การพัฒนาตัวแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล

นายปรีชา ลิ้มตระกูล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
พ.ศ. 2558

Development of a Model to Predict Cassava Yield Using Data Mining Techniques

Mr. Preecha Limtrakul

A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Master of Science in Information and Communication Technology
School of Science and Technology
Sukhothai Thammathirat Open University
2015

หัวข้อวิทยานิพนธ์ การพัฒนาตัวแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล

ชื่อและนามสกุล นายปรีชา ลิ้มตระกูล

แขนงวิชา เทคโนโลยีสารสนเทศและการสื่อสาร

สาขาวิชา วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

อาจารย์ที่ปรึกษา 1. รองศาสตราจารย์ ดร. วิภา เจริญภัณฑารักษ์

2. อาจารย์วิทยา พรพัชรพงศ์

วิทยานิพนธ์นี้ ได้รับความเห็นชอบให้เป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรระดับปริญญาโท เมื่อวันที่ 25 กุมภาพันธ์ 2559

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(อาจารย์ ดร. ดวงดาว วิชาดากุล)

..... กรรมการ
(รองศาสตราจารย์ ดร. วิภา เจริญภัณฑารักษ์)

..... กรรมการ
(อาจารย์วิทยา พรพัชรพงศ์)

..... ประธานกรรมการบัณฑิตศึกษา
(รองศาสตราจารย์ ดร. สุจินต์ วิศวธีรานนท์)

ชื่อวิทยานิพนธ์ การพัฒนาตัวแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล
ผู้วิจัย นายปรีชา ลิ้มตระกูล **รหัสนักศึกษา** 2549600068 **ปริญญา** วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีสารสนเทศและการสื่อสาร) **อาจารย์ที่ปรึกษา** (1) รองศาสตราจารย์ ดร. วิภา เจริญภัณฑารักษ์ (2) อาจารย์วิทยา พรพัชรพงศ์ **ปีการศึกษา** 2558

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อพัฒนาตัวแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล 2) เพื่อพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจของผู้ใช้งานระบบผ่านเว็บไซต์

โดยใช้ข้อมูลปัจจัยการผลิตมันสำปะหลังจากสำนักงานเกษตรจังหวัดกำแพงเพชร เพื่อสร้างตัวแบบพยากรณ์ โดยใช้เทคนิคการจำแหนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) โดยจะใช้อัลกอริทึม จำนวน 5 ตัว ได้แก่ J48, Random Tree, Simple Cart, Naïve Bayes, และLAD Tree แล้วทำการทดสอบตัวแบบการพยากรณ์ด้วยวิธี Cross-validation Test พบว่ายังให้ค่าแม่นยำเพียง 70.96% ซึ่งผู้วิจัยเห็นว่ายังไม่ดีพอ จึงได้ปรับปรุงวิธีการทดสอบโดยการแบ่งชุดข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) จำนวน 5 ชุด แล้วทำการสร้างตัวแบบการพยากรณ์ด้วยอัลกอริทึม J48, Random Tree, Simple Cart, Naïve Bayes และ LAD Tree อีกครั้ง พบว่ามีค่าความแม่นยำสูงขึ้นกว่าวิธี Cross-validation Test ในทุกอัลกอริทึม ดังนั้นผู้วิจัยจึงเลือกตัวแบบการพยากรณ์ที่ให้ค่าความแม่นยำสูงสุด ในส่วนของข้อมูลทดสอบ (Test Set) อัลกอริทึม J48 ให้ค่าความแม่นยำสูงสุดที่ 75.64% อัลกอริทึม Simple Cart ให้ค่าความแม่นยำสูงสุดที่ 80.12% และอัลกอริทึม LAD Tree ให้ค่าความแม่นยำสูงสุดที่ 89.55 % ส่วนอัลกอริทึม Random Tree และอัลกอริทึม Naïve Bayes ให้ค่าความแม่นยำต่ำกว่า 70% จึงไม่นำมาพิจารณา

จากนั้นนำอัลกอริทึมทั้งสามตัวที่ให้ค่าความแม่นยำสูงสุดไปพัฒนาตัวแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล ระบบสารสนเทศ และการสืบค้นข้อมูล ในรูปแบบเว็บแอพพลิเคชั่น ซึ่งระบบสามารถสืบค้น ปรับปรุง เพิ่มเติม บันทึก และแสดงรายงานข้อมูลมันสำปะหลังได้ และผลการตอบแบบประเมินความพึงพอใจผ่านเว็บไซต์ของระบบสารสนเทศดังกล่าว จากเจ้าหน้าที่สำนักงานเกษตรจังหวัดกำแพงเพชร ผู้ใช้งานทั่วไป และผู้ดูแลระบบ รวมทั้งสิ้น 30 คน มีความพึงพอใจในการใช้งานระบบดังกล่าวเฉลี่ยที่ 91% ซึ่งถือว่าอยู่ในระดับที่ดีมาก

คำสำคัญ การทำเหมืองข้อมูล การพยากรณ์ ต้นไม้ตัดสินใจ

Thesis title: Development of a Model to Predict Cassava Yield Using Data Mining Techniques
Researcher: Mr. Preecha Limtrakul; **ID:** 2549600068;
Degree: Master of Science (Information and Communication Technology);
Thesis advisors: (1) Dr. Vipa Jaroenpuntaruk, Associate Professor;
(2) Wittaya Pornpatcharapong, Lecturer; **Academic year:** 2015

Abstract

The objective of this research were 1) to development of model to predict the yield of cassava using data mining techniques, 2) to develop a forecast yield of cassava. With queries system to evaluate satisfaction of users of the system through the website.

The process on data from the Kamphaengphet Provincial Agriculture Extension Office. A predictive model was based on techniques Classification with a decision tree, using an algorithm of five, including the J48, Random Tree, Simple Cart, Naïve Bayes, and LAD Tree. Then tested the predictive models provide a technical Cross-validation The result found that the accuracy was only 70.96%. Which was not good enough. Thevefove, testing methods had improved by splitting the series into two parts Including training set and 5 test set and then create predictive models. Using the folloning: Algorithm J48, SimpleCart, LAD Tree, Random Tree and Naïve Bayes again. The model obtained from Training set method were more accurate than from Cross-validation Test ins in all algorithms used, which were J48, Random Tree, Simple Cart, NaïveBayes, and LAD Tree Thus the selection of predictive models that provided the highest accuracy. In terms of Test Set, Algorithm J48 for maximum accuracy at 75.64%, Simple Cart with 80.12%, and LAD Tree with 89.55% while algorithms Random Tree and Naïve Bayes the accuracy lower than 70% were not taken into consideration.

Then only the first three models were used to implement the development of model to predict the yield of cassava. The model was integrated with the web-based information system for information search on cassava, which was designed and developed in this research to provide basic information such as update, add, and reports The for decision suppost respondents were rated their satisfaction over the site of such information. Officials of Kamphaengphet Provincial Agriculture Extension, users, and administrator a total of 30 people. There was satisfaction in the use of such systems by averaging the responses to a satisfaction rating of over 91%, which was considered very good.

Keywords: Data Mining, Prediction, Decision Tree

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาอย่างสูงจาก รองศาสตราจารย์ ดร.วิภา เจริญภัณฑารักษ์ และอาจารย์ วิทยา พรพัชรพงศ์ ซึ่งเป็นอาจารย์ที่ปรึกษา ได้สละเวลาให้คำปรึกษา แนะนำในการแก้ไขข้อบกพร่องต่างๆ ด้วยความเมตตาและเอาใจใส่เป็นอย่างดียิ่ง ขอกราบขอบพระคุณในความกรุณาของอาจารย์ทั้งสองมา ณ โอกาสนี้

ขอกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่าน ที่กรุณาให้คำแนะนำ เพื่อแก้ไขให้วิทยานิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น ขอขอบพระคุณเจ้าหน้าที่ประจำสาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช ทุกท่าน โดยเฉพาะคุณวันเพ็ญ บุญนุกูล ที่ช่วยกรุณาประสานงานในส่วนต่างๆ ที่เกี่ยวข้องเป็นอย่างดียิ่ง และขอขอบคุณเจ้าหน้าที่ของสำนักงานเกษตรจังหวัดกำแพงเพชรทุกท่าน ที่กรุณาอนุเคราะข้อมูลเพื่อนำมาใช้ประกอบในการจัดทำวิทยานิพนธ์ฉบับนี้

สุดท้ายขอกราบเท้าบิดา มารดา ผู้มีพระคุณอันยิ่งใหญ่และเป็นผู้ให้ทุกสิ่งทุกอย่างโดยไม่หวังผลตอบแทนอันใด อีกทั้งครอบครัวลิ้มตระกูลอันเป็นที่รัก และบุคคลที่ใกล้ชิดที่ให้กำลังใจ และห่วงใยเสมอมา ส่งผลให้ข้าพเจ้าประสบความสำเร็จในวันนี้

ปรีชา ลิ้มตระกูล
กุมภาพันธ์ 2559

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญตาราง ..... ฌ
สารบัญภาพ ..... ญ
บทที่ 1 บทนำ 1
ความเป็นมาและความสำคัญของปัญหา 1
วัตถุประสงค์ของการวิจัย 2
ขอบเขตของงานวิจัย 2
ประโยชน์ที่คาดว่าจะได้รับ 2
บทที่ 2 วรรณกรรมที่เกี่ยวข้อง 3
การปลูกมันสำปะหลัง 3
ความหมายของการทำเหมืองข้อมูล 8
เทคนิคการทำเหมืองข้อมูล 12
ต้นไม้ตัดสินใจ (Decision Tree) 14
แนวคิดและทฤษฎีเกี่ยวกับตัวแบบพยากรณ์แบบต้นไม้ตัดสินใจ 17
การประเมินประสิทธิภาพของโมเดล 18
การวิเคราะห์ประสิทธิภาพ 19
งานวิจัยที่เกี่ยวข้อง 20
บทที่ 3 วิธีดำเนินการวิจัย 25
ข้อมูลและกลุ่มตัวอย่างที่ใช้ในการวิจัย 26
การทำข้อมูลให้สมบูรณ์ 26
การคัดเลือกข้อมูล 29
การแปลงข้อมูล 30
การพัฒนาตัวแบบพยากรณ์ 30
การออกแบบระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ 36
สถิติที่ใช้ในงานวิจัย 38

สารบัญ (ต่อ)

หน้า

บทที่ 4 ผลการวิจัย 40
ผลการพัฒนาตัวแบบพยากรณ์ 40
ผลการทดสอบตัวแบบพยากรณ์ 40
ผลการออกแบบระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง
การสืบค้นข้อมูลและการประเมินความพึงพอใจผ่านเว็บไซต์ 43
บทที่ 5 สรุปผลการวิจัย 54
สรุปการวิจัย 54
การอภิปรายผล 55
ข้อเสนอแนะ 56
บรรณานุกรม 57
ภาคผนวก 61
ก แบบประเมินความพึงพอใจการใช้งานระบบพยากรณ์ผลผลิตมันสำปะหลัง
และระบบสืบค้นข้อมูล 62
ข หนังสือถึงสำนักงานเกษตรจังหวัดกำแพงเพชร เพื่อขอความอนุเคราะห์เก็บข้อมูล 65
ประวัติผู้วิจัย 67

สารบัญตาราง

หน้า

ตารางที่ 2.1 ค่าของคอนฟิวชันเมทริกซ์ (Confusion Matrix) แบบ 2 กลุ่ม 19
ตารางที่ 3.1 แทนความหมายในแอททริบิวต์ 27
ตารางที่ 3.2 กำหนดจังหวัด 28
ตารางที่ 3.3 กำหนดรหัสพันธุ์ 28
ตารางที่ 3.4 กำหนดรหัสปุ๋ย 28
ตารางที่ 3.5 กำหนดรหัสหน่วยนับ 29
ตารางที่ 3.6 ผลการเคราะห์สหสัมพันธ์ (Correlation Analysis) 29
ตารางที่ 4.1 แสดงค่าระดับผลผลิต 40
ตารางที่ 4.2 แสดงผลการทดสอบตัวแบบพยากรณ์ด้วยเทคนิค Cross-validation Test 41
ตารางที่ 4.3 ค่าแสดงการแบ่งสัดส่วนของข้อมูล 41
ตารางที่ 4.4 สรุปความแม่นยำในการพยากรณ์ของข้อมูลโดยวิธีการแบ่งข้อมูลทั้งชุดเป็นข้อมูลเรียนรู้และข้อมูลทดสอบ (Data Set = Tranining Set + Test Set) 42
ตารางที่ 4.5 สรุปผลการทดสอบตัวแบบพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) 43
ตารางที่ 4.6 สรุปผลการตอบแบบประเมินความพึงพอใจการใช้งานระบบพยากรณ์ผลผลิตมันสำปะหลัง และระบบสืบค้นข้อมูล 53

สารบัญภาพ

หน้า

ภาพที่ 2.1 แหล่งความรู้จากศาสตร์ต่างๆ ที่นำมาทำเหมืองข้อมูล 9
ภาพที่ 2.2 ขั้นตอนการทำเหมืองข้อมูล 10
ภาพที่ 2.3 แผนภาพต้นไม้เพื่อการตัดสินใจ 13
ภาพที่ 2.4 โครงสร้างเครือข่ายใยประสาท 14
ภาพที่ 2.5 การแบ่งข้อมูลสำหรับทดสอบแบบ k-Fold Cross Validation 18
ภาพที่ 3.1 แผนผังขั้นตอนการดำเนินงานวิจัย 25
ภาพที่ 3.2 ไฟล์ข้อมูลที่นำเข้าสู่โปรแกรม WEKA 31
ภาพที่ 3.3 ขั้นตอนการนำเข้ามูลสู่โปรแกรม WEKA 32
ภาพที่ 3.4 ขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิค 5-fold cross-validation ของอัลกอริทึม J48 33
ภาพที่ 3.5 ขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะเป็นในส่วนของ Data Set ของอัลกอริทึม J48 34
ภาพที่ 3.6 ขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะเป็นในส่วนของ Traning Set ของอัลกอริทึม J48 35
ภาพที่ 3.7 ขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะเป็นในส่วนของ Test Set ของอัลกอริทึม J48 36
ภาพที่ 4.1 แผนภาพบริบทของระบบพยากรณ์ผลผลิตมันสำปะหลังและระบบสืบค้นข้อมูล 44
ภาพที่ 4.2 แผนภาพกระแสข้อมูลที่ระดับ 0 (Data Flow Diagram Level 0) 44
ภาพที่ 4.3 แผนภาพกระแสข้อมูลที่ระดับ 1 (Data Flow Diagram Level 1) 45
ภาพที่ 4.4 แผนภาพกระแสข้อมูลที่ระดับ 2 (Data Flow Diagram Level 2) 45
ภาพที่ 4.5 แผนภาพกระแสข้อมูลที่ระดับ 3 (Data Flow Diagram Level 3) 46
ภาพที่ 4.6 แผนภาพกระแสข้อมูลที่ระดับ 4 (Data Flow Diagram Level 4) 46

สารบัญภาพ (ต่อ)

หน้า

ภาพที่ 4.7 หน้าจอหลักของระบบ 47
ภาพที่ 4.8 หน้าจอจัดการสิทธิ์การใช้งาน 47
ภาพที่ 4.9 หน้าจอจัดการข้อมูลระบบ 48
ภาพที่ 4.10 หน้าจอการพยากรณ์ผลผลิตมันสำปะหลัง 49
ภาพที่ 4.11 หน้าจอผลการพยากรณ์ผลผลิตมันสำปะหลัง 40
ภาพที่ 4.12 หน้าจอค้นหาประวัติการปลูก 50
ภาพที่ 4.13 หน้าจอผลค้นหาประวัติการปลูก 50
ภาพที่ 4.14 หน้าจอรายงานปริมาณผลผลิตมันสำปะหลัง 51
ภาพที่ 4.15 หน้าจอรายงานปริมาณการใช้ปุ๋ย 51
ภาพที่ 4.16 หน้าจอแบบประเมินความพึงพอใจการใช้งานระบบ 52

บทที่ 1
บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

มันสำปะหลังเป็นพืชอาหารที่สำคัญเป็นอันดับ 5 ของโลกรองจากข้าวสาลี ข้าวโพด ข้าว และมันฝรั่ง เป็นพืชอาหารที่สำคัญของประเทศในเขตร้อน โดยเฉพาะประเทศต่างๆ ในทวีปอัฟริกา และทวีปอเมริกาใต้ ในทวีปเอเชียประเทศอินโดนีเซียและอินเดียมีการบริโภคมันสำปะหลังกันเป็นจำนวนมาก ปริมาณผลผลิตที่ได้ในแต่ละปีร้อยละ 60 ใช้เป็นอาหารของมนุษย์ ร้อยละ 27.5 ใช้ทำเป็นอาหารสัตว์ และร้อยละ 12.5 ใช้ประโยชน์ในด้านอื่นๆ มันสำปะหลังเป็นพืชที่ทำรายได้ให้เกษตรกรมากเป็นอันดับที่ 4 รองจากยางพารา อ้อย และข้าว ผลผลิตมันสำปะหลังภายในประเทศนำไปใช้ทำมันเส้นและมันอัดเม็ดร้อยละ 45-50 ใช้แปรรูปเป็นแป้งร้อยละ 50-55

สำหรับประเทศไทยเป็นประเทศที่ส่งผลิตภัณฑ์มันสำปะหลังออกมากที่สุดในโลก ประเทศที่ไทย ส่งผลิตภัณฑ์มันสำปะหลังในรูปของมันอัดเม็ดไปขายมากที่สุดคือ ประเทศในกลุ่มประชาคมยุโรป (เนเธอร์แลนด์ สเปน เยอรมัน โปรตุเกส) เกาหลีใต้และญี่ปุ่น ส่วนในรูปของแป้งมันสำปะหลัง ประเทศญี่ปุ่นสั่งซื้อ มากที่สุด รองลงมาคือฮ่องกง สหรัฐอเมริกา มาเลเซีย สิงคโปร์ และไต้หวัน (กรมวิชาการเกษตร, 2555) จากที่กล่าวมาข้างต้นจะเห็นว่ามันสำปะหลังเป็นพืชเกษตรที่มีความสำคัญของประเทศไทย โดยในปัจจุบันการปลูกมันสำปะหลังต้องใช้ปัจจัยการผลิตที่สูงขึ้นมากไม่ว่าจะค่าใช้จ่ายในส่วนของปุ๋ย สารเคมีกำจัดวัชพืชและแมลงศัตรูพืช ค่าจ้างแรงงาน ค่าขนส่ง ค่าน้ำมัน เป็นต้น อีกทั้งผลผลิตที่ได้ยังมีปริมาณผลผลิตต่อไร่ลดลง และราคาขายก็ต่ำอีกด้วย ดังนั้นหากมีการพยาการณ์หรือคาดเดาผลผลิตมันสำปะหลังได้ล่วงหน้า จะเป็นประโยชน์อย่างยิ่ง ทำให้ทราบถึงปัจจัยการผลิตที่เกี่ยวข้อง อาทิเช่น พันธุ์ที่ปลูก อายุปลูกถึงการเก็บ การใช้ปุ๋ย การใช้สารเคมีกำจัดวัชพืชและแมลงศัตรูพืช เป็นต้น ส่งผลทำให้สามารถวางแผนบริหารจัดการเพื่อลดต้นทุนในการปลูกลงได้

วิทยานิพนธ์นี้จึงได้เสนอการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูลโดยการจำแนกประเภท (classification) ซึ่งใช้อัลกอริธึมจำนวน 5 ตัว ได้แก่ J48, Random Tree, Simple Cart, Naïve Bayes, และ LAD Tree เพื่อสร้างต้นไม้ตัดสินใจ (Decision Tree) จากนั้นทำการทดสอบความแม่นยำในการพยากรณ์ แล้วคัดเลือกอัลกอริธึมที่ให้ความแม่นยำดีที่สุดไปใช้ในการออกแบบและพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูลและการประเมินความพึงพอใจผ่านเว็บไซต์ ต่อไป

2. วัตถุประสงค์ของการวิจัย

2.1 เพื่อศึกษาและพัฒนาตัวแบบพยากรณ์ผลผลิตมันสำปะหลังโดยใช้เทคนิคการทำเหมืองข้อมูล

2.2 เพื่อพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจของผู้ใช้งานระบบผ่านเว็บไซต์

3. ขอบเขตของงานวิจัย

3.1 ขอบเขตข้อมูล จะใช้ข้อมูลปัจจัยผลผลิตมันสำปะหลัง ตั้งแต่ปี 2551-2555 จากสำนักงานเกษตรจังหวัดกำแพงเพชร ซึ่งประกอบด้วยข้อมูลปัจจัยการผลิตมันสำปะหลังของจังหวัดในภาคเหนือที่มีการปลูกมันสำปะหลังมากที่สุด และมีสภาพภูมิอากาศที่ใกล้เคียงกับจังหวัดกำแพงเพชร จำนวน 5 จังหวัด ได้แก่ กำแพงเพชร นครสวรรค์ เพชรบูรณ์ พิษณุโลก และอุทัยธานี รวม 325 ครัวเรือน และกลุ่มตัวอย่างในการประเมินความพึงพอใจในการใช้งานระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง ประกอบด้วย เจ้าหน้าที่สำนักงานเกษตรจังหวัดกำแพงเพชร ผู้ใช้งานทั่วไป และผู้ดูแลระบบ รวม 30 คน

3.2 ขอบเขตด้านเทคโนโลยี จะใช้ซอฟต์แวร์โอเพนซอร์สเวกา (WEKA) ในการสร้างตัวแบบพยากรณ์โดยใช้การจำแนกประเภท (classification) และใช้อัลกอริธึมจำนวน 5 ตัว ได้แก่ J48, Random Tree, Simple Cart, Naïve Bayes, และ LAD Tree เพื่อสร้างต้นไม้ตัดสินใจ (Decision Tree) จัดทำระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ โดยใช้ภาษา PHP ในการพัฒนาเว็บไซต์ ใช้โปรแกรม My SQL ในการจัดการฐานข้อมูล และใช้โปรแกรม Apache เป็นเว็บเซิร์ฟเวอร์

4. ประโยชน์ที่คาดว่าจะได้รับ

4.1 ทำให้ได้ทราบถึงปัจจัยที่มีผลต่อการปลูกมันสำปะหลัง

4.2 ทำให้ได้ระบบสารสนเทศพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์

4.3 สามารถนำไปใช้ในการวางแผนบริหารจัดการเพื่อลดต้นทุนในการปลูกมันสำปะหลังลงได้

บทที่ 2
วรรณกรรมที่เกี่ยวข้อง

1. การปลูกมันสำปะหลัง

มันสำปะหลังจัดเป็นพืชหัวชนิดหนึ่งมีชื่อวิทยาศาสตร์ คือ Manihot esculenta Crantz มีชื่อสามัญเรียกหลายชื่อตามภาษาต่างๆ ที่ได้ยินกันมากได้แก่ Cassava, Yuca, Mandioa, Manioc, Tapioca มันสำปะหลังมีแหล่งกำเนิดแถบที่ลุ่มเขตร้อน (Lowland tropics) มีหลักฐานแสดงว่าปลูกกันในโคลัมเบียและเวเนซูเอลา มานานกว่า 3,000 – 7,000 ปีมาแล้วสันนิษฐานว่าแหล่งกำเนิดมันสำปะหลังมี 4 แหล่งด้วยกันคือ แถบประเทศกัวเตมาลาและเม็กซิโก ทางทิศตะวันตกเฉียงเหนือของทวีปอเมริกาใต้ ทางทิศตะวันออกของประเทศโบลิเวียและทางทิศตะวันตกเฉียงเหนือของประเทศอาร์เจนตินา และทางทิศตะวันออกของประเทศบราซิล

ในทวีปเอเชียมีการนำมันสำปะหลังมาปลูกครั้งแรกที่ประเทศฟิลิปปินส์ในคริสต์ศตวรรษที่ 17 โดยชาวสเปนได้นำมาจากเม็กซิโกและในเวลาต่อมาก็มีการปลูกที่ อินโดนีเซีย และเมื่อ พ.ศ. 2337 ได้มีการนำมันสำปะหลังจาก อัฟริกามาปลูกที่อินเดียเพื่อใช้ในการทดลองสำหรับประเทศไทยไม่มีหลักฐานที่แน่นอนว่ามีการนำมันสำปะหลังเข้ามาปลูกเมื่อใดคาดว่าคงเข้ามาในระยะเดียวกันกับการเข้าสู่ศรีลังกาและฟิลิปปินส์คือประมาณ พ.ศ. 2329–2383 มันสำปะหลังเดิมเรียกกันว่ามันสำโรง มันไม้ ทางภาคตะวันออกเฉียงเหนือเรียกว่ามันต้นเตี้ย ทางภาคใต้เรียกว่ามันเทศ (แต่เรียกมันเทศว่ามันหลา) มันสำปะหลังเป็นพืชอาหารที่สำคัญเป็นอันดับ 5 ของโลกรองจากข้าวสาลี ข้าวโพด ข้าว และมันฝรั่ง เป็นพืชอาหารที่สำคัญของประเทศในเขตร้อน โดยเฉพาะประเทศต่างๆ ในทวีปอัฟริกาและทวีปอเมริกาใต้ ในทวีปเอเชียประเทศอินโดนีเซียและอินเดียมีการบริโภคมันสำปะหลังกันเป็นจำนวนมาก ปริมาณผลผลิตที่ได้ในแต่ละปีร้อยละ 60 ใช้เป็นอาหารของมนุษย์ ร้อยละ 27.5 ใช้ทำเป็นอาหารสัตว์ และร้อยละ 12.5 ใช้ประโยชน์ในด้านอื่นๆ มันสำปะหลังเป็นพืชที่ทำรายได้ให้เกษตรกรมากเป็นอันดับที่ 4 รองจากยางพารา อ้อย และข้าว ผลผลิตมันสำปะหลัง ภายในประเทศนำไปใช้ทำมันเส้นและมันอัดเม็ดร้อยละ 45-50 ใช้แปรรูปเป็นแป้งร้อยละ 50-55

ประเทศไทยเป็นประเทศที่ส่งผลิตภัณฑ์มันสำปะหลังออกมากที่สุดในโลก ประเทศที่ไทยส่งผลิตภัณฑ์มันสำปะหลังในรูปของมันอัดเม็ดไปขายมากที่สุดคือ ประเทศในกลุ่มประชาคมยุโรป (เนเธอร์แลนด์ สเปน เยอรมัน โปรตุเกส) เกาหลีใต้และญี่ปุ่น ส่วนในรูปของแป้งมันสำปะหลังประเทศญี่ปุ่นสั่งซื้อ มากที่สุด รองลงมาคือฮ่องกง สหรัฐอเมริกา มาเลเซีย สิงคโปร์ และไต้หวัน (กรมวิชาการเกษตร, 2555)

1.1 พันธุ์มันสำปะหลัง

1.1.1 พันธุ์ระยอง 5 ได้จากการผสมพันธุ์และคัดเลือกที่ศูนย์วิจัยพืชไร่ระยอง เมื่อปี พ.ศ.2525 ระหว่างพันธุ์ 27-77-10 กับพันธุ์ ระยอง 3 นำเมล็ดมาปลูกคัดเลือก เปรียบเทียบเบื้องต้น เปรียบเทียบมาตรฐาน เปรียบเทียบในท้องถิ่น เปรียบเทียบในไร่เกษตรกร และนำมาทดสอบ

ในไร่เกษตรกรได้พันธุ์ดี ลำต้นสีเขียวอมน้ำตาล สูงประมาณ 170 เซนติเมตร มีระดับการแตกกิ่ง 2-3 ระดับความสูงของการแตกกิ่งระดับแรก 100-120 เซนติเมตร มุมของกิ่ง 15-30 องศา แผ่นใบมีรูปร่างเป็นแบบใบหอก ใบแก่สีเขียวเข้ม ก้านใบสีแดงเข้ม ยอดอ่อนสีม่วงอมน้ำตาล หัวรูปร่างป้อมอ้วน เปลือกสีน้ำตาลอ่อน เนื้อสีขาว ผลผลิตหัวสดสูง 4,420 กิโลกรัมต่อไร่ ซึ่งสูงกว่าพันธุ์ ระยอง 1 23 เปอร์เซ็นต์ ให้ผลผลิตแป้งสูง 1,027 กิโลกรัมต่อไร่ ซึ่งสูงกว่าพันธุ์ระยอง 1 ระยอง 3 เท่ากับ 44 และ 35 เปอร์เซ็นต์ ตามลำดับ ให้ผลผลิตมันแห้งสูง 1,554 กิโลกรัมต่อไร่ สูงกว่าพันธุ์ระยอง 1 ระยอง 3 เท่ากับ 23 และ 37 เปอร์เซ็นต์ ตามลำดับ ปรับตัวเข้ากับสภาพแวดล้อมได้ดีมีความงอกดีและอยู่รอดถึงการเก็บเกี่ยวมากถึง 93 เปอร์เซ็นต์ ต้านทานโรคใบไหม้ปานกลาง เป็นโรคใบไหม้ได้ง่ายกว่าพันธุ์อื่น ๆ แต่อาการไม่รุนแรงถึงกับทำให้ต้นตาย

1.1.2 พันธุ์ระยอง 60 ปี พ.ศ.2524 ศูนย์วิจัยพืชไร่ระยอง ได้ทำการผสมพันธุ์มันสำปะหลังระหว่างพันธุ์ Mcol 1684 กับพันธุ์ ระยอง 1 เพื่ออายุเก็บเกี่ยวสั้น ต้นฝนปลายฝนจำนวน 89 คู่ผสม นำเมล็ดที่ได้ไปเพาะเพื่อปลูกคัดเลือกพันธุ์ได้จำนวน 6,839 ต้น และได้สายพันธุ์ CMR 24-63-43 ซึ่งให้ผลผลิตแป้งสูง ยอดอ่อนและใบแรกเจริญเติบโตเต็มที่มีสีเขียวปนม่วง ก้านใบสีเขียวปนม่วง ยาวประมาณ 25-30 เซนติเมตร ใบมีลักษณะแบบใบหอก ลำต้นสีน้ำตาลอ่อน เนื้อในสีขาวครีม ความสูงของต้นประมาณ 275 เซนติเมตร แตกกิ่งสูงจากพื้นดินประมาณ 170 เซนติเมตร กิ่งทำมุมกับลำต้นประมาณ 45-60 องศา อายุเก็บเกี่ยว 8-12 เดือน ขยายพันธุ์ด้วยลำต้นให้ผลผลิตดีในภาคตะวันออก เก็บเกี่ยวอายุ 8 เดือน ผลผลิตหัวสด ผลผลิตแป้ง ผลผลิตมันเส้น ผลผลิตต่อวันและค่าดัชนีการเก็บเกี่ยวสูงกว่าพันธุ์ ระยอง 1 24.5, 31.3, 41.9, 24.8 และ 21.8 เปอร์เซ็นต์ตามลำดับ โดยมีปริมาณกรดไซยานิค 3 ส่วนในล้าน (ppm) มีความต้านทานต่อโรคใบไหม้ปานกลางผลผลิตหัวสดเมื่ออายุ 8 เดือน 3,148 กิโลกรัมต่อไร่ (มันแห้ง 1,217 กิโลกรัมต่อไร่) และผลผลิตหัวสด เมื่ออายุ 12 เดือน 4,224 กิโลกรัมต่อไร่ (มันแห้ง 1,404 กิโลกรัมต่อไร่) เก็บเกี่ยวในฤดูฝน เปอร์เซ็นต์แป้งลดลง 20 เปอร์เซ็นต์ และเนื้อมีสีครีมบางครั้งทำให้โรงงานตัดราคา

1.1.3 พันธุ์ระยอง 7 ได้จากการผสมข้ามพันธุ์ระหว่างพันธุ์ CMR30-71-25 กับพันธุ์ OMR29-20-118 ตั้งแต่ปี 2535 ที่ศูนย์วิจัยพืชไร่ระยอง ภายใต้ชื่อ CMR35-48-196 โดยผ่านการคัดเลือกและประเมินพันธุ์ที่ศูนย์วิจัยพืชไร่ สถานีทดลองพืชไร่ และไร่เกษตรกรในแหล่งปลูกมันสำปะหลังที่สำคัญของประเทศไทย รวม 13 จังหวัด แปลงทดลอง 51 แปลง ทำการคัดเลือกและประเมินพันธุ์ตามขั้นตอนการปรับปรุงพันธุ์ ตั้งแต่ปี พ.ศ. 2535-2548 รวมระยะเวลาการวิจัย 14 ปี ลำต้นมีสีน้ำตาลอ่อนและใบสีเขียวอ่อนคล้ายกับพันธุ์ระยอง 90 ลำต้นตั้งตรง ไม่โค้งงอ ไม่มีการแตกกิ่ง มีจำนวนลำต้นที่แตกจากท่อนปลูกประมาณ 2-4 ลำต้นต่อหลุมซึ่งมากกว่าพันธุ์อื่นทั่วไป และเนื้อของหัวมันสำปะหลังมีสีขาว ให้ผลผลิตหัวสด 6.08 ตันต่อไร่ ปริมาณแป้งในหัวสด 27.7 เปอร์เซ็นต์ ผลผลิตแป้ง 1.71 ตันต่อไร่ และผลผลิตมันเส้น 2.35 ตันต่อไร่ ผลผลิตและปริมาณแป้งในหัวสดเมื่อปลูกทั้งต้นฤดูฝนและปลายฤดูฝนสูงกว่าพันธุ์รับรองทุกพันธุ์ โดยให้ผลผลิตหัวสดเฉลี่ย 6.08 ตันต่อไร่ ปริมาณแป้งในหัวสด 27.7 เปอร์เซ็นต์ ผลผลิตแป้ง 1.71 ตันต่อไร่ และผลผลิตมันเส้น 2.35 ตันต่อไร่ งอกเร็วมากประมาณ 6 วันหลังจากปลูก ในขณะที่พันธุ์ทั่วไปใช้เวลางอก 14 วัน หลังปลูก ดังนั้นสามารถปลูกปลายฤดูฝนได้ดีเนื่องจากให้ความงอกเร็วกว่าทุกพันธุ์จึงเป็นโอกาสให้การเจริญเติบโตได้เร็วในช่วงระยะแรกของการเจริญเติบโตคือ ในช่วงอายุ 1-2 เดือนหลังปลูก ซึ่งในขณะนั้นดินยังมี

ความชื้นอยู่ และสามารถฟื้นตัวหรือเจริญเติบโตได้เร็วเมื่อได้รับน้ำฝนอีกครั้งหลังจากผ่านช่วงฤดูแล้งอันยาวนานด้วยเหตุผลดังกล่าวจึงเป็นพันธุ์ที่ใช้สำหรับปลูกทั้งปลายฤดูฝนและต้นฤดูฝนได้ดี จำนวนลำต้นที่แตกออกจากท่อนปลูก 2-4 ลำต้นต่อหลุม ทำให้คลุมวัชพืชได้ดีในช่วง 3 เดือนหลังจากปลูก นอกจากนี้ยังเป็นพันธุ์ที่มีทรงต้นดีไม่แตกกิ่ง เป็นผลทำให้ลำต้นไม่หักล้มสะดวกในการดูแลรักษาและเก็บเกี่ยว จำนวนหัวต่อต้นมากถึง 11 หัวต่อต้น มีขนาดของหัวใกล้เคียงกันโดยหัวออกรอบโคนลำต้นและเรียงกันเป็นชั้นและไม่มีก้านหัวคือหัวขาดยากเมื่อเก็บเกี่ยว ดังนั้น จึงเหมาะกับการใช้เครื่องขุดในการเก็บเกี่ยวผลผลิต 5.) เมื่อปลูกในสภาพดินที่มีความอุดมสมบูรณ์สูง จะให้ลำต้นแข็งแรง ไม่แตกกิ่งและไม่เกิดปัญหาหักล้มเมื่อเทียบกับพันธุ์รับรองทุกพันธุ์ เหมาะสมในสภาพดินร่วนเหนียวและดินร่วนทรายที่มีความชื้นของดินดีตลอดช่วงของการเจริญเติบโต โดยเฉพาะดินที่มีระดับน้ำใต้ดินอยู่ใกล้ผิวดินแต่ไม่เหมาะกับสภาพดินที่แห้งแล้ง

1.1.4 พันธุ์ระยอง 72 ได้จากการผสมพันธุ์ระหว่างพันธุ์ระยอง 1 กับพันธุ์ระยอง 5 ในปี พ.ศ.2533 ที่ศูนย์วิจัยพืชไร่ระยอง มีชื่อเดิมว่า CMR33-57-81 ดำเนินการตามขั้นตอนการปรับปรุงพันธุ์ โดยมีจำนวนแปลงทดลอง 91 แปลงทดลอง มีระยะเวลาการทดลอง 9 ปี ลำต้นสีเขียว สูง 200 เซนติเมตร มีระดับการแตกกิ่ง 0-1 ระดับ ความสูงของการแตกกิ่งระดับแรก 130-140 เซนติเมตร มุมของกิ่ง 60-75 องศา ใบแก่สีเขียวเข้ม ก้านใบสีแดงเข้ม ความยาวก้านใบ 25-30 เซนติเมตร ยอดอ่อนสีม่วง เปลือกนอกของหัวสีขาวนวล เนื้อสีขาว ผลผลิตหัวสดสูง 5.09 ตันต่อไร่ สูงกว่าพันธุ์มาตรฐาน คือพันธุ์ระยอง 1 ร้อยละ 27 ผลผลิตแป้งสูง 1.07 ตันต่อไร่ สูงกว่าพันธุ์ระยอง 1 ร้อยละ 36 ผลผลิตมันแห้งสูงถึง 1.71 ตันต่อไร่ สูงกว่าพันธุ์ระยอง 1 ร้อยละ31 ปรับตัวได้ดีในสภาพแวดล้อมทางภาคตะวันออกเฉียงเหนือ โดยให้ผลผลิตหัวสดสูงถึง 5.55 ตันต่อไร่ ผลผลิตแป้ง 1.23 ตันต่อไร่ และผลผลิตมันแห้ง 1.91 ตันต่อไร่ ซึ่งสูงกว่าการปลูกในท้องที่อื่น ท่อนพันธุ์มีความอยู่รอดถึงเก็บเกี่ยวสูง 92 เปอร์เซ็นต์ ทรงต้นดี แตกกิ่งบ้างเล็กน้อยในระดับที่สูงจากโคนต้น ทำให้สามารถขยายพันธุ์ได้มากขึ้น ต้านทานโรคใบจุดและโรคใบไหม้ปานกลาง

1.1.5 พันธุ์ระยอง 9 ได้จากการผสมข้ามระหว่างพันธุ์ที่มีเปอร์เซ็นต์แป้ง คือ พันธุ์ CMR31-19-23 และพันธุ์ OMR29-20-118 ศูนย์วิจัยพืชไร่ระยอง โดยดำเนินการคัดเลือกและประเมินศักยภาพของพันธุ์ในจังหวัดที่เป็นแหล่งผลิตมันสำปะหลังที่สำคัญรวม 14 จังหวัด ตั้งแต่ปี 2535-2542 แปลงทดลอง 38 แปลงทดลอง และผ่านการประเมินผลผลิตเอทานอลในระดับห้องปฏิบัติการ และระดับโรงงานต้นแบบ จากสถาบันวิจัยวิทยาศาสตร์และเทคโนโลยีแห่งประเทศไทย ในปี 2544-2547 ทำการคัดเลือกและประเมินพันธุ์ตามขั้นตอนการปรับปรุงพันธุ์ ตั้งแต่ปี 2535-2547 รวมระยะเวลาการวิจัย 13 ปี มีลำต้นสีน้ำตาลอ่อน ลำต้นสูงตั้งตรงไม่ค่อยแตกกิ่ง ลำต้นมีความสูง เฉลี่ย 235 เซนติเมตร หากมีการแตกกิ่งจะแตกกิ่งที่ระดับความสูง 160-180 เซนติเมตร กิ่งทำมุมแคบกับลำต้น ก้านใบสีเขียวอ่อนอมชมพู ยอดอ่อนสีเขียวอ่อนอมน้ำตาล ใบแก่สีเขียวอ่อน หัวสีน้ำตาลอ่อน เนื้อของหัวสีขาว ให้ผลผลิตแป้งและผลผลิตมันแห้งสูง 1.24 และ 2.11 ตันต่อไร่ ตามลำดับ ให้ผลผลิตเอทานอลสูงทุกอายุเก็บเกี่ยว โดยเก็บเกี่ยวเมื่ออายุ 8 12 และ 18 เดือน จะให้ผลผลิตเอทานอล 191 208 และ 194 ลิตร จากหัวสด 1 ตัน ซึ่งจะสูงกว่าพันธุ์ระยอง 90 ที่ให้เอทานอลเพียง 170 174 และ 155 ลิตร จากหัวสด 1 ตัน จากการทดสอบการผลิตในโรงงานต้นแบบ พบว่าการผลิตเอทานอล 1 ลิตร จะต้องใช้หัวมันสดพันธุ์ระยอง 9 ที่มีน้ำหนัก 5.05 กิโลกรัม เปรียบเทียบ

กับพันธุ์ระยอง 90 ที่ใช้หัวสด 5.68 กิโลกรัม จากการศึกษาจะเห็นได้ว่ามันสำปะหลังพันธุ์ระยอง 9 จะสามารถลดต้นทุนวัตถุดิบในการผลิตเอทานอลได้ลิตรละ 40 และ 84 สตางค์ เมื่อหัวมันสำปะหลังสดราคากิโลกรัมละ 1.00 บาท และ 1.70 บาท ตามลำดับ ทรงต้นดี สูงตั้งตรง ได้ต้นสำหรับขยายพันธุ์มาก อัตราการขยายพันธุ์สูงกว่า 1 : 8 ควรเก็บเกี่ยวเมื่ออายุประมาณ 1 ปี ถ้าเก็บเกี่ยวเร็วจะให้ผลผลิตหัวสดต่ำกว่าพันธุ์มาตรฐานอื่นๆ เนื่องจากพันธุ์ระยอง 9 มีเปอร์เซ็นต์แป้งสูงแต่สะสมน้ำหนักช้า

1.1.6 พันธุ์ระยอง 90 ได้จากการผสมพันธุ์ระหว่างพันธุ์ CMC76 และพันธุ์ V43 ในปีพ.ศ.2521 ที่สถานีทดลองพืชไร่ระยองปีพ.ศ.2521-2522 ทำการคัดเลือกครั้งที่ 1 และครั้งที่ 2 ปีพ.ศ.2523-2532 เปรียบเทียบพันธุ์เบื้องต้น เปรียบเทียบมาตรฐานพันธุ์ เปรียบเทียบพันธุ์ในท้องถิ่น เปรียบเทียบพันธุ์ในไร่กสิกรและทดสอบพันธุ์ในไร่เกษตรกรในสถานีทดลองและเกษตรกรจังหวัดระยอง จังหวัดชลบุรี และจังหวัดปราจีนบุรี ลำต้นมีลักษณะโค้ง สีน้ำตาลอ่อน สูงประมาณ 165 เซนติเมตร มีระดับการแตกกิ่ง 0-2 ระดับความสูงของการแตกกิ่งระดับแรก 120-14O เซนติเมตร มุมของกิ่งกว้าง 75-60 องศา แผ่นใบรูปร่างเป็นแบบใบหอก ใบแก่สีเขียวเข้ม ยอดอ่อนและก้านใบสีเขียวอ่อนหัวมันรูปร่างยาวเรียว เปลือกสีน้ำตาลเข้ม เนื้อสีขาวเนื้อมันแห้ง 36.4 เปอร์เซ็นต์ ผลผลิตหัวสดสูง เก็บเกี่ยวอายุ 12 เดือน ให้ผลผลิต 3.81 ตันต่อไร่ สูงกว่าพันธุ์ระยอง 1 ร้อยละ 5 มีเปอร์เซ็นต์แป้งสูงประมาณ 25 เปอร์เซ็นต์ เมื่อเก็บเกี่ยวในฤดูฝนผลผลิตมันแห้งสูง 1,400 กิโลกรัมต่อไร่ ผลผลิตแป้งสูง 966 กิโลกรัมต่อไร่ ต้านทานต่อโรคไหม้ ปลูกปลายฤดูฝน ในดินที่สูญเสียความชื้นง่าย อาจมีปัญหาจำนวนท่อนพันธุ์หรือต้นงอกต่ำกว่า 80 เปอร์เซ็นต์ ซึ่งจะทำให้ได้ผลผลิตต่ำไม่เหมาะสมกับท้องที่ที่พบการแพร่ระบาดของแมลงหวี่ขาวอยู่เสมอ ควรใส่ปุ๋ยตามคำแนะนำ สูตร 15-7-18 หรือ 15-15-15 อัตรา 50 กิโลกรัมต่อไร่ ลำต้นมีลักษณะโค้ง ถ้าหากแตกกิ่งจะทำให้ปฏิบัติดูแลยาก ในฤดูแล้งต้นพันธุ์สำหรับนำไปใช้ปลูกเสื่อมคุณภาพเร็ว เมื่อตัดต้นพันธุ์แล้วควรรีบปลูกภายในเวลาไม่เกิน 2 สัปดาห์ เพราะความงอกจะลดลง

1.1.7 พันธุ์เกษตรศาสตร์ 50 มีลักษณะเด่น งอกดี หัวดก และมีลักษณะเป็นกลุ่มมีปริมาณแป้งในหัวสูง ลักษณะประจำพันธุ์ ลำต้นโค้งเล็กน้อย สีเขียวเงิน สูง 180- 250 ซม. แตกกิ่งระดับแรกที่ความสูง 80-150 ซม. ฤดูปลูกที่เหมาะสม ต้นฤดูฝน เดือนพฤษภาคมถึงเดือนมิถุนายน ปลายฤดูฝน เดือนกันยายนถึงเดือนตุลาคม สามารถปลูกได้ทั่วประเทศ

1.1.8 พันธุ์ระยอง 3 ได้มาจากการผสมพันธุ์ระหว่างพันธุ์ Mmex 55 กับพันธุ์ Mven 307 เรียกชื่อคู่ผสมนี้ว่า CM 407 กองพืชไร่นำเมล็ดลูกผสมมาจากศูนย์วิจัยเกษตรเขตร้อนนานาชาติ (CIAT) ประเทศโคลัมเบีย เมื่อปีพ.ศ.2518 นำมาปลูกคัดเลือกตามขั้นตอนการปรับปรุงพันธุ์ที่สถานีทดลองพืชไร่ระยองได้ต้นที่ค่อนข้างเตี้ยเก็บเกี่ยวผลผลิตง่าย ยอดและใบแรกที่เจริญเต็มที่สีเขียวอ่อน ก้านใบสีเขียวอ่อนปนแดง ใบแหลมแบบใบหอก ลำต้น หัวเปลือกมีสีน้ำตาลอ่อนเนื้อในสีขาวความสูงของต้นประมาณ 73 เซนติเมตร ลักษณะการเกิดของหัวจะรวมกันแน่นอายุเก็บเกี่ยวประมาณ 2 เดือน ขยายพันธุ์ด้วยลำต้น ผลผลิตหัวสด 3,899 กิโลกรัมต่อไร่ ผลผลิตแป้งสูงถึง 914 กิโลกรัมต่อไร่ สูงกว่าพันธุ์ระยอง 1 19.8 เปอร์เซ็นต์ เหมาะสำหรับอุตสาหกรรมทำแป้งและอาหารสัตว์ หัวสดมีแป้งสูง 23.4 เปอร์เซ็นต์ ซึ่งสูงกว่าพันธุ์ ระยอง 1 ซึ่งมีแป้งเพียง 18.3 เปอร์เซ็นต์ ให้ผลผลิต

มันเส้นหรือมันแห้ง สูงถึง 1,489 กิโลกรัมต่อไร่ ซึ่งสูงกว่าพันธุ์ระยอง 1 ถึง 11.5 เปอร์เซ็นต์ ต้านทานต่อโรคใบไหม้ปานกลาง ไม่ควรปลูกช่วงฝนตกหนักหรือแล้งจัด เพราะอัตราการตายสูง และให้ผลผลิตต่ำ

1.2 โรคมันสำปะหลัง

1.2.1 โรคในมันสำปะหลัง ได้แก่ โรคใบไหม้ (Cassava Bacterial Blight: CBB) โรคใบจุดสีน้ำตาล (Brown Leaf Spot) โรคใบจุดไหม้ (Blight Leaf Spot) โรคใบจุดขาว (White Leaf Spot) โรคลำต้นเน่าที่เกิดจากเชื้อรา (Stem Rot) โรคที่เกิดจากเชื้อรา Botryodiplodia theobromae โรคขี้เถ้าหรือราแป้ง (Cassava Ash Disease) โรคแอนแทรคโนส (Anthracnose) และโรครากหรือหัวเน่า (Root and Tuber Rot Diseases)

1.2.2 แมลงศัตรูของมันสำปะหลัง ได้แก่ ไรแดง เพลี้ยแป้งลาย แมลงหวี่ขาว และแมลงศัตรูประเภทปากกัด

1.3 การป้องกันกำจัดวัชพืช จะทำการไถ 1 ครั้ง ตากดินไว้ 7–10 วัน พรวน 1 ครั้ง แล้วคราดเก็บเศษซาก ราก เหง้า หัว และไหลของวัชพืชข้ามปี ออกจากแปลง กำจัดวัชพืชไม่น้อยกว่า 2 ครั้ง ตลอดฤดูปลูก คือ ครั้งแรกพ่นสารกำจัดวัชพืชทันทีหลังปลูกก่อนวัชพืชงอก หรือใช้จอบเครื่องกลขนาดเล็ก หรือแรงงานสัตว์ เพื่อกำจัดวัชพืชระหว่างแถวปลูก เมื่อมันสำปะหลังอายุ 1 – 2 เดือน ก่อนใส่ปุ๋ย ครั้งที่สองใช้จอบดาย หรือพ่นสารกำจัดวัชพืชอีกครั้ง ถ้ามีวัชพืชฤดูเดียวประเภทใบแคบมากกว่า 50 เปอร์เซ็นต์ของพื้นที่

แมลงศัตรูธรรมชาติ ได้แก่ แมลงห้ำ มี 4 ชนิด ซึ่งทั้งระยะหนอนและตัวเต็มวัยเป็นตัวห้ำ ยกเว้นแมลงช้างปีกใส เฉพาะระยะหนอนเท่านั้นที่เป็นตัวห้ำ ด้วงเต่าสีดำ ตัวเต็มวัยยาวประมาณ 1 มิลลิเมตร รูปร่างกลม หัวและท้ายเล็กกว่าส่วนลำตัวเล็กน้อย ปีกสีน้ำตาลเป็นมัน เป็นตัวห้ำของไรแดงและเพลี้ยแป้งลาย ด้วงเต่าสีน้ำตาล หนอนมีลักษณะคล้ายเพลี้ยแป้ง หัวท้ายเรียว เคลื่อนไหวรวดเร็ว ตัวเต็มวัยยาวประมาณ 2 มิลลิเมตร หลังโค้งนูนสีน้ำตาล หัวสีน้ำตาลเข้ม ส่วนอกสีเหลืองทอง ปลายปีกมีรูปยาวรีสีเหลืองทอง เป็นตัวห้ำของเพลี้ยแป้งลาย ด้วงปีกสั้น ตัวเต็มวัยสีน้ำตาลดำ ยาวประมาณ 1 มิลลิเมตร รูปร่างยาวเรียว ปีกสั้นกว่าท้อง เป็นตัวห้ำของไรแดง แมลงช้างปีกใส ตัวเต็มวัยลำตัวเรียวยาว ปีกโค้งบางใสขนาดใหญ่และ ยาวกว่าลำตัว สีเขียวอ่อนหรือสีน้ำตาลอ่อน มีปีก 2 คู่ หนอนมีลำตัว เรียวยาว สีน้ำตาลอ่อน มีแถบสีน้ำตาลพาดผ่านลำตัว มีกรามคล้ายเขี้ยว และบางชนิดจะมีซากเหยื่อที่กินแล้วอยู่บนหลังเพื่อพรางตัว หนอนเป็นตัวห้ำของไรแดง เพลี้ยแป้งลาย และแมลงหวี่ขาว ไรตัวห้ำ ตัวเต็มวัยสีแดงเข้ม มี 8 ขา ตัวมันวาว วิ่งค่อนข้างเร็ว ขายาวกว่าไรแดง เพศผู้มีขนาดเล็กกว่าเพศเมีย ไข่มีสีขาว รูปยาวรี ตัวอ่อนมีสีขาวและสีเหลือง เป็นไรตัวห้ำที่สำคัญของไรแดง แตนเบียนเพลี้ยแป้ง เป็นแมลงเบียนขนาดเล็กมาก เข้าทำลายตัวอ่อน วัยสุดท้ายของเพลี้ยแป้งลาย โดยแตนเบียนจะวางไข่เข้าไปในตัวอ่อนของ เพลี้ยแป้งลาย ทำให้เพลี้ยแป้งลายที่ถูกทำลาย มีลักษณะเป็นมัมมี่แข็งตายติดที่ผิวใบพืช

1.4 การปลูก

จะปลูกในช่วงต้นฤดูฝน เดือนเมษายน–มิถุนายน หรือปลายฤดูฝน เดือนกันยายน–พฤศจิกายน การเตรียมดินให้โล่งกลบวัชพืชและเศษใบ-ต้น มันสำปะหลังส่วนที่ไม่ได้ใช้ทำพันธุ์ พื้นที่ดอนไม่จำเป็นต้องยกร่อง ส่วนพื้นที่ต่ำอาจมีน้ำขังได้บ้าง จึงควรยกร่องปลูก สำหรับพื้นที่ลาดเอียงควรยกร่องปลูกขวางแนวลาดเอียง

การเตรียมท่อนพันธุ์ ใช้ท่อนพันธุ์จากต้นที่มีอายุ 8-12 เดือน เพราะท่อนพันธุ์จากลำต้นเจริญเติบโตและอยู่รอดดีกว่าท่อนพันธุ์จากกิ่ง ท่อนพันธุ์ใหม่ สด ไม่บอบช้ำ และไม่มีโรคแมลงทำลาย ตัดท่อนพันธุ์ยาวประมาณ 20 เซนติเมตรสำหรับปลูกในฤดูฝน หรือ 25 เซนติเมตรสำหรับปลูกในช่วงปลายฝน และมีจำนวนตาอย่างน้อย 5-10 ตาต่อท่อนพันธุ์

วิธีปลูก ปลูกแบบปักท่อนพันธุ์ตั้งหรือเอียง โดยในฤดูฝนควรปักให้ลึก 5- 10 เซนติเมตร ในช่วงปลายฤดูฝนควรปักให้ลึก 10- 15 เซนติเมตร ในพื้นที่ต่ำหรือพื้นที่ที่มีความลาดเอียงควรปลูกมันสำปะหลังบนสันร่อง

ระยะปลูก พื้นที่ราบใช้ระยะปลูกระหว่างแถว 80-100 เซนติเมตรระหว่างต้น 80-100 เซนติเมตร ซึ่งมีจำนวนต้น 1,600-2,500 ต้นต่อไร่ พื้นที่ลาดเอียงใช้ระยะปลูกระหว่างร่อง 80 เซนติเมตร ระหว่างต้น 80 เซนติเมตรเพื่อช่วยลดปัญหาการชะล้างพังทลายของดิน

การใส่ปุ๋ย แนะนำให้ใช้ปุ๋ยเคมีที่มีอัตราส่วน N:P:K 2:1:2 ในทางปฏิบัติ แนะนำให้ใส่ปุ๋ยสูตร 15-15-15 อัตรา 25 กิโลกรัมต่อไร่ ผสมกับปุ๋ยยูเรียและโปรแตสเซียมคลอไรด์อย่างละ 10 กิโลกรัมต่อไร่ หรืออาจใช้ปุ๋ยเคมีที่มีอัตราส่วนปุ๋ยใกล้เคียง เช่น สูตร 15-7-18 ใส่อัตรา 50 กิโลกรัมต่อไร่ ใส่เพียงครั้งเดียวเมื่อมันสำปะหลังอายุ 1-2 เดือนในขณะที่ดินมีความชื้น เพียงพอ โดยขุดหลุมใส่ 2 ข้างต้นระยะพุ่มใบแล้วกลบดิน

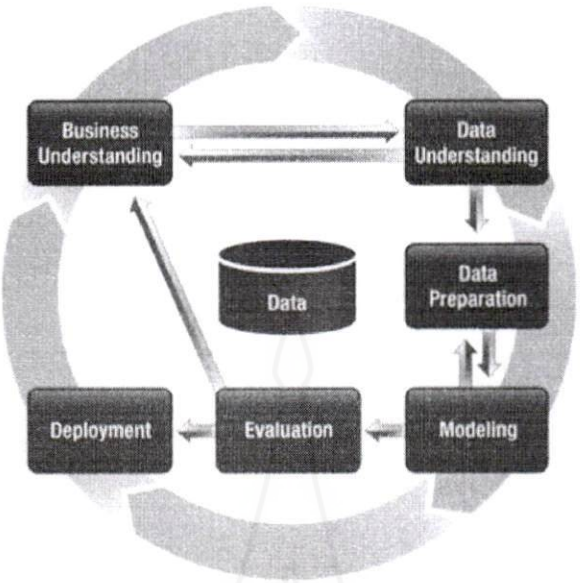
การเก็บเกี่ยว มันสำปะหลังเป็นพืชที่ไม่จำกัดอายุการเก็บเกี่ยวแต่ควรเก็บเกี่ยวเมื่ออายุครบ 8 เดือนขึ้นไป อายุเก็บเกี่ยวที่เหมาะสมคือ 12 เดือน หลังจากเก็บเกี่ยวแล้วควรรีบส่งโรงงานโดยเร็ว ไม่ควรทิ้งเกิน 4 วันเพราะหัวมันจะเน่าเสีย (กรมวิชาการเกษตร ระบบข้อมูลทางวิชาการ, 2555)

2. ความหมายของการทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining) คือ การสำรวจและวิเคราะห์ข้อมูลที่มีขนาดใหญ่เพื่อค้นหารูปแบบหรือกฎที่ซ่อนอยู่ในข้อมูลขนาดใหญ่นั้น และนำความรู้ที่ค้นพบได้นั้นไปใช้ให้เกิดประโยชน์เพื่อพัฒนาองค์กร เช่น ด้านการตลาด ลูกค้าสัมพันธ์ การแพทย์ อุตสาหกรรม การเกษตร การศึกษา เป็นต้น (Berry and Linnoff 2004: 7) ซึ่งต้องใช้ความรู้จากศาสตร์หลายแขนง ได้แก่ ฐานข้อมูล (Database technology) เป็นแหล่งรวบร่วมความรู้ สถิติ (Statistics) สำหรับวิเคราะห์ข้อมูลเบื้องต้น การเรียนรู้ของเครื่อง (Machine learning) เป็นการใช้อัลกอริทึมเพื่อค้นหารูปแบบและกฎที่ซ่อนอยู่ การมองเห็น (Visualization) เป็นการแสดงผลลัพธ์ในรูปแบบและความสัมพันธ์เพื่อให้เข้าใจได้ง่ายขึ้น และความรู้ด้านวิทยาศาสตร์สารสนเทศ (Information science)

ภาพที่ 2.1 แหล่งความรู้จากศาสตร์ต่างๆ ที่นำมาทำเหมืองข้อมูล

2.1 ขั้นตอนการทำเหมืองข้อมูล การวิเคราะห์ข้อมูลด้วยเทคนิค data mining ได้มีการกำหนดมาตรฐานในปี ค.ศ.1996 โดยบริษัท DaimlerChrysler บริษัท SPSS และบริษัท NCR ได้ร่วมกันตั้งมาตรฐานในการวิเคราะห์ข้อมูล ที่เรียกว่า "Cross-Industry Standard Process for Data Mining" หรือเรียกย่อๆ ว่า "CRISP-DM" โดยในกระบวนการ CRISP-DM นี้จะประกอบด้วย 6 ขั้นตอนดังแสดงในภาพที่ 2.2 ซึ่งในแต่ละขั้นตอนจะเป็นการต่อเนื่องกันคือขั้นตอนถัดไปจะรอผลลัพธ์จากขั้นตอนก่อนหน้าซึ่งแสดงด้วยลูกศรที่เชื่อมระหว่างกล่องสี่เหลี่ยมแต่ละกล่อง ตัวอย่างเช่น เมื่อได้ผลลัพธ์จากขั้นตอนการเตรียมข้อมูล (Data Preparation) แล้วจะนำไปสร้างตัวทำนาย (Modeling) และหลังจากสร้างตัวทำนายได้แล้วอาจจะย้อนกลับมาเปลี่ยนแปลงข้อมูลให้ถูกต้องมากขึ้นเพื่อหวังว่าจะได้ตัวทำนายที่ให้ความถูกต้องมากขึ้นก็ได้



ภาพที่ 2.2 ขั้นตอนการทำเหมืองข้อมูล

2.1.1 การทำความเข้าใจปัญหา (Problem Understanding) เป็นการศึกษาปัญหาที่เกิดขึ้นและแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูลทาง data mining พร้อมทั้งวางแผนในการดำเนินการ ดังนี้

1) Determine objective เป็นการตั้งเป้าหมายว่าการทำเหมืองข้อมูลครั้งนี้ต้องการแก้ปัญหาใด

2) Define success criteria เป็นเกณฑ์วัดความสำเร็จ ซึ่งอาจเป็นความสำเร็จในด้านรูปธรรม เช่น เพิ่มยอดขาดได้ 5% หรือในด้านนามธรรม เช่น การค้นพบความรู้ใหม่จากข้อมูล

3) Assess situation เป็นการประเมินสถานการณ์ในด้านต่างๆ เช่น ความรู้พื้นฐานในเรื่องการทำเหมืองข้อมูลมีเพียงพอหรือไม่ และผลประโยชน์จะคุ้มค่าหรือไม่

4) Determine data mining goals เป็นการตั้งเป้าหมายในเชิงการทำเหมืองข้อมูล ซึ่งต่างจากเป้าหมายหลักในการแก้ปัญหา เช่น เป้าหมายหลักคือต้องการเพิ่มยอดขายสินค้า เป้าหมายการทำเหมืองข้อมูลคือ การหาลักษณะของลูกค้าที่มีแนวโน้มที่จะซื้อสินค้า

5) Produce a project plan เป็นการวางแผนการทำเหมืองข้อมูลว่าจะเก็บข้อมูลอย่างไร และใช้ อัลกอริทึมอะไร

2.1.2 การทำความเข้าใจข้อมูล (Data Understanding) เป็นขั้นตอนในการเก็บรวบรวมข้อมูล หลังจากนั้นจะเป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาเพื่อพิจารณาความถูกต้องของข้อมูลว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์ ดังนี้

1) รวบรวมข้อมูลเริ่มต้น (Collect initial data)

2) กำหนดคุณสมบัติที่เก็บมา (Define success criteria)

3) อธิบายข้อมูล (Describe data)

4) สำรวจข้อมูล (Explore data)

5) ตรวจสอบความถูกต้องและความสมบูรณ์ของข้อมูล (Verify data quality)

2.1.3 การเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา (raw data) ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง (data cleaning) การแปลงข้อมูลให้อยู่ในช่วง (scale) เดียวกัน หรือการเติมข้อมูลที่ขาดหายไป ดังนี้

1) การคัดเลือกข้อมูลที่จะนำมาใช้ (Select data)

2) ทำความสะอาดข้อมูล (Clean data)

3) เตรียมข้อมูลให้เหมาะสมที่สุดเพื่อนำไปใช้ในขั้นตอนต่อไป

4) แก้ไขข้อมูลให้ถูกต้องสมบูรณ์ เช่น การแก้ไขค่าว่างของข้อมูลโดยใส่ค่า 0

5) ปรับเปลี่ยนข้อมูลให้มีค่าที่เหมาะสมในการตัดสินใจ เช่น ข้อมูลที่มีค่า "Coke" และ "Pepsi" อาจเปลี่ยนค่าเป็น "น้ำอัดลม" เป็นต้น

6) เลือกข้อมูลเฉพาะที่สนใจ เช่น ต้องการหาลักษณะลูกค้าที่ซื้อรถสปอรต์ ไม่ควรนำรายชื่อพนักงานเข้ามาเกี่ยวข้อง

7) คอลัมน์ที่มีค่าสำหรับทุกแถวเป็นค่าเดียว เช่น "สัญชาติไทย" หรือ คอลัมน์ที่มีค่าไม่ซ้ำกันเลย เช่น "หมายเลขสมาชิก" ไม่ควรนำมาใช้ เนื่องจากไม่สามารถบอกรูปแบบของข้อมูลได้

8) ปรับเปลี่ยนรูปแบบข้อมูล เช่น นำสองตารางในฐานข้อมูลมาเชื่อมต่อกัน (Transform data)

2.1.4 การสร้างตัวแบบ (Modeling) เป็นการวิเคราะห์ข้อมูลด้วยเทคนิค data mining โดยการสร้างตัวทำนาย (prediction model) ซึ่งในขั้นตอนนี้หลายๆ เทคนิคของ data mining จะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด ดังนั้นในบางครั้งอาจจะต้องมีการย้อนกลับไปที่ขั้นตอน data preparation เพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิคด้วย ดังนี้

1) เลือกอัลกอริทึมที่เหมาะสมในการทำเหมืองข้อมูล (Select modeling technique)

2) กำหนดรูปแบบการทดสอบผลลัพธ์ (Generate test design)

3) สร้างตัวแบบตามอัลกอริทึม (Build a model)

4) ทดสอบตัวแบบที่ได้มานั้นว่ามีความถูกต้องและน่าเชื่อถือเพียงใด (Assess the model)

2.1.5 การประเมินผล (Evaluation) เป็นวัดประสิทธิภาพของผลลัพธ์ที่ได้ ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรก หรือ มีความน่าเชื่อถือมากน้อยเพียงใด ซึ่งอาจจะย้อนกลับไปยังขั้นตอนก่อนหน้าเพื่อเปลี่ยนแปลงแก้ไขเพื่อให้ได้ผลลัพธ์ตามที่ต้องการได้

2.1.6 การนำไปใช้งาน (Deployment) เมื่อได้ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยเทคนิคทาง data mining แล้ว จะต้องนำองค์ความรู้นั้นไปใช้ได้จริงในองค์กรหรือบริษัท เพื่อพัฒนาธุรกิจให้ประสบความสำเร็จต่อไป

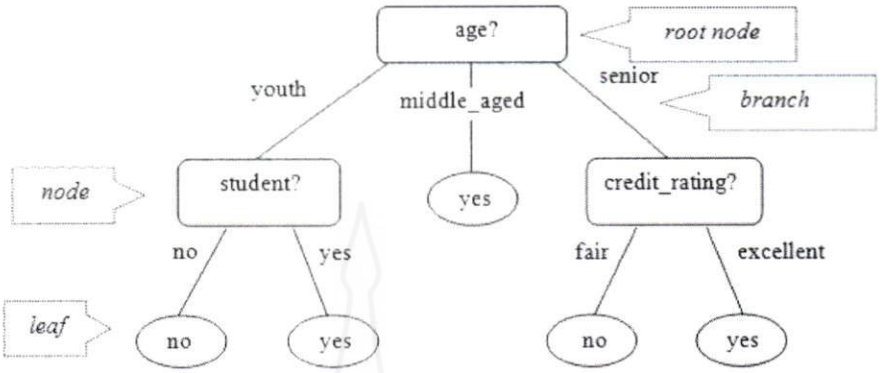
3. เทคนิคการทำเหมืองข้อมูล

การทำเหมืองข้อมูลเป็นเทคนิคในการค้นความรู้จากข้อมูลขนาดใหญ่ การทำเหมืองข้อมูล จึงเป็นการรวมเอาศาสตร์ต่างๆ หลายแขนงมารวมไว้ด้วยกันโดยไม่จำกัดวิธีการที่จะใช้ ตัวอย่างศาสตร์ที่ใช้ เช่น เทคโนโลยีฐานข้อมูล (Database technology) วิทยาศาสตร์สารสนเทศ (Information science) สถิติ (Statistics) และระบบการเรียนรู้ (Machine learning) เป็นต้นซึ่งศาสตร์ต่างๆ เหล่านี้จะทำให้เกิดกระบวนการค้นความรู้ในแบบต่างๆ ซึ่งเทคนิคที่นิยมใช้กันอย่างแพร่หลายในปัจจุบันมีดังนี้

3.1 กฎความสัมพันธ์ (Association Rules) เป็นการค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปใช้วิเคราะห์หรือทำนายปรากฏการณ์ต่างๆ หรือมาจากการวิเคราะห์การซื้อสินค้าของลูกค้าหรือที่เรียกว่า การวิเคราะห์ตะกร้าตลาด (market basket analysis) ซึ่งประเมินจากข้อมูลในตารางที่รวบรวมไว้ ผลการวิเคราะห์ที่ได้จะเป็นคำตอบของปัญหา ซึ่งการวิเคราะห์รูปแบบนี้เป็นการใช้กฎความสัมพันธ์ (association rule) เพื่อหาความสัมพันธ์ของข้อมูล ตัวอย่างเช่น ระบบแนะนำหนังสือให้กับลูกค้าแบบอัตโนมัติของ Amazon ข้อมูลสั่งซื้อทั้งหมดของ Amazon ซึ่งมีขนาดใหญ่มากจะถูกนำมาประมวลผลเพื่อหาความสัมพันธ์ของข้อมูลลูกค้า ลูกค้าที่ซื้อหนังสือเล่มหนึ่งๆ มักจะซื้อหนังสือเล่มใดพร้อมกันด้วยเสมอ ซึ่งความสัมพันธ์ที่ได้จากกระบวนการนี้จะสามารถนำไปใช้คาดเดาได้ว่าควรแนะนำหนังสือเล่มใดเพิ่มเติมให้กับลูกค้าที่เพิ่งซื้อหนังสือจากร้าน เช่น buys (x, database) -> buys (x, data mining) [80%, 60%] หมายความว่าเมื่อซื้อหนังสือฐานข้อมูล (database) แล้วมีโอกาสที่จะซื้อหนังสือการทำเหมืองข้อมูลด้วย 60% และมีการซื้อทั้งหนังสือฐานข้อมูลและหนังสือการทำเหมืองข้อมูลพร้อมๆ กัน 80%

3.2 การจำแนกกลุ่มและการทำนาย (Classification and Prediction) เป็นการทำเหมืองข้อมูลเพื่อจำแนกว่าข้อมูลที่ต้องการวิเคราะห์ว่าอยู่ในประเภท (class) ใด โดยอาศัยแบบจำลองหรือตัวแบบที่สร้างขึ้นมา การจำแนกประเภท (classification) จะใช้ข้อมูลที่มีอยู่จำนวนหนึ่ง หรือชุดข้อมูลเพื่อสร้างแบบจำลอง (training data set) หลังจากที่ได้แบบจำลองแล้วก็จะต้องทำการทดสอบแบบจำลองดังกล่าวด้วยข้อมูลอีกกลุ่ม (testing data set) เพื่อตรวจสอบความเที่ยงหรือความน่าเชื่อถือของแบบจำลองนั้น การทำเหมืองข้อมูลโดยวิธีจำแนกประเภท (classification) จะใช้สำหรับข้อมูลที่มีค่าไม่ต่อเนื่อง (discrete data) เช่น ข้อมูลที่ระบุเป็นผลการตัดสินใจว่า ใช่ หรือไม่ใช่ เสี่ยงหรือไม่เสี่ยง (วิภา เจริญภัณฑารักษ์, 2555: 20) ซึ่งตัวแบบที่ได้อาจแสดงในรูปแบบของ 2 ลักษณะ คือ แผนภาพต้นไม้เพื่อการตัดสินใจ (decision tree) และโครงข่ายประสาท (neural network)

3.2.1 แผนภาพต้นไม้เพื่อการตัดสินใจ (decision tree) เป็นขั้นตอนการสร้างตัวแบบ (model) ที่นิยมใช้กันมาก ทำความเข้าใจได้ง่าย โดยมีลักษณะเหมือนแผนภูมิองค์กร ภายในต้นไม้จะประกอบไปด้วยโหนด (node) ซึ่งแต่ละโหนดจะมีคุณลักษณะ (attribute) เป็นตัวทดสอบกิ่งของต้นไม้ (branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบและใบ (leaf) ซึ่งเป็นสิ่งที่อยู่ล่างสุดของต้นไม้ตัดสินใจแสดงถึงกลุ่มของข้อมูล (class) หรือนั่นก็คือผลลัพธ์ที่ได้จากการทำนาย โหนดที่อยู่บนสุดของต้นไม้เรียกว่าโหนดราก (root node) ดังแสดงโครงสร้างของต้นไม้ตัดสินใจตัดสินใจดังภาพที่ 2.3

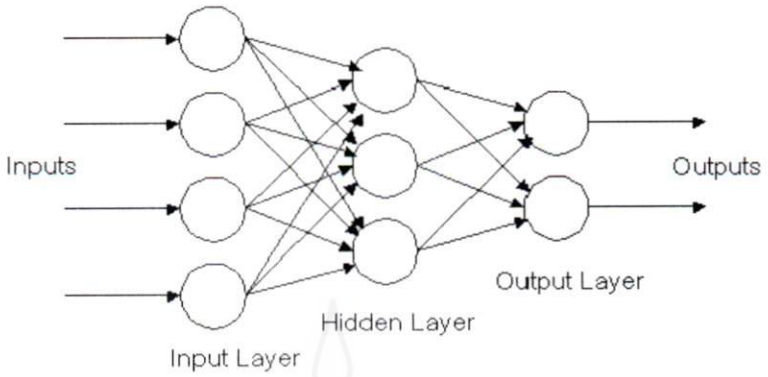


ภาพที่ 2.3 แผนภาพต้นไม้เพื่อการตัดสินใจ

ที่มา : http://scriptslines.com/blog/decision-tree/ Retrieved January 19, 2015

จากรูปเป็นต้นไม้ที่ใช้ในการตัดสินใจว่าจะเลือกซื้อคอมพิวเตอร์หรือไม่ คุณลักษณะที่พิจารณาคืออายุ (age) นักศึกษา (student) และอัตราเครดิต (credit_rating) โดยที่โหนดสี่เหลี่ยมมุมโค้งจะเป็นการทดสอบคุณลักษณะของข้อมูล ท้ายสุดจะได้ผลลัพธ์ของการทำนายว่าจะซื้อคอมพิวเตอร์ (yes) หรือไม่ซื้อคอมพิวเตอร์ (no) จากการทดสอบตามเส้นทางของต้นไม้ตัดสินใจตั้งแต่โหนดรากไปจนถึงใบ

3.2.2 โครงข่ายประสาท (Neural Network) เป็นเทคโนโลยีที่มีที่มาจากงานวิจัยด้านปัญญาประดิษฐ์ (Artificial Intelligence : AI) เพื่อใช้ในการคำนวณค่าฟังก์ชันจากกลุ่มข้อมูล วิธีการของเครือข่ายใยประสาทเป็นวิธีการที่ให้เครื่องเรียนรู้จากตัวอย่างต้นแบบแล้วฝึก (Train) ให้ระบบได้รู้จักที่จะคิดแก้ปัญหาที่กว้างขึ้นได้ ในโครงสร้างของเครือข่ายใยประสาทจะประกอบด้วยโหนด (Node) สำหรับ Input - Output และการประมวลผลกระจายอยู่ในโครงสร้างเป็นชั้นๆ ได้แก่ input layer , output layer และ hidden layers การประมวลผลของเครือข่ายใยประสาทจะอาศัยการส่งการทำงานผ่านโหนดต่างๆ ใน layer เหล่านี้ ตัวอย่างโครงสร้างแบบเครือข่ายใยประสาท



ภาพที่ 2.4 โครงสร้างเครือข่ายใยประสาท

ที่มา : http://www.no-poor.com/SAandNetwork/nu1.htm Retrieved January 19, 2015

3.3 **การจัดกลุ่มของข้อมูล (Clustering)** เป็นการดำเนินการกับข้อมูลที่ไม่ทราบรายละเอียดเกี่ยวกับลักษณะและไม่ทราบจำนวนกลุ่มที่แน่นนอน โดยจะดำเนินการร่วมส่วนต่างๆ ของข้อมูล ที่ต่างชนิดกันให้อยู่รวมกันเป็นกลุ่มย่อย หรือคลัสเตอร์ (Clusters) เดียวกัน จะดำเนินการหาตัวแทนแต่ละกลุ่มขึ้นมาเป็นตัวหลัก จากนั้นก็นำเอาข้อมูลมาเปรียบเทียบความเหมือนกับตัวแทนกลุ่มแต่ละกลุ่ม ถ้าเหมือนกลุ่มใด ก็จัดอยู่ในกลุ่มนั้น

4. ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคในการสร้างแบบจำลองวิธีหนึ่งที่นิยมใช้ในการพยากรณ์ (Prediction) หรือการจำแนกข้อมูล (Classification) ที่มีลักษณะการทำงานเหมือนโครงสร้างต้นไม้และมีการสร้างกฎต่างๆ เพื่อนำไปช่วยใช้ในการประกอบการตัดสินใจ ซึ่งต้นไม้ตัดสินใจเป็นแบบ จำลองที่ง่ายต่อการเข้าใจ และง่ายต่อการปรับเปลี่ยนเป็นกฎการจำแนก (Classification Rules) โดยทั่วไปอัลกอริทึมพื้นฐานของการสร้างต้นไม้ตัดสินใจ คือ อลักอริทึมละโมบ (Greedy Algorithm) โดยจะสร้างต้นไม้จากบนลงล่างแบบวนซ้ำ (Recursive) ด้วยวิธีการแบ่งปัญหาใหญ่เป็นปัญหาย่อย (Divide-and-Conquer) รูปแบบของต้นไม้ตัดสินใจจะประกอบด้วยโหนดแรกสุดที่เรียกว่าโหนดราก (Root node) จากโหนดรากจะแตกออกเป็นโหนดลูก และที่โหนดลูกก็จะมีลูกของตัวเอง ซึ่งโหนดในระดับสุดท้ายจะเรียกว่าใบ หรือลีฟโหนด (Leaf Node) แต่ละโหนดแสดงคุณลักษณะ (Attribute) ที่ใช้ทดสอบข้อมูลแต่ละกิ่งแสดงผลลัพธ์ในการทดสอบตามเงื่อนไขแ ละลีฟโหนดแสดงกลุ่มข้อมูลหรือคลาส (Class) ที่กำหนดไว้

ต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคที่ค่อนข้างแพร่หลาย เนื่องจากผู้ใช้สามารถทำ ความเข้าใจผลลัพธ์ได้ง่าย เทคนิคต้นไม้ตัดสินใจจะจำกัดข้อมูลที่เป็นตัวแปรตาม (Dependent Variable) 1 ตัวต่อ 1 แบบจำลองถ้าต้องการทำนายตัวแปรตามหลาย ๆ ตัวจะต้องสร้างแบบจำลองสำหรับตัวแปรตามแต่ละตัวอัลกอริทึมของเทคนิคแบบต้นไม้ตัดสินใจ ส่วนใหญ่ไม่รองรับแบบต่อเนื่อง (Continuous Data) จะต้องมีการแบ่งให้เป็นข้อมูลแบบไม่ต่อเนื่อง (Discrete Data)

เสียก่อน ตัวอย่างอัลกอริทึม เช่นClassification and Regression Trees (CART), ID3, C4.5 และ C5.0 algorithm เป็นต้น (สุมิตรา นวลมีศรี, 2555: 9)

4.1 อัลกอริทึม ID3 เป็นอัลกอริทึมในการสร้างต้นไม้ตัดสินใจ ที่ใช้หลักการของทฤษฎีข่าวสารค่าที่วัดได้จะนำมาใช้ตัดสินใจว่าจะใช้ตัวแปรใดในการแบ่งข้อมูลโดยวิธีการกำหนดโครงสร้างต้นไม้ตัดสินใจ จะเป็นการเลือกข้อมูลตามลำดับของตัวชี้วัดหรือค่าเกน (Gain) สูงที่สุดเป็นข้อมูลเริ่มต้นและข้อมูลถัดไปที่มีค่าลดหลั่นกันตามลำดับตัวอย่าง เช่น การพิจารณาจากกลุ่มข้อมูล 2 คลาส คือ P หรือ N โดยจำนวนตัวอย่างในคลาส P คือ p ตัวและจำนวนตัวอย่างในคลาส N คือ n ตัว ส่วนค่าของกลุ่มข้อมูลคือค่าคาดคะเนที่กลุ่มตัวอย่างต้องใช้จำนวนบิตในการแยกคลาส P และ N โดยนิยามตามสมการที่ 2-1

$$I(p,n) = -\frac{p}{p+n}\log_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n}\log_2\left(\frac{n}{p+n}\right) \qquad (2\text{-}1)$$

ค่าคาดคะเนข้อมูล (Entropy) การใช้ลักษณะประจำ A ซึ่งกำหนด A คือ ลักษณะประจำที่แบ่ง S ออกเป็น {S1, S 2, ,Sv} โดยให้ S1 มีตัวอย่างจากคลาส P จำนวน p1 และตัวอย่างจากคลาส N จำนวน n1 ดังสมการที่ 2-2

$$E(A) = \sum_{i=1}^{v} \frac{p_1 + n_1}{p+n} I(p_1, n_1) \qquad (2\text{-}2)$$

ดังนั้นค่าเกนของข้อมูล (Data Gain) ที่ได้จากการแยกข้อมูลด้วยลักษณะประจำ A เป็นดังสมการที่ 2-3

$$Gain(A) = I(p,n) - E(A) \qquad (2\text{-}3)$$

ค่าที่เป็นรูปแบบหนึ่งของการใช้ต้นไม้ตัดสินใจที่พัฒนาเพิ่มเติมโดยไม่ได้ใช้ค่าเกนเป็นตัวแบ่ง แต่ ใช้อัตราค่าเกน (Gain ratio) เป็นตัวแบ่งดังสมการที่ 2-4

$$Gain\ ratio = \frac{Gain}{SplitInfo} \qquad (2\text{-}4)$$

4.2 อัลกอริธึม C4.5 (J48) เป็นอัลกอริทึมการสร้างกฎจากต้นไม้ตัดสินใจ (Decision Tree) มีพื้นฐานเดียวกับอัลกอริทึม ID3 ซึ่งได้พัฒนาเพิ่มเติมจาก ID3 ดังนี้

สามารถหลีกเลี่ยงการสร้างโครงสร้างต้นไม้ที่ใหญ่เกินไป เนื่องจากมีข้อมูล จำนวนมาก อย่างไรก็ตามขึ้นอยู่กับการกำหนดความลึก เมื่อมีการเจริญเติบของ Decision Tree

ความผิดพลาดลดลง เพราะมีการดัดทอนความผิดพลาดออกไป (Pruning node)

มีการสร้างกฎหลังการดัดทอนข้อมูลที่ผิดพลาดออก

สามารถใช้กับข้อมูลที่มีความต่อเนื่อง (Continuous Attributes) ที่เป็นตัวเลขได้ เช่น ค่าอุณหภูมิ จำนวนเงิน เป็นต้น

การเลือก Attribute ที่วัดการเลือกให้เหมาะสม

สามารถใช้กับชุดข้อมูล (Training Data) ที่มีค่าผิดพลาด (Missing Attribute)

สามารถใช้สำหรับ Attribute กับ Costs ที่แตกต่างกันได้มีการยอมรับประสิทธิภาพกระบวนการคำนวณที่ปรับปรุง

4.3 เนอีฟเบย์ (Naive-Bayes) เป็นวิธีสร้างโมเดลที่ง่ายและไม่ซับซ้อน โดยอาศัยทฤษฎีความน่าจะเป็น (probability) เป็นหลัก ซึ่งจัดเป็นเทคนิคในการแก้ปัญหาแบบ classification ที่สามารถคาดการณ์ ผลลัพธ์ได้และสามารถอธิบายได้ด้วย มันจะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ การเรียนรู้เบย์อย่างง่ายเป็นวิธีจำแนกประเภทขอมูลที่มีประสิทธิภาพวิธีหนึ่ง ที่การทำงานไม่ซับซ้อน เหมาะกับกรณีของเซตตัวอย่างมีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างไมขึ้นต่อกัน โดยกำหนดให้ความน่าจะเป็นของข้อมูลที่จะเป็นดังสมการที่ 2-5

$$P(a_1, a_2, \ldots a_n v_j) = \prod_{i=1}^{n} P(a_i \mid v_j) \qquad (2\text{-}5)$$

กลุ่ม V_j สำหรับข้อมูลที่มีคุณสมบัติ n ตัว X={a_1, a_2, ..., a_n} หรือ ใช้สัญลักษณ์ว่า P(a_1, a_2,..., a_n I V_j) โดยที่ $\prod$ หมายถึง ผลคูนของค่า P(a_i I v_j) ทั้งหมด I = 1,2,3,...,n แล j = 1,2,3,...,n ดังนั้นเราจะได้ว่า วิธีการจำแนกประเภทแบบเบย์อย่างง่าย ดังสมการที่ 2-6

$$V_{NB} = \underset{v_i \epsilon V}{argmag P(v_j)} x \prod_{i=1}^{n} P(a_i \mid v_j) \qquad (2\text{-}6)$$

4.4 อัลกอริทึม CART (Classification and Regression Tree) ถูกคิดค้นโดย Breiman ในปี ค.ศ.1984 ต้นไม้ตัดสินใจ ที่สร้างจากอัลกอริทึม CART จะเป็นต้นไม้แบบ binary ซึ่งประกอบด้วย กิ่งหรือแขนง 2 กิ่ง สำหรับแต่ละโหนดเทคนิคแบบ CART จะทำการแบ่งเรคคอร์ดใน Training Data Set ออกเป็น เรคคอร์ดย่อย ที่ให้ค่าเป้าหมาย (target) ที่เหมือนกัน (พวงทิพย์ แท่นแสง, 2550: 19) ดังสมการที่ 2-6

$$\emptyset(s\backslash t) = 2P_L P_R \sum_{j=1}^{\#classes} |P(j\backslash t_L) - P(j\backslash t_R)| \qquad (2\text{-}6)$$

โดยกำหนดให้ $\emptyset(s\backslash t)$ คือ หน่วยวัดค่าที่ดีที่สุดของการแตกโหนดคู่แข่ง s ที่โหนด t

t_L คือ โหนดลูกทางซ้ายของโหนด t

t_R คือ โหลดลูกทางขวาของโหนด t

$$P_L = \frac{\text{จำนวนเรคคอร์ดที่โหนด } t_L}{\text{จำนวนเรคคอร์ดใน } Training\ Set}$$

$$P_R = \frac{\text{จำนวนเรคคอร์ดที่โหนด } t_R}{\text{จำนวนเรคคอร์ดใน } Training\ Set}$$

$$P(j\backslash t_L) = \frac{\text{จำนวนเรคคอร์ดของคลาส } j \text{ ที่ } t_L}{\text{จำนวนเรคคอร์ดที่ } t}$$

$$P(j\backslash t_R) = \frac{\text{จำนวนเรคคอร์ดของคลาส } j \text{ ที่ } t_R}{\text{จำนวนเรคคอร์ดที่ } t}$$

5. แนวคิดและทฤษฎีเกี่ยวกับตัวแบบพยากรณ์แบบต้นไม้ตัดสินใจ

ต้นไม้ตัดสินใจ เป็นเทคนิคที่ให้ผลลัพธ์ในลักษณะของโครงสร้างต้นไม้ ซึ่งเมื่อมีข้อมูลที่ต้องการจะจัดกลุ่ม ก็จะนำ attribute ต่างๆ ของข้อมูลนั้นไปเทียบกับ Decision tree ตามเส้นทางใน tree จนกระทั่งคลาสปลายทางซึ่งก็คือ กลุ่มของข้อมูลที่เหมือนกัน โดยปกติมักประกอบถ้วยกฎในรูปแบบ “ถ้า เงื่อนไข แล้ว ผลลัพธ์” เช่น

“If Income = High and Married = No THEN Risk = Poor”
“If Income = High and Married = Yes THEN Risk = Good”

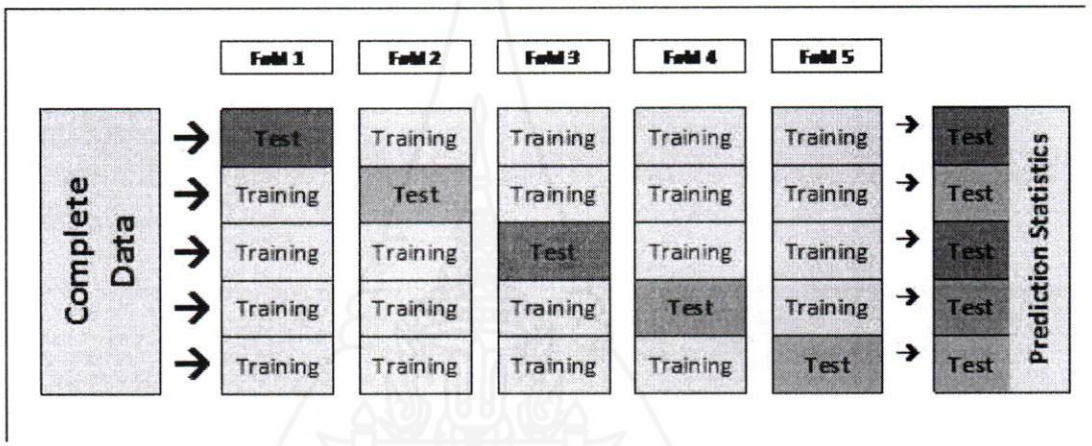
เนื่องจากผู้ใช้สามารถทำความเข้าใจผลลัพธ์ได้ง่าย เทคนิค Decision tree จะจำกัดข้อมูลที่เป็นตัวแปรตาม (dipendent variable) 1 ตัวต่อ 1 แบบจำลอง ถ้าต้องการทำนายตัวแปรตามหลายๆ ตัว ก็จะต้องสร้างแบบจำลองสำหรับตัวแปรตามแต่ละตัว ซึ่ง algoritm ของเทคนิคแบบ Decision tree นี้ ส่วนใหญ่ไม่รองรับข้อมูลแบบต่อเนื่อง (continous data) จะต้องมีการแบ่งให้เป็นข้อมูลแบบไม่ต่อเนื่องหรือแบบช่วง (discrete date) เสียก่อน ซึ่ง algoritm ที่ใช้ในการดำเนินการของ Decision tree นั้น ได้แก่ Chi-squared Automatic Interactiion Detection (CHAID), Classification and Regression Trees (CART), C4.5 และ C5.0 โดย algorithm เหล่านี้ส่วนมากมักเหมาะกับปัญหาแบบ classification (เกรียงไกร พิพิธหิรัญการ, 2550: 35)

ในงานวิจัยนี้ได้ใช้เทคนิคต้นไม้ตัดสินใจ ในการจำแนกข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปของต้นไม้ตัดสินใจซึ่งต้นไม้ตัดสินใจนั้นมีการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยหลักการสร้างต้นแบบคือการแบ่งข้อ มูลเป็น 2 ส่วน คือข้อมูลเรียนรู้ (Training Data) เป็นกลุ่มตัวอย่างของข้อมูลที่ได้กำหนดไว้ก่อนล่วงหน้า และนำแบบจำลองที่ได้ไปใช้กับ ข้อมูลทดสอบ (Test Data) เพื่อทดสอบความแม่นยำของโมเดล โดยสามารถพยากรณ์กลุ่มของข้อมูลที่ยังไม่เคยนามาจัดหมวดหมู่ได้ และในกระบวนการเหมืองข้อมูล ใช้ในการพยากรณ์ (Prediction) หรือการจำแนกเทคนิคการทำเหมืองข้อมูลกลุ่ม (Classification) แผนผังต้นไม้ตัดสินใจในรูปแบบของโหนดและแสดง

ผลลัพธ์จากการกระทำหรือตัดสินใจในเงื่อนไขต่างๆ แล้วเชื่อมต่อกันเป็นเส้นที่แตกแขนงออกไป เทคนิคนี้จะช่วยให้ผู้ใช้เข้าใจถึงความสัมพันธ์และคุณสมบัติของข้อมูลได้ง่ายกว่าเทคนิคอื่น

6. การประเมินประสิทธิภาพของโมเดล

จะใช้การทดสอบประสิทธิภาพของโมเดล แบบ k-Fold Cross Validation คือ การแบ่งข้อมูลเป็นส่วนสำหรับสอน (Training Set) และสำหรับทดสอบ (Testing Set) ออกเป็น k ส่วนเท่าๆ กัน ซึ่งค่า k ที่เหมาะสม คือ 5 (k = 5) หรือ 5-Fold Cross Validation ดังแสดงในภาพที่ 2.5



ภาพที่ 2.5 การแบ่งข้อมูลสำหรับทดสอบแบบ k-Fold Cross Validation
ที่มา : http://blog.goldenhelix.com/bchristensen/cross-validation-for-genomic-prediction-in-svs/ Retrieved January 20, 2015

จากภาพที่ 2.5 กำหนดค่า k = 5 สามารถแบ่งข้อมูลออกเป็น 5 ส่วนเท่าๆ กัน แต่ละส่วนประกอบด้วยข้อมูลจำนวน 5 เรคอร์ด และมีจำนวนรอบการทดสอบประสิทธิภาพทิ้งสิ้น 5 รอบ โดยในการทำงานรอบที่ 1 จะสุ่มเลือกข้อมูล 1 ส่วนสำหรับเป็นข้อมูลชุดทดสอบ และ 4 ส่วนที่เหลือจะใช้เป็นข้อมูลชุดสอน การทำงานรอบที่ 2 จะใช้หลักการเดียวกันโดยจะสุ่มเลือกจากข้อมูล 4 ส่วนที่เหลือสำหรับสร้างเป็นข้อมูลทดสอบ และจะทำเช่นนี้จนกระทั่งข้อมูลทุกส่วนถูกนำมาเป็นชุดทดสอบ และนำค่า ความถูกต้อง (Accuracy) ที่ได้ในแต่ละรอบมาหาค่าเฉลี่ยเพื่อเป็นค่าความถูกต้องโดยรวมของการจำแนกข้อมูลของแบบจำลอง

7. การวิเคราะห์ประสิทธิภาพ

ประสิทธิภาพ (Performance) คือ การทำงานของขั้นตอนวิธีแต่ละวิธีซึ่งสามารถวัดจากผลลัพธ์ของการจัดกลุ่มข้อมูล (Classification) ค่าของผลลัพธ์ที่ได้จากการจัดกลุ่ม คือ ค่า True Positive (TP) ค่า True Negative (TN) ค่า False Positive (FP) และ ค่า False Negative (FN) ตามลำดับ สามารถแสดงดังตารางที่ 2.1 (Written and Frank, 2005)

ตารางที่ 2.1 ค่าของคอนฟิวชันเมทริกซ์ (Confusion Matrix) แบบ 2 กลุ่ม

ค่าที่แท้จริง (Actual Class)	**ค่าที่ทำนายได้ (Predicted Class)**	
	Class YES	**Class NO**
Class YES	True Positive : TP	False Negative : FN
Class NO	False Positive: FP	True Negative : TN

ค่าที่ได้จากการทำนาย (Prediction) ในตารางที่ 2.1 อธิบายรายละเอียดได้ดังนี้

1) ค่า True Positive (TP) คือค่าที่บอกความถูกต้องในการจำแนกข้อมูลซึ่งมี ค่าที่แท้จริงอยู่ใน Class YES และมีการทำนายว่าอยู่ใน Class YES (ทำนายถูกต้อง)

2) ค่า False Negative (FN) คือ ค่าที่บอกความถูกต้องในการจำแนกข้อมูลซึ่งมีค่าที่แท้จริงอยู่ใน Class YES และมีการทำนายว่าอยู่ใน Class NO (ทำนายผิด)

3) ค่า False Positive (FP) คือ ค่าที่บอกความถูกต้องในการจำแนกข้อมูลซึ่งมีค่าที่แท้จริงอยู่ใน Class NO และมีการทำนายว่าอยู่ใน Class YES (ทำนายผิด)

4) ค่า True Negative (TN) คือ ค่าที่บอกความถูกต้องในการจำแนกข้อมูลซึ่งมีค่าที่แท้จริงอยู่ใน Class NO และมีการทำนายว่าอยู่ใน Class NO (ทำนายถูกต้อง)

จากค่าผลลัพธ์ที่ได้ทั้ง 4 ค่า สามารถนิยามการวัดประสิทธิภาพได้ดั้งนี้

7.1 Precision คือ ค่าของการทำนายค่าที่จริงในกลุ่มที่พบและถูกต้อง ซึ่งพิจารณาจาก จำนวนกลุ่มข้อมูลที่งหมด คำนวณดังสมการ (2-6)

$$\boldsymbol{Precision} = \frac{TP}{TP+FN} \qquad (2\text{-}6)$$

เมื่อ TP + FN มีค่าเท่ากับ 0 ให้ตัวหาร (TP + FN) มีค่าเป็น 1

7.2 Recall คือ ค่าจำนวนกลุ่มข้อมูลที่พบ ซึ่งพิจารณาจากจำนวนกลุ่มข้อมูลทั้งหมด ดังสมการที่ 2-7

$$Recall = \frac{TP}{TP+TN} \quad (2\text{-}7)$$

เมื่อ TP + TN มีค่าเท่ากบ 0 ให้ตัวหาร (TP + TN) มีค่าเป็น 1

7.3 F-measure คือ การวัดค่า Precision และ Recall พร้อมกันของโมเดล โดยพิจารณาแยกที่ละคลาส

7.4 Accuracy คือ การวัดความถูกต้องของโมเดล โดยพิจารณารวมทุกคลาส

8. งานวิจัยที่เกี่ยวข้อง

ศุภชัย ประคองศิลปี (2551) ได้นำเสนอการพัฒนาระบบสนับสนุนการตัดสินใจในการอนุมัติ ลูกบ้านเข้าโครงการ โดยใช้เทคนิคต้นไม้ตัดสินใจและในอัลกอริธึม ID3 ในการเรียนรู้จากข้อมูลลูกบ้านในอดีตเพื่อสร้างโมเดลสำหรับตัดสินใจอนุมัติลูกบ้านเข้าโครงการ โดยทำการทดสอบความแม่นยำของอัลกอริธึม ID3 เปรียบเทียบกับอัลกอริธึมอื่นที่เป็นเทคนิคต้นไม้ตัดสินใจเหมือนกัน ด้วยการนำข้อมูลใน**อดีต**ของบุคคลที่เคยเข้าร่วมโครงการทั้งหมด 963 รายการและวัดความแม่นยำ ของอัลกอริธึม ID3 โดยใช้ Confusion Matrix ได้ผลความแม่นยำคือ 93.67 % ซึ่งมากกว่าอัลกอริธึม ตัวอื่นๆ และผู้วิจัยได้วัดประสิทธิภาพโมเดลของระบบที่พัฒนาขึ้น เปรียบเทียบกับโปรแกรม Weka 3.5.8 โดยใช้ข้อมูลสำหรับการสอน (Training Data) ที่เหมือนกัน ผลที่ได้คือสามารถสร้างโมเดล ได้เหมือนกับระบบที่ผู้วิจัยพัฒนา จึงสรุปได้ว่าโมเดลที่พัฒนาขึ้นมีความถูกต้องและแม่นยำอยู่ใน ระดับดีมาก สามารถนำไปสนับสนุนงาน ได้จริง

พยูน พาณิชย์กุล (2548) ได้พัฒนาระบบดาต้าไมน์นิ่งโดยใช้ ต้นไม้ตัดสินใจ (Decision Tree) โดยนำเทคนิคของโครงสร้างต้นไม้การตัดสินใจมาพัฒนาระบบงาน มีวิธีการคือ ระบบจะรับข้อมูล (Data) มาประมวลผล เพื่อสร้างโครงสร้างต้นไม้ ซึ่งโครงสร้างของต้นไม้ที่ได้จะแสดงเป็นโครงสร้างที่มีกฎต่างๆ ตามเป้าหมายของการใช้งาน เพื่อนำมาแบ่งกลุ่มข้อมูลตามที่กำหนดไว้โดย ผลการพัฒนาระบบสามารถแบ่งประเภทได้ในระดับดี และสามารถหาความน่าเชื่อถือของการแบ่งประเภทได้ในระดับดี และระบบที่พัฒนาสามารถนำไปใช้ได้กับข้อมูลอื่นๆ เพราะไม่ได้กำหนดขอบเขตไว้กับงานใดงานหนึ่ง

มลธิดา ฤทธิสมบูรณ์ (2550) ได้นำเสนอการพัฒนาระบบสนับสนุนการพิจารณาอนุมัติให้ สินเชื่อเพื่อการเช่าซื้อสินค้า โดยใช้เทคนิคการแบ่งกลุ่มแบบโครงสร้างต้นไม้และใช้อัลกอริธึมในการเรียนรู้แบบ ID3 โดยใช้โปรแกรม Weka 3.4ในการทดลองเพื่อสร้างโมเดล อัลกอริธึมแบบ ID3 จะนำค่าชี้วัดมาใช้ในการแบ่งข้อมูล โดยกลุ่มตัวอย่าง คือชุดข้อมูลที่ใช้ในการเรียนรู้ เป็นตัว แปรเป้าหมาย และแอททริบิว คือตัวแปรอื่นๆ ที่ใช้ในการสร้างโหนดในต้นไม้ และไม่ใช่ตัวแปร เป้าหมาย ส่วนต้นไม้ตัดสินใจจะนำมาใช้ในการเลือกรากของต้นไม้โดยดูจากแอททริบิวที่มี ความสำคัญมากที่สุดก่อนแล้วลดลงตามลำดับเพื่อช่วยในการพิจารณาการอนุมัติ การให้สินเชื่อโดย วิธีการตังกล่าวเหมาะกับการจำแนกข้อมูลออกเป็นกลุ่มอย่างชัดเจน โดยระบบที่ได้พัฒนาขึ้นนั้นอยู่ ในระดับดี

วาสนา วงษ์กำภู (2552) ได้พัฒนาระบบสืบค้นงานวิจัยด้านจิตเวชศาสตร์ สำหรับจัดเก็บข้อมูลงานวิจัย และช่วยในการสืบค้นงานวิจัยได้อย่างเป็นระบบ โดยนำวิธีการของต้นไม้ตัดสินใจ (Decision Tree) เข้ามาช่วยในการจัดแบ่งหมวดหมู่งานวิจัย ทำให้สามารถแบ่งแยกข้อมูลงานวิจัยออกเป็นประเภทต่างๆ และสามารถแสดงผลลัพธ์ออกเป็นเงื่อนไขที่ผู้ใช้ต้องการ ทำให้การสืบค้นข้อมูลวิจัยมีประสิทธิภาพมากขึ้น

ปฏิยากร โคผดุง (2551) ได้นำเสนองานวิจัยเรื่องต้นไม้ตัดสินใจสำหรับการวิเคราะห์ภาวะตัวรับฮอร์โมนของผู้ป่วยมะเร็งเต้านมและอัตราการอยู่รอดชีวิตของผู้ป่วยมะเร็งเต้านมในโรงพยาบาล ขอนแก่น พบว่าโรคมะเร็งเต้านมเป็นสาเหตุการตายมากเป็นอันดับสองในหญิงไทย จากข้อมูลผู้ป่วย เพศหญิง ที่มีผลการวินิจฉัยโรคเป็นมะเร็งเต้านม และเข้ารับการรักษาในโรงพยาบาลขอนแก่นตั้งแต่วันที่ 1 มกราคม พ.ศ. 2550 วันที่ 31 ธันวาคม พ.ศ. 2550 ซึ่งงานวิจัยนี้ยกเว้นข้อมูลผู้ป่วยที่ไม่สามารถติดตามการรักษาหรือส่งต่อการรักษาไปที่สถานพยาบาลอื่น วัตถุประสงค์ในการวิจัยเพื่อใช้ในการจำแนกระดับภาวะตัวรับฮอร์โมนสาหรับแนวทางการรักษาผู้ป่วยมะเร็งเต้านม และใช้เป็นระบบสนับสนุนการตัดสินใจด้านคลินิก โดยการประยุกต์ใช้อัลกอริทึม J48 ในโปรแกรมเวกาสร้างตัวแบบต้นไม้ตัดสินใจโดยใช้ทฤษฎีการทำเหมืองข้อมูลกับข้อมูลผู้ป่วยจำนวน 57 คน ซึ่งทำการวิเคราะห์ข้อมูลจำนวน 57 instance มีแอททริบิวต์จำนวน 6 แอททริบิวต์ สำหรับการประเมินความถูกต้องของตัวแบบที่ใช้ทำนายใช้วิธี 10-fold cross validation และวิธี Train test พบว่าวิธี 10-fold cross validation มีความถูกต้อง 57.89 % และ วิธี Train test มีความถูกต้อง 66.67% ซึ่งวิธีนี้จะมีประสิทธิภาพสูงสุด ความรู้ที่ได้จากการทำเหมืองข้อมูลนี้จะเป็นประโยชน์ต่อแพทย์ผู้เชี่ยวชาญในการพัฒนาตัวแบบ การทำนายค่าและสนับสนุนกระบวนการตัดสินใจของแพทย์ได้อย่างถูกต้อง โดยเฉพาะการประยุกต์ใช้การจำแนกภาวะตัวรับฮอร์โมนในการรักษามะเร็งเต้านมซึ่งข้อมูลนี้สามารถทำนายการแพร่กระจายของมะเร็งเต้านมไปยังต่อมน้ำเหลืองต่อไป

พิจิตรา จอมศรี (2549) ได้นำเสนองานวิจัยเรื่องการทำนายเนื้อหาของเว็บ โดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษามหาวิทยาลัยศิลปากร โดยใช้เทคนิคกฎการค้นหาความสัมพันธ์ข้อมูลมาประยุกต์ใช้สร้างตัวแบบเพื่อทำนายข้อมูลการเรียกใช้เว็บในอนาคต โดยผู้วิจัยได้จัดเก็บข้อมูล 2 ส่วนคือการเรียกใช้เว็บภายในมหาวิทยาลัยศิลปากรจากระบบพร็อกซีเซิร์ฟเวอร์ และจัดทำฐานข้อมูลเว็บเพื่อจัดหมวดหมู่ของเว็บ แล้วนำข้อมูลทั้ง 2 ส่วนมาสร้างความสัมพันธ์ซึ่งงานวิจัยนี้นำข้อมูลวันเวลา หมวดเว็บ และเว็บมาค้นหาความสัมพันธ์เพื่อสร้างตัวแบบ และพิจารณาตัวแบบจากค่าความเชื่อมั่น และค่าสนับสนุนโดยผลของการศึกษาตัวแบบพบว่าโมเดลที่สร้างขึ้นสามารถทำนายเนื้อหาเว็บที่จะถูกเรียกใช้ในวันถัดมาได้ ผลของการใช้เทคนิคเหมืองข้อมูลพบว่าตัวแบบที่สร้างขึ้นสามารถทำนายเนื้อหาเว็บที่จะถูกเรียกใช้ได้ โดยมีความถูกต้องร้อยละ 66.67 % นอกจากนี้ยังนำเสนอระบบการทำนายเนื้อหาของเว็บ โดยใช้เทคนิคเหมืองข้อมูลซึ่งสร้างอยู่บนขั้นตอนวิธีที่นำเสนอและอธิบายถึงผลการทดลองบนข้อมูลจริงด้วย

ไพฑูรย์ จันทร์เรือง (2550) ได้นำเสนองานวิจัยเรื่องระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคต้นไม้ พบว่าการเลือกสาขาการเรียนของผู้ที่ต้องการสมัครเข้าศึกษาต่อในระดับปริญญาตรีในสถาบันอุดมศึกษาต่างๆ เป็นสิ่งที่สำคัญเนื่องจากส่งผลโดยตรงกับผู้สมัครเองทางด้านคุณภาพหลังจบการศึกษา โดยได้พัฒนาระบบสนับสนุนการ

ตัดสินใจเลือกสาขาการเรียนของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคต้นไม้ตัดสินใจ ซึ่งจากการทดลองพบว่าในการสร้างตัวแบบสำหรับพัฒนาระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคต้นไม้ตัดสินใจนั้น ควรแยกสร้างตัวแบบสำหรับแต่ละสาขาการเรียน เนื่องจากคุณสมบัติของผู้เรียนแต่ละสาขามีความแตกต่างกันเพื่อให้ได้ตัวแบบที่สามารถทำนายแนวโน้มของผลการเรียนที่เหมาะสมสำหรับแต่ละสาขา แต่เนื่องจากคะแนนเฉลี่ยของนักศึกษาที่นำมาพัฒนาตัวแบบนั้นส่วนใหญ่จะมีเกณฑ์คะแนนเกาะกลุ่มกันอยู่ในช่วงกลางของข้อมูล (2.00 - 3.00) ทำให้ผลการตัดสินส่วนใหญ่จะโน้มเอียงไปในเกณฑ์พอใช้ (ช่วงคะแนน 2.00 - 2.49) และปานกลาง (ช่วงคะแนน 2.50 - 2.99)

เสกสรรค์ วิลัยลักษณ์ (2558) ได้นำเสนอการใช้เทคนิคการทำเหมืองข้อมูลเพื่อการพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา โดยการพัฒนาคลังข้อมูลโดยใช้โครงสร้างแบบสโนว์เฟลกสกีมา ซึ่งใช้ข้อมูลนักเรียนระดับมัธยมศึกษาปีที่ 4 ระหว่างปีการศึกษา 2553-2556 จำนวน 525 ระเบียน ประกอบด้วย 16 คุณลักษณะ มาสร้างตัวแบบพยากรณ์และการเรียนโดยใช้ชุดข้อมูล 2 แบบ คือ ข้อมูลแบบไม่จัดกลุ่ม (Original Data) และข้อมูลแบบจัดกลุ่ม (Cluster Data) จากนั้นนำไปผ่านกระบวนการคัดเลือกคุณลักษณะ (Feature Selection) ซึ่งใช้วิธี Carrelation-based Feature Selection (GFS) และวิธี Information Gain (IG) แล้วใช้เทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ซัพพอร์เวกเตอร์แมชชีน (SVM) และต้นไม้ตัดสินใจ (Decision Tree) มาสร้างตัวแบบพยากรณ์และเปรียบเทียบตัวแบบ ด้วยการทดสอบประสิทธิภาพแบบ 10-fold Cross Validation ซึ่งผลการวิจัยพบว่า ผู้บริหาร หัวกลุ่มกลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้นมีความพึงพอใจการใช้งานคลังข้อมูลอยู่ในระดับดี และในการทำเหมืองข้อมูลพยากรณ์ พบว่าชุดข้อมูลแบบไม่จำกัดกลุ่มนำมาคัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Selection (CFS) ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 94.48 และมีค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดที่ 0.1880 เหมาะสมสำหรับการสร้างระบบพยากรณ์ผลการเรียนของนักเรียน

ธวัชชัย อดิเทพสถิต (2557) ได้นำเสนองานวิจัยการเปรียบเทียบการพยากรณ์ราคายางพาราด้วยวิธีการวิเคราะห์การถดถอยแบบพหุคูณกับวิธีโครงข่ายประสาทเทียม และหาวิธีพยากรณ์ราคายางพาราที่เหมาะสมที่สุด ซึ่งมีวิธีดำเนินการวิจัยโดยรวบรวมข้อมูลต่างๆ ของแต่ละปัจจัยซึ่งการเก็บรวบรวมตั้งแต่วันที่ 1 มกราคม พ.ศ. 2552 ถึง 31 ธันวาคม พ.ศ. 2556 เป็นจำนวน 1,826 ข้อมูล แล้วหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน เพื่อหาปัจจัยที่ส่งผลกระทบต่อราคายางพารา ทำการเปรียบเทียบประสิทธิภาพของการพยากรณ์ ด้วยวิธีการวิเคราะห์การถดถอยแบบพหุคูณและวิธีโครงข่ายประสาทเทียม ด้วยค่าวัดความถูกต้องของการพยากรณ์ คือ รากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง จากนั้นทำการทดสอบกับสถานการณ์จริงเพื่อหาข้อผิดพลาด ซึ่งเก็บรวบรวมตั้งแต่วันที่ 1 มกราคม พ.ศ. 2557 ถึง 30 มิถุนายน พ.ศ. 2557 เป็นจำนวน 181 ข้อมูล และนำโมเดลที่ได้มาพัฒนาต้นแบบระบบสารสนเทศพยากรณ์ราคายางพาราตามแบบจำลองที่ได้ ซึ่งผลการวิจัยพบว่า วิธีที่เหมาะสมสำหรับการพยากรณ์ราคายางพารา คือ วิธีการวิเคราะห์วิธีโครงข่ายประสาทเทียม ซึ่งรากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของวิธีการวิเคราะห์

การถดถอยแบบพหุคูณ เท่ากับ 6.4214 ซึ่งมีค่ามากกว่าการวิเคราะห์ด้วยวิธีโครงข่ายประสาทเทียม ที่มีค่าเท่ากับ 4.5151 ผลการวิจัย ซึ่งสรุปได้ว่าวิธีที่เหมาะสมสำหรับการพยากรณ์ราคายางพารา คือ วิธีโครงข่ายประสาทเทียม

ชฎิภกรน์ ทรายหมอ (2557) ได้นำเสนองานวิจัยการสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นในผู้ป่วยโรคเบาหวาน และทดสอบแบบจำลองโดยใช้กับกลุ่มงานอายุรกรรม ห้องตรวจ ผู้ป่วยเบาหวาน โรงพยาบาลปทุมธานี โดยมีวิธีการดำเนินการวิจัยโดยใช้ประวัติผลการตรวจรักษาผู้ป่วยเบาหวานโรงพยาบาลปทุมธานี ตั้งแต่วันที่ 1 มกราคม 2553 ถึงวันที่ 31 ธันวาคม 2556 จำนวน 17,043 คน 1) ทำความเข้าใจปัญหาที่ดำเนินงานการตรวจรักษาของแพทย์ต่อผู้ป่วย และโครงสร้างข้อมูล เพื่อการสร้างแบบจำลองให้ตรงตามความเป็นจริงมากที่สุด 2) จัดเตรียมข้อมูลผลการตรวจวินิจฉัยโรคเบาหวานของผู้ป่วยตั้งแต่เริ่มตรวจพบจนถึงปัจจุบันให้มีความถูกต้อง เพื่อให้แบบจำลองมีความแม่นยำ ทั้งนี้ข้อมูลที่ใช้ในการสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ ผู้ป่วยเป็นโรคเบาหวาน คือข้อมูลผลตรวจการวินิจฉัยโรคเบาหวานตั้งแต่ที่ผู้ป่วยเป็นโรคเบาหวาน เพื่อนำไปใช้ในการสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน 3) สร้างแบบจำลองประกอบไปด้วยตัวแปรที่สำคัญ ได้แก่ จำนวนครั้งการมาตรวจรักษากับแพทย์ จำนวนรหัสโรคที่ลงผลการตรวจวินิจฉัยโรคกับผู้ป่วยและเป็นผู้ป่วยเบาหวานเท่านั้น 4) เปรียบเทียบผลการพยากรณ์ โดยวิธีการหาความสัมพันธ์ของข้อมูลระหว่าง AprioriAlgorithm และ FPGrowth Algorithm 5) พยากรณ์ภาวะแทรกซ้อนโดยใช้เครื่องมือที่ประกอบด้วยโปรแกรม WEKA เพื่อวิเคราะห์ข้อมูลพยากรณ์ภาวะแทรกซ้อนโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน MicroSoft Visual Foxpro เพื่อจัดเก็บข้อมูลผลการตรวจวินิจฉัยโรคเบาหวาน Microsoft Execl 2010 สำหรับสร้างแบบจำลองพยากรณ์ภาวะแทรกซ้อนโรคอื่นหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน 6) ทดสอบแบบจำลองกับสถานการณ์จริง โดยพัฒนาระบบนำเสนอผลการพยากรณ์โอกาสการเกิดโรคอื่นที่เกี่ยวข้องจากการเป็นโรคเบาหวาน โดยใช้ตั้งแต่ 1 เมษายน 2557 ถึง 15 กรกฎาคม 2557 กับกลุ่มงานอายุรกรรม ห้องตรวจผู้ป่วยเบาหวานโรงพยาบาลปทุมธานี ข้อมูลที่นำเสนอมีประวัติผู้ป่วยผลการพยากรณ์ภาวะแทรกซ้อนของโรคเบาหวาน อาการตามกลุ่มโรค ระดับความเสี่ยงต่อการเป็นอัมพฤกษ์ อัมพาต ระดับความเสี่ยงต่อการเสียชีวิต ข้อควรปฏิบัติ/คำแนะนำจากแพทย์ผลการพยากรณ์ความสัมพันธ์ภาวะแทรกซ้อนของโรคเบาหวานหลังจากที่ผู้ป่วยเป็นโรคเบาหวาน ในผู้ป่วยเบาหวานโรงพยาบาลปทุมธานี พบว่าผู้ป่วยที่เป็นโรค 1) เบาหวานชนิดที่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา (รหัสโรค E103) แล้วจะเป็นโรคแทรกซ้อนดังนี้ 1.1) โรคจอตาจากโรคเบาหวานชนิดไม่งอกขยายหรือเอ็นพีดีอาร์ (รหัสโรค H3600) คิดเป็น 65% 1.2) โรคต้อกระจกในผู้สูงอายุ (รหัสโรค H251) คิดเป็น 46% 1.3) โรคจอตาจากโรคเบาหวานชนิดงอกขยายหรือพีดีอาร์ (รหัสโรค H3602) คิดเป็น 44% 2) โรคเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางตา (รหัสโรค E113) แล้วจะเป็นโรคแทรกซ้อน 2.1) โรคจอตาจากโรคเบาหวานชนิดไม่งอกขยาย (รหัสโรค H3600) คิดเป็น 77% 2.2) โรคต้อกระจกในผู้สูงอายุ (รหัสโรค H251) คิดเป็น 49% 2.3) โรคจอตาจากโรคเบาหวานชนิดงอกขยาย (รหัสโรค H3602) คิดเป็น 36% 2.4) โรคไตวายเรื้อรังที่มิได้ระบุรายละเอียด (รหัสโรค N189) คิดเป็น 20% 3) โรคเบาหวานชนิดที่ไม่ต้องพึ่งอินซูลิน ร่วมกับภาวะแทรกซ้อนทางประสาท (รหัสโรค E114) แล้วจะเป็นโรคแทรกซ้อน 3.1) โรคต้อกระจกใน

ผู้สูงอายุ (รหัสโรค H251) คิดเป็น 24% 3.2) โรคหัวใจขาดเลือดเรื้อรังที่มิได้ ระบุรายละเอียด (รหัสโรค I251) คิดเป็น 20%

ธีรพงศ์ สงผัด (2556) ได้นำเสนองานวิจัยการใช้เทคนิคการทาเหมืองข้อมูล (Data Mining) เพื่อสนับสนุนการให้บริการสารสนเทศของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ วัตถุประสงค์ของงานวิจัยคือ 1) การสร้างคลังข้อมูลผู้ใช้บริการห้องสมุด 2) การทาเหมืองข้อมูลเพื่อวิเคราะห์และจาแนกกลุ่มผู้ใช้บริการห้องสมุด 3) การทาเหมืองข้อมูลเพื่อวิเคราะห์และศึกษารูปแบบการยืมหนังสือของผู้ใช้บริการห้องสมุด 4) การจัดทำรายงานหลายมิติจากข้อมูลที่จัดเก็บในคลังข้อมูลตามความต้องการของผู้ใช้ เครื่องมือที่ใช้ในงานวิจัย ได้แก่ โปรแกรม SQL Server 2008 และ SQL Service Analysis กระบวนการเริ่มจากการนาข้อมูลการใช้ห้องสมุดจากฐานข้อมูลห้องสมุดตั้งแต่ปี พ.ศ. 2553 - 2556 และข้อมูลทั่วไปของนักศึกษาจากฐานข้อมูลระบบทะเบียนนักศึกษามาสร้างคลังข้อมูลด้วยกระบวนการอีทีแอล (ETL) วัตถุประสงค์ของกระบวนการอีทีแอลคือ เพื่อให้ข้อมูลมีความถูกต้องสมบูรณ์ จากนั้นนำข้อมูลจากคลังข้อมูลมาสร้างรายงานหลายมิติตามความต้องการผ่านกระบวนการโอแลป สำหรับการทาเหมืองข้อมูลใช้แบบจาลองคริสพ์-ดีเอ็ม (CRISP-DMModel) และเทคนิคการทาเหมืองข้อมูลที่ใช้ประกอบด้วย 3 เทคนิค ได้แก่ 1) การจัดกลุ่ม เพื่อแบ่งกลุ่มสมาชิกที่มีลักษณะการยืมหนังสือคล้ายกันออกเป็นกลุ่มๆ 2) กฎการหาความสัมพันธ์ วิเคราะห์ข้อมูลการยืมหนังสือเพื่อค้นหาความสัมพันธ์ที่น่าสนใจระหว่างหนังสือที่ถูกยืมว่า มีหนังสือเล่มใดบ้างที่มักถูกยืมไปด้วยกันและ 3) อนุกรมเวลา เพื่อพยากรณ์ปริมาณผู้ใช้บริการสารสนเทศห้องสมุดจาแนกตามประเภทของผู้ใช้บริการผลการวิจัยที่ได้คือ คลังข้อมูลผู้ใช้บริการห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษที่ใช้โครงสร้างแบบสโนว์เฟลก (Snowflake schema Model) ผู้ใช้บริการที่มีคุณลักษณะคล้ายกันถูกจัดออกเป็นทั้งหมด 6 กลุ่ม การหาความสัมพันธ์ของหนังสือที่มักถูกยืมไปด้วยกัน พบว่าที่ค่าความเชื่อมั่นไม่น้อยกว่าร้อยละ 90 และค่าสนับสนุนไม่น้อยกว่า 9 ได้กฎความสัมพันธ์ทั้งหมด 20 กฎ และอนุกรมเวลาทาให้สามารถพยากรณ์ปริมาณผู้ใช้บริการสารสนเทศห้องสมุดในแต่ละกลุ่มหรือประเภทได้ การทำเหมืองข้อมูลทำให้บุคลากรทำงานในห้องสมุดได้รับความรู้ใหม่ ซึ่งสามารถนำมาประกอบในการแนะนำหนังสือหรือจัดกิจกรรมส่งเสริมการใช้บริการสารสนเทศของห้องสมุดตามกลุ่มของผู้ใช้ ยิ่งกว่านั้นยังช่วยเพิ่มความสะดวกในการค้นคืนสารสนเทศหรือรายชื่อหนังสือที่ต้องการ ลดเวลาในการสืบค้นช่วยสนับสนุนการวางแผนจัดกิจกรรมเพื่อส่งเสริมการใช้บริการและเพิ่มปริมาณผู้ใช้ นอกจากนี้แล้วยังทำให้ผู้บริหารได้รับรายงานหลายมิติที่สามารถมองภาพรวมและเจาะลึกในรายละเอียดตามมิติที่สนใจได้

บทที่ 3
วิธีดำเนินการวิจัย

งานวิจัยนี้ได้นำเทคนิคการทำเหมืองข้อมูล เพื่อพยากรณ์ผลผลิตมันสำปะหลัง โดยได้กำหนดขั้นตอนการดำเนินการวิจัยดังภาพที่ 3.1

ภาพที่ 3.1 แผนผังขั้นตอนการดำเนินงานวิจัย

1. ข้อมูลและกลุ่มตัวอย่างที่ใช้ในการวิจัย

ข้อมูลที่ใช้ในงานวิจัยการพัฒนาตัวแบบพยากรณ์ผลผลิตมันสำปะหลังโดยใช้เทคนิคการทำเหมืองข้อมูล และพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูลและการประเมินความพึงพอใจผ่านเว็บไซต์ จะใช้ข้อมูลปัจจัยที่เกี่ยวข้องกับผลผลิตมันสำปะหลัง ซึ่งเป็นข้อมูล แบบทุติยภูมิ ตั้งแต่ปี 2551 - 2555 จากสำนักงานเกษตรจังหวัดกำแพงเพชร ซึ่งประกอบด้วยข้อมูลผลผลิตมันสำปะหลังของจังหวัดในภาคเหนือที่มีการปลูกมันสำปะหลังมากที่สุด และมีสภาพภูมิอากาศที่ใกล้เคียงกับจังหวัดกำแพงเพชร จำนวน 5 จังหวัด ได้แก่ กำแพงเพชร นครสวรรค์ เพชรบูรณ์ พิษณุโลก และอุทัยธานี รวม 325 ครัวเรือน และกลุ่มตัวอย่างในการประเมินความพึงพอใจในการใช้งานระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง ประกอบด้วย เจ้าหน้าที่สำนักงานเกษตรจังหวัดกำแพงเพชร ผู้ใช้งานทั่วไป และผู้ดูแลระบบ รวม 30 คน

2. การทำข้อมูลให้สมบูรณ์

ข้อมูลปัจจัยการผลิตที่นำมาจากสำนักงานเกษตรจังหวัดกำแพงเพชร บางส่วนยังมีความไม่สมบูรณ์จำเป็นต้องดำเนินการทำให้ข้อมูลมีความสมบูรณ์ เพื่อนำมาเข้ากระบวนการในการจัดทำเหมืองข้อมูล โดยทำการจัดข้อมูลให้เป็นรูปแบบมาตรฐานเดียวกัน (Normalization) ปรับเปลี่ยนค่าของแอตทริบิวต์ให้เป็นหน่วยนับเดียวกัน เช่น หน่วย ไร่ งาน ตารางวา ให้เป็นหน่วย ไร่ ปริมาณปุ๋ยปรับหน่วยเป็นกิโลกรัม ส่วนปริมาณผลผลิตปรับหน่วยเป็นตัน และผลผลิตให้อยู่ในหน่วยมาตรฐาน คือ ตันต่อไร่ หลังจากนั้นกำจัดข้อมูลที่ในแถวเป็นค่าว่าง (NULL) มีค่าข้อมูลเป็น 0 หรือระเบียนที่มีค่าผลผลิตเป็น 0 ออกและเลือกแอตทริบิวต์ที่มีคาดว่าสามารถนำมาใช้ประโยชน์ได้ และมีข้อมูลครบถ้วน เช่น จังหวัด อำเภอ ตำบล พันธุ์ที่ใช้ อายุปลูกถึงเก็บ เดือนเก็บผลผลิต ปีเก็บผลผลิต เนื้อที่เก็บเกี่ยว เนื้อที่เสียหาย เนื้อที่ปลูก เนื้อที่ ชนิดปุ๋ย ปริมาณปุ๋ย และปริมาณผลผลิต เป็นต้น จากนั้นทำการกำหนดรหัสข้อมูล ดังแสดงในตารางที่ 3.1 - 3.6 ดังนี้

ตารางที่ 3.1 แทนความหมายในแอททริบิวต์

ชื่อข้อมูล	ความหมาย
Year	ปี
Province	จังหวัด
Amphoe	อำเภอ
Tambon	ตำบล
Village	หมู่บ้าน
Species	พันธุ์ที่ใช้
CassAge	อายุปลูกถึงเก็บ (เดือน)
AreaGrownV	เนื้อที่ปลูก (ตร.ว.)
AreaGrownR	เนื้อที่ปลูก (ไร่)
AreaHarvestV	เนื้อที่เก็บเกี่ยว (ตร.ว.)
AreaHarvestR	เนื้อที่เก็บเกี่ยว (ไร่)
YieldProdTN	ปริมาณผลผลิต (ตัน)
YieldProdKG	ปริมาณผลผลิต (กิโลกรัม)
YieldProdAndAreaGrownx400TN	ผลผลิตต่อไร่ (ตัน/ไร่)
YieldProdAndAreaGrownxNEW	ระยะช่วงผลผลิต (ตัวอักษร)
YieldProdAndAreaGrownxNUM	ระยะช่วงผลผลิต (ตัวเลข)
AreaDamageV	เนื้อที่เสียหาย (ตร.ว.)
AreaDamageR	เนื้อที่เสียหาย (ไร่) (งาน)
FertAreaCHV	เนื้อที่ใส่ปุ๋ยเคมี (ไร่)
FertAreaCHR	เนื้อที่ใส่ปุ๋ยเคมี (งาน)
FertQuanCHKG	ปริมาณปุ๋ยเคมีที่ใช้ (กก.)
FertQuanCHx400	ปริมาณปุ๋ยเคมีต่อเนื้อที่ใช้ปุ๋ยเคมี (ไร่)
FertAreaCKV	เนื้อที่ใส่ปุ๋ยคอก (ตร.ว.)
FertAreaCKR	เนื้อที่ใส่ปุ๋ยคอก (ไร่)
FertQuanCK(KG)	ปริมาณปุ๋ยคอกที่ใช้ (กก.)
FertQuanCLx400	ปริมาณปุ๋ยอินทรีย์ต่อเนื้อที่ปุ๋ยอินทรีย์ (ไร่)
FertQuanCKx400	ปริมาณปุ๋ยคอกต่อเนื้อที่ใช้ปุ๋ยคอก (ไร่)

ตารางที่ 3.2 กำหนดจังหวัด

รหัสจังหวัด	จังหวัด
K amphaengphet	กำแพงเพชร
Nakhon Sawan	นครสวรรค์
Phitsanulok	พิษณุโลก
Phetchabun	เพชรบูรณ์
Uthai Thani	อุทัยธานี

ตารางที่ 3.3 กำหนดรหัสพันธุ์

หมายเลข	รหัสประเภทพันธุ์	พันธุ์
1	A	ระยอง 1
2	B	ระยอง 3
3	C	ระยอง 5
4	D	ระยอง 60
5	E	ระยอง 90
6	F	ศรีราชา 1
7	G	เกษตรศาสตร์ 50
8	H	ห้วยบง 60
9	I	ระยอง 7
10	J	ระยอง 9
11	K	ระยอง 72
12	L	ห้วยบง 80
99	M	พันธุ์อื่น ๆ

ตารางที่ 3.4 กำหนดรหัสปุ๋ย

รหัสปุ๋ย	พันธุ์
CH	ปุ๋ยเคมี
CK	ปุ๋ยคอก
CB	ปุ๋ยชีวภาพ
CL	ปุ๋ยอินทรีย์

ตารางที่ 3.5 กำหนดรหัสหน่วยนับ

รหัสหน่วยนับ	หน่วยนับ
RR	ไร่
NG	งาน
TW	ตารางวา
KG	กิโลกรัม
TN	ตัน

4. การคัดเลือกข้อมูล

การคัดเลือกข้อมูลปัจจัยการผลิตมันสำปะหลัง ได้พิจารณาจาก การสอบถามข้อมูลเจ้าหน้าที่สำนักงานเกษตรจังหวัดกำแพงเพชร และจากเกษตรกรผู้ปลูกมันสำปะหลัง จึงได้คัดเลือกข้อมูลแอตทริบิวต์ที่เกี่ยวข้องกับมันสำปะหลัง จำนวน 11 ปัจจัย ได้แก่ จังหวัด พันธุ์ที่ใช้ อายุปลูก ผลผลิตต่อไร่ เนื้อที่ปลูก เนื้อที่เก็บเกี่ยว เนื้อที่เสียหาย ปริมาณปุ๋ยเคมี ปริมาณปุ๋ยคอก ปริมาณปุ๋ยชีวภาพ และปริมาณปุ๋ยอินทรีย์ แล้วทำการวิเคราะห์สัมพันธ์ของข้อมูลทั้ง 11 ปัจจัย ทั้งนี้จะไม่นำจังหวัดมาวิเคราะห์หาความสัมพันธ์เนื่องจากจังหวัดมีลักษณะเป็นการแบ่งเขตพื้นที่การปลูกซึ่งก็ชัดเจนว่าเป็นตัวแปรที่มีความสำคัญและสัมพันธ์กับตัวแปรอื่นๆ อยู่แล้ว จากนั้นจะกำหนดให้ผลผลิตต่อไร่เป็นตัวแปรตาม ส่วนปัจจัยการผลิตที่เหลืออีก 9 ปัจจัย กำหนดให้เป็นตัวแปรต้น ซึ่งผลการเคราะห์สหสัมพันธ์ (Correlation Analysis) โดยใช้โปรแกรม SPSS แสดงดังตารางที่ 3.6

ตารางที่ 3.6 ผลการเคราะห์สหสัมพันธ์ (Correlation Analysis)

ปัจจัยการผลิต	ค่าสัมประสิทธิ์สหสัมพันธ์
พันธุ์ที่ใช้	-0.194*
อายุปลูก	0.228*
เนื้อที่ปลูก	0.047**
เนื้อที่เก็บเกี่ยว	0.051**
เนื้อที่เสียหาย	-0.025
ปริมาณปุ๋ยเคมี	0.084*
ปริมาณปุ๋ยคอก	-0.014
ปริมาณปุ๋ยชีวภาพ	0.000
ปริมาณปุ๋ยอินทรีย์	0.025*

* P – Value < 0.01, ** P Value <0.05

จากตารางที่ 3.6 จะพบว่าตัวแปรที่มีนัยสำคัญระดับ 0.01 มีจำนวน 4 ตัวแปร ได้แก่ พันธุ์ที่ใช้ อายุปลูก ปริมาณปุ๋ยเคมี และปริมาณปุ๋ยอินทรีย์ และตัวแปรที่มีนัยสำคัญระดับ 0.05 มีจำนวน 2 ตัวแปร ได้แก่ เนื้อที่ปลูก และเนื้อที่เก็บเกี่ยว ฉะนั้น สรุปได้ว่าตัวแปรหรือปัจจัยการผลิตที่มีความสัมพันธ์กัน จำนวน 5 ตัวแปร ได้แก่ พันธุ์ที่ใช้ อายุปลูก ปริมาณปุ๋ยเคมี ปริมาณปุ๋ยอินทรีย์ และจังหวัด

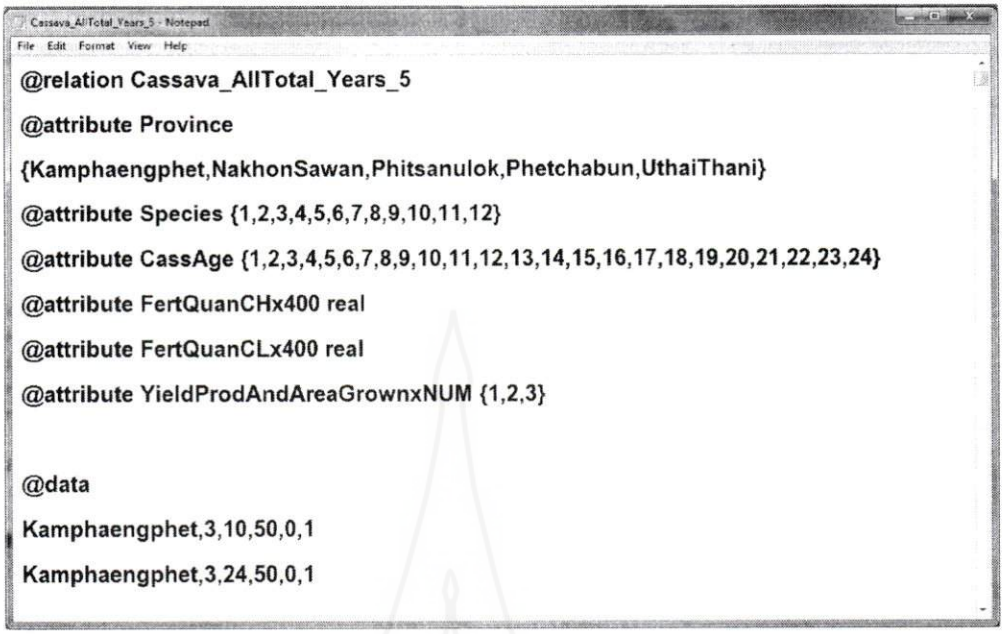
5. การแปลงข้อมูล

จากข้อมูลปัจจัยการผลิตของสำนักงานเกษตรจังหวัดกำแพงเพชร จำนวน 1,764 ระเบียน 40 แอตทริบิวต์ ผู้วิจัยได้ทำการจัดข้อมูลให้มีความสัมบูรณ์และทำการคัดเลือกข้อมูลได้ออกมาเป็นจำนวน1,677 ระเบียน 11 แอทริบิวต์ จากนั้นทำการวิเคราะห์หาความสัมพันธ์ของปัจจัยการผลิต จำนวน 11 ปัจจัย (แอตทริบิวต์) จะได้ปัจจัยการผลิตที่มีความสัมพันธ์กันจำนวน 5 ตัวแปร (แอตทริบิวต์) ทั้งนี้งานวิจัยนี้ได้ใช้วิธีการทำเหมืองข้อมูล โดยใช้เทคนิคต้นไม้ตัดสินใจในการสร้างตัวแบบการพยากรณ์ ซึ่งผลลัพธ์ที่ได้จะแสดงออกมาเป็นค่าระดับ หรือค่าช่วง ฉะนั้น ผู้วิจัยจึงทำการกำหนดค่าระดับผลผลิตผลิตมันสำปะหลัง โดยอาศัยการสอบถามข้อมูลค่าเฉลี่ยผลผลิตสูงสุด และต่ำสุด ต่อพื้นที่การปลูกหนึ่งไร่ จากเจ้าหน้าที่สำนักงานเกษตรจังหวัดกำแพงเพชร และเกษตรกรผู้ปลูกมันสำปะหลัง ซึ่งได้ค่าระดับผลผลิตดังนี้

ระดับ 1 = ผลผลิตมากกว่า 6 ตัน/ไร่
ระดับ 2 = ผลผลิตตั้งแต่ 3 – 6 ตัน/ไร่
ระดับ 3 = ผลผลิตตั้งแต่ 0 – 3 ตัน/ไร่

6. การพัฒนาตัวแบบพยากรณ์

ผู้วิจัยได้ใช้โปรแกรมเวกา (WEKA) เพื่อสร้างตัวแบบการพยากรณ์ผลผลิตมันสำปะหลัง โดยใช้ข้อมูลปัจจัยการผลิตจำนวน 1,677 ระเบียน 5 แอตทริบิวต์ โดยใช้เทคนิคการจำแนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) และใช้อัลกอริทึม J48, Random Tree, Simple Cart, Naïve Bayes และ LAD Tree ซึ่งในการนำเข้าข้อมูลสู่โปรแกรมเวกา (WEKA) ต้องทำการแปลงข้อมูลให้อยู่ในรูปของเท็กซ์ไฟล์ (Text File) ซึ่งแสดงดังภาพที่ 3.2



ภาพที่ 3.2 ไฟล์ข้อมูลที่นำเข้าสู่โปรแกรม WEKA

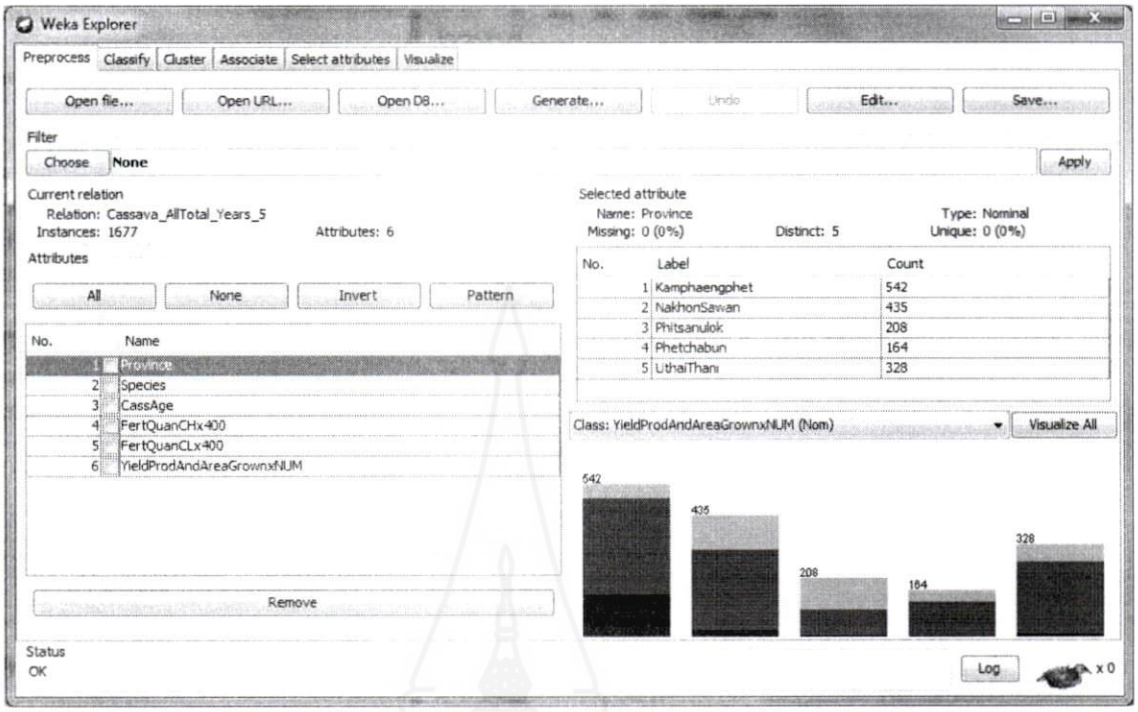
จากภาพที่ 3.2 จะพบว่าไฟล์ข้อมูลที่นำเข้าสู่โปรแกรม WEKA จะต้องทำการจัดให้อยู่ในรูปแบบไฟล์ .arff โดยอธิบายรายละเอียดได้ดังนี้

@relation name เป็นการบอกชื่อตารางข้อมูลเชิงสัมพันธ์

@attribute att-name type เป็นการบอกชื่อลักษณะประจำแต่ละชนิด เช่น ลักษณะประจำเก็บเป็นตัวเลข และลักษณะประจำเก็บค่าไม่ต่อเนื่อง

@data เป็นการบอกถึงแถว และตามมาด้วยข้อมูล โดยคั่นด้วย comma (,)

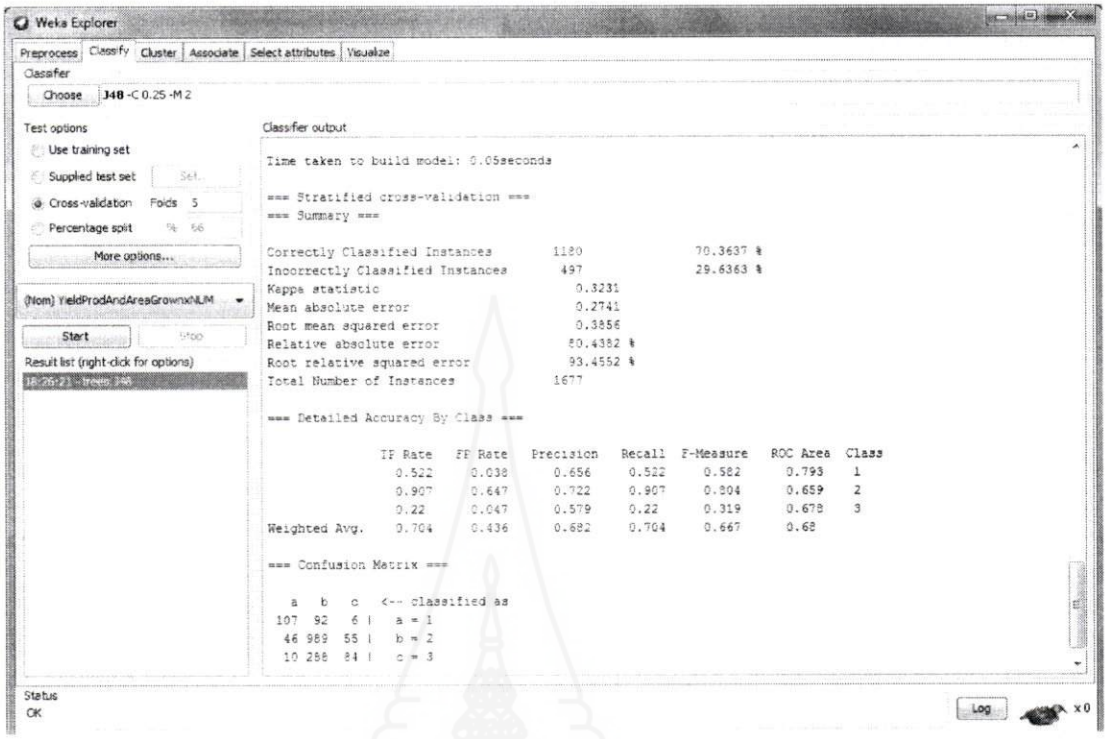
และเมื่อเข้าข้อมูลเสร็จเรียบร้อยแล้วจะแสดงในภาพที่ 3.3



ภาพที่ 3.3 ขั้นตอนการนำเข้ามูลสู่โปรแกรม WEKA

จากภาพที่ 3.3 แสดงจำนวนระเบียน ได้แก่ จังหวัดกำแพงเพชร 542 ระเบียน จังหวัดนครสวรรค์ 435 ระเบียน จังหวัดพิษณุโลก 208 ระเบียน จังหวัดเพชรบูรณ์ 164 ระเบียน และจังหวัดอุทัยธานี 328 ระเบียน

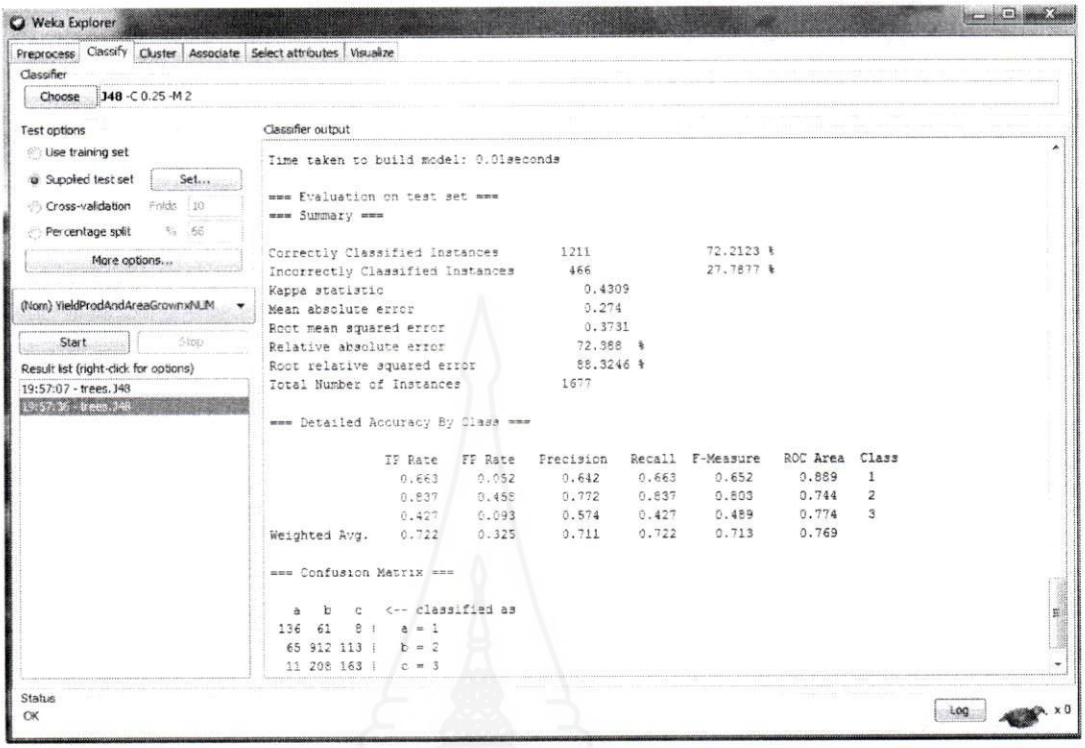
จากนั้นทำการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิค Cross-validation Test โดยทำการแบ่งข้อมูลออกเป็น 5 ส่วน (5-fold cross-validation) และ 10 ส่วน (10-fold cross-validation) จากตัวแบบการพยากรณ์ที่ได้จากการใช้เทคนิคการจำแหนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) ใช้อัลกอริทึม J48, Random Tree, Simple Cart, Naïve Bayes และ LAD Tree ซึ่งแสดงดังภาพที่ 3.4



ภาพที่ 3.4 ขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิค 5-fold cross-validation ของอัลกอริทึม J48

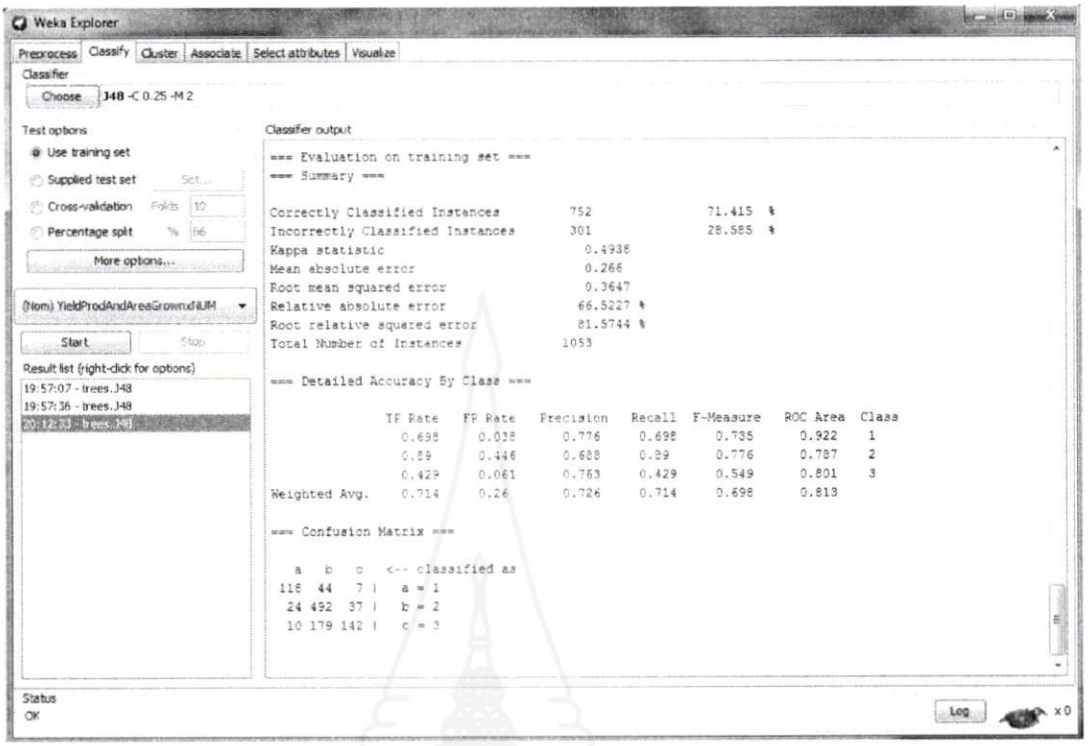
จากภาพที่ 3.4 แสดงขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิค 5-fold cross-validation ของอัลกอริทึม J48 ซึ่งได้ค่าแม่นยำ 70.36% จากนั้นจะทำการทดสอบเช่นนี้ในทุกอัลกอริทึม และเมื่อทำการทดสอบด้วยเทคนิค 5-fold cross-validation แล้วก็จะดำเนินการทดสอบด้วยเทคนิค 10-fold cross-validation ต่อไป

จากนั้นจะทำการปรับปรุงวิธีการทดสอบให้มีความแม่นยำที่ดีขึ้น โดยทำการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะรักษาสัดส่วนของข้อมูล และจะทำการสุ่มข้อมูลตามสัดส่วนดังกล่าว ทำให้จะได้ชุดข้อมูลที่จะทำการทดสอบจำนวน 5 ชุด แล้วทำการทดสอบจากอัลกอริทึม J48, Random Tree, Simple Cart, Naïve Bayes และ LAD Tree ซึ่งแสดงภาพที่ 3.5



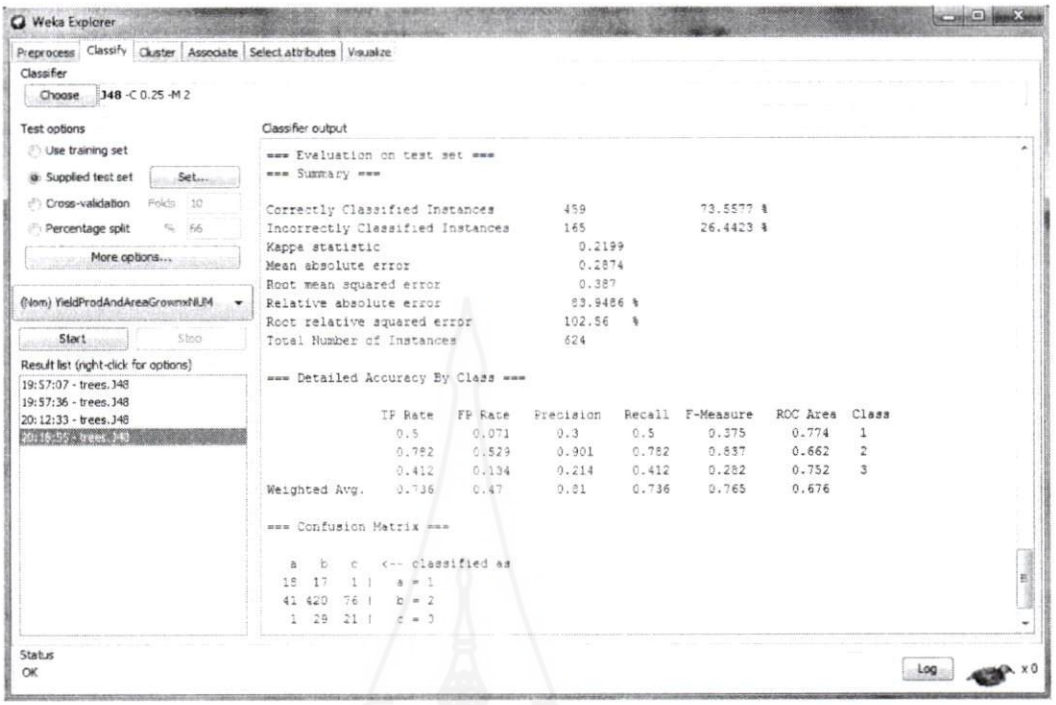
ภาพที่ 3.5 ขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะเป็นในส่วนของ Data Set ของอัลกอริทึม J48

จากภาพที่ 3.5 แสดงขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะเป็นในส่วนของ Data Set ของอัลกอริทึม J48 ซึ่งจะได้ค่าความแม่นยำ 72.21%



ภาพที่ 3.6 ขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะเป็นในส่วนของ Traning Set ของอัลกอริทึม J48

จากภาพที่ 3.6 แสดงขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะเป็นในส่วนของ Traning Set ของอัลกอริทึม J48 ซึ่งจะได้ค่าความแม่นยำ 71.41%



ภาพที่ 3.7 ขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะเป็นในส่วนของ Test Set ของอัลกอริทึม J48

จากภาพที่ 3.7 แสดงขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะเป็นในส่วนของ Test Set ของอัลกอริทึม J48 ซึ่งจะได้ค่าความแม่นยำ 73.55%

ทั้งนี้ จากภาพที่ 3.5 – 3.7 เป็นการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) ในชุดข้อมูลที่ 1 ของอัลกอริทึม J48 เท่านั้น ซึ่งผู้วิจัยได้ยกมาแสดงให้เห็นถึงขั้นตอนการพัฒนาตัวแบบพยากรณ์ โดยในทางปฏิบัติได้ทำการทดสอบต่อไปในทุกอัลกอริทึม และชุดข้อมูลที่เหลืออีก 4 ชุด ซึ่งผลการพัฒนาตัวแบบการพยากรณ์ได้ปรากฏรายละเอียดในบทที่ 4

7. การออกแบบระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูลและการประเมินความพึงพอใจผ่านเว็บไซต์

เมื่อทำการทดสอบความถูกต้องของตัวแบบพยากรณ์แล้ว จะนำเอาอัลกอริทึมที่ได้จากตัวแบบพยากรณ์มาพัฒนาเป็นระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูลและการประเมินความพึงพอใจผ่านเว็บไซต์ โดยได้แบ่งขั้นตอนการการดำเนินงาน ดังนี้

7.1 ศึกษาระบบงานเดิม

ในระบบงานเดิมข้อมูลที่เกี่ยวกับผลผลิตมันสำปะหลัง จะเก็บอยู่ในรูปเอกสารการจดบันทึก บางส่วนจัดเก็บในรูปแบบตารางทำการ ประกอบกับข้อมูลมีจำนวนมาก มีลักษณะการจัดเก็บข้อมูลที่แตกต่างกันไม่เป็นมาตรฐานเดียวกัน ทำให้ไม่สะดวกในการนำข้อมูลดังกล่าวไปใช้งาน ในการวิเคราะห์หรือคาดการณ์ผลผลิตได้อย่างสะดวก รวดเร็ว ถูกต้อง และแม่นยำ

7.2 ศึกษาระบบงานใหม่

เมื่อทราบถึงปัญหาในการที่จะนำข้อมูลที่เกี่ยวข้องกับผลผลิตมันสำปะหลังมาใช้งานแล้ว ผู้วิจัยจึงมีแนวคิดที่จะนำระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ มาใช้ในการบริหารจัดการข้อมูลที่เกี่ยวข้องกับผลผลิตมันสำปะหลังให้มีประสิทธิภาพมากยิ่งขึ้น และเกิดประโยชน์สูงสุดต่อไป

7.3 ออกแบบระบบงาน

จากข้อมูลการศึกษาระบบงานเดิม และระบบงานใหม่ พบว่ายังมีไม่การบริหารจัดการข้อมูลจำนวนมากด้วยระบบคอมพิวเตอร์มาใช้งาน ผู้วิจัยจึงได้พัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ มาใช้งาน โดยใช้เทคนิคการทำเหมืองข้อมูลโดยการจำแนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) และใช้อัลกอริทึม J48, Random Tree, Simple Cart, Naïve Bayes และ LAD Tree มาเป็นตัวแบบพยากรณ์ โดยระบบสารสนเทศดังกล่าวจะมีผู้ใช้งาน 3 ส่วน ได้แก่ เจ้าหน้าที่ของสำนักงานเกษตรจังหวัดกำแพงเพชร ผู้ดูแลระบบ และผู้ใช้งานทั่วไป

7.4 การพัฒนาระบบ

จะนำตัวแบบการพยากรณ์ที่ได้จากโปรแกรมเวกา (WEKA) ซึ่งเป็นเงื่อนไขที่ได้จากต้นไม้ตัดสินใจ มาพัฒนาเป็นระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ โดยใช้ภาษา PHP ในการพัฒนาเว็บไซต์ ใช้โปรแกรม MySQL ในการจัดการฐานข้อมูล และใช้โปรแกรม Apache เป็นเว็บเซิร์ฟเวอร์ โดยระบบสามารถการสืบค้นข้อมูลสามารถกำหนดสิทธิ์การใช้งาน ได้แก่ เจ้าหน้าที่ของสำนักงานเกษตรจังหวัดกำแพงเพชร ผู้ดูแลระบบ และผู้ใช้งานทั่วไป สามารถค้นหา ปรับปรุง แก้ไข แสดงผล และพิมพ์ข้อมูลจังหวัด อำเภอ ตำบล เนื้อที่ปลูก พันธุ์ที่ใช้ ปุ๋ย

7.5 การประเมินผลระบบ

ระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ ที่ผู้วิจัยได้พัฒนาขึ้นจะมีในของการประเมินความพึงพอใจผ่านเว็บไซต์ ซึ่งจะให้ผู้ที่ใช้งานระบบสามารถประเมินความพึงพอใจในการใช้งานได้ทันทีหลังการใช้งาน ซึ่งในงานวิจัยนี้ได้มีผู้ตอบแบบประเมินความพึงพอใจการใช้งานระบบพยากรณ์ผลผลิตมันสำปะหลังและระบบสืบค้นข้อมูล จำนวน 30 คน ได้แก่ เจ้าหน้าที่ของสำงานเกษตรจังหวัดกำแพงเพชร ผู้ดูแลระบบ และผู้ใช้งานทั่วไป

8. สถิติที่ใช้ในงานวิจัย

จะใช้สถิติเชิงพรรณนา (Descriptive Statistics) ในการวัดค่ากลางของข้อมูลโดยใช้ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) หรือค่าเฉลี่ย (Mean: $\overline{X}$) การหาค่าร้อยละ (Percentage) และวัดการกระจายของข้อมูลโดยใช้ ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: SD) รวมทั้งการประมวลข้อมูลดังนี้

8.1 ค่าเฉลี่ย (Mean)

$$\overline{X} = \frac{\sum x}{n} \qquad (3\text{-}1)$$

เมื่อ $\overline{X}$ คือ ค่าเฉลี่ย
$\sum X$ คือ ผลรวมของข้อมูลทั้งหมด
n คือ จำนวนข้อมูลทั้งหมด

8.2 ร้อยละ (Percentage)

$$p = \frac{f}{N} \times 100 \qquad (3\text{-}2)$$

เมื่อ p คือ ร้อยละ
f คือ ความถี่ที่ต้องการแปลงเป็นร้อยละ
n คือ ขนาดของตัวอย่าง

8.3 ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

$$S = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} \qquad (3\text{-}3)$$

เมื่อ S คือ ค่าส่วนเบี่ยงเบนมาตรฐาน
X คือ ข้อมูลแต่ละจำนวน
$\overline{X}$ คือ จำนวนคะแนนในแต่ละกลุ่ม
n คือ จำนวนข้อมูลจากกลุ่มตัวอย่าง

8.4 ร้อยละของความพึงพอใจ (Percentage of satisfaction)

ร้อยละของความพึงพอใจ = (B x 100) / A -4)

เมื่อ A คือ คะแนนเต็มจากแบบสอบถาม คำนวณได้จาก ค่าของระดับคะแนนที่มากที่สุด x จำนวนข้อคำถามในแบบสอบถาม x จำนวนผู้ตอบแบบสอบถาม

B คือ คะแนนรวมจริงที่ได้จากแบบสอบถามของผู้ตอบแบบสอบถามทั้งหมด

บทที่ 4
ผลการวิจัย

ผลการวิจัยการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล ผู้วิจัยได้ดำเนินการดังนี้

1. ผลการพัฒนาตัวแบบพยากรณ์

จากข้อมูลปัจจัยการผลิตมันสำปะหลังจำนวน 1,764 ระเบียน 40 แอตทริบิวต์ แล้วทำความสะอาดข้อมูลโดยทำการตัดระเบียนที่มีข้อมูลไม่สมบูรณ์ออก จะได้ข้อมูลชุดใหม่จำนวน 1,677 ระเบียน 11 แอตทริบิวต์ และผ่านการทดสอบข้อมูลทางสถิติจะได้ปัจจัยการผลิตมีความสัมพันธ์กันจำนวน 5 ตัวแปร (แอตทริบิวต์) ได้แก่ พันธุ์ที่ใช้ อายุปลูก ปริมาณปุ๋ยเคมี ปริมาณปุ๋ยอินทรีย์ และจังหวัด จากนั้นทำการกำหนดค่าระดับผลผลิตผลิตมันสำปะหลัง ซึ่งสามารถสรุปข้อมูลระดับผลผลิตมันสำปะหลัง แสดงในตารางที่ 4.1

ตารางที่ 4.1 แสดงค่าระดับผลผลิต

จังหวัด	ระดับผลผลิต			จำนวนระเบียน
	1 (มากกว่า 6 ตัน/ไร่)	2 (ตั้งแต่ 3-6 ตัน/ไร่)	3 (ตั้งแต่ 0-3 ตัน/ไร่)	
กำแพงเพชร	151	342	49	542
นครสวรรค์	128	295	12	435
เพชรบูรณ์	3	120	41	164
พิษณุโลก	9	90	109	208
อุทัยธานี	14	253	61	328
รวม	305	1100	272	1677

2. ผลการทดสอบตัวแบบพยากรณ์

เมื่อสร้างตัวแบบพยากรณ์ผลผลิตมันสำปะหลัง โดยใช้ข้อมูลปัจจัยการผลิตจำนวน 1,677 ระเบียน 5 แอตทริบิวต์ โดยใช้เทคนิคการจำแนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) และใช้อัลกอริทึม J48, Random Tree, Simple Cart, Naïve Bayes และ LAD Tree รวมเป็น 5 อัลกอริทึม จากนั้นทำการทดสอบความแม่นยำในการพยากรณ์ด้วย

เทคนิค Cross-validation Test โดยทำการแบ่งข้อมูลออกเป็น 5 ส่วน (5-fold cross-validation) และ 10 ส่วน (10-fold cross-validation) ซึ่งผลลัพธ์แสดงในตารางที่ 4.2

ตารางที่ 4.2 แสดงผลการทดสอบตัวแบบพยากรณ์ด้วยเทคนิค Cross-validation Test

ข้อมูล	เทคนิค	J48	RandomTree	SimpleCart	NaiveBayes	LADTree
(ระเบียน)		(%)	(%)	(%)	(%)	(%)
1677	5- folds	70.36	65.59	71.37	67.62	68.69
	10- folds	69.82	66.30	70.96	68.27	68.51

จากตารางที่ 4.2 ผลการทดสอบค่าความแม่นยำของการพยากรณ์ในแต่ละเทคนิคด้วยวิธี 5 – folds Cross-Validation และ 10 – folds Cross-Validation พบว่าทั้งสองเทคนิคยังให้ค่าความแม่นยำสูงสุดเพียง 70.96% ดังนั้นผู้วิจัยจึงทำการปรับปรุงวิธีการทดสอบให้มีความแม่นยำที่ดีขึ้น โดยการทำการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะรักษาสัดส่วนของข้อมูล ดังแสดงในตารางที่ 4.3

ตารางที่ 4.3 แสดงการแบ่งสัดส่วนของข้อมูล

ข้อมูล	จำนวนระเบียน	อัตราส่วน (%)
Tranining Set	1053	63
Test Set	624	37
Data Set	1677	100

จากตารางที่ 4.3 ได้กำหนดสัดส่วนของข้อมูลแล้ว ผู้วิจัยได้ทำการสุ่มข้อมูลตามสัดส่วนเพื่อสร้างชุดข้อมูการเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) ใหม่ (Data Set = Training Set + Test Set) จำนวน 5 ชุด แล้วทำการทดสอบค่าความแม่นยำของตัวแบบพยากรณ์โดยใช้เทคนิคการจำแนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) และใช้อัลกอริทึมเดิม ผลลัพธ์ที่ได้แสดงในตารางที่ 4.4

ตารางที่ 4.4 สรุปความแม่นยำในการพยากรณ์ของข้อมูลโดยวิธีการแบ่งข้อมูลทั้งชุดเป็นข้อมูลเรียนรู้และข้อมูลทดสอบ (Data Set = Tranining Set + Test Set)

อัลกอริทึม	ชุดข้อมูล (ระเบียน)	ชุดที่ 1 (%)	ชุดที่ 2 (%)	ชุดที่ 3 (%)	ชุดที่ 4 (%)	ชุดที่ 5 (%)	ค่าเฉลี่ย (%)
J48	Data Set (1677)	72.21	74.06	72.92	72.74	72.33	72.33
	Traning Set (1053)	71.41	73.78	74.26	71.41	70.37	70.37
	Test Set (624)	73.55	74.51	70.67	75.48	75.64	**75.64**
Random	Data Set (1677)	81.09	80.02	80.56	80.14	80.02	80.37
Tree	Traning Set (1053)	89.17	88.22	89.17	87.93	87.46	88.39
	Test Set (624)	67.46	66.66	65.70	66.98	67.46	**66.85**
Simple	Data Set (1677)	70.48	72.03	72.80	72.15	70.30	71.55
Cart	Traning Set (1053)	69.70	72.93	71.41	70.75	64.48	69.85
	Test Set (624)	71.79	70.67	75.16	74.51	80.12	**74.45**
Naïve	Data Set (1677)	66.54	65.77	66.96	63.80	68.33	66.28
Bayes	Traning Set (1053)	65.71	66.28	66.85	63.24	64.29	65.27
	Test Set (624)	67.94	64.90	66.66	64.90	75.16	**67.91**
LAD	Data Set (1677)	70.18	67.85	64.87	70.18	67.44	68.10
Tree	Traning Set (1053)	65.15	66.47	64.95	65.14	63.15	64.97
	Test Set (624)	78.67	70.35	63.94	78.68	74.67	**73.26**

จากตารางที่ 4.4 จะพบว่าการทดสอบความแม่นยำในการพยากรณ์ของชุดข้อมูลโดยการวิธีการแบ่งข้อมูลออกเป็นข้อมูการเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Tranining Set + Test Set) จำนวน 5 ชุด โดยอัลกอริทึม J48 อัลกอริทึม Simple Cart และอัลกอริทึม LAD Tree จะให้ความแม่นยำที่ดีขึ้นอย่างชัดเจนเมื่อเทียบการทดสอบความแม่นยำแบบ Cross-validation Test ที่ทำการแบ่งข้อมูลออกเป็น 5 ส่วน (5-fold cross-validation) และ 10 ส่วน (10-fold cross-validation) ส่วนอัลกอริทึม Random Tree และอัลกอริทึม Naïve Bayes มีค่าความแม่นยำที่ต่ำใกล้เคียงกับค่าที่ได้จากการทดสอบแบบ Cross-validation Test

ตารางที่ 4.5 สรุปผลการทดสอบตัวแบบพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set)

เทคนิค	J48	RandomTree	SimpleCart	NaiveBayes	LADTree
	(%)	(%)	(%)	(%)	(%)
Data Set =	75.64	66.85	74.45	67.91	73.26
Training					
+ Test Set					
(พิจารณาจาก % ค่าเฉลี่ย Test Set)					

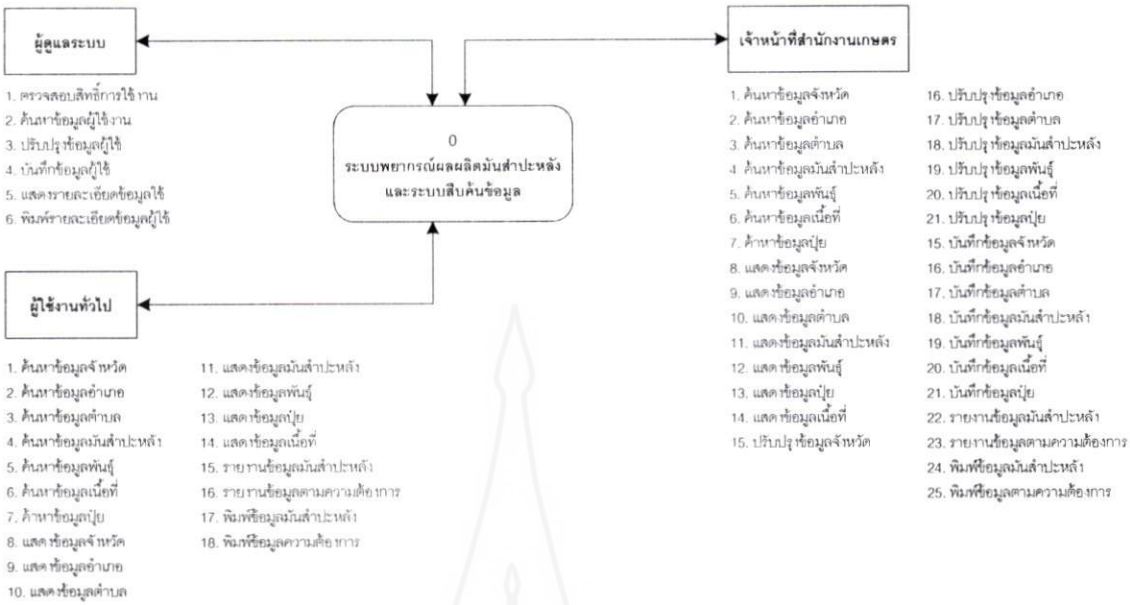
จากตารางที่ 4.5 จะพบว่าค่าเฉลี่ย Test Set ของอัลกอริทึม J48 ให้ค่าความแม่นยำสูงสุดที่ 75.64% อัลกอริทึม Simple Cart ให้ค่าความแม่นยำสูงสุดที่ 74.45% อัลกอริทึม LAD Tree ให้ค่าความแม่นยำสูงสุดที่ 73.26% ส่วนอัลกอริทึม Random Tree ให้ค่าความแม่นยำที่ 66.85% และอัลกอริทึม Naïve Bayes ให้ค่าความแม่นยำที่ 67.91% ซึ่งค่าความแม่นยำที่ได้ต่ำ ผู้วิจัยจึงคัดเลือกอัลกอริทึมที่มีค่าความแม่นยำที่ดีที่สุด จำนวน 3 ตัว ได้แก่ อัลกอริทึม J48, Simple Cart และ LAD Tree มาเป็นตัวแบบการพยากรณ์ผลผลิตมันสำปะหลัง โดยพิจารณาจากชุดข้อมูล Test Set ดังนี้

ชุดข้อมูลที่ 5 ของอัลกอริทึม J48 ให้ค่าความแม่นยำสูงสุดที่	75.64%
ชุดข้อมูลที่ 5 ของอัลกอริทึม Simple Cart ให้ค่าความแม่นยำสูงสุดที่	80.12%
ชุดข้อมูลที่ 4 ของอัลกอริทึม LAD Tree ให้ค่าความแม่นยำสูงสุดที่	78.68%

3. ผลการออกแบบระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์

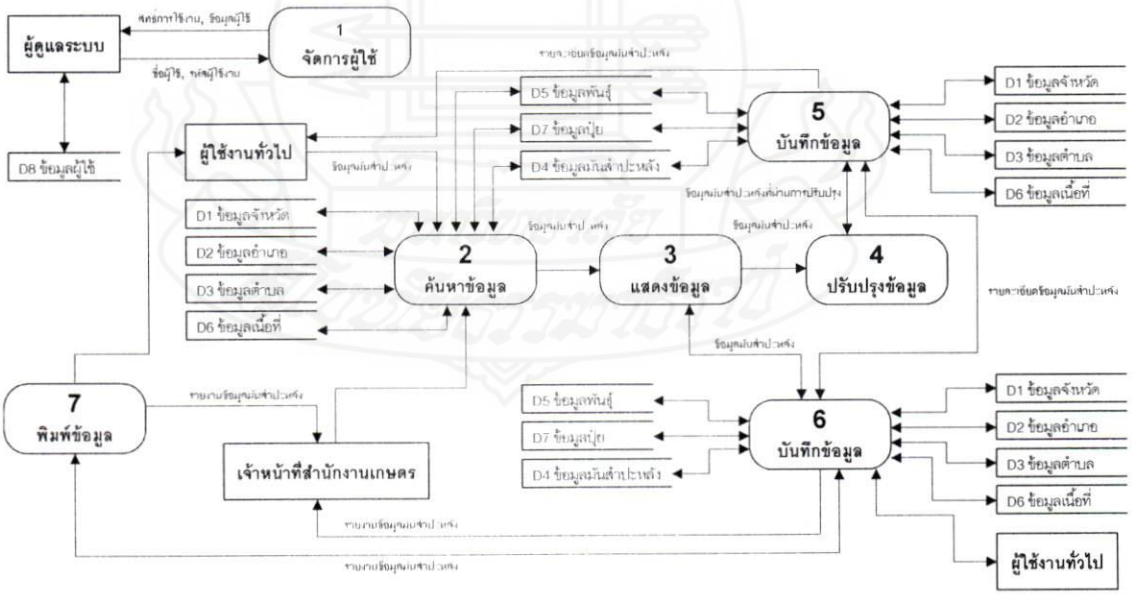
เมื่อได้ตัวแบบการพยากรณ์แล้ว จะนำเงื่อนไขที่ได้จากต้นไม้ตัดสินใจ ของอัลกอริทึม J48 จำนวน 171 บรรทัด อัลกอริทึม Simple Cart จำนวน 71 บรรทัด และอัลกอริทึม LAD Tree จำนวน 1321 บรรทัด มาพัฒนาเป็นระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ โดยใช้ภาษา PHP ในการพัฒนาเว็บไซต์ ใช้โปรแกรม My SQL ในการจัดการฐานข้อมูล และใช้โปรแกรม Apache เป็นเว็บเซิร์ฟเวอร์

3.1 ทำการออกแบบระบบโดยใช้ผนภาพกระแสข้อมูล (Data Flow Diagram: DFD) เป็นเครื่องมือในการวิเคราะห์ระบบ โดยแสดงการไหลของข้อมูล และความสัมพันธ์ของข้อมูลที่ได้จากการวิเคราะห์ระบบ เพื่อแสดงการนำเข้าข้อมูล (Input) และแสดงผลลัพธ์ (Output) ของระบบในการทำงานแต่ละส่วน โดยเส้นทางการไหลของข้อมูลดังกล่าว ซึ่งลำดับภาพรวมของกระบวนการทำงาน (Context Diagram) แสดงดังภาพที่ 4-1



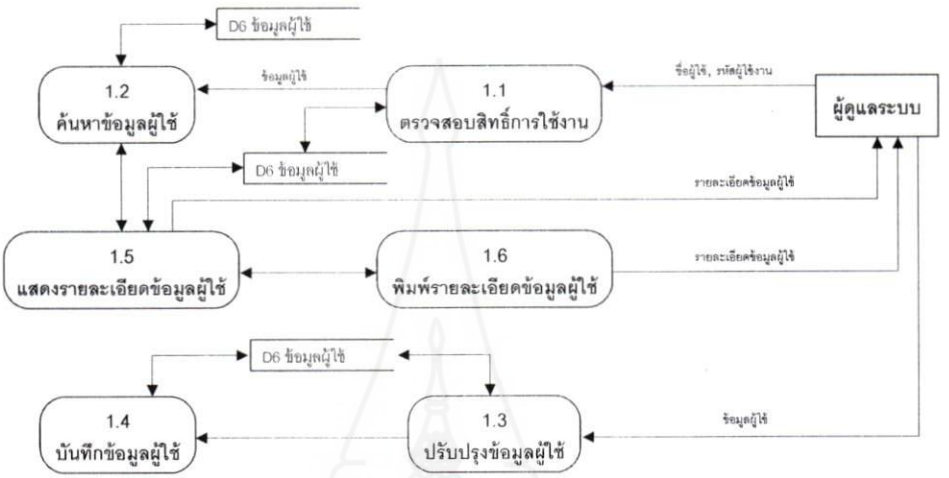
ภาพที่ 4.1 แผนภาพบริบทของระบบพยากรณ์ผลผลิตมันสำปะหลังและระบบสืบค้นข้อมูล

จากภาพที่ 4.1 แสดงเส้นทางของข้อมูลที่เข้าและออกจากแหล่งข้อมูลที่มีผลกระทบต่อระบบโดยมีผู้ที่เกี่ยวข้องกับการใช้งาน ซึ่งประกอบ เจ้าหน้าที่ของสำนักงานเกษตร ผู้ดูแลระบบ และผู้ใช้งานทั่วไป

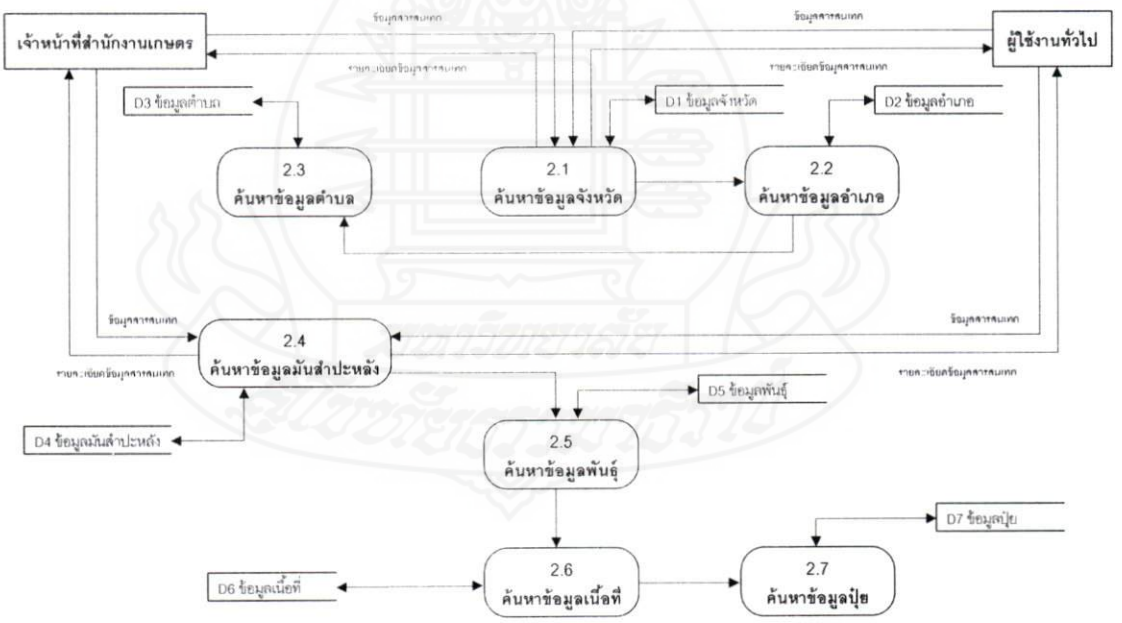


ภาพที่ 4.2 แผนภาพกระแสข้อมูลที่ระดับ 0 (Data Flow Diagram Level 0)

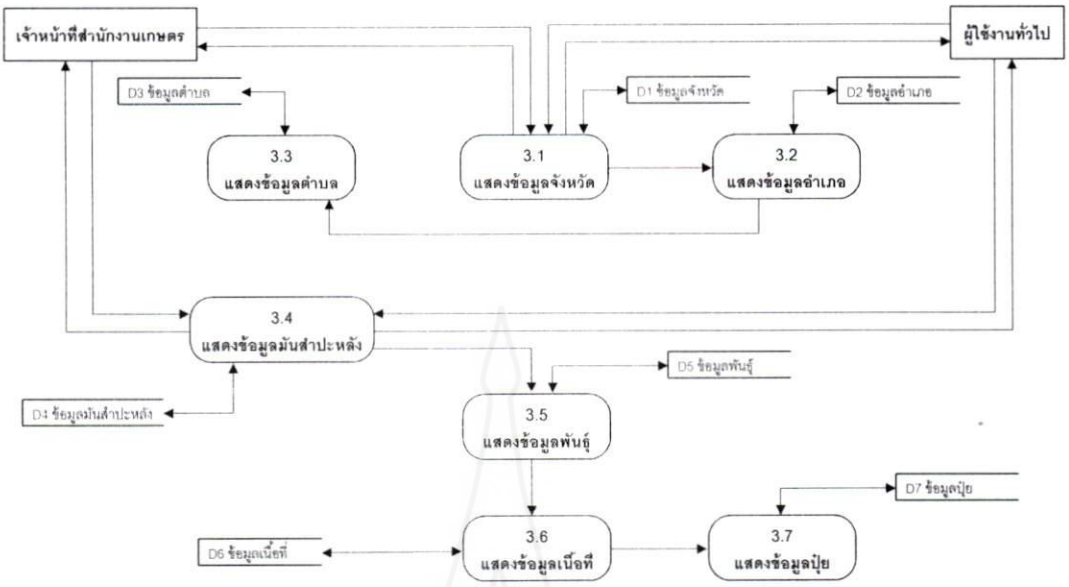
จากภาพที่ 4.2 แสดงกระแสข้อมูลในส่วนของการจัดการผู้ใช้งานระบบ ได้แก่ เจ้าหน้าที่ของสำนักงานเกษตร ผู้ดูแลระบบ และผู้ใช้งานทั่วไป สามารถค้นหาข้อมูล จังหวัด อำเภอ ตำบล และเนื้อที่ปลูก แสดงข้อมูล พันธุ์ที่ใช้ ปุ๋ย เนื้อที่ปลูก ปรับปรุงข้อมูล การบันทึกข้อมูล แสดงรายงาน และสั่งพิมพ์ข้อมูลในระบบได้ เป็นต้น



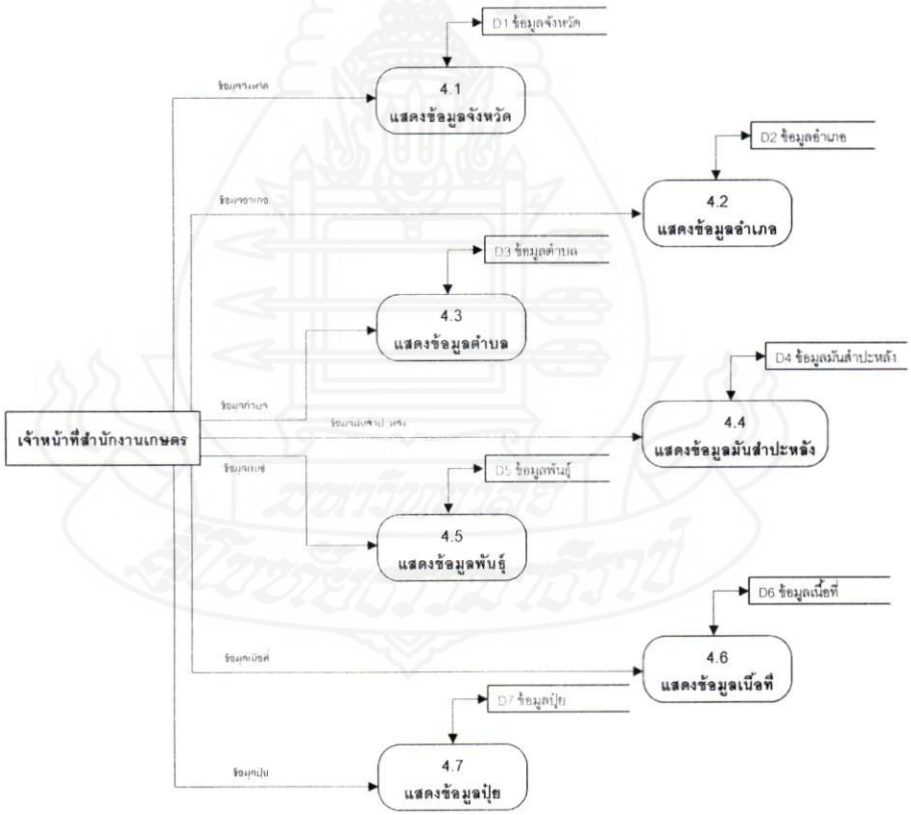
ภาพที่ 4.3 แผนภาพกระแสข้อมูลที่ระดับ 1 (Data Flow Diagram Level 1)



ภาพที่ 4.4 แผนภาพกระแสข้อมูลที่ระดับ 2 (Data Flow Diagram Level 2)

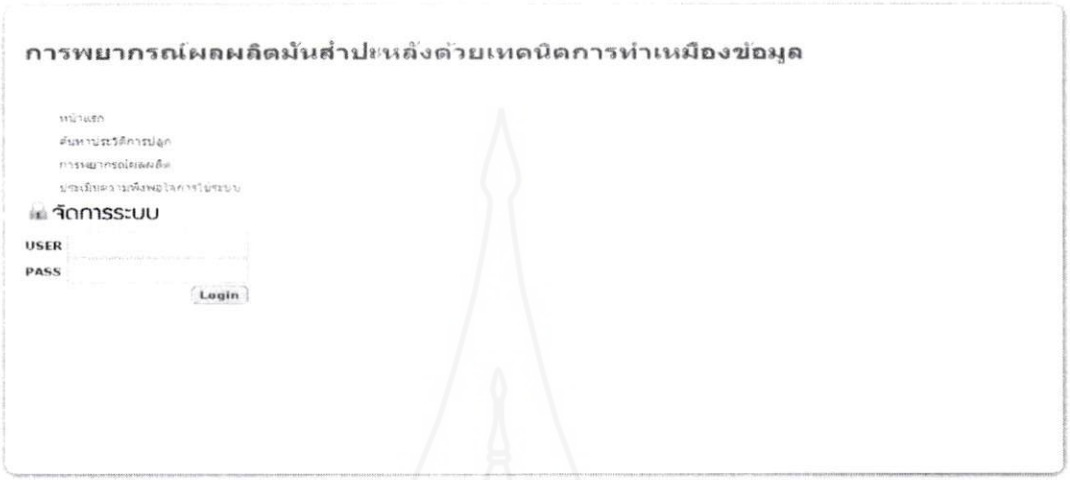


ภาพที่ 4.5 แผนภาพกระแสข้อมูลที่ระดับ 3 (Data Flow Diagram Level 3)



ภาพที่ 4.6 แผนภาพกระแสข้อมูลที่ระดับ 4 (Data Flow Diagram Level 4)

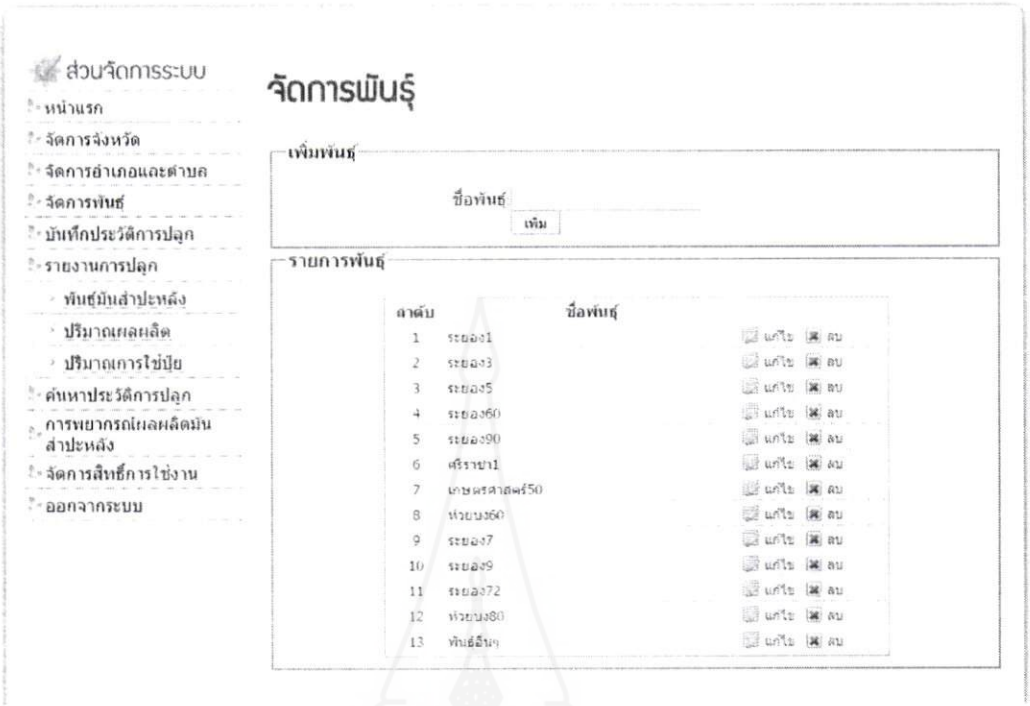
3.2 หน้าจอหลักของระบบสารสนเทศการพยาการ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจ ซึ่งจะมีเมนูแสดงค้นหาประวัติการปลูก การพยากรณ์ผลผลิต ประเมินความพึงพอใจการใช้ระบบ และจัดการระบบ ดังแสดงในภาพที่ 4.7



ภาพที่ 4.7 หน้าจอหลักของระบบ

ภาพที่ 4.8 หน้าจอจัดการสิทธิ์การใช้งาน

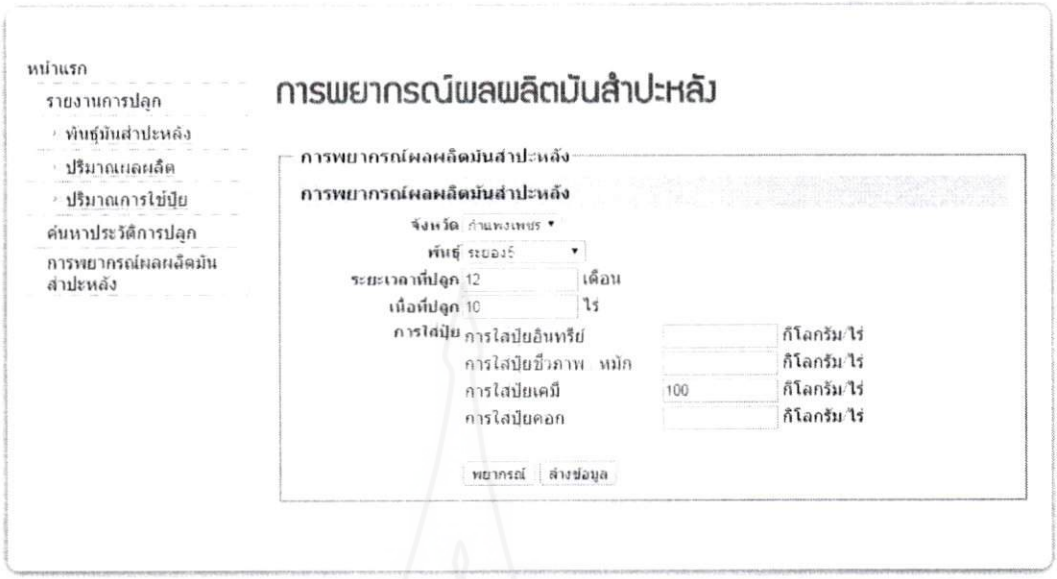
จากภาพที่ 4.8 เป็นการจัดการสิทธิ์การใช้งาน สามารถแก้ไข เพิ่มเติม และลบ ผู้ใช้งานระบบสารสนเทศการพยาการ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจ



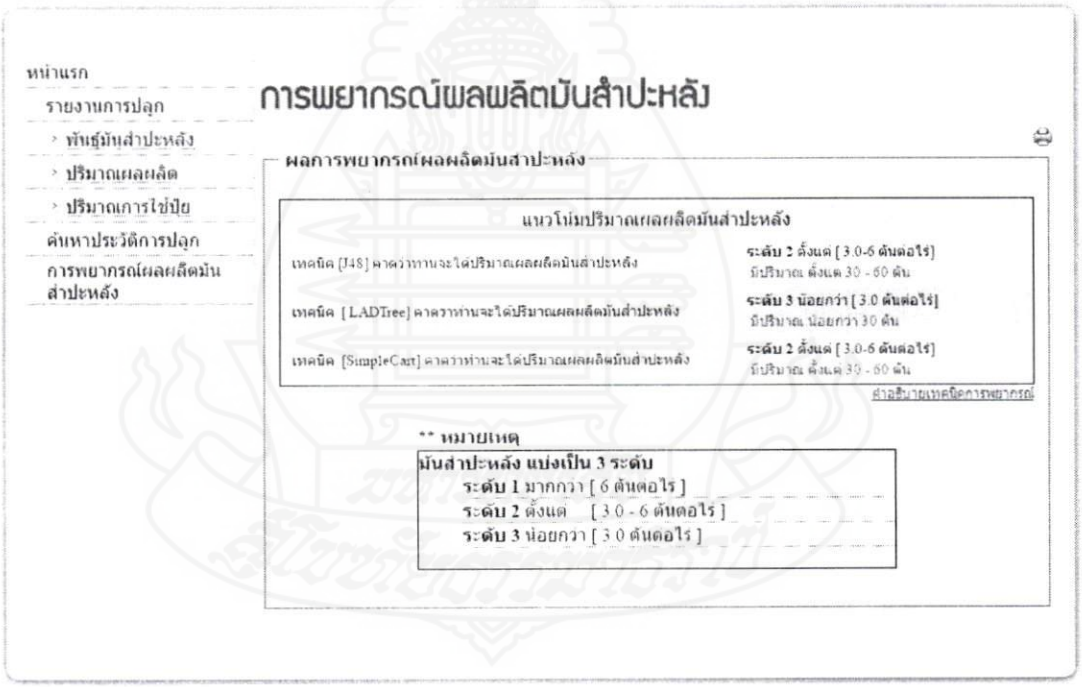
ภาพที่ 4.9 หน้าจอจัดการข้อมูลระบบ

จากภาพที่ 4.9 จะแสดงเมนูการจัดการระบบ โดยสามารถเพิ่มเติม แก้ไข ปรับลด พันธุ์มันสำประหลังได้ จัดการข้อมูลพื้นที่ปลูก บันทึกประวัติการปลูก รายการปลูก ค้นหาประวัติการปลูกได้เป็นต้น

3.3 จากภาพที่ 4.10 หน้าจอการพยากรณ์ผลผลิตมันสำปะหลัง โดยทำการเลือกจังหวัดที่ปลูกมันสำปะหลัง พันธุ์ที่ใช้ ระยะเวลาปลูก ปุ๋ยที่ใช้ จากนั้นทำการคลิกปุ่มพยากรณ์ ระบบก็จะทำการพยากรณ์ผลผลิตมันสำปะหลัง ดังแสดงในภาพที่ 4.11



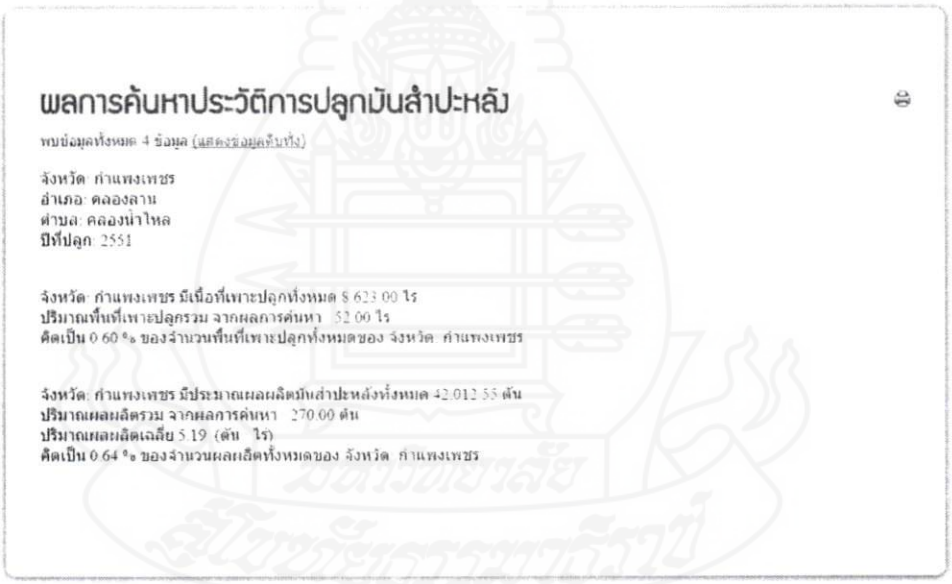
ภาพที่ 4.10 หน้าจอการพยากรณ์ผลผลิตมันสำปะหลัง



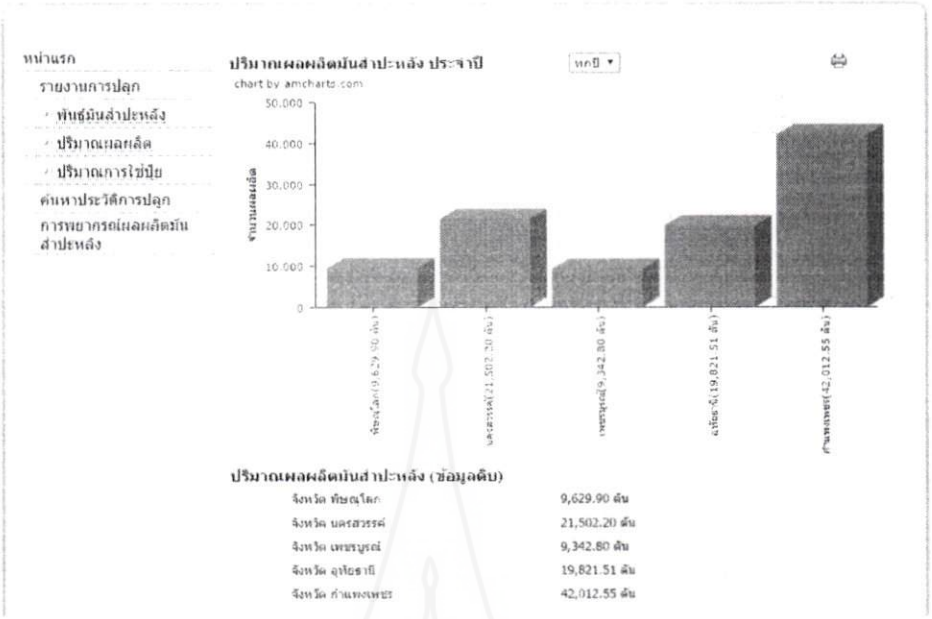
ภาพที่ 4.11 หน้าจอผลการพยากรณ์ผลผลิตมันสำปะหลัง

จากภาพที่ 4.11 แสดงผลการพยากรณ์ผลผลิตมันสำประหลังออกมาเป็นค่าระดับตามที่ได้กล่าวไว้แล้วในบทที่ 3 โดยจะแสดงผลการพยากรณ์จากอัลกอริทึมทั้ง 3 ตัว ได้แก่ อัลกอริทึม J8 อัลกอริทึม LAD Tree และอัลกอริทึม Simple Cart

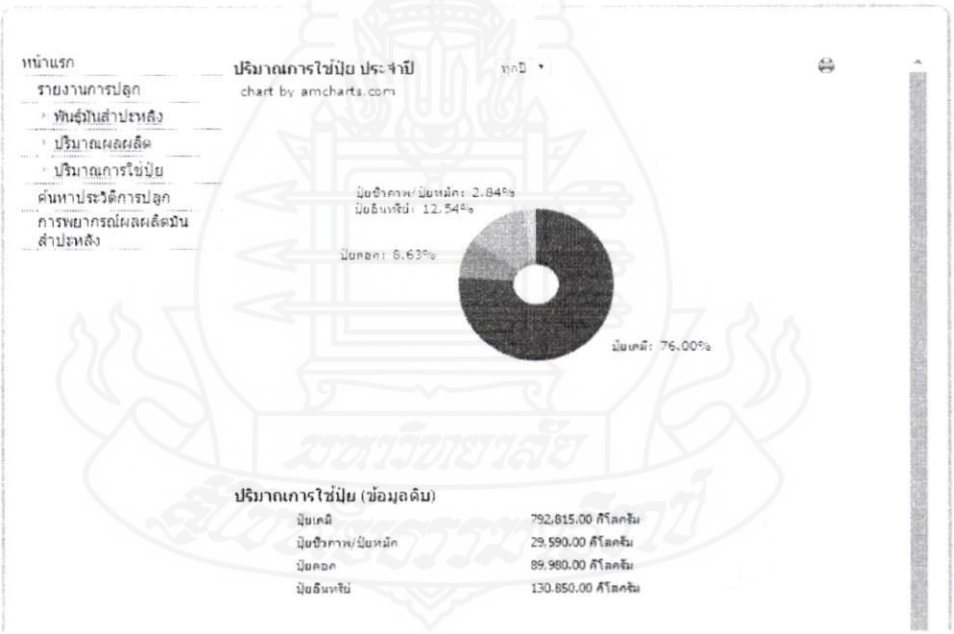
ภาพที่ 4.12 หน้าจอค้นหาประวัติการปลูก



ภาพที่ 4.13 หน้าจอผลค้นหาประวัติการปลูก



ภาพที่ 4.14 หน้าจอรายงานปริมาณผลผลิตมันสำปะหลัง



ภาพที่ 4.15 หน้าจอรายงานปริมาณการใช้ปุ๋ย

แบบประเมินความพึงพอใจการใช้งานระบบ
พยากรณ์ผลผลิตมันสำปะหลัง และระบบสืบค้น
ข้อมูล

1. เพศ
- ชาย
- หญิง

2. อายุ
- ต่ำกว่า 20 ปี
- 20-30 ปี
- 31-40 ปี
- 41-50 ปี
- มากกว่า 50 ปี

3. ระดับการศึกษา
- ต่ำกว่าปริญญาตรี
- ปริญญาตรี
- สูงกว่าปริญญาตรี
- อื่นๆ

4. สถานะของผู้ใช้งาน
- ผู้ดูแลระบบ
- ผู้ใช้ทั่วไป
- เจ้าหน้าที่ของสำนักงานเกษตร

5. ความพึงพอใจในการสืบค้นข้อมูล

	น้อยที่สุด	น้อย	ปานกลาง	มาก	มากที่สุด
ความแม่นยำในการค้นหา					
ความรวดเร็วในการแสดงผล					
ระบบมีความทันสมัย					
ความเชื่อใจใน					

ภาพที่ 4.16 หน้าจอแบบประเมินความพึงพอใจการใช้งานระบบ

3.4 เข้ามาตอบคำถาม แล้วนำค่าคะแนนที่ได้จากประเมินไปหาค่าร้อยละของความพึงพอใจ (Percentage **หน้าจอการประเมินความพึงพอใจจากภาพที่ 4.16** จะมีข้อมูลให้ผู้ใช้งานระบบ of satisfaction) แล้วนำข้อมูลที่ได้ไปพัฒนาและปรับปรุงระบบต่อไป ซึ่งแสดงในตารางที่ 4.6

ตารางที่ 4.6 สรุปผลการตอบแบบประเมินความพึงพอใจการใช้งานระบบพยากรณ์ผลผลิตมันสำปะหลัง และระบบสืบค้นข้อมูล

รายการ		จำนวน (%)
ผู้ตอบแบบประเมิน จำนวน 30 คน		
1. เพศ	ชาย	63.30
	หญิง	36.70
2. อายุ	ต่ำกว่า 20 ปี	13.35
	20 – 30 ปี	33.30
	31 – 40 ปี	33.35
	41 – 50 ปี	20.00
	มากกว่า 50 ปี	0.00
3. ระดับการศึกษา	ต่ำกว่าปริญญาตรี	65.50
	ปริญญาตรี	34.50
	สูงกว่าปริญญาตรี	0.00
	อื่นๆ	0.00
4. สถานะผู้ใช้งาน	ผู้ดูแลระบบ	7.70
	ผู้ใช้งานทั่วไป	84.60
	เจ้าหน้าที่ของสำนักงานเกษตร	7.70
5. ความพึงพอใจในการสืบค้นข้อมูล		88.00
6. ความพึงพอใจในการใช้งานระบบการพยากรณ์ผลผลิตมันสำปะหลัง		94.00

จากตารางที่ 4.6 จะพบว่ามีผู้ตอบแบบประเมินความพึงพอใจการใช้งานระบบพยากรณ์ผลผลิตมันสำปะหลัง และระบบสืบค้นข้อมูล จำนวน 30 คน มีความพึงพอใจในการสืบค้นข้อมูล 88.00% และพึงพอใจในการใช้งานระบบการพยากรณ์ผลผลิตมันสำปะหลัง 94.00% จากนั้นทำการหาค่าเฉลี่ย จึงสามารถสรุปได้ว่าผู้ใช้งานระบบพยากรณ์ผลผลิตมันสำปะหลัง และระบบสืบค้นข้อมูลมีความพึงพอใจที่ 91% ซึ่งถือว่าอยู่ในระดับที่ดีมาก

บทที่ 5
สรุปผลการวิจัย

การดำเนินการวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อและพัฒนาตัวแบบพยากรณ์ผลผลิตมันสำปะหลังโดยใช้เทคนิคการทำเหมืองข้อมูล และพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ ได้ข้อสรุปผลการวิจัย การอภิปรายผล และข้อเสนอแนะดังนี้

1. สรุปผลงานวิจัย

1.1 การพัฒนาตัวแบบการพยากรณ์

จากข้อมูลปัจจัยการผลิตมันสำปะหลังจำนวน 1,764 ระเบียน 40 แอตทริบิวต์ แล้วทำความสะอาดข้อมูลโดยทำการตัดระเบียนที่มีข้อมูลไม่สมบูรณ์ออก จะได้ข้อมูลชุดใหม่จำนวน 1,677 ระเบียน 11 แอตทริบิวต์ และผ่านการทดสอบข้อมูลทางสถิติจะได้ปัจจัยการผลิตมีความสัมพันธ์กันจำนวน 5 ตัวแปร (แอตทริบิวต์) ได้แก่ พันธุ์ที่ใช้ อายุปลูก ปริมาณปุ๋ยเคมี ปริมาณปุ๋ยอินทรีย์ และจังหวัด จากนั้นทำการกำหนดค่าระดับผลผลิตผลิตมันสำปะหลัง โดยกำหนดให้ ระดับ 1 = ผลผลิตมากกว่า 6 ตัน/ไร่ ระดับ 2 = ผลผลิตตั้งแต่ 3-6 ตัน/ไร่ และระดับ 3 = ผลผลิตตั้งแต่ 0-3 ตัน/ไร่ แล้วทำการสร้างตัวแบบการพยากรณ์ด้วยโปรแกรมเวกา (WEKA) และทำการทดสอบความแม่นยำของตัวแบบการพยากรณ์ แล้วนำตัวแบบการพยากรณ์ที่มีค่าความแม่นยำสูงสุด มาพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ ต่อไป

1.2 การออกแบบระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์

จะนำตัวแบบการพยากรณ์ที่ได้จากโปรแกรมเวกา (WEKA) ซึ่งเป็นเงื่อนไขที่ได้จากต้นไม้ตัดสินใจ มาพัฒนาเป็นระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ โดยใช้ภาษา PHP ในการพัฒนาเว็บไซต์ ใช้โปรแกรม My SQL ในการจัดการฐานข้อมูล และใช้โปรแกรม Apache เป็นเว็บเซิร์ฟเวอร์ โดยระบบสามารถการสืบค้นข้อมูลสามารถกำหนดสิทธิ์การใช้งาน ได้แก่ เจ้าหน้าที่ของสำนักงานเกษตรจังหวัดกำแพงเพชร ผู้ดูแลระบบ และผู้ใช้งานทั่วไป สามารถค้นหา ปรับปรุง แก้ไข แสดงผล และพิมพ์ข้อมูล จังหวัด อำเภอ ตำบล เนื้อที่ปลูก พันธุ์ที่ใช้ ปุ๋ย ได้เป็นต้น

ในส่วนของการประเมินความพึงพอใจผ่านเว็บไซต์ ทำการพัฒนาแบบประเมินความพึงพอใจผ่านทางเว็บไซต์ โดยผู้ใช้งานระบบ สามารถตอบแบบสอบถามความพึงพอใจในการใช้งานระบบได้ทันที

2. การอภิปรายผล

2.1 การพัฒนาตัวแบบการพยากรณ์

ตัวแบบการพยากรณ์ที่ได้จากโปรแกรมเวกา (WEKA) โดยใช้เทคนิคการทำเหมืองข้อมูลจะใช้การจำแนกประเภท (classification) และใช้อัลกอริธึมจำนวน 5 ตัว ได้แก่ J48, Random Tree, SimpleCart, Naïve Bayes, และ LAD Tree เพื่อสร้างต้นไม้ตัดสินใจ (Decision Tree) จากนั้นทำการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิค Cross-validation Test โดยทำการแบ่งข้อมูลออกเป็น 5 ส่วน (5-fold cross-validation) และ 10 ส่วน (10-fold cross-validation) ซึ่งผลการทดสอบพบว่าทั้งสองเทคนิคยังให้ค่าความแม่นยำสูงสุดเพียง 70.96% ดังนั้นผู้วิจัยจึงทำการปรับปรุงวิธีการทดสอบให้มีความแม่นยำที่ดีขึ้น โดยการทำการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) จำนวน 5 ชุด ผลทดสอบพบว่ามีความแม่นยำเพิ่มขึ้นสูงสุดถึง 80.12% โดยสรุปได้ดังนี้ อัลกอริทึมของชุดข้อมูล Test Set ที่ได้ความแม่นยำสูงสุดจำนวน 3 อัลลอริทึม ได้แก่ ชุดข้อมูลที่ 5 อัลกอริทึม J48 ให้ค่าความแม่นยำที่ 75.64% ชุดข้อมูลที่ 5 ของอัลกอริทึม Simple Cart ให้ค่าความแม่นยำที่ 80.12% และชุดข้อมูลที่ 4 ของอัลกอริทึม LAD Tree ให้ค่าความแม่นยำที่ 78.68% ส่วนอัลลอริทึมที่เหลือได้แก่ อัลกอริทึม Random Tree ให้ค่าความแม่นยำที่ 66.85% และอัลกอริทึม Naïve Bayes ให้ค่าความแม่นยำที่ 67.91%
ซึ่งให้ค่าความแม่นยำที่ต่ำ จึงไม่นำมาใช้ในการสร้างตัวแบบการพยากรณ์ จากนั้นจึงดำเนินการนำอัลกอริธึมที่ให้ความแม่นยำดีที่สุดไปใช้ในการออกแบบและพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ ต่อไป

ซึ่งสอดคล้องกับงานวิจัยของปฏิยากร โคผดุง (2551) ได้นำเสนองานวิจัยเรื่องต้นไม้ตัดสินใจสำหรับการวิเคราะห์ภาวะตัวรับฮอร์โมนของผู้ป่วยมะเร็งเต้านมและอัตราการอยู่รอดชีวิตของผู้ป่วยมะเร็งเต้านมในโรงพยาบาลขอนแก่น โดยการจำแนกระดับภาวะตัวรับฮอร์โมนสาหรับแนวทางการรักษาผู้ป่วยมะเร็งเต้านม และใช้เป็นระบบสนับสนุนการตัดสินใจด้านคลินิก โดยการประยุกต์ใช้อัลกอริทึม J48 ในโปรแกรมเวกา สร้างตัวแบบต้นไม้ตัดสินใจโดยใช้ทฤษฎีการทำเหมืองข้อมูลกับข้อมูลผู้ป่วยจำนวน 57 คน ซึ่งทำการวิเคราะห์ข้อมูลจำนวน 57 instance มีแอททริบิวต์จำนวน 6 แอททริบิวต์ สำหรับการประเมินความถูกต้องของตัวแบบที่ใช้ทำนายใช้วิธี 10-fold cross validation และวิธี Train test พบว่าวิธี 10-fold cross validation มีความถูกต้อง 57.89 % และวิธี Train test มีความถูกต้อง 66.67% ซึ่งวิธีนี้จะมีประสิทธิภาพสูงสุด ซึ่งผู้วิจัยได้นำเอาเทคนิค 10-fold cross validation และวิธี Train test มาปรับใช้ในงานวิจัยพัฒนาตัวแบบพยากรณ์ผลผลิตมันสำปะหลังโดยใช้เทคนิคการทำเหมืองข้อมูล และพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์

2.2 การออกแบบระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์

การออกแบบระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์จะนำเงื่อนไขที่ได้จากต้นไม้ตัดสินใจ ของอัลกอริทึม J48

อัลกอริทึม Simple Cart และอัลกอริทึม LAD Tree มาพัฒนาเป็นระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ โดยใช้ภาษา PHP ในการพัฒนาเว็บไซต์ ใช้โปรแกรม My SQL ในการจัดการฐานข้อมูล และใช้โปรแกรม Apache เป็นเว็บเซิร์ฟเวอร์ และในส่วนของประเมินความพึงพอใจจากเจ้าหน้าที่สำนักงานเกษตรจังหวัดกำแพงเพชร ผู้ใช้งานทั่วไป และผู้ดูแลระบบ รวม 30 คน มีความพึงพอใจในการใช้งานระบบดังกล่าว โดยการหาค่าเฉลี่ยจากการตอบแบบประเมินความพึงพอใจผ่านเว็บไซต์คิดเป็น 91% ซึ่งถือว่าอยู่ในระดับที่ดีมาก

3. ข้อเสนอแนะ

แนวทางการพัฒนาต่อของงานวิจัยนี้ เนื่องจากเป็นการสร้างตัวแบบการพยากรณ์ด้วยเทคนิคการจำแนกประเภท (classification) ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) ไม่สามารถให้คำตอบที่เป็นค่าต่อเนื่องได้ ซึ่งจะแสดงเป็นค่าระดับ หรือ ค่าช่วง โดยอาจเลือกใช้วิธีโครงข่ายประสาท (Neural Network) ที่ให้ค่าการพยากรณ์แบบต่อเนื่องได้ และอาจพัฒนาไปเป็นแอปพลิเคชันเพื่อใช้กับสมาร์ทโฟน และแท็บเล็ต เพื่อสะดวกและคล่องตัวในการใช้งานต่อไป

ภาคผนวก

ภาคผนวก ก

(แบบประเมินความพึงพอใจการใช้งานระบบพยากรณ์ผลผลิตมันสำปะหลัง และระบบสืบค้นข้อมูล)

แบบประเมินความพึงพอใจการใช้งานระบบพยากรณ์ผลผลิตมันสำปะหลัง และระบบสืบค้นข้อมูล

1. **1. เพศ**
ทำเครื่องหมายเพียงหนึ่งช่อง
   - ◯ ชาย
   - ◯ หญิง

2. **2. อายุ**
ทำเครื่องหมายเพียงหนึ่งช่อง
   - ◯ ต่ำกว่า 20 ปี
   - ◯ 20-30 ปี
   - ◯ 31-40 ปี
   - ◯ 41-50 ปี
   - ◯ มากกว่า 50 ปี

3. **3. ระดับการศึกษา**
ทำเครื่องหมายเพียงหนึ่งช่อง
   - ◯ ต่ำกว่าปริญาตรี
   - ◯ ปริญญาตรี
   - ◯ สูงกว่าปริญญาตรี
   - ◯ อื่นๆ

4. **4. สถานะของผู้ใช้งาน**
ทำเครื่องหมายเพียงหนึ่งช่อง
   - ◯ ผู้ดูแลระบบ
   - ◯ ผู้ใช้ทั่วไป
   - ◯ เจ้าหน้าที่ของสำงานเกษตร

5. 5. ความพึงพอใจในการสืบค้นข้อมูล

ทำเครื่องหมายแถวละหนึ่งช่องเท่านั้น

	น้อยที่สุด	น้อย	ปานกลาง	มาก	มากที่สุด
ความแม่นยำในการค้นหา	◯	◯	◯	◯	◯
ความรวดเร็วในการแสดงผล	◯	◯	◯	◯	◯
ระบบมีความทันสมัย	◯	◯	◯	◯	◯
ความเข้าใจในเนื้อหา	◯	◯	◯	◯	◯
ความสะดวกในการใช้งาน	◯	◯	◯	◯	◯
ประโยชน์ที่ได้รับ	◯	◯	◯	◯	◯
ความพึงพอใจในการใช้งานระบบ	◯	◯	◯	◯	◯

6. 6. ความพึงพอใจในการใช้งานระบบการพยากรณ์ผลผลิตมันสำปะหลัง

ทำเครื่องหมายแถวละหนึ่งช่องเท่านั้น

	น้อยที่สุด	น้อย	ปานกลาง	มาก	มากที่สุด
ความแม่นยำในการพยากรณ์	◯	◯	◯	◯	◯
ความเร็วเร็วในการแสดงผล	◯	◯	◯	◯	◯
ระบบมีความทันสมัย	◯	◯	◯	◯	◯
ความเข้าใจในเนื้อหา	◯	◯	◯	◯	◯
ความสะดวกในการใช้งาน	◯	◯	◯	◯	◯
ประโยชน์ที่ได้รับ	◯	◯	◯	◯	◯
ความพึงพอใจในการใช้งานระบบ	◯	◯	◯	◯	◯

ขับเคลื่อนโดย
Google Forms

ภาคผนวก ข

หนังสือถึงสำนักงานเกษตรจังหวัดกำแพงเพชร เพื่อขอความอนุเคราะห์เก็บข้อมูล

ที่ ศธ 0522.25/ 1786

มหาวิทยาลัยสุโขทัยธรรมาธิราช
ตำบลบางพูด อำเภอปากเกร็ด
จังหวัดนนทบุรี 11120

19 ธันวาคม 2555

เรื่อง ขอความอนุเคราะห์ให้นักศึกษาเก็บข้อมูลเพื่อการวิจัย

เรียน เกษตรจังหวัดกำแพงเพชร

สิ่งที่ส่งมาด้วย โครงการวิทยานิพนธ์ จำนวน 1 แผ่น

ด้วยนายปรีชา ลิ้มตระกูล นักศึกษาระดับบัณฑิตศึกษา แขนงวิชาสารสนเทศและการสื่อสาร สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช กำลังทำวิทยานิพนธ์ เรื่อง การพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล และงานวิจัยดังกล่าวจำเป็นต้องใช้ข้อมูลปัจจัยการผลิตต่าง ๆ ที่เกี่ยวข้องกับการปลูกมันสำปะหลังในพื้นที่จังหวัดกำแพงเพชร โดยมี รองศาสตราจารย์ ดร.วิภา เจริญภัณฑารักษ์ เป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก และอาจารย์วิทยา พรพัชรพงศ์ เป็นอาจารย์ที่ปรึกษาร่วม

ในการนี้ นักศึกษาจำเป็นจะต้องเก็บข้อมูลเพื่อการวิจัย เพื่อพยากรณ์ผลผลิตมันสัมปะหลังด้วยเทคนิคการทำเหมืองข้อมูล ตั้งแต่วันที่ 1 มกราคม - กุมภาพันธ์ 2556 จึงขอความอนุเคราะห์จากท่านในการอนุญาตให้ นายปรีชา ลิ้มตระกูล ได้ดำเนินการเก็บข้อมูลเพื่อการวิจัย ตามวัน เวลา และรายละเอียดที่นักศึกษาเสนอมาพร้อมนี้ หวังเป็นอย่างยิ่งจะได้รับความกรุณาจากท่าน และขอขอบคุณมา ณ โอกาสนี้

จึงเรียนมาเพื่อโปรดให้ความอนุเคราะห์ด้วย จะขอบคุณยิ่ง

ขอแสดงความนับถือ

(อาจารย์ปิยพร บุรารักษ์)
ประธานกรรมการประจำสาขาวิชาวิทยาศาสตร์และเทคโนโลยี

21 มก 2556

บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์และเทคโนโลยี
โทร. 02-5048192

บรรณานุกรม

บรรณานุกรม

กรมวิชาการเกษตร. (2555). "ฐานข้อมูลมันสำปะหลัง" ค้นวันที่ 19 เมษายน 2556 จาก เวิลด์ไวด์เว็บ: http://at.doa.go.th/cassava/.

เกรียงไกร พิพิธหิรัญการ. (2550). *การเปรียบเทียบประสิทธิภาพในการทำนายพฤติกรรมผู้บริโภคโดยใช้เทคนิคการทำเหมืองข้อมูลระหว่างกฎความสัมพันธ์จำแนก และต้นไม้ตัดสินใจ กรณีศึกษา: ศูนย์หนังสือมหาวิทยาลัยเกษตรศาสตร์* (สารนิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

ชลนิศา สาระ. (2550). *การจำแนกกลุ่มสถานภาพการสำเร็จการศึกษาโดยแบบจำลองต้นไม้ตัดสินใจ* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

ชฎิภกรน์ ทรายหมอ. (2557). "การพัฒนาแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นในผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล :กรณีโรงพยาบาลปทุมธานี" การประชุมเสนอผลงานวิจัยระดับบัณฑิตศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช ครั้งที่ 4.

ณัฐกรณ์ ตั้งพูนทรัพย์. (2551). *ระบบช่วยในการตัดสินใจเพื่อหาพื้นที่สร้างที่พักโดยอ้างอิงจากสถานที่ที่เหมาะสม* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ.

ดิษฐพล มั่นธรรม. (2553). *การประยุกต์อัลกอริทึมต้นไม้ตัดสินใจกับการวินิจฉัยโรคระบบทางเดินหายใจ* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ.

ธันยพัฒน์ วงศ์รัตน์. (2555). *การประยุกต์ใช้โปรแกรม SPSS 17.0 วิเคราะห์ข้อมูลทางสถิติ* กรุงเทพฯ สวัสดี ไอที.

ธีรพงศ์ สงผัด. (2556). *การทำเหมืองข้อมูลเพื่อสนับสนุนการให้บริการสารสนเทศของห้องสมุดมหาวิทยาลัยราชภัฏศรีสะเกษ* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยสุโขทัยธรรมาธิราช, นนทบุรี.

ธวัชชัย อดิเทพสถิต. (2557). "การเปรียบเทียบประสิทธิภาพในการการพยากรณ์ราคายางพาราระหว่างวิธีการวิเคราะห์การถดถอยแบบพหุคูณกับวิธีโครงข่ายประสาทเทียม" การประชุมเสนอผลงานวิจัยระดับบัณฑิตศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช ครั้งที่ 4.

ปฏิยากร โคผดุง. (2551). *ต้นไม้ตัดสินใจสำหรับการวิเคราะห์ภาวะตัวรับฮอร์โมนของผู้ป่วยมะเร็งเต้านมและอัตราการอยู่รอดชีวิตของผู้ป่วยมะเร็งเต้านมในโรงพยาบาลขอนแก่น* (ค้นคว้าศึกษาอิสระปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยขอนแก่น, ขอนแก่น.

พยูน พาณิชย์กุล. (2548). "การพัฒนาระบบดาต้าไมน์นิ่งโดยใช้ Decision Tree" โครงการพัฒนาระบบงานปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ แขนงวิทยาการสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง.

พวงทิพย์ แท่นแสง. (2550). "การทดสอบประสิทธิภาพการทำงานของอัลกอริทึมการไมนิ่งกฎสำหรับจำแนก" ปัญหาพิเศษปริญญาวิทยาศาสตรมหาบัณฑิต วิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

พิจิตรา จอมศรี. (2549). *การทำนายเนื้อหาของเว็บโดยใช้เทคนิคเหมืองข้อมูล: กรณีศึกษามหาวิทยาลัยศิลปากร* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยศิลปากร, กรุงเทพฯ.

ไพฑูรย์ จันทร์เรือง. (2550). *ระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคต้นไม้ตัดสินใจ* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. กรุงเทพฯ.

มลธิดา ฤทธิสมบูรณ์. (2550). *การพัฒนาระบบสนับสนุนการพิจารณาอนุมัติให้สินเชื่อเพื่อการเช่า ซื้อสินค้าโดยใช้ดิซิชันทรี* (ปัญหาพิเศษปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

วาสนา วงษ์กาภู. (2552). *ระบบสืบค้นงานวิจัยทางด้านจิตเวชศาสตร์โดยใช้ต้นไม้ตัดสินใจ กรณีศึกษา สถาบันจิตเวชศาสตร์สมเด็จเจ้าพระยา* (ปัญหาพิเศษปริญญาวิทยาศาสตร์มหาบัณฑิต ไม่ได้ตีพิมพ์). สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

วิภา เจริญภัณฑารักษ์. (2555). "หลักการพื้นฐานของการทำเหมืองข้อมูล" ใน *ประมวลสาระชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ = Data Warehouse, Data Mining and Business Intelligence* หน่วยที่ 8 หน้า 1-32 นนทบุรี บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช.

ศุภชัย ประคองศิลป์. (2551). *การออกแบบและพัฒนาระบบสนับสนุนการตัดสินใจในการอนุมัติลูกบ้านเข้าโครงการโดยใช้เทคนิคต้นไม้ตัดสินใจ กรณีศึกษา มูลนิธิที่อยู่อาศัยเพื่อมนุษยชาติ* (ปัญหาพิเศษปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). สาขาวิชาสถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

สุมิตรา นวลมีศรี. (2555). "การพัฒนาระบบสารสนเทศเพื่อการพยากรณ์ผู้เข้าศึกษาโดยผ่านเครือข่ายอินเทอร์เน็ต" รายงานการวิจัย มหาวิทยาลัยราชภัฏสวนสุนันทา.

เสกสรรค์ วิลัยลักษณ์. (2558). *การใช้เทคนิคการทำเหมืองข้อมูลเพื่อการพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา* (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยสุโขทัยธรรมาธิราช, นนทบุรี.

Frank, E. (2010). *Machine Learning with WEKA*. http://www.cs.waikato.ac.nz/ml/WEKA [Online].

Geoff Webb. (1999). Decision *Tree Grafting From the All-Tests- But-One Partition. In*: , San Francisco, CA.

Han and Kamber.(2006). Data *Mining Concepts and Techniques* , San Francisco, Morgan Kaufmann Publishers.

Ian H. Witten.(2011). *Eibe Frank, Mark A. Hall, "Practical Machine Learning Tools and Techniques* , 3rd Edition, Morgan Kaufman Publishers.

http://scriptslines.com/blog/decision-tree/ Retrieved January 19, 2015.

http://www.no-poor.com/SAandNetwork/nu1.htm Retrieved January 19, 2015.

http://blog.goldenhelix.com/bchristensen/cross-validation-for-genomic-prediction-in-svs/ Retrieved January 20, 2015.

ประวัติผู้วิจัย

ชื่อ	นายปรีชา ลิ้มตระกูล
วัน เดือน ปีเกิด	27 มีนาคม 2520
สถานที่เกิด	อำเภอเมือง จังหวัดกำแพงเพชร
ประวัติการศึกษา	ครุศาสตรอุตสาหกรรมบัณฑิต (วิศวกรรมไฟฟ้า)
	สถาบันเทคโนโลยีราชมงคล 2542
	วิทยาศาสตรบัณฑิต (เทคโนโลยีอุตสาหกรรม)
	วิชาเอกเทคโนโลยีคอมพิวเตอร์อุตสาหกรรม มหาวิทยาลัยราชภัฏพระนคร 2553
สถานที่ทำงาน	กรมอุทยานแห่งชาติ สัตว์ป่า และพันธุ์พืช กรุงเทพมหานคร
ตำแหน่ง	หัวหน้าฝ่ายโครงการและแผนงานวิศวกรรม