Predicting Study or Work From Home Satisfaction using Data Mining Technique: A Case Study in Malaysia

Florence Sia [1], Lim Hong Yao [2], Tuen Yong Hao [3], Peter Chin [4], Lim Min Xuan [5]
Faculty of Computing and Informatics
Universiti Malaysia Sabah
Kota Kinabalu, Malaysia
[1]florence.sfs@ums.edu.my, [2]lim_hong_bi20@iluv.ums.edu.my, [3]tuen_yong_bi20@iluv.ums.edu.my,
[4]peter_chin_bi20@iluv.ums.edu.my, [5]lim_min_bi20@iluv.ums.edu.my

2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST) | 978-1-6654-8663-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/GECOST55694.2022.10010647

***Abstract*— The COVID-19 pandemic has urged the government of Malaysia to implement Movement Control Order (MCO) which forces working people to work from home while students to study from home. People's satisfaction on work from home is crucial in determining their work productivity and efficiency whereas student's satisfaction on study from home is important for their learning effectiveness. There is no work has been done yet for exploring data mining techniques to build a model for predicting work or study from home satisfaction using Malaysia as a case study. This paper aimed to identify the best data mining model for predicting the work or study from home satisfaction. The prediction model is learned by analyzing the demographic, the personality traits, and the work from home experience collected from a group of Malaysia people. This study attempts to investigate four data mining techniques that are the decision tree, linear kernel support vector machine, polynomial support vector machine, and radial basis support vector machine. Experiment results show that the radial basis support vector machine outperformed other techniques in predicting the work or study from home satisfaction of Malaysia's community.**

Keywords*— data mining, work from home, study from home, covid-19, prediction, decision tree, support vector machine

I. INTRODUCTION

In December 2019, an infectious disease called Coronavirus or Covid-19 has spread to many countries worldwide. World Health Organization (WHO) has announced Covid-19 as pandemic disease [1]. Malaysia has implemented movement control order (MCO) in order to control the spreading of the diseases among community. During MCO, people are urged to stay at home, thus business, education, and other industry sectors are forced to be operated from home. After vaccination has been introduced in early 2021, the Covid-19 case has shown a continued decrease with low fatality rate.

Although MCO has been fully lifted since early May 2022, some of the companies are still opted for work from home rotation. That is employees are required to work from home or work at office according to a work from home rotation schedule. Same to the education sector, there are some courses where the teaching and learning are still conducted through virtual class. Hence, students can study the course from home instead of having to go to the campus. However, the employee's work from home satisfaction may affect his/her work productivity and efficiency. While the student's study from home satisfaction can affect his/her learning productivity and effectiveness.

The personal traits depict a person characteristic of thoughts, behaviours, and feelings. It has been widely used for solving real world problems in many domains. For example, in hotel business, personal traits of hotel management employee have been used for predicting the attractiveness of the hotel [2]. In education, personal traits of students are used to predict the academic performance of the students [3]. The personal traits of an individual may also affect the work or study from home satisfaction. Besides that, the demographic and the work or study from home experience may contribute to the determination of work or study from home satisfaction [4]. The main aim of this paper is to identify the best data mining model for predicting the work or study from home satisfaction of Malaysia people based on the individual personal traits, demographic, and work from home experience. A dataset comprised of personal traits, demographic, and work from home experience of a group of Malaysia people which can be employees or students, are gathered and pre-processed. The personal traits refer to the four aspects of personality that are mind (introvert or extrovert), energy (intuitive or observant), nature (thinking or feeling), and tactics (judging or prospecting). Data mining techniques are then applied to learn prediction models from the dataset. Herein, four well known data mining techniques are used which including the decision tree, the linear kernel support vector machine, the radial basis support vector machine, and the polynomial kernel support vector machine.

Several experiments have been carried out to evaluate the accuracy of the work or study from home satisfaction prediction models built from the application of the different data mining techniques. Results demonstrate that the radial basis support vector machine performs well compared to other data mining techniques. This also indicates that there is correlation between the personal traits and the work or study from home satisfaction.

The rest of the paper is structured as follows. Section II presents the review of existing data mining techniques used for solving various prediction problems. Section III describes the methodology involved in learning the work or study from home satisfaction prediction model using data mining techniques from the collected dataset. Section IV presents the experimental setup, dataset description, and the analysis of experiment results. Lastly, section V presents the conclusion and some future works.

Authorized licensed use limited to: UNIVERSITY SABAH MALAYSIA. Downloaded on February 07,2023 at 08:33:57 UTC from IEEE Xplore. Restrictions apply.

II. LITERATURE REVIEW

Sheybani [5] has carried out the study of job satisfaction in United States and evaluated the factors affecting it in the studied sample. Decision tree has been used to learn a job satisfaction prediction model and concluded that the perception of workspace and the personal traits can affect the job satisfaction. Besides, the study conducted by Kuzey [6] used support vector machine to determine the factors that contribute significantly to the job satisfaction of employee in a healthcare institution which is located in Istanbul, Turkey. Four different types of kernel function are used which including radial basis function, polynomial kernel, linear kernel and sigmoid kernel. Bashir et. al [7] applied decision tree and regression to predict the grade and mark respectively, based on the students' academic historic obtained from the Federal Board of Intermediate and Secondary Education Islamabad Pakistan. The prediction model is essential to improve the quality of education in Pakistan. Shouwu et. al [8] used a combination of decision tree and random forest to forecast the employment of graduates from Guilin University of Technology. Saputra et. al [9] have implemented decision tree to predict the graduation rate of undergraduate college students based on the students' demographic, activities, and academic performance.

In addition to that, Rajni et.al [10] applied several data mining techniques to predict the susceptibility of an individual to Covid-19 mental stress by learning from COVIDiSTRESS survey data. It was found that the gradient boosting classifier achieved the highest prediction accuracy. Indy et.al [11] reported that support vector machine showed good performance in predicting satisfaction of undergraduate students on remote learning during Covid-19 pandemic. The prediction modelling is carried out on the collected data containing the students' remote working and learning experience with some addition of demographic information. Francesca et. al [12] identified that gaussian naïve bayes able to predict the employee attrition well based on real dataset obtained from IBM analytic. The dataset comprised of the employees' working life and personal characteristics.

To the best of our knowledge, there is no work has been conducted yet to investigate the most effective data mining technique for learning model used to predict the work or study from home satisfaction for Malaysia community. Based on this review, it can be seen that there is no one universe data mining technique can be apply to solve different prediction problems effectively.

III. METHODOLOGY

The methodology of this research study consists of four main phases which are data collection, data pre-processing, data mining, and evaluation. Fig. 1 illustrates the methodology of this research study.

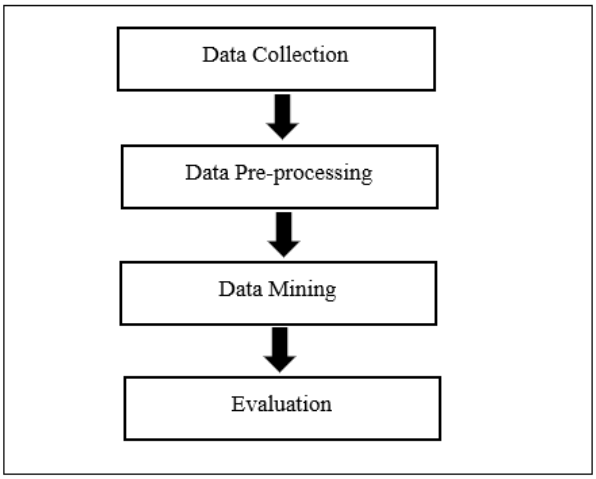

Figure 1 The Four Phases Methodology

A. Data Collection

A questionnaire is built using Google form to collect data online from random people in Malaysia who could be either employee or student. It consists three main parts of questions related to the demographic, the personal traits, and the work from home experience. A total of 100 respondents' answer were collected.

B. Data Preprocessing

From the collected data, a total of 100 records with 15 attributes are identified. One of the attributes is the class label referring to the satisfaction status that are satisfied or neutral or unsatisfied, with work or study from home. Table I presents the list of attributes. All numeric attributes are discretized and transformed onto categorical attributes. Any missing attribute value is then checked and imputed by using the mean value for the attribute at hand.

TABLE I. LIST OF ATTRIBUTES

Attribute	Value
Age	> 10 years old
Gender	Male Female
Race	All types of races in Malaysia
Origin	All states in Malaysia
Employment	Working Student
Marital Status	Single Married
Mind	Introvert Extravert
Energy	Intuitive Observant
Nature	Thinking Feeling
Tactics	Judging Prospecting
Work/Study Duration	> 30 minutes
Internet Speed	Very Fast Fast Moderate Slow
Activity Preference	Outdoor Indoor
Communication	Very Easy Easy

Authorized licensed use limited to: UNIVERSITY SABAH MALAYSIA. Downloaded on February 07,2023 at 08:33:57 UTC from IEEE Xplore. Restrictions apply.

	Moderate Difficult Very Difficult

C. Data Mining

Data mining is a process of analyzing and extracting unknown useful information from a large amount of data. It has been widely used in prediction, classification, summarization, and etc, problems [10-12]. In this study, the typically well performed data mining techniques, the decision tree and the support vector machine, are used for learning the work or study from home satisfaction prediction model from the prepared dataset [13-15]. For support vector machine, three types of kernel functions are applied to build the prediction model. They are the linear kernel support vector machine, the radial basis support vector machine, and the polynomial support vector machine.

a) Decision Tree: It learns a prediction model in a flowchart-tree structure by recursively selecting attribute value as a test attribute to partition data into nodes of records until a termination criterion is met. The top node is the root node of the tree which consists of whole data. It is partitioned into nodes called internal nodes where each node with an outgoing edge that represents the result of test. While the terminal node is also known as a leaf node. A new record is traverse through the learned tree model starting from the root node to the leaf node. The class of the new record is then predicted following the dominant class in the leaf node.

b) Support Vector Machine: It constructs a model by finding the hyperplane with maximum margin to separate records of different classess. Maximum margin gives maximum distance between records of different classes. The class label of a new record is predicted according to the respective class of hyperplane region where the record falls. Linear, polynomial, and radial basis are most commonly used kernel function in identifying the optimal hyperplane depending on the distribution of the data. Linear kernel function works for linearly separable data. While polynomial kernel and radial basis are for dealing with high dimensional and non linearly seperable data.

D. Evaluation

The prediction accuracy of the learned model from the prepared dataset using is assessed based on the two commonly used evaluation metrics namely the prediction/classification accuracy and kappa value. The prediction accuracy is the number of correct classifications divided by the total classification. Prediction accuracy with kappa value greater than 0.4 is considered as moderate to good prediction performance.

IV. Experiment Setup and Result

Several experiments have been conducted to evaluate the prediction accuracy of the data mining techniques in predicting the work or study from home satisfaction. The data mining techniques are the decision tree, linear kernel support vector machine (LK-SVM), polynomial kernel support vector machine (PK-SVM), and radial basis kernel support vector machine (RBK-SVM). The prediction modelling using the data mining techniques are implemented in R. The collected dataset is split into two partitions where 70% of data used as a training data and 30% of data taken as a testing data, as applied in many research works [16-17].

The procedure of the experiment begins by applying a data mining technique to learn a prediction model from the training data. The prediction accuracy of the resulting model on the testing data is then measured. This procedure repeats for the remaining data mining techniques.

Table II shows the prediction accuracy and kappa value of the data mining techniques in predicting the work or study from home satisfaction.

TABLE II. Prediction Accuracy (%)

Data Mining Model	Accuracy (%)	Kappa
Decision Tree	48.3	0.2
LK-SVM	75.9	0.6
PK-SVM	65.5	0.5
RBK-SVM	82.8	0.7

The results show that RBK-SVM achieved the highest accuracy, 82.8% with Kappa 0.7 in predicting the work or study from home satisfaction. LK-SVM attained 75.9% prediction accuracy whereas PK-SVM obtained 65.5% with kappa 0.5 and 0.6 respectively. On the other hand, decision tree achieved the lowest accuracy, 48.3% and kappa 0.2, which implies worst prediction performance. It is worth to note that SVM often perform well compared to decision tree regardless of the type of kernel function. RBK-SVM performs well due to its capability in handling the non-linear distribution of the collected data as it comprised of diverse personal traits. Decision tree performs poorly because it is highly depending on the attributes in distinguishing records of different classes. The prepared dataset might have irrelevant attributes that can deteriorate the performance of the decision tree. This also indicates that there is correlation between the personal traits and the work or study from home satisfaction.

V. Conclusion

Work or study from home satisfaction can affect the work or study productivity and effectiveness. Hence, by knowing the satisfaction, a right decision about the implementation of fully or partially work from home can be made especially for this prolonged Covid-19 pandemic. This paper aimed to identify the best data mining model for predicting the work or study from home satisfaction for Malaysia community. A questionnaire has been distributed to collect data related to the demographic, the personal traits, and the work from home experience. Based on the

Authorized licensed use limited to: UNIVERSITY SABAH MALAYSIA. Downloaded on February 07,2023 at 08:33:57 UTC from IEEE Xplore. Restrictions apply.

respondents' answer, a dataset is then generated consists of 15 attributes and 100 records. Data mining techniques which including the decision tree, linear kernel support vector machine, polynomial kernel support vector machine, and radial basis support vector machine, are applied to learn prediction model from the prepared dataset. Experiment results show that radial basis kernel support vector machine performs well compared to other techniques in predicting the work or study from home satisfaction. For future work, more data from student and worker will be collected and analyzed separately in order to avoid any biases. Other data mining techniques will be explored for building the prediction model. Feature selection will be performed to identify relevant attributes for improving the prediction accuracy of decision tree.

REFERENCES

[1] H. Feng, Y. Deng, and W. Li. "Coronavirus disease 2019: What we know?." Journal of medical virology 92, no. 7, pp. 719-725, 2020.

[2] B. Victoria, N. Stylos, and R. Rahimi. "Predicting hotel attractiveness via personality traits of applicants: the moderating role of self-esteem and work experience." International Journal of Contemporary Hospitality Management, 2018.

[3] R. Mahmoud, A. MK Abdel Aal, Ali O. Jifri, and N. F. Omran. "Enhancement of predicting students performance model using ensemble approaches and educational data mining techniques." Wireless Communications and Mobile Computing, 2021.

[4] Shahiri, A. Mohamed, and W. Husain. "A review on predicting student's performance using data mining techniques." Procedia Computer Science 72, pp.414-422, 2015.

[5] S. Farhad. "Predicting the Individuals' job satisfaction and determining the factors affecting it using the CHAID Decision Tree Data Mining Algorithm Case Study: the National Opinion Research Center of the United States." European Journal of Engineering and Technology Research 4, no. 3, pp. 6-9, 2019.

[6] K. Cemil. "Impact of Health Care Employees' Job Satisfaction On Organizational Performance Support Vector Machine Approach." European Journal of Economic & Political Studies 5, no. 1, 2012.

[7] Yousafzai, B. Khan, M. Hayat, and S. Afzal. "Application of machine learning and data mining in predicting the performance of intermediate and secondary education level student." Education and Information Technologies 25, no. 6 pp. 4677-4697, 2020.

[8] H. Shouwu, X.Y. Li, and J. Chen. "Application of data mining in predicting college graduates employment." In 2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD), pp. 65-69. IEEE, 2021.

[9] Saputra, J. P. Bangkit, and R. Waluyo. "Data Mining Implementation with Algorithm C4. 5 for Predicting Graduation Rate College Student." Journal of Applied Data Sciences 2, no. 3, pp. 74-83, 2021.

[10] J. Rajni, C. Kumar, G. Jawla, and H. Goyal. "Predicting Susceptibility to Covid Stress Using Data Mining." In 2022 International Conference on Emerging Smart Computing and Informatics (ESCI), pp. 1-6. IEEE, 2022.

[11] Ho, I. M. Kit, K. Y. Cheong, and A. Weldon. "Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques." Plos one 16, no. 4, 2021.

[12] F. Francesca, M. Coladangelo, R. Giuliano, and E. W. D. Luca. "Predicting employee attrition using machine learning techniques." Computers 9, no. 4, pp. 86, 2020.

[13] M. Azimi-Pour, H. Eskandari-Naddaf, and A. Pakzad. "Linear and non-linear SVM prediction for fresh properties and compressive strength of high volume fly ash self-compacting concrete." Construction and Building Materials, 230, 117021, 2020.

[14] R. Hasan and S. Palaniappan, A. R. A. Raziff, S. Mahmood, and K. U. Sarker. "Student academic performance prediction by using decision tree algorithm." In 2018 4th international conference on computer and information sciences (ICCOINS), pp. 1-5, 2018.

[15] F. Sia, and R. Alfred. "Tree-based mining contrast subspace." International Journal of Advances in Intelligent Informatics, 5(2), pp. 169-178, 2019.

[16] Z. Zhao, A. Chen, W. Hou, J. M. Graham, H. Li, P.S. Richman, and T. Q. Duong. "Prediction model and risk scores of ICU admission and mortality in COVID-19." PloS one, 15(7), 2020.

[17] E. M. M. Van der Heide, R. F. Veerkamp, M. L. Van Pelt, C. Kamphuis, I. Athanasiadis, and B. J. Ducro. "Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle." Journal of dairy science, 102(10), pp. 9409-9421, 2019.

Authorized licensed use limited to: UNIVERSITY SABAH MALAYSIA. Downloaded on February 07,2023 at 08:33:57 UTC from IEEE Xplore. Restrictions apply.