with incomplete data (Uusitalo, 2007) and could yield good predictions even with small data sizes (Kontkanen et al., 1997), making them quite suitable for analyses with limited (or incomplete) geotechnical data.

## 2 Database Description

*Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR)*
*Editors:* Jianye Ching, Dian-Qing Li and Jie Zhang
Copyright © ISGSR 2019 Editors. All rights reserved.
*Published by* Research Publishing, Singapore.
ISBN: 978-981-11-2725-0; doi:10.3850/978-981-11-2725-0_IS8-9-cd

For slopes subjected to circular failures, the major factors affecting the slope stability include the geometry of the slope (i.e., slope height $H$ and slope angle $\alpha$), shear strength of the geomaterial (i.e., cohesion $c$ and friction angle $\varphi$), gravity (i.e., unit weight $\gamma$), and water condition (i.e., pore pressure ratio $r_u$, which is defined as the ratio of the pore pressure to the overburden pressure) (Liu et al. 2014; Michalowski 1995; Rukhaiyar et al. 2017).

We compiled a database comprising 69 slope cases with 41 cases being stable slopes and 28 cases being failed slopes (Feng 2000; Sah et al. 1994; Wang et al. 2005; Xu et al. 1999; Zhou and Chen 2009). Note that the values of $r_u$ for 10 slope cases were not reported, i.e., the input data are "incomplete." The "incomplete data" means that there is some vacancy for some factors. In this study, we choose to predict the actual condition of the slopes (i.e., 0 = Failed and 1 = Stable) instead of predicting the specific values of the factor of safety (FoS).

Fig. 1 shows the histograms, cumulative distributions, and additional statistics of all the six input parameters in the database. Note that the minimum and maximum values of each parameter define the ranges within which the predictions can be conducted.

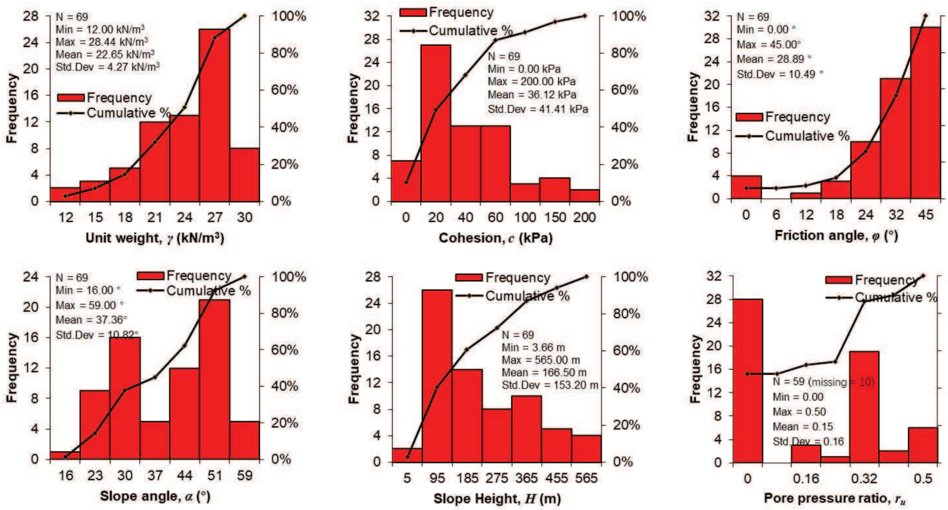


**Figure 1.** Histograms, CDF's, and statistics of the six factors in the database.

## 3 Naive Bayes Classifier

### 3.1 Naive Bayes classifier

$\mathbf{X} = (x_1, x_2, \ldots, x_n)$ is the input vector representing the six independent factors affecting the slope stability (i.e., $\gamma$, $c$, $\varphi$, $\alpha$, $H$, and $r_u$), and $(C_1, C_2, \ldots, C_k)$ denotes the two outcomes (stable or failed) of the slope stability. When $x_1, x_2, \ldots, x_n$ are discrete, using the Bayes' theorem, the conditional probability of the $k^{\text{th}}$ possible outcome can be expressed as follows (Domingos and Pazzani 1997; Friedman et al. 1997; Shirzadi et al. 2017).

$$\mathrm{P}(C_k \mid \mathbf{X}) = \frac{\mathrm{P}(\mathbf{X} \mid C_k)\mathrm{P}(C_k)}{\mathrm{P}(\mathbf{X})} \tag{1}$$

Based on the conditionally independent assumption, we have:

$$\mathrm{P}(\mathbf{X} \mid C_k) = \prod_{i=1}^{n} \mathrm{P}(x_i \mid C_k) \tag{2}$$

Substituting Eq. (2) into Eq. (1) yield:

$$\mathrm{P}(C_k \mid \mathbf{X}) = \frac{\mathrm{P}(C_k)\prod_{i=1}^{n}\mathrm{P}(x_i \mid C_k)}{\mathrm{P}(C_1)\prod_{i=1}^{n}\mathrm{P}(x_i \mid C_1) + \mathrm{P}(C_2)\prod_{i=1}^{n}\mathrm{P}(x_i \mid C_2)} \tag{3}$$

The NBCs aim to determine the class by maximizing the posteriori probability $\mathrm{P}(C_k|\mathbf{X})$ (Chen et al. 2017). It is necessary to choose a suitable threshold probability for classification. As there are only two classes in this study, it is common to use a threshold of 1/2 (Wu and Kumar 2009). Fig. 2 shows the structure of the NBC, wherein each of the six input parameters is connected to the "Slope_Stability" node using arrows.

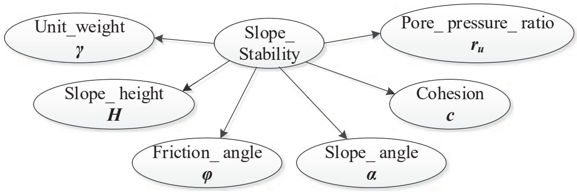


**Figure 2.** Structure of the Naïve Bayes Classifier

### 3.2 Discretization of the continuous factors

In this study, equal-frequency binning algorithm (Kotsiantis and Kanellopoulos 2006), which is a commonly used unsupervised discretization algorithms, is used to discretize the six continuous input factors. Table 1 lists the intervals and corresponding definitions of the states.

**Table 1.** Overview of the intervals and the corresponding state names

| Factors | Set of intervals / States | | |
|---|---|---|---|
| **Slope stability** | 0/Failed | 1/Stable | / |
| $\gamma$ | [12.0, 21.4]/Low | (21.4, 25.5]/Medium | [25.5, 28.44]/High |
| $c$ | [0, 12.0]/Low | (12.0, 43.0]/Medium | (43.0, 200]/High |
| $\varphi$ | [0, 26.8]/Small | (26.8, 34.0]/Medium | (34.0, 45.0]/ Large |
| $\alpha$ | [8, 30.5]/Small | (30.5, 44.8]/Medium | (44.8, 59.0]/Large |
| $H$ | [3.66, 67.0]/Low | (67.0, 205.0]/Medium | (205.0, 565.0]/High |
| $r_u$ | 0/Dry | (0, 0.5]/Wet | / |

Note that the $r_u$ values were not reported in 10 out of the 69 cases. The expectation maximization (EM) algorithm can be used to estimate the conditional probabilities in the NBCs based on these "incomplete" data (Jensen and Nielsen 2007).

## 4 Results and Discussions

### 4.1 Model performance

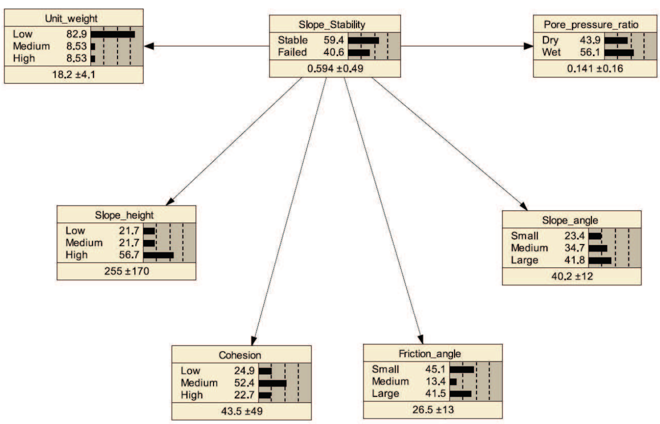


**Figure 3.** The NBC after parameter learning using EM algorithm.

The parameters in the conditional probability tables (CPTs) for each node are trained using the EM algorithm built in Netica (Norsys Software Corporation 1998). Fig. 3 shows the NBC after parameter learning using EM algorithm. Given the trained NBC, the probabilistic inference can be conveniently obtained using Eq. (3). For example, the following factors are assumed for a slope design case: X = ($\gamma$ = 24.6 kN/m3 (medium), $c$ = 50 kPa (high), $\varphi$ = 35° (large), $\alpha$ = 45° (large), $H$ = 290 m (high), $r_u$ = NA). Then the probability of the slope stability, P(Stable | X), can be computed as

$$\mathrm{P(Stable} \mid \mathbf{X}) = \frac{\mathrm{P(Stable)} \prod_{i=1}^{5} \mathrm{P}(x_i \mid \mathrm{Stable})}{\mathrm{P(Stable)} \prod_{i=1}^{5} \mathrm{P}(x_i \mid \mathrm{Stable}) + \mathrm{P(Failed)} \prod_{i=1}^{5} \mathrm{P}(x_i \mid \mathrm{Failed})}$$
$$= \frac{0.5942 \cdot 0.1111 \cdot 0.6785 \cdot 0.2632 \cdot 0.5000 \cdot 0.4286}{0.5942 \cdot 0.1111 \cdot 0.6785 \cdot 0.2632 \cdot 0.5000 \cdot 0.4286 + 0.4058 \cdot 0.0476 \cdot 0.4035 \cdot 0.1739 \cdot 0.2917 \cdot 0.4035}$$
$$= 0.9406 \tag{4}$$

Note that the input data is "incomplete" in this example and the NBC can still compute the value of P(Stable|**X**). In other words, the NBC can predict slope stability with any subset of the six input factors, making it more flexible than other soft computing techniques such as the ANN or SVM methods.

The NBC was first tested using all the 69 slope cases in the database. The overall accuracy is approximately 97.1%, which is quite satisfactory for practical engineering.

The sensitivity analysis could be performed using Netica, and the results show that $H$ is the factor with highest influence on slope stability and $r_u$ is the least important input factor.

### 4.2 Validation with new cases

To validate the proposed NBC, it is tested with 13 new cases obtained from literature that were not included in the training data set. Table 2 lists the results predicted using the proposed NBC and the empirical equation proposed by Sah et al. (1994). Only two cases (Nos. 3 and 11) are misclassified, indicating an acceptable performance of the NBC with new cases.

Three cases (Nos. 3, 11, and 12) were wrongly predicted using the empirical equation proposed by Sah et al. (1994). However, it cannot be applied to case No. 13 wherein the input data are incomplete. Therefore, the results, listed in Table 2, show that the proposed NBC performs slightly better than the empirical equation proposed by Sah et al. (1994) and can be applied to a wider range of slope cases, particularly to the ones with incomplete data or information.

**Table 2.** Results of validation with 13 new cases collected from literatures

| No. | $\gamma$ [kN/m$^3$] | $c$ [kPa] | $\varphi$ [°] | $\alpha$ [°] | $H$ [m] | $r_u$ | Actual (FoS) | Predicted with NBC (P(Stable)) | Empirical equation by Sah et al. (1994) | Reference |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 21.00 | 20.00 | 40.00 | 40.00 | 12.00 | 0.00 | Stable (1.84) | Stable (58%) | Stable (1.82) | Hoek and Bray (1981) |
| 2 | 21.00 | 30.00 | 35.00 | 40.00 | 12.00 | 0.40 | Stable (1.49) | Stable (54%) | Failed (1.19) | |
| 3 | 21.00 | 35.00 | 28.00 | 40.00 | 12.00 | 0.50 | Stable (1.43) | Failed (31%) | Failed (0.98) | |
| 4 | 19.00 | 30.00 | 35.00 | 35.00 | 11.00 | 0.20 | Stable (2.00) | Stable (54%) | Stable (1.80) | |
| 5 | 20.00 | 40.00 | 40.00 | 40.00 | 10.00 | 0.20 | Stable (2.31) | Stable (54%) | Stable (1.94) | |
| 6 | 18.80 | 20.00 | 10.00 | 25.00 | 50.00 | 0.30 | Failed (0.97) | Failed (27%) | Failed (0.52) | Lin et al. (1988) |
| 7 | 19.10 | 10.00 | 10.00 | 25.00 | 50.00 | 0.40 | Failed (0.65) | Failed (20%) | Failed (0.41) | |
| 8 | 18.80 | 20.00 | 20.00 | 30.00 | 50.00 | 0.30 | Failed (1.00) | Failed (27%) | Failed (0.78) | |
| 9 | 19.10 | 10.00 | 20.00 | 30.00 | 50.00 | 0.40 | Failed (0.65) | Failed (20%) | Failed (0.63) | |
| 10 | 22.00 | 20.00 | 22.00 | 20.00 | 180.00 | 0.00 | Stable (1.12) | Stable (54%) | Stable (1.74) | Madzic (1988) |
| 11 | 22.00 | 20.00 | 22.00 | 20.00 | 180.00 | 0.10 | Failed (0.99) | Stable (50%) | Stable (1.57) | |
| 12 | 27.00 | 16.80 | 28.00 | 50.00 | 90.50 | 0.25 | Stable (-) | Stable (55%) | Failed (0.54) | Yan and Li (2011) |
| 13 | 22.00 | 15.00 | 18.00 | - | - | - | Stable (1.84) | Stable (71%) | NA | Zhao (2008) |

## 5 Conclusions

An NBC was developed to predict the probability of slope stability based on six input factors: slope height ($H$), slope angle ($\alpha$), cohesion ($c$), friction angle ($\varphi$), unit weight ($\gamma$), and pore pressure ratio ($r_u$). An EM algorithm was employed to learn the conditional probabilities from the "incomplete" training data set including 69 slope cases. The "learned" NBC was tested with the training data, and the results show that the overall accuracy is 97.1%. In addition, 13 new cases were used to validate the proposed NBC model and only two cases were misclassified, which is considered acceptable for practical engineering. The proposed NBC shows some improvements over the conventional empirical equation proposed by Sah et al. (1994), as it yields higher accuracy and allows using incomplete data for the prediction. Moreover, it helps in estimating the probabilities of slope stability that are of interest to reliability-based design of slopes.

**Acknowledgments**

This work was financially supported by National Natural Science Foundation of China [grant number 51809115], the Research Award Fund for Young and Middle-aged Scientists of Shandong Province [grant number ZR2016EEB11], and Qilu Construction Projects of Science and Technology in 2016 [grant number 2016B20]. These financial supports are gratefully acknowledged.

## References

Alimohammadlou, Y., Najafi, A., and Gokceoglu, C. (2014). Estimation of rainfall-induced landslides using ANN and fuzzy clustering methods: A case study in Saeen Slope, Azerbaijan province, Iran. *CATENA*, 120, 149-162.

Bye, A. and Bell, F. (2001). Stability assessment and slope design at Sandsloot open pit, South Africa. *International Journal of Rock Mechanics and Mining Sciences*, 38(3), 449-466.

Chen, W.-F. (1975). *Limit Analysis and Soil Plasticity*, Developments in Geotechnical Engineering, 7. Elsevier Scientific Publishing Co., New York.

Chen, W., Xie, X., Peng, J., Wang, J., Duan, Z., and Hong, H. (2017). GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models. *Geomatics, Natural Hazards and Risk*, 1-24.

Domingos, P. and Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 29(2-3), 103-130.

Erzin, Y. and Cetin, T. (2013). The prediction of the critical factor of safety of homogeneous finite slopes using neural networks and multiple regressions. *Computers & Geosciences*, 51, 305-313.

Feng, X. (2000). *Introduction of Intelligent Rock Mechanics*. Science Press, Beijing, China.

Ferentinou, M.D. and Sakellariou, M.G. (2007). Computational intelligence tools for the prediction of slope performance. *Computers and Geotechnics*, 34(5), 362-384.

Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29(2-3), 131-163.

Gordan, B., Jahed Armaghani, D., Hajihassani, M., and Monjezi, M. (2016). Prediction of seismic slope stability through combination of particle swarm optimization and neural network. *Engineering with Computers*, 32(1), 85-97.

Hoek, E. and Bray, J.W. (1981). *Rock Slope Engineering,* 3rd edition, Institution of Mining and Metallurgy, London.

Jensen, F.V. and Nielsen, T.D. (2007). *Bayesian Networks and Decision Graphs* , Springer, New York.

Kontkanen, P., Myllymäki, P., Silander, T., and Tirri, H. (1997). Comparing predictive inference methods for discrete domains, *Proceedings of the Sixth International Workshop on Artificial Intelligence and Statistics*, Ft. Lauderdale, USA, 311-318.

Kotsiantis, S. and Kanellopoulos, D. (2006). Discretization techniques: A recent survey. *GESTS International Transactions on Computer Science and Engineering*, 32(1), 47-58.

Lenchman, J.B. and Griffiths, D.V. (2000). Analysis of the progression of failure of the earth slopes by finite elements. *Slope Stability 2000: Proceedings of Sessions of Geo-Denver 2000*, D.V. Griffiths, G.A. Fenton and T.R. Martin (eds.), ASCE, Denver, 250-265.

Li, X. and Kong, J. (2014). Application of GA-SVM method with parameter optimization for landslide development prediction. *Natural Hazards and Earth System Sciences*, 14(3), 525-533.

Lin, P.S., Lin, M.H., Su, M.B., and Lee, T.M. (1988). An investigation on the failure of a building constructed on hillslope. *Proceedings of 2nd International Symposium on Field Measurements in Geomechanics*. A.A. Balkema, Kobe (eds.), 445-449.

Liu, Z., Shao, J., Xu, W., Chen, H., and Zhang, Y. (2014). An extreme learning machine approach for slope stability evaluation and prediction. *Natural Hazards*, 73(2), 787-804.

Lu, P. and Rosenbaum, M.S., 2003. Artificial Neural Networks and Grey Systems for the Prediction of Slope Stability. *Natural Hazards*, 30(3), 383–398.

Madzic, E. (1988). Stability of unstable final slope in deep open iron mine. *Proceedings of 2nd International Symposium on Field Measurements in Geomechanics*. A.A. Balkema, Kobe, 455-458.

Michalowski, R. (1995). Slope stability analysis: A kinematical approach. *Geotechnique,* 45(2), 283-93.

Norsys Software Corporation (1998). *Netica Application User's Guide*, Norsys Software Corporation, Vancouver, BC, Canada.

Rukhaiyar, S., Alam, M., and Samadhiya, N. (2017). A PSO-ANN hybrid model for predicting factor of safety of slope. *International Journal of Geotechnical Engineering*, 1-11.

Sah, N., Sheorey, P., and Upadhyaya, L. (1994). Maximum likelihood estimation of slope stability. *International journal of rock mechanics and mining sciences & geomechanics abstracts*, 31(1), 47-53.

Sakellariou, M. and Ferentinou, M. (2005). A study of slope stability prediction using neural networks. *Geotechnical & Geological Engineering*, 23(4), 419-445.

Samui, P. (2008). Slope stability analysis: a support vector machine approach. *Environmental Geology*, 56(2), 255-267.

Shirzadi, A., Bui, D.T., Pham, B.T., Solaimani, K., Chapi, K., Kavian, A., Shahabi, H., and Revhaug, I. (2017). Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environmental Earth Sciences*, 76(2), 1-18.

Taheri, A. and Tani, K. (2010). Assessment of the stability of rock slopes by the slope stability rating classification system. *Rock Mechanics and Rock Engineering*, 43(3), 321-333.

Ting, S.L., Ip, W.H., and Tsang, A.H.C. (2011). Is Naïve Bayes a good classifier for document classification? *International Journal of Software Engineering and Its Applications*, 5(3), 37-46.

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling*, 203(3–4), 312-318.

Wang, H., Xu, W., and Xu, R. (2005). Slope stability evaluation using back propagation neural networks. *Engineering Geology*, 80(3), 302-315.

Wu, X. and Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*, CRC Press, New York.

Xu, W., Xie, S., Jean-Pascal, D., Nicolas, B., and Imbert, P. (1999). Slope stability analysis and evaluation with probabilistic artificial neural network method. *Site Investigation Science and Technology*, 3, 19-21.

Xue, X. (2017). Prediction of slope stability based on hybrid PSO and LSSVM. *Journal of Computing in Civil Engineering*, 31(1), 1-10.

Yan, X. and Li, X. (2011). Bayes discriminant analysis method for predicting the stability of open pit slope. *Proceedings of the International Conference on Electric Technology and Civil Engineering (ICETCE)*, Lushan, China, 147-150.

Zhao, H.-b. (2008). Slope reliability analysis using a support vector machine. *Computers and Geotechnics*, 35(3), 459-467.

Zhou, K. and Chen, Z. (2009). Stability prediction of tailing dam slope based on neural network pattern recognition. *Proceedings of Second International Conference on Environmental and Computer Science (ICECS'09)*, IEEE, Dubai, the United Arab Emirates, 380-383.