**Received:** 9 ก.ค. 2566 **Revised:** 20 ต.ค. 2566 **Accepted:** 31 ต.ค. 2566

# การพัฒนาแบบจำลองเพื่อการตัดสินใจรับซื้อโคขุน โดยใช้เทคนิคการทำเหมืองข้อมูล กรณีศึกษากลุ่มวิสาหกิจชุมชนโคเนื้อบ้านบูเกะจือฆา

# The Development of Decision-Making Models for Buying the Fattening Cattle Using Data Mining Techniques: A case study of the Buket-Chekha Beef Cattle Social Enterprise.

อับดุลเลาะ บากา[1*], สุลัยมาน เภอโส๊ะ[1], ฟูไดละห์ ดือมอง[1] และอรรถพล อดุลยศาสน์[1]
[1]คณะวิทยาศาสตร์เทคโนโลยีและการเกษตร มหาวิทยาลัยราชภัฏยะลา อำเภอเมือง จังหวัดยะลา

Abdulloh Baka[1*], Sulaiman Persoh[1], Fudailah Demong[1] and Attapol Adulyasas[1]
[1]Faculty of Science Technology and Agriculture, Yala Rajabhat University, Yala
*Corresponding author: abdulloh.b@yru.ac.th

## Abstract

"The Buket-Chekha Beef Cattle Social Enterprise" is a group that operates its own business of food transformation of frozen beef products. Frequently, some farmers face financial loss due to a lack of experience in selecting appropriate cattle as input in the transformation process. Therefore, researchers have used the RapidMiner Studio to develop the 5 data mining models, including Decision Tree (C4.5), Iterative Dichotomiser 3 (ID3), Naive Bayes (NB), Random Forest (RF), and K-Nearest Neighbors (k-NN) models, for purchasing decisions of fattened cattle. The case study was based on the 416 purchase records of the Buket-Chekha Beef Cattle Social Enterprise from 2018 to 2022. The data cleaning and feature selection were pre-processed before building the models. Then, the data were separated into 10 folds to cross-validate between the train and test sets. As the decision accuracy measures the performance of the models, the experimental results show that the RF decision tree model has given the highest accuracy at 98.53 % with 9 correlation rules, followed by the accuracy of C4.5, NB, ID3, and k-NN models at 97.02 %, 95.00 %, 94.75 %, and 93.54 %, respectively. Furthermore, the significant factors in the model were weight before slaughtering, body length, tooth structure, height, species, networking, fattening mean, fattening period, age, and purchase price.

***Keywords:*** *Data mining, Decision Tree, Feature Selection, Data Discretization*

**บทคัดย่อ**

กลุ่มวิสาหกิจชุมชนโคเนื้อบ้านบูเกะจือฆาเป็นกลุ่มที่ดำเนินธุรกิจแปรรูปผลิตภัณฑ์เนื้อโคขุนแช่แข็ง ซึ่งบ่อยครั้งที่เกษตรกรบางรายประสบปัญหาขาดทุนจากการเลือกซื้อโคขุนที่ผิดพลาด เนื่องจากขาดความเชี่ยวชาญในการเลือกซื้อโคขุนเพื่อเข้ามาแปรรูป ดังนั้นผู้วิจัยจึงได้พัฒนาแบบจำลองเพื่อการตัดสินใจรับซื้อโคขุน โดยใช้เทคนิคการทำเหมืองข้อมูล 5 วิธี ประกอบด้วย Decision Tree (C4.5), Iterative Dichotomiser 3 (ID3), Naive Bayes (NB), Random Forest (RF) และ K-Nearest Neighbors (k-NN) ทำการวิเคราะห์ด้วยโปรแกรม RapidMiner Studio กรณีศึกษากลุ่มวิสาหกิจชุมชนโคเนื้อบ้านบูเกะจือฆา โดยใช้ข้อมูลการรับซื้อโคขุนตั้งแต่ปี พ.ศ 2561 - 2565 จำนวน 416 แถวข้อมูล มีการกำจัดข้อมูลที่ผิดพลาด มีการคัดเลือกคุณลักษณะของข้อมูล มีการแบ่งช่วงข้อมูล และใช้ 10-Fold Cross Validation ในการแบ่งข้อมูลออกเป็นชุดข้อมูลสอนและชุดข้อมูลทดสอบ โดยใช้ค่าความถูกต้องในการวัดประสิทธิภาพการพยากรณ์ของแบบจำลอง ผลการทดลองพบว่า เทคนิคการทำเหมืองข้อมูลด้วยวิธีการของ RF มีค่าความถูกต้องสูงที่สุดคือ 98.53% และได้กฎความสัมพันธ์จำนวน 9 กฎ รองลงมาเป็น C4.5, NB, ID3 และ k-NN คือ 97.02%, 95.00%, 94.75% และ 93.54% ตามลำดับ โดยมีคุณลักษณะข้อมูลสำคัญคือ น้ำหนักก่อนเชือด ความยาวลำตัว ลักษณะฟัน ความสูง สายพันธ์ กลุ่มเครือข่าย ลักษณะขุน ระยะเวลา อายุ และราคารับซื้อ ในการเรียนรู้และสร้างแบบจำลองเพื่อการตัดสินใจรับซื้อโคขุน

***คำสำคัญ :*** *เหมืองข้อมูล, ต้นไม้ตัดสินใจ, การคัดเลือกคุณลักษณะข้อมูล, การแบ่งช่วงข้อมูล*

**บทนำ**

ภาคการเกษตรในประเทศไทยถือว่ามีบทบาทสำคัญในการขับเคลื่อนระบบเศรษฐกิจของประเทศเป็นอย่างมาก ทั้งในรูปแบบการเกษตรโดยตรงและการอุตสาหกรรมที่เกี่ยวของกับการเกษตร โดยเฉพาะการเลี้ยงโคเนื้อเปนอาชีพทางการเกษตรที่สำคัญอีกอาชีพหนึ่ง แต่เกษตรกรผู้เลี้ยงโคในประเทศไทยมีการนำข้อมูลและเทคโนโลยีมาใช้ในการพัฒนาภาคเกษตรอยู่ในระดับที่น้อยมาก เมื่อเทียบกับภาคอุตสาหกรรมอื่นๆ ดังนั้นการนำข้อมูลและเทคโนโลยีมาช่วยในการผลักดันธุรกิจการโคเนื้อในประเทศไทยถือว่าเป็นสิ่งที่จำเป็นอย่างยิ่ง โดยเฉพาะกลุ่มวิสาหกิจชุมชนด้านการผลิตปศุสัตว์ที่มีการจดทะเบียนกับกองส่งเสริมวิสาหกิจชุมชน กรมส่งเสริมการเกษตร (2566) รายงานว่ามีจำนวนมากถึง 23,686 กลุ่ม แต่น้อยมากที่กลุ่มวิสาหกิจชุมชนเหล่านั้นจะประสบความสำเร็จในการประกอบกิจการ ด้วยหลายปัจจัย เช่น องค์ความรู้ที่มี เงินลงทุน รวมถึงรูปแบบการจัดเก็บข้อมูล และการนำเทคโนโลยีมาใช้งาน กลุ่มวิสาหกิจชุมชนโคเนื้อบ้านบูเกะจือฆา เป็นอีกกลุ่มวิสาหกิจที่มีการทำธุรกิจปศุสัตว์ โดยเฉพาะการแปรรูปผลิตภัณฑ์โคขุนพรีเมี่ยมแช่แข็ง เช่น เนื้อสไลด์ เนื้อสเตก และเนื้อบด เป็นต้น ซึ่งมีการจัดเก็บข้อมูลที่เป็น

ระบบตั้งแต่ปี พ.ศ. 2561 จนถึงปัจจุบัน แต่ยังขาดในเรื่องของการนำข้อมูลไปประยุกต์ใช้ประโยชน์โดยเฉพาะองค์ความรู้ของเกษตรกรในการตัดสินใจรับซื้อโคขุนเข้าโรงฆ่าสัตว์เพื่อแปรรูปเป็นผลิตภัณฑ์ต่างๆ ของกลุ่มวิสาหกิจชุมชน ซึ่งจำเป็นต้องมีผู้เชี่ยวชาญและผู้ที่มีประสบการณ์เฉพาะด้านในการรับซื้อโคขุน เพื่อพิจารณาคุณลักษณะของโคขุนที่จะรับซื้อ เช่น ความสูง ความยาว สายพันธ์ ลักษณะการขุน ระยะเวลาการขุน น้ำหนัก ลักษณะฟัน เป็นต้น เพื่อลดความเสี่ยงในการขาดทุนของเกษตรกรในการรับซื้อโคขุน ซึ่งบ่อยครั้งที่เกษตรกรบางรายประสบปัญหาขาดทุนในการเลือกซื้อโคขุน เนื่องจากขาดความเชี่ยวชาญในการเลือกซื้อโคขุนเพื่อเข้ามาแปรรูป ส่งผลให้กลุ่มวิสาหกิจชุมชนมีผลประกอบการที่ไม่เป็นไปตามที่คาดหวังไว้ และบางครั้งมีผลประกอบการอยู่ในภาวะขาดทุน

เหมืองข้อมูล (Data Mining) คือกระบวนการของการกลั่นกรองสารสนเทศที่ซ่อนอยู่ในคลังข้อมูลหรือฐานข้อมูลขนาดใหญ่ เพื่อทำนายแนวโน้มและพฤติกรรมของข้อมูล โดยอาศัยข้อมูลในอดีต เพื่อค้นหากฎความสัมพันธ์ (Association Rule) จดจำรูปแบบ (Pattern Recognition) และจำแนกประเภทข้อมูล (Classification) เพื่อใช้สนับสนุนการตัดสินใจ คุณลักษณะโดยทั่วไปของเหมืองข้อมูลสามารถแบ่งออกเป็น 3 คุณลักษณะใหญ่ๆ (Liu, et al., 1998) คือ 1) การเรียนรู้แบบมีผู้สอน (Supervised Learning) คอมพิวเตอร์จะทำการเรียนรู้เพื่อแบ่งกลุ่มข้อมูลจากชุดข้อมูลตัวอย่างที่ใช้ในการสอนที่มีการระบุคลาสแล้ว ตัวอย่างเช่น ต้นไม้ตัดสินใจ (Decision Trees) โครงข่ายประสาทเทียม (Artificial Neural Network) และนาอีฟเบย์ (Naïve Bayes) เป็นต้น 2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) คอมพิวเตอร์จะทำการเรียนรู้เพื่อแบ่งกลุ่มข้อมูลที่ยังไม่มีการระบุคลาส ซึ่งอาศัยความสัมพันธ์หรือความคล้ายของข้อมูลหรือระยะห่างของข้อมูลในการแบ่งกลุ่ม ตัวอย่างเช่น แผนที่การจัดกลุ่มเอง (Self-Organizing: SOM) การแบ่งกลุ่มด้วยวิธีเคมีน (K-Means) เป็นต้น และ 3) การวิเคราะห์ตะกร้าทางการตลาด (Market Basket Analysis) เพื่อค้นหากฎความสัมพันธ์ (Association Rules) ตัวอย่างเช่น ขั้นตอนวิธีอพริโอริ (Apriori Algorithm) (Fudailah, et al., 2009)

ดังนั้นผู้วิจัยจึงเล็งเห็นถึงความสำคัญในการนำข้อมูลที่มีไปประยุกต์ใช้งาน โดยการพัฒนาแบบจำลองเพื่อการตัดสินใจรับซื้อโคขุน โดยใช้เทคนิคการทำเหมืองข้อมูล กรณีศึกษา กลุ่มวิสาหกิจชุมชนโคเนื้อบ้านบูเกะจือฆา เพื่อเป็นแบบจำลองที่จะช่วยในการตัดสินใจรับซื้อโคขุนให้กับกลุ่มวิสาหกิจชุมชนที่ทำธุรกิจเกี่ยวกับการเลี้ยงโคเนื้อหรือเกษตรกรที่สนใจได้อย่างมีประสิทธิภาพถูกต้องและแม่นยำ

**วัตถุประสงค์การวิจัย**

1. เพื่อพัฒนาและเปรียบเทียบประสิทธิภาพของแบบจำลองในการตัดสินใจรับซื้อโคขุนของกลุ่มวิสาหกิจชุมชนโคเนื้อบ้านบูเกะจือฆา โดยใช้เทคนิคการทำเหมืองข้อมูลสำหรับการเรียนรู้แบบมีผู้สอน

**ทฤษฎีและงานวิจัยที่เกี่ยวข้อง**

ต้นไม้ตัดสินใจ (Decision Tree) เป็นแบบจำลองในการจำแนกข้อมูลที่ใช้โครงสร้างของต้นไม้ ประกอบด้วยโหนด (nodes) แทนการตัดสินใจในการจำแนกข้อมูลตามคุณลักษณะข้อมูล กิ่ง (branches) แทนเงื่อนไขที่จะเชื่อมโหนดกับโหนดย่อย และใบ (leaf) แทนคำตอบหรือคลาสที่ถูกทำนาย โดยประเมินจากสิ่งที่เรียกว่า ค่าไม่บริสุทธิ์ (Impurity measure) เช่น Entropy และ Gini เป็นต้น ซึ่ง Entropy จะมีประสิทธิภาพในการทำงานได้ดีกับสถานการณ์ที่คลาสมีจำนวนไม่เท่ากันและมีการแจกแจงที่แตกต่างกันมาก จะมีค่าอยู่ระหว่าง 0-1 ซึ่งค่า Entropy ยิ่งน้อย แสดงถึงการจำแนกข้อมูลที่ดีและชัดเจนมากขึ้น ส่วน Gini จะมีประสิทธิภาพในการทำงานได้ดีในสถานการณ์ที่มีคลาสมากกว่าสองและคลาสมีจำนวนที่ใกล้เคียงกัน จะมีค่าอยู่ระหว่าง 0-0.5 โดยส่วนใหญ่ Entropy จะให้ผลลัพธ์ของต้นไม้ตัดสินใจที่มีจำนวนโหนดมากกว่า Gini (ภรัณยา ปาลวิสุทธิ์, 2559)

**Iterative Dichotomiser3 (ID3)** เป็นอัลกอริทึมในการสร้างต้นไม้ตัดสินใจโดยใช้หลักการหา Information Gain เป็นตัวชี้วัดในการสร้างต้นไม้ตัดสินใจ เพื่อเลือกคุณลักษณะข้อมูลที่เหมาะสมที่สุดในการจำแนกข้อมูล โดยคำนวณค่าความสัมพันธ์ของข้อมูลในแต่ละคุณลักษณะข้อมูล เพื่อหาคุณลักษณะข้อมูลที่มีความสัมพันธ์กับคลาสข้อมูลมากที่สุด โดย Information Gain คำนวณได้ดังสมการที่ (1) และ Entropy คำนวณได้ดังสมการที่ (2) (R. Quinlan., 1986)

$$IG = Entropy(t_i) - \sum_{i=1}^{n} [p(t_i)] * p(t_i) \quad (1)$$

$$Entropy(t_i) = 1 - \sum_{i=1}^{n} [p(t_i)] * log_2 p(t_i) \quad (2)$$

โดยที่ $t_i$ คือ คุณลักษณะข้อมูลที่นำมาวัดค่า Entorpy

$p(t_i)$ คือ สัดส่วนจำนวนสมาชิก $i$ กับจำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง

**Decision Tree (C4.5)** เป็นอัลกอริทึมที่พัฒนาต่อจาก ID3 ที่มีความสามารถในการจัดการกับข้อมูลที่เป็นตัวเลข (continuous data) และมีความสามารถมากขึ้นในการจัดการข้อมูลที่ไม่สมบูรณ์และความแปรปรวนของคุณลักษณะข้อมูล ซึ่ง RapidMiner ได้ใช้ C4.5 ในโมดูลของ Decision Tree (DT) โดยจะทำการคัดเลือกคุณลักษณะที่มีความสัมพันธ์กับคลาสมากที่สุดเป็นโหนดราก (Root node) หลังจากนั้นก็จะหาคุณลักษณะข้อมูลที่มีความสัมพันธ์กับคลาสถัดไปเรื่อย ๆ ในการหาความสัมพันธ์ของ

คุณลักษณะข้อมูลนี้จะใช้ตัวชี้วัดในการจำแนกข้อมูลด้วยสมการ Gini แสดงดังสมการที่ (3) (C. Kaewchinporn., 2010)

$$Gini(t_i) = 1 - \sum_{i=1}^{n} p(t_i)^2 \quad (3)$$

โดยที่ $t_i$ คือ คุณลักษณะข้อมูลที่นำมาวัดค่า Gini
$p(t_i)$ คือ สัดส่วนจำนวนสมาชิก $i$ กับจำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง

**Random Forest (RF)** เป็นการสร้างแบบจำลองต้นไม้ตัดสินใจขึ้นมาหลายๆ แบบจำลอง โดยวิธีการสุ่มตัวแปร แล้วนำผลที่ได้แต่ละแบบจำลองมารวมกัน แล้วนับจำนวนผลที่มีจำนวนซ้ำกันมากที่สุดจะถูกโหวตออกมาเป็นผลลัพธ์สุดท้าย โดยมีขั้นตอนดังนี้ (L. Breiman and R. Forests, 2001)

1) ต้นไม้แต่ละต้นจะถูกสอนโดยใช้เซตย่อยจากข้อมูลตัวอย่าง

2) เมื่อต้นไม้โตขึ้น จะค้นหาโหนดที่อยู่ในกิ่งที่ดีที่สุดของต้นไม้ โดยการสุ่มเลือกคุณสมบัติจาก N คุณสมบัติ

3) ต้นไม้แต่ละต้นจะไม่มีการตัดออก แต่จะปล่อยให้ต้นไม้โตขึ้นไปเรื่อย ๆ จนได้ผลลัพธ์ที่ดีที่สุด หลักจากการสร้างป่าแล้ว ทำการให้คะแนน (Vote) โดยต้นไม้ภายในป่า หากต้นไม้ต้นใดได้คะแนนสูงสุดก็จะนำเอาต้นไม้ต้นนั้นออกมาสร้างเป็นแบบจำลอง

**K-Nearest Neighbors (k-NN)** เป็นวิธีการแบ่งคลาสสำหรับใช้จัดหมวดหมู่ข้อมูล (Classification) โดยมีหลักการดังต่อไปนี้ (เกรียงศักดิ์ รักภักดี และคณะ, 2560)

1) กำหนดค่าของ k

2) คำนวณระยะห่างระหว่างชุดข้อมูลสอน (Training set) กับชุดข้อมูลทดสอบ (Test set) โดยใช้ทฤษฎีระยะห่างแบบยคูลิเดียน (Euclidean Distance) แสดงดังสมการที่ (4)

3) กำหนดคลาสให้กับข้อมูลใหม่ โดยอาศัยระยะห่างที่สั้นที่สุดจำนวน k ข้อมูล

$$d(x,y) = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2} \quad (4)$$

โดยที่ $x$ คือ ข้อมูลที่ต้องการจำแนก
$y$ คือ ข้อมูลในแต่ละกลุ่ม
$N$ คือ จำนวนคุณลักษณะที่มีในข้อมูลทั้ง 2 ชุด

**Naïve Bayes (NB)** เป็นเทคนิคการจำแนกข้อมูลที่มีการตั้งสมมติฐานเพื่อกำหนดให้การเกิดของเหตุการณ์ต่างๆ ที่ใช้ในการจัดกลุ่มนั้นเป็นอิสระต่อกัน ซึ่งจะทำการวิเคราะห์ความสัมพันธ์ระหว่างคุณลักษณะข้อมูลและคลาสโดยใช้สมการเบย์ (Bayes' Theorem) สามารถคำนวณได้ดังสมการที่ (5) เพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นของแต่ละความสัมพันธ์ จุดประสงค์เพื่อต้องการสร้างแบบจำลองที่อยู่ในรูปของความน่าจะเป็น เพื่อหาว่าสมมติฐานใดถูกต้องที่สุด ข้อดีของวิธีการแบบเบย์คือสามารถใช้ข้อมูลและความรู้ก่อนหน้านี้ เพื่อใช้ในการเรียนรู้ นอกจากนี้ยังเหมาะสมกับชุดข้อมูลที่มีขนาดใหญ่และคุณลักษณะข้อมูลที่เป็นอิสระต่อกัน (Richard, J. R. and Geatz, M. W., 2003)

$$\mathrm{P(H|E)} = [\mathrm{P(E|H)} \times \mathrm{P(H)}]/\mathrm{P(E)} \quad (5)$$

โดยที่ $\mathrm{P(H)}$ คือ ความน่าจะเป็นที่จะเกิดสมมติฐาน H
$\mathrm{P(E)}$ คือ ความน่าจะเป็นของชุดข้อมูล E
$\mathrm{P(H|E)}$ คือ ความน่าจะเป็นของ H เมื่อทราบ E
$\mathrm{P(E|H)}$ คือ ความน่าจะเป็นของ E เมื่อทราบ H

**การคัดเลือกคุณลักษณะ (Feature Selection)** เป็นเทคนิคที่ช่วยลดมิติของข้อมูลและช่วยให้การเรียนรู้วิธีการพยากรณ์ดำเนินการได้เร็วขึ้นและมีประสิทธิภาพมากขึ้น (วนัสดา สิริวุฒิชาภิรัชต์ และคณะ, 2564) กระบวนการของการคัดเลือกคุณลักษณะแบ่งได้เป็น 2 กลุม ได้แก่ Filter approach และ Wrapper approach (ปะพาดา ณ วิเชียร และคณะ, 2563) สำหรับงานวิจัยนี้เลือกใช Filter approach ด้วยหลักการคัดเลือกคุณลักษณะ โดยใช้การคำนวณหาคาน้ำหนักหรือค่าความสัมพันธ์ระหว่างแต่ละคุณลักษณะข้อมูลและคลาสต่าง ๆ การคัดเลือกคุณลักษณะข้อมูลด้วยการเรียงลำดับตามคาน้ำหนักที่คำนวณ แล้วทำการเลือกคุณลักษณะข้อมูลที่มีคาน้ำหนักมากกวาที่ต้องการมาใชงาน

**งานวิจัยที่เกี่ยวข้อง**

วนัสดา สิริวุฒิชาภิรัชต์ และคณะ (2564) นำเสนอระบบสนับสนุนการตัดสินใจสำหรับการเลี้ยงโคเนื้อ เพื่อเป็นแนวทางในการเลี้ยงโคสายพันธุ์ชาโรเลส์และบาร์ห์มัน ให้ผู้ที่สนใจริเริ่มเลี้ยงได้ศึกษาถึงวิธีกระบวนการเลี้ยง และสามารถนำองค์ความรู้ไปใช้ก่อนลงมือปฏิบัติได้จริง งานวิจัยโดยมีวัตถุประสงค์คือ 1) เพื่อพัฒนาระบบสนับสนุนการตัดสินใจสำหรับการเลี้ยงโคเนื้อ 2) เพื่อประเมินความพึงพอใจของผู้ใช้ระบบสนับสนุนการตัดสินใจสำหรับการเลี้ยงโคเนื้อผ่านเว็บแอปพลิเคชัน โดยใช้ภาษา PHP และออกแบบเป็นเรซสปอนต์ซีพเว็บเพื่อรองรับสมาร์ทโฟน ใช้ฐานข้อมูล MySQL ในการจัดการฐานข้อมูล โดยระบบจะสอบถามข้อมูลดังนี้ ข้อมูลสายพันธ์ที่ต้องการเลี้ยง ข้อมูลระยะเวลาการเลี้ยง ข้อมูลอายุ และข้อมูลต้นทุนการผลิต แล้วระบบจะรายงานข้อมูลค่าใช้จ่ายต่างๆ สิ่งที่ผู้เลี้ยงต้องเตรียมตัว การเลือกอาหารสัตว์หรือวัตถุดิบแต่ละระยะเวลาในการเลี้ยงของโค ผลการประเมินประสิทธิภาพการทำงานของระบบโดยผู้เชี่ยวชาญจำนวน 5 คน ผลประเมินอยู่ในระดับปานกลาง และความพึงพอใจของผู้ใช้งานระบบจำนวน 30 คน ผลการประเมินมีความพึงพอใจอยู่ในระดับมาก

Cáceres, E. N, et al. (2011) นำเสนอระบบปศุสัตว์แม่นยำ เพื่อเพิ่มประสิทธิภาพในกระบวนควบคุมการผลิตของอุตสาหกรรมโคเนื้อ ช่วยในเรื่องของการตัดสินใจ การคาดการณ์ และจัดการความเสี่ยงให้กับผู้ผลิตในอุตสาหกรรมโคเนื้อในด้านต่างๆ อย่างมีประสิทธิภาพ วัตถุประสงค์หลักของงานวิจัยนี้คือ ศึกษาปัญหาของการจัดการฝูงสัตว์โดยใช้อุปกรณ์อิเล็กทรอนิกส์ การจำลองการเจริญเติบโตของทุ่งหญ้า การพยากรณ์ราคาสินค้า การวินิจฉัยโรคสัตว์ลวงหน้า การศึกษาสายพันธุกรรมของสัตว์ การจำแนกซากอัตโนมัติ เช่น เนื้อ หนังวัว และการเลือกเส้นทางการขนส่งเป็นต้น การพัฒนาความเชี่ยวชาญของระบบติดตามและการจัดการฝูงสัตว์ที่ชาญฉลาด จะส่งผลให้ได้โคเนื้อที่มีคุณภาพ และเพิ่มขีดความสามารถในการแข่งขันของธุรกิจเกษตร ปศุสัตว์แม่นยำ สามารถกำหนดเป็นแนวปฏิบัติในการจัดการฝูงวัวซึ่งใช้เทคโนโลยีสารสนเทศและการสื่อสารเพื่อให้แน่ใจว่ามีแนวทางปฏิบัติที่ดีในการผลิตเนื้อสัตว์จากข้อมูลเฉพาะของฝูงสัตว์ พื้นที่เลี้ยงสัตว์ที่ใช้อ้างอิงทางภูมิศาสตร์ ระบบอัตโนมัติในทุกขั้นตอนของอุตสาหกรรมเนื้อสัตว์ การปศุสัตว์ที่แม่นยำมีจุดมุ่งหมายเพื่อปรับต้นทุนการผลิตให้เหมาะสม เพื่อให้ได้เนื้อที่มีคุณภาพ นอกจากนี้ระบบปศุสัตว์แม่นยำยังศึกษาผลกระทบด้านสิ่งแวดล้อมและสังคมในการผลิตเนื้อสัตว์และผลผลิตที่เพิ่มขึ้น

Gonzalez-Araya and Marcela C. (2016) ได้นำเสนอระบบสนับสนุนการตัดสินใจสำหรับการวางแผนการผลิตในโรงฆ่าสุกรในประเทศชิลี เพื่อสนับสนุนการตัดสินใจในการปฏิบัติงานในโรงฆ่าสุกร โดยมีเป้าหมายเพื่อลดต้นทุนการผลิตการเก็บรักษาและลดการสูญเสียของเนื้อ สำหรับการพัฒนาแบบจำลองนี้ประกอบด้วย 4 รูปแบบ คือ 1) แบบจำลองการวางแผนของการตัดแต่งและการบรรจุรายวัน ช่วยให้ผู้มีอำนาจตัดสินใจสามารถกำหนดจำนวนซากหมูที่จะใช้ในแต่ละวันได้ ตามน้ำหนักและชนิดของ

การตัดแต่ง 2) แบบจำลองการวางแผนการตัดแต่งและการบรรจุรายสัปดาห์ ช่วยให้ผู้ใช้สามารถประเมินวันที่ในการจัดส่งสินค้าของลูกค้าแต่ละราย สถานะของการขนส่ง และจำนวนซากที่จะใช้ตามผลิตภัณฑ์ในแต่ละรอบการผลิต 3) แบบจำลองการจัดการระบบทำความเย็น เนื่องจากผลลัพธ์ที่ได้รับจากแบบจำลองการวางแผนรายวันและรายสัปดาห์ มีผลต่อการจัดการระบบทำความเย็นที่จะต้องกำหนดเวลาการนำสุกรออกจากห้องเย็นไปยังห้องเชือดเพื่อเพิ่มประสิทธิภาพในการเชือดและบรรจุ และ 4) แบบจำลองการบริหารจัดการของระบบจัดเก็บและจัดระเบียบผลิตภัณฑ์ในห้องเย็นตามประเภทของสินค้า เช่น ตระกูลแฮม เนื้อสันนอก ซี่โครง พุง สันไหล่ และหมูแผ่น โดยแบบจำลองจะกำหนดเส้นทางเดินที่จะใช้ตามคลาส ผลลัพธ์ที่ได้ทำให้มีอัตรากำไรที่สูงขึ้น 0.5 เซ็นต์ ต่อกิโลกรัมที่สัตว์มีชีวิต ลดการสูญเสียลูกค้าต่างประเทศ การลดลงของสินค้าคงคลังประมาณร้อยละ 0.10 ระบบสนับสนุนการตัดสินใจช่วยในการปรับปรุงระบบบริหารจัดการทั้งหมดนี้มีผลกำไรเพิ่มขึ้นต่อปีประมาณ 3.5 ล้านดอลลาร์

วัชรวิทย์ มีหนองใหญ่ (2557) ได้นำเสนอปัจจัยที่มีผลต่อความนุ่มของเนื้อโค พบว่า โคสายพันธ์เขตร้อน เช่น โคบรามัน หรือโคพื้นเมืองในประเทศไทย เนื้อจะมีความนุ่มน้อยกว่า โคสายพันธ์เขตหนาว เช่น ชาโลเล่ หรือแองกัส เป็นต้น ส่วนปัจจัยด้านระยะเวลาการขุนต้องมีระยะเวลาอย่างน้อย 80 วันขึ้นไป ถึงจะทำให้โคมีน้ำหนักตัวที่อ้วนสมบูรณ์ โดยเฉพาะคุณลักษณะไขมันแทรกสายพันธ์เขตหนาวจะมีคุณภาพของไขมันแทรกที่ดีกว่าเช่นกัน ถึงแม้จะใช้อาหารเลี้ยงขุนในลักษณะเดียวกันก็ตาม

**วิธีการดำเนินการวิจัย**

ขั้นตอนในการการพัฒนาแบบจำลองสำหรับการตัดสินใจรับซื้อโคขุน โดยใช้เทคนิคการทำเหมืองข้อมูล กรณีศึกษา กลุ่มวิสาหกิจชุมชนโคเนื้อบ้านบูเกะจือฆามีทั้งหมด 5 ขั้นตอนดังนี้

**ขั้นตอนที่ 1** การรวบรวมข้อมูลที่ใช้ในการวิจัยเป็นการรวบรวมข้อมูลจากฐานข้อมูลการรับซื้อโคเนื้อจากสมาชิก ตั้งแต่เดือนมกราคม พ.ศ 2561 ถึงเดือนธันวาคม พ.ศ. 2565 จำนวน 416 แถวข้อมูล (Data Example) ประกอบด้วย 3 ฐานข้อมูลดังนี้

1.1) ฐานข้อมูลการรับซื้อโค เป็นข้อมูลพื้นฐานของโคที่รับซื้อจากสมาชิกเพื่อเข้าโรงฆ่าสัตว์ เช่น สายพันธ์ อายุ น้ำหนัก ส่วนสูง ระยะเวลาการขุน และราคารับซื้อ เป็นต้น มีจำนวนทั้งหมด 21 คุณลักษณะ (Attribute) แสดงตัวอย่างดังภาพที่ 1

รหัส:1001,วันเดือนปี:12 มกราคม 2018,สายพันธ์:ชาโลเล่,อายุ (เดือน):32,สี:ขาว,เพศ:ผู้, น้ำหนักก่อนเชือด(กก):568,ความสูง(ซม.):158,ความยาวลำตัว(ซม.):220,เครือข่าย/กลุ่ม:ท่าธงพัฒนา, เบอร์โทร:ไม่มี,ตำบล:ท่าธง,อำเภอ: รามัน,จังหวัด:ยะลา,ลักษณะขุน:ยืนโรง,อาหารหยาบ:หญ้าสด, อาหารข้น:อาหารผง,ทำหมัน:ไม่ทำ, ลักษณะฟัน:ผลัดฟัน 2 ซี่,ระยะเวลาขุน:3-4 เดือน,ราคาซื้อต่อ(กก):90, ราคารวม:51120;

**ภาพที่ 1** ตัวอย่างข้อมูลบันทึกข้อมูลโคก่อนเข้าเชือด

1.2) ฐานข้อมูลการขาย เป็นการเก็บข้อมูลราคาขาย โดยคำนวณจากน้ำหนักการตัดแต่งของซากโคในแต่ละส่วนคูณด้วยราคาขายต่อกิโลกรัม เช่น สันคอ สันไหล่ ใบพาย และเสือร้องไห้เป็นต้น มีจำนวนทั้งหมด 38 คุณลักษณะ แสดงตัวอย่างดังภาพที่ 2

รหัส:1001,หัว:1165,สันคอ:5600,สันไหล่:2240,ใบพาย:2400,ปลาช่อน:2100,เสื้อร้องไห้:4500,สันใน:2870, สันนอก:5973,ริบอาย:4375,ร่องซี่โครง:1660,รวมเศษเนื้อ:1520,น่องลาย:4060,สะโพก:5320,พับใน:1652, พิคานยา:1932,พับนอก:2380,มะพร้าว:2296,พื้นท้อง:2400,เนื้อสามชั้น:1600,กระดูกซี่โครง:1400, กระดูกสันคอ:2030,กระดูกสันหลัง:2310,กระดูกก้น:260,กระดูกขา:3540,ช่อมาลี:486,หาง:375, รวมเครื่องใน:1550, รวมไส้:2350,หนังวัว:200,เศษไขมัน:117.5,รวมน้ำหนักทั้งหมด:457,น้ำหนักของเสีย:111, อัตราซากสดเบื้องต้น:300.2,%อัตตาซากสดเบื้องต้น:52.85,อัตราซากสดไม่มีกระดูก:202.2, % อัตราซากสดไม่มีกระดูก:35.60, ราคาขาย:67521.50

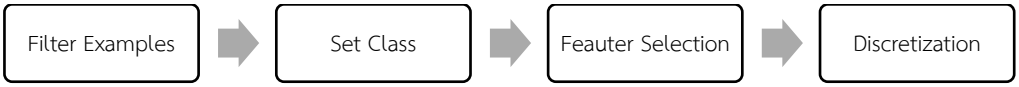
**ภาพที่ 2** ตัวอย่างข้อมูลบันทึกข้อมูลซากโคหลังเข้าเชือด

1.3) ฐานข้อมูลค่าใช้จ่าย เป็นการเก็บข้อมูลค่าแรงงาน ค่าธรรมเนียม และค่าขนส่งเป็นต้น มีจำนวนทั้งหมด 8 คุณลักษณะ แสดงตัวอย่างดังภาพที่ 3

รหัส:1001,ค่าเชือดโค:1704,ค่าธรรมเนียมโรงฆ่า:250,ค่ารถขนส่ง:500,ค่าแรงตัดแต่ง:1870,รวมค่าแรงงาน:4324, ค่าบริหารจัดการ:5112,รวมค่าใช้จ่าย:9436,กำไรเบื้องต้น:6965.5

**ภาพที่ 3** ตัวอย่างข้อมูลค่าจ้างแรงงานการผลิตและการแปรรูป

**ขั้นตอนที่ 2** การเตรียมข้อมูลประกอบด้วยขั้นตอนย่อย 4 ขั้นตอน แสดงดังภาพที่ 4 โดยมีรายละเอียดดังนี้




**ภาพที่ 4** ขั้นตอนและกระบวนการในการเตรียมข้อมูล

2.1) การคัดกรองข้อมูล (Filter Examples) เป็นการคัดเลือกแถวข้อมูลโดยการกำจัดข้อมูลที่ผิดพลาด (Missing Value) เป็นการกำจัดแถวข้อมูลที่ไม่สมบูรณ์ออกจากฐานข้อมูล เช่น บางแถวข้อมูลขาดข้อมูล อายุ และลักษณะขุน เป็นต้น ที่อาจเกิดจาการเก็บข้อมูลที่ไม่ครบถ้วน ขั้นตอนนี้ได้ทำการลบแถวข้อมูลที่ไม่สมบูรณ์ออกจำนวน 22 แถวข้อมูล จากทั้งหมด 416 แถวข้อมูล ทำให้คงเหลือจำนวน 394 แถว แสดงดังตารางที่ 1

**ตารางที่ 1** ตัวอย่างคุณลักษณะข้อมูลเบื้องต้นในการสร้างแบบจำลองเพื่อตัดสินใจรับซื้อโคขุน

| รหัส | สายพันธ์ | อายุ | น้ำหนัก | สูง | ยาว | ลักษณะขุน | อาหาร | ... | กำไรขาดทุน |
|---|---|---|---|---|---|---|---|---|---|
| 001 | พื้นเมือง | 20 | 320 | 160 | 182 | ยืนโรง | หญ้าสด | ... | -113 |
| ... | บรามัน | 26 | 585 | 165 | 210 | ธรรมชาติ | หญ้าสด | ... | 2280 |
| ... | ชาโลเล่ | 32 | 568 | 158 | 220 | ยืนโรง | TMR | ... | 4965.5 |
| ... | แองกัส | - | 650 | 162 | 253 | - | ฟาง | ... | 6306 |
| 416 | ชาโลเล่ | 37 | 625 | 167 | 250 | กั้นคอก | ฟาง | ... | 10251.5 |

2.2) การกำหนดคลาส (Class) จะกำหนดคลาสจากผลกำไรขาดทุนในตารางที่ 1 ของการรับซื้อโคขุนแต่ละตัว โดยมีรายละเอียดดังตารางที่ 2

**ตารางที่ 2** ตัวอย่างเงื่อนไขที่ใช้ในการคำนวณเพื่อกำหนดคลาส

| รหัส | กำไรขาดทุน | เงื่อนไขกำหนดคลาส | class |
|---|---|---|---|
| 001 | -113 | < 0 | ขาดทุน |
| ... | 2280 | 0 - 2500 | น้อย |
| ... | 4965.5 | 2501 - 5000 | ปานกลาง |
| ... | 6306 | 5001 - 7500 | มาก |
| 416 | 10251.5 | > 7501 | มากที่สุด |

2.3) การคัดเลือกคุณลักษณะข้อมูล (Feature Selection) เป็นการประเมินประสิทธิภาพของคุณลักษณะข้อมูลด้วยวิธีการ Weight by Information Gain ซึ่งจะทำการคำนวณหาค่าน้ำหนักหรือค่าความสัมพันธ์ระหว่างคุณลักษณะข้อมูลและคลาสต่าง ๆ โดยใช้วิธีการของ Information Gain ในการคำนวณหาค่าน้ำหนัก และกำหนดค่าน้ำหนักของคุณลักษณะขั้นต่ำ (select by weights) ในการสร้างแบบจำลองอยู่ที่ 0.5 ขึ้นไป ทำให้จำนวนของคุณลักษณะข้อมูลทั้งหมดจาก 67 เหลือเพียงแค่ 10 คุณลักษณะข้อมูลที่ผ่านเกณฑ์คัดเลือกจะอยู่ในกรอบเส้นประ แสดงดังตารางที่ 3

**ตารางที่ 3** ค่าน้ำหนักของแต่ละคุณลักษณะโดยใช้วิธีการของ Information Gain

| Attribute (คุณลักษณะข้อมูล) | Weight (น้ำหนัก) |
|---|---|
| น้ำหนักก่อนเชือด (กก.) | 0.998 |
| ความยาวลำตัด (ซม.) | 0.998 |
| ลักษณะฟัน | 0.987 |
| ความสูง (ซม.) | 0.877 |
| สายพันธ์ | 0.772 |
| เครือข่าย/กลุ่มผู้เลี้ยง | 0.754 |
| ลักษณะการขุนเลี้ยง | 0.702 |
| ระยะเวลาการขุน | 0.631 |
| อายุโค (เดือน) | 0.615 |
| ราคารับซื้อต่อ กก. (บาท) | 0.550 |
| อาหารหยาบ | 0.379 |
| อาหารข้น | 0.379 |
| ทำหมันวัว | 0.349 |

2.4) การแบ่งช่วงข้อมูล (Data Discretization) เป็นการแปลงค่าข้อมูลจากข้อมูลเชิงปริมาณ (Quantitative data) ให้เป็นข้อมูลเชิงคุณภาพ (Qualitative data) เพื่อลดความซับซ้อนของข้อมูลและเป็นไปตามเงื่อนไขของการสร้างแบบจำลองในแต่ละวิธี เช่น ต้นไม้ตัดสินใจ สำหรับขั้นตอนนี้ได้ทำการแบ่งช่วงข้อมูลของคุณลักษณะข้อมูลอายุ น้ำหนัก ความสูง และความยาวของลำตัว โดยใช้วิธีการแบ่งช่วงด้วยความถี่ที่เท่ากัน (Equal-Frequency) (Wong and Chiu, 1987) คือแต่ละช่วงข้อมูลจะมีจำนวนข้อมูลที่เท่ากัน (ใช้ความถี่เป็นเกณฑ์) ซึ่งจะช่วยแก้ปัญหาเรื่องการกระจุกตัวของข้อมูลได้ งานวิจัยนี้ได้กำหนดจำนวนของช่วงข้อมูล (number of bins) เท่ากับ 2 แสดงตัวอย่างดังตารางที่ 4

**ตารางที่ 4** ตัวอย่างข้อมูลที่ผ่านการทำ Data Discretization

| รหัส | อายุ | น้ำหนัก | ส่วนสูง | ความยาว |
|---|---|---|---|---|
| 001 | [ < 32.5] | [ < 637.5] | [ < 163.5] | [ < 245.0] |
| ... | [ >= 32.5] | [ >= 637.5] | [ < 163.5] | [ >= 245.0] |
| ... | [ < 32.5] | [ < 637.5] | [ >= 163.5] | [ < 245.0] |
| ... | [ >= 32.5] | [ < 637.5] | [ >= 163.5] | [ >= 245.0] |
| 416 | [ < 32.5] | [ >=637.5] | [ < 163.5] | [ < 245.0] |

**ขั้นตอนที่ 3** การแบ่งชุดข้อมูลทดสอบในการสร้างแบบจำลอง ใช้วิธีการแบ่งชุดข้อมูลออกเป็น 10 ชุด โดยใช้ข้อมูล 9 ชุดในการสอนและข้อมูล 1 ชุดในการทดสอบ ซึ่งชุดข้อมูลแต่ละชุดจะถูกสอนและทดสอบสลับกันไป เรียกวิธีการนี้ว่า 10 Fold Cross Validation

**ขั้นตอนที่ 4** การพัฒนาแบบจำลองสำหรับการตัดสินใจรับซื้อโคขุน โดยใช้เทคนิคการทำเหมืองข้อมูล กรณีศึกษา กลุ่มวิสาหกิจชุมชนโคเนื้อบ้านบูเกะจือฆา จะใช้ 5 เทคนิคของการทำเหมืองข้อมูล คือ C4.5, ID3, NB, RF และ k-NN โดยใช้โปรแกรม RapidMiner Studio โดยแบ่งชุดทดสอบประสิทธิภาพของเทคนิคการทำเหมืองข้อมูล ออกเป็นเป็น 4 รูปแบบคือ ชุดทดสอบ A เป็นชุดข้อมูลดิบ ที่ไม่ผ่านกระบวนการใดๆ ในขั้นตอนการเตรียมข้อมูล ชุดทดสอบ B เป็นชุดข้อมูลที่ผ่านกระบวนการกำจัดข้อมูลที่ผิดพลาด (Missing Value) ชุดทดสอบ C เป็นชุดข้อมูลที่ผ่านการกำจัดข้อมูลที่ผิดพลาด และผ่านการคัดเลือกคุณลักษณะของข้อมูล (Feature Selection) ชุดทดสอบ D เป็นชุดข้อมูลที่ผ่านการกำจัดข้อมูลที่ผิดพลาด มีการคัดเลือกคุณลักษณะของข้อมูล และมีการแบ่งช่วงข้อมูล (Data Discretization) แสดงรายละเอียดดังตารางที่ 5

**ตารางที่ 5** รูปแบบการทดสอบประสิทธิภาพของที่ใช้ในการทำเหมืองข้อมูล

| data set | num of attribute | num of rows | missing value | feature selection | data discretization | cross validation | classification method | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | C4.5 | ID3 | NB | RF | k-NN |
| A | 68 | 416 | X | X | X | 10-fold | ✓ | ✓ | ✓ | ✓ | ✓ |
| B | 68 | 394 | ✓ | X | X | 10-fold | ✓ | ✓ | ✓ | ✓ | ✓ |
| C | 11 | 394 | ✓ | ✓ | X | 10-fold | ✓ | ✓ | ✓ | ✓ | ✓ |
| D | 11 | 394 | ✓ | ✓ | ✓ | 10-fold | ✓ | ✓ | ✓ | ✓ | ✓ |

**ขั้นตอนที่ 5** การเปรียบเทียบผลการทดลอง โดยใช้วิธีการประเมินค่าความถูกต้อง (Accuracy) ในการพยากรณ์ดังแสดงในสมการที่ (6)

$$\text{ค่าความถูกต้อง} = \frac{\text{จำนวนข้อมูลที่ทำนายถูก}}{\text{จำนวนข้อมูลทั้งหมด}} * 100 \quad (6)$$

**ผลการวิจัย**

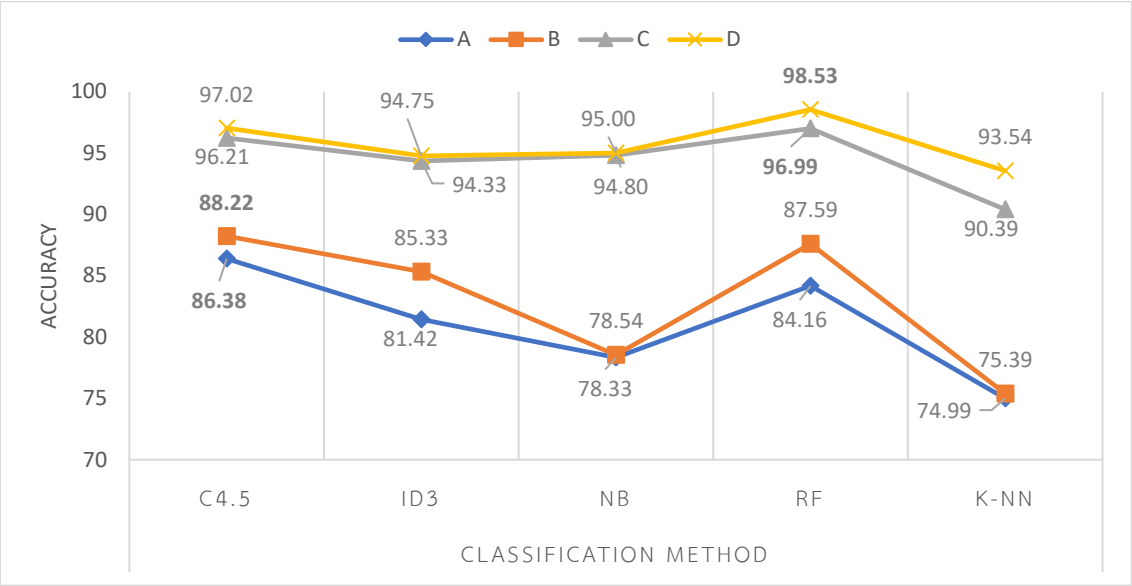
ผลการทดลองเพื่อประเมินประสิทธิภาพของแบบจำลองสำหรับการตัดสินใจรับซื้อโคขุน โดยใช้เทคนิคการทำเหมืองข้อมูล กรณีศึกษา กลุ่มวิสาหกิจชุมชนโคเนื้อบ้านบูเกะจือฆา สรุปได้ว่า ชุดการทดลอง D เป็นชุดข้อมูลที่ผ่านการกำจัดข้อมูลที่ผิดพลาด มีการคัดเลือกคุณลักษณะของข้อมูล และมีการแบ่งช่วงข้อมูล ได้ค่าความถูกต้องเฉลี่ยสูงสุด และรองลงมาเป็นชุดการทดลอง C, B และ A ตามลำดับ แสดงดังภาพที่ 5 โดยมีรายละเอียดดังนี้

ผลการทดลอง A สรุปได้ว่า C4.5 ได้ค่าความถูกต้องสูงสุดคือ 86.38% รองลงมาคือ RF, ID3, NB และ k-NN คือ 84.16%, 81.42%, 78.33% และ 74.99% ตามลำดับ

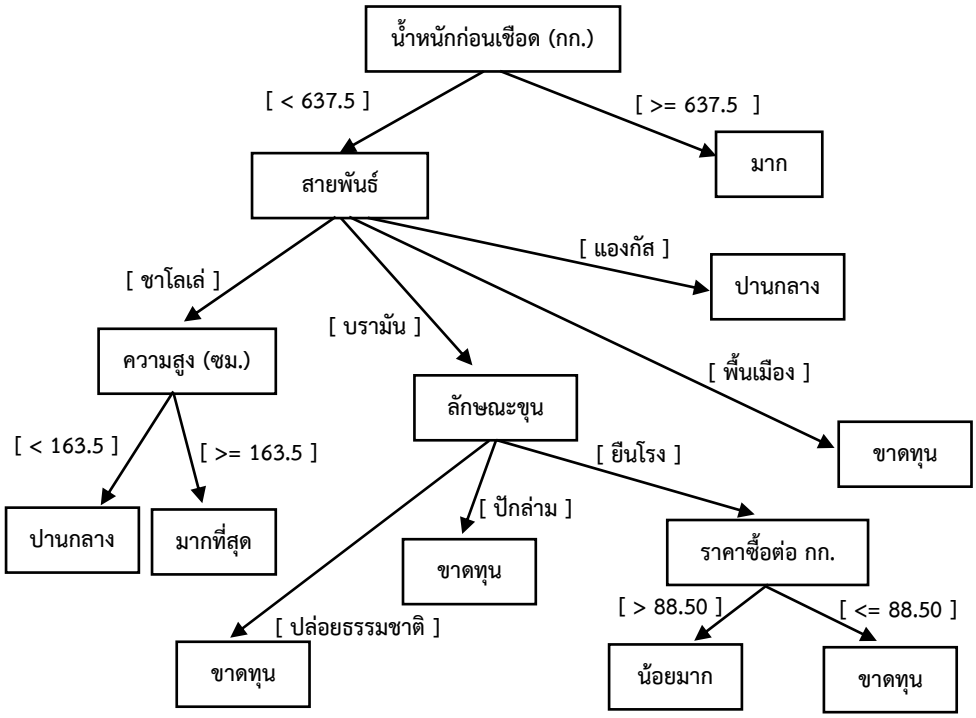
ผลการทดลอง B สรุปได้ว่า C4.5 ได้ค่าความถูกต้องสูงสุดคือ 88.22% รองลงมาคือ RF, ID3, NB และ k-NN คือ 87.59%, 85.33%, 78.54% และ 75.39% ตามลำดับ

ผลการทดลอง C สรุปได้ว่า RF ได้ค่าความถูกต้องสูงสุดคือ 96.99% รองลงมาคือ C4.5, NB, ID3 และ k-NN คือ 96.21%, 94.80%, 94.33% และ 90.39% ตามลำดับ

ผลการทดลอง D สรุปได้ว่า RF ได้ค่าความถูกต้องสูงสุดคือ 98.53% รองลงมาคือ C4.5, NB, ID3 และ k-NN คือ 97.20%, 95.00%, 94.75% และ 93.54% ตามลำดับ




**ภาพที่ 5** ผลการทดลองประสิทธิภาพของแบบจำลองสำหรับการตัดสินใจรับซื้อโคขุน โดยใช้เทคนิคการทำเหมืองข้อมูล




**ภาพที่ 6** กฎความสัมพันธ์ที่ได้จากต้นไม้ตัดสินใจ Random Forest (RF)

จากภาพที่ 6 ชุดทดลอบ D มีค่าความถูกต้องของแบบจำลองสูงที่สุด โดยใช้เทคนิคต้นไม้ตัดสินใจ RF และได้กฎความสัมพันธ์จำนวน 9 โดยมีรายละเอียดดังนี้

**R1:** ถ้าน้ำหนักก่อนเชือด { < 637.50} และสายพันธ์ {ชาโลเล่} และความสูง { < 163.5} แล้วทำนายว่าถ้าซื้อวัวกลุ่มนี้จะมีผลกำไรอยู่ในระดับ **{ปานกลาง}**

**R2:** ถ้าน้ำหนักก่อนเชือด { < 637.50} และสายพันธ์ {ชาโลเล่} และความสูง { >=163.5} แล้วทำนายว่าถ้าซื้อวัวกลุ่มนี้จะมีผลกำไรอยู่ในระดับ **{มากที่สุด}**

**R3:** ถ้าน้ำหนักก่อนเชือด { < 637.50} และสายพันธ์ {บรามัน} และลักษณะขุน {ปล่อยธรรมชาติ} แล้วทำนายว่า ถ้าซื้อวัวกลุ่มนี้จะมีผลกำไรอยู่ในระดับ **{ขาดทุน}**

**R4:** ถ้าน้ำหนักก่อนเชือด { < 637.50} และสายพันธ์ {บรามัน} และลักษณะขุน {ปักล่าม} แล้วทำนายว่าถ้าซื้อวัวกลุ่มนี้จะมีผลกำไรอยู่ในระดับ **{ขาดทุน}**

**R5:** ถ้าน้ำหนักก่อนเชือด { < 637.50} และสายพันธ์ {บรามัน} และลักษณะขุน {ยืนโรง} และราคาซื้อต่อกก. {>88.50} แล้วทำนายว่า ถ้าซื้อวัวกลุ่มนี้จะมีผลกำไรอยู่ในระดับ **{น้อยมาก}**

**R6:** ถ้าน้ำหนักก่อนเชือด { < 637.50} และสายพันธ์ {บรามัน} และลักษณะขุน {ยืนโรง} และราคาซื้อต่อกก. { <=88.50} แล้วทำนายว่า ถ้าซื้อวัวกลุ่มนี้จะมีผลกำไรอยู่ในระดับ **{ขาดทุน}**

**R7:** ถ้าน้ำหนักก่อนเชือด { < 637.50} และสายพันธ์ {พื้นเมือง} แล้วทำนายว่า ถ้าซื้อวัวกลุ่มนี้จะมีผลกำไรอยู่ในระดับ **{ขาดทุน}**

**R8:** ถ้าน้ำหนักก่อนเชือด { < 637.50} และสายพันธ์ {แองกัส} แล้วทำนายว่า ถ้าซื้อวัวกลุ่มนี้จะมีผลกำไรอยู่ในระดับ **{ปานกลาง}**

**R9:** ถ้าน้ำหนักก่อนเชือด { >= 637.50} ถ้าซื้อวัวกลุ่มนี้จะมีผลกำไรอยู่ในระดับ **{มาก}**

สำหรับค่าความแม่นยำ (class precision) สูงสุดจะอยู่ในส่วนของคลาส กำไรมากที่สุด และคลาสขาดทุน รองลงมาเป็น คลาสกำไรปานกลาง คลาสกำไรมาก และคลาสกำไรน้อยมาก ตามลำดับ ค่าระลึก (class recall) มีค่าสูงสุดคือคลาสกำไรมาก กำไรน้อยมาก และกำไรมากที่สุด รองลงมาเป็นคลาสกำไรปานกลาง และคลาสขาดทุน ตามลำดับ แสดงดังตารางที่ 6

**ตารางที่ 6** ค่าความแม่นยำ ค่าระลึกและค่าความถูกต้องของการเลือกคุณลักษณะด้วย Information Gain กับเทคนิคการจำแนกข้อมูลด้วย Random Forest

| | true มากที่สุด | true มาก | true ปานกลาง | true น้อยมาก | true ขาดทุน | class precision |
|---|---|---|---|---|---|---|
| **pred. มากที่สุด** | 52 | 0 | 0 | 0 | 0 | **100.00%** |
| **pred. มาก** | 0 | 161 | 0 | 0 | 0 | **100.00%** |
| **pred. ปานกลาง** | 0 | 1 | 132 | 0 | 1 | **98.51%** |
| **pred. น้อยมาก** | 0 | 0 | 0 | 31 | 2 | **93.94%** |
| **pred. ขาดทุน** | 0 | 0 | 0 | 2 | 26 | **92.86%** |
| **class recall** | **100.00%** | **99.38%** | **100.00%** | **93.94%** | **89.66%** | |

**สรุปผลการวิจัย/อภิปรายผล**

การพัฒนาแบบจำลองสำหรับการตัดสินใจรับซื้อโคขุน โดยใช้เทคนิคการทำเหมืองข้อมูล กรณีศึกษา กลุ่มวิสาหกิจชุมชนโคเนื้อบ้านบูเกะจือฆา สามารถสรุปได้ดังนี้

ประเด็นที่ 1) ชุดทดสอบที่ผ่านการกำจัดข้อมูลที่ผิดพลาด มีการคัดเลือกคุณลักษณะของข้อมูล และมีการแบ่งช่วงข้อมูล นั้นคือชุดทดสอบ D จะให้ค่าเฉลี่ยความถูกต้องสูงที่สุด ซึ่งสอดคล้องกับงานวิจัย เรื่อง การเพิ่มประสิทธิภาพการคัดเลือกคุณลักษณะข้อมูล สำหรับการจำแนกกลุ่มบทความวิจัย (ปะพาดา ณ วิเชียร และคณะ, 2563) ที่ใช้วิธีการคัดเลือกคุณลักษณะด้วยการประเมินประสิทธิภาพของแต่ละคุณลักษณะโดยการคำนวณหาค่าน้ำหนักหรือค่าความสัมพันธ์ระหว่างคุณลักษณะข้อมูล และคลาสด้วย Information Gain และงานวิจัยเรื่อง การเปรียบเทียบประสิทธิภาพของการจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ ที่ใช้เทคนิคการแบ่งช่วงข้อมูลที่แตกต่างกัน (สุระสิทธิ์ ทรงม้า, 2559) กรณีที่มีการกำหนดช่วงข้อมูล หรือค่า k น้อย วิธีการแบ่งช่วงข้อมูลด้วยความถี่ที่เท่ากัน จะได้ค่าความถูกต้องสูงกว่าวิธีการแบ่งช่วงข้อมูลด้วยขนาดความกว้างที่เท่ากัน ซึ่งผลการทดลองการแบ่งช่วงข้อมูลด้วยวิธีกำหนดความถี่ของข้อมูลที่เท่ากันกับชุดการทดลอง D โดยการกำหนดจำนวนช่วงข้อมูลเป็น 3 และ 2 ช่วงข้อมูล ผลการทดลองพบว่า การกำหนดช่วงข้อมูลจำนวน 2 ช่วงข้อมูลได้ค่าความถูกต้องที่สูงกว่าเมื่อเทียบกับการกำหนดช่วงข้อมูล จำนวน 3 ช่วงข้อมูล แสดงดังตารางที่ 7

**ตารางที่ 7** ผลการทดลองการแบ่งช่วงข้อมูลด้วยความถี่ที่เท่ากันของข้อมูลชุดการทดลอง D

| number of bins | classification method | | | | |
|---|---|---|---|---|---|
| | C4.5 | ID3 | NB | RF | k-NN |
| 3 | 96.89 | 94.65 | 94.30 | 97.11 | 92.93 |
| **2** | **97.02** | **94.75** | **95.00** | **98.53** | **93.54** |

ประเด็นที่ 2) แบบจำลองที่สร้างด้วยต้นไม้ตัดสินใจ Random Forest (RF) ให้ค่าเฉลี่ยความถูกต้องสูงสุด เมื่อเทียบกับต้นไม้ตัดสินใจอื่นๆ ซึ่งสอดคล้องกับงานวิจัยเรื่อง การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อทำนายการได้รับทุนการศึกษาของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคการทำเหมืองข้อมูล (พีรศุษม์ ทองพ่วง และจรัญ แสนราช, 2564)

ประเด็นที่ 3 การคัดเลือกคุณลักษณะข้อมูลที่สำคัญด้วย Information Gain เพื่อใช้ในการตัดสินใจรับซื้อโคขุนที่มีค่าน้ำหนักมากกว่า 0.5 มีจำนวน 10 คุณลักษณะคือ น้ำหนักก่อนเชือด ความยาวลำตัว ลักษณะฟัน ความสูง สายพันธ์ เครือข่ายผู้เลี้ยง ลักษณะการขุน ระยะเวลาการขุน อายุ และราคารับซื้อ มีผลทำให้ค่าความถูกต้องเฉลี่ยของชุดการทดลอง B ในการสร้างแบบจำลองเพิ่มขึ้น 11.53% เมื่อเทียบกับค่าความถูกต้องเฉลี่ยของ แบบจำลองที่ใช้ชุดข้อมูลการทดลอง C ที่ไม่ผ่านการคัดเลือกคุณลักษณะข้อมูลด้วย Information Gain

ประเด็นที่ 4) ได้แบบจำลองสำหรับการตัดสินใจรับซื้อโคขุน โดยใช้เทคนิคการทำเหมืองข้อมูลจำนวน 9 กฎ โดยได้กฎที่น่าสนใจกรณีที่จะทำให้ได้กำไรต่อตัวมากกว่า 7,501 บาท ก็คือกฎ R2 กรณีที่เกษตรกรเลือกซื้อโคขุนที่น้ำหนักน้อยกว่า 637.50 กิโลกรัม จะต้องเลือกสายพันธ์ลูกผสมชาโลเล่ ที่มีความสูงของลำตัวมากกว่า 163.5 เซนติเมตร หรือจะเป็นโคขุนสายพันธ์ใดก็ได้ ที่มีน้ำหนักเกิน 637.50 กิโลกรัม ก็จะมีกำไรต่อตัวอยู่ 5,001-7,500 บาท ตามกฎ R9 ซึ่งสามารถนำกฎที่ได้ไปสร้างเป็นแอพพลิเคชั่นสำหรับการตัดสินใจรับซื้อโคขุน เพื่อช่วยให้เกษตรกรที่ไม่มีความชำนาญการในเรื่องการวิเคราะห์คุณลักษณะเบื้องต้นของโคขุน และลดความเสี่ยงภาวะขาดทุนในการรับซื้อโคขุนได้อย่างมีประสิทธิภาพ

**เอกสารอ้างอิง**


กรมส่งเสริมการเกษตร. (2566). **รายงานสรุปประเภทกิจการวิสาหกิจชุมชน.** ค้นเมื่อ 2 กุมภาพันธ์ 2566 ค้นจาก https://smce.doae.go.th/ProductCategory/SmceCategory.php.

เกรียงศักดิ์ รักภักดี, ปราโมทย์ นามวงศ์, ไมตรี ริมทอง, และวชิระ โมราชาติ. (2560). **ระบบสนับสนุนการตัดสินใจเพื่อการเดินทางท่องเที่ยวในจังหวัดอุบลราชธานีผ่านระบบดาวเทียมบอกพิกัด**. วารสารวิจัยและพัฒนา มหาวิทยาลัยราชภัฏบุรีรัมย์ ปีที่ 12 ฉบับที่ 2 กรกฎาคม-ธันวาคม 2560.

ปะพาดา ณ วิเชียร, ภาคภูมิ มันแอ, ญาณพัฒน์ ชูชื่น, และสุภาวดี มากอ้น. (2563). **การเพิ่มประสิทธิภาพ Feature Selection สำหรับการจำแนกกลุ่มบทความวิจัย**. วารสารวิชาการวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสงขลา ปีที่ 1 ฉบับที่ 1 มกราคม-มิถุนายน 2563.

พีรศุษม์ ทองพ่วง และจรัญ แสนราช. (2564). การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อทำนายการได้รับทุนการศึกษาของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคการทำเหมืองข้อมูล. วารสาร journal of Professional Routine to Research Volume 8, กรกฎาคม - ธันวาคม 2564.

ภรัณยา ปาลวิสุทธิ์. (2559). **การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุลโดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อย สำหรับข้อมูลการเป็นโรคติดอินเทอร์เน็ต.** วารสารเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ ปีที่ 12 ฉบับที่ 1 มกราคม-มิถุนายน 2559.

วนัสดา สิริวุฒิชาภิรัชต์, สุนทรียา จอมผักแว่น, รัตนา ลีรุ่งนาวารัตน์, และพรทิพย์ เหลียวตระกูล. (2564). **ระบบสนับสนุนการตัดสินใจสำหรับการเลี้ยงโคเนื้อ** วารสารแม่โจ้เทคโนโลยีสารสนเทศและนวัตกรรม มหาวิทยาลัยแม่โจ้ ปีที่ 7 ฉบับที่ 2 กรกฎาคม-ธันวาคม 2564.

วัชรวิทย์ มีหนองใหญ่. (2557). ปัจจัยที่มีผลต่อความนุ่มของเนื้อโค. วารสารแก่นเกษตร มหาวิทยาลัยขอนแก่น ปีที่ 42 ฉบับที่ 3 ปี 2557.

สุระสิทธิ์ ทรงม้า. (2559). การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจรุ่นที่ 3 ที่ใช้เทคนิคการแบ่งช่วงข้อมูลที่แตกต่างกัน. วารสาร SDU RESEARCH JOURNAL OF SCIENCES AND TECHNOLOGY ปีที่ 8 ฉบับที่ 3 กันยายน – ธันวาคม 2559.

C. Kaewchinporn. (2010). **Data Classification with Decision Tree and Clustering Techniques**. Thesis in Computer Science, King Mongkut's Institute of Technology Ladkrabang, Thailand.

Cáceres, E. N., Pistori, H., Turine, M. A. S., Pires, P. P., Soares, C. O., & Carromeu, C. (2011). **Computational precision livestock- position paper**. In II Workshop of the Brazilian Institute for Web Science Research (No. 02-03, p. 9).

F. Duemong, L. Preechaveerakul and S. Vanichayobon. (2009). **FIAST : A Novel Algorithm for Mining Frequent Itemsets**. In Proceedings of 2009 International Conference on Future Computer and Communication, Kuala Lumpur, Malasia, pp.140-144.

Gonzalez-Araya and Marcela C. (2016). **Turning pork into profits in Chile: a decision support system for production planning in a Chilean swine slaughterhouse**. OR/MS Today, vol. 43, no. 2.

L. Breiman and R. Forests. (2001). **Machine Learning**, Vol. 45, No. 1, pp.5-32, 2001.

Liu, B., Hsu, W., and Ma, Y. (1998). **Integrating classification and association rule mining.** Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, p. 80-86.

R. Quinlan. (1986). **Induction of decision trees**. Machine Learning, Vpl. 1, No. 1, pp. 81-106.

Richard, J. R. and Geatz, M. W. (2003). **Data Mining a Tutorial-Based Primer**. Pearson Education Inc.

Wong A. K. C., and Chiu, D. K. Y. (1987). **Synthesizing Statistical Knowledge from Incomplete Mixed-Mode Data**. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 9, pp. 796-805.