Prediction of Hyacinth Handicraft Trends in New Era

Pinthusorn Pasanajano

Department of Computer Science, Faculty of Science and Technology,
Rajamangala University of Technology Suvarnabhumi,
Nonthaburi, THAILAND
pinthusorn.p@rmutsb.ac.th

Abstract

The purpose of this research is to predict trends of Hyacinth handicraft products based on web application with make order, select products, register member, customer service, summary report and payment notification. The research method is System Development Life Cycle (SDLC) that involves many phases, including planning, design, building, testing, implementing and evaluating. The result of this research is useful for customer who want to order hyacinth handicraft products and it is better way to get their orders. The conclusion of this research showed that Decision Tree (J48) is the appropriate algorithms that can predict hyacinth handicraft trends in new era. It can be said that web application of Hyacinth handicraft is easy to use and can also promote Hyacinth handicraft all time.

Keywords: Classification, Hyacinth, Web application

I. INTRODUCTION

Water hyacinth can be made trouble for people in many countries around the world. It can also spread as fast as they can. Even in drought area that it can exist for many years. So, water hyacinth become a water weed that causes problems for water sources. It is still an obstacle to irrigation, resulting in the speed of water in the canal and river. On the other hand, their plants can be used for many products in house such as basket, tissue box, pencil case, and so on [1]. Some of hyacinth can be fed for animals.

Recently, there are many research contribute data analysis method for predict the trend [2]. In this research, we focused on creating web application to promote hyacinth handicraft products from water hyacinth. This can also sell handicraft products from water hyacinth in new era.

II. PROPOSED WORK

Machine learning algorithms are part of data mining and use to extract remarkable pattern for decision making [3]. There are supervised and unsupervised learning and they are also separated into classification and clustering. Classification function is used to predict training data set but clustering is known as unsupervised learning [4]. In this research, the selected classification is used for supervised learning. These classification algorithms are data analysis techniques to forecast the hyacinth products trend in new era.

A. Decision Tree: J48

Decision tree is one of classifications technique to consist decision nodes and leaf nodes. There are many type of decision tree such as ID3, J48, LMT, and so on. J48 is the most accuracy rate [6], thus we use it in this research.

B. OneR

OneR algorithm is one rule of classifications technique to create each attribute into training dataset. It will be shown minimum error rate [7].

C. Support Vector Machine

Support Vector Machine (SVM) is statistic algorithm and supervised learning that can be used for classification [8].

III. METHODOLOGY

In this research, data is collected from 120 participants (70 females and 50 males) and the detailed of demographics shown in these following:

TABLE 1 DEMOGRAPHICS OF PARTICIPANTS

Item	Detail	Number of
Sex	Female	70
	Male	50
Age	0-15	16
	16-30	74
	31-45	19
	46-60	11
Education	High School	38
	Bachelor degree	73
	Master degree	9
Occupation	Business	52
	Employee	16
	Government	22
	Salesperson	11

	Student	13
	Teacher	6

The data set consisted of four parts: demographics of participants (Table 1), use of content, presentation of content, display usage, and processing usage. Each part is separated into sub items such as relevance of content, the outstanding display of content, the attractive of content, the design of display, the correctness of information that propose in Table 2.

TABLE 2 ATTRIBUTE OF DATA SET

Item	Detail	Number of items
Use of content	- Relevance - Easy to use - Outstanding display - Knowledge of content	4
Presentation of content	- Suitable of content - Correct of content - Attractive content - Easy to read	4
Display usage	- Appropriate screen - Outstanding picture - Appropriate font - Design of display - Each menu is easy to use	5
Processing usage	- Easy to access information - Searching information - Level of processing speed - Correctness of information - Clarity of picture	5

The next step is prepared data set for analysed and used appropriate algorithms. There are 19 attributes, the attribute of sex is class. Then, data set (120 instances) are required for pre-processing and cleaning. All features in data set can be shown on the Visualize all (Figure 1) such as women think that hyacinth web application are easy to use, it is also presented outstanding picture, etc. Lastly, the appropriate algorithms are used that shown in the following table.

The confusion matrix is presented the actual and predicted classification and defined as True positive (TP), False positive (FP), Precision, and Recall [5].
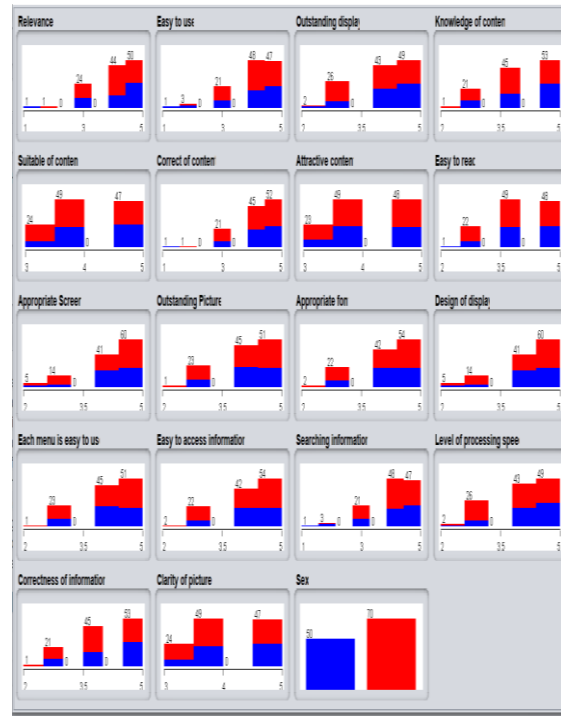

Figure 1. Visualize all

A. Application of Decision Tree: J48

When the Decision Tree (J48) algorithm is classified, it is found that women prefer appropriate font to correct of content, but men like to see more attractive content than outstanding picture. Table 3 shows classifier output of Decision Tree (J48).

TABLE 3 CLASSIFIER OUTPUT OF DECISION TREE: J48

	Positive	Negative	Total
Positive	30	20	50
Negative	15	55	70
Total	45	75	120

B. Application of OneR

OneR algorithm is proposed one attribute that relates with class and lead to the prediction. The output describes female value is less than 4.5 but male value is greater than and equal to 4.5. The correctly classified instances are 72. Table 4 shows classifier output of OneR.

TABLE 4 CLASSIFIER OUTPUT OF ONER

	Positive	Negative	Total
Positive	26	24	50
Negative	24	46	70
Total	50	70	120

C. Application of Support Vector Machine

When Support Vector Machine (SVM) is classified, it is explained the value of negative.

63

Table 5 shows classifier output of Support Vector Machine.

TABLE 5 CLASSIFIER OUTPUT OF SUPPORT VECTOR MACHINE

	Positive	Negative	Total
Positive	10	40	50
Negative	0	70	70
Total	10	110	120

As mentioned above, the Decision Tree (J48) algorithm shows the best output. True positive for class a = male is 30, False positive is 20 whereas True positive for class b = female is 55, False positive is 15. The diagonal calculate is 30+55 = 85 that represents the correctly classified instances and the other one is 20+15 = 35 (the incorrectly classified instances).

The OneR algorithm shows: True positive for class a = male is 26, False positive is 24 whereas True positive for class b = female is 46, False positive is 24. The diagonal calculate is 26+46 = 72 that represents the correctly classified instances and the other one is 24+24 = 48 (the incorrectly classified instances).

The Support Vector Machine algorithm shows: True positive for class a = male is 10, False positive is 40 whereas True positive for class b = female is 70, False positive is 0. The diagonal calculate is 10+70 = 80 that represents the correctly classified instances and the other one is 40+0 = 40 (the incorrectly classified instances).

TABLE 6 THE COMPARISON PERFORMANCE OF ALGORITHMS

No.	Parameters	J48	OneR	SVM
1	Percent of correctness	70.83%	60%	66.67%
2	Correctly Classified Instances	85	72	80
3	Incorrectly Classified Instances	35	48	40
4	Kappa statistic	0.3913	0.1771	0.2258
5	Mean absolute error	0.3748	0.4	0.3333
6	Root mean squared error	0.4329	0.6325	0.5774
7	Relative absolute error	77.06%	82.25%	68.54%
8	Root relative squared error	87.80%	128.29%	117.11%

IV. CONCLUSION

The conclusion of this research showed that the overall correctly classified instances are 70.83 % (Decision Tree: J48), 60% (OneR), and 66.67% (SVM). Thus, Decision Tree (J48) is the appropriate algorithms that can predict hyacinth handicraft trends in new era.

REFERENCES

[1] Rakotoarisoa, T. F., Waeber, P. O., Richter, T., & Mantilla-Contreras, J. (2015). Water hyacinth (Eichhornia crassipes), any opportunities for the Alaotra wetlands and livelihoods?. *Madagascar Conservation & Development*, *10*(3), 128-136.

[2] Aibinu, A. M., Salami, M. J. E., & Shafie, A. A. (2010, November). Application of modeling techniques to diabetes diagnosis. In *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)* (pp. 194-198). IEEE.

[3] Farhad, A., & Sanjay, P. (2017). Comparative Study of J48, Naive Bayes and One-R Classification Technique for Credit Card Fraud Detection using WEKA [J]. *Advances in computational sciences and technology*, *10*(62), 1731-1743.

[4] Syarif, I., Prugel-Bennett, A., & Wills, G. (2016). SVM parameter optimization using grid search and genetic algorithm to improve classification performance. *Telkomnika*, *14*(4), 1502.

[5] Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, *5*(2), 1.

[6] Jain, Y. K. (2012). Upendra,"An Efficient Intrusion Detection based on Decision Tree Classifier Using Feature Reduction,". *International Journal of Scientific and Research Publication*, *2*(1), 1-6. [4] Muntean, M., & Diana, T. (2009). *Information technology & organizational knowledge management*.

[7] Buddhinath, G., & Derry, D. (2006). A simple enhancement to one rule classification. *Department of Computer Science & Software Engineering. University of Melbourne, Australia*, 40.

[8] Al Amrani, Y., Lazaar, M., & El Kadiri, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, *127*, 511-520.