REVENUE BASED MODEL OF CUSTOMER SEGMENTATION FOR INTERNATIONAL LOGISTIC BUSINESS

DONNAPHA MEEPHOL

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2014

Thesis
entitled

REVENUE BASED MODEL OF CUSTOMER SEGMENTATION FOR INTERNATIONAL LOGISTIC BUSINESS

…..………….……………………………..
Ms. Donnapha Meephol
Candidate

……………………..……..………………
Asst. Prof. Adisorn Leelasantitham,
Ph.D. (Electrical Engineering)
Major advisor

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

……………………..……..………………
Lect. Chanattha Thongsuk,
Ph.D. (Information Technology)
Co-advisor

…………………………..……..………
Prof. Banchong Mahaisavariya,
M.D., Dip. Thai Board of Orthopedics
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 Management
Faculty of Engineering,
Mahidol University

Thesis
entitled

REVENUE BASED MODEL OF CUSTOMER SEGMENTATION FOR INTERNATIONAL LOGISTIC BUSINESS

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science
(Technology of Information System Management)
on
October 11, 2014

……………….….…..……………………
Ms. Donnapha Meephol
Candidate

……………….….…..……………………
Lect. Sotarat Thammaboosadee,
Ph.D. (Information Technology)
Chair

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

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

…………………….….….…..………
Lect. Chanattha Thongsuk,
Ph.D. (Information Technology)
Member

……………….….…..……………………
Lect.Waranyu Wongseree,
Ph.D. (Electrical Engineering)
Member

…………………….….….…..………
Prof. Banchong Mahaisavariya,
M.D., Dip. Thai Board of Orthopedics
Dean
Faculty of Graduate Studies
Mahidol University

…………………….….….…..……………
Lect. Worawit Israngkul,
M.S. (Technical Management)
Dean
Faculty of Engineering
Mahidol University

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my thesis advisor, Asst. Prof. Adisorn Leelasantitham and my thesis co-advisor Asst. Prof. Supaporn Kiattisin and Lect. Chanattha Thongsuk for their invaluable help and constant encouragement throughout the course of this research.

I would also like to extend my sincere thanks to Lect. Sotarat Thammaboosadee, Lect.Waranyu Wongseree and all staff of the Technology of Information System Management Program, Faculty of Engineering, Mahidol University for their helpful suggestions about my research methodology.

Finally, I most gratefully acknowledge my parents and my friends for all their support throughout the period of this research.

Donnapha Meephol

REVENUE BASE MODEL OF CUSTOMER SEGMENTATION FOR INTERNATIONAL LOGISTIC BUSINESS

DONNAPHA MEEPHOL 5436440 EGTI / M

M.Sc. (TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)

THESIS ADVISORY COMMITTEE: ADISORN LEELASANTITHAM, Ph.D., SUPAPORN KIATTISIN, Ph.D., CHANATHA THONGSUK, Ph.D.

ABSTRACT

Customer segmentation is usually made by the salesperson who focusing on international business logistic. There are three significant factors that are used for distinguishing customers by category, which are large, medium, and small size businesses. However, the customer segmentation that arises from the decision of salesperson which can be subjective and ambiguous as a result of individual differences in factor prioritization or factor weighting involved in the process of problem solving and decision-making are problematic. This study proposes to establish a decision-making revenue base model for customer segmentation by using a decision tree and the fuzzy logic approach to problem solving. Data which consisted of 1) outbound revenue, 2) inbound revenue, and 3) third party revenue from an international logistic company in Thailand was used as a sample for this experiment, The results of this experiment showed that the decision tree representation model of actual tree with 52 leaves had a validity of 94.44% regarding predictions, while the reliability was 96.9% of ROC (Receiver Operator Characteristic) area. The fuzzy logic representation model showed a validity of 58.15% with a reliability of 71% of ROC area. This provides business with a suitable decision-making model for the salesperson in regards to customer care and meeting arrangements, which leads to more effective sale performance.

KEY WORDS: DECISION TREE/ FUZZY/LOGISTIC/CUSTOMER SEGMENTATION

84 pages

REVENUE BASED MODEL เพื่อการแบ่งกลุ่มลูกค้าในธุรกิจขนส่งสินค้าระหว่างประเทศ

REVENUE BASE MODEL OF CUSTOMER SEGMENTATION FOR INTERNATIONAL LOGISTIC BUSINESS

ดลนภา มีผล 5436440 EGTI/M

วท.ม. (เทคโนโลยีการจัดการระบบสารสนเทศ)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์: อดิศร ลีลาสันติธรรม, Ph.D., สุภาภรณ์ เกียรติสิน, Ph.D., ชนัฏฐา ทองสุข, Ph.D.

บทคัดย่อ

การแบ่งกลุ่มลูกค้าในธุรกิจขนส่งสินค้าระหว่างประเทศนั้น โดยปกติแล้วพนักงานขายจะเป็นผู้กำหนดกลุ่มของลูกค้า โดยพิจารณาจาก 3 ปัจจัยหลัก ได้แก่ ยอดขายจากการส่งออก (Outbound revenue) ยอดขายจากการนำเข้า (Inbound revenue) และยอดขายที่เกิดจากตัวกลางผู้ส่งสินค้า(Third Party revenue) โดยแบ่งเป็น 3 กลุ่มตามขนาดของลูกค้า ได้แก่ขนาดเล็ก ขนาดกลาง และขนาดใหญ่ แต่การตัดสินใจของพนักงานนั้นอาจแปรผันไปได้ตามประสบการณ์ของแต่ละบุคคล งานวิจัยนี้จึงศึกษาใช้ต้นไม้การตัดสินใจ (Decision Tree) และ ตรรกศาสตร์คลุมเครือ (Fuzzy Logic) มาสร้างตัวแบบเพื่อเป็นเครื่องมือช่วยตัดสินใจให้พนักงานขายในการแบ่งกลุ่มลูกค้า

ผลการวิจัยพบว่าตัวแบบที่ได้จากต้นไม้ตัดสินใจพบว่า มีขนาดกิ่งของต้นไม้ ที่ 52 กิ่ง ค่าความถูกต้อง (Correctly Classified Instances) อยู่ที่ 94.44 % ค่า ROC ที่ 96.9% และตัวแบบที่ได้จากตรรกศาสตร์คลุมเครือนั้น มีค่าความถูกต้องที่ 58.9% และมีค่า ROC ที่ 71% ตัวแบบดังกล่าวสามารถนำไปใช้ในการแบ่งกลุ่มลูกค้าเพื่อพนักงานขายสามารถวางแผนการให้บริการลูกค้าได้อย่างมีประสิทธิภาพมากที่สุดต่อไป

84 หน้า

CONTENTS

CONTENTS (cont.)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (cont.)

Figure **Page**

CHAPTER I
INTRODUCTION

1.1 Background and Problems

At present, it is accepted that the export and import sector has significant role and importance to the economic development of the country. Due to the international trade promotion policy and regional trade liberalization, Thailand's export and import revenue for the year 2013 was more than 14 billion baht. For this reason, the business of international logistics is playing an important role for economic development as a transport connection that is convenient, cost and time saving for exporter and importer. The number of exporters and importers has dramatically increased during the last few years (figure 1.1). Which causes a variety in demand for logistical services. The result in supply of providers of international logistics of goods between countries is increasing.

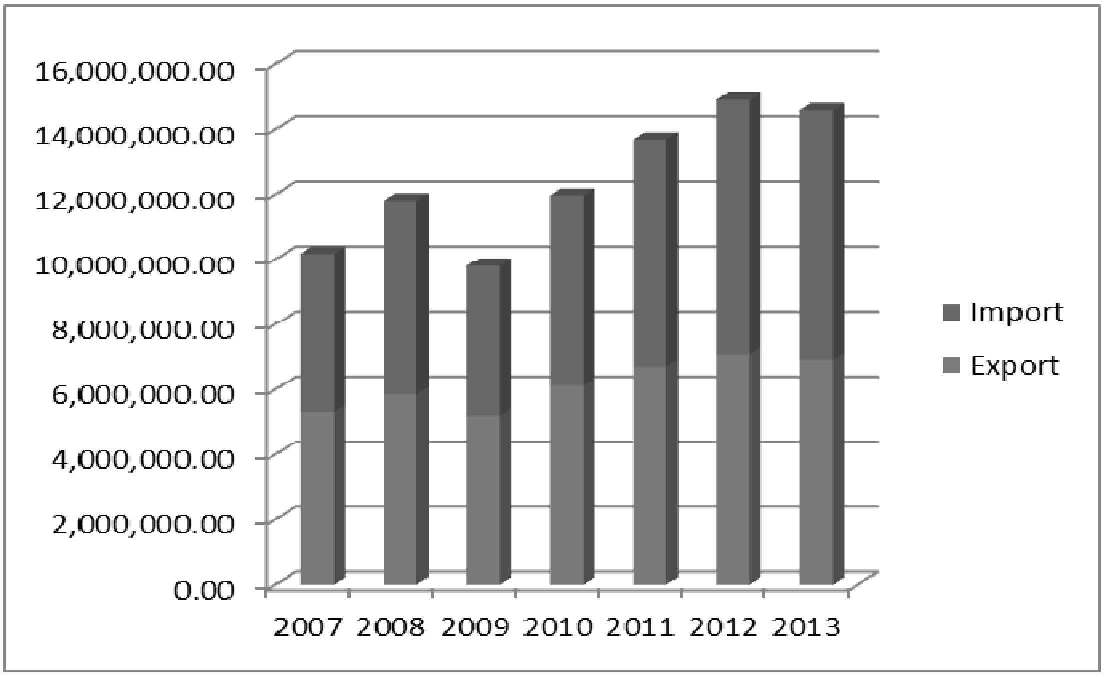


Figure 1.1: Thailand's Import-Export statistics from 2007-2013 (million baht)

As a result of the steadily rising number of entrepreneurs in this business sector, it will cause a higher business competition between these companies. Any company that's aware of customer needs would satisfy the customers and create a competitive advantage in the market.

We use customer segmentation to classify different groups of customer base accordingly to individuals that bear similarity of specific marketing-related ways to each other. A company can't only apply customer segmentation to aim at specific customers' groups sufficiently, but also designate marketing resources for following successive results; to determine relationship strategy that is suited for individual customer, customer groups and market segments resulting in successful communication across related-business units (such as customer service and marketing) to sustain product purchases, 2) to asses individual customer, groups of customers and market segments which will lead to appropriate marketing plan, and 3) to distribute the eligible company's resources.

However, the criteria for customer segmentation are different in each company. It is also very subjective and dependent on the decision of the salesperson. For example, Salesperson "A" would segment the company "ABC" in a medium-sized company, whereas, salesperson "B" probably define the company "ABC" as a large-sized company, which can be lead to ineffective strategy planning, such as the customer visit's frequency which is different depend on company size. (At least every 3 months for large-sized company or at least every year for small-sized company and so on).

By the way, each international logistic company has different factors for customer segmentation. The study brought three factors under consideration;1) outbound revenue (OB revenue), 2) inbound revenue (IB revenue) and 3) third party revenue (B3P revenue). To establish a decision-making revenue base model for customer segmentation by using a decision tree and the fuzzy logic approach to problem solving.

1.2 Objectives

The objective of this work is

1.2.1 To segment customers in the international logistic services business.

1.2.2 To study the use of the “decision tree” and “fuzzy logic” for helping salesperson to make the decision on customer segmentation

1.3 Scope of Work

The scope of this work included the following:

1.3.1 Population of this research is the customer data from international shipping companies which are totally 20,280 data.

1.3.2 Variables of this research are a) outbound revenue, b) inbound revenue and c) third party revenue.

1.4 Definitions

1.4.1 International logistics means the management of physical and information flows of goods between the point of origins and the point of usage, applying the process of planning, implementing and controlling. The resources from a company’s supply chain are well-managed in order to implement International logistics across international borders.

1.4.2 Outbound revenue means amounts that company received from customers that using the exporting shipment services.

1.4.3 Inbound revenue means amounts that company received from customers that using the importing shipment services.

1.4.4 Third party revenue means amounts that company received from customers that not either shipper or receiver.

1.4.5 Customer segmentation (market segmentation) takes place when we want to divide customers which bear similarity of some characteristics to each other into groups.

1.5 Expected Result

The outcomes of this work proved that the method of “decision tree” or/and “fuzzy logic” would be good for customer segmentation for international logistics company in the means of a) alternative method for customer segmentation b) supply the information for sale person in customer segmentation.

CHAPTER II
LITERATURE REVIEW

This chapter will explain about Theory and detail of Decision Tree, Fuzzy logic and including theory about customer segmentation

2.1 Decision Tree Introduction

A decision tree is a model that is used to classify a recursive partitioning of the instance space. The decision tree is formed by multiple nodes. A kind of node which has no incoming but outgoing edges is called an internal or test node, whereas all other nodes which have exactly one incoming edge are called leaves (or so called, “terminal node” or “decision nodes”). Each interior node corresponds to one of the input variables. In the simplest and most frequent case, each node represents a test on an attribute which partitions the instance space according to the attribute’s value. To illustrate, the condition implies a range in the case of numeric attributes.

Each leaf represents a value of the most appropriate target variable. Alternatively, the leaf of the tree is labeled with a probability distribution over the target attribute with a certain value. The outcome of the tests along the path indicates the root of the tree down to a leaf which navigates instances. Figure 2.1 represents a decision tree which gives out reasons and possibilities whether a potential customer will respond to a direct mailing or not. The circles denote internal nodes, while the triangles signify leaves. This decision is a consolidation of both numeric and nominal attributes. The analyst can use this as a classifier to predict multiple responses of a potential customer (by viewing the tree downward) and understand why each potential customer behaves differently in responding direct mailing. The nodes and their branches are labeled with the attribute they test and corresponding values, respectively.

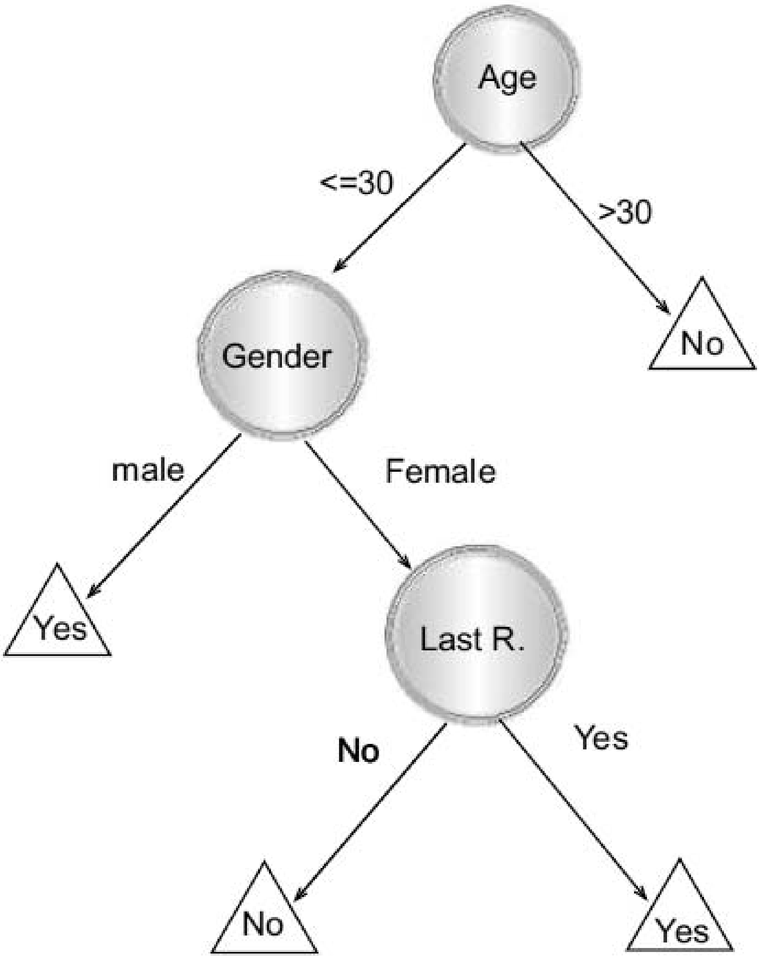


Figure 2.1: Decision Tree Presenting Response to Direct Milling.

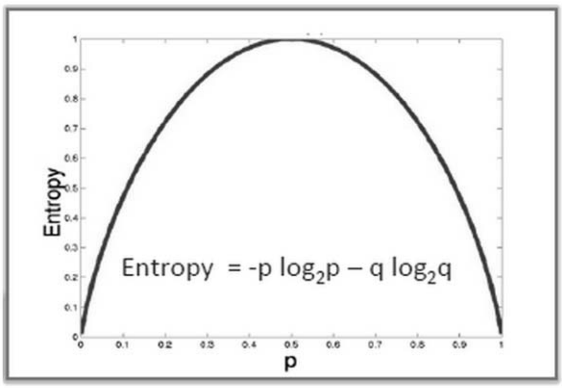
Decision tree methods are tuned in a form of hyperplanes especially for attributes that are numeric. Obviously, less complicated decision trees are way more preferable because it's considered easier to understand. Moreover, Breiman et al. said in 1984 that the complexity of the tree plays a key role in determining accuracy. The use of stopping criteria and the employment of the pruning method control tree complexity. Some metrics which usually measure tree complexity include the nodes' total number, leaves' total number, tree depth and number of attributes used. Decision tree and rule deduction are closely interrelated to each other. A decision tree has several paths connecting to its leaves. Each of the paths is capable of combining the tests for the formation of the antecedent part and considering the leaf's class prediction and the class value in order to describe a rule. To illustrate in the Figure 9.1, the path describes the rule “If the customer is “Male” who is aged 30 years old, or younger, then the customer will respond to the mail”. The rule can then be adjusted and improved for better understanding and easily-used application among human users. It’s also expected to raise standard of accuracy. (Quinlan, 1987)

2.2 Algorithm of Decision Tree

J.R. Quinlan invented ID3 which is an algorithm in a decision tree. It uses a top-down and greedy approach into the space of possible branches without backtracking. In order to build up a decision tree, the ID 3 also needs Entropy and Information Gain.

2.3 Entropy

In order to establish a decision tree, the line must be drawn top-down from a root node. In addition, the data is partitioned into subsets that involve instances with similar values (homogenous). Entropy is also necessary for ID3 to calculate a sample’s homogeneity. The entropy is always zero, when the sample is homogeneous, and always one, when the sample is equally divided.



Entropy = $-0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1$

Building a decision tree need to calculate two types of entropy using frequency tables as follows:

a) Entropy using the frequency table of one attribute:

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

Play Golf	
Yes	No
9	5

Entropy(PlayGolf) = Entropy (5,9)
= Entropy (0.36, 0.64)
= - (0.36 $\log_2$ 0.36) - (0.64 $\log_2$ 0.64)
= 0.94

b) Entropy using the frequency table of two attributes:

$$E(T, X) = \sum_{c \in X} P(c) E(c)$$

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

E(PlayGolf, Outlook) = **P**(Sunny)***E**(3,2) + **P**(Overcast)***E**(4,0) + **P**(Rainy)***E**(2,3)
= (5/14)*0.971 + (4/14)*0.0 + (5/14)*0.971
= 0.693

2.4 Information Gain

The information gain is an expected reduction in entropy caused by a dataset according to a given attribute. A decision tree can assist in finding attribute that results in the highest information gain.

Step 1: Calculate entropy of the target.

$$\begin{aligned} \text{Entropy (PlayGolf)} &= \text{Entropy } (5,9) \\ &= \text{Entropy } (0.36, 0.64) \\ &= -(0.36 \log_2 0.36) - (0.64 \log_2 0.64) \\ &= 0.94 \end{aligned}$$

Step 2: The dataset can be divided into the different attributes. After the entropy for each branch is calculated, it is added proportionally to come up with total entropy which is deducted from the preliminary entropy before being split up. A decrease of entropy is so-called the "Information Gain".

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain =0.247			

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1
Gain =0.029			

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1
Gain =0.152			

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3
Gain =0.048			

$$Gain(T,X) = Entropy(T) - Entropy(T,X)$$

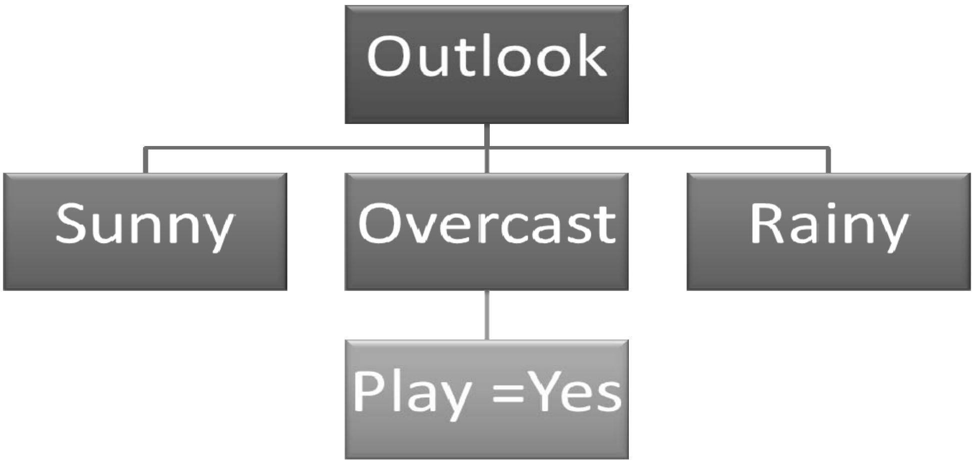
$$\begin{aligned} G(\text{PlayGolf, Outlook}) &= E(\text{PlayGolf}) - E(\text{PlayGolf, Outlook}) \\ &= 0.940 - 0.693 \\ &= 0.247 \end{aligned}$$

Step 3: Choose attribute with the largest information gain as the decision node.

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain =0.247			

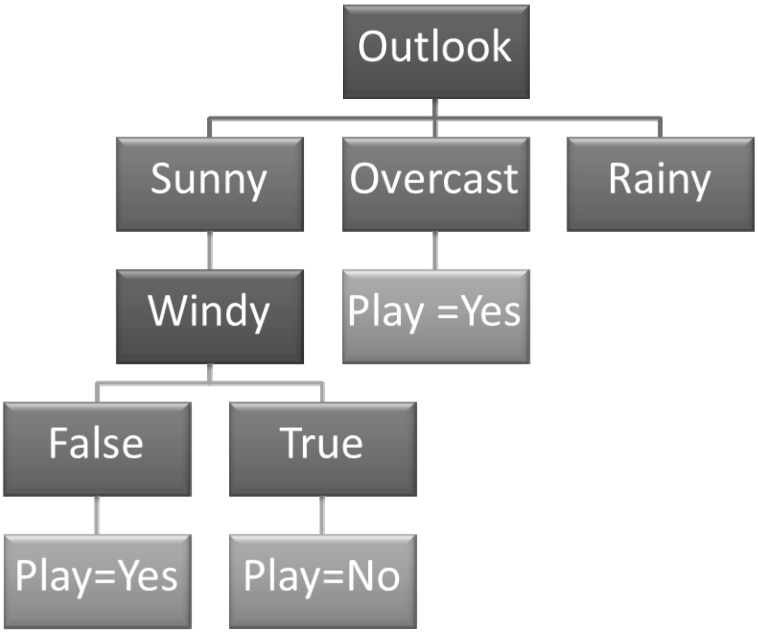
Step 4a: A branch with entropy of 0 is a leaf node.

Temp	Humidity	Windy	Play Golf
Hot	High	False	Yes
Cool	Normal	True	Yes
Mind	High	True	Yes
Hot	Normal	False	Yes
Hot	High	False	Yes



Step 4b: A branch with entropy more than 0 needs further splitting.

Temp	Humidity	Windy	Play Golf
Mind	High	False	Yes
Cool	Normal	False	Yes
Mind	Normal	False	Yes
Cool	Normal	True	No
Mind	High	True	No



Step 5: The ID3 algorithm is run recursively on the non-leaf branches, until all data is classified.

2.5 Decision Tree to Decision Rules

A decision tree can easily be transformed to a set of rules by mapping from the root node to the leaf nodes one by one.

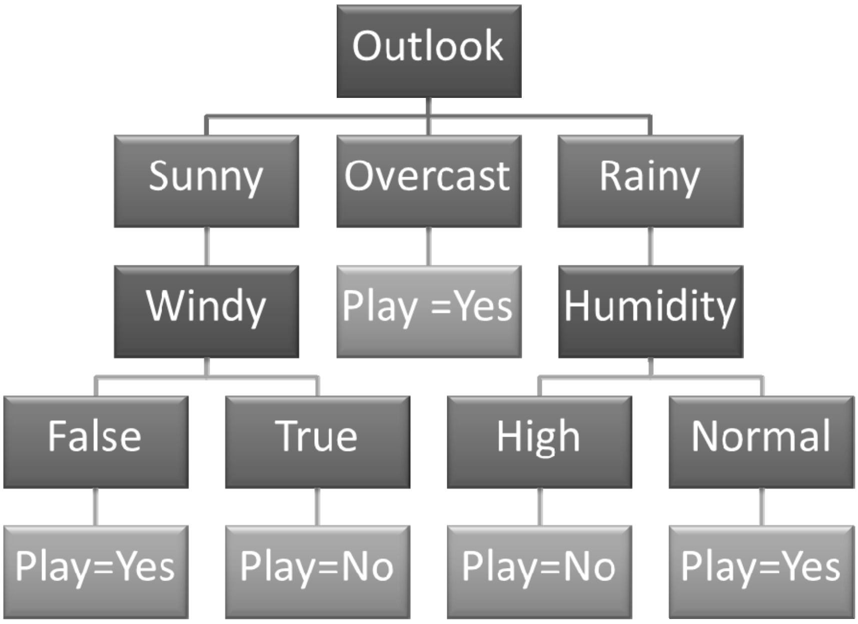
R1: IF (Outlook = Sunny) AND (Windy = FALSE) THEN Play = Yes

R2: IF (Outlook = Sunny) AND (Windy = TRUE) THEN Play = No

R3: IF (Outlook = Overcast) THEN Play = Yes

R4: IF (Outlook = Rainy) AND (Humidity = High) THEN Play = No

R5: IF (Outlook =Rainy) AND (Humidity = Normal) THEN Play = Yes



2.6 Fuzzy Logic Introduction

Fuzzy logic is originated by a form of humans linguistics variables. The rules are applied to the mathematical equivalents of fuzzy systems. The system designer and the computer also alter this method to simplify their jobs, as well as to ensure the accuracy of the results.

Thanks to fuzzy logic's simplicity and flexibility, it can not only efficiently deal with vague and incomplete data, but also explore arbitrary complexity through nonlinear functions. Bob Varley, a Senior Systems Engineer at Harris Corp., an aerospace company in Palm Bay, Florida cited that the fuzzy can provide more effective solution than conventional control techniques for those who lack of good plant model, or encounter the changes of the system.

In term of flexibility, fuzzy system is invented to be able to match any set of input-output data. The Fuzzy Logic Toolbox provides techniques such as adaptive neuro-fuzzy inference systems (ANFIS) and fuzzy substractive clustering which can be easily adapted to any kinds of data.

The if-then rules statements are used to formulate the conditional statements that comprise fuzzy logic models or fuzzy interference systems. The well-trained designer can easily write the rules with unlimited numbers of rules used to explain the system adequately (however, a moderate number of rules are typically needed).

The fuzzy logic differs from conditional logic. The truth of any statement ranges in several degrees. (How cold is it? How high should we set the heat?) We mostly get used to the rules of the form p -> q (p implies q). For fuzzy logic, however, it can probably be (.5*p) -> (.5 * q). To simplify, if (the weather is cold) then (heat is on), ranges of values will then be mapped accordingly to both variables (cold and on). Fuzzy interference systems apply membership functions to instruct the computer on how get the accurate value between 0 and 1 which verifies if any fuzzy statement is true.

The rule-based approach and flexible membership function scheme are implemented to create fuzzy systems straightforwardly. In addition, they also facilitate the designs of the systems and ensure that the system can be updated and maintained over time.

2.7 Fuzzy Set

Professor Lofti Zadeh from the University of California was the one who proposed fuzzy set in 1965 which first gained acceptance in the Far East. The rule was later applied successfully and has been adopted worldwide.

A paradigm is a set of rules that establishes boundaries and suggests how to behave inside those boundaries in order to be successful in solving problems. For instance, when a person chooses transistors over vacuum tube, a paradigm shift is properly applied – likewise the development of fuzzy set theory from conventional bivalent set theory is also considered a paradigm shift.

Bivalent Set Theory cannot be applied flexibly and freely in order to mathematically describe a 'humanistic' problem. Figure 2.2 exemplifies bivalent sets to demonstrate a room's temperature.

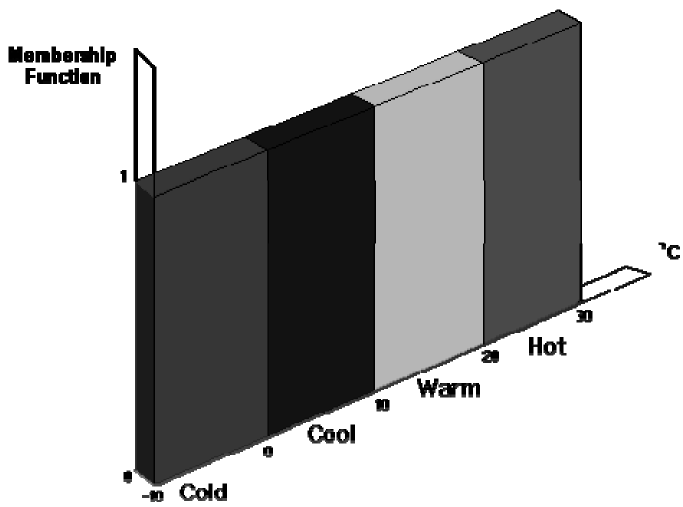


Figure 2.2: Bivalent Sets to Characterize the Temp. of a room

From the diagram above, the most obvious limiting feature is that they are mutually exclusive – it's impossible to see membership of more than one set (it can be very subjective to judge whether 50 degrees Fahrenheit is 'cold' or 'cool', the expert knowledge needed to define the system is, therefore, in disagreement with the real world). Obviously, one degree Fahrenheit of heat is incapable of defining a quantity such as 'warm' to 'hot' accurately. In reality, a smooth drift from warm to hot would be apparent.

In order to describe this natural phenomenon accurately, the Fuzzy set theory is recommended. Figure 2.3 proves the fuzzy sets' efficiency in quantifying the same information to describe this natural drift.

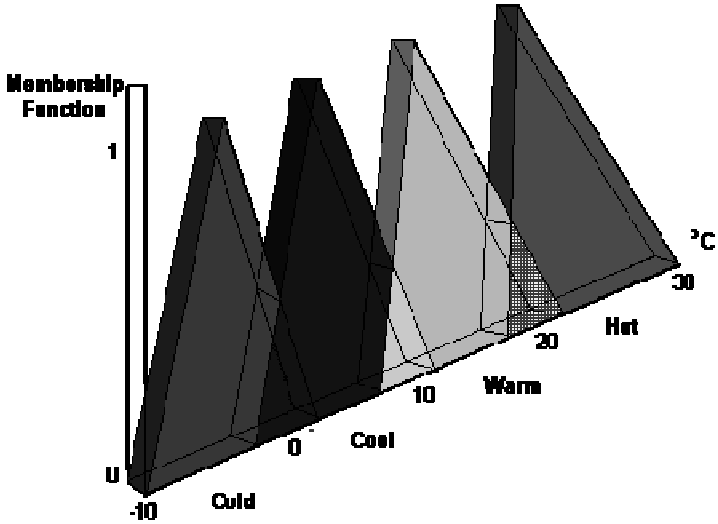


Figure 2.3: Fuzzy Set to characterize the Temp. of a room.

The entire concept can also be applied to the example of people and "youngness". The set S (the universe of discourse) refers to the set of people. A fuzzy subset YOUNG is also defined to find out what degree of ages a person x can be considered young? A degree of membership in the subset YOUNG will be labeled to each person in the universe of discourse. We can easily apply a membership function based on the person's age.

young(x) = { 1, if age(x) <= 20,
(30-age(x))/10, if 20 < age(x) <= 30,
0, if age(x) > 30 }

A graph of this looks like:

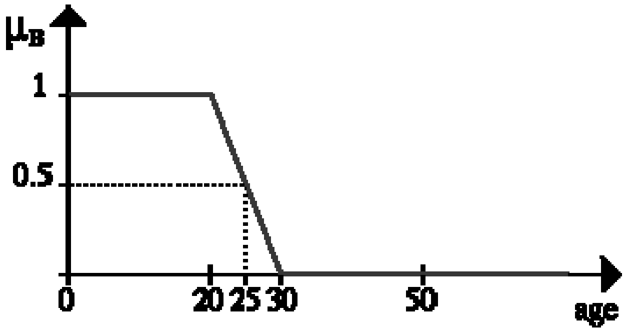


Figure 2.4: Graph of youngness people

Table 2.1: Age and Degree of youth

[redacted]		
Johan	10	1.00
Edwin	21	0.90
Parthiban	25	0.50
Arosha	26	0.40
Chin Wei	28	0.20
Rajkumar	83	0.00

Conclusively, the degree of truth of the statement “Parthiban is YOUNG” is 0.50.

Note : It’s very rare to see the membership functions shaped as simple as age (x). They can probably be triangles pointing up, or even much more complicated than that. In addition, as far as it has been discussed, membership functions are always based on a single criterion, but this isn’t always true. For example, the membership functions can be adapted for YOUNG depend on both a person’s age and their height (Arosha’s short for his age). The case is occasionally but legitimate. It’s referred to as a two-dimensional membership function. Membership function allows us to invoke elements from two completely different universes of discourse.

2.8 Membership functions

μA:X → [0,1], is used to define a membership function for a fuzzy set A on the universe of discourse X. Each element of X is adhered to a value (known as membership value or degree of membership) from 0 to 1 which is capable of quantifying a range of grade for membership in the X element to the fuzzy set A

We can apply membership functions to understand a fuzzy set in graphical form. The x axis and y axis represents the universe of discourse and the degrees of membership in the [0,1] interval, respectively.

Membership functions cannot be completed without simple functions. As we are now trying to understand fuzzy concept precisely, it's not a very good idea to apply too complicated functions.

Type of Member ship function

- **Triangular function**: defined by a lower limit a, an upper limit b, and a value m, where a < m < b.

$$\mu_A(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{m-a}, & a < x \le m \\ \dfrac{b-x}{b-m}, & m < x < b \\ 0, & x \ge b \end{cases}$$

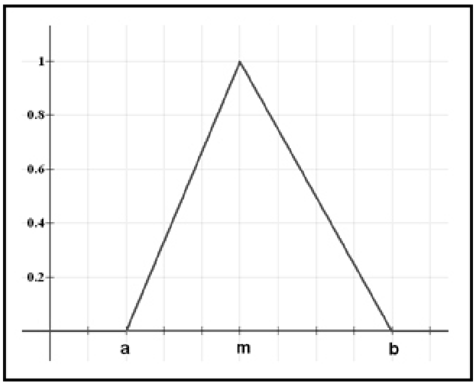


Figure 2.5 : Triangular function

- **Trapezoidal function**: defined by a lower limit a, an upper limit d, a lower support limit b, and an upper support limit c, where a < b < c < d.

$$\mu_A(x) = \begin{cases} 0, & (x < a) \text{ or } (x > d) \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \dfrac{d-x}{d-c}, & c \le x \le d \end{cases}$$

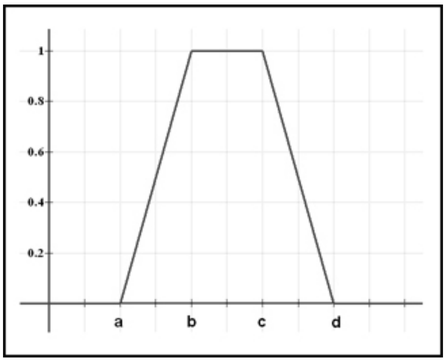


Figure 2.6 : Trapezoidal function

There are two special cases of a trapezoidal function, which are called R-functions and L-functions:

- **R-functions**: with parameters a = b = - ∞

$$\mu_A(x) = \begin{cases} 0, & x > d \\ \dfrac{d-x}{d-c}, & c \le x \le d \\ 1, & x < c \end{cases}$$

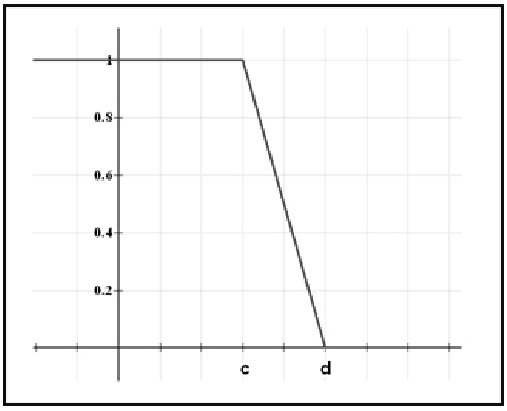


Figure 2.7 : R-functions

- **L-Functions**: with parameters c = d = + ∞

$$\mu_A(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$$

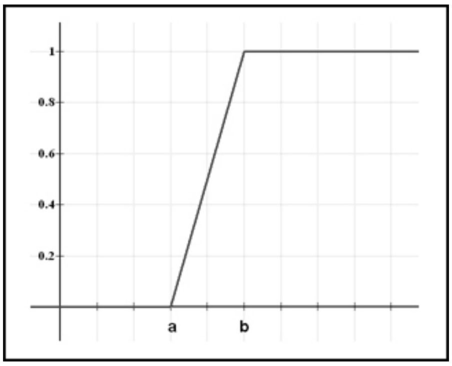


Figure 2.8 : L-Functions

- **Gaussian function**: defined by a central value m and a standard deviation k > 0. The smaller k is, the narrower the "bell" is.

$$\mu_A(x) = e^{-\frac{(x-m)^2}{2k^2}}$$

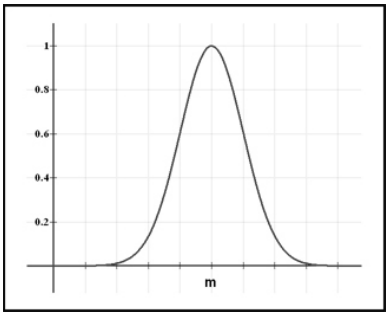


Figure 2.9 : Gaussian function

2.9 Fuzzy Set Operations

The most widely used operations in classical set of theory consist of three operations: union, intersection, and complement. These operations are generalization of crisp set operations in several ways. However, there's a particular generalization which causes operations that are usually called "standard fuzzy set operations". It plays a very significant role in fuzzy set theory. The standard operations are introduced as follows.

- **Union**: We will consider two fuzzy sets A and B represented by the membership functions μ_A and μ_B as the maximum of the two individual membership functions, or so called *maximum* criterion.

$$\mu_{A \cup B} = \text{MAX}\ (\mu_A, \mu_B)$$

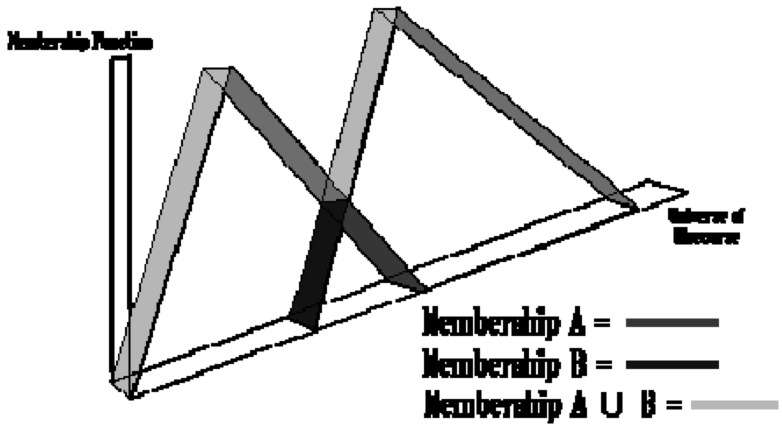


Figure 2.10 : The Union operation in Fuzzy set theory is the equivalent of the OR operation in Boolean algebra.

- **Intersection**: We will consider two fuzzy sets A and B represented by the membership functions μ_A and $\mu_{B\ a}$ as the minimum of the two individual membership functions, or so called *minimum* criterion.

$$\mu_{A \cap B} = \text{Min}\ (\mu_A, \mu_B)$$

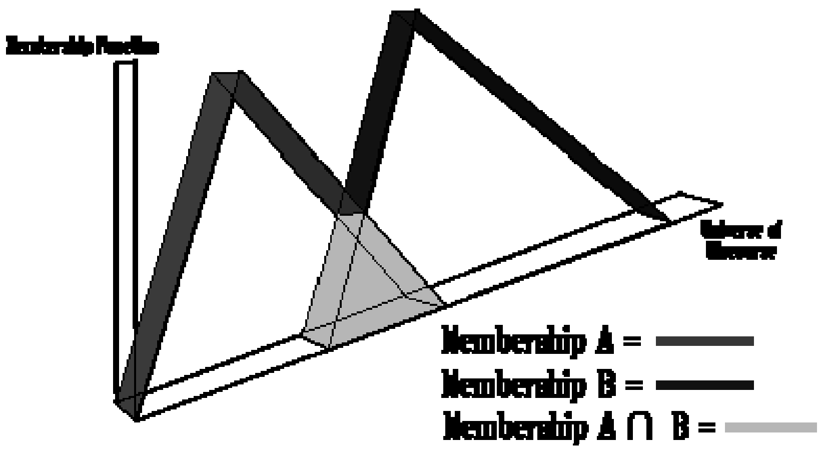


Figure 2.11: The Intersection operation in Fuzzy set theory is the equivalent of the AND operation in Boolean algebra.

- **Complement**: We will consider two fuzzy sets A represented by the membership functions μ_A as the negation of the specified membership function, or so called *negation* criterion.

$$\mu_{\overline{A}} = 1 - \mu_A$$

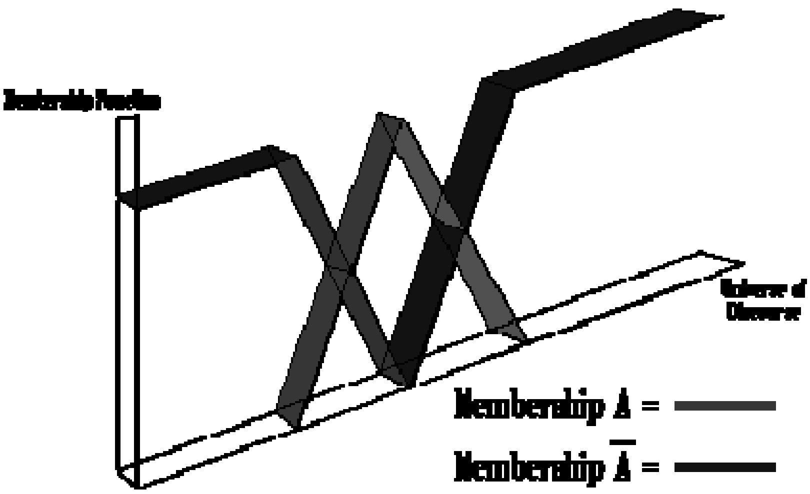


Figure 2.12 : The Complement operation in Fuzzy set theory is the equivalent of the NOT operation in Boolean algebra.

2.10 Fuzzy Rule

Unaware of it, human beings always apply rules to make decision. The if-then statements also unavoidably play a key role in making decision. If the weather looks pleasant, then we will go outside. If the weather forecasting warns that it's miserable out, then we'll decide to stay in the house until the weather is fine enough. Rules integrate ideas and vary in different circumstances in consequence.

Fuzzy machines are known to mirror the behavior of human beings. Nevertheless, when a person makes a decision, the fuzzy sets conjure up to replace the means of choosing that decision, while the rules are replaced by fuzzy rules alongside a series of if-then statements. To illustrate, if X then A, if y then B, where A and B represent sets of X and Y. Fuzzy rules define fuzzy patches, which is the key idea in fuzzy logic.

Thanks to the concept invented by Bart Kosko, a machine called Fuzzy Approximation Theorem (FAT) is far smarter. As seen in figure 2.13, the FAT theorem which generally states a finite number of patches can cover curve. The larger the patches are, the sloppier the rules are. The rules are fine only when the patches get smaller.

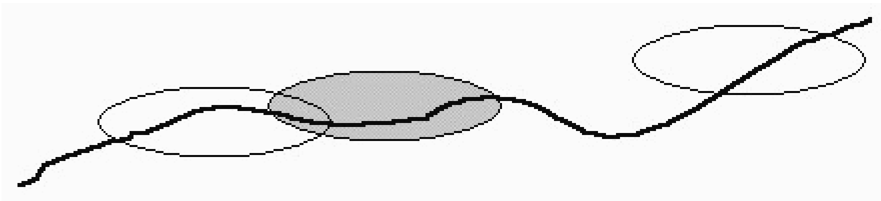


Figure 2.13 : Curve of finite number of patches

2.11 Fuzzy Control

Structure of a fuzzy logic controller L.A. Zadeh (Zadeh, 1973) studied on fuzzy algorithms and proposed the idea of the application of logical rules in a bid to formulate the control algorithm. An FLC is compatible with a set of the form's rules.

IF (a set of conditions are satisfied) *THEN* (a set of consequences can be in ferred)

The Fuzzy concepts (linguistic terms) are closely related with the premises and conclusions of IF-THEN statement, so that they are usually called "fuzzy conditional statements". As described in FLC terminology, the fuzzy control rule becomes fuzzy conditional statement when the premise is a condition in its application domain and the conclusion is the action that controls the system under control. In general, the input to a fuzzy rule is the current value for the fuzzy sets, while the output is an entire fuzzy set. This set will later be defuzzified, assigning one value to the output. Fuzzy logic control systems are often formed by four main parts, Fuzzy interface, Fuzzy rule base, Fuzzy interference engine and Defuzzification interface.

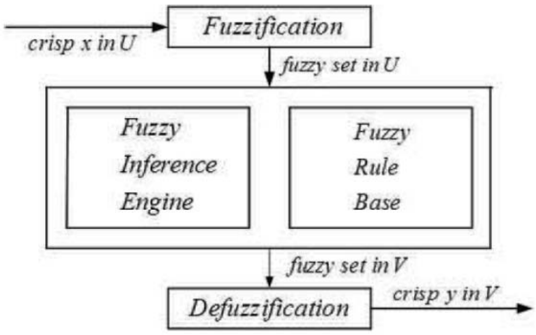


Figure 2.14: Fuzzy Logic Controller

One of the most popular methods of fuzzy interference was proposed by Mamdani and Assilian in 1975 in an effort to control steam engine and combination of boiler. To achieve their objectives, they integrated a set of rules for linguistic control derived from direct experiences of human operators. A big inspiration of their effort to discover fuzzy control came from Zadeh's 1973 paper. The literature about the subject has also grown speedily afterward. Some surveys related to the field include those of Lee in 1990, or more recently in Sala et al in 2005. According to Mamdani's model, the Mamdani's minimum operator models the fizzy implication. Both conjunction operator and t-nom from compositional rule are min, while the aggregation of the rules is max. In 2010, Rakic proposed an example of FLC to examine a simple two-input one-output problem that includes three rules.

Rule1 : IF x is A_3 OR y is B_1 THEN z is C_1
Rule2 : IF x is A_2 AND y is B_2 THEN z is C_2
Rule3 : IF x is A_1 THEN z is C_3.

Step 1: Fuzzification

The first step is to take the crisp inputs, x_0 and y_0, and determine the degree to which these inputs belong to each of the appropriate fuzzy sets. According to Fig 2.15(a) one obtains.

$$\mu A_1 (x_0) = 0.5, \mu A_2 (x_0) = 0.2, \mu B_1 (y_0) = 0.1, \mu B_2 (y_0) = 0.7$$

Step 2: Rules evaluation

The inputs that are fuzzified will depend on the premise of the fuzzy rules. If the current fuzzy rule results in diverse conclusions, the fuzzy operator (AND or OR) will be applied to come up with a single number which means the result of the premises assessment. Those who use OR fuzzy operation expect the evaluation of the disjunction of the rule premises. The classical fuzzy operation union is typically applied :

$$\mu_{A \cup B}(x) = \max\{\mu_A(x), \mu_B(x)\}.$$

Similarly, in order to evaluate the conjunction of the rule antecedents, the AND fuzzy operation intersection is applied:

$$\mu_{\mathrm{A} \cap \mathrm{B}}(x) = \min\{\mu_{\mathrm{A}}(x), \mu_{\mathrm{B}}(x)\}.$$

The result is given in the Figure 2.15(b).

The result of the premises assessment can now be to the membership function of the conclusion. One of the most widely used methods is called “clipping” which can be done easily by cutting the conclusion membership function at the premise truth’s level. Though sliced top of the membership function results in the loss of the clipped fuzzy set’s information, clipping is still widely used as it’s not complicated and results in a total output surface that is easier to be defuzzified. There’s another method called scaling which performs better in preserving the original shape of fuzzy set : the original membership function of the rule consequent is adjusted by multiplying all its membership degrees by the truth value of the rule antecedent (see Figure 2.15(c)).

Step 3: Aggregation of the rule outputs

The membership functions of all rule consequents previously clipped or scaled are combined into a single fuzzy set (see Figure 2.16 (a)).

Step 4: Defuzzification

The most popular defuzzification method is the centroid technique. It finds a point representing the center of gravity (COG) of the aggregated fuzzy set *A*, on the interval [*a*, *b*]. A reasonable estimate can be obtained by calculating it over a sample of points. According to Figure 2.16 (b), in our case results.

$$\text{COG} = \frac{\int_a^b \mu_A(x)x\,dx}{\int_a^b \mu_A(x)\,dx}$$

$$\text{COG} = \frac{(0 + 10 + 20) \times 0.1 + (30 + 40 + 50 + 60) \times 0.2 + (70 + 80 + 90 + 100) \times 0.5}{0.1 + 0.1 + 0.1 + 0.2 + 0.2 + 0.2 + 0.2 + 0.5 + 0.5 + 0.5 + 0.5} = 67.4$$

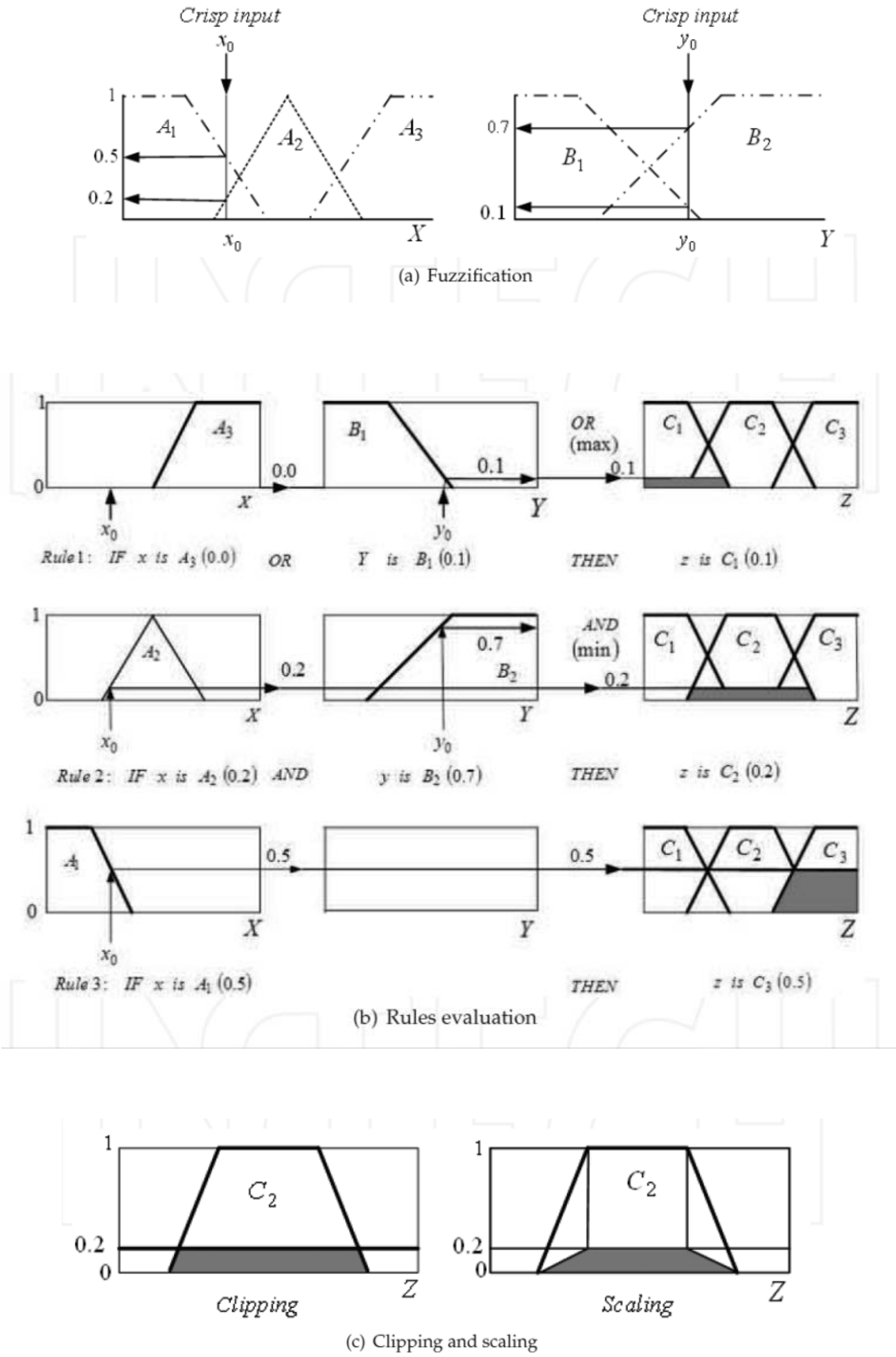


Figure 2.15: Mamdani fuzzy logic controller step 1-2

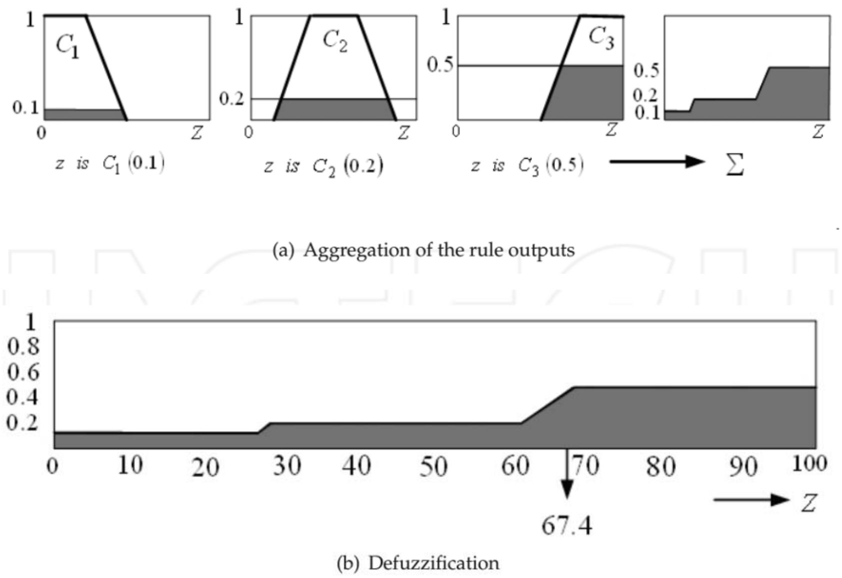


Figure 2.16: Mamdani fuzzy logic controller step 3-4

2.12 Customer Segmentation Overview

Customer segmentation (also known as "market segmentation) takes place when we want to divide customers which bear similarity of some characteristics to each other into groups. For example, for an online pet supply store, its customers can be very diverse with dynamics of needs and preferences. Each customer also has different lifestyles and earns different income. Most importantly, they interact with their pets differently. From a market perspective, a pet supply store can't communicate with its entire customer base in the same way. However, customer segmentation doesn't only signify the product offers which please customers, but also demonstrates different ways of communication to the customers based on what you learn about them. It can also be applied to identify the most profitable customers and choose the most appropriate products which are capable of fulfilling their specific needs. On top of that, customer segmentation can eventually raise awareness of brand loyalty among customers thanks to relevant shopping experiences.

When segmentation is in process, the customers will be split up into groups. The members of each group share similar behaviors, characteristics or needs. Segmentation can be divided into several levels including geographic, demographic, psychographic, and behavioral. For instance, a segment can refer to “all customers who live in San Francisco”, or “All customers who purchased handbags”.

The customers must be classified into groups reasonably with two main functions of the market which are to improve customer acquisition and retention. The form of segmentation should be based on the one objective – to have a better understanding of customers so that we can improve customer acquisition and retention.

Segmentation lies somewhere in the middle of the spectrum of customer differentiation ranging from the entire customer base (no differentiation at all) to complete personalization (differentiation on the individual level).

You can adopt segmentation to alter your retention and conversion marketing campaigns (via email, promotions or other ways) to specific customer segments. You can customize messages to promote different products such as Sale on Men Shoes”, “Sale on Women Shoes”, and “Sale on Kids Shoes” to customers who previously bought relevant products or have otherwise expressed interest in relevant product segments instead of blasting the same “Sale” message to all customers. And because each customer has their own story, we shouldn’t create segments of size one but segments that are “similar within and different across”. At the same time, “keeping them REAL” — Relevant, Efficient, Actionable and Lasting shouldn’t be overlooked.

- Relevant: Segments ideally give explanation of difference in customer behavior from various attributes, meaning the uniqueness of the segment should be capable of being explained properly, not just co-occurring with it. For instance, a geographic segmentation that results in customers from Miami buying more beach towels than those from Montreal sounds reasonable than those from Montreal is better than a geographic segmentation that results in customers from Miami buying more yellow highlighters than those from Montreal.
- Efficient: As mentioned earlier, we shouldn’t create too granular segments because they are impractical and incapable of specifying common behaviors and preferences. Because segments’ number will never be correct, we should create as few segments as possible to prevent outnumbered overhead in association with treating the segments differently.

- Actionable: Likewise, ideal marketers can measure and respond to the dimensions of segments. For example, an online car rental site may rely on anecdotal evidence to presume that taller-than-average customers have tendencies to prefer sedans with extra legroom to compact cars. However, the segmentation on customer height abruptly becomes pretty useless, if the company isn't aware that up front can't collect that information easily
- Lasting: What bring along a good segmentation are dimensions that will remain relatively stable over time. Marketers cannot rely on the segmentation of customers based on how they responded to one-time promotion to categorize other groups of customers who weren't exposed to that promotion. It's also meaningless to compare between the performance of these segments overtime and the customers in that segment that are changing very quickly.

2.13 Segmentation Criteria

- **Identifiable:** Each attribute of the segments can ideally be measured because they are identifiable.
- **Accessible:** The segments must be accessible by the implementation of communication and distribution channels.
- **Sizeable:** The good segments are large enough so that they are capable of demonstrating the resources required to target them. Too small segment may not be able to achieve the requirement.
- **Profitable:** Sizable segments still not come in handy enough, if they are not profitable.
- **Unique needs:** The segments must be in response accordingly to differences in marketing combination in order to justify diverse offerings.
- **Durable:** In an effort to save cost of changes, the segments are expected to be relatively stable.
- **Measurable:** The results of the segments use as well as the impact on specific marketing groups must be able to be measured.

- **Compatible:** Segments must be compatible with the company's capacities and resources.

2.14 Benefits of Segmentation

Custom segmentation based on their similar preference can offer a lot of benefits as follows.

- To help split up the groups of most and least profitable customers.
- To help us target the customers who have higher tendencies to buy the products and services.
- To count the strategies this may not be profitable enough out from the business.
- To raise awareness of loyalty among customers by optimizing and supplying products and services in response to their preference.
- To raise standard of customer service
- To be an effective weapon against other competitors in specific market areas.
- To help to apply the resources sufficiently.
- To conjure up the ideas to offer new products.
- To develop products especially for customers satisfaction.
- To make profit and minimize costs. If it's wisely applied, the customers are even willing to buy the products and services at higher price.
- To gives out several factors to classify the customers such as geographical location, size and sort of organization, customers lifestyles, as well as customers attitudes and behaviors.

2.15 Literature Review

Andreas Hermann published "Applying decision trees for value-based customer relations management: Predicting airline customers' future values" in 2007 which addressed a method of customers segmentation based on lifetime value and prediction of their future variables, considering from demographic and behavioral features. A number of scientific studies concerning question without single-customer dataset were brought alongside the problem applying a decision tree. The way was considered very rare then. The methodology was embarked into a group of customers from a major European airline in a bid to generate prediction in long-, middle-, and short-run, and evaluate the levels of difficulty in predicting the value of each customer in consequence of inadequate inputs. In general, the marketing manager applies the method to develop customer equity in the organization as well as increase returns on marketing in long-run.

Later in 2008, Nattapong applied fuzzy logic to come up with a model of real estate in Thailand. The mean was proved to efficiently troubleshoot any value in vacant land and submarkets. The clustering methods derived from classic (or crisp) set theory were also adapted. The banker would keep track of inspection result clarifying details of such land's location and condition in order to compare between such valuation data and others before classifying the value based on its significance, such as, A, B, C, D. However, some criteria especially weight-related details of those properties can still not be classified clearly. The selection of algorithm in an effort to apply fuzzy clustering to 101 metropolitan areas in Thailand becomes a key agenda to be deliberated.

In 2010, Patinee proposed her analysis. Likewise, she used fuzzy logic to model a merchant discount rate for a kind of credit card service in a commercial bank. In general, the bank staff are assigned to search for and gather up information for the further analysis of eligible factors for the most satisfactory rate. As a result, each bank has different merchandise discount rate (MDR). In order to have the lowest rate, the merchants require different MDR. The issue had gradually had impact on banking industry, which later raised demand in the method to analyze the merchants for the lowest MDR, while they also had to attentively consider issues of costs control and strategies against competitors. Patinee's research discovers a way which applies fuzzy logic for merchant discount factors analysis. Six industries in Thailand, namely, automobiles, apparels, clinics, restaurants, materials, and schools were taken as samples. The MDR analysis focuses on four areas which are sales volume, over drown, deposit, and time-period. The experiment of the model border resulted in 94% higher MDR for each bank. The merchants can be evaluated potentially by the use of MDR.

In 2012, Kit Yan Chan conducted study into market segmentation as well as ideal point identification in an effort to optimize new products by applying fuzzy data compression and fuzzy clustering methods. A number of methodologies were altered alternatively for specific market segmentation. Each group of customers was divided based on their similar demands into clusters. The main point of market segmentation is to figure out the eligible products design in response to the customers' needs and preference. The secondary objective is to determine strategy in an effort to fulfill requirements of customers. Still, other methodologies unfortunately overlook the fuzziness on customers' demands. His publication addressed a new method of market segmentation considering from customers' requirement which involves fuzziness. The method is enabled by synchronizing a fuzzy compression technique for multiple dimension reduction with a fuzzy clustering technique. The fuzzy data stating the customers requirement were firstly compressed from multiple to only two dimensions. The central points of market segmentation become a role model for product optimization, after the fuzzy data is clustered into segments of the market. Kit Yan Chan took a case study of a new way of design for camera to conduct more research on his methodology which later became very effective in market segmentation and the way to identify ideal points for new product design.

In 2012, Jinsoo Hwang and Young Gin Choi studied an issue named "Customer Segmentation Based on Dining Preferences in Full-Service Restaurants". The purpose of their studies was to demonstrate factors which affect five different groups of customers based on decision factors (menu, atmosphere, price, health, and brand reputation). The customers take these factors into consideration before they choose a full-service restaurant. The Decision Tree Analysis was applied properly to specify differences between each group of customers in order to focus only the most desirable ones. The researchers asked 390 questions in total to different respondents to record the data analysis and realized from the result that the variables are classified differently according to the group. In the manner of speaking, five groups have their specific attributes. The marketers or restaurants managers found this method very useful as they can understand more about customer segmentation.

In 2012, Shui Hua Ha, Shui Xiu Lu and Stephen C.H. Leung altered decision tree model to study customer value, which thereafter crystalized "Segmentation of telecom customers based on customer value by decision tree model" research. Their study was focused on the marketing paradigm in the telecom services marketing in an effort to sustain desirable customers. Traditionally, Both customers' future revenue and the cost of servicing customers of different type don't play a major role in customer segmentation methods based on experience or ARPU (Average Revenue per User). As a consequence, it's not easy to specifically point out desirable customers. The publication considers customer lifecycle to deploy a novel customer segmentation method. Five decision models including current value, historic value, prediction of long-term value, credit and loyalty are in need to activate the methods for the presentation of a judgment matrix calculated from characteristics of data and the experience of business experts. When the model is properly used, the researchers could establish an easy and effective customer value assessment system. The telecom operators in a province in China applied the methodology and successfully accomplished satisfactory result.

CHAPTER III
RESEARCH METHODOLOGY

The purpose of this study is to examine the customer segmentation in the international logistic services business, and study the use of the “decision tree” and “fuzzy logic” for helping salesperson to make the decision on customer segmentation.

The two purposes of this chapter are to (1) describe the materials use for this study (2) describe the population of this study, and (3) describe the procedure used in “decision tree” and “fuzzy logic”.

3.1 Materials

3.1.1 Hardware

Personal Computer

- CPU	:	AMD Dual-Core A4-3300M
- RAM	:	4 GB
- Hard Disk	:	500 GB
- Monitor	:	Acer

3.1.2 Software

- Operation System	:	WINDOW 7
- Programming Language	:	WEKA
- Statistical Analysis	:	WEKA
- Documentation	:	MS EXCEL

3.2 Populations

Populations of this research are the customer data from international shipping companies. Totally 20,280 data are take in to the analysis.

3.3 Customer Segmentation Criteria

Customer Segmentation Criteria obtained from survey method by sending questionnaire to 30 sale persons. The three highest criteria was chosen to be the criteria for customer segmentation in this research.

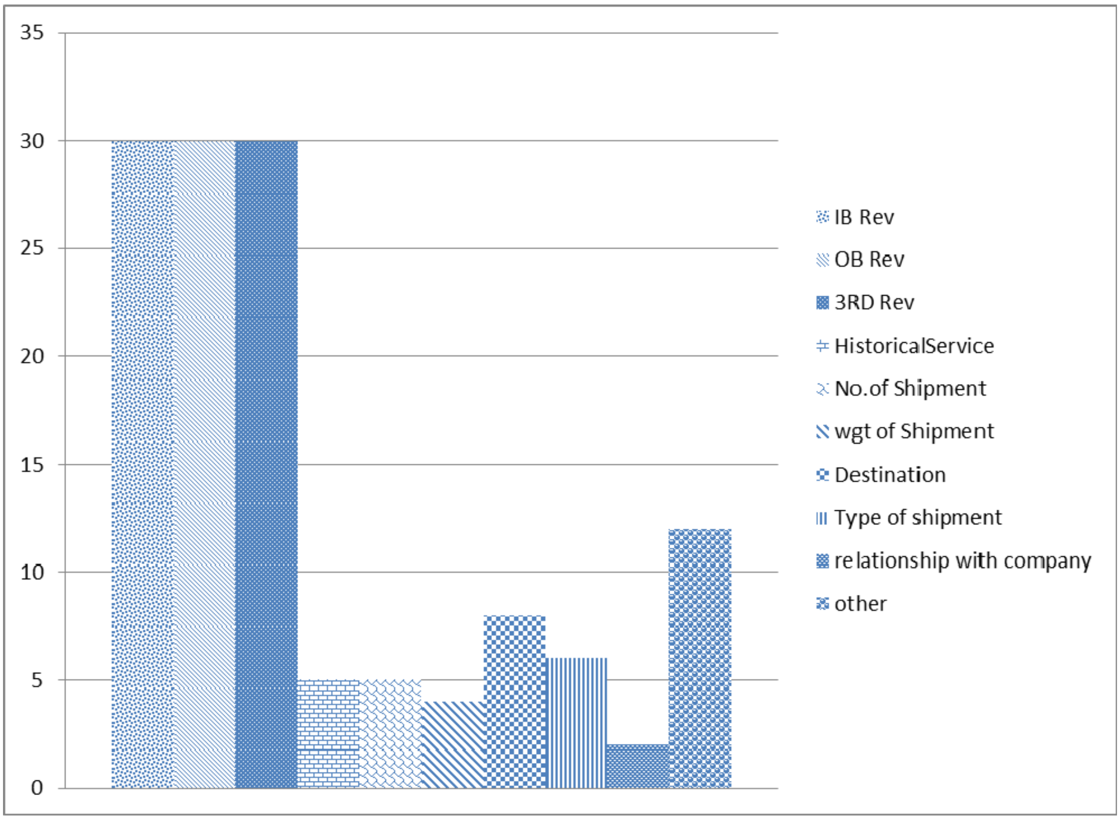


Figure 3.1: Shown that the highest three criteria was selected as follows; Inbound Revenue (IB), Outbound Revenue (OB) and Third Party Revenue

Due to the information reviewed from an international logistic company in Thailand, there were a few criteria, mainly sale revenue that effects the customer segmentation.

1) Outbound revenue means amounts that a company receives from customers that using exporting shipment services.

2) Inbound revenue means amounts that a company receives from customers that are using import shipment services.

3) Third party revenue means amounts that a company receives from customers that are neither shipper nor receiver.

3.4 Step of Work

Research procudure as follow:

3.4.1 Decision Tree

Process of creating Decision Tree Model is shown in Figure 3.2. Due to the complex of data in this research, WEKA software was apply for the analysis.

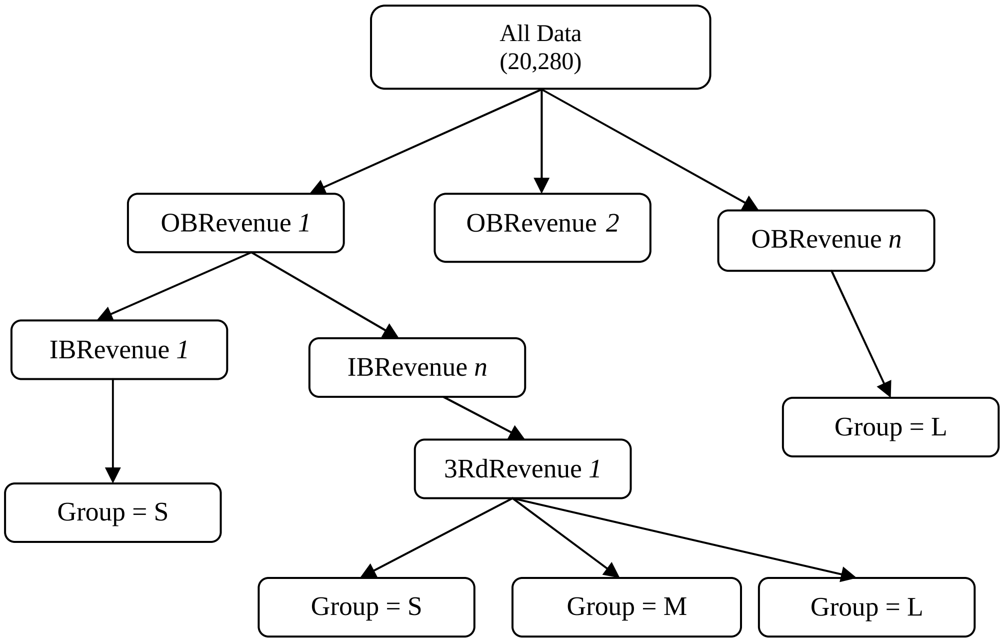


Figure 3.2 : Process of creating Decision Tree Model

3.4.2 Fuzzy Logic

1) Variable definition of 3 input variables and 1 output variable as shown in table 3.1.

Table 3.1: Variable and categories

Item	Variable	Variable meaning	categories	Categories meaning
Input #1	Outbound Revenue (OB Rev)	Amounts that company received from customers that using the exporting shipment services.	L (Low) M (Medium) H (High)	OB Rev Low OB Rev Medium OB Rev High
Input#2	Inbound Revenue (IB Rev)	Amounts that company received from customers that using the importing shipment services.	L (Low) M (Medium) H (High)	IB Rev Low IB Rev Medium IB Rev High
Input#3	Third Party Revenue (3^{rd} Rev)	Amounts that company received from customers that not either shipper or receiver.	L (Low) M (Medium) H (High)	3^{rd} Rev Low 3^{rd} Rev Medium 3^{rd} Rev High
Output	Customer Segmentation	System output	S (Small) M (Medium) L (Large)	Small Group Medium Group Large Group

2) Making rule for analysis of Fuzzy Logic (Table 3.2)

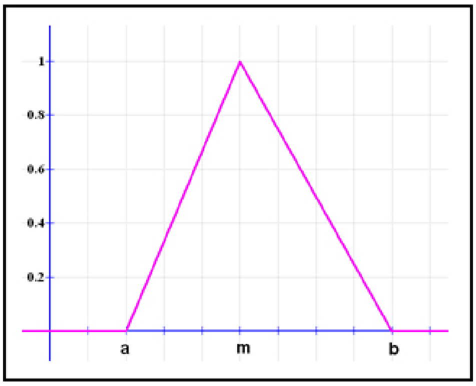
Table3.2 : Shows the variable determination including 3 input variables and 1 output variable.

RULE TEXT			
1	IF	Outbound Revenue	= Low
	THEN	Customer Segmentation	= Small
2	IF	Outbound Revenue	= Medium
	THEN	Customer Segmentation	= Medium
3	IF	Outbound Revenue	= High
	THEN	Customer Segmentation	= Large
4	IF	Inbound Revenue	= Low
	THEN	Customer Segmentation	= Small
5	IF	Inbound Revenue	= Medium
	THEN	Customer Segmentation	= Medium
6	IF	Inbound Revenue	= High
	THEN	Customer Segmentation	= Large
7	IF	Third Party Revenue	= Low
	THEN	Customer Segmentation	= Small
8	IF	Third Party Revenue	= Medium
	THEN	Customer Segmentation	= Medium
9	IF	Third Party Revenue	= High
	THEN	Customer Segmentation	= Large

3) Type of Member ship function

This research using Triangular function defined by a lower limit a, an upper limit b, and a value m, where a < m < b.

$$\mu_A(x) = \begin{cases} 0, & x \leq a \\ \dfrac{x-a}{m-a}, & a < x \leq m \\ \dfrac{b-x}{b-m}, & m < x < b \\ 0, & x \geq b \end{cases}$$



From table 3.3 we can determine the membership function that effect Customer Segmentation as shown in Figure 3.3 (a), (b), (c) and Output membership function shown in Figure 3.3 (d)

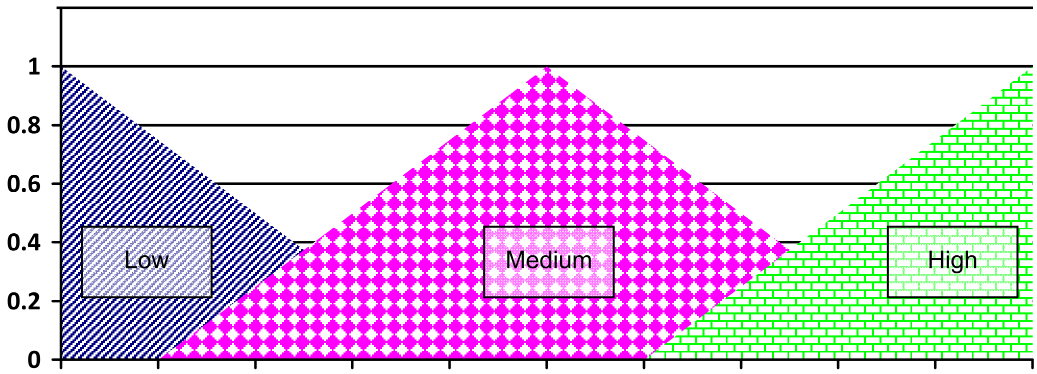


Figure 3.3: (a) X1 OB revenue

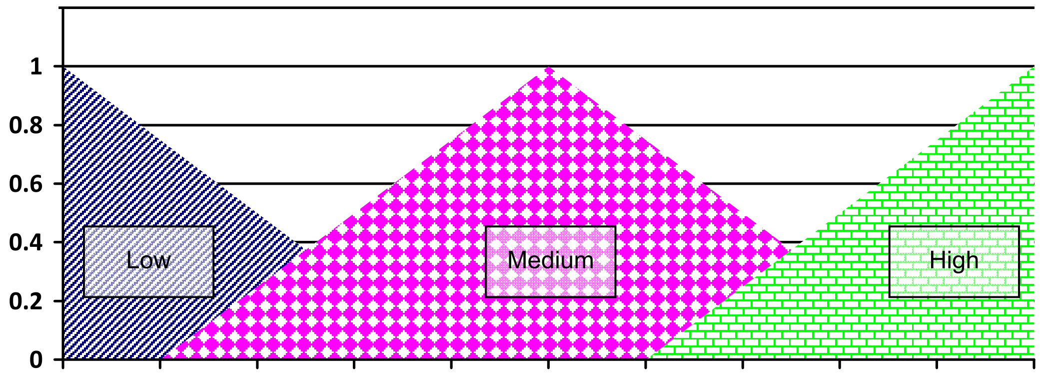


Figure 3.3: (b) X2 IB revenue

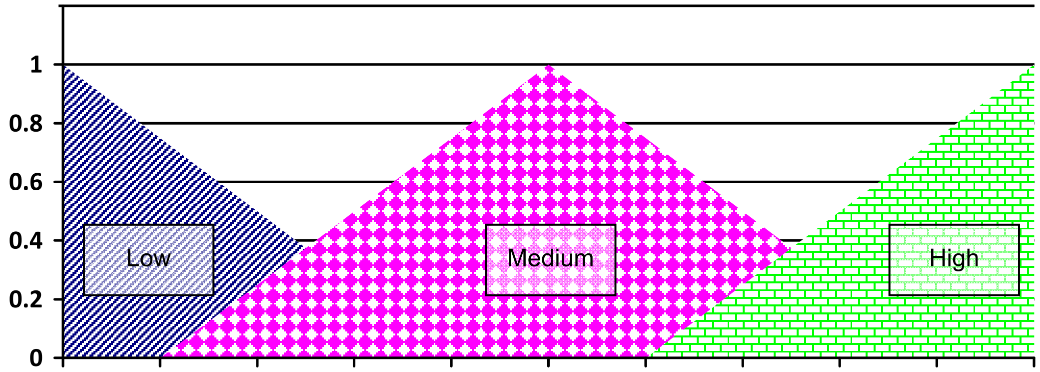


Figure 3.3: (c) X3B3P revenue

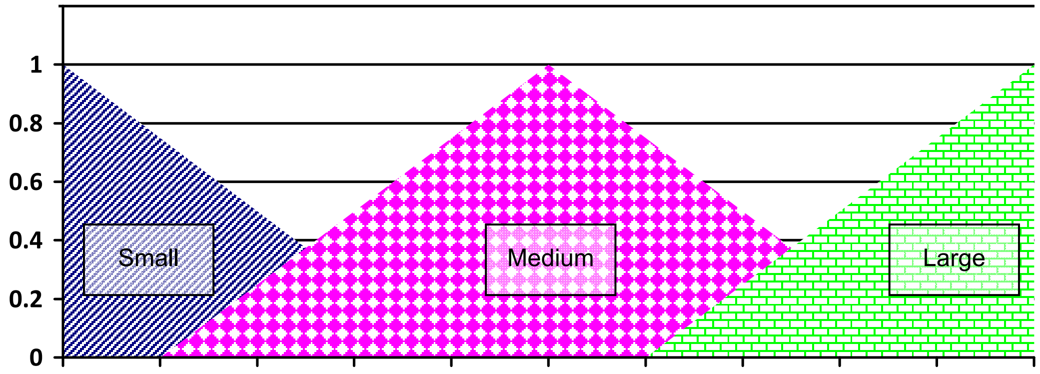


Figure 3.3: (d) Outputs for effective factor of customer segmentation

4) Fuzzy logic procedures.

Fuzzy control, which directly uses fuzzy rules, is the most important application in fuzzy theory. Using a procedure originated by Ebrahim Mamdani in the late 70s, four steps are taken to create a fuzzy controlled machine. Figure 3.4 shows 4 steps for processing fuzzy logic, which can be described as follows.

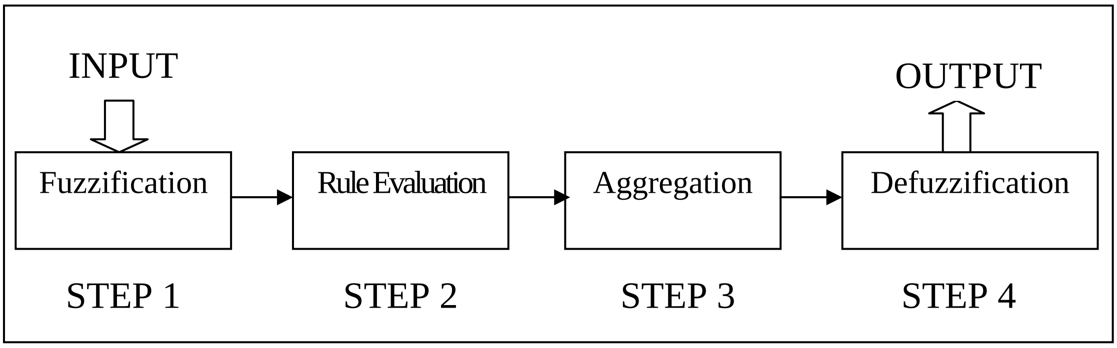


Figure 3.4: Four steps for processing fuzzy logic

Step 1: Fuzzification

This step comprises the process of transforming crisp inputs, outbound revenue, inbound revenue and third-party revenue into the degree to find out which appropriate fuzzy sets these inputs should belong to.

Step 2: Rule Evaluation

Rule evaluation applies fuzzified inputs to the premises of the fuzzy rules. In case of multiple premises, the AND/OR fuzzy operator is applied to get a single number representing the result of premise assessment. This number is then applied to the consequent membership function.

Step 3: Aggregation of the rule outputs

In order to aggregate the output, the membership functions of all rule which were previously clipped or scaled are compressed into a single fuzzy set. The input and output of aggregation process represents the clipped or scaled consequent membership functions, and the fuzzy set for each output, respectively.

Step 4: Defuzzification

Lastly, the defuzzification step is capable of evaluating the rules under a strict condition; the final output must be crisped into a single number. The input of the process is the aggregate output fuzzy set, while the output is a single number.

3.4.3 Analysis Process

Since the data preparation in above process was done. WEKA software was apply for the analysis.

CHAPTER IV
RESULTS AND DISCUSSION

This chapter will conclude about the result of applying Fuzzy logic model for customer segmentation. The details are below

4.1 Experimental Results

Threr are 20,208 data that was input to weka programe as in Fugure 4.1 and the detail of the data in each criteria is in table 4.1

Table 4.1 : Input data to Waka

Class	Description	Number of Data
A	Small	10,644
B	Medium	5,548
C	Large	4,088
	Total	**20,280**

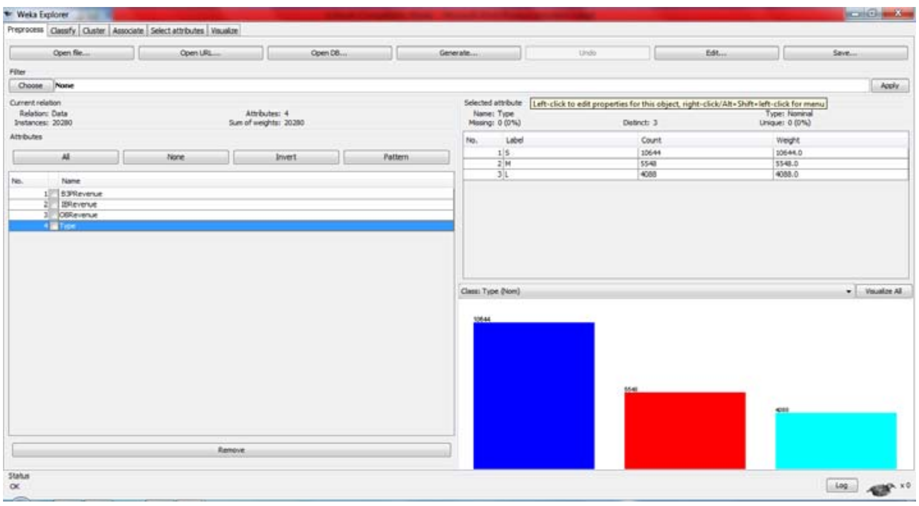


Figure 4.1: input data process to Weka

Analysis by J48 (evaluate on training data) shows that the use of one input feature is sufficient for the differentiation between three groups with overall classification accuracy of 95.18% there is 52 number of leaves and 103 size of the tree. The confusion matrix of J48 with evaluate on training data as in the Table 4.2

Table 4.2: The confusion matrix of J48 with evaluate on training data.

Actual Class	Identified class		
	S	M	L
S	10,524	120	0
M	352	5,101	95
L	184	225	3,679

The detailed accuracy by class is showed in table 4.3

Table 4.3: The detailed accuracy by class of J48 with evaluate on training data.

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.989	0.056	0.952	0.989	0.97	0.978	S
	0.919	0.023	0.937	0.919	0.928	0.971	M
	0.9	0.006	0.975	0.9	0.936	0.965	L
Weighted Avg.	0.952	0.037	0.952	0.952	0.952	0.973	

Analysis by J48 (10-fold cross-validation) shows that the use of one input feature is sufficient for the differentiation between three groups with overall classification accuracy of 94.43% there is 52 number of leaves and 103 size of the tree. The confusion matrix of J48 with 10-fold cross-validation as in the Table 4.4

Table 4.4 : The confusion matrix of J48 with 10-fold cross-validation.

Actual Class	Identified class		
	S	**M**	**L**
S	10,491	151	2
M	371	5,037	140
L	181	283	3,624

The detailed accuracy by class is showed in table 4.5

Table 4.5 : The detailed accuracy by class of J48 with 10-fold cross-validation.

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.986	0.057	0.95	0.986	0.967	0.975	S
	0.908	0.029	0.921	0.908	0.914	0.964	M
	0.886	0.009	0.962	0.886	0.923	0.959	L
Weighted Avg.	0.944	0.04	0.944	0.944	0.944	0.969	

Analysis by JRip (evaluate on training data) shows that the use of one input feature is sufficient for the differentiation between three groups with overall classification accuracy of 94.98% there are 17 number of rules . The confusion matrix of JRip with evaluate on training data as in the Table 4.6

Table 4.6 : The confusion matrix of JRip with evaluate on training data.

Actual Class	Identified class		
	S	**M**	**L**
S	10,526	118	0
M	358	5,081	109
L	188	245	3,655

The detailed accuracy by class is showed in table 4.7

Table 4.7 : The detailed accuracy by class of JRip with evaluate on training data.

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.989	0.057	0.951	0.989	0.969	0.971	S
	0.916	0.025	0.933	0.916	0.924	0.963	M
	0.894	0.038	0.971	0.894	0.931	0.96	L
Weighted Avg.	0.95	0.038	0.95	0.95	0.949	0.967	

Analysis by JRip (10-fold cross-validation) shows that the use of one input feature is sufficient for the differentiation between three groups with overall classification accuracy of 94.51% there are 17 number of rules. The confusion matrix of JRip 10-fold cross-validation as in the Table 4.8

Table 4.8 : The confusion matrix of JRip with 10-fold cross-validation.

Actual Class	Identified class		
	S	M	L
S	10,498	146	0
M	379	5,046	123
L	185	281	3,622

The detailed accuracy by class is showed in table 4.9

Table 4.9 : The detailed accuracy by class of JRip with 10-fold cross-validation.

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.986	0.059	0.949	0.986	0.967	0.97	S
	0.91	0.029	0.922	0.91	0.916	0.96	M
	0.886	0.008	0.967	0.886	0.925	0.956	L
Weighted Avg.	0.945	0.04	0.945	0.945	0.945	0.964	

Analysis by FLR (evaluate on training data) shows that the use of one input feature is sufficient for the differentiation between three groups with overall classification accuracy of 91.04% and the result of the rule as following.

Total Number of Rules: 4

Rules pointing in Class S :1

Rules pointing in Class M :1

Rules pointing in Class L :2

The confusion matrix of FLR with evaluate on training data as in the Table 4.10 below

Table 4.10 : The confusion matrix of FLR with evaluate on training data.

Actual Class	Identified class		
	S	M	L
S	10,644	0	0
M	942	4,606	0
L	207	668	3,213

The detailed accuracy by class is showed in table 4.11

Table 4.11 : The detailed accuracy by class of FLR with evaluate on training data.

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0.119	0.903	1	0.949	0.94	S
	0.83	0.045	0.873	0.83	0.851	0.892	M
	0.786	0	1	0.786	0.88	0.893	L
Weighted Avg.	0.91	0.075	0.914	0.91	0.908	0.918	

Analysis by FLR (10-fold cross-validation) shows that the use of one input feature is sufficient for the differentiation between three groups with overall classification accuracy of 58.15% and the result of the rule as following.

Total Number of Rules: 4

Rules pointing in Class S :1

Rules pointing in Class M :1

Rules pointing in Class L :2

The confusion matrix of FLR with 10-fold cross-validation as in the Table 4.12 below

Table 4.12 : The confusion matrix of FLR with 10-fold cross-validation.

Actual Class	Identified class		
	S	M	L
S	4,257	5,323	1,064
M	367	4,155	1,026
L	87	620	3,381

The detailed accuracy by class is showed in table 4.13

Table 4.13 : The detailed accuracy by class of FLR with 10-fold cross-validation.

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.4	0.047	0.904	1	0.949	0.94	S
	0.749	0.403	0.411	0.83	0.851	0.892	M
	0.827	0.129	0.618	0.786	0.88	0.893	L
Weighted Avg.	0.582	0.161	0.711	0.91	0.908	0.918	

Analysis by j48, JRip between 10-fold cross-validation and training data found that Correctly Classified Instances is greater than 90% and 90 % of ROC area. Which means there was a congruence between 10-fold cross-validation and training data.

Whereas, the analysis by FLR (evaluate on training data) the Correctly Classified Instances is 91.04%. Using the analysis by FLR (10-fold cross-validation) the Correctly Classified Instances is 58.15%. Both FLR and j48/JRip was show the ROC Area of 91.80%.

Discritization by Weka was shown in Figure 4.2

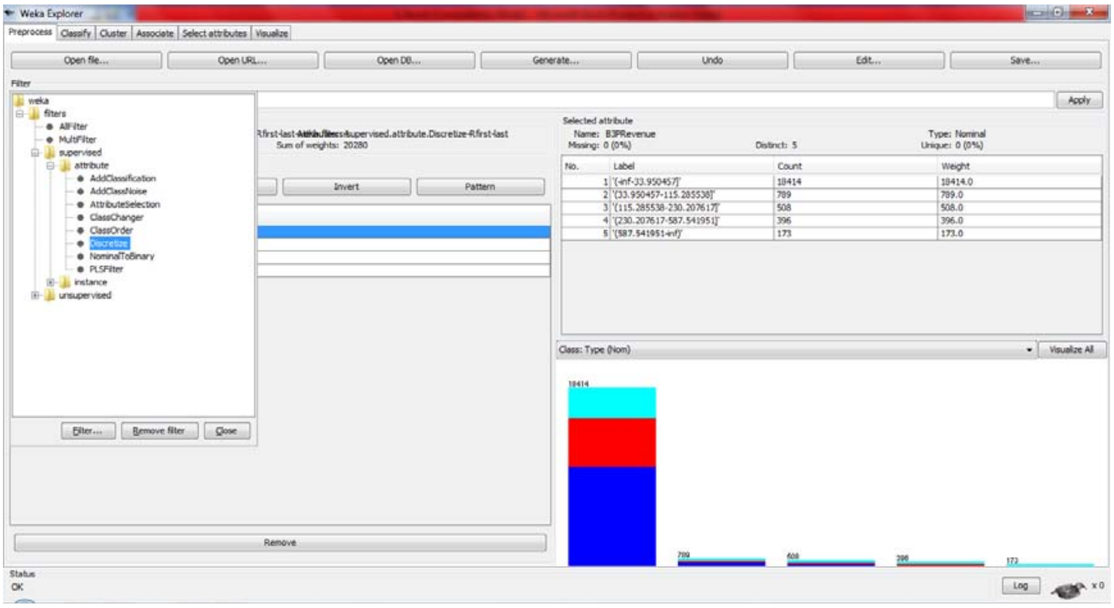


Figure 4.2 : Discretize data before classify

Analysis by J48 (10-fold cross-validation) shows that the use of one input feature is sufficient for the differentiation between three groups with overall classification accuracy of 94.57% there is 143 number of leaves and 168 size of the tree. The confusion matrix of J48 with 10-fold cross-validation as in the Table 4.14

Table 4.14 : The confusion matrix of J48 with 10-fold cross-validation.

Actual Class	Identified class		
	S	M	L
S	10,521	123	0
M	397	5,053	98
L	190	294	3,604

The detailed accuracy by class is showed in table 4.15

Table 4.15 : The detailed accuracy by class of J48 with 10-fold cross-validation.

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.988	0.061	0.947	0.988	0.967	0.981	S
	0.911	0.028	0.924	0.911	0.917	0.976	M
	0.882	0.006	0.974	0.882	0.925	0.963	L
Weighted Avg.	0.946	0.041	0.946	0.946	0.945	0.976	

Analysis by JRip (10-fold cross-validation) shows that the use of one input feature is sufficient for the differentiation between three groups with overall classification accuracy of 94.40% there are 34 number of rules. The confusion matrix of JRip 10-fold cross-validation as in the Table 4.16

Table 4.16 : The confusion matrix of JRip with 10-fold cross-validation.

Actual Class	Identified class		
	S	M	L
S	10,533	111	0
M	430	5,029	89
L	218	287	3,583

The detailed accuracy by class is showed in table 4.17

Table 4.17 : The detailed accuracy by class of JRip with 10-fold cross-validation.

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.99	0.067	0.942	0.99	0.967	0.965	S
	0.906	0.027	0.927	0.906	0.916	0.96	M
	0.876	0.005	0.976	0.876	0.923	0.955	L
Weighted Avg.	0.944	0.044	0.945	0.944	0.943	0.961	

The above result found that there was no different of correctly classified Instances between discretized and non-discretized data. However, non-discretized data was more practical due to generate smaller size of model in decision tree method.

The tree of revenue based model as detail below.

```
J48 pruned tree
------------------

OBRevenue <= 229.3091
|   IBRevenue <= 230.074543
|   |   B3PRevenue <= 230.023771
|   |   |   OBRevenue <= 159.643974
|   |   |   |   IBRevenue <= 161.781164: S (9981.0/361.0)
|   |   |   |   IBRevenue > 161.781164
|   |   |   |   |   OBRevenue <= 55.955057: S (431.0/59.0)
|   |   |   |   |   OBRevenue > 55.955057
|   |   |   |   |   |   IBRevenue <= 184.009415
|   |   |   |   |   |   |   OBRevenue <= 138.468671
|   |   |   |   |   |   |   |   B3PRevenue <= 42.382383: S (58.0/18.0)
|   |   |   |   |   |   |   |   B3PRevenue > 42.382383: M (4.0/1.0)
|   |   |   |   |   |   |   OBRevenue > 138.468671: M (13.0/4.0)
|   |   |   |   |   |   IBRevenue > 184.009415: M (116.0/38.0)
|   |   |   OBRevenue > 159.643974
|   |   |   |   IBRevenue <= 56.11
|   |   |   |   |   OBRevenue <= 204.231778: S (435.0/35.0)
|   |   |   |   |   OBRevenue > 204.231778
|   |   |   |   |   |   B3PRevenue <= 0: S (155.0/63.0)
|   |   |   |   |   |   B3PRevenue > 0
|   |   |   |   |   |   |   B3PRevenue <= 49.648278: L (3.0/1.0)
|   |   |   |   |   |   |   B3PRevenue > 49.648278: M (6.0/1.0)
|   |   |   |   IBRevenue > 56.11: M (230.0/44.0)
|   |   B3PRevenue > 230.023771
```

```
| | | B3PRevenue <= 587.140638
| | | | B3PRevenue <= 449.512076: M (89.0/5.0)
| | | | B3PRevenue > 449.512076
| | | | | OBRevenue <= 12.477992
| | | | | | B3PRevenue <= 459.373256: L (3.0/1.0)
| | | | | | B3PRevenue > 459.373256: M (6.0)
| | | | | OBRevenue > 12.477992
| | | | | | OBRevenue <= 103.99053: L (4.0)
| | | | | | OBRevenue > 103.99053: M (3.0/1.0)
| | | B3PRevenue > 587.140638: L (39.0)
| IBRevenue > 230.074543
| | IBRevenue <= 569.155701
| | | B3PRevenue <= 488.65
| | | | IBRevenue <= 435.480444: M (1643.0/70.0)
| | | | IBRevenue > 435.480444
| | | | | OBRevenue <= 36.753087
| | | | | | B3PRevenue <= 97.15034: M (252.0/19.0)
| | | | | | B3PRevenue > 97.15034: L (11.0/4.0)
| | | | | OBRevenue > 36.753087
| | | | | | OBRevenue <= 85.878644
| | | | | | | B3PRevenue <= 50.86: M (33.0/15.0)
| | | | | | | B3PRevenue > 50.86: L (4.0)
| | | | | | OBRevenue > 85.878644: L (49.0/6.0)
| | | B3PRevenue > 488.65: L (26.0/1.0)
| | IBRevenue > 569.155701: L (1040.0/16.0)
OBRevenue > 229.3091
| OBRevenue <= 569.865642
| | IBRevenue <= 301.707798
| | | OBRevenue <= 506.28281
| | | | IBRevenue <= 114.301341: M (2545.0/86.0)
| | | | IBRevenue > 114.301341
| | | | | OBRevenue <= 394.51023
```

```
|   |   |   |   |   |   B3PRevenue <= 95.44
|   |   |   |   |   |   |   IBRevenue <= 241.72288: M (235.0/5.0)
|   |   |   |   |   |   |   IBRevenue > 241.72288
|   |   |   |   |   |   |   |   OBRevenue <= 343.983503: M (54.0/4.0)
|   |   |   |   |   |   |   |   OBRevenue > 343.983503
|   |   |   |   |   |   |   |   |   OBRevenue <= 384.063882: L (12.0/4.0)
|   |   |   |   |   |   |   |   |   OBRevenue > 384.063882: M (3.0)
|   |   |   |   |   |   B3PRevenue > 95.44
|   |   |   |   |   |   |   IBRevenue <= 124.153679: L (3.0)
|   |   |   |   |   |   |   IBRevenue > 124.153679
|   |   |   |   |   |   |   |   B3PRevenue <= 205.33: M (10.0)
|   |   |   |   |   |   |   |   B3PRevenue > 205.33: L (4.0/1.0)
|   |   |   |   |   OBRevenue > 394.51023
|   |   |   |   |   |   B3PRevenue <= 35.8
|   |   |   |   |   |   |   B3PRevenue <= 12.03992
|   |   |   |   |   |   |   |   OBRevenue <= 475.564469
|   |   |   |   |   |   |   |   |   IBRevenue <= 169.657211: M (16.0/2.0)
|   |   |   |   |   |   |   |   |   IBRevenue > 169.657211: L (22.0/7.0)
|   |   |   |   |   |   |   |   OBRevenue > 475.564469: L (8.0)
|   |   |   |   |   |   |   B3PRevenue > 12.03992: M (2.0)
|   |   |   |   |   |   B3PRevenue > 35.8: L (8.0)
|   |   |   OBRevenue > 506.28281
|   |   |   |   IBRevenue <= 82.113808
|   |   |   |   |   OBRevenue <= 522.514021: M (36.0/3.0)
|   |   |   |   |   OBRevenue > 522.514021
|   |   |   |   |   |   IBRevenue <= 41.511355: M (117.0/44.0)
|   |   |   |   |   |   IBRevenue > 41.511355: L (6.0/1.0)
|   |   |   |   IBRevenue > 82.113808: L (36.0)
|   |   IBRevenue > 301.707798
|   |   |   IBRevenue <= 560.750117
|   |   |   |   OBRevenue <= 335.673714
|   |   |   |   |   B3PRevenue <= 72.277565
```

| | | | | | IBRevenue <= 351.401345: M (23.0/3.0)

| | | | | | IBRevenue > 351.401345

| | | | | | | OBRevenue <= 237.535639: L (5.0)

| | | | | | | OBRevenue > 237.535639

| | | | | | | | OBRevenue <= 257.583038: M (8.0)

| | | | | | | | OBRevenue > 257.583038: L (32.0/15.0)

| | | | | B3PRevenue > 72.277565

| | | | | | OBRevenue <= 256.604242

| | | | | | | OBRevenue <= 245.228656: L (3.0)

| | | | | | | OBRevenue > 245.228656: M (2.0)

| | | | | | OBRevenue > 256.604242: L (10.0)

| | | | OBRevenue > 335.673714: L (90.0/16.0)

| | | IBRevenue > 560.750117: L (241.0)

| OBRevenue > 569.865642: L (2115.0/22.0)

CHAPTER V
CONCLUSION

This chapter will describe the conclusion of applying decision tree model and fuzzy logic model with the customer segmentation and also so the recommendation for further study. The detail as follows:

5.1 Conclusion

This research purposed the method for making decision about customer segmentation by using decision tree and fuzzy logic. The model is able to classify the customer in the following groups; large size, medium size and small size by using 3 indicators which are inbound revenue, outbound revenue and third party revenue. Data was taken from the database of international logistic company, 3 indicators obtained by the focus group of 20 senior sale staff of that company.

This study also applies the concurrent validity using the known group technic. The results showed that the number of customer that segment by the theory of decision tree and fuzzy logic is congruence with the decision made by the sale person. Decision tree represent model of tree with 52 number of leaves which 94.44 % of correctly classified instances and 96.9% of ROC area and fuzzy logic represent model with 58.15% of correctly classified instances and 71% of ROC area, which is acceptable for preliminary segmentations of sales person.

5.2 Future work

5.2.1 The further study needs to compare the customer satisfactions, gain revenue between the segmentation that generated from decision tree model and the segmentation that come from sales person decision.

5.2.2 The further study needs to study other indicators that can affect the customer segmentation apart from inbound revenue, outbound revenue and third party revenue.

REFERENCES

Andreas Herrmann, “Applying decision trees for value-based customer relations management: Predicting airline customers’ future values”, Journal of Database Marketing & Customer Strategy Management (2007) 14, 130 – 142, 2007.

Chan, S. L., & Ip, W. H., “A dynamic decision support system to predict thevalue of customer for new product development”, Decision Support Systems, 52,178–188, 2011.

Jinsoo Hwang and Young Gin Choi, “Customer Segmentation Based on Dining Preferences in Full-Service Restaurants” The Journal of Foodservice Business Research, 15:226–246, 2012.

Nattapong Thipayawat, Adisorn Leelasantitham, Supaporn Kiattisin and Prawat Chaiprapa, “An Appraisal Model of Real Estate in Thailand Using Fuzzy Lattice Reasoning”, The 2009 International Conference on Computer Design and Applications (ICCDA 2009), Singapore, 2009.

Pang-Ning Tan, Michael Steinbach and Vipin Kumar, “Introduction to Data Mining”, Pearson Eduction, In., 2006

Patinee Sakdanupab, Davich Banthao, Adisorn Leelasantitham and Supaporn Kiattisin, “An Analysis Model of Merchant Discount Rate for Services of Credit Card in Commercial Bank Using Fuzzy Logic”, The 2010 IEEE International Conference on Information and Automation, Harbin, China, 2010.

Shui Hua Ha, Shui Xiu Lu and Stephen C.H. Leung, “Segmentation of Telecom Customers Based on Customer Value by Decision Tree Model”, Expert Systems with Applications, Volume 39, Issue 4, March 2012, Pages 3964-3973, 2012.

Kotler, Philip, and Kevin Lane Keller, “What is geographic segmentation”, Marketing Management. Prentice Hall, 2006.

http://dataminingtrend.com/2014/decision-tree-model/

http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol2/jp6/article2.html

APPENDIX

Running data from Weka Program.

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     Data
Instances:    20280
Attributes:   4
              B3PRevenue
              IBRevenue
              OBRevenue
              Type
Test mode:    evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree
------------------

OBRevenue <= 229.3091
|   IBRevenue <= 230.074543
|   |   B3PRevenue <= 230.023771
|   |   |   OBRevenue <= 159.643974
|   |   |   |   IBRevenue <= 161.781164: S (9981.0/361.0)
|   |   |   |   IBRevenue > 161.781164
|   |   |   |   |   OBRevenue <= 55.955057: S (431.0/59.0)
|   |   |   |   |   OBRevenue > 55.955057
|   |   |   |   |   |   IBRevenue <= 184.009415
|   |   |   |   |   |   |   OBRevenue <= 138.468671
|   |   |   |   |   |   |   |   B3PRevenue <= 42.382383: S (58.0/18.0)
|   |   |   |   |   |   |   |   B3PRevenue > 42.382383: M (4.0/1.0)
|   |   |   |   |   |   |   OBRevenue > 138.468671: M (13.0/4.0)
|   |   |   |   |   |   IBRevenue > 184.009415: M (116.0/38.0)
|   |   |   OBRevenue > 159.643974
|   |   |   |   IBRevenue <= 56.11
|   |   |   |   |   OBRevenue <= 204.231778: S (435.0/35.0)
|   |   |   |   |   OBRevenue > 204.231778
|   |   |   |   |   |   B3PRevenue <= 0: S (155.0/63.0)
|   |   |   |   |   |   B3PRevenue > 0
|   |   |   |   |   |   |   B3PRevenue <= 49.648278: L (3.0/1.0)
|   |   |   |   |   |   |   B3PRevenue > 49.648278: M (6.0/1.0)
|   |   |   |   IBRevenue > 56.11: M (230.0/44.0)
|   |   B3PRevenue > 230.023771
|   |   |   B3PRevenue <= 587.140638
|   |   |   |   B3PRevenue <= 449.512076: M (89.0/5.0)
|   |   |   |   B3PRevenue > 449.512076
|   |   |   |   |   OBRevenue <= 12.477992
|   |   |   |   |   |   B3PRevenue <= 459.373256: L (3.0/1.0)
```

```
|   |   |   |   |   |   B3PRevenue > 459.373256: M (6.0)
|   |   |   |   |   OBRevenue > 12.477992
|   |   |   |   |   |   OBRevenue <= 103.99053: L (4.0)
|   |   |   |   |   |   OBRevenue > 103.99053: M (3.0/1.0)
|   |   |   B3PRevenue > 587.140638: L (39.0)
|   IBRevenue > 230.074543
|   |   IBRevenue <= 569.155701
|   |   |   B3PRevenue <= 488.65
|   |   |   |   IBRevenue <= 435.480444: M (1643.0/70.0)
|   |   |   |   IBRevenue > 435.480444
|   |   |   |   |   OBRevenue <= 36.753087
|   |   |   |   |   |   B3PRevenue <= 97.15034: M (252.0/19.0)
|   |   |   |   |   |   B3PRevenue > 97.15034: L (11.0/4.0)
|   |   |   |   |   OBRevenue > 36.753087
|   |   |   |   |   |   OBRevenue <= 85.878644
|   |   |   |   |   |   |   B3PRevenue <= 50.86: M (33.0/15.0)
|   |   |   |   |   |   |   B3PRevenue > 50.86: L (4.0)
|   |   |   |   |   |   OBRevenue > 85.878644: L (49.0/6.0)
|   |   |   B3PRevenue > 488.65: L (26.0/1.0)
|   |   IBRevenue > 569.155701: L (1040.0/16.0)
OBRevenue > 229.3091
|   OBRevenue <= 569.865642
|   |   IBRevenue <= 301.707798
|   |   |   OBRevenue <= 506.28281
|   |   |   |   IBRevenue <= 114.301341: M (2545.0/86.0)
|   |   |   |   IBRevenue > 114.301341
|   |   |   |   |   OBRevenue <= 394.51023
|   |   |   |   |   |   B3PRevenue <= 95.44
|   |   |   |   |   |   |   IBRevenue <= 241.72288: M (235.0/5.0)
|   |   |   |   |   |   |   IBRevenue > 241.72288
|   |   |   |   |   |   |   |   OBRevenue <= 343.983503: M (54.0/4.0)
|   |   |   |   |   |   |   |   OBRevenue > 343.983503
|   |   |   |   |   |   |   |   |   OBRevenue <= 384.063882: L (12.0/4.0)
|   |   |   |   |   |   |   |   |   OBRevenue > 384.063882: M (3.0)
|   |   |   |   |   |   B3PRevenue > 95.44
|   |   |   |   |   |   |   IBRevenue <= 124.153679: L (3.0)
|   |   |   |   |   |   |   IBRevenue > 124.153679
|   |   |   |   |   |   |   |   B3PRevenue <= 205.33: M (10.0)
|   |   |   |   |   |   |   |   B3PRevenue > 205.33: L (4.0/1.0)
|   |   |   |   |   OBRevenue > 394.51023
|   |   |   |   |   |   B3PRevenue <= 35.8
|   |   |   |   |   |   |   B3PRevenue <= 12.03992
|   |   |   |   |   |   |   |   OBRevenue <= 475.564469
|   |   |   |   |   |   |   |   |   IBRevenue <= 169.657211: M (16.0/2.0)
|   |   |   |   |   |   |   |   |   IBRevenue > 169.657211: L (22.0/7.0)
|   |   |   |   |   |   |   |   OBRevenue > 475.564469: L (8.0)
|   |   |   |   |   |   |   B3PRevenue > 12.03992: M (2.0)
|   |   |   |   |   |   B3PRevenue > 35.8: L (8.0)
```

```
|   |   |   OBRevenue > 506.28281
|   |   |   |   IBRevenue <= 82.113808
|   |   |   |   |   OBRevenue <= 522.514021: M (36.0/3.0)
|   |   |   |   |   OBRevenue > 522.514021
|   |   |   |   |   |   IBRevenue <= 41.511355: M (117.0/44.0)
|   |   |   |   |   |   IBRevenue > 41.511355: L (6.0/1.0)
|   |   |   |   IBRevenue > 82.113808: L (36.0)
|   |   IBRevenue > 301.707798
|   |   |   IBRevenue <= 560.750117
|   |   |   |   OBRevenue <= 335.673714
|   |   |   |   |   B3PRevenue <= 72.277565
|   |   |   |   |   |   IBRevenue <= 351.401345: M (23.0/3.0)
|   |   |   |   |   |   IBRevenue > 351.401345
|   |   |   |   |   |   |   OBRevenue <= 237.535639: L (5.0)
|   |   |   |   |   |   |   OBRevenue > 237.535639
|   |   |   |   |   |   |   |   OBRevenue <= 257.583038: M (8.0)
|   |   |   |   |   |   |   |   OBRevenue > 257.583038: L (32.0/15.0)
|   |   |   |   |   B3PRevenue > 72.277565
|   |   |   |   |   |   OBRevenue <= 256.604242
|   |   |   |   |   |   |   OBRevenue <= 245.228656: L (3.0)
|   |   |   |   |   |   |   OBRevenue > 245.228656: M (2.0)
|   |   |   |   |   |   OBRevenue > 256.604242: L (10.0)
|   |   |   |   OBRevenue > 335.673714: L (90.0/16.0)
|   |   |   IBRevenue > 560.750117: L (241.0)
|   OBRevenue > 569.865642: L (2115.0/22.0)

Number of Leaves  :   52

Size of the tree :    103


Time taken to build model: 0.89 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances       19304               95.1874 %
Incorrectly Classified Instances       976                4.8126 %
Kappa statistic                          0.9202
Mean absolute error                      0.0565
Root mean squared error                  0.1681
Relative absolute error                 13.9209 %
Root relative squared error             37.311  %
Coverage of cases (0.95 level)          97.0513 %
Mean rel. region size (0.95 level)      37.2354 %
Total Number of Instances            20280
```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.989	0.056	0.952	0.989	0.97	0.978	S
	0.919	0.023	0.937	0.919	0.928	0.971	M
	0.9	0.006	0.975	0.9	0.936	0.965	L
Weighted Avg.	0.952	0.037	0.952	0.952	0.952	0.973	

=== Confusion Matrix ===

a	b	c	<-- classified as
10524	120	0	a = S
352	5101	95	b = M
184	225	3679	c = L

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     Data
Instances:    20280
Attributes:   4
              B3PRevenue
              IBRevenue
              OBRevenue
              Type
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------

OBRevenue <= 229.3091
|   IBRevenue <= 230.074543
|   |   B3PRevenue <= 230.023771
|   |   |   OBRevenue <= 159.643974
|   |   |   |   IBRevenue <= 161.781164: S (9981.0/361.0)
|   |   |   |   IBRevenue > 161.781164
|   |   |   |   |   OBRevenue <= 55.955057: S (431.0/59.0)
|   |   |   |   |   OBRevenue > 55.955057
|   |   |   |   |   |   IBRevenue <= 184.009415
|   |   |   |   |   |   |   OBRevenue <= 138.468671
|   |   |   |   |   |   |   |   B3PRevenue <= 42.382383: S (58.0/18.0)
|   |   |   |   |   |   |   |   B3PRevenue > 42.382383: M (4.0/1.0)
|   |   |   |   |   |   |   OBRevenue > 138.468671: M (13.0/4.0)
|   |   |   |   |   |   IBRevenue > 184.009415: M (116.0/38.0)
|   |   |   OBRevenue > 159.643974
|   |   |   |   IBRevenue <= 56.11
|   |   |   |   |   OBRevenue <= 204.231778: S (435.0/35.0)
|   |   |   |   |   OBRevenue > 204.231778
|   |   |   |   |   |   B3PRevenue <= 0: S (155.0/63.0)
|   |   |   |   |   |   B3PRevenue > 0
|   |   |   |   |   |   |   B3PRevenue <= 49.648278: L (3.0/1.0)
|   |   |   |   |   |   |   B3PRevenue > 49.648278: M (6.0/1.0)
|   |   |   |   IBRevenue > 56.11: M (230.0/44.0)
|   |   B3PRevenue > 230.023771
|   |   |   B3PRevenue <= 587.140638
|   |   |   |   B3PRevenue <= 449.512076: M (89.0/5.0)
|   |   |   |   B3PRevenue > 449.512076
|   |   |   |   |   OBRevenue <= 12.477992
|   |   |   |   |   |   B3PRevenue <= 459.373256: L (3.0/1.0)
|   |   |   |   |   |   B3PRevenue > 459.373256: M (6.0)
|   |   |   |   |   OBRevenue > 12.477992
```

```
|   |   |   |   |   |   OBRevenue <= 103.99053: L (4.0)
|   |   |   |   |   |   OBRevenue > 103.99053: M (3.0/1.0)
|   |   |   B3PRevenue > 587.140638: L (39.0)
|   IBRevenue > 230.074543
|   |   IBRevenue <= 569.155701
|   |   |   B3PRevenue <= 488.65
|   |   |   |   IBRevenue <= 435.480444: M (1643.0/70.0)
|   |   |   |   IBRevenue > 435.480444
|   |   |   |   |   OBRevenue <= 36.753087
|   |   |   |   |   |   B3PRevenue <= 97.15034: M (252.0/19.0)
|   |   |   |   |   |   B3PRevenue > 97.15034: L (11.0/4.0)
|   |   |   |   |   OBRevenue > 36.753087
|   |   |   |   |   |   OBRevenue <= 85.878644
|   |   |   |   |   |   |   B3PRevenue <= 50.86: M (33.0/15.0)
|   |   |   |   |   |   |   B3PRevenue > 50.86: L (4.0)
|   |   |   |   |   |   OBRevenue > 85.878644: L (49.0/6.0)
|   |   |   B3PRevenue > 488.65: L (26.0/1.0)
|   |   IBRevenue > 569.155701: L (1040.0/16.0)
OBRevenue > 229.3091
|   OBRevenue <= 569.865642
|   |   IBRevenue <= 301.707798
|   |   |   OBRevenue <= 506.28281
|   |   |   |   IBRevenue <= 114.301341: M (2545.0/86.0)
|   |   |   |   IBRevenue > 114.301341
|   |   |   |   |   OBRevenue <= 394.51023
|   |   |   |   |   |   B3PRevenue <= 95.44
|   |   |   |   |   |   |   IBRevenue <= 241.72288: M (235.0/5.0)
|   |   |   |   |   |   |   IBRevenue > 241.72288
|   |   |   |   |   |   |   |   OBRevenue <= 343.983503: M (54.0/4.0)
|   |   |   |   |   |   |   |   OBRevenue > 343.983503
|   |   |   |   |   |   |   |   |   OBRevenue <= 384.063882: L (12.0/4.0)
|   |   |   |   |   |   |   |   |   OBRevenue > 384.063882: M (3.0)
|   |   |   |   |   |   B3PRevenue > 95.44
|   |   |   |   |   |   |   IBRevenue <= 124.153679: L (3.0)
|   |   |   |   |   |   |   IBRevenue > 124.153679
|   |   |   |   |   |   |   |   B3PRevenue <= 205.33: M (10.0)
|   |   |   |   |   |   |   |   B3PRevenue > 205.33: L (4.0/1.0)
|   |   |   |   |   OBRevenue > 394.51023
|   |   |   |   |   |   B3PRevenue <= 35.8
|   |   |   |   |   |   |   B3PRevenue <= 12.03992
|   |   |   |   |   |   |   |   OBRevenue <= 475.564469
|   |   |   |   |   |   |   |   |   IBRevenue <= 169.657211: M (16.0/2.0)
|   |   |   |   |   |   |   |   |   IBRevenue > 169.657211: L (22.0/7.0)
|   |   |   |   |   |   |   |   OBRevenue > 475.564469: L (8.0)
|   |   |   |   |   |   |   B3PRevenue > 12.03992: M (2.0)
|   |   |   |   |   |   B3PRevenue > 35.8: L (8.0)
|   |   |   OBRevenue > 506.28281
|   |   |   |   IBRevenue <= 82.113808
```

```
|   |   |   |   |   OBRevenue <= 522.514021: M (36.0/3.0)
|   |   |   |   |   OBRevenue > 522.514021
|   |   |   |   |   |   IBRevenue <= 41.511355: M (117.0/44.0)
|   |   |   |   |   |   IBRevenue > 41.511355: L (6.0/1.0)
|   |   |   |   IBRevenue > 82.113808: L (36.0)
|   |   IBRevenue > 301.707798
|   |   |   IBRevenue <= 560.750117
|   |   |   |   OBRevenue <= 335.673714
|   |   |   |   |   B3PRevenue <= 72.277565
|   |   |   |   |   |   IBRevenue <= 351.401345: M (23.0/3.0)
|   |   |   |   |   |   IBRevenue > 351.401345
|   |   |   |   |   |   |   OBRevenue <= 237.535639: L (5.0)
|   |   |   |   |   |   |   OBRevenue > 237.535639
|   |   |   |   |   |   |   |   OBRevenue <= 257.583038: M (8.0)
|   |   |   |   |   |   |   |   OBRevenue > 257.583038: L (32.0/15.0)
|   |   |   |   |   B3PRevenue > 72.277565
|   |   |   |   |   |   OBRevenue <= 256.604242
|   |   |   |   |   |   |   OBRevenue <= 245.228656: L (3.0)
|   |   |   |   |   |   |   OBRevenue > 245.228656: M (2.0)
|   |   |   |   |   |   OBRevenue > 256.604242: L (10.0)
|   |   |   |   OBRevenue > 335.673714: L (90.0/16.0)
|   |   |   IBRevenue > 560.750117: L (241.0)
|   OBRevenue > 569.865642: L (2115.0/22.0)

Number of Leaves  :     52

Size of the tree :      103


Time taken to build model: 1.2 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       19152               94.4379 %
Incorrectly Classified Instances      1128                5.5621 %
Kappa statistic                          0.9078
Mean absolute error                      0.0614
Root mean squared error                  0.181
Relative absolute error                 15.1237 %
Root relative squared error             40.1676 %
Coverage of cases (0.95 level)          96.6617 %
Mean rel. region size (0.95 level)      38.003  %
Total Number of Instances            20280
```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.986	0.057	0.95	0.986	0.967	0.975	S
	0.908	0.029	0.921	0.908	0.914	0.964	M
	0.886	0.009	0.962	0.886	0.923	0.959	L
Weighted Avg.	0.944	0.04	0.944	0.944	0.944	0.969	

=== Confusion Matrix ===

a	b	c	<-- classified as
10491	151	2	a = S
371	5037	140	b = M
181	283	3624	c = L

=== Run information ===

Scheme: weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1
Relation: Data
Instances: 20280
Attributes: 4
B3PRevenue
IBRevenue
OBRevenue
Type
Test mode: evaluate on training data

=== Classifier model (full training set) ===

JRIP rules:
===========

(OBRevenue >= 520.034417) and (OBRevenue >= 604.868792) => Type=L (2019.0/6.0)
(IBRevenue >= 471.234838) and (IBRevenue >= 590.538145) => Type=L (1232.0/4.0)
(OBRevenue >= 479.704023) and (IBRevenue >= 35.394203) => Type=L (126.0/9.0)
(OBRevenue >= 484.014205) and (OBRevenue >= 555.029035) => Type=L (99.0/38.0)
(IBRevenue >= 421.59839) and (OBRevenue >= 43.282874) and (IBRevenue >= 473.23) => Type=L (117.0/18.0)
(B3PRevenue >= 142.834236) and (B3PRevenue >= 508.435904) => Type=L (83.0/4.0)
(OBRevenue >= 374.216775) and (IBRevenue >= 87.140443) and (IBRevenue >= 184.13883) => Type=L (63.0/21.0)
(IBRevenue >= 307.39392) and (OBRevenue >= 167.003476) and (B3PRevenue >= 20.98752) => Type=L (25.0/9.0)
(OBRevenue >= 196.522298) and (OBRevenue >= 248.475772) => Type=M (2801.0/116.0)
(IBRevenue >= 193.784392) and (IBRevenue >= 239.846763) => Type=M (1902.0/108.0)
(OBRevenue >= 164.053371) and (IBRevenue >= 57.39) => Type=M (300.0/41.0)
(IBRevenue >= 155.132024) and (IBRevenue >= 208.914136) and (IBRevenue >= 229.885041) => Type=M (67.0/18.0)
(OBRevenue >= 191.990676) and (OBRevenue >= 227.489109) => Type=M (154.0/31.0)
(IBRevenue >= 178.329873) and (OBRevenue >= 55.012011) and (IBRevenue >= 185.275597) => Type=M (116.0/36.0)
(B3PRevenue >= 180.84) and (B3PRevenue >= 230.391462) => Type=M (87.0/8.0)
(IBRevenue >= 135.81337) and (OBRevenue >= 94.029048) and (OBRevenue >= 135.608224) and (OBRevenue <= 145.343721) => Type=M (17.0/5.0)
=> Type=S (11072.0/546.0)

Number of Rules : 17

Time taken to build model: 13.2 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances	19262	94.9803 %
Incorrectly Classified Instances	1018	5.0197 %
Kappa statistic	0.9167	
Mean absolute error	0.0612	
Root mean squared error	0.175	
Relative absolute error	15.084 %	
Root relative squared error	38.8384 %	
Coverage of cases (0.95 level)	96.5335 %	
Mean rel. region size (0.95 level)	38.4122 %	
Total Number of Instances	20280	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.989	0.057	0.951	0.989	0.969	0.971	S
	0.916	0.025	0.933	0.916	0.924	0.963	M
	0.894	0.007	0.971	0.894	0.931	0.96	L
Weighted Avg.	0.95	0.038	0.95	0.95	0.949	0.967	

=== Confusion Matrix ===

a	b	c	<-- classified as
10526	118	0	a = S
358	5081	109	b = M
188	245	3655	c = L

=== Run information ===

Scheme: weka.classifiers.fuzzy.FLR -R 0.5 -Y -B
Relation: Data
Instances: 20280
Attributes: 4
B3PRevenue
IBRevenue
OBRevenue
Type
Test mode: evaluate on training data

=== Classifier model (full training set) ===

FLR classifier
======================
Rhoa = 0.5
Extracted Rules (Fuzzy Lattices):

Rule: 0 [0.0 248.78] [0.0 249.2] [0.0 247.6990495] in Class: S
Rule: 1 [0.0 587.1406382] [0.0 614.5533333] [0.0 622.7100393] in Class: M
Rule: 2 [0.0 66604.85352] [0.0 21108.25543] [0.0 84721.27437] in Class: L
Rule: 3 [0.0 0.0] [0.0 0.0] [129731.4972 129731.4972] in Class: L

Metric Space:
[0.0 66604.85352] [0.0 21108.25543] [0.0 129731.4972] in Class: Metric Space

Total Number of Rules: 4
Rules pointing in Class S :1
Rules pointing in Class M :1
Rules pointing in Class L :2

Time taken to build model: 0.09 seconds

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances       18463               91.0404 %
Incorrectly Classified Instances      1817                8.9596 %
Kappa statistic                          0.8486
Mean absolute error                      0.0597
Root mean squared error                  0.2444
Relative absolute error                 14.7104 %
Root relative squared error             54.2413 %
Coverage of cases (0.95 level)          91.0404 %
Mean rel. region size (0.95 level)      33.3333 %
Total Number of Instances            20280

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0.119      0.903     1         0.949      0.94     S
                 0.83      0.045      0.873     0.83      0.851      0.892    M
                 0.786     0          1         0.786     0.88       0.893    L
Weighted Avg.    0.91      0.075      0.914     0.91      0.908      0.918

=== Confusion Matrix ===

     a     b     c   <-- classified as
 10644     0     0 |     a = S
   942  4606     0 |     b = M
   207   668  3213 |     c = L
```

=== Run information ===

Scheme: weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1
Relation: Data
Instances: 20280
Attributes: 4
B3PRevenue
IBRevenue
OBRevenue
Type
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

JRIP rules:
===========

(OBRevenue >= 520.034417) and (OBRevenue >= 604.868792) => Type=L (2019.0/6.0)
(IBRevenue >= 471.234838) and (IBRevenue >= 590.538145) => Type=L (1232.0/4.0)
(OBRevenue >= 479.704023) and (IBRevenue >= 35.394203) => Type=L (126.0/9.0)
(OBRevenue >= 484.014205) and (OBRevenue >= 555.029035) => Type=L (99.0/38.0)
(IBRevenue >= 421.59839) and (OBRevenue >= 43.282874) and (IBRevenue >= 473.23) => Type=L (117.0/18.0)
(B3PRevenue >= 142.834236) and (B3PRevenue >= 508.435904) => Type=L (83.0/4.0)
(OBRevenue >= 374.216775) and (IBRevenue >= 87.140443) and (IBRevenue >= 184.13883) => Type=L (63.0/21.0)
(IBRevenue >= 307.39392) and (OBRevenue >= 167.003476) and (B3PRevenue >= 20.98752) => Type=L (25.0/9.0)
(OBRevenue >= 196.522298) and (OBRevenue >= 248.475772) => Type=M (2801.0/116.0)
(IBRevenue >= 193.784392) and (IBRevenue >= 239.846763) => Type=M (1902.0/108.0)
(OBRevenue >= 164.053371) and (IBRevenue >= 57.39) => Type=M (300.0/41.0)
(IBRevenue >= 155.132024) and (IBRevenue >= 208.914136) and (IBRevenue >= 229.885041) => Type=M (67.0/18.0)
(OBRevenue >= 191.990676) and (OBRevenue >= 227.489109) => Type=M (154.0/31.0)
(IBRevenue >= 178.329873) and (OBRevenue >= 55.012011) and (IBRevenue >= 185.275597) => Type=M (116.0/36.0)
(B3PRevenue >= 180.84) and (B3PRevenue >= 230.391462) => Type=M (87.0/8.0)
(IBRevenue >= 135.81337) and (OBRevenue >= 94.029048) and (OBRevenue >= 135.608224) and (OBRevenue <= 145.343721) => Type=M (17.0/5.0)
=> Type=S (11072.0/546.0)

Number of Rules : 17

Time taken to build model: 9.89 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	19166	94.5069 %
Incorrectly Classified Instances	1114	5.4931 %
Kappa statistic	0.9089	
Mean absolute error	0.0633	
Root mean squared error	0.1818	
Relative absolute error	15.5867 %	
Root relative squared error	40.3431 %	
Coverage of cases (0.95 level)	96.6519 %	
Mean rel. region size (0.95 level)	42.1466 %	
Total Number of Instances	20280	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.986	0.059	0.949	0.986	0.967	0.97	S
	0.91	0.029	0.922	0.91	0.916	0.96	M
	0.886	0.008	0.967	0.886	0.925	0.956	L
Weighted Avg.	0.945	0.04	0.945	0.945	0.945	0.964	

=== Confusion Matrix ===

a	b	c	<-- classified as
10498	146	0	a = S
379	5046	123	b = M
185	281	3622	c = L

=== Run information ===

Scheme: weka.classifiers.fuzzy.FLR -R 0.5 -Y -B
Relation: Data
Instances: 20280
Attributes: 4
B3PRevenue
IBRevenue
OBRevenue
Type
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

FLR classifier
======================
Rhoa = 0.5
Extracted Rules (Fuzzy Lattices):

Rule: 0 [0.0 248.78] [0.0 249.2] [0.0 247.6990495] in Class: S
Rule: 1 [0.0 587.1406382] [0.0 614.5533333] [0.0 622.7100393] in Class: M
Rule: 2 [0.0 66604.85352] [0.0 21108.25543] [0.0 84721.27437] in Class: L
Rule: 3 [0.0 0.0] [0.0 0.0] [129731.4972 129731.4972] in Class: L

Metric Space:
[0.0 66604.85352] [0.0 21108.25543] [0.0 129731.4972] in Class: Metric Space

Total Number of Rules: 4
Rules pointing in Class S :1
Rules pointing in Class M :1
Rules pointing in Class L :2

Time taken to build model: 0.14 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	11793	58.1509 %
Incorrectly Classified Instances	8487	41.8491 %
Kappa statistic	0.3913	
Mean absolute error	0.279	
Root mean squared error	0.5282	
Relative absolute error	68.7104 %	
Root relative squared error	117.2275 %	
Coverage of cases (0.95 level)	58.1509 %	
Mean rel. region size (0.95 level)	33.3333 %	
Total Number of Instances	20280	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.4	0.047	0.904	0.4	0.554	0.676	S
	0.749	0.403	0.411	0.749	0.531	0.673	M
	0.827	0.129	0.618	0.827	0.707	0.849	L
Weighted Avg.	0.582	0.161	0.711	0.582	0.579	0.71	

=== Confusion Matrix ===

a	b	c	<-- classified as
4257	5323	1064	a = S
367	4155	1026	b = M
87	620	3381	c = L

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     Data-weka.filters.supervised.attribute.Discretize-Rfirst-last
Instances:    20280
Attributes:   4
              B3PRevenue
              IBRevenue
              OBRevenue
              Type
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------

OBRevenue = '(-inf-9.670941]'
|   IBRevenue = '(-inf-16.340389]'
|   |   B3PRevenue = '(-inf-33.950457]': S (43.0/14.0)
|   |   B3PRevenue = '(33.950457-115.285538]': S (39.0/4.0)
|   |   B3PRevenue = '(115.285538-230.207617]': S (31.0/2.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (33.0/2.0)
|   |   B3PRevenue = '(587.541951-inf)': L (19.0)
|   IBRevenue = '(16.340389-155.123512]'
|   |   B3PRevenue = '(-inf-33.950457]': S (1568.0/35.0)
|   |   B3PRevenue = '(33.950457-115.285538]': S (51.0/2.0)
|   |   B3PRevenue = '(115.285538-230.207617]': S (32.0/8.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (13.0/2.0)
|   |   B3PRevenue = '(587.541951-inf)': L (4.0)
|   IBRevenue = '(155.123512-192.613058]'
|   |   B3PRevenue = '(-inf-33.950457]': S (227.0/3.0)
|   |   B3PRevenue = '(33.950457-115.285538]': S (4.0/1.0)
|   |   B3PRevenue = '(115.285538-230.207617]': S (2.0/1.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (5.0/1.0)
|   |   B3PRevenue = '(587.541951-inf)': L (2.0)
|   IBRevenue = '(192.613058-230.123548]'
|   |   B3PRevenue = '(-inf-33.950457]': S (125.0/32.0)
|   |   B3PRevenue = '(33.950457-115.285538]': M (3.0)
|   |   B3PRevenue = '(115.285538-230.207617]': S (2.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (1.0)
|   |   B3PRevenue = '(587.541951-inf)': S (0.0)
|   IBRevenue = '(230.123548-249.215152]': M (71.0/16.0)
|   IBRevenue = '(249.215152-471.008761]'
|   |   B3PRevenue = '(-inf-33.950457]': M (905.0/14.0)
|   |   B3PRevenue = '(33.950457-115.285538]': M (22.0)
|   |   B3PRevenue = '(115.285538-230.207617]': M (30.0/3.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (25.0/6.0)
```

```
|   |   B3PRevenue = '(587.541951-inf)': L (7.0)
|   IBRevenue = '(471.008761-569.082407]'
|   |   B3PRevenue = '(-inf-33.950457]': M (132.0/14.0)
|   |   B3PRevenue = '(33.950457-115.285538]': L (1.0)
|   |   B3PRevenue = '(115.285538-230.207617]': L (2.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (4.0/1.0)
|   |   B3PRevenue = '(587.541951-inf)': L (3.0)
|   IBRevenue = '(569.082407-614.619731]': L (50.0/14.0)
|   IBRevenue = '(614.619731-inf)': L (527.0)
OBRevenue = '(9.670941-106.198851]'
|   IBRevenue = '(-inf-16.340389]'
|   |   B3PRevenue = '(-inf-33.950457]': S (4498.0/148.0)
|   |   B3PRevenue = '(33.950457-115.285538]': S (54.0/1.0)
|   |   B3PRevenue = '(115.285538-230.207617]': S (26.0/1.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (17.0/4.0)
|   |   B3PRevenue = '(587.541951-inf)': L (3.0)
|   IBRevenue = '(16.340389-155.123512]'
|   |   B3PRevenue = '(-inf-33.950457]': S (1807.0/41.0)
|   |   B3PRevenue = '(33.950457-115.285538]': S (130.0/6.0)
|   |   B3PRevenue = '(115.285538-230.207617]': S (39.0/7.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (7.0/2.0)
|   |   B3PRevenue = '(587.541951-inf)': L (1.0)
|   IBRevenue = '(155.123512-192.613058]'
|   |   B3PRevenue = '(-inf-33.950457]': S (168.0/27.0)
|   |   B3PRevenue = '(33.950457-115.285538]': S (5.0/1.0)
|   |   B3PRevenue = '(115.285538-230.207617]': M (3.0/1.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (1.0)
|   |   B3PRevenue = '(587.541951-inf)': S (0.0)
|   IBRevenue = '(192.613058-230.123548]': M (95.0/46.0)
|   IBRevenue = '(230.123548-249.215152]': M (40.0/6.0)
|   IBRevenue = '(249.215152-471.008761]': M (408.0/28.0)
|   IBRevenue = '(471.008761-569.082407]': L (42.0/17.0)
|   IBRevenue = '(569.082407-614.619731]': L (17.0/2.0)
|   IBRevenue = '(614.619731-inf)': L (255.0)
OBRevenue = '(106.198851-159.448064]'
|   IBRevenue = '(-inf-16.340389]'
|   |   B3PRevenue = '(-inf-33.950457]': S (1156.0/42.0)
|   |   B3PRevenue = '(33.950457-115.285538]': S (29.0/3.0)
|   |   B3PRevenue = '(115.285538-230.207617]': M (7.0/3.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (5.0)
|   |   B3PRevenue = '(587.541951-inf)': L (1.0)
|   IBRevenue = '(16.340389-155.123512]'
|   |   B3PRevenue = '(-inf-33.950457]': S (348.0/30.0)
|   |   B3PRevenue = '(33.950457-115.285538]': S (20.0/5.0)
|   |   B3PRevenue = '(115.285538-230.207617]': S (8.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (7.0)
|   |   B3PRevenue = '(587.541951-inf)': L (1.0)
|   IBRevenue = '(155.123512-192.613058]': M (41.0/16.0)
```

```
|   IBRevenue = '(192.613058-230.123548]': M (44.0/11.0)
|   IBRevenue = '(230.123548-249.215152]': M (19.0/1.0)
|   IBRevenue = '(249.215152-471.008761]': M (143.0/9.0)
|   IBRevenue = '(471.008761-569.082407]': L (17.0/2.0)
|   IBRevenue = '(569.082407-614.619731]': L (8.0)
|   IBRevenue = '(614.619731-inf)': L (107.0)
OBRevenue = '(159.448064-191.777345]'
|   IBRevenue = '(-inf-16.340389]': S (339.0/22.0)
|   IBRevenue = '(16.340389-155.123512]'
|   |   B3PRevenue = '(-inf-33.950457]': S (87.0/36.0)
|   |   B3PRevenue = '(33.950457-115.285538]': M (4.0/1.0)
|   |   B3PRevenue = '(115.285538-230.207617]': S (2.0/1.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (7.0)
|   |   B3PRevenue = '(587.541951-inf)': L (1.0)
|   IBRevenue = '(155.123512-192.613058]': M (21.0/5.0)
|   IBRevenue = '(192.613058-230.123548]': M (22.0/2.0)
|   IBRevenue = '(230.123548-249.215152]': M (10.0)
|   IBRevenue = '(249.215152-471.008761]': M (64.0/8.0)
|   IBRevenue = '(471.008761-569.082407]': L (8.0)
|   IBRevenue = '(569.082407-614.619731]': L (3.0)
|   IBRevenue = '(614.619731-inf)': L (38.0)
OBRevenue = '(191.777345-229.521161]'
|   IBRevenue = '(-inf-16.340389]'
|   |   B3PRevenue = '(-inf-33.950457]': S (213.0/68.0)
|   |   B3PRevenue = '(33.950457-115.285538]': S (0.0)
|   |   B3PRevenue = '(115.285538-230.207617]': M (5.0/1.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (1.0)
|   |   B3PRevenue = '(587.541951-inf)': L (3.0)
|   IBRevenue = '(16.340389-155.123512]': M (79.0/17.0)
|   IBRevenue = '(155.123512-192.613058]': M (36.0)
|   IBRevenue = '(192.613058-230.123548]': M (31.0/2.0)
|   IBRevenue = '(230.123548-249.215152]': M (11.0/1.0)
|   IBRevenue = '(249.215152-471.008761]': M (45.0/7.0)
|   IBRevenue = '(471.008761-569.082407]': L (9.0/1.0)
|   IBRevenue = '(569.082407-614.619731]': L (4.0)
|   IBRevenue = '(614.619731-inf)': L (31.0)
OBRevenue = '(229.521161-247.699677]'
|   IBRevenue = '(-inf-16.340389]': M (131.0/26.0)
|   IBRevenue = '(16.340389-155.123512]': M (43.0/4.0)
|   IBRevenue = '(155.123512-192.613058]': M (13.0)
|   IBRevenue = '(192.613058-230.123548]': M (11.0)
|   IBRevenue = '(230.123548-249.215152]': M (7.0)
|   IBRevenue = '(249.215152-471.008761]': M (21.0/5.0)
|   IBRevenue = '(471.008761-569.082407]': L (5.0/1.0)
|   IBRevenue = '(569.082407-614.619731]': L (1.0)
|   IBRevenue = '(614.619731-inf)': L (20.0)
OBRevenue = '(247.699677-376.880103]'
|   IBRevenue = '(-inf-16.340389]': M (1259.0/24.0)
```

```
|   IBRevenue = '(16.340389-155.123512]': M (364.0/8.0)
|   IBRevenue = '(155.123512-192.613058]': M (53.0/2.0)
|   IBRevenue = '(192.613058-230.123548]': M (45.0/2.0)
|   IBRevenue = '(230.123548-249.215152]': M (22.0/1.0)
|   IBRevenue = '(249.215152-471.008761]'
|   |   B3PRevenue = '(-inf-33.950457]': M (92.0/27.0)
|   |   B3PRevenue = '(33.950457-115.285538]': M (4.0/1.0)
|   |   B3PRevenue = '(115.285538-230.207617]': L (2.0)
|   |   B3PRevenue = '(230.207617-587.541951]': L (9.0/1.0)
|   |   B3PRevenue = '(587.541951-inf)': L (1.0)
|   IBRevenue = '(471.008761-569.082407]': L (27.0/8.0)
|   IBRevenue = '(569.082407-614.619731]': L (7.0)
|   IBRevenue = '(614.619731-inf)': L (100.0)
OBRevenue = '(376.880103-504.176667]'
|   IBRevenue = '(-inf-16.340389]': M (741.0/18.0)
|   IBRevenue = '(16.340389-155.123512]': M (130.0/20.0)
|   IBRevenue = '(155.123512-192.613058]': M (14.0/6.0)
|   IBRevenue = '(192.613058-230.123548]': L (13.0/3.0)
|   IBRevenue = '(230.123548-249.215152]': M (10.0/5.0)
|   IBRevenue = '(249.215152-471.008761]': L (44.0/15.0)
|   IBRevenue = '(471.008761-569.082407]': L (9.0/1.0)
|   IBRevenue = '(569.082407-614.619731]': L (9.0)
|   IBRevenue = '(614.619731-inf)': L (76.0)
OBRevenue = '(504.176667-570.299148]'
|   IBRevenue = '(-inf-16.340389]'
|   |   B3PRevenue = '(-inf-33.950457]': M (141.0/41.0)
|   |   B3PRevenue = '(33.950457-115.285538]': L (3.0)
|   |   B3PRevenue = '(115.285538-230.207617]': M (0.0)
|   |   B3PRevenue = '(230.207617-587.541951]': M (0.0)
|   |   B3PRevenue = '(587.541951-inf)': L (1.0)
|   IBRevenue = '(16.340389-155.123512]': L (31.0/8.0)
|   IBRevenue = '(155.123512-192.613058]': L (4.0)
|   IBRevenue = '(192.613058-230.123548]': L (6.0)
|   IBRevenue = '(230.123548-249.215152]': L (1.0)
|   IBRevenue = '(249.215152-471.008761]': L (27.0)
|   IBRevenue = '(471.008761-569.082407]': L (8.0)
|   IBRevenue = '(569.082407-614.619731]': L (2.0)
|   IBRevenue = '(614.619731-inf)': L (24.0)
OBRevenue = '(570.299148-622.938386]': L (169.0/22.0)
OBRevenue = '(622.938386-inf)': L (1946.0)

Number of Leaves  :   143

Size of the tree :    168


Time taken to build model: 0.13 seconds
```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	19178	94.5661 %
Incorrectly Classified Instances	1102	5.4339 %
Kappa statistic	0.9097	
Mean absolute error	0.0596	
Root mean squared error	0.175	
Relative absolute error	14.6795 %	
Root relative squared error	38.8287 %	
Coverage of cases (0.95 level)	97.7367 %	
Mean rel. region size (0.95 level)	40.4389 %	
Total Number of Instances	20280	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.988	0.061	0.947	0.988	0.967	0.981	S
	0.911	0.028	0.924	0.911	0.917	0.976	M
	0.882	0.006	0.974	0.882	0.925	0.963	L
Weighted Avg.	0.946	0.041	0.946	0.946	0.945	0.976	

=== Confusion Matrix ===

a	b	c	<-- classified as
10521	123	0	a = S
397	5053	98	b = M
190	294	3604	c = L

=== Run information ===

Scheme: weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1
Relation: Data-weka.filters.supervised.attribute.Discretize-Rfirst-last
Instances: 20280
Attributes: 4
B3PRevenue
IBRevenue
OBRevenue
Type
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

JRIP rules:
===========

(OBRevenue = '(622.938386-inf)') => Type=L (1946.0/0.0)
(IBRevenue = '(614.619731-inf)') => Type=L (1191.0/0.0)
(OBRevenue = '(570.299148-622.938386]') => Type=L (156.0/22.0)
(OBRevenue = '(504.176667-570.299148]') and (IBRevenue = '(249.215152-471.008761]') => Type=L (27.0/0.0)
(IBRevenue = '(569.082407-614.619731]') => Type=L (101.0/16.0)
(IBRevenue = '(471.008761-569.082407]') and (OBRevenue = '(106.198851-159.448064]') => Type=L (17.0/2.0)
(B3PRevenue = '(587.541951-inf)') => Type=L (68.0/0.0)
(IBRevenue = '(471.008761-569.082407]') and (OBRevenue = '(9.670941-106.198851]') => Type=L (40.0/17.0)
(OBRevenue = '(504.176667-570.299148]') and (IBRevenue = '(16.340389-155.123512]') => Type=L (30.0/8.0)
(IBRevenue = '(471.008761-569.082407]') and (OBRevenue = '(247.699677-376.880103]') => Type=L (26.0/8.0)
(OBRevenue = '(504.176667-570.299148]') and (IBRevenue = '(192.613058-230.123548]') => Type=L (6.0/0.0)
(OBRevenue = '(504.176667-570.299148]') and (IBRevenue = '(471.008761-569.082407]') => Type=L (7.0/0.0)
(IBRevenue = '(249.215152-471.008761]') and (OBRevenue = '(247.699677-376.880103]') and (B3PRevenue = '(230.207617-587.541951]') => Type=L (9.0/1.0)
(IBRevenue = '(249.215152-471.008761]') and (OBRevenue = '(376.880103-504.176667]') => Type=L (43.0/15.0)
(IBRevenue = '(471.008761-569.082407]') and (OBRevenue = '(376.880103-504.176667]') => Type=L (9.0/1.0)
(OBRevenue = '(504.176667-570.299148]') and (B3PRevenue = '(33.950457-115.285538]') => Type=L (4.0/0.0)
(OBRevenue = '(504.176667-570.299148]') and (IBRevenue = '(155.123512-192.613058]') => Type=L (3.0/0.0)

(IBRevenue = '(471.008761-569.082407]') and (OBRevenue = '(191.777345-229.521161]') => Type=L (9.0/1.0)
(OBRevenue = '(247.699677-376.880103]') => Type=M (1840.0/66.0)
(IBRevenue = '(249.215152-471.008761]') => Type=M (1658.0/75.0)
(OBRevenue = '(376.880103-504.176667]') => Type=M (906.0/57.0)
(OBRevenue = '(229.521161-247.699677]') => Type=M (208.0/32.0)
(OBRevenue = '(191.777345-229.521161]') and (IBRevenue = '(16.340389-155.123512]') => Type=M (79.0/17.0)
(IBRevenue = '(230.123548-249.215152]') => Type=M (151.0/24.0)
(IBRevenue = '(192.613058-230.123548]') and (OBRevenue = '(191.777345-229.521161]') => Type=M (30.0/1.0)
(IBRevenue = '(471.008761-569.082407]') and (OBRevenue = '(-inf-9.670941]') and (B3PRevenue = '(-inf-33.950457]') => Type=M (132.0/14.0)
(IBRevenue = '(192.613058-230.123548]') and (OBRevenue = '(106.198851-159.448064]') => Type=M (44.0/11.0)
(OBRevenue = '(504.176667-570.299148]') => Type=M (141.0/41.0)
(B3PRevenue = '(230.207617-587.541951]') => Type=M (105.0/12.0)
(IBRevenue = '(192.613058-230.123548]') and (OBRevenue = '(9.670941-106.198851]') => Type=M (92.0/45.0)
(OBRevenue = '(191.777345-229.521161]') and (IBRevenue = '(155.123512-192.613058]') => Type=M (36.0/0.0)
(IBRevenue = '(192.613058-230.123548]') and (OBRevenue = '(159.448064-191.777345]') => Type=M (22.0/2.0)
(OBRevenue = '(191.777345-229.521161]') and (B3PRevenue = '(115.285538-230.207617]') => Type=M (5.0/1.0)
=> Type=S (11139.0/608.0)

Number of Rules : 34

Time taken to build model: 66.65 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	19145	94.4034 %
Incorrectly Classified Instances	1135	5.5966 %
Kappa statistic	0.9069	
Mean absolute error	0.0666	
Root mean squared error	0.1838	
Relative absolute error	16.4116 %	
Root relative squared error	40.7905 %	
Coverage of cases (0.95 level)	97.998 %	
Mean rel. region size (0.95 level)	54.8636 %	
Total Number of Instances	20280	

```
=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.99      0.067      0.942     0.99      0.965      0.965    S
                 0.906     0.027      0.927     0.906     0.916      0.96     M
                 0.876     0.005      0.976     0.876     0.923      0.955    L
Weighted Avg.    0.944     0.044      0.945     0.944     0.943      0.961

=== Confusion Matrix ===

     a     b     c   <-- classified as
 10533   111     0 |     a = S
   430  5029    89 |     b = M
   218   287  3583 |     c = L
```

BIOGRAPHY

NAME	Donnapha Meephol
DATE OF BIRTH	14 June 1980
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTIONS ATTENDED	Chiang Mai University, 19939-2003
	Bachelor of Science (Agriculture)
	Chulalongkorn University, 2005-2007
	Master Degree of Education
	Mahidol University, 2011-2014
	Master of Science (Technology Information System Management)
RESEARCH GRANTS	This thesis is partially supported by Graduate Studies of Mahidol University Alumni Association
HOME ADDRESS	9 Soi Petchakasem 40 Junction 9, Banjak, Phaseejarean, Bangkok 10160
	Tel. 08-9131-4408
	E-mail:donnapha@gmail.com
PUBLICATION / PRESENTATION	Analysis Model of Customer Segmentation for Customer Care and Meeting Arrangement in International Logistic Business Using Fuzzy Logic, 2014 International Congress on Natural Sciences and Engineering (ICNSE 2014), Kyoto, Japan