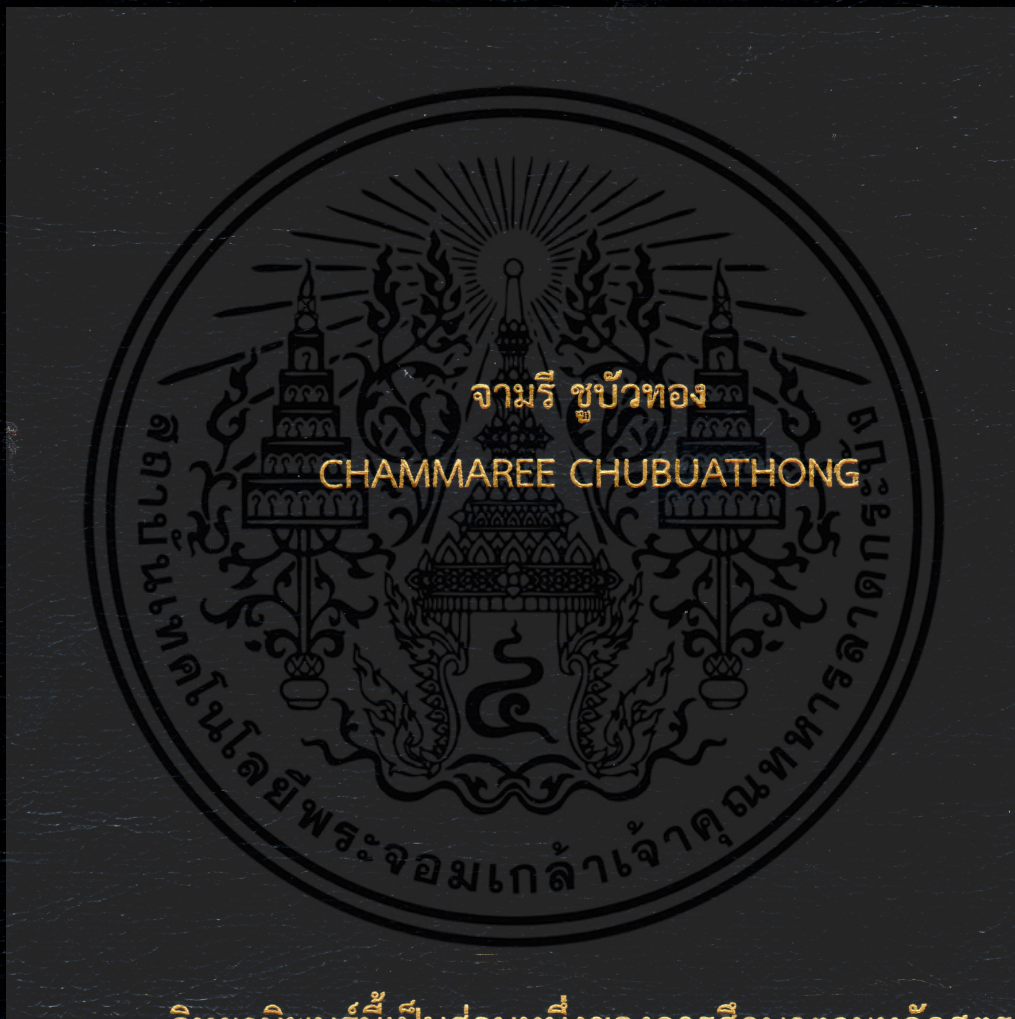


การพัฒนาตัวแบบเพื่อพยากรณ์คุณภาพผลิตภัณฑ์ฮาร์ดดิสก์โดยใช้การ
วิเคราะห์เหมืองข้อมูล

MODEL DEVELOPMENT TO PREDICT QUALITY OF HARD DISK
DRIVE USING DATA MINING



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์
ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2559

KMITL-2016-SC-M-050-032

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การพัฒนาตัวแบบเพื่อพยากรณ์คุณภาพผลิตภัณฑ์ฮาร์ดดิสก์โดยใช้การ

วิเคราะห์เหมืองข้อมูล

MODEL DEVELOPMENT TO PREDICT QUALITY OF HARD DISK

DRIVE USING DATA MINING



T143999

จามรี ชুবัวทอง

CHAMMAREE CHUBUATHONG

๒๐๐ ๒๖๖ ๒๔๖

.b.....
.f.....

สาขา.....
เลขทะเบียน..... 143999
รับเดือนปี 10 ต.ค. 2559

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์

ภาควิชาสถิติ คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2559

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับ KMITL-2016-SC-M-050-032 อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

MODEL DEVELOPMENT TO PREDICT QUALITY OF HARD DISK
DRIVE USING DATA MINING



A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE
DEGREE OF MASTER OF SCIENCE IN APPLIED STATISTICS
DEPARTMENT OF STATISTICS
FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2016

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับก **KMITL-2016-SC-M-050-032** อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2016

FACULTY OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์

“การพัฒนาตัวแบบเพื่อพยากรณ์คุณภาพผลิตภัณฑ์ฮาร์ดดิสก์โดยใช้การวิเคราะห์เหมืองข้อมูล”

(MODEL DEVELOPMENT TO PREDICT QUALITY OF HARD DISK DRIVE USING DATA MINING)

ชื่อนักศึกษา

นางสาวจามรี ชูบัวทอง

รหัสประจำตัว

57605097

ปริญญา


วิทยาศาสตรมหาบัณฑิต (สาขาสถิติประยุกต์)

ภาควิชา

สถิติ

อาจารย์ที่ปรึกษาวิทยานิพนธ์

ผศ.ดร.สมศรี บัณฑิตวิไล

คณะกรรมการสอบวิทยานิพนธ์	ลายมือชื่อ
ดร.บุญญสิทธิ์ วรจันทร์ ประธานกรรมการ รศ.ดร.วัลย์ลักษณ์ อัครีรวงศ์ อาจารย์บัณฑิตประจำ (ในสาขาวิชาที่เกี่ยวข้อง) ผศ.ดร.รวิวรรณ เทนอิสสระ ผู้ทรงคุณวุฒิจากภายนอกสถาบันฯ ผศ.ดร.สมศรี บัณฑิตวิไล อาจารย์ที่ปรึกษาวิทยานิพนธ์	 คณบดี คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง 

วัน/ เดือน/ ปี ที่สอบ 15 กรกฎาคม พ.ศ. 2559 เวลา 13.00 - 16.00 น.

สถานที่สอบ ณ ห้อง 115 ตึกจุฬารกรณ์ 1

คณะวิทยาศาสตร์รับรองแล้ว
(รองศาสตราจารย์ ดร.ตชนันท์ ธนะบริพัฒน์)
คณบดี คณะวิทยาศาสตร์
วันที่... ๑๖...เดือน... พ.ศ. ๕๙

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การพัฒนาตัวแบบเพื่อพยากรณ์คุณภาพผลิตภัณฑ์ฮาร์ดดิสก์โดยใช้การวิเคราะห์เหมืองข้อมูล
ชื่อนักศึกษา	นางสาวจามรี ชูบัวทอง
รหัสนักศึกษา	57605097
ปริญญา	วิทยาศาสตรมหาบัณฑิต (สาขาสถิติประยุกต์)
ภาควิชา	สถิติ
พ.ศ.	2559
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผศ. ดร. สมศรี บัณฑิตวิไล

บทคัดย่อ

ในยุคที่มีการแข่งขันทางด้านเทคโนโลยีสูง บริษัทผลิตฮาร์ดดิสก์จึงต้องมีการคิดค้นวิธีการหรือกระบวนการผลิตฮาร์ดดิสก์ เพื่อให้ได้ฮาร์ดดิสก์ที่มีคุณภาพดีขึ้น อีกวิธีหนึ่งซึ่งใช้กันอย่างแพร่หลายคือการลดจำนวนของเสีย งานวิจัยนี้เป็นการพัฒนาตัวแบบเพื่อพยากรณ์คุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ในกระบวนการทดสอบความเชื่อมั่นก่อนการส่งมอบแก่ผู้บริโภค ซึ่งช่วยลดการสูญเสียทรัพยากร เช่น ลดต้นทุนและเวลาในการทดสอบ เป็นต้น เนื่องจากจำนวนข้อมูลในการทำงานวิจัยนี้มีขนาดใหญ่จึงใช้การวิเคราะห์เหมืองข้อมูลเข้ามาช่วย และแบบจำลองที่ใช้ในการพยากรณ์คุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ที่พิจารณามีด้วยกัน 2 แบบคือ การวิเคราะห์การถดถอยโลจิสติก และโครงข่ายประสาทเทียม โดยมีจำนวนตัวแปรอิสระทั้งหมดคือ 197 ตัวแปรซึ่งเป็นตัวแปรเชิงปริมาณทั้งหมดเพื่อลดจำนวนตัวแปรอิสระลงจึงทำการทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระกับคุณภาพฮาร์ดดิสก์ พบว่ามีตัวแปรอิสระเพียง 25 ตัว ที่มีความสัมพันธ์กับคุณภาพของฮาร์ดดิสก์ จึงนำตัวแปรอิสระทั้ง 25 ตัวไปสร้างแบบจำลองโครงข่ายประสาทเทียม ส่วนแบบจำลองโลจิสติกจะต้องกำจัดปัญหาตัวแปรอิสระที่มีความสัมพันธ์กันเองออก คงเหลือตัวแปรที่ใช้ในการสร้างแบบจำลอง 17 ตัวแปร จากงานวิจัยพบว่าแบบจำลองโครงข่ายประสาทเทียมให้เปอร์เซ็นต์ความถูกต้อง 95.98% ส่วนแบบจำลองการวิเคราะห์การถดถอยโลจิสติกให้เปอร์เซ็นต์ความถูกต้อง 72.37% ดังนั้น แบบจำลองที่ได้จากโครงข่ายประสาทเทียมให้ความถูกต้องในการพยากรณ์คุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ได้ดีกว่าแบบจำลองที่ได้จากการวิเคราะห์การถดถอยโลจิสติก

คำสำคัญ : เหมืองข้อมูล โครงข่ายประสาทเทียม การถดถอยโลจิสติก

Thesis Title Model Development to Predict Quality of Hard Disk Drive Using
Data Mining

Student Name Miss Chammaree Chubuathong

Student ID 57605097

Degree Master of Science in Applied Statistics

Department Statistics

Year 2016

Thesis AdvisorAsst. Prof. Dr. Somsri Banditvilai

Abstract

In a highly competitive technology era, the company that produces hard disk drive must be considered about inventing a new method or process for producing hard disk drive with higher quality. Another method, which is widely used, is to reduce waste. This thesis is developed model to predict the quality of hard disk drive in an outgoing reliability test process. This reduces the loss of resources, such as cost reduction and testing time. The amount of data in this research is huge, therefore 2 data mining techniques which are logistic regression analysis and neural network are employed in this study. There are 197 quantitative independent variables. In order to reduce the independent variables, the relationship between each independent variable and the quality of hard disk drive was tested and found that only 25 variables are related to the quality of hard disk drive. Therefore, 25 independent variables were used to build the neural network. Since logistic regression analysis need to eliminate independent variables that has multicollinearity problem. Then it was left 17 variables to build logistic regression analysis. The result shows that neural network model has 95.98% accuracy and logistic regression model has 72.37% accuracy. Therefore, neural network model performs better prediction in hard disk drive quality than logistic regression analysis model.

Keywords: Data Mining, Neural Network, Logistic Regression

กิตติกรรมประกาศ

การจัดทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยบุคคลและการสนับสนุนจากหลายฝ่าย ขอกราบขอบพระคุณอาจารย์ที่ปรึกษา ผศ.ดร.สมศรี บัณฑิตวิไล ประธานกรรมการสอบวิทยานิพนธ์ ดร.บุญญสิทธิ์ วรจันทร์ กรรมการสอบวิทยานิพนธ์ รศ.ดร.วลัยลักษณ์ อัครีรวงศ์ และ ผศ.ดร.รวิวรรณ เทนอิสสระ ที่กรุณาให้ความรู้และคำแนะนำที่ดีในกระบวนการทำวิทยานิพนธ์ฉบับนี้ รวมทั้งกรุณาช่วยแก้ไขและตรวจสอบวิทยานิพนธ์นี้ให้สำเร็จลุล่วงไปด้วยดี

ขอขอบพระคุณหัวหน้างานของข้าพเจ้าที่ให้การสนับสนุนในด้านการศึกษาและการทำวิทยานิพนธ์ รวมไปถึงบริษัทที่ข้าพเจ้าทำงานด้วยที่อนุเคราะห์ให้ใช้โปรแกรม SAS Enterprise Miner Workstation และขอขอบคุณบริษัทกรณีศึกษาที่อนุเคราะห์ข้อมูลในการทำวิทยานิพนธ์ฉบับนี้

สุดท้ายขอขอบคุณครอบครัวที่ให้กำลังใจและให้การสนับสนุนในด้านการการศึกษาหาวิชาความรู้ รวมถึงขอขอบคุณ เพื่อน ๆ พี่ ๆ น้อง ๆ ที่คอยเป็นกำลังใจและห่วงใยเสมอมา

จามรี ชูบัวทอง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	ช
สารบัญรูป.....	ซ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	2
1.3 ขอบเขตของงานวิจัย.....	3
1.4 เกณฑ์การตัดสินใจ.....	5
1.5 ขั้นตอนการวิจัย.....	5
1.6 ประโยชน์ที่คาดว่าจะได้รับ.....	7
1.7 นิยามศัพท์.....	7
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	9
2.1 การวิเคราะห์การถดถอยโลจิสติก.....	9
2.1.1 การจำแนกชนิดของการวิเคราะห์การถดถอยโลจิสติก.....	9
2.1.2 ประเภทของการวิเคราะห์การถดถอยโลจิสติก.....	10
2.1.3 วัตถุประสงค์ของการวิเคราะห์การถดถอยโลจิสติก.....	10
2.1.4 ข้อตกลงเบื้องต้นการวิเคราะห์การถดถอยโลจิสติก.....	10
2.1.5 เหตุผลที่ใช้การวิเคราะห์การถดถอยโลจิสติก	
แทนการวิเคราะห์การถดถอยเชิงเส้น.....	11
2.1.6 โมเดลการวิเคราะห์การถดถอยโลจิสติก.....	11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
2.1.7 การประมาณค่าโมเดลโลจิสติก.....	12
2.1.8 การทดสอบค่าพารามิเตอร์ในแบบจำลองโลจิสติก.....	13
2.2 โครงข่ายประสาทเทียม.....	14
2.2.1 เพอร์เซพตรอน.....	15
2.2.2 การเรียนรู้โครงข่ายประสาทเทียมแบบเพอร์เซพตรอน.....	16
2.2.3 โครงข่ายประสาทเทียมแบบหลายชั้น.....	18
2.2.4 การเรียนรู้ของโครงข่ายประสาทเทียม.....	20
2.2.5 ข้อดีและข้อเสียของเครือข่ายประสาทเทียม.....	22
2.3 การวิเคราะห์องค์ประกอบหลัก.....	22
2.4 การทดสอบโคสแควร์.....	23
2.4.1 วัตถุประสงค์ของการทดสอบโคสแควร์.....	23
2.4.2 วิธีของเพียร์สัน.....	24
2.5 งานวิจัยที่เกี่ยวข้อง.....	25
บทที่ 3 วิธีการดำเนินงานวิจัย.....	27
3.1 ลักษณะของปัญหา.....	27
3.2 การเตรียมข้อมูล.....	27
3.3 การแบ่งชุดข้อมูล.....	28
3.4 การทดสอบความสัมพันธ์ระหว่างตัวแปร.....	29
3.4.1 ตรวจสอบปัญหาตัวแปรอิสระมีความสัมพันธ์กันเอง.....	30
3.5 การสร้างแบบจำลอง.....	33
3.5.1 สร้างแบบจำลองการวิเคราะห์การถดถอยโลจิสติก.....	33
3.5.2 สร้างแบบจำลองโครงข่ายประสาทเทียม.....	33
3.6 ประเมินความถูกต้อง.....	36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
3.7 พัฒนาโปรแกรมทดสอบและประยุกต์ใช้ตัวแบบ.....	36
บทที่ 4 ผลการวิจัย	37
4.1 ผลการวิเคราะห์	37
4.1.1 ผลการพยากรณ์ของแบบจำลองการวิเคราะห์การถดถอยโลจิสติก.....	37
4.1.2 ผลการพยากรณ์ของแบบจำลองการวิเคราะห์โครงข่ายประสาทเทียม	39
4.1.3 การเปรียบเทียบผลการพยากรณ์.....	39
4.2 สร้างโปรแกรมการทดสอบและประยุกต์ใช้ตัวแบบ	40
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	42
5.1 สรุปผลการวิจัย	42
5.2 ข้อเสนอแนะ	42
เอกสารอ้างอิง	44
ภาคผนวก.....	46
ภาคผนวก ก.....	47
ภาคผนวก ข	57
ประวัติผู้เขียน.....	84

สารบัญตาราง

ตารางที่	หน้า
2.1 ผลการทำนายการเกิดเหตุการณ์.....	13
2.2 ชุดข้อมูลแบบการจัดหมวดหมู่.....	15
3.1 ลักษณะอาการผิดปกติของฮาร์ดดิสก์ในหมวดหมู่ต่างๆในจำนวนข้อมูล 8 เดือน.....	27
3.2 สัดส่วนจำนวนข้อมูลของชุดข้อมูลการเรียนรู้ ชุดทดสอบ และชุดทดสอบ.....	29
3.3 ค่าโคสแควร์และ P-Value ที่ได้จากการทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระ กับคุณภาพของฮาร์ดดิสก์.....	30
3.4 ความสัมพันธ์ของตัวแปรอิสระทั้ง 25 ตัวแปร.....	31
3.5 ตัวแปรอิสระใหม่ที่สร้างจากตัวแปรอิสระที่มีความสัมพันธ์กัน ด้วยวิธีการวิเคราะห์องค์ประกอบหลัก.....	32
3.6 ความสัมพันธ์ของตัวแปรอิสระจำนวน 17 ตัวแปรหลังจากแก้ปัญหา Multicollineality.....	33
3.7 ผลลัพธ์การปรับค่าจำนวนหน่วยช้อนของโครงข่ายประสาทเทียม.....	34
4.1 ผลการพยากรณ์คุณภาพของฮาร์ดดิสก์ด้วยวิธีการวิเคราะห์การถดถอยโลจิสติก ของชุดเรียนรู้ ชุดทดสอบ และชุดทดสอบ.....	37
4.2 ตัวแปรอิสระที่ถูกเลือกโดยการวิเคราะห์การถดถอยโลจิสติกแบบเป็นขั้นตอน.....	38
4.3 ผลการพยากรณ์คุณภาพของฮาร์ดดิสก์ด้วยวิธีโครงข่ายประสาทเทียม ของชุดเรียนรู้ ชุดทดสอบ และชุดทดสอบ.....	39
4.4 ประสิทธิภาพการพยากรณ์ของแบบจำลองการวิเคราะห์การถดถอยโลจิสติก และโครงข่ายประสาทเทียม.....	40
4.5 ผลการพยากรณ์คุณภาพผลิตภัณฑ์ฮาร์ดดิสก์ก่อนกระบวนการทดสอบ ความเชื่อมั่นของข้อมูลใหม่ 2 เดือน.....	41

สารบัญรูป

รูปที่	หน้า
1.1 จำนวนและเปอร์เซ็นต์อาการผิดปกติของฮาร์ดดิสก์ในกระบวนการทดสอบความเชื่อมั่น	3
1.2 กระบวนการทดสอบฮาร์ดดิสก์ก่อนเพิ่มกระบวนการพยากรณ์.....	4
1.3 กระบวนการทดสอบฮาร์ดดิสก์หลังเพิ่มกระบวนการพยากรณ์.....	4
1.4 ขั้นตอนการดำเนินงานวิจัย	6
1.5 ส่วนประกอบของฮาร์ดดิสก์.....	7
2.1 การจำแนกชนิดของการวิเคราะห์การถดถอยโลจิสติก.....	9
2.2 ฟังก์ชันโลจิสติก	11
2.3 สถาปัตยกรรมโครงข่ายประสาทเทียมเปอร์เซพตรอน.....	15
2.4 ขอบเขตการตัดสินใจโดยการประยุกต์ใช้วิธีการเรียนรู้แบบเปอร์เซพตรอน	17
2.5 ข้อมูลที่ไม่มีไฮเปอร์เพลนเชิงเส้น.....	18
2.6 ตัวอย่างโครงข่ายประสาทเทียมหลายชั้นแบบ Feed-Forward	18
2.7 ชนิดของฟังก์ชันกระตุ้นในโครงข่ายประสาทเทียม	19
2.8 ขอบเขตการตัดสินใจโดยการใช้ไฮเปอร์เพลนสองเส้นในข้อมูลนำเข้า.....	20
2.9 โครงสร้างของโครงข่ายประสาทเทียมแบบสองชั้น	20
2.10 พื้นผิวความผิดพลาดของตัวอย่างที่มีสองพารามิเตอร์	21
2.11 การแพร่ย้อนกลับด้วยเทคนิค Back Propagation.....	22
3.1 กระบวนการเตรียมข้อมูล.....	28
3.2 กราฟแสดงความสัมพันธ์ระหว่างหน่วยซ่อนและเปอร์เซ็นต์ความถูกต้องของการพยากรณ์	35
3.3 ลักษณะการแบ่งแยกข้อมูลจริงกับข้อมูลที่เกิดจากการทำนาย	36
4.1 ผลการพยากรณ์จำนวนฮาร์ดดิสก์ผิดปกติเทียบกับจำนวนฮาร์ดดิสก์ที่ผิดปกติจริง	41

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ฮาร์ดดิสก์เป็นอุปกรณ์จัดเก็บข้อมูลอิเล็กทรอนิกส์ที่มีความสำคัญมากในปัจจุบัน โดยมีหน้าที่ในการจัดเก็บข้อมูลต่างๆ บนปริมาณพื้นที่ที่เพิ่มขึ้น จาก 500 จิกะไบต์ (GB) ไปจนถึง 6 เทระไบต์ (TB) และมีแนวโน้มในการเพิ่มจำนวนความจุอย่างต่อเนื่อง ทำให้อุตสาหกรรมฮาร์ดดิสก์ (Hard Disk) ไม่ได้หยุดอยู่กับที่บนตลาดคอมพิวเตอร์เพียงอย่างเดียว มีการขยายตลาดฮาร์ดดิสก์ไปใช้ในอุปกรณ์อื่นๆ อย่างแพร่หลาย อาทิเช่น โทรศัพท์มือถือ อุปกรณ์ที่ใช้ในการเก็บข้อมูลขนาดใหญ่หรือที่เรียกว่า เซิร์ฟเวอร์/เอ็นเตอร์ไพรส์ (Server/Enterprise) เครื่องเล่นดีวีดี (DVD) โทรศัพท์ และกล้องวงจรปิด เป็นต้น สำหรับฮาร์ดดิสก์ (Hard Disk) ถูกจัดเป็นองค์ประกอบที่สำคัญไม่น้อยไปกว่าระบบคอมพิวเตอร์ และเป็นอุปกรณ์ที่ง่ายต่อการอัปเดตสเปค อาทิเช่น ความจุ ความเร็วรอบ ขนาดของหน่วยความจำแคช เป็นต้น

ในยุคที่มีการแข่งขันทางด้านเทคโนโลยีสูง บริษัทผลิตฮาร์ดดิสก์จึงต้องมีการคิดค้นวิธีการกระบวนการผลิตฮาร์ดดิสก์ เพื่อให้ได้ฮาร์ดดิสก์ที่มีคุณภาพที่ดีที่สุดในราคาถูก อีกวิธีหนึ่งซึ่งใช้กันอย่างแพร่หลายคือการลดจำนวนของเสีย ซึ่งสามารถลดต้นทุนการผลิตได้เป็นอย่างดี

การผลิตฮาร์ดดิสก์ได้ร่วโดยทั่วไปมีกระบวนการที่สำคัญ 2 ขั้นตอน คือ Head Disk Assembly (HDA) และ Printed Circuit Board Assembly (PCBA) ซึ่งส่วนแรกจะเป็นส่วนการประกอบชุดหัวอ่าน/เขียนกับแผ่นดิสก์เก็บข้อมูล และส่วนประกอบที่เกี่ยวข้องทางกลไกการทำงานในขณะที่ยานที่ผลิตฮาร์ดดิสก์จะเป็นการประกอบชิ้นส่วนทางด้านวงจรอิเล็กทรอนิกส์ที่ใช้ในการควบคุมการทำงานของฮาร์ดดิสก์ไดรฟ์ และเชื่อมต่อสัญญาณกับอุปกรณ์อื่น ๆ เช่น คอมพิวเตอร์ เป็นต้น เมื่อได้ครบทั้งสองส่วนแล้วจึงทำการประกอบเข้าด้วยกันเป็นฮาร์ดดิสก์ไดรฟ์

โดยปกติแล้วฮาร์ดดิสก์แต่ละตัวก่อนที่จะออกจากโรงงานผู้ผลิตมาสู่มือผู้บริโภค ต้องผ่านกระบวนการต่างๆ เหล่านี้คือ กระบวนการผลิต และกระบวนการทดสอบที่ยุ่ยากซับซ้อน เพื่อให้ได้ผลิตภัณฑ์ที่มีคุณภาพมากที่สุด ซึ่งกระบวนการทดสอบยังแบ่งออกเป็น กระบวนการทดสอบระหว่างการประกอบฮาร์ดดิสก์ (Assembly Process) กระบวนการทดสอบหลังประกอบฮาร์ดดิสก์ (Backend Test : BE) กระบวนการทดสอบคุณภาพ (Final Quality Audit : FAQ) และกระบวนการทดสอบความเชื่อมั่นก่อนการส่งมอบให้ผู้บริโภค (Outgoing Reliability Testing : ORT) อย่างไรก็ตามในการผลิตฮาร์ดดิสก์ในกระบวนการต่างๆ อาจมีขั้นตอนหรือชิ้นส่วนบางอย่างที่มีการทำงานผิดพลาดซึ่งทำให้ฮาร์ดดิสก์มีคุณภาพไม่ตรงตามความต้องการส่งผลให้เกิดเป็นของเสีย (Failure) ในกระบวนการผลิตนั้นๆ เพื่อลดจำนวนของเสียดังกล่าวให้เกิดขึ้นน้อยที่สุดจึงจำเป็นต้องมีกระบวนการป้องกันการเกิดของเสียในกระบวนการเพื่อดำเนินการแก้ไขและปรับปรุงก่อนที่จะเกิดของเสียในกระ

บวนการผลิตและกระบวนการทดสอบ ซึ่งก่อให้เกิดผลกระทบในส่วนของต้นทุนการผลิต และระยะเวลาในการผลิตรวมไปถึงการส่งมอบผลิตภัณฑ์ให้แก่ผู้บริโภคด้วย

ดังนั้นงานวิจัยฉบับนี้จึงมุ่งเน้นการป้องกันการเกิดของเสียในกระบวนการทดสอบความเชื่อมั่นก่อนการส่งมอบให้แก่ผู้บริโภค (ORT) ซึ่งถือเป็นกระบวนการทดสอบลำดับสุดท้ายก่อนการส่งมอบผลิตภัณฑ์ โดยการสร้างแบบจำลองเพื่อพยากรณ์คุณภาพผลิตภัณฑ์ฮาร์ดดิสก์ เนื่องจากข้อมูลที่จะใช้ในงานวิจัยนี้มีจำนวนตัวแปรหลายตัวแปรและเป็นข้อมูลขนาดใหญ่ จึงเลือกใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) หรืออาจจะเรียกว่า การค้นหาความรู้ในฐานข้อมูล (Knowledge Discovery in Databases : KDD) คือกระบวนการที่กระทำกับข้อมูลที่มีจำนวนมาก เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยใช้การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) ซึ่งสามารถใช้กับตัวแปรตามที่เป็นตัวแปรเชิงกลุ่มหรือเชิงคุณภาพ ซึ่งมีวัตถุประสงค์และแนวคิดเหมือนกับการวิเคราะห์การถดถอยเชิงเส้น คือเพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ และนำสมการที่ได้ไปประมาณค่าหรือพยากรณ์ตัวแปรตาม วิธีที่สองโครงข่ายประสาทเทียม โครงข่ายประสาทเทียม (Neural network) เป็นศาสตร์แขนงหนึ่งทางด้านปัญญาประดิษฐ์ ที่สามารถนำไปประยุกต์ใช้กับงานหลายด้านได้อย่างมีประสิทธิภาพ เช่นการจำแนกรูปแบบ การทำนาย การควบคุม การหาความเหมาะสม และการจัดกลุ่ม เป็นต้น

ด้วยเหตุที่ฮาร์ดดิสก์หนึ่งตัวมีส่วนประกอบมากมายและการทดสอบหลายขั้นตอนจนกว่าจะส่งมอบให้แก่ผู้บริโภคและขั้นตอนเหล่านั้นได้มีการเก็บข้อมูลไว้ทั้งสิ้น จึงจำเป็นต้องศึกษาถึงปัจจัยที่สำคัญหรือปัจจัยที่ส่งผลกระทบต่อความผิดปกติหรือคุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ ในงานวิจัยนี้ ผู้วิจัยทำการเก็บข้อมูลอาการผิดปกติของฮาร์ดดิสก์ที่กระบวนการทดสอบความเชื่อมั่นก่อนการส่งมอบ พบว่าลักษณะอาการเสียของชุดหัวอ่าน/เขียนผิดปกติมีเป็นจำนวนมาก จึงจะทำการศึกษาและสร้างแบบจำลองเพื่อพยากรณ์ฮาร์ดดิสก์ว่าชุดหัวอ่าน/เขียน จะมีความผิดปกติหรือไม่ และในปัจจุบันยังไม่มียานวิจัยใดที่ทำการศึกษาและสร้างแบบจำลองในการทำนายคุณภาพของฮาร์ดดิสก์ ดังนั้นผู้วิจัยจึงได้ทำการศึกษางานวิจัย ขั้นตอนและวิธีการวิเคราะห์ข้อมูล เพื่อที่จะได้มาซึ่งรูปแบบการพยากรณ์ของเสียในกระบวนการทดสอบความเชื่อมั่นก่อนการส่งมอบผลิตภัณฑ์

1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อทำการศึกษาถึงปัจจัยที่ส่งผลกระทบต่อคุณภาพฮาร์ดดิสก์หรือโอกาสที่ฮาร์ดดิสก์จะเกิดอาการผิดปกติที่หัวอ่านเขียน พร้อมทั้งศึกษาระดับความสัมพันธ์ของปัจจัย ในกระบวนการทดสอบผลิตภัณฑ์ฮาร์ดดิสก์ก่อนที่จะเป็นผลิตภัณฑ์ที่เสร็จสมบูรณ์

2. เพื่อเปรียบเทียบความสามารถในการทำนายคุณภาพความเชื่อมั่นของผลิตภัณฑ์ฮาร์ดดิสก์ ระหว่างแบบจำลองการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) และโครงข่ายประสาทเทียม (Neural Network)

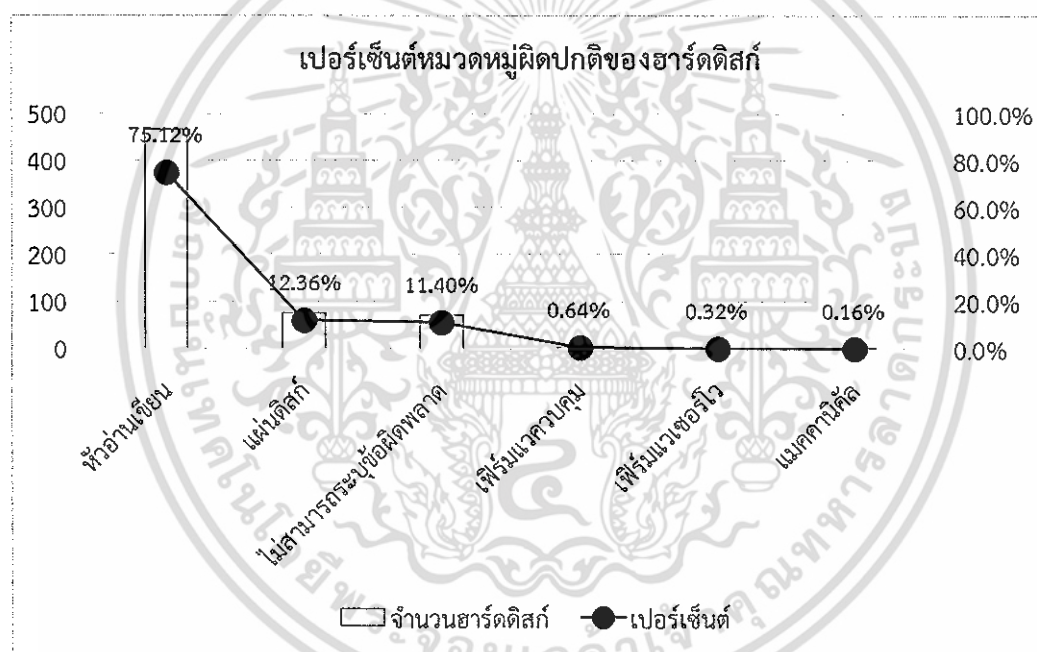
3. นำแบบจำลองที่เหมาะสมที่สร้างขึ้นมาใช้ในการทำนายคุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ ก่อนขั้นตอนการทดสอบคุณภาพความเชื่อมั่นก่อนการส่งมอบแก่ผู้บริโภค (ORT)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ขอบเขตของงานวิจัย

ในงานวิจัยนี้จะทำการศึกษาปัจจัยที่ส่งผลกระทบต่อคุณภาพผลิตภัณฑ์ฮาร์ดดิสก์ และสร้างแบบจำลองในการพยากรณ์คุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ โดยการเลือกแบบจำลองที่มีความเหมาะสมมากที่สุดเพื่อให้ผลการพยากรณ์มีความถูกต้อง จากวิธีการวิเคราะห์การถดถอยโลจิสติกและโครงข่ายประสาทเทียม โดยมีขอบเขตการวิจัยดังนี้

1. ทำการศึกษาปัจจัยที่ส่งผลกระทบต่อฮาร์ดดิสก์ที่มีลักษณะอาการผิดปกติที่ชุดหัวอ่าน/เขียน (HSA) ในกระบวนการทดสอบความเชื่อมั่นก่อนการส่งมอบ เนื่องจากอาการผิดปกติของฮาร์ดดิสก์มีได้หลายลักษณะอาการ จึงรวบรวมข้อมูลอาการผิดปกติที่เกิดขึ้นพบว่าอาการผิดปกติที่หัวอ่าน/เขียนนั้นมีปริมาณมากกว่าอาการอื่นๆ ผู้วิจัยจึงทำการวิจัยลักษณะอาการผิดปกติที่ชุดหัวอ่าน/เขียนเท่านั้น



รูปที่ 1.1 จำนวนและเปอร์เซ็นต์อาการผิดปกติของฮาร์ดดิสก์ในกระบวนการทดสอบความเชื่อมั่น

2. ตัวแปรที่ใช้ในงานวิจัยประกอบด้วย

ก. ตัวแปรอิสระหรือปัจจัย ซึ่งมาจากกระบวนการทดสอบฮาร์ดดิสก์สองกระบวนการก่อนหน้ากระบวนการทดสอบความเชื่อมั่นก่อนการส่งมอบผลิตภัณฑ์ ดังนี้

- กระบวนการทดสอบชุดหัวอ่าน/เขียน (Head Stack Assembly: HSA) จำนวน 14 ตัวแปร

- กระบวนการทดสอบฮาร์ดดิสก์ (Backend Testing: BE) จำนวน 183 ตัวแปร

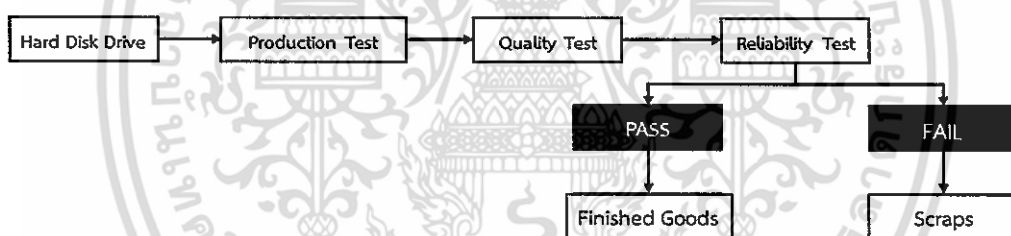
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข. ตัวแปรตาม คือฮาร์ดดิสก์ที่มีอาการผิดปกติที่หัวอ่าน/เขียนและฮาร์ดดิสก์ที่ไม่มีอาการผิดปกติ โดยกำหนดให้ค่าเป็น 1 เท่ากับฮาร์ดดิสก์ที่มีอาการผิดปกติ และ 0 เท่ากับฮาร์ดดิสก์ที่ไม่มีอาการผิดปกติ หรือตัวแปรตามเป็นข้อมูลแบบไบนารี จำนวนตัวแปรอิสระทั้งหมดแสดงในตารางภาคผนวก ก

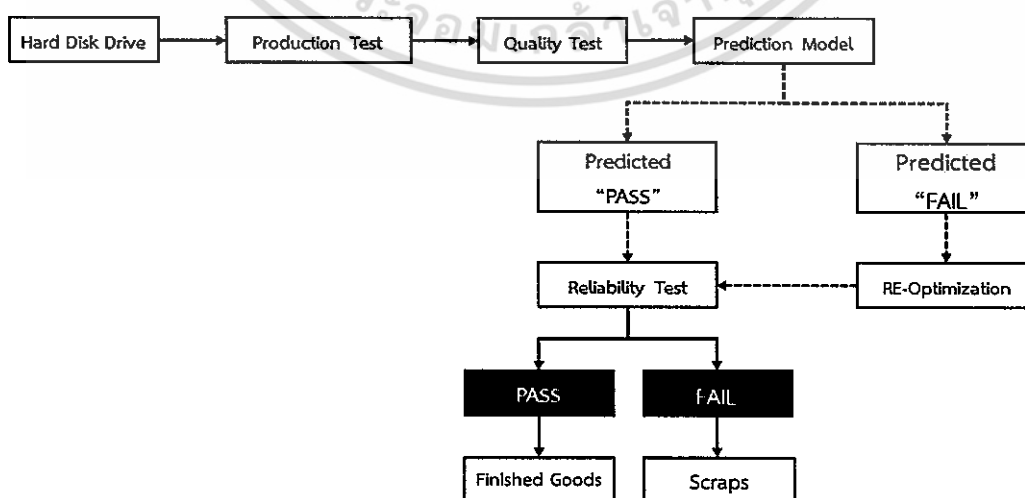
3. ปริมาณของข้อมูลที่ใช้ในการทำวิจัยจำนวน 167,934 ข้อมูล แบ่งเป็นชุดการเรียนรู้ (Training Data Set) 60% ชุดการทดสอบ (Validation Data Set) 20% และชุดการทวนสอบ (Testing Data Set) 20%

4. เปรียบเทียบความสามารถในการทำนายคุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ระหว่างแบบจำลองการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) และโครงข่ายประสาทเทียม (Neural Network)

5. นำแบบจำลองที่ดีที่สุดจากข้อ 4 มาสร้างเป็นรูปแบบการทำนายคุณภาพผลิตภัณฑ์ (Prediction Model) ในกระบวนการทดสอบคุณภาพผลิตภัณฑ์ฮาร์ดดิสก์ก่อนการส่งมอบให้แก่ผู้บริโภค ดังแสดงในรูปที่ 1.2 แสดงกระบวนการทดสอบคุณภาพฮาร์ดดิสก์ก่อนเพิ่มกระบวนการพยากรณ์ และรูปที่ 1.3 แสดงกระบวนการทดสอบคุณภาพฮาร์ดดิสก์หลังเพิ่มกระบวนการพยากรณ์



รูปที่ 1.2 กระบวนการทดสอบฮาร์ดดิสก์ก่อนเพิ่มกระบวนการพยากรณ์



รูปที่ 1.3 กระบวนการทดสอบฮาร์ดดิสก์หลังเพิ่มกระบวนการพยากรณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นับญาติเห็นาไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลังจากที่มีกระบวนการพยากรณ์คุณภาพของฮาร์ดดิสก์แล้ว ฮาร์ดดิสก์ที่มีผลการพยากรณ์ว่ามีโอกาสจะผิดปกติ จะถูกนำไปปรับค่าพารามิเตอร์เพื่อให้มีประสิทธิภาพที่ดีขึ้นแล้วจึงนำไปทดสอบผลิตภัณฑ์ก่อนการส่งมอบให้แก่ผู้บริโภคตามปกติ เพื่อลดจำนวนฮาร์ดดิสก์ผิดปกติหรือฮาร์ดดิสก์เสียที่จะเกิดขึ้นในกระบวนการทดสอบความเชื่อมั่นก่อนการส่งมอบให้เหลือน้อยลง

6. โปรแกรมที่ใช้ในงานวิจัยครั้งนี้คือ SAS Enterprise Miner Workstation 13.2

1.4 เกณฑ์การตัดสินใจ

ในงานวิจัยนี้จะทำการเปรียบเทียบแบบจำลองที่ได้จากการวิเคราะห์การถดถอยโลจิสติกหลายตัวแปรและโครงข่ายประสาทเทียมแบบหลายชั้น ที่ให้ผลการพยากรณ์ถูกต้องมากที่สุดโดยจะใช้เกณฑ์ในการประเมินความถูกต้องจาก Confusion Matrix โดยจะพิจารณาจากค่าความถูกต้อง (Accuracy) และค่าพยากรณ์การเกิดเหตุการณ์ที่ถูกต้อง (Recall)

1.5 ขั้นตอนการวิจัย

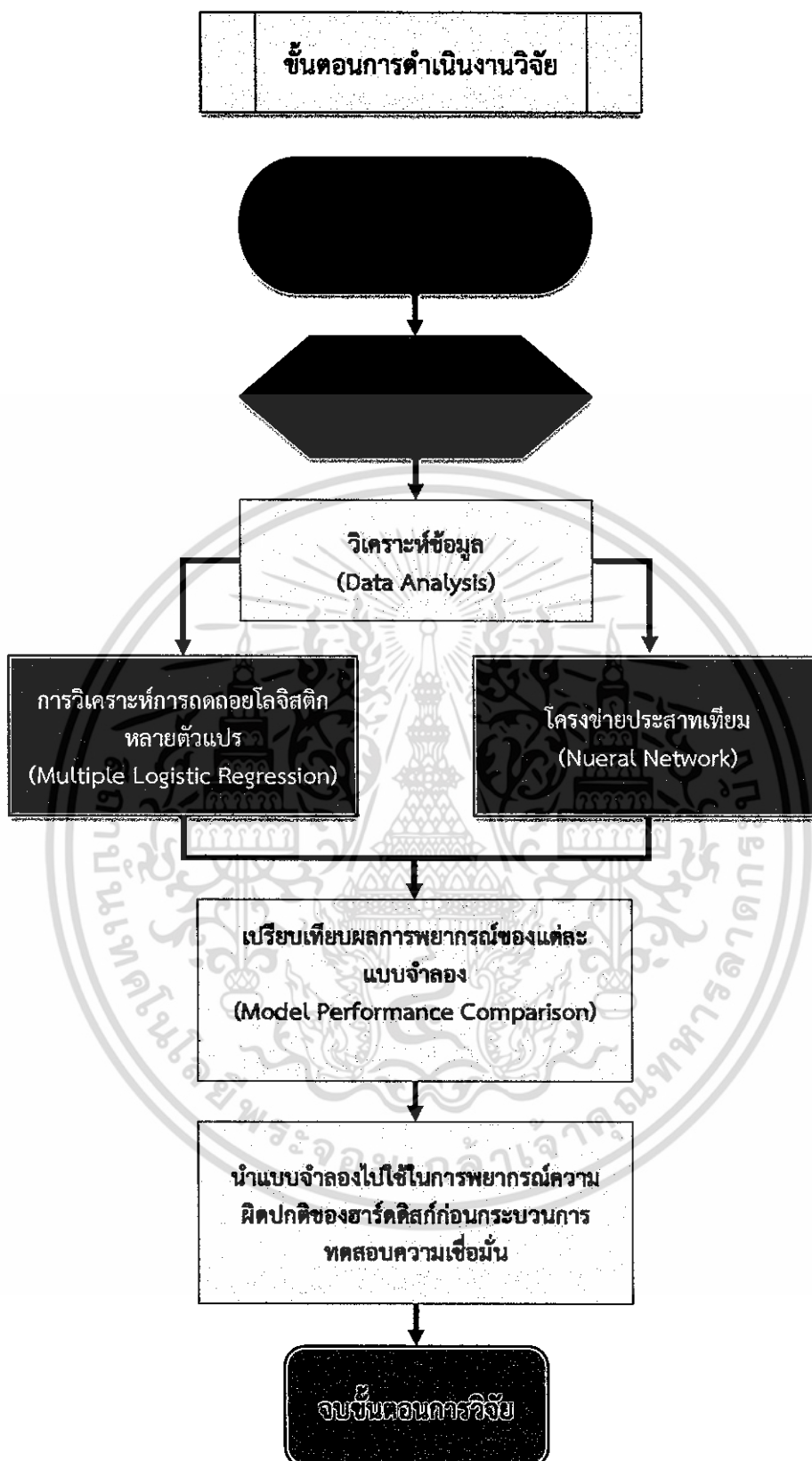
1. เก็บรวบรวมข้อมูลจากแหล่งที่เกี่ยวข้องกับงานวิจัย ในที่นี้เก็บข้อมูลตัวแปรอิสระที่ได้จากกระบวนการทดสอบชุดหัวอ่าน/เขียน และกระบวนการทดสอบฮาร์ดดิสก์หลังการประกอบ ข้อมูลตัวแปรตามคือผลการทดสอบความเชื่อมั่นก่อนการส่งมอบผลิตภัณฑ์ฮาร์ดดิสก์ ตั้งแต่เดือนเมษายน 2558 ถึงเดือนพฤศจิกายน 2558 เป็นจำนวน 8 เดือน

2. เตรียมข้อมูล คัดเลือกข้อมูลที่เกี่ยวข้อง ตรวจสอบความสมบูรณ์ของข้อมูล รวบรวมข้อมูลจากหลายแหล่งข้อมูลและเปลี่ยนข้อมูลให้อยู่ในรูปแบบที่เหมาะสมกับเทคนิคการสร้างแบบจำลองที่เลือกใช้

3. การวิเคราะห์ข้อมูล หาความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามเพื่อศึกษาถึงปัจจัยที่ส่งผลกระทบต่อฮาร์ดดิสก์ที่มีอาการผิดปกติที่หัวอ่าน/เขียน และทำการวิเคราะห์การถดถอยโลจิสติกตัวแปรหลายตัวและโครงข่ายประสาทเทียมแบบหลายชั้น เพื่อหาแบบจำลองการพยากรณ์คุณภาพของฮาร์ดดิสก์ที่เหมาะสมที่สุด

4. สรุปผลการวิเคราะห์ และเปรียบเทียบความสามารถในการพยากรณ์ของแบบจำลองการวิเคราะห์การถดถอยโลจิสติกและโครงข่ายประสาทเทียม

5. นำแบบจำลองที่เหมาะสมไปใช้ในการพยากรณ์คุณภาพผลิตภัณฑ์ฮาร์ดดิสก์ก่อนกระบวนการทดสอบความเชื่อมั่นก่อนการส่งมอบผลิตภัณฑ์ ดังแสดงในรูปที่ 1.4 แสดงขั้นตอนการดำเนินงานวิจัย



รูปที่ 1.4 ขั้นตอนการดำเนินงานวิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.6 ประโยชน์ที่คาดว่าจะได้รับ

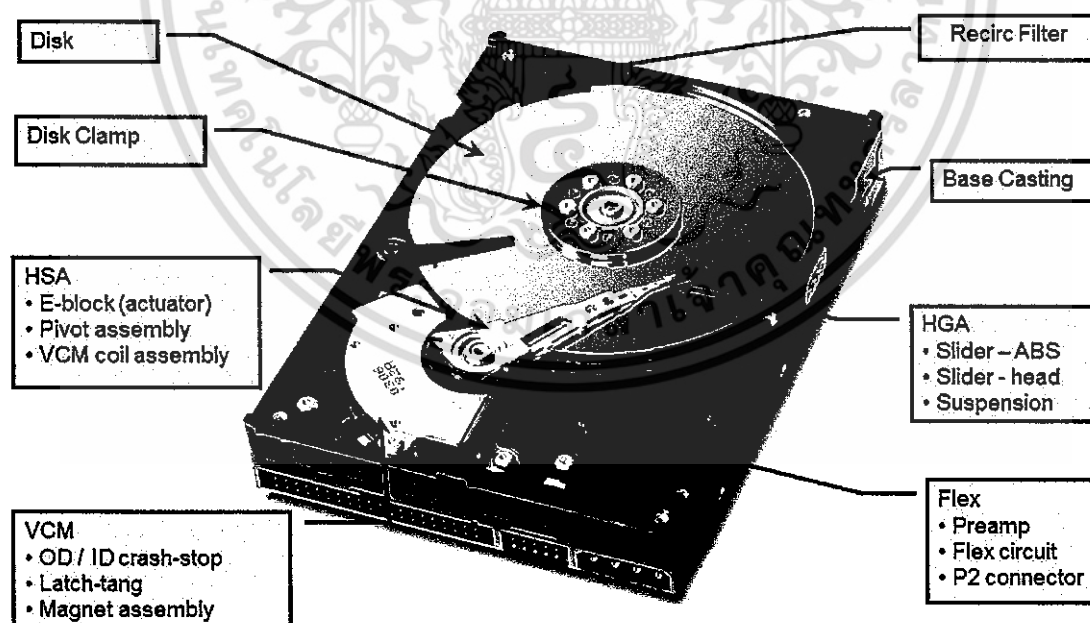
1. ทำให้ทราบว่ามีตัวแปรอิสระหรือปัจจัยตัวใดที่ส่งผลกระทบต่อหรือมีความสัมพันธ์กับฮาร์ดดิสก์ที่มีอาการผิดปกติที่ซุดหัวอ่าน/เขียน รวมทั้งรูปแบบความสัมพันธ์ระหว่างปัจจัยกับฮาร์ดดิสก์ที่มีอาการผิดปกติที่ซุดหัวอ่าน/เขียนด้วย

2. สามารถนำแบบจำลองที่ได้จากงานวิจัยไปใช้ในการพยากรณ์คุณภาพของฮาร์ดดิสก์ได้อย่างมีประสิทธิภาพ

3. ช่วยลดปริมาณของเสียที่เกิดจากกระบวนการผลิตและกระบวนการทดสอบผลิตภัณฑ์ฮาร์ดดิสก์ ลดต้นทุนในการผลิต และสามารถป้องกันการเกิดของเสียในกระบวนการได้ก่อนที่จะเกิดของเสียในกระบวนการผลิต/ทดสอบจริง จะทำให้ช่วยลดภาระงาน (Work Load) ในส่วนของการสืบค้นและการวิเคราะห์ความล้มเหลว (Failure Analysis) ของผลิตภัณฑ์ได้

1.7 นิยามศัพท์

โน้ตบุ๊ก (Notebook) คือ คอมพิวเตอร์ที่มีขนาดเล็กกว่าไมโครคอมพิวเตอร์ ถูกออกแบบไว้เพื่อนำติดตัวไปใช้ตามที่ต่าง ๆ มีขนาดเล็ก และน้ำหนักเบา ในปัจจุบันมีขนาดพอ ๆ กับสมุด มีเส้นผ่าศูนย์กลางของฮาร์ดดิสก์ประมาณ 2.5 นิ้ว ซึ่งภายในฮาร์ดดิสก์จะประกอบด้วยองค์ประกอบ ดังรูปที่ 1.5 ตามรายละเอียดดังต่อไปนี้



รูปที่ 1.5 ส่วนประกอบของฮาร์ดดิสก์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มอเตอร์เบสหรือสปินเดิลมอเตอร์ (Motor base หรือ Spindle motor) คือชุดมอเตอร์หลักที่หมุนแผ่นบันทึกข้อมูลให้เคลื่อนที่ไปรอบ ๆ ที่ความเร็วหนึ่งๆ

แผ่นดิสก์ (Disk) คือแผ่นบันทึกข้อมูล ใช้สำหรับจัดเก็บข้อมูล

ดิสก์แคลมป์ (Disc clamp) คืออุปกรณ์ยึดติดชุดมอเตอร์กับแผ่นดิสก์เข้าด้วยกัน

ซัสเพนชัน (Suspension) คือแขนยึดหัวอ่าน-เขียนเข้ากับแขนยึดชุดหัวอ่านเขียน

เอทจีเอ (Head Gimbal Assembly : HGA) คือชุดหัวอ่าน/เขียน

เอทเอสเอ (Head Stack Assembly : HSA) คือชุดหัวอ่าน/เขียน มีด้านหนึ่งยึดติดกับซัสเพนชัน และอีกด้านหนึ่งติดกับขดลวดแม่เหล็ก (VCM Coil)

วอยล์คอยล์มอเตอร์ (Voice Coil Motor : VCM Coil) คือ ขดลวดแม่เหล็กที่ใช้หลักการเหนี่ยวนำของแม่เหล็กให้เกิดการเคลื่อนที่ของเอทเอสเอ

เฟล็ก (Flex) คือชุดประกอบวงจรชนิดอ่อนของหัวอ่าน เพื่อทำการเชื่อมต่อระหว่างชุดหัวอ่านภายในตัวฮาร์ดดิสก์กับแผงวงจรด้านนอก

เบสคาสติง (Base casting) เป็นแผ่นโลหะชิ้นล่างที่มีความแข็งแรงมากที่สุด

มัลติแพลตเทียร์ล (Multi Platter) คือฮาร์ดดิสก์ที่ประกอบด้วยแผ่นดิสก์มากกว่าหนึ่งแผ่น

ซิงเกิลแพลตเทียร์ล (Single Platter) คือฮาร์ดดิสก์ที่ประกอบด้วยแผ่นดิสก์เพียงหนึ่งแผ่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

วัตถุประสงค์ของงานวิจัยนี้คือ การสร้างรูปแบบการทำนายคุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ โดยใช้เทคนิคการทำเหมืองข้อมูลมาใช้คัดแยกรูปแบบฮาร์ดดิสก์ที่มีอาการผิดปกติในกระบวนการทดสอบคุณภาพความเชื่อมั่นก่อนการส่งมอบแก่ผู้บริโภค (ORT) โดยการเปรียบเทียบความสามารถในการทำนายคุณภาพของฮาร์ดดิสก์ระหว่างแบบจำลอง การวิเคราะห์การถดถอยโลจิสติกและโครงข่ายประสาทเทียม ทั้งนี้พบว่ามิงงานวิจัยเรื่องการวิเคราะห์อาการเสียของฮาร์ดดิสก์โดยวิธีการทำเหมืองข้อมูล ซึ่งได้นำเสนอถึงแนวทางการนำเทคนิคการทำเหมืองข้อมูลมาใช้เพื่อช่วยวิเคราะห์อาการเสียของฮาร์ดดิสก์โดยวิธีการคัดแยกรูปแบบของเสียที่เกิดขึ้นบนแผ่นบันทึกข้อมูล (Recoding Media) และงานวิจัยเรื่องการสร้างโมเดลเพื่อทำนายจากโครงข่ายประสาทเทียมสำหรับการผลิตชุดประกอบหัวอ่านข้อมูล (Head Gimbal Assembly : HGA) โดยงานวิจัยมุ่งเน้นในส่วนการออกแบบระบบจำลองการทำนายเปอร์เซ็นต์ผลผลิตเป็นสำคัญ ดังนั้นจึงสามารถนำแนวคิดที่ได้จากการศึกษางานวิจัยมาประยุกต์ใช้เพื่อเป็นแนวทางสำหรับงานวิจัยนี้ ก่อนจะเข้าสู่กระบวนการและขั้นตอนของการดำเนินงานวิจัย ผู้วิจัยทำการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องดังต่อไปนี้

2.1 การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis)

การวิเคราะห์การถดถอยโลจิสติกถูกนำมาใช้เพื่อทำนายว่า จะเกิดเหตุการณ์หนึ่งขึ้นหรือไม่ โดยมีการกำหนดค่าตัวแปรตัวหนึ่งหรือหลายตัวที่คาดว่าจะส่งผลกระทบต่อเหตุการณ์นั้นๆ

2.1.1 การจำแนกชนิดของการถดถอยโลจิสติก

การแบ่งชนิดของการวิเคราะห์การถดถอยโลจิสติก จะใช้ลักษณะหรือธรรมชาติของตัวแปรตอบสนองหรือตัวแปรตาม (Response :Y) เป็นตัวกำหนด ดังแผนภาพต่อไปนี้ (นำชัย, 2553)



รูปที่ 2.1 การจำแนกชนิดของการวิเคราะห์การถดถอยโลจิสติก (นำชัย, 2553)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถึงแม้ตัวแปรตอบสนองจะเป็นแบบไม่ต่อเนื่องก็ตาม แต่การถดถอยโลจิสติกก็ไม่กำหนดว่าตัวแปรอิสระ (X) จะต้องเป็นแบบไม่ต่อเนื่อง สามารถเป็นได้ทั้งแบบค่าต่อเนื่องที่วัดได้โดยเครื่องมือวัดและแบบไม่ต่อเนื่อง (Category) ก็ได้

2.1.2 ประเภทของการวิเคราะห์การถดถอยโลจิสติก

การวิเคราะห์การถดถอยโลจิสติกแบ่งออกเป็น 2 ประเภท (กัลยา, 2546)

1. โลจิสติกแบบทวิ (Binary Logistic Regression) จะใช้เมื่อตัวแปรตามเป็นตัวแปรเชิงกลุ่มที่มีค่าได้เพียงสองค่าเท่านั้น (Dichotomous Variable)
2. โลจิสติกแบบพหุกลุ่ม (Multinomial Logistic Regression) จะใช้เมื่อตัวแปรตามเป็นตัวแปรเชิงกลุ่มที่มีค่ามากกว่า 2 ค่า

2.1.3 วัตถุประสงค์ของการวิเคราะห์การถดถอยโลจิสติก

1. เพื่อศึกษาความสัมพันธ์ของตัวแปรอิสระที่มีผลต่อโอกาสที่จะเกิดเหตุการณ์ (ตัวแปรตาม) พร้อมทั้งศึกษาระดับความสัมพันธ์ของตัวแปรอิสระแต่ละตัว
2. เพื่อพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจ จากสมการที่เหมาะสมหรือใช้สมการโดยการเลือกตัวแปรอิสระที่เหมาะสมเพื่อให้เปอร์เซ็นต์ของความถูกต้องในการพยากรณ์มีค่าสูงสุด

2.1.4 ข้อตกลงเบื้องต้นการวิเคราะห์การถดถอยโลจิสติก

การวิเคราะห์การถดถอยโลจิสติก มีข้อตกลงเบื้องต้นดังนี้

1. ตัวแปรอิสระ เป็นตัวแปรที่ระดับข้อมูลอยู่ในระดับช่วง (interval Scale) เป็นอย่างต่ำ กรณีที่เป็นข้อมูลเชิงกลุ่มให้แปลงเป็นตัวแปรหุ่น (Dummy Variable) ส่วนตัวแปรตามหรือตัวแปรเกณฑ์ กรณีที่เป็นการวิเคราะห์โลจิสติกแบบทวิ (Binary Logistic Regression) จะกำหนด 2 ค่าคือ 0 กับ 1 ส่วนกรณีการวิเคราะห์โลจิสติกแบบพหุกลุ่ม (Multinomial Logistic Regression) จะกำหนดตามจำนวนกลุ่มของตัวแปรตาม

2. ค่าเฉลี่ยความคลาดเคลื่อนเป็นศูนย์และไม่มีความสัมพันธ์กัน นั่นคือ $E(e) = 0$

3. ตัวแปรอิสระไม่มีความสัมพันธ์กันหรือไม่เกิดปัญหา Multicollinearity

ตัวแบบการถดถอยเชิงเส้นพหุเป็นตัวแบบที่มีตัวแปรอิสระจำนวนมากกว่า 1 ตัว ดังนั้นอาจเป็นไปได้ที่ตัวแปรอิสระที่มีอยู่อาจมีความสัมพันธ์กันเองได้ซึ่งจะส่งผลในการเลือกตัวแปรอิสระที่มีความสำคัญต่อตัวแปรตาม เนื่องจากจะพบว่าตัวแปรเหล่านั้นแสดงค่าความสัมพันธ์ต่อตัวแปรตาม หากนักวิจัยเก็บตัวแปรอิสระเหล่านั้นไว้ทั้งหมดในตัวแบบจะทำให้เกิดการซ้ำซ้อนและตัวแบบจะใหญ่เกินความจำเป็น ความสัมพันธ์ระหว่างตัวแปรอิสระเหล่านี้เรียกว่า Multicollinearity

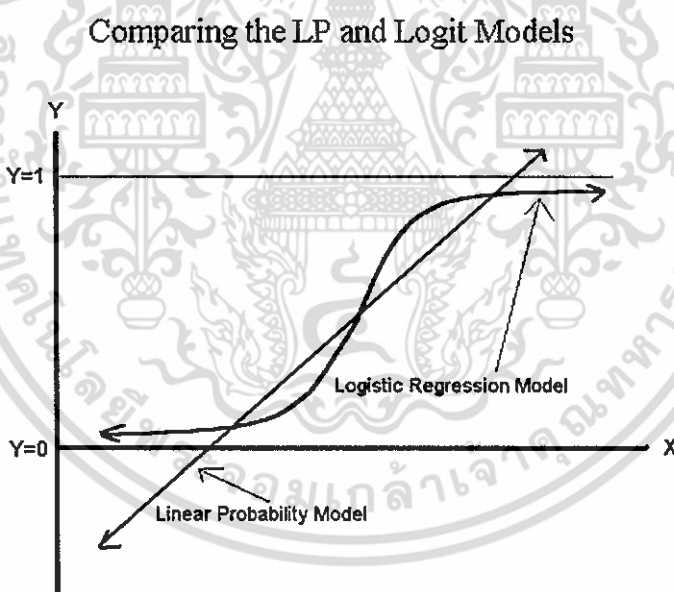
4. การวิเคราะห์การถดถอยโลจิสติกจะต้องใช้ขนาดตัวอย่าง n มากกว่าการวิเคราะห์การถดถอยแบบปกติ โดยจะใช้ขนาดตัวอย่างเท่ากับ $n \geq 30p$ โดยที่ p คือ จำนวนตัวแปรทำนาย

2.1.5 เหตุผลที่ใช้การวิเคราะห์การถดถอยโลจิสติกแทนการวิเคราะห์การถดถอยเชิงเส้น

1. เมื่อตัวแปรตามมีค่าได้เพียง 2 ค่า ทำให้ค่าประมาณของตัวแปรตามเป็นโอกาสที่จะเกิดเหตุการณ์ที่สนใจ ซึ่งมีค่าระหว่าง 0 ถึง 1 ถ้าใช้สมการการถดถอยเชิงเส้นตรง คือ $Y = a + bX$ ค่า Y ที่ได้อาจจะไม่ได้อยู่ในช่วง 0 ถึง 1 หรือ อาจมีค่าน้อยกว่า 0 หรือ มากกว่า 1

2. เงื่อนไขข้อผิดพลาด (Non-Normal Error Terms) ในการวิเคราะห์การถดถอยเชิงเส้น มีเงื่อนไขว่าค่าความคลาดเคลื่อนต้องมีการแจกแจงแบบปกติ แต่เมื่อตัวแปรตามมีค่าเพียง 2 ค่า คือ 0 กับ 1 จะทำให้ค่าความคลาดเคลื่อน e มีค่าได้เพียง 2 ค่าด้วย ซึ่งเป็นไปไม่ได้ที่ e จะมีการแจกแจงแบบปกติ จึงทำให้ไม่สามารถใช้การวิเคราะห์การถดถอยเชิงเส้นตรงโดยทั่วไปได้

3. ค่าความแปรปรวนไม่คงที่ (Non-Constant Error Variance) เนื่องจากเงื่อนไขของการวิเคราะห์การถดถอยคือ ค่าความแปรปรวนของค่าความคลาดเคลื่อนหรือ $V(e)$ ต้องคงที่ทุกค่าของ X แต่ในโลจิสติกนั้น เมื่อตัวแปรตามมีค่าได้เพียง 2 ค่าและตัวแปรตามมีการแจกแจงแบบเบอร์นูลลี ซึ่งทำให้ค่าความแปรปรวนและค่าเฉลี่ยมีความสัมพันธ์กัน จึงทำให้เงื่อนไขที่ว่า $V(e)$ คงที่ไม่เป็นจริง ซึ่งทำให้ไม่สามารถใช้การวิเคราะห์การถดถอยเชิงเส้นตรงตามปกติได้ (สวส์ดีชัย, 2548) โดยฟังก์ชันการวิเคราะห์การถดถอยโลจิสติกเทียบกับการวิเคราะห์การถดถอยเชิงเส้นดังแสดงในรูปที่ 2.2



รูปที่ 2.2 ฟังก์ชันโลจิสติก (Logistic Function)

2.1.6 โมเดลการวิเคราะห์การถดถอยโลจิสติก

โลจิสติกก็คือ Natural Log Odds ซึ่ง Odds หรือ Odds Ratio หมายถึงอัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์ที่สนใจ ($Y=1$) กับโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ ($Y=0$) หรือจะได้ว่า

$$\text{Odds} = \frac{P_Y}{Q_Y} \quad \text{โดยที่} \quad P_Y = \frac{1}{1+e^{-f(x)}}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ P_y คือ ความน่าจะเป็นของการเกิดเหตุการณ์ y

e คือ 2.71828

$f(x)$ คือ ฟังก์ชันของตัวแปรอิสระ

$$P_y = \frac{e^{b_0 + b_1x_1 + \dots + b_px_p}}{1 + e^{b_0 + b_1x_1 + \dots + b_px_p}} \quad \text{และ} \quad Q_y = 1 - P_y$$

ดังนั้น $\text{Log(Odds)} = \text{Log} \frac{P_y}{Q_y} = \text{Log} \frac{P_y}{1 - P_y} = b_0 + b_1x_1 + \dots + b_px_p$

2.1.7 การประมาณค่าโมเดลโลจิสติก (Estimating Logistic Model)

การวิเคราะห์การถดถอยโลจิสติกแตกต่างจากการวิเคราะห์การถดถอยทั่วไป คือการวิเคราะห์การถดถอยใช้การประมาณค่าแบบกำลังสองน้อยที่สุด (Ordinal Least Square : OLS) แต่สำหรับการวิเคราะห์การถดถอยโลจิสติกใช้วิธีประมาณค่าแบบความน่าจะเป็นสูงสุด (Maximum Likelihood : ML) (นำชัย, 2553)

วิธีการประมาณค่าแบบความน่าจะเป็นสูงสุดใช้วิธีการคำนวณซ้ำๆ โดยเริ่มจากการประมาณค่าสัมประสิทธิ์ในสมการการถดถอยโลจิสติก เพื่อให้สามารถแก้สมการได้ แล้วพิจารณาผลการพยากรณ์เพื่อนำมาประมาณค่าสัมประสิทธิ์ใหม่ที่จะทำให้เกิดความน่าจะเป็นสูงสุด เพื่อพยากรณ์ค่าของตัวแปรตามได้อย่างถูกต้องใกล้เคียงกับข้อมูลจริงมากที่สุด

การวิเคราะห์การถดถอยทั่วไป ประมาณค่าแบบกำลังสองน้อยที่สุดโดยประเมินคุณภาพของแบบจำลองหรือสมการจากค่าผลรวมกำลังสองของความคลาดเคลื่อน (Residual Sum of Squares : RSS) แต่สำหรับโลจิสติกประเมินจากค่า Log-Likelihood : LL ซึ่งเทียบเคียงได้กับค่าความคลาดเคลื่อน (Residuals) นั่นก็คือ ยิ่งค่า LL น้อยยิ่งแสดงว่าสมการหรือแบบจำลองที่สร้างขึ้นมีคุณภาพหรือกลมกลืนกับข้อมูลดี เมื่อนำเอา LL คูณกับ -2 กลายเป็น -2LL จะมีการแจกแจงแบบไคสแควร์ (Chi Square) หรือบางที่เรียกว่าค่า Deviance

$$\log - \text{likelihood} = \sum_{i=1}^N [y_i \ln(\hat{y}_i) + (1 - y_i) \ln(1 - \hat{y}_i)]$$

Likelihood Ration Test

$$\chi^2 = [(-2 * \log \text{likelihood for smaller model}) - (-2 * \log \text{likelihood for bigger model})]$$

$$\chi^2 = 2[LL(B) - LL(0)]$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.8 การทดสอบค่าพารามิเตอร์ในแบบจำลองโลจิสติก

1. Hosmer and Lemeshow Test ใช้สถิติไคสแควร์ในการทดสอบว่าแบบจำลองโลจิสติกที่สร้างขึ้นนั้นสามารถสร้างผลการพยากรณ์ความน่าจะเป็นของการเกิดเหตุการณ์สอดคล้องกับความน่าจะเป็นของการเกิดเหตุการณ์จากข้อมูลที่เก็บมาจริงหรือไม่ ถ้าสถิติทดสอบไคสแควร์ ไม่มีนัยสำคัญ แสดงว่าแบบจำลองโลจิสติกมีความกลมกลืน

2. Residual Statistics ใช้ค่า Standardized Residuals ตรวจสอบว่าไม่เกิน 5% ของกลุ่มตัวอย่างมีค่าเกิน 2 และไม่เกิน 1% มีค่าไม่เกิน 2.5 และรายใดที่มีค่าตั้งแต่ 3 ขึ้นไปน่าจะเป็น Outliers

3. Influential Statistics ใช้ Cook's Distance, DFBeta ไม่ควรมีรายใดที่มีค่าเกิน 1 และ Leverage ไม่ควรมีรายใดที่มีค่าเกินกว่าค่าคาดหวัง (Expected Value) คือ $(k+1)/n$ เมื่อ k คือ จำนวนตัวแปรอิสระ

4 Classification Table เป็นตารางที่แสดงผลการทำนายการเกิดเหตุการณ์หรือความเป็นกลุ่มของกลุ่มตัวอย่างจากสมการโลจิสติกที่สร้างขึ้นมา (นำชัย, 2553)

ตารางที่ 2.1 ผลการทำนายการเกิดเหตุการณ์ (Classification Table)

Classification Table

		Predicted		
		decision		Percentage
Observed		stop	continue	Correct
	Step 1 decision	stop	140	47
Continue		60	68	53.1
Overall Percentage				66.0

a. The cut value is .500

ก. เปอร์เซ็นต์ความถูกต้อง (Overall Percentage Correct) ร้อยละของการพยากรณ์ถูกต้องโดยรวมทั้งกรณีที่เกิดเหตุการณ์และไม่เกิดเหตุการณ์ เช่น $(140+68)/315 = 66\%$

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

ข. ค่าความไว (Sensitivity/Recall) ร้อยละของการพยากรณ์การเกิดเหตุการณ์ที่ถูกต้อง เช่น $68/(60+68) = 53.1\%$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

- ค. ค่าจำเพาะ (Specificity) ร้อยละของการพยากรณ์การไม่เกิดเหตุการณ์ถูกต้อง เช่น $140/(140+47) = 74.9\%$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- ง. ค่าทำนายผิดของการเกิดเหตุการณ์ (False Positive) ร้อยละของการพยากรณ์ผิดว่าเกิดเหตุการณ์ เช่น $47/(47+68) = 41.1\%$
- จ. ค่าพยากรณ์ผิดของการไม่เกิดเหตุการณ์ (False Negative) ร้อยละของการพยากรณ์ผิดว่าไม่เกิดเหตุการณ์ เช่น $60/(140+60) = 30\%$

เมื่อ True Positive (TP) คือพยากรณ์ว่าจะเกิดเหตุการณ์และผลลัพธ์บอกว่าเกิดเหตุการณ์
 True Negative (TN) คือพยากรณ์ว่าจะไม่เกิดเหตุการณ์และผลลัพธ์บอกว่าไม่เกิดเหตุการณ์
 False Positive (FP) คือพยากรณ์ว่าจะเกิดเหตุการณ์และผลลัพธ์บอกว่าไม่เกิดเหตุการณ์
 False Negative (FN) คือพยากรณ์ว่าจะไม่เกิดเหตุการณ์และผลลัพธ์บอกว่าเกิดเหตุการณ์

2.2 โครงข่ายประสาทเทียม (Artificial Neural Network)

โครงข่ายประสาทเทียมได้รับแรงบันดาลใจจากการพยายามจำลองระบบประสาททางชีวภาพของมนุษย์ สมองมนุษย์โดยส่วนใหญ่จะประกอบด้วยเซลล์ประสาทหลักที่เรียกว่านิวรอน (Neurons) การเชื่อมโยงของเซลล์ประสาทผ่านทางเส้นใยที่เรียกว่าแอกซอน (axons) ถูกใช้สำหรับส่งกระแสประสาทจากเซลล์ประสาทหนึ่งไปยังอีกเซลล์ประสาทหนึ่ง เมื่อใดก็ตามที่เซลล์ประสาทได้รับการกระตุ้น เซลล์ประสาทจะเชื่อมต่อกับแอกซอนของเซลล์ประสาทอื่นๆ ผ่านทางเดนไดรต์ (Dendrites) จุดเชื่อมต่อระหว่างเดนไดรต์และแอกซอนเรียกว่าไซแนปส์ (Synapse) นักประสาทวิทยาได้ค้นพบว่าสมองมนุษย์เรียนรู้โดยการเปลี่ยนแปลงความแข็งแรงของไซแนปส์ที่ใช้สำหรับการเชื่อมต่อระหว่างเซลล์ประสาท เมื่อกระตุ้นซ้ำโดยการใช้แรงกระตุ้นเดียวกัน (Fausett, 1994)

โครงข่ายประสาทเทียมมีโครงสร้างคล้ายคลึงกับสมองของมนุษย์ โครงข่ายประสาทเทียมประกอบด้วยโหนดที่เชื่อมต่อกันหลายๆ กลุ่มและโหนดที่เชื่อมต่อกันโดยตรง รูปแบบของโครงข่ายประสาทเทียม ที่ง่ายที่สุดเรียกว่าเปอร์เซพตรอน (Perceptron) และจะแสดงให้เห็นว่าโครงข่ายรูปแบบนี้สามารถฝึกให้แก้ปัญหาในการจัดหมวดหมู่ได้ (Classification Problems) (Tan. et al, 2005)

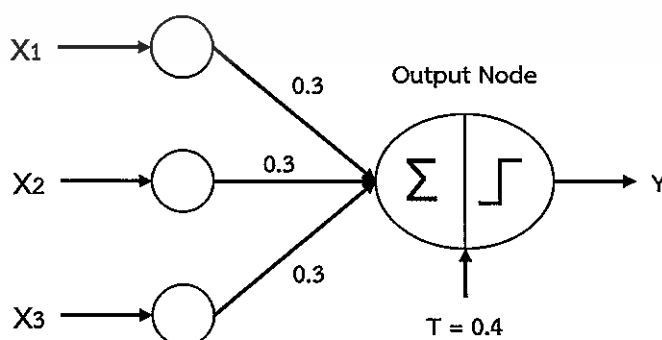
2.2.1 เปอร์เซพตรอน (Perceptron)

จากตารางที่ 2.2 แสดงชุดข้อมูลที่มีตัวแปรอิสระ 3 ตัวแปร (x_1, x_2, x_3) และตัวแปรตาม 1 ตัวแปร (Y) โดยที่ Y จะเท่ากับ -1 ถ้าตัวแปรอิสระอย่างน้อยสองตัวมีค่าเท่ากับ 0 และ Y จะเท่ากับ $+1$ ถ้าตัวแปรอิสระอย่างน้อยสองตัวมีค่ามากกว่า 0

ตารางที่ 2.2 ชุดข้อมูลแบบการจัดหมวดหมู่ (Classification)

X_1	X_2	X_3	Y
1	0	0	-1
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	-1
0	1	0	-1
0	1	1	1
0	0	0	-1

รูปที่ 2.3 แสดงให้เห็นถึงสถาปัตยกรรมของโครงข่ายประสาทเทียมอย่างง่ายที่เรียกว่าเปอร์เซพตรอน (Perceptron) ประกอบด้วยโหนดสองชนิดคือ อินพุตโหนด (Input Node) ซึ่งจะใช้เป็นตัวแทนของการป้อนข้อมูลนำเข้า และเอาต์พุตโหนด (Output Node) ซึ่งจะใช้เป็นตัวแทนแสดงรูปแบบผลลัพธ์ (Model Output) โหนดในสถาปัตยกรรมของโครงข่ายประสาทเทียมเปรียบได้ว่าเป็นเซลล์ประสาทหรือหน่วยงานในเปอร์เซพตรอน แต่ละอินพุตโหนดจะเชื่อมต่อกันด้วยสายเชื่อมโยงโดยมีค่าน้ำหนักกับเอาต์พุตโหนด การเชื่อมโยงโดยมีค่าน้ำหนักใช้ในการจำลองความแข็งแรงของไซแนปส์ ซึ่งเป็นจุดเชื่อมต่อระหว่างเซลล์ประสาท ดังระบบประสาททางชีวภาพ



รูปที่ 2.3 สถาปัตยกรรมโครงข่ายประสาทเทียมแบบเปอร์เซพตรอน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การเรียนรู้ของเปอร์เซพตรอน จะทำการปรับค่าน้ำหนักการเชื่อมโยง จนกว่าจะได้ค่าน้ำหนักที่พอดีกับความสัมพันธ์ของข้อมูลนำเข้าและผลลัพธ์ เปอร์เซพตรอนจะคำนวณค่าเอาท์พุท โดยการแสดงผลรวมการถ่วงน้ำหนักของปัจจัยหรือตัวแปรอิสระลบด้วยปัจจัยเอนเอียง (Bias Factor) T ดังแสดงในรูปที่ 2.3 มี 3 อินพุทโหนด (x_1, x_2, x_3) โดยที่แต่ละตัวแปรมีค่าน้ำหนักเท่ากับ 0.3 เชื่อมต่อไปยังโหนดผลลัพธ์ และปัจจัยเอนเอียง T เท่ากับ 0.4 สามารถคำนวณผลลัพธ์ได้ดังนี้

$$\hat{y} = 1 \text{ if } 0.3x_1 + 0.3x_2 + 0.3x_3 - 0.4 > 0$$

$$\hat{y} = -1 \text{ if } 0.3x_1 + 0.3x_2 + 0.3x_3 - 0.4 < 0$$

เช่น ถ้า $x_1=1, x_2=1$ และ $x_3=0$ จะได้เท่ากับ +1 และถ้า $x_1=0, x_2=1$ และ $x_3=0$ จะได้เท่ากับ -1 ความแตกต่างระหว่างอินพุทและเอาท์พุทโหนดของเปอร์เซพตรอน อินพุทโหนดจะทำการส่งผ่านชุดข้อมูลโดยไม่ได้ดำเนินการเปลี่ยนแปลงใด ในทางกลับกันเอาท์พุทโหนดเป็นอุปกรณ์ทางคณิตศาสตร์ที่จะคำนวณหรือประมวลผลรูปแบบที่เกิดขึ้น โดยเฉพาะอย่างยิ่งเอาท์พุทในรูปแบบของเปอร์เซพตรอนสามารถแสดงสมการทางคณิตศาสตร์ได้ดังนี้

$$\hat{y} = \text{Sign}(w_n x_m + w_{n-1} w_{n-1} + \dots + w_2 x_2 + w_1 x_1 - T)$$

เมื่อ $w_1, w_2, w_3, \dots, w_n$ คือค่าน้ำหนักของแต่ละการเชื่อมโยงปัจจัยนำเข้า และ $x_1, x_2, x_3, \dots, x_n$ คือตัวแปรอิสระ Sign Function จะทำหน้าที่เป็นฟังก์ชันกระตุ้น (Activation Function) สำหรับเซลล์ประสาทเอาท์พุท รูปแบบเปอร์เซพตรอนสามารถเขียนในรูปแบบกะทัดรัดได้ดังนี้

$$\hat{y} = \text{Sign}(w_0 x_0 + w_1 x_1 + \dots + w_n x_n) = \text{Sign}(w \cdot x)$$

เมื่อกำหนดให้ $w_0 = -T$ และ $x_0 = 1$ และ $w \cdot x$ คือดอทโปรดักต์ระหว่างเวกเตอร์ค่าถ่วงน้ำหนัก W และ เวกเตอร์ตัวแปรอิสระ X

2.2.2 การเรียนรู้โครงข่ายประสาทเทียมแบบเปอร์เซพตรอน

ขั้นตอนของการเรียนรู้ของเปอร์เซพตรอน พารามิเตอร์ถ่วงน้ำหนัก W จะถูกปรับค่าจนกว่าผลลัพธ์ที่ได้จะมีความสอดคล้องกับผลลัพธ์ที่แท้จริง หลักสำคัญในการคำนวณหาค่าถ่วงน้ำหนักเป็นดังสมการ

$$w_j^{(k+1)} = w_j^{(k)} + \lambda(y_i - \hat{y}_i^{(k)})x_{ij}$$

เมื่อ $w^{(k)}$ คือพารามิเตอร์น้ำหนักที่เกี่ยวข้องกับ x_i เชื่อมต่อข้อมูลนำเข้า หลังจากการทำซ้ำ k ครั้ง λ คือพารามิเตอร์ที่ทราบค่า นั่นก็คืออัตราการเรียนรู้ (Learning Rate) และ x_{ij} คือค่าครั้งที่ j ของตัวแปรที่ i ในการปรับค่าน้ำหนักดังสมการด้านบน ซึ่งแสดงค่าน้ำหนักใหม่คือ $w^{(k+1)}$ เป็นการรวมกันของค่าน้ำหนักเก่า $w^{(k)}$ และเทอมของสัดส่วนในการพยากรณ์ผิดพลาด ($y - \hat{y}$) ถ้าค่าพยากรณ์ถูกต้องแล้วค่าถ่วงน้ำหนักจะมีค่าเท่าเดิม อย่างไรก็ตามยังสามารถที่จะปรับปรุงแก้ไขได้ดังต่อไปนี้

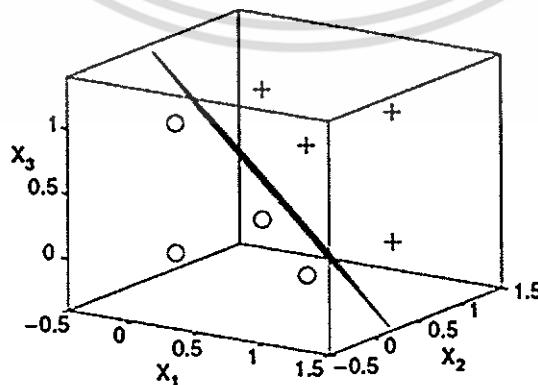
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถ้า $y = +1$ และ $\hat{y} = -1$ แล้วค่าพยากรณ์ผิดพลาดเท่ากับ $(y - \hat{y}) = 2$ เพื่อชดเชยข้อผิดพลาดต้องทำการเพิ่มค่าพยากรณ์โดยการเพิ่มค่าน้ำหนักของทุกสายที่เชื่อมต่อกับข้อมูลนำเข้ามีค่าเป็นบวก และจะทำการลดค่าน้ำหนักของทุกสายการเชื่อมต่อกับค่าผลลัพธ์ที่เป็นลบ

ถ้า $y = -1$ และ $\hat{y} = +1$ แล้ว $(y - \hat{y}) = -2$ เพื่อชดเชยข้อผิดพลาดต้องทำการลดค่าพยากรณ์โดยทำการลดค่าน้ำหนักของทุกสายการเชื่อมต่อกับอินพุตที่มีค่าเป็นบวก และจะเพิ่มค่าน้ำหนักของทุกสายที่เชื่อมต่อกับอินพุตที่เป็นลบ

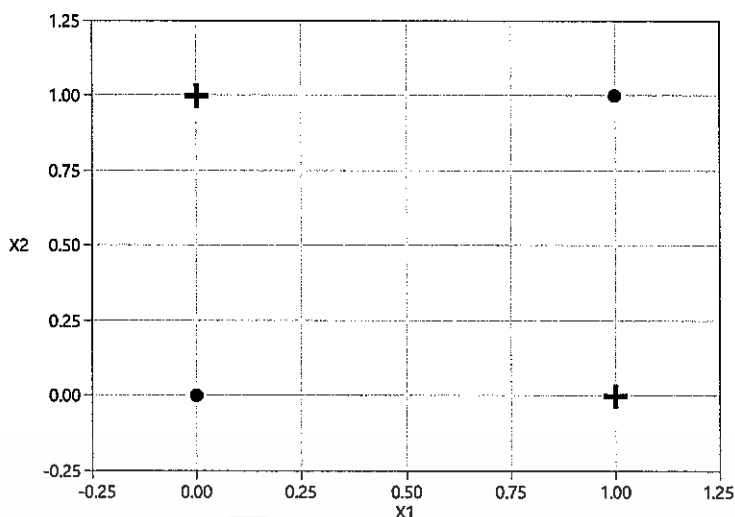
ในสมการการปรับปรุงค่าน้ำหนัก การเชื่อมโยงที่มีส่วนร่วมมากที่สุดกับเทอมของข้อผิดพลาดนั้นจำเป็นต้องมีการปรับปรุงมากที่สุด อย่างไรก็ตามค่าน้ำหนักไม่ควรมีการเปลี่ยนแปลงมากเกินไป เพราะเทอมของข้อผิดพลาด (Error Term) คำนวณจากข้อมูลปัจจุบันที่ใช้ในการเรียนรู้เท่านั้น อีกอย่างหนึ่งการปรับค่าก่อนหน้าจะถูกยกเล็ก อัตราการเรียนรู้ λ คือพารามิเตอร์ที่มีค่าอยู่ระหว่าง 0 และ 1 สามารถใช้ในการควบคุมการปรับค่าในแต่ละการทำซ้ำ ถ้าอัตราการเรียนรู้เข้าใกล้ 0 แล้วโดยส่วนใหญ่ค่าน้ำหนักใหม่จะเป็นผลสืบเนื่องมาจากค่าน้ำหนักเดิม ในทางกลับกันถ้าอัตราการเรียนรู้เข้าใกล้ 1 แล้วค่าน้ำหนักใหม่จะมีความไวต่อปริมาณการปรับค่าในการทำซ้ำ ในบางกรณีการปรับค่าอัตราการเรียนรู้สามารถนำมาใช้ในตอนเริ่มต้น โดยที่อัตราการเรียนรู้จะมีค่าลดลงในการทำซ้ำหลายๆครั้ง (Ben. et al, 1996)

เปอร์เซพตรอนเป็นสมการเชิงเส้นในพารามิเตอร์ W และ ตัวแปร X กำหนดค่า $\hat{y} = 0$ คือไฮเปอร์เพลนเชิงเส้นที่ทำการแยกข้อมูลออกเป็นสองกลุ่มคือ -1 และ $+1$ ดังรูปที่ 2.4 แสดงขอบเขตการตัดสินใจโดยการประยุกต์ใช้วิธีการเรียนรู้แบบเปอร์เซพตรอน วิธีการเรียนรู้แบบเปอร์เซพตรอนสามารถใช้หาค่าตอบที่ดีที่สุดสำหรับการจำแนกหมวดหมู่แบบเส้นตรง ถ้าปัญหาการจำแนกแบบไม่เป็นเส้นตรงจะทำให้เกิดความผิดพลาดขึ้นได้ รูปที่ 2.5 แสดงตัวอย่างของข้อมูลที่ไม่เป็นเชิงเส้น เปอร์เซพตรอนไม่สามารถหาการจำแนกที่สมบูรณ์ได้ในการเรียนรู้ข้อมูลนี้ เพราะไม่มีไฮเปอร์เพลนเชิงเส้นที่สามารถแยกหมวดหมู่ของกรณีศึกษาได้



รูปที่ 2.4 ขอบเขตการตัดสินใจโดยการประยุกต์ใช้วิธีการเรียนรู้แบบเปอร์เซพตรอน

X1	X2	Y
0	0	-1
1	0	1
0	1	1
1	1	-1

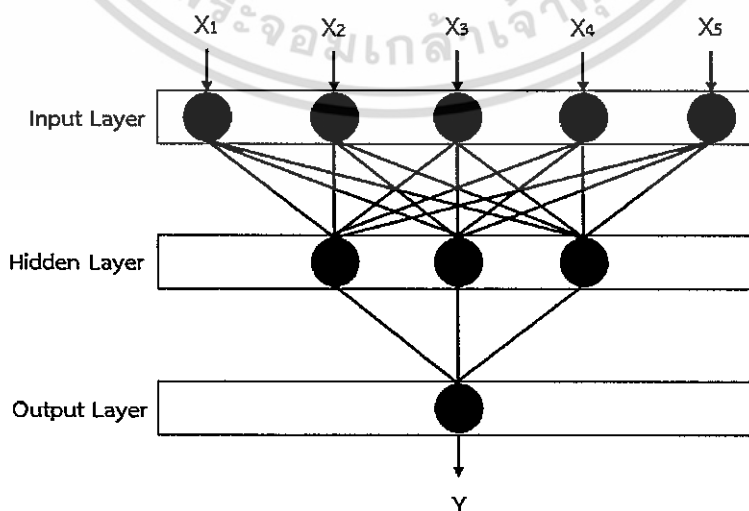


รูปที่ 2.5 ข้อมูลที่ไม่มีไฮเปอร์เพลนเชิงเส้น

2.2.3 โครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer Artificial Neural Network)

โครงข่ายประสาทเทียมรูปแบบนี้มีโครงสร้างที่ซับซ้อนมากกว่าแบบเปอร์เซพตรอน ความซับซ้อนที่เพิ่มขึ้นมีดังนี้

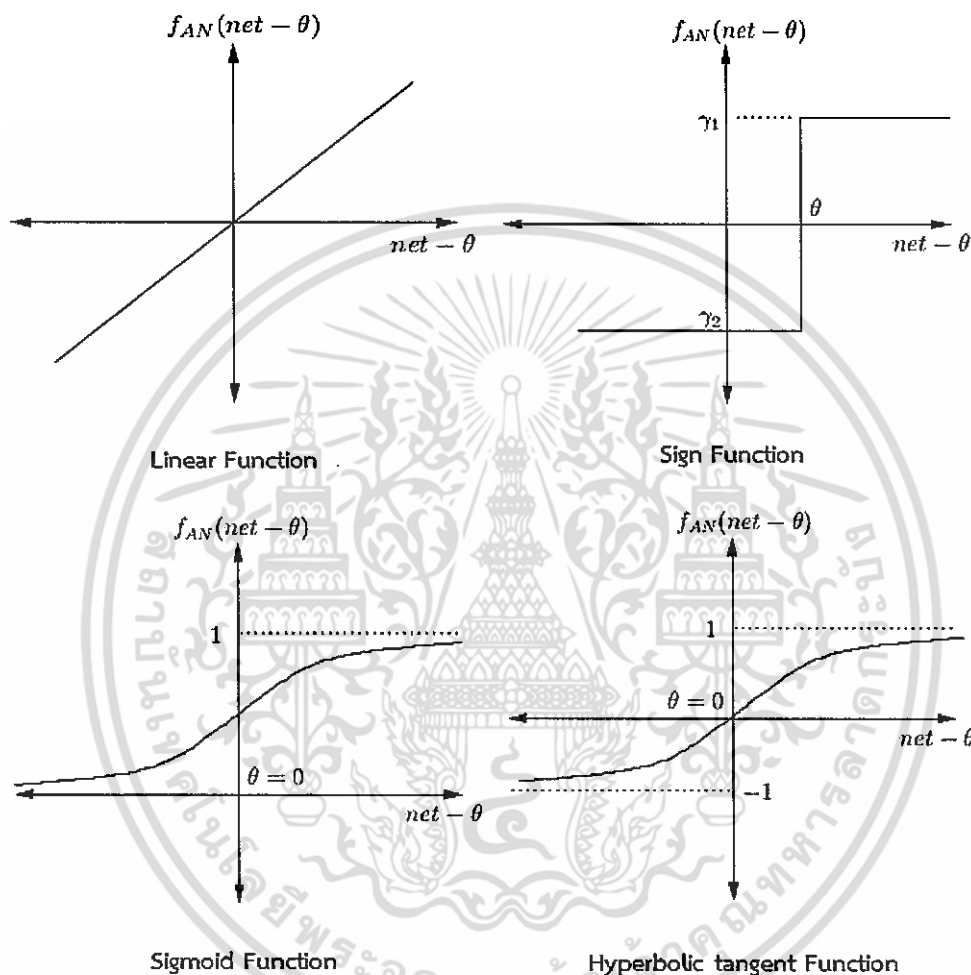
1. โครงข่ายอาจมีหลายชั้น ซึ่งชั้นที่อยู่ตรงกลางระหว่างชั้นข้อมูลนำเข้าและชั้นผลลัพธ์ เรียกว่าชั้นซ่อน (Hidden Layers) และโหนดที่ฝังอยู่ในชั้นนี้เรียกว่าโหนดซ่อน (Hidden Node) ในโครงข่ายประสาทเทียมแบบ Feed-Forward โหนดในชั้นหนึ่งจะเชื่อมต่อกับโหนดในชั้นถัดไปเท่านั้น ในโครงข่ายประสาทการเชื่อมโยงอาจเชื่อมต่อโหนดภายในชั้นเดียวกันหรือโหนดจากชั้นหนึ่งไปยังชั้นก่อนหน้านี้อีกได้



รูปที่ 2.6 ตัวอย่างโครงข่ายประสาทเทียมหลายชั้นแบบ Feed-Forward

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษเท่านั้น ไม่อนุญาตให้เผยแพร่ไปยังประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

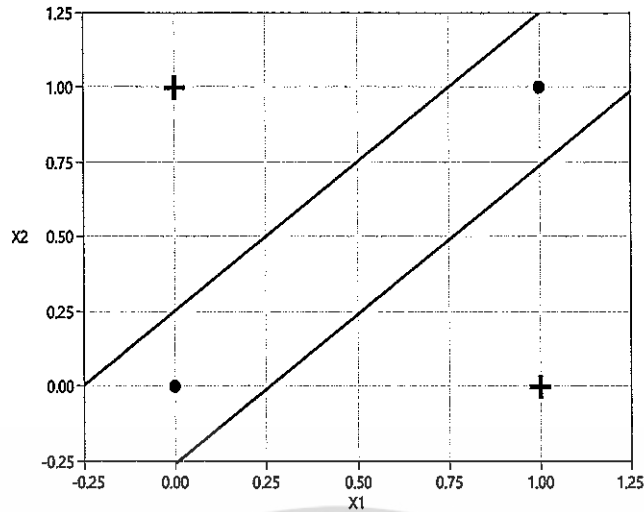
2 โครงข่ายประสาทเทียมอาจจะใช้ฟังก์ชันกระตุ้นอื่นๆนอกเหนือจากไซน์ฟังก์ชัน (Sign Function) ตัวอย่างของฟังก์ชันกระตุ้นอื่นๆ เช่นฟังก์ชันเชิงเส้น ซิกมอยด์ฟังก์ชัน (Sigmoid) และไฮเปอร์โบลิกแทนเจนท์ฟังก์ชัน (Hyperbolic Tangent) ดังแสดงในรูปที่ 2.7 ฟังก์ชันกระตุ้นเหล่านี้ช่วยให้โหนดซ่อนและโหนดผลลัพธ์ดำเนินการผลิตค่าผลลัพธ์ของพารามิเตอร์นำเข้าไปที่ไม่เป็นเชิงเส้น



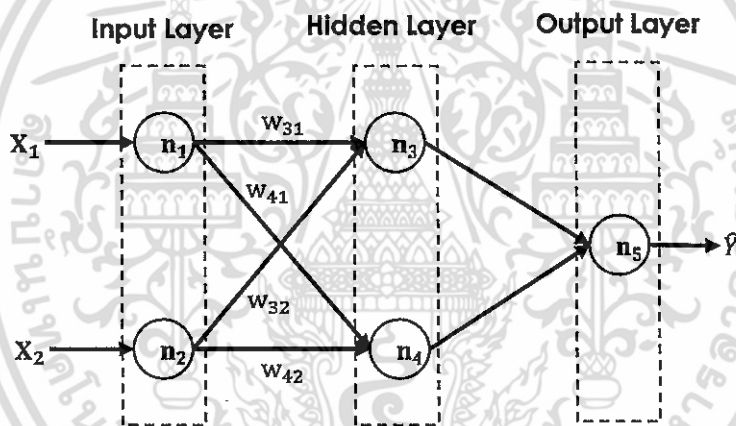
รูปที่ 2.7 ชนิดของฟังก์ชันกระตุ้นในโครงข่ายประสาทเทียม (Ben. et al, 1996)

ความซับซ้อนที่เพิ่มขึ้นเหล่านี้ทำให้โครงข่ายประสาทเทียมแบบหลายชั้นมีความสัมพันธ์ที่ซับซ้อนมากขึ้นระหว่างข้อมูลนำเข้าและตัวแปรเอาต์พุต จากตัวอย่างในรูปที่ 2.8 กรณีนี้สามารถจำแนกหมวดหมู่ได้โดยการใช้ไฮเปอร์เพลน (Hyperplane) สองเส้น ในการแบ่งข้อมูลนำเข้า เพอร์เซพตรอนสามารถสร้างไฮเปอร์เพลนได้เพียงเส้นเดียวเท่านั้น ซึ่งไม่สามารถหาคำตอบที่ดีที่สุดได้ ในปัญหานี้สามารถแก้ไขได้โดยการใช้วิธีการแก้ปัญหาแบบสองชั้น ดังแสดงในรูปที่ 2.9 แต่ละโหนดซ่อนของเพอร์เซพตรอนพยายามที่จะสร้างไฮเปอร์เพลนหนึ่งในสองเส้น ในขณะที่โหนดผลลัพธ์จะทำการรวบรวมผลจากแต่ละเพอร์เซพตรอน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.8 ขอบเขตการตัดสินใจโดยการใช้ไฮเปอร์เพลนสองเส้นในข้อมูลนำเข้า



รูปที่ 2.9 โครงสร้างของโครงข่ายประสาทเทียมแบบสองชั้น

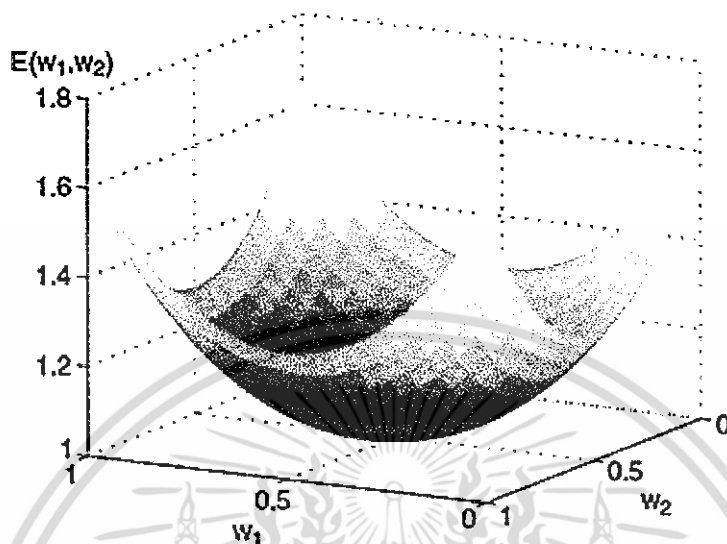
2.2.4 การเรียนรู้ของโครงข่ายประสาทเทียม

เป้าหมายของการเรียนรู้โครงข่ายประสาทเทียมคือ เพื่อตรวจสอบการตั้งค่าน้ำหนักของโครงข่ายโดยให้มีผลรวมความผิดพลาดที่น้อยที่สุด (Total Sum of Square Error)

$$E(w) = \frac{1}{2} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

ผลรวมความผิดพลาดขึ้นอยู่กับค่าน้ำหนักเพราะค่าพยากรณ์ที่ได้ \hat{y} คือฟังก์ชันของค่าน้ำหนักที่กำหนดให้กับโหนดซ่อนและโหนดผลลัพธ์ รูปที่ 2.10 แสดงพื้นผิวความผิดพลาดของตัวอย่างที่มีสองพารามิเตอร์คือ w_1 และ w_2 ชนิดของพื้นผิวข้อผิดพลาดมักจะพบเมื่อ \hat{y}_i คือฟังก์ชันเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เชิงเส้นของ w ถ้าแทน $y = W.X$ ในสมการแล้วฟังก์ชันความผิดพลาดจะกลายเป็นกำลังสองของพารามิเตอร์และสามารถหาคำตอบได้



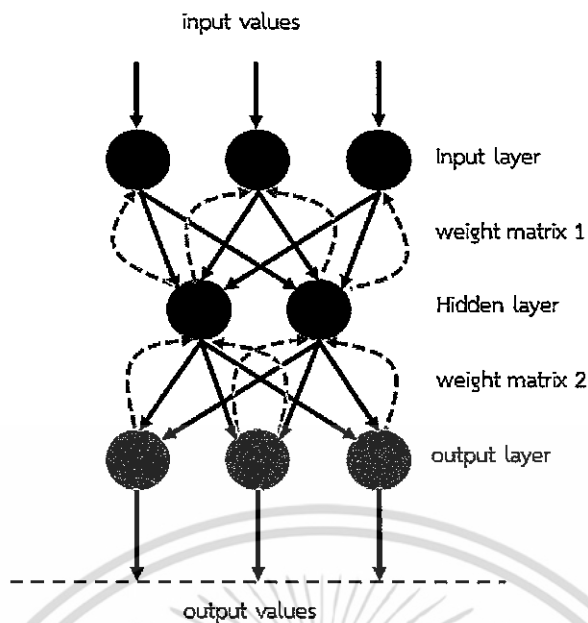
รูปที่ 2.10 พื้นผิวความผิดพลาดของตัวอย่างที่มีสองพารามิเตอร์

ในหลายกรณีผลลัพธ์ของโครงข่ายประสาทเทียมไม่เป็นฟังก์ชันเชิงเส้นของพารามิเตอร์ เพราะการเลือกฟังก์ชันกระตุ้น ผลลัพธ์ที่ไม่ถูกต้องจะถูกปรับค่าน้ำหนักที่ทำให้ได้คำตอบที่ดีที่สุด วิธีการเช่นนั้นขึ้นอยู่กับวิธีการ Gradient Descent Method ทำให้มีประสิทธิภาพในการแก้ปัญหา การปรับค่าน้ำหนักที่ใช้สำหรับวิธีการ Gradient Descent Method สามารถเขียนได้ดังนี้ (Tan, et al, 2005)

$$w_{j+1} = w_j - \lambda \frac{\partial E(w)}{\partial w_j}$$

เมื่อ λ คือ อัตราการเรียนรู้ ในเทอมที่สองของสมการค่าน้ำหนักจะเพิ่มขึ้นถ้าข้อผิดพลาดทั้งหมดลดลง Gradient Descent Method สามารถใช้ในการศึกษาถึงค่าน้ำหนักของผลลัพธ์และโหนดซ่อนในโครงข่ายประสาทเทียม สำหรับค่าน้ำหนักของโหนดซ่อนนั้นสามารถคำนวณได้ยาก เทคนิค Back Propagation จึงถูกสร้างขึ้นมาเพื่อแก้ปัญหานี้ เทคนิคนี้จะมีสองช่วงในแต่ละการทำซ้ำของขั้นตอนคือ Forward Phase และ Backward Phase ใน Forward Phase ค่าน้ำหนักซึ่งได้มาจากการทำซ้ำในครั้งก่อนหน้าจะถูกใช้คำนวณค่าผลลัพธ์ การคำนวณจะดำเนินไปด้านหน้าเพียงอย่างเดียว ผลลัพธ์ของโครงข่ายในการทำซ้ำครั้งที่ k จะถูกคำนวณก่อนผลลัพธ์การทำซ้ำครั้งที่ $k+1$ ส่วน Backward Phase การปรับค่าน้ำหนักจะถูกนำมาใช้ในทิศทางย้อนกลับ กล่าวคือค่าน้ำหนักในครั้งที่ $k+1$ จะถูกปรับปรุงก่อนที่ค่าน้ำหนักในครั้งที่ k จะถูกปรับปรุง วิธีการ Back-Propagation ใช้ข้อผิดพลาดของโครงข่ายในครั้งที่ $k+1$ คำนวณการย้อนกลับของโครงข่ายในครั้งที่ k (Satisch, 2012)

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้สำหรับใช้ในเชิงวิชาการเท่านั้น ไม่สามารถนำเอกสารไปใช้
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.11 การแพร่ย้อนกลับด้วยเทคนิค Back Propagation

2.2.5 ข้อดีและข้อเสียของโครงข่ายประสาทเทียม

ข้อดี

1. ทนทานต่อชุดข้อมูลสอนที่มีค่าผิดพลาด
2. ให้ความแม่นยำในการทำนายสูง
3. ให้ผลลัพธ์เป็นเวกเตอร์คุณลักษณะค่าจริงหรือค่าไม่ต่อเนื่อง
4. ใช้ได้ดีกับปัญหาการรู้จำเสียง (Voice) ลายมือ (Handwriting) หรือรูปภาพ (Image)

ข้อเสีย

1. ใช้เวลาสอนหรือเรียนรู้นาน
2. ผลลัพธ์หรือค่าน้ำหนักของตัวแปรฟังก์ชันที่ได้จากการเรียนรู้ค่อนข้างเข้าใจยาก

2.3 การวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis)

แนวคิดการวิเคราะห์องค์ประกอบหลัก (PCA) คือการลดมิติของชุดข้อมูลที่ประกอบด้วยตัวแปรที่มีความสัมพันธ์กันมาก โดยการสร้างตัวแปรใหม่ซึ่งจะไม่มีความสัมพันธ์กัน โดยสร้างมาจากตัวแปรเดิมทั้งหมดที่มีความสัมพันธ์กันนั่นเอง วิธี PCA เป็นการนำความสัมพันธ์เชิงเส้นระหว่างตัวแปรที่ใช้เป็นข้อมูลกับองค์ประกอบหลัก เพื่ออธิบายความแปรปรวนของกลุ่มข้อมูล ที่มีองค์ประกอบปัจจัยหลักมีความแปรปรวนมากไปน้อยตามลำดับ จนได้องค์ประกอบหลักหรือปัจจัยใหม่ ที่สามารถเป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวแทนอธิบายการผันแปรของทุกตัวแปรได้ครบถ้วน สำหรับสมการที่ใช้ในการประมาณค่าปัจจัยที่ j คือ (I.T. Jolliffe, 2005)

$$F_j = w_{j1}x_1 + w_{j2}x_2 + \dots + w_{jp}x_p + e$$

โดยที่ x_i คือตัวแปรที่ i

w_j คือ สัมประสิทธิ์ของตัวแปรที่ j

F_j คือตัวแปรหรือปัจจัยหรือองค์ประกอบหลักตัวที่ j

นอกจากนี้แล้วเทคนิค PCA ต้องมีการประมาณค่าน้ำหนักของปัจจัย (Factor Loading) ซึ่งจะนำมาใช้พิจารณาว่ามีตัวแปรใดบ้างที่ควรอยู่ในปัจจัยเดียวกัน โดยดูจากค่าน้ำหนักของปัจจัยที่มากที่สุดเป็นหลัก

การหมุนแกนปัจจัย (Factor Rotation) ถือเป็นหลักการที่ใช้จัดการกับค่าน้ำหนักของปัจจัยที่ไม่ชัดเจน คือเป็นค่ากลางๆ ซึ่งวัตถุประสงค์ของการหมุนปัจจัยคือ เพื่อให้ค่าน้ำหนักของปัจจัยของตัวแปรมีค่ามากขึ้นหรือลดลงจนกระทั่งทำให้ทราบว่าตัวแปรนั้นควรอยู่ในปัจจัยใด หรือไม่ควรอยู่ในปัจจัยใด

2.4 การทดสอบไคสแควร์ (Chi-Square Test)

การทดสอบไคสแควร์ เป็นวิธีการทดสอบเพื่อเปรียบเทียบข้อมูลที่อยู่ในรูปของความถี่หรือในรูปของสัดส่วน ตัวอย่างเช่น การศึกษาเจตคติความคิดเห็น ความสนใจ หรือการยอมรับ เป็นต้น ซึ่งไม่สามารถวัดค่าออกมาเป็นตัวเลขที่แน่นอนได้ แต่สามารถจำแนกออกเป็นหมวดหมู่ได้ เช่น มากที่สุด มาก ปานกลาง น้อย น้อยที่สุด หรือดี ไม่ดี เป็นต้น ถ้าหากต้องการเปรียบเทียบตัวแปร 2 กลุ่มหรือมากกว่า 2 กลุ่มว่ามีความสัมพันธ์กันหรือไม่ การทดสอบไคสแควร์จะเหมาะสมกว่าการทดสอบแบบ Z เนื่องจากการทดสอบแบบ Z เหมาะสำหรับการทดสอบสัดส่วนของประชากรเพียงกลุ่มเดียวหรือการทดสอบความแตกต่างระหว่างสัดส่วนของสิ่งที่สนใจจากประชากร 2 กลุ่มเท่านั้น การทดสอบไคสแควร์จึงนิยมใช้มากในการเปรียบเทียบหรือทดสอบข้อมูลที่เป็นความถี่หรือข้อมูลที่อยู่ในรูปของสัดส่วน

2.4.1 วัตถุประสงค์ของการทดสอบไคสแควร์

การทดสอบไคสแควร์ มีวัตถุประสงค์สำคัญ 3 ประการคือ

1. การทดสอบภาวะสารูปดี (The goodness of fit test) เป็นการทดสอบไคสแควร์เพื่อศึกษาว่าการแจกแจงความถี่ของตัวแปรเป็นไปตามรูปแบบที่กำหนดไว้หรือไม่ โดยศึกษาจากตัวแปรเพียงตัวเดียว โดยการเปรียบเทียบระหว่างข้อมูลจากตัวแปรกับข้อมูลที่ได้จากความคาดหมายหรือจากทฤษฎีใด ๆ ว่ามีความสอดคล้องกันหรือไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. การทดสอบความสัมพันธ์ระหว่างตัวแปร (Test of Association) หรือเรียกอีกอย่างหนึ่งว่าการทดสอบความเป็นอิสระ (Test of Independence) เป็นการทดสอบไคสแควร์เพื่อศึกษาว่าตัวแปรต่าง ๆ สัมพันธ์กันหรือไม่โดยการศึกษาค่าความสัมพันธ์ระหว่างตัวแปรที่ละคู่ ซึ่งตัวแปรแต่ละตัวอาจจำแนกออกเป็นหลายกลุ่มหรือหลายพวกที่แจกแจงอยู่ในตารางมิติต่าง ๆ เช่น 2x2, 3x2 หรือ 2x3 เป็นต้น เมื่อต้องการทดสอบความสัมพันธ์ระหว่างตัวแปรที่ละคู่จะต้องนำข้อมูลมาใส่ในตารางเพื่อหาความสัมพันธ์ระหว่างตัวแปรทั้งสอง สำหรับการทดสอบสมมติฐานว่าตัวแปรแต่ละคู่จะมีความสัมพันธ์กันหรือไม่ มีหลักการทดสอบความสัมพันธ์ระหว่างตัวแปรเพื่อให้สามารถหาค่าที่คาดหมายได้โดยการกำหนดสมมติฐานเป็นกลางว่าจะไม่มีความสัมพันธ์ระหว่างตัวแปรทั้งสอง

3. การทดสอบความเป็นเอกภาพ (Test of Homogeneity) การทดสอบความเป็นเอกภาพ หรือเรียกว่า การทดสอบความเป็นเอกพันธ์หรือการทดสอบความคล้ายคลึงกันของตัวแปร (Test of Homogeneity) เป็นการทดสอบความเหมือนกัน (หรือไม่แตกต่างกัน) ของตัวแปร โดยพิจารณาจากความน่าจะเป็นหรืออัตราส่วนของตัวแปรทั้งสอง ถ้ามีค่าใกล้เคียงกันแสดงว่าตัวแปรมีความเหมือนกัน (Campbell, 2007)

2.4.2 วิธีของเพียร์สัน

การวัดความสัมพันธ์ระหว่างตัวแปรด้วยวิธีของเพียร์สัน มีสูตรในการคำนวณดังนี้

$$C = \frac{\sqrt{\chi^2}}{\sqrt{\chi^2 + n}}$$

เมื่อ C = ค่าสัมประสิทธิ์ความสัมพันธ์ (มีค่าไม่เกิน 1)

N = จำนวนสมาชิก (ซึ่งจะต้องมีค่ามากกว่า 0)

จากสูตร ถ้า $C = 0$ แสดงว่าไม่มีความสัมพันธ์ระหว่างตัวแปรทั้งสอง และถ้า C ยิ่งมีค่ามากแสดงว่าระดับความสัมพันธ์ยิ่งมีค่ามากโดยที่ค่า C สูงสุดสามารถหาได้จากสูตรต่อไปนี้

$$C_{\max} = \sqrt{\left(\frac{k-1}{k}\right)}$$

เมื่อ C_{\max} = ค่าสูงสุดของสัมประสิทธิ์ของความสัมพันธ์

k = จำนวนของแถวหรือคอลัมน์ที่มีค่าน้อยที่สุด

2.5 งานวิจัยที่เกี่ยวข้อง

2.5.1 วราชัย สิงห์จันทริด (2550) ได้ทำการศึกษาถึงการวิเคราะห์อาการเสียวของฮาร์ดดิสก์ โดยการใช้เทคนิคของการทำเหมืองข้อมูลมาใช้ตัดแยกรูปแบบของเสียว เพื่อช่วยหาสาเหตุของอาการเสียวที่เกิดขึ้นในกระบวนการผลิต ในการพิจารณาลักษณะของตำแหน่งที่จุดเสียวเกิดขึ้นนั้น จะพิจารณาโดยการอาศัยเครื่องมือซึ่งได้นำเอาความรู้และเทคนิคการวิเคราะห์ข้อมูลโดยวิธีการทำเหมืองข้อมูล เพื่อช่วยเพิ่มความสามารถที่จะใช้ช่วยวิเคราะห์หารูปแบบของเสียวที่มาจากลักษณะการกระจายตัวของเสียวแบบต่างๆซึ่งสามารถบอกถึงแนวโน้มของสาเหตุของอาการเสียวในลักษณะต่างๆ จากการทดสอบปรากฏว่าเครื่องมือที่พัฒนาขึ้นสามารถทำการตัดแยกข้อมูลและมีความรวดเร็วในการวิเคราะห์สูง โดยการเปรียบเทียบกับวิธีการตัดแยกโดยผู้เชี่ยวชาญ

2.5.2 ประสิทธิ์ชัย บุญเสริม (2552) งานวิจัยนี้มุ่งเน้นในส่วนของการออกแบบระบบจำลองการทำนายเปอร์เซ็นต์ผลผลิตเป็นสำคัญ โดยในการทำนายนั้นจะเกิดขึ้นหลังกระบวนการวิเคราะห์สาเหตุและแก้ปัญหาเพื่อสนับสนุนกระบวนการดังกล่าว การทำนายผลจะยืนยันแนวทางการแก้ปัญหาที่ผ่านการวิเคราะห์ว่าสามารถนำไปปรับปรุงสายการผลิตได้ โดยได้ทำการพิสูจน์สมมติฐานทดสอบสมการถดถอยพหุคูณกับข้อมูลที่ได้มาจากกระบวนการผลิตฮาร์ดดิสก์ จากผลการทดลองให้ผลการทำนายที่มีความคลาดเคลื่อนสูงและข้อมูลที่ใช้ไม่มีความเป็นเส้นตรง ดังนั้นจึงได้เลือกใช้ทฤษฎีโครงข่ายประสาทเทียมแบบลุ่ม พบว่าสามารถใช้กับข้อมูลที่มีความซับซ้อนสูง รวมทั้งข้อมูลที่มีคุณสมบัติไม่เป็นเส้นตรงได้ ผลการทำนายที่ได้มีความคลาดเคลื่อนน้อยมาก ซึ่งสามารถสรุปได้ว่าตัวแบบการทำนายที่สร้างบนพื้นฐานทฤษฎีโครงข่ายประสาทเทียมเหมาะสมที่จะใช้เป็นระบบจำลองการทำนายในอุตสาหกรรมฮาร์ดดิสก์

2.5.3 สิริยาภรณ์ ไกรมาก (2552) เป็นการศึกษาการทำนายคุณภาพของผลิตภัณฑ์กาแฟคั่วและบดของโรงงานกรณีศึกษา ด้วยการสร้างแบบจำลองเพื่อทำนายคุณภาพเบื้องต้นก่อนการทดสอบโดยผู้เชี่ยวชาญในห้องปฏิบัติการ ซึ่งจะช่วยลดการสูญเสียทรัพยากร ได้แก่ เวลาและแรงงานที่ใช้ในระหว่างการทดสอบในห้องปฏิบัติการโดยผู้เชี่ยวชาญ โดยการตัดการทดลองใช้ปัจจัยในการผลิตที่แบบจำลองทำนายคุณภาพเบื้องต้นของผลิตภัณฑ์ทำนายว่า น่าจะให้ผลิตภัณฑ์สุดท้ายที่ด้อยคุณภาพออกจากการทดลองผลิต แบบจำลองที่นำมาใช้ทำนายคุณภาพของผลิตภัณฑ์กาแฟคั่วบดที่พิจารณามี 4 แบบ ได้แก่ แบบจำลองการถดถอยแบบพหุ แบบจำลองการถดถอยแบบโพลิโนเมียล แบบจำลองเครือข่ายประสาทเทียม และแบบจำลองเครือข่ายประสาทเทียมที่มีการนำเทคนิคพื้นผิวตอบสนองมาช่วยหาโครงสร้างที่เหมาะสมที่ให้ผลการทำนายที่ดีที่สุด โดยมีข้อมูลตัวแปรตามซึ่งเป็นคะแนนจากการทดสอบคุณลักษณะทางประสาทสัมผัส 5 ตัวแปร ได้แก่ สี กลิ่น ความขม รสชาติ และความชอบโดยรวมของน้ำกาแฟ ซึ่งผู้วิจัยได้ทำการวิเคราะห์ปัจจัยของตัวแปรตามทั้งห้านี้ เพื่อสร้างตัวแปรใหม่ที่ไม่มีความสัมพันธ์กัน และลดจำนวนตัวแปรลง ข้อมูลตัวแปรต้นประกอบด้วย ระยะเวลาในการเก็บการรักษามาเล็ดกาแฟก่อนนำมาคั่ว สี ความมันวาว กลิ่น ขนาดของมาเล็ดกาแฟ สัดส่วนของกาแฟพันธุ์อราบิก้าจาก 9 แหล่ง และสัดส่วนของกาแฟพันธุ์โรบัสต้าจาก 7 แหล่ง จากการศึกษาพบว่า การลดจำนวนตัวแปรตามด้วยการวิเคราะห์ปัจจัยไม่ได้ช่วยปรับปรุงความสามารถในการทำนายของเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบจำลอง นอกจากนี้พบว่าแบบจำลองเครือข่ายประสาทเทียมมีความสามารถในการทำนายค่าคะแนนคุณลักษณะทางประสาทสัมผัสของกาแฟคั่วและบดได้ถูกต้อง และใกล้เคียงกับค่าจริงมากกว่าแบบจำลองการถดถอยแบบพหุและแบบจำลองการถดถอยแบบโพลีโนเมียล และการสร้างแบบจำลองเครือข่ายประสาทเทียมด้วยเทคนิคพื้นผิวตอบสนอง สามารถช่วยลดระยะเวลาในการหาโครงสร้างที่เหมาะสมของแบบจำลองเครือข่ายประสาทเทียมลงได้

2.5.4 นิตาชล แซ่แห้ว (2555) สร้างตัวแบบสำหรับการพยากรณ์ปริมาณผลผลิตลำไยนอกฤดูภาค โดยใช้เทคนิคโครงข่ายประสาทเทียม และการวิเคราะห์ถดถอยพหุคูณ เพื่อพยากรณ์ปริมาณผลผลิตลำไยนอกฤดูภาค และเปรียบเทียบ ประเมินประสิทธิภาพของตัวแบบพยากรณ์ในการหาตัวแบบที่มีความแม่นยำ เหมาะสมสำหรับการนำไปใช้งาน โดยออกแบบแบบสอบถามในปัจจุบันที่มีผลต่อการผลิตลำไยนอกฤดู 23 ปัจจัย ซึ่งที่มาของปัจจัยนั้นมาจากการทบทวนวรรณกรรมและสอบถามผู้เชี่ยวชาญ จากนั้นจึงเก็บข้อมูลกรณีศึกษาจากกลุ่มเกษตรกรผู้ปลูกลำไยนอกฤดูภาค ในจังหวัดเชียงใหม่และลำพูน ทำการวิเคราะห์ข้อมูล และเตรียมข้อมูลเพื่อออกแบบตัวแบบพยากรณ์ ซึ่งผลจากการออกแบบตัวแบบพยากรณ์ มีร้อยละของค่าความคลาดเคลื่อนเฉลี่ยสมบูรณ์ (MAPE) ของวิธีการวิเคราะห์ถดถอยพหุคูณแบบปกติ แบบเดินหน้า แบบถอยหลัง แบบขั้นบันได และเทคนิคโครงข่ายประสาทเทียมตามลำดับดังนี้ 18.39, 25.63, 21.21, 25.63 และ 15.96 สรุปได้ว่าเทคนิคโครงข่ายประสาทเทียมมีประสิทธิภาพในการพยากรณ์ปริมาณผลผลิตลำไยนอกฤดูที่ดีที่สุด

บทที่ 3

วิธีการดำเนินงานวิจัย

จากวัตถุประสงค์ของงานวิจัย เพื่อสร้างแบบจำลองที่สามารถใช้ในการทำนายคุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ก่อนขั้นตอนการทดสอบคุณภาพความเชื่อมั่นก่อนการส่งมอบแก่ผู้บริโภค (ORT) โดยใช้ข้อมูลการทดสอบหัวอ่าน/เขียนและการทดสอบฮาร์ดดิสก์ งานวิจัยนี้มีขั้นตอนการดำเนินงานต่าง ๆ ดังนี้

3.1. ลักษณะของปัญหา

ตรวจสอบอาการผิดปกติของฮาร์ดดิสก์ที่เกิดขึ้นในกระบวนการทดสอบความเชื่อมั่นก่อนการส่งมอบแก่ผู้บริโภค โดยตรวจสอบข้อมูลทั้งหมดจำนวน 8 เดือน พบว่า ลักษณะอาการผิดปกติของฮาร์ดดิสก์ในหมวดหมู่ชุดหัวอ่าน/เขียน (HSA) มีจำนวนมากที่สุด ดังนั้นงานวิจัยนี้จึงมุ่งเน้นการสร้างแบบจำลองการพยากรณ์ผลการทดสอบคุณภาพของฮาร์ดดิสก์ในลักษณะอาการผิดปกติที่ชุดหัวอ่าน/เขียน

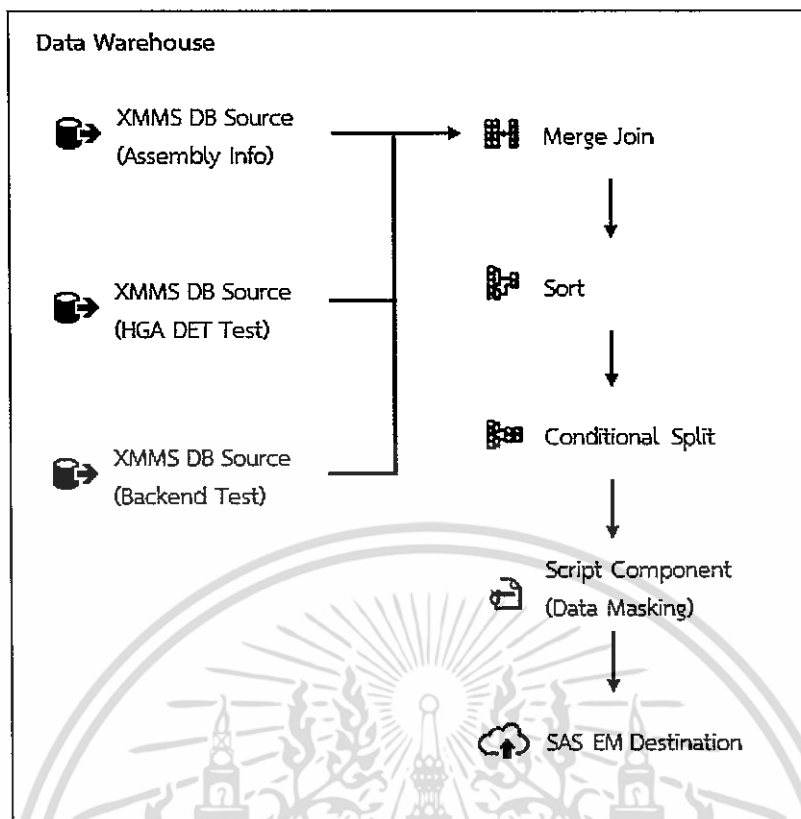
ตารางที่ 3.1 ลักษณะอาการผิดปกติของฮาร์ดดิสก์ในหมวดหมู่ต่างๆในจำนวนข้อมูล 8 เดือน

Drive Failure Category (หมวดหมู่ความผิดปกติของฮาร์ดดิสก์)	จำนวนฮาร์ดดิสก์	เปอร์เซ็นต์
หัวอ่านเขียน	468	75.12%
แผ่นดิสก์	77	12.36%
ไม่สามารถระบุข้อผิดพลาดได้	71	11.40%
เฟิร์มแวร์ควบคุม	4	0.64%
เฟิร์มแวร์เซอร์โว	2	0.32%
แมคคานิคัล	1	0.16%
รวม	623	100.00%

3.2 การเตรียมข้อมูล

คัดเลือกข้อมูลที่เกี่ยวข้อง ตรวจสอบความสมบูรณ์ของข้อมูล รวมข้อมูลจากหลายแหล่งข้อมูลและเปลี่ยนรูปแบบข้อมูลให้อยู่ในรูปแบบที่เหมาะสมกับเทคนิคการสร้างแบบจำลองที่เลือกใช้ รูปที่ 3.1 แสดงกระบวนการเตรียมข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.1 กระบวนการเตรียมข้อมูล

3.3 การแบ่งชุดข้อมูล

งานวิจัยนี้จะใช้ข้อมูลที่มีความสมบูรณ์ครบถ้วนเท่านั้น จึงทำการกำจัดข้อมูลที่ผิดปกติ (Outlier) ออกไป แล้วทำการแบ่งชุดข้อมูลออกเป็น 3 ชุดข้อมูล โดยทำการแบ่งชุดข้อมูลดังนี้

1. ชุดการเรียนรู้ (Training Data) เพื่อนำมาใช้ในการสร้างแบบจำลอง โดยใช้ข้อมูลทั้งหมด 60 เปอร์เซ็นต์
2. ชุดการทดสอบ (Validation Data) เพื่อประเมินความเพียงพอของแบบจำลองในการเปรียบเทียบผลการพยากรณ์ของแบบจำลอง และใช้ในการปรับปรุงแบบจำลอง โดยใช้ข้อมูลทั้งหมด 20 เปอร์เซ็นต์
3. ชุดการทวนสอบ (Testing Data) ใช้ในการตรวจสอบแบบจำลองว่ามีความถูกต้องมากน้อยเพียงใด โดยใช้ข้อมูลทั้งหมด 20 เปอร์เซ็นต์

โดยแต่ละชุดข้อมูลจะมีสัดส่วนของฮาร์ดติสก์ที่มีความผิดปกติเท่ากันทั้งสามชุด ดังแสดงในตารางที่ 3.2 แสดงสัดส่วนการแบ่งชุดข้อมูล ซึ่งจะเห็นได้ว่าข้อมูลทั้งสามชุดมีอัตราส่วนฮาร์ดติสก์ผิดปกติที่เท่ากันทุกชุดข้อมูล โดยจำนวนข้อมูลทั้งหมดคือ จำนวนข้อมูลที่เหลือหลังจากทำการกำจัดข้อมูลที่ไม่สมบูรณ์ออกไปแล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 สัดส่วนจำนวนข้อมูลของชุดข้อมูลการเรียนรู้ ชุดทวนสอบ และชุดทดสอบ

ชุดข้อมูล (Data Set)	ฮาร์ดดิสก์ผิดปกติ	ฮาร์ดดิสก์ปกติ	รวม	อัตราส่วนของ ฮาร์ดดิสก์ผิดปกติ	สัดส่วนข้อมูล
เรียนรู้ (Train)	445	100314	100759	0.44%	60.00%
ทดสอบ (Validate)	148	33438	33586	0.44%	20.00%
ทวนสอบ (Test)	149	33440	33589	0.44%	20.00%
รวม Total	742	167192	167934	0.44%	100.00%

3.4 การทดสอบความสัมพันธ์ระหว่างตัวแปร

เนื่องจากข้อมูลที่ใช้ทำการวิจัยมีปริมาณตัวแปรอิสระเป็นจำนวนมาก จึงจะทำการคัดเลือกตัวแปรอิสระทั้งหมด 197 ตัวแปร ที่ส่งผลกระทบต่อหรือมีความสัมพันธ์กับฮาร์ดดิสก์ที่มีอาการผิดปกติที่หัวอ่าน/เขียน ก่อนที่นำตัวแปรอิสระดังกล่าวไปสร้างแบบจำลอง วิธีการคัดเลือกตัวแปรที่จะใช้คือวิธีไคสแควร์ เป็นการทดสอบความสัมพันธ์ระหว่างตัวแปร (Test of Association) หรือเป็นการทดสอบว่าตัวแปรต่างๆ มีความสัมพันธ์กับคุณภาพของฮาร์ดดิสก์หรือไม่ โดยทำการศึกษาความสัมพันธ์ดังกล่าวระหว่างตัวแปรที่ละคู่และกำหนดค่าจำนวนการแบ่งหมวดหมู่ (Number of Bins) เท่ากับ 10 โดยพิจารณาจากค่า P-Value คือเมื่อตัวแปรอิสระตัวใดมีความสัมพันธ์กับคุณภาพของฮาร์ดดิสก์จะให้ค่า P-Value ที่น้อยกว่า 0.05 หรือมีนัยสำคัญ จะนำตัวแปรอิสระตัวนั้นมาสร้างแบบจำลองต่อไป จากงานวิจัยมีตัวแปรอิสระทั้งหมด 25 ตัวแปรที่มีค่า P-Value < 0.05 จากตัวแปรอิสระทั้งหมด 197 ตัวแปร ดังแสดงในตารางที่ 3.3

ตารางที่ 3.3 ค่าไคสแควร์ และค่า P-Value ที่ได้จากการทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระ กับคุณภาพของฮาร์ดดิสก์

Input	Chi-Square	Df	Prob	Input	Chi-Square	Df	Prob
X102	540.0478	9	<.0001	X176	68.5231	7	<.0001
X49	452.5088	9	<.0001	X38	65.4601	8	<.0001
X147	261.8552	9	<.0001	X178	64.9795	7	<.0001
X54	234.5153	8	<.0001	X58	61.6656	9	<.0001
X94	127.9397	9	<.0001	X136	54.7973	9	<.0001
X28	111.2103	5	<.0001	X121	49.3637	6	<.0001
X133	106.6299	7	<.0001	X111	38.4993	7	<.0001
X35	94.6792	9	<.0001	X53	34.6681	9	<.0001
X50	90.9331	9	<.0001	X146	31.8817	9	0.0002
X16	86.4445	9	<.0001	X157	21.8255	6	0.0013
X34	77.509	6	<.0001	X52	17.7995	9	0.0376
X51	73.1884	9	<.0001	X61	8.3238	2	0.0156
X73	69.5486	9	<.0001				

3.4.1 ตรวจสอบปัญหาตัวแปรอิสระที่มีความสัมพันธ์กันเอง (Multicollinearity)

เนื่องจากการวิเคราะห์การถดถอยโลจิสติกมีข้อตกลงเบื้องต้นที่ว่า ตัวแปรอิสระจะต้องไม่มีความสัมพันธ์กันหรือไม่มีปัญหา Multicollinearity เพื่อลดความซ้ำซ้อนและขนาดของตัวแปรที่ใหญ่เกินความจำเป็น จึงทำการหาความสัมพันธ์ระหว่างตัวแปรอิสระแต่ละตัวดังในตารางที่ 3.4 แสดงความสัมพันธ์ของตัวแปรอิสระทั้ง 25 ตัวแปร จากความสัมพันธ์ระหว่างตัวแปรอิสระทั้ง 25 ตัวแปร โดยกำหนดให้ตัวแปรอิสระที่มีความสัมพันธ์กันตั้งแต่ 80 เปอร์เซ็นต์ขึ้นไปถือว่ามีความสัมพันธ์กันมาก ซึ่งจะต้องทำการแก้ไขตัวแปรเหล่านี้ ด้วยวิธีการสร้างตัวแปรใหม่ด้วยการวิเคราะห์องค์ประกอบหลัก

ตารางที่ 3.4 ความสัมพันธ์ของตัวแปรอิสระทั้ง 25 ตัวแปร

Variable	X102	X111	X121	X133	X136	X146	X147	X175	X176	X178	X28	X34	X35	X38	X49	X50	X51	X52	X53	X54	X58	X61	X73	X94
X102																								
X111	-0.1379																							
X121	0.0341	-0.0174																						
X133	0.1699	0.0726	0.2754																					
X136	-0.0450	0.1869	0.2850	0.1616																				
X146	-0.0032	-0.1470	0.1869	0.1616	0.6387																			
X147	-0.0369	0.2878	0.2850	0.1616	0.6387	0.7065																		
X175	0.0075	0.0051	0.0027	0.0051	0.0020	0.0020	0.0020																	
X176	0.0452	-0.0357	0.0344	0.1092	-0.0054	0.0028	0.0028	-0.0054																
X178	-0.0087	0.1613	0.1749	0.1488	0.0912	0.0098	0.0098	0.0098	0.0124															
X179	0.0075	0.0344	0.1092	0.1092	0.0904	0.4988	0.4988	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
X28	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037	-0.0037
X34	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647	0.0647
X35	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453	0.0453
X38	-0.0002	0.0284	-0.0133	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098
X49	-0.2281	0.1230	-0.0396	-0.1492	-0.0097	-0.1005	-0.0281	-0.0090	-0.0692	-0.0094	-0.0094	-0.0094	-0.0094	-0.0094	-0.0094	-0.0094	-0.0094	-0.0094	-0.0094	-0.0094	-0.0094	-0.0094	-0.0094	-0.0094
X50	-0.4101	0.0315	-0.0098	-0.0456	-0.0160	-0.4802	-0.6818	0.0116	-0.2428	0.3293	0.3293	0.3293	0.3293	0.3293	0.3293	0.3293	0.3293	0.3293	0.3293	0.3293	0.3293	0.3293	0.3293	0.3293
X51	0.0208	0.1975	-0.1364	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098
X52	0.0302	0.1975	-0.1364	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098
X53	0.0439	0.1975	-0.1364	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098
X54	0.0184	0.1655	-0.1364	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098
X58	0.0320	0.1918	-0.1340	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098	-0.0098
X61	-0.0306	0.0062	-0.0011	0.0062	-0.0049	-0.0005	-0.0005	-0.0045	-0.0078	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058
X73	-0.0066	0.0062	-0.0011	0.0062	-0.0049	-0.0005	-0.0005	-0.0045	-0.0078	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058	0.0058
X94	-0.3668	0.1218	-0.0584	-0.1321	-0.0357	-0.0236	-0.0399	-0.0061	-0.0387	0.0315	0.0315	0.0315	0.0315	0.0315	0.0315	0.0315	0.0315	0.0315	0.0315	0.0315	0.0315	0.0315	0.0315	0.0315

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยพิจารณาจากค่าความสัมพันธ์ของตัวแปรด้วยทฤษฎีหรือความรู้ทางเทคนิคของผลิตภัณฑ์ว่าตัวแปรใดควรจะมีความสำคัญมากกว่าก็เลือกตัวแปรนั้น เพื่อไม่ให้เกิดปัญหา Multicollinearity ในงานวิจัยนี้จะดำเนินสร้างตัวแปรอิสระตัวใหม่ขึ้นมาแทนตัวแปรอิสระที่มีความสัมพันธ์กันเอง ซึ่งจากตารางที่ 3.4 จะเห็นตัวแปรอิสระ X_{136} และ X_{147} มีความสัมพันธ์กันมาก จึงจะรวมตัวแปรอิสระทั้ง 2 ตัวแปรนี้เป็นตัวแปรอิสระใหม่ และยังคงมีตัวแปรใหม่อีกสามตัวที่สร้างจากตัวแปรอิสระดัง แสดงในตารางที่ 3.5 แสดงตัวแปรอิสระที่นำไปสร้างตัวแปรอิสระตัวใหม่ ด้วยวิธีการการวิเคราะห์องค์ประกอบหลัก

ตารางที่ 3.5 ตัวแปรอิสระใหม่ที่สร้างจากตัวแปรอิสระที่มีความสัมพันธ์กันด้วยวิธีการวิเคราะห์องค์ประกอบหลัก

ชื่อตัวแปรอิสระใหม่	ตัวแปรอิสระเดิมที่ใช้สร้างตัวแปรอิสระใหม่
X_{136N}	X_{136}, X_{147}
X_{16N}	X_{16}, X_{34}, X_{35}
X_{49N}	X_{49}, X_{50}
X_{51N}	$X_{51}, X_{52}, X_{53}, X_{54}, X_{58}$

หลังจากสร้างตัวแปรใหม่แล้วจะทำการตรวจสอบหา Multicollinearity อีกครั้ง จากการวิเคราะห์พบว่าไม่มีตัวแปรใดที่มีความสัมพันธ์กันเองหลงเหลืออยู่ ดังนั้นหลังจากแก้ไขปัญหานี้จะทำให้เหลือตัวแปรอิสระที่จะนำไปสร้างแบบจำลองการวิเคราะห์การถดถอยโลจิสติกจำนวน 17 ตัวแปร ดังแสดงในตารางที่ 3.6

ตารางที่ 3.6 ความสัมพันธ์ของตัวแปรอิสระจำนวน 17 ตัวแปรหลังจากแก้ปัญหา Multicollinearity

VARIABLE	X73	X94	X102	X133	X146	X111	X121	X175	X176	X178	X28	X38	X61	X49N	X16N	X51N	X136N
X73		-0.0022	-0.0066	0.0062	-0.0649	0.0062	-0.0011	0.0045	0.0438	0.0469	0.0007	-0.0218	-0.0022	0.0345	-0.0126	0.0212	-0.0170
X94	-0.0022		-0.3668	-0.1321	-0.0236	0.1218	-0.0584	-0.0061	0.0315	0.0241	-0.0044	0.0020	0.0204	0.2817	-0.0372	0.0585	-0.0374
X102	-0.0066	-0.3668		0.1609	-0.0052	-0.1379	0.0341	0.0075	-0.0204	-0.0087	-0.0037	-0.0002	-0.0306	-0.5898	0.0527	0.0296	-0.0410
X133	0.0062	-0.1321	0.1609		0.1616	0.0726	0.2754	0.0051	-0.0213	-0.0189	0.0026	-0.0053	-0.0150	-0.0999	0.0957	-0.1408	0.2872
X146	-0.0649	-0.0236	-0.0052	0.1616		-0.2257	0.1256	-0.0019	-0.1375	-0.1235	0.0113	-0.0760	-0.0005	-0.1379	0.1183	-0.4765	0.6743
X111	0.0062	0.1218	-0.1379	0.0726	-0.2257		-0.0174	-0.0055	0.1798	0.1613	-0.0047	0.0284	0.0068	0.0801	-0.0241	0.1764	-0.1679
X121	-0.0011	-0.0584	0.0341	0.2754	0.1256	-0.0174		-0.0027	-0.0473	-0.0515	0.0073	-0.0133	-0.0029	-0.0256	0.0291	-0.1349	0.1904
X175	0.0045	-0.0061	0.0075	0.0051	-0.0019	-0.0055	-0.0027		0.0012	0.0030	0.0005	-0.0004	0.0023	-0.0150	-0.0076	0.0152	0.0049
X176	0.0438	0.0315	-0.0204	-0.0213	-0.1375	0.1798	-0.0473	0.0012		0.4988	-0.0455	0.0528	0.0076	0.0003	0.0251	0.3966	-0.2121
X178	0.0469	0.0241	-0.0087	-0.0189	-0.1235	0.1613	-0.0515	0.0030	0.4988		-0.0640	0.0320	0.0058	-0.0012	-0.0803	0.3718	-0.1860
X28	0.0007	-0.0044	-0.0037	0.0026	0.0113	-0.0047	0.0073	0.0005	-0.0455	-0.0640		0.0020	0.0238	-0.0037	-0.0010	-0.0203	0.0131
X38	-0.0218	0.0020	-0.0002	-0.0053	-0.0760	0.0284	-0.0133	-0.0004	0.0528	0.0320	0.0020		0.0061	0.0209	-0.0566	0.0719	-0.0683
X61	-0.0022	0.0204	-0.0306	-0.0150	-0.0005	0.0068	-0.0029	0.0023	0.0076	0.0058	0.0238	0.0061		0.0204	-0.0075	0.0019	-0.0050
X49N	0.0345	0.2817	-0.5898	-0.0999	-0.1379	0.0801	-0.0256	-0.0150	0.0003	-0.0012	-0.0037	0.0209	0.0204		-0.0658	-0.0003	-0.0247
X16N	-0.0126	-0.0372	0.0527	0.0957	0.1183	-0.0241	0.0291	-0.0076	0.0251	-0.0803	-0.0010	-0.0566	-0.0075	-0.0658		-0.1810	0.1298
X51N	0.0212	0.0585	0.0296	-0.1408	-0.4765	0.1764	-0.1349	0.0152	0.3966	0.3718	-0.0203	0.0719	0.0019	-0.0003	-0.1810		-0.6532
X136N	-0.0170	-0.0374	-0.0410	0.2872	0.6743	-0.1679	0.1904	0.0049	-0.2121	-0.1860	0.0131	-0.0683	-0.0050	-0.0247	0.1298	-0.6532	

3.5 การสร้างแบบจำลอง

นำตัวแปรอิสระ 25 ตัวแปร ที่ได้จากขั้นตอนก่อนหน้า มาสร้างแบบจำลองเพื่อพยากรณ์คุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ โดยวิธีโครงข่ายประสาทเทียม และนำตัวแปรอิสระจำนวน 17 ตัวแปรหลังจากที่กำจัดปัญหา Multicollinearity แล้ว มาสร้างแบบจำลองการถดถอยโลจิสติก

3.5.1 สร้างแบบจำลองการวิเคราะห์การถดถอยโลจิสติกด้วยโปรแกรม SAS Enterprise Miner Workstation 13.2

การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression) โดยเลือกรูปแบบการวิเคราะห์การถดถอยแบบเป็นขั้นตอน (Stepwise) และใช้เกณฑ์ในการคัดเลือกตัวแปรอิสระเป็น Cross Validation Misclassification ซึ่งก็คือจะพิจารณาจากค่าอัตราการจำแนกผิด (Misclassification) ของการทำการทดสอบซ้ำๆ (Cross Validation) โดยจะเลือกแบบจำลองที่เหมาะสมจากอัตราการจำแนกผิดกลุ่มที่น้อยที่สุด

3.5.2 สร้างแบบจำลองโครงข่ายประสาทเทียมด้วยโปรแกรม SAS Enterprise Miner Workstation 13.2

การวิเคราะห์โครงข่ายประสาทเทียม (Neural Network) โดยจะกำหนดค่าดังนี้

1. โครงข่าย (Network) เลือกสถาปัตยกรรม (Architecture) แบบเปอร์เซพตรอนหลายชั้น (Multilayer Perceptron)
2. จำนวนหน่วยซ่อน (Number of hidden Unit) ในงานวิจัยนี้ จะกำหนดหน่วยซ่อนเป็น 2 ถึง 10 หน่วยตามลำดับแล้วทำการเปรียบเทียบผลลัพธ์ที่ได้เพื่อเลือกจำนวนหน่วยซ่อนที่ดีที่สุด (สามารถกำหนดค่าหน่วยซ่อนได้ตั้งแต่ 1 จนถึง 64 หน่วย)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

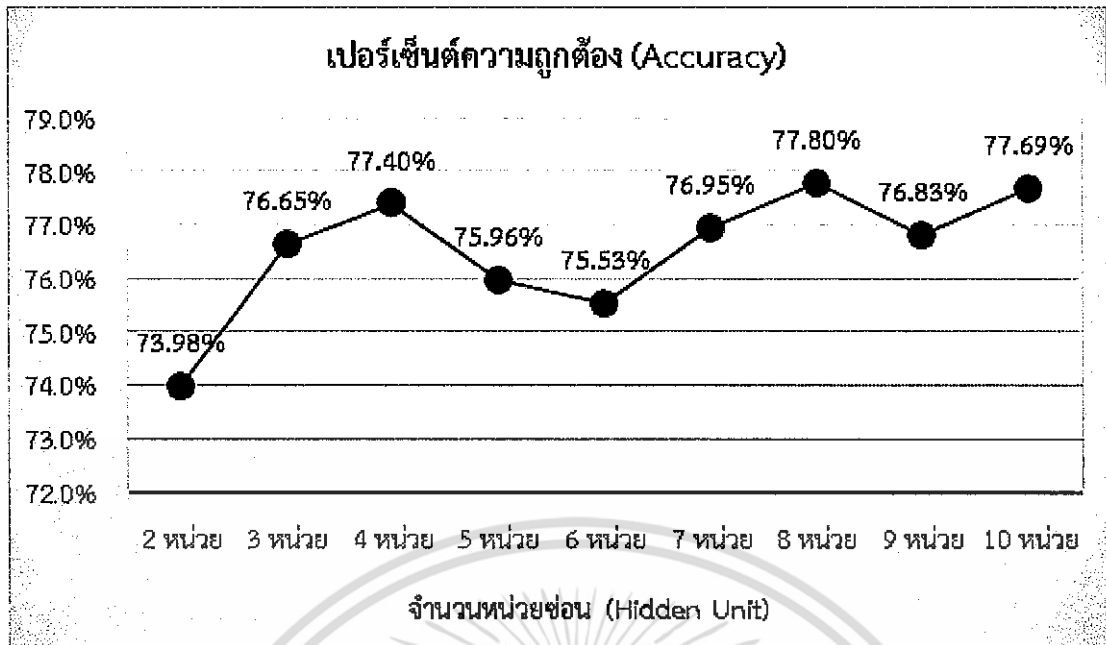
ในการกำหนดจำนวนหน่วยซ่อนจะไม่มีทฤษฎีตายตัว โดยส่วนใหญ่จะทำการทดลองซ้ำๆ โดยการเปลี่ยนจำนวนหน่วยซ่อนไปเรื่อยๆจนกว่าจะได้ข้อมูลที่เหมาะสม ซึ่งมีหลักการกำหนดค่าดังนี้คือจะกำหนดจำนวนหน่วยซ่อนไปเรื่อยๆจนถึงค่า $n-1$ เมื่อ n คือจำนวนตัวแปรอิสระ (Satish K, 2012) จากการศึกษาถึงความสัมพันธ์ของตัวแปรอิสระที่มีต่อตัวแปรตามพบว่า มีตัวแปรอิสระจำนวน 25 ตัวแปรที่เข้าเงื่อนไขการคัดเลือกตัวแปร ดังนั้นจึงสามารถทำการทดลองซ้ำๆโดยกำหนดจำนวนหน่วยซ่อนได้ตั้งแต่ 2-24 หน่วยได้ แต่เนื่องจากว่าการกำหนดหน่วยซ่อนสูงๆนั้นจะทำให้แบบจำลองมีการเรียนรู้ข้อมูลมากเกินไปแต่ไม่สามารถพยากรณ์ข้อมูลที่ใช้ทำการทดสอบได้หรือเรียกว่า Over fitting นั้นเอง ดังนั้นในงานวิจัยนี้จะเลือกทดสอบค่าหน่วยซ่อนที่ 2 ถึง 10 หน่วยเท่านั้น และเลือกจำนวนหน่วยซ่อนที่ให้ค่าการพยากรณ์ที่ดีที่สุด เพื่อนำไปเปรียบเทียบกับการวิเคราะห์การถดถอยต่อไป

3. การกระจายของกลุ่มตัวอย่าง (Randomization Distribution) ใช้ในการกำหนดค่าการกระจายเพื่อนำไปใช้กับการปรับปรุงค่าน้ำหนัก โดยจะเลือกเป็นการกระจายแบบปกติ (Normal Distribution)

เมื่อกำหนดค่าโครงข่ายเรียบร้อยแล้วจะใช้เกณฑ์ในการคัดเลือกแบบจำลองเป็นอัตราการทำผิดกลุ่ม (Misclassification) โดยจะเลือกแบบจำลองที่เหมาะสมจากอัตราการทำผิดกลุ่มที่น้อยที่สุดของชุดข้อมูลการทดสอบ (Validation Data)

ตารางที่ 3.7 ผลลัพธ์การปรับค่าจำนวนหน่วยซ่อนของโครงข่ายประสาทเทียม

จำนวนหน่วยซ่อน Hidden Unit	ผลลัพธ์ (Result)				เกณฑ์การวัด (Measurements)	
	FN	TN	FP	TP	Accuracy	Recall
H2	84	74180	26134	361	73.98%	81.12%
H3	92	76883	23431	353	76.65%	79.33%
H4	95	77638	22676	350	77.40%	78.65%
H5	94	76188	24126	351	75.96%	78.88%
H6	102	75763	24551	343	75.53%	77.08%
H7	86	77175	23139	359	76.95%	80.67%
H8	83	78033	22281	362	77.80%	81.35%
H9	80	77044	23270	365	76.83%	82.02%
H10	77	77912	22402	368	77.69%	82.70%



รูปที่ 3.2 กราฟแสดงความสัมพันธ์ระหว่างหน่วยซ่อนและเปอร์เซ็นต์ความถูกต้องของการพยากรณ์

จากผลการทดลองปรับค่าหน่วยซ่อนของโครงข่ายประสาทเทียมดังแสดงในตารางที่ 3.7 เมื่อพิจารณาจากค่าความถูกต้องในการพยากรณ์คุณภาพของฮาร์ดดิสก์ (Accuracy) ในรูปที่ 3.2 จะเห็นว่าค่าความถูกต้องมีค่าเพิ่มขึ้นเมื่อเพิ่มจำนวนหน่วยซ่อนไปจนถึง 4 หน่วยหลังจากนั้นจะมีค่าลดลงเหมือนเป็นวัฏจักร (Cycle) และพิจารณาค่าน้ำหนักความถูกต้องมีค่าลดลงที่จำนวนหน่วยซ่อน 5 หน่วย ก่อนที่จะเพิ่มขึ้นอีกครั้งและเมื่อใช้จำนวนหน่วยซ่อน 8 หน่วยจะให้ค่าความถูกต้องสูงสุด แต่งานวิจัยนี้จะไม่เลือกหน่วยซ่อนนี้เนื่องจากถ้าเลือกหน่วยซ่อนนี้จะเกิดปัญหา Over fitting คือเหตุการณ์ที่โครงข่ายประสาทเทียม (Neural Networks) เรียนรู้ข้อมูลได้ดีเกินไปในระหว่างกระบวนการเรียนรู้ (Training) แต่ทำข้อผิดพลาดอย่างมากในระหว่างกระบวนการทดสอบ (Test) หรือสรุปได้ว่าถ้าใช้หน่วยซ่อนใหญ่เกินความจำเป็นจะทำให้เกิดปัญหา Over Fitting ได้ (Satish K, 2012) ดังนั้นหน่วยซ่อนที่ให้ค่าความถูกต้องที่สูงสุดในวัฏจักรแรกจะถูกเลือกเนื่องจากการป้องกันการเกิดปัญหา Over fitting เมื่อเลือกหน่วยซ่อนมากเกินความจำเป็น ดังนั้นงานวิจัยนี้จะเลือกกำหนดค่าหน่วยซ่อนเท่ากับ 4 หน่วย

เมื่อทำการกำหนดค่าที่เหมาะสมที่สุดได้แล้ว หลังจากนั้นจะทำการสร้างแบบจำลองในการพยากรณ์ฮาร์ดดิสก์ในแต่ละแบบคือ การวิเคราะห์การถดถอยโลจิสติกและโครงข่ายประสาทเทียม นำผลลัพธ์จากทั้งสองวิธีการมาเปรียบเทียบกันเพื่อหาแบบจำลองที่ให้ประสิทธิภาพในการพยากรณ์ผลิตภัณฑ์ฮาร์ดดิสก์ที่ดีที่สุด

3.6 ประเมินความถูกต้อง

ประเมินความถูกต้องของแบบจำลองที่ได้จากการวิเคราะห์การถดถอยแบบโลจิสติก และโครงข่ายประสาทเทียม โดยนำมาเปรียบเทียบกันทั้งผลการทดลองของชุดข้อมูลการเรียนรู้ ชุดการทดสอบ และชุดการทวนสอบ เพื่อคัดเลือกตัวแบบที่ให้ผลการพยากรณ์ที่ดีที่สุด ซึ่งสามารถพิจารณาได้ดังต่อไปนี้

Confusion Matrix เป็นการเก็บข้อมูลที่เกี่ยวกับการแบ่งแยกข้อมูลจริง กับข้อมูลที่เกิดจากการพยากรณ์ ด้วยระบบการแบ่งแยก (Classification System) เช่น การหาค่าข้อมูลแบบธรรมดาในแมทริก ดังตารางของ Confusion Matrix ซึ่งมีด้วยกัน 2 ระดับ

		Prediction Outcome		
		P	N	
Actual Value	P*	True Positive (TP)	False Negative (FN)	P*
	N*	False Positive (FP)	True Negative (TN)	N*
Total		P	N	

รูปที่ 3.3 ลักษณะการแบ่งแยกข้อมูลจริงกับข้อมูลที่เกิดจากการทำนาย (Confusion Matrix)

โดยในงานวิจัยนี้จะพิจารณาจากค่า Accuracy คือเปอร์เซ็นต์ความถูกต้อง (Overall Percentage Correct) ร้อยละของการพยากรณ์ถูกต้องโดยรวมทั้งกรณีที่เกิดเหตุการณ์และไม่เกิดเหตุการณ์ และค่าพยากรณ์การเกิดเหตุการณ์ที่ถูกต้อง (Recall)

3.7 พัฒนาโปรแกรมทดสอบและประยุกต์ใช้ตัวแบบ

โดยการนำตัวแบบที่ดีที่สุดมาพัฒนาต่อเป็นโปรแกรมทดสอบคุณภาพผลิตภัณฑ์ฮาร์ดดิสก์เพื่อการทำนายคุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ก่อนขั้นตอนการทดสอบความเชื่อมั่นก่อนการส่งมอบแก่ผู้บริโภค (ORT) ในกระบวนการผลิตฮาร์ดดิสก์ และทำการเปรียบเทียบผลการทดสอบฮาร์ดดิสก์ระหว่างก่อนการประยุกต์ใช้แบบจำลองและหลังการประยุกต์ใช้แบบจำลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิจัย

หลังจากทำการศึกษาลักษณะของปัญหาแล้วพบว่าลักษณะอาการผิดปกติของฮาร์ดดิสก์ที่มีเป็นจำนวนมากได้แก่ อาการผิดปกติที่ซุดหัวอ่าน/เขียน และได้ทำการเตรียมข้อมูลในการสร้างแบบจำลองทั้งสิ้น 167,934 ข้อมูล มีตัวแปรอิสระทั้งสิ้น 197 ตัวแปร ซึ่งแบ่งเป็นชุดการเรียนรู้ 60 เปอร์เซ็นต์ ชุดการทดสอบ 20 เปอร์เซ็นต์ และชุดการทวนสอบ 20 เปอร์เซ็นต์ คัดเลือกตัวแปรเบื้องต้นด้วยวิธีโคสแควร์พบว่า มี 25 ตัวแปรที่มีนัยสำคัญกับฮาร์ดดิสก์ผิดปกติที่ซุดหัวอ่าน/เขียน โดยจะนำทั้ง 25 ตัวแปรนี้ไปสร้างแบบจำลองโครงข่ายประสาทเทียม ส่วนแบบจำลองการถดถอยโลจิสติกจะต้องนำตัวแปรทั้ง 25 ตัวแปรไปกำจัดปัญหา Multicollinearity ก่อนจึงจะนำไปสร้างแบบจำลองได้ หลังจากกำจัดปัญหา Multicollinearity ออกไปแล้วจะคงเหลือตัวแปร 17 ตัวแปรที่จะนำไปสร้างแบบจำลองการถดถอยโลจิสติก ในบทนี้จะแสดงผลการวิจัยของแบบจำลองแต่ละประเภทคือการวิเคราะห์การถดถอยโลจิสติกและโครงข่ายประสาทเทียม ดังต่อไปนี้

4.1 ผลการวิเคราะห์

จากการวิเคราะห์การถดถอยโลจิสติกและโครงข่ายประสาทเทียมของข้อมูลทั้งสามชุดคือ ชุดการเรียนรู้ ชุดการทดสอบและชุดการทวนสอบ มีผลลัพธ์ดังนี้

4.1.1 ผลการพยากรณ์ของแบบจำลองการวิเคราะห์การถดถอยโลจิสติก

จากตาราง 4.1 แสดงผลการพยากรณ์ด้วยวิธีการวิเคราะห์การถดถอยโลจิสติก ซึ่งจะเห็นว่าค่าความถูกต้องและค่าความไวของแต่ละชุดข้อมูลมีค่าใกล้เคียงกัน ซึ่งจะแสดงให้เห็นว่าไม่มีปัญหาการเข้ากันเกินไป (Over fitting) ซึ่งก็คือการที่แบบจำลองสามารถพยากรณ์ได้ดีในชุดการเรียนรู้แต่ไม่สามารถพยากรณ์ได้ดีในชุดการทวนสอบนั่นเอง (Chen. et al, 2004) ดังนั้นผลลัพธ์ที่ได้จากการพยากรณ์ของแบบจำลองจากการวิเคราะห์โลจิสติกสามารถนำมาใช้ในการเปรียบเทียบกับผลการวิเคราะห์ด้วยวิธีโครงข่ายประสาทเทียมได้

ตารางที่ 4.1 ผลการพยากรณ์คุณภาพของฮาร์ดดิสก์ด้วยวิธีการวิเคราะห์การถดถอยโลจิสติกของชุดการเรียนรู้ ชุดทดสอบ และชุดทวนสอบ

ชุดข้อมูล	ผลลัพธ์ (Result)				เกณฑ์การวัด (Measurement)	
	FN	TN	FP	TP	Accuracy	Recall
ชุดการเรียนรู้	110	72337	27977	335	72.37%	75.28%
ชุดการทดสอบ	40	24152	9286	108	72.47%	72.97%
ชุดการทวนสอบ	41	24089	9351	108	72.27%	72.48%
รวม	191	120578	46614	551	72.37%	74.26%

ตารางที่ 4.2 ตัวแปรอิสระที่ถูกเลือกโดยการวิเคราะห์การถดถอยโลจิสติกแบบเป็นขั้นตอน
(Stepwise logistics Regression)

Analysis of Maximum Likelihood Estimates							
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept	1	24.7079	1.5638	249.64	<.0001		999
X102	1	-0.0555	0.00311	317.37	<.0001	-0.4659	0.946
X111	1	0.1434	0.0268	28.6	<.0001	0.1101	1.154
X133	1	-0.1523	0.0388	15.4	<.0001	-0.0882	0.859
X136N	1	-0.1752	0.0332	27.91	<.0001	-0.1363	0.839
X16N	1	-0.0786	0.027	8.47	0.0036	-0.0735	0.924
X176	1	-0.1588	0.0239	44.04	<.0001	-0.118	0.853

ตัวแปรอิสระหรือปัจจัยที่ส่งผลกระทบต่อคุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ที่มีอาการผิดปกติของหัวอ่านเขียนคือ X_{102} , X_{111} , X_{133} , X_{136N} , X_{16N} และ X_{176} รวมทั้งสิ้น 6 ตัวแปร โดยได้ทำการตรวจสอบปัจจัยดังกล่าวกับผู้เชี่ยวชาญด้านเทคนิคของผลิตภัณฑ์เป็นที่เรียบร้อยแล้ว ปัจจัยเหล่านี้มีความสำคัญกับคุณภาพของฮาร์ดดิสก์จริง และเมื่อพิจารณาจากค่าการทดสอบทางสถิติ P-value ของการวิเคราะห์การถดถอยโลจิสติกแบบเป็นขั้นตอน (Stepwise) ในขั้นตอนสุดท้าย จากตารางที่ 4.2 จะเห็นได้ว่า ตัวแปรทั้ง 6 ตัวแปร มีค่า P-value น้อยกว่า 0.05 แสดงว่าตัวแปรเหล่านี้มีผลกระทบต่อหรือมีความสัมพันธ์กับคุณภาพของฮาร์ดดิสก์อย่างมีนัยสำคัญ จากค่าสัมประสิทธิ์ที่ได้สามารถเขียนสมการการวิเคราะห์การถดถอยโลจิสติก เพื่อพยากรณ์ฮาร์ดดิสก์ผิดปกติได้ดังนี้

$$P(\text{ฮาร์ดดิสก์ผิดปกติ}) = \frac{1}{1 + e^{-f(x)}}$$

โดยที่

$$f(x) = -0.0555X_{102} + 0.1434X_{111} - 0.1523X_{133} - 0.1751X_{136N} - 0.0787X_{16N} - 0.1588X_{176}$$

เมื่อ $f(x)$ คือ ฟังก์ชันเชิงเส้นของตัวแปรอิสระ

P (ฮาร์ดดิสก์ผิดปกติ) คือ ความน่าจะเป็นของการเกิดเหตุการณ์หรือความน่าจะเป็นของฮาร์ดดิสก์ผิดปกติ

e คือ 2.71828

X คือ ตัวแปรอิสระ

4.1.2 ผลการพยากรณ์ของแบบจำลองการวิเคราะห์โครงข่ายประสาทเทียม

ผลการพยากรณ์ด้วยการวิเคราะห์โครงข่ายประสาทเทียมแบบหลายชั้นกำหนดหน่วยซ่อนเท่ากับ 4 หน่วยซ่อน อัตราการเรียนรู้ที่เริ่มต้น 0.1 จำนวนตัวแปรอิสระหรือตัวแปรนำเข้า 25 ตัวแปร ได้ผลดังตารางที่ 4.3 จะเห็นว่าผลลัพธ์ที่ได้มีค่าความถูกต้องและค่าระลึกของแต่ละชุดข้อมูลมีความใกล้เคียงกันหรือไม่มีปัญหาการเข้ากันเกินไป (Over fitting) เช่นเดียวกับการวิเคราะห์การถดถอยโลจิสติก ดังนั้นสามารถนำผลการวิเคราะห์นี้ไปเปรียบเทียบกับการวิเคราะห์การถดถอยโลจิสติกได้ โดยสามารถดูสมการของแบบจำลองโครงข่ายประสาทเทียมได้ในภาคผนวก ข รวมทั้งผลลัพธ์ของแต่ละขั้นตอนของกระบวนการวิจัยอย่างละเอียด

ตารางที่ 4.3 ผลการพยากรณ์คุณภาพของฮาร์ดดิสก์ด้วยวิธีโครงข่ายประสาทเทียมของชุดเรียนรู้ ชุดทดสอบ และชุดทวนสอบ

การวิเคราะห์โครงข่ายประสาทเทียม						
ชุดข้อมูล	ผลลัพธ์ (Result)				เกณฑ์การวัด (Measurement)	
	FN	TN	FP	TP	Accuracy	Recall
ชุดการเรียนรู้	13	95803	4511	432	95.92%	97.08%
ชุดการทดสอบ	12	32010	1428	136	96.10%	91.89%
ชุดการทวนสอบ	14	31999	1441	135	96.05%	90.60%
รวม	39	159812	7380	703	95.98%	94.74%

4.1.3 การเปรียบเทียบผลการพยากรณ์

ทำการเปรียบเทียบผลการพยากรณ์ระหว่างแบบจำลองจากวิธีการวิเคราะห์การถดถอยโลจิสติกและแบบจำลองจากโครงข่ายประสาทเทียม โดยจะพิจารณาจากค่าการวัดทั้งหมด จากผลการวิเคราะห์พบว่า ค่าความถูกต้องของแบบจำลอง (Accuracy) และค่าพยากรณ์การเกิดเหตุการณ์ที่ถูกต้อง (Recall) จะเห็นได้ว่าแบบจำลองที่ได้จากโครงข่ายประสาทเทียมให้ผลการพยากรณ์ทั้งสองค่าได้ดีกว่าแบบจำลองจากการวิเคราะห์การถดถอยโลจิสติกของจำนวนข้อมูลทั้งหมดคือค่าความถูกต้อง $95.98\% > 72.37\%$ และค่าความไว $94.97\% > 74.26\%$ เมื่อพิจารณาในแต่ละชุดการทดสอบทั้งสามชุดข้อมูล ค่าความถูกต้องของการวิเคราะห์โครงข่ายประสาทเทียมให้ค่าสูงกว่าการวิเคราะห์การถดถอยโลจิสติกทั้งสามชุดข้อมูล จึงสรุปได้ว่าวิธีวิเคราะห์โครงข่ายประสาทเทียมให้ผลการพยากรณ์ที่ดีกว่าการวิเคราะห์การถดถอยโลจิสติก ดังแสดงในตารางที่ 4.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.4 ประสิทธิภาพการพยากรณ์ของแบบจำลองการวิเคราะห์การถดถอยโลจิสติกและ
โครงข่ายประสาทเทียม

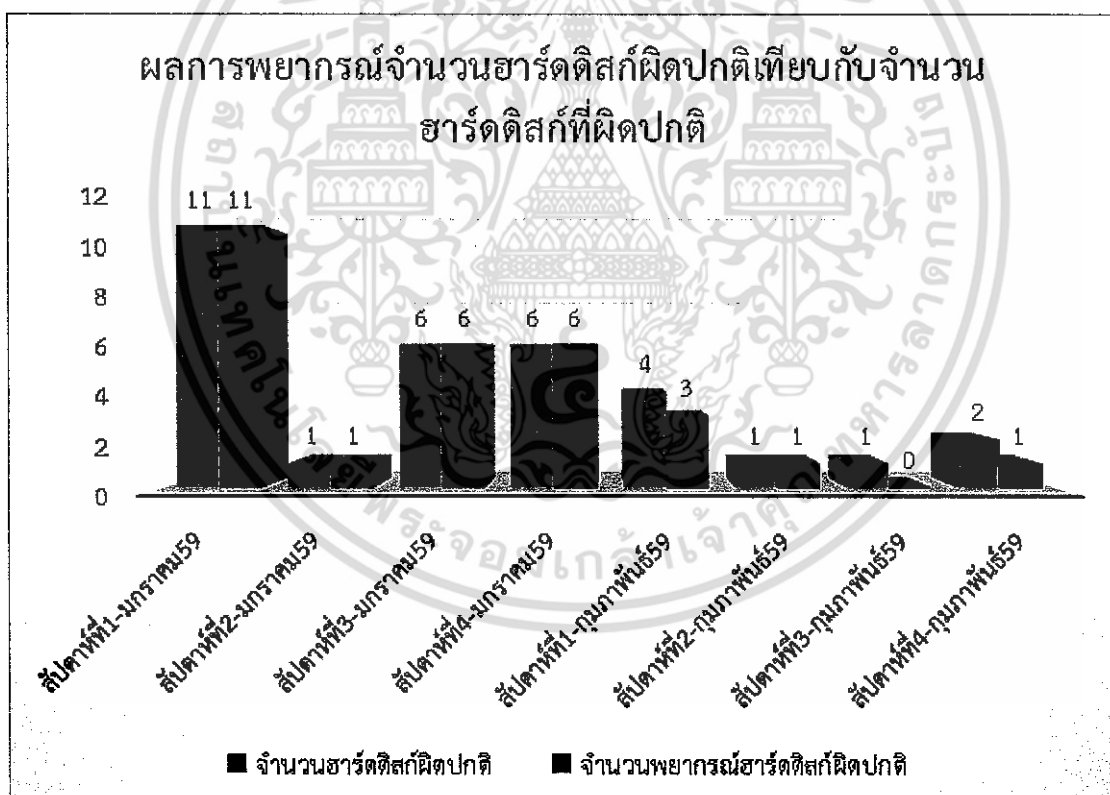
ชุดข้อมูล	เกณฑ์การวัด (Measurement)			
	การวิเคราะห์การถดถอยโลจิสติก		การวิเคราะห์โครงข่ายประสาทเทียม	
	Accuracy	Recall	Accuracy	Recall
ชุดการเรียนรู้	72.37%	75.28%	95.92%	97.08%
ชุดการทดสอบ	72.47%	72.97%	96.10%	91.89%
ชุดการทวนสอบ	72.27%	72.48%	96.05%	90.60%
รวม	72.37%	74.26%	95.98%	94.74%

4.2 สร้างโปรแกรมการทดสอบและประยุกต์ใช้ตัวแบบ

สร้างตัวแบบเพื่อพยากรณ์คุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ จากแบบจำลองโครงข่ายประสาทเทียม โดยการเขียนโปรแกรมแล้วนำไปประยุกต์ใช้ในกระบวนการทดสอบคุณภาพฮาร์ดดิสก์ก่อนการทดสอบความเชื่อมั่นตามรูปที่ 1.4 โดยผู้วิจัยจะทำการเก็บข้อมูลเพิ่มเติมเพื่อประเมินผลความถูกต้องของแบบจำลองต่อไป อ้างอิงข้อมูลใหม่จำนวน 2 เดือนเพื่อให้แน่ใจว่าแบบจำลองดังกล่าวสามารถนำมาใช้งานได้จริง จากผลการประเมินแบบจำลองที่สร้างด้วยวิธีการโครงข่ายประสาทเทียมพบว่า สามารถป้องกันการเกิดฮาร์ดดิสก์ผิดปกติที่หัวอ่านเขียนได้ 90% จากข้อมูลพบว่าสามารถป้องกันฮาร์ดดิสก์ที่มีอาการผิดปกติที่หัวอ่านเขียนจำนวน 29 ตัวจากจำนวนทั้งหมด 32 ตัว จากตารางที่ 4.5 แสดงผลการพยากรณ์คุณภาพผลิตภัณฑ์ฮาร์ดดิสก์ก่อนการทดสอบความเชื่อมั่นของข้อมูลใหม่ 2 เดือน และรูปที่ 4.1 แสดงผลการพยากรณ์ฮาร์ดดิสก์ผิดปกติเทียบกับฮาร์ดดิสก์ที่ผิดปกติจริง ดังนั้นจึงสามารถสรุปได้ว่าแบบจำลองที่ได้จากการวิเคราะห์โครงข่ายประสาทเทียมมีประสิทธิภาพในการพยากรณ์คุณภาพของฮาร์ดดิสก์และสามารถนำมาใช้งานได้จริงกับกระบวนการทดสอบฮาร์ดดิสก์ก่อนการส่งมอบแก่ผู้บริโภค

ตารางที่ 4.5 ผลการพยากรณ์คุณภาพผลิตภัณฑ์ฮาร์ดดิสก์ก่อนกระบวนการทดสอบความเชื่อมั่นของข้อมูลใหม่ 2 เดือน

ข้อมูล	ผลการพยากรณ์(Prediction Result)				เกณฑ์การวัด (Measurements)	
	FN	TN	FP	TP	Accuracy	Recall
สัปดาห์ที่1-มกราคม59	0	4200	191	11	96%	100%
สัปดาห์ที่2-มกราคม59	0	4216	209	1	95%	100%
สัปดาห์ที่3-มกราคม59	0	7018	367	6	95%	100%
สัปดาห์ที่4-มกราคม59	0	3160	157	6	95%	100%
สัปดาห์ที่1-กุมภาพันธ์59	1	4228	194	3	96%	75%
สัปดาห์ที่2-กุมภาพันธ์59	0	6314	298	1	95%	100%
สัปดาห์ที่3-กุมภาพันธ์59	1	2637	118	0	96%	0%
สัปดาห์ที่4-กุมภาพันธ์59	1	2546	159	1	94%	50%
รวม	3	34319	1693	29	95%	91%



รูปที่ 4.1 ผลการพยากรณ์จำนวนฮาร์ดดิสก์ที่ผิดปกติเทียบกับจำนวนฮาร์ดดิสก์ที่ผิดปกติจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

จากการทดสอบตัวแปรอิสระที่มีความสัมพันธ์กับคุณภาพของฮาร์ดดิสก์จำนวน 197 ตัวแปร ผลการทดสอบพบว่าตัวแปรอิสระจำนวน 25 ตัวแปรที่มีความสัมพันธ์กับฮาร์ดดิสก์ที่มีอาการผิดปกติที่หัวอ่าน/เขียนอย่างมีนัยสำคัญและได้ทำการยืนยันตัวแปรเหล่านั้นด้วยผู้มีความรู้ทางด้านเทคนิคของผลิตภัณฑ์ฮาร์ดดิสก์แล้วว่าตัวแปรเหล่านี้ส่งผลกับคุณภาพของฮาร์ดดิสก์จริง ดังนั้นจึงสามารถนำตัวแปรอิสระทั้ง 25 ตัวแปรนี้ไปสร้างแบบจำลองโครงข่ายประสาทเทียม ส่วนแบบจำลองการวิเคราะห์การถดถอยโลจิสติกนั้นจะต้องกำจัดตัวแปรอิสระที่มีความสัมพันธ์กันเองออกก่อนจึงสามารถนำมาสร้างแบบจำลองได้ จากการวิเคราะห์ความสัมพันธ์พบว่าตัวแปรอิสระสี่กลุ่มที่มีความสัมพันธ์กันเอง จึงได้ทำการกำจัดตัวแปรอิสระเหล่านั้นออกไป โดยใช้วิธีการการสร้างตัวแปรใหม่ด้วยวิธีการวิเคราะห์องค์ประกอบหลัก เพื่อป้องกันไม่ให้เกิดความซ้ำซ้อนและมีขนาดใหญ่จนเกินไป ดังนั้นคงเหลือตัวแปรอิสระที่นำไปสร้างแบบจำลองการวิเคราะห์การถดถอยโลจิสติกเพียง 17 ตัวแปร

ผลการพยากรณ์คุณภาพผลิตภัณฑ์ฮาร์ดดิสก์จากแบบจำลองการถดถอยโลจิสติก ซึ่งเลือกตัวแปรอิสระทั้งสิ้นจำนวน 6 ตัวแปรจากตัวแปรอิสระ 17 ตัวแปรและโครงข่ายประสาทเทียมสร้างแบบจำลองจากตัวแปรอิสระทั้งสิ้น 25 ตัวแปร จากการเปรียบเทียบความสามารถในการพยากรณ์คุณภาพผลิตภัณฑ์ฮาร์ดดิสก์ที่มีอาการผิดปกติที่หัวอ่าน/เขียนระหว่างแบบจำลองโลจิสติกและโครงข่ายประสาทเทียมพบว่าแบบจำลองที่ได้จากการวิเคราะห์โครงข่ายประสาทเทียมจะให้ประสิทธิภาพในการพยากรณ์คุณภาพของฮาร์ดดิสก์ได้ดีกว่าแบบจำลองที่ได้จากการวิเคราะห์การถดถอยโลจิสติก โดยพิจารณาจากค่าความถูกต้อง (Accuracy) และค่าความถูกต้องของการพยากรณ์การเกิดเหตุการณ์ (Recall) ของการพยากรณ์คุณภาพฮาร์ดดิสก์ผิดปกติ ดังนั้นจึงสรุปได้ว่าแบบจำลองโครงข่ายประสาทเทียมให้ผลการพยากรณ์ที่มีประสิทธิภาพในการพยากรณ์ได้ดีกว่าแบบจำลองการถดถอยโลจิสติก

5.2 ข้อเสนอแนะ

แบบจำลองที่ใช้พยากรณ์คุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ทั้งสองแบบจำลองนั้น สามารถปรับเปลี่ยนและแก้ไขได้ เนื่องจากในแต่ละช่วงระยะเวลาอาจมีการปรับปรุงกระบวนการผลิตหรือเปลี่ยนแปลงเทคโนโลยีในการผลิตให้ทันต่อเทคโนโลยีใหม่ๆ ทำให้ข้อมูลต่างๆหรือตัวแปรอิสระมีการเปลี่ยนแปลงไปด้วย ดังนั้นจึงต้องมีการเรียนรู้แบบจำลองใหม่หรือที่เรียกว่า Retraining Model คือการใช้ข้อมูลชุดใหม่ไปสร้างแบบจำลองเพื่อหาสมการใหม่โดยยังคงใช้ตัวแปรเดิม แต่ถ้าหากมีการเปลี่ยนแปลงของข้อมูลมากเกินไปหรือมีการเปลี่ยนแปลงตัวแปรอิสระ จะต้องทำการสร้างแบบจำลองใหม่หรือที่เรียกว่า Rebuilding Model คือการสร้างแบบจำลองใหม่โดยใช้ข้อมูลใหม่และตัวแปรใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ด้วย โดยที่ทั้งสองวิธีการจะยังคงสามารถใช้กระบวนการหรือวิธีการสร้างแบบจำลองเช่นเดิมได้ (Tan. et al, 2005)



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

กัลยา วานิชย์บัญชา. 2546. การวิเคราะห์สถิติขั้นสูงด้วย SPSS for Windows.

พิมพ์ครั้งที่ 3 : บริษัทธรรมสาร

นิตาชล แซ่แห้ว. 2555. การประยุกต์ใช้โครงข่ายประสาทเทียมและการวิเคราะห์การถดถอยเชิง

พหุคูณในการพยากรณ์. วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมอุตสาหการ บัณฑิตวิทยาลัย มหาวิทยาลัยเชียงใหม่

นำชัย ศุภฤกษ์ชัยสกุล. 2553. การวิเคราะห์ Logistic Regression. สถาบันวิจัยพฤติกรรมศาสตร์

มหาวิทยาลัยศรีนครินทรวิโรฒ. [ออนไลน์]. เข้าถึงได้จาก :

http://rlc.nrct.go.th/ewt_dl.php?nid=701

ประสิทธิ์ชัย บุญเสริม. 2552. การสร้างโมเดลเพื่อทำนายจากเครือข่ายประสาทแบบสุ่มสำหรับการ

การผลิตเซมิคอนดักเตอร์. วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์

ภาควิชาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

วรชัย สิงห์จั้งหรีด. 2550. การวิเคราะห์อาการเสียของฮาร์ดดิสก์โดยวิธีการทำเหมืองข้อมูล.

วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า

บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

สิริยาภรณ์ ไกรมาก. 2552. แบบจำลองการทำนายเพื่อการประเมินคุณภาพของผลิตภัณฑ์

กาแพคค์และบัด. วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต สาขาการจัดการเทคโนโลยี

อุตสาหกรรมเกษตร ภาควิชาเทคโนโลยีอุตสาหกรรมเกษตร บัณฑิตวิทยาลัย

มหาวิทยาลัยเกษตรศาสตร์

Ben, K. and Patrick, S. 1996. An Introduction to Neural Networks. Eighth edition.

Amsterdam: Oberpfaffenhofen.

Campbell, I, 2007. Chi-squared and Fisher-Irwin tests of two-by-two tables

Statistics in Medicine: Aug 30. 26(19) p. 3661-3675.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Chen, C. Liaw, A. and Breiman, L. 2004, **Using Random Forest to Learn Imbalanced Data** : University of California. Report ID: 666

Fausett, L. 1994. **Fundamentals of Neural Networks: Architectures, Algorithms, and Application**. Prentice Hall. Inc., New Jersey.

Jolliffe I.T. 2002, **Principal Component Analysis** : Second Edition. Springer-Verlag. New York. Inc.

Satish, K. 2012. **Neural Networks: A Classroom Approach (Second Edition)**: McGraw Hill Education. Columbus.

Tan, Steinbach and Kumar. 2005. **Introduction to Data Mining**. Pearson International Edition, Inc., Addison Wesley. Boston.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

ภาคผนวก ก. ตัวแปรอิสระ

ตารางที่ ก.1 ตัวแปรอิสระจากกระบวนการทดสอบชุดหัวอ่าน/เขียน

Independent Variable (ตัวแปรอิสระ)	Variable Name (ชื่อตัวแปร)	Unit (หน่วย)
V1	ความแตกต่างความต้านทาน	ohm
V2	ความกว้างของคลื่นความถี่ต่ำ	MHz
V3	ความกว้างของคลื่นความถี่กลาง	MHz
V4	ความกว้างของสนามแม่เหล็กของหัวอ่าน	nm
V5	ความกว้างของสนามแม่เหล็กของหัวเขียน	nm
V6	การเขียนทับไปมา	dB
V7	ค่าความกว้างเฉลี่ยสมดุของหัวอ่าน	percent
V8	กำลังแรงกตการบินของหัวอ่าน	mW
V9	เซ็นเซอร์การบินแต่ละพื้น	ohm
V10	แรงดันไฟฟ้าของแรงกตการบินของหัวอ่าน	Volt
V11	ความกว้างของสนามแม่เหล็กของหัวอ่าน	nm
V12	การแปลงข้อมูลอัตราส่วนสัญญาณต่อเสียงรบกวน	ratio
V13	ความกว้างการลบสนามแม่เหล็กของหัวเขียน	nm
V14	อัตราส่วนสเปกตรัมสัญญาณต่อเสียงรบกวน	ratio

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.2 ตัวแปรอิสระจากกระบวนการทดสอบฮาร์ดดิสก์

Independent Variable (ตัวแปรอิสระ)	Variable Name (ชื่อตัวแปร)	Unit (หน่วย)
X1	ผลรวมข้อผิดพลาดจากการอ่าน ทดสอบครั้งที่ 1 พื้นที่ A	COUNT
X2	ผลรวมข้อผิดพลาดจากการอ่าน ทดสอบครั้งที่ 1 พื้นที่ B	COUNT
X3	ผลรวมข้อผิดพลาดจากการอ่าน ทดสอบครั้งที่ 1 พื้นที่ C	COUNT
X4	อัตราความถูกต้องเทียบกับข้อผิดพลาดที่แย่ที่สุดของแทร็คหลังการเขียน ทดสอบครั้งที่ 1 พื้นที่ A	(EM*10)DB
X5	อัตราความถูกต้องเทียบกับข้อผิดพลาดที่แย่ที่สุดของแทร็คหลังการเขียน ทดสอบครั้งที่ 1 พื้นที่ B	(EM*10)DB
X6	อัตราความถูกต้องเทียบกับข้อผิดพลาดที่แย่ที่สุดของแทร็คหลังการเขียน ทดสอบครั้งที่ 1 พื้นที่ C	(EM*10)DB
X7	อัตราความถูกต้องเทียบกับข้อผิดพลาดที่แย่ที่สุดของแทร็คหลังการเขียน ทดสอบครั้งที่ 1 พื้นที่ D	(EM*10)DB
X8	อัตราความถูกต้องเทียบกับข้อผิดพลาดที่แย่ที่สุดของแทร็คหลังการเขียน ทดสอบครั้งที่ 2 พื้นที่ A	(EM*10)DB
X9	อัตราความถูกต้องเทียบกับข้อผิดพลาดที่แย่ที่สุดของแทร็คหลังการเขียน ทดสอบครั้งที่ 2 พื้นที่ B	(EM*10)DB
X10	อัตราความถูกต้องเทียบกับข้อผิดพลาดที่แย่ที่สุดของแทร็คหลังการเขียน ทดสอบครั้งที่ 2 พื้นที่ C	(EM*10)DB
X11	อัตราความถูกต้องเทียบกับข้อผิดพลาดที่แย่ที่สุดของแทร็คหลังการเขียน ทดสอบครั้งที่ 2 พื้นที่ D	(EM*10)DB
X12	การเขียนซ้ำ ทดสอบครั้งที่1 พื้นที่ A	DB
X13	การเขียนซ้ำ ทดสอบครั้งที่1 พื้นที่ B	DB
X14	การเขียนซ้ำ ทดสอบครั้งที่1 พื้นที่ C	DB
X15	การเขียนซ้ำ ทดสอบครั้งที่2 พื้นที่ A	DB
X16	การเขียนซ้ำ ทดสอบครั้งที่2 พื้นที่ B	DB
X17	การเขียนซ้ำ ทดสอบครั้งที่2 พื้นที่ C	DB
X18	ผลรวมสะสมความผิดพลาดของสัญญาณรบกวนพื้นฐานแต่ละรัศมีของ แผ่นบันทึกข้อมูล ทดสอบครั้งที่ 1 พื้นที่ A	COUNT
X19	ผลรวมสะสมความผิดพลาดของสัญญาณรบกวนพื้นฐานแต่ละรัศมีของ แผ่นบันทึกข้อมูล ทดสอบครั้งที่ 1 พื้นที่ B	COUNT
X20	ผลรวมสะสมความผิดพลาดของสัญญาณรบกวนพื้นฐานแต่ละรัศมีของ แผ่นบันทึกข้อมูล ทดสอบครั้งที่ 1 พื้นที่ C	COUNT
X21	ผลรวมสะสมความผิดพลาดของสัญญาณรบกวนพื้นฐานแต่ละรัศมีของ แผ่นบันทึกข้อมูล ทดสอบครั้งที่ 2 พื้นที่ A	COUNT

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.2 ตัวแปรอิสระจากกระบวนการทดสอบฮาร์ดดิสก์ (ต่อ)

Independent Variable (ตัวแปรอิสระ)	Variable Name (ชื่อตัวแปร)	Unit (หน่วย)
X22	ผลรวมสะสมความผิดพลาดของสัญญาณรบกวนพื้นฐานแต่ละรัศมีของแผ่นบันทึกข้อมูล ทดสอบครั้งที่ 2 พื้นที่ B	COUNT
X23	ผลรวมสะสมความผิดพลาดของสัญญาณรบกวนพื้นฐานแต่ละรัศมีของแผ่นบันทึกข้อมูล ทดสอบครั้งที่ 2 พื้นที่ C	COUNT
X24	ผลรวมความผิดพลาดของสัญญาณรบกวนพื้นฐานแต่ละรัศมีของแผ่นบันทึกข้อมูล ทดสอบครั้งที่ 1 พื้นที่ A	COUNT
X25	ผลรวมความผิดพลาดของสัญญาณรบกวนพื้นฐานแต่ละรัศมีของแผ่นบันทึกข้อมูล ทดสอบครั้งที่ 1 พื้นที่ B	COUNT
X26	ผลรวมความผิดพลาดของสัญญาณรบกวนพื้นฐานแต่ละรัศมีของแผ่นบันทึกข้อมูล ทดสอบครั้งที่ 1 พื้นที่ C	COUNT
X27	ผลรวมความผิดพลาดของสัญญาณรบกวนพื้นฐานแต่ละรัศมีของแผ่นบันทึกข้อมูล ทดสอบครั้งที่ 2 พื้นที่ A	COUNT
X28	ผลรวมความผิดพลาดของสัญญาณรบกวนพื้นฐานแต่ละรัศมีของแผ่นบันทึกข้อมูล ทดสอบครั้งที่ 2 พื้นที่ B	COUNT
X29	ผลรวมความผิดพลาดของสัญญาณรบกวนพื้นฐานแต่ละรัศมีของแผ่นบันทึกข้อมูล ทดสอบครั้งที่ 2 พื้นที่ C	COUNT
X30	ความผิดพลาดของสัญญาณรบกวนพื้นฐานของหัวอ่านเขียน	COUNT
X31	ค่าต่ำสุดของการเขียนซ้ำของแตรคครั้งที่ 1 พื้นที่ A	DB*100
X32	ค่าต่ำสุดของการเขียนซ้ำของแตรคครั้งที่ 1 พื้นที่ B	DB*101
X33	ค่าต่ำสุดของการเขียนซ้ำของแตรคครั้งที่ 1 พื้นที่ C	DB*102
X34	ค่าต่ำสุดของการเขียนซ้ำของแตรคครั้งที่ 2 พื้นที่ A	DB*103
X35	ค่าต่ำสุดของการเขียนซ้ำของแตรคครั้งที่ 2 พื้นที่ B	DB*104
X36	ค่าต่ำสุดของการเขียนซ้ำของแตรคครั้งที่ 2 พื้นที่ C	DB*105
X37	ความหนาแน่นของข้อบกพร่อง	COUNT
X38	ค่าความต่างระหว่างอัตราบิตข้อผิดพลาดของการทดสอบครั้งที่1และ2 พื้นที่ A	LOG PER RATIO * 1000
X39	ค่าความต่างระหว่างอัตราบิตข้อผิดพลาดของการทดสอบครั้งที่1และ2 พื้นที่ B	LOG PER RATIO * 1000
X40	ค่าความต่างระหว่างอัตราบิตข้อผิดพลาดของการทดสอบครั้งที่1และ2 พื้นที่ C	LOG PER RATIO * 1000
X41	ค่าความต่างระหว่างอัตราบิตข้อผิดพลาดของการทดสอบครั้งที่1และ2 พื้นที่ D	LOG PER RATIO * 1000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.2 ตัวแปรอิสระจากกระบวนการทดสอบฮาร์ดดิสก์ (ต่อ)

Independent Variable (ตัวแปรอิสระ)	Variable Name (ชื่อตัวแปร)	Unit (หน่วย)
X42	ค่าความต่างระหว่างอัตราบิตข้อผิดพลาดของการทดสอบครั้งที่1และ2 พื้นที่ E	LOG PER RATIO * 1000
X43	ความผิดพลาดของข้อมูลหน่วยความจำแอดเดรส พื้นที่ A	COUNT
X44	ความผิดพลาดของข้อมูลหน่วยความจำแอดเดรส พื้นที่ B	COUNT
X45	ความผิดพลาดของข้อมูลหน่วยความจำแอดเดรส พื้นที่ C	COUNT
X46	ความผิดพลาดของข้อมูลหน่วยความจำแอดเดรส พื้นที่ D	COUNT
X47	ความผิดพลาดของข้อมูลหน่วยความจำแอดเดรส พื้นที่ E	COUNT
X48	ค่าการย้ายที่ของหัวอ่าน ทดสอบครั้งที่ 1	Pico meters per volt
X49	ค่าการย้ายที่ของหัวอ่าน ทดสอบครั้งที่ 2	Pico meters per volt
X50	ค่าการย้ายที่ของหัวอ่าน ทดสอบครั้งที่ 3	Pico meters per volt
X51	ค่าต่ำสุดของค่าเฉลี่ยจำนวนที่เกิดความผิดพลาด ทดสอบครั้งที่ 1	dB x 100
X52	ค่าต่ำสุดของค่าเฉลี่ยจำนวนที่เกิดความผิดพลาด ทดสอบครั้งที่ 2	dB x 100
X53	ค่าต่ำสุดของค่าเฉลี่ยจำนวนที่เกิดความผิดพลาด ทดสอบครั้งที่ 3	dB x 100
X54	จำนวนที่เกิดความผิดพลาด ทดสอบครั้งที่ 1	dB x 100
X55	จำนวนที่เกิดความผิดพลาด ทดสอบครั้งที่ 2	dB x 100
X56	จำนวนที่เกิดความผิดพลาด ทดสอบครั้งที่ 3	dB x 100
X57	ค่าต่ำสุดของจำนวนที่เกิดความผิดพลาด ทดสอบครั้งที่ 1	dB x 100
X58	ค่าต่ำสุดของจำนวนที่เกิดความผิดพลาด ทดสอบครั้งที่ 2	dB x 100
X59	ค่าต่ำสุดของจำนวนที่เกิดความผิดพลาด ทดสอบครั้งที่ 3	dB x 100
X60	ค่าสูงสุดของค่าความร้อน	TRACK
X61	ความอึดตัวของหัวอ่าน	COUNT
X62	ค่าสูงสุดของข้อมูลบกพร่อง	COUNT/ TRACK
X63	ข้อผิดพลาดของข้อมูลลอการิทึม	COUNT
X64	ข้อผิดพลาดของเซอร์โวลอการิทึม	COUNT
X65	การเขียนคำสั่งยกเลิก	COUNT
X66	ข้อบกพร่องของแผ่นดิสก์ตอนหัวอ่านเข้าระยะยาว	COUNT
X67	ข้อบกพร่องของแผ่นดิสก์ตอนหัวอ่านออกระยะยาว	COUNT
X68	ข้อบกพร่องของแผ่นดิสก์ตอนหัวอ่านเข้าระยะสั้น	COUNT
X69	ข้อบกพร่องของแผ่นดิสก์ตอนหัวอ่านออกระยะสั้น	COUNT

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.2 ตัวแปรอิสระจากกระบวนการทดสอบฮาร์ดดิสก์ (ต่อ)

Independent Variable (ตัวแปรอิสระ)	Variable Name (ชื่อตัวแปร)	Unit (หน่วย)
X70	ความสมมาตรของกระแสไฟฟ้ากับความต้านทาน ทดสอบครั้งที่ 1 พื้นที่ A	DAC
X71	ความสมมาตรของกระแสไฟฟ้ากับความต้านทาน ทดสอบครั้งที่ 1 พื้นที่ B	DAC
X72	ความสมมาตรของกระแสไฟฟ้ากับความต้านทาน ทดสอบครั้งที่ 1 พื้นที่ C	DAC
X73	ความสมมาตรของกระแสไฟฟ้ากับความต้านทาน ทดสอบครั้งที่ 2 พื้นที่ A	DAC
X74	ความสมมาตรของกระแสไฟฟ้ากับความต้านทาน ทดสอบครั้งที่ 2 พื้นที่ B	DAC
X75	ความสมมาตรของกระแสไฟฟ้ากับความต้านทาน ทดสอบครั้งที่ 2 พื้นที่ C	DAC
X76	การเปลี่ยนความต้านทาน	percent
X77	ค่าเฉลี่ยการอ่านข้อมูลแทร็ค ทดสอบครั้งที่ 1 พื้นที่ A	% of data TRACK
X78	ค่าเฉลี่ยการอ่านข้อมูลแทร็ค ทดสอบครั้งที่ 1 พื้นที่ B	% of data TRACK
X79	ค่าเฉลี่ยการอ่านข้อมูลแทร็ค ทดสอบครั้งที่ 1 พื้นที่ C	% of data TRACK
X80	ค่าเฉลี่ยการอ่านข้อมูลแทร็ค ทดสอบครั้งที่ 2 พื้นที่ A	% of data TRACK
X81	ค่าเฉลี่ยการอ่านข้อมูลแทร็ค ทดสอบครั้งที่ 2 พื้นที่ B	% of data TRACK
X82	ค่าเฉลี่ยการอ่านข้อมูลแทร็ค ทดสอบครั้งที่ 2 พื้นที่ C	% of data TRACK
X83	ค่าเฉลี่ยการอ่านข้อมูลแทร็ค ทดสอบครั้งที่ 3 พื้นที่ A	% of data TRACK
X84	ค่าเฉลี่ยการอ่านข้อมูลแทร็ค ทดสอบครั้งที่ 3 พื้นที่ B	% of data TRACK
X85	ค่าเฉลี่ยการอ่านข้อมูลแทร็ค ทดสอบครั้งที่ 3 พื้นที่ C	% of data TRACK
X86	ค่าสูงสุดของข้อผิดพลาดของแทร็ค	COUNT
X87	ผลรวมค่าผิดพลาดจากการอ่านเวดจ์ พื้นที่ A	COUNT
X88	ผลรวมค่าผิดพลาดจากการอ่านเวดจ์ พื้นที่ B	COUNT
X89	ผลรวมค่าผิดพลาดจากการอ่านเวดจ์ พื้นที่ C	COUNT
X90	ข้อผิดพลาดของเซอร์โว	COUNT
X91	การสั่นชนิดที่ 2 การทดสอบครั้งที่ 1 พื้นที่ A	dB
X92	การสั่นชนิดที่ 2 การทดสอบครั้งที่ 1 พื้นที่ B	dB
X93	การสั่นชนิดที่ 2 การทดสอบครั้งที่ 1 พื้นที่ C	dB
X94	การสั่นชนิดที่ 2 การทดสอบครั้งที่ 2 พื้นที่ A	dB
X95	การสั่นชนิดที่ 2 การทดสอบครั้งที่ 2 พื้นที่ B	dB

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.2 ตัวแปรอิสระจากกระบวนการทดสอบฮาร์ดดิสก์ (ต่อ)

Independent Variable (ตัวแปรอิสระ)	Variable Name (ชื่อตัวแปร)	Unit (หน่วย)
X96	การสั่นชนิดที่2 การทดสอบครั้งที่ 2 พื้นที่ C	dB
X97	การสั่นชนิดที่1 การทดสอบครั้งที่ 1 พื้นที่ A	dB
X98	การสั่นชนิดที่1 การทดสอบครั้งที่ 1 พื้นที่ B	dB
X99	การสั่นชนิดที่1 การทดสอบครั้งที่ 1 พื้นที่ C	dB
X100	การสั่นชนิดที่1 การทดสอบครั้งที่ 2 พื้นที่ A	dB
X101	การสั่นชนิดที่1 การทดสอบครั้งที่ 2 พื้นที่ B	dB
X102	การสั่นชนิดที่1 การทดสอบครั้งที่ 2 พื้นที่ C	dB
X103	ความแปรปรวนแรงกดการบินในการอ่าน	-
X104	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านใน ทดสอบครั้งที่ 1 พื้นที่ A	MW
X105	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านใน ทดสอบครั้งที่ 1 พื้นที่ B	MW
X106	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านใน ทดสอบครั้งที่ 1 พื้นที่ C	MW
X107	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านใน ทดสอบครั้งที่ 1 พื้นที่ D	MW
X108	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านใน ทดสอบครั้งที่ 2 พื้นที่ E	MW
X109	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านใน ทดสอบครั้งที่ 2 พื้นที่ A	MW
X110	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านใน ทดสอบครั้งที่ 2 พื้นที่ B	MW
X111	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านใน ทดสอบครั้งที่ 2 พื้นที่ C	MW
X112	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านใน ทดสอบครั้งที่ 2 พื้นที่ D	MW
X113	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านใน ทดสอบครั้งที่ 2 พื้นที่ E	MW
X114	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านใน ทดสอบครั้งที่ 3 พื้นที่ A	MW
X115	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านใน ทดสอบครั้งที่ 3 พื้นที่ B	MW

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.2 ตัวแปรอิสระจากกระบวนการทดสอบฮาร์ดดิสก์ (ต่อ)

Independent Variable (ตัวแปรอิสระ)	Variable Name (ชื่อตัวแปร)	Unit (หน่วย)
X116	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านใน ทดสอบครั้งที่ 3 พื้นที่ C	MW
X117	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านใน ทดสอบครั้งที่ 3 พื้นที่ D	MW
X118	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านใน ทดสอบครั้งที่ 3 พื้นที่ E	MW
X119	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 1 พื้นที่ A	MW
X120	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 1 พื้นที่ B	MW
X121	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 1 พื้นที่ C	MW
X122	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 1 พื้นที่ D	MW
X123	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 1 พื้นที่ E	MW
X124	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 2 พื้นที่ A	MW
X125	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 2 พื้นที่ B	MW
X126	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 2 พื้นที่ C	MW
X127	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 2 พื้นที่ D	MW
X128	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 2 พื้นที่ E	MW
X129	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 3 พื้นที่ A	MW
X130	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 3 พื้นที่ B	MW
X131	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลซ์ พื้นที่ด้านนอก ทดสอบครั้งที่ 3 พื้นที่ C	MW

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.2 ตัวแปรอิสระจากกระบวนการทดสอบฮาร์ดดิสก์ (ต่อ)

Independent Variable (ตัวแปรอิสระ)	Variable Name (ชื่อตัวแปร)	Unit (หน่วย)
X132	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านนอก ทดสอบครั้งที่ 3 พื้นที่ D	MW
X133	ความแตกต่างของการบินของหัวอ่านในการอ่านระหว่างแรกเริ่มกับค่านอร์มอลไชน์ พื้นที่ด้านนอก ทดสอบครั้งที่ 3 พื้นที่ E	MW
X134	การบินของหัวอ่านในการอ่านแรกของพื้นที่ด้านใน พื้นที่ A	MW
X135	การบินของหัวอ่านในการอ่านแรกของพื้นที่ด้านใน พื้นที่ B	MW
X136	การบินของหัวอ่านในการอ่านแรกของพื้นที่ด้านใน พื้นที่ C	MW
X137	การบินของหัวอ่านในการอ่านแรกของพื้นที่ด้านใน พื้นที่ D	MW
X138	การบินของหัวอ่านในการอ่านแรกของพื้นที่ด้านใน พื้นที่ E	MW
X139	ความแตกต่างสูงสุดของการบินของหัวอ่านในการอ่านแรก	MW
X140	การบินของหัวอ่านในการอ่านแรกของพื้นที่ด้านใน พื้นที่ A	MW
X141	การบินของหัวอ่านในการอ่านแรกของพื้นที่ด้านใน พื้นที่ B	MW
X142	การบินของหัวอ่านในการอ่านแรกของพื้นที่ด้านใน พื้นที่ C	MW
X143	การบินของหัวอ่านในการอ่านแรกของพื้นที่ด้านใน พื้นที่ D	MW
X144	การบินของหัวอ่านในการอ่านแรกของพื้นที่ด้านใน พื้นที่ E	MW
X145	ความแปรปรวนแรงกวดการบินในการเขียน	
X146	ความแตกต่างสูงสุดของการบินของหัวอ่านในการเขียนแรก	MW
X147	กำลังการอ่านของพื้นที่ด้านใน พื้นที่ A	mW
X148	กำลังการอ่านของพื้นที่ด้านใน พื้นที่ B	mW
X149	กำลังการอ่านของพื้นที่ด้านใน พื้นที่ C	mW
X150	กำลังการอ่านของพื้นที่ด้านใน พื้นที่ D	mW
X151	กำลังการอ่านของพื้นที่ด้านใน พื้นที่ E	mW
X152	กำลังการอ่านของพื้นที่ด้านนอก พื้นที่ A	mW
X153	กำลังการอ่านของพื้นที่ด้านนอก พื้นที่ B	mW
X154	กำลังการอ่านของพื้นที่ด้านนอก พื้นที่ C	mW
X155	กำลังการอ่านของพื้นที่ด้านนอก พื้นที่ D	mW
X156	กำลังการอ่านของพื้นที่ด้านนอก พื้นที่ E	mW
X157	ค่าความต้านทานเริ่มต้นของเซ็นเซอร์การบินหัวอ่าน	ohm
X158	ค่าแตกต่างความต้านทานเริ่มต้นและสุดท้ายของเซ็นเซอร์การบินหัวอ่าน	ohm
X159	จำนวนผิดปกติของเซ็นเซอร์การบินหัวอ่าน	COUNT
X160	ค่าสูงสุดเซ็นเซอร์การบินหัวอ่าน	COUNT
X161	จำนวนผิดปกติของการสแกนเสียง	COUNT

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.2 ตัวแปรอิสระจากกระบวนการทดสอบฮาร์ดดิสก์ (ต่อ)

Independent Variable (ตัวแปรอิสระ)	Variable Name (ชื่อตัวแปร)	Unit (หน่วย)
X162	ความแตกต่างของข้อมูลการเพิ่มความกว้างของคลื่นเสียง ทดสอบครั้งที่ 1 พื้นที่ A	DAC
X163	ความแตกต่างของข้อมูลการเพิ่มความกว้างของคลื่นเสียง ทดสอบครั้งที่ 1 พื้นที่ B	DAC
X164	ความแตกต่างของข้อมูลการเพิ่มความกว้างของคลื่นเสียง ทดสอบครั้งที่ 1 พื้นที่ C	DAC
X165	ความแตกต่างของข้อมูลการเพิ่มความกว้างของคลื่นเสียง ทดสอบครั้งที่ 2 พื้นที่ A	DAC
X166	ความแตกต่างของข้อมูลการเพิ่มความกว้างของคลื่นเสียง ทดสอบครั้งที่ 2 พื้นที่ B	DAC
X167	ความแตกต่างของข้อมูลการเพิ่มความกว้างของคลื่นเสียง ทดสอบครั้งที่ 2 พื้นที่ C	DAC
X168	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตรีกหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่1 ทดสอบครั้งที่ 1 พื้นที่ A	(EM*10)DB
X169	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตรีกหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่1 ทดสอบครั้งที่ 1 พื้นที่ B	(EM*10)DB
X170	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตรีกหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่1 ทดสอบครั้งที่ 1 พื้นที่ C	(EM*10)DB
X171	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตรีกหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่1 ทดสอบครั้งที่ 1 พื้นที่ D	(EM*10)DB
X172	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตรีกหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่1 ทดสอบครั้งที่ 2 พื้นที่ A	(EM*10)DB
X173	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตรีกหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่1 ทดสอบครั้งที่ 2 พื้นที่ B	(EM*10)DB
X174	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตรีกหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่1 ทดสอบครั้งที่ 2 พื้นที่ C	(EM*10)DB
X175	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตรีกหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่1 ทดสอบครั้งที่ 2 พื้นที่ D	(EM*10)DB
X176	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตรีกหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่2 ทดสอบครั้งที่ 1 พื้นที่ A	(EM*10)DB
X177	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตรีกหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่2 ทดสอบครั้งที่ 1 พื้นที่ B	(EM*10)DB

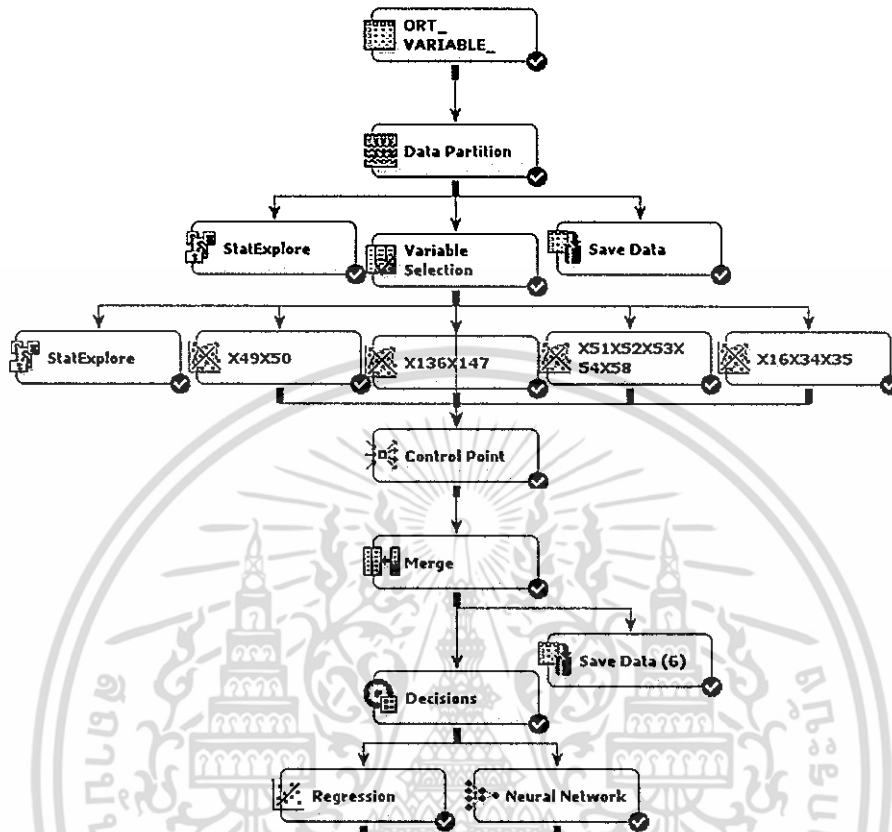
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.2 ตัวแปรอิสระจากกระบวนการทดสอบฮาร์ดดิสก์ (ต่อ)

Independent Variable (ตัวแปรอิสระ)	Variable Name (ชื่อตัวแปร)	Unit (หน่วย)
X178	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตร็คหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่2 ทดสอบครั้งที่ 1 พื้นที่ C	(EM*10)DB
X179	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตร็คหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่2 ทดสอบครั้งที่ 1 พื้นที่ D	(EM*10)DB
X180	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตร็คหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่2 ทดสอบครั้งที่ 2 พื้นที่ A	(EM*10)DB
X181	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตร็คหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่2 ทดสอบครั้งที่ 2 พื้นที่ B	(EM*10)DB
X182	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตร็คหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่2 ทดสอบครั้งที่ 2 พื้นที่ C	(EM*10)DB
X183	ขนาดความผิดพลาดที่เลวร้ายที่สุดของแตร็คหลังจากผ่านการเขียนครั้งที่2 ในบริเวณที่2 ทดสอบครั้งที่ 2 พื้นที่ D	(EM*10)DB

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข



รูปที่ ข.1 แผนผังแสดงขั้นตอนในการสร้างแบบจำลอง

ตารางที่ ข.1 จำนวนข้อมูลและตัวแปรอิสระ

Node=ORT_VARIABLE_
Data Attributes

Attribute	Value	Attribute	Value	Attribute	Value
Data Name	ORT_VARIABLE_	Date Created	15Jun2016:08:42:26	Data Size	271844352
Data Type	DATA	Date Modified	15Jun2016:09:43:15	Role	RAW
Data Label		Number Rows	167934	Segment	
Engine	BASE	Number Columns	200	Data Library	THESIS

Node=ORT_VARIABLE_
Variable Summary

Role	Level	Frequency Count	Name
TARGET	BINARY	1	HEAD_P_F
INPUT	INTERVAL	197	V1 V10 V11 V12 V13 V14 V2 V3 V4 V5 V6 V7 V8 V9 X1 X10 X100 X101 X102 X103 X104 X105 X106 X107 X108 X109 X11 X110 X111 X112 X113 X114 X115 X116 X117 X118 X119 X12 X120 X121 X122 X123 X124 X125 X126 ...
ID	NOMINAL	2	HEAD_ID SERIAL_NUMBER

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.2 การแบ่งชุดการเรียนรู้ ชุดการทดสอบ และชุดการทวนสอบ

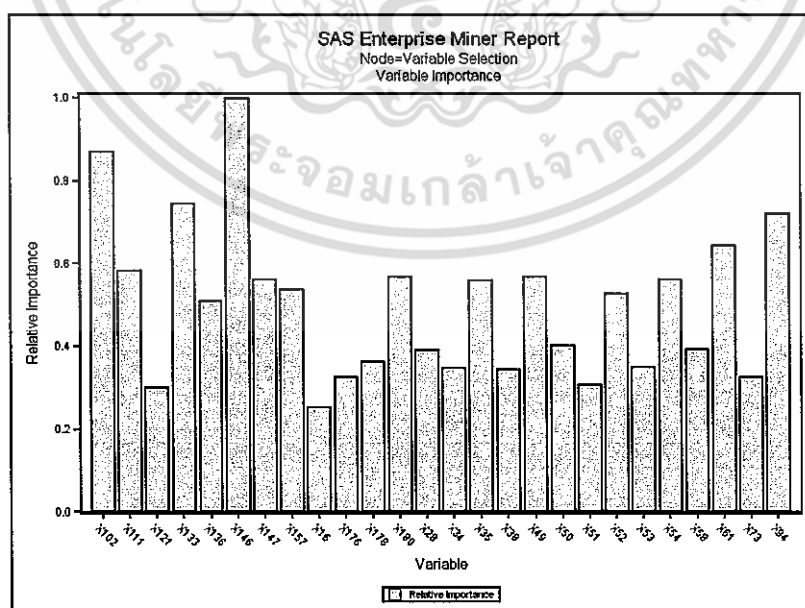
Node=Data Partition Properties

Property	Value	Default	Property	Value	Default	Property	Value	Default
Component	Partition		Method	DEFAULT		TestPct	20	30
ClassDistribution	Y		OutputType	DATA		TrainPct	60	40
IntervalDistribution	Y		RandomSeed	12345		ValidatePct	20	30

ตารางที่ ข.3 การตั้งค่าการทดสอบไคสแควร์

Node=Variable Selection (Chi-Square Test) Properties

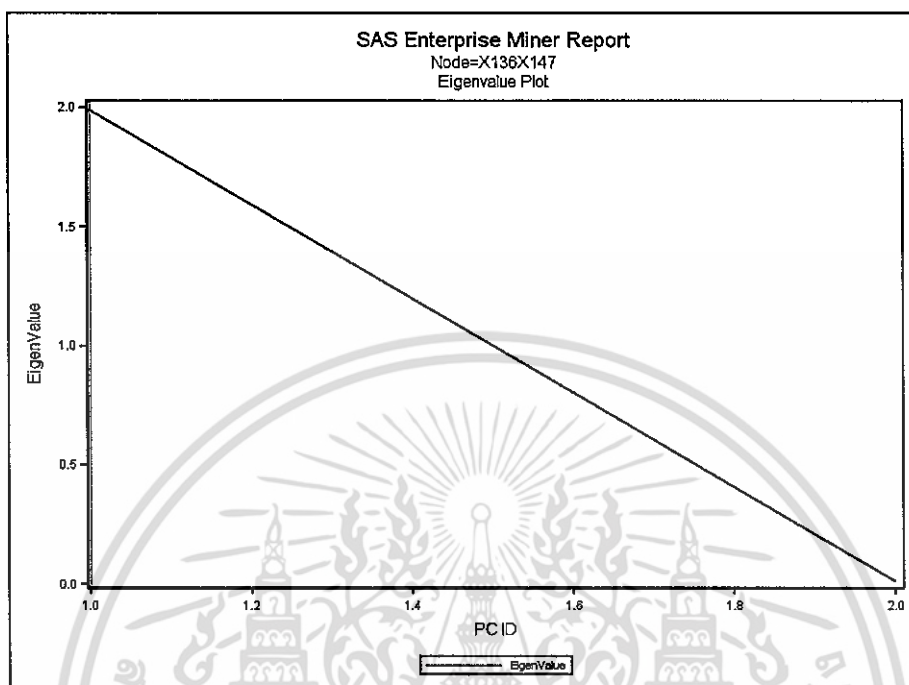
Property	Value	Default	Property	Value	Default	Property	Value	Default
Component	VariableSelection		MaxLevel	100		RoleUnusedVars	DEFAULT	
Bins	50		MaxMissingPercent	50		SASSPDS	Y	
ByPassVars	NONE		MaxRows	3000		StopR2	0.0005	
ByPassVarsRole	INPUT		MinR2	0.005		TargetModel	CHISQUARE	DEFAULT
ChiSquare	3.84		Passes	6		UseAov16	N	
HideRejectedVars	Y		PrintOption	DEFAULT		UseGroups	Y	
HideUnusedInputVars	Y		RejectUnusedInputVars	Y		UseInteractions	N	



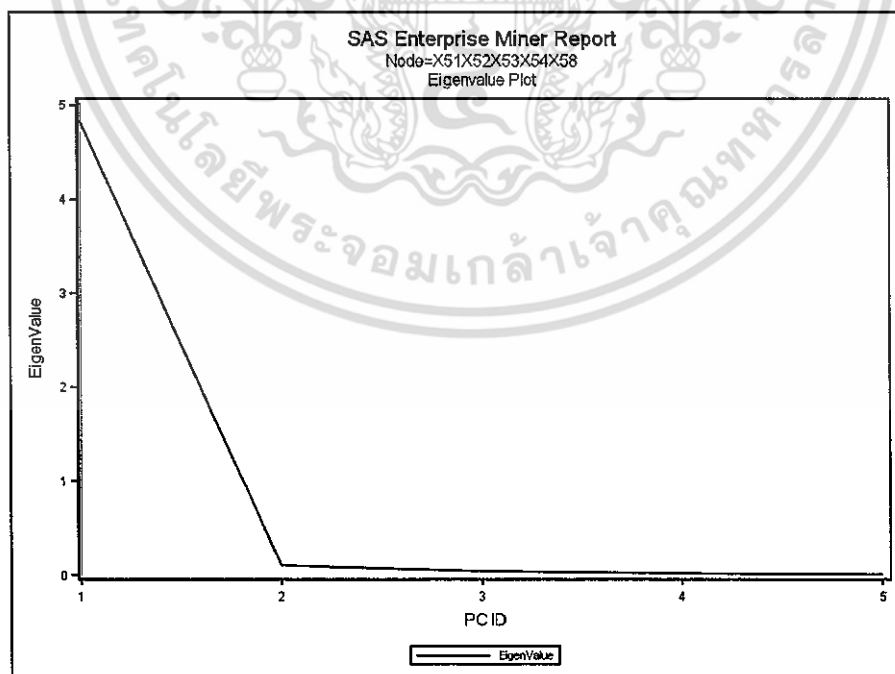
รูปที่ ข.2 จำนวนตัวแปรอิสระที่ผ่านการทดสอบไคสแควร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าลักษณะเฉพาะ (Eigenvalue) ของการสร้างตัวแปรใหม่ด้วยวิธีการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis)

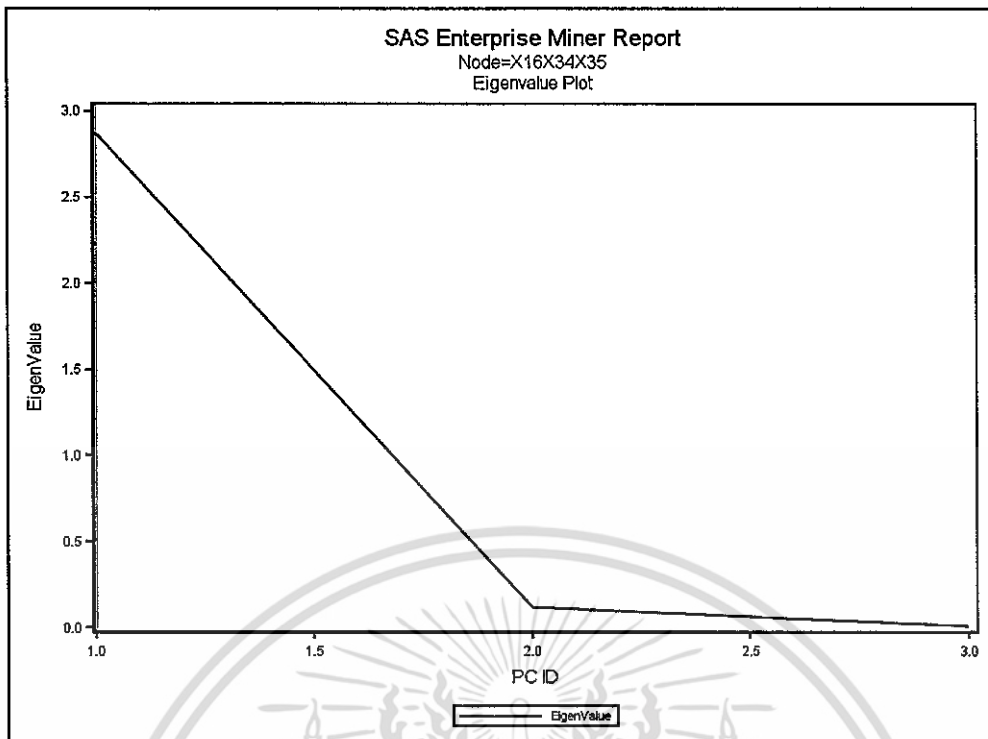


รูปที่ ข.3 ค่าลักษณะเฉพาะของตัวแปร X_{136} , X_{147}

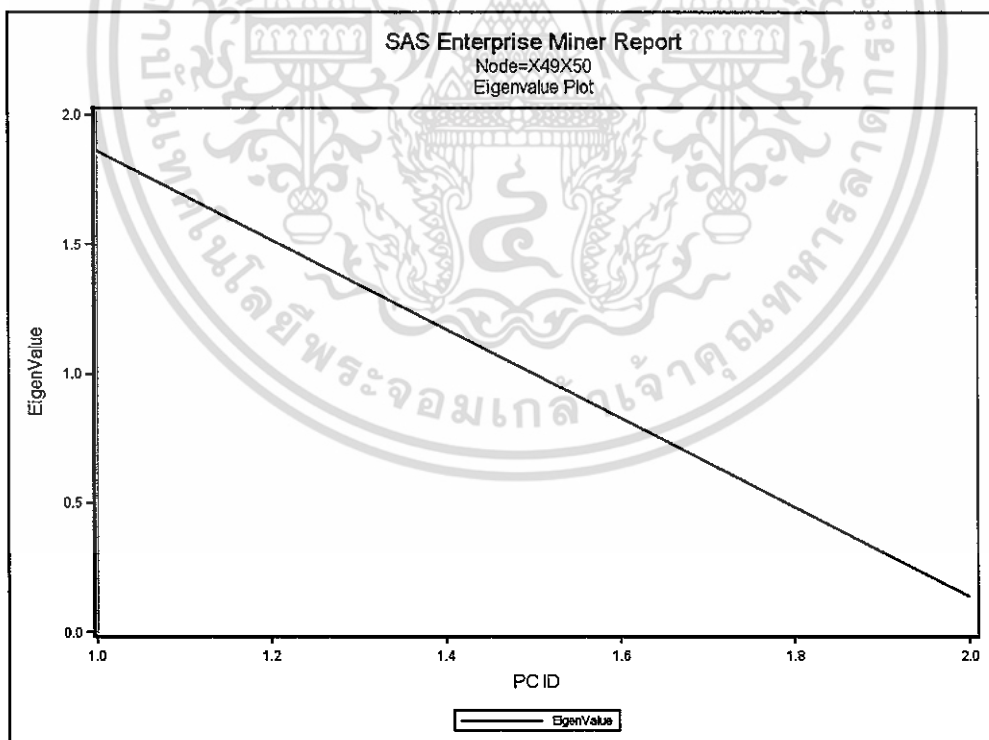


รูปที่ ข.4 ค่าลักษณะเฉพาะของตัวแปร X_{51} , X_{52} , X_{53} , X_{54} และ X_{58}

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ข.5 ค่าลักษณะเฉพาะของตัวแปร X_{16} , X_{34} , และ X_{35}



รูปที่ ข.6 ค่าลักษณะเฉพาะของตัวแปร X_{49} , X_{50}

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สมการตัวแปรใหม่ที่สร้างจากวิธีการวิเคราะห์องค์ประกอบหลัก

X_{49N}

$$X_{49N} = 0.7071067812 * ((X_{49}-14872.8328586032)/722.486654176743) \\ + 0.7071067812 * ((X_{50}-14804.1986820035)/495.16081844793)$$

X_{136N}

$$X_{136N} = 0.7071067812 * ((X_{136}-87.9392610089421)/12.3029280466786) \\ + 0.7071067812 * ((X_{147}-93.3775245883742)/14.4872033238276)$$

X_{51N}

$$X_{51N} = 0.4477511562 * ((X_{51}-304.217221290405)/36.0872744624743) \\ + 0.4484086799 * ((X_{52}-291.350211891741)/35.3609247143665) \\ + 0.4472088051 * ((X_{53}-302.394158338213)/35.5539027898313) \\ + 0.4460769526 * ((X_{54}-321.2668843478)/35.7724166838206) \\ + 0.4466186237 * ((X_{58}-286.788902232058)/36.0381273115602)$$

X_{16N}

$$X_{16N} = 0.5833557893 * ((X_{16}-3719.45732887384)/247.433956233994) \\ + 0.5658828616 * ((X_{34}-3436.5478518048)/272.891165688682) \\ + 0.5826427809 * ((X_{35}-3530.10376244305)/223.060633308424)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.4 การตั้งค่าการวิเคราะห์การถดถอยโลจิสติก (Logistics Regression Analysis)

Node=Regression
Properties

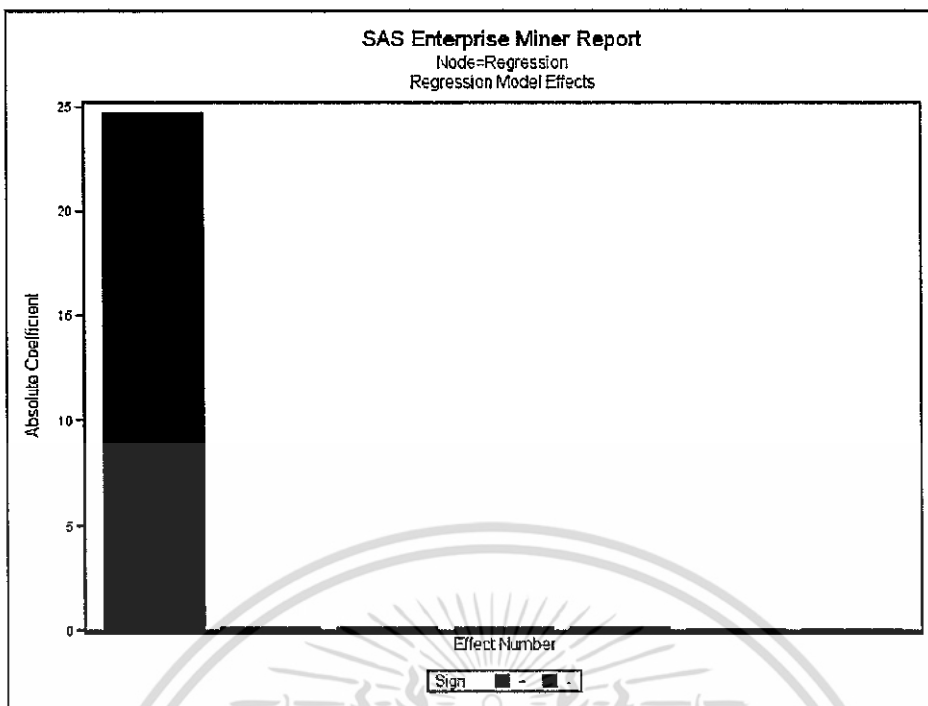
Property	Value	Default	Property	Value	Default	Property	Value	Default
Component	Regression		Force	0		PolynomialDegree	2	
AbsConvValue	-1.34078E154	-7.237008E75	GConvTimes	1		PrintDesignMatrix	N	
AbsFTime	1		GConvValue	1E-8		Rule	NONE	
AbsFValue	0		Hierarchy	CLASS		SASSPDS	N	
AbsGTime	1		InputCoding	DEVIATION		SelectionCriterion	DEFAULT	
AbsGValue	0.00001		Interactions			SelectionDefault	Y	
AbsXTime	1		LinkFunction	LOGIT		Sequential	N	
AbsXValue	1E-8		MainEffect	Y		Simple	Y	N
CIParam	N		MaxCPUTime	1 HOUR		SIEntry	0.05	
ConvDefaults	Y		MaxFunctionCalls	.		SISlay	0.05	
CorB	Y	N	MaxIterations	.		Start	0	
CorB	N		MaxStep	.		StepOutput	N	
Covout	N		MinResourceUse	N		Stop	0	
Details	Y	N	ModelDefaults	Y		SuppressIntercept	N	
Error	LOGISTIC		ModelSelection	STEPWISE	NONE	SuppressOutput	N	
ExcludedVariable	REJECT		OptimizationTechnique	DEFAULT		Terms	N	
FConvTimes	1		Performance	N		TwoFactor	N	
FConvValue	0		Polynomial	N				

ตารางที่ ข.5 ตัวแปรอิสระที่ใช้ในการวิเคราะห์การถดถอยโลจิสติก จำนวน 17 ตัวแปร

Node=Regression
Variable Summary

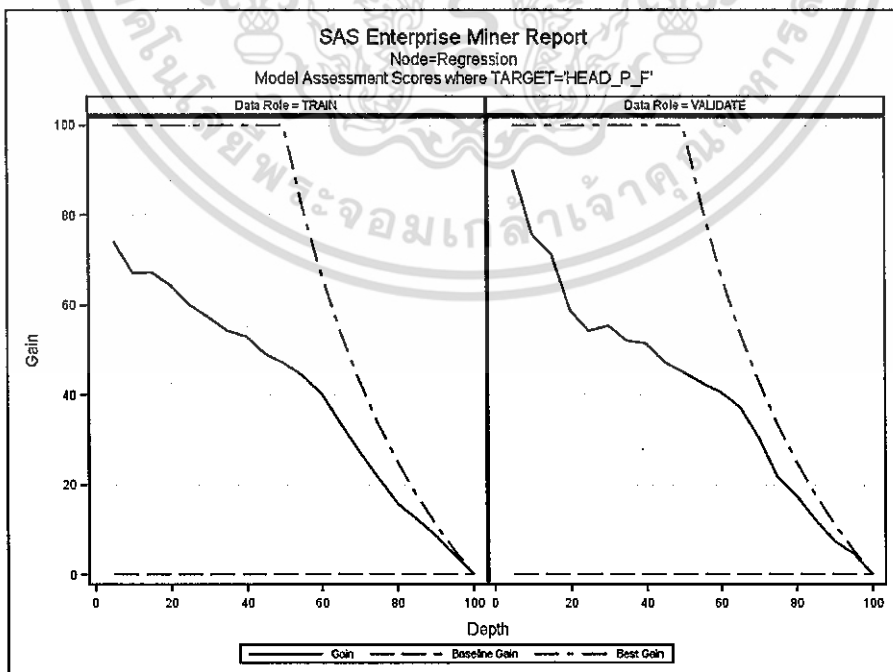
Role	Level	Frequency Count	Name
TARGET	BINARY	1	HEAD_P_F
INPUT	INTERVAL	17	X102 X111 X121 X133 X130N_1 X146 X16N_1 X175 X176 X178 X28 X38 X49N_1 X51N_1 X61 X73 X94

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



Effect Number	Variable	Level	Coefficient	T-value	P Value	Effect Number	Variable	Level	Coefficient	T-value	P Value
1	Intercept	1	24.7035	15.7983	3.1977E-50	5	X111		0.14344	5.3487	.000000080
2	X136N_1		-0.1751	-5.2782	.000000130	6	X16N_1		-0.07872	-2.9152	.003555148
3	X178		-0.1588	-6.6353	3.2378E-11	7	X102		-0.05548	-17.8138	5.5200E-71
4	X133		-0.1523	-3.9260	.000093364						

รูปที่ ข.7 ตัวแปรอิสระที่ได้จากการวิเคราะห์การถดถอยโลจิสติก



รูปที่ ข.8 Gain Chart การวิเคราะห์การถดถอยโลจิสติก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.6 Score การวิเคราะห์การถดถอยโลจิสติก

Node=Regression
Score Distributions

Target Variable=HEAD_P_F Data Role=TRAIN

Posterior Probability Range	Number of Events	Percentage of Events	Percentage of Nonevents	Cumulative Percentage of Events	Cumulative Percentage of Nonevents
0.95-1.00	228.42	0.4484	0.0199	0.449	0.0199
0.90-0.95	1471.78	2.9213	0.2831	3.371	0.3030
0.85-0.90	2830.31	5.6180	1.0828	8.989	1.3858
0.80-0.85	5207.77	10.3371	2.3157	19.326	3.7014
0.75-0.80	8879.53	13.2594	3.3575	32.584	7.0588
0.70-0.75	5094.58	10.1124	4.0742	42.697	11.1330
0.65-0.70	4415.28	8.7640	4.2738	51.461	15.4068
0.60-0.65	5980.62	11.2360	4.1928	62.697	19.5995
0.55-0.60	2943.52	5.8427	4.0712	68.539	23.6707
0.50-0.55	3398.37	6.7418	4.2188	75.281	27.8894
0.45-0.50	2830.31	5.6180	4.7700	80.899	32.6594
0.40-0.45	1924.81	3.8202	5.6991	84.719	38.3588
0.35-0.40	1245.34	2.4719	7.3160	87.191	45.6746
0.30-0.35	1245.34	2.4719	9.1931	89.663	54.8677
0.25-0.30	1358.55	2.6966	11.5158	92.360	66.3836
0.20-0.25	2377.48	4.7181	12.8588	97.079	79.2422
0.15-0.20	905.70	1.7978	11.6016	98.876	90.8438
0.10-0.15	566.06	1.1236	7.1478	100.000	97.9913
0.05-0.10	0.00	0.0000	1.9828	100.000	99.9541
0.00-0.05	0.00	0.0000	0.0459	100.000	100.000

ตารางที่ ข.7 การตั้งค่าการวิเคราะห์โครงข่ายประสาทเทียม

Node=Neural Network
Properties

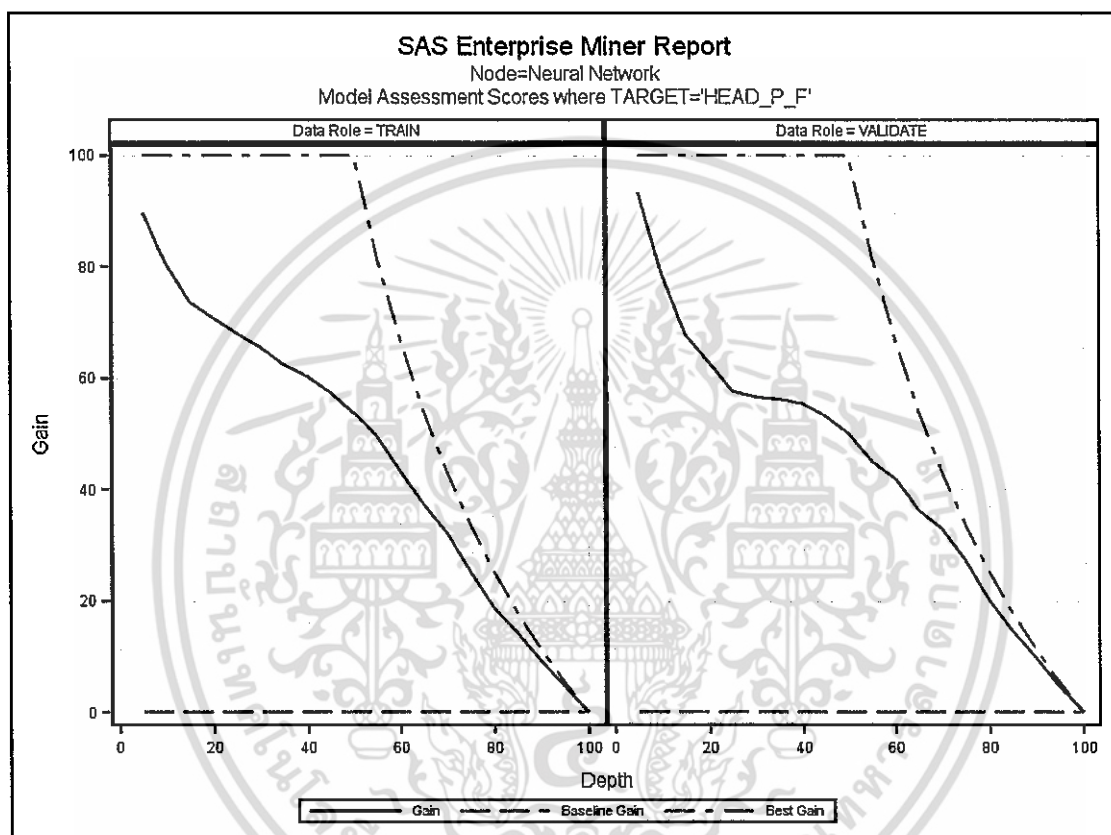
Property	Value	Default	Property	Value	Default	Property	Value	Default
Component	NeuralNetwork		Hidden	4	3	Prelim	Y	
AbsConvValue	-1.34078E154	-7.237006E75	HiddenActivation	DEFAULT		PrelimMaxTime	1 HOUR	
AbsPTime	1		HiddenBias	Y		PrelimMaxIter	10	
AbsPValue	0		HiddenCombFunction	DEFAULT		PrelimOutest		
AbsGTime	1		HiddenUnits	Y	N	PreliminaryRuns	5	
AbsGValue	0.00001		InitialDs			RandDist	NORMAL	
AbsXTime	1		InitialSeed	12345		RandLoc	0	
AbsXValue	1E-8		InputStandardization	STD		RandScale	0.1	
Accelerate	1.2		Learn	0.1		Residuals	Y	
AddHidden	Y		MaxLearn	50		Standardizations	N	
CodefileNoRes			MaxMomentum	1.75		SuppressOutput	N	
CodefileRes			MaxIter	50		TargetActivation	DEFAULT	
ConvDefaults	Y		Maxtime	4 HOURS		TargetBias	Y	
Decelerate	0.5		MinLearn	0.00001		TargetCombFunction	DEFAULT	
DirectConnection	Y	N	ModelSelectionCriterion	MISCLASSIFICATION	PROFITLOSS	TargetError	DEFAULT	
FConvTime	1		Momentum	0		TIR	0	
FConvValue	0		NetworkArchitecture	MLP		TrainCode		
GConvTime	1		Outest			TrainingTechnique	DEFAULT	
GConvValue	1E-6		OutIt			UseEstimates	N	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.8 จำนวนตัวแปรอิสระที่ใช้ในการวิเคราะห์โครงข่ายประสาทเทียม จำนวน 25 ตัวแปร

Node=Neural Network
Variable Summary

Role	Level	Frequency Count	Name
TARGET	BINARY	1	HEAD_P_F
INPUT	INTERVAL	25	X102 X111 X121 X133 X136 X146 X147 X16 X175 X176 X178 X28 X34 X35 X36 X49 X50 X51 X52 X53 X54 X56 X61 X73 X14



รูปที่ ข.9 Gain Chart ของการวิเคราะห์โครงข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.9 Score การวิเคราะห์โครงข่ายประสาทเทียม

Node=Neural Network Score Distributions

Target Variable=HEAD_P_F Data Role=TRAIN

Posterior Probability Range	Number of Events	Percentage of Events	Percentage of Nonevents	Cumulative Percentage of Events	Cumulative Percentage of Nonevents
0.95-1.00	1245.34	2.4719	0.0748	2.472	0.0748
0.90-0.95	3283.16	6.5169	0.3788	8.989	0.4536
0.85-0.90	4415.28	8.7640	1.3797	17.753	1.8332
0.80-0.85	6226.68	12.3596	3.0155	30.112	4.8488
0.75-0.80	7358.80	14.6067	4.0333	44.719	8.8821
0.70-0.75	6000.26	11.9101	4.1141	56.629	12.9962
0.65-0.70	4528.49	8.9888	3.4751	65.618	16.4713
0.60-0.65	2717.10	5.3933	2.7274	71.011	19.1987
0.55-0.60	1811.40	3.5955	2.2798	74.607	21.4786
0.50-0.55	1132.12	2.2472	2.0705	76.854	23.5491
0.45-0.50	1018.91	2.0225	2.2739	78.876	25.8229
0.40-0.45	2037.82	4.0449	2.5231	82.921	28.3460
0.35-0.40	1018.91	2.0225	3.2598	84.944	31.6058
0.30-0.35	905.70	1.7978	4.4052	86.742	36.0109
0.25-0.30	1698.19	3.3708	6.1407	90.112	42.1516
0.20-0.25	1471.76	2.9213	8.9030	93.034	51.0547
0.15-0.20	792.49	1.5730	13.3820	94.607	64.4367
0.10-0.15	2151.03	4.2697	20.3730	98.876	84.8097
0.05-0.10	566.06	1.1236	14.7218	100.000	99.5315
0.00-0.05	0.00	0.0000	0.4685	100.000	100.000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Code สมการการวิเคราะห์การถดถอยโลจิสติก

```

*****;

*** begin scoring code for regression;

*****;

length _WARN_ $4;

label _WARN_ = 'Warnings' ;

length I_HEAD_P_F $ 12;

label I_HEAD_P_F = 'Into: HEAD_P_F' ;

*** Target Values;

array REGDRF [2] $12 _temporary_ ('1' '0' );

label U_HEAD_P_F = 'Unnormalized Into: HEAD_P_F' ;

format U_HEAD_P_F BEST12.;

*** Unnormalized target values;

ARRAY REGDRU[2] _TEMPORARY_ (1 0);

drop _DM_BAD;

_DM_BAD=0;

*** Check X102 for missing values ;

if missing( X102 ) then do;

    substr(_warn_,1,1) = 'M';

    _DM_BAD = 1;

end;

*** Check X111 for missing values ;

if missing( X111 ) then do;

    substr(_warn_,1,1) = 'M';

    _DM_BAD = 1;

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

end;

*** Check X133 for missing values ;

if missing( X133 ) then do;

    substr(_warn_,1,1) = 'M';

    _DM_BAD = 1;

end;

*** Check X136N_1 for missing values ;

if missing( X136N_1 ) then do;

    substr(_warn_,1,1) = 'M';

    _DM_BAD = 1;

end;

*** Check X16N_1 for missing values ;

if missing( X16N_1 ) then do;

    substr(_warn_,1,1) = 'M';

    _DM_BAD = 1;

end;

*** Check X176 for missing values ;

if missing( X176 ) then do;

    substr(_warn_,1,1) = 'M';

    _DM_BAD = 1;

end;

*** If missing inputs, use averages;

if _DM_BAD > 0 then do;

    _P0 = 0.0044164789;

    _P1 = 0.9955835211;

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

goto REGDR1;

end;

*** Compute Linear Predictor;

drop _TEMP;

drop _LP0;

_LP0 = 0;

*** Effect: X102 ;

_TEMP = X102 ;

_LP0 = _LP0 + ( -0.05545580557093 * _TEMP);

*** Effect: X111 ;

_TEMP = X111 ;

_LP0 = _LP0 + ( 0.14344264468352 * _TEMP);

*** Effect: X133 ;

_TEMP = X133 ;

_LP0 = _LP0 + ( -0.15230892624802 * _TEMP);

*** Effect: X136N_1 ;

_TEMP = X136N_1 ;

_LP0 = _LP0 + ( -0.17506820169739 * _TEMP);

*** Effect: X16N_1 ;

_TEMP = X16N_1 ;

_LP0 = _LP0 + ( -0.0787201605603 * _TEMP);

*** Effect: X176 ;

_TEMP = X176 ;

_LP0 = _LP0 + ( -0.15881779148638 * _TEMP);

```

เอกสารนี้เป็นเอกสารที่สวอนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

*** Naive Posterior Probabilities;

drop _MAXP _IY _P0 _P1;

_TEMP = 24.7055210195399 + _LP0;

if (_TEMP < 0) then do;

    _TEMP = exp(_TEMP);

    _P0 = _TEMP / (1 + _TEMP);

end;

else _P0 = 1 / (1 + exp(-_TEMP));

_P1 = 1.0 - _P0;

REGDR1:

*** Update Posterior Probabilities;

_P0 = _P0 * 0.5 / 0.00441647892495;

_P1 = _P1 * 0.5 / 0.99558352107504;

drop _sum; _sum = _P0 + _P1 ;

if _sum > 4.135903E-25 then do;

    _P0 = _P0 / _sum;

    _P1 = _P1 / _sum;

end;

*** End Decision Processing ;

*** Posterior Probabilities and Predicted Level;

label P_HEAD_P_F1 = 'Predicted: HEAD_P_F=1' ;

label P_HEAD_P_F0 = 'Predicted: HEAD_P_F=0' ;

P_HEAD_P_F1 = _P0;

_MAXP = _P0;

_IY = 1;

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

P_HEAD_P_F0 = _P1;

if (_P1 > _MAXP + 1E-8) then do;
    _MAXP = _P1;
    _IY = 2;
end;

I_HEAD_P_F = REGDRF[_IY];
U_HEAD_P_F = REGDRU[_IY];

*****
***** end scoring code for regression;
*****

```



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Code สมการการวิเคราะห์โครงข่ายประสาทเทียม

```

*****;
*** Begin Scoring Code for Neural;
*****;

DROP _DM_BAD_EPS_NOCL__MAX__MAXP__SUM__NTRIALS;

_DM_BAD = 0;

_NOCL_ = .;

_MAX_ = .;

_MAXP_ = .;

_SUM_ = .;

_NTRIALS = .;

_EPS = 1E-10;

LENGTH _WARN_ $4
;

label S_X102 = 'Standard: X102' ;
label S_X111 = 'Standard: X111' ;
label S_X121 = 'Standard: X121' ;
label S_X133 = 'Standard: X133' ;
label S_X136 = 'Standard: X136' ;
label S_X146 = 'Standard: X146' ;
label S_X147 = 'Standard: X147' ;
label S_X157 = 'Standard: X157' ;
label S_X16 = 'Standard: X16' ;
label S_X176 = 'Standard: X176' ;
label S_X178 = 'Standard: X178' ;

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

label S_X180 = 'Standard: X180' ;
 label S_X28 = 'Standard: X28' ;
 label S_X34 = 'Standard: X34' ;
 label S_X35 = 'Standard: X35' ;
 label S_X38 = 'Standard: X38' ;
 label S_X49 = 'Standard: X49' ;
 label S_X50 = 'Standard: X50' ;
 label S_X51 = 'Standard: X51' ;
 label S_X52 = 'Standard: X52' ;
 label S_X53 = 'Standard: X53' ;
 label S_X54 = 'Standard: X54' ;
 label S_X58 = 'Standard: X58' ;
 label S_X61 = 'Standard: X61' ;
 label S_X73 = 'Standard: X73' ;
 label S_X94 = 'Standard: X94' ;
 label H11 = 'Hidden: H1=1' ;
 label H12 = 'Hidden: H1=2' ;
 label H13 = 'Hidden: H1=3' ;
 label H14 = 'Hidden: H1=4' ;
 label I_HEAD_P_F = 'Into: HEAD_P_F' ;
 label U_HEAD_P_F = 'Unnormalized Into: HEAD_P_F' ;
 label P_HEAD_P_F1 = 'Predicted: HEAD_P_F=1' ;
 label P_HEAD_P_F0 = 'Predicted: HEAD_P_F=0' ;
 label _WARN_ = "Warnings";

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

*** *****;

*** Checking missing input Interval

*** *****;

IF NMISS(

X102 ,

X111 ,

X121 ,

X133 ,

X136 ,

X146 ,

X147 ,

X157 ,

X16 ,

X176 ,

X178 ,

X180 ,

X28 ,

X34 ,

X35 ,

X38 ,

X49 ,

X50 ,

X51 ,

X52 ,

X53 ,



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

X54 ,
X58 ,
X61 ,
X73 ,
X94 ) THEN DO;
SUBSTR(_WARN_, 1, 1) = 'M';
_DM_BAD = 1;
END;
*** *****;
*** Writing the Node intvl ;
*** *****;
IF _DM_BAD EQ 0 THEN DO;
S_X102 = -30.3737392751463 + 0.06562545054903 * X102 ;
S_X111 = 2.05895676309441 + 0.71837370707551 * X111 ;
S_X121 = 0.06707868375181 + 1.28104266416773 * X121 ;
S_X133 = 0.10398395262241 + 0.95179134105034 * X133 ;
S_X136 = -7.14783185557867 + 0.08128146374634 * X136 ;
S_X146 = -2.64073739412519 + 0.27800183583548 * X146 ;
S_X147 = -6.44551764071625 + 0.06902643509912 * X147 ;
S_X157 = -8.93751948574923 + 0.10615053873128 * X157 ;
S_X16 = -15.0321216436292 + 0.00404148248373 * X16 ;
S_X176 = -21.5312621052115 + 0.74213348897747 * X176 ;
S_X178 = -22.5643018279315 + 0.77455465929999 * X178 ;
S_X180 = -390.725938108647 + 19.5366846957974 * X180 ;
S_X28 = -0.03540188267168 + 0.01093877604277 * X28 ;

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

S_X34 = -12.5931077436389 + 0.00366446454019 * X34 ;
S_X35 = -15.8257587189847 + 0.00448308598952 * X35 ;
S_X38 = -0.7795505830249 + 0.04845360702585 * X38 ;
S_X49 = -20.5856160423481 + 0.00138410861186 * X49 ;
S_X50 = -29.8977587289886 + 0.00201954589851 * X50 ;
S_X51 = -8.43004149861056 + 0.02771059923186 * X51 ;
S_X52 = -8.23932672137867 + 0.02827980342928 * X52 ;
S_X53 = -8.50523106072899 + 0.02812630742428 * X53 ;
S_X54 = -8.98085491923458 + 0.02795449937974 * X54 ;
S_X58 = -7.95793021520472 + 0.02774838968059 * X58 ;
S_X61 = -0.03913803376998 + 0.69184370958422 * X61 ;
S_X73 = 0.28164009305875 + 1.08274921349564 * X73 ;
S_X94 = -26.9737207592359 + 0.05171280810459 * X94 ;
END;
ELSE DO;
IF MISSING( X102 ) THEN S_X102 = . ;
ELSE S_X102 = -30.3737392751463 + 0.06562545054903 * X102 ;
IF MISSING( X111 ) THEN S_X111 = . ;
ELSE S_X111 = 2.05895676309441 + 0.71837370707551 * X111 ;
IF MISSING( X121 ) THEN S_X121 = . ;
ELSE S_X121 = 0.06707868375181 + 1.28104266416773 * X121 ;
IF MISSING( X133 ) THEN S_X133 = . ;
ELSE S_X133 = 0.10398395262241 + 0.95179134105034 * X133 ;
IF MISSING( X136 ) THEN S_X136 = . ;
ELSE S_X136 = -7.14783185557867 + 0.08128146374634 * X136 ;

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

IF MISSING( X146 ) THEN S_X146 = . ;
ELSE S_X146 = -2.64073739412519 + 0.27800183583548 * X146 ;
IF MISSING( X147 ) THEN S_X147 = . ;
ELSE S_X147 = -6.44551764071625 + 0.06902643509912 * X147 ;
IF MISSING( X157 ) THEN S_X157 = . ;
ELSE S_X157 = -8.93751948574923 + 0.10615053873128 * X157 ;
IF MISSING( X16 ) THEN S_X16 = . ;
ELSE S_X16 = -15.0321216436292 + 0.00404148248373 * X16 ;
IF MISSING( X176 ) THEN S_X176 = . ;
ELSE S_X176 = -21.5312621052115 + 0.74213348897747 * X176 ;
IF MISSING( X178 ) THEN S_X178 = . ;
ELSE S_X178 = -22.5643018279315 + 0.77455465929999 * X178 ;
IF MISSING( X180 ) THEN S_X180 = . ;
ELSE S_X180 = -390.725938108647 + 19.5366846957974 * X180 ;
IF MISSING( X28 ) THEN S_X28 = . ;
ELSE S_X28 = -0.03540188267168 + 0.01093877604277 * X28 ;
IF MISSING( X34 ) THEN S_X34 = . ;
ELSE S_X34 = -12.5931077436389 + 0.00366446454019 * X34 ;
IF MISSING( X35 ) THEN S_X35 = . ;
ELSE S_X35 = -15.8257587189847 + 0.00448308598952 * X35 ;
IF MISSING( X38 ) THEN S_X38 = . ;
ELSE S_X38 = -0.7795505830249 + 0.04845360702585 * X38 ;
IF MISSING( X49 ) THEN S_X49 = . ;
ELSE S_X49 = -20.5856160423481 + 0.00138410861186 * X49 ;
IF MISSING( X50 ) THEN S_X50 = . ;

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

ELSE S_X50 = -29.8977587289886 + 0.00201954589851 * X50 ;
IF MISSING( X51 ) THEN S_X51 = . ;
ELSE S_X51 = -8.43004149861056 + 0.02771059923186 * X51 ;
IF MISSING( X52 ) THEN S_X52 = . ;
ELSE S_X52 = -8.23932672137867 + 0.02827980342928 * X52 ;
IF MISSING( X53 ) THEN S_X53 = . ;
ELSE S_X53 = -8.50523106072899 + 0.02812630742428 * X53 ;
IF MISSING( X54 ) THEN S_X54 = . ;
ELSE S_X54 = -8.98085491923458 + 0.02795449937974 * X54 ;
IF MISSING( X58 ) THEN S_X58 = . ;
ELSE S_X58 = -7.95793021520472 + 0.02774838968059 * X58 ;
IF MISSING( X61 ) THEN S_X61 = . ;
ELSE S_X61 = -0.03913803376998 + 0.69184370958422 * X61 ;
IF MISSING( X73 ) THEN S_X73 = . ;
ELSE S_X73 = 0.28164009305875 + 1.08274921349564 * X73 ;
IF MISSING( X94 ) THEN S_X94 = . ;
ELSE S_X94 = -26.9737207592359 + 0.05171280810459 * X94 ;

END;

*** *****;

*** Writing the Node H1 ;

*** *****;

IF _DM_BAD EQ 0 THEN DO;

H11 = 0.14466121028496 * S_X102 + -0.02577599256309 * S_X111
      + -0.18861156294529 * S_X121 + -0.02017308497972 * S_X133
      + 0.21046595493311 * S_X136 + 0.23199105899316 * S_X146

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\begin{aligned}
& + -0.03505967348244 * S_X147 + -0.06816679074015 * S_X157 \\
& + 0.13927267979002 * S_X16 + -0.02004439117038 * S_X176 \\
& + -0.13477347340156 * S_X178 + -0.2759485068746 * S_X180 \\
& + -0.18809308165817 * S_X28 + 0.10553232412925 * S_X34 \\
& + 0.21930944326147 * S_X35 + -0.34585085162091 * S_X38 \\
& + 0.30034627947745 * S_X49 + -0.30041524929267 * S_X50 \\
& + -0.24848664836139 * S_X51 + 0.00993126676763 * S_X52 \\
& + 0.26813094999875 * S_X53 + -0.09434635137839 * S_X54 \\
& + -0.2682467622534 * S_X58 + -0.24857342563863 * S_X61 \\
& + 0.27127156706181 * S_X73 + -0.05202883426154 * S_X94 ; \\
H12 = & -0.31876016683107 * S_X102 + -0.00596833004245 * S_X111 \\
& + 0.18052723093763 * S_X121 + 0.08840064359622 * S_X133 \\
& + 0.06206889166337 * S_X136 + 0.08600932746943 * S_X146 \\
& + -0.29133849604732 * S_X147 + 0.24606127859197 * S_X157 \\
& + 0.14597113837863 * S_X16 + -0.44492484969788 * S_X176 \\
& + 0.22540161510335 * S_X178 + -0.01616723830539 * S_X180 \\
& + -0.11602215372271 * S_X28 + -0.18735728262056 * S_X34 \\
& + -0.07923244555719 * S_X35 + -0.15092355468655 * S_X38 \\
& + -0.13403800506874 * S_X49 + 0.01313838341089 * S_X50 \\
& + -0.16522200783156 * S_X51 + -0.04780585971692 * S_X52 \\
& + 0.13126204016916 * S_X53 + 0.1650804165982 * S_X54 \\
& + -0.13618023584927 * S_X58 + 0.21636886411246 * S_X61 \\
& + 0.02837589581552 * S_X73 + -0.41891681314402 * S_X94 ; \\
H13 = & 0.16707204152008 * S_X102 + -0.04257445770947 * S_X111 \\
& + -0.20127014352461 * S_X121 + 0.38110388550463 * S_X133.
\end{aligned}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\begin{aligned}
& + 0.16523179066702 * S_X136 + 0.12526245485494 * S_X146 \\
& + 0.0538601870637 * S_X147 + 0.26590286023546 * S_X157 \\
& + 0.17261048450984 * S_X16 + -0.21706814426932 * S_X176 \\
& + 0.46424357076669 * S_X178 + 0.12647824356427 * S_X180 \\
& + -0.27480419631334 * S_X28 + 0.09149212559907 * S_X34 \\
& + 0.3389938968161 * S_X35 + -0.11065309701283 * S_X38 \\
& + 0.3620571227729 * S_X49 + -0.28050863453465 * S_X50 \\
& + -0.20805895283705 * S_X51 + -0.07545168991954 * S_X52 \\
& + 0.33212425644201 * S_X53 + -0.13572182003681 * S_X54 \\
& + -0.13767265624499 * S_X58 + -0.14093812980047 * S_X61 \\
& + -0.00538750030407 * S_X73 + -0.0025656217434 * S_X94 ; \\
H14 = & -0.25589496129325 * S_X102 + 0.77534736143727 * S_X111 \\
& + -0.308612688439 * S_X121 + -0.16780397063093 * S_X133 \\
& + 0.00290860966726 * S_X136 + -0.06484956571156 * S_X146 \\
& + 0.1052997869901 * S_X147 + 0.19513458692327 * S_X157 \\
& + 0.20039177085039 * S_X16 + 0.01947584915419 * S_X176 \\
& + -0.2568942419454 * S_X178 + 0.16067508195468 * S_X180 \\
& + 0.02176690758904 * S_X28 + 0.04531576929095 * S_X34 \\
& + 0.23783204448933 * S_X35 + 0.02906399759605 * S_X38 \\
& + 0.48389331305862 * S_X49 + -0.61580043039116 * S_X50 \\
& + -0.51166872590275 * S_X51 + 0.06169879522564 * S_X52 \\
& + -0.36942382063494 * S_X53 + -0.09124598010947 * S_X54 \\
& + -0.10615720478817 * S_X58 + 0.17856187453152 * S_X61 \\
& + -0.38516671496033 * S_X73 + 0.45808686370354 * S_X94 ; \\
H11 = & -0.55987920707624 + H11 ;
\end{aligned}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

H12 = 0.02178828217646 + H12 ;
H13 = -1.14038370310514 + H13 ;
H14 = -0.68571347180421 + H14 ;

H11 = TANH(H11 );
H12 = TANH(H12 );
H13 = TANH(H13 );
H14 = TANH(H14 );

END;

ELSE DO;

H11 = ;
H12 = ;
H13 = ;
H14 = ;

END;

*** *****
*** Writing the Node HEAD_P_F ;
*** *****

IF _DM_BAD EQ 0 THEN DO;

P_HEAD_P_F1 = 0.56216867083047 * H11 + 0.05671051977203 * H12
+ 0.52681091469945 * H13 + 0.89226135824664 * H14 ;

P_HEAD_P_F1 = P_HEAD_P_F1 + -0.20990567194658 * S_X102
+ -0.06430966786354 * S_X111 + 0.07627514124303 * S_X121
+ -0.08603248186272 * S_X133 + -0.11479812018437 * S_X136
+ -0.30639456127822 * S_X146 + 0.01898988702206 * S_X147
+ -0.05429124552377 * S_X157 + 0.17353439039058 * S_X16

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

+ -0.12601811489547 * S_X176 + -0.04427425722958 * S_X178
+ 0.00161021185361 * S_X180 + 0.01040414820507 * S_X28
+ -0.07210527271615 * S_X34 + -0.66472255537007 * S_X35
+ -0.050425252589 * S_X38 + 1.2897008079716 * S_X49
+ -0.8703457339886 * S_X50 + -0.06723550893411 * S_X51
+ 0.68564480585326 * S_X52 + 0.01766006067622 * S_X53
+ 0.18730432285308 * S_X54 + -0.47181186725153 * S_X58
+ 0.02179606029248 * S_X61 + -0.00772278432239 * S_X73
+ -0.18145528183427 * S_X94 ;

```

```
P_HEAD_P_F1 = -5.69210921959662 + P_HEAD_P_F1 ;
```

```
P_HEAD_P_F0 = 0;
```

```
_MAX_ = MAX (P_HEAD_P_F1 , P_HEAD_P_F0 );
```

```
_SUM_ = 0.;
```

```
P_HEAD_P_F1 = EXP(P_HEAD_P_F1 - _MAX_);
```

```
_SUM_ = _SUM_ + P_HEAD_P_F1 ;
```

```
P_HEAD_P_F0 = EXP(P_HEAD_P_F0 - _MAX_);
```

```
_SUM_ = _SUM_ + P_HEAD_P_F0 ;
```

```
P_HEAD_P_F1 = P_HEAD_P_F1 / _SUM_;
```

```
P_HEAD_P_F0 = P_HEAD_P_F0 / _SUM_;
```

```
END;
```

```
ELSE DO;
```

```
P_HEAD_P_F1 = .;
```

```
P_HEAD_P_F0 = .;
```

```
END;
```

```
IF _DM_BAD EQ 1 THEN DO;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

P_HEAD_P_F1 = 0.00441647892495;
P_HEAD_P_F0 = 0.99558352107504;

END;

*** Update Posterior Probabilities;

P_HEAD_P_F1 = P_HEAD_P_F1 * 0.5 / 0.00441647892495;
P_HEAD_P_F0 = P_HEAD_P_F0 * 0.5 / 0.99558352107504;

drop _sum; _sum = P_HEAD_P_F1 + P_HEAD_P_F0 ;

if _sum > 4.135903E-25 then do;

    P_HEAD_P_F1 = P_HEAD_P_F1 / _sum;
    P_HEAD_P_F0 = P_HEAD_P_F0 / _sum;

end;

*** End Decision Processing ;
*** *****,
*** Writing the I_HEAD_P_F AND U_HEAD_P_F ;
*** *****,
_MAXP_ = P_HEAD_P_F1 ;
I_HEAD_P_F = "1" ;
U_HEAD_P_F = 1;
IF( _MAXP_ LT P_HEAD_P_F0 ) THEN DO;
    _MAXP_ = P_HEAD_P_F0 ;
    I_HEAD_P_F = "0" ;
    U_HEAD_P_F = 0;
END;

*****,
*** End Scoring Code for Neural;

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ นางสาวจามรี ชูบัวทอง

วัน เดือน ปีเกิด 24 กุมภาพันธ์ 2529

ที่อยู่ปัจจุบัน 1/1 ม.1 ต.วະตะแบก อ.เทพสถิต จ.ชัยภูมิ 36230
โทร. 094-679-1313 E-Mail: suay-physics@hotmail.com

ประวัติการศึกษา พ.ศ.2551 วิทยาศาสตรบัณฑิต สาขาวิชาฟิสิกส์ เกรตเฉลี่ย 3.19
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
พ.ศ.2558 วิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์
เกรตเฉลี่ย 3.61 สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้