

การพัฒนาตัวแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล*

Development of a Model to Predict Cassava Yield Using Data Mining

ปรีชา ลิ้มตระกูล (Preecha Limtrakul)**

วิภา เจริญภักดิ์ (Vipa Jaroenpuntaruk)***

วิทยา พรพัชรพงศ์ (Wittaya Pornpatcharapong)****

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อพัฒนาตัวแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล 2) เพื่อพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจของผู้ใช้งานระบบผ่านเว็บไซต์

โดยใช้ข้อมูลปัจจัยการผลิตมันสำปะหลังจากสำนักงานเกษตรจังหวัดกำแพงเพชร เพื่อสร้างตัวแบบพยากรณ์ โดยใช้เทคนิคการจำแนกประเภทข้อมูล(Classification) ด้วยวิธีต้นไม้ตัดสินใจ (DecisionTree) โดยจะใช้อัลกอริทึม จำนวน 5 ตัว ได้แก่ J48, RandomTree, SimpleCart, NaïveBayes, และLADTree แล้วทำการทดสอบตัวแบบการพยากรณ์ด้วยวิธี Cross-validation Test พบว่ายังให้ค่าแม่นยำเพียง 70.96% ซึ่งผู้วิจัยเห็นว่ายังไม่ดีพอ จึงได้ปรับปรุงวิธีการทดสอบโดยการแบ่งชุดข้อมูลออกเป็น 2 ส่วนได้แก่ ข้อมูลเรียนรู้ (Training Set)และข้อมูลทดสอบ (Test Set) จำนวน 5 ชุด แล้วทำการสร้างตัวแบบการพยากรณ์ด้วยอัลกอริทึม J48, RandomTree, SimpleCart, NaïveBayes และ LADTree อีกครั้งพบว่ามีความแม่นยำสูงขึ้นกว่าวิธี Cross-validation Test ในทุกอัลกอริทึม ดังนั้นผู้วิจัยจึงเลือกตัวแบบการพยากรณ์ที่ให้ค่าความแม่นยำสูงสุดในส่วนของข้อมูลทดสอบ (Test Set) อัลกอริทึม J48 ให้ค่าความแม่นยำสูงสุดที่ 75.64% อัลกอริทึม SimpleCart ให้ค่าความแม่นยำสูงสุดที่ 80.12% และอัลกอริทึม LADTree ให้ค่าความแม่นยำสูงสุดที่ 89.55%

* บทความนี้เป็นส่วนหนึ่งของวิทยานิพนธ์ เรื่อง การพัฒนาตัวแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล

** นักศึกษาระดับปริญญาโท หลักสูตรวิทยาศาสตรมหาบัณฑิต แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยสุโขทัยธรรมาธิราช

M.Sc. Information and Communication Technology, Sukhothai Thammathirat Open University. Email : mailpreecha.lim@gmail.com

*** อาจารย์ที่ปรึกษาหลัก รองศาสตราจารย์ ดร. สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
ADVISORS: Asst.Prof. Dr. at School of Science and Technology, Sukhothai Thammathirat Open University.
Email : vipa.jar@stou.ac.th

**** อาจารย์ที่ปรึกษาร่วม ผู้จัดการทั่วไป บริษัท กลุ่มแอตวานซ์ รีเสิร์ช จำกัด
CO-ADVISORS: general manager, Advanced Research Group Co., Ltd. Email : wittayap@ar.co.th

ส่วนอัลกอริทึม RandomTree และอัลกอริทึม Naïve Bayes ให้ค่าความแม่นยำต่ำกว่า 70% จึงไม่นำมาพิจารณา

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

คำสำคัญ การทำเหมืองข้อมูล การพยากรณ์ ต้นไม้ตัดสินใจ

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, RandomTree, SimpleCart, NaïveBayes, and LAD Tree. Then tested the predictive models provide a technicalCross-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, LADTree, 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, RandomTree, SimpleCart, NaïveBayes, and LADTree Thus the selection of predictive models that provided the highest accuracy. In terms of Test Set, Algorithm J48 for maximum accuracy at 75.64%, SimpleCart with 80.12%, and LADTree with 89.55% while algorithms RandomTree 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 support 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

บทนำ

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

สำหรับประเทศไทยเป็นประเทศที่ส่งผลิตภัณฑ์มันสำปะหลังออกมากที่สุดในโลก ประเทศไทยส่งผลิตภัณฑ์มันสำปะหลังในรูปของมันอัดเม็ดไปขายมากที่สุดคือ 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 ดังนั้นหากมีการพยากรณ์หรือคาดเดาผลผลิตมันสำปะหลังได้ล่วงหน้า จะเป็นประโยชน์อย่างยิ่ง ทำให้ทราบถึงปัจจัยการผลิตที่เกี่ยวข้อง อาทิเช่น พันธุ์ที่ปลูก อายุปลูกถึงการเก็บ การใช้ปุ๋ย การใช้สารเคมีกำจัดวัชพืชและแมลงศัตรูพืช เป็นต้น ส่งผลทำให้สามารถวางแผนบริหารจัดการเพื่อลดต้นทุนในการปลูกได้

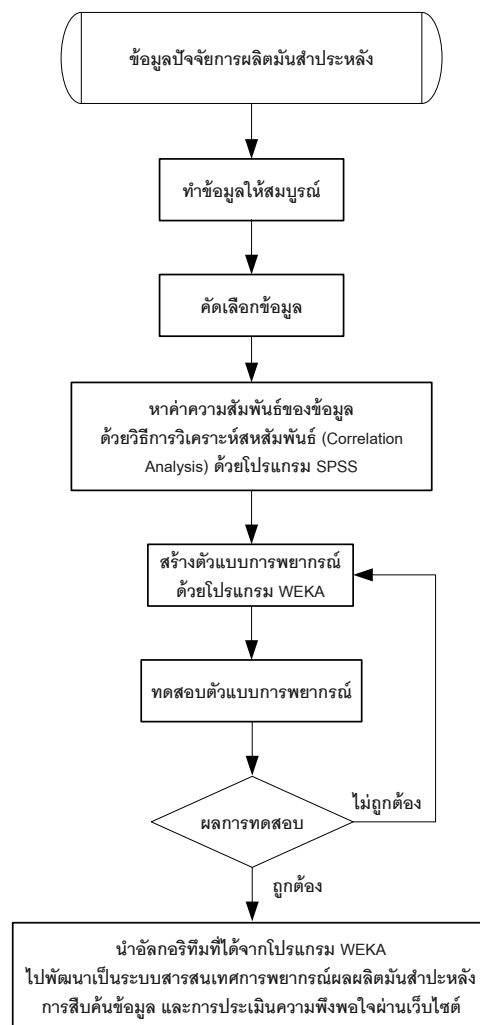
วิทยานิพนธ์นี้จึงได้เสนอการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูลโดยการจำแนกประเภท (classification) ซึ่งใช้อัลกอริทึมจำนวน 5 ตัว ได้แก่ J48, RandomTree, SimpleCart, NaïveBayes, และLADTree เพื่อสร้างต้นไม้ตัดสินใจ (Decision Tree) จากนั้นทำการทดสอบความแม่นยำในการพยากรณ์ แล้วคัดเลือกอัลกอริทึมที่ให้ความแม่นยำดีที่สุดไปใช้ในการออกแบบและพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ ต่อไป

วิธีการวิจัย

งานวิจัยนี้ได้นำเทคนิคการทำเหมืองข้อมูล เพื่อพยากรณ์ผลผลิตมันสำปะหลัง โดยได้กำหนดขั้นตอนการดำเนินการวิจัยดังภาพที่ 1

1. ข้อมูลและกลุ่มตัวอย่างที่ใช้ในการวิจัย

ข้อมูลที่ใช้ในงานวิจัยการพัฒนาตัวแบบพยากรณ์ผลผลิตมันสำปะหลังโดยใช้เทคนิคการทำเหมืองข้อมูล และพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ จะใช้ข้อมูลปัจจัยที่เกี่ยวข้องกับผลผลิตมันสำปะหลังซึ่งเป็นข้อมูลแบบทฤษฎีภูมิ ตั้งแต่ปี 2551-2555 จากสำนักงานเกษตรจังหวัดกำแพงเพชร ซึ่งประกอบด้วยข้อมูลผลผลิตมันสำปะหลังของจังหวัดในภาคเหนือที่มีการปลูกมันสำปะหลังมากที่สุด และมีสภาพภูมิอากาศที่ใกล้เคียงกับจังหวัดกำแพงเพชร จำนวน 5 จังหวัด ได้แก่ กำแพงเพชร นครสวรรค์ เพชรบูรณ์ พิษณุโลก และอุทัยธานีรวม 325 ครัวเรือน และกลุ่มตัวอย่างในการประเมินความพึงพอใจในการใช้งานระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง ประกอบด้วย เจ้าหน้าที่สำนักงานเกษตรจังหวัดกำแพงเพชร ผู้ใช้งานทั่วไปและผู้ดูแลระบบ รวม 30 คน



ภาพที่ 1 แผนผังขั้นตอนการดำเนินงานวิจัย

2. การทำข้อมูลให้สมบูรณ์

ข้อมูลปัจจัยการผลิตที่นำมาจากสำนักงานเกษตรจังหวัดกำแพงเพชร บางส่วนยังมีความไม่สมบูรณ์ จำเป็นต้องดำเนินการทำให้ข้อมูลมีความสมบูรณ์ เพื่อนำมาเข้ากระบวนการในการจัดทำเหมืองข้อมูล โดยทำการจัดข้อมูลให้เป็นรูปแบบมาตรฐานเดียวกัน (Normalization) เปลี่ยนแปลงค่าของแอตทริบิวต์ให้เป็นหน่วยนับเดียวกัน เช่น หน่วย ไร่ งาน ตารางวา ให้เป็นหน่วย ไร่ ปริมาณปุ๋ยปรับหน่วยเป็นกิโลกรัม ส่วนปริมาณผลผลิตปรับหน่วยเป็นตัน และผลผลิตให้อยู่ในหน่วยมาตรฐาน คือ ตันต่อไร่ หลังจากนั้นกำจัดข้อมูลที่ในแถวเป็นค่าว่าง (NULL) มีค่าข้อมูลเป็น 0 หรือระยะเบี่ยงที่มีค่าผลผลิตเป็น 0 ออกและเลือกแอตทริบิวต์ที่มีค่าที่สามารถนำมาใช้ประโยชน์ได้ และมีข้อมูลครบถ้วน เช่น จังหวัด อำเภอ ตำบล พันธุ์ที่ใช้ อายุปลูกถึงเก็บ เดือนเก็บผลผลิต ปีเก็บผลผลิต เนื้อที่เก็บเกี่ยว เนื้อที่เสียหาย เนื้อที่ปลูก เนื้อที่ ชนิดปุ๋ย ปริมาณปุ๋ย และปริมาณผลผลิต เป็นต้น จากนั้นทำการกำหนดรหัสข้อมูล ดังแสดงในตารางที่ 1 ดังนี้

ตารางที่ 1 แทนความหมายในแอตทริบิวต์

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	ปริมาณปุ๋ยคอกต่อเนื้อที่ใส่ปุ๋ยคอก (ไร่)

3. การคัดเลือกข้อมูล

การคัดเลือกข้อมูลปัจจัยการผลิตมันสำปะหลัง ได้พิจารณาจาก การสอบถามข้อมูลเจ้าหน้าที่สำนักงานเกษตรจังหวัดกำแพงเพชร และจากเกษตรกรผู้ปลูกมันสำปะหลัง จึงได้คัดเลือกข้อมูล แอตทริบิวต์ที่เกี่ยวข้องกับมันสำปะหลัง 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 ซึ่งผลการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) โดยใช้โปรแกรม SPSS แสดงดังตารางที่ 2

ตารางที่ 2 ผลการวิเคราะห์สหสัมพันธ์ (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 2 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

4. การแปลงข้อมูล

จากข้อมูลปัจจัยการผลิตของสำนักงานเกษตรจังหวัดกำแพงเพชร จำนวน 1,764 ไร่ปีน 40 แอตทริบิวต์ ผู้วิจัยได้ทำการจัดข้อมูลให้มีความสมบูรณ์และทำการคัดเลือกข้อมูลได้ออกมาเป็นจำนวน 1,677 ไร่ปีน 11 แอตทริบิวต์ จากนั้นทำการวิเคราะห์หาความสัมพันธ์ของปัจจัยการผลิตจำนวน 11 ปัจจัย (แอตทริบิวต์) จะได้ปัจจัยการผลิตที่มีความสัมพันธ์กันจำนวน 5 ตัวแปร (แอตทริบิวต์) ทั้งนี้งานวิจัยนี้ได้ใช้วิธีการทำเหมืองข้อมูล โดยใช้เทคนิคต้นไม้ตัดสินใจในการสร้างตัวแบบการพยากรณ์ ซึ่งผลลัพธ์ที่ได้จะแสดงออกมาเป็นค่าระดับ หรือค่าช่วง ฉะนั้น ผู้วิจัยจึงทำการกำหนดค่าระดับผลผลิตผลผลิตมันสำปะหลัง โดยอาศัยการสอบถามข้อมูลค่าเฉลี่ยผลผลิตสูงสุด และต่ำสุด ต่อพื้นที่การปลูกหนึ่งไร่ จากเจ้าหน้าที่สำนักงานเกษตรจังหวัดกำแพงเพชร และเกษตรกรผู้ปลูกมันสำปะหลัง ซึ่งได้ค่าระดับผลผลิต ดังนี้

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

5. การพัฒนาตัวแบบพยากรณ์

ผู้วิจัยได้ใช้โปรแกรมเวกา (WEKA) เพื่อสร้างตัวแบบการพยากรณ์ผลผลิตมันสำปะหลัง โดยใช้ข้อมูลปัจจัยการผลิตจำนวน 1,677 ไร่ปีน 5 แอตทริบิวต์ โดยใช้เทคนิคการจำแนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) และใช้อัลกอริทึม J48, RandomTree, SimpleCart, NaïveBayes และ LADTree ซึ่งในการนำข้อมูลสู่โปรแกรมเวกา (WEKA) ต้องทำการแปลงข้อมูลให้อยู่ในรูปของเท็กซ์ไฟล์ (Text File) ซึ่งแสดงดังภาพที่ 2



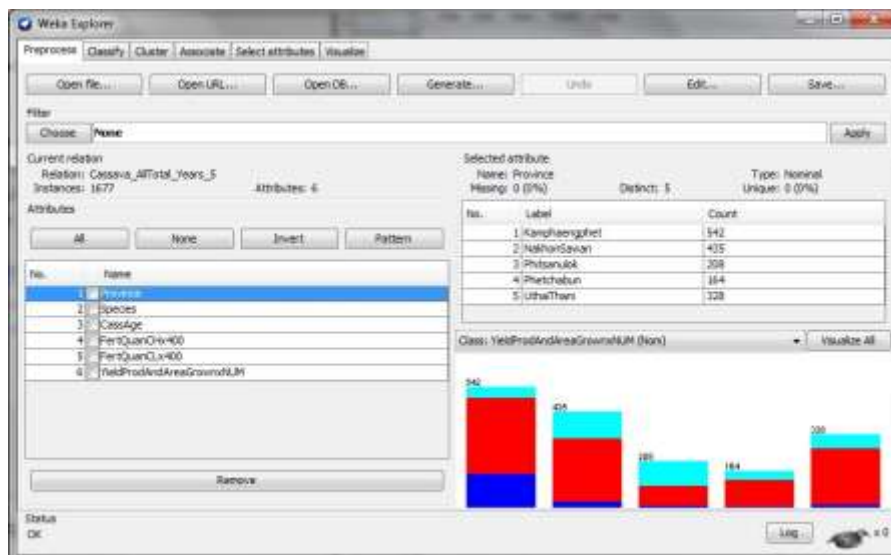
ภาพที่ 2 ไฟล์ข้อมูลที่นำเข้าสู่โปรแกรม WEKA

จากภาพที่ 2 จะพบว่าไฟล์ข้อมูลที่น่าเข้าสู่โปรแกรม WEKA จะต้องทำการจัดให้อยู่ในรูปแบบไฟล์ .arff โดยอธิบายรายละเอียดได้ดังนี้

@relation name เป็นการบอกชื่อตารางข้อมูลเชิงสัมพันธ์

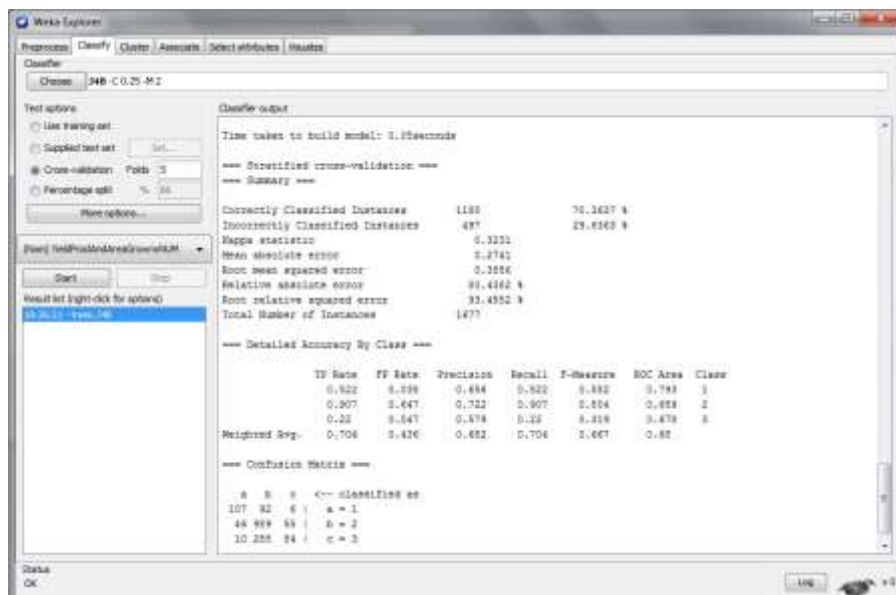
@attribute att-name type เป็นการบอกชื่อลักษณะประจำแต่ละชนิด เช่น ลักษณะประจำเก็บเป็นตัวเลข และลักษณะประจำเก็บค่าไม่ต่อเนื่อง

@data เป็นการบอกถึงแถว และตามมาด้วยข้อมูล โดยคั่นด้วย comma (,) และเมื่อเข้าข้อมูลเสร็จเรียบร้อยแล้วจะแสดงในภาพที่ 3



ภาพที่ 3 ขั้นตอนการนำเข้าข้อมูลสู่โปรแกรม WEKA

จากภาพที่ 3 แสดงจำนวนระเบียบ ได้แก่ จังหวัดกำแพงเพชร 542 ระเบียบ จังหวัดนครสวรรค์ 435 ระเบียบ จังหวัดพิษณุโลก 208 ระเบียบ จังหวัดเพชรบูรณ์ 164 ระเบียบ และจังหวัดอุทัยธานี 328 ระเบียบ จากนั้นทำการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิค Cross-validation Test โดยทำการแบ่งข้อมูลออกเป็น 5 ส่วน (5-fold cross-validation) และ 10 ส่วน (10-fold cross-validation) จากตัวแบบการพยากรณ์ที่ได้จากการใช้เทคนิคการจำแนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) ใช้อัลกอริทึม J48, RandomTree, SimpleCart, NaïveBayes และLADTree ซึ่งแสดงดังภาพที่ 4



ภาพที่ 4 ขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิค 5-fold cross-validation ของอัลกอริทึม J48

จากภาพที่ 4 แสดงขั้นตอนการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิค 5-fold cross-validation ของอัลกอริทึม J48 ซึ่งได้ค่าแม่นยำ 70.36% จากนั้นจะทำการทดสอบเช่นนี้ในทุกอัลกอริทึม และเมื่อทำการทดสอบด้วยเทคนิค 5-fold cross-validation แล้วก็จะดำเนินการทดสอบด้วยเทคนิค 10-fold cross-validation จากนั้นจะทำการปรับปรุงวิธีการทดสอบให้มีความแม่นยำที่ดีขึ้น โดยทำการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะรักษาสัดส่วนของข้อมูล และจะทำการสุ่มข้อมูลตามสัดส่วนดังกล่าว ทำให้จะได้ชุดข้อมูลที่จะทำการทดสอบจำนวน 5 ชุด แล้วทำการทดสอบจากอัลกอริทึม J48, RandomTree, SimpleCart, NaïveBayes และLADTree

6. การออกแบบระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์

เมื่อทำการทดสอบความถูกต้องของตัวแบบพยากรณ์แล้ว จะนำเอาอัลกอริทึมที่ได้จากตัวแบบพยากรณ์มาพัฒนาเป็นระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ โดยได้แบ่งขั้นตอนการดำเนินงาน ดังนี้

6.1 ศึกษาระบบงานเดิม

ในระบบงานเดิมข้อมูลที่เกี่ยวข้องกับผลผลิตมันสำปะหลัง จะเก็บอยู่ในรูปเอกสาร การจดบันทึก บางส่วนจัดเก็บในรูปแบบตารางทำการ ประกอบกับข้อมูลมีจำนวนมาก มีลักษณะการจัดเก็บข้อมูลที่แตกต่างกันไม่เป็นมาตรฐาน

เดียวกัน ทำให้ไม่สะดวกในการนำข้อมูลดังกล่าวไปใช้งาน ในการวิเคราะห์หรือคาดการณ์ผลผลิตได้อย่างสะดวก รวดเร็ว ถูกต้อง และแม่นยำ

6.2 ศึกษากระบวนการใหม่

เมื่อทราบถึงปัญหาในการที่จะนำข้อมูลที่เกี่ยวข้องกับผลผลิตมันสำปะหลังมาใช้งานแล้ว ผู้วิจัยจึงมีแนวคิดที่จะนำระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ มาใช้ในการบริหารจัดการข้อมูลที่เกี่ยวข้องกับผลผลิตมันสำปะหลังให้มีประสิทธิภาพมากยิ่งขึ้น และเกิดประโยชน์สูงสุดต่อไป

6.3 ออกแบบระบบงาน

จากข้อมูลการศึกษาระบบงานเดิม และระบบงานใหม่ พบว่ายังมิได้มีการบริหารจัดการข้อมูลจำนวนมากด้วยระบบคอมพิวเตอร์มาใช้งาน ผู้วิจัยจึงได้พัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ มาใช้งาน โดยใช้เทคนิคการทำเหมืองข้อมูลโดยการจำแนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) และใช้อัลกอริทึม J48, RandomTree, SimpleCart, NaïveBayes และLADTree มาเป็นตัวแทนพยากรณ์ โดยระบบสารสนเทศดังกล่าวจะมีผู้ใช้งาน 3 ส่วน ได้แก่ เจ้าหน้าที่ของสำนักงานเกษตรจังหวัดกำแพงเพชร ผู้ดูแลระบบ และผู้ใช้งานทั่วไป

6.4 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 และพิมพ์ข้อมูล จังหวัด อำเภอ ตำบล เนื้อที่ปลูก พันธุ์ที่ใช้ ปุ๋ย

6.5 การประเมินผลระบบ

ระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ ที่ผู้วิจัยได้พัฒนาขึ้นจะมีในของการประเมินความพึงพอใจผ่านเว็บไซต์ ซึ่งจะให้ผู้ที่ใช้ระบบสามารถประเมินความพึงพอใจในการใช้งานได้ทันทีหลังการใช้งาน ซึ่งในงานวิจัยนี้ได้มีผู้ตอบแบบประเมินความพึงพอใจการใช้งานระบบพยากรณ์ผลผลิตมันสำปะหลัง และระบบสืบค้นข้อมูล จำนวน 30 คน ได้แก่ เจ้าหน้าที่ของสำนักงานเกษตรจังหวัดกำแพงเพชร ผู้ดูแลระบบ และผู้ใช้งานทั่วไป

7. สถิติที่ใช้ในงานวิจัย

จะใช้สถิติเชิงพรรณนา (Descriptive Statistics) ในการวัดค่ากลางของข้อมูลโดยใช้ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) หรือค่าเฉลี่ย (Mean: $\bar{X}$) การหาตัวร้อยละ (Percentage) และวัดการกระจายของข้อมูลโดยใช้ ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: SD) รวมทั้งการประมวลผลข้อมูลดังนี้

7.1 ค่าเฉลี่ย (Mean)

$$\bar{X} = \frac{\sum x}{n} \quad (1)$$

เมื่อ $\bar{X}$ คือ ค่าเฉลี่ย
 $\sum X$ คือ ผลรวมของข้อมูลทั้งหมด
 N คือ จำนวนข้อมูลทั้งหมด

7.2 ร้อยละของความพึงพอใจ (Percentage of satisfaction)

$$\text{ร้อยละของความพึงพอใจ} = (B \times 100) / A \quad (2)$$

เมื่อ A คือ คะแนนเต็มจากแบบสอบถาม
 จำนวนได้จาก ค่าของระดับคะแนนที่มากที่สุด $\times$
 จำนวนข้อคำถามในแบบสอบถาม $\times$ จำนวนผู้ตอบ
 แบบสอบถาม
 B คือ คะแนนรวมจริงที่ได้จากแบบสอบถามของผู้ตอบแบบสอบถาม
 ทั้งหมด

ผลการวิจัย

1. ผลการพัฒนาตัวแบบพยากรณ์

จากข้อมูลปัจจัยการผลิตมันสำปะหลังจำนวน 1,764 ไร่เป็น 40 แอตทริบิวต์ แล้วทำความสะอาดข้อมูลโดยการตัดระยะเบี่ยงที่มีข้อมูลไม่สมบูรณ์ออก จะได้ข้อมูลชุดใหม่จำนวน 1,677 ไร่เป็น 11 แอตทริบิวต์ และผ่านการทดสอบข้อมูลทางสถิติจะได้ปัจจัยการผลิตมีความสัมพันธ์กัน จำนวน 5 ตัวแปร (แอตทริบิวต์) ได้แก่ พันธุ์ที่ใช้ อายุปลูก ปริมาณปุ๋ยเคมี ปริมาณปุ๋ยอินทรีย์ และจังหวัด จากนั้นทำการกำหนดค่าระดับผลผลิตมันสำปะหลัง ซึ่งสามารถสรุปข้อมูลระดับผลผลิตมันสำปะหลัง แสดงในตารางที่ 4.1

ตารางที่ 3 แสดงค่าระดับผลผลิต

จังหวัด	ระดับผลผลิต			จำนวนไร่
	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 ไร่เป็น 5 แอตทริบิวต์ โดยใช้เทคนิคการจำแนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) และใช้อัลกอริทึม J48, RandomTree, SimpleCart, NaïveBayes และLADTree รวมเป็น 5 อัลกอริทึม จากนั้นทำการทดสอบความแม่นยำในการพยากรณ์ด้วยเทคนิค Cross-validation Test โดยทำการแบ่งข้อมูลออกเป็น 5 ส่วน (5-fold cross-validation) และ 10 ส่วน (10-fold cross-validation) ซึ่งผลลัพธ์แสดงในตารางที่ 4

ตารางที่ 4 แสดงผลการทดสอบตัวแบบพยากรณ์ด้วยเทคนิค Cross-validation Test

ข้อมูล (ไร่เป็น)	เทคนิค (%)	J48 (%)	RandomTree (%)	SimpleCart (%)	NaiveBayes (%)	LADTree
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 ผลการทดสอบค่าความแม่นยำของการพยากรณ์ในแต่ละเทคนิค ด้วยวิธี 5 – folds Cross-Validation และ 10 – folds Cross-Validation พบว่าทั้งสองเทคนิคยังให้ค่าความแม่นยำสูงสุดเพียง 70.96% ดังนั้นผู้วิจัยจึงทำการปรับปรุงวิธีการทดสอบให้มีความแม่นยำที่ดีขึ้น โดยทำการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) โดยจะรักษาสัดส่วนของข้อมูล ดังแสดงในตารางที่ 5

ตารางที่ 5 แสดงการแบ่งสัดส่วนของข้อมูล

ข้อมูล	จำนวนไร่เป็น	อัตราส่วน (%)
Traning Set	1053	63
Test Set	624	37
Data Set	1677	100

จากตารางที่ 5 ได้กำหนดสัดส่วนของข้อมูลแล้ว ผู้วิจัยได้ทำการสุ่มข้อมูลตามสัดส่วน เพื่อสร้างชุดข้อมูลการเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) ใหม่ (Data Set = Training Set + Test Set) จำนวน 5 ชุด แล้วทำการทดสอบค่าความแม่นยำของตัวแบบพยากรณ์ โดยใช้เทคนิค การจำแนกประเภทข้อมูล (Classification) ด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) และใช้อัลกอริทึมเดิม ผลลัพธ์ที่ได้แสดงในตารางที่ 6

ตารางที่ 6 สรุปความแม่นยำในการพยากรณ์ของข้อมูลโดยวิธีการแบ่งข้อมูลทั้งชุดเป็นข้อมูลเรียนรู้ และข้อมูลทดสอบ (Data Set = Training 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
	Training 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	Training 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	Training 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	Training 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	Training 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 6 จะพบว่าการทดสอบความแม่นยำในการพยากรณ์ของชุดข้อมูลโดยการวิธีการแบ่งข้อมูลออกเป็นข้อมูลการเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set) จำนวน 5 ชุด โดยอัลกอริทึม J48 อัลกอริทึม SimpleCart และอัลกอริทึม LADTree จะให้ความแม่นยำที่ดีขึ้นอย่างชัดเจนเมื่อเทียบการทดสอบความแม่นยำแบบ Cross-validation Test ที่ทำการแบ่งข้อมูลออกเป็น 5 ส่วน (5-fold cross-validation) และ 10 ส่วน (10-fold cross-validation) ส่วนอัลกอริทึม RandomTree และอัลกอริทึม Naïve Bayes มีค่าความแม่นยำที่ต่ำใกล้เคียงกับค่าที่ได้จากการทดสอบแบบ Cross-validation Test

ตารางที่ 7 สรุปผลการทดสอบตัวแบบพยากรณ์ด้วยเทคนิคการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลเรียนรู้ (Training Set) และข้อมูลทดสอบ (Test Set) (Data Set = Training Set + Test Set)

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

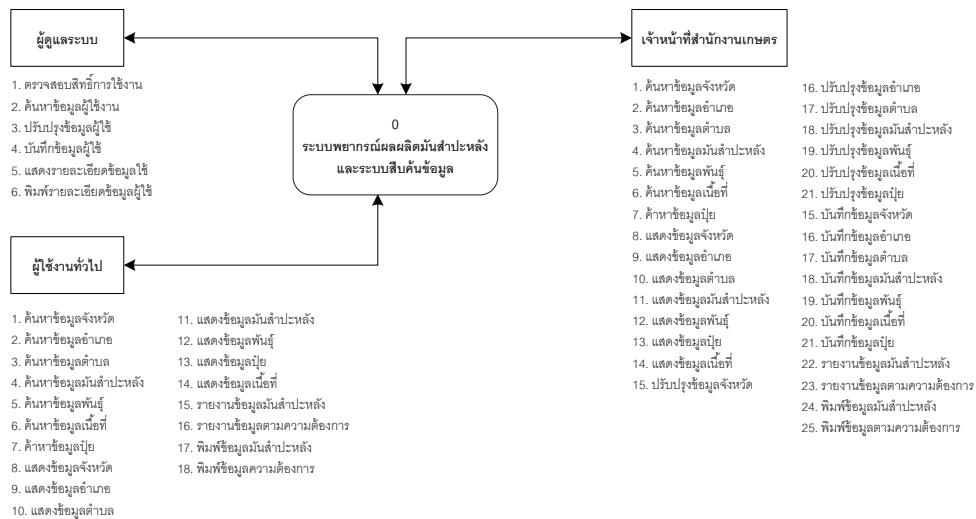
จากตารางที่ 7 จะพบว่าค่าเฉลี่ย Test Set ของอัลกอริทึม J48 ให้ค่าความแม่นยำสูงสุดที่ 75.64% อัลกอริทึม SimpleCart ให้ค่าความแม่นยำสูงสุดที่ 74.45% อัลกอริทึม LADTree ให้ค่าความแม่นยำสูงสุดที่ 73.26% ส่วนอัลกอริทึม RandomTree ให้ค่าความแม่นยำที่ 66.85% และอัลกอริทึม NaiveBayes ให้ค่าความแม่นยำที่ 67.91% ซึ่งค่าความแม่นยำที่ได้ต่ำ ผู้วิจัยจึงคัดเลือกอัลกอริทึมที่มีค่าความแม่นยำที่ดีที่สุด จำนวน 3 ตัว ได้แก่ อัลกอริทึม J48, SimpleCart และ LADTree มาเป็นตัวแบบการพยากรณ์ผลผลิตมันสำปะหลัง โดยพิจารณาจากชุดข้อมูล Test Set ดังนี้

ชุดข้อมูลที่ 5 ของอัลกอริทึม J48 ให้ค่าความแม่นยำสูงสุดที่	75.64%
ชุดข้อมูลที่ 5 ของอัลกอริทึม SimpleCart ให้ค่าความแม่นยำสูงสุดที่	80.12%
ชุดข้อมูลที่ 4 ของอัลกอริทึม LADTree ให้ค่าความแม่นยำสูงสุดที่	78.68%

3. ผลการออกแบบระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์

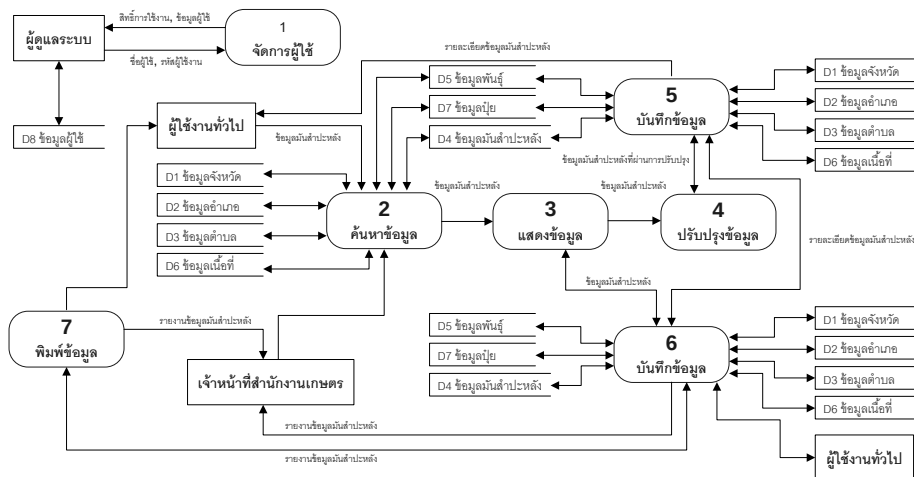
เมื่อได้ตัวแบบการพยากรณ์แล้ว จะนำเงื่อนไขที่ได้จากต้นไม้ตัดสินใจ ของอัลกอริทึม J48 จำนวน 171 บรรทัด อัลกอริทึม SimpleCart จำนวน 71 บรรทัด และอัลกอริทึม LADTree จำนวน 1321 บรรทัด มาพัฒนาเป็นระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ โดยใช้ภาษา PHP ในการพัฒนาเว็บไซต์ ใช้โปรแกรม MySQL ในการจัดการฐานข้อมูล และใช้โปรแกรม Apache เป็นเว็บเซิร์ฟเวอร์

3.1 ทำการออกแบบระบบโดยใช้แผนภาพกระแสข้อมูล (Data Flow Diagram : DFD) เป็นเครื่องมือในการวิเคราะห์ระบบ โดยแสดงการไหลของข้อมูล และความสัมพันธ์ของข้อมูลที่ได้จากการวิเคราะห์ระบบ เพื่อแสดงการนำเข้าข้อมูล (Input) และแสดงผลลัพธ์ (Output) ของระบบในการทำงานแต่ละส่วน โดยเส้นทางการไหลของข้อมูลดังกล่าว ซึ่งลำดับภาพรวมของกระบวนการทำงาน (Context Diagram) แสดงดังภาพที่ 5



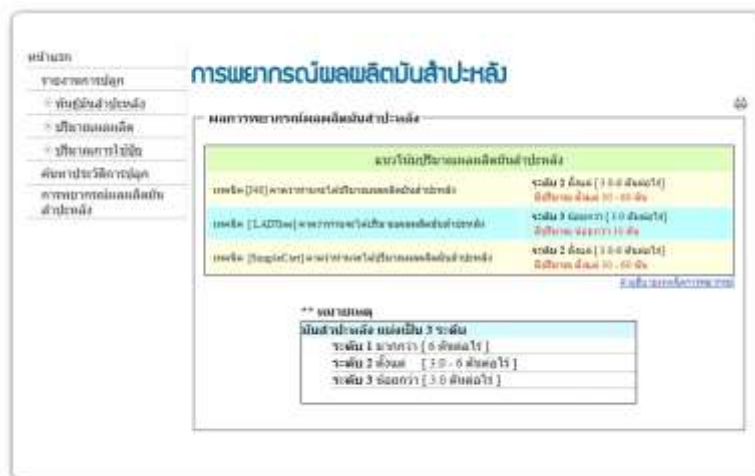
ภาพที่ 5 แผนภาพบริบทของระบบพยากรณ์ผลผลิตมันสำปะหลังและระบบสืบค้นข้อมูล

จากภาพที่ 5 แสดงเส้นทางของข้อมูลที่เข้าและออกจากแหล่งข้อมูลที่มีผลกระทบต่อระบบโดยมีผู้ที่เกี่ยวข้องกับการใช้งาน ซึ่งประกอบ เจ้าหน้าที่ของสำนักงานเกษตร ผู้ดูแลระบบ และผู้ใช้งานทั่วไป



ภาพที่ 6 แผนภาพกระแสข้อมูลที่ระดับ 0 (Data Flow Diagram Level 0)

จากภาพที่ 6 แสดงกระแสข้อมูลในส่วนของการจัดการผู้ใช้งานระบบ ได้แก่ เจ้าหน้าที่ของสำนักงานเกษตร ผู้ดูแลระบบ และผู้ใช้งานทั่วไป สามารถค้นหาข้อมูล จังหวัด อำเภอ ตำบล และเนื้อที่ปลูก แสดงข้อมูลพันธุ์ที่ใช้ ปุ๋ย เนื้อที่ปลูก ปรับปรุงข้อมูล การบันทึกข้อมูล แสดงรายงาน และสั่งพิมพ์ข้อมูลในระบบได้ เป็นต้น



ภาพที่ 7 หน้าจอผลการพยากรณ์ผลผลิตมันสำปะหลัง

จากภาพที่ 7 แสดงผลการพยากรณ์ผลผลิตมันสำปะหลังออกมาเป็นค่าระดับตามที่ได้กล่าวไว้แล้วในบทที่ 3 โดยจะแสดงผลการพยากรณ์จากอัลกอริทึมทั้ง 3 ตัว ได้แก่ อัลกอริทึม J8 อัลกอริทึม LADTree และอัลกอริทึม SimpleCart

ค้นหาประวัติการปลูก

กรุณาระบุข้อมูลการค้นหา

จังหวัด : กรุงเทพมหานคร

อำเภอ : กรุงเทพมหานคร

พื้นที่ปลูก : กรุงเทพมหานคร

ปีปลูก : 2001

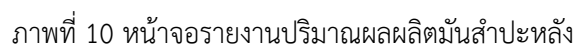
พืชปลูก : ข้าว

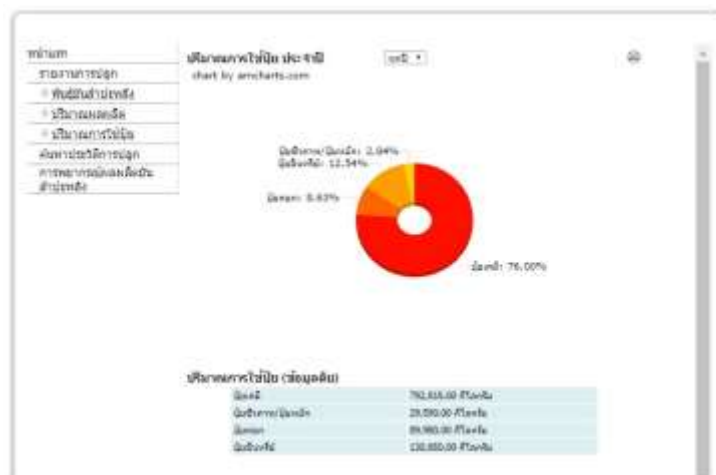
พืชปลูก : ข้าว

ผลการปลูก : ข้าว

การปลูก : ☐ ปลูก ☐ ปลูก ☐ ปลูก

ภาพที่ 8 หน้าจอค้นหาประวัติการปลูก





ภาพที่ 11 หน้าจอรายงานปริมาณการใช้ปุ๋ย

ตารางที่ 8 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. สถานะผู้ใช้งานผู้ดูแลระบบ		7.70
	ผู้ใช้งานทั่วไป	84.60
	เจ้าหน้าที่ของสำนักงานเกษตร	7.70
5. ความพึงพอใจในการสืบค้นข้อมูล		88.00
6. ความพึงพอใจในการใช้งานระบบพยากรณ์ผลผลิตมันสำปะหลัง		94.00

จากตารางที่ 8 จะพบว่าผู้ตอบแบบประเมินความพึงพอใจการใช้งานระบบพยากรณ์ผลผลิตมันสำปะหลัง และระบบสืบค้นข้อมูล จำนวน 30 คน มีความพึงพอใจในการสืบค้นข้อมูล 88.00% และพึงพอใจในการใช้งานระบบการพยากรณ์ผลผลิตมันสำปะหลัง 94.00% จากนั้นทำการหาค่าเฉลี่ยจึงสามารถสรุปได้ว่าผู้ใช้งานระบบพยากรณ์ผลผลิตมันสำปะหลัง และระบบสืบค้นข้อมูล มีความพึงพอใจที่ 91% ซึ่งถือว่าอยู่ในระดับที่ดีมาก

สรุปและอภิปรายผล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาตัวแบบการพยากรณ์ผลผลิตมันสำปะหลังด้วยเทคนิคการทำเหมืองข้อมูล และเพื่อพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจของผู้ใช้งานระบบผ่านเว็บไซต์ โดยการพัฒนาตัวแบบการพยากรณ์จะใช้ตัวแบบการพยากรณ์ที่ได้จากโปรแกรมเวกา (WEKA) โดยใช้เทคนิคการทำเหมืองข้อมูลจะใช้การจำแนกประเภท (classification) และใช้อัลกอริทึมจำนวน 5 ตัว ได้แก่ J48, RandomTree, SimpleCart, NaïveBayes, และ LADTree 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 อัลกอริทึม ได้แก่ ชุดข้อมูลที่ 5 อัลกอริทึม J48 ให้ค่าความแม่นยำที่ 75.64% ชุดข้อมูลที่ 5 ของอัลกอริทึม SimpleCart ให้ค่าความแม่นยำที่ 80.12% และชุดข้อมูลที่ 4 ของอัลกอริทึม LADTree ให้ค่าความแม่นยำที่ 78.68% ส่วนอัลกอริทึมที่เหลือได้แก่ อัลกอริทึม RandomTree ให้ค่าความแม่นยำที่ 66.85% และอัลกอริทึม NaïveBayes ให้ค่าความแม่นยำที่ 67.91% ซึ่งให้ค่าความแม่นยำที่ต่ำ จึงไม่นำมาใช้ในการสร้างตัวแบบการพยากรณ์ จากนั้นจึงดำเนินการนำอัลกอริทึมที่ให้ค่าความแม่นยำที่ดีที่สุดไปใช้ในการออกแบบและพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ ต่อไป

ซึ่งสอดคล้องกับงานวิจัยของปฎิยากร โคผดุง (2551) ได้นำเสนองานวิจัยเรื่องต้นไม้ตัดสินใจสำหรับการวิเคราะห์ภาวะตัวรับฮอร์โมนของผู้ป่วยมะเร็งเต้านมและอัตราการอยู่รอดชีวิตของผู้ป่วยมะเร็งเต้านมในโรงพยาบาลขอนแก่น โดยการจำแนกระดับภาวะตัวรับฮอร์โมนสำหรับแนวทางการรักษาผู้ป่วยมะเร็งเต้านม 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 % และ วิธี Train test มีความถูกต้อง 66.67% ซึ่งวิธีนี้จะมีประสิทธิภาพสูงสุด ซึ่งผู้วิจัยได้นำเอาเทคนิค 10-fold

cross validation และวิธี Train test มาปรับใช้ในงานวิจัยพัฒนาตัวแบบพยากรณ์ผลผลิตมันสำปะหลังโดยใช้เทคนิคการทำเหมืองข้อมูล และพัฒนาระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์

การออกแบบระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์จะนำเสนอขั้นตอนที่ได้จากต้นไม้มัดตัดสินใจ ของอัลกอริทึม J48 อัลกอริทึม SimpleCart และอัลกอริทึม LADTree มาพัฒนาเป็นระบบสารสนเทศการพยากรณ์ผลผลิตมันสำปะหลัง การสืบค้นข้อมูล และการประเมินความพึงพอใจผ่านเว็บไซต์ โดยใช้ภาษา PHP ในการพัฒนาเว็บไซต์ ใช้โปรแกรม MySQL ในการจัดการฐานข้อมูล และใช้โปรแกรม Apache เป็นเว็บเซิร์ฟเวอร์ และในส่วนของการประเมินความพึงพอใจจากเจ้าหน้าที่สำนักงานเกษตรจังหวัดกำแพงเพชร ผู้ใช้งานทั่วไป และผู้ดูแลระบบ รวม 30 คน มีความพึงพอใจในการใช้งานระบบดังกล่าวโดยการหาค่าเฉลี่ยจากการตอบแบบประเมินความพึงพอใจผ่านเว็บไซต์คิดเป็น 91% ซึ่งถือว่าอยู่ในระดับที่ดีมาก

ข้อเสนอแนะในการวิจัยพบว่า การสร้างตัวแบบการพยากรณ์ด้วยเทคนิคการจำแนกประเภท (Classification) ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) ไม่สามารถให้คำตอบที่เป็นคำตอบต่อเนื่องได้ ซึ่งจะแสดงเป็นค่าระดับ หรือ ค่าช่วง โดยอาจเลือกใช้วิธีโครงข่ายประสาท (Neural Network) ที่ให้ค่าการพยากรณ์แบบต่อเนื่องได้ และอาจพัฒนาไปเป็นแอปพลิเคชันเพื่อใช้กับสมาร์ทโฟน และแท็บเล็ต เพื่อสะดวกและคล่องตัวในการใช้งานต่อไป

เอกสารอ้างอิง

ภาษาไทย

กรมวิชาการเกษตร (2555) “ฐานข้อมูลมันสำปะหลัง” ค้นวันที่ 19 เมษายน 2556 จาก เว็บไซต์ :
<http://at.doa.go.th/cassava/>

เกรียงไกร พิพิธวิทยการ (2550) “การเปรียบเทียบประสิทธิภาพในการทำนายพฤติกรรมผู้บริโภคโดยใช้เทคนิคการทำเหมืองข้อมูลระหว่างกฎความสัมพันธ์จำแนก และต้นไม้ตัดสินใจกรณีศึกษา: ศูนย์หนังสือมหาวิทยาลัยเกษตรศาสตร์” สารนิพนธ์ปริญญาโทบริหารธุรกิจ สาขาบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ชลนิศา สาระ (2550) “การจำแนกกลุ่มสถานภาพการสำเร็จการศึกษาโดยแบบจำลองต้นไม้ตัดสินใจ”

วิทยานิพนธ์ วิทยาศาสตรมหาบัณฑิต มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ชฎิภรณ์ ทราชมอ วิชา เจริญภรณ์ธารักษ์ และวิทยา พรพชรพงศ์ (2557) “การพัฒนาแบบจำลองพยากรณ์ภาวะแทรกซ้อนของโรคอื่นในผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล :กรณีโรงพยาบาลปทุมธานี” การประชุมเสนอผลงานวิจัยระดับบัณฑิตศึกษามหาวิทยาลัยสุโขทัยธรรมาธิราช ครั้งที่ 4
ณัฐกรณ์ ตั้งพูนทรัพย์ (2551) “ระบบช่วยในการตัดสินใจเพื่อหาพื้นที่สร้างที่พักโดยอ้างอิงจากสถานที่ที่เหมาะสม” วิทยานิพนธ์ วิทยาศาสตรมหาบัณฑิต มหาวิทยาลัยเกษตรศาสตร์

- ดิษฐพล มั่นธรรม (2553) “การประยุกต์อัลกอริทึมต้นไม้ตัดสินใจกับการวินิจฉัยโรคระบบทางเดินหายใจ”
วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต มหาวิทยาลัยเกษตรศาสตร์
- ฉันทพัฒน์ วงศ์รัตน์ (2555) การประยุกต์ใช้โปรแกรม SPSS 17.0 วิเคราะห์ข้อมูลทางสถิติกรุงเทพมหานคร
สวัสดี ไอที
- ธีรพงศ์ สงผัด (2556) “การทำเหมืองข้อมูลเพื่อสนับสนุนการให้บริการสารสนเทศของห้องสมุด
มหาวิทยาลัยราชภัฏศรีสะเกษ” วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิทยาศาสตร์
และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
- ธวัชชัย อติเทพสถิต วิชา เจริญภักดิ์ และวิทยา พรพชรพงศ์ (2557) “การเปรียบเทียบประสิทธิภาพในการ
การพยากรณ์ราคาขายพาราระหว่างวิธีการวิเคราะห์การถดถอยแบบพหุคูณกับวิธีโครงข่ายประสาท
เทียม” การประชุมเสนอผลงานวิจัยระดับบัณฑิตศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช ครั้งที่ 4
- ปฎิยากร โคมดุง (2551) “ต้นไม้ตัดสินใจสำหรับการวิเคราะห์ภาวะตัวรับฮอร์โมนของผู้ป่วยมะเร็งเต้านมและ
อัตราการอยู่รอดชีวิตของผู้ป่วยมะเร็งเต้านมในโรงพยาบาลขอนแก่นค้นคว้าศึกษาอิสระ
วิทยาศาสตร์มหาบัณฑิต มหาวิทยาลัยขอนแก่น
- พยุณ พาณิษฐ์กุล (2548) “การพัฒนาระบบตัดสินใจโดยใช้ Decision Tree” โครงการพัฒนาระบบงาน
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ แขนงวิทยาการสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง
- พวงทิพย์ แท่นแสง (2550) “การทดสอบประสิทธิภาพการทำงานของอัลกอริทึมการไม่แบ่งกลุ่มสำหรับจำแนก”
ปัญหาพิเศษปริญญาวิทยาศาสตรมหาบัณฑิต วิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการ
คอมพิวเตอร์และสารสนเทศ บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ
- พิจิตรา จอมศรี (2549) “การทำนายเนื้อหาของเว็บโดยใช้เทคนิคเหมืองข้อมูล: กรณีศึกษา
มหาวิทยาลัยศิลปากร” วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต มหาวิทยาลัยศิลปากร
- ไพฑูรย์ จันทรเรือง (2550) “ระบบสนับสนุนการตัดสินใจเลือกสาขาการเรียนของนักศึกษาในระดับปริญญาตรี
โดยใช้เทคนิคต้นไม้ตัดสินใจ” วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
- มลธิดา ฤทธิสมบุญ (2550) “การพัฒนาระบบสนับสนุนการพิจารณาอนุมัติให้สินเชื่อเพื่อการเช่า ซื้อสินค้าโดยใช้
ใช้ดัชนีชี้วัด” ปัญหาพิเศษปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาระบบสารสนเทศเพื่อการจัดการ
คณะเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ
- วาสนา วงษ์กาญ (2552) “ระบบสืบค้นงานวิจัยทางด้านจิตเวชศาสตร์โดยใช้ต้นไม้ตัดสินใจกรณีศึกษา สถาบันจิต
เวชศาสตร์สมเด็จพระเจ้าพระยา” ปัญหาพิเศษปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาระบบ
สารสนเทศเพื่อการจัดการ คณะเทคโนโลยีสารสนเทศบัณฑิตวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ

- วิภา เจริญภัณฑารักษ์ (2555) “หลักการพื้นฐานของการทำเหมืองข้อมูล” ประมวลสาระชุดวิชาคลังข้อมูล
เหมืองข้อมูล และธุรกิจอัจฉริยะ หน่วยที่ 8 หน้า 1-32 นนทบุรี บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์
และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
- ศุภชัย ประครองศิลป์ (2551) “การออกแบบและพัฒนาระบบสนับสนุนการตัดสินใจในการ อนุมัติลูกบ้านเข้า
โครงการโดยใช้เทคนิคต้นไม้ตัดสินใจ กรณีศึกษา มูลนิธิที่อยู่อาศัยเพื่อ มนุษยชาติ”
ปัญหาพิเศษปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศคณะเทคโนโลยี
สารสนเทศ บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ
- สุมิตรา นวลมีศรี (2555) “การพัฒนาระบบสารสนเทศเพื่อการพยากรณ์ผู้เข้าศึกษาโดยผ่านเครือข่าย
อินเทอร์เน็ต” รายงานการวิจัย มหาวิทยาลัยราชภัฏสวนสุนันทา
- เสกสรรค์ วิสัยลักษณ์ วิภา เจริญภัณฑารักษ์ และดวงดาว วิชาดากุล (2558) “การใช้เทคนิคการทำเหมืองข้อมูล
เพื่อการพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์
วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา” วารสารวิชาการ Veridian E-Journal
Science and Technology Silpakorn Universityปีที่ 2 ฉบับที่ 2 เดือนกรกฎาคม-ธันวาคม 2558

ภาษาต่างประเทศ

- 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