

**ANALYSIS OF DAILY LIFE OF STUDENTS  
BY DATA MINING TECHNIQUE CASE  
STUDY: THE STUDENTS AT THE  
UNIVERSITY OF THE THAI CHAMBER OF  
COMMERCE**

**Sirithorn Jalernrat and Sirinard Tantakasem**

*School of Science and Technology,  
The University of The Thai Chamber of Commerce,  
Bangkok, Thailand  
e-mail: sirithorn\_jal@utcc.ac.th*

**Abstract**

This research aims to analyze the daily life of students at the University of the Thai Chamber of Commerce related to health by Data Mining techniques, using Decision Tree method. Cross-Industry Standard Process for Data Mining (CRISP -DM) concept was applied for the data analysis. Start from the clarifying the objectives, gathering data, data preparation, by converting it into a data that can be analyzed, modeling, evaluation and deployment. In the modeling phase, the data is classified by the Decision Tree, based on the Accuracy, Precision, Recall. The results of research have shown that most students will not take supplementary food. But the ones who take supplementary food mostly for skin care, better health, and increasing muscles respectively. For students who take supplementary food for decreasing weight, BMI means Healthy, Moderately Obese, which the Accuracy is 91.06%. Most female students take supplementary food for facial skincare. For male students who want to take lean Food and BMI means Healthy. They need to have supplementary food for increasing muscles which the Accuracy is 53.85%.

---

**Key words:** Data Mining, Decision Tree, Health.

## 1. Introduction

Nowadays, the consumers not just only want to have good health but also concern about beauty; complexion and shape. As it can be seen, there are more consumers having varieties of food, more than the basic 5 groups of food, which are clean food, lean food, and supplementary food, also having more exercises such as running, fitness and yoga.

According to the data above, the researcher is interested in analyzing the data of the students involving health by using Data Mining with the technique of Decision Tree method. Using Cross - Industry Standard Process of Data Mining (CRISP- DM) is the study and analysis of the students daily life data concerning with health for examples; the food consuming behavior and exercises, the range of bed time and wakeup time, and health interests. Then calculate the Body Mass Index (BMI) and bring all the variations and analyze the data with the data-mining technique. At this point, the researcher chooses the classification by using Decision Tree method and then evaluates the result for the efficiency.

## 2. The procedure of the research

The researcher does the research by using CRISP-DM method for bringing the knowledge of using Data-Mining which consists of 6 steps as follows:

(1) Business Understanding (2) Data Understanding (3) Data Preparation (4) Modeling (5) Evaluation (6) Deployment. In each step may have to go back to the step earlier to change or adjust for the best result. [4], [6].

### (1) Business Understanding

It is the first step of CRISP-DM. It focuses on the problem understanding, the purposes of procedures and converts the problem to the data-mining analysis pattern, also plans the future procedure roughly. The other researches concerning this are as follows.

Yimprasert (2017) studied and compared food consumption behavior of 1st year students in Rajamangala University of Technology Isan, Nakhon Ratchasima based on their sex, faculty, cost of living per month, habitat and body mass index. The results of research found that students with different sex, faculty, cost of living per month, habitat, and BMI did not changed consumption behavior [11].

Pongpipat, Yantaragorn and Pukkama (2017) studied the health behavior of CMUs students who enrolled in the sport for health course, the second

semester of an academic year 2016. The research findings were, on food consuming behavior, there was high percentage of unhealthy food consumption of students; sugar, salt and greasy foods were highly consumed [5].

Vimonwattana, Sangkapong and Panriansaen (2017) studied health promotion behaviors and the relationship among bio-social factors, predisposing factors, enabling factors, reinforcing factors and health promotion behaviors of professional nurses in Medicine Vajira Hospital, Navamindradhiraj University. The results of research found that Bio-social factors and predisposing factors were not correlated with health promoting behaviors while enabling factors and reinforcing factors were significantly correlated with health promoting behaviors of professional nurses [9].

Vinichaiyanun, Vichitthamaros (2017) studied factors affecting the consumption of weight control dietary supplements of people in Bangkok. The study sample consists of people aged 20-59 years and living in Bangkok. Results, analyzed by logistic regression show that demographic factors, the persons aged between 30-39 years are 3.7 times more likely to consume the weight control dietary supplement product than those aged between 20-29 years [7].

Healthy food is the food that is useful for your body. In nutrition, it supports the working system of the body to decrease the risk of diseases which is worth for the body [2].

Clean Food is the non-toxic seasoning or any transforming. This food is fresh. It is less seasoning or non-seasoning. It focuses on natural food, no any fermented procedures over seasoning [1].

Lean Food is consist of less fat, focuses on protein. It is the food that is less in carbohydrate, less in sugar. Having food in low fat or nothing is good for people who love having good shape, constructing the body, building body muscles or someone who regularly exercises. Having lean food will give you in good shape [10].

## **(2) Data Understanding**

In this step the researcher has collected the data of the students at the University of the Thai Chamber of Commerce by distributing the questionnaire using Random Sampling. There are 410 student samplings. Then the researcher has revised the collecting data to check the data correctness, and considers to use all or partially data to be used in this analysis.

The topics using in this questionnaire are:

1. Basic Personal Data; sex, age, height, weight.

2. Daily Life behavior in eating, exercising, and resting data. For examples, Breakfast-Type, Dinner- Type, Dietary Supplements, Buy-Food-Reason, Wake-up Time, Exercise-Type, etc.

3. The researcher evaluates the Body Mass Index (BMI) and the result calculating differentiates according to the table 1. Body mass index is defined as the individual body weight divided the square of his/her height using equation (1).

$$BMI = \frac{Weight(kg)}{Height(m)^2} \quad (1)$$

Table 1: The meaning of body mass index

BMI	Description
BMI < 18.5	Underweight
18.5 - 22.9	Healthy
23.0 - 24.9	Moderately Obese
25.0 - 29.9	Overweight
BMI ≥ 30	Severely Obese

4. Then the researcher has considered all the data collecting to check the correctness and revised the important of the data to use in this analysis; all or partially data by rechecking and choosing. At this point, weight, height are not used in this evaluation as BMI is used instead. Age is not used here as there is no significantly difference.

### (3) Data Preparationng

In this step is the converting the raw data to the data cleaning as it can be used in next step, examples; convert the data to usable data, add the missing data, and delete Outlier which means to delete the different value to make it correct ( Data Cleaning). So the researcher has done this method from this data.

### (4) Modeling

In this step is the data analysis using Data Mining technique. The researcher has used supervised learning in this data analysis by using technique of Decision Tree method, which are easy, more convenience to analyze the problem. The researcher has selected RapidMiner Software to calculate data by using Decision Tree method.

The Procedure of data analysis using data classification techniques with Decision Tree method.

The researcher has chosen Decision Tree method as it has easily understanding way, the patterns are easy to learn according to the structured data and also chosen Split Test to test the model. After that the researcher has divided the data for the test. The 70% of the data is for constructing the model which is called the Training Set. The 30% of the data is for the Testing the model, which is called Test Set. By constructing the Decision Tree model, it is used Information Gain calculation of each Node comparing with Node in Class to find which Node has the maximum Information Gain to be the Root Node of Decision Tree. The researcher has used the Gain Ratio in solving the unfair problems in comparing the variation in Split Node from Information Gain by adjusting the Information Gain by dividing SplitINFO which can be calculating as in equation (2) When bringing the SplitINFO divided from Information Gain, the Gain Ratio will come out as the equation (3) The researcher also has used Accuracy to evaluate the model to check the correctness of the model considering all classes. Using Precision to evaluate the model prediction, considering separately class and using Recall to evaluate the model correctness, considering separately class.

$$SplitINFO = - \sum_{i=1}^k \frac{n_i}{n} \log \frac{n_i}{n} \quad (2)$$

where  $n_i$  = sample amount at Node # $i$ , and  $n$  = sample amount at parent node.

$$GainRATIO_{split} = \frac{GAIN_{split}}{SplitINFO} \quad (3)$$

## (5) Evaluation

In this step, the researcher has had the analysis data result by using Data-Mining technique. Before bringing the result to use in next step, the researcher has to verify the effectiveness of the result by looking at the Accuracy value. It may go back to the step before to adjust the result for the exactly wants. Tree method.

## (6) Deployment

The result from this research has shown the useful knowledge to the researchers, health and nutrition specialists, and marketing persons to better understanding and could bring this result to do further marketing, health and nutrition analysis.

### 3. The research results

The basic data analysis can be seen in table 2.

Table 2: The basic statistical analysis

Variables	Description	Statistics
Sex	Male, Female	Female = 57.32 % Male = 42.68%
BMI	Body Mass Index	Healthy 52.20 % Underweight 18.78 % Overweight 12.44 % Moderately Obese 10.48 % Severely Obese 6.10 %
Breakfast	Y= have breakfast N = have not breakfast	have breakfast = 68.05 % have not breakfast = 31.95%
Reason – Breakfast (N)	The reason for not eating breakfast - no time, not hungry, save money, wake up late	no time = 76.59% not hungry = 15.61% save money = 4.63% wake up late = 3.17%
Feeling – Breakfast	Feeling when not eating breakfast - so-so, hungry, feel dizzy, etc.	hungry = 66.34% , so-so = 32.68%, etc. = 0.98%
Dinner-Type	Normal Foods, High-Protein Foods, Diet Foods	Normal Foods = 83.17%, Diet foods = 15.12% High-Protein Foods = 1.71%
Supper	Y= have supper N = have not supper	have not supper = 54.63 % have supper = 45.37%
Dietary supplements	Y= take dietary supplements N = take not dietary supplements	take not dietary supplements = 57.56% take dietary supplements = 42.44%
dietary supplements-Reason	The reason for take dietary supplements - better health, gain muscle, facial skincare, to lose weight, etc.	The most facial skincare followed by better health, gain muscle, to lose weight, respectively
Bed time	before 3 pm, 3 pm – 4 pm, 4 pm – midnight, after midnight	The most after midnight, 4 pm – midnight, 3 pm - 4pm, respectively
Wake up time	before 6 am,	The most 6 am - 7 am, 7 am - 8 am, after 9 am .

The analysis of using Data Tree Decision Method has Node as dietary supplementary Reason followed by BMI-meaning, Read- Nutrition information (in figure1.). Target Variable is take dietary supplements. Conclusion of efficiency of Decision Tree is Accuracy 91.06% (in figure1.), predicted that take dietary supplement = Yes, Class Precision = 87.04%, Class Recall = 92.16%, predicted

Table 2 (cont.): The basic statistical analysis

	6 am - 7 am, 7 am - 8 am, 8 am - 9 am, after 9 am	8 am - 9 am ,before 6 am, respectively
Exercise-often	everyday, 1-3 times per week at least, 1-3 times per month at least, don't exercise	The most 1-3 times per month at least, don't exercise, 1-3 times per week at least, everyday, respectively
Exercise-type	play football, running, play badminton, yoga, going to gym, etc.	The most running , going to gym, respectively
Buy-Food-Reason	tasty, food safety and quality, price, hungry	tasty = 59.76 % food safety and quality = 26.34% price = 13.41%, hungry = 0.49%
Food-Type-Want	Delicious Food, Clean Food, Lean Food, Vegetarian food	Healthy Food = 43.17% Delicious Food = 26.10% Lean Food = 16.10% Clean Food = 12.20% Vegetarian Food = 2.43%
Level-important-healthy food	Low, Medium, High	Medium = 69.02% High = 19.27% Low = 11.71%
Interest-food safety(Y,N)	Y = interest in food safety N = have no interest in food safety	interest in food safety = 94.39% have no interest in food safety = 5.61%

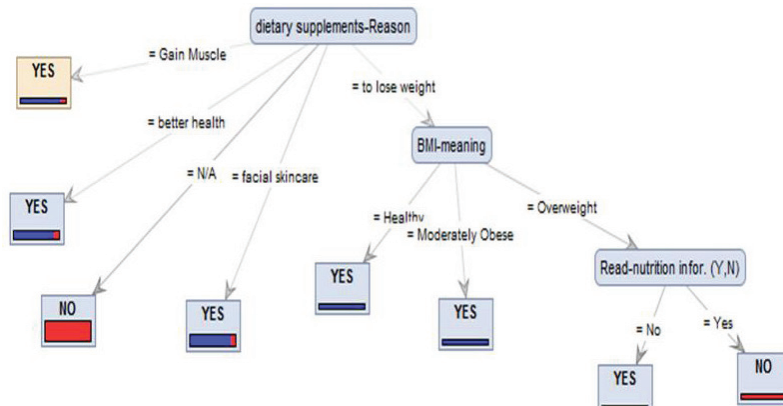


Figure 1: Decision Tree, Target Variable is dietary supplements

that take dietary supplement = No, Class Precision = 94.20%, Class Recall = 90.28%.

The tree from the root node to the leaf node can be converted in to an if-then Rule. From figure 1, Decision Tree can be built Rule Models in the rule of if-then as follows.

```

if dietary supplements-Reason = gain muscle then dietary supplements = YES
  if dietary supplements-Reason = N/A then dietary supplements = NO
  if dietary supplements-Reason = better health then dietary supplements = YES
    if dietary supplements-Reason = facial skincare then dietary supplements = YES
      if dietary supplements-Reason = to lose weight and BMI-meaning = Healthy then dietary supplements = YES
        if dietary supplements-Reason = to lose weight and BMI-meaning = Moderately Obese then dietary supplements = YES
          if dietary supplements-Reason = to lose weight and BMI-meaning = Overweight and Read-nutrition infor. (Y,N) = No then dietary supplements = YES
            if dietary supplements-Reason = to lose weight and BMI-meaning = Overweight and Read-nutrition infor. (Y,N) = Yes, then dietary supplements = NO
  
```

From Decision Tree can be concluded that most students do not take supplementary food. For the ones who take it mostly is for skin care, better health, and gain muscle respectively. Considering Node is BMI- Meaning, found that the students who take supplementary food for lose weight, BMI means Healthy, Moderately Obese.

Bringing Model Test by using Target Variable is Dietary Supplements- Reason which is focusing on the students who only take supplementary food to analyze. The result is Decision Tree in figure 2.

Conclusion of Efficiency of Decision Tree in figure 2 is Accuracy at 55.77% predicted that facial skin care with Class Precision = 71.88%, Class Recall = 69.70%, predicted that gain muscle with Class Precision = 37.50%, Class Recall = 42.86%, predicted that better health with Class Precision = 25.00%, Class Recall = 25.00%.



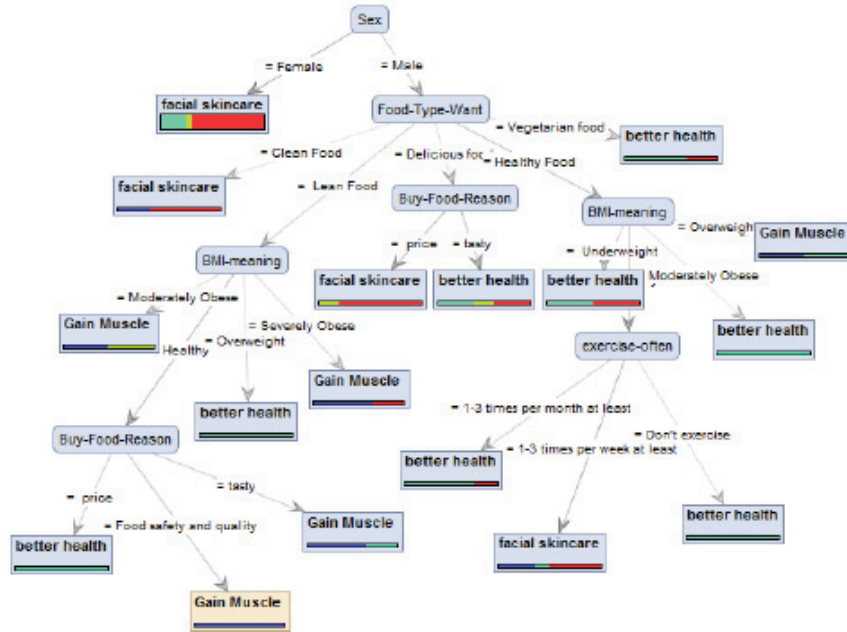


Figure 2: Decision Tree, Target variable is diet dietary supplements-Reason.

Looking at the Decision Tree, it is found the reason of taking supplementary food for female students is for facial skincare as in Rule Model which is If- Sex = Female then facial skincare. In male students have many variation issues such as Food Type Want, Buy Food Reason, BMI. The researcher has created Decision Tree by looking at other variations and focusing the Accuracy value at 50% up. The result is followed in table 3.

Table 3: The Result of Decision Tree to Tree to Rules

Target Variable	Nodes	Tree to Rules	Accuracy
Reason Breakfast (N)	Feeling - Breakfast(N)	if Feeling - Breakfast(N) = hungry then Reason - Breakfast(N)= no time	66.67%
Supper	Reason - Breakfast(N)	if Feeling - Breakfast(N) = so-so then Reason - Breakfast(N) = no time	66.67%
	Supper, Bed time, Wake up time	if Bed time = 4 pm - midnight and Wake up time = 6 am - 7 am then Supper = NO if Bed time = 4 pm - midnight and Wake up time = 7 am - 8 am then	

Table 3 (cont.): The Result of Decision Tree to Tree to Rules

		<p>Supper = NO</p> <p>if Bed time = 4 pm - midnight and Wake up time = before 6 am then Supper = NO</p> <p>if Bed time = after midnight and Wake up time = 6 am - 7 am then Supper = YES</p> <p>if Bed time = after midnight and Wake up time = 7 am - 8 am then Supper = YES</p> <p>if Bed time = after midnight and Wake up time = 8 am - 9 am then Supper = YES</p> <p>if Bed time = after midnight and Wake up time = after 9 am then Supper = YES</p>	
Dietary supplements Reason	Sex dietary supplements Reason, Food-Type-Want	<p>if Sex = Female then dietary supplements-Reason = facial skincare</p> <p>if Sex = Male and Food-Type-Want = Lean Food and BMI-meaning Healthy then dietary supplements Reason = gain muscleaning</p>	53.85%
Exercise-often	Exercise-often, Exercise-type	<p>if Exercise-type = running then Exercise-often = 1-3 times per month at least</p> <p>if Exercise-type = going to gym then Exercise-often = 1-3 times per month at least</p> <p>if Exercise-type = play football then Exercise-often = 1-3 times per week at least</p>	53.70%
Buy-Food-Reason	Buy-Food-Reason, BMI-meaning Food-Type-Want, Level-important- Level-important- healthy foodant-	<p>if BMI-meaning = Healthy and Food- Type-Want = Healthy Food and Level-important-healthy food = Medium then Buy-Food-Reason = tasty</p> <p>if BMI-meaning = Healthy and Food Type-Want = Healthy Food and Level-important-healthy food = High then Buy-Food-Reason = food safety and quality</p>	51.22%

## 4. Conclusions

This research is the analysis of daily life of students related to health by using data classification techniques with the 410 student samplings. The characteristics are Female = 57.32%, and Male = 42.68%. Mostly the BMI means Healthy = 52.20%, Underweight = 18.78% and Overweight = 12.44%, respectively.

Dividing data for effectively measurement of data classification model is tested by using Split Test, which the Training Set = 70% from all data and then Test Set = 30% for evaluated Accuracy models. Looking at If-then Rule from Decision Tree can be concluded as follows.

Most students do not take supplementary food. For the ones who take supplementary food mostly for skincare, better health, and gain muscle, respectively. Concerning with Node BMI - Meaning found that the ones who take supplementary food for lose weight, The BMI means Healthy, Moderately Obese, which the Accuracy is 91.06%.

Most female students take supplementary food for facial skin care. While male students who want to take supplementary food is for muscle increasing. They want Lean food and BMI means healthy, which the Accuracy is 53.85%

Students who feel hungry and not hungry when not having breakfast, have the same reason, that is there is no time, which the Accuracy is 66.67%.

For the reason buying food, which Accuracy is 51.22%, can be said; 1) If the students BMI means healthy and do need healthy food and level of the important of healthy food at average level then they would buy food because of the taste. 2) If the students BMI means healthy and do need healthy food and level of the important of healthy food at high level then they would buy food with safety and quality.

## References

- [1] C. KUSUMA, Faculty of Medicine Siriraj Hospita, Clean Food, Retrieved from [http://www.si.mahidol.ac.th/sidoctor/e-pl/admin/article.files/1205\\_1.pdf](http://www.si.mahidol.ac.th/sidoctor/e-pl/admin/article.files/1205_1.pdf)
- [2] Department of Health, Ministry of Public Health. Good Body, Good Health, Just using 4 Behaviors, (2018), Retrieved from <http://nutrition.anamai.moph.go.th/images/files/HLworkingage.pdf>
- [3] G. Kesavaraj and S. Sukumaran, *A study on classification techniques in data mining*, Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), (2013).
- [4] P. Eakasit, *Data Mining Trend*, (2014), Retrieved from <http://dataminingtrend.com/2014/data-mining-techniques/crisp-dm-example>.
- [5] P. Sarawut, Y.Patara and P. Thachakorn, *Health Behavior of Students in Chiang Mai University*, CMU Journal of Education, Vol. 1 No. 1, (2017), pp. 34-45.

- [6] T. Pang-Ning, S. Michael and K. Vipin, *Introduction to Data Mining*, Pearson Addison Wesley, (2015).
- [7] V. Chawan and V. Preecha, *Factors Affecting Weight Control Dietary Supplements Consumption of People in Bangkok*, WMS Journal of Management Walailak University. Vol. 6 No.1, (2017), pp. 84-90.
- [8] VOGUE, New Trend is Coming Eat Lean, Retrieved from <https://www.vogue.co.th/eat-lean>
- [9] V. Nonthacha, S. Tipapan and P. Rattana, *Factors Affecting the Health Promotion Behaviors of Professional Nurses in Faculty of Medicine, Vajira Hospital, Navamin-dradhiraaj, University*, Kuakarun Journal of Nursing, Vol.24 No.2, (2017), pp. 67-81.
- [10] We Fitness Society, EAT CLEAN VS EAT LEAN, Retrieved from <https://society.wefitnesssociety.com/eat-clean-vs-eat-lean->
- [11] Y. Siripaisarn, *Food Consumption Behavior of Undergraduate Student Level 1 in Rajamangala University of Technology Isan Nakhon Ratchasima*, Ratchaphruek Journal, Vol.15, No.1, (January April 2017), pp. 33-41.