การประยุกต์ใช้เทคโนโลยีการทำเหมืองข้อมูลกับธุรกิจเครื่องจำหน่ายสินค้าอัตโนมัติ

DATA MINING FOR AUTOMATE VENDING MACHINE BUSINESS

นายชยุต ตั้งสกุลวัฒนา
นางสาวพรฉาย อภิชาติพันธุ์

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2560

การประยุกต์ใช้เทคโนโลยีการทำเหมืองข้อมูลกับธุรกิจเครื่องจำหน่ายสินค้าอัตโนมัติ

DATA MINING FOR AUTOMATE VENDING MACHINE BUSINESS

นายชยุต ตั้งสกุลวัฒนา

นางสาวพรฉาย อภิชาติพันธุ์

ID 00265371

TB00106

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2560

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปีการศึกษา 2560

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การประยุกต์ใช้เทคโนโลยีการทำเหมืองข้อมูลกับธุรกิจเครื่องจำหน่ายสินค้าอัตโนมัติ

DATA MINING FOR AUTOMATE VENDING MACHINE BUSINESS

ผู้จัดทำ

1. นายชยุต ตั้งสกุลวัฒนา รหัสนักศึกษา 57010265

2. นางสาวพรฉาย อภิชาติพันธุ์ รหัสนักศึกษา 57010835

_____ อาจารย์ที่ปรึกษา

(ดร.วัชระ ฉัตรวิริยะ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การประยุกต์ใช้เทคโนโลยีการทำเหมืองข้อมูล กับธุรกิจเครื่องจำหน่ายสินค้าอัตโนมัติ

นายชยุต ตั้งสกุลวัฒนา 57010265

นางสาวพรฉาย อภิชาติพันธุ์ 57010835

ดร.วัชระ ฉัตรวิริยะ อาจารย์ที่ปรึกษา

ปีการศึกษา 2560

บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์เพื่อสร้างโมเดลคาดการณ์ความเสี่ยงในการลงทุนติดตั้งเครื่องจำหน่ายสินค้าอัตโนมัติ ณ ตำแหน่งนั้น ๆ กลุ่มตัวอย่างเป็นข้อมูลเครื่องจำหน่ายสินค้าอัตโนมัติประเภทตู้เติมเงิน 106 ตู้ 6 เดือน โดยการนำข้อมูลที่ได้รับมาวิเคราะห์หาปัจจัยที่ส่งผลต่อโมเดล และเก็บข้อมูลระยะห่างจากตำแหน่งสำคัญโดยใช้ Google Map API คัดเลือกข้อมูล ใช้วิธี Principal component analysis (PCA) ประกอบกับสร้างโมเดลโดยการทำ Classification โดยใช้ Multilayer Neural Network นอกจากนี้ยังทดสอบประสิทธิภาพด้วย K-Fold Cross Validation ดูผลจากค่า Accuracy เพื่อให้ได้โมเดลที่สามารถคาดการณ์ได้เหมาะสมที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Data Mining for Automate Vending Machine Business

Mr. Chayut Tangsagulwattana 57010265

Ms. Pornchay Apichardpanth 57010835

Dr. Watchara Chatwiriya Advisor

Academic Year 2017

ABSTRACT

The goal of this research is to build model to predict an investment of install a vending machine at that location will profit or won't we got a sample of top-up vending machine around 106 cabinet for 6 month by analysis a feature that affect to model and nearby location data which we collect from Google Map API and we perform feature selection by using Principal Component Analysis (PCA) and building classification model using Multilayer Neural Network. We test model using K-Fold Cross Validation to make a model that most suitable for predict the result.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้จะไม่สามารถสำเร็จลุล่วงไปได้ หากมิได้รับความกรุณาจากอาจารย์ดร. วัชระ ฉัตรวิริยะอาจารย์ที่ปรึกษาโครงงาน ซึ่งเป็นผู้ให้การสนับสนุนในด้านการติดต่อผู้ประกอบการ การเดินทางและให้คำปรึกษาเกี่ยวกับการทำงานเป็นอย่างดี ตลอดจนให้คำแนะนำในการแก้ไขข้อบกพร่องต่าง ๆ มาโดยตลอด คณะผู้จัดทำรู้สึกซาบซึ้งเป็นอย่างมากในความอนุเคราะห์จากอาจารย์ที่คอยให้การสนับสนุนในการทำปริญญานิพนธ์นี้เสมอมา นอกจากนี้ขอขอบคุณผู้ประกอบการธุรกิจเครื่องจำหน่ายสินค้าอัตโนมัติที่ให้ข้อมูลมาใช้งาน ให้คำปรึกษา และสอนให้เข้าใจตัวธุรกิจ รวมถึงห้องวิจัย ESL (Embedded System Laboratory) ของภาควิชาวิศวกรรมคอมพิวเตอร์ ที่เอื้อเฟื้อสถานที่ ให้เป็นสถานที่ทำงาน ศึกษาหาความรู้ประกอบการทำโครงงานในครั้งนี้ ขอบคุณครอบครัว และเพื่อนๆพี่ๆ ที่คอยให้คำปรึกษา สนับสนุนในเรื่องต่าง ๆ ซึ่งทำให้คณะผู้จัดทำมีกำลังใจในการทำงานได้อยู่เสมอ

คณะผู้จัดทำขอกราบขอบพระคุณเป็นอย่างสูง ในความกรุณาของทุกท่านที่ได้กล่าวมานี้ และหวังเป็นอย่างยิ่งว่า ปริญญานิพนธ์ฉบับนี้จะเป็นประโยชน์ต่อทุกท่านที่ต้องการศึกษาทางด้านนี้ และสามารถเป็นคำแนะนำ หรือเป็นแนวทางให้แก่ผู้ที่สนใจ

ชยุต ตั้งสกุลวัฒนา

พรฉาย อภิชาติพันธุ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IIII

สารบัญตาราง VI

สารบัญรูป VII

บทที่ 1 บทนำ 1

1.1 ความสำคัญและที่มาของโครงงาน 1

1.2 วัตถุประสงค์ของโครงงาน 1

1.3 ขอบเขตของโครงงาน 1

1.4 วิธีการดำเนินงาน 2

1.5 ประโยชน์ที่คาดว่าจะได้รับ 2

1.6 ส่วนประกอบของปริญญานิพนธ์ 2

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง 3

2.1 ธุรกิจเครื่องจำหน่ายสินค้าอัตโนมัติ 3

2.2 งานวิจัยที่เกี่ยวข้อง 3

2.3 CRISP-DM Methodology 4

2.4 Classification 5

2.5 โครงข่ายประสาทเทียม (Multilayer Neural Network) 6

2.6 Normalization by Standard Deviation 9

2.7 Google map API 9

2.8 K-fold Cross-Validation 10

2.9 Principal Component Analysis (PCA) 10

2.10 Confusion Matrix 11

2.11 Feature Selection 12

2.12 โปรแกรมและภาษาที่ใช้พัฒนา 13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3 การออกแบบและพัฒนา....14
3.1 Business Understanding....14
3.2 Data Understanding....15
3.3 Data Preparation....16
3.4 Modeling....16
3.5 Evaluation....17
3.6 โครงสร้างของการจัดการข้อมูล....17
3.7 การออกแบบโปรแกรมสำหรับหาระยะทางโดยใช้ Google Map API....18
3.8 การออกแบบการทำงานสำหรับ Modeling....19
3.9 เครื่องมือที่ใช้พัฒนา....20

บทที่ 4 การทดลองและสรุปผลการทดลอง....21
4.1 ทดลองเขียนเว็ปและใช้ Google Map API เพื่อนำมาช่วยในการเก็บข้อมูล....21
4.2 ทดลองใช้ Feature Selection และ Dimension Reduction....25
4.3 การปรับค่าโมเดล Neural Network....26
4.4 สรุปผลการทดลองทั้งหมด....28

บทที่ 5 บทสรุปและข้อเสนอแนะ....30
5.1 บทสรุปและบทวิจารณ์....30
5.2 ปัญหาและอุปสรรค....30
5.3 แนวทางในการแก้ไขปัญหา....30
5.4 แนวทางในการพัฒนาต่อ....30

บรรณานุกรม....31

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตาราง หน้า

4.1 ผลการทดลองด้วย Feature ทั้ง 21 ตัว 25

4.2 ผลการทดลองใช้ Feature Selection แบบ Correlation 25

4.3 ผลการทดลองใช้ Feature Selection แบบ ReliefF 26

4.4 ผลการทดลองใช้ Dimension Reduction แบบ PCA 26

4.5 ผลการทดลองเปลี่ยน Transfer Function 27

4.6 ผลการทดลองเปลี่ยน Train Function 28

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูป หน้า

2.1 Multilayer Neural Network 6
2.2 Tan-Sigmoid Transfer Function 7
2.3 Log-Sigmoid Transfer Function 7
2.4 Linear Transfer Function 8
2.5 K-fold Cross Validation 10
2.6 ตัวอย่าง Confusion Matrix 11
2.7 Releif 13
3.1 โครงสร้างของระบบ 17
3.2 Flow Chart โปรแกรมสำหรับหาระยะทางโดยใช้ Google Map API 18
3.3 Modeling 19
4.1 หน้า Index ของ website 24
4.2 หน้าแสดงผลสถานที่ใกล้เคียง 24
4.3 ผลการทดลองเปลี่ยน Hidden Layer 27
4.4 Multilayer Neural Network 28
4.5 Confusion Matrix แสดงรวมทั้ง 5 Fold Cross Validation 29

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของโครงงาน

ในปัจจุบันนั้นการใช้งานสินค้าและบริการอัตโนมัตินั้นได้เข้ามามีส่วนกับการใช้ชีวิตของคนเรามากขึ้นเพราะว่าการซื้อสินค้าและบริการอัตโนมัตินั้นมีทั้งความความสะดวกสบายความรวดเร็ว โดยบริการของเครื่องจำหน่ายสินค้าอัตโนมัติมีทั้งการซื้ออาหารหรือเครื่องดื่มจากตู้จำหน่ายสินค้าอัตโนมัติ หรือแม้กระทั่งตู้เติมเงิน-ชำระบิล นั้นก็เป็นเครื่องจำหน่ายสินค้าอัตโนมัติ แต่เนื่องจากการจะติดตั้งเครื่องจำหน่ายสินค้าอัตโนมัตินั้นต้องใช้ทั้งต้นทุนที่สูงและยังมีค่าการบำรุงรักษาอีก จึงทำให้การเลือกสถานที่ที่จะติดตั้งนั้นจะต้องเป็นสถานที่ที่มีความเหมาะสม โดยการจะหาสถานที่แบบนั้นจะต้องวิเคราะห์ปัจจัยต่าง ๆที่คาดว่าจะเกี่ยวข้อง เช่น สถานที่ต่าง ๆในบริเวณใกล้เคียง ที่จะต้องนำมาวิเคราะห์ก่อนที่จะทำการติดตั้งเครื่องจำหน่ายสินค้าและบริการอัตโนมัติ

กลุ่มโครงงานของพวกเรานั้นได้เห็นถึงปัญหานี้จึงคิดโครงงาน การประยุกต์ใช้เทคโนโลยีการทำเหมืองข้อมูลกับธุรกิจเครื่องจำหน่ายสินค้าอัตโนมัติขึ้นมาเพื่อศึกษาและสร้างโมเดลที่ใช้ในการหาว่าการตั้งเครื่องจำหน่ายสินค้าอัตโนมัติตามสถานที่ต่าง ๆนั้นว่าคุ้มค่าต่อการลงทุนหรือไม่

1.2 วัตถุประสงค์ของโครงงาน

1) เพื่อทำนายการได้กำไร ขาดทุน หรือเท่าทุนของเครื่องจำหน่ายสินค้าในตำแหน่งที่เลือก
2) เพื่อหาว่าปัจจัยใดบ้างที่ส่งผลต่อการได้กำไร ขาดทุน หรือเท่าทุน

1.3 ขอบเขตของโครงงาน

ข้อมูลที่นำมาใช้งาน

1) เครื่องจำหน่ายสินค้าอัตโนมัติในที่นี้ หมายถึง ตู้เติมเงินโทรศัพท์อัตโนมัติ
2) ได้รับข้อมูลมาจากผู้ประกอบการรายหนึ่ง
3) มีจำนวนตู้ 106 ตู้ ที่ติดตั้งอยู่ใน 3 จังหวัด คือ เชียงใหม่,สุโขทัย และพิษณุโลก ใช้ข้อมูลในปี พ.ศ.2560
4) ข้อมูลที่นำมาใช้หลักๆ จะเป็นตำแหน่งที่ตั้งข้อตู้ และจำนวนผู้ใช้งาน โดยตำแหน่งที่ตั้งของตู้สามารถหารายละเอียดได้จาก Google Map

ใช้เทคโนโลยีการทำเหมืองข้อมูล (Data Mining) ในการเขียนวิจัยที่มีจุดมุ่งหมายในการสร้างโมเดล โดยโมเดลนั้นจะสามารถใช้วิเคราะห์ได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) ทราบว่าปัจจัยใด ที่ส่งผลให้ได้กำไร ขาดทุน หรือเท่าทุน
2) สามารถคาดการณ์ว่า ณ ตำแหน่งที่ตั้งนั้น จะได้กำไรขาดทุน หรือเท่าทุน

1.4 วิธีการดำเนินงาน

1) ทำความเข้าใจตัวธุรกิจ
2) ทำความเข้าใจข้อมูล และว่าข้อมูลใดที่ต้องนำมาใช้
3) คัดเลือกข้อมูลที่เกี่ยวข้อง และนำมาดัดแปลงบางส่วน เพื่อง่ายต่อการทำขั้นตอนถัดไป
4) สร้างโมเดล
5) ประเมินผลเพื่อทดสอบประสิทธิภาพของโมเดล

1.5 ประโยชน์ที่คาดว่าจะได้รับ

สามารถวิเคราะห์ได้ว่าในตำแหน่งที่ติดตั้งเครื่องจำหน่ายสินค้าอัตโนมัติจะคุ้มค่าต่อการลงทุนหรือไม่อีกทั้งยังทราบว่าปัจจัยเกี่ยวกับตำแหน่งที่ตั้งใดบ้างที่ทำให้ได้รับจำนวนเงินมาก

นอกจากนี้ยังสามารถนำโมเดลที่ได้รับไปประยุกต์ใช้งานกับเครื่องจำหน่ายสินค้าประเภทอื่นอีกด้วย

1.6 ส่วนประกอบของปริญญานิพนธ์

ปริญญานิพนธ์ฉบับนี้ได้แบ่งเนื้อหาออกเป็น 5 บทด้วยกันคือ

บทที่ 1 บทนำ กล่าวถึงความสำคัญและที่มาของโครงงาน วัตถุประสงค์ของโครงงาน ขอบเขตของโครงงาน วิธีการดำเนินการ ประโยชน์ที่คาดว่าจะได้รับ และส่วนประกอบของปริญญานิพนธ์

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง กล่าวถึงในเรื่องของตู้จำหน่ายสินค้าอัตโนมัติ งานวิจัยที่เกี่ยวข้อง และทฤษฎีที่นำมาใช้งาน

บทที่ 3 การออกแบบและพัฒนา กล่าวถึงปัจจัยสำคัญที่ใช้ในการตัดสินใจ การวางแผนการทำงานและเครื่องมือที่ใช้

บทที่ 4 การทดลองและสรุปผลการทดลอง กล่าวถึงการทดลองต่าง ๆ ที่ใช้ในการทำงาน การเปรียบเทียบ การวัดประสิทธิภาพของโมเดลและสรุปผลการทดลอง

บทที่ 5 บทสรุปและข้อเสนอแนะ กล่าวถึงบทสรุปของโครงงาน วิจารณ์สิ่งที่ได้รับจากโครงงาน ข้อจำกัด รวมถึงปัญหาอุปสรรคต่าง ๆ ของโครงงาน และข้อเสนอแนะสำหรับเป็นแนวทางในการพัฒนาต่อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

2.1 ธุรกิจเครื่องจำหน่ายสินค้าอัตโนมัติ

เครื่องจำหน่ายสินค้าอัตโนมัตินั้นคือเครื่องที่จะให้สินค้าและบริการหลังจากที่เราได้ใส่เงินหรือบัตรเครดิตเข้าไปโดยเครื่องจำหน่ายสินค้าอัตโนมัตินั้นจะมีหลายแบบทั้งเครื่องจำหน่ายเครื่องดื่มเครื่องจำหน่ายอาหาร เครื่องจำหน่ายบุหรี่ รวมไปถึงเครื่องจำหน่ายสินค้าอัตโนมัติ

2.2 งานวิจัยที่เกี่ยวข้อง

Jia-Lang Seng และ T.C. Chen (2010) ได้วิจัยเรื่อง วิธีการวิเคราะห์เพื่อเลือกการทำเหมืองข้อมูลเพื่อการตัดสินใจทางธุรกิจ (An analytic approach to select data mining for business decision) ที่มีลักษณะเป็นงานวิจัยที่เขียนเกี่ยวกับรายละเอียดทุกขั้นตอนในการทำธุรกิจและวิเคราะห์ว่า ในกระบวนการไหนควรใช้เทคนิคใดมาใช้งานกับงานทางด้านธุรกิจ โดยมีจุดเด่นคือ สามารถนำไปใช้ได้กับธุรกิจได้จริง และทำให้ผู้นำไปใช้ไม่สับสนเกี่ยวกับการเลือกใช้เทคนิคต่าง ๆ ทำให้รู้ว่าแต่ละขั้นตอนของธุรกิจมีความละเอียดอ่อนอย่างมาก เมื่อได้อ่านงานวิจัยนี้แล้วนอกจากจะได้ความรู้ทางด้าน Data Analysis แล้ว ยังได้ความรู้ทางด้านธุรกิจอย่างครบถ้วน หากแต่ยังคงมีข้อจำกัดคือ ยังไม่ครอบคลุมทุกประเด็นของช่วงธุรกิจ และยังขาดการวิเคราะห์เกี่ยวกับตำแหน่งที่ตั้งตลาด

Fatemeh Rahimi-Ajdadi และ Yousef Abbaspour-Gilandeh (2011) ได้วิจัยเรื่อง Artificial Neural Network and stepwise multiple range regression methods for prediction of tractor fuel consumption ที่มีลักษณะ เป็นงานวิจัยเขียนเกี่ยวกับการทำนายการใช้เชื้อเพลิงของรถแทรกเตอร์ เพื่อรักษาอุปกรณ์ในฟาร์ม โดยใช้วิธี Artificial Neural Network และ Stepwise Regression แล้วนำมาเปรียบเทียบความแม่นยำของการคาดการณ์ พบว่าวิธี Neural Network มีความแม่นยำที่มากกว่า ข้อดีคือ มีกระบวนการทำในส่วนของเทคนิคการทำเหมืองข้อมูลได้ดี มีการแสดงขั้นตอนการทำ Artificial Neural Network และ Stepwise Regression ดี ตารางแสดงผลต่าง ๆ อ่านเข้าใจง่าย

Feng-Cheng Lin, Hsin-Wen Yu, Chih-Hao Hsu และ Tzu-Chun Weng b (2011) ได้วิจัยเรื่อง Recommendation system for localized products in vending machines ที่มีลักษณะ เป็นงานวิจัยสร้างโปรแกรมประยุกต์ (Application) จัดหาบริการเสริม เพื่อเพิ่มรายได้ให้กับเครื่องจำหน่ายสินค้าอัตโนมัติตามคำแนะนำของเทคนิคทาง การทำเหมืองข้อมูล หลายอย่าง คือเทคนิค K-means Clustering, Decision Tree และ Bayesian Network โดยโปรแกรมประยุกต์มีอินพุต (Input) เป็นรายละเอียดของสินค้า และเอาต์พุต (Output) เป็น สินค้าที่แนะนำให้วางไว้ในตู้ มีสิ่งที่น่าสนใจคือ เป็นการทำการทำนายข้อมูลที่ครบวงจร สามารถใช้ได้จริง เพราะมีการพัฒนาส่วนของ User Interface และยังมีการติดตามผลกำไร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก่อน-หลังการใช้โปรแกรมประยุกต์ที่พัฒนา แต่ในประเทศที่เขียนงานวิจัยนี้เครื่องจำหน่ายสินค้าอัตโนมัติมีตู้ขนาดใหญ่มากและวางติดกัน ทำให้สามารถมีสินค้าภายในตู้หลายชนิด หากนำมาใช้ในประเทศไทยอาจได้ผลที่ต่างกัน

จากงานวิจัยข้างต้น แสดงให้เห็นถึงความสำคัญของธุรกิจเครื่องจำหน่ายสินค้าอัตโนมัติที่มีอยู่ในทั่วโลก ได้รับการแนะนำกระบวนการในทางธุรกิจต่าง ๆ ว่าควรใช้เทคนิคทางการทำเหมืองข้อมูลแบบใดเหมาะสมกับส่วนไหนในทางธุรกิจ และถ้าหากต้องการทำงานในด้านนี้สามารถนำไปพัฒนางานวิจัยนี้เพื่อนำไปใช้จริงโดยที่ให้คนทั่วไป สามารถใช้งานได้อย่างง่ายดาย ในงานวิจัยนี้ จะต้องพัฒนาให้สามารถทำนายตำแหน่งที่ตั้งของเครื่องจำหน่ายสินค้าอัตโนมัติ ซึ่งเป็นตำแหน่งที่มีผู้ใช้งานมาก

2.3 CRISP-DM Methodology

2.3.1 Business Understanding

1) กำหนดเป้าหมายของธุรกิจที่เราจะทำการวิเคราะห์ เป้าหมายแรกก่อนการวิเคราะห์จากมุมมองของการทำธุรกิจว่าลูกค้านั้นต้องการอะไรส่วนมากแล้วลูกค้าจะมีการแข่งขันกับคนอื่นสูง เป้าหมายของการวิเคราะห์นั้นคือ การที่เราสามารถหาข้อมูลที่มีอิทธิพลต่อผลลัพธ์
2) ประเมินสถานการณ์ สร้างสมมุติฐานและส่วนอื่นที่จะนำมาพิจารณาหาเป้าหมายของการวิเคราะห์ข้อมูล
3) ตั้งเป้าหมายของการทำเหมืองข้อมูล
4) ทำแผนงาน ระบุแผนที่จะทำให้เราสามารถบรรลุเป้าหมายทั้งการทำเหมืองข้อมูล และธุรกิจ

2.3.2 Data Understanding

1) รวบรวมข้อมูลจากแหล่งข้อมูลที่เชื่อถือได้
2) ระบุประเภทของข้อมูล เป็นขั้นตอนตรวจสอบข้อมูลทั้งหมดและรายงานผล
3) สำรวจข้อมูล ขั้นตอนนี้นี้ช่วยตอบปัญหาของการทำเหมืองข้อมูลในเรื่องของการ Querying, Visualization และการทำแบบรายงาน การวิเคราะห์เหล่านี้ช่วยระบุวัตถุประสงค์ของการทำเหมืองข้อมูล แต่ถึงอย่างนั้นก็ยังมีส่วนช่วยในเรื่องคุณภาพของงาน
4) ตรวจสอบคุณภาพของข้อมูล เป็นขั้นตอนการตรวจสอบข้อมูลว่าข้อมูลนี้พร้อมจะนำมาใช้งานหรือไม่และมีข้อผิดพลาด หรือ มีข้อมูลขาดหายหรือไม่

2.3.3 Data Preparation

1) เลือกข้อมูล (Select Data) ตัดสินใจเลือกข้อมูลที่จะใช้ในการวิเคราะห์ กำหนดเกณฑ์มีความเกี่ยวข้องกับเป้าหมายการทำเหมืองข้อมูลให้มีคุณภาพ และข้อจำกัดด้านเทคนิค เช่น ข้อจำกัดของปริมาณข้อมูลหรือชนิดข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) ความสะอาดข้อมูล (Clean Data) เช่น พิจารณาวิธีจัดการกับข้อมูล Noise พิจารณาความถูกต้อง, การลบ หรือเพิกเฉยต่อ Noise
3) Construct Data เป็นผลลัพธ์ที่ได้รับมาจากคุณลักษณะ (Attribute) ต่าง ๆ การบันทึกข้อมูลใหม่ การเปลี่ยนค่าจากคุณลักษณะที่มีอยู่
4) Integrate Data วิธีการสำหรับการรวมข้อมูลจากหลาย ๆ ตารางหรือแหล่งข้อมูลอื่น ๆ เพื่อสร้างค่าใหม่
5) Format Data การดัดแปลงที่ไม่ทำให้ข้อมูลเปลี่ยนความหมาย

2.3.4 Modeling

1) การเลือกเทคนิคการสร้างแบบจำลอง หากใช้เทคนิคหลาย ๆ เทคนิคให้ปฏิบัติงานแยกกันสำหรับแต่ละเทคนิค
2) สร้างการออกแบบการทดสอบ ก่อนที่จะสร้างโมเดลจำเป็นต้องกำหนดขั้นตอนเพื่อทดสอบคุณภาพของรูปแบบและความถูกต้อง การออกแบบทดสอบระบุว่าแต่ละชุดข้อมูล (Datasets) ควรแยกออกจากกันใน Training Set และ Test Set แบบจำลองหรือโมเดลนั้นสร้างขึ้นจากชุดฝึกอบรม และทดสอบคุณภาพโดย ชุดทดสอบ
3) สร้างโมเดล รวมถึงปรับเปลี่ยนมักมีพารามิเตอร์จำนวนมากที่สามารถปรับเปลี่ยนได้ ระบุพารามิเตอร์และค่าที่เลือกพร้อมด้วยเหตุผลสำหรับการเลือก
4) ประเมินแบบจำลอง ควรประเมินแบบจำลองนี้เพื่อให้มั่นใจว่าตรงตามเกณฑ์ความสำเร็จในการทำเหมืองข้อมูลหรือไม่

2.3.5 การประเมินผล

ขั้นตอนนี้จะประเมินโมเดลว่าคุณสมบัติตรงตามวัตถุประสงค์ทางธุรกิจ และพยายามที่จะตรวจสอบว่ามีเหตุผลทางธุรกิจบางอย่างที่ทำให้บางส่วนของโมเดลขาดหายไปหรือไม่ อีกทางเลือกหนึ่งคือการทดสอบรูปแบบการใช้งานทดสอบในแอพพลิเคชั่นจริง ถ้าหากมีข้อ จำกัด ด้านเวลาและงบประมาณ

2.4 Classification

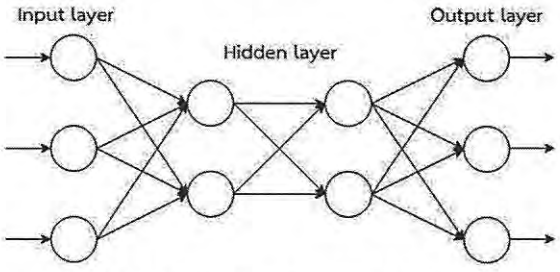
เป็นหนึ่งในเทคนิคที่ใช้กันอย่างแพร่หลายในการทำ Machine Learning คือ การทำนายประเภท (Category) หรือ Class จากอินพุตเป็นการทำนายโดยให้เอาต์พุตออกมาเป็นประเภทต่าง ๆ

2.4.1 การนำมาใช้งาน

ได้นำมาสร้างโมเดลโดยใช้เทคนิค Classification โดยแยกประเภทของ เอาต์พุต เป็น 3 ประเภทคือ ได้กำไร,เท่าทุน และขาดทุน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5 โครงข่ายประสาทเทียม (Multilayer Neural Network)



รูป 2.1 Multilayer Neural Network

จากรูป 2.1

Input Layer คือ ชั้นที่รับอินพุตจากภายนอก

Hidden Layer คือ ชั้นที่ตรวจหารูปแบบของข้อมูลอินพุต โดยการปรับค่าน้ำหนัก (Weight) ของโครงข่ายประสาทเทียม อยู่ตรงกลางระหว่าง Input Layer กับ Output Layer

Output Layer คือ ชั้นที่รับข้อมูลจาก Hidden Layer และแปลงข้อมูลให้เป็น Output Pattern

วิธีการ Training Model

ขั้นตอนที่ 1 ใส่ค่า Weight และ Bias โดยค่าสุ่ม

ขั้นตอนที่ 2 Activation (Feed Forward Computation) โดยการคำนวณ หาค่าเอาต์พุตจาก Hidden Layer ตามสมการดังนี้

$$y_i(p) = sigmoid\left[\sum_{i=1}^{n}\left(x_{i(p)} * w_{ij(p)}\right) + \theta_j\right] \qquad (2.1)$$

โดย จากในตัวอย่างสมการด้านบน ใช้ Train Function หรือ Activation Function เป็น Sigmoid

$x_{i(p)}$ = อินพุตของโมเดล

$w_{ij(p)}$ = ค่าน้ำหนักระหว่าง Node i ถึง Node j

$y_i(p)$ = เอาต์พุตที่ผ่าน Node j

θ_j = ค่า Bias ของ Node j

n = จำนวนอินพุต

ขั้นตอนที่ 3 Weight Training (Backpropagation) คำนวณค่า Error จาก ค่าที่คำนวณได้ และค่าจริงที่ควรเป็น

ขั้นตอนที่ 4 ทำขั้นตอนที่กล่าวไว้ข้างต้นวนไป จนกว่าจะได้ค่า Error ที่พอใจ

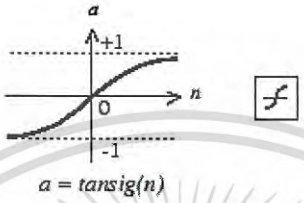
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.1 การนำมาใช้งาน

นำมาทำ Classification โมเดลโดยใช้วิธี Multilayer Neural Network

2.5.2 ฟังก์ชันที่เกี่ยวข้อง

1) Transfer Function คำหน้าที่คำนวณอินพุต ณ Hidden Layer มีหลายวิธี ดังนี้ Hyperbolic Tangent Sigmoid Transfer Function (tansig)



Tan-Sigmoid Transfer Function

รูป 2.2 Tan-Sigmoid Transfer Function

คำนวณได้ดังนี้

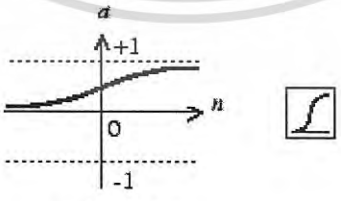
$$tansig(n) = \frac{2}{(1 + exp(-2 * n)) - 1} \quad (2.2)$$

โดย n = อินพุตของ Function และ เอาต์พุตที่ออกมาจาก Transfer Function จะมีค่าอยู่ระหว่าง -1 ถึง +1 สามารถเรียกใช้งานใน MATLAB ได้ดังนี้

net.layers{i}.transferFcn ='tansig ';

โดย i เป็นตำแหน่ง Hidden Layer

Log-Sigmoid Transfer Function (logsig)



Log-Sigmoid Transfer Function

รูป 2.3 Log-Sigmoid Transfer Function

คำนวณได้ดังนี้

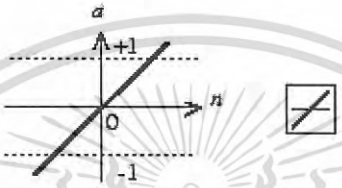
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$logsig(n) = \frac{1}{(1 + exp(-n))} \quad (2.3)$$

โดย n = อินพุตของ Function และ เอาต์พุตที่ออกมาจาก Transfer Function จะมีค่าอยู่ระหว่าง 0 ถึง +1 สามารถเรียกใช้งานใน MATLAB ได้ดังนี้

net.layers{i}.transferFcn = 'logsig'; โดย i เป็นตำแหน่ง Hidden Layer

Linear Transfer Function (purelin)



รูป 2.4 Linear Transfer Function

คำนวณได้ดังนี้

$$purelin(n) = n \quad (2.4)$$

โดย n = อินพุตของ Function และ เอาต์พุตที่ออกมาจาก Transfer Function จะเท่ากับค่าอินพุต สามารถเรียกใช้งานใน MATLAB ได้ดังนี้

net.layers{i}.transferFcn = 'purelin'; โดย i เป็นตำแหน่ง Hidden Layer

2) Train Function หน้าที่เป็นตัวที่อัพเดทค่าน้ำหนัก และ Bias ด้วยวิธีต่างๆ

Scaled Conjugate Gradient Backpropagation (trainscg) ใช้สำหรับงานที่ต้องการใช้ทรัพยากรน้อย สามารถตั้งค่าใน MATLAB ได้ดังนี้

net.trainFcn = 'trainscg'

Levenberg-Marquardt Backpropagation (trainlm) เป็นวิธีที่เร็วที่สุด แต่ใช้ทรัพยากร มากกว่าวิธีอื่น สามารถตั้งค่าใน MATLAB ได้ดังนี้

net.trainFcn = 'trainlm'

Bayesian Regularization Backpropagation (trainbr) ใช้เวลาประมวลผลงานที่สุด เหมาะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับงานที่มีความซับซ้อนสามารถตั้งค่าใน MATLAB ได้ดังนี้

net.trainFcn = 'trainbr'

2.6 Normalization by Standard Deviation

เป็นการ Normalize ข้อมูลโดยใช้ส่วนเบี่ยงเบนมาตรฐาน สามารถคำนวณได้ดังนี้

โดย x = ค่าที่ต้องการทำ Normalize

$mean$ = ค่าเฉลี่ยเลขคณิตของข้อมูลทั้งหมด

$Deviation$ = ค่าส่วนเบี่ยงเบนมาตรฐานของข้อมูลทั้งหมด

y = ค่าที่ผ่านการทำ Normalize

$$y = \frac{x - mean}{Deviation} \tag{2.5}$$

2.6.1 การนำมาใช้งาน

เพื่อใช้ในการเปรียบเทียบข้อมูลแต่ละตัวได้ง่ายมากขึ้น และการทำใช้วิธีโครงข่ายประสาทเทียม จำเป็นต้อง Normalize ข้อมูล เพื่อให้มีช่วงของข้อมูลที่อยู่ในช่วงเดียวกัน

2.7 Google map API

เป็นตัว API ของ Google สำหรับช่วยในการดึงข้อมูลจากแผนที่ใน Google Map

2.7.1 การนำมาใช้งาน

นำมาใช้สร้างเว็บไซต์ เพื่อดึงข้อมูลระยะทางห่างจากสถานที่สำคัญ ที่ใช้ในอินพุตของโมเดล

2.7.2 ฟังก์ชันที่เกี่ยวข้อง

1) Nearby Search Requests ช่วยในการค้นหาสถานที่ใกล้เคียงกับสถานที่ที่ระบุ โดย เอาต์พุตที่ได้รับออกมา จะเป็นประเภท JavaScript Object Notation (JSON)

   Parameters ที่ต้องการ

   Key คือ API Key ใช้ระบุตัวตนของโปรแกรมประยุกต์ที่เราใช้ เพื่อจัดการโควตา

   Location คือ ลองจิจูด (Longitude) และละติจูด (Latitude) ที่ต้องการหาตำแหน่งใกล้เคียง

   Radius คือ รัศมี หน่วยเมตร ที่ต้องการให้ได้สถานที่ใกล้เคียงในรัศมีนั้น

2) Distance Matrix API หาระยะทางและเวลาในการเดินทางระหว่างสองสถานที่

   Parameters ที่ต้องการ

   Key คือ API Key ใช้ระบุตัวตนของโปรแกรมประยุกต์ที่เราใช้ เพื่อจัดการโควตา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

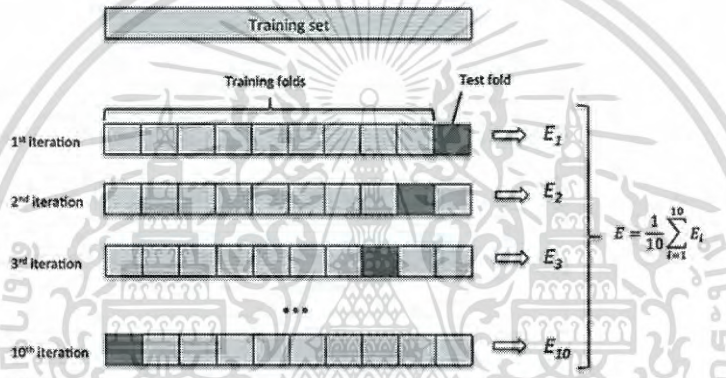
Origins คือ ลองจิจูด และละติจูด ของจุดเริ่มต้น

Destinations คือ ลองจิจูด และละติจูด ของจุดหมาย

2.8 K-fold Cross-Validation

เป็นการแบ่งส่วนของข้อมูลเพื่อนำไป ทดสอบประสิทธิภาพโมเดลโดยมีตัวอย่างดังนี้ การทำ 10 Fold Cross Validation คือการแบ่งข้อมูลเป็น 10 ส่วน

ในตัวอย่างได้นำข้อมูลที่มีทั้งหมด มาทำการแบ่งเป็น Training Set 9 ส่วน และ Test set 10 ส่วน ทำการ Train โมเดล 10 ครั้ง ทุกครั้งที่ Train โมเดล จะเปลี่ยนส่วนที่เป็น Training Set และ Test set ทุกครั้ง ในแต่ละครั้ง จะเก็บค่า Error แต่ละครั้งมาทำการคำนวณหาค่า Error เฉลี่ย ดังรูปด้านล่าง



รูป 2.5 K-fold Cross Validation

2.8.1 การนำมาใช้งาน

นำมาใช้ในการประเมินผลโมเดล เพื่อให้ได้ใช้ข้อมูลทุกตัวให้เป็น ทั้ง ข้อมูลที่ใช้ Train และ Test

2.8.2 ฟังก์ชันที่เกี่ยวข้อง

Cvpartition ใช้ในการแบ่งส่วนของข้อมูลสำหรับทำ Cross Validation ตัวอย่างการใช้งาน c = cvpartition(n,'KFold',k) โดย N = จำนวนแถวของข้อมูล, k = จำนวนที่ต้องการ Fold

2.9 Principal Component Analysis (PCA)

เป็นเทคนิคการทำ Feature Extraction เพื่อสร้าง Feature ใหม่ เป็นวิธีการลดมิติของข้อมูลให้น้อยลง โดยมิติที่สร้างมาใหม่จะเรียกว่า Principal Components (PC) การลดมิติโดยใช้ PCA เพื่อช่วยลดความซ้ำซ้อนของข้อมูล และทำให้ค่าความแม่นยำในการทำนายแบบ Classification สูงขึ้น

2.9.1 การนำมาใช้งาน

เพื่อลดมิติของข้อมูล ทำให้โมเดลมีประสิทธิภาพมากยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.9.2 ฟังก์ชันที่เกี่ยวข้อง

ตัวอย่างการใช้งาน Data = pca(input,outputsize);

โดย Input = ข้อมูลอินพุตประเภท Table, Outputsize = จำนวนมิติที่ต้องการ

2.10 Confusion Matrix

เป็นตารางที่ใช้ในการอธิบายประสิทธิภาพของโมเดล Classification

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

รูป 2.6 ตัวอย่าง Confusion Matrix

จากตัวอย่าง หมายถึง มีข้อมูลทั้งหมด 165 ตัว ถูกทำนายว่า YES 110 ตัว ถูกทำนายว่า NO 55 ตัว แต่ค่าจริง มีจำนวน 105 ตัว NO 60 ตัว

มีคำศัพท์ที่สำคัญดังนี้

True Positives (TP): สิ่งที่ทำนายว่า YES และค่าจริงคือ YES

True Negatives (TN): สิ่งที่ทำนายว่า NO และค่าจริงคือ NO

False Positives (FP): สิ่งที่ทำนายว่า YES และค่าจริงคือ NO

False Negatives (FN) สิ่งที่ทำนายว่า NO และค่าจริงคือ YES

Accuracy คือ ค่าที่บอกว่าโมเดลมีความแม่นยำมากน้อยแค่ไหน ถ้าค่ามาก คือมีประสิทธิภาพดี คำนวณได้ดังนี้

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{2.6}$$

Recall (True Positive Rate) คือ ค่าที่บอกว่าโมเดลทำนายได้ค่าจริง เป็นอัตราส่วนเท่าไหร่ของค่าจริงทั้งหมด คำนวณได้ดังนี้

$$Recall = \frac{TP}{(TP + FN)} \tag{2.7}$$

Precision คือ ค่าที่บอกว่าโมเดลทำนายได้ค่าจริง มีความถูกต้องเท่าไหร่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$Precision = \frac{TP}{(TP + FP)} \quad (2.8)$$

2.10.1 การนำมาใช้งาน

ได้นำ Confusion matrix มาแสดงผล นำค่า Accuracy มาประเมินผลโมเดล และนำ Precision, Recall มาวิเคราะห์

2.10.2 ฟังก์ชันที่เกี่ยวข้อง

plotconfusion(t,y) เป็น Function สำหรับสร้าง Confusion Matrix โดย t คือ ค่าจริง, y คือผลการทำนาย

2.11 Feature Selection

เป็นวิธีการเลือก Feature ที่นำมาใช้สร้างโมเดล โดยในงานวิจัยนี้ ได้เลือกใช้สองวิธีหลัก ๆ ดังนี้

วิธีแรก Correlation เป็นการหาค่าความสัมพันธ์ระหว่างตัวแปร โดยใช้ Pearson's Correlation Coefficient คำนวณได้ดังนี้

$$r = \frac{\sum((x - \bar{x})\,(y - \bar{y}))}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \quad (2.9)$$

โดย x = ค่า x

$\bar{x}$ = ค่าเฉลี่ยของ x ทั้งหมด

y = ค่า y

$\bar{y}$ = ค่าเฉลี่ยของ y ทั้งหมด

หากได้ Correlation เป็นบวก หมายถึง ตัวแปรทั้งสองตัวนั้น มีความสัมพันธ์แบบแปรผันตรง

Correlation เป็นลบ หมายถึง ตัวแปรทั้งสองตัวนั้น มีความสัมพันธ์แบบแปรผกผัน

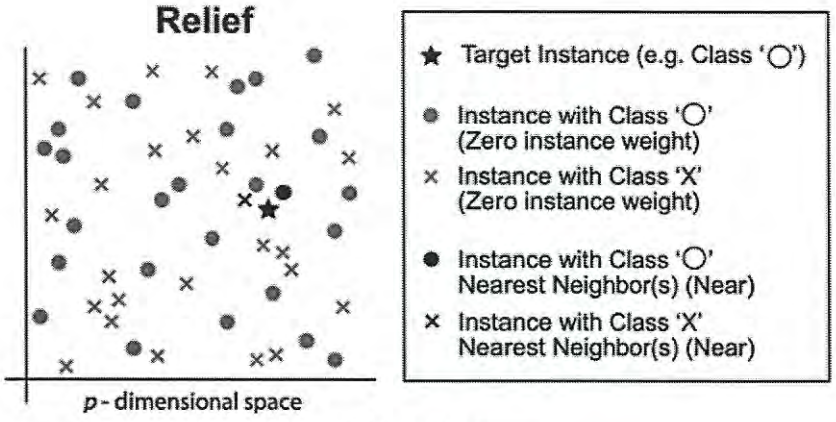
Correlation เป็นศูนย์ หมายถึง ตัวแปรทั้งสองตัวนั้น ไม่มีความสัมพันธ์กัน

วิธีที่สอง Releif เป็นการคำนวณคะแนน Feature แต่ละ Feature และจะนำมาจัดอันดับและเลือกอันที่ดีไปใช้ ซึ่งการให้คะแนน จะขึ้นอยู่กับการระบุความแตกต่างของเพื่อนบ้านที่ใกล้ที่สุด (Nearest Neighbor)

ถ้าค่าความแตกต่างของ Feature ในคู่ที่อยู่ติดกันมี Class เดียวกัน คะแนนจะลดลงเรียกว่า ฮิต (Hit)

ถ้าค่าความแตกต่างของ Feature ในคู่ที่อยู่ติดกันมีคนละ Class คะแนนจะเพิ่มขึ้นเรียกว่า มิส (Miss)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 2.7 Releif

ส่วนอัลกอริทึมของ ReleifF จะหาตัวฮิตที่ใกล้ที่สุดและ ตัวมิสที่ใกล้ที่สุดโดยใช้ Manhattan (L1) norm มากกว่า Euclidean (L2) norm ในช่วงนี้จะเห็นความแตกต่างระหว่าง X กับตัวฮิตที่ใกล้ X ที่สุด และ ตัวมิสที่ใกล้ X ที่สุดเมื่อมีการปรับค่าเวกเตอร์น้ำหนัก

2.11.1 การนำไปใช้งาน

นำไปใช้ในการเลือก Feature เพื่อหาว่า Feature มีความสำคัญ และให้โมเดลมีประสิทธิภาพมากขึ้น โดยเรียกใช้งานโดยใช้ โปรแกรม Weka

2.12 โปรแกรมและภาษาที่ใช้พัฒนา

2.12.1 MATLAB

เป็นซอฟต์แวร์ในการคำนวณและการเขียนโปรแกรม พัฒนาโดย MathWorks ใช้ในการพัฒนาอัลกอริธึม การสร้างแบบจำลองทางคณิตศาสตร์ และการทำซิมูเลชั่นของระบบ การทำงาน มีทั้งแบบเขียนคำสั่งทีละบรรทัด และแบบรวบรวมชุดคำสั่ง

2.12.2 PHP

เป็นภาษาสำหรับเขียนเว็บที่แสดงผลในรูปแบบ HTML ทางฝั่ง Server Side เป็น Open Source สามารถใช้งานได้โดยไม่เสียค่าใช้จ่าย มีรากฐานคำสั่งมาจากภาษาซี นำมาใช้งานในส่วนการพัฒนาเว็บไซต์ดึงข้อมูลจาก Google Map API

2.12.3 Weka

เป็นโปรแกรมที่รวบรวม Machine Learning Algorithm ต่าง ๆไว้ ถูกพัฒนาโดยภาษา Java มีเครื่องมือต่าง ๆ ไม่ว่าจะเป็น Data Pre-Processing, Classification, Regression, Clustering, Association Rules และ Visualization เป็น Open Source ที่อยู่ภายใต้ลิขสิทธิ์ของ GNU General Public License นำมาใช้งานในเรื่องการทำ Feature Selection เพื่อช่วยในการตัดสินใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การออกแบบและพัฒนา

3.1 Business Understanding

ในขั้นตอนนี้จะทำความเข้าใจปัญหาในธุรกิจ เรื่องการวางตำแหน่งตู้เติมเงิน โดยการนัดคุยงานกับผู้ประกอบการตู้เติมเงิน เพื่อทราบถึงการวิเคราะห์ทางด้านการตลาด และนำไปใช้ในการตัดสินใจเลือก Attribute ที่นำไปสร้างโมเดล โดยมีรายละเอียดดังนี้

3.1.1 รายละเอียดของตู้เติมเงิน

มีจำนวน 106 ตู้ ตั้งอยู่บริเวณต่างจังหวัดคือจังหวัดเชียงใหม่ สุโขทัย พิษณุโลก ค่าขนส่งเงินและสินค้า จะมากหรือน้อยขึ้นอยู่กับจุดติดตั้งที่ห่างจากตัวเมือง

3.1.2 สินค้าและบริการของตู้เติมเงิน

1) เติมเงินมือถือ
2) ซื้อบัตรเติมเงิน
3) จ่ายบิล
4) ส่งข้อความ

3.1.3 กลุ่มลูกค้า

ส่วนมากเป็นกลุ่มลูกค้าระดับล่าง เพราะจากรายการทำธุรกรรมมีการเติมเงินในจำนวนที่น้อยในแต่ละครั้ง ดังนั้นทำให้ปัจจัยในตำแหน่งของการติดตั้ง จำเป็นต้องมุ่งเป้าหมายไปยังกลุ่มลูกค้าระดับล่าง โดยจะติดตั้งในสถานที่ที่คาดว่ากลุ่มลูกค้าระดับล่างมีจำนวนมาก

3.1.4 ตำแหน่งที่ติดตั้งตู้เติมเงินใหม่

ใช้วิธีการสำรวจโดยจ้างคนไปสำรวจ โดยมีปัจจัยที่สังเกตดังนี้ติดกับร้านสะดวกซื้อ เพราะป้องกันการขโมยตู้และมีคนดูแลโดยการแบ่งเปอร์เซ็นให้ร้านนั้น ๆ แต่ละตู้ตั้งในถนนเส้นเดียวกันและใกล้กันเพื่อสะดวกต่อการเก็บเงินที่อยู่ภายในตู้ใกล้อำเภอเมือง เพื่อสะดวกในการเก็บเงินภายในตู้ และมีจำนวนผู้ใช้งานมากใกล้บริเวณชุมชน เพื่อมีจำนวนผู้ใช้บริการเยอะใกล้เครื่องให้บริการสินค้าอัตโนมัติประเภทอื่นไม่ตั้งบริเวณร้านค้าสะดวกซื้อเซเว่นอีเลฟเว่น เพราะมีตู้เติมเงินที่เป็นตู้คู่แข่ง ที่ผูกขาดกับเซเว่นอีเลฟเว่นไม่ตั้งที่บริเวณในรั้ว เพราะเป็นการจำกัดจำนวนคนที่มาใช้บริการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 Data Understanding

ขั้นตอนนี้ทำถัดจากขั้นตอนการเข้าใจในธุรกิจ โดยการนำความรู้จากกระบวนการ Business Understanding รวมกับข้อมูลที่ได้รับ นำมาวิเคราะห์ต่อว่าจะมีปัจจัยใดบ้างที่เกี่ยวข้องกับการวางตำแหน่งตู้เติมเงิน

1) เลขตู้ คือ รหัสประจำตู้เติมเงิน ประเภทข้อมูล Numeric
2) สถานที่ตั้ง คือ ตำแหน่งที่วางตู้เติมเงิน ประเภทข้อมูล Text
3) จำนวนเงินที่ได้รับ คือ จำนวนเงินที่ได้รับแต่ละตู้ในแต่ละเดือน หน่วยบาท ประเภทข้อมูล Numeric

นอกจากนี้ยังได้หาข้อมูลเพิ่มเติมมาจาก Google Map API โดยมีการเลือก Attribute จาก Place Types มา 21 Attribute ดังนี้

1) Atm คือ ระยะห่างจากตำแหน่งตู้ถึงตู้ ATM
2) Bank คือ ระยะห่างจากตำแหน่งตู้ถึงธนาคาร
3) Beauty_salon คือ ระยะห่างจากตำแหน่งตู้ถึงร้านตัดผม
4) Bus_station คือ ระยะห่างจากตำแหน่งตู้ถึงอู่รถ
5) Department_store คือ ระยะห่างจากตำแหน่งตู้ถึงห้างสรรพสินค้า
6) Gas_station ระยะห่างจากตำแหน่งตู้ถึงปั๊มน้ำมัน
7) Hardware_store ระยะห่างจากตำแหน่งตู้ถึงร้านค้าเกี่ยวกับคอมพิวเตอร์
8) Hospital ระยะห่างจากตำแหน่งตู้ถึงโรงพยาบาล
9) Pharmacy ระยะห่างจากตำแหน่งตู้ถึงร้านขายยา
10) Police ระยะห่างจากตำแหน่งตู้ถึงสถานีตำรวจ
11) Post_office ระยะห่างจากตำแหน่งตู้ถึงที่ทำการไปรษณีย์
12) Restaurant ระยะห่างจากตำแหน่งตู้ถึงร้านอาหาร
13) School ระยะห่างจากตำแหน่งตู้ถึงโรงเรียน
14) Shopping_mall ระยะห่างจากตำแหน่งตู้ถึงศูนย์การค้า
15) Store ระยะห่างจากตำแหน่งตู้ถึงร้านค้า
16) Supermarket ระยะห่างจากตำแหน่งตู้ถึงซุปเปอร์มาร์เก็ต
17) Local_government_office ระยะห่างจากตำแหน่งตู้ถึงสำนักงานรัฐบาลท้องถิ่น
18) Travel_agency ระยะห่างจากตำแหน่งตู้ถึงบริษัทตัวแทนท่องเที่ยว
19) Convenience_store ระยะห่างจากตำแหน่งตู้ถึงร้านสะดวกซื้อ
20) Train_station ระยะห่างจากตำแหน่งตู้ถึงสถานีรถไฟ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

21) Transit_station ระยะห่างจากตำแหน่งตู้ถึงสถานีขนส่ง

3.3 Data Preparation

3.3.1 จัดการกับ Missing Values

จัดการกับ Missing Values โดยการไม่นำข้อมูลนั้นมาใช้งาน

3.3.2 Data Transformation

ในกระบวนการนี้ได้ทำการ Data Transformation 2 ส่วน

ส่วนที่ 1 ข้อมูลสถานที่ตั้งจากที่เป็น สถานที่ แปลงเป็น Longtitude, Latitude เพื่อนำ Longtitude, Latitude ที่ได้มาใส่เป็นอินพุตของส่วนการหาระยะห่างจากตำแหน่งสถานที่สำคัญโดยใช่ Google Map API แต่เนื่องจากระยะทาง หากมีระยะเกิน 5 กิโลเมตร คาดว่าจะไม่มีผลต่อจำนวนเงินที่ได้รับของตู้เติมเงินอัตโนมัติ ทำให้บาง Attribute ที่กล่าวมา ไม่มีข้อมูลเป็นจำนวนมาก จึงทำการลบ Attribute ออก 7 ตัว คือ Bus_station, Department_store, Train_station, Transit_station, Travel_agency, Hardware_store และ Post_office ทำให้เหลือ Attribute ที่นำไปใช้ในข้อมูล 14 Attribute

ส่วนที่ 2 จำนวนเงินที่ได้รับของทั้ง 6 เดือน นำมาหาค่าเฉลี่ยให้เหลือเพียงค่าเดียวหลังจากนั้นเปลี่ยน Type จาก Numeric เป็น Ordinal โดยแปลงเป็นตัวเลข 1,2,3
โดย 1 นั้นจะแทนการขาดทุนโดยมีจำนวนเงินที่ได้เฉลี่ย 1 - 9000 บาท
และ 2 นั้นจะแทนการเท่าทุนโดยมีจำนวนเงินที่ได้เฉลี่ย 9001 – 18000 บาท
และ 3 นั้นจะแทนการได้กำไร โดยมีจำนวนเงินที่ได้เฉลี่ยมากกว่า 18000 บาท
และจำเป็นต้องแปลงเลข 1 เป็น 100, 2 เป็น 010, 3 เป็น 001 เพื่อนำไปใช้งานกับการทำ Modeling โดย Multilayer Neural Network

3.3.3 Normalization

ทำการ Normalization โดยใช้วิธี Standard Deviation

3.3.4 Dimension Reduction

โดยได้ทดลองใช้ PCA เพื่อลดจำนวน Feature เพื่อให้ได้ Model ที่มีประสิทธิภาพมากขึ้น

3.3.5 Feature Selection

ทำการ Feature Selection โดยใช้วิธีการ Correlation และ Relief

3.4 Modeling

ข้อมูลที่นำมาใช้ มีทั้งหมด 52 แถว

อินพุตมีทั้งหมด 14 Attribute

เอาต์พุตมีทั้งหมด 1 Attribute

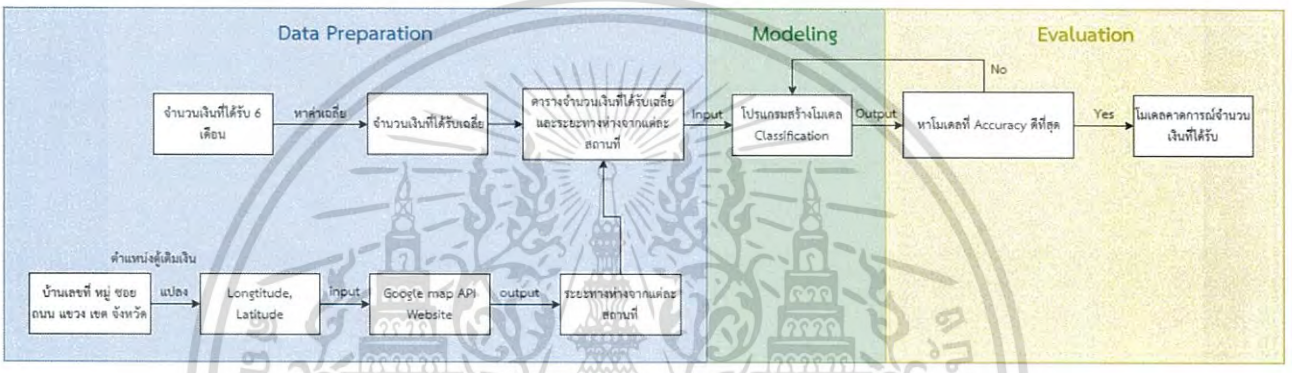
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำ Classification ทดลองใช้วิธี Multilayer Neural Network และปรับค่าต่าง ๆ เลือกวิธีการที่ได้ค่า Accuracy มากกว่ามาทำ Modeling

3.5 Evaluation

ทำการประเมินผลโดยการเปรียบเทียบค่า Accuracy จาก Confusion Matrix จากการทำ 5 Fold Cross Validation ที่ได้ทำการทดลองทำ Modeling

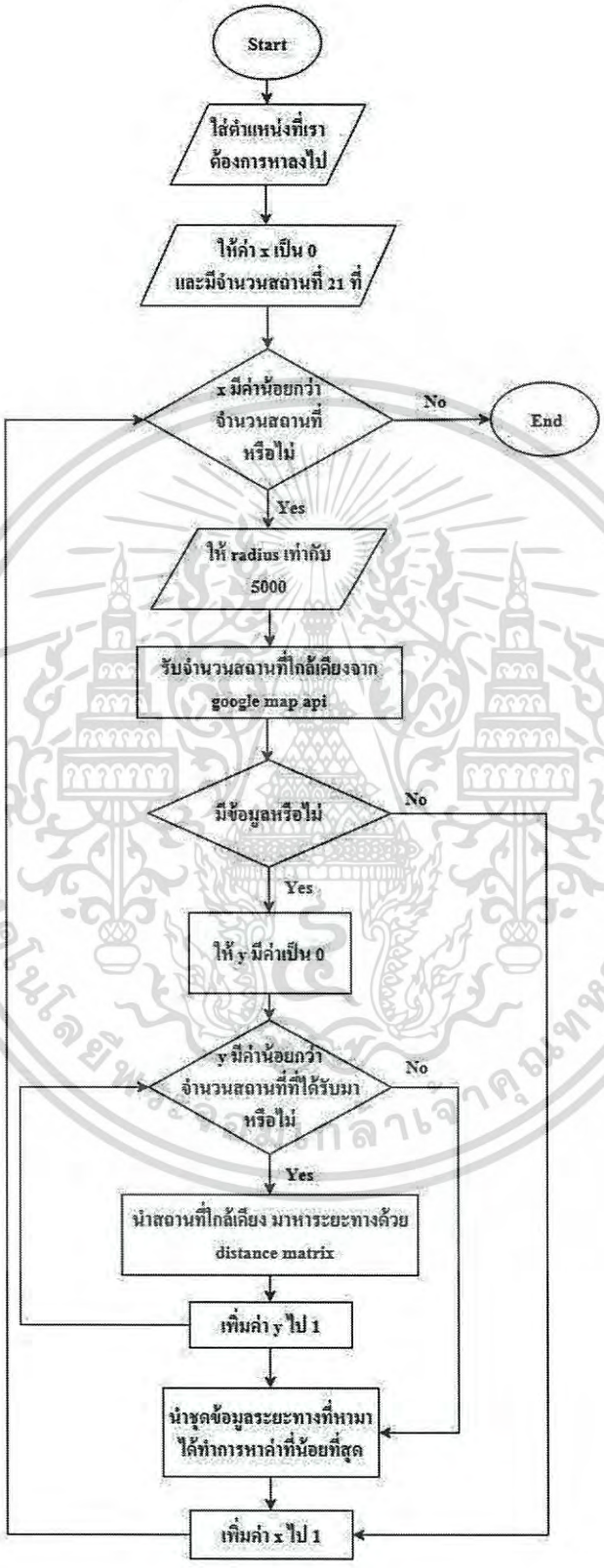
3.6 โครงสร้างของการจัดการข้อมูล



รูป 3.1 โครงสร้างของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

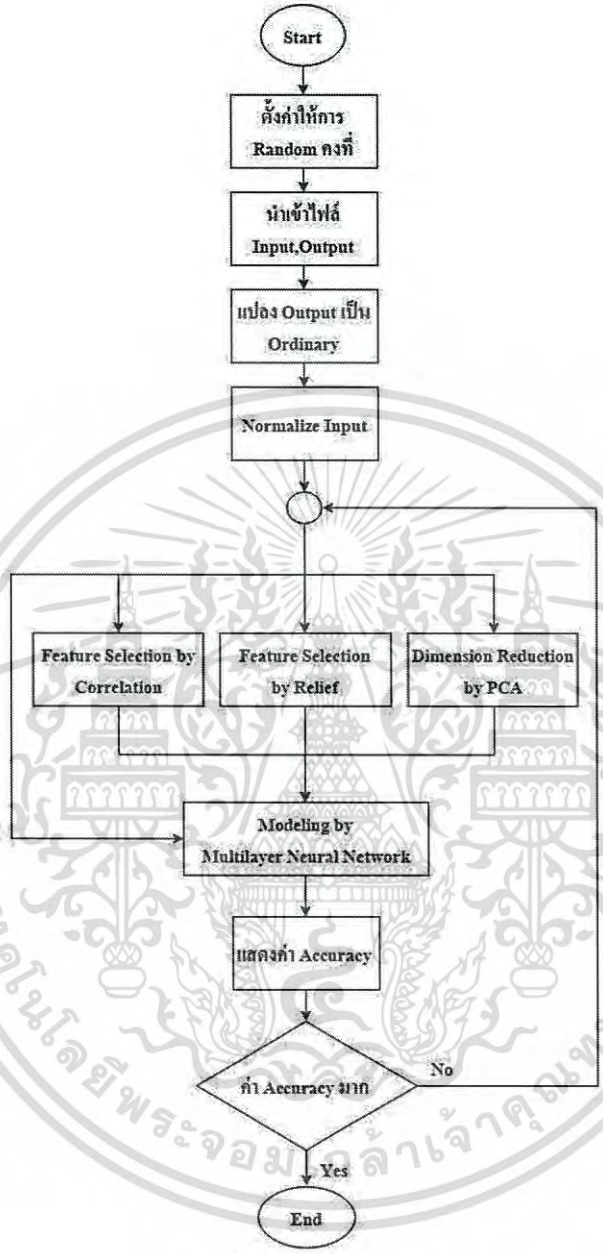
3.7 การออกแบบโปรแกรมสำหรับหาระยะทางโดยใช้ Google Map API



รูป 3.2 Flow Chartโปรแกรมสำหรับหาระยะทางโดยใช้ Google Map API

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.8 การออกแบบการทำงานสำหรับ Modeling



รูป 3.3 Modeling

1) ตั้งค่าให้การ Random คงที่ ทำให้การทดลองทุกครั้งคงที่ เพื่อที่จะนำค่ามาเปรียบเทียบได้
2) นำเข้าไฟล์อินพุตและ เอาต์พุตโดยใช้สกุลไฟล์ CSV โดยอินพุตหมายถึงปัจจัยต่าง ๆ (Feature) และ เอาต์พุต
3) แปลงเอาต์พุตเป็น Ordinary คือทำให้เป็นประเภท กำไร ขาดทุน และเท่าทุน
4) Normalize อินพุตโดยใช้ Standard Deviation
5) เปรียบเทียบค่า Accuracy โดยทดลองใช้ Feature ทั้งหมดสร้างโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6) เปรียบเทียบค่า Accuracy ของการทำ Classification โดย Simple Tree และ Neural Network
7) เปรียบเทียบค่า Accuracy โดยทดลองใช้ Feature Selection และ Dimension Reduction
8) ปรับค่าโมเดลให้มี Accuracy ที่น่าพอใจ

3.9 เครื่องมือที่ใช้พัฒนา

3.9.1 สภาพแวดล้อมในการพัฒนา

1) หน่วยประมวลผล Intel® Core i7-3612QM CPU @ 2.10GHz 2.10GHz
2) หน่วยความจำหลัก 8 GB
3) ระบบปฏิบัติการ Windows 10 64-bit
4) Hard Disk Western Digital Scorpio Blue 1TB
5) กราฟิกการ์ด NVIDIA GeForce GT 630M

3.9.2 ซอฟต์แวร์ที่ใช้ในการพัฒนา

1) XAMPP Control Panel v3.2.2
2) MATLAB R2017a (9.2.0.538062)
3) Excel 2016
4) Weka 3.8

3.9.3 ภาษาที่ใช้ในการพัฒนา

1) MATLAB
2) HTML
3) PHP

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การทดลองและสรุปผลการทดลอง

4.1 ทดลองเขียนเว็ปและใช้ Google Map API เพื่อนำมาช่วยในการเก็บข้อมูล

4.1.1 จุดประสงค์

เพื่อให้เราสามารถรวบรวมข้อมูลได้รวดเร็วมากขึ้น โดยทดลองนำ Google Map API มาช่วยในการทำเว็บไซต์

4.1.2 วิธีการทดลอง

เราใช้ Google map API Nearby Search เพื่อหาสถานที่ประเภทต่าง ๆ ที่อยู่ใกล้เคียงกับตู้จำหน่ายบัตรเติมเงินของเราโดยเรียกใช้งาน API ดังนี้

https://maps.googleapis.com/maps/api/place/nearbysearch/json?location=Origins&radius=radius&type=placetype&key=YOUR_API_KEY

โดย Location เป็นตำแหน่งที่ตู้เราอยู่, Radius เป็นขอบเขตของสถานที่ที่เราต้องการหา, Type เป็นประเภทของสถานที่ที่เราต้องการหา, Key เป็น API Key ของ Google ที่เราทำการขอมา

ตัวอย่างการใช้งาน Nearby Search โดยเรียกใช้งาน API ดังนี้

https://maps.googleapis.com/maps/api/place/nearbysearch/json?location=18.6202733,98.9437574&radius=5000&type=atm&key=YOUR_API_KEY

ตัวอย่าง 4.1 ข้อมูล (บางส่วน) ในรูปแบบ JSON ที่ได้จากการใช้งาน API Nearby Search


```
{
  "html_attributions" : [],
  "results" : [
    {
      "geometry" : {
        "location" : {
          "lat" : 18.617652,
          "lng" : 98.94899199999999
        },
        "viewport" : {
```


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        "northeast" : {
          "lat" : 18.6190009802915,
          "lng" : 98.95034098029149
        },
        "southwest" : {
          "lat" : 18.6163030197085,
          "lng" : 98.94764301970848
        }
      }
    },
    "icon" : "https://maps.gstatic.com/mapfiles/place_api/icons/atm-71.png",
    "id" : "3d48d49e21de51abb3874770a540759c25e6843e",
    "name" : "ATM Bank for Agriculture and Agricultural Cooperatives",
    "place_id" : "ChIJCxZS0T8z2jARDe8SRQoTN1E",
    "reference" : "CmRRAAAAo-hYjruHpjoWp2s-
EcwDUOy64eFPJwVHcdxQYd4JSYmbwxMOT1BmrEYCxatfJT9kgJ2F4kio6G3aAyWzyNg
P6tOhN3d5Ty8Bigeq5TD9OMBqWonG9DON8yuaLRqqKvfYEhCyKfSDE_TwhOPrVdiHj
NPYGhRxd9gFvboZ5ZgDo1faQsouk4jR4A",
    "scope" : "GOOGLE",
    "types" : [ "atm", "finance", "point_of_interest", "establishment" ],
    "vicinity" : "223 Moo 3, San Pa Tong - Nong Tong Road, Tambon Nong Tong, Amphoe
Hang Dong, Chiang Mai, 50340"
  },
```


นอกจากนี้เราได้ใช้ Distance Matrix API ในการหาระยะห่างของสถานที่ที่เราหากับตู้ของเราโดยเรียกใช้งาน API ดังนี้

https://maps.googleapis.com/maps/api/distancematrix/json?origins=Origin&destinations=destination&key=YOUR_API_KEY

โดย Origin เป็นตำแหน่งของของตู้ของเรา, Destination เป็นตำแหน่งของสถานที่ที่เราหามาได้, Key เป็น API Key ของ Google ที่เราทำการขอมา

ตัวอย่างการใช้งาน Nearby Search โดยเรียกใช้งาน API ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

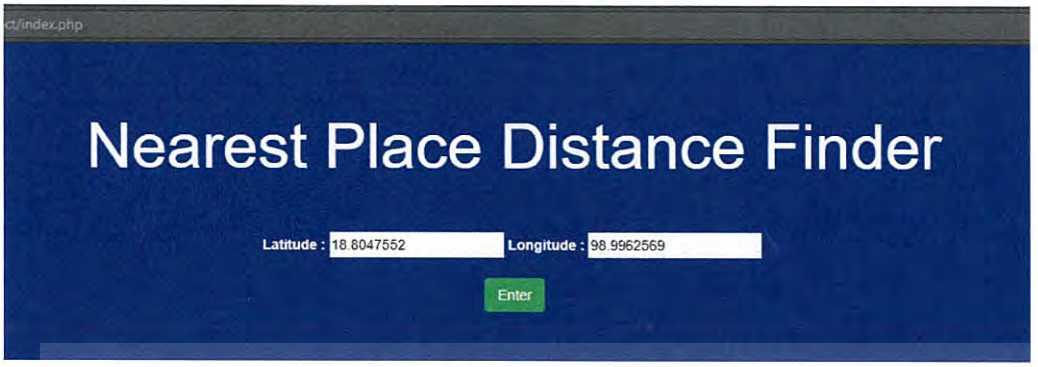
https://maps.googleapis.com/maps/api/distancematrix/json?origins=18.6202733,98.9437574&destinations=18.617652,98.94899199999999&key=YOUR_API_KEY

ตัวอย่าง 4.2 ข้อมูลในรูปแบบ JSON ที่ได้จากการใช้งาน Distance Matrix API

```
{
  "destination_addresses" : [
    "1015 ตำบล หนองตอง อำเภอ หางดง เชียงใหม่ 50340 ประเทศไทย"
  ],
  "origin_addresses" : [
    "1015 ตำบล หนองตอง อำเภอ หางดง เชียงใหม่ 50340 ประเทศไทย"
  ],
  "rows" : [
    {
      "elements" : [
        {
          "distance" : {
            "text" : "0.6 กม.",
            "value" : 626
          },
          "duration" : {
            "text" : "1 นาที",
            "value" : 58
          },
          "status" : "OK"
        }
      ]
    }
```

และหลังจากที่เราได้ระยะทางจาก Distance Matrix API แล้วเราก็นำระยะทางนั้นมาหาสถานที่ที่ห่างจากตู้จำหน่ายบัตรเติมเงินน้อยที่สุดและแสดงค่าออกมาโดยการในการใช้งานเว็ปนั้นจะเริ่มจากหน้า Index

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 4.1 หน้า Index ของ website

โดยในหน้า Index นั้นเราใช้ในการใส่สถานที่ตู้ที่เราต้องการหาสถานที่น่าสนใจใกล้เคียง โดยค่า Latitude นั้นจะใส่ค่าละติจูดของตู้นั้น ๆ , ค่า Longitude นั้นจะใส่ค่าลองจิจูดของตู้นั้น ๆ และปุ่ม Enter มีไว้ส่งค่าไปหน้าถัดไปเพื่อหาสถานที่ใกล้เคียง และในหน้าถัดมาจะเป็นหน้าแสดงผลสถานที่ใกล้เคียงที่หามาได้จาก Google Map API

Origins: 18.8047552, 98.9962569

Place Type	Name	Distance
atm	ATM K- Bank	303
bank	กรุงศรี ออโต้ (สินเชื่อรถยนต์) ดาษาเชียงใหม่	1025
beauty_salon	Maikaew Beauty Salon	429
bus_station	Chiang Mai Bus Terminal 1 (Chang Phueak Bus Terminal)	1550
department_store	Candy	1667
gas_station	สน ปตท. บริษัท เชียงใหม่ลักษณพงษ์ จำกัด	987
hardware_store	บริษัท นานา แมชชีนพาร์ท จำกัด	1124
hospital	Saint Peter Eye Hospital โรงพยาบาลตาเซ็นต์ปีเตอร์	1313
pharmacy	ณัฐพงศ์เภสัช	342
police	ตู้ยามจราจรเชียงใหม่	1698
post_office	Talat Kam Tieng Post Office	659
restaurant	ร้านอาหารทิวไผ่	73
school	โรงเรียนสอนภาษาญี่ปุ่น แนวธุรกิจ ทายาฬ่า เอ็นจีโอ จิสเซน	535
shopping_mall	Greenheart (JJ Market Chiang Mai)	426
store	Lanna InterPack Limited Partnership	97

รูป 4.2 หน้าแสดงผลสถานที่ใกล้เคียง

โดย Origin นั้นจะเป็นละติจูดและลองจิจูดของตู้ที่เราได้ใส่เข้ามา, Place Type นั้นจะแสดงถึงสถานที่ประเภทต่าง ๆ ที่เราได้รับมาจาก Google Map API, Name นั้นจะแสดงชื่อของสถานที่ ๆ เราได้รับมาจาก Google Map API และ Distance นั้นจะแสดงถึงระยะที่เราได้รับมาจาก Google Map API และ Distance Metrix API โดยจะแสดงในหน่วยเมตร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 ทดลองใช้ Feature Selection และ Dimension Reduction

4.2.1 จุดประสงค์

เพื่อเพิ่มประสิทธิภาพของโมเดล และลดความซ้ำซ้อนของตัวแปร โดยการเปรียบเทียบ การทำ Feature Selection ด้วยวิธี Correlation, Relief, การทำ Dimension Reduction และการใช้ Feature ทั้งหมดที่มีอยู่

4.2.2 วิธีการทดลอง

ตัวแปรต้น วิธีในการทำ Feature Selection ต่าง ๆ และ การทำ Dimension Reduction โดย PCA ตัวแปรตาม ค่า Accuracy ของโมเดลตัวแปรที่ต้องควบคุม สร้างโมเดล Classification โดยวิธี Multilayer Neural Network อินพุตจำนวน 14 Feature จำนวน 52 แถว ตรวจสอบความถูกต้องโดยวิธี 5 Fold Cross Validation

4.2.3 ผลการทดลอง

ตาราง 4.1 ผลการทดลองด้วย Feature ทั้ง 21 ตัว

Hidden Layer	Train function	Transfer Function	Accuracy (%)
6	trainscg	logsig	61.5

การใช้ Feature ทั้งหมดที่มีอยู่ทำให้ได้ ค่า Accuracy อยู่ที่ 61.5% แต่ยังไม่น่าพึงพอใจ จึงต้องใช้การทำ Feature Selection และ Dimension Reduction

ตาราง 4.2 ผลการทดลองใช้ Feature Selection แบบ Correlation

Hidden Layer	Train Function	Transfer Function	Accuracy (%)	Amount of Feature
6	trainscg	logsig	63.08	12
6	trainscg	logsig	44.68	10
6	trainscg	logsig	45.38	8
6	trainscg	logsig	62.38	6

จากการทดลองใช้ Feature Selection แบบ Correlation ได้ทำการทดลองให้มีจำนวน Feature 1 - 13 พบว่าหากมีจำนวน Feature 12 ตัว และ 6 ตัว จะได้ค่า Accuracy มากกว่าการใช้ Feature ทั้งหมด แต่เมื่อใช้จำนวน Feature อื่นทำให้ Accuracy น้อยกว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 4.3 ผลการทดลองใช้ Feature Selection แบบ ReliefF

Hidden Layer	Train Function	Transfer Function	Accuracy (%)	Amount of Feature
6	trainscg	logsig	60.77	12
6	trainscg	logsig	54.23	10
6	trainscg	logsig	63.08	8
6	trainscg	logsig	65.77	6

จากการทดลองใช้ Feature Selection แบบ ReliefF ได้ทำการทดลองให้มีจำนวน Feature 1 - 13 พบว่าที่จำนวน Feature 8 และ 6 มีค่า Accuracy มากกว่าการใช้ Feature ทั้งหมด

ตาราง 4.4 ผลการทดลองใช้ Dimension Reduction แบบ PCA

Hidden Layer	Train Function	Transfer Function	Accuracy (%)	Amount of Feature
6	trainscg	logsig	63.08	14
6	trainscg	logsig	61.54	12
6	trainscg	logsig	58.46	10
6	trainscg	logsig	61.92	8
6	trainscg	logsig	62.69	6

จากการทดลองใช้ Dimension Reduction โดย PCA ได้ทำการทดลองให้มีจำนวน Feature 1 - 13 พบว่ามีหลายตัวที่มีค่า Accuracy มากกว่าการใช้ Feature ทั้งหมด

จากการใช้ Feature Selection และ Dimension Reduction พบว่า ได้ค่า Accuracy ที่ดีกว่าแบบใช้ Feature ทั้งหมดที่มี ซึ่ง ReliefF มีค่า Accuracy ได้สูงสุด ตามมาด้วย PCA และ Correlation จึงนำมาพิจารณาในการทดลองถัดไปด้วย

4.3 การปรับค่าโมเดล Neural Network

4.3.1 จุดประสงค์

จุดประสงค์ เพื่อเพิ่มประสิทธิภาพของโมเดล เพื่อให้ได้ค่า Accuracy ที่เหมาะสมกับปัญหานี้

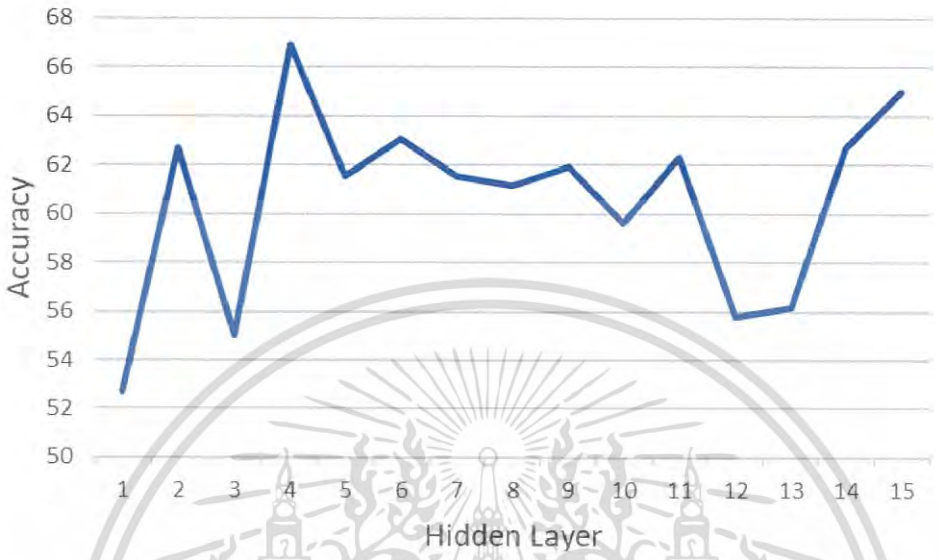
4.3.2 วิธีการทดลอง

ตัวแปรต้น Train Function, จำนวน Hidden layer, Transfer Function ตัวแปรตาม ค่า Accuracy ของโมเดล ตัวแปรที่ต้องควบคุม สร้างโมเดล Classification โดยวิธี Multilayer Neural Network , อินพุต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จำนวน 14 Feature จำนวน 52 แถว, จำนวน Principal Components, ตรวจสอบความถูกต้องโดยวิธี 5 Fold Cross Validation ได้ทดลองเปลี่ยนค่า Hidden Layer, Transfer Function, Train Function

4.3.3 ผลการทดลอง



รูป 4.3 ผลการทดลองเปลี่ยน Hidden Layer

ทดลองเปลี่ยนค่า Hidden Layer ตั้งแต่ 1 - 15 พบว่า Hidden layer ช่วงประมาณ 5-8 มีค่า Accuracy ที่มากที่สุด

ตาราง 4.5 ผลการทดลองเปลี่ยน Transfer Function

Hidden Layer	PCA Size	Train Function	Transfer Function	Accuracy (%)
6	14	trainscg	tansig	63.08
6	14	trainscg	logsig	64.62
6	14	trainscg	purelin	63.85
6	6	trainscg	tansig	65.77
6	6	trainscg	logsig	60.77
6	6	trainscg	purelin	48.08

ทดลองเปลี่ยน Transfer function เป็น Hyperbolic Tangent Sigmoid Transfer Function (tansig), Log-Sigmoid Transfer Function (logsig), Linear Transfer Function (purelin) พบว่า ที่ PCA Size = 14 Transfer Function ทั้งสามตัวได้ค่า Accuracy ใกล้เคียงกัน จึงนำไปพิจารณาทั้งหมด ส่วน PCA Size = 6 tansig ได้ Accuracy ที่มากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

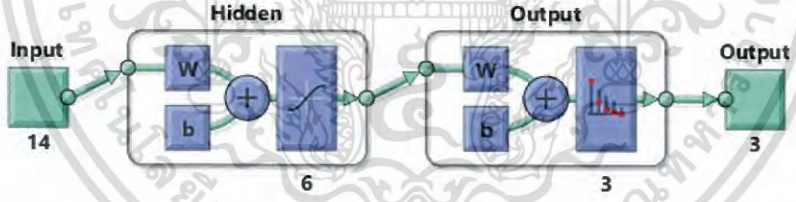
ตาราง 4.6 ผลการทดลองเปลี่ยน Train Function

Hidden Layer	PCA Size	Train Function	Transfer Function	Accuracy (%)
6	14	trainscg	tansig	63.08
6	14	trainlm	tansig	60.77
6	14	trainbr	tansig	79.23
8	14	trainscg	tansig	55.77
8	4	trainlm	tansig	60.38
8	14	trainbr	tansig	71.92

ทดลองเปลี่ยน Train Function เป็น Scaled Conjugate Gradient Backpropagation (trainscg), Levenberg-Marquardt Backpropagation (trainlm), Bayesian Regularization Backpropagation (trainbr) พบว่า trainbr ได้ Accuracy ที่มากที่สุด

4.4 สรุปผลการทดลองทั้งหมด

จากการทดลองทั้งหมดที่ผ่านมา ทำให้เห็นถึงความสอดคล้อง หรือคุณสมบัติของตัวแปร จึงได้ทำการทดลองมากขึ้น จนได้ค่า Accuracy มากที่สุด คือ 79.23% จากการสร้างโมเดลดังนี้



รูป 4.4 Multilayer Neural Network

อินพุต = 14 Feature

เอาต์พุต = 1 Feature

มีจำนวนทั้งหมด 52 แถว

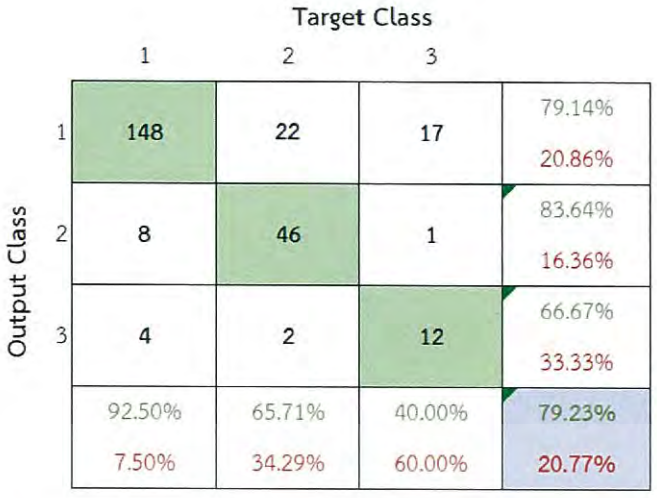
จำนวน Principal Components = 14 ตัว

ทำ Classification ด้วยวิธี Multilayer Neural Network มี Hidden Layer = 6 Layer

ใช้ Hyperbolic Tangent Sigmoid Transfer Function และ Bayesian Regularization Backpropagation

ประเมินผลโดย 5 Fold Cross validation ได้ผลดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 4.5 Confusion Matrix แสดงรวมทั้ง 5 Fold Cross Validation

จาก Confusion Matrix แสดงค่า Accuracy = 79.23%

Precision ของการขาดทุน (คลาส 1) = 79.14%

Precision ของการเท่าทุน (คลาส 2) = 83.64%

Precision ของการได้กำไร (คลาส 3) = 66.67%

Recall ของการขาดทุน (คลาส 1) = 92.50%

Recall ของการเท่าทุน (คลาส 2) = 65.71%

Recall ของการได้กำไร (คลาส 3) = 40.00%

สังเกตได้ว่า Recall ของการได้กำไรมีค่าน้อยที่สุด แสดงถึงคลาสการได้กำไร มีทำนายได้ถูกต้องน้อยที่สุด อาจจะเป็นเพราะ ชุดข้อมูลมีจำนวนคลาสที่ได้กำไรไม่มากพอที่จะทำให้โมเดลทำนายได้แม่นยำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
บทสรุปและข้อเสนอแนะ

5.1 บทสรุปและบทวิจารณ์

งานวิจัยนี้ สร้างโมเดลที่นำมาทำนายว่า หากติดตั้งเครื่องจำหน่ายสินค้า ณ ตำแหน่งนั้น จะได้กำไรเท่าทุน หรือขาดทุน เพื่อลดความเสี่ยงในการลงทุนติดตั้งสำหรับธุรกิจเครื่องจำหน่ายสินค้าอัตโนมัติประเภทตู้เติมเงิน โดยใช้หลักการและวิธีการทำเหมืองข้อมูล ในการวิเคราะห์ข้อมูลที่ได้รับมา เพื่อหาโมเดลที่สามารถทำนายข้อมูลได้แม่นยำที่สุด

5.2 ปัญหาและอุปสรรค

1) ระยะทางระหว่างสถานที่ที่ใกล้เคียงในการวิจัยนี้นั้นได้รับมาจากใน Google Map เท่านั้น ทำให้บางสถานที่ที่มีอยู่จริงแต่ไม่ได้ปรากฏใน Google Map จะไม่ได้นำไปเป็นอินพุตในโมเดล
2) ในการนำโมเดลไปใช้งานจริง อาจจะต้องพิจารณาในปัจจัยอื่น ๆ ก่อนที่จะนำไปติดตั้งด้วย

5.3 แนวทางในการแก้ไขปัญหา

1) ลงพื้นที่จริงเพื่อเก็บข้อมูลปัจจัยที่เกี่ยวข้องในการสร้างโมเดล
2) ใช้ข้อมูลจากแหล่งที่อื่นนอกเหนือจากใน Google Map

5.4 แนวทางในการพัฒนาต่อ

1) เพิ่มปัจจัยที่เกี่ยวข้องกับโมเดล หรือเก็บข้อมูลเพิ่ม เพื่อให้โมเดลมีประสิทธิภาพมากขึ้น
2) สามารถนำข้อมูลที่ได้มา ทำนายสิ่งอื่นนอกเหนือจากงานวิจัยนี้ เช่น วิเคราะห์การติดตั้งในระดับจังหวัด หรือทำระบบแนะนำบริการให้กับผู้ใช้งาน
3) นำกระบวนการ และวิธีการไปใช้กับเครื่องจำหน่ายสินค้าประเภทอื่นได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

Jia-Lang Seng and T.C. Chen. 2010.

"An analytic approach to select data mining for business decision". Expert Systems with Applications 37(2010) 8042-8057

Fatemeh Rahimi-Ajdadi and Yousef Abbaspour-Gilandeh. 2011.

"Artificial Neural Network and stepwise multiple range regression methods for prediction of tractor fuel consumption.". Measurement 44(2011) 2104-2111

Feng-Cheng Lin, Hsin-Wen Yu ,Chih-Hao Hsu and Tzu-Chun Weng. 2011.

"Recommendation system for localized products in vending machines". Expert Systems with Applications 38(2011) 9129-9138

Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer and Rüdiger Wirth. 2010. "An analytic approach to select data mining for business decision". Expert Systems with Applications 37(2010) 8042-8057

Dorian Pyle. 2010. **Data Preparation for Data Mining.** San Francisco : Morgan Kaufmann Publishers.

NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA), and OHRA Verzekeringen en BankGroep B.V. (The Netherlands).1999. **CRISP-DM 1.0.** USA : SPSS.Inc.

Jiawei Han, Micheline Kamber, Jian Pei. 2556. **Data Mining Concepts and Techniques.** [Online]. Available : https://www.slideshare.net/salahecom/data-mining-concepts-and-techniques-3rd-ed-chapter-3-preprocessing

K.Keerthivasan and B.Surendiran. 2016.

"Dimensionality reduction using Principal Component Analysis for network intrusion detection". Perspectives in Science 8(2016) 510-512

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Mathworks. 2018. **tansig.**

[Online]. Available :_https://www.mathworks.com/help/nnet/ref/tansig.html

Mathworks. 2018. **logsig.**

[Online]. Available :_https://www.mathworks.com/help/nnet/ref/logsig.html

Mathworks. 2018. **purelin.**

[Online]. Available :_https://www.mathworks.com/help/nnet/ref/purelin.html

Mathworks. 2018. **trainscg.**

[Online]. Available :_https://www.mathworks.com/help/nnet/ref/trainscg.html

Mathworks. 2018. **trainlm.**

[Online]. Available :_https://www.mathworks.com/help/nnet/ref/trainlm.html

Mathworks. 2018. **trainbr.**

[Online]. Available :_https://www.mathworks.com/help/nnet/ref/trainbr.html

Mathworks. 2018. **cvpartition.**

[Online]. Available :_https://www.mathworks.com/help/stats/cvpartition.html

Tutorialspoint. 2018. **Data Mining - Decision Tree Induction.**

[Online]. Available :_https://www.tutorialspoint.com/data_mining/dm_dti.htm

The Pennsylvania State University. 2018. **Principal Component Analysis (PCA) Procedure.**

[Online]. Available :_https://onlinecourses.science.psu.edu/stat505/node/51

Researchgate. 2018. **The K-fold cross-validation scheme.** [Online].

Available :_https://www.researchgate.net/figure/The-K-fold-cross-validation-scheme-133-Each-of-the-K-partitions-is-used-as-a-test_fig10_323969239

Dataschool. 2014. **Simple guide to confusion matrix terminology.** [Online].

Available : http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Wikipedia. **PHP.** [Online]. Available : https://en.wikipedia.org/wiki/PHP

Wikipedia. **MATLAB.** [Online]. Available : https://en.wikipedia.org/wiki/MATLAB

Machine Learning Group at the University of Waikato. **Data Mining Software in Java.** [Online]. Available : https://www.cs.waikato.ac.nz/ml/weka/

Google Maps APIs. 2018. **Distance & duration for multiple destinations and transport modes.** [Online]. Available : https://developers.google.com/maps/documentation/distance-matrix/

Google Maps APIs. 2018. **Overview.** [Online]. Available : https://developers.google.com/maps/documentation/javascript/tutorial

Wikipedia. **Relief (feature selection).** [Online]. Available : https://en.wikipedia.org/wiki/Relief_(feature_selection)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้