THE DECISION TREE ANALYSIS MODEL ON A REMOTE POPULAR SCIENCE LEARNING SYSTEM FOR IN-SERVICES EDUCATION OF TEACHERS

Lung-Hsing Kuo[1], Pei-Hua Tsai[1], Tzuchau Yeh[2], Chenyu Chiang[3], & Kuo-Ming Tu[3]
[1]Center for General Education, National Kaohsiung Normal University, Taiwan (R.O.C.)
[2]Penghu County Jiang Jun Elementary School, Taiwan (R.O.C.)
[3]National Center for High-performance Computing (NCHC) of NARLabs, Taiwan (R.O.C.)

Abstract

The in-service training learning system for primary and secondary school teachers in Taiwan breaks through the time and space limitations and uses a distance learning method to promote primary and secondary school and kindergarten teachers on understanding the scientific development, cultivating scientific knowledge literacy and life-long learning habits, and then giving feedback to students, making popular science knowledge activities a part of teachers' in-service training and social rational cultural activities. Nowadays, Taiwan's largest teachers' in-service education distance science learning website is "Knowledge Lecture Hall" (https://knowledge.colife.org.tw/), which was built by the Co-Life team of National Center for High-performance Computing (NCHC) of NAR Labs. The purpose of this study was to establish a data repository based on the in-service training data of 5,291 teachers participating in distance science education learning in the past five years (2018-2022) collected by the Knowledge Lecture website. The data contains 16 domains of variables, including basic information about teachers and the schools they work for, online viewing time, and information about teachers' in-service training courses. In the process of knowledge discovery in the database, the decision tree algorithm C5.0 was used for modeling, the AUC value was used to evaluate the predictive ability, and the concept of tree branch was used as the decision model. According to the class and relationship of variables, the explanation model of good prediction results was found, and the following research objectives were achieved as follows: First, analyze the preferences and trends of in-service teachers' participation in learning courses during the past five years (2018-2022). Second, explore the decision tree model of the key predictors of in-service faculty participation in learning courses in the Knowledge Forum over the past five years (2018-2022). Finally, recommendations for improvement based on the findings are provided as a reference for future management of in-service distance learning courses on the Knowledge Lecture website and for future research by future researchers.

Keywords: *Decision tree, remote education, teacher in-service training.*

1. Introduction

Improvements in distance learning systems are an excellent way to improve the professional development of teachers. Online popular science course learning can promote the popularization of science more quickly, so as to break through the space-time dilemma of learning, promote the understanding of scientific development of primary and secondary school teachers and kindergarten teachers, cultivate scientific knowledge literacy and lifelong learning habits, and give back to students. In this study, the Co-Life team of the NCHC together with the National Science Council and other academic institutions, built a "Knowledge Lecture Hall". In addition to providing teacher learning certification courses through the speaker portal, the site also records in-service learning information for teachers participating in digital learning activities. This study will be used as a reference for the Knowledge Lecture Hall website for advance distance learning courses for in-service teachers, and for additional discussions for future researchers.

With the advent of the big data era, besides relying on continuous advances in information devices, it is even more critical to obtain valuable information and transform it into knowledge through the exploration of massive data. With data mining models, we can discover hidden knowledge in the collected data. The key, however, is not the amount of data, but the goal and direction of first defining the problem and then following the process of knowledge discovery, etc. Additionally, discover the key knowledge (Liao & Wen, 2019).

2. Literature review

2.1. Remote education

Remote education is a form of remote teaching and learning that breaks through the formalities of time and space. This mode of learning with the help of networks continues to flourish around the world. With a growing variety of curriculum topics available online, the range of flexible and affordable educational options is growing. In fact, distance learning is full of numerous advantages over traditional modes of teaching. At this time, the domain of distance learning and distance education lacks such a precise vocabulary. King et al. (2001) hopes to start the movement toward a common vocabulary by offering precise definitions of distance learning and distance education, and their interrelationship. Both are treated in this study as tele-education.

Yueh & Liang (2015) pointed out that the development of ICT makes teaching and learning increasingly diversified, and the application of technology in digital learning, mobile and ubiquitous learning, cloud or open course is all aimed at facilitating the realization of supporting specific teaching implementation and providing different oriented learning resources.

As public administration programs extend their online education offerings to reach more time- and place-bound students, and as accredited institutions become interested in documenting teaching and learning effectiveness, the degree to which online students are successful as compared to their classroom counterparts is of interest to teaching faculty and others charged with assessment (Ni, 2013).

2.2. Decision Tree

From one or more predictor variables, the decision tree aimed to predict class-dependent variables and predict relationships between cases or objects. The goal of the decision tree is to predict or explain the results of category dependent variables (Liao & Wen, 2012). Decision trees are a common technique for data exploration, which is easy to understand and implement. It mainly uses the concept of tree branches as a decision model. There are two types of models, one is classification trees, which are mainly used to predict classification results. Regression trees are also used to predict the mean value of each classification. The decision tree will display a series of nodes as a tree structure, and the middle nodes of the tree (non-leaf nodes) represent their test attributes. The branches of the tree represent the classification of attributes; leaf nodes of a tree represent results obtained after classification (Tsai et al., 2021). Decision tree is a widely used classification and prediction tool at present. In the construction of the decision tree, it is based on the amount of information that is beneficial. Its advantage is that it can be simple, fast and accurate with fewer parameters set (Tsai et al., 2022).

There are numerous different algorithms for decision trees, including ID3, C4.5, C5.0, CART, CHAID, QUEST, etc. The main difference between these algorithms is that they use different attribute indices for classification. Common attribute indexes are Information Gain, Gini index and Chi square independence test (Tseng & Huang, 2017). The main indicator of ID3, C4.5 and C5.0 is information profit. The main indicator of CART is the Gini value, while the index of CHAID is Chi-square independent test (Pandya & Pandya, 2015). In this study, the C5.0 algorithm was used because the attributes of the analyzed data are categorical variables, node branches can have multiple branches, and the attribute index used information gain to determine each branch of the decision tree. Each decision was made to maximize the message gain.

2.3. Aim of the study

- Discuss recent five years (2018-2022) in the knowledge lecture hall in-service teachers to participate in seminar C5.0 model of important indicators.
- According to the results put forward suggestions for improvement, as a reference for in-service teachers' distance study course, and future researchers.

3. Study design

3.1. Definition of terms

- Attributes of the curriculum: Humanities, arts and philosophy (Attribute 1), Nature and life science and technology (Attribute 2), Social science (Attribute 3), Healthcare (Attribute 4).
- Educational background: college/university, Master, Ph.D.
- Time for surfing the Internet (week): from Sunday to Monday.
- School district: Northern group(1_N), Central group(2_M), Southern group(3_S), Eastern and offshore islands group(4_R)
- Remote area attribute: Non-remote area, Highly remote area, Remote area.

3.2. Data analysis

The decision tree will generate a tree based on the training data and predict new samples based on the trained rule. In this study, C5.0 was used as the decision tree algorithm, as shown in Table 1. The C5.0 model of in-service teachers participating in the learning course of knowledge lecture Hall in recent five years (2018-2022) was shown in Figure 1.

Table 1. Prediction effect of C5.0 Decision tree data exploration algorithm (2018-2022).

Algorithm	Accuracy	Errors	Adjusted propensity score	Boosting	Cross-validation
C5.0	59.95%	0.341	0.227	0.760	0.716

4. Results

4.1. The result of C5.0 decision tree course attribute type modeling

Model 2 was used to model the sample in this study. Table 2 and Figure 1 showed the decision tree model of in-service teachers participating in the course C5.0 in the recent five years (2018-2022) in Model 2. In-service teachers participated in Curriculum Model 2, where weekly online viewing time was 0.5200, school district was 0.2100, remote area attributes were 0.1400 and educational background was 0.1300.

Table 2. Parameter results of C5.0 decision tree modeling.

parameter	Model 1	Model 2	Model 3	Model 4
Boosting	10	10	10	10
Cross-validation	10	10	10	10
Deletion importance	75	75	75	75
Minimum record for each branch	1	2	3	4
Accuracy	55.70%	59.95%	57.01%	56.26%
Training area AUC	0.621	0.721	0.511	0.653
Testing area AUC	0.671	0.842	0.755	0.759
Importance of prediction				
Time for surfing the Internet(week)	0.1254	0.5200	0.3523	0.4822
School district	0.2195	0.2100	0.1823	0.1658
Remote area attribute	0.1223	0.1400	0.1725	0.0621
Educational background	0.1217	0.1300	0.1605	0.0648

According to Figure 1, this study predicted that curriculum attributes are the most important decision metrics. During the online viewing period of teachers' participation in the knowledge Lecture Hall website, 82.650% (4,373 people) watched the Internet on weekdays, Tuesdays, Thursdays, Fridays and weekends. The largest number of in-service teachers (60.348%) (2,639) viewed "Nature and Life technology" online. However, 17.350% (918) of the teachers participated in online training on weekdays and Wednesdays, and the "science and technology of nature and life" accounted for 46.187% (424), followed by the "social science" teachers, accounting for 46.078% (423).

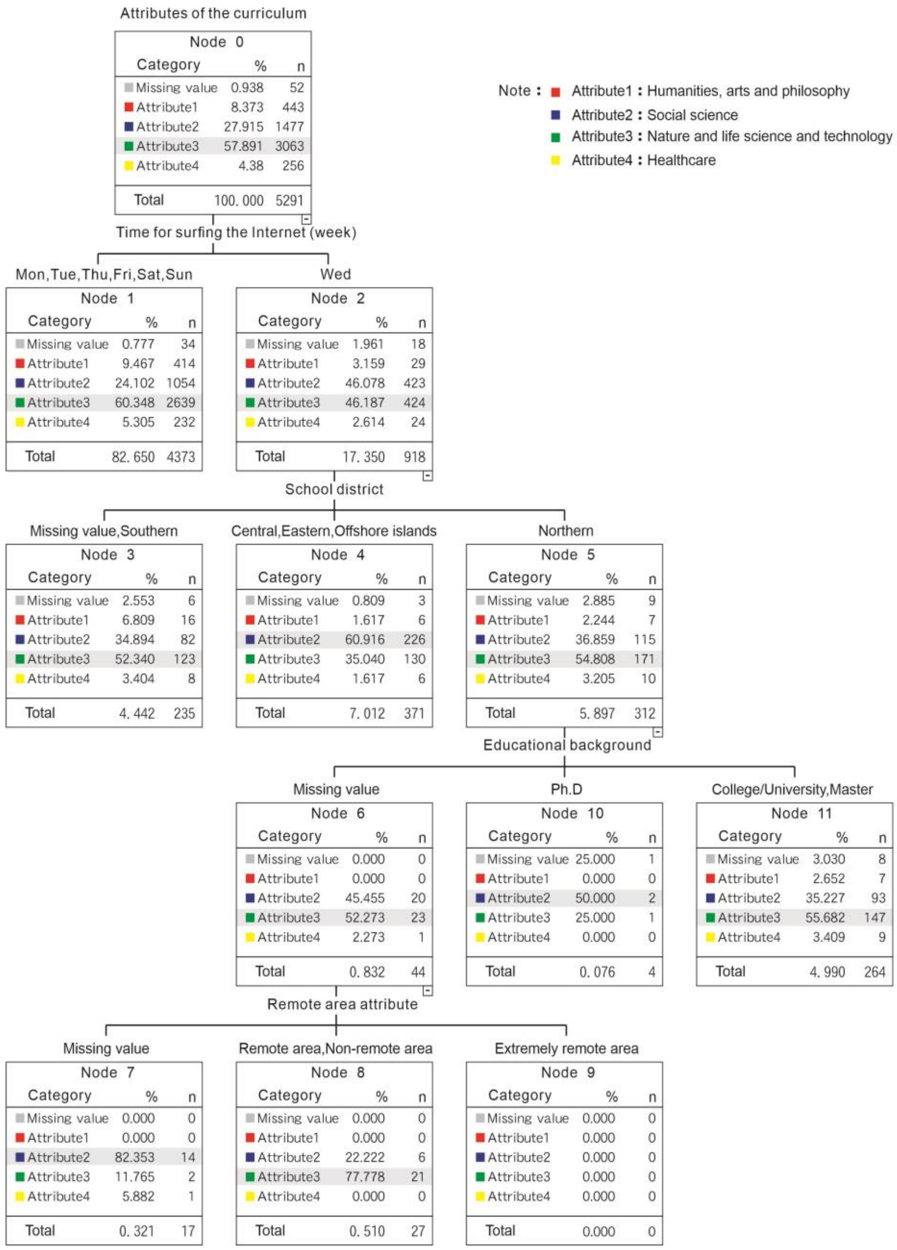
In addition, 4.442% (235) of the teachers in the southern district and 52.340% (123) of the teachers in the southern district studied remote teachers on Wednesday, while 52.340% (123) of the teachers in the southern district watched "Nature and life technology" on the Internet. 7.012% of the respondents said their school districts were located in central, Eastern and outlying islands, and 60.916% (226) said social science. 5.897 percent (312) of the respondents said their school is located in Buk-gu, and 54.808 percent (171) said science and technology in nature and life. So on Wednesday, go online to view the Nature and Life Technology category, School Districts in the North, Sustainable Forecasting for the distant Science Education Learning System C5.0 analytical model for the subsequent secondary indicators.

The proportion of in-service teachers with a doctor's degree was 0.076% (4 people), 50% (2 people), 4.990% (264 people) of in-service teachers with a university or master's degree, and 55.682% (147 people) of in-service teachers with a science and life science degree. If in-service teachers are eligible to participate in the online viewing period of the Knowledge Lecture Hall website on weekdays on Wednesdays and the school district is located in the North district, the proportion of online viewing of the course attributes of the in-service teachers is different according to their educational background. Among them, 0.832% (44) of in-service teachers whose educational background was omitted, and

52.273% (23) of them watched science and technology on the Internet. So, on Wednesday, I went online and looked up nature and life science and technology. the school district is located in the north. education background is a secondary index. missing value sustainable prediction of the distant science education learning system C5.0 analysis model.

If the online viewing time of in-service teachers participating in the Knowledge Lecture Hall website is weekdays on Wednesdays, the school area is located in the North district, the educational background is missing value, and the remote area is classified as remote or non-remote area, only 0.510% (27 people), 77.778% (21 people) watched "nature and life technology" online. 0.321% (17 people) of in-service learning teachers participated in the knowledge lecture hall website, and the remote area attribute was also omitted. 82.353% (14 people) of in-service learning teachers watched social science online. So on Wednesdays, watch nature and life science and technology classes online, with the school district to the north. Educational background is the missing value for sustainable prediction of remote science education learning system C5.0 analytical model for secondary index remote area attributes.

Figure 1. C5.0 model of in-service teachers participating in the curriculum (2018- 2022).



5. Conclusions and discussions

Online viewing time for teacher in-service training and participation in the Knowledge Lecture Hall Web site study was an important predictor in the C5.0 analysis model of the distance science education learning system.

In the C5.0 analysis model of the distance science education learning system, it was pointed out that the attributes of school area, educational background and remote area were secondary predictors, and they were also consistent with the preferences and trends of in-service teachers to participate in the course.

Regardless of whether the service school was located in a remote or non- remote area, the properties of the distance learning curricula of the in-service teachers also indicated that the teacher's concept of mobile technology learning was diverse, including the acquisition and development of concepts in natural and life technology and social sciences.

Acknowledgements

We would like to thank the NARLabs (Taiwan) Co-Life team of the National Center for High-Performance Computing (NCHC) for providing valuable information for this study.

References

King, F. B., Young, M. F., Drivere-Richmond, K., & Schrader, P. G. (2001). Defining Distance Learning and Distance Education. *AACE Review (formerly AACE Journal), 9*(1), 1-14. Norfolk, VA: Association for the Advancement of Computing in Education (AACE). May 4, 2023 Retrieved from https://www.learntechlib.org/primary/p/17786/

Liao, S., & Wen, C. (2019). *Data Mining: In Case of SPSS Modeler*. DrMaster Press.

Ni, A. Y. (2013). Comparing the Effectiveness of Classroom and Online Learning: Teaching Research Methods. *Journal of Public Affairs Education, 19*(2), 199-215.

Pandya, R., & Pandya, J. (2015). C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. *International Journal of Computer Applications*, *117*(16), 18-21. https://doi.org/10.5120/20639-3318

Tsai, H., Wei, H., & Kuo, L. (2021). Data Mining Technique for Predicting the Model of In-service Teachers' Bilingual Instruction Training. *Journal of Sciences & Humanities, 2021, 9*(1), 14-29. https://doi.org/10.6535/JSH.202111_9(1).0002

Tsai, H., Wei, H., & Kuo, L. (2022). Data Mining Technique for Constructing the Prediction Model of University Teachers' Evaluation of Schools. *Journal of Sciences & Humanities*, *9*(2), 24-38. https://doi.org/10.6535/JSH.202202_9(2).0002

Tseng, S., & Huang, S. (2017). The Application of Big Data on the Impact between Workplace Educational Training and Employees' Background - in case of Hotel Industry. *Yu Da Academic Journal, 45*, 33-49.

Yueh, H., & Liang, C. (2015). Integrating Considerations of Students, Teachers, and Instructional Contexts in a Predictive Model of Distance Education. *Journal of Educational Media & Library Sciences, 52*(1), 33-57.