

ห้องสมุดคณะเทคโนโลยีสารสนเทศ ผจธ.

การพัฒนาระบบงานเพื่อการค้นพบความรู้จากระบบบริหารสินค้าคงคลัง

Knowledge Discovery in an Inventory Management System



611704218  
112842758

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน  
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ  
ภาคเรียนที่ 2 ปีการศึกษา 2547  
คณะเทคโนโลยีสารสนเทศ  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับ..... ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลง \*H002301\* ต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อหัวข้อ	การพัฒนาระบบงานเพื่อการค้นพบความรู้จากระบบบริหารสินค้าคงคลัง
นักศึกษา	นายณรงค์ฤทธิ์ อนันต์ชัยพัชฌนา
อาจารย์ที่ปรึกษา	ผศ.ดร. ภัทรชัย ลลิตโรจน์วงศ์
ระดับการศึกษา	วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2547

### บทคัดย่อ

ในโครงการนี้กล่าวถึงกระบวนการในการพัฒนาระบบด้วยเทคโนโลยีดาต้าไมนิ่ง ซึ่งจะมีการยกตัวอย่างของข้อมูลการบริหารจัดการระบบสินค้าคงคลัง โดยมีวัตถุประสงค์เพื่อการสืบค้นความรู้จากฐานข้อมูลตัวอย่าง แล้วนำความรู้ที่ได้ไปช่วยแก้ปัญหาในการบริหารจัดการระบบบริหารจัดการสินค้าคงคลัง ให้สามารถลดระดับสินค้าภายในคลังสินค้าให้เหลือปริมาณในระดับที่น้อยที่สุด ในขณะที่เดียวกันก็มีความสามารถในการสำรองวัตถุดิบให้เพียงพอต่อการผลิตเพื่อตอบสนองความต้องการของลูกค้าได้อย่างเพียงพอ ซึ่งเป็นเป้าหมายหลักที่ระบบบริหารสินค้าคงคลังทั่วไปต้องการ โดยเริ่มตั้งแต่ขั้นตอนการเตรียมข้อมูลจากข้อมูลดิบโดยใช้เครื่องมือดาต้าแวร์เฮาส์และ OLAP เข้ามาช่วยเพื่อกลั่นกรองข้อมูลก่อนที่จะนำข้อมูลที่ได้เข้าสู่กระบวนการวิเคราะห์ด้วยเทคนิคของดาต้าไมนิ่ง เพื่อให้ได้ผลลัพธ์ที่มีประสิทธิภาพน่าเชื่อถือ และจะมีการกล่าวถึงการตีความผลลัพธ์ที่ได้เพื่อนำไปวิเคราะห์และแก้ปัญหาได้อย่างมีประสิทธิภาพมากที่สุด

Title Knowledge Discovery in an Inventory Management System  
Student Mr.Narongrit Ananchaipattana  
Advisor Asst.Prof.Dr. Pattarachai Lalitrojawanong  
Level of Study Master of Science in Information Technology  
Major Information Science  
Academic Year 2004



## ABSTRACT

In this Project, we present the process of system development with data mining technology. It also highlights the case study on inventory management system to provide the rationale of the strategy to reduce the total of inventory, while maintaining the same level of probability that a particular customer's demand will be satisfied, which the main objective of inventory management. Before adopting data mining techniques, preliminary the process of data preparation, integration and transformation by use tool of data warehouse and OLAP server to prepare data for process of analysis by technique of data mining for the efficiently and reliability. In additional we present the description of result to bring knowledge from data to solve the specific problem efficiently.

## กิตติกรรมประกาศ

ความสำเร็จของการพัฒนาโครงการพัฒนาระบบงานฉบับนี้ สำเร็จขึ้นได้จากความช่วยเหลือของบุคคลหลายๆท่าน ข้าพเจ้ามีความรู้สึกขอบคุณทุกท่านที่มีส่วนช่วยเหลือในด้านต่างๆ ด้วยความจริงใจ หากขาดบุคคลที่จะกล่าวถึงต่อไปนี้ไป ก็จะส่งผลกระทบต่อความสำเร็จของโครงการพัฒนาระบบงานฉบับนี้

ขอขอบคุณ บิดา มารดาผู้ให้กำเนิดผู้ให้กำเนิดซึ่งทุกอย่าง

ขอขอบคุณ อาจารย์ ดร.ภทรชัย ลลิตโรจน์วงศ์ อาจารย์ที่ปรึกษา ที่ให้ความรู้ แนะนำหนังสือ และให้คำปรึกษา ตลอดจนถึงแนะแนวทางการแก้ปัญหาโดยตลอด

สุดท้าย ขอขอบคุณ เพื่อนๆทุกคนที่คอยช่วยเหลือและเป็นกำลังใจให้เสมอมา

ณรงค์ฤทธิ์ อนันต์ชัยพัชฌา

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ	IV
บทที่	
1. บทนำ	1
1.1 ความเป็นมาของปัญหา	1
1.2 วัตถุประสงค์	1
1.3 ขอบเขตของการศึกษา	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	2
2. ทฤษฎีที่เกี่ยวข้อง	3
2.1 คาด้าไมน์นิ่ง	3
2.2 กระบวนการทำงานของคาด้าไมน์นิ่ง	3
2.3 การจำแนกประเภทข้อมูล (Data Classification)	4
2.4 Naive Bayes Data Mining Classifier	4
2.5 การพยากรณ์ข้อมูล	7
2.6 การวิเคราะห์อนุกรมเวลา	7
3. การวิเคราะห์ออกแบบระบบ	13
3.1 การสร้างแบบจำลองด้วยเทคนิคคาด้าไมน์นิ่ง	13
3.2 การแสดงผลลัพธ์	15
3.3 รายละเอียดการทำงานของโปรแกรม	15
3.4 ขั้นตอนการทำงานของระบบแสดงด้วย Activity Diagram	17
3.5 สภาพแวดล้อมของการพัฒนาระบบ	19

## สารบัญ(ต่อ)

	หน้า
3.6 การพัฒนาโปรแกรม	19
4. กรณีศึกษาการค้นพบความรู้จากระบบบริหารสินค้าคงคลัง	20
4.1 ลักษณะการดำเนินธุรกิจ	20
4.2 วัตถุประสงค์	20
4.3 การเตรียมข้อมูล	24
4.4 การคัดเลือกข้อมูล	24
4.5 การวิเคราะห์และออกแบบระบบงาน	29
5. การศึกษารูปแบบและการพยากรณ์	35
6. บทสรุป	44
6.1 สรุปหลักการที่ใช้ในระบบ	44
6.2 สรุปกระบวนการในการทำงาน	45
6.3 สรุปผลการทดสอบ	45
6.4 ข้อเสนอแนะและแนวทางในการพัฒนาต่อไป	45
บรรณานุกรม	46
ประวัติผู้เขียน	47

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาของปัญหา

ในปัจจุบันความก้าวหน้าทางเทคโนโลยีสารสนเทศได้รับการพัฒนาขึ้นมากทำให้มีข้อมูลจำนวนมากถูกจัดเก็บไว้ในฐานข้อมูลขนาดใหญ่โดยไม่ถูกนำมาใช้ประโยชน์อย่างเต็มที่ด้วยวิวัฒนาการของเทคโนโลยีในการพัฒนาการจัดเก็บข้อมูลได้ถูกพัฒนาเรื่อย มาทำให้เกิดเทคนิคดาต้าไมน์นิ่งหรือการสืบค้นความรู้ที่เป็นประโยชน์และน่าสนใจบนฐานข้อมูลขนาดใหญ่ (Knowledge Discovery from very large Database: KDD) เป็นสาขาหนึ่งในวิทยาศาสตร์คอมพิวเตอร์ที่กำลังได้รับความสนใจอย่างสูงในปัจจุบัน เมื่อใช้เทคนิคดาต้าไมน์นิ่ง ข้อมูลขนาดใหญ่จะถูกวิเคราะห์และสืบค้นความรู้หรือสิ่งที่สำคัญออกมา จากนั้นจะรวบรวมความรู้ที่ได้ให้อยู่ในฐานความรู้ (Knowledge Base) เพื่อนำไปใช้ประโยชน์ต่อไป โดยในปัจจุบันได้มีการนำเทคนิคดาต้าไมน์นิ่งไปประยุกต์ใช้ในงานด้านต่างๆมากขึ้น ทั้งในด้านการส่งเสริมการขายสินค้าในห้างสรรพสินค้า ด้านการวิเคราะห์เครดิตลูกค้าในธนาคาร และด้านอื่นๆ อีกมาก โดยในบทความนี้ได้นำเสนอเทคนิคต่างๆที่สำคัญของดาต้าไมน์นิ่งมาประยุกต์ใช้ในการสืบค้นสิ่งที่น่าสนใจออกมาจากข้อมูลสินค้าคงคลัง

การพัฒนาโปรแกรมแบบจำลองโดยอาศัยเทคนิคดาต้าไมน์นิ่งที่ได้นำเทคนิคที่น่าสนใจ

2 ประการมาประยุกต์ใช้คือ การจำแนกข้อมูล (Data Classification) และ การพยากรณ์ข้อมูล (Data Prediction) เพื่อเพิ่มความสะดวกในการนำมาประยุกต์ใช้งานสำหรับการช่วยในการบริหารและพยากรณ์ข้อมูลทางธุรกิจ น่าจะเป็นประโยชน์ในการตัดสินใจและเพื่อเป็นการหาทิศทางของเป้าหมายทางธุรกิจได้

### 1.2 วัตถุประสงค์ของโครงการ

การศึกษาโครงการนี้มีวัตถุประสงค์เพื่อ

1. เพื่อศึกษาการทำดาต้าไมน์นิ่ง ศึกษาเทคนิคที่น่าสนใจ 2 ประการคือ การจำแนกข้อมูล และ การพยากรณ์ข้อมูล
2. พัฒนาโปรแกรมแบบจำลอง ตามรูปแบบของเทคนิคทั้ง 2 ประการที่กล่าวข้างต้น
3. เพื่อประยุกต์ใช้งาน ในการช่วยงานบริหารและพยากรณ์ทางธุรกิจ โดยในโครงการนี้จะประยุกต์ใช้ในการบริหารจัดการและพยากรณ์ปริมาณสินค้าคงคลังของธุรกิจสินค้าเกษตรแปรรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 1.3 ขอบเขตของการศึกษา

การศึกษานี้กำหนดขอบเขตในการศึกษาเป็นการพัฒนาระบบเพื่อช่วยในการบริหารปริมาณสินค้าคงคลัง โดยอาศัยเทคนิคของคาด้าไมน์นิ่งทั้ง 2 ประการที่กล่าวไว้ข้างต้นเพื่อเปรียบเทียบระหว่างผลลัพธ์ที่ได้กับกับค่าความเป็นจริงของข้อมูลที่ใช้ในการทดสอบ โดยข้อมูลที่น่ามาใช้ในการทดสอบกับแบบจำลองนี้เป็นข้อมูลย้อนหลังในอดีตเกี่ยวกับข้อมูลปริมาณสินค้าคงคลังของธุรกิจผลิตสินค้าเกษตรแปรรูป โดยมีขอบเขตหลักๆ ที่จะศึกษาดังต่อไปนี้

- การสืบค้นความรู้จากฐานข้อมูลขนาดใหญ่
- การจำแนกข้อมูล เพื่อจัดกลุ่มปริมาณสินค้าสำเร็จรูป โดยต้นไม้การตัดสินใจ (Decision Tree)
- การพยากรณ์ข้อมูล เพื่อใช้ในการพยากรณ์ปริมาณในการสั่งซื้อของวัตถุดิบในช่วงเวลาถัดไป โดยการวิเคราะห์อนุกรมเวลา (Time Series Analysis)

### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

ผลจากการศึกษาจะทำให้ทราบกระบวนการการทำงานคาด้าไมน์นิ่งมีหลักการทำงานอย่างไร และแต่ละเทคนิคให้ผลลัพธ์ที่แตกต่างกันตามความต้องการในการค้นพบความรู้ในรูปแบบที่แตกต่างกัน เรียนรู้วิธีการในการพัฒนาระบบการค้นพบความรู้จากฐานข้อมูลด้วยเทคนิคต่างๆ วิเคราะห์ว่าระบบที่พัฒนาขึ้นสามารถใช้วิเคราะห์ได้ดีเพียงไร โดยการเปรียบเทียบการการทำงานจริง และมีความเหมาะสมที่จะนำไปใช้เป็นเครื่องมือในทางธุรกิจหรือไม่อย่างไร

## บทที่ 2

### ทฤษฎีที่เกี่ยวข้อง

#### 2.1 ดาต้าไมน์นิ่ง

ดาต้าไมน์นิ่ง คือ กระบวนการในการวิเคราะห์ข้อมูล เพื่อนำมาใช้ในระบบสนับสนุนการตัดสินใจ ดาต้าไมน์นิ่งเป็นขั้นตอนในการดึงเอาสารสนเทศมาจากข้อมูลทำให้ได้ความรู้ใหม่ๆที่ไม่เคยรู้มาก่อน หรือข้อมูลที่มีความถูกต้อง เป็นข้อมูลที่สามารถนำไปใช้ประโยชน์ในทางปฏิบัติได้จริง ดาต้าไมน์นิ่งเป็นส่วนหนึ่งของกระบวนการค้นพบความรู้ในฐานข้อมูล (Knowledge Discovery in Database : KDD) การนำแนวโน้มของข้อมูลและสารสนเทศที่ซ่อนอยู่ในฐานข้อมูลขนาดใหญ่ออกมาเป็นสิ่งสำคัญ เพราะถ้าไม่รู้จักใช้ประโยชน์จากข้อมูลเหล่านั้นก็จะสูญเปล่า ดาต้าไมน์นิ่งเป็นเครื่องมือที่สามารถค้นหาข้อมูลในฐานข้อมูลขนาดใหญ่หรือข้อมูลที่เป็นประโยชน์ที่อาจซ่อนเร้นอยู่ในฐานข้อมูล ซึ่งเป็นการเพิ่มคุณค่าให้กับฐานข้อมูลที่มีอยู่ (Berry and Gordon, 2000.)

#### 2.2 กระบวนการทำงานของดาต้าไมน์นิ่ง

การสืบค้นความรู้ที่เป็นประโยชน์และน่าสนใจบนฐานข้อมูลขนาดใหญ่ หรือที่เรียกกันว่า ดาต้าไมน์นิ่ง เป็นเทคนิคที่ใช้ในการจัดการกับฐานข้อมูลขนาดใหญ่ โดยจะนำข้อมูลที่มีอยู่มาวิเคราะห์แล้วดึงความรู้หรือสิ่งสำคัญออกมา เพื่อใช้ในการวิเคราะห์ หรือทำนายสิ่งต่างๆที่อาจจะเกิดขึ้น และกระบวนการ KDD ประกอบด้วย 3 ส่วนหลัก คือ

1) การประมวลผลก่อน (Pre-processing) คือขั้นตอนในการเตรียมข้อมูลให้เหมาะสมและอยู่ในรูปแบบที่สามารถนำไปใช้งานได้

2) การทำดาต้าไมน์นิ่ง คือขั้นตอนเลือกเทคนิคที่เหมาะสมสำหรับงานที่ต้องการ โดยสามารถนำเทคนิคมากกว่า 1 เทคนิคมาประมวลผล เพื่อดึงความรู้หรือสิ่งที่น่าสนใจจากข้อมูลที่ผ่านขั้นตอนการประมวลผลก่อน แล้วโดยผลลัพธ์ที่ได้จากขั้นตอนนี้คือฐานความรู้

3) การประมวลผลหลัง (Post-processing) คือขั้นตอนการนำฐานความรู้ที่ได้จากขั้นตอนดาต้าไมน์นิ่งมาทดสอบและพิจารณาว่าถูกต้องตามความต้องการหรือไม่ ซึ่งบางครั้งอาจต้องปรับแก้ค่าและนำเข้าสู่ขั้นตอนดาต้าไมน์นิ่งใหม่อีกครั้ง จนกว่าจะได้ความรู้หรือสิ่งที่น่าสนใจตามที่ต้องการออกมา

จะเห็นได้ว่า คาด้า ไม่นิ่งเป็นองค์ประกอบหลักที่สำคัญในกระบวนการ KDD ในลำดับต่อไป ซึ่งจะกล่าวถึงเทคนิคคาด้า ไม่นิ่งที่สำคัญ 2 เทคนิค ที่ได้นำมาประยุกต์ใช้ คือ การจำแนกประเภทข้อมูล และการพยากรณ์ข้อมูล

### 2.3 การจำแนกประเภทข้อมูล

การจำแนกประเภทข้อมูล (Data Classification) เป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ โดยจะนำข้อมูลส่วนหนึ่งมาสอนให้ระบบเรียนรู้ (Training Data) เพื่อจำแนกข้อมูลออกเป็นกลุ่มตามที่กำหนดไว้ ผลลัพธ์ที่ได้จากการเรียนรู้คือ โมเดลจำแนกประเภทข้อมูล (Classifier Model) และจะนำข้อมูลส่วนที่เหลือจากข้อมูลสอนระบบเป็นข้อมูลที่ใช้ทดสอบ (Testing Data) ซึ่งกลุ่มที่แท้จริงของข้อมูลที่ใช้ทดสอบนี้จะถูกนำมาเปรียบเทียบกับกลุ่มที่หามาได้จากโมเดลเพื่อทดสอบความถูกต้องและปรับปรุงโมเดลจนกว่าจะได้ค่าความถูกต้องในระดับที่น่าพอใจ หลังจากนั้น เมื่อมีข้อมูลใหม่เข้ามา เราจะนำข้อมูลมาผ่านโมเดล โดยโมเดลจะสามารถทำนายกลุ่มของข้อมูลนี้ได้

ต้นไม้การตัดสินใจ เป็นวิธีหนึ่งที่สำคัญในการจำแนกข้อมูล โดยต้นไม้การตัดสินใจมีลักษณะคล้ายโครงสร้างต้นไม้ที่แต่ละโหนดแสดงคุณลักษณะ (แอตทริบิวต์) แต่ละกิ่งแสดงผลในการทดสอบ และลิฟโหนดแสดงกลุ่มที่กำหนดไว้ ซึ่งต้นไม้ช่วยตัดสินใจนี้ง่ายต่อการปรับเปลี่ยนเป็นกฎการจำแนกประเภทข้อมูล (Classification Rule)

### 2.4 Naive Bayes Data Mining Classifier

อัลกอริทึม Naive Bayes ใช้ในการประมาณการความน่าจะเป็นที่ใช้ในการตัดสินใจว่าสิ่งที่ต้องการศึกษามีความเป็นไปได้หรือไม่ โดยสามารถอธิบายด้วยตัวอย่างเพื่อความเข้าใจดังนี้ (Tagbo, 2004.)

ตัวอย่างการประยุกต์ใช้งานอัลกอริทึม Naive Bayes

พิจารณาความน่าจะเป็นของลูกค้าที่จะซื้อสินค้าผ่านทางร้านค้าออนไลน์ โดยมีขั้นตอนดังนี้

1. นำข้อมูลของร้านค้าออนไลน์ ซึ่งเก็บข้อมูลการตัดสินใจว่าจะซื้อหรือไม่ซื้อสินค้าตัวใหม่ที่ทำการจัดรายการ
2. ลูกค้าแต่ละคนที่ได้รับอีเมลเกี่ยวกับสินค้าของบริษัท จะถูกเก็บข้อมูลเกี่ยวกับคุณลักษณะต่างๆลงในฐานข้อมูลดังนี้

- รายได้ของลูกค้า (ต่ำ ปานกลาง สูง)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ความเร็วในการติดต่อสื่อสารผ่านระบบอินเทอร์เน็ต (เร็ว ช้า)
- ระดับการศึกษา (มัธยม วิทยาลัย)

ข้อมูลที่ลูกค้าที่ได้รับทางอีเมลว่าจะทำการตัดสินใจซื้อหรือไม่จะถูกเก็บในฐานะข้อมูลด้วยเช่นกัน

3. ตัวอย่างของข้อมูลด้านล่างประกอบด้วย 12 ลูกค้าที่มีการตัดสินใจที่จะซื้อหรือไม่ซื้อสินค้าจากการที่ได้รับ อีเมลจตุรกายการพิเศษจากทางบริษัท

ตารางที่ 2.5 ตัวอย่างข้อมูลลูกค้า

	CUSTOMER INCOME	CUSTOMER USES HIGH SPEED CONNECTION	EDUCATION LEVEL	BUY DECISION
1	LOW	NO	HIGH SCHOOL	NO-BUY
2	LOW	YES	HIGH SCHOOL	NO-BUY
3	LOW	NO	COLLEGE	NO-BUY
4	LOW	YES	COLLEGE	BUY
5	MEDIUM	NO	HIGH SCHOOL	NO-BUY
6	MEDIUM	YES	HIGH SCHOOL	NO-BUY
7	MEDIUM	NO	COLLEGE	BUY
8	MEDIUM	YES	COLLEGE	BUY
9	HIGH	NO	HIGH SCHOOL	NO-BUY
10	HIGH	YES	HIGH SCHOOL	BUY
11	HIGH	NO	COLLEGE	BUY
12	HIGH	YES	COLLEGE	BUY

4. ผลการคำนวณ

ตารางที่ 2.6 ตัวอย่างผลการคำนวณ

Attribute	ATTRIBUTE VALUE	BUY DECISION	NO-BUY DECISION	BUY PROBABILITY	NO-BUY PROBABILITY
CUSTOMER INCOME	LOW	1	3	1/6	3/6
CUSTOMER INCOME	MEDIUM	2	2	2/6	2/6
CUSTOMER INCOME	HIGH	3	1	3/6	1/6
<b>TOTAL</b>		<b>6</b>	<b>6</b>	<b>1.0</b>	<b>1.0</b>
CUSTOMER USES HIGH SPEED CONNECTION	YES	4	2	4/6	2/6
CUSTOMER USES HIGH SPEED CONNECTION	NO	2	4	2/6	4/6
<b>TOTAL</b>		<b>6</b>	<b>6</b>	<b>1.0</b>	<b>1.0</b>
EDUCATION LEVEL	HIGH SCHOOL	1	5	1/6	5/6
EDUCATION LEVEL	COLLEGE	5	1	5/6	1/6
<b>TOTAL</b>		<b>6</b>	<b>6</b>	<b>1.0</b>	<b>1.0</b>
<b>PREDICTION</b>		<b>6</b>	<b>6</b>	<b>6/12</b>	<b>6/12</b>

ขั้นตอนการคำนวณของอัลกอริทึมมีดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยกตัวอย่างจากรายได้ของลูกค้าในการอธิบายวิธีการคำนวณ

- รายได้ของลูกค้าจัดกลุ่มอยู่ในระดับ ต่ำ ปานกลาง สูง
- จากข้อมูลตัวอย่างทั้ง 12 เรคคอร์ด

### การหาอัตราส่วนการตัดสินใจซื้อ

พิจารณาจากรายได้ของลูกค้า จำนวนทั้งหมดของการตัดสินใจซื้อสินค้า เท่ากับ 6 โดยมีอัตราส่วนดังนี้

รายได้สูง มีอัตราส่วน 1/6

รายได้ปานกลาง มีอัตราส่วน 2/6

รายได้ต่ำ มีอัตราส่วน 3/6

### การหาอัตราส่วนการตัดสินใจไม่ซื้อ

พิจารณาจากรายได้ของลูกค้า จำนวนทั้งหมดของการตัดสินใจไม่ซื้อสินค้า เท่ากับ 6 โดยมีอัตราส่วนดังนี้

รายได้สูง มีอัตราส่วน 3/6

รายได้ปานกลาง มีอัตราส่วน 2/6

รายได้ต่ำ มีอัตราส่วน 1/6

ทำการพยากรณ์การตัดสินใจซื้อสินค้าของลูกค้า โดยยกตัวอย่างจาก ลูกค้าที่มีรายได้ประจำปี 15,000 เหรียญ ใช้โมเด็มที่มีความเร็ว 24 กิโลไบต์ และจบการศึกษาจากวิทยาลัยการบริหาร

จากตัวอย่างข้อมูลของลูกค้าสามารถจำแนกได้ดังนี้

- รายได้จัดอยู่ในระดับที่ต่ำ
- ใช้อินเทอร์เน็ตความเร็วต่ำ
- การศึกษาระดับวิทยาลัย

ต้องการที่จะพยากรณ์ว่าลูกค้ารายนี้มีความน่าจะเป็นที่จะซื้อสินค้าหรือไม่ โดยพิจารณาจากความน่าจะเป็นในการตัดสินใจซื้อหรือไม่ซื้อแล้วนำมาเปรียบเทียบกัน

### โดยใช้สูตร

$$\begin{aligned} \text{ความน่าจะเป็นในการตัดสินใจซื้อสินค้า} &= (\text{ความน่าจะเป็นที่จะซื้อสินค้าของลูกค้ารายได้ต่ำ}) \\ & * (\text{ความน่าจะเป็นที่จะซื้อสินค้าของลูกค้าที่ใช้อินเทอร์เน็ตความเร็วต่ำ}) * \\ & (\text{ความน่าจะเป็นที่จะซื้อสินค้าของลูกค้าที่มีการศึกษาระดับวิทยาลัย}) \end{aligned}$$

$$\text{ความน่าจะเป็นในการตัดสินใจซื้อสินค้า} = 1/6 * 2/6 * 5/6$$

$$\text{ความน่าจะเป็นในการตัดสินใจซื้อสินค้า} = 0.0463$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความน่าจะเป็นในการตัดสินใจไม่ซื้อสินค้า = (ความน่าจะเป็นที่จะซื้อสินค้าของลูกค้ารายได้ต่ำ)

\* (ความน่าจะเป็นที่จะซื้อสินค้าของลูกค้าใช้ความเร็ว internet ต่ำ) \*

(ความน่าจะเป็นที่จะซื้อสินค้าของลูกค้าที่มีการศึกษาระดับวิทยาลัย)

$$\text{ความน่าจะเป็นในการตัดสินใจไม่ซื้อสินค้า} = 3/6 * 4/6 * 1/6$$

$$\text{ความน่าจะเป็นในการตัดสินใจไม่ซื้อสินค้า} = 0.0556$$

จากผลลัพธ์ที่ได้สรุปได้ว่าลูกค้ามีความน่าจะเป็นที่จะตัดสินใจไม่ซื้อสินค้า

จากค่าที่ได้สามารถหาความน่าจะเป็นได้ดังนี้

$$\text{ความน่าจะเป็นในการตัดสินใจซื้อสินค้า} = 0.0463 / (0.0463 + 0.0556) = 45.44\%$$

$$\text{ความน่าจะเป็นในการตัดสินใจไม่ซื้อสินค้า} = 0.0556 / (0.0463 + 0.0556) = 54.56\%$$

## 2.5 การพยากรณ์ข้อมูล

การพยากรณ์ข้อมูลเป็นกระบวนการสร้าง โมเดลเพื่อทำนายค่าที่ต้องการจากข้อมูลที่มีอยู่ โดยมีกระบวนการสร้าง โมเดลคล้ายกับการจำแนกประเภทข้อมูลดังที่ได้กล่าวมาข้างต้น ต่างกันตรงที่ การพยากรณ์ข้อมูล ไม่มีการจัดข้อมูลเข้ากลุ่มตามที่กำหนด แต่การพยากรณ์ข้อมูลนี้เป็นการพยากรณ์ค่าที่ต้องการออกมาเป็นตัวเลข ตัวอย่างเช่น หายอดขายของเดือนถัดไป จากข้อมูลการขายทั้งหมดที่ผ่านมา หรือทำนายปริมาณการสั่งซื้อวัตถุดิบในไตรมาสต่อไป จากข้อมูลการสั่งซื้อทั้งหมด โดยจะศึกษาถึงเทคนิคที่ใช้เพื่อการพยากรณ์แบบอนุกรมเวลา (Time series) โดยเทคนิคการพยากรณ์ทางด้านสถิติ

## 2.6 การวิเคราะห์อนุกรมเวลา

ข้อมูลอนุกรมเวลาคือข้อมูลที่เกิดขึ้นในเวลาที่มีระยะห่างเท่าๆกันและต่อเนื่องกัน ข้อมูลอนุกรมเวลาจึงเป็นค่าที่แสดงการเปลี่ยนแปลงของตัวแปรใดตัวแปรหนึ่งที่เปลี่ยนไปตามเวลา

ปัญหาของอนุกรมเวลาเป็นการคาดหวั่งค่าที่ได้มาจากความสัมพันธ์ของค่าก่อนหน้า การพยากรณ์อนุกรมเวลาเป็นหนึ่งในเทคนิคการพยากรณ์เชิงปริมาณ ใช้ข้อมูลเชิงปริมาณ โดยสร้างตัวแปรแบบทางคณิตศาสตร์โดยใช้การพยากรณ์อนุกรมเวลา เป็นเทคนิคที่ใช้เฉพาะข้อมูลในอดีตของตัวแปรที่ต้องการพยากรณ์ เพื่อพยากรณ์ค่าของตัวแปรนั้นในอนาคต (สมจิต วัฒนาชยากุล, 2532.)

### 2.6.1 วัตถุประสงค์ของการวิเคราะห์อนุกรมเวลา

1. เพื่อศึกษาถึงลักษณะหรือสภาพของตัวแปร เช่น ศึกษาถึงระดับปริมาณของวัตถุดิบที่คงเหลือภายในคลังสินค้าในแต่ละเดือนของบริษัท หรือการศึกษาถึงระดับปริมาณของสินค้าคงเหลือภายในคลังสินค้าของช่วงเวลาแต่ละปี ซึ่งสามารถช่วยประมาณการการสั่งซื้อสินค้าและ การวางแผนการผลิตภายในช่วงเวลาที่ต้องการพยากรณ์
2. เพื่อพยากรณ์ค่าของตัวแปร ในอนาคต ทำให้สามารถวางแผนการดำเนินงานได้อย่างมีประสิทธิภาพ

### 2.6.2 ส่วนประกอบของอนุกรมเวลา

ข้อมูลของอนุกรมเวลาแบ่งออกเป็น 4 ส่วนดังนี้

1. แนวโน้ม หรือแนวโน้มระยะยาว เป็นข้อมูลที่มีการเปลี่ยนแปลงในระยะเวลาที่นานพอที่จะเห็นแนวโน้มของข้อมูลว่าในอนาคตจะเพิ่มขึ้นหรือลดลง
2. ความผันแปรตามวัฏจักร เป็นการเคลื่อนไหวของข้อมูลที่มีลักษณะขึ้นๆลงๆ มีลักษณะคล้ายกับการความแปรผันตามฤดูกาล แต่มีช่วงเวลาที่ยาวกว่าในการครบวงจร
3. ความผันแปรตามฤดูกาล การเปลี่ยนแปลงของข้อมูลเกิดขึ้นเนื่องจากอิทธิพลของฤดูกาล ซึ่งเกิดขึ้นซ้ำๆ ในช่วงเวลาเดียวกันของแต่ละปี
4. ความผันแปรที่ไม่แน่นอน

### 2.6.3 เทคนิคการวิเคราะห์อนุกรมเวลา

วัตถุประสงค์ในการศึกษาหรือวิเคราะห์เกี่ยวกับอนุกรมเวลานั้น ก็เพื่อประโยชน์ในการกำหนดนโยบายและดูทางแก้ปัญหาต่างๆ ซึ่งในการกำหนดนโยบายนั้นอาจเป็นการกำหนดนโยบายระยะสั้นเช่นเพียงช่วงเวลา 1 ปีหรือน้อยกว่านั้น การพยากรณ์ระยะยาวออกไปถึงอนาคตถือว่าเป็นสิ่งจำเป็นในกรณีที่จะเป็นการเตรียมพร้อมสำหรับฝ่ายโรงงาน ฝ่ายขาย ฝ่ายการคลังสินค้า และฝ่ายอื่นๆ ภายในองค์กร

ข้อมูลแต่ละค่าในอนุกรมเวลาจะประกอบด้วยส่วนประกอบต่างๆ ซึ่งอาจมีทั้ง 4 ชนิด หรือมีเพียงบางชนิดก็ได้ สำหรับโครงงานนี้สนใจที่ 2 ส่วนประกอบหลักๆ ดังนี้

1. การแยกค่าทางแนวโน้มระยะยาวออกจากอนุกรมเวลา ซึ่งวิธีการในการที่จะแยกเอาทางโน้มออกจากอนุกรมเวลา มีหลายแบบ แบบที่สนใจก็คือ Exponential Smoothing ซึ่งจะพิจารณาต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. การคำนวณหาความเคลื่อนไหวอันเนื่องมาจากฤดูกาล ซึ่งมีวิธีคำนวณหลายแบบ เช่นเดียวกัน โดยทั่วไปการวัดความเคลื่อนไหวอันเนื่องมาจากฤดูกาลมักวัดออกมาในรูปของดัชนี ซึ่งเรียกว่าดัชนีฤดูกาล(Season Index) เมื่อเราทราบค่าของดัชนีฤดูกาล เราอาจนำค่าของดัชนีฤดูกาลที่ได้ไปปรับให้ความเคลื่อนไหวเนื่องจากฤดูกาลหมดไปจากตัวเลขเดิม โดยการหารตัวเลขเดิมในอนุกรมเวลา ด้วยดัชนีฤดูกาลของช่วงเวลานั้นจะได้ข้อมูลที่ขจัดอิทธิพลของฤดูกาลออกไป เพื่อจะได้ทำให้การพิจารณาและแปลความหมายของทางโน้มทำได้ง่ายมากขึ้น

ในการวิเคราะห์การเปลี่ยนแปลงตามฤดูกาล มักจะวัดการเปลี่ยนแปลงออกมาในรูปของดัชนีฤดูกาล ซึ่งวิธีที่ใช้ในการคำนวณดัชนีฤดูกาลมีอยู่หลายแบบด้วยกัน โดยการทำโครงการนี้เลือกใช้วิธีหาอัตราส่วนเทียบกับทางโน้ม (The Ratio to Trend Method)

### 2.6.3.1 การทำให้เรียบด้วยเทคนิคเอกซ์โพเนนเชียล (Exponential Smoothing)

เป็นเทคนิคที่ให้ความสำคัญแก่ข้อมูลล่าสุดมากที่สุด และความสำคัญจะลดลงสำหรับข้อมูลที่มีระยะเวลาห่างออกไป สำหรับเทคนิคการทำให้เรียบด้วยเทคนิคเอกซ์โพเนนเชียล จะแยกเป็น 3 เทคนิคโดยแต่ละเทคนิคยังคงยึดเกณฑ์ที่ว่า จะให้ความสำคัญแก่ข้อมูลล่าสุดมากที่สุด และความสำคัญจะค่อยๆ ลดลงตามเวลาที่ห่างออกไป แต่ละเทคนิคเหมาะสมกับรูปแบบข้อมูลคนละชนิด เทคนิคที่จะกล่าวถึงมี

1. Simple (Simple Exponential Smoothing)
2. Holt (Double Exponential Smoothing: Holt 's Two-Parameter Method)
3. Winter (Triple Exponential Smoothing : Winter 's Three-Parameter Trend and Seasonality Method)

โดยจะยกตัวอย่างวิธีการทำงานแบบง่าย (Simple Exponential Smoothing หรือ Single Exponential Smoothing) เทคนิคการทำให้เรียบอย่างง่าย(Simple Exponential Smoothing) เป็นเทคนิคที่ใช้ในการหาค่าเฉลี่ยแบบถ่วงน้ำหนัก โดยให้น้ำหนักหรือความสำคัญของข้อมูลคือ  $\alpha$  (alpha) หรือเรียกกันว่า ค่าคงที่ของการทำให้เรียบ (Smoothing Constant) เทคนิคนี้จะใช้เมื่อข้อมูลมีลักษณะการเคลื่อนไหวในระดับคงที่ หรือข้อมูลมีแนวโน้ม (Trend) และไม่มีอิทธิพลของฤดูกาล (Seasonality)

### รูปแบบของข้อมูล

$$F_{t+1} = \alpha X_t + (1-\alpha)F_t ; t = 1, 2, \dots, N$$

โดยที่

$$X_t = \text{ข้อมูล ณ. เวลา } t; t = 1, 2, 3, \dots, N$$

$$\alpha = \text{น้ำหนักหรือความสำคัญที่ให้แก่อข้อมูล ณ. เวลา } t$$

$$F_{t+1} = \text{ค่าปริมาณหรือค่าพยากรณ์ของข้อมูล ณ. เวลา } t+1$$

สำหรับน้ำหนักหรือความสำคัญ  $\alpha$  (alpha) จะมีค่าระหว่าง 0 ถึง 1

#### 2.6.3.2 วิธีหาค่าส่วนเทียบกับทางโน้ม

การหาค่าดัชนีฤดูกาลโดยวิธีนี้ เราจะหาค่าของแนวโน้มโดยวิธีกำลังสองน้อยที่สุดออกจากข้อมูลเสียก่อน จากนั้นจึงใช้การเฉลี่ยขจัดกาเปลี่ยนแปลงเนื่องจากวัฏจักรและเหตุการณ์ผิดปกติออกไป ดังนั้นในการหาค่าดัชนีฤดูกาลโดยวิธีหาค่าส่วนเทียบกับทางโน้ม เราต้องดำเนินการดังนี้

1. หาค่าเฉลี่ยต่อคาบของปีต่างๆ ในอนุกรม โดยถือว่าค่าดังกล่าวที่ได้เป็นค่าของทางโน้ม
2. นำค่าที่ได้จากข้อ 1 มาหาสมการทางโน้ม
3. หาค่าต่างๆ ที่ควรจะได้จากทางโน้มในแต่ละคาบ
4. เทียบค่าของข้อมูลเดิมในคาบต่างๆ ของอนุกรมว่าเป็นที่เปอร์เซนต์ของแนวโน้ม
5. นำค่าสัมพัทธ์ที่ได้จาก 4 มาหาค่าเฉลี่ยตามคาบ ซึ่งจะใช้ค่าเฉลี่ยเลขคณิตหรือมัธยฐานก็ได้
6. ตรวจสอบผลรวมของค่าเฉลี่ยว่าเท่ากับที่เปอร์เซนต์ของจำนวนคาบที่เก็บในแต่ละปีหรือไม่ ถ้าเท่ากันค่าเฉลี่ยที่ได้คือดัชนีฤดูกาล ถ้าไม่เท่ากันต้องปรับค่าที่ได้เสียก่อน ซึ่งค่าที่ได้หลังจากการปรับจะเป็นดัชนีฤดูกาลตามที่ต้องการ

สามารถอธิบายโดยตัวอย่างเพื่อความเข้าใจ ได้ดังนี้

ต้องการหาสมการทางโน้มแบบเส้นตรงของข้อมูลยอดขายสินค้าของห้างสรรพสินค้าแห่งหนึ่งระหว่างปี 2525 – 2532 โดยสร้างสมการทางโน้มได้จากข้อมูลในตารางที่ 2.7

ตารางที่ 2.7 ข้อมูลอนุกรมเวลา และค่าคำนวณต่างๆที่จะนำมาช่วยในการหา  
สมการทางโน้มแบบเส้นตรง

ปี	X	ค่าเฉลี่ยต่อเดือน	XY	X*X
2525	-7	273.7	-1915.9	49
2526	-5	293.5	-1467.5	25
2527	-3	315	-945	9
2528	-1	336.8	-336.8	1
2529	1	364.4	364.4	1
2530	3	394.8	1184.4	9
2531	5	424.2	2121	25
2532	7	458.7	3210.9	49
		2861.1	2215.5	168

จาก  $Y = a + bX$

เนื่องจาก  $\sum X = 0$

ดังนั้น  $b = \frac{\sum XY}{\sum X^2}$  และ  $a = \frac{\sum Y}{n}$

ดังนั้น  $b = 2215.5/168 = 13.188$

$a = 2861.1/8 = 357.6$

ได้สมการทางโน้มดังนี้  $Y = 357.6 + 13.188 X$

$X = 0$  ตรงกับวันที่ 1 มกราคม 2529

และหน่วยของ  $X = 6$  เดือน

### 2.6.3.3 การประยุกต์ใช้งานการวิเคราะห์อนุกรมเวลา

ตัวอย่างโจทย์ของการนำการวิเคราะห์อนุกรมเวลาไปใช้งาน ยกตัวอย่างเช่น

- ต้องการวิเคราะห์ปริมาณสินค้าคงเหลือของสินค้าสำเร็จรูปภายในคลังสินค้าของเดือน มกราคม ปี 2005 โดยใช้ข้อมูลปริมาณสินค้าคงเหลือของเดือน มกราคม – พฤศจิกายน ปี 2004
- ต้องการพยากรณ์ยอดขายของสินค้าชนิดหนึ่งประจำไตรมาสแรกของปี 2005 โดยมีข้อมูลยอดขายของสินค้าชนิดนั้นตั้งแต่ไตรมาสที่ 1 ปี 2000 ถึงไตรมาสที่ 4 ปี 2004 ผลลัพธ์ที่ได้จะแสดงในรูปแบบของกราฟเพื่อแสดงแนวโน้มของข้อมูลที่สนใจ ประกอบด้วยรายงานสรุปผลของผลลัพธ์ที่ต้องการในรูปแบบที่สามารถเข้าใจง่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 2.6.4 ขั้นตอนการสร้างแบบจำลองอนุกรมเวลา

แบบจำลองอนุกรมเวลาสามารถหาค่าพยากรณ์ในอนาคต จากการศึกษาข้อมูลของตัวแปรนั้นในอดีต เพื่อหารูปแบบของข้อมูลแล้วนำรูปแบบนั้นไปพยากรณ์ค่าในอนาคต ขั้นตอนการสร้างแบบจำลองมีดังนี้

1. กำหนดหรือเลือกรูปแบบที่เหมาะสมกับข้อมูลอนุกรมเวลาที่ต้องการศึกษา โดยใช้ความรู้ทางด้านสถิติ และประสบการณ์ในอดีต หรือทำการเปรียบเทียบกับข้อมูลอื่นที่มีลักษณะคล้ายกัน
2. ประเมินค่าพารามิเตอร์ของรูปแบบที่ได้จากขั้นที่ 1 โดยใช้หลักการทางสถิติมาทดสอบ ถ้าสรุปผลการทดสอบได้ว่าเหมาะสมก็สามารถนำแบบจำลองนั้นไปใช้ในการพยากรณ์ได้ แต่ถ้าสรุปได้ว่าไม่เหมาะสม ก็ต้องสร้างแบบจำลองใหม่ โดยต้องย้อนกลับไปทำขั้นตอนที่ 1 ใหม่

## บทที่ 3

### การวิเคราะห์ออกแบระบบ

จากวัตถุประสงค์ของการทำดาต้าไมนนิ่งหรือการค้นพบความรู้จากฐานข้อมูลที่มีขนาดใหญ่ และมีข้อมูลเป็นจำนวนมาก เพื่อต้องการที่จะพัฒนาเครื่องมือหรือเทคนิคที่มีประสิทธิภาพ โดยใช้ อุปกรณ์คอมพิวเตอร์ในการรับข้อมูลเป็นจำนวนมากเพื่อนำไปประมวลผลวิเคราะห์ข้อมูลจากแบบจำลองที่ได้มาจากการสร้างรูปแบบจากฐานข้อมูลขนาดใหญ่ ซึ่งในการสร้างแบบจำลองที่สามารถวิเคราะห์ข้อมูลได้อย่างมีประสิทธิภาพนั้นต้องอาศัยข้อมูลที่มีคุณภาพ สิ่งที่สำคัญอย่างยิ่งในการออกแบบระบบดาต้าไมนนิ่งคือกระบวนการในการเตรียมข้อมูลซึ่งมีขั้นตอนดังนี้

- Extraction การเข้าถึงฐานข้อมูลในหลายๆรูปแบบ
- Cleaning แก้ปัญหาความไม่สอดคล้องกันของข้อมูล
- Transformation แก้ไขปรับปรุงข้อมูลให้อยู่ในรูปแบบ และ ภาษาเดียวกัน
- Replication การทำสำเนาหรือนำเข้าข้อมูลจากฐานข้อมูล
- Analyzing การตรวจสอบความถูกต้องของข้อมูล เช่น ชนิดของข้อมูล, ความสัมพันธ์ของข้อมูล
- Checking for data quality การตรวจสอบความถูกต้องและสมบูรณ์ของข้อมูล
- Update metadata ทำการปรับปรุงที่เก็บรายละเอียดของฐานข้อมูลถ้าจำเป็น

#### 3.1 การสร้างแบบจำลองด้วยเทคนิคดาต้าไมนนิ่ง

ในส่วนนี้เป็นการนำเอาข้อมูลที่ได้ผ่านการจัดเตรียมและจัดกลุ่มเรียบร้อยแล้วผ่านกระบวนการดาต้าไมนนิ่งซึ่งประกอบด้วยเทคนิค 2 ประการดังนี้

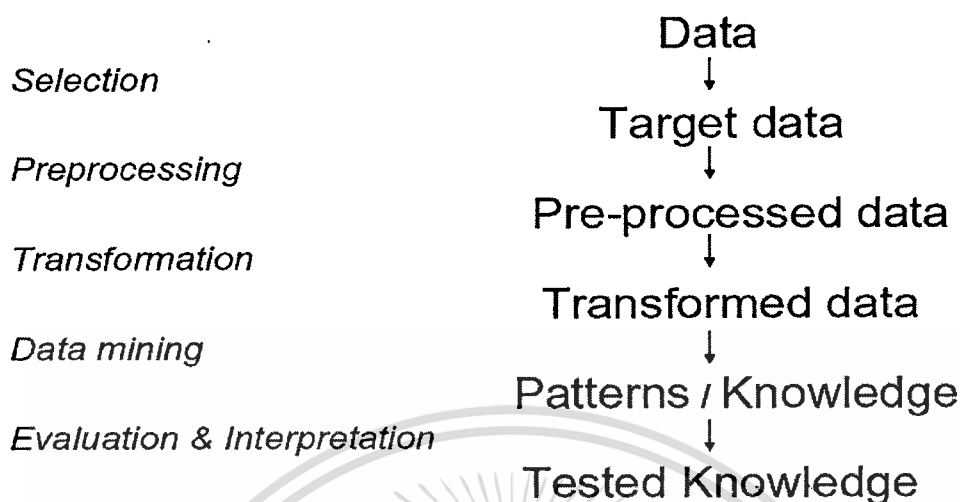
1. การจำแนกประเภทข้อมูลโดยใช้ Naive Bayes Data Mining Classifier ในการจัดกลุ่มข้อมูล โดยจะนำข้อมูลส่วนหนึ่งมาสอนให้ระบบเรียนรู้ เพื่อจำแนกข้อมูลออกเป็นกลุ่มตามที่กำหนดไว้ ผลลัพธ์ที่ได้จากการเรียนรู้คือ โมเดลจำแนกประเภทข้อมูล และจะนำข้อมูลส่วนที่เหลือจากข้อมูลสอนระบบเป็นข้อมูลที่ใช้ทดสอบ ซึ่งกลุ่มที่แท้จริงของข้อมูลที่ใช้ทดสอบนี้จะถูกนำมาเปรียบเทียบกับกลุ่ม

ที่นำมาได้จากโมเดลเพื่อทดสอบความถูกต้องและปรับปรุงโมเดล จนกว่าจะได้ค่าความถูกต้องในระดับที่น่าพอใจ

ข้อมูลที่นำมาทำการสำรวจมีวัตถุประสงค์เพื่อต้องการวิเคราะห์แบ่งกลุ่มของข้อมูลวัตถุดิบที่ทำการสั่งซื้อเข้ามาว่าแต่ละปัจจัยที่ใช้ในการตรวจวัดคุณภาพ มีปัจจัยใดบ้างที่มีผลกระทบต่อผลลัพธ์ของการผ่านกระบวนการตรวจวัดคุณภาพของสินค้า ซึ่งการแบ่งกลุ่มด้วยวิธีของ Naïve Data Mining Classifier นี้จะสามารถนำไปพยากรณ์ผลลัพธ์ของการตรวจวัดคุณภาพของวัตถุดิบแต่ละ Lot ที่จะทำการรับในอนาคตว่าจะผ่านหรือไม่โดยพิจารณาจากปัจจัยที่เกี่ยวข้อง อีกทั้งยังสามารถที่จะวิเคราะห์ได้ว่าปัจจัยใดมีผลกระทบต่อคุณภาพมากน้อยเพียงใด เพื่อนำข้อมูลที่ได้แจ้งให้กับทางผู้จัดจำหน่ายทราบว่าวัตถุดิบที่มีลักษณะอย่างไรที่เป็นที่ต้องการ และวัตถุดิบที่มีลักษณะอย่างไรที่ไม่เป็นที่ต้องการ

2. การพยากรณ์และหาแนวโน้มของข้อมูลโดยใช้วิธีการทางสถิติโดยเลือกใช้เทคนิคการวิเคราะห์อนุกรมเวลา โดยสามารถที่จะพยากรณ์ข้อมูลในเชิงปริมาณ ซึ่งข้อมูลที่ใช้ในการวิเคราะห์จะเป็นข้อมูลเชิงปริมาณ ยกตัวอย่างเช่น ยอดขายของสินค้าสำเร็จรูปประเภทต่างๆ ยอดในการสั่งซื้อวัตถุดิบ โดยพิจารณาเป็นช่วงเวลานำข้อมูลมาผ่านแบบจำลองเพื่อพยากรณ์หาค่าในอนาคต แบบจำลองที่นำมาใช้จะอาศัยหลักการอนุกรมเวลาเพื่อหาแนวโน้มของข้อมูลที่จัดอิทธิพลของฤดูกาลออก

สำหรับโครงการนี้จะทำการวิเคราะห์ข้อมูลปริมาณของสินค้าสำเร็จรูปที่มีอยู่ในคลังสินค้า โดยจะทำการวิเคราะห์ทาง ไน้มของข้อมูลที่ทำการจัดอิทธิพลของฤดูกาลออกไปแล้ว เพื่อทำการพยากรณ์ปริมาณของสินค้าคงคลังในช่วงเวลาต่างๆ ในอนาคต เพื่อที่จะนำข้อมูลที่ได้ไปใช้ในการปรับปรุงยอดปริมาณของสินค้าคงคลังเพื่อให้มีปริมาณเพียงพอต่อความต้องการของผู้บริโภค และป้องกันไม่ให้สินค้ามีมากเกินไปในคลังซึ่งอาจทำให้เกิดค่าใช้จ่ายในการจัดเก็บหรืออาจทำให้เกิดของเสียได้



รูปที่ 3.1 กระบวนการในการพัฒนาระบบค้ำไม้หนึ่ง

### 3.2 การแสดงผลลัพธ์

ในส่วนนี้เป็นส่วนการแสดงผลลัพธ์ที่ได้จากการใช้เทคนิคค้ำไม้หนึ่งทั้ง 2 ประการ โดยสามารถจำแนกได้ทั้งหมด 2 ส่วนดังนี้

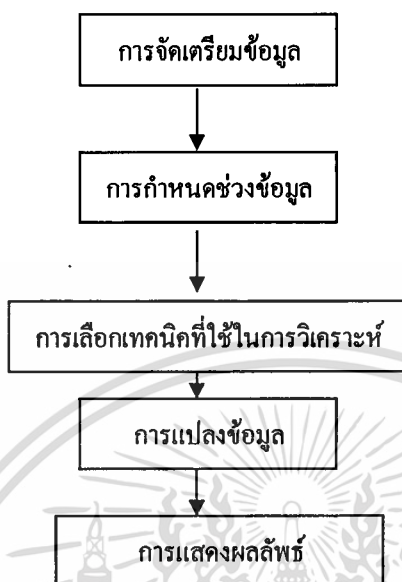
1. การแสดงผลการจำแนกประเภทข้อมูล จะแสดงผลอยู่ในลักษณะตารางความน่าจะเป็นของแต่ละปัจจัยที่มีผลต่อการตรวจวัดคุณภาพเพื่อให้ผู้ใช้สามารถวิเคราะห์ได้ว่าปัจจัยใดมีผลต่อการตรวจวัดคุณภาพมากที่สุดเพียงใด และมีหน้าจอสําหรับการพยากรณ์การรับเข้าของสินค้าแต่ละ Lot ว่าคุณสมบัติที่ระบุผ่านจะสามารถผ่านการตรวจสอบคุณภาพหรือไม่
2. การพยากรณ์และหาแนวโน้มของข้อมูล โดยใช้ Time series ผลลัพธ์ที่ได้จะแสดงในรูปแบบของกราฟเพื่อแสดงให้เห็นถึงแนวโน้มของข้อมูล และแสดงเป็นตารางเพื่อแสดงค่าที่ใช้ในการพยากรณ์ในรูปแบบที่สามารถเข้าใจง่าย และมีหน้าจอสําหรับการพยากรณ์ปริมาณของสินค้าแต่ละชนิดในช่วงเวลาต่างๆ

### 3.3 รายละเอียดการทำงานของโปรแกรม

ระบบงานแบ่งตามหน้าที่การวิเคราะห์ข้อมูลออกเป็น 2 ส่วนหลักๆดังนี้

- การจำแนกประเภทข้อมูล
- การพยากรณ์ข้อมูล

โดยแต่ละส่วนประกอบด้วยลำดับการทำงานดังนี้



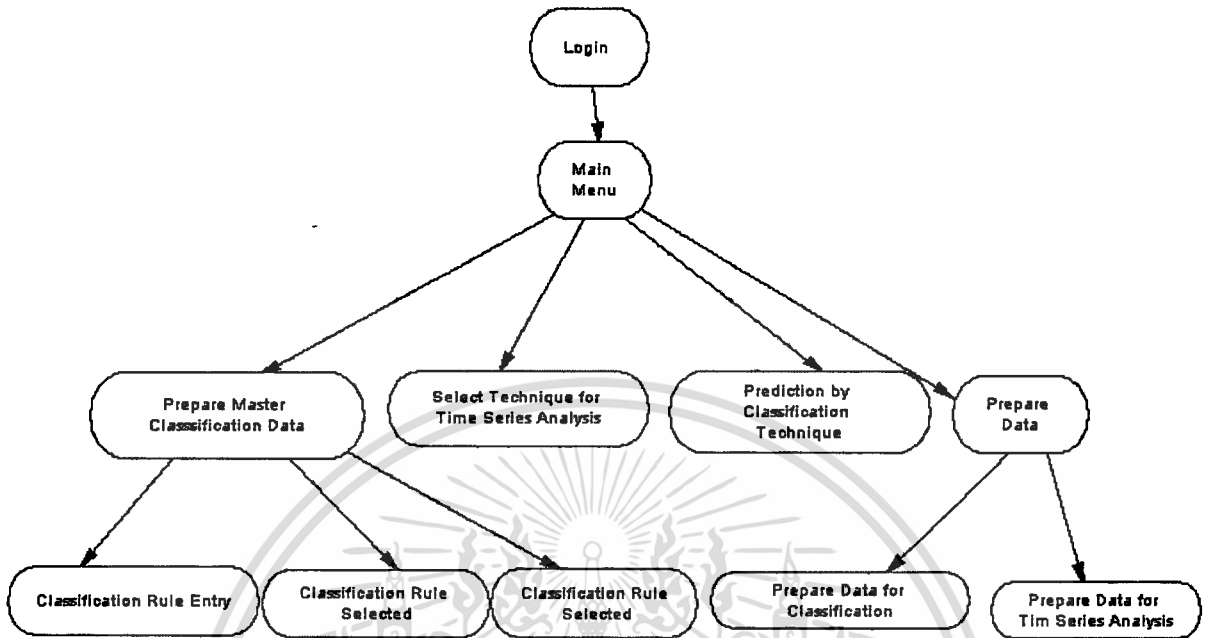
รูปที่ 3.2 ขั้นตอนการทำงานของโปรแกรม

งานหลักๆของระบบการค้นพบความรู้จากฐานข้อมูลสามารถแบ่งได้เป็นส่วนต่างๆดังนี้

- 1) การนำเข้าข้อมูลจากแหล่งข้อมูลที่เตรียม (Data Input)
- 2) การคำนวณผลทางสถิติหรือ การประมวลผลอัลกอริทึม (Processing)
- 3) การแสดงผลข้อมูล (Output View)
- 4) ส่วนติดต่อกับผู้ใช้ระบบ (User Interface)

นอกเหนือจากงานฟังก์ชันงานหลักๆของระบบที่กล่าวไปแล้วยังมีส่วนเพิ่มเติมเพื่อสนับสนุนสภาพแวดล้อมในการทำงานอีก 3 ฟังก์ชันคือ

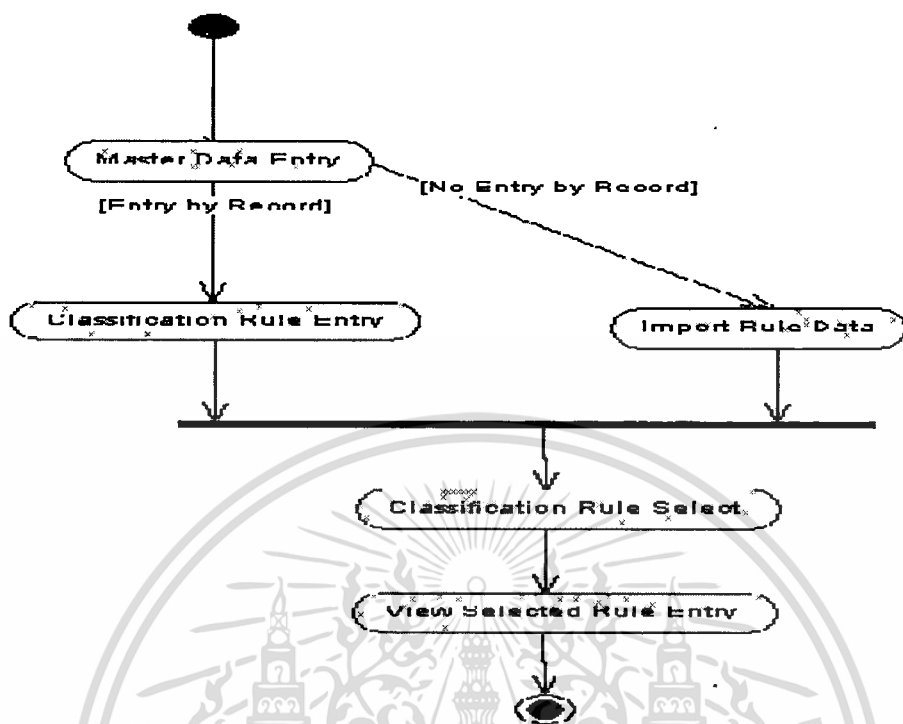
- 1) ฟังก์ชันฐานข้อมูลช่วยสนับสนุนการสืบค้นความรู้จากฐานข้อมูลขนาดใหญ่
- 2) ส่วนที่ใช้ในการเก็บข้อมูลเพื่อการประมวลผลและนำไปแสดงเป็นผลลัพธ์



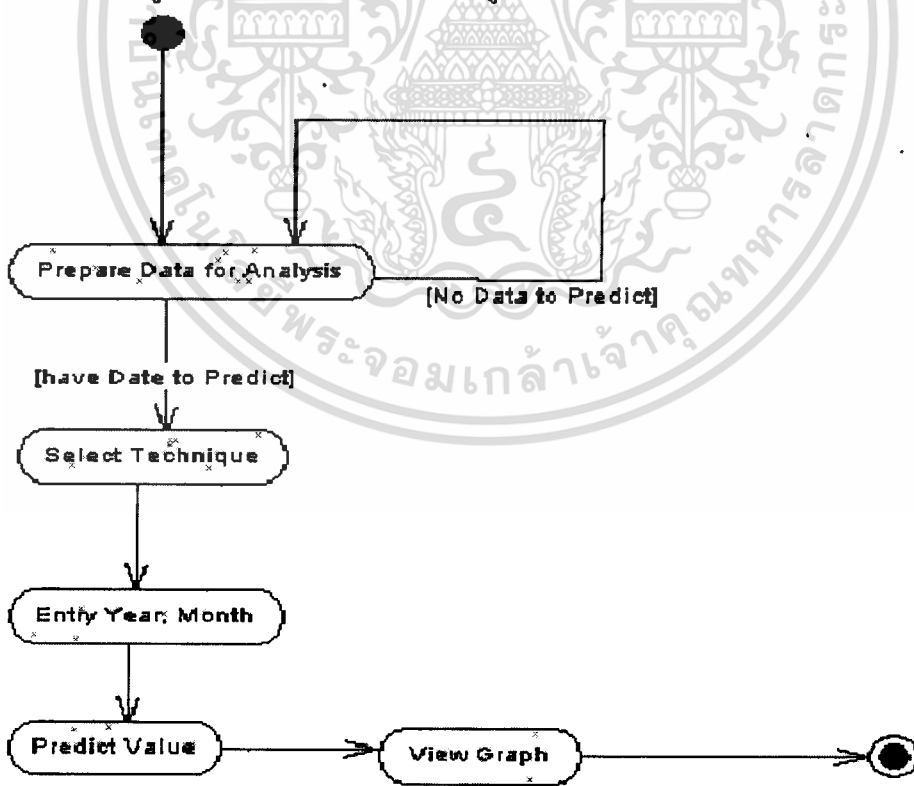
รูปที่ 3.4 แผนผังแสดงขั้นตอนการใช้งานโปรแกรม

### 3.4 ขั้นตอนการทำงานของระบบ

ระบบประกอบด้วยขั้นตอนการทำงาน 3 ส่วนหลักๆ คือ 1) การเตรียมข้อมูล 2) พยากรณ์โดยการวิเคราะห์อนุกรมเวลา 3) พยากรณ์โดยเทคนิคการจัดกลุ่มข้อมูล โดยแต่ละขั้นตอนมีรายละเอียดการทำงานดังรูปที่ 3.5 ถึง 3.7

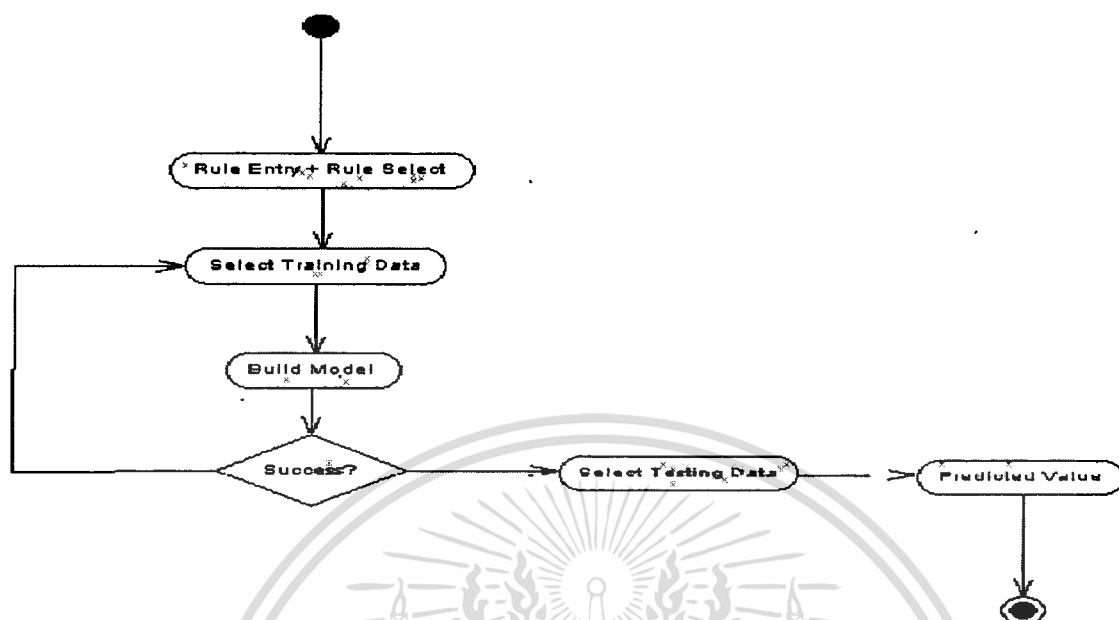


รูปที่ 3.5 ขั้นตอนการเตรียมข้อมูลในการสร้างแบบจำลอง



รูปที่ 3.6 ขั้นตอนการพยากรณ์โดยการวิเคราะห์อนุกรมเวลา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.7 ขั้นตอนการพยากรณ์โดยเทคนิคการจัดกลุ่มข้อมูล

### 3.5 สภาพแวดล้อมของการพัฒนาระบบ

สภาพแวดล้อมในการพัฒนาระบบ ประกอบด้วย 2 ส่วนได้แก่

#### 3.5.1 รายละเอียดด้านฮาร์ดแวร์ประกอบด้วย

- เครื่องคอมพิวเตอร์ ความเร็ว CPU-Pentium IV 2.4 GHz
- หน่วยความจำ (RAM) 512 MB
- ฮาร์ดดิสก์ ไม่ต่ำกว่า 5 GB
- เม้าส์ และ คีย์บอร์ด
- ซีดี-รอม
- Floppy Disk Drive

#### 3.5.2 รายละเอียดด้านซอฟต์แวร์ ประกอบด้วย

- ระบบปฏิบัติการ Windows XP
- โปรแกรม Microsoft SQL 2000 ใช้เป็นฐานข้อมูล
- โปรแกรม Microsoft Visual Basic .Net ในการพัฒนาโปรแกรม
- โปรแกรม Crystal Report

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.6 การพัฒนาโปรแกรม

การพัฒนาโปรแกรมในครั้งนี้ใช้ระบบจัดการฐานข้อมูลเป็น Microsoft SQL Server 2000 ส่วนโปรแกรมพัฒนาด้วย Microsoft Visual Basic .Net อยู่บนระบบปฏิบัติการ Windows XP โดยใช้ Crystal Reports ในการสร้างรายงาน โดยข้อมูลที่นำมาใช้จะเป็นข้อมูลจากระบบ ERP ซึ่งจะต้องมีหน้าต่างในการจัดการข้อมูลเพื่อดึงข้อมูลจากระบบตามช่วงเวลาที่กำหนดพร้อมทั้งทำการ ทำความสะอาด และ แปลงข้อมูล ก่อนที่จะนำไปสร้างแบบจำลองของกระบวนการค้าไม้หนึ่งต่อไป



## บทที่ 4

### กรณีศึกษาการค้นพบความรู้จากระบบบริหารสินค้าคงคลัง

#### 4.1 ลักษณะการดำเนินธุรกิจ

บริษัทที่นำข้อมูลมาวิเคราะห์ในครั้งนี้ เป็นบริษัทที่ทำธุรกิจทางการผลิตและจัดจำหน่ายสินค้าเกษตรแปรรูปโดยมีการรับซื้อวัตถุดิบที่เป็นผลผลิตทางการเกษตรทั้งจากเกษตรกรและพ่อค้าคนกลาง เพื่อนำมาทำการแปรรูปและจัดจำหน่ายให้กับลูกค้าทั้งในและต่างประเทศ โดยเน้นที่การส่งออกเป็นหลัก สินค้าที่นำแปรรูปยกตัวอย่างเช่น ข้าว ถั่ว น้ำมัน งา เป็นต้น เพื่อนำมาแปรรูปเป็นสินค้าสำเร็จรูปเช่น น้ำมันพืช วนเส้น ข้าวที่ผ่านการคัดและบรรจุ น้ำมันงา เป็นต้น ซึ่งพบว่าในปัจจุบันมีสินค้าและวัตถุดิบหลากหลายประเภทถูกจัดเก็บอยู่ภายในคลังสินค้า ดังนั้นการบริหารและจัดการและบริหารระบบคลังสินค้าจึงเป็นสิ่งจำเป็น และข้อมูลเป็นจำนวนมากภายในฐานข้อมูลไม่ได้ถูกนำมาใช้ประโยชน์ การนำเทคนิคดาต้าไมน์นิ่งเข้ามาช่วยในการวิเคราะห์ข้อมูลจึงเป็นสิ่งที่น่าสนใจและน่าจะเป็นประโยชน์สำหรับผู้บริหารในการบริหารจัดการคลังสินค้า และช่วยแสดงให้เห็นว่าเทคนิคที่นำมาใช้นั้น ได้ผลถูกต้องใกล้เคียงกับความเป็นจริงมากน้อยแค่ไหน

กระบวนการทางธุรกิจที่เกี่ยวข้องกับโครงการนี้ได้แก่ กระบวนการในการจัดซื้อวัตถุดิบเพื่อนำมาใช้ในการผลิต และ กระบวนการในการตรวจสอบคุณภาพและส่งคืนสินค้าที่ด้อยคุณภาพ ซึ่งสามารถอธิบายรายละเอียดได้ดังรูปที่ 4.1 และ 4.2

#### 4.2 วัตถุประสงค์ของการทำดาต้าไมน์นิ่ง

ในการศึกษาโครงการนี้ เป็นการนำเทคนิคดาต้าไมน์นิ่ง 2 อย่างมาทดสอบซึ่งประกอบด้วย

1. เทคนิคการจำแนกประเภทข้อมูล เพื่อใช้ในการจำแนกประเภทของวัตถุดิบที่ผ่านกระบวนการตรวจสอบคุณภาพแล้ว ได้ของเสียหรือของที่ไม่ผ่านการตรวจสอบคุณภาพมีปริมาณแตกต่างกันไป โดยมีปัจจัยในการพิจารณาหลายข้อตามที่ผู้ใช้กำหนด

2. เทคนิคการพยากรณ์ข้อมูลโดยใช้ทฤษฎี Time Series ในการสร้างแบบจำลอง เพื่อใช้ในการศึกษาถึงระดับปริมาณของวัตถุดิบที่คงเหลือภายในคลังสินค้าในแต่ละช่วงเวลา. และการศึกษาถึงระดับปริมาณของสินค้าคงเหลือภายในคลังสินค้าในแต่ละช่วงเวลา เพื่อช่วยประมาณการการสั่งซื้อสินค้าและ การวางแผนการผลิตภายในช่วงเวลาที่ต้องการพยากรณ์

ตารางที่ 4.1 กระบวนการในการจัดซื้อและรับเข้าวัตถุดิบ

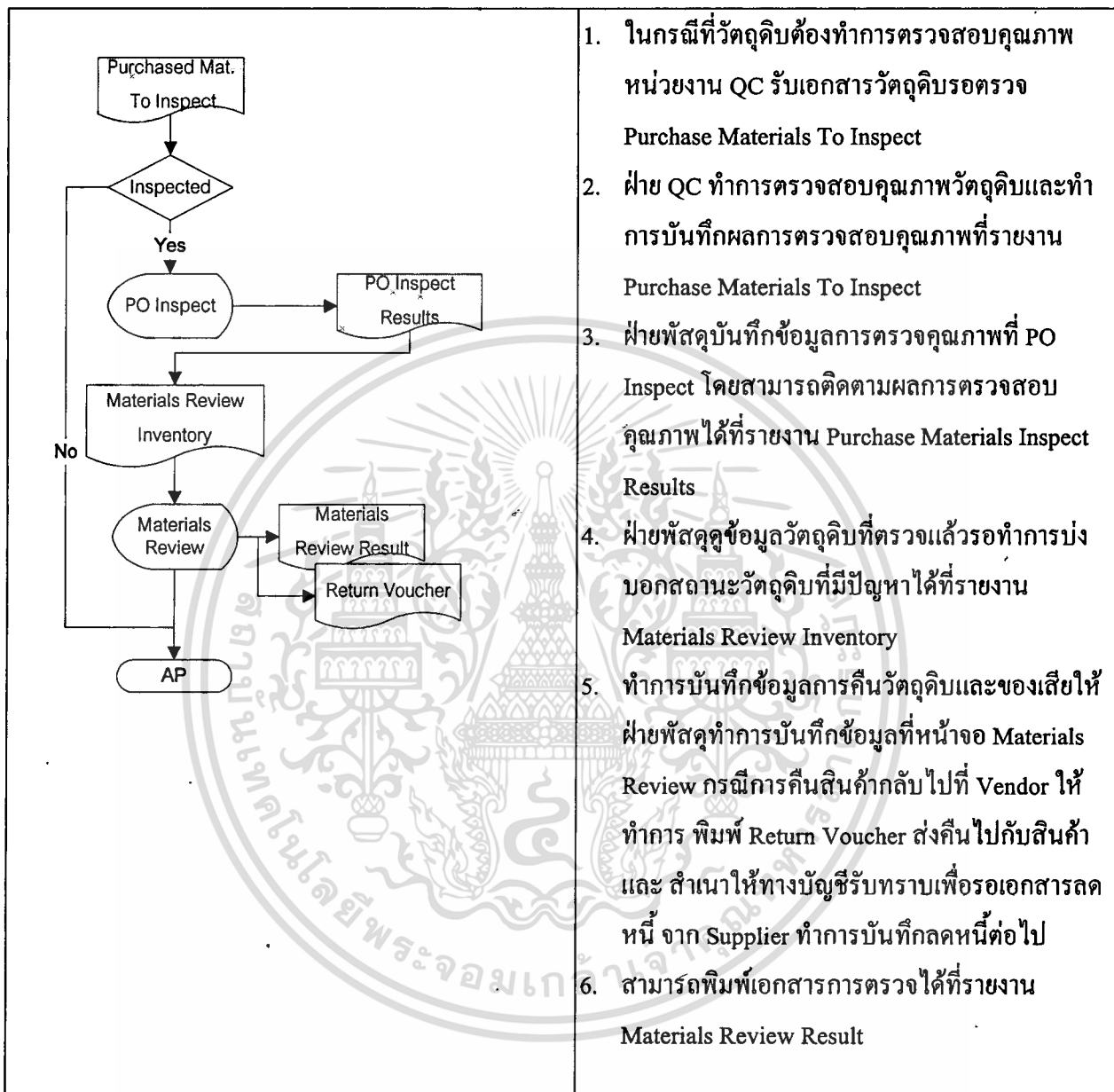
System Flow	Requirement
<pre> graph TD     A[Plan Order To Start (Materials)] --&gt; B((Convert PP to PO))     B --&gt; C[PO Receipt Listing]     B --&gt; D((Purchase Order Card))     D --&gt; E[Purchase Order (R/M)]     D --&gt; F[/Incoming Receipt/]     F --&gt; G((PO Receipt))     G --&gt; H[PO Receipt Listing]     G --&gt; I[Packing Card]     G --&gt; J[Invoice]     H --&gt; K[PO Receipt Listing]     H --&gt; L[Purchased Mat. To Inspect]     H --&gt; M[Invoice]     </pre>	<ol style="list-style-type: none"> <li>ฝ่ายพัสดุตรวจสอบแผนการผลิตได้ที่รายงาน Plan Order To Start</li> <li>ฝ่ายพัสดุทำการแปลงแผนสั่งซื้อเป็นคำสั่งซื้อที่หน้าจอ Convert Plan PO To PO โดยสามารถดูข้อมูลการแปลงแผนสั่งซื้อได้ที่รายงาน Convert Plan Purchased Order</li> <li>ฝ่ายพัสดุทำการได้แปลงจากแผนสั่งซื้อที่หน้าจอ Purched Order Card และพิมพ์ใบสั่งซื้อได้ที่ Purchase Order Report ส่ง Fax ให้ Supplier โดยสามารถตรวจสอบข้อมูลการสั่งซื้อได้ที่รายงาน Purched Order Detail</li> <li>กรณีวัตถุดิบส่งถึง โรงงานพร้อมเอกสาร Invoiced และ Packing Card ให้ทำการบันทึกที่รับเข้าที่หน้าจอ PO Receipt โดยสามารถดูข้อมูลการรับเข้าได้ที่ PO Receipt Listing และพิมพ์ส่งบัญชียืนยันการรับ</li> <li>กรณีวัตถุดิบที่ต้องตรวจสอบคุณภาพเช่น เหล็ก ให้ทำการพิมพ์รายงาน Purched Materials To Inspect ส่งหน่วยงาน QC</li> </ol>

4.3 การเตรียมข้อมูล

ในงานวิจัยนี้ใช้ข้อมูลที่เกี่ยวข้องกับระบบบริหารสินค้าคงคลังของบริษัทขายสินค้าผลิตภัณฑ์ทางการเกษตร แหล่งข้อมูลที่ใช้เป็นกรณีศึกษาสำหรับการสร้างและทดสอบโครงงานนี้เป็นข้อมูลในส่วนของ การจัดซื้อและตรวจสอบคุณภาพวัตถุดิบเพื่อนำมาใช้ในการผลิต ซึ่งข้อมูลในส่วนนี้มาวิเคราะห์และสร้างแบบจำลองในการแบ่งกลุ่มของข้อมูล โดยเริ่มจากการข้อมูลการออกใบสั่งซื้อ การรับสินค้า การตรวจสอบคุณภาพของสินค้า ตามลำดับ ข้อมูลอีกส่วนหนึ่งเป็นข้อมูลปริมาณสินค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 กระบวนการในการการตรวจสอบคุณภาพ, การรับสินค้าเข้าและการทำรายการคืนวัตถุดิบ



สำเร็จรูปที่เกิดจากกระบวนการผลิตและคงค้างอยู่ในคลังสินค้าในแต่ละเดือน ซึ่งข้อมูลในส่วนนี้นำมาวิเคราะห์และสร้างแบบจำลองในการพยากรณ์ระดับปริมาณสินค้าคงคลัง แหล่งข้อมูลที่ใช้ในการพัฒนาระบบแบ่งเป็น 2 ส่วนดังนี้

- ข้อมูลแม่แบบ(Master Data) เป็นกฎที่ใช้ตรวจสอบคุณลักษณะของวัตถุดิบในการสร้างแบบจำลอง มีการตั้งค่าเริ่มต้น เช่น สิ่งปลอมปน, การจับตัว เป็นต้น เป็นข้อมูลจากภายนอกออกแบบเพื่อให้เหมาะสมสามารถใช้ในการสร้างแบบจำลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ข้อมูลทำการ(Transaction Data) เป็นข้อมูลที่เกิดขึ้นประจำวันนำมาใช้สร้างแบบจำลอง Training, Testing Data เป็นข้อมูลที่มีขนาดใหญ่แล้วนำมาจากระบบ ERP โดยมีการปรับแต่งให้สามารถใช้ในการวิเคราะห์ความน่าจะเป็นของกฎต่างๆ

#### 4.4 การคัดเลือกข้อมูล

ข้อมูลที่นำมาใช้ในการดำเนินงานทั้งหมด รวบรวมมาจากฐานข้อมูลของระบบ ERP ซึ่งจะ เป็นข้อมูลในส่วนของการบริหารสินค้าคงคลัง ทำให้ข้อมูลที่ได้มีความน่าเชื่อถือและมีความถูกต้อง โดยข้อมูลที่นำมาศึกษาเป็นข้อมูลตั้งแต่ เดือนมกราคม 2547 ถึงเดือนธันวาคม 2547 โดยมีการคัดเลือก ข้อมูลมาเก็บไว้ใน Microsoft SQL Server 2000 โดยเลือกเฉพาะข้อมูลบางส่วนของที่จะนำมาใช้ใน กระบวนการทำดาต้าไมน์นิ่ง ซึ่งสามารถแสดงออกมาเป็นตารางที่เกี่ยวข้องได้ดังนี้

1. ตารางสินค้าและวัตถุดิบ (IMA) ซึ่งเก็บข้อมูลเกี่ยวกับรายละเอียดของสินค้าและวัตถุดิบ ได้แก่ รหัสสินค้า ชื่อสินค้า หน่วยนับ ราคาต่อหน่วย ปริมาณสินค้าคงคลัง เป็นต้น ดัง ตารางที่ 4.1
2. ตารางทรานแซกชัน (ITR) จัดเก็บข้อมูลที่เป็นทรานแซกชันการเบิกจ่ายวัตถุดิบและรับเข้า สินค้าสำเร็จรูป ดังตารางที่ 4.2
3. ตารางเก็บค่าผลลัพธ์ของการตรวจวัดคุณภาพของสินค้ารับเข้า (M\_Incoming\_QC) เก็บ ข้อมูลเกี่ยวกับรายละเอียดของการรับสินค้าที่เป็นวัตถุดิบ ดังตารางที่ 4.3
4. ตาราง (M\_Incoming\_QC\_Detail) เก็บข้อมูลเกี่ยวกับรายละเอียดของการตรวจสอบ คุณภาพของการรับวัตถุดิบ ดังตารางที่ 4.4
3. ตารางรายละเอียดการสั่งซื้อวัตถุดิบ (POM) เก็บข้อมูลรายละเอียดของใบสั่งซื้อแต่ละใบ ดังตารางที่ 4.5
4. ตารางรายละเอียดการสั่งซื้อวัตถุดิบแต่ละรายการ (POI) เก็บข้อมูล ปริมาณ ราคา ของแต่ละ รายการของวัตถุดิบที่สั่งซื้อในแต่ละใบสั่งซื้อ ดังตารางที่ 4.6
5. ตารางแยกประเภทผลการตรวจสอบคุณภาพ (M\_Defect) เก็บข้อมูลเกี่ยวกับเกณฑ์การวัด คุณภาพของวัตถุดิบที่ทำการตรวจสอบ ดังตารางที่ 4.7
6. ตารางเก็บค่าที่ใช้ในการวัดค่าของการตรวจสอบคุณภาพของวัตถุดิบ (M\_Defect\_Value) เก็บข้อมูลของค่าที่ใช้เป็นเกณฑ์ในการวัดเพื่อนำไปหาค่าผลลัพธ์ที่ได้จากการวัดดังตาราง ที่ 4.8

รายละเอียดต่างๆ แสดงในตาราง จะกำหนดว่าข้อมูลในแอตทริบิวต์ นั้นสามารถเป็นค่าว่างได้หรือไม่และเป็นข้อมูลประเภทใดโดยจะระบุเฉพาะแอตทริบิวต์ ที่จำเป็นดังต่อไปนี้

ตารางที่ 4.1 ตารางรายละเอียดสินค้าและวัตถุดิบ IMA

ชื่อ แอตทริบิวต์	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
IMA_ItemID	nvarchar	50	N
IMA_ItemName	nvarchar	200	Y
IMA_Classification	nvarchar	60	Y
IMA_UnitMeasure	nvarchar	20	N
IMA_Price	money	8	N
IMA_OnHandQty	Float	8	N
IMA_Weight	Float	53	Y
IMA_ProdFam	nvarchar	60	N
IMA_ProdModel	nvarchar	60	Y
IMA_MaxStockQty	Float	8	N
IMA_MinStockQty	Float	8	N
IMA_SafetyStockQty	Float	8	N
IMA_MinOrderQty	Float	8	N
IMA_MaxOrderQty	Float	8	N
IMA_StdCostAmt	Float	8	N

ตารางที่ 4.2 ตารางแสดงทรานแซกชันการเบิกจ่ายสินค้า ITR

ชื่อ แอตทริบิวต์	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
ITR_ItemID	Nvarchar	50	N
ITR_TransNbr	Int	4	N
ITR_TransQty	Float	8	N
ITR_TransDate	Datetime	8	Y
ITR_TransType	Nvarchar	40	Y
ITR_Reason	Nvarchar	100	Y
ITR_OnHandQtyPrior	Float	8	N

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 ตารางแสดงทรานแซกชันการเบิกจ่ายสินค้า ITR (ต่อ)

ชื่อ แอตทริบิวต์	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
ITR_UnitCostAmt	Money	8	N
ITR_ReceiverDate	Datetime	8	Y
ITR_ShipDate	Datetime	8	Y

ตารางที่ 4.3 ตารางเก็บค่าผลลัพธ์ของการตรวจวัดคุณภาพของสินค้ารับเข้า(M\_Incoming\_QC)

ชื่อ แอตทริบิวต์	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
IRI_ID	Nvarchar	30	N
po_no	Nvarchar	50	Y
line_no	Int	4	Y
Receipt_Date	Datetime	15	Y
Car_license	Nvarchar	30	Y
Tank_no	Nvarchar	50	Y
Qty_inspected	decimal	18, 2	Y
Net_weight	decimal	18, 2	Y
Gross_weight	decimal	18, 2	Y
Sizel	decimal	18, 2	Y
Toxic_level	decimal	18, 2	Y
QC_Result	nvarchar	80	Y
Remark	nvarchar	255	Y
QC_note	nvarchar	255	Y
product_id	nvarchar	80	Y
Required_Date	datetime	15	Y
Receive_qty	decimal	18, 2	Y
Lot_no	nvarchar	255	Y
Itemid_DG	varchar	80	Y

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.4 ตารางเก็บรายละเอียดของการตรวจวัดคุณภาพของสินค้า (M\_Incoming\_QC\_Detail)

ชื่อ แอตทริบิวต์	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
ID	int	4	N
IRI_ID	nvarchar	100	N
Defect_Type	nvarchar	100	Y
QC_Value	decimal	9	Y

ตารางที่ 4.5 ตารางรายละเอียดการสั่งซื้อวัตถุดิบ (POM)

ชื่อ แอตทริบิวต์	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
POM_PurchOrderID	nvarchar	20	N
POM_VendorID	nvarchar	20	N
POM_PurchOrderDate	datetime	8	N
POM_ShipName	nvarchar	60	N
POM_ShipAddress	nvarchar	100	Y

ตารางที่ 4.6 ตารางรายละเอียดการสั่งซื้อวัตถุดิบแต่ละรายการ (POI)

ชื่อ แอตทริบิวต์	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
POI_PurchOrderID	nvarchar	20	N
POI_POLineNbr	smallint	2	N
POI_LineItemValAmt	money	8	N
POI_UnitMeasure	nvarchar	20	N
POI_ItemName	nvarchar	200	N
POI_ItemID	nvarchar	50	N

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 ตารางแยกประเภทผลการตรวจสอบคุณภาพ (M\_Defect)

ชื่อ แอตทริบิวต์	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
Defect_ID	nvarchar	20	N
Defect_Name	nvarchar	100	N
Std_Value	nvarchar	100	N
Defect_name_eng	nvarchar	100	Y
std_value_id	int	4	N

ตารางที่ 4.8 ตารางเก็บค่าที่ใช้ในการวัดค่าของการตรวจสอบคุณภาพของวัตถุดิบ(M\_Defect\_Value)

ชื่อ แอตทริบิวต์	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
Defect_Value_ID	int	4	N
Symbol	char	10	N
Defect_Value_One	decimal	9	N
Defect_Value_Two	decimal	9	Y

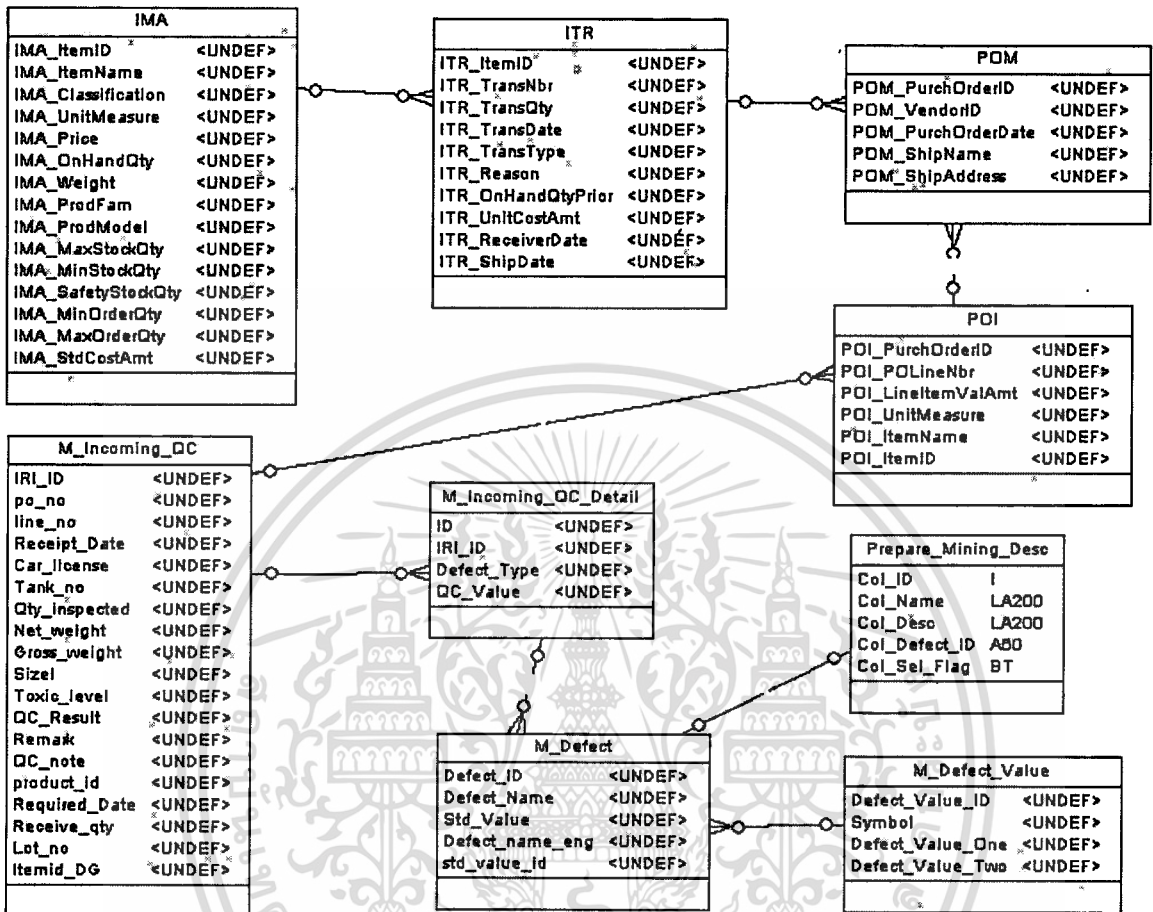
โดยสามารถแสดงความสัมพันธ์ของแต่ละตารางด้วย อีอาร์ไดอะแกรมดังรูปที่ 4.1

#### 4.5 การวิเคราะห์และออกแบบระบบงาน

การวิเคราะห์ และออกแบบระบบงานประกอบด้วยขั้นตอนการทำงานหลักดังนี้

##### 4.5.1 การคัดเลือกข้อมูล

ในส่วนนี้เป็นการเตรียมข้อมูลให้พร้อมสำหรับการนำไปสร้างแบบจำลอง โดยจะต้องกำจัดข้อมูลที่ผิดพลาดหรือไม่สมบูรณ์ทิ้งไป เช่น ปริมาณสินค้าคงคลังไม่สามารถเป็นค่าติดลบได้ ปริมาณการเบิกจ่ายสินค้าไม่ควรมีค่าเป็นศูนย์ เป็นต้น และคัดเลือก แอตทริบิวต์ ที่ต่างๆเพื่อนำมาใช้ในการวิเคราะห์ เพื่อที่จะได้ แอตทริบิวต์ ที่มีแนวโน้มว่าจะเกี่ยวข้องกับผลลัพธ์เป้าหมายที่ต้องการ เพื่อที่จะนำแอตทริบิวต์ เหล่านั้นมาใช้ในการสร้างแบบจำลองในขั้นตอนต่อไป โดยได้ทำการคัดเลือก แอตทริบิวต์ จากตารางต่างๆมีรายละเอียดดังนี้



รูปที่ 4.1 อีอาร์ไดอะแกรม แสดงความสัมพันธ์ของแต่ละตารางในฐานข้อมูล

1. จากตาราง IMA มี แอตทริบิวต์ ที่เลือกมาใช้ดังต่อไปนี้  
 IMA\_ItemID, IMA\_ItemName, IMA\_Price, IMA\_OnHandQty
2. จากตาราง ITR มี แอตทริบิวต์ ที่เลือกมาใช้ดังต่อไปนี้  
 ITR\_ItemID, ITR\_TransNbr, ITR\_TransQty, ITR\_PriorOnHandQty, ITR\_transDate,  
 ITR\_TransType
3. จากตาราง M\_Incoming\_QC มี แอตทริบิวต์ ที่เลือกมาใช้ดังต่อไปนี้  
 IRI\_ID, po\_no, line\_no, QC\_Result, Lot\_no, Qty\_inspected
4. จากตาราง M\_Incoming\_QC\_Detail มี แอตทริบิวต์ ที่เลือกมาใช้ดังต่อไปนี้  
 IRI\_ID, Defect\_Type, QC\_Value
5. จากตาราง POM มี แอตทริบิวต์ ที่เลือกมาใช้ดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

POM\_PurchOrderID, POM\_PurchOrderDate, POM\_OrderValAmt

6. จากตาราง POI มี แอตทริบิวต์ ที่เลือกมาใช้ดังต่อไปนี้

POI\_PurchOrderID, POI\_POLineNbr, POI\_LineItemValAmt, POI\_ItemID

7. จากตาราง M\_Defect มี แอตทริบิวต์ ที่เลือกมาใช้ดังต่อไปนี้

defect\_id, defect\_name

8. จากตาราง M\_Defect\_Value มี แอตทริบิวต์ ที่เลือกมาใช้ดังต่อไปนี้

Defect\_Value\_ID, Symbol, Defect\_Value\_One, Defect\_Value\_Two

#### 4.5.2 การจัดกลุ่มข้อมูล

ในส่วนนี้เป็นการจัดกลุ่มของข้อมูลของสินค้าและวัตถุดิบหลายๆประเภทให้อยู่กลุ่มเดียวกัน เพื่อลดการกระจายของข้อมูล การจัดกลุ่มมีวัตถุประสงค์เพื่อจัดรายการสินค้าและวัตถุดิบที่มีอยู่หลายประเภทให้มาอยู่กลุ่มเดียวกัน เพื่อใช้สำหรับหาความสัมพันธ์ระหว่างสินค้าและวัตถุดิบได้หลากหลายมากขึ้น

#### 4.5.3 การทำความสะอาดข้อมูล

เป็นกระบวนการที่อยู่ในส่วนการเตรียมข้อมูล โดยจะทำการทำความสะอาดข้อมูลหรือตัดข้อมูลที่ไมถูกต้องออกด้วย เพื่อให้ได้ข้อมูลเฉพาะข้อมูลที่ดี ที่จะนำไปใช้ในการสร้างแบบจำลอง โดยการดำเนินการตัดข้อมูลที่ไมถูกต้องทั้งหมดมีเงื่อนไขดังนี้

- แอตทริบิวต์ ที่มีค่าเท่ากับ 0 หรือ Null
- แอตทริบิวต์ ที่มีค่าไม่ตรงกับความต้องการ เช่น N/A

#### 4.5.4 การนำข้อมูลเข้าตารางข้อมูลสำหรับการสร้างแบบจำลอง

โดยหลังจากทำการคัดเลือก แอตทริบิวต์ ที่ต้องการได้แล้ว จึงใช้คำสั่ง SQL เพื่อนำข้อมูลจากตารางต่างๆมารวมกันเป็นตารางเดียวกัน โดยมีคำสั่ง SQL ดังนี้

1. คำสั่งคิวรี สำหรับรวบรวมข้อมูลเพื่อสร้างแบบจำลอง Classification

```

select po_no + ':' + lot_no, defect_type, qc_value,
'QC Result' = case
  when ltrim(rtrim(symbol)) = '<' then
    --'Less than'
    case when qc_value < Defect_Value_One then
      1
    else
      0
    end
  when ltrim(rtrim(symbol)) = '>' then
    --'More than'
    case when qc_value > Defect_Value_One then
      1
    else
      0
    end
  when ltrim(rtrim(symbol)) = '=' then
    --'Equal'
    case when qc_value = Defect_Value_One then
      1
    else
      0
    end
  when ltrim(rtrim(symbol)) = 'Between' then

```

```

--'Between'
case when qc_value < Defect_Value_One and qc_value < Defect_Value_Two then
    1
else
    0
end
End, QC_Result
from m_incoming_qc, m_incoming_qc_detail, m_defect, poi, m_defect_value
where m_incoming_qc.iri_id = m_incoming_qc_detail.iri_id
and defect_id = defect_type
and poi_purchorderid = po_no
and poi_polinenbr = line_no
and std_value_id = Defect_Value_ID
and poi_itemid = @ItemID
order by po_no, line_no, defect_type

```

จากชุดคำสั่ง SQL จะเชื่อมโยงตารางต่างๆแล้วนำข้อมูลที่ได้คัดเลือกมาเฉพาะตามเงื่อนไขที่ตั้งไว้แล้วนำค่าข้อมูลที่ได้นำมาจัดเก็บในตาราง Prepare\_Mining โดยจะเก็บค่าผลลัพธ์ของความน่าจะเป็นในแต่ละ Lot ว่าแต่ละเงื่อนไขใดบ้างของ Lot นั้นที่เป็นจริงซึ่งมีข้อจำกัดในการเก็บข้อมูลเงื่อนไขได้ไม่เกิน 60 เงื่อนไขและเก็บผลลัพธ์ของการทดสอบของแต่ละ Lot ด้วย โดยมีลักษณะโครงสร้างของตารางดังนี้

ตารางที่ 4.9 ตาราง Prepare\_Mining เก็บข้อมูลผลลัพธ์ของแต่ละ Lot

ชื่อ Column	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
Lot_No	nvarchar	160	Y
Defect1	Int	4	Y
Defect2	Int	4	Y
.....	Int	4	Y
.....	Int	4	Y
Defect60	Int	4	Y
Test_result	Nvarchar	30	Y

จากนั้นทำการแปลงผลลัพธ์ที่ได้จากตาราง Prepare\_Mining มาจำแนกเป็นค่าความน่าจะเป็นของแต่ละเงื่อนไขว่ามีค่าความน่าจะเป็นที่จะผ่านหรือไม่ผ่านการทดสอบคุณภาพเป็นเท่าไร ทำให้ได้ตารางใหม่ตามโครงสร้างที่วางไว้ดังตารางที่ 4.10

ตารางที่ 4.10 ตารางที่ใช้ในการสร้างแบบจำลอง Prepare\_Mining\_Probability

ชื่อ Column	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
Criteria	nvarchar	160	Y
Pass_Value	int	4	Y
NoPass_Value	int	4	Y
Pass_Prob	decimal	9	Y
NoPass_Prob	decimal	9	Y

ตารางที่ 4.11 ตารางที่ใช้ในการเก็บข้อมูลที่ได้จากการประมวลผล training Prepare\_Mining\_Training

ชื่อ Column	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
ItemID	nvarchar	100	Y
Training_Date	datetime	8	Y
Criteria	nvarchar	160	Y
Pass_Value	int	4	Y
NoPass_Value	int	4	Y
Pass_Prob	decimal	9	Y
NoPass_Prob	decimal	9	Y

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. คำสั่ง SQL สำหรับรวบรวมข้อมูลเพื่อสร้างแบบจำลองพยากรณ์ข้อมูลโดยอาศัยหลักการอนุกรมเวลา

```

SELECT MonthNbr, round(avg(ITR_OnHandQtyPrior), 2) As MonthAverage,
round(avg(ITR_OnHandQtyPrior), 2) * round(avg(ITR_OnHandQtyPrior), 2),
MAX(ITR_TransDate) As MonthLastDay FROM ITR,
(select ITR_ItemID, MAX(ITR_TransDate) As MonthLastDay, month(ITR_TransDate) As
MonthNbr, year(ITR_TransDate) As YearNbr
, max(ITR_TransNbr) As TransNbr
from itr
where ITR_ItemID = @ProductID
AND Year(ITR_TransDate) = 2004
Group By ITR_ItemID, month(ITR_TransDate), year(ITR_TransDate)) TransSummary
WHERE ITR_TransNbr = TransSummary.TransNbr
Group By YearNbr, MonthNbr, ITR_transNbr, TransSummary.ITR_ItemID
Order By YearNbr, MonthNbr

```

จากชุดคำสั่ง SQL จะเชื่อมโยงตารางต่างๆแล้วนำข้อมูลที่ได้คัดเลือกมาเฉพาะตามเงื่อนไขที่ตั้งไว้โดยผ่าน Stored Procedure ทำให้ได้ตารางใหม่ตามโครงสร้างที่วางไว้

ตารางที่ 4.10 ตารางที่ใช้ในการสร้างแบบจำลองการวิเคราะห์ข้อมูลด้วยทฤษฎี Time Series

ชื่อ Column	ประเภทข้อมูล	ความยาว	เป็นค่าว่าง(Y/N)
UnitNbr	int	4	Y
Xcol	int	4	Y
Ycol	decimal	9	Y
XYCol	decimal	9	Y
XSqr	decimal	9	Y
LastDay	datetime	8	Y

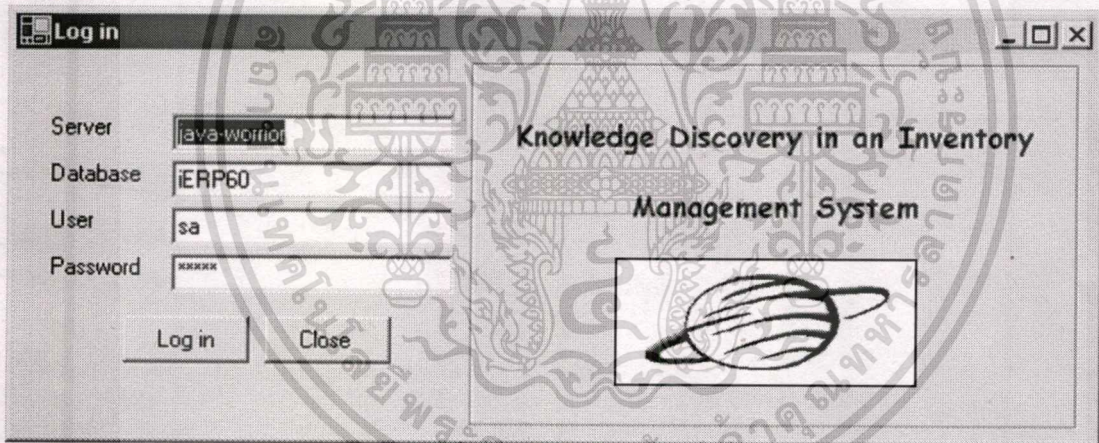
ข้อมูลจากตารางนี้ไปหาดัชนีฤดูกาลเพื่อขจัดอิทธิพลของฤดูกาลออกจากข้อมูลแล้วทำการวิเคราะห์ทางโน้มของข้อมูลต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 5

### การศึกษารูปแบบและการพยากรณ์

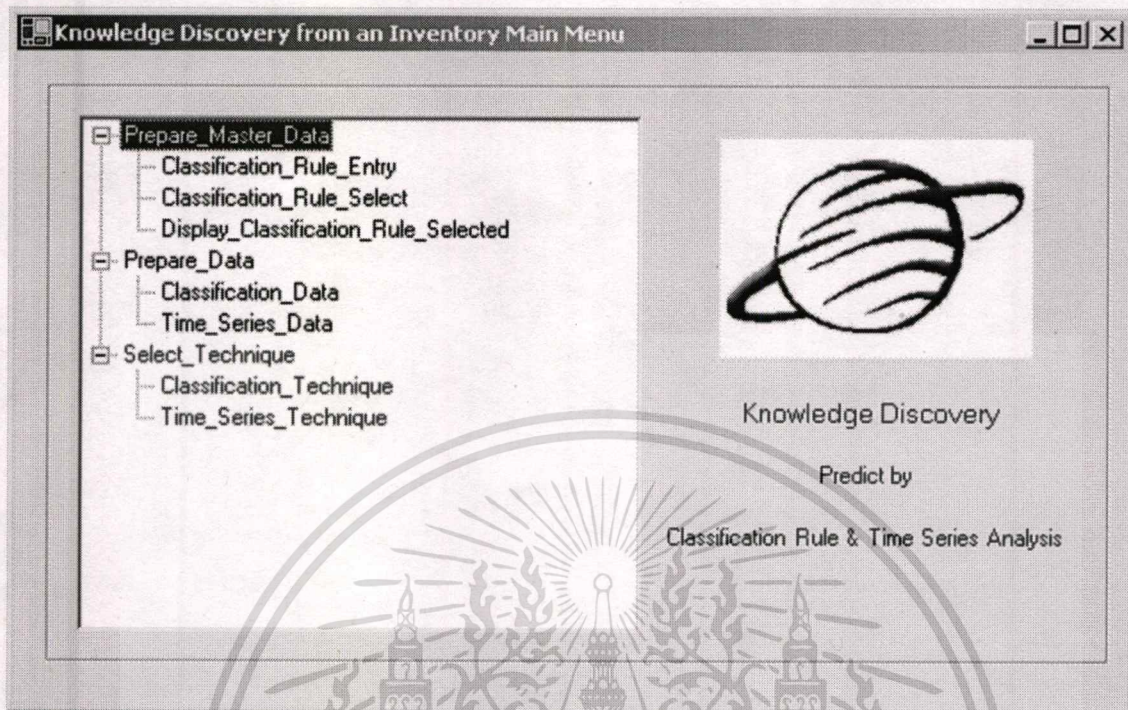
จากบทที่ผ่านมาเป็นการนำเสนอกระบวนการในการเตรียมข้อมูลเพื่อให้ได้ข้อมูลที่มีคุณภาพ ซึ่งเมื่อจัดเตรียมข้อมูลเรียบร้อยแล้วจึงนำข้อมูลที่ได้มาทำการสร้างแบบจำลองสำหรับการค้นหาข้อมูล วัตถุประสงค์ที่ไม่ผ่านการตรวจสอบคุณภาพในส่วนของการจัดกลุ่มข้อมูลด้วยแบบจำลองการแบ่งกลุ่ม ข้อมูล และการพยากรณ์ยอดขายของสินค้าสำเร็จรูปด้วยภาวะวิเคราะห์รูปแบบของข้อมูลด้วยวิธีการทาง สถิติ โดยในบทนี้จะกล่าวถึงกระบวนการในการเตรียมข้อมูล สร้างแบบจำลองและนำเอาผลจากการ สร้างแบบจำลองไปใช้ในการพยากรณ์ต่อไป



รูปที่ 5.1 หน้าจอล็อกอินเพื่อเริ่มต้นระบบ

หน้าจอแรกของระบบเพื่อทำการล็อกอิน เข้าสู่ระบบมีองค์ประกอบดังนี้

- Server: กรอกชื่อเซิร์ฟเวอร์ ที่ต้องการติดต่อเพื่อเข้าใช้งานฐานข้อมูล
- Database: ชื่อของฐานข้อมูลที่ต้องการติดต่อ
- User: ล็อกอินของผู้ใช้ที่มีสิทธิ์ในการเข้าใช้ฐานข้อมูล
- Password: รหัสผ่าน



รูปที่ 5.2 หน้าจอหลัก

หน้าจอหลัก จะจัดแบ่งเมนูต่างๆตามหมวดหมู่ดังนี้

1. Prepare Master Data: หมวดหมู่ของเมนูสำหรับการเตรียมตั้งค่าข้อมูลเริ่มต้นข้อมูลในการจัดกลุ่มข้อมูลและการสร้างแบบจำลองในการพยากรณ์

1.1 Classification Rule Entry

เมนูลิงค์ไปยังหน้าจอการเพิ่มกฎการจัดกลุ่มข้อมูล

1.2 Classification Rule Select

เมนูลิงค์ไปยังหน้าจอการเลือกกฎการจัดกลุ่มข้อมูล

1.3 Display Classification Rule Selected

เมนูลิงค์ไปยังหน้าจอการแสดงผลการเลือกกฎที่ต้องการใช้ในการจำแนกประเภทของข้อมูล

2. Prepare Data: หมวดหมู่ของเมนูสำหรับการเตรียมข้อมูลเริ่มต้นข้อมูลในการสร้างแบบจำลอง

2.1 Classification Data

เมนูลิงค์ไปยังหน้าจอเตรียมข้อมูลเพื่อใช้ในการเตรียมข้อมูลการสร้าง โมเดลการจัดกลุ่มข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.2 Time Series Data

เมนูลิงค์ไปยังหน้าจอเตรียมข้อมูลเพื่อใช้ในการพยากรณ์ด้วยเทคนิค Time Series

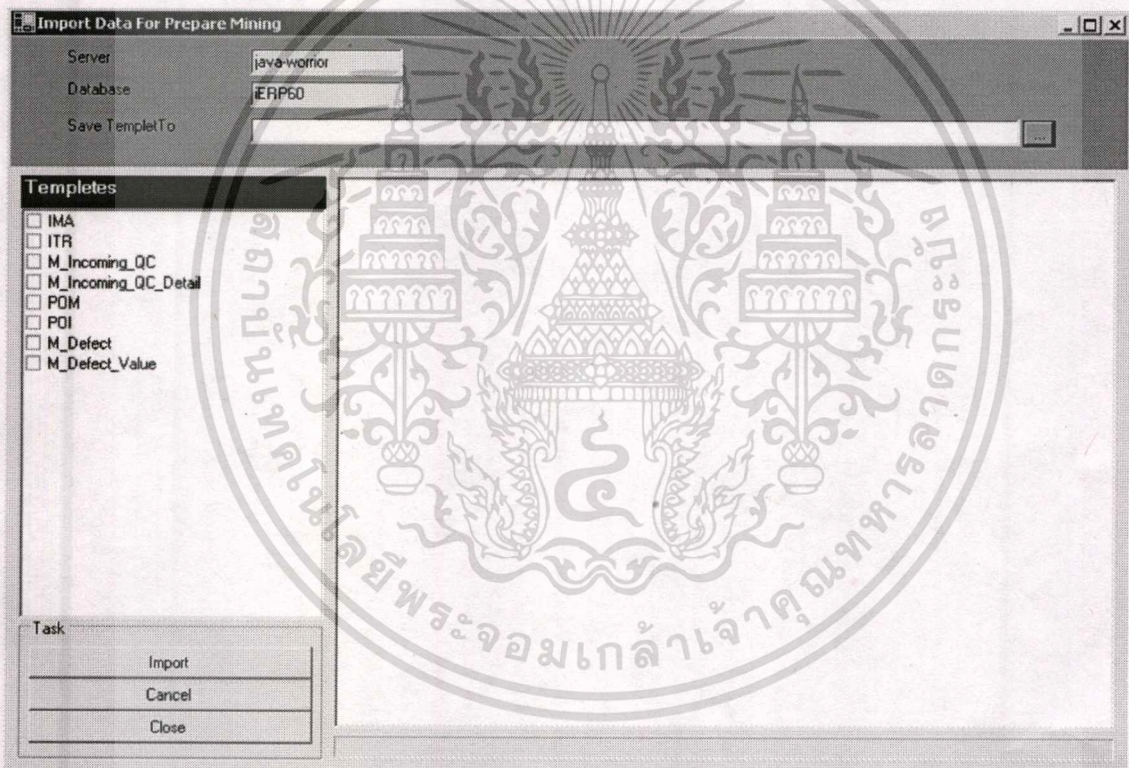
## 3. Select Technique

### 3.1 Classification Technique

เมนูลิงค์ไปยังหน้าจอการพยากรณ์ข้อมูลจากแบบจำลองการแบ่งกลุ่มของข้อมูล

### 3.2 Time Series Technique

เมนูลิงค์ไปยังหน้าจอการพยากรณ์ข้อมูลด้วยทฤษฎี Time Series



รูปที่ 5.3 หน้าจอนำเข้าข้อมูลเข้า

ทำหน้าที่ในการนำเข้าข้อมูลภายนอกในรูปแบบไฟล์ Excel เพื่อนำไปใช้ในการสร้างแบบจำลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Prepare Master Data

Classification Rule Entry | Classification Rule Select | Display Classification Rule Selected

RuleID	Rule_Name	Rule_Descript	Symbol	Defect_Value	Defect_Value
1	LotNbr	-	<	0.1000	-
2	อะไหล่	-	<	0.2000	-
3	ความชื้น	-	<	1.0000	-
4	ข้าวขาวปน.เต็ม	-	<	10.0000	-
5	พื้นข้าวความยาว	-	<	12.0000	-
6	พื้นข้าวความยาว	-	<	14.0000	-
58	Test	TestTest	>	0.9000	0.0000
10	ข้าวหัก4.3mm	-	<	5.0000	-
11	ข้าวหัก4.2mm	-	<	7.0000	-
12	ปลายข้าว	-	<	8.0000	-
13	เส้นแฉก	-	>	6.0000	-
14	ท้องไร่	-	>	60.0000	-
15	ข้าวเสีย	-	=	0	-

Rule ID:

Rule Name:

Rule Description:

Symbol:

Value 1:  Value 2:

New | Accept | Cancelled | Delete | Close

รูปที่ 5.4 หน้าจอสร้างกฎสำหรับจัดกลุ่มข้อมูล

หน้าจอการเพิ่มค่ากฎในการจัดกลุ่มข้อมูลด้วยแบบจำลอง Classification โดยจะทำเป็นขั้นตอนแรกเพื่อกำหนดกฎและค่าที่ใช้ในการแบ่งกลุ่มข้อมูล โดยจะกำหนดชื่อของกฎ คำอธิบาย เครื่องหมายที่ใช้ และค่าที่ต้องการกำหนด

Prepare Master Data

Classification Rule Entry | Classification Rule Select | Display Classification Rule Selected

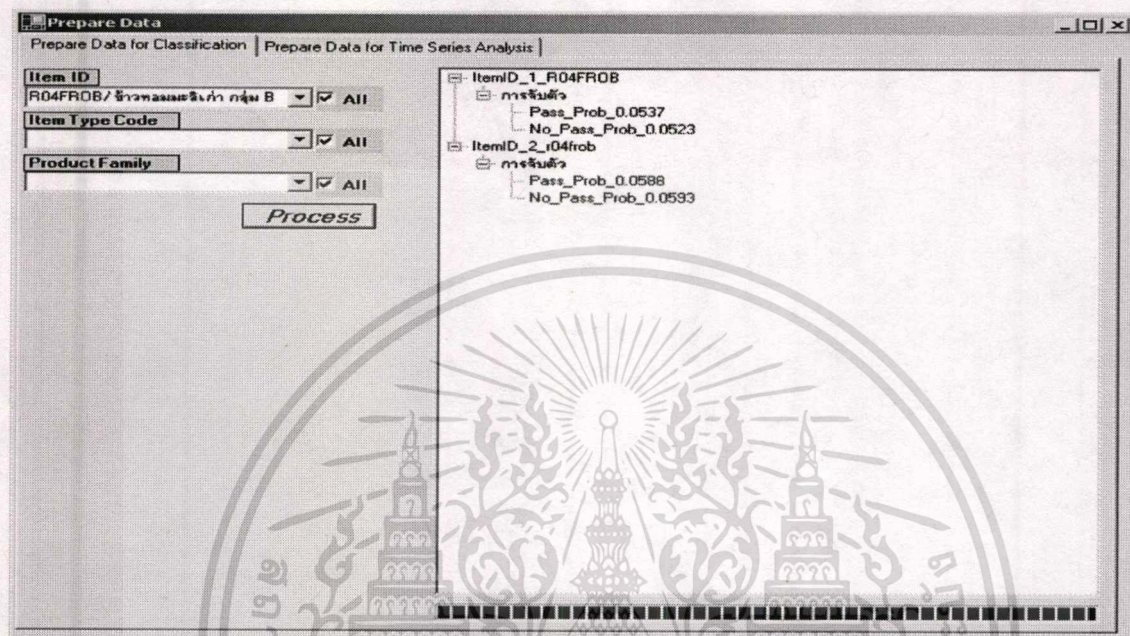
Source Criteria	Selected Criteria
1	LotNbr
2	อะไหล่
3	ความชื้น
4	ข้าวขาวปน.เต็ม
5	พื้นข้าวความยาวมากกว่า7มม
6	พื้นข้าวความยาวน้อยกว่า6.2มม
58	Test
10	ข้าวหัก4.3mm
11	ข้าวหัก4.2mm
12	ปลายข้าว
13	เส้นแฉก
14	ท้องไร่
15	ข้าวเสีย
16	เมล็ดเหลือง
17	ข้าวเหนียว
18	เมล็ดสีปนเมล็ดอ่อนและวัตถุอื่นๆ
19	สี
20	กลิ่น
21	ขนาดเมล็ด
22	การจับตัว
23	เนื้อสัมผัส
24	กายสมบัติโดยรวม
25	ข้าวข้าวมัน
26	ข้าวหัก3.42mm.
27	ข้าวหัก4.0mm.
28	ข้าวขาวปน.สายในต่าง
29	เมล็ดเร็ก
30	เมล็ดเขียว

Accept | Cancelled | Close

รูปที่ 5.5 หน้าจอการเลือกกฎ

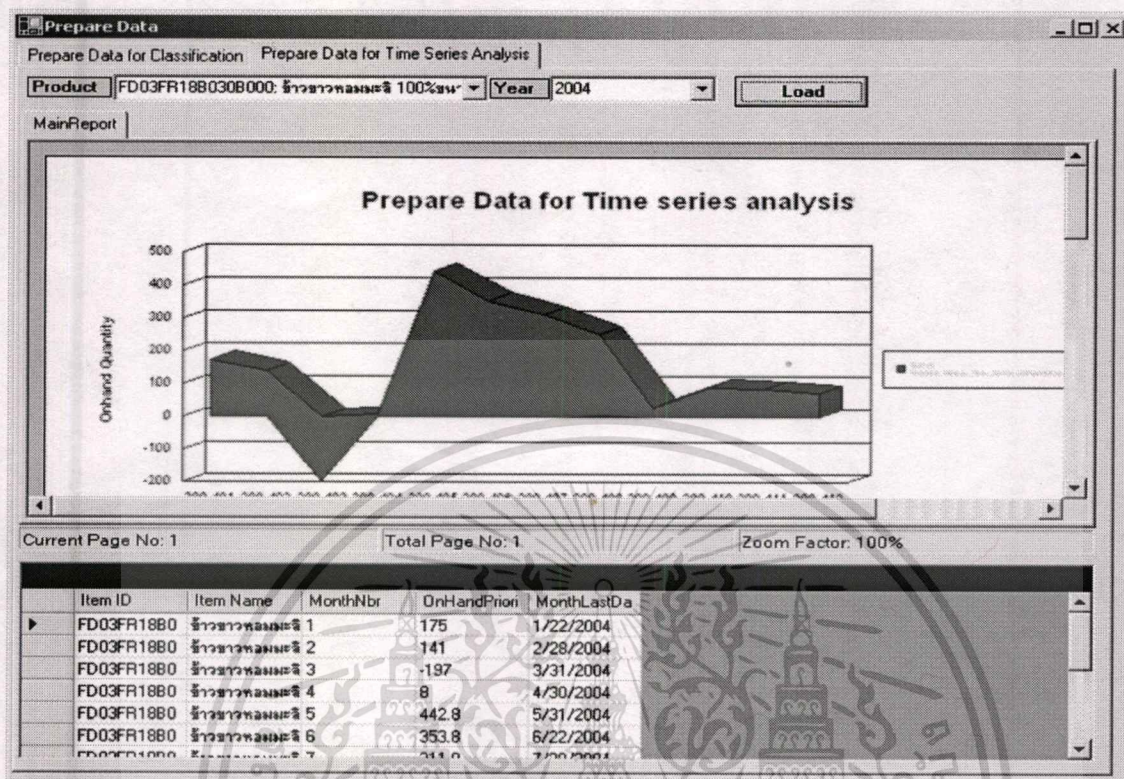
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หน้าจอการเลือกกฎที่ต้องการใช้ในการจัดกลุ่มข้อมูล โดยจะเลือกได้จากกฎที่สร้างขึ้นในหน้าจอแรกเพื่อนำมาใช้ในการพิจารณาในการจำแนกข้อมูลและการพยากรณ์



รูปที่ 5.6 หน้าจอเตรียมข้อมูลเพื่อจัดกลุ่ม โดยใช้กฎที่เลือก

หน้าจอแสดงการเตรียมข้อมูลในการสร้างแบบจำลองการแบ่งกลุ่มข้อมูล หลังจากเลือกรหัสสินค้าแล้วกดปุ่ม Process โปรแกรมจะทำการประมวลผลเพื่อหาค่าความน่าจะเป็นของผลลัพธ์ทั้งผ่านและไม่ผ่านของกฎที่เลือกแต่ละข้อ โดยทำการประมวลผลจากข้อมูลที่น่ามาทำการเทรนนิ่งของรหัสสินค้าที่เลือก ผลลัพธ์ที่ได้จะเป็นตารางแสดงความน่าจะเป็นของแต่ละปัจจัยว่ามีความน่าจะเป็นที่จะผ่านและไม่ผ่านการตรวจวัดคุณภาพของวัตถุดิบเป็นอัตราส่วนเท่าไร



รูปที่ 5.7 หน้าจอการเตรียมข้อมูลสำหรับการพยากรณ์เชิงอนุกรมเวลา

หน้าจอแสดงยอดขายของสินค้าในแต่ละช่วงเวลาที่กำหนดโดยผู้ใช้ โดยผู้ใช้ทำการเลือกสินค้า แล้วเลือกปีเพื่อดูยอดขายในแต่ละเดือนภายในปีนั้น หลังจากนั้นทำการเลือกเดือนเพื่อดูยอดขายภายในเดือน เมื่อกดปุ่ม Load จะได้ข้อมูลเพื่อนำไปพยากรณ์โดยใช้ การวิเคราะห์เชิงอนุกรมเวลาซึ่งข้อมูลที่ได้สามารถอธิบายได้ดังนี้

Item ID: รหัสสินค้า

Item Name: ชื่อของสินค้า

Month Nbr: ลำดับของเดือนภายในปีนั้น

OnhandPriorityofMonth: ปริมาณสินค้าคงคลังในแต่ละสิ้นเดือน

MonthLastDay: วันสุดท้ายที่เกิดการเคลื่อนไหวของยอดของสินค้าคงคลังในแต่ละเดือน

**Classification Result**

Result of Classification Model

Purchase Order ID: 03003/0031

Purchase Order Line Number: 1

Lot Number: 0202/0031

Prediction

PD Number	Line Number	Lot Number	Item ID	Item Name	Qty Inspected	Criteria	QC Value
03003/0031	1	0202/0031	R04WHQA	ข้าวขาวเก่า กลุ่มA	2,000	[ข้าวขาวปน:ตั้ง]	40.0000
03003/0031	1	0202/0031	R04WHQA	ข้าวขาวเก่า กลุ่มA	2,000	[ข้าวหัก 4.6 mm.]	0.9000
03003/0031	1	0202/0031	R04WHQA	ข้าวขาวเก่า กลุ่มA	2,000	[ข้าวหัก 5.2 mm.]	1.8100
03003/0031	1	0202/0031	R04WHQA	ข้าวขาวเก่า กลุ่มA	2,000	[ข้าวหัก 3.42 mm.]	7.0000
03003/0031	1	0202/0031	R04WHQA	ข้าวขาวเก่า กลุ่มA	2,000	[พื้นข้าวความยาวน้อยกว่า]	0.4000
03003/0031	1	0202/0031	R04WHQA	ข้าวขาวเก่า กลุ่มA	2,000	[พื้นข้าวความยาวมากกว่า]	3.2000
03003/0031	1	0202/0031	R04WHQA	ข้าวขาวเก่า กลุ่มA	2,000	LotNbr	12.4000
03003/0031	1	0202/0031	R04WHQA	ข้าวขาวเก่า กลุ่มA	2,000	เมล็ดสีฟ้า	8.0000

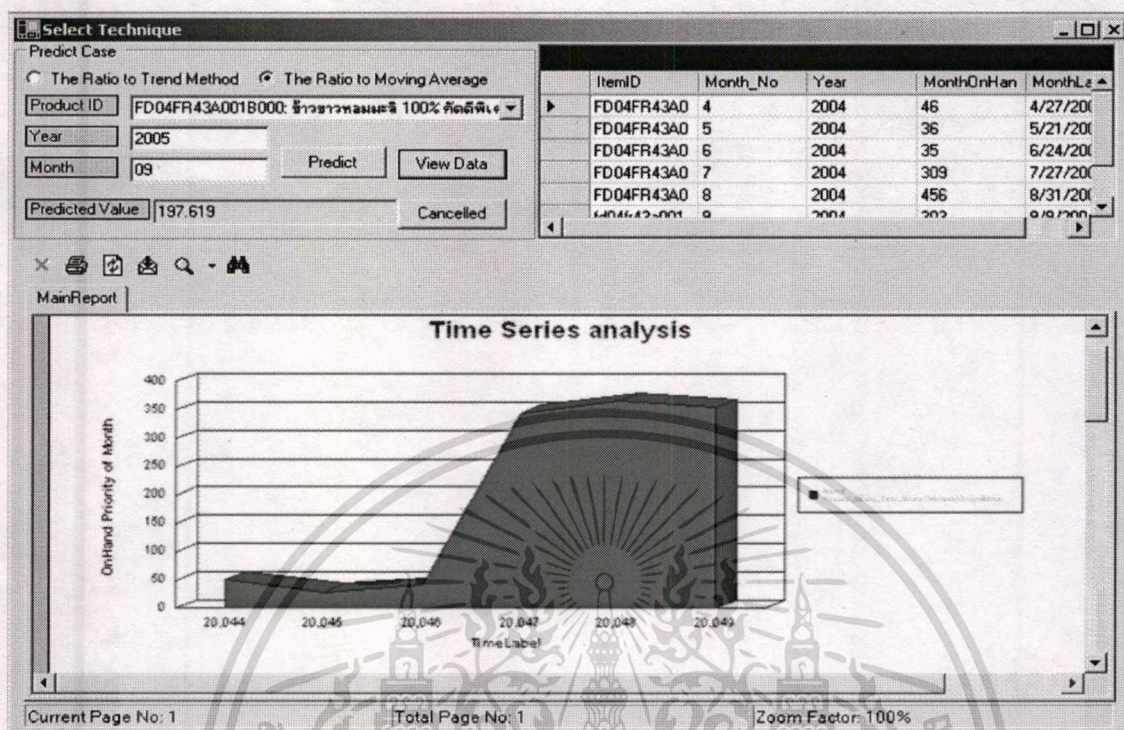
Classification Prediction

R04WHQA: ข้าวขาวเก่า กลุ่มA Pass Probability 0.6083  
R04WHQA: ข้าวขาวเก่า กลุ่มA No Pass Probability 0.0830

### รูปที่ 5.8 หน้าจอแสดงผลจากการสร้างกลุ่ม

เมื่อทำการเตรียมข้อมูลในการสร้างแบบจำลองของกฎการจำแนกข้อมูล(Classification Rule) เรียบร้อยแล้วสามารถนำแบบจำลองที่ได้มาทำการพยากรณ์ผลลัพธ์ในการตรวจสอบคุณภาพของสินค้าในแต่ละ ล็อตการรับของได้ โดยทำการเลือกใบสั่งซื้อ รายการในใบสั่งซื้อ และ ล็อตการรับของในแต่ละครั้ง แต่รหัสสินค้าของ ล็อตที่เลือกจะต้องเป็นรหัสสินค้าเดียวกับที่ทำการสร้างแบบจำลองไว้ในขั้นตอนการเตรียมข้อมูล ผลลัพธ์แสดงในส่วนของ Classification Prediction ในรูปแบบของคำอธิบายความน่าจะเป็นที่ผลลัพธ์จะออกมาผ่านและไม่ผ่านเกณฑ์ที่กำหนดไว้ในกฎที่ตั้งขึ้น

จากรูปแสดงผลการวิเคราะห์หว่าข้าวขาวกลุ่ม A มีความน่าจะเป็นที่จะผ่านการตรวจสอบคุณภาพร้อยละ 60.83 และ มีความน่าจะเป็นที่จะไม่ผ่านการตรวจสอบคุณภาพร้อยละ 8.3

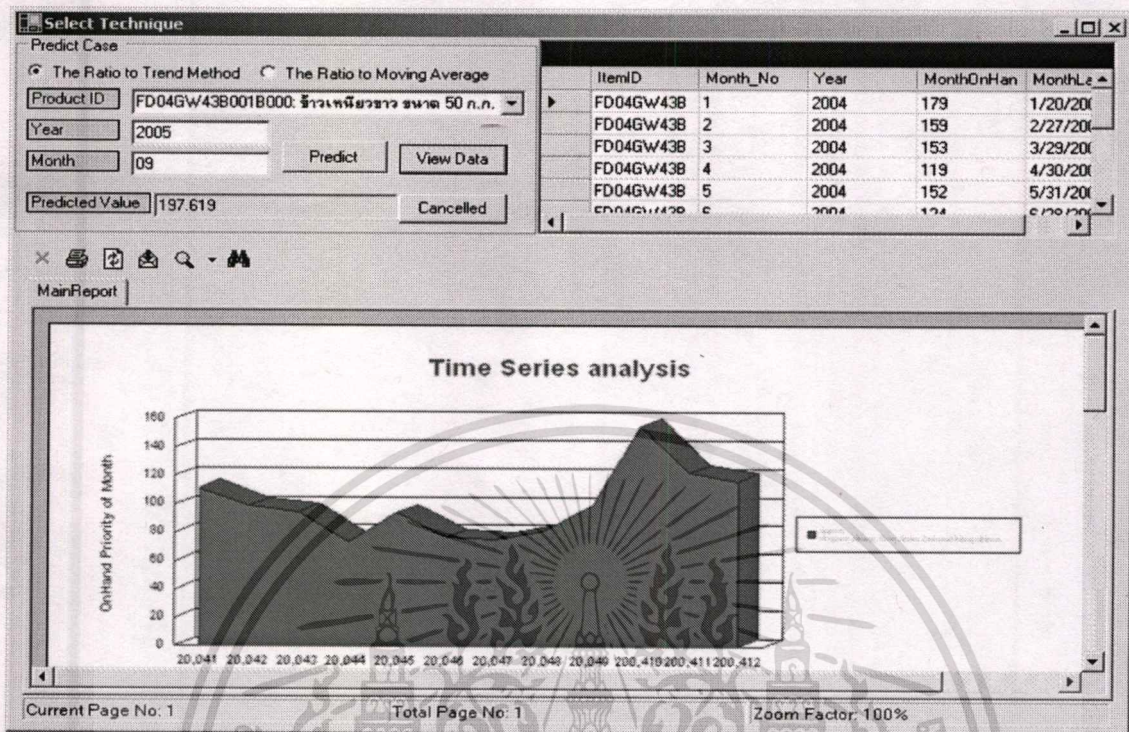


รูปที่ 5.10 หน้าจอแสดงผลการพยากรณ์เชิงอนุกรมเวลา แบบวิเคราะห์ทางโน้ม

หน้าจอสำหรับการพยากรณ์ยอดขายของสินค้าในแต่ละช่วงเวลาที่ต้องการ โดยอาศัยทฤษฎี Time Series Analysis ซึ่งเป็นวิธีการทางสถิติช่วยในการพยากรณ์ เลือกใช้ 2 เทคนิค ในการขจัดอิทธิพลของฤดูกาลที่มีผลต่อค่าแนวโน้มของข้อมูล

1. The Ratio to Trend Method
2. The Ratio to Moving Average

เพื่อใช้ในการเปรียบเทียบผลลัพธ์ว่าวิธีไหน ใกล้เคียงที่สุด



รูปที่ 5.11 หน้าจอแสดงผล

หน้าจอแสดงผลลัพธ์ของข้อมูลที่ใช้ในการสร้างแบบจำลองการพยากรณ์ด้วย Time Series Analysis และแสดงกราฟเส้นประกอบเพื่อแสดงแนวโน้มของข้อมูลเพื่อให้ผู้ใช้สามารถมองเห็นแนวโน้มของข้อมูล

จากผลลัพธ์ที่ได้ทั้ง 2 แบบ จะเห็นว่าลักษณะของกราฟจากวิธีที่ 2 จะราบเรียบกว่าวิธีแรก เพราะได้ขจัดอิทธิพลของฤดูกาลออกไปแล้ว ซึ่งเหมาะสำหรับการนำไปใช้ในการพยากรณ์

## บทที่ 6

### บทสรุป

โครงการพัฒนาโปรแกรมประยุกต์สำหรับช่วยในการบริหารสินค้าคงคลังโดยใช้ Naïve Bays Data Mining Classification ในการจำแนกประเภทของข้อมูล และ Time Series Analysis ในการพยากรณ์ข้อมูลในช่วงเวลาต่างๆนี้เป็น โครงการที่จัดทำขึ้นเพื่อนำเสนอให้เห็นถึงประโยชน์ของการนำเอาทฤษฎีาค่าไมน์นิ่งมาใช้เพิ่มประสิทธิภาพในการบริหารสินค้าคงคลัง ทำให้สามารถวิเคราะห์คุณลักษณะของวัตถุดิบที่ไม่ผ่านการตรวจสอบคุณภาพ โดยสามารถเน้นตรวจสอบบางคุณสมบัติเบื้องต้นก่อนการรับสินค้าทำให้การคืนของเสียทำได้รวดเร็วขึ้น และแจ้งเตือนผู้จัดจำหน่ายว่าควรคัดเลือกสินค้าให้มีคุณลักษณะเช่นไรจึงจะสามารถผ่านการตรวจสอบคุณภาพเพื่อป้องกันปัญหาสินค้าขาดเนื่องจากปัญหาการคืนของเสีย และด้วยเทคนิคของ Time Series Analysis ช่วยให้การคาดคะเนประมาณสินค้าภายในคลังสินค้ามีความแม่นยำมากยิ่งขึ้น ทำให้สามารถพยากรณ์ยอดการผลิตสินค้าให้เพียงพอต่อความต้องการของผู้บริโภคและป้องกันไม่ให้มีสินค้าค้างในคลังสินค้ามากเกินไป

#### 6.1 สรุปหลักการที่ใช้ในระบบ

1. เทคนิคที่ใช้ในการหาคุณลักษณะของวัตถุดิบที่มีคุณภาพสามารถผ่านการตรวจวัดคุณภาพได้นำเอาเทคนิคของการจัดกลุ่ม มาใช้ในการพัฒนาระบบซึ่งเป็นเทคนิคของการสร้างรูปแบบการทำนาย โดยอาศัยประสบการณ์การเรียนรู้หรือความสนใจในบางสิ่งบางอย่างที่จะเป็นหลักสำคัญในการตัดสินใจ โดยอัลกอริทึมที่เลือกใช้คือ Naïve Bayes Classification ซึ่งเป็นกระบวนการเรียนรู้แบบมีรูปแบบของอินพุตและเอาท์พุทไว้ก่อนแล้วจากข้อมูลการตรวจสอบคุณภาพของวัตถุดิบจากนั้นเอาข้อมูลที่ได้มาเลือกหาปัจจัยที่คาดว่าจะมีผลต่อการตรวจสอบคุณภาพ มาทำการเทรนนิ่ง เพื่อให้ได้แบบจำลองในการจัดกลุ่ม หลังจากที่ได้แบบจำลองแล้วจึงนำเอาแบบจำลองที่ได้ไปทำการทดสอบกับข้อมูลอื่นๆ ที่ไม่ใช่ข้อมูลในการสร้างแบบจำลอง เพื่อทำการทดสอบว่าแบบจำลองที่ได้มีความถูกต้องหรือไม่

2. เทคนิคที่ใช้ในการพยากรณ์ประมาณคงค้างของสินค้าสำเร็จรูปภายในคลังสินค้า โดยอาศัยเทคนิคทางสถิติในส่วนของ Time Series Analysis ซึ่งเป็นอีกรูปแบบหนึ่งของการทำค่าไมน์นิ่งที่ช่วย

ในการพยากรณ์โดยอาศัยข้อมูลที่มีอยู่เป็นจำนวนมากทั้งในอดีตจนถึงปัจจุบันนำมาสร้างแบบจำลองเพื่อคำนวณหาค่าที่ต้องการพยากรณ์ของช่วงเวลาต่างๆในอนาคต

## 6.2 สรุปกระบวนการในการทำงาน

6.2.1 การจำแนกประเภทข้อมูล เริ่มต้นจากการเลือกกฎหรือปัจจัยที่คาดว่าจะมีผลต่อผลลัพธ์ของการตรวจสอบคุณภาพของวัตถุดิบ จากนั้นจึงนำข้อมูลมาทำความสะอาด พร้อมทั้งแปลงข้อมูลให้อยู่ในรูปแบบที่แบ่งกลุ่มไว้ แล้วจึงนำเอาข้อมูลเหล่านั้นไปทำการเทรนนิ่งเพื่อให้ได้ผลลัพธ์ออกมาเป็นแบบจำลอง แล้วแสดงผลในรูปแบบของตารางแสดงความน่าจะเป็น นอกจากนั้นส่วนงานการทำนายสามารถนำเอาข้อมูลการรับเข้าวัตถุดิบ Lot ใหม่ที่ยังไม่สิ้นสุดกระบวนการตรวจสอบคุณภาพเพื่อหาคำตอบว่าวัตถุดิบดังกล่าวจะผ่านการตรวจวัดคุณภาพหรือไม่

6.2.2 การพยากรณ์ข้อมูล เริ่มจากการรวบรวมและกลั่นกรองข้อมูลเพื่อนำมาหาสมการทางโมเดลที่ขจัดอิทธิพลของฤดูกาลออกแล้ว นำมาใช้ในการคำนวณหาค่าที่ต้องการพยากรณ์ของช่วงเวลาต่างๆในอนาคต โดยแสดงผลเป็นการสรุปยอดสินค้าในแต่ละเดือนและนำเสนอเป็นกราฟของข้อมูลที่ขจัดอิทธิพลของฤดูกาลออกแล้ว ช่วยเพิ่มประสิทธิภาพในการประมาณการปริมาณของวัตถุดิบที่ควรจะเป็นในอนาคต

## 6.3 สรุปผลการทดสอบ

เมื่อนำเอาผลจากการเทรนนิ่งไปทดสอบกับข้อมูลจริงพบว่าผลการทำนายว่าวัตถุดิบในการรับแต่ละครั้งจะผ่านหรือไม่ผ่านการตรวจวัดคุณภาพหรือไม่มีความถูกต้องร้อยละ 79 โดยองค์กรสามารถนำความรู้ที่ได้ไปใช้ในการช่วยตรวจรับวัตถุดิบให้มีประสิทธิภาพมากยิ่งขึ้นส่งผลให้การบริหารสินค้าคงคลังมีประสิทธิภาพมากยิ่งขึ้น

การพยากรณ์ปริมาณของสินค้าคงคลังของเดือนถัดไปต่อจากเดือนสุดท้ายของข้อมูลที่นำมาใช้สร้างสมการแบบจำลองยังไม่ใกล้เคียงกับความเป็นจริง ทั้งนี้อาจเป็นเพราะเกิดจากปัจจัยอื่นที่ไม่ใช่อิทธิพลของฤดูกาลเช่น ภาวะเศรษฐกิจ การเมือง หรือวัฏจักรอื่นๆ

## 6.4 ข้อเสนอแนะและแนวทางในการพัฒนาต่อไป

จากระบบเดิมใช้เพื่อการตรวจสอบและแบ่งกลุ่มของวัตถุดิบจะทำให้สามารถพยากรณ์ผลลัพธ์ได้จากคุณลักษณะต่างๆของวัตถุดิบนั้นๆ ซึ่งเมื่อมีการเก็บข้อมูลมากขึ้นเราสามารถที่จะพัฒนาไปสู่การตรวจวัดคุณภาพของวัตถุดิบได้จากคุณลักษณะต่างๆของผู้จัดจำหน่าย หรือสร้างแบบจำลอง

เพื่อวิเคราะห์หาปัจจัยที่ส่งผลให้เกิดของเสียจากการผลิตสินค้าสำเร็จรูปซึ่งจะช่วยให้การคลังสินค้า รวมถึงการวางแผนการผลิตมีประสิทธิภาพมากยิ่งขึ้น

จากการพยากรณ์ปริมาณสินค้าคงคลังโดย Time Series Analysis ในอนาคตควรเพิ่มการวิเคราะห์ปัจจัยอื่นเข้าไปด้วยเพื่อให้ครอบคลุมทุกปัจจัย และให้ผลลัพธ์ที่ถูกต้องแม่นยำมากยิ่งขึ้น



## บรรณานุกรม

สมจิต วัฒนาชยากุล, 2532. สถิติเพื่อการวิเคราะห์. กรุงเทพฯ: ประชาชนพริก.

Two Crows Corporation. 2000 **Introduction to Data Mining and Knowledge Discovery**.

[Online]. Available : <http://www.twocrows.com/KDMining.html>.

Berry, Michael and Gordon, Linoff. 2000. **Data Mining Techniques**. [online].

Available : <http://www.data-miners.com/services/consulting.html>.

Weiss, Sholom and Indurkha, Nitin. 1998. **Predictive Data Mining A Practical Guide**. [Online].

Available : <http://www.data-miners.com/services/brochure-churn-modeling.html>.

Wipro Technologies. 2001. **Mining for Gold**. [Online].

Available : <http://www.wipro.com/prodesign/ieee1394/index.html>.

Hahsler Michael. 2001. Knowledge Management: Data Warehouses and Data Mining. [Online].

Available: <http://www.wi.wu.ac.at/index.html>.

Perisic, Igor and Posse, Christian. 2002. **Combining Advanced Dynamic Semantic Profiling with Data Mining Technologies to Create Unified Enterprise Information Management Solutions**. [Online]. Available : <http://www.DMReview.com/portals/DBMiner.pdf>.

Anjali, Dhond and Amar, Gupta and Sanjeev, Vadhavkar. 2000. Data Mining Techniques for Optimizing Inventories for Electronic Commerce. [Online].

Available : <http://www.DMReview.com/portals/CaseMiner.pdf>.

Chen, Chung-Min and Cochinwala, Munir and Yueh, Elsa. 1999. **A Case Study on Inventory Data Warehousing**. [Online]. Available :

<http://www.research.telcordia.com/demos/CaseInv.pdf>.

Han, Jiawei and Chiang, Jenny and Chee Sonny. 1999. **A System for Data Mining in Relational Databases and Data Warehouses**. [Online].

Available : <http://db.cs.sfu.ca/DBMiner/Warehouse.html>.

Tagbo Kingsley. 2004. **Naïve Bayes Data Mining Classifiers**.

[Online]. Available : <http://www.kdkeys.net/forums/4284/ShowPost.aspx>.

## ประวัติผู้เขียน

ชื่อผู้เขียน นายณรงค์ฤทธิ์ อนันต์ชัยพัทธนา  
 วันเดือนปีเกิด 6 พฤศจิกายน พ.ศ. 2521  
 สถานที่เกิด เชียงราย  
 อีเมล [narongrit6@hotmail.com](mailto:narongrit6@hotmail.com)

### ประวัติการศึกษา

- สำเร็จการศึกษาระดับประถมศึกษา โรงเรียนอนุบาลเชียงราย
- เข้ารับการศึกษาระดับมัธยมศึกษา โรงเรียนสามัคคีวิทยาคม เชียงราย
- สำเร็จการศึกษามัธยมศึกษา จากโรงเรียนเตรียมทหาร
- สำเร็จการศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยธรรมศาสตร์
- กำลังศึกษาระดับปริญญาโท คณะเทคโนโลยีสารสนเทศ สาขาเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

### ประวัติการทำงาน

- บริษัทเอ็ม โฟกัสจำกัด (ปี 2545) ตำแหน่ง Programmer