



A Data Mining-driven Model for Job Satisfaction Prediction of School Administrators in DepEd Surigao del Norte Division

Ronita E. Talingting

Vice President for Academic Affairs, Surigao State College of Technology, Surigao City, Philippines,
talingting2013@gmail.com

ABSTRACT

Among managerial researches, job satisfaction has long been acclaimed as the prime index of worker's wellbeing as well as a core motivational construct that has key implications for organizational and administrative policy. Talking about administrative policy, some existing problems identified in every organization especially in schools and institutions are the proliferation of data and how those data will be an aid for decision making and intervention programs. To bridge the gap, research is the viewed solution. However, despite the advent of technology, there are still researchers who rely on old methods and still used predefined queries and charts. They are often unlikely to accomplish a well-established knowledge extraction from databases using the old and traditional methods. As the regarded solution, the use of data mining techniques in the educational context is encouraged as it optimized the process of data extraction and knowledge discovery. In this paper, the use of predictive models such as Naïve Bayes, C4.5, and KNN algorithms were observed. The simulation result shows that the C4.5 algorithm is the optimal model for prediction of job satisfaction as perceived by the 157 school administrator-respondents from the Department of Education, Division of Surigao del Norte, Philippines. An accuracy of 80.89%, 74.52%, and 71.97% was obtained using C4.5, Naïve Bayes, and K-Nearest Neighbor (KNN) algorithms, respectively.

Key words: C4.5 algorithm, DepEd, Job satisfaction, Prediction

1. INTRODUCTION

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 [1]. The progress on the learning situation and instruction are realized through the kind of leadership the management has to offer as well as how they perform to achieve their organizational goals. The development of a unified school program, the enriching of working environment by taking into consideration the emotional atmosphere of every employee, collaborative workgroup, and 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 evidently observed from the school leaders, effective teaching performance and a demonstration of satisfaction and content in the teaching service are manifested. In addition, the school administrators contribute much to the continuing productivity and self-preservation of teachers especially when teachers are recognized by the administration; fortifying educator's fulfillment from their doled out assignments and obligations [2].

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 state of the art 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 the conveyance of the quality teaching services for the purpose of quality education, the school management must take steps to increase their job satisfaction.

Recently, numerous researches about job satisfaction were conducted [3]. The methods employed by other researchers are mostly grounded on predefined queries and charts. The qualitative and quantitative methods are some of the existing procedures used to create knowledge from the predefined queries out 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 in the quest to effectively mine and explore raw data and discover knowledge from huge databases [4].

Data mining (DM) or also called as Knowledge Discovery from Databases (KDD) is the process of extracting implicit information or knowledge from huge databases through the use of data mining and machine learning algorithms. Some of the renowned function in data mining includes association, classification, clustering, and prediction, among others. These approaches encourage data analysis in its full potential with the help of 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].

This study aimed to show the effectiveness of DM analysis in predicting the accuracy of the responses of the school administrators and teachers obtained from the actual research survey conducted. The actual research survey used and employed a researcher-made questionnaire on the evaluation of job satisfaction of school administrators and teachers in the Department of Education (DepEd) – Surigao del Norte division. This study is hoped to contribute on the two major areas of literature; first, on the educational data mining to how the data of survey responses can be processed and mined and second, as a basis for an intervention program by the DepEd. In this study, the famous Naïve Bayes, C4.5, and K-Nearest Neighbor (KNN) Algorithms were used as predictive models. Different prediction models were observed, and the best model with the highest prediction accuracy was selected from it.

2. LITERATURE REVIEW

Educational Data mining (EDM) [6] is a new powerful tool offering solid platform for applications in the context of the educational environment. Educational data mining can mine educational data and extract knowledge related to learning activities within institutions. A wide empirical literature on EDM and researches of the likes are prevalent today.

To name some, the study of [7] employed data preprocessing techniques and algorithms such as C4.5 and K-means [8], [9] 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 [10] 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 [11]. The use of a sequential pattern and classification technique were observed in analyzing teacher behaviors.

Lastly, the study of [12] employed statistical and data mining techniques in investigating the impact of leadership styles such as transactional, transformational and laissez-faire 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 Datasets

In this study, the responses of 157 school administrators from the different schools under the Department of Education – Surigao del Norte Division were utilized. 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 1 shows the profile of the school administrator-respondents in terms of sex, age, years in service, highest educational attainment, and item position.

Table 1: 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.

To predict the accuracy of job satisfaction of the school administrators, their responses from the forty (40) items survey questionnaire having divided into four (4) parts viz., security, work environment, job responsibilities, and community attachment or linkages were utilized using the 10-folds cross-validation scheme. The software called WEKA was instrumental in training the dataset using different predictive models. Table 2 depicts the variables used in the study.

Table 2: Variables used

Category	Statement	Variable	Possible Value
	<i>On my present occupation, this is the way I feel about...</i>		
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}

3.2 Naïve Bayes Algorithm

The Naïve Bayes classifier is a probabilistic classifier based on Bayes’ theorem. It attributes are fully independent [13] 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 [14].

3.3 C4.5 Algorithm

According to [15], 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. Attributes whose gain ratio is at maximum will be identified as the tree’s root node. The algorithm uses a pessimistic pruning approach in removing redundant branches of the tree in increasing accuracy of the classification method

3.4 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 [16].

4. RESULTS AND DISCUSSION

The use of Waikato Environment for Knowledge Analysis (WEKA) was instrumental to the training of datasets using the three prediction models, namely Naïve Bayes, C4.5, and KNN algorithms. To test the predictive capability of each algorithm, the dataset was loaded using the 10-folds cross-validation scheme. The prediction model with the highest prediction accuracy is identified as the optimal model for the prediction and evaluation of responses on the survey conducted for job satisfaction evaluation as perceived by the school administrators in the DepEd. Table 3 shows the comparative results of prediction models with corresponding percentage accuracy, precision, recall, and f-measures.

Table 3: Prediction model accuracy evaluation on the level of job satisfaction as perceived by the school administrators

Model	Accuracy %	Precision	Recall	F-Measure
Naïve Bayes	74.5223 %	0.730	0.745	0.737
C4.5	80.8917 %	0.673	0.809	0.735
KNN	71.9745 %	0.720	0.720	0.720

The result showed that among the three famous prediction models, the optimal model for prediction of school administrators’ job satisfaction is the C4.5 algorithm with the accuracy of 80% followed by the Naïve Bayes algorithm.

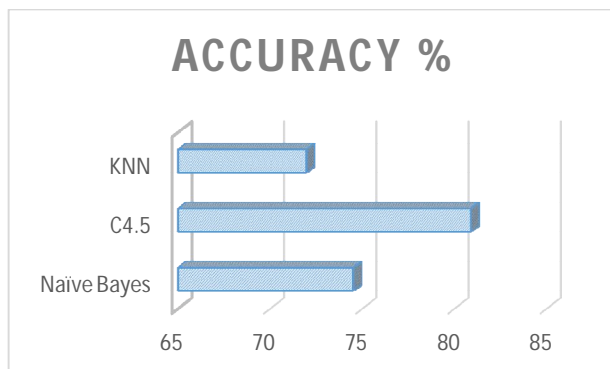


Figure 1: Prediction model accuracy

Figure 1 shows the graphical representation of the comparative models with their corresponding prediction percentage accuracy.

5. CONCLUSION AND RECOMMENDATION

Through this study, a comparative approach among the three famous prediction models was conducted. To predict the accuracy of the responses of school administrators in the study performed in relation to job satisfaction of school administrators in the Department of Education – Division of Surigao del Norte, the Naïve Bayes, C4.5, and KNN algorithms were utilized. Simulation results showed that the C4.5 algorithm remains to be the optimal model used for prediction may it be used in huge or fewer datasets. The algorithm depicted a high 80.89% prediction accuracy compared to the other two algorithms. The prediction accuracy of Naïve Bayes, C4.5 and KNN algorithms is already acceptable when the job satisfaction datasets are used; but with the advent of preprocessing techniques such as variable reduction since forty variables are quite a number to be trained, an increase on the prediction is hoped to be established for future researches.

REFERENCES

- [1] L. P. Antiojo, “Employability of Education Graduates of Cavite State University Naic,” *Soc. Sci. Humanit. J.*, vol. 2, no. 4, pp. 423–431, 2018.
- [2] G. R. Morrison, S. J. Ross, J. R. Morrison, and H. K. Kalman, *Second Edition*, vol. 2, 2019.
- [3] 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.
- [4] 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>
- [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] 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

- [7] 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>
- [8] 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.
- [9] 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>
- [10] 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.
- [11] J. Barracosa and C. Antunes, "Anticipating Teachers' Performance," in *KDD Workshop: Knowledge Discovery in Educational Data*, 2011, pp. 77–82.
- [12] 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>
- [13] 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>
- [14] 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>
- [15] 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.
- [16] 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>