Classification and Prediction on School Children about their Food Intake Attitude towards Food and Beverage Advertising on Television: KFC as a Case Study

Ahmad Fikri Anuar, Khairul Nizam Abd Halim
Faculty of Computer & Mathematical Sciences
Universiti Teknologi MARA
Cawangan Melaka, Kampus Jasin, Malaysia
boyfikri86@gmail.com

Roger Canda
Faculty of Computer & Mathematical Sciences
Universiti Teknologi MARA
Cawangan Pahang, Kampus Raub, Malaysia

***Abstract*— Serious health problems that occur in adults such as diabetes, hypertension, cardiovascular diseases are closely related to obesity in early childhood. Obesity has become a problem in Malaysia in the context of healthy lifestyle and in estimation, Malaysia has the highest rate of obesity in South-East Asia involving children. One of the most dominant mediums who promote unhealthy food is the Television Food Advertising (TVFA). An approach using Artificial Intelligence (AI) strategy, the Naïve Bayes (NB) technique is used to predict the eating behaviour of children toward TVFA. Five independent variables were used in the prediction model, which are advertisement recognition, favourite advertisement, purchase request, product preference and time spent watching TV. A total of 105 Year 6 pupils of SK Merlimau have been chosen as the target subjects to fulfil the objectives of the prediction model. 80% of the collected data were used as training data, and the other 20% are new data used to be tested. 31 prediction models were produced using this technique, and the results indicate a 78% accuracy from the data learnt. Although the accuracy result is not as expected (80% and above), Naïve Bayes could be implemented and may be continued by using other methods such as Support Vector Machine (SVM) and Artificial Neural Network (ANN). In the near future, hopefully there will be an extended work that utilises different techniques and independent variables used to increase the accuracy of results.**

Keywords*— *eating behaviour; naïve bayes; school children; television advertisement

I. INTRODUCTION

Obesity is becoming a widespread problem in Malaysia within the context of healthy lifestyle. Consequently, the accumulation of excess body fat as a result of obesity is also a cause for multiple other diseases [1]. In 2015, Malaysia became the most prevalent nation with the highest rate of obesity in South East Asia [2]. Obesity is a state that can be determined by the Body Mass Index (BMI) of an individual. Food advertising is one of the driving factors towards children obesity as the marketing on food is focused to achieve their target sales [3]. The purpose of food advertising is to induce craving for the advertised food among the viewers by using persuasive messages. Besides that, advertising also gives a positive impact on their labels or brands [3]. One of the most dominant determinants that encourages unhealthy food upon children is through Television Food Advertising (TVFA) [4]. Animation is used in TV commercial, especially in food marketing purposes to attract children's attention by combining elements of the fictional world of advertisements with reality [2]. Children who are exposed to TVFA have a higher rate of food consumption and are more likely to be obese compared to children who don't [5].

This is a study involving the obesogenic environment in Malaysia, aiming to evaluate five inducing factors: (1) advertisement recognition (AR) by children, (2) what are their favourite advertisements (FA), (3) purchase requested (PR) by them, (4) product preference (PP) and (5) TV viewing duration (TW) [2]. Taking these factors into account, children behaviour on food intake such as involvement in consuming the product can be determined. The children are categorized into subclasses according to their food intake behaviour with fast food. First subclass is for the non-consumer that rarely ate fast food, and the second will be for the few consumer that ate in moderation, and the final subclass will be for the heavy consumer that often feast on fast food [6].

The aim of this study is to make a prediction on school children's food intake and behaviour. Their behaviour prediction depends on what they watch on television; food advertisement in accordance to the IV stated before. The previous researches have been using quantitative approach such as correlation and survey research. A new approach based on Artificial Intelligence (AI) have been introduced in this study. This study contributes in term of knowledge in specific case study as well as experimental research approach was applied by using Artificial Intelligence strategy.

II. METHODOLOGY

Naïve Bayes (NB) is a simpler classifier if compared to other sophisticated classification algorithms, but much better than the others in terms of best performance in accuracy, precision, recall

and F-measure [7]. Naïve Bayes algorithm will be used as a method of prediction for this case study.

A. Dataset (Data collection)

The subjects chosen to fulfill the objectives of this study are 12 year-old students from SK Merlimau. 105 of the students' information were collected through the questionnaire given. Questionnaire is used to collect information regarding the food intake by the target subjects. The accuracy of the collected data is tested by using NB method.

B. Naïve Bayes classifier

The Naïve Bayes method uses knowledge of probability and statistics based on the application of Bayes' theorem, which can predict the class membership probabilities. Abstractly, Naïve Bayes is a conditional probability model given the problem needs to be instantly classified. The conditional probability is described as:

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)} \quad (1)$$

In plain English, using Bayesian probability terminology, the above equation can be written as

$$posterior = \frac{(prior)(likelihood)}{evidence} \quad (2)$$

Figure 1 below shows the key data structure of Naïve Bayes. It shows how the independent and dependent variables are constructed and joint together. In each category, there are associated independent variable with values attached to it which particularize the number of joints that category has with each categories of the dependent variable.

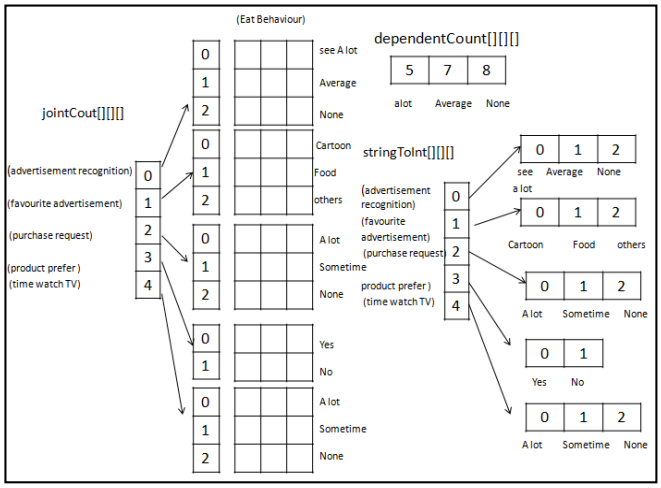

Fig. 1. Figure 1 Key Data Structure of Naïve Bayer for 5 Independent variables and 1 Dependent Variable.

Based on the combination of the IV as shows in Figure 1, a total of 31 prediction models were constructed for this particular case study. In each of the jointCount array, there are associated independent variables attached to it. For each of the variable, there are values attached. For example, the advertisement recognition variable has 'seen a lot', 'average', and 'none' values attached to it that shows the level of the variable. The jointCount array is used to store dependencies of IV and DV. Next, the dependentCount array is used to get the probability of the DV. The stringToInt array is used to assign value for each of the IV stated.

C. Statistical analysis

The performance of each model was evaluated by calculating the accuracy of the training data and the accuracy for new data testing. 105 data were collected. These data were collected from the same target subject. The data then will be used for the testing.

Percentage accuracy for training data was calculated from 85 data, which constitute around 80% of the overall data collected from the sampling. The 85 data were picked randomly to be utilized as training data so that the training of the model can be diversified and not just centered on an isolated arrangement of the sampling.

Meanwhile, percentage accuracy for new data was calculated from 20 data, which constitute around 20% from the overall data. These data are being tested to ensure the strength and accuracy of all models.

III. RESULT AND DISCUSSION

The 31 prediction models were established as sets of combination of five independent variables which include advertisement recognition (AR), favorite advertisement (FA), purchase request (PR), product preference (PP), and TV viewing duration (TW). All of these models are designed to predict eating behavior level that act as the dependent variable which consists of these categories; 'eat a lot of the product' (A lot), 'eat in a considerable amount' (Average) , and 'didn't eat that much' (Few).

A. Result

Table 1 below shows the test cases of the prediction model and also its accuracy for training data and new data.

Table 1 Percentage of accuracy testing for 80% (Training data) and 20% (New data).

	Independent Variables					Accuracy (%)	
Case	***AR***	***FA***	***PR***	***PP***	***TW***	***80% (training Data)***	***20% (New Data)***
1						67.86	61.90
2						66.67	61.90
3						73.91	61.90
4						67.86	61.90
5						69.04	52.38
6						67.86	61.90
7						75.00	61.90
8						70.24	61.90
9						70.24	52.38
10						73.81	61.90
11						69.04	61.90
12						66.67	57.14
13						75.00	61.90
14						73.81	61.90
15						72.62	57.14
16						75.00	61.90
17						70.24	61.90
18						67.86	57.14

	Independent Variables					Accuracy (%)	
Case	***AR***	***FA***	***PR***	***PP***	***TW***	***80% (training Data)***	***20% (New Data)***
19						77.38	61.90
20						72.62	57.14
21						72.62	57.14
22						76.19	61.90
23						72.62	61.90
24						72.62	57.14
25						76.19	57.14
26						75.00	61.90
27						73.81	61.90
28						73.81	57.14
29						76.19	61.90
30						78.57	61.90
31						78.57	61.90

Figure 2 below shows the class diagram for prediction model case 31 as in Table 1 with an accuracy of 78.57%. The UML diagram designed for the prediction model in this study domain and can be an example model design for others in developing their own system.

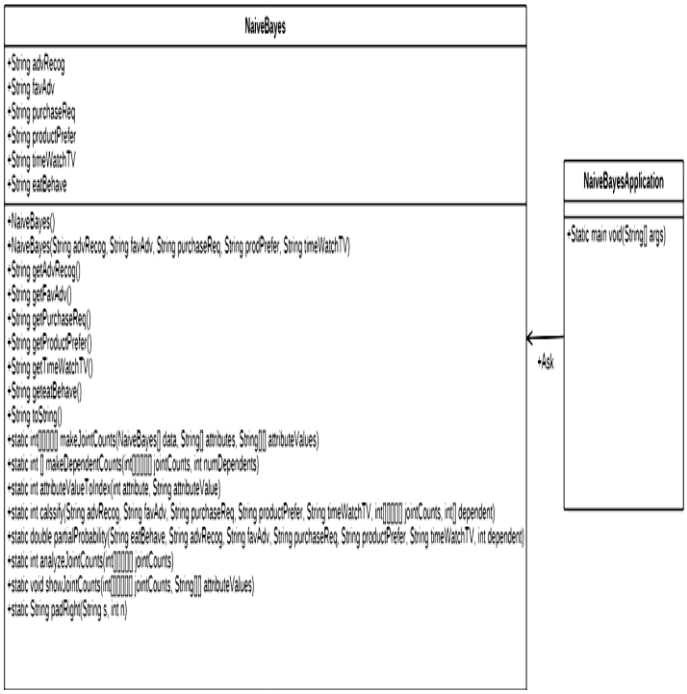

Fig. 2. Figure 2 UML Class Diagram for 5 IV based on Prediction Model.

B. Discussion

The combination of all of the independent variables and one of the four combinations record the highest accuracy for both 80% training data and 20% new data with the accuracy percentage of 78.57% and 61.90%. This shows that there are models capable of making predictions with almost 80% accuracy. However, there are several models that recorded the result of the lowest accuracy for both training data and new data accuracy percentage.

IV. CONCLUSION

In conclusion, Naïve Bayes is a method that can predict the behaviour of school children on their food intake toward television advertisement. Prediction Model can be constructed by using Naïve Bayes approach. The best result was obtained from the prediction model with all of the independent variables and it shows a percentage of accuracy of 78.57%. Another prediction model with only four independent variables, that is favourite advertisement, purchase request, product preference and TV viewing duration, shared the same percentage as the best result. Therefore, it can be concluded that both models can be used for the prediction.

The best result using the Naïve Bayes approach is to get an accuracy of 80% and above. However, the highest recorded accuracy is at 78.57%, just shy of 80. The result shows that some parts of the study still need improvement. Therefore, there is a need to do a further study with the same scope of study as there is a need to achieve the optimal accuracy percentage of the result.

ACKNOWLEDGMENT

We thank Universiti Teknologi MARA (UiTM) for all the helps and supports.

REFERENCES

[1] A. Must and R. S. Strauss, “Risks and consequences of childhood and adolescent obesity.,” *Int. J. Obes. Relat. Metab. Disord.*, vol. 23 Suppl 2, pp. S2-11, Mar. 1999.

[2] S. H. Ng, B. Kelly, C. H. Se, S. Sahathevan, K. Chinna, M. N. Ismail, and T. Karupaiah, “Reading the mind of children in response to food advertising: a cross-sectional study of Malaysian schoolchildren’s attitudes towards food and beverages advertising on television,” *BMC Public Health*, vol. 15, no. 1, p. 1047, 2015.

[3] F. Folkvord, D. J. Anschütz, E. Boyland, B. Kelly, and M. Buijzen, “Food advertising and eating behavior in children,” *Curr. Opin. Behav. Sci.*, vol. 9, pp. 26–31, 2016.

[4] G. Cairns, K. Angus, and G. Hastings, “The extent, nature and effects of food promotion to children: A review of the evidence to 2008,” no. December, 2009.

[5] J. C. Halford, E. J. Boyland, G. M. Hughes, L. Stacey, S. McKean, and T. M. Dovey, “Beyond-brand effect of television food advertisements on food choice in children: the effects of weight status.,” *Public Health Nutr.*, vol. 11, no. 9, pp. 897–904, 2008.

[6] J. M. Poti, K. J. Duffey, and B. M. Popkin, “The association of fast food consumption with poor dietary outcomes and obesity among children: Is it the fast food or the remainder of the diet?,” *Am. J. Clin. Nutr.*, vol. 99, no. 1, pp. 162–171, 2014.

[7] M. Bilal, H. Israr, M. Shahid, and A. Khan, “Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and {KNN} classification techniques,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 28, no. 3, p. , 2015.