Children Shoes Suggestion System Using Data Mining

P. Sudta, K. Kanchan and C. Chantrapornchai

Department of Computing, Faculty of Science
Silpakorn University, Thailand, 73000
ctana@su.ac.th

Abstract

We present a prototype web application for children shoes suggestion. We are interested in attributes of foot and shoes that can comfort the children's feet. The key attributes are types of shoes, size, and brands. For the feet, there are width, length, thickness, sole curve, toe and others. The application has the database that stores these children shoes and foot characteristics for training. The data mining model is created using decision tree, KNN, neural net and integrated to the application.

Keywords: *Shoe suggestion, data mining, decision tree, KNN, Neural net*

1. Introduction

For children, there are various kinds of attributes that are needed to consider for buying their shoes. The children shoes need to be comfort. The size needs to be suitable to them as well as the shape of the shoes. In this research, we are interested in shoe sizing and major attributes that affect the decision on buying shoes for Thai students. In particular, we would like to develop a prototype system that suggests proper shoes for small kids based on certain criteria.

Data mining is a popular technique used in classification and predictions. Many research works used it as a tool for decision support system. Palaniappan S. and Awang R. [7] developed a heart disease decision support system. It used naïve bayes, decision tree and neural net to model the classifications. It is developed as a web application and used DMX query tool. C.Y. Ma, Frances, V Buontempo and X. Z Wang [6] used inductive data mining for historic data analysis. M.J. Aitkenhead [1] developed a co-evolution decision tree method. It combines an evolution method with the decision tree model for a better classification. Hudson S., and Ritchie B. [4] used cluster analysis for tourism in Alberta. They applied the model CRISP-DM. Jiao J. R et. al. [5] developed Kensei Mining which uses association rules to design user interface. Hsieh C., Huang S. [3] applied data mining technique to design new products. Apriori algorithm was used here. The attributes investigated are customer needs and product properties. Prassas G., et. al., [8] data mining for online shopping suggestion. Chen Y.L., Chen J.M., Tung C.W. [2] applied data mining to see the effect of shelf-space adjacency.

In our paper, we develop a children shoe suggestion system. The data mining engine is used. Particularly the decision tree, KNN, neural net techniques are applied. We collect samples data from the neighborhood school from K1-K3 and P1-P3, each of which is 20-30 students. About the students, we collect the student age, shoe shape, sex. About their shoes, we collect the shoe sizes, brands, and types. Assume the types are student shoes, sport shoes and leisure shoes. We design the database to save all the student data and shoe attributes. After the modeling for classification is developed, it is integrated to the web-based system.

The paper is organized as follows: the next section is the backgrounds. Then we present the data collection and preparation. Next, in Section 4, we present some results from data mining and compare the three approaches. In Section 5, the web application is presented and Section 6 presents the conclusion.

2. Backgrounds

Data mining technique is used often with large database, dataware house etc. It is mainly applied in decision support systems for modeling and prediction. There are several kinds of data mining: classification, clustering, association, sequencing etc.

Typically, for classification, the classifier model is needed. Data are divided into training set and test set. The training data is used to create the model. Then the test set applied for checking the model correctness. Until satisfied, the model is trained and adjusted by training data. Common techniques used in classification are decision tree, neural network, naïve bayes, etc.

In this work, we collect the data from one of the public kindergarden school. The school is a small size. There are 20 boys each in K1, K2 and K3 respectively. There are 30 boys in P1, P2, and P3 respectively. There are 20 girls each in K1, K2, and K3 respectively. Also, there are 30 girls each in P1,P2 and P3 respectively. There are 3 types of shoe types such as sport shoes, student shoes and leisure shoes. We collect 120 pairs of student shoes dividing into 60 boys' and 60 girl's, 70 pairs of sport shoes from 35 boys and 35 girls, 50 pairs of leisure shoes from 25 boys and 25 girls.

We design a form for collecting these data from students. The attributes that are specified are student information such as name, sex, age, foot information such as sole curve, toes characteristics, thickess, and height. We also draw the foot figures for left and right sides for references. The shoe sizes for each type are collected as well. The sample form is shown in Figure 12 in Appendix. From the data collected, we design an E-R diagram which corresponds to the database table for the application in Figure 1. Then these data are divided into training set and test set as in Figure 2.

In Figure 2, the attributed used to train are age, sex, foot width, foot length, foot thickness, sold curve, foot type, toe type, and length upto the angle. The last two columns are the brands and shoe sizes.

3. Data Collection

From the above data collection, we analyze the attributes of the feet as following.

- Foot length. The students for these ages mostly have the same foot length. There are very few samples that have a distinct foot size.
- Sole curve. Usually, there are three types of sole curve. However, since these are very small children, the sole curves are in only flat level and small curve.

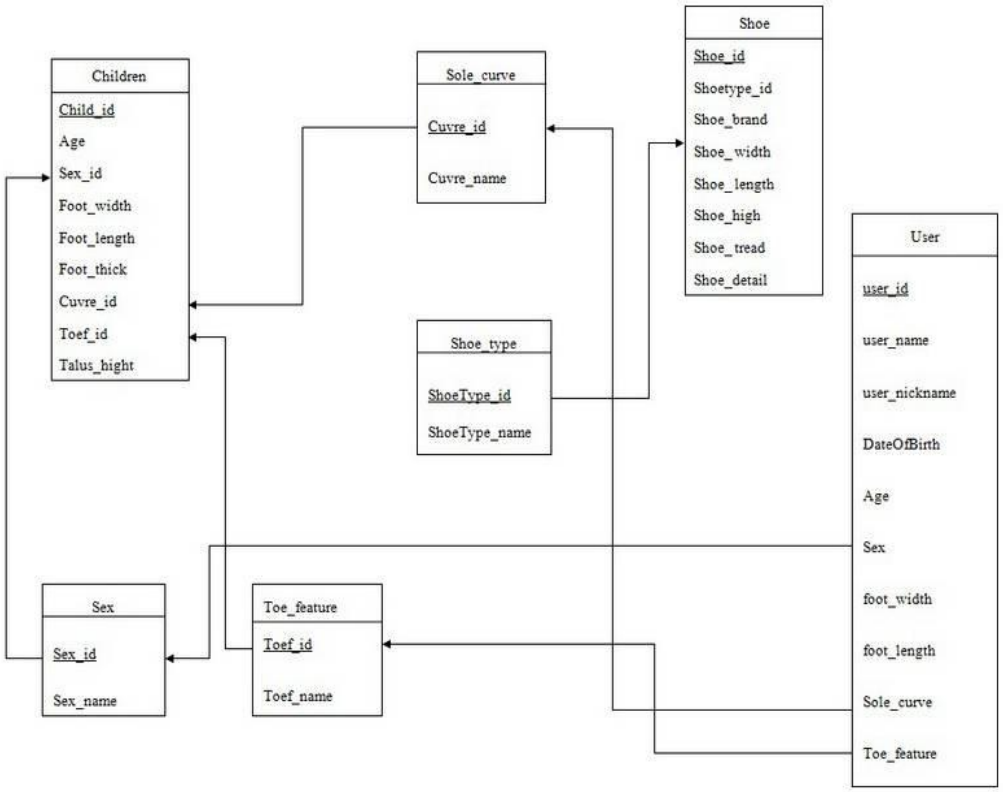

Figure 1. E-R Diagram of the Collected Data

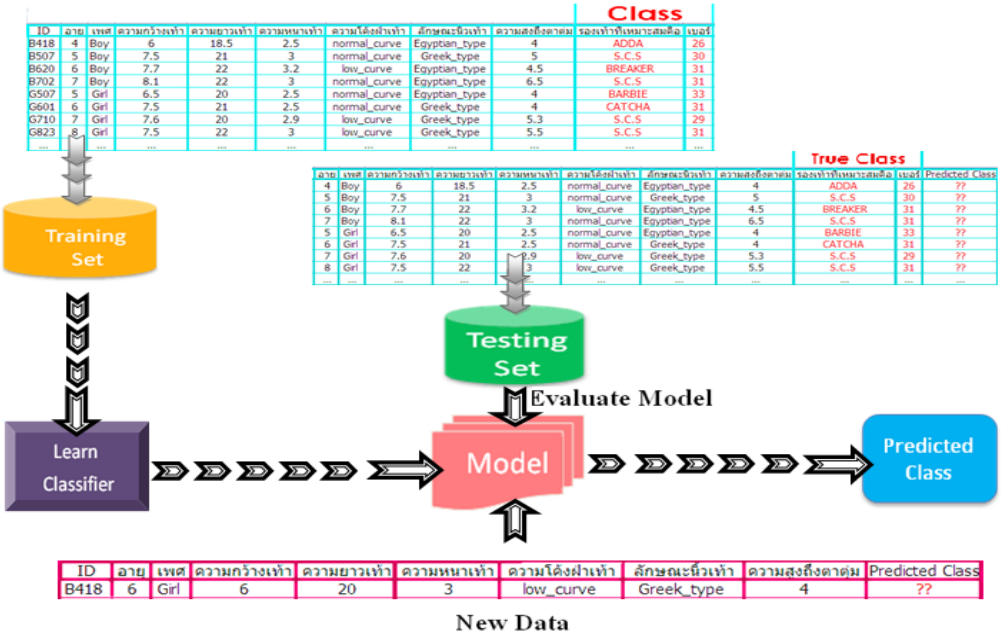

Figure 2. Training the Model

- Toe feature. Typically, there are three types: Egyptian style (where left one in the right side is the longest one and the others are shorter in the decreasing order), Greek style (where the second one is the longest) and the square style (where every finger is of the same height). From our data, most of the students are in the Greek style, then Egyptian style. Then the for the student shoe type, it can be matched with all kinds of feet. For the sport shoes, by inspecting the head of the shoes, it may be matched only Greek style feet. For the leisure shoes, by inspecting the head of the shoes, it may also be matched only Greek style feet.

From our data, we still have problems of insufficient data for children. Most likely, children in the same area and nationality, have the same style of feet. Also, the available shoes are very similar to each other. Then, the foot styles are not influence the classification so much.

4. Results and Analysis

We use Weka (http://www.cs.waikato.ac.nz/ml/weka/) tool to create the decision tree. Figure 3 is an example result for student shoes. In this example, 1) foot length is the most important attribute. Then it is sole curve (small curve) and sex (female). We obtain CATCHA_31 (6.0) which is the brand CATCHA, size 31, and there are 6 students on this leaf. In 2), also, foot length is the most important attribute. Then it is sole curve (small curve) and sex (male). The shoes are SCS brand whose size is 33 and there are 5 students in this leaf.

In Figure 4, it is the sample results from sport shoes. In 1), only the foot length is used. We show the path from the root to leaves, ADDA_30(21.0). From our data, thereare 21 children who wear this type of shoes. In 2), the attributes are foot length, sex (male). In the tree, we step through the foot length attribute first, and then the sex attribute is used. After that, the foot length attribute is used again. From the result, PAN_36 (3.0) is selected.

In Figure 5, it is the tree for leisure shoes. For 1) and 2), the sex and foot length are the major attributes for the decision. For 1), we use the sex attribute first, then the foot length is considered. From the result, we obtain the leaf BATMAN_35 (15.0). There are 15 students on this leaf. For 2), similarly, the sex and foot length are being used. KITTY_33 (6.0) is the resulting leaf.

For the given data, we test the accuracy results for different models: KNN, Neural net and decision tree. These are shown in Table 1. It is seen that the decision tree performs best among the three models. We suspect this is because the mentioned reason in Section 4.

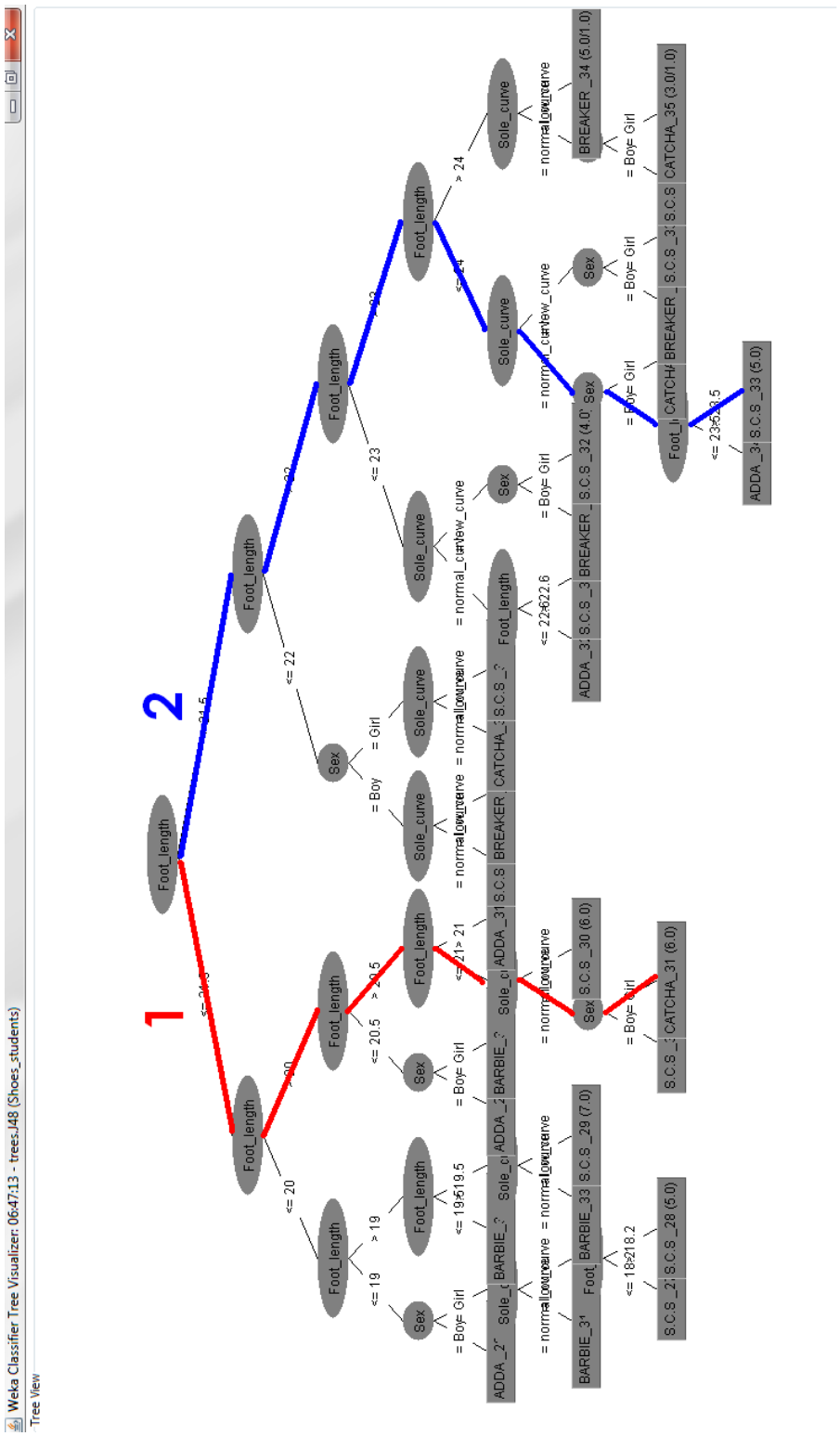

Figure 3. Sample Tree Results for Student Shoe Classification

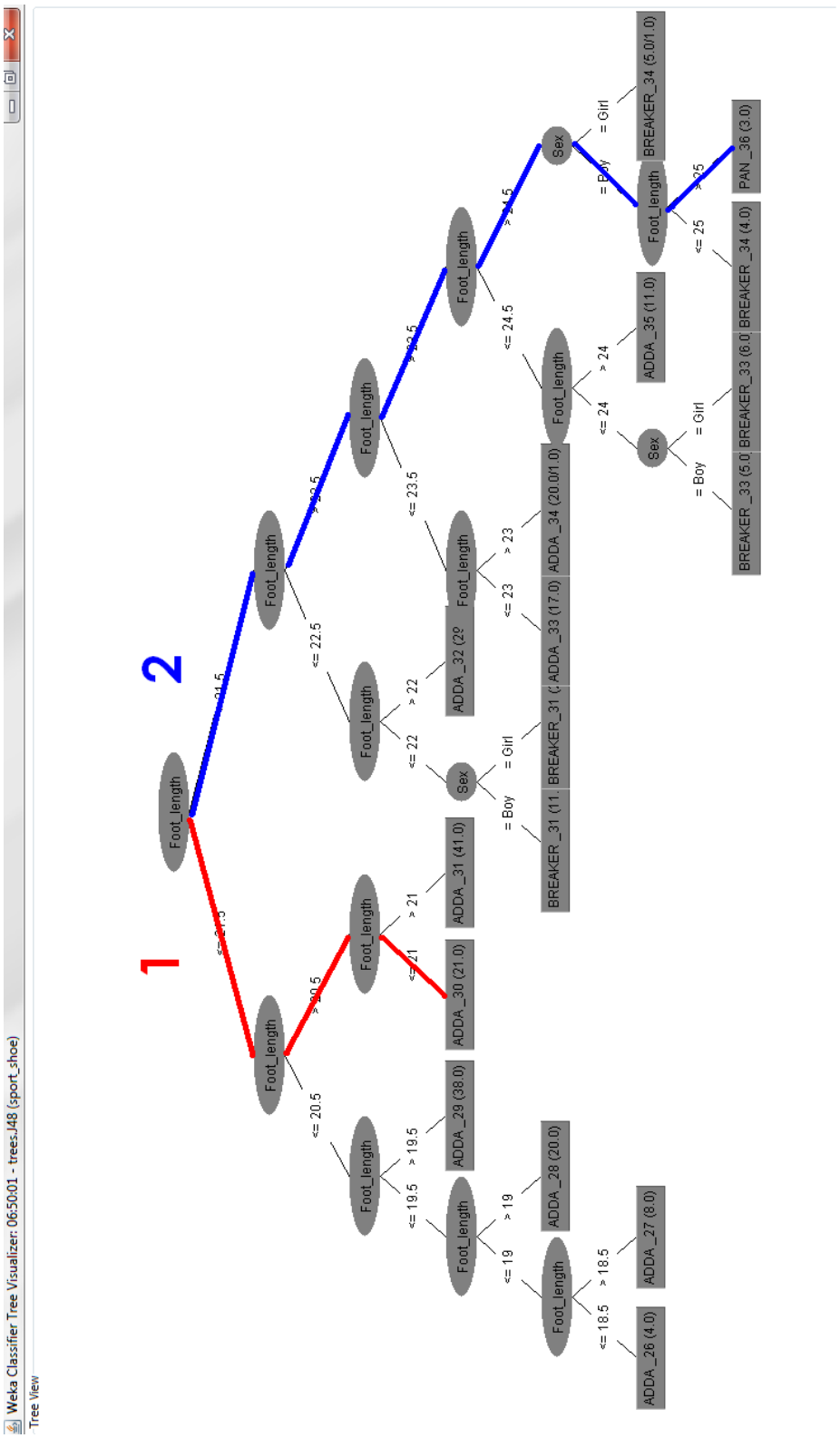

Figure 4. Sample Tree Results for Sport Shoe Classification

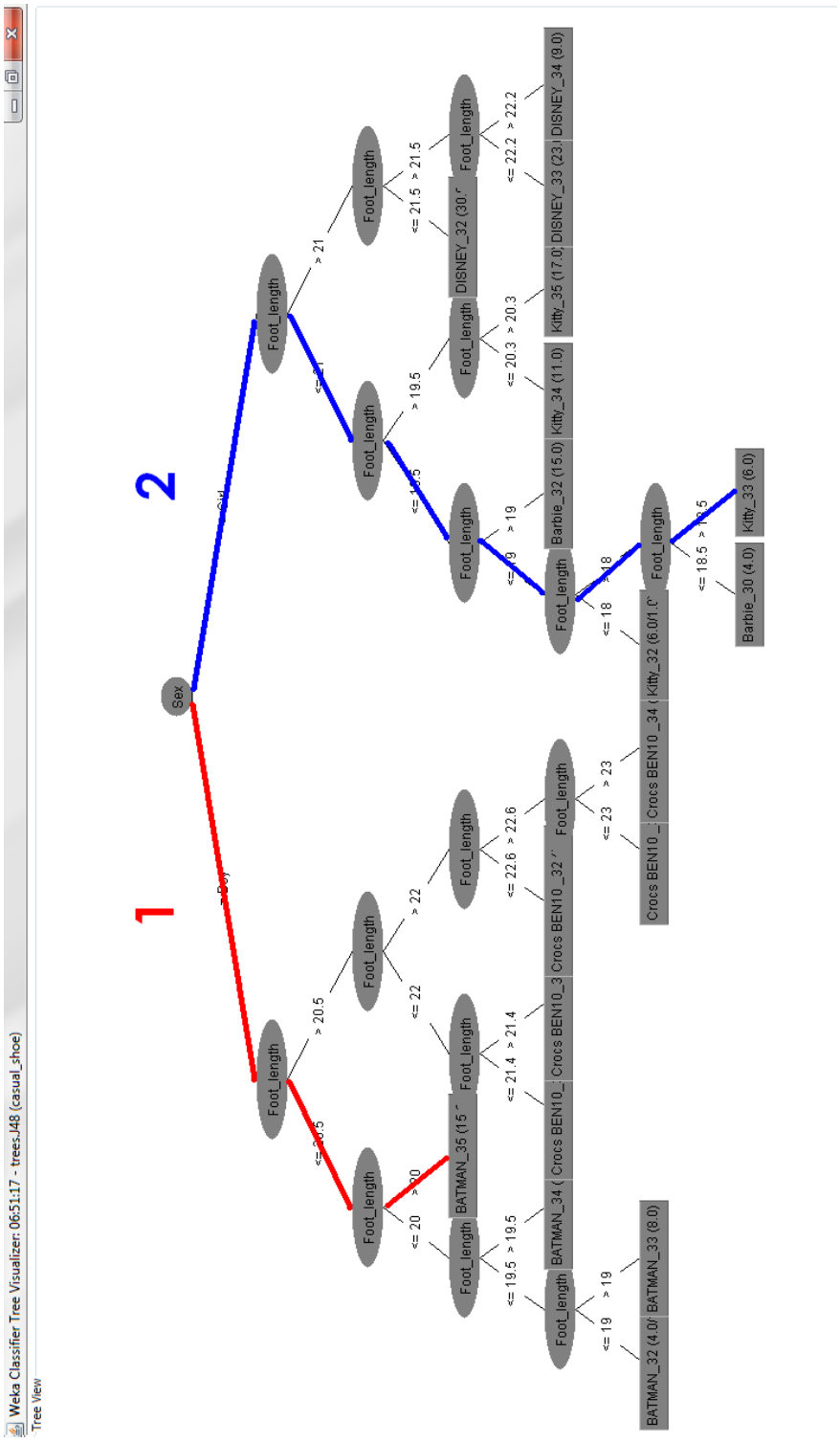

Figure 5. Sample Tree Results for Leisure Shoe Classification

Table 1. Accuracy for Different Models

Type of shoes	Accuracy		
	NN	KNN	Decision tree
Student shoes	53%	56%	82%
Sport shoes	76%	54%	98%
Leisure shoes	80%	55%	99%

The data from this table is taken from the example running of tests in Table 1 shown in Figures 13-21 in Appendix.

5. Web Application

We develop the web application for suggesting the shoes using our data mining engine as well. Sample user interface is shown in Figures 6-11. In Figure 6, it is an introductory page. the user enters his profile about name, date of birth, sex. Then, in Figures 7-8, we show the shoe types in our database. Figure 9 is an interface for a user profile. The user must register to enter his information on name, age, and sex. In Figure 10, the user then enters his foot information such as thickness, length, width, foot types etc. These information are required for our data mining model. After that, the program reports proper shoes for each type as shown in Figure 11. The types are student shoes, sport shoes and leisure shoes. The answer is the brand and size as well as the percentage of accuracy from our database in the last column.

Figure 6. Sample Introductory Page

Figure 7. Shoes Examples (I)

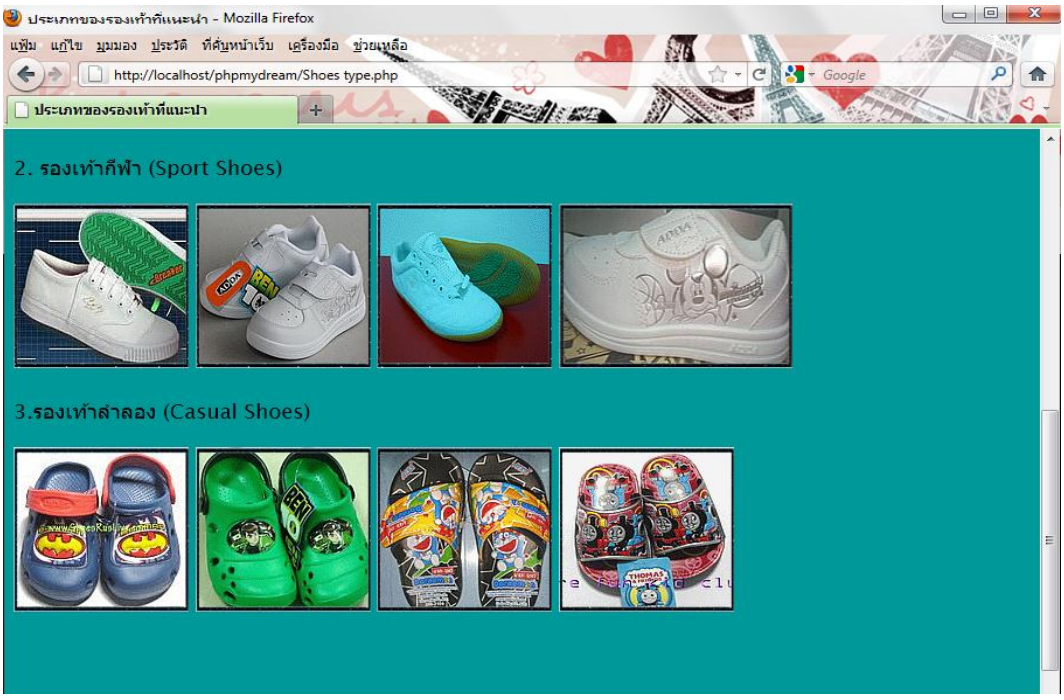

Figure 8. Shoes Examples (II)

Figure 9. User Profile

Figure 10. User Foot Information

Fig. 11. Suggested Shoes with Accuracy

5. Conclusion

We present an application for children shoes suggestion. The application is developed as a web application. It includes a data mining engine for predicting the proper shoe brand and size. The goal of this work is to study the attributes necessary to decide proper shoes for children since there are various foot attributes and shoe attributes.

Currently, the length of the shoes is the major concern. Due to this, decision tree performs well among the three approaches. In reality, sole curve, width, thickness are important to the comfort as well. To reflect this, data about the shoes need to be collected more intensively and in more detailed. This will be investigated further.

References

[1] Aitkenhead MJ, "A co-evolving decision tree classification method", Journal Expert Systems with Applications: An International Journal, vol. 34 Issue 1, (**2008**) January, DOI:10.1016/j.eswa.2006.08.008.

[2] Chen YL, Chen JM, Tung CW, "A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales", Journal Decision Support Systems, vol. 42, Issue 3, (**2006**) December, DOI: 10.1016/j.dss.2005.12.004..

[3] Hsieh C, Huang S, "Mining product maps for new product development Expert Systems with Applications", vol.34, Issue 1, Wiley Publishing Inc., (**2008**), pp. 50-62.

[4] Hudson S, Ritchie B, "Understanding the domestic market using cluster analysis", Journal of Vacation Marketing, vol. 8, no. 3, (**2002**) July, pp. 263-276. DOI: 10.1177/135676670200800305.

[5] Jiao JR, Zhang Y, Helander MG, "A Kansei mining system for affective design", Expert Syst. Appl., (**2006**), pp. 658-673.

[6] Ma CY, Buontempo FV, Wang Z, "Inductive Data Mining: Automatic Generation of Decision Trees from Data for QSAR Modeling and Process Historical Data Analysis", 18th European Symposium on Computer Aided Process Engineering Vol. 25, Issue. Idm, Publisher: Elsevier, (**2008**), pp. 581-586.

[7] Palaniappan S, Awang R, "Web-based Heart Disease Decision Support System Using Data Mining Classification Modeling Techniques", In Proceedings of iiWAS'2007, (**2007**), pp.157-167.

[8] Prassas G, et. al., "Recommender System for Online Shopping Based on Past Customer Behaviour, e-Everything: e-Commerce, e-Government, e-Household, e-Democracy", The 14th Bled Electronic Commerce Conference, Bled, Slovenia, (**2001**) June 25–26.

Appendix

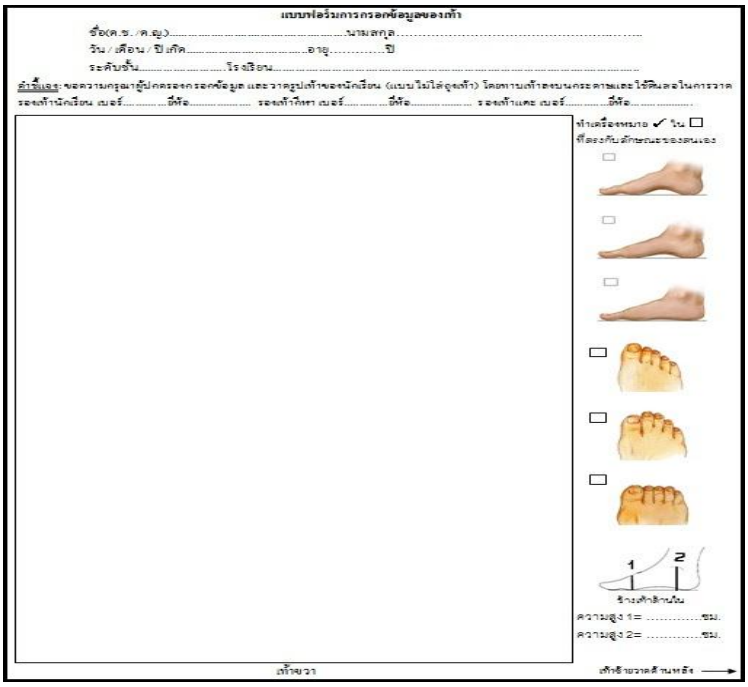

Figure 12. Questionaire Example

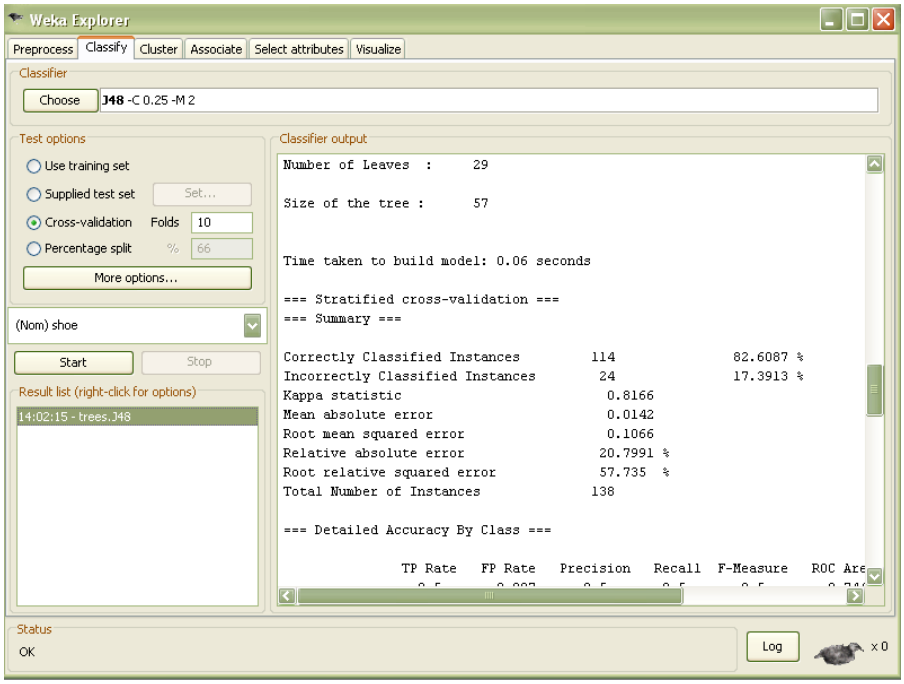

Figure 13. Results for Decision Tree from Weka (student shoes)

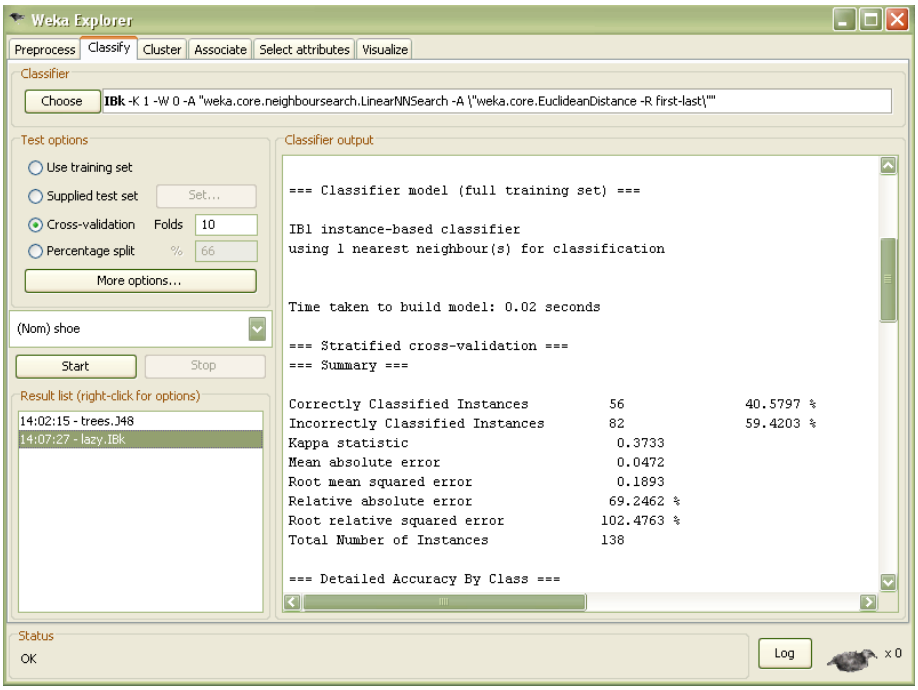

Figure14. Results for KNN from Weka (student shoes)

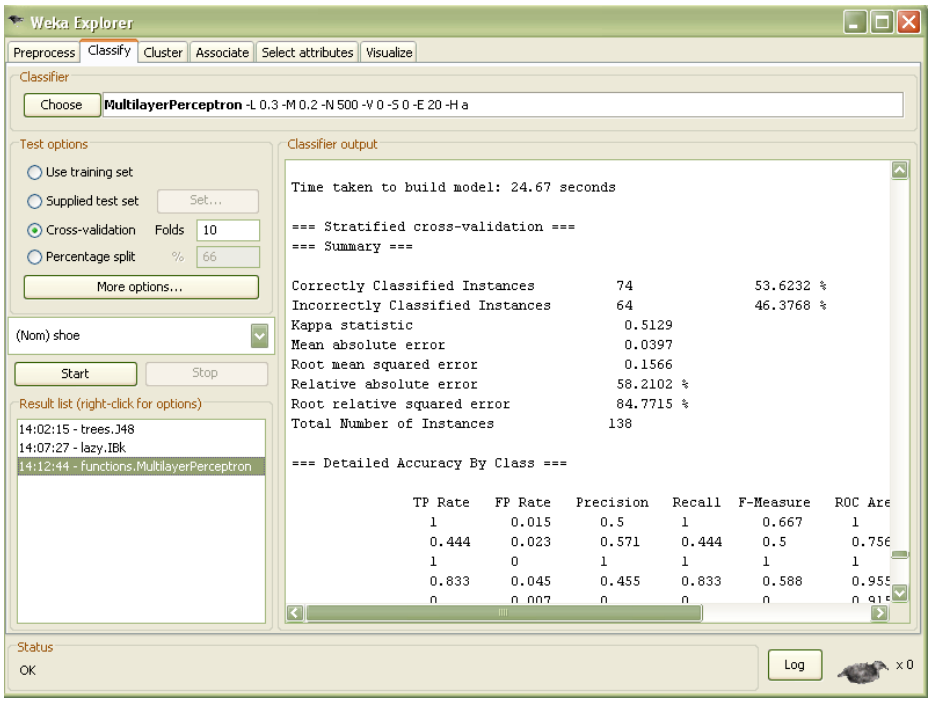

Figure 15. Results for Neural Net from Weka (student shoes)

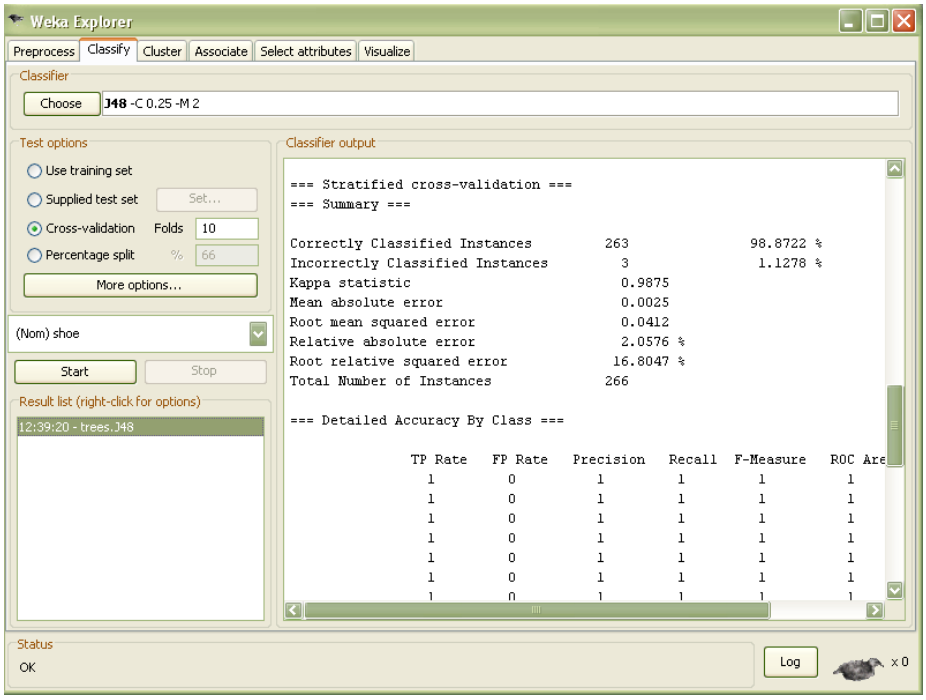

Figure 16. Results for Decision Tree from Weka (sport shoes)

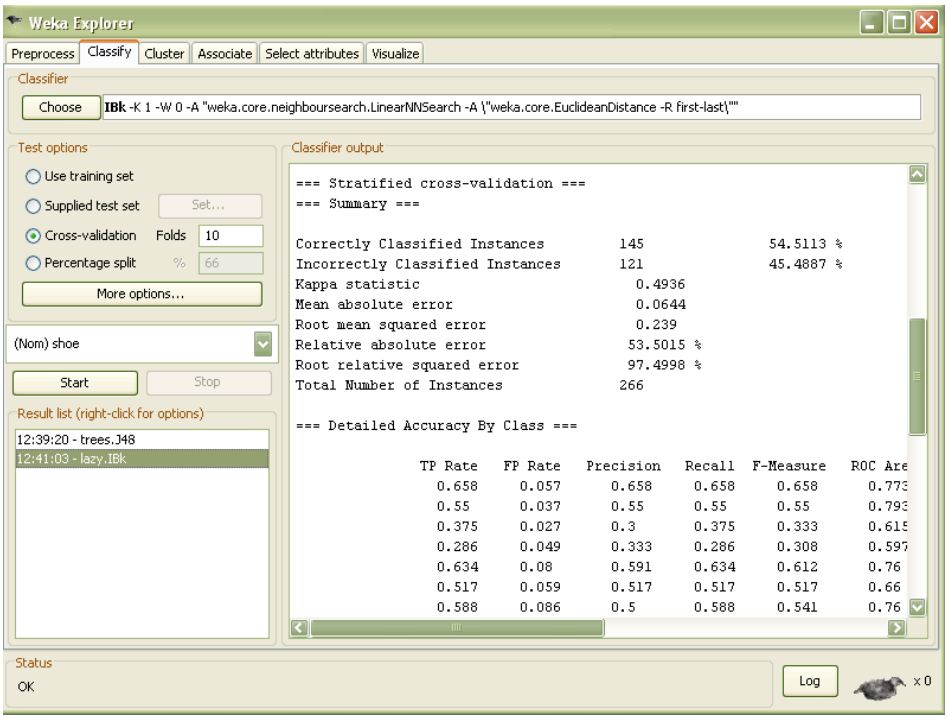

Figure 17. Results for KNN from Weka (sport shoes)

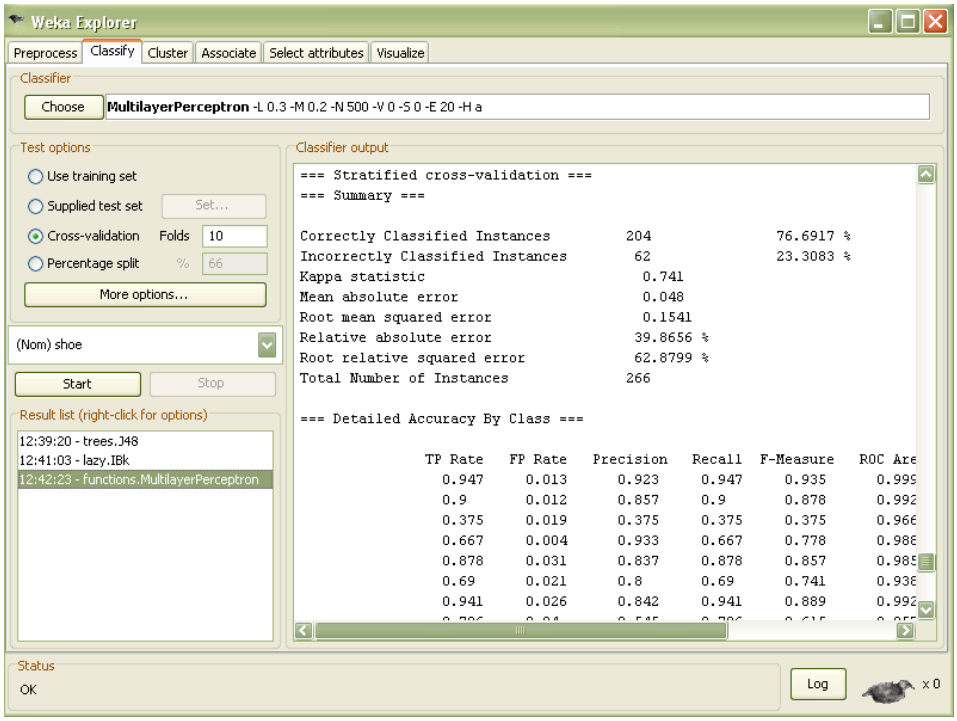

Figure 18. Results for Neural Net from Weka (sport shoes)

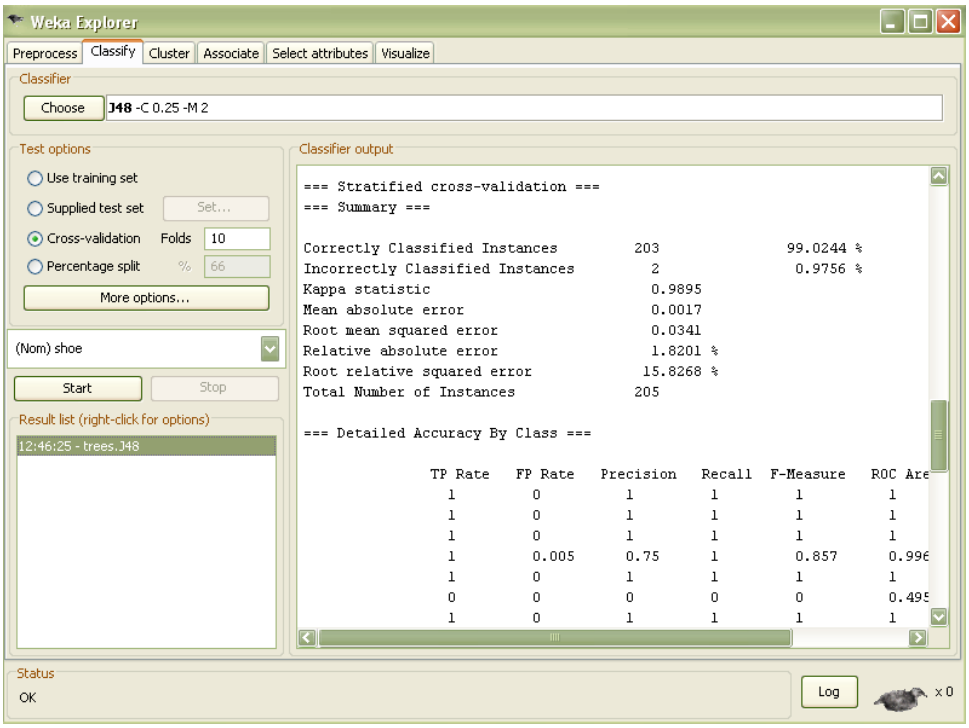

Figure 19. Results for Decision Tree from Weka (leisure shoes)

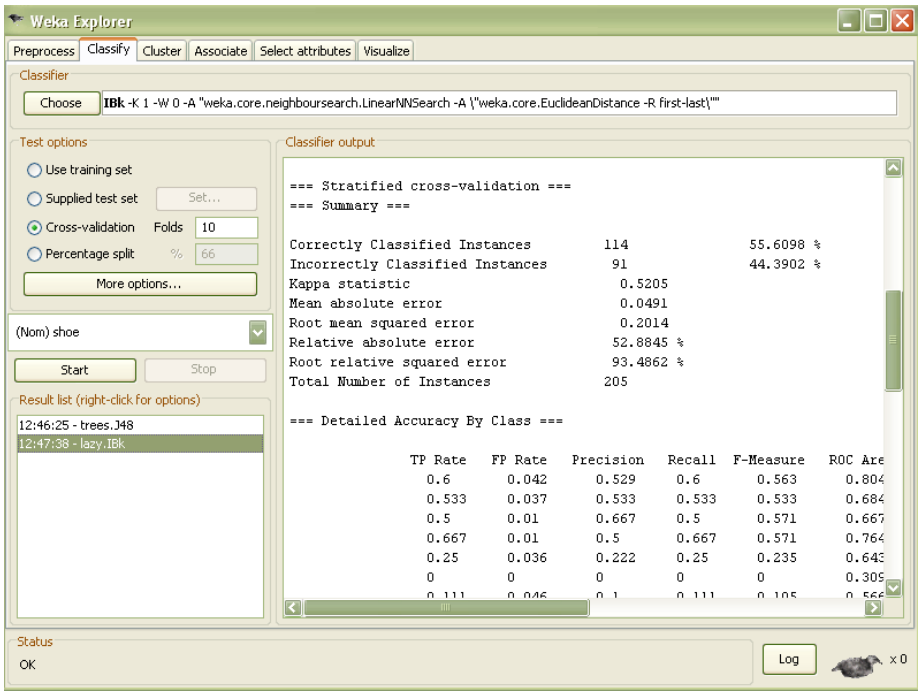

Figure 20. Results for KNN from Weka (leisure shoes)

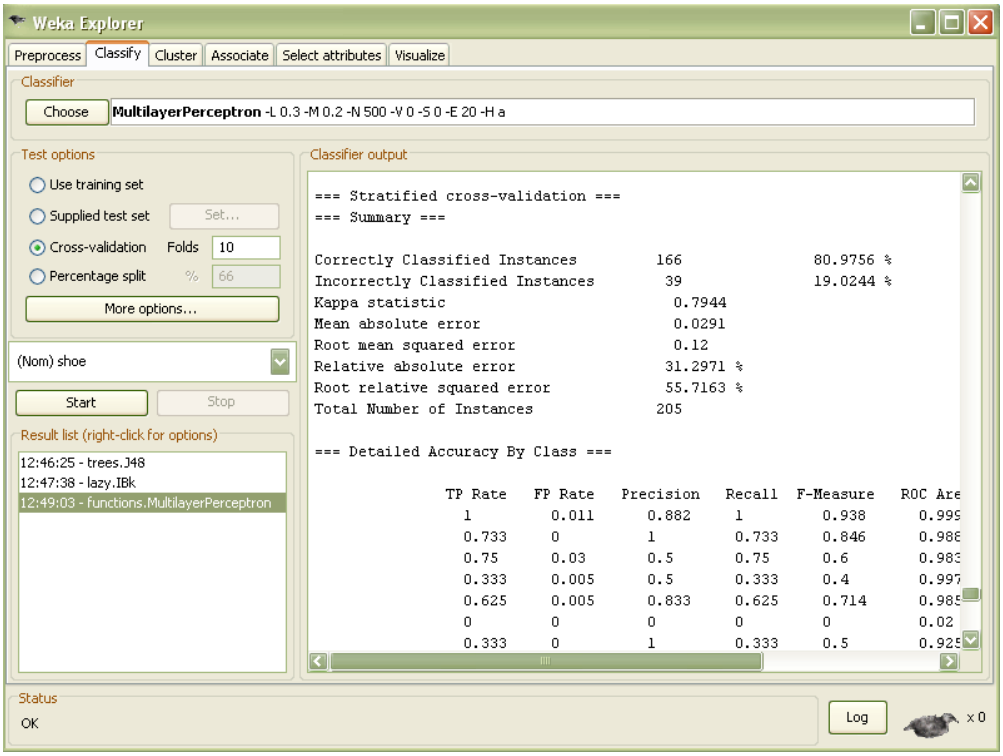

Figure 21. Results for Neural Net from Weka (leisure shoes)