ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

# การศึกษาปัจจัยที่แสดงแนวโน้มจะเป็นหนี้ที่ไม่ก่อให้เกิดรายได้ โดยใช้ Predictive Model

**Predictive Model for Non Performing Loan Trend**

โดย

**นางสาวปรียากร สัปตา**

**รหัสนักศึกษา 43067254**

**อาจารย์ที่ปรึกษา**

**ผศ. ดร. วรพจน์ กรีสุระเดช**

*H002922*

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการศึกษากรณีพิเศษ
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
ภาคเรียนที่ 2 ปีการศึกษา 2545
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

วัน เดือน ปี 02 ม.ค. 2550
เลขทะเบียน 02922
เลขเรียกหนังสือ วท.ม. ป 473ก 2545
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ชื่อหัวข้อ** การศึกษาปัจจัยที่แสดงแนวโน้มจะเป็นหนี้ที่ไม่ก่อให้เกิดรายได้ โดยใช้ Predictive Model

**นักศึกษา** นางสาวปรียากร สัปตา

**อาจารย์ที่ปรึกษา** ผศ. ดร. วรพจน์ กรีสุระเดช

**ระดับการศึกษา** วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

**แขนงวิชา** การจัดการเทคโนโลยีสารสนเทศ

**ปีการศึกษา** 2545

## บทคัดย่อ

จากการเกิดวิกฤตเศรษฐกิจของประเทศตั้งแต่ปี 2540 ทำให้เศรษฐกิจอยู่ในภาวะซบเซาอย่างต่อเนื่อง ส่งผลให้ลูกหนี้ของสถาบันการเงินจำนวนมากประสบปัญหาจากการดำเนินงานจนไม่สามารถชำระหนี้ได้ตามกำหนดและกลายเป็นหนี้ด้อยคุณภาพหรือหนี้ที่ไม่ก่อให้เกิดรายได้ Non-Performing Loans (NPLs) ซึ่งผลต่อกระทบผลการดำเนินงานของสถาบันการเงินและภาคธุรกิจอย่างต่อเนื่อง

ดังนั้นเพื่อศึกษาถึงปัจจัยซึ่งมีผลต่อแนวโน้มที่จะเป็น NPLs โดยอาศัยขั้นตอนของ Data Mining โดยใช้เทคนิค Predictive มาใช้ในการศึกษาถึงรูปแบบความสัมพันธ์โดยอาศัยข้อมูลจากงบการเงินของลูกหนี้ที่เป็น NPLs ของตลาดหลักทรัพย์ในปัจจุบันเพื่อสามารถนำข้อมูลดังกล่าวมาวิเคราะห์ถึงแนวโน้มที่ลูกหนี้จะเป็น NPLs และหาทางป้องกันหรือแก้ไขปัญหาของลูกหนี้ก่อนที่จะเป็น NPLs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**Title** Predictive Model for Non-Performing Loans Trend

**Student** Miss Preeyakorn Subta

**Advisor** Asst. Prof. Dr.Worapoj Greesuradej

**Level of Study** Master of Science in Information Technology

**Major** Information Technology Management

**Academic Year** 2002

# ABSTRACT

According to the economic crisis known as Tom Yum Kong crisis in 1996, Thailand is continuing facing an economic set back. The crisis caused an inability to service debt in the private sector and they became Non-

Non - Performing Loans (NPLs). Therefore, this NPLs directly affect the performance of both financial sector and private sector until now.

This study will determine factors that most likely to create NPLs by using the process of Data Mining and Predictive technique based on the financial statement of existing NPLs companies in the Stock Exchange of Thailand. The result will direct us the way to prevent or how to help them overcome the problem before becoming NPLs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# กิตติกรรมประกาศ

โครงการศึกษานี้ สำเร็จลุล่วงได้ ข้าพเจ้าขอขอบพระคุณ ผศ. ดร. วรพจน์ กรีสุระเดช อาจารย์ที่ปรึกษาวิชาโครงการศึกษาที่ได้กรุณาสละเวลาให้ความรู้ คำปรึกษา และคำแนะนำต่างๆ อันเป็นประโยชน์ต่อการศึกษาโครงการนี้

นอกจากนี้ ข้าพเจ้าขอกราบขอบพระคุณบุพการี และบุคคลในครอบครัวที่ได้ให้การสนับสนุนส่งเสริมเป็นกำลังใจในการเรียนตลอดมา รวมทั้งเพื่อนร่วมงานและเพื่อนร่วมรุ่น ITM8 ภาคสมทบที่มีส่วนให้ความช่วยเหลือ เป็นกำลังใจและสนับสนุนให้โครงการศึกษาฉบับนี้สำเร็จลุล่วงด้วยดี

ข้าพเจ้าหวังเป็นอย่างยิ่งว่าโครงการศึกษาฉบับนี้ จะเป็นประโยชน์แก่ผู้ที่สนใจ สำหรับข้อบกพร่องของโครงการศึกษานี้ ข้าพเจ้าขอรับไว้ เพื่อนำไปปรับปรุงแก้ไขในโอกาสต่อไป

ปรียากร สัปตา

กุมภาพันธ์ 2546

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ


| | หน้า |
|---|---|
| บทคัดย่อภาษาไทย | I |
| บทคัดย่อภาษาอังกฤษ | II |
| กิตติกรรมประกาศ | III |
| สารบัญ | IV |
| สารบัญตาราง | VI |
| สารบัญภาพ | VII |
| บทที่ | |
| 1. บทนำ | |
| 1.1 หลักการและเหตุผล | 1 |
| 1.2 วัตถุประสงค์ของการศึกษา | 1 |
| 1.3 ขอบเขตของการศึกษา | 2 |
| 1.4 ขั้นตอนและวิธีการดำเนินงาน | 2 |
| 1.5 ประโยชน์ที่คาดว่าจะได้รับ | 2 |
| 2. หนี้ NPLs | |
| 2.1 ความหมายของ Non-Performing Loans หรือ NPLs | 3 |
| 2.2 สาเหตุการเกิด NPLs | 3 |
| 2.3 ผลกระทบ NPLs ต่อเศรษฐกิจ | 5 |
| 3. ดาต้าไมนิ่งและทฤษฎีที่เกี่ยวข้อง | |
| 3.1 ดาต้าไมนิ่ง | 8 |
| 3.2 ขั้นตอนในการทำดาต้าไมนิ่ง | 8 |
| 3.3 ลักษณะของดาต้าไมนิ่ง | 11 |
| 4. การพยากรณ์ (Predictive) | |
| 4.1 การพยากรณ์ (Predictive) | 13 |
| 4.2 CHAID Algorithm | 14 |


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ(ต่อ)


| | | หน้า |
|---|---|---|
| บทที่ | | |
| 5 วิธีการดำเนินการศึกษา | | |
| | 5.1 ข้อมูลที่ศึกษา | 23 |
| | 5.2 การทำดาตาไมนิ่ง | 27 |
| | 5.3 การวิเคราะห์ผลลัพธ์ที่ได้ | 29 |
| 6 สรุปผลการศึกษาและข้อเสนอแนะ | | |
| | 6.1 บทสรุป | 37 |
| | 6.2 ปัญหาและอุปสรรค | 37 |
| บรรณานุกรม | | 38 |
| ภาคผนวก ก | | 39 |
| ประวัติผู้เขียน | | 47 |


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญตาราง


| ตารางที่ | | หน้า |
|---|---|---|
| 3.1 | แสดงโมเดลของ Data Mining กับการประยุกต์ใช้งาน | 12 |
| 4.1 | แสดง Contingentcy ของ การคำนวณ | 15 |
| 4.2 | แสดงการกระจายของข้อมูลตัวแปร Y ใน ROOT NODE | 16 |
| 4.3 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X1 | 16 |
| 4.4 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X1 กรณี 1 และ 2 | 16 |
| 4.5 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X1 กรณี 1 และ 3 | 17 |
| 4.6 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X1 กรณี 1 และ 4 | 17 |
| 4.7 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X1 กรณี 2 และ 3 | 17 |
| 4.8 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X1 กรณี 2 และ 4 | 17 |
| 4.9 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X1 กรณี 3 และ 4 | 18 |
| 4.10 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X1 กรณี 1 และ 3, 4 | 18 |
| 4.11 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X1 กรณี 2 และ 3, 4 | 18 |
| 4.12 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X1 กรณี 1 และ 2 | 19 |
| 4.13 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X1 กรณี 1 และ 2,3,4 | 19 |
| 4.14 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X2 | 20 |
| 4.15 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X2 กรณี 0 และ 1 | 20 |
| 4.16 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X2 กรณี 0 และ 2 | 20 |
| 4.17 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X2 กรณี 1 และ 2 | 21 |
| 4.18 | แสดงค่าความสัมพันธ์ระหว่าง Y และ X2 กรณี 0, 1 และ 2 | 21 |
| 5.1 | แสดงความสัมพันธ์ระหว่างชื่อของข้อมูลกับประเภทข้อมูล | 26 |


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญภาพ


| ภาพที่ | | หน้า |
|---|---|---|
| 2.1 | แสดงแผนภูมิแก้ไข NPLs ของไทย | 6 |
| 3.1 | แสดงเปอร์เซ็นต์การทำงานในแต่ละขั้นตอนในการทำ Data Mining | 11 |
| 4.1 | แสดงภาพ Decision Tree | 13 |
| 4.2 | แสดง Tree ของข้อมูลจากการใช้ CHAID Algorithm | 22 |
| 5.1 | แสดงหน้าจอของ Tree ที่สร้างจาก Answer Tree | 30 |
| 5.2 | แสดง Tree ที่ได้จากการใช้ตัวแปรทั้งหมดกรณีที่ 1.1 | 31 |
| 5.3 | แสดง Tree ที่ได้จากการสร้างตัวแปรทั้งหมดกรณีที่ 1.2 | 32 |
| 5.4 | แสดงหน้าจอของ Tree ที่สร้างจากตัวแปรงบดุล | 33 |
| 5.5 | แสดง Tree ที่ได้จากการใช้ตัวแปรจากงบดุลกรณีที่ 2.1 | 33 |
| 5.6 | แสดง Tree ที่ได้จากการใช้ตัวแปรจากงบดุลกรณีที่ 2.2 | 34 |
| 5.7 | แสดงหน้าจอการสร้าง Tree จากตัวแปรจากงบกำไร(ขาดทุน) | 36 |
| 5.8 | แสดง Tree ที่ได้จากการใช้ตัวแปรจากงบกำไร(ขาดทุน) | 36 |
| ก.1 | แสดงหน้าจอการนำประเภทข้อมูลเข้าสู่โปรแกรม Answer Tree | 41 |
| ก.2 | แสดงหน้าจอการเลือกไฟล์ข้อมูลเพื่อเข้าสู่โปรแกรม Answer Tree | 41 |
| ก.3 | แสดงหน้าจอการเลือกไฟล์ข้อมูลที่จะนำไปใช้ | 42 |
| ก.4 | แสดงหน้าจอการเลือก Algorithm ที่ต้องการใช้สร้าง Tree | 42 |
| ก.5 | แสดงหน้าจอการเลือกตัวแปรที่จะนำเข้าโปรแกรม | 43 |
| ก.6 | แสดงหน้าจอการเลือก Options ในการสร้าง Tree | 43 |
| ก.7 | แสดงหน้าจอเปอร์เซ็นต์การกำหนด Test และ Train | 44 |
| ก.8 | แสดงหน้าจอการกำหนด Options ในการสร้าง Tree | 44 |
| ก.9 | แสดงหน้าจอ Node แรก ก่อนการแตกกิ่งของ Tree | 45 |
| ก.10 | แสดงหน้าจอ Tree ที่สร้างเสร็จแล้ว | 45 |


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญภาพ(ต่อ)

หน้า

ภาพที่

ก.11 แสดงหน้าจอข้อมูลที่ได้จากการสร้าง Tree 46

ก.12 แสดงหน้าจอค่าความเสี่ยงในการพยากรณ์จากการสร้าง Tree 46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 1
# บทนำ

## 1.1 หลักการและเหตุผล

จากภาวะเศรษฐกิจที่ชะลอตัวตั้งแต่ปี 2540 ส่งผลให้ธุรกิจในประเทศไทยได้รับผลกระทบอย่างมาก จนก่อให้เกิดปัญหาในการชำระหนี้จนลูกหนี้ส่วนใหญ่ไม่สามารถชำระหนี้ได้ และกลายเป็นหนี้ที่ไม่ก่อให้เกิดรายได้ หรือ NPLs (Non-Performing Loans) :ซึ่งส่งผลกระทบต่อการดำเนินงานของสถาบันการเงินและส่งผลกระทบเป็นลูกโซ่ก่อให้เกิดปัญหาของประเทศชาติ

หนี้ที่ไม่ก่อให้เกิดรายได้ หรือ NPLs จึงเป็นปัญหาที่แต่ละสถาบันการเงินให้ความสำคัญในการแก้ไขโดยใช้วิธีต่างๆในการเจรจาหรือปรับโครงสร้างหนี้ให้กับลูกหนี้ซึ่งเป็น NPLs เหล่านั้น รวมทั้งภาครัฐก็ได้ให้ความสำคัญจัดตั้งหน่วยงานเพื่อมาช่วยในการเร่งการปรับโครงสร้างหนี้เพื่อให้ความต้องการที่แตกต่างระหว่างสถาบันเงินที่ติดตามหนี้กับลูกหนี้ที่ต้องรับภาระหนี้ให้สามารถตกลงในแนวทางที่ยอมรับกันทั้งสองฝ่าย ซึ่งในการแก้ไขปัญหาหนี้ดังกล่าวต้องใช้ทรัพยากรทั้งบุคคลากร ระยะเวลาในการดำเนินการต่างๆ อีกทั้งในบางโครงการการเจรจาอาจได้ข้อยุติที่พอใจทั้งสองฝ่ายหรือในทางตรงข้ามบางรายอาจต้องใช้ขั้นตอนของกฎหมาย ดังนั้นหากมีการศึกษาถึงรูปแบบของโครงการที่เป็นหนี้ไม่ก่อให้เกิดรายได้หรือ NPLs เพื่อใช้ในการ Predictive ได้ล่วงหน้าว่ารูปแบบใดจะมีแนวโน้มเป็น NPLs เพื่อหาทางป้องกันหรือวิธีการแก้ไขปัญหาให้กับลูกหนี้ได้ ซึ่งเป็นที่มาของรายงานฉบับนี้ โดยในรายละเอียดจะนำข้อมูลทางการเงินโดยใช้งบการเงินของบริษัทที่อยู่ในตลาดหลักทรัพย์มาใช้ในการวิเคราะห์ รวมทั้งพิจารณาถึงสิ่งแวดล้อมที่เกี่ยวข้อง

## 1.2 วัตถุประสงค์ของการศึกษา

1) เพื่อศึกษาถึงรูปแบบของธุรกิจซึ่งเป็นหนี้ที่ไม่ก่อให้เกิดรายได้รวมถึงสภาวะแวดล้อมที่เกี่ยวข้อง
2) เพื่อเป็นการช่วยในการป้องกันและแก้ไขปัญหาก่อนที่จะเป็นลูกหนี้ที่มีปัญหา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**1.3 ขอบเขตของการศึกษา**

1) ศึกษาและรวบรวมจากข้อมูลจากงบการเงินของบริษัทในตลาดหลักทรัพย์ แล้วนำมาวิเคราะห์ข้อมูลต่อไป
2) การศึกษาจะครอบคลุมเฉพาะปี 2543 เนื่องจากคาดว่าเป็นปีที่มีหนี้เป็น NPLs จำนวนมาก

**1.4 ขั้นตอนและวิธีการดำเนินงาน**

1) ศึกษาค้นคว้าข้อมูลที่เกี่ยวข้องกับ NPLs จากหนังสือ วารสาร หนังสือพิมพ์และอินเตอร์เน็ต
2) เก็บรวบรวมข้อมูล โดยใช้ข้อมูลจากงบการเงินของบริษัทในตลาดหลักทรัพย์เนื่องจากงบการเงินในตลาดหลักทรัพย์ได้จัดทำให้อยู่ในมาตรฐานเดียวกันและมีผู้สอบบัญชีได้รับอนุญาตรับรองความถูกต้อง รวมทั้งเป็นข้อมูลที่เปิดเผยต่อบุคคลทั่วไปด้วย
3) สรุปผลการศึกษาของข้อมูล Data Mining จากการทำ Predictive

**1.5 ประโยชน์ที่คาดว่าจะได้รับ**

1) ศึกษารูปแบบของโครงการที่เป็นลูกหนี้ด้อยคุณภาพเพื่อนำมา Predictive ข้อมูลงบการเงินในปัจจุบัน สำหรับหาทางป้องกันหรือแก้ไขลูกหนี้ก่อนที่จะเป็นหนี้ที่มีปัญหา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 2
# หนี้ NPLs

## 2.1 ความหมายของ Non-Performing Loan s หรือ NPLs

Non-Performing Loans หรือที่เรียกว่ากันว่า NPLs (หนี้ที่ไม่ก่อให้เกิดรายได้) หมายถึง เงินที่ให้สินเชื่อที่ค้างชำระต้นเงินและหรือดอกเบี้ย นับจากวันที่ครบกำหนดชำระเงินตามงวดที่กำหนดในสัญญา สำหรับสินเชื่อที่ต้องชำระเมื่อทวงถามหรือแจ้งให้ชำระหนี้ และได้ทวงถามหรือเรียกให้ชำระหนี้แล้วเป็นระยะเวลาเกินกว่า 3 เดือน โดยพิจารณาการค้างชำระเป็นรายสัญญาหรือรายบัญชี

## 2.2 สาเหตุการเกิด NPLs

มีหลายสาเหตุที่ทำให้เกิดวิกฤติ NPLs ดังนี้

1) **สาเหตุจากปัจจัยภายนอก** เป็นปัจจัยที่ผลักดันจากวิกฤติการณ์ทางเศรษฐกิจที่ส่งผลให้ธุรกิจทางการเงินและการค้าตกต่ำ ในขณะที่อัตราดอกเบี้ยสูง และอัตราแลกเปลี่ยนผันผวนโดยอัตราการเจริญเติบโตของเศรษฐกิจเริ่มมีแนวโน้มลดลงร้อยละ 9 ในปี 2537 และลดลงมาตามลำดับเป็นร้อยละ 5.9 ในปี 2539 และติดลบร้อยละ 10.4 ในปี 2541 ในขณะที่อัตราดอกเบี้ยเงินให้กู้ยืมร้อยละ 11.75 ต่อปี เป็นร้อยละ 13.75 ต่อปี ในปี 2537 และ 2539 ตามลำดับ ในขณะเดียวกันอัตราดอกเบี้ยในต่างประเทศยังอยู่ในระดับต่ำกว่าภายในประเทศมาก โดยอัตราดอกเบี้ย LIBOR ระยะเวลา 3 เดือนในช่วงปี 2537-2539 นั้น มีอัตราเพียงร้อยละ 3.25-6.5 ต่อปีเท่านั้น

จากส่วนต่างของอัตราดอกเบี้ยต่างประเทศและในประเทศดังกล่าว ทำให้ธุรกิจภายในประเทศต้องการแหล่งเงินเงินกู้ที่ต้นทุนที่ต่ำประกอบกับธุรกิจภายในประเทศอยู่ในช่วงที่มีการเติบโตสูงมากทำให้มีความต้องการเงินทุนสูง จึงมีการกู้ยืมจากต่างประเทศมากขึ้น โดยเพิ่มจาก 25 พันล้านดอลลาร์สหรัฐอเมริกา ในปี 2534 เป็น 74 พันล้านดอลลาร์สหรัฐอเมริกา ในปี 2539 หรือมากกว่า 2 เท่าในระยะเวลา 6 ปี

ในขณะเดียวกันกับสินเชื่อในประเทศได้ลดต่ำลงจากที่เคยขยายตัวสูงสุดร้อยละ 30 ในปี 2537 เป็นร้อยละ14 ในปี 2539 โดยประเทศไทยได้อาศัยเงินทุนจากภายนอกมาชดเชยเป็นจำนวนสูงมาก คือเป็นสินเชื่อของวิเทศธนกิจได้เพิ่มขึ้นจาก 197 พันล้านบาท ในปี 2536 เป็น 497 พันล้านบาท 686 พันล้านบาท และ 808 พันล้านบาท ในปี 2537 ปี 2538 และในปี 2539 ตามลำดับ ส่งผลให้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บัญชีเดินสะพัดขาดดุลเพิ่มขึ้นอย่างรวดเร็วจากร้อยละ 5 ของผลิตภัณฑ์ในประเทศ(GDP) ในปี 2535 เป็นร้อยละ 8 ในปี 2538 ตามลำดับ กดดันให้รัฐบาลต้องประกาศลอยตัวค่าเงินบาทเมื่อเดือนกรกฎาคม 2540 ลูกหนี้ที่ใช้เงินตราต่างประเทศมีภาระหนี้เพิ่มขึ้นอีก 2 เท่าตัวภายหลังการแปลงหนี้เป็นเงินบาท ซึ่งส่งผลกระทบต่อโครงสร้างทางการเงินของลูกหนี้จำนวนมากเนื่องจากรายได้จากการดำเนินธุรกิจไม่เพียงพอต่อภาระหนี้ที่เพิ่มขึ้น

2) **สาเหตุจากปัญหาภายในประเทศ** ที่เกิดจากการดำเนินงานและการบริหารของสถาบันการเงินและของลูกหนี้ ซึ่งปัญหาหลักของสถาบันการเงินในช่วงที่เฟื่องฟูนั้น ธุรกิจมีการแข่งขันกันมาก ส่งผลให้มีการกระจุกตัวของการปล่อยสินเชื่อในบางอุตสาหกรรม การปล่อยสินเชื่อขาดความระมัดระวัง ปล่อยสินเชื่อด้วยความเกรงใจ และขาดการติดตามอย่างใกล้ชิด ดังนั้นเมื่อเกิดวิกฤติทางเศรษกิจ จึงทำให้สินเชื่อกลายเป็น NPLs อย่างที่ไม่เคยเกิดขึ้นมาก่อน

ในส่วนของภาคธุรกิจนั้นได้มีการแข่งขันกันอย่างรุนแรงเช่นเดียวกัน ประกอบกับเป็นช่วงที่ประเทศไทยเปิดการค้าเสรีทางการเงิน จึงทำให้มีเงินทุนดอกเบี้ยต่ำจากต่างประเทศไหลเข้าในประเทศซึ่งมีอัตราดอกเบี้ยที่สูงกว่า ในขณะที่รัฐบาลมีนโยบายอัตราแลกเปลี่ยนคงที่ จึงจูงใจให้มีการนำเข้าเงินตราต่างประเทศเพื่อแสวงหารายได้ส่วนต่างของอัตราดอกเบี้ย และลงทุนเก็งกำไรในธุรกิจต่างๆอย่างมาก โดยมีการลงทุนทั้งการผลิตและธุรกิจอสังหาริมทรัพย์เกินความต้องการ และสัดส่วนหนี้สินต่อทุนได้เพิ่มจากร้อยละ 8 ในปี 2536 เป็นร้อยละ 155 ในปี 2539 ซึ่งในธุรกิจของไทยมีสัดส่วนหนี้สินต่อทุนสูงสุดเมื่อเปรียบเทียบกับบประเทศอื่นๆ เช่น อินโดนีเซีย ร้อยละ 92 มาเลเซีย ร้อยละ 62 สิงค์โปร์ ร้อยละ 58 และฮ่องกง ร้อยละ 39 โดยเฉพาะการสร้างอาคารสูงเพื่อสนองอุปสงค์เทียม (artifical demand) นั้นมีมากเกินความต้องการจริงๆ(effective demand) และเมื่อเกิดวิกฤติอาคารสูงหยุดการก่อสร้างถึง 389 โครงการในเขตกรุงเทพมหานคร และมีบ้านร้างในเขตกรุงเทพมหานครและปริมณฑลถึง 330,000 หน่วย

ประกอบกับธุรกิจในประเทศไทยมีการบริหารโครงสร้างทางการเงินผิดพลาดโดยนำเงินกู้ระยะสั้นไปลงในโครงการระยะยาว ใช้เงินกู้ผิดวัตถุประสงค์ มีการกู้ยืมเงินตราต่างประเทศมาใช้กับกิจการที่มีรายได้เป็นบาท รวมทั้งมีการลงทุนหรือการปล่อยสินเชื่อในภาคธุรกิจที่มีมากอยู่แล้ว ภายหลังได้รับผลกระทบจากวิกฤติทางเศรษฐกิจ ธุรกิจเกิดการขาดสภาพคล่อง ขณะที่สถาบันการเงินไม่สามารถปล่อยสินเชื่อเพิ่มเติมเนื่องจากเกรงว่าจะมีปัญหาของ NPLs เพิ่มขึ้น ซึ่งทำให้สถาบันการเงินต้องสำรองเพิ่มขึ้น การปล่อยสินเชื่อจึงเข้มงวดส่งผลให้การประกอบการขาดทุนและธุรกิจต้องปิดกิจการ ปัญหา NPLs จึงรุนแรงขึ้นตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**2.3 ผลกระทบ NPLs ต่อเศรษฐกิจ**

1. ธุรกิจขาดเงินทุนหมุนเวียน ธุรกิจที่เป็นลูกหนี้ NPLs แม้ว่าจะแก้ไขหนี้กับเจ้าหนี้สถาบันการเงินไปแล้ว ก็ไม่มีเงินทุนหมุนเวียนมาดำเนินกิจการเพราะการจะขอเพิ่มวงเงินสินเชื่อกับสถาบันการเงินมีความเข้มงวด ทรัพย์สินที่มีอยู่ก็ติดภาระจำนำจำนองจนหมดสิ้น และมูลค่าที่ดินก็ลดลงเนื่องจากการปรับลดราคาประเมินของทางการ ซึ่งก่อให้เกิด NPLs ย้อนกลับ
2. กำลังซื้อในประเทศลดลงเนื่องจากธุรกิจมีปัญหาขาดเงินทุนหมุนเวียน การแก้ไขปัญหาอันดับแรก คือ การลดค่าใช้จ่ายของธุรกิจซึ่งส่วนใหญ่ คือ เงินเดือนและค่าจ้าง โดยการปลดหรือเลิกจ้างพนักงานและลูกจ้าง แม้กระทั่งธนาคารและธุรกิจขนาดใหญ่ก็ยังต้องมีการปลดหรือเลิกจ้างพนักงาน ซึ่งทำให้เกิดการว่างงานในจำนวนที่สูง
3. สภาพคล่องของธนาคารล้นระบบ ธนาคารจะต้องลดต้นทุนของเงินลง โดยการปรับลดอัตราดอกเบี้ยเงินฝาก ซึ่งจะส่งผลกระทบถึงบรรดาข้าราชการ พนักงานองค์การรัฐวิสาหกิจที่เกษียณอายุโดยการรับบำเหน็จ และพนักงานคนงานที่ได้รับเงินชดเชย เงินจากกองทุนสำรองเลี้ยงชีพ หรือเงินที่ได้รับจาก Early Retire ซึ่งบุคคลเหล่านี้หวังว่าเงินที่ได้รับมานั้นสามารถนำไปฝากกับธนาคารเพื่อนำดอกเบี้ยมายังชีพในบั้นปลายชีวิต หรือในขณะว่างงาน เมื่ออัตราดอกเบี้ยเงินฝากของธนาคารอยู่ในอัตราต่ำ รายได้บุคคลเหล่านั้นจะตกต่ำไปด้วย ซึ่งมีผลต่ออำนาจซื้อในระบบเศรษฐกิจเช่นกัน
4. ธุรกิจอสังหาริมทรัพย์ ที่มีหนี้ NPLs กับระบบสถาบันการเงินเป็นจำนวนมาก โครงการที่ดำเนินการแล้วเสร็จและโครงการที่ค้างคาอยู่ก็ไม่สามารถขายออกไปได้ เนื่องจากขาดกำลังซื้อจากประชาชนผู้บริโภค แม้ที่ผ่านมาจะมีการแก้โดยการขายหนี้ไปภายใต้การดำเนินการขององค์การเพื่อการปฏิรูประบบสถาบันการเงิน (ปรส.) แต่ปัญหาของภาคอสังหาริมทรัพย์ยังคงค้างอยู่ปัจจุบันภาครัฐได้พยายามกระตุ้นโดยใช้กลไกผ่านหน่วยงานต่างๆ เช่น การสนับสนุนอัตราดอกเบี้ยต่ำสำหรับผู้ที่ต้องการซื้อบ้านของธนาคารอาคารสงเคราะห์ เพื่อกระตุ้นให้ผู้บริโภคมีการใช้จ่าย

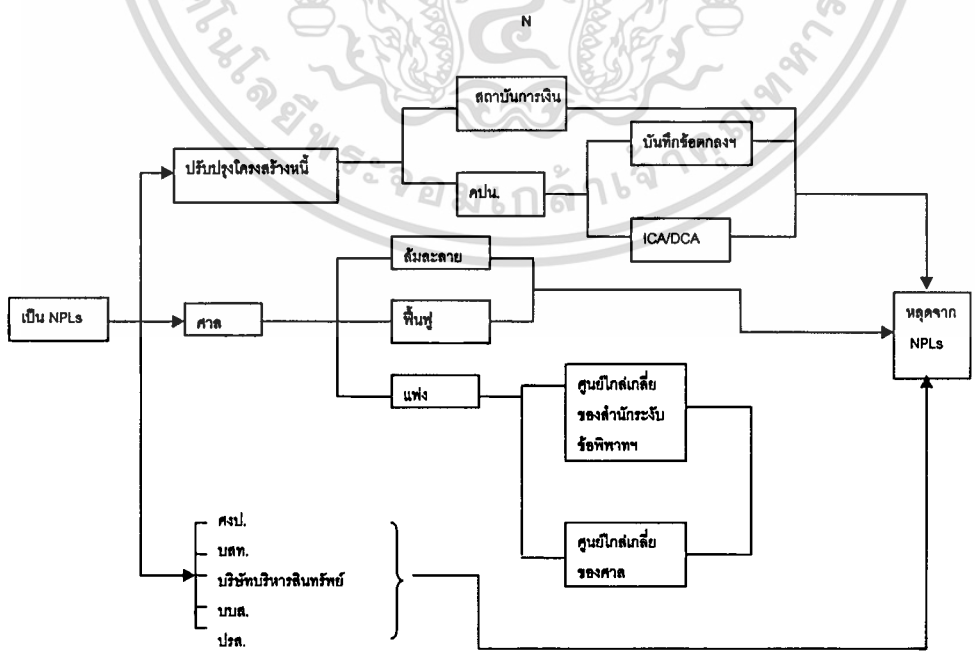
**2.4 การแก้ปัญหา NPLs**

ผลกระทบจากการที่หนี้ที่ไม่ก่อให้เกิดรายได้สูงมากทำให้การฟื้นตัวทางเศรษฐกิจเป็นไปอย่างลำบาก อย่างไรก็ตามภาครัฐและภาคเอกชนได้พยายามแก้ไขปัญหา NPLs โดยการจัดตั้งองค์กรและออกมาตรการต่างๆหลายมาตรการ รวมทั้งปรับปรุงแก้ไขกฎหมายล้มละลาย เพื่อเปิดโอกาสให้ธุรกิจสามารถฟื้นฟูในชั้นศาล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

องค์กรที่ตั้งขึ้นเพื่อปรับปรุงโครงสร้างหนี้และแก้ไขสินทรัพย์ด้อยคุณภาพ ประกอบด้วย ศาลล้มละลายกลาง องค์การเพื่อการปฏิรูประบบสถาบันการเงิน (ปรส.) บรรษัทบริหารสินทรัพย์สถาบันการเงิน (บบส.) คณะกรรมการเพื่อส่งเสริมการปรับปรุงโครงสร้างหนี้ (คปน.) คณะอนุกรรมการปรับปรุงโครงสร้างหนี้จังหวัด บริษัทบริหารสินทรัพย์เอกชน ศูนย์ไกล่เกลี่ยข้อพิพาททางการเงิน สำนักงานศาลยุติกรรม บรรษัทบริหารสินทรัพย์ไทย (บสท.) และศูนย์ให้คำปรึกษาทางการเงินสำหรับวิสาหกิจขนาดกลางขนาดย่อมและประชาชน (ศงป.)

นอกจากนั้น ยังมีข้อตกลงของเจ้าหนี้และลูกหนี้ เพื่อใช้แนวทางในการเจรจาปรับปรุงโครงสร้างหนี้เรียกว่า หลักการปรับปรุงโครงสร้างหนี้ (Framework for Corporate Debt Restructuring in Thailand-Bangkok Framework) สัญญาระหว่างเจ้าหนี้ว่าด้วยการลงมติในการปรับปรุงโครงสร้างหนี้และผู้ชี้ขาด (Inter-Creditor Agreement on Restructure Plan Votes and Executive Decision Panel-ICA) สัญญาระหว่างลูกหนี้และเจ้าหนี้ว่าด้วยกรรมวิธีการปรับโครงสร้างหนี้ (Debtor-Creditor Agreement on Restructure Process -DCA) บันทึกข้อตกลงในการปรับปรุงโครงสร้างหนี้ และธนาคารแห่งประเทศไทยได้ออกหลักเกณฑ์การปรับปรุงโครงสร้างหนี้ พร้อมทั้งมีมาตรการด้านกำกับสถาบันการเงินและสิทธิประโยชน์ต่างๆ เพื่อจูงใจให้มีการเจรจาปรับปรุงโครงสร้างหนี้ หรืออีกนัยหนึ่งทางการได้ส่งเสริมการปรับปรุงโครงสร้างหนี้เพื่อการแก้ปัญหา NPLs โดยอาศัยทั้งกระบวนการศาลและนอกกระบวนการศาลตามแผนภูมิการแก้ไขปัญหา NPLs ดังรูป




**รูปที่ 2.1** แผนภูมิแก้ไข NPLs ของประเทศไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 3
# ดาต้าไมนิ่งและทฤษฎีที่เกี่ยวข้อง

## 3.1 ดาต้าไมนิ่ง

ดาต้าไมนิ่งเป็นกระบวนการและเทคโนโลยีในการค้นหาข้อมูลที่ซ่อนไว้จากฐานข้อมูลที่มีขนาดใหญ่ ซึ่งเป็นเครื่องมือที่มีประสิทธิภาพมากสำหรับองค์กรในการช่วยค้นหาสารสนเทศที่สำคัญเพื่อใช้ในการกำหนดแผนในการดำเนินธุรกิจเพื่อชิงความได้เปรียบจากคู่แข่ง พยากรณ์แนวโน้มหรือพฤติกรรม ซึ่งอาศัยความสัมพันธ์ของข้อมูล โดยดาต้าไมนิ่งจะค้นหาความรู้จากฐานข้อมูลเดิมและนำมาทดสอบสมมติฐานภายหลัง สารสนเทศที่ได้มาจากการดาต้าไมนิ่งต้องมีลักษณะเป็นข้อมูลที่ 1)ไม่รู้มาก่อนล่วงหน้า (Unknown) 2)เป็นข้อมูลที่มีความถูกต้อง (Valid) และ3) สามารถนำไปใช้ประโยชน์ได้จริง (Actionable) โดยมีรายละเอียดดังนี้

1.) ข้อมูลที่ไม่รู้มาก่อนล่วงหน้า (Unknown) เป็นข้อมูลที่ผู้ใช้งานไม่รู้มาก่อนและไม่ชัดเจน ไม่สามารถตั้งสมมติฐานล่วงหน้าได้ว่าเป็นแบบใด เช่น เจ้าของห้างสรรพสินค้าเพิ่งจะค้นพบพฤติกรรมของลูกค้าว่าซึ่งเป็นพ่อบ้านมักจะซื้อผ้าออมและเบียร์พร้อมกันในวันศุกร์ ซึ่งข้อมูลดังกล่าวอาจก่อให้เกิดความได้เปรียบต่อห้างสรรพสินค้าได้โดยการกำหนดแผนการขายหรือจุดวางสินค้าที่เหมาะสมสำหรับผู้ซื้อ ขณะที่คู่แข่งไม่รู้พฤติกรรมดังกล่าวของผู้บริโภค

2.) ข้อมูลที่มีความถูกต้อง (Valid) เป็นข้อมูลที่มีความถูกต้อง เนื่องจากเมื่อมีผู้นำข้อมูลไปใช้ด้วยเทคนิ่งดาต้าไมนิ่งจะค้นพบสิ่งที่น่าสนใจตลอดเวลา แต่ต้องพิจารณาด้วยว่าสิ่งนั้นถูกต้องหรือไม่

3.) ข้อมูลที่สามารถนำไปใช้ประโยชน์ได้จริง (Actionable) เป็นข้อมูลที่ได้จากการนำดาต้าไมนิ่งมาใช้ซึ่งอาจนำมาช่วยในการตัดสินใจเพื่อชิงความได้เปรียบของธุรกิจ แต่ในบางครั้งข้อมูลที่ค้นพบอาจเป็นข้อมูลที่ผิดกฎหมายหรือคู่แข่งได้ค้นพบไปแล้ว ซึ่งหากนำมาใช้อาจไม่ก่อให้เกิดประโยชน์หรือไม่อาจกระทำผิดกฎหมาย ซึ่งการนำข้อมูลมาใช้ต้องคำนึงถึงความเหมาะสมด้วยการนำดาต้าไมนิ่งมาใช้ในในการทำงานหลายด้านเช่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ด้านการแพทย์ เช่น ทราบผลข้างเคียงจากการใช้ยา การวิเคราะห์เกี่ยวกับยีน หรือใช้ในการพยากรณ์แนวโน้มของผู้ป่วยจากอาการของโรคที่แสดงออก เพื่อลดภาระของแพทย์ในขั้นต้น
- ด้านการเงิน เช่น การพยากรณ์เกี่ยวกับข้อมูลการค้าหุ้น หรือการประเมินความเสี่ยงทั้งในการปล่อยสินเชื่อ
- ด้านวิทยาศาสตร์ เช่น การค้นหาทฤษฎีใหม่ๆ
- ด้านวิศวกรรม เช่น การสืบค้นความผิดปกติในการทำงาน (fault detection)

**3.2 ขั้นตอนในการทำดาต้าไมนิ่ง**

สำหรับขั้นตอนในการทำดาต้าไมนิ่งประกอบด้วย 5 ขั้นตอน ดังนี้

1. การกำหนดวัตถุประสงค์ของธุรกิจ (Business Objectives Determination)
2. การเตรียมข้อมูล (Data Prepareation)
3. การทำดาต้าไมนิ่ง (Data Mining)
4. การวิเคราะห์ผลลัพธ์ที่ได้(Analysis of Result)
5. การนำมาประยุกต์ใช้กับธุรกิจ (Assimilation of knowledge)

**3.2.1 การกำหนดวัตถุประสงค์ของธุรกิจ (Business Objectives Determination)**

ทำความเข้าใจต่อปัญหาที่แท้จริงของธุรกิจหรือการกำหนดเป้าหมายที่ท้าท้ายเพื่อที่จะไปให้ถึงจุดมุ่งหมายนั้น ซึ่งมีความจำเป็นอย่างมากในการนำดาต้าไมนิ่งมาใช้เพราะต้องมีการกำหนดเป้าหมายที่ชัดเจน จึงจะสามารถนำดาต้าไมนิ่งจะช่วยในการแก้ปัญหาหรือบรรลุเป้าหมายได้ โดยในขั้นตอนนี้จะประกอบด้วยการวิเคราะห์ทางธุรกิจ และการวิเคราะห์ข้อมูลเบื้องต้นว่าข้อมูลที่มีอยู่ประกอบด้วยอะไรบ้างและต้องการอะไรจากข้อมูล การกำหนดแบบจำลองที่เหมาะสม ซึ่งในการสร้างแบบจำลองจะแตกต่างตามเป้าหมายของธุรกิจนั้นๆ

**3.2.2 การเตรียมข้อมูล (Data Prepareation)**

เป็นขั้นตอนที่ใช้เวลาส่วนใหญ่ของในการทำดาต้าไมนิ่ง โดยแบ่งเป็น 3 ขั้นตอน ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) การเลือกข้อมูล (Data Selection)

2) การสุ่มเลือกข้อมูลและการทดสอบข้อมูล (Data Preprocessing)

3) การแปลงข้อมูลไปสู่ Model สำหรับการวิเคราะห์ (Data Transformation)

**1) การเลือกข้อมูล (Data Selection)** จุดประสงค์ คือ การจำแนกแหล่งที่มาของข้อมูลและการดึงข้อมูลที่ต้องการใช้เบื้องต้นสำหรับการวิเคราะห์ก่อนที่จะทำการไมนิ่ง ซึ่งในการเลือกข้อมูลที่จะนำมาใช้นั้นขึ้นอยู่กับวัตถุประสงค์ของธุรกิจซึ่งกำหนดขึ้นมานั้นว่าต้องการใช้ข้อมูลลักษณะใด ซึ่งข้อมูลนั้นต้องทำความเข้าใจกับชนิดของข้อมูล ค่าที่เป็นไปได้ แหล่งกำเนิดข้อมูล รูปแบบของข้อมูล และลักษณะอื่นๆ ตัวแปรข้อมูลมี 2 ลักษณะ

a) Categorical แบ่งเป็น

- Nominal เป็นตัวแปรที่ลำดับข้อมูลไม่มีผลกับค่า เช่น เพศ ชาย หญิง เป็นต้น
- Ordinal เป็นตัวแปรที่ลำดับข้อมูลมีผลกับค่า เช่น เกรด 4,3,2,1 เป็นต้น

b) Quantitative ข้อมูลเชิงคุณภาพแบ่งเป็น

- Continuous ค่าที่เก็บเป็นเลขจำนวนจริง ( Real Number) หรือค่าที่ต่อเนื่อง เช่น ค่าเฉลี่ย เป็นต้น
- Discrete ค่าที่เก็บเป็นเลขจำนวนเต็ม (Interger) เช่น ข้อมูลอายุ 15, 27 เป็นต้น

**2) การสุ่มเลือกข้อมูลและการทดสอบข้อมูล ( Data Preprocessing )** จุดประสงค์เพื่อการเลือกข้อมูลนั้นมีความเชื่อมั่นว่าเลือกข้อมูลที่ถูกต้องที่จะนำไปใช้ทำดาต้าไมนิ่งต่อไป โดยใช้หลักการทางสถิติและเทคนิคการนำเสนอข้อมูลที่น่าสนใจ ( Data Visualization Techniques) นอกจากนี้สิ่งสำคัญที่จะต้องพิจารณาในขั้นตอนนี้ คือ การพิจารณาข้อมูลที่เป็น 1.) Missing Data และ 2.) Noisy Data

1.) Missing Data คือ ค่าข้อมูลที่ขาดหายไปไม่ครบถ้วนในชุดข้อมูลที่จะมาวิเคราะห์ สาเหตุอาจเกิดจากความผิดพลาดของมนุษย์ เช่น ในกรณีที่เป็นการกรอกข้อมูลของแบบสอบถาม ในบางหัวข้อที่อาจไม่มีการกรอกข้อมูลหรือไม่อาจกรอกข้อมูลที่ผิดพลาด ทำให้ไม่สามารถนำข้อมูลมาใช้ได้ การแก้ไขอาจตัดข้อมูลดังกล่าวทิ้งไป หรือกำหนดค่าให้กับข้อมูลที่เป็น Missing โดยใช้ค่าเฉลี่ย หรือ กำหนดค่าใหม่ซึ่งไม่มีผลต่อการนำข้อมูลมาวิเคราะห์

2.) Noisy Data คือ ค่าข้อมูลที่ผิดปกติ หรือ มีความแตกต่างจากกลุ่มของข้อมูลที่นำมาใช้ ซึ่งในบางครั้งการแก้ไขปัญหา คือ การนำข้อมูลมา plot graph เพื่อตรวจสอบว่ามีข้อมูลมากน้อยแค่ไหนที่เป็น Noisy และตัดรายการดังกล่าวทิ้งไป เพื่อไม่ให้ข้อมูลดังกล่าวสร้างความบิดเบือนในการทำดาต้ามายนิ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**3) การแปลงข้อมูลไปสู่ Model สำหรับการวิเคราะห์ (Data Transformation)** เป็นการแปลงข้อมูลให้อยู่ในรูปแบบที่จะนำไปวิเคราะห์ตามอัลกอริทึมของดาต้าไมนิ่งที่ใช้ เช่น การแปลงตัวแปรแบบ Categorical ให้เป็น Numeric เช่น สีแดง สีน้ำเงิน และ สีเขียว อาจแทนค่าเป็นตัวเลขเท่ากับ 1 2 และ3

**3.2.3.การทำดาต้าไมนิ่ง (Data Mining)**

เป็นการนำข้อมูลที่จัดเตรียมไว้มาทำดาต้าไมนิ่ง เพื่อแปลงสภาพข้อมูลดิบให้เป็นความรู้ในลักษณะของรูปแบบและกฎเกณฑ์ (Pattern And Rule Finder) ซึ่งขั้นตอนนี้จะมีความสัมพันธ์กับขั้นตอนที่ผ่านมา เนื่องจากวิธีการทำดาต้าไมนิ่งใช้วิธีการทำซ้ำจนกว่าจะสามารถหาความสัมพันธ์หรือรูปแบบที่ต้องการซึ่งต้องอาศัยการใช้อัลกอริทึมเพื่อหาความสัมพันธ์ของข้อมูลที่ต้องการศึกษา

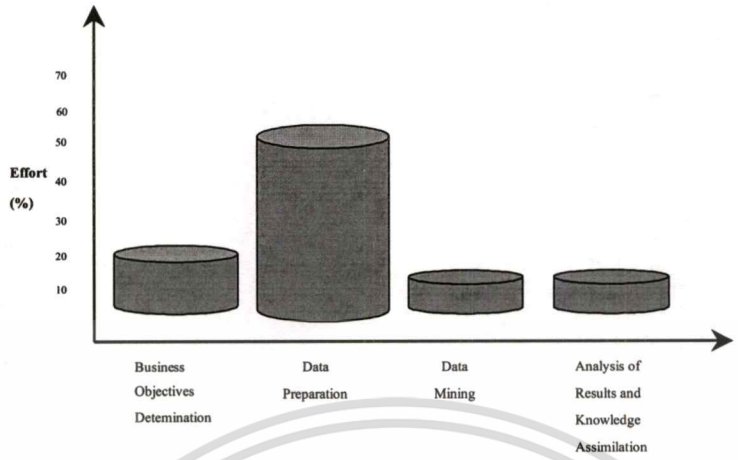
**3.2.4 การวิเคราะห์ผลลัพธ์ที่ได้ (Analysis of Result)**

เป็นการวิเคราะห์และตีความผลลัพธ์ที่ได้จากการทำดาต้าไมนิ่ง การทำงานในขั้นตอนนี้ต้องใช้ทักษะในการวิเคราะห์ข้อมูล และการวิเคราะห์ทางธุรกิจ รวมทั้งพื้นฐานความเข้าใจของข้อมูลเนื่องจากการทำดาต้ามายนิ่งได้นั้นจะได้ค่าเป็นตัวเลขหรือช่วงเพราะเป็นไปตามลักษณะข้อมูลที่ใช้ในการวิเคราะห์ แต่การนำผลจากการทำดาต้ามายนิ่งมาใช้ประโยชน์หรือแปลความหมายนั้นต้องอาศัยความรู้และความเข้าใจในข้อมูลหรือธุรกิจที่จะนำไปใช้เป็นหลัก นอกจากนี้การนำเอาแบบจำลองที่ได้ไปทดสอบกับข้อมูลชุดอื่น ที่ไม่ใช่ข้อมูลที่ใช้ในการสร้างแบบจำลอง เพื่อนำเอาผลลัพธ์ที่ได้มาเปรียบเทียบกับผลตามแบบจำลอง ว่ามีความแม่นยำและยอมรับได้หรือไม่ ซึ่งถ้าผลการทดสอบไม่สามารถยอมรับได้ หากต้องเพิ่มข้อมูลให้มากขึ้นเพื่อให้มีความแม่นยำในการหาความสัมพันธ์ได้หรือในบางครั้งอาจต้องเปลี่ยนอัลกอริทึมให้เหมาะสม

**3.2.5 การนำมาประยุกต์ใช้กับธุรกิจ (Assimilation of knowledge)**

สำหรับในขั้นตอนนี้เป็นการนำผลที่ได้จากขั้นตอนที่ 3.2.4 มารวมกับส่วนความรู้เพื่อนำไปใช้ต่อไป เนื่องจากในการทำ Data Mining รูปแบบที่ได้หรือ Pattern ที่เรียนรู้ต้องนำมาปรับกับความรู้ที่อยู่เพื่อก่อให้เกิดประโยชน์ คือทำแล้วต้องได้อะไร ในขั้นตอนนี้มีหลักอยู่ 2 ประการ คือ การนำเสนอแนวคิดทางธุรกิจที่ค้นพบใหม่ และหาแนวทางที่จะใช้กฎเกณฑ์ใหม่ที่ค้นพบเพื่อให้เกิดประโยชน์สูงสุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




**รูปที่ 3.1** เปอร์เซ็นต์การทำงานในแต่ละขั้นตอนในการทำ Data Mining

**3.3 ลักษณะของดาต้าไมนิ่งที่นิยมใช้มี 4 รูปแบบ**

- Database Segmentation หรือ Database Clustering การแยกข้อมูลในฐานข้อมูลออกเป็นกลุ่มตามคุณสมบัติที่เหมือนกัน โดยไม่รู้ล่วงหน้าได้ว่าจะมีข้อมูลเป็นจำนวนมากี่กลุ่ม กลุ่มใดบ้าง
- Link Analysis การตรวจหาความสัมพันธ์ (Associations) ระหว่าง Record หรือ กลุ่ม Record ใน Database
- Deviation Detection การตรวจหาค่าที่เบี่ยงเบนไปจากค่ามาตรฐาน
- Predictive Modeling กระบวนการสร้างการทำนาย มี 2 แบบ คือ
  1) Value Predictive การพยากรณ์หรือประมาณค่าข้อมูลที่มีความสัมพันธ์กันใน Database
  2) Classification การจัดหมวดหมู่ข้อมูล โดยพิจารณาข้อมูลที่กำหนดไว้ล่วงหน้า

ซึ่งจากข้อมูลข้างต้นสามารถสรุปเป็นตาราง 3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ 3.1** โมเดลของ Data Mining กับการประยุกต์ใช้งาน

| | Market Management | Risk Management | Prauad Management |
|---|---|---|---|
| Application | Target Marketing<br>Customer Relationship Management<br>Marketing Basket Analysis<br>Cross Selling<br>Market Segmentation | Forecasting<br>Customer Relation<br>Improved Underwriting<br>Quality Control<br>Competitive Analysis | Fraud Datаction |

| | | | | |
|---|---|---|---|---|
| Operation | Predictive Modeling | Database Segmentation | Link Analysis | Daviation Detection |
| Techniques | Classification | Demographic Clustering | Associations Discovery | Visualization |
| | Value Prediction | Neural Clustering | SeQuential Pattern Discovery<br>Similar Time Sequence Discovery | Statistic |

สำหรับในโครงการนี้จะนำเสนอกระบวนการของการพยากรณ์ (Predictive Model) ในการหาความสัมพันธ์จากข้อมูลว่าแนวโน้มของบริษัทจะเป็นหนี้ที่ไม่ก่อให้เกิดรายได้หรือ NPLs โดยใช้ Decision Trees เนื่องจาก Decision Trees จะสามารถศึกษาหรือเรียนรู้ถึงรูปแบบหรือปัจจัยจากข้อมูล โดยสามารถแสดงภาพที่ชัดเจนออกมาแตกเป็นต้นไม้ กิ่งก้าน ทำให้ในการนำไปใช้หรือการแปรความหมายกระทำได้ค่อนข้างเข้าใจได้ง่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

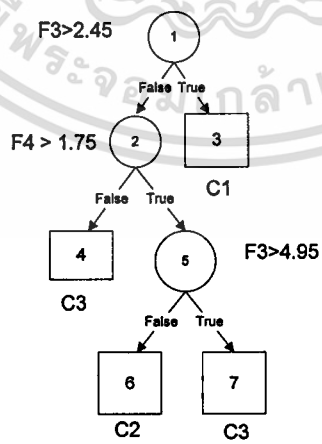
# บทที่ 4

## การพยากรณ์(Predictive)

### 4.1 การพยากรณ์ (Predictive)

Predictive Data Mining คือการค้นหารูปแบบความสัมพันธ์จากข้อมูลที่มีอยู่เป็นจำนวนมากเพื่อนำมาใช้ในการพยากรณ์หรือคำนวณการตัดสินใจในอนาคตได้อย่างถูกต้อง โดยเทคนิคในการทำ Predictive คือการศึกษา decision criteria จากข้อมูลที่มีในอดีตซึ่งเหมือนกับการเรียนรู้จากประสบการณ์ของคนเราเพื่อใช้ในการสังเกตรูปแบบของ Model ที่มีลักษณะที่สำคัญของปรากฏการณ์บางอย่าง (phenomenon) ซึ่งมีหลาย Solution ในการทำ Predictive โดย Decision Trees เป็น Solution หนึ่งซึ่งนำมาใช้ในการจำแนก Case ต่างๆ

Decision Trees คือการกำหนด Rules เพื่อใช้ในการแบ่งกลุ่มของข้อมูลโดย Rules แรก จะแบ่งข้อมูลส่วนใหญ่ออกมา และจะใช้ Rules ต่อมาในการแบ่งข้อมูลออกเป็นลำดับชั้นคล้ายการแตกรากของต้นไม้จากบนลงล่างหรือ Organization chart ขององค์กร Trees ประกอบด้วย Node เพื่อทดสอบค่า ว่า ผิด หรือ ถูก สำหรับในจุดที่ต้องตัดสินใจ โดย Branch เป็นฐานของผลที่ผ่านการทดสอบมาแล้ว กระบวนการนี้จะดำเนินไปเรื่อยๆจนกระทั่งจุดสุดท้ายหรือเรียกว่า Terminal Node ซึ่งจะเป็นที่เก็บค่าของ Node ที่เป็นคำตอบ โดยคำตอบดังกล่าวจะต้องใกล้เคียงกับข้อมูลที่คาดหวังไว้ ดังนั้น trees คือการนำเสนอการจำแนกข้อมูลออกเป็นกลุ่มข้อมูล ตัวอย่างของ Decission Trees ดังรูปที่ 4.1




**รูปที่ 4.1** แสดงภาพของ Decision Tree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Trees สามารถรองรับลักษณะของข้อมูลได้หลายลักษณะ ดังนี้

- Nominal ข้อมูลตัวเลข เช่น 1,2 ,3 เป็นต้น
- Ordinal ข้อมูลแยกเป็นประเภท เช่น อากาศ เย็น หนาว ร้อน และอบอุ่น .
- Interval ข้อมูลที่เป็นค่าต่อเนื่องหรือค่าเฉลี่ย เช่น อุณหภูมิ เป็นต้น

ดังนั้นในการใช้ Trees จึงขึ้นอยู่กับลักษณะข้อมูลที่นำมาใช้และ Trees สามารถใช้งานได้กับ Spreadsheet ได้ โดยไม่ต้องมีการ normalized ตารางข้อมูล นอกจากนี้การสร้าง Trees สามารถใช้ Algorithm ได้หลาย Algorithm ได้แก่ CHAID (Chisquare-Automatic-Interaction-Detection) , Exhaustive CHAID, CART (Classification and Regression Tree), Quest (quick Unbiased,Statistical,Efficient,Statiscal Tree) เป็นต้น สำหรับการพยากรณ์ในครั้งนี้ใช้ Algorithm คือ Exhaustive CHAID ซึ่งเป็นการใช้หลักการของ CHAID Algorithm คือการคำนวณค่า Chisquare ในการหาความสัมพันธ์ในการแตกหรือยุบ Tree เพียงแต่ Exhaustive CHAID ได้ถูกปรับปรุงส่วนสำคัญในการแตกกิ่ง ซึ่ง Exhaustive CHAID จะแตกกิ่งได้เยอะมากกว่า CHAID

**4.2 CHAID Algorithm**

สำหรับ Model ที่นำมาใช้ในการพยากรณ์ คือ CHAID ซึ่งเป็น Model ที่มีความนิยมในการนำมาใช้งาน Data Mining ประกอบกับข้อมูลที่นำมาใช้พยากรณ์ส่วนใหญ่เป็นตัวเลขดังนั้นในการหาความสัมพันธ์ CHAID (Chisquare-Automatic-Interaction-Detection) เป็นโมเดลที่ออกแบบมาเพื่อจัดการแยกแยะค่าตัวแปร variable โดยเฉพาะ ซึ่งสามารถใช้ได้ทั้งข้อมูลที่เป็นจำนวนตัวเลขปกติ ข้อมูลตัวเลขต่อเนื่องโดย CHAID จะทำการแยกชุดข้อมูลโดยพิจารณาจากความสัมพันธ์ที่ได้จากการคำนวณ โดยจะแบ่งจากชุดข้อมูลรวมเป็นซับเซ็ตย่อย ระดับ Child Node ตามจำนวนที่กำหนด โดยการทำจะทำจนกว่าตัวแปรที่มีไม่มีความสัมพันธ์กับชุดข้อมูลที่เป็น Target แล้วก็จะยุติระบวนการหาความสัมพันธ์ ซึ่งผลที่ได้เป็นทางเลือกที่จะใช้ในการพยากรณ์

องค์ประกอบของการใช้ CHAID ในการวิเคราะห์

- ตัวแปรที่ใช้ในการพยากรณ์
- ตัวแปรที่เป็น Target
- การกำหนดค่า CHAID parameters ที่จะใช้ในการวิเคราะห์ชุดข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการทำงานของ CHAID ก็คือ

1. คำนวณการกระจายของตัวแปรY ใน ROOT NODE

**ตารางที่ 4.2** 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 |

2. ทุกค่าของตัวแปร X จะหาคู่ของกลุ่มของข้อมูลของ X โดยการหาค่าเปรียบเทียบจากความแตกต่างน้อยที่สุดทีละคู่โดยดูจากคู่ที่มีค่า p-value มากที่สุด ที่เกี่ยวข้องกับการกระจายของ Y ในแต่ละ Node การคำนวณ p-value การวัดระดับของ Y ซึ่งในตัวอย่าง

ความสัมพันธ์ระหว่างค่า X1 และ ค่า Y ภายใน Node แสดงรายละเอียดตามตารางด้านล่าง

**ตารางที่ 4.3** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1

| X1/Y | 1 | 2 | 3 | 4 | RowToll |
|---|---|---|---|---|---|
| 1 | 23 | 5 | 19 | 4 | 51 |
| 2 | 12 | 8 | 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 d.f.=9(p=0.00234955)

จากตารางที่ 4.3 แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 สามารถคำนวณหาหาค่า$(\chi^2)$ d.f. และ p-value ตามลำดับ ได้ผลลัพธ์ตามตารางที่ 4.4 – 4.9.

**ตารางที่ 4.4** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 1 และ2

| X1/Y | 1 | 2 | 3 | 4 | RowToll |
|---|---|---|---|---|---|
| 1 | 23 | 5 | 19 | 4 | 51 |
| 2 | 12 | 8 | 15 | 13 | 42 |
| ColToll | 35 | 7 | 34 | 17 | 93 |

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

**ตารางที่ 4.5** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 1 และ3

| X1/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 d.f.=3(p=0.0456149)

**ตารางที่ 4.6** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 1 และ4

| X1/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 d.f.=3(p=0.0001939266)

**ตารางที่ 4.7** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 2 และ3

| X1/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 d.f.=3(p=0.0648145)

**ตารางที่ 4.8** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 2 และ4

| X1/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 d.f.=3(p=0.1745651)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ 4.9** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 3 และ4

| X1/Y | 2 | 3 | 4 | RowToll |
|---|---|---|---|---|
| 3 | 1 | 0 | 1 | 2 |
| 4 | 0 | 1 | 4 | 5 |
| ColToll | 1 | 1 | 5 | 7 |

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

2.2 algorithm จะคำนวณหาค่าของ p-value ของความสัมพันธ์ในกลุ่มของตัวแปรเพื่อนำมาเปรียบเทียบกับค่า alpha level $\alpha$merge (0.05 คือค่าที่กำหนดไว้) ซึ่งจากการพิจารณาค่าของ p-value ในคู่ตัวแปรของ 3 และ 4 ของ X1 เท่ากับ 0.2143 ซึ่งมีค่ามากกว่า $\alpha$merge คือ 0.05 ทำให้สามารถยุบรวมกลุ่มของ 3 และ 4 เป็นกลุ่มเดียวกันได้ ซึ่งผลที่ได้จะสามารถจัดกลุ่มข้อมูลของ X1 ใหม่

**ตารางที่ 4.10** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 1 และ3,4

| X1/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 d.f.=3(p=0.0001502343)

**ตารางที่ 4.11** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 2 และ3,4

| X1/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 d.f.=3(p=0.09333908)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ 4.12** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 1 และ2

| X1/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 d.f.=3(p=0.02682849)

2.3. กระบวนการจะทำซ้ำต่อไป โดยจะหาค่า p-value ของความสัมพันธ์ของข้อมูลในแต่ละคู่ว่าคู่ใดที่มีค่า p-valueสูงกว่า alpha level $\alpha$merge (0.05 คือค่าที่กำหนดไว้) ซึ่งจะพบว่าของ p-value ในคู่ของชุดข้อมูล 2 3 และ 4 ของ X1 เท่ากับ 0.09333908 ซึ่งมีค่ามากกว่า $\alpha$merge คือ 0.05 ทำให้สามารถยุบรวมกลุ่มของ 2 3 และ 4 เป็นกลุ่มเดียวกันได้ ทำให้การยุบรวมของชุดข้อมูลยุติลงเนื่องจากสามารถแบ่งกลุ่มข้อมูลของ X1 ได้แล้ว

2.4 algorithm จะคำนวณเพื่อปรับค่าของ p-value สำหรับเซ็ตที่ได้ยุบรวมกัน โดยใช้ Bonferroni multiplier

**ตารางที่ 4.13** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 1 และ2,3,4

| X1/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 d.f.=3(p=0.004448843)

ค่าของ $(\chi^2)$ p-value ได้จากคำนวณข้างต้นจะถูกปรับโดยการใช้ Bonferroni multiplier โดย X1 เป็นข้อมูลลักษณะ Nominal โดยสูตรคำนวณ Bonferroni multiplier เป็นดังนี้

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

ซึ่งค่า c = จำนวนกลุ่มข้อมูลของ X1 และ r= จำนวนของกลุ่มที่ยุบรวมกัน ซึ่งจากสูตรคำนวณจะปรับค่าของ p-value เท่ากับ 0.0311 (=0.00448843x7)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ขั้นตอนตั้งแต่ 2.1-2.4 จะถูกทำซ้ำโดยการเปลี่ยนค่าจาก X1 เป็น X2
ตารางแสดงความสัมพันธ์ระหว่าง X2 และ Y ปรากฏตามด้านล่าง

**ตารางที่ 4.14** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X2

| X2/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 d.f.=6(p=0.8372577)

3.2 X2 ประกอบด้วยข้อมูล 3 กลุ่ม ปรากฏตามตารางด้านล่าง

**ตารางที่ 4.15** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X2 กรณี 0 และ1

| X2/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 d.f.=3(p=0.8282962)

**ตารางที่ 4.16** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X2 กรณี 0 และ2

| X2/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 d.f.=3(p=0.5930452)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ตารางที่ 4.17** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X2 กรณี 1 และ2

| X2/Y | 1 | 3 | RowToll |
|---|---|---|---|
| 1 | 1 | 1 | 2 |
| 2 | 1 | 0 | 1 |
| ColToll | 2 | 1 | 3 |

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

3.3 จากการพิจารณาค่าของ p-value ที่มากที่สุดก็คือ 0.8283 ทำให้สามารถยุบรวมกลุ่มข้อมูลของ 0 และ 1 เป็นกลุ่มเดียวกัน และยุติการยุบรวมเนื่องจากสามารถแยกกลุ่มข้อมูลได้อย่างชัดเจนแล้ว

3.4 การคำนวณเพื่อปรับค่าของ p-value ของ X2 เช่นเดียวกับ X1 (0.5985587 x 3 )

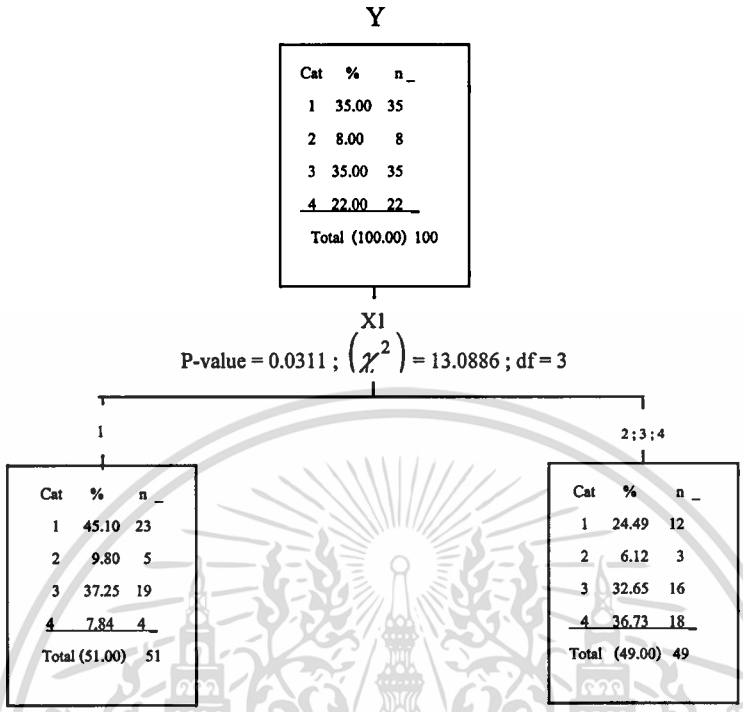
**ตารางที่ 4.18** แสดงค่าความสัมพันธ์ระหว่างค่า Y และ X1 กรณี 0,1 และ2

| X1/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 d.f.=3(p=0.5985587)

4. ในขั้นตอนสุดท้ายที่จะแตก node โดยใช้พื้นฐานของการยุบรวมโดยพิจารณาค่าของ p-value ที่เล็กที่สุดที่ปรับค่าแล้วซึ่งมีค่าน้อยกว่า 0.05 ทำให้สามารถสรุปได้ว่าจะแตก Root Node ได้ 2 Subnode โดยกลุ่มข้อมูลของ X1 นั้นคือ หนึ่ง subnode กรณี X1=1 จะมีกรณีได้ 35 กรณี โดยส่วนที่เหลืออีก 65 กรณี เป็น X1 = 2 3 หรือ 4
5. การสร้าง Trees จะดำเนินต่อไปจนกระทั่งหมด Criteria ที่กำหนดหรือหยุดเมื่อได้ Trees ที่พอใจ โดยการสร้าง Trees โดยใช้ CHAID ตามขั้นตอนที่กล่าวมาข้างต้นแสดงโดยรูปที่ 4.2 ด้านล่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




**รูปที่ 4.2** Tree ของข้อมูลจากการใช้ CHAID Algorithm

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 5
# วิธีการดำเนินการศึกษา

## 5.1 ข้อมูลที่ศึกษา

### 5.1.1 การคัดเลือกข้อมูล

การคัดเลือกข้อมูลใช้ข้อมูลงบการเงินของบริษัทที่อยู่ในตลาดหลักทรัพย์แห่งประเทศไทยของปี 2543 เนื่องจากข้อมูลงบการเงินของกิจการที่ดำเนินงานในประเทศไทยส่วนใหญ่หากไม่ได้อยู่ในตลาดหลักทรัพย์ฯแม้จะได้รับการรับรองจากผู้สอบบัญชีแล้วก็ตามแต่จะจัดทำงบการเงินไม่ถูกต้องกับความเป็นจริงหรือเปิดเผยฐานะการเงินของกิจการที่แท้จริงไม่ครบถ้วน ซึ่งหากนำข้อมูลมาใช้อาจก่อให้เกิดความผิดพลาดได้ ดังนั้นเพื่อให้ข้อมูลที่นำมาใช้ในการศึกษามีมาตรฐานเดียวกันประกอบกับข้อมูลงบการเงินของตลาดหลักทรัพย์ฯจะต้องมีผู้สอบบัญชีซึ่งได้รับอนุญาตและอยู่ในรายชื่อที่ตลาดหลักทรัพย์ยอมรับด้วย โดยงบการเงินที่นำมาใช้นั้นแบ่งข้อมูลเป็น 2 กลุ่ม คือ 1.งบการเงินของบริษัทในตลาดหลักทรัพย์ที่ไม่อยู่ระหว่างการฟื้นฟูกิจการ 2..งบการเงินของบริษัทในตลาดหลักทรัพย์ที่อยู่ระหว่างการฟื้นฟูกิจการ ซึ่งมีสถานะเป็นหนี้ไม่ก่อให้เกิดรายได้ที่ชัดเจนเนื่องข้อมูลในของบริษัทที่เป็น NPLs ในปี 2543 ไม่สามารถหาข้อมูลที่ชัดเจนเนื่องจากแหล่งข้อมูลที่น่าเชื่อถือที่สุดคือธนาคารแห่งประเทศไทยนั้น ไม่พึงเปิดเผยข้อมูลดังกล่าว แต่เพื่อต้องการศึกษาถึงรูปแบบของลูกหนี้ที่เป็น NPLs ดังนั้นจึงจำเป็นต้องใช้ข้อมูลที่สามารถหาได้ คือ จากตลาดหลักทรัพย์โดยรูปแบบที่ทำการศึกษาเป็นบริษัทที่อยู่ระหว่างฟื้นฟูกิจการซึ่งถือเป็นตัวแทนของกิจการที่เป็น NPLs ได้เนื่องจากกิจการที่อยู่ระหว่างฟื้นฟูกิจการ หมายถึง ฐานะการเงินของกิจการในปัจจุบันไม่สามารถรับภาระหนี้ที่มีอยู่ได้และไม่สามารถประกอบธุรกิจได้ตามปกติ

งบการเงิน (Financial Statement or Financial Report) คือรายงานข้อมูลทางการเงินที่ได้จัดทำขึ้นจากข้อมูลทางการเงินที่จดบันทึกไว้ในรอบระยะเวลาหนึ่ง ประกอบด้วย รายละเอียด 5 ส่วน คือ งบดุล งบกำไรขาดทุน งบกระแสเงินสด งบกำไรสะสม และรายละเอียดและหมายเหตุประกอบงบการเงิน

1.) งบดุล (Balance Sheet) เป็นรายงานที่แสดงถึงฐานะการเงินของกิจการ ณ วันใดวันหนึ่ง เพื่อทราบว่ากิจการมีฐานะการเงินเป็นอย่างไร ณ วันนั้น ดังนั้น ในรายงานนี้จึงประกอบด้วยข้อมูลทางการเงินที่แสดงถึงฐานะของกิจการซึ่งได้แก่ข้อมูลสินทรัพย์ หนี้สิน และส่วนของผู้ถือหุ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.) งบกำไรขาดทุน(Income Statement) เป็นรายงานที่แสดงถึงผลการดำเนินงาน (กำไรหรือขาดทุน) ของกิจการในรอบระยะเวลาใดเวลาหนึ่ง ซึ่งจะประกอบด้วยข้อมูลทางการเงินที่เป็นรายได้และค่าใช้จ่าย ถ้าในรอบระยะเวลาใดเวลาหนึ่งรายได้สูงกว่าค่าใช้จ่าย แสดงว่ากิจการนั้นมีกำไร หมายความว่ามีผลสำเร็จในการดำเนินงานเนื่องจากได้รับกำไร แต่ถ้าค่าใช้จ่ายสูงกว่ารายได้ แสดงว่ากิจการนั้นมีผลขาดทุน หมายความว่า ไม่มีผลสำเร็จในการดำเนินงาน เนื่องจากได้รับผลขาดทุน การจัดทำงบกำไรขาดทุนจะเสนอข้อมูลรายได้หักด้วยข้อมูลค่าใช้จ่าย ผลต่างจะเป็นกำไรสุทธิหรือขาดทุนสุทธิแล้วแต่กรณี

3.) งบกระแสเงินสด (Statement of Cash Flow) เป็นรายงานที่แสดงถึงการเปลี่ยนแปลงเงินสดของกิจการในรอบระยะเวลาใดเวลาหนึ่ง การจัดทำงบกระแสเงินสดจะรายงานให้ทราบว่าเงินสดในปีปัจจุบันเพิ่มขึ้นหรือลดลงจากปีก่อนเท่าใด เพราะเหตุใด เนื่องจากข้อมูลจากงบกระแสเงินสดเป็นข้อมูลในเชิงของการเปรียบเทียบของงบการเงิน 2 ปี คือ ปีปัจจุบันกับปีที่ผ่านมา ดังนั้นในรายงานฉบับนี้จึงไม่ได้มีการนำข้อมูลส่วนนี้มาทำการศึกษาเนื่องจากต้องการจะใช้ข้อมูลในงบดุลและงบกำไรขาดทุนเฉพาะปีเท่านั้น

4.) งบกำไรสะสม (Retained Earning Statement) เป็นรายงานที่แสดงถึงการเปลี่ยนแปลงของกำไรสะสมในระยะเวลาใดเวลาหนึ่งซึ่งยอดของกำไร(ขาดทุน)สะสมจะแสดงอยู่ในส่วนของผู้ถือหุ้น ดังนั้นจึงไม่ได้มีการนำข้อมูลส่วนนี้มาทำการศึกษาโดยตรง

5.) รายละเอียดและหมายเหตุประกอบงบการเงิน (Note to Financial Statement) เป็นรายละเอียดหมายเหตุ ซึ่งเป็นคำอธิบายหรือรายละเอียดเพิ่มเติมเกี่ยวกับรายงานทั้ง 4 ข้างต้น

เนื่องจากข้อมูลที่นำมาจากงบการเงินมีจำนวนรายละเอียดค่อนข้างมาก รวมทั้งธุรกิจในการดำเนินงานของแต่ละบริษัทไม่เหมือนกัน การแสดงข้อมูลในงบการเงินรายละเอียดอาจมีความแตกต่างกัน ดังนั้นในการคัดเลือกข้อมูลที่นำมาใช้มาใช้ในการศึกษาซึ่งมีสาระสำคัญ ดังนี้

**5.1.2 ข้อมูลที่เลือกมาใช้งาน**

**1. ข้อมูลที่จากงบดุลที่นำมาใช้ ได้แก่**

1) ด้านสินทรัพย์ (Assets) หมายถึง สิทธิและทรัพยากรที่กิจการมีอยู่ซึ่งเกิดจากการประกอบการและสามารถแสดงค่าเป็นตัวเงิน และจะให้ประโยชน์ในอนาคต

   a) เงินสดและเงินฝากธนาคาร (Cash In Hand and at bank) หมายถึงเงินสดที่กิจการดำรงอยู่ไว้ ณ ที่กิจการเองหรือฝากไว้ธนาคารพาณิชย์

   b) ลูกหนี้การค้า (Account Receivable) หมายถึง เงินที่ลูกค้าค้างชำระค่าบริการหรือค่าสินค้าที่เกิดจากการดำเนินการค้าปกติของธุรกิจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

c) สินค้าคงเหลือ (Inventory) หมายถึง สินค้าสำเร็จรูป งานหรือสินค้าระหว่างทำ วัตถุดิบ หรือวัสดุที่ใช้ในการผลิตตามปกติของธุรกิจ

d) ที่ดิน อาคารและอุปกรณ์-สุทธิ (Property Plant and Equipment) หมายถึง สินทรัพย์ที่มีลักษณะคงทนถาวร มีอายุการใช้งาน 1 ปี

e) สินทรัพย์รวม (Total Assets)

ซึ่งในการเลือกข้อมูลในส่วนของสินทรัพย์มาใช้ในการศึกษาเนื่องจากสินทรัพย์จะเป็นตัวสะท้อนถึงตัวกิจการเองว่ามีสภาพคล่องดีแค่ไหน หรือ มีมูลค่าสินทรัพย์ที่เกิดจากการดำเนินงานมากแค่ไหน

2) ด้านหนี้สินและส่วนของผู้ถือหุ้น

- ด้านหนี้สิน (Liabilities) หมายถึง พันธะผูกพันของกิจการอันเกิดจากการค้า การกู้ยืม หรือจากการอื่น ซึ่งจะต้องชำระคืนในภายหน้าด้วยสินทรัพย์หรือบริการ
  1. หนี้สินหมุนเวียน (Current Liabilities) หมายถึง หนี้สินที่ต้องชำระให้หมดภายใน 1 ปี นับตั้งแต่วันที่ในงบดุล
  2. หนี้สินระยะยาว (Long-term Liabilities) หมายถึง เงินกู้ยืมจากบุคคลธรรมดาหรือนิติบุคคลที่มีกำหนดชำระคืนในระยะยาวโดยมีสัญญากู้ยืมเป็นหลักฐาน
  3. หนี้สินทั้งหมด (Total Liabilities)

ซึ่งในการเลือกข้อมูลในส่วนของหนี้สินมาใช้ในการศึกษาเนื่องจากหนี้สินจะเป็นตัวสะท้อนฐานะของกิจการว่ามีการก่อภาระหนี้มากแค่ไหน กิจการสามารถรับภาระหนี้ได้หรือไม่

- ส่วนของผู้ถือหุ้น (Shareholder's equity) หมายถึง สิทธิเรียกร้องของผู้ถือหุ้นที่มีต่อสินทรัพย์ของกิจการ
  1. ทุนที่ออกและเรียกชำระแล้ว (Issued and paid-up share capital) หมายถึง หุ้นที่นำออกจำหน่ายและเรียกชำระค่าหุ้นแล้ว
  2. กำไร(ขาดทุน)สะสม (Retained Earning or Deficit) หมายถึง กำไรที่เกิดขึ้นแต่ยังไม่แบ่งคืนผู้ถือหุ้น

ซึ่งในการเลือกข้อมูลในส่วนของส่วนของผู้ถือหุ้นมาใช้ในการศึกษาเนื่องจากข้อมูลจากสมการของงบการเงิน คือ สินทรัพย์ = หนี้สิน + ส่วนของผู้ถือหุ้น นั้นหมายถึงส่วนของผู้ถือหุ้นจะเป็นส่วนที่แสดงฐานะของกิจการเช่น หากกิจการดำเนินกิจการไปได้ด้วยดี ส่วนของผู้ถือหุ้นจะเพิ่มขึ้นเนื่องจากกำไรจากกการดำเนินงานที่สะสมในแต่ละปี ในทางตรงข้ามหากผลการดำเนินงานไม่ดีจะส่งผลให้มีผลขาดทุนซึ่งจะสะสมไปเรื่อยจนทำให้ส่วนของผู้ถือหุ้นลดลงหรือติดลบในที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**2. ข้อมูลที่นำมาจากงบกำไรขาดทุน ได้แก่**

1. รายได้ (Revenue) หมายถึง รายได้ที่เกิดขึ้นจากการส่งมอบสินค้า ทรัพย์สิน สิทธิหรือบริการเพื่อแลกเปลี่ยนกับเงินสด สิทธิเรียกร้องให้ชำระเงิน
2. ค่าใช้จ่าย (Expenses) หมายถึง ต้นทุนและค่าใช้จ่ายในการดำเนินงานของธุรกิจ
3. กำไรก่อนดอกเบี้ยและค่าเสื่อมราคา (Earning Before Interest and Depreciation) หมายถึง รายได้จากการดำเนินการหลังหักค่าใช้จ่ายต่างยกเว้นภาระดอกเบี้ยที่เกิดขึ้นและค่าเสื่อมราคา
4. ดอกเบี้ยจ่าย (Interest) หมายถึง ดอกเบี้ยหรือค่าตอบแทนที่คิดให้เนื่องจากการใช้ประโยชน์จากเงินหรือเงินทุน
5. ค่าเสื่อมราคา (Depreciation) หมายถึง ค่าใช้จ่ายตัดจ่ายจากสินทรัพย์ถาวรซึ่งทยอยตัดตามอายุการใช้งาน ค่าใช้จ่ายดังกล่าวไม่ถือเป็นตัวเงิน
6. กำไร(ขาดทุน)สุทธิ (Net income Profit(Loss)) หมายถึง ยอดรวมรายได้หักยอดรวมค่าใช้จ่าย

ซึ่งในการเลือกข้อมูลในส่วนของงบกำไร(ขาดทุน) ใช้ในการศึกษาเนื่องจากผลการดำเนินงานของกิจการจะบอกถึงการที่กิจการสามารถทำมาหาได้เพียงพอต่อภาระค่าใช้จ่ายที่มี รวมทั้งเพียงพอต่อภาระดอกเบี้ยที่เกิดขึ้นหรือไม่

ข้อมูลการเงินที่นำมาศึกษาปัจจัยที่มีผลต่อการเป็นลูกหนี้ที่ไม่ก่อให้เกิดรายได้ มีดังนี้

**ตารางที่ 5.1** แสดงความสัมพันธ์ของชื่อข้อมูลกับประเภทของข้อมูล

| ชื่อข้อมูล | ประเภทข้อมูล |
|---|---|
| Cash In Hand and at bank | Number |
| Account Receivable | Number |
| Inventory | Number |
| Property Plant and Equipment (net) | Number |
| Total Assets | Number |
| Current Liabilities | Number |
| Long-term Liabilities | Number |
| Total Liabilities | Number |
| Shareholder's equity | Number |
| Issued and paid-up share capital (Coommon Stock) | Number |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

| ชื่อข้อมูล | ประเภทข้อมูล |
|---|---|
| Retained Earning or Deficit | Number |
| Revenue | Number |
| Expenses | Number |
| Earning Before Interest and Depreciation(EBITDA) | Number |
| Interest | Number |
| Depreciation | Number |
| Net income Profit(Loss) | Number |
| Target | Text |

และTarget Attribute ที่จะใช้ในการพยากรณ์ คือ เป็น NPLs ซึ่งอยู่ในระหว่างการฟื้นฟูกิจการ แทนค่าด้วย a และไม่อยู่แทนค่าด้วย b

**5.1.2 การสุ่มเลือกข้อมูลและการทดสอบข้อมูล (Data Preprocessing)** คือ การนำข้อมูลที่คัดเลือกมาทำความสะอาดเพื่อกำจัดข้อมูลที่มีลักษณะ Noisy Data และ Missing Data ซึ่งชุดข้อมูลที่นำมาใช้ไม่พบข้อมูลที่มีลักษณะดังกล่าว จึงไม่ต้องมีการทำใดๆกับชุดข้อมูล

**5.1.3 การแปลงข้อมูลไปสู่ Model (Data Transformation)** คือ ซึ่งในการทำดาต้าไมนิ่งในครั้งนี้ใช้โปรแกรมสำเร็จรูป คือ Answer Tree 3.0 ซึ่งสามารถรับข้อมูลจาก Worksheet ที่เป็น Excel ได้เพียงแต่ก่อนการใช้ข้อมูลเนื่องจากข้อมูลที่นำมาใช้กำหนดหน่วยเป็นหลักบาท จึงได้มีการกำหนดหน่วยใหม่เป็นหลักล้านบาท เพื่อไม่ให้จำนวนตัวเลขที่ปรากฏมีจำนวนมากเกินไป

## 5.2 การทำดาต้าไมนิ่ง

การทำดาต้าไมนิ่งในการศึกษาใช้โปรแกรม คือ Answer Tree 3.0 ซึ่งเป็นโปรแกรมคอมพิวเตอร์ที่เรียนรู้จาก Rule จากข้อมูลเดิมเพื่อการแบ่งกลุ่มข้อมูล( Classification ) โดยแสดงในรูปแบบของ Decision Trees โปรแกรมถูกพัฒนาขึ้นโดย SPSS ซึ่งเป็นบริษัทชั้นนำในการทำงานเกี่ยวกับดาต้ามายนิ่งใช้สำหรับการสร้าง Decision Trees ซึ่งก็คือการหา rule ซึ่งได้จากประสบการณ์การสังเกต การทดลองเพื่อใช้ในการพยากรณ์ Class ของกลุ่มข้อมูลโดยใช้ค่าของตัวแปรที่นำมาใช้ในการพยากรณ์ โปรแกรม Answer Tree 3.0 สามารถใช้กับข้อมูลได้ทั้ง

- SPSS data files

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ODBC data sources
- Excel spreadsheets
- Text files

วิธีการในการสร้าง Tree ของ Answer Tree มี 4 วิธี คือ

I. Merging การหาความสัมพันธ์ของตัวแปรที่ใช้ในการพยากรณ์กับข้อมูลที่เป็น Target โดยโปรแกรมจะยุบรวมตัวแปรที่ไม่มีความสำคัญเป็นกลุ่มหนึ่ง และตัวแปรที่มีความสำคัญเป็นกลุ่มหนึ่ง

II. Splitting การเลือกจุดที่จะแตกกิ่งของ Tree โดยอาจใช้ตัวแปรประชากรซึ่งถูกเลือกโดยการเปรียบเทียบกับประชากรทั้งหมด

III. Stopping กฎซึ่งถูกกำหนดว่าเมื่อไหร่จึงจะหยุดสร้าง Tree

IV. Pruning การยุบกิ่งซึ่งเมื่อแตกออกไปแล้วมีค่าหรือกลุ่มข้อมูลที่แสดงข้อมูลน้อยและนำมาใช้ในการพยากรณ์ไม่เกิดประโยชน์

Algorithm ใน Answer Tree มี 4 Algorithm ดังนี้

- CHAID (Chi-squared Automatic Interaction Detector) เป็นการหาค่าความสัมพันธ์โดยใช้วิธีการของ chi-squared statistics เพื่อใช้ในการแตกกิ่งที่ดีที่สุด (Kass, 1980).
- Exhaustive CHAID เป็นการปรับปรุงการทำงานของ CHAID เพื่อให้สามารถหาความสัมพันธ์ของชุดข้อมูลเพื่อใช้ในการแตกกิ่งอย่างละเอียดมากขึ้น (Biggs et al., 1991).
- CART (Classification and Regression Trees) จะสร้าง binary tree โดยอาศัยการคำนวณของ Ginni index และ " twoing" (Breiman et al., 1984).
- QUEST ( Quick, Unbiased, Efficient Statistical Tree) เป็นวิธีการในการคำนวณอย่างรวดเร็วเพื่อหลีกเลี่ยงการขัดแย้งระหว่างตัวแปรที่ใช้ในการพยากรณ์ (Loh and Shih, 1997).

ผลที่ได้จาก Algorithm ทั้ง 4 วิธี นั้นจะแสดงค่าเหมือนกัน คือ การอธิบายความสัมพันธ์ของกลุ่มข้อมูลเพื่อนำมาใช้ในการพยากรณ์ (Predictive) หรือ การแยกกลุ่ม (Classification) โดย Algorithm จะเลือกค่าของตัวแปรที่ดีที่สุดในการดำเนินการซึ่งกระบวนการทำจะอาศัยการทำซ้ำหลายครั้งจนกว่าจะได้ Tree ที่คิดว่าเหมาะสมในการอธิบายค่าความสัมพันธ์ของกลุ่มข้อมูล แม้ว่า Algorithm ทั้ง 4 วิธี จะมีความแตกต่างทั้งกระบวนการคิด วิธีการในการได้ความสัมพันธ์มาและคุณลักษณะต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 5.3 การวิเคราะห์ผลลัพธ์ที่ได้

การแปรความหมายของ Tree ที่ได้จากการทำ Data Mining ขึ้นอยู่กับการวิเคราะห์ข้อมูลและความเข้าพื้นฐานต่อข้อมูลที่นำมาใช้ในการ Predictive ซึ่งในการศึกษาปัจจัยใดที่มีผลต่อการเป็น NPLs ลักษณะฟื้นฟูกิจการ จากข้อมูลทั้งหมด 202 รายการ เป็นบริษัทที่อยู่ระหว่างฟื้นฟูกิจการ 55 บริษัท และไม่อยู่ระหว่างการฟื้นฟู 147 บริษัท

โดยข้อมูลทั้งหมดไม่ได้มีการทำ Validation ของข้อมูล คือ การทดสอบการสร้าง Tree โดยการแบ่งข้อมูลเป็น 2 ชุด โดยข้อมูลชุดแรก เพื่อการ Training สำหรับให้โปรแกรมเรียนรู้ความสัมพันธ์ของกลุ่มข้อมูลที่นำมาใช้ สำหรับข้อมูลชุดที่สอง เพื่อการ Testing หรือทดสอบว่า Tree ที่สร้างนั้นมีความเหมาะสมหรือไม่ เนื่องจากจำนวนข้อมูลที่เป็น Target ค่อนข้างน้อยหรือเพียง ร้อยละ 27 จากจำนวนข้อมูลทั้งหมด ซึ่งอาจทำให้การ Test ข้อมูลอาจได้ผลที่ไม่ถูกต้องนัก โดยในการศึกษาอาจแบ่ง 3 กรณี ดังนี้

1. การศึกษาจากตัวแปรทั้งหมดทั้งจากงบดุลและงบกำไร(ขาดทุน) เพื่อหาว่าตัวแปรใดจะเป็นตัวแปรที่สำคัญที่บ่งชี้ว่าบริษัทนั้นมีแนวโน้มที่จะเป็น NPLs ลักษณะต้องฟื้นฟูกิจการ
2. การศึกษาเฉพาะตัวแปรจากงบดุล เพื่อหาตัวแปรที่สำคัญจากงบดุล ที่บ่งชี้ว่าบริษัทนั้นมีแนวโน้มที่จะเป็น NPLs ลักษณะต้องฟื้นฟูกิจการ
3. การศึกษาเฉพาะตัวแปรจากงบกำไร(ขาดทุน) เพื่อหาตัวแปรที่สำคัญจากงบกำไร(ขาดทุน) ที่บ่งชี้ว่าบริษัทนั้นมีแนวโน้มที่จะเป็น NPLs ลักษณะต้องฟื้นฟูกิจการ

**1. การศึกษาจากตัวแปรทั้งหมดทั้งจากงบดุลและงบกำไร(ขาดทุน)**

ศึกษาเป็น 2 กรณี ดังนี้

<u>กรณีศึกษา 1.1</u> ผลที่ได้ คือ โปรแกรมจะเลือกตัวแปร Share Holder หรือ ส่วนของผู้ถือหุ้น = สินทรัพย์ –หนี้สินนั้น หมายถึง ส่วนของผู้ถือหุ้นจะเป็นส่วนที่แสดงฐานะของกิจการ กล่าวคือ

- หากกิจการมีสินทรัพย์มากกว่าหนี้สิน ส่วนของผู้ถือหุ้นแสดงค่าเป็นบวก
- หากกิจการมีหนี้สินสูงกว่าสินทรัพย์ ส่วนของผู้ถือหุ้นแสดงค่าติดลบ

ซึ่งโปรแกรมใช้ Share Holder เป็นตัวหลักที่ใช้ในการแบ่งข้อมูล ซึ่งจากการอ่านจาก Tree จะได้รูปแบบ 3 รูปแบบ เป็นดังนี้

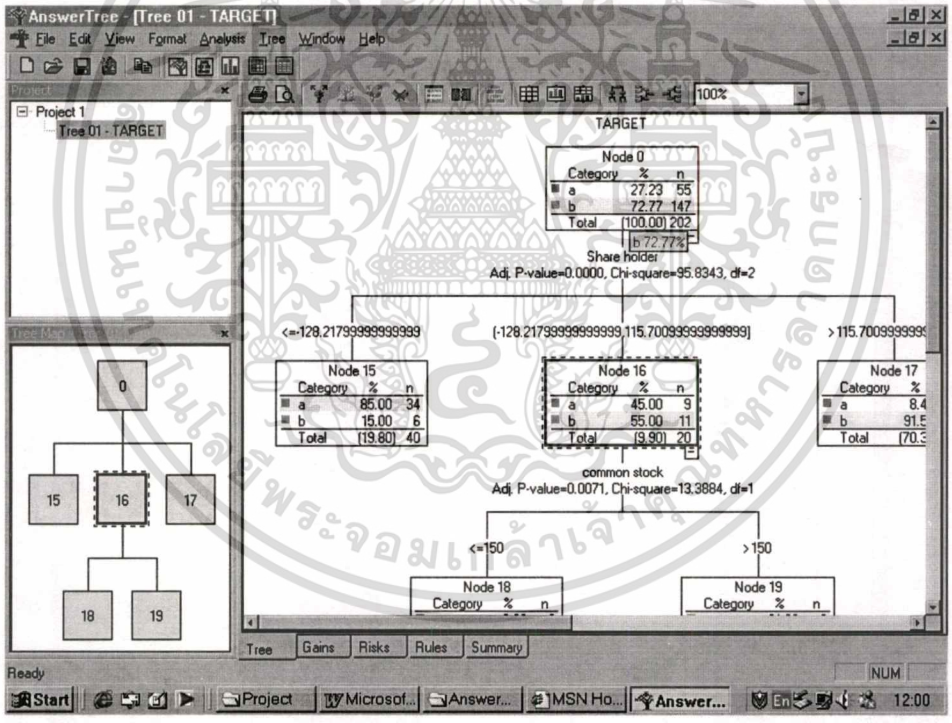
(a) หากกิจการมีส่วนของมีผู้ถือหุ้นมากกว่า 115.700 ล้านบาท จะพยากรณ์ได้ว่าไม่เป็นกิจการที่อยู่ระหว่างฟื้นฟูร้อยละ 91 ของจำนวนข้อมูลใน Node

(b) หากกิจการมีส่วนของผู้ถือหุ้นติดลบตั้งแต่ 128.217 ล้านบาท จะพยากรณ์ได้ว่าเป็นกิจการที่เป็น NPLs ที่อยู่ระหว่างฟื้นฟูกิจการร้อยละ 85 ของจำนวนข้อมูลใน Node

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

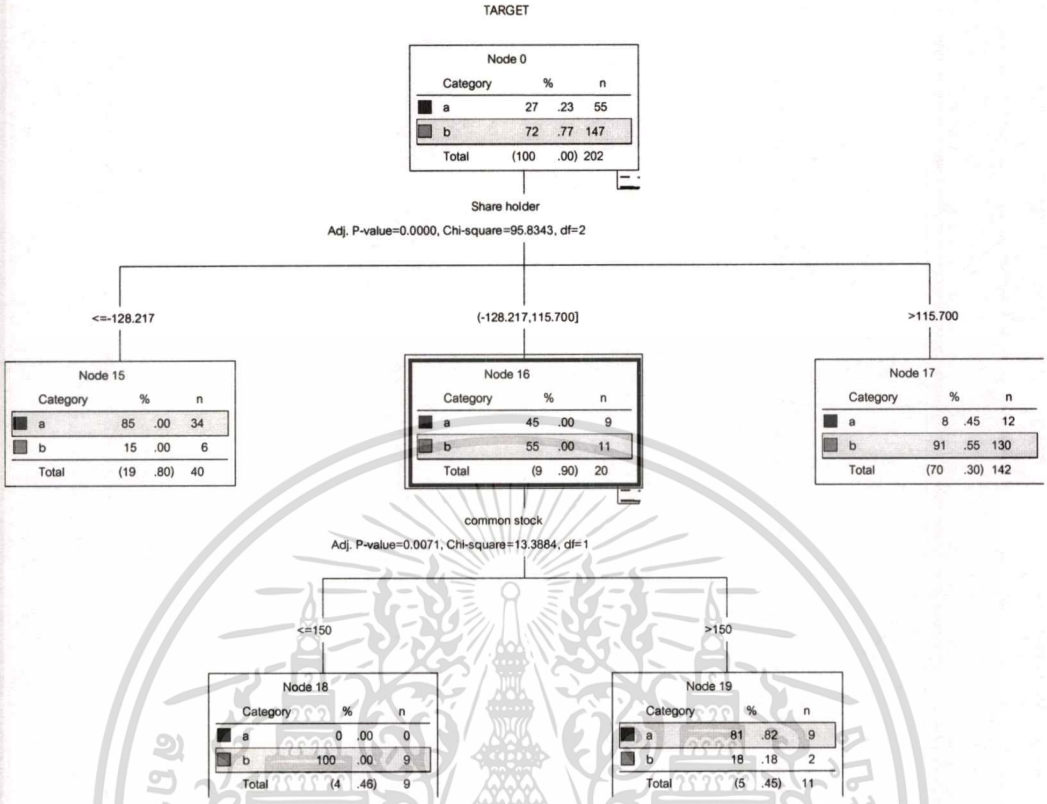
(c) หากกิจการมีทุนเรือนหุ้น (Common Stock) น้อยกว่าเท่ากับ 150 ล้านบาท และมีส่วนของผู้ถือหุ้นติดลบตั้งแต่ 128.217 ล้านบาท ถึง 115.700 ล้านบาท จะพยากรณ์ได้ว่าเป็นกิจการไม่เป็น NPLs ที่อยู่ระหว่างฟื้นฟูกิจการร้อยละ 100 ของจำนวนข้อมูลใน Node ซึ่งเมื่อพิจารณาจากข้อมูลทั้ง 9 รายการพบว่ากิจการเหล่านั้นมีส่วนของผู้ถือหุ้นเป็นบวกและมีทุนเรือนหุ้นน้อยกว่า 150 ล้านบาท ซึ่งถือเป็นรูปแบบที่ปกติ

ซึ่งจากรูปแบบ (a) และ (b) เป็นรูปแบบปกติที่เกิดขึ้นอยู่แล้ว สำหรับรูปแบบ (c) มีรูปแบบปกติแสดงไว้ด้วย นั้นหมายถึง ส่วนของผู้ถือหุ้นสามารถที่จะบ่งชี้การดำรงอยู่ของกิจการได้ นอกจากนี้ Tree ดังกล่าวมีค่าความเสี่ยงที่จะพยากรณ์ผิดพลาดเพียง ร้อยละ 9.9 เท่านั้น ซึ่งถือว่ามีความน่าเชื่อในระดับหนึ่ง รูปของ Tree แสดงดังรูปที่ 5.1 และ 5.2




**รูปที่ 5.1** แสดงหน้าจอของ Tree ที่สร้างจาก Answer Tree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




**รูปที่ 5.2** แสดง Tree ที่ได้จากการใช้ตัวแปรทั้งหมดกรณีที่ 1.1

**<u>กรณีศึกษา 1.2</u>** ผลที่ได้ คือ โปรแกรมจะเลือกตัวแปร Profit(Loss) หรือ กำไร(ขาดทุน)สุทธิ หมายถึง ผลการดำเนินงานของกิจการในรอบปี กล่าวคือ

- หากกิจการมีกำไร(ขาดทุน)สุทธิ เป็นบวก แสดงว่ากิจการมีผลการดำเนินงานดี มีกำไรหลังจากหักค่าใช้จ่ายที่เกิดขึ้น
- หากกิจการมีกำไร(ขาดทุน)สุทธิ เป็นลบ แสดงว่ากิจการมีปัญหาในการดำเนินงาน รายได้ที่มีไม่เพียงพอต่อค่าใช้จ่ายที่เกิดขึ้น

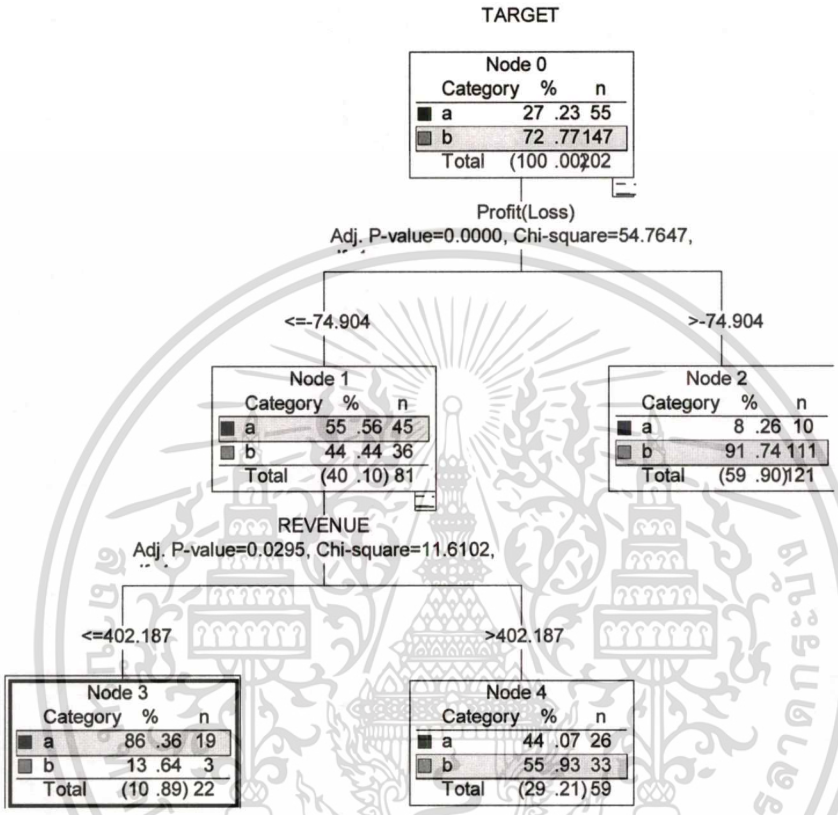
**รูปแบบที่เกิดขึ้น** จากการอ่านจาก Tree จะได้รูปแบบ 2 รูปแบบ เป็นดังนี้

(d) หากกิจการมีกำไร(ขาดทุน)สุทธิ > –74.90 ล้านบาท จะพยากรณ์ได้ว่าไม่เป็นกิจการที่อยู่ระหว่างฟื้นฟูร้อยละ 91.74 ของจำนวนข้อมูลใน Node

(e) หากกิจการกำไรสะสม(ขาดทุนสะสม) <= –74.90 ล้านบาท และมีรายได้ <= 402.18 ล้านบาท จะพยากรณ์ได้ว่าเป็นกิจการที่อยู่ระหว่างฟื้นฟูร้อยละ 86.36 ของจำนวนข้อมูลใน Node

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นั้นหมายถึงกิจการใดมีผลขาดทุนสุทธิจากการดำเนินงาน กิจการนั้นมีรายได้ไม่เพียงพอต่อภาระที่เกิดขึ้นและมีปัญหาในการดำเนินการ จึงมีแนวโน้มที่จะเป็น NPLs รูปของ Tree แสดงดังรูปที่ 5.3



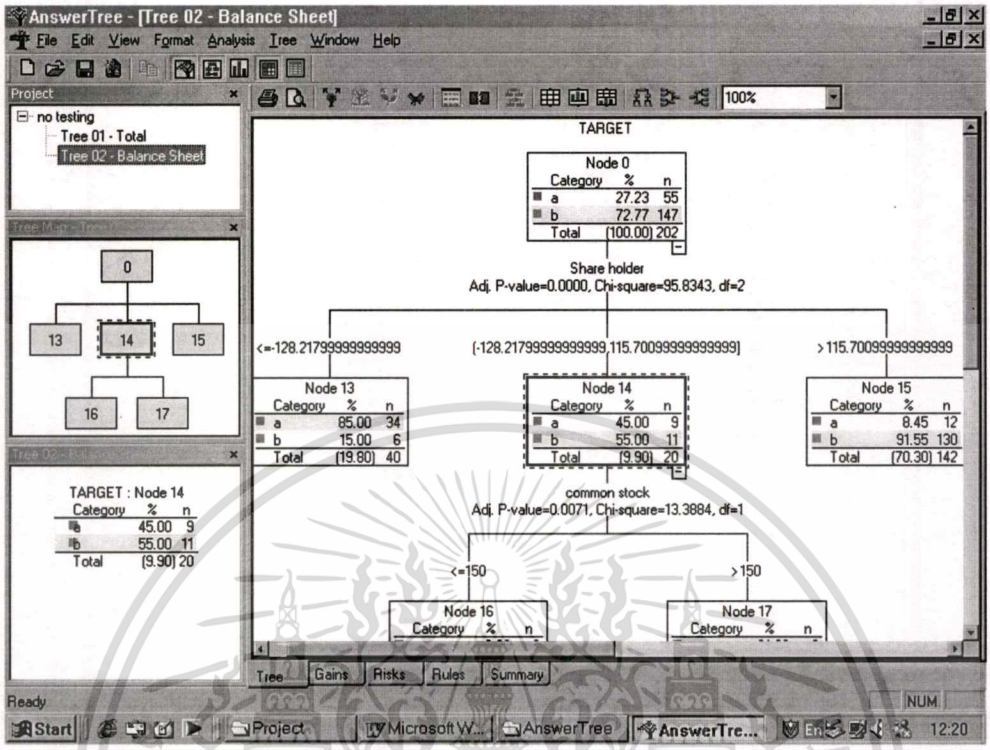

**รูปที่ 5.3** แสดง Tree ที่ได้จากการสร้างจากตัวแปรรวมกรณีที่ 1.2

**2. การศึกษาจากตัวแปรจากงบดุล**

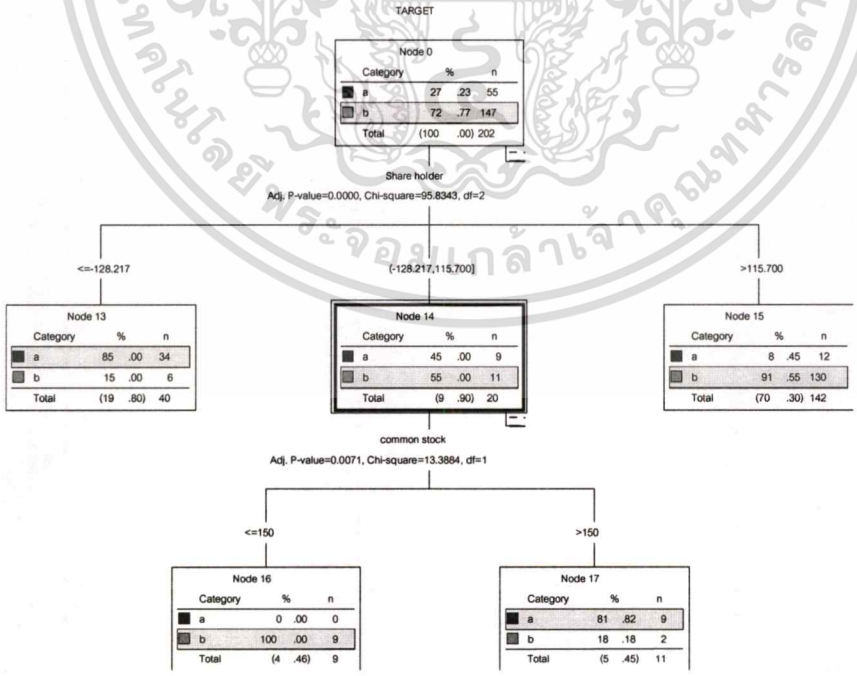
ศึกษาเป็น 2 กรณี ดังนี้

**<u>กรณีศึกษา 2.1</u>** ผลที่ได้ คือ โปรแกรมจะเลือกตัวแปร Share Holder หรือ ส่วนของผู้ถือหุ้น ซึ่งให้ผลเหมือนกับกรณีที่ 1 ซึ่งใช้ตัวแปรทั้งหมดจากงบดุลและงบกำไร(ขาดทุน) และมี Tree เป็นรูปเดียวกัน ซึ่งมีความเสี่ยงในการพยากรณ์ผิดพลาดเท่ากับร้อยละ 9.9 นั้นหมายถึง เป็นการย้ำว่าส่วนของผู้ของผู้ถือหุ้น สามารถพยากรณ์ได้ว่ากิจการใดมีแนวโน้มที่จะเป็น NPLs ลักษณะที่ต้องฟื้นฟูกิจการ ตามที่กล่าวมาแล้วข้างต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




**รูปที่ 5.4** แสดงหน้าจอการสร้าง Tree จากตัวแปรจากงบดุล




**รูปที่ 5.5** แสดง Tree ที่ได้จากการสร้างจากตัวแปรจากงบดุลกรณีที่ 2.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**<u>กรณีศึกษาที่ 2.2</u>** ผลที่ได้ คือ โปรแกรมจะเลือกตัวแปร Profit(Deficit) หรือ กำไรสะสม(ขาดทุนสะสม) หมายถึง ผลการดำเนินงานสะสมของกิจการ กล่าวคือ

- หากกิจการมีกำไรสะสม(ขาดทุนสะสม) เป็นบวก แสดงว่ากิจการสามารถดำเนินงานได้ผลดี
- หากกิจการมีกำไรสะสม(ขาดทุนสะสม)เป็นลบ แสดงว่ากิจการมีปัญหาในการดำเนินงาน

จากรูปแบบที่เกิดขึ้น จากการอ่านจาก Tree จะได้รูปแบบ 3 รูปแบบ เป็นดังนี้

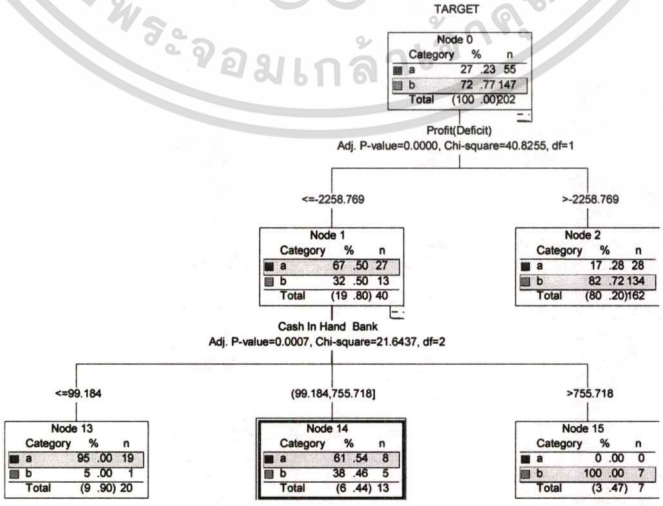
(f) หากกิจการมีกำไรสะสม(ขาดทุนสะสม) > –2,258.77 ล้านบาท จะพยากรณ์ได้ว่าไม่เป็นกิจการที่อยู่ระหว่างฟื้นฟูร้อยละ 100 ของจำนวนข้อมูลใน Node

(g) หากกิจการกำไรสะสม(ขาดทุนสะสม) <= –2,258.77 ล้านบาท เนื่องจากค่าที่ได้ค่อนข้างใกล้เคียงกัน โปรแกรมจึงนำตัวแปร Cash In Hand & Bank เข้ามาใช้ในการพยากรณ์เพิ่ม โดยจะได้ผลว่า

กิจการที่มี Cash In Hand & Bank <= 99.18 ล้านบาท จะพยากรณ์ได้ว่าเป็นกิจการที่อยู่ระหว่างฟื้นฟูร้อยละ 100 ของจำนวนข้อมูลใน Node

แต่ถ้ากิจการมี Cash In Hand & Bank > 755.72 ล้านบาท จะพยากรณ์ได้ว่าเป็นกิจการที่ไม่อยู่ระหว่างฟื้นฟูร้อยละ 100 ของจำนวนข้อมูลใน Node

นั้นหมายถึงกิจการใดที่มีผลขาดทุนสะสมจำนวนมากและมีเงินสดคงเหลือต่ำมีแนวโน้มที่จะเป็นกิจการที่จะเป็น NPLs รูปของ Tree แสดงดังรูปที่ 5.6




**รูปที่ 5.6** แสดง Tree ที่ได้จากการสร้างจากตัวแปรจากงบดุลกรณีที่ 2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**3. การศึกษาจากตัวแปรจากงบกำไร(ขาดทุน)**

ผลที่ได้ คือ โปรแกรมจะเลือกตัวแปร EBITDA ( Earning Before Interest Tax and Depreciation) หรือ กำไรก่อนดอกเบี้ยและค่าเสื่อมราคา หมายถึง กำไรจากการดำเนินงานหลังค่าใช้จ่ายในการดำเนินงานทั้งหมดแล้ว กล่าวคือ

- หากกิจการมีกำไรก่อนดอกเบี้ยและค่าเสื่อมราคามากกว่าภาระดอกเบี้ยที่เกิดขึ้น หมายถึงกิจการสามารถรับภาระหนี้ที่มีอยู่ได้ ณ ขณะนั้นได้
- หากกิจการมีกำไรก่อนดอกเบี้ยและค่าเสื่อมราคาน้อยกว่าภาระดอกเบี้ยที่เกิดขึ้น หมายถึงกิจการไม่สามารถรับภาระหนี้ที่มีอยู่ได้ ณ ขณะนั้นได้

ซึ่งโปรแกรมใช้ EBITDA เป็นตัวหลักที่ใช้ในการแบ่งข้อมูล ซึ่งจากการอ่านจาก Tree จะได้รูปแบบ 3 รูปแบบ เป็นดังนี้

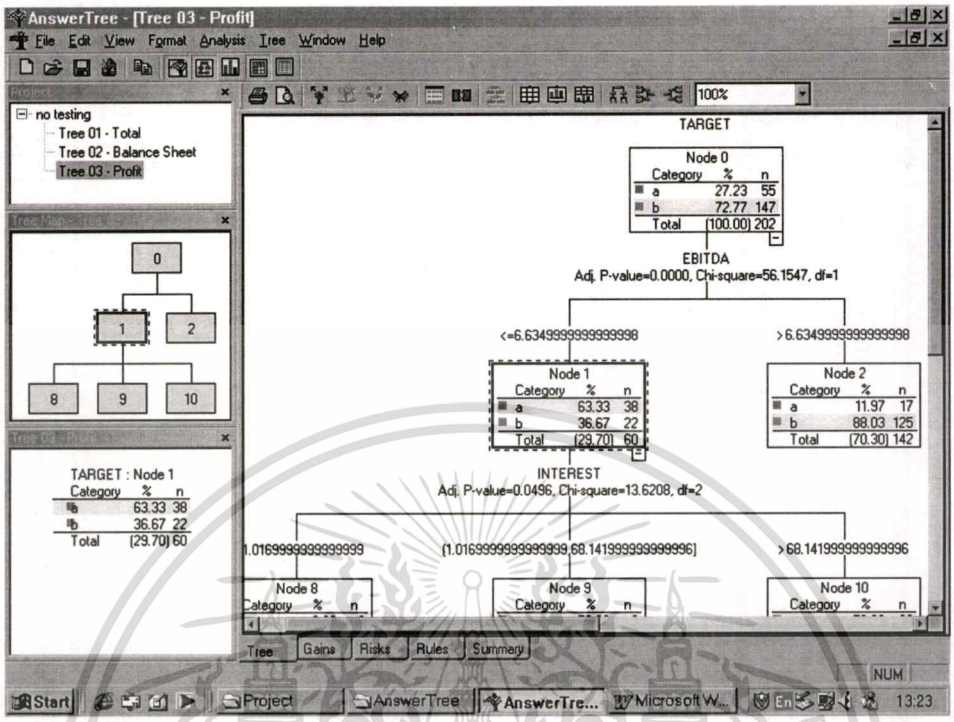
(h) หากกิจการมีดอกเบี้ยจ่ายมากกว่า 68.141 ล้านบาท ขณะที่มี EBITDA น้อยกว่า 6.634 ล้านบาท จะพยากรณ์ได้ว่าเป็นกิจการที่อยู่ระหว่างฟื้นฟูร้อยละ 78 ของจำนวนข้อมูลใน Node

(i) หากกิจการมีดอกเบี้ยจ่ายน้อยกว่า 1.016 ล้านบาท ขณะที่มี EBITDA น้อยกว่า 6.634 ล้านบาท จะพยากรณ์ได้ว่าเป็นกิจการที่ไม่อยู่ระหว่างฟื้นฟูร้อยละ 100 ของจำนวนข้อมูลใน Node

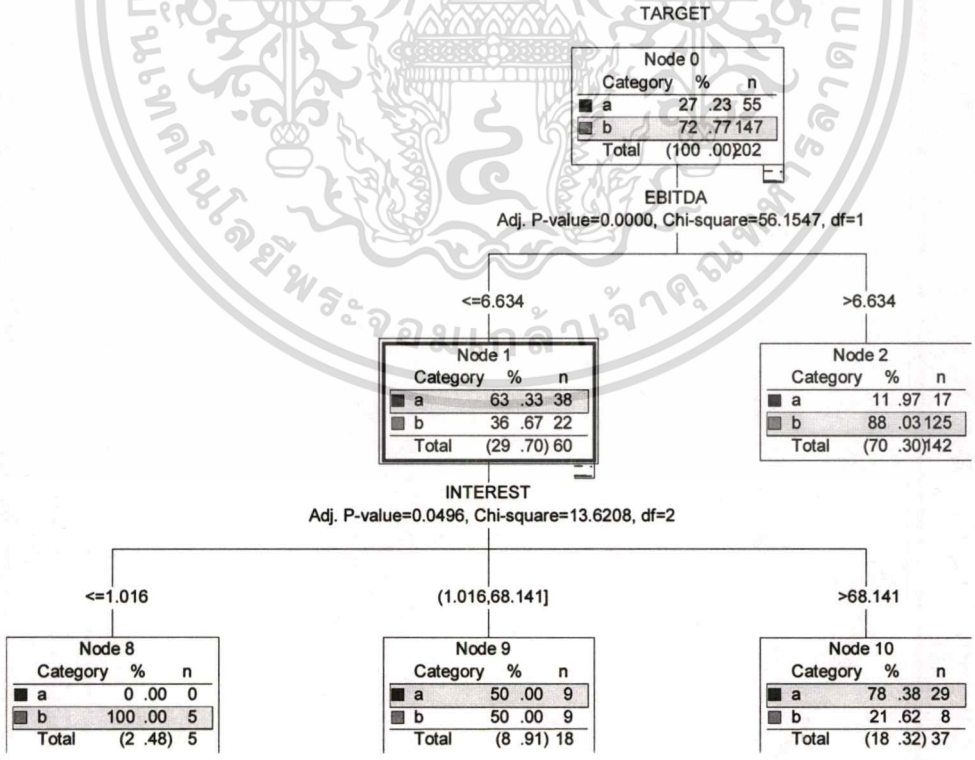
(j) หากกิจการมี EBITDA มากกว่า 6.634 ล้านบาท จะพยากรณ์ได้ว่าเป็นกิจการที่อยู่ไม่อยู่ระหว่างฟื้นฟูร้อยละ 88 ของจำนวนข้อมูลใน Node

นั้นหมายถึงกิจการใดที่มี EBITDA น้อยกว่า ดอกเบี้ยจ่าย แสดงว่ามีแนวโน้มเป็น NPLs รูปของ Tree แสดงดังรูปที่ 5.7 และ 5.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




**รูปที่ 5.7** แสดงหน้าจอการสร้าง Tree จากตัวแปรจากงบกำไร(ขาดทุน)




**รูปที่ 5.8** แสดงรูป Tree ที่สร้างจากตัวแปรในงบกำไร(ขาดทุน)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 6
# บทสรุป

## 6.1 บทสรุป

สำหรับการศึกษาถึงปัจจัยใดที่สามารถบ่งชี้ได้ว่ากิจการใดมีแนวโน้มที่จะเป็นหนี้ที่ไม่ก่อให้เกิดรายได้เพื่อสามารถนำรูปแบบที่ศึกษามา Predictive จากงบการเงินของลูกหนี้ในปัจจุบัน และหาทางป้องกันหรือแก้ไขก่อนที่ลูกหนี้จะกลายเป็นหนี้ที่ไม่ก่อให้เกิดรายได้ (NPLs) โดยใช้รูปแบบของดาต้าไมนิ่ง คือ Predictive Model โดยอาศัยเทคนิคของ Decision Trees และใช้ Exhaustive CHAID Algorithm ในการศึกษา ซึ่ง Algorithm ดังกล่าวมีการใช้งานอย่างกว้างขวาง สำหรับข้อมูลที่นำมาใช้คืองบการเงินในตลาดหลักทรัพย์ของบริษัทที่อยู่ระหว่างฟื้นฟูกิจการซึ่งถือเป็นหนี้ที่ไม่ก่อให้เกิดรายได้รูปแบบหนึ่งที่สามารถนำข้อมูลมาใช้ได้และบริษัทที่ไม่อยู่ระหว่างฟื้นฟู

ผลจากการศึกษาพบว่าปัจจัยที่โปรแกรมนำมาใช้ในการพยากรณ์เป็นปัจจัยแรกและเป็นปัจจัยหลักที่บ่งชี้กิจการที่มีแนวโน้มจะเป็น NPLs คือ ส่วนผู้ถือหุ้นและกำไรก่อนดอกเบี้ยและค่าเสื่อมราคา โดยหากกิจการใดมีส่วนของผู้ถือหุ้นติดลบหรือมีกำไรก่อนดอกเบี้ยและค่าเสื่อมราคาดอกเบี้ยน้อยกว่าภาระดอกเบี้ยจ่ายที่เกิดขึ้น นั้นแสดงว่ากิจการนั้นไม่สามารถดำเนินกิจการได้เป็นปกติคืออาจมีภาระหนี้สูงกว่าความสามารถของกิจการจะรับได้ จึงต้องมีการปรับโครงสร้างหนี้ใหม่ให้เหมาะสม โดยจากปัจจัยทั้งสองดังกล่าวมีค่าความผิดพลาดในการพยากรณ์ ร้อยละ 9.9 และ ร้อยละ 16 ตามลำดับ ผลที่ได้จากการทำดาต้าไมนิ่งในครั้งนี้นั้นออกมาตรงตามความเป็นจริงที่ใช้งานอยู่ในปัจจุบัน แต่ทั้งนี้เนื่องจากข้อมูลที่นำมาเป็นตัวแทนทางสถิติอาจมีจำนวนไม่มากพอ ทำให้ไม่สามารถเป็นตัวแทนได้อย่างสมบูรณ์ได้

## 6.2 ปัญหาและอุปสรรค

ธนาคารแห่งประเทศไทยซึ่งเป็นหน่วยงานที่ดูแลเก็บข้อมูลกิจการที่เป็นลูกหนี้ที่ไม่ก่อให้เกิดรายได้ ( Npls ) ไม่สามารถเปิดเผยข้อมูลให้ทราบได้ ในการศึกษาครั้งนี้เป็นกิจการที่อยู่ในตลาดหลักทรัพย์ที่อยู่ระหว่างฟื้นฟูเท่านั้น ทำให้ในการศึกษารูปแบบของกิจการที่เป็น NPLs อาจไม่เพียงพอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บรรณานุกรม


คณาจารย์ภาควิชาคณิตศาตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย 2530 กรกฎาคม 22 **ความน่าจะเป็นและสถิติ** โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย

จรัญ จันทลักขณาและอนันต์ชัย เชื่อนธรรม 2535 **สถิติเบื้องต้นแบบประยุกต์** บริษัทโรงพิมพ์ไทยวัฒนาพานิช จำกัด

ชนินทร์ พิทยาวิวิธและชาญชัย คำรงบุล. สิงหาคม 2545. **วิกฤติเศรษฐกิจ ธุรกิจ NPL สถาบันการเงินจะอยู่รอดได้อย่างไร** กรุงเทพฯ

ธนาคารแห่งประเทศไทย สายปรับปรุงโครงสร้างหนี้. ธันวาคม 2544. **ปรับปรุงโครงสร้างหนี้ บทเรียนและประสบการณ์จากวิกฤติเศรษฐกิจไทย** กรุงเทพฯ

อรุณี อย่างธารา และคณะ มหาวิทยาลัยธรรมศาสตร์. 2539. **การบัญชีการเงิน** โรงพิมพ์ธรรมศาสตร์

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

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

**Decision Trees for Predictive Modeling**. 2002. December 15 [Online]. Available: http://www..datamining.com

Information Discovery, Inc.Marketing Communications, **" A Characterization of Data Mining Technologies and Process"** 2002 December 15 [Online]. Available: http://www.datamining.com.

Information Discovery, Inc.Marketing Communications, **" Rules are Much More than Decision Trees"** 2002 December 15 [Online]. Available: http://www.datamining.com.

**Predictive Dataming Selecting the right solution.** 2002 December 15 [Online]. Available: http://www.datamining.


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# ภาคผนวก ก

# ขั้นตอนการใช้งานโปรแกรมสำเร็จรูป Answer Tree 3.0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**ขั้นตอนการใช้งานของโปรแกรม Answer Tree**

1. เริ่ม Start New Project โดยการเลือกคำสั่ง Start New Project และเลือกประเภทข้อมูลที่ต้องการเรียกใช้ ซึ่งในกรณีนี้ใช้ข้อมูลที่เป็นตารางของ Excel ซึ่งหน้าจอในการทำงานแสดงตามรูปที่ ก.1 และ ก.2
2. การเลือกไฟล์ข้อมูลของ Excel ที่ต้องการและเรียก Sheet ที่ต้องการนำมาใช้งาน ตามรูปที่ ก.3
3. การสร้าง Tree เลือกคำสั่ง New Tree เพื่อที่จะสร้าง Tree โดยผู้ใช้สามารถเลือก Option ในการสร้าง Tree

3.1. ในขั้นแรก เลือกวิธีการในการสร้าง Tree ซึ่ง Answer Tree มี Algorithm 4 วิธี ตามที่กล่าวไว้ข้างต้น สำหรับการศึกษาในครั้งนี้เลือก Exhuastive CHAID ในการสร้าง Tree เนื่องจากข้อมูลที่ใช้ส่วนใหญ่เป็นตัวเลข ดังแสดงตามรูปที่ ก.4

3.2 การเลือก Target และการเลือกตัวแปร (Variable) ที่จะใช้ในการสร้าง Tree ดังแสดงตามรูปที่ ก.5

3.3 การกำหนด Validation ของข้อมูล คือ การทดสอบการสร้าง Tree โดยการแบ่งข้อมูลเป็น 2 ชุด โดยข้อมูลชุดแรก เพื่อการ Training สำหรับให้โปรแกรมเรียนรู้ความสัมพันธ์ของกลุ่มข้อมูลที่นำมาใช้ สำหรับข้อมูลชุดที่สอง เพื่อการ Testing หรือทดสอบว่า Tree ที่ได้จากการสร้างนั้นมีความถูกต้องตรงกันไหม โดยเปอร์เซนต์ที่ใช้ในการทดสอบนั้นขึ้นอยู่กับจำนวนข้อมูลที่มีว่าต้องมีการทดสอบที่ระดับใด ดังแสดงตามรูปที่ ก.6 และ ก.7

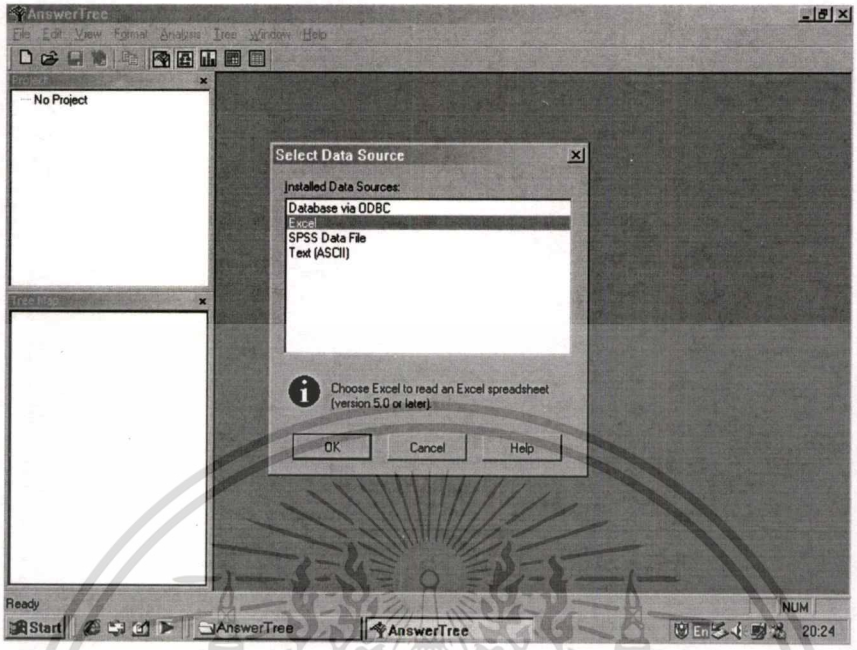
3.4 การกำหนด Options ในการกำหนดค่าที่หยุดในการแตกกิ่งหรือสร้าง Treeg เนื่องจากข้อมูลที่อยู่ใน Node ข้างล่างลงไปอาจมีจำนวนข้อมูลที่น้อยและไม่สามารถแสดงถึงความสัมพันธ์ได้เพียงพอ โดยการกำหนดจำนวนข้อมูลขั้นต่ำใน รวมทั้งจำนวน Parent Node และ Child Nodeซึ่งในการเปลี่ยนแปลงค่าดังกล่าวจะมีผลต่อการสร้าง Tree ดังแสดงในรูปที่ ก.8

4 เมื่อกำหนดเงื่อนไขในการสร้าง Tree เสร็จแล้ว โปรแกรมจะทำการคำนวณเพื่อหาค่าความสัมพันธ์ของกลุ่มข้อมูลโดย Rule ในการสร้าง Tree ซึ่งเมื่อโปรแกรมดำเนินการเสร็จแล้วจะแสดง Root Node ดังแสดงตามรูป ก.9

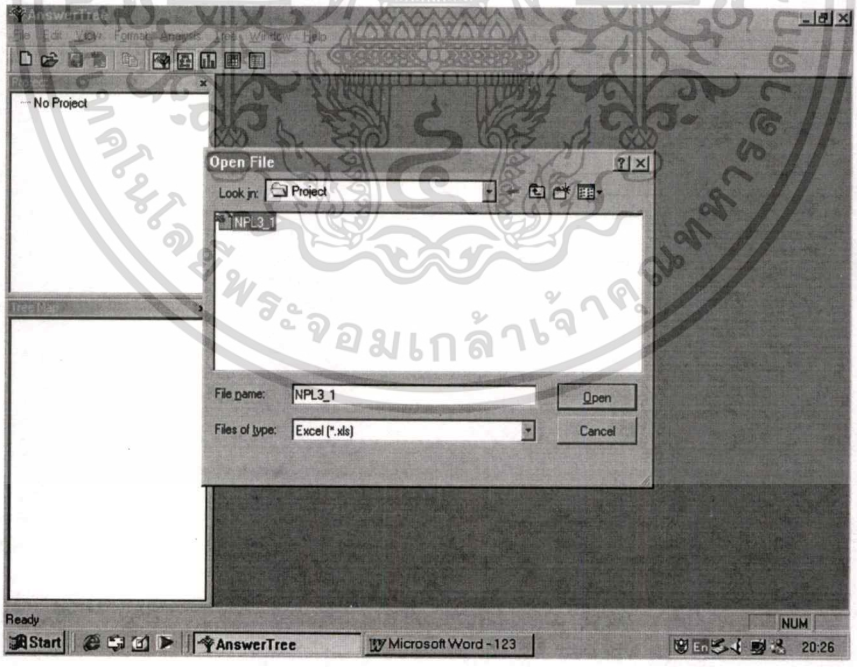
5 ขั้นต่อไป คือ การ Grow Tree เพื่อแตกกิ่งตามที่โปรแกรมได้มีการคำนวณมา ดังแสดงตามรูปที่ ก.10

6 หลังจากได้ Tree แล้ว โปรแกรมจะแสดงข้อมูลที่ได้จากการสร้าง Tree ดังแสดงตามรูปที่ ก.11 รวมทั้งคำนวณ Risk ของ Tree ที่สร้างว่ามีค่าความเสี่ยงในการพยากรณ์ผิดพลาดกี่เปอร์เซนต์ ดังแสดงตามรูปที่ ก.12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

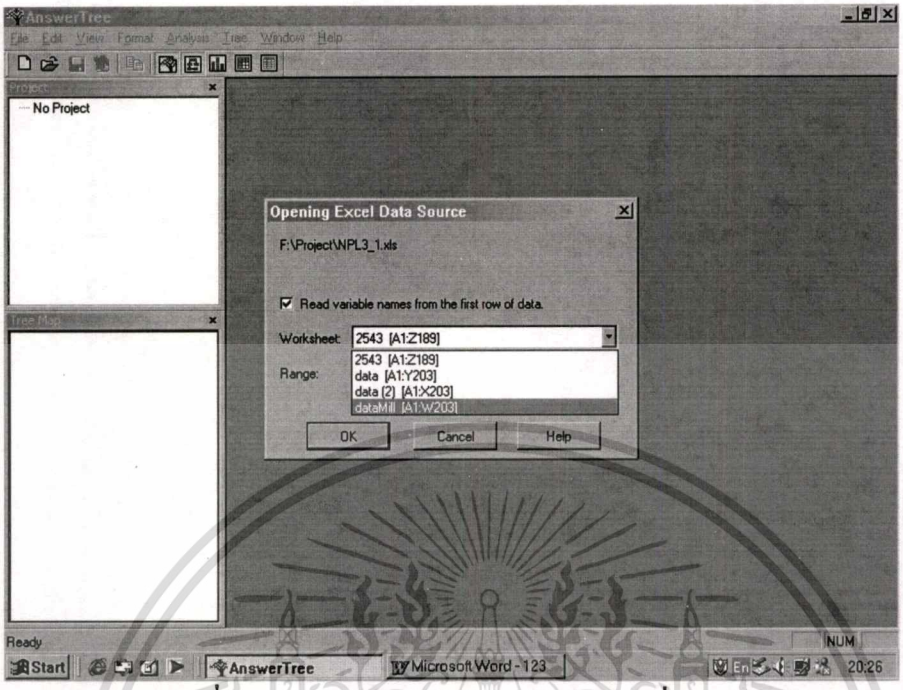



**รูปที่ ก.1** แสดงหน้าจอการนำประเภทข้อมูลเข้าสู่โปรแกรม Answer Tree

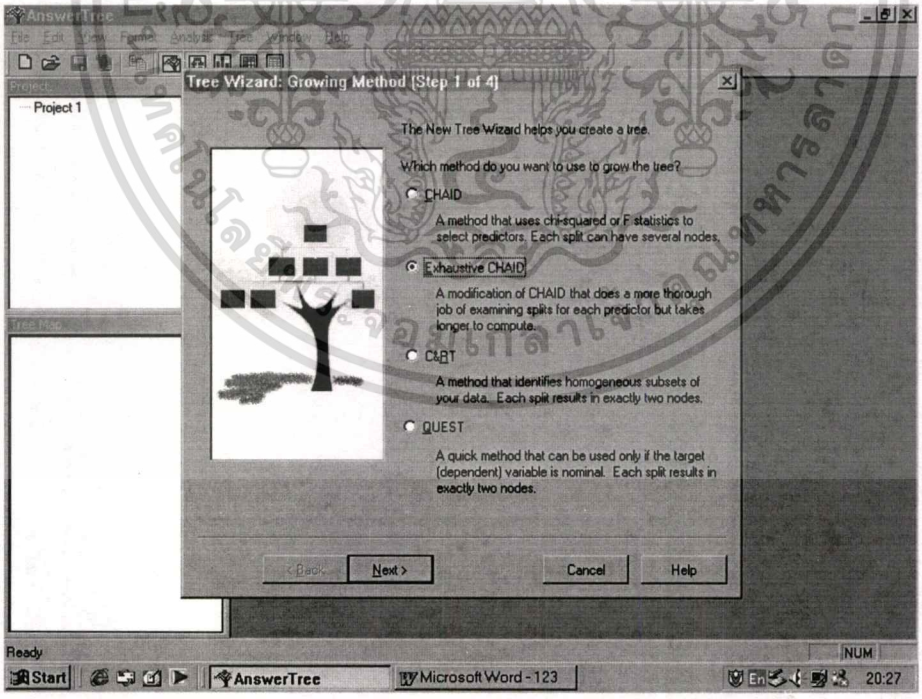



**รูปที่ ก.2** แสดงหน้าจอการเลือกไฟล์ข้อมูลเพื่อเข้าสู่โปรแกรม Answer Tree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

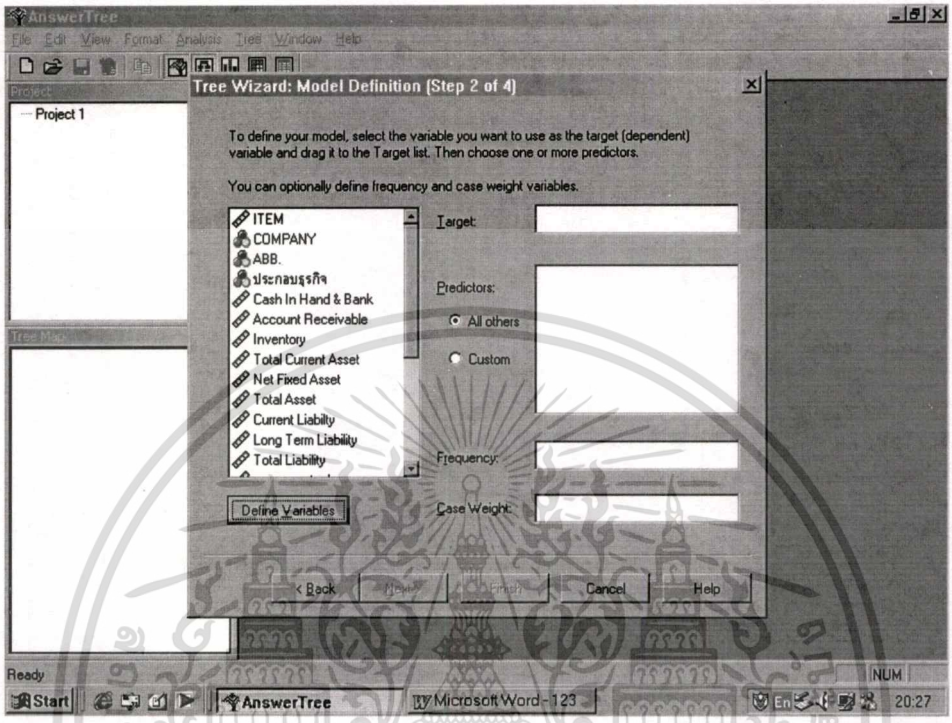



**รูปที่ ก.3** แสดงหน้าจอการเลือกไฟล์ข้อมูลที่จะนำไปใช้

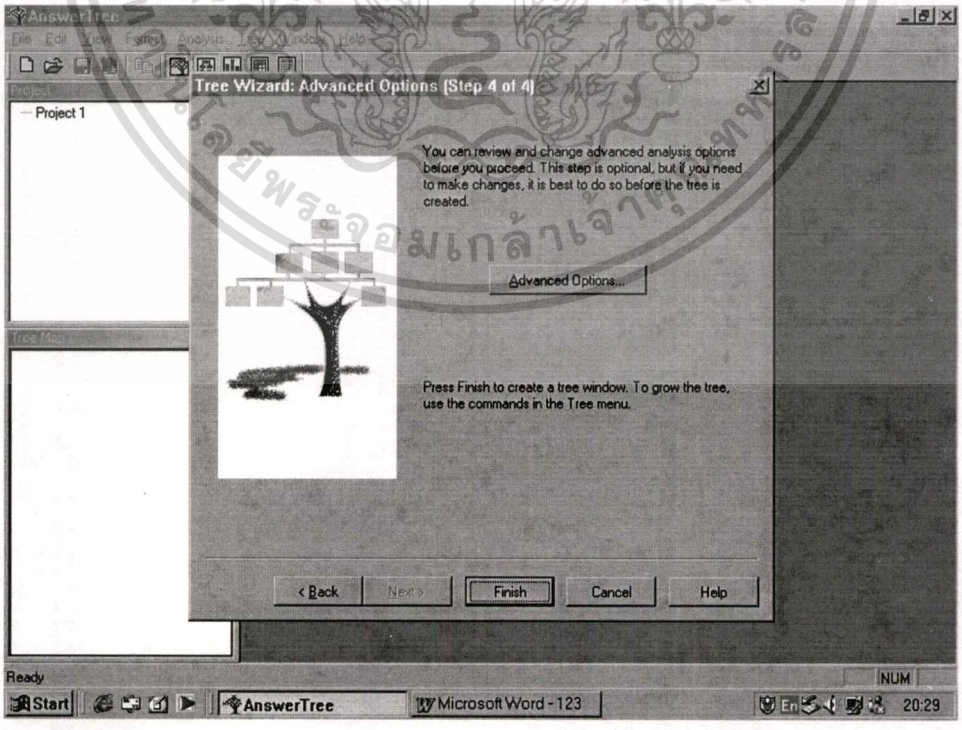



**รูปที่ ก.4** แสดงหน้าจอการเลือก Algorithm ที่ต้องการใช้สร้างTree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

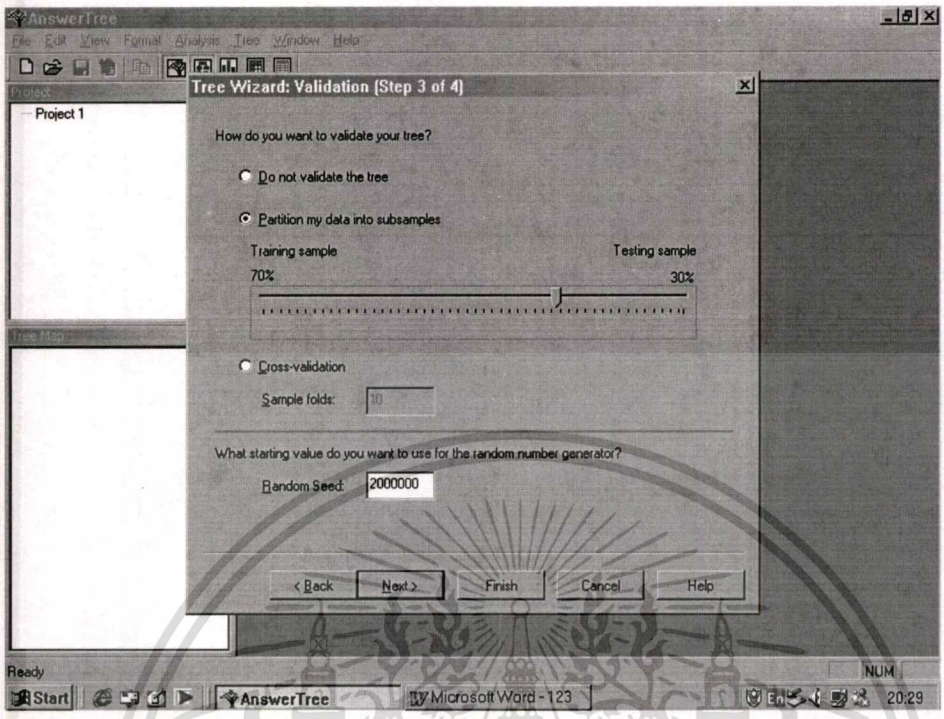



**รูปที่ ก.5** แสดงหน้าจอการเลือกตัวแปรที่นำเข้ามาใช้ในโปรแกรม

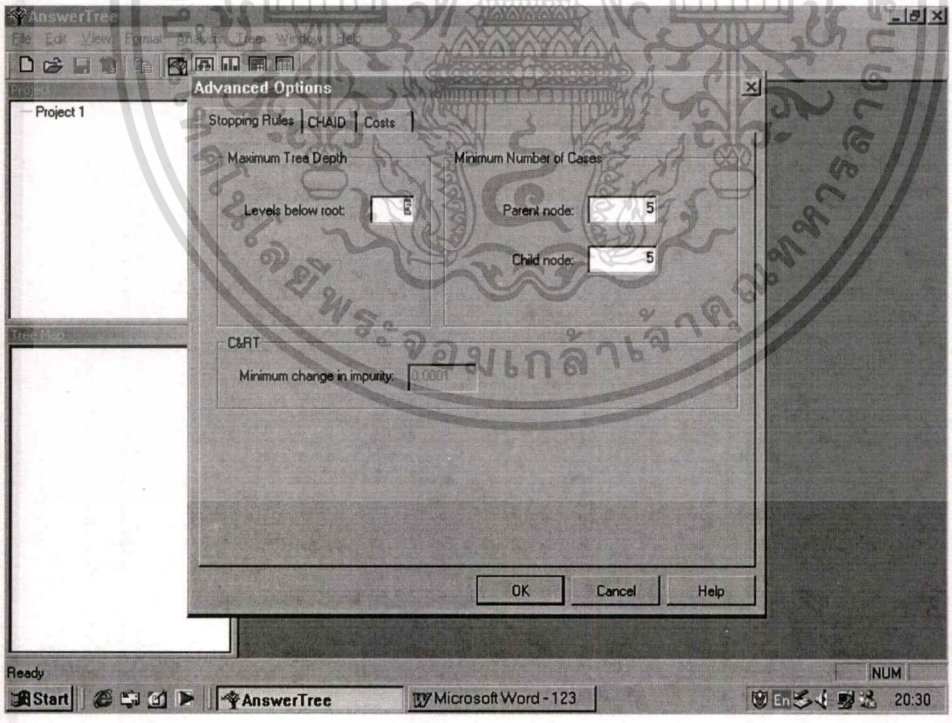



**รูปที่ ก.6** แสดงหน้าจอการเลือก Options ในการสร้าง Tree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

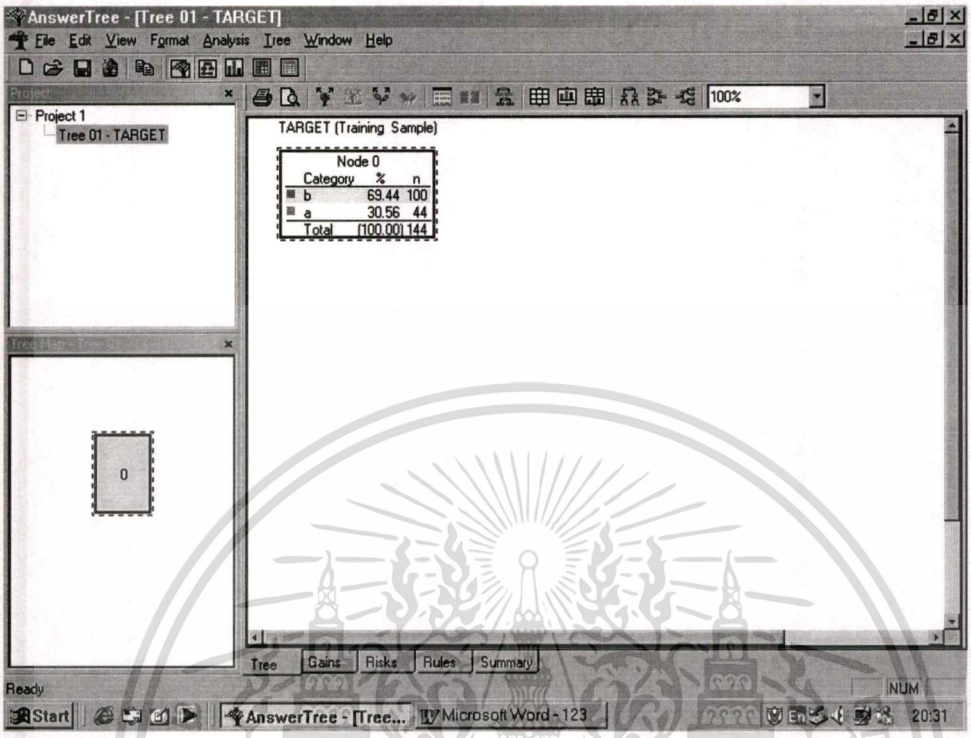



**รูปที่ ก.7** แสดงหน้าจอเปอร์เซ็นการกำหนด Test และ Train

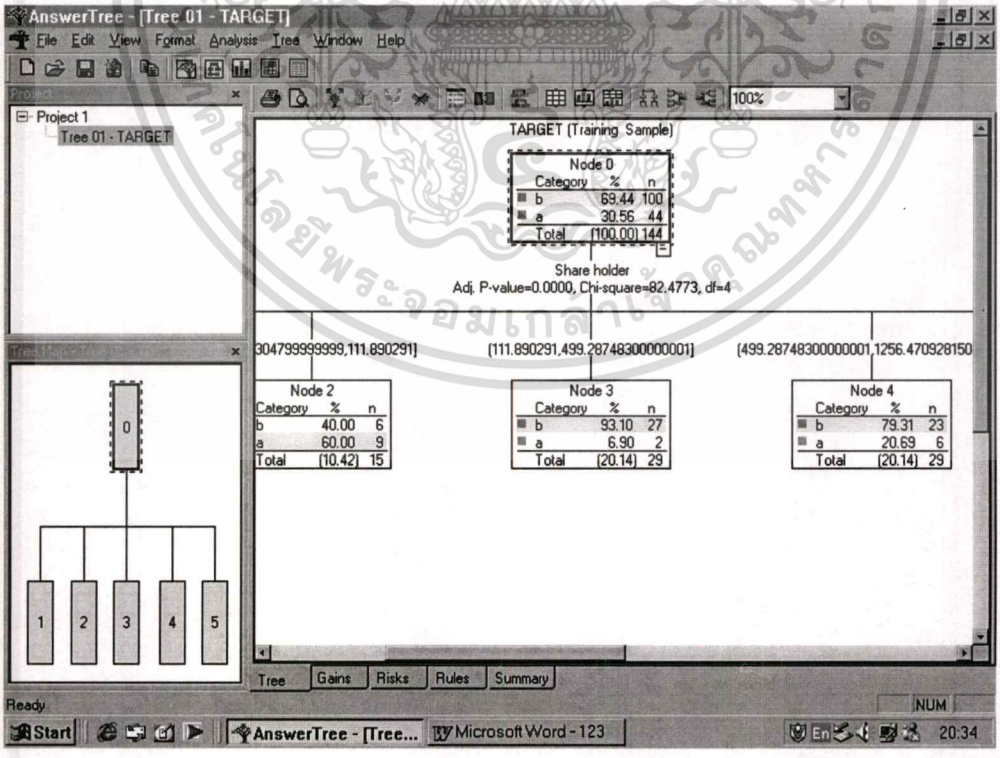



**รูปที่ ก.8** แสดงหน้าจอการกำหนด Options ในการสร้าง Tree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

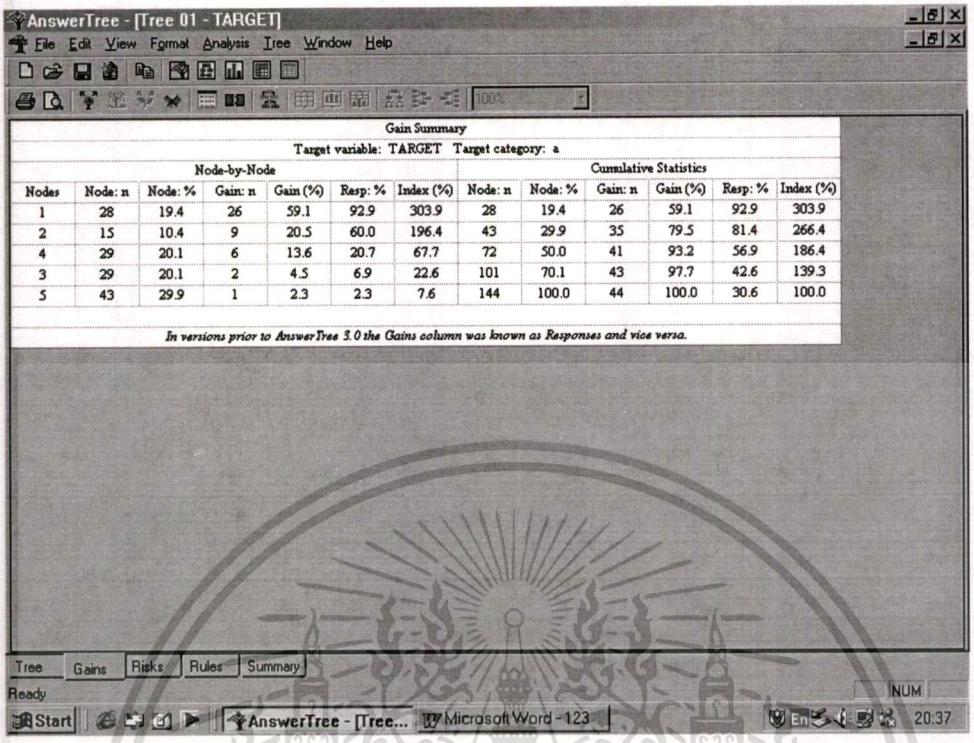



**รูปที่ ก.9** แสดงหน้าจอ Node แรกก่อนการแตกกิ่งของ Tree

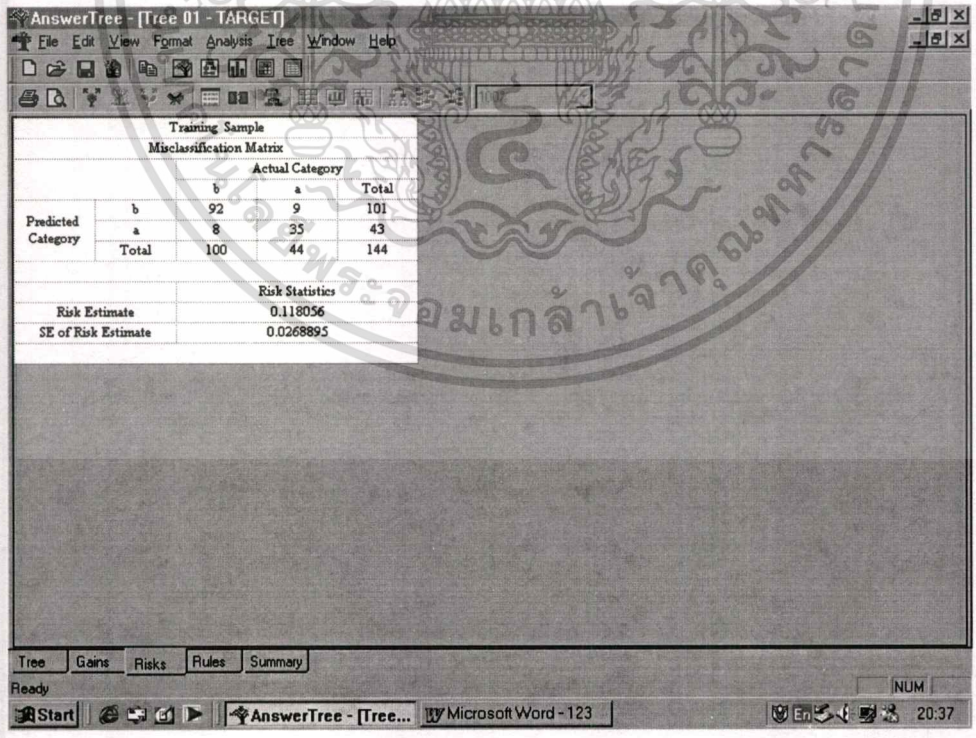



**รูปที่ ก.10** แสดงหน้าจอ Tree ที่สร้างเสร็จแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้




**รูปที่ ก.11** แสดงหน้าจอข้อมูลที่ได้จากการสร้าง Tree




**รูปที่ ก.12** แสดงหน้าจอค่าความเสี่ยงในการพยากรณ์จากการสร้าง Tree

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# ประวัติผู้เขียน

นางสาวปรียากร สัปตา เกิดวันที่ 15 กันยายน พ.ศ.2514 สำเร็จการศึกษาบัญชีบัณฑิต จาก คณะพาณิชย์ศาสตร์และการบัญชี มหาวิทยาลัยธรรมศาสตร์ ในปีการศึกษา 2536 และศึกษาต่อใน ระดับปริญญามหาบัณฑิต สาขาการจัดการเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ในปีการศึกษา 2545 ปัจจุบันทำงานใน ตำแหน่ง ผู้จัดการ(หน่วยสินเชื่อ) ให้กับ บรรษัทเงินทุนอุตสาหกรรมแห่งประเทศไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้