

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

การประยุกต์ดาต้าไมนิ่งเพื่อทำนายลูกค้าที่มีแนวโน้มจะยกเลิก
การใช้บริการโทรศัพท์พื้นฐาน

Using Data Mining for Churn Predictive in Telephone Services



H003133

วัน เดือน ปี.....	18 พ.ค. 2550
เลขทะเบียน.....	03133
เลขเรียกหนังสือ.....	ด.ท.ด.142ก. 2547
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."	

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการศึกษาระดับปริญญาตรี
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
ภาคเรียนที่ 1 ปีการศึกษา 2547
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อหัวข้อ	การประยุกต์ค้าไม้เนื้อเพื่อทำนาลูกค้าที่มีแนวโนมจะยกเลิกการใช้บริการโทรศัพท์พื้นฐาน
นักศึกษา	นางสาวคลใจ กระแสเสน
อาจารย์ที่ปรึกษา	ผศ.ดร.วรพจน์ กริสุระเดช
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	การจัดการเทคโนโลยีสารสนเทศ
ปีการศึกษา	2547

บทคัดย่อ

ภายใต้การแข่งขันทางการตลาดสื่อสารโทรคมนาคมที่มีการเปลี่ยนแปลงอย่างต่อเนื่อง และการเปิดเสรีโทรคมนาคมในปี 2549 บมจ. ทศท ผู้ให้บริการโทรศัพท์พื้นฐานจะมีผู้ให้บริการรายใหม่ ทั้งจากภายในประเทศ และภายนอกประเทศ เข้าสู่ตลาดบริการโทรคมนาคมมากขึ้น ในขณะที่ผู้ให้บริการรายเดิมยังมีศักยภาพทั้งด้านการเงินและเทคโนโลยี มีผลทำให้การแข่งขันทวีความรุนแรงยิ่งขึ้น ซึ่งจะส่งผลกระทบต่อส่วนแบ่งทางการตลาดของโทรศัพท์พื้นฐาน ดังนั้น บมจ. ทศท จึงได้กำหนดเป้าหมายเชิงยุทธศาสตร์ไว้คือรักษาลูกค้า ขยายฐานลูกค้าและเพิ่มส่วนครองการตลาด รวมถึงรายได้ให้กับ บมจ. ทศท โดยใช้การตลาดเชิงรุกที่มุ่งเข้าหาลูกค้า และการตอบสนองลูกค้าให้ตรงความต้องการของลูกค้าที่มีความแตกต่างในแต่ละกลุ่ม โดยเฉพาะกลุ่มลูกค้ารายใหญ่และกลุ่มลูกค้าที่มีศักยภาพสูงในการจะเพิ่มการใช้บริการด้านโทรคมนาคมในอนาคต

ดังนั้นจำเป็นต้องดำเนินการศึกษาพฤติกรรม และความต้องการของลูกค้า เพื่อเป็นแนวทางในการปรับปรุงแผนกลยุทธ์ทางการตลาด เพิ่มศักยภาพการแข่งขัน และชิงความได้เปรียบทางการตลาด โดยการประยุกต์ Data Mining เพื่อทำนายลูกค้าที่มีแนวโน้มจะยกเลิกการใช้บริการโทรศัพท์พื้นฐาน

Title	Using Data Mining for Churn Predictive in Telephone Services
Student	Miss Doljai Krasaesen
Advisor	Asst. Prof. Dr. Worapoj Kreesuradej
Level of Study	Master of Science in Information Technology
Major	Information Technology Management
Academic Year	2004

ABSTRACT

During the last decade, Thailand telecommunication industry as fixed-line telephone has been dramatically growing. Due to a liberalization plan of the Government in this industry, private sectors have been allowed to be subcontractors as the telecommunication agency. A competition in this industry becomes reality and will be fierce after liberalization of telecommunication concession in 2006. Then customers have more choices for telephone services among telecommunication providers especially from private subcontractors; therefore, churn rate in TOT service will increase. It is then vital for TOT to find a attack strategy to maintain customers, extend customers base, increase market share, serve more focus group customers satisfaction especially big or high potential customers and other methods to reduce the churn rate. TOT should to study data mining method and apply it on churn rate prediction.

This article presents a study of customer analysis, principle and concept of data mining as well as its applications on churn rate prediction, and on strategic improvement in customer services. The improvement aims at increasing customer retention and reducing customer defection by reacting appropriately and efficiently to customer demand.

กิตติกรรมประกาศ

โครงการศึกษานี้ สำเร็จลุล่วงได้ ข้าพเจ้าขอขอบพระคุณ ผศ.ดร.วรพจน์ กริสุระเดช อาจารย์ที่ปรึกษาวิชาโครงการศึกษาที่ได้กรุณาใช้เวลาให้ความรู้ คำปรึกษา และคำแนะนำต่างๆ อันเป็นประโยชน์ต่อการศึกษาโครงการนี้

นอกจากนี้ข้าพเจ้าขอกราบขอบพระคุณบุพการี และบุคคลในครอบครัวที่ได้ให้การสนับสนุนส่งเสริมเป็นกำลังใจในการเรียนตลอดมา รวมทั้งเพื่อนร่วมงาน บริษัท ทศท คอร์ปอเรชั่น จำกัด (มหาชน) และเพื่อนร่วมรุ่น ITM12 ภาคสมทบที่มีส่วนช่วยเหลือ และสนับสนุนให้โครงการศึกษานี้สำเร็จลุล่วงด้วยดี

ข้าพเจ้าหวังเป็นอย่างยิ่งว่าโครงการศึกษานี้ จะเป็นประโยชน์แก่ผู้ที่สนใจ สำหรับข้อบกพร่องของโครงการศึกษานี้ ข้าพเจ้าขอรับไว้ เพื่อนำไปปรับปรุงแก้ไขในโอกาสต่อไป

คลใจ กระแสเสน

ตุลาคม 2547

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ.....	IV
สารบัญตาราง	VI
สารบัญรูป.....	IX
บทที่	
1. บทนำ.....	1
1.1 ความเป็นมา.....	1
1.2 วัตถุประสงค์ของการศึกษา.....	1
1.3 ขอบเขตของการศึกษา.....	1
1.4 ประโยชน์ที่คาดว่าจะได้รับ	2
1.5 ขั้นตอนในการศึกษา.....	2
2. บริการโทรศัพท์พื้นฐาน	3
2.1 ข้อมูลทั่วไป.....	3
2.2 การประกอบธุรกิจ.....	3
2.3 สถานภาพการให้บริการ โทรศัพท์พื้นฐาน	5
3. คาด้าไมนิ่งและทฤษฎีที่เกี่ยวข้อง.....	10
3.1 การทำคาด้าไมนิ่ง	10
3.2 ความหมายของคาด้าไมนิ่ง	10
3.3 วิวัฒนาการของคาด้าไมนิ่ง	11
3.4 ขั้นตอนในการทำคาด้าไมนิ่ง.....	12
3.5 แหล่งที่มาของข้อมูลในการทำคาด้าไมนิ่ง.....	16
3.6 การพิจารณาเลือกใช้วิธีคาด้าไมนิ่ง.....	17
3.7 การประยุกต์ใช้คาด้าไมนิ่ง.....	21
3.7 Algorithm ที่ใช้ในการศึกษา.....	22

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้เผยแพร่โดยไม่เสียค่าใช้จ่าย

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
4. วิธีการดำเนินการศึกษา.....	34
4.1 ข้อมูลที่ศึกษา.....	34
4.2 การเตรียมข้อมูล	37
4.3 การทำคาด้าไม่นิ่ง	60
4.4 การวิเคราะห์ผลลัพธ์ที่ได้	65
4.4 การนำมาประยุกต์ใช้กับธุรกิจ	87
5. สรุปผลการศึกษาและข้อเสนอแนะ	88
5.1 การทำคาด้าไม่นิ่ง	88
5.2 ปัญหาและข้อจำกัด.....	89
บรรณานุกรม.....	90
ภาคผนวก ก ตัวอย่างการทำงานของโปรแกรม.....	91
ประวัติผู้เขียน.....	121

สารบัญตาราง

หน้า

ตารางที่

2.1	แสดงสถานภาพเลขหมายเปิดใช้ และเลขหมายที่ยกเลิกของบริษัท ทศท คอร์ปอเรชั่น จำกัด (มหาชน) และ บริษัท โทร คอร์ปอเรชั่น จำกัด (มหาชน)	9
3.1	ค่าใดบ้างที่มีรูปแบบตามแอปพลิเคชัน การดำเนินงาน และเทคนิคที่ใช้	18
3.2	K-Means Input Values	22
3.3	Several Applications of the K-Means Algorithm (K=2)	24
3.4	ตาราง Contingency ของ (? ²)	25
3.5	แสดงการกระจายของข้อมูลตัวแปร Y ใน Root Node	26
3.6	แสดงความสัมพันธ์ระหว่าง Y และ X1	27
3.7	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 1 และ 2	27
3.8	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 1 และ 3	28
3.9	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 1 และ 4	28
3.10	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 2 และ 3	28
3.11	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 2 และ 4	28
3.12	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 3 และ 4	29
3.13	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 1 และ 3,4	29
3.14	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 2 และ 3,4	29
3.15	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 1 และ 2	30
3.16	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 1 และ 2,3,4	30
3.17	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X2	31
3.18	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X2 กรณี 0 และ 1	31
3.19	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X2 กรณี 0 และ 2	31
3.20	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X2 กรณี 1 และ 2	32
3.21	แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 0,1 และ 2	32
4.1	แสดงตารางข้อมูลพื้นฐานลูกค้า	35
4.2	แสดงตารางข้อมูลการใช้โทรศัพท์	37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่หรือใช้ซ้ำโดยไม่ได้รับอนุญาต

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
4.3 แสดงข้อมูลการรับชำระค่าใช้บริการ	37
4.4 ตารางแสดง โครงสร้างข้อมูลพื้นฐานลูกค้า	39
4.5 ข้อมูลพื้นฐานลูกค้า	40
4.6 โครงสร้างข้อมูลการใช้โทรศัพท์	41
4.7 ข้อมูลการใช้โทรศัพท์	41
4.8 โครงสร้างข้อมูลการชำระเงิน	42
4.9 ข้อมูลการชำระเงิน	42
4.10 โครงสร้างข้อมูลตารางความสัมพันธ์ของข้อมูลพื้นฐานลูกค้า และข้อมูลการ ใช้โทรศัพท์	44
4.11 ข้อมูลพื้นฐานและข้อมูลการใช้โทรศัพท์	45
4.12 โครงสร้างข้อมูลจากข้อมูลลูกค้า ข้อมูลการใช้โทรศัพท์ และข้อมูลการชำระเงิน ..	46
4.13 ข้อมูลข้อมูลการใช้โทรศัพท์ และข้อมูลการชำระเงิน	47
4.14 ข้อมูลสำหรับทำค่างานนี้	47
4.15 ค่าสถิติของข้อมูล TENURE	49
4.16 ค่าสถิติของข้อมูล MONTH	49
4.17 ค่าสถิติของข้อมูล P_METHOD	50
4.18 ค่าสถิติของข้อมูล ACC_TYPE	50
4.19 ค่าสถิติของข้อมูล METER100	51
4.20 ค่าสถิติของข้อมูล METER8M	51
4.21 ค่าสถิติของข้อมูล LDCHG8M	52
4.22 ค่าสถิติของข้อมูล PAID9M	52
4.23 ค่าสถิติของข้อมูล T_LATE	53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

	หน้า
ตารางที่	
4.24 ตารางข้อมูลที่ผ่านการ Recode ข้อมูลจากข้อมูลเชิงกลุ่มเป็นข้อมูลเชิงปริมาณ	57
4.25 ตารางข้อมูลที่ผ่านการ Normalization	58
4.26 การแบ่งกลุ่มลูกค้ำเป็น 8 กลุ่ม	60
4.27 จำนวนสมาชิกในการแบ่งกลุ่มลูกค้ำเป็น 8 กลุ่ม	60
4.28 การแบ่งกลุ่มลูกค้ำเป็น 8 กลุ่มตัด Attribute : FEMETR4	61
4.29 จำนวนสมาชิกในการแบ่งกลุ่มลูกค้ำเป็น 8 กลุ่ม ตัด Attribute : FEMETR4	61
4.30 การแบ่งกลุ่มลูกค้ำเป็น 6 กลุ่ม	62
4.31 จำนวนสมาชิกในการแบ่งกลุ่มลูกค้ำเป็น 6 กลุ่ม	62
4.32 การแบ่งกลุ่มลูกค้ำเป็น 4 กลุ่ม	62
4.33 จำนวนสมาชิกในการแบ่งกลุ่มลูกค้ำเป็น 4 กลุ่ม	63

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

หน้า

รูปที่

2.1	แสดงส่วนแบ่งเลขหมายที่มีลูกค้า	6
2.2	แสดงสถิติเลขหมายโทรศัพท์พื้นฐานในเขตโทรศัพท์นครหลวงตั้งแต่ปี 2536 – 2547	7
2.3	แสดงสถิติอัตราการสูญเสียเลขหมายของ บมจ. ทศท ตั้งแต่ปี 2541 - 2546	7
3.1	แสดงขั้นตอนการทำค้ำไม้หนึ่ง	12
3.2	สัดส่วนของเวลาในแต่ละขั้นตอนการทำค้ำไม้หนึ่ง	16
3.3	รูปแบบการทำนาย (Predictive Modeling)	19
3.4	แสดงการจัดกลุ่มข้อมูล (Database Segmentation)	19
3.5	แสดงแบบแผนของข้อมูล (Pattern Matching)	20
3.6	แสดงแบบคิซึซันทรี (Decision Tree)	20
3.7	แสดงนิวรอลเน็ตเวิร์ค (Neural Network)	21
3.8	กลุ่มของข้อมูลจากตารางที่ 3.2	23
3.9	การจัดกลุ่มข้อมูลของตารางที่ 3.2	24
3.10	แสดง Tree ของข้อมูลจากการใช้ CHAID Algorithm	33
4.1	การกระจายของข้อมูล	48
4.2	การกำหนดค่า Interval ของ METER8M	53
4.3	New Histogram METER8M	54
4.4	การกำหนดค่า Interval ของ LDCHG8M	54
4.5	New Histogram LDCHG8M	55
4.6	การกำหนดค่า Interval ของ PAID9M	55
4.7	New Histogram PAID9M	56
4.8	Histogram ใหม่หลังจาก Normalization	59
4.9	ผลลัพธ์การหา Pattern ของกลุ่มลูกค้าที่ยกเลิกการใช้บริการโทรศัพท์	64
4.10	Node 4 ลักษณะลูกค้ากลุ่มที่ 1	66

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.11 Node 23 ลักษณะลูกค้ำกลุ่มที่ 1	66
4.12 Node 75 ลักษณะลูกค้ำกลุ่มที่ 1	67
4.13 Node 13 ลักษณะลูกค้ำกลุ่มที่ 2	68
4.14 Node 15 ลักษณะลูกค้ำกลุ่มที่ 2	69
4.15 Node 17 ลักษณะลูกค้ำกลุ่มที่ 2	69
4.16 Node 19 ลักษณะลูกค้ำกลุ่มที่ 2	70
4.17 Node 90 ลักษณะลูกค้ำกลุ่มที่ 2	70
4.18 Node 102 ลักษณะลูกค้ำกลุ่มที่ 2	71
4.19 Node 16 ลักษณะลูกค้ำกลุ่มที่ 3	74
4.20 Node 34 ลักษณะลูกค้ำกลุ่มที่ 3	74
4.21 Node 40 ลักษณะลูกค้ำกลุ่มที่ 3	75
4.22 Node 50 ลักษณะลูกค้ำกลุ่มที่ 3	75
4.23 Node 53 ลักษณะลูกค้ำกลุ่มที่ 3	76
4.24 Node 59 ลักษณะลูกค้ำกลุ่มที่ 3	76
4.25 Node 77 ลักษณะลูกค้ำกลุ่มที่ 3	77
4.26 Node 78 ลักษณะลูกค้ำกลุ่มที่ 3	77
4.27 Node 80 ลักษณะลูกค้ำกลุ่มที่ 3	78
4.28 Node 81 ลักษณะลูกค้ำกลุ่มที่ 3	78
4.29 Node 83 ลักษณะลูกค้ำกลุ่มที่ 3	79
4.30 Node 87 ลักษณะลูกค้ำกลุ่มที่ 3	79
4.31 Node 95 ลักษณะลูกค้ำกลุ่มที่ 3	80
4.32 Node 97 ลักษณะลูกค้ำกลุ่มที่ 3	80
4.33 Node 103 ลักษณะลูกค้ำกลุ่มที่ 3	81

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.34 Node 106 ลักษณะลูกค้ำกลุ่มที่ 3	81
4.35 Node 18 ลักษณะลูกค้ำกลุ่มที่ 4	83
4.36 Node 32 ลักษณะลูกค้ำกลุ่มที่ 4	83
4.37 Node 38 ลักษณะลูกค้ำกลุ่มที่ 4	84
4.38 Node 63 ลักษณะลูกค้ำกลุ่มที่ 4	84
4.39 Node 65 ลักษณะลูกค้ำกลุ่มที่ 4	85
4.40 Node 79 ลักษณะลูกค้ำกลุ่มที่ 4	85
4.41 Node 84 ลักษณะลูกค้ำกลุ่มที่ 4	86
4.42 ค่าความเสี่ยงในการพยากรณ์	86

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมา

ในปัจจุบันธุรกิจสื่อสารโทรคมนาคมได้เจริญเติบโตอย่างรวดเร็ว การแข่งขันในธุรกิจโทรศัพท์พื้นฐานมีโอกาสเพิ่มสูงขึ้น เนื่องมาจากการให้สัมปทานกับบริษัทเทเลคอมเอเชีย จำกัด (มหาชน) ผู้ให้บริการโทรศัพท์ในเขตนครหลวงจำนวน 2.6 ล้านเลขหมาย และบริษัท ทีทีเอ็นดีที จำกัด (มหาชน) จำนวน 1.5 ล้านเลขหมาย รวมทั้งการเปิดเสรีโทรคมนาคมที่จะเกิดขึ้นในปี 2549

โทรศัพท์พื้นฐานเป็นรายได้หลักของ บมจ. ทศท ซึ่งใช้ในการติดต่อกับโทรศัพท์พื้นฐานด้วยกันภายในเขตพื้นที่เดียวกัน, ต่างพื้นที่, ต่างประเทศ และติดต่อไปยังโทรศัพท์เคลื่อนที่ โดยให้บริการทั้งในเขตนครหลวงและภูมิภาคจำนวน 1.5 และ 1.8 ล้านกว่าเลขหมายตามลำดับ ข้อมูลลูกค้าที่ใช้บริการโทรศัพท์พื้นฐานในระบบงานต่างๆ ขององค์กรยังไม่ได้ถูกนำมาศึกษาและวิเคราะห์อย่างจริงจัง ทำให้องค์กรมีส่วนแบ่งทางการตลาดโทรศัพท์พื้นฐานในเขตนครหลวงน้อยกว่าบริษัทคู่แข่ง ดังนั้นจำเป็นต้องโดยใช้ Data Mining เพื่อศึกษาและวิเคราะห์ข้อมูลในการทำนายพฤติกรรมลูกค้าที่มีแนวโน้มจะยกเลิกการใช้บริการโทรศัพท์พื้นฐาน (Churn) และนำไปปรับปรุงกลยุทธ์ในการดำเนินธุรกิจให้สามารถรักษาลูกค้าในปัจจุบัน และลดการสูญเสียลูกค้าให้กับบริษัทคู่แข่ง โดยสามารถตอบสนองความต้องการของลูกค้าได้อย่างเหมาะสมและมีประสิทธิภาพ

1.2 วัตถุประสงค์การศึกษา

การทำนายลูกค้าที่มีแนวโน้มจะยกเลิกการใช้โทรศัพท์พื้นฐาน โดยใช้ Data Mining มีวัตถุประสงค์เพื่อศึกษา และวิเคราะห์พฤติกรรมลูกค้า จากการทำนายข้อมูลดังกล่าว เพื่อเป็นข้อมูลสนับสนุนการจัดทำแผนการตลาดในการรักษาลูกค้าปัจจุบัน และดึงลูกค้าเก่ากลับมาใช้บริการตามเดิม

1.3 ขอบเขตของการศึกษา

ศึกษา และวิเคราะห์ข้อมูลเลขหมายของฝ่ายปฏิบัติการโทรศัพท์ประจำที่นครหลวง ซึ่งเป็นข้อมูลที่อยู่กระจัดกระจายในฐานข้อมูล แต่ละระบบงานจำนวนมากไม่ว่าจะเป็น Relational

Database หรือ Transactional Database ได้แก่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. ข้อมูลพื้นฐานลูกค้า(Customer Information)
2. ข้อมูลการใช้โทรศัพท์
3. ข้อมูลรับชำระค่าใช้บริการทุกๆ 15 วัน

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถเข้าใจหลักการทำ Data Mining และการเลือกใช้รูปแบบและเทคนิค ที่เหมาะสมกับวัตถุประสงค์
2. ศึกษารูปแบบพฤติกรรมของลูกค้าที่มีแนวโน้มจะยกเลิกการใช้บริการ โทรศัพท์
3. ช่วยให้องค์กรสามารถนำเสนอสารสนเทศที่ได้ไปปรับปรุงกลยุทธ์ทางการตลาด เช่น รักษาลูกค้าในปัจจุบัน โดยทำให้การยกเลิกการใช้บริการโทรศัพท์ลดลง ดึงลูกค้าเก่ากลับมาใช้บริการ โทรศัพท์เหมือนเดิม

1.5 ขั้นตอนในการศึกษา

1. ศึกษาค้นคว้าเทคนิคที่จะมาช่วยในการทำนายพฤติกรรมของลูกค้าที่มีแนวโน้มจะยกเลิกการใช้บริการ โทรศัพท์
2. เก็บรวบรวมข้อมูลเลขหมายของฝ่ายปฏิบัติการ โทรศัพท์ประจำที่นครหลวง ซึ่งเป็นข้อมูลที่อยู่กระจัดกระจายในฐานข้อมูลต่างๆ เช่น ข้อมูลพื้นฐานลูกค้า(Customer Information) ข้อมูลการใช้โทรศัพท์ และข้อมูลรับชำระค่าใช้บริการ เป็นต้น
3. ศึกษา และวิเคราะห์ข้อมูลดังกล่าวตามขั้นตอนของ Data Mining
4. สรุปผลการศึกษาที่ได้จากการทำ Predictive

บทที่ 2

การให้บริการโทรศัพท์พื้นฐาน

2.1 ข้อมูลทั่วไป

บริษัท ทศท คอร์ปอเรชั่น จำกัด (มหาชน) จัดทะเบียนจัดตั้งเป็นบริษัทมหาชนจำกัด เมื่อวันที่ 31 กรกฎาคม พ.ศ. 2545 โดยแปลงสภาพมาจากองค์การโทรศัพท์แห่งประเทศไทย (ทศท.) ซึ่งเป็นรัฐวิสาหกิจที่จัดตั้งขึ้นภายใต้พระราชบัญญัติองค์การโทรศัพท์แห่งประเทศไทย พ.ศ. 2497 ที่ได้รับทุนสนับสนุนจากรัฐบาลไทย โดยแต่เดิม ทศท. จะดำเนินงานภายใต้กฎและข้อบังคับของ รัฐวิสาหกิจ โดยไม่ต้องเสียภาษีเงินได้จากกำไรสุทธิ แต่จะต้องจัดสรรกำไรสุทธิและนำส่ง กระทรวงการคลัง โดยนับตั้งแต่วันที่ 31 กรกฎาคม พ.ศ. 2545 เป็นต้นไป บริษัทได้รับโอนกิจการ สิทธิ หนี้ ความรับผิดชอบ สินทรัพย์ ส่วนของทุนและพนักงานทั้งหมดจาก ทศท. ภายใต้พระราชบัญญัติ ทุนรัฐวิสาหกิจ พ.ศ. 2542 และบริษัทได้รับ ำเนินไม่ต้องปฏิบัติตามคำสั่ง ระเบียบ ข้อบังคับ และมติคณะรัฐมนตรีที่ใช้กับรัฐวิสาหกิจทั่วไป

2.2 การประกอบธุรกิจ

บริษัท ทศท คอร์ปอเรชั่น จำกัด (มหาชน) เป็นผู้ให้บริการโทรศัพท์รายแรกของประเทศ ตั้งแต่ปี 2497 จนถึงปัจจุบัน ประกอบธุรกิจหลักในด้านโทรคมนาคม อันได้แก่ การให้บริการ โทรศัพท์ประจำที่ โทรศัพท์สาธารณะทั้งแบบใช้เหรียญและแบบใช้บัตร โทรศัพท์รหัสส่วนตัว บริการสื่อสารร่วมระบบดิจิทัล (ISDN) โทรศัพท์ไร้สาย โทรศัพท์ติดตามตัว บริการสื่อสารข้อมูล บริการวีดีโอ บริการคู่สายเช่า และวงจรเช่า บริการอินเทอร์เน็ต และธุรกิจอื่นที่เกี่ยวข้องและ ใกล้เคียงกัน ทั้งในประเทศ ระหว่างประเทศ และนอกประเทศ โดยแบ่งกลุ่มการให้บริการ ดังนี้

1. การบริการโทรศัพท์พื้นฐาน ประกอบด้วย โทรศัพท์ประจำที่ (Fixed Line) และ โทรศัพท์สาธารณะ (Public Payphone)

- โทรศัพท์ประจำที่ (Fixed Line) สำหรับลูกค้าบ้านพักอาศัย หน่วยงานราชการ และ ธุรกิจ

- โทรศัพท์สาธารณะ (Public Payphone) ทศท ดำเนินการติดตั้งทั้งที่เป็นเครื่องแบบ หยอดเหรียญ แบบใช้บัตร แบบหยอดเหรียญและใช้บัตรในเครื่องเดียวกัน เพื่ออำนวยความสะดวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการใช้งาน และตอบสนองความต้องการของลูกค้าที่หลากหลาย เป็นผลให้มีโทรศัพท์สาธารณะใช้ทั่วประเทศจำนวนทั้งสิ้น 270,917 เลขหมาย

2. บริการคู่สายเช่า (Leased Lines) และวงจรเช่า (Leased Circuits) บริการคู่สายเช่า และวงจรเช่าความเร็วสูงในการรับส่งข้อมูล

3. บริการสื่อสารร่วมระบบดิจิทัล (Integrated Service Digital Network – ISDN) เป็นบริการสื่อสารระบบดิจิทัลความเร็วสูงที่มีความเร็วตั้งแต่ 64 Kbps ถึง 2.048 Mbps สามารถรองรับการส่งสัญญาณเสียง ข้อมูล และภาพ

4. บริการโทรศัพท์เคลื่อนที่ ให้บริการโทรศัพท์เคลื่อนที่ระบบ 470 NMT และ 1900 MHz ของกิจการร่วมการค้าไทย – โมบาย

5. บริการโทรศัพท์รหัสส่วนตัว 108 (PIN Phone 108) PINPHONE หรือ PERSONAL IDENTIFICATION NUMBER PHONE ทำงานด้วยระบบคอมพิวเตอร์ มีความแม่นยำ และปลอดภัยสูง รหัสมีมูลค่าตามหน้าบัตร สามารถโทรได้จนรหัสหมดมูลค่า มีรหัสตัวเลข 14 หลัก สามารถติดต่อปลายทางที่ใดก็ได้ไม่จำกัด แม้เครื่องโทรศัพท์ ที่ทำงานถือการโทรทางไกล หักค่าใช้จ่ายบริการจากมูลค่ารหัส

6. บริการโทรศัพท์ระหว่างประเทศ เปิดให้บริการโทรศัพท์ระหว่างประเทศกับประเทศเพื่อนบ้านติดชายแดน คือ มาเลเซีย ลาว กัมพูชา และพม่า ได้แก่บริการโทรศัพท์ระหว่างประเทศ 007 (IDD 007) บริการโทรประหยัดระหว่างประเทศ 002 (I SAVE CALL 002) บริการผ่านพนักงานสลับสาย 101 (Operator Assisted Call 101) บริการโทรกลับบ้านเพื่อคนไทยในต่างแดน (Calling Home Service) บริการโทรศัพท์ระหว่างประเทศเรียกเก็บเงินปลายทาง เพื่อนักท่องเที่ยวต่างชาติในประเทศไทย (Collect Call Service) บริการวงจรเช่าระบบดิจิทัล (International Private Leased Circuits) และบริการโทรฟรีระหว่างประเทศ (International FreePhone Services : IFS)

7. บริการบนโครงข่าย IP (Internet Protocol) ทศท มีโครงข่าย IP ที่สามารถรองรับบริการประเภท Voice และ Data การให้บริการประเภท Voice ประกอบด้วย บริการ Y-tel 1234 และ I SAVE CALL 002 และบริการ Data ประกอบด้วยบริการ Remote Internet Access และ VPDN

8. บริการบนโครงข่ายอัจฉริยะ (Intelligence Network – IN) เปิดบริการใหม่ 3 บริการคือ บริการเลขหมายเดียวทั่วไทย (One Number 1800) บริการทายผลกีฬาหรือเกมส์ต่างๆ (Vote Now) บริการเลขหมายพิเศษ (Follow Me)

9. บริการรูปแบบใหม่และทันสมัย บริการรูปแบบใหม่ๆ เช่น บริการรับฝากข้อความ (i-BOX 1278) บริการอินเทอร์เน็ตไร้สายความเร็วสูง (Wi-Fi Flexinet) บริการแสดงเลขหมายเรียกเข้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Caller ID) บริการสอบถามข้อมูลภาครัฐเพื่อประชาชน (1111) บริการออกใบรับรองอิเล็กทรอนิกส์ (Certificate Authority) และบริการส่งข้อความสั้น (SMS) บนโครงข่ายโทรศัพท์ประจำที่

10. บริการที่ให้บริษัทเอกชนร่วมให้บริการ บริษัทได้เปิดโอกาสให้บริษัทเอกชนเข้าร่วม การงานและร่วมลงทุน เพื่อขยายโทรศัพท์พื้นฐานและบริการ โทรคมนาคมที่เกี่ยวข้องต่างๆ ได้แก่

- บริษัท เทเลคอมเอเชียคอร์ปอเรชั่น จำกัด (มหาชน) ปัจจุบันเปลี่ยนชื่อเป็นบริษัท ทูรคอร์ดคอร์ปอเรชั่น จำกัด (มหาชน) ให้บริการโทรศัพท์พื้นฐานจำนวน 2.6 ล้านเลขหมาย ในเขตนครหลวงครอบคลุมพื้นที่กรุงเทพฯ และปริมณฑล (นนทบุรี สมุทรปราการ และปทุมธานี)

- บริษัท ไทยเทเลโฟนแอนด์เทเลคอมมิวนิเคชั่น จำกัด (มหาชน) ให้บริการโทรศัพท์พื้นฐานจำนวน 1.5 ล้านเลขหมายในเขตโทรศัพท์ภูมิภาค

- บริษัท ฮัทชีสัน เทเลคอมมิวนิเคชั่นส์ (ประเทศไทย) จำกัด ให้บริการโทรศัพท์ติดตามตัวเพจโฟน

- บริษัท แอดวานซ์ ดาต้าเน็ตเวิร์ค คอมมิวนิเคชั่น จำกัด ให้บริการสื่อสารข้อมูลทุกประเภทโดยใช้ระบบ Datakit Virtual Circuit Switch

- องค์กรรับส่งสินค้าและพัสดุภัณฑ์ ให้บริการโทรศัพท์เคลื่อนที่ระบบ Common Base Radio Telephone

- บริษัท แอดวานซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน) ให้บริการโทรศัพท์เคลื่อนที่เซลลูลาร์ 900 ระบบ NMT, GSM และ Prepaid

- บริษัท คอม – ลิงค์ จำกัด ให้บริการระบบเชื่อมโยงโครงข่ายเคเบิลใยแก้วนำแสงตามเส้นทางรถไฟ

2.3 สถานภาพการให้บริการโทรศัพท์พื้นฐาน

บมจ. ทศท ให้บริการโทรศัพท์พื้นฐานดำเนินการทั้งในเขตโทรศัพท์นครหลวงและภูมิภาค ซึ่งในปัจจุบันได้มีการพัฒนา และขยายการติดตั้งอย่างต่อเนื่อง เพื่อให้ลูกค้าได้รับบริการที่รวดเร็ว และมีความสะดวกเพิ่มขึ้น- นอกจากนี้ บมจ. ทศท ได้ยกเลิกการเก็บ และคืนเงินประกันให้กับลูกค้าในรูปแบบต่างๆ และปรับปรุงการคุณภาพการให้บริการให้ดีขึ้น

คู่แข่งทางการค้า และการแข่งขัน บมจ. ทศท มีคู่แข่งทางการค้าในธุรกิจบริการโทรศัพท์พื้นฐาน ได้แก่

1. บริษัท เทเลคอมเอเชียคอร์ปอเรชั่น จำกัด (มหาชน) ปัจจุบันเปลี่ยนชื่อเป็นบริษัท ทูรคอร์ดคอร์ปอเรชั่น จำกัด (มหาชน) เข้าร่วมการงานและร่วมลงทุน เพื่อขยายบริการโทรศัพท์พื้นฐานในเขตโทรศัพท์นครหลวงครอบคลุมพื้นที่กรุงเทพฯ และปริมณฑล (นนทบุรี สมุทรปราการ และ

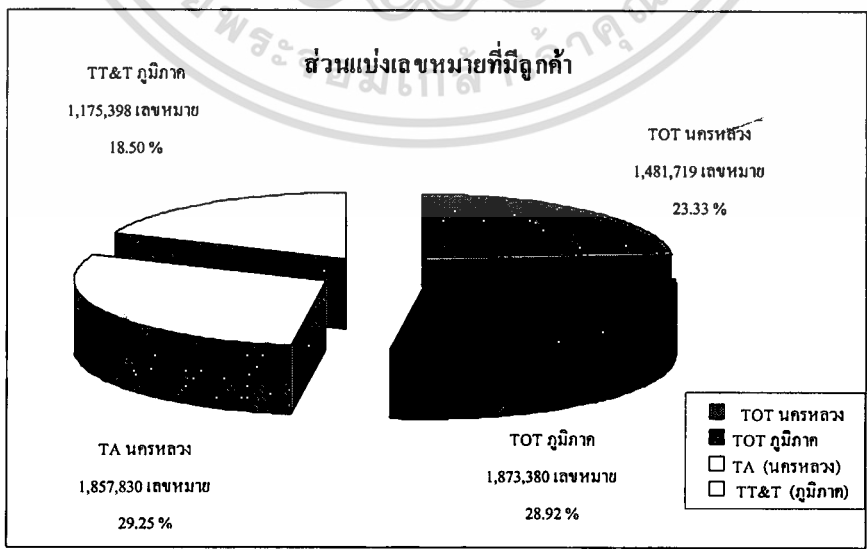
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปทุมธานี) จำนวน 2.6 ล้านเลขหมาย บริษัท ทู ได้ดำเนินการทางการตลาดอย่างต่อเนื่อง นับว่าเป็นคู่แข่งที่สำคัญ เนื่องจากบริษัทฯ มีความคล่องตัว มีประสิทธิภาพในการดำเนินธุรกิจ และการตลาดมากกว่า ทำให้สามารถแย่งชิงลูกค้ารายใหม่ และลูกค้ารายเดิมจาก บมจ. ทศท ได้

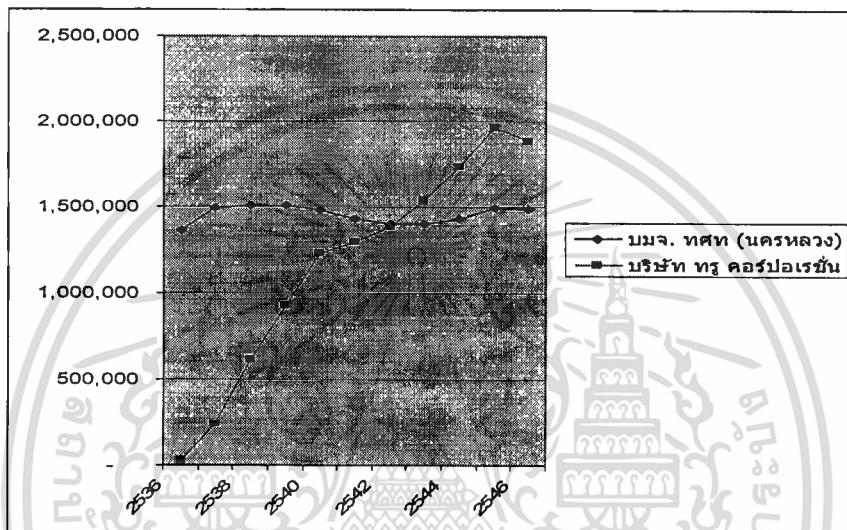
2. บริษัท ไทยเทเลโฟนแอนด์เทเลคอมมิวนิเคชั่น จำกัด (มหาชน) (TT&T) ให้บริการโทรศัพท์พื้นฐานในเขตโทรศัพท์ภูมิภาค จำนวน 1.5 ล้านเลขหมาย ปัจจุบันบริษัทมีปัญหาในการดำเนินธุรกิจ โดยเฉพาะอย่างยิ่งในเรื่องฐานะทางการเงิน นอกจากนี้ต้องจ่ายส่วนแบ่งรายได้ให้ บมจ. ทศท ถึงร้อยละ 43.1 – 44.5 ของรายได้ ค่าบริการเป็นอีกปัจจัยที่เป็นอุปสรรคต่อการแข่งขันในการให้บริการในเขตโทรศัพท์ภูมิภาค ทำให้ บมจ. ทศท ต้องเผชิญการแข่งขันในเขตโทรศัพท์นครหลวงมากกว่า

สถานการณ์ให้บริการโทรศัพท์ของ บมจ. ทศท นับตั้งแต่ให้บริษัทเอกชนเข้าร่วมการทำงานและร่วมลงทุน เพื่อขยายโทรศัพท์พื้นฐานในเขตโทรศัพท์นครหลวงและภูมิภาค พบว่า ณ เดือนมี.ค. 2547 เลขหมายโทรศัพท์ประจำที่ บริษัท ทู คอร์ปอเรชั่น จำกัด (มหาชน) มีส่วนแบ่งเลขหมายที่มีลูกค้ามากที่สุดจำนวนทั้งสิ้น 1,857,830 เลขหมายคิดเป็นร้อยละ 29.25 อันดับที่สองคือ บมจ. ทศท ในเขตโทรศัพท์ภูมิภาคมีจำนวนทั้งสิ้น 1,837,380 เลขหมาย คิดเป็นร้อยละ 28.92 อันดับที่สาม คือ บมจ. ทศท ในเขตโทรศัพท์นครหลวงมีจำนวนทั้งสิ้น 1,481,719 เลขหมายคิดเป็นร้อยละ 23.33 และอันดับสุดท้าย คือ บริษัท ไทยเทเลโฟนแอนด์เทเลคอมมิวนิเคชั่น จำกัด (มหาชน) (TT&T) มีจำนวนทั้งสิ้น 1,175,398 เลขหมายคิดเป็นร้อยละ 18.50 ของเลขหมายที่มีลูกค้าทั้งหมด ดังแสดงในรูปที่ 2.1

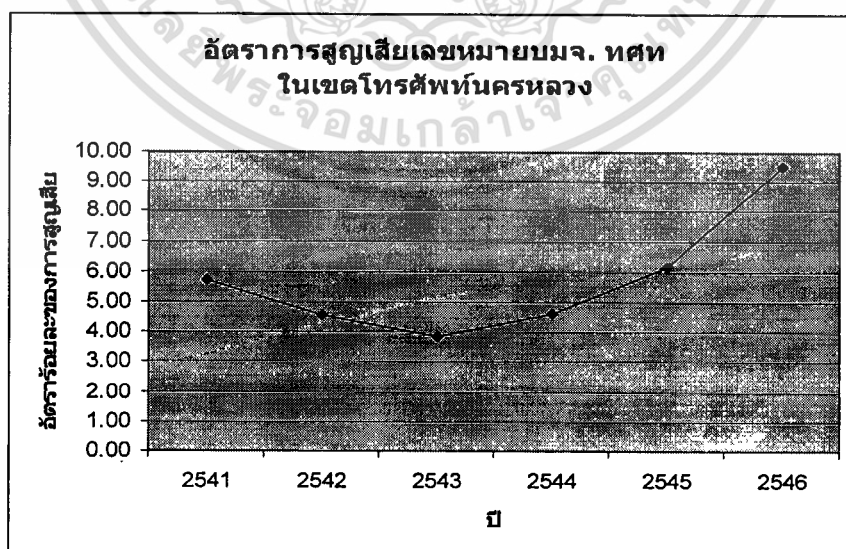


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับรูปที่ 2.1 แสดงส่วนแบ่งเลขหมายที่มีลูกค้าให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สถานภาพการให้บริการโทรศัพท์พื้นฐาน โดยเฉพาะในเขตโทรศัพท์นครหลวง บริษัท โทรคอร์ปอเรชั่น จำกัด (มหาชน) มีการขยายตัวของการให้บริการเพิ่มขึ้นอย่างต่อเนื่องเมื่อเปรียบเทียบกับ บมจ. ทศท ดังตารางที่ 2.1 และ รูปที่ 2.2 ส่วนอัตราการสูญเสียเลขหมาย (จำนวนเลขหมายที่ยกเลิก/จำนวนเลขหมายที่มีอยู่ทั้งหมด) ของ บมจ. ทศท เพิ่มขึ้นอย่างต่อเนื่อง นับตั้งแต่ปี 2543 – 2546 ดังแสดงในตารางที่ 2.1 และรูปที่ 2.3



รูปที่ 2.2 แสดงสถิติเลขหมายโทรศัพท์พื้นฐานในเขตโทรศัพท์นครหลวงตั้งแต่ปี 2536 – 2547



รูปที่ 2.3 แสดงสถิติอัตราการสูญเสียเลขหมายของ บมจ. ทศท ตั้งแต่ปี 2541 - 2546

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปัจจุบันการแข่งขันในการให้บริการโทรคมนาคม โดยเฉพาะอย่างยิ่งในธุรกิจโทรศัพท์พื้นฐาน มีแนวโน้มสูงขึ้น เช่น บริษัท ทู คอร์ปอเรชั่น จำกัด (มหาชน) ออกแพ็คเกจค่าโทรศัพท์ใหม่ บริษัท ทศท. คอร์ปอเรชั่น (ทศท.) ลดค่าโทรทางไกล และบริษัท ทีทีเอ็นดีที ปรับราคาโทรศัพท์ และค่าบริการรายเดือน นอกจากนี้ยังมีบริการโทรศัพท์เคลื่อนที่ที่ถือว่าเป็นบริการสำคัญที่เข้ามาทดแทน ซึ่งบริการดังกล่าวมีแนวโน้มขยายตัวเพิ่มขึ้นอย่างรวดเร็ว อีกทั้งยังมีศักยภาพที่ดีขึ้น และสามารถให้บริการที่หลากหลาย ส่งผลให้สามารถแย่งชิงลูกค้าโทรศัพท์พื้นฐานไปมากขึ้น และภายหลังจากการเปิดเสรีโทรคมนาคมในปี 2549 บมจ. ทศท. จะมีผู้ให้บริการรายใหม่ ทั้งจากภายในประเทศ และภายนอกประเทศ เข้าสู่ตลาดบริการโทรคมนาคมมากขึ้น ในขณะที่ผู้ให้บริการรายเดิมยังมีศักยภาพทั้งด้านการเงินและเทคโนโลยี มีผลทำให้การแข่งขันทวีความรุนแรงยิ่งขึ้น ซึ่งจะส่งผลกระทบต่อส่วนแบ่งทางการตลาดของโทรศัพท์พื้นฐาน ดังนั้น บมจ. ทศท. จึงได้กำหนดเป้าหมายเชิงยุทธศาสตร์ไว้คือรักษาลูกค้า ขยายฐานลูกค้าและเพิ่มส่วนครองการตลาด รวมถึงรายได้ให้กับทศท. โดยใช้การตลาดเชิงรุกที่มุ่งเข้าหาลูกค้า และการตอบสนองลูกค้าให้ตรงความต้องการของลูกค้าที่มีความแตกต่างในแต่ละกลุ่ม โดยเฉพาะกลุ่มลูกค้ารายใหญ่และกลุ่มลูกค้าที่มีศักยภาพสูงในการจะเพิ่มการใช้บริการด้านโทรคมนาคมในอนาคต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปี	2536	2537	2538	2539	2540	2541	2542	2543	2544	2545	2546
	เลขหมายที่มีลูกค้า										
บมจ. ทศท (นครหลวง)	1,363,420	1,489,965	1,505,982	1,500,967	1,472,134	1,423,842	1,397,971	1,395,004	1,425,800	1,492,211	1,480,166
บริษัท โทร คอร์ปอเรชั่น	30,973	241,169	615,770	933,191	1,235,147	1,299,702	1,381,539	1,535,634	1,736,774	1,961,764	1,886,547
เลขหมายรูดอน											
บมจ. ทศท (นครหลวง)	-	-	-	-	-	81,460	63,786	54,014	65,637	92,092	140,744
บริษัท โทร คอร์ปอเรชั่น	-	-	-	-	-	77,504	71,025	67,828	112,260	269,361	326,382
อัตราการสูญเสียลูกค้า ของบมจ. ทศท (%)						5.72	4.56	3.87	4.60	6.17	9.51

ตารางที่ 2.1 แสดงสถานภาพเลขหมายเปิดใช้ และเลขหมายที่ยกเลิกของบริษัท ทศท คอร์ปอเรชั่น จำกัด (มหาชน)
และ บริษัท โทร คอร์ปอเรชั่น จำกัด (มหาชน)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

ดาต้าไมนิ่งและทฤษฎีที่เกี่ยวข้อง

3.1 การทำดาต้าไมนิ่ง

การทำดาต้าไมนิ่ง โดยทั่วไปเป็นกระบวนการค้นหาความสัมพันธ์ของข้อมูลสารสนเทศที่มีอยู่ในฐานข้อมูลของบริษัทหรือองค์กรที่มีรูปแบบเฉพาะของข้อมูล ปริมาณข้อมูลที่มีอยู่จำนวนมาก มีบางลักษณะของข้อมูลที่มีความสัมพันธ์กันซ่อนอยู่ แต่ผู้ใช้ระบบทั่วไปรวมถึงผู้มีอำนาจการตัดสินใจไม่ทราบว่าลักษณะข้อมูลสารสนเทศที่เขามีนั้น มีความสัมพันธ์หรือเป็นข้อมูลที่มีความใกล้เคียงกัน เช่น โดยพื้นฐานทั่วไปของบุคคล ถ้าต้องการทำนายการขายสินค้า จะพิจารณาจากข้อมูลของยอดขายในอดีต ประกอบกับประสบการณ์ที่สะสมมา ซึ่งเป็นการคาดเดาแบบไม่มีหลักวิชาการ และเปอร์เซ็นต์ความถูกต้องต่ำ การทำดาต้าไมนิ่งจะช่วยให้การคาดเดาของการทำธุรกิจมีความเป็นจริงมากขึ้น รวมทั้งช่วยหารูปแบบการตัดสินใจ หรือวางแผนการตลาด และส่งเสริมการขาย

การทำดาต้าไมนิ่งประกอบด้วยหลายส่วน หลายสาขาวิชา นับตั้งแต่เทคนิคฐานข้อมูล (Database Technology) ปัญญาประดิษฐ์ (Artificial Intelligence) การเรียนรู้ด้วยตนเองของเครื่องคอมพิวเตอร์ (Machine Learning) นิวรอนเน็ตเวิร์ค (Neural Network) สถิติ (Statistics) รูปแบบที่สนใจ (Pattern Recognition) ระบบความรู้พื้นฐาน (Knowledge-Based Systems) ความรู้ได้เรียนผ่านมา (Knowledge Acquisition) การสืบค้นข้อมูล (Information Retrieval) การคำนวณประสิทธิภาพสูง (High-Performance Computing) และข้อมูลเสมือน (Data Visualization)

3.2 ความหมายของดาต้าไมนิ่ง

ดาต้าไมนิ่งเป็นกระบวนการนำข้อมูลที่ไม่รู้มาก่อนล่วงหน้า (Unknown) ข้อมูลที่มีความถูกต้อง (Valid) และข้อมูลที่สามารถนำไปใช้ประโยชน์ได้จริง (Actionable) จากฐานข้อมูลขนาดใหญ่ วิเคราะห์ให้ได้รับความรู้ใหม่ๆ หลังจากได้สาระสำคัญแล้ว จะใช้ข้อมูลสารสนเทศเหล่านี้เพื่อการตัดสินใจในการดำเนินธุรกิจ มีรายละเอียดของข้อมูลดังนี้

1) ข้อมูลที่ไม่รู้มาก่อนล่วงหน้า (Unknown) เป็นข้อมูลที่ผู้ใช้งานไม่รู้มาก่อนและไม่ชัดเจน ไม่สามารถตั้งสมมติฐานล่วงหน้าได้ว่าเป็นแบบใด เช่น เจ้าของห้างสรรพสินค้าเพิ่งจะค้นพบพฤติกรรมของลูกค้าว่า พ่อบ้านมักซื้อผ้าอ้อมและเบียร์พร้อมกันในวันศุกร์ ซึ่งข้อมูลดังกล่าว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สามารถสร้างความได้เปรียบให้กับห้างสรรพสินค้า โดยสามารถกำหนดแผนการขายหรือจุดวางสินค้าที่เหมาะสมสำหรับผู้ซื้อ ขณะที่คู่แข่งไม่รู้พฤติกรรมดังกล่าวของผู้บริโภค

2) ข้อมูลที่มีความถูกต้อง (Valid) เป็นข้อมูลที่มีความถูกต้อง เนื่องจากเมื่อมีผู้นำข้อมูลไปใช้ด้วยเทคนิคดาต้าไมนิ่งจะค้นพบสิ่งที่น่าสนใจตลอดเวลา แต่ต้องพิจารณาด้วยว่าสิ่งนั้นถูกต้องหรือไม่

3) ข้อมูลที่สามารถนำมาใช้ประโยชน์ได้จริง (Actionable) เป็นข้อมูลที่ได้จากการนำดาต้าไมนิ่งมาใช้ ซึ่งอาจนำมาช่วยในการตัดสินใจเพื่อชิงความได้เปรียบของธุรกิจ แต่ในบางครั้งข้อมูลที่ค้นพบอาจเป็นข้อมูลที่ผิดกฎหมายหรือคู่แข่งได้ค้นพบไปแล้ว ซึ่งหากนำมาใช้อาจไม่ก่อให้เกิดประโยชน์หรือไม่อาจกระทำผิดกฎหมาย ซึ่งการนำข้อมูลมาใช้ต้องคำนึงถึงความเหมาะสมด้วยการนำดาต้าไมนิ่งมาใช้ในการทำงานหลายด้านเช่น

ด้านการแพทย์ เช่น ทราบผลข้างเคียงจากการใช้ยา การวิเคราะห์เกี่ยวกับยีน หรือใช้ในการพยากรณ์แนวโน้มของผู้ป่วยจากการของโรคที่แสดงออก เพื่อลดภาระของแพทย์ในขั้นต้น

ด้านการเงิน เช่น การพยากรณ์เกี่ยวกับข้อมูลการค้าหุ้น หรือประเมินความเสี่ยงทั้งในการปล่อยสินเชื่อ

ด้านวิทยาศาสตร์ เช่น การค้นหาทฤษฎีใหม่ๆ

ด้านวิศวกรรม เช่น การสืบค้นความคิดปกติในการทำงาน (Fault Detection)

3.3 วิวัฒนาการของดาต้าไมนิ่ง

ในปี 1960 เทคโนโลยีฐานข้อมูลได้เริ่มพัฒนามาจาก file processing พื้นฐาน การค้นคว้าและพัฒนาระบบฐานข้อมูลมีมาเรื่อย ๆ

ปี 1970 ได้นำไปสู่การพัฒนาการเก็บข้อมูลในรูปแบบตาราง (Relational Database System) มีเครื่องมือจัดการโมเดลข้อมูล และมีเทคนิคการใช้อินเด็กซ์และการบริหารข้อมูล นอกจากนี้ผู้ใช้งานได้รับความสะดวกในการเข้าถึงข้อมูลโดยการใช้ภาษาในการเรียกข้อมูล (Query Language)

ปี 1980 เทคโนโลยีฐานข้อมูลได้เริ่มมีการปรับปรุงและพัฒนาในการหาระบบจัดการที่มีศักยภาพมากขึ้น ความก้าวหน้าในเทคโนโลยี hardware ใน 30 ปีที่ผ่านมา ได้นำไปสู่การจัดเก็บข้อมูลจำนวนมากที่มีความซับซ้อนได้อย่างมีประสิทธิภาพเพิ่มขึ้น

ปี 1990 - ปัจจุบัน สามารถจัดเก็บข้อมูลได้ในหลายรูปแบบ แตกต่างกันทั้งระบบปฏิบัติการหรือการจัดเก็บฐานข้อมูล ซึ่งการนำข้อมูลทั้งหมดมารวมและจัดเก็บไว้ในรูปแบบเดียวกันเรียกว่า Data Warehouse เพื่อความสะดวกในการจัดการต่อไป ซึ่งเทคโนโลยี Data Warehouse รวมไปถึงเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Data Cleansing , Data Integration และ On-Line Analytical Processing (OLAP) เป็นเทคนิคการวิเคราะห์ข้อมูลในหลาย ๆ มิติ นั้นได้เกิดขึ้นมาตามลำดับ

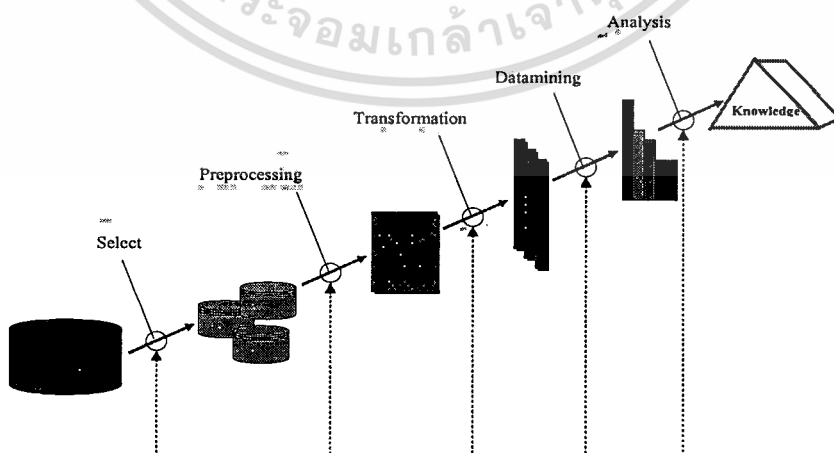
การละเลยข้อมูล ควบคู่ไปกับการขาดเครื่องมือวิเคราะห์ข้อมูลที่มีศักยภาพ นำไปสู่คำสถานการณ์ที่ว่า " ข้อมูลมาก แต่ความรู้ น้อย " data rich but information poor) การเติบโตขึ้นอย่างรวดเร็วของข้อมูลจำนวนมากที่สะสมไว้ในฐานข้อมูลขนาดใหญ่มาก ซึ่งเกินกว่าที่กำลังคนจะสามารถจัดการได้ เป็นผลทำให้มีความจำเป็นที่ต้องมีเครื่องมือที่ช่วยในการวิเคราะห์ข้อมูลและหาความเป็นไปได้ของข้อมูลทั้งหมดที่เป็นประโยชน์ออกมา ซึ่งก็คือ Data Mining

3.4 ขั้นตอนในการทำดาต้าไมนิ่ง (Data Mining Process)

โดยทั่วไปเมื่อกล่าวถึงการทำดาต้าไมนิ่ง คนส่วนใหญ่จะนึกถึงเฉพาะในส่วนการทำไมนิ่งข้อมูลเท่านั้น แต่ความเป็นจริงการทำไมนิ่งเป็นขั้นตอนหนึ่งในกระบวนการทำดาต้าไมนิ่งทั้งหมด ในรูปที่ 3.1 แสดงขั้นตอนการทำดาต้าไมนิ่ง โดยเริ่มจากการได้มาของข้อมูล การเลือกข้อมูลที่เกี่ยวข้อง การเตรียมข้อมูลเฉพาะที่ใช้ การแปลงข้อมูลเป็นข้อมูลมาตรฐาน การทำไมนิ่งข้อมูล และการวิเคราะห์สรุปผลลัพธ์ที่ได้ เราสามารถวนกลับไปทำซ้ำใหม่ได้ในทุกขั้นตอน

แม้ว่าทุกขั้นตอนจะเรียงตามลำดับ ซึ่งในแต่ละขั้นตอนต่างๆ จะมีการทำซ้ำไปซ้ำมา แต่สามารถวนกลับไปทำขั้นตอนเดิมใหม่ (Loopback) ได้ โดยถือว่าขั้นตอนต่างเป็นอิสระต่อกัน

วัตถุประสงค์ตามแนวทางการดำเนินธุรกิจ (Business Objective) จะเป็นตัวขับเคลื่อนการทำดาต้าไมนิ่ง ซึ่งกระทำบนพื้นฐานว่าโครงการนั้นๆ จะต้องมีการทำอยู่จริงและมีเครื่องมือที่สามารถวัดผลลัพธ์ที่เกิดขึ้นได้จริง



รูปที่ 3.1 แสดงขั้นตอนการทำดาต้าไมนิ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยสรุปแล้วกระบวนการของค้ำไม่นี้ประกอบด้วย 5 ขั้นตอนหลักๆ ดังนี้

- 1) กำหนดวัตถุประสงค์ให้ชัดเจนของธุรกิจนั้นๆ (Business Objectives Determination)
- 2) การเตรียมข้อมูล (Data Preparation)
 - 2.1 การเลือกข้อมูล (Data Selection)
 - 2.2 การสุ่มเลือกข้อมูลและการทดสอบข้อมูล (Data Preprocessing)
 - 2.3 การแปลงข้อมูลไปสู่ Model สำหรับการวิเคราะห์ (Data Transformation)
- 3) การทำค้ำไม่นี้ (Data Mining)
- 4) การวิเคราะห์ผลลัพธ์ที่ได้ (Analysis of Result)
- 5) การนำมาประยุกต์ใช้กับธุรกิจ (Assimilation of knowledge)

รายละเอียดขั้นตอนการทำค้ำไม่นี้ มีดังนี้

1) การกำหนดวัตถุประสงค์ของธุรกิจ (Business Objectives Determination) เป็นขั้นตอนแรกของกระบวนการทำ Data Mining เป็นขั้นตอนที่สร้าง Object ขึ้นมาให้ได้ ซึ่งเมื่อนำเทคนิค Data Mining มาใช้ในการแก้ปัญหขององค์กร จะต้องเริ่มศึกษาว่าองค์กรนั้นทำอะไรและมีปัญหาอะไรเกิดขึ้นบ้าง ซึ่งการได้ข้อมูลมาอาจได้มาจากการสัมภาษณ์บุคคลที่ทำงานอยู่ในองค์กรนั้นๆ แล้วจึงปัญหาเหล่านั้นออกมาเป็นข้อๆ หลังจากนั้นจึงนำมาพิจารณาว่าปัญหาใดที่ไม่สามารถแก้ไขได้ด้วยเทคนิคของ Data Mining จะต้องตัดออกเพราะไม่ได้หมายความว่าทุกปัญหาสามารถแก้ไขได้ด้วยเทคนิค Data Mining โดยจุดประสงค์สำคัญของขั้นตอนนี้คือ จะต้องรู้ปัญหาและความต้องการของธุรกิจนั้น อย่างชัดเจน และถูกต้อง เพื่อกำหนดทิศทางในการทำ Data Mining และเลือกใช้ Algorithm ได้อย่างถูกต้องและเหมาะสม

2) การเตรียมข้อมูล (Data Preparation) หน้าที่ของขั้นตอนนี้คือการจัดเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสม โดยรูปแบบที่นิยมใช้กันมากคือ การทำข้อมูลให้อยู่ในรูปแบบตารางให้สามารถนำเข้าสู่กระบวนการทำ data mining ขั้นตอนการเตรียมข้อมูล แบ่งออกเป็น 3 ขั้นตอน ดังนี้

2.1 การเลือกข้อมูล (Data Selection) จุดประสงค์ คือ การจำแนกแหล่งที่มาของข้อมูล และการดึงข้อมูลที่ต้องการใช้เบื้องต้นสำหรับการวิเคราะห์ก่อนที่จะทำการค้ำไม่นี้ ซึ่งในการเลือกข้อมูลที่จะนำมาใช้นั้นขึ้นอยู่กับวัตถุประสงค์ของธุรกิจ ซึ่งกำหนดขึ้นว่าต้องการใช้ข้อมูลลักษณะใด ต้องทำความเข้าใจกับชนิดของข้อมูล ค่าที่เป็นไปได้ แหล่งกำเนิดของข้อมูล รูปแบบและลักษณะอื่นๆ ซึ่งแหล่งที่มาของข้อมูลอาจมาจากที่ต่างๆ เช่น Relational database, Transactional database , Data warehouses, Spatial databases , [www](#) เป็นต้น ซึ่งข้อมูลจะมี 2 ลักษณะ คือ

2.1.1 ข้อมูลประเภท Quantitative แบ่งเป็น

- Continuous คือข้อมูลค่าเป็นตัวเลขได้แก่ค่าที่เก็บเป็นเลขจำนวนจริง (Real Number) หรือค่าที่ต่อเนื่อง เช่น 9.2 ,12.1,1.2
- Discrete คือ ค่าที่เก็บเป็นเลขจำนวนเต็ม (Integer) เช่น ข้อมูลอายุ 20 ,32

2.1.2 ข้อมูลประเภท Categorical แบ่งเป็น

- Nominal คือข้อมูลที่ลำดับของข้อมูลไม่มีความสำคัญ เช่น เพศ ชาย-หญิง
- Ordinal คือ ข้อมูลที่ลำดับของข้อมูลมีความสำคัญ เช่น เกรด 4 3 2 1

2.2 การสุ่มเลือกข้อมูลและการทดสอบข้อมูล (Data Preprocessing) วัตถุประสงค์เพื่อให้มั่นใจว่าคุณภาพของข้อมูลที่ถูกเลือกนั้นถูกต้องและเหมาะสม ที่จะนำไปทำค้ำไ่มนึ่ง โดยใช้หลักการทางสถิติและเทคนิคการนำเสนอข้อมูลที่น่าสนใจ (Data Visualization Techniques) นอกจากนี้สิ่งสำคัญที่ต้องพิจารณาในขั้นตอนนี้คือการพิจารณาเพื่อจัดการกับข้อมูลที่เป็น Noisy Data (ค่าข้อมูลที่ขาดหายไปไม่ครบถ้วน) และ Missing Data (ค่าข้อมูลที่ผิดปกติ หรือมีความแตกต่างจากกลุ่มของข้อมูลที่น่ามาใช้)

2.2.1 Missing Data คือ ค่าข้อมูลที่ขาดหายไปไม่ครบถ้วนในชุดข้อมูลที่จะมาวิเคราะห์สาเหตุอาจเกิดจากความผิดพลาดของมนุษย์ เช่น ในกรณีที่เป็นกรอกข้อมูลของแบบสอบถาม ในบางหัวข้อที่อาจไม่มีการกรอกข้อมูลหรือไม่อาจกรอกที่ผิดพลาด ทำให้ไม่สามารถนำข้อมูลมาใช้ได้ การแก้ไขอาจตัดข้อมูลดังกล่าวทิ้งไป หรือกำหนดค่าให้กับข้อมูลที่เป็น Missing โดยค่าเฉลี่ยหรือกำหนดค่าใหม่ซึ่งไม่มีผลต่อการนำข้อมูลมาวิเคราะห์

2.2.2 Noisy Data คือ ค่าข้อมูลที่ผิดปกติ หรือมีความแตกต่างจากกลุ่มข้อมูลที่น่ามาใช้ ซึ่งในบางครั้งการแก้ไขปัญหาคือ การนำข้อมูลมา Plot Graph เพื่อตรวจสอบว่ามีข้อมูลมากน้อยแค่ไหนที่เป็น Noisy และตัดรายการดังกล่าวทิ้งไป เพื่อไม่ให้ข้อมูลดังกล่าวสร้างความบิดเบือนในการทำค้ำไ่มนึ่ง

2.3 การแปลงข้อมูลไปสู่ Model สำหรับการวิเคราะห์ (Data Transformation) เป็นการแปลงข้อมูล ให้อยู่ในรูปแบบของข้อมูลที่พร้อมที่จะนำไปวิเคราะห์ตาม algorithms ของค้ำไ่มนึ่งที่ใช้ เช่น การแปลงข้อมูลที่เป็นแบบ Categorical ให้เป็น Numeric เช่น ระดับการศึกษา อาจแปลงเป็น 1 2 และ 3

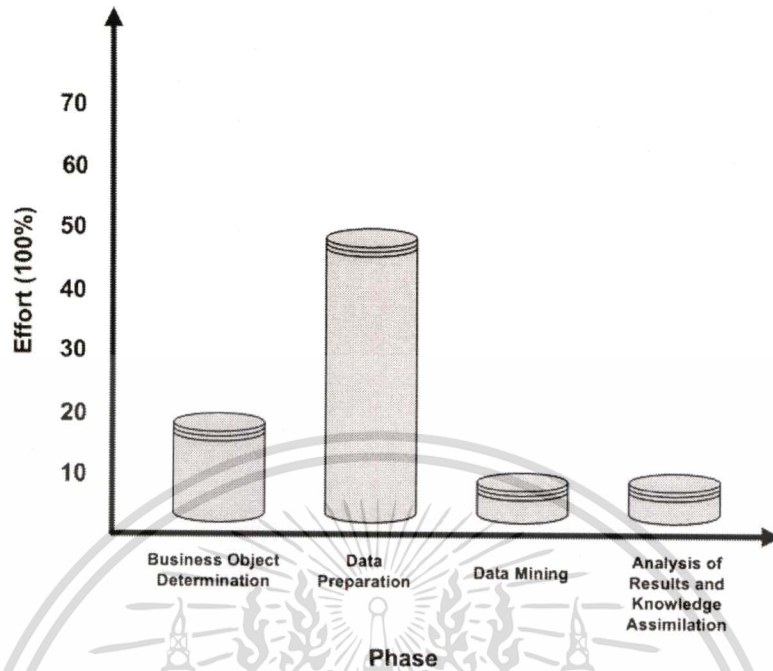
3) การทำค้ำไ่มนึ่ง (Data Mining) เป็นการนำข้อมูลผ่านกระบวนการเตรียมข้อมูล (Data Preparation) ในขั้นตอนที่ 2 แล้ว มาทำการวิเคราะห์ และประมวลผลตาม Operation ที่

กำหนดไว้ โดยมี Operation ในการทำ มี 4 รูปแบบ ได้แก่ Database Segmentation หรือ Database Clustering , Link Analysis , Deviation Detection และ Predictive Modeling

4) การวิเคราะห์ผลลัพธ์ที่ได้ (Analysis of Result) เป็นการวิเคราะห์และตีความผลลัพธ์ที่ได้จากการทำดาต้าไมนิ่ง การทำงานในขั้นตอนนี้ต้องใช้ทักษะในการวิเคราะห์ข้อมูล และการวิเคราะห์ทางธุรกิจ รวมทั้งพื้นฐานความเข้าใจของข้อมูล เนื่องจากการทำดาต้าไมนิ่งจะได้ค่าเป็นตัวเลขหรือช่วง เป็นไปตามลักษณะข้อมูลที่ใช้ในการวิเคราะห์ แต่การนำผลจากการทำดาต้าไมนิ่งมาใช้ประโยชน์หรือแปลความหมายนั้นต้องอาศัยความรู้และความเข้าใจในข้อมูลหรือธุรกิจที่จะนำไปใช้เป็นหลัก นอกจากนี้สามารถนำเอาแบบจำลองที่ได้ไปทดสอบกับข้อมูลชุดอื่น ที่ไม่ใช่ข้อมูลที่ใช้ในการสร้างแบบจำลอง เพื่อนำเอาผลลัพธ์ที่ได้มาเปรียบเทียบกับผลตามแบบจำลองว่ามีความแม่นยำและยอมรับได้หรือไม่ ถ้าผลการทดสอบไม่สามารถยอมรับได้ ต้องเพิ่มข้อมูลให้มากขึ้น เพื่อให้มีความแม่นยำในการหาความสัมพันธ์ หรือเปลี่ยนอัลกอริทึมเพื่อความเหมาะสม

5) การนำมาประยุกต์ใช้กับธุรกิจ (Assimilation of knowledge) สำหรับในขั้นตอนนี้เป็น การนำผลที่ได้จากขั้นตอนที่ 4 มารวมกับส่วนความรู้ เพื่อนำไปใช้ต่อไป เนื่องจากในการทำ Data Mining รูปแบบที่ได้หรือ Pattern ที่เรียนรู้ต้องนำมาปรับกับความรู้ที่มีอยู่ เพื่อก่อให้เกิดประโยชน์คือทำแล้วต้องได้อะไร ในขั้นตอนนี้มีหลักอยู่ 2 ประการ คือ การนำเสนอแนวคิดทางธุรกิจที่ค้นพบใหม่ และหาแนวทางที่จะใช้กฎเกณฑ์ใหม่ที่ค้นพบ เพื่อให้เกิดประโยชน์สูงสุด

โดยทั่วไปวัตถุประสงค์ดาต้าไมนิ่ง ทำให้เริ่มกระบวนการทำดาต้าไมนิ่ง ขั้นตอนเหล่านี้เป็นพื้นฐานที่สำคัญตั้งแต่เริ่มต้นจนกระทั่งได้ผลลัพธ์ ถ้าเปรียบเทียบอัตราส่วนการใช้เวลาทั้งหมดของขั้นตอนต่างๆ จะได้ว่าในขั้นตอนการตั้งวัตถุประสงค์คิดเป็น 20 เปอร์เซ็นต์ ขั้นตอนการเตรียมข้อมูลคิดเป็น 60 เปอร์เซ็นต์ ขั้นตอนการทำดาต้าไมนิ่งคิดเป็น 10 เปอร์เซ็นต์ และขั้นตอนการวิเคราะห์ผลลัพธ์คิดเป็น 10 เปอร์เซ็นต์ ดังรูปที่ 3.2 จะเห็นได้ว่าในขั้นตอนที่ต้องใช้เวลามากที่สุดคือ ขั้นตอนการเตรียมข้อมูล เนื่องจากในขั้นตอนนี้อาจจำเป็นต้องได้ข้อมูลถูกต้อง (Clean) และตรงประเด็น (Relevant) ของการทำมากที่สุด



รูปที่ 3.2 สัดส่วนของเวลาในแต่ละขั้นตอนการทำดาต้าไมนิ่ง

3.5 แหล่งที่มาของข้อมูลที่ทำดาต้าไมนิ่ง

1) รีเลชันนอล ดาต้า (Relational Database) ฐานข้อมูลที่จัดเก็บอยู่ในรูปแบบของตาราง โดยในแต่ละตารางจะประกอบไปด้วยแถวและคอลัมน์ ความสัมพันธ์ของข้อมูลทั้งหมดสามารถแสดงได้โดย entity-relationship (ER) model

ข้อมูลรีเลชันนอลสามารถเข้าถึงหรือดึงข้อมูล โดยผ่านทางภาษา Structure Query Language (SQL) ซึ่งเป็นภาษาที่ช่วยในการทำงานสะดวกขึ้น โดยใช้ภาษา SQL ทำงานเพียงแค่ระบุชื่อ ตาราง เอ็ทริบิวท์ของตาราง ฟังก์ชันพื้นฐาน และเงื่อนไขในการเข้าถึงข้อมูลหรือดึงข้อมูล ก็สามารถได้ข้อมูลมาโดยที่ผู้ใช้ไม่ต้องเข้าไปในโครงสร้างข้อมูล หรือลักษณะการเก็บข้อมูลในรีเลชันนอลดาต้าเบสนั้น

เมื่อนำรีเลชันนอล ดาต้าเบสมาประยุกต์ใช้งานดาต้าไมนิ่ง มีสิ่งที่สามารถค้นพบได้คือการค้นหาโอกาสหรือรูปแบบข้อมูลที่น่าจะเป็นไป (Searching for trends or data patterns) ตัวอย่างเช่น การทำดาต้าไมนิ่งวิเคราะห์ข้อมูลลูกค้าสำหรับทำนายความเสี่ยงที่สมาคมเป็นสมาชิกใหม่ โดยอาศัยพื้นฐานของรายได้ อายุ และข้อมูลการใช้จ่ายเงินก่อนหน้า เพื่อหารูปแบบการให้หรือไม่ให้ต่อการสมัครสมาชิกบัตรใหม่

รีเลชันนอล ดาต้าเบสเป็นหนึ่งในแหล่งข้อมูลที่เหมาะนำมาทำดาต้าไมนิ่ง เพราะมีรูปแบบที่เข้าใจ เรียนรู้ได้ง่าย ทำให้มีการใช้อย่างแพร่หลาย

นอกจากนี้ยังมีแหล่งข้อมูลอื่น ๆ อีกมากมายที่สามารถนำมาใช้ประโยชน์ด้านการค้า ไม่ว่าจะเป็นข้อมูลจากสื่อสังคมออนไลน์ ข้อมูลจากเว็บไซต์ ข้อมูลจากแอปพลิเคชัน ข้อมูลจากเซ็นเซอร์ ข้อมูลจากอุปกรณ์ IoT ข้อมูลจากดาวเทียม ข้อมูลจากหน่วยงานราชการ ข้อมูลจากภาคเอกชน ข้อมูลจากภาคประชาสังคม ข้อมูลจากภาคการศึกษา ข้อมูลจากภาคการวิจัย ข้อมูลจากภาคการแพทย์ ข้อมูลจากภาคการเกษตร ข้อมูลจากภาคการอุตสาหกรรม ข้อมูลจากภาคการบริการ ข้อมูลจากภาคการขนส่ง ข้อมูลจากภาคการพลังงาน ข้อมูลจากภาคการสิ่งแวดล้อม ข้อมูลจากภาคการสุขภาพ ข้อมูลจากภาคการกีฬา ข้อมูลจากภาคการบันเทิง ข้อมูลจากภาคการสื่อสาร ข้อมูลจากภาคการเทคโนโลยี ข้อมูลจากภาคการนวัตกรรม ข้อมูลจากภาคการวิจัยและพัฒนา ข้อมูลจากภาคการความร่วมมือระหว่างประเทศ ข้อมูลจากภาคการความร่วมมือระหว่างองค์กร ข้อมูลจากภาคการความร่วมมือระหว่างชุมชน ข้อมูลจากภาคการความร่วมมือระหว่างภาคประชาสังคม ข้อมูลจากภาคการความร่วมมือระหว่างภาคเอกชน ข้อมูลจากภาคการความร่วมมือระหว่างภาคการศึกษา ข้อมูลจากภาคการความร่วมมือระหว่างภาคการวิจัย ข้อมูลจากภาคการความร่วมมือระหว่างภาคการแพทย์ ข้อมูลจากภาคการความร่วมมือระหว่างภาคการเกษตร ข้อมูลจากภาคการความร่วมมือระหว่างภาคการอุตสาหกรรม ข้อมูลจากภาคการความร่วมมือระหว่างภาคการบริการ ข้อมูลจากภาคการความร่วมมือระหว่างภาคการขนส่ง ข้อมูลจากภาคการความร่วมมือระหว่างภาคการพลังงาน ข้อมูลจากภาคการความร่วมมือระหว่างภาคการสิ่งแวดล้อม ข้อมูลจากภาคการความร่วมมือระหว่างภาคการสุขภาพ ข้อมูลจากภาคการความร่วมมือระหว่างภาคการกีฬา ข้อมูลจากภาคการความร่วมมือระหว่างภาคการบันเทิง ข้อมูลจากภาคการความร่วมมือระหว่างภาคการสื่อสาร ข้อมูลจากภาคการความร่วมมือระหว่างภาคการเทคโนโลยี ข้อมูลจากภาคการความร่วมมือระหว่างภาคการนวัตกรรม ข้อมูลจากภาคการความร่วมมือระหว่างภาคการวิจัยและพัฒนา ข้อมูลจากภาคการความร่วมมือระหว่างภาคการความร่วมมือระหว่างประเทศ ข้อมูลจากภาคการความร่วมมือระหว่างภาคการความร่วมมือระหว่างองค์กร ข้อมูลจากภาคการความร่วมมือระหว่างภาคประชาสังคม ข้อมูลจากภาคการความร่วมมือระหว่างภาคเอกชน ข้อมูลจากภาคการความร่วมมือระหว่างภาคการศึกษา ข้อมูลจากภาคการความร่วมมือระหว่างภาคการวิจัย ข้อมูลจากภาคการความร่วมมือระหว่างภาคการแพทย์ ข้อมูลจากภาคการความร่วมมือระหว่างภาคการเกษตร ข้อมูลจากภาคการความร่วมมือระหว่างภาคการอุตสาหกรรม ข้อมูลจากภาคการความร่วมมือระหว่างภาคการบริการ ข้อมูลจากภาคการความร่วมมือระหว่างภาคการขนส่ง ข้อมูลจากภาคการความร่วมมือระหว่างภาคการพลังงาน ข้อมูลจากภาคการความร่วมมือระหว่างภาคการสิ่งแวดล้อม ข้อมูลจากภาคการความร่วมมือระหว่างภาคการสุขภาพ ข้อมูลจากภาคการความร่วมมือระหว่างภาคการกีฬา ข้อมูลจากภาคการความร่วมมือระหว่างภาคการบันเทิง ข้อมูลจากภาคการความร่วมมือระหว่างภาคการสื่อสาร ข้อมูลจากภาคการความร่วมมือระหว่างภาคการเทคโนโลยี ข้อมูลจากภาคการความร่วมมือระหว่างภาคการนวัตกรรม ข้อมูลจากภาคการความร่วมมือระหว่างภาคการวิจัยและพัฒนา

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

2) **ดาต้าแวร์เฮาส์ (Data Warehouses)** เป็นคลังข้อมูลขนาดใหญ่ที่เก็บรวบรวมข้อมูลจากหลายๆ สาขา มาไว้ที่ศูนย์กลางเดียวกัน การเก็บแต่ละสาขาจะมีลักษณะโครงสร้างเป็นหนึ่งเดียว (Unified Schema) และมีข้อมูลกระจายเก็บตามสาขาต่างๆ ดาต้าแวร์เฮาส์มีโครงสร้างที่ช่วยในการทำดาต้าคลีนนิ่ง (Data Cleaning) ดาต้าทรานฟอร์มเมชัน (Data Transformation) ดาต้าอินทิเกรชัน (Data Integration) ดาต้าโหลดคิง (Data Loading) และมีการปรับปรุงข้อมูลตามช่วงเวลา (Periodic Data Refreshing)

ดาต้าแวร์เฮาส์เหมาะสำหรับการตัดสินใจ (Decision Making) ข้อมูลที่อยู่ในดาต้าแวร์เฮาส์มีการจัดเก็บตามหัวข้อที่สำคัญ เช่น ข้อมูลลูกค้า สินค้า ผู้ผลิต และกิจกรรมข้อมูลดังกล่าวจะอยู่ในรูปข้อมูลสารสนเทศที่ผ่านมาแล้ว (Historical Perspective) ระยะเวลาประมาณ 5 – 10 ปี และอยู่ในรูปของผลรวม

รูปแบบของดาต้าแวร์เฮาส์มีโครงสร้างหลายมิติ (Multidimensional Database) ในแต่ละมิติจะมีการเก็บข้อมูลในชุดแอตทริบิวต์ (Set of Attributes) ในสกรีม่า (Schema) และแต่ละค่าที่เก็บจะมีการรวมค่าบางอย่าง เช่น การนับ (Count) หรือยอดขายรวมเงิน (Sales_amount) และลักษณะทางกายภาพของดาต้าแวร์เฮาส์เป็นข้อมูลรีเลชันนอลเป็นชั้นๆ หรือมีลักษณะเป็นลูกบาศก์หลายมิติ ทำให้ผู้ใช้สามารถสร้างได้หลายมุมมองข้อมูล

3) **ทรานแซกชันนอล ดาต้าเบส (Transactional Database)** ประกอบด้วยแฟ้ม (File) โดยที่แต่ละเรคคอร์ด (Record) แทนความหมายของแต่ละทรานแซกชัน รูปแบบทั่วไปของทรานแซกชันประกอบด้วยเลขประจำทรานแซกชัน (Transaction_ID) และรหัสสินค้า โดยในทรานแซกชันนอล ดาต้าเบส ยังมีข้อมูลอื่นๆ อีกเช่น วันที่ขายสินค้า รหัสลูกค้า รหัสพนักงานขาย และสาขาที่ทำการขาย

4) **Advanced Database** เป็นฐานข้อมูลที่จัดเก็บในรูปแบบอื่น ๆ เช่น ข้อมูลแบบ object-oriented , ข้อมูลที่เป็น text file , ข้อมูลมัลติมีเดีย , ข้อมูลในรูปแบบของ web

3.6 การพิจารณาเลือกใช้วิธีดาต้าไมนิ่ง

รูปแบบความเป็นไปได้ของการทำดาต้าไมนิ่งขึ้นกับวัตถุประสงค์ของการทำ และลักษณะการทำธุรกิจขององค์กร (Business Objective Determination)

สามารถแบ่งตามการจัดการเป็นด้านต่างๆ ได้ดังนี้

1. การจัดการด้านการตลาด (Market Management)
2. การจัดการด้านความเสี่ยง (Risk Management)
3. การจัดการด้านการตรวจสอบการทำทุจริต (Fraud Management)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และสามารถแบ่งอีกแนวทางหนึ่งคือ

1. ด้านแอปพลิเคชัน (Application)
2. ด้านการดำเนินการ (Operation)
3. ด้านเทคนิค (Techniques)

ตารางที่ 3.1 คาด้าไมนิ่งที่มีรูปแบบตามแอปพลิเคชัน การดำเนินงาน และเทคนิคที่ใช้

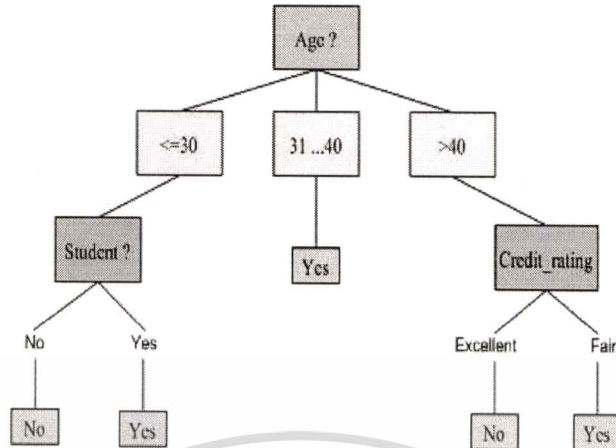
Application	Market Management	Risk Management		Fraud Management
	- Target Marketing - Customer Relationship Management - Market Basket Analysis - Cross Selling - Market Segmentation	- Forecasting - Customer Relation - Improved Underwriting - Quality Control - Competitive Analysis		Fraud Detection
Operation	Predictive Modeling	Database Segmentation	Link Analysis	Deviation Detection
Techniques	Classification	Demographic Clustering	Association Discovery	Visualization
	Value Prediction	Neural Clustering	Sequential Pattern Discovery Similar Time Sequence discovery	Statistic

สามารถแสดงรายละเอียดได้ดังนี้

1) ด้านแอปพลิเคชัน (Application) ในด้านนี้จะเน้นแอปพลิเคชันทางธุรกิจ (Business Application) ดังแสดงในตารางที่ 3.1 เช่น เป้าหมายทางการตลาด (Target Marketing) ลูกค้าสัมพันธ์ (Customer Relation Management) ทำนายยอดขาย (Forecasting) และการตรวจสอบการทำทุจริต (Fraud Detection) ทั้งนี้ขึ้นกับวัตถุประสงค์การทำคาด้าไมนิ่งของบริษัทหรือองค์กร

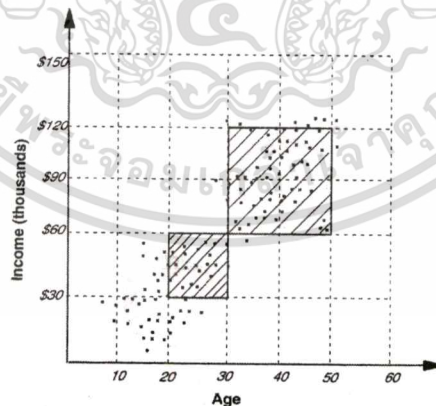
2) ด้านการดำเนินการ (Operation) จะมี 4 ด้านหลักๆ ดังนี้

2.1 Predictive Modeling ในด้านนี้ทำงานคล้ายกับการเรียนรู้จากประสบการณ์ของมนุษย์ แล้วนำมาสร้างรูปแบบเพื่อตอบคำถามสำหรับข้อมูลที่เข้ามาใหม่ รูปแบบที่นิยมใช้คือ IF – THEN – ELSE หรือสามารถเขียนในรูปแบบทรี (Tree) เพื่อง่ายแก่การทำความเข้าใจ ดังรูปที่ 3.3 ตัวอย่างการสร้างรูปแบบการทำนาย



รูปที่ 3.3 รูปแบบการทำนาย (Predictive Modeling)

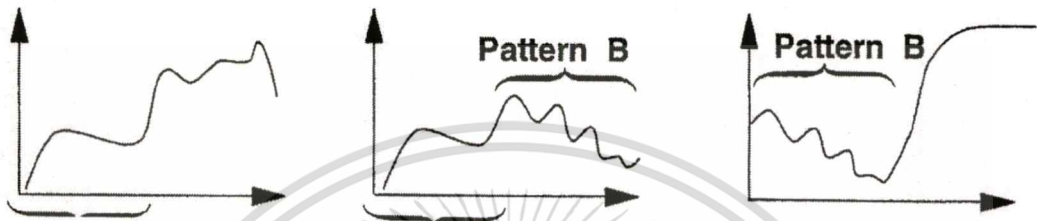
2.2 Database Segmentation ในด้านนี้จะเน้นการแบ่งกลุ่มหรือจัดกลุ่มของข้อมูลที่มีความเหมือนหรือคล้ายกันหรือมีค่าใกล้เคียงกัน ให้เป็นข้อมูลกลุ่มเดียวกัน บางครั้งมีการใช้คำว่า Segmentation กับ Clustering ซึ่งคำทั้งสองมีความหมายเดียวกัน สามารถใช้แทนกันได้ ถ้าเน้นเฉพาะคำว่า Segmentation หมายถึง การดำเนินการทางดาต้าไมนิ่ง (Data Mining Operation) ส่วนคำว่า Segmentation และ Clustering หมายถึง ผลลัพธ์จากการดำเนินการทางดาต้าไมนิ่ง ซึ่งเป็นกลุ่มต่างๆ ในรูปที่ 3.4 แสดงการจัดกลุ่มของข้อมูลประชากร โดยแกน X แสดงอายุ และแกน Y แสดงรายได้



รูปที่ 3.4 แสดงการจัดกลุ่มข้อมูล (Database Segmentation)

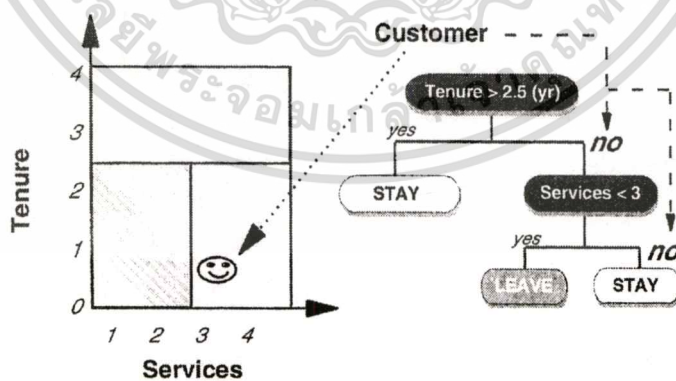
2.3 Link Analysis ในด้านนี้ทำงานตรงกันข้ามกับ 2 รูปแบบที่ผ่านมาคือ Predictive Modeling และ Database Segmentation ซึ่งในด้านนี้จะเน้นการศึกษาลักษณะ คุณสมบัติ เนื้อหา ข้อมูลใน Database ทั้งหมด เพื่อหาความสัมพันธ์ของข้อมูลแต่ละ Record หรือ Tuple หรือกลุ่มของ เอกสารนี้เป็นการหาความสัมพันธ์ที่ตรงกันหรือคล้ายกันระหว่างเอกสารที่มีเนื้อหาใกล้เคียงกัน ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Record ความสัมพันธ์นี้เรียกว่า Associations นิยมที่จะนำ Link Analysis ค้นหาความสัมพันธ์ของสินค้าและบริการ เพื่อนำความรู้ที่ได้ไปวางแผนการตลาด ในรูปที่ 3.5 แสดงแบบแผนที่ค้นพบได้ ข้อมูลเป็นแบบ A และแบบ B ซึ่งข้อมูลที่ได้มาจากประวัติการขายสินค้าเฉพาะ หรือการเคลื่อนไหวเปลี่ยนแปลงราคาหุ้น



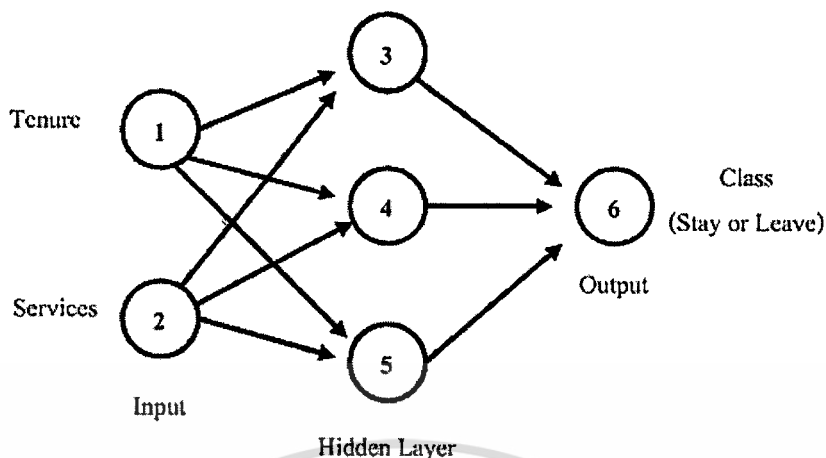
รูปที่ 3.5 แสดงแบบแผนของข้อมูล (Pattern Matching)

3. ด้านเทคนิค (Techniques) ด้านนี้จะมีการใช้งานจริงตามด้านการดำเนินการ (Operation) และมีการระบุ Algorithms ที่ใช้งานความสัมพันธ์ของด้านเทคนิคกับด้านการดำเนินการมีความสัมพันธ์ที่ใกล้ชิดกันมากกว่าด้านแอปพลิเคชัน เพราะเมื่อมีการใช้งานหรือกระทำจริงจะมีการใช้ Algorithms ต่างๆตามด้านการดำเนินการ เพราะให้ได้ผลลัพธ์ตามวัตถุประสงค์มีรูปแบบทั่วไปที่นิยมคือ แบบตัดสินใจ (Decision Tree) ดังรูปที่ 3.6 และนิวรอลเน็ตเวิร์ค (Neural Network) ดังรูปที่ 3.7



รูปที่ 3.6 แสดงแบบตัดสินใจ (Decision Tree)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.7 แสดงนิวรอลเน็ตเวิร์ค (Neural Network)

3.7 การประยุกต์ใช้ค้ำไม้หนึ่ง

ธุรกิจต่างๆ มีการนำเอาวิธีการของค้ำไม้หนึ่งมาประยุกต์ใช้ดังนี้

- 1) ธุรกิจปลีกใช้ค้ำไม้หนึ่งในการพิจารณาหากลยุทธ์ให้เป็นที่สนใจกับผู้บริโภคในรูปแบบต่างๆ เช่น ที่ว่างในชั้นวางของจัดการอย่างไรจึงจะเพิ่มยอดขายได้
- 2) กิจการโทรคมนาคม ใช้ตรวจสอบทุจริต โดยวิเคราะห์รูปแบบการใช้งานของสมาชิกลูกค้าในการใช้งานโทรศัพท์ และคาดการณ์ข้อบกพร่องที่เป็นไปได้ในการชำระเงินเทคนิคนี้ยังใช้ได้กับลูกค้าโทรศัพท์เคลื่อนที่ในการตรวจสอบพื้นที่ที่เสี่ยงต่ออาชญาเสี่ยลูกค้า โดยค้นหาวิธีรวมกลุ่มผู้ใช้ให้เป็นหนึ่งเดียว ด้วยการสร้างแรงดึงดูดในเรื่องค่าใช้จ่าย และพัฒนาเรื่องความรักภักดีต่อสินค้า
- 3) การวิเคราะห์ผลิตภัณฑ์ เก็บรวบรวมลักษณะและราคาของผลิตภัณฑ์ทั้งหมด สร้างโมเดลเพื่อทำนายราคาผลิตภัณฑ์ตัวอื่นๆ
- 4) การวิเคราะห์บัตรเครดิต
 - ช่วยในการตัดสินใจให้เครดิตการ์ดกับลูกค้า
 - ประเภทลูกค้าที่เสี่ยงในเรื่องเครดิต ต่ำ ปานกลาง หรือสูง
 - ป้องกันปัญหาเรื่องทุจริตบัตรเครดิต
- 5) การวิเคราะห์ลูกค้า
 - แบ่งกลุ่มและวิเคราะห์ลูกค้า เพื่อผลิตและเสนอสินค้าได้ตรงตามกลุ่มเป้าหมายแต่ละกลุ่ม
 - ทำนายลูกค้าคนใดจะเลิกใช้บริการจากบริษัทภายใน 6 เดือนหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 6) การวิเคราะห์การขาย
 - หาความสัมพันธ์ของการซื้อสินค้าร่วมกัน
 - ช่วยในการโฆษณาสินค้าได้อย่างเหมาะสมและตรงตามเป้าหมาย
 - ช่วยในการจัดวางสินค้าได้อย่างเหมาะสม
- 7) Text Mining เป็นเครื่องมือตรวจระดับความพึงพอใจของผู้ที่เข้าชมนิทรรศการ โดยผ่านการประมวลผลจากการสอบถาม
- 8) E-Commerce
 - ช่วยให้เข้าใจพฤติกรรมลูกค้า เช่น ลูกค้าเข้าไป Web ใดตามลำดับก่อนหลัง
 - ช่วยในการปรับปรุง Web Site เช่น พิจารณาส່วนใดของ Web ที่ควรปรับปรุง หรือ ควรเรียงลำดับการเชื่อมโยงในแต่ละหน้าอย่างไร เพื่อให้สะดวกกับผู้ใช้ชม

3.8 Algorithm ที่ใช้ในการศึกษา

ในการศึกษาการทำนายลูกค้าที่มีแนวโน้มจะยกเลิกการใช้โทรศัพท์พื้นฐานกำหนดขั้นตอนการทำงานไว้ 2 ขั้นตอนคือ

- 1) จัดกลุ่มลูกค้าโดยใช้ Database Segmentation ซึ่งใช้ K-Means Algorithm มีหลักการทำงานดังนี้
 - แบ่งกลุ่มข้อมูลตามที่ต้องการเป็น k Clusters
 - คำนวณจุดศูนย์กลางแต่ละกลุ่ม
 - ย้ายกลุ่มให้เข้าใกล้จุดศูนย์กลางให้มากที่สุด
 - การย้ายกลุ่มทำให้จุดศูนย์กลางเปลี่ยนไป ดังนั้นจำเป็นต้องคำนวณหาจุดศูนย์กลางใหม่อีกครั้ง
 - ทำขั้นตอนการคำนวณจุดศูนย์กลางใหม่ และทำงานกว่าจะครบเงื่อนไขที่จะหยุดได้ คือ จุดศูนย์กลางไม่เปลี่ยน

ตัวอย่างการใช้ K-Means Algorithm ใช้ข้อมูลจากตารางที่ 3.2 แสดงดังรูปกราฟที่ 3.8

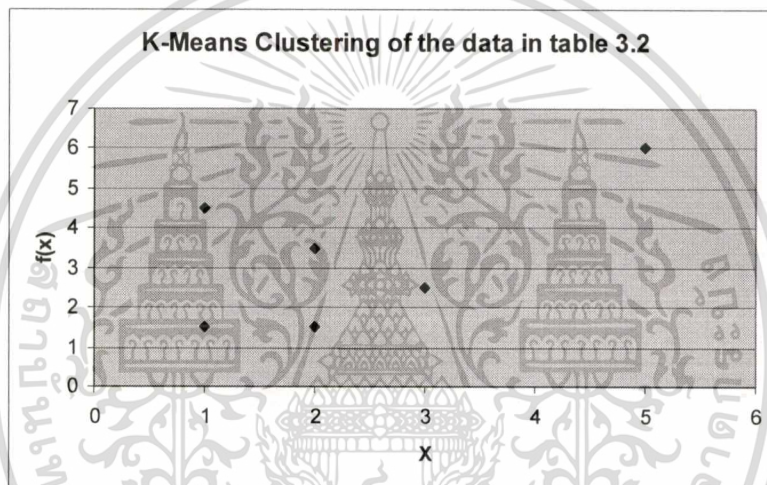
ตารางที่ 3.2 K-Means Input Values

Instance	X	Y
1	1.0	1.5
2	1.0	4.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 K-Means Input Values (ต่อ)

Instance	X	Y
3	2.0	1.5
4	2.0	3.5
5	3.0	2.5
6	5.0	6.0



รูปที่ 3.8 กลุ่มของข้อมูลจากตารางที่ 3.2

ต้องการแบ่งข้อมูล n ตัวเป็น K กลุ่ม

K

? $n_i = n$

$i = 1$

ใช้สูตรคำนวณดังนี้

- หาค่า Mean Vector หรือ Center of Cluster

$$m^{(K)} = \frac{1}{n_k} \sum_{i=1}^{n_k} X_i^{(K)}$$

- Square error เป็นผลรวมของ Euclidean distance

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$e^2 K = \sum_{i=1}^{n_k} (x_i - m^{(K)})^T + (x_i^{(K)} - m^{(K)})$$

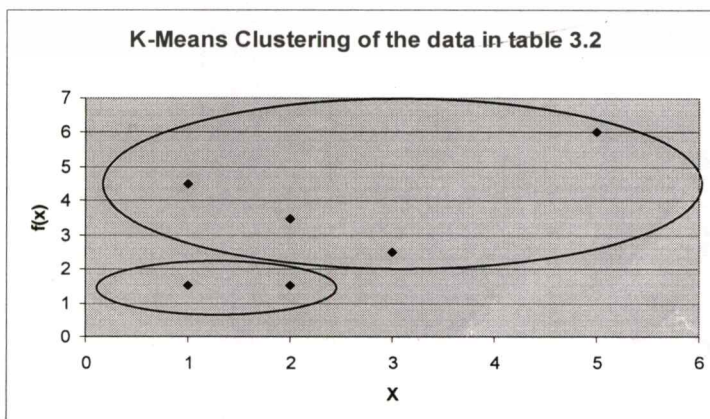
ขั้นตอนการทำงานของ K-Means Algorithm มีดังนี้

- 1) เลือก Instance ที่ 1 และ 2 เป็น Initial Cluster
- 2) หาจุดศูนย์กลาง Mean Vector และค่า $e^2 K$ หากให้ค่า $e^2 K$ น้อยข้อมูลย้ายเข้าไปใกล้จุด Mean ดังนั้นข้อมูลนั้นจึงเป็นสมาชิกของกลุ่ม
- 3) คำนวณหาค่า Mean Vector ใหม่ และค่า $e^2 K$ ใหม่
- 4) เปรียบเทียบขั้นตอนที่ 3 กับรอบก่อนหน้า หากค่า $e^2 K$ เท่ากันหยุดทำต่อไป หากไม่เท่ากันให้เริ่มทำขั้นตอน 2 ใหม่

ผลลัพธ์จากการทำ K-Means Cluster แสดงดังตารางที่ 3.3 และรูปกราฟที่ 3.9

ตารางที่ 3.3 Several Applications of the K-Means Algorithm (K=2)

Outcome	Cluster Centers	Cluster Points	Squared Error
1	(2.67, 4.67)	2, 4, 6	14.5
	(2.00, 1.83)	1, 3, 5	
2	(1.5, 1.5)	1, 3	15.94
	(2.75, 4.125)	2, 4, 5, 6	
3	(1.8, 2.7)	1, 2, 3, 4, 5	9.6
	(5, 6)	6	



รูปที่ 3.9 การจัดกลุ่มข้อมูลของตารางที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

K-Means Algorithm เป็นเทคนิคที่นิยมใช้กันมาก เนื่องจากเป็นเทคนิคที่ทำให้เข้าใจง่าย แต่ไม่สามารถใช้กับข้อมูลที่เป็นตัวเลข การแปลงข้อมูลอาจทำให้ความหมายผิดเพี้ยนไป จึงมีผู้พัฒนา K-Modes มาใช้กับข้อมูลที่เป็นตัวเลข และต่อมาจึงมีการพัฒนาเป็น K-prototype ที่สามารถใช้ได้กับข้อมูลที่เป็นตัวเลข และตัวอักษรทำให้สามารถทำงานได้อย่างเหมาะสมมากขึ้น

2) หา Pattern ลักษณะของแต่ละกลุ่ม จากการแบ่งกลุ่ม โดยใช้ Predictive Modeling ในการศึกษาจะใช้ CHAID Algorithm เป็น Model ที่มีความนิยมในการนำมาใช้งาน ประกอบกับข้อมูลที่นำมาใช้พยากรณ์ส่วนใหญ่เป็นข้อมูลตัวเลข ดังนั้นในการหาความสัมพันธ์ CHAID (Chi Square Automatic Interaction Detection) สามารถใช้ได้ทั้งข้อมูลที่เป็นตัวเลขปกติ ข้อมูลที่เป็นตัวเลขต่อเนื่อง โดย CHAID จะทำการแยกชุดข้อมูลพิจารณาความสัมพันธ์ที่ได้จากการคำนวณ แบ่งข้อมูลจากชุดข้อมูลรวมเป็นชั้นเซตย่อย ระดับ Child Node ตามจำนวนที่กำหนด ทำงานกว่าตัวแปรไม่มีความสัมพันธ์กับชุดข้อมูลที่เป็น Target แล้วจะยุติ

องค์ประกอบของการใช้ CHAID ในการวิเคราะห์คือ

1. ตัวแปรที่ใช้ในการพยากรณ์
2. ตัวแปรที่เป็น Target
3. การกำหนดค่า CHAID Parameters ที่จะใช้ในการวิเคราะห์ชุดข้อมูล

การทดสอบไคสแควร์ (Chi-Square Test : (χ^2) Test) เป็นการศึกษาความเป็นอิสระของตัวแปร การทำงานของ CHAID จะใช้การคำนวณของ Chi-Square (χ^2) Degree of freedom (d.f.) และ p-value มีสูตรการคำนวณดังนี้

สมมติตารางแจกแจงความถี่ อธิบายวิธีการคำนวณของ Chi-Square

ตารางที่ 3.4 ตาราง Contingency ของ (χ^2)

Row Variable	Column Variable					Total
	A_1	A_2	A_3	A_c	
B_1	O_{11}	O_{12}	O_{13}	O_{1c}	R_1
B_2	O_{21}	O_{22}	O_{23}	O_{2c}	R_2
.
.
B_r	O_{r1}	O_{r2}	O_{r3}	O_{rc}	R_r
	C_1	C_2	C_3	C_c	N

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรณีฉุกเฉินและกรณีการดำเนินงานที่เกี่ยวข้องกับการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \dots\dots\dots(1)$$

O_{ij} = ค่าของข้อมูลที่เก็บมาได้

E_{ij} = ค่าของเหตุการณ์ที่น่าจะเกิดขึ้น

ความถี่ที่คาดว่าจะได้ หากจากการคูณความน่าจะเป็นแต่ละค่าด้วยผลรวมของความถี่ทั้งหมด N คือ

$$E_{ij} = N \cdot \Pr(A_i | B_j) \dots\dots\dots(2)$$

เมื่อ R และ C เป็นผลรวมของความถี่ในแถวและสดมภ์ที่ต้องการ N คือ ผลรวมของความถี่ทั้งหมด

$$v = (R-1)(C-1) \dots\dots\dots(3)$$

หลังจากนั้นนำค่า Chi-Square (χ^2) และ degree of freedom (v) มาใช้ในการคำนวณหาค่า p-value ต่อไป

ตัวอย่างการหาค่าความสัมพันธ์โดยใช้ CHAID Algorithms

จากข้อมูลที่มีค่าตัวแปร 3 ตัวแปรคือ X_1 , X_2 , Y โดยมีค่า Y เป็น Target CHAID จะพิจารณาค่าความสัมพันธ์ดังนี้

ตัวแปร Y จะมีค่า 4 ประเภทคือ 1,2,3,4

ตัวแปร X_1 จะมีค่า 4 ประเภทคือ 1,2,3,4

ตัวแปร X_2 จะมีค่า 4 ประเภทคือ 0,1,2

ขั้นตอนการทำงานของ CHAID

1. คำนวณการกระจายของตัวแปร Y ใน Root Node

ตารางที่ 3.5 แสดงการกระจายของข้อมูลตัวแปร Y ใน Root Node

Cat	%	N
1	35.00	35
2	8.00	8
3	35.00	35
4	22.00	22
Total	(100.00)	100

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ทุกค่าของตัวแปร X จะหาคู่ของกลุ่มข้อมูล X โดยการหาค่าเปรียบเทียบจากความแตกต่างที่น้อยที่สุดทีละคู่ จากคู่ที่มีค่า p -value มากที่สุดที่เกี่ยวข้องกับการกระจายของ Y ในแต่ละ Node ความสัมพันธ์ระหว่างค่า X_i และ ค่า Y ภายใน Node แสดงรายละเอียดตามตารางด้านล่าง

ตารางที่ 3.6 แสดงความสัมพันธ์ระหว่าง Y และ X_i

X_i/Y	1	2	3	4	RowToll
1	23	5	19	4	51
2	12	2	15	13	42
3	0	1	0	1	2
4	0	0	1	4	5
ColToll	35	8	35	22	100

$$(\chi^2) = 25.63559 \text{ d.f.} = 9 \text{ (} p = 0.00234955 \text{)}$$

2.1 จากตารางที่ 3.6 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_i สามารถคำนวณหาค่า (χ^2) d.f. และ p -value ตามลำดับ ได้ผลลัพธ์ตามตารางที่ 3.7 – 3.12

ตารางที่ 3.7 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_i กรณี 1 และ 2

X_i/Y	1	2	3	4	RowToll
1	23	5	19	4	51
2	12	2	15	13	42
ColToll	35	7	34	17	93

$$(\chi^2) = 9.193281 \text{ d.f.} = 3 \text{ (} p = 0.02682849 \text{)}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.8 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_1 กรณี 1 และ 3

X_1/Y	1	2	3	4	RowToll
1	23	5	19	4	51
3	0	1	0	1	2
ColToll	23	6	19	5	53

$$(\chi^2) = 8.019281 \text{ d.f.} = 3 \text{ (} p = 0.0456149 \text{)}$$

ตารางที่ 3.9 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_1 กรณี 1 และ 4

X_1/Y	1	2	3	4	RowToll
1	23	5	19	4	51
4	0	0	1	4	5
ColToll	23	5	20	8	56

$$(\chi^2) = 19.72078 \text{ d.f.} = 3 \text{ (} p = 0.0001939266 \text{)}$$

ตารางที่ 3.10 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_1 กรณี 2 และ 3

X_1/Y	1	2	3	4	RowToll
2	12	2	15	13	42
3	0	1	0	1	2
ColToll	12	3	15	14	44

$$(\chi^2) = 7.23356 \text{ d.f.} = 3 \text{ (} p = 0.0648145 \text{)}$$

ตารางที่ 3.11 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_1 กรณี 2 และ 4

X_1/Y	1	2	3	4	RowToll
2	12	2	15	13	42
4	0	0	1	4	5
ColToll	12	2	16	17	47

$$(\chi^2) = 4.962482 \text{ d.f.} = 3 \text{ (} p = 0.1745651 \text{)}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.12 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_i กรณี 3 และ 4

X_i/Y	1	2	3	4	RowToll
3	0	1	0	1	2
4	0	0	1	4	5
ColToll	0	1	1	5	7

$$(\chi^2) = 3.08 \text{ d.f.} = 2 \text{ (} p = 0.2143811 \text{)}$$

2.2 คำนวณหาค่า p-value ของความสัมพันธ์ในกลุ่มของตัวแปร เพื่อนำมาเปรียบเทียบกับ alpha level ? merge (0.05 คือค่าที่กำหนดไว้) พบว่า ค่า p-value ในกรณี 3 และ 4 มีค่าเท่ากับ 0.2143811 ซึ่งมากกว่า ? merge 0.05 จึงสามารถยุบรวมกรณี 3 และ 4 รวมกันได้ เช่นเดียวกับกรณีอื่นๆ

ตารางที่ 3.13 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_i กรณี 1 และ 3,4

X_i/Y	1	2	3	4	RowToll
1	23	5	19	4	51
3,4	0	1	1	5	7
ColToll	23	6	20	9	58

$$(\chi^2) = 20.25577 \text{ d.f.} = 3 \text{ (} p = 0.0001502343 \text{)}$$

ตารางที่ 3.14 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_i กรณี 2 และ 3,4

X_i/Y	1	2	3	4	RowToll
2	12	2	15	13	42
3,4	0	1	1	5	7
ColToll	12	3	16	18	49

$$(\chi^2) = 6.408565 \text{ d.f.} = 3 \text{ (} p = 0.09333908 \text{)}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.15 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_1 กรณี 1 และ 2

X_1/Y	1	2	3	4	RowToll
1	23	5	19	4	51
2	12	2	15	13	42
ColToll	35	7	34	17	93

$$(\chi^2) = 9.193281 \text{ d.f.} = 3 \text{ (} p = 0.002682849 \text{)}$$

2.3 กระบวนการทำซ้ำเหมือนขั้นตอนที่ 2.2 พบว่า กรณี 2 , 3 และ 4 มีค่า $p = 0.09333908$ ซึ่งมากกว่าค่า $\alpha = 0.05$ จึงสามารถยุบรวมกรณี 2 , 3 และ 4 รวมกันได้ ได้ผลลัพธ์ตามตารางที่ 3.13 – 3.16

ตารางที่ 3.16 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_1 กรณี 1 และ 2,3,4

X_1/Y	1	2	3	4	RowToll
1	23	5	19	4	51
2,3,4	12	3	16	18	49
ColToll	35	8	35	22	100

$$(\chi^2) = 13.08861 \text{ d.f.} = 3 \text{ (} p = 0.004448843 \text{)}$$

2.4 ปรับค่า p-value สำหรับซัพเซตที่ได้ยุบรวมกัน โดยใช้ Bonferroni multiplier โดยมีสูตรคำนวณดังนี้

$$B_{free} = \sum_{i=0}^{r-1} \frac{(-1)^i (r-i)^c}{r! (r-i)!} \dots \dots \dots (4)$$

$c =$ จำนวนกลุ่มข้อมูลของ X_1

$r =$ จำนวนของกลุ่มที่ยุบรวมกัน

ค่าของ (χ^2) p-value ที่ได้จากการคำนวณข้างต้นจะถูกปรับ โดยใช้ Bonferroni multiplier โดย X_1 เป็นข้อมูลลักษณะ Nominal จากสูตรปรับค่า p-value เท่ากับ 0.0311 (0.004448843×7)

3. ขั้นตอนที่ 2.1 – 2.4 จะถูกทำซ้ำ โดยเปลี่ยนค่าจาก X_1 เป็น X_2 เปรียบเทียบกับค่า Y จำนวนได้ดังตารางที่ 3.17 – 3.20 นี้คือ

ตารางที่ 3.17 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_2

X_2/Y	1	2	3	4	RowToll
0	33	8	34	22	97
1	1	0	1	0	2
2	1	0	0	0	1
ColToll	35	8	35	22	100

$$(\chi^2) = 2.768778 \text{ d.f.} = 6 \text{ (} p = 0.8372577 \text{)}$$

3.1 X_2 ประกอบด้วยข้อมูล 3 กลุ่ม คือ

ตารางที่ 3.18 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_2 กรณี 0 และ 1

X_2/Y	1	2	3	4	RowToll
0	33	8	34	22	97
1	1	0	1	0	2
ColToll	34	8	35	22	99

$$(\chi^2) = 0.8881097 \text{ d.f.} = 3 \text{ (} p = 0.8282962 \text{)}$$

ตารางที่ 3.19 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_2 กรณี 0 และ 2

X_2/Y	1	2	3	4	RowToll
0	33	8	34	22	97
2	1	0	0	0	1
ColToll	34	8	34	22	98

$$(\chi^2) = 1.901759 \text{ d.f.} = 3 \text{ (} p = 0.5930452 \text{)}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.20 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_2 กรณี 1 และ 2

X_2/Y	1	2	3	4	RowToll
1	1	0	1	0	2
2	1	0	0	0	1
ColToll	2	0	1	0	3

$$(\chi^2) = 0.75 \text{ d.f.} = 1 \text{ (p} = 0.3864762)$$

3.2 ค่าของ p-value ที่มากที่สุดคือ 0.8283 สามารถยุบรวมกลุ่ม 0 และ 1 ได้ และยุติการยุบรวมเนื่องจากสามารถแยกกลุ่มข้อมูลได้อย่างชัดเจน

3.3 ปรับค่า p-value ของ X_2 (0.5985587×3) ดังตารางที่ 3.21

ตารางที่ 3.21 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X_1 กรณี 0,1 และ 2

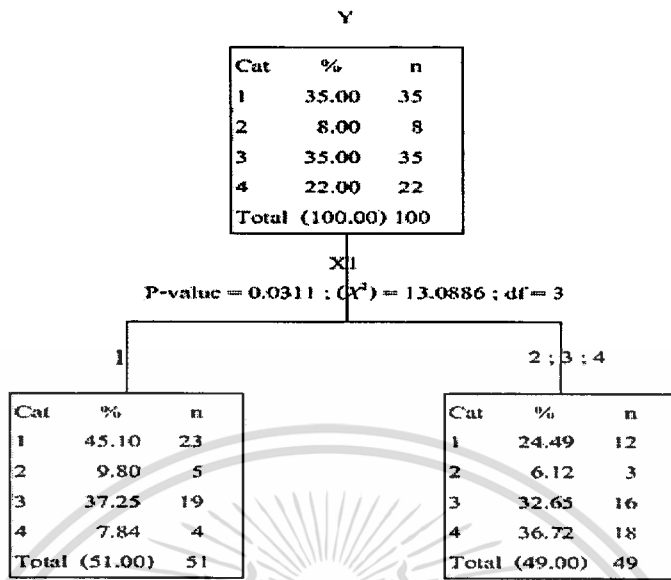
X_1/Y	1	2	3	4	RowToll
0,1	34	8	35	22	99
2	1	0	0	0	1
ColToll	35	8	35	22	100

$$(\chi^2) = 1.875902 \text{ d.f.} = 3 \text{ (p} = 0.5985587)$$

4. ในขั้นตอนสุดท้ายที่จะแตก Node ทำให้สรุปได้ว่าจะแตก Root Node ได้ 2 Subnode โดยกลุ่มข้อมูลของ X_1 คือ หนึ่ง Subnode กรณี $X_1 = 1$ จะมีได้ 35 กรณี ส่วนที่เหลืออีก 65 กรณีเป็น $X_1 = 2, 3$ หรือ 4

5. การสร้าง Trees จะดำเนินต่อไปจนกระทั่งหมด Criteria ที่กำหนดหรือหยุดเมื่อได้ Trees ที่พอใจ ตามขั้นตอนที่กล่าวมาข้างต้นแสดงตามรูปที่ 3.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.10 แสดง Tree ของข้อมูลจากการใช้ CHAID Algorithm

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

วิธีการดำเนินการศึกษา

4.1 ข้อมูลที่ศึกษา

ข้อมูลการใช้บริการโทรศัพท์ แบ่งออกเป็น 2 ส่วนคือ ข้อมูลการใช้บริการโทรศัพท์ในเขตนครหลวง (กรุงเทพมหานครและปริมณฑล) และข้อมูลการใช้บริการโทรศัพท์ในเขตโทรศัพท์ภูมิภาค ในการศึกษาเรื่องการประยุกต์ Data Mining เพื่อทำนายลูกค้าที่มีแนวโน้มจะยกเลิกการใช้บริการโทรศัพท์พื้นฐาน โดยใช้ Predictive Modeling นั้น จะเลือกข้อมูลเลขหมายโทรศัพท์ประจำที่ในเขตโทรศัพท์นครหลวง จากฐานข้อมูลต่างๆ ดังนี้

1. ข้อมูลพื้นฐานลูกค้า (Customer information data) ข้อมูลประวัติลูกค้าตามสัญญาที่ทำไว้กับบริษัทฯ ประกอบด้วย เลขหมาย วันที่ติดตั้ง สถานะ จำนวนเดือนที่ไม่ได้จ่ายเงิน ชื่อลูกค้า ที่อยู่ประเภทลูกค้า ชนิดอัตราค่าบริการ จำนวนครั้งในการใช้ฟรี วิธีการชำระเงิน เงินประกันการใช้โทรศัพท์ และประเภทการให้บริการ เป็นต้น แสดงตามตารางที่ 4.1
2. ข้อมูลการใช้โทรศัพท์ (Call data) ข้อมูลที่แสดงพฤติกรรมการใช้โทรศัพท์ ประกอบด้วย เลขหมาย ชื่อชุมสาย วันที่ติดตั้ง สถานะของเลขหมาย ประเภทลูกค้า จำนวนมิเตอร์ในการใช้ และค่าทางไกล เป็นต้น แสดงตามตารางที่ 4.2
3. ข้อมูลการชำระเงิน (Billing and payment data) ข้อมูลเกี่ยวกับค่าบริการ และการชำระค่าบริการ ประกอบด้วย เลขหมาย วันที่ชำระเงิน ค่าใช้บริการประจำเดือน สถานที่ชำระค่าบริการจำนวนค่าใช้บริการ และจำนวนเงินที่ชำระ เป็นต้น แสดงตามตารางที่ 4.3

ข้อมูลข้างต้นที่เลือกมาใช้ประกอบการศึกษาเป็นข้อมูลการใช้บริการโทรศัพท์ตั้งแต่เดือน ก.ค. 46 – มี.ค. 47

ตารางที่ 4.1 แสดงตารางข้อมูลพื้นฐานลูกค้า

Field	Field Name	Type	Description
1	AREA_CODE	Character	รหัสทางไกล
2	TELNO	Character	เลขหมายโทรศัพท์
3	IN_SERVICE	Character	รหัส ทศท. วันที่ติดตั้ง (ปีเดือนวัน)
4	STAT_OF_AC	Character	สถานะของเลขหมาย
			“ ” = เปิดใช้งาน
			F = รีดอน
5	NO_OF_FBIL	Character	จำนวนเดือนที่ถอนเลิกใช้
6	OWNER_NAME	Character	ชื่อคู่สัญญา
7	OWNER_ADDR	Character	ที่อยู่ที่ตั้งใบแจ้งหนี้ทางไปรษณีย์
8	CITY	Character	ชื่อจังหวัด
9	ZIP_CODE	Character	รหัสไปรษณีย์
10	BANK_IND	Character	สถานะการหักบัญชีธนาคาร
			“ ” = จ่ายเงินสด
			B = หักผ่านธนาคาร
11	BANK_BRAN	Character	รหัสธนาคารและสาขา
12	ACCT_NO	Character	เลขที่บัญชีธนาคาร
13	NAME_OF_AC	Character	ชื่อเจ้าของบัญชีธนาคาร
14	CREDIT_CL	Character	สถานะลูกค้าเกี่ยวกับการชำระเงิน
			A = ลูกค้าที่มีอิทธิพล
			B = ลูกค้าที่ชำระเงินทันที
			C = ลูกค้าที่ชำระเงินสม่ำเสมอ
			D = ลูกค้าที่เก็บเงินไม่ค่อยได้
15	ACCT_TYPE	Character	ประเภทลูกค้า
			R = บ้านพัก
			B = ร้านค้า , สำนักงาน , บริษัท
			G = หน่วยราชการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 แสดงตารางข้อมูลพื้นฐานลูกค้า (ต่อ)

Field	Field Name	Type	Description
			H = รัฐวิสาหกิจ
			A = ผู้มีอุปการะคุณ , สำนักพระราชวัง
16	CL_OF_SERV	Character	ประเภทการให้บริการ
			R = เครื่องโทรศัพท์แบบธรรมดา
			T = เครื่องโทรศัพท์แบบกดปุ่ม
			S = วงจรพิเศษ รวมทั้งวงจรเช่าต่างๆ
			L = สายตรง (คู่สายต่างๆ)
			M = ตู้สาขาแบบธรรมดา ใช้กับโทรศัพท์ธรรมดา
			N = ตู้สาขาแบบธรรมดา ใช้กับโทรศัพท์กดปุ่ม
			P = ตู้สาขาอัตโนมัติ ใช้กับโทรศัพท์ธรรมดา
			Q = ตู้สาขาอัตโนมัติ ใช้กับโทรศัพท์ธรรมดา
			Z = โทรศัพท์ประเภท ISDN
17	BILL_RATE	Character	อัตราค่าบริการ (Bill Rate)
18	FREE_METER	Character	จำนวนครั้งในการใช้ฟรี
			100 = ใช้ฟรี 100 ครั้ง
			150 = ใช้ฟรี 150 ครั้ง
			200 = ใช้ฟรี 200 ครั้ง
19	DEP_AMOUNT	Numeric	เงินประกันการใช้โทรศัพท์ (บาทและสตางค์)
20	DEP_DATE	Character	วันที่รับเงินประกัน
21	DEP_RECEIP	Character	เลขที่ใบเสร็จเงินประกัน
22	DATE_OF_SO	Character	วันที่ให้บริการ (ปีเดือนวัน)
23	TYPE_OF_SO	Character	ประเภทการให้บริการ
24	NO_OF_SO	Character	เลขที่ใบสั่งบริการ
25	DATE_CANCE	Character	วันที่ถอนเลิกใช้โทรศัพท์
26	GOV_CODE	Character	รหัสราชการ
27	BILL_MONTH	Character	ปีเดือนของบิลล่าสุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 แสดงข้อมูลการใช้โทรศัพท์

Field	Field Name	Type	Description
1	TELNO	Character	เลขหมายโทรศัพท์
2	INS_DATE	Character	วันที่ติดตั้ง
3	STATUS	Character	สถานะของเลขหมาย
4	ACC_TYPE	Character	ประเภทผู้เช่า
5	M_USAGE	Numeric	จำนวนครั้งในการใช้
6	L_DIST_CHG	Numeric	ค่าทางไกลประจำเดือน (บาทและสตางค์)

ตารางที่ 4.3 แสดงข้อมูลการรับชำระค่าใช้บริการ

Field	Field Name	Type	Description
1	TELNO	Character	เลขหมายโทรศัพท์
2	PAID_DATE	Character	วันที่ชำระเงิน
3	BILL_MONTH	Character	ค่าใช้บริการประจำเดือน
4	LOCATION	Character	สถานที่ชำระค่าใช้บริการ
5	LOC_NAME	Character	ชื่อสถานที่ชำระเงิน
6	AMOUNT_PAID	Numeric	จำนวนเงินที่ชำระ

4.2 การเตรียมข้อมูล (Data Preparation)

เมื่อได้ข้อมูลที่จะทำการศึกษาแล้ว ขั้นตอนต่อมาคือ การเตรียมข้อมูล เพื่อพร้อมที่จะนำไปทำคาด้าไมนิ่ง ซึ่งแบ่งเป็นขั้นตอนต่างๆ ดังนี้

4.2.1 การคัดเลือกข้อมูล (Data Selection)

เลือกข้อมูลเฉพาะเลขหมายที่มีการยกเลิกการใช้บริการ กล่าวคือ CHURN = “YES” จำนวน 17,473 รายการ โดย

1) ข้อมูลพื้นฐานลูกค้า เลือกข้อมูลลูกค้าที่มีประเภทลูกค้าเป็น R , B , G , H , A แต่ละประเภทหมายถึง R บ้านพักอาศัย , B ธุรกิจ , G หน่วยงานราชการ , H รัฐวิสาหกิจ และ A ผู้มีพระคุณ สำนักพระราชวัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) ข้อมูลการใช้โทรศัพท์ เลือกข้อมูลลูกค้าที่มีการใช้โทรศัพท์ภายในพื้นที่เดียวกัน และหรือลูกค้าที่มีการใช้โทรศัพท์ทางไกล กล่าวคือ ลูกค้าที่มีจำนวนมิเตอร์ และหรือลูกค้าที่มีค่าใช้โทรศัพท์ทางไกล

3) ข้อมูลการชำระเงิน เลือกข้อมูลลูกค้าที่ชำระค่าใช้บริการโทรศัพท์เท่านั้น

4.2.2 การทำข้อมูลให้มีคุณภาพ (Data Preprocessing)

การทำข้อมูลให้มีคุณภาพมีกิจกรรมหลัก 3 กิจกรรม คือ

1. การทำข้อมูลให้สมบูรณ์ (Data Cleaning) ข้อมูลที่ได้มาทั้งหมดเป็นข้อมูลที่ไม่สมบูรณ์ ยังไม่สามารถนำไปทำคาน่าไมนิ่งได้จำเป็นต้องมีการจัดการข้อมูล การเตรียมข้อมูลเบื้องต้น มีวิธีการดังนี้

1.1 ข้อมูลพื้นฐานลูกค้า มีวิธีการ ดังนี้

? เลือกเฉพาะ Attribute ที่สำคัญที่สามารถนำไปใช้ประโยชน์ได้ และมีความครบถ้วน เช่น

- เลขหมายโทรศัพท์ (TELNO) ประกอบด้วยเลขหมาย 7 ตัว
- วันที่ติดตั้ง (IN_SERVICE) แสดงในรูปแบบ ปีเดือนวัน
- สถานะของเลขหมาย (STAT_OF_AC) คือ “F” ยกเลิกการใช้บริการ
- ประเภทลูกค้า (ACC_TYPE) เช่น R , B , G และ H เป็นต้น
- สถานการณ์หักบัญชีธนาคาร (BANK_IND) มี 2 สถานะ คือ ค่าว่างกับ “B” หักผ่านธนาคาร
- จำนวนครั้งในการใช้ฟรี (FREE_METER) จำนวนครั้งในการใช้ฟรีมี 4 ประเภทคือ ค่าว่าง = ไม่ใช้ฟรี , 100 = ใช้ฟรี 100 ครั้ง , 150 = ใช้ฟรี 150 ครั้ง และ 200 = ใช้ฟรี 200 ครั้ง

? แก้ไขข้อมูลให้ถูกต้องสมบูรณ์ และปรับเปลี่ยนข้อมูลบาง Attribute ให้มีค่าเหมาะสมในการตัดสินใจ เช่น

- รวม Attribute : TELNO+IN_SERVICE เป็น TELINS เพื่อใช้เป็น Key ในการเชื่อมโยงตารางที่เกี่ยวข้อง
- วันที่ติดตั้ง (IN_SERVICE) แยกเป็น Attribute ใหม่ 3 Attribute คือ TENURE แสดงระยะเวลาการครอบครองเลขหมายโทรศัพท์ Y_INS แสดงปีที่ติดตั้ง และ M_INS แสดงเดือนที่ติดตั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- สถานะของเลขหมาย (STAT_OF_AC) เปลี่ยนเป็น Attribute “CHURN” ให้มีค่าเป็น YES
- สถานการณ์หักบัญชีธนาคาร (BANK_IND) มีข้อมูล 2 ค่าคือ ค่าว่าง กับ “B” เปลี่ยนเป็น Attribute “P_METHOD” และปรับข้อมูลเป็น CASH จ่ายเงินสด และ BANK หักผ่านธนาคาร
- จำนวนครั้งในการใช้ฟรี FREE_METER เปลี่ยนเป็น Attribute : METER100 คือ จำนวนครั้งที่ใช้ฟรีตั้งแต่ 100 ครั้งขึ้นไป เป็น “YES” ส่วนที่ไม่ใช่กรณีดังกล่าวเป็น “NO”

จากการทำข้อมูลพื้นฐานลูกค้าทำให้ได้ตารางที่มีโครงสร้างข้อมูลดังตารางที่ 4.4 และตัวอย่างข้อมูลดังตารางที่ 4.5

ตารางที่ 4.4 ตารางแสดงโครงสร้างข้อมูลพื้นฐานลูกค้า

Field	Field Name	Type	Width	Nulls
1	TELINS	Character	13	No
2	TENURE	Numeric	6	No
3	Y_INS	Character	4	No
4	M_INS	Character	4	No
5	CHURN	Character	10	No
6	P_METHOD	Character	4	No
7	ACC_TYPE	Character	1	No
8	METER100	Character	3	No

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 ข้อมูลพื้นฐานลูกค้า

Telins	Tenure	Y_ins	M_ins	Churn	P_method	Acc_type	Meter100
2000843430908	4	2543	09	Yes	Cash	B	No
2000844430908	4	2543	09	Yes	Cash	B	No
2000845430908	4	2543	09	Yes	Cash	B	No
2000846430908	4	2543	09	Yes	Cash	B	No
2000847430908	4	2543	09	Yes	Cash	B	No
2000848430908	4	2543	09	Yes	Cash	B	No
2000849430908	4	2543	09	Yes	Cash	B	No
2000860451009	2	2545	10	Yes	Cash	G	No
2000861451009	2	2545	10	Yes	Cash	G	No
2000862451009	2	2545	10	Yes	Cash	G	No
2000863451009	2	2545	10	Yes	Cash	G	No
2000864451009	2	2545	10	Yes	Cash	G	No
2000865451009	2	2545	10	Yes	Cash	G	No
2000866451009	2	2545	10	Yes	Cash	G	No
2000868451009	2	2545	10	Yes	Cash	G	No
2000869451009	2	2545	10	Yes	Cash	G	No

1.2 ข้อมูลการใช้โทรศัพท์ มีวิธีการ ดังนี้

? สร้าง หรือกำหนดเพิ่มข้อมูลการใช้โทรศัพท์จำนวน 8 เดือน (ก.ค. 46

- ก.พ. 47) โดยใช้โปรแกรม ACL For Windows (Audit Command Language) Version 8 ซึ่งมีขั้นตอน 5 ขั้นตอนหลัก คือ

- เลือกเพิ่มข้อมูล (Select the Data File)
- กำหนดคุณลักษณะของเพิ่มข้อมูล (The Wizard Identifies Properties of Data File)
- เริ่มต้นการสร้างเพิ่มข้อมูล (The Wizard Definition)
- กำหนดคุณลักษณะของ Field (Edit Field Properties)
- สิ้นสุดการสร้างเพิ่มข้อมูล (The Wizard Finishes Creating)

ขั้นตอนการสร้าง หรือกำหนดเพิ่มข้อมูลโดยใช้โปรแกรม ACL For Windows (Audit Command Language) Version 8 แสดงในภาคผนวก ก

? ปรับเปลี่ยนข้อมูลบาง Attribute ให้มีค่าเหมาะสมในการตัดสินใจ เช่น รวม Attribute : TELNO+IN_SERVICE เป็น TELINS เพื่อใช้เป็น Key ในการเชื่อมโยงตารางที่เกี่ยวข้อง

เพิ่มข้อมูลจะมีโครงสร้างแสดงดังตารางที่ 4.6 รวมทั้งรายละเอียดข้อมูล

ตัวอย่างแสดงดังตารางที่ 4.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 โครงสร้างข้อมูลการใช้โทรศัพท์

Field	Field Name	Type	Width	Dec	Nulls
1	TELINS	Character	13		No
2	ACC_TYPE	Character	1		No
3	M_USAGE	Numeric	13		No
4	LDCHG	Numeric	20	2	No

ตารางที่ 4.7 ข้อมูลการใช้โทรศัพท์

Telins	Acc_type	M_usage01	Ldchg01
2000001250945	R	176	0.00
2000003281137	R	0	0.00
2000005220638	B	1	9.00
2000008070838	B	23	81.00
2000013070843	B	0	0.00
2000014230543	B	0	0.00
2000015030745	B	4	0.00
2000016030745	B	0	0.00
2000017240645	B	24	0.00
2000022271144	B	0	0.00
2000029020544	B	0	0.00
2000030270845	B	6	0.00
2000031160845	B	2	0.00
2000032270845	B	10	0.00
2000033140937	B	0	0.00
2000034091239	B	0	0.00

1.3 ข้อมูลการรับชำระเงิน มีวิธีการ ดังนี้

- ? สร้าง หรือกำหนดเพิ่มข้อมูลการรับชำระเงินจำนวน 9 เดือน (ก.ค. 46 – มี.ค. 47) โดยใช้โปรแกรม ACL For Windows (Audit Command Language) Version 8 ซึ่งมีขั้นตอน 5 ขั้นตอนหลัก คือ
- เลือกเพิ่มข้อมูล (Select the Data File)
 - กำหนดคุณลักษณะของเพิ่มข้อมูล (The Wizard Identifies Properties of Data File)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับเริ่มต้นการสร้างเพิ่มข้อมูล (The Wizard Definition) ระโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- กำหนดคุณลักษณะของ Field (Edit Field Properties)
- สิ้นสุดการสร้างเพิ่มข้อมูล (The Wizard Finishes Creating)

ขั้นตอนการสร้าง หรือกำหนดเพิ่มข้อมูลโดยใช้โปรแกรม ACL For Windows (Audit Command Language) Version 8 แสดงในภาคผนวก ก

? ปรับเปลี่ยนข้อมูลบาง Attribute ให้มีค่าเหมาะสมในการตัดสินใจ เช่น รวม Attribute : TELNO+IN_SERVICE เป็น TELINS เพื่อใช้เป็น Key ในการเชื่อมโยงตารางที่เกี่ยวข้อง

เพิ่มข้อมูลจะมีโครงสร้างแสดงดังตารางที่ 4.8 รวมทั้งรายละเอียดข้อมูล ตัวอย่างแสดงดังตารางที่ 4.9

ตารางที่ 4.8 โครงสร้างข้อมูลการชำระเงิน

Field	Field Name	Type	Width	Dec	Nulls
1	TEL_INS	Character	13		No
2	PAID_DATE	Character	6		No
3	BILL_MONTH	Character	4		No
4	AMT_PAID	Numeric	20	2	No

ตารางที่ 4.9 ข้อมูลการชำระเงิน

◆◆◆ Dcpm	Telins	Paid_date	Bill_month	Amt_paid
▶	2000001250945	160147	1246	1509.77
	2000003281137	060147	1246	267.50
	2000005220638	300147	1246	690.15
	2000008070838	270147	1246	366.27
	2000015030745	100147	1246	252.52
	2000016030745	100147	1246	214.00
	2000017240645	050147	1246	319.93
	2000017240645	200147	0147	291.04
	2000022271144	070147	1146	214.00
	2000029020544	230147	1246	214.00
	2000031160845	270147	0147	220.42
	2000032270845	270147	0147	246.10
	2000033140937	280147	0147	802.50
	2000048230937	140147	1246	661.26
	2000050080737	300147	1246	914.85

2) การรวมข้อมูล (Data Integration) นำข้อมูลที่มีความสมบูรณ์ทั้ง 3 ข้อมูล กล่าวคือ ข้อมูลพื้นฐานลูกค้า ข้อมูลการใช้โทรศัพท์ และข้อมูลการชำระเงิน มาสร้างความสัมพันธ์ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ให้เป็นตารางเดียวกัน โดยใช้ Attribute : TELINS เป็น Key ในการเชื่อมโยงตารางข้อมูลทั้งสาม ดังนี้

2.1 ข้อมูลการใช้โทรศัพท์ มีวิธีการดังนี้

- ? นำข้อมูลการใช้โทรศัพท์จำนวน 8 เดือนสร้างความสัมพันธ์ โดยการทำให้ Relate Tables กับข้อมูลพื้นฐานลูกค้า
- ? ตัดบาง Attribute ที่ซ้ำกับข้อมูลพื้นฐานลูกค้า เช่น สถานะเลขหมาย (ACCT_TYPE) และวันที่ติดตั้ง
- ? เพิ่ม Attribute เช่น
 - วันที่ครบกำหนดชำระ (DUE_DATE) กำหนดไว้ดังนี้
 1. DUE_DATE = '05' เลขหมายใน Cycle Bill ที่ 1 กำหนดชำระภายในวันที่ 5 ของทุกเดือน
 2. DUE_DATE = '10' เลขหมายใน Cycle Bill ที่ 2 กำหนดชำระภายในวันที่ 10 ของทุกเดือน
 3. DUE_DATE = '15' เลขหมายใน Cycle Bill ที่ 3 กำหนดชำระภายในวันที่ 15 ของทุกเดือน
 4. DUE_DATE = '20' เลขหมายใน Cycle Bill ที่ 4 กำหนดชำระภายในวันที่ 20 ของทุกเดือน
 - จำนวนมิเตอร์ที่ใช้ตั้งแต่เดือน ก.ค. 2546 – ก.พ. 2547 (METER8M) โดยรวมจำนวนมิเตอร์ตั้งแต่เดือนดังกล่าว
 - จำนวนค่าใช้โทรศัพท์ทางไกล มือถือ ตั้งแต่เดือน ก.ค. 2546 – ก.พ. 2547 (LDCHG8M) โดยรวมค่าใช้โทรศัพท์ตั้งแต่เดือนดังกล่าว

ดังนั้นจะได้ตารางความสัมพันธ์ใหม่ ซึ่งมีโครงสร้างข้อมูลดังตารางที่ 4.10 และตัวอย่างข้อมูลดังตารางที่ 4.11

ตารางที่ 4.10 โครงสร้างข้อมูลตารางความสัมพันธ์ของข้อมูลพื้นฐานลูกค้า
และข้อมูลการใช้โทรศัพท์

Field	Field Name	Type	Width	Dec	Nulls
1	TELINS	Character	13		No
2	TENURE	Numeric	6		No
3	Y_INS	Character	4		No
4	M_INS	Character	4		No
5	CHURN	Character	10		No
6	P_METHOD	Character	4		No
7	ACC_TYPE	Character	1		No
8	METER100	Character	3		No
9	RENT_CHG07	Numeric	20	2	No
10	M_USAGE07	Numeric	19		No
11	LDCHG07	Numeric	20	2	No
12	DUE_DATE	Character	4		No
13	RENT_CHG08	Numeric	20	2	No
14	M_USAGE08	Numeric	16		No
15	LDCHG08	Numeric	20	2	No
16	RENT_CHG09	Numeric	20	2	No
17	M_USAGE09	Numeric	16		No
18	LDCHG09	Numeric	20	2	No
19	RENT_CHG10	Numeric	20	2	No
20	M_USAGE10	Numeric	16		No
21	LDCHG10	Numeric	20	2	No
22	RENT_CHG11	Numeric	20	2	No
23	M_USAGE11	Numeric	16		No
24	LDCHG11	Numeric	20	2	No

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในเท่านั้น ไม่สามารถเผยแพร่ภายนอกได้โดยไม่ได้รับอนุญาตให้ใช้

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 ข้อมูลการชำระเงิน มีวิธีการ ดังนี้

? นำข้อมูลการชำระเงินสร้างความสัมพันธ์กับข้อมูลพื้นฐานลูกค้าและข้อมูลการใช้โทรศัพท์ จะได้โครงสร้างตารางความสัมพันธ์ของทั้ง3 ตารางดังกล่าว

? เพิ่ม Attribute เช่น

- จำนวนเงินรวมที่ชำระค่าใช้บริการโทรศัพท์ (PAID9M) ตั้งแต่เดือน ก.ค. 2546 – มี.ค. 2547 โดยรวมจำนวนเงินที่ชำระค่าใช้บริการโทรศัพท์ตั้งแต่เดือนดังกล่าว
- จำนวนครั้งที่มาชำระช้า (T_LATE) โดยเปรียบเทียบวันที่ถึงกำหนดชำระ (DUE_DATE)

ดังนั้นจะได้ตารางใหม่ ซึ่งมีโครงสร้างข้อมูลดังตารางที่ 4.12 และตัวอย่างข้อมูลดังตารางที่ 4.13

ตารางที่ 4.12 โครงสร้างข้อมูลจากข้อมูลลูกค้า ข้อมูลการใช้โทรศัพท์ และข้อมูลการชำระเงิน

Field	Field Name	Type	Width	Dec	Nulls
1	TELINS	Character	13		No
2	TENURE	Numeric	6		No
3	Y_INS	Character	4		No
4	M_INS	Character	4		No
5	CHURN	Character	10		No
6	P_METHOD	Character	4		No
7	ACC_TYPE	Character	1		No
8	METER100	Character	3		No
9	METER8M	Numeric	16		No
10	LDCHG8M	Numeric	20	2	No
11	PAID9M	Numeric	20	2	No
12	T_LATE	Numeric	12		No

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.13 ข้อมูลข้อมูลการใช้โทรศัพท์ และข้อมูลการชำระเงิน

Telins	Tenure	Y_ins	M_ins	Churn	P_method	Acc_type	Meter100	Meter8m	Ldchg8m	Paid9m	T_late
2000275121043	4	2543	10	YES	CASH	B	NO	423	0.00	2213.83	0
2000541271245	2	2545	12	YES	CASH	R	NO	874	0.00	2404.29	2
2000683090945	2	2545	09	YES	CASH	R	NO	2125	0.00	8533.25	0
2000733070743	4	2543	07	YES	CASH	B	NO	1	0.00	428.00	0
2000761070743	4	2543	07	YES	CASH	B	NO	245	0.00	1405.98	0
2000780110743	4	2543	07	YES	CASH	B	NO	40	0.00	770.40	0
2000863091045	2	2545	10	YES	CASH	G	NO	1828	29905.50	33661.90	6
2000864091045	2	2545	10	YES	CASH	G	NO	2	0.00	1016.50	5
2000865091045	2	2545	10	YES	CASH	G	NO	3	0.00	1019.55	5
2000866091045	2	2545	10	YES	CASH	G	NO	2	0.00	1016.50	5
2000868091045	2	2545	10	YES	CASH	G	NO	1	0.00	1019.55	5
2000869091045	2	2545	10	YES	CASH	G	NO	1	0.00	1019.55	5
2000870091045	2	2545	10	YES	CASH	G	NO	1	0.00	1019.55	5
2000871091045	2	2545	10	YES	CASH	G	NO	32	0.00	1071.39	5
2000872091045	2	2545	10	YES	CASH	G	NO	16	0.00	1034.80	5
2000873091045	2	2545	10	YES	CASH	G	NO	15	0.00	1031.75	5
2001122091045	2	2545	10	YES	CASH	G	NO	18	3.00	1836.82	5

4.2.3 การประเมินค่าข้อมูล (Evaluating the data)

ขั้นตอนในการทำความเข้าใจกับข้อมูลที่จะนำไปทำค่าทำนาย โดยควรพิจารณา
ลักษณะข้อมูลดังนี้

- ? ลักษณะการแจกแจงของข้อมูล (Data distribution)
- ? ค่าของข้อมูลที่ผิดปกติ (Missing and invalid values)
- ? ค่าของข้อมูลสูงมากหรือต่ำมาก (Outliers)

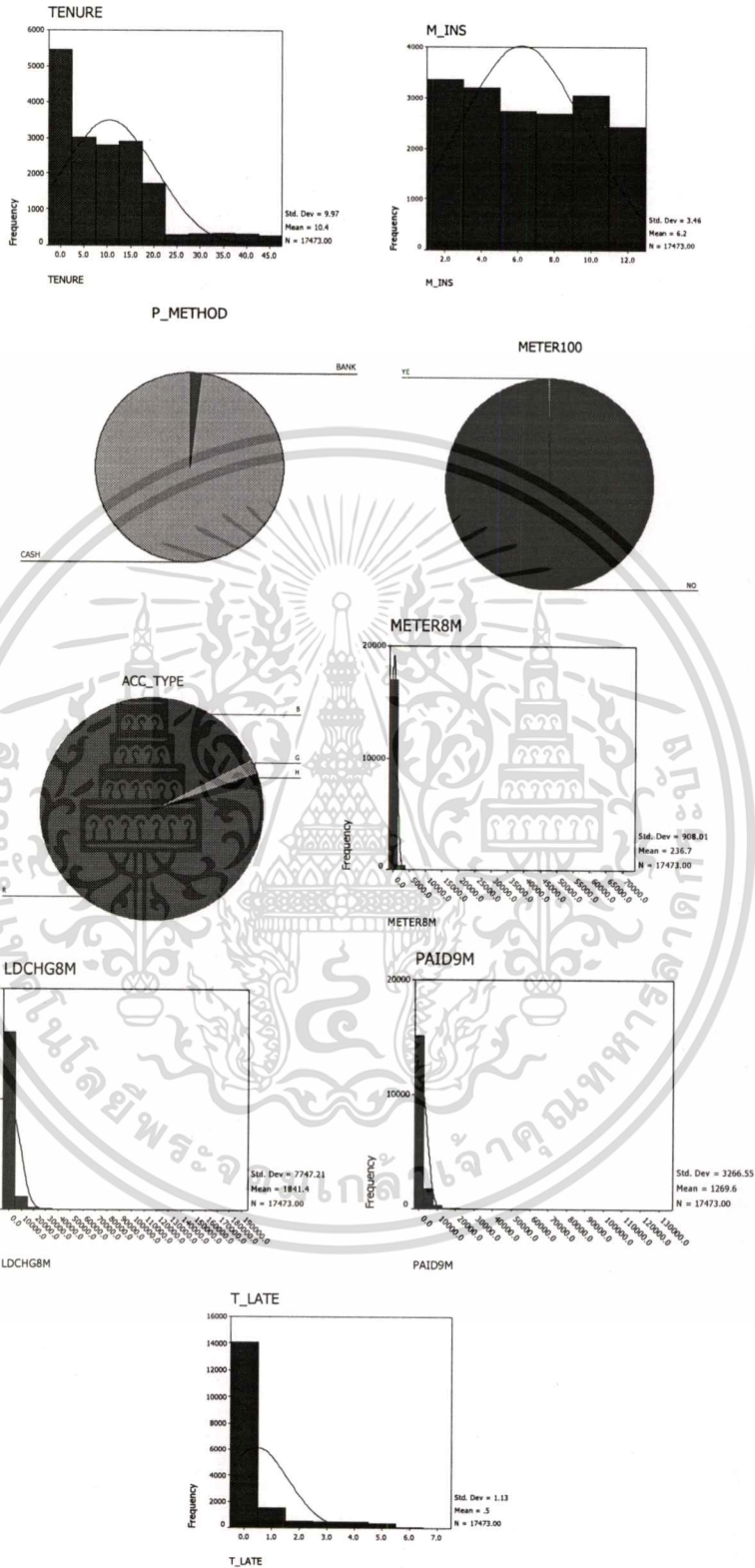
จากตารางที่ 4.13 นำข้อมูลเข้าสู่โปรแกรม SPSS เลือกข้อมูลที่จะนำไปทำค่าทำนาย
นี้ได้ตารางใหม่ดังตารางที่ 4.14 และตรวจสอบลักษณะการกระจายของข้อมูลดังรูปที่ 4.1

ตารางที่ 4.14 ข้อมูลสำหรับทำค่าทำนาย

	tenure	m_ins	p_method	acc_type	meter100	meter8m	ldchg8m	paid9m	t_late
1	4.0	10.0	CASH	B	NO	423.00	.00	2213.8	.00
2	2.0	12.0	CASH	R	NO	874.00	.00	2404.3	2.00
3	2.0	9.0	CASH	R	NO	2125.00	.00	8533.3	.00
4	4.0	7.0	CASH	B	NO	245.00	.00	1406.0	.00
5	2.0	10.0	CASH	G	NO	1828.00	29906	33662	6.00
6	2.0	10.0	CASH	G	NO	395.00	2868.0	5166.9	5.00
7	3.0	12.0	CASH	R	NO	19.00	237.00	999.38	5.00
8	2.0	1.0	CASH	R	NO	18.00	213.00	.00	.00
9	3.0	12.0	CASH	R	NO	17.00	595.00	.00	.00
10	2.0	4.0	CASH	R	NO	3.00	282.00	.00	.00
11	5.0	3.0	CASH	R	NO	318.00	5939.0	500.00	1.00
12	5.0	3.0	CASH	R	NO	272.00	1554.0	.00	.00
13	5.0	3.0	CASH	R	NO	.00	15334	12506	5.00
14	5.0	3.0	CASH	R	NO	142.00	222.00	.00	.00
15	9.0	12.0	CASH	B	NO	62.00	594.00	1373.9	.00
16	9.0	12.0	CASH	B	NO	61.00	45.00	851.72	.00
17	2.0	8.0	CASH	B	NO	374.00	4106.0	5858.3	.00
18	2.0	8.0	CASH	B	NO	153.00	4761.0	5887.1	.00
19	2.0	8.0	CASH	R	NO	6.00	549.00	.00	.00
20	3.0	12.0	CASH	B	NO	130.00	650.00	1861.8	5.00
21	12.0	12.0	CASH	R	NO	406.00	3.00	2055.5	.00
22	8.0	5.0	CASH	R	NO	72.00	860.00	1636.0	.00
23	5.0	1.0	CASH	B	NO	3143.00	18992	23832	.00

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์หรือสงวนเพื่อการศึกษาเท่านั้น และอาจถูกนำไปใช้โดยไม่ได้รับอนุญาต

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.1 การกระจายของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าสถิติของแต่ละ Attribute มีดังนี้

- 1) TENURE เป็นข้อมูลเชิงปริมาณ (Quantitative Data) มีการแจกแจงข้อมูล ดังรูปที่ 4.1 กล่าวคือ ค่าเฉลี่ย (Mean) สูงกว่าค่าฐานนิยม (Mode) และค่ามัธยฐาน (Median) เนื่องจากมีข้อมูลบางค่าที่มีความถี่สูงมาก แสดงดังตารางที่ 4.15

ตารางที่ 4.15 ค่าสถิติของข้อมูล TENURE

Statistics		
TENURE		
N	Valid	17473
	Missing	0
Mean		10.4329
Median		9.0000
Mode		2.00
Std. Deviation		9.97493
Skewness		1.453
Std. Error of Skewness		.019
Minimum		.00
Maximum		46.00
Percentiles	25	2.0000
	50	9.0000
	75	15.0000

- 2) M_INS เป็นข้อมูลเชิงปริมาณ (Quantitative Data) มีการแจกแจงข้อมูลดังรูปที่ 4.1 กล่าวคือ ค่าเฉลี่ย (Mean) และค่ามัธยฐาน (Median) ใกล้เคียงกัน แสดงดังตารางที่ 4.16

ตารางที่ 4.16 ค่าสถิติของข้อมูล MONTH

Statistics		
M_INS		
N	Valid	17473
	Missing	0
Mean		6.200
Median		6.000
Mode		3.0
Std. Deviation		3.4556
Skewness		.076
Std. Error of Skewness		.019
Minimum		1.0
Maximum		12.0
Percentiles	25	3.000
	50	6.000
	75	9.000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 3) P_METHOD เป็นตัวแปรเชิงกลุ่ม (Categorical Data) มีการแจกแจงความถี่ของข้อมูลแยกตามประเภทการชำระค่าใช้บริการ ดังรูปที่ 4.1 กล่าวคือ ชำระด้วยเงินสด จำนวน 17,123 คิดเป็น 98% และชำระโดยหักผ่านบัญชีธนาคาร จำนวน 350 คิดเป็น 2% แสดงดังตารางที่ 4.17

ตารางที่ 4.17 ค่าสถิติของข้อมูล P_METHOD

P_METHOD					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	BANK	350	2.0	2.0	2.0
	CASH	17123	98.0	98.0	100.0
	Total	17473	100.0	100.0	

- 4) ACC_TYPE เป็นตัวแปรเชิงกลุ่ม (Categorical Data) มีการแจกแจงความถี่ของข้อมูลแยกตามประเภทลูกค้า ดังรูปที่ 4.1 กล่าวคือ
- ? ธุรกิจ (B) จำนวน 3,036 คิดเป็น 17.4%
 - ? ราชการ (G) จำนวน 418 คิดเป็น 2.4%
 - ? รัฐวิสาหกิจ (H) จำนวน 216 คิดเป็น 1.2%
 - ? บ้านพักอาศัย (R) จำนวน 13,803 คิดเป็น 79%
- แสดงดังตารางที่ 4.18

ตารางที่ 4.18 ค่าสถิติของข้อมูล ACC_TYPE

ACC_TYPE					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	B	3036	17.4	17.4	17.4
	G	418	2.4	2.4	19.8
	H	216	1.2	1.2	21.0
	R	13803	79.0	79.0	100.0
	Total	17473	100.0	100.0	

- 5) METER100 เป็นตัวแปรเชิงกลุ่ม (Categorical Data) มีการแจกแจงความถี่ของข้อมูลแยกตามประเภทการใช้มิเตอร์ฟรี ดังรูปที่ 4.1 กล่าวคือ
- ? YES ใช้มิเตอร์ฟรี จำนวนตั้งแต่ 100 - 200 จำนวน 39 คิดเป็น 0.2%
 - ? NO ไม่ใช้มิเตอร์ฟรี จำนวน 17434 คิดเป็น 99.8%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แสดงคั้งตารางที่ 4.19

ตารางที่ 4.19 ค่าสถิติของข้อมูล METER100

METER100

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid NO	17434	99.8	99.8	99.8
YE	39	.2	.2	100.0
Total	17473	100.0	100.0	

6) METER8M เป็นข้อมูลเชิงปริมาณ (Quantitative Data) มีการแจกแจงข้อมูล ดังรูปที่ 4.1 กล่าวคือ ค่าเฉลี่ย (Mean) สูงกว่าค่าฐานนิยม (Mode) และ ค้ำมัธยฐาน (Median) เนื่องจากมีข้อมูลบางค่าที่มีความถี่สูงมาก แสดงคั้ง ตารางที่ 4.20

ตารางที่ 4.20 ค่าสถิติของข้อมูล METER8M

Statistics

METER8M		
N	Valid	17473
	Missing	0
Mean		236.6619
Median		70.0000
Mode		1.00
Std. Deviation		908.0071
Skewness		40.167
Std. Error of Skewness		.019
Minimum		.00
Maximum		72247.00
Percentiles	25	18.0000
	50	70.0000
	75	225.0000

7) LDCHG8M เป็นข้อมูลเชิงปริมาณ (Quantitative Data) มีการแจกแจงข้อมูล ดังรูปที่ 4.1 กล่าวคือ ค่าเฉลี่ย (Mean) สูงกว่าค่าฐานนิยม (Mode) และ ค้ำมัธยฐาน (Median) เนื่องจากมีข้อมูลบางค่าที่มีความถี่สูงมาก แสดงคั้ง ตารางที่ 4.21

ตารางที่ 4.21 ค่าสถิติของข้อมูล LDCHG8M

Statistics		
LDCHG8M		
N	Valid	17473
	Missing	0
Mean		1841.367
Median		391.0000
Mode		.00
Std. Deviation		7747.210
Skewness		17.964
Std. Error of Skewness		.019
Minimum		.00
Maximum		185195.0
Percentiles	25	39.0000
	50	391.0000
	75	1567.500

- 8) PAID9M เป็นข้อมูลเชิงปริมาณ (Quantitative Data) มีการแจกแจงข้อมูล ดังรูปที่ 4.1 กล่าวคือ ค่าเฉลี่ย (Mean) สูงกว่าค่าฐานนิยม (Mode) และค่ามัธยฐาน (Median) เนื่องจากมีข้อมูลบางค่าที่มีค่าสูงมาก แสดงดังตารางที่ 4.22

ตารางที่ 4.22 ค่าสถิติของข้อมูล PAID9M

Statistics		
PAID9M		
N	Valid	17473
	Missing	0
Mean		1269.602
Median		299.6000
Mode		.00
Std. Deviation		3266.549
Skewness		10.264
Std. Error of Skewness		.019
Minimum		.00
Maximum		134888.5
Percentiles	25	.0000
	50	299.6000
	75	1288.760

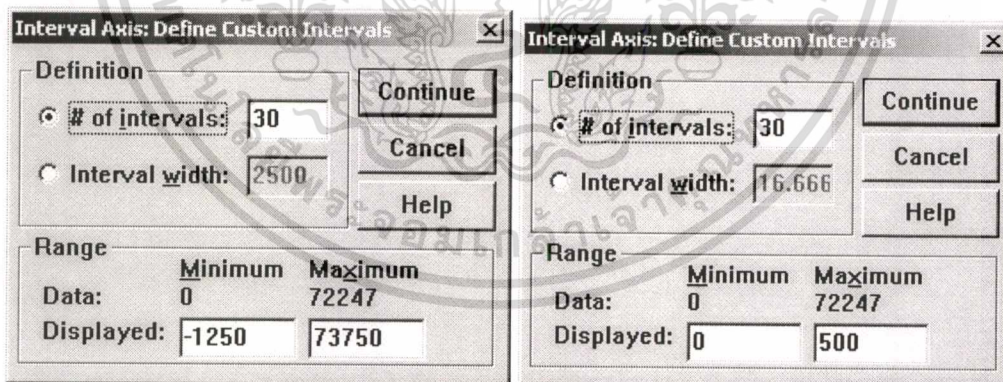
- 9) T_LATE เป็นข้อมูลเชิงปริมาณ (Quantitative Data) มีการแจกแจงข้อมูลดังรูปที่ 4.1 และมีค่าเฉลี่ย (Mean) สูงกว่าค่าฐานนิยม (Mode) และค่ามัธยฐาน (Median) เนื่องจากมีข้อมูลบางค่าที่มีค่าความถี่สูงมาก แสดงดังตารางที่ 4.23

ตารางที่ 4.23 ค่าสถิติของข้อมูล T_LATE

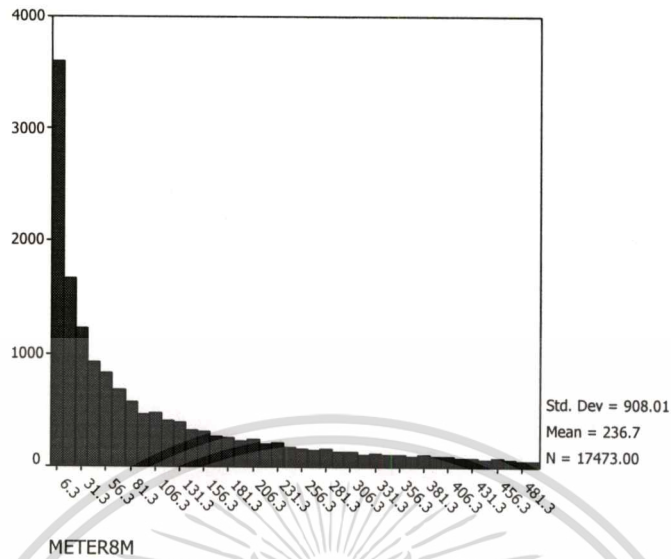
Statistics		
T_LATE		
N	Valid	17473
	Missing	0
Mean		.4524
Median		.0000
Mode		.00
Std. Deviation		1.13436
Skewness		2.829
Std. Error of Skewness		.019
Minimum		.00
Maximum		7.00
Percentiles	25	.0000
	50	.0000
	75	.0000

จากการศึกษาลักษณะของข้อมูล พบว่ามีตัวแปรที่มีค่าของข้อมูลที่สูงมากหรือต่ำมาก (Outlier) ทำให้มีการกระจายของข้อมูลผิดปกติ ดังนั้นปรับค่าของข้อมูลให้มีการกระจายมากขึ้น โดยใช้ Histogram ดังนี้

1. METER8M กำหนดค่า Maximum ใหม่ โดยเปลี่ยนค่าต่ำสุดจาก -2,500 เป็น 0 และค่าสูงสุดจาก 73,750 เป็น 500 ดังรูปที่ 4.2 ได้ Histogram ใหม่แสดงดังรูปที่ 4.3



รูปที่ 4.2 การกำหนดค่า Interval ของ METER8M



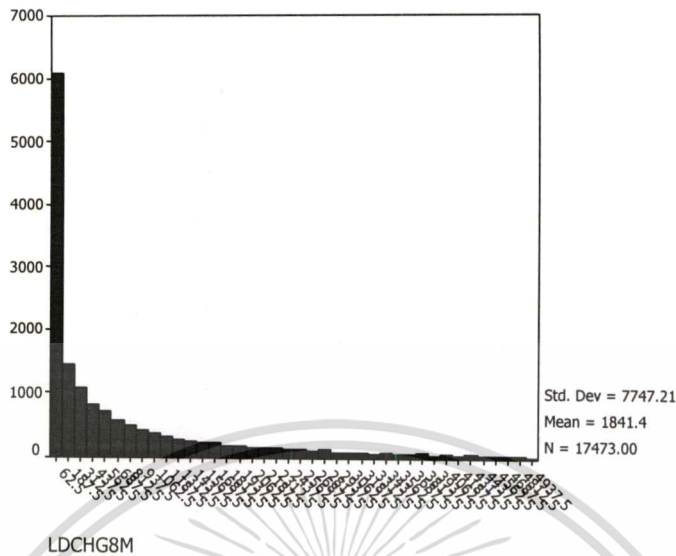
รูปที่ 4.3 New Histogram METER8M

2. LDCHG8M กำหนดค่า Maximum ใหม่ โดยเปลี่ยนค่าต่ำสุดจาก -5,000 เป็น 0 และค่าสูงสุดจาก 195,000 เป็น 5,000 ดังรูปที่ 4.4 ได้ Histogram ใหม่แสดงดังรูปที่ 4.5

Interval Axis: Define Custom Intervals		
Definition		
<input checked="" type="radio"/> # of intervals:	20	
<input type="radio"/> Interval width:	10000	
Continue		
Cancel		
Help		
Range		
	Minimum	Maximum
Data:	0	185195
Displayed:	-5000	195000

Interval Axis: Define Custom Intervals		
Definition		
<input checked="" type="radio"/> # of intervals:	40	
<input type="radio"/> Interval width:	10000	
Continue		
Cancel		
Help		
Range		
	Minimum	Maximum
Data:	0	185195
Displayed:	0	5000

รูปที่ 4.4 การกำหนดค่า Interval ของ LDCHG8M



รูปที่ 4.5 New Histogram LDCHG8M

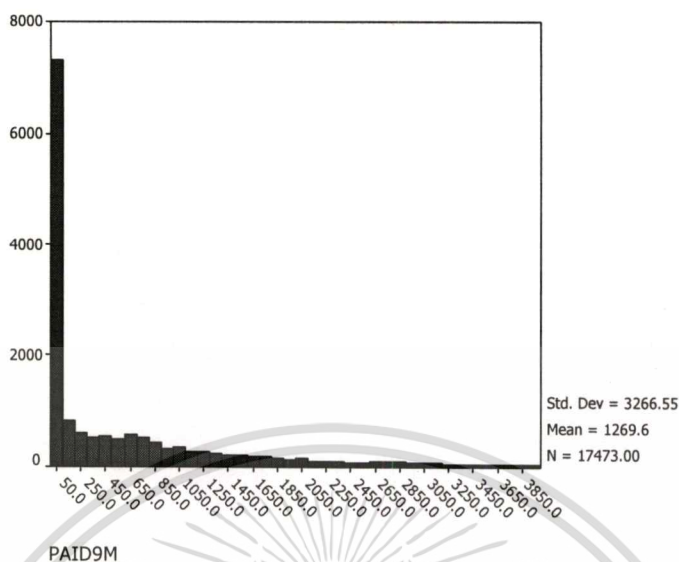
3. PAID9M กำหนดค่า Maximum ใหม่ โดยเปลี่ยนค่าต่ำสุดจาก -2,500 เป็น 0 และค่าสูงสุดจาก 137,500 เป็น 4,000 ดังรูปที่ 4.6 ได้ Histogram ใหม่แสดงดังรูปที่ 4.7

Interval Axis: Define Custom Intervals		
Definition		
<input checked="" type="radio"/> # of intervals:	28	Continue
<input type="radio"/> Interval width:	5000	Cancel
Help		
Range		
	Minimum	Maximum
Data:	0	1.35E+05
Displayed:	-2500	137500

Interval Axis: Define Custom Intervals		
Definition		
<input checked="" type="radio"/> # of intervals:	40	Continue
<input type="radio"/> Interval width:	5000	Cancel
Help		
Range		
	Minimum	Maximum
Data:	0	1.35E+05
Displayed:	0	4000

รูปที่ 4.6 การกำหนดค่า Interval ของ PAID9M

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.7 New Histogram PAID9M

4.2.3 การปรับเปลี่ยนรูปแบบข้อมูล (Data Transformation)

หากนำข้อมูลข้างต้นไปทำ K-Means Cluster ในขั้นตอนต่อไป ผลลัพธ์จากการแบ่งกลุ่มอาจไม่สามารถแยกความแตกต่างของแต่ละกลุ่มได้อย่างชัดเจน ดังนั้นจำเป็นต้องแปลงค่าข้อมูลทุก Attribute ให้อยู่ระหว่างค่า 0 - 1 โดยใช้สูตร Min - Max Normalization ดังนี้

$$Y = \frac{(X - a)}{(b - a)} + 1$$

โดย b = ค่าสูงสุด (Maximum Value)

a = ค่าต่ำสุด (Minimum Value)

X = ค่าของข้อมูล

1) เปลี่ยนแปลงรูปแบบข้อมูล (Recode) จากตารางที่ 4.14 ข้อมูลที่เป็นข้อมูลเชิงกลุ่ม (Categorical Data) ให้เป็นข้อมูลเชิงปริมาณ (Quantitative Data) ซึ่งมีขั้นตอนการทำตามภาคผนวก ก จะได้ตารางข้อมูลดังตารางที่ 4.24

ตารางที่ 4.24 ตารางข้อมูลที่ผ่านมาการ Recode ข้อมูลจากข้อมูลเชิงกลุ่มเป็นข้อมูลเชิงปริมาณ

	tenure	month	p_method	acc_type	meter100	meter8m	ldchg8m	paid9m	t_late	pay	acctype	fmeter
1	4.00	10	CASH	B	NO	423.00	.00	2213.83	.00	1.00	2.00	2.00
2	2.00	12	CASH	R	NO	874.00	.00	2404.29	2.00	1.00	1.00	2.00
3	2.00	9.0	CASH	R	NO	2125.0	.00	8533.25	.00	1.00	1.00	2.00
4	4.00	7.0	CASH	B	NO	1.00	.00	428.00	.00	1.00	2.00	2.00
5	4.00	7.0	CASH	B	NO	245.00	.00	1405.98	.00	1.00	2.00	2.00
6	4.00	7.0	CASH	B	NO	40.00	.00	770.40	.00	1.00	2.00	2.00
7	2.00	10	CASH	G	NO	1828.0	29906	33661.90	6.00	1.00	3.00	2.00
8	2.00	10	CASH	G	NO	2.00	.00	1016.50	5.00	1.00	3.00	2.00
9	2.00	10	CASH	G	NO	3.00	.00	1019.55	5.00	1.00	3.00	2.00
10	2.00	10	CASH	G	NO	2.00	.00	1016.50	5.00	1.00	3.00	2.00
11	2.00	10	CASH	G	NO	1.00	.00	1019.55	5.00	1.00	3.00	2.00
12	2.00	10	CASH	G	NO	1.00	.00	1019.55	5.00	1.00	3.00	2.00
13	2.00	10	CASH	G	NO	1.00	.00	1019.55	5.00	1.00	3.00	2.00
14	2.00	10	CASH	G	NO	32.00	.00	1071.39	5.00	1.00	3.00	2.00
15	2.00	10	CASH	G	NO	16.00	.00	1034.80	5.00	1.00	3.00	2.00
16	2.00	10	CASH	G	NO	15.00	.00	1031.75	5.00	1.00	3.00	2.00
17	2.00	10	CASH	G	NO	18.00	3.00	1836.82	5.00	1.00	3.00	2.00
18	2.00	10	CASH	G	NO	7.00	9.00	1797.17	5.00	1.00	3.00	2.00
19	2.00	10	CASH	G	NO	395.00	2868	5166.86	5.00	1.00	3.00	2.00
20	3.00	12	CASH	R	NO	19.00	237.0	999.38	5.00	1.00	1.00	2.00

2) Normalization ข้อมูลที่ได้จากการทำ Histogram ใหม่ โดยใช้สูตร Min – Max Normalization ข้างต้น โดยกำหนดค่าสูงสุดใหม่ ดังนี้ และจะได้ตารางที่ผ่านมาการ Normalization ดังตารางที่ 4.25

$$\text{TENURE : } Y_{\text{TENURE}} = \frac{(X - 46) + 1}{(46 - 0)}$$

$$\text{MONTH : } Y_{\text{MONTH}} = \frac{(X - 12) + 1}{(12 - 1)}$$

$$\text{PAY : } Y_{\text{PAY}} = \frac{(X - 2) + 1}{(2 - 1)}$$

$$\text{ACCTYPE : } Y_{\text{ACCTYPE}} = \frac{(X - 2) + 1}{(2 - 1)}$$

$$\text{FMETER : } Y_{\text{FMETER}} = \frac{(X - 2) + 1}{(2 - 1)}$$

$$\text{METER8M : } Y_{\text{METER8M}} = \frac{(X - 500) + 1}{(500 - 0)}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\text{LDCHG8M} : \quad \text{YLDCHG8M} = (X - 5,000) + 1$$

$$(5,000 - 0)$$

$$\text{PAID9M} : \quad \text{YPAID9M} = (X - 4,000) + 1$$

$$(4,000 - 0)$$

$$\text{T_LATE} : \quad \text{YT_LATE} = (X - 7) + 1$$

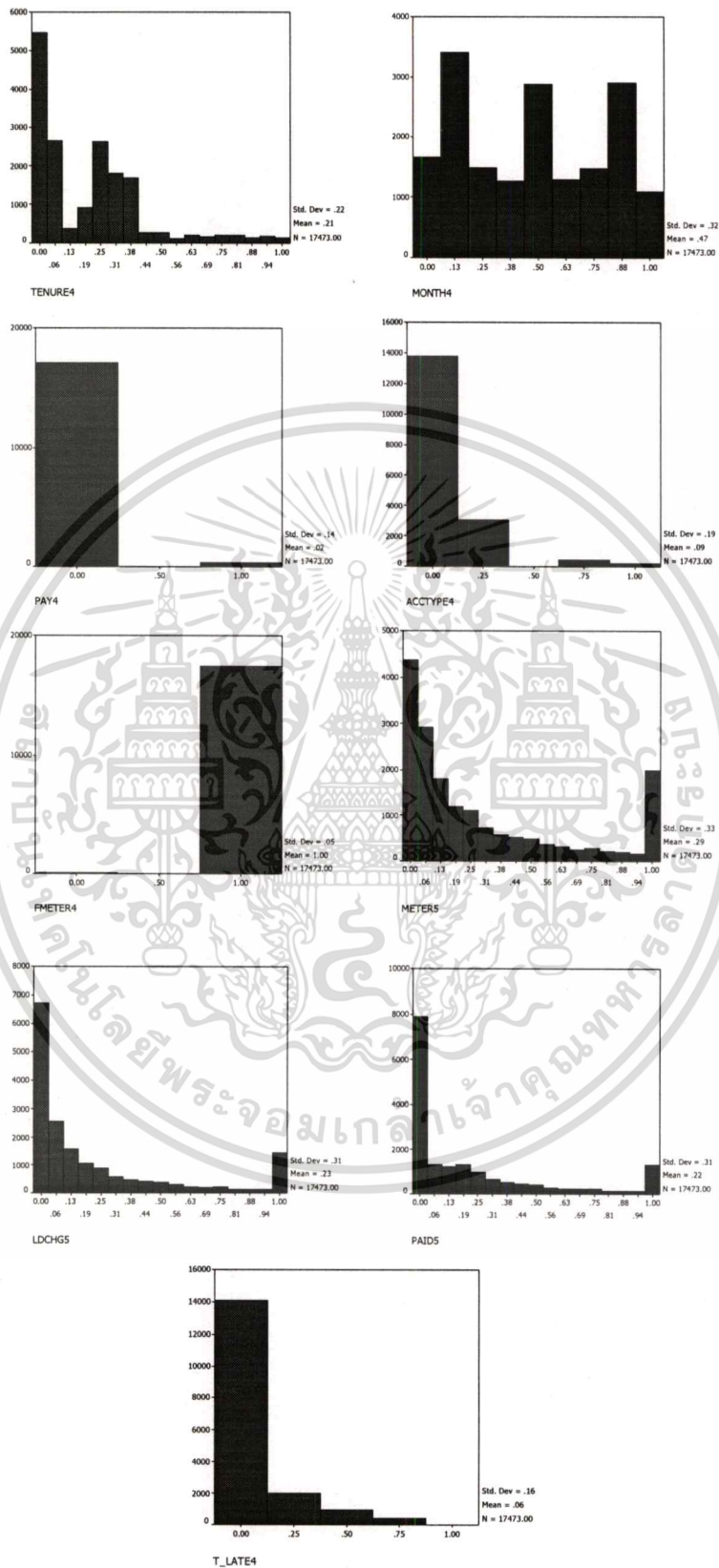
$$(7 - 0)$$

ตารางที่ 4.25 ตารางข้อมูลผ่านการ Normalization

	tenure	month	p_meth	acc_by	met	meter8m	ldchg8m	paid9m	t_late	pay	accty	fmete	lanur	month	pay+	acctyp	fmet	meter	ldchg	paid5	t_late+
1	4.00	10.0	CASH	B	NO	423.00	.00	2213.83	.00	1.0	2.00	2.00	.07	82	.00	.33	1.0	.85	.00	.55	.00
2	2.00	12.0	CASH	R	NO	674.00	.00	2404.29	2.00	1.0	1.00	2.00	.02	1.00	.00	.00	1.0	1.00	.00	.60	.29
3	2.00	9.0	CASH	R	NO	2125.0	.00	8533.25	.00	1.0	1.00	2.00	.02	73	.00	.00	1.0	1.00	.00	1.00	.00
4	4.00	7.0	CASH	B	NO	1.00	.00	428.00	.00	1.0	2.00	2.00	.07	55	.00	.33	1.0	.00	.00	.11	.00
5	4.00	7.0	CASH	B	NO	245.00	.00	1405.99	.00	1.0	2.00	2.00	.07	55	.00	.33	1.0	.49	.00	.35	.00
6	4.00	7.0	CASH	B	NO	40.00	.00	770.40	.00	1.0	2.00	2.00	.07	55	.00	.33	1.0	.08	.00	.19	.00
7	2.00	10.0	CASH	G	NO	1828.0	29905.50	39661.90	6.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	1.00	1.00	1.00	.86
8	2.00	10.0	CASH	G	NO	2.00	.00	1016.50	5.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	.00	.00	.25	.71
9	2.00	10.0	CASH	G	NO	3.00	.00	1019.55	5.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	.01	.00	.25	.71
10	2.00	10.0	CASH	G	NO	2.00	.00	1016.50	5.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	.00	.00	.25	.71
11	2.00	10.0	CASH	G	NO	1.00	.00	1019.55	5.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	.00	.00	.25	.71
12	2.00	10.0	CASH	G	NO	1.00	.00	1019.55	5.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	.00	.00	.25	.71
13	2.00	10.0	CASH	G	NO	1.00	.00	1019.55	5.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	.00	.00	.25	.71
14	2.00	10.0	CASH	G	NO	32.00	.00	1071.39	5.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	.06	.00	.27	.71
15	2.00	10.0	CASH	G	NO	16.00	.00	1034.80	5.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	.03	.00	.26	.71
16	2.00	10.0	CASH	G	NO	15.00	.00	1031.75	5.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	.03	.00	.26	.71
17	2.00	10.0	CASH	G	NO	18.00	3.00	1896.82	5.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	.04	.00	.46	.71
18	2.00	10.0	CASH	G	NO	7.00	9.00	1797.17	5.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	.01	.00	.45	.71
19	2.00	10.0	CASH	G	NO	395.00	2868.00	5166.86	5.00	1.0	3.00	2.00	.02	82	.00	.67	1.0	.79	.57	1.00	.71
20	3.00	12.0	CASH	R	NO	19.00	237.00	999.38	5.00	1.0	1.00	2.00	.04	1.00	.00	.00	1.0	.04	.06	.25	.71
21	2.00	1.0	CASH	R	NO	18.00	213.00	.00	.00	1.0	1.00	2.00	.02	.00	.00	.00	1.0	.04	.04	.00	.00
22	3.00	12.0	CASH	R	NO	17.00	595.00	.00	.00	1.0	1.00	2.00	.04	1.00	.00	.00	1.0	.03	.12	.00	.00
23	2.00	4.0	CASH	R	NO	3.00	282.00	.00	.00	1.0	1.00	2.00	.02	.27	.00	.00	1.0	.01	.06	.00	.00
24	2.00	12.0	CASH	R	NO	20.00	166.00	.00	.00	1.0	1.00	2.00	.02	1.00	.00	.00	1.0	.04	.03	.00	.00
25	2.00	2.0	CASH	R	NO	11.00	24.00	.00	.00	1.0	1.00	2.00	.02	.09	.00	.00	1.0	.02	.00	.00	.00
26	5.00	3.0	CASH	R	NO	318.00	5939.00	500.00	1.00	1.0	1.00	2.00	.09	18	.00	.00	1.0	.64	1.00	.12	.14
27	5.00	3.0	CASH	R	NO	272.00	1554.00	.00	.00	1.0	1.00	2.00	.09	18	.00	.00	1.0	.54	.31	.00	.00
28	5.00	3.0	CASH	R	NO	.00	15334.00	12505.64	5.00	1.0	1.00	2.00	.09	18	.00	.00	1.0	.00	1.00	1.00	.71
29	5.00	3.0	CASH	R	NO	142.00	222.00	.00	.00	1.0	1.00	2.00	.09	18	.00	.00	1.0	.28	.04	.00	.00
30	9.00	12.0	CASH	B	NO	2.00	.00	9741.13	.00	1.0	2.00	2.00	.18	1.00	.00	.33	1.0	.00	.00	1.00	.00
31	9.00	12.0	CASH	B	NO	62.00	594.00	1373.88	.00	1.0	2.00	2.00	.18	1.00	.00	.33	1.0	.12	.12	.34	.00

3) ตรวจสอบการกระจายของข้อมูลหลังจากทำ Normalization พบว่ามีลักษณะการกระจายข้อมูลต่างๆ มากขึ้นดังรูปที่ 4.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับรูปที่ 4.8 Histogram ใหม่หลังจาก Normalization นี้ไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 การทำดาต้าไมนิ่ง

การทำนายลูกค้าที่มีแนวโน้มจะยกเลิกการใช้บริการโทรศัพท์ที่มีการดำเนินงาน 2 ขั้นตอนหลัก ดังนี้

4.3.1 จัดกลุ่มลูกค้าที่ยกเลิกการใช้บริการ ที่มีลักษณะเหมือนกันไว้ด้วยกันโดยใช้ K-Means Cluster ซึ่งมีขั้นตอนการทำ Cluster ในภาคผนวก ก

- แบ่งกลุ่มลูกค้าเริ่มต้นที่ 8 กลุ่ม ผลลัพธ์การแบ่งกลุ่มแสดงดังตารางที่ 4.26 และมีจำนวนสมาชิกในแต่ละ Cluster ดังตารางที่ 4.27 แต่พบว่า Attribute : FMETER4 การใช้มิเตอร์โดยไม่เสียค่าใช้จ่ายไม่มีผลต่อการแบ่งกลุ่ม ดังนั้นจึงไม่นำ Attribute ดังกล่าวมาใช้ในการแบ่งกลุ่ม

ตารางที่ 4.26 การแบ่งกลุ่มลูกค้าเป็น 8 กลุ่ม

	Cluster							
	1	2	3	4	5	6	7	8
TENURE4	.34	.16	.13	.19	.77	.18	.22	.25
MONTH4	.54	.46	.21	.44	.50	.77	.48	.51
PAY4	.00	.00	.02	.03	.04	.02	.04	.04
ACCTYPE4	.66	.03	.04	.07	.09	.05	.14	.16
FMETER4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
METER5	.19	.50	.12	.27	.18	.12	.87	.93
LDCHG5	.06	.82	.09	.67	.09	.09	.13	.81
PAID5	.20	.07	.08	.78	.15	.09	.49	.94
T_LATE4	.44	.01	.04	.18	.03	.04	.07	.11

ตารางที่ 4.27 จำนวนสมาชิกในการแบ่งกลุ่มลูกค้าเป็น 8 กลุ่ม

Number of Cases in each Cluster

Cluster	1	653.000
	2	1375.000
	3	5800.000
	4	854.000
	5	980.000
	6	4904.000
	7	1805.000
	8	1102.000
Valid		17473.00
Missing		.000

- แบ่งกลุ่มลูกค้าเป็น 8 กลุ่มอีกครั้ง โดยตัด Attribute : FMETER4 ที่ไม่มีผลต่อการแบ่งกลุ่ม ผลลัพธ์การแบ่งกลุ่มแสดงดังตารางที่ 4.28 และมีจำนวนสมาชิกในแต่ละ Cluster ดังตารางที่ 4.29 แต่พบว่า กลุ่มที่ 5 และ 7 มีลักษณะใกล้เคียงกัน ดังนั้นต้องแบ่งกลุ่มให้น้อยลง

ตารางที่ 4.28 การแบ่งกลุ่มลูกค้าเป็น 8 กลุ่มตัด Attribute : FEMETR4

	Cluster							
	1	2	3	4	5	6	7	8
TENURE4	.22	.16	.77	.18	.12	.34	.18	.26
MONTH4	.48	.46	.50	.44	.21	.54	.77	.51
PAY4	.04	.00	.04	.03	.02	.00	.02	.04
ACCTYPE4	.14	.03	.09	.07	.04	.67	.05	.16
METER5	.87	.49	.18	.27	.12	.20	.12	.93
LDCHG5	.13	.82	.09	.67	.09	.06	.09	.81
PAID5	.49	.07	.15	.77	.08	.20	.09	.94
T_LATE4	.07	.01	.03	.18	.04	.44	.04	.11

ตารางที่ 4.29 จำนวนสมาชิกในการแบ่งกลุ่มลูกค้าเป็น 8 กลุ่ม ตัด Attribute : FEMETR4

Number of Cases in each Cluster

Cluster	1	1802.000
	2	1386.000
	3	981.000
	4	849.000
	5	5805.000
	6	643.000
	7	4900.000
	8	1107.000
Valid		17473.00
Missing		.000

- แบ่งกลุ่มลูกค้าเป็น 6 กลุ่ม ผลลัพธ์การแบ่งกลุ่มแสดงดังตารางที่ 4.30 และมีจำนวนสมาชิกในแต่ละ Cluster ดังตารางที่ 4.31 พบว่า กลุ่มที่ 5 และ 7 มีลักษณะใกล้เคียงกัน ดังนั้นต้องแบ่งกลุ่มให้น้อยลง

ตารางที่ 4.30 การแบ่งกลุ่มลูกค้าเป็น 6 กลุ่ม

Final Cluster Centers

	Cluster					
	1	2	3	4	5	6
TENURE4	.23	.23	.18	.76	.13	.16
MONTH4	.48	.49	.77	.51	.21	.45
PAY4	.04	.03	.02	.03	.02	.00
ACCTYPE4	.16	.13	.08	.17	.05	.03
METER5	.85	.72	.12	.18	.12	.48
LDCHG5	.14	.81	.09	.09	.10	.82
PAID5	.51	.92	.10	.15	.09	.09
T_LATE4	.10	.14	.06	.06	.05	.01

ตารางที่ 4.31 จำนวนสมาชิกในการแบ่งกลุ่มลูกค้าเป็น 6 กลุ่ม

Number of Cases in each Cluster

Cluster	1	2001.000
	2	1597.000
	3	5209.000
	4	1082.000
	5	6086.000
	6	1498.000
Valid		17473.00
Missing		.000

- แบ่งกลุ่มลูกค้าเป็น 4 กลุ่ม และตัด Attribute : MONTH4 ซึ่งไม่มีผลต่อการแบ่งกลุ่มออก ผลลัพธ์การแบ่งกลุ่มแสดงดังตารางที่ 4.32 และมีจำนวนสมาชิกในแต่ละ Cluster ดังตารางที่ 4.33 พบว่า ไม่มีกลุ่มใดที่มีลักษณะที่คล้ายกัน ดังนั้นจึงสรุปได้ว่ากลุ่มลูกค้าแบ่งได้เป็น 4 กลุ่ม

ตารางที่ 4.32 การแบ่งกลุ่มลูกค้าเป็น 4 กลุ่ม

Final Cluster Centers

	Cluster			
	1	2	3	4
TENURE4	.20	.27	.16	.24
MONTH4	.48	.48	.21	.77
PAY4	.02	.03	.02	.02
ACCTYPE4	.09	.16	.06	.09
METER5	.61	.85	.13	.13
LDCHG5	.89	.18	.11	.11
PAID5	.58	.53	.09	.10
T_LATE4	.09	.10	.05	.06

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.33 จำนวนสมาชิกในการแบ่งกลุ่มลูกค้าเป็น 4 กลุ่ม

Number of Cases in each Cluster

Cluster	1	2401.000
	2	2360.000
	3	6721.000
	4	5991.000
Valid		17473.00
Missing		.000

4.3.2 ทา Pattern ลักษณะของลูกค้าแต่ละกลุ่ม จากการแบ่งกลุ่มลูกค้าเป็น 4 กลุ่มข้างต้น โดยใช้โปรแกรม Answer Tree 3.0 ซึ่งเป็นโปรแกรมคอมพิวเตอร์ที่เรียนรู้จาก Rule ของข้อมูลเดิม เพื่อแบ่งกลุ่มข้อมูล (Classification) โดยแสดงในรูปแบบของ Decision Tree โปรแกรมนี้สามารถใช้ได้กับข้อมูล ดังนี้

- ? SPSS Data File
- ? ODBC Data Source
- ? Excel Spreadsheets
- ? Text Files

วิธีการในการสร้าง Tree ของ Answer Tree มี 4 วิธี คือ

- 1) Merging การหาความสัมพันธ์ของตัวแปรที่ใช้ในการพยากรณ์กับข้อมูลที่เป็น Target โดยโปรแกรมจะยุบรวมตัวแปรที่ไม่มีความสำคัญเป็นกลุ่มหนึ่ง และตัวแปรที่มีความสำคัญเป็นกลุ่มหนึ่ง
- 2) Splitting การเลือกจุดที่จะแตกกิ่งของ Tree โดยอาจใช้ตัวแปรประชากร ซึ่งถูกเลือกโดยการเปรียบเทียบกับประชากรทั้งหมด
- 3) Stopping กฎซึ่งถูกกำหนดว่าเมื่อไหร่จึงจะหยุดสร้าง Tree
- 4) Pruning การยุบกิ่ง ซึ่งเมื่อแตกออกไปแล้วมีค่าหรือกลุ่มข้อมูลที่แสดงข้อมูลน้อย และนำมาใช้ในการพยากรณ์ไม่เกิดประโยชน์

Algorithms ที่ใช้ใน โปรแกรม Answer Tree มี 4 Algorithms คือ

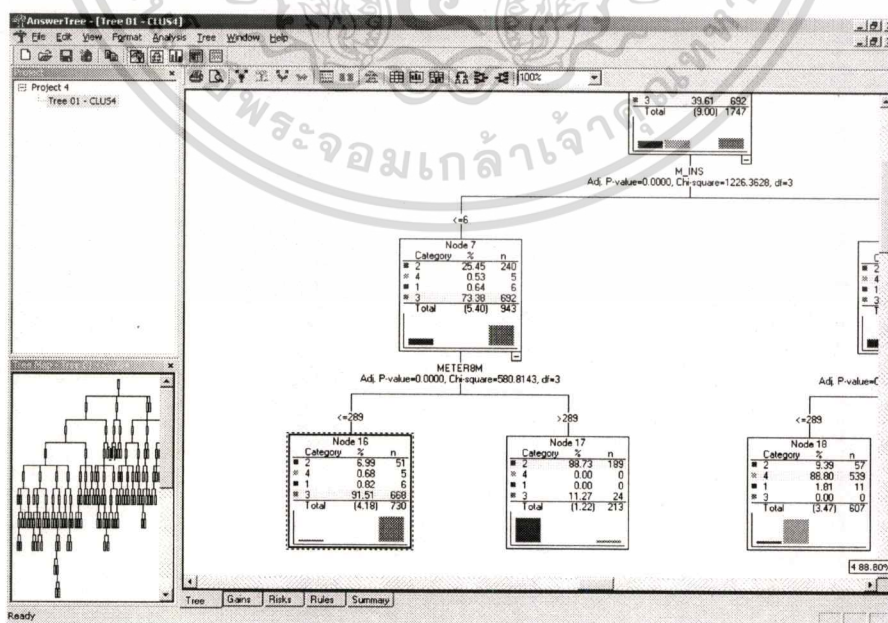
- ? CHAID (Chi-squared Automatic Interaction Detector) เป็นการหาค่าความสัมพันธ์ โดยใช้วิธีการของ Chi-squared Statistics เพื่อใช้ในการแตกกิ่งที่ดีที่สุด (Kass, 1980).

เอกสารนี้เป็นเอกสารที่สวอนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ? Exhaustive CHAID เป็นการปรับปรุงการทำงานของ CHAID เพื่อให้สามารถหาความสัมพันธ์ของชุดข้อมูล เพื่อใช้ในการแตกกิ่งอย่างละเอียดมากขึ้น (Biggs et al., 1991).
- ? CART (Classification and Regression Trees) จะสร้าง Binary Tree โดยอาศัยการคำนวณของ Ginni Index และ “towing” (Breiman et al., 1984)
- ? QUEST (Quick , Unbiased , Efficient Statistical Tree) เป็นวิธีการในการคำนวณอย่างรวดเร็ว เพื่อหลีกเลี่ยงการขัดแย้งระหว่างตัวแปรที่ใช้ในการพยากรณ์ (Loh and Shih , 1997).

ผลที่ได้จาก Algorithms ทั้ง 4 วิธีนั้นแสดงค่าที่เหมือนกันคือ การอธิบายความสัมพันธ์ของกลุ่มข้อมูล เพื่อนำมาใช้ในการพยากรณ์ (Predictive) หรือ การแยกกลุ่ม (Classification) โดย Algorithms จะเลือกค่าของตัวแปรที่ดีที่สุดในการดำเนินการ กระบวนการทำงานจะอาศัยการทำซ้ำหลายครั้งจนกว่าจะได้ Tree ที่คิดว่าเหมาะสมในการอธิบายค่าความสัมพันธ์ของกลุ่มข้อมูล แม้ว่า Algorithms ทั้ง 4 วิธีจะมีความแตกต่างทั้งกระบวนการคิด วิธีการในการได้ความสัมพันธ์มาและคุณลักษณะต่างๆ

จากการหา Pattern ของลักษณะลูกค้าแต่ละกลุ่ม โดยใช้โปรแกรม Answer Tree ซึ่งมีวิธีการทำดังภาคผนวก ก จะได้ผลลัพธ์ดังรูปที่ 4.9 แต่ผลลัพธ์ยังสามารถยุบกิ่งได้ โดยการเปลี่ยนแปลงค่าใน Node ต่างๆ และพิจารณาความเสี่ยง Error ควบคู่ไปด้วย



รูปที่ 4.9 ผลลัพธ์การค้นหา Pattern ของกลุ่มลูกค้าที่ยกเลิกการใช้บริการโทรศัพท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 การวิเคราะห์ผลลัพธ์ที่ได้

การวิเคราะห์ผลลัพธ์จากการทำค้ำไมนิ่งในการทำนายลูกค้าที่มีแนวโน้มจะยกเลิกการใช้บริการโทรศัพท์ แบ่งเป็น 2 ส่วนคือ

4.4.1 การจัดกลุ่มลูกค้าที่ยกเลิกการใช้บริการโทรศัพท์ โดยใช้ K-Means Cluster สามารถแบ่งลูกค้าเป็น 4 กลุ่ม คือ

- 1) กลุ่มลูกค้าที่มีการใช้โทรศัพท์ทางไกลสูง มีค่าเฉลี่ยเท่ากับ 0.89 จำนวน 2,401 ราย
- 2) กลุ่มลูกค้าที่มีการใช้โทรศัพท์ทางไกลภายในกรุงเทพฯ และปริมณฑลสูง มีค่าเฉลี่ยเท่ากับ 0.85 จำนวน 2,360 ราย
- 3) กลุ่มลูกค้าที่ใช้บริการตามปกติ แต่ติดตั้งโทรศัพท์ ณ เดือนกุมภาพันธ์ จำนวน 6,721 ราย
- 4) กลุ่มลูกค้าที่ใช้บริการตามปกติ แต่ติดตั้งโทรศัพท์ ณ เดือนกรกฎาคม จำนวน 5,991 ราย

4.4.2 ลักษณะของลูกค้าที่ยกเลิกการใช้บริการโทรศัพท์จากกลุ่มลูกค้าทั้ง 4 กลุ่มในข้อ

4.4.1 คือ

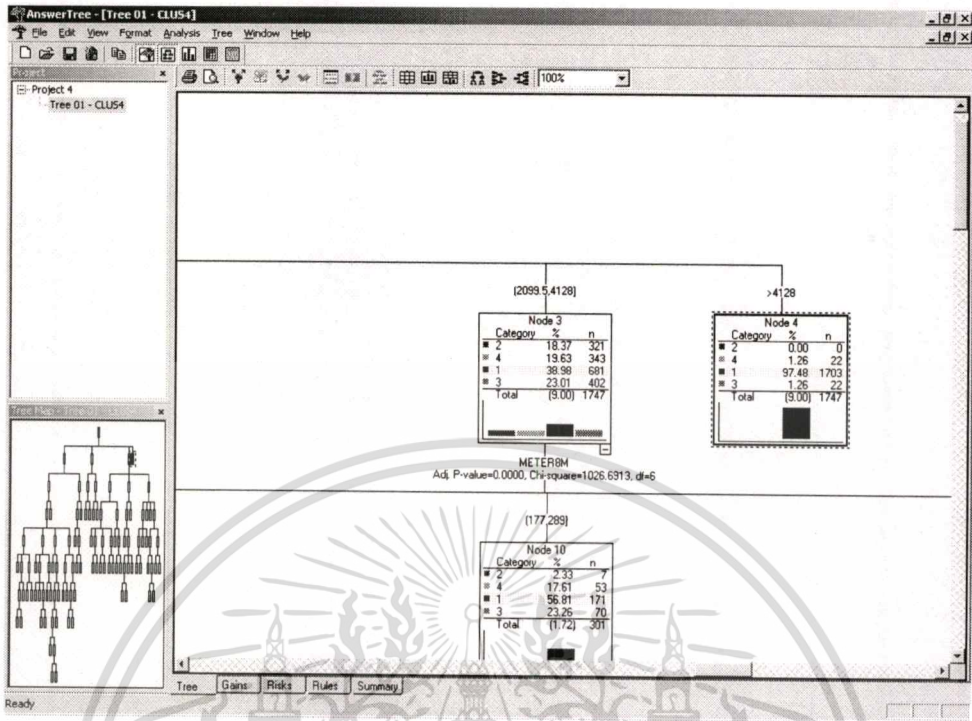
กลุ่มที่ 1 มีลักษณะดังนี้

- Node 4 ลูกค้าจำนวน 1,703 ราย (97.48 %) ที่มีการใช้โทรศัพท์ทางไกลมากกว่า 4,128 บาท ดังรูปที่ 4.10

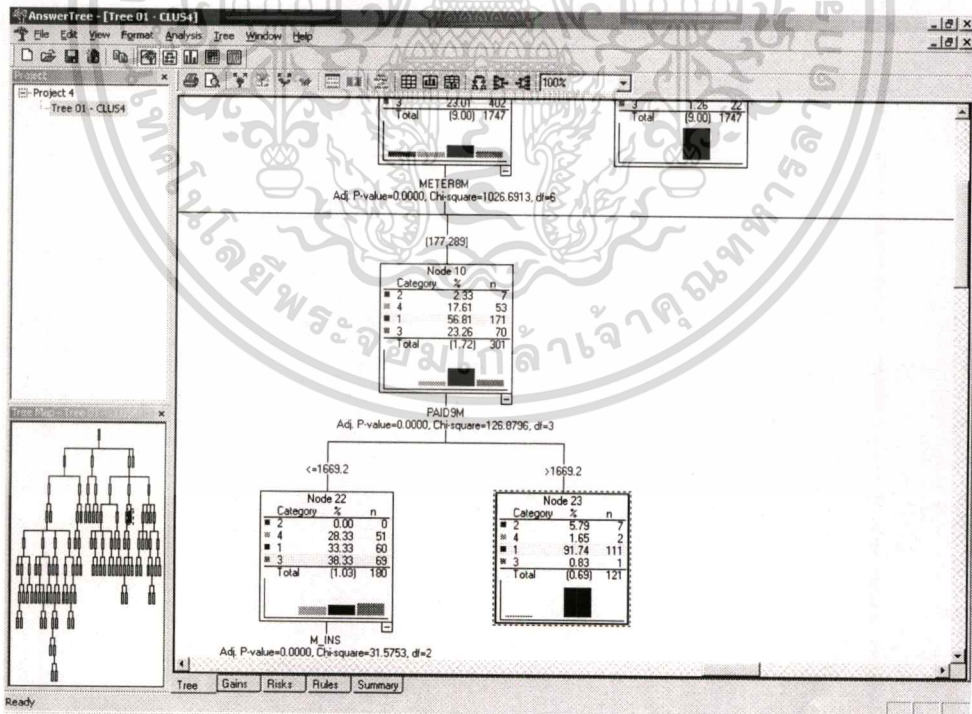
- Node 23 ลูกค้าจำนวน 111 ราย (91.74%) ที่มีการจ่ายชำระเงินค่าใช้บริการโทรศัพท์มากกว่า 1,669.20 บาท การใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลอยู่ระหว่าง 177 - 289 ครั้ง และค่าใช้โทรศัพท์ทางไกลระหว่าง 2,099.5 - 4,128 บาท ดังรูปที่ 4.11

- Node 75 ลูกค้าจำนวน 30 ราย (100%) ที่มีการติดตั้งโทรศัพท์ตั้งแต่เดือนกรกฎาคม – ตุลาคม มีการจ่ายชำระเงินค่าบริการโทรศัพท์น้อยกว่า 3,141.52 บาท การใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่า 177 ครั้ง และใช้โทรศัพท์ทางไกลระหว่าง 2,099.5 - 4,128 บาท ดังรูปที่ 4.12

สรุปลักษณะของลูกค้ากลุ่มที่ 1 เป็นลูกค้าที่มีการใช้โทรศัพท์ทางไกลมากกว่า 2,099.5 บาท และจ่ายชำระเงินมากกว่า 1,669.20 บาท ติดตั้งระหว่างเดือนกรกฎาคม – ตุลาคม

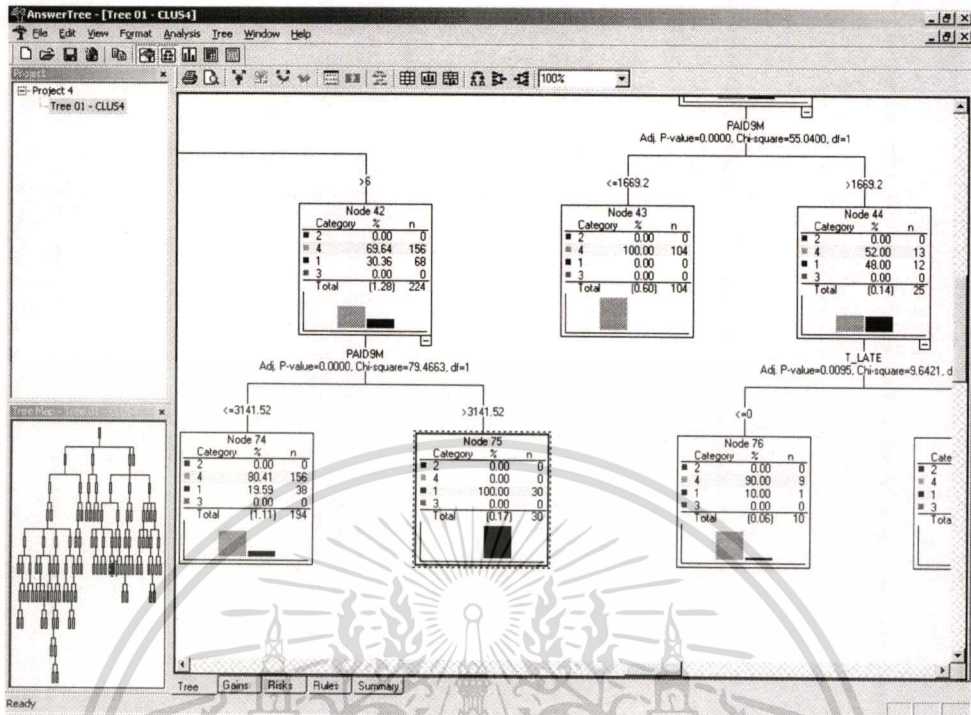


รูปที่ 4.10 Node 4 ลักษณะลูกค้ำกลุ่มที่ 1



รูปที่ 4.11 Node 23 ลักษณะลูกค้ำกลุ่มที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.12 Node 75 ลักษณะลูกค้ากลุ่มที่ 1

กลุ่มที่ 2 มีลักษณะดังนี้

- Node13 ลูกค้านำจำนวน 705 ราย (91.32 %) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลมากกว่า 289 ครั้ง ติดตั้งระหว่างเดือนมกราคม- มิถุนายน และมีการใช้โทรศัพท์ทางไกลน้อยกว่าเท่ากับ 1,188 บาท ดังรูปที่ 4.13

- Node15 ลูกค้านำจำนวน 675 ราย (91.34%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลมากกว่า 289 ครั้ง ติดตั้งระหว่างเดือนกรกฎาคม - ธันวาคม และมีการใช้โทรศัพท์ทางไกลน้อยกว่าเท่ากับ 1,188 บาท ดังรูปที่ 4.14

- Node17 ลูกค้านำจำนวน 189 ราย (88.73%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลมากกว่า 289 ครั้ง ติดตั้งระหว่างเดือนมกราคม- มิถุนายน และมีการใช้โทรศัพท์ทางไกลอยู่ระหว่าง 1,188 – 2,099.5 บาท ดังรูปที่ 4.15

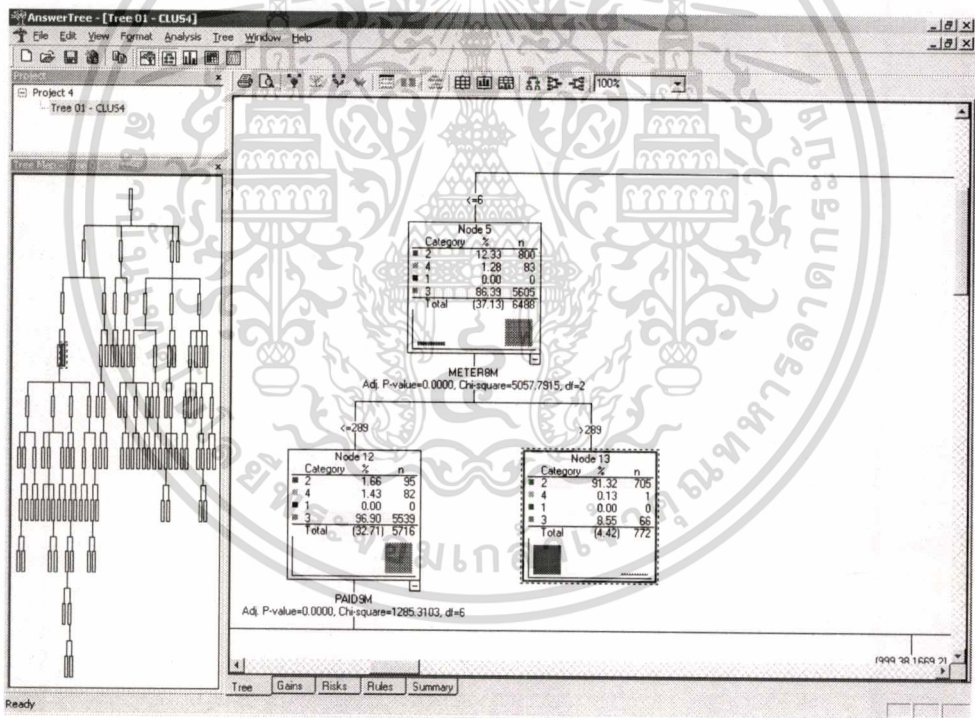
- Node19 ลูกค้านำจำนวน 185 ราย (93.91%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลมากกว่า 289 ครั้ง ติดตั้งระหว่างเดือนกรกฎาคม - ธันวาคม และมีการใช้โทรศัพท์ทางไกลอยู่ระหว่าง 1,188 – 2,099.5 บาท ดังรูปที่ 4.16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Node 90 ลูกค้ายจำนวน 24 ราย (96%)ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯ และปริมณฑลระหว่าง 177 - 289 ครั้ง ติดตั้งระหว่างเดือนพฤษภาคม - มิถุนายน มีการใช้โทรศัพท์ทางไกลอยู่น้อยกว่าเท่ากับ 1,188 บาท และจ่ายชำระเงินมากกว่า 1,669.20 บาท ดังรูปที่ 4.17

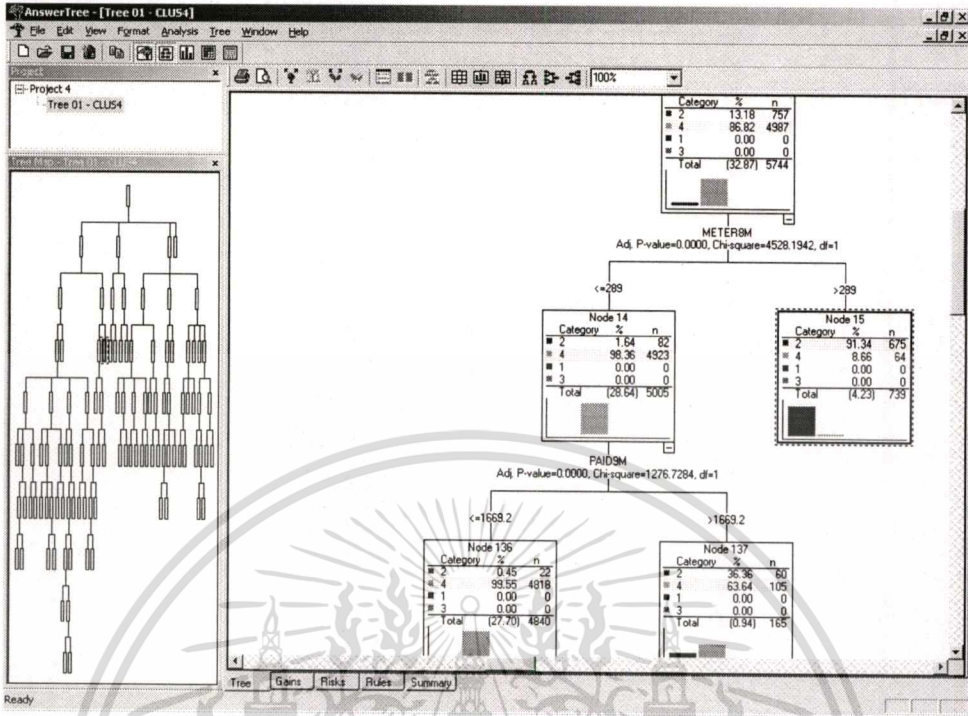
- Node 102 ลูกค้ายจำนวน 20 ราย (83.33%) มีระยะเวลาการเป็นลูกค้ายมากกว่า 13 ปี ใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลระหว่าง 177 - 289 ครั้ง ติดตั้งระหว่างเดือนมกราคม - มิถุนายน มีการใช้โทรศัพท์ทางไกลอยู่น้อยกว่าเท่ากับ 1,188 บาท และจ่ายชำระเงินมากกว่า 1,669.20 บาท ดังรูปที่ 4.18

สรุปลักษณะของลูกค้ายกลุ่มที่ 2 เป็นลูกค้ายที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯ และปริมณฑลมากกว่า 177 ครั้ง (มากกว่า 531 บาท) ใช้โทรศัพท์ทางไกลน้อยกว่า 2,099.5 บาทและจ่ายชำระค่าใช้บริการโทรศัพท์มากกว่า 1,700 บาทและมีระยะเวลาการเป็นลูกค้ายมากกว่า 13 ปี

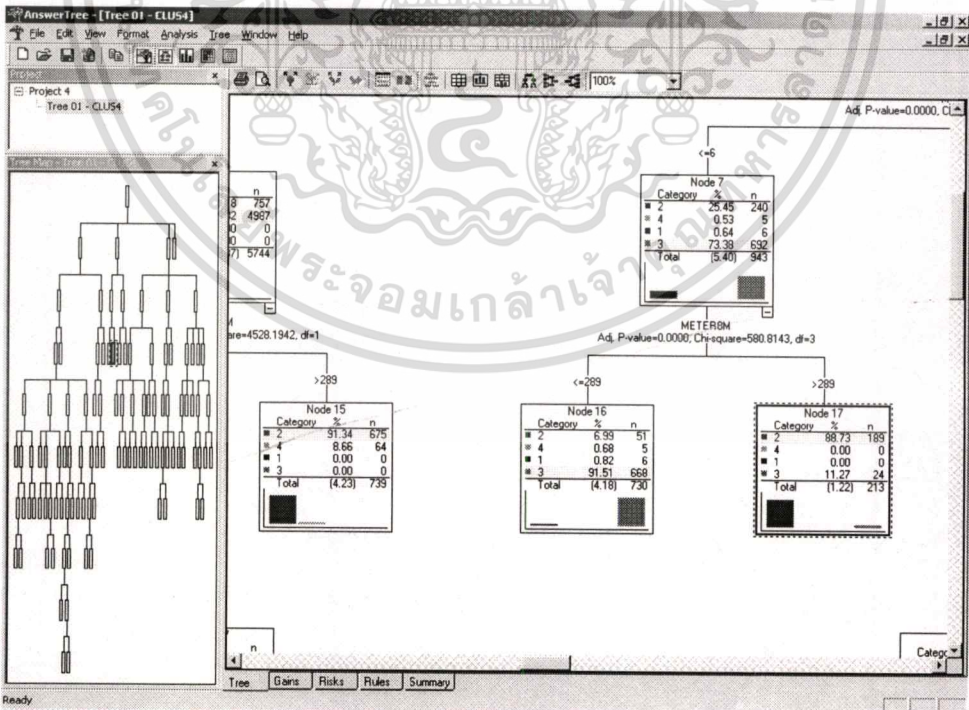


รูปที่ 4.13 Node 13 ลักษณะลูกค้ายกลุ่มที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

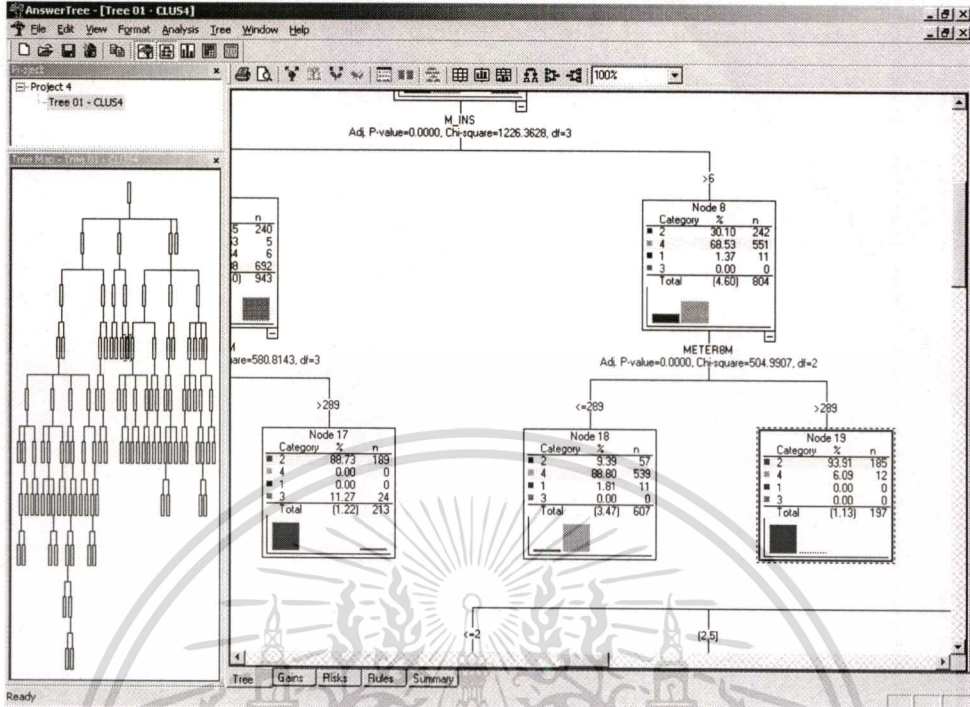


รูปที่ 4.14 Node 15 ลักษณะลูกค้ำกลุ่มที่ 2

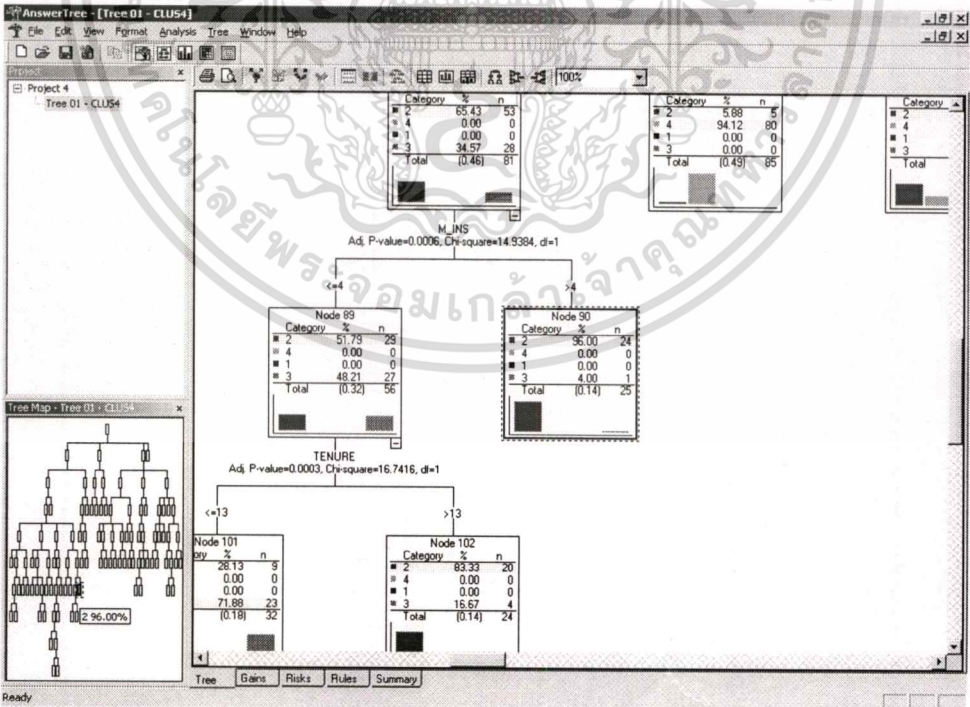


รูปที่ 4.15 Node 17 ลักษณะลูกค้ำกลุ่มที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

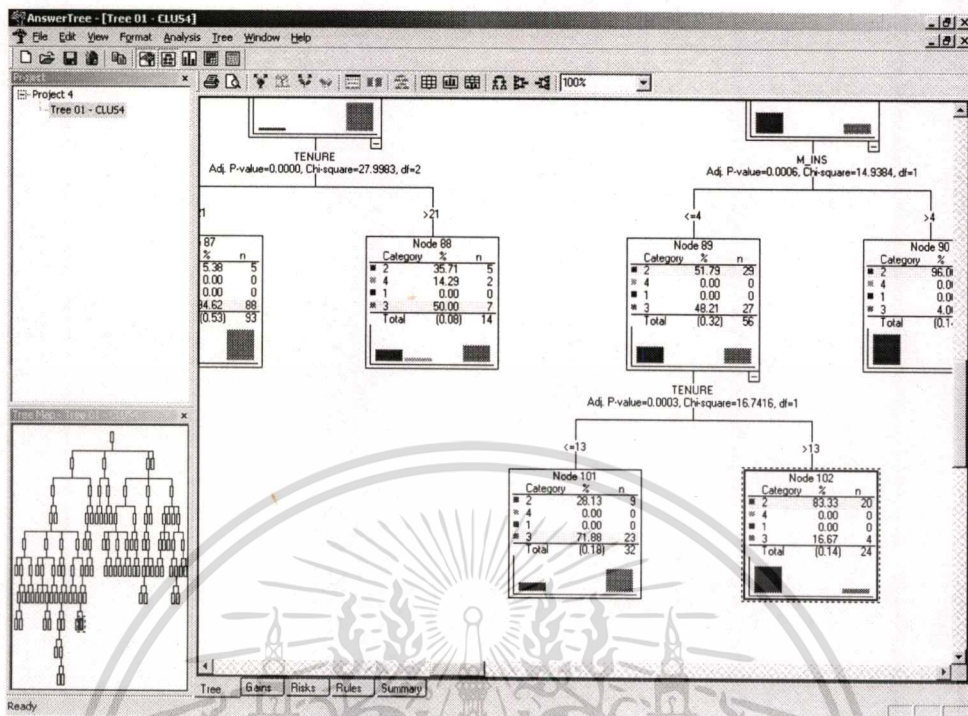


รูปที่ 4.16 Node 19 ลักษณะลูกค้ำกลุ่มที่ 2



รูปที่ 4.17 Node 90 ลักษณะลูกค้ำกลุ่มที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.18 Node 102 ลักษณะลูกค้ากลุ่มที่ 2

กลุ่มที่ 3 มีลักษณะดังนี้

- Node 16 ลูกค้าจำนวน 668 ราย (91.51%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมาณน้อยกว่าเท่ากับ 289 ครั้ง ติดตั้งระหว่างเดือนมกราคม- มิถุนายน และมีการใช้โทรศัพท์ทางไกลน้อยกว่า 1,188 บาท ดังรูปที่ 4.19

- Node 34 ลูกค้าจำนวน 146 ราย (82.49%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมาณน้อยกว่าเท่ากับ 177 ครั้ง ติดตั้งระหว่างเดือนมกราคม- ตุลาคมและมีการใช้โทรศัพท์ทางไกลระหว่าง 2,099.5 – 4,128 บาท ดังรูปที่ 4.20

- Node 40 ลูกค้าจำนวน 21 ราย (91.30%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมาณระหว่าง 177 - 289 ครั้ง ติดตั้งระหว่างเดือนมกราคม จ่ายชำระค่าบริการโทรศัพท์น้อยกว่าเท่ากับ 1,669.2 บาท และมีการใช้โทรศัพท์ทางไกลระหว่าง 2,099.5 – 4,128 บาท ดังรูปที่ 4.21

- Node 50 ลูกค้าจำนวน 836 ราย (100%) มีระยะเวลาการเป็นลูกค้าน้อยกว่าเท่ากับ 3 ปี ใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมาณน้อยกว่าเท่ากับ 289 ครั้ง ติดตั้งระหว่างเดือนมกราคม- มิถุนายน จ่ายชำระค่าบริการโทรศัพท์ระหว่าง 0 – 999.38 บาทและมีการใช้โทรศัพท์ทางไกลน้อยกว่า 1,188 บาท ดังรูปที่ 4.22

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Node 53 ลูกค้ายจำนวน 474 ราย (98.54%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมาณน้อยกว่าเท่ากับ 289 ครั้ง จ่ายชำระค่าบริการโทรศัพท์ระหว่าง 999.38 – 1,669.2 บาทติดตั้งระหว่างเดือนมกราคม- มิถุนายน และมีการใช้โทรศัพท์ทางไกลน้อยกว่า 1,188 บาท ดังรูปที่ 4.23

- Node 59 ลูกค้ายจำนวน 113 ราย (100%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมาณน้อยกว่าเท่ากับ 177 ครั้ง จ่ายชำระค่าบริการโทรศัพท์น้อยกว่าเท่ากับ 999.38 บาทติดตั้งระหว่างเดือนกุมภาพันธ์ - ตุลาคม และมีการใช้โทรศัพท์ทางไกลระหว่าง 2,099.5 – 4,128 บาท ดังรูปที่ 4.24

- Node 77 ลูกค้ายจำนวน 352 ราย (100%) เป็นลูกค้ายตามบ้านพักอาศัย มีระยะเวลาการเป็นลูกค้าย 13 – 21 ปี ใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมาณน้อยกว่าเท่ากับ 289 ครั้ง ติดตั้งระหว่างเดือนมกราคม- มิถุนายน และมีการใช้โทรศัพท์ทางไกลน้อยกว่า 1,188 บาท ดังรูปที่ 4.25

- Node 78 ลูกค้ายจำนวน 146 ราย (99.32%) มีระยะเวลาการเป็นลูกค้ายมากกว่า 21 ปี ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมาณน้อยกว่าเท่ากับ 289 ครั้ง ติดตั้งระหว่างเดือนมกราคม- พฤษภาคม และมีการใช้โทรศัพท์ทางไกลน้อยกว่า 1,188 บาท ดังรูปที่ 4.26

- Node 80 ลูกค้ายจำนวน 240 ราย (98.36) ประเภทธุรกิจที่มีระยะเวลาการเป็นลูกค้าย 3- 21 ปี ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมาณน้อยกว่าเท่ากับ 289 ครั้ง ติดตั้งระหว่างเดือนมกราคม- มิถุนายน จ่ายชำระค่าบริการโทรศัพท์ระหว่าง 0 – 999.38 บาทและมีการใช้โทรศัพท์ทางไกลระหว่าง 1,188 – 2,099.5 บาท ดังรูปที่ 4.27

- Node 81 ลูกค้ายจำนวน 773 ราย (100%) ประเภทบ้านพักที่มีระยะเวลาการเป็นลูกค้าย 3- 21 ปี ติดตั้งระหว่างเดือนมกราคม- มิถุนายน ใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมาณน้อยกว่าเท่ากับ 289 ครั้ง จ่ายชำระค่าบริการโทรศัพท์ระหว่าง 0 – 999.38 บาทและมีการใช้โทรศัพท์ทางไกลน้อยกว่าเท่ากับ 1,188 บาทดังรูปที่ 4.28

- Node 83 ลูกค้ายจำนวน 173 ราย (99.43%) เป็นลูกค้ายมากกว่า 21 ปี ติดตั้งระหว่างเดือนมกราคม- มิถุนายน ใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมาณน้อยกว่าเท่ากับ 289 ครั้ง จ่ายชำระค่าบริการโทรศัพท์ระหว่าง 0 – 999.38 บาท และมีการใช้โทรศัพท์ทางไกลน้อยกว่าเท่ากับ 1,188 บาท ดังรูปที่ 4.29

- Node 87 ลูกค้ายจำนวน 88 ราย (94.62%) เป็นลูกค้ายน้อยกว่าเท่ากับ 21 ปี ติดตั้งระหว่างเดือนมกราคม- มิถุนายน ใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมาณน้อยกว่าเท่ากับ

289 ครั้ง จ่ายชำระค่าบริการโทรศัพท์มากกว่า 1,669.2 บาท และมีการใช้โทรศัพท์ทางไกลอยู่น้อยกว่าเท่ากับ 1,188 บาท ดังรูปที่ 4.30

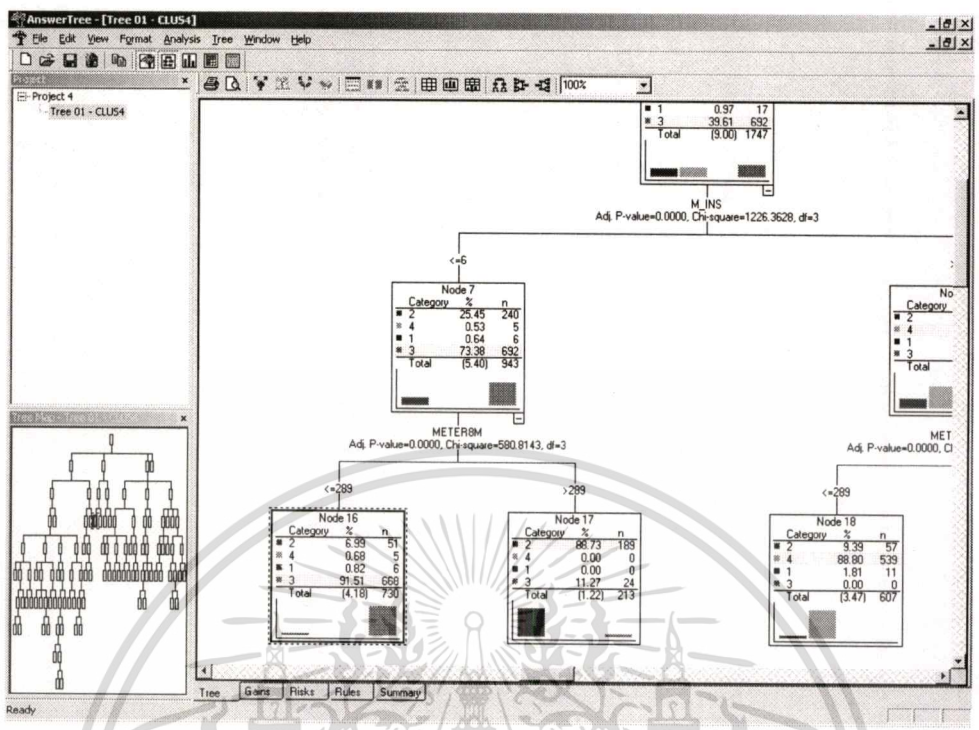
- Node 95 ลูกค้ายจำนวน 43 ราย (100%) ประเภทธุรกิจและรัฐวิสาหกิจ ที่มีระยะเวลาการเป็นลูกค้า 3- 21 ปี ติดตั้งระหว่างเดือนมกราคม- มิถุนายน ใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่าเท่ากับ 289 ครั้ง และมีการใช้โทรศัพท์ทางไกลน้อยกว่าเท่ากับ 1,188 บาทดังรูปที่ 4.31

- Node 97 ลูกค้ายจำนวน 47 ราย (100%) ประเภทรัฐบาลและรัฐวิสาหกิจ ที่มีระยะเวลาการเป็นลูกค้า 3- 21 ปี ติดตั้งระหว่างเดือนมกราคม- มิถุนายน ใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่าเท่ากับ 289 ครั้ง จ่ายชำระค่าบริการโทรศัพท์ระหว่าง 0 – 999.38 บาท และมีการใช้โทรศัพท์ทางไกลน้อยกว่าเท่ากับ 1,188 บาท ดังรูปที่ 4.32

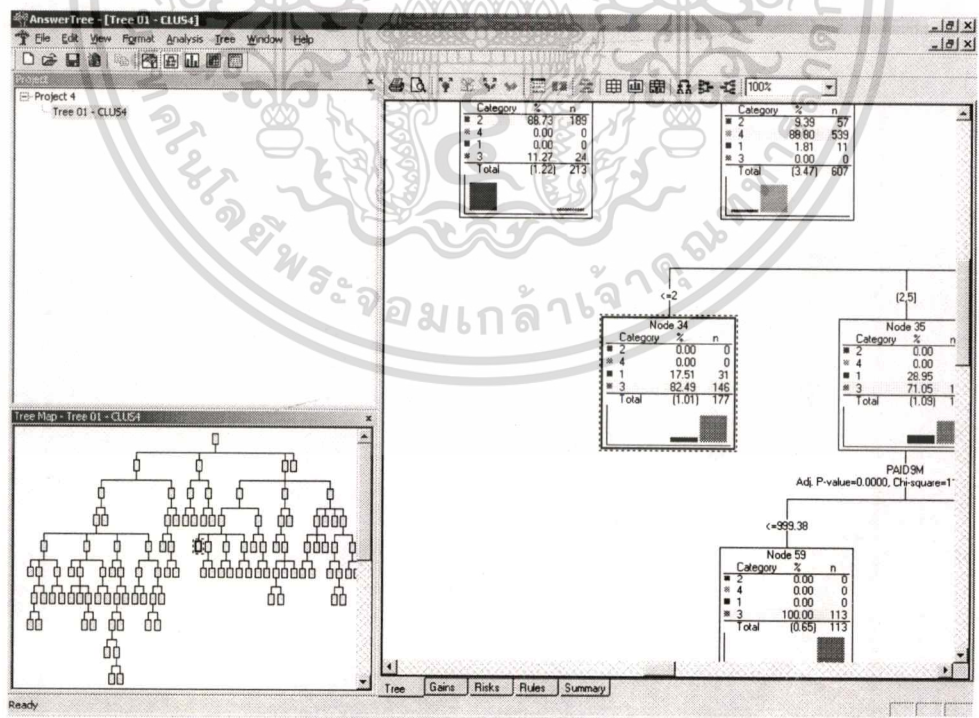
- Node 103 ลูกค้ายจำนวน 45 ราย (100%) ที่มีระยะเวลาการเป็นลูกค้าน้อยกว่า 21 ปี ติดตั้งระหว่างเดือนมกราคม- มิถุนายน ใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่าเท่ากับ 289 ครั้ง จ่ายชำระค่าบริการโทรศัพท์ระหว่าง 999.38 – 1,669.2 บาท โดยชำระเงินซ้ำ 2 ครั้ง และมีการใช้โทรศัพท์ทางไกลน้อยกว่าเท่ากับ 1,188 บาท ดังรูปที่ 4.33

- Node 107 ลูกค้ายจำนวน 24 ราย (100%) ประเภทบ้านพักอาศัย ติดตั้งโทรศัพท์ตั้งแต่เดือนมีนาคม จ่ายชำระค่าบริการโทรศัพท์ระหว่าง 999.38 – 1,669.2 บาท จ่ายชำระค่าบริการโทรศัพท์ระหว่าง 999.38 – 1,669.2 บาท ใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่าเท่ากับ 289 ครั้ง และมีการใช้โทรศัพท์ทางไกลน้อยกว่าเท่ากับ 1,188 บาท ดังรูปที่ 4.34

สรุปลักษณะของลูกค้ากลุ่มที่ 3 ประเภทบ้านพัก ธุรกิจ รัฐบาล รัฐวิสาหกิจ ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่า 289 ครั้ง (น้อยกว่า 867 บาท) ใช้โทรศัพท์ทางไกลน้อยกว่า 2,099.5 บาทและจ่ายชำระค่าใช้บริการโทรศัพท์น้อยกว่า 1,669.20 บาทและมีจ่ายชำระค่าใช้บริการซ้ำน้อยกว่า 2 ครั้ง

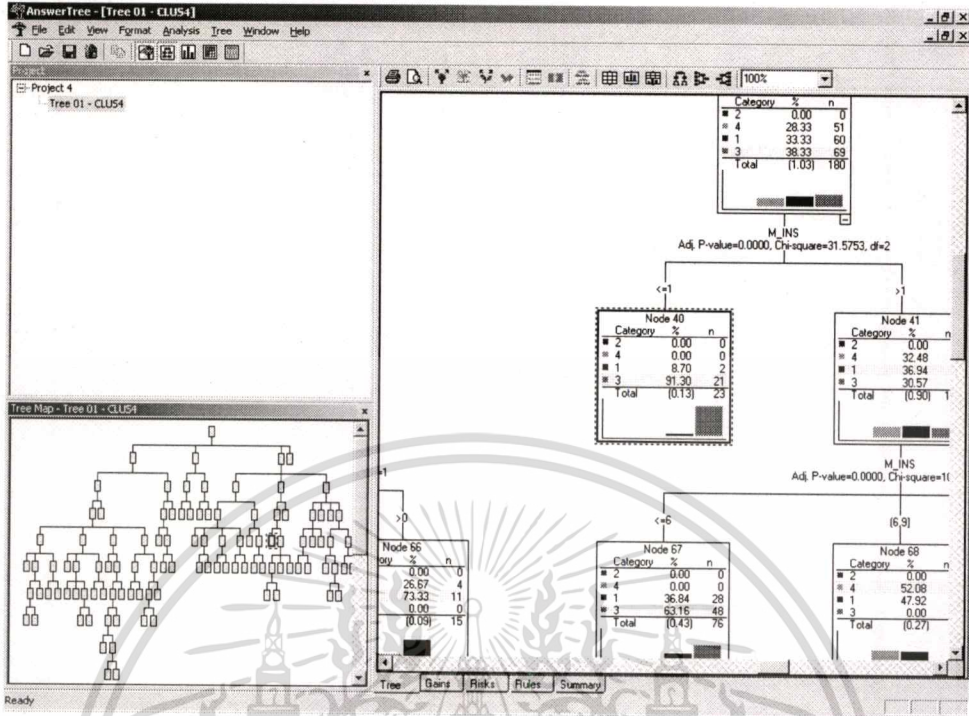


รูปที่ 4.19 Node 16 ลักษณะลูกค้ำกลุ่มที่ 3

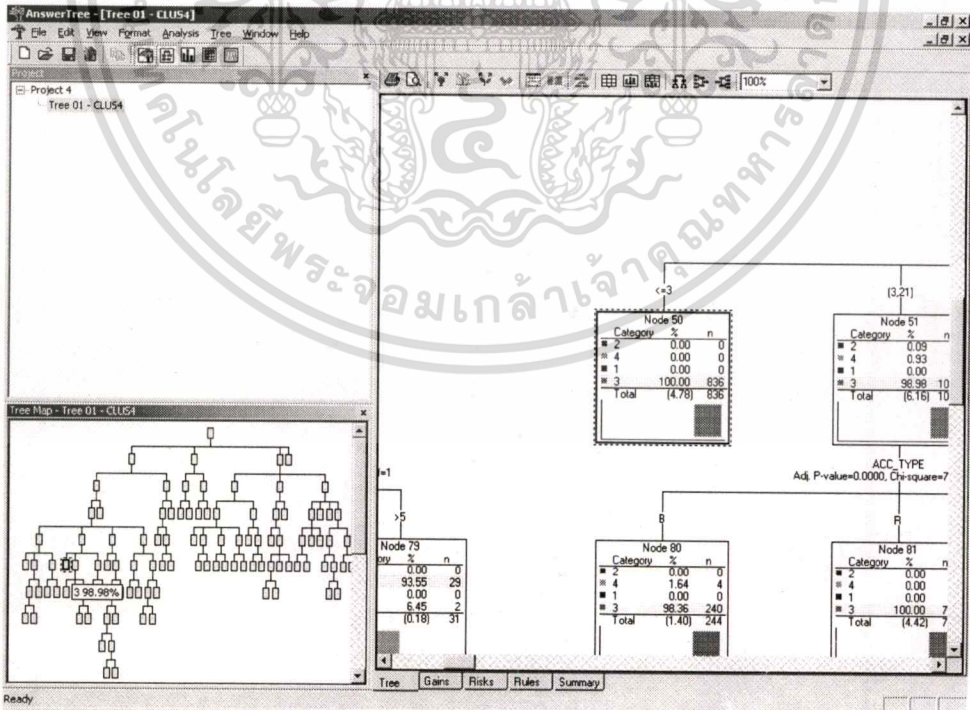


รูปที่ 4.20 Node 34 ลักษณะลูกค้ำกลุ่มที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

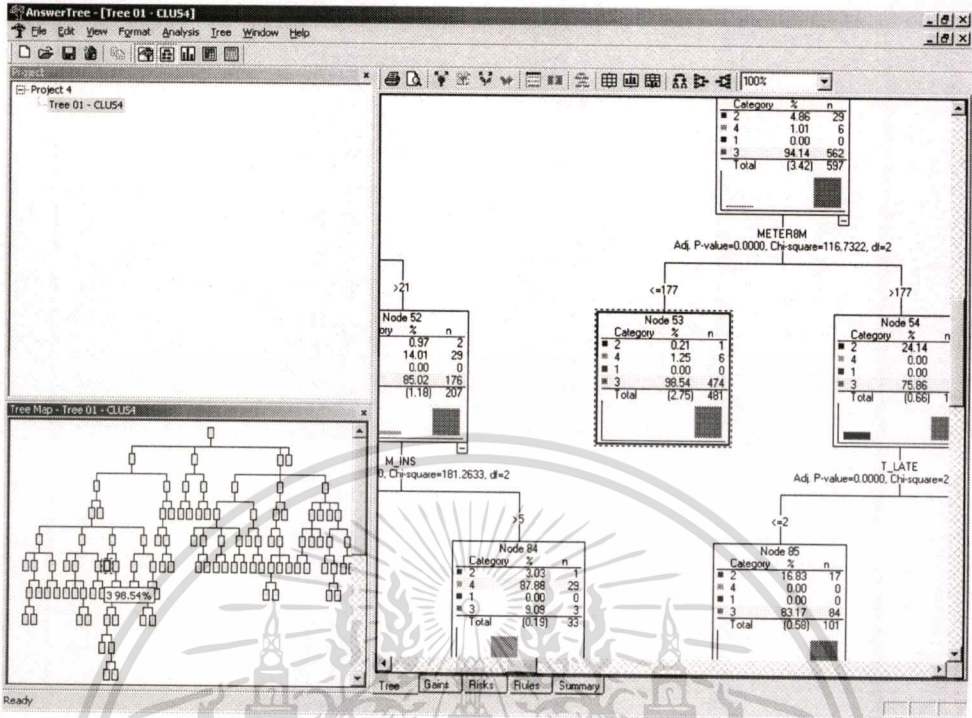


รูปที่ 4.21 Node 40 ลักษณะลูกค้ำกลุ่มที่ 3

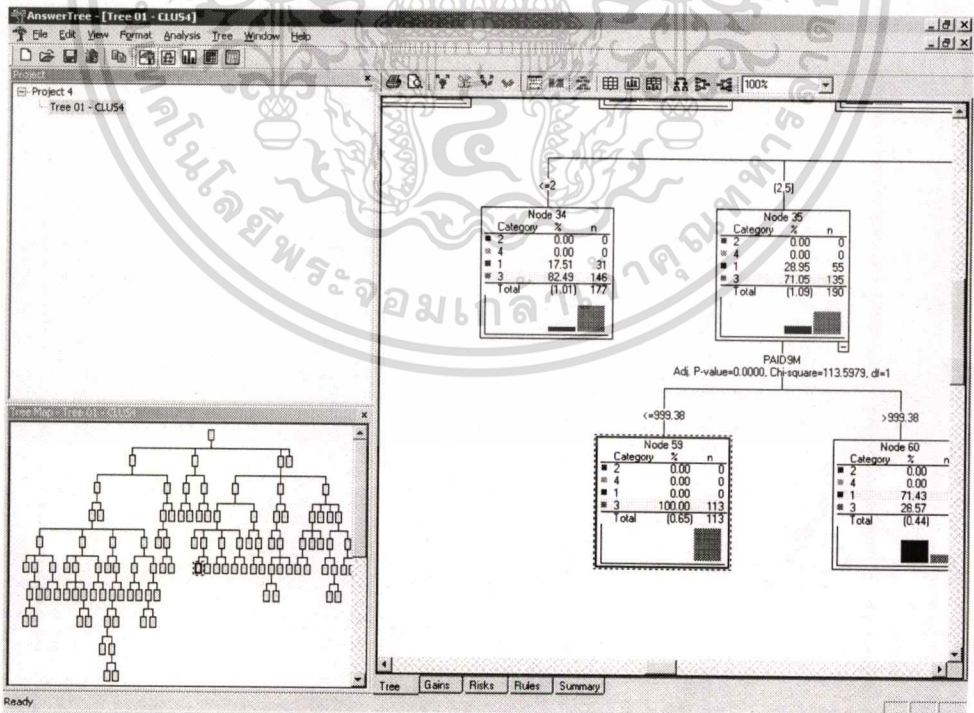


รูปที่ 4.22 Node 50 ลักษณะลูกค้ำกลุ่มที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

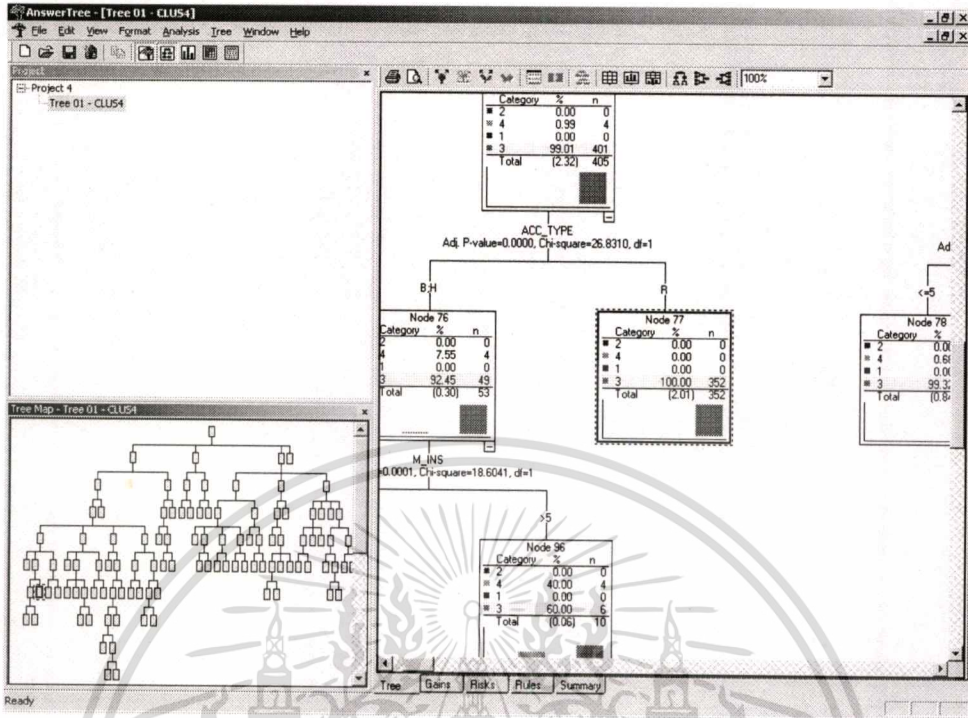


รูปที่ 4.23 Node 53 ลักษณะลูกค้ำกลุ่มที่ 3

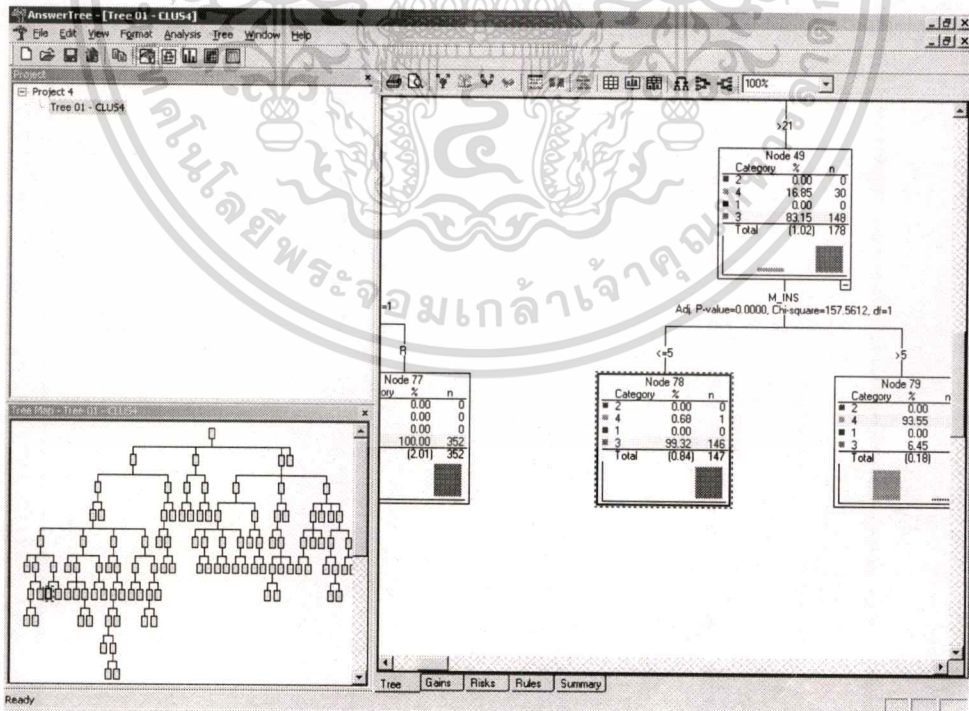


รูปที่ 4.24 Node 59 ลักษณะลูกค้ำกลุ่มที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

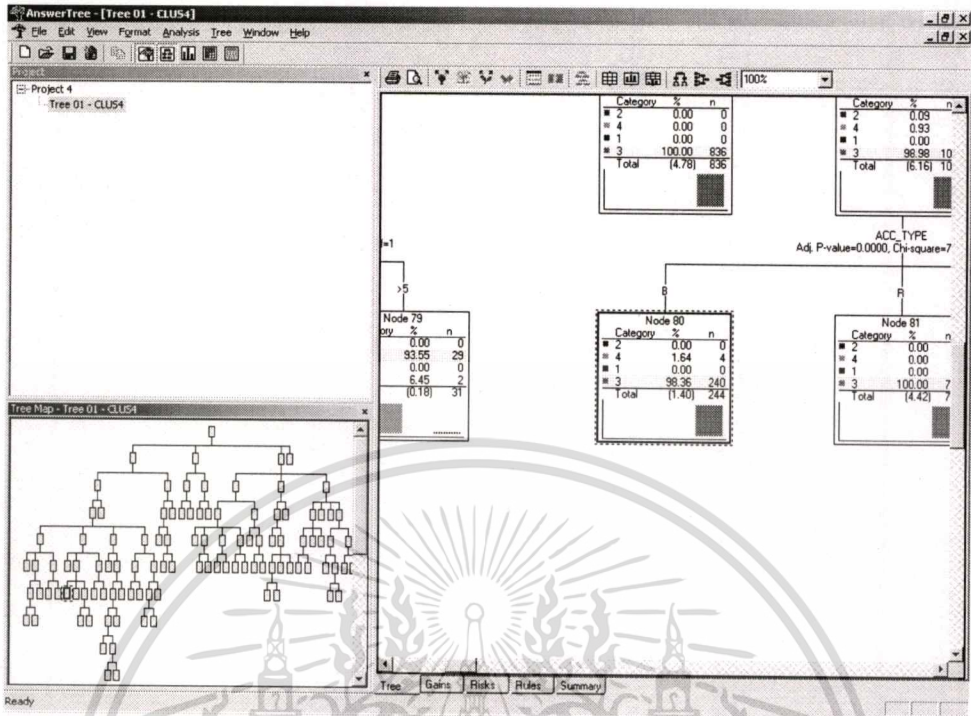


รูปที่ 4.25 Node 77 ลักษณะลูกค้ำกลุ่มที่ 3

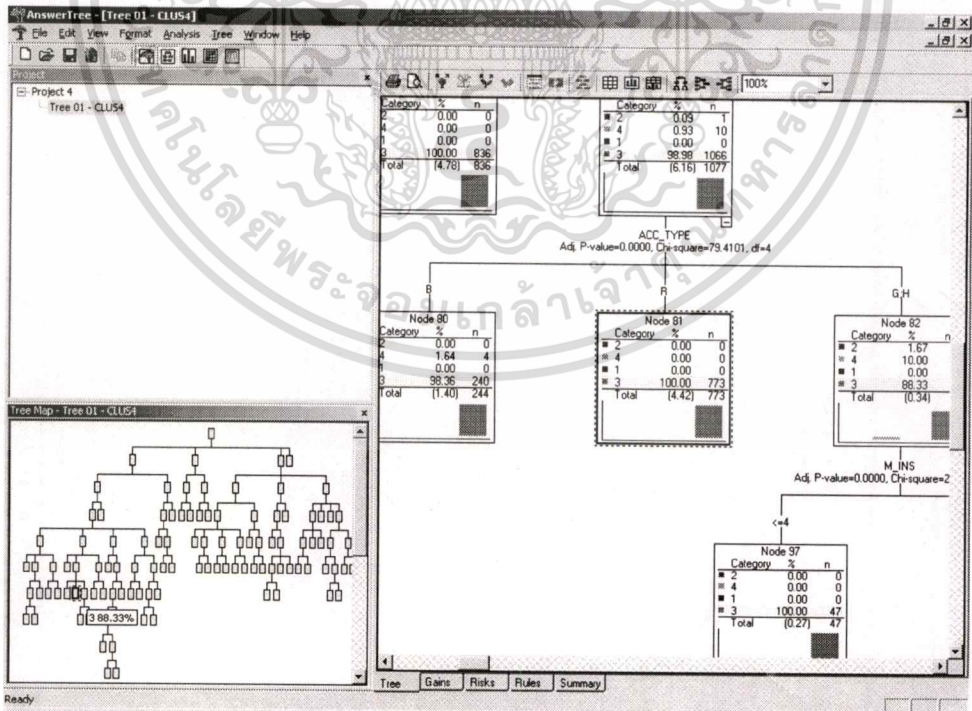


รูปที่ 4.26 Node 78 ลักษณะลูกค้ำกลุ่มที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

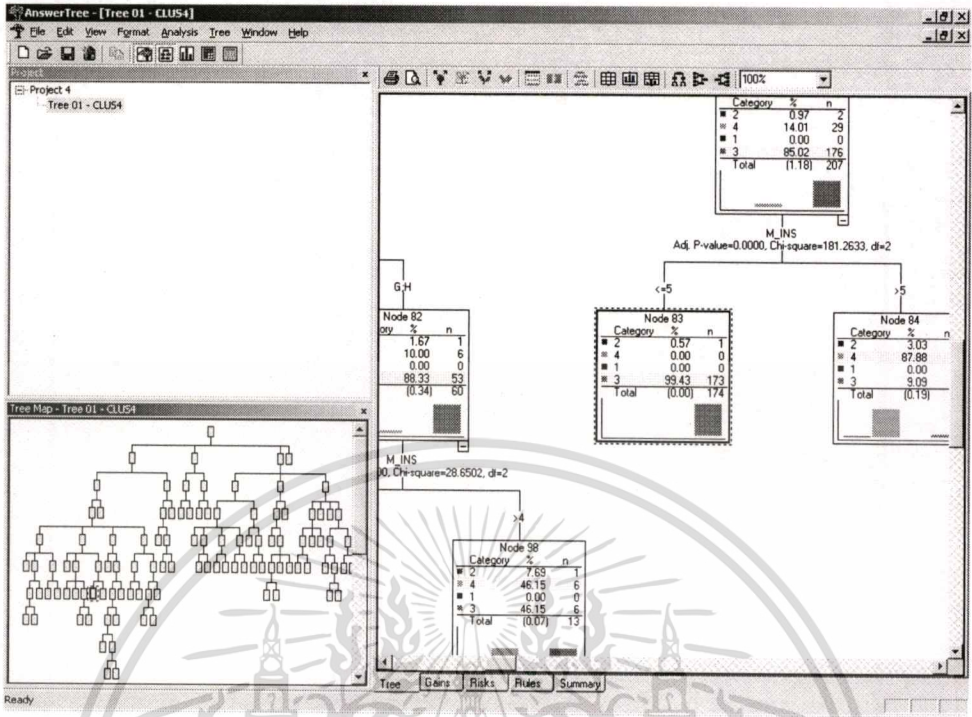


รูปที่ 4.27 Node 80 ลักษณะลูกค้ำกลุ่มที่ 3

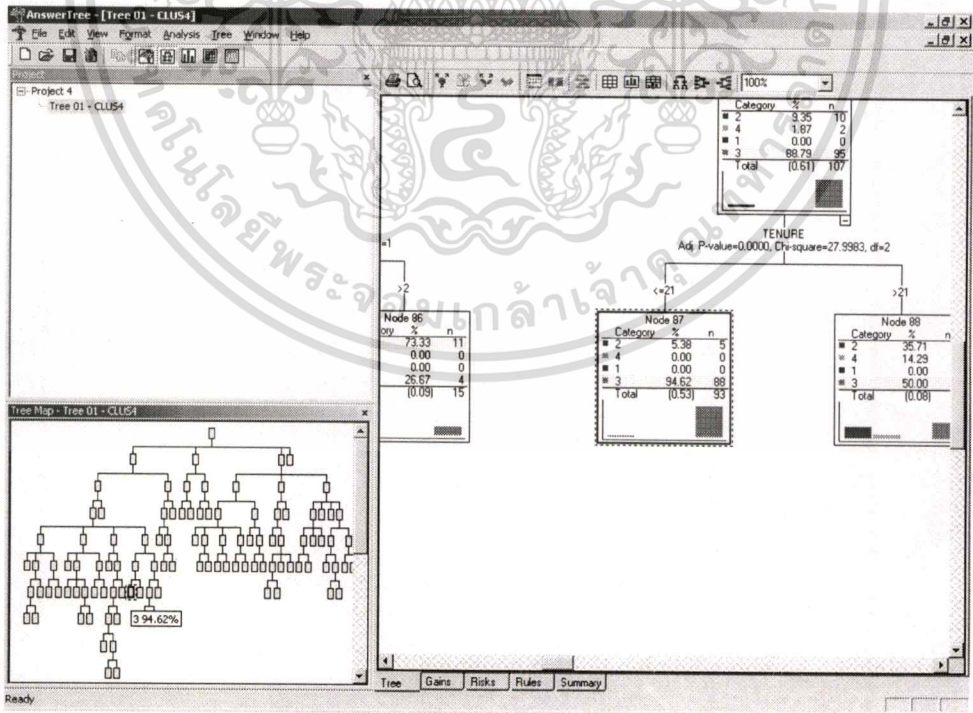


รูปที่ 4.28 Node 81 ลักษณะลูกค้ำกลุ่มที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

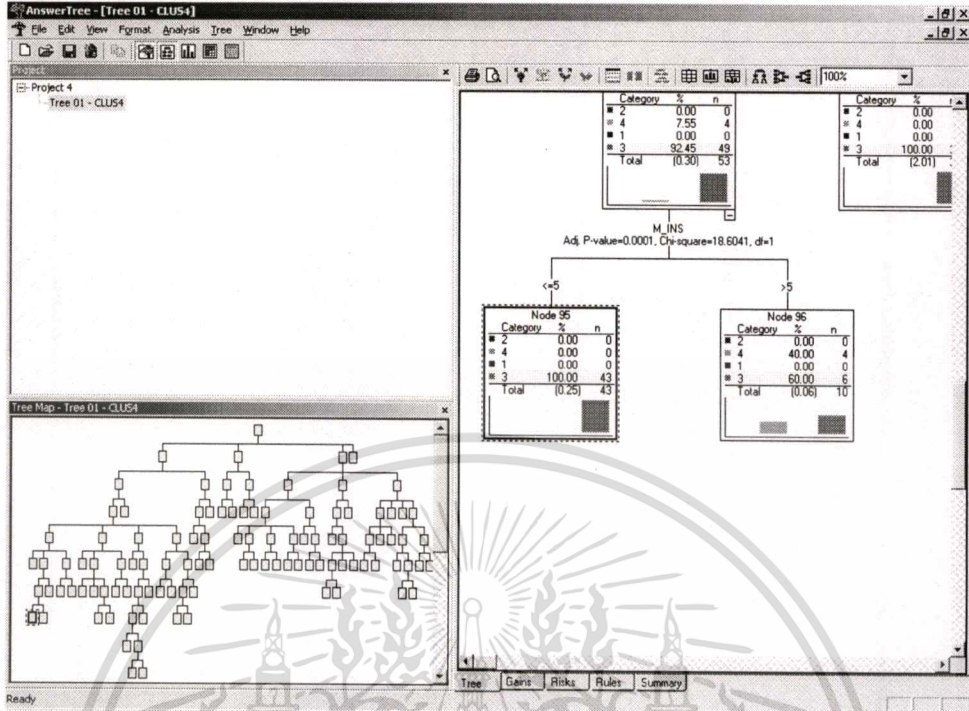


รูปที่ 4.29 Node 83 ลักษณะลูกค้ำกลุ่มที่ 3

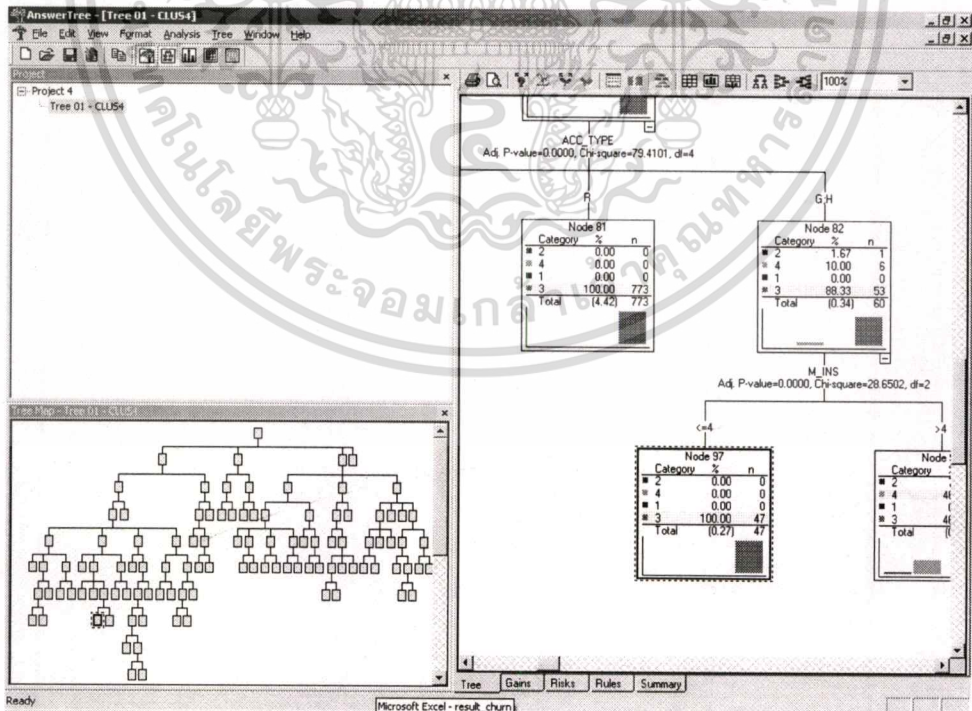


รูปที่ 4.30 Node 87 ลักษณะลูกค้ำกลุ่มที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

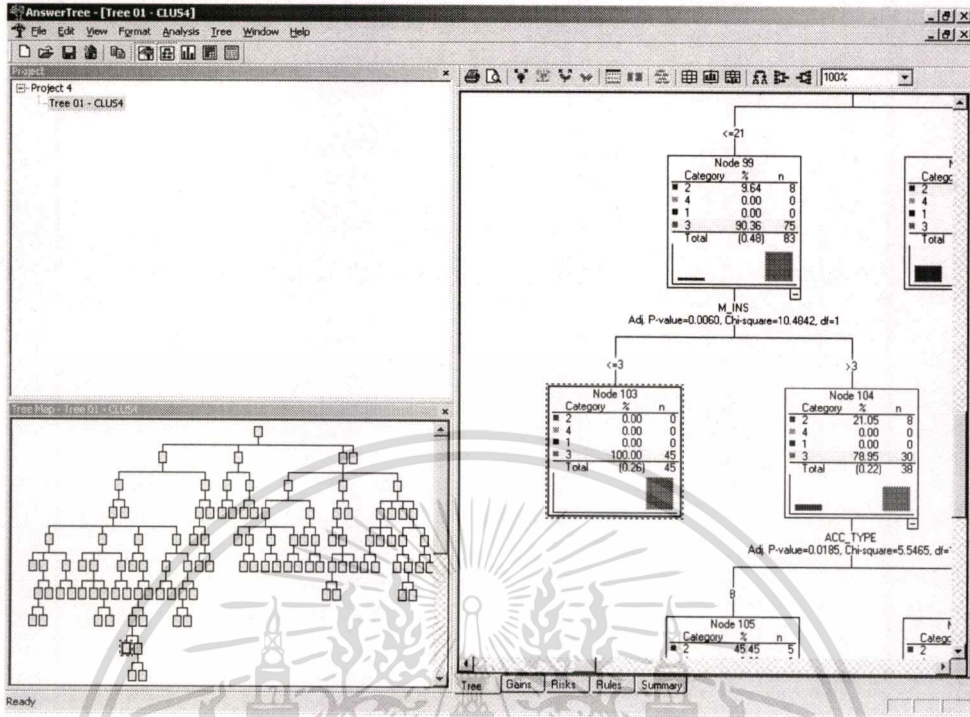


รูปที่ 4.31 Node 95 ลักษณะลูกค่ากลุ่มที่ 3

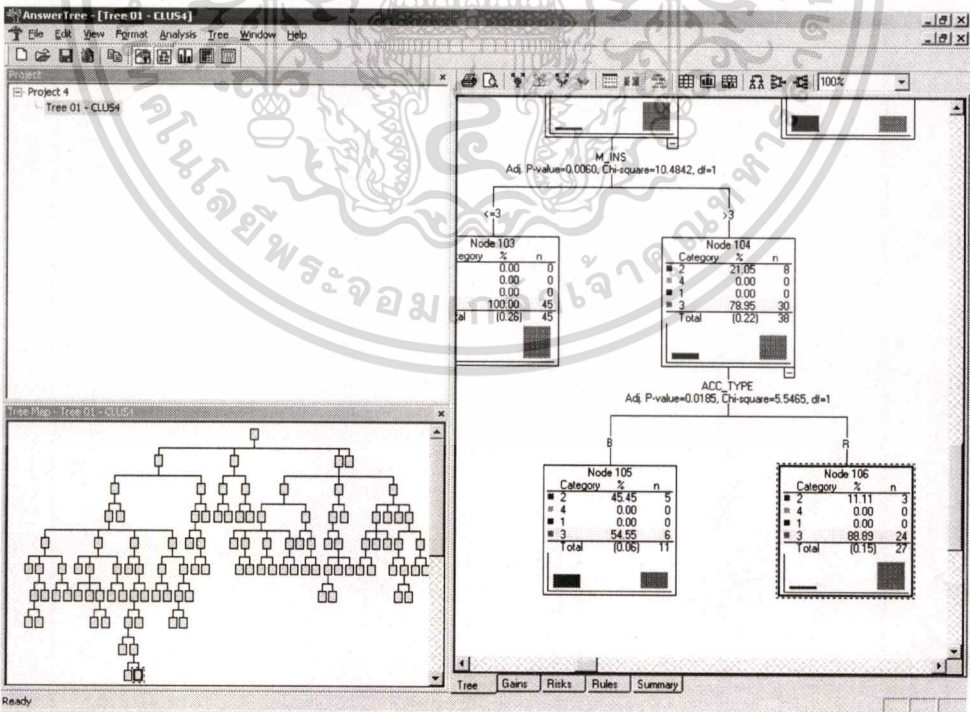


รูปที่ 4.32 Node 97 ลักษณะลูกค่ากลุ่มที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.33 Node 103 ลักษณะลูกค้ำกลุ่มที่ 3



รูปที่ 4.34 Node 106 ลักษณะลูกค้ำกลุ่มที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลุ่มที่ 4 มีลักษณะดังนี้

- Node 18 ลูกค้าย่อยจำนวน 539 ราย (88.8%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่า 289 ครั้ง ติดตั้งระหว่างเดือนกรกฎาคม - ธันวาคม และมีการใช้โทรศัพท์ทางไกลระหว่าง 1,188 – 2,099.5 บาทดังรูปที่ 4.35

- Node 32 ลูกค้าย่อยจำนวน 4,818 ราย (99.15%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่า 289 ครั้ง ระหว่างเดือนกรกฎาคม - ธันวาคม และมีการใช้โทรศัพท์ทางไกลน้อยกว่า 1,188 บาทดังรูปที่ 4.36

- Node 38 ลูกค้าย่อยจำนวน 104 ราย (100%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่า 177 ครั้ง ติดตั้งระหว่างเดือนพฤศจิกายน - ธันวาคม จ่ายชำระค่าใช้บริการโทรศัพท์น้อยกว่า 1,669.2 บาท และมีการใช้โทรศัพท์ทางไกลอยู่ระหว่าง 2,099.5 – 4,128 บาทดังรูปที่ 4.37

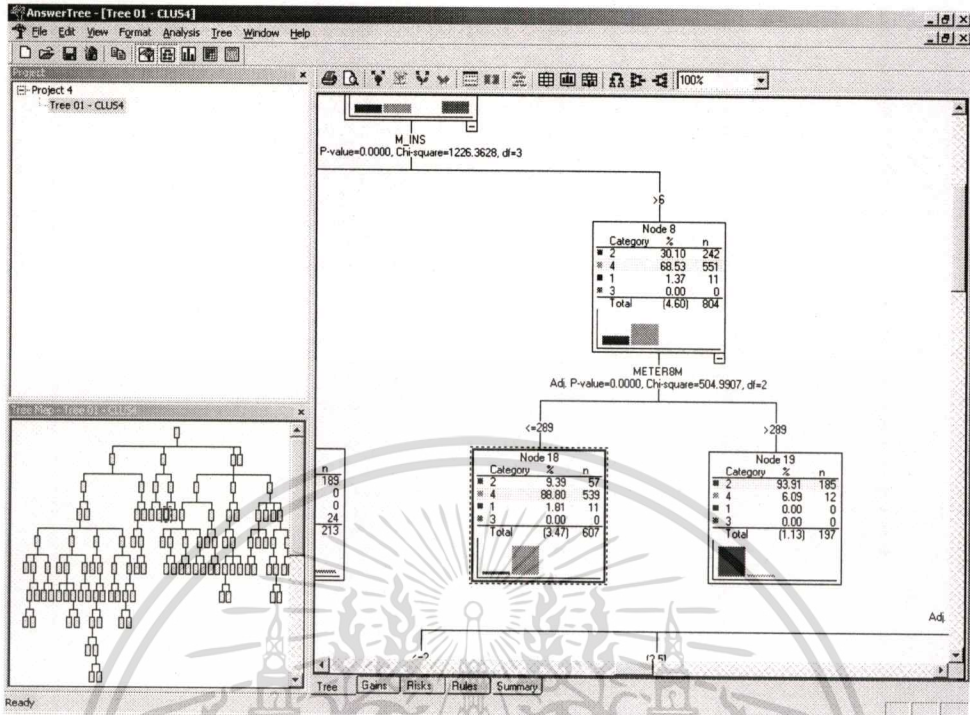
- Node 63 ลูกค้าย่อยจำนวน 156 ราย (100%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่า 177 ครั้ง ติดตั้งระหว่างเดือนกรกฎาคม - ตุลาคม จ่ายชำระค่าใช้บริการโทรศัพท์น้อยกว่า 3,141.52 บาท และมีการใช้โทรศัพท์ทางไกลอยู่ระหว่าง 2,099.5 – 4,128 บาท ดังรูปที่ 4.38

- Node 65 ลูกค้าย่อยจำนวน 9 ราย (90%) ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่า 177 ครั้ง ติดตั้งระหว่างเดือนตุลาคม - ธันวาคม จ่ายชำระค่าใช้บริการโทรศัพท์มากกว่า 1,669.2 บาท และมีการใช้โทรศัพท์ทางไกลระหว่าง 2,099.5 – 4,128 บาทดังรูปที่ 4.39

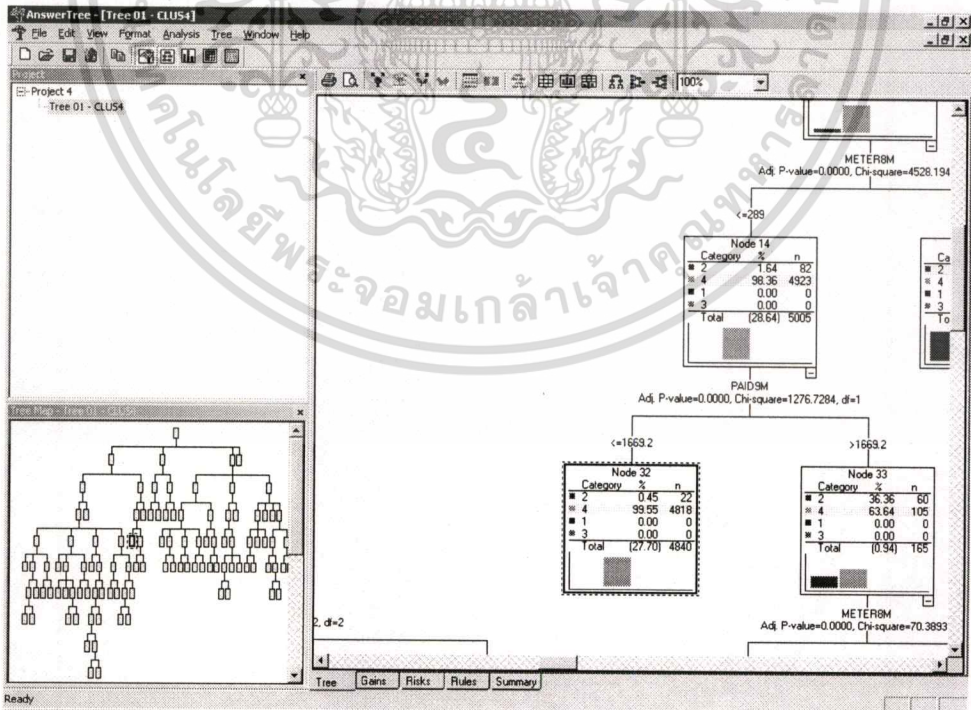
- Node 79 ลูกค้าย่อยจำนวน 29 ราย (93.55%) ระยะเวลาการเป็นลูกค้าย่อยมากกว่า 21 ปี ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่า 289 ครั้ง ติดตั้งเดือนมิถุนายน และมีการใช้โทรศัพท์ทางไกลน้อยกว่าเท่ากับ 1,188 บาทดังรูปที่ 4.39

- Node 84 ลูกค้าย่อยจำนวน 29 ราย (87.88%) ระยะเวลาการเป็นลูกค้าย่อยมากกว่า 21 ปี ที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่า 289 ครั้ง ติดตั้งเดือนมิถุนายน และมีการใช้โทรศัพท์ทางไกลน้อยกว่าเท่ากับ 1,188 บาท ดังรูปที่ 4.40

สรุปลักษณะของลูกค้าย่อยกลุ่มที่ 4 เป็นลูกค้าย่อยที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่า 289 ครั้ง (น้อยกว่า 867 บาท) ใช้โทรศัพท์ทางไกลมากกว่า 4,128 บาท และจ่ายชำระค่าใช้บริการโทรศัพท์ระหว่าง 0 - 3141.52 บาท ติดตั้งโทรศัพท์ภายในเดือนมิถุนายน - ธันวาคม ระยะเวลาการเป็นลูกค้าย่อยมากกว่า 21 ปี

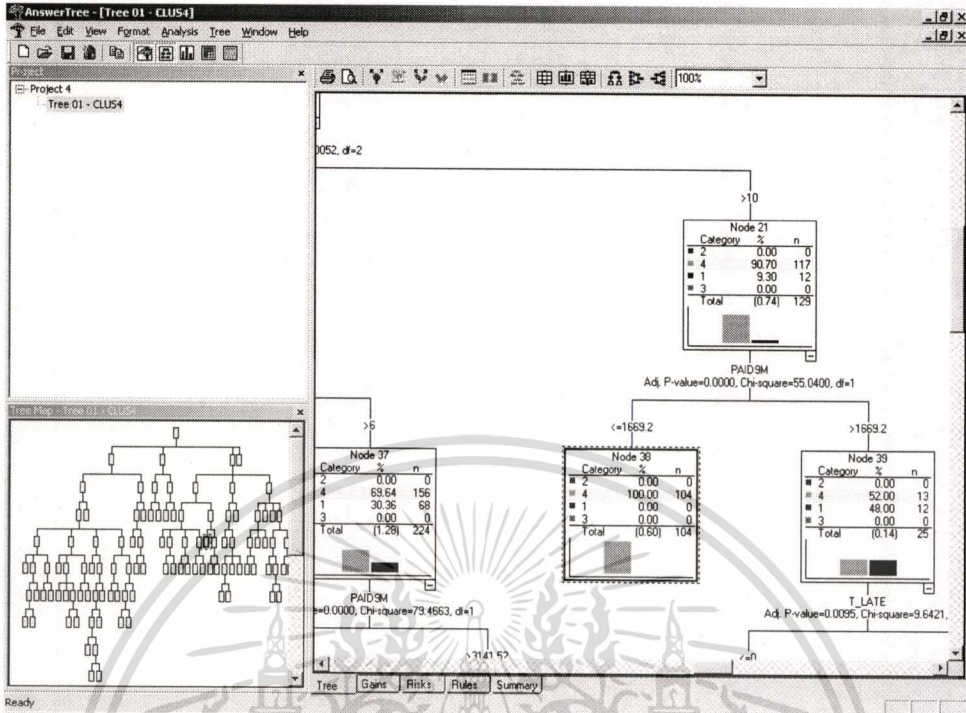


รูปที่ 4.35 Node 18 ลักษณะลูกค้ำกลุ่มที่ 4

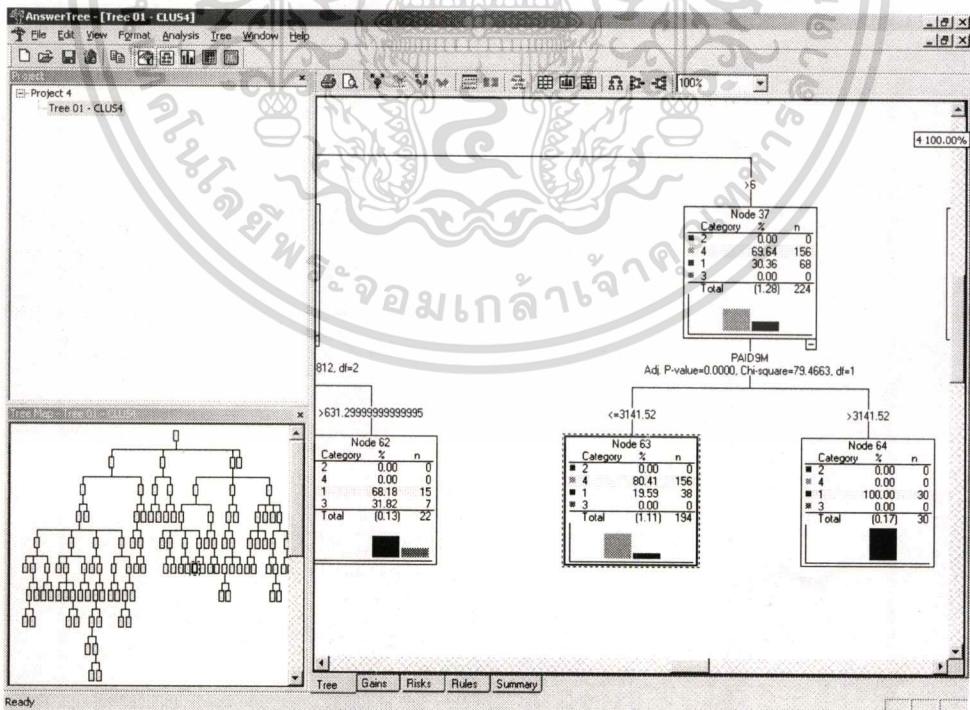


รูปที่ 4.36 Node 32 ลักษณะลูกค้ำกลุ่มที่ 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

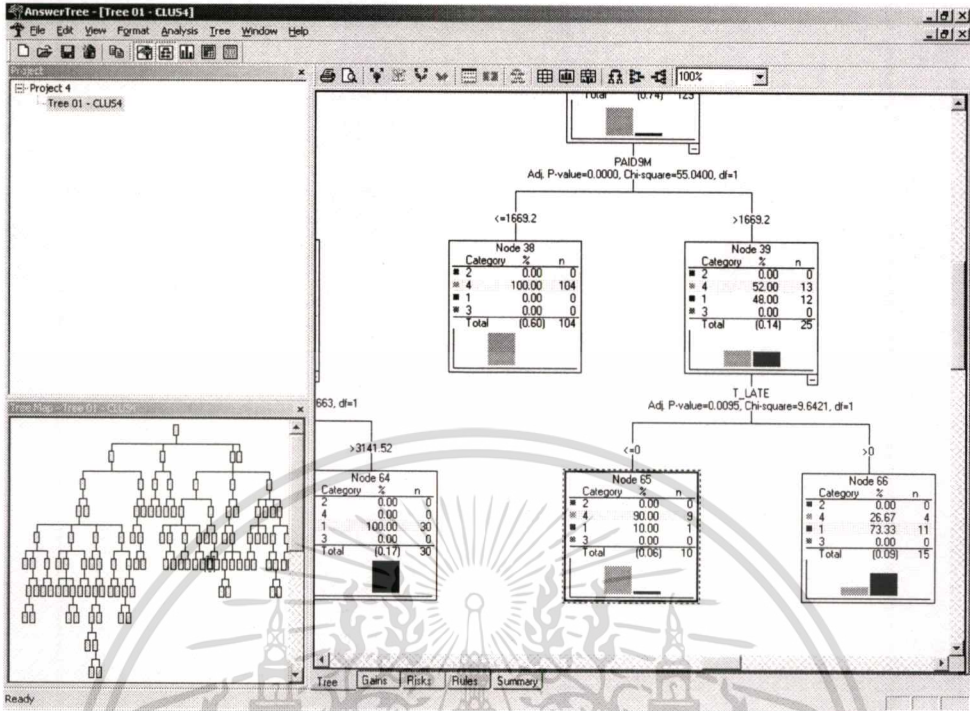


รูปที่ 4.37 Node 38 ลักษณะลูกค้ายุ่มที่ 4

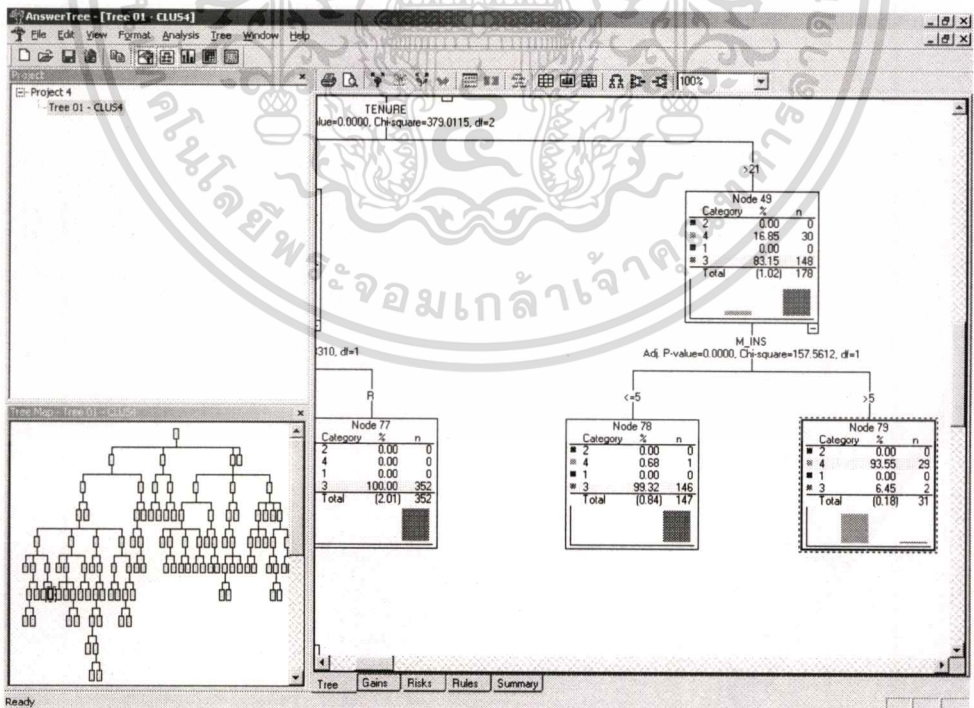


รูปที่ 4.38 Node 63 ลักษณะลูกค้ายุ่มที่ 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

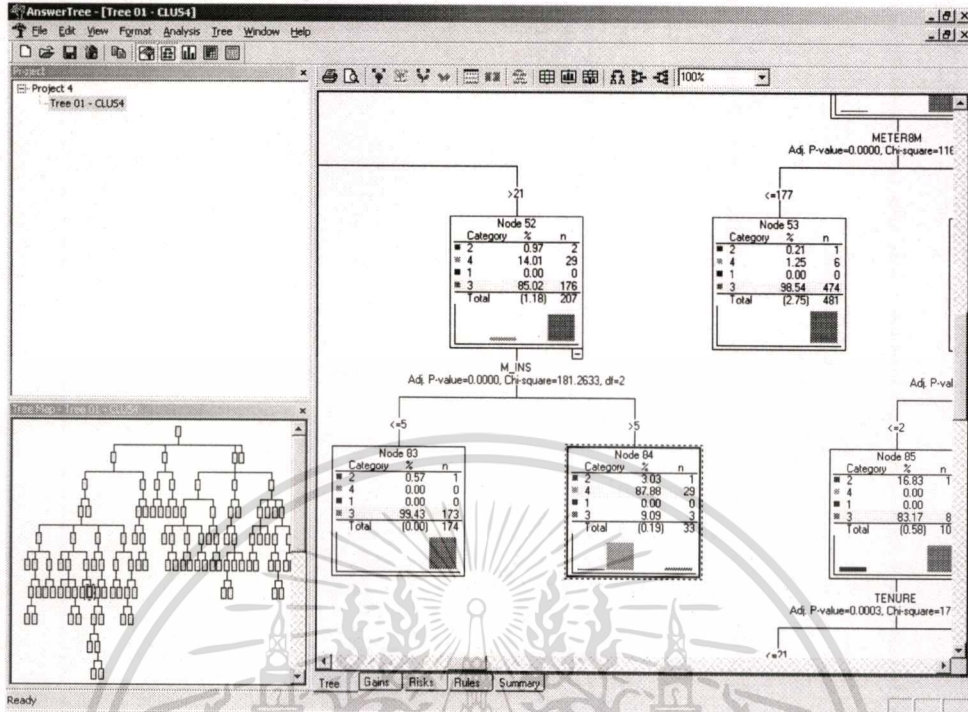


รูปที่ 4.39 Node 65 ลักษณะลูกค้ำกลุ่มที่ 4



รูปที่ 4.40 Node 79 ลักษณะลูกค้ำกลุ่มที่ 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.41 Node 84 ลักษณะลูกค้ำกลุ่มที่ 4

ผลลัพธ์จากการศึกษาการทำนายลูกค้ำที่มีแนวโน้มจะยกเลิกการใช้โทรศัพท์พื้นฐาน โดยใช้ Predictive Modeling จากแบบจำลองนี้ให้ความถูกต้องประมาณ 95% มีค่าความเสี่ยงในการพยากรณ์เท่ากับ 0.05 ดังรูปที่ 4.42

Misclassification Matrix						
		Actual Category				Total
		2	4	1	3	
Predicted Category	2	2075	108	122	111	2416
	4	86	5821	80	11	5998
	1	119	33	2143	67	2362
	3	80	29	56	6532	6697
Total		2360	5991	2401	6721	17473
Risk Statistics						
Risk Estimate		0.0516225				
SE of Risk Estimate		0.00167389				

รูปที่ 4.42 ค่าความเสี่ยงในการพยากรณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5 การนำมาประยุกต์ใช้กับธุรกิจ

การนำผลลัพธ์จากการทำค้ำไมนิ่งข้างต้นมาประยุกต์ใช้กับธุรกิจ เพื่อป้องกันการยกเลิกการใช้บริการ โทรศัพท์ มีแนวทางดังนี้

- 1) ใช้โปรแกรมการส่งเสริมการขายระยะยาว สำหรับลูกค้าที่มียอดการใช้โทรศัพท์จำนวนมาก เช่น ให้ส่วนลดกับลูกค้าที่มียอดการใช้โทรศัพท์ 1,000 บาทขึ้นไป โดยกำหนดส่วนลดเป็นเปอร์เซ็นต์ ทั้งนี้ขึ้นระยะเวลาการเป็นลูกค้ากับบริษัทด้วย
- 2) ออกแคมเปญสำหรับลูกค้ากลุ่มใหญ่ โดยมีจุดเด่นตรงที่สิ่งที่ลูกค้าแต่ละรายจะได้รับจากแคมเปญนั้นไม่เหมือนกัน เช่น สะสมแต้มจากการใช้โทรศัพท์ และสามารถนำแต้มที่สะสมมาแลกเปลี่ยนของรางวัลตามที่ลูกค้าต้องการได้ เป็นต้น
- 3) สร้างต้นทุนหรืออุปสรรคในการเปลี่ยนไปเป็นลูกค้าของบริษัทอื่น (Switching Cost) เช่น แคมเปญสะสมแต้มที่มีรางวัลมูลค่าสูงๆ และเมื่อสะสมคะแนนถึงระดับหนึ่ง แต้มคะแนนจะเพิ่มขึ้นในอัตราก้าวหน้า
- 4) ใช้ระบบศูนย์บริการลูกค้า (Call Center) ที่สนับสนุนข้อมูลเกี่ยวกับข้อมูลลูกค้า ข้อมูลการใช้งานโทรศัพท์ ข้อมูลผลิตภัณฑ์ และข้อมูลการส่งเสริมการขาย ทั้งนี้เพื่อเป็นสร้างความสัมพันธ์ และบริการที่ดีแก่ลูกค้า

การนำแนวทางดังกล่าวไปปรับใช้กับธุรกิจจำเป็นต้องพิจารณาความคุ้มค่าในการลงทุน โดยคำนวณต้นทุนค่าใช้จ่ายเปรียบเทียบกับกำไรที่ได้รับ เพื่อตัดสินใจเลือกแนวทางที่เหมาะสมต่อไป

บทที่ 5

บทสรุป

โครงการศึกษาระณีพิเศษฉบับนี้ ได้นำเสนอการศึกษาการทำนายลูกค้าที่มีแนวโน้มจะยกเลิกการใช้โทรศัพท์พื้นฐาน โดยดำเนินการตามขั้นตอนการทำดาต้าไมนิ่ง และทฤษฎีอื่นๆ ที่เกี่ยวข้อง พอสรุปได้ดังนี้

5.1 การทำดาต้าไมนิ่ง

1) กำหนดวัตถุประสงค์เพื่อศึกษา และวิเคราะห์พฤติกรรมลูกค้า จากการทำนายลูกค้าที่มีแนวโน้มจะยกเลิกการใช้โทรศัพท์พื้นฐาน เพื่อจัดทำแผนการตลาดในการรักษาลูกค้าปัจจุบัน และดึงลูกค้าเก่ากลับมาใช้บริการตามเดิม

2) เตรียมข้อมูลที่ใช้ในการศึกษา โดยการคัดเลือกข้อมูล ทำข้อมูลให้สมบูรณ์ ทดสอบข้อมูลและเปลี่ยนแปลงข้อมูลให้เหมาะสมที่จะนำไปทำดาต้าไมนิ่ง

3) เลือก Operation ที่เหมาะสมในการศึกษานี้ใช้ 2 Operation คือ Database Segmentation ในการจัดกลุ่มลูกค้า และ Predictive Modeling ในการหารูปแบบของแต่ละกลุ่มลูกค้า

4) วิเคราะห์ผลลัพธ์จากการศึกษานี้พบว่า พฤติกรรมลูกค้าที่มีแนวโน้มจะยกเลิกการใช้บริการแบ่งเป็น 4 กลุ่ม ดังนี้

- กลุ่มที่ 1 ลูกค้าที่มีการใช้โทรศัพท์ทางไกลมากกว่า 2,099.5 บาท และจ่ายชำระเงินมากกว่า 1,669.20 บาท ติดตั้งระหว่างเดือนกรกฎาคม – ตุลาคม

- กลุ่มที่ 2 ลูกค้าที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลมากกว่า 177 ครั้ง (มากกว่า 531 บาท) ใช้โทรศัพท์ทางไกลน้อยกว่า 2,099.5 บาทและจ่ายชำระค่าใช้บริการโทรศัพท์มากกว่า 1,669.20 บาท และเป็นลูกค้ามากกว่า 13 ปี

- กลุ่มที่ 3 ลูกค้าประเภทบ้านพัก ธุรกิจ รัฐบาล รัฐวิสาหกิจ ที่มีการใช้โทรศัพท์ภายในกรุงเทพฯ และปริมณฑลน้อยกว่า 289 ครั้ง (น้อยกว่า 867 บาท) ใช้โทรศัพท์ทางไกลน้อยกว่า 2,099.5 บาท จ่ายชำระค่าใช้บริการโทรศัพท์น้อยกว่า 1,669.20 บาทและมีจ่ายชำระค่าใช้บริการน้อยกว่า 2 ครั้ง

- กลุ่มที่ 4 ลูกค้าที่มีการใช้โทรศัพท์โทรภายในกรุงเทพฯและปริมณฑลน้อยกว่า 289 ครั้ง (น้อยกว่า 867 บาท) ใช้โทรศัพท์ทางไกลมากกว่า 4,128 บาท จ่ายชำระค่าใช้บริการโทรศัพท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปเผยแพร่ขึ้นต้นการคำ

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระหว่าง 0 – 3,141.52 บาท และติดตั้งโทรศัพท์ภายในเดือนมิถุนายน – ธันวาคม เป็นลูกค้ามากกว่า 21 ปี

5) การนำผลลัพธ์จากการทำค้ำไ่มิ่งข้างต้นมาประยุกต์ใช้กับธุรกิจ เพื่อป้องกันการยกเลิกการใช้บริการโทรศัพท์ มีหลายแนวทาง ทั้งนี้จำเป็นต้องพิจารณาความคุ้มค่าในการลงทุน โดยคำนวณต้นทุนค่าใช้จ่ายเปรียบเทียบกับกำไรที่ได้รับ เพื่อตัดสินใจเลือกแนวทางที่เหมาะสมต่อไป

5.2 ปัญหาและอุปสรรค

1) ที่ผ่านมามีบริษัทฯ ไม่ได้จัดเก็บข้อมูลที่สำคัญ ทำให้การศึกษานี้ยังขาดตัวแปรอื่นๆ ที่เกี่ยวข้องที่สามารถให้ผลลัพธ์ได้ละเอียด ถูกต้อง และแม่นยำขึ้น ได้แก่

- ข้อมูลเกี่ยวกับลูกค้า เช่น รายได้ลูกค้า เพศ อายุ และอาชีพ เป็นต้น

- ข้อมูลการใช้โทรศัพท์ เช่น การใช้บริการเสริมต่างๆ สายเรียกช้อน โอนเลขหมาย

สนทนา 3 สาย และโทรซ้ำอัตโนมัติ เป็นต้น

- ข้อมูลคุณภาพบริการ เช่น การเรียกเลขหมายปลายทางสำเร็จ เป็นต้น

- ข้อมูลไม่เพียงพอต่อการพยากรณ์ เนื่องจากหน่วยงานไม่จัดเก็บข้อมูลย้อนหลัง

- ข้อมูลอื่น เช่น สาเหตุที่ลูกค้ายกเลิกการใช้บริการโทรศัพท์ เป็นต้น

2) สภาพแวดล้อมในการทำงานไม่สนับสนุนต่อการทำงานกับข้อมูลจำนวนมาก กล่าวคือ การประมวลผลข้อมูลจำนวนมากบนเครื่อง Notebook แรม 256 ฮาร์ดดิสก์ 30 GB ทำให้เกิดข้อผิดพลาดบ่อยครั้ง ในการทำงาน รวมถึงทำให้การทำงานช้าลง

บรรณานุกรม

กัลยา วานิชย์บัญชา. 2545. การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล. กรุงเทพฯ: โรงพิมพ์ จุฬาลงกรณ์มหาวิทยาลัย

Cavena ,Peter.et al.1995. **Discover Data Mining**. New Jersey: Prentice Hall PRT.

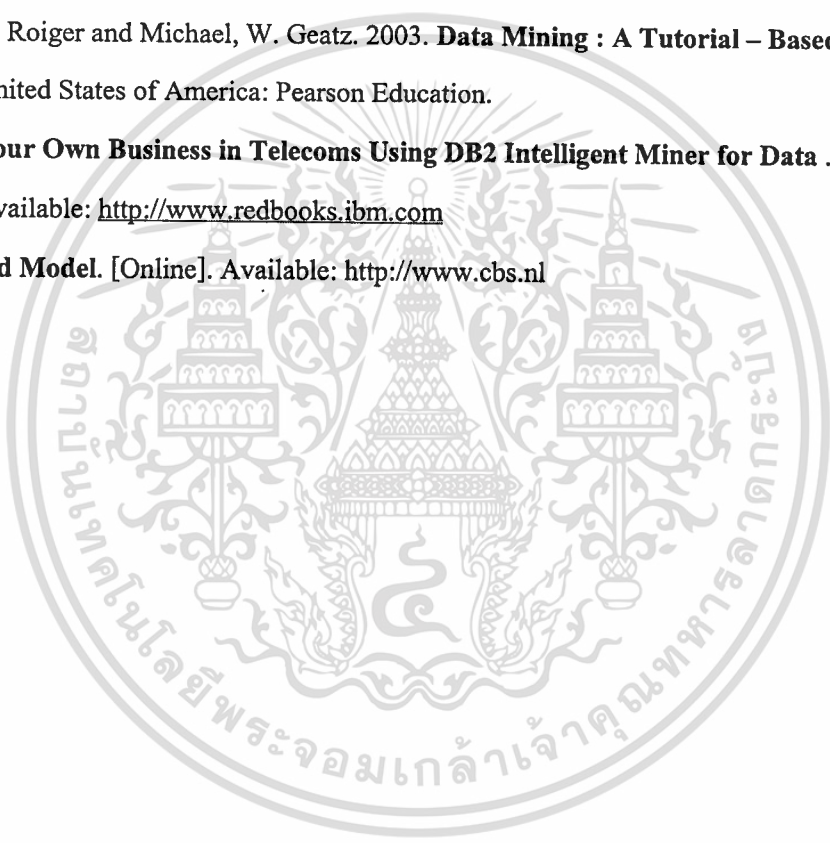
Jiawei Han and Micheline Kamber. 2001. **Data Mining: Concepts and Techniques**. Simon Fraser University: MORGAN KAUFMANN.

Richard, J. Roiger and Michael, W. Geatz. 2003. **Data Mining : A Tutorial – Based Primer**. United States of America: Pearson Education.

Mining Your Own Business in Telecoms Using DB2 Intelligent Miner for Data . [Online].

Available: <http://www.redbooks.ibm.com>

Tree Based Model. [Online]. Available: <http://www.cbs.nl>





ภาคผนวก ก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



**การใช้งานโปรแกรมสำเร็จรูป ACL
ในขั้นตอนการกำหนดเพิ่มข้อมูลในรูปแบบตาราง
(Defining Table)**

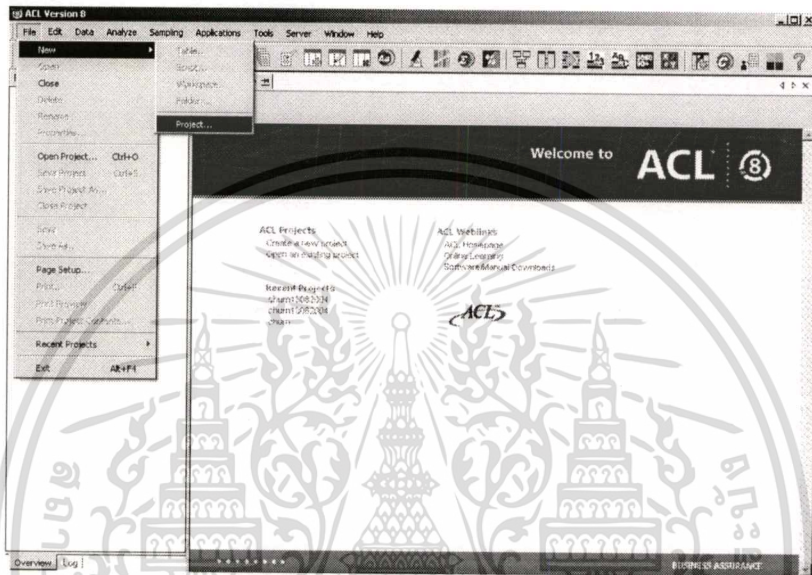
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการสร้าง และ กำหนดเพิ่มข้อมูล (Defining Tables) มี 5 ขั้นตอนหลัก ดังนี้

1. การเลือกเพิ่มข้อมูล (Select the data file)

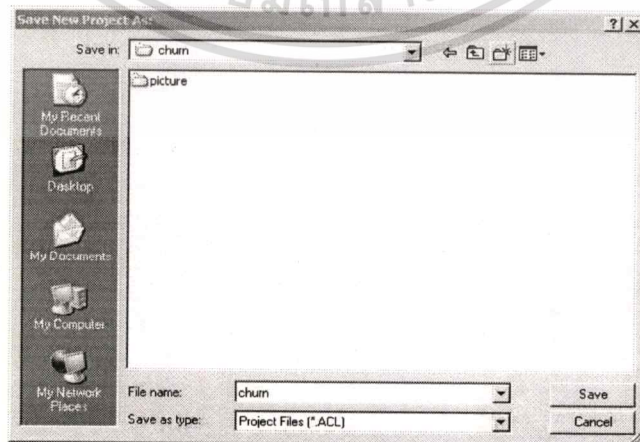
1.1 สร้าง Project (Creating a New Project)

1.1.1 เลือก Menu 'File' , 'New' ดังรูปที่ ก.1



รูปที่ ก.1 แสดงหน้าจอการสร้าง Project

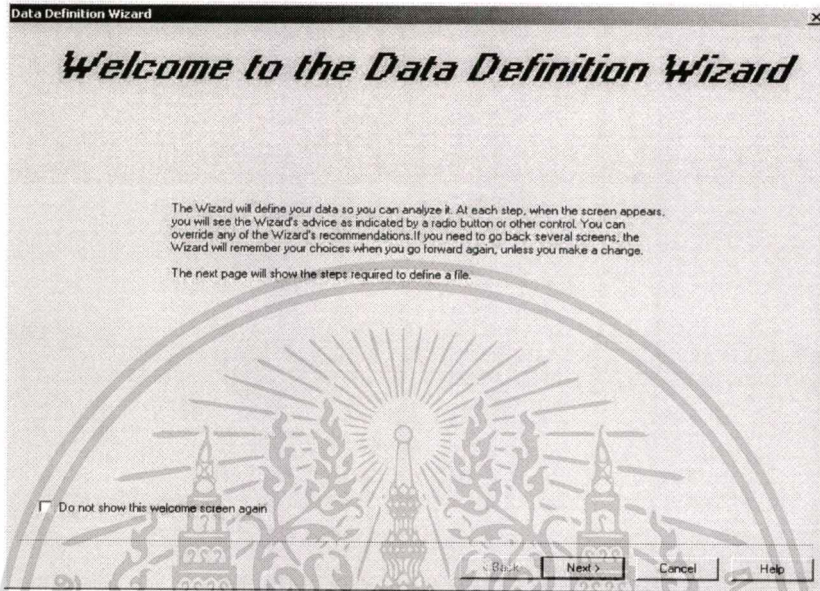
1.1.2 เลือก 'Project' จะปรากฏหน้าจอ หลังจากนั้นให้กำหนดชื่อ Project ใหม่ ในช่อง File name ดังรูปที่ ก.2



รูปที่ ก.2 แสดงหน้าจอการบันทึกเพิ่มข้อมูล Project

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

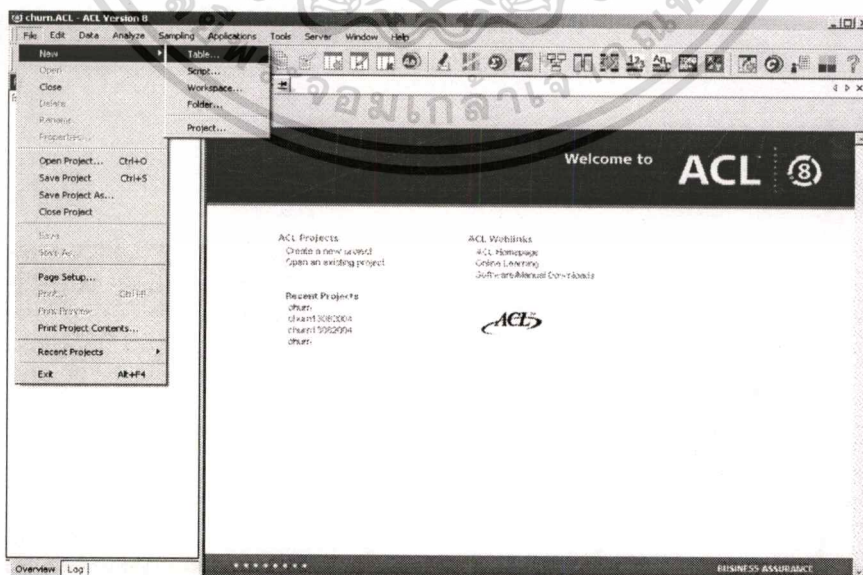
1.1.3 หลังจากนั้น เลือกปุ่ม 'Save' เพื่อบันทึก Project และเริ่มปฏิบัติงานกับ Project นี้ ดังรูปที่ ก.3



รูปที่ ก.3 แสดงหน้าจอสิ้นสุดการสร้าง Project

1.2 การสร้างเพิ่มข้อมูล

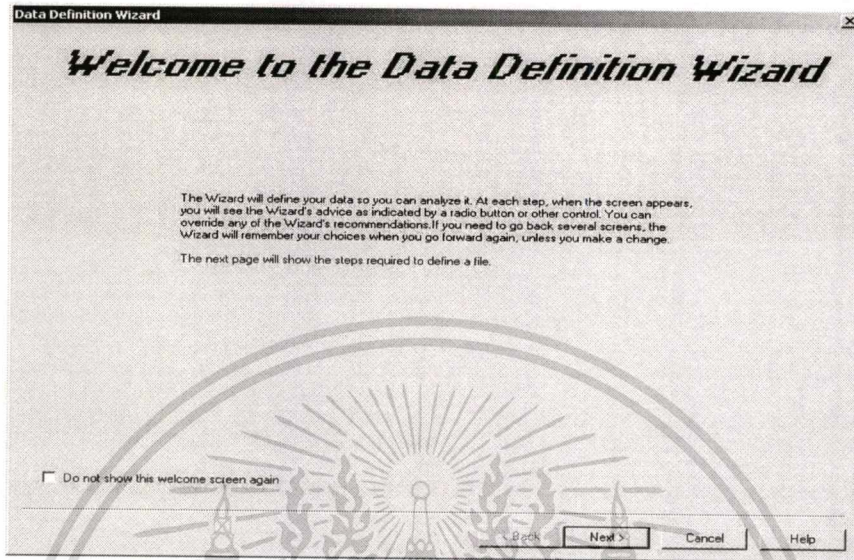
1.2.1 กำหนดเพิ่มข้อมูลสำหรับ Project เลือก 'File', 'New' ดังรูปที่ ก.4



รูปที่ ก.4 แสดงหน้าจอการสร้างเพิ่มข้อมูล

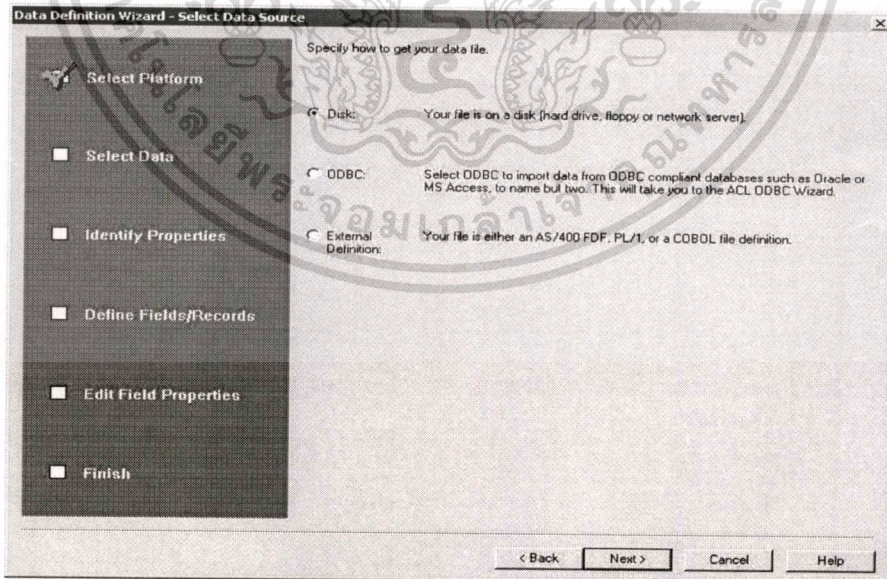
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ภายในเท่านั้น ไม่ควรเผยแพร่หรือทำซ้ำโดยไม่ได้รับอนุญาต หากต้องการข้อมูลเพิ่มเติม กรุณาติดต่อฝ่ายสนับสนุนลูกค้า

1.2.2 เลือก 'Table' เพื่อเริ่มต้นการสร้างแฟ้มข้อมูล ดังรูปที่ ก.5



รูปที่ ก.5 แสดงหน้าจอการเริ่มสร้างแฟ้มข้อมูล

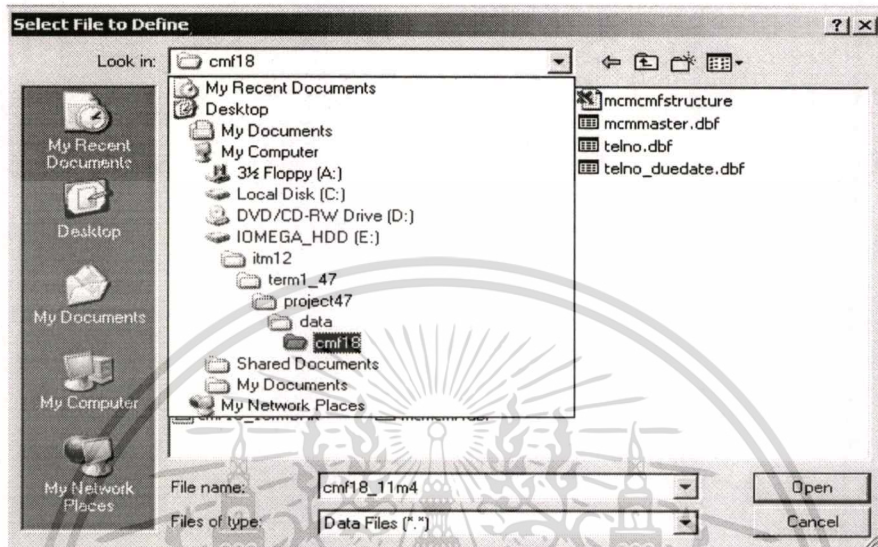
1.2.3 เลือกปุ่ม 'Next' เพื่อเลือกแหล่งข้อมูล ดังรูปที่ ก.6



รูปที่ ก.6 แสดงหน้าจอการเลือกแหล่งที่มาของข้อมูล

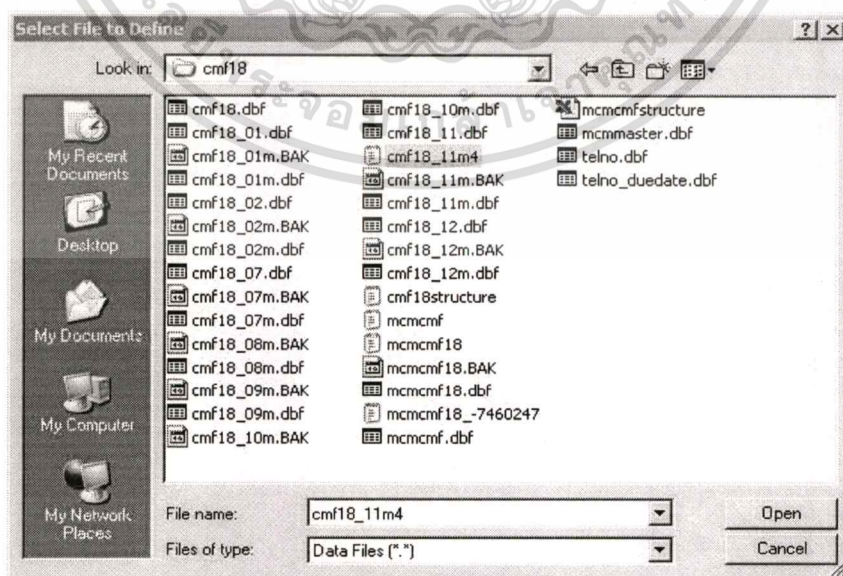
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2.4 เลือกแหล่งข้อมูลจาก Disk แล้ว เลือกปุ่ม 'Next' เพื่อเลือกเพิ่มข้อมูลที่ต้องการนำมาสร้าง ดังรูปที่ ก.7



รูปที่ ก.7 แสดงหน้าจอการเลือกข้อมูลจาก Disk

1.2.5 เลือกชื่อข้อมูลที่จะนำมาสร้างเพิ่มข้อมูล โดยเลือกที่ชื่อเพิ่มข้อมูลที่ต้องการดังรูปที่ ก.8



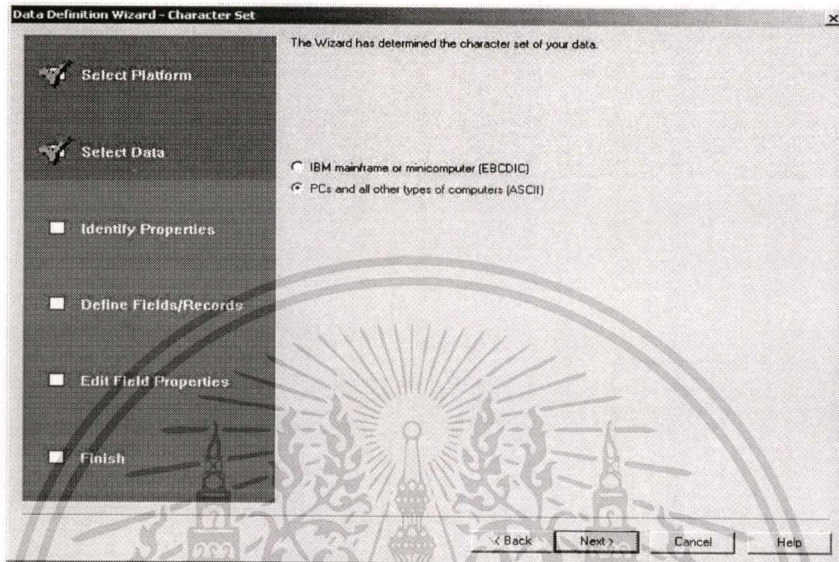
รูปที่ ก.8 แสดงหน้าจอการเลือกเพิ่มข้อมูลที่ต้องการ

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการศึกษาเท่านั้น เมื่ออนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

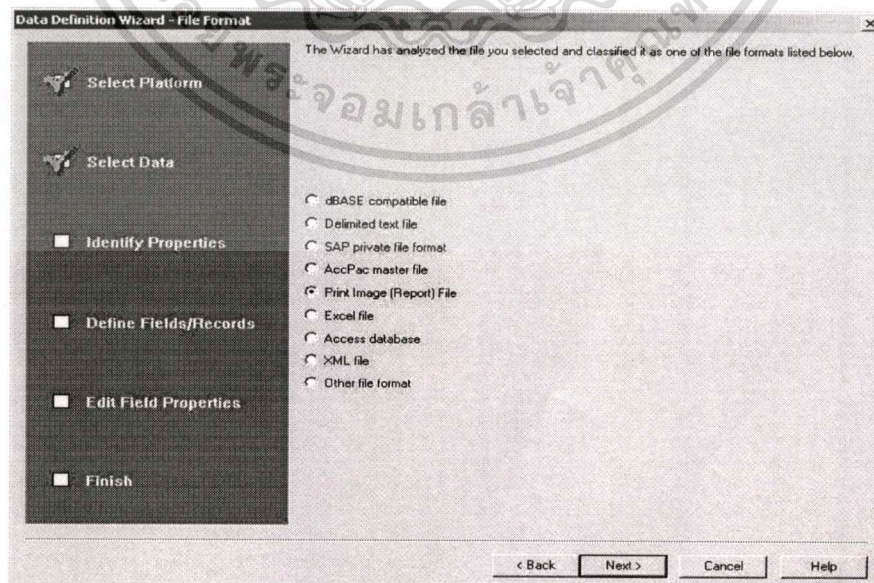
2. การกำหนดคุณลักษณะของแฟ้มข้อมูล (The Wizard identifies properties of data file)

2.1 จากข้อ 1.2.5 เลือกปุ่ม 'Open' จะปรากฏหน้าจอ ดังรูปที่ ก.9



รูปที่ ก.9 แสดงหน้าจอการเลือกลักษณะของแฟ้มข้อมูล

2.2 เลือกกลุ่มข้อมูล 'PC and all other types of computer (ASCII)' และ เลือกปุ่ม 'Next' จะพบหน้าจอ ดังรูปที่ ก.10



เอกสารนี้เป็นเอกสารที่สงวนรูปที่ ก.10 แสดงหน้าจอการเลือกประเภทของแฟ้มข้อมูล ไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. เริ่มต้นการสร้างเพิ่มข้อมูล เลือกรูปแบบ 'Next' โปรแกรมจะแสดงลักษณะของข้อมูล ดังรูปที่

ก.11

Hint: Select a field by clicking and dragging. A field may span more than one line. Record Length: 185

IRUN DATE	30/11/46	PROGRAM NO.	CMF18RNC	BILLING REGISTER					
OSTD CODE	002	EXCH	220	THAI INTER AIR.	UNIT 01	LONG DIST.	CURRENT ADJUSTMEN		
OLINE NO.	INS. DATE	ST. TYPE OR ZIP	ACC GOV. ZIP	RENTAL CHARGE	METER USAGE	METER CHARGE	CHARGE	CHARGE	
4000	140636	H1	Q 6038	30000.00	35559	106,677.00	0.00	136,677.00	6833.8
5000	090945	A H1	T 6032	30000.00	36470	109,410.00	0.00	139,410.00	6970.5
8100	180142	C B1	T10100	30000.00	0	0.00	0.00	30,000.00	0.0
8110	070143	C B1	T10100	7500.00	0	0.00	0.00	7,500.00	0.0
8111	180142	C B1	T10500	60000.00	1	3.00	0.00	60,003.00	0.0
8112	090742	C B1	T10500	7500.00	1	3.00	0.00	7,503.00	0.0
8150	290942	C B1	T10100	7500.00	0	0.00	0.00	7,500.00	0.0
8190	121142	C B1	T10330	30000.00	0	0.00	0.00	30,000.00	0.0
8200	180142	C B1	T10330	7500.00	0	0.00	0.00	7,500.00	0.0
8230	270442	C B1	T10330	7500.00	0	0.00	0.00	7,500.00	0.0
8300	220644	A G1	T 1907	7500.00	0	0.00	0.00	7,500.00	375.0
8800	090643	F9C B1	T 2000	0.00	0	0.00	0.00	0.00	0.0
8801	031143	F9C B1	T 2000	0.00	0	0.00	0.00	0.00	0.0
8802	061243	F9C B1	T 2000	0.00	0	0.00	0.00	0.00	0.0
8803	061243	F9C B1	T 2000	0.00	0	0.00	0.00	0.00	0.0
8900	020643	C B1	T10400	7500.00	0	0.00	0.00	7,500.00	0.0
8940	210944	C B1	T10400	7500.00	0	0.00	0.00	7,500.00	0.0
EXCHANGE TOTAL				240,000.00	72,031	216,093.00	0.00	456,093.00	14,179.3

< Back Next > Cancel Help

รูปที่ ก.11 แสดงหน้าจอประเภทเพิ่มข้อมูลที่เป็น Print Image File

4. กำหนดคุณลักษณะของ Field

4.1 เลือก Field เริ่มต้น กำหนดชื่อ (Name) และชนิด (Type) ดังรูปที่ ก.12

Hint: Select a field by clicking and dragging. A field may span more than one line. Record Length: 195

IRUN DATE	30/11/46	Field Definition						NT ADJUSTMEN	
OSTD CODE	002	Name	Type					GE	
4000	140636	code	Character					00	
5000	090945	Advanced Options >>						6833.8	
8100	180142							00	
8110	070143							00	
8111	180142	C B1	T10500	60000.00	1	3.00	0.00	60,003.00	0.0
8112	090742	C B1	T10500	7500.00	1	3.00	0.00	7,503.00	0.0
8150	290942	C B1	T10100	7500.00	0	0.00	0.00	7,500.00	0.0
8190	121142	C B1	T10330	30000.00	0	0.00	0.00	30,000.00	0.0
8200	180142	C B1	T10330	7500.00	0	0.00	0.00	7,500.00	0.0
8230	270442	C B1	T10330	7500.00	0	0.00	0.00	7,500.00	0.0
8300	220644	A G1	T 1907	7500.00	0	0.00	0.00	7,500.00	375.0
8800	090643	F9C B1	T 2000	0.00	0	0.00	0.00	0.00	0.0
8801	031143	F9C B1	T 2000	0.00	0	0.00	0.00	0.00	0.0
8802	061243	F9C B1	T 2000	0.00	0	0.00	0.00	0.00	0.0
8803	061243	F9C B1	T 2000	0.00	0	0.00	0.00	0.00	0.0
8900	020643	C B1	T10400	7500.00	0	0.00	0.00	7,500.00	0.0
8940	210944	C B1	T10400	7500.00	0	0.00	0.00	7,500.00	0.0
EXCHANGE TOTAL				240,000.00	72,031	216,093.00	0.00	456,093.00	14,179.3

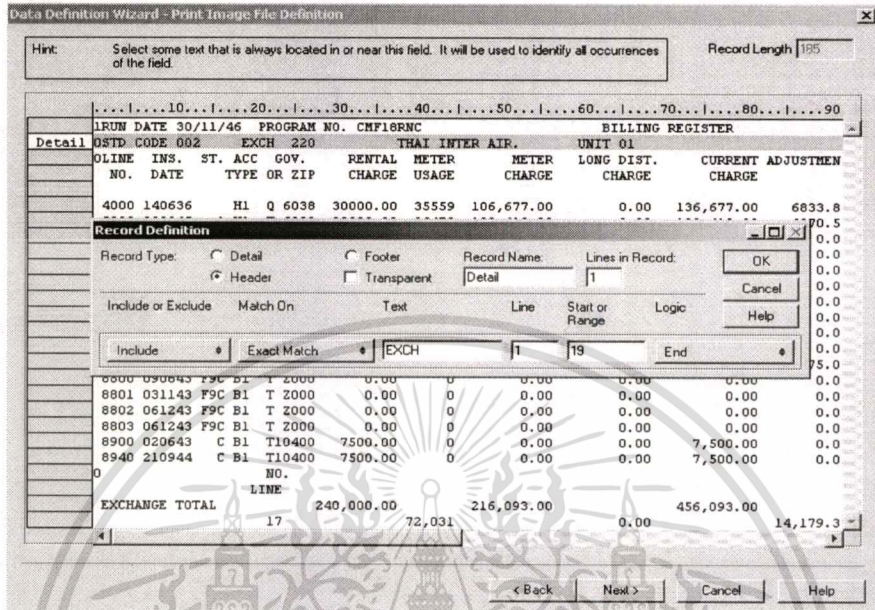
< Back Next > Cancel Help

รูปที่ ก.12 แสดงหน้าจอการกำหนด Filed เริ่มต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

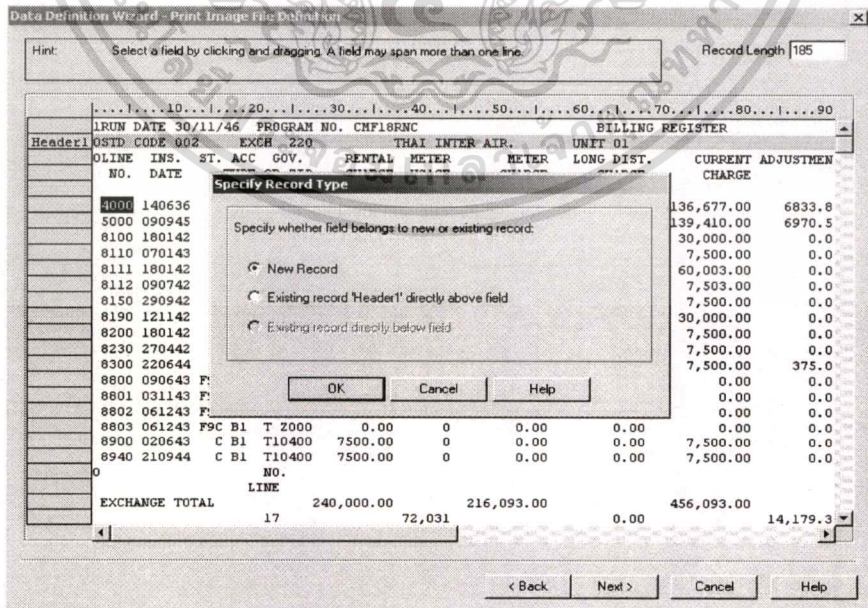
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 เลือก Key Field สำหรับ Header ที่มีข้อมูลซ้ำกันทุก Header ดังรูปที่ ก.13



รูปที่ ก.13 แสดงหน้าจอการกำหนด Header โดยใช้ Key Field ที่ซ้ำกัน

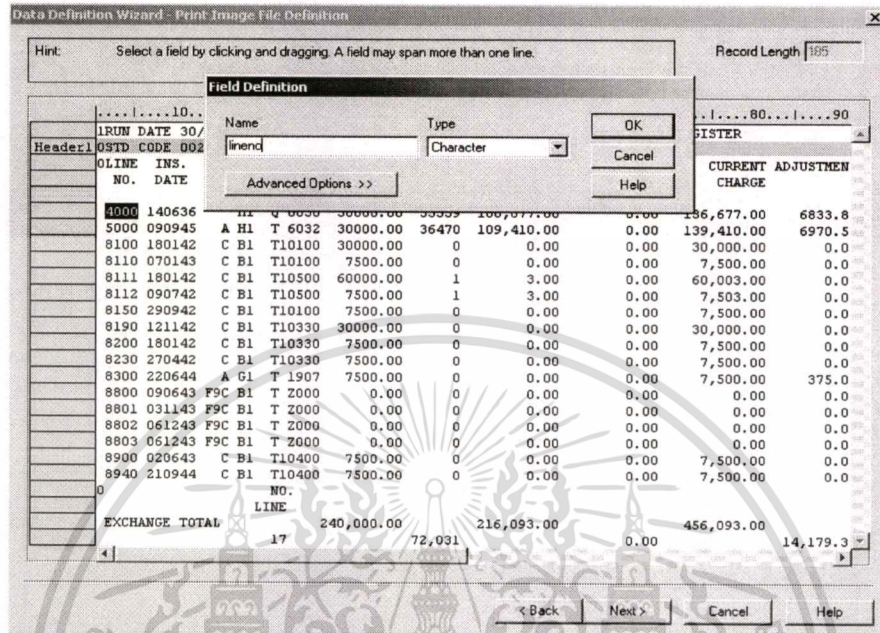
4.3 เลือก Field ต่อไป กำหนดให้เป็น New Record ดังรูปที่ ก.14



รูปที่ ก.14 แสดงหน้าจอการกำหนด Record ต่อไป

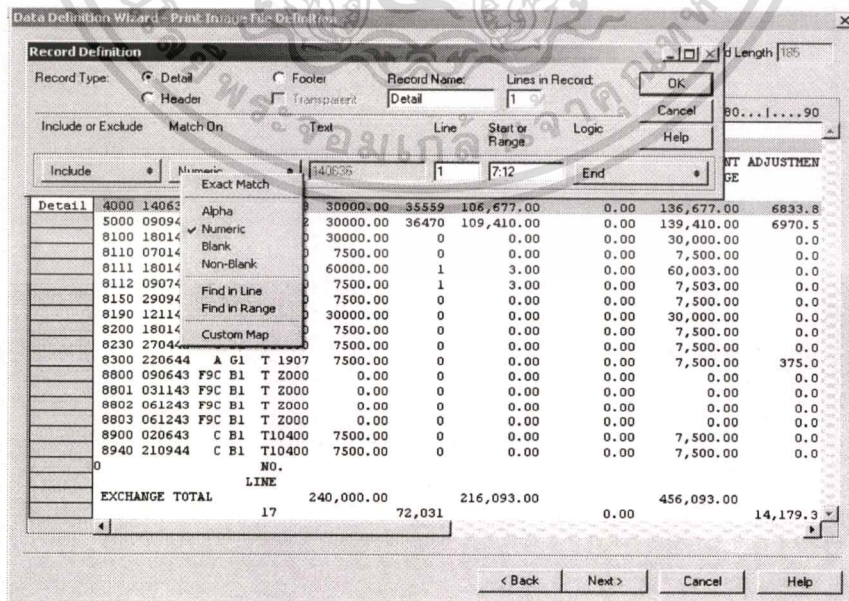
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 กำหนดชื่อ (Name) และชนิด (Type) ของ Field ดังรูปที่ ก.15



รูปที่ ก.15 แสดงหน้าจอการกำหนด Field ใน Detail

4.5 เลือก Key Field สำหรับ Detail ที่มีข้อมูลซ้ำกันทุก Detail ดังรูปที่ ก.16



รูปที่ ก.16 แสดงหน้าจอการกำหนด Detail โดยใช้ Key Field ที่ซ้ำกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้เผยแพร่หรือใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6 โปรแกรมจะเลือกข้อมูลตามที่กำหนดเงื่อนไข ดังรูปที่ ก.17

Data Definition Wizard - Print Image File Definition

Hint: Select a field by clicking and dragging. A field may span more than one line. Record Length 185

Header	OLINE NO.	INS. DATE	ST. TYPE OR ZIP	ACC GOV.	RENTAL CHARGE	METER USAGE	THAI INTER AIR. METER CHARGE	UNIT 01 LONG DIST. CHARGE	CURRENT ADJUSTMEN CHARGE	
Detail	4000	140636	HI	Q 6038	30000.00	35559	106,677.00	0.00	136,677.00	6833.8
Detail	5000	090945	A	HL T 6032	30000.00	36470	109,410.00	0.00	139,410.00	6970.5
Detail	8100	180142	C	BL T10100	30000.00	0	0.00	0.00	30,000.00	0.0
Detail	8110	070143	C	BL T10100	7500.00	0	0.00	0.00	7,500.00	0.0
Detail	8111	180142	C	BL T10500	60000.00	1	3.00	0.00	60,003.00	0.0
Detail	8112	090742	C	BL T10500	7500.00	1	3.00	0.00	7,503.00	0.0
Detail	8150	290942	C	BL T10100	7500.00	0	0.00	0.00	7,500.00	0.0
Detail	8190	121142	C	BL T10330	30000.00	0	0.00	0.00	30,000.00	0.0
Detail	8200	180142	C	BL T10330	7500.00	0	0.00	0.00	7,500.00	0.0
Detail	8230	270442	C	BL T10330	7500.00	0	0.00	0.00	7,500.00	0.0
Detail	8300	220644	A	GI T 1907	7500.00	0	0.00	0.00	7,500.00	375.0
Detail	8600	090643	F9C	BL T 2000	0.00	0	0.00	0.00	0.00	0.0
Detail	8601	031143	F9C	BL T 2000	0.00	0	0.00	0.00	0.00	0.0
Detail	8602	061243	F9C	BL T 2000	0.00	0	0.00	0.00	0.00	0.0
Detail	8603	061243	F9C	BL T 2000	0.00	0	0.00	0.00	0.00	0.0
Detail	8900	020643	C	BL T10400	7500.00	0	0.00	0.00	7,500.00	0.0
Detail	8940	210944	C	BL T10400	7500.00	0	0.00	0.00	7,500.00	0.0
EXCHANGE TOTAL										
	17				240,000.00	72,031	216,093.00	0.00	456,093.00	14,179.3

< Back Next > Cancel Help

รูปที่ ก.17 แสดงหน้าจอการเลือกข้อมูลตามเงื่อนไข

4.7 เลือกข้อมูลที่ต้องการ กำหนดชื่อ (Name) และชนิดของข้อมูล (Type) ดังรูปที่ ก.18

Data Definition Wizard - Print Image File Definition

Hint: Select a field by clicking and dragging. A field may span more than one line. Record Length 185

Header	OLINE NO.	INS. DATE	ST. TYPE OR ZIP	ACC GOV.	RENTAL CHARGE	METER USAGE	THAI INTER AIR. METER CHARGE	UNIT 01 LONG DIST. CHARGE	CURRENT ADJUSTMEN CHARGE	
Detail	4000	140636	HI	Q 6038	30000.00	35559	106,677.00	0.00	136,677.00	6833.8
Detail	5000	090945	A	HL T 6032	30000.00	36470	109,410.00	0.00	139,410.00	6970.5
Detail	8100	180142	C	BL T10100	30000.00	0	0.00	0.00	30,000.00	0.0
Detail	8110	070143	C	BL T10100	7500.00	0	0.00	0.00	7,500.00	0.0
Detail	8111	180142	C	BL T10500	60000.00	1	3.00	0.00	60,003.00	0.0
Detail	8112	090742	C	BL T10500	7500.00	1	3.00	0.00	7,503.00	0.0
Detail	8150	290942	C	BL T10100	7500.00	0	0.00	0.00	7,500.00	0.0
Detail	8190	121142	C	BL T10330	30000.00	0	0.00	0.00	30,000.00	0.0
Detail	8200	180142	C	BL T10330	7500.00	0	0.00	0.00	7,500.00	0.0
Detail	8230	270442	C	BL T10330	7500.00	0	0.00	0.00	7,500.00	0.0
Detail	8300	220644	A	GI T 1907	7500.00	0	0.00	0.00	7,500.00	375.0
Detail	8600	090643	F9C	BL T 2000	0.00	0	0.00	0.00	0.00	0.0
Detail	8601	031143	F9C	BL T 2000	0.00	0	0.00	0.00	0.00	0.0
Detail	8602	061243	F9C	BL T 2000	0.00	0	0.00	0.00	0.00	0.0
Detail	8603	061243	F9C	BL T 2000	0.00	0	0.00	0.00	0.00	0.0
Detail	8900	020643	C	BL T10400	7500.00	0	0.00	0.00	7,500.00	0.0
Detail	8940	210944	C	BL T10400	7500.00	0	0.00	0.00	7,500.00	0.0
EXCHANGE TOTAL										
	17				240,000.00	72,031	216,093.00	0.00	456,093.00	14,179.3

Field Definition

Name: Type:

OK Cancel Help

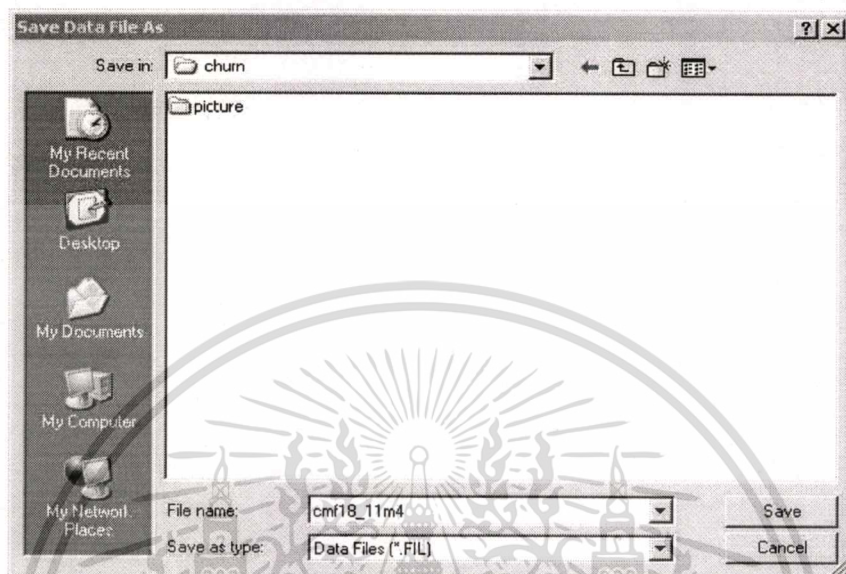
Advanced Options >>

< Back Next > Cancel Help

รูปที่ ก.18 แสดงหน้าจอการกำหนดชื่อ และชนิดของ Field

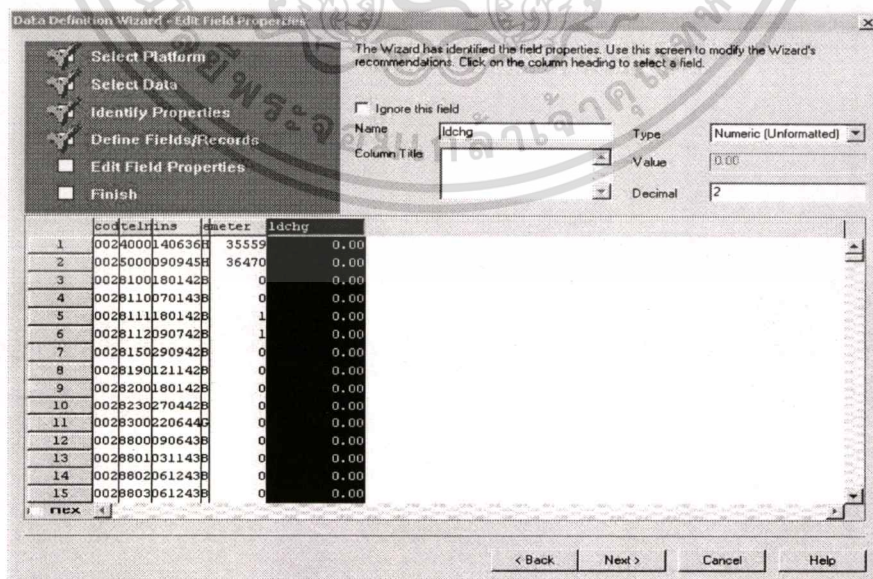
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.8 กำหนดชื่อ และชนิดของ Field ที่ต้องการแล้ว เลือกปุ่ม 'Next' เพื่อตั้งชื่อ และบันทึก
 เพิ่มข้อมูล ดังรูปที่ ก.19



รูปที่ ก.19 แสดงหน้าจอการตั้งชื่อ และบันทึกเพิ่มข้อมูล

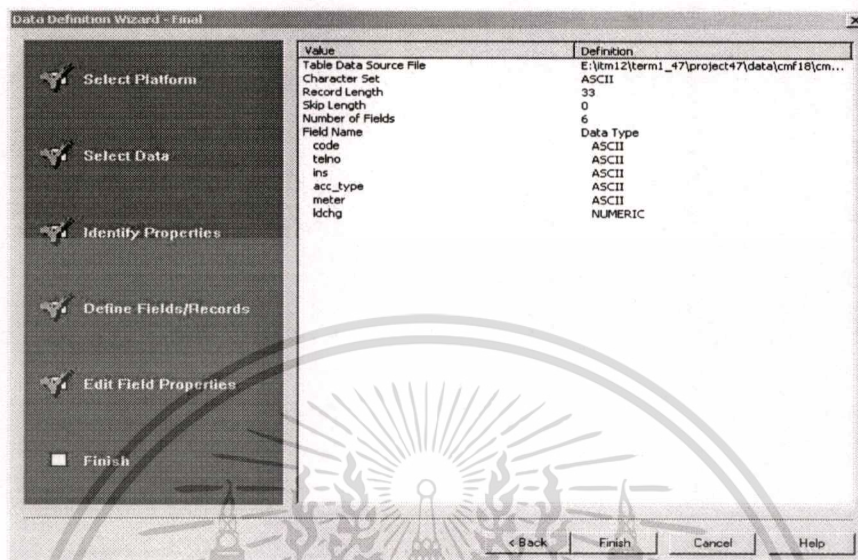
4.9 ตรวจสอบความถูกต้องของการกำหนดชื่อ และชนิดของแต่ละ Field ดังรูปที่ ก.20 หาก
 กำหนดไม่ถูกต้องสามารถแก้ไขได้



รูปที่ ก.20 แสดงหน้าจอการตรวจสอบความถูกต้องของข้อมูลแต่ละ Field

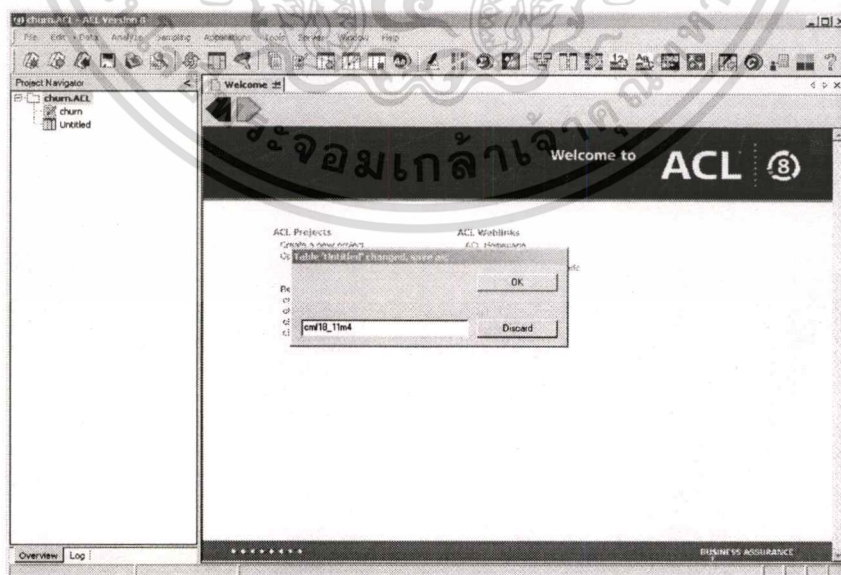
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.10 เมื่อตรวจสอบความถูกต้องแล้ว จะพบหน้าจอ ดังรูปที่ ก.21



รูปที่ ก.21 แสดงหน้าจอการยืนยันความถูกต้องของข้อมูล

4.11 สิ้นสุดการสร้างเพิ่มข้อมูล เลือกปุ่ม 'OK' เพื่อยืนยันความถูกต้อง เลือกปุ่ม 'Discard' เพื่อยกเลิกการสร้างเพิ่มข้อมูล ดังรูปที่ ก.22



รูปที่ ก.22 แสดงหน้าจอสิ้นสุดการสร้างเพิ่มข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. ขั้นตอนการสร้างเพิ่มข้อมูล (The Wizard finishes creating) จากข้อ 4.11 เลือกปุ่ม 'OK' เพื่อยืนยันการสร้าง หรือกำหนดเพิ่มข้อมูล พบหน้าจอดังรูปที่ ก.23

	code	hmgd	ins	act	hsp	mnter	lchga
	002	4000	140836	H		3558	0.00
2	002	5000	090945	H		36470	0.00
3	002	8100	180142	B		0	0.00
4	002	8110	070143	B		0	0.00
5	002	8111	180142	B		1	0.00
6	002	8112	090742	B		1	0.00
7	002	8150	290942	B		0	0.00
8	002	8190	121142	B		0	0.00
9	002	8200	180142	B		0	0.00
10	002	8230	270442	B		0	0.00
11	002	8300	220644	O		0	0.00
12	002	8800	080643	B		0	0.00
13	002	8801	031143	B		0	0.00
14	002	8802	081243	B		0	0.00
15	002	8803	061243	B		0	0.00
16	002	8900	020643	B		0	0.00
17	002	8940	210844	B		0	0.00
18	002	0000	020710	T		0	92.00
19	002	0001	030710	R		68	33.00
20	002	0002	040710	H		6	0.00
21	002	0004	060710	B		208	0.00
22	002	0005	070710	R		1	0.00
23	002	0007	190246	B		21	3.00
24	002	0008	140825	R		6	31.50
25	002	0009	080710	R		608	246.00
26	002	0010	300646	R		7	0.00
27	002	0011	160710	R		97	670.00
28	002	0012	040646	R		81	514.50
29	002	0013	241230	R		117	3231.00
30	002	0014	120710	B		39	36.00
31	002	0015	140710	R		451	1508.00
32	002	0016	090631	R		44	108.00
33	002	0017	160710	R		0	0.00

รูปที่ ก.23 แสดงหน้าจอเพิ่มข้อมูลที่สร้างขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



การใช้งานโปรแกรมสำเร็จรูป SPSS
ในขั้นตอนการเปลี่ยนแปลงรูปแบบข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

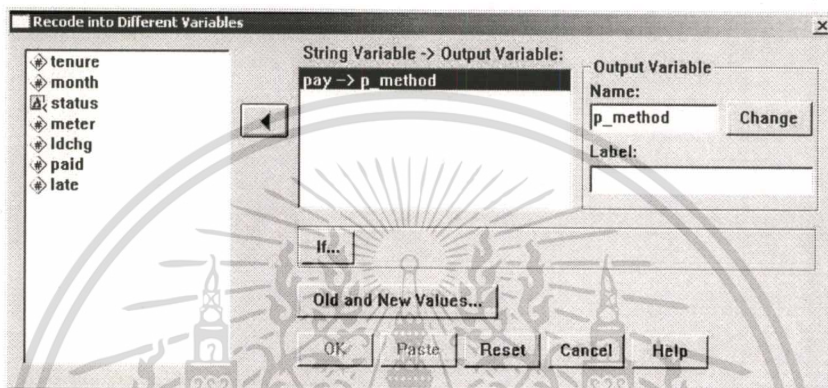
ขั้นตอนในการเปลี่ยนแปลงรูปแบบข้อมูลโดยใช้ K-Means Clustering ประกอบด้วย

เปลี่ยนแปลงรูปแบบข้อมูล โดยสร้างตัวแปรใหม่ โดยใช้คำสั่ง Transform คือ การสร้างตัวแปรใหม่ เพื่อเก็บค่าของตัวแปรที่เปลี่ยนไป โดยตัวแปรเดิมยังคงมีค่าเดิม มีขั้นตอนดังนี้

1. จะได้น้ำจืดรูปที่

ก.24


Transform  Recode  Into Different Variables



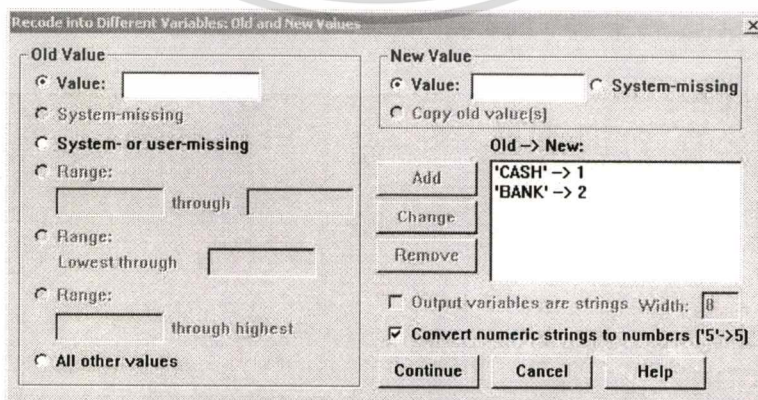
รูปที่ ก.24 แสดงหน้าจอการเปลี่ยนแปลงรูปแบบข้อมูล

2. เลือกตัวแปรเดิม และใส่ชื่อตัวแปรใหม่และ Label ของตัวแปรใหม่ในส่วนของ Output Variable

3. เลือกปุ่ม 

4. เลือกปุ่ม 

จะได้น้ำจืดรูปที่ ก.25



รูปที่ ก.25 แสดงหน้าจอการกำหนด Old and New Values

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. กำหนดค่าเดิมใน Old Value และค่าใหม่ใน New Value โดยเลือกเปลี่ยนจากตัวแปรที่เป็นตัวอักษรแปลงเป็นตัวเลข ดังรูปที่ ก.25 แล้วเลือก **Continue**

6. เลือก **OK** จะได้ตัวแปรใหม่ในเพิ่มข้อมูล ดังรูปที่ ก.26

	tenure	month	p_method	acc_type	meter100	meter8m	ldchg8m	paid9m	t_late	pay	acctype	frmeter
1	4.00	10	CASH	B	NO	423.00	.00	2213.83	.00	1.00	2.00	2.00
2	2.00	12	CASH	R	NO	874.00	.00	2404.29	2.00	1.00	1.00	2.00
3	2.00	9.0	CASH	R	NO	2125.0	.00	8533.25	.00	1.00	1.00	2.00
4	4.00	7.0	CASH	B	NO	1.00	.00	428.00	.00	1.00	2.00	2.00
5	4.00	7.0	CASH	B	NO	245.00	.00	1405.98	.00	1.00	2.00	2.00
6	4.00	7.0	CASH	B	NO	40.00	.00	770.40	.00	1.00	2.00	2.00
7	2.00	10	CASH	G	NO	1828.0	29906	33661.90	6.00	1.00	3.00	2.00
8	2.00	10	CASH	G	NO	2.00	.00	1016.50	5.00	1.00	3.00	2.00
9	2.00	10	CASH	G	NO	3.00	.00	1019.55	5.00	1.00	3.00	2.00
10	2.00	10	CASH	G	NO	2.00	.00	1016.50	5.00	1.00	3.00	2.00
11	2.00	10	CASH	G	NO	1.00	.00	1019.55	5.00	1.00	3.00	2.00
12	2.00	10	CASH	G	NO	1.00	.00	1019.55	5.00	1.00	3.00	2.00
13	2.00	10	CASH	G	NO	1.00	.00	1019.55	5.00	1.00	3.00	2.00
14	2.00	10	CASH	G	NO	32.00	.00	1071.39	5.00	1.00	3.00	2.00
15	2.00	10	CASH	G	NO	16.00	.00	1034.80	5.00	1.00	3.00	2.00
16	2.00	10	CASH	G	NO	15.00	.00	1031.75	5.00	1.00	3.00	2.00
17	2.00	10	CASH	G	NO	18.00	3.00	1836.82	5.00	1.00	3.00	2.00
18	2.00	10	CASH	G	NO	7.00	9.00	1797.17	5.00	1.00	3.00	2.00
19	2.00	10	CASH	G	NO	395.00	2868	5166.86	5.00	1.00	3.00	2.00
20	3.00	12	CASH	R	NO	19.00	237.0	999.38	5.00	1.00	1.00	2.00

รูปที่ ก.26 แสดงหน้าจอเพิ่มข้อมูลที่มีการเปลี่ยนแปลงรูปแบบข้อมูล



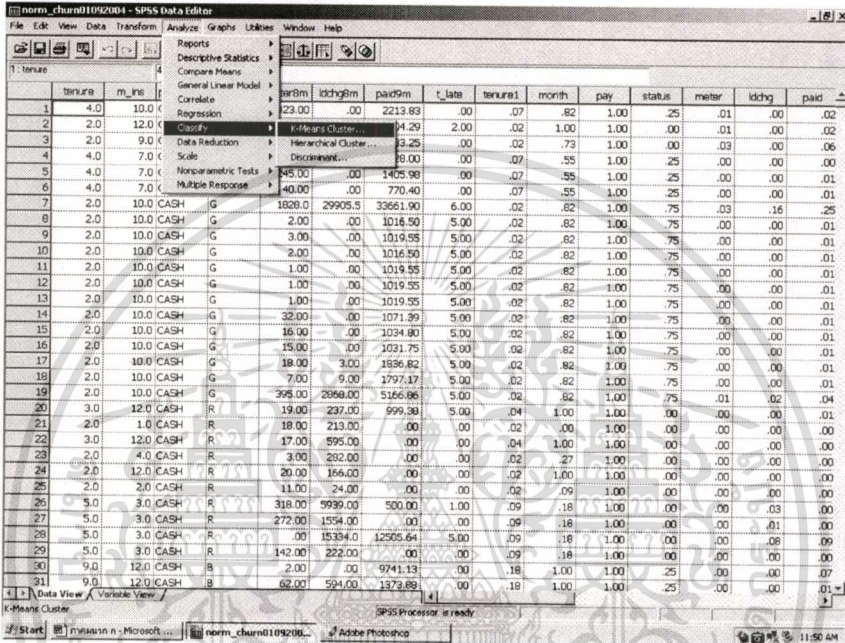
**การใช้งานโปรแกรมสำเร็จรูป SPSS
ในขั้นตอนการจัดกลุ่มลูกค้าโดยใช้ K-Means Clustering**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการจัดกลุ่มลูกค้าโดยใช้ K-Means Clustering ประกอบด้วย

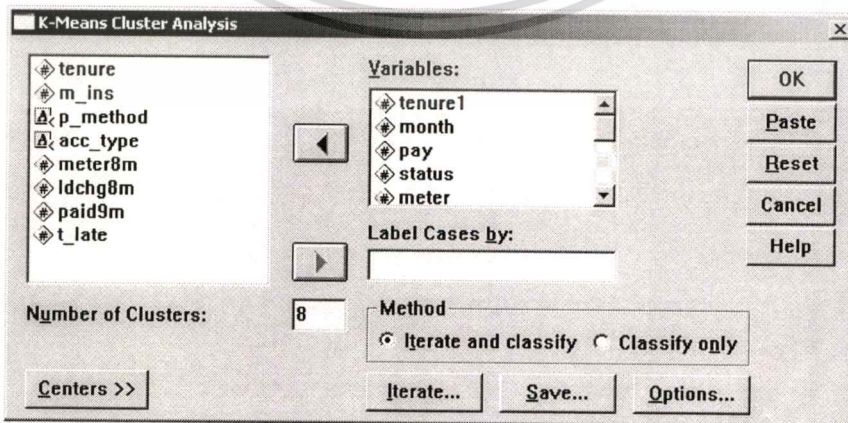
การแบ่งข้อมูลออกเป็นกลุ่มย่อยหลายๆ กลุ่ม โดยภายในกลุ่มย่อยมีลักษณะที่คล้ายกัน โดยใช้คำสั่ง Classify K-Means Cluster

1. Analyze Classify K-Means Cluster จะได้น้ำจอ ดังรูปที่ ก.27



รูปที่ ก.27 แสดงหน้าจอการจัดกลุ่มข้อมูล

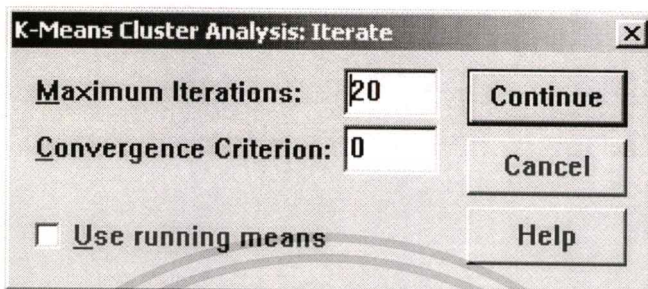
2. เลือกตัวแปรที่เป็นตัวเลข และเลือกจำนวนกลุ่มเริ่มต้น แสดงดังหน้าจอรูปที่ ก.28



รูปที่ ก.28 แสดงหน้าจอการเลือกจำนวน Cluster และเงื่อนไขอื่นๆ

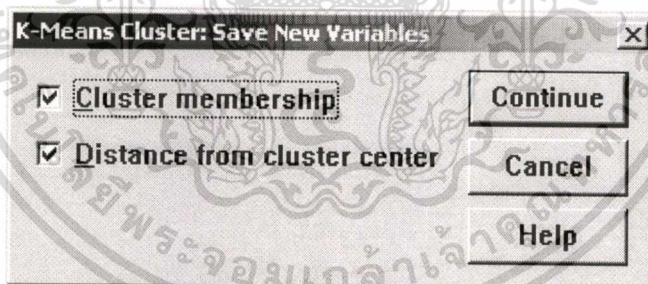
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้งานเพื่อการศึกษาเท่านั้น เมื่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. กำหนดการวนรอบในการจัดกลุ่ม เลือกปุ่ม **Iterate...** จะแสดงหน้าจอรูปที่ ก.29 และเลือกปุ่ม **Continue** กลับมาที่หน้าจอรูปที่ ก.28



รูปที่ ก.29 แสดงหน้าจอการกำหนดการวนรอบการทำ Cluster

4. กำหนดให้บันทึก Cluster Membership และ Distance from Cluster Center เลือกปุ่ม **Save...** จะแสดงหน้าจอรูปที่ ก.30 และเลือกปุ่ม **Continue** กลับมาที่หน้าจอรูปที่ ก.28



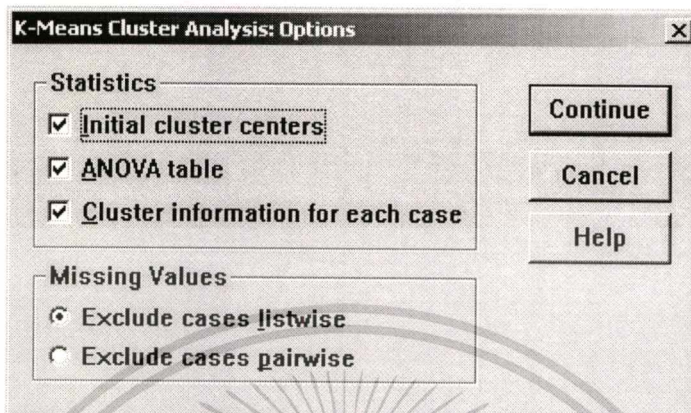
รูปที่ ก.30 แสดงหน้าจอการกำหนด Cluster Membership และ Distance from cluster center

5. กำหนดค่าทางสถิติเลือกปุ่ม **Options...** จะแสดงหน้าจอรูปที่ ก.31 โดยเลือกค่าสถิติเพื่อให้โปรแกรมแสดงผลดังนี้

- Initial Cluster Center ค่าเริ่มต้นของ Cluster
- ANOVA Table
- Cluster information for each case แสดงข้อมูลในแต่ละ Cluster
- Exclude cases listwise ตัดกรณีที่มีค่า Missing Value ของตัวแปรใดๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เลือกปุ่ม **Continue** กลับมาหน้าจอรูปที่ ก.28



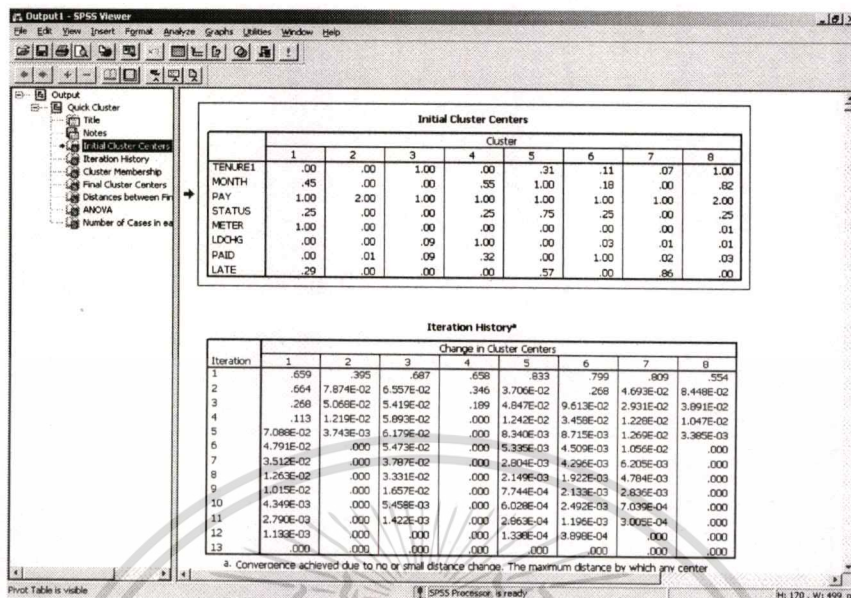
รูปที่ ก.31 แสดงหน้าจอการกำหนดค่าทางสถิติ และการแสดงผลลัพธ์

6. เลือกปุ่ม **OK** โปรแกรมประมวลผลที่ด้านล่างหน้าจอปรากฏคำว่า Running Quick Cluster ดังหน้าจอรูปที่ ก.32
7. ผลลัพธ์ที่ได้จากคำสั่งต่างๆ จะอยู่ที่หน้าจอ Viewer ดังแสดงในรูปที่ ก.33

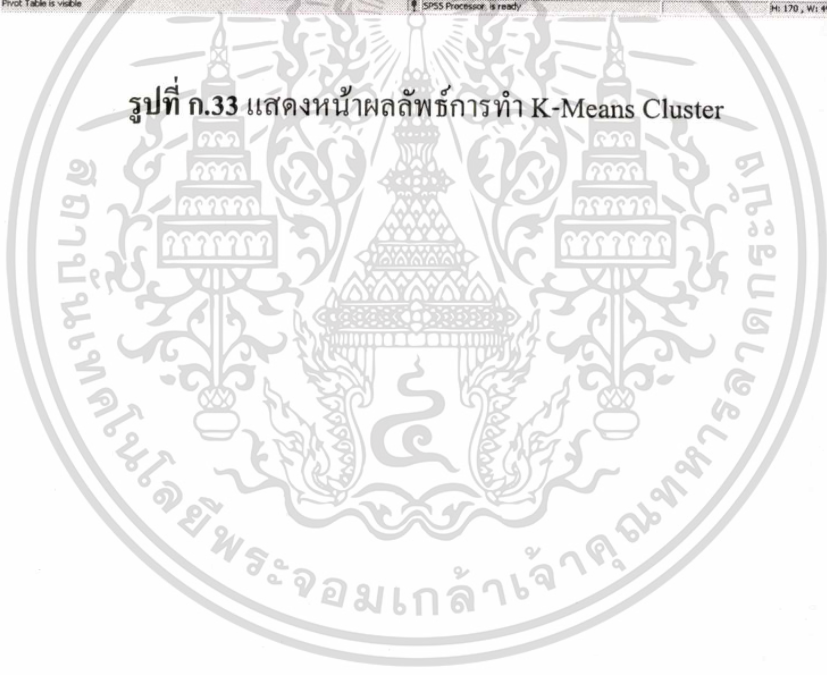
	tenure	tenure1	month	pay	status	meter	kchg	paid							
1	4.0	10.0	CASH	B	423.00	.00	2213.83	2.00	.07	.82	1.00	.25	.01	.00	.02
2	2.0	12.0	CASH	R	874.00	.00	2404.29	2.00	.02	1.00	1.00	.00	.01	.00	.02
3	2.0	9.0	CASH	R	2125.0	.00	8533.25	.00	.02	.73	1.00	.00	.03	.00	.06
4	4.0	7.0	CASH	B	1.00	.00	426.00	.00	.07	.55	1.00	.25	.00	.00	.00
5	4.0	7.0	CASH	B	245.00	.00	1405.96	.00	.07	.55	1.00	.25	.00	.00	.01
6	4.0	7.0	CASH	B	40.00	.00	770.40	.00	.07	.55	1.00	.25	.00	.00	.01
7	2.0	10.0	CASH	G	1828.0	29905.5	33661.90	6.00	.02	.82	1.00	.75	.03	.16	.25
8	2.0	10.0	CASH	G	2.00	.00	1016.50	5.00	.02	.82	1.00	.75	.00	.00	.01
9	2.0	10.0	CASH	G	3.00	.00	1019.55	5.00	.02	.82	1.00	.75	.00	.00	.01
10	2.0	10.0	CASH	G	2.00	.00	1016.50	5.00	.02	.82	1.00	.75	.00	.00	.01
11	2.0	10.0	CASH	G	1.00	.00	1019.55	5.00	.02	.82	1.00	.75	.00	.00	.01
12	2.0	10.0	CASH	G	1.00	.00	1019.55	5.00	.02	.82	1.00	.75	.00	.00	.01
13	2.0	10.0	CASH	G	1.00	.00	1019.55	5.00	.02	.82	1.00	.75	.00	.00	.01
14	2.0	10.0	CASH	G	32.00	.00	1071.39	5.00	.02	.82	1.00	.75	.00	.00	.01
15	2.0	10.0	CASH	G	16.00	.00	1034.80	5.00	.02	.82	1.00	.75	.00	.00	.01
16	2.0	10.0	CASH	G	15.00	.00	1031.75	5.00	.02	.82	1.00	.75	.00	.00	.01
17	2.0	10.0	CASH	G	18.00	3.00	1836.82	5.00	.02	.82	1.00	.75	.00	.00	.01
18	2.0	10.0	CASH	G	7.00	9.00	1797.17	5.00	.02	.82	1.00	.75	.00	.00	.01
19	2.0	10.0	CASH	G	395.00	2868.00	5166.86	5.00	.02	.82	1.00	.75	.01	.02	.04
20	3.0	12.0	CASH	R	19.00	237.00	999.38	5.00	.04	1.00	1.00	.00	.00	.00	.01
21	2.0	1.0	CASH	R	18.00	213.00	.00	.00	.02	.00	1.00	.00	.00	.00	.00
22	3.0	12.0	CASH	R	17.00	595.00	.00	.00	.04	1.00	1.00	.00	.00	.00	.00
23	2.0	4.0	CASH	R	3.00	282.00	.00	.00	.02	.27	1.00	.00	.00	.00	.00
24	2.0	12.0	CASH	R	20.00	166.00	.00	.00	.02	1.00	1.00	.00	.00	.00	.00
25	2.0	2.0	CASH	R	11.00	24.00	.00	.00	.02	.09	1.00	.00	.00	.00	.00
26	5.0	3.0	CASH	R	318.00	5939.00	500.00	1.00	.09	.18	1.00	.00	.00	.03	.00
27	5.0	3.0	CASH	R	272.00	1554.00	.00	.00	.09	.18	1.00	.00	.00	.01	.00
28	5.0	3.0	CASH	R	.00	15334.0	12505.64	5.00	.09	.18	1.00	.00	.00	.00	.00
29	5.0	3.0	CASH	R	142.00	222.00	.00	.00	.09	.18	1.00	.00	.00	.00	.00
30	9.0	12.0	CASH	B	2.00	.00	9741.13	.00	.18	1.00	1.00	.25	.00	.00	.07
31	9.0	12.0	CASH	B	62.00	594.00	1373.88	.00	.18	1.00	1.00	.25	.00	.00	.01

รูปที่ ก.32 แสดงหน้าจอการทำงานของ K-Means Cluster

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ก.33 แสดงหน้าผลลัพธ์การทำ K-Means Cluster



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

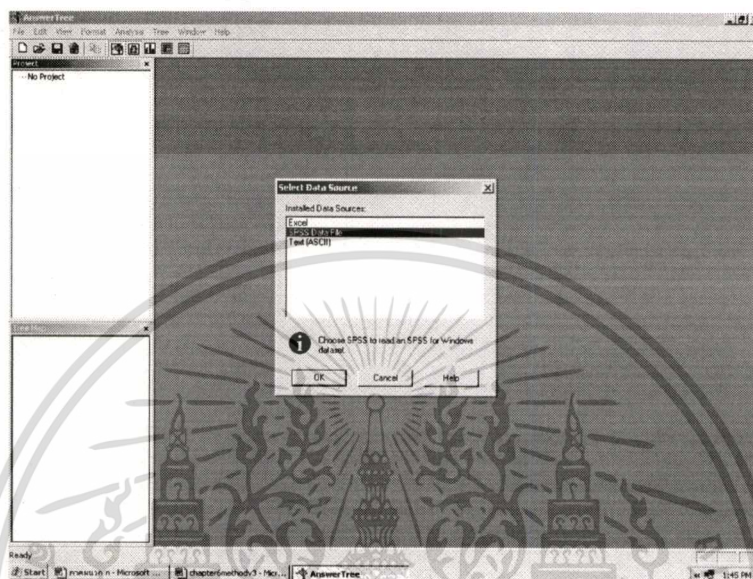


**การใช้งานโปรแกรมสำเร็จรูป Answer Tree
ในขั้นตอนการทำนายลักษณะลูกค้าโดยใช้ CHAID Algorithm**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

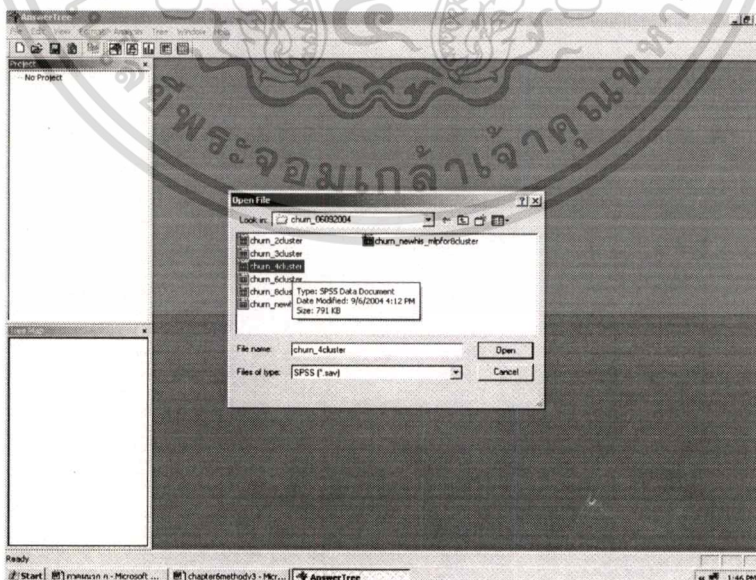
ขั้นตอนในการใช้งานของโปรแกรม Answer Tree

1. เลือกคำสั่ง Start New Project และเลือกประเภทข้อมูลที่ต้องการเรียกใช้ ซึ่งในกรณีนี้จะเรียกใช้ข้อมูลจาก SPSS ดังรูปที่ ก.34



รูปที่ ก.34 แสดงหน้าจอการนำประเภทข้อมูลเข้าสู่โปรแกรม Answer Tree

8. เลือกปุ่ม **OK** ปราบกฏหน้าจอรูปที่ ก.35 เลือกเพิ่มข้อมูลที่จะนำเข้าสู่โปรแกรม

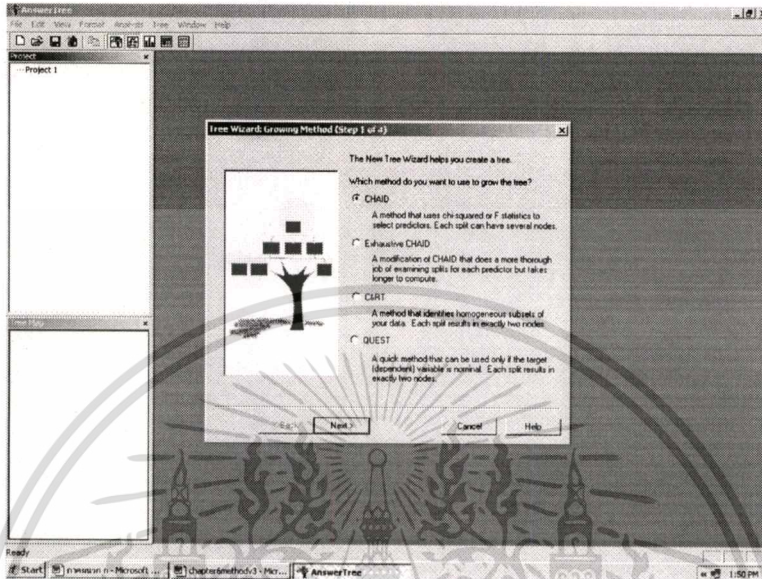


รูปที่ ก.35 แสดงหน้าจอการเลือกไฟล์ข้อมูล

9. เลือกปุ่ม **Open** เปิดเพิ่มข้อมูลที่นำไปใช้

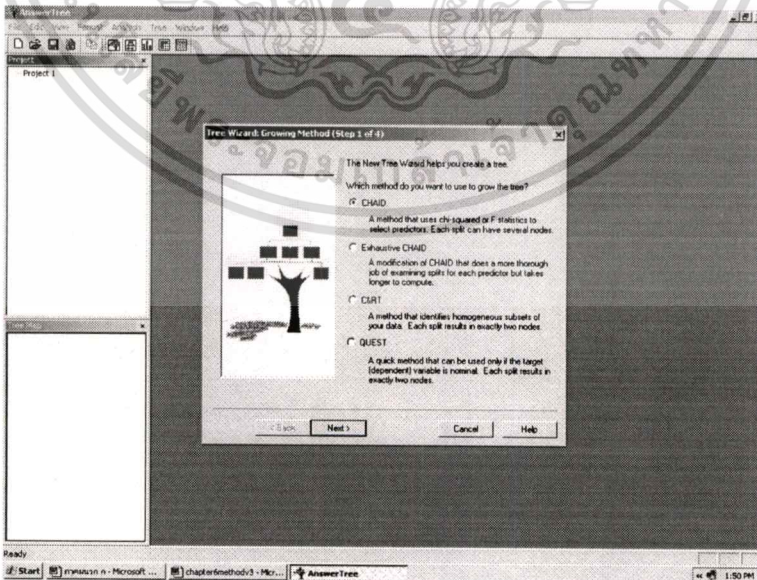
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

10. สร้าง Tree โดยกำหนด Algorithm หรือ Method ในโปรแกรม ซึ่งมีทั้งหมด 4 วิธี สำหรับการศึกษานี้เลือก CHAID ในการสร้าง Tree เป็น Step 1 ดังรูปที่ ก.36



รูปที่ ก.36 แสดงหน้าจอการเลือก Algorithm

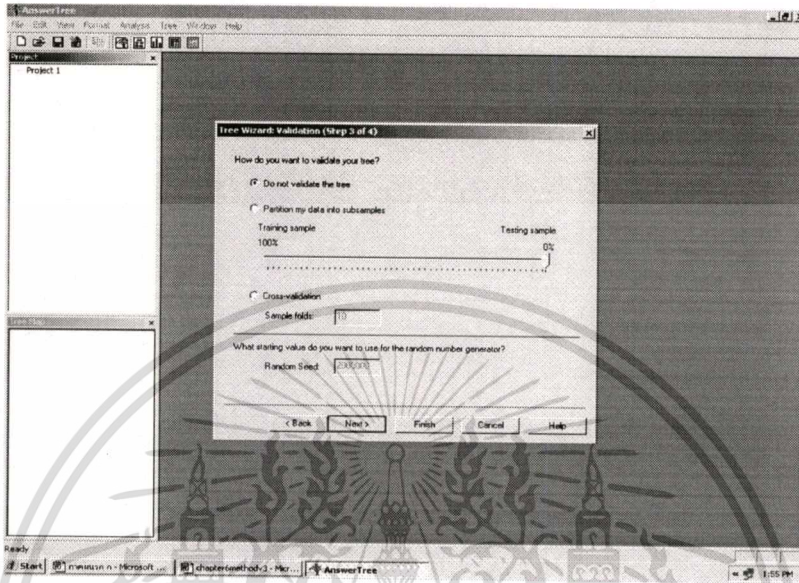
11. เลือกปุ่ม **Next >** ปราบกฏหน้าจอให้กำหนด Target Attribute เลือก Cluster Number of case เป็น Step 2 แสดงดังรูปที่ ก.37



รูปที่ ก.37 แสดงหน้าจอการกำหนด Target Attribute

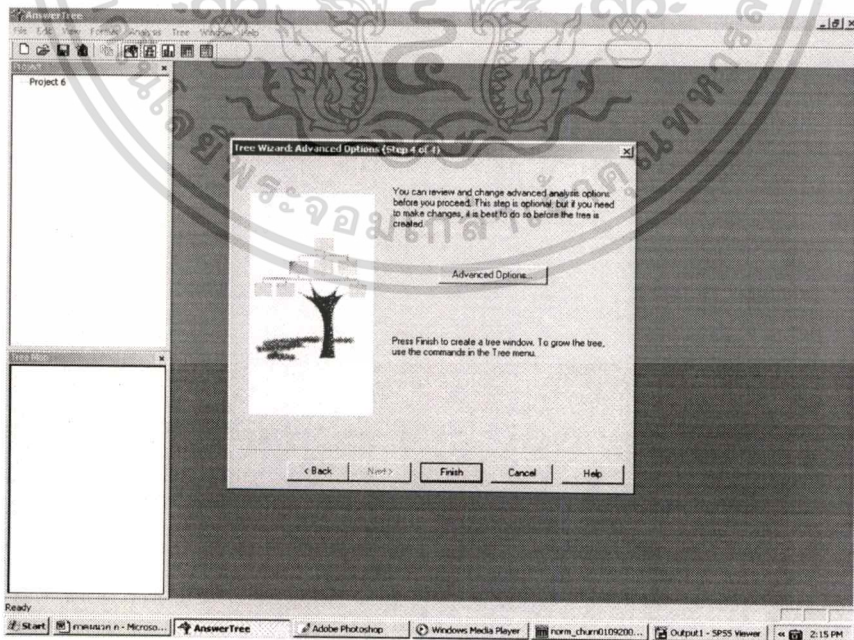
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

12. เลือกปุ่ม **Next >** ปราบกฏหน้าจอให้กำหนด Validation ของข้อมูล ไม่แบ่งข้อมูล
เพื่อการ Training และ Testing ใช้วิธีชูปกิ่งโดยวิธีเปลี่ยนแปลงค่าในแต่ละ Node แสดงดังรูปที่ ก.38



รูปที่ ก.38 แสดงหน้าจอการกำหนดข้อมูล Training และ Testing

13. เลือกปุ่ม **Next >** ปราบกฏหน้าจอการกำหนด Advanced Options ดังรูปที่ ก.39

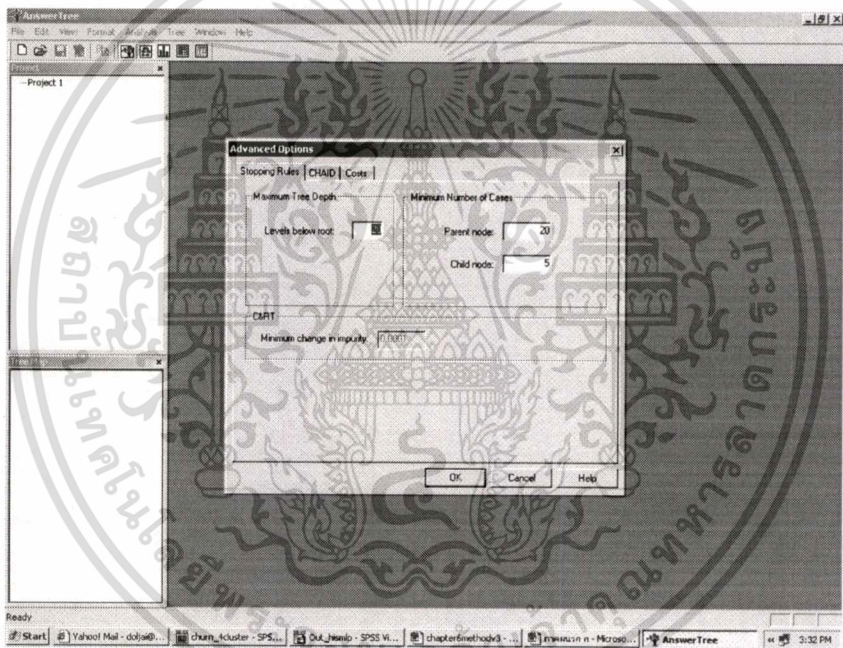


รูปที่ ก.39 แสดงหน้าจอการกำหนด Advanced Option

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

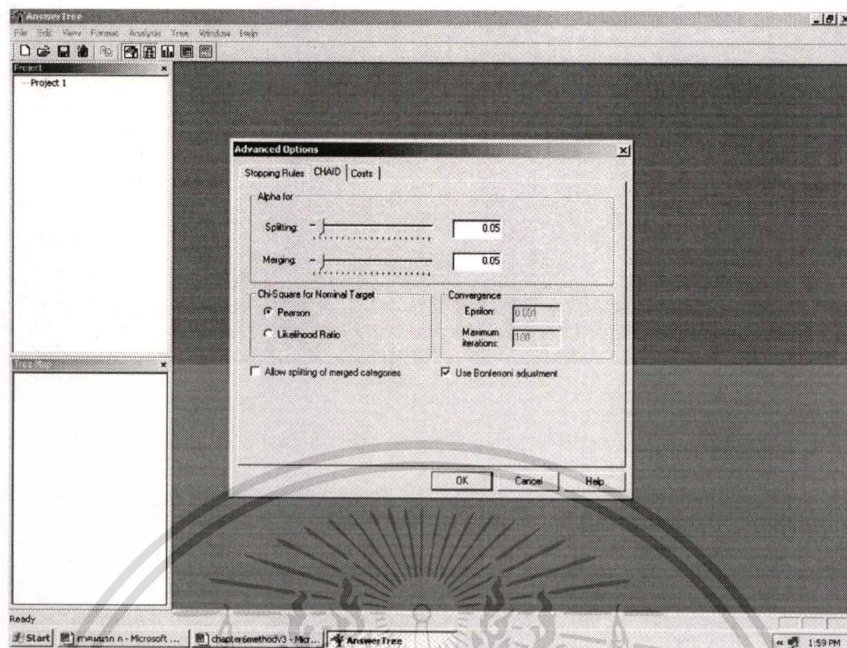
14. เลือกปุ่ม **Advanced Options...** ปรากฏหน้าจอตั้งรูปที่ ก.40 และรูปที่ ก.41 ในการกำหนดเงื่อนไขดังนี้

- กำหนดการหยุดแตกกิ่ง หรือสร้าง Tree เนื่องจากข้อมูลที่อยู่ใน Node ข้างล่างลงไปอาจมีจำนวนข้อมูลที่น้อย และไม่สามารถแสดงความสัมพันธ์ได้เพียงพอ รวมทั้งกำหนดค่า Parent Node และ Child Node
- ค่า Chi Square เลือกค่า Alpha for splitting and merging เท่ากับ 0.5 และกำหนดค่า Chi - Square for Nominal Target เป็น Pearson รวมทั้งเลือก Bonferroni Adjustment



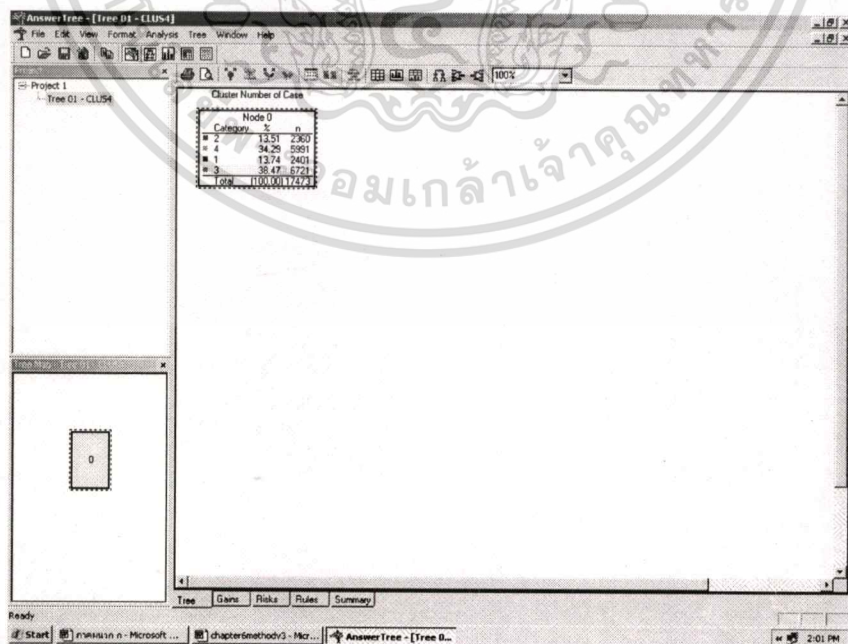
รูปที่ ก.40 แสดงหน้าจอการกำหนดเงื่อนไขในการหยุดและParent Node , Child Node

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ก.41 แสดงหน้าจอการกำหนดเงื่อนไขในการหยุด Alpha for splitting and merging

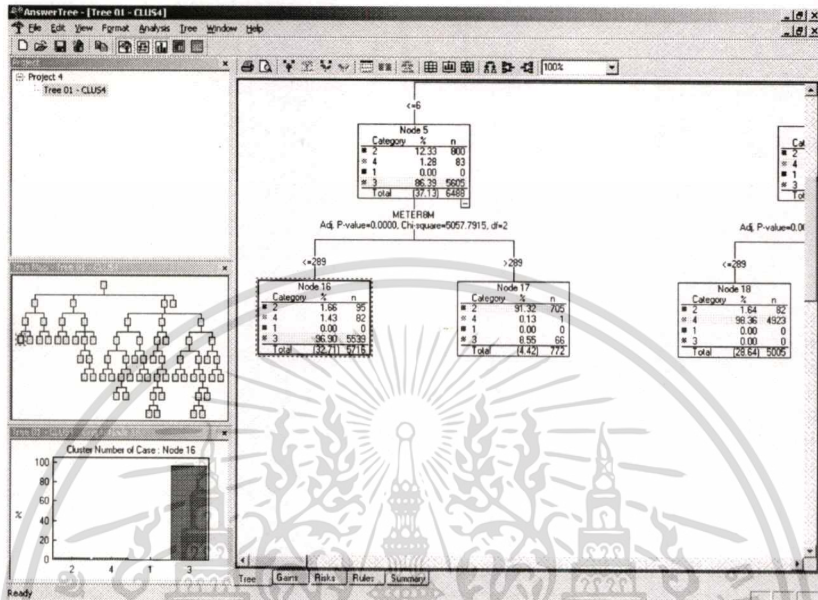
15. เลือกปุ่ม **OK** กลับไปที่หน้าจอรูปที่ ก.40 เลือกปุ่ม **Finish** ปรากฏหน้าจอรูปที่ ก.42



รูปที่ ก.42 แสดงหน้าจอ Node แรก ก่อนการแตกกิ่ง

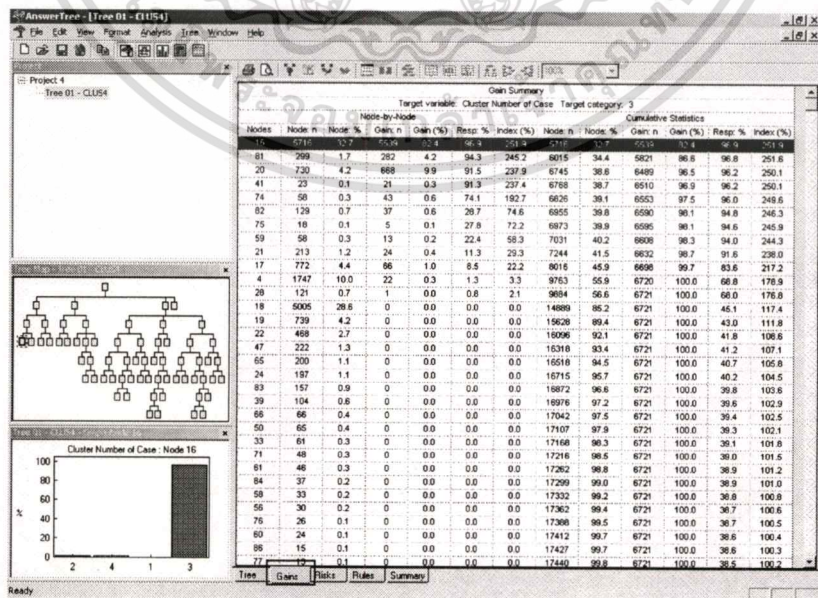
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรศึกษาเพื่อการศึกษาเท่านั้น ไม่นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

16. เลือกเมนู Grow Tree หรือ Grow Branch เพื่อแตกกิ่งข้อมูลตามที่โปรแกรมจัดการดังรูปที่ ก.43



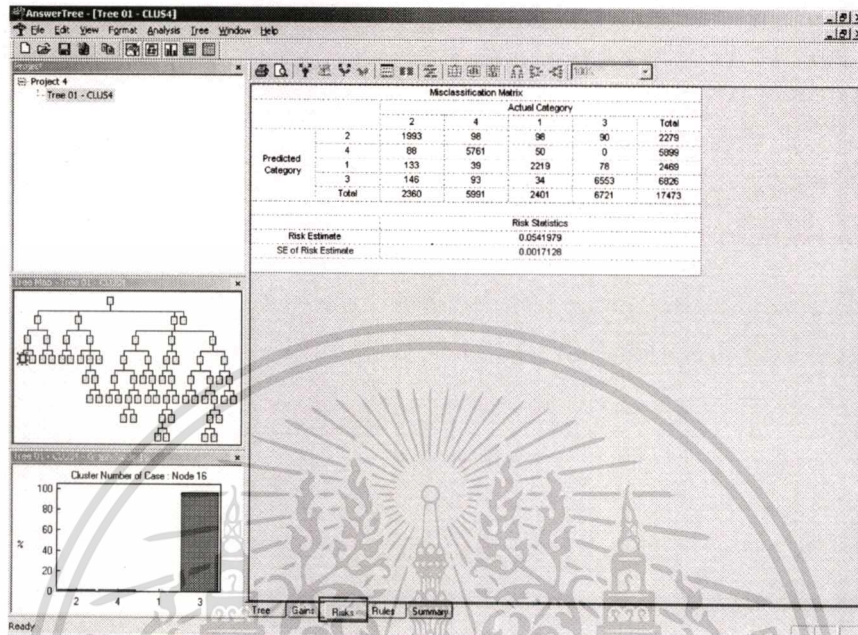
รูปที่ ก.43 แสดงหน้าจอการแตกกิ่ง

17. เลือกเมนูข้างล่างจอ Gains แสดงข้อมูลที่ได้อากการสร้าง Tree ดังรูปที่ ก.44



เอกสารนี้เป็นเอกสารที่สงวนรูปที่ ก.44 แสดงหน้าจอข้อมูลที่ได้จากการสร้าง Tree มาไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

18. เลือกเมนูข้างล่างจ้อ Risk แสดงค่าความเสี่ยงในการพยากรณ์ผิดพลาดดังรูปที่ ก.45



รูปที่ ก.45 แสดงหน้าจอค่าความเสี่ยงในการพยากรณ์การสร้าง Tree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ – นามสกุล	นางสาวคลใจ กระแสเสน
วัน เดือน ปีเกิด	21 ก.ค. 2512
สถานที่เกิด	จังหวัดสระบุรี
ประวัติการศึกษา	ปริญญาตรีวิทยาศาสตร์บัณฑิต สาขาเศรษฐศาสตร์ มหาวิทยาลัยเกษตรศาสตร์
ประวัติการทำงาน	ฝ่ายตรวจสอบภายใน บริษัท ทศท คอร์ปอเรชั่น จำกัด (มหาชน) ตรวจสอบงานบริหาร พ.ศ. 2536 – 2539 ตรวจสอบงานคอมพิวเตอร์ พ.ศ. 2539 – ปัจจุบัน



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้