ISSN 2278-3091
Volume 8, No.3, May - June 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse26832019.pdf
https://doi.org/10.30534/ijatcse/2019/26832019

Predictability of School Administrator's Job Satisfaction through Hybrid Segmentation-based Prediction Model

Elvis P. Patulin
College of Arts and Sciences, Surigao State College of Technology, Surigao City, Philippines,
elvispatulin@yahoo.com

ABSTRACT

Today, various researchers have amplified the use of data mining techniques in almost all sectors of society. In this study, the effectiveness of data mining techniques, particularly the use of K-Means segmentation in aid of increasing prediction accuracy when used to job satisfaction dataset is observed. Researches on job satisfaction have long been acclaimed as the prime index of worker's wellbeing as well as a core motivational construct that plays key implications for organizational and administrative policy. Due to the advent of technology, well-established mining and extraction of knowledge from databases are optimized. In this paper, the use of predictive models such as C4.5, DT, Naïve Bayes, OneR, KNN, and JRip algorithms was observed. However, prior to prediction, the variables from the dataset obtained from the 157 school administrator-respondents from the Department of Education, Division of Surigao del Norte, Philippines was segmented using the K-Means algorithm. This study strengthens the notion that a reduced number of variables in the dataset increases the accuracy of a prediction model. Simulation results showed that with the integration of K-Means segmentation, an accuracy of 83.54% using Decision Table algorithm is depicted from the previously identified 82.16% using the DT algorithm alone.

Key words: DepEd, K-Means, Job satisfaction, Prediction

1. INTRODUCTION

Recently, numerous researches about job satisfaction were already conducted [1]. However, despite the advent of technology, there are still researchers who used the old predefined queries and charts. They are often unlikely to accomplish a well-established knowledge extraction from databases using the old and traditional methods. The methods employed by other researchers are qualitative and quantitative methods. This is realized with the help of researcher-made questionnaires. The responses on these evaluations are tallied and are kept in databases, while interpretation is drawn using statistics and basic mathematical methods. However, the abovementioned methods are outdated, and the use of the latter limits the researchers in achieving their quality objectives. Now, with the advent of technology, a new paradigm called data mining has emerged [2].

Data mining (DM), also called Knowledge Discovery from Databases (KDD) is the process of extracting implicit information or knowledge from massive databases using data mining and machine learning algorithms [3]. Some of the renowned function in data mining used diverse approaches of DM analysis such as decision tree, Bayes classifiers, association rules, clustering, neural networks, genetic algorithms, support vector machines, predicting or forecasting and many more [4]. These approaches encourage data analysis in its full potential with the help of the right algorithms. Among the many data mining approaches, the prediction is considered as the prime method that is commonly used in educational data mining, business, health, and even in almost all sectors of the society [5].

In the era of data mining and machine learning, data reduction procedures play vital importance [6] in increasing the accuracy of a prediction model. It aims to obtain a fast, accurate, and adaptable model that quickly respond to incoming changes due to low computational complexity [7]. The goal of this paper is to cluster the variables from the survey questionnaire used on a job satisfaction survey. The respondents of the survey are the school administrators in DepEd Surigao del Norte Division, Philippines. By creating clusters, the total number of variables to be trained on DT, OneR, JRip, Naïve Bayes, C4.5, and KNN prediction algorithms are reduced in the quest to increase the prediction accuracy of the abovementioned models using the job satisfaction dataset. Different prediction models were observed, and the best model with the highest prediction accuracy after the K-Means segmentation is selected. This study is hoped to contribute on the two major areas of literature; first, on data reduction and segmentation and second, as a basis for an intervention program by the DepEd.

2. LITERATURE REVIEW

The improvement of the learning and instruction in the academe are realized through the kind of leadership the management has to offer as well as how they perform to achieve their organizational goals. Enriched working environment, a well outlined unified school program, confidence in the school system, and honest evaluation are some of the concerns in the workplace in the context of the educational setting. When all these concerns are observed from the school leaders, effective teaching performance, and a demonstration of satisfaction and content in the teaching service are manifested [8]. Also, the school's effectiveness depends on the competence and commitment of its employees, particularly the faculty and the school administrators. Being regarded as the heart of the educational process, they serve as the most valuable and essential aides of the school in delivering quality education [9].

Despite the abovementioned ideas, a notion is recommended that for quality education to become a reality, the level of satisfaction of the school administrators in terms of security, work environment, job responsibilities, and community attachments or linkages must first be established for them to be effective during their governance. The faculty and staff must not just be the focal point of all formative plans, projects, and programs but as well as the management. An exceptionally responsive curriculum and educational programs, as well as a physically well-equipped school with upgraded technology, are nothing if the school administrator, teachers, and staff do not convey their maximum potentials and capabilities. Since this school personnel assumes an indispensable job in conveying quality teaching services for quality education, it is suggested that the school management take a step up for their job satisfaction.

Researches on job satisfaction in the context of education are amplified with the use of Educational Data mining (EDM). This new powerful tool offers a platform for applications in the context of the educational environment by using the right machine learning algorithms. Educational data mining can mine educational data and extract knowledge related to learning activities within institutions. A vast empirical literature on EDM and researches of the likes are prevalent today [10].

To name some, the study of [11] employed data preprocessing techniques and algorithms such as C4.5 and K-means segmentation in examining teaching concepts that influenced student satisfaction as well as identifying predictors for effective teaching performance. The result showed that many of the attributes in the datasets are found effective as an input in the performance prediction.

A recent study of [12] applied text mining and machine learning techniques to measure job satisfaction based on employees tweeter contents. The use of Support Vector Machine (SVM) was instrumental as the classifier. 5-fold cross-validation was used for parameter tuning and obtained an accuracy of 76% for labeling relevance texts. The relatively high accuracy of the classifier is evident that determining job satisfaction based on worker's tweets is feasible.

Meanwhile, a new methodology for predicting teachers performance based on the analysis of educational surveys was proposed in the study of [13]. The use of a sequential pattern and classification technique were observed in analyzing teacher behaviors.

Lastly, the study of [14] employed statistical and data mining techniques in investigating the impact of leadership styles on leadership outcome as perceived in the survey conducted. Regression and correlation analysis were the statistical methods used in the study, while decision trees and rule-based algorithms were the data mining algorithms utilized for extracting knowledge from the datasets. Results showed that OneR algorithm, which is a type of rule-base algorithm obtained a high accuracy of 66% while Moldem obtained the highest accuracy of 91%. The output of this study was used as input for a managerial decision support system.

3. METHODOLOGY

3.1 K-Means Algorithm

The experimental result for clustering [15], [16] was implemented using KNIME (Konstanz Information Miner) [23] analytics platform. Figure 1 shows the node structure of the K-means clustering executed in KNIME. The node for the K-Means is connected and then positioned after the node of the imported CSV file of the dataset. The node color manager comes after as it put distinctions to the results to be generated later. The node scatter plot shows the scatter plot of the clusters while the interactive table is used to view the result in a table manner.

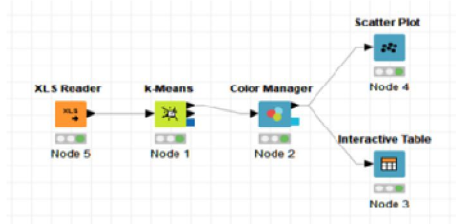

Figure 1: Node structure of K-means in KNIME

3.2 C4.5 Algorithm

According to [17], J. Ross Quinlan created the C4.5 algorithm, which is a descendant to ID3 algorithm. It is considered as the most famous decision tree algorithm in data mining. To execute, first is to compute the gain ratio of each attribute. The tree's root node is determined as the attributes whose gain ratio is at maximum value. The algorithm uses a pessimistic pruning approach in removing redundant branches of the tree in increasing accuracy of the classification method.

3.3 Decision Table (DT) Algorithm

Decision tables (DTs) uses a tabular representation for describing and analyzing situations. The decision, i.e., an action is taken depending upon the number of conditions and their inter-relationships.

3.4 Naïve Bayes Algorithm

The Naïve Bayes classifier is a probabilistic classifier based on Bayes' theorem. It attributes are fully independent [18] making it reliable and simple to use. Some of the notable advantages of the NB Algorithm includes simplicity to execute and its ability to cope up with noise and irrelevant data. For the complete formula of Naïve Bayes Algorithm, please refer to the study of [19].

3.5 OneR Algorithm

One Rule algorithm is the most straightforward learning algorithm for discrete attributes under rule-based algorithm.

3.6 KNN Algorithm

Another well-known data mining algorithm that adopts instance-based learning for prediction is the K-Nearest neighbor (KNN) algorithm, which was introduced by Fix and Hodges. The KNN is simple and can be executed by assigning k values of the nearest neighbor of an instance and perform the calculation on the Euclidian distance. The K neighboring attributes that have the lowest Euclidian distance is chosen [20].

3.7 JRip Algorithm

William Cohen proposed the JRip algorithm. This algorithm tries to add every possible rule until it becomes accurate. It Optimizes rule set using discretion length.

3.8 Datasets

In this study, a total of 157 records of school administrator-respondents in the evaluation of job satisfaction of DepEd Surigao del Norte Division's school administrators. There were forty (40) variables that represent the job satisfaction from the survey questionnaire having divided into four (4) parts viz., security, work environment, job responsibilities, and community attachment or linkages. The forty variables from the job satisfaction survey questionnaire were reduced to obtain a maximized accuracy using 10-folds cross-validation scheme. The use of the Waikato Environment for Knowledge Analysis (WEKA) was instrumental in training the dataset using different predictive models. Table 1 depicts the variables used in the study.

Table 1: Variables used

Category	Statement *On my present occupation, this is the way I feel about…*	Variable	Possible Value
Security	The salary I receive for my work.	A1	{4,3,2,1}
	The chance to be reclassified / be promoted	A2	{4,3,2,1}
	The benefits I receive are good as most other organizations can offer.	A3	{4,3,2,1}
	Efforts are not rewarded the way it should be.	A4	{4,3,2,1}
	The way my job provides a secured future	A5	{4,3,2,1}
	The way I get recognition for the things I do	A6	{4,3,2,1}
	Being proud to the job I do.	A7	{4,3,2,1}
	The way how my pay compares with that for similar jobs in other companies	A8	{4,3,2,1}
	The way how my pay compares with other coworkers in school.	A9	{4,3,2,1}
	The opportunities for advancement are ably supported.	A10	{4,3,2,1}
Work Environment	The policies & practice towards employees of the school.	B1	{4,3,2,1}
	The way my immediate head & I understand each other.	B2	{4,3,2,1}
	The spirit of cooperation among my co-workers	B3	{4,3,2,1}
	The working conditions (heating, lighting, ventilation etc.)	B4	{4,3,2,1}
	The way my colleagues make friends with.	B5	{4,3,2,1}
	The way my immediate head trains his/ her subordinates.	B6	{4,3,2,1}
	The contentment I get from the job.	B7	{4,3,2,1}
	The way my immediate head takes care of the complaints of his / her employees.	B8	{4,3,2,1}
	The pleasantness of the working conditions,	B9	{4,3,2,1}
	The way my immediate provides help on hard problems.	B10	{4,3,2,1}
Job Responsibilities	The chance to "rub elbows" with important people.	C1	{4,3,2,1}
	Being able to accomplish works that don't go against my will.	C2	{4,3,2,1}
	The chance to do work that well suited to my abilities.	C3	{4,3,2,1}
	The chance to tell other co-workers how to do things.	C4	{4,3,2,1}
	The chance to try something different in my job.	C5	{4,3,2,1}
	The chance to do something that makes use of my abilities.	C6	{4,3,2,1}
	The chance to develop new and better ways to do the job.	C7	{4,3,2,1}
	The chance to do things that don't harm my other co-workers.	C8	{4,3,2,1}
	The freedom to use my own judgment.	C9	{4,3,2,1}
	The chance to the job without the feeling I am cheating anyone.	C10	{4,3,2,1}
Community Attachment / Linkages	The opportunity to have a secured place in the community.	D1	{4,3,2,1}
	The opportunity to be of help to other people.	D2	{4,3,2,1}
	The chance to encourage the stakeholders' participation in all school-related activities.	D3	{4,3,2,1}
	The chance to be somebody in the community.	D4	{4,3,2,1}
	The chance to do the community outreach programs (i.e., linis barangay, coastal clean-up, tree planting.	D5	{4,3,2,1}
	The chance to help people's concern in the community.	D6	{4,3,2,1}
	The linkages of the school in the immediate community.	D7	{4,3,2,1}
	The way my immediate head takes care of the complaints of some parents in the community.	D8	{4,3,2,1}
	The pleasantness of the school community towards external stakeholders.	D9	{4,3,2,1}
	The social position in the community that goes with the job.	D10	{4,3,2,1}

The Division of Surigao del Norte is comprised of 11 mainland municipalities of the second congressional district. It has 13 districts namely Alegria, Anao-aon, Bacuag, Claver, Gigaquit, Mainit I, Mainit II, Malimono, Placer I, Placer II, Sison, Taganaan, and Tubod. It has 158 public and three private elementary schools, and 35 public and six private secondary schools where the 157 school administrator-respondents are employed. Table 2 shows the profile of the school administrator-respondents in terms of sex, age, years in service, highest educational attainment, and item position.

Table 2: Profile of the respondents

Profile		f(n=157)	Percent
Sex	Male	60	38.2
	Female	97	61.8
Age	39 or below	53	33.8
	40-49	79	50.3
	50 and above	25	15.9
Years in Service	0-5	52	33.1
	6-10	77	49.0
	11 and above	28	17.8
Highest Educational Attainment	Baccalaureate Degree Holder or with Master's Units	87	55.4
	Master’s degree holder	53	33.8
	At least Phd/EdD Units Earner	17	10.8
Item Position	SIC or HT1	49	31.2
	HT2	17	10.8
	HT3	46	29.3
	P1	29	18.5
	P2 or P3	16	10.2

Out of 157 school administrator-respondents, there are 60 or 38.2% who are males and 97 or 61.8% are females. There are 53 or 3.8% who are 39 years old or below, 79 or 50.3% who are 40-49 years old, and 25 or 15.9% who are 50 years old and above. There are also 52 or 33.1% who have been in the service for 0-5 years, 77 or 49% for 6-10 years, and 28 or 17.8% for 11 years and above. In terms of their highest educational attainment, there are 87 or 55.4% who have Master’s units or just a Baccalaureate degree, 53 or 33.8% who are Master’s degree holder, and 17 or 10.8% who are at least PhD/EdD units earners. For their item position, there are 49 or 31.2% who are either school-in-charge of Head Teacher I, 17 or 10.8% who are Head Teacher II, 46 or 29.3% who are Head Teacher III, 29 or 18.5% who are Principal I, and 16 or 10.2% who are either Principal II or Principal III.

4. RESULTS AND DISCUSSION

4.1 Variable Segmentation using K-Means Algorithm

The first step of the process before the prediction is by clustering the variables into two. This is to divide the variables in the dataset to obtain an increase in the prediction. The KNIME analytics was instrumental for the cluster analysis. Variables within each cluster share almost identical attributes. Table 3 shows the groupings of the variables generated by KNIME analytics.

Table 3: Cluster results using KNIME

Variable	Mean Value		Cluster
	Male Respondent	Female Respondent	
A1	3.13230	3.1323	cluster_1
A2	3.0117	3.32296	cluster_2
A3	2.579	3.39300	cluster_2
A4	2.9222	3.2933	cluster_2
A5	3.1712	3.05444	cluster_1
A6	3.2763	3.13333	cluster_1
A7	3.24903	3.29333	cluster_1
A8	3.0267	3.1666	cluster_1
A9	3.4133	3.29333	cluster_1
A10	3.2667	3.2763	cluster_1
B1	3.23333	3.24903	cluster_1
B2	3.3151	3.1284	cluster_1
B3	3.39300	3.48249	cluster_1
B4	3.46667	3.4786	cluster_1
B5	3.1323	3.32685	cluster_1
B6	3.32296	3.1634	cluster_1
B7	3.39300	3.08949	cluster_1
B8	3.13230	3.25	cluster_1
B9	3.16667	3.05837	cluster_1
B10	3.08667	3.0856	cluster_1
C1	3.43333	3.3618	cluster_1
C2	3.26	3.2451	cluster_1
C3	3.4	3.23346	cluster_1
C4	3.28	3.2607	cluster_1
C5	3.20667	3.15953	cluster_1
C6	3.14	3.2179	cluster_1
C7	3.1267	3.0778	cluster_1
C8	3.1	3.1245	cluster_1
C9	2.5798	3.39300	cluster_2
C10	2.9212	3.29333	cluster_2
D1	3.1712	3.05444	cluster_1
D2	3.2763	3.13333	cluster_1
D3	2.5798	3.3930	cluster_2
D4	2.9222	3.29333	cluster_2
D5	3.1712	3.0544	cluster_1
D6	3.1333	2.9222	cluster_1
D7	3.39300	3.1712	cluster_1
D8	3.13333	3.05444	cluster_1
D9	3.39300	2.9222	cluster_1
D10	3.13333	3.0117	cluster_1

Simulation results showed that variables A2, A3, A4, C9, C10, D3, and D4 shares identical traits as depicted by the algorithm; hence, they are grouped as one. From the total of 40 variables, 33 of them falls under cluster 1 while the rest belong to cluster 2. Therefore, feeding the job satisfaction dataset containing variables that are grouped in cluster 1 and cluster 2 will be conducted.

Table 4: Indexed cluster analysis

Clusters	List of Variables	Total Number of Variables
1	A1, A5, A6, A7, A8, A9, A10, B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, C1, C2, C3, C4, C5, C6, C7, C8, D1, D2, D5, D6, D7, D8, D9, D10	33
2	A2, A3, A4, C9, C10, D3, D4	7

4.2 Prediction Model Accuracy Evaluation

The use of 10-folds cross-validation was instrumental in predicting the accuracy of job satisfaction dataset. The predictive capability of C4.5, DT, NB, OneR, KNN, and JRip algorithms was also tested without the K-Means segmentation, as shown in Table 5. The result showed that the highest prediction accuracy, with 82.16%, was depicted using the DT algorithm. Table 6 shows the simulation result in the prediction using the data containing the variables from cluster 1 with 157 instances. The result showed that hybrid K-means and Decision Table (DT) algorithms obtained the highest prediction accuracy of 83.54%. The prediction with k-means segmentation using the variables from cluster 1 increased the prediction accuracy of every model. Lastly, Table 7 shows the simulation result when prediction using variables from cluster 2 was used along with the identified prediction models. The simulation result showed that the highest prediction accuracy is obtained using hybrid k-means and JRip algorithms with 82.61%.

Table 5: Prediction model accuracy evaluation without K-means segmentation

Model	Accuracy %	Precision	Recall	F-Measure
C4.5	80.8917 %	0.673	0.809	0.735
DT	**82.1656 %**	-	**0.822**	-
Naïve Bayes	74.5223 %	0.730	0.745	0.737
OneR	81.5287	0.674	0.815	0.738
KNN	71.9745 %	0.720	0.720	0.720
JRip	80.8917	0.673	0.809	0.735

Table 6: Prediction model accuracy evaluation without K-means segmentation

Model	Accuracy %	Precision	Recall	F-Measure
C1+C4.5	82.1656 %	0.833	0.822	0.82
C1 + DT	**83.541 %**	-	**0.833**	-
C1+ Naïve Bayes	76.4331 %	0.740	0.764	0.751
C1+ OneR	82.193 %	0.81	0.815	0.83
C1+ KNN	72.584 %	0.740	0.740	0.723
C1 + JRip	82.6138	0.871	0.82	0.82

Table 7: Prediction model accuracy evaluation without K-means segmentation

Model	Accuracy %	Precision	Recall	F-Measure
C2+C4.5	81.5287 %	0.674	0.815	0.738
C2 + DT	82.1656 %	-	0.82	-
C2+ Naïve Bayes	80.2548 %	0.711	0.803	0.742
C2+ OneR	81.5287 %	0.674	0.815	0.738
C2+ KNN	71.0637 %	0.696	0.701	0.698
C2 + JRip	**82.8917 %**	**0.893**	**0.809**	**0.875**

5. CONCLUSION

Through this study, various prediction algorithms were integrated with k-means segmentation and have proven to be effective in elevating the accuracy rate of the used prediction models. A maximized accuracy is evident in the simulation results making this study a success. The result showed that the hybrid k-means and decision table algorithms are the optimal models for prediction using the job satisfaction dataset as perceived by the school administrators in the Department of Education, Surigao del Norte Division, Philippines.

REFERENCES

[1] A. D. Marmol, "Dimensions of Teachers' Work -Life Balance and School Commitment : Basis for Policy Review," *Ioer Int. Multidiscip. Res. J.*, vol. 1, no. 1, pp. 110–120, 2019.

[2] K. Rajalakshmi, S. S. Dhenakaran, and N. Roobini, "Comparative Analysis of K-Means Algorithm in Disease Prediction," *Int. J. Sci. Eng. Technol. Res.*, vol. 4, no. 7, pp. 2697–2699, 2015.

[3] A. J. P. Delima, A. M. Sison, and R. P. Medina, "Variable Reduction-based Prediction through Modified Genetic Algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 356–363, 2019. https://doi.org/10.14569/IJACSA.2019.0100544

[4] E. Susnea, "Using data mining techniques in higher education," in *The 4th International COnference on Virtual Learning ICVL 2009*, 2009, vol. 1, no. 1, pp. 371–375.

[5] A. J. P. Delima, "An Experimental Comparison of Hybrid Modified Genetic Algorithm-based Prediction Models," *Int. J. Recent Technol. Eng.*, vol. 8, no. 1, pp. 1756–1760, 2019.

[6] I. Czarnowski and J. Piotr, "Data Reduction Algorithm for Machine Learning and Data Mining," pp. 276–285, 2008. https://doi.org/10.1007/978-3-540-69052-8_29

[7] S. Ramírez-Gallego, B. Krawczyk, S. García, M. Woźniak, and F. Herrera, "A survey on data preprocessing for data stream mining: Current status and future directions," *Neurocomputing*, vol. 239, pp. 39–57, 2017. https://doi.org/10.1016/j.neucom.2017.01.078

[8] G. R. Morrison, S. J. Ross, J. R. Morrison, and H. K. Kalman, *Second Edition*, vol. 2. 2019.

[9] L. P. Antiojo, "Employability of Education Graduates of Cavite State University Naic," *Soc. Sci. Humanit. J.*, vol. 2, no. 4, pp. 423–431, 2018.

[10] R. Ahuja, A. Jha, R. Maurya, and R. Srivastava, "Analysis of Educational Data Mining," *Adv. Intell. Syst. Comput.*, pp. 897–907, 2018. https://doi.org/10.1007/978-981-13-0761-4_85

[11] A. Al-Masri, "Experiences in Mining Educational Data to Analyze Teacher's Performance: A Case Study with High Educational Teachers," *Int. J. Hybrid Inf. Technol.*, vol. 10, no. 12, pp. 1–12, 2017. https://doi.org/10.14257/ijhit.2017.10.12.01

[12] L. Hickman, K. Saha, M. De Choudhury, and L. Tay, "Automated Tracking of Components of Job Satisfaction via Text Mining of Twitter Data," in *ML Symposium, SIOP*, 2019, pp. 1–11.

[13] J. Barracosa and C. Antunes, "Anticipating Teachers' Performance," in *KDD Workshop: Knowledge Discovery in Educational Data*, 2011, pp. 77–82.

[14] W. Ahmad, M. Akhtaruzamman, U. Zahra, C. Ohri, and B. Ramakrishnan, "Investigation on the Impact of Leadership Styles Using Data Mining Techniques," *Leadership*, pp. 139–155, 2018. https://doi.org/10.5772/intechopen.78660

[15] S. B. B, K. K. V, and A. N. Ahmed, "Semantically enriched Tag clustering and image feature based image retrieval system," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1, pp. 138–141, 2019.

[16] J. Goyal and B. Kishan, "Progress on Machine Learning Techniques for Software Fault Prediction," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 2, pp. 305–311, 2019. https://doi.org/10.30534/ijatcse/2019/33822019

[17] S. K. Yadav and S. Pal, "Data Mining : A Prediction

for Performance Improvement of Engineering Students using Classification," *World Comput. Sci. Inf. Technol. J. WCSIT*, vol. 2, no. 2, pp. 51–56, 2012.

[18] D. Soria, J. M. Garibaldi, F. Ambrogi, E. M. Biganzoli, and I. O. Ellis, "A 'non-parametric' version of the naive Bayes classifier," *Knowledge-Based Syst.*, vol. 24, no. 6, pp. 775–784, 2011. https://doi.org/10.1016/j.knosys.2011.02.014

[19] W. Chen, X. Yan, Z. Zhao, H. Hong, D. T. Bui, and B. Pradhan, "Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBFNetwork models for the Long County area (China)," *Bull. Eng. Geol. Environ.*, pp. 1–20, 2018. https://doi.org/10.1007/s10064-018-1256-z

[20] M. Huang, R. Lin, S. Huang, and T. Xing, "A novel approach for precipitation forecast via improved K-nearest neighbor algorithm," *Adv. Eng. Informatics*, vol. 33, pp. 89–95, 2017. https://doi.org/10.1016/j.aei.2017.05.003