

การวิเคราะห์ปัจจัยผู้บริโภคที่มีผลต่อการซื้อผลิตภัณฑ์จากดักแด้ไหม
ของชุมชนบ้านหัวสะพาน ตำบลบ้านยาง อำเภอพุทไธสง
จังหวัดบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล

Analysis of Consumer Factors towards the Purchase of Silk Pupa
in Ban Hua Saphan, Ban Yang, Phutthaisong District,
Buriram Province with Data Mining Techniques

จิรวดี โยรัมย์^{1*} สุธีรา สุนทรารักษ์² กมลรัตน์ สมใจ³

Jiravadee yoyram^{1*} suteera suntaralak² kamonrat somchai³

สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์^{1 3} และสาขาวิชาวิทยาศาสตร์สิ่งแวดล้อม
คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์²

Jirawadee.yr@bru.ac.th¹, tangmay-jaa@hotmail.com²

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยผู้บริโภคที่มีผลต่อการซื้อผลิตภัณฑ์จากดักแด้ไหมของชุมชนบ้านหัวสะพาน ตำบลบ้านยาง อำเภอพุทไธสง จังหวัดบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล โดยมีขั้นตอนการดำเนินงานโดยการเตรียมข้อมูลที่ได้จากแบบสอบถาม จำนวน 222 ชุด ซึ่งมีปัจจัยที่สนใจนำมาพิจารณาจำนวน 8 ปัจจัย ทำการวิเคราะห์ข้อมูลเพื่อสร้างแบบจำลองโดยใช้เทคนิคการทำเหมืองข้อมูลแบบการจำแนกข้อมูลด้วยฟังก์ชันไม้ตัดสินใจ โดยใช้อัลกอริทึม Random Forest ในการทดสอบ และทำการเปรียบเทียบแบบจำลองโดยทำการคัดเลือกคุณลักษณะของข้อมูล 4 วิธี ได้แก่ IR IG CFS และ SU โดยใช้โปรแกรม WEKA ผลการศึกษาพบว่า โมเดลการพยากรณ์แบบ Random Forest เมื่อใช้ร่วมกับเทคนิคการลดคุณลักษณะของข้อมูลแบบ SU ทำให้ประสิทธิภาพในการพยากรณ์เพิ่มขึ้น โดยวัดค่าความถูกต้องได้ค่าร้อยละ 98.74 ทำให้ได้ปัจจัยที่ส่งผลกระทบต่อผลิตภัณฑ์มากที่สุด จำนวน 4 ปัจจัย ได้แก่ ปัจจัยด้านผลิตภัณฑ์และราคาของผลิตภัณฑ์ ปัจจัยด้านช่องทางการจัดจำหน่าย ปัจจัยการนำเสนอและการปฏิบัติต่อผู้บริโภค และปัจจัยด้านประชากรศาสตร์ เพื่อเป็นสารสนเทศในการพัฒนาผลิตภัณฑ์จากดักแด้ไหมเพื่อยกระดับรายได้ของชุมชนบ้านหัวสะพาน ตำบลบ้านยาง อำเภอพุทไธสง จังหวัดบุรีรัมย์ต่อไป

คำสำคัญ: ปัจจัยผู้บริโภค ดักแด้ไหม การทำเหมืองข้อมูล การจำแนกข้อมูล

ABSTRACT

The purposes of the research were to analyzed of consumer factors towards the purchase of silk pupa from Ban Hua Saphan, Ban Yang, Phutthaisong District, Buriram Province with data mining techniques which the operation process by prepared Data from the 222 sets of questionnaire with 8 factors influencing to considered to create data model by the tree-based classification, Random Forest algorithm were used for testing the data and to compared the data model classification by used four methods which IR, IG, SU and CFS to by WEKA program. The results show that the Random Forest model when combined with SU data technique predict that more effectiveness. The accuracy was at 98.74 . The most for four factors were the products and price, distribution channel, factors of the presentation, the treatment of consumers and demographic factors to be the information to develop the products to raise revenue in the community.

Keyword: Consumer factors, Silk Pupa, Data mining Classification

บทนำ

ชุมชนบ้านหัวสะพาน ตั้งอยู่ในพื้นที่ตำบลบ้านยาง อำเภอพุทไธสง จังหวัดบุรีรัมย์ จัดเป็นอีกพื้นที่หนึ่งที่สามารถเป็นแหล่งท่องเที่ยวเชิงวัฒนธรรมและภูมิปัญญาที่น่าสนใจ เนื่องจากเป็นแหล่งปลูกหม่อนเลี้ยงไหมและทอผ้ารวมถึงแปรรูปผลิตภัณฑ์อย่างครบวงจร เป็นแหล่งผลิตผ้าชิ้นต้นแดนคุณภาพ และยังเป็นพื้นที่ปลูกหม่อนที่ใหญ่ที่สุดในประเทศไทยอีกด้วย โดยในปี พ.ศ. 2553 เกษตรกรได้รวมกลุ่มกันจัดตั้งเป็นสหกรณ์ผู้ปลูกหม่อนเลี้ยงไหมบ้านหัวสะพาน จำกัดขึ้น ทำให้การปลูกหม่อนเลี้ยงไหมและทอผ้าไหมถือเป็นรายได้หลักของหมู่บ้านในปัจจุบัน

ด้วยสภาพแวดล้อมทางกายภาพที่เอื้ออำนวย ในการปลูกหม่อนและภูมิปัญญาในการเลี้ยงไหม การทอผ้าที่สืบสานต่อเนื่องกันมาอย่างยาวนาน ทำให้ชุมชนบ้านหัวสะพานเป็นที่รู้จักและยอมรับจากภายนอก และจากข้อมูลข้างต้นคณะผู้วิจัยพิจารณาเห็นความสำคัญในการผลักดันให้เกิดการพัฒนาผลิตภัณฑ์ใหม่เพื่อยกระดับรายได้ ประกอบกับเป็นสินค้าทางการท่องเที่ยวของชุมชนตัวใหม่โดยมุ่งเน้นในการแปรรูปจากดักแด้ไหมเป็นหลัก โดยปัจจุบันราคาขายโดยราคาเฉลี่ยอยู่ที่ 80 บาท/กก. ซึ่งจะเห็นได้ว่าราคาที่ขายได้นั้นมีราคาต่ำกว่ามูลค่าและคุณค่าจริงของหนอนไหมดักแด้ไหมเนื่องด้วยดักแด้ไหมมีคุณค่าทางโภชนาการสูงอุดมไปด้วยโปรตีนที่เป็นประโยชน์ต่อร่างกาย (จินตนาภรณ์ วัฒนธร, 2559)

คณะผู้วิจัยจึงมีความสนใจในการวิเคราะห์ปัจจัยผู้บริโภคที่มีผลต่อการซื้อผลิตภัณฑ์จากดักแด้ไหมของชุมชนบ้านหัวสะพาน ตำบลบ้านยาง อำเภอพุทไธสง จังหวัดบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล เพื่อเป็นข้อมูลในการพัฒนาผลิตภัณฑ์จากดักแด้ไหมในการส่งเสริมการตลาดเพื่อยกระดับรายได้ของชุมชนบ้านหัวสะพาน ตำบลบ้านยาง อำเภอพุทไธสง จังหวัดบุรีรัมย์ ให้ชุมชนได้เกิดการแลกเปลี่ยนเรียนรู้และพัฒนาไปสู่การพึ่งพาตนเอง การสร้างงาน สร้างอาชีพให้กับคนในชุมชนเกิดการกระจายรายได้สู่ชุมชน

1. วัตถุประสงค์การวิจัย

วิเคราะห์ปัจจัยผู้บริโภคที่มีผลต่อการซื้อผลิตภัณฑ์จากดักแด้ไหมของชุมชนบ้านหัวสะพาน ตำบลบ้านยาง อำเภอพุทไธสง จังหวัดบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ผู้วิจัยได้นำอัลกอริทึม (Algorithm) ต้นไม้ตัดสินใจ (Decision Tree) และแรนดอมฟอเรส (Random Forest) มาทำการวิเคราะห์เชิงเปรียบเทียบ ซึ่งมีทฤษฎีดังต่อไปนี้

การทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining) เป็นการวิเคราะห์ข้อมูลเพื่อแยกประเภท จำแนกรูปแบบ และความสัมพันธ์ของข้อมูลจากฐานข้อมูลที่มีขนาดใหญ่ หรือคลังข้อมูล โดยมีวิธีต่าง ๆ หลายวิธี ซึ่งรูปแบบการทำเหมืองข้อมูลนั้นได้รวบรวมความรู้จากหลายแขนงเข้าไว้ด้วยกันประกอบด้วย การเรียนรู้ของเครื่อง (Machine Learning) ร่วมกับวิทยาศาสตร์สารสนเทศ (Information Science) สถิติ (Statistics) เทคนิคการสร้างภาพ (Visualization Techniques) ทำให้ค้นพบรูปแบบความรู้ใหม่ ๆ และสามารถนำเสนอความรู้ในลักษณะที่เข้าใจได้ง่าย ในอีกความหมายหนึ่งการทำเหมืองข้อมูลคือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อสกัดสารสนเทศ รูปแบบ และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลขนาดใหญ่ (สุรวุฑ ศรีเปารยะ และ สายชล สันสมบุรณ์ทอง, 2560)

ต้นไม้ตัดสินใจ (Decision Tree)

เป็นเทคนิคหนึ่งของ Classification ที่ได้รับความนิยมนำมาใช้ ในการแบ่งประเภทหรือแยกหมวดหมู่ของข้อมูลโดยมีลักษณะคล้ายโครงสร้างต้นไม้ โดยแสดงการเก็บคุณสมบัติของคุณลักษณะ แสดงเงื่อนไขการทดสอบ โดยแสดงกลุ่มข้อมูลที่กำหนดไว้ซึ่งจะมีการเรียนรู้จากข้อมูลการสอนในระบบ (Training Data) และใช้การคำนวณค่า Information Gain สำหรับแต่ละคุณลักษณะโดยแบ่งข้อมูลตามลำดับ แต่ข้อจำกัด คือ ไม่สามารถจำแนกข้อมูลได้อย่างต่อเนื่องจะการจำแนกข้อมูลที่มีอยู่ในกฎเท่านั้น ส่วนแรนดอมฟอเรส (Random Forest) เป็นโครงสร้างข้อมูลแบบต้นไม้ตัดสินใจ (Decision Tree) โดยจะใช้วิธีการสุ่มตัวแปรแล้วนำผลลัพธ์ที่ได้ข้อมูลที่ซ้ำกันออก ซึ่งวิธีการนี้ช่วยจำแนกข้อมูลได้อย่างต่อเนื่อง และผลการพยากรณ์มีความ

แม่นยำและลดการความซ้ำซ้อนของข้อมูลทำให้สามารถสร้างตัวทำนายที่เหมาะสมและแม่นยำ(สายชล สินสมบูรณ์ทอง, 2558)

สำหรับการวิจัยที่เกี่ยวข้อง มีดังนี้

วันวิสาข์ ชนะประเสริฐ (2560) ได้ศึกษาเทคนิคการทำเหมืองข้อมูลด้วยอัลกอริทึม(Algorithm) ต้นไม้ตัดสินใจ (Decision Tree) และนำมาใช้สำหรับทำนายอาชีพของนักศึกษาที่จบสาขาโบราณคดี มีค่าผลเฉลี่ยความแม่นยำร้อยละ 80

อัจฉิมา มณฑพาพันธ์ (2562) ได้ศึกษาเปรียบเทียบวิธีการคัดเลือกคุณลักษณะที่สำคัญในการปรับปรุงการพยากรณ์มะเร็งเต้านม โดยใช้วิธีการคัดเลือกคุณลักษณะจากเทคนิคต่าง ๆ จำนวน 7 เทคนิค ได้แก่ เทคนิค Correlation Based Feature Selection เทคนิค Information Gain เทคนิค Gain Ratio เทคนิค Chi-Square เทคนิค Forward Selection เทคนิค Backward Elimination และเทคนิค Evolutionary Selection หลังจากคัดเลือกคุณลักษณะที่สำคัญจึงนำผลที่ได้จากแต่ละเทคนิคมาคำนวณหาค่าประสิทธิภาพในการพยากรณ์การเป็นมะเร็งเต้านมโดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ผลการทดลองพบว่าร้อยละของความถูกต้องในการพยากรณ์การเป็นมะเร็งเต้านม จากจำนวนคุณลักษณะของข้อมูลทั้งหมด 30 คุณลักษณะเท่ากับ 91.39 ขณะที่เทคนิค Evolutionary Selection ให้ผลดีที่สุดโดยสามารถลดคุณลักษณะที่สำคัญเหลือเพียง 16 คุณลักษณะ และให้ผลการวัดค่าความถูกต้องในการพยากรณ์ได้ดีถึงร้อยละ 95.26

วิธีดำเนินการวิจัย

ขั้นตอนการดำเนินการวิจัยประกอบด้วย 5 ขั้นตอน ดังนี้

1. กลุ่มเป้าหมาย ได้แก่

ผู้บริโภครั่วไปจังหวัดบุรีรัมย์ที่เคยรับประทานผลิตภัณฑ์จากดักแด่ ซึ่งไม่ทราบจำนวนที่แน่นอนคำนวณขนาดของกลุ่มตัวอย่างที่ระดับความเชื่อมั่นร้อยละ 95 ได้กลุ่มตัวอย่างเท่ากับ 475 คน สุ่มเพื่อความสะดวก 15 ตัวอย่าง รวมเป็นขนาดตัวอย่างทั้งสิ้นจำนวน 500 ตัวอย่างทำการสุ่มโดยกำหนดกลุ่มตัวอย่างตามตารางของ ทาโร ยามาเน่ (Yamane, 1973 อ้างอิงใน ประสิทธิ์ สุวรรณรักษ์, 2555) ที่ระดับความเชื่อมั่น 95% ค่าความคลาดเคลื่อน $\pm 5\%$ ได้ขนาดกลุ่มตัวอย่างจำนวน 222 คน แล้วทำการจับฉลากเพื่อเลือกพื้นที่ชุมชนในเขตจังหวัดบุรีรัมย์

2. วิธีดำเนินการวิจัย

วิเคราะห์ปัจจัยผู้บริโภคที่มีผลต่อการซื้อผลิตภัณฑ์จากดักแด่ใหม่ของชุมชนบ้านหัวสะพาน ตำบลบ้านยาง อำเภอพุทไธสง จังหวัดบุรีรัมย์ด้วยเทคนิคการทำเหมืองข้อมูลเนื่องจากการทำเหมืองข้อมูลเป็นเทคนิคเพื่อค้นหารูปแบบของความรู้จากฐานข้อมูลโดยอัตโนมัติ เพื่อนำข้อมูลที่มีอยู่มาวิเคราะห์แล้วดึงความรู้หรือสิ่งสำคัญออกมาเพื่อใช้ในการวิเคราะห์ข้อมูลแยกประเภท จำแนกรูปแบบ และความสัมพันธ์ของข้อมูล จากคลังข้อมูลและนำเสนอสารสนเทศที่ได้ไปใช้ในการตัดสินใจ หากมีการเพิ่มข้อมูลในภายหลังก็สามารถวิเคราะห์รูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลได้ โดยการเตรียมข้อมูล หลังจากนั้นหาโมเดลในการพยากรณ์ข้อมูล และทำการคัดเลือกคุณลักษณะของข้อมูลเพื่อเพิ่มประสิทธิภาพให้โมเดลที่ได้ เมื่อได้คุณลักษณะของข้อมูลแล้วนำคุณลักษณะมาจัดลำดับความสำคัญ ดังนี้

2.1 การเตรียมข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลที่ได้จากแบบสอบถาม สอบถามบริโภคทั่วไปเกี่ยวกับปัจจัยที่มีผลต่อการซื้อผลิตภัณฑ์จากดักแด่ใหม่ของชุมชนบ้านหัวสะพาน ตำบลบ้านยาง อำเภอพุทไธสง จังหวัดบุรีรัมย์ จำนวน 8 ปัจจัย ประกอบด้วย ปัจจัยด้านประชากรศาสตร์ ปัจจัยด้านผลิตภัณฑ์ ปัจจัยด้านราคา ปัจจัยด้านช่องทางการจัดจำหน่าย ปัจจัยด้านการส่งเสริมทางการตลาด ปัจจัยด้านบุคคล ปัจจัยด้านกระบวนการตัดสินใจ และปัจจัยด้านการนำเสนอลักษณะทางกายภาพ

2.2 การสร้างโมเดลในการพยากรณ์ข้อมูล

การพยากรณ์ข้อมูลใช้การวัดค่าประสิทธิภาพจากเทคนิคในการจำแนกประเภทข้อมูล 5 กลุ่มเทคนิคได้แก่ กลุ่มของ Bayes (Naïve Bayes, Bayes Network) กลุ่มของ Function (LivSVM, MLP, SMO) กลุ่มของ Lazy (IBk, KStar กลุ่มของ Rules (DecisionTable, JRIP, OneR, PART) และกลุ่มของ Tree (J48, Random Forest, RandomTree, REFTree)

2.3 การคัดเลือกคุณลักษณะของข้อมูล

มีปัจจัยที่สนใจทั้งหมด 8 ปัจจัย นำมาลดมิติของข้อมูล หรือการกำหนดลำดับความสำคัญของคุณลักษณะของข้อมูลใหม่เพื่อเพิ่มประสิทธิภาพการพยากรณ์โดยการคัดเลือกคุณลักษณะของข้อมูล 4 วิธี ได้แก่ IR (OneR Attribute Avaluation), IG (Information Gain), SU (Symmetrical Uncertainty) และ CFS (Correlation-based Feature Selection)

2.4 การจัดลำดับความสำคัญของคุณลักษณะ

ในแต่ละเทคนิคของการคัดเลือกคุณลักษณะของข้อมูลจะได้ผลลัพธ์ของปัจจัยที่มีความสำคัญที่ไม่เท่ากัน ดังนั้นต้องนำปัจจัยผลลัพธ์ที่ได้มาจัดลำดับความสำคัญใหม่โดยดูว่าจากทั้ง 4 เทคนิคนั้น เหตุปัจจัยใดที่ถูกกำหนดให้มีความสำคัญมากที่สุด

2.5 การวัดค่าประสิทธิภาพ

ขั้นตอนการเปรียบเทียบโมเดลและวิธีเลือกคุณลักษณะข้อมูลพิจารณาจากค่าความถูกต้อง (Accuracy) ความแม่นยำ (Precision) และค่าการที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE)

3. เครื่องมือที่ใช้ในการวิจัย

แบบสอบถามที่ผ่านการเสนอต่อผู้เชี่ยวชาญเพื่อตรวจสอบความครบถ้วนและความสอดคล้องของเนื้อหาของแบบสอบถามที่ตรงกับเรื่องที่จะศึกษา และ มีการตรวจสอบความเชื่อมั่นจากการเก็บแบบสอบถามโดยนำไปทดสอบใช้กับกลุ่มทดลองจำนวน 30 ชุด พบว่า มีค่าความเชื่อมั่นรวม เท่ากับ 0.982

4. การวิเคราะห์ข้อมูล

วิเคราะห์ข้อมูลโดยใช้เทคนิคการทำเหมืองข้อมูลโดยการคัดเลือกคุณลักษณะข้อมูล (Feature Selection) 4 วิธี ได้แก่ IR IG SU และ CFS โดยใช้โปรแกรม WEKA

ผลการวิจัยและอภิปรายผล

วิเคราะห์ปัจจัยผู้บริโภคที่มีผลต่อการซื้อผลิตภัณฑ์จากผักแต่ใหม่ของชุมชนบ้านหัวสะพาน ตำบลบ้านยาง อำเภอพุทไธสง จังหวัดบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล มีผลการวิจัยดังนี้

1.1 การเตรียมข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลที่ได้จากแบบสอบถาม สอบถามผู้บริโภคทั่วไปเกี่ยวกับวิเคราะห์ปัจจัยผู้บริโภคที่มีผลต่อการซื้อผลิตภัณฑ์จากผักแต่ใหม่ของชุมชนบ้านหัวสะพาน ตำบลบ้านยาง อำเภอพุทไธสง จังหวัดบุรีรัมย์ จำนวน 222 ชุด จากข้อมูลประกอบด้วยปัจจัยที่มีผลต่อการซื้อ ดังนี้

ตารางที่ 1 ปัจจัยผู้บริโภคที่มีผลต่อการซื้อผลิตภัณฑ์จากผักแต่ใหม่

ปัจจัย	คำอธิบาย
1	ปัจจัยด้านประชากรศาสตร์
2	ปัจจัยด้านผลิตภัณฑ์
3	ปัจจัยด้านราคา
4	ปัจจัยด้านช่องทางการจัดจำหน่าย
5	ปัจจัยด้านการส่งเสริมทางการตลาด
6	ปัจจัยด้านบุคคล
7	ปัจจัยด้านกระบวนการตัดสินใจ
8	ปัจจัยด้านการนำเสนอลักษณะทางกายภาพ

1.2 การสร้างโมเดลในการพยากรณ์ข้อมูล

การพยากรณ์ข้อมูลใช้การวัดค่าประสิทธิภาพจากเทคนิคในการจำแนกประเภทข้อมูล 5 กลุ่มเทคนิคได้แก่ กลุ่มของ Bayes (Naïve Bayes, Bayes Network) กลุ่มของ Function (LivSVM, MLP, SMO) กลุ่มของ Lazy (IBk, KStar) กลุ่มของ Rules (DecisionTable, JRIP, OneR, PART) และกลุ่มของ Tree (J48, Random Forest, RandomTree, REFTree)

การเปรียบเทียบประสิทธิภาพโมเดลพยากรณ์ จากเทคนิคการจำแนกประเภทข้อมูล 5 กลุ่มเทคนิค นั้นพบว่าโมเดลที่มีประสิทธิภาพความแม่นยำในการจำแนกประเภทข้อมูลมากที่สุดคือโมเดลที่ได้จากกลุ่ม Tree ได้แก่ Random Forest มีความถูกต้องในการพยากรณ์ 98.71% หลังจากนั้นนำโมเดลการพยากรณ์แบบ Random Forest ไปเพิ่มประสิทธิภาพความแม่นยำในการพยากรณ์ให้มากยิ่งขึ้น โดยทำการคัดเลือกคุณลักษณะของข้อมูล เพื่อให้ได้คุณลักษณะข้อมูลที่สำคัญในการพยากรณ์ ดังตารางที่ 2

ตารางที่ 2 : ผลการวัดค่าประสิทธิภาพโมเดลพยากรณ์

Classifiers	10- fold cross validation		
	Accuracy	Precision	RMSE
Bayes			
- Naïve Bayes	69.32 %	0.538	0.1594
- Bayes Network	69.75 %	0.773	0.1159
Function			
- LivSVM	80.31 %	0.783	0.1046
- MLP	76.45 %	0.726	0.1021
- SMO	57.08%	0.428	0.1608
Lazy			
- IBk	98.27%	0.983	0.0298
- KStar	93.63%	0.935	0.0547
Rules			
- DecisionTable	78.86%	0.785	0.132
- JRIP	86.55%	0.861	0.0814
- OneR	73.37%	0.637	0.1216
- PART	90.20%	0.900	0.0636
Tree			
- J48	89.67%	0.891	0.0648
- <u>Random Forest</u>	<u>98.71%</u>	<u>0.987</u>	<u>0.0369</u>
- RandomTree	98.16%	0.0309	0.9820
- REFTree	85.97%	0.847	0.0775

1.3 การคัดเลือกคุณลักษณะของข้อมูล

เป็นวิธีที่ใช้เพื่อลดมิติของข้อมูล หรือการกำหนดลำดับความสำคัญของคุณลักษณะของข้อมูลใหม่เพื่อเพิ่มประสิทธิภาพการพยากรณ์ และได้ปัจจัยที่มีอำนาจในการจำแนกผลลัพธ์มากที่สุด ในการทดลองนี้เปรียบเทียบการคัดเลือกคุณลักษณะของข้อมูล 4 วิธี ได้แก่ IR IG CFS และ SU หลังจากทดลองนำโมเดลการพยากรณ์แบบ Random Forest มาเพิ่มประสิทธิภาพในการพยากรณ์นั้นพบว่า โมเดลการพยากรณ์แบบ Random Forest เมื่อใช้ร่วมกับเทคนิคการลดคุณลักษณะของข้อมูลแบบ SU ทำให้ประสิทธิภาพในการพยากรณ์เพิ่มขึ้นดังตารางที่ 3

ตารางที่ 3 : ผลการเปรียบเทียบประสิทธิภาพของเทคนิคการเลือกคุณลักษณะของข้อมูล เมื่อใช้กับโมเดลการพยากรณ์
Random Forest

Model	Accuracy	Precision	RMSE
Random Forest	98.71%	0.987	0.0369
Random Forest +IR	98.65%	0.987	0.0366
Random Forest+IG	98.65%	0.987	0.0365
Random Forest+CFS	97.99%	0.98	0.039
Random Forest+SU	98.74%	0.987	0.0366

1.4 การจัดลำดับความสำคัญของคุณลักษณะ

ในแต่ละเทคนิคของการคัดเลือกคุณลักษณะของข้อมูลจะได้ผลลัพธ์ของปัจจัยที่มีความสำคัญที่ไม่เท่ากัน ดังนั้นต้องนำปัจจัยผลลัพธ์ที่ได้มาจัดลำดับความสำคัญใหม่โดยดูว่าจากทั้ง 4 เทคนิคนั้นเหตุปัจจัยใดที่ถูกกำหนดให้มีความสำคัญมากที่สุด ผลและการวิเคราะห์ลำดับความสำคัญของปัจจัยที่เลือก โดยขึ้นอยู่กับวิธีการคัดเลือกคุณลักษณะที่แตกต่างกัน ดังตารางที่ 4

ตารางที่ 4 ลำดับความสำคัญของแต่ละปัจจัย

ปัจจัย	IR	GR	CFS	SU	Sum of priority	Final Order
ปัจจัยด้านช่องทางการจัดจำหน่าย	2	3	1	2	8	2
ปัจจัยการนำเสนอ และการปฏิบัติต่อผู้บริโภค	4	1	3	3	14	3
ปัจจัยด้านผลิตภัณฑ์และราคาของผลิตภัณฑ์	1	2	2	1	6	1
ปัจจัยด้านประชากรศาสตร์	3	4	4	4	15	4

จากตาราง 4 พบว่าปัจจัยผู้บริโภคที่มีต่อการซื้อผลิตภัณฑ์ผ้าไหมทอมือผ่านแอปพลิเคชัน 4 อันดับแรก ได้แก่ ปัจจัยด้านผลิตภัณฑ์และราคาของผลิตภัณฑ์ ปัจจัยด้านช่องทางการจัดจำหน่าย ปัจจัยการนำเสนอ และการปฏิบัติต่อผู้บริโภค ปัจจัยด้านประชากรศาสตร์ ตามลำดับ

1.5 การวัดค่าประสิทธิภาพ

การเปรียบเทียบโมเดลและวิธีเลือกคุณลักษณะข้อมูลพิจารณาจากค่าความถูกต้อง (Accuracy) มีค่าเท่ากับ 98.74% ความแม่นยำ (Precision) มีค่าเท่ากับ 0.987 และค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) มีค่าเท่ากับ 0.0366

ข้อเสนอแนะ

การใช้เทคนิคการพยากรณ์ข้อมูลสามารถใช้เทคนิคอื่น ๆ ที่น่าสนใจที่สามารถนำมาประเมินประสิทธิภาพและการวัดค่าความถูกต้องของการพยากรณ์ในการนำมาเปรียบเทียบเพื่อให้ได้ประสิทธิภาพสูงที่สุด

เอกสารอ้างอิง

- จินตนาภรณ์ วัฒนธร. (2559). “ดักแด้ใหม่” อาหารสุขภาพยุคใหม่. สืบค้นเมื่อ 2 มกราคม 2563 จาก <http://www.icam.kku.ac.th/Newletter/bombyx.pdf>.
- ประสิทธิ์ สุวรรณรักษ์. (2555). ระเบียบวิธีวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์. (พิมพ์ครั้งที่10). บุรีรัมย์ : คณะครุศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์.
- อัจฉิมา มณฑาทันธุ์. (2562). การเปรียบเทียบวิธีการคัดเลือกคุณลักษณะที่สำคัญในการปรับปรุงการพยากรณ์ มะเร็งเต้านม. รายงานการวิจัย มหาวิทยาลัยศรีปทุม.
- วันวิสาข์ ชนะประเสริฐ. (2560). การประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อแนะนำอาชีพสำหรับนักศึกษาปริญญาตรี คณะโบราณคดี มหาวิทยาลัยศิลปากร. วิทยานิพนธ์ ศิลปศาสตรมหาบัณฑิต มหาวิทยาลัยศิลปากร.
- สายชล สินสมบูรณ์ทอง. (2558). การทำเหมืองข้อมูล. กรุงเทพฯ : จามจุรีโปรดักท์.
- สุรวีชร ศรีเปารยะ, และ สายชล สินสมบูรณ์ทอง. (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง : กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย. วารสารวิทยาศาสตร์และเทคโนโลยี, 25(5), 839-853.
- Han, J., & Kamber, M. (2006). *Data mining : Concepts and techniques*. (2nd ed.).San Francisco: Morgan Kaufmann.