CONSTRUCTING A RISK BEHAVIOR GUIDELINE FOR ADOLESCENT STUDENT USING DECISION TREE

SURIN AUNSAN

A THEMATIC PAPER SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE (INFORMATION TECHNOLOGY MANAGEMENT) FACULTY OF GRADUATE STUDIES MAHIDOL UNIVERSITY 2016

Thematic Paper
entitled

CONSTRUCTING A RISK BEHAVIOR GUIDELINE FOR ADOLESCENT STUDENT USING DECISION TREE

Surin Aunsan
.....
Miss. Surin Aunsan
Candidate

.....
Lect. Sotarat Thammaboosadee,
Ph.D. (Information Technology)
Major advisor

.....
Asst. Prof. Supaporn Kiattisin,
Ph.D. (Electrical and Computer
Engineering)
Co-advisor

.....
Prof. Patcharee Lertrit,
M.D., Ph.D. (Biochemistry)
Dean
Faculty of Graduate Studies
Mahidol University

.....
Asst. Prof. Supaporn Kiattisin,
Ph.D. (Electrical and Computer
Engineering)
Program Director
Master of Science Program in
Information Technology Management
Faculty of Engineering
Mahidol University

Thematic Paper
entitled

CONSTRUCTING A RISK BEHAVIOR GUIDELINE FOR ADOLESCENT STUDENT USING DECISION TREE

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science
(Information Technology Management)
on
January 4, 2016

Surin Aunsan
.....
Miss. Surin Aunsan
Candidate

.....
Lect. Taweesak Samanchuen, Ph.D.
(Electrical Engineering)
Chair

.....
Lect. Sotarat Thammaboosadee, Ph.D.
(Information Technology)
Member

.....
Asst. Prof. Supaporn Kiattisin, Ph.D.
(Electrical and Computer Engineering)
Member

Kairoek Choeychuen.
.....
Asst. Prof. Kairoek Choeychuen, Ph.D.
(Electrical and Computer Engineering)
Member

.....
Prof. Patcharee Lertrit,
M.D., Ph.D. (Biochemistry)
Dean
Faculty of Graduate Studies
Mahidol University

.....
Asst. Prof. Jackrit Suthakorn, Ph.D.
(Robotics)
Dean
Faculty of Engineering
Mahidol University

ACKNOWLEDGEMENTS

This thematic paper would not be success without advices, supports and encouragements from many people. I would like to thanks my research supervisors Dr. Sotarat Thammaboosadee, Asst. Prof. Supaporn Kiattisin and Dr. Taweesak Samanchuen. For their support through my research, they suggested me on research of this topic and necessary knowledge on topic of data mining, and using tools. They was gave me a good advices and taught me about thinking process.

Most importantly, none of this could research happened without my family, my friend and colleague who always stays beside and encourage me when I was under pressure and discouraged. Although they do not know deeply in this topic but they tend to support me as much as possible.

Surin Aunsan

CONSTRUCTING A RISK BEHAVIOR GUIDELINE FOR ADOLESCENT STUDENT USING DECISION TREE

SURIN AUNSAN 5737291 EGIT/M

M.Sc. (INFORMATION TECHNOLOGY MANAGEMENT)

THEMATIC PAPER ADVISORY COMMITTEE: SOTARAT THAMMABOOSADEE, Ph.D., SUPAPORN KIATTISIN, Ph.D.

ABSTRACT

This research proposes construction of six risk behavior models and creates the risk student behavior guideline. This study adapted the concept of filleting student behavior process of the Student Care and Support System, and applied it to student profile database. The risk behavior is separated into six types: Ability, Healthy, Economy and Family, Drug, Safety and Sexuality. Six risk behavior determinations are then defined the student behavior into three groups: Normal, Risk and Problem. The C4.5 decision tree algorithm of data mining technique is to build the six risk behavior models and provides an overall accuracy over 90%. The models based on risk behavior groups are then used to create a student behavior guideline to support the decision making of teacher. Finally, after creating the guideline to observe the behavior of student, the apparent information and knowledge provide more beneficial for consideration in Student Care and Support System.

KEY WORDS: DECISION TREE / RISK BEHAVIOR / ADOLESCENT

50 pages

การสร้างแนวทางปฏิบัติสำหรับพฤติกรรมความเสี่ยงสำหรับนักเรียนวัยรุ่นโดยใช้ต้นไม้ตัดสินใจ
CONSTRUCTING A RISK BEHAVIOR GUIDELINE FOR ADOLESCENT STUDENT USING DECISION TREE

สุรินทร์ อุ่นแสน 5737291 EGIT/M

วท.ม. (การจัดการเทคโนโลยีสารสนเทศ)

คณะกรรมการที่ปรึกษาสารนิพนธ์ : โยฑศ์รัตต ธรรมบุษดี, Ph.D., สุภาภรณ์ เกียรติสิน, Ph.D.

บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์เพื่อสร้างแบบจำลองพฤติกรรมเสี่ยงของนักเรียนจำนวน 6 รูปแบบ และสร้างแนวทางปฏิบัติสำหรับพฤติกรรมเสี่ยงของนักเรียน ซึ่งในงานวิจัยนี้ได้นำหลักการคัดกรองพฤติกรรมนักเรียนของระบบดูแลช่วยเหลือนักเรียนมาประยุกต์ใช้กับฐานข้อมูลประวัตินักเรียนโดยจำแนกพฤติกรรมนักเรียนออกเป็น 6 ด้าน ได้แก่ ด้านความสามารถ, ด้านสุขภาพกาย, ด้านครอบครัวและเศรษฐกิจ, ด้านยาเสพติด, ด้านความปลอดภัย และด้านพฤติกรรมทางเพศ พฤติกรรมแต่ละด้านประกอบไปด้วยนักเรียนจำนวน 3 กลุ่ม ได้แก่ กลุ่มปกติ, กลุ่มเสี่ยงและกลุ่มมีปัญหา การนำวิธีการทำเหมืองข้อมูล โดยใช้วิธีการจำแนกข้อมูลด้วยอัลกอริทึม C4.5 มาสร้างแบบจำลองซึ่งมีค่าความน่าเชื่อถือมากกว่า 90% และสร้างแนวทางปฏิบัติสำหรับจำแนกพฤติกรรมนักเรียนเพื่อมอบให้ครูที่ปรึกษาใช้เป็นเครื่องมือประกอบการตัดสินใจการคัดกรองและจำแนกนักเรียน สุดท้ายงานวิจัยนี้การสร้างแบบจำลองและแนวทางปฏิบัติสำหรับจำแนกพฤติกรรมนักเรียนทำให้ได้รูปแบบพฤติกรรมของนักเรียนที่มีความเหมาะสม ถูกต้อง ชัดเจนและ สามารถนำไปประยุกต์ใช้งานได้ดียิ่งขึ้น

50 หน้า

CONTENTS

CONTENTS (cont.)

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

1.1 Statement of the problem

Nowadays, the terrible situations of Thai youth represent demoralization in society. A survey of childbirth in private hospital, the United Nations Population fund (UN) and the office of the national economic and social development board found that a teenager (15-19 years old) pregnant 130,000 in 2014. In addition, 15,000 of teenage mothers late delivered a child [1]. The interesting point of view on this research is that the teenager mother late delivered a child per hour or every two days or late nineteen moms delivered a child. The results of research represent that teenager's pregnancy is a continuous problem [2]. The research Thanyarak Institute in 2013, this year surveying 7,960 patients (86.11 % of male and 13.89 % of female) found that 18.12 % or 1,422 of patients aged 15 to 19 years old [3]. Based on surveys from Department of Mental Health in 2014, the teenager's violence rate is 879 cases per year [3]. The results of those researches showed that emotional states of teenagers are in lack of control. The problem impacts on society, therefore, all parts of society must find the way. However, the governance must seriously drive the educational policy of the School children counselling of the teacher in school. More specifically, the Office of the Basic Education Commission of Thailand [4] defines six of information groups including student abilities, health, family and economy, drug, safety, and sexuality to provide the student group classification which is a process used in Student Care and Support System. Hence, it could be beneficial if there is some of systematic system to identify and classify the level of each risk behaviors from student general information.

To analyze and define factors of problem, Applying technology is one of the ways for supporting the system. The Data mining is the extraction of implicit, previously unknown, and potentially useful information from data [5]. The most useful data mining technique is to determine and classify the root or the factors of problem related to student's behaviors. The Decision Tree, one of classification method, is a

predictive data mining technique that makes data prediction using the known results from different data [6].

This paper provides a construction of risk behavior classification models using the decision tree with student behavior guideline. These models support data analysis on the online student's profile to provide the student counselling. In addition, adviser can use this guideline to improve the behavior level of students.

The remainder of the paper is organized as follows: the next section discusses the related works of education domain. Section 3 introduces the methodologies and experimental result. Finally, we conclude the paper and suggestion directions for future works.

1.2 Research purpose

- To build the six adolescent student behavior models.
- To create the student behavior guideline from the six students behavior models.

1.3 Scope of research

- This paper investigates many factors related to student's behaviors, and generates the accurate model to provide both the Student Care and Support System.
- Student classification group is based on care and support system guideline of the Department of Mental Health.
- A source of data is taken from the student profile database (2012-2014).
- A groups of student's behaviors are classified into 3 groups, give as normal, risk and problem groups.

1.4 Benefits of research

This proposed at risk student behaviors guideline that would be applied to the Student Care and Support System of institution.

CHAPTER II
LITERATURE REVIEW

This chapter presents the literature review related to the Student Care and Support System, Data mining theory and worked, give as:

2.1 The Student Care and Support System

2.2 Data Mining Definitions

2.3 Cross-Industry Standard Process: CRISP–DM

2.4 Data Mining Task

2.5 The Decision Tree

2.6 Related worked

2.1 The Student Care and Support System

The Student Care and Support System [4] is defined as a system approach improvement, development and carefulness a student to fulfill student's potential and improve student's property that is responsible for good quality of life.

The system is a process of prediction risk behaviors to improve and develop student behavior. The standard tool for operation provides a teacher, internal and external academic personnel (Parent, Community and Manager).

The system concept includes three concerning points, as followed:

1. All of the students need attention, opportunity, intelligence, and happiness in life.

2. The time and the method of learning provide to the difference of student's ability.

3. The institution corporation is related to the successful methodology for providing student.

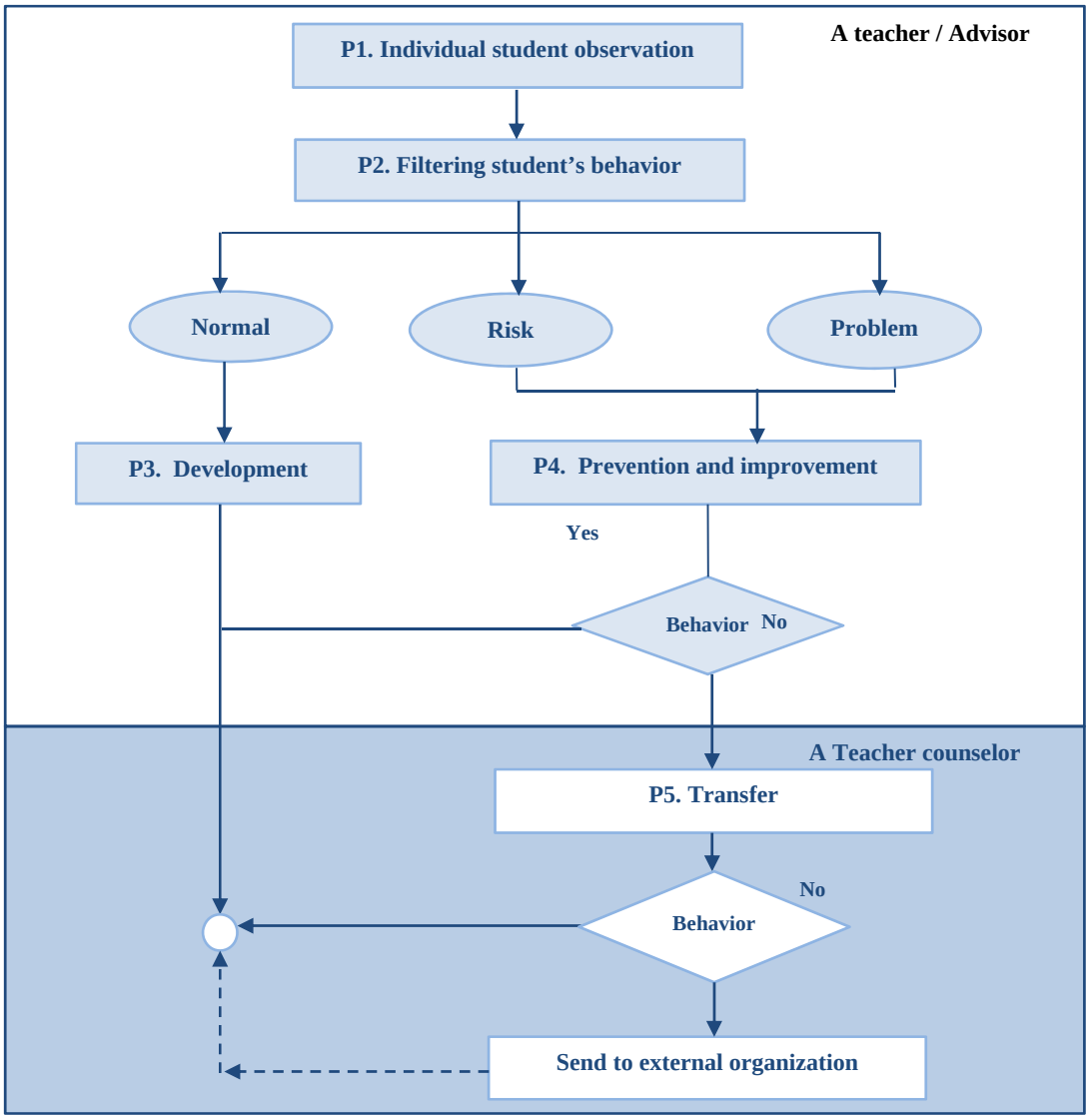


Figure 2.1 The process of Student Care and Support System.

The key main person for driving system is a teacher or an adviser. The system process is including five processes which provided the behavior details of a student by the duty. A process separated to two parts of controlling are a teacher or advisor and teacher counselor. The definition of all process with be described as follows.

P1. An individual student observation

Basically life of students is different. Therefore, their students have several behaviors give as: positive and negative behavior. Then, advisor should learn about student's behavior that it is the most important thing to understand the students and student development.

P2. Filtering student's behavior

This step, advisor classifies a student into three groups (Normal, Risk, Problem) by analyzing the student information of observation step. The student guideline is created by institute. The standard guidelines are represented in Fig 2.2.

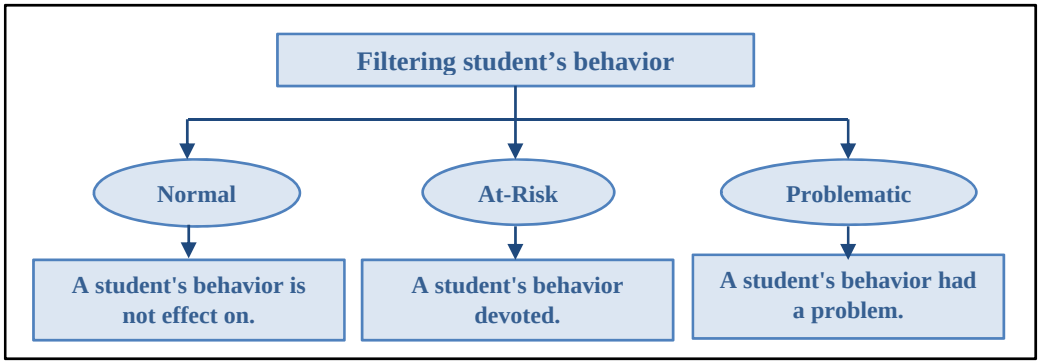


Figure 2.2 The characteristic of three behavior domain.

P3. Development

Development is to support the students receiving the quality to education and proud of them. The expected result of this process is to prevent the normal students becoming the at-risk and problematic students.

P4. Prevention and improvement

To prevent a student, advisor must takes care the students closely. For At-risk and problematic students, the advisor must observe their behaviors closely and considers the way to solve the problems.

P5. Transfer

The last step of system provides the worst case that advisor cannot help At-risk and problematic student. Advisor should transfer the student to the expert or the teacher counselor who will contact to internal or external organization for solving problems.

2.2 Data Mining Definitions

According to the Gartner Group, "Data mining is the process of determining the meaningful correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques" [6]. Other definitions of data mining are as follows:

Hand et al, “Data mining is the analysis of (often large) observing data sets to find unsuspected relationships and to summarize the data in novel ways that are understandable and useful to the data owner”.

Evangelos Simoudis in Cabena et al, “Data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issue of information extraction from large data bases”.

In conclusion, data mining is a process of analysis data sets from a large of database or the other sources to find the relationship between data and extracted the knowledge from the machine learning techniques.

2.3 Cross-Industry Standard Process: CRISP–DM

CRISP–DM given data mining project which has its life cycle consisting of six phases [7], including business understanding, data understanding, data preprocessing, modeling, evaluation, and deployment. The process is illustrated in Figure 2.3.

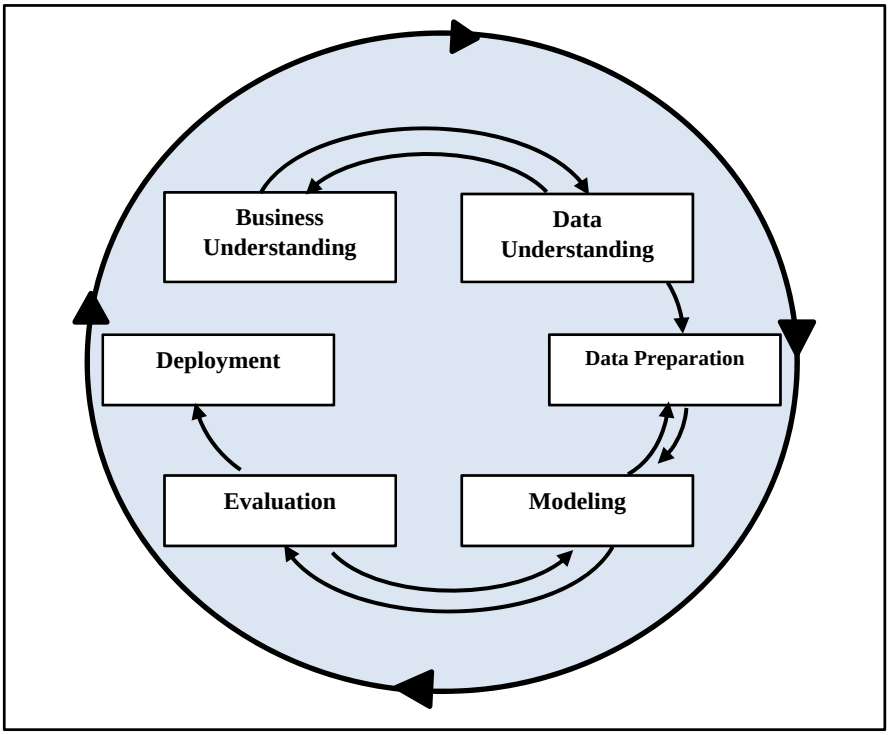


Figure 2.3 The CRISP-DM process.

In the following, the outline definition of life cycle phase can be given as:

- Business understanding phase is the initial phase, which focuses on understanding the project objectives and requirements from a business perspective [7].
- Data understanding phase is the collection data and proceeds with the activities related to preparation or identification to the factor, quality of the information to concern the problem of analysis information, and requirement.
- Data preparation is phase preparation data that covers all activities needed to construct the final dataset from the raw data. Tasks include table, record, and attribute selection, as well as transformation and cleaning of data for modeling tools [7].
- Modeling phase is to identify the techniques which are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data [7].
- Evaluation phase is the consideration activity used to compare between the accuracy and another term. The result provides the suitable tool or the methodology identification topic.
- Deployment phase is the creation of the model that is generally not the end of the project [7]. The knowledge is considered to extract the knowledge to user making decision.

2.4 Data mining Task

The main tasks, data mining is usually called upon to accomplish. The following lists shows the three common data mining tasks is Classification, Clustering and Association.

- **Classification** is to analyze the group of data from class or type of data set, such as the age of the patient and the patient's sodium/potassium ratio. Figure 2.4 is a scatter plot of patients' sodium/potassium ratio against patients' ages for the samples of 200 patients. The particular drug prescribed is symbolized by the shade of the points. Light gray points indicate drug Y; medium gray points indicate drug A or X [7].

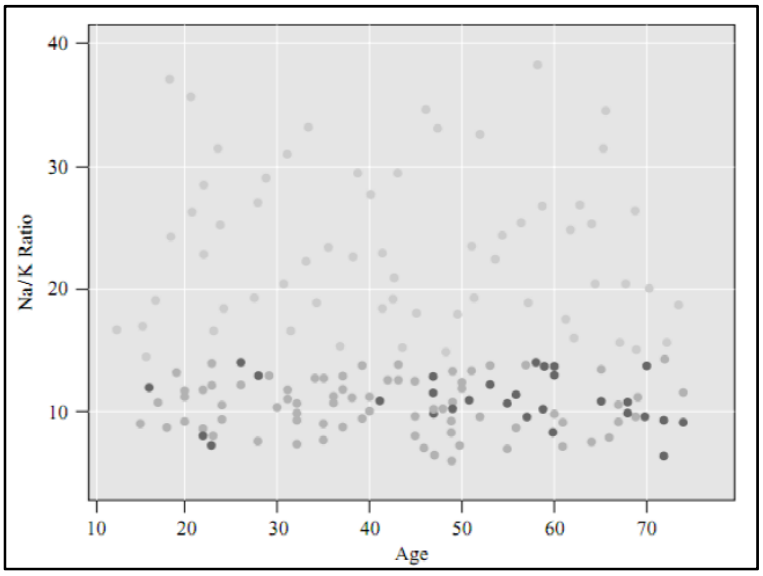


Figure 2.4 The type of drug a patient.

- **Clustering** refers to the grouping of records, observations, or cases into classes of similar objects. A cluster is a collection of records that are similar to one another, and are dissimilar to the record in other clusters [7]. This technique differs from classification in that there is no target variable for clustering.

- **Association** is the job of finding which attributes "go together". Most prevalent in the business world, known as affinity analysis or market basket analysis, the task of association seeks to uncover rules for quantifying the relationship between two more attributes [7].

2.5 The Decision Tree

This machine learning scheme has been developed from the simple division and conquer algorithm for producing decision trees [12]. Description performs in training set and generates the model depending on test set. The method of analyzing decision tree is numerical attributes, missing value, pruning and estimating error rate.

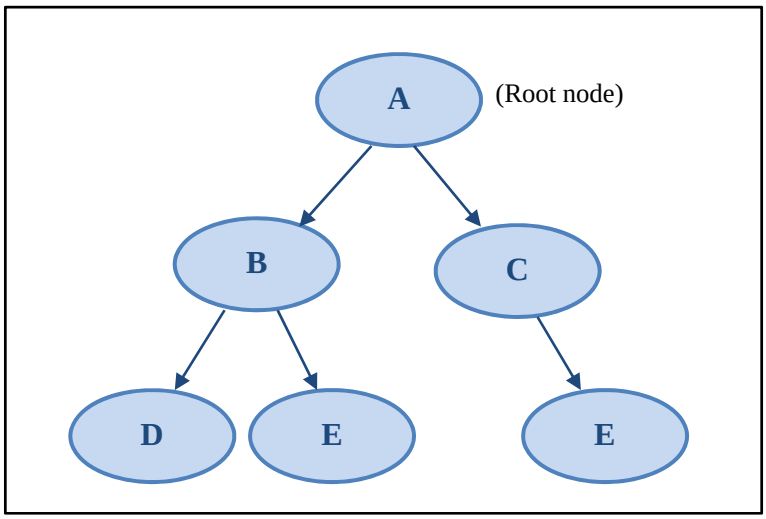


Figure 2.5 The decision tree model.

2.6 Related work

Baradwaj et al [7] described that the data mining in education field, called Educational Data Mining (EDM). The EDM emphasize methodology development to apply data mining technique which is increased popular. The definition method can be used to predict student performance and alienation of classroom teaching model. The determinate knowledge can be used for prediction regarding the student's risk performance. They used data mining classification method to predict student's performance at the end of semester. The three decision tree algorithms (ID3, ASSISTANT and C4.5) are used to generate the possible scenario decision. The algorithm provides the better result for prediction. It concerned the classification method to provide the way of extraction the risk behavior for study.

Ahmed et al [8] applied the ID3 decision tree method to predict the student's performance on the basic of student database. For risk behavior, they used top-level method (C4.5) and applied collection data method regarding the attribute that extracted knowledge to described student's risk behaviors.

Prasartwanakit et al [9] study the patterns of sexual behavior of Thai adolescents and youths in education institutions in Songkhla Province. The details of the survey provided the adolescent information that was useful for creating an attribute

of the data set including type-1 (healthy), type-3 (family and economical) and type-6 (sexual).

Approaching to this paper, the attributes of dataset of type-1 (healthy), type-3 (family and economical) and type-6 (sexual) are based on Prasartwanakit's work [9]. Moreover, Baradwaj's works [7,8] provides the classification technique for risk behavior pattern extraction. The classification model is generated by top-level algorithm (C4.5) of the decision tree.

CHAPTER III
RESEARCH METHODOLOGY

The classification models are developed by the process model improved from the data mining standard or CRISP-DM methodology [10] including business understanding, data understanding, modeling, evaluation and deployment, as shown in Figure.3.1.

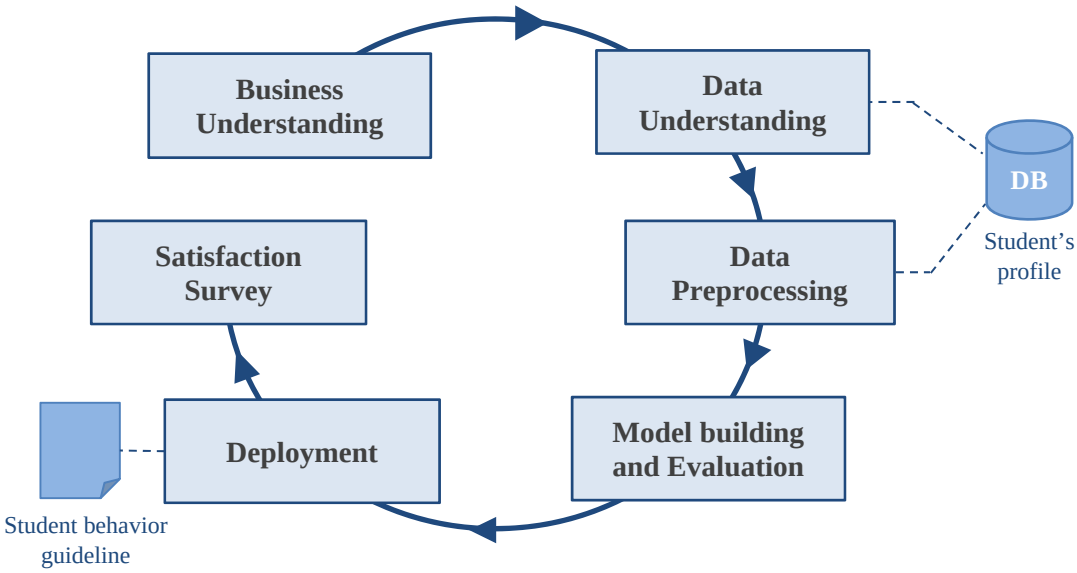


Figure 3.1 Research methodology.

3.1 Business Understanding

For the filtering process of the Student Care and Support System, the advisor observes students to classify student's behaviors into three groups (normal, At-Risk, problematic). In the process of phrase, the six consideration factors are ability, health, family and economy, drug, safety and sexuality.

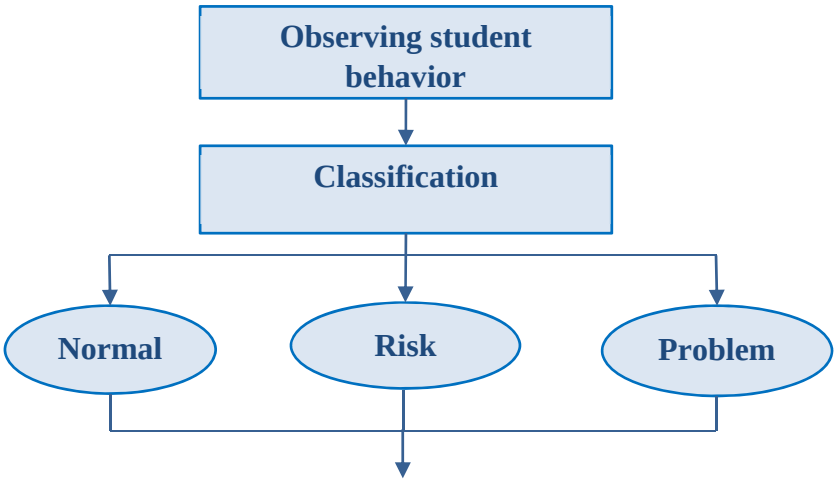


Figure 3.2 Related processes of Student Care and Support System.

3.2 Data Understanding

The developed classification model covered the data of student's profile database. The attributes of the entities (six types of student's behavior) represent in ER-Diagram (see in Fig 3.4). "Group" is the main attribute for all types, classify three types of student behaviors are normal, risk and problem.

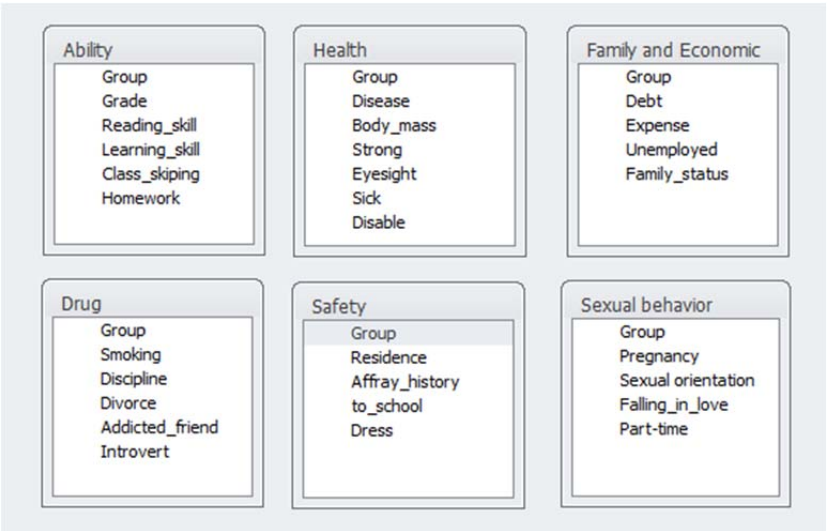


Figure 3.4 ER-Diagram of the risk student behavior.

The details of six entities, which specify the types of the behaviors, are presented as below. The attribute description is identified to the possible value.

Type-1 (Ability)

Table to Type-1 or Ability, as shown in Table 3.1, represents the attribute, type, description, and possible values. Type-one of student's behavior or ability extracted from student database. A set of determination factor is related to education ability and learning skill of student.

Table 3.1 Type-1(Ability).

No.	Attribute	Type	Description	Possible values
1	Group	String	A result of filtering student's behavior	{Normal , Risk , Problem}
2	Grade	String	Final grade of student	{A >= 3.00 , B >= 2.00 & <= 2.99, C < 2.00}
3	Absent	String	A number of the student absent from school, and late to study.	{always > 5 day/term , sometimes >=3 & <=5 day/term , never <=2}
4	Learn	Boolean	A level of student concentration in study	{Yes , No}
5	Assignment	String	Student assignment	{punctual , unpunctual}
6	Skip_class	Boolean	Student skip class.	{yes , No}
7	Reading_skill	Boolean	Reading and spelling skill	{yes , No}

Type -2 (Health)

Table to Type-2, as shown in Table 3.2 represents the attributes, type, description, and possible values. Type-two of student's behavior factors extracted from the student health profile and observation. The determination factors are provided the student's health prediction.

Table 3.2 Type-2(Health).

No.	Attribute	Type	Description	Possible values
1	Group	String	A result of filtering student's behavior	{Normal , Risk , Problem}
2	Strong	String	Body efficiency	{Strong , Weak}
3	Congenital_ disease	String	Congenital disease	{Yes , No}
4	Eeight_ height	Boolean	BMI (body mass language)	{Yes , No}
5	Eyesight	String	Eyesight and sense	{Yes , No}
6	Disabled	Boolean	Disabled	{Yes , No}
7	Sick	Boolean	Painful	{Yes , No}

Type-3 (Family and Economy)

Table to Type-3, as shown in Table 3.3 represents the attributes, type, description, and possible values. Type-three of student's behavior or ability extracted from the student database. The determination factors are related to the environment and the life of student family.

Table 3.3 Type-3(Family and Economy).

No.	Attribute	Type	Description	Possible values
1	Group	String	A result of filtering student's behavior.	{Normal , Risk , Problem}
2	Expense	String	Tuition fee	{Yes , No}
3	Unemployed	String	Parent is unemployment	{Yes , No}
4	Debt	Boolean	Parent is in dept.	{Yes , No}
5	Family_status	String	Family status	{together, divorce, dead}

Table 3.3 Type-3(Family and Economy). (cont.)

No.	Attribute	Type	Description	Possible values
6	Residence	Boolean	Student residence	{rel_house, apartment, house}
7	F_problem	Boolean	Student problem	{Yes , No}

Type- 4 (Drug)

Table to Type-4, as shown in Table 3.4, represents the attribute, type, description and possible values. Type-4 of student's behavior or ability extracted from student database and observed by survey. The determination factors are related to student smoke drug.

Table 3.4 Type-4(Drug).

No.	Attribute	Type	Description	Possible values
1	Group	String	A result of filtering student's behavior.	{Normal , Risk , Problem}
2	Smoke	String	Drug smoking	{Yes , No}
3	Friend_drug	String	Student's friend smokes drug.	{Yes , No}
4	Rent_devorce	Boolean	Parents divorced.	{Yes , No}
5	Beak_ discipline	String	Student break discipline	{Always >=4 ,Sometime >= 1 & <=3 ,Never = 0}
6	Introvert	Boolean	Depressing or seclusion	{Yes , No}

Type- 5 (Safety)

Table to Type-5, as shown in Table 3.5, represents the attribute, type, description, and possible values. Type-5 of student's behavior or ability extracted from the student database and observed by survey. A set of the determination is related to the safety factor determination.

Table 3.5 Type-5 (Safety).

No.	Attribute	Type	Description	Possible values
1	Group	String	A result of filtering student's behavior.	{Normal , Risk , Problem}
2	Residence	String	Student habitual	{rel_house , apartment , house}
3	Dress	String	Dressing	{dangerous , polite}
4	School_ motoring	Boolean	School motoring	{bus , motorcycle , walk}
5	Brawl_history	String	Student quarrel list	{Yes , No}

Type-6 (Sexual)

Table to Type-6 represents the attribute, type, description, and possible values. Type-6 of student's behavior or ability extracted from student database and observed by survey. The determination factors are related to risk behavior of adolescent (Sexuality).

Table 3.6 Type-6 (Sexuality).

No.	Attribute	Type	Description	Possible values
1	Group	String	A result of filtering student's behavior.	{Normal , Risk , Problem}
2	Pregnant	String	Pregnant	{Yes , No}

Table 3.6 Type-6 (Sexuality). (cont.)

No.	Attribute	Type	Description	Possible values
3	Homosexual	String	Sexually deviant , homosexual	{Yes , No}
4	Adulterous	Boolean	Romance	{Yes , No}
5	Live_ boyfriend	String	Living with boyfriend and girlfriend	{Yes , No}
6	Part_time	Boolean	Risk part time	{Yes , No}

3.3 Data Preprocessing

In this data mining process, the data set contains a specific domain which is selected from the plain text, the original data is in student's profile database. Collecting original data is regarded to fill in related attributes of data set, as shown in Figure 3.4. The missing values are filled by Min, Max and Average methodology. Each attribute value of each model of the dataset is heuristically mapped by their semantically.

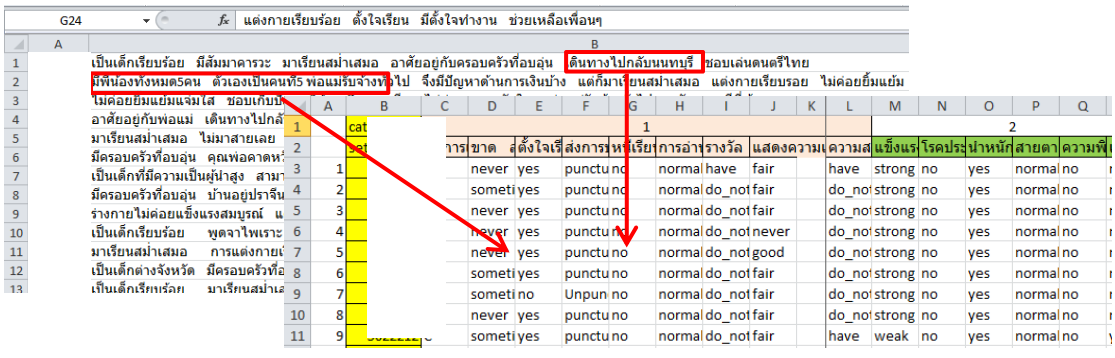


Figure 3.4 Filtering attributing from plain text.

3.4 Model Building and Evaluation

To generate the rules for the classification method, we use the decision tree algorithm (C4.5), as the learning algorithm to build the models and to lean a mapping from attribute values to the class [11]. A set of data is collected by the cross validation technique [12]. This technique repeats the sample of data set until a perfect score is representation. To avoiding a failure of predicting or overfitting by performed the available data set before automatically models generating. To define the best model, the cross validation technique compare to the hole-out technique [13] which the proportion of hold-out is 70:30, This method reserve amount of data for training set and test set.

3.5 Deployment

The six decision tree models are used to constructed the student's behavior guideline that provides characteristic of student behaviors for six types predicting information. The general guideline consists of characteristic and the improvement for solving problem.

3.6 Satisfaction Survey

The target group or the population of this research is an advisor at the college. A sample is selected by cluster random sampling methodology. The behavior guideline questionnaire is proved by specialist and is provided on web application. The statistic to estimate the result of satisfaction survey are the average (X) and the standard deviation (SD).

The details of questionnaire factors presented in the Appendix A which are separated to three parts, given as: advisor personality, guideline evaluation, and overall guideline evaluation.

3.8 Research Schedule

Table 3.7 Research Schedule.

Tasks	Time in moths, 2015																							
	Apr				May				Jun				Jul				Aug				Sep			
1. Business understanding and defining a problem.																								
2. Studying the approach theory and methodology.																								
3. Data collection and data preprocessing																								
4. Data mining tasks and model evaluation.																								
5. Building the classification models.																								
6. Creating student's behavior guideline.																								
7. Satisfaction survey.																								
8. Documentation.																								

CHAPTER IV
RESULTS

This chapter presented results of experiment related experiment result, classification model, the student risk behavior guideline and satisfaction survey.

4.1 Experiment Result

Discovery knowledge from C4.5 performs a multiple decision model. The quantitative results of each model are compared between cross validation technique and Hold-out technique in terms of classification accuracy, results are shown in Table 4.1.

Table 4.1 Experiment accuracy result.

Type	**Hold-out (70 : 30)**	**Cross-validation (10 folds)**
Ability	94.09%	90.82%
2 Health	91.86%	92.09%
3 Family and Economy	94.12%	92.72%
4 Drug	97.74%	96.52%
5 Safety	94.57%	94.62%
6 Sexuality	91.86%	92.72%

The overall accuracy of experiments is greater than 90%, the result is the important evidence for confirming, models generating without data overfitting. Moreover the overall accuracy results (greater than 90%) represent the decision tree based on sequence factors determination. The six initial models represent in the next section.

4.2 Classification Models

Generating six models used cross-validation (10-folds) technique, performing model through Data mining tool on computer. In Fig. 4.1, the illustration model provides the significant result, with the accuracy level of 90 percent or higher. The six models are illustrated in Figures 4.1 to 4.6.

Type-1(Ability)

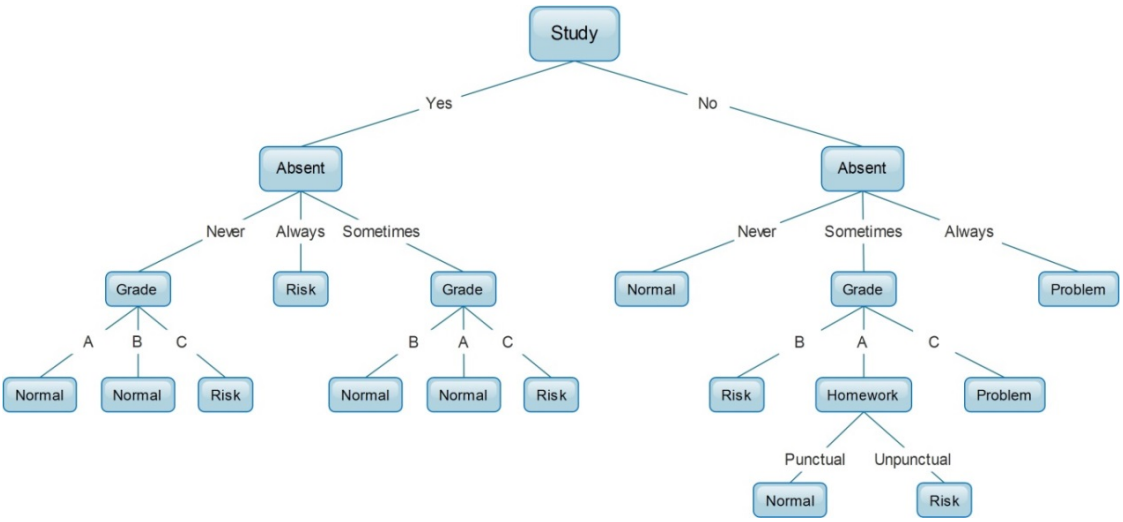


Figure 4.1 Type-1 models.

The generated model based on computing preformation, is to provide the significant result that defines the risk behavior of Type-1. The decision tree path is not reasonably because cannot specify the meaning of student behavior. For example, student sometimes absent, some grades are given as “A” and unpunctual homework, this character is “Risk” behavior group. However, overall of model indicates the characteristic student behavior for Type-1.

Type -2 (Health)

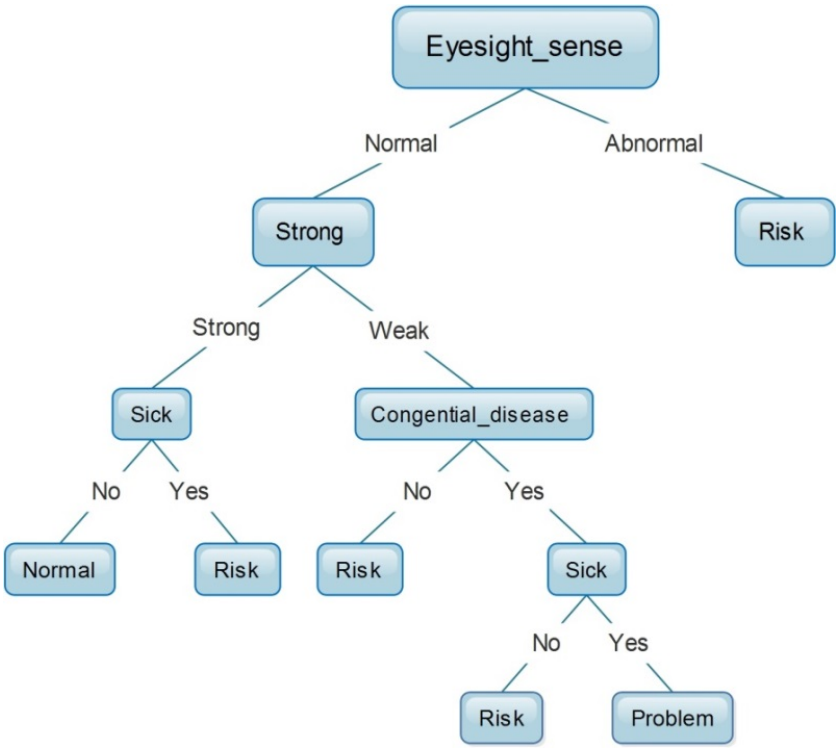


Figure 4.2 Type-2 models.

The Type-2 model is used to predict the health behavior, many relationships of some node on the decision tree provide the significant meaning, For example, eyesight of student is normal, the body is strong, but student get sick. Then we assign student to “At-Risk” group.

Type-3 (Family and Economy)

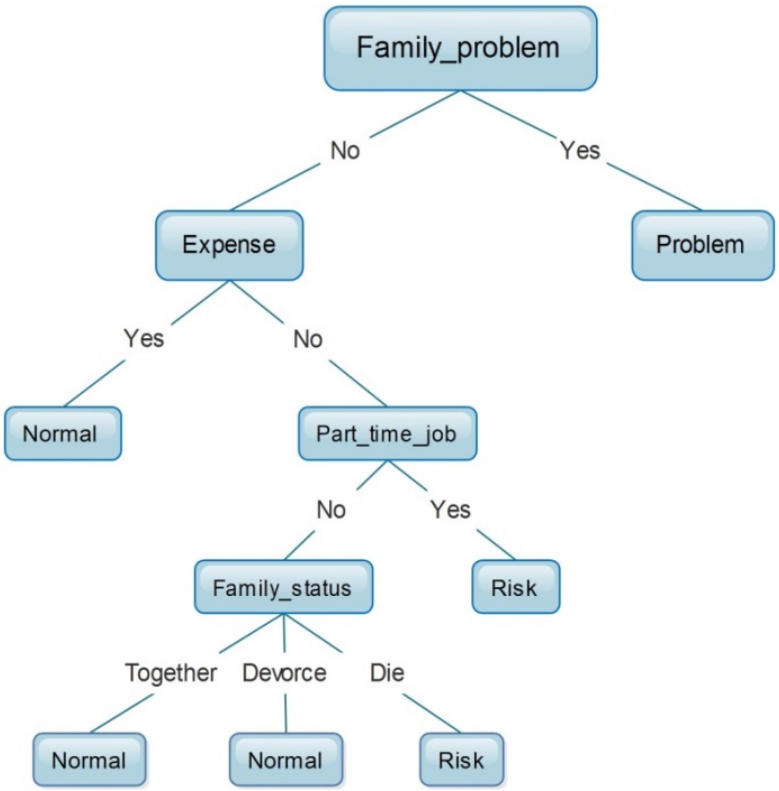


Figure 4.3 Type-3 models.

The Type-3 model is used to predict the family and Economy behavior. Many relationships of some nodes on the decision tree provide the significant meaning, For example, student family has no a problem, no expense, no part time job and related to parents. Then we assign student to “At-Risk” group.

Type- 4 (Drug)

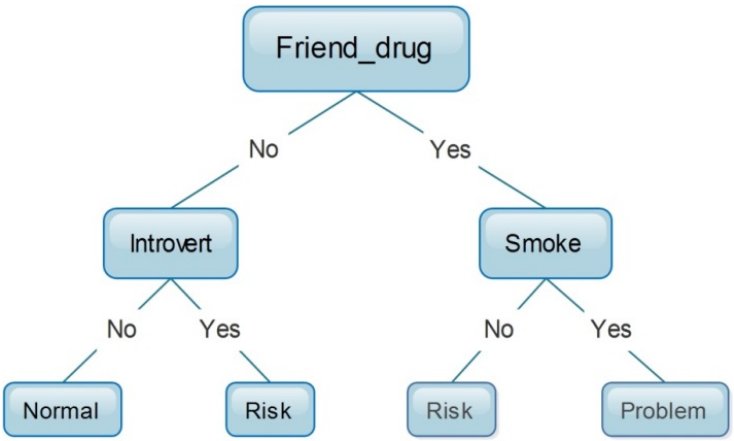


Figure 4.4 Type-4 Drug.

The Type-4 model is used to predict the drug behavior. Many relationships of some nodes on the decision tree provide the significant meaning. For example, student has the smoking's friend, but student does not smoke. Then we assign student to "At-Risk" group.

Type- 5 (Safety)

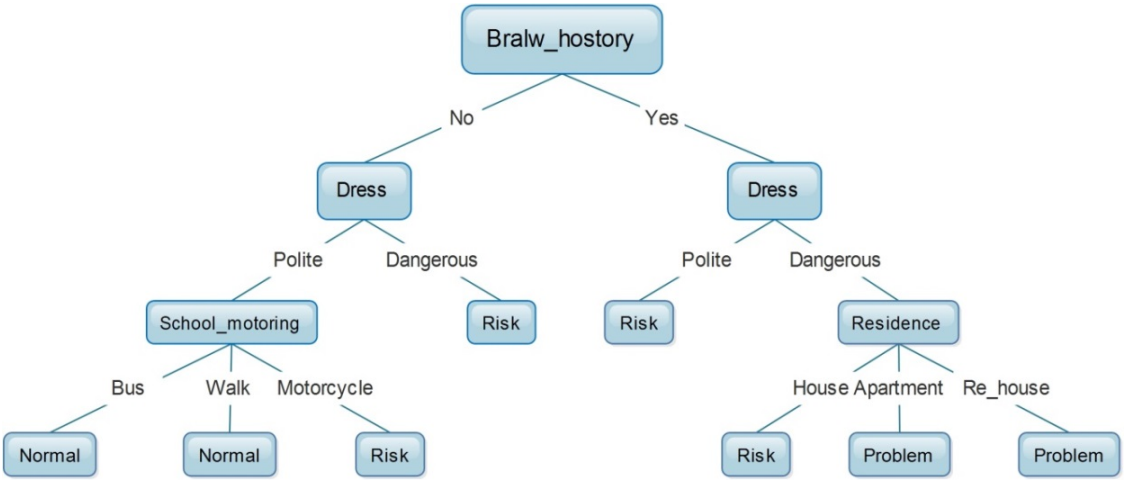


Figure 4.5 Type-5 safety.

The Type-5 model is used to predict the safety behavior. Many relationships of some node on the decision tree provide significant meaning. For example, student never brawl with the others, go to school by motorcycle and use politely dress. Then we assign student to “At-Risk” group.

Type-6 (Sexuality)

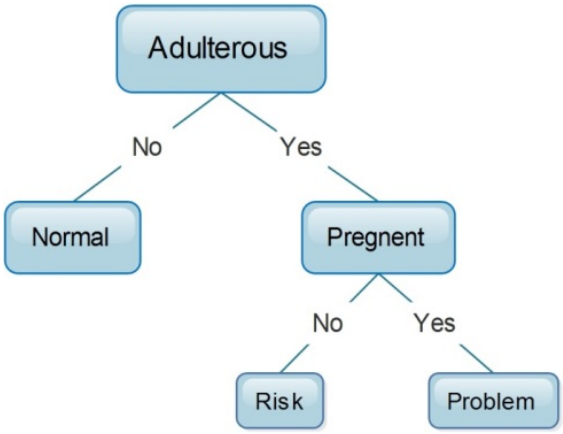


Figure 4.6 Type-6 Sexuality.

The Type-6 model is used to predict the Sexuality behavior. Many relationships of some nodes on the decision tree provide significant meaning. For example, student has boyfriend or girlfriend and there is pregnant. Then we assign student to “Problematic” group.

4.3 Student Risk Behavior Guideline

After the models generating, he risk behavior models would be paraphrased to make the guideline for observing and improving the student behavior. The guideline regarded following a process step by step.

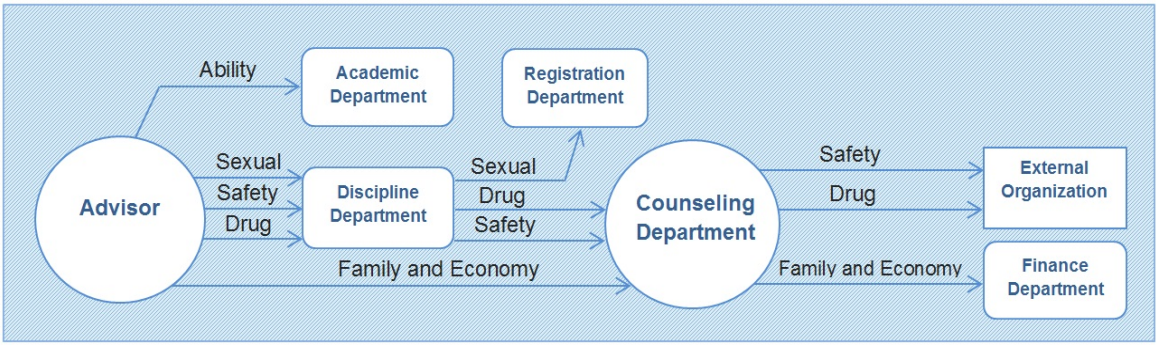


Figure 4.7 risk behavior process.

According to general guideline, as shown in Fig 4.7, the advisor separates At-Risk students to transfer to the related department for improving their behaviors. A list of risk behavior and improvement based on this work flows is represented in Table 4.2.

4.4 Student Risk Behavior Guideline Separated by Type

After the building models process, the student behavior guideline is generated by the information from the risk behavior model. The method for improving the student skill provides to each of student behavior group. The summary of student behavior guideline presented as shown in Tables 4.2 to Table 4.7.

Type-1 (Ability)

Table 4.2 Type-1 guideline.

Group	Normal	At-Risk	Problematic
Behavior	**Step 1.** A student always attends class, never absents from school, and GPA 2.00 - 2.99.	**Step1.** A student always attends class, never absents from school, and GPA is 2.00.	**Step1.** A student not concentrates on study, sometimes absents, and GPA is lower than 2.00.
	Step2. A student always attends class, never absents, and GPA is greater than 3.00. **Step3.** A student always attends class, sometime absents, and GPA 2.00 – 2.99. **Step4.** A student always attends class, and GPA is greater than 3.00.	**Step2.** A student always attends class, sometimes absents, and GPA is lower than 2.00. **Step3.** A student concentrates on study and sometimes absents from school. **Step4.** A student not concentrates on study, sometimes absents, and GPA 2.00 – 2.99.	**Step2.** A student not concentrates on study, and always absents from school.
Methodology	For special student (emphasized special ability), a teacher coordinates academic department to develop skill.	A teacher always evaluates student grade during the semester, and advices grade improvement.	A teacher transfers a student to Academic Department for improving study behavior.

The Student behavior guideline for Type-1 (Ability), retrieved from classification model, consists of behaviors workflow that advisor can identify student behavior by considering a step of behavior. For example, student always attends class, sometimes misses' class (3 times per semester). Moreover, GPA is greater than 3.00 (considering student to Normal group). The method provides the solvation the problem suggested by expert. Many factors in workflow obtain the insignificant meaning and insufficient details to identify the group of student behavior.

Type-2 (Health)

Table 4.3 Type-2 guideline.

Group	**Normal**	**At-Risk**	**Problematic**
Behavior	**Step1.** A student has normal eyesight, and health.	**Step1.** A student has normal eyesight but always has disease. **Step2.** A student has normal eyesight, body weakness, and has not a congenital disease. **Step3.** A student has normal eyesight, body weakness, and a congenital disease.	**Step1.** A student has normal eyesight, body weakness, a congenital disease and always has a disease.
Methodology	The institution should regard health checking for student less than one times per semester.	An advisor must emphasize the student care, For example, the advisor coordinates with a teacher to reduce the assignment for the student who has a heart disease.	

The identify factor of normal behavior group is short detail to describe or consider the behavior specified by data or node of model. However, workflow of risk group is good meaning for considering the student behavior.

Type-3 (Family and Economy)

Table 4.4 Type-3 guideline.

Group	Normal	At-Risk	Problematic
Behavior	**Step1.** A student is a good family, and has enough money for study. **Step2.** The student's parents live together. A student does not make a part time job.	**Step1.** A student is a good family, but student must make a part time job to support a fee.	**Step1.** A student's family has a problem.
Methodology		The advisor transfers student to the Counseling Department for finding a job, receiving a scholarship, and getting pocket money from school charity.	The advisor should visit to student's house for observing the situation, and then continuously contact to student's parents for planning and reducing the problem.

The factor to identify a risk and problem behavior group is specified by database and information extraction of classification technique. The details of Type-3 guideline may not enough to identify the student behavior group.

Type-4 (Drug)

Table 4.5 Type-4 guideline.

Group	Normal	At-Risk	Problematic
Behavior	**Step1:** A student has no friend taking drug. **Step2:** A student is not depressing.	**Step1:** A student has friend who smokes a drug, but student has never been a smoker. **Step2:** A student sometimes lives alone and introvert.	**Step1:** A student smokes a drug.
Methodology	The intuition should always provide the activity or training about drug addiction for all student groups.	The teacher checks smoking drug 4 times per semester to classify the student into three groups. Then, randomly repeat to check the smoking status of At-Risk and problematic group from all students. A teacher uses the result to consider the problem level.	Transferring to external drug institution (i.e. Thanyarak Institution and hospital).

The Type-4 factors insignificant to identify student behavior, the detail of each student behavior group of the guideline does not enough for consideration. Then the drug factor is sympathetic information. In addition, the determination factor of data set is limited by database and the survey.

Type-5 (Safety)

Table 4.6 Type-5 guideline.

Group	Normal	At-Risk	Problematic
Behavior	**Step1:** A student dresses politely, never fight another, and go to school by bus or walk.	**Step1:** A student dresses impolitely, never flight another, and go to school by drive a motorcycle. **Step2:** A student had ever flight another.	**Step1:** A student had ever flight another, dresses impolitely, and lives in the apartment or lives with the relative.
Methodology	The institute should provide the knowledge to student about the safety, and always intends safety campaign with the external organization.	For the brawl situation, a teacher should contact to the Discipline Department for mediation, and notify student's family. To controls the worst situation, the Discipline Department cooperates to a police to assuage.	

The factor of Type-5 guideline (risk group) provides decision making but details of normal and At-Risk are specified which is not enough to consider or identify student to these group, many factors are applied by observation activity or behavior of student.

Type-6 (Sexualityity)

Table 4.7 Type-6 guideline.

Group	Normal	At-Risk	Problematic
Behavior	**Step1:** A student no couple, and no Sexualityity.	**Step1:** A student has a couple, and Sexualityity. **Step2:** A student has a couple, and not pregnant.	Step1: A student has a couple, Sexualityity and pregnant.
Methodology	The institution should campaign about Sexualityity.	The advisor must advices the suitable behavior when student lives with a couple and student pregnant. An advisor should coordinate to student's parent for decision making.	The advisor must contact to the Discipline Department, and student's parents. For the pregnancy case, the advisor should contact to the Registration Department for study plan.

The factor of Type-6 guideline can provide decision for making the Sexuality behavior of general student. The detail is not enough to identify the special student behavior (bi-Sexuality), which this factor affects for students.

4.5 Satisfaction Survey

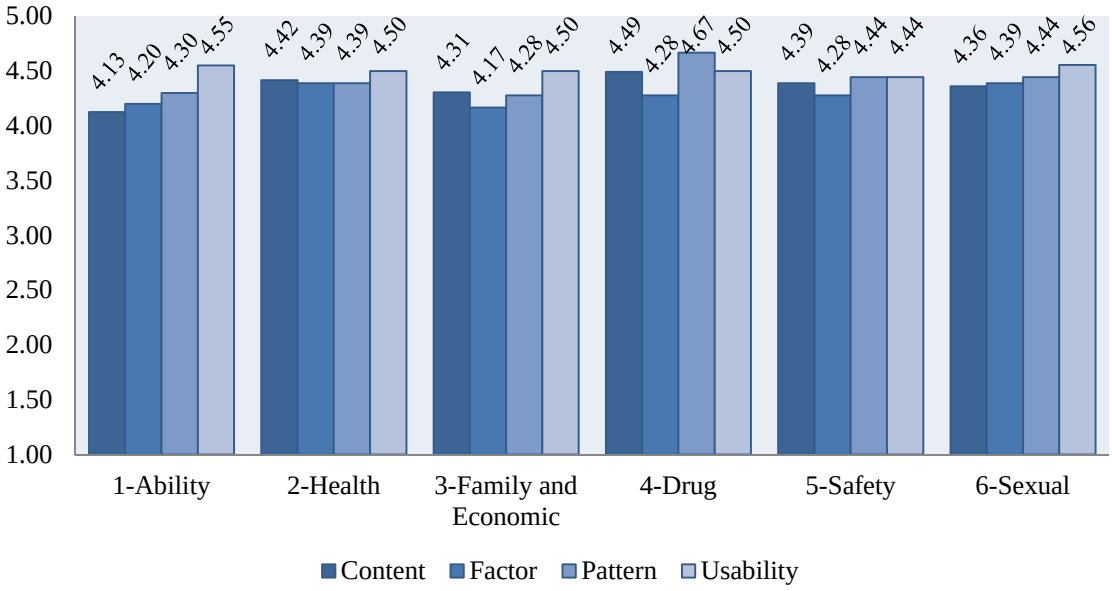


Figure 4.8 The result summary of student risk behavior.

In Fig 4.8, the bar chart indicates a satisfaction survey on six groups of student risk behaviors on the factors contributing to concern about the advisor satisfaction for the student behavior guideline. These factors are divided into four factors, including, content, risk behavior factor, pattern and usability. Overall, the result of all factor above 4.00 are agreed the details of the guideline are essential to provide information observation information of the student behavior. In conclusion, the 4-Drug has the highest significant response result of all factors. It is noticeable that the advisor concerned in the advantage of the guideline.

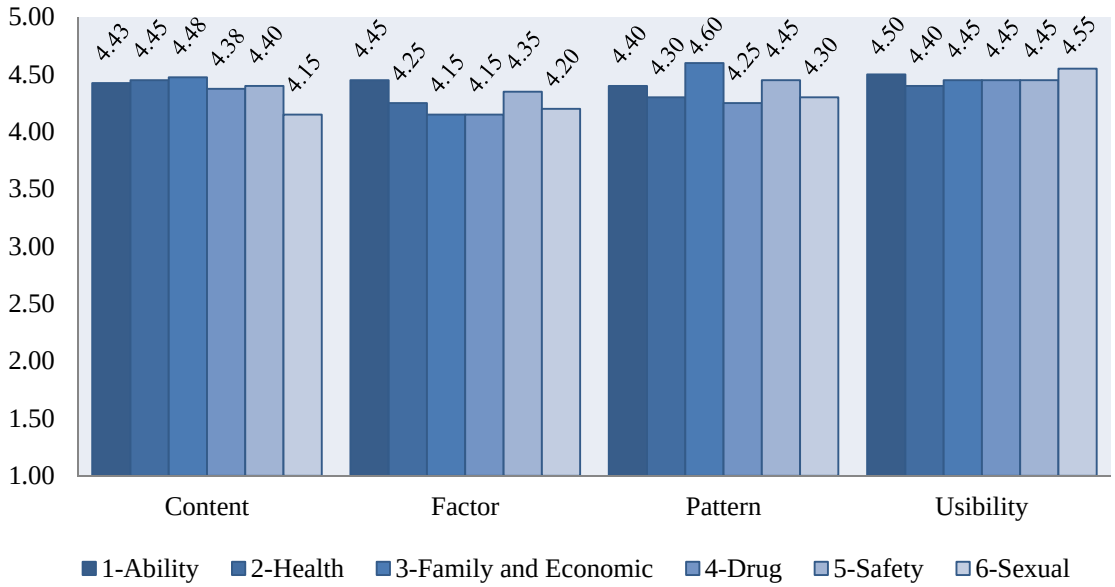


Figure 4.9 Summary result of the factor determination.

From Fig 4.9, the bar chart presents the summary result of the risk behavior factor determination to estimate the student behavior guideline. The summary of the evaluation term is the information of pattern presentation. Moreover, the highest number is Type 3-Family and Economy (4.60) that provides apparently the information of risk behavior characteristic for the advisor. Furthermore, the statistic result is confirmed by the expert in terms of the information based on the knowledge.

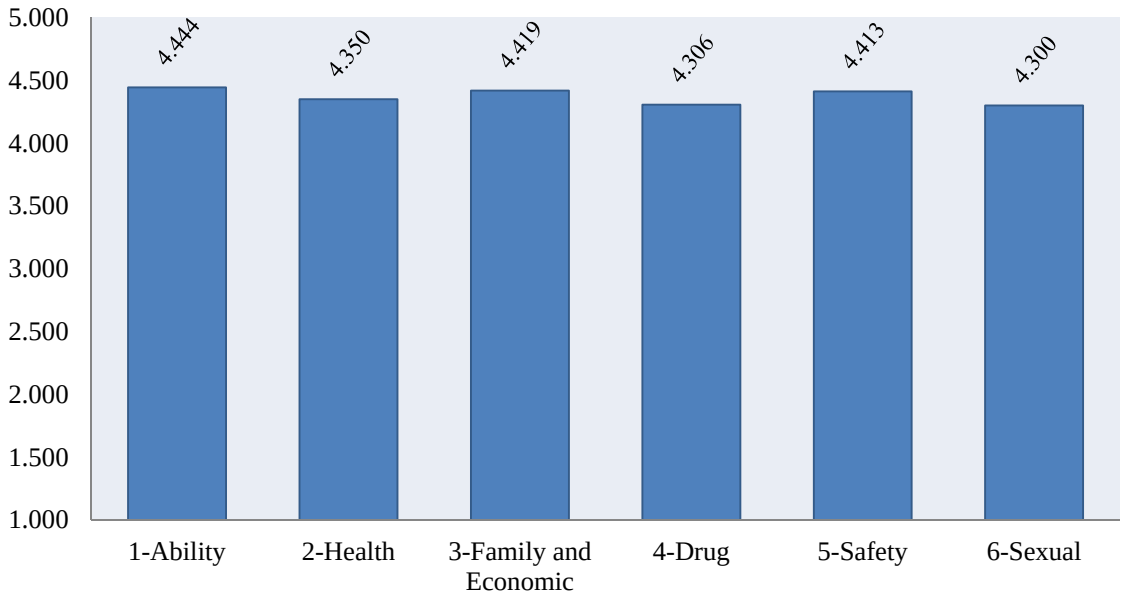


Figure 4.10 The overall summary result of the student risk behavior.

In Fig 4.10, the overall summary result of the factor determination of type 1-Ability is the highest. The information apparently descripted the student risk behavior that aligns with the factor determination including: obvious ability content, factor complement, apparent pattern and usability for observation student behavior.

CHAPTER V
CONCLUSION

This study represents that the decision tree is applied to construct an adolescent student risk behavior guideline and six classification models based on information of student database which is beneficial for the prediction of the student's behaviors. The risk levels can be classified into 3 groups including: Normal, At-Risk, and Problem. The accuracy of all classification models is greater than 90%.

Challenge of the student behavior guideline is influenced by six student behaviors type factors that defines as the concerning problem of adolescent student. The student's behavior guideline would facilitate the advisor for decision making in the Student Care and Support System.

However, the guideline provides the student behavior knowledge to the student girl in high school level of adolescent student that cannot apply to consider student behavior of a children and student boy. In addition, some factors or information insufficient for observation risk behavior and the information not defined meaning apparently.

For the future work, the risk behavior guideline develops to provide the advantage for a children and the information covers the other group of student with will be applied to the knowledge to development the application.

REFERENCES

1 Chatchai N. (2010), The statistic of adolescence pregnancy. Thai Health Promotion Foundation website:http://www.thaihealth.or.th/Content/25667.html.

2 No writer (2013), Quality of life :Drug statistic of adolescent, ASTV Manager.

3 Naruemon U. (2014), Stress Reduction. Nation News Bureau Of Thailand Website.

4 Bureau of Academic Affairs and Educational Standards Thailand (2006), Student care and help system guideline, Bangkok Transportation Organization of Thailand, pp.20-30.

5 Qasem A. Al-Radaideh, Emad M. Al-Shawakfa and Mustafa I. Al-Najjar (2006), Mining Student Data using Decision trees. International Arab Conference on Information Technology, Yarmouk University, Jordan.

6 Ian H. W. and Eibe F. (2006), Data Mining : Practical Machine Learning Tools and Techniques (Second Edition), Hoboken, New Jersey, John Wiley & Sons.

7 Brijesh K. B. and Saurabh P. (2011), Mining Educational Data to Analyze Students' Performance, International Journal of Advanced Computer Science and Applications (IJACSA), Singhaniya University and VBS Purvanchal University, India.

8 Abeer Eddr EL Din A. and I brahim Sayed E. (2014), Data mining : A Prediction for Student's Performance Using Classification Method, Cairo university, Egypt.

9 Prasartwanakit, Songwathana and Phetcharat (2010), Sexual beliefs and patterns among Thai adolescents and youths in educational institutions in Songkhla Province, Prince of Songkla University, Thailand.

10 Alex A. F. (2009), Data Mining and Knowledge Discovery with Evolutionary Algorithms, Computing Laboratory, University of Kent, UK.

11 Xindong W. and Vipin K. (2009), The Top Ten Algorithms in Data Mining, Taylor & Francis Group, pp.3.

12 Daniel T. Larose,Chantal D. Larose (2015), Data Mining and Predictive Analytics second edition, Canada : John Wiley & Sons, Inc., Hoboken, New Jersey, pp.161.

13 K. P. S. ,Shyam D. and V. A. (2006), Data Mining : Theory and Practice, Prentice of India PrivateLimited, New Delhi, pp.104.

APPENDICES

APPENDIX A

The Satisfaction Survey of the Risk Student Behavior

แนวทางปฏิบัติสำหรับคัดกรองพฤติกรรมนักเรียน
The Risk Behaviors Guideline for Adolescent Student

การประเมินแนวทางปฏิบัติสำหรับคัดกรองพฤติกรรมนักเรียน

คำชี้แจง
แบบสอบถามนี้เป็นส่วนหนึ่งของการค้นคว้าอิสระ เรื่อง การสร้างแนวทางปฏิบัติสำหรับพฤติกรรมความเสี่ยงสำหรับนักเรียนวัยรุ่นโดยใช้ต้นไม้ตัดสินใจ สาขาวิทยาศาสตรมหาบัณฑิต (การจัดการเทคโนโลยีสารสนเทศ) มหาวิทยาลัยมหิดล เพื่อนำข้อมูลไปพัฒนาและปรับปรุงการจัดทำแนวทางปฏิบัติสำหรับคัดกรองพฤติกรรมนักเรียนให้บรรลุตามวัตถุประสงค์ของระบบดูแลช่วยเหลือนักเรียนต่อไป
แบบสอบถามนี้ประกอบไปด้วย 3 ส่วน ได้แก่
ส่วนที่ 1 ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม
ส่วนที่ 2 การประเมินแนวทางปฏิบัติสำหรับคัดกรองพฤติกรรมนักเรียน 6 ด้าน ได้แก่ ด้านความสามารถ, ด้านสุขภาพกาย, ด้านครอบครัว เศรษฐกิจ, ด้านยาเสพติด และด้านพฤติกรรมทางเพศ
ส่วนที่ 3 การประเมินภาพรวมแนวทางปฏิบัติสำหรับคัดกรองพฤติกรรมนักเรียน

*จำเป็น

ส่วนที่ 1 ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม
โปรดเลือกคำตอบที่ตรงกับความข้อมูลหรือความคิดเห็นของท่านมากที่สุด

เพศ *
- ชาย
- หญิง

ระยะเวลาการการปฏิบัติงาน *
- น้อยกว่า 1 ปี
- 1 – 5 ปี
- 6 – 10 ปี
- 11 – 15 ปี
- 16 – 20 ปี
- มากกว่า 21 ปี

ความสามารถ					
สุขภาพกาย					
ครอบครัวและเศรษฐกิจ					
ยาเสพติด					
ความปลอดภัย					
พฤติกรรมทางเพศ					

ต่อไป »

สำเร็จแล้ว 12%

ขับเคลื่อนโดย Google Forms

เนื้อหานี้มิได้ถูกสร้างขึ้นหรือรับรองโดย Google
รายงานการละเมิด - ข้อกำหนดในการให้บริการ - ข้อกำหนดเพิ่มเติม

แนวทางปฏิบัติสำหรับคัดกรองพฤติกรรมนักเรียน
The Risk Behaviors Guideline for Adolescent Student

การประเมินแนวทางปฏิบัติสำหรับคัดกรองพฤติกรรมนักเรียน

*จำเป็น

การประเมินพฤติกรรมด้านที่ 1 ความสามารถ

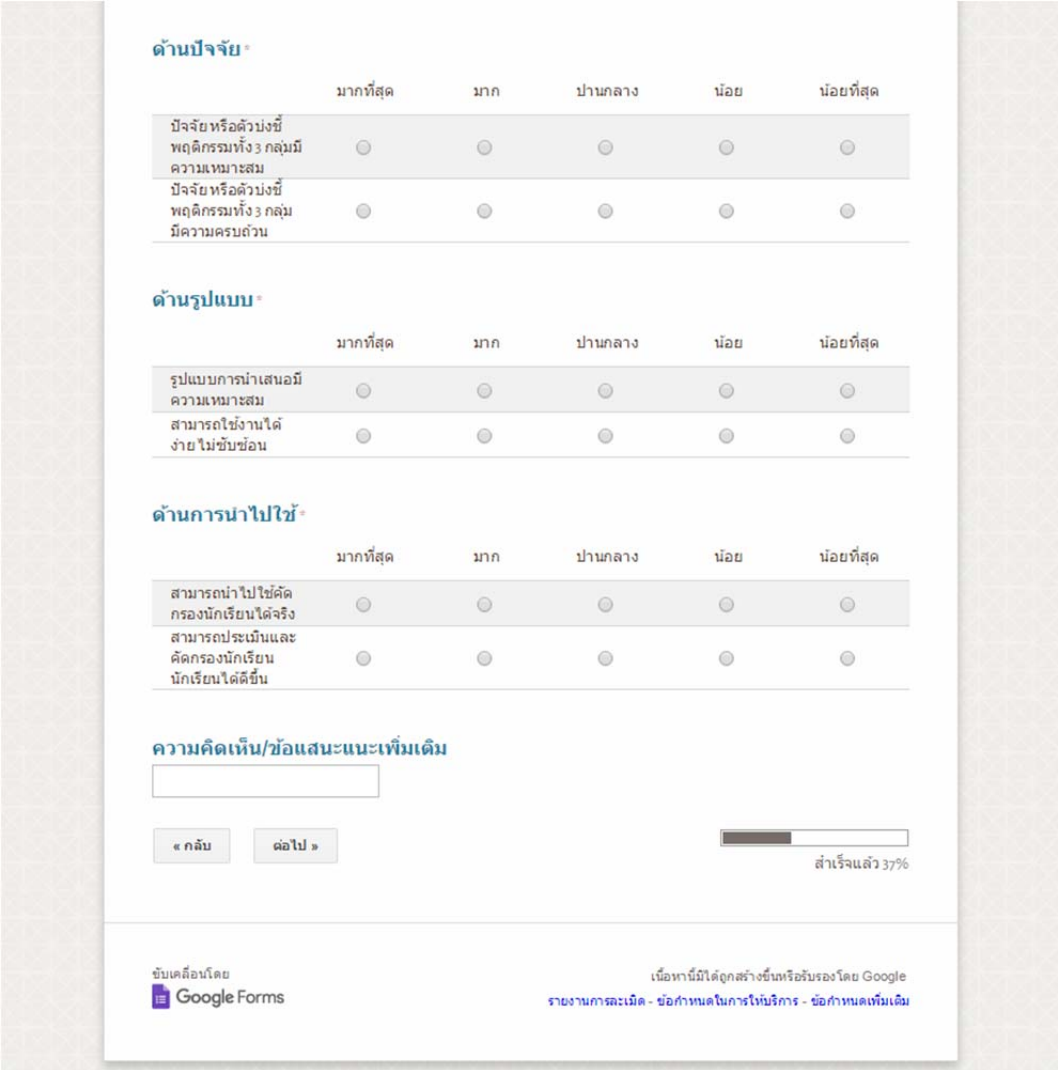
การใช้คู่มือ แนวทางปฏิบัติสำหรับคัดกรองนักเรียน

- ครูที่ปรึกษานำผลการสังเกตพฤติกรรมนักเรียนมาเปรียบเทียบกับลักษณะพฤติกรรมในแต่ละกลุ่ม (กลุ่มปกติ, เสี่ยง, มีปัญหา) เพื่อจำแนกกลุ่มนักเรียน
- พฤติกรรมแต่ละกลุ่มประกอบไปหลักเกณฑ์การพิจารณาหลายแบบ (step) หากพบว่าเข้าเกณฑ์ข้อใดข้อหนึ่งให้ถือว่านักเรียนจัดอยู่ในพฤติกรรมกลุ่มนั้น
- ตัวอย่าง นร. มีตั้งใจเรียน เข้าเรียนทุกครั้ง เกรดเฉลี่ย 3.10 แสดงว่านักเรียนจัดอยู่ในกลุ่ม "ปกติ" เป็นต้น

อนึ่ง เกณฑ์การพิจารณาที่ปรากฏใช้วิธีการทางคอมพิวเตอร์ในการจำแนกข้อมูล หากข้อมูลไม่ถูกต้องประการใด ผู้วิจัยขออภัยมานะที่นี้ด้วย และหวังเป็นอย่างยิ่งว่าผู้ตอบแบบสอบถามจะให้ความกรุณาให้ข้อเสนอแนะเพื่อการปรับปรุงแก้ไข

ด้านเนื้อหา*

	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
เนื้อหามีความถูกต้องเที่ยงตรง	○	○	○	○	○
เนื้อหามีความสมบูรณ์ ครบถ้วน	○	○	○	○	○
เนื้อหามีความเหมาะสม	○	○	○	○	○
เนื้อหาสอดคล้องตามเกณฑ์การประเมินพฤติกรรมนักเรียน	○	○	○	○	○



ด้านปัจจัย *

	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
ปัจจัยหรือตัวบ่งชี้พฤติกรรมทั้ง 3 กลุ่มมีความเหมาะสม	○	○	○	○	○
ปัจจัยหรือตัวบ่งชี้พฤติกรรมทั้ง 3 กลุ่มมีความครบถ้วน	○	○	○	○	○

ด้านรูปแบบ *

	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
รูปแบบการนำเสนอมีความเหมาะสม	○	○	○	○	○
สามารถใช้งานได้ง่ายไม่ซับซ้อน	○	○	○	○	○

ด้านการนำไปใช้ *

	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
สามารถนำไปใช้คัดกรองนักเรียนได้จริง	○	○	○	○	○
สามารถประเมินและคัดกรองนักเรียนนักเรียนได้ดีขึ้น	○	○	○	○	○

ความคิดเห็น/ข้อแสนะแนะเพิ่มเติม

« กลับ ต่อไป »

สำเร็จแล้ว 37%

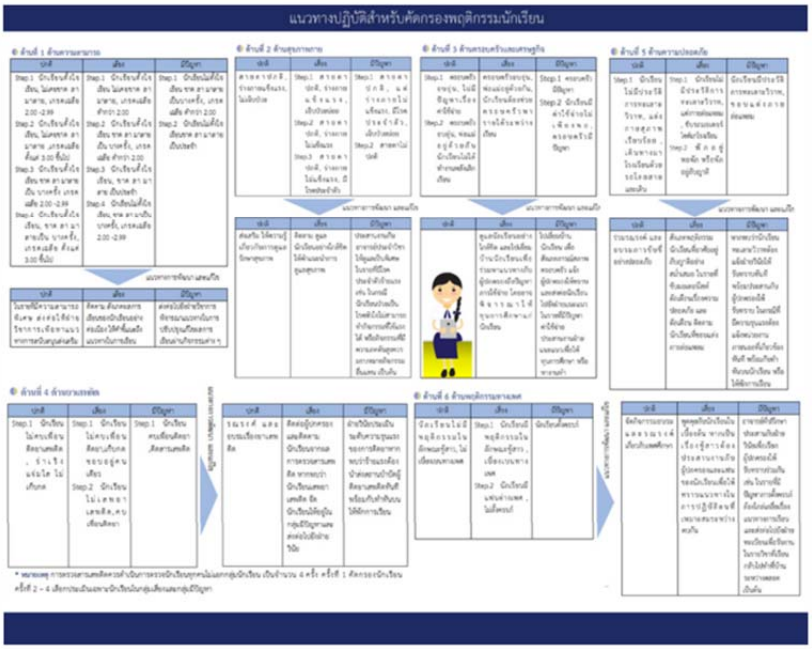
ขับเคลื่อนโดย Google Forms

เนื้อหานี้มิได้ถูกสร้างขึ้นหรือรับรองโดย Google
รายงานการละเมิด - ข้อกำหนดในการให้บริการ - ข้อกำหนดเพิ่มเติม

การประเมินแนวทางปฏิบัติสำหรับคัดกรองพฤติกรรมนักเรียน

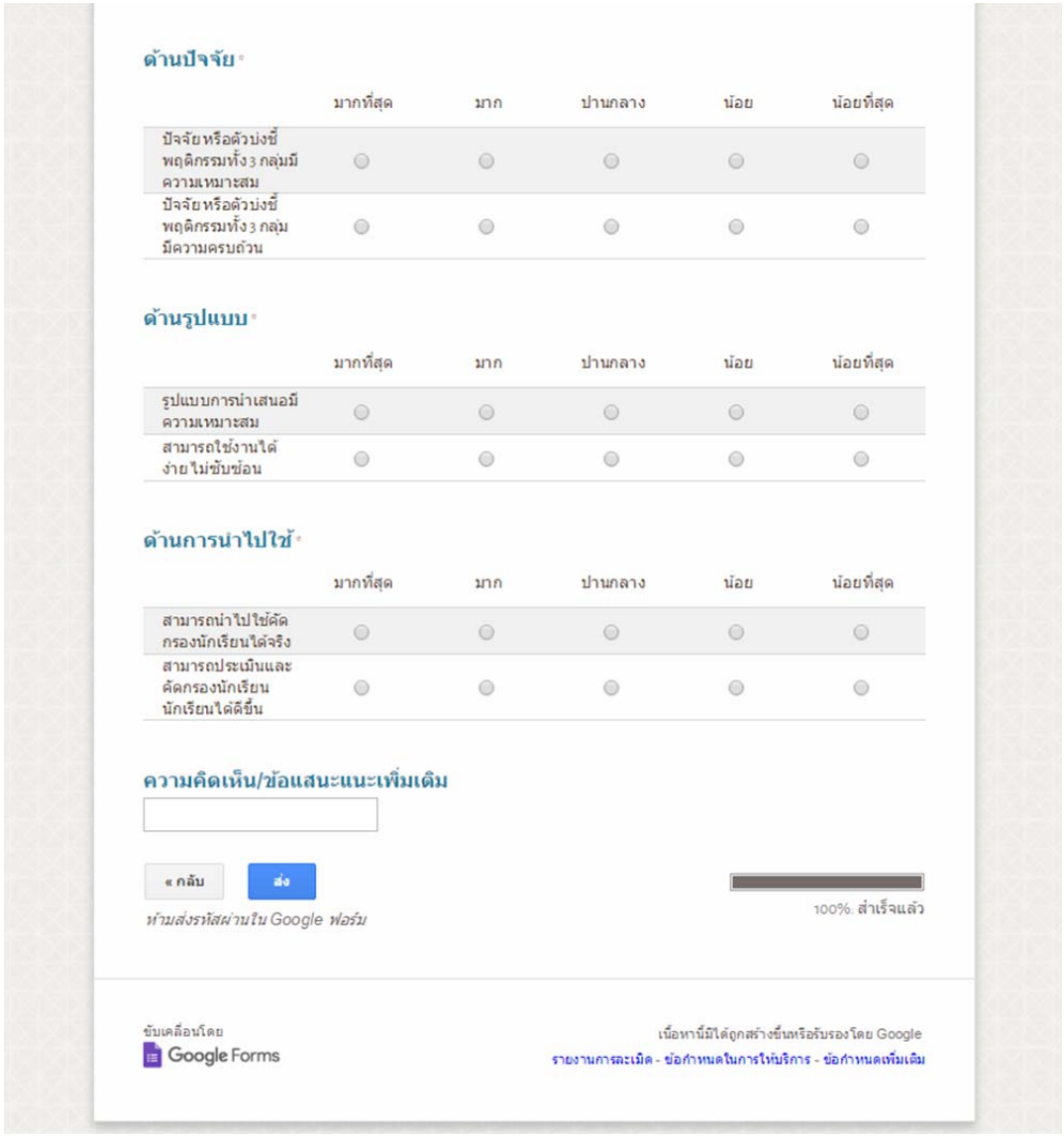
*จำเป็น

ส่วนที่ 3 การประเมินภาพรวมแนวทางปฏิบัติสำหรับคัดกรองพฤติกรรมนักเรียน



ด้านเนื้อหา*

	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
เนื้อหามีความถูกต้องเที่ยงตรง	○	○	○	○	○
เนื้อหามีความสมบูรณ์ ครบถ้วน	○	○	○	○	○
เนื้อหามีความเหมาะสม	○	○	○	○	○
เนื้อหาสอดคล้องตามเกณฑ์การประเมินพฤติกรรมนักเรียน	○	○	○	○	○



ด้านปัจจัย *

	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
ปัจจัยหรือตัวบ่งชี้พฤติกรรมทั้ง 3 กลุ่มมีความเหมาะสม	○	○	○	○	○
ปัจจัยหรือตัวบ่งชี้พฤติกรรมทั้ง 3 กลุ่มมีความครบถ้วน	○	○	○	○	○

ด้านรูปแบบ *

	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
รูปแบบการนำเสนอมีความเหมาะสม	○	○	○	○	○
สามารถใช้งานได้ง่ายไม่ซับซ้อน	○	○	○	○	○

ด้านการนำไปใช้ *

	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
สามารถนำไปใช้คัดกรองนักเรียนได้จริง	○	○	○	○	○
สามารถประเมินและคัดกรองนักเรียนนักเรียนได้ดีขึ้น	○	○	○	○	○

ความคิดเห็น/ข้อแสนะแนะเพิ่มเติม

« กลับ ส่ง

ห้ามส่งรหัสผ่านใน Google ฟอร์ม

100%: สำเร็จแล้ว

ขับเคลื่อนโดย Google Forms

เนื้อหานี้มิได้ถูกสร้างขึ้นหรือรับรองโดย Google

รายงานการละเมิด - ข้อกำหนดในการให้บริการ - ข้อกำหนดเพิ่มเติม

APPENDIX B

The 2nd Management Innovation Technology International Conference (MITiCON2015)

Constructing a Risk Behavior Guideline for Adolescent Students using Decision Tree

Surin Aunsan
Technology of Information System Management Division
Faculty of Engineering, Mahidol University,
Nakornpathom, Thailand
it_surina@hotmail.com

Sotarat Thammaboosadee
Technology of Information System Management Division
Faculty of Engineering, Mahidol University,
Nakornpathom, Thailand
sotarat.tha@mahidol.ac.th

***Abstract*—The main objective of education is to offer the best quality of life to student. One way to achieve the best quality of life is reduce and improve adolescent problem by discovering knowledge for providing the problem factors to institution. In this research, the knowledge is extractable through student database. The decision tree data mining technique is used to build six adolescent students behavior models and provide an overall accuracy over 90%. Six classification models are constructed based on risk behavior groups and are then used create student behavior guideline to support teacher decision making.**

Keywords—Decision Tree; Risk Behavior; Adolescent.

I. INTRODUCTION

Nowadays, the terrible situations of Thai youth represent demoralization in society. A survey of childbirth in private hospital, the United Nations Population fund (UN) and the office of the national economic and social development board found that 130,000 of the teenager mothers (15-19 years old) delivered a child and presents that 15,000 of teen mothers late delivered a child [1]. The interesting point of view on this research is that a teenager mother lately delivered a child per hour or every two days or nineteen moms late delivered a child. The results of research represent that teenager's pregnancy is a continuous problem [2]. Thanyarak Institute in 2013 surveyed 7,960 patients (86.11 % of male and 13.89 % of female) found that 18.12 % or 1,422 of patients are 15 through 19 years old [3]. Based on surveys from Department of Mental Health in 2014, the teenager's violence rate is 879 per year [3]. The result of those researches showed that teenagers lack of control their emotional state. The problem impacts on society, therefore, all parts of society must find the way solving problem. However, the governance must seriously drive the educational policy of the School children counselling of the teacher in school. To be more clarified, the Office of the Basic Education Commission of Thailand [4] define six of information groups including student abilities, health, family and economy, drug, safety, and sexual to provide classification for a group of student which are a step in student care and support health system. Hence, it could be beneficial if there is some of systematic system to identify and cclassify the level of each risk behaviors from general student information.

Applying technology to analyze and define factors of problem is one of the ways for supporting the system. The Data mining is the extraction of implicit, previously unknown, and potentially useful information from data [5]. The most useful data mining technique is to discover the root or the factors of problem which related to student's behaviors is classification. The Decision Tree, one of classification method, is a predictive data mining technique, makes prediction about values of data using known results found from different data [6].

This paper provides a construction of risk behavior classification models using the decision tree, and student behavior guideline. These models support data analysis on the student's profile online to provide the student counselling. In addition, adviser can use this guideline to improve behavior level of student.

The remainder of the paper is organized as follows: the next section discuss relevant in the education of domain related past works in the education domain. Section 3 introduces the methodologies and experimental result. Finally, we conclude the paper and suggest directions for future works.

II. RELATED WORKS

Baradwaj and Pal [7] described the data mining in education field, called Educational Data Mining, concern with developing methods that apply data mining technique that this field is increased popular. The discovery method can be used to predict student performance and alienation of classroom teaching method model. The discovered knowledge can be used for prediction regarded student's risk performance. They used data mining classification method to predict student's performance at the end of semester. The three decision tree algorithms (ID3, ASSISTANT and C4.5) are used to generate possible scenario decision. The algorithm provides better result to prediction. It concerned classification method to provide the way of extraction the risk behavior for study.

Ahmed and Elaraby [8] applied the ID3 decision tree method to predict the student's performance on the basic of student database. For risk behavior, they used top-level method (C4.5) and applied collection data method to regard the attribute that extracted knowledge to described student's risk behaviors.

Prasartwanakit, Songwathana, and Phetcharat [9] study the patterns of sexual behavior of Thai adolescents and youths in education institutions in Songkhla Province. The details of the survey are provided adolescent information that useful for creating an attribute of the data set including type-1 (healthy), type-3 (family and economical) and type-6 (sexual).

Approaching to this paper, the attribute of dataset of type-1 (healthy), type-3 (family and economical) and type-6 (sexual) is based on Prasartwanakit's work [9]. Baradwaj's works [7,8] concern extraction methodology for data mining, the classification technique is used to extract risk behavior, and classification models generate by top-level algorithm (C4.5) of the decision tree.

III. RESEARCH METHODOLOGY AND EXPERIMENTAL RESULTS

The classification models developed by the process model that extended from the data mining standard or CRISP-DM methodology [10] including business understanding, data understanding, modeling, evaluation and deployment, as shown in Fig.1.

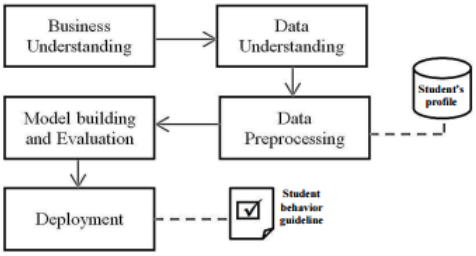


Fig.1. Research methodology process.

A. Business Understanding

Usually, the advisor observes student's behaviors to classify the student in three groups of student are risk, problem and normal. In the process of phrase, consideration the six factors are ability, health, family and economy, drug, safety and sexual.

B. Data Understanding

Developing classification model covered data in student's profile database. The attributes of the entities (six types of student's behavior) represent below, "group" is the main attribute for all types, classify three types of student behaviors are normal, risk and problem. The details are shown in the Table 1.

TABLE I. THE RELATED FACTORS.

Type	Related factors
1-Ability	final_grade, absent, learn, readding_skill, skip_class, assignment
2-Health	strong, congential_disease, eyesight_sense, disabled, sick, weight_height
3-Family and Economic	expense, unemployed, family_problem, family_status, residence, dept
4-Drug	smoke, friend_drug, break_disclipline, rent_divorce, introvert
5-Safety	residence, scholl_motoring, brawl_history, dress
6-Sexual	pregnancy, homosexual, live_boyfriend, adulterous, part_time_risk.

C. Data Preprocessing

This data mining task, the data set contains a specific domain which is selected from the plain text, the original data is in student's profile database. Collecting original data is regarded to fill in related attributes of data set, as shown in Fig 2. The missing values are filled by Min, Max and Average methodology. Each attribute value of each model of the dataset is heuristically mapped by their semantic.

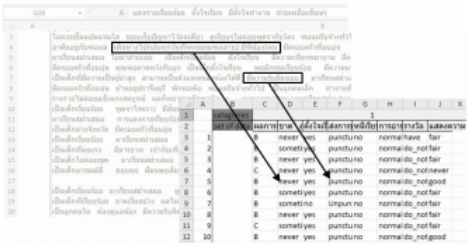


Fig.2. Filtering attributing from plain text.

D. Data Modeling and Evaluation

Generating rule for the classification method use the decision tree algorithm (C4.5), the learning algorithm construct the model and leans a mapping from attribute values to class [11]. Collecting a set of data by cross validation technique [12], repeating the sample of data set until a perfect score is representation, avoiding a failure of predicting or overfitting by performed the available data set before automatically models generating. This paper, the proportion of hold-out is 70:30, the method reserve amount of data for training set and test set [13].

Discovery knowledge from C4.5 performs a multiple decision model. The quantitative results of each model are compared between cross validation technique and Hold-out

technique in terms of classification accuracy, results are shown in Table 2.

TABLE II. EXPERIMANTAL REULTS.

Type	Hold-out (70 : 30)	Cross-validation (10 folds)
1-Ability	94.09%	90.82%
2-Health	91.86%	92.09%
3-Family and Economic	94.12%	92.72%
4-Drug	97.74%	96.52%
5-Safety	94.57%	94.62%
6-Sexual	91.86%	92.72%

The accuracy of experiments is greater than 90%. Result of Hold-out technique, type-1 (Ability), type-3 (Family and economic) and type-4 (Drug) are more significant in cross validation technique. This occurrence may be caused by cross validation is compared all of attribute that generates a possible path or scenario for risk behavior models.

E. Deployment

The six decision tree models are used to constructed the student's behavior guideline that provides characteristic of student behaviors for six types predicting information. The general guideline consists of characteristic and the way to improve or solve a problem, all of observing types that represents in a sequential of step as shown in Figure 3.

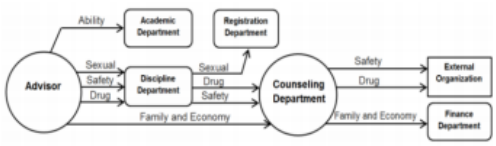


Fig.3. Risk behavior process.

According to general guideline shown in Fig 3, the advisor is separated and regarded risk students to transfer them to related department for improving their behavior. To exemplify, the students of a drug type risk behavior transferred to discipline department and they evaluated drug behavior by teacher random checks 3 times a semester. After the last evaluating, teacher considers a student who cannot change risk behavior to drug institute. The factor of determining a student to a group represented in decision tree model by the Fig 4.

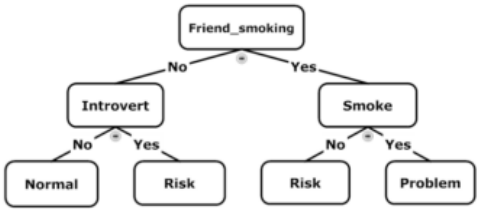


Fig. 4. Decision tree of type-5 model.

Finally, the decision tree model of the drug factors are conceptualized into more interpretable guideline that is summarized in the Table 3 below, the information provides the teacher for determining a group of student.

TABLE III. THE RISK BEHAVIOR GUIDELINE OF TYPE-5 DRUG

Student's behavior		Methods
Normal	- Student has no friend who smokes a drug. - Student is not depressing.	The institution should always provide the activities or training about drug addition for all student groups. Checking smoking drug of student 4 times to classify a group of student to tree type of risk behaviors. Then, random repeat check smoking drug to reflect a status of risk and problem student group from all student. A teacher use checking result and consider problem level, transition student to institution.
Risk	- Student has a friend who smoke drug but she or her never smokes. Student sometimes lives alone and introvert.	
Problem	Student smokes a drug.	

IV. CONCLUSIONS

This paper, the decision tree is applied to construct an adolescent student risk behavior guideline and six classification models based on information of student database which is beneficial for the prediction of the student's behaviors. The risk is separated into three levels: Normal, Risk, and Problem. The accuracy of all classification models is greater than 90%. The student's behavior guideline would facilitate the advisor in decision making in student care and help system. For future work, the way to management education is change to 21st century trend that emphasize skill to living. Then, the help student system is more necessary for school, risk behavior guideline is a tool that can provide a teacher for classification student process. Moreover, the current guideline would be evaluated the satisfaction by an additional survey.

REFERENCES

[1] C. Nokdee, "The statistic of adolescence pregnancy". Retrieved January 5,2015, from Thai Health Promotion Foundation Website: http://www.thaihealth.or.th/Content/25667.html, 2010 .

[2] No writer. Quality of life :Drug statistic of adolescent. ASTV Manager, 2013, p. A3 .

[3] N. Udomporn, "Stress Reduction. Retrieved December 27,2014", from Nation News Bureau Of Thailand Website, 2014 .

[4] Bureau of Academic Affairs and Educational Standards Thailand, "Student care and help system guideline. Bangkok: Transportation Organization of Thailand", 2006 .

[5] Q. A. Al-Radaideh, E. M. Al-Shawakfa and M. I Al-Najjar, "Mining Student Data using Decision trees". International Arab Conference on Information Technology, Yarmouk University, Jordan, 2006 .

[6] I. H. Witten and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques (Second Edition)", Canada: John Wiley & Sons, Inc., Hoboken, New Jersey, 2006 .

[7] B. Kumar Baradwaj and S. Pal , "Mining Educational Data to Analyze Students' Performance", International Journal of Advanced Computer Science and Applications (IJACSA), Singhaniya University and VBS Purvanchal University, India, 2011.

[8] A. Eddr EL Din Ahmed and I. brahim Sayed Elaraby, "Data mining : A Prediction for Student's Performance Using Classification Method", Cairo university, Egypt, 2014.

[9] Prasartwanakit, Songwathana and Phetcharat, "Sexual beliefs and patterns among Thai adolescents and youths in educational institutions in Songkhla Province", Prince of Songkla University, Thailand, 2010.

[10] A. A. Freitas, "Data Mining and Knowledge Discovery with Evolutionary Algorithms", Computing Laboratory, University of Kent, UK, 1964 .

[11] X. Wu and V. Kumar, "The Top Ten Algorithms in Data Mining", Taylor & Francis Group,2009.

[12] D. T. Larose and C. D. Larose, "Data Mining and Predictive Analytics", Canada : John Wiley & Sons, Inc., Hoboken, New Jersey, 2015.

[13] K. P. Soman ,S. Diwakae and V.Ajey, "Data Mining: Theory and Practice, India : Prentice of India PrivateLimited, New Delhi, 2006.

BIOGRAPHY

NAME	Miss Surin Aunsan
DATE OF BIRTH	28 April 1990
PLACE OF BIRTH	Chaing Mai, Thailand
INSTITUTIONS ATTENDED	Maejo University, 2008-2012 Bachelor of Science (Information Technology) Mahidol University, 2014-2016 Master of Science (Information Technology Management)
RESEARCH GRANTS	Grant to Support Graduate Students in Academic Presentation in Thailand Academic Year 2015
HOME ADDRESS	84 M.12, Banmae, Sanpatong, Chaing Mai, Thailand, 50120 Tel. 088-400-7801 E-mail : it_surina@hotmail.com
PUBLICATION / PRESENTATION	Aunsan, S., Thammaboosadee, S., (2015), Constructing a Risk Behavior for Adolescent Student Using Decision Tree, The Proceeding of the 2nd Management and Innovation Technology International Conference (MITiCON2015), pp. 256-259