

**RATING COMPUTATION USING DECISION SUPPORT
SYSTEM: A CASE STUDY OF LEASING COMPANY**

PATCHARAPONG SAMANTARN

**A THEMATIC PAPER SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)
FACULEY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2015**

COPYRIGHT OF MAHIDOL UNIVERSITY

Thematic Paper
entitled
**RATING COMPUTATION USING DECISION SUPPORT
SYSTEM: A CASE STUDY OF LEASING COMPANY**

.....
Mr. Patcharapong Samantarn
Candidate

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

.....
Lect. Taweesak Samanchuen, Ph.D.
(Electrical 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
Technology of Information System
Managment
Faculty of Engineering
Mahidol University

Thematic Paper
entitled
**RATING COMPUTATION USING DECISION SUPPORT
SYSTEM: A CASE STUDY OF LEASING COMPANY**

was submitted to the Faculty of Graduate Studies, Mahidol University
For the degree of Master of Science
(Technology of Information System Management)

on
May 23, 2015

.....
Mr. Patcharapong Samantarn
Candidate

.....
Asst. Prof Adisorn Leelasantitham, Ph.D.
(Electrical Engineering)
Chair

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

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

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

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

.....
Lect. Worawit Israngkul, M.S.
Dean
Faculty of Engineering
Mahidol University

ACKNOWLEDGEMENTS

This thematic paper was successfully completed by the assistance of Assistant Professor Dr. Supaporn Kiattisin, Major of Thematic Paper and Lect. Dr. Taweesak Samanchuen, Co-advisor of the Thematic Paper who have given numerous advices and instructions as well as reviewed the paper to ensure correctness of the contents.

I would like to thank the professors at the Technology of information system management, The Faculty of Engineering who have furnished me with valued knowledge and experiences. I also would like to express my thanks to the officers, my seniors and my friends who have given full supports both on my study and my personal life while I was working on this thematic paper.

Thank my parents and relatives for their permanently supports.

Last but not least, I would like to thank the sample company for providing and allowing me to use the data even though the name of the company was not disclosed. Thank my colleagues for facilitating and providing access to the data for this research.

Patcharapong Samantarn

RATING COMPUTATION USING DECISION SUPPORT SYSTEM: A CASE STUDY OF LEASING COMPANY

PATCHARAPONG SAMANTARN 5537929 EGTI/M

M.Sc. (TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)

**THEMATIC PAPER ADVISORY COMMITTEE: SUPAPORN KIATTISIN, Ph.D.,
TAWEESAK SAMANCHUEN, Ph.D.**

ABSTRACT

This thematic paper has the purpose of creating a decision support system to compute the interest rate in the case of a Sample Company, which is in the leasing business. The sample company has a problem, which delays the execution of new contracts because the rating method is calculated from the forecasts and past experience of a quotation maker. Most of the quotations that are sent to the customer are revised interest rates even from the customer or approval stage of the Sample Company. To improve workflow of the business to become more efficient, we developed a decision support system, which chooses the appropriate interest rate. The rate chosen is the most suitable for each transaction using historical data from approved transactions. The system uses data mining to build a decision tree using the C4.5 algorithm and calculation from the WEKA program. This result is used to build the decision support system for an interface to an Excel file for universal access which the user can easily understand. The result from the system, when tested with a testing set, has an accuracy rate of 82.33%, and a system satisfaction questionnaire showed great satisfaction in every segment, except the processing accuracy segment, which was ranked as satisfied.

**KAY WORDS: DATA MINING / LEASING / C4.5 / DECISION SUPPORT
SYSTEM / DECISION TREE**

78 pages

การคำนวณอัตราดอกเบี้ยโดยระบบสนับสนุนการตัดสินใจ กรณีศึกษาบริษัทตัวอย่างในธุรกิจลีสซิ่ง
RATING COMPUTATION USING DECISION SUPPORT SYSTEM: A CASE STUDY OF
LEASING COMPANY

พรพงษ์ สมนานธรัตน์ 5537929 EGTI/M

วท.ม. (เทคโนโลยีการจัดการระบบสารสนเทศ)

คณะกรรมการที่ปรึกษาสารนิพนธ์ : สุภาภรณ์ เกียรติสิน, Ph.D., ทวีศักดิ์ สมนานชื่อ, Ph.D.

บทคัดย่อ

งานค้นคว้าอิสระฉบับนี้ มีวัตถุประสงค์เพื่อจัดทำระบบสนับสนุนการตัดสินใจในการคำนวณอัตราดอกเบี้ยของบริษัทตัวอย่างในกลุ่มธุรกิจลีสซิ่ง โดยปัจจุบันบริษัทดังกล่าวประสบปัญหาความล่าช้าในการทำสัญญาเนื่องจากการคำนวณอัตราดอกเบี้ยแบบเดิมนั้นยังคงใช้วิธีการคาดคะเนจากประสบการณ์ของผู้จัดทำ ซึ่งมักจะเกิดปัญหาการแก้ไขอัตราดอกเบี้ยดังกล่าวจากทั้งฝ่ายลูกค้า และฝ่ายผู้อนุมัติของบริษัทตัวอย่างเอง เพื่อช่วยพัฒนากระบวนการในการดำเนินธุรกิจให้มีประสิทธิภาพยิ่งขึ้น ด้วยเหตุนี้จะได้สร้างระบบสนับสนุนการตัดสินใจเพื่อคัดเลือกอัตราดอกเบี้ยที่เหมาะสมที่สุดสำหรับการทำสัญญาลีสซิ่งแต่ละสัญญา โดยใช้ข้อมูลสัญญาที่ได้รับอนุมัติแล้วของบริษัทมาจัดทำระบบสนับสนุนการตัดสินใจ และใช้วิธีการทำเหมืองข้อมูลเพื่อจัดทำต้นไม้ตัดสินใจ โดยใช้วิธีการ C4.5 ในการจัดทำต้นไม้ตัดสินใจ และใช้โปรแกรม WEKA ในการคำนวณดังกล่าว จากนั้นนำผลลัพธ์ที่ได้จากต้นไม้ตัดสินใจดังกล่าว จัดทำระบบสนับสนุนการตัดสินใจในรูปแบบของ Excel ไฟล์ เพื่อให้ผู้จัดทำสามารถเข้าถึง และเข้าใจระบบได้โดยง่าย สรุปค่าผลลัพธ์ที่ได้จากการทดสอบกับ Testing Set มีค่า Accuracy Rate เท่ากับ 82.33% และผลจากการตอบแบบสอบถามในด้านต่างๆ อยู่ในระดับ ดีมาก ยกเว้นในส่วนของ ความถูกต้องของระบบ ซึ่งอยู่ในระดับดี

78 หน้า

CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
ABSTRACT (ENGLISH)	iv
ABSTRACT (THAI)	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER I INTRODUCTION	1
1.1 Definition of leasing business.	2
1.1.1 Differences between Hire purchase and Leasing contract	3
1.1.2 Benefit from leasing contract	4
1.1.3 Types of leasing contract	7
1.1.4 Competitive in leasing business	8
1.2 A sample company	9
1.2.1 Internal system structure	10
1.2.2 The analysis of a system from the sample company	10
1.3 Problem issues	12
1.3.1 Many revision of quotations	12
1.3.2 Delaying of each process	12
1.4 Purpose and benefits	13
1.5 Limitations of the research	13
CHAPTER II LITERATURE REVIEW	14
2.1 Related researches	14
2.2 Definition of Decision Support System	15
2.3 Data Mining	17
2.4 Creation of Decision Tree prototype	24
2.5 Algorithm C4.5	36
2.6 Instruction to program “WEKA” for build decision tree	36

CONTENTS (cont.)

	Page
CHAPTER III ANALYSIS AND DESIGN THE SYSTEM	41
3.1 Methodology	41
3.2 Determine objectives	42
3.3 Data understanding	47
3.4 Data preparation	47
3.4.1 Data selection	47
3.4.2 Data processing	52
3.4.3 Data cleaning and exploration	53
3.4.4 Data transformation	53
3.4.5 Data Engineering	58
CHAPTER IV SYSTEM DEVELOPMENT	59
4.1 The result from Weka Program	59
4.2 Creating decision tree	60
4.3 Conversion of Decision Tree into spreadsheet	62
4.4 Interface Creation	63
4.5 VBA coding	65
4.6 Testing DSS in testing set data	68
4.7 The accuracy testing by experts	68
4.8 Satisfaction test of the system	72
CHARTER V CONCLUSIONS	75
5.1 Summary of system development	75
5.2 Recommendations and future work	76
REFERENCES	77
BIOGRAPHY	78

LIST OF TABLES

Table	Page	
2.1	Data of sample case which selling of air conditioner	26
2.2	Attribute “Age” Clarification	28
2.3	Attribute “Income” Clarification	29
2.4	Attribute “Working” Clarification	29
2.5	Attribute “Credit” Clarification	30
2.6	Limb “Age<=30” by “Age” Clarification	30
2.7	Limb “Age<=30” by “Income” Clarification	31
2.8	Limb “Age<=30” by “Working” Clarification.	31
2.9	Limb “Age<=30” by “Credit” Clarification.	32
2.10	Limb “Age>=40” by “Age” Clarification.	33
2.11	Limb “Age>=40” by “Income” Clarification	33
2.12	Limb “Age>=40” by “Working” Clarification	34
2.13	Limb “Age>=40” by “Credit” Clarification	34
2.14	Sample data for import to program “WEKA”	36
3.1	Result of observation factor interview	50
3.2	Structure explanation of data	57
4.1	Userform1 structure.	64
4.2	Conclusion testing set classification	68
4.3	The questionnaire for expert	69
4.4	Result from questionnaire	70
4.5	The System Satisfaction Questionnaire	73
4.6	Conclusion of System satisfaction	74

LIST OF FIGURES

Figure		Page
1.1	Leasing business investor by nationality	9
2.1	Sample of decision tree	24
2.2	Root node of decision tree	27
2.3	1 st node of decision tree	28
2.4	Completed decision tree from sample data	35
2.5	Welcome interface of program “WEKA”	37
2.6	“WEKA” explorer window.	37
2.7	After import data to “WEKA”	38
2.8	Sample of numeric data in “WEKA”	38
2.9	Calculated result from “WEKA”	39
2.10	Decision tree from “WEKA”	40
3.1	Research methodology	42
3.2	Sample company work flow	43
3.3	Workflow after included DSS	45
3.4	Flow chart of DSS	46
3.5	Sample of importing data VBA code	48
3.6	Sample RAW data	49
3.7	Sample data after data transformation process.	58
4.1	Result in Run Information section	59
4.2	Result in Summary section	60
4.3	Rode node of tree	60
4.4	Decision Tree Node (Cost \leq 1,003,210.2)	61
4.5	Decision Tree Node (Cost $>$ 1,003,210.2)	61
4.6	Sample data in page “Tree”	62
4.7	Data in page “DATA”	63
4.8	Userform1	64

LIST OF FIGURES (cont.)

Figure		Page
4.9	DSS interface with sample data	67
4.10	DSS interface with result	68

CHAPTER I

INTRODUCTION

In the present day, the financial institution in Thailand has been developed from the past to the present day which can be noticed from new financial products that continuously enter the market. This can be applied to both capital market and currency market; for example, Single Stock Future, Gold Future, and other derivative instruments. This becomes a motivation to push economy within the country to continuously grow. The index for national economic is for example; the product with the nation total mass. The GDP in the year 2013 is worth 387,250 million dollars increasing from year 2012 at 365.97 million dollars (increased for 5.8%). Moreover, it also has increasingly grown rate within the latest four years. The index at SET or Stock Exchange of Thailand which is the main market for buying and selling capital in the country has the index of 1,500 points. This has increased from the previous year which was only at 1,200 points.

The indirect financial institutions; such as, Commercial Bank, Capital Property Company and other financial institutions have a growth rate according to the economic conditions. The main commerce banks in Thailand arrange in order of Total Asset including; Bangkok Bank, Krungthai Bank, Siam Commercial Bank, Kasikorn Bank, respectively. These financial institutions have been making a total profit of 134.49, 137.30, 168.15 and 145.96 billion, respectively. Although most of the profit derived from interest of credit products, this is because credit products are considering the main product that created profit for commerce banks. However, the other financial activities that should not be overlook is hire-purchase or leasing business which has create profit no less than other types of product for the banks. Even if Thai people does not have wide knowledge of Leasing business and it still has not expand in individual type market. However, if we focus in the term of leasing, it is a continuously grow business in the country. The example is leasing business in the type of car hire-purchase or what it is usually called as “Car Finance” business.

1.1 Definition of leasing business.

It is simple to say that when a customer or individual would like to purchase a car, however that person does not want to buy the asset, which in this case is a car with cash or there is not enough cash for purchase. The person will use the financial service of a leasing company. The leasing company is a representative to purchase the car for the customer. The property right in ownership is of the leasing company, but the person who uses the service has a right to use that asset. The process is to pay installment to the leasing company instead. The leasing company will add the additional interest rate to the installment. In this case, it will mean an interest rate that is constant for the entire contract or Flat Rate which is different from the loan interest rate. Other types of loan interest rate including; Over Draft, Home Loan or Equity Loan which will have the characteristic of Fixed Rate or Float Rate, this will depends the referral interest of that certain bank or the standard interest. They will not count it as Float Rate. Moreover, after a complete installment with leasing company up until the end of the contract, the leasing company will transfer the ownership of asset back to the customer. Although hire-purchase contract or leasing contract is very similar with the steps as mentioned but there are some differences between a hire-purchase contract and leasing contract.

A characteristic of Hire Purchase is similar to purchasing product with partial installment. However the difference is the ownership of the asset will not belong to ours until the installment is completed. This is that until there is a contract or agreement between hire-purchasee person and hire-purchasor regarding partial installment according to amount of money and time limit. During that time, we can use the product or asset for using. However the ownership remain with the hire-purchasor up until we have completed the installment then the ownership become ours. The characteristic of this contract has been usually done with asset; such as automobile; for example, hire-purchase of different types of cars or motor bicycle. Moreover, the characteristic of a leasing contract has a similar contract to hire-purchase contract which is a person who requested leasing must pay a partial installment according to the amount of money and time schedule. However, the differences are that after the contract is completed, a person requested leasing can choose whether to purchase the asset, continue with the contract of the asset or return the asset back to the leaser.

Mainly, those who commit these types of credit usually are a company, corporate, or business that would like to hire-purchase asset that is expensive or large quantity of asset. The examples are machines, automobile, or it could be a product that has a characteristic that may change fast in technology; such as, office computer, office tools, or photocopy machine.

1.1.1 Differences between Hire purchase and Leasing contract

The difference between Hire purchase contract and leasing contract can be distinctly divided as follow.

A law that uses to enforce that certain contract, for the case of hire purchase, there is an article stated in civil and commerce law. However, for the leasing case, there has not been clearly specify, but they did give the meaning in separating types of contract as Financial Lease and Operating Lease

The customer for the leasing company mainly will undergo contract under corporate; but for the hire purchase customers, it can be both individual and corporate within the same ratio.

The purpose of hire purchase is to become the ownership of the asset. However, for the leasing, the person who requested leasing can choose at the end of the contract, whether to complete residual value or they can extend the contract or they can return assets back to the leaser. This will depend on the characteristic of that leasing contract; such as, operating lease, usually the leaser will not be able to purchase the asset after the contract is completed. This is because the contract itself did not specify the residual value. Therefore, the leaser will not be able to purchase the asset.

The person who requested leasing can make a total amount of contract according to the value of the asset that will be used for leasing contract. However, for hire purchase, the credit can only be requested for 70%-80% according to an article stated in law of civil and commerce.

The propriety right in ownership of the asset as well as the record of the account and acknowledge of the corporate's income that use both types of hire purchase contracts.

The cancellation of contract can be done by the person who requested leasing by returning the asset back to the leaser. However, for the leasing type it is depending on the contract; such as, leasing contract in the form of finance leasing. The person requested leasing must pay the full installment as well as residual value of the contract from the day of the cancellation until the day the contract is completed. Then they will gain the right of cancellation of the asset.

The hire purchase company will know the rate of return for the hire purchase contract from calculating the Flat rate. However, for a leasing contract, it will be calculate with the effective rate. In the perspective of purchaser or a person who requested leasing, they will only know the monthly installment or residual value after the contract is completed. In the case of a leasing contract for finance lease, and they will know only part of leasing for the operation lease.

A leaser has a duty to give the ownership in a perfect condition to the person requested a leasing and a leaser also is responsible to take care of asset that is dysfunction or the payment occur after the day the ownership is transferred. For the leasing contract, it is depending on the specification and type of contract; such as operation lease combine with maintenance lease. The leaser will be responsible for the payment derive from the use of that asset and if the asset is an automobile. The leaser will be responsible for insurance, maintenance, as well as registration fee.

1.1.2 Benefits from leasing contract

A leasing company in Thailand has a continuous grow, not only automobile leasing that many people know, but there is also equipment leasing, office tools leasing and other type of assets. It can be said that almost every types of asset use in business can be used for leasing. The main reasons that corporate choose a leasing contract instead of purchasing it with cash or hire purchase can be seen as follow.

1) Tax benefits

A tax benefits is considering the main reason for corporate to be interest in doing leasing type of contract. This is because when comparing between purchasing asset with cash and leasing in some case, after a deduction of corporate budget that is being display to Revenue Department. The amount that has been deducted according to a leasing contract will worth more and the deduction of residual value of the certain

asset comparing to the corporate purchase the asset with cash, which will make the company able to save cost of tax from the event mentioned.

2) Able to find the asset within a short period of time

It also give a leaser a chance to find asset within shorter period of time comparing to purchasing or building it; such as, building or warehouse. Also, if a person requested leasing or the corporate would like to invest to build by themselves, they may need to request for large sum of financial structure and credit. Therefore, the responsible for the project and the risk for all the responsibilities will depend on that corporate. For the hire purchase, the person requested leasing can choose warehouse and hire purchase contract in the form that they preferred. They will also be able to benefits instantly from the asset. It can be said that the corporate can use the asset immediately and does not need to take responsibilities and risk from to the construction project.

3) Suitable for asset that has seasonal needs

For an asset that is needed temporary; for example, sometimes the corporate need to construct or repair land, building, or machinery. And the tools for repair could be expensive and for the company to invest to purchase the machinery for temporary may not worth the investment. Therefore, the operation leasing contract could be a better solution.

4) Suitable for asset that can easily deteriorate

Leasing contract is suitable for asset that is easily deteriorated or out-of-date; such as, computer or electronic machines. Also, if the corporate decide to purchase the asset, they may need to be responsible for the cost in selling the asset or eliminate the asset when it is out-of-date. Therefore, if a corporate did a leasing contract, they will be able to solve the problems regarding elimination costs or risk in selling out-of-date asset will be given to the leasing company instead.

5) Able to choose according to needs of the company

There are different types of leasing contract, so a person requested leasing is able to choose the suitable type of leasing that is appropriate for their businesses. This is including operation or finance type as well as financial policy and company account policy as well.

6) Reduce the risk of interest rate

This is to say that the corporate who requested for a leasing contract, they will only know the partial installment and residual value after the contract is completed. This is to compare with the loan credit to purchase the asset. The loan contract may depend on the interest float rate of that bank; such as, MRL (Minimum Loan Rate). The person who requested loan has to be responsible for interest and the responsibility for the changes of interest rate by themselves. This is different from a hire purchase contract that will remain constant until the contract is finished.

7) Off balance sheet financing

Another advantage for a leasing contract is the off balance sheet financing. This means that the rate for debt will remain constant. This is because a leasing contract in the form of operation lease and other types of lease, they do not have to record the account of corporate. Therefore, the corporate does not need to deduct the residual value which result in the entire asset value did not increase. Thus, the financial ratio is being compared between the income and ROA (Return on Assets) will not be reduced.

8) Result in good liquidity of the company

The company that has good financial management usually will not neglect the financial liquidity of the company. This means to have cash or asset enough to circulate to use within the company. This also means if the company would like to purchase asset with cash, the company will need to change it to permanent asset or even to add the investment or convert other types of asset to cash to be able to purchase the asset. However, if the corporate does not possess enough cash to purchase the asset, a leasing contract can make the corporate benefit from using the asset. It will also not affect the liquidity of the company.

9) Remove the problem of account record

It also removes the problem of recording account regarding asset of the company. This means that if its large company that use large amount of office tools, the purchase of the asset could result in the corporate have to record the asset account. It may also include calculating the entire residual value which may be a difficult task to evaluate the value of asset and record the asset. A leasing contract will enable the corporate to pay an installment only for the certain asset. And the corporate do not

need to record the asset in the finance budget of the company and it will also not affect the debt rate of the company as mentioned earlier.

10) Cheaper prices

The capital for asset that the corporate has included in leasing contract usually has a cheaper capital comparing to purchasing it with the cash. This means that the leasing company will be able to purchase a large quantity of asset within one time. This is because leasing companies have their own supplier. Therefore, a leasing company would get discount from purchasing more than the corporate whenever they purchase asset; such as automobile.

From the advantage of leasing mentioned above, a leasing contract still has some limitations that make it unsuitable for some companies. The example is that the system is not suitable for a large company with asset management system and good finance system. Therefore, a leasing company will mainly focus on the medium to small size company. This is especially with the international company with Japanese nationalities. They will have their mother company in Japan and the main shareholders are Japanese and their main management policies will prefer leasing to purchasing asset permanently.

1.1.3 Types of leasing contract

The leasing contract can be separate into 2 types; which are, Financial Lease and Operating Lease. Each type has both advantage and disadvantage, so the person requested leasing should study the conditions of contract carefully before decide the right type. The main component in separating these types are that If its financial lease, the contract will create a risk transferal and the entire or almost the entire return will be given to the owner. It does not matter if the contract is finished, whether the ownership is completed or not. The type of leasing contract will enable these situations to happen. It will also considering as financial leasing.

The contract to transfer ownership of asset to the person requested leasing after the contract is completed.

A person requested leasing has a choice to Bargain Purchase Option with the lower price than the value on the date it is set. The price according to the right has to be lower than justify value of the asset to be sure at the day started of the contract.

The person requested leasing has a choice to bargain purchase option the asset. However the % has not been specified for the fair value yet.

Time limits for the leasing contract will cover the economic life. For the asset, although there is no ownership transferal, the economic life will depend on the useful life of the manufactural as well as shelf life

At the start of contract, the present value of the minimum amount of money that need to be paid is equal or almost equal to the justify value of the leasing asset.

If a person requested leasing can cancel the contract, he/she will be responsible for the damages caused due to the cancellation.

A person requested leasing is the one who received benefit or losses from the changes in justify value of the remaining value.

A person requested leasing can continue the contract for a second time by pay the installment that worth less than the market installment.

Thus, the situations above does not consider as the settlement for separating types of leasing to a financial leasing type. However, if it is clear that there is another characteristic that does not include risk transferal and owner benefit from the asset, then the lease might be characterize as operating lease. The operating lease is a lease that is not a financial lease.

1.1.4 Competitive in leasing business

The companies that rerolled under Department of business development in sector of leasing business ranking by total asset of companies. The most of companies are classify in sector of automotive business that impact form the several year of government policy. The leader of aforesaid business is Toyota Leasing Co., Ltd., Isuzu Leasing Co., Ltd and Honda Leasing Co., Ltd. which have the total asset more than 1 billion baht (from the financial statement in 2013). From the data for department of business development we rank investing by nationality of investors in aforesaid sector by investing value.

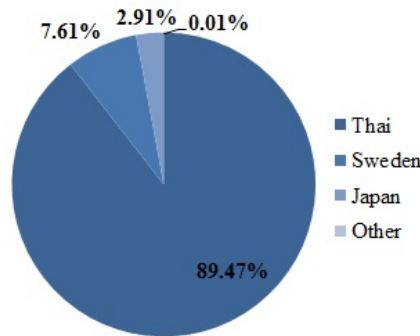


Figure 1.1 Leasing business investor by nationality.

In part of financial statement of companies, in 2012 it have 75 companies sent completed financial report with summary of total asset to 231,565.13 million baht, sales to 16,385.26 million baht and net profit to 1,767.73 million baht which classified in small, medium and large company by

- Small company : is company that has fixed asset not over 50 million baht.
- Medium company : is company that has fixed asset over 50 million baht but not over 200 million baht.
- Large company : is company that has fixed asset over 200 million baht.

1.2 A sample company

This research has collect information for the report from one of the leasing companies that has several services regarding hire purchase contract and different types of leasing contract; such as, financial lease, operating lease, hire-purchase contract, and maintenance contract as well as financial services. They also provide hire purchase for environment and also this company has been operated in Thailand as of year 2014, they have 60 million baht registered capital stocks. It has 51% Thai shareholders and 49% Japanese shareholders. Their main customers are Japanese Nationality Company and they also have business that collaborate with Japanese

nationality companies. The ratio for their present customers is 97.51 percent of Japanese customers and other nationalities at 2.49%. This will be divided according to the company conditions. Moreover, they also have experience in leasing business for over 20 years.

1.2.1 Internal system structure

The Marketing Promotion department is a department that contact and coordinate with customers as well as find new customers and maintain the relationship with the existing customers. They will be the one that coordinate between customers and other department in the company. They will also be responsible for finding new contract and they will have power in specify the return rate for each contract in the beginning stage.

The Operation & Credit department is responsible for analyze the information for customers regarding Credit, making different types of contract, and document verification. It will also include verify the customer's credit and make the financial report that is not related to the finance department. They will also overlook the company internal system; such as, approval process, contract approval, as well as the flow of the documents.

The Marketing Coordination department works as Marketing Promotion; however, they are the department that works after the customer has contract approval. They will follow up with customers so they will not have problem with the asset or the payment with the supplier.

The Rent Collection department is a department who collects rent from the customers as well as manages different types of financial documents; such as, invoice or receipt.

The Accounting & Treasury department is a department that responsible for manage budget and taking care of cash flow in the company. It will also be responsible for the loan of the company and debt payment of the company.

1.2.2 The analysis of a system from the sampled company

The 5 departments that have been mentioned earlier will collaborate together to drive the company toward success and for the company's benefits. The

business process is the process that they used to operate their businesses from starter to final stage and each stage differs. This study will focus on the requesting for lease from the company from starter to the final stage where the customers have been approved of the leasing contract. This will be the end of this study and it can be described via process as follow.

Step 1

It will start with the marketing department have a meeting with the customer to contact and collaborate. They will divided into new customers and existing customers. For the case of new customers, the marketing officer will start by present about the company's product for sales. For the case of existing customers, the marketing officer will propose other suggestions or offers customer to lease other asset or continue with the lease. This will depend on the case, after the process is done. The marketing officer received the information regarding requesting for hire purchase contract, customers name, and the capital of the asset that customer preferred.

Step 2

From the step mentioned earlier in the step 1, the marketing officer will send the information that they have received to the Operation & Credit department. This is to calculate the rent cost to be able to handle the quotation and return it back to the customers. This is to reaffirm the lease contract. The calculation consists of different factors; such as customer rating, type of lease, time of lease, as well as nationalities of customers. The rent cost determination will be determined by marketing officer.

Step 3

After the marketing officer has received the Quotation, they will send the Quotation to the management team in the company to sign the quotation. This means they have accepted the new transaction and then send the quotation back to the company of the customer to sign the quotation. However, according to the interview, this is rather difficult to proceed; this is because the delivering of quotation for both takes long time. Moreover, most customers tend to like Shop rate which means they did not contact with only our companies.

1.3 Problem issues.

From the study regarding information and company operating system, the researcher can identify the problems into topics as follow.

1.3.1 Many revision of quotations.

After the Requirements have been received from the customer, the marketing officer has to choose interest rate for each transaction which is most suitable for each customer. This is because customer will have characteristic as can shopping rate and with the competition between another leasing companies, there are several competition regarding price. Therefore, if a customer receives a quotation and there has been comparison with other company it may leads to customer decide not to sign the quotation. They may confirm the quotation with other company instead. Another issue is if the Marketing has given lower number of Rating in hoping that customer would confirm, however after the quotation sent, it maybe not approve in approval stage of company. From both of the situations mentioned, it leads to several problems in revise the document. This means that when the customer decide not to use the interest rate. Additionally, the Marketing officer needs to revise the document as well as the case of unapproved from the management that the marketing also need to redo the document from the beginning to the final stage.

1.3.2 Delaying of each process.

From the analysis of the process of work, it shows that the minimum day is 3 days. This is from the meeting with the customer until the day the customer decided on the service. It is usually take about 3-5 days in case of amend of the Quotation will take 3-4 days. Also, if it is needed to gain approval from the mother company, it may take longer period of time. This will result in the decision making of the customer. Due to high competition in the market, customer may change their minds to other leasing competitors in the market.

1.4 Purpose and benefits

To create system to support the decision making from the existing historical company data. This can be done by data mining through selecting the factors and suitable method for decision tree to find the condition which lead to variables. It will be done with the use of interest rate of the contract as the following variable.

To use the conditions that is the result from the decision tree in the previous process. This is to create system to support the decision in selecting return rate and conduct a calculation program with File Excel. This is so the users of the system can predict the return rate by input different factors as well as enable user to calculate return rate according to the conditions.

1.5 Limitations of the research

This research has collected all the information regarding the sampled company from 2012-2014. The data was in Excel Files that has been record separately in each contract. The researcher has been separating the data into 2 parts and uses the first part as training set to create a decision tree. After it has been done, the researcher uses the data from second part as testing set to experiment with the decision tree to find the confidence of the decision tree. Then 50 transactions of training set are collected from random and expert person in leasing field will calculate the result from their opinion. Next the result from expert person and decision support system will be compare and discussions. At last decision support system will be evaluated in part of satisfaction of using by user.

CHAPTER II

LITERATURE REVIEW

From purposes of this research, the researcher has been studying about theory and related researches of this study. The main target is on the Decision Support System by combining data from the sampled company to analyze and further develop into a present work flow of sampled company.

2.1 Related researches

Wah YB, Ismail NH, Fong S.[1] They researched about behavior of buying car by created Decision Support System for forecasting executed contract rate and unexecuted contract rate after customers booked a car. They created systems by several methods that Logistic Regression (LR), Decision Tree (DT) and Neural Network (NN) and compare result of each method which from same data used. They separate raw data to 2 groups, Training Set and Testing Set by 70% and 30% respectively. Result of systems shown accuracy rate not much different and they had conclusion which accuracy rate depend on characteristic of data used and choosing factors. LR method had the most accuracy rate and DT method had the second most accuracy rate which 91.79% and 91.17% respectively.

Somnuk Krajang [2] reported research about improving decision support system which approval credit transaction in case of Pattanasin Leasing (CPL) Co., Ltd.. The company is in leasing business especially auto mobile and main customer is individual. Created DSS was built for classifying customer to segment rank that good, medium and poor. The system was built by decision tree method by information gain value, algorithm ID3 and separated data to 2 set. The first is training set and the second is testing set. Then built decision tree by training set and find accuracy value by testing set. As a result 74.70% accuracy and the system can help user achieve the purpose.

Paitoon Janreung [3] reported research which has purpose to forecasting trends of student grade in bachelor degree student in faculty of business administration and information technology. DSS is the result of decision tree that created by ID3 algorithm. Forecast factor of tree is each student grade which quantity value. So he separated forecast factor to intervals that Excellent, Good, Medium and Poor. Then he created 12 decision trees which has accuracy value from 78.64% to 92.31%.

Fan J, Wen P [4] they research which behavior of online learning behavior. More than 90% of teachers and students in program were satisfied with result that provided an objective, reasonable method for assessing online learning. They created Decision Support System by Algorithm C4.5 for grouping learning behavior. Criteria factor is final grade of each student. The result of system can provide appropriate studying plan for each student.

According to researched aforesaid. Researcher can conclude that Wah YB, Ismail NH, Fong S.[1] they research several methods for created DSS by the same data used. Logistic Regression has the most accuracy rate but criteria factors of this method has to be only categorical variable and the second most accuracy is Decision Tree which more suitable for this paper. Research Somnuk Krajang [2] it use algorithm ID3 for created DSS which widely used and Sample Company is in similar business with this paper. However criteria factor is categorical variable. In case of Paitoon Janreung [3] it use algorithm which improved from algorithm ID3. Although criteria factor can be Quantitative variable but DSS that was create by Data Mining Program which specific software. It may have problem with installing program and policy of Sample Company, company skill and knowledge of users. Final research Fan J, Wen P [4] they created DSS from Decision Tree by algorithm C4.5 which criteria factor can be quantitative variable in this case is Final grade of each student that similar criteria factor of this paper.

2.2 Definition of Decision Support System

Taweesak Nakmoung [5] explains that Decision Support System (DSS) is software that supports the decision making regarding management and data collection as well as a complex structural under the same software. Moreover, DSS also will help

with works between personnel and technological software. It will works as respond to solve the unstructured problem. It will be under the control of the user from the beginning to the response process by using computer to find an easy solution, convenient and quick from a problem with uncertain structure. Therefore, a decision support system will consist of tools, data, model and other resources that user or analyst use to evaluate and solve problems. The principle of DSS is then to use the necessary tool for the management for the analysis of complex form of data. However, the operation is flexible, so DSS is designed to increase efficiency in work performance not only to respond the needs of data collection.

The DSS components can be separated into 4 parts; as follow.

1) Tool

Tool is the first component and basic structure of DSS. The related tool of Decision Support System will include 3 groups; which are, evaluation tools, communication tools, and monitor tools.

2) Operating system

Operating system is the main component of DSS as it is considered the main component in the operating of DSS. According to the aim and the needs of user, the operating system will include important components; which are, data base, model base and order system of DSS.

3) Data

Data is the data that is used with DSS will be differing from other information technology. An appropriate DSS data will have characteristic that is suitable for usage. It will have accuracy and up to date related to the needs of user. It is also convenient, quick, and complete as well as flexible and able to put into pattern for further analysis.

4) People

People that use DSS can be divided into 2 groups; which are user and supporter. A good DSS needs to have skilled personnel that have a suitable ability to adapt to the system. The personnel need to be creative and able to work to correspond with each other. This is to reach the target of the organization efficiently and effectively.

The characteristic of a suitable DSS should be as follow.

1) Easy to learn and use, this is because a user may have limit ability in information technology. Also, the urgent usage and need of the problem result in DSS need to be convenient to users.

2) Able to respond to users quickly and effectively, a good DSS need to be able to communicate with the user immediately. It needs to respond to needs and response to the user in time. This is especially with the present situation that needs limit time in solving problem.

3) Enough information and demonstration model for suitable support the decision making and corresponding to characteristic of problems.

4) Support the decision making both structural and nonstructural model. This will be different from the information technology system for operation that uses to manage daily information.

5) Flexible and respond to the needs that may change of users, this is because the problems may not be constant and change according to time. Moreover, there are several problems to face that needs information technology that able to manage the data form which is not complex and easy to make decision.

6) For this research, the researcher has divided theory for creating related DSS into 2 parts; which are Data Mining with C4.5 and Multiple Regression Analysis (MRA) with the details as following.

2.3 Data Mining

Shi G. [6] explains that Data Mining is a process that performs with large numbers of data to search for pattern and relation that is hidden in the data base. In the present day, data mining has been applied to use in various types of work; both in business field for the decision making of management department. It also has been use in science and medical as well as economic and social usage. The data mining can be compare to one of innovation in collect and interpret data. Previously, data has been collected simply; it has been evolved to a database that can take the information technology to use to create data mining. This will enable the user to able to find the hidden knowledge within the data. Data mining can be compare to mining that use machines to separate minerals from other unnecessary matters. However, the data

mining is when we gather data from various sources which is knowledge hidden under the database. This will help us understand characteristic of data and understand factors that cause certain matter to occur in some group of data. This can also help us to predict tendency of new data that may occur in the future. The data mining will help us to acknowledge the beneficial knowledge that hidden under large amount of information. The information may be in a form that is visible or invisible through observation. The data mining is then comparing to a tool that help gather pattern of the data and display it distinctly. This is different than other methods; such as report or data evaluation because the data mining does not possess a clear hypothesis.

The evolution of data mining has begun with collect information in a suitable manner with the reliable tools and able to protect from the loss of data which occurred in the year 1960. Data Collection & Database Creation has been developed from the collection of original information to arrange in the appropriate tools and pattern and larger database. In the year 1980, Database Management Systems has begun to create relation between numbers of database and combine them together for the benefits of analyze and effective decision making. Later on in the Data Warehouse & Data Mining generation, the data has been collected from larger database. The data came from all the performance of the business to help in decision making, and the present day, the data has been use to analyze to evaluate the result through the model by using statistical relation.

Han J, Kamber M. (7) explain general process of data mining is as follow.

1) Determine objectives

The purpose specification is to understand the problems and needs of creating data mining. It is to specify purpose to make data mining through analysis and specify the target for problem solving of data mining. The example is to help in decision making for giving credit and increase sales and also to establish standard criterion to evaluate data mining. This can be in both abstract and concrete object; fo example, to increase product sales by 5% or abstract is for this research is to complete predictive modeling or prediction prototype. This is a prototype that is created with the use of human learning characteristic which will include analyze existing data to specify important quality for the data. Therefore, existing data must be data that is

completed, in order to make the modeling to be able to accurately predict the result. And then the connection will be created with the new relation which has a similar characteristic as If-then.

2) Data understanding

This is to collect information and specify quality of data that has been collected. It also include to observe data briefly; such as, statistic of data as well as investigate the completeness of collected data. In this process, the data will be in RAW DATA type which is not yet suitable to use for data mining. There is a process in preparation and selection of data as follow.

3) Data preparation

3.1 Selection

The purpose of Data Selection is to identify existing data source and put it to use for basic analysis to prepare for mining in the next steps. The data selection will be different according to the purpose of each business that has been indicated from the start. Also the data selection will also have been indicated with the type of work that is taken to use and the variables will need to be interpreted. This will include the understanding of the meaning and components of variables. It does not only include word or business description but it is also including a clear detail regarding data, possibilities, data source, data pattern and other characteristics. There are 2 types of variables.

Categorical variable

Nominal Variable is a characteristic type of data. There is no order in possibilities; for example; person marriage status has possibilities from married, divorce, single or gender orientation; such as, male, female, or unidentified.

Ordinal Variable is also a characteristic type of data, however there is a chronological order in possibilities; such as order of customers has possibilities; which are, excellent, good, fair, poor or characteristic of weight that did not display in numerical term; such as; obese, overweight, and average, below average.

Quantitative variable

Quantitative variable which is to measure the level of differences between the possibilities.

Continuous; such as, income, ratio, and numerical amount; such as, examination rate of students or number of customers that purchased the product.

Discrete has a continuous characteristic; such as annual possibilities; month, week, season, and quarter.

3.2 Preprocessing

The purpose of Data preprocessing is to create confidence in quality of chosen data. That the data is suitable and completed as a guaranteed that data mining is successful. This process is a process where problems have been found comparing to preparation process. This is due to existing data in an organization has been not been prepare for Data mining. The data will be taken from various sources and has not been collected properly. Therefore the data will be taken from external source combined with the internal data. The main problem of Data is quality and Data integrity.

Before this process is completed, the structure need to be review and measure quality of data through statistical method or random sample from tools and screen the information as follow.

Categorical variable is the frequency division of method. It will result in the understanding in Data Content and the graphic tools will supporting and indicate the missing value.

Quantitative variable usually use the measurement; such as, highest value to lowest value, the mean, medium, and mode as well as other statistic values. After these values have been calculated it will result in incomplete value or problem value.

The obstacles for screening data process that often show is Noisy Data which is a type of data that exceeds the expectations. This could be advantage and disadvantage, for the advantage, it will clearly display the chance we are searching for. However, for the disadvantage, it may be incomplete information. The reason this happened may be because the carelessness of human; such as Operator

input age of person to 300 years or input income value as negative value. These values need to be revising or remove from the analysis. There should be process to investigate before the data has been taken to use. Another obstacle is Missing Value or value that has not been display or incomplete value that has been deleted during the Noise Detection process. This may leads to missing value due to carelessness of human.

3.3 Cleaning and Exploration

After data has been collected and screen, the next process is data cleaning. This process consist of 2 parts; the first one is analyst should be familiar with data not only the attribute or its meaning. But the analyst should know the content or real motivation of data. The second one is there may be mistake of data collection during the process of collecting data from various source to one larger source for further analysis. A good analyst should investigate these data, the example of mistake occurred in analysis was the mistake in data collection from unnecessary attribute. This result from the mislabeling of field of attribute; for example, we need to collect level of education for candidate. In reality, it has been collected in an attribute name "LEVEL_EDU", however, the database happened to contain another attribute name "EDUCATION" which collect level of education for candidate. If we did not properly investigate the relation and real motivation of each attribute. This may lead to confusion, the data collection of attribute "EDUCATION" instead. After the data has been taken to input in Data mining the result will be incorrect.

3.4 Transformation

Between the processes of Data transformation, the information that has been screened will be transformed to data that is ready to be analyzed. The form of data that is ready for analysis is a form of no oppose data which has been arrange properly and screen from an external and internal source.

This is an importance process because the accuracy and complete of the final result depend on how the analyst decided to specify structure and propose the input characteristic. The example is the screened data will be suitable for each process of Data mining. Moreover, the data transformation will also include Data Recording and Data Format Conversion; such as, date transformation.

According to the statistic, data transformation also consist of data reduction technique and the purpose to reduce the variable for the process is to take 2 or more variables to combine and complete the Process. The advantage is it can reduce the number of variable and can be easily managed. Another technique is called Discretization which means to transform Quantitative variable to Categorical variable. The dividing of variable value that will include in input to smaller section; such as, salary transformation or age, also there is another technique call One of N. This will transform variable from Categorical to Numeric; such as type of cars; Ford, Lincoln, Nissan to 100, 010, 001. Usually these are input of neural network.

3.5 Engineering

The process before is the creation and verification process, now we need to do data engineering. There will be 3 main issues; which are, database might consist of large number of attributes that is beneficial but has been neglected. The selection of attributes that need to use is also another important issue. Moreover, database might consist of exceeding number of record to be able complete analyze within suitable time period. IN this case, we have to random sample data to use instead and last, some data may be beneficial only when it has been proposed in the form of specific analysis. The making of Data engineering need to be repeated several time for test and the need to use different attribute, size of group sample; such as we would like to predict future after time has passed for 1, 2, 3, or 4 months only to use attribute for prediction. Or we can use every data that we possessed for prediction.

4) Modeling

It can be separated into 2 major types;

4.1 Predictive Data Mining is the prediction of characteristic or assumption of distinct value of the data by using the existed data.

4.2 Descriptive Data Mining is the searching for model to explain characteristic of existing data mainly it is a type of data division

This is because data mining is a knowledge selection from large database. Therefore, data mining can be compare to the combination of each knowledge sectors together; such as, Statistics, Machine Learning. The pattern of research is as follow.

1) Association Analysis

The relation analysis in the form of search for knowledge is also called 'Association Rules'. It shows the relation of correlate value that has conditions that is identical to regulations and characteristic of data that has been learned in the form of market basket or transaction. In the picture, generally, an association rules, if X and Y, and X is conditions that may cause Y and Y is result of calculation of X. For the analysis of association rule it needs to indicate support value and confidence value. This will be the indicator if the rule will be have certain relation with the value and it also help restrain unnecessary rule that is least related to the value.

2) Classification and Prediction

Classification is a process that help search for pattern of data that is similar and identical to use in predicting the category of data. The data group that has been classified originated from training data. Training data can be display in different ways; such as Classification rules, IF-THEN, Decision Tree, and Mathematical formula. The making of decision tree will display in the diagram of tree structure which the branch will be able to display the knowledge received. Whilst, leaves will display type of data that has been classified, also we can see that the prediction is about the numeric value and type of that data by looking at the tendency that may occur.

3) Clustering Analysis

Clustering analysis is different from classification & prediction that analyze data that is similar by classify without identify the name of group. Generally, these method can be used with manage the data that is uncertain about the category. Moreover, this analysis can also explain the name of classified group. In order to make this analysis it needs similar principle and least similarity. This is data that has been collected in the same group with possess high similarity which is differ from the data that has been separately classified.

5) Evaluation

It is a process to translate the meaning and evaluate the result whether the meaning is suitable with the needed purpose or not. Generally, it should be display in

the form that is easy to understand with the relation that has been found. It need to be tested for rate of incorrect and analysis the complex of pattern for the model. And if the incorrect still occur, it may need to improve the model to be more accurate. In the same sense, if the model is too complex that it is too difficult to understand. It is also necessary to find new model that contains same accuracy but less complex pattern.

2.4 Creation of Decision Tree prototype

It is one of the classification techniques which has characteristic as decision structure. Each branch will display type of experimented data and leaves will display group or class that has been specified. The tree prototype will consist of the root which has a characteristic to best separate the group. The data group will be separate until the classification has no significance in statistic or can no longer be classify into smaller groups. There are many processes to separate data and it is usually in the rule of IF-then; such as, “If income = High and Married = No THEN Risk = Poor”.

The principle algorithm of making decision tree is Greedy Algorithm. This will began from a recursive process with the method of Divided-and-Conquer. The pattern of tree will consist of starter node or root node. Root Node will be divided into small node and small node will have their kids' node. And the last node is leaf node which is the result of the process that can be display as follow.

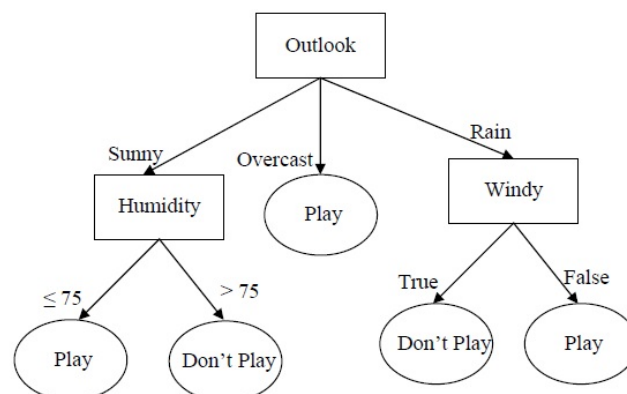


Figure 2.1 Sample of decision tree.

From the example, the result display as if the weather is Sunny and Humidity has less or equal value to 75, then the result from the condition is Play.

Decision Tree is a widespread technique this is because user can understand the result easily. Decision tree technique will limit the dependent variable 1 variable for 1 model. And if there is the need to predict various variable, the model need to be created according to each algorithm. Decision tree technique usually does not handle continuous data, there has to be a separation of discrete data. The example of algorithm; such as, Classification and Regression Trees (CART), ID3, C4.5, and C5.0 algorithm.

This research chose ID3 method from Quinlan JR. (8) for measure the value Entropy (Entropy-based measure) or what is known as Information Gain. The selection of the classification is information gain value or another is ID3. This is a process to separate data group to create decision tree. The measured value will be used to make decision regarding which variable is needed to separate the data. The method to decision tree specification is the selection of data from chronological order. The highest gain will be the beginning and the next data will be lessening, respectively. The principle of this exploration is as follow.

Assign P and N are data for separate

p is sample data in P

n is ample date in N

A is value of information gain in A, an equation is

$$\text{Gain}(A) = I(p,n) - E(A) \tag{2.1}$$

Where I(p,n) is information gain of set date that estimating sample data use amount of bit to separate P and N defined by

$$I(p,n) = -\frac{p}{p+n} \log_2 \left(\frac{p}{p+n} \right) - \frac{n}{n+p} \log_2 \left(\frac{n}{p+n} \right) \tag{2.2}$$

E(A) is Entropy which separate attribute A by A is separate from attribute S by {S1, S2, S3,..., Sv} by Si has sample data from data P available in Pi and N available Ni by equation.

$$E(A) = \sum_{i=1}^v \frac{p_i - n_i}{p+n} I(p_i, n_i) \tag{2.3}$$

In case of separated data has 2 number, information gain of group of data and estimate of data calculate from

$$I(S) = \sum_{i=1}^n -\frac{P_i}{S} \log_2 \left(\frac{P_i}{S} \right) \quad (2.4)$$

Where

S: amount of whole track in data

P: amount of track that divide by set of data.

n: amount of track that divide by sample of data

$$E(A) = \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} I(S_v) \quad (2.5)$$

Where

A: an attribute that want to estimate

V: value of attribute A

Table 2.1 Data of sample case which selling of air conditioner.

Age	Income	Working	Credit	Buy
<=30	High	No	Medium	No
<=30	High	No	High	No
31-39	High	No	Medium	Yes
>=40	Medium	No	Medium	Yes
>=40	Low	Yes	Medium	Yes
>=40	Low	Yes	High	No
31-39	Low	Yes	High	Yes
<=30	Medium	No	Medium	No
<=30	Low	Yes	Medium	Yes
>=40	Medium	Yes	Medium	Yes
<=30	Medium	Yes	High	Yes
31-39	Medium	No	High	Yes
31-39	High	Yes	Medium	Yes
>=40	Medium	No	High	No

Attributes is Age, Income, Working and Credit

Age consist of “ ≤ 30 ”, “31-39”, “ ≥ 40 ”

Income consist of

“High” mean income more than 30,000 baht per month

“Medium” mean income more than 10,000 - 30,000 baht per month

“Low” mean income not over 10,000 baht per month

Working consist of “Yes” and “No”

Credit consist of “Medium”, “High”

Set of data is Buying status consist of “Yes” and “No”

As decision tree method us information gain to define all attributes are Age, Income, Working and Credit. Then considering buying status of an air conditioner which “Yes” or “No” by step

Step 1

Find root node by considering attribute that has highest information gain value from sample data and calculate the highest information gain amount. It can draw as Fig 2.1

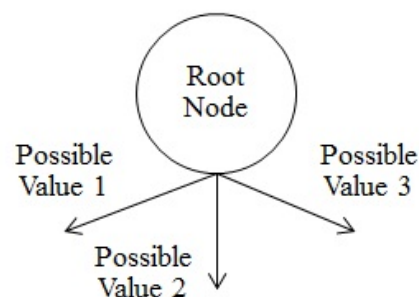


Figure 2.2 Root node of decision tree.

Step 2

Find attribute that second highest from root node by considering possible value of root node. In this case has 3 ways of possible value (3 limbs). So we consider each limb for separating next node. It can draw as Fig 2.2

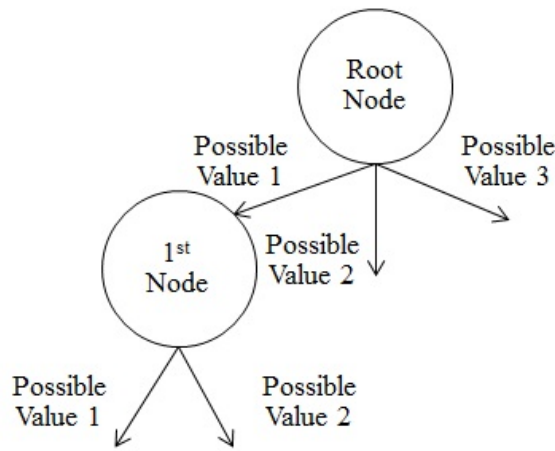


Figure 2.3 1st node of decision tree.

Step 3

Find next attribute with the same method of step 2 in rest limb and sub-limb. This is finishing of making decision tree

From an example data, we can build decision tree with this method and the detail of calculated are as follow.

From RAW data sample of selling air conditioners has 14 transactions. So we can calculate information gain of data set by

Information gain = $I(p,n) = I(9,5)$ By $P=9$ (yes), $n=5$ (no)

$$I(ALL)=I(9,5)=-\frac{9}{9+5} \log_2 \left(\frac{9}{9+5} \right) - \frac{5}{9+5} \log_2 \left(\frac{5}{9+5} \right) = 0.9403 \quad (2.6)$$

Next step is finding root node

Attribute “Age”

Table 2.2 Attribute “Age” Clarification

Age	Total	Yes	No
<=30	5	2	3
31-39	4	4	0
>=40	5	3	2

Then can calculate estimate by attribute “Age” by

$$E(\text{Age}) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} I(p_i, n_i) = \frac{5}{14} I(2,3) + \frac{4}{14} I(4,0) + \frac{5}{14} I(3,2) = 0.6935 \quad (2.7)$$

The equation of information gain by attribute “Age” is

$$\text{Gain}(\text{Age}) = I(\text{All}) - E(\text{Age}) = 0.9403 - 0.6935 = 0.2468 \quad (2.8)$$

Attribute “Income”

Table 2.3 Attribute “Income” Clarification

Income	Total	Yes	No
High	4	2	2
Medium	6	4	2
Low		3	1

Then can calculate estimate by attribute “Income” by

$$E(\text{Income}) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} I(p, n) = \frac{4}{14} I(2,2) + \frac{6}{14} I(4,2) + \frac{4}{14} I(3,1) = 0.9111 \quad (2.9)$$

The equation of information gain by attribute “Income” is

$$\text{Gain}(\text{Income}) = I(\text{All}) - E(\text{Income}) = 0.9403 - 0.9111 = 0.0292 \quad (2.10)$$

Attribute “Working”

Table 2.4 Attribute “Working” Clarification

Working	Total	Yes	No
Yes	7	6	1
No	7	3	4

Then can calculate estimate by attribute “Working” by

$$E(\text{Income}) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} I(p, n) = \frac{7}{14} I(6,1) + \frac{7}{14} I(3,4) = 0.7885 \quad (2.11)$$

The equation of information gain by attribute “Working” is

$$\text{Gain}(\text{Income}) = I(\text{All}) - E(\text{Working}) = 0.9403 - 0.7885 = 0.1518 \quad (2.12)$$

Attribute “Credit”

Table 2.5 Attribute “Credit” Clarification

Credit	Total	Yes	No
Medium	8	6	2
High	6	3	3

Then can calculate estimate by attribute “Credit” by

$$E(\text{Income}) = \sum_{i=1}^v \frac{p_i + n_i}{p + n} I(p, n) = \frac{8}{14} I(6, 2) + \frac{6}{14} I(3, 3) = 0.8922 \quad (2.13)$$

The equation of information gain by attribute “Credit” is

$$\text{Gain}(\text{Income}) = I(\text{All}) - E(\text{Credit}) = 0.9403 - 0.8922 = 0.0481 \quad (2.14)$$

For calculate aforesaid we can conclude information gain value as

From Eq.(2.8) $\text{Gain}(\text{Age}) = 0.2468$

From Eq.(2.10) $\text{Gain}(\text{Income}) = 0.0292$

From Eq.(2.12) $\text{Gain}(\text{Working}) = 0.1518$

From Eq.(2.14) $\text{Gain}(\text{Credit}) = 0.0481$

We can chose highest information gain. In this case is “Age” which root node. Then we calculate next attribute from root node by consider each of the rest attributes. The detail of calculation is follow.

Limb “Age<=30”

Information gain data of “Age<=30” can be calculated by $I(p, n) = I(2, 3)$ by $p = 2$ “Buy”, $n = 3$ “Not buy”

$$I(\text{Age} \leq 30) = I(2, 3) = -\frac{2}{2+3} \log_2 \left(\frac{2}{2+3} \right) - \frac{3}{2+3} \log_2 \left(\frac{3}{2+3} \right) = 0.9709 \quad (2.15)$$

Limb “Age<=30” by “Age”

Table 2.6 Limb “Age<=30” by “Age” Clarification.

Age	Total	Yes	No
<=30	5	2	3

We can find estimate of separated data by attributed “Age” by

$$E(\text{Age} \leq 30 \rightarrow \text{Age}) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} I(p,n) = \frac{5}{5} I(2,3) = 0.9709 \tag{2.16}$$

As a result, information gain of attribute “Age” is

$$\begin{aligned} \text{Gain}(\text{Age} \leq 30 \rightarrow \text{Age}) &= I(\text{Age} \leq 30) - E(\text{Age} \leq 30 \rightarrow \text{Age}) \\ &= 0.9709 - 0.9709 = 0 \end{aligned} \tag{2.17}$$

Limb “Age<=30” by “Income”

Table 2.7 Limb “Age<=30” by “Income” Clarification.

Income	Total	Yes	No
High	2	0	2
Medium	2	1	1
Low	1	1	0

We can find estimate of separated data by attributed “Income” by

$$\begin{aligned} E(\text{Age} \leq 30 \rightarrow \text{Income}) &= \\ \sum_{i=1}^v \frac{p_i + n_i}{p+n} I(p,n) &= \frac{2}{5} I(0,2) + \frac{2}{5} I(1,1) + \frac{1}{5} I(1,0) = 0.4000 \end{aligned} \tag{2.18}$$

As a result, information gain of attribute “Income” is

$$\begin{aligned} \text{Gain}(\text{Age} \leq 30 \rightarrow \text{Income}) &= I(\text{Age} \leq 30) - E(\text{Age} \leq 30 \rightarrow \text{Income}) \\ &= 0.9709 - 0.4000 = 0.5709 \end{aligned} \tag{2.19}$$

Limb “Age<=30” by “Working”

Table 2.8 Limb “Age<=30” by “Working” Clarification.

Working	Total	Yes	No
Yes	2	2	0
No	3	0	3

We can find estimate of separated data by attributed “Working” by

$$E(\text{Age} \leq 30 \rightarrow \text{Working}) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} I(p,n) = \frac{2}{5} I(2,0) + \frac{3}{5} I(0,3) = 0 \quad (2.20)$$

As a result, information gain of attribute “Working” is

$$\begin{aligned} \text{Gain}(\text{Age} \leq 30 \rightarrow \text{Working}) &= I(\text{Age} \leq 30) - E(\text{Age} \leq 30 \rightarrow \text{Working}) \\ &= 0.9709 - 0 = 0.9709 \end{aligned} \quad (2.21)$$

Limb “Age ≤ 30” by “Credit”

Table 2.9 Limb “Age ≤ 30” by “Credit” Clarification.

Credit	Total	Yes	No
Medium	3	1	2
High	2	1	1

We can find estimate of separated data by attributed “Credit” by

$$E(\text{Age} \leq 30 \rightarrow \text{Credit}) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} I(p,n) = \frac{3}{5} I(1,2) + \frac{2}{5} I(1,1) = 0.9510 \quad (2.22)$$

As a result, information gain of attribute “Credit” is

$$\begin{aligned} \text{Gain}(\text{Age} \leq 30 \rightarrow \text{Credit}) &= I(\text{Age} \leq 30) - E(\text{Age} \leq 30 \rightarrow \text{Credit}) \\ &= 0.9709 - 0.9510 = 0.0199 \end{aligned} \quad (2.23)$$

Conclude of information gain value by limb “Age ≤ 30” is

$$\text{From Eq. (2.17) } \text{Gain}(\text{Age} \leq 30 \rightarrow \text{Age}) = 0$$

$$\text{From Eq. (2.19) } \text{Gain}(\text{Age} \leq 30 \rightarrow \text{Income}) = 0.5709$$

$$\text{From Eq. (2.21) } \text{Gain}(\text{Age} \leq 30 \rightarrow \text{Working}) = 0.9709$$

$$\text{From Eq. (2.23) } \text{Gain}(\text{Age} \leq 30 \rightarrow \text{Credit}) = 0.0199$$

As a result, information gain of “Age ≤ 30” by “Working” is highest value. Then we use limb “Working” to calculate in next step

Limb “Age31-39”

Information gain data of “Age=31-39” can be calculated by $I(p,n) = I(4,0)$ by $p = 4$ “Buy”, $n = 0$ “Not buy”

$$I(\text{Age}=31-40) = I(4,0) = -\frac{4}{4+0} \log_2 \left(\frac{4}{4+0} \right) - \log_2 \left(\frac{0}{4+0} \right) = 1 \quad (2.24)$$

According to data it show all of transaction is “Buy”. It mean Limb “Age 31-39” can’t spate to sub limb.

Limb “Age>=40” by “Age”

Table 2.10 Limb “Age>=40” by “Age” Clarification.

Age	Total	Yes	No
>=40	5	3	2

We can find estimate of separated data by attributed “Age” by

$$E(\text{Age} \geq 40 \rightarrow \text{Age}) = \sum_{i=1}^v \frac{p_i + n_i}{p + n} I(p,n) = \frac{5}{5} I(3,2) = 0.9709 \quad (2.25)$$

As a result, information gain of attribute “Age” is

$$\begin{aligned} \text{Gain}(\text{Age} >= 40 \rightarrow \text{Age}) &= I(\text{Age} >= 40) - E(\text{Age} >= 40 \rightarrow \text{Age}) \\ &= 0.9709 - 0.9709 = 0 \end{aligned} \quad (2.26)$$

Limb “Age>=40” by “Income”

Table 2.11 Limb “Age>=40” by “Income” Clarification.

Income	Total	Yes	No
High	0	0	0
Medium	3	2	1
Low	2	1	1

We can find estimate of separated data by attributed “Income” by

$$\begin{aligned} E(\text{Age} \geq 40 \rightarrow \text{Income}) &= \\ \sum_{i=1}^v \frac{p_i + n_i}{p + n} I(p,n) &= \frac{0}{5} I(0,0) + \frac{3}{5} I(2,1) + \frac{2}{5} I(1,1) = 0.9510 \end{aligned} \quad (2.27)$$

As a result, information gain of attribute “Age” is

$$\begin{aligned} \text{Gain}(\text{Age} \geq 40 \rightarrow \text{Income}) &= I(\text{Age} \geq 40) - E(\text{Age} \geq 40 \rightarrow \text{Income}) \\ &= 0.9709 - 0.9510 = 0.0199 \end{aligned} \quad (2.28)$$

Limb $\text{Age} \geq 40$ by “Working”

Table 2.12 Limb “Age ≥ 40 ” by “Working” Clarification.

Income	Total	Yes	No
Yes	3	2	1
No	2	1	1

We can find estimate of separated data by attributed “Working” by

$$\begin{aligned} E(\text{Age} \geq 40 \rightarrow \text{Working}) &= \\ \sum_{i=1}^v \frac{p_i + n_i}{p + n} I(p, n) &= \frac{3}{5} I(2, 1) + \frac{2}{5} I(1, 1) = 0.9510 \end{aligned} \quad (2.29)$$

As a result, information gain of attribute “Working” is

$$\begin{aligned} \text{Gain}(\text{Age} \geq 40 \rightarrow \text{Working}) &= I(\text{Age} \geq 40) - E(\text{Age} \geq 40 \rightarrow \text{Working}) \\ &= 0.9709 - 0.9510 = 0.0199 \end{aligned} \quad (2.30)$$

Limb “Age ≥ 40 ” by “Credit”

Table 2.13 Limb “Age ≥ 40 ” by “Credit” Clarification.

Income	Total	Yes	No
Medium	3	3	0
High	2	0	2

We can find estimate of separated data by attributed “Credit” by

$$\begin{aligned} E(\text{Age} \geq 40 \rightarrow \text{Credit}) &= \\ \sum_{i=1}^v \frac{p_i + n_i}{p + n} I(p, n) &= \frac{3}{5} I(3, 0) + \frac{2}{5} I(0, 2) = 0 \end{aligned} \quad (2.31)$$

As a result, information gain of attribute “Credit” is

$$\begin{aligned} \text{Gain}(\text{Age} \geq 40 \rightarrow \text{Credit}) &= I(\text{Age} \geq 40) - E(\text{Age} \geq 40 \rightarrow \text{Credit}) \\ &= 0.9709 - 0 = 0.9709 \end{aligned} \quad (2.32)$$

Conclude of information gain value by climb “Age \geq 40” is

From Eq. (2.36) $\text{Gain}(\text{Age}\geq 40 \rightarrow \text{Age}) = 0$

From Eq. (2.38) $\text{Gain}(\text{Age}\geq 40 \rightarrow \text{Income}) = 0.0199$

From Eq. (2.30) $\text{Gain}(\text{Age}\geq 40 \rightarrow \text{Working}) = 0.0199$

From Eq. (2.32) $\text{Gain}(\text{Age}\geq 40 \rightarrow \text{Credit}) = 0.9709$

As a result, information gain of “Age \geq 40” by “Credit” is highest value.

Then we use limb “Credit” to calculate in nest step

Follow result of calculating from sample data that selling of air conditioner can draw decision tree as follow

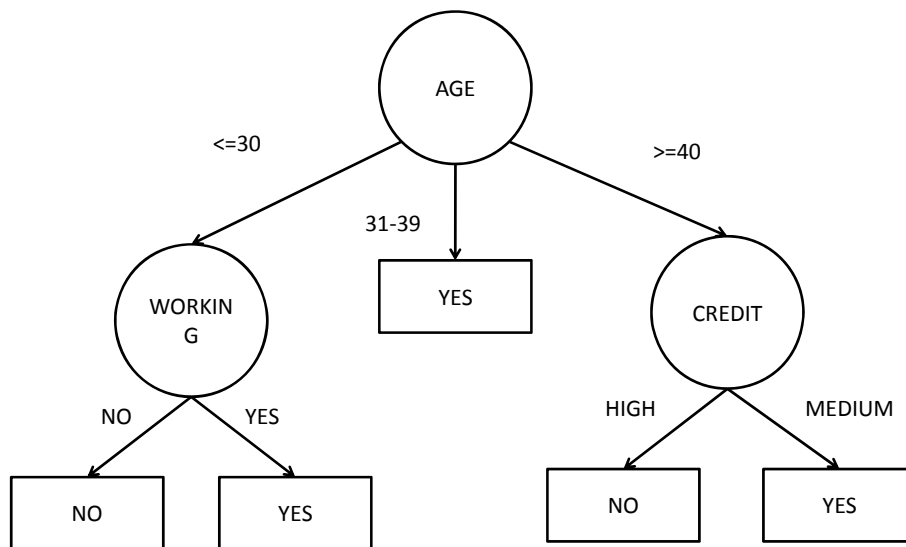


Figure 2.4 Completed decision tree from sample data.

From graph we can define the quotation in platform of “If-Then” is as follow.

IF Age = “<=30” and Working = “No” THEN Buy = “No”

IF Age = “<=30” and Working = “Yes” THEN Buy = “Yes”

IF Age = “<=30” and Working = “Yes” THEN Buy = “Yes”

IF Age = “31-39” THEN Buy = “Yes”

IF Age = “>=40” and Credit = “High” THEN Buy = “No”

IF Age = “>=40” and Credit = “Medium” THEN Buy = “Yes”

2.5 Algorithm C4.5

The C4.5 process is the extension of ID3 method which has been developed by Quinlan JR. (9). The creation of decision tree for the decision in use of classification of data needs to use gain value and Entropy data prediction similar to ID3. However, there is an additional method from ID3 which is it can be use with both continuous characteristic and discrete type. For the continuous data, C4.5 method will create Threshold and separate the characteristic into 2 parts; the higher value or lower value and equivalent to the beginning value. It also can avoid the creation of decision tree that is too large due to large number of data and it depends on the depth of data.

2.6 Instruction to program “WEKA” for build decision tree

This research use program “Weka” version 3.6 for build decision tree and step of using can explain as follow.

Record all data in format CSV, An example of all data is

Table 2.14 Sample data for import to program “WEKA”

Outlook	temperature	humidity	windy	Play
sunny	85	85	FALSE	no
sunny	80	90	TRUE	no
overcast	83	86	FALSE	yes
rainy	70	96	FALSE	yes
rainy	68	80	FALSE	yes
rainy	65	70	TRUE	no
overcast	64	65	TRUE	yes
sunny	72	95	FALSE	no
sunny	69	70	FALSE	yes
rainy	75	80	FALSE	yes
sunny	75	70	TRUE	yes
overcast	72	90	TRUE	yes
overcast	81	75	FALSE	yes
rainy	71	91	TRUE	No

When open program choose “Explorer” for start build decision tree.

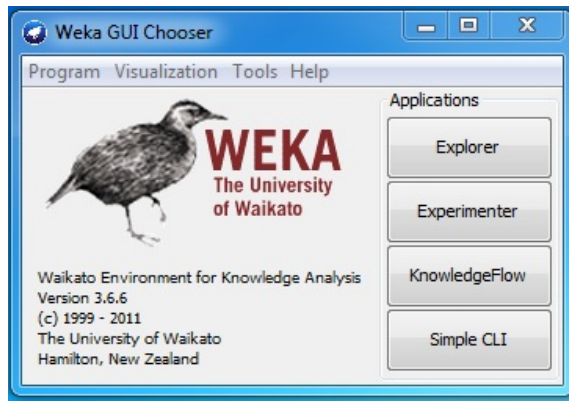


Figure 2.5 Welcome interface of program “WEKA”

In Weka Explorer tab “Preprocess” choose “Open file...” for choose CSV file in aforesaid step

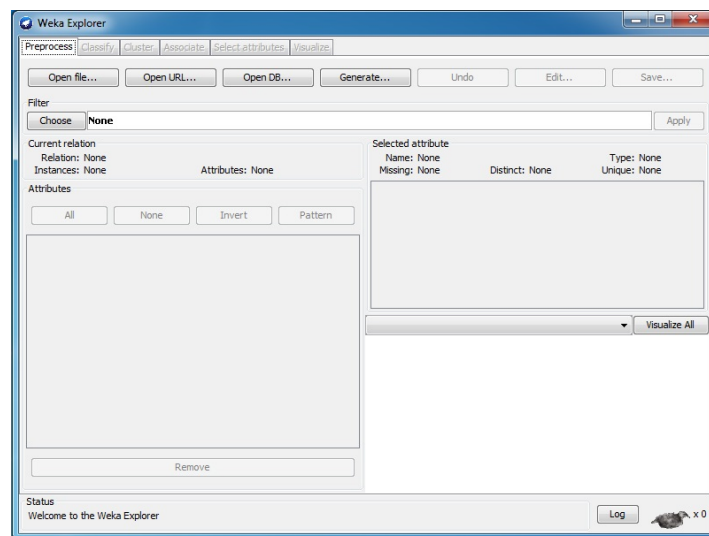


Figure 2.6 Weka explorer window.

After that, program will show detail of each attributes and type of attributes. For instance, Attribute: Outlook has “Nominal” type and has 5 records of “Sunny”, 4 records of “Overcast” and 5 records of “Rainy” etc.

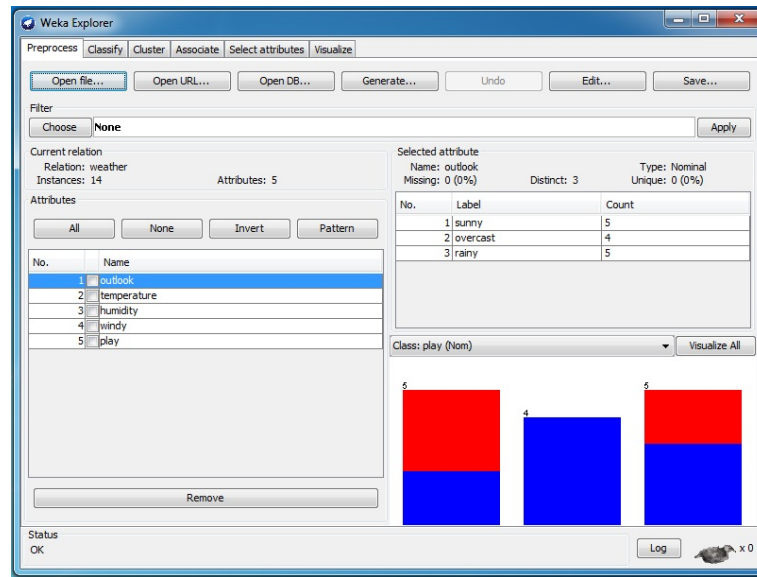


Figure 2.7 After import data to “WEKA”

Look at attribute “Temperature” has “Numeric” type and has minimum value is “64” and maximum value is “85”. So in sample data has attribute type is which gathering of “Nominal” and “Numeric”, researcher choose algorithm C4.5 which can calculate and build decision tree with data like this but algorithm ID3 data must be “Nominal” only.

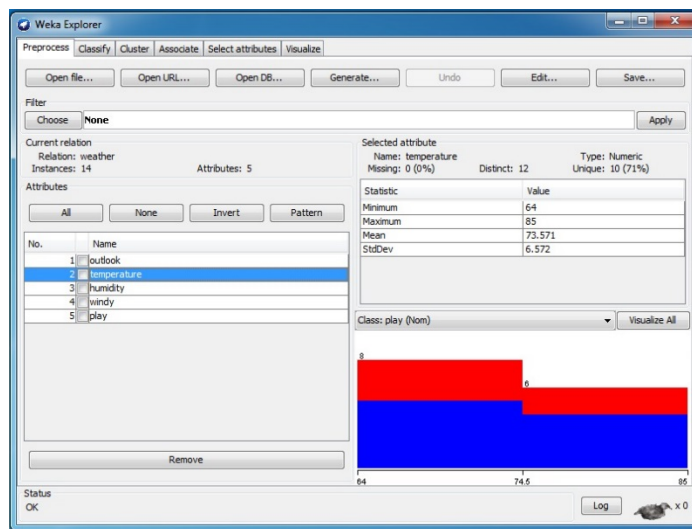


Figure 2.8 Sample of numeric data in “WEKA”

Then choose tab “Classify” and choose method C4.5 (in program “Weka” will show “J48” as the same). Next choose “Use training set” in part of “Test options”

then choose estimate attribute by “Play”. Then “Start” and program will calculate and show result in part of “Classifier output” as

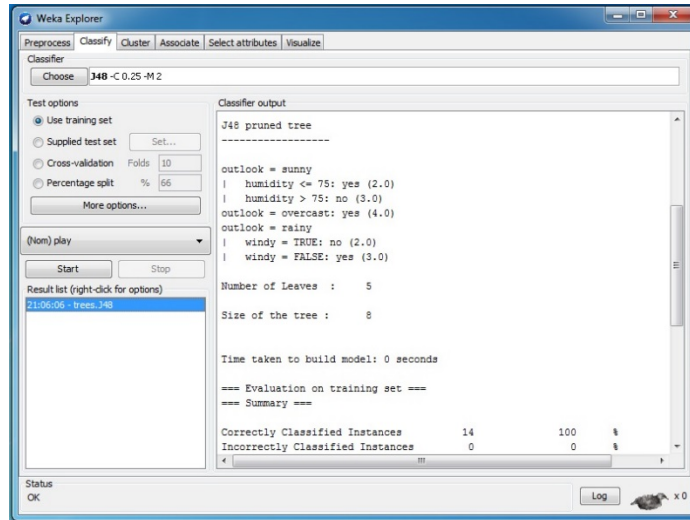


Figure 2.9 Calculated result from “WEKA”

We can explain result as

Calculated Decision Tree is as follow

outlook = sunny
 / *humidity <= 75: yes (2.0)*
 / *humidity > 75: no (3.0)*
outlook = overcast: yes (4.0)
outlook = rainy
 / *windy = TRUE: no (2.0)*
 / *windy = FALSE: yes (3.0)*

Correctly of calculated decision tree is 100% accuracy

<i>Correctly Classified Instances</i>	<i>14</i>	<i>100</i>	<i>%</i>
<i>Incorrectly Classified Instances</i>	<i>0</i>	<i>0</i>	<i>%</i>

We can draw decision tree as follow

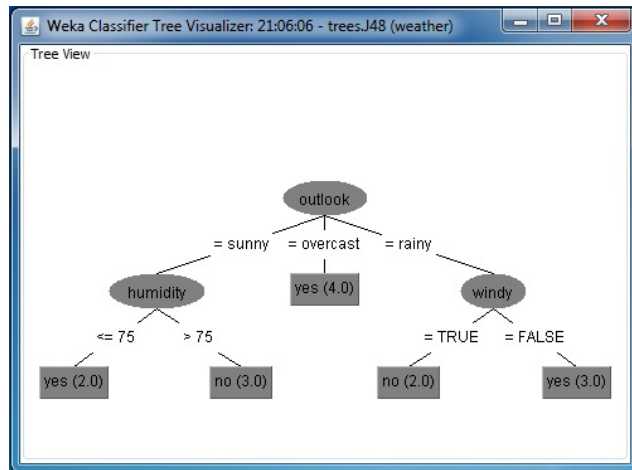


Figure 2.10 Decision tree from “WEKA”

CHAPTER III

ANALYSIS AND DESIGN THE SYSTEM

3.1 Methodology

According to principle of Data Mining process and historical data of Sample Company can explain the methodology of this research. The first step is understanding purpose of Decision support system and data of Sample Company. Then prepare data aforesaid in the pattern that can import to Weka program. Next select factors from all factors by build questionnaire that include all factors of data and send to expert of Sample Company who exclusive managers and union all of factors that checked in questionnaire. Use selected factors to build decision tree by C4.5 algorithm by Weka program.

Convert result from Weka program that in term of “IF THEN” and store in text file, to an excel spread sheet. Then create interface of Decision support system from Userform function in Microsoft officer excel program. Next coding calculates data from interface and conditions of result from Weka program for decision support system can automatically calculate. So decision support system is created.

Estimation of system come from accuracy rate between result condition from training set test with testing set data. Moreover researcher random some records from training set and build questionnaire. Then questionnaire will complete by experts of Sample. Next compute correlation between result from experts and decision support system.

Finally, researcher create query of the system to accept satisfaction of decision support system by divide questioner to 2 parts. The first is users in operating segment and the second is users in management segment and conclude the result from questionnaire. The whole step which explained can show as.

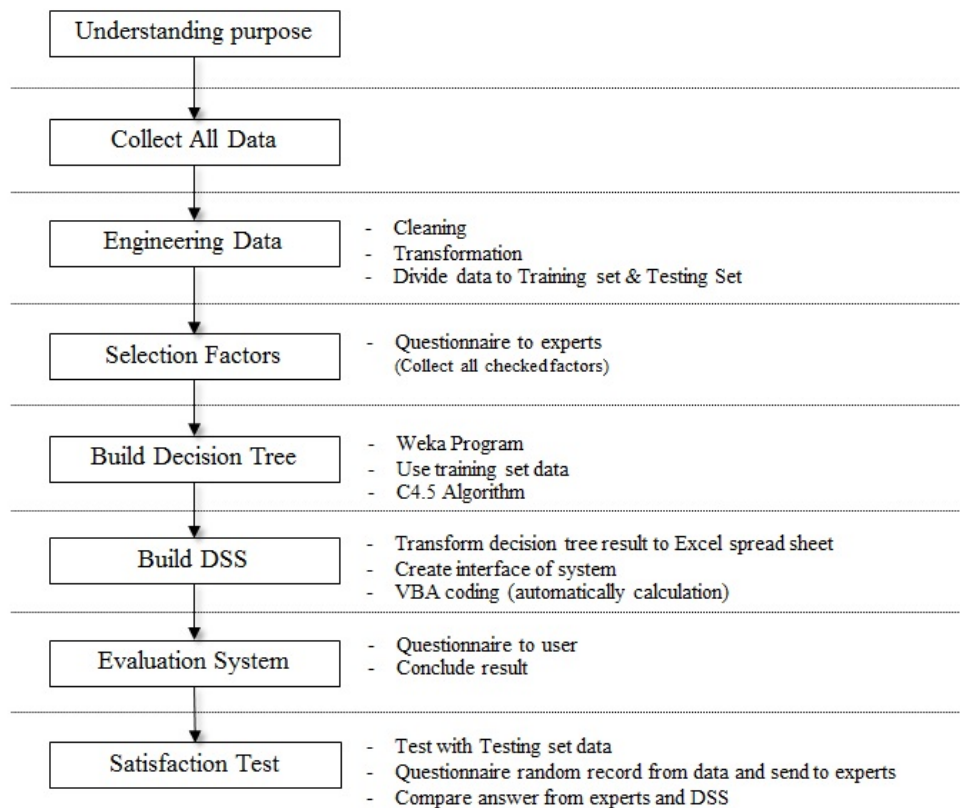


Figure 3.1 Research methodology

3.2 Determine objectives

The purpose of the making of decision support system is to create decision tree to select the return rate that is most appropriate for each contract. In the original system of the company, it is rather slow in operation and Quotation until the quotation has been approved by Flow chart of the company internal work. It can be display as follow.

V. Customer Sign is the waiting process while the customer's company makes decision to use leasing contract from the invoice.

In case that the customer's company decided to use leasing according to the invoice received, it will be signed by the management department at the customer's company and it will return back to the sampled company.

In case that the customer's company decided not to sign the contract or request a revise in the invoice, the invoice will not be sign and will be send back to Marketing Officer. So the Marketing Officer will analyze the customer and evaluate new price in the process of First Analysis. In case of change of return rate of contract, it will go back in the original flow which is invoice and return it back to the customer.

VI. In case the invoice has been signed, the document will be send back to the process of Check Authorize, which the contract department of the company will verify the correctness of the document and invoice and other documents that is related to the customer's company.

If the document is incorrect, will be send back to the Quotation Maker process again to request a new invoice under the same condition.

If the document passes the verification, it will be forward to the State.

VII. Approval is the company approves processes which consist of the making of Application and request document regarding company and the making of the contract within the internal system. The requesting of contract and related company document according to the internal system as well as the approve process with the company Authorized person.

If the authorized person signs the contract, the process of approval for leasing contract is completed.

If the authorized person does not sign the contract, all the document including contract and application as well as internal system of the company will be send back to Revise Condition State.

VIII. Revise Condition State is the process to revise the condition of leasing contract that has been denied from the authorized person of the company. The marketing officer will revise the invoice with new condition and bring it to the original

flow again which is to request invoice n Quotation maker state with the revised condition. The document will be send to the customer with the original flow until he contract has been approved.

From the information mentioned, the process takes a long time to revise the document as well as internal system process and approval power from the sampled company. The making of decision support system will be a system that has been insert which the system will be after the step of First Analysis before the quotation is completed.

This means that after the analysis and evaluation of customer credit is finalized, the customer’s company will pass the first evaluation. The marketing officer will take data from requirements evaluation that customer needed and input into the decision support system and it will be display as show in the diagram.

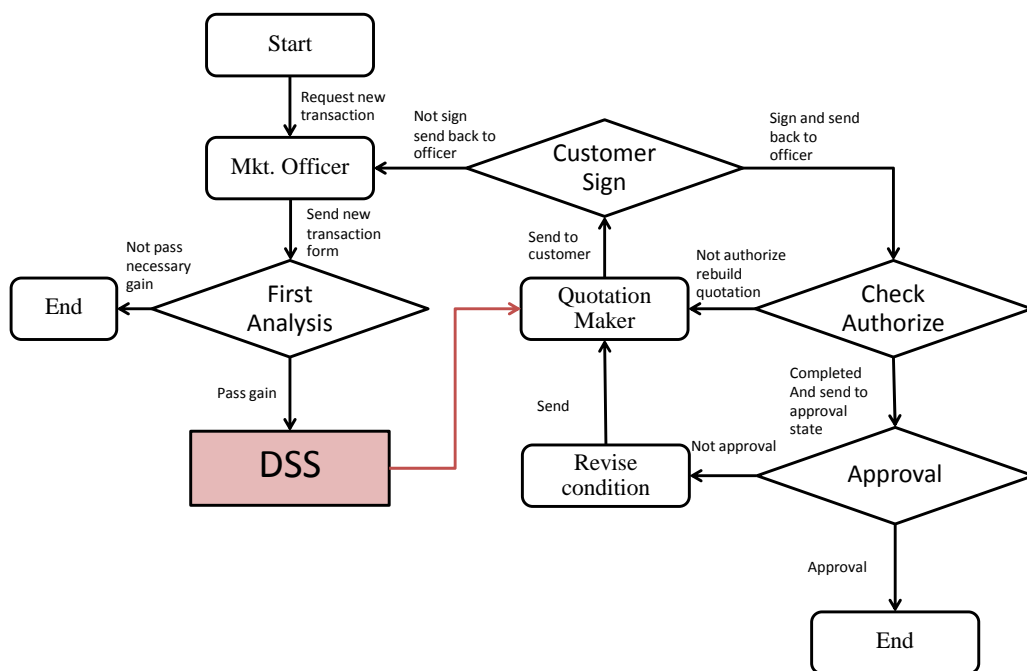


Figure 3.3 Workflow after included DSS.

After the evaluation of Customer Company is completed, normally the Marketing Officer will evaluate return rate of customer contract with their experiences and speculate. However, the data from the evaluation will enter the decision support

system and it will evaluate the suitable return rate with the decision tree method of Algorithm C4.5. The previous data from the company will be use to evaluate return rate that is most appropriate for the contract and the return rate will be calculate for further invoice.

Thus, decision support system will calculate the interest rate under a certain standard that when a similar ype of customer with similar type of request can use the similar type of return rate. This is because the return rate already has other customer from other company with a similar characteristic, therefore; it is acceptable to use the return rate and contract. Moreover, the return rate has already been approved from the company.

The decision support system can display the process of calculation for interest rate as show in the diagram.

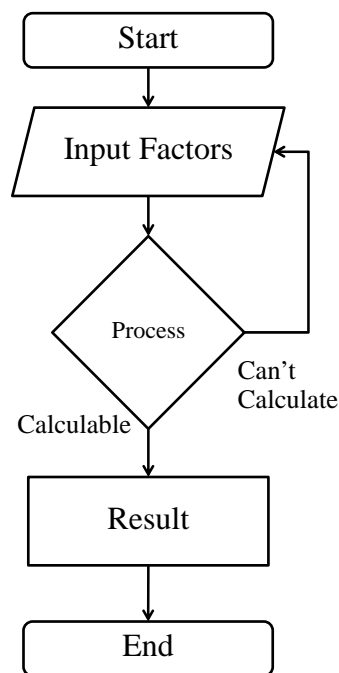


Figure 3.4 Flow chart of DSS.

3.3 Data understanding

The data that has been used to research is an internal data from the sampled company. It has a characteristic whereas name and information of the company has been disclosed. The data has been collected in the Data Centre of the company and each database will consist of 2 databases; which are,

- a. Customer database which consist of details of customers; such as, corporate name, type of business, rating of the company.
- b. Contract database which will consist of characteristic of each type of contract; such as, type of contract, value of asset, and return rate of that certain contract.

3.4 Data preparation

3.4.1 Data selection

The researcher has use 2 databases that have field that connect between data and data base which is contract data base that is in the form of Soft file xls (Microsoft Excel). The researcher has been collected altogether 4,503 Files to be within the same database. The data from each File will be combining within the Spreadsheet. The method for the making of code to take the data by separate into each factor is Visual Basic for Application; such as, Code sample that use to separate data has the characteristic as follow.

```

Sub operate()
On Error Resume Next
Dim t As Date
t = Now()
Application.ScreenUpdating = False
Application.Calculation = xlCalculationManual
Dim allcontract As Double
allcontract = Application.WorksheetFunction.CountA(Range("A:A")) - 1
For i = 1 To allcontract
'For i = 1 To 20
Workbooks.Open Range("C2").Offset(i - 1, 0), False, True
Workbooks("All Data").Sheets("RAW").Range("D2").Offset(i - 1, 0) = Range("B8").Value
Workbooks("All Data").Sheets("RAW").Range("E2").Offset(i - 1, 0) = Range("CQ2").Value
Workbooks("All Data").Sheets("RAW").Range("F2").Offset(i - 1, 0) = Range("CQ1").Value
Workbooks("All Data").Sheets("RAW").Range("G2").Offset(i - 1, 0) = Range("CQ3").Value
Workbooks("All Data").Sheets("RAW").Range("H2").Offset(i - 1, 0) = Range("CQ4").Value
Workbooks("All Data").Sheets("RAW").Range("I2").Offset(i - 1, 0) = Range("A28").Value
Workbooks("All Data").Sheets("RAW").Range("J2").Offset(i - 1, 0) = Range("CQ10").Value
Workbooks("All Data").Sheets("RAW").Range("K2").Offset(i - 1, 0) = Range("CQ12").Value
Workbooks("All Data").Sheets("RAW").Range("L2").Offset(i - 1, 0) = Range("A212").Value
Workbooks("All Data").Sheets("RAW").Range("M2").Offset(i - 1, 0) = Range("B12").Value
If Range("A35") = "Total" Then
Workbooks("All Data").Sheets("RAW").Range("N2").Offset(i - 1, 0) = Range("AV35").Value
Workbooks("All Data").Sheets("RAW").Range("O2").Offset(i - 1, 0) = Range("BL35").Value
Workbooks("All Data").Sheets("RAW").Range("P2").Offset(i - 1, 0) = Range("BX35").Value
Workbooks("All Data").Sheets("RAW").Range("Q2").Offset(i - 1, 0) = Range("A37").Value
Workbooks("All Data").Sheets("RAW").Range("R2").Offset(i - 1, 0) = Range("J37").Value
Workbooks("All Data").Sheets("RAW").Range("S2").Offset(i - 1, 0) = Range("S37").Value
Workbooks("All Data").Sheets("RAW").Range("T2").Offset(i - 1, 0) = Range("AB37").Value
Workbooks("All Data").Sheets("RAW").Range("U2").Offset(i - 1, 0) = Range("AP37").Value
Workbooks("All Data").Sheets("RAW").Range("V2").Offset(i - 1, 0) = Range("AV37").Value
Workbooks("All Data").Sheets("RAW").Range("W2").Offset(i - 1, 0) = Range("BH37").Value
Workbooks("All Data").Sheets("RAW").Range("X2").Offset(i - 1, 0) = Range("BP37").Value
Workbooks("All Data").Sheets("RAW").Range("Y2").Offset(i - 1, 0) = Range("CF37").Value
Workbooks("All Data").Sheets("RAW").Range("Z2").Offset(i - 1, 0) = Range("CG37").Value
Workbooks("All Data").Sheets("RAW").Range("AA2").Offset(i - 1, 0) = Range("BN51").Value
Else
Workbooks("All Data").Sheets("RAW").Range("N2").Offset(i - 1, 0) = Range("AV25").Value
Workbooks("All Data").Sheets("RAW").Range("O2").Offset(i - 1, 0) = Range("BL25").Value

```

Figure 3.5 Sample of importing data VBA code

After Run code is completed, we will receive all the data from File contract which is under the same Spreadsheet. Next, the researcher will take the data to combine with the 1 more related data base which is customer characteristic database by using the factor as customer name to connect between two databases

The process is completed we will get the raw data which is File Excel for the total of 1 file name Data to use in the next process of analysis as show in the diagram.

Figure 3.6 Sample RAW data

The factor from combine the data is the data with the characteristic both Nominal Variable and Quantitative which depends on each factor. The factors that related are as follow.

App Date	Insurance	Rating
Approval	Net Profit	Register Capital
Business	NPV	Rental
Business Area	Parent Net Profit	RV
Completed Documents	Parent Rate	Sale & Lease Back
Contract Type	Type	Sales
Cost	Parent Sales	Scheme
Deposit	Payment Method	Spread
Due Date	Percentage Owner	Status
Eff. Rate	Person in charge	Term
Funding	Privilege	

The researcher has selected the related factor with the calculation of return rate for each contract. The unrelated factor has been removed and the researcher has used interview method from expert and experienced person by using interview regarding factor that use to calculate basic interest rate. The experts are as follow.

- Manager of operation and credit department
- Manager of marketing and promotion department
- General Manager
- Assist manager of operation and credit department
- Team lead of marketing promotion officer

From the observation and data collection from interview, the person has been asked to arrange data mention and the question use for interview is “which factor affects the evaluation of return rate of each contract. The result is as follow.

Table 3.1 Result of observation factor interview.

Factors	No. 1	No. 2	No. 3	No. 4	No. 5
Approval	X		X		
Business		X	X	X	
Business Area					

Table 3.1 Result of observation factor interview.(cont.)

Factors	No. 1	No. 2	No. 3	No. 4	No. 5
Completed Documents					
Contract Type	X	X	X	X	X
Cost	X	X	X	X	X
Deposit	X	X	X	X	X
Due Date					
Eff. Rate	X	X	X	X	X
Funding	X	X	X	X	X
Insurance	X				
Net Profit					
NPV					
Parent Net Profit					
Parent Rate	X	X		X	
Parent Rate					
Parent Sales					
Payment Method					
Percentage Owner					
Person in charge					
Privilege	X	X	X	X	X
Rating	X	X	X	X	X
Register Capital					
Rental					
RV	X	X	X		
Sale & Lease Back	X		X	X	X
Sales					
Scheme		X			
Status	X	X	X	X	
Term		X	X		X
Type	X	X	X	X	X

After the result has been concluded, we will removed the factor with no expert out and the factor that is the result of return rate (EFF. RATE and FUNDING factor which is factors that result from return rate of the company. This means the return rate of the company or Spread which is equal to EFF.RATE minus FUNDING). After the factor has been removed, it remains the factor to evaluate the basic return rate as follow.

Approval	- Authorized level of each transaction
Business	- Business type of customer
Contract Type	- Leasing type of each transaction
Cost	- Cost of asset which leased of each transaction
Deposit	- Amount of guarantee of each transaction
Insurance	- Condition of insurer of each transaction
Parent Rate	- Rating of parent's customer company
Privilege	- Special condition by government policy
Rating	- Credit rating of Customer Company
RV	- Amount of residual value of each transaction
Sale & Lease Back	- Leasing for purchased asset optional
Scheme	- Scheme of installment of each transaction
Status	- Customer status of Sample Company
Term	- Term of each transaction
Type	- Type of customer divided by nationality

The researcher has use the mentioned factor to create system to support the decision making in the next process.

3.4.2 Data Processing

The researcher has selected the data that has blank value or incomplete value. The review of data has been completed such as Term factor should not have value more than 120 months or 10 years. This is because the company does not allow the contract; therefore, the contract with the term more than 120 months may result in error data record. So, it need to be change or removed from the data base. And the database that is blank value or not precise with Categorical of the factor will help reduce the experiment to be incorrect.

3.4.3 Data Cleaning and Exploration

This process the researcher has review each factor in details to remove the error value from recording or mistaking in recording data. Also, the data should be adjust to suit the possibility of each factor; such as Type factor has the possibility only 2 words; which is Japanese or Non-Japanese. However, the analyst found that after it has been count the number of value of Japanese altogether with Non-Japanese, the result is uneven. So the research removed the redundant value of the factor and found that the unrepeated word of the factor is not only Japanese and Non-Japanese. However, there is also an error variable such as “Non – Japanese” (space between “-“and letter or “Japaness” (misspell). The value mentioned will need to be transform within the possibilities of the value of the certain factor before it has been use to further evaluate.

3.4.4 Data Transformation

After the data cleaning and exploration, the next process is data transformation to make it in an appropriate form for the making of Data mining. This means that this study the researcher has used the method to create decision tree with Cd.5 method. This has been developed from ID3 method and the different factors that have been used to calculate can be in both Categorical and Quantitative. The researcher can transformed the data with the details as follow.

After the transformation of the value, we will receive new data table structure to be put into Weka program to evaluate. All the data will be collected into CSV (Comma-separated Value) file which can be calculated with Weka program. The method C4.5 by the structure of data after it has been transform consists of details as follow.

Contract Type is type of each contract has possible value start from “A” to “F” and detail of meaning as follow

- A : Lease (Financial lease but has scheme similar operating lease)
- B : Leasing (Financial Lease)
- C : Leasing with Undertaking (Financial lease with undertaking in extra)

- D : Hire Purchase
- E : Operating Lease
- F : Operating Lease with Undertaking (Operating lease with undertaking in extra)

Approval is level of approval state has possible value is “P” and “PP” and detail of meaning as follow

- P : President in Thailand
- PP : President in Japan

Status is status of customer by sample company has possible value is “Existing”, “New” and “Return” and detail of meaning as follow

- Existing : The customer has already been a customer of sample company.
- New : The customer hasn't had any contract with sample company before.
- Return : Then customer use to have contract with sample company and hasn't had any contract for now.

Parent Rating : Rating of parent company has possible value start from 1 to 8. The data is quantitative type and has detail as follow.

- 1 is mean the best
- 8 is mean the worst

Rating : rating of customer company has possible value start from 1 to 10. The data is quantitative type and has detail as follow.

- 1 is mean the best
- 10 is mean the worst

Type : type of customer company divided by nationality and has possible value is “Japan” and “Thai”. The detail meaning as follow

- Japanese is company that has parent company located in Japan (More than 50% direct or indirect own by parent company)
- Thai is company that has more than 50% major shareholder are Thai.

Business : business type of customer company has possible value is “Construction, Manufacturing and Other” and detail of meaning as follow.

Con is company which main business in construction
 Man is company which main business in manufacturing
 Other is except in group of construction and manufacturing

Cost : value of cost of asset in each contract with Thai baht currency. The data is quantitative start form 0.01 baht.

RV or Residual Value: amount of value of asset when end contract. We can calculate by RV divide Cost and possible value will be 0% to 100% in quantitative type of data.

Term : Amount of total time in each contract has quantitative type start from 1 to 120 months.

Deposit : amount of cash that customer transferred to sample company in start term of lease can calculate by amount of deposit divide cost. So the possible value is 0 to 1 or 1% to 100% in quantitative type.

Scheme : calculate installment of each contract has possible value is “Beginning” and “Ending”. The detail of meaning is follow.

Beginning is contract that customer must pay installment at the first date of contract and when customer pay last installment. The customer has right to use asset until maturity date.

Ending is contract that customer pay first installment after use the asset 1 month and when maturity date of contract the customer must pay the last installment in the same date.

Sale & Lease Back : type of contract which customer has already purchased asset then sale back to sample company and lease back again. So it has possible value to “True” and “False” as

True is Sale & Lease Back contract

False isn't Sale & Lease Back contract

Privilege : the special condition that government provide to customer company has possible value is “None”, “BOI” and “EPZ” as

None is not have special condition from government

BOI is have special condition form board of investment of Thailand in part of import machinery

EPZ is getting exclusive access to the location of the property in promotion area

Insurance is insurance policy of each contract has possible value to “Company” or “Customer” as

Com is sample company will pay insurance fee for asset in each contract.

Cus is customer company will responsibility in insurance of asset.

Spread is the surplus between contract rating and funding rate of each contract. So it is prediction attribute of data. However the type of data is quantitative that can't calculate to decision tree with algorithm C4.5 but researcher divide possible value of this attribute by 1 to 7 class interval and detail of each interval are as follow

Contract rate – Funding cost	Interval
>0,<=1.5%	1S
>1.5,<=3.0%	2S
>3.0,<=4.5%	3S
>4.5,<=6.0%	4S
>6.0,<=7.5%	5S
>7.5,<=9.5%	6S
>9.5%	7S

Date: the date that each contract has been approved from Sample Company. The data is numeric and show in date format. However this attribute not use for calculate the decision tree but only filter and separate data only.

After data was transformed the structure of new date for calculate in Weka program will be save in CSV (Comma-separate value) format and wait for import to program in next step. The structure of data is show below.

Table 3.2 Structure explanation of data

Transaction	Possible Data	Sample Data
Contract Type	A,B,C,D,E,F	A
Approval	P,PP	P
Statue	Existing, New, Return	Existing
Parent Rate	1,2,3,4,5,6,7,8	1
Rate	1,2,3,4,5,6,7,8,9,10	1
Customer Type	Japanese,Thai	Japanese
Business	Con, Man, Other	Man
Cost	All Number	1,000,000
RV	0%-100%	0.00
Term	All Number between 1 – 120	36
Deposit	0%-100%	0.00
Scheme	Beginning, Ending	Beginning
Sale & Lease Back	True, False	False
Privilege	None, BOI, EPZ	BOI
Insurance	Com, Cus	Cus
Spread	1S,2S,3S,4S,5S,6S,7S	2S

Some part of data after transformed in spreadsheet as show below.

Contract T	Approval	Status	Parent	Rat	Rate	Type	Business	Cost	RV	Term	Deposit	Scheme	Sale & Lea	Privilege	Insurance	Spread	Date
1	C	P	Existing	2P	4R	Japanese	Other	11C	2V	3T	0D	Beginning	FALSE	None	Customer	25	07-Apr-11
2	B	P	Existing	2P	4R	Japanese	Other	11C	2V	3T	0D	Beginning	FALSE	None	Customer	25	27-Apr-11
3	B	P	Existing	2P	4R	Japanese	Other	11C	2V	3T	0D	Beginning	FALSE	None	Customer	25	17-Feb-12
4	B	P	Existing	2P	4R	Japanese	Other	11C	2V	3T	0D	Beginning	FALSE	None	Customer	25	23-Mar-12
5	B	P	Existing	2P	4R	Japanese	Other	11C	2V	3T	0D	Beginning	FALSE	None	Customer	25	23-Mar-12
6	B	P	Existing	2P	4R	Japanese	Other	11C	2V	3T	0D	Beginning	FALSE	None	Customer	25	13-Dec-10
7	C	P	Existing	2P	4R	Japanese	Other	13C	2V	5T	0D	Beginning	FALSE	None	Customer	25	27-Dec-10
8	B	P	Existing	2P	4R	Japanese	Other	12C	2V	5T	0D	Beginning	FALSE	None	Customer	25	27-Dec-10
9	B	P	Existing	2P	4R	Japanese	Other	12C	2V	5T	0D	Beginning	FALSE	None	Customer	25	19-Jun-12
10	B	P	New	2P	5R	Japanese	Other	13C	2V	3T	0D	Beginning	FALSE	None	Customer	15	19-Jun-12
11	B	P	New	2P	5R	Japanese	Other	13C	2V	3T	0D	Beginning	FALSE	None	Customer	15	19-Jun-12
12	B	P	New	2P	5R	Japanese	Other	13C	2V	3T	0D	Beginning	FALSE	None	Customer	15	19-Jun-12
13	B	P	New	2P	5R	Japanese	Other	11C	2V	3T	0D	Beginning	FALSE	None	Customer	15	19-Jun-12
14	B	P	New	2P	5R	Japanese	Other	11C	2V	3T	0D	Beginning	FALSE	None	Customer	15	19-Jun-12
15	B	P	New	2P	5R	Japanese	Other	11C	2V	3T	0D	Beginning	FALSE	None	Customer	15	19-Jun-12
16	B	P	New	2P	5R	Japanese	Other	11C	2V	3T	0D	Beginning	FALSE	None	Customer	15	19-Jun-12
17	B	P	New	2P	5R	Japanese	Other	11C	2V	3T	0D	Beginning	FALSE	None	Customer	15	19-Jun-12
18	B	P	New	2P	5R	Japanese	Other	11C	2V	3T	0D	Beginning	FALSE	None	Customer	15	19-Jun-12
19	B	P	Existing	2P	5R	Japanese	Other	13C	2V	3T	0D	Beginning	FALSE	None	Customer	15	03-3-12
20	B	P	Existing	2P	5R	Japanese	Other	13C	2V	3T	0D	Beginning	FALSE	None	Customer	15	03-3-12
21	B	P	Existing	2P	5R	Japanese	Other	13C	2V	3T	0D	Beginning	FALSE	None	Customer	15	03-3-12
22	B	P	Existing	4P	4R	Japanese	Manufactu	15C	1V	3T	0D	Beginning	TRUE	None	Customer	25	19-Dec-13
23	A	P	New	0P	3R	Thai	Manufactu	10C	0V	4T	0D	Beginning	FALSE	None	Company	35	21-Feb-13
24	A	P	Existing	0P	3R	Thai	Manufactu	9C	0V	3T	0D	Beginning	FALSE	None	Company	35	17-Jun-13
25	A	P	Existing	0P	3R	Thai	Manufactu	9C	0V	4T	0D	Beginning	FALSE	None	Company	35	11-Oct-13
26	A	P	Existing	0P	3R	Thai	Manufactu	9C	0V	4T	0D	Beginning	FALSE	None	Company	35	07-Feb-11
27	E	P	Existing	4P	4R	Japanese	Manufactu	12C	4V	3T	0D	Beginning	FALSE	None	Customer	25	29-3-11
28	B	P	Existing	4P	7R	Japanese	Manufactu	8C	1V	3T	0D	Beginning	FALSE	None	Company	25	23-Aug-11
29	B	P	Existing	3P	7R	Japanese	Manufactu	11C	1V	4T	0D	Beginning	FALSE	None	Company	25	16-Dec-11
30	B	P	Existing	3P	7R	Japanese	Manufactu	10C	1V	3T	0D	Beginning	FALSE	None	Company	25	23-Aug-12
31	B	P	Existing	3P	7R	Japanese	Manufactu	9C	1V	3T	0D	Beginning	FALSE	None	Company	25	23-Aug-11
32	B	P	Existing	3P	7R	Japanese	Manufactu	9C	1V	3T	0D	Beginning	FALSE	None	Company	25	10-Apr-13
33	B	P	Existing	4P	7R	Japanese	Manufactu	10C	1V	3T	0D	Beginning	FALSE	None	Company	25	09-Sep-13
34	B	P	Existing	4P	7R	Japanese	Manufactu	9C	1V	3T	0D	Beginning	FALSE	None	Company	25	28-Feb-11
35	B	P	Existing	4P	4R	Japanese	Manufactu	10C	1V	3T	0D	Beginning	FALSE	None	Company	35	06-Dec-11
36	B	P	Existing	4P	4R	Japanese	Manufactu	7C	1V	3T	0D	Beginning	FALSE	None	Company	35	06-Dec-11
37	B	P	Existing	4P	4R	Japanese	Manufactu	10C	1V	3T	0D	Beginning	FALSE	None	Company	35	06-Dec-11
38	A	P	Existing	2P	6R	Japanese	Other	8C	0V	3T	0D	Beginning	FALSE	None	Company	35	06-Dec-11
39	A	P	Existing	2P	6R	Japanese	Other	10C	0V	3T	0D	Beginning	FALSE	None	Company	35	06-Dec-11
40	A	P	Existing	2P	6R	Japanese	Other	8C	0V	3T	0D	Beginning	FALSE	None	Company	35	06-Dec-11
41	A	P	Existing	2P	6R	Japanese	Other	9C	0V	3T	0D	Beginning	FALSE	None	Company	35	06-Dec-11
42	A	P	Existing	2P	6R	Japanese	Other	8C	0V	3T	0D	Beginning	FALSE	None	Company	35	06-Dec-11

Figure 3.7 Sample data after data transformation process.

3.4.5 Data Engineering

After the database has been transformed, the researcher will completed the data engineering by separate data into 2 groups by random which the first group as training set is 80% random (4,777 records) and 20% rest is testing set (1,194 records)

Training set will use to created decision tree in Weka program and use the conditions result from training set to test on testing set for accuracy rate.

CHAPTER IV

SYSTEM DEVELOPMENT

4.1 The result from Weka Program

After completion of data management, the “Training Set” data was processed in Weka Version 3.6 by using “Spread” as forecast factor. The results from the program are as follows:

```
=== Run information ===  
  
Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2  
Relation: Training2  
Instances: 4777  
Attributes: 11  
    Cost  
    Type  
    Rate  
    Status  
    PP Rate  
    Previlage  
    Business  
    Insurance  
    RV  
    Term  
    Spread  
Test mode:evaluate on training data
```

Figure 4.1 Result in Run Information section

Weka program processed ran with totally 4,777 records of data by using Weka.Classifiers.Trees.J48 method. There were 11 attributes and Evaluate on Training Data Process was used all records of data were used in generating “Decision Tree”

=== Summary ===		
Correctly Classified Instances	4223	88.4028 %
Incorrectly Classified Instances	554	11.5972 %
Kappa statistic	0.8521	
Mean absolute error	0.0565	
Root mean squared error	0.1681	
Relative absolute error	25.22 %	
Root relative squared error	50.2229 %	
Total Number of Instances	4777	

Figure 4.2 Result in Summary section

With “Decision Tree”, the results from processing the “Training Set” showed that 4,223 records or 88.40% were correct and 554 records or 11.60% were incorrect.

4.2 Creating decision tree.

The generated decision tree was a 118 node tree and has 67 final leaf nodes. The results from the program were used to draw the Decision Tree contained “Rode Node” with “Cost” less than or equal to 1,003,210.2 or greater than 1,003,210.2 as follows:

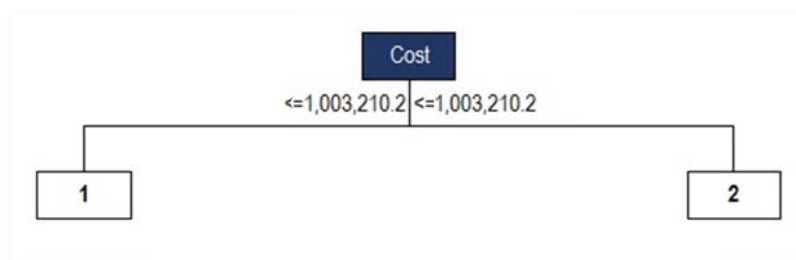


Figure 4.3 Rode node of tree

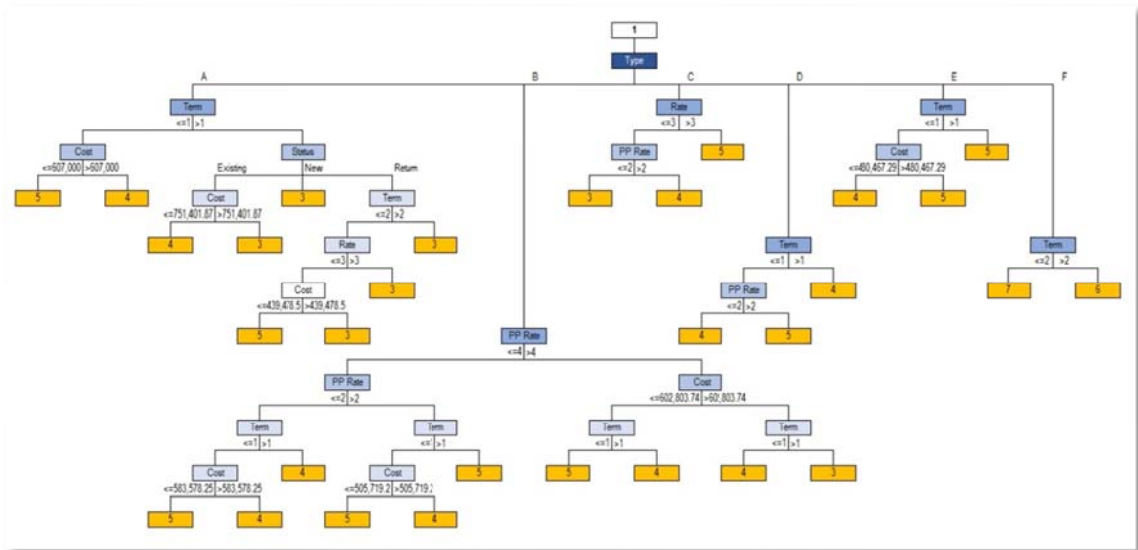


Figure 4.4 Decision Tree Node (Cost ≤ 1,003,210.2)

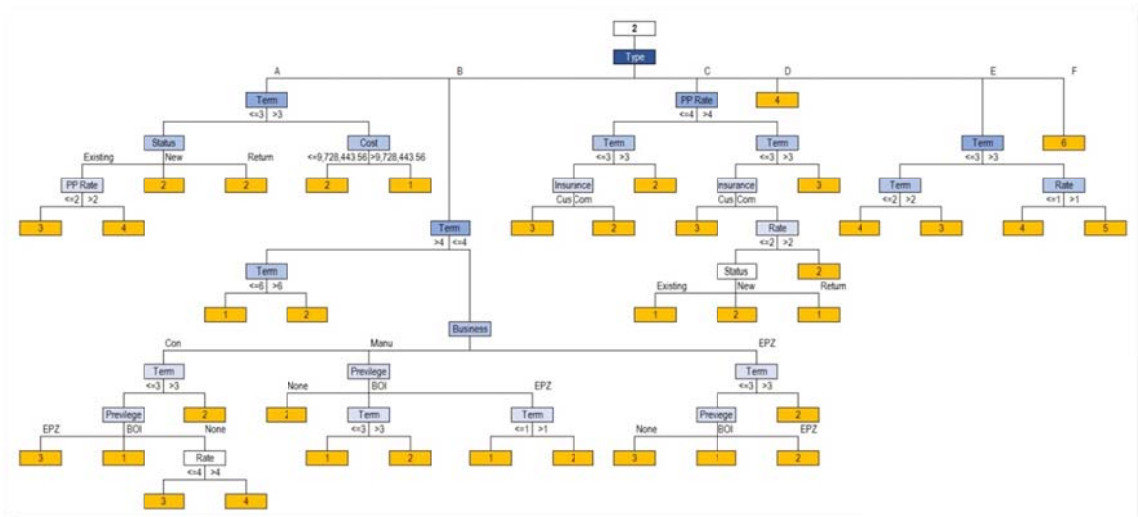


Figure 4.5 Decision Tree Node (Cost > 1,003,210.2)

By observing the generated Decision Tree, Root Node was separated with cost factor into ≤ 1 million Baht or > 1 million Baht and the next Node of Leaf Node at both sides were separated by factor “Type”.

4.3 Conversion of Decision Tree into spreadsheet

When the results were taken for doing “DSS” in Excel program, the results from Weka program were properly adjusted in the spread sheet of Excel program. The data was recorded into the page named “Tree” and the workbook was saved under the name “Dss.xlms” an Excel Macro-Enabled Workbook to support VBA writing. The sample of data received from the Decision Tree conversion are as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Cost	\leq	1003210.2	Type = B	PPRate	\leq	4	PPRate	\leq	2			Term	\leq	1	Cost	\leq	583578.25					5S
2	Cost	\leq	1003210.2	Type = B	PPRate	\leq	4	PPRate	\leq	2			Term	\leq	1	Cost	$>$	583578.25					4S
3	Cost	\leq	1003210.2	Type = B	PPRate	\leq	4	PPRate	\leq	2			Term	$>$	1								4S
4	Cost	\leq	1003210.2	Type = B	PPRate	\leq	4	PPRate	$>$	2			Term	\leq	1	Cost	\leq	505719.2					5S
5	Cost	\leq	1003210.2	Type = B	PPRate	\leq	4	PPRate	$>$	2			Term	\leq	1	Cost	$>$	505719.2					4S
6	Cost	\leq	1003210.2	Type = B	PPRate	\leq	4	PPRate	$>$	2			Term	$>$	1								5S
7	Cost	\leq	1003210.2	Type = B	PPRate	$>$	4	Cost	\leq	602803.74			Term	\leq	1								5S
8	Cost	\leq	1003210.2	Type = B	PPRate	$>$	4	Cost	\leq	602803.74			Term	$>$	1								4S
9	Cost	\leq	1003210.2	Type = B	PPRate	$>$	4	Cost	$>$	602803.74			Term	\leq	1								4S
10	Cost	\leq	1003210.2	Type = B	PPRate	$>$	4	Cost	$>$	602803.74			Term	$>$	1								3S
11	Cost	\leq	1003210.2	Type = C	Rate	\leq	3	PPRate	\leq	2													3S
12	Cost	\leq	1003210.2	Type = C	Rate	\leq	3	PPRate	$>$	2													4S
13	Cost	\leq	1003210.2	Type = C	Rate	$>$	3																5S
14	Cost	\leq	1003210.2	Type = A	Term	\leq	1	Cost	\leq	607000													5S
15	Cost	\leq	1003210.2	Type = A	Term	\leq	1	Cost	$>$	607000													4S
16	Cost	\leq	1003210.2	Type = A	Term	$>$	1	Status	=	Existing	Cost	\leq	751401.87										4S
17	Cost	\leq	1003210.2	Type = A	Term	$>$	1	Status	=	Existing	Cost	$>$	751401.87										3S
18	Cost	\leq	1003210.2	Type = A	Term	$>$	1	Status	=	New													3S
19	Cost	\leq	1003210.2	Type = A	Term	$>$	1	Status	=	Return	Term	\leq	2	Rate	\leq	3	Cost	\leq	439478.5				5S
20	Cost	\leq	1003210.2	Type = A	Term	$>$	1	Status	=	Return	Term	\leq	2	Rate	\leq	3	Cost	$>$	439478.5				3S
21	Cost	\leq	1003210.2	Type = A	Term	$>$	1	Status	=	Return	Term	\leq	2	Rate	$>$	3							3S
22	Cost	\leq	1003210.2	Type = A	Term	$>$	1	Status	=	Return	Term	$>$	2										3S
23	Cost	\leq	1003210.2	Type = E	Term	\leq	1	Cost	\leq	480467.29													5S
24	Cost	\leq	1003210.2	Type = E	Term	\leq	1	Cost	$>$	480467.29													6S
25	Cost	\leq	1003210.2	Type = E	Term	$>$	1																5S
26	Cost	\leq	1003210.2	Type = D	Term	\leq	1	PPRate	\leq	2													4S
27	Cost	\leq	1003210.2	Type = D	Term	\leq	1	PPRate	$>$	2													5S
28	Cost	\leq	1003210.2	Type = D	Term	$>$	1																4S
29	Cost	\leq	1003210.2	Type = F	Term	\leq	2																7S
30	Cost	\leq	1003210.2	Type = F	Term	$>$	2																6S
31	Cost	$>$	1003210.2	Type = B	Term	\leq	4	Business	=	Con	Term	\leq	3	Prevlige	=	None	Rate	\leq	4				3S
32	Cost	$>$	1003210.2	Type = B	Term	\leq	4	Business	=	Con	Term	\leq	3	Prevlige	=	None	Rate	$>$	4				4S

Figure 4.6 Sample data in page “Tree”

The factors in decision tree were reduced from 15 factors to 10 factors as 5 factors didn’t have effect to criteria factor which Approval, Customer Type, Deposit, Scheme and Sale & Lease Bank will not be calculate in system.

Next “DATA” page was created to input the data for forecasting as follows: in column “A” are factors for entering into the system, in column “B” are values for entering into the system and in column “C” are the data transformed from

the input in column “B” to be the calculable data according to the condition in page “Tree”. The data in “DATA” page are as follows:

	A	B	C
1	Category	Value	T Value
2	Cost	1,000,000.00	1000000
3	Type	Leasing	B
4	Term	36	3.000
5	Status	Existing	Existing
6	Rate	1	1
7	PPRate	1	1
8	Business	Manufacturing	Manu
9	Privilege	None	None
10	Insurance	Customer	Cus

Figure 4.7 Data in page “DATA”

Data in column “C” or “T Value” were converted from column “Value” or “Input Value” such as row “Type” was entered by leasing values but in column “T Value” the data was transformed to be “B” which referred to “Leasing” in Decision Tree. The same was applied to row of “Business”, “Privilege” and “Insurance”.

4.4 Interface Creation

Next, the system interface was created for easily operated by the users. The “Userform” was created from Microsoft Visual Basic for Application program. The “Userform” was similar to a questionnaire that enable user to input data factors and perform calculation by using “Calculate” button. The last part was the result and its explanation. The “Userform” features are as follows:

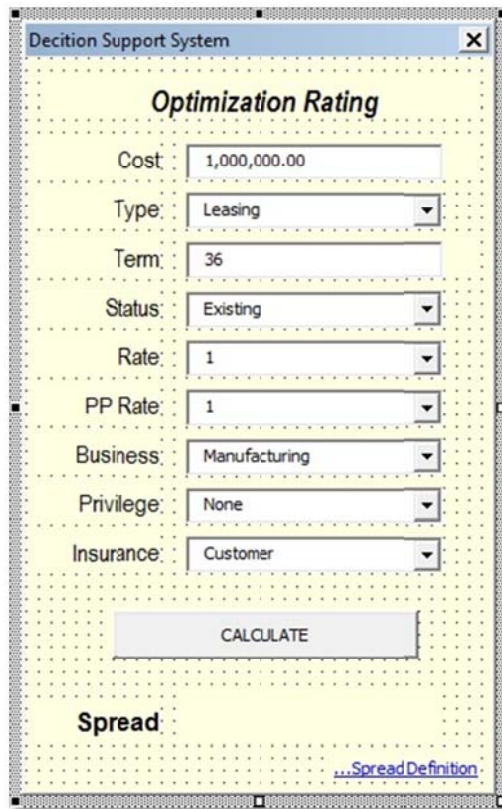


Figure 4.8 Userform1

Table 4.1 Userform1 structure.

No.	Field	Type	Data Validation
1	Cost	Textbox	Whole Number (> 0)
2	Type	Combobox	Lease
			Leasing
			Leasing + Undertaking
			Hire Purchase
			Operating Lease
			Operating + Undertaking
3	Term	Textbox	Whole Number (>0)
4	Status	Combobox	Existing
			New
			Return

Table 4.1 (cont.) Userform1 structure.

No.	Field	Type	Data Validation
5	Rate	Combobox	1-10
6	PPRate	Combobox	1-8
7	Business	Combobox	Construction
			Manufacturing
			Other
8	Privilege	Combobox	None
			BOI
			EPZ
9	Insurance	Combobox	Company
			Customer
10	Calculate	CommandButton1	Call "Calculaterating"
11	Spread	Label	1S-7S

4.5 VBA Coding

After completion of the data arrangement in page "DATA", "Tree" and "Userform", a new Module was created for writing Code into Microsoft Visual Basic for Application program for "the calculation" according to the Decision Tree by defining column "T Value" in page "DATA" as input data and named it as "calculaterating". The VBA codes are as follows:

```

Sub calculaterating()
On Error GoTo result1
FF = 1
For i = 1 To 8
aa = Application.WorksheetFunction.Match(Sheets("TREE").Range("A" &
FF).Offset(0, 3 * i - 3), _
Sheets("DATA").Range("A:A"), 0)
bb = Sheets("DATA").Range("A1").Offset(aa - 1, 2)
If Application.WorksheetFunction.Or(Sheets("TREE").Range("A" & FF).Offset(0,

```

```

1).Offset(0, 3 * i - 3) _
= "<=", Sheets("TREE").Range("A" & FF).Offset(0, 1).Offset(0, 3 * i - 3) = ">" =
True Then
CC = "1-9"
    If bb <= Sheets("TREE").Range("A1").Offset(0, 2).Offset(FF - 1, 3 * i - 3) Then
        DD = "<="
        EE = Application.WorksheetFunction.Match("<=",
Sheets("TREE").Range(Sheets("TREE").Range("B" & FF) .Offset(0, 3 * i - 3),
Sheets("TREE").Range("B10000").Offset(0, 3 * i - 3)), 0) + FF - 1
        FF = EE
    Else
        DD = ">"
        EE = Application.WorksheetFunction.Match(">",
Sheets("TREE").Range(Sheets("TREE").Range("B" & FF). Offset(0, 3 * i - 3),
Sheets("TREE").Range("B10000").Offset(0, 3 * i - 3)), 0) + FF - 1
        FF = EE
    End If
Else
CC = "A-Z"
    EE = Application.WorksheetFunction.Match(bb,
Sheets("TREE").Range(Sheets("TREE").Range("C" & FF). Offset(0, 3 * i - 3),
Sheets("TREE").Range("C10000").Offset(0, 3 * i - 3)), 0) + FF - 1
    FF = EE
End If
Next i
Exit Sub
result1:
Sheets("DATA").Range("B11") = Sheets("TREE").Range("AE1")
End Sub

```

After that, “Code” was written to link between pages of “DATA” and “Userform” to enable change of figures in column “Value” in page “DATA” in

compliance with the change of each field in Userform. Then setting to click the “Command Button 1” to operate calculation according to Code “Calculaterating” recorded earlier and displayed the results in Label of Spread. For example when the following values were entered:

Cost	=	1,000,000
Type	=	Leasing
Term	=	36
Status	=	Existing
Rate	=	1
PPRate	=	1
Business	=	Manufacturing
Privilege	=	None
Insurance	=	Customer

When opening the file and entering the values as per the example, the feature is as follow:



The screenshot shows a window titled "Decision Support System" with a close button (X). The main content area is titled "Optimization Rating" and contains a form with the following fields and values:

Cost	1,000,000.00
Type	Leasing
Term	36
Status	Existing
Rate	1
PP Rate	1
Business	Manufacturing
Privilege	None
Insurance	Customer

Below the form is a "CALCULATE" button. At the bottom of the window, the word "Spread" is displayed, followed by a link "...Spread Definition".

Figure 4.9 DSS interface with sample data

After that when click “Calculate” button, the system then operated calculation according to the recorded Decision Tree and gave result as follows:

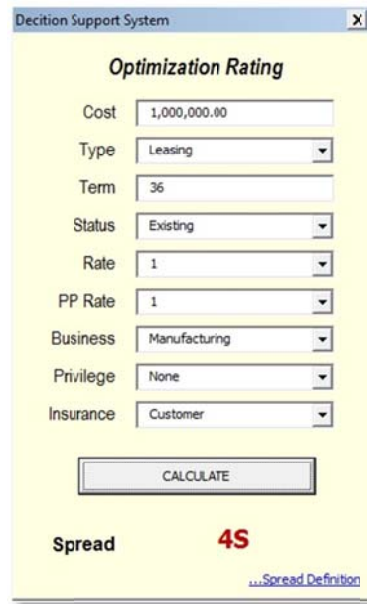


Figure 4.10 DSS interface with result

The result showed that Spread was equal to 4S or equal to 4.5% - 6.0 %, etc.

4.6 Testing DSS in testing set data.

When took the Testing Set separated earlier to process with created system, the result is as follow:

Table 4.2 Conclusion testing set classification

Correctly Classified Instances	983	82.33%
Incorrectly Classified Instances	211	17.67%
Total	1,194	100%

4.7 The Accuracy Testing by experts

A questionnaire for testing system reliability was established by randomly picking up 50 records from Training Set to be evaluated by the specialists according to

their experiences and their expectation, then compare with the result received from the created system. The 5 specialists are

- Manager of operation and credit department
- Manager of marketing and promotion department
- General Manager
- Assist manager of operation and credit department
- Team lead of marketing promotion officer

Table 4.3 The questionnaire for expert

No.	Cost	Type	Rate	Status	PP Rate	Privilege	Business	Insurance	RV	Term	Spread
1	1,236,318.53	A	2	Existing	4	None	Other	Cus	60%	4	
2	530,000.00	B	2	New	3	None	Manu	Cus	60%	7	
3	3,270,000.00	A	4	Return	6	None	Manu	Cus	0%	1	
4	56,781.60	A	2	New	4	None	Manu	Com	1%	4	
5	7,454,202.20	A	1	Existing	3	None	Manu	Com	61%	4	
6	341,321.14	E	3	Existing	4	None	Manu	Cus	55%	4	
7	3,689.55	B	3	New	4	None	Manu	Cus	0%	3	
8	972,051.40	F	2	Return	5	None	Other	Com	35%	1	
9	3,590,736.17	A	2	New	5	BOI	Con	Com	62%	4	
10	700,000.00	A	2	Return	2	None	Manu	Cus	54%	3	
11	1,043,244.38	A	3	New	5	None	Manu	Cus	15%	5	
12	10,691,723.00	A	4	Existing	5	BOI	Manu	Cus	14%	6	
13	3,092,509.12	A	7	Return	5	None	Manu	Com	50%	4	
14	405,934.58	D	2	New	3	None	Other	Cus	20%	6	
15	670,093.46	C	6	Existing	5	None	Manu	Com	0%	6	
16	1,364,485.98	A	2	Existing	1	None	Manu	Cus	0%	3	
17	20,709.14	B	4	New	4	None	Manu	Cus	71%	4	
18	12,343,750.00	E	1	Return	5	EPZ	Manu	Com	60%	3	
19	23,263,200.00	B	3	New	4	None	Con	Com	49%	4	
20	1,004,482.13	B	6	Return	5	None	Con	Cus	15%	4	
21	550,000.00	A	3	New	3	None	Manu	Com	0%	3	
22	542,000.00	B	10	New	4	None	Manu	Com	0%	3	
23	1,311,635.51	E	3	New	2	BOI	Manu	Cus	0%	1	
24	1,526,165.12	E	6	Existing	6	None	Con	Cus	15%	4	
25	872,787.00	A	4	New	3	None	Manu	Com	64%	6	
26	489,000.00	E	10	Return	6	None	Manu	Cus	55%	6	

Table 4.3 (cont.) The questionnaire for expert

No.	Cost	Type	Rate	Status	PP Rate	Privilege	Business	Insurance	RV	Term	Spread
27	462,400.00	A	3	Return	3	None	Manu	Cus	10%	2	
28	102,244,900	B	1	Return	2	None	Con	Com	50%	7	
29	7,668,225.00	A	2	New	4	None	Manu	Com	0%	1	
30	1,155,022.03	A	4	Existing	4	None	Manu	Com	15%	5	
31	10,592,193.54	D	6	New	2	None	Manu	Com	0%	5	
32	572,897.20	B	4	New	8	BOI	Manu	Cus	0%	3	
33	1,372,897.19	B	4	Existing	4	BOI	Manu	Com	59%	7	
34	10,150.00	C	2	New	2	None	Con	Cus	0%	3	
35	30,900.00	D	4	New	2	None	Manu	Cus	0%	3	
36	444,000.00	E	4	New	4	None	Manu	Cus	10%	3	
37	179,000.00	B	6	Existing	4	None	Manu	Com	58%	3	
38	230,400.00	A	4	New	2	None	Manu	Cus	40%	7	
39	17,504,455.00	A	3	Existing	3	None	Manu	Cus	0%	3	
40	14,290.86	C	1	New	6	None	Manu	Com	15%	4	
41	9,327,184.73	A	4	Existing	4	None	Con	Com	31%	4	
42	343,000.00	A	3	Existing	6	None	Con	Com	0%	2	
43	4,657,695.28	A	6	Return	5	BOI	Other	Cus	30%	4	
44	29,000.00	B	1	Existing	5	None	Con	Cus	20%	3	
45	23,061,697.34	A	4	Existing	6	None	Other	Com	56%	4	
46	78,300.00	B	4	New	3	None	Other	Cus	49%	3	
47	1,137,383.17	A	1	Existing	4	None	Manu	Com	0%	2	
48	503,500.00	A	4	Existing	7	BOI	Manu	Com	0%	3	
49	1,286,511.82	A	8	Existing	4	None	Con	Com	0%	5	
50	705,000.00	A	4	Return	4	None	Manu	Com	20%	7	

The results from the 5 specialists were later taken to find the average result and the questionnaire's result conclusion is as follow:

Table 4.4 Result from questionnaire

No.	1	2	3	4	5	Expert	DSS
1	2	1	1	2	1	1	2
2	5	6	4	5	6	5	5
3	1	2	2	1	1	1	2
4	6	5	6	6	4	5	3

Table 4.4 (cont.) Result from questionnaire

No.	1	2	3	4	5	Expert	DSS
5	1	2	2	1	2	2	2
6	5	6	4	5	4	5	5
7	5	6	6	6	4	5	5
8	7	7	7	6	6	7	7
9	3	2	2	2	3	2	2
10	5	4	5	4	3	4	3
11	1	1	1	1	1	1	2
12	1	2	1	2	2	2	1
13	3	3	3	3	2	3	2
14	3	4	3	5	3	4	4
15	5	4	5	4	5	5	5
16	4	2	3	4	4	3	3
17	4	5	5	4	6	5	5
18	4	3	2	3	3	3	3
19	2	4	3	4	2	3	2
20	3	1	1	3	3	2	2
21	2	4	4	2	2	3	3
22	6	4	6	4	4	5	5
23	5	4	3	5	3	4	4
24	5	5	4	4	6	5	5
25	4	3	3	2	2	3	3
26	5	5	4	4	5	5	5
27	2	3	3	2	2	2	3
28	1	2	1	1	2	1	1
29	2	2	2	2	1	2	2
30	3	2	1	1	3	2	2
31	5	3	5	3	5	4	4
32	3	4	4	2	2	3	4
33	1	1	1	1	1	1	1

Table 4.4 (cont.) Result from questionnaire

No.	1	2	3	4	5	Expert	DSS
34	4	4	2	3	2	3	3
35	4	4	4	4	4	4	4
36	6	4	4	4	5	5	5
37	5	4	3	4	4	4	5
38	5	5	4	4	4	4	3
39	4	3	4	5	5	4	4
40	5	6	5	5	5	5	4
41	2	2	1	3	2	2	2
42	4	3	5	5	4	4	4
43	1	3	1	1	2	2	2
44	3	3	3	3	2	3	4
45	1	2	1	1	1	1	1
46	5	6	4	4	6	5	5
47	6	4	6	6	4	5	4
48	5	3	4	4	5	4	4
49	2	2	2	3	1	2	2
50	4	2	4	2	3	3	3

From the results, comparing the recommendations of the specialists with the decision support system, it was found that 35 records or 70% were comparable.

4.8 Satisfaction test of the system

This test is an overall system testing that conducted by letting 9 people who were specialist and users to use the decision making system by separating them into a group of 5 people, included 4 marketing officers and a team leader, who directly concerned with making decision on the interest rate and regularly used the system. The other group included 4 people from management team of the sample company. The assessment was conducted by setting the qualitative criteria of opinions as follows:

5 = very satisfied 4 = satisfied 3 = neutral 2 = dissatisfied 1 = very dissatisfied

The total score of the questionnaire was calculated by multiplying the top score with the number of questions, for example: the opinion top score is 5 that translated into “very satisfied” in qualitative criteria, multiplied by number of questions which is 15 numbers so the total score of the questionnaire is 5 X 15 = 15. The assessment of the results from the questionnaire was conducted by finding the average of the results with following calculation pattern:

$$\frac{\text{Sumation of each question score} \times 100}{\text{Total score}} \tag{4.1}$$

And average “ \bar{x} ” is translated as follow

< 50% is very dissatisfied,

51 – 60% is dissatisfied,

61 -70% is neutral,

71-80 % is satisfied,

80 100 % is very satisfied.

The questionnaire feature is as follows:

Table 4.5 The System Satisfaction Questionnaire

Issue	Satisfaction Level				
	Very satisfied (5)	Satisfied (4)	Neutral (3)	Dissatisfied (2)	Very dissatisfied (1)
Convenience of system access					
Uncomplicated and easy usage					
Interface’s beauty and attractiveness					
Processing accuracy					
Processing speed					

Table 4.6 Conclusion of System satisfaction

Issue	Manager		Officer	
	\bar{x}	Meaning	\bar{x}	Meaning
Convenience of system access	95%	Very satisfied	92%	Very satisfied
Uncomplicated and easy usage	100%	Very satisfied	96%	Very satisfied
Interface's beauty and attractiveness	90%	Very satisfied	92%	Very satisfied
Processing accuracy	75%	satisfied	88%	Very satisfied
Processing speed	100%	Very satisfied	100%	Very satisfied

CHAPTER V

CONCLUSIONS

5.1 Summary of system development

From the development of making decision support system for increasing the competency of interest rate processing by using Decision Tree with method C4.5 over the sample company in leasing business by using historical recorded data in generating Decision Tree by using Weka program, the data were separated into 2 sets: Training Set for generating Decision Tree and testing set for testing the created Decision Tree. The created Decision Tree was taken to establish decision making support system by using Microsoft Excel program for calculating the conditions of the Decision Tree.

The result of the established system showed that, when tested the Decision Tree with Training Set used for generating the tree, from totally 4,777 records, the Decision Tree could reveal correct results of 4,223 records or 88.40% and revealed incorrect result on 554 records or 11.60%. For the result of the test used Testing Set, from totally 1,194 records, the Decision Tree could reveal correct results of 983 records or 82.33% and revealed incorrect results on 211 records or 17.67%. The correct results of both Training Set and Testing Set were in good criteria.

When comparing the results from 5 specialists of management team with the system results by randomly picked up 50 records from the Training Set. It was found that the results of 35 records or 70% were comparable. However, the result was the comparison of the average results from the specialists that were personal opinion of particular person.

The System Satisfaction Questionnaire used to survey the satisfaction of 4 users in management team and 5 operation officers over the system inclusively resulted excellent level of satisfaction. Both management and operation officers were very satisfied with “convenience of system access”, “uncomplicated and easy usage”, “interface’s beauty and attractiveness” and “processing speed” but for “processing

accuracy” issue, the score was in good criteria at 75 % for the management and 88% for the operation officers.

5.2 Recommendations and future work

The problem found in this research was since the system was developed by Microsoft Excel Macro Enabled so the system is not supported by the unit without the program and the type of the file can be opened on PC or Laptop only. It is not openable by other communication devices.

The data management is the most time consuming step when establishing the system due to the unsystematic of data storing of the company. That caused long period of time for data cleaning step. For generating Decision Tree in the future the data base structure should be properly created to speed up the system establishment.

The data used for developing the system were historical data. Actually new records are added all the time, therefore, for system accuracy, the processing of data for generating Decision Tree should be operated on monthly or annual basis.

The system is not required in the company workflow. It was done as an experiment only and due to company policy, the rate of interest rate correction and the duration of the comparison before and after system development are not recorded.

In part of data processing process of this paper, it uses screening factors by past experience of experts of Sample Company. The chosen factors come from all factors that checked in questionnaire. However some of factors probably have correlation with another. So it have to find correlation coefficient by process for identify predictor variable and criterion variable. Then analyze chosen factors before build decision tree.

REFERENCES

- 1 Wah YB, Ismail NH, Fong S. Predicting car purchase intent using Data Mining approach. FSKD 2011:1994-9.
- 2 Somnuk Krajang. Development of credit scoring Decision Support System for Pattanasin Leasing (CPL) Co., Ltd.. [Thematic Paper]. Chiang Mai: Chiang Mai University; 2008.
- 3 Paitoon Janreung. The Decision Supported System to choose the ungraduate students' field using Decision Tree techniques. [Thematic Paper]. Bangkok: King Mongkut's University of Technology North Bangkok; 2007.
- 4 Fan J, Wen P. Application of C4.5 algorithm in web-based learning assessment system. Proceedings of the Sixth International Conference on Machine Learning and Cybernetics.2007; 4139-43
- 5 Taweesak Nakmoung. Decision Support System. Available from: http://www.sirikitdam.egat.com/WEB_MIS/107/
- 6 Shi G. Data Mining and knowledge discovery for geoscientists. Research Institute of Petroleum Exploration and Development, Beijing, China; 2013:1-22
- 7 Han J, Kamber M. Data Mining: concepts and techniques. 2nd ed. San CA:Diane Cerra. 2006.
- 8 Quinlan JR. Induction of decision trees. Machine Learning 1. 1986; 81-106
- 9 Quinlan JR. C4.5: Programs for machine learning. CA:Morgan Kaufmann Publishers Inc. 1993.

BIOGRAPHY

NAME	Mr. Pacharapong Samantarn
DATE OF BIRTH	1 March 1988
PLACE OF BIRTH	Nonthaburi, Thailand
INSTITUTIONS ATTENDED	Kasetsart University, 2006-2009 Bachelor degree of Business Administration Mahidol University, 2012-2015 Master of Science (Technology of Information System Management)
RESEARCH GRANTS	Optimization Rating Method Using Decision Support System (DSS) In Case of a Leasing Company
HOME ADDRESS	59/341 Moo1, Sao Thonghin, Bang Yai, Nonthaburi 11140, Thailand Tel. 099 904 9094 E-mail : smt.bes@gmail.com
EMPLOYMENT ADDRESS	Bangkok Mitsubishi UFJ Lease Co., Ltd., 139 Bangrak, Bangkok 10500 Tel. 02 266 6040 E-mail : pacharapong@th.lf.mufg.jp
PUBLICATION / PRESENTATION	-