Development of the Edible and Poisonous Mushrooms Classification Model by using the Feature Selection and the Decision Tree Techniques

Article *in* International Journal of Engineering and Advanced Technology · December 2019

DOI: 10.35940/ijeat.B4115.129219

Sumitra Nuanmeesri
Suan Sunandha Rajabhat University

W. Sriurai
Ubon Ratchathani University

Development of the Edible and Poisonous Mushrooms Classification Model by using the Feature Selection and the Decision Tree Techniques

Sumitra Nuanmeesri, Wongkot Sriurai

Abstract**:** ***This research aims to develop a classification model for edible and poisonous mushrooms by applying the feature selection approach together with the decision tree technique. Two feature selection methods were applied, including 1) Chi-square and 2) Information Gain, while the effectiveness of the model was compared by three decision tree methods such as Iterative Dichotomiser3, C4.5 and Random Forest. The data used for classifying the edible and poisonous mushrooms derived from the Encyclopedia of Thai mushrooms and the book entitled "Diversity of Mushrooms and Macrofungi in Thailand". The results of the model's effectiveness evaluation revealed that the model using the Information Gain technique alongside with the Random Forest technique provided the most accurate classification outcomes at 94.19%; therefore, this model could be further applied in the future studies.***

Keywords**:** ***Classification, Feature Selection, Decision Tree, Mushrooms, Poisonous.***

I. INTRODUCTION

Humans have appreciated and consumed mushrooms throughout the history. Apart from being used as cooking ingredients, mushrooms also help the ecosystem by decomposing plant residues and animal droppings; they are beneficial to the growth of plants and reduce the amount of waste created by plants and animals. In general, there are mushrooms depending on other living organisms or nutrients from various types of plant roots in the nature; these mushrooms can be either edible or poisonous [1]. In Thailand, mushrooms have been consumed for centuries. Statistically, a number of Thai citizens have died or gotten sick because of poisonous mushrooms each year. Nonetheless, mushrooms are a source of food which can be easily found and is full of nutrients, so it is better to resolve the problem by promoting safe consumption. There should be public communication on poisonous mushrooms to help the population differentiate them from edible mushrooms. Presently, the information about mushrooms available on most websites focuses on the characteristics of mushrooms and whether they are poisonous or not, but it still takes time for the consumers to identify poisonous mushrooms.

Revised Manuscript Received on December 30, 2019.
* Correspondence Author

SumitraNuameesri*, Assistant Professor, Department of Information Technology, Science and Technology, Suan Sunandha Rajabhat University, Thailand. Email: Sumitra.nu@ssru.ac.th

WongkotSriurai, Assistant Professor, Department of Mathematics Statistics and Computer Science Ubon Ratchathani University, Thailand. Email: wongkot.s@ubu.ac.th

Data mining is a popular technique used for analyzing big data. It is the process of sorting through large data sets to identify patterns and establish relationships within them. expert system for diagnosing rice diseases was developed by applying the C4.5 decision tree technique to create a model analyzing 16 rice diseases; according to the test results of the model, the model's analysis achieved 94.5% accuracy when compared to the experts' analysis [2].

To tackle the problem of mushroom consumption based on the review of literatures, the research team has developed a model classifying edible and poisonous mushrooms by applying the feature selection approach and the decision tree technique. This model is expected to help the users to differentiate edible mushrooms from poisonous mushrooms faster.

II. RELATED WORKS

A. Feature selection

Feature selection [3] is a key procedure in the data mining process since it is likely to increase the accuracy of a classification model. There are different feature selection methods, but they all select the most significant data sets. These data sets fasten the model synthesization and enhance the effectiveness of data selection. In this study, two feature selection methods were selected, including:

- *Chi-square*(X^2):[3] is a feature selection method which compares the values of the features andX_{ij}derived from the calculation of the entire data sets in order to find the significance level of the variables of each feature. The final value indicates the significance order of the features as formulated in (1).

$$X^2 = \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{\left(A_{ij}-X_{ij}\right)^2}{X_{ij}} \qquad (1)$$

Where$X_{ij} = \frac{R_i x\, C_j}{N}$

By which$m$ refers to the number of data ranges and$k$refers to the number of data samples. To illustrate, if$i$is the number of data samples and$j$ is the class, then:

R_i refersthe number of data sets in i range.

C_j refers to the number of data sets in j range.

Nrefers to the total number of data sets.

Therefore, X_{ij}isthe expected frequency ofA_{ij}. If the output has the highest value of X^2, it means thatfeatureis the most significant among the data sets.

- *Information Gain (IG):*is a feature selection method based on theInformation theorywhich identifies the feature variables by ranking the feature variables according to theInformationGain value as formulated in (2) [3].

$$IG(Y, X) = H(Y) - H(Y|X) \tag{2}$$

Where:

Y indicatesclass.

X indicates features.

$H(Y)$refers to Entropy of Y.

$H(Y|X)$ refers to Entropy of $Y$when following the condition of$X$.

The calculation of $H(Y)$ is formulated in (3), and the calculation of$H(Y|X)$is formulated in (4).

$$H(Y) = -\sum_{y \in Y} p(y)\, log_2\, p\,(y) \tag{3}$$

$$H(Y|X) = -\sum_{x \in x} p(x) \sum_{y \in Y} p(y|x)\, log_2\, p\,(y|x) \tag{4}$$

Where:

$p(y)$indicatesthe probability of$y$.

$p(x)$indicates the probability of$x$.

$p(y|x)$indicatesthe probability of $y$when following the condition of$x$.

B. Decision tree

Decision tree [3],[4] is a tree-like classification method. A tree consists of decision nodes and end nodes. Each decision node represents a decision to be made, while each branch represents the final outcome of a decision path and each leaf node represents a class label. There are several decision tree methods such as:

- *Iterative Dichotomiser 3 (ID3):*is an algorithm used for creating a decision tree based on the Information Theory by which the Information Gain value is used for sorting data. The information value is used to measure the variables for sorting data. This method constructs the tree structure by selecting the data set by order or setting the attribute that has the maximum information gain as the starting data, following by the attributes with lower information gain value, respectively. According to the Information Theory, the information gain value depends on the probability of the data, as formulated in (5) [3],[4].

$$Info(D) = -\sum_{i=1}^{m} p_i\, log_2(p_i) \tag{5}$$

Where:

p_i is the probability of having a data set of Ddatabase in ClassC_i which is equal to $|C_{i,D}|/|D|$.

m indicates the total number of different classes in that data set.

C_i indicates the class at Sequence i by which$i$ is between1 and$m$.

$|C_{i,D}|$indicates the number of data sets of D database in Class C_i.

$|D|$ indicates the number of data sets in$D$database.

$Info(D)$ refers to'Entropy' of$D$database. The information gain value, according to the Information Theory, contributes to the classification as it reduces the number of tests. Moreover, it also guarantees the simplicity of the decision tree. Once a feature is selected as the test choice, the information gain value can be formulated as illustrated in (6) [3],[4].

$$Info_A = \sum_{j=1}^{k} \frac{|D_j|}{|D|} x Info(D_j) \tag{6}$$

Where:

k indicates the probability value of the feature.

$|D|$ indicates the number of data sets in D database.

$|D_j|$ indicates the number of data sets in$D$database that have A feature valued at$j$.

The standard gain value used for selecting A feature as a tree node is equal to the difference between the information gain from Information Theory and the information gain from the Feature Information Theory. This is formulated in (7) [3],[4].

$$Gain(A) = Info(D) - Info_A(D) \tag{7}$$

- *C4.5:*[3], [5] is an algorithm used for creating decision trees, which was developed by Ross Quinlan. It has the same foundation as the ID3 algorithm by which the development of decision trees for classification is based on the Information Gain principle same as ID3, but it adds the Gain Ratio criterion to select the root nodes or nodes at different levels in the tree, as shown in (8).

$$SplitInfo(A) = -\sum_{j=1}^{k} \frac{|D_j|}{|D|} x\, log_2\left(\frac{|D_j|}{|D|}\right) \tag{8}$$

Where:

k indicates the level of data dispersion.

$|D|$ indicates the number of data sets in D database.

$|D_j|$ indicates the number of data sets inD database that have j as the level of data dispersion of Adata.

The split information value indicates the level of data dispersion. When dividing the Gain value with it, the gain ratio standard value will be calculated, as shown in (9).

$$GainRatio(D) = \frac{Gain(A)}{SplitInfo(A)} \tag{9}$$

The gain ratio standard value is used for resolving the problem of gain bias by reducing the gain ratio that has a high rate of dispersion.

- *Random Forest:*is a feature randomization technique. It randomizes features from various data sets and then use the randomized features to build a model by using the decision trees based on the Random Forest technique. This technique was originally proposed in 1998 by Kam and then developed by Breiman [6]. The features of trees in the Random Forest are controlled by 3 stages: 1) each tree is trained by the sub-set of sample data. 2) When the tree grows up, it can find the nodes from the best branches by randomizing from N features. 3) Each tree will not be excluded. The trees are left to grow continuously in order to get the best outputs. After the forest is created, the trees are voted for developing a model by which the tree receiving the highest number of votes will be selected as a model.

Data mining technique in order to create a model predicting the chances of readmission of diabetes patients. The effectiveness of the model was compared by three techniques, including Decision tree, Naïve Bayes and K-Nearest Neighbors. The findings showed that the model developed by the decision tree technique had prediction effectiveness with 85.5% accuracy [7], meaning it could be used for supporting medical care in the future. Development of a model for predicting the youth's internet addiction, and then compared the model with decision trees created by J48, ID3, LMT, CART and Random Forest with 10-fold cross validation, in order to sort the data into instruction data sets and test data sets. The evaluation of the model's prediction effectiveness demonstrated that the Random Forest technique could provide better prediction than J48, ID3, LMT and CART with 87.15% accuracy [8]. A model for predicting basic diseases in cows was developed by applying the decision tree technique in order to create an application for diagnosing basic diseases in cows on smartphones. This was done by comparing three algorithms, including J48, RandomTree and REPTree. The findings revealed that the RandomTree algorithm provided the most effective prediction with 99.47% accuracy, showing that the developed application was effective at a high level.

III. RESEARCH METHODOLOGY

Based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) [9], the research process in this study consists of 6 stages as illustrated in Figure 1.

A. Business understanding

In Thailand, there are patients and casualties of mushroom poisoning annually. This trend tends to grow since mushrooms are a source of food that can be easily found in the nature and is full of nutrients. Therefore, there should be the public communication on safe mushroom consumption and how to identify poisonous mushrooms in order to solve the problem. At present, the information about poisonous mushrooms on the existing websites mostly required the users a lot of time to study in order to be able to identify poisonous mushrooms. The research team have collected information about edible and poisonous mushrooms to analyze the features of poisonous mushrooms and developed a classification model to warn the users to avoid poisonous mushroom consumption faster and more conveniently.

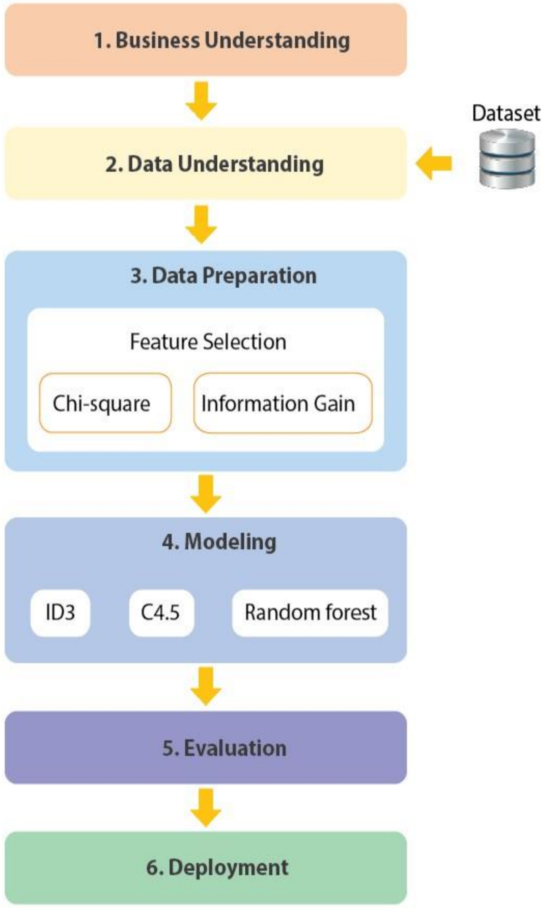
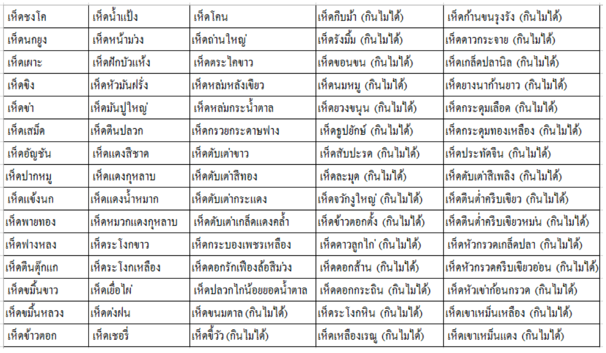


Fig. 1. The research process based on the CRISP-DM data mining development standards.

B. Data understanding

This research has collected data on edible and poisonous mushrooms from the Encyclopedia of Thai mushrooms [10] and the book entitled "Diversity of Mushrooms and Macrofungi in Thailand [11]" which provided a list of 75 mushrooms show as Figure 2 as well as the features for classifying edible and poisonous mushrooms show as Table I.



เห็ดชงโค	เห็ดน้ำแป้ง	เห็ดโคน	เห็ดกีบม้า (กินไม่ได้)	เห็ดก้านขนรุงรัง (กินไม่ได้)
เห็ดนกยูง	เห็ดหน้าม่วง	เห็ดถ่านใหญ่	เห็ดรังมิ้ม (กินไม่ได้)	เห็ดดาวกระจาย (กินไม่ได้)
เห็ดเผาะ	เห็ดฝักบัวแห้ง	เห็ดตระไคขาว	เห็ดขอนขน (กินไม่ได้)	เห็ดเกล็ดปลานิล (กินไม่ได้)
เห็ดขิง	เห็ดหัวมันฝรั่ง	เห็ดหล่มหลังเขียว	เห็ดนมหมู (กินไม่ได้)	เห็ดยางนาก้านยาว (กินไม่ได้)
เห็ดข่า	เห็ดมันปูใหญ่	เห็ดหล่มกระน้ำตาล	เห็ดยวงขนุน (กินไม่ได้)	เห็ดกระดุมเลือด (กินไม่ได้)
เห็ดเสม็ด	เห็ดตีนปลวก	เห็ดกรวยกระดาษฟาง	เห็ดฐูปยักษ์ (กินไม่ได้)	เห็ดกระดุมทองเหลือง (กินไม่ได้)
เห็ดอัญชัน	เห็ดแดงสีชาด	เห็ดตับเต่าขาว	เห็ดสับปะรด (กินไม่ได้)	เห็ดประทัดจีน (กินไม่ได้)
เห็ดปากหมู	เห็ดแดงกุหลาบ	เห็ดตับเต่าสีทอง	เห็ดละมุด (กินไม่ได้)	เห็ดตับเต่าสีเพลิง (กินไม่ได้)
เห็ดแข้งนก	เห็ดแดงน้ำหมาก	เห็ดตับเต่ากระแดง	เห็ดจวักงูใหญ่ (กินไม่ได้)	เห็ดตีนต่ำครีบเขียว (กินไม่ได้)
เห็ดพายทอง	เห็ดหมวกแดงกุหลาบ	เห็ดตับเต่าเกล็ดแดงคล้ำ	เห็ดข้าวตอกตั้ง (กินไม่ได้)	เห็ดตีนต่ำครีบเขียวหม่น (กินไม่ได้)
เห็ดฟางหลง	เห็ดระโงกขาว	เห็ดกระบองเพชรเหลือง	เห็ดดาวลูกไก่ (กินไม่ได้)	เห็ดหัวกรวดเกล็ดปลา (กินไม่ได้)
เห็ดตีนตุ๊กแก	เห็ดระโงกเหลือง	เห็ดดอกรักเฟืองล้อสีม่วง	เห็ดดอกส้าน (กินไม่ได้)	เห็ดหัวกรวดครีบเขียวอ่อน (กินไม่ได้)
เห็ดขมิ้นขาว	เห็ดเยื่อไผ่	เห็ดปลวกไก่น้อยยอดน้ำตาล	เห็ดดอกกระถิน (กินไม่ได้)	เห็ดหัวเข่าก้อนกรวด (กินไม่ได้)
เห็ดขมิ้นหลวง	เห็ดต่งฝน	เห็ดขนมตาล (กินไม่ได้)	เห็ดระโงกหิน (กินไม่ได้)	เห็ดเขาเหมี้นเหลือง (กินไม่ได้)
เห็ดข้าวตอก	เห็ดเชอรี่	เห็ดขี้วัว (กินไม่ได้)	เห็ดเหลืองเรณู (กินไม่ได้)	เห็ดเขาเหมี้นแดง (กินไม่ได้)

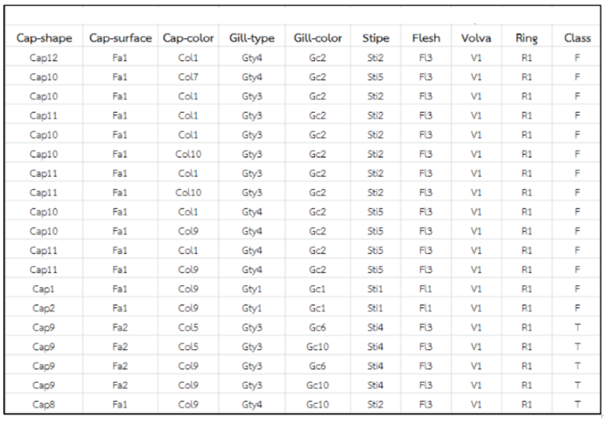
Fig. 2.List of Thai name of mushrooms.

Table- I: Features used for classifying edible and poisonous mushrooms

Features	Feature Variables
Cap Shapes	- round - cylindrical - conical - oval - tubal - paraboloid - ovate - plane - flat - depressed - bellshaped - funnelshaped - hemispherical - campanulate - umbonate
Cap Textures	- smooth - flat scales - raised scales - velvety - hair or fibrous - patches
Cap Colors	- white - red - pink - orange - yellow - green - purple - grey - brown - black
Gills	- free -attached - decurrent - notched
Gill Colors	- translucent -white - red - pink - orange - yellow - green - purple - grey - brown
Stems	- none -cylindrical - venticose - attenuated aloft - rooting
Stem Bodies	- none - stuffed - hollow
Volva	- none - topshaped - marginate
Rings	- none -pendent - flaring - sheathing - double - cobwebby - ring zone
Class	- edible, - poisonous

C. Data preparation

After collecting the data on edible and poisonous mushrooms (Figure 1 and Table I), the research team analyzed and cross-checked the data. Then, the data was transformed into .CSV file in order to be loaded in Weka version 3.8.3 (See Figure 3).



Cap-shape	Cap-surface	Cap-color	Gill-type	Gill-color	Stipe	Flesh	Volva	Ring	Class
Cap12	Fa1	Col1	Gty4	Gc2	Sti2	Fl3	V1	R1	F
Cap10	Fa1	Col7	Gty4	Gc2	Sti5	Fl3	V1	R1	F
Cap10	Fa1	Col1	Gty3	Gc2	Sti2	Fl3	V1	R1	F
Cap11	Fa1	Col1	Gty3	Gc2	Sti2	Fl3	V1	R1	F
Cap10	Fa1	Col1	Gty3	Gc2	Sti2	Fl3	V1	R1	F
Cap10	Fa1	Col10	Gty3	Gc2	Sti2	Fl3	V1	R1	F
Cap11	Fa1	Col1	Gty3	Gc2	Sti2	Fl3	V1	R1	F
Cap11	Fa1	Col10	Gty3	Gc2	Sti2	Fl3	V1	R1	F
Cap10	Fa1	Col1	Gty4	Gc2	Sti5	Fl3	V1	R1	F
Cap10	Fa1	Col9	Gty4	Gc2	Sti5	Fl3	V1	R1	F
Cap11	Fa1	Col1	Gty4	Gc2	Sti5	Fl3	V1	R1	F
Cap11	Fa1	Col9	Gty4	Gc2	Sti5	Fl3	V1	R1	F
Cap1	Fa1	Col9	Gty1	Gc1	Sti1	Fl1	V1	R1	F
Cap2	Fa1	Col9	Gty1	Gc1	Sti1	Fl1	V1	R1	F
Cap9	Fa2	Col5	Gty3	Gc6	Sti4	Fl3	V1	R1	T
Cap9	Fa2	Col5	Gty3	Gc10	Sti4	Fl3	V1	R1	T
Cap9	Fa2	Col9	Gty3	Gc6	Sti4	Fl3	V1	R1	T
Cap9	Fa2	Col9	Gty3	Gc10	Sti4	Fl3	V1	R1	T
Cap8	Fa1	Col9	Gty4	Gc10	Sti2	Fl3	V1	R1	T

Fig. 3.Data used for modeling.

D. Modeling

This stage is the development of the model classifying edible and poisonous mushrooms. Before using the data to develop the classification model, the research team had applied two feature selection methods which were 1) Chi-square and 2) Information Gain. After selecting the features, the data was used for modeling with three decision tree techniques, including ID3, C4.5 and Random Forest, by Weka software in order to discover the most effective model.

E. Evaluation

The model's effectiveness was evaluated by the 10-fold Cross validation approach. Three decision trees were created by ID3, C4.5 and Random Forest, respectively. Next, they were compared in order to find the most effective one. Finally, the most effective decision tree was applied for developing the edible and poisonous mushrooms classification model.

The evaluation applied the accuracy calculation formula as formulated in (10) [5].

$$Accuracy = \frac{TP+TN}{(TP+FP+FN+TN)} \qquad (10)$$

The equation above describes the following variables: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

IV. EXPERIMENTAL RESULTS

The result of presents the development of the edible and poisonous mushrooms classification model by using the feature selection and the decision tree techniques. Two feature selection methods were applied, which were Chi-square and Information Gain, while three decision tree techniques were employed, including ID3, C4.5, and Random Forest. The research methodology was conducted based on CRISP-DM. The research findings revealed that the model developed by Information Gain integrated with Random Forest achieved the highest accuracy is 94.19%.

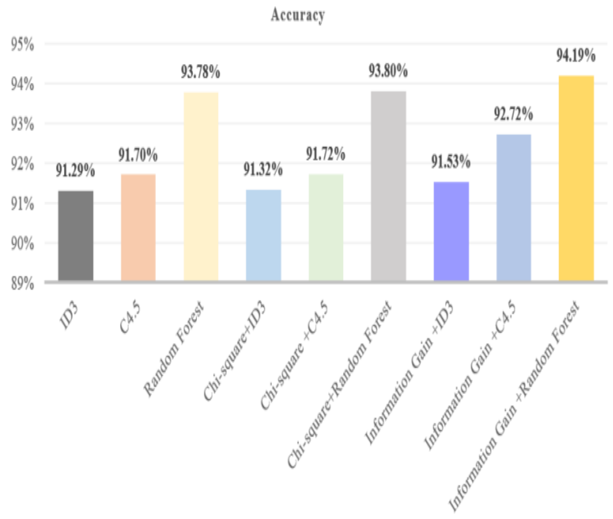
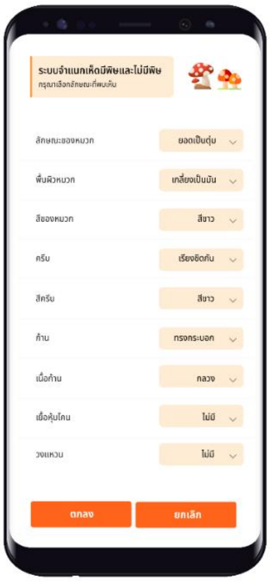


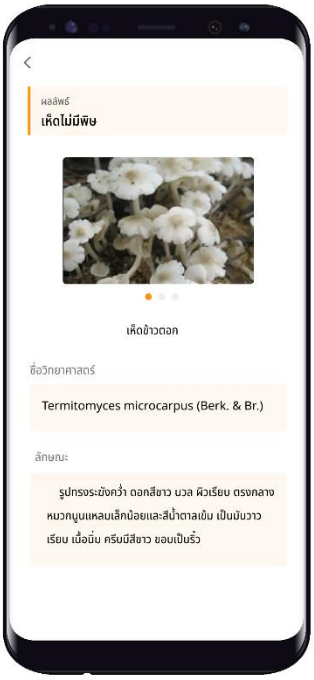
Fig. 4.Results of the evaluation of the model's effectiveness.

After obtaining the most effective decision tree, which was created by the Information Gain method and the Random Forest method, the research team applied the decision tree to develop the prototype model of edible and poisonous mushrooms classification. At this stage, the most effective decision tree was applied for developing the edible and poisonous mushrooms classification model that was available on Android smartphones for Thai user, as illustrated in Figure 3 and 4.


ระบบจำแนกเห็ดมีพิษและไม่มีพิษ
กรุณาเลือกลักษณะที่พบเห็น

ลักษณะของหมวก ยอดเป็นตุ่ม
พื้นผิวหมวก เกลี้ยงเป็นมัน
สีของหมวก สีขาว
ครีบ เรียงชิดกัน
สีครีบ สีขาว
ก้าน ทรงกระบอก
เนื้อก้าน กลวง
เยื่อหุ้มโคน ไม่มี
วงแหวน ไม่มี

Fig. 5.The screen displaying the edible and poisonous mushrooms classification system.


ผลลัพธ์
เห็ดไม่มีพิษ

เห็ดข้าวตอก

ชื่อวิทยาศาสตร์
Termitomyces microcarpus (Berk. & Br.)

ลักษณะ
รูปทรงระฆังคว่ำ ดอกสีขาว นวล ผิวเรียบ ตรงกลางหมวกนูนแหลมเล็กน้อยและสีน้ำตาลเข้ม เป็นมันวาวเรียบ เนื้อนิ่ม ครีบมีสีขาว ขอบเป็นริ้ว

Fig. 6.The screen displaying the classification results.

V. CONCLUSION AND DISCUSTION

This article presents the development of the edible and poisonous mushrooms classification model by using the feature selection and the decision tree techniques. Two feature selection methods were applied, which were Chi-square and Information Gain, while three decision tree techniques were employed, including ID3, C4.5 and Random Forest. The research methodology was conducted based on CRISP-DM. The research findings revealed that the model developed by Information Gain integrated with Random Forest achieved the highest accuracy (94.19%). After applying the model to the development of an edible and poisonous mushrooms classification model, it was found out that the model worked effectively. This conforms to the study conducted by Palwisut [8], which introduced the application of the decision tree technique to the youth's internet addiction prediction and compared the effectiveness of several decision tree techniques such as J48, ID3, LMT, CART and Random Forest; the research results showed that the Random Forest technique could provide better prediction than J48, ID3, LMT and CART.

ACKNOWLEDGMENT

We would like to express our gratitude to the Institute for Research and Development, SuanSunandhaRajabhat University and Faculty of Science, UbonRatchathani University, UbonRatchathani Province, who offered us the opportunities to conduct this research.

REFERENCES

1. Common Knowledge about Mushrooms. (2017, November 5). [Online]. Available: http://www.aopdh02.doae.go.th/wonlop_het.pdf
2. P. Ratchapra, and C. Chantrapornchai, “Expert system for rice disease diagnosis,” E-Journal SU, vol. 6, no. 1, 2013.
3. H. Jiawei, and K. Micheline, “Data Mining: Concepts and Techniques,” United States of America: Morgan Kaufmann, 2006.
4. S. N. Chary, and B. Rama, “A Survey on Comparative Analysis ofDecision Tree Algorithms in Data Mining,” International Journal of Advanced Scientific Technologies, Engineering and Management Sciences, vol. 3, no. 1, Special Issue 1, 2017.
5. H. W. Ian, F. Eibe, and A. H. Mark. “Data Mining: Practical Machine Learning Tools and Techniques,” 3th Edition, Burlington, 2011.
6. L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, 2001, pp. 5–32.
7. W. Sangmanee, V. Rattanacharoenlert, N. Pottirat, and P. Noosawat, “Building the Predicting Model of the Chance of Repeating the Treatment of Diabetes Patient Using Data Mining,” In Proceeding of the5th ASEAN Undergraduate Conference in Computing (AUC2), 2017.
8. P. Palwisut. “Improving Decision Tree Technique in Imbalanced Data Sets Using SMOTE for Internet Addiction Disorder Data,” Information Technology Journal, vol. 12, no. 1, 2016.
9. C. Shearer, “The CRISP-DM model: The new blueprint for data mining,” Journal of Data Warehousing, vol. 5, no. 4, 2000, pp. 13–22.
10. A. Chandrasrikul, Thai Mushrooms and Mushroom Cultivation Technology, Bangkok: Thai Wattanapanich Publishing, 1999.
11. A. Chandrasrikul, P. Suwanarit, and U. Sangwanit, Diversity of Mushrooms and Macrofungi in Thailand, Bangkok: Kasetsart University Press, 2008.

AUTHORS PROFILE

Sumitra Nuanmeesri, received the Ph.D. in Information Technology at King Mongkut’s University of Technology North Bangkok, Thailand. She is Assistant Professor in Information Technology Department, Faculty of Science and Technology at SuanSunandhaRajabhat University, Thailand. Her research interests include speech recognition, data mining, deep learning, image processing, mobile application, supply chain management system, internet of things (IoT), robotics, augmented reality (AR), and virtual reality (VR).

Wongkot Sriurair, eceived the Ph.D. in Information Technology at King Mongkut’s University of Technology North Bangkok. Bangkok, Thailand. She is Assistant Professor in Mathematics Statistics and Computer Department, Faculty of Science, UbonRatchathani University, UbonRatchathani Province, Thailand. Her research interests include data mining, text mining, web mining, recommender system, information filtering, information retrieval, decision support system, expert system, multimedia technology and computer education.